

Марковські та байєсівські моделі у вебаналітиці поведінки користувачів

<https://doi.org/10.31713/MCIT.2025.060>

Мороз Дмитро

Національний університет «Львівська політехніка»

Львів, Україна

dmytro.a.moroz@lpnu.ua

Анотація — У статті розглянуто застосування статистичних методів для аналізу поведінкових патернів користувачів вебплатформ. Проведено порівняльний аналіз моделей марковських ланцюгів та байєсівських підходів, які використовуються для виявлення закономірностей переходів між станами користувацької взаємодії. На прикладі даних ОТТ-платформи Netflix продемонстровано можливості кожного методу для прогнозування дій користувачів. Встановлено, що моделі марковських ланцюгів ефективні для відтворення послідовності переходів і побудови карти навігації, тоді як байєсівський підхід дозволяє визначати ймовірності цільових дій і працювати з неповними або нерівномірними даними. Результати дослідження свідчать про доцільність розроблення гібридних моделей, що поєднують часову логіку марковських процесів і гнучкість байєсівських оцінок, для підвищення точності прогнозування поведінки користувачів вебплатформ.

Ключові слова — вебаналітика, поведінка користувачів, марковський ланцюг, байєсівський метод, ОТТ-платформи

I. ВСТУП

У сучасних умовах цифровізації суспільства вебаналітика стає ключовим інструментом для дослідження поведінки користувачів, оптимізації онлайн-платформ і підвищення ефективності електронних сервісів. Аналіз дій користувачів дозволяє виявляти закономірності навігації, прогнозувати наступні кроки та формувати персоналізовані рекомендації. Для цього дедалі частіше застосовуються статистичні та ймовірнісні моделі, здатні описувати послідовність і контекст користувацької взаємодії.

Серед таких підходів особливу увагу привертають марковські ланцюги та байєсівські методи, які забезпечують різні аспекти моделювання поведінкових процесів: часову динаміку та оцінку умовних ймовірностей. Їхнє порівняння дозволяє визначити найбільш ефективні інструменти для аналізу користувацьких патернів і підвищення точності прогнозування у сфері вебаналітики.

II. ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Мета роботи: Метою роботи є порівняння ефективності марковських і байєсівських моделей у

задачах аналізу та прогнозування поведінкових патернів користувачів вебплатформ, визначення їхніх переваг, обмежень і практичних напрямів застосування для оптимізації користувацької взаємодії.

Для досягнення поставленої мети було використано комплексний статистичний підхід, що включає побудову та порівняння марковських ланцюгів і байєсівських ймовірнісних моделей.

Дослідження здійснювалося на основі відкритого набору даних, що містить детальну інформацію про взаємодію користувачів із сервісом Netflix (<https://www.kaggle.com/datasets/arjunajn/netflix-watch-log>). Цей набір даних є репрезентативним джерелом для досліджень у галузі вебаналітики, поведінкового моделювання та побудови систем персоналізованих рекомендацій.

З вихідного датасету було сформовано набір послідовностей подій для кожного користувача. Кожна подія відображала певний стан взаємодії — наприклад, вибір контенту (*browseTitles*), перегляд деталей (*movieDetails*), початок відтворення (*playback*) тощо.

У результаті попереднього аналізу ідентифіковано 32 унікальні стани навігації, які охоплюють повний життєвий цикл взаємодії користувача з платформою. Отримані послідовності слугували основою для побудови моделей переходів між станами.

На першому етапі проведено аналіз поведінкових послідовностей за допомогою марковських ланцюгів. Цей метод дозволив описати ймовірність переходу користувача від одного стану до іншого та побудувати матрицю переходів, що відображає структуру навігаційних дій (Рис. 1).

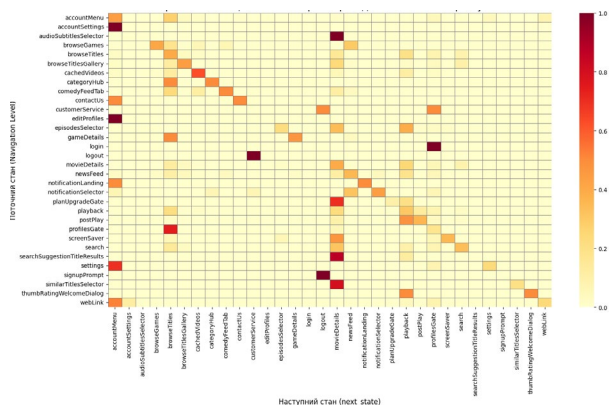


Рисунок 1. Переходи користувачів

На основі отриманих результатів візуалізовано типові шляхи користувачів, зокрема цикл *browseTitles* → *movieDetails* → *playback* → *postPlay* → *browseTitles*, який є характерним для платформ потокового відео.

Цей цикл відображає природну логіку користування сервісом — від вибору контенту до його перегляду і повернення до пошуку нового відео.

На другому етапі було застосовано байєсівський підхід, який дозволяє уточнювати ймовірнісні оцінки переходів у процесі надходження нових даних і враховувати невизначеність у спостереженнях. Для кожного стану було обчислено ймовірність переходу до ключової дії — перегляду відео (*playback*). Отримані результати подано у вигляді теплової карти (рис. 2), що ілюструє, з яких станів користувач найчастіше переходить до перегляду.

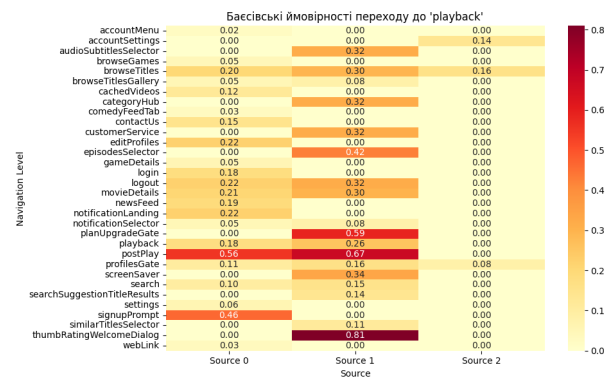


Рисунок 2. Байєсівські ймовірності переходу користувачів до стану перегляду (*playback*)

Найвищі ймовірності виявлено для станів *thumbRatingWelcomeDialog* (0,81), *postPlay* (0,65) та *planUpgradeGate* (0,59), що вказує на сильну кореляцію між оцінкою контенту, завершенням попереднього перегляду і запуском нового відео.

Обидва розглянутих методи мають свої переваги та недоліки, які обумовлені їх певною обмеженістю. Загалом, вони обидва відтворили реальну картину взаємодії користувача з сайтом.

Метод марковських ланцюгів показав, як користувачі переходять між екранами Netflix, тобто дозволяє побудувати карту взаємодій. У випадку

даного датасету переходи виглядають високоструктурованими, що свідчить про типовість користувацьких сценаріїв (наприклад, *profilesGate* → *browseTitles* → *movieDetails* → *playback*).

Байєсівський підхід оцінює ймовірність того, що користувач розпочне перегляд (*playback*), залежно від поточного стану; це дає можливість оцінити ефективність кожного етапу навігації щодо стимулювання переглядів.

Порівняльний аналіз двох підходів наведений в таблиці 1.

Таблиця 1. Порівняльний аналіз методів марковських ланцюгів та байєсівського підходу		
Критерій	Марковський ланцюг	Байєсівський підхід
Тип моделі	Послідовна, стохастична.	Ймовірнісна, класифікаційна.
Що показує	Як користувач рухається між станами навігації (перехід між сторінками, меню, переглядами).	Яка ймовірність того, що користувач почне перегляд (<i>playback</i>) з певного стану або контексту.
Основна мета	Моделювання ланцюжків переходів.	Моделювання умовної ймовірності події.
Переваги	Враховує часову послідовність дій. Добре показує "шляхи користувача". Підходить для побудови графів переходів.	Гнучкий до додаткових ознак (пристрій, країна, тип користувача). Дає інтерпретовані ймовірності. Може працювати навіть із неповними даними.
Недоліки	Ігнорує фактори контексту (час, пристрій, профіль). Якщо даних мало – переходи виглядають детермінованими (1.0).	Не враховує часову послідовність (стани незалежні). Припущення незалежності ознак може спотворювати результати.
В аналізованому датасеті Netflix	Модель виявила сталі, послідовні переходи: <i>profilesGate</i> → <i>browseTitles</i> , <i>searchSuggestionTitleResults</i> → <i>movieDetails</i> . Поведінка структурована.	Модель показала ймовірність початку перегляду, найвищу після <i>thumbRatingWelcomeDialog</i> , <i>postPlay</i> , <i>episodesSelector</i> .
Приклад результату	<i>profilesGate</i> → <i>browseTitles</i> (ймовірність ≈ 0.75).	<i>thumbRatingWelcomeDialog</i> → <i>playback</i> (ймовірність ≈ 0.81).
Тип інформації, що дає модель	Структура шляху користувача, навігаційна поведінка.	Ймовірність конверсії до перегляду з різних точок взаємодії.
Ефективність методу	Добре підходить для аналізу сценаріїв використання (UX, flow користувача).	Корисний для оцінки готовності користувача до перегляду (модель конверсії).

III. ВИСНОВКИ

Проведене дослідження підтвердило, що обидва методи ефективні у виявленні поведінкових закономірностей, проте мають різну аналітичну спрямованість. Марковські моделі є доцільними для аналізу послідовності дій і побудови структурних моделей навігації. Байєсівські підходи краще відображають контекстуальні фактори та ймовірність конверсійних подій.

Перспективним напрямом подальших досліджень є створення гібридних моделей, що поєднують часову логіку марковських процесів і гнучкість байєсівських оцінок, що дозволить підвищити точність прогнозування поведінки користувачів і ефективність вебаналітичних систем.

СПИСОК ЛІТЕРАТУРИ

- [1] Naprawski, T. (2023). The impact of web analytics tools on knowledge management. *Procedia Computer Science*, 225, 3404–3414. <https://doi.org/10.1016/j.procs.2023.01.123>
- [2] Bekavac, I., & Garbin Praničević, D. (2015). Web analytics tools and web metrics tools: An overview and comparative analysis. *Croatian Operational Research Review*, 6(2), 373–386. <https://doi.org/10.17535/croirr.2015.0029>
- [3] Hausmann, V., Williams, S. P., & Schubert, P. (2012). Developing a framework for web analytics. In *BLED 2012 Proceedings* (Paper 11). <https://aisel.aisnet.org/bled2012/11>
- [4] Önder, I., & Berbekova, A. (2022). Web analytics: More than website performance evaluation? *International Journal of Tourism Cities*, 8(3), 603–615. <https://doi.org/10.1108/IJTC-10-2021-0213>
- [5] Kalan, R. (2025). Website analysis in qualitative research. *International Journal of Advanced Research*, 13(2), 536–544. <https://doi.org/10.21474/IJAR01/000000>
- [6] Qing, H., Ibrahim, R., & Nies, H. W. (2024). Analysis of web design visual element attention based on user educational background. *Scientific Reports*, 14, 4657. <https://doi.org/10.1038/s41598-024-54444-8>
- [7] Morales-Vargas, A., Pedraza-Jimenez, R., & Codina, L. (2023). Website quality evaluation: A model for developing comprehensive assessment instruments based on key quality factors. *Journal of Documentation*, 79(7), 95–114. <https://doi.org/10.1108/JD-11-2022-0246>