

РЕЦЕНЗІЯ

рецензента, д.ф.-м.н., професора Галузи Олексія Анатолійовича

на дисертаційну роботу Ямкового Кліма Сергійовича

«Інформаційні технології побудови композитних індикаторів

на основі методів машинного навчання»

подану на здобуття наукового ступеня доктора філософії

за спеціальністю 122 – Комп'ютерні науки

Детальний аналіз дисертаційної роботи Ямкового Кліма Сергійовича на тему «Інформаційні технології побудови композитних індикаторів на основі методів машинного навчання», що представлена для захисту на здобуття наукового ступеня доктора філософії у Національному технічному університеті «Харківський політехнічний інститут», дає змогу зробити комплексний висновок щодо її актуальності, ступеня обґрунтованості наукових положень, висновків, рекомендацій, достовірності та значущості отриманих результатів, наукової новизни, теоретичної та практичної цінності, надати загальну оцінку дисертації.

Актуальність теми

Однією з найскладніших проблем при побудові моделей композитних індикаторів є синтез їх нелінійних моделей, що відображають можливу складну структуру функцій переваг, які відображають уподобання експертів, оскільки прості лінійні моделі часто не забезпечують необхідної передбачуваної здатності. Таким чином, актуальною є задача розробки методів побудови нелінійних моделей композитних індикаторів обмеженої складності методами машинного навчання з узгодженням різномірної експертної та статистичної інформації, що і обумовлює вибір теми досліджень.

З практичної точки зору важливим є розробка та впровадження нових інформаційних технологій побудови композитних індикаторів методами машинного навчання.

Зв'язок роботи з науковими програмами, планами, темами

Зв'язок полягає в тому, що здобувач, як виконавець окремих розділів, брав участь у науково-дослідних роботах за договором № 99/332-2021 від 22.04.2021 р. (замовник ТОВ “Клауд Воркс”). Також результати роботи впроваджено у науково-дослідних роботах та навчальному процесі кафедри комп'ютерної математики і аналізу даних Національного технічного університету “Харківський політехнічний інститут”.

Наукова новизна одержаних результатів

Дисертація містить наукову новизну, з найбільш суттєвих доробок роботи можна назвати:

- вперше розроблено метод побудови композитних індикаторів з використанням нелінійної функції переваг, що відрізняється поєднанням ядерних методів машинного навчання з рідж-регресією;
- отримали подальший розвиток методи узгодження експертної і статистичної інформації, а саме вперше запропоновано метод оптимального узгодження з використанням регуляризуючого функціоналу, що містить експертні оцінки, що забезпечує підвищення точності отриманих моделей;
- вперше розроблено алгоритм двоетапної агрегації даних, який використовує ідею пошуку глобальних і локальних шаблонів у вибірці та регуляризовану кластеризацію;
- вперше запропоновано метод регуляризації при побудові нелінійних моделей індикаторів, який використовує немарковані експертами спостереження з використанням напівконтрольованого графового підходу у поєднанні з ядерними методами, що забезпечує можливість використання даних з обмеженою експертною інформацією;
- вдосконалена інформаційна технологія для побудови композитних індикаторів методами машинного навчання, що відрізняється комбінацією підходів.

Вважаю, що робота дисертанта є внеском в області побудови композитних індикаторів методами машинного навчання.

Практична цінність одержаних результатів та рекомендації щодо їх подальшого використання

Запропоновані методи та інформаційна технологія можуть бути повною мірою впроваджені в організаціях всіх галузей економіки для підвищення якості вибору альтернатив та автоматизації процесів побудови композитних індикаторів і ранжування.

Повнота викладення матеріалів дисертації в наукових працях, які опубліковані автором

Основні наукові та практичні результати досліджень опубліковані у 9 роботах, серед яких: 4 статті у наукових фахових виданнях України (1 у фаховому виданні, включеному до міжнародної наукометричної бази Scopus), 3 – у матеріалах конференції включених до наукометричної бази Scopus, 2 – у матеріалах науково-практичних конференцій.

Результати досліджень доповідались і були схвалені на: 13th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) (Дортмунд, Німеччина, 2023 p.), 12th International Conference on Dependable Systems, Services and Technologies (DESSERT) (Афіни, Греція, 2022 p.); XXXVI International Conference Problems of Decision Making Under Uncertainties (PDMU-2021) (Київ, Україна, 2021 p.); XIII Міжнародна науково-практична конференція магістрантів та аспірантів (Харків, Україна, 2019 p.); IEEE First International Conference on System Analysis & Intelligent Computing (SAIC) (Київ, Україна, 2018 p.).

Аналіз змісту дисертації. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих в дисертації

Робота Ямкового К. С. є завершеною науковою роботою, містить анотацію – українською та англійською мовами, вступ, чотири розділи, висновки, список використаних джерел і додатки.

У дисертаційній роботі вирішено науково-практичне завдання розробки методів та інформаційних технологій побудови композитних індикаторів на

основі ядерних методів машинного навчання та оптимального узгодження експертної та статистичної інформації.

Об'єкт дослідження – процеси побудови композитних індикаторів у задачах ранжування і багатокритеріального оцінювання та вибору.

Предмет дослідження – методи та інформаційні технології побудови композитних індикаторів з на основі оптимального узгодження експертної і статистичної інформації та агрегації даних.

Мета і задачі дослідження – розробка методів та інформаційних технологій побудови композитних індикаторів основі ядерних методів машинного навчання та оптимального узгодження експертної та статистичної інформації з метою підвищення точності отриманих моделей та обмеження їх складності.

У вступі обґрунтовано актуальність теми дисертаційної роботи, зазначено зв'язок роботи з науковими темами, сформульовано мету, задачі та цілі дослідження, визначено об'єкт, предмет та методи дослідження, показано наукову новизну та практичне значення отриманих результатів.

У першому розділі проведено аналіз задачі побудови композитних індикаторів та огляд різних підходів до їх побудови, зокрема, методів машинного навчання. Розглянуто метрики оцінки якості роботи алгоритмів ранжування, які використовуються для оцінювання запропонованих алгоритмів. Визначено особливості збору та використання експертно-статистичної інформації для побудови композитних індикаторів. Обґрунтовано вибір мети і задач роботи.

У другому розділі задача побудови композитного індикатора сформульована в термінах машинного навчання, отримано розв'язання задачі побудови нелінійної моделі композитного індикатору на основі ядерної рідж-регресії. Проаналізовано методи узгодження різнорідної експертної інформації, що дозволяють знайти компроміс між експертними оцінками композитних індикаторів та статистичними оцінками часткових показників. Обґрунтовано запропонований метод оптимального узгодження експертної і

статистичної інформації за допомогою регуляризації ядерної регресії з використанням апіорної інформації щодо важливості часткових показників, що суттєво підвищує точність отриманих моделей.

У третьому розділі запропоновано використовувати методи агрегування даних для зменшення складності ядерної моделі. Розглянуто методи групування та кластеризації для агрегації даних. Визначено проблему недостатньої розмітки даних, яка особливо часто виникає при збільшенні кількості даних. Для розв'язання цієї проблеми запропоновано використовувати методи напівкерованого навчання на основі графової регуляризації та ядерного трюку під час оптимізації нелінійної функції переваг. Для розв'язання зазначених проблем розроблено двоетапний алгоритм агрегації даних, який використовує як глобальні, так і локальні шаблони у структурі набору під час агрегації. Такий підхід дозволяє значно скоротити розмір вибірки при збереженні всіх властивостей та шаблонів.

У четвертому розділі наведено опис запропонованої інформаційної технології для побудови композитних індикаторів методами машинного навчання, яка реалізує розроблені у роботі методи та алгоритми. Розроблена інформаційна технологія реалізована у вигляді бібліотеки мовою програмування Python з відкритим похідним кодом та наслідуванням інтерфейсів бібліотеки scikit-learn. Перевірено працездатність розробленої інформаційної технології, точність запропонованих алгоритмів та проаналізовано отримані результати дослідження. Результати показали працездатність та ефективність запропонованих у роботі методів та алгоритмів.

У висновках наведено основні результати дисертаційної роботи щодо вирішення поставлених наукових задач дослідження.

Список літератури досить широко охоплює предметне поле дослідження, певною мірою відображає опрацювання автором значної кількості джерел.

Додаток містить інформацію про практичне впровадження результатів дисертації та приклади використання реалізованої інформаційної технології для вирішення задач в області діагностування виробничих процесів..

Достовірність отриманих результатів та висновків

Достовірність результатів, які представлені у дисертаційному дослідженні, підтверджується результатами відповідних експериментальних досліджень. А також обґрунтовується комплексним підходом у вивченні визначеного об'єкта та описам результатів у четвертому розділі.

Оформлення дисертації, дотримання вимог академічної доброчесності та повнота викладення наукових положень та результатів в опублікованих працях

Дисертація виконана з дотримання вимог академічної доброчесності, отримані результати дають підстави говорити про оригінальність роботи. У тексті містяться авторські ідеї, і не виявлено використання ідей інших науковців без посилання на їх роботи.

Основні ідеї автора та результати дослідження викладено у 4 статтях у наукових фахових виданнях України (1 у фаховому виданні, включеному до міжнародної наукометричної бази Scopus), 3 – у матеріалах конференції включених до наукометричної бази Scopus, 2 – у матеріалах науково-практичних конференцій.

Недоліки та зауваження до дисертаційної роботи

По дисертаційній роботі можна зробити наступні зауваження:

- Бажано більш чітко і конкретно формулювати актуальність та новизну дисертаційної роботи. Необхідно підкреслити новизну.
- У четвертій частині, запропоновані й існуючі методи узгодження порівняні лише за часом роботи алгоритмів. Не вистачає порівняння точності методів. А також варто описати сильні і слабкі сторони методів узгодження з прикладом задач, де який метод краще використовувати.
- У роботі зустрічаються стилістичні та друкарські помилки.

Вказані недоліки не впливають на загальну позитивну оцінку виконаної роботи. Дисертація є актуальною і має високу наукову цінність та практичну значущість.

Висновки

Дисертаційна робота Ямкового К. С. є завершеною науково-дослідною роботою, яка містить науково обґрунтовані результати, має наукову новизну та дає перспективи подальших досліджень. Тема дослідження відповідає галузі знань 12 – «Інформаційні технології» та спеціальності 122 – «Комп'ютерні науки».

Отже, враховуючи актуальність теми, отримані результати та певну практичну значущість вважаю, що дисертаційна робота Ямкового Кліма Сергійовича «Інформаційні технології побудови композитних індикаторів на основі методів машинного навчання» відповідає вимогам 6, 7, 8, 9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціальної вченої ради Закладу вищої освіти, наукової установи про присудження ступеня доктора філософії» від 12.01.2022 р. № 44 та вимогам до оформлення дисертації МОН України від 12.01.2017 № 40, а сам автор, Ямковий Клим Сергійович, заслуговує присудження йому наукового ступеня доктора філософії зі спеціальності 122 «Комп'ютерні науки».

Рецензент – доктор фізико-математичних наук,
професор кафедри комп'ютерної математики і аналізу даних
Національного технічного університету
«Харківський Політехнічний Інститут»

Олексій ГАЛУЗА

