Information Fusion 46 (2019) 141–146

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

EARS: Emotion-aware recommender system based on hybrid information fusion

Yongfeng Qian[a,b], Yin Zhang[*,c], Xiao Ma[c], Han Yu[c], Limei Peng[d]

[a] School of Computer Science, China University of Geosciences, Wuhan, China
[b] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
[c] School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, China
[d] School of Computer Science and Engineering, Kyungpook National University, South Korea

ARTICLE INFO

Keywords:
Hybrid information fusion
Emotion-aware intelligent system
Recommender systems
Matrix factorization

ABSTRACT

Recommender systems suggest items that users might like according to their explicit and implicit feedback information, such as ratings, reviews, and clicks. However, most recommender systems focus mainly on the relationships between items and the user's final purchasing behavior while ignoring the user's emotional changes, which play an essential role in consumption activity. To address the challenge of improving the quality of recommender services, this paper proposes an emotion-aware recommender system based on hybrid information fusion in which three representative types of information are fused to comprehensively analyze the user's features: user rating data as explicit information, user social network data as implicit information and sentiment from user reviews as emotional information. The experimental results verify that the proposed approach provides a higher prediction rating and significantly increases the recommendation accuracy.

1. Introduction

Information overload (information overload) is an increasing problem that cannot be ignored. Recommendation systems were developed to reduce the time that users spend browsing useless information. A recommendation system recommends interests and merchandise to the user by observing the user's interest characteristics and selection behavior, and even provides personalized services [1]. In recent years, research on recommendation systems has expanded, and the huge amount of accompanying data has brought new challenges for recommendation systems.

With the development of big data, cloud computing, mobile computing and other advanced information technologies [2], many types of data are used in recommender systems, and the information that is advantageous to the system must be identified. Therefore, researchers have begun to integrate all types of information and make recommendations based on the fused information. However, the fused information comes from different dimensions, such as personal information, social information, emotional information and whether the activity area of users is explicit [3,4]. Various works have proved that fusion information can significantly improve the availability of the recommendation system [5,6].

The core resources that support the recommendation system are the user's historical behavior data, including explicit feedback and implicit feedback [7]. Most explicit feedback-based collaborative filtering recommendation systems are based on user ratings and trust information to improve the accuracy of recommendation [8]. However, this will miss a lot of implicit feedback data. Implicit feedback data is more common than the additional inputs required for explicit feedback data, and its collection costs are low and do not affect the user experience. To address the shortcomings of explicit feedback data recommendation systems, recommendation systems based on implicit feedback data have been introduced. However, implicit feedback data can express the user's positive feedback but are less able to express the user's negative feedback. Solving the problem of the lack of negative feedback is thus very important.

At present, recommendation systems based on implicit feedback data face the following three challenges:

1. **Sample imbalance**. In implicit feedback data, there are usually only positive feedback and no negative feedback. In contrast to explicit feedback data, which directly reflect the tendency of the user's likes and dislikes, implicit data include "selected" and "non-selected" categories. Although the & quot; selected & quot; may indicate a user & apos; s tendency, the & quot; unselected & quot; cannot directly represent a user & apos; s negative tendency because "not selected" includes not only products in which the user is not really

* Corresponding author.
E-mail addresses: yongfengqian@ieee.org (Y. Qian), yin.zhang.cn@ieee.org (Y. Zhang), cindyma@hust.edu.cn (X. Ma), hay@zuel.edu.cn (H. Yu), auroraplm@ajou.ac.kr (L. Peng).

https://doi.org/10.1016/j.inffus.2018.06.004
Received 13 February 2018; Received in revised form 26 May 2018; Accepted 17 June 2018
Available online 19 June 2018

interested but also products in which the user is interested but has not yet found. This lack of positive examples adds difficulty to the model.
2. **Noise**. In contrast to an explicit rating, in implicit feedback data, the user can produce a lot of noise due to various misuses.
3. **Large scale**. The actual scenario of the recommendation model will involve large-scale data, which requires the model to have sufficiently efficient performance and excellent scalability to handle vast amounts of data.

In this paper, the model proposed in the above problem is used to model the implicit feedback data directly by converting the recommended task into the probability of maximizing the probability of user selection behavior. The model will be able to "unselect" the information that is fully utilized while avoiding the introduction of negative cases and noise to improve recommendation quality. Specifically, this article makes the following contributions: (1) it develops an effective approach to fuse social information, rating information and emotional information for recommender systems; (2) assisted by the hybrid features from implicit and explicit data, the performance of the proposed recommender system is significantly improved.

2. Related work

2.1. Recommender systems

Traditional recommendation systems are difficult to apply directly to implicit feedback data because the data contain only positive feedback from the user and lack negative feedback. In [9], Pan et al. defines this issue as a single type of collaborative filtering (One Class Collaborative Filtering, OCCF), which is generally summarized as an imbalance problem (Unbalanced Class Problem, UCP) [10]. The main way to deal with this problem is to introduce negative feedback in the following three ways:

1. The specificity of the application environment and addition of negative artificial rules [11]. For example, an item that is forwarded in the other two but are not forwarded between microblogs can be considered negative samples as users browse them, but there is no forwarding, which may indicate a lack of interest. This method relies on domain knowledge and cannot be extended.
2. Label an unknown sample as a negative sample in a random sample [12]. This method generally assumes that most samples labeled as unknown are negative; thus, most of the randomly sampled samples are negative.
3. The samples labeled as unknown are negative samples, but a lower weight is set. The weight of the sample is negative, reflecting the confidence level of the sample [13]. The disadvantage of this approach is that it cannot guarantee "true".

In fact, these three methods attempt to add negative samples in the experiment but cannot guarantee the authenticity of negative cases. In addition, the introduction of negative samples will increase the burden of training, and the larger scale will affect the efficiency of the experiment. The model proposed in this paper is based on the idea of a probability generation model. By maximizing the observed probability of user feedback to directly select a user to choose the modeling trend, no negative feedback can be trained. This model is therefore applicable to a variety of implicit feedback recommendation scenarios.

2.2. Collaborative filtering

Personalized recommender systems include user-based and content-based collaborative filtering [14]. Among them, the basic idea of user-based collaborative filtering is to analyze a user's general rating of an item, including an analysis of the user's historical situation, preferences, and behavior, and then find the best similarity with those of another user. However, poor scalability and a user data transferring method that is based on filters are problems. To solve these problems, another collaborative filtering method is based on content of the recommendation. This method mainly uses the Bayesian probability model, genetic algorithm and other machine learning methods [15]. With the expansion of this application, collaborative filtering also encounters some problems, which are mainly confined to the following three aspects: sparsity, scalability and synonymy.

In general, the collaborative filtering algorithm searches a large group of people and finds a smaller set who are similar to the target user's preferences. The basic mechanism of collaborative filtering is as follows [16]: (i) based on behavior, preferences and other factors, identify a group of people's preferences for analysis; (ii) using a similarity measure, select a sub-group that has the greatest similarity with the target users; (iii) weight the user inside the sub-group in the calculation; and (iv) use the resulting preference function to make a recommendation to the target user that is more in line with the user's preferences. Commonly used methods are metered similarity, cosine similarity, and Pearson correlation coefficient calculation. Collaborative filtering recommendation systems have many advantages. For example, from the establishment of the model to the preparation of the program, the whole process is very clear; the applicable range is relatively large and includes movie, music, social, commodity and many other aspects; and higher degree. However, shortcomings inevitably remain: in the initial system, there is a lack of basic calculation data to relate to the recommended data, i.e., the cold start problem; when the system is being used to recommend items and for user group secondary filtering, other problems must wait.

2.3. Information fusion

In recent years, with the development of data mining, hidden information can be extracted from all types of data. Using effective information can improve the solution of related problems [17]. In recommender systems, the use of converged information is also becoming popular.

In [18], a Hybrid Multigroup CoClustering recommendation framework integrated information from user-item rating records, user social networks, and item features extracted from the DBpedia knowledge base. The experimental results demonstrated the superior performance of our approach in top-n recommendation in terms of MAP, NDCG, and F1 compared with other clustering-based CF models. Sun et al. [3] developed a probabilistic factor analysis framework, named RMSQ-MF, which has the ability to exploit multi-source information. Cheng et al. [19] integrated personal information, network structure information and social information and then proposed a fusion recommendation framework. Hence, information fusion is available to improve the performance of recommender systems, especially with the assistance of advanced mobile network to acquits various context data [20].

3. System design

3.1. System architecture

Fig. 1 shows the architecture of the proposed recommender system, which consists of the following four components:

1. **Data Collection**: In the proposed scheme, three types of raw data are collected for information fusion, i.e., social network data, rating data and user review data. With the development of crowd-sourced review websites such as Yelp,[1] these data are easy to access.
2. **Information Fusion**: Among the collected data, rating data are

[1] http://www.yelp.com/datasetchallenge/.