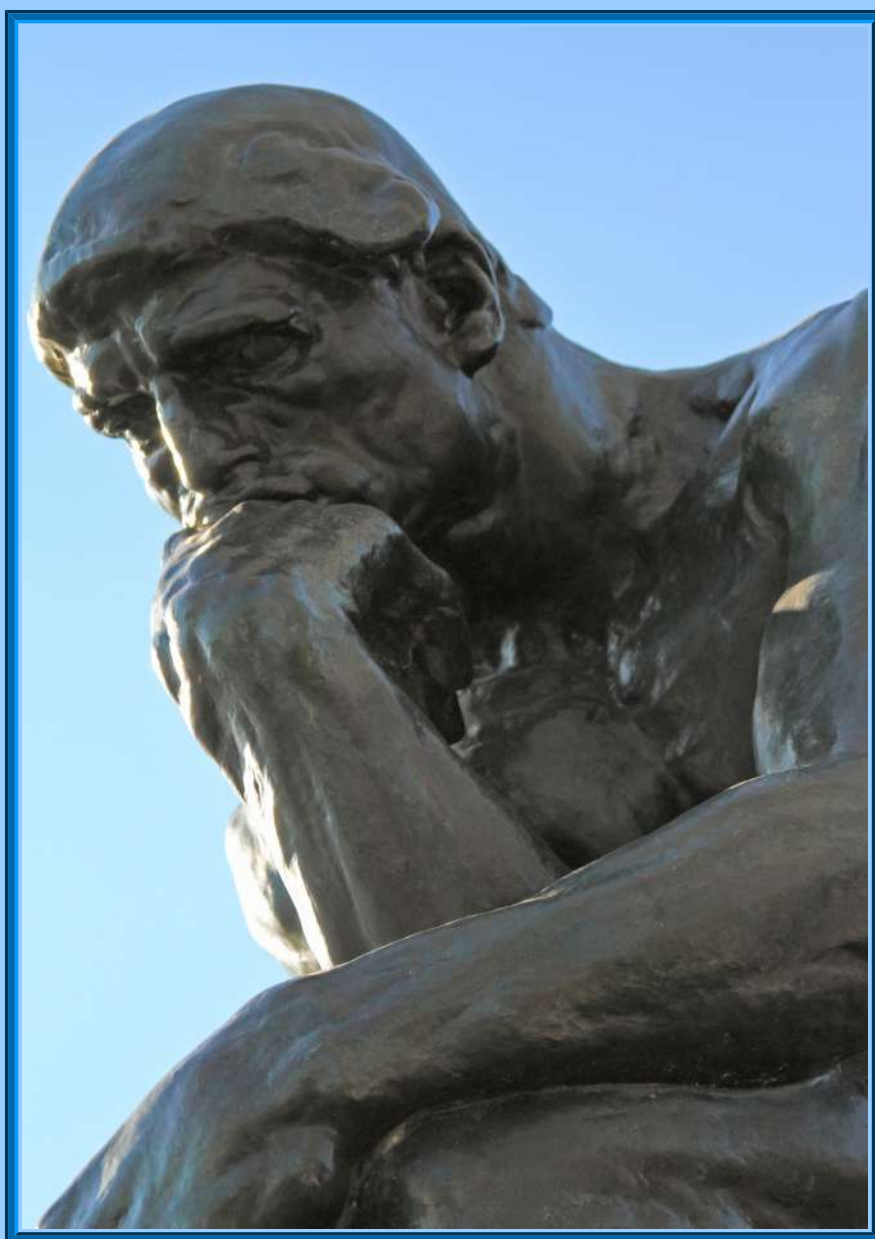


Т. Л. КАЦАДЗЕ

ЕКСПЕРТНІ СИСТЕМИ ПРИЙНЯТТЯ РІШЕНЬ В ЕНЕРГЕТИЦІ



Т. Л. КАЦАДЗЕ

**ЕКСПЕРТНІ СИСТЕМИ
ПРИЙНЯТТЯ РІШЕНЬ
В ЕНЕРГЕТИЦІ**

НАВЧАЛЬНИЙ ПОСІБНИК

КИЇВ — ЛОГОС — 2014

УДК 621.311:004.89(82)
ББК 31я73
К30

*Гриф надано Методичною радою НТУУ «КПІ»
(Протокол № 3 від 19.11.2013 р)*

Кацадзе Т. Л.

К30 Експертні системи прийняття рішень в енергетиці: навч. посіб. /
Т. Л. Кацадзе. – К.: ЛОГОС, 2014. – 173 с. – Бібліогр.: с. 167-173.
ISBN 978–966–171–768–7

В посібнику розглянуті основні положення застосування інтелектуальних технологій, зокрема експертних систем в галузі електроенергетики. Теоретичний матеріал проілюстрований прикладами розв'язання практичних електроенергетичних задач із застосуванням моделей та методів штучного інтелекту. Наведено питання для самоконтролю знань за тематикою розділів.

Рецензенти:

О. Ф. Буткевич – д-р техн. наук, проф.

П. Д. Лежнюк – д-р техн. наук, проф.

Відповідальний редактор:

В. М. Сулейманов – канд. техн. наук, проф.

УДК 621.311:004.89(82)
ББК 31я73

ISBN 978–966–171–768–7

© Т. Л. Кацадзе, 2014

ВСТУП.....	5
РОЗДІЛ 1. ЗАГАЛЬНІ ВІДОМОСТІ ПРО ЕКСПЕРТНІ СИСТЕМИ ПРИЙНЯТТЯ РІШЕНЬ.....	7
§1.1. Загальна характеристика задач прийняття рішень в процесі управління електроенергетичними системами.....	7
§1.2. Визначення експертної системи.....	12
§1.3. Класифікація експертних систем.....	17
§1.4. Узагальнена структура експертної системи.....	20
§1.5. Базові функції експертної системи.....	22
§1.6. Короткі характеристики деяких відомих експертних систем.....	26
РОЗДІЛ 2. МОДЕЛІ ФОРМАЛІЗАЦІЇ ТА РОЗВ'ЯЗАННЯ ПРАКТИЧНИХ ЗАДАЧ В СЕРЕДОВИЩІ ЕКСПЕРТНИХ СИСТЕМ.....	28
§2.1. Загальна характеристика способів формалізації практичних задач.....	28
§2.2. Загальна характеристика алгоритмів розв'язання задач у просторі станів.....	34
§2.3. Алгоритми неінформативного пошуку у просторі станів.....	36
§2.4. Алгоритми евристичного пошуку у просторі станів.....	44
§2.5. Стратегії пошуку у просторі станів.....	50
§2.6. Алгоритми пошуку у просторі задач.....	52
§2.6. Генетичні алгоритми пошуку.....	58
РОЗДІЛ 3. МОДЕЛІ ФОРМАЛІЗАЦІЇ ЗНАНЬ В ЕКСПЕРТНИХ СИСТЕМАХ.....	68
§3.1. Загальна характеристика задач інженерії знань.....	68
§3.2. Логічні моделі знань.....	76
§3.3. Продукційні моделі знань.....	85
§3.4. Семантичні мережі.....	90
§3.5. Скриптові моделі знань.....	95
§3.6. Фреймові моделі знань.....	96
РОЗДІЛ 4. ЛОГІЧНЕ ВИВЕДЕННЯ В УМОВАХ НЕВИЗНАЧЕНОСТІ.....	100
§4.1. Загальна характеристика проблеми невизначеності інформації.....	100
§4.2. Багатозначна логіка Лукасевича.....	103
§4.3. Байєсова логіка.....	105

§4.4. Апарат коефіцієнтів упевненості.....	110
§4.5. Теорія свідочств Демстера-Шефера.....	114
§4.6. Нечітка логіка.....	121
РОЗДІЛ 5. ШТУЧНІ НЕЙРОННІ МЕРЕЖІ.....	132
§5.1. Загальні відомості про штучні нейронні мережі.....	132
§5.2. Модель штучного нейрону.....	135
§5.3. Загальна характеристика архітектури штучних нейронних мереж.....	140
§5.4. Методи навчання штучних нейронних мереж.....	144
§5.5. Штучна нейронна мережа Хопфілда.....	153
§5.6. Нейромережеві експертні системи.....	161
ПРЕДМЕТНИЙ ПОКАЖЧИК.....	164
ЛІТЕРАТУРА.....	167

Підготовка сучасного спеціаліста в області виробництва, передавання та розподілу електричної енергії немислима без привиття йому навичок використання комп'ютеризованих інтелектуальних технологій розв'язання складних, погано формалізованих електроенергетичних задач. Саме тому навчальний план спеціальності «Електричні системи і мережі» передбачає вивчення дисципліни «Експертні системи прийняття рішень в енергетиці», основна мета якої полягає у наданні студентам уяви про основні принципи побудови експертних систем підтримки рішень з елементами штучного інтелекту. Інженер-електрик, що спеціалізується в галузі передачі та розподілу електричної енергії, повинен орієнтуватися в передових інтелектуальних методах аналізу процесів в енергетичних системах, які ґрунтуються на принципах штучного інтелекту, володіти базовими поняттями експертних систем, методами реалізації та технологіями побудови систем підтримки рішень. Під час вивчення дисципліни студент повинен отримати стійкі знання та навички, обсяг яких повинен бути достатнім для вирішення проектно-конструкторських задач, які виникають під час проектування та впровадження технологій експертних систем підтримки рішень.

Матеріал навчального посібника базується на лекційному курсі «Експертні системи прийняття рішень в енергетиці», який автор читає в НТУУ «КПІ» студентам спеціальності «Електричні системи і мережі» протягом більше 10 років. Головна мета, яку мав на увазі автор під час написання навчального посібника, полягає в ознайомленні студентів електроенергетичних спеціальностей, які не є фахівцями в області комп'ютерних наук, із можливими способами використання технологій штучного інтелекту, зокрема експертних систем, для розв'язання практичних електроенергетичних задач. Звісно, обмежений обсяг та специфіка цільової аудиторії читачів навчального посібника, не дозволяє повністю охопити весь спектр проблем, пов'язаних з розробкою та експлуатацією експертних систем прийняття рішень. Тут розглянуті лише загальні, часто у спрощеному вигляді, моделі та методи формування ланцюжків логічних міркувань, які використовують в технологіях експертних систем.

Для детального ознайомлення зі спеціальним матеріалом в області застосування штучного інтелекту для розв'язання практичних задач слід скористатися спеціальною літературою, список якої наведено наприкінці посібника. Список рекомендованої літератури умовно поділений на дві

частини – основну та додаткову літературу. До основного списку увійшли підручники, навчальні посібники та монографії, опубліковані протягом останніх 10-12 років, матеріали яких, в основному, були використані для підготовки лекцій та тексту навчального посібника. Уся інша література, у тому числі англomовна, увійшла до додаткового списку.

Матеріал посібника поділено на п'ять змістовних розділів, кожен з яких складається з декількох підрозділів. У першому розділі посібнику наведено загальну інформацію про використання технологій експертних систем для розв'язання практичних задач, зокрема в області електроенергетики. Тут наведено типову структуру експертної системи та охарактеризовано основні функції, які виконує експертна система в процесі свого функціонування. Другий розділ навчального посібника присвячений формальним моделям та методам розв'язання інтелектуальних задач в середовищі програм штучного інтелекту, зокрема в експертних системах. У третьому розділі розглянуті основні проблеми інженерії знань, пов'язані з використанням технологій експертних систем, наведені основні моделі представлення знань в інтелектуальних комп'ютерних системах та методи розв'язання практичних задач на їх основі. Четвертий розділ посібника присвячений розв'язанню задач логічного виведення в умовах неповноти або невизначеності вихідної інформації. В п'ятому розділі посібника дана коротка характеристика технологій штучних нейронних мереж та їх використання у складі експертних систем.

Матеріал навчального посібника містить приклади розв'язання практичних електроенергетичних задач із застосуванням моделей та методів штучного інтелекту, що полегшує сприйняття теоретичного матеріалу посібника. Наприкінці кожного підрозділу посібника наведені питання для самоконтролю та списки посилань на спеціальну літературу, де теоретичний матеріал відповідного підрозділу розглянуто більш детально.

Наприкінці навчального посібника представлений предметний покажчик, який містить список основних понять з області штучного інтелекту, експертних систем та інженерії знань, які використано у тексті посібника.

Автор висловлює щире подяку рецензентам д-ру техн. наук, проф. Буткевичу О. Ф., д-ру техн. наук, проф. Лежнюку П. Д., а також колективу кафедри електричних мереж та систем, зокрема проф. Сулейманову В. М. за ретельний аналіз тексту посібника, цінні зауваження та рекомендації щодо викладення матеріалу.

ЗАГАЛЬНІ ВІДОМОСТІ ПРО ЕКСПЕРТНІ СИСТЕМИ ПРИЙНЯТТЯ РІШЕНЬ

§1.1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ЗАДАЧ ПРИЙНЯТТЯ РІШЕНЬ В ПРОЦЕСІ УПРАВЛІННЯ ЕЛЕКТРОЕНЕРГЕТИЧНИМИ СИСТЕМАМИ

Головне призначення електроенергетичних систем полягає у забезпеченні надійного постачання споживачів електричною енергією нормованої якості та у необхідних обсягах за умови мінімально можливих витрат. Технологічні особливості електроенергетичних систем, зокрема тісний взаємний зв'язок між режимами окремих елементів, визначають необхідність організації єдиного централізованого управління. Тут, зазвичай, розрізняють три рівня управління енергетичною системою: оперативний, тактичний та стратегічний.

Оперативне управління енергетичною системою має місце у разі виникнення неузгодженості між програмою виробництва, передачі і розподілу електричної енергії та реальним режимом енергосистеми, відбувається у режимі реального часу та полягає у виробленні управляючих впливів, направлених на нормалізацію робочого режиму енергосистеми. Таке управління здійснюється автоматичними або автоматизованими засобами.

До автоматичних засобів відносять пристрої, які самостійно вироблюють управляючі дії у разі виходу контрольованих параметрів за допустимі межі відповідно до програми дій, закладених у логічному блоці цих пристроїв. До таких засобів відносять, наприклад, пристрої регулювання напруги без розриву кіл навантаження (РПН) силових трансформаторів, системи автоматичного збудження синхронних генераторів, пристрої автоматичного частотного розвантаження енергосистеми, струмового та диференційного захисту елементів електричних мереж тощо. До автоматичних можна також віднести системи управління, побудовані за сучасними технологіями «розумних мереж» (Smart Grid). Таке управління направлене на оптимальне регулювання процесів виробництва і споживання електричної енергії, в основному з метою оптимального використання потужностей відновлюваних джерел енергії, зокрема ві-

трових та сонячних електричних станцій. Зазначимо, що автоматичне регулювання режимів енергосистем носить, зазвичай, локальний характер і враховує обмежену кількість показників поточного стану енергосистеми.

В автоматизованих оперативних засобах управління у циклі вироблення управляючих дій завжди приймає участь людина, якій при цьому належить вирішальна роль. Таке управління режимами енергосистем називають диспетчерським, а всю людино-машинну систему управління — автоматизованою системою диспетчерського управління.

До функцій оперативно-диспетчерського управління відносять:

- регулювання частоти змінного струму та оптимальний розподіл активних потужностей між електричними станціями системи;
- регулювання режиму напруги в основних вузлових точках електричних мереж та оптимальний розподіл реактивних потужностей;
- встановлення, підтримка та оптимальний розподіл резерву активної потужності в енергосистемі;
- зміну робочої схеми основних електричних мереж та станцій з урахуванням технологічних обмежень;
- виведення в ремонт та увімкнення у роботу після ремонту найважливішого обладнання енергосистеми;
- запобігання та ліквідація системних аварій в енергосистемі;
- ліквідацію великих системних аварій із втратою електропостачання частини споживачів.

Тактичне управління полягає у вирішенні задачі планування режимів роботи електроенергетичних систем, тобто у виборі заходів та послідовності їх реалізації для забезпечення нормального функціонування системи на найближчу та віддалену перспективу. Планування (прогнозування) режимів енергосистеми здійснюють для таких періодів часу: півгодини, одна година, дві години та більше, одна доба, один місяць, квартал, рік, 5 років.

Планування перспективних режимів енергосистем полягає в наступному:

- визначенні прогнозованих сумарних графіків навантаження енергосистеми;
- формуванні балансу енергії та визначення складу генеруючих потужностей для покриття навантаження з урахуванням обмежень за готовністю силового обладнання електричних станцій, наявністю енергоресурсів, нормами витрат гідроресурсів, техніко-економічними характеристиками силового обладнання;
- формуванні робочої схеми електричних з'єднань елементів електричних мереж з урахуванням обмежень за пропускною здат-

ністю, максимальним та аварійно допустимим навантаженням устаткування;

- визначенні параметрів робочих режимів енергосистем, які необхідно підтримувати протягом усього інтервалу планування;
- виробленні вказівок щодо налагодження пристроїв релейного захисту, режимної та противоаварійної автоматики з урахуванням характеристик перспективних режимів енергосистем;
- формуванні вказівок щодо дій диспетчерів та чергового персоналу під час ведення перспективних режимів енергосистем.

Стратегічне управління полягає у розв'язанні задачі планування і проектування розвитку енергосистем, що передбачає прогнозування розвитку енергетичної галузі у тісному зв'язку з розвитком інших галузей економіки на віддалену перспективу 10-15 років і більше. Тут необхідно передбачити зміну попиту на електричну енергію, тенденції перерозподілу використання енергетичних ресурсів різних типів, розвиток зовнішніх зв'язків з енергосистемами суміжних держав, можливості застосування новітніх технологій в енергетиці тощо.

Метою розв'язання задач прогнозування та проектування розвитку енергетичних систем є дослідження динаміки зміни максимумів та мінімумів навантажень енергосистеми, визначення та обґрунтування необхідного енергетичного будівництва, зокрема реконструкції існуючих та будівництва нових електричних станцій, підстанцій та ліній електропередачі тощо.

Звісно, що наведені вище задачі не охоплюють всього спектру завдань, пов'язаних з управлінням енергетичними системами, проте вони представляють найбільш характерні задачі, які виникають під час експлуатації електроенергетичних систем.

Сучасні електроенергетичні системи відносяться до класу великих людино-машинних систем кібернетичного типу, оскільки характеризуються великою кількістю регульованих параметрів. Це означає, що задачі управління енергосистемами є нетривіальними і у більшості практичних випадків визначити оптимальний розв'язок таких задач неможливо. В умовах оперативного управління режимами енергосистем це пов'язано із жорстким обмеженням часу для прийняття рішення, а під час прогнозування перспективних режимів та проектування розвитку енергосистем – із невизначеністю та неоднозначністю вихідної інформації, наприклад, даних про перспективні навантаження енергосистеми.

Таким чином, під час управління електроенергетичними системами зазвичай мають місце лише прийнятні рішення, якість яких, переважно, залежить від кваліфікації персоналу, який приймає управляючі рішення.

Очевидно, що задачі управління електроенергетичними системами, так само, як і будь-якою іншою складною технічною, соціальною або природною системою, є комплексними. У загальному випадку процес управління складними системами складається із виконання наступних етапів:

1) етап аналізу станів контрольованої системи, що полягає у виявленні неузгодженості між бажаним та реальним станами системи;

2) етап виявлення цілей та критеріїв ефективності, що полягає у виявленні необхідності зміни або збереження поточного стану системи, тобто у виявленні цілі, яку необхідно досягнути в процесі управління;

3) етап вироблення рішень, що полягає у виявленні можливих способів або шляхів досягнення цілі та переходів у цільові стани, тобто формуванні множини альтернативних рішень задачі управління;

4) етап прийняття рішень, що полягає у виборі найкращого, найбільш ефективного способу досягнення цільового стану серед множини альтернатив;

5) етап реалізації рішення, що полягає у зміні стану або траєкторії руху контрольованої системи;

6) етап оцінювання результатів управління, що полягає у визначенні ефективності прийнятих управляючих дій і, у разі необхідності, корекції прийнятого рішення.

Таким чином, управління електроенергетичними системами пов'язано із виробленням управляючих дій, направлених на корекцію руху системи у просторі станів з метою досягнення поставлених цілей функціонування системи. Тут на кожному етапі управління необхідно здійснювати вибір управляючих дій серед множини можливих варіантів.

Задачею прийняття рішень називають задачу вибору способу дії серед множини допустимих альтернатив під час управління складною системою

Задачі прийняття рішень відносять до найбільш складних та відповідальних задач з області прикладного системного аналізу. Розв'язання таких задач пов'язано із управлінням складними технічними або природними системами і часто ускладнюється жорстким обмеженням часу та недостатністю або ненадійністю вихідної інформації.

Ефективність розв'язання задач прийняття рішень багато в чому визначається кваліфікацією особи, яка приймає рішення, її досвідом у розв'язанні подібних задач. Найбільш кваліфікованих фахівців у певній предметній області називають експертами. Їх знання та досвід дають змогу ефективно розв'язувати практичні задачі в області прийняття рішень. Разом з тим підготовка таких висококваліфікованих фахівців для обслу-

говування сучасних надскладних технічних систем потребує дуже багато часу та витрат ресурсів, передбачає наявність певних психофізичних властивостей у кандидатів тощо. Все це визначає обмежену кількість висококваліфікованих експертів, здатних приймати найбільш ефективні рішення в процесі вироблення управляючих дій.

Одним з перспективних шляхів підвищення ефективності прийняття рішень під час управління великими системами кібернетичного типу є впровадження технологій комп'ютерних експертних систем. Такі системи виконують роль експертів у певній предметній області та призначені для підтримки рішень осіб під час вироблення управляючих дій.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть типи задач управління електроенергетичними системами.
2. Дайте характеристику автоматичних засобів оперативного управління режимами енергосистем.
3. Дайте характеристику задач оперативно-диспетчерського управління режимами енергосистем.
4. Дайте характеристику задач планування режимів роботи енергосистем на найближчу та віддалену перспективу.
5. Дайте характеристику задач планування та проектування розвитку енергосистем.
6. Поясніть положення про практичну неможливість отримання оптимальних рішень в задачах управління енергосистемами.
7. Дайте характеристику основних етапів управління складними системами.
8. Дайте визначення та охарактеризуйте задачу прийняття рішень.



ЛІТЕРАТУРА

[14], [19], [21], [24], [52], [74], [75].

§1.2. ВИЗНАЧЕННЯ ЕКСПЕРТНОЇ СИСТЕМИ

Досвід розробки та експлуатації комп'ютерних обчислювальних систем різного призначення показав, що їх ефективність визначається не лише якістю закладених алгоритмів, але й обсягом знань, які вони використовують під час свого функціонування. Реалізація даної концепції привела до розвитку спеціалізованих програмних систем, які є, свого роду, «експертами» у певній предметній галузі — експертних систем.

Експертні системи являють собою інтелектуальні програмні засоби, які, внаслідок інтерактивної взаємодії з користувачем, здатні отримувати, акумулювати та коригувати знання у певній предметній області, виводити нові знання, розв'язувати на основі цих знань практичні задачі та пояснювати спосіб їх розв'язання

Найважливіша ознака експертних систем полягає в орієнтації на процеси накопичення та організації знань. Стратегії експертних систем базуються на використанні високоякісних елітних знань людей — висококваліфікованих спеціалістів-експертів у певній предметній області. Такі знання сконцентровані в базах знань експертних систем та використовуються для розв'язання поточних практичних задач. Таким чином, експертні системи виконують роль асистента-радника для особи, яка приймає рішення в процесі управління або обслуговування складних технічних, соціальних або природних систем, у тому числі — електроенергетичних.

Перелік типових задач, які можна розв'язувати за допомогою експертних систем включає інтерпретацію, діагностику, прогнозування, моніторинг тощо. Такі завдання можна вирішувати не тільки засобами експертних систем, але і за допомогою традиційних алгоритмічних прикладних програм. Проте експертні системи характеризуються низкою суттєвих особливостей, за якими вони відрізняються від традиційних алгоритмічних програм (див. табл. 1.1):

1. Експертні системи моделюють не стільки фізичну (або іншу) природу предметної області, скільки механізм мислення людини щодо розв'язання задач в цій області. Це, звісно, не означає, що експертні системи повністю відтворюють психофізичну модель мислення людини, але основна увага тут приділяється відтворенню методики розв'язання проблем, якою користується експерт у певній предметній області.

2. Для розв'язання задач експертна система використовує знання експертів, зосереджені в базах знань. Знання зберігаються окремо від програмного коду, що дозволяє будувати гнучкі експертні системи, не «прив'язані» до конкретних прикладних задач.

3. Для розв'язання задач основними є евристичні наближені методи. Евристики являють собою емпіричні знання, придбані людиною під час накопичення практичного досвіду розв'язання подібних задач, але не завжди підкріплені суворим математичним обґрунтуванням. Евристичні методи розв'язання задач, на відміну від алгоритмічних, не завжди гарантують отримання результату. Разом з тим вони не потребують вичерпної вихідної інформації про об'єкти дослідження і дають змогу знайти прийнятне рішення навіть за умови браку вихідної інформації, що неможливе за використання алгоритмічних моделей.

Таблиця 1.1. Зіставлення характеристик алгоритмічних програм та експертних систем

АЛГОРИТМІЧНІ ПРОГРАМИ	ЕКСПЕРТНІ СИСТЕМИ
Моделювання фізичної природи предметної області	Моделювання механізму мислення людини щодо розв'язання практичних задач у предметній області
Алгоритми розв'язання задач є складовою частиною програмного коду	Алгоритми розв'язання задач містяться у базі знань та зберігаються окремо від програмного коду
Використання детермінованих моделей даних та точних методів розв'язання задач	Використання евристичних, наближених методів розв'язання задач

Перша з перелічених особливостей позиціонує експертні системи як складову області штучного інтелекту, тобто області створення програмних засобів моделювання розумової діяльності людини. Однак слід виділити низку особливостей експертних систем, які виділяють експертні системи з сукупності програм штучного інтелекту (див. табл. 1.2):

1. Експертні системи мають яскраво виражену практичну направленість у певній комерційній або науковій сфері. Вони завжди пов'язані з об'єктами реального світу, операції з якими потребують наявності суттєвого досвіду, накопиченого людиною.

На відміну, програми штучного інтелекту є дослідницькими і основна увага в них приділяється абстрактним математичним проблемам, або спрощеним варіантам реальних завдань. Метою функціонування таких систем є не отримання розв'язку, як в експертній системі, а розробка методики розв'язання задачі.

2. Однією з основних характеристик експертних систем є її продуктивність, тобто швидкість отримання результату та його достовірність. Це означає, що протягом прийняттого часу експертна система має знайти вирішення проблеми, не гірше за те, що може запропонувати спеціаліст з предметної області.

Разом з тим програми штучного інтелекту є дослідницькими та не виставляють жорстких вимог за часом отримання результату, а часто – і за точністю отриманого результату.

3. Експертні системи повинні обґрунтовувати отримані результати. Оскільки експертні системи проектуються для взаємодії з різними користувачами, то в них має бути закладений механізм пояснення способу розв'язання задачі, тобто експертна система має бути, по можливості, прозорою.

Дослідницькі програми штучного інтелекту, навпроти, взаємодіють лише зі своїм розробником, який і так знає, на чому базується отримуваний результат.

Таблиця 1.2. Зіставлення характеристик експертних систем та програм штучного інтелекту

ЕКСПЕРТНІ СИСТЕМИ	ПРОГРАМИ ШТУЧНОГО ІНТЕЛЕКТУ
Пов'язані з об'єктами реального світу та направлені на розв'язання практичних задач із використанням моделей та методів штучного інтелекту	Направлені на розв'язання абстрактних математичних задач з метою дослідження та відпрацювання моделей штучного інтелекту
Висока продуктивність, тобто швидкість отримання результату та його достовірність	Швидкість, а інколи – й достовірність отримання результату, не мають значення
Наявність механізму пояснення способу отримання результату	Відсутня потреба пояснення способу отримання результату

Таким чином, основними відмінними ознаками експертних систем є:

1. Розв'язання складно формалізованих практично важливих задач у певній предметній області.

2. Наявність спеціальних експертних знань в даній проблемній області.

3. Використання для розв'язання задач підходів та методів, які застосовує людина-експерт в реальних ситуаціях.

4. Наявність можливості пояснення шляху отримання рішення.

5. Застосування методів та моделей штучного інтелекту для розв'язання задач.

В табл. 1.3 наведено порівняльні властивості прикладних задач, які визначають доцільність застосування технологій експертних систем для їх розв'язання.

Таблиця 1.3. Зіставлення характеристик прикладних задач, які визначають доцільність використання експертних систем

ДОЦІЛЬНО	НЕ ДОЦІЛЬНО
Наявність евристичних методів розв'язання задач разом із повною або частковою відсутністю строгих алгоритмів	Наявність ефективних алгоритмічних методів розв'язання задач
Наявність експертних знань у предметній області	Експертні знання відсутні, або доступ до них обмежений
Задачі носять оцінювальний характер	Задачі носять розрахунковий характер
Розв'язання задач в умовах невідомості вихідної інформації	Наявність «чітких» даних та алгоритмів
Можливість розв'язання задач методами формальних міркувань	Для розв'язання задач застосовують алгоритмічні методи
Знання статичні	Знання динамічні

Слід зазначити, що на сьогодні експертні системи не знайшли такого широкого застосування, як прогнозувалося на початку їх використання. Це пояснюється рядом об'єктивних причин:

1. Більшість експертних систем потребує спеціального апаратного забезпечення (платформи). Перенесення їх на універсальні платформи призводить до суттєвого зниження продуктивності.

2. Громіздкі експертні системи є дуже складними в управлінні. Якщо їх бази знань містять десятки тисяч правил, а для їх управління потрібно декілька операторів, то їх не тільки складно адаптувати до конкретних ситуацій та організацій, але й дуже складно опанувати їх структуру.

3. Повільність деяких експертних систем не дає змоги використовувати їх для розв'язання практичних задач.

4. Впровадження експертних систем зазвичай потребує певних організаційних змін.

Разом з тим, до основних переваг використання експертних систем можна віднести такі їх властивості.

1. Експертні системи забезпечують зберігання експертних знань висококваліфікованих фахівців після зміни місця роботи, виходу на пенсію, смерті.

2. Вони забезпечують можливість акумуляції експертних знань, отриманих з декількох джерел. За допомогою експертних систем можна зібрати та використовувати знання багатьох експертів у предметній області, що суттєво підвищує ефективність розв'язання практичних задач.

3. Ефективність роботи експертної системи не залежить від психофізичного та емоційного стану людини-експерта, що виключає можливість помилки через втому, недостатню увагу та інші людські фактори та забезпечує підвищену надійність та якість отримання розв'язку.

4. Для забезпечення доступу до експертних знань можна використовувати будь-які наявні комп'ютерні та апаратні засоби, що визначає підвищену доступність інформації.

5. Експертні системи забезпечують зниження витрат на отримання експертних знань під час розв'язання практичних задач.

6. Експертні системи надають можливості докладного пояснення шляху отримання результату, використаних під час розв'язання задачі експертних знань тощо.

7. Такі технології забезпечують можливість використання експертної системи як інтелектуальної навчальної програми шляхом діагностики знань учнів та навчання на прикладах розв'язання складних задач предметної області.

8. Експертні системи можна використовувати для організації доступу до баз даних із використанням алгоритмів штучного інтелекту, що суттєво підвищує ефективність роботи з базами даних.

Проте, незважаючи на певні недоліки, експертні системи знайшли широке застосування в багатьох областях людської діяльності, зокрема, в енергетиці, медицині, високомолекулярній хімії тощо, та стали незамінними внаслідок їх здатності накопичувати та представляти у зручній для сприйняття формі вузькоспеціалізовані знання в конкретній предметній області, якими володіє вкрай обмежена кількість людей-експертів.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Дайте визначення експертної системи.

2. Наведіть основні відмінні риси між експертними системами та алгоритмічними програмами.

3. Наведіть основні відмінні риси між експертними системами та програмами штучного інтелекту.

4. Наведіть відмінні ознаки експертних систем.
5. Наведіть основні причини недостатнього поширення експертних систем.
6. Наведіть основні переваги використання експертних систем.



ЛІТЕРАТУРА

[2], [4], [5], [10], [11], [13], [14], [17], [31], [55], [63], [69].

§1.3. КЛАСИФІКАЦІЯ ЕКСПЕРТНИХ СИСТЕМ

На сьогодні експертні системи являють собою широкий спектр програмних комплексів, які можна класифікувати за різними ознаками. Найбільш корисними можуть виявитися способи класифікації, представлені на рис. 1.1.

Класифікація за типом задачі

Експертні системи орієнтовані на задачі, які важко розв'язувати за допомогою традиційних математичних методів. Перелік типових задач, для розв'язання яких залучають засоби експертних систем містить:

- *інтерпретація* – витягнення узгодженої інформації з первинних даних;
- *діагностика* – виявлення несправностей і причин їх появи в деякій системі;
- *моніторинг* – безперервна інтерпретація даних в режимі реального часу із сигналізацією виходу контрольованих параметрів за допустимі межі;
- *прогнозування* – передбачення ймовірних наслідків на основі минулих і поточних подій;
- *планування* – визначення послідовності дій, направлених на досягнення попередньо поставлених цілей;
- *проектування* – визначення конфігурації технічної системи за заданих обмежень;
- *налагодження та ремонт* – визначення послідовності дій з приведення системи до необхідних режимів функціонування;
- *навчання* – інтерпретація, діагностика і корекція знань та умінь учнів;
- *управління* – формування керуючих впливів, які визначають процеси функціонування складних технічних систем;

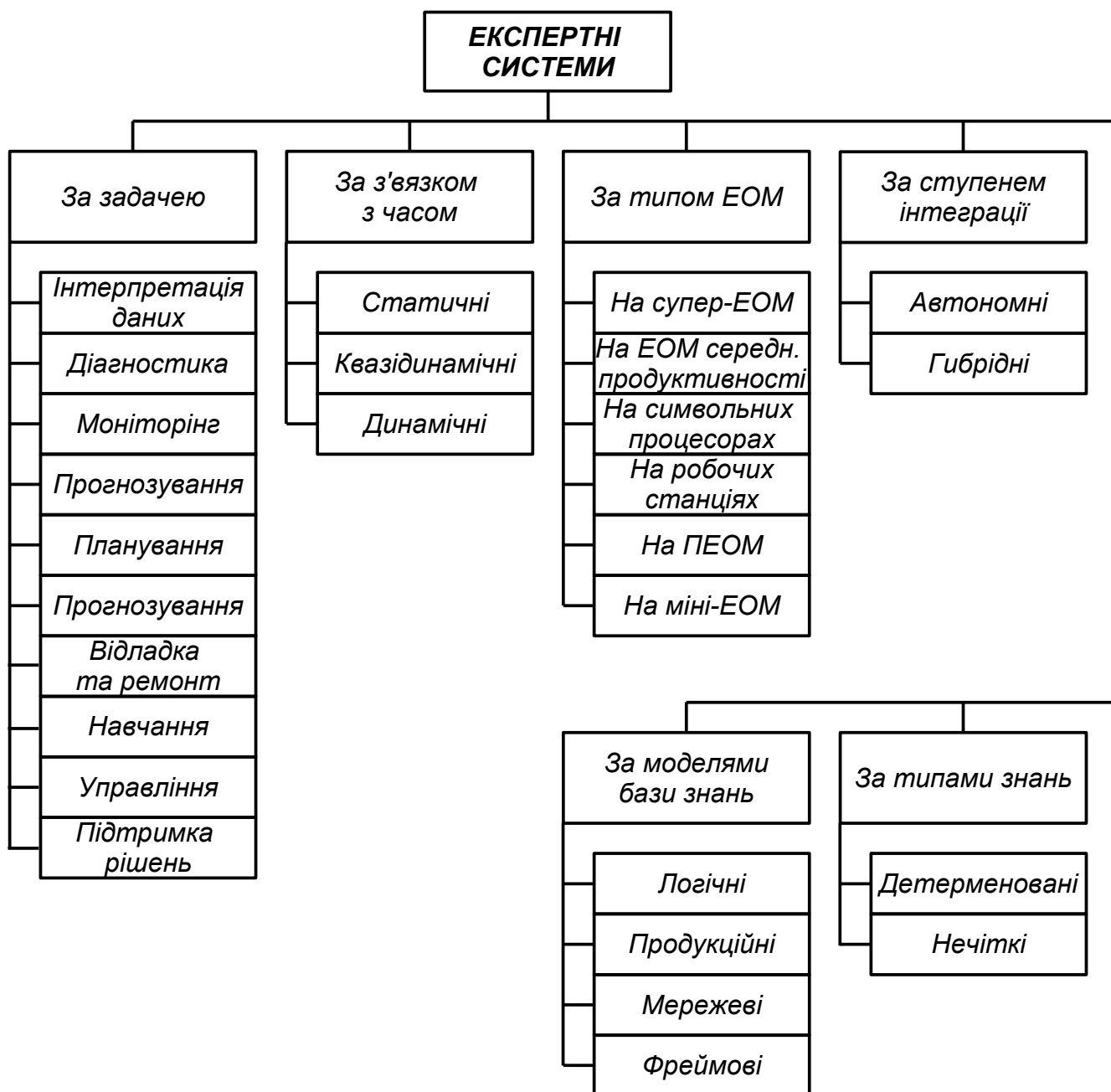


Рис. 1.1. Класифікація експертних систем

- *підтримка прийняття рішень* – забезпечення особи, яка приймає рішення, необхідною інформацією та рекомендаціями для полегшення процесу прийняття рішень.

Класифікація за зв'язком із реальним часом

- статичні експертні системи передбачають незмінність часу, статичність моделі предметної області;
- динамічні експертні системи функціують у сполученні із датчиками об'єктів у режимі реального часу з безперервною інтерпретацією даних, які поступають в систему;

- квазідинамічні експертні системи інтерпретують ситуацію, яка змінюється відповідно до внутрішнього таймеру.

Класифікація за типом ЕОМ

Експертні системи є унікальними програмними продуктами та їх продуктивність тісно пов'язана із ефективністю взаємодії із апаратною частиною обчислювальної техніки. Тут розрізняють такі типи експертних систем:

- експертні системи на супер-ЕОМ для унікальних, стратегічно важливих задач (CRAY, CONVEX тощо);
- експертні системи на ЕОМ середньої продуктивності (mainframe);
- експертні системи на символьних процесорах та робочих станціях (SUN, Silicon Graphics, APOLLO);
- експертні системи на персональних комп'ютерах (IBM PC, MAC);
- експертні системи на міні- та суперміні-ЕОМ (VAX, micro-VAX).

Класифікація за ступенем інтеграції з іншими програмами

- автономні експертні системи функціонують безпосередньо в режимі консультацій з користувачем для “виключно експертних” задач, розв'язання яких не вимагає залучення традиційних методів обробки інформації (розрахунки, моделювання тощо);
- гібридні експертні системи агрегують традиційні пакети прикладних програм та засоби обробки експертних знань.

Класифікація за моделями баз знань

- логічні експертні системи, бази знань та машини логічного виведення яких побудовані на принципах формальної логіки, зазвичай на базі апарату числення предикатів;
- продукційні експертні системи у своїй роботі використовують апарат продукцій, тобто правил типу «Якщо ... То ...»;
- мережеві експертні системи для зберігання та обробки знань базуються на апараті семантичних мереж;
- фреймові експертні системи для роботи зі знаннями використовують апарат фреймів – об'єктозорієнтованих інформаційних структур.

Класифікація за видами знань

- експертні системи із детермінованими знаннями передбачають жорстке точне визначення усієї вихідної інформації то точного визначення розв'язку практичної задачі;
- експертні системи із невизначеними (нечіткими) знаннями передбачають розв'язання практичних задач за умов неповноти, недовіри, невизначеності або нечіткості вихідної інформації.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Охарактеризуйте основні підходи до класифікації експертних систем.
2. Дайте характеристику типових задач, для розв'язання яких залучають експертні системи.
3. Поясніть класифікацію експертних систем за зв'язком із реальним часом.
4. Наведіть приклади статичних, динамічних та квазідинамічних експертних електроенергетичних задач.
5. Поясніть класифікацію експертних систем за ступенем інтеграції з іншими програмними засобами.
6. Наведіть приклади електроенергетичних задач, для розв'язання яких можна залучати автономні та гібридні експертні системи.



ЛІТЕРАТУРА

[7], [9], [10], [11], [13], [14], [31], [45], [62], [64]

§1.4. УЗАГАЛЬНЕНА СТРУКТУРА ЕКСПЕРТНОЇ СИСТЕМИ

Структуру експертної системи у загальному вигляді можна представити діаграмою, представленою на рис. 1.2. Ядро експертної системи утворюють база знань, машина логічного виводу та підсистема пояснень.

База знань – сукупність експертних знань предметної області експертної системи, формально поданих таким чином, щоб за їх допомогою можна було б реалізовувати логічні міркування.

Машина логічного виведення (підсистема пошуку рішень) – сукупність програмних засобів, призначених для моделювання ходу міркувань експертів в процесі розв'язання задачі з використанням інформації, яка зберігається у базі знань.

Підсистема пояснень – сукупність програмних засобів, призначених для відстежування роботи машини логічного виведення, що надає змогу користувачу з'ясувати, яким чином отримано розв'язок задачі та які знання були при цьому залучені.

Експертні системи завжди працюють в інтерактивному режимі з людиною. Тому в циклі функціонування експертної системи обов'язково необхідно враховувати людський фактор. Тут розрізняють три основних

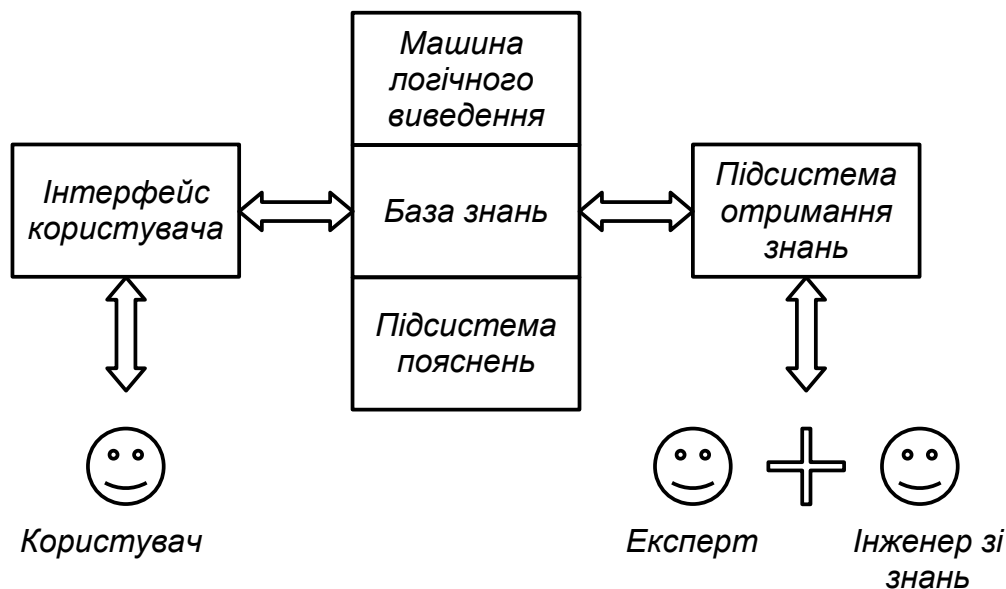


Рис. 1.2. Структура експертної системи

типи людей, які взаємодіють з експертною системою – експерт, інженер зі знань та користувач.

Експерт – висококваліфікований спеціаліст в предметній області, знання та практичний досвід якого використовують для наповнення бази знань експертної системи.

Інженер зі знань – спеціаліст в області штучного інтелекту, який є посередником між експертом та базою знань експертної системи.

Користувач – спеціаліст предметної області, для професійної діяльності якого призначена експертна система. Зазвичай кваліфікація користувача недостатньо висока і тому йому потрібна допомога експерта або експертної системи.

Взаємодія людини та експертної системи здійснюється за допомогою спеціальних програмних засобів забезпечення інтерактивного функціонування системи. До таких засобів відносять підсистему отримання знань та інтерфейс користувача.

Підсистема отримання знань призначена для автоматизації процесу наповнення, контролю та корекції бази знань експертної системи.

Інтерфейс користувача – комплекс програмних засобів, які реалізують інтерактивну взаємодію користувача і експертної системи на стадії введення вихідної інформації та на стадії отримання результату.

З діаграми на рис. 1.2 видно, що експертна система може функціонувати в двох режимах – в режимі отримання знань та в режимі розв'язання задач. В першому разі експерт взаємодіє із експертною системою за допомогою інженера зі знань (правий тракт структурної схеми рис. 1.2). В другому разі в режимі розв'язання задач із експертною системою взаємодіє користувач, якого цікавить результат та спосіб його отримання (лівий тракт структурної схеми рис. 1.2). Експертні системи, на відміну від

алгоритмічних систем, не виключають користувача з процесу розв'язання задач, а навпаки, залишають за ним ініціативу. Разом з тим експертні системи не є пасивними джерелами необхідної інформації, а активно підказують можливі шляхи розв'язання задачі, розвивають ланцюги міркувань, пояснюють свої дії.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть типову структуру експертної системи.
2. Охарактеризуйте основні структурні елементи експертної системи.
3. Охарактеризуйте можливі режими роботи експертної системи.



ЛІТЕРАТУРА

[1], [4], [7], [9], [10], [11], [13], [14], [15], [31], [45], [62], [63], [68].

§1.5. БАЗОВІ ФУНКЦІЇ ЕКСПЕРТНОЇ СИСТЕМИ

До основних функцій експертних систем відносять:

- 1)отримання знань;
- 2)представлення знань;
- 3)управління процесом пошуку рішень;
- 4)пояснення прийнятого рішення.

Отримання знань

Під отриманням знань слід розуміти процес передавання потенційного досвіду розв'язання практичних задач від експертів предметної області і перетворення його до вигляду, який дає змогу використовувати ці знання в експертній системі.

Найбільш поширеним методом передавання знань є співбесіди між спеціалістами з проектування експертних систем (інженером зі знань) та експертом з предметної області. Слід зазначити, що процес передавання знань є самим “вузьким” місцем у створенні промислових експертних систем, оскільки характеризується вкрай малою продуктивністю. Це пояснюється такими причинами:

- спеціалісти у вузькій предметній області, зазвичай, користуються власним жаргоном, що суттєво ускладнює процес передавання корисної інформації від експерта до інженера зі знань;

- факти та принципи, на яких базуються специфічні області знань, знання експертів не завжди можна чітко сформулювати в термінах математичної теорії або детермінованої моделі. Це потребує використання евристичних правил в базі знань. При цьому адекватно сформулювати евристичні правила, зазвичай, значно складніше, ніж традиційні логічні моделі;
- для розв'язання проблем у предметній області експерту, зазвичай, недостатньо мати певну суму знань про факти та принципи цієї області. Необхідно також знати, якого виду інформацією необхідно володіти для формування того або іншого міркування, чи надійні джерела інформації та яким чином можна поділити складну проблему на більш прості задачі. Виявити під час співбесіди знання такого типу, які базуються на власному досвіді експерта та погано піддаються формалізації, значно складніше, ніж отримати перелік певних фактів та загальних принципів;
- експертний аналіз, який виконує людина, зазвичай необхідно вмістити у доволі широкий контекст, який містить багато фактів, які здаються експерту само собою зрозумілими, проте для стороннього в жодному разі такими не є. Досить складно відокремити певну кількість та природу знань, які мають бути залучені в контекст задач певної предметної області.

Загалом, теорія витягання знань охоплює широкий спектр задач, які визначаються психологічними, лінгвістичними та гносеологічними аспектами тощо.

Формалізація знань

Проблема формалізації знань охоплює методи асоціативного зберігання інформації, на кшталт тих, що використовує людина в процесі розумової діяльності. При цьому основна увага приділяється логічній, а не біологічній стороні процесу. Тут основним є формування методів формального подання великих обсягів корисної інформації з метою подальшої обробки методами символічних обчислень.

Основні вимоги до методів формалізації знань полягають у такому:

- *логічна адекватність* – формалізація знань має володіти здібністю розпізнавати всі відмінності, закладені у вихідну сутність;
- *евристична потужність* – наряду з наявністю виразної мови формалізації знань, мають існувати певні засоби використання формалізмів, побудованих таким чином, щоб за їх допомогою можна було б ефективно розв'язувати задачі предметної області;
- *природність нотації* передбачає використання для формалізації знань мови, максимально наближеної до природної, що суттєво

спрощує процедуру редагування баз знань з метою доповнення новою інформацією або коригування існуючої.

Управління процесом пошуку рішень

Із застосуванням технологій штучного інтелекту, зокрема, в області експертних систем, пов'язане поняття логічного виведення

Логічне виведення – процес отримання нових знань із вихідних фактів за заданими правилами

Тут розрізняють такі види логічного виведення:

- *дедукція* – логічний розсуд, у якому висновки випливають із відповідних посилок за правилами виду *ЯКЩО А, ТО В*;
- *індукція* – логічне виведення від окремого випадку до загального, відповідно до якого розв'язання поставлених задач виконують шляхом узагальнення відомих способів розв'язання задач аналогічного змісту;
- *інтуїція* – метод логічного виведення, який базується на підсвідомому розпізнаванні певних образів і не опирається на перевірені теорії (в системах штучного інтелекту ще не реалізований);
- *евристика* – метод логічного виведення, заснований на емпіричних правилах, отриманих на основі практичного досвіду і, часто, не підтверджених жорсткою теорією;
- метод *спроб та помилок* полягає у цілеспрямованому переборі можливих способів розв'язання задач з метою визначення оптимального або прийняттого рішення;
- *абдукція* – метод формування суджень у зворотному напрямку від висновку до посилок, які підтверджують або спростовують істинність висновку;
- *судження за аналогією* – метод логічного виведення, який базується на використанні практичного досвіду розв'язання подібних задач;
- *судження за замовченням* – спосіб використання загальновідомих знань у разі відсутності конкретної інформації щодо умов розв'язання поставленої задачі;
- *немонотонні судження* – судження, які необхідно застосувати в умовах, коли отримані раніше знання виявляються невірними після отримання нових свідчень.

Під час проектування експертних систем особлива увага має бути надана тому, яким чином здійснюється доступ до знань і яким чином вони використовуються під час пошуку рішень. Знання про те, які знання потрібні за тієї або іншої ситуації, називають метазнаннями, тобто

знаннями про знання. Такі метазнання дозволяють реалізувати ефективний пошук рішення за рахунок використання певної стратегії перебору знань, що визначає, яким чином експертна система здійснює пошук розв'язання проблеми у просторі альтернатив.

Пояснення прийнятого рішення

Здатність експертної системи пояснювати методику прийняття рішення інколи називають прозорістю системи. Під прозорістю розуміють наскільки просто персоналу з'ясувати, що робить програма і чому. Представлення інформації про поведінку експертної системи важливо на різних рівнях її експлуатації:

- користувачі, які працюють з експертною системою потребують підтвердження того, що у кожному конкретному випадку висновки, до якого прийшла експертна система, є коректним;
- інженери зі знань мають переконатися, що сформульовані ними знання, застосовані правильно;
- експертам з предметної області бажано прослідити хід міркувань і спосіб використання тих відомостей, які з їх слів були введені в базу знань;
- програмістам, які супроводжують, відлагоджують і модернізують експертну систему важливо мати в своєму розпорядженні інструмент, який дає змогу відстежити роботу експертної системи на більш високому рівні, ніж виклик окремих процедур та функцій програмного коду;
- менеджер системи, який несе повну відповідальність за наслідки прийнятих рішень, потребує підтвердження, що рішення, пропоновані експертною системою достатньо обґрунтовані.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Дайте характеристику основних функцій експертної системи.
2. Перелічить та пояснить основні причини складності процесу отримання знань.
3. Дайте характеристику основних вимог до методів представлення знань.
4. Дайте визначення та охарактеризуйте задачі логічного виведення.
5. Дайте характеристику основних типів логічного виведення.
6. Дайте визначення метазнанням.
7. Пояснить необхідність пояснення пропонованих рішень експертною системою.

[5], [11], [45]

§1.6. КОРОТКІ ХАРАКТЕРИСТИКИ ДЕЯКИХ ВІДОМИХ ЕКСПЕРТНИХ СИСТЕМ

До теперішнього часу розроблена велика кількість експертних систем, які працюють в різних галузях експертних знань. Розглянемо деякі класичні експертні системи.

Експертна система DENDRAL

Вважається першою експертною системою, розробка якої розпочата в 1965 році в Стендфордському університеті (США). Проблемна область експертної системи – хімічний аналіз складних органічних сполук на основі обробки мас-спектральних графів молекул. Система DENDRAL реалізує пряме логічне виведення на базі апарата Аристотелевої логіки із використанням простих правил вигляду «якщо – то». Експертна система реалізує процес міркування подібно досвідченому хіміку і вирішує, яка молекула могла дати відповідні мас-спектрограми.

Підхід, реалізований в експертній системі DENDRAL виявився настільки ефективним, що «нащадки» системи до сьогодні використовуються в хімічних і фармацевтичних лабораторіях.

Експертна система MYCIN

Розроблена в середині 1970-х років в Стендфордському університеті. Призначена для діагностики інфекційних захворювань крові і менінгіту. В експертній системі MYCIN, в одній з перших реалізовано вирішення проблеми прийняття рішення на основі ненадійної або недостатньої інформації. Експертна система виводить ясні і логічні пояснення своїх міркувань, використовуючи структуру логіки, відповідну специфіці предметної області і критерії для надійного оцінювання своєї роботи.

Дослідження, проведені в Стенфордській медичній школі, виявили, що MYCIN пропонує прийнятну терапію приблизно в 69% випадків, що краще, ніж у експертів з інфекційних хвороб. Разом з тим, експертна система MYCIN ніколи не використовувалася на практиці. Це пояснюється, у першу чергу, етично-правовими міркуваннями про відповідальність у разі неправильного діагнозу або неправильного дозування ліків.

Проте, багато сучасних методик розроблення експертних систем вперше були розроблені саме в рамках проекту MYCIN.

Експертна система PROSPECTOR

Розробка експертної системи PROSPECTOR проводилась компанією SRI International (США) з 1974 по 1983 роки. Предметна область експертної системи – дослідження в геології. Система отримує дані про характеристики досліджуваної породи та дає висновок про те, які мінерали можуть знаходитися в родовищі, оцінюючи при цьому перспективність геологічного району щодо наявності корисних копалин. Логічне виведення в експертній системі PROSPECTOR реалізовано на базі апарату Байєсової логіки.

Система PROSPECTOR – одна з перших експертних систем, доведених до промислового використання. Після того, як в 1984 році за її допомогою було знайдено велике родовище молібдену, технології експертних систем отримали широке визнання, що призвело до масових розробок в цій області по всьому світу.

Експертна система МИМИР

Система МИМИР (Малая Информационная Модель Интеллектуальных Решений) є складовою частиною інформаційно-управляючої системи «Диспетчер» та призначена для побудови на її базі інформаційних систем типу АСДУ, тренажерів та навчальних систем. Розробки проводились починаючи з 1980-х років у ВНИИЭ.

На базі МИМИР розроблено ряд прикладних експертних систем для задач управління енергосистемами та енергооб'єднаннями, зокрема з моніторингу стану устаткування електричних мереж, радника з виведення в ремонт обладнання, з ситуаційної діагностики стану обладнання тощо.

Серед сучасних слід зазначити експертну систему OMEGAMON, розроблену корпорацією Candle (США). Призначення цієї системи полягає у моніторингу позаштатних ситуацій в корпоративних інформаційних мережах та генерації управляючих дій, направлених на підтримку нормальної роботи таких мереж. Система OMEGAMON є динамічною, мультиагентною, здатною працювати в режимі реального часу.



ЛІТЕРАТУРА

[4], [7], [45], [76], [49], [55], [64], [9], [17], [63]

МОДЕЛІ ФОРМАЛІЗАЦІЇ ТА РОЗВ'ЯЗАННЯ ПРАКТИЧНИХ ЗАДАЧ В СЕРЕДОВИЩІ ЕКСПЕРТНИХ СИСТЕМ

§2.1. ЗАГАЛЬНА ЗАРАКТЕРИСТИКА СПОСОБІВ ФОРМАЛІЗАЦІЇ ПРАКТИЧНИХ ЗАДАЧ

Для розв'язання практичних задач, зазвичай, використовують одну з чотирьох стратегій:

- використання явної формули (наприклад, для визначення струму за законом Ома);
- використання рекурентної формули (наприклад, для ітераційного уточнення режиму напруги із застосуванням методу Зейделя до вузлової розрахункової моделі усталеного режиму електричної системи);
- використання алгоритму (наприклад, під час реалізації оперативних перемикань на розподільчому пристрої електричної підстанції);
- використання методу перебору (спроб та помилок).

В розділі 1 йшла мова про те, що експертні системи призначені для розв'язання складних нетипових задач, для яких неможливо, або дуже складно сформулювати адекватну строгу математичну модель. Це означає, що експертні системи спираються, в основному, саме на останню стратегію розв'язання задач. Разом з тим, вони, за необхідності, можуть застосовувати й інші стратегії.

Відомо, що експертні системи моделюють процеси мислення людей-експертів під час розв'язання практичних задач предметної області. Слід зазначити, що людина, зазвичай, не розв'язує задачу у тому вигляді, в якому вона від початку сформульована, а представляє її таким чином, щоб її було зручно розв'язувати. Таке переформулювання умов практичних задач називають формалізацією задач.

До способів формалізації задач висувають дві основні вимоги:

- 1) подання задачі має достатньо точно моделювати предметну область експертної системи;
- 2) подання має давати змогу реалізовувати просте та ефективне розв'язання практичних задач.

Коло задач, які розв'язують за допомогою експертних систем, є дуже широким. Відповідно до їх специфіки використовують різні способи подання, серед яких найбільшого поширення набули наступні:

- подання у просторі станів;
- подання у просторі задач;
- подання у вигляді доведення теорем.

Наведемо короткі характеристики цих способів.

Подання задач у просторі станів містить такі елементи:

- початковий стан системи;
- цільовий стан системи;
- сукупність операторів, за допомогою яких реалізують переходи між станами системи.

Розв'язання задачі тут полягає у визначенні послідовності переходів між початковим та цільовим станами.

Очевидно, існує багато задач, розв'язання яких можна інтерпретувати як пошук у просторі станів, наприклад:

- визначення послідовності операцій по переведенню робочого режиму енергосистеми в інший в процесі оперативного управління;
- визначення послідовності оперативних перемикачів для виведення в ремонт (або введення в експлуатацію) устаткування електричних мереж;
- визначення раціональної стратегії перспективного розвитку електроенергетичної системи;
- визначення оптимальної конфігурації робочої схеми електричної мережі

та багато інших.

У поданні задач у просторі станів широкого поширення набув математичний апарат теорії графів. Тут кожному можливому стану предметної області відповідає певна вершина, а можливим переходам між станами – орієнтовані дуги графу. Крім того, для розв'язання деяких практичних задач кожній дузі графу ставлять у відповідність певний ваговий коефіцієнт, який характеризує, наприклад, витрати ресурсів на реалізацію переходів між станами предметної області. Таким чином, розв'язання задачі у просторі станів зводиться до пошуку в орієнтованому зваженому графі, який моделює стани предметної області. Такий граф називають деревом пошуку.

На рис. 2.1 наведено приклад орієнтованого графу простору станів деякої предметної області. Тут початковому стану відповідає вершина *A*, цільовому – вершина *I*. Можливі переходи між станами на рис. 2.1 представлено стрілками.

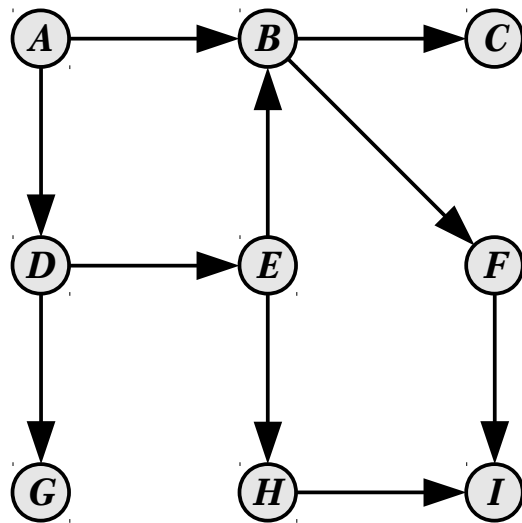


Рис. 2.1. Граф простору станів

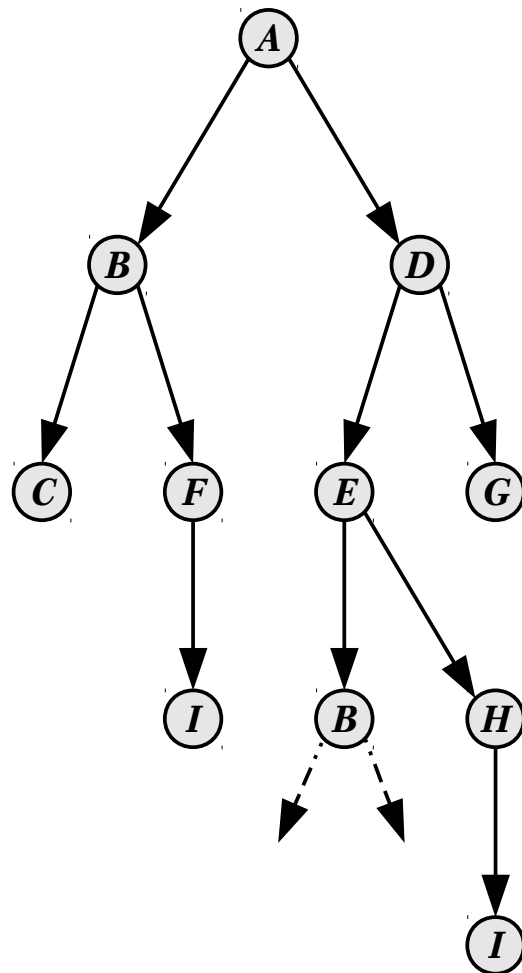


Рис. 2.2. Дерево пошуку у просторі станів

Процес розв'язання задачі полягає у формуванні дерева пошуку, вершина (корінь) якого відповідає початковому стану предметної області. Інші вершини дерева пошуку відповідають всім можливим станам предметної області, впорядкованим за ярусами відповідно до кількості переходів із початкового стану. Ребрам дерева пошуку відповідають можливі переходи між станами предметної області.

На рис. 2.2 наведено дерево пошуку, яке відповідає графу простору станів на рис. 2.1. Вершини дерева пошуку, які породжуються можливими переходами із поточної вершини, називають дочірніми. Так, для вершини A дерева пошуку на рис. 2.2 дочірніми є вершини B та D . Вершину, яка породжує поточну називають батьківською. Так, для вершини B батьківською є вершина A . Вершину, яка не має жодної дочірньої називають тупиковою. Так, тупиковими є вершини C та G .

Аналіз дерева пошуку, наведеного на рис. 2.2 свідчить про наявність альтернативних шляхів на графі простору станів. Так, стану B предметної області відповідає дві вершини дерева станів у другому та четвертому ярусах відповідно. Дійсно, до стану B можна потрапити безпосередньо з вихідного стану A , або через два проміжні стани D і E . Часто в таких випадках із складу дерева пошуку виключають можливі альтернативні шляхи, залишаючи лише єдиний (перший з визначених або оптимальний). На дереві пошуку рис. 2.2 вилучення альтернативного шляху показано пунктирними стрілками. Відповідні вершини дерева пошуку умовно розглядають як тупикові. На рис. 2.2 це вершина B третього ярусу.

Процес розв'язання задачі закінчують, коли до складу дерева пошуку потрапляє цільова вершина. На рис. 2.2 це відбувається під час формування третього ярусу – дочірніх вершин стану F або четвертого ярусу – дочірніх вершин стану H .

Подання у просторі задач полягає у багатократному розбитті задач на дрібніші доки вихідна задача не буде зведена до сукупності елементарних, які мають тривіальний розв'язок, або доки не буде доведено, що задача не має розв'язку. Процес розбиття (декомпозиції) задачі називають редукцією.

Очевидно, що спосіб зведення задач до підзадач ефективний у розв'язанні складних комплексних задач, декомпозиція яких породжує взаємно незалежні підзадачі, наприклад, пов'язані із проектуванням електричних станцій, підстанцій, електричних мереж.

Для математичного моделювання зведення задач до підзадач широкого застосування також набув апарат теорії графів. Вершинами такого графу виступають вихідна задача та підзадачі, отримані в результаті редукції. Ребрам графу відповідають оператори декомпозиції, які породжують нові підзадачі. Отриманий граф називають графом редукції.

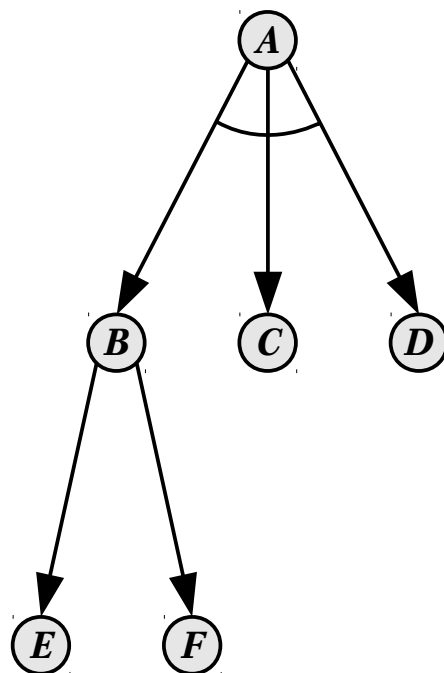


Рис. 2.3. Дерево редукції

У графі редукції розрізняють два види вершин – кон'юнктивні та диз'юнктивні. Кон'юнктивні вершини означають, що для розв'язання задачі необхідно розв'язати всі дочірні підзадачі. Такі вершини часто називають вершинами типу «І». Диз'юнктивні вершини, у свою чергу, позначають розбиття на альтернативні підзадачі. Для розв'язання таких задач достатньо розв'язати будь-яку з дочірніх підзадач. Диз'юнктивні вершини часто називають вершинами типу «АБО».

Граф редукції задачі називають графом «І-АБО» типу. На рис. 2.3 наведено приклад графу редукції деякої задачі. Тут вершині *A* відповідає вихідна задача, для розв'язання якої необхідно розв'язати задачі, яким відповідають вершини *B*, *C* та *D*. Вершина *A* є кон'юнктивною, що позначено на рис. 2.1 дужкою, яка поєднує ребра *A-B*, *A-C* та *A-D*. Для розв'язання задачі, яка відповідає вершині *B* необхідно розв'язати одну з її підзадач – *E* або *F*. Підзадачі *E* та *F* є альтернативними, тому вершина *B* є диз'юнктивною.

Початкова вершина (корінь) дерева редукції відповідає вихідній задачі, яку необхідно розв'язати. Вершини, які відповідають елементарним задачам із тривіальним розв'язком називають термінальними. Вершини, які відповідають задачам, які не мають розв'язку, називають тупиковими. Очевидно, що термінальні вершини є розв'язними, а тупикові – нерозв'язними. Розв'язними є також кон'юнктивні вершини, усі дочірні підзадачі яких є розв'язними, а також диз'юнктивні вершини, в яких хоча б одна дочірня підзадача є розв'язною. Таким чином, задача пошуку на графі дерева редукції полягає у визначенні розв'язності

кореневої вершини та вибору способу розв'язання цільової задачі. Зазначимо, що у разі відсутності кон'юнктивних вершин у складі дерева редукції, процес розв'язання задачі при її зведенні до підзадач повністю збігається із розв'язанням задачі у просторі станів.

Подання задач у вигляді доведення теорем полягає у послідовному багатократному переформулюванні умов задачі доки не буде доведена (або спростована) істинність певних положень. Процес розв'язання задачі тут полягає у послідовному виконанні наступних процедур:

- 1) формують множину вихідних істинних тверджень (аксіом);
- 2) формують гіпотезу відносно результату розв'язку задачі;
- 3) аксіоми комбінують між собою за допомогою правил виведення та отримують нові істинні твердження. Якщо під час виконання процедури нові твердження не з'явилися, це свідчить про неможливість розв'язання задачі;
- 4) перевіряють, чи є серед нових тверджень цільова гіпотеза, або її спростування. Якщо існує, то теорему вважають доведеною та розв'язання задачі закінчують. В інакшому випадку управління передають процедурі 3 алгоритму для формування нових істинних тверджень.

Очевидно, що розглянутий спосіб подання задач є окремим випадком представлення у просторі станів із логічною формалізацією ходу рішення. Дійсно, всі можливі сукупності істинних тверджень, які виникають в процесі розв'язання задачі, визначають можливі стани предметної області, тобто вершини графу простору станів, а правила виведення, за допомогою яких аксіоми поєднуються визначають можливі переходи між станами, тобто ребра графу.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Які вимоги висувають до способів формалізації задач?
2. Перелічіть найбільш поширені способи подання задач.
3. Дайте характеристику подання задач у просторі станів.
4. Наведіть перелік задач з області електроенергетики, які можна подати у просторі станів.
5. Що називають графом простору станів?
6. Що називають деревом пошуку?
7. У чому полягає розв'язання задач за їх подання у просторі станів?
8. Дайте характеристику зведення задач до підзадач.
9. Наведіть перелік задач з області електроенергетики, які можна звести до підзадач.
10. Що називають деревом редукції?

11. У чому полягає розв'язання задач за їх зведення до підзадач?
12. За яких умов процеси розв'язання задач у просторі станів та за зведення до підзадач збігаються?
13. Дайте характеристики подання задач у вигляді доведення теорем.
14. Наведіть перелік задач з області електроенергетики, які можна подати у вигляді доведення теорем.
15. Доведіть, що подання задач у вигляді доведення теорем є окремим випадком подання у просторі станів.



ЛІТЕРАТУРА

[1], [2], [3], [4], [5], [7], [9], [11]

§2.2. ЗАГАЛЬНА ХАРАКТЕРИСТИКА АЛГОРИТМІВ РОЗВ'ЯЗАННЯ ЗАДАЧ У ПРОСТОРІ СТАНІВ

Розв'язання задач пошуку у просторі станів полягає у визначенні сукупності переходів між вихідним та цільовим станами предметної області. Формування та реалізація відповідних алгоритмів базується на двох основних положеннях.

1. Граф пошуку задається у пам'яті експертної системи у неявному вигляді сукупності можливих станів предметної області та переходів між ними. В результаті дерево пошуку будується послідовно, починаючи з вихідного стану системи і проглядаючи всі можливі переходи з цього стану.

2. Кожний цикл пошуку у просторі станів полягає у виборі одного переходу серед множини альтернатив. В теорії систем штучного інтелекту множину можливих альтернативних переходів називають конфліктною множиною. Саме ефективний вибір наступного переходу з конфліктної множини характеризує інтелектуальність експертної системи.

Розв'язання задач пошуку у просторі станів часто пов'язано із проблемою комбінаторного вибуху. Комбінаторний вибух полягає у різкому зростанні розмірності конфліктної множини із просуванням по графу станів. Це пов'язано з тим, що кожний цикл пошуку пов'язаний із можливістю реалізації лише одного переходу, який часто породжує багато (часто – дуже багато) нових альтернатив. Так, наприклад, під час гри у шахи виконання першого ходу пов'язано із вибором однієї з 20 альтернатив (16 ходів пішаками на одну або дві клітини вперед та 4 ходи

конями). Після обміну традиційними першими ходами партії e2–e4; e7–e5 кількість альтернативних ходів білими фігурами зростає до 28 (14 ходів пішаками, 4 ходи конями, 5 ходів слоном, 4 ходи ферзем та 1 хід королем). Далі, із розвитком гри розмірність конфліктної множини різко збільшується таким чином, що для прорахунку всіх можливих комбінацій на 6 ходів вперед сучасному комп'ютеру, здатному прорахувати мільйон комбінацій на секунду, буде потрібна 1 секунда, на 12 ходів – 11 діб, на 18 ходів – біля 32 тисяч років, а для формування повного дерева пошуку, яке охоплює всі можливі стани на шаховій дошці, потрібний час, порівняний з часом, який минув з моменту «великого вибуху».

Таким чином, алгоритми пошуку мають додатково штучно обмежувати розмірність дерев пошуку для отримання прийняттого результату в умовах обмеженого часу. Ефективність такого обмеження також свідчить про інтелектуальність експертних систем.

Методи пошуку у просторі станів поділяють на дві групи: неінформативного («сліпого») та направленого (евристичного) пошуку. Методи «сліпого» пошуку не враховують інформацію про близькість поточного стану до цільового. Це призводить до необхідності повного перебору всього простору станів, що для нетривіальних задач пов'язано із проблемою комбінаторного вибуху. Евристичні методи навпаки використовують додаткову інформацію, яка дозволяє обмежити конфліктну множину та обирати черговий перехід таким чином, щоб із кожним кроком гарантовано наближатися до цільового стану.

? КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає розв'язання задач у просторі станів?
2. Сформулюйте основні положення, на яких базуються алгоритми пошуку у просторі станів.
3. У чому полягає проблема комбінаторного вибуху?
4. Наведіть класифікацію методів пошуку у просторі станів.



ЛІТЕРАТУРА

[1], [2], [5], [9], [11], [49], [64], [62], [63]

§2.3. АЛГОРИТМИ НЕІНФОРМАТИВНОГО ПОШУКУ У ПРОСТОРІ СТАНІВ

Методи неінформативного («сліпого») пошуку передбачають реалізацію повного перебору всіх можливих станів у просторі предметної області, не враховуючи додаткової оціночної інформації щодо ефективності кожного можливого переходу. До цієї групи методів відносять:

- випадковий пошук;
- пошук у глибину;
- пошук у ширину;
- пошук за алгоритмом Дейкстри.

Стратегія *випадкового пошуку* полягає в тому що оператор чергового переходу обирають випадково. Пошук виконують доки не буде досягнуто цільового стану. Очевидно, що метод випадкового пошуку характеризується простотою реалізації. Разом з тим випадковий вибір переходу не гарантує збіжності алгоритму. Такий метод можна використовувати для розв'язання простих задач з малою розмірністю простору станів. Для більш складних задач час пошуку стає невиправдано великим, що суттєво обмежує використання алгоритму.

Стратегія *пошуку у глибину* полягає у послідовному розгляданні одного ланцюга переходів між станами предметної області. Якщо під час такого розглядання будуть виявлені тупикові стани або замкнені цикли, то система пошуку відкочується до останньої нереалізованої альтернативи та продовжує розглядання нового ланцюга переходів. Пошук продовжують до виявлення цільового стану.

Проілюструємо реалізацію пошуку у глибину на прикладі пошуку у просторі станів, граф якого наведено на рис. 2.1. На кожному кроці процесу пошуку будемо розглядати альтернативні переходи з поточної вершини за годинниковою стрілкою, починаючи з «дванадцятої години». Реалізовувати при цьому будемо перехід, виявлений останнім. Процес пошуку альтернативних переходів та їх реалізацію називають розкриттям вершини графу.

Відповідно до графу станів, наведеному на рис. 2.1 з початкової вершини *A* можна перейти до станів *B* або *D*. Відповідно до прийнятої стратегії пошуку реалізуємо перехід до стану *D*, як показано на рис. 2.4 а. Розкриття вершини *D* породжує переходи до станів *E* та *G*. Відповідно до стратегії пошуку у глибину виконуємо перехід до вершини *G*, що показано на рис. 2.4 б. Стан *G* є тупиковим, що свідчить про необхідність повернення до стану *D* та реалізацію останньої нереалізованої альтернативи – переходу до стану *E* (див. рис. 2.4 в). Розкриття вершини *E* визначає переходи до станів *B* та *H*. Тут допустимим є лише перехід

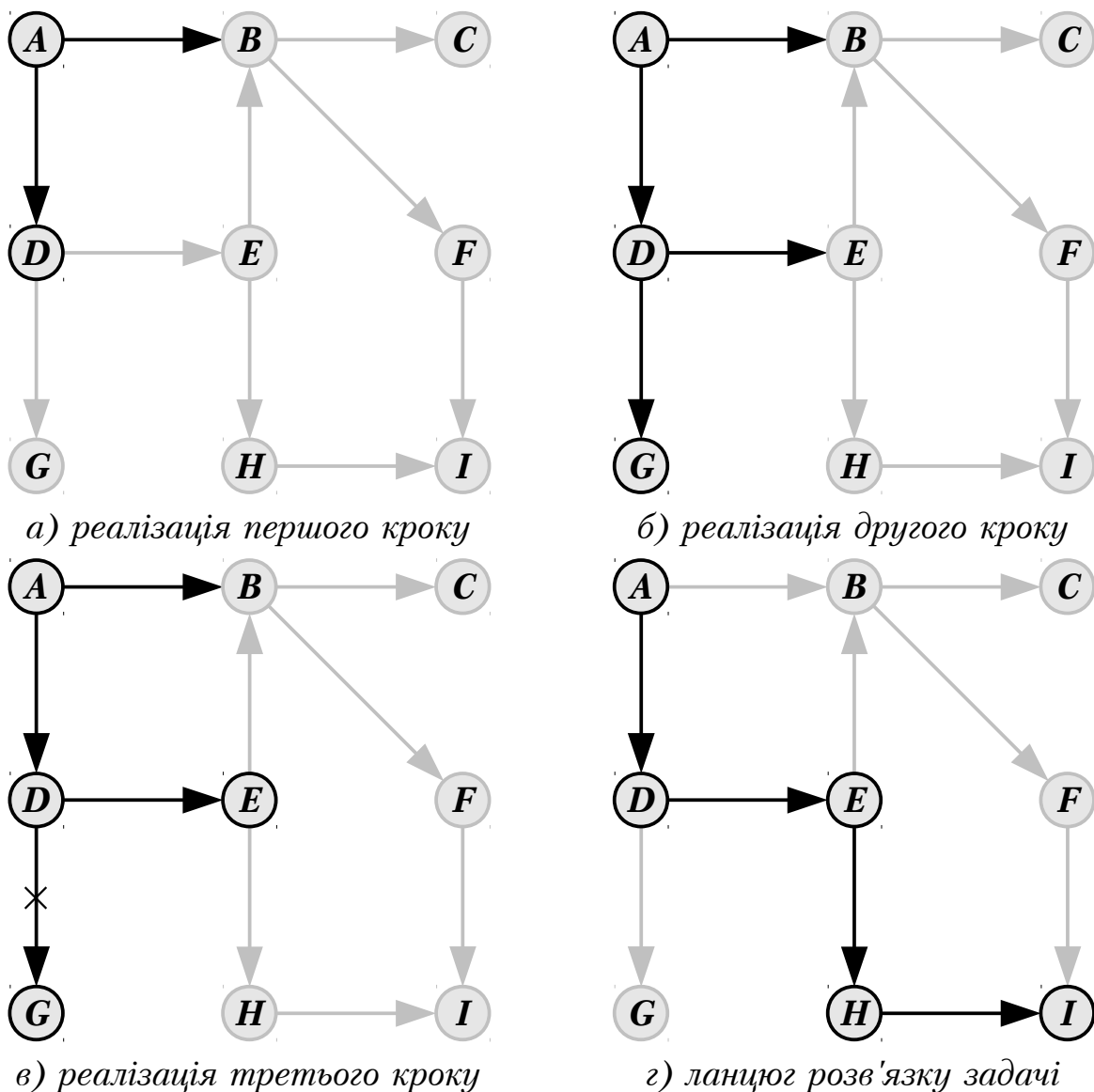


Рис. 2.4. Ілюстрація пошуку у глибину

до стану H , оскільки перехід в вершину B вже знаходиться в черзі потенційних переходів із початкового стану A . Далі необхідно реалізувати безальтернативний перехід до цільового стану, що свідчить про знаходження рішення у вигляді ланцюга переходів $A-D-E-H-I$, як показано на рис. 2.4 г.

Формалізація алгоритму пошуку у глибину потребує використання спискової структури конфліктної множини, організованої за принципом стеку (LIFO – Last In First Out – останній прийшов, перший вийшов). До конфліктної множини після розкриття чергової вершини вносять породжені потенційні підграфи розв'язку. Для розкриття використовують кінцеву вершину останнього підграфу конфліктної множини.

Тут на попередньому етапі роботи алгоритму до складу конфліктної множини вносять потенційний підграф розв'язку задачі, який складає-

ться з кореневої вершини: $\{A\}$. Розкриття цієї вершини породжує альтернативні переходи $\{A-B\}$ та $\{A-D\}$, які вносять до складу конфліктної множини замість кореневої вершини $\{A\}$. Таким чином, після виконання першого циклу пошуку, конфліктна множина складається з двох потенційних підграфів:

$$\{\{A-B\}; \{A-D\}\}.$$

Останній потенційний підграф закінчується вершиною D , яку слід розкривати на другому етапі роботи алгоритму. Така операція породжує переходи до станів E або G , що визначає потенційні підграфи розв'язку $\{A-D-E\}$ та $\{A-D-G\}$, які заміщують підграф $\{A-D\}$ в конфліктній множині:

$$\{\{A-B\}; \{A-D-E\}; \{A-D-G\}\}.$$

Далі розкриттю підлягає вершина G , якою закінчується останній потенційний підграф $\{A-D-G\}$. Така вершина є тупиковою і її розкриття не породжує жодного нового підграфу. Це означає, що на цьому етапі потенційний підграф $\{A-D-G\}$ заміщують пустою множиною:

$$\{\{A-B\}; \{A-D-E\}\}.$$

Наступний етап роботи алгоритму пошуку полягає у розкритті вершини E , що породжує новий потенційний підграф $\{A-D-E-H\}$:

$$\{\{A-B\}; \{A-D-E-H\}\}.$$

Тут альтернативний перехід до стану B ($\{A-D-E-B\}$) не розглядають, оскільки вершина B вже знаходиться серед кінцевих вершин потенційних підграфів конфліктної множини – $\{A-B\}$.

На останньому етапі роботи алгоритму здійснюють розкриття вершини H , що породжує шуканий розв'язок задачі у вигляді підграфу $\{A-D-E-H-I\}$.

Стратегія *пошуку у ширину* полягає у послідовному розгляданні сукупності дерев (лісу) переходів із початкового стану довжиною один перехід, два переходи і так далі, аж до досягнення цільового стану системи. При цьому слід відфільтровувати тупикові стани та замкнені цикли на графі простору станів.

Проілюструємо реалізацію пошуку у ширину на прикладі пошуку у просторі станів, граф якого наведено на рис. 2.1. Відповідно до графу станів, наведеному на рис. 2.1 початкова вершина A породжує куц дерева з кінцевими станами B та D , як показано на рис. 2.5 а. Розкриття кінцевих вершин потенційних підграфів B та D породжує переходи $B-C$ та $B-F$, а також $D-E$ і $D-G$. Тут вершини C і G є тупиковими (див. рис. 2.5 б). Таким чином, подальшому розкриттю підлягають вершини E та F . Їх розкриття породжує переходи $E-B$, $E-H$ і $F-I$, при чому перехід

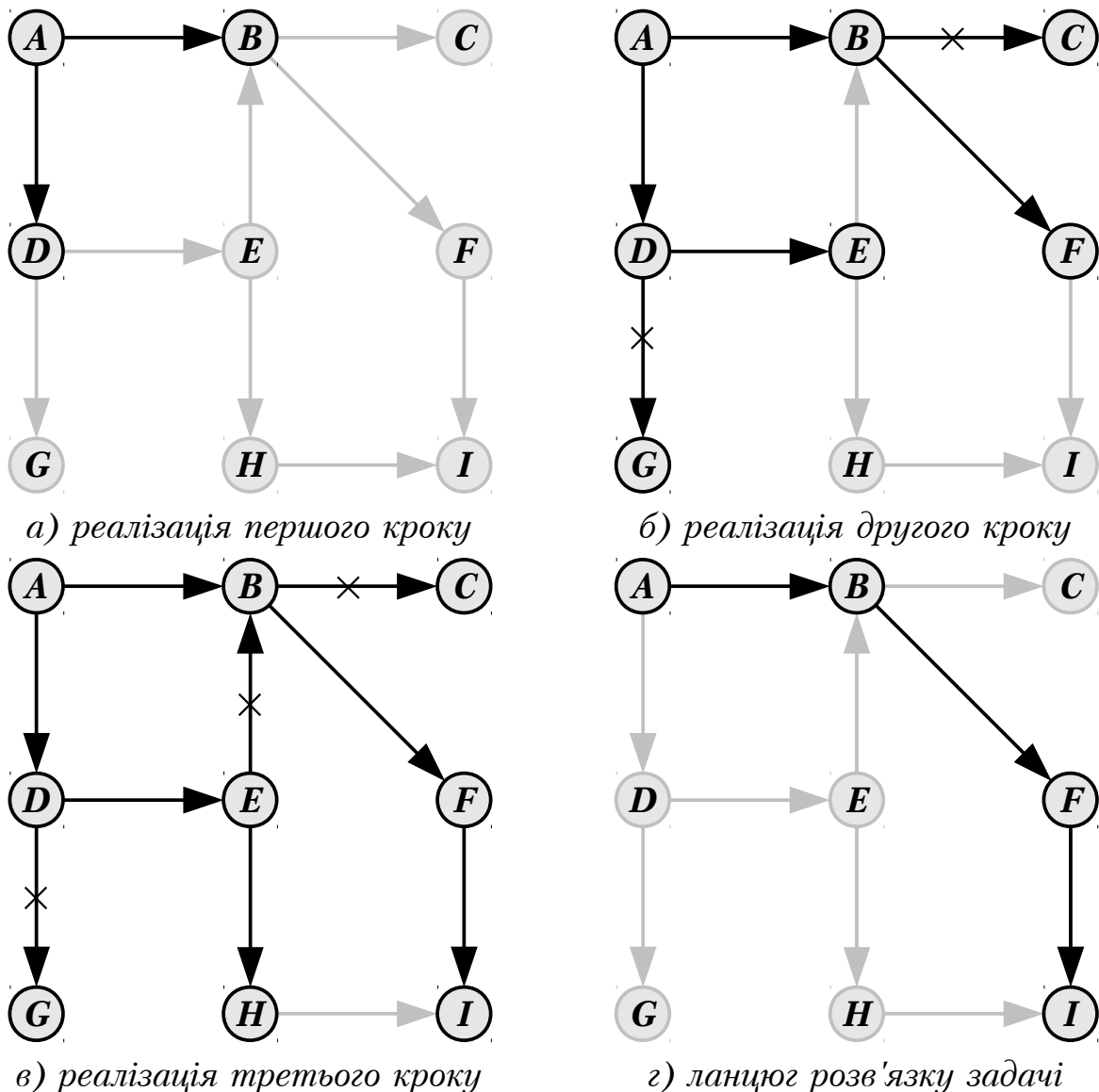


Рис. 2.5. Ілюстрація пошуку у ширину

$E-B$ є недопустимим, оскільки породжує замкнений цикл (див. рис. 2.5 в). Перехід $F-I$ призводить до цільового стану, що свідчить про знаходження розв'язку у вигляді ланцюга переходів $A-B-F-I$, як показано на рис. 2.5 г.

Формалізація алгоритму пошуку у ширину потребує використання спискової структури конфліктної множини, організованої за принципом черги (FIFO – First In First Out – перший прийшов, перший вийшов). До конфліктної множини після розкриття чергової вершини вносять породжені потенційні підграфи розв'язку. Для подальшого розкриття використовують кінцеву вершину першого підграфу конфліктної множини.

Тут на попередньому етапі роботи алгоритму до складу конфліктної множини вносять потенційний підграф розв'язку задачі, який складає-

ться з кореневої вершини: $\{A\}$. Розкриття цієї вершини породжує альтернативні переходи $\{A-B\}$ та $\{A-D\}$, які вносять до складу конфліктної множини замість кореневої вершини $\{A\}$ так само, як і під час виконання алгоритму пошуку у глибину. Таким чином, після виконання першого циклу пошуку, конфліктна множина складається з двох потенційних підграфів:

$$\{\{A-B\}; \{A-D\}\}.$$

Далі розкриттю підлягає вершина B , якою закінчується перший потенційний підграф конфліктної множини. Така операція породжує переходи до станів C або F , що визначає потенційні підграфи розв'язку $\{A-B-C\}$ та $\{A-B-F\}$, які записують в кінець конфліктної множини а підграф $\{A-B\}$ виключають із неї:

$$\{\{A-D\}; \{A-B-C\}; \{A-B-F\}\}.$$

Перший потенційний підграф закінчується вершиною D , яку слід розкривати на третьому етапі роботи алгоритму. Така операція породжує переходи до станів E або G , що визначає потенційні підграфи розв'язку $\{A-D-E\}$ та $\{A-D-G\}$, які заміщують підграф $\{A-D\}$ в конфліктній множині:

$$\{\{A-B-C\}; \{A-B-F\}; \{A-D-E\}; \{A-D-G\}\}.$$

Далі розкриттю підлягає вершина C , якою закінчується преший потенційний підграф $\{A-B-C\}$. Така вершина є тупиковою і її розкриття не породжує жодного нового підграфу. Таким чином, після виконання поточного етапу конфліктна множина набуває вигляду

$$\{\{A-B-F\}; \{A-D-E\}; \{A-D-G\}\}.$$

На останньому етапі роботи алгоритму здійснюють розкриття вершини F , що породжує шуканий розв'язок задачі у вигляді підграфу $\{A-B-F-G\}$.

Розв'язання задач пошуку у просторі станів часто потребує урахування витрат ресурсів на реалізацію переходів між станами предметної області. Тут кожному можливому переходу однозначно зіставляють деяке додатне число, яке характеризує витрати ресурсів на реалізацію цього переходу. Очевидно, що така задача є оптимізаційною з мінімізації сумарних витрат ресурсів на переходи між вихідним та цільовим станами. Для розв'язання такої задачі пошуку використовують *алгоритм Дейкстри*. Суть цього алгоритму полягає у наступному. Характеристики витрат ресурсів на реалізацію переходів між станами дозволяють отримати значення цін станів предметної області як витрати ресурсів на досягнення відповідного стану із початкового за рекурентною формулою:

$$g_j = g_i + c_{ij},$$

де g_i – ціна i -го стану; c_{ij} – витрати ресурсів на реалізацію переходу $i-j$.

Очевидно, що через наявні замкнені контури в графі пошуку можливе визначення цін станів за різними траєкторіями. Тут, виходячи з принципів розв'язання оптимізаційних задач, ціною стану є найменше з можливих значень. Пошук шляху на графі станів відповідно до алгоритму Дейкстри починають із початкової вершини, ціна якої дорівнює нулю і розкриваючи послідовно всі вершини з найменшими значеннями ціни, що забезпечує мінімізацію витрат ресурсів.

Проілюструємо роботу алгоритму Дейкстри на прикладі розв'язання задачі пошуку на графі станів, наведеному на рис. 2.1. У табл. 2.1 наведено значення витрат ресурсів на реалізацію переходів.

Таблиця 2.1. Витрати ресурсів на реалізацію переходів

<i>A-B</i>	<i>A-D</i>	<i>B-C</i>	<i>B-F</i>	<i>D-E</i>	<i>D-G</i>	<i>E-B</i>	<i>E-H</i>	<i>H-I</i>	<i>F-I</i>
9	3	8	4	4	2	1	6	4	8

На першому етапі необхідно здійснити розкриття вихідної вершини A графу станів, що породжує два альтернативні переходи до станів B та D . Витрати ресурсів на реалізацію цих переходів визначають ціни відповідних вершин:

$$g_B = c_{AB} = 9; \quad g_D = c_{AD} = 3.$$

Найменшою ціною характеризується вершина D , тому на першому етапі необхідно реалізувати перехід до стану D (див. рис. 2.6 а). Розкриття вершини D дозволяє визначити ціни суміжних станів E та G :

$$g_E = g_D + c_{DE} = 3 + 4 = 7;$$

$$g_G = g_D + c_{DG} = 3 + 2 = 5.$$

Серед нерозкритих вершин найменшою ціною характеризується вершина G ($g_G = 5$), тому перехід слід здійснити до стану G і на наступному етапі розкрити саме цю вершину. Проте вершина G є тупиковою та її розкриття не призводить до визначення цін наступних станів.

Таким чином, на наступному етапі реалізують перехід $D-E$ та здійснюють розкриття вершини E :

$$g_B = g_E + c_{EB} = 7 + 1 = 8;$$

$$g_H = g_E + c_{EH} = 7 + 6 = 13.$$

Зазначимо, що проведена операція дозволила перевизначити ціну стану B . Дійсно, на попередніх етапах за траєкторією $A-B$ ціна вершини B була визначена на рівні 9 одиниць, а за траєкторією $A-D-E-B$ – 8 одиниць. Тому у подальших розрахунках ціну вершини B приймають

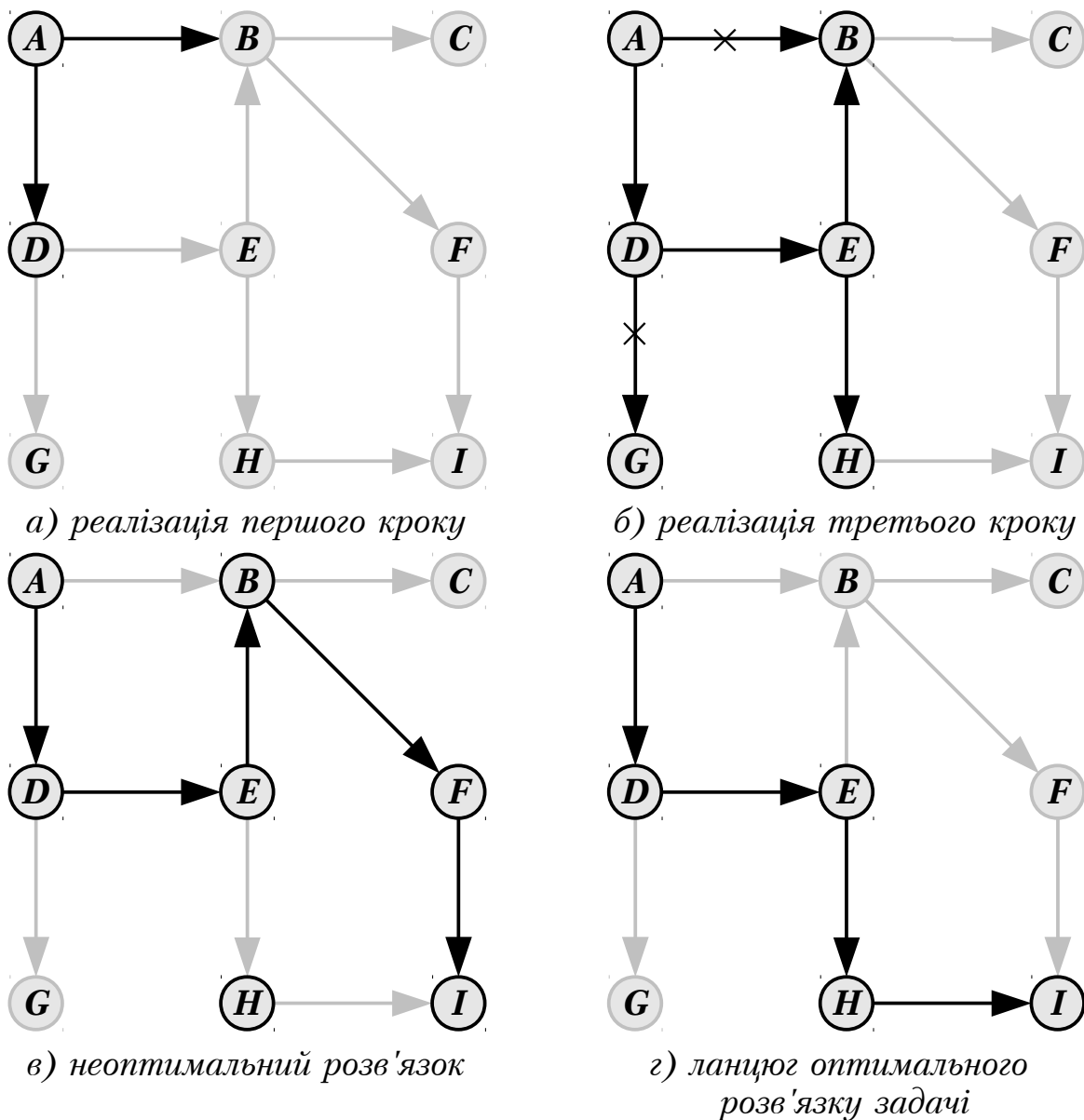


Рис. 2.6. Ілюстрація пошуку за алгоритмом Дейкстри

на рівні 8 одиниць, а перехід $A-B$ помічають як тупиковий (див. рис. 2.6 б).

На подальших етапах виконують пошук за наведеним алгоритмом. Тут розкриття вершини B дозволяє визначити ціни станів C та F :

$$g_C = g_B + c_{BC} = 8 + 8 = 16;$$

$$g_F = g_B + c_{BF} = 8 + 4 = 12.$$

Наступною розкривають вершину F , що визначає перехід до цільового стану I із ціною

$$g_I = g_F + c_{FI} = 12 + 8 = 20.$$

Формально на цьому виконання алгоритму можна завершити, оскільки було знайдено сукупність переходів між вихідною та цільовою вершинами за траєкторією $A-D-E-B-F-I$, що показано на рис. 2.6 в.

Разом з тим, це ще не визначає оптимальний шлях між вихідною та цільовою вершинами, оскільки на графі пошуку залишились не розкриті вершини, переходи з яких можуть мінімізувати витрати ресурсів та знайти більш оптимальний розв'язок задачі. Дійсно, розкриття вершини H визначає більш оптимальний шлях на графі пошуку, показаний на рис. 2.6 г із ціною

$$g_I = g_H + c_{HI} = 13 + 4 = 17.$$

Таким чином, реалізація алгоритму Дейкстри дозволяє розв'язати задачу пошуку у вигляді ланцюга переходів $A-D-E-H-I$.

Очевидно, що алгоритм Дейкстри являє собою модифікацію алгоритму пошуку у ширину. Це означає, що формалізація алгоритму Дейкстри базується за принципом черги. Різниця тут полягає у тому, що перед розкриттям чергової вершини чергу конфліктної множини сортують за збільшенням цін вершин, якими закінчуються відповідні потенційні підграфи. Далі, так само, як і для алгоритму пошуку у ширину для розкриття обирають вершину, якою закінчується перший потенційний граф конфліктної множини.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає алгоритм випадкового пошуку?
2. У чому полягає стратегія пошуку у глибину?
3. Сформулюйте формальний алгоритм пошуку у глибину.
4. У чому полягає стратегія пошуку у ширину?
5. Сформулюйте формальний алгоритм пошуку у ширину.
6. У чому полягає стратегія пошуку за алгоритмом Дейкстри?
7. Сформулюйте формальний алгоритм Дейкстри.
8. Поясніть положення про те, що алгоритм Дейкстри є модифікацією алгоритму пошуку у ширину.
9. Проведіть покрокове розв'язання наведеної задачі пошуку за алгоритмом Дейкстри із використанням формалізмів організації спискових структур конфліктної множини.



ЛІТЕРАТУРА

[1], [2], [3], [7], [8], [9], [10], [11]

§2.4. АЛГОРИТМИ ЕВРИСТИЧНОГО ПОШУКУ У ПРОСТОРИ СТАНІВ

Реалізація методів евристичного (направленого) пошуку базується на використанні додаткової інформації, яка дозволяє оцінити очікувані витрати ресурсів на реалізацію переходів між поточним та цільовим станами предметної області. До евристичних методів пошуку відносять:

- градієнтний алгоритм;
- алгоритм глобального врахування відповідності цілі;
- А-алгоритм.

Градієнтний алгоритм являє собою модифікацію пошуку у глибину, де розкриття чергової вершини здійснюють з кінця поточного ланцюга пошуку у напрямку мінімального значення оцінювальної функції суміжного стану таким чином, щоб якомога ближче наблизитись до цілі на кожному кроці пошуку.

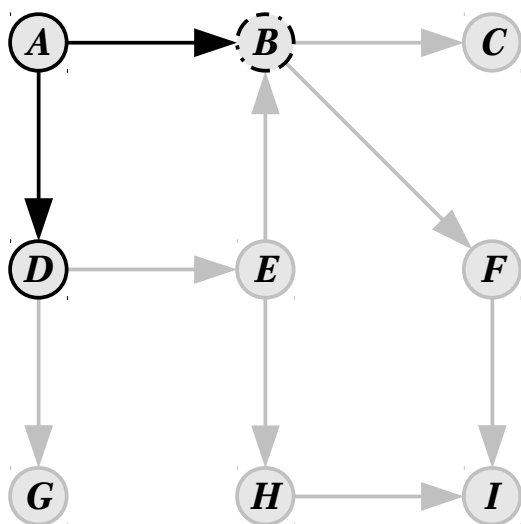
Проілюструємо роботу градієнтного алгоритму на прикладі розв'язання задачі пошуку на графі станів, наведеному на рис. 2.1. У табл. 2.2 наведено значення оцінювальної функції витрат ресурсів на реалізацію переходів між поточним та цільовим станами.

Таблиця 2.2. Значення оцінювальної функції вершин графу пошуку

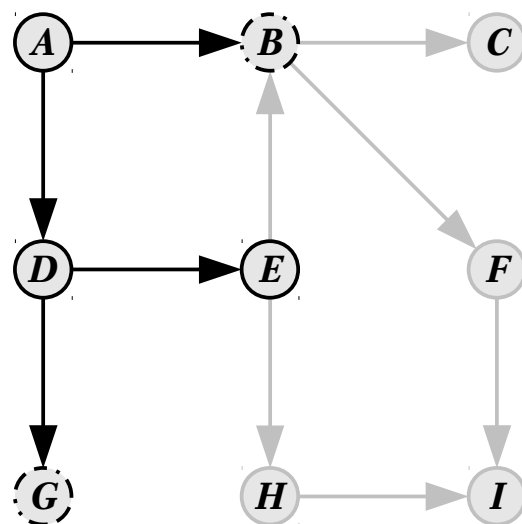
A	B	C	D	E	F	G	H
18	14	7	12	5	3	20	2

Відповідно до графу станів, наведеному на рис. 2.1, з початкової вершини *A* можна перейти до станів *B* або *D*. Серед наведених альтернатив найменшим значенням характеризується оцінювальна функція для стану *D* ($h_D=12 < 14=h_B$), тому, відповідно до прийнятої стратегії пошуку, реалізуємо перехід до стану *D*, як показано на рис. 2.7 а. Розкриття вершини *D* породжує переходи до станів *E* та *G*. Серед цих станів найменшим значенням характеризується оцінювальна функція для вершини *E* ($h_E=5 < 9=h_G$), що визначає перехід до стану *E*, як показано на рис. 2.7 б. Далі необхідно здійснити два переходи до станів *H* та *I* відповідно (див. рис. 2.7 в та 2.7 г). Останній перехід призводить до цільового стану, що свідчить про знаходження рішення у вигляді ланцюга переходів *A–D–E–H–I*.

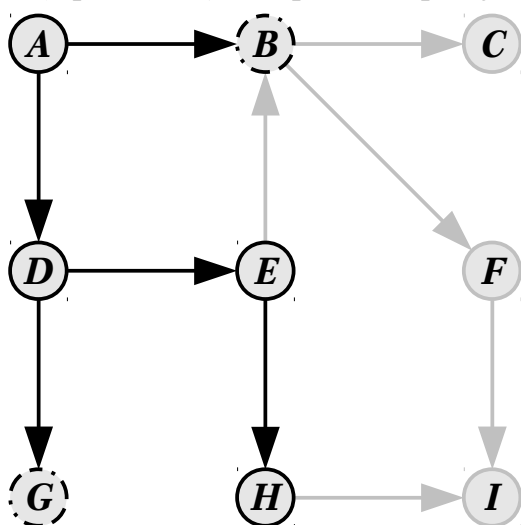
Очевидно, що формалізація градієнтного алгоритму базується на тих самих принципах, що й алгоритму пошуку у глибину, тобто за принципом організації стеку (LIFO). Різниця тут полягає у тому, що перед внесенням у конфліктну множину нові потенційні підграфи, породжені розкриттям поточної вершини, сортують за зменшенням



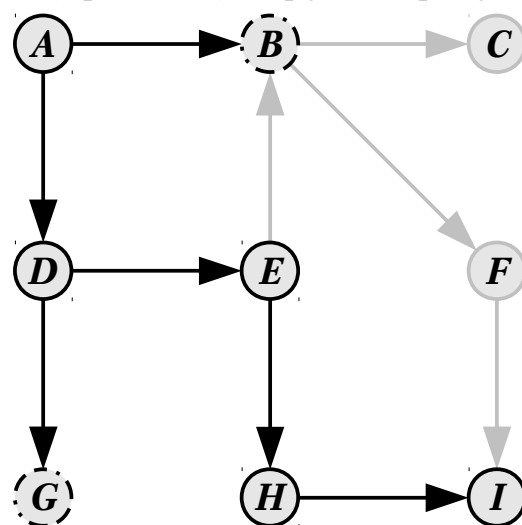
а) реалізація першого кроку



б) реалізація другого кроку



в) реалізація третього кроку



г) реалізація останнього кроку

Рис. 2.7. Ілюстрація пошуку за градієнтним алгоритмом

значень оцінювальної функції для вершин, якими закінчуються відповідні потенційні підграфи. Далі, так само, як і для алгоритму пошуку у глибину для розкриття обирають вершину, якою закінчується останній потенційний граф конфліктної множини.

Алгоритм глобального урахування цілі являє собою модифікацію пошуку у ширину, де розкриття чергової вершини здійснюють виходячи з мінімального значення оцінювальної функції відповідного стану.

Проілюструємо роботу алгоритму глобального урахування цілі на прикладі розв'язання задачі пошуку на графі станів, наведеному на рис. 2.1 із урахуванням значень оцінювальної функції для вершин графу, наведених у табл. 2.2.

Перший крок пошуку виконуємо так само, як за градієнтним алгоритмом, тобто розкриваємо початкову вершину A , що породжує альте-

рнативні переходи до станів B або D . Тут найменшим значенням характеризується оцінювальна функція для стану D ($h_B=12 < 14=h_D$), тому, відповідно до прийнятої стратегії пошуку, реалізуємо перехід до вершини D , що породжує нові альтернативні переходи $D-E$ і $D-G$, як показано на рис. 2.8 а. На наступному кроці для розкриття обирають одну з нерозкритих вершин, суміжних із розкритими, тобто серед вершин B , E та G . Тут найменшим значенням характеризується оцінювальна функція стану E – $g_E=5$, розкриття якої породжує новий альтернативний перехід до стану H , що показано на рис. 2.8 б (перехід $E-B$ не розглядаємо, оскільки вершина B вже знаходиться серед потенційних для розкриття). Далі здійснюють вибір між переходами до станів B , G та H . Тут найменшим значенням характеризується оцінювальна функція стану H –

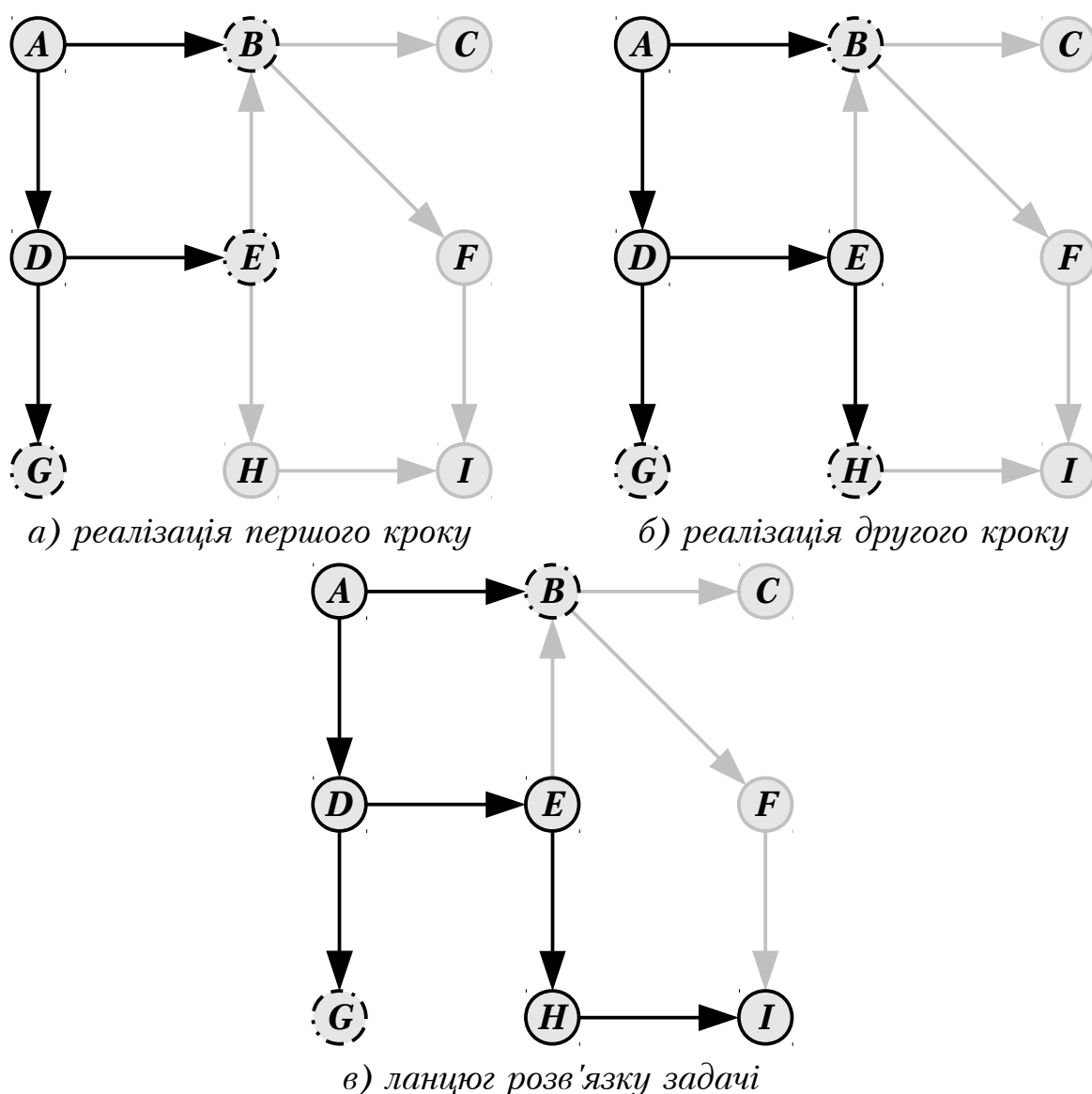


Рис. 2.8. Ілюстрація пошуку за алгоритмом глобального урахування цілі

$g_H=2$, розкриття якої породжує новий альтернативний перехід до цільового стану I , що показано на рис. 2.8 в. Оскільки такий перехід призводить до цільового стану, розв'язком задачі пошуку є ланцюг переходів $A-D-E-H-I$.

Очевидно, що формалізація алгоритму глобального урахування цілі базується на тих самих принципах, що й алгоритму пошуку у ширину, тобто за принципом організації стеку (FIFO). Різниця тут полягає у тому, що після внесення у конфліктну множину всі потенційні підграфи сортують за збільшенням значень оцінювальної функції для вершин, якими закінчуються відповідні потенційні підграфи. Далі, так само, як і для алгоритму пошуку у ширину для розкриття обирають вершину, якою закінчується перший потенційний граф конфліктної множини.

A-алгоритм суміщує в собі алгоритми Дейкстри та глобального урахування цілі. Тут для кожної вершини формують значення оцінювальної функції, яка дорівнює сумі витрат ресурсів на досягнення поточного стану починаючи із вихідного та оцінювальної функції витрат ресурсів на реалізацію переходів від поточного стану до цільового:

$$f_i = g_i + h_i.$$

Наступну вершину для розкриття обирають виходячи з мінімального значення такої оцінювальної функції.

Проілюструємо роботу *A-алгоритму* на прикладі розв'язання задачі пошуку на графі станів, наведеному на рис. 2.1 із урахуванням витрат ресурсів на реалізацію переходів, наведених у табл. 2.1 та значень оцінювальної функції для вершин графу, наведених у табл. 2.2.

На першому етапі необхідно здійснити розкриття вихідної вершини A графу станів, що породжує два альтернативні переходи до станів B та D . Витрати ресурсів на реалізацію цих переходів визначають ціни відповідних вершин:

$$\begin{aligned} g_B &= c_{AB} = 9; \\ g_D &= c_{AD} = 3. \end{aligned}$$

Значення оцінювальної функції *A-алгоритму* тут дорівнює

$$\begin{aligned} f_B &= g_B + h_B = 9 + 14 = 23; \\ f_D &= g_D + h_D = 3 + 12 = 15. \end{aligned}$$

Найменшим значенням оцінювальної функції характеризується вершина D , тому на першому етапі необхідно реалізувати перехід до стану D та реалізувати розкриття цієї вершини. Ця операція визначає оцінювальні функції для станів E та G (див. рис. 2.9 а):

$$\begin{aligned}
 g_E &= g_D + c_{DE} = 3 + 4 = 7; \\
 g_G &= g_D + c_{DG} = 3 + 2 = 5; \\
 f_E &= g_E + h_E = 7 + 5 = 12; \\
 f_G &= g_G + h_G = 5 + 20 = 25.
 \end{aligned}$$

Для розкриття наступної вершини обираємо стан із найменшим значенням оцінювальної функції А-алгоритму серед вершин B , G та E . Такою є вершина E $f_E=12$, розкриття якої дозволяє розрахувати оцінювальну функцію для стану H (див. рис. 2.9 б):

$$\begin{aligned}
 g_H &= g_E + c_{EH} = 7 + 6 = 13; \\
 f_H &= g_H + h_H = 13 + 2 = 15.
 \end{aligned}$$

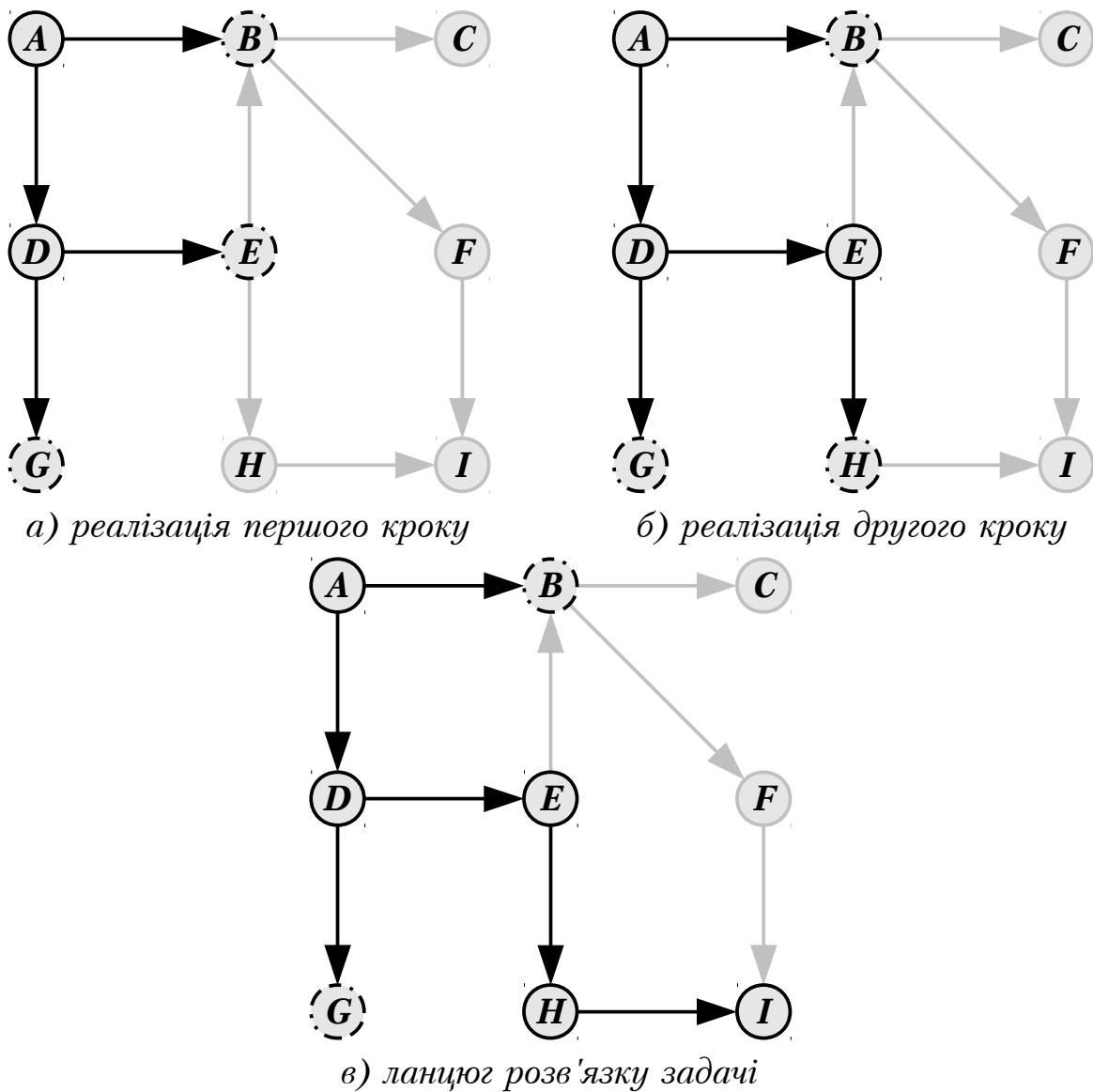


Рис. 2.9. Ілюстрація пошуку за А-алгоритмом

Розкриття наступної вершини H визначає перехід до цільового стану I , тобто розв'язком задачі є ланцюг переходів $A-D-E-H-I$.

Очевидно, що формалізація A -алгоритму базується на тих самих принципах, що й алгоритму пошуку у ширину, тобто за принципом організації стеку (FIFO). Різниця тут полягає у тому, що після внесення у конфліктну множину всі потенційні підграфи сортують за збільшенням значень оцінювальної функції A -алгоритму для вершин, якими закінчуються відповідні потенційні підграфи. Далі, так само, як і для алгоритму пошуку у ширину для розкриття обирають вершину, якою закінчується перший потенційний граф конфліктної множини.

За умови нульових значень оцінювальної функції витрат ресурсів на реалізацію переходів між поточним та цільовим станами A -алгоритм перетворюється на алгоритм Дейкстри. Можна показати, що за умови недооцінювання витрат ресурсів на реалізацію переходів між поточним та цільовим станами, тобто за умови

$$h_i < \sum c_{i-T},$$

де $\sum c_{i-T}$ – сума витрат ресурсів на реалізацію переходів між поточним та цільовим станами, A -алгоритм визначає оптимальний шлях між вихідною та цільовою вершинами. Таку стратегію називають A^* -алгоритмом. A^* -алгоритм буде тим ефективніший, чим ближче значення оцінювальної функції до реальних витрат ресурсів на реалізацію переходів у просторі станів.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає стратегія пошуку за градієнтним алгоритмом?
2. Сформулюйте формальний градієнтний алгоритм.
3. Поясніть положення про те, що градієнтний алгоритм є модифікацією алгоритму пошуку у глибину.
4. Проведіть покрокове розв'язання наведеної задачі пошуку за градієнтним алгоритмом із використанням формалізмів організації спискових структур конфліктної множини.
5. У чому полягає стратегія пошуку алгоритмом глобального урахування цілі?
6. Сформулюйте формальний алгоритм глобального урахування цілі.
7. Поясніть положення про те, що алгоритм глобального урахування цілі є модифікацією алгоритму пошуку у ширину.
8. Проведіть покрокове розв'язання наведеної задачі пошуку за алгоритмом глобального урахування цілі із використанням формалізмів організації спискових структур конфліктної множини.

9. У чому полягає стратегія пошуку за A -алгоритмом?

10. Сформулюйте формальний A -алгоритм.

11. Поясніть положення про те, що A -алгоритм є модифікацією алгоритму пошуку у ширину.

12. Проведіть покрокове розв'язання наведеної задачі пошуку за A -алгоритмом із використанням формалізмів організації спискових структур конфліктної множини.

13. Дайте характеристику A^* -алгоритму.



ЛІТЕРАТУРА

[1], [3], [5], [7], [8], [9], [10], [11], [56], [71]

§2.5. СТРАТЕГІЇ ПОШУКУ У ПРОСТОРІ СТАНІВ

Пошук у просторі станів може бути побудований за однією з двох стратегій:

- пошук під керуванням даних;
- пошук під керуванням цілі.

Стратегія пошуку під керуванням даних передбачає реалізацію пошуку із вихідної вершини у напрямку цільового стану. Тут задача пошуку полягає у визначенні ланцюжка переходів із вихідного стану, який приводить до цільової вершини. Таку стратегію часто називають пошуком із початку в кінець, або прямим ланцюжком міркувань.

Стратегія пошуку під керуванням цілі, навпаки, передбачає реалізацію пошуку починаючи із цільової вершини у напрямку вихідного стану. Тут задача пошуку полягає у визначенні ланцюжка переходів, який приводить до цільового стану із вихідної вершини. Таку стратегію часто називають пошуком із кінця в початок, або зворотним ланцюжком міркувань.

Обидві стратегії пошуку є рівноправними. Вибір стратегії визначається особливостями предметної області експертної системи, прийнятою моделлю формалізації знань та зручністю розв'язання поставлених задач. Так, прямий пошук, зазвичай (але не обов'язково), використовують у продукційних експертних системах, а зворотний пошук – в системах логічного програмування.

Інколи використовують комбіновану (двонаправлену) стратегію пошуку. Тут одночасно виконують пошук із початку в кінець та із кінця в початок. Таким чином формують два дерева пошуку із коренями у ви-

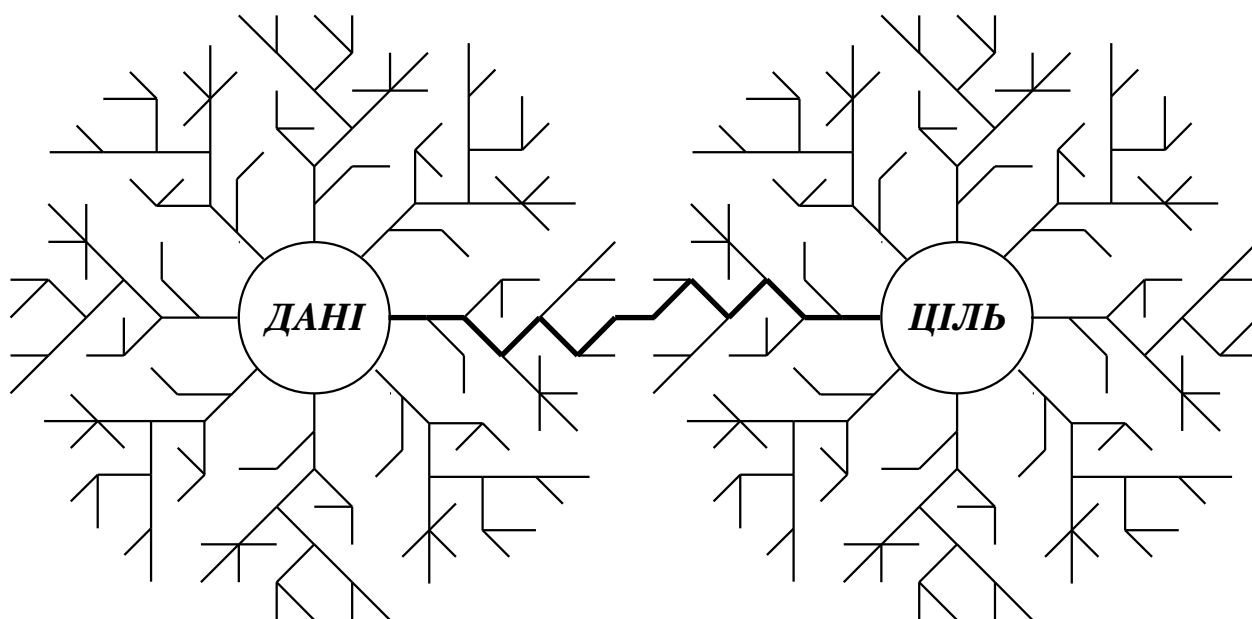


Рис. 2.10. Ілюстрація комбінованої стратегії пошуку

хідній та цільовій вершинами графу простору станів. Пошук закінчують, коли знаходять перетин двох дерев пошуку.

Використання комбінованої стратегії часто дає змогу уникнути проблеми комбінаторного вибуху. Відомо, що проблема комбінаторного вибуху пов'язана із різким зростанням розмірності конфліктної множини із збільшенням глибини пошуку. Використання комбінованої стратегії дає змогу знизити у два рази глибину пошуку за рахунок виконання двох зустрічних пошуків, як показано на рис. 2.10.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає стратегія пошуку під керуванням даними?
2. У чому полягає стратегія пошуку під керуванням мети?
3. У чому полягає комбінована стратегія пошуку?
4. Поясніть основну перевагу використання комбінованої стратегії пошуку.



ЛІТЕРАТУРА

[1], [3], [4], [5], [7], [8], [11], [13], [14], [31], [56], [69]

§2.6. АЛГОРИТМИ ПОШУКУ У ПРОСТОРІ ЗАДАЧ

Пошук у просторі задач реалізують із використанням графів редукції. Розв'язок задачі тут являє собою підграф, корінь якого відповідає вихідній задачі, а кінцеві вершини є термінальними та відповідають тривіальним підзадачам з відомими розв'язками.

Алгоритми пошуку в графах редукції повністю аналогічні алгоритмам пошуку у просторі станів, тобто базуються на стратегіях пошуку у глибину та у ширину, але враховують особливості формування «І-АБО» графів. Це стосується, насамперед, наявності кон'юнктивних вершин (типу «І»), розкриття яких вимагає розширення дерева пошуку одночасно по декількох напрямках, відповідно до змісту підзадач, які необхідно розв'язати одночасно.

В загальному випадку алгоритм пошуку в графі редукції полягає у циклічному виконання наступних процедур:

- 1) перевіряють, чи є вершини потенційного підграфу термінальними, нерозкритими, або тупиковими; за необхідності розраховують ціни відповідних підграфів;
- 2) розширюють потенційний підграф відповідно до прийнятої стратегії пошуку.

Роботу алгоритму закінчують, коли визначають, що усі кінцеві вершини потенційного підграфу є термінальними.

Розглянемо особливості пошуку в графі редукції за алгоритмом пошуку у глибину. Відповідно до стратегії пошуку тут послідовно розширюють один з ланцюгів підграфу редукції вихідної задачі доки не будуть знайдені термінальні або тупикові вершини. У разі виявлення тупикових вершин здійснюють повернення пошуку до останньої нереалізованої альтернативи диз'юнктивної вершини (типу «АБО») та продовжують пошук у новому напрямі.

Проілюструємо роботу алгоритму пошуку у глибину на прикладі графу редукції, наведеному на рис. 2.11. Тут вершині S відповідає вихідна комплексна задача, для розв'язання якої використовують метод декомпозиції. Вершини $A...F$, а також H відповідають проміжним підзадачам, декомпозиція яких породжує нові підзадачі. Вершини G , I та K є тупиковими і відповідають підзадачам, які неможливо розв'язати. Вершини $t_1...t_6$ є термінальними і відповідають тривіальним підзадачам з відомими розв'язками.

Під час реалізації пошуку у глибину будемо послідовно розкривати вершини потенційних підграфів, визначених останніми при розгляданні дерева пошуку зліва праворуч, тобто тих, які останніми потрапили до конфліктної множини (відповідно до принципу LIFO).

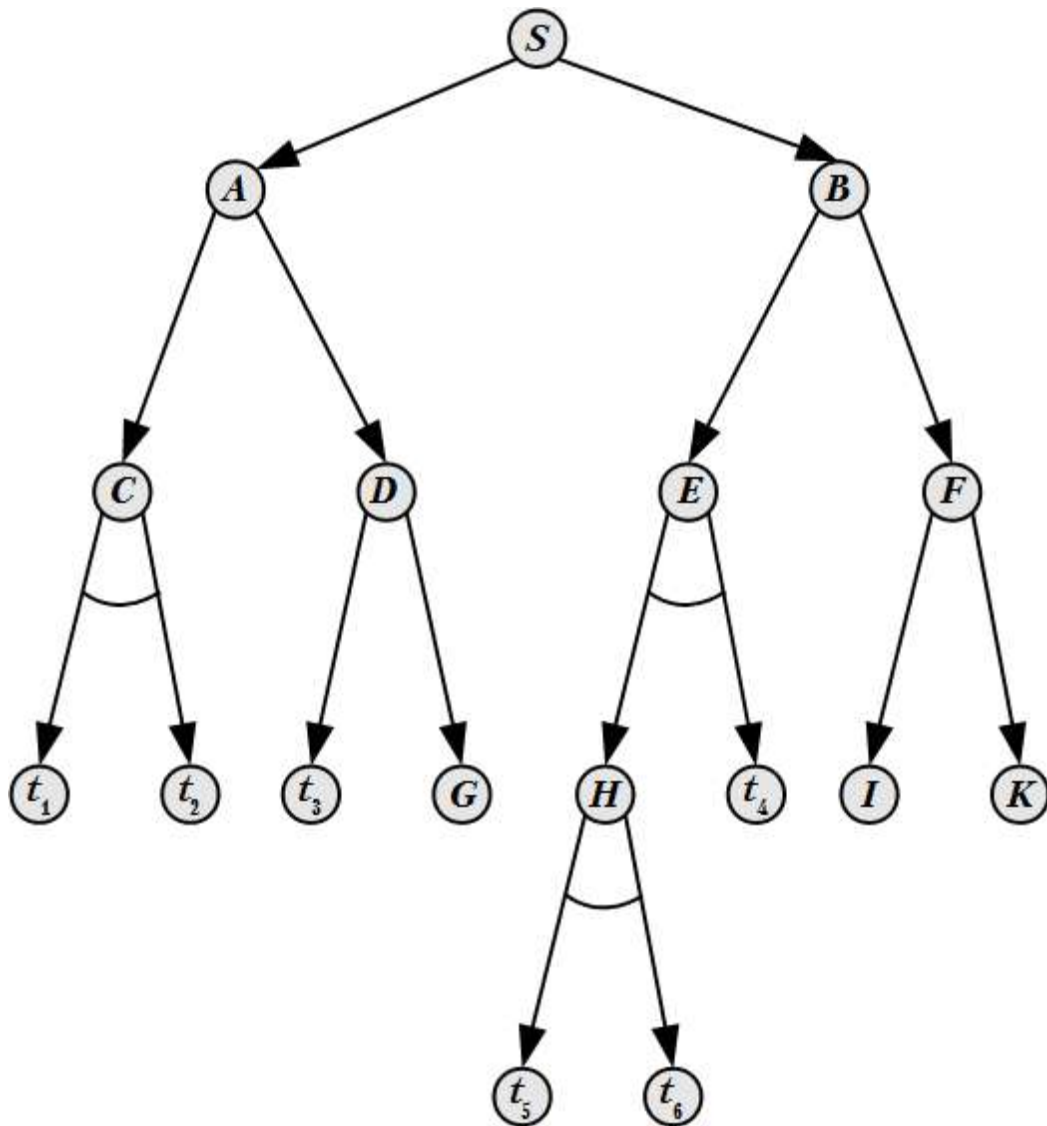


Рис. 2.11. Граф редукції складної задачі

На рис. 2.12 проілюстровано зміну складу конфліктної множини в процесі розв'язання задачі пошуку. На початку роботи алгоритму до складу конфліктної множини вносять кореневу вершину $\{S\}$, яка відповідає вихідній задачі (див. рис. 2.12, крок 1). Така вершина є диз'юнктивною і її розкриття породжує два потенційні підграфи $\{S-A\}$ і $\{S-B\}$ (див. рис. 2.12, крок 2). Наступною слід розкривати вершину B , яка є кінцевою останнього потенційного підграфу конфліктної множини. Така вершина також є диз'юнктивною. Її розкриття породжує ще два нових потенційних підграфи $\{S-B-E\}$ та $\{S-B-F\}$ (див. рис. 2.12, крок 3). Кінцева вершина останнього потенційного підграфу конфліктної множини F знов є кон'юнктивною. Внаслідок розкриття цієї вершини конфліктна множина поповнюється двома потенційними підграфами $\{S-B-F-I\}$ та $\{S-B-F-K\}$ (див. рис. 2.12, крок 4). Кінцеві вершини двох останніх потенційних підграфів конфліктної множини (I та K) є тупиковими. Тому

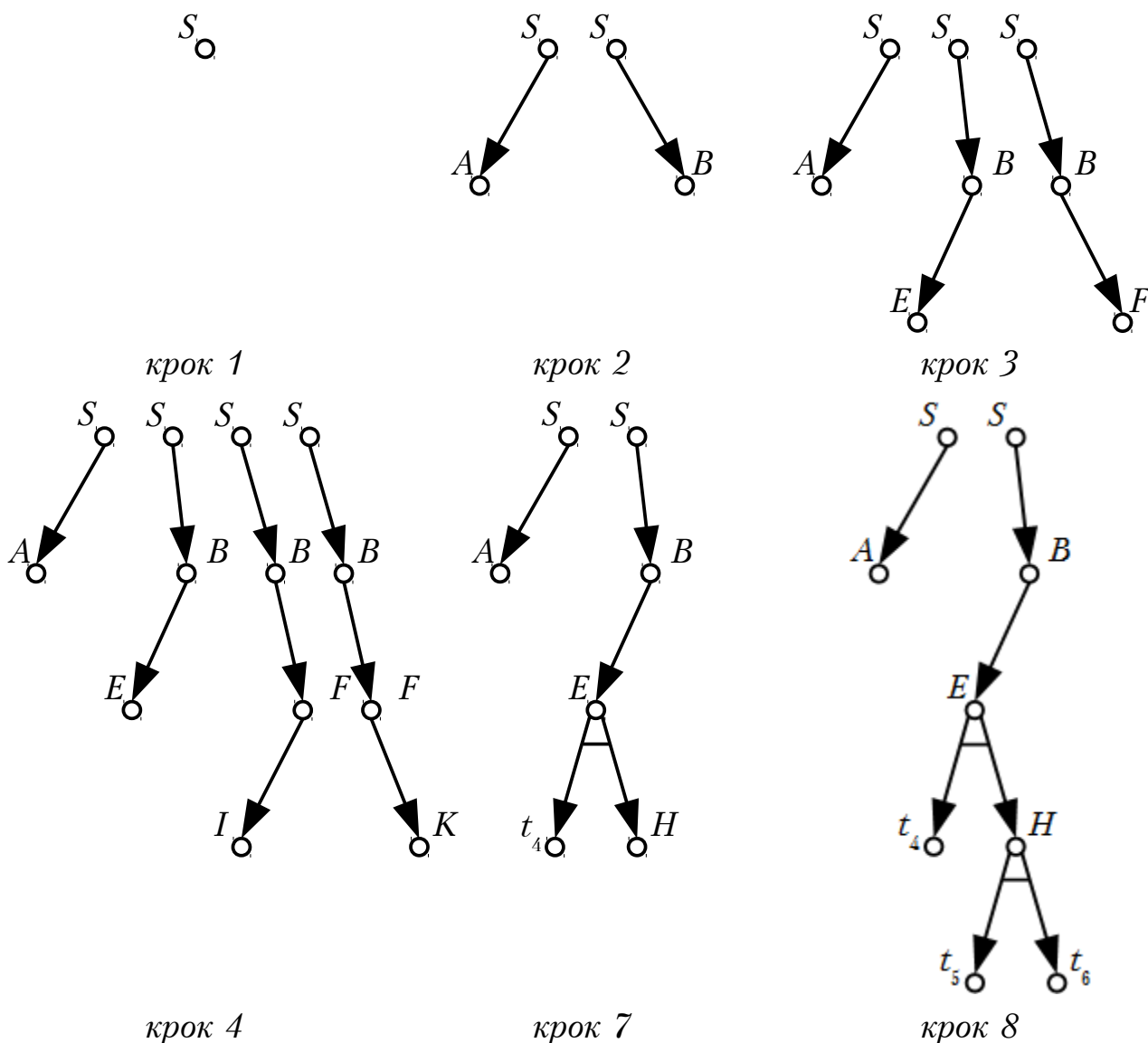


Рис. 2.12. Зміна складу конфліктної множини під час пошуку у глибину. Два наступні кроки алгоритму пов'язані лише із видаленням цих потенційних підграфів з конфліктної множини. Наступний крок алгоритму пошуку пов'язаний із розкриттям кон'юнктивної вершини E . Така операція породжує розгалужений потенційний підграф $\{S-B-E \begin{smallmatrix} t_4 \\ H \end{smallmatrix}\}$ (див. рис. 2.12, крок 7).

Аналіз цього підграфу, який стає останнім в конфліктній множині свідчить про те, що для розв'язання цільової задачі необхідно розв'язати підзадачі, яким відповідають вершини t_4 та H дерева пошуку. Вершина t_4 є термінальною і подальший пошук пов'язаний із розкриттям вершини H . Ця вершина також є кон'юнктивною. Її розкриття визначає подальше розгалуження потенційного підграфу до форми

$\{S-B-E \begin{smallmatrix} \swarrow t_4 \\ \searrow t_6 \end{smallmatrix} H \begin{smallmatrix} \swarrow t_5 \end{smallmatrix} \}$ (див. рис. 2.12, крок 8). Всі кінцеві вершини отримано-

го підграфу є термінальними, що свідчить про досягнення мети пошуку.

Таким чином, в результаті розв'язання задачі пошуку було визначено, що для розв'язання цільової задачі необхідно одночасно розв'язати тривіальні підзадачі, яким відповідають термінальні вершини t_4 , t_5 та t_6 .

Аналогічним чином реалізують пошук у графі редукції задачі за алгоритмом пошуку у ширину. Тут реалізують послідовне розкриття вершин «І-АБО» графу першого, другого, третього ярусу і так далі доки всі кінцеві вершини одного з потенційних підграфів не виявляться термінальними.

Проілюструємо роботу алгоритму пошуку у ширину на графі редукції, наведеному на рис. 2.11. Тут вибір чергової вершини з конфліктної множини для розкриття виконують виходячи із стратегії FIFO, тобто кожного разу обирають перший потенційний підграф, а результат розкриття вершин записують в кінець конфліктної множини. Заповнення конфліктної множини будемо проводити послідовно, проглядаючи дерево пошуку зліва праворуч.

На рис. 2.13 проілюстровано зміну складу конфліктної множини в процесі розв'язання задачі пошуку. На першому етапі роботи алгоритму до складу конфліктної множини вносять кореневу вершину $\{S\}$, яка відповідає вихідній задачі (див. рис. 2.13, крок 1). Розкриття диз'юнктивної вершини S породжує два потенційні підграфи $\{S-A\}$ і $\{S-B\}$, які вносять до складу конфліктної множини (див. рис. 2.13, крок 2). На наступному етапі роботи алгоритму здійснюють розкриття диз'юнктивної вершини A , якою закінчується перший потенційний підграф конфліктної множини. Така операція породжує два нові потенційні підграфи $\{S-A-C\}$ і $\{S-A-D\}$ (див. рис. 2.13, крок 3). В такий самий спосіб здійснюють розкриття диз'юнктивної вершини B , якою закінчується перший потенційний підграф конфліктної множини, що породжує потенційні підграфи $\{S-B-E\}$ і $\{S-B-F\}$ (див. рис. 2.13, крок 4). Далі необхідно реалізувати розкриття кінцевої вершини C першого потенційного підграфу. Така вершина є кон'юнктивною і її розкриття породжує розгалужений

потенційний підграф $\{S-A-C \begin{smallmatrix} \swarrow t_1 \\ \searrow t_2 \end{smallmatrix} \}$ (див. рис. 2.13, крок 5). Обидві кі-

нцеві вершини отриманого підграфу є термінальними, що свідчить про завершення роботи алгоритму.

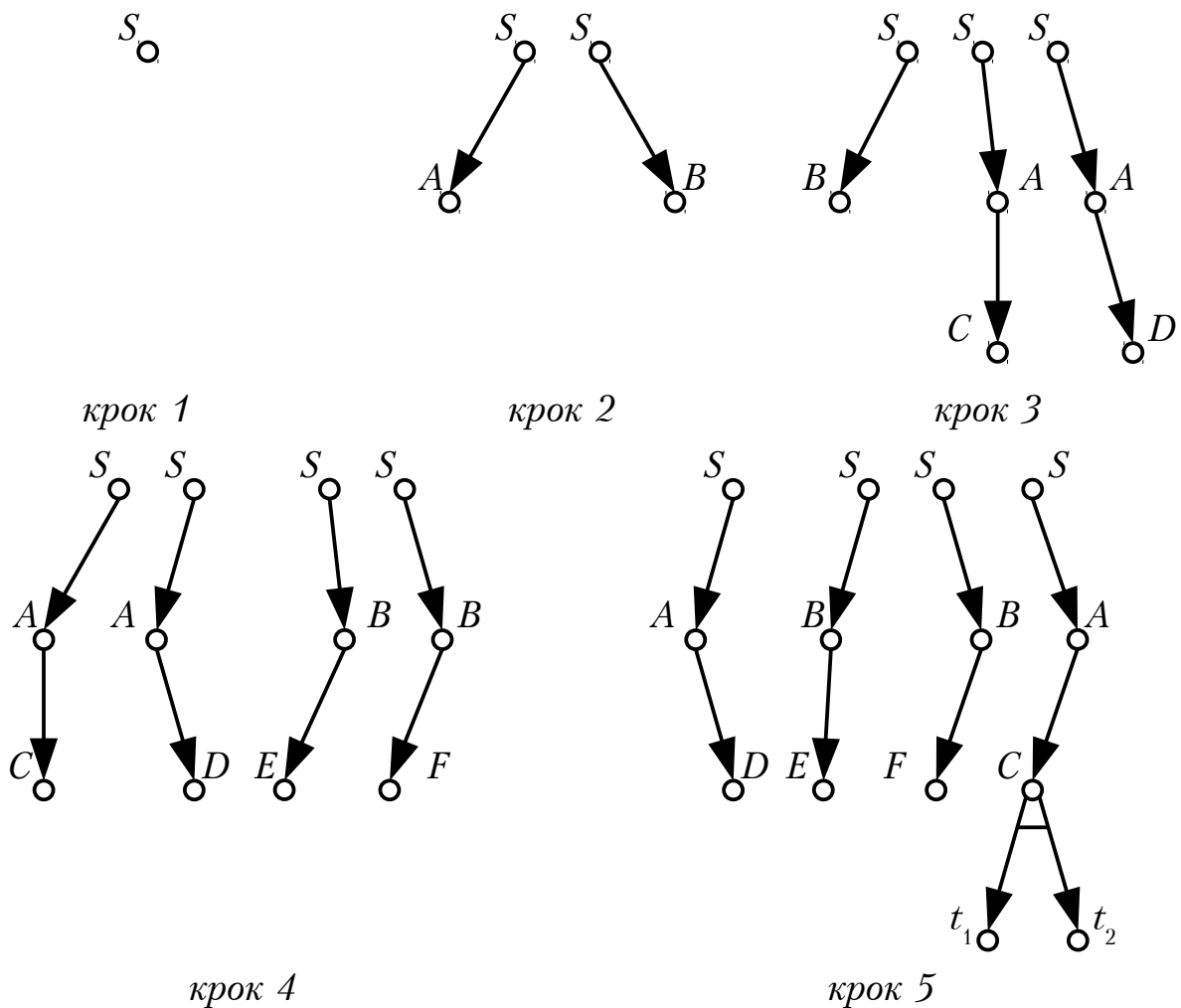


Рис. 2.13. Зміна складу конфліктної множини під час пошуку у ширину

Таким чином, в результаті розв'язання задачі пошуку було визначено, що для розв'язання цільової задачі необхідно одночасно розв'язати тривіальні підзадачі, яким відповідають термінальні вершини t_1 та t_2 .

Евристичний пошук в графі редукції базується на використанні оцінювальних функцій для вершин графу пошуку. Тут кожній вершині «І-АБО» графу ставлять у відповідність деяке додатне число, за допомогою якого оцінюють витрати ресурсів на декомпозицію та розв'язання відповідної підзадачі. Для термінальних вершин оцінювальна функція дорівнює нулю, а для тупикових – є нескінченно великою. В процесі розкриття вершин дерева редукції на кожному кроці обирають вершину з найменшим значенням оцінювальної функції.

Проілюструємо роботу евристичного алгоритму на прикладі графу редукції, наведеного на рис 2.11 з урахуванням значень оцінювальної функції, наведених в табл. 2.3.

Таблиця. 2.3. Значення оцінювальної функції вершин дерева редукції

<i>S</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>H</i>	<i>G, I, K</i>	$t_1...t_6$
12	6	9	4	2	5	8	2	∞	0

На рис. 2.14 проілюстровано зміну складу конфліктної множини в процесі розв'язання задачі пошуку. Тут на першому кроці роботи алгоритму до складу конфліктної множини вносять кореневу вершину дерева пошуку $\{S\}$, яка відповідає вихідній задачі (див. рис. 2.14, крок 1). Вершина S є диз'юнктивною, її розкриття породжує два потенційні підграфи $\{S-A\}$ і $\{S-B\}$, які вносять до складу конфліктної множини (див. рис. 2.14, крок 2). Серед кінцевих вершин потенційних підграфів конфліктної множини найменшим значенням оцінювальної функції характеризується вершина A ($h_A=6$). Тому на наступному етапі здійснюють розкриття саме цієї вершини. Така операція породжує два нові потенційні підграфи $\{S-A-C\}$ і $\{S-A-D\}$ (див. рис. 2.14, крок 3). Далі необхідно реалізувати розкриття кінцевої вершини D , оскільки ця вершина характеризується найменшим значенням оцінювальної функції серед кінцевих вершин потенційних підграфів конфліктної множини ($h_D=2$). Така вершина є диз'юнктивною і її розкриття породжує підграфи $\{S-A-D-t_3\}$ і $\{S-A-D-G\}$ (див. рис. 2.14, крок 4). Один з цих підграфів закінчується термінальною вершиною t_3 , що визначає завершення роботи алгоритму.

Таким чином, в результаті розв'язання задачі пошуку було визначено, що розв'язання цільової задачі зводиться до розв'язання елементарної підзадачі, якій відповідає термінальна вершина t_3 .

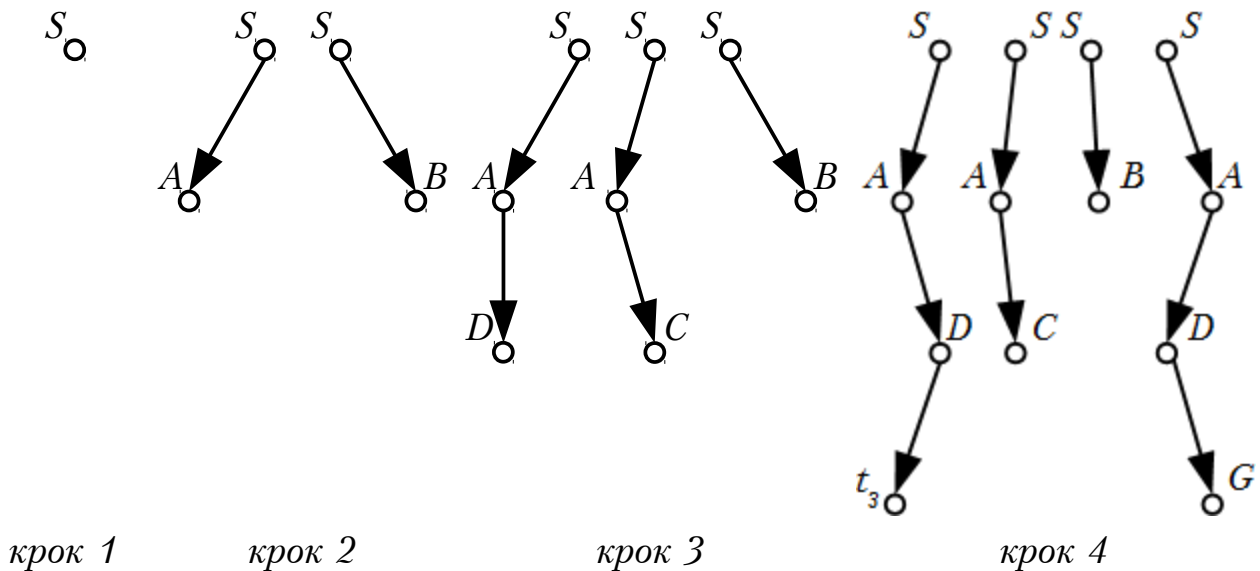


Рис. 2.14. Зміна складу конфліктної множини під час евристичного пошуку

Очевидно, що залучення оцінювання можливих витрат ресурсів на декомпозицію та розв'язання задач та породжуваних підзадач дозволяє суттєво скоротити простір пошуку та ефективно розв'язувати складні комплексні задачі.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть узагальнений алгоритм пошуку у графі редуції.
2. Наведіть особливості застосування алгоритму пошуку у глибину для графу редуції задачі.
3. Наведіть особливості застосування алгоритму пошуку у ширину для графу редуції задачі.
4. Наведіть особливості застосування евристичного алгоритму пошуку для графу редуції задачі.



ЛІТЕРАТУРА

[1], [2], [10], [9], [62], [63]

§2.7. ГЕНЕТИЧНІ АЛГОРИТМИ ПОШУКУ

Практичне застосування розглянутих раніше алгоритмів пошуку оптимальних рішень у просторі станів обмежене наступними факторами. Методи сліпого пошуку гарантують виявлення оптимального розв'язку поставленої задачі лише після повного перебору всіх можливих станів. Це пов'язано із проблемою комбінаторного вибуху, коли через велику розмірність простору пошуку, розв'язання задачі вимагає забагато часу та витрат ресурсів. Застосування традиційних оптимізаційних алгоритмів, наприклад, методу щонайскорішого спуску, зазвичай, пов'язано із додатковими вимогами, які висуваються до форми цільової функції задачі оптимізації. Виконання таких вимог не завжди можна забезпечити, особливо в умовах багатокрітеріального простору станів, де розв'язання задачі вимагає забезпечення оптимальності взаємоконфліктуючих, а, інколи, ще й несумірних факторів. Найбільш поширений підхід розв'язання таких задач базується на штучному обмеженні простору пошуку та використанні методів сліпого або евристичного пошуку серед обмеженого набору альтернативних варіантів можливих розв'язків поставленої задачі. Використання такого підходу характеризується потенційними загрозами,

пов'язаними з підвищеною ймовірністю відбракування найефективнішого рішення ще на попередньому етапі обмеження простору пошуку.

Зазначених недоліків певною мірою позбавлені методи, які базуються на апараті генетичних алгоритмів. Останні реалізують пошук у просторі станів, який базується на принципах біологічної еволюції, тобто виживання найсильніших особин у популяції в процесі її розвитку. Основна ідея тут полягає у моделюванні механізмів еволюційного розвитку біологічних організмів, забезпечених методами природного відбору та генетичного спадкування властивостей. Пристосування до умов існування, у загальному випадку, можна розглядати як задачу оптимізації щодо вдосконалення певних властивостей протягом декількох поколінь біологічних організмів.

Відповідно до принципів застосування генетичних алгоритмів, інформацію про характеристики та властивості складних штучних систем подають у вигляді кодових ланцюжків (зазвичай двійкових), які певною мірою відповідають спрощеній моделі геному біологічних організмів. Це дозволяє застосувати основні генетичні процедури для породження нових моделей штучних систем, які успадковують властивості відомих рішень, але, водночас, відрізняються від них. В цілому еволюційний процес призводить до покращення властивостей породжуваних моделей, що дозволяє реалізувати розв'язання поставленої задачі оптимізації.

Генетичний алгоритм – це еволюційний алгоритм пошуку у просторі станів, призначений для розв'язання задач оптимізації та моделювання шляхом випадкового підбору, комбінування та варіації шуканих параметрів з використанням механізмів, які моделюють біологічну еволюцію

В загальному випадку розв'язання практичних задач з використанням генетичних алгоритмів складається з наступних етапів:

1) кодування параметрів потенційних розв'язків поставленої задачі та представлення їх у вигляді, зручному для виконання еволюційних операцій змінювання та відбору;

2) реалізація генетичних операторів схрещування та мутацій, направлених на породження нових рішень з рекомбінованими характеристиками батьківських пар;

3) відбір найкращих рішень, параметри яких найбільш відповідають умовам поставленої задачі для виживання та формування батьківських пар з метою породження нового покоління потенційних рішень задачі пошуку.

Процедури 2 та 3 повторюють багатократно аж до досягнення прийнятного (оптимального або близького до нього) розв'язку поставленої задачі.

Розглянемо представлені етапи генетичних алгоритмів та проблеми, пов'язані з їх виконанням більш детально.

Кодування генетичної інформації полягає у визначенні сукупності параметрів, які характеризують потенційні рішення поставленої задачі та представленні їх у певному вигляді. Тут, зазвичай, використовують бінарні ланцюжки двійкових кодів визначених параметрів, як представлено, наприклад, на рис. 2.15. На рис. 2.15 показано приклад кодування інформації про потенційні рішення деякої задачі, яка складається з 6 параметрів. Це означає, що відповідний генотип складається з 6 хромосом, кожна з яких представляє закодовану інформацію про конкретне значення відповідного параметру. Перша з хромосом являю собою 7-бітний ланцюжок двійкового коду. Тобто перша хромосома складається з 7 генів, кожен з яких може приймати одне з двох можливих значень «0» або «1». Довжина бінарного ланцюжка з 7 бітів визначає можливість кодування 2^7 значень першого параметру потенційного розв'язку задачі. В такий самий спосіб визначаються інші хромосоми, які визначають кодування інших параметрів розв'язку. Зазначимо, що довжини бінарних ланцюжків окремих хромосом можуть відрізнятися відповідно до необхідної інформаційної ємності відповідного параметру.

1011001	000010101	10101011	111000101	00011	10000101
⏟	⏟	⏟	⏟	⏟	⏟
параметр 1	параметр 2	параметр 3	параметр 4	параметр 5	параметр 6

Рис. 2.15. Кодування генетичної інформації бінарними ланцюжками

Основна операція, направлена на породження нових потенційних розв'язків задачі пошуку методами генетичних алгоритмів полягає у схрещуванні генотипів батьківських пар (кросинговер). Операцію кросинговеру умовно представлено на рис. 2.16 а. Тут випадковим чином обирають точку розриву геномів батьківських пар та реалізують обмін частинами інформаційних ланцюжків. Один з отриманих генотипів в подальшому розглядають як дочірній, яким поповнюють популяцію потенційних розв'язків задачі. Крім одноточкового кросинговеру, представлено на рис. 2.16 а, апарат генетичних алгоритмів передбачає також виконання двоточкового та багатоточкового схрещування. Двоточковий кросинговер умовно представлено на рис. 2.16 б. Тут випадковим чином обирають дві точки розриву батьківських генотипів та реалізують обмін центральною частиною інформаційного ланцюжка даних. У випадку багатоточкового кросинговеру дочірній генотип визначають внаслідок повного перемішування батьківських генотипів і випадкового вибору значення кожного гену всіх хромосом. Багатоточковий кросинговер умовно представлено на рис. 2.16 в.

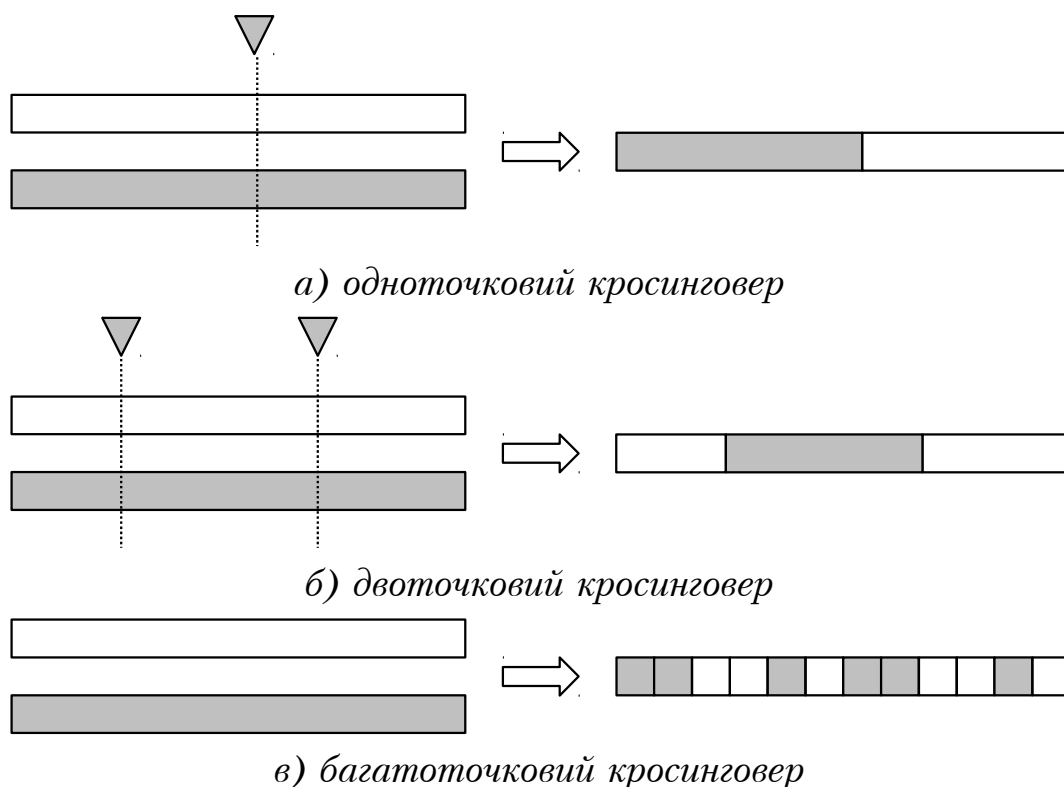


Рис. 2.16. Ілюстрація операції кросингверу

Крім операцій схрещування генетичні алгоритми передбачають інші операції зміни генотипів у популяції – мутації. Виконання таких операцій направлене, зазвичай, на виведення процесу генетичного пошуку з можливих локальних оптимумів шляхом випадкової зміни генотипів породжуваного покоління потенційних розв'язків поставленої задачі.

Проста мутація полягає у випадковій зміні значення одного гену в інформаційному генетичному ланцюжку породженого рішення. Операцію простої мутації умовно проілюстровано на рис. 2.17 а.

Інший вид мутації – інверсія – полягає у випадковій зміні послідовності генів у генотипі породженого рішення. Виконання такої операції умовно проілюстровано на рис. 2.17 б.

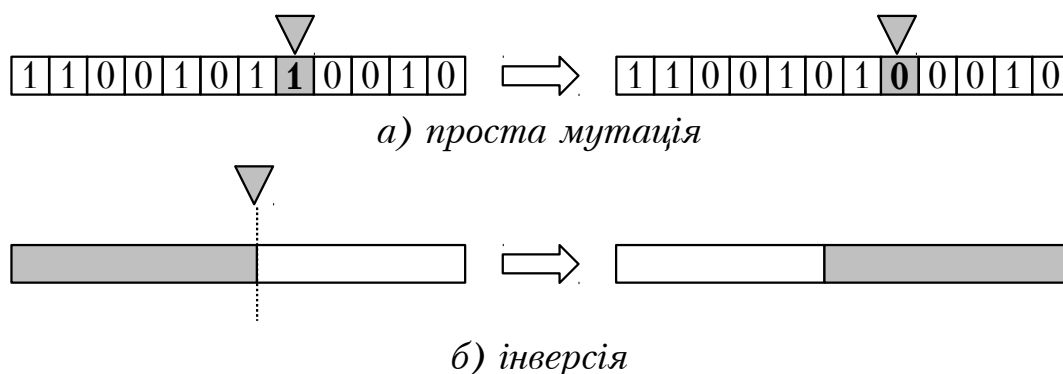


Рис. 2.17. Ілюстрація операцій мутації

Реалізація цілеспрямованого генетичного пошуку оптимального рішення поставленої задачі базується на організації селекції, тобто вибору найбільш ефективних потенційних розв'язків та формування з них батьківських пар з метою породження нових рішень, які успадковують властивості «батьків». Тут слід розглядати дві основні проблеми:

1) задача оцінювання ефективності породжуваних потенційних розв'язків задачі;

2) методика формування батьківських пар.

Для оцінювання ефективності породжуваних рішень використовують, так звані, фітнес-функції. Аргументами таких функцій виступають декодовані значення параметрів пропонованих рішень, а результатом виконання – кількісні показники якості рішень.

Суть селективного відбору батьківських пар полягає в тому, що пріоритет для подальшого схрещування мають потенційні рішення, які характеризуються найбільшими значеннями фітнес-функції. Загалом можна запропонувати безліч способів формування батьківських пар, проте у практиці застосування генетичних алгоритмів найбільшого поширення набули методи організації турнірів та рулеткового відбору.

Метод турнірів полягає у тому, що в поточній популяції потенційних рішень обирається два чи більше елементів і для подальшого схрещування серед них обирають рішення з найбільшим значенням фітнес-функції. У свою чергу, ідея рулеткового відбору полягає в тому, що батьківські пари обирають за допомогою генератора випадкових чисел. При цьому ймовірність використання генотипу деякого потенційного рішення для подальшого схрещування пропорційна значенню фітнес-функції для цього елементу популяції.

Розглянемо застосування генетичного алгоритму на прикладі розв'язання транспортної задачі забезпечення живлення споживачів деякого району. На рис. 2.18 представлено надлишкову схему електричної мережі, що забезпечує живлення споживачів у пунктах 1, 2 і 3 від електростанції, розташованої в пункті 0. На схемі вказані також навантаження споживачів в мегаватах і відстані між підстанціями в кілометрах. Розв'язання задачі полягає у виборі мінімальної оптимальної конфігурації електричної мережі та визначенні перерізів проводів на ділянках схеми.

Інформаційний генетичний ланцюжок проекрованої електричної мережі складається з шести ділянок – хромосом, кожна з яких несе інформацію про конструктивне виконання відповідної лінії електропередачі. Під конструктивним виконанням лінії будемо розуміти марку і переріз проводів на відповідній ділянці мережі. Як показано в табл. 2.4 така інформація може бути представлена одним з восьми мож-

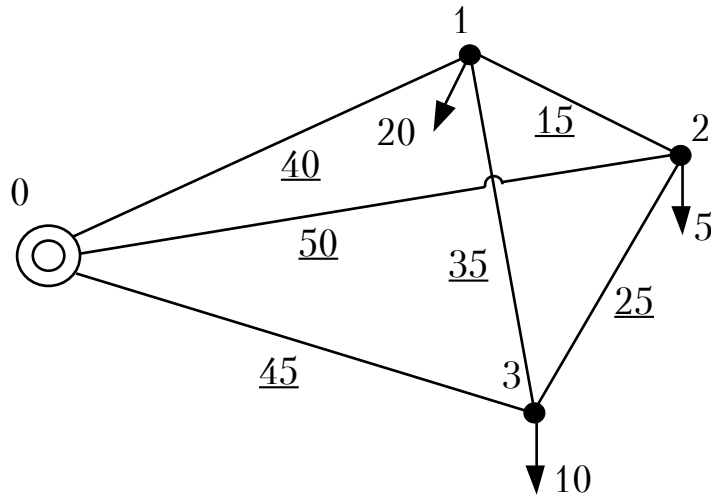


Рис. 2.18. Надлишкова схема електричної мережі

ливих значень. Тут коду «000» відповідає відсутність лінії електропередач на відповідній ділянці між підстанціями.

Таблиця. 2.4. Кодування генетичної інформації про проектувану електричну мережу

№ з/п	Марка проводу	Індекс	Код
1	—	0	000
2	АС 70/11	1	001
3	АС 95/16	2	010
4	АС 120/19	3	011
5	АС 150/24	4	100
6	АС 185/29	5	101
7	АС 240/32	6	110
8	2×АС 240/32	7	111

Ефективність проектного рішення будемо оцінювати за величиною сумарних зведених дисконтованих витрат на спорудження електричної мережі та покриття втрат потужності на передачу електричної енергії по лініях. Для спрощення задачі та підвищення її наочності, будемо нехтувати витратами на експлуатацію лінії, а також постійною складовою втрат. У свою чергу, фітнес-функція генетичного алгоритму дорівнює зверненій величині зведених дисконтованих витрат:

$$F = \left(\sum_{i=1}^6 3_{0i} l_i \right)^{-1} = \left(\sum_{i=1}^6 (K_{0i} E + I_i^2 r_{0i} \tau C) l_i \right)^{-1},$$

де Z_{0i} – погонні значення зведених дисконтованих витрат; K_{0i} – погонні значення капіталовкладень у будівництво ліній електропередач; E – норма дисконту; I_i – робочий струм в лінії в режимі максимальних навантажень; r_{0i} – погонні активні опори ліній електропередач; τ – час максимальних витрат; C – тариф на покриття витрат потужності в електричній мережі; l_i – довжина ліній електропередач.

Тут для кожного можливого конструктивного виконання ділянок електричної мережі зведені дисконтовані витрати визначаються поліномами:

$$\begin{aligned} Z_{70} &= (27,5 + 0,015 I_p^2) l; & Z_{95} &= (28,5 + 0,011 I_p^2) l; \\ Z_{120} &= (29,5 + 8,78 \cdot 10^{-3} I_p^2) l; & Z_{150} &= (31 + 7,34 \cdot 10^{-3} I_p^2) l; \\ Z_{185} &= (32,5 + 5,72 \cdot 10^{-3} I_p^2) l; & Z_{240} &= (35 + 4,25 \cdot 10^{-3} I_p^2) l; \\ Z_{2 \times 240} &= (50 + 2,12 \cdot 10^{-3} I_p^2) l. \end{aligned}$$

Визначимо початкову популяцію з десяти випадкових рішень, параметри яких наведено в табл. 2.5.

Таблиця. 2.5. Дані початкової популяції конструктивних рішень

№	Код	F	№	Код	F
1	553614	1/110167	6	751327	1/106390
2	113331	1/150229	7	424111	1/119773
3	464064	1/108956	8	354451	1/116127
4	125075	1/137697	9	656110	1/95320
5	700375	1/93356	10	532560	1/113378

Тут під кодом варіанту спорудження електричної мережі будемо розуміти кодування конструктивного виконання повітряних ліній електропередач згідно даних табл. 2.4. Відповідно до даних табл. 2.5 найкращим серед запропонованих є рішення 5, зведені дисконтовані витрати для якого дорівнюють 93356 тис. грн.

Для визначення батьківських пар наступного покоління проектних рішень будемо застосовувати рулетковий підхід, проілюстрований на діаграмі рис. 2.19. Тут кожному елементу поточної популяції відповідає сектор «рулетки», ширина якого звернено пропорційна значенню фітнес-функції. Такий підхід забезпечує формування батьківських пар таким чином, щоб найкращі рішення з більшою ймовірністю потрапляли до батьківського пулу.

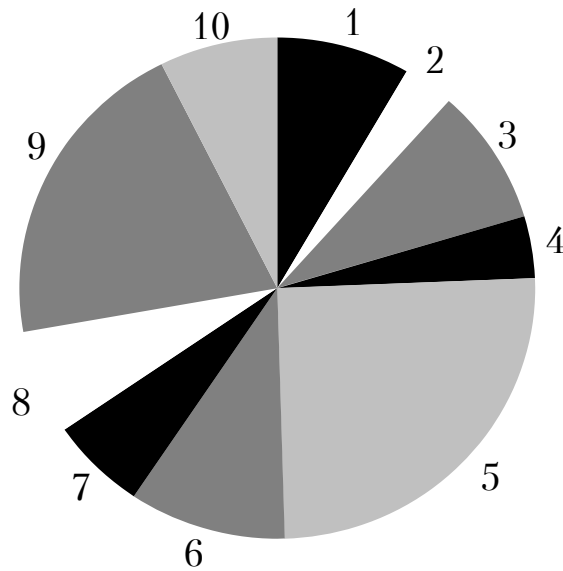


Рис. 2.19. Ілюстрація рулеткового відбору

За допомогою генератора випадкових чисел визначимо першу батьківську пару. Нехай, наприклад, до батьківського пулу потрапляють варіанти спорудження електричної мережі 5 та 8. Після виконання операції односточкового кросинговеру після, наприклад, п'ятого біту інформаційного ланцюжка отримуємо нове проектне рішення 340375, як показано на рис. 2.20.

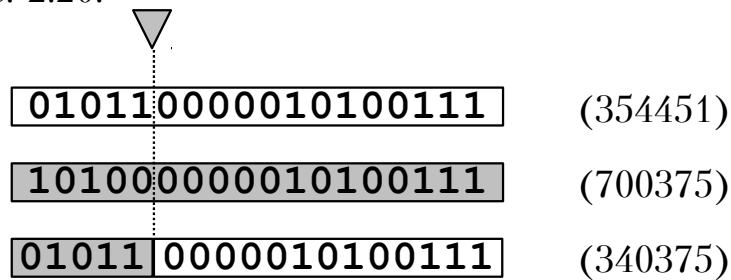


Рис. 2.20. Ілюстрація операції кросинговеру

В такий самий спосіб отримуємо дев'ять нових проектних рішень, які заміщують всі елементи популяції, крім найкращого. Результати виконання першого циклу генетичного алгоритму зведено до табл. 2.6.

Таблиця. 2.6. Дані другого покоління проектних рішень

№	Код	F	№	Код	F
1	340375	1/139199	6	532564	1/119198
2	464064	1/108956	7	751214	1/100961
3	424105	1/111654	8	764064	1/95449
4	654064	1/101576	9	553616	1/110989
5	453614	1/114503	10	700375	1/93356

Аналіз даних табл. 2.6 свідчить про те, що виконання першого циклу генетичного алгоритму ще не породжує кращого проектного рішення, проте друге покоління рішень характеризується меншими максимальним та середнім значеннями зведених дисконтованих витрат. В такий самий спосіб виконуємо генетичні операції з генерації нових поколінь проектних рішень, що проілюстровано на діаграмі рис. 2.21.

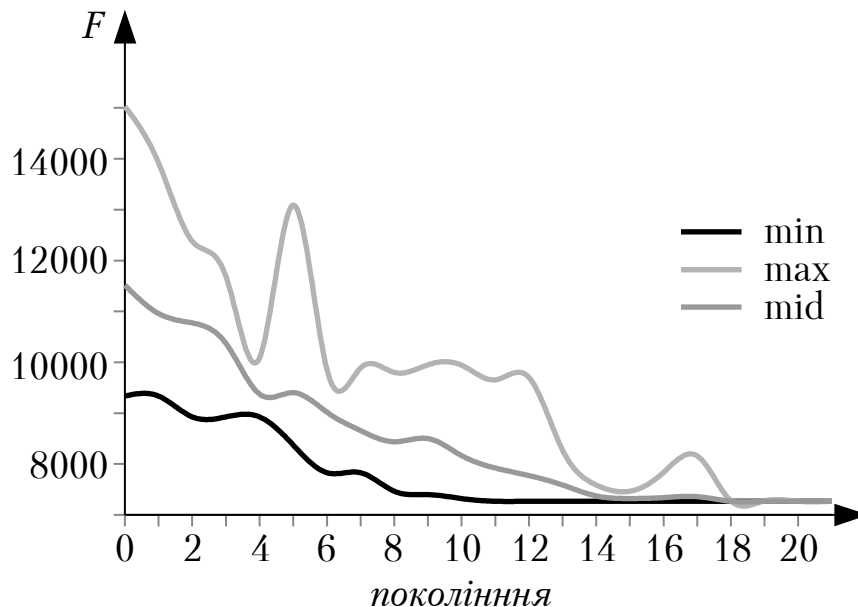


Рис. 2.21. Ілюстрація виконання генетичного алгоритму

Відповідно до даних рис. 2.21 після виконання 20 ітерацій генетичний алгоритм збігається до оптимального рішення 703600, схема якого наведена на рис. 2.22, зі зведеними дисконтованими витратами 72671 тис. грн.

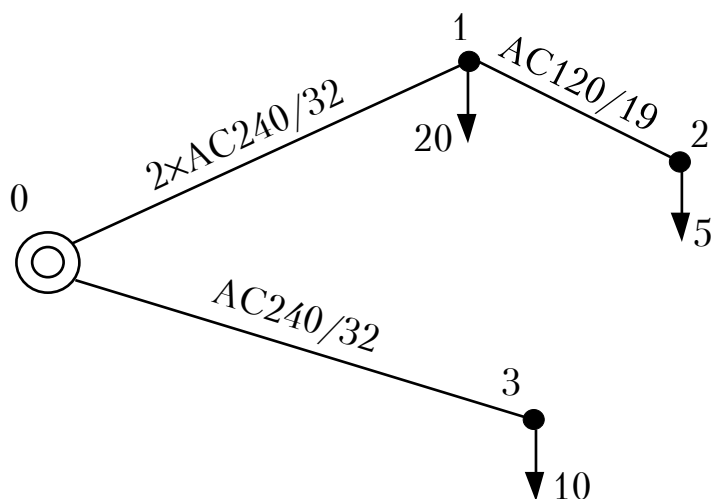


Рис. 2.22. Оптимальна схема електричної мережі

Аналіз розв'язання задачі за допомогою апарату генетичних алгоритмів дозволяє зробити такі висновки. Інформаційний ланцюжок складається з 6 хромосом по три біти кожна, що визначає розмір простору пошуку з 2^{18} можливих станів. Генетичний алгоритм збігся за 20 ітерацій, що потребувало проглядання менше 200 варіантів спорудження електричної мережі (менше 0,1% простору пошуку). При цьому не було задіяно жодної додаткової експертної інформації щодо доцільності можливих проектних рішень.

Загалом, до позитивних властивостей апарату генетичних алгоритмів слід віднести:

- простоту та прозорість реалізації;
- можливість розпаралелювання процесу пошуку;
- простота кодування та декодування інформації;
- знижена ймовірність зациклювання пошуку поблизу локального мінімуму.

Разом з тим, генетичні алгоритми не позбавлені ряду недоліків:

- висока ітеративність алгоритму;
- суттєва залежність ефективності пошуку від параметрів генетичного алгоритму;
- висока ймовірність передчасної збіжності циклічного процесу пошуку.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть визначення генетичного алгоритму.
2. Перелічіть основні етапи генетичних алгоритмів.
3. Поясніть способи кодування генетичної інформації технічних рішень.
4. Поясніть сутність генетичних операцій кросинговеру та мутації.
5. Поясніть сутність селекції.
6. Дайте визначення фітнес-функції.
7. Перелічіть та охарактеризуйте методи формування батьківських пар в апараті генетичних алгоритмів.
8. Перелічіть позитивні та негативні риси апарату генетичних алгоритмів.



ЛІТЕРАТУРА

[7], [8], [66]

МОДЕЛІ ФОРМАЛІЗАЦІЇ ЗНАНЬ В ЕКСПЕРТНИХ СИСТЕМАХ

§3.1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ЗАДАЧ ІНЖЕНЕРІЇ ЗНАНЬ

З розробкою і використанням експертних систем тісно пов'язані такі поняття, як «знання» і «бази знань». Взагалі, термін «знання» являє собою складну філософську категорію, яка визначає, у самому широкому сенсі, результат систематизації пізнавальної діяльності людини. Разом з тим, у контексті інтелектуальних комп'ютерних систем знання можна визначити наступним чином.

Знання – це складноорганізовані дані про об'єкти, процеси, явища та відношення предметної області, зв'язки, закономірності та правила здійснення логічного виведення, а також інша інформація, отримана у результаті практичної діяльності та професійного досвіду людини, що дає змогу спеціалістам ставити та розв'язувати практичні задачі

Важливим етапом у роботі зі знаннями для використання у комп'ютерних системах є формування поля знань, що полягає у виявленні та визначенні об'єктів предметної області, їх властивостей та зв'язків між ними, а також представленні у наочній інтуїтивно зрозумілій формі. У загальному випадку поле знань можна представити багаторівневою моделлю, представленою на рис. 3.1. Таку модель можна розглядати по відношенню до структури знань людини так само як і до структури знань систем штучного інтелекту, зокрема експертних систем.

Тут *рівню 0* відповідає первинна інформація, яку отримує система від датчиків. Така інформація визначає цілі рішень конкретних задач та обмеження, які накладають на такі рішення. Для людини рівню 0 відповідають елементарні рефлексорні реакції організму на зовнішні подразники.

Рівень I утворюють «пасивні» знання, або «знання-описи». Такі знання характеризують структури об'єктів реального світу, їх взаємне розташування у просторі, а також якісні та кількісні властивості об'єктів.

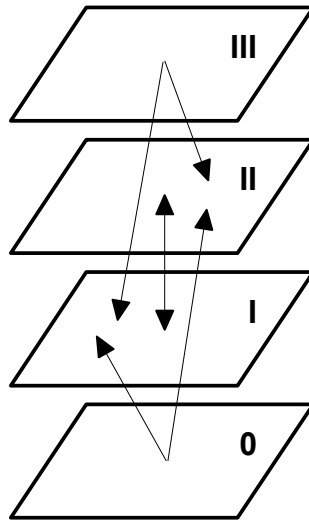


Рис. 3.1. Багаторівнева модель поля знань

Рівень II складається з «активних» знань. Такі знання характеризують реальний світ з погляду розвитку процесів. Також ці знання містять правила функціонування окремих об'єктів та їх сукупностей, відбивають причинно-наслідкові характеристики функціонування різних систем.

На *рівні III* знаходяться «абстрактні» знання або метазнання. Такі знання містять інформацію про методи використання знань нижчих рівнів, способах їх систематизації та доступу до них, про придбання нових та знищення (забуття) зайвих знань.

В задачах моделювання електроенергетичних систем рівню 0 відповідає система датчиків, які надають інформацію про поточний стан системи, положення комутаційного обладнання, результати вимірювань режимних характеристик тощо. На рівні I міститься інформація про схему електричних з'єднань в енергосистемі, параметри елементів робочої схеми, дані про електричні навантаження тощо. Рівень II містить таку інформацію як матриці струморозподілу, опорів впливу, чутливості тощо. На рівні III знаходиться інформація про правила використання даних нижчих рівнів, наприклад, алгоритми розрахунків робочих режимів енергосистем, правила оперативного керування робочими режимами енергосистем, правила виконання комутаційних перемикачів тощо.

Класифікація знань дозволяє визначити вісім основних їх типів:

1. *Базові елементи реального світу* – пов'язані з безпосереднім сприйняттям, не вимагають обговорення і додаються до бази фактів у тому вигляді, в якому вони отримані. Прикладами базових елементів реального світу є лінії електропередач, електричні станції та підстанції, вузли навантаження тощо.

2. *Твердження і визначення* – засновані на базових елементах і заздалегідь розглядаються як достовірні. До тверджень та визначень можна віднести знання про поточний режим синхронного генератора, визначення силового трансформатора, визначення неповнофазного режиму лінії електропередач.

3. *Концепції* – являють собою перегруповання або узагальнення базових об'єктів. До концепцій слід віднести узагальнені поняття електроенергетичної системи – лінії електропередач, режиму короткого замкнення тощо.

4. *Відношення* – виражають елементарні властивості базових елементів і відношення між концепціями. Так, за допомогою відношень визначають характеристики енергосистеми, зокрема опори ліній електропередач, коефіцієнти трансформації трансформаторів, типи ізоляторів та опор повітряних ліній і так далі. Відношення визначають також способи поєднання об'єктів енергосистеми у мережу.

5. *Теорема і правила перезапису* – являють собою правила поведінки, засновані на використанні інформації в порядку її отримання. До такого типу знань можна віднести записи законів Ома та Кірхгофа, які дають змогу виразити одні режимні характеристики через інші.

6. *Алгоритми* – визначають жорстку послідовність дій для досягнення поставленої цілі. Алгоритми визначають послідовність розрахунків режимних параметрів електричних систем, жорстку послідовність дій під час комутаційних перемикань тощо.

7. *Стратегії та евристики* – являють собою правила поведінки, засновані на використанні інформації в порядку, зворотному тому, в якому вона була отримана. До такого типу знань слід віднести принципи організації зустрічного регулювання режиму напруги в енергосистемі.

8. *Метазнання* – являють собою знання про знання. Метазнання забезпечують організацію кожного типу знань і вказівок, коли і як вони можуть бути використані в процесі розв'язання практичних задач.

Інший підхід до класифікації знань поділяє їх на поверхневі та глибинні.

Поверхневі знання характеризують лише видимі взаємозв'язки між подіями та фактами предметної області.

Глибинні знання містять абстракції, аналогії, схеми та іншу інформацію, яка відбиває структуру та природу процесів предметної області.

Різницю між поверхневими та глибинними знаннями проілюструємо на прикладі зустрічного регулювання напруги в електричних мережах. Відомо, що для забезпечення належної якості електричної енергії за напругою в режимах максимальних навантажень необхідно зменшити кількість регулювальних витків пристрою РПН живильного трансформатора.

Такі знання є поверхневими, оскільки тут зв'язок між зростанням навантаження в енергосистемі та перемиканням регулювальних витків РПН трансформаторів прихований.

Глибинні знання мають відбити цей зв'язок таким логічним ланцюгом. «Зростання навантаження енергосистеми призводить до збільшення робочих струмів на ділянках електричної мережі. Це призводить до збільшення падінь напруги на ділянках схеми та зниження рівнів напруги на шинах споживчих підстанцій. Тому для забезпечення нормованих рівнів напруги в електричній мережі необхідно збільшити напругу на вторинних обмотках живильних трансформаторів. Для цього необхідно зменшити коефіцієнти трансформації живильних трансформаторів, тобто зменшити кількість регулювальних витків пристроїв РПН».

Розвиток комп'ютеризованих засобів обробки інформаційних структур, на певному етапі привів до якісного переходу від «даних» до «знань». Традиційні комп'ютерні засоби оброблення інформації на основі алгоритмічних програм можна представити сукупністю двох основних складових: даних та алгоритмів. Разом з тим, інтелектуальні комп'ютеризовані засоби передбачають розв'язання практичних задач на основі поєднання знань та стратегій логічного виведення. При цьому знання мають ряд специфічних ознак, що дозволяють визначити співвідношення понять «знання» і «база знань» з поняттями «дані» і «база даних».

Відмінні якісні властивості знань обумовлені наявністю у знань надлишкових можливостей порівняно з даними. Основні специфічні ознаки знань полягають у наступному:

1. *Внутрішня інтерпретованість* передбачає встановлення для елементів даних зв'язаної системи ідентифікаторів, за допомогою якої надається можливість інтерпретації вміст відповідних комірок пам'яті.

2. *Структурованість* визначає властивість декомпозиції інформаційних одиниць і дозволяє встановити між ними відношень класифікації типу род-вид, клас-елемент, частина-ціле тощо.

4. *Зв'язність* характеризує можливість встановлення між інформаційними одиницями найрізноманітніших відношень, які пов'язують між собою інформаційні елементи предметної області та визначають семантику та прагматику системи в цілому.

4. *Семантичні простори з метрикою* характеризують близькість-віддаленість інформаційних одиниць їх взаємне розташування у просторі станів.

5. *Активність* визначає первинність знань у виконанні тих чи інших дій. Це означає, що експертні системи виконують певні дії не тому, що такі дії передбачені закладеними в них процедурами, а тому що в си-

стемі знань склалася відповідна ситуація, і відповідне співвідношення між інформаційними одиницями спонукає до певних дій, для реалізації яких необхідно виконати ті чи інші процедури.

6. *Функціональна цілісність* відбиває можливість вибору бажаного результату, часу його отримання, засобів отримання результату, засобів аналізу достатності отриманого результату.

Слід зауважити, що дві перші ознаки – внутрішня інтерпретованість та рекурсивна структурованість – властиві також і традиційним структурам даних. Таким чином, в системах штучного інтелекту, зокрема в експертних системах, знання виступають концептуальною моделлю реального світу предметної області у той час, коли дані не виконують функції моделювання, а лише характеризують параметри моделей. При цьому дані можна розглядати як окремий випадок знань.

Знання, які використовують експертні системи для розв'язання практичних задач, сконцентровані у базах знань. У загальному випадку бази знань являють собою цифрові сховища знань, представлених таким чином, щоб на їх основі можна було б організувати розв'язання задач предметної області експертної системи.

Розділ теорії штучного інтелекту, присвячений проблемам отримання, подання і формалізації знань в інтелектуальних комп'ютерних системах, зокрема, в експертних системах, отримав назву «інженерія знань». Остання вивчає процеси та методи виявлення знань щодо певної предметної області, наповнення ними спеціальних баз знань і представлення у формі таким чином, щоб на їх основі можна було б організувати ефективно розв'язання практичних задач.

Відповідно до типу носіїв, знання можна поділити на об'єктивізовані та суб'єктивні. Об'єктивізовані знання містяться у документах та текстах фахової літератури. Суб'єктивними є експертні знання фахівців.

Задача витягнення знань полягає у передачі потенційного досвіду розв'язання практичних задач від певного носія знань і перетворення його у вигляд, що дає змогу ефективно використовувати знання в інтелектуальних комп'ютерних системах.

На рис. 3.2 наведена ієрархічна діаграма класифікації методів витягнення знань. Методи витягнення знань поділяють на комунікативні та текстологічні.

Комунікативні методи передбачають витягнення знань під час безпосереднього контакту інженера зі знань із фахівцем – носієм експертних знань з предметної області. Такі методи поділяють на пасивні та активні.

Пасивні комунікативні методи передбачають передачу активної ролі в процесі витягнення знань до експерту. Тут фахівець сам вирішує, які

знання слід передати до бази знань для ефективного розв'язання практичних задач. Найбільш поширеними пасивними методами є спостереження за діями експерту під час розв'язання типових задач, протоколювання «думок уголос», лекції.

Метод спостереження полягає у тому, що інженер зі знань знаходиться біля експерту під час його професійної діяльності. При цьому фахівець має коментувати всі свої дії. Метод протоколювання «думок уголос» полягає у тому, що експерт не тільки коментує всі свої дії, але ще й пояснює, яким чином було рішення. Витягнення знань методом лекцій призначено для введення інженера зі знань у курс предметної області.

Активні комунікативні методи витягнення знань передбачають передавання активуючої функції у діалозі з експертом до інженера зі знань. Такі методи поділяють на індивідуальні і групові.

Найбільш поширеними індивідуальними методами витягнення знань є анкетування, інтерв'ю та діалоги. Анкетування полягає у тому, що інженер зі знань заздалегідь складає список питань, який використовує для опитування декількох експертів. Тут експерти самостійно заповнюють отримані анкети після попереднього інструктажу. Витягнення знань методом інтерв'ю полягає у спілкуванні з експертом, під час якого інженер зі знань задає фахівцю серію заздалегідь підготовлених питань. Метод вільного діалогу полягає у не регламентованому спілкуванні інженера зі знань з експертом.

Групові методи витягнення знань забезпечують можливість одночасного використання знань декількох експертів, взаємодія яких часто призводить до суттєвого поліпшення якості пропонуваних рішень практичних задач для наповнення баз знань. Тут використовують такі методи, як «круглий стіл», «мозковий штурм», рольові та ділові ігри. Метод «круглого столу» передбачає обговорення певної проблеми предметної області групою експертів з рівними правами. Тут задача дискусії полягає у колективному дослідженні спірних гіпотез з різних точок зору. Метод «мозкового штурму» полягає у генерації ідей щодо розв'язання проблем предметної області у замкненій групі експертів в умовах обмеженого часу. При цьому результати оцінює інша група фахівців, які не брали участь в обговоренні. Ігрові методи витягнення знань полягають у відтворенні певних ситуацій, наприклад, із виробничої практики, під час якої учасники гри на основі знань і життєвого досвіду приймають певні рішення, які потім аналізуються.

Текстологічні методи витягнення знань базуються на аналізі текстів фахової літератури (підручників, монографій, статей) та документів з предметної області. Під час такого процесу інженер зі знань отримує

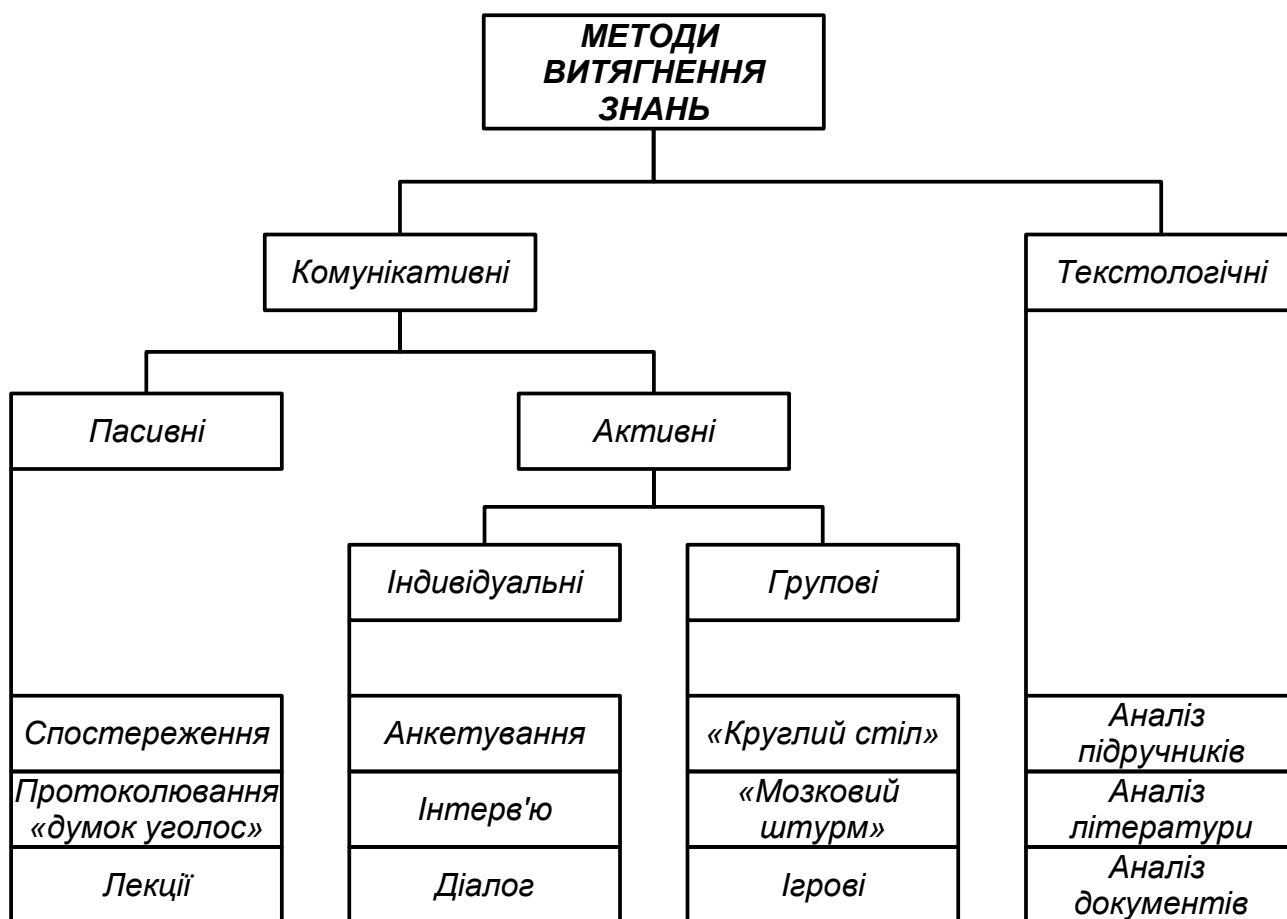


Рис. 3.2. Класифікація методів витягнення знань

певні знання, які, в подальшому, використовує для наповнення баз знань експертних систем.

Наведені методи витягнення знань не є антагоністичними по відношенню один до одного. Інженер зі знань, зазвичай, комбінує різні підходи до отримання необхідних експертних знань, наприклад, спочатку вивчає фахову літературу, а потім спілкується з експертами у предметній області.

Представлення знань в системах, які базуються на знаннях, являє собою систематизований спосіб формального вираження знань про предметну область. Практичний досвід спеціалістів, вкладений в базу знань експертної системи, представлений сукупністю фактів, правил, процедур і евристик (емпіричних правил) з предметної області. Формалізми, які забезпечують подання знань, називають моделями формалізації знань.

Модель представлення знань повинна відбивати найважливіші характеристики практичної задачі, на розв'язання якої направлене функціонування експертної системи. Разом з тим така модель має відбивати

важливі деталі предметної області, бути достатньо ефективною з погляду пошуку рішень тощо.

Моделі формалізації знань поділяють на процедурні та декларативні. В процедурних моделях знання представлені алгоритмами (процедурами) та містять необхідні описи інформаційних об'єктів а також способи та методи обробки такої інформації. Декларативні моделі знань передбачають розділення описів (декларацій) інформаційних структур та механізмів виводу, які оперують із цими даними.

Історично первинними були процедурні моделі знань. Тут для зміни або корекції інформаційної одиниці необхідно було змінювати програмний код інтелектуальної системи. Із розвитком систем штучного інтелекту пріоритет даних змінювався, що привело до розвитку декларативних моделей знань, які на сьогодні набули найбільшого поширення в інтелектуальних комп'ютерних системах.

В системах штучного інтелекту, зокрема в експертних системах, найбільшого поширення набули такі моделі формалізації знань:

- логічні моделі;
- продукційні моделі;
- мережеві моделі;
- фреймові моделі.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Дайте визначення поняття «Знання».
2. Наведіть та охарактеризуйте багаторівневу модель поля знань.
3. Наведіть та охарактеризуйте класифікацію типів знань.
4. Поясніть різницю між поверхневими та глибинними знаннями.
5. Наведіть та охарактеризуйте відмінні якості знань.
6. Що називають інженерією знань?
7. Наведіть класифікацію та пояснить особливості методів витягнення експертних знань.
8. Поясніть різницю між процедурними та декларативними моделями формалізації знань.



ЛІТЕРАТУРА

[1], [2], [4], [5], [9], [11], [13], [21], [31], [45], [62], [63], [69]

§3.2. ЛОГІЧНІ МОДЕЛІ ЗНАНЬ

Логічні системи знань базуються на методах «правильних» міркувань, які дають змогу формувати ланцюжки висновків, представлених в логічно послідовній формі. Тут міркування досліджуються не з погляду вкладеного до них змісту, а з погляду форми, що визначає принципи функціонування формальної логічної системи.

Формальна логіка полягає у конструюванні та дослідженні правил перетворення висловів, які зберігають свої істиноносні значення незалежно від закладеного до них конкретного змісту

Формальні логічні системи характеризуються високим рівнем абстракції та чітким визначенням методів, правил та законів оброблення інформації.

В основі сучасних логічних систем лежить розроблена Аристотелем силогістика – теорія логічних виведень, яка досліджує категоричні судження. Тут на підставі дедуктивного аналізу посилянь формують логічне виведення, яке є новим знанням. Так, наприклад, з двох посилянь

- 1) всі метали є провідниками електрики;
- 2) мідь є металом

випливає висновок про те, що мідь є провідником.

Сучасні логічні системи представлення знань базуються на апараті числень математичної логіки. Тут знання про предметну область представляють сукупністю логічних формул. Тотожні перетворення формул дозволяють отримувати нові знання.

Будь-яка формальна логічна система складається з таких елементів:

- 1) алфавіт, який утворює множина базових елементів предметної області;
- 2) синтаксичні правила, які визначають побудову допустимих висловлювань логічної системи – правильно побудованих формул;
- 3) правила логічного виведення, які визначають способи перетворення правильно побудованих формул з метою отримання нових знань.

Так, для розглянутого силогізму алфавітом є множина речовин (мідь), множина груп хімічних елементів (метали) та властивості речовин (електрична провідність). Синтаксичні правила дозволяють сформувати правильно побудовані формули, зокрема висловлювання «Всі метали є провідниками електрики» та «Мідь є металом», які визначають початковий стан бази знань логічної системи та розглядаються як аксіоми. Правила виведення разом із правилами синтаксису дозволя-

ють отримати нову правильно побудовану формулу «Мідь є провідником», яка є результатом роботи логічної системи.

Розв'язання практичних задач за допомогою формальних логічних систем здійснюють за наступним двотактним алгоритмом.

1. Перевіряють, чи є серед аксіом правильно побудована формула, яка визначає цільовий стан системи. Якщо така формула є, мета розв'язання задачі вважається досягнутою. В інакшому випадку управління передають процедурі 2.

2. До множини аксіом застосовують правила виведення. Якщо це призводить до формування нових правильно побудованих формул, останні додають до множини аксіом та передають управління процедурі 1. В інакшому випадку слід вважати, що задача не має розв'язку.

Серед логічних моделей, які використовують для представлення знань в системах штучного інтелекту, зокрема в експертних системах, найбільшого поширення набули апарати числення висловлювань, числення предикатів та псевдовізначні логіки.

Розглянемо основні положення апарату числення висловлювань.

Висловлюванням називають речення, зміст якого можна оцінити як істинне або хибне

На основі заданих висловлювань за допомогою логічних зв'язок утворюють складні висловлювання. До логічних зв'язків відносять кон'юнкцію (\wedge) – правило типу «І»; диз'юнкцію (\vee) – правило типу «АБО»; імплікацію (\rightarrow) – правило типу «ЯКЩО ... ТО ...»; та заперечення (\neg) – правило типу «НЕВІРНО, ЩО ...». Кожну логічну зв'язку розглядають як операцію, яка з простих висловлювань утворює складне.

Результати застосування логічних операцій до простих висловлювань можна визначити за таблицями істинності. Для найпростішого випадку з одного та двох висловлювань таблиця істинності наведена в табл. 3.1. Тут «І» означає істину, а «Х» – хибність.

Таблиця 3.1. Таблиця істинності простих висловлювань

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \rightarrow B$	$A \leftrightarrow B$
X	X	I	X	X	I	I
X	I	I	X	I	I	X
I	X	X	X	I	X	X
I	I	X	I	I	I	I

До таблиці 3.1 включено логічну зв'язку $A \leftrightarrow B$, яку називають подвійною імплікацією (правило типу «ЯКЩО, І ТІЛЬКИ ЯКЩО А ТО В»).

Множину базових елементів (алфавіт) числення висловлювань складають:

- 1) пропозиційні літери, за допомогою яких позначають окремі висловлювання;
- 2) логічні константи «ІСТИНА» та «ХИБНІСТЬ»;
- 3) символи логічних операцій « \neg », « \wedge », « \vee », « \rightarrow » та « \leftrightarrow »;
- 4) знаки пунктуації «(» та «)», які визначають пріоритет виконання логічних операцій.

Правильно побудовані формули числення висловлювань визначають наступні синтаксичні правила:

- 1) будь-яка пропозиційна літера є правильно побудованою формулою;
- 2) логічні константи є правильно побудованими формулами;
- 3) якщо F_1 та F_2 – правильно побудовані формули, то логічні вирази « $\neg F_1$ », « $F_1 \wedge F_2$ », « $F_1 \vee F_2$ », « $F_1 \rightarrow F_2$ » та « $F_1 \leftrightarrow F_2$ » також є правильно побудованими формулами.

Перші два правила визначають формування елементарних (атомарних) формул, а третє правило визначає способи формування нових логічних формул.

Серед логічних формул числення висловлювань виділяють правильно побудовані формули, які є істинними за всіх можливих інтерпретацій складових. Такі формули називають тавтологіями. Прикладом тавтології є вираз « $A \vee \neg A$ ». Формули, які приймають хибне значення за будь-яких інтерпретацій складових називають протиріччя, наприклад, « $A \wedge \neg A$ ».

Логічне виведення в апараті числення висловлювань базується на наступних правилах.

1. *Модус поненс* (modus ponens):

$$\{A, (A \rightarrow B)\} \Rightarrow B,$$

тобто, якщо висловлювання A істинне, а з істинності висловлювання A випливає істинність висловлювання B , то це означає, що висловлювання B також істинне. Тут висловлювання A та імплікацію $A \rightarrow B$ розглядають як посилення, а висловлювання B – заключення правила логічного виведення.

Наприклад, нехай база знань містить правило «якщо речовина є металом, то вона є провідником електрики», а також факт про те, що речовина є міддю, тобто металом. Відповідно до правила modus ponens мідь є провідником електрики.

Звернемо увагу на те, що зворотнє виведення є некоректним. Тобто, якщо відомо, що висловлювання B істинне, то це ще не означає істинність висловлювання A . Дійсно, для наведеного прикладу логічного виведення, електрична провідність речовини ще не означає її належність до металів (наприклад, графіт є провідником електрики, але не є металом). Тут коректним є правило модус толленс.

2. *Модус толленс* (modus tollens):

$$\{(A \rightarrow B), \neg B\} \Rightarrow \neg A,$$

тобто, якщо з істинності висловлювання A випливає істинність висловлювання B , а висловлювання B є хибним, то це означає хибність висловлювання A .

Наприклад, якщо для розглянутого правила відомо, що деяка речовина (наприклад, скло) не проводить електрику, то це означає, що ця речовина не є металом.

3. *Силогізм* (syllogism):

$$\{(A \rightarrow B), (B \rightarrow C)\} \Rightarrow (A \rightarrow C),$$

тобто, якщо з істинності висловлювання A випливає істинність висловлювання B , а з істинності висловлювання B випливає істинність висловлювання C , то це означає з істинності висловлювання A випливає істинність висловлювання C .

4. *Контрпропозиція* (contraposition):

$$(A \rightarrow B) \Rightarrow (\neg B \rightarrow \neg A),$$

тобто, якщо з істинності висловлювання A випливає істинність висловлювання B , то це означає з хибності висловлювання B випливає хибність висловлювання A .

5. *Виключення кон'юнкту*:

$$(A_1 \wedge A_2 \wedge \dots \wedge A_n) \Rightarrow A_i,$$

тобто, якщо кон'юнкція деяких висловлювань є істинним, то це означає що кожне з цих висловлювань є істинним.

6. *Введення кон'юнкції*:

$$(A_1, A_2, \dots, A_n) \Rightarrow (A_1 \wedge A_2 \wedge \dots \wedge A_n),$$

тобто, якщо певні висловлення є істинними, то кон'юнкція цих висловлювань також є істинною.

7. *Введення диз'юнкції*:

$$(A) \Rightarrow (A \vee B_1 \vee B_2 \vee \dots \vee B_n),$$

тобто, якщо деяке висловлення є істинним, то диз'юнкція цього висловлювання з будь-якими іншими висловленнями також є істинною.

8. Виключення подвійного заперечення:

$$(\neg\neg A) \Rightarrow A,$$

тобто, якщо подвійне заперечення деякого висловлення є істинним, то саме висловлювання також є істинним.

9. Проста резолюція:

$$[(A \vee B), \neg B] \Rightarrow A,$$

тобто, якщо відомо, що диз'юнкція висловлювань A та B істинна, а висловлювання B хибне, то це означає, що висловлювання A є істинним.

10. Резолюція:

$$[(A \vee B), (\neg B \vee C)] \Rightarrow (A \vee C),$$

тобто, якщо відомо, що диз'юнкції висловлювань A та B , а також заперечення B та C істинні, то це означає істинність диз'юнкції висловлювань A та C .

Наведені правила логічного виведення досить прості та зрозумілі. Разом з тим вони дозволяють організувати ефективне формальне розв'язання практичних задач.

Розглянемо застосування апарату числення висловлювань на прикладі розв'язання наступної задачі.

Нехай необхідно перевірити, чи будуть отримувати живлення споживачі НЗ вузлової підстанції 110/10 кВ, спрощену схему електричних з'єднань якої представлено на рис. 3.3. На рис. 3.3 використано такі позначення: Н1...Н8 – групи споживачів, які отримують живлення від відповідних фідерів розподільного пристрою 10 кВ підстанції; В1...В8 – вимикачі, через які фідери під'єднані до шин розподільного пристрою; Ш1, Ш2 – секції шин розподільного пристрою 10 кВ; СВ – секційний вимикач; Т1, Т2 – силові трансформатори; Л1, Л2 – лінії електропередач мережі 110 кВ; ЛР1, ЛР2 – лінійні роз'єднувачі; ВВ1, ВВ2 – вимикачі, через які лінії електропередач під'єднані до силових трансформаторів; ВН1, ВН2 – вимикачі, через які силові трансформатори під'єднані до секцій шин розподільного пристрою 10 кВ.

Очевидно, що споживачі НЗ будуть отримувати живлення за умов наявності напруги на шинах Ш1 та замкнених контактів вимикача В3. Це можна представити логічною формулою

$$B1 \wedge S3 \Rightarrow C3, \quad (3.1)$$

де пропозиційні літери відповідають таким висловлюванням: $B1$ – «НА СЕКЦІЮ ШИН Ш1 ПОДАНО НАПРУГУ»; $S3$ – «КОНТАКТИ ВИМИКАЧА В3 ЗАМКНЕНО»; $C3$ – «СПОЖИВАЧІ НЗ ОТРИМУЮТЬ ЖИВЛЕННЯ».

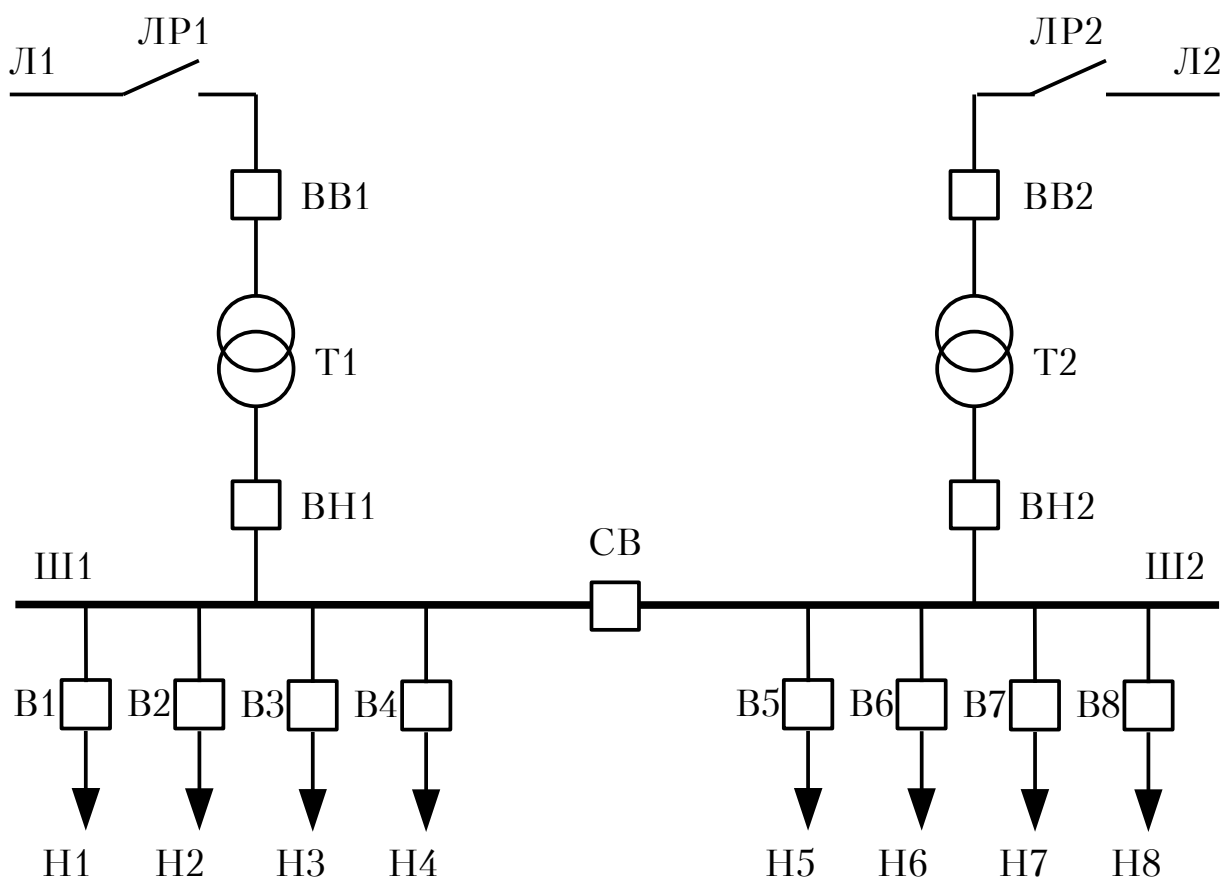


Рис. 3.3. Схема електричних з'єднань вузлової підстанції

Умовою наявності напруги на шинах Ш1 розподільного пристрою 10 кВ підстанції є замкнення контактів вимикача ВВ1 або наявність напруги на шинах Ш2 та замкнення контактів секційного вимикача СВ. Це дозволяє представити логічний вираз (3.1) у вигляді

$$\left. \begin{array}{l} B1 \wedge S3 \Rightarrow C3; \\ SW1 \vee (B2 \wedge SW3) \Rightarrow B1 \end{array} \right\} \Rightarrow \{(SW1 \vee (B2 \wedge SW3)) \wedge S3 \Rightarrow C3\}, \quad (3.2)$$

де $SW1$ та $SW3$ позначають висловлювання «КОНТАКТИ ВИМИКАЧА ВВ1 ЗАМКНЕНО» та «КОНТАКТИ СЕКЦІЙНОГО ВИМИКАЧА ЗАМКНЕНО»; $B2$ – «НА СЕКЦІЮ ШИН Ш2 ПОДАНО НАПРУГУ».

У свою чергу наявність напруги на шинах Ш2 розподільного пристрою визначається замкненістю контактів вимикача ВВ2. Це визначає перетворення логічної формули (3.2) до вигляду

$$(SW1 \vee (SW2 \wedge SW3)) \wedge S3 \Rightarrow C3. \quad (3.3)$$

де $SW2$ позначає висловлювання «КОНТАКТИ ВИМИКАЧА ВВ2 ЗАМКНЕНО».

Таким чином, для визначення отримання живлення споживачами Н3 необхідно перевірити істинність логічного виразу (3.3).

Апарат числення висловлювань характеризується обмеженими можливостями представлення знань. Тут розглядаються логічні зв'язки між окремими твердженнями, але структура тверджень не аналізується. Тому в експертних системах більшого поширення набув апарат числення предикатів.

Предикатом називають неоднорідну двозначну логічну функцію довільної кількості аргументів

Так, наприклад, для розв'язання розглянутої вище задачі визначення отримання живлення споживачами можна запропонувати логічну функцію

BREAKER_CLOSED(sw),

яка визначає положення контактів вимикачів та після підставлення назви конкретного вимикача, перетворюється у висловлювання «КОНТАКТИ ВИМИКАЧА sw ЗАМКНЕНО».

У загальному вигляді предикат записують у вигляді $P(x_1, x_2, \dots, x_n)$, де x_1, x_2, \dots, x_n належать одній або різним множинам об'єктів предметної області. Аргументи x_1, x_2, \dots, x_n називають предметними змінними, а їх конкретні значення – предметними сталими. Після підставлення у предикат всіх предметних сталих, останній перетворюється у висловлювання.

Алфавіт апарату числення предикатів складається з таких елементів:

- 1) предикатних літер, призначених для позначення конкретного предикату. Тут, зазвичай, використовують прописні латинські літери P, Q, R, S, T, V, N ;
- 2) функціональних символів за допомогою яких позначають логічні функції. Зазвичай, це малі латинські літери f, g, h ;
- 3) предметних змінних, які визначають зміст логічних функцій;
- 4) предметних сталих, які перетворюють предикати у висловлювання;
- 5) символів логічних операцій « \wedge », « \vee », « \rightarrow », « \leftrightarrow » та « \neg »;
- 6) символів кванторів « \forall » та « \exists »;
- 7) знаків пунктуації «(» та «)».

Синтаксичні правила, які визначають правильно побудовані формули апарату числення предикатів, базуються на поняттях терм та елементарна формула.

Терм визначають наступні рекурентні співвідношення:

- 1) будь-яка предметна змінна та будь-яка предметна стала є терм;
- 2) якщо t_1, t_2, \dots, t_n – терми, а f – n -місна функціональна літера, то $f(t_1, t_2, \dots, t_n)$ є термом.

Якщо P – предикатна літера, а t_1, t_2, \dots, t_n – терми, то предикат $P(t_1, t_2, \dots, t_n)$ є елементарною формулою (атомом). Окремим випадком елементарної формули є висловлювання (0-місний предикат).

Правильно побудовані формули числення предикатів визначаються наступними синтаксичними правилами:

- 1) будь-яка елементарна формула є правильно побудована формула;
- 2) якщо P і Q – правильно побудовані формули, а x – предметна змінна, то кожне з висловлювань $\neg P(x)$, $P(x) \wedge Q(x)$, $P(x) \vee Q(x)$, $P(x) \rightarrow Q(x)$, $\forall x P(x)$, $\exists x P(x)$ є правильно побудованими формулами.

Алфавіт апарату числення предикатів містить символи кванторів \forall і \exists – узагальнення та існування відповідно. Ці квантори разом із предметними змінними дозволяють визначити, що для всіх або деяких об'єктів предметної області виконується певне відношення або властивість.

Так, наприклад, запис $\forall sw \text{ BREAKER_CLOSED}(sw)$ визначає, що контакти всіх вимикачів на підстанції замкнені. У свою чергу запис $\exists sw \text{ BREAKER_CLOSED}(sw)$ означає, що на підстанції є хоча б один вимикач із замкненими контактами.

Квантор узагальнення можна інтерпретувати як узагальнення кон'юнкції над усіма елементами множини об'єктів предметної області. Це означає, що виразу $\forall x P(x)$ еквівалентно $P(x_1) \wedge P(x_2) \wedge \dots \wedge P(x_n)$. Так само квантор існування можна інтерпретувати як узагальнення диз'юнкції, тобто вираз $\exists x P(x)$ еквівалентний логічній формулі $P(x_1) \vee P(x_2) \vee \dots \vee P(x_n)$.

Розглянемо ілюстрацію використання апарату числення предикатів на прикладі фрагменту бази знань щодо регулювання режиму напруги в енергосистемі.

- 1) $(\forall x)(\forall y)(\text{OPER}(\text{RROFF}(y)) \wedge \text{CNT}(x, y, R_0) \rightarrow \text{INC}(U(x)))$;
- 2) $(\forall y)(\text{RR}(y, \text{ON}) \rightarrow \text{RES}(\text{RROFF}(y)))$.

Предикатні літери позначають: INC – бути збільшеним; CNT – мати з'єднання; OPER – бути операцією; RES – мати резерв. Терми: $\text{INC}(U(x))$ – збільшити напругу у пункті x ; $\text{CNT}(x, y, R_0)$ – пункти x та y поєднані через електричний радіус R_0 ; $\text{OPER}(\text{RROFF}(y))$ – операція по вимкненню реактора у пункті y ; $\text{RES}(\text{RROFF}(y))$ – резерв операції з регулювання шляхом вимкнення реактору у пункті y ; $\text{RR}(y, \text{ON})$ – реактор розташований у пункті y та увімкнено.

Зміст першого з логічних виразів полягає в наступному: якщо у пункті y є можливість вимкнути реактор, а сам пункт поєднано із пунктом x через електричну відстань R_0 , то реалізація такої операції призведе до збільшення напруги у пункті x .

Другий вираз означає, що якщо реактор підключено у пункті *y*, то існує резерв з регулювання напруги шляхом вимкнення реактору у пункті *y*.

Використання апарату числення предикатів обмежено жорсткими синтаксичними правилами утворення логічних формул. Це ускладнює, а часто унеможлиблює представлення просторових, часових, причинно-наслідкових та інших подібних зв'язків між об'єктами предметної області. Тут широкого застосування набули псевдофізичні логіки, які відбивають особливості сприйняття людиною фізичного світу. Загалом розрізняють такі види псевдофізичних логік:

- часова логіка – вивчає взаємний зв'язок часових співвідношень;
- просторова логіка – вивчає просторові співвідношення;
- каузальна логіка – вивчає причинно-наслідкові зв'язки;
- логіка дії – вивчає співвідношення типу суб'єкт–дія, дія–місце тощо.

Основною перевагою логічних моделей представлення знань є наявність чіткого синтаксису і широко прийнятої формальної семантики, а також теоретично обґрунтованих процедур автоматичного виводу. Разом з тим, основним недоліком даних моделей є неможливість отримання висновків в областях, де потрібна правдоподібні висновки, коли результат виходить з певною оцінкою впевненості в його істинності. Крім того, такі моделі характеризуються монотонним характером виведення, тобто в базу знань додаються тільки справжні твердження, що виключає можливість протиріч. На практиці часто зустрічаються немонотонні міркування, які важко, а, іноді, взагалі неможливо реалізувати в рамках логічної моделі.

Тем не менш логічні моделі виступають як теоретична основа опису самої системи подання знань і поступово розширюють свої можливості.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть визначення формальної логічної системи та охарактеризуйте її складові елементи.
2. Наведіть алгоритм розв'язання практичних задач за допомогою формальних логічних систем.
3. Дайте характеристику апарату числення висловлювань.
4. Наведіть таблицю істинності простих висловлювань.
5. Перерахуйте базові елементи апарату числення висловлювань.
6. Сформулюйте синтаксичні правила апарату числення висловлювань.

7. Наведіть основні правила логічного виведення.
8. Дайте характеристику апарату числення предикатів.
9. Перерахуйте базові елементи апарату числення предикатів.
10. Сформулюйте синтаксичні правила апарату числення предикатів.
11. Дайте характеристику апарату псевдофізичних логік.
12. Перелічіть види псевдофізичних логік.
13. Дайте коротку характеристику переваг та недоліків використання логічних моделей знань.



ЛІТЕРАТУРА

[1], [2], [3], [4], [5], [7], [8], [9], [10], [11], [41], [42], [45], [49], [56], [57], [60], [62], [63], [65], [69], [70], [76]

§3.3. ПРОДУКЦІЙНІ МОДЕЛІ ЗНАНЬ

Продукційна модель знань є самою розповсюдженою в системах штучного інтелекту і, у першу чергу, в експертних системах. Тут знання представляють у вигляді продукційних правил, або продукцій

Продукцією називають спосіб формалізації знань за допомогою правил типу «ЯКЩО A , ТО B »

Тут під A і B розуміють «ситуація–дія», «причина–наслідок», «посилання–висновок» тощо. Часто продукційне правило записують за допомогою знаку логічного слідування:

$$A \Rightarrow B.$$

На рис 3.4 наведено фрагмент продукційної бази знань, яка описує принципи регулювання режиму напруги в електричних мережах енергосистем. Аналіз наведеної на рис. 3.4 продукційної системи свідчить, що в разі збільшення навантаження споживачів експертна система видасть рекомендацію про зниження кількості регулювальних витків трансформатора живильної підстанції, що відповідає принципам зустрічного регулювання напруги.

В загальному випадку, продукційна система представлення знань містить такі основні елементи:

- базу фактів (робочу пам'ять);
- базу продукційних правил (базу знань);
- інтерпретатор (машину логічного виводу).

1. ЯКЩО	{навантаження споживачів збільшується}
ТО	{збільшуються струми на ділянках схеми}
2. ЯКЩО	{струми на ділянках схеми збільшуються}
ТО	{збільшуються падіння напруги на ділянках схеми}
3. ЯКЩО	{падіння напруги на ділянках схеми збільшуються}
ТО	{напруга на шинах споживчих підстанцій знижується}
4. ЯКЩО	{напруга на шинах споживчих підстанцій Нижче гранично допустимого значення}
ТО	{підняти напругу з боку живлення}
5. ЯКЩО	{необхідно підняти напругу}
ТО	{зменшити кількість регулювальних витків пристрою РПН живлячого трансформатора}

Рис. 3.4. Фрагмент продукційної бази знань

Робоча пам'ять (база фактів) продукційної системи представлена сукупністю істинних фактів, які визначають поточний стан предметної області експертної системи, або її образ.

База знань експертної системи представлена сукупністю продукційних правил вигляду

$$P = a_1, a_2, \dots, a_n \rightarrow a_m,$$

що означає, що новий факт a_m має місце за умови істинності всіх фактів a_1, a_2, \dots, a_n .

Продукційні правила складаються з двох частин – умовної (передумови) та виконавчої (висновку). Умовну частину продукційних правил називають зразком (шаблоном), який визначає узагальнену ситуацію в предметній області, коли це правило може бути активізовано та за-

стосовано для розв'язання певного етапу задачі. Частина дії визначає відповідний крок у розв'язанні задачі.

Інтерпретатор (машина логічного виведення) призначений для обробки бази знань експертної системи і безпосереднього розв'язання практичних задач. Машина логічного виведення реалізує дві наступні функції:

- 1) зіставлення бази правил із робочою пам'яттю та додавання нових фактів;
- 2) визначення порядку проглядання та застосовування продукційних правил.

Процес логічного виведення в продукційних експертних системах часто називають пошуком за зразком. Такий пошук виконують за наступним алгоритмом.

1. Зразки (передумови) продукційних правил зіставляють із поточним образом предметної області. В результаті визначають сукупність правил, які можуть бути активовані за поточних умов. Такі правила розміщують у конфліктній множині.

2. Якщо за поточної ситуації можна застосувати більше одного правила, то серед них, із застосуванням деякого критерію, обирають таке, що найбільш точно підходить до поточних умов. Таку операцію називають розв'язанням конфлікту.

3. Обране правило активують. Його спрацьовування породжує нові факти, які відповідають висновку правила. Нові факти розміщують у робочій пам'яті, що визначає доповнення образу предметної області.

Наведені процедури виконують циклічно, аж доки не буде знайдено розв'язок задачі, тобто коли висновок чергового активованого правила не співпадає з ціллю, або коли у базі знань не знайдеться жодного правила, передумови яких співставні з образами бази фактів.

Роботу інтерпретатора продукційної експертної системи проілюстровано на рис. 3.5. Представлений алгоритм роботи продукційної експертної системи базується на стратегії пошуку під керуванням даних, тобто прямого ланцюжку логічного виведення. Можливий також і зворотній процес пошуку під керівництвом цілі, тобто за зверненням ланцюжком. Тут серед продукційних правил бази знань здійснюють пошук правил, висновки яких призводять до цілі поставленої задачі. Для знайдених правил перевіряють істинність фактів, які містяться у передумовних частинах. Для цього реалізують пошук продукційних правил, які підтверджують істинність цих фактів. Таку операцію проводять циклічно доки не буде підтверджена, або, навпаки, спростована істинність цільового факту. Представлений циклічний про-

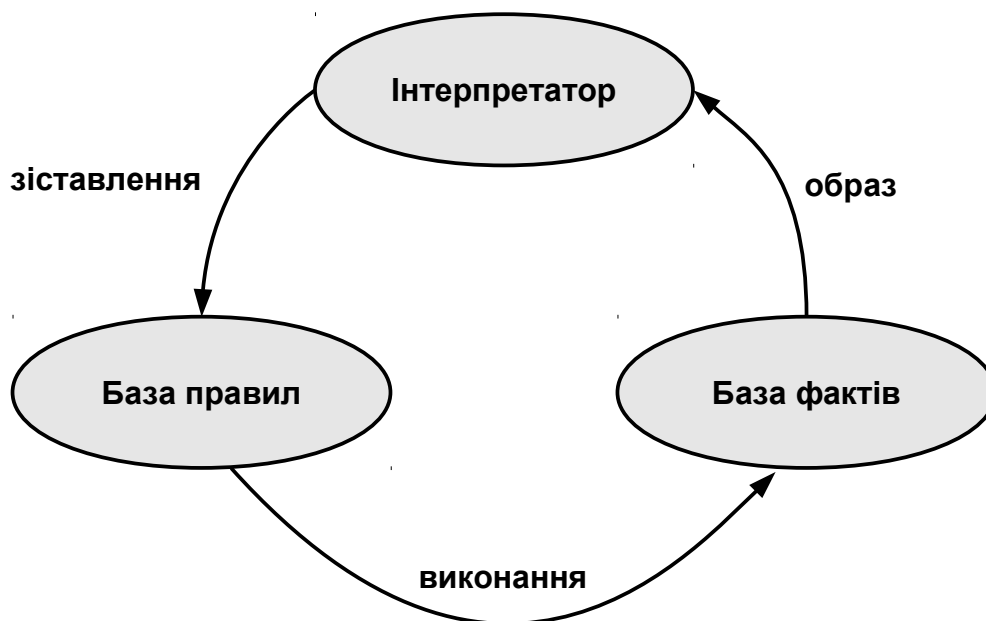


Рис. 3.5. Ілюстрація роботи продукційної системи

цес пошуку за стратегією зворотнього ланцюжку логічного виведення повністю відповідає діаграмі, наведеній на рис. 3.5.

Реалізація продукційних експертних систем пов'язана із розв'язанням проблеми конфліктної множини. Справа в тому, що кожний цикл роботи машини логічного виведення пов'язаний із активацією лише одного правила. Разом з тим, база знань експертної системи може містити декілька правил, для котрих передумови є істинними та містяться у робочій пам'яті. Такі правила розміщують у конфліктній множині і на кожному циклі роботи здійснюють вибір одного правила для активації.

Тут вибір правила з конфліктної множини може здійснюватися із застосуванням різних стратегій. Найпростіші серед них є неінформативними і базуються на алгоритмах пошуку у ширину і глибину. До інших стратегій відносять:

- принцип «стопки книжок»;
- принцип найбільш довгої умови;
- принцип пріоритетного вибору;
- принцип метапродукцій.

Принцип «стопки книжок» полягає в тому, що конфліктна множина продукційних правил впорядковується за частотою використання продукцій. У першу чергу використовують продукцію з найбільшою частотою використання.

Принцип найбільш довгої умови полягає в тому, що у першу чергу виконують продукції, передумовна частина яких охоплює більшу кількість

фактів. Такі продукції, зазвичай, відповідають специфічним ситуаціям предметної області і застосування даного принципу часто призводить до прискорення пошуку розв'язання задачі.

Принцип пріоритетного вибору полягає в тому, що з кожною продукцією зв'язують певне додатне число (статичне або динамічне), яке визначає пріоритет використання правила.

Принцип метапродукції полягає в тому, що до бази знань вводять спеціальні метаправила, за допомогою яких здійснюється впорядкування конфліктної множини та вибір продукції для активації.

Інколи у постумовах продукції зазначають ім'я наступної продукції, яку необхідно виконати. Це перетворює продукційну систему в алгоритмічну.

На сьогодні в технології конструювання експертних систем продукційні моделі знань отримали найбільш широке розповсюдження. Це пояснюється наступними основними причинами:

- універсальністю – практично будь-яку предметну область можна подати у формі системи продукцій;
- модульністю – кожна продукція являє собою окремий елемент знань, додавання або вилучення окремих продукцій здійснюється незалежно;
- декларативністю – продукції визначають ситуації предметної області, а не механізм управління виведенням;
- природністю процесу виведення заключень, аналогічному процесу міркувань експерта під час розв'язання практичних задач;
- асинхронністю та природним паралелізмом реалізації логічних міркувань;
- простотою програмної реалізації механізму логічного виведення;
- можливістю підтримки нечітких знань.

Разом з тим, продукційні моделі представлення знань не позбавлені певних недоліків:

- зі збільшенням розмірності бази продукційних правил суттєво знижується ефективність виведення, оскільки за великої кількості продукцій значна частина часу витрачається на непродуктивну перевірку умов активації правил;
- до продукційних моделей виставляють жорсткі вимоги щодо несуперечності баз знань, контроль якої стає дуже складним завданням із збільшенням розмірності бази знань.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає формалізація знань у вигляді продукційних систем?
2. Що називають продукцією?
3. Наведіть та охарактеризуйте складові продукції.
4. Наведіть та охарактеризуйте складові елементи продукційної експертної системи.
5. Наведіть алгоритм роботи машини логічного виведення продукційної системи під керуванням даними.
6. Наведіть алгоритм роботи машини логічного виведення продукційної системи під керуванням ціллю.
7. Перерахуйте та охарактеризуйте стратегії розв'язання конфліктів продукційних систем.
8. Наведіть та охарактеризуйте позитивні та негативні властивості продукційних експертних систем.
9. Спробуйте згенерувати продукційну систему знань для розв'язання будь-якої електроенергетичної задачі.



ЛІТЕРАТУРА

[1], [2], [4], [5], [7], [9], [10], [11], [13], [14], [15], [31], [45], [56], [60], [62], [63], [65], [69], [76]

§3.4. СЕМАНТИЧНІ МЕРЕЖІ

Практично будь-яку предметну область можна представити із використанням апарату теорії графів. Тут об'єкти предметної області асоціюють із вершинами графу, а відношення між об'єктами – дугами. Такий спосіб подання інформації про предметну область називають семантичними мережами.

Семантичними мережами називають сімейство способів формалізації знань за допомогою апарату теорії графів, вершини якого, як правило, позначають об'єкти предметної області, а дуги відповідають зв'язкам між об'єктами

На рис. 3.6 представлено фрагмент семантичної мережі, призначений для зберігання в базі знань експертної системи факту про те, що трансформатор ТМ-250/10, встановлений на підстанції «Володимирська», працює в режимі перезбудження.

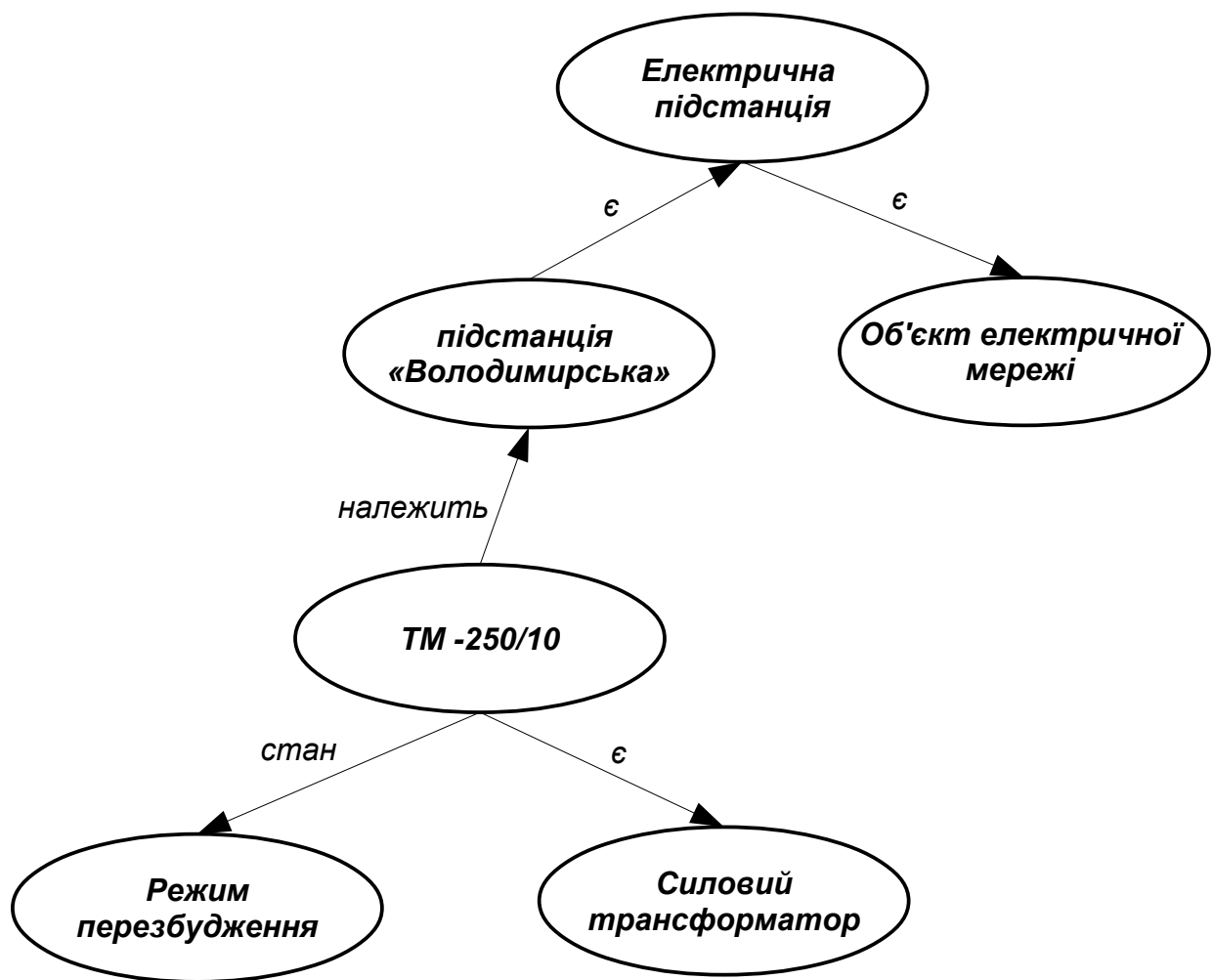


Рис. 3.6. Фрагмент семантичної мережі

Основними структурними елементами апарату семантичних мереж є концепти – об'єкти предметної області, поняття, події, властивості, які визначають вершини семантичної мережі та відношення між ними, які утворюють дуги.

Об'єкти предметної області, які відображаються в семантичній мережі, поділяють на три групи:

- узагальнені;
- індивідні;
- агрегатні.

Узагальнені об'єкти відповідають деякій збірній абстракції реально існуючих об'єктів, процесів або явищ предметної області. Прикладами узагальнених об'єктів є поняття «повітряна лінія електропередачі», «електрична підстанція», «режим короткого замкнення» тощо.

Індивідний об'єкт являє собою відокремлений певним чином одиничний представник (екземпляр) узагальненого класу. Наприклад, під-

станція «Володимирська» є екземпляром узагальненого класу «електрична підстанція».

Агрегатним називають об'єкт предметної області, який складається з інших. Наприклад, повітряна лінія складається з фазних проводів, опор, ізоляторів, грозозахисних тросів, лінійної арматури тощо.

Наведена класифікація об'єктів є умовною та визначається контекстом поставленої задачі. Відповідно до змісту задачі той самий об'єкт може належати до груп узагальнених, індивідних або агрегатних. Наприклад, концепт «трансформатор ТМ-250/10» можна розглядати як індивідний об'єкт, тобто як екземпляр класу «силові трансформатори». Разом з тим, даний концепт можна розглядати як узагальнення силових трансформаторів даного класу. Крім того, відповідно до контексту задачі трансформатор ТМ-250/10 можна розглядати як агрегатний об'єкт, який складається з баку, осереддя, обмоток і так далі.

Між об'єктами семантичних мереж можуть бути встановлені зв'язки будь-якого типу. Такі зв'язки можуть бути бінарними (пов'язують між собою два об'єкти предметної області) та *N*-арні (пов'язують між собою *N* об'єктів).

В апараті семантичних мереж використовують такі типи відношень між об'єктами:

- *родо-видові відношення* встановлюють між узагальненими об'єктами різного ієрархічного рівня. Наприклад, поняття «лінія електропередачі» є родовим поняттям для узагальненого об'єкту «повітряна лінія електропередачі»;
- *відношення «є представником»* встановлюють між узагальненим та індивідним об'єктами, якщо індивідний об'єкт є екземпляром відповідного класу. Наприклад, повітряна лінія «Трихати–Аджалик» є представником (екземпляром) класу «повітряні лінії електропередач»;
- *відношення «є частиною»* встановлюють між агрегатним об'єктом та його складовими. Наприклад, силовий трансформатор містить фазні обмотки;
- *атрибутивні відношення* встановлюють між об'єктами предметної області та концептами, які виражають властивості цих об'єктів. Наприклад, силовий трансформатор має властивість – номінальну потужність;
- *відношення значень* встановлюють зв'язки між властивостями об'єктів та їх конкретними значеннями. Наприклад, номінальна потужність силового трансформатора 250 кВА;
- *функціональні відношення* визначають прагматичні аспекти функціонування об'єктів предметної області. Наприклад, факт «силовий трансформатор передає електричну енергію потужністю

100 кВА» встановлює функціональне відношення між концептами «силовий трансформатор» та «потік потужності»;

- *кількісні* встановлюють такі відношення як *більше, менше, дорівнює* тощо;
- *просторові* встановлюють такі відношення як *близько, далеко, за, перед* тощо;
- *часові* встановлюють такі відношення як *раніш, пізніш, протягом* тощо;
- *логічні відношення* встановлюють зв'язки *та, або, ні*.

Семантичні мережі зазвичай розглядають на трьох ієрархічних рівнях, як показано на рис. 3.7. Тут повній базі знань відповідає вся семантична мережа. Окремі концептуальні одиниці (окремому твердженню, або окремому поняттю) – концептуальний підграф. Так, семантична мережа, представлена на рис. 3.6 відповідає окремому твердженню і по суті є концептуальним підграфом, який є складовою семантичної мережі. Нижчому рівню (тріаді «об'єкт-атрибут-значення») відповідають дві пов'язані вершини семантичної мережі разом із дугою між ними.

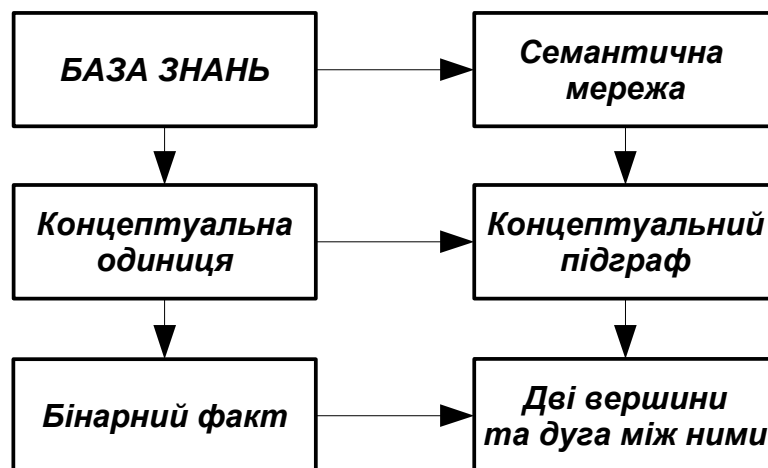


Рис. 3.7. Трирівнева архітектура семантичних мереж

Існує багато підходів до способів обробки експертної інформації, представленої за допомогою семантичних мереж. Найбільш поширеними тут є метод перехресного пошуку та метод зіставлення зі зразком.

Метод перехресного пошуку полягає у просуванні графом семантичної мережі з метою пошуку відповіді на поставлений запит. Тут основними процедурами обробки семантичних мереж є:

- 1) визначення та анулювання екземпляру класу (вершини семантичної мережі);
- 2) встановлення та анулювання зв'язку між вершинами;
- 3) підрахунок кількості екземплярів класу;
- 4) підрахунок кількості вершин, пов'язаних із заданою;

- 5) перевірка належності екземпляру до класу;
- 6) перевірка наявності, або відсутності зв'язку між вершинами;
- 7) пошук вершини або зв'язку за ідентифікатором;
- 8) перехід від однієї вершини до іншої;
- 9) об'єднання підграфів.

Метод зіставлення зі зразком полягає в тому, що запит до семантичної мережі оформлюють у вигляді автономної мережі запиту, яка побудована за тими ж правилами і відбиває ті самі об'єкти і відношення, що й в базі знань основної семантичної мережі. Пошук відповіді реалізують шляхом зіставлення мережі запиту з основною мережею та визначення необхідної інформації, яка міститься у базі знань.

Апарат семантичних мереж є простим, виразним та потужним засобом представлення інформації в базі знань. Основні переваги використання семантичних мереж полягають у:

- простоті і наочності представлення інформації;
- практично необмежених можливостях розширення бази знань;
- структурованості інформації;
- можливості відображення складної ієрархічної системи концептів.

Разом з тим, до недоліків семантичних мереж слід віднести:

- відображення переважно статичної описової інформації;
- складність, часто – неможливість, реалізації складних каузальних (причинно-наслідкових) взаємодій між концептами.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає формалізація знань у вигляді семантичних мереж?
2. Наведіть та охарактеризуйте формальні складові апарату семантичних мереж.
3. Дайте характеристику класифікації об'єктів предметної області.
4. Дайте характеристику класифікації відношень між об'єктами предметної області.
5. Дайте характеристику трирівневій архітектурі семантичних мереж.
6. Перелічить основні процедури обробки семантичних мереж.
7. Наведіть та охарактеризуйте позитивні та негативні властивості апарату семантичних мереж.
8. Спробуйте згенерувати семантичну мережу для формалізації знань і розв'язання будь-якої електроенергетичної задачі.

[1], [2], [4], [5], [7], [9], [10], [11], [13], [15], [31], [45], [56], [60], [62], [63], [65], [69], [76]

§3.5. СКРИПТОВІ МОДЕЛІ ЗНАНЬ

Скрипти (сценарії) являють собою мережеві моделі формалізації знань, та є, по суті, структурованим представленням, яке описує стереотипну послідовність подій за участю певних об'єктів

Скрипти використовують в системах розуміння природної мови для організації бази знань в термінах ситуацій, які система має розуміти. Скрипти можуть взаємодіяти з іншими скриптами і мають великі можливості для опису динамічних аспектів знання.

В загальному випадку сценарії містять наступні елементи:

- початкові умови, які мають бути істинними для активації скрипту;
- результати або факти, які стають істинними після завершення сценарію;
- припущення, які підтримують контекст сценарію. Множина припущень визначає прийняті за замовчуванням умови реалізації сценарію;
- ролі є діями, які здійснюють окремі учасники сценарію;
- сцени являють собою окремі атомарні елементи сценарію, кожна з яких представляє часові аспекти сценарію а їх послідовність – весь сценарій в цілому.

На рис. 3.8, як приклад скриптової моделі, представлено спрощений типовий сценарій оперативних перемикачів з виведення в ремонт силового трансформатора Т1 підстанції, умовна схема головних з'єднань якої наведена на рис. 3.3.

Позитивними властивостями скриптової моделі формалізації знань є:

- простота інструментальної реалізації;
- безпосередня участь користувача на кожному етапі функціонування системи;
- простота формату команд користувача;
- стандартизація дій користувача.

До недоліків слід віднести:

- жорстку встановленість дій системи;



Рис. 3.8. Типовий сценарій оперативних переключень

- необхідність зберігання в пам'яті великих обсягів інформації, яка не стосується безпосередньо до бази знань;
- суттєву обмеженість процедур логічного вибору.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає формалізація знань за допомогою скриптів?
2. Перелічить та охарактеризуйте складові елементи системи скриптів.
3. Наведіть та охарактеризуйте позитивні та негативні властивості скриптової моделі представлення знань.



ЛІТЕРАТУРА

[7], [11]

§3.6. ФРЕЙМОВІ МОДЕЛІ ЗНАНЬ

Фреймом називають складно організовану структуру даних, призначену для подання типових понять або стереотипних ситуацій

Фрейм можна представити у вигляді деякого каркасу, який адаптують та заповнюють у відповідності до конкретних об'єктів, процесів або явищ реального світу. Фрейми являють собою блоки знань, ко-

жним з яких можна маніпулювати як єдиним цілим, що дає змогу ефективно структурувати базу знань.

Формально деякий об'єкт в термінах фреймової моделі можна описати такою структурою:

$$\begin{array}{l} \text{Ім'я_фрейму} \quad \{ \\ \quad <\text{атрибут_1: значення_1}> \\ \quad <\text{атрибут_2: значення_2}> \\ \quad \dots \\ \quad <\text{атрибут_n: значення_n}> \} \end{array}$$

Тут *ім'я_фрейму* являє собою унікальний ідентифікатор, який визначає конкретний фрейм; *атрибут_1*, *атрибут_2*, ... — множина атрибутів, які визначають властивості об'єктів реального світу; *значення_1*, *значення_2*, ... — конкретні значення атрибутів, притаманні конкретному об'єкту.

Кожна пара *<атрибут: значення>* являє собою окрему інформаційну одиницю, яку називають *слот*. Після заповнення слотів конкретними даними, фрейм буде представляти ту чи іншу ситуацію, явище чи об'єкт предметної області.

За змістовною формою фрейми поділяють на три типи:

1) *фрейми-поняття* являють собою інформаційні структури кон'юнктивного типу, які містять набір слотів для визначення конкретних властивостей відповідної інформаційної одиниці, наприклад, фрейм «*лінія електропередач*» містить слоти «*номінальна напруга*», «*довжина*», «*марка проводу*» тощо, які визначають характеристики конкретної лінії;

2) *фрейми-меню* являють собою інформативні структури диз'юнктивного типу, призначені для організації процедурних знань із використанням операції вибору, наприклад, фрейм «*провести аналіз*» може містити слоти «*розв'язати рівняння*», «*підставити дані*», «*уточнити завдання*» тощо;

3) *фрейми-сценарії* являють собою ієрархічні інформаційні структури, які визначають послідовність подій для певного процесу або для певної ситуації, або причинно-наслідкові зв'язки між подіями. Таким чином, розглянуті раніше скриптові моделі знань, по суті, є фреймоподібними структурами.

Фреймові моделі формалізації знань передбачають наявність фреймів-прототипів, які відповідають узагальненим об'єктам предметної області. Такі фрейми можна розглядати як «пусті» каркаси, заповнення яких призводить до формування фреймів-примірників, які відповідають індивідним об'єктам.

На рис. 3.9 представлено систему фреймів для опису лінії електропередач та електричної підстанції робочої схеми електричної ме-

режі. У фреймових моделях подання знань як значення слотів можуть виступати імена інших фреймів. Так, наприклад, фрейм опису лінії електропередачі містить слоти ідентифікаторів початкового та кінцевого пунктів. Як значення цих слотів ступають посилання на відповідні фрейми вузлових підстанцій (див. рис. 3.9). Так само, субфреймами можуть виступати значення слотів «Марка проводу» і «Марка тросу» фрейму «Лінія електропередачі», або слоту «Тип трансформатору» фрейму «Електрична підстанція». Такі взаємні посилання між фреймами надають можливість представлення агрегатних об'єктів предметної області та формування мережі фреймів будь-якої складності та ієрархічності. Таким чином, фрейм можна розглядати як фрагмент семантичної мережі, призначений для опису деякого об'єкту (події, ситуації) предметної області, тобто як концептуальний підграф семантичних мереж.

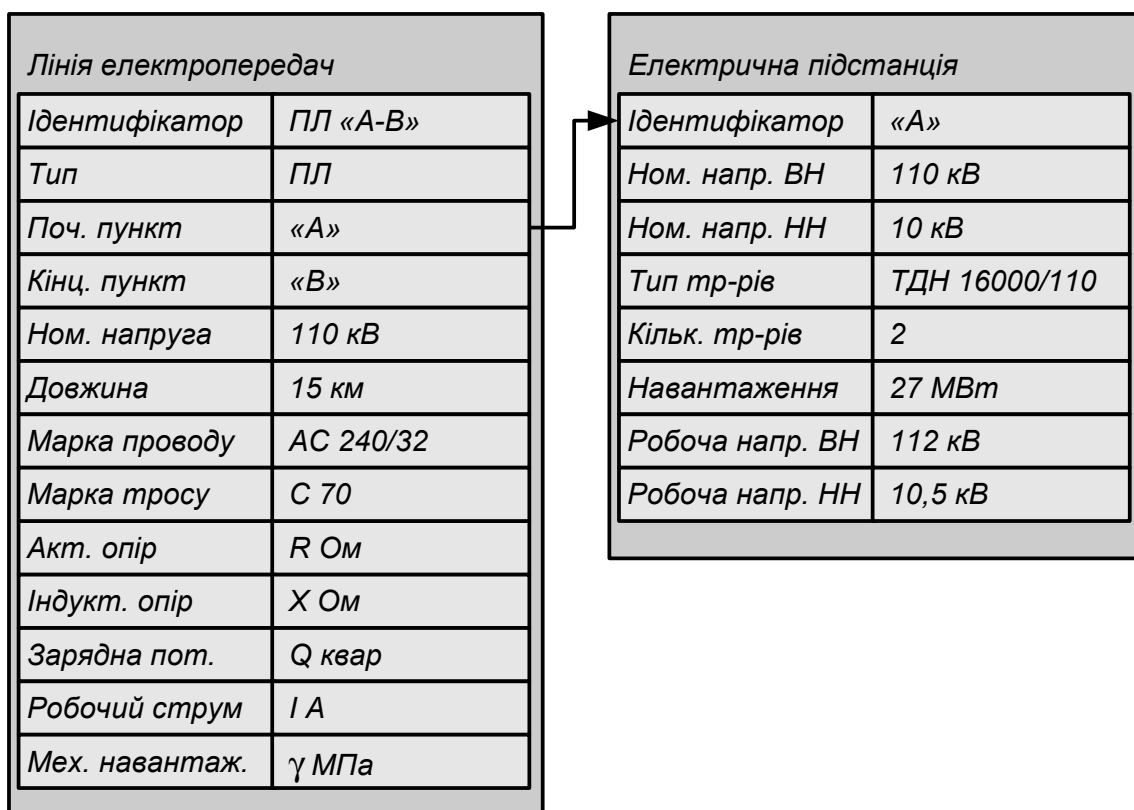


Рис. 3.9. Типова система фреймів

Значенням слотів може бути також деяка процедура, пов'язана з фреймом. Так, наприклад, фрейм «Лінія електропередачі» містить слот «Робочий струм», значенням якого може бути приєднана процедура розрахунку струму за законом Ома, або вимірювання поточного значення струму лінії.

Під час формування фреймів-примірників не обов'язково заповнювати всі слоти, а лише тільки ті, які потрібні для розв'язання конкретної практичної задачі. Так, наприклад, фрейм «Лінія електропередачі»

містить слот «*Механічні навантаження на проводи лінії*», значення якого має сенс лише в задачах дослідження механічної конструкції лінії і може бути пустим у разі розв'язання задачі аналізу робочого режиму електричної системи.

Оскільки фреймові моделі представлення знань є мережевими, таким моделям притаманні ті самі позитивні та негативні властивості, що і розглянутим вище мережевим моделям, зокрема семантичним мережам. Разом з тим, слід зважати на жорстку структурованість фреймових моделей, що з одного боку, суттєво обмежує гнучкість моделі, але з іншого боку дає змогу використовувати типові процедури обробки знань в процесі логічного виведення.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає формалізації знань за допомогою фреймів?
2. Наведіть типову узагальнену структуру фрейму та охарактеризуйте його складові елементи.
3. Наведіть та поясніть класифікацію фреймів за змістовною формою.
4. Поясніть положення про те, що скриптові моделі представлення знань можна подати у термінах фреймових структур.
5. Що називають фреймом-прототипом та фреймом-примірником?
6. Поясніть процес утворення фрейму-примірника?
7. Поясніть особливості заповнення слотів фреймів-примірників.
8. Поясніть, яким чином за допомогою фреймових структур моделюють агрегатні об'єкти?
9. Поясніть положення про те, що фреймові моделі по суті є мережевими.
10. Перелічіть позитивні та негативні властивості фреймових моделей подання знань.
11. Спробуйте згенерувати фреймоподібну структуру для подання знань і розв'язання будь-якої електроенергетичної задачі.



ЛІТЕРАТУРА

[1], [2], [4], [5], [7], [9], [10], [11], [13], [15], [31], [45], [60], [54], [62], [63], [65], [69], [76]

ЛОГІЧНЕ ВИВЕДЕННЯ В УМОВАХ НЕВИЗНАЧЕНОСТІ

§4.1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ПРОБЛЕМИ НЕВИЗНАЧЕНОСТІ ІНФОРМАЦІЇ

Розв'язання більшості практичних задач пов'язано із недостатніми обсягами вихідної інформації про предметну область, що унеможлиблює використання традиційних алгоритмічних методів аналізу. Так, наприклад, для визначення характеристик робочого режиму електричної системи необхідно мати точну інформацію про навантаження у всіх вузлах розрахункової схеми. Проте, через стохастичність процесів споживання електричної енергії така інформація завжди є неточною та може бути подана лише з певною ймовірністю.

Невизначеність можна розглядати як недостатність адекватної інформації для прийняття рішення. Це породжує значну проблему, оскільки може стати причиною прийняття невірного рішення. Так, наприклад, через невизначеність можливо не буде знайдено ефективного шляху відновлення електроенергетичної системи після тяжкої системної аварії.

Класифікація видів невизначеності умовно наведена на рис. 4.1. Загалом невизначеність може бути пов'язана із неповнотою інформації або з її неоднозначністю. Неповнота інформації зазвичай визначається недостатньою повнотою знань про предметну область, тобто з відсутністю адекватної моделі об'єкта дослідження або з недостатністю інформації про конкретну ситуацію, наприклад, через недостатню оснащеність датчиками. Неоднозначність знань зазвичай пов'язана з багатоваріантністю трактування інформації про предметну область та неможливістю точно встановити істинність певних положень. Неоднозначність зазвичай породжується фізичною або лінгвістичною невизначеністю. На фізичну невизначеність впливають невраховані випадкові події або неточність вимірювань. Лінгвістична невизначеність пов'язана із використанням природної мови для подання інформації про предметну область. Тут розрізняють невизначеність слів та невизначеність фраз. Невизначеність слів зазвичай пов'язана з використанням таких модальних міркувань, як «часто», «багато», «високо» тощо, значення яких необхідно трактувати відповідно до контексту поточної задачі. Так само виділяють

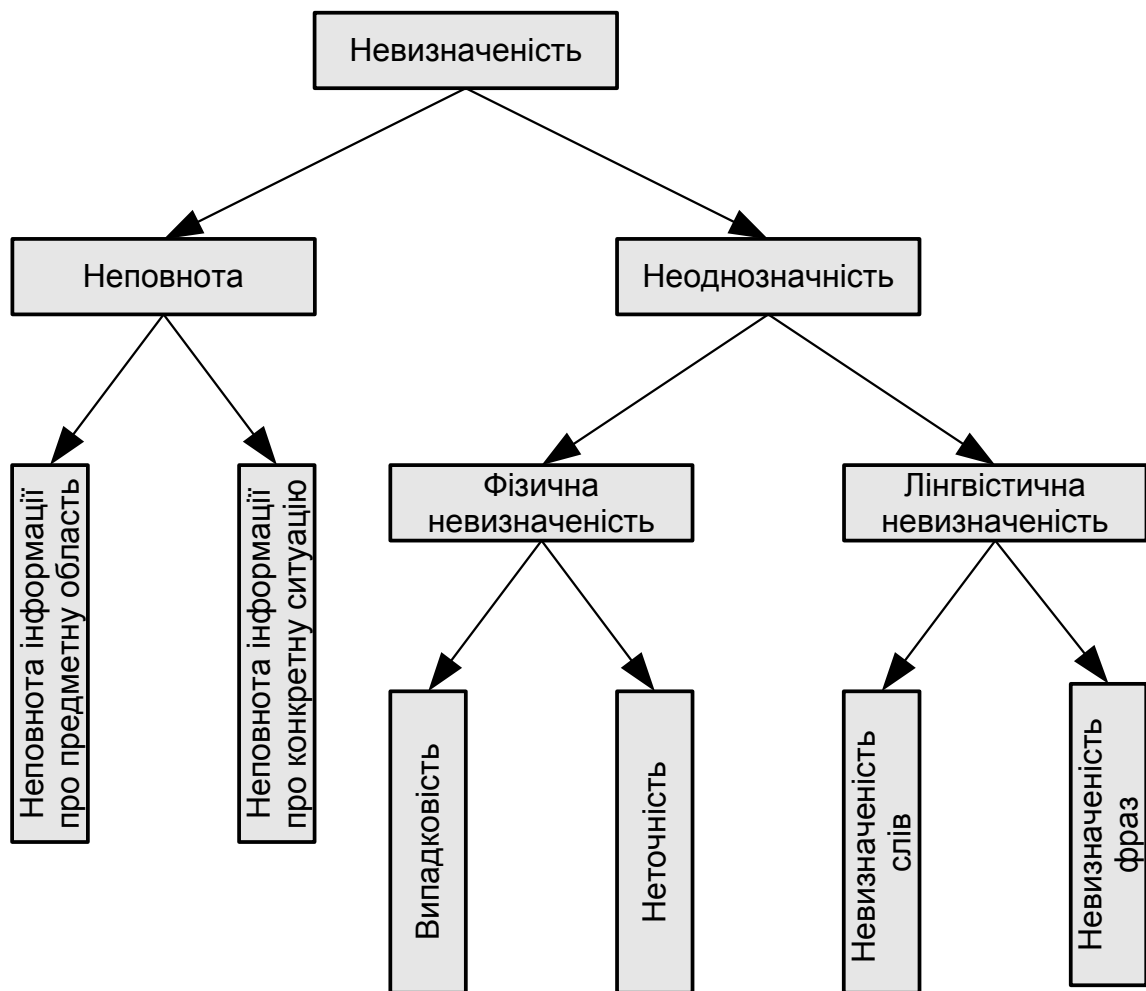


Рис. 4.1. Класифікація видів невизначеностей у поданні знань

невизначеність фраз, зміст яких визначається контекстом поточної практичної задачі.

Наведені типи невизначеності в загальному випадку можуть бути пов'язаними із об'єктивними або суб'єктивними причинами. Об'єктивна невизначеність пов'язана з принципово випадковим характером процесів у досліджуваній системі, що визначається впливом невідомих або неконтрольованих факторів. Тут, для прийняття якісних рішень необхідно враховувати певні ймовірнісні показники стохастичних процесів, які мають місце в предметній області.

Суб'єктивна невизначеність, у свою чергу, визначається надійністю суджень експертів, знання яких використано для формування бази знань експертної системи. Так, недостатня компетенція, занижена або завищена авторитетність джерела знань може призводити до хибних або неефективних рішень задач управління. Тут, для прийняття якісних рішень необхідно враховувати певні коефіцієнти довіри до експертної інформації бази знань. Разом з цим слід зважати на те, що визначення

таких коефіцієнтів довіри також є суб'єктивним, що породжує, так звану, проблему невизначеності проблеми невизначеності.

Під час розробки експертних систем, які функціонують в умовах невизначеності інформації, необхідно додатково передбачати механізм розв'язання конфліктів між різними джерелами знань. Справа в тому, що декларативність баз знань може призводити до взаємопротилежних висновків, отриманих через різні ланцюги логічних міркувань. Більшість сучасних систем логічного виведення в умовах невизначеності вихідної інформації передбачають формування розв'язку практичних задач не у вигляді єдиного рішення, що можливо лише для детермінованих моделей знань, а шляхом генерації декількох альтернативних гіпотез із подальшим ранжуванням відповідно до ступеня їх достовірності.

На сьогодні розроблено декілька різних теорій, які дають змогу організувати ефективне логічне виведення в умовах невизначеності інформації. Тут найбільшого поширення набули такі методи:

- багатозначна логіка Лукасевича;
- ймовірнісний метод (байєсова логіка);
- метод коефіцієнтів упевненості;
- теорія функцій довіри (свідчень Демстера-Шефера);
- апарат нечіткої логіки

та інші.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Поясніть зміст поняття «невизначеність інформації».
2. Наведіть приклади задач прийняття рішень в умовах невизначеності.
3. Наведіть класифікацію та охарактеризуйте види невизначеності.
4. Поясніть поділ невизначеності на об'єктивну та суб'єктивну.
5. У чому полягає задача розв'язання конфліктів в логічному виведенні в умовах невизначеності.



ЛІТЕРАТУРА

[1], [2], [3], [4], [5], [7], [8], [9], [11]

§4.2. БАГАТОЗНАЧНА ЛОГІКА ЛУКАСЕВИЧА

Тризначна логіка Лукасевича є розширенням класичної булевої алгебри. Тут істинність будь-якого висловлювання оцінюється одним з трьох можливих значень:

- неможливе твердження (0);
- можливе, але не необхідне твердження (0,5);
- необхідне твердження (1).

У тризначній логіці Лукасевича діє закон виключного четвертого: будь-яке твердження є або неможливим, або можливим, але не необхідним, або необхідним.

Пояснімо застосування логіки Лукасевича на прикладі визначення стану комутаційного обладнання. У підрозділі 3.2 для визначення стану вимикача оцінювалась істинність висловлювання «КОНТАКТИ ВИМИКАЧА *sz* ЗАМКНЕНО». Дійсно, контакти вимикача можуть бути замкненими або розімкненими. Таки чином висловлювання «КОНТАКТИ ВИМИКАЧА *sz* ЗАМКНЕНО» може набувати одного з двох значень «ІСТИНА» або «ХИБНІСТЬ» відповідно. Разом з тим слід зважати на те, що експертні системи моделюють не фізичну природу предметної області, а процес мислення людини щодо цієї області. Тут до двох наведених значень додають третє, яке відповідає факту про те, що стан вимикача наразі невідомий. В тризначній логіці Лукасевича цьому факту відповідає константа «МОЖЛИВО». Дійсно, у разі відсутності інформації про стан вимикача істинність висловлювання «КОНТАКТИ ВИМИКАЧА *sz* ЗАМКНЕНО» можна інтерпретувати як можливість.

Для реалізації тризначної логіки Лукасевича використовують відповідні таблиці істинності. В табл. 4.1 наведено правила визначення істинності для одного та двох простих висловлювань.

Таким чином, тризначна логіка Лукасевича дозволяє моделювати найпростіші модальні оцінки висловлювань щодо предметної області експертної системи.

Наведені принципи організації тризначної логіки можна розширити на будь-яку кількість істиноносних значень. Реалізація такого підходу дозволяє отримати багатозначну логіку Лукасевича, яка характеризується єдиним хибним та будь-якою обмеженою кількістю істиноносних значень. Фізичний зміст таких істиноносних значень визначається, зазвичай, контекстом задачі, для розв'язання якої залучають апарат багатозначної логіки, наприклад, вони можуть характеризувати ступінь істинності відповідних фактів предметної області.

Таблиця 4.1. Таблиця істинності простих висловлювань

<i>A</i>	<i>B</i>	$\neg A$	$A \wedge B$	$A \vee B$	$A \rightarrow B$
0	0	1	0	0	1
0	0,5	1	0	0,5	1
0	1	1	0	1	1
0,5	0	0,5	0	0,5	0,5
0,5	0,5	0,5	0,5	0,5	1
0,5	1	0,5	0,5	1	1
1	0	0	0	1	0
1	0,5	0	0,5	1	0,5
1	1	0	1	1	1

Реалізація багатозначної логіки Лукасевича базується на правилах виконання логічних операцій:

$$\begin{aligned}\neg A &= 1 - A; \\ A \wedge B &= \min\{A, B\}; \\ A \vee B &= \max\{A, B\}; \\ A \rightarrow B &= \min\{1, 1 + B - A\}.\end{aligned}$$

Очевидно, що для тризначної логіки наведені вирази породжують таблицю істинності 4.1, а для двозначної – таблицю істинності 3.1.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає апарат тризначної логіки Лукасевича?
2. Поясніть різницю між апаратами числення висловлювань та тризначною логікою Лукасевича з погляду побудови експертних систем.
3. Наведіть таблицю істинності тризначної логіки Лукасевича для простих висловлювань.
4. Наведіть вирази для визначення істинності логічних операцій у багатозначній логіці Лукасевича.



ЛІТЕРАТУРА

[2], [4]

§4.3. БАЙЄСОВА ЛОГІКА

Байєсова логіка базується на математичному апараті теорії ймовірності. Тут кожному факту бази знань експертної системи ставиться в однозначну відповідність деяке додатне число в діапазоні $0...1$, яке характеризує достовірність (ймовірність істинності) цього факту. Так само оцінюють і правила логічного висновку. Тут числові значення в діапазоні $0...1$ характеризують ступінь довіри до відповідного правила. В термінах теорії ймовірності це можна трактувати як умовну ймовірність, тобто ступінь довіри до продукційного правила $A \rightarrow B$ можна оцінити умовною ймовірністю $p(B|A)$. Дійсно, відповідно до такого правила факт B має місце у тому разі, якщо має місце факт A . У свою чергу, умовна ймовірність $p(B|A)$ характеризує достовірність того, що подія B відбудеться у разі виникнення події A .

В задачах логічного виведення в умовах невизначеності з використанням ймовірнісних моделей широкого застосування набув підхід, який базується на відомій теоремі Байєса. Саме тому ймовірнісну модель логічного виведення часто називають байєсовою логікою.

Байєсова логіка – це метод формування логічних міркувань з використанням математичного апарату теорії ймовірності

Теорема Байєса (інколи – правило Байєса) дає змогу визначити ймовірність настання деякої події, коли відома лише часткова інформація про цю подію. Тут справедливий вираз

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

Особливість використання правила Байєса полягає у можливості реалізації двонаправленого логічного виведення від причини до наслідків та від наслідків до причин. У першому випадку це дозволяє реалізувати традиційне логічне виведення, а у другому – розв'язати задачу діагностики.

Проілюструємо розв'язання задачі діагностики із використанням байєсової логіки на наступному прикладі. Нехай відомо, що міжвиткові замкнення двигунів призводять до зниження продуктивності їх роботи у 70% випадків. Також відомо, що міжвиткові замикання спостерігаються у 0,5% двигунів, а зниження продуктивності – у 35%.

Позначимо події:

- A – міжвиткові замикання у двигуні;
- B – зниження продуктивності двигуна.

Тоді можна визначити наступні ймовірнісні показники

$$p(A)=0,005; \quad p(B)=0,35; \quad P(B|A)=0,7.$$

Застосування правила Байєса дозволяє визначити умовну ймовірність

$$p(A|B)=\frac{p(B|A)p(A)}{p(B)}=\frac{0,7 \cdot 0,005}{0,35}=0,01.$$

Отриманий результат свідчить про те, що зниження продуктивності двигунів пов'язано із міжвитковими замиканнями в 1% випадків.

Розглянемо особливості застосування апарату теорії ймовірності для реалізації логічного виведення.

Нехай задана проста імплікація $A \rightarrow B$. База знань експертної системи містить інформацію про достовірність факту A ($p(A)$), а також про надійність самого правила ($p(B|A)$). Робота машини логічного виведення полягає у визначенні достовірності (ймовірності) факту B .

Відповідно до правила композиції таку ймовірність можна оцінити за виразом

$$p(B)=p(B|A)p(A)+p(B|\bar{A})p(\bar{A})=p(B|A)p(A)+\underbrace{p(B|\bar{A})}_{?}(1-p(A)).$$

Аналіз отриманого виразу свідчить про те, що для точного оцінювання достовірності факту B необхідно додатково знати, чи може відбутися подія B , якщо подія A не відбудеться. Зазвичай бази знань таку інформацію у явному вигляді не містять, що унеможливорює коректне застосування ймовірнісної моделі логічного виведення. Тут можна лише оцінити граничні значення достовірності факту B :

$$\begin{aligned} p(B)_{\min} &= p(B|A)p(A); \\ p(B)_{\max} &= p(P|A)p(A)+(1-p(A)). \end{aligned}$$

На практиці, під час реалізації ймовірнісної моделі логічного виведення, зазвичай, умовно вважають, що подія B взагалі не може відбутися, якщо не відбулася подія A . За цих умов достовірність результуючого факту оцінюють за виразом

$$p(B) \approx p(B|A)p(A),$$

тобто як результат добутку достовірності посилення на ступінь довіри до правила.

Розглянемо далі більш складний випадок поєднання посилення імплікації кон'юнктивним правилом

$$(A \wedge B) \rightarrow C.$$

Тут база знань містить інформацію про достовірність фактів A та B , а також про ступінь довіри до самого правила. Задача машини логічного виведення полягає у визначені достовірності факту C .

Відповідно до правила композиції достовірність факту C визначають за виразом

$$p(C) = p(C|(A \wedge B)) \underbrace{p(A \wedge B)}_{?} + \underbrace{p(C|\overline{A \wedge B})}_{?} \cdot (1 - p(A \wedge B)),$$

де $p(A \wedge B)$ – ймовірність одночасності подій A та B .

Аналіз отриманого виразу свідчить про наявність двох невизначеностей:

1) невідомо, чи може статися подія C , якщо події A та B не відбулися одночасно;

2) невідома ймовірність одночасності подій A та B .

Проблема першої з наведених невизначеностей вирішується так само, як і для простої імплікації, тобто умовно вважається, що подія C не може відбутися, якщо події A та B не відбулися одночасно.

Ймовірність одночасності подій A та B , очевидно, не перевищує найменшу з ймовірностей подій A та B :

$$p(A \wedge B) \leq \min\{p(A), p(B)\}.$$

Саме таке значення зазвичай використовують для оцінювання ймовірності одночасності подій під час реалізації логічного виведення. Таким чином, ймовірність факту C розраховують за виразом

$$p(C) \approx p(C|(A \wedge B)) \min\{p(A), p(B)\}.$$

Розглянемо ще більш складний випадок розрахунку достовірності логічного висновку імплікації, яка містить диз'юнктивне посилення:

$$(A \vee B) \rightarrow C.$$

Тут база знань містить інформацію про достовірність фактів A та B , а також про ступінь довіри до самого правила. Задача машини логічного виведення полягає у визначені достовірності факту C .

Відповідно до правила композиції достовірність факту C визначають за виразом

$$p(C) = p(C|(A \vee B)) p(A \vee B) + p(C|\overline{A \vee B}) (1 - p(A \vee B)),$$

де $p(A \vee B)$ – ймовірність виникнення події A або події B . Таку ймовірність розраховують за третьою аксіомою Колмогорова:

$$p(A \vee B) = p(A) + p(B) - p(A \wedge B).$$

З урахуванням наведених вище допущень, достовірність логічного виведення слід визначати за виразом

$$p(C) \approx p(C|(A \vee B))(p(A) + p(B) - \min\{p(A), p(B)\}) = p(C|(A \vee B)) \max\{p(A), p(B)\}.$$

Аналіз отриманих виразів для визначення ймовірнісних показників свідчить про неминучу наявність похибок в отриманні результатів через додаткові невизначеності вихідної інформації.

Проілюструємо застосування ймовірнісних правил логічного виведення на наступному прикладі. Розглянемо систему продукційних правил, пов'язаних із виникненням лавини напруги в електричній системі:

Правило 1.

ЯКЩО відбулося різке збільшення навантаження,
АБО аварійно відключилася одна із живильних ліній
електропередачі,
ТО спостерігається провал напруги на шинах підстанції
споживача.

Правило 2.

ЯКЩО увімкнено всі секції конденсаторної установки,
ТО діапазон регулювання режиму реактивної потужності
обмежений.

Правило 3.

ЯКЩО спостерігається провал напруги на шинах підстанції
споживача
ТА діапазон регулювання режиму реактивної потужності
обмежений,
ТО в електричній системі спостерігається лавина напруги.

Нехай відомо, що за певних умов роботи електричної системи ймовірність різкого збільшення навантаження дорівнює 0,6; ймовірність аварійного відключення живлячої лінії електропередач складає 0,5; ймовірність увімкнення всіх секцій конденсаторної установки дорівнює 0,7. Відомо також, що ступені достовірності розглянутих правил складають 0,9; 0,8 та 0,95 відповідно. Необхідно визначити ймовірність виникнення лавини напруги в електричній системі.

Позначимо події системи логічного виведення:

- *A* – різке збільшення навантаження;
- *B* – аварійне відключення живильної лінії електропередач;
- *C* – провал напруги на шинах підстанції споживача;
- *D* – увімкнення всіх секцій конденсаторної установки;
- *E* – обмеження діапазону регулювання реактивної потужності;
- *F* – лавина напруги.

Тоді

$$\begin{aligned} p(A) &= 0,6; & p(B) &= 0,5; & p(D) &= 0,7; \\ p(C|A \vee B) &= 0,9; & p(E|D) &= 0,8; & p(F|C \wedge E) &= 0,95. \end{aligned}$$

Реалізація першого продукційного правила дозволяє визначити ймовірність провалу напруги на шинах підстанції споживача:

$$p(C) = p(C|(A \vee B)) \max\{p(A), p(B)\} = 0,9 \max\{0,6; 0,5\} = 0,9 \cdot 0,6 = 0,36.$$

Застосування другого правила дозволяє визначити ймовірність обмеження діапазону регулювання реактивної потужності:

$$p(E) = p(E|D) p(D) = 0,8 \cdot 0,7 = 0,56.$$

Третє продукційне правило дозволяє визначити ймовірність виникнення лавини напруги в електричній системі:

$$p(F) = p(F|(C \wedge E)) \min\{p(C), p(E)\} = 0,95 \min\{0,36; 0,56\} = 0,95 \cdot 0,36 = 0,342.$$

Таким чином, ймовірність виникнення лавини напруги в електричній системі складає 0,342.

В загальному випадку робота машини логічного виведення експертної системи, яка реалізує байєсову логіку, полягає у визначенні ймовірнісних показників можливих гіпотез розв'язання поставленої практичної задачі. Це дає змогу ранжувати гіпотези за зниженням ймовірності і видати користувачу список можливих рішень.

Головною гідністю апарату байєсової логіки є наявність розвинутої математичної теорії та повне математичне обґрунтування методу обчислень. Разом з тим, застосування класичної теорії ймовірності в задачах логічного виведення в умовах невизначеності пов'язано із додатковими невизначеностями, що призводить до похибок отриманих результатів. Такі похибки мають тенденцію накопичуватися і стають тим більшими, чим довший ланцюжок логічних міркувань.

? **КОНТРОЛЬНІ ЗАПИТАННЯ**

1. Дайте загальну характеристику ймовірнісних моделей логічного виведення в умовах невизначеності.

2. Поясніть можливості застосування байєсової логіки для розв'язання задач діагностики.

3. Поясніть причини виникнення додаткових невизначеностей під час реалізації логічного виведення за моделлю байєсової логіки.

4. Сформулюйте допущення, яке приймають під час реалізації логічного виведення простої імплікації.

5. Сформулюйте допущення, які приймають під час реалізації логічного виведення імплікації з кон'юнктивними посиленнями.

6. Сформулюйте допущення, які приймають під час реалізації логічного виведення імплікації з диз'юнктивними посиленнями.

7. У чому полягає робота машини логічного виведення експертних систем, які реалізують апарат байєсової логіки?

8. Наведіть позитивні та негативні властивості апарату байєсової логіки.



ЛІТЕРАТУРА

[1], [4], [5], [7], [8], [11], [14], [65]

§4.4. АПАРАТ КОЕФІЦІЄНТІВ УПЕВНЕНОСТІ

Метод коефіцієнтів упевненості було розроблено у Стенфордському університеті під час роботи над експертною системою MYCIN. Цей метод базується на евристичному числовому оцінюванні таких нечітких понять, як «точно», «скоріше за все», «можливо» та інших за допомогою шкали коефіцієнтів упевненості в діапазоні $-1...+1$. На рис. 4.2 представлена типова шкала таких коефіцієнтів.

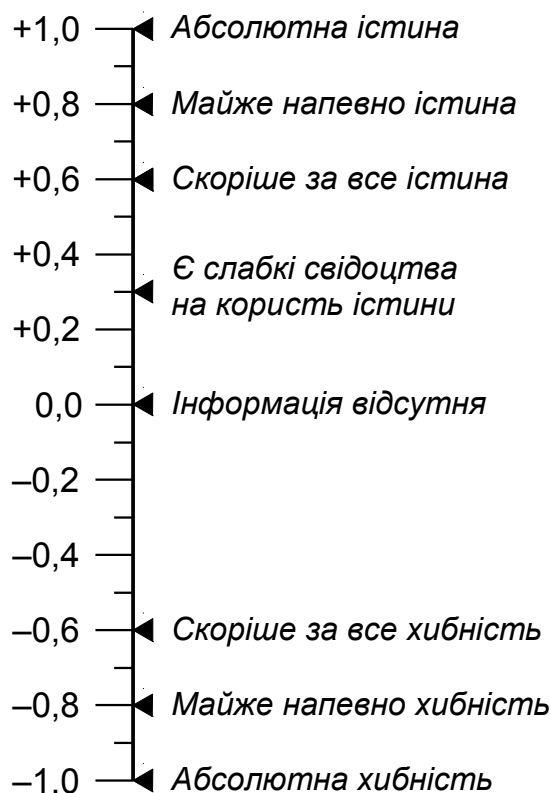


Рис. 4.2. Типова шкала коефіцієнтів упевненості

Тут оцінці «+1» відповідає абсолютна істинність відповідного міркування; оцінка «0» свідчить про відсутність інформації щодо істинності або хибності міркування; оцінці «-1» відповідає абсолютна хибність. Проміжні значення коефіцієнтів упевненості характеризують ступінь істинності або хибності міркування. Такі величини дещо схожі на ймовірності показники, але носять чисто емпіричний, суб'єктивний характер та не підпорядковуються жорстким вимогам теорії ймовірності.

Для реалізації методу коефіцієнтів упевненості введено дві оцінки, які відбивають ступінь істинності та ступінь хибності кожного міркування відповідно. Кожна з цих оцінок може приймати довільне числове значення в діапазоні 0...1:

$$\begin{aligned} 0 \leq MB \leq 1, \text{ якщо } MD = 0; \\ 0 \leq MD \leq 1, \text{ якщо } MB = 0, \end{aligned}$$

де MB – оцінка істинності; MD – оцінка хибності міркування.

Результуючий коефіцієнт упевненості визначають за виразом

$$CF = MB - MD.$$

Апарат коефіцієнтів упевненості передбачає встановлення оціночних функцій не тільки для окремих фактів, але і для правил логічного виведення, так само, як і в апараті байєсової логіки. Тут коефіцієнт упевненості заключення визначають за виразом:

$$CF = CF_{\text{пос}} CF_{\text{прав}},$$

де $CF_{\text{пос}}$ – коефіцієнт упевненості посилення; $CF_{\text{прав}}$ – коефіцієнт упевненості правила.

Для складних правил, посилення яких містять операції кон'юнкції та диз'юнкції, апарат коефіцієнтів упевненості передбачає використання таких співвідношень:

$$\begin{aligned} CF(P_1 \wedge P_2) &= \min\{CF(P_1), CF(P_2)\}; \\ CF(P_1 \vee P_2) &= \max\{CF(P_1), CF(P_2)\}. \end{aligned}$$

Проілюструємо застосування апарату коефіцієнтів упевненості до системи продукційних правил щодо виникнення лавини напруги в електричній системі, розглянутої в підрозділі 4.3.

Будемо вважати, що коефіцієнти упевненості окремих фактів та продукційних правил дорівнюють:

$$\begin{aligned} CF_A &= 0,6; & CF_B &= 0,5; & CF_D &= 0,7; \\ CF_{(A \vee B \rightarrow C)} &= 0,9; & CF_{(D \rightarrow E)} &= 0,8; & CF_{(C \wedge E \rightarrow F)} &= 0,95. \end{aligned}$$

Реалізація першого продукційного правила дозволяє визначити коефіцієнт упевненості факту провалу напруги на шинах споживацької підстанції:

$$CF_C = CF_{(A \vee B \rightarrow C)} \max\{CF_A, CF_B\} = 0,9 \max\{0,6; 0,5\} = 0,9 \cdot 0,6 = 0,36.$$

Застосування другого правила дає змогу визначити коефіцієнт упевненості факту обмеження діапазону регулювання реактивної потужності:

$$CF_E = CF_{(D \rightarrow E)} CF_D = 0,8 \cdot 0,7 = 0,56.$$

Третє продукційне правило дає змогу визначити коефіцієнт упевненості виникнення лавини напруги в електричній системі:

$$CF_F = CF_{(C \wedge E \rightarrow F)} \min\{CF_C, CF_E\} = 0,95 \min\{0,36; 0,56\} = 0,95 \cdot 0,36 = 0,342.$$

Таким чином, коефіцієнт упевненості факту виникнення лавини напруги в електричній системі становить 0,342.

Звернемо увагу на збіг числових характеристик фактів, отриманих з використанням апарату коефіцієнтів упевненості та за байєсовою логікою. Разом з тим такі характеристики мають різний фізичний зміст. Зокрема, апарат байєсової логіки оперує з об'єктивними ймовірностями, отриманими із результатів незалежних спостережень із використанням методів математичної статистики, на відміну від суб'єктивних коефіцієнтів упевненості, отриманих в результаті емпіричних експертних оцінок достовірності фактів та продукційних правил предметної області.

Декларативний характер баз знань експертних систем визначає можливість підтримки одного заключення декількома правилами. Тут, у разі застосування апарату коефіцієнтів упевненості, кожне з таких правил визначає своє значення коефіцієнта упевненості заключення. За цих умов необхідно організувати можливість комбінування свідочств на підтримку (або спростування) деякого заключення. Тут відокремлюють чотири випадки щодо комбінації значень коефіцієнтів упевненості:

- 1) $CF = CF_1 + CF_2 - CF_1 CF_2$, якщо $CF_1 > 0$, $CF_2 > 0$;
- 2) $CF = CF_1 + CF_2 + CF_1 CF_2$, якщо $CF_1 < 0$, $CF_2 < 0$;
- 3) $CF = \frac{CF_1 + CF_2}{1 - \min(|CF_1|, |CF_2|)}$, якщо $CF_1 CF_2 < 0$, та $CF_1 \neq \pm 1$, $CF_2 \neq \pm 1$;
- 4) $CF = \pm 1$, якщо $CF_1 = \pm 1$, або $CF_2 = \pm 1$.

Коефіцієнти упевненості заключень, які підтримуються трьома або більше правилами, розраховують за наведеною схемою послідовно додаючи по одному свідочству.

Розглянемо застосовування наведеної методики визначення результуючого значення коефіцієнтів упевненості на наступному прикладі. Нехай машина логічного виведення експертної системи виявила два незалежні ланцюги міркувань щодо можливості виникнення лавини

напруги, які привели до коефіцієнтів упевненості 0,342 та 0,25 відповідно:

$$CF_1=0,342; \quad CF_2=0,25.$$

Позначимо через CF_3 результуючий коефіцієнт упевненості. Оскільки обидва знайдені коефіцієнти упевненості є додатними, то для визначення результуючого значення слід скористатися першим виразом:

$$CF_3=CF_1+CF_2-CF_1CF_2=0,342+0,25-0,342\cdot 0,25=0,5065.$$

Нехай далі машина логічного виведення виявила ще один ланцюг, який заперечує можливість виникнення лавини напруги, тобто коефіцієнт упевненості досліджуваного факту є від'ємним:

$$CF_4=-0,175.$$

Тут для визначення результуючого значення коефіцієнта упевненості слід скористатися третім виразом:

$$CF_5=\frac{CF_3+CF_4}{1-\min(|CF_3|,|CF_4|)}=\frac{0,5065-0,175}{1-\min(|0,5065|,|-0,175|)}=0,294.$$

Якщо в подальшому машина логічного виведення виявить ще один або більше ланцюгів, які підтверджують або спростовують можливість виникнення лавини напруги, коефіцієнт упевненості відповідного факту слід скоригувати за наведеною вище схемою.

Апарат коефіцієнтів упевненості по суті є альтернативою байєсової логіки. Це визначає загальні риси в організації роботи машини логічного виведення на базі байєсової логіки та апарату коефіцієнтів упевненості. Зокрема, машина логічного виведення експертної системи, яка реалізує апарат коефіцієнтів упевненості, видає не єдине рішення поставленої задачі, а цілий список, ранжований за зниженням коефіцієнтів упевненості відповідних гіпотез.

Найбільш привабливою властивістю апарату коефіцієнтів упевненості є реалізація алгебри суб'єктивної ймовірності на відміну від об'єктивної ймовірності байєсової логіки. Загалом метод коефіцієнтів упевненості характеризується простотою реалізації, наочністю обчислень, відповідає здоровому глузду. Головний недолік методу полягає у відсутності математичного обґрунтування правомірності використання такого апарату. Також певні труднощі часто викликає визначення коефіцієнтів упевненості фактів та правил бази знань.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає суть методу коефіцієнтів упевненості?
2. Наведіть вираз для визначення коефіцієнту упевненості заключення.
3. Наведіть вираз для визначення коефіцієнту упевненості посилок складних правил.
4. Наведіть вираз для визначення коефіцієнту упевненості заключень, які підтримуються декількома правилами.
5. Наведіть порівняльну характеристику методів байєсової логіки та коефіцієнтів упевненості.
6. Наведіть позитивні та негативні риси апарату коефіцієнтів упевненості.



ЛІТЕРАТУРА

[1], [2], [4], [5], [7], [11], [65]

§4.5. ТЕОРІЯ СВІДОЦТВ ДЕМСТЕРА-ШЕФЕРА

Розглянуті раніше моделі реалізації логічного виведення в умовах інформаційної невизначеності – байєсова логіка та апарат коефіцієнтів упевненості – ґрунтуються на точковому оцінюванні нечітких понять та невизначених даних. На відміну від них теорія свідомств Демстера-Шефера є інтервальним методом з нижньою та верхньою границями оцінок. Тут кожній гіпотезі ставиться у відповідність певний діапазон значень, в якому знаходиться ступінь достовірності.

Нижня межа такого діапазону характеризує ступінь довіри до гіпотези. Така величина знаходиться в діапазоні $0...1$. Нуль означає відсутність свідомств на користь відповідної гіпотези; «1» означає повне підтвердження гіпотези. Зазвичай нижню межу діапазону позначають через *Bel* (від англійського *belief* – довіра).

Верхня межа інтервалу визначається ступенем правдоподібності гіпотези, тобто наявністю (або відсутністю) свідомств, які спростовують цю гіпотезу. Така величина також знаходиться в діапазоні $0...1$. Тут «1» свідчить про відсутність свідомств, які спростовують гіпотезу; «0» означає повне спростування гіпотези. Зазвичай верхню межу діапазону позначають через *Pls* (від англійського *plausibility* – правдоподібність). Верхню

межу інтервалу достовірності можна трактувати як найбільше можливе значення ступеню довіри до гіпотези та визначати за виразом

$$Pls(A) = 1 - Bel(\bar{A}),$$

де A – гіпотеза, для якої формують діапазон достовірності; $Bel(\bar{A})$ – ступінь довіри до гіпотез, альтернативних A .

На початку роботи механізм логічного виведення формує сукупність гіпотез можливих рішень задачі. Для кожної з гіпотез встановлюється діапазон достовірності $[0, 1]$, що свідчить про відсутність свідочств на користь та на спростування відповідної гіпотези. Далі, у разі виявлення свідочств на користь деякої гіпотези, відповідна ступінь довіри збільшується, а у разі виявлення свідчень, які спростовують гіпотезу, зменшується відповідна ступінь правдоподібності, в результаті чого інтервальна оцінка нечіткої інформації звужується. В результаті роботи машини логічного виведення діапазони достовірності можливих гіпотез звужуються, що дозволяє ранжувати ці гіпотези за ступенем довіри та (або) правдоподібності і видати користувачу отриманий список гіпотез.

Важливою характеристикою теорії свідочств Демстера-Шефера є розділення апіорної та апостеріорної ймовірностей. Пояснімо це на наступному прикладі. Нехай розглядається задача аналізу причин аварійних відключень повітряних ліній. Тут, часто, розрізняють три групи причин:

- 1) знос та старіння обладнання повітряних ліній;
- 2) несприятливі погодні умови в районі траси лінії;
- 3) впливи сторонніх організацій та фізичних осіб.

За відсутності додаткової інформації вважають, що апіорні ймовірності аварійних відключень однакові та дорівнюють 0,33. Звісно, що ступінь довіри до такої інформації є найнижчим, оскільки вона не підкріплена жодним свідочством. Якщо ж рівномірний розподіл ймовірності аварійних відключень підтверджено статистичними даними спостережень, то ступінь довіри є набагато вищий за тих самих ймовірнісних показників. Класичний апарат теорії ймовірності не передбачає спеціального механізму розділення апіорної та апостеріорної ймовірності, тобто підтверджений або не підтверджений рівномірний розподіл ймовірностей подається однаковими показниками.

Навпаки, в теорії Демпстера-Шефера свідчення на користь або на спростування деякої гіпотези призводять до звуження інтервалу достовірності. Тобто ширина інтервалу може служити мірою невпевненості гіпотези за наявного набору свідочств. І чим вужчий такий інтервал, тим більший ступінь довіри до відповідної інформації.

Визначення числових характеристик інтервалів достовірності базується на понятті маси упевненості. Маса упевненості асоціюється із середовищем множини альтернативних гіпотез та їх можливих комбіна-

цій, розподіляється між елементами цієї множини та характеризує ступені довіри до гіпотез.

Певним чином маса упевненості схожа на ймовірнісні показники істинності відповідних гіпотез. Разом з тим, на відміну від класичної теорії ймовірності, маса упевненості характеризує лише достовірність однієї гіпотези (або комбінації гіпотез) та не визначає достовірності альтернативних. Так, наприклад, якщо база знань експертної системи містить свідчення про те, що частка аварійних відключень повітряних ліній через несприятливі погодні умови складає 55%, то відповідно до класичної теорії ймовірності це свідчення автоматично означає, що доля аварійних відключень з інших причин дорівнює 45%. Таке заключення є справедливим для об'єктивних ймовірностей, підкріплених тривалими спостереженнями та пов'язаними, наприклад, з задачами ретроспективного аналізу причин аварій на лініях. Однак, теорія Демпстера-Шефера орієнтована на обробку суб'єктивних даних і може застосовуватися для прогнозування можливих аварійних відключень на найближчу та віддалену перспективу. Відповідно до цієї теорії, наведене свідчення лише встановлює масу упевненості для відповідної гіпотези 0,55. Залишок $1 - 0,55 = 0,45$ розглядають як «вільну» масу, не пов'язану з жодною гіпотезою, яку буде розподілено у разі виявлення нових свідчень на користь альтернативних гіпотез. Разом з тим, наведене свідчення визначає зниження верхніх меж альтернативних гіпотез та їх комбінацій до значення 0,45. Тобто ступінь довіри до таких гіпотез, наприклад, до гіпотези про те, що аварійні відключення відбуваються з причин зносу та старіння обладнання, не може перевищувати 0,45, що відповідає здоровому глузду.

Для розрахунку числових характеристик ступенів довіри до гіпотез або груп гіпотез використовують вираз

$$Bel(A) = \sum_{B \subseteq A} m(B),$$

де $m(B)$ – маса достовірності гіпотези, або групи гіпотез B ; запис $B \subseteq A$ означає, що підсумовування слід виконувати для мас всіх гіпотез B , які підтверджують достовірність гіпотези A .

Проілюструємо розрахунок інтервалів достовірності на наступному прикладі. Нехай база знань експертної системи містить наступні свідчення щодо розподілу причин аварійних відключень повітряних ліній:

- 1) з причин зносу та старіння обладнання – 29%;
- 2) з причин несприятливих погодних умов – 55%;
- 3) з причин дій сторонніх організацій та фізичних осіб – 7%;
- 4) з невизначених причин – 9%.

В табл. 4.2 наведено значення мас упевненості, ступенів довіри та правдоподібності окремих гіпотез та їх комбінацій. Тут $H1$, $H2$ та $H3$ –

гіпотези щодо аварійних відключень повітряних ліній з причин старіння обладнання, несприятливих погодних умов та дій фізичних осіб відповідно.

Таблиця 4.2. Характеристики гіпотез відповідно до теорії Демпстера-Шефера

<i>H</i>	<i>m(H)</i>	<i>Bel(H)</i>	<i>Pls(H)</i>
\emptyset	0	0	0
$\{H1\}$	0,29	0,29	0,38
$\{H2\}$	0,55	0,55	0,64
$\{H3\}$	0,07	0,07	0,16
$\{H1, H2\}$	0	0,84	0,93
$\{H1, H3\}$	0	0,36	0,45
$\{H2, H3\}$	0	0,62	0,71
$\{H1, H2, H3\}$	0,09	1	1

Пояснимо наведені результати обчисленнями:

$$Bel(\{H1\}) = m(\emptyset) + m(\{H1\}) = 0 + 0,29 = 0,29;$$

$$Bel(\{H1, H2\}) = m(\emptyset) + m(\{H1\}) + m(\{H2\}) + m(\{H1, H2\}) = 0 + 0,29 + 0,55 + 0 = 0,84;$$

$$Bel(\{H1, H2, H3\}) = m(\emptyset) + m(\{H1\}) + m(\{H2\}) + m(\{H3\}) + m(\{H1, H2\}) + m(\{H1, H3\}) + m(\{H2, H3\}) + m(\{H1, H2, H3\}) = 0 + 0,29 + 0,55 + 0,07 + 0 + 0 + 0 + 0,09 = 1;$$

$$Pls(\{H1\}) = 1 - Bel(\neg\{H1\}) = 1 - Bel(\{H2, H3\}) = 1 - 0,62 = 0,38;$$

$$Pls(\{H1, H2\}) = 1 - Bel(\neg\{H1, H2\}) = 1 - Bel(\{H3\}) = 1 - 0,07 = 0,93;$$

$$Pls(\{H1, H2, H3\}) = 1 - Bel(\neg\{H1, H2, H3\}) = 1 - Bel(\emptyset) = 1 - 0 = 1.$$

Дійсно, свідчення про те, що аварійні відключення через старіння та знос обладнання відбуваються у 29% випадків означає, що маса достовірності відповідної гіпотези, а також ступінь довіри до неї дорівнюють 0,29. Одночасно ступінь правдоподібності визначається свідченнями до альтернативних гіпотез, тобто свідченнями про те, що аварійні відключення відбуваються з причин несприятливих погодних умов, або дій сторонніх організацій та фізичних осіб. Поєднання таких свідчень дає ступінь довіри 0,62, що визначає ступінь правдоподібності 0,38.

Зазначимо, що в теорії Демпстера-Шефера масу достовірності універсальної множини (у прикладі це $\{H1, H2, H3\}$) слід розглядати як «вільну» масу, не поділену між альтернативними гіпотезами.

Декларативний характер баз знань експертних систем може призводити до виникнення конфліктів між свідченнями, отриманими з різних джерел. Тут для комбінації свідочств теорія Демпстера-Шефера передбачає використання евристичного правила Демпстера

$$m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B)m_2(C),$$

де $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$ – міра конфлікту між наборами мас; множник

$\frac{1}{1-K}$ називають коефіцієнтом нормалізації; запис $B \cap C = A$ означає, що підсумовування слід виконувати за тими групами гіпотез, перетинання яких визначає гіпотезу A ; $B \cap C = \emptyset$ означає, що підсумовування слід виконувати за групами гіпотез, які не перетинаються.

Проілюструємо застосування правила Демпстера на наступному прикладі. Нехай база знань експертної системи містить два незалежні свідчення про причини аварійних відключень ліній електропередач:

1) частка відключень з причин старіння обладнання, або через несприятливі природні умови складає 84%;

2) частка відключень через старіння обладнання, або через дії сторонніх організацій та фізичних осіб складає 36%.

Перше свідчення визначає маси упевненості:

$$\begin{aligned} m_1(\{H1, H2\}) &= 0,84; \\ m_1(\{H1, H2, H3\}) &= 0,16. \end{aligned}$$

Другому свідченню відповідають маси упевненості:

$$\begin{aligned} m_2(\{H1, H3\}) &= 0,36; \\ m_2(\{H1, H2, H3\}) &= 0,64. \end{aligned}$$

Тут $m_1(\{H1, H2, H3\})$ та $m_2(\{H1, H2, H3\})$ являють собою «вільні» маси відповідних свідчень. Всі інші маси упевненості можливих гіпотез та їх поєднань дорівнюють нулю.

Застосування правила Демпстера дає результати, наведені в табл. 4.3.

Додамо до отриманого результату ще одне свідчення про те, що частка аварій з причин несприятливих погодних умов складає 55%, тобто

$$\begin{aligned} m_4(\{H2\}) &= 0,55; \\ m_4(\{H1, H2, H3\}) &= 0,45. \end{aligned}$$

Результати обчислень за правилом Демпстера дають результати, наведені в табл. 4.4.

Таблиця 4.3. Результати застосування правила Демпстера

1	2	3
$m_1(\{H1, H2\}) = 0,84$	$m_2(\{H1, H3\}) = 0,36$	$m_3(\{H1\}) = 0,3024$
$m_1(\{H1, H2\}) = 0,84$	$m_2(\{H1, H2, H3\}) = 0,64$	$m_3(\{H1, H2\}) = 0,5376$
$m_1(\{H1, H2, H3\}) = 0,16$	$m_2(\{H1, H3\}) = 0,36$	$m_3(\{H1, H3\}) = 0,0576$
$m_1(\{H1, H2, H3\}) = 0,16$	$m_2(\{H1, H2, H3\}) = 0,64$	$m_3(\{H1, H2, H3\}) = 0,1024$

Таблиця 4.4. Результати повторного застосування правила Демпстера

3	4	5
$m_3(\{H1\}) = 0,3024$	$m_4(\{H2\}) = 0,55$	$m_5(\emptyset) = 0,16632$
$m_3(\{H1\}) = 0,3024$	$m_4(\{H1, H2, H3\}) = 0,45$	$m_5(\{H1\}) = 0,13602$
$m_3(\{H1, H2\}) = 0,5376$	$m_4(\{H2\}) = 0,55$	$m_5(\{H2\}) = 0,29568$
$m_3(\{H1, H2\}) = 0,5376$	$m_4(\{H1, H2, H3\}) = 0,45$	$m_5(\{H1, H2\}) = 0,24192$
$m_3(\{H1, H3\}) = 0,0576$	$m_4(\{H2\}) = 0,55$	$m_5(\emptyset) = 0,03168$
$m_3(\{H1, H3\}) = 0,0576$	$m_4(\{H1, H2, H3\}) = 0,45$	$m_5(\{H1, H3\}) = 0,02592$
$m_3(\{H1, H2, H3\}) = 0,1024$	$m_4(\{H2\}) = 0,55$	$m_5(\{H2\}) = 0,05632$
$m_3(\{H1, H2, H3\}) = 0,1024$	$m_4(\{H1, H2, H3\}) = 0,45$	$m_4(\{H1, H2, H3\}) = 0,04608$

Отриманий результат містить дві маси, пов'язані із гіпотезою $H2$, тому результуюча маса дорівнює їх сумі

$$m_5(\{H2\}) = 0,29568 + 0,05632 = 0,325.$$

Пусті множини у результатах застосовування правила Демпстера свідчать про наявність конфлікту між свідченнями. Міра конфлікту тут складає

$$K = \sum m_s(\emptyset) = 0,16632 + 0,03168 = 0,198,$$

а коефіцієнт нормалізації дорівнює

$$\frac{1}{1-K} = \frac{1}{1-0,198} = 1,24688.$$

Таким чином, маси упевненості після двократного використання правила Демпстера складають:

$$\begin{aligned}
m_5(\{H1\}) &= 1,24688 \cdot 0,13608 = 0,1697; \\
m_5(\{H2\}) &= 1,24688 \cdot 0,325 = 0,405236; \\
m_5(\{H1, H2\}) &= 1,24688 \cdot 0,24162 = 0,30127; \\
m_5(\{H1, H3\}) &= 1,24688 \cdot 0,02592 = 0,03232; \\
m_5(\{H1, H2, H3\}) &= 1,24688 \cdot 0,04608 = 0,05746.
\end{aligned}$$

Визначені в такий спосіб маси упевненості гіпотез та їх поєднань дозволяють розрахувати інтервали достовірності гіпотез за наведеною вище схемою, що представлено в табл. 4.5.

Таблиця 4.5. Характеристики гіпотез після повторного застосування правила Демпстера

<i>H</i>	<i>m(H)</i>	<i>Bel(H)</i>	<i>Pls(H)</i>
∅	0	0	0
{H1}	0,1697	0,1697	0,5948
{H2}	0,4052	0,4052	0,8303
{H3}	0	0	0,4251
{H1, H2}	0,3013	0,5749	1
{H1, H3}	0,0323	0,1697	0,5948
{H2, H3}	0	0,4052	0,8303
{H1, H2, H3}	0,0575	0,966	1

Головна перевага теорії Демстера-Шефера полягає у використанні інтервальних оцінок достовірності інформації, що дає змогу реалізувати алгебру суб'єктивних ймовірностей та розділити апіорні та апостеріорні дані.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть основні характеристики апарату теорії Демпстера-Шефера.

2. Поясніть різницю між масою упевненості та ймовірносними показниками достовірності.

3. Наведіть вирази для визначення нижньої та верхньої границь інтервальних оцінок достовірності.

4. Наведіть вираз емпіричного правила Демпстера.

5. Наведіть позитивні властивості апарату теорії Демпстера-Шефера.

[1], [4], [5], [7], [11], [65], [8]

§4.6. НЕЧІТКА ЛОГІКА

В основі апарату нечіткої логіки лежать поняття лінгвістичної змінної та нечіткої множини.

Лінгвістичною називають змінну, значення якої визначають сукупністю вербальних (словесних) характеристик певної властивості, наприклад, «збільшене навантаження», «знижена напруга», «дефіцит потужності»

Нечітку множину зазвичай визначають від супротивного по відношенню до поняття традиційної (чіткої) множини.

Під традиційною множиною розуміють сукупність об'єктів предметної області, відокремлених за певною ознакою. Наприклад, множини повітряних ліній електропередачі, несиметричних режимів електричних систем, кліматичних умов в районі траси лінії тощо. Для визначення об'єктів предметної області, які входять до складу тієї чи іншої множини можна запропонувати двозначну логічну характеристичну функцію, яка визначає належність певного об'єкту до множини та приймає значення «+1» («TRUE»/«ИСТИНА») у разі належності та «0» («FALSE»/«ХИБНІСТЬ») у протилежному випадку:

$$\mu(x) = \begin{cases} +1 \text{ (TRUE)}, & \text{якщо } x \in A; \\ 0 \text{ (FALSE)}, & \text{якщо } x \notin A, \end{cases}$$

де A – множина, належність до якої визначає характеристична функція.

На відміну від цього положення для нечітких множин визначають функцію належності, яка може приймати довільне значення в діапазоні $0...1$, та характеризувати таким чином суб'єктивну ступінь упевненості експерта у належності елементу до відповідної множини.

Нечіткою множиною називають сукупність пар елементів предметної області та відповідних значень функції належності цих елементів до множини: $M = \left\{ \frac{x_i}{\mu(x_i)} \right\}$.

Таким чином, апарат нечітких множин дає змогу визначити числові оцінки для будь-яких лінгвістичних змінних, зокрема для таких модальних характеристик, як «багато», «швидко», «часто» тощо. Наприклад, можна визначити нечіткі множини для лінгвістичних змінних «знижена напруга», «нормальна напруга», «підвищена напруга», що дає змогу реалізувати оцінювання режиму напруги в електричній мережі:

ЗНИЖЕНА НАПРУГА, кВ:

$$\left\{ \frac{96}{1}, \frac{98}{1}, \frac{100}{1}, \frac{102}{0,8}, \frac{104}{0,6}, \frac{106}{0,4}, \frac{108}{0,2}, \frac{110}{0}, \dots \right\};$$

НОРМАЛЬНА НАПРУГА, кВ:

$$\left\{ \dots, \frac{100}{0}, \frac{102}{0,2}, \frac{104}{0,4}, \frac{106}{0,6}, \frac{108}{0,8}, \frac{110}{1}, \frac{112}{0,8}, \frac{114}{0,6}, \frac{116}{0,4}, \frac{118}{0,2}, \frac{120}{0}, \dots \right\};$$

ПІДВИЩЕНА НАПРУГА, кВ:

$$\left\{ \dots, \frac{110}{0}, \frac{112}{0,2}, \frac{114}{0,4}, \frac{116}{0,6}, \frac{118}{0,8}, \frac{120}{1}, \frac{122}{1}, \frac{124}{1}, \dots \right\}.$$

На рис. 4.3 представлено графічну інтерпретацію нечітких множин, які визначають розглянуті лінгвістичні змінні.

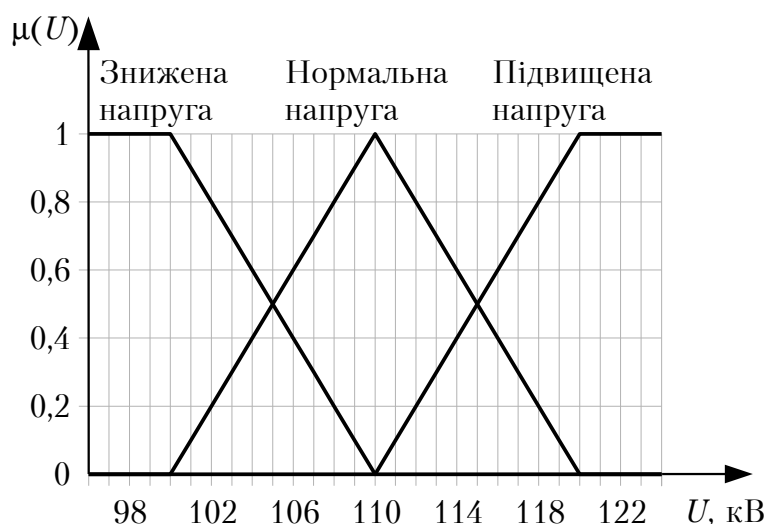


Рис. 4.3. Нечіткі множини аналізу режиму напруги

Основні особливості нечітких множин полягають у наступному.

1. Функція належності є суб'єктивною оцінкою і, на відміну від об'єктивної ймовірності, не може бути визначена однозначно. Це означає, що різні експерти можуть назначати різні значення функції належності для одних і тих самих елементів предметної області. Так, наприклад, за думкою одного експерта напругу 114 кВ слід вважати підвищеною із упевненістю 0,3, а за думкою іншого – із упевненістю 0,4.

2. Самі функції належності слід розглядати як нечіткі. Дійсно, дуже рідко експерт може чітко та аргументовано пояснити, наприклад, чому

він оцінив напругу 114 кВ як підвищену з упевненістю 0,4, а не 0,3, або 0,5.

3. Незважаючи на суб'єктивність функції належності, її не можна визначати абсолютно довільно. Природно, що для лінгвістичної змінної «підвищена напруга» ступінь упевненості для 114 кВ не може бути більшою за ступінь упевненості для 118 кВ.

Для представлення функції належності нечітких множин найбільшого поширення набули два основних способи: прямий та аналітичний.

Прямий спосіб подання функцій належності полягає у безпосередньому призначенні експертом кожному елементу нечіткої множини відповідного значення функції належності. Саме таким способом вище було представлено нечіткі множини зниженої, нормальної та підвищеної напруги.

Аналітичний спосіб полягає у використанні LR -функцій належності нечітких множин. Тут LR -функції являють собою двоскладові функції лівого (L – left) та правого (R – right) крил функції належності.

Загалом функції належності можуть бути довільними. Найпростішою тут є кусочно-лінійні функції:

$$\mu(x) = \begin{cases} a_1(x-x_0) + b_1, & \text{якщо } x \leq X; \\ a_2(x-x_0) + b_2, & \text{якщо } x \geq X, \end{cases}$$

або

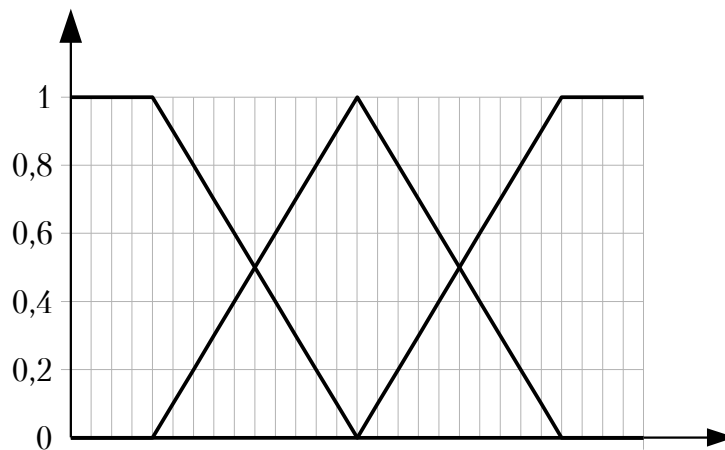
$$\mu(x) = \begin{cases} 1, & \text{якщо } x \leq (\geq) X; \\ a(x-x_0) + b, & \text{якщо } x \geq (\leq) X. \end{cases}$$

Проте найбільшого поширення набули s -подібні функції належності, наприклад, сигмоїдні функції:

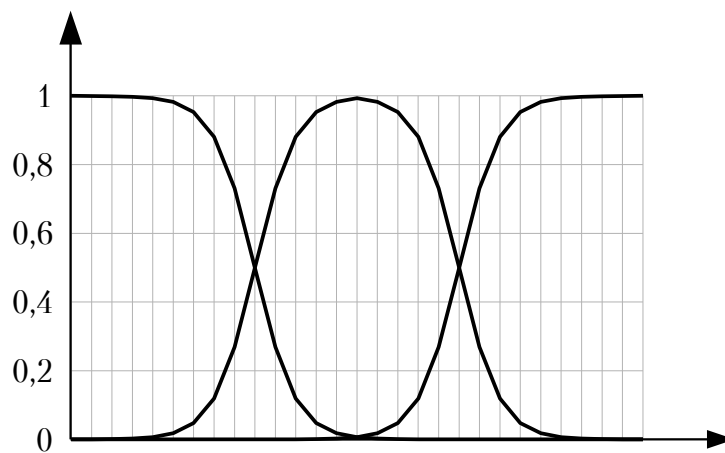
$$\mu(x) = \begin{cases} \frac{1}{1 + e^{-(x-x_0)}}, & \text{якщо } x \leq X; \\ \frac{1}{1 + e^{-(x_0-x)}}, & \text{якщо } x \geq X. \end{cases}$$

Нечіткі множини з кусочно-нелінійними та s -подібними функціями належності представлено на рис 4.4 а та б відповідно.

Лінгвістичні змінні та нечіткі множини дають змогу реалізувати моделі нечіткої логіки, тобто моделі логічного виведення, в яких відсутнє чітке визначення операцій та функцій, які поєднують вхідні та вихідні параметри.



а)



б)

Рис. 4.4. Нечіткі множини з кусочно-лінійними (а) та s-подібними (б) функціями належності

Нечітка логіка – це метод формування логічних міркувань з використанням лінгвістичних змінних, представлених нечіткими множинами

Тут поняття «ІСТИНА» та «ХИБНІСТЬ» розглядають як лінгвістичні змінні, представлені нечіткими множинами із такими, наприклад, функціями належності:

$$\begin{cases} \text{ІСТИНА:} & \mu_{\text{ІСТИНА}}(x) = x & \text{для } 0 \leq x \leq 1; \\ \text{ХИБНІСТЬ:} & \mu_{\text{ХИБНІСТЬ}}(x) = 1 - \mu_{\text{ІСТИНА}}(x) = 1 - x & \text{для } 0 \leq x \leq 1. \end{cases}$$

Апарат нечіткої логіки базується на понятті відношення між нечіткими множинами. Звісно, такі відношення також є нечіткими. Нечіткі відношення встановлюють на декартовому добутку відповідних нечітких

множин, тобто на всіх можливих комбінаціях елементів нечітких множин. В загальному випадку нечітке відношення визначається виразом

$$P = \left\{ \frac{(x_1, x_2)}{\mu_P(x_1, x_2)} \right\},$$

де $\mu_P(x_1, x_2)$ – функція належності елементів x_1 та x_2 до відношення P , тобто ступінь виконання відношення P між елементами x_1 та x_2 .

Одним з найпоширеніших способів визначення нечіткого відношення між множинами є використання логічного добутку відповідних множин:

$$P_{A \times B} = \left\{ \frac{(x_1, x_2)}{\mu_P(x_1) \wedge \mu_P(x_2)} \right\} = \left\{ \frac{(x_1, x_2)}{\min(\mu_P(x_1), \mu_P(x_2))} \right\}.$$

Часто нечіткі відношення використовують для моделювання продукційних правил «ЯКЩО A ТО B » та реалізації на їх основі логічного виведення.

Розглянемо формування нечіткої бази знань для розв'язання задачі оцінювання показників надійності функціонування електроустаткування. Із загальної теорії надійності відомо, що життєвий цикл експлуатації електротехнічних об'єктів можна розділити на три характерні зони, як представлено на рис. 4.5:

I) інтервал приробітку, коли через неякісну збірку схем складних пристроїв, недотримання вимог монтажу та інші внутрішні дефекти інтенсивність відмов висока;

II) інтервал нормальної експлуатації устаткування з малою інтенсивністю відмов;

III) інтервал старіння та зносу обладнання електроустаткування з високою інтенсивністю відмов.

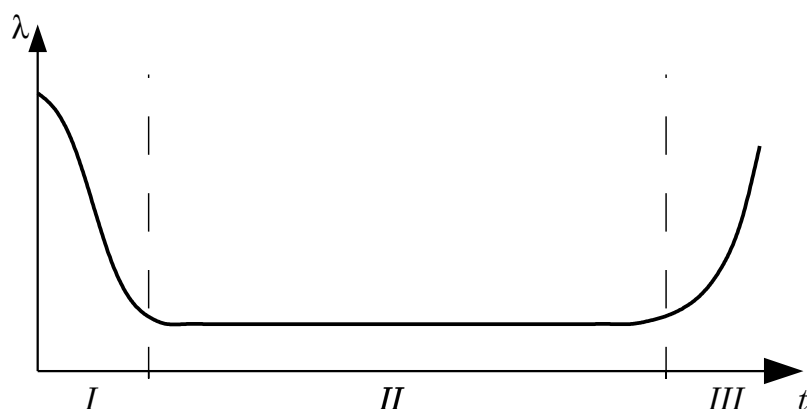


Рис. 4.5. Інтервали життєвого циклу експлуатації електроустаткування

Залежність інтенсивності відмов електроустаткування від строку експлуатації можна представити сукупністю нечітких правил наступного змісту.

Правило 1.

ЯКЩО період експлуатації електроустаткування дуже малий
ТО інтенсивність відмов висока

Правило 2.

ЯКЩО період експлуатації електроустаткування середній
ТО інтенсивність відмов мала

Правило 3.

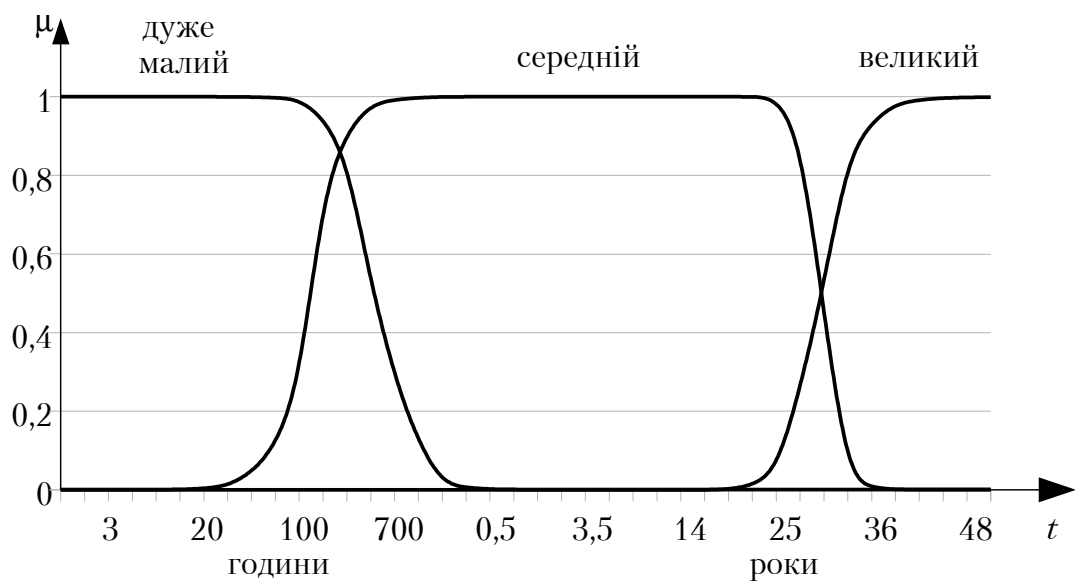
ЯКЩО період експлуатації електроустаткування великий
ТО інтенсивність відмов висока

Аналіз представлених правил свідчить про те, що для реалізації логічного виведення необхідно визначити лінгвістичні змінні, які характеризують період експлуатації електроустаткування (дуже малий, середній та великий) та інтенсивність відмов (мала та висока). Нечіткі множини, які відповідають таким лінгвістичним змінним, представлені функціями належності, наведеними на рис. 4.6. Тут, на рис. 4.6 а і в подальшому використана псевдологаріфмічна шкала часу експлуатації устаткування.

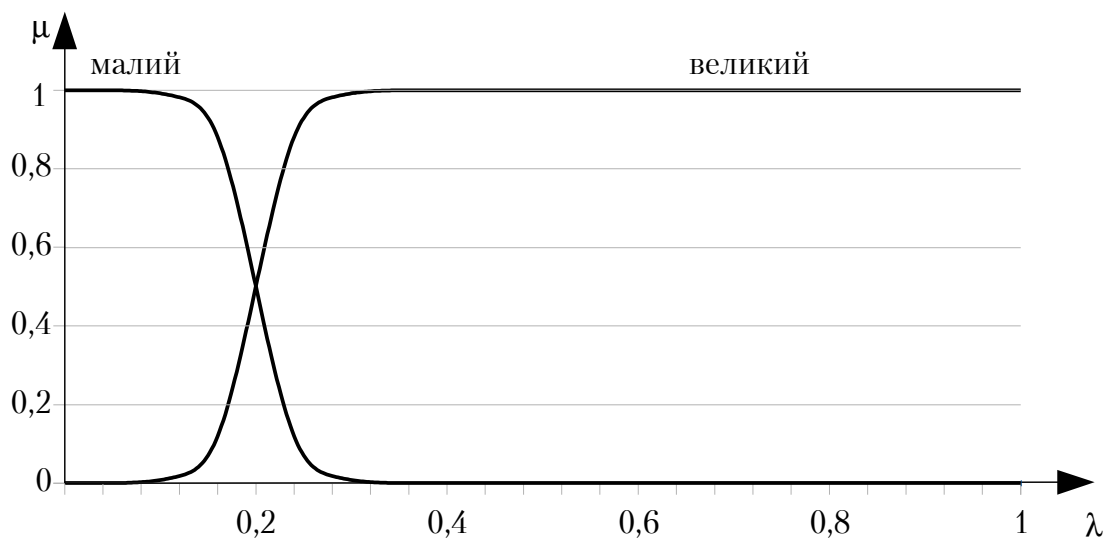
Нечіткі відношення, які моделюють продукційні правила оцінювання надійності функціонування електроустаткування визначаються правилом логічного добутку між значеннями функцій належності відповідних елементів нечітких множин та можуть бути представленими тривимірними діаграмами, наведеними на 4.7.

Такі відношення часто подають у табличному вигляді. Наприклад, в таблиці 4.6 представлено нечітке відношення, яке моделює правило 1 (*якщо період експлуатації електроустаткування дуже малий, то інтенсивність відмов висока*).

Часто реалізація нечіткої логіки полягає у логічному виведенні за продукцією $A \rightarrow B$ для нечіткої множини A' , яка лише певною мірою подібна до A . Таке логічне виведення, зазвичай, виконують із застосуванням композиційного правила, принципи застосування якого наведені нижче.



а)



б)

Рис. 4.6. Нечіткі множини лінгвістичних змінних оцінювання інтенсивності відмов:

а) дуже малий, середній та великий періоди експлуатації;

б) малі та великі інтенсивності відмов

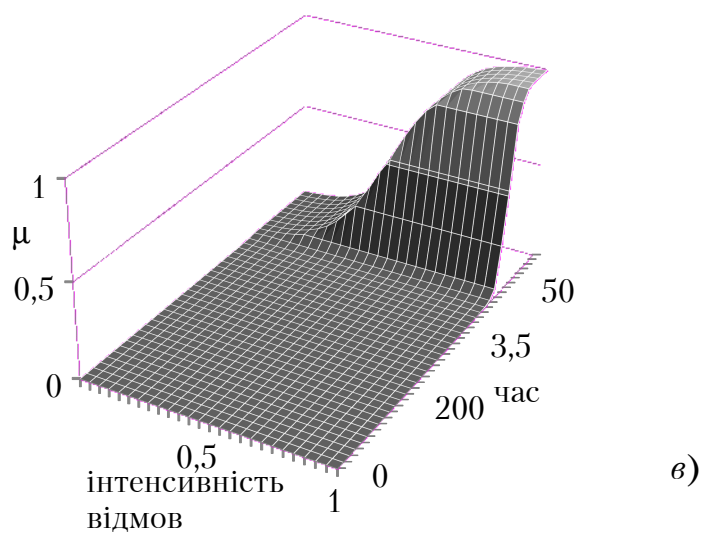
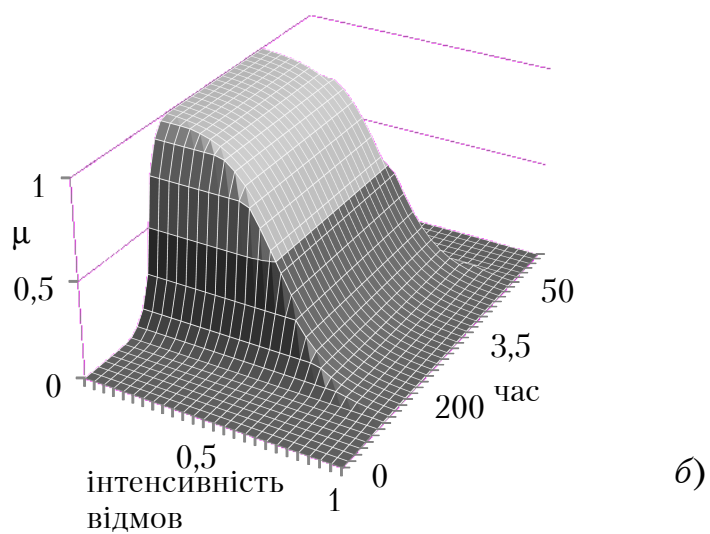
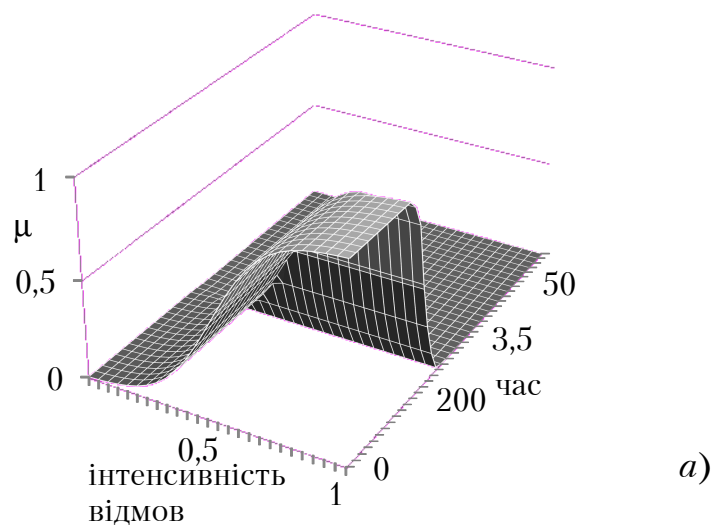


Рис. 4.7. Ілюстрація нечітких відношень продукційних правил оцінювання надійності: а) правило 1; б) правило 2; в) правило 3.

Таблиця 4.6. Нечітке відношення високої інтенсивності відмов за дуже малого періоду експлуатації

	$\frac{0}{0,00}$	$\frac{0,12}{0,02}$	$\frac{0,16}{0,12}$	$\frac{0,2}{0,50}$	$\frac{0,24}{0,88}$	$\frac{0,28}{0,98}$	$\frac{0,32}{1,00}$	$\frac{0,64}{1,00}$	$\frac{1}{1,00}$
10/1,00	0,00	0,02	0,12	0,50	0,88	0,98	1,00	1,00	1,00
50/1,00	0,00	0,02	0,12	0,50	0,88	0,98	1,00	1,00	1,00
100/0,99	0,00	0,02	0,12	0,50	0,88	0,98	0,99	0,99	0,99
200/0,95	0,00	0,02	0,12	0,50	0,88	0,95	0,95	0,95	0,95
300/0,84	0,00	0,02	0,12	0,50	0,84	0,84	0,84	0,84	0,84
500/0,53	0,00	0,02	0,12	0,50	0,53	0,53	0,53	0,53	0,53
700/0,29	0,00	0,02	0,12	0,29	0,29	0,29	0,29	0,29	0,29
1000/0,12	0,00	0,02	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2000/0,02	0,00	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
3000/0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
5000/0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Нехай задано два нечітких відношення між нечіткими множинами A та B , а також B та C :

$$P_1: A \times B; \quad P_2: B \times C.$$

Комбінація цих відношень дозволяє визначити відношення між множинами A та C

$$P = P_1 \circ P_2: A \times C.$$

Для розрахунку функцій належності композиції двох відношень використовують операцію максимінної згортки за виразом

$$\mu_{(P_1 \circ P_2)}(a, c) = \max \min(\mu_{P_1}(a, b); \mu_{P_2}(b, c)),$$

де a, b, c – елементи нечітких множин A, B та C відповідно.

На рис. 4.8 проілюстрована процедура логічного виведення за композиційним правилом для нечіткої множини A' , подібної до A в продукції $A \rightarrow B$. Тут елементи вектора нечіткої множини A' послідовно накладають на стовпці таблиці відношення $A \times B$ і для кожної пари елементів обирають найменше. Далі, у кожному стовпці обирають максимальний елемент, який записують у вихідний вектор функції належності.

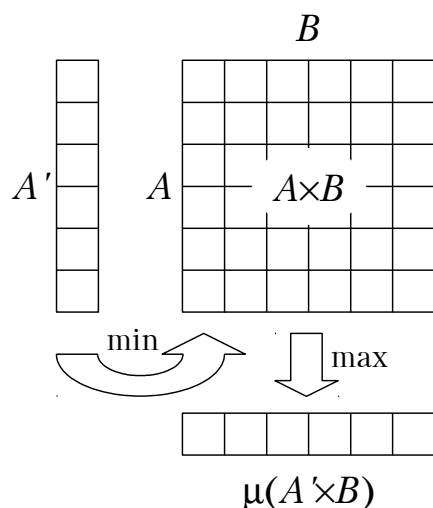


Рис. 4.8. Ілюстрація процедури максимінної згортки

Розглянемо, наприклад, реалізацію логічного виведення за наведеними вище продукційними правилами для малого періоду експлуатації електротехнічного устаткування. Визначимо лінгвістичну змінну «малий період експлуатації» по відношенню до поняття «дуже малий період»:

$$\mu_{\text{м.п.}} = \sqrt{\mu_{\text{д.м.п.}}}$$

Відповідну нечітку множину можна подати прямим способом:

МАЛИЙ ПЕРІОД:

$$\left\{ \frac{10}{1}; \frac{50}{1}; \frac{100}{0,997}; \frac{200}{0,973}; \frac{300}{0,917}; \frac{500}{0,726}; \frac{700}{0,573}; \frac{1000}{0,345}; \frac{2000}{0,127}; \frac{3000}{0,069}; \frac{5000}{0} \right\}$$

Проведемо максимінну згортку з лінгвістичною змінною «малий період експлуатації» та таблицею 4.6 відношення за першим продукційним правилом. Таку операцію виконуємо за наступною схемою. Накладаємо нечітку множину малого періоду експлуатації послідовно на всі стовпці таблиці 4.6. З кожної пари значень функції належності обираємо мінімальне, а з отриманого набору значень – максимальне. В результаті отримуємо нечітку множину, яка визначає інтенсивність відмов за малого періоду експлуатації устаткування (див. рис. 4.9):

$$\left\{ \frac{0}{0}; \frac{0,12}{0,018}; \frac{0,2}{0,5}; \frac{0,24}{0,881}; \frac{0,28}{0,982}; \frac{0,32}{0,998}; \frac{0,64}{1}; \frac{1}{1} \right\}.$$

Таким чином, апарат нечіткої логіки дозволяє організувати розв'язання практичних задач на основі вербального представлення вихідної інформації. Разом з тим, слід пам'ятати, що головним недоліком апарату, зокрема операції максимінної згортки та її аналогів є суттєва нечутливість моделі нечіткого логічного виведення.

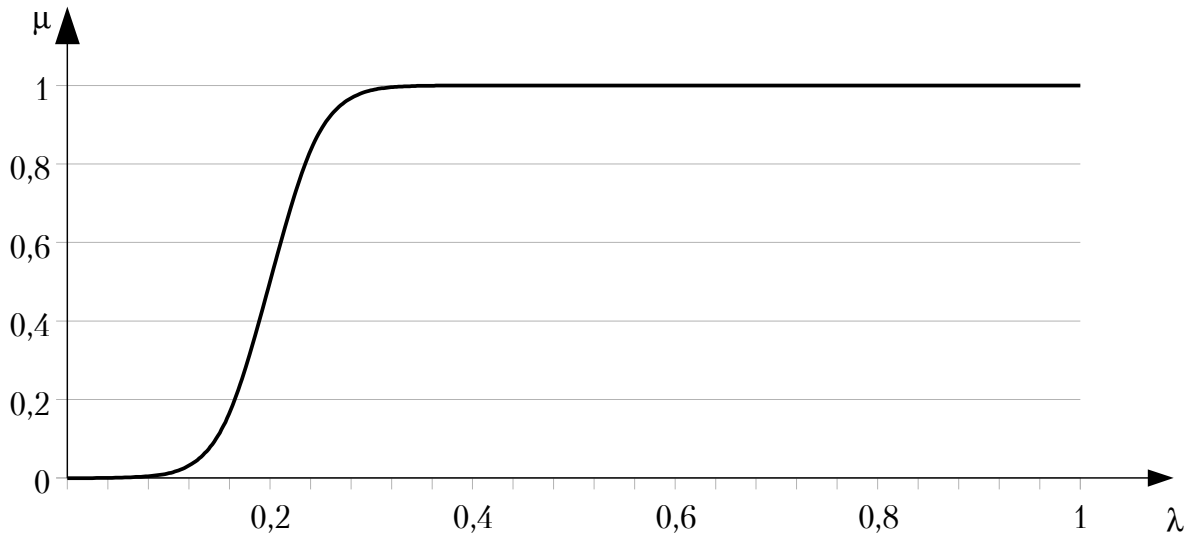


Рис. 4.9. Функція належності до нечіткої множини інтенсивності відмов за малого періоду експлуатації устаткування

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Дайте визначення лінгвістичної змінної.
2. Дайте визначення нечіткої множини.
3. Сформулюйте основні особливості нечітких множин.
4. Дайте визначення нормальних та субнормальних нечітких множин.
5. Наведіть характеристики основних операцій над нечіткими множинами.
6. Що називають відношенням між нечіткими множинами?
7. Як визначають значення функції належності для нечітких відношень?
8. Дайте визначення нечіткого логічного виведення.
9. Що називають комбінацією нечітких множин?
10. Сформулюйте принципи максимінної згортки у визначенні комбінації нечітких множин.



ЛІТЕРАТУРА

[1], [2], [3], [4], [5], [6], [7], [8], [9], [11], [13], [24], [31], [35], [44], [51], [58], [59], [61], [65], [66], [72]

ШТУЧНІ НЕЙРОННІ МЕРЕЖІ

§5.1. ЗАГАЛЬНІ ВІДОМОСТІ
ПРО ШТУЧНІ НЕЙРОННІ МЕРЕЖІ

Розглянуті раніше методи розв'язання практичних задач базуються на алгоритмах числового або символьного оброблення інформації, зокрема пошуку в просторі станів або просторі задач. Головний недолік таких підходів полягає у збільшенні часу, необхідного для розв'язання задачі зі збільшенням розміру бази знань. Давно відомо, що час, необхідний комп'ютеру для розв'язання задач, збільшується пропорційно збільшенню обсягів інформації, яку він оброблює. Навпаки, чим більшими обсягами інформації володіє людина, тим менший час йому потрібний для розв'язання практичних задач. Це пояснюється принциповою різницею організації мислення людини і комп'ютера. Традиційні комп'ютерні методи розв'язання практичних задач базуються на централізованій моделі управління, в той час, коли мозок людини реалізує модель паралельних обчислень.

В останні десятиріччя в технологіях штучного інтелекту, зокрема експертних системах, широкого поширення набув конекціоністський підхід, який базується на безпосередньому моделюванні розумової діяльності людського мозку. Такі системи штучного інтелекту отримали назву штучних нейронних мереж.

Штучні нейронні мережі – це математичні моделі, а також їх програмні або апаратні реалізації, побудовані за принципом організації та функціонування біологічних мереж нервових клітин живих організмів

Організація штучних нейронних мереж принципово відрізняється від традиційних комп'ютерних засобів оброблення інформації. Порівняння цих архітектур наведено в табл. 5.1. Аналіз даних табл. 5.1 дає змогу зробити висновок про те, що традиційні комп'ютерні системи орієнтовані на ефективне оброблення числової та символьної інформації під централізованим управлінням головного процесора за заданим алгоритмом відповідної програми. Напроти, штучні нейронні мережі реалізують паралельну багатопроцесорну обробку сукупності вхідних сигналів.

Таблиця 5.1. Порівняння архітектур традиційних комп'ютерних засобів та штучних нейронних мереж

ХАРАКТЕРИСТИКИ	ТРАДИЦІЙНІ КОМП'ЮТЕРНІ ЗАСОБИ ОБРОБЛЕННЯ ІНФОРМАЦІЇ	ШТУЧНІ НЕЙРОННІ МЕРЕЖІ
Процесор	Складний Високошвидкісний Один чи декілька	Простий Низькошвидкісний Велика кількість
Пам'ять	Локалізована Відділена від процесора Адресація за адресою	Розподілена Інтегрована в процесор Адресація за змістом
Управління	Централізоване	Розподілене
Режим функціонування	Послідовний	Паралельний
Способи розв'язання задач	Алгоритмізовані відповідно до програми	Із застосуванням самонавчання
Надійність	Висока вразливість	Живучість
Спеціалізація	Числові та символні операції	Проблеми сприйняття
Метод навчання	За правилами (дидактично)	За прикладами (сократично)
Застосування	Числова та символна обробка інформації	Розпізнавання образів, асоціативна пам'ять

Головні переваги використання технологій штучних нейронних мереж полягають у відмовостійкості пам'яті нейронної мережі. Видалення частини нейроноподібних елементів викликає лише часткове зниження якості інформації, яка зберігається у пам'яті, і катастрофічної втрати інформації не відбувається. Як наслідок, штучні нейронні мережі характеризуються пластичністю. Це означає, що після втрати деякої кількості нейроноподібних елементів можна провести повторне навчання нейронної мережі та повністю відновити її пам'ять. Крім того, розподілена архітектура штучної нейронної мережі дозволяє організувати моделі асоціативної пам'яті. Тут з'являється можливість повністю відновити початкову інформацію за її частиною після втрати або спотворення іншої частини.

Наведемо короткі характеристики типових задач, для ефективного розв'язання яких застосовують технології штучних нейронних мереж:

- *класифікація образів* – визначення приналежності вхідного образу, представленого вектором ознак, одному чи декільком заздалегідь визначеним класам;
- *кластеризація/категоризація* – визначення класів та розподіл за ними вхідних образів, тобто така задача є класифікацією без попереднього визначення класів;
- *апроксимація функцій* – визначення невідомої функції, яка поєднує наперед задані точки в деякому фазовому просторі;
- *діагностика* – виявлення несправностей і причин їх появи в контрольованій системі;
- *прогнозування* – передбачення ймовірних наслідків на основі минулих і поточних подій;
- *оптимізація* – визначення такого розв'язку деякої задачі, що задовольняє заданій системі обмежень та максимізує (або мінімізує) деяку цільову функцію;
- *реалізація асоціативної пам'яті* – адресація пам'яті за змістом, що дає змогу викликати вміст пам'яті за частковим або за спотвореним входом;
- *фільтрація* – виділення корисного сигналу із фонового зашумлення;
- *управління* – формування керуючих впливів, які визначають процеси функціонування складних технічних систем.

Незважаючи на переваги нейронних мереж в деяких галузях над традиційними обчисленнями, існуючі штучні нейронні мережі все ще є недосконалими рішеннями. Вони навчаються і можуть робити «помилки». Крім того, не можна гарантувати, що розроблена мережа є оптимальною для розв'язання певної практичної задачі.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Сформулюйте основні відмінності стратегій централізованих та паралельних обчислень.
2. Дайте визначення штучної нейронної мережі.
3. Проведіть порівняльний аналіз традиційних комп'ютерних засобів оброблення інформації та технологій штучних нейронних мереж.
4. Наведіть характеристики та приклади типових задач, для розв'язання яких застосовують технології штучних нейронних мереж.
5. Поясніть причини труднощів, пов'язаних із залученням технологій штучних нейронних мереж до розв'язання типових прикладних задач.

[1], [2], [4], [6], [7], [10], [12], [15], [39]

§5.2. МОДЕЛЬ ШТУЧНОГО НЕЙРОНУ

Розглянемо спрощену структуру біологічного нейрону, представлену на рис. 5.1. Нервова клітина (нейрон) складається з тіла (соми) та відростків двох типів – дендритів та аксону. Нервові імпульси інших нейронів передаються у клітину через дендрити, що призводить до збільшення потенціалу соми. У момент, коли потенціал досягає певного порогового значення, нейрон розряджається, передаючи нервовий імпульс через аксон на інші суміжні нейрони. Місце поєднання волокон наприкінці аксону з дендритами інших нейронів називають синапсами. Біохімічні властивості синапсів визначають розподіл потенціалу соми між дендритами. Таким чином у нервовій системі здійснюється передавання сигналів у вигляді коротких електричних імпульсів. Такі процеси визначають нервову діяльність тварин та людей, починаючи від забезпечення життєдіяльності (дихання, серцебиття) та елементарних реакцій організму на зовнішні подразники до складних процесів абстрактного мислення.

У 1943 році американські вчені У. Маккаллох та У. Піттс запропонували модель штучного нейрону, формальна структура якої представлена на рис. 5.2. Тут x_1, x_2, \dots, x_n – вхідні сигнали; w_1, w_2, \dots, w_n – вагові коефіцієнти синаптичних зв'язків; θ – порогове значення чутливості нейроноподібного елементу; Σ – суматор, який реалізує мережеву функцію штучного нейрону; Γ – нелінійний перетворювач, який генерує вихідний сигнал відповідно до закладеної функції активації.

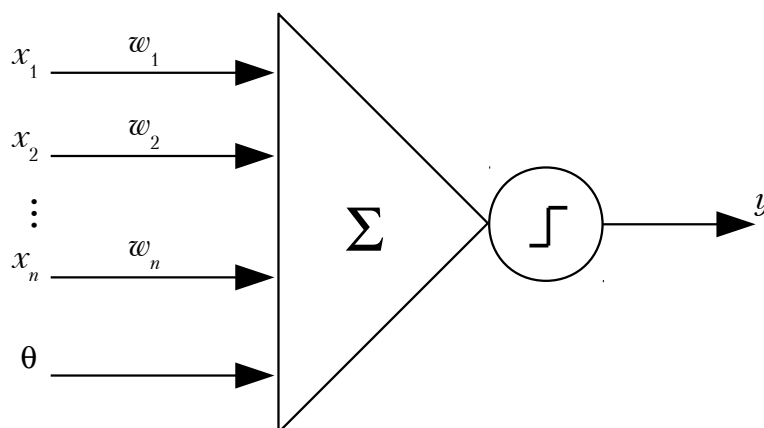


Рис. 5.1. Формальна модель штучного нейрону

Сигнали, які поступають на вхід нейроноподібного елементу, множуються на синаптичні ваги та сумуються з пороговим значенням чутливості. Далі результуючий сигнал поступає на вхід нелінійного перетворювача, де формується вихідний сигнал. Таким чином, структурними елементами штучного нейрону є блок синаптичних вагових коефіцієнтів, суматор та нелінійний перетворювач, які визначають процеси перетворення сигналів у нейронній мережі.

Математична модель найпростішого штучного нейрону має вигляд

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right),$$

де x_i та y – вхідний та вихідний сигнали відповідно; w_i – ваговий коефіцієнт відповідного синапсу; n – кількість вхідних сигналів; θ – порогове значення чутливості нейроноподібного елементу; f – функція нелінійного перетворення (функція активізації). Аргумент функції активації називають мережевою функцією.

Вхідні та вихідні сигнали нейроноподібних елементів штучних нейронних мереж зазвичай є бінарними уніполярними (+1 / 0) або біполярними (+1 / -1). Разом з тим, в сучасних нейронних мережах використовують сигнали, які можуть мати довільні дійсні значення.

Довільними є також вагові коефіцієнти синаптичних зв'язків та пороги чутливості нейроноподібних елементів, які визначають мережеву функцію штучного нейрону.

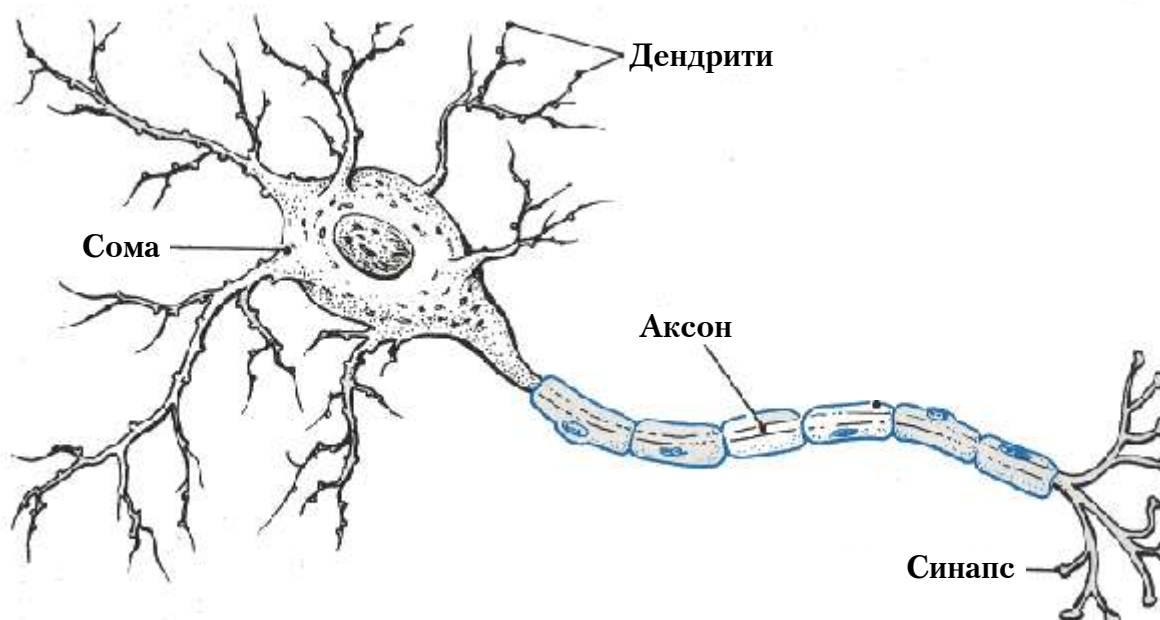


Рис. 5.2. Структура біологічного нейрону

Синаптичні зв'язки з додатними ваговими коефіцієнтами призводять до збільшення потенціалу нейронного елементу. Такі синаптичні зв'язки називають збуджувальними. Навпаки, зв'язки із від'ємними значеннями вагових коефіцієнтів знижують потенціал нейронного елементу та мають назву гальмівних.

Модель штучного нейрону, запропонована Маккалохом та Пітсом використовує лінійну модель мережевої функції

$$u = \sum_{i=1}^n w_i x_i - \theta.$$

Разом з тим сучасні нейромережеві технології передбачають інші способи формування мережевих функцій. Тут, крім лінійної, найбільшого поширення набули квадратична

$$u = \sum_{i=1}^n w_i x_i^2$$

та радіальна (сферична)

$$u = \frac{1}{R^2} \sum_{i=1}^n (x_i - w_i)^2$$

форми мережевої функції.

Довільною може також бути і функція активації нейроноподібного елементу. Така функція перетворює сигнал мережевої функції на вихідний. Найпростіша функція активізації – функція Хевісайду – має ступінчасту форму (див. рис. 5.3 а):

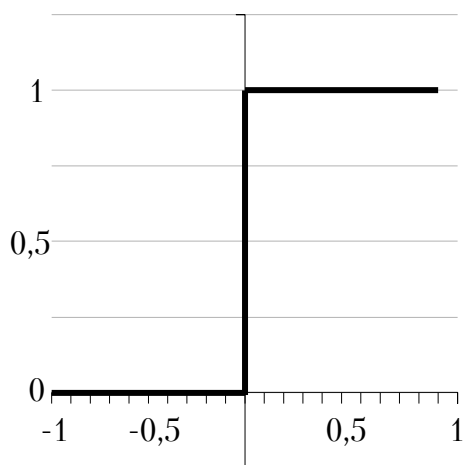
$$f(u) = \begin{cases} 1, & \text{якщо } u \geq 0; \\ 0, & \text{якщо } u < 0. \end{cases}$$

Очевидно, що у разі застосування функції Хевісайду, вихідний сигнал нейронного елементу має місце, якщо сума добутків вхідних сигналів на відповідні вагові коефіцієнти синаптичних зв'язків перевищує встановлене порогове значення.

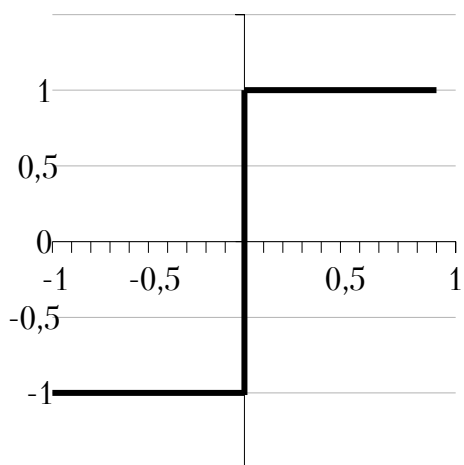
Крім функції Хевісайду в моделях штучних нейронів часто використовують такі форми функції активації:

– знакова функція (див. рис. 5.3 б)

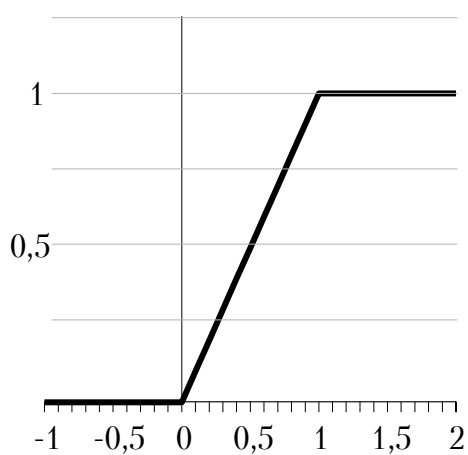
$$f(u) = \text{sgn}(u) = \begin{cases} +1, & \text{якщо } u \geq 0; \\ -1, & \text{якщо } u < 0; \end{cases}$$



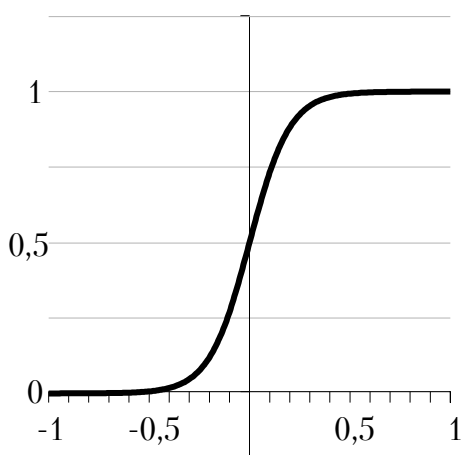
a



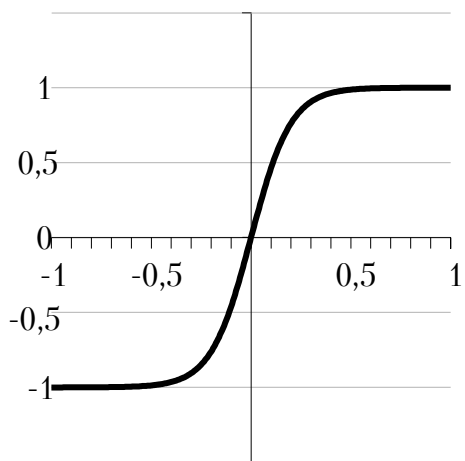
б



в



г



д

Рис. 5.3. Функції активації нейроноподібних елементів:
a) функція Хевісайду; *б)* знакова функція; *в)* лінійна функція з насиченням;
г) уніполярна сигмоїдна функція; *д)* біполярна сигмоїдна функція.

– лінійна функція з насиченням (див. рис. 5.3 в)

$$f(u) = \begin{cases} 1, & \text{якщо } u \geq 1; \\ u, & \text{якщо } 0 \leq u < 1; \\ 0, & \text{якщо } u < 0; \end{cases}$$

– уніполярна сигмоїдна функція (див. рис. 5.3 г)

$$f(u) = \frac{1}{1 + e^{-2\beta u}},$$

де β – постійний коефіцієнт;

– біполярна сигмоїдна функція (див. рис. 5.3 д)

$$f(u) = \frac{1 - e^{-2\beta u}}{1 + e^{-2\beta u}}$$

та інші форми.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Наведіть структуру та охарактеризуйте принципи функціонування біологічного нейрону.

2. Наведіть функціональну схему та математичну модель штучного нейрону.

3. Наведіть характеристики складових елементів штучного нейрону.

4. Наведіть та охарактеризуйте розповсюджені моделі мережеских функцій штучних нейроноподібних елементів.

5. Наведіть та охарактеризуйте розповсюджені моделі функцій активації штучних нейроноподібних елементів.

6. Спробуйте сформулювати моделі штучних нейронів, які реалізують логічні функції кон'юнкції, диз'юнкції та заперечення.



ЛІТЕРАТУРА

[1], [2], [4], [6], [7], [8], [9], [12], [15], [39], [66]

§5.3. ЗАГАЛЬНА ХАРАКТЕРИСТИКА АРХІТЕКТУРИ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Штучні нейронні мережі містять велику кількість елементарних процесорів (штучних нейроноподібних елементів), поєднаних між собою таким чином, щоб вихідні сигнали одних нейронів були входними для інших. Функціонування такої мережі полягає у розповсюдженні входних сигналів та перетворення їх на вихідні відповідно до математичних моделей штучних нейронів, які беруть участь у передаванні сигналів.

В загальному випадку функціонування штучної нейронної мережі полягає у реалізації математичного відображення

$$X \rightarrow Y,$$

таким чином, щоб для кожної можливої сукупності входних сигналів формувалася правильна сукупність вихідних сигналів. Тут X , Y – вектори входних та вихідних сигналів, а «правильність» вихідних сигналів визначається особливостями задачі, яку розв'язують за допомогою технологій штучних нейронних мереж.

Для відображення структури нейронних мереж, зазвичай, використовують апарат теорії графів. Тут вершинами зваженого орієнтованого графу є окремі нейроноподібні елементи, а дугами – зв'язки між ними. Кожний зв'язок між нейроноподібними елементами характеризується напрямком та ваговим синаптичним коефіцієнтом.

Відповідно до топології розрізняють чотири види нейронних мереж, представлених на рис. 5.4:

- 1) одношарові штучні нейронні мережі із прямими зв'язками – персептрони (рис. 5.4 а);
- 2) одношарові рекурентні штучні нейронні мережі із зворотними зв'язками (рис. 5.4 б);
- 3) багатошарові штучні нейронні мережі із прямими зв'язками – багатошарові персептрони (рис. 5.4 в);
- 4) багатошарові рекурентні штучні нейронні мережі із зворотними зв'язками (рис. 5.4 г).

Очевидно, що найпростішими є одношарові нейронні мережі із прямими зв'язками. Такі мережі отримали назву персептронів. По суті одношарові персептрони являють собою сукупність взаємозалежних нейроноподібних елементів із загальною системою входних сигналів та призначені для перетворення їх в систему вихідних сигналів. Кількість вихідних сигналів завжди дорівнює кількості нейроноподібних елементів у складі мережі і, зазвичай, не дорівнює кількості входних сигналів.

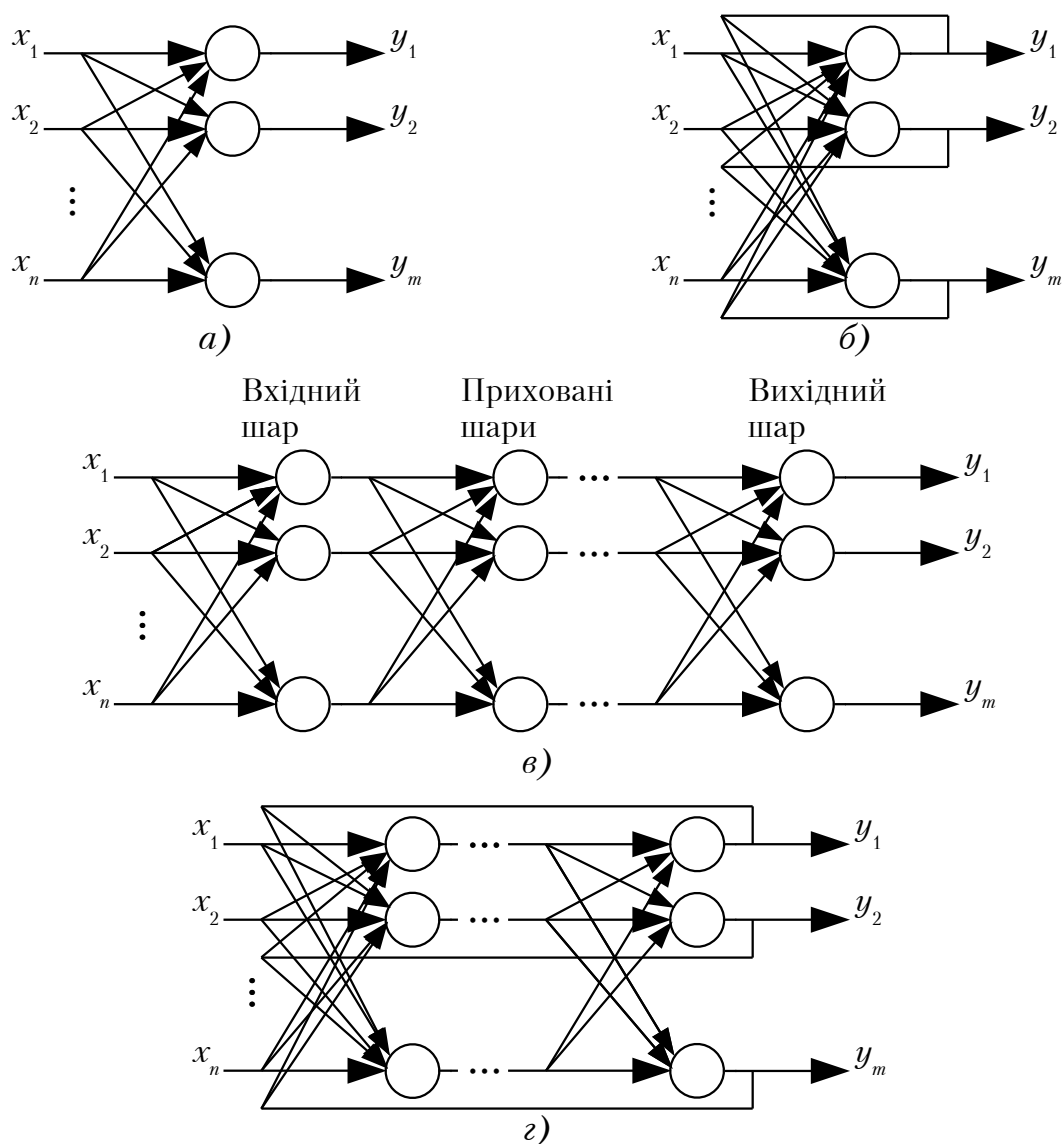


Рис. 5.4. Структури штучних нейронних мереж: а) одношарова мережа з прямими зв'язками (персептрон); б) одношарова рекурентна мережа; в) багатошарова мережа з прямими зв'язками (багатошаровий персептрон); г) багатошарова рекурентна мережа.

Персептрони дають змогу реалізувати моделі простих логічних функцій. Так, до класу персептронів можна віднести нейроноподібні елементи, які реалізують функції диз'юнкції, кон'юнкції, логічного заперечення щодо входних сигналів тощо. Як наслідок, одношарові персептрони дають змогу розв'язувати задачі розпізнавання простих образів, поділяючи простір ознак лінійними функціями.

Разом з тим, використання одношарових персептронів обмежено принциповою неможливістю організації нелінійних перетворень входних сигналів. Доведено, що одношарові нейронні мережі дають змогу розв'язати задачу розпізнавання образів лише у тому разі, коли для входних сигналів мережі можна побудувати лінійні функції розділення. Так,

наприклад, одношаровий персептрон не здатен реалізувати логічну функцію «ВИКЛЮЧНЕ АБО» щодо вхідних сигналів.

Вимога лінійного розділення вхідних сигналів суттєво послаблюється у багатошарових нейронних мережах із прямими зв'язками (багатошарових персептронах). Такі мережі складаються із вхідного шару нейроноподібних елементів, на які подаються вхідні сигнали мережі, вихідного шару нейронів, які формують систему вихідних сигналів, а також прихованих внутрішніх шарів нейроноподібних елементів (див. рис. 5.4 в). Багатошарові персептрони можуть мати декілька прихованих шарів, або не мати їх взагалі. В таких персептронах вхідні сигнали розповсюджуються по мережі від одного шару до наступного, послідовно перетворюючись в систему вихідних сигналів.

Інколи в штучних нейронних мережах відсутній жорсткий поділ нейроноподібних елементів на шари. Тут можливе передавання вихідного сигналу штучного нейрону на вхід нейрону не наступного шару, а через декілька прихованих шарів.

В задачах реалізації асоціативної пам'яті та розпізнавання образів на її основі часто використовують одно- та багатошарові нейронні мережі із зворотними зв'язками (див. рис. 5.4 б, г). Наявність зворотних зв'язків визначає певні динамічні процеси, пов'язані із циркуляцією сигналів по мережі нейроноподібних елементів. До цього класу відносять, наприклад, штучні нейронні мережі Хопфілда, розглянуті у підрозділі 5.5.

Штучні нейронні мережі часто поділяють за ступенем зв'язності нейроноподібних елементів на слабо- та повнозв'язні. Так, до слабо- зв'язних нейронних мереж можна віднести, наприклад, багатошарові персептрони, а до повнозв'язних – штучні нейронні мережі Хопфілда із зворотними зв'язками.

В сучасних технологіях штучних нейронних мереж часто використовують динамічні синаптичні зв'язки, які встановлюють або анулюють в процесі функціонування мережі.

Відповідно до положення нейронного елементу у структурі нейронної мережі розрізняють три типи нейронів:

1) вхідні нейрони, на входи яких поступають зовнішні сигнали. Такі нейрони утворюють вхідний шар нейронної мережі;

2) вихідні нейрони, сукупність вихідних сигналів яких, формує вихідні сигнали нейронної мережі. Такі нейрони утворюють вихідний шар нейронної мережі;

3) проміжні нейрони, вхідні сигнали яких формуються нейронами попередніх шарів, а вихідні поступають на входи нейронів наступних шарів. Такі нейрони утворюють приховані шари нейронної мережі.

Наведена класифікація нейроноподібних елементів є умовною. Відповідно топології нейронної мережі одні й ті самі нейрони можуть нале-

жати до різних груп. Так, в одношарових нейронних мережах всі нейрони є одночасно вхідними та вихідними; у багатошарових рекурентних мережах нейрони вихідного шару одночасно є вихідними та проміжними; а у повнозв'язних мережах нейрони можуть одночасно належати до всіх трьох типів.

Вибір топології штучної нейронної мережі для розв'язання практичних задач завжди пов'язаний з двома взаємоконфліктуючими факторами. З одного боку, збільшення кількості нейронних елементів у складі мережі, введення додаткових прихованих шарів, застосування зворотних зв'язків та збільшення зв'язності мережі призводить до збільшення можливостей нейронної мережі та її потужності. З іншого боку, застосування зворотних зв'язків та посилення зв'язності часто ускладнює налагодження штучної нейронної мережі та погіршує умови динамічної стійкості.

Проблема визначення необхідних та достатніх властивостей нейронної мережі є дуже складною. Структура та розмірність мережі зазвичай визначається тією задачею, яку мережа буде розв'язувати. Це унеможливорює надання типових рекомендацій щодо синтезу штучних нейронних мереж. Разом з тим, багатий досвід використання нейромережевих технологій дозволяє обрати типові структури штучних нейронних мереж для розв'язання деяких типових задач.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає функціонування штучної нейронної мережі?
2. Наведіть класифікацію штучних нейронних мереж за топологічними ознаками.
3. Наведіть класифікацію нейроноподібних елементів за їх положенням у структурі мережі. Поясніть умовність такої класифікації.
4. Охарактеризуйте проблеми, пов'язані з синтезом структури штучної нейронної мережі.



ЛІТЕРАТУРА

[1], [2], [6], [8], [9], [12], [15], [66]

§5.4. МЕТОДИ НАВЧАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Ефективне розв'язання практичних задач засобами штучних нейронних мереж визначається, у першу чергу, обсягами знань, якими володіє нейроноподібна мережа. Це означає, що синтез штучних нейронних мереж складається з двох етапів. На першому етапі визначають топологію нейронної мережі, а на другому – проводять її навчання.

Під **навчанням штучної нейронної мережі** розуміють процес визначення вагових коефіцієнтів всіх синаптичних зв'язків між нейроноподібними елементами, що призводить до зміни пам'яті системи

Навчання є найбільш складною задачею у процесі створення штучної нейронної мережі. Очевидно, що складність такої задачі визначається розмірністю та зв'язністю мережі – чим більша кількість синаптичних зв'язків нейронної мережі, тим складнішою є задача її навчання.

В загальному випадку процес навчання штучної нейронної мережі є ітераційним та полягає в цілеспрямованій зміні вагових коефіцієнтів синаптичних зв'язків. Тут розрізняють три способи навчання:

- 1) навчання з вчителем;
- 2) навчання з підкріпленням;
- 3) навчання без вчителя.

Під час навчання з вчителем для кожної сукупності вхідних сигналів наперед відомі відповідні вектори вихідних сигналів. Це дає змогу сформувати різницеві вектори між дійсними та бажаними вихідними сигналами для кожних наборів вхідних сигналів за поточних значень вагових коефіцієнтів синаптичних зв'язків. Такий різницевий вектор визначає поправки до вагових коефіцієнтів та дозволяє скоригувати характеристики синаптичних зв'язків.

Процес навчання штучної нейронної мережі із вчителем проілюстровано на рис. 5.5. Тут $\vec{x}(\mu)$ – вектор сигналів навчальної вибірки μ , які подають на вхід штучної нейронної мережі; $\vec{y}(\mu)$ – вектор вихідних сигналів за поточних значень вагових коефіцієнтів синаптичних зв'язків; $\vec{t}(\mu)$ – вектор бажаних вихідних сигналів; $\Delta \vec{w}$ – вектор поправок до вагових коефіцієнтів синаптичних зв'язків.

Спосіб навчання штучної нейронної мережі з вчителем часто називають методом зворотного розповсюдження помилки. Тут загальну множину векторів вхідних сигналів поділяють на дві частини: навчальну та тестову вибірки. Навчальну вибірку використовують для багатокрокової корекції синаптичних зв'язків нейронної мережі. Для контролю правильності налагодження нейронної мережі використовують

тестову вибірку вхідних сигналів. Тут необхідно, щоб множини навчальних та тестових вибірок не перетиналися.

Алгоритм навчання штучної нейронної мережі за методом зворотнього розповсюдження помилки полягає у послідовному виконанні таких процедур.

1. Загальну множину векторів вхідних сигналів поділяють на навчальну та тестові вибірки. Ваговим коефіцієнтам синаптичних зв'язків привласнюють випадкові значення.

2. Організують цикл налагодження штучної нейронної мережі за множиною навчальної вибірки. Тут обирають черговий вектор вхідних сигналів та подають його на вхід мережі.

3. Для поточного вектору вхідних сигналів навчальної вибірки та за поточних значень вагів синаптичних зв'язків визначають вектор вихідних сигналів штучної нейронної мережі.

4. За виразом

$$\Delta \vec{w} = \eta (\vec{t}(\mu) - \vec{y}(\mu)) \vec{x}(\mu),$$

де η – коефіцієнт навчання, визначають вектор поправок до вагових коефіцієнтів.

5. Якщо норма (максимальний за модулем елемент) вектора поправок вагових коефіцієнтів перевищує задану інженерну точність

$$\max_i \{|\Delta w_i|\} > \varepsilon,$$

то визначені у процедурі 4 поправки вносять до складу вектору вагових коефіцієнтів та передають управління процедурі 3 алгоритму.

В іншому випадку продовжують налаштування штучної нейронної мережі за наступним вектором вхідних сигналів навчальної вибірки починаючи з процедури 2. Якщо ж було проведено навчання із всіма векторами вхідних сигналів навчальної вибірки, управління передають процедурі 6 для перевірки за сигналами тестової вибірки.

6. Організують цикл перевірки налаштування штучної нейронної мережі за множиною тестової вибірки. Тут обирають черговий вектор вхідних сигналів та подають його на вхід мережі.

7. Для поточного вектору вхідних сигналів тестової вибірки визначають вектор вихідних сигналів штучної нейронної мережі.

8. Перевіряють правильність налагодження нейронної мережі. Якщо виконується умова

$$|\vec{t}(\mu') - \vec{y}(\mu')| \rightarrow 0,$$

де μ' – індекс поточного вектору вхідних сигналів тестової вибірки, то обирають наступний вектор вхідних сигналів та передають управління процедурі 7. В іншому випадку слід скоригувати склад множин на-

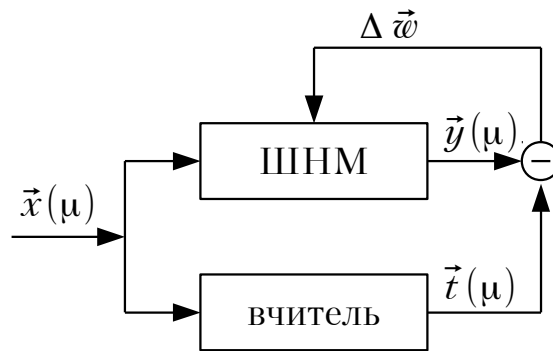


Рис. 5.5. Ілюстрація навчання нейронної мережі із вчителем

вчальної та тестової вибірок та повторити налагодження мережі, починаючи з процедури 2 алгоритму.

Процедури 7 та 8 повторюють циклічно для всіх елементів множини тестової вибірки.

Проілюструємо застосування методу навчання штучної нейронної мережі з вчителем на прикладі встановлення вагових коефіцієнтів синаптических зв'язків одноелементного персептрону, призначеного для визначення отримання електропостачання споживачами лінії з двостороннім живленням (див. рис. 5.6). Тут ДЖ1, ДЖ2 – джерела живлення 1 та 2; Л1, Л2 – лінії електропередач 1 та 2; Н – навантаження.

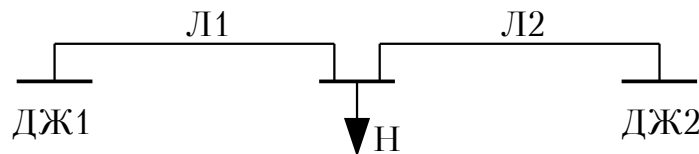


Рис. 5.6. Електропостачання із застосуванням лінії електропередач з двостороннім живленням

Для розв'язання логічної задачі визначення отримання живлення можна використати одноелементний персептрон з двома входними каналами, представлений на рис. 5.7. Такий персептрон має два входні бінарні уніполярні канали, які характеризують подачу живлення по відповідним лініям електропередач. Очевидно, що вихідний сигнал матиме місце у разі наявності сигналів на будь-якому із входів, або на обох входах одночасно, тобто персептрон має реалізовувати диз'юнктивну логічну функцію щодо входних сигналів.

Математична модель персептрону має вигляд

$$y = \begin{cases} 1, & \text{якщо } w_1 x_1 + w_2 x_2 - \theta \geq 0; \\ 0, & \text{якщо } w_1 x_1 + w_2 x_2 - \theta < 0. \end{cases}$$

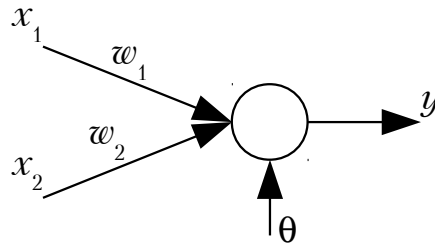


Рис. 5.7. Одноелементний персептрон з двома вхідними каналами

Нехай порогове значення функції активації нейроподібного елементу дорівнює одиниці, а початкові наближення синаптичних вагів вхідних сигналів – $+0,5$ та $-0,1$ відповідно. Коефіцієнт навчання приймемо дорівнюючим $0,4$:

$$\begin{aligned}w_1 &= 0,5; \\w_2 &= -0,1; \\ \theta &= 1; \\ \eta &= 0,4.\end{aligned}$$

Можливі групи вхідних сигналів поділимо на навчальну та тестову вибірки. До навчальної вибірки віднесемо групи сигналів $\{1; 0\}$ та $\{0; 1\}$, які відповідають подаванню живлення по одній з ліній електропередачі. Тестову вибірку, відповідно, складають групи сигналів $\{1; 1\}$ та $\{0; 0\}$, що відповідає одночасному подаванню, або не подаванню живлення по лініям.

Проведемо цикл навчання з першою групою вхідних сигналів. Відповідно до математичної моделі персептрону для поточних значень вагових коефіцієнтів синаптичних зв'язків маємо:

$$0,5 \cdot 1 - 0,1 \cdot 0 - 1 = -0,5 \rightarrow y = 0.$$

Отриманий результат свідчить про необхідність коригування синаптичних вагів. Тут проправку визначають лише для синаптичного зв'язку першого входу, на який подається сигнал. Дійсно:

$$\Delta \vec{w} = \eta (\vec{t}(\mu) - \vec{y}(\mu)) \vec{x}(\mu) = 0,4(1 - 0) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,4 \\ 0 \end{bmatrix},$$

тобто

$$\begin{aligned}\Delta w_1 &= 0,4; & \Delta w_2 &= 0; \\ w_1 &= 0,5 + 0,4 = 0,9; & w_2 &= -0,1.\end{aligned}$$

Повторюємо розрахунковий експеримент для нового значення вагового коефіцієнту:

$$0,9 \cdot 1 - 0,1 \cdot 0 - 1 = -0,1 \rightarrow y=0;$$

$$\Delta w_1 = 0,4(1-0)1 = 0,4;$$

$$w_1 = 0,9 + 0,4 = 1,3.$$

Тепер одноелементний персептрон буде спрацьовувати:

$$1,3 \cdot 1 - 0,1 \cdot 0 - 1 = 0,3 \rightarrow y=1.$$

За аналогічною схемою проводимо цикл навчання за другою навчальною вибіркою. Для поточних значень синаптичних вагів маємо:

$$1,3 \cdot 0 - 0,1 \cdot 1 - 1 = -1,3 \rightarrow y=0;$$

$$\Delta w_2 = 0,4(1-0)1 = 0,4;$$

$$w_2 = -0,1 + 0,4 = 0,3.$$

Далі проводимо навчання за наведеною схемою. Після виконання третього циклу корекції синаптичних зв'язків за другою навчальною вибіркою маємо $w_2 = 1,1$ і

$$1,3 \cdot 0 + 1,1 \cdot 1 - 1 = 0,1 \rightarrow y=1,$$

що свідчить про вірне спрацьовування штучного нейрону.

Перевіряємо правильність налаштування синаптичних зв'язків персептрону на тестових вибірках.

У разі відсутності входних сигналів маємо

$$1,3 \cdot 0 + 1,1 \cdot 0 - 1 = -1 \rightarrow y=0,$$

а подавання сигналів на обидва входи дає такий результат:

$$1,3 \cdot 1 + 1,1 \cdot 1 - 1 = 1,4 \rightarrow y=1.$$

Таким чином після навчання вагові коефіцієнти синаптичних зв'язків штучного нейрону набувають значень 1,3 та 1,1 відповідно.

Зауважимо, що отриманий розв'язок задачі навчання персептрону не є єдиним можливим. Очевидно, що нейроноподібний елемент буде реалізовувати функцію диз'юнкції за будь-яких, але не нижчих за одиницю, значень вагових коефіцієнтів синаптичних зв'язків.

Загалом, швидкість навчання методом зворотного розповсюдження помилки визначається вдалим вибором коефіцієнту навчання. На практиці обирають коефіцієнт навчання в діапазоні $10^{-3} \dots 10$. Очевидно, що збільшення коефіцієнту призводить до прискорення процесу навчання. Разом з тим, для складних нейронних мереж таке збільшення може призводити до порушення стійкості, тобто до відсутності збіжності навчального ітераційного процесу. З метою запобігання такого явища часто використовують динамічну зміну коефіцієнту навчання, зменшуючи його зі збільшенням кількості циклів коригування синаптичних вагів.

Навчання із підкріпленням має місце у тому разі, коли обсяги інформації про бажану реакцію штучної нейронної мережі обмежені. Ця ситуація, зазвичай, пов'язана із відсутністю точних характеристик вихідних сигналів для векторів навчальних вибірок, а правильність роботи нейронної мережі можна оцінити лише якісно, наприклад, за допомогою оцінок «вірно»/«невірно».

Навчання штучної нейронної мережі з підкріпленням проілюстровано на рис. 5.8. Тут сигнал підкріплення відповідає якійсь оцінці результату перетворення вхідної вибірки за поточних значень вагових коефіцієнтів.

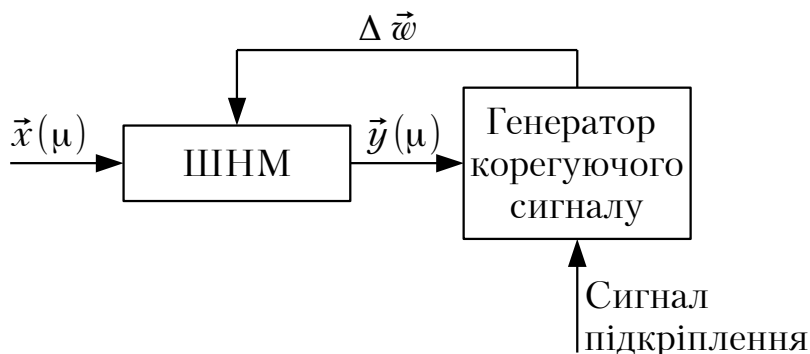


Рис. 5.8. Ілюстрація навчання нейронної мережі із підкріпленням

Метод навчання із підкріпленням базується на системі заохочень та штрафів для правильних і неправильних реакцій штучної нейронної мережі. Системою підкріплення нейронної мережі називають сукупність правил, за допомогою яких можна змінювати вагові коефіцієнти синаптичних зв'язків, тобто змінювати стан пам'яті мережі. Загалом існує велика кількість різних систем підкріплення. Тут найбільш поширеною є система підкріплення з корекцією помилок.

Відповідно до цього методу вагові коефіцієнти синаптичних зв'язків не змінюються доки поточна реакція нейронної мережі залишається вірною. У разі виявлення неправильної реакції вагові коефіцієнти синаптичних зв'язків нейронів, які були активовані в процесі оброблення сигналу, змінюються на певну дискретну величину у напрямку, протилежному до знаку помилки:

$$\Delta \vec{w} = -h \operatorname{sgn}(\vec{y}(\mu)) \vec{x}(\mu),$$

де h – навчальний крок зміни вагових коефіцієнтів; $\operatorname{sgn}(\vec{y}(\mu))$ – сигнальна функція, яка визначає знак помилки за поточних значень вагових коефіцієнтів.

Проілюструємо застосування методу навчання із підкріпленням на прикладі переналаштування одноелементного персептрону з виконання функції диз'юнкції на кон'юнктивну функцію щодо вхідних сигналів.

Нехай тепер, одноелементний перспетрон, представлений на рис. 5.7 буде налаштований на визначення режиму короткого замкнення в лінії електропередач (див. рис. 5.9). Тут ДЖ – джерело живлення; Н – навантаження; ЛЕП – лінія електропередавання; к.з. – коротке замкнення. Відомо, що режим короткого замкнення супроводжується різким збільшенням фазного струму в лінії та глибокою посадкою напруги. Призначимо ці ознаки режиму короткого замкнення відповідним входам перспетрону. Очевидно, що нейроноподібний елемент має спрацьовувати лише у тому разі, коли мають місце обидві ознаки одночасно, тобто перспетрон має реалізовувати кон'юнктивну функцію щодо входних сигналів.

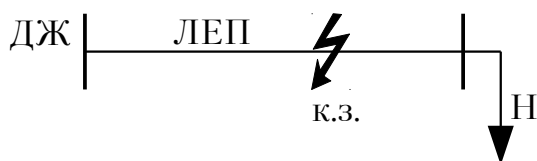


Рис. 5.9. Ілюстрація режиму короткого замкнення

Відповідно до результатів навчання виконанню диз'юнктивної функції синаптичні ваги входних сигналів перспетрону дорівнюють 1,3 та 1,1 відповідно (див. приклад навчання із вчителем). Задача полягає в переналаштуванні вагових коефіцієнтів таким чином, щоб нейроноподібний елемент спрацьовував лише у разі подавання сигналів на обидва входи одночасно. Нехай крок навчання дорівнює 0,3.

Для поточної задачі навчальні та тестові вибірки відповідають наявності або відсутності відповідних ознак режиму короткого замкнення.

Проведемо цикл навчання за вибіркою входних сигналів $\{1; 0\}$, що відповідає лише збільшенню струму в лінії, яке не супроводжується глибокою посадкою напруги. Очевидно, що нейроноподібний елемент у цьому разі не має спрацьовувати.

Для поточних значень синаптичних вагів маємо хибне спрацьовування штучного нейрону. Дійсно

$$1,3 \cdot 1 + 1,1 \cdot 0 - 1 = 0,3 \rightarrow y = 1.$$

Такий результат свідчить про необхідність послаблення синаптичного зв'язку першого входного сигналу:

$$w_1 = 1,3 - 0,3 = 1.$$

Нове значення вагового коефіцієнту ще не призводить до бажаного результату:

$$1 \cdot 1 + 1,1 \cdot 0 - 1 = 0 \rightarrow y = 1.$$

Після виконання другого кроку навчання маємо вже правильну роботу штучного нейрону:

$$\begin{aligned}w_1 &= 1 - 0,3 = 0,7; \\ 0,7 \cdot 1 + 1,1 \cdot 0 - 1 &= -0,3 \rightarrow y = 0.\end{aligned}$$

Далі проводимо цикл навчання за другою навчальною вибіркою $\{0; 1\}$, яка відповідає лише глибокій посадці напруги, яка не супроводжується різким збільшенням фазного струму. У цьому разі нейроноподібний елемент також не має спрацьовувати.

Поточні значення вагових коефіцієнтів визначають необхідність послаблення синаптичного зв'язку другого вхідного сигналу:

$$\begin{aligned}0,7 \cdot 0 + 1,1 \cdot 1 - 1 &= 0,1 \rightarrow y = 1; \\ w_2 &= 1,1 - 0,3 = 0,8.\end{aligned}$$

Отриманий результат дає правильну роботу штучного нейрону:

$$0,7 \cdot 0 + 0,8 \cdot 1 - 1 = -0,2 \rightarrow y = 0.$$

Перевіримо правильність налаштування персептрону на тестових вибірках, які відповідають відсутності та наявності обох ознак режиму короткого замкнення одночасно

$$\begin{aligned}\{0; 0\}: \\ 0,7 \cdot 0 + 0,8 \cdot 0 - 1 &= -1 \rightarrow y = 0; \\ \{1; 1\}: \\ 0,7 \cdot 1 + 0,8 \cdot 1 - 1 &= 0,5 \rightarrow y = 1.\end{aligned}$$

Очевидно, що отриманий результат навчання не є єдиним можливим. Персептрон буде давати вірне спрацьовування за будь-яких додатних значень вагових коефіцієнтів, менших за одиницю, але сума яких не менша одиниці.

Швидкість навчання штучної нейронної мережі із підкріпленням визначається величиною кроку навчання. З метою збільшення швидкості необхідно збільшувати величину кроку, проте така операція небезпечна втратою стійкості навчального процесу. Загалом міркування щодо вибору оптимального кроку навчання в методі із підкріпленням повністю аналогічні наведеним вище міркуванням щодо вибору коефіцієнту навчання в методі із вчителем.

Методи навчання без вчителя взагалі не передбачають наявності інформації про бажану реакцію штучної нейронної мережі на вектори вхідних сигналів навчальної вибірки. Така ситуація унеможливорює визначення помилки функціонування мережі з метою корекції її пам'яті. Тут нейронна мережа навчається самостійно, виявляючи та вивчаючи закономірності простору вхідних сигналів.

Процес навчання штучної нейронної мережі без вчителя проілюстровано на діаграмі рис. 5.10.

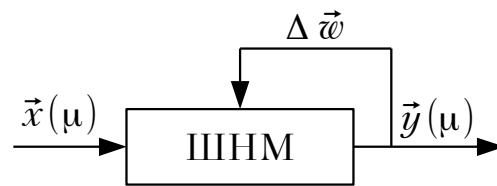


Рис. 5.10. Ілюстрація навчання нейронної мережі без вчителя

В основу навчання штучної нейронної мережі без вчителя покладено гіпотезу Хебба щодо біологічних нейронних систем про те, що вагові коефіцієнти синаптичних зв'язків змінюються пропорційно кореляції між активністю нейронів, пов'язаних цими зв'язками. Це означає, що якщо вихідний імпульс нейрону *A* збуджує нейрон *B*, і це постійно повторюється, то синаптичний зв'язок між цими нейронами посилюється.

Алгоритм навчання штучної нейронної мережі за правилом Хебба полягає у послідовному виконанні таких процедур.

1. Проводять початкову ініціалізацію нейронної мережі шляхом присвоєння випадкових невеликих значень вагових коефіцієнтів всіх синаптичних зв'язків мережі.

2. На вхід мережі подають черговий образ, сигнали якого розповсюджуються по всім шарам нейроноподібних елементів, формуючи систему вихідних сигналів.

3. За виразом

$$\Delta w_{ij} = \eta y_i x_j,$$

де η – коефіцієнт навчання; y_i – вихідний сигнал *i*-го нейрону; x_j – вхідний сигнал *j*-го нейрону, який формується *i*-м нейроном, здійснюють корекцію синаптичних вагів нейронної мережі.

Наведений вираз являє собою математичну модель гіпотези Хебба.

Процедури 2-3 проводять циклічно для всіх образів навчальної вибірки доки вихідні сигнали мережі не стабілізуються із заданою точністю для кожного вхідного образу.

Очевидно, що зміна вагового коефіцієнту синаптичного зв'язку за правилом Хебба не потребує залучення інформації про бажану реакцію нейрону, що дає змогу організувати ефективне самонавчання нейронної мережі.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. У чому полягає процес навчання штучних нейронних мереж?

2. Наведіть класифікацію методів навчання штучних нейронних мереж.

3. Дайте характеристику та наведіть алгоритм навчання штучних нейронних мереж із вчителем.

4. Дайте характеристику та наведіть алгоритм навчання штучних нейронних мереж із підкріпленням.

5. Дайте характеристику та наведіть алгоритм навчання штучних нейронних мереж без вчителя.



ЛІТЕРАТУРА

[1], [6], [7], [8], [9], [12], [15], [39], [66]

§5.5. ШТУЧНА НЕЙРОННА МЕРЕЖА ХОПФІЛДА

Мережа Хопфілда являє собою одношарову мережу із симетричними зворотними зв'язками, призначену для реалізації асоціативної пам'яті. Структуру мережі Хопфілда у загальному вигляді представлено на рис. 5.11.

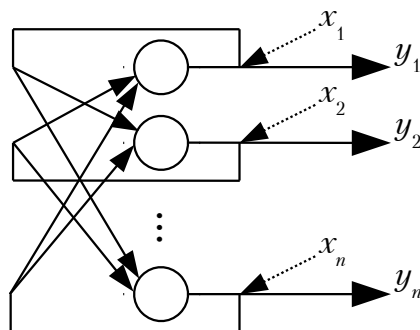


Рис. 5.11. Архітектура мережі Хопфілда

Ініціалізація мережі відбувається шляхом подання вхідних сигналів на вихід мережі. Очевидно, що кількість вхідних сигналів мережі Хопфілда завжди дорівнює кількості вихідних сигналів та кількості нейроноподібних елементів у складі мережі. Далі, сигнали з виходу мережі через зворотні зв'язки подаються на її вхід та перетворюються на нові вихідні сигнали, які знов подаються на вхід мережі. Така операція проводиться циклічно, доки сигнали, які циркулюють у мережі не зафіксуються, а мережа не перейде до одного з можливих усталених станів.

Запам'ятовування інформації (навчання) нейронних мереж Хопфілда здійснюється за безітераційним алгоритмом, тобто відбувається за один прохід шляхом встановлення синаптичних вагів зворотніх зв'язків між

виходами та входами нейроноподібних елементів. Доведено, що мережа Хопфілда буде стійкою, тобто завжди буде збігатися до одного з усталених станів, у разі відсутності зворотних зв'язків між власними виходами та входами окремих нейронів та за симетрії зворотніх зв'язків між різними нейронами:

$$w_{ii}=0; \quad w_{ij}=w_{ji}.$$

В мережах Хопфілда, зазвичай, використовують уніполярні або біполярні функції активації нейроноподібних елементів. У загальному випадку математична модель нейронної мережі Хопфілда має вигляд

$$y_i(t)=f\left(\sum_{j=1}^n w_{ij} y_j(t-1)\right),$$

де t – індекс поточного циклу проходження сигналів по мережі; $f(u)$ – функція активації:

$$f(u)=\begin{cases} +1, & \text{якщо } u \geq 0; \\ 0, & \text{якщо } u < 0 \text{ та функція активації є уніполярною;} \\ -1, & \text{якщо } u < 0 \text{ та функція активації є біполярною.} \end{cases}$$

Налаштування мережі Хопфілда здійснюють наступним чином. У разі використання біполярної функції активації, вагові коефіцієнти синаптичних зв'язків визначають за виразами:

$$w_{ij}=\begin{cases} \eta \sum_{\mu=1}^P x_i(\mu)x_j(\mu), & \text{якщо } i \neq j; \\ 0, & \text{якщо } i = j, \end{cases}$$

де P – загальна кількість образів, які запам'ятовує мережа Хопфілда.

Пороги спрацьовування нейроноподібних елементів такої мережі дорівнюють нулю.

У разі використання уніполярної функції активації вагові коефіцієнти синаптичних зв'язків та пороги спрацьовування нейронів визначають за виразами:

$$w_{ij}=\begin{cases} \eta \sum_{\mu=1}^P (2x_i(\mu)-1)(2x_j(\mu)-1), & \text{якщо } i \neq j; \\ 0, & \text{якщо } i = j; \end{cases}$$

$$\theta_i=\frac{1}{2} \sum_{j=1}^N w_{ij},$$

де N – загальна кількість нейроноподібних елементів у складі мережі Хопфілда.

У наведених виразах коефіцієнт навчання являє собою довільне додатне число. Такий параметр використовують, зазвичай, з метою обмеження, наприклад, нормалізації, вагових коефіцієнтів синаптичних зв'язків штучної нейронної мережі.

Експериментально доведено, що стійке розв'язання задач із використанням мереж Хопфілда гарантується у разі, коли загальна кількість образів, які зберігаються у пам'яті мережі задовольняє умові

$$P \leq \frac{N}{2 \ln N}.$$

Мережа Хопфілда дає змогу ефективно реалізовувати асоціативну пам'ять, запам'ятовуючи певні образи предметної області, які відповідають станам, до яких буде стійко збігатися мережа у процесі свого функціонування. Таким чином, мережа Хопфілда дає змогу знайти попередньо запам'ятований образ, найбільш схожий на той, що подається на вхід мережі.

Проілюструємо використання апарату мереж Хопфілда для розв'язання задач розпізнання образів на основі асоціативної пам'яті.

Відомо, що розрахунок режимних параметрів електричних систем є дуже чутливим до вибору початкових наближень невідомих. У разі невдалого вибору початкових наближень ітераційний процес може виявитися розбіжним навіть у разі існування розв'язку режимної задачі. У більшості практичних випадків для отримання стійкого збіжного ітераційного процесу достатньо обрати номінальні значення невідомих як початкові наближення, наприклад, номінальні напруги незалежних вузлів розрахункової схеми. Проте, у разі розв'язання задач аналізу переобтяжених режимів, зображуючі точки яких знаходяться поблизу межі існування, задача вибору початкових наближень набуває особливої актуальності. Для отримання стійких ітераційних процесів необхідно обрати початкові наближення невідомих параметрів якомога ближче до їх реальних значень. Тут, наприклад, використовують дані минулих розрахунків режимних параметрів електричної системи, певною мірою схожі на поточну задачу. Для цього організовують банки архівованих зрізів режимів електричних систем, дані яких використовують для розв'язання поточних задач. Тут перед початком кожного циклу розрахунку здійснюють пошук зрізу, найближчого до поточного розрахункового режиму. Такий підхід дає змогу суттєво підвищити надійність розрахункових моделей в задачах аналізу переобтяжених режимів електричних систем.

Задачу пошуку режимного зрізу, максимально наближеного до поточної розрахункової схеми можна розглядати як задачу класифікації із використанням моделей асоціативної пам'яті. Розглянемо розв'язання такої задачі із використанням мереж Хопфілда для тестової схеми

електричної системи IEEE-30, яка містить 30 вузлів, представлену на рис. 5.12.

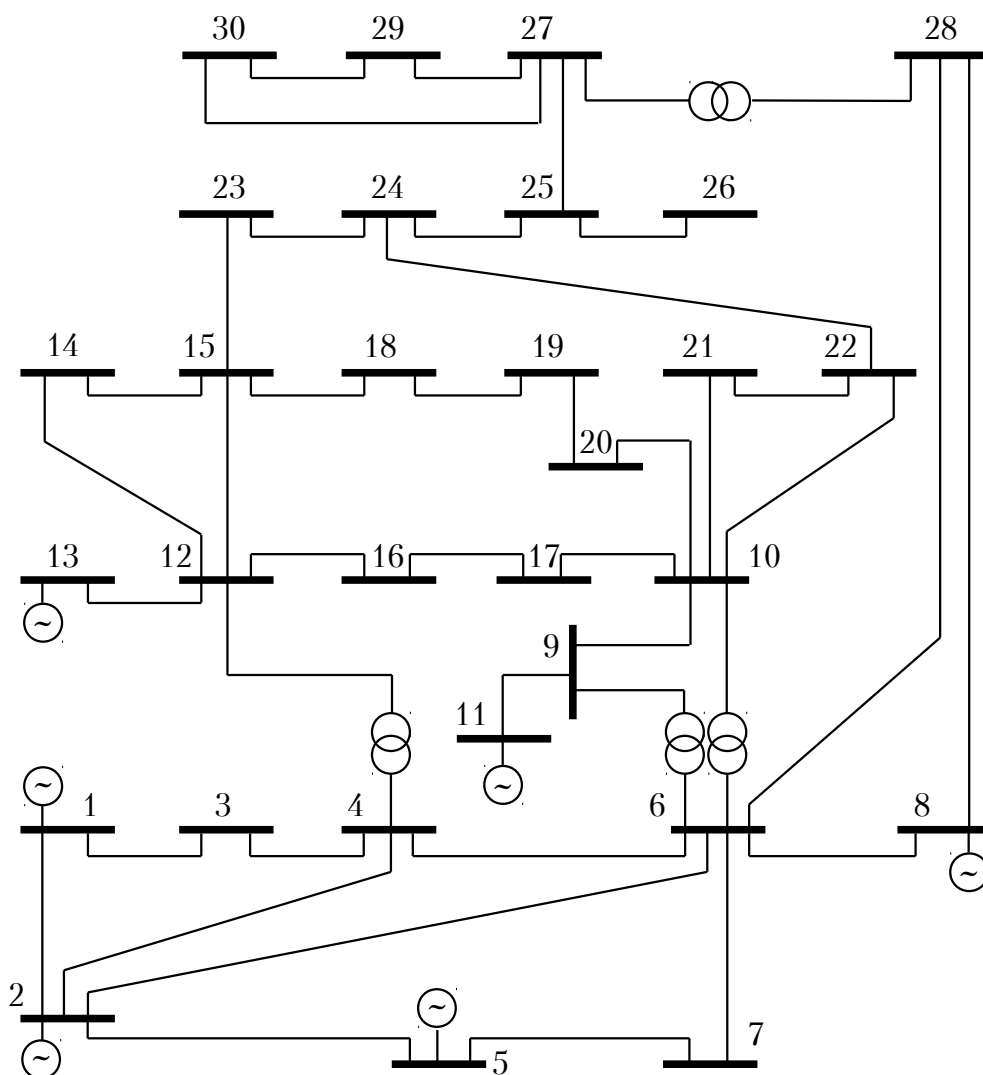


Рис. 5.12. Тестова схема IEEE-30

Переобтяження режиму здійснюється шляхом збільшення навантаження у вузлах розрахункової схеми. Це означає, що мережа Хопфілда, яка буде зберігати інформацію про архівовані режими має містити 30 нейроноподібних елементів. Така мережа зможе зберегти інформацію про чотири образи зрізів переобтяжених режимів. Дійсно

$$P \leq \frac{N}{2 \ln N} = \frac{30}{2 \ln 30} = 4,41 \approx 4.$$

Нехай банк архівованих зрізів містить інформацію про чотири переобтяжені режими, отримані шляхом збільшення навантаження у таких вузлах схеми:

- 1) 3, 4, 6, 8, 9, 10, 12, 14, 15, 16;
- 2) 4, 6, 7, 9, 10, 20, 21, 22, 28;

3) 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22;

4) 21, 22, 23, 24, 25, 26, 27, 28, 29, 30.

Такі режими умовно можна представити у вигляді діаграм образів наведених на рис. 5.13. Тут затемнені комірки відповідають вузлам із збільшеними навантаженнями. Рахунок комірок здійснюють зліва направо, починаючи із верхнього лівого кута.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

Рис. 5.13. Діаграми образів архівованих зрізів переобтяжених режимів

Проведемо навчання штучної нейронної мережі із застосуванням біполярної функції активації нейроноподібних елементів. З метою нормалізації синаптичних вагів коефіцієнт навчання приймемо таким, що дорівнює 0,25 (виходячи з чотирьох образів, які будуть зберігатися у пам'яті мережі).

Вага синаптичного зв'язку між першим та другим нейронами складає

$$w_{1,2}=w_{2,1}=0,25((-1)(-1)+(-1)(-1)+(-1)(-1)+(-1)(-1))=1.$$

Для синаптичного зв'язку між першим та другим нейронами маємо

$$w_{1,3}=w_{3,1}=0,25(1(-1)+(-1)(-1)+(-1)(-1)+(-1)(-1))=0,5.$$

В такий самий спосіб визначаємо синаптичні ваги всіх зв'язків між нейроноподібними елементами мережі Хопфілда. В результаті отримуємо матрицю синаптичних вагів, яка має розмірність 30×30. Таку матрицю можна представити у вигляді таблиці 5.2.

Таблиця 5.2. Матриця синаптичних вагів мережі Хопфілда

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	1	0,5	0	1	0	0,5	0,5	-0,5	-0,5	1	0	1	0	0
2	1	0	0,5	0	1	0	0,5	0,5	-0,5	-0,5	1	0	1	0	0
3	0,5	0,5	0	0,5	0,5	0,5	0	1	0	0	0,5	0,5	0,5	0,5	0,5
4	0	0	0,5	0	0	1	0,5	0,5	0,5	0,5	0	0	0	0	0
5	1	1	0,5	0	0	0	0,5	0,5	-0,5	-0,5	1	0	1	0	0
6	0	0	0,5	1	0	0	0,5	0,5	0,5	0,5	0	0	0	0	0
7	0,5	0,5	0	0,5	0,5	0,5	0	0	0	0	0,5	-0,5	0,5	-0,5	-0,5
8	0,5	0,5	1	0,5	0,5	0,5	0	0	0	0	0,5	0,5	0,5	0,5	0,5
9	-0,5	-0,5	0	0,5	-0,5	0,5	0	0	0	1	-0,5	0,5	-0,5	0,5	0,5
10	-0,5	-0,5	0	0,5	-0,5	0,5	0	0	1	0	-0,5	0,5	-0,5	0,5	0,5

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	1	1	0,5	0	1	0	0,5	0,5	-0,5	-0,5	0	0	1	0	0
12	0	0	0,5	0	0	0	-0,5	0,5	0,5	0,5	0	0	0	1	1
13	1	1	0,5	0	1	0	0,5	0,5	-0,5	-0,5	1	0	0	0	0
14	0	0	0,5	0	0	0	-0,5	0,5	0,5	0,5	0	1	0	0	1
15	0	0	0,5	0	0	0	-0,5	0,5	0,5	0,5	0	1	0	1	0
16	0	0	0,5	0	0	0	-0,5	0,5	0,5	0,5	0	1	0	1	1
17	0,5	0,5	0	-0,5	0,5	-0,5	0	0	0	0	0,5	0,5	0,5	0,5	0,5
18	0,5	0,5	0	-0,5	0,5	-0,5	0	0	0	0	0,5	0,5	0,5	0,5	0,5
19	0,5	0,5	0	-0,5	0,5	-0,5	0	0	0	0	0,5	0,5	0,5	0,5	0,5
20	0	0	-0,5	0	0	0	0,5	-0,5	0,5	0,5	0	0	0	0	0
21	-0,5	-0,5	-1	-0,5	-0,5	-0,5	0	-1	0	0	-0,5	-0,5	-0,5	-0,5	-0,5
22	-0,5	-0,5	-1	-0,5	-0,5	-0,5	0	-1	0	0	-0,5	-0,5	-0,5	-0,5	-0,5
23	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
24	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
25	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
26	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
27	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
28	0	0	-0,5	0	0	0	0,5	-0,5	-0,5	-0,5	0	-1	0	-1	-1
29	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
30	0,5	0,5	0	-0,5	0,5	-0,5	0	0	-1	-1	0,5	-0,5	0,5	-0,5	-0,5
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	0,5	0,5	0,5	0	-0,5	-0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0,5
2	0	0,5	0,5	0,5	0	-0,5	-0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0,5
3	0,5	0	0	0	-0,5	-1	-1	0	0	0	0	0	-0,5	0	0
4	0	-0,5	-0,5	-0,5	0	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	0	-0,5	-0,5
5	0	0,5	0,5	0,5	0	-0,5	-0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0,5
6	0	-0,5	-0,5	-0,5	0	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	0	-0,5	-0,5
7	-0,5	0	0	0	0,5	0	0	0	0	0	0	0	0,5	0	0
8	0,5	0	0	0	-0,5	-1	-1	0	0	0	0	0	-0,5	0	0
9	0,5	0	0	0	0,5	0	0	-1	-1	-1	-1	-1	-0,5	-1	-1
10	0,5	0	0	0	0,5	0	0	-1	-1	-1	-1	-1	-0,5	-1	-1
11	0	0,5	0,5	0,5	0	-0,5	-0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0,5
12	1	0,5	0,5	0,5	0	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-1	-0,5	-0,5
13	0	0,5	0,5	0,5	0	-0,5	-0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0,5
14	1	0,5	0,5	0,5	0	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-1	-0,5	-0,5
15	1	0,5	0,5	0,5	0	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-1	-0,5	-0,5
16	0	0,5	0,5	0,5	0	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-1	-0,5	-0,5
17	0,5	0	1	1	0,5	0	0	0	0	0	0	0	-0,5	0	0

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
18	0,5	1	0	1	0,5	0	0	0	0	0	0	0	-0,5	0	0
19	0,5	1	1	0	0,5	0	0	0	0	0	0	0	-0,5	0	0
20	0	0,5	0,5	0,5	0	0,5	0,5	-0,5	-0,5	-0,5	-0,5	-0,5	0	-0,5	-0,5
21	-0,5	0	0	0	0,5	0	1	1	1	1	1	1	0,5	1	1
22	-0,5	0	0	0	0,5	1	0	0	0	0	0	0	0,5	0	0
23	-0,5	0	0	0	-0,5	0	0	0	1	1	1	1	0,5	1	1
24	-0,5	0	0	0	-0,5	0	0	1	0	1	1	1	0,5	1	1
25	-0,5	0	0	0	-0,5	0	0	1	1	0	1	1	0,5	1	1
26	-0,5	0	0	0	-0,5	0	0	1	1	1	0	1	0,5	1	1
27	-0,5	0	0	0	-0,5	0	0	1	1	1	1	0	0,5	1	1
28	-1	-0,5	-0,5	-0,5	0	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0,5	0,5
29	-0,5	0	0	0	-0,5	0	0	1	1	1	1	1	0,5	0	1
30	-0,5	0	0	0	-0,5	0	0	1	1	1	1	1	0,5	1	0

Нехай тепер необхідно реалізувати розрахунок характеристик переобтяженого режиму електричної системи, який характеризується збільшеними навантаженнями у вузлах 6, 10, 12, 18, 20, 21, 22, 24, 27, 28, 29 та 30. Для цього у банку архівованих режимів необхідно знайти переобтяжений режим, найбільш близький до шуканого. Таку задачу будемо розв'язувати із використанням нейронної мережі Хопфіда. Вектор вхідних сигналів тут визначається переліком вузлових точок електричної системи із збільшеними навантаженнями, для яких встановлюють сигнал «+1», а для інших – сигнал «-1»:

$$\mathbf{X} = \begin{bmatrix} -1; & -1; & -1; & -1; & -1; & 1; & -1; & -1; & -1; & 1; & \dots \\ \dots -1; & 1; & -1; & -1; & -1; & -1; & -1; & 1; & -1; & 1; & \dots \\ \dots 1; & 1; & -1; & 1; & -1; & -1; & 1; & 1; & 1; & 1. \end{bmatrix}^T$$

Такий вектор подають на вихід мережі, що через наявність зворотних зв'язків означає появу цих сигналів на вході мережі Хопфілда. Проходження цих сигналів через нейронну мережу призводить до перетворення сигналів на виході. Так вихідний сигнал першого нейрону складає

$$\begin{aligned} u_1 = & 0(-1) + 1(-1) + 0,5(-1) + 0(-1) + 1(-1) + 0 \cdot 1 + 0,5(-1) + \\ & + 0,5(-1) - 0,5(-1) - 0,5 \cdot 1 + 1(-1) + 0 \cdot 1 + 1(-1) + 0(-1) + 0(-1) + \\ & + 0(-1) + 0,5(-1) + 0,5 \cdot 1 + 0,5(-1) + 0 \cdot 1 - 0,5 \cdot 1 - 0,5 \cdot 1 + 0,5(-1) + \\ & + 0,5 \cdot 1 + 0,5(-1) + 0,5(-1) + 0,5 \cdot 1 + 0 \cdot 1 + 0,5 \cdot 1 + 0,5 \cdot 1 = -6,5 \rightarrow \\ & y_1 = -1. \end{aligned}$$

В такий самий спосіб визначають вихідні сигнали всіх інших нейроноподібних елементів. В результаті отримуємо вектор вихідних сигналів

$$\mathbf{Y} = \begin{bmatrix} -1; & -1; & -1; & -1; & -1; & -1; & -1; & -1; & 1; & -1; & \dots \\ \dots & -1; & -1; & -1; & -1; & -1; & -1; & -1; & -1; & 1; & \dots \\ \dots & 1; & 1; & 1; & -1; & 1; & 1; & -1; & 1; & -1; & \dots \end{bmatrix}^T$$

Це означає активізацію вихідних сигналів нейроноподібних елементів за номерами 9, 20, 21, 22, 23, 25, 26 та 28. Ці сигнали знов подають на вхід мережі Хопфілда, що знов призводить до перетворення сигналів на виході мережі. Таку операцію проводять циклічно і після виконання чотирьох циклів отримують усталений стан мережі, який відповідає четвертому архівованому режиму. Послідовність таких перетворень умовно наведена на рис. 5.14.

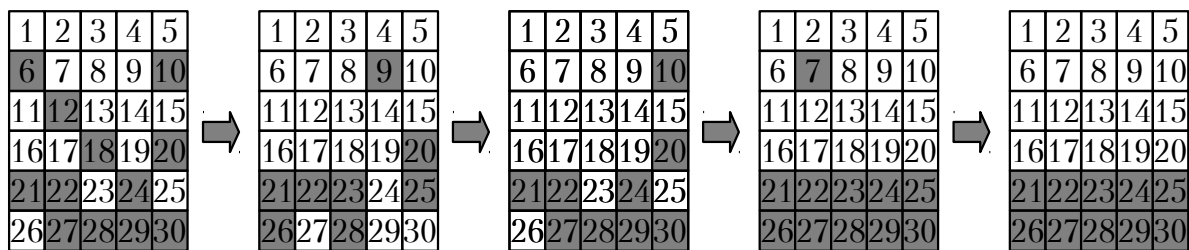


Рис. 5.14. Послідовність перетворень вихідних сигналів мережі Хопфілда

Таким чином, для надійної реалізації ітераційного розрахунку переобтяженого робочого режиму електричної системи тестової схеми IEEE-30 необхідно використовувати дані четвертого архівованого режиму як початкові наближення невідомих параметрів.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

1. Дайте визначення мережі Хопфілда.
2. Наведіть структурну схему мережі Хопфілда.
3. Наведіть алгоритм функціонування мережі Хопфілда.
4. Сформулюйте умови стійкості мережі Хопфілда.
5. Наведіть математичні моделі уніполярної та біполярної мережі Хопфілда.



ЛІТЕРАТУРА

[1], [2], [6], [7], [12], [15], [39]

§5.6. НЕЙРОМЕРЕЖЕВІ ЕКСПЕРТНІ СИСТЕМИ

Останнім часом для побудови експертних систем часто використовують нейромережеві технології. Комбінування використання експертної системи та апарату штучних нейронних мереж забезпечує необхідну гнучкість та самонавчання такої системи. Одночасно отримані від експертів знання дають змогу суттєво спростити структуру нейронної мережі, зменшити її розмірність та зв'язність.

Тут слід звернути увагу на загальні риси нейромережевих та продукційних моделей подання знань. Обидва підходи передбачають розповсюдження інформаційних сигналів по мережах нейроноподібних елементів або продукційних правил. Такі процеси пов'язані із активацією штучних нейронів або спрацьовуванням правил продукційних систем. Це дає змогу реалізувати взаємне заміщення штучних нейронних мереж та продукційних систем подання знань. Так, для реалізації продукційної експертної системи можна сформувати відповідну штучну нейронну мережу, яка буде ефективно розв'язувати практичні задачі предметної області. І навпаки, для деякої штучної нейронної мережі можна сформувати фіктивну систему продукційних правил, яка буде розв'язувати ті самі задачі, що й штучна нейронна мережа.

Так, наприклад, на рис. 5.15 представлена штучна нейронна мережа, яка відповідає системі продукційних правил, пов'язаних із виникненням лавини напруги в електричній системі, наведеної в § 4.3. Нагадаємо, що така продукційна система містить три правила, перше з яких визначає зниження напруги на шинах підстанції у разі збільшення навантаження або аварійного вимкнення однієї з живильних ліній електропередач. Друге правило визначає зв'язок між фактами увімкнення всіх секцій конденсаторної установки та обмеженням діапазону регулювання реактивної потужності. Останнє, третє правило визначає виникнення лавини напруги у разі спрацьовування перших двох правил.

На рис. 5.15 кожному продуктивному правилу відповідає нейроноподібний елемент штучної мережі. Вхідні сигнали x_1 , x_2 та x_3 активізуються у разі збільшення навантаження, аварійного вимкнення однієї з живильних ліній та увімкнення всіх секцій конденсаторної установки відповідно. Проміжні сигнали u_1 та u_2 мають місце у разі активізації нейроноподібних елементів вхідного шару та відповідають фактам провалу напруги на шинах підстанції та обмеження діапазону регулювання реактивної потужності відповідно. Вихідний сигнал штучної нейронної мережі має місце у разі активізації вихідного нейроноподібного елемента, тобто у разі виникнення лавини напруги.

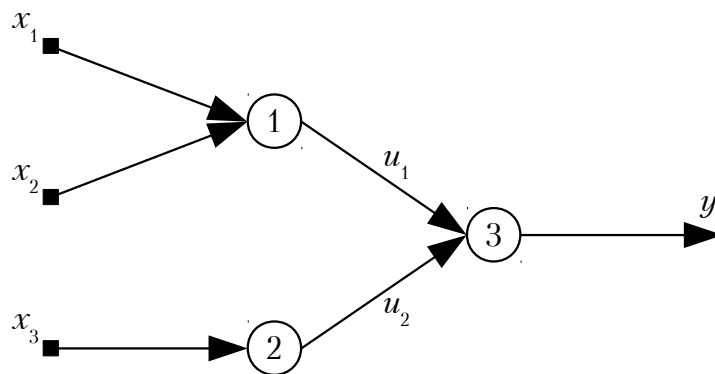


Рис. 5.15. Штучна нейронна мережа системи продукційних правил щодо визначення лавини напруги

Дійсно, представлена на рис. 5.15 штучна нейронна мережа формує вихідний сигнал y у разі виникнення сигналів x_1 або x_2 та x_3 на входах, що відповідає принципам організації вихідної системи продукційних правил.

Наведений підхід дає змогу ефективно використовувати нейромережеві технології для формування експертних систем та розв'язання за їх допомогою практичних задач відповідної предметної області. Такі експертні системи, зазвичай, призначені для розв'язання задач діагностики та прогнозування. Тут штучні нейронні мережі являють собою бази знань, які містять запам'ятовану інформацію щодо розв'язання типових задач предметної області експертної системи за різних можливих вихідних даних, а також реалізують асоціативну пам'ять в процесі розв'язання поточних практичних задач. Такі експертні системи дозволяють запам'ятовувати вдалі розв'язки практичних задач та використовувати таку інформацію в своїй подальшій роботі. Таким чином в нейромережевих експертних системах реалізується логічне виведення інформації на основі прикладів, які містяться в її пам'яті. Це означає, що в таких системах реалізуються принципи індуктивного навчання на основі виведення загального із окремого.

Індуктивний підхід дає змогу суттєво скоротити витрати часу на навчання експертної системи та на наповнення її бази знань, оскільки необхідність представлення закономірностей предметної області в явному вигляді тут відсутня. Бази знань містять лише певні співвідношення між вхідними та вихідними параметрами для типових задач. В результаті нейромережеві експертні системи спроможні самостійно визначати закономірності та виводити нові знання, які ще не були відомі людині. Це визначає широкі можливості застосування нейромережевих експертних систем для розв'язання складно формалізованих задач в областях, де сформулювати адекватну математичну модель дуже важко, або взагалі неможливо.

Типові нейромережеві експертні системи використовують прямі ланцюги логічного виведення для формування заключень в процесі розв'язання задач та зворотнє логічне виведення для формування запитів щодо додаткової інформації, необхідної для розв'язання практичних задач.

❓ КОНТРОЛЬНІ ЗАПИТАННЯ

- 1. Поясніть, у чому полягають переваги комбінування технологій експертних систем та апарату штучних нейронних мереж?*
- 2. Наведіть порівняльний аналіз технологій штучних нейронних мереж та продукційних систем подання знань.*
- 3. Поясніть особливості та переваги використання індуктивного підходу у нейромережевих експертних системах.*



ЛІТЕРАТУРА

[1], [4], [6], [9], [39]

ПРЕДМЕТНИЙ ПОКАЖЧИК

- Абдукція, 24
- Аксон, 135
- Алгоритм, А, 47
 - А*, 49
 - генетичний, 59
 - глобального урахування цілі, 45
 - градієнтний, 44
 - Дейкстри, 40
- Апроксимація, 134
- Асоціативна пам'ять, 134
- Баєсова логіка, 105
- База знань, 20, 72, 86
- База фактів, 86
- Введення диз'юнкції, 79
- Введення кон'юнкції, 79
- Вершина графу термінальна, 32
- Вершина графу тупикова, 32
- Вибірка навчальна, 144
- Вибірка тестова, 144
- Виключення кон'юнкту, 79
- Виключення подвійного заперечення, 80
- Висловлювання, 77
- Витягнення знань, 72
 - – методи комунікативні, 72
 - – – – активні, 73
 - – – – пасивні, 72
 - – – – текстологічні, 73
- Вихідний шар, 142
- Відбір рулековий, 62
- Відношення, 92
 - «є представником», 92
 - «є частиною», 92
 - атрибутивні, 92
 - значень, 92
 - кількісні, 93
 - логічні, 93
 - між нечіткими множинами, 124
 - просторові, 93
 - родо-видові, 92
 - функціональні, 92
 - часові, 93
- Гіпотеза Хебба, 152
- Граф редукції, 31, 52
- Дедукція, 24
- Декомпозиція, 31
- Дендрит, 135
- Дерево пошуку, 29
- Діагностика, 17, 134
- Евристика, 24
- Експерт, 10, 21
- Експертна система, 12
 - – автономна, 19
 - – гібридна, 19
 - – динамічна, 18
 - – система квазідинамічна, 19
 - – нейромережева, 161
 - – статична, 18
- Елементарна формула, 82
- Засоби управління автоматизовані, 8
- Засоби управління автоматичні, 7
- Знання, 68
 - глибинні, 70
 - об'єктивізовані, 72
 - поверхневі, 70
 - суб'єктивні, 72
- Індуктивне навчання, 162
- Інверсія, 61
- Індукція, 24
- Інженер зі знань, 21
- Інженерія знань, 72
- Інтерпретатор, 87
- Інтерпретація, 17
- Інтерфейс користувача, 21
- Інтуїція, 24
- Інформація генетична, 60
- Категоризація, 134
- Квантор, 82
- Класифікація образів, 134
- Кластеризація, 134
- Коефіцієнт навчання, 148
- Комбінаторний вибух, 34
- Конекціоністський підхід, 132
- Контрпропозиція, 79
- Конфліктна множина, 88
- Концепт, 91
- Користувач, 21

- Крок навчання, 151
- Кросинговер, 60
 - одноточковий, 60
 - двоточковий, 60
 - багатоточковий, 60
- Ланцюжок міркувань зворотній, 50
 - – прямий, 50
- Лінгвістична змінна, 121
- Логіка Лукасевича багатозначна, 103
 - – тризначна, 103
- Логічна система формалізації знань, 76
- Логічне виведення, 24
- Максимінна згортка, 129
- Маса упевненості, 115
- Машина логічного виведення, 20, 87
- Мережева функція, 136
- Метазнання, 24, 69
- Метод зворотного розповсюдження помилки, 144
 - зіставлення зі зразком, 94
 - коефіцієнтів упевненості, 110
 - перехресного пошуку, 93
 - спроб та помилок, 24
 - турнірів, 62
- Множина, 121
- Формалізація знань, 68
 - –, моделі декларативні, 75
 - –, моделі процедурні, 75
- Модус поненс, 78
- Модус толленс, 79
- Моніторинг, 17
- Мутація, 61
- Навчання, 17
 - без вчителя, 151
 - з вчителем, 144
 - із підкріпленням, 149
 - штучної нейронної мережі, 144
- Налагодження та ремонт, 17
- Невизначеність, 100
 - лінгвістична, 100
 - об'єктивна, 101
 - суб'єктивна, 101
 - фізична, 100
- Нейрон біологічний, 135
 - вихідний, 142
 - вхідний, 142
 - проміжний, 142
 - штучний, 135
 - –, математична модель, 136
- Немонотонні судження, 24
- Неоднозначність інформації, 100
- Неповнота інформації, 100
- Нечітка логіка, 124
 - множина, 121
- Об'єкт агрегатний, 92
 - індивідний, 91
 - узагальнений, 91
- Оптимізація, 134
- Отримання знань, 22
- Персептрон, 140
 - багатошаровий, 142
- Підсистема отримання знань, 21
- Підсистема пояснень, 20
- Підтримка прийняття рішень, 18
- Планування, 17
- Подання задач у вигляді доведення теорем, 33
 - – у просторі задач, 31
 - – у просторі станів, 29
- Поле знань, 68
- Пошук випадковий, 36
 - двонаправлений, 50
 - евристичний, 44, 56
 - із кінця в початок, 50
 - із початку в кінець, 50
 - комбінований, 50
 - направлений, 44
 - неінформативний, 36
 - під керуванням даних, 50
 - під керуванням цілі, 50
 - сліпий, 36
 - у глибину, 36, 52
 - у ширину, 38, 55
- Пояснення прийнятого рішення, 25
- Правило Демпстера, 118
- Предикат, 82
- Прийняття рішень, 10
- Принцип «стопки книжок», 88
 - метапродукцій, 89
 - найбільш довгої умови, 88
 - пріоритетного вибору, 89
- Прогнозування, 17, 134
- Продукційне правило, 85
- Продукція, 85
- Проектування, 17
- Прозорість експертної системи, 25
- Проста резолюція, 80
- Протиріччя, 78
- Псевдофізичні логіки, 84
- Редукція, 31

- Резолюція, 80
- Робоча пам'ять, 86
- Розкриття вершини графу, 36
- Селекція, 62
- Семантична мережа, 90
- Силогізм, 79
- Силогістика, 76
- Синапс, 135
- Синаптичний зв'язок 137
 - – гальмівний, 137
 - – збуджувальний, 137
- Система підкріплення 149
 - – з корекцією помилок, 149
- Скрипт, 95
 - , початкові умови, 95
 - , припущення, 95
 - , результати, 95
 - , ролі, 95
 - , сцени, 95
- Слот, 97
- Стек, 37
- Ступінь довіри, 114
- Ступінь правдоподібності, 114
- Судження за аналогією, 24
- Судження за замовченням, 24
- Сценарій, 95
- Тавтологія, 78
- Теорема Байєса, 105
- Теорія свідомств Демстера-Шефера, 114
- Терм, 82
- Управління, 17, 134
 - оперативне, 7
 - пошуком рішення, 24
 - стратегічне, 9
 - тактичне, 8
 - централізоване, 7
- Фільтрація, 134
- Фітнес-функція, 62
- Формалізація задач, 28
- Формалізація знань, 23, 74
- Формальна логічна система, 76
 - – –, аксіоми, 76
 - – –, алфавіт, 76
 - – –, логічне виведення, 76
 - – –, синтаксичні правила, 76
- Фрейм 96
 - -меню, 97
 - -поняття, 97
 - -примірник, 97
 - -прототип, 97
 - -сценарій, 97
- Функція активації, 137
 - належності 121, 123
 - – кусочно-лінійна, 123
 - – сигмоїдна, 123
 - – LR , 123
 - – s -подібна, 123
- Характеристична функція, 121
- Ціна стану, 40
- Черга, 39
- Числення висловлювань, 77
- Числення предикатів, 82
- Шар внутрішній, 142
 - вхідний, 142
 - прихований, 142
- Штучна нейронна мережа, 132
 - – – багат шарова, 140
 - – – – з прямими зв'язками, 140
 - – – – із зворотними зв'язками, 140
 - – – одношарова, 140
 - – – – з прямими зв'язками, 140
 - – – – із зворотними зв'язками, 140
 - – – Хопфілда, 153
- Штучний інтелект, 13

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

ОСНОВНА ЛІТЕРАТУРА

1. Бондарев В. Н. Искусственный интеллект [Текст]: учеб. пособ. / В. Н. Бондарев, Ф. Г. Аде. - Севастополь : Изд-во Севастоп. нац. техн. ун-та, 2002. - 616 с.: ил. - Библиогр.: с. 605-608 (95 назв.). - ISBN 966-7473-45-7.
2. Глибовець М. М. Штучний інтелект [Текст]: підручник для студ. вищ. навч. закладів, що навч. за спец."Комп'ютерні науки" та "Прикладна математика" / М. М. Глибовець, О. В. Олецкий. - Київ : Академія, 2002. - 366 с. : іл.; 21 см. - ISBN 966-518-153-X.
3. Девятков В. В. Системы искусственного интеллекта [Текст] / В. В. Девятков. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2001. — 352 с. - ISBN 5-7038-1727-7.
4. Джарратано Д. Экспертные системы: принципы разработки и программирование [Текст] / Д. Джарратано, Г. Райли; пер. с англ. - 4-е изд. - М. : Вильямс, 2007. - 1152 с. : ил. - ISBN 978-5-8459-1156-8.
5. Джексон П. Введение в экспертные системы [Текст]: пер. с англ. / П. Джексон. - 3-е изд. - М.; СПб.; К.: Вильямс, 2001. - 622, [2] с.: ил. - Библиогр.: с. 597-616. - Предм. указ.: с. 617-622. - ISBN 5-8459-0150-2.
6. Круглов В. В. Искусственные нейронные сети. Теория и практика [Текст] / В. В. Круглов, В. В. Борисов. — М.: Горячая линия — Телеком, 2002. — 382 с. — ISBN 5-93517-031-0.
7. Люгер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем [Текст]: пер. с англ. / Дж. Ф. Люгер. - 4-е изд. — М.; Спб.; К.: Издательский дом "Вильямс", 2003. — 864 с. - ISBN 5-8459-0437-4.
8. Рассел С. Искусственный интеллект: современный подход [Текст]: пер. с англ. / С. Рассел, П. Норвиг. - 2-е изд. - М.; СПб.; К.: Вильямс, 2006. - 1408 с. : ил. - Библиогр.: с. 1302-1372 . - Предм. указ.: с. 1373-1407. - ISBN 5-8459-0887-6.
9. Ручкин В. Н. Универсальный искусственный интеллект и экспертные системы [Текст] / В. Н. Ручкин, В. А. Фулин. — СПб.: БХВ - Петербург, 2009. - 240 с. - ISBN 978-5-9775-0460-7.
10. Смолин Д. В. Введение в искусственный интеллект: конспект лекций. [Текст] / Д. В. Смолин — М.: ФИЗМАТЛИТ, 2004. — 208 с. — ISBN 5-9221-0513-2.

- 11.Субботін С.О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень: Навчальний посібник. [Текст] / С. О. Субботін. – Запоріжжя: ЗНТУ, 2008 – 341 с. – ISBN 978-966-7809-87-4.
- 12.Хайкин С. Нейронные сети: полный курс, 2-е издание. : Пер. с англ. [Текст] / Саймон Хайкин. – М.: Издательский дом “Вильямс”, 2006. – 1104 с. – ISBN 5-8459-0890-6.
- 13.Частиков А. П. Разработка экспертных систем. Среда CLIPS [Текст] / А. П. Частиков, Т. А. Гаврилова, Д. Л. Белов. - СПб.: БХВ - Петербург, 2003. - 608 с. : ил. - Глоссарий: с. 582-606. - ISBN 5-94157-248-4
- 14.Черноруцкий И. Г. Методы принятия решений [Текст] / И. Г. Черноруцкий. - СПб.: БХВ-Петербург, 2005. - 416 с. - ISBN 5-94157-481-9.
- 15.Ясницкий Л. Н. Введение в искусственный интеллект: учеб. пособие для студ. вузов [Текст] / Л. Н. Ясницкий. - 2-е изд., испр. - М.: Издательский центр «Академия», 2008. - 176 с. - ISBN 978-5-7695-5390-5.

ДОДАТКОВА ЛІТЕРАТУРА

- 16.Аверкин А. Н. Нечёткие множества в моделях управления и искусственного интеллекта [Текст] / А. Н. Аверкин и др. – Под ред. Д. А. Поспелова. – М.: Наука, 1986. – 312 с.
- 17.Адаменко А. Н. Логическое программирование и Visual Prolog [Текст] / А. Н. Адаменко, А. М. Кучуков. – СПб, 2003. – 992 с. – ISBN 5-94157-156-9.
- 18.Балтрашевич В. Э. Реализация инструментальной экспертной системы [Текст] / В. Э. Балтрашевич. – СПб: Политехника, 1993. – 237 с. — ISBN 5-7325-0037-5.
- 19.Баркан Я. Д. Эксплуатация электрических систем [Текст]: учебн. пособие / Я. Д. Баркан. – М.: Высш. шк., 1990. – 304 с. – ISBN 5-06-000448-1
- 20.Барский А. Б. Нейронные сети: распознавание, управление, принятие решений [Текст] / А. Б. Барский. — М.: Финансы и статистика, 2004. — 176 с. — ISBN 5-279-02757-X.
- 21.Башлыков А. А. Проектирование систем принятия решений в энергетике [Текст] / А. А. Башлыков. - М.: Энергоатомиздат, 1986. - 120 с.

22. Башмаков А. И. Интеллектуальные информационные технологии: Учеб. Пособие [Текст] / А. И. Башмаков, И. А. Башмаков. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 304 с : ил. – ISBN 5-7038-2544-X.
23. Беркинблит М. Б. Нейронные сети: Учебное пособие [Текст] / М. Б. Беркинблит. – М.: МИРОС и ВЗМШ РАО, 1993. – 96 с. – ISBN 5-7084-0026-9.
24. Блюмин С. Л. Модели и методы принятия решений в условиях неопределенности [Текст] / С. Л. Блюмин, Л. А. Шуйкова. – Липецк: ЛЭГИ, 2001. – 138 с. – ISBN 5-900037-19-3.
25. Борисов А. Н. Обработка нечёткой информации в системах принятия решений [Текст] / А. Н. Борисов и др. – М.: Радио и связь, 1989. – 304 с. – ISBN 5-256-00178-7.
26. Борисов А. Н. Принятие решений на основе нечётких моделей: Примеры использования [Текст] / А. Н. Борисов, О. А. Крумберг, И. П. Федоров. – Рига: Зинатне, 1990. – 184 с. – ISBN 5-7966-0489-7.
27. Братко И. Алгоритмы искусственного интеллекта на языке PROLOG [Текст]: пер. с англ. / Иван Братко. – М.: Вильямс, 2004. – 640 с. : ил. – ISBN 5-8459-0064-4.
28. Вагин В. Н. Достоверный и правдоподобный вывод в интеллектуальных системах [Текст] / В. Н. Вагин, Е. Ю. Головина, А. А. Загорянская, М. В. Фомина – под ред. В. Н. Вагина, Д. А. Поспелова. – М.: ФИЗМАТЛИТ, 2004. – 704 с. – ISBN 5-9221-0474-8.
29. Васенев Ю. Б. Экспертные системы в GURU. Справочное пособие [Текст] / Ю. Б. Васенев, А. Г. Бабанин, М. Б. Бессонов, М. Н. Блинов – СПб: Санкт-Петербургский государственный университет, 1993. – 112 с.
30. Гаврилов А. В. Гибридные интеллектуальные системы [Текст] / А. В. Гаврилов. – Новосибирск: Изд-во НГТУ, 2003. – 164 с – ISBN 5-7782-0413-2
31. Гаврилова Т. А. Базы знаний интеллектуальных систем [Текст]: учеб. пособие для студ. вузов, обуч. по напр. «Прикладная математика и информатика», «Информатика и вычислительная техника», «Прикладная математика» / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб. и др.: Питер, 2000, 2001. – 382 [2] с. : ил. – Библиогр.: с. 358-382 (457 назв.). – ISBN 5-272-00071-4.
32. Галушкин А. И. Теория нейронных сетей. Кн. 1: Учебн. пособие для вузов [Текст] / А. И. Галушкин, – М.: ИПРЖР, 2000. – 416 с. – ISBN 5-93108-05-8.
33. Джонс М. Т. Программирование искусственного интеллекта в приложениях: пер. с англ. [Текст] / М. Тим Джонс. – М.: ДМК Пресс, 2004. – 312 с. – ISBN 5-94074-275-0.

- 34.Дюбуа Д. Теория возможности. Приложения к представлению знаний в информатике [Текст]: пер. с фр. / Д. Дюбуа, А. Прад. – М.: Радио и связь, 1990. – 288 с. – ISBN 5-256-00184-1.
- 35.Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений [Текст]: пер. с англ. / Л. Заде. – М.: Мир, 1976. – 167 с.
- 36.Искусственный интеллект: в 3-х кн. Кн. 1. Системы общения и экспертные системы: Справочник [Текст] / под ред. Э. В. Попова. – М.: Радио и связь, 1990. – 464 с. – ISBN 5-256-00365-8.
- 37.Искусственный интеллект: в 3-х кн. Кн. 2. Модели и методы: Справочник [Текст] / под ред. Д. А. Поспелова. – М.: Радио и связь, 1990. – 304 с. – ISBN 5-256-00368-2.
- 38.Искусственный интеллект: в 3-х кн. Кн. 3. Программные и аппаратные средства: Справочник [Текст] / под ред. В. Н. Захарова, В. Ф. Хорошевского. – М.: Радио и связь, 1990. – 368 с. – ISBN 5-256-00366-6.
- 39.Каллан Р. Основные концепции нейронных сетей: пер. с англ. [Текст] / Роберт Каллан. – М.: Издательский дом “Вильямс”, 2001. – 287 с. – ISBN 5-8459-0210-X.
- 40.Кини Р. Л. Принятие решений при многих критериях: Предпочтения и замещения [Текст]: пер. с англ./ Р. Л. Кини, Х. Хайфа. – М.: Радио и связь, 1981. – 560 с.
- 41.Клини С. К. Математическая логика [Текст]: пер. с англ. / С. К. Клини. – М.: Мир, 1973. – 480 с.
- 42.Ковальски Р. Логика в решении проблем [Текст]: пер. с англ. / Р. Ковальски. – М.: Мир, 1990. – 280 с. – ISBN 5-02-014148-8.
- 43.Комашинский В. И. Нейронные сети и их применение в системах управления и связи [Текст] / В. И. Комашинский, Д. А. Смирнов. – М.: Горячая линия – Телеком, 2003. – 94 с. – ISBN 5-93517-094-9.
- 44.Кофман А. Введение в теорию нечётких множеств [Текст]: пер. с фр. / А. Кофман. – М.: Радио и связь, 1982. – 432 с.
- 45.Крисевич В. С. Экспертные системы для персональных компьютеров [Текст]: методы, средства, реализации: Справ. Пособие / В. С. Крисевич и др. - Минск: Выш. шк., 1990. - 190 [7] с.: ил; - Библиогр.: с. 191-192 (35 назв.). - Предм. указ.: с. 193-195. - ISBN 5-339-00492-9
- 46.Ларычев О. И. Выявление экспертных знаний [Текст] / О. И. Ларычев и др. – М.: Наука, 1989. – 128 с. – ISBN 5-02-006736-9
- 47.Левин Р. Практическое введение в технологию искусственного интеллекта и экспертных систем с иллюстрациями на Бейсике [Текст]: пер. с англ. / Р. Левин, Д. Дранг, Б. Эделсон. – М.: Финансы и статистика, 1990. – 239 с. – ISBN 5-279-00448-0.

48. Леоненков А. В. Нечёткое моделирование в среде MATLAB и fuzzyTECH [Текст] / А. В. Леоненков. – СПб.: БХВ-Петербург, 2005. – 736 с. – ISBN 5-94157-087-2.
49. Лорьер Ж.-Л. Системы искусственного интеллекта — М.: Мир, 1991. – 568 с. – ISBN 5-03-001408-X.
50. Луценко Е. В. Интеллектуальные информационные системы [Текст] / Е. В. Луценко. – Краснодар: КубГАУ, 2006. – 645 с. – ISBN 5-94672-060-0
51. Малышев Н. Г. Нечёткие модели для экспертных систем в САПР [Текст] / Н. Г. Малышев, Л. С. Берштейн, А. В. Боженюк. – М.: Энергоатомиздат, 1991. – 136 с. – ISBN 5-283-01592-0.
52. Маркович И. М. Режимы энергетических систем [Текст] / И. М. Маркович. – М.: Энергия, 1969. – 352 с.
53. Минский М. Перцептроны [Текст]: пер. с англ. / М. Минский, С. Пейперт. – М.: Мир, 1971. – 261 с.
54. Минский М. Фреймы для представления знаний [Текст]: пер. с англ. / М. Минский. – М.: Энергия, 1979. – 152 с.
55. Нейлор К. Как построить свою экспертную систему [Текст] : пер. с англ. / К. Нейлор. - М. : Энергоатомиздат, 1991. - 286 с. - ISBN 5-283-02502-0.
56. Нильсон Н. Принципы искусственного интеллекта [Текст]: пер. с англ. / Н. Нильсон. – М.: Радио и связь, 1985. – 376 с.
57. Новиков П. П. Элементы математической логики [Текст] / П. П. Новиков. – М.: Наука, 1973. – 400 с.
58. Орловский С. А. Проблемы принятия решений при нечеткой исходной информации [Текст] / С. А. Орловский. – М.: Наука, 1981. – 208 с.
59. Осовский С. Нейронные сети для обработки информации: пер. с польского [Текст] / Станислав Осовский. – М.: Финансы и статистика, 2002. – 344 с. – ISBN 5-279-02567-4.
60. Осуга С. Обработка знаний [Текст] : пер. с япон. / С. Осуга. - М.: Мир, 1989. - 293 с. - ISBN 5-03-000638-9.
61. Пономарев А. С. Нечёткие множества в задачах автоматизированного управления и принятия решения: Навчальний посібник [Текст] / О. С. Пономарьов. - Харків: НТУ “ХПІ”, 2005. - 232 с. - Рос. мовою – ISBN 966-593-393-3.
62. Попов Э. В. Статические и динамические экспертные системы: Учеб. пособие [Текст] / Э. В. Попов, И. Б. Фоминых, Е. Б. Кисель, М. Д. Шапот. – М.: Финансы и статистика, 1996. - 320 с. – ISBN 5-279-01598-9.
63. Попов Э. В. Экспертные системы. Решение неформализованных задач в диалоге с ЭВМ [Текст] / Э. В. Попов. – М.: Наука, 1987. – 288 с.

64. Построение экспертных систем [Текст]: пер. с англ. / под ред. Ф. Хайеса-Рота, Д. Уотермана, Д. Лената. – М.: Мир, 1987. – 441 с.
65. Представление и использование знаний [Текст] : пер. с япон. / под ред. Х. Уэно, М. Исидзука. – М.: Мир, 1989. – 220 с. – ISBN 5-03-000685-0.
66. Рутковская Д. Нейронные сети, генетические алгоритмы и нечёткие системы: пер. с польского [Текст] / Д. Рутковская, М. Пилиньский, Л. Рутковский. – М.: Горячая линия – Телеком, 2006. – 452 с. – ISBN 5-93517-103-1.
67. Слэйгл Дж. Искусственный интеллект. Подход на основе эвристического программирования. Пер. с англ. [Текст] / Дж. Слэйгл. – М.: Мир, 1973 - 319 с.
68. Сойер Б. Программирование экспертных систем на Паскале [Текст]: пер. с англ. / Б. Сойер, Д. Л. Фостер. – М.: Финансы и статистика, 1990. – 191 с. – ISBN 5-279-00449-9.
69. Таунсенд К. Проектирование и программная реализация экспертных систем на персональных ЭВМ [Текст]: Пер. с англ. / Таунсенд К., Фохт Д. – М.: Финансы и статистика, 1990. - 320 с. – ISBN 5-279-00255-0.
70. Тейз А. Логический подход к искусственному интеллекту [Текст] : пер. с франц. / А. Тейз, П. Грибомон, Ж. Луи и др. – М.: Мир, 1990. – 432 с. – ISBN 5-03-001636-8.
71. Хант Э. Искусственный интеллект. Пер. с англ. [Текст] / Э. Хант. – М.: Мир, 1978 – 558 с.
72. Штовба С. Д. Проектирование нечетких систем средствами MATLAB [Текст] / С. Д. Штовба. – М.: Горячая линия – Телеком, 2007. – 288 с. – ISBN 5-93517-359-X.
73. Экспертные системы. Принципы работы и примеры [Текст]: пер. с англ. / А. Брукинг, П. Джонс, Ф. Кокс и др.; под ред. Р. Форсайта. – М.: Радио и связь, 1987. – 224 с.
74. Электрические системы. Автоматизированные системы управления режимами энергосистем [Текст]: учебн. пособие / под ред. В. А. Веникова. – М.: Высшая школа, 1979. – 447 с.
75. Электрические системы. Кибернетика электрических систем [Текст]: учебн. пособие / под ред. В. А. Веникова. – М.: Высшая школа, 1974. – 328 с.
76. Элти Дж. Экспертные системы: концепции и примеры [Текст]: пер. с англ. / Дж. Элти, М. Кумбс. – М.: Финансы и статистика, 1987. – 191 с.
77. Эндрю А. Искусственный интеллект [Текст]: пер. с англ. / А. Эндрю. – М.: Мир, 1985. – 264 с.

78. Expert Systems [Текст] / edited by Petrica Vizureanu. – Vukovar, Croatia: Intech, 2010. – 238 pp. – ISBN 978-953-307-032-2
79. Fuzzy Logic – Algorithms, Techniques and Implementations [Текст] / Edited by Elmer P. Dadios. – Rijeka, Croatia: InTech, 2012. – 282 pp. – ISBN 978-953-51-0393-6
80. Haykin S. Neural Networks: A Comprehensive Foundation [Текст] / Simon Haykin. – Pearson Education, Inc, 1999. – 823 pp. – ISBN 81-7808-300-0
81. Jones M. T. Artificial Intelligence: A Systems Approach [Текст] / M. Tim Jones. – Infinity Science Press LLC, 2008. – 498 pp. – ISBN 978-0763773373
82. Munakata T. Fundamentals of the New Artificial Intelligence [Текст] / Toshinori Munakata. – Springer, 2008. – 255 pp. – ISBN 978-1-84628-838-8
83. Pool D. L. Artificial Intelligence [Текст] / David L. Poole, Alan K. Mackworth. – Cambridge University Press, 2010. – 662 pp. – ISBN 978-0-521-51900-7
84. Rabunal J. R. Artificial Neural Networks in Real-Life Applications [Текст] / Juan R. Rabunal, Julian Dorado. – Idea Group Publishing, 2006. – 375 pp. – ISBN 1-59140-902-0
85. Ross T. J. Fuzzy Logic With Engineering Applications [Текст] / Timothy J. Ross. – John Wiley & Sons, 2010. – 585 pp. – ISBN 978-0-470-74376-8
86. Russel S. J. Artificial Intelligence: A Modern Approach [Текст] / Stuart J. Russel, Peter Norving. – Pearson Education, Inc, 2003. – 177 pp. – ISBN 0-13-090376-0
87. Shi Z. Advanced Artificial Intelligence [Текст] / Zhongzhi Shi. – World Scientific, 2011. – 613 pp. – ISBN 978-981-4291-34-7
88. Siler W. Fuzzy Expert Systems and Fuzzy Reasoning [Текст] / William Siler, James J. Buckley. – New Jersey: John Wiley & Sons, 2005. – 405 pp. – ISBN 0-471-38859-9.



Кацадзе Теймураз Луарсабович

кандидат технічних наук, доцент
кафедри електричних мереж та систем
Національного технічного університету України
«Київський політехнічний інститут»

В навчальному посібнику розглянуті основні положення застосування інтелектуальних технологій, зокрема експертних систем в галузі електроенергетики. Теоретичний матеріал проілюстрований прикладами розв'язання практичних електроенергетичних задач із застосуванням моделей та методів штучного інтелекту. Наведено питання для самоконтролю знань за тематикою розділів.

