

ВІДГУК
офіційного опонента
на дисертаційну роботу Єгошкіна Данила Ігоровича
**«Розробка методів і алгоритмів автоматичної генерації та дослідження
структури нечіткої бази знань»**, представлену на здобуття ступеня
доктора філософії за спеціальністю 113 Прикладна математика

Актуальність дисертаційної роботи

Нечіткою базою знань називають сукупність нечітких правил «якщо - то», які описують взаємозв'язок між входами та виходами деякого об'єкта з використанням лінгвістичних термів. База правил є основним елементом нечіткої моделі, оскільки саме в ній міститься інформація щодо структури моделі. Нечітка база правил є основним компонентом і нечіткого класифікатора, який є актуальним способом розв'язання задач класифікації та використовує нечіткі множини і нечітку логіку в якості інструменту для подання знань про розв'язувану проблему. Одним із завдань проектування нечіткої бази знань є вибір правил з деякої, наперед визначеної, множини кандидатів. Правила-кандидати можуть бути сформовані експертом або отримані шляхом оброблення відповідних експериментальних даних. В ідеальному випадку нечітка база знань має бути і компактною, і адекватною. Досягти цього в реальних задачах неможливо, тому на практиці намагаються обрати базу знань з коректним балансом між цими суперечливими критеріями.

Отже, на практиці задача формування нечіткої бази знань зводиться до відповіді на питання про те, скільки правил має бути в базі та як сформувати антецеденти та консеквенти правил. Мінімальна кількість правил у нечіткому класифікаторі сприяє кращому розумінню його роботи, зменшує обчислювальні витрати на налаштування його параметрів. Однак на мінімальній кількості правил не завжди вдається отримати необхідну точність класифікації.

Дисертаційна робота Єгошкіна Д.І. присвячена розробці та обґрунтуванню методів і алгоритмів автоматичної генерації нечіткої бази знань для розв'язання задач класифікації, а також дослідженню її структури. Беручи до уваги все вищезгадане, можна стверджувати, що дисертаційна робота є актуальною, як з теоретичної так із практичної точок зору.

Загальна характеристика роботи та отриманих у ній результатів

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, переліку використаних джерел, що містить 137 найменувань на 19 сторінках та додатків на 9 сторінках. Загальний обсяг дисертації – 171 сторінки, обсяг основного тексту – 121 сторінка. Робота містить 29 рисунків та 17 таблиць.

У **вступі** обґрунтовані актуальність теми, мета та завдання роботи, наведені наукова новизна, практичне значення, особистий внесок дисертанта та апробація роботи. У **першому розділі** проведено аналіз літературних джерел, що охоплюють тематику дослідження, та обґрунтовано вибір напрямку. Досліджено сучасні моделі представлення знань, методи логічного виведення, підходи до налаштування параметрів нечітких моделей, а також різні підходи до редукції баз правил у системах логічного виведення.

У **другому розділі** роботи сформульовано постановку задачі класифікації та описано продукційну модель представлення знань, яка комбінує моделі Мамдані і Сугено та вдосконалена введенням функції вагового коефіцієнту для консеквенту правил. Розроблено алгоритм для автоматичного формування бази правил на основі навчальної вибірки для розв'язання задачі класифікації. Вдосконалено метод логічного виведення для вирішення конфліктів при прийнятті рішень, обрано метрики для оцінки якості побудованої нечіткої системи та результатів класифікації. Розроблено адаптивний алгоритм редукції бази правил із використанням методу дихотомії з динамічним кроком.

У **третьому розділі** розглядаються методи налаштування параметрів нечіткої продукційної моделі. Пропонується процедура автоматичного визначення границь термів за допомогою статистичних методів, що дозволяє зменшити вплив аномальних об'єктів на результат класифікації. В роботі пропонується автоматична перевірка повноти бази правил за допомогою логіки Хоара, методу резолюцій та програмного додатку Simplify, який автоматизує перевірку коректності бази правил на етапі навчання системи. Для оцінки якості результатів класифікації використано матрицю невідповідностей, що доповнює стандартні метрики accuracy, precision, recall та f1-score.

У **четвертому розділі** представлені результати розв'язання модельних задач класифікації за допомогою розроблених алгоритмів. Наведено порівняння отриманих результатів класифікації з результатами класичного алгоритму

нечіткої класифікації. Проведено аналіз впливу розмірів навчальної вибірки на якість класифікації.

Обґрунтованість і достовірність наукових положень і висновків дисертаційного дослідження

Наукові положення дисертації є достатньо обґрунтованими та відображеними у висновках. Основні результати, які були отримані здобувачем, відповідають меті та завданням дисертаційної роботи і добре висвітлені в опублікованих наукових працях. Ступінь вірогідності отриманих результатів забезпечується такими факторами:

- використання відомих та загальновизнаних положень теорії нечітких множин та нечіткої логіки;
- використання математичних моделей представлення знань та апробованих алгоритмів нечіткого логічного виведення;
- використання методів машинного навчання та нечіткої класифікації, логіки Хоара, методу резолюцій;
- коректністю формулювань задач, контрольованою якістю отриманих результатів класифікації, зіставленням отриманих результатів у часткових випадках із вже відомими розв'язками інших авторів.

Обґрунтованість наукових висновків, сформульованих автором дисертаційної роботи, підтверджується апробацією результатів роботи на міжнародних наукових конференціях.

Новизна одержаних теоретичних та/або експериментальних результатів досліджень

У рамках поставленого та вирішеного у дисертації завдання отримано такі нові наукові результати:

1. *Дістав подальшого розвитку* метод побудови нечіткої продукційної моделі для представлення знань про об'єкти предметної області на основі експериментальних даних.

2. *Запропоновано* вид нечіткого продукційного правила на основі моделей Мамдані та Такагі-Сугено-Канг для формалізації чітких та нечітких залежностей в даних, що підлягають аналізу.

3. *Удосконалено* методику побудови сукупності правил бази знань та алгоритм їх автоматичного формування.

4. *Дістав подальшого розвитку* метод налаштування значень параметрів нечіткої продукційної моделі, зокрема автоматизовано визначення границь термів, вагових коефіцієнтів правил, кількості правил.

5. *Вперше* для перевірки коректності автоматично побудованої бази правил експертної системи за критеріями повноти, мінімальності, зв'язності та несуперечності застосовано логіку Хоара, метод резолюцій та програмне забезпечення з автоматичного доведення правильності програм Simplify.

6. *Запропоновано* адаптивний алгоритм редукції бази правил із використанням методу дихотомії з динамічним кроком.

Рівень виконання поставленого наукового завдання, оволодіння здобувачем методологією наукової діяльності

Дисертаційна робота написана державною мовою, викладена грамотно. Кожен розділ дисертації відповідає сформульованому завданню, містить вичерпний виклад матеріалу та закінчується висновками. Це свідчить про достатній рівень оволодіння здобувачем методології проведення наукових досліджень. Поставлені в дисертаційній роботі наукові завдання дисертант реалізував на достатньому науково-методичному рівні з дотриманням методологічних вимог до наукової діяльності. Автор оволодів необхідними для рівня доктора філософії компетентностями.

Повнота викладу сформульованих у дослідженні наукових положень, висновків і результатів в опублікованих працях

Представлена дисертаційна робота є самостійним, завершеним науковим дослідженням, результати якого мають теоретичне та практичне значення.

Основні результати дисертаційного дослідження опубліковані в 13 наукових працях, з них одна стаття у провідному міжнародному журналі, що входить до наукометричної бази даних Scopus, 3 статті у наукових фахових виданнях України категорії Б з фізико-математичних наук, 9 тез доповідей у збірниках матеріалів міжнародних наукових конференцій.

Зауваження стосовно змісту та оформлення дисертаційної роботи

1. У розділі 1, де автором проведено аналіз моделей представлення знань, методів логічного виведення та ін., подано великий обсяг інформації, яка, на нашу думку, має більш довідковий характер (розділ 1 майже вдвічі більший за кількістю сторінок, ніж інші). У розділі немає формалізованих математичних аспектів розглянутих моделей та методів. Також, на нашу думку, доцільно було б узагальнити наведені переваги та недоліки розглянутих моделей та методів у

вигляді таблиці або діаграми, що, на наш погляд, спростило б сприйняття матеріалу та дозволило би більш чітко визначити авторську позицію з низки питань та щодо обґрунтування вибору напрямку дослідження.

2. У співвідношенні (2.1) на с. 76 потребує уточнення інформація щодо функції $f(k_1, \dots, k_L, p)$:

- по-перше, указано, що це «нормована» функція, але не указано, як саме;
- по-друге, указано, що вона є неперервною на відрізку $[a, b]$, але це функція від $(L+1)$ змінних.

3. У п. 2.2 на с. 73 потребує уточнення зміст такого речення: «У разі, якщо значення коефіцієнту кореляції наближається до 1, то це може означати, що обидві ознаки сильно впливають на цільову змінну.» Не указано між якими змінними розглядається коефіцієнт кореляції. Якщо між факторними ознаками, то твердження помилкове.

У наступному абзаці таке саме за змістом речення: «Якщо значення коефіцієнту кореляції наближається до -1, тоді обидві ознаки сильно впливають на цільову змінну, але у протилежних напрямках» є помилковим.

Взагалі застосування коефіцієнту кореляції Пірсона, яке пропонується у роботі для вибору ознак об'єктів експериментальної вибірки, вважаємо недоцільним, оскільки нечітке моделювання частіше застосовується для складних нелінійних залежностей, а не лінійних, до виявлення яких застосовується коефіцієнт кореляції Пірсона. Така недоцільність підтверджується далі у п.4.2.1 та п.4.2.3, де автор розраховує коефіцієнти кореляції, але ніяк не ураховує їх і нечіткі моделі будує за всіма наявними ознаками об'єктів експериментальних вибірок.

4. У п. 2.6 розроблений у роботі алгоритм розв'язання задачі класифікації описано дуже узагальнено. Наприклад, «Крок 2. Побудувати систему правил.», «Крок 6. Об'єкт неможливо класифікувати – необхідно створити новий клас.» потребують більш детального та математично формалізованого пояснення, що має виконуватися на цих кроках.

На жаль, у розділі 4, де показано роботу алгоритму на двох тестових наборах даних, теж не наведено результати, коли об'єкт неможливо класифікувати та створюється новий клас.

5. У п. 2.8, де описано розроблений адаптивний алгоритм редукції бази правил та указано, що для організації роботи алгоритму застосовується метод дихотомії з динамічним кроком, необхідно було б, на наш погляд, навести математичну постановку задачі оптимізації, для розв'язання якої застосовано метод дихотомії.

6. У п. 4.2.3 для ілюстрації роботи розроблених алгоритмів на тестовому наборі даних про арктичних пінгвінів застосовується термін «навчання нечіткої бази правил (до і після)». Чи є це синонімом терміну «налаштування параметрів нечіткої продукційної моделі», який застосовувався для тестового набору даних про іриси Фішера? Чому для тестового набору даних про арктичних пінгвінів не наведено результати редукції нечіткої бази правил?

7. У роботі немає єдиного підходу щодо позначення змінних:

- кількість ознак об'єкту вибірки позначається і як l , і як L ;
- кількість термів у терм-множинах лінгвістичних змінних позначається і як t , і як T , і як T_l ;
- не зрозумілі індекси елементів матриці A на с. 78: у третьому рядку повторюються індекси елементів A_{11} , A_{22} , A_{31} та ін.;
- на с. 95 у формулі (3.2) та на с. 97 у формулі (3.4) не пояснено, що означає змінна \bar{k}_l та як за її допомогою розраховуються параметри a_{lt} , b_{lt} , c_{lt} , d_{lt} функцій належності (3.1) для «відображення чітких вхідних значень ознак k_l в нечіткі множини» (с. 93).

Зауваження щодо рисунків та таблиць:

- рис. 3.2., с. 92 та рис. 4.5, с. 120 однакові;
- рис. 3.3, с. 98 та рис. 4.2, с. 115 теж однакові, крім того на них представлені функції належності термів лінгвістичних змінних, а у назві рисунків указано, що це «Границі термів...», теж саме і для рис. 4.4, с. 119;
- табл. 4.10, с. 128 та табл. 4.11, с. 129 з автоматично сформованими правилами відповідно розробленому алгоритму краще було б, на нашу думку, подати у вигляді задекларованих у роботі «матричного представлення антецедентів та векторного представлення консеквентів правил».

Окрім того, у роботі є технічні неточності, помилки, стилістичні огріхи:

- по тексту роботи багато формул знаходяться не в рядку тексту, а вище, що суттєво ускладнює їх розуміння;
- на с. 79 запис нечіткої множини \tilde{a}_{pl} містить зайві символи dk ;
- на с. 80 запис нечіткої множини \tilde{C} містить зайві символи dC , не пояснено, що позначено як \bar{C} , \underline{C} , $\mu_{\tilde{C}}(C)$;
- на с. 80 речення «Для дефазифікації \tilde{C} застосовується алгоритм Мамдані...» є помилковим;
- у тексті роботи зустрічаються не зовсім вдалі словосполучення, наприклад, на с. 2 «...опис моделей для погано формалізованих об'єктів», «...побудови

складної і кошовної математичної моделі», на с. 63 «...ступінь належності об'єктів навчальної вибірки правилу».

Відеутність порушень академічної доброчесності

Порушень академічної доброчесності в дисертаційній роботі Єгошкіна Д.І. та його наукових публікаціях не виявлено.

Заклучна оцінка дисертаційної роботи

На основі вищевикладеного вважаю, що дисертаційна робота Єгошкіна Данила Ігоровича є завершеним науковим дослідженням. Робота містить нові науково обґрунтовані результати, які у сукупності є важливими для розвитку методів нечіткого логічного виведення, а також прикладних застосувань для розв'язання практичних задач класифікації у будь-якій сфері людської діяльності. За актуальністю, науковою новизною, теоретичним та практичним значенням дисертаційна робота «Розробка методів і алгоритмів автоматичної генерації та дослідження структури нечіткої бази знань» повністю відповідає вимогам Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого постановою Кабінету Міністрів України № 44 від 12 січня 2022 року зі змінами, а її автор, Єгошкін Данило Ігорович, заслуговує присудження йому наукового ступеня доктора філософії в галузі знань 11 Математика та статистика за спеціальністю 113 Прикладна математика.

Офіційний опонент

доктор фізико-математичних наук, професор,
професор кафедри математичного
моделювання та статистики
Київського національного економічного
університету імені Вадима Гетьмана

Ольга ПРИТОМАНОВА

