

СЕКЦІЯ 10 МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 004.451.53

Гавриленко О.В.

*кандидат фізико-математичних наук,
доцент кафедри автоматизованих систем обробки інформації та управління
Національного технічного університету України
«Київський політехнічний інститут»*

ОГЛЯД МЕТОДІВ ГІБРИДИЗАЦІЇ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ІНФОРМАЦІЇ

Стаття присвячена розгляду методів гібридизації інтелектуальної обробки інформації. Приведено архітектуру гібридного використання правил та прецедентів. Розглянуто гібридне поєднання технологій правил та нейроммереж в інтелектуальній системі, а саме гібридний підхід до побудови орієнтованих на знання систем підтримки прийняття рішень. Визначено, що орієнтовані на знання системи підтримки прийняття рішень здебільшого будуються для таких предметних областей, які характеризуються необхідністю актуалізації та розширення знань. Розглянуто особливості застосування технології дейтамайнінгу, що зумовлюють отримання додаткових конкурентних переваг фірмам та корпораціям, а отже і підвищення їх інвестиційної привабливості і капіталізації.

Ключові слова: інтелектуальний аналіз даних, технології штучного інтелекту, системи підтримки прийняття рішень, використання прецедентів.

Гавриленко Е.В. ОБЗОР МЕТОДОВ ГИБРИДИЗАЦИИ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ИНФОРМАЦИИ

Статья посвящена рассмотрению методов гибридной обработки информации. Приведена архитектура гибридного использования правил и прецедентов. Рассмотрено гибридное сочетание технологий правил и нейросетей в интеллектуальной системе, а именно гибридный подход к построению ориентированных на знание систем поддержки принятия решений. Определено, что ориентированные на знания системы поддержки принятия решений в основном строятся для таких предметных областей, которые характеризуются необходимостью актуализации и расширения знаний. Рассмотрены особенности применения технологии дейтамайнинга, которые обуславливают получение дополнительных конкурентных преимуществ фирмам и корпорациям, а следовательно и повышение их инвестиционной привлекательности и капитализации.

Ключевые слова: интеллектуальный анализ данных, технологии искусственного интеллекта, системы поддержки принятия решений, использование прецедентов.

Gavrylenko O.V. REVIEW BY HYBRIDIZATION INTELLIGENT PROCESSING OF INFORMATION

The article is devoted to the methods of hybridization intelligent processing of information. Powered hybrid architecture using rules and precedents. We consider a hybrid combination of technology rules and neural networks in intelligent systems, namely hybrid approach to building knowledge-oriented decision support systems. Determined that the focus on knowledge decision support system mainly built for these subject areas, which are characterized by the need of updating and expanding knowledge. Features of the application of technology Data Mining additional competitive advantage to firms and corporations, and thus increasing their investment attractiveness and capitalization.

Keywords: data mining, technology, artificial intelligence, decision support systems, the use of precedents.

Постановка проблеми. Більшість підприємств роками накопичують та записують величезні обсяги різномірної інформації щодо усіх аспектів своєї діяльності. Однак сподіватися, що ці масиви фактичних даних суттєво допоможуть підприємствам в ухваленні якісних стратегічних рішень без відповідних технологій обробки інформації, марно. Адже для того, щоб сформулювати коректні посилки для ухвалення ефективних управлінських рішень, потрібні об'єктивні знання про суть, зв'язки і закономірності досліджуваної предметної області. Відшукати необхідні факти в базі чи сховищі даних не так вже складно, проте у сучасному інформаційному суспільстві потрібні не тільки факти самі по собі, а особливо нові знання. Важливе значення знань для організації та подальше ефективне їх використання вже багато років визнається провідними теоретиками і практиками менеджменту, проте сама концепція управління знаннями вимагає переосмислення традиційної управлінської думки.

Аналіз останніх досліджень і публікацій. Аналіз низки наукових праць, зокрема [1-4], показує,

що для позначення цього напрямку найчастіше використовуються терміни «Data Mining» і «Knowledge Discovery», в україномовних публікаціях з даної тематики застосовується термін «дейтамайнінг» (ДМ) або «інтелектуальний аналіз даних» (ІАД). Порівняно недавно на перехресті декількох областей інформаційних систем і технологій сформувався і активно розвивається цей науковий напрям, метою якого є пошук корисних знань у корпоративних базах і сховищах даних.

Постановка завдання. Динамічність і невизначеність процесів в умовах глобальної економіки потребують розвитку та вдосконалення концептуальних положень та методів інтелектуальної обробки інформації.

Виклад основного матеріалу дослідження. Сьогодні найбільший потенціал і сподівання покладають на гібридні інтелектуальні системи, які є потужним інструментом рішення складних проблем, які не під силу «чистим» підходам. Гібридизація методів інтелектуальної обробки інформації – девіз, під яким пройшли 90-ті роки у європейських і американських дослідників. У результаті об'єднання декількох тех-

нологій штучного інтелекту з'явився спеціальний термін – «м'які обчислення» (soft computing), який ввів Л. Заде у 1994 році. Проте для ефективної інтеграції різних технологій в одній системі необхідне уніфіковане подання даних та знань [2; 3].

Розглянемо комбінування парадигм використання правил та прецедентів. Цей підхід дозволяє підвищити ефективність обробки виключень, не ускладнюючи набір правил. Таким чином, кожен із компонентів займається тим, на що найкраще він орієнтований: правила мають справу з узагальненнями предметної області, а прецеденти – з окремими нетиповими випадками. Ця технологія передбачає пошук прецедентів, які аналогічні поточному випадку, якщо він є виключенням із правил. Такий підхід вимагає, щоб база прецедентів була проіндексована по правилам, з яких вони є виключеннями. Алгоритм передбачає наступний порядок: спочатку для вирішення поточної проблеми застосовуються правила, в результаті чого формується деяке рішення. Потім проглядається бібліотека прецедентів щодо наявності виключень із відповідного правила (рис. 1) [2].

Цей підхід для побудови реальних орієнтованих на знання систем заслуговує на увагу, оскільки побудова набору правил для складної предметної області (наприклад, для хімічної промисловості) є складним завданням. Крім складностей з вилученням та поданням знань існує проблема повноти охоплення предметної області. Отже, в ідеалі база знань має бути коректною, повною та не містити протиріч. Але кількість правил все зростає, вони ускладнюються, і до того ж врахувати у правилах всі можливі виключення дуже складно.

Крім того, існують розробки щодо застосування прецедентів для навчання правилоорієнтованої системи підтримки прийняття рішень (СППР).

Розглянемо гібридне поєднання технологій правил та нейромереж у інтелектуальній системі. Гібридний підхід до побудови орієнтованих на знання СППР здатен вирішити важливу проблему навчання системи. Адже орієнтовані на знання СППР здебільшого будуються для тих предметних областей, які характеризуються необхідністю актуалізації та розширення знань. З огляду на це, орієнтована на знання СППР повинна містити блок, який дозволяє завантажувати у систему нові знання та редагувати старі знання. Усе більший інтерес викликає автоматизоване набуття знань (самонавчання), або створення нових понять чи правил на базі існуючих даних та знань, перевірку їх на несперечливість існуючим та включення до БЗ. Іншими словами, набуття знань 1-го роду на базі неповних знань (знань 2-го роду). Повністю автоматизувати цей процес занадто проблематично, адже нові знання потребують дуже ретельної перевірки експертами [2].

Розглянемо питання використання комбінації штучних нейромереж з технологією нечіткої логіки. Класичні нечіткі системи мають той недолік, що для формулювання логічних правил і функцій приналежності необхідно привертати об'єктивних експертів предметної області, що доволі складно забезпечити. Концепція нечітких нейромереж (або адаптивних нечітких систем) передбачає реалізацію логічного доведення на основі апарату нечіткої

логіки, але його необхідні складові (набір доречних лінгвістичних правил та оптимальна форма функцій приналежності) підбираються у процесі навчання нейромережі на експериментальних даних. Для цього модуль нечіткого управління представляється у формі багатошарової нейронної мережі, і застосовується алгоритм оберненого розповсюдження помилки. Нечітка нейронна мережа, як правило, складається з чотирьох шарів: шару фазифікації вхідних змінних, шару агрегації значень умови активації, шару агрегації нечітких правил і вихідного шару. Найбільше поширення у даний час набула архітектура нечітких нейромереж виду ANFIS і TSK, саме такі мережі є універсальними апроксиматорами. Такі переваги адаптивних нечітких систем, як можливість не тільки використовувати накопичені раніше дані та знання, але і можливість набувати нові знання, прозора інтерпретованість накопичених знань зробили сьогодні такі системи найбільш перспективними й ефективними інструментами «м'яких» обчислень.

Комбінація штучних нейромереж з технологією генетичних алгоритмів передбачає вибір входів для нейромережі у формі природного відбору та навчання нейромережі шляхом підбору адекватних параметрів на двох етапах: для пошуку загальних параметрів нейромережі (кількості прихованих шарів та нейронів); на заключному етапі навчання (для пошуку всіх значень ваг нейромережі та функцій активації). В основному генетичні алгоритми використовуються для покращення нейромереж, вже створених та навчених за допомогою будь-якого алгоритму.

Збір та зберігання знань експертів найкраще виконується орієнтованою на знання системою, але вона не пристосовується до змін негайно. Нейромережі, з іншого боку, не є дуже гарними зберігачами експертного досвіду, але їх можна тренувати для подовження навчання. Нейромережі можуть переглядати великі обсяги даних та знаходити причинно-наслідкові відношення, які допоможуть їм адаптуватися до змін у своєму середовищі. Зокрема, для подання гаусової функції приналежності може бути використаний нейрон з наступною функцією активації:

$$f(net) = \exp\left(\frac{-(net - m)^2}{\sigma^2}\right),$$

де net – сума добутоків входів нейрона;
 m – значення нечіткої змінної – максимуму функції приналежності;
 σ – середньоквадратичне відхилення функції приналежності від \max значення.

Нейромережевий підхід особливо ефективний у задачах експертної оцінки з тієї причини, що він сполучить у собі спроможність комп'ютера до опрацювання чисел і спроможність мозку до узагальнення і розпізнавання (наприклад, цей підхід варто використовувати у задачах оцінки активів нафтогазових компаній) [4].

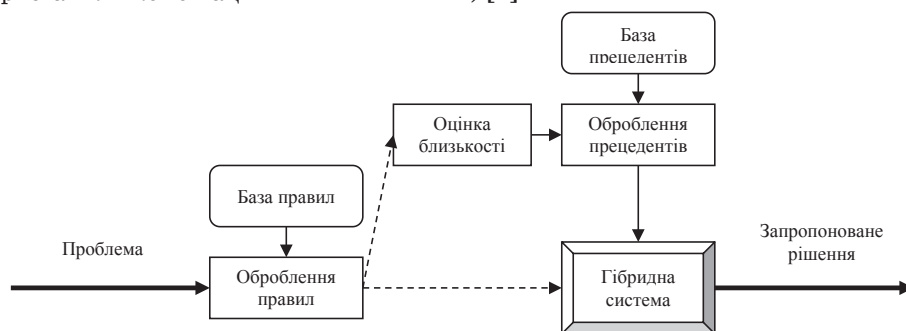


Рис. 1. Архітектура гібридного використання правил та прецедентів

Отже, подібно нейромережевим алгоритмам і системам з нечіткою логікою технологія експертних систем (що використовується у орієнтованій на знання системі) теж здатна до навчання (тобто до формування своєї бази знань) на основі прикладів із предметної області. Очевидним достоїнством таких орієнтованих на знання систем, як і нейромереж, є поліпшення характеристик системи за рахунок удосконалення структури і нарощування бази знань. Необхідно відзначити, однак, що, на відміну від нейромереж, експертні оболонки поки не роблять узагальнень ситуацій по параметрах і критеріям вибору дій, а роблять тільки порівняння набору значень аналізованих параметрів з базою прецедентів. Реалізація узагальнень у орієнтованих на знання системах можлива при використанні гібридного підходу до побудови орієнтованих на знання систем (використання елементів нечітких множин, нейромереж тощо) [4].

Як приклад можна запропонувати таку архітектуру нейронної орієнтованої на знання СППР, за якої один шар вузлів мережі є висхідними техніко-економічними умовами проекту з розробки нафтогазового родовища, інший шар є прецедентами результатів буріння, а ребра нейромережі є зваженими зв'язками.

Разом дві технології можуть забезпечити повноцінну підтримку у межах правило-орієнтованої системи, оперуючи як формальним знанням (орієнтована на знання технологія), так і неформальним (технологія нейромереж). Комбіноване використання експертної системи й апарату штучних нейронних мереж забезпечує необхідну гнучкість і самонавчання на основі знань, у той же час отримані від експертів знання дозволяють істотно спростити структуру нейронних мереж, зменшити число нейронів і зв'язків мережі.

Комбінація штучних нейромереж з технологією генетичних алгоритмів передбачає вибір входів для нейромережі у формі природного відбору та навчання нейромережі шляхом підбору адекватних параметрів на двох етапах: для пошуку загальних параметрів нейромережі (кількості прихованих шарів та нейронів); на заключному етапі навчання (для пошуку всіх значень ваг нейромережі та функцій активації). В основному генетичні алгоритми використовуються для покращення нейромереж, вже створених та навчених за допомогою будь-якого алгоритму.

Вплив нечіткої логіки на гібридизацію методів ДМ виявилось одним з найбільших. Подібно до того, як нечіткі множини розширили рамки класичної математичної теорії множин, нечітка логіка втрутилася практично в більшість звичайних методів ІАД, наділивши їх новою функціональністю. Нижче наводяться ще декілька найцікавіших прикладів таких об'єднань.

Нечіткі запити до баз даних дають можливість формулювати запити на природній мові (наприклад: «Вивести список недорогих пропозицій щодо оренди житла близько до центру міста»), що неможливо при використанні стандартного механізму запитів. Для цієї мети розроблена нечітка реляційна алгебра і спеціальні розширення мови SQL для нечітких запитів. Велика частина досліджень в цій області проведена західноєвропейськими вченими Д. Дюбуа і Р. Праде.

Нечіткі асоціативні правила – інструмент для витягання з баз даних закономірностей асоціації, які формулюються у вигляді лінгвістичних висловів. Передбачає використання спеціальних понять: нечіткої транзакції, підтримки і достовірності нечіткого асоціативного правила.

Нечіткі когнітивні карти (або нечіткі довірчі мережі) були запропоновані Б. Косьюко у 1986 р. і використо-

вуються для моделювання причинних взаємозв'язків, виявлених між об'єктами деякої предметної області. На відміну від звичайних, нечіткі когнітивні карти є нечітким орієнтованим графом, вузли якого є нечіткими множинами. Направлені ребра графа не тільки відображають причинно-наслідкові зв'язки між об'єктами, але і визначають ступінь впливу (вагу) зв'язуваних об'єктів. Активне використання нечітких когнітивних карт як засіб моделювання систем обумовлено можливістю наочного представлення аналізованої системи і легкістю інтерпретації причинно-наслідкових зв'язків між об'єктами. Основні проблеми цієї технології такі: процес побудови когнітивної карти не піддається формалізації; важливо довести, що побудована когнітивна карта адекватна реальній модельованій системі.

Нечіткі методи кластеризації, на відміну від чітких методів, дозволяють одному і тому ж об'єкту належати одночасно декільком кластерам, але з різним ступенем. Нечітка кластеризація у багатьох ситуаціях є більш «природною» для використання і розуміння користувачем, особливо для об'єктів, розташованих на межі кластерів. Найбільш поширеним нечітким методом кластеризації є алгоритм нечіткої самоорганізації c-means і його узагальнення у вигляді алгоритму Густафсона-Кесселя.

До сучасних засобів Business Intelligence відносяться: генератори звітів, засоби аналітичної обробки даних, засоби інтелектуального аналізу даних, BI-Platforms і Enterprise BI Suites (інтегровані засоби аналізу й обробки даних масштабу підприємства, які дозволяють здійснювати комплекс дій, пов'язаних з інтелектуальним аналізом даних, і включають інтегрований набір інструментів і засобів розробки).

Засоби ДМ як невід'ємна складова концепції Business Intelligence традиційно відносяться до дорогих програмних інструментів – ціна деяких з них доходить до декількох десятків тисяч доларів. Тому до недавнього часу основними споживачами цієї технології були банки, фінансові і страхові компанії, великі торгові підприємства, а основними завданнями, що вимагають застосування дейтамайнінгу, вважалися: оцінка кредитних і страхових ризиків, розробка маркетингової політики, тарифних планів і інших принципів роботи з клієнтами. Останніми роками ситуація зазнала певних змін: на ринку програмного забезпечення з'явилися відносно недорогі інструменти ДМ від декількох виробників, що зробило доступною цю технологію для підприємств малого і середнього бізнесу.

Наразі індустрія систем ДМ є галуззю, що бурхливо розвивається, і про приналежність своїх систем до даного класу програмних продуктів заявляють багато виробників. Крім того, оскільки сама область ДМ є мультидисциплінарною, це призводить до необхідності використання великої кількості методів і алгоритмів.

Виділяють п'ять загальних типів закономірностей (паттернів), що можуть бути отримані засобами дейтамайнінгу: класифікація – робить висновок щодо визначення характеристик конкретної групи (наприклад, споживачі, які були втрачені через дії конкурентів); кластеризація (групування) – ототожнює групи елементів, які використовують спільно зображуючий параметр сигналу даних (кластеризація відрізняється від класифікації, бо наперед визначені характеристики не приводиться); асоціація – ідентифікує зв'язки або відношення між подіями, які відбувалися колись (наприклад, зміст кошика відвідання магазину за покупкою); послідовність – подібна

до асоціації, крім того зв'язок ідентифікується за певні періоди часу (наприклад, повторний візит до супермаркету або фінансове планування продукту); прогнозування – оцінює майбутні значення, основані на взірцях всередині великого набору даних (наприклад, прогнозування попиту).

При виборі системи ДМ слід приділити увагу наступним моментам:

1. Система повинна автоматично виконувати тести, що визначають статистичну значущість моделі, що опрацьовується. Існує спостереження, що довільний зовнішній об'єкт, який генерує випадкові дані, може бути пояснений за допомогою певної моделі, якщо ситуація передбачає врахування достатньо великої кількості вільних параметрів. Проте така штучна модель не має ніякої семантичної та прогностичної суті. І в цьому випадку існує загроза «підгонки», яка часто є недоліком систем, заснованих на нейронних мережах.

2. Одержана в процесі ДМ модель повинна бути легко інтерпретована. Якщо ви не можете зрозуміти, які знання модель містить, як цільова змінна залежить від незалежних змінних, ви практично не здатні контролювати результати використання моделі. Практично неможливо буде розпізнати у знайдений моделі можливі суперечності.

3. Система ДМ повинна знаходити шаблони різноманітного типу. Перед початком дослідження ви не можете сказати точно, який саме тип відносин між змінними прихований у даних. Тому потрібно бути готовим до перебору в процесі ДМ різних видів залежностей, щоб не пропустити оптимальний варіант.

4. Ефективність та керованість процесом обробки даних посилиться, якщо використовувати універсальну ДМ систему. Така система має комплекс взаємодоповнюючих інструментів, які дозволяють користувачу аналізувати дані виходячи з різних умов. Поєднання методів ДМ дозволяє значно підвищити значущість одержуваних результатів і загальні характеристики системи ДМ.

5. Важливе значення має також час інтелектуального аналізу даних. Проте цей параметр складно точно оцінити, оскільки він дуже залежить від характеристик досліджуваних даних. Обмеження

експоненціального зростання числа гіпотез, що висувуються і вимагають перевірки, істотним чином зменшує час процесінгу даних.

6. Обрана ДМ система повинна легко застосовувати знайдений шаблон до нової порції даних. Якщо ця операція вимагає додаткової роботи, то така система вже має недолік.

7. Необхідною є візуалізація результатів дослідження даних. Це найбільш швидкий і зрозумілий шлях в оцінці точності і надійності виявлених шаблонів.

8. При виборі ДМ системи також важливий вигляд, вид формованих кінцевих звітів і можливість їх групового вивчення. Може знадобитися як для того, щоб провести презентацію результатів досліджень колегам, так і просто щоб мати графіки і висновки «на руках» для власного вжитку. Тому система повинна володіти додатковими механізмами для впорядкування і форматування звітів.

Висновки з проведеного дослідження. Отже, можна стверджувати, що у сучасних умовах розвитку глобальної економіки і у зв'язку з появою нових галузей економічної діяльності застосування технології дейтамайнінгу зумовлює отримання додаткових конкурентних переваг фірмами та корпораціями, а отже і підвищення їх інвестиційної привабливості і капіталізації. Саме інструментом реалізації такої стратегії і виступають засоби дейтамайнінгу.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Голосани В. А. и др. Интеллектуальные системы принятия решений в нештатных ситуациях с использованием информации о состоянии природной среды. – М. : Здиториал УРСС, 2001. – 304 с.
2. Ситник В.Ф., Краснюк М.Т. Интеллектуальный анализ данных (дейтамайнінг) : навч. посібник. – К. : КНЕУ, 2007. – 376 с.
3. Краснюк М.Т. Интернет-орієнтований підхід до створення орієнтованих на знання СППР як обов'язковий фактор підвищення економічної ефективності підприємств нафтогазової індустрії // Моделювання та інформаційні системи в економіці : міжвідом. наук. зб. Заснов. у 1965 р. Вип. 67 / Відп. ред. М.Г. Твердохліб. – К. : КНЕУ, 2002. – 247 с.
4. Люгер Дж.Ф. Искусственный интеллект. – М. : Изд. дом «Вильямс», 2005. – 864 с.