A NEW CLUSTERING-BASED CF (CBCF) METHOD USING AN INCENTIVIZED/PENALIZED USER (IPU) MODEL

PAPPURI KOTESWARARAO[1], A.RAMESH[2]

[1]Student[CSE], NOVA COLLEGE OF ENGINEERING AND TECHNOLOGY, A.P., India.

[2] ASSISTANT PROFESSOR & HEAD OF THE DEPARTMENT, Dept of CSE, NOVA COLLEGE OF ENGINEERING AND TECHNOLOGY, A.P., India.

ABSTRACT:

The purpose of clustering based CF (CBCF) with the IPU model is to improve recommendation performance such as precision, recall, and F1 score by carefully exploiting different preferences among users. Specifically, we formulate a constrained optimization problem in which we aim to maximize the recall (or equivalently F1 score) for a given precision. To this end, users are divided into several clusters based on the actual rating data and Pearson correlation coefficient. Afterward, we give each item an incentive/penalty according to the preference tendency by users within the same cluster. Our experimental results show a significant performance improvement over the baseline CF scheme without clustering in terms of recall or F1 score for a given precision.

KEYWORDS: collaborative filtering (CF), clustering based CF (CBCF), incentivized/penalized user (IPU) model.

INTRODUCTION

collaborative filtering (CF) is one of the most prominent and popular techniques used for recommender systems. CF methods are generally classified into memory-based CF and model-based CF. In model-based CF, training datasets are used to develop a model for predicting user preferences. Different machine learning techniques such as Bayesian networks, clustering, and rule-based approaches can also be utilized to build models. An alternating least squares with weighted λ-regularization (ALS-WR) scheme is a representative example of model-based CF. ALS-WR is performed based on a matrix factorization algorithm and is tolerant of the data sparsity and scalability [6], [7]. The main advantages of model-based CF are an improvement of prediction performance and the robustness against the data sparsity. However, it has some shortcomings such as an expensive cost for building a model [5]. On the other hand, memory-based CF does not build a specific model, but directly computes the similarity between users or items using the entire rating matrix or its samples. Hence, memory-based CF is easy to implement and effective to manage. However, it has also some drawbacks such as dependence on human ratings, performance decrement when data are sparse, and disability of recommendation for new users (i.e., cold-start users) and items [5]. Memory-based CF approaches are again classified into user-based CF and item-based CF. The main ideas behind the user-based CF and item-based CF approaches are to find the user similarity and the item similarity, respectively, according to the ratings (or preferences). After finding similar users, called neighbors, user-based CF recommends the top-N most preferable items that an active user has not accessed yet. User-based CF has limitations related to scalability, especially when the number of users is much larger than the number of items. Item-based CF was proposed to mitigate this scalability problem, but cannot still entirely solve the problem when the numbers of users and items are large. Despite such limitations, CF has been employed as one of the most representative recommender systems leveraged in online commerce.

LITERATURE SURVEY

1.G. Adomavicius and A. Tuzhilin,

In this work, we will introduce the recommended field field and the years that is often presented for three senior students. The content above content, the Sisergy, cross, the cross is pointing out that they will be fucked. The current expanding system due to the possibility of expanded one can increase the energy that provides for increasing amounts.

These developments include, among others, understand the customer understanding and things that combine information that has a relationship. Positions presented, supporting for evaluation of evaluation and set the type of advice that is more efficient and use it.

2.G. Linden, B. Smith, and J. York,

Suggestion calculation are the most employed in online marketing and use clients to summarize what is presented. Most apps use what customers have bought and explicitly evaluate them to fix issues, but also what they see, domain data, preferences, favorites, etc., use different features. Amazon.com uses query statistics to design an online store for all customers. The store changes completely depending on the customer’s interest and the theme of the program. For programmers, toys for other mothers.

There are three common approaches to solving the problem presented: traditional filtering, organizational modeling, and research strategies. Here we will compare this strategy with mathematics. Unlike our traditional custom catalog, our online catalogs are automatically updated based on the number of customers and the number of content items. Our analysis results in step-by-step recommendations, we weighed them in a database, and we offer quality recommendations.

PROBLEM DEFINTION

On the other hand, several companies, e.g., Pandora Internet Radio, Netflix, and Artsy, have developed their own clustering-based recommendation methods, called Music Genome Project, Micro-Genres of Movies, and Art Genome Project, respectively. These clustering-based recommendation methods have successfully led to satisfactory performance, but the processing cost for clustering is very expensive.

Unlike the aforementioned clustering-based recommendation methods that take long processing time to recommend items, we aim to design a simple but novel clustering-based CF (CBCF) method only with ratings given by users, which is thus easy to implement.

On the other hand, memory-based CF does not build a specific model, but directly computes the similarity between users or items using the entire rating matrix or its samples. Hence, memory-based CF is easy to implement and effective to manage. However, it has also some drawbacks such as dependence on human ratings, performance decrement when data are sparse, and disability of recommendation for new users and items.

Disadvantages of Existing system

- There is no accurate analysis on lack of Classification.

PROPOSED APPROACH

Advantages of Proposed System

- An easy-to-implement CBCF method using the IPU model is proposed to further enhance the performance related to UX.
- To design our CBCF method, we first formulate a constrained optimization problem, in which we aim to maximize the recall (or equivalently F1 score) for a given precision.
- We numerically find the amount of incentive/penalty that is to be given to each item according to the preference

tendency by users within the same cluster.

- We evaluate the performance of the proposed method via extensive experiments and demonstrate that F1 score of the CBCF method using the IPU model is improved compared with the baseline CF method without clustering, while recall for given (fixed) precision can be significantly improved by up to about 50%.

CONCLUSION:

Specifically, in the proposed CBCF method, we formulated a constrained optimization problem in terms of maximizing the recall (or equivalently F1 score) for a given precision. To this end, clustering was applied so that not only users are divided into several clusters based on the actual rating data and Pearson correlation coefficient but also an incentive/penalty is given to each item according to the preference tendency by users within a same cluster. As a main result, it was demonstrated that the proposed CBCF method using the IPU model brings a remarkable gain in terms of recall or F1 score for a given precision. A possible direction of future research in this area includes the design of a new clustering-based CF method by exploiting the properties of model-based CF approaches (e.g., matrix factorization).

REFERENCES:

[1] G. Adomavicius and A. Tuzhilin, ''Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,'' IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005.

[2] G. Linden, B. Smith, and J. York, ''Amazon.com recommendations: Itemto-item collaborative filtering,'' IEEE Internet Comput., vol. 7, no. 1, pp. 76–80, Jan./Feb. 2003.

[3] Y. Koren, R. Bell, and C. Volinsky, ''Matrix factorization techniques for recommender systems,'' Computer, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[4] J. A. Konstan and J. Riedl, ''Recommender systems: From algorithms to user experience,'' User Model. User-Adapted Interact., vol. 22, nos. 1–2, pp. 101–123, Apr. 2012.

[5] X. Su and T. M. Khoshgoftaar, ''A survey of collaborative filtering techniques,'' Adv. Artif. Intell., vol. 2009, pp. 1–19, Jan. 2009, Art. no. 421425. [6] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, ''Large-scale parallel collaborative filtering for the Netflix prize,'' in Proc. Int. Conf. Algorithmic Appl. Manage., Shanghai, China, Jun. 2008, pp. 337–348.

[7] Y. Hu, Y. Koren, and C. Volinsky, ''Collaborative filtering for implicit feedback datasets,'' in Proc. 8th IEEE Int. Conf. Data Mining, Pisa, Italy, Dec. 2008, pp. 263–272.

[8] Y. Cai, H.-F. Leung, Q. Li, H. Min, J. Tang, and J. Li, ‘‘Typicality-based collaborative filtering recommendation,’’ IEEE Trans. Knowl. Data Eng., vol. 26, no. 3, pp. 766–779, Mar. 2014.