

## **ВІДГУК**

офіційного опонента

Недашківської Надії Іванівни

на дисертаційну роботу Ямкового Кліма 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**

При побудові моделей композитних індикаторів, як і в інших галузях, важливо використовувати нелінійні моделі, які повною мірою відображають думки експертів та складні зв'язки змінних, оскільки прості лінійні моделі не здатні забезпечити необхідної точності. Таким чином, актуальним завданням є розробка методів побудови нелінійних моделей композитних індикаторів з використанням методів машинного навчання, а саме моделей обмеженої складності з узгодженням різномірної експертної та статистичної інформації. Велике значення має розробка та впровадження сучасних інформаційних технологій побудови композитних, які використовують методи штучного інтелекту.

Результати дисертації впроваджено у науково-дослідних роботах та навчальному процесі кафедри комп'ютерної математики і аналізу даних (КМАД) Національного технічного університету «Харківський політехнічний інститут». Запропоновані методи та інформаційна технологія можуть бути впроваджені на підприємствах та організаціях різних галузей економіки з метою підвищення якості вибору альтернатив рішень та автоматизації процесів побудови композитних індикаторів і ранжування.

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

Положення та висновки, представлені в дисертаційній роботі Ямкового Кліма Сергійовича, є обґрунтованими, базуються на аналізі та систематизації наукових

джерел та використанні сучасних методів машинного навчання. Дослідження виконані з використанням інформаційних технологій. Результати роботи перевірено шляхом реалізації практичних експериментів, які підтверджують обґрунтованість наукових положень, а також висновків і рекомендацій, що були сформульовані в дисертаційній роботі.

#### **Достовірність результатів досліджень.**

Достовірність результатів теоретичних досліджень підтверджено результатами відповідних експериментальних досліджень.

Наукові результати застосовано під час створення програмних засобів побудови моделей композитних індикаторів. Результати впровадженні у науково-дослідній та освітній діяльності кафедри КМАД.

#### **Основні нові наукові результати дисертації наступні:**

- вперше запропоновано нелінійні моделі для побудови функції переваги, що відрізняються поєднанням ядерних методів машинного навчання з рідж-регресією;
- вперше запропоновано метод оптимального узгодження для нелінійних функцій переваги з використанням регуляризації, що містить експертні оцінки, таким чином забезпечується підвищення точності отриманих моделей;
- вперше розроблено алгоритм двоетапної агрегації даних, який використовує ідею пошуку глобальних і локальних шаблонів для підвищення репрезентативності агрегованого набору даних;
- вперше запропоновано метод регуляризації при побудові нелінійних моделей індикаторів з використанням напівконтрольованого графового підходу у поєднанні з ядерними методами, що забезпечує можливість використання даних з обмеженою експертною інформацією.

#### **Значимість отриманих результатів для науки і практичного використання:**

- розроблено метод узгодження, який базується на ядерній рідж-регресії з регуляризацією. Такий підхід дозволяє використовувати узгодження з нелінійною функцією переваг;
- запропоновано використовувати методи двоетапної агрегації для вирішення проблем пов'язаних з великими обсягами даних. Для підвищення точності та

ефективності кластеризації запропоновано використовувати регуляризацію у просторі цільових змінних;

- запропоновано використовувати методи напівконтрольованого навчання на основі регуляризації та ядерного трюку для вирішення проблеми недостатньої кількості розмічених даних;
- розроблено інформаційну технологію для побудови композитних індикаторів методами машинного навчання.

#### **Повнота викладення результатів досліджень в опублікованих працях.**

Результати досліджень доповідались і були схвалені на: 13th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS-2023) (Дортмунд, Німеччина, 2023 р.), 12th International Conference on Dependable Systems, Services and Technologies (DESSERT-2022) (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.).

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

#### **Оцінка змісту дисертаційної роботи**

Робота Ямкового К. С. є завершеною науковою роботою, складається з анотації, переліку позначень і скорочень, вступу, чотирьох розділів, висновків, списку використаних джерел і трьох додатків.

У вступі проаналізовано і обґрунтовано актуальність теми дисертації, вказано зв'язок роботи з науковими темами. Сформульовано мету, часткові цілі і задачі дослідження. Визначено об'єкт, предмет і методи дослідження. Наведено наукову новизну і практичне значення отриманих результатів.

У першому розділі роботи проведено аналіз завдання побудови комплексних

показників та огляд різних підходів до їх побудови, зокрема методів машинного навчання. Розглянуто метрики оцінки якості алгоритмів ранжирування, які використовуються для оцінки запропонованих алгоритмів. Визначено особливості збирання та використання експертної статистичної інформації для побудови комплексних показників. Обґрунтовано вибір мети та завдань роботи.

У другому розділі сформульовано задачу побудови складового індикатора в термінах машинного навчання та одержано розв'язання задачі побудови нелінійної моделі складового індикатора на основі ядерної рідж-регресії. Проаналізовано методи гармонізації розрізненої експертної інформації, що дозволяють знайти компроміс між експертними оцінками складових показників та статистичними оцінками часткових показників. Обґрунтовано запропонований метод оптимального зіставлення експертної та статистичної інформації за допомогою регуляризації ядерної регресії з використанням апріорної інформації про важливість часткових показників, що суттєво підвищує точність одержуваних моделей.

У третьому розділі запропоновано використовувати методи агрегування даних для зниження складності ядерної моделі. Розглянуто методи групування та кластеризації для агрегування даних. Виявлено проблему недостатнього маркування даних, що особливо часто виникає зі збільшенням обсягу даних. Для вирішення цієї проблеми запропоновано використовувати методи напівконтрольованого навчання, що базуються на регуляризації графа та трюку з ядром при оптимізації нелінійної функції переваги. Для вирішення цих проблем було розроблено двоетапний алгоритм агрегації даних, який при агрегації використовує як глобальні, так і локальні закономірності у структурі множини. Такий підхід дозволяє істотно зменшити розмір вибірки, зберігши всі властивості та закономірності.

У четвертому розділі представлено опис запропонованої інформаційної технології побудови комплексних показників з використанням методів машинного навчання, що реалізує розроблені в роботі методи та алгоритми. Розроблена інформаційна технологія реалізована у вигляді бібліотеки мовою програмування Python з відкритим вихідним кодом та імітацією інтерфейсів бібліотеки scikit-learn. Проаналізовано функціональні можливості розробленої інформаційної технології, точність запропонованих алгоритмів та отримані результати дослідження.

Результати показали працездатність та результативність запропонованих у роботі методів та алгоритмів.

Наприкінці представлені основні результати дисертаційної роботи з вирішення поставлених завдань наукового дослідження.

Список літератури містить 129 джерел і в достатній мірі відображає опрацювання автором значної кількості джерел.

Додатки містять 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 дисертаційній роботі і наукових публікаціях, у яких висвітлено основні наукові результати дисертації, не виявлено.

Всі результати, які винесено автором на захист, отримано Ямковим К.С. самостійно і містяться в опублікованих працях. У роботах, що опубліковані у співавторстві, використано тільки ідеї, положення і розрахунки, які є результатом особистих наукових пошуків здобувача.

### **Зауваження по дисертаційній роботі:**

- У першій частині недостатньо детально описано процес збору експертно-статистичної інформації. Цей процес важливий, бо має суттєвий вплив на подальші вибір та оптимізацію моделей для побудови функцій переваг. Як отримано експертні оцінки в таблицях 4.3, 4.6 (стор.75, 77)? Відомо, що метод безпосереднього оцінювання – дуже неточний.
- Під час демонстрації роботи методів узгодження не вказано яким чином формується вектор експертних переваг ознак.
- В практичному значенні одержаних результатів дисертації не вказано який у кількісному вираженні ефект забезпечують розроблені методи.
- Недостатньо описано результати порівняння розроблених методів з відомими методами, які вирішують аналогічні задачі.
- Бракує математичного опису функції втрат, яка враховує відносний порядок у кожній парі документів при попарному підході та порядок документів у

остаточному відранжованому списку (стор.17 і далі). Чи є ця функція неперервною відносно параметрів моделі? Не описано яким методом здійснюється оптимізація такої функції. Те саме стосується функцій втрат у списковому підході та LambdaMART / LambdaRank.

- На стор.31 вказано, що ваги критеріїв отримуються шляхом експертної оцінки за допомогою методу попарних порівнянь або статистичної оцінки за допомогою підходу PCA. Бажано було навести приклад розрахунку цих ваг на практиці.
- Чи є обмеження на вигляд модельних координатних функцій (стор.31)?
- Яка методика вибору параметра «ширина ядра» (стор.36)?
- Про яку порядкову рідж-регресію йде мова на стор.41?
- Недостатньо обгрунтовано чому пропонований метод узгодження за допомогою регуляризації ядерної регресії зменшує використання ресурсів під час побудови КІ (стор.41).
- Чому зменшується час на обробку у випадку пропонованої двоетапної схеми агрегування (стор.50)? Спектральна кластеризація та методи кластеризації на основі щільності, що пропонується використовувати на першому етапі двоетапної схеми (стор.50), потребують великої кількості часу, і витрати часу стрімко зростають із ростом розміру набору даних.
- Що мається на увазі під «зберігає глобальні та локальні патерни у вибірці» (стор.58)?
- Який метод кластеризації з регуляризацією за відстанню між композитними індикаторами використовувався (стор.82)?
- Варто конкретніше формулювати актуальність та новизну дисертаційної роботи.
- У четвертій частині процес попередньої обробки даних та вибору ознак описано без важливих деталей, що зменшує сприйняття та розуміння експерименту. Задача вибору ознак достатньо важлива, для її розв'язання існують підходи feature extraction і feature selection, які не висвітлено в роботі. Вибір множини ознак суттєво впливає на результат роботи моделі.

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 Кліма Сергійовича «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.

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 Ямковий К. С. заслуговує присудження йому наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки.

### Офіційний опонент

доктор технічних наук, доцент кафедри  
математичних методів системного аналізу  
Національного технічного університету  
України «Київський політехнічний  
інститут імені Ігоря Сікорського»

«22» листопада 2023 р.

Підпис д.т.н., доцента Недашківської Н. І. засвідчую:

«29» листопада 2023 р.



*V. D. Romanenko*