International Journal of Research
Available at
https://www.internationaljournalofresearch.org/

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

l-Injection Toward Effective Collaborative Filtering

[1]CHOPPARI KARUNA, [2]Dr VIJAY REDDY MADIREDDY

M.Tech, Department of Computer Science and Engineering,

Associate Professor, Department of CSE,

Swami Ramananda Tirtha Institute of Science & Technology, Nalgonda.

Abstract - We develop a novel framework, named as l-injection, to address the sparsity problem of recommender systems. By carefully injecting low values to a selected set of unrated user-item pairs in a user-item matrix, we demonstrate that top-N recommendation accuracies of various collaborative filtering (CF) techniques can be significantly and consistently improved. We first adopt the notion of pre-use preferences of users toward a vast amount of unrated items. Using this notion, we identify uninteresting items that have not been rated yet but are likely to receive low ratings from users, and selectively impute them as low values. As our proposed approach is method-agnostic, it can be easily applied to a variety of CF algorithms. Through comprehensive experiments with three real-life datasets (e.g., Movie lens, Ciao, and Watcha), we demonstrate that our solution consistently and universally enhances the accuracies of existing CF algorithms (e.g., item-based CF, SVD-based CF, and SVD++) by 2.5 to 5 times on average. Furthermore, our solution improves the running time of those CF methods by 1.2 to 2.3 times when its setting produces the best accuracy. The datasets and codes that we used in the experiments are available at: https://goo.gl/KUrmip

1. INTRODUCTION

The goal of recommender systems (RS) is to suggest appealing items (e.g., movies, books, or news articles) to a user by analyzing her prior preferences. As a large number of online applications use RS as a core component, improving the quality of RS becomes a critically important problem to businesses. Among existing solutions in RS, in particular, collaborative filtering (CF) methods (e.g., [2], [3], [4], [5], [6], [7]) have been shown to be widely effective. Based on the past behavior of users such as explicit user ratings and implicit click logs, CF methods exploit the similarities between users' behavior patterns.

However, when the fraction of known ratings in a rating matrix R is overly small (so-called data sparsity problem), CF methods tend to suffer. For an R withmusers and n items, if we assume that each user has rated k items on average, the fraction of rated items in R is k n (= mk mn). Asymptotically, such a fraction of rated items in R is extremely small (i.e., k n). It is common for an e-business to sell millions of items with a very long tail, and many users rate very few items (i.e., cold-start users). The goal of this work is to mitigate such a data sparsity problem to improve top- N recommendation accuracies of CF methods. Our proposal is based on the following hypothesis in CF:

TABLE 1
Rating distributions of three real-life datasets.

Dataset	Low ratings (1 or 2)	High ratings (3, 4, or 5)
Movielens	17%	83%
Ciao	10%	90%
Watcha	13%	87%

Hypothesis 1. *Filling some values into empty cells, i.e., unrated items, in a rating matrix R can improve the accuracy of CF methods for top-N recommendation.*

We first argue that ratings in R be often a reflection of the satisfaction of users. Therefore, users tend to rate (high) only the items that they like, and those who are dissatisfied tend not to rate items in R. Corroborating this point, Table 1 illustrates severe imbalance between low (i.e., 1 or 2) and high (i.e., 3, 4, or 5) ratings from three real-life datasets that we used in our experiments. Note that only a small fraction (i.e., 10–17%) of ratings are low values. Then, a natural question to raise is: how can we identify the unknown opinions of those users who were dissatisfied with and did not leave ratings for items?

To answer this question, note that unrated items in R can be classified into three different types: (1) unrated items whose existence users were not aware of, (2) unrated items that users knew and purchased but did not rate, and (3) unrated items that users knew but did not like and thus did not purchase. We note that the unrated items of the third type, called uninteresting items (denoted by Iun), clearly indicate users' latent negative preferences on them. Therefore, it is better not to recommend those uninteresting items. In order to identify such uninteresting items, we propose to use a new notion of pre-use preference, i.e., an impression of items before purchasing and using them. That is, by definition, uninteresting items indicate the items with low pre-use preferences. Unfortunately, the ratings in R do not indicate pre-use preferences but the preferences after using the items, called post-use preference. Based on this novel notion of pre-use preference and uninteresting item, we develop a solution that consists of three steps: (1) infer the pre-use preferences of unrated items by solving the one-class collaborative filtering (OCCF) problem [8], [9], (2) assign "low" values to uninteresting items in R, yielding an augmented matrix L, and (3) apply existing CF methods to L, instead of R, to recommend top-N appealing items. This simple-yet-novel imputation solution significantly alleviates the data sparsity problem by augmenting R. Extending our prior work [1], in this work, we develop a more general l-injection to infer different user preferences for uninteresting items for users, and show that l-injection mostly outperforms 0-injection in [1]. The proposed l-injection approach can improve the accuracy of top-N recommendation based on two strategies: (1) preventing uninteresting items from being included in the top-N recommendation, and (2) exploiting both uninteresting and rated items to predict the relative preferences of unrated items more accurately. With the first strategy, because users are aware of the existence of uninteresting items but do not like them, such uninteresting items are likely to be false positives if included in top-N recommendation.

2. Literature Survey

Improving the Accuracy of Top-N Recommendation using a Preference Model:

In this paper, we study the problem of retrieving a ranked list of top-N items to a target user in recommender systems. We first develop a novel preference model by distinguishing different rating patterns of users, and then apply it to existing collaborative filtering (CF) algorithms. Our preference model, which is inspired by a voting method, is well-suited for representing qualitative user preferences. In particular, it can be easily implemented with less than 100 lines of codes on top of existing CF algorithms such as user-based, item-based, and matrix-factorization based algorithms. When our preference model is combined to three kinds of CF algorithms, experimental results demonstrate that the preference model can improve the accuracy of all existing CF algorithms such as ATOP and NDCG@25 by 3%–24% and 6%–98%, respectively.

Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions

This paper presents an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. This paper also describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multcriteria ratings, and a provision of more flexible and less intrusive types of recommendations.

Matrix factorization techniques for recommender systems

In this aspect, content filtering is superior. The two primary areas of collaborative

International Journal of Research
Available at
https://www.internationaljournalofresearch.org/
ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

filtering are the neighborhood methods and latent factor models. Neighborhood methods are centered on computing the relationships between items or, alternatively, between users. The item oriented approach evaluates a user's preference for an item based on ratings of "neighboring" items by the same user. A product's neighbors are other products that tend to get similar ratings when rated by the same user. For example, consider the movie Saving Private Ryan. Its neighbors might include war movies, Spielberg movies, and Tom Hanks movies, among others. To predict a particular user's rating for Saving Private Ryan, we would look for the movie's nearest neighbors that this user actually rated. As Figure 1 illustrates, the user-oriented approach identifies like-minded users who can complement each other's ratings.

3. OVERVIEW OF THE SYSTEM

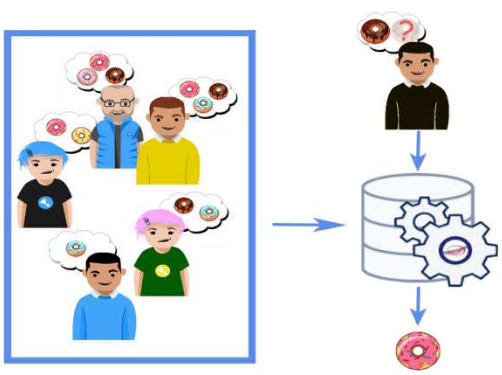

Fig 3.1 System Architecture

3.1 EXISTING SYSTEM

- Among existing solutions in recommender systems RS, in particular, collaborative filtering (CF) methods have been shown to be widely effective. Based on the past behavior of users such as explicit user ratings and implicit click logs, CF methods exploit the similarities between users' behavior patterns.
- Most CF methods, despite their wide adoption in practice, suffer from low accuracy if most users rate only a few items (thus producing a very sparse rating matrix), called the data sparsity problem. This is because the number of unrated items is significantly more than that of rated items.
- To address this problem, some existing work attempted to infer users' ratings on unrated items based on additional information such as clicks and bookmarks.

3.2 DISADVANTAGES OF EXISTING SYSTEM

- These works require an overhead of collecting extra data, which itself may have another data sparsity problem.
- 0-injection simply considers all uninteresting items as zero, it may neglect to the characteristics of users or items. In contrast, l-injection not only maximizes the impact of filling missing ratings but also considers the characteristics of users and items, by imputing uninteresting items with low peruse preferences.

3.3 PROPOSED SYSTEM

- In this work, we develop a more general l-injection to infer different user preferences for uninteresting items for users, and show that l-injection mostly outperforms 0-injection.

International Journal of Research
Available at
https://www.internationaljournalofresearch.org/

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

- The proposed l-injection approach can improve the accuracy of top-N recommendation based on two strategies: (1) preventing uninteresting items from being included in the top-N recommendation, and (2) exploiting both uninteresting and rated items to predict the relative preferences of unrated items more accurately.
- With the first strategy, because users are aware of the existence of uninteresting items but do not like them, such uninteresting items are likely to be false positives if included in top-N recommendation. Therefore, it is effective to exclude uninteresting items from top-N recommendation results.
- Next, the second strategy can be interpreted using the concept of typical memory based CF methods.

3.4 ADVANTAGES OF PROPOSED SYSTEM:

- We introduce a new notion of uninteresting items, and classify user preferences into pre-use and post-use preferences to identify uninteresting items.
- We propose to identify uninteresting items via peruse preferences by solving the OCCF problem and show its implications and effectiveness.
- We propose low-value injection (called l-injection) to improve the accuracy of top-N recommendation in existing CF algorithms.
- While existing CF methods only employ user preferences on rated items, the proposed approach employs both peruse and post-use preferences. Specifically, the proposed approach first infers pre-use preferences of unrated items and identifies uninteresting items

4. SYSTEM DESIGN

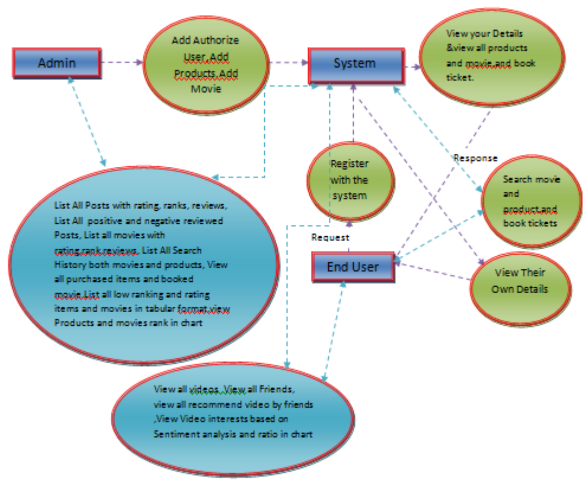

Fig 4.1: Data Flow Diagram

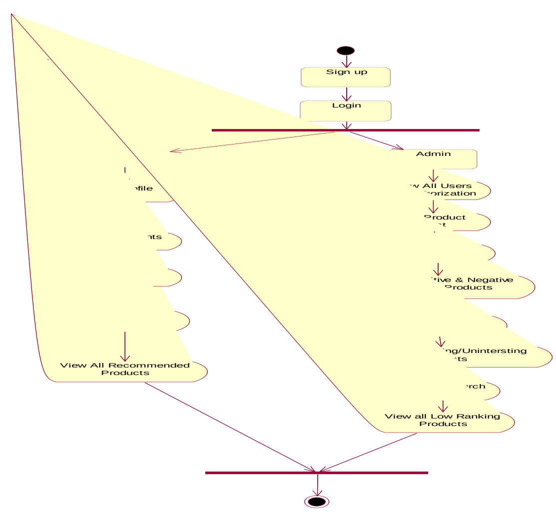

Fig 4.2: Activity Diagram

5. OUTPUT SCREEN SHOTS

International Journal of Research
Available at
https://www.internationaljournalofresearch.org/

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

Fig 5.1: Home Page

Fig 5.2: View All Users Page

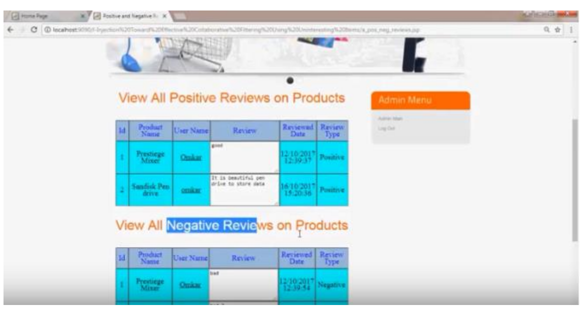

Fig 5.3: View all Positive and Negative Reviews Page

6. CONCLUSION AND FUTURE SCOPE

In this paper, we proposed a novel approach, l-injection, for uninteresting items by using a new notion of pre-use preferences. This approach not only significantly alleviates the data sparsity problem but also effectively prevents those uninteresting items from being recommended. Because the proposed approach is method-agnostic, it can be easily applied to a wide variety of existing CF methods. Through comprehensive experiments, we successfully demonstrated that the proposed approach is effective and practical, dramatically improving the accuracies of existing CF methods (e.g., item-based CF, SVD-based CF, and SVD++) by 2.5 to 5 times. Furthermore, our approach improves the running time of those CF methods by 1.2 to 2.3 times when its setting produces the best accuracy.

7. REFERENCES

[1] W. Hwang, J. Parc, S. Kim, J. Lee, and D. Lee, "Told you i didn't like it: Exploiting uninteresting items for effective collaborative filtering," in Proc. of IEEE ICDE, 2016, pp. 349–360.

[2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[3]Y. Koren et al., "Matrix factorization techniques for recommender systems," IEEE Computer, vol. 42, no. 8, pp. 30–37, 2009.

[4] B. Sarwar et al., "Item-based collaboration filtering recommendation algorithms," in Proc. of IEEE WWW, 2001, pp. 285–295.

[5]S. Zhang et al., “Using singular value decomposition approximation for collaborative filtering,” in Proc. of IEEE CEC, 2005, pp. 257–264.

[6] P. Resnick et al., “Grouplens: an open architecture for collaborative filtering of netnews,” in Proc. of ACM CSCW, 1994, pp. 175–186.