

МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИЗНАЧЕННЯ РЕЙТИНГУ ФІЛЬМІВ НА ОСНОВІ ДАНИХ ПРО ПОПЕРЕДНІ ОЦІНКИ КОРИСТУВАЧА

Войтюк І.Ф.¹⁾, Павшок Ю.В.²⁾

Західноукраїнський національний університет

^{1)к.т.н., доцент; 2) магістрант}

I. Постановка проблеми

Протягом останніх декількох десятиліть при зростанні веб-сервісів системи рекомендацій зайняли важливе місце в нашому онлайн житті. Рекомендаційні системи пропонують користувачам інформацію на основі даних про їх попередні дії. Головною перевагою цих систем є те, що вони навчаються автоматично, як тільки отримують нову інформацію про уподобання користувача. Тому, чим більше користувач взаємодіє із системою (кількість вподобань, переглядів певних даних у системі), тим кращі пропозиції (більш наближені до інтересу користувача) робитиме система.

II. Мета роботи

Метою дослідження є розробка математичного забезпечення для визначення рейтингу фільмів, які користувач ще не бачив на основі даних про фільми, які він оцінив раніше.

III. Особливості визначення рейтингу фільму

Припустимо, існує інформація про те, скільки відсотків певного жанру становить фільм (наприклад, 80% - мелодрама та 20% - детектив) і оцінки (наприклад, 1-5), які ставлять користувачі [1]. Враховуючи рейтинги користувачів і з яких жанрів (у відсотках) складається фільм, потрібно дізнатися, наскільки користувачам сподобалися жанри фільмів або яка тенденція уподобань користувачів щодо жанрів на основі їх попередніх оцінок. Припустимо, що відомо дані про оцінки користувачів і їхні вподобання щодо різних жанрів [2]. На основі цих даних потрібно дізнатися жанр фільму. Якщо існують дані про відсотковий вміст різних жанрів (бойовик, жахи, фентезі) у певному фільмі, на основі цієї інформації можна передбачити рейтинг. Зокрема, слід з'ясувати, як певний жанр впливає на оцінку користувачів. Тобто потрібно дізнатися вагові коефіцієнти (θ) кожного жанру для певних уподобань користувача. Таким чином опишемо формулу для визначення рейтингу фільму:

$$R_m = \sum_{i=1}^n (X_i^1 \cdot \theta^1 + X_i^2 \cdot \theta^2 + \dots + X_i^k \cdot \theta^k), \quad (1)$$

де n – кількість фільмів, k – кількість жанрів, θ^j – ваговий коефіцієнт жанру j (потрібно знайти),

X_i^j – відсотковий вміст жанру j i -го фільму (відомо), R_m – оцінка користувача m .

Коли буде знайдений ваговий коефіцієнт (θ), тоді можна знайти рейтинги для всіх фільмів, які користувач ще не оцінив. Тепер потрібно застосувати вагові коефіцієнти (θ) та типи жанрів (X), до рівняння, поданого вище. Припустимо, що відомо ваговий коефіцієнт (θ) або вподобання користувачів, але відсутній опис фільмів. Тоді, вагові коефіцієнти (θ) замінимо жанрами. Потрібно буде знайти відсотковий вміст жанру j i -го фільму (X_i^j). Отримаємо рівняння таке ж, як і вище, але вагові коефіцієнти (θ) не є змінними, хоча вони відомі. Отож, коли з'ясуємо відсотковий вміст певного жанру (X) у фільмі, тоді буде можливість знайти оцінки користувача для всіх фільмів, які він ще не оцінив. Тепер, окрім відомих уподобань користувача, буде відомо також детальну інформацію (жанри) про фільми. Тому можна легко визначити, чи наближений певний фільм до уподобань користувача, чи ні. Для цього потрібно застосувати вагові коефіцієнти (θ) та типи жанрів (X) до рівняння, поданого вище. Далі необхідно об'єднати два випадки, що і є метою майбутніх досліджень.

Висновок

У роботі наведено частину розробленого математичного апарату для визначення рейтингу фільмів, які користувач не бачив і не оцінював раніше на основі жанрів фільмів та попередніх вподобань користувача.

Список використаних джерел

1. Movie Lenssmall dataset: 100,000 ratings and 3,600 tag application applied to 9,000 movies by 600 users [Електронний ресурс] – Режим доступу до ресурсу: <http://files.grouplens.org/datasets/movielens/ml-latest-small-README.html>
2. Su X. A Survey of Collaborative Filtering Techniques / X. Su, T. M. Khoshgoftaar. // Advances in Artificial Intelligence. – 2009. – №2009 (4). – P. 11–14.