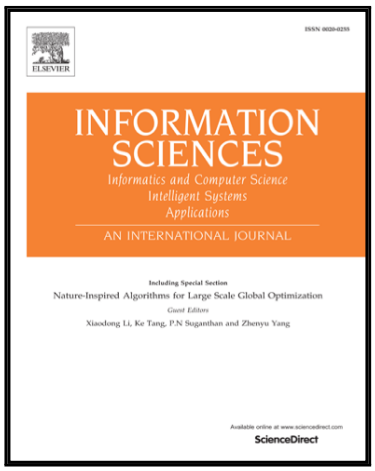

A Hybrid User Similarity Model for Collaborative Filtering

Yong Wang , Jiangzhou Deng , Jerry Gao , Pu Zhang

PII: S0020-0255(17)30862-9
DOI: 10.1016/j.ins.2017.08.008
Reference: INS 13012

To appear in: *Information Sciences*

Received date: 2 April 2017
Revised date: 21 June 2017
Accepted date: 2 August 2017

Please cite this article as: Yong Wang , Jiangzhou Deng , Jerry Gao , Pu Zhang , A Hybrid User Similarity Model for Collaborative Filtering, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.08.008

A Hybrid User Similarity Model for Collaborative Filtering

Yong Wang [a,b,*] Jiangzhou Deng [a] Jerry Gao [b] Pu Zhang [a]

[a] Key Laboratory of Electronic Commerce and Logistics, Chongqing University of Posts and Telecommunications, Chongqing, China

[b] Computer Engineering Department, San Jose State University, San Jose, US

Abstract

In the neighborhood-based Collaborative Filtering (CF) algorithms, the user similarity has an important effect on the result of CF. In order to evaluate the user similarity comprehensively and objectively, we proposed a hybrid model. In the model, an item similarity measure is designed based on the Kullback-Leibler (KL) divergence, which is used as a weight to correct the output of an adjusted Proximity-Significance-Singularity model. Meanwhile, a user preference factor and an asymmetric factor are considered in our model to distinguish the rating preference between difference users and improve the reliability of the model output. The tests on different datasets show that the proposed user similarity model is suitable for the sparse data and effectively improves the prediction accuracy and the recommendation quality.

Keywords: User similarity; Collaborative filtering; Common rated items; Kullback-Leibler divergence; Recommendation algorithm

1. Introduction

Recommendation algorithms have been widely applied to deal with information overload problems in e-commerce sites [19]. The purpose of recommendation algorithms is to constantly learn the prior preference behaviors of users, and models the interaction of both users and items to provide the personalized recommendation for users. Broadly speaking, recommendation algorithms use the following type of filtering: content-based filtering, demographic filtering, collaborative filtering, social-based filtering, context-aware filtering and hybrid filtering [4].

It is well-known that collaborative filtering (CF) is one of the commonly used and successful techniques in recommendation systems. The CF algorithm was first described by Goldberg et al. in 1992 [13]. The basic assumption is the past preference behaviors of users have a significant influence on their future behaviors, and their previous behaviors are basically consistent with future behaviors [12]. The goal of CF is to calculate the unknown ratings based on the known information. In CFs, neighborhood based CF is widely used to recommend items for users.

Neighborhood based CF, also called *K*-nearest neighbors (KNN) method, depends on a simple intuition that an active user may have preference behaviors on some item if this item is appreciated by a set of similar users (neighbors) in recommendation systems [21]. The neighborhood based CF involves three main steps. The first step is to calculate similarity value between two users or items by using a similarity measure. The second step is that finding the nearest neighbors of an active user or item according to the similarity value. The last step is to generate a predicted rating value for an item according to the preference of the neighbors. Moreover, the previous researches have indicated that the similarity measure is an important factor to the prediction accuracy of recommendation algorithm.

* Corresponding author.
E-mail addresses: wangyong_cqupt@163.com (Y. Wang).