***IMPROVEMENT OF PEARSON CORRELATION TO OVERCOME THE MATRIX SPARSITY PROBLEM IN RECOMMENDER SYSTEMS***

|  |
| --- |
| Sherstuik Ihor, Holubiev Leontii  Igor Sikorsky Kyiv Polytechnic Institute  Kiev, Ukraine  igor.sherstuk8229@gmail.com, golubev.leontiy@lll.kpi.ua |
| ***Abstract.* The article examines the issue of reduced recommendation accuracy under conditions of rating matrix sparsity. An improved approach to Pearson correlation is proposed by introducing a regularization factor that accounts for the number of common ratings between users. Experimental results demonstrate the effectiveness of the method in cases with 30–55% missing values. The proposed solution leads to improved accuracy metrics and reduced prediction errors.**  ***Keywords*: *ecommender system, Pearson correlation, regularization, matrix sparsity, prediction accuracy.*** |

***УДОСКОНАЛЕННЯ КОРЕЛЯЦІЇ ПІРСОНА ДЛЯ ПОДОЛАННЯ ПРОБЛЕМИ РОЗРІДЖЕННОСТІ МАТРИЦІ В РЕКОМЕНДАЦІЙНИХ СИСТЕМАХ***

Шерстюк Ігор, Голубєв Леонтій

КПІ імені Ігоря Сікорського

Київ, Україна

igor.sherstuk8229@gmail.com, golubev.leontiy@lll.kpi.ua

***Анотацiя*. У статтi розглянуто проблему зниження точностi рекомендацiй в умовах розрiдженостi матрицi. Запропоновано вдосконалення методу кореляцiї Пірсона шляхом введення регуляризацiйного множника, що враховує кiлькiсть спiльних оцiнок. Експериментально доведено ефективнiсть пiдходу при 30–55% вiдсутнiх значень. Метод забезпечує покращення метрик точностi та зменшення похибок.**

***Ключові слова: рекомендацiйна система, кореляцiя Пірсона, регуляризацiя, розрiдженiсть матрицi, точнiсть рекомендацiй.***

Вступ

У сучасних рекомендаційних системах колаборативна фільтрація є одним з найефективніших і найпоширеніших методів персоналізації. Проте однією з її основних проблем залишається розрідженість матриці оцінок, яка призводить до погіршення якості рекомендацій [1]. Зокрема, класичний підхід на основі кореляції Пірсона дає нестабільні або хибні результати при недостатній кількості спільно оцінених об'єктів [2]. Це створює потребу в удосконаленні методу з урахуванням обмеженості даних.Пірсона.

Математичні моделі

У колаборативній фільтрації персоналізовані рекомендації формуються на основі схожості між користувачами. Одним з базових підходів є використання коефіцієнта кореляції Пірсона, який оцінює ступінь подібності між користувачами на основі спільно оцінених об’єктів [1].

Класична формула має вигляд:

(1)

де – оцінка користувача  для об’єкта ;

  – середнє значення оцінок користувача *u*;

  – множина об’єктів, оцінених обома користувачами  та .

Цей метод працює добре, коли є достатньо даних, однак при високій розрідженості матриці оцінок виникає серйозна проблема: навіть незначна кількість спільних оцінок може призводити до високої, але хибної схожості [3]. У таких умовах класична формула стає ненадійною.

Щоб зменшити вплив подібних похибок, у роботі запропоновано вдосконалити цей підхід шляхом введення регуляризації, яка враховує кількість спільних оцінок при обчисленні схожості [4]. Регуляризована формула виглядає так:

 (2)

де: – кількість спільно оцінених об’єктів;

 – регуляризаційний параметр, що визначає ступінь згладжування.

Таким чином, чим менше спільних оцінок між користувачами, тим сильніше згладжується значення схожості. Це дозволяє зменшити вплив випадкових або статистично ненадійних збігів.

На основі скоригованої схожості обчислюється прогнозована оцінка для користувача u щодо об'єкта i:

 (3)

– множина користувачів, які оцінили об’єкт та мають схожість із користувачем .

Запропонована математична модель дозволяє сформувати більш стійкі та надійні рекомендації у випадках, коли інформація про користувачів є частково відсутньою, а матриця оцінок має високий ступінь розрідженості.

Результати експериментальних досліджень

У межах дослідження було проведено серію експериментів із моделюванням різних рівнів розрідженості даних. Матриця оцінок складалася з 50 користувачів і 20 об’єктів, де поступово збільшувався відсоток відсутніх значень — від 10% до 55%. Основна увага приділялася оцінці якості рекомендацій у випадках, коли частка пропущених оцінок перевищувала 30%, оскільки саме в цих умовах класичний метод Пірсона демонструє найменшу стабільність [3].

На рис. 1. чітко спостерігається тенденція: зі зростанням рівня розрідженості показники MAE і RMSE для класичного підходу зростають значно швидше, ніж для регуляризованої версії. Найбільш виражена різниця фіксується після позначки у 40% пропущених оцінок. У цьому діапазоні регуляризований метод утримує більш плавне зростання похибки, тоді як класичний втрачає точність різко і нерівномірно.

Ще один помітний ефект — це стабільність метрик Precision і F1-score при зміні рівня відсутніх даних. Якщо для класичного алгоритму ці значення знижуються вже при 30–35% пропусків, то у вдосконаленого варіанту спостерігається стійкий результат навіть при 50% порожніх комірок. Це свідчить не лише про поліпшення абсолютних значень похибки, але й про більш надійне функціонування алгоритму за умов нестачі інформації.

Особливу увагу було приділено впливу параметра регуляризації λ. Як показали експерименти, надто малі значення цього параметра не усувають проблему нестабільності, тоді як надто великі можуть призводити до заниження оцінки схожості навіть у релевантних парах. Оптимальним виявився діапазон λ від 10 до 40, де баланс між надійністю і чутливістю алгоритму є найбільш ефективним.

Загалом, результати свідчать про помітну зміну динаміки похибок і точності залежно від розрідженості, причому саме вдосконалений метод демонструє більш передбачувану поведінку, що особливо важливо в практичних сценаріях з неповними даними.

Рис. 1. Порівняння критеріїв оцінки рекомендацій для кореляції Пірсона та удосконаленого методу для різної частки відсутніх оцінок

Висновки

Запропонований у роботі метод удосконалення кореляції Пірсона шляхом регуляризації демонструє підвищену стійкість до проблеми розрідженості матриці. Завдяки врахуванню кількості спільно оцінених об’єктів в обчисленні схожості вдалося зменшити похибки та покращити якість рекомендацій. Отримані результати підтверджують доцільність застосування цього підходу у практичних системах, де дані є неповними або частково відсутні.

##### Лiтература

1. Anwar T., Uma V. Comparative study of recommender system approaches and movie recommendation using collaborative filtering // Int. J. Syst. Assurance Eng. Manage. – 2021. – Vol. 12, No. 3. – P. 426–436.
2. Haruna K., Akmar I. M., Damiasih D. et al. A collaborative approach for research paper recommender system // PLoS ONE. – 2017. – Vol. 12, No. 10. – e0184516.
3. Lin K., Sonboli N., Mobasher B., Burke R. Calibration in collaborative filtering recommender systems: a user-centered analysis // Proc. 31st ACM Conf. on Hypertext and Social Media. – 2020. – P. 197–206.
4. Kumar P., Gupta M., Rao C. et al. A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems // Int. J. Recent and Innovation Trends in Computing and Communication. – 2023. – Vol. 11. – P. 184–192.