Information Sciences 327 (2016) 1–20

Contents lists available at ScienceDirect

Information Sciences

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

Representing conditional preference by boosted regression trees for recommendation

Juntao Liu[a], Chenhong Sui[b], Dewei Deng[a], Junwei Wang[a], Bin Feng[b], Wenyu Liu[b], Caihua Wu[c,*]

[a] 709th Research Institute, China Shipbuilding Industry Corporation, Wuhan 430074, China
[b] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[c] Section of Automatic Command, Huang Pi NCO School, Air Force Early-warning Academy, Wuhan 430345, China

ARTICLE INFO

Article history:
Received 21 January 2015
Revised 30 July 2015
Accepted 3 August 2015
Available online 12 August 2015

Keywords:
Conditional preference
Boosted regression trees
Matrix factorization
Collaborative filtering
Recommender system

ABSTRACT

Widely existing conditional preference is seldom taken into consideration in recommender systems. This may lead to unsatisfying recommendation results. To address this issue, in this paper, we propose to use boosted regression trees to represent conditional preference in recommendation systems, which is more expressive than linear and quadratic function for conditional preference. Compared with the existing conditional preference model, boosted regression trees can process large amounts of data in recommendation systems due to the reasonable storage space and low learning complexity. We integrate boosted regression trees into the framework of matrix factorization, and propose an algorithm combining gradient boosting and coordinate descent to learn the model. The proposed method is evaluated on four real world datasets, and compared with other matrix factorization based state-of-the-art methods. The experimental results show that the proposed method outperforms most of the comparison methods.

1. Introduction

Nowadays, recommender systems are widely deployed in online business systems [21,37]. News [26], friends [50], blog articles [27] and so on are recommended to the users in these systems. How to represent the user's preference to improve the performance of the recommender systems is one of the fundamental problems. It is revealed that most of the users' preference are conditional [32]. This means that user may like different things under different conditions. But most of the existing recommendation methods ignore the conditional preference. For example, the matrix factorization based methods [13,25,40–42], which are the most popular recommendation methods, use linear function to represent the users' preference. However, It is proved that the linear function cannot represent conditional preference [32]. Ignoring the conditional preference may lead to unsatisfying recommendation results.

The existing conditional preference models, including CP-nets [2] and its variations [1,3,16,23], are seldom used in recommender systems because of high storage complexity and computational complexity of the learning algorithms [8,9]. Due to the

* Corresponding author. Tel.: +8613971025396.
E-mail addresses: prolay@163.com (J. Liu), sui6662008@163.com (C. Sui), dengddw@163.com (D. Deng), wjw.20012001@163.com (J. Wang), fengbin@mail.hust.edu.cn (B. Feng), liuwy@hust.edu.cn (W. Liu), wucaihua2009@163.com, wucaihua_1999@yahoo.com.cn (C. Wu).

http://dx.doi.org/10.1016/j.ins.2015.08.001
0020-0255/

huge amount of data in recommender systems, even the computational complexity of the approximate CP-nets learning methods [30,31] is still high.

Recently, Liu et al. [32] propose to use quadratic polynomial to approximate conditional preference. However, in this paper, we find that quadratic polynomial cannot represent conditional preference in some cases. While the regression tree can always represent conditional preference relation. (For more details, please see Example 1 and Theorem 1). That means regression tree is more expressive than linear and quadratic function. On the other hand, regression tree is suitable for recommender systems due to the low storage complexity and low computational complexity of the learning algorithm [12]. For these reasons, we propose to use boosted regression trees, the weighted sum of regression trees, to represent conditional preference in recommender systems, and design an algorithm combining gradient boosting and coordinate descent to learn the proposed model. The proposed method is evaluated on Eipinions [36], MovieLen (1M), MovieLen (100k) and MovieLen+ [17] datasets, and compared with other recommendation methods. The performance of the recommendation methods is measured by the Normalized Discounted Cumulative Gain (NDCG) [20] and the Expected Reciprocal Rank (ERR) [5]. The experimental results show that our method gets more satisfying recommendation results than most of the comparison recommendation methods. Complexity analysis indicates that the proposed recommendation method is efficient and can be used in real life recommender systems.

The main contributions of our work are two-fold. (1) We prove that boosted regression trees can represent conditional preference better than quadratic polynomial. (2) We propose to use boosted regression trees to represent conditional preference in the framework of matrix factorization for recommendation, and combine gradient boosting and coordinate descent to learn the proposed model, which is discontinuous, nonlinear and non-differential.

The rest of this paper is organized as follows. In Section 2, a brief survey of conditional preference and recommendation methods based on matrix factorization is provided. In Section 3, The expressibility of boosted regression trees for conditional preference is investigated. In Section 4, the proposed model and the learning algorithm are introduced. The experimental results are presented and analyzed in Section 5 followed by the conclusions and further work in Section 6.

2. Background and related work

In this section, the related research fields of conditional preference and matrix factorization based recommendation are introduced briefly.

2.1. Conditional preference

CP-net [2], a digraph, is proposed to represent conditional preference. Each node, corresponding to a variable, is annotated with a conditional preference table. The preferential dependency among variables is represented by the graph structure. Since CP-nets is unable to represent the total order, the original model is extended. For example, Brafman et al. [3] add the conditional importance tables into CP-nets and propose TCP-nets (Trade-off Conditional Preference Networks). Wilson [48] proposes the stronger conditional preference statement. In CP-net and its extensions, the number of conditional preference rules is exponential to the number of the feature variables. The space complexity of these models is high. To address this issue, some approximate models are proposed. For example, Domshlak [10] uses the soft constraint to represent conditional preference approximately. Lang and Mengin [23] ignore the conditional preferential dependency, and propose the separable ceteris paribus structures (SCP-structures). Because of the high computational complexity of dominance testing and consistency testing in CP-nets, which are the most commonly used operations in recommender systems, CP-nets and its extensions are seldom used in recommender systems. Here, dominance testing means deciding the preference relation between two given outcomes. Consistency testing means deciding whether a given CP-net is consistent. To reduce the computational complexity, CP-nets is simplified under some assumptions. For example, the conditional preference rule is replaced with the utility function under the assumption of additive independence in UCP-nets [1]. Under the same assumption, GAI-network [16] is proposed. LCP-nets [6] is proposed by integrating TCP-nets [3] and UCP-nets [1]. UCI-nets [11] is proposed under the assumption of conditional independence. Although these models do reduce the computational complexity, the assumptions they rely on are too strict.

In terms of learning algorithm, learning the fixed structure CP-nets is addressed by Boutilier et al. [1]. The learning algorithm for SCP-structures is proposed by Lang and Mengin [23]. The algorithm for generating an acyclic CP-net entailing all samples under the transparency assumption (see Definition 5 in [9]) is proposed by Dimopoulos et al. [9]. The problem of learning CP-nets from inconsistent samples is addressed in [30]. To reduce the computational complexity, an approximate learning method based on hypothesis testing is proposed [31]. Besides the passive learning algorithms above, the active one is also proposed [22].

2.2. Matrix factorization

The matrix factorization based recommendation methods [13,25,40–42] gain great success. They are the most popular collaborative filtering methods. To alleviate the data sparsity problem and the cold-start problem [21], side information, such as the social relation and the tag information, is considered in the matrix factorization based recommendation methods [7,19,24,28,33–35,44,49,52,52,53,55,56].

Recently, some researchers propose to optimize information retrieval measures instead of the sum-of-squares of factorization errors, which is the objective function in the traditional matrix factorization based recommendation methods. These measures