

РЕЦЕНЗІЯ

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

Актуальність теми дисертації. Сьогоденню властиво широке використання апарату нечіткої логіки і теорії нечітких множин в різних галузях природничих і гуманітарних наук, що дозволяє формалізувати якісні поняття і судження для більш повного опису знань фахівцями. Складність вибору адекватного способу представлення накопиченого досвіду в таких галузях як біологія, медицина, соціологія, криміналістика, лінгвістика та багатьох інших – це відомий факт. Не завжди такі знання можуть бути перетворені у кількісні категорії, що дозволить застосувати до них мову логіки предикатів. Тому велика частина важливих знань фахівців, що носять якісний, інтуїтивний характер, часто залишається поза можливістю формального опису. Проблеми неточності, неповноти і невизначеності багатьох даних для таких задач призводить до використання методів теорії нечіткої логіки та системи нечіткого висновку. Одним з обмежень систем нечіткого висновку є ускладнення структури нечітких правил баз знань і зростання їх числа при зростанні кількості вхідних змінних.

Генерація бази несуперечливих нечітких продукційних правил є важливим і складним завданням, але не менш актуальним є завдання визначення найбільш ефективного алгоритму генерації. Ефективність цього процесу може бути підвищена шляхом автоматизованого аналізу існуючих моделей предметної області.

Ось чому розробка нових методів і алгоритмів автоматичної генерації нечітких баз правил та дослідження структури нечіткої бази знань залишається важливим і актуальним напрямком наукових досліджень.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконано у відповідності до планів наукових досліджень кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара в рамках наукових тем «Дослідження математичних моделей фізичних процесів методами ідентифікації та рекурентного аналізу із застосуванням інформаційних технологій» (№ держреєстрації 0119U101053, 2019-2021 р.р.), «Детерміновані та стохастичні алгоритми комп'ютерного моделювання об'єктів та процесів різної природи» (№ держреєстрації 0122U001467, 2022-2024 р.р.) при кафедрі комп'ютерних технологій у відповідності до тематичних планів науково-дослідних робіт Дніпровського національного університету імені Олеся Гончара.

Формулювання наукової задачі, нове вирішення якої одержане в дисертації. Наукова задача роботи полягає у розробці методів та алгоритмів автоматичної генерації бази знань системи логічного виведення для задач класифікації на основі експериментальних даних, вдосконалення методів логічного виведення, методів налаштування параметрів моделі представлення знань, методів доведення коректності бази знань та дослідження якості побудованої системи логічного виведення.

Об'єктом дослідження є нечіткі бази знань експертних систем.

Предметом дослідження є моделі, методи та алгоритми автоматичного формування бази знань системи логічного виведення для розв'язання задачі класифікації з використанням експериментальної інформації.

Основними завданнями дослідження є:

- створити нечітку продукційну модель, що об'єднує моделі Мамдані та Такагі-Сугено-Канг, для представлення знань про об'єкти предметної області з використанням навчальної вибірки та алгоритмів навчання;
- обрати спосіб представлення правил, розробити методику побудови сукупності правил бази знань та алгоритм їх автоматичного формування;
- розробити алгоритми ідентифікації значень параметрів нечіткої продукційної моделі;

- розробити підхід для перевірки коректності побудованої бази правил за критеріями повноти, мінімальності, зв'язності і несуперечності;
- вдосконалити метод логічного виведення із урахуванням того, що інформація про об'єкти предметної області може бути нечіткою та неповною;
- обрати метрики для оцінки якості автоматично побудованої нечіткої бази правил;
- розробити на основі побудованих алгоритмів програмне забезпечення для формування систем нечітких правил та реалізації механізму нечіткого логічного виведення;
- застосувати розроблені методи та алгоритми до розв'язання модельних задач.

Наукова новизна отриманих автором результатів. У дисертаційній роботі вирішено актуальну наукову проблему, що полягає у розробці нових і подальшому розвиненні існуючих методів та алгоритмів автоматичної генерації бази знань та дослідження структури нечіткої бази знань.

Основні наукові результати дисертації полягають у наступному.

- дістав подальшого розвитку метод побудови нечіткої продукційної моделі для представлення знань про об'єкти предметної області на основі навчальної вибірки за допомогою розроблених алгоритмів навчання;
- запропоновано вид нечіткого продукційного правила, що поєднує моделі Мамдані та Такагі-Сугено-Канг та містить у консеквенті правила в якості вагового коефіцієнту функцію, що визначає ступінь належності вихідної змінної до результуючої терм-множини;
- удосконалено методику побудови сукупності правил бази знань та алгоритм їх автоматичного формування за допомогою попереднього аналізу ознак навчальної вибірки, матричного представлення антецедентів та векторного представлення консеквентів правил;
- дістав подальшого розвитку метод логічного виведення з використанням бази знань нечіткої продукційної моделі для розв'язання задачі класифікації в умовах нечіткості та повноти інформації про об'єкти предметної області; за наявності конфлікту при прийнятті рішення, що полягає

у неспроможності системи однозначно визначити належність об'єкта певному класу, запропоновано процедуру порівняння відстані між об'єктами та класами;

- дістав подальшого розвитку метод ідентифікації та налаштування значень параметрів нечіткої продукційної моделі, зокрема автоматизовано визначення лінгвістичних змінних, терм-множин, границь термів, параметрів функцій належності, вагових коефіцієнтів правил, кількості правил;

- вперше для перевірки коректності автоматично побудованої бази правил за критеріями повноти, мінімальності, зв'язності і несуперечності застосовано логіку Хоара, метод резолюцій та програмне забезпечення Simplify;

- вперше запропоновано адаптивний алгоритм редукції бази правил із використанням методу дихотомії з динамічним кроком та інтегральної метрики, що враховує кількість активацій певного правила та ступінь належності об'єктів навчальної вибірки правилу. Множина правил поділяється на підмножину правил, що виконуються, та підмножину правил, що є надмірними для досягнення критерію якості;

- досліджено якість побудованої нечіткої продукційної моделі з використанням матриці невідповідностей та метрик, що визначають частину правильних відповідей моделі (accuracy), точність (precision), повноту (recall), f-міру (f1-score);

- створено нові ефективні алгоритми розв'язання вищеназваних задач;

- створено програмний продукт на об'єктно-орієнтованих мовах програмування C++, JavaScript та Python, що реалізує розроблені алгоритми;

- методи і алгоритми застосовано до розв'язання модельних задач класифікації.

Практичне значення одержаних результатів. Розроблені і розвинені в роботі існуючі методи й алгоритми можуть бути застосовані до розв'язання прикладних задач із різних предметних областей, що зводяться в математичній постановці до задач класифікації.

Показано, що класифікаційні нечіткі правила, які з'єднують міри значимостей причин і наслідків, представляють множину розв'язків системи нечітких логічних рівнянь.

Розроблений підхід та відповідне програмне забезпечення для автоматизації побудови системи логічного виведення, що базується на нечіткій продукційній моделі представлення знань, вдосконалює алгоритми логічного виведення та налаштування параметрів моделі задля отримання адекватних і точних розв'язків задач класифікації, сприяє покращенню точності та ефективності процесів обробки та аналізу даних.

Саме застосування нечіткої продукційної моделі є вдалим рішенням, завдяки її відносній простоті, наочності, високій модульності, легкості до внесення змін та доповнень, простоті схеми логічного виводу.

Деякі теоретичні результати дослідження, а також запрограмовані розроблені і розвинені в роботі алгоритми, включено до змісту дисципліни «Методи Machine Learning», що викладається на кафедрі комп'ютерних технологій факультету прикладної математики Дніпровського національного університету імені Олеся Гончара для студентів, що навчаються за другим (магістерським) рівнем вищої освіти за спеціальністю 113 «Прикладна математика», освітня програма «Інформатика». Результати, одержані при виконанні даного дослідження використовуються також при виконанні курсових та кваліфікаційних робіт студентами, які навчаються за спеціальністю 113 «Прикладна математика», що є Важливою ланкою практичної підготовки, коли під час виконання цих видів робіт поглиблюються та закріплюються теоретичні знання студентів з усіх дисциплін навчального плану, формуються навички наукової діяльності.

Обґрунтованість і достовірність наукових положень, висновків і рекомендацій, які захищаються. Дисертаційну роботу виконано на сучасному науковому рівні. Достатньо високий рівень обґрунтованості наукових положень, висновків, рекомендацій, сформульованих в дисертації, їх

достовірність забезпечені строгістю і коректністю математичних постановок задач в межах теорії нечіткої логіки і систем нечіткого виводу.

Особистий внесок здобувача. Усі результати дисертаційної роботи, що виносяться на захист, отримані автором особисто або за безпосередньої участі. Дисертаційна робота є завершеним дослідженням, виконаним автором самостійно відповідно до програми спланованих, проведених і узагальнених досліджень.

Дисертантом проведено пошук та аналіз літературних даних, розроблено і розвинено існуючі методи й алгоритми автоматичної генерації бази знань системи логічного виведення для задач класифікації на основі експериментальних даних, вдосконаленню методи логічного виведення, методи налаштування параметрів, методи доведення коректності бази знань та дослідженню якості побудованої системи логічного виведення. В роботі розроблено адаптивний алгоритм редукції бази правил, який дозволяє значно скоротити розмір бази правил без втрати якості результатів класифікації

Апробація матеріалів дисертації. Результати дисертаційної роботи доповідались і обговорювались на семінарі «Актуальні питання оптимізації та дискретної математики» при Науковій раді НАН України з проблеми «Кібернетика», який функціонує при Дніпровському національному університеті імені Олеся Гончара; на міжнародних конференціях «Прикладна математика та інформаційні технології ПМ&ІТ» (Чернівці, 2022 р.); «Математика та інформаційні технології» (Чернівці, 2023 р.); «Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті» (Дніпро, 2020 р.); «Молодь у світі сучасних технологій за тематикою: Використання інформаційних та комунікаційних технологій в сучасному цифровому суспільстві» (Херсон, 2020 р.); «Математичне та програмне забезпечення інтелектуальних систем» (Дніпро, 2018, 2021, 2022, 2023 рр.); «Сучасні науково-технічні дослідження у контексті мовного простору» (Дніпро, 2019 р.).

Мова і стиль роботи. Матеріал роботи викладений логічно, представлений достатньою мірою і написаний гарною українською мовою з використанням сучасної наукової термінології. Зміст дисертації відповідає вимогам, що висуваються до наукових праць такого рівня, а також вирізняється послідовністю, системністю та обґрунтованістю та висвітлює основні результати наукових досліджень.

Публікації. За темою дисертації опубліковано 13 наукових праць: 3 статті у наукових фахових виданнях України категорії Б з фізико-математичних наук, 1 - у науковому виданні, яке входить до міжнародних наукометричних баз даних, що індексується науково-метричною базою Scopus, 9 тез доповідей у збірниках матеріалів міжнародних наукових конференцій.

Структура та обсяг дисертації.

Дисертація складається із вступу, чотирьох розділів та висновків, переліку використаних джерел, що містить 137 найменувань на 19 сторінках та додатків на 9 сторінках. Загальний обсяг дисертації – 171 сторінка, обсяг основного тексту – 121 сторінка. У роботі наявні 29 рисунків та 17 таблиць.

У вступі обґрунтовані актуальність теми дисертації, мета та завдання роботи, наведені відомості про наукову новизну, практичне значення, особистий внесок дисертанта та відомості про апробацію роботи. В основу дисертаційної роботи покладено результати, що отримано автором під час виконання наукових досліджень в рамках наукових тем кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара.

У першому розділі було розглянуто літературні джерела, що стосуються тематики дослідження, та обґрунтовано вибір напрямку дослідження. Огляд літературних джерел було здійснено з метою аналізу сучасних моделей представлення знань, методів логічного виведення, підходів налаштування параметрів моделей, дослідженню коректності побудованих нечітких баз знань та підходів до аналізу їх якості. Розглянуто сучасні методи отримання знань, архітектуру компонентів системи нечіткого логічного виведення, різні підходи до розв'язання задач редукції бази правил в системах логічного виведення.

У другому розділі надано постановку задачі класифікації, описано продукційну модель представлення знань, що поєднує моделі Мамдані та Такагі-Сугено-Канг, та метод логічного виведення. Продукційну модель представлення знань вдосконалено введенням функції як вагового коефіцієнту для консеквенту правила. Із застосуванням кореляційного аналізу пропонується виявляти взаємозв'язків між змінними та приховані закономірності, патерни чи аномалії у наборах даних. Особлива увага приділяється обчисленню коефіцієнта кореляції Пірсона та виявленню мультиколінеарності як ключових аспектів аналізу великих обсягів навчальних даних. Додатково, запропоновано використання діаграм точкового розсіювання класів для візуалізації розподілу даних, що сприятиме більш глибокому розумінню їхньої суті та виявленню взаємозв'язків. Розроблено алгоритм автоматичного формування бази правил на основі навчальної вибірки з метою розв'язання задачі класифікації. Матриця антецедентів створюється за допомогою ознак об'єктів з деякої скінченної множини з використанням декартового добутку. Для формування вектору консеквентів використовується навчальна вибірка, що забезпечує можливість створювати нечітку базу правил та здійснювати налаштування моделі в автоматичному режимі. Вдосконалено метод логічного виведення в частині вирішення конфліктів при прийнятті рішень. Обрано метрики для оцінки якості побудованої нечіткої системи логічного виведення та отриманих результатів класифікації. Розроблено адаптивний алгоритм редукції бази правил із використанням методу дихотомії з динамічним кроком та запропонованої метрики.

У третьому розділі наведено методи налаштування параметрів нечіткої продукційної моделі. Запропоновано процедуру автоматичного визначення границь термів за допомогою статистичних методів з використанням вибіркового та міжквартильного середніх значень, що дозволило зменшити вплив наявних у навчальній вибірці аномальних об'єктів на результат класифікації та зробити процедуру визначення границь термів робастною. Здійснено аналіз коректності бази правил враховуючи її повноту,

несуперечливість, зв'язаність та мінімальність. В роботі запропоновано здійснювати автоматичну перевірку повноти бази правил з використанням логіки Хоара та предикату найслабшої передумови $Q \Rightarrow WP(S, R)$. Доведення проведено з використанням методу резолюцій та програмного додатку Simplify, що забезпечує автоматизацію перевірки коректності бази правил на етапі навчання системи. Для оцінки роботи експертної системи та її здатності забезпечити якість результатів класифікації, додатково до стандартних метрик accuracy, precision, recall та f1-score, запропоновано використання матриці невідповідностей.

У четвертому розділі наведено результати обчислювального експерименту з розв'язання модельних задач класифікації за допомогою запропонованої нечіткої моделі представлення знань на базі навчальної вибірки та системи нечіткого логічного виведення. Наведено порівняння результатів класифікації з використанням запропонованого в роботі підходу і класичного алгоритму нечіткої класифікації. Виконано чисельний аналіз впливу параметрів моделі та розмірів вибірки на якість класифікації. Також у розділі розглядається процедура редукції нечіткої бази правил для розв'язання задачі мультикласової класифікації з використанням критерію якості.

Висновки по роботі чіткі, логічні і витікають із теоретичних та експериментальних даних, отриманих автором.

Окремі дискусійні питання і зауваження.

До дисертаційної роботи є деякі зауваження та рекомендації:

1. Під час попередньої обробки даних обираються параметри об'єкту та з'ясовується ступінь їх залежності один від одного. Однак, кількість параметрів, що розглядається, є незначною, тому бажано збільшити їх кількість, ранжувати їх в порядку спадання та зробити аналіз, які з них є найбільш впливовими на результат класифікації.

2. Для налаштування границь термів в роботі пропонується застосовувати вибіркове та міжквартильне середні значення, але не наводиться порівняння з іншими підходами до вибору границь термів.

3. В разі, коли об'єкт належить кільком класам, утворюється конфліктна ситуація, для вирішення якої у роботі пропонується використовувати відстань між об'єктами на основі евклідової метрики. Але у роботі не пояснюється чому саме здійснено вибір зазначеної метрики, також бажано було б дослідити, в який ще спосіб можна обчислити відстань між об'єктом та іншими об'єктами класу.

4. Бажано було би розширити клас задач, до яких застосовується запропонована у роботі підходи до класифікації.

Вказані зауваження не знижують загальний рівень поданої дисертаційної роботи. Здобувачем отримані значні за науковою цінністю результати. Це дає можливість оцінити дисертаційну роботу як цілісну, закінчену та актуальну наукову працю.

Загальний висновок

Дисертаційна робота Єгошкіна Данила Ігоровича на тему «Розробка методів і алгоритмів автоматичної генерації та дослідження структури нечіткої бази знань» є актуальною за змістом, містить наукову новизну, основні результати та висновки дисертації обґрунтовані, мають теоретичне та практичне значення. Рецензована дисертаційна робота є завершеною науковою працею.

Оформлення дисертації відповідає вимогам, затверджених наказом Міністерства освіти і науки України № 40 від 12.01.2017 року. Методичний рівень, наукова новизна і практичне значення, рівень оприлюднення результатів рецензованої роботи відповідає вимогам до дисертацій ступеня доктора філософії згідно з вимогами Постанови Кабінету Міністрів України від 12 січня 2022 року №44 «Про затвердження Порядку присудження ступеня доктора

філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», а здобувач Єгошкін Данило Ігорович заслуговує на присудження ступеня доктора філософії в галузі знань 11 Математика та статистика за спеціальністю 113 Прикладна математика.

Рецензент

Завідувачка кафедри
обчислювальної математики
та математичної кібернетики
Дніпровського національного університету
імені Олеся Гончара
кандидат фіз.-мат. наук, доцент



Валентина ТУРЧИНА

Підпис канд. фіз.-мат. наук, доцента *Турчиної В. А. засвідчую*

Вчена секретарка
Дніпровського національного університету
імені Олеся Гончара
кандидат фіз.-мат. наук, доцент



Тетяна ХОДАНЕН