ВІДГУК

офіційного опонента
Машталіра Сергія Володимировича
на дисертаційну роботу Ямкового Клима Сергійовича
«Інформаційні технології побудови композитних індикаторів
на основі методів машинного навчання»,
представлену на здобуття наукового ступеня доктора філософії
за спеціальністю 122 – Комп'ютерні науки

Актуальність теми

Розробка різного роду систем інформаційного пошуку нерозривно пов'язана з побудовою різного роду індикаторів, які є критерієм ранжування даних. При цьому одним з найцікавіших напрямів є побудова моделей композитних індикаторів. Однак однією з найскладніших проблем при побудові подібних моделей є синтез їх нелінійних моделей, що відображають складну структуру функції переваг, яка відображає думки експертів, оскільки прості лінійні моделі часто не забезпечують необхідної оціночної здатності. Таким чином, актуальним завданням є розробка методів побудови нелінійних моделей інтегральних індикаторів обмеженої складності з використанням методів машинного навчання з узгодженням різнорідної експертної та статистичної інформації, що визначає вибір теми дослідження.

З практичної точки зору важливе значення має розробка та впровадження нових інформаційних технологій побудови композитних показників за допомогою сучасними методами машинного навчання.

Дана дисертаційна робота моє зв'язки з науково-дослідної діяльністю кафедри комп'ютерної математики і аналізу даних, а саме здобувач, як виконавець окремих розділів, брав учать у науково-дослідних роботах за договором № 99/332-2021 від 22.04.2021 р. (замовник ТОВ "Клауд Воркс").

Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих в дисертаційній роботі.

Положення та висновки, що наведено в дисертаційній роботі Ямкового К. С. є достатньо обґрунтованими та логічно сформульованими. Їх основою є системність та послідовність у проведенні критичного аналізу методів та засобів побудови композитних індекаторів. Робота відзначається широким спектром використаних

сучасних методів машинного навчання. Сформульовані алгоритми стосуються важливих аспектів, відзначаються науковою новизною та свідчать про вагомий внесок автора у розвиток інформаційних технологій побудови композитних індикаторів. Апробація нових наукових результатів дисертації проводилася на міжнародних конференціях.

Достовірність результатів досліджень.

Достовірність результатів теоретичних досліджень підтверджується результатами відповідних експериментальних досліджень.

Наукові результати застосовані під час створення програмних засобів побудови моделей композитних індикаторів.

До основних нових наукових результатів дисертації слід віднести наступне:

- 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 патернів 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 поєднані з ядерними методами, що забезпечує можливість використання даних з обмеженою експертною інформацією;

Значимість отриманих результатів для науки і практичного використання.

Серед практичного значення отриманих результатів можна виділити наступні:

- запропоновано метод узгодження за допомогою регуляризації ядерної регресії. Це дозволяє поєднувати узгодження з нелінійною функцією переваг, що підвищує точність і зменшує використання ресурсів під час побудови композитних індикаторів;

- запропоновано використовувати методи агрегацію для вирішення проблем пов'язаних з великими обсягами даних. Агрегація дозволяє зменшити кількість даних, при чому зберегти властивості та значущі характеристики набору даних. Для підвищення точності та ефективності кластеризації запропоновано використовувати регуляризацію за допомогою цільової змінної;
- для розв'язання проблеми недостатньої розмітки даних запропоновано використовувати методи напівавтоматичного навчання на основі графової регуляризації та ядерного трюку під час оптимізації нелінійної функції переваг. Даний підхід змінює функцію переваг використовуючи доступні не розмічені спостереження;
- розроблено інформаційну технологію для побудови композитних індикаторів методами машинного навчання, яка містить усі запропоновані методи.

Повнота викладення результатів досліджень в опублікованих працях.

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 р.), 12th International Conference on Dependable Systems, Services and Technologies (DESSERT) (Афіни, Греція, 2022 р.); XXXVI International Conference Problems of Decision Making Under Uncertainties (PDMU-2021) (Київ, Україна, 2021 р.); XIII Міжнародна науково-практична конференція магістрантів та аспірантів (Харків, Україна, 2019 р.); IEEE First International Conference on System Analysis & Intelligent Computing (SAIC) (Київ, Україна, 2018 р.).

Оцінка змісту дисертаційної роботи

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.

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 апріорної 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 патерни 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.

Академічна доброчесність

Порушень академічної доброчесності в дисертації та наукових публікаціях, у яких висвітлені основні наукові результати дисертації, не виявлено.

Усі результати, які винесено автором на захист, отримані самостійно і містяться в опублікованих роботах. У роботах, опублікованих у співавторстві, використані тільки ті ідеї, положення та розрахунки, які є результатом особистих наукових пошуків.

По дисертаційній роботі можна зробити наступні зауваження:

- необхідно конкретніше формулювати актуальність та новизну дисертаційної роботи, підкреслити важливість та переваги нелінійних методів;
- у четвертому розділі запропоновані та існуючі методи узгодження мають порівняння лише за часом роботи алгоритмів. Не вистачає використання метрик для порівняння точності методів. До того ж виконане порівняння має лише табличний вигляд, що трохи ускладнює сприйняття інформації;

- у четвертому розділі під час опису набору даних не зазначено, що одна з ознак набору має текстове значення та не використовувалась в подальшому під час побудови моделей;
- 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.

ВИСНОВОК

Дисертаційна робота Ямкового К. С. «Інформаційні технології побудови композитних індикаторів на основі методів машинного навчання» за своїм змістом відповідає спеціальності 122 – Комп'ютерні науки. Дисертація є завершеною науково-дослідною роботою, яка розв'язує важливу науково-практичну задачу, яка полягає в створенні пристрою контролю і діагностування стану складних промислових об'єктів, який підвищує ефективність технічних процедур контролю та діагностування.

Подана дисертаційна робота «Інформаційні технології побудови композитних індикаторів на основі методів машинного навчання» Ямкового К. С. відповідає вимогам до дисертацій на здобуття наукового ступеня доктора філософії, а саме вимогам пунктів 6, 7, 8 і 9 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого Постановою КМУ від 12.01.2022 р. №44, а здобувач заслуговує присудження наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки.

Офіційний опонент

доктор технічних наук, професор кафедри інформатики Харківського національного університету радіоелектроніки — Сергій МАШТАЛІР

«25» листопада 2023 р.

Підпис засвідчую:

Учений секретар ХНУРЕ — Ігор МАГДАЛІНА

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ * ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ * №02071197 *