

УДК 004.032.26

DOI: 10.31673/2412-9070.2020.066465

Д. Д. ДЕМИДОВ, студент;

І. С. ЩЕРБИНА, канд. техн. наук, доцент;

Н. А. ТРИНТИНА, доцент,

А. М. ШТИММЕРМАН, ст. викладач;

С. М. ІЩЕРЯКОВ, канд. техн. наук, доцент,

Державний університет телекомунікацій, Київ

РОЗРОБЛЕННЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ВИБОРУ МУЗИЧНИХ КОМПОЗИЦІЙ

Проаналізовано метод сингулярного розкладання матриці (SVD) як ефективний спосіб побудови рекомендаційної системи. Розвиток інформаційних технологій та їх упровадження в суспільне життя зумовлює потребу пошуку акцентованої інформації за умов невизначеності. Для розв'язання таких завдань останнім часом створюються інтелектуальні рекомендаційні системи. Популярність рекомендаційних систем зростає в кожному сегменті товарів і послуг, зокрема музичних. Із соціально-економічного погляду такі системи є основним інструментом поширення нових композицій у сфері музики, сприяє просуванню цих композицій відповідно до вподобань цільової аудиторії і стимулює користувачів набувати нових музичних треків. Окрім цього такі системи значно скорочують час і полегшують пошук відповідних музичних композицій за умов невизначеності.

Ключові слова: матриця; машинне навчання; рекомендаційна система.

Вступ

Постановка проблеми. Основною проблемою розроблення алгоритмів машинного навчання є відсутність індивідуального підходу до кожного з користувачів. Усі рекомендації створюються на базі статистичної поведінки більшості, у результаті чого формується деякий відсоток людей, які не отримують рекомендації, що відповідають їх особистим вподобанням. У разі окремого аналізу кожного з користувачів, а також впровадження рекомендацій відповідно до їх особистого використання інтернет-ресурсів кількість якісних і точніших пропозицій у списку рекомендацій значно б збільшилась.

Аналіз останніх досліджень і публікацій. Теоретичні та методологічні основи побудови рекомендаційних систем, дослідження алгоритмів формування списку рекомендацій, розроблення та доцільність практичних рекомендацій знайшли відображення в працях багатьох відомих зарубіжних вчених: А. Гомзіна, А. Пройдла, А. Азарія, А. Хасидим, С. Крауса, А. Еш-кол, В. Джонса, Г. Фур-наса, С. Чої, С. Тапперта, Б. Хань, С. Ро, С. Джуна, Е. Хван, П. Мелвілліанда, В. Сіндхвані, Л. Лі, Д. Вана, Д. Лі, Д. Нокса, Б. Падманабхан, Дж. Констана, Б. Міллера, Д. Мальца, Дж. Херлокера, Л. Гордона, Д. Рідла та ін.

Мета статті полягає в ознайомленні з методом сингулярного розкладання матриці та обґрунтуванні його ефективності в процесі побудови рекомендаційної системи.

Основна частина

З розвитком інформаційних технологій та їх упровадженням у суспільне життя постає потреба пошуку акцентованої інформації за умов невизначеності. Для вирішення таких завдань сьогодні створюються інтелектуальні рекомендаційні системи [1]. Популярність рекомендаційних систем зростає в кожному сегменті товарів і послуг, зокрема музичних. З соціально-економічного погляду такі системи є основним інструментом поширення нових композицій у сфері музики, що сприяє просуванню цих композицій згідно із вподобаннями цільової аудиторії і стимулює користувачів набувати нових музичних треків. Також такі системи значно скорочують час і полегшують пошук відповідних музичних композицій за умов невизначеності.

Для побудови рекомендаційних систем ефективно використовують методи машинного навчання, а саме: метод k -найближчих сусідів, алгоритм Байеса та метод сингулярного розкладання матриці (SVD).

Серед цих методів найбільш широкого застосування дістав на практиці метод SVD [2]. Цей метод використовується для зменшення кількості факторів набору даних, що не є значущими. Факторами в рекомендаційних системах є властивості, які описують користувача або предмет. У музичних рекомендаційних системах це може бути жанр. SVD зменшує розмірність матриці завдяки вилученню її прихованих чинників. Сингулярне розкладання матриці має такий вигляд:

$$\mathbf{R} \approx \mathbf{U}\mathbf{S}\mathbf{V}^t, \quad (1)$$

де \mathbf{U} — ліва сингулярна матриця, що унаочнює взаємозв'язок між користувачами і прихованими факторами; \mathbf{S} — діагональна матриця, що описує вплив кожного прихованого фактора; \mathbf{V}^t — права сингулярна матриця, яка зазначає подібність між елементами і прихованими факторами; \mathbf{R} — матриця вихідних оцінок.

Машинне навчання використовують для пошуку векторів оцінок, що нам невідомі. Наша модель буде залежати від багатьох параметрів — векторів користувачів і музичних треків. Щоб передбачити оцінку, ми візьмемо вектор користувача, вектор фільму і дістанемо їх скалярний добуток:

$$r(\theta) = p_u^T q_i,$$

$$\theta = \{p_u, q_i \mid u \in U, i \in I\},$$

де p_u — вектор факторів конкретного користувача; q_i — вектор факторів конкретного музичного треку; θ — скалярний добуток векторів.

Основна мета такої моделі — мінімізувати похибку:

$$E(\hat{r}(\theta) - r) \rightarrow \min, \quad (2)$$

де r — оцінка музичної композиції, $r \in \mathbf{R}$.

Далі визначаємо такий параметр θ , з яким квадрат похибки буде найменший. Щоб підібрати параметр θ у формулі (2), скористаємось методом градієнтного спуску.

Метод градієнтного спуску — це ітеративний алгоритм, що багаторазово бере параметри певної точки і з огляду на градієнт змінює значення проти його напрямку:

$$\theta(t+1) = \theta_t - \mu \nabla(\theta), \quad (3)$$

де $\nabla(\theta)$ — градієнт напрямку; μ — задає швидкість градієнтного спуску.

Основною проблемою цього методу є повільна швидкість оброблення даних в онлайн-системах, а отже, пропонується удосконалити застосування методу SVD поєднанням його з методом градієнтного бустингу [3]. Це дає можливість скоротити час та збільшити швидкість оброблення даних. Основною метою алгоритму градієнтного бустингу є відновлення функції у загальному контексті навчання з учителем. Тобто, відшукується наближене значення функції:

$$\hat{f}(x) = \operatorname{argmin} L(y, f(x)), \quad (4)$$

де $\hat{f}(x)$ — наближена функція; $L(y, f(x))$ — функція втрат.

Слід зауважити, що простір функцій $f(x)$ прямує до нескінченності, тому для розв'язання задачі градієнтного бустингу обмежимо простір пошуку певним конкретним параметризованим сімейством функцій $f(x, \theta), \theta \in R^n$, де R^n — простір дійсних векторів.

Сформуємо підсумкову модель градієнтного бустингу:

$$\hat{f}(x) = \sum_{i=0}^M \hat{f}_i(x), \quad (5)$$

де M — кількість ітерацій підбору наближеної функції.

Результатом розрахунку стане модель, що оптимально прогнозує вподобання користувача та пропонує музичні треки, що найбільш схожі на минулі композиції.

Висновки

Таким чином, упровадження рекомендаційних систем у музичні сервіси є важливим фактором задоволення потреб людей у музичних композиціях в їх повсякденному житті.

Список використаної літератури

1. **Melville P., Sindhvani V.** Recommender Systems // *Encyclopedia of Machine Learning / Claude Sammut and Geoffrey Webb (Eds).* Springer, 2010. 9 с.
2. **Королева Д. Е., Филиппов М. В.** Анализ алгоритмов обучения коллаборативных рекомендательных систем // *Наука и инновации: электрон. версія журн.* 2013. Вип. 6. URL: <http://engjournal.ru/catalog/it/hidden/816.html>
3. **Hartshorn S.** Machine Learning With Boosting: A Beginner's Guide. Springer, 2017. 227 с.

Д. Д. Демидов, И. С. Щербина, Н. А. Тринтина, А. Н. Штimmerман, С. М. Ищеряков РАЗРАБОТКА АЛГОРИТМОВ МАШИНОГО ОБУЧЕНИЯ ДЛЯ РЕКОМЕНДАЦИОННОЙ СИСТЕМЫ ВЫБОРА МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ

Проанализирован метод сингулярного разложения матрицы (SVD) как эффективный способ построения рекомендательной системы. С развитием информационных технологий и их внедрением в общественную жизнь возникает потребность поиска акцентированной информации в условиях неопределенности. Для решения таких задач в последнее время создаются интеллектуальные рекомендательные системы. Популярность рекомендательных систем растет в каждом сегменте товаров и услуг, в частности музыкальных.

С социально-экономической точки зрения такие системы являются основным инструментом распространения новых композиций в сфере музыки, способствуют продвижению этих композиции в соответствии с предпочтениями целевой аудитории и стимулируют пользователей приобретать новые музыкальные треки. Кроме этого такие системы значительно сокращают время и облегчают поиск соответствующих музыкальных композиций в условиях неопределенности.

Ключевые слова: матрица; машинное обучение; рекомендательная система.

D. D. Demydov, I. S. Sherbina, N. A. Trintina, A. M. Shtimmerman, S. M. Ishcheryakov DEVELOPMENT OF MACHINE LEARNING ALGORITHMS FOR A RECOMMENDATION SYSTEM FOR SELECTING MUSICAL COMPOSITIONS

The article analyzes the method of singular value decomposition (SVD) as an effective way to build a recommender system. With the development of information technologies and their introduction into public life, there is a need to search for accentuated information in conditions of uncertainty. To solve such problems, recently created intelligent recommendation systems [1]. The popularity of recommendation systems is growing in every segment of goods and services, in particular music.

From a socio-economic point of view, such systems are the main tool for the dissemination of new compositions in the field of music promotes the promotion of these compositions in accordance with the preferences of the target audience and encourages users to purchase new music tracks. In addition, such systems significantly reduce the time and facilitate the search for appropriate musical compositions under conditions of uncertainty.

The main problem in developing machine learning algorithms is the lack of an individual approach to each user. All recommendations are based on the statistical behavior of the majority, resulting in a percentage of people who do not receive recommendations that match their personal preferences. In the case of a separate analysis of each of the users and the implementation of recommendations in accordance with their personal use of Internet resources, the number of quality and more accurate proposals in the list of recommendations would increase significantly.

Machine learning methods are effectively used to build recommendation systems, namely: the k-nearest neighbors method, the Bayesian algorithm and the singular matrix decomposition method. Among these methods, the SVD method is the most widely used in practice. This method is used to reduce the number of non-significant data set factors. Factors in recommendation systems are properties that describe the user or subject. In music recommendation systems, this can be a genre. SVD reduces the dimension of the matrix by removing its hidden factors.

Keywords: matrix; machine learning; recommendation system.