

зростає, тому що алгоритм починає частіше вибирати dataset з вищою ймовірністю отримання винагороди.

Потрібно уникнути повного ігнорування одного з запропонованих масивів даних, для вирішення цієї задачі було прийняте рішення поставити нижню межу для коефіцієнту успішності кожної матриці, що буде рівна 0,3. Таким чином, ми залишаємо можливість залишатися продуктам з невдалої матриці на рекомендаційній каруселі, у меншій кількості, але з надією, що вона у перспективі зможе повернути свою успішність.

Таким чином було розроблено нову рекомендаційну систему, яка вирішує ряд проблем, що присутні у її аналогів, за допомогою додавання нових операцій до уже існуючих алгоритмів та розширення уже існуючого функціоналу.

Список літератури:

1. Authors: Peng Hu , Rong Du , Yao Hu and Nan Li - Hybrid Item-Item Recommendation via Semi-Parametric Embedding - [Електронний ресурс] – режим доступу до ресурсу: <https://www.ijcai.org/proceedings/2019/0350.pdf>
2. Розподіл Бернуллі - [Електронний ресурс] – режим доступу до ресурсу: <http://people.umass.edu/~biep540w/pdf/bernoulli.pdf>
3. Authors: Daniel J. Russo , Benjamin Van Roy , Abbas Kazerouni , Ian Osband and Zheng Wen - Thompson Sampling algorithm - [Електронний ресурс] – режим доступу до ресурсу: https://web.stanford.edu/~bvr/pubs/TS_Tutorial.pdf

УДК 004.01

Дужак А. О.,
студент СО «Бакалавр»,
Нескородева Т.В д.т.н, доцент
кафедри інформаційних технологій

ОСОБЛИВОСТІ РОБОТИ РЕКОМЕНДАЦІЙНИХ АЛГОРИТМІВ YOUTUBE

Донецький національний університет імені Василя Стуса, м. Вінниця

Вступ. YouTube – популярний відеохостинг, заснований у 2005 році. Станом на березень 2020 року за даними Alexa Company сайт Youtube.com став другим за відвідуваністю у світі та Україні, поступаючись лише Google.com [1,2]. Сайт дозволяє завантажувати, редагувати, дивитися та коментувати відеоролики кожному користувачу. Протягом років свого існування YouTube розширював свій функціонал і ставав все більш універсальною та багатофункціональною інформаційною системою. За даними Hootsuite, станом на 2020 рік платформою користується більше 2 мільярдів чоловік щомісяця, а щохвилини на сайті з'являється 500 годин відеоконтенту.

Актуальність. Такий масив інформації потребує обробки. На YouTube велике значення відіграють алгоритми, що здатні виокремити серед завантажуваних на сайт відео ті, які варто рекомендувати до перегляду індивідуально кожному користувачу з урахування безлічі параметрів, основними з яких є мова відео та тематика. Окремо від системи рекомендацій існує пошуковий механізм YouTube, який працює аналогічно до пошукової системи Google та обирає найбільш актуальні результати пошуку згідно з запитом користувача.

Важливо розуміти та вивчати особливості пошукових та рекомендаційних алгоритмів YouTube, бо вони мають прямий вплив на кількість можливих переглядів відео, тобто ефективність комунікації між автором контенту та аудиторією. Ці знання можуть використовувати не тільки блогери, а й приватні телеканали чи державні медіа, YouTube-канали яких на сьогоднішній день функціонують, але не завжди дістаються аудиторії. Світова статистика говорить, що користувач YouTube в середньому проводить на сайті 11 хвилин 24 секунди на день, а 70% з того, що він переглядає, визначено алгоритмом рекомендацій. Сенса боротися за увагу глядача на найбільш популярному відео сервісі світу є у кожного засобу інформації. YouTube – всесвітня платформа для поширення відеоконтенту, що активно функціонує у наші часи, тому вивчення особливостей її роботи є надзвичайно актуальним питанням.

Об’єкт дослідження – всесвітня платформа для поширення відеоконтенту у Інтернеті YouTube.

Предмет дослідження – властивості рекомендаційних алгоритмів YouTube. Наукова спільнота активно вивчає YouTube з багатьох сторін. Базою для цього дослідження стала стаття від групи працівників компанії YouTube Ковінгтона, Адамса та Сорджина «Deep Neural Networks for YouTube Recommendations» [3], у якій уперше публічно докладно розказано про особливості роботи рекомендаційних алгоритмів відеохостингу. В Україні рекомендаційні інформаційні системи для пошуку відеоматеріалів вивчали Ю. Парфененко, А. Ковтун та А. Вербицька. Методи побудови рекомендаційних систем в Інтернеті досліджували Є. Мелешко, С. Семенов, В. Хох., Нескородєва Т.В., Федоров Є.Є. [4]. Рекомендаційні механізми YouTube мають багато відмінностей у порівнянні із аналогічними алгоритмами інших систем, вони є сучасними та продовжують удосконалюватися. Наявні дослідження не розкривають цю тему, необхідне більш детальне її вивчення. Ще однією досі невирішеною проблемою залишаються методики використання відомих алгоритмів YouTube з метою більш ефективного поширення відеоконтенту телеканалів чи інших засобів масової інформації.

Новизна дослідження. І хоча Youtube та його рекомендаційні алгоритми вже досліджувалися раніше, питання практичного застосування наявних знань у журналістиці ніхто не розглядав. Це дослідження дозволяє комплексно підійти до питання: від моменту розуміння того, як працює система рекомендацій на актуальних прикладах, до конкретних рекомендацій щодо більш ефективного

підходу до створення та просування відеоконтенту на платформі YouTube у майбутньому.

Мета дослідження полягає у виявленні актуальних ефективних стратегій роботи з рекомендаційним алгоритмом на платформі YouTube, які дозволять позитивно вплинути на показники переглядів на відео. Під час дослідження вирішувалися такі завдання: відтворення історії змін до системи рекомендацій YouTube. опис актуальної на момент проведення дослідження схеми роботи рекомендаційних алгоритмів YouTube. виявлення найбільш ефективних з точки зору впливу на рекомендаційний рейтинг параметрів відео.

10 серпня 2012 р. команда авторів YouTube повідомила користувачів про зміни у механізмі роботи розділу рекомендованих до перегляду відео: було зміщено пріоритет від кількості переглядів до параметру watch time, тобто тривалості перегляду. «Відтепер, коли ми рекомендуємо відео, то акцентуємо увагу на тому, щоб збільшити час, який користувач проводить за переглядом... Найкраще, що може зробити автор, щоб бути успішним на YouTube, це створювати відео, які люди будуть хотіти дивитися». У цей же момент розробники додали і параметр session time, тобто час, проведений на сайті Youtube в цілому. І хоча цей крок призвів до чудових результатів, з часом з'являлися і мінуси. Деякі автори почали цілеспрямовано розтягувати відеоролики та переходили до суті тільки ближче до кінця відео, бо це дозволяло отримувати перевагу від нових параметрів, що відігравали ключову роль у поширенні відео. Також від нової системи страждали автори, які не мали змоги створювати тривалі відео (бо, наприклад, цей процес був дуже непростим). І навіть якщо цей контент отримував багато схвальних відгуків, він не мав змоги збирати велику кількість переглядів, бо поступався за ключовими показниками: тривалості перегляду та сесії.

Тому YouTube продовжив розвиток рекомендаційних алгоритмів. У 2016 р. трійка працівників YouTube (Пол Ковінгтон, Джей Адамс та Емре Серджін) опублікувала статтю, де вперше публічно поділилася інформацією про глибокі нейронні мережі та машинне навчання, які використовуються для формування рекомендацій YouTube. Алгоритм визначення відео для рекомендацій, який створили у компанії, базувався на приблизно мільярді параметрів, які умовно було поділено на дві категорії: ті, що стосуються глядача, або глядацькі (допомагають створити уявлення про інтереси з метою підібрати цікаву для нього тематику відео) та ті, що стосуються відео, або рейтингові (допомагають визначити тематику, якість та привабливість контенту, тобто його рейтинг). Глядацькі параметри сильно відрізняються: деякі є простими (наприклад, чи користувач увійшов у систему), а інші мають мільйони можливих значень (наприклад, останні пошукові запити користувача). Також параметри було поділено за кількістю отримуваних значень: одновалентні (наприклад, ID відео, яке задовольнило глядача) чи мультивалентні (наприклад, реакція користувача на групу відео). До уваги беруться параметри не тільки предмету рекомендації (власне відеоролику), а й параметри користувача (його інтереси та запити). На основі цих параметрів кожне відео проходить шлях до глядача, при цьому мета

залишається незмінною: заохотити аудиторію якомога більше часу проводити на YouTube.

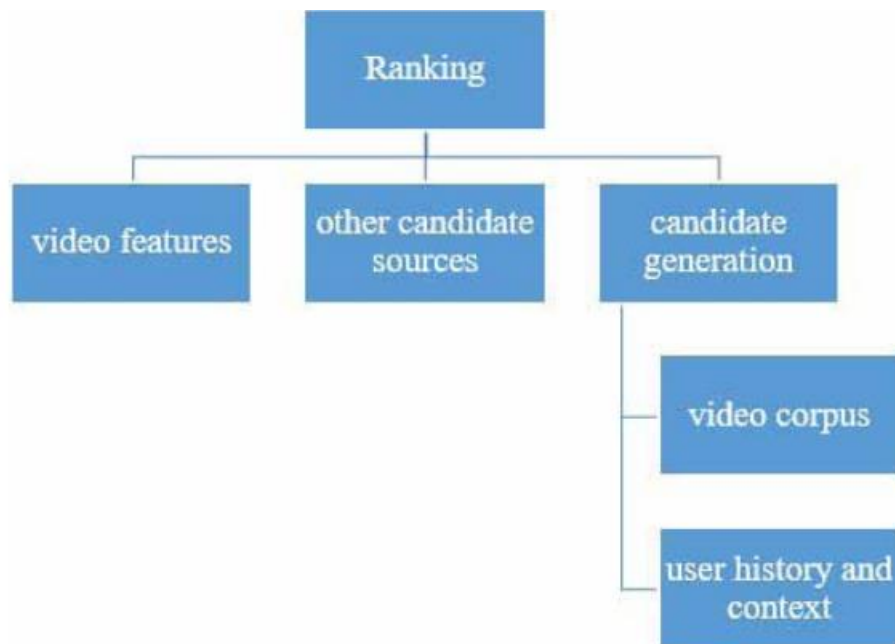


Рисунок 1 – Система формування рекомендаційного рейтингу YouTube

Ця схема демонструє процес досягнення відеороликом глядача. Алгоритм на основі історії та контексту кожного окремого користувача (user history and context) генерує список з мільйонів можливих кандидатів для рекомендації (candidate generation), потім виділяє тисячі кандидатів за допомогою рейтингу (ranking), що формується з урахуванням особливостей та параметрів відеоролику (video features). І лише десятки користувачів, які згідно з рейтингом є найбільш вдалим кандидатами, бачать відеоролик у списку своїх рекомендацій. Оберуть вони це відео для перегляду чи ні, відреагують на нього чи одразу вимкнуть – ці та ще безліч інших параметрів впливають не тільки на майбутні рекомендації цих конкретних користувачів, але й впливають на рейтинг цього ж відеоролику (other candidate sources). Алгоритм побудовано таким чином, щоб відеоролики, які отримали позитивну реакцію від глядача, отримували більше охоплення. І хоча з виходом статті «Deep Neural Networks for YouTube Recommendations» у багатьох відкрилися очі на алгоритми YouTube, більшість параметрів, які впливають на оцінку відеороликів, досі тримаються у секреті. Це спричинило багато критики на адресу YouTube: экс-працівник компанії називав алгоритм «локомотивом дезінформації», та «великим радикалайзером» (через те, що користувачам часто рекомендувалися відео про теорії заговорів, фейкові новини та тривожний контент).

Тож YouTube продовжив рухатися у напрямку розвитку своїх рекомендаційних алгоритмів. У 2017 р. компанія почала розширювати штат модераторами, видаляти небезпечний контент та позбавляти монетизації проблемні канали. У грудні 2019 р. працівники YouTube у своєму блозі повідомили про те, що протягом тільки останнього року зробили сотні змін у рекомендаційному алгоритмі, зокрема зменшили кількість рекомендацій відео,

що межують з Community Guidelines (правилами спільноти), але не порушують їх.

Список літератури

1. P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
2. X. Amatriain. Building industrial-scale real-world recommender systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems* <https://medialab.online/news/alhorytmy-youtube/>
3. Covington, Paul, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations." *Proceedings of the 10th ACM conference on recommender systems*. 2016.
4. Т. В. Нескородєва, Є. Є. Федоров, Т. В. Січко, А. Р. Нескородєва. Експертні та рекомендаційні системи: навч. посіб. для здобувачів вищої освіти спеціальностей 122 «Комп'ютерні науки», 125 «Кібербезпека», 113 «Прикладна математика». – ДонНУ імені Василя Стуса. 2021. - 320с.

УДК 004.89

*Жеребцов О.М., студент СО «Магістр»
спеціальності 122 «Комп'ютерні науки»
Нескородєва Т.В., д.т.н., доцент, зав.
кафедри комп'ютерних наук та
інформаційних технологій*

РЕКОМЕНДАЦІЙНІ СИСТЕМИ ДЛЯ ІНТЕРНЕТ-МАГАЗИНУ

Донецький національний університет імені Василя Стуса, м. Вінниця

В наші дні рекомендаційні системи вже досить поширені і мають велику кількість застосувань. В першу чергу, рекомендаційні системи використовуються в інтернет-комерції для того, щоб допомогти користувачам вибрати відповідні товари. Такі сервіси збирають інформацію про переваги користувачів і намагаються запропонувати їм корисні товари. Існує багато методів для формування рекомендацій, але всі вони мають як свої переваги так і свої недоліки. Саме тому дослідження в даній області актуальні.

Рекомендації формуються персонально для кожної людини, опираючись на її попередні дії на конкретному веб-ресурсі чи на основі минулої активності. Окрім того, значення має і поведінка попередніх учасників процесу.[1]

На сьогоднішній день при створенні рекомендаційних систем використовуються такі основні стратегії: колаборативна фільтрація, контентна фільтрація та фільтрація на основі вмісту.

Колаборативна фільтрація

Методи, які зазвичай використовуються в колаборативній фільтрації це