

*Короленко О. О., здобувач вищої освіти 4 курсу спеціальності 122 Комп'ютерні науки,
Ротштейн О. П., д-р техн. наук, професор,
професор кафедри інформаційних технологій*

РОЗРОБКА ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ БАЄСІВСЬКИХ МЕРЕЖ

Донецький національний університет імені Василя Стуса, м. Вінниця

У сучасних умовах стрімкого розвитку цифрових технологій та активного впровадження інтелектуальних рішень у практику бізнесу, медицини, освіти й інших сфер життєдіяльності особливого значення набувають питання ефективного і швидкого прийняття рішень в умовах високої невизначеності. У ситуації, коли дані можуть бути частковими, неповними, зашумленими або суперечливими, класичні детерміновані алгоритми часто виявляються недостатньо гнучкими. Саме тому дедалі більшої популярності набувають ймовірнісні методи, зокрема баєсівські мережі (Bayesian Networks, БМ), що дають змогу моделювати причинно-наслідкові зв'язки між змінними та робити обґрунтовані висновки на основі наявної інформації [1; 2].

Баєсівські мережі поєднують у собі потужність графових структур і математичний апарат теорії ймовірностей. Їх суть полягає у представленні складної системи у вигляді орієнтованого ациклічного графу, де вузли відповідають змінним, а дуги – причинно-наслідковим залежностям. Завдяки можливості враховувати попередні знання (апостеріорні ймовірності) та динамічно оновлювати оцінки після отримання нових даних (апостеріорні ймовірності) такі мережі ідеально підходять для задач, де повнота інформації змінюється у часі [3; 4].

У межах виконаної роботи було реалізовано повноцінну систему підтримки прийняття рішень на основі баєсівських мереж, зосереджену на прикладі медичної діагностики. Актуальність обраної предметної області обумовлена критичною важливістю точності та пояснюваності рішень у медицині, де навіть незначна помилка у прогнозі може мати серйозні наслідки для здоров'я або життя людини.

Для реалізації системи було обрано технологічний стек на основі мови програмування Python та бібліотеки `rgmpy`, яка забезпечує інструменти для побудови графу, створення умовних таблиць ймовірностей (CPT), виконання параметричного навчання (використано алгоритми Hill-Climbing, Expectation-Maximization) та проведення ймовірнісного виведення за допомогою методу Variable Elimination [5].

Система має модульну структуру, яка включає: модуль збору та обробки вхідних даних (симптоми пацієнта), модуль побудови мережі та навчання, механізм прогнозування на основі виведення ймовірностей, інтерфейс користувача (реалізовано за допомогою Flask і Streamlit), а також REST API для подальшої інтеграції із зовнішніми платформами. Така архітектура дає змогу не лише локально аналізувати дані, але й підключатися до віддалених медичних баз чи IoT-сенсорів у майбутньому.

Особливу увагу приділено пояснюваності результатів. Користувач (лікар, аналітик чи системний оператор) має змогу побачити не лише фінальний прогноз, але й дізнатися, які саме симптоми або комбінації факторів вплинули на результат, які ймовірності змінилися після введення нових даних, а також отримати візуальне відображення змін у графі. Саме цей аспект – інтерпретованість – виділяє БМ серед інших підходів, наприклад, нейронних мереж, які часто функціонують як «чорна скринька» [6].

Результати численних сучасних досліджень підтверджують ефективність використання баєсівських мереж у медичній практиці. Зокрема, у роботах [7; 8] наведено приклади застосування БМ для виявлення онкологічних захворювань, оцінки ризиків серцево-судинних хвороб та аналізу електронних медичних записів. До того ж останнім часом активно досліджуються гібридні підходи, де БМ поєднуються з нейромережами для покращення точності за умови збереження логічної пояснюваності [9].

Водночас перспективним напрямом розвитку таких систем є впровадження динамічних баєсівських мереж (DBN), які дають змогу враховувати часову залежність параметрів. Це відкриває можливість розробки моніторингових медичних систем, що здатні оцінювати динаміку стану пацієнта в реальному часі (наприклад, на основі показників пульсу, температури, тиску тощо) [10].

За результатами тестування реалізованої системи з використанням навчального набору симптомів та діагнозів встановлено, що точність прогнозу у більшості випадків перевищує 85 %, навіть за неповної інформації. Система здатна гнучко реагувати на зміни вхідних даних, надаючи нові оцінки з мінімальним часом обчислення.

Отже, розроблена система підтверджує доцільність використання баєсівських мереж як основи для створення сучасних, надійних та пояснюваних рішень у сфері інтелектуального аналізу даних. Її застосування є перспективним не лише в медицині, але й у суміжних сферах: кібербезпеці, логістиці, праві, фінансовому моніторингу. Подальше розширення функціоналу, інтеграція із зовнішніми джерелами даних, а також масштабування до хмарних платформ дасть змогу використовувати такі системи як ключовий інструмент для підтримки прийняття рішень у різних галузях цифрової трансформації.

Список використаних джерел

1. Koller D., Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009. 1231 p.
2. Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.
3. Ankan A., Panda A. pgmpy: Probabilistic Graphical Models using Python. *Proceedings of the Python in Science Conference*. 2015. P. 1–6. URL: https://www.researchgate.net/publication/328778465_pgmpy_Probabilistic_Graphical_Models_using_Python
4. A Survey of Methods for Explaining Black Box Models / R. Guidotti, A. Monreale, S. Ruggeri et al. *ACM Computing Surveys*. 2019. Vol. 51(5). P. 93:1–93:42. URL: <https://dl.acm.org/doi/10.1145/3236009>
5. Fenton N., Neil M. *Risk Assessment and Decision Analysis with Bayesian Networks*. 2nd ed. CRC Press, 2018. URL: http://bayesianrisk.com/sample_chapters/K31678_C000.pdf

6. A Hybrid Deep Learning and Bayesian Network Model for Decision-Making under Uncertainty / S. Zhu, B. Chen, S. Gao et al. *2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*. URL: <https://ieeexplore.ieee.org/document/9486527>
7. Barber D. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

УДК 004.82:004.85]:791.6(043.2)

*Левченко М. Р., здобувачка вищої освіти 3 курсу спеціальності 122 Комп'ютерні науки,
Хмелівський Ю. С., асистент кафедри інформаційних технологій*

АНАЛІЗ ХАРАКТЕРИСТИК ФІЛЬМІВ ТА ЇХ УСПІШНОСТІ ЗА ДОПОМОГОЮ DATASET MOVIES У МОВІ R

Донецький національний університет імені Василя Стуса, м. Вінниця

Аналіз даних є важливим інструментом для виявлення закономірностей у великих масивах інформації. У цій роботі досліджується набір даних про фільми (movies) з пакету ggplot2movies [1] у середовищі R, який містить інформацію про жанри, рік випуску, тривалість, рейтинг та інші характеристики понад 58 тисяч фільмів. Метою дослідження є вивчення структури даних за допомогою кластерного аналізу та побудова моделей класифікації для прогнозування успішності фільмів. Робота базується на методах ієрархічної кластеризації, К-середніх та логістичної регресії, а також демонструє можливості мови R для статистичного аналізу та візуалізації.

Почнемо з кластерного аналізу [2].

Лістинг програми:

```
library(ggplot2movies)
library(dplyr)
data("movies")
movies_clean <- movies %>%
  select(title, year, length, rating, votes, Action, Comedy, Drama, Romance,
Short) %>%
  na.omit()
summary(movies_clean[, c("year", "length", "rating", "votes")])
colMeans(movies_clean[, c("Action", "Comedy", "Drama", "Romance", "Short")])
```

Результат:

Після очистки залишилось 58 788 фільмів і 10 змінних. Цільової змінної немає, тому задача кластеризації є без учителя. У більшості фільмів тривалість приблизно 82 хв, рейтинг – 5.9, а медіана голосів – 30. Найпоширеніші жанри: драма (37 %), комедія (29 %) та короткометражки (16 %).