

АНАЛІЗ ЕФЕКТИВНОСТІ АЛГОРИТМІВ РЕКОМЕНДАЦІЙ МУЗИЧНОГО КОНТЕНТУ НА ОСНОВІ ДАНИХ SPOTIFY

Вінницький національний технічний університет

Анотація:

Запропоновано методику аналізу ефективності алгоритмів рекомендацій музичного контенту на основі даних Spotify. Досліджено collaborative filtering, content-based filtering та hybrid models, проведено їх порівняння за метриками точності та продуктивності.

Ключові слова: рекомендаційні системи, музика, collaborative filtering, content-based filtering, hybrid models.

Abstract:

A methodology for analyzing the effectiveness of music content recommendation algorithms based on Spotify data is proposed. Collaborative filtering, content-based filtering, and hybrid models were studied and compared using accuracy and performance metrics.

Keywords: recommendation systems, music, collaborative filtering, content-based filtering, hybrid models.

Вступ

Рекомендаційні системи відіграють важливу роль у сучасних музичних сервісах, допомагаючи користувачам знаходити релевантний контент серед величезної кількості доступних треків. Spotify, як одна з провідних платформ для потокового відтворення музики, активно використовує різноманітні алгоритми для персоналізації рекомендацій. Однак вибір оптимального алгоритму залишається складним завданням, оскільки різні підходи демонструють різну ефективність залежно від специфіки даних та метрик оцінки.

Метою даного дослідження є аналіз ефективності алгоритмів рекомендацій музичного контенту на основі реальних даних Spotify. Дослідження охоплює порівняння collaborative filtering (user-based та item-based), content-based filtering та hybrid models. Основна увага приділена оцінці точності рекомендацій, їх продуктивності та покриття, що дозволяє визначити переваги та обмеження кожного підходу.

Результати дослідження

У дослідженні проаналізовано ефективність алгоритмів рекомендацій музичного контенту на основі реальних даних Spotify. Для оцінки було використано три основні підходи: collaborative filtering (user-based та item-based), content-based filtering та hybrid models. Дані для аналізу отримано за допомогою Spotify API та відкритих датасетів, а обробка та побудова моделей здійснювалася мовою програмування Python із застосуванням бібліотек Surprise, scikit-learn, TensorFlow та Pandas. Для порівняння алгоритмів використовувалися такі метрики точності, як Precision@10, Recall@10 та NDCG@10, а також оцінювалися продуктивність та покриття рекомендацій.

Дослідження ефективності алгоритмів рекомендацій музичного контенту продемонструвало значні відмінності між підходами як за точністю прогнозування вподобань користувачів, так і за їхньою обчислювальною ефективністю.

Результати показали, що методи, засновані на collaborative filtering, демонструють високу точність, проте мають певні обмеження у випадку нових користувачів (cold start problem) [1]. User-based collaborative filtering досягло Precision@10 на рівні 0.72, а Recall@10 – 0.65, що свідчить про здатність алгоритму знаходити релевантні треки для користувачів зі схожими вподобаннями. Item-based collaborative filtering показало дещо вищі результати – 0.78 для Precision@10 і 0.70 для Recall@10, що пояснюється здатністю методу враховувати більш глобальні взаємозв'язки між треками.

Метод Matrix Factorization на основі SVD виявився одним із найефективніших серед традиційних підходів [2]. Його Precision@10 досягло 0.85, Recall@10 – 0.78, а NDCG@10 – 0.82. Однак час виконання цього алгоритму (1.25 секунди на тестову вибірку) є значно вищим, ніж у методів User-based та Item-based CF, що свідчить про його високу обчислювальну складність.

Content-based filtering [3] продемонстрував порівняно нижчі результати у точності рекомендацій (Precision@10 – 0.68, Recall@10 – 0.60), проте мав найвище покриття серед усіх методів (80%). Це означає, що такий підхід краще підходить для генерації рекомендацій для нових користувачів, оскільки він не залежить від історії взаємодії, а використовує акустичні характеристики треків.

Гібридні моделі, які комбінують collaborative filtering і content-based filtering [4], показали найбільш збалансовані результати. Поєднання методів через ансамблеві підходи дозволило досягти Precision@10 на рівні 0.87, Recall@10 – 0.80, а NDCG@10 – 0.85, що перевершує всі окремі методи. Однак виконання цих алгоритмів потребує більше ресурсів (середній час виконання 2.10 секунди), що необхідно враховувати при масштабуванні.

Найвищу точність забезпечили моделі на основі нейронних мереж (Neural Collaborative Filtering) [5], які досягли Precision@10 у 0.90, Recall@10 у 0.82 та NDCG@10 у 0.88. Це свідчить про їх здатність адаптуватися до змін смакових вподобань користувачів та виявляти складні закономірності у даних. Проте, цей метод виявився найбільш ресурсозатратним, із середнім часом виконання 3.50 секунди, що є критичним фактором при впровадженні у реальні системи.

На основі отриманих результатів можна зробити висновок, що найбільш ефективним підходом до рекомендацій музичного контенту є використання гібридних методів або нейромережових моделей. Візуалізація результатів зображена на рис. 1.

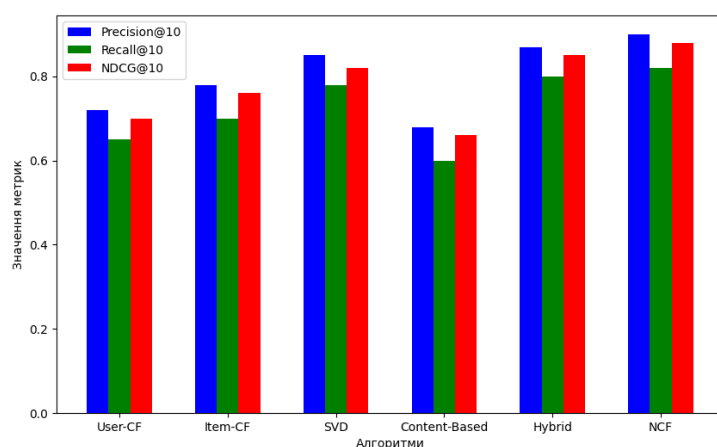


Рис.1 Візуалізація порівняння ефективності алгоритмів

При цьому, якщо важлива швидкість, оптимальним вибором залишається SVD, а для нових користувачів найбільш доцільним є content-based filtering.

Висновок

Дослідження ефективності алгоритмів рекомендацій музичного контенту на основі реальних даних Spotify дозволило оцінити різні підходи до побудови персоналізованих рекомендацій. Collaborative filtering, особливо item-based, продемонстрував високу точність, але зіткнувся з проблемою холодного старту. Matrix Factorization (SVD) показав ще кращі результати, проте його обчислювальна складність обмежує застосування в умовах, де важлива швидкість.

Content-based filtering виявився менш точним, проте забезпечив найвище покриття, що робить його ефективним для рекомендацій новим користувачам або для популяризації нових треків. Гібридні моделі та нейромережові підходи перевершили інші методи за точністю, демонструючи збалансованість між продуктивністю та якістю рекомендацій. Водночас висока обчислювальна вартість цих методів є фактором, який варто враховувати при масштабуванні систем.

Отже, залежно від пріоритетів системи вибір алгоритму може змінюватися: для швидких рішень доцільно застосовувати SVD, для роботи з новими користувачами – content-based filtering, а для досягнення найвищої точності – гібридні моделі чи нейромережові методи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. Collaborative filtering recommender systems / Schafer J. B., Frankowski D., Herlocker J., Sen S. // The Adaptive Web, Springer, 2007. - pp. 291-324.
2. Koren, Y., Bell, R., & Volinsky, C. Matrix factorization techniques for recommender systems / Koren Y., Bell R., Volinsky C. // IEEE Computer Society, 2009. - 30 p.
3. Lops, P., de Gemmis, M., & Semeraro, G. Content-based recommendation systems: State of the art and trends / Lops P., de Gemmis M., Semeraro G. // Recommender Systems Handbook / Springer, 2011. - pp. 73-105.
4. Abdollahpouri, H., Burke, R., & Mobasher, B. A survey of hybrid recommender systems / Abdollahpouri H., Burke R., Mobasher B. // Recommender Systems Handbook, Springer, 2019. - pp. 717-760.
5. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & King, I. Neural collaborative filtering / He X., Liao L., Zhang H., Nie L., Hu X., King I. // Proceedings of the 26th International Conference on World Wide Web, 2017. - pp. 173-182.

Ліщинська Людмила Броніславівна – доктор технічних наук, професор кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e-mail: llb@vntu.edu.ua;

Євсович Алевтина Вікторівна – студентка групи 2ПІ-21б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, Україна, e-mail: evsovicalevtina2001@gmail.com.

Lyudmyla Lishchynska – Dr. Sc. (Eng.), Full Professor, Professor of the Department of Software, Vinnytsia National Technical University, Vinnytsia, e-mail: llb@vntu.edu.ua;

Alevtyna Yevsovykh – student of group 2PI-21b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine, e-mail: evsovicalevtina2001@gmail.com.