Журнал обчислювальної та прикладної математики **2025, № 2** Journal of Numerical & Applied Mathematics

УДК 004.9:007
MSC 90B50, 68T10

APPLICATION OF MULTI-CRITERION DECISION-MAKING METHODS FOR BOT CLASSIFICATION IN SOCIAL NETWORKS

Mykhailo Makhno, Oleksandr Borysenko

Faculty of Computer Science and Cybernetics,
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine,
E-mail: mykhailo.makhno@knu.ua, ORCID: 0000-0001-5045-1705
E-mail: borysenko@knu.ua, ORCID: 0009-0009-7852-3227

ЗАСТОСУВАННЯ МЕТОДІВ БАГАТОКРИТЕРІАЛЬНОГО АНАЛІЗУ ДЛЯ КЛАСИФІКАЦІЇ БОТІВ У СОЦІАЛЬНИХ МЕРЕЖАХ

Михайло Махно, Олександр Борисенко

Факультет комп'ютерних наук та кібернетики,
Київський національний університет імені Тараса Шевченка, Київ, Україна,
E-mail: mykhailo.makhno@knu.ua, ORCID: 0000-0001-5045-1705
E-mail: borysenko@knu.ua, ORCID: 0009-0009-7852-3227

Abstract. *The aim of the article* is to develop a methodology for classifying social network accounts into «bot», «non-bot», and «suspicious» categories using Multi-Criteria Decision-Making methods (MCDM).

Research methodology. The study employs a hybrid MCDM approach, combining the Analytic Hierarchy Process (AHP) and entropy method to determine feature weights, and the TOPSIS method for final classification. The criteria integrate behavioral, structural, attributive, and content-based features.

Results of the research. The proposed model was tested on a synthetic dataset of 100 accounts. It demonstrated high effectiveness, achieving 90% classification accuracy, with a precision of 0.85 and a recall of 0.89. The results confirm the model's ability to reliably detect bots while minimizing false classifications of genuine users.

Practical significance. The developed methodology provides a transparent, explainable, and adaptable tool for bot detection that can be integrated into social network monitoring systems, digital security tools, and information analytics platforms without the need for complete model retraining.

Keywords: social networks, bots, multi-criteria decision-making, TOPSIS.

Corresponding author: Олександр Борисенко (borysenko@knu.ua).

Анотація. *Метою статті* є розробка методології класифікації акаунтів у соціальних мережах на категорії «бот», «не бот» та «підозрілий» на основі методів багатокритеріального аналізу.

Методика дослідження. Дослідження ґрунтується на використанні комбінованого підходу до багатокритеріального аналізу: методу аналітичної ієрархії (AHP) та ентропійного методу для визначення ваг ознак, і методу TOPSIS для фінальної класифікації. Критерії включають поведінкові, структурні, атрибутивні та контентні ознаки.

Результати дослідження. Запропоновану модель було перевірено на синтетичній вибірці з 100 акаунтів. Модель продемонструвала високу ефективність: точність класифікації сягнула 90%, при цьому значення precision та recall склали 0.85 та 0.89 відповідно. Результати підтверджують здатність моделі до надійного виявлення ботів при мінімізації помилкових класифікацій справжніх користувачів.

Практична значимість. Розроблена методологія надає прозорий, пояснюваний та адаптивний інструмент для виявлення ботів, який може бути інтегрований у системи моніторингу соціальних мереж, інструменти цифрової безпеки та платформи інформаційної аналітики без необхідності повного перенавчання моделі.

Ключові слова: соціальні мережі, боти, багатокритеріальний аналіз, TOPSIS.

1. Вступ

У сучасному цифровому середовищі соціальні мережі стали не лише засобом спілкування, а й платформою для формування суспільної думки, політичного впливу та комерційної взаємодії. Однак із зростанням ролі цих платформ збільшується і загроза з боку автоматизованих акаунтів — ботів, які імітують людську поведінку з метою дезінформації, маніпуляцій або масової автоматизації дій [1]. Надійне виявлення таких ботів є критично важливим для забезпечення безпечного та достовірного інформаційного простору.

Значна частина сучасних досліджень зосереджена на використанні методів машинного навчання або аналізу часових патернів активності. Проте, для ефективної класифікації акаунтів необхідно враховувати множину різнотипних ознак — таких як частота постів, наявність аватару, дата реєстрації, топологія зв'язків, мовні особливості тощо. Кожна з цих характеристик може мати різний вплив на ймовірність того, що акаунт є ботом.

У цьому контексті доцільним є застосування методів багатокритеріального аналізу (БКА), які дозволяють формалізовано враховувати різні показники з відповідними ваговими коефіцієнтами. Такий підхід забезпечує прозору й інтерпретовану процедуру прийняття рішень, що є особливо важливим у галузях, де необхідна пояснюваність алгоритмів.

Ця стаття пропонує методологію класифікації акаунтів у соціальних мережах на основі БКА з поділом на три категорії: «бот» / «не бот» / «підозрілий». У роботі систематизовано релевантні ознаки та продемонстровано, як багатокритеріальний підхід може бути інтегрований у практичні системи моніторингу. Запропонований підхід є міждисциплінарним, поєднуючи інформатику, соціальну аналітику та методи прийняття рішень, і має потенційне застосування у сферах кібербезпеки, цифрового маркетингу та досліджень соціальних мереж.

2. Огляд літератури

Сучасні дослідження з виявлення ботів у соціальних мережах значною мірою зосереджені на методах машинного навчання та аналізі часових патернів активності [1, 2]. Однак, такі підходи часто є «чорними скринями» з низькою пояснюваністю результатів. Для ефективної класифікації необхідно враховувати множину різнотипних ознак — поведінкових, структурних, атрибутивних та контентних. Методи багатокритеріального аналізу (MCDM), такі як AHP [3] та TOPSIS [4, 7], дозволяють інтегрувати ці показники з відповідними ваговими коефіцієнтами, забезпечуючи прозору процедуру прийняття рішень.

Дослідження також показують, що боти значно впливають на поширення низькоякісного контенту [2, 6], що підкреслює важливість їх своєчасного виявлення.

3. Мета

Метою дослідження є розробка методики класифікації акаунтів у соціальних мережах на категорії «бот», «не бот» та «підозрілий» з використанням методів багатокритеріального аналізу для забезпечення високої точності та пояснюваності результатів.

4. Методика

Було розроблено модель класифікації на основі методів багатокритеріального аналізу (MCDM). Робота включала наступні ключові етапи:

1. Визначення релевантного набору ознак (критеріїв) для оцінки акаунтів.
2. Визначення вагових коефіцієнтів для кожного критерію.
3. Побудова математичної моделі для інтеграції ознак та їх ваг у єдину оцінку кожного акаунту за методом TOPSIS.
4. Встановлення порогових значень для фінальної класифікації.
5. Тестування моделі на синтетичних даних та оцінка її ефективності за стандартними метриками (accuracy, precision, recall, F1-score).

Дослідження, включаючи аналіз літератури, розробку методології, проведення експериментів та аналіз результатів, проводилося протягом трьох місяців.

Робота була теоретико-експериментальною. Для моделювання та аналізу даних використовувалося програмне забезпечення для математичного

та статистичного аналізу (Python з бібліотеками NumPy, Pandas, SciPy). Для аналізу структурних ознак, таких як мережева центральність були використані бібліотеки для аналізу графів, такі як NetworkX та igraph.

Математичний апарат дослідження ґрунтувався на методах багатокритеріального прийняття рішень (MCDM), зокрема:

- **Метод аналітичної ієрархії (AHP)** для визначення ваг критеріїв на основі експертних оцінок.
- **Метод TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)** для агрегації ознак та ранжування акаунтів.

Для первинної перевірки ефективності моделі було згенеровано синтетичний набір даних, що імітує 100 облікових записів (40 ботів та 60 не ботів). Кожен акаунт у цьому наборі даних представлений вектором значень за дев'ятьма критеріями, нормалізованими до діапазону [0, 1].

Оскільки на етапі дослідження використовувались синтетичні дані, вони були згенеровані штучно. Однак, методологія розроблена для роботи з реальними даними, які можна отримати через API соціальних мереж (наприклад, X (Twitter), Telegram, Facebook) або шляхом парсингу відкритих джерел. Збір даних для кожного акаунту мав би охоплювати різні аспекти. По-перше, важливо враховувати метадані профілю, такі як наявність аватара, дата реєстрації та інші базові характеристики. По-друге, слід аналізувати історію активності користувача, включно з частотою та часом публікацій, а також змістом повідомлень. Нарешті, важливо враховувати мережеву структуру акаунту — кількість підписників і підписок, взаємні зв'язки та загальну інтегрованість у соціальну мережу.

Аналіз проводився у кілька етапів:

1. **Нормалізація даних:** Усі значення критеріїв були нормалізовані за формулою векторної нормалізації:
$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}$$
де x_{ij} — значення j-го критерію для i-го акаунту.
2. **Зважування:** До нормалізованих значень застосовувалися ваги w_j, отримані методом AHP:
$$v_{ij} = w_j \cdot r_{ij}$$
3. **Агрегація за методом TOPSIS:** Для кожного акаунту обчислювалася відстань до ідеального (D_i^+) та антиідеального (D_i^-) рішення, а потім інтегральна оцінка:
$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}, \quad 0 \leq C_i \leq 1$$
4. **Класифікація:** Акаунти класифікувались на основі порогових значень C_i:
   - $C_i > 0.7 \Rightarrow$ "не бот"
   - $0.4 < C_i \leq 0.7 \Rightarrow$ "підозрілий"
   - $C_i \leq 0.4 \Rightarrow$ "бот"

5. **Оцінка якості:** На синтетичних даних була побудована матриця плутанини та обчислені метрики accuracy, precision, recall та F1-score.

В якості критеріїв було обрано 9 ознак, об'єднаних у чотири групи: поведінкові, атрибутивні, структурні та контентні. Їхні ваги, визначені комбінованим методом (AHP та ентропійна оцінка), представлені в Таблиці 1.

Табл. 1. Пріоритетність критеріїв

№	Критерій	Вага AHP	Вага ентропійна
1	Частота постів (на день)	0.18	0.21
2	Коефіцієнт асиметрії зв'язків	0.14	0.13
3	Вік акаунту	0.12	0.10
4	Наявність фото профілю	0.09	0.07
5	Щільність зв'язків (кластерний коеф.)	0.11	0.12
6	Частка дублікатів або однакових повідомлень	0.10	0.11
7	Активність у нічний час	0.07	0.06
8	Середня кількість посилань у постах	0.06	0.08
9	Центральність у графі	0.13	0.12

Для демонстрації роботи моделі на синтетичних даних трьох акаунтів були розраховані TOPSIS-оцінки та визначені їхні класи (Таблиця 2).

Табл. 2. Синтетичні дані

Користувач	**user_001**	**user_002**	**user_003**
Критерії (нормалізовані)			
Частота постів	0.95	0.30	0.55
Фото	0.00	1.00	1.00
Вік (днів)	0.05	0.90	0.35
Відношення підписок/підписники	0.92	0.35	0.60
Центральність у графі	0.10	0.65	0.43
Повтори	0.90	0.05	0.50
Посилання	0.80	0.15	0.50
Кластерний коеф.	0.15	0.72	0.42
Нічна активність	0.75	0.20	0.40
Результат			
TOPSIS score C_i	0.29	0.76	0.58
Клас	бот	не бот	підозрілий

Ефективність моделі оцінена на вибірці з 100 акаунтів (Таблиця 3). Результати показали високу точність (90%) та збалансованість між precision (0.85) і recall (0.89).

Метрики якості:

- **Accuracy** $= (34 + 56)/100 = \mathbf{0.90}$
- **Precision** $= 34/(34 + 6) = \mathbf{0.85}$
- **Recall** $= 34/(34 + 4) = \mathbf{0.89}$

Табл. 3. Результати класифікатора

	Бот (факт)	**Не бот (факт)**
Бот (модель)	34	6
Не бот (модель)	4	56

– **F1-score** $= 2 \cdot (0.85 \cdot 0.89)/(0.85 + 0.89) \approx \mathbf{0.87}$

5. Результати

Запропонована модель показала високу ефективність класифікації. На синтетичних даних досягнуто точність 90%. Метрики якості склали:

$$\text{precision} = 0.85, \quad \text{recall} = 0.89, \quad \text{F1-score} \approx 0.87.$$

Модель коректно ідентифікувала 34 з 40 ботів та 56 з 60 реальних користувачів.

6. Обговорення

Отримані результати демонструють, що комбінація AHP та TOPSIS є ефективним інструментом для виявлення ботів, перевершуючи за пояснюваністю традиційні методи машинного навчання.

Висока точність та збалансованість метрик свідчать про надійність моделі. Основною перевагою є прозорість прийняття рішень та можливість швидкої адаптації до різних платформ шляхом коригування ваг критеріїв.

Недоліком дослідження є використання синтетичних даних замість реальних датасетів, що може обмежувати застосовність моделі в реальних умовах. Також модель не враховує текстучі семантичні ознаки на повну.

Майбутні дослідження мають бути спрямовані на тестування моделі на реальних даних, включення лінгвістичних ознак та інтеграцію з машинним навчанням для автоматизації визначення ваг.

Практична цінність результату полягає у можливості його використання в системах моніторингу соціальних мереж та інструментах кібербезпеки.

7. Висновки

Метою роботи була розробка пояснюваного методу класифікації ботів у соціальних мережах. Запропонована гібридна модель на основі методів AHP та TOPSIS продемонструвала високу ефективність (точність 90%) на синтетичних даних.

Основним результатом є створення адаптивної та масштабованої методики, яка може бути використана для забезпечення інформаційної безпеки.

Автори заявляють про відсутність конфлікту інтересів щодо публікації цієї статті.

Література

1. Ferrara E., Varol O., Davis C., Menczer F., Flammini A. The rise of social bots. *Communications of the ACM*. 2016. Vol. 59, no. 7. P. 96–104. doi: 10.1145/2818717

2. Shao C., Ciampaglia G.L., Varol O., Yang K.C., Flammini A., Menczer F. The spread of low-credibility content by social bots. *Nature Communications*. 2018. Vol. 9, no. 1. P. 4787. doi: 10.1038/s41467-018-06930-7
3. Hwang C.L., Yoon K. Multiple Attribute Decision Making: Methods and Applications. Berlin, Germany: Springer-Verlag, 1981. doi: 10.1007/978-3-642-48318-9
4. Chen S.J., Hwang C.L. Fuzzy Multiple Attribute Decision Making: Methods and Applications. Berlin, Germany: Springer-Verlag, 1992. doi: 10.1007/978-3-642-46768-4
5. Pote M. Computational Propaganda Theory and Bot Detection System: Critical Literature Review. *arXiv*, Apr. 2024. doi: 10.48550/arXiv.2404.05240
6. Stella M., Ferrara E., De Domenico M. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*. 2018. Vol. 115, no. 49. P. 12435–12440. doi: 10.1073/pnas.1803470115
7. Романченко І.С., Потьомкін М.М. Метод TOPSIS та його використання для багатокритеріального порівняння альтернатив. *Системи обробки інформації*. 2016. 138. С. 104–106.

Надійшла: 14.10.2025 / Прийнята: 26.11.2025