кластерному аналізі./ Т. Савчук, С. Петришин – м. Вінниця – 2010 р – (Тези XXXIX науково-технічної конференції професорсько-викладацького складу, співробітників та студентів університету з участю працівників науково-дослідницьких організацій та інженерно-технічних працівників підприємств м.Вінниці та області. ВНТУ). 3. Савчук Т.О. Порівняльний аналіз використання методів кластеризації для ідентифікації надзвичайних ситуацій на залізничному транспорті / Т. Савчук, С. Петришин – 2010. – Вип. 11(134). – С. 135–140 – (Наукові праці Донецького національного технічного університету. – Серія "Інформатика, кібернетика і обчислювальна техніка"). 4. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н. Загоруйко – Новосибирск. – 1999. – с. 270. 5. Савчук Т.О. Оцінювання результатів моделювання процесу кластерного аналізу надзвичайних ситуацій на залізничному транспорті / Т.О. Савчук, С.І. Петришин. – 2012. – №1, С. 18–24 – (Інформаційні технології та комп'ютерна інженерія).

УДК 004.8

О.Ю. Сєдушев, Є.В. Буров
Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

МЕТОДИ ВИДОБУВАННЯ ДАНИХ З БАЗ НЕЧІТКИХ ЗНАНЬ

Досліджено нечіткі методи видобування даних. Акцент при цьому робиться на інтелектуальному аналізі баз нечітких знань та задачах, які при цьому виникають. Описано найпопулярніші сьогодні методи, їхні переваги та отримані за їх допомогою результати. Наведено узагальнені варіанти використання таких методів.

Ключові слова: нечіткі методи видобування даних, база нечітких правил, нечітка логіка, база знань.

The paper aims to study the fuzzy data mining techniques. The emphasis is put on an intelligent analysis of fuzzy knowledge bases and problems that arise. Most popular methods are described, their advantages and results obtained with their assistance are highlighted. Generalized use cases of such methods are given.

Key words: fuzzy data mining methods, fuzzy rule base, fuzzy logic, knowledge base.

Вступ та постановка проблеми

Бази нечітких знань являють собою сукупність фактів, лінгвістичних змінних та відповідних функцій приналежності (сукупно трактуватимемо їх як знання), якими можна оперувати, та нечітких висловлювань "ЯКЩО–ТО", що мають назву нечітких продукційних правил виведення. Такі бази знань є цінним джерелом для опису нечітких понять, видобування даних та прийняття різнорідних рішень у різних галузях науки, бізнесу та виробництва, а також є ефективним засобом моделювання у багатьох задачах кібернетики та штучного інтелекту, що мають справу з нечіткостями, серед яких управління технологічними процесами, різного роду діагностики, розпізнавання образів та мови, прогнозування часових рядів тощо.

Чіткі та нечіткі бази знань використовуються сьогодні в багатьох напрямках застосування інформаційних технологій: для побудови експертних та інтелектуальних систем, систем дистанційного навчання та контролю знань тощо. Сьогодні поширені системи підтримки прийняття рішень, які використовують знання, отримані від експертів. Такі знання зберігаються у базах знань, які зазвичай слугують для різнорідного інтелектуального аналізу та виявлення (виведення) певних закономірностей. Усе частіше моделювання складних залежностей в економіці, медицині, будівництві та в інших областях здійснюється за допомогою саме нечітких баз знань. А тому

розроблення та дослідження методів видобування даних із такого роду баз знань є актуальним завданням.

Методи видобування даних можна поділити на певну кількість базових груп, кожна з яких має справу з тими чи іншими задачами: класифікації, кластеризації, прогнозування, асоціації тощо [1]. Сьогодні існують різноманітні методи та алгоритми інтелектуального аналізу даних, що є придатними для вирішення конкретних завдань:

- дерева рішень та символьні правила класифікації і прогнозування;
- штучні нейронні мережі;
- карти Кохонена (як спеціальний тип штучних нейронних мереж);
- статистичні методи, серед яких байєсівські мережі, кореляційний, регресійний та дисперсний аналізи, лінійна та нелінійна регресії;
- ієрархічні та неієрархічні (ітераційні) методи кластерного аналізу;
- методи найближчого і k-найближчого сусіда;
- методи пошуку асоціативних правил та часових шаблонів;
- генетичні алгоритми та еволюційне програмування;
- метод опорних векторів.

Проблема полягає у придатності та можливості застосування вищенаведених методів та алгоритмів у тому чи іншому вигляді/варіанті до задач видобування даних із нечітких баз знань, оскільки традиційні методи та алгоритми інтелектуального аналізу розраховані насамперед для застосування у середовищах, котрі використовують класичні бази знань, бази та сховища даних. Під класичними тут мається на увазі репозиторії, що не містять різного роду нечітких, невизначених та неточних даних та знань.

Аналіз останніх досліджень і публікацій

У [2] автори розповідають про можливість лінгвістичної інтерпретації фактів, що зберігаються у нечітких базах знань, за допомогою механізму виведення нечітких продукційних правил, тобто наводять приклад використання нечіткої логіки. Розглядаються обмеження щодо генерування нових правил та їхньої якості, а також можливі шляхи та підходи до утворення правил.

Дуже поширені різноманітні методи та алгоритми інтелектуального аналізу даних для пошуку нечітких класифікаційних правил, які відрізняються від звичайних нечітких продукційних правил тим, що застосовуються для вирішення саме класифікаційних проблем та завдань і місять у результуючій частині певну мітку, яка відповідає результуючому класу або результуючій змінній з певної наперед визначеної множини тощо. З деякими цікавими підходами можна ознайомитися у [3]. Тоді як у [4] використовується поняття лінгвістичних правил. За суттю вони не відрізняються від нечітких класифікаційних правил, але використовуються здебільшого для подібної до класифікації задачі розпізнавання образів та утворюються на основі використання та інтерпретації лінгвістичних змінних.

Бази нечітких знань нерозривно пов'язані із системами, які їх використовують. У 1990-ті рр. системи нечітких баз знань (під таким терміном маються на увазі інтелектуальні системи, ядром яких є саме нечіткі бази знань та які активно використовують теорію нечітких множин і механізми нечіткої логіки) зазнали суттєвого впливу апарату штучних нейронних мереж. У результаті цього виникли [5]:

- Нейронечіткі системи (мережі), що являють собою нечіткі системи, доповнені нейронними мережами для підвищення таких характеристик, як гнучкість, швидкість, інтерпретація та адаптація;
- Нечіткі штучні нейронні мережі, які являють собою нейронні мережі з можливістю обробки нечіткої інформації, поданої на вхід. У цьому випадку нечітка логіка є лише інструментом нейронних мереж, а тому така мережа не може бути інтерпретована в нечіткі правила, оскільки являє собою "чорну скриньку". Такі мережі у цій статті не розглядаються. У [6] автор наводить варіант використання нечітких нейронних мереж для прогнозування курсу акцій на валютній біржі. При цьому він використовує різні нечіткі та агреговані економічні показники за попередні періоди, які занесені у базу знань.

Застосувавши теорію нечіткої логіки, утворені технології суттєво підвищили точність (знизили похибку) при задачах класифікації та регресії.

Дещо пізніше властивість глобальної оптимізації напряму еволюційних алгоритмів (складена назва, куди входять генетичні алгоритми, еволюційне програмування, еволюційні стратегії та інші подібні алгоритми) також була вбудована у системи нечітких баз знань. Відтак подібні системи отримали назву еволюційних нечітких систем. Тоді як нейронечіткі системи зазвичай могли оптимізувати тільки неперервні параметричні значення, еволюційні нечіткі системи вирвалися уперед та давали змогу оптимізувати багато аспектів систем нечітких баз знань, включаючи комбінаційну оптимізацію (підбір та корегування різних параметрів, вибір придатніших нечітких правил тощо).

На перетині нечіткої логіки та еволюційних обчислень утворилися також і нечіткі еволюційні алгоритми. У таких алгоритмах, як правило, фазифіковані певні параметри чи компоненти (фітнес функція, критерій зупинки, генетичні оператори тощо) для того, щоб мати змогу обробляти нечіткі дані. Знову ж таки, як і нечіткі штучні нейронні мережі, у цій статті такі алгоритми не розглядаємо.

Багато статей присвячено нечітким деревам рішень та алгоритмам, які покладено в основу їхньої побудови [7–9]. Один цікавий підхід розглянуто у [9]. Йдеться про застосування матричного методу для систем нечітких баз знань, який виконує процес нечіткого виведення правил за наявності нечітких дерев рішень. Автори використовують його вже після того, як дерево побудовано. В основу метода покладено використання перехідних матриць (transition matrices), що пришвидшує вихідні обчислення та усуває деякі небажані ефекти нечітких дерев рішень завдяки проведенню попередніх обчислень.

Загалом можна зауважити, що кількість статей та досліджень, присвячених задачам видобування даних, традиційним та нечітким методам видобування даних, застосуванню таких методів до баз даних та знань, є досить великою та невпинно продовжує зростати кожного року, що свідчить про актуальність вищезазначених напрямів.

Формулювання цілі статті

Метою статті є огляд, аналіз та узагальнення нечітких методів видобування даних та їхнього можливого застосування для інтелектуального аналізу баз нечітких знань, наведення прикладів таких застосувань. Робота не має на меті огляд та аналіз усіх існуючих сьогодні методів видобування даних, а скоріше концентрує увагу на множині найдієвіших та придатніших для поставленої мети, тому поза увагою залишено багато інших методів та технологій видобування даних.

Виклад основного матеріалу

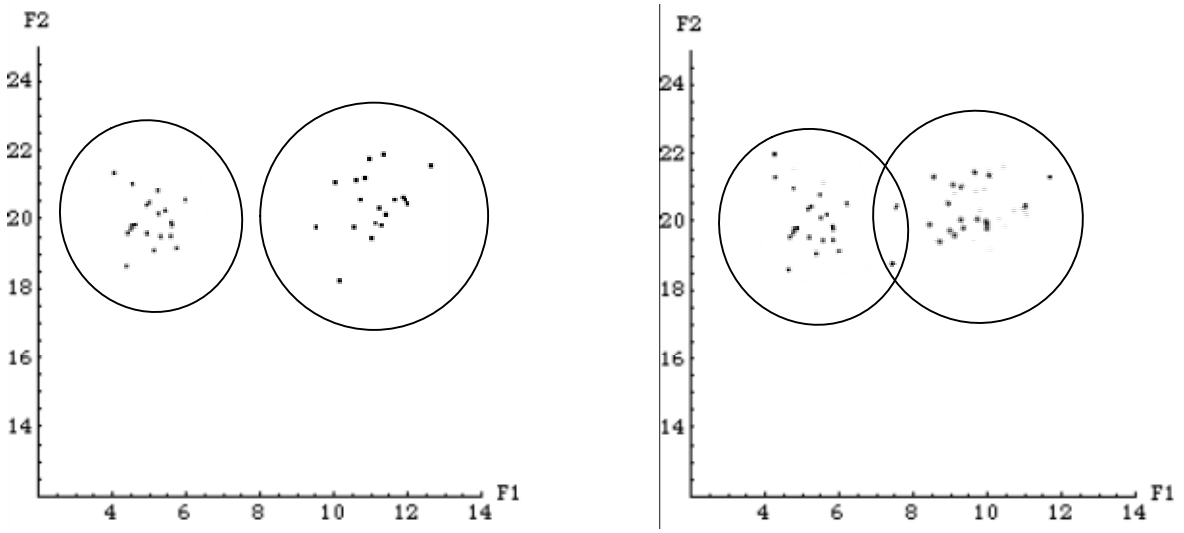
Ця робота концентрує увагу на таких методах та технологіях:

- нечітка кластеризація;
- нечіткі дерева рішень;
- нечіткий асоціативний аналіз;
- нейронечіткі мережі;
- генетичні нечіткі технології.

Методи кластеризації є одними з найважливіших, якщо йдеться про неконтрольоване навчання. В аналізі даних вони часто застосовуються на перших кроках для того, щоб отримати метадані та візуальний вигляд кластерного розміщення поданої на вхід множини об’єктів. Зазвичай під кластеризацією розуміють процес групування колекції об’єктів у класи-кластери так, щоб примірники у кожному класі були найбільш подібними у певному аспекті на своїх сусідів по кластеру, а відмінність з примірниками-об’єктами інших класів була суттєвою. Кластеризацію можна застосовувати для побудови ієрархії таксонів, описового та візуального аналізу даних тощо.

За стандартної кластеризації кожен об’єкт-примірник належить лише до одного кластера. Тому окремі кластери відділяються явними границями (зазвичай порожніми областями) або плавно переходять один в одного та не можуть перетинатися. За нечіткої кластеризації об’єкт може бути віднесений до кількох кластерів одночасно, міра його приналежності до того чи іншого кластера

виражається через функцію приналежності. Як правило, також вводиться обмежено, за якого сума значень ступенів приналежності об'єкта *X* до усіх кластерів, до які він входить, дорівнює щонайбільше 1. Результати уявної кластеризації в обох випадках можна зобразити графічно способом (рис. 1).



а *б*

Рис. 1. Результати: класичної кластеризації (а); нечіткої кластеризації (б)

Варто зауважити, що на рис. 1 не показано функції належності. Присутність функції належності додає елементам відповідного відтінка/об'єму/тощо.

Якщо говорити про застосування нечіткого кластерного аналізу до баз нечітких знань, то слід сказати, що у такому випадку його можна використовувати з такою метою:

1. Використання розпізнаних кластерів для побудови нечітких правил (кількість кластерів дорівнює кількості правил, функції приналежності об'єктів можна відшукати завдяки проектуванню нечітких кластерів на відповідні координатні осі), зокрема визначення областей-кластерів у деякому заданому просторі (зазвичай вхідному), які можна використати надалі для формування умовної частини нечіткого правила. Тут маються на увазі антецеденти правила вигляду "x ∈ A", де x може бути певним атрибутом, лінгвістичною змінною тощо; A являє собою розпізнаний кластер, нечітку множину або проекцію нечіткої області.
2. Використання розпізнаних кластерів як певної функції належності для характеристики нечіткої множини чи її елементів.
3. Використання кластерів для мінімізації наявного банку правил [10].

Нечітку кластеризацію переважно доволі успішно застосовують при аналізі та розвідці різного роду баз даних.

Для баз знань дуже характерними задачами є виведення правил (нових чи на основі існуючих), редукція наявних правил або моделі правил та адаптація моделі правил до нового середовища [11]. Розв'язання таких задач дозволяє базі знань успішно показувати класифікаційні та регресійні (прогнозуючі) функції. Серед методів видобування даних важливе місце посідає метод дерев рішень, який є доволі універсальним для розв'язання задач класифікації та регресії.

Для виведення дерев рішень розроблено багато алгоритмів, серед яких ID3, C4.5/C5.0, CART, та ін. В основу цього методу покладено рекурсивне розбиття навчальної множини примірників, допоки отримані підмножини не стануть однорідними (звичайно ж відносно вихідного атрибута, де в задачі класифікації таким є мітка класу, а в задачі регресії – неперервне числове значення). Кожен внутрішній вузол дерева рішень визначає подальше розбиття множини примірників, тобто аналізує значення їхніх певних атрибутів та апостеріорі відправляє кожен примірник на нижчий рівень

дерева. Розбиття відбувається з огляду на отримання найбільшої однорідності та побудови якнайпростішого для інтерпретації дерева рішень.

Як тільки дерево рішень буде побудоване, кожен шлях (від кореня до результуючого листка) можна інтерпретувати як окреме незалежне правило. Нові примірники класифікуються за такими шляхами/правилами. Отже, виведення дерева рішень можна розглядати як виведення моделі правил.

Нечіткі варіанти дерев рішень (нечіткі дерева рішень) розробляються вже давно. Щоб пояснити, чому вони називаються нечіткими, наведемо такий приклад [11]. У звичайних дерев рішень критерій розбиття у внутрішніх вузлах є чітким та точним (для прикладу, *Зарплата* ≤ *2000*). Такі порогові значення призводять до чітких граничних рішень, тому якщо *Зарплата = 2010*, то екземпляр буде класифіковано по-іншому. Хоча насправді його варто було б віднести до першого випадку, оскільки різниця у 10 у. о. є незначною. Крім того, навчальний процес дерева є у певному роді нестабільним, оскільки невелика варіація вхідних значень (в околі порогового значення) дуже сильно впливає на класифікацію/регресію та на вигляд дерева.

Для того, щоб усунути цей недолік та згладити вищеописаний приклад, виникла ідея застосовувати у внутрішніх вузлах дерева нечіткі предикати розбиття (наприклад, Зарплата $\in$ Середня, де Середня є значенням, що задане нечіткою множиною). Оскільки вхідні приклади можуть задовольняти нечіткий предикат лише певною мірою (тобто приналежність не дорівнює строго 1), то вони можуть також і розбиватися у нечіткий спосіб. Це означає, що приклад не приписується унікально лише до одного вузла, а швидше до декількох водночас, але з певною мірою (сума мір приналежності не може перевищувати 1). Наприклад, зарплата у 2000 у. о. може належати до середнього класу як 0.4, а до нижче середнього – як 0.6.

Наведемо приклад та вигляд нечіткого дерева рішень. Нехай існує псевдопотреба у класифікації людей за деякими параметрами (вага, зріст, колір очей). Існує деяка множина цільових класів $Y = \{Y_1, Y_2\}$ та певна навчальна вибірка екземплярів для конструювання дерева. Тоді, використавши алгоритм Fuzzy ID3, можна отримати класифікаційне нечітке дерево рішень (рис. 2).

Як видно з рис. 2, деякі результуючі рішення є однаковими (1 та 3). Через те, що дерева рішень будуються згори донизу та використовують жадібні стратегії розвитку (мається на увазі той факт, що алгоритм побудови дерева не може повернутися на попередні кроки та переписати їх), їх прийнято вважати ефективними, але не оптимальними. На рис. 3 не показано, як відбувається розбиття атрибутів на основі нечітких множин, оскільки це займає багато місця та потребує додаткових роз'яснень. У літературі можна зустріти нечіткі дерева рішень, які виглядають дещо інакше, ніж показано у цій статті.

Цей метод можна використовувати для виведення та редукції правил у базі знань. Редукції правил можна досягти при відсіканні неважливих піддерев дерева рішень, оптимізації процесу розбиття, розмірів дерева. Важливо розуміти, що якнайменше дерево гарантує найменшу кількість правил, проте не слід забувати, що у такому разі можна втратити семантично важливі, нетривіальні та практично корисні правила. Цей метод є неефективним для адаптації моделі правил до нового середовища, оскільки у такому разі доведеться перебудовувати/перенавчати дерево.

Сьогодні популярний метод пошуку асоціативних правил (асоціативний аналіз), який також дає змогу знаходити залежності у формі правил. Такі правила відрізняються від класичних продукційних правил, є за своєю суттю описовими та розглядаються не як складова баз знань, а скоріше відокремлено (ізольовано), локально.

Розпочнемо огляд асоціативних правил із простих ("булевих") асоціативних правил. Нехай існує дві множини бінарних характеристик, або дві події X та Y. Тоді під правилом "Якщо X, тоді Y" (альтернативний запис $X \Rightarrow Y$) мається на увазі наступне:

- об'єкт, який володіє характеристиками з X, скоріше за все володіє характеристиками з Y;
- якщо настала (відбулася) подія X, тоді настане (відбудеться) і подія Y;
- поява X асоціюється з появою Y.

Для того, щоб вирішувати, чи знайдене правило є цікавим, практично корисним, нетривіальним, вводяться якісні характеристики для правила: його підтримка (кількість об'єктів, фактів у базі даних чи знань, які задовольняють *X* та *Y* водночас) та впевненість (частотність появи об'єктів, фактів, що задовольняють *Y*, серед таких, що задовольняють *X*). Тоді необхідно віднайти усі асоціації, що відповідають (≥) заданим підтримці та впевненості.

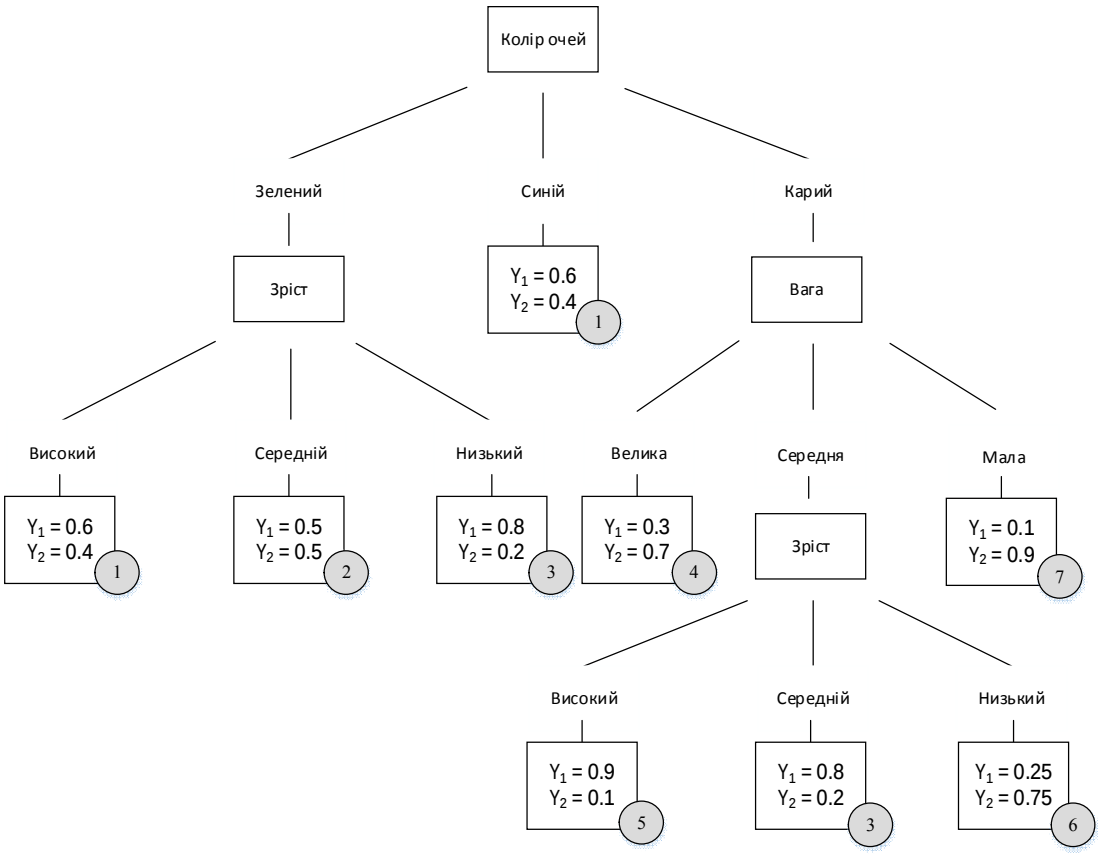


Рис. 2. Приклад нечіткого дерева рішень

Асоціативні правила найчастіше застосовуються для аналізу транзакційних баз даних (баз даних, що містять транзакції ринкових покупок), оскільки були винайдені та запропоновані для розв'язання задачі ринкового кошика. У даному контексті об'єктом є покупка, чиї характеристики представлені продуктами/речами. Асоціація вигляду {Молоко, Хліб} ⇒ {Масло} повідомляє, що якщо покупець купує молоко та хліб, то він також придбає і масло (підтримка та впевненість такого правила розраховуються окремо на основі певної вибірки даних).

Використання апарату нечітких множин та нечіткої логіки до процесу генерування асоціативних правил є дещо схожим до використання такого апарату відносно нечітких дерев рішень. Тобто дозволяється вироблення нечітких границь, нечітких інтервалів значень, лінгвістичних змінних, які вкупі допомагають уникати чітких порогових значень як у самих правилах, так і при якісному оцінюванні правила через його підтримку та впевненість. Крім того, лінгвістичні терми надають асоціативним правилам простої інтерпретації.

Нечіткий варіант булевого асоціативного правила міг би мати такий формальний вигляд: {Молоко/0.6, Хліб/0.65} ⇒ {Масло/0.7}, де після назви харчового продукту вказується його нечітка міра належності до цього правила.

Пошук простих булевих асоціативних правил є частковим випадком пошуку узагальнених асоціативних правил (Generalized Association Rules) [12], які в цій статті не розглядаються, та числових асоціативних правил (Quantitative Association Rules), оскільки при такому пошуку сама задача є спрощеною. Можна вважати, що все зводилося до присутності того чи іншого елемента у

транзакції і зазначалося це значеннями із [0, 1] у звичайних правилах, та значеннями у межах {0, 1} у нечітких аналогах. Проте бази даних та знань оперують не лише булевими, а й числовими та категорійними даними. Саме для цього було розроблено числові асоціативні правила.

У числових асоціативних правилах антецеденти та консеквенти правила, як правило, містять інтервальні значення. Прикладами таких правил є:

- якщо людина купує від 7 до 15 пляшок пива, то вона також купить від 4 до 10 пачок чіпсів;
- працівники зі стажем роботи від 5 до 10 років отримують надбавки до зарплати у розмірах від 1000 до 2000 у. о.

На основі вищенаведених числових правил можна приблизно сформулювати їхні нечіткі аналоги:

- купуючи немало пляшок пива, людина також купляє багато пачок чіпсів;
- багаторічні працівники отримують солідні надбавки до зарплати.

Відповідно апарат нечіткості може бути накладений і на інші типи асоціативних правил. Про це та інші подробиці щодо нечітких асоціативних правил можна детальніше дізнатися у [13, 14].

Особливо важливими при задачі виведення чи адаптації моделі правил є гібридні методи, які поєднують теорію нечітких множин з іншими технологіями, серед яких еволюційні і генетичні алгоритми та нейронні мережі. Для прикладу, еволюційні алгоритми можуть використовуватися для оптимізації, налаштування (корегування) чи навчання нечітких баз знань або для систематичнішого пошуку у просторі бази знань.

Досить цікавими є також нейронечіткі технології. Можна закодувати нечітку систему правил у вигляді нейронної мережі та застосувати стандартні методи та алгоритми таких мереж, щоб навчити (скорегувати) цю мережу. Як правило, просту типову нейронечітку систему можна розглядати як спеціальну 3- або 5-шарову нейронну мережу (рис. 3).

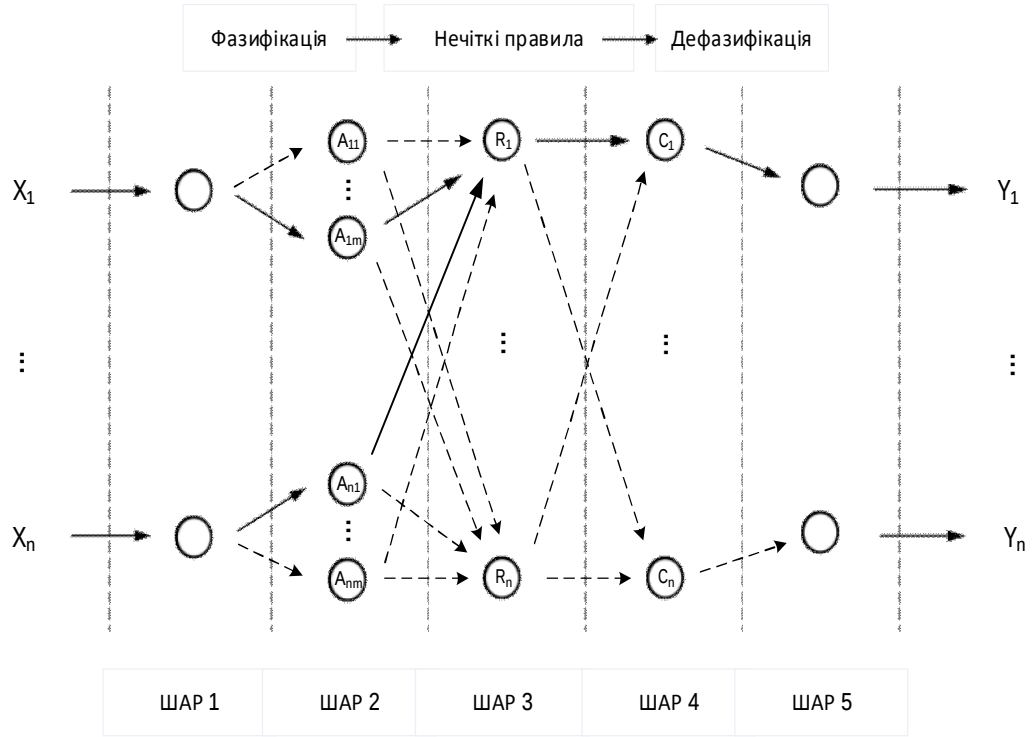


Рис. 3. Приклад 5-шарової нейронечіткої мережі

При 5-шаровій схемі кожен шар відповідає за певні функції:

1) шар вхідних даних;
2) шар фазифікації (антецедентів);
3) шар нечітких правил, який може складатися з *AND* та *OR* нейронів;
4) шар дефазифікації (консеквентів);
5) вихідний шар.

Кожен нейрон шару фазифікації являє собою вхідну функцію належності антецедента нечіткого правила та сам антецедент. Натомість кожен нейрон вихідного шару являє собою функцію приналежності консеквента правила, яка може бути реалізована у вигляді активаційної функції, та сам консеквент.

Нечіткі множини при цьому кодуються у вигляді (нечітких) вагових зв'язків. Зазвичай після навчання таку мережу можна трактувати як систему нечітких правил.

Варто зазначити, що зображення елементів нейронечіткої мережі на рис. 3 можуть відрізнятися у іншій науковій літературі. Крім того, існують різні типи нейронечітких мереж, тому для кожної з них є своє схематичне умовне позначення.

Здатність нейронних мереж робити узагальнення на основі існуючого набору навчальних даних, використовуючи навчальний алгоритм, є їхньою найціннішою властивістю. На основі цього можливі такі шляхи застосування таких мереж при використанні їх у нечітких системах та при роботі з базами нечітких знань [10]:

- навчання чи корегування функцій приналежності;
- виведення (визначення) та удосконалення нечітких правил;
- оптимізація кількості продукційних правил;
- здійснення нечіткого логічного виведення.

До цього можна також додати постійну адаптацію до змінного зовнішнього середовища (нові вхідні дані, змінені існуючі залежності тощо), що потребує постійного процесу зміни значень ваг, структурних зв'язків (топології мережі) та відповідних реакцій мережі. Такі реакції можна подавати у вигляді зворотних зв'язків. На рис. 3 зворотних зв'язків не показано.

Недоліком простих нейронних мереж є складність аналізу процесу виведення та отримання результатів і кінцевих правил мережею. Тобто вони не здатні пояснити користувачу, чому було отримано саме такі правила, а не інакші. Такий недолік усувається при використанні нейронечітких технологій. Про різні типи нейронечітких систем (паралельні, конкуруючі, гібридні), характер навчання, способи відображення нечітких множин у структурі мережі можна дізнатися у [15].

Суміжними до нейронечітких технологій щодо використання та застосування є еволюційні нечіткі технології, які, як було зазначено вище, є гібридними технологіями, в основу яких покладено нечітку логіку та еволюційні алгоритми. Найпоширенішими серед еволюційних алгоритмів є генетичні алгоритми. Вони являють собою пошуково-оптимізаційні технології і побудовані на формальних механізмах природної генетики. Крім того, суттєвою перевагою генетичних алгоритмів є легкість, з якою вони можуть бути розпаралелені виконавчою системою. Як правило, такі алгоритми характеризуються такими ознаками [16]:

1. Закодованою схемою для кожного можливого вирішення поставленої проблеми. Схема являє собою набір бітових стрічок, якщо йдеться про дискретні генетичні алгоритми, або набір числових значень, якщо розглядаються неперервні генетичні алгоритми або еволюційні стратегії. У будь-якому випадку такий набір є хромосомою, утвореною з генів (бітів чи чисел).

2. Фітнес-функцією (функцією пристосовуваності, адаптації), яка визначає якість кожного вирішення, яке входить у множину вирішень (популяцію).

3. Початковою множиною вирішень (початковою популяцією), отриманою випадково або на основі певних апріорних знань.

4. Множиною генетичних операторів, які допомагають утворювати нову популяцію рішень на основі існуючої. До таких операторів належать мутація, схрещення (кросовер, кросинговер), розмноження (репродуктивність, вибірка).

5. Термінальною умовою, що визначає кінець генетичного процесу.

Дискретні та неперервні генетичні алгоритми оперують або фіксованими за довжиною бінарними рядками (наприклад, "001110"), або ж числовими *n*-елементними рядками-векторами (наприклад, <12 10 6 2>). Таке обмеження можна обійти завдяки використанню еволюційних програм, які повністю грунтуються на генетичних алгоритмах, проте дають змогу кодувати рішення будь-якими структурами даних (часто використовуються дерева та І-АБО графи) та застосовувати

будь-який набір генетичних операторів. Згідно із [17] еволюційні програми можна сприймати як узагальнення генетичних алгоритмів.

Генетичні нечіткі системи, які містять базу нечітких знань, використовують різноманітні генетичні та еволюційні технології для отримання нових, кращих, удосконалених правил, параметрів та знань. Такі системи мають можливість кодувати правила та знання (рядками бітів, чисел чи іншою структурою даних), на основі чого проводять подальший пошук, оптимізацію чи налаштування як набору правил, так і поодиноких правил, лінгвістичних змінних чи функцій приналежності. Приклад кодування бітовими стрічками та еволюції нечіткої бази знань подано у [16].

Генетичну нечітку систему можна подати так (рис. 4).

Рис. 4. Генетична нечітка система

Існують декілька розроблених варіантів (підходів) для кодування правил бази знань хромосомами [18], якщо стоять завдання оптимізації чи навчання бази знань:

- Пітсбургський підхід, коли хромосома визначає базу правил одночасно. При цьому створюється багато хромосом (версій набору правил), які еволюціонують та змагаються одночасно.
- Підхід, коли одна хромосома кодує одне правило. Тут розрізняють:
- Мічіганський підхід, коли генетичний алгоритм визначає нові правила, що замінять старі, через змагання між хромосомами під час єдиного загального еволюційного процесу.
- Підхід ітеративного навчання (IRL – Iterative Rule Learning), коли хромосоми змагаються на кожній ітерації алгоритму. При цьому вибирається найкраща хромосома (правило). У результаті глобальне рішення буде сформоване вибраними хромосомами.
- Генетичне кооперативно-конкурентне навчання (GCCL – Genetic Cooperative-Competitive Learning), коли хромосоми змагаються та співпрацюють одночасно.

У [19] автори пропонують використовувати достатньо новітні бактеріальні еволюційні алгоритми, які значно скорочують час пошуку та оптимізації баз знань. Бактеріальні алгоритми використовують не хромосоми, а бактерії, тому кожне можливе рішення формується саме у вигляді бактерії. Зауважимо, що для бактерії характерні дещо інші варіації генетичних операторів.

Поєднання нейронних мереж із нечіткою логікою та нечіткої логіки із еволюційними алгоритмами привносить потужний синергічний ефект, тобто таке об’єднання суттєво перевершує потужність простої суми технологій, узятих поодинці.

Серед розглянутих методів (та й взагалі серед відомих сьогодні) немає універсального. Усі вони чимось відрізняються, мають свої переваги та недоліки, можуть бути використані по-різному.

Тому необхідно враховувати конкретну специфіку задачі та предметної області, перш ніж застосовувати той чи інший метод, технологію.

Узагальнені відомості та підсумки щодо проаналізованих у цій роботі нечітких методів та технологій в контексті баз нечітких знань подано нижче (таблиця).

Нечіткі методи та технології видобування даних в контексті баз нечітких знань

№ з/п	Назва моделі	Основна суть та особливості	Ймовірні шляхи використання
1	2	3	4
1	Нечіткий кластерний аналіз	Формування нечітких кластерів на основі вхідного простору інформації для їхньої подальшої різноманітної інтерпретації. Кластери зазвичай перетинають вхідну інформацію у спільних областях.	• Трактування розпізнаних кластерів як нечітких правил чи їхніх частин • Трактування розпізнаних кластерів як функцій приналежності • Мінімізація (редукція) правил та їхньої кількості
2	Нечіткі дерева рішень	Побудова дерева рішень як процес виведення моделі правил для вирішення класифікаційних та регресійних задач. Використовують нечіткі критерії розбиття у внутрішніх вузлах.	• Виведення моделі правил • Мінімізація (редукція) правил та їхньої кількості шляхом відсікання неважливих піддерев, оптимізації процесу розбиття, оптимізації розмірів дерева
3	Нечіткі асоціативні правила	Знаходження асоціативних залежностей у формі правил. Нечіткість дозволяє вироблення нечітких границь та інтервалів значень, лінгвістичних змінних.	• Асоціативний аналіз вмісту бази знань
4	Нейронечіткі мережі	Синергізм нечіткої логіки та штучних нейронних мереж. Кодування нечіткої системи правил у вигляді мережі. Застосування алгоритмів навчання для її корегування.	• Навчання функцій приналежності • Виведення та оптимізація нечітких правил та їхньої кількості • Адаптація до змінного зовнішнього середовища
5	Еволюційні нечіткі системи	Синергізм нечіткої логіки та еволюційних алгоритмів. Являють собою оптимізаційно-пошукову технологію. Навчають базу знань завдяки генетичним операторам через еволюцію.	• Навчання, оптимізація та адаптація бази знань (набору правил, одиничних нечітких правил, функцій приналежності, лінгвістичних змінних) • Пошук у просторі бази знань

Висновки

Необхідність застосування нечіткої логіки та теорії нечітких множин у процесах видобування даних спричинена тим, що:

- нечіткість як невід'ємне явище притаманна для методів отримання та представлення даних та знань;
- управлінці, топ-керівники компаній, системи прийняття рішень часто мають справу з узагальненими розмитими концептами та лінгвістичними виразами, які за своєю природою є нечіткими.

Завдяки цим причинам методи видобування даних було розширено для врахування аспектів нечіткості на різних рівнях аналізу. Більшість сьогоднішніх робіт та пошуків спрямовано в бік нечітких баз та сховищ даних. Досліджено існуючі нечіткі методи видобування даних та показано, що вони придатні для використання при інтелектуальному аналізі баз нечітких знань для глибокого внутрішнього аналізу та корегування збережених знань, виведення правил (нових чи на основі

існуючих), мінімізації та оптимізації наявних правил або їхньої загальної кількості у базі, та адаптації бази правил до нового зовнішнього середовища.

Подальші дослідження стосуватимуться поняття інтерпретації отриманих результатів нечіткого інтелектуального аналізу даних.

1. Han J. Data Mining: Concepts and Techniques / Han, J., Kamber, M., Pei, J. – Elsevier Inc., 3rd Edition, 2012. – 740 p. 2. Kruse R. Data Mining with Fuzzy Methods: Status and Perspectives / Kruse, R., Nauck, D., Borgelt, C. // Proceedings of the EUFIT'99, Aachen, Germany, 1999. – P. 488–495. 3. Hu Y. Finding Fuzzy Classification Rules using Data Mining Techniques / Hu, Y., Chen, R., Tzeng, G. // Pattern Recognition Letters, Vol. 24, 2003. – P. 509–519. 4. Ishibuchi H. Pattern Classification with Linguistic Rules / Ishibuchi, H., Nojima, Y. // Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, Vol. 220, 2008. – P. 377–395. 5. Maimon O. Data Mining and Knowledge Discovery Handbook / Maimon, O., Rokach, L. – Springer, 2nd Edition, 2010. – 1306 p. 6. Tang Y. Web-based Fuzzy Neural Networks for Stock Prediction / Tang, Y., Xu, F., Wan, X., Zhang, Y. // Proceedings of the 2nd International workshop on Intelligent Systems Design and Application, Atlanta, USA, 2002. – P. 169–174. 7. Fernandez S. Matrix Inference in Fuzzy Decision Trees / Fernandez, S., Lopez, C. // Proceedings of the Joint 4th Conference of the European Society for Fuzzy Logic and Technology, Barcelona, Spain, 2005. – P. 979-985. 8. Beynon M. Utilizing Fuzzy Decision Trees in Decision Making / Beynon, M. // Encyclopedia of Data Warehousing and Mining, 2nd Edition, 2008. – P. 2024–2030. 9. Liu X. Extraction of Fuzzy Rules from Fuzzy Decision Trees: An Axiomatic Fuzzy Sets Approach / Liu, X., Feng, X., Pedrycz, W. // Data and Knowledge Engineering, Vol. 84, 2013. – P. 1-25. 10. Chi Z. Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition / Chi, Z., Pham, T., Yan, H. // Advances in Fuzzy Systems – Applications and Theory, Vol. 10, 1996. – 240 p. 11. Hullermeier E. Fuzzy Methods for Machine Learning and Data Mining: State of the Art and Prospects / Hullermeier, E. // Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, Vol. 220, 2008. – P. 357–377. 12. Agrawal R. Mining Generalized Association Rules / Agrawal, R., Srikant, R. // Proceedings of the 21st International Conference of Very Large Databases, Zurich, Switzerland, 1995. – P. 407–419. 13. Chen G. Overview of Fuzzy Associations Mining / Chen, G., Wei, Q., Kerre, E., Wets, G. // Proceedings of the 4th International Symposium on Advanced Intelligent Systems, Jeju, Korea, 2003. – P. 209–214. 14. Delgado M. Fuzzy Association Rules: General Model and Applications / Delgado, M., Marin, N., Sanchez, D., Vila, MA. // IEEE Transactions on Fuzzy Systems №11 (2), 2003. – P. 214–225. 15. Суботін С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень: Навчальний посібник / С. О. Суботін. – Запоріжжя: ЗНТУ, 2008. – 341 с. 16. Magdalena L. A Fuzzy Logic Controller with Learning through the Evolution of its Knowledge Base / Magdalena, L., Monasterio-Huelin, F. // International Journal of Approximate Reasoning, Vol. 16 (3-4), 1997. – P. 335–358. 17. Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs / Michalewicz, Z. – Springer-Verlag, 1992. – 387 p. 18. Cordon H. Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases / Cordon, H., Herrera, F., Hoffmann, F., Magdalena, L. // Advances in Fuzzy Systems – Applications and Theory, Vol. 19, 2001. – 489 p. 19. Drobics M. Selecting the Optimal Rule Set Using a Bacterial Evolutionary Algorithm / Drobics, M., Botzheim, J., Adlassnig, K. // Proceedings of the 5th EUSFLAT Conference, Ostrava, Czech Republic, 2007. – P. 361–366.