COUSIN: A network-based regression model for personalized recommendations

Mingxin Gan *

Article history:
Received 17 November 2014
Received in revised form 9 November 2015
Accepted 1 December 2015
Available online 11 December 2015

Keywords:
Recommender systems
Network-based regression
Accuracy
Diversity

A B S T R A C T

Recently, such state-of-the-art methods as collaborative filtering, content-based, model-based and graph-based approaches have achieved remarkable success in recommendations. However, most of them make recommendations based on either information from users or objects, or bipartite relationships between them, without explicitly exploring object, user and object-user relationships simultaneously. Meanwhile, recent discoveries in sociology and behavior science have demonstrated that similar users tend to select similar objects, usually referred to the n-degree of influence. However, such understandings have not been systematically incorporated into recommendations yet. With these understandings, we propose a novel method named COUSIN (Correlating Object and User Similarity profiles to personalized recommendation), adopting a regression model to incorporate object, user and object-user associations simultaneously in a global way for personalized recommendation. We also construct a power-law adjusted heterogeneous network for COUSIN to prevent adversely influence of popular nodes. We demonstrate the effectiveness of our method through comprehensive cross-validation experiments across two data sets (MovieLens and Netflix). Results show that our method outperforms the state-of-the-art methods in both accuracy and diversity performance, indicating its promising future for recommendation.

1. Introduction

Over the past few years, information overload accompanying the explosion of the world-wide-web has been recognized as a serious problem in not only business areas but also our daily life. As one of the promising solutions, recommender systems have been proposed to help people filtering out irrelevant information efficiently and identifying their personalized preferences effectively [1], resulting in successful applications in a variety of fields such as the online recommendation books [2], movies [3], news [4,5], TV programs [6], microblogs [7], friends [8,9], tourism [10], taxi [11], and many others [12,13].

A recommendation method is typically designed based on the collaborative filtering principle, grounded on the understanding that users who agreed on preferred objects in the past will tend to agree in the future [14–18]. With this assumption, a user-based filtering approach calculates discriminant scores for candidate objects relying on user similarities that are derived from historical data and then ranks candidates accordingly [19]. An object-based filtering design, as a counterpart of the user-based formulation, relies on object similarities that are also derived from historical data [20]. In contrast, information such as descriptions, annotations and properties of objects can also be utilized to characterize similarities between objects, resulting in a class of content-based approaches [21–23]. To promote respective advantages of these two categories, hybrid methods have also been proposed [19]. The recent advancement has also suggested that sophisticated mathematical modeling of latent relationships between users or objects could greatly benefit recommendation performance, resulting in such state-of-the-art model-based methods as the probabilistic latent semantic analysis [21,24], non-negative matrix factorization [25] and singular value decomposition [26]. Another branch of actively studied graph-based approaches tries to construct a network of users and/or objects and then makes recommendations via simulating random walk [8,27], diffusion [28,29] and heat conduction [30] processes in the network.

A hallmark of the classical collaborative filtering or content-based approaches is that they make recommendations based on the information of either users or objects alone in a local way. For example, a typical object-based collaborative filtering method relies on only object similarities and overlooks potential relationships between users. In such an approach, the prediction score of a candidate object for a query user is calculated by considering only objects that have been selected by the user in history, coinciding with a local manner. Therefore, in the scenario that similarities can only be reliably inferred between the candidate object and a small set of other objects, and it happens that these objects are not frequently selected by the query user, such an item-based method can hardly be effective due to its intrinsic local characteristic. In such a situation, simultaneous consideration of both object similarity and user similarity is desired. A toy example for the scenario is illustrated in the supplementary
material (section 1). Furthermore, model-based methods such as the probabilistic latent semantic analysis, and graph-based approaches such as the random walk with restart model, have demonstrated that the use of object or user relationships in a global way will benefit the recommendation performance [31,32]. Nevertheless, none of these existing methods has explicitly explored object relationships and user relationships simultaneously in a global way. Recent discoveries in sociology and behavior science have demonstrated that similar users tend to select similar objects, in studying statistical properties of bipartite graphs of actors and movies [33,34] as well as scientists and papers they co-authored [35]. These findings, usually referred to the n-degree of influence [36–38], suggest that not only direct associations but also indirect relationships between users and objects could contribute to a recommendation process. However, such discoveries have not been systematically incorporated into a recommendation method yet.

With the above considerations, we propose in this paper a novel method called COUSIN (Correlating Object and User Similarity profiles for personalized recommendation), adopting a regression model to simultaneously incorporate both of object relationships and user relationships in a global way into personalized recommendation. Specifically, our method treats the user similarity as the response variable, derive the predictor variable from the object similarity, and adopt a regression through the origin model to explain the user similarity using the object similarity. Furthermore, the computation involved in our method can be further simplified to the calculation of the cosine value between a user similarity vector and an object similarity vector, thereby greatly reduce the computational burden. We demonstrate the effectiveness of our method through comprehensive cross-validation experiments across two data sets. Results show that our method outperforms the state-of-the-art methods in both accuracy and diversity of recommendation.

2. Methods

2.1. Overview of COUSIN

The basic premise of our method is that two users with high similarity in preference often rate similar objects, and thus the concordance between two users on preference coincides with the relatedness of objects they preferred. As illustrated in Fig. 1, our approach includes four main components. First, in the similarity calculation procedure, we calculate object similarities and user similarities based on the historical data, obtaining two matrices representing pairwise similarities. Note that in principle, object similarities can also be calculated using a content-based approach [39] or based on annotations, such as tags [40]. Second, in the network construction procedure, we apply a power-law adjustment strategy [41] to both of the object and user similarity matrices, obtaining sparse similarity networks for objects and users, respectively. Third, in the extraction process of concordance vectors, we construct two concordance vectors for the target user and a candidate object, respectively. Particularly, the vector for the target user is composed of similarities between the user and all other users, which is represented as the “user similarity profile”, and thus is of the same length as the number of users. The vector for the candidate object is derived from object similarities in the following way. For each of the other user, we collect all objects that have been selected by the user in history and sum over similarities between such objects and the candidate object to obtain a score, represented as the “associated object similarity”. Repeating this procedure for all the users, we obtain the “associated object similarity profile” between users and the object, which is represented as a vector whose length is the same as the number of users and whose elements contain scores calculated in the above way. Forth, in the concordance score calculation procedure, we establish a regression through the origin model using the two vectors, known as the user similarity profile of the user and the associated object similarity profile of the object, and adopt the goodness of fit of this model to measure the concordance between the target user and the candidate object. Specifically, this procedure can be simplified to the calculation of a concordance score (the same as the cosine of the angle) between the two vectors (the user similarity profile of the user and the associated object similarity profile of the object), which is defined as the “global matching degree”. Repeating the above four procedures for each pair of a target user and a candidate object, we generate their corresponding “global matching degree”. Finally, we prioritize candidate objects according to their “global matching degrees” (the concordance scores) for each target user and obtain the ranking list to make recommendations. A toy example for COUSIN is illustrated in the supplementary material (section 1).

2.2. Similarity calculation

We adopt two definitions of similarities: the cosine similarity (CS) and the Jaccard index (JC). Let matrix $X = (X_{uv})_{n \times m}$ be preferences of $n$ users on $m$ objects, where $X_{uv} = 1$ if object $o$ is preferred by user $u$ and $X_{uv} = 0$ otherwise. We calculate the cosine similarity between users $u$ and $v (1 \leq u \leq n, 1 \leq v \leq n)$ and denote it as,

$$
S_{(cosine)}(u,v) = \frac{\sum_{x=1}^{m} X_{ux}X_{vx}}{\sqrt{\sum_{x=1}^{m} X_{ux}^2} \sqrt{\sum_{x=1}^{m} X_{vx}^2}},
$$

That is, the cosine of the angle between the two column vectors corresponding to the users. We denote similarities between all users as matrix $S_{(cosine)} = (s_{uv})_{n \times n}$ in a similar way, we calculate cosine similarities between all objects as matrix $T_{(cosine)} = (t_{uv})_{m \times m}$.

Furthermore, treating preferences of user $u$ and $v$ as two sets, $x_u = \{ o : X_{uv} = 1 \}$ and $x_v = \{ o : X_{uv} = 1 \}$, respectively, we calculate the Jaccard index between the two users as

$$
S_{(jaccard)}(u,v) = \frac{|x_u \cap x_v|}{|x_u \cup x_v|},
$$

That is, the number of elements in the intersection of the two sets over that in the union. We denote similarities between all users as matrix $S_{(jaccard)} = (s_{uv})_{n \times n}$ and similarly, we calculate similarity matrix between all objects using the Jaccard index and denote it as

$$
T_{(jaccard)} = (t_{uv})_{m \times m}.
$$

2.3. Network construction

Although the above similarity calculation have been widely used in the existing collaborative filtering approaches, the existence of unreliable relationships may result in small similarity scores and hence adversely influence the downstream inference [42]. Recent studies have shown that the application of a power-law transformation to user similarities can greatly benefit recommendation performance by effectively magnifying the difference between strong similarities resulting from the share of a large fraction of objects and weak similarities due to the share of a small number of popular objects [41]. We therefore adopt such a power-law transformation strategy in our model. Specifically, given the original similarity $s_{uv}$ between two users $u$ and $v$, we raise it by the exponent $\beta$, obtaining the adjusted similarity $s_{uv}^{\beta}$. Repeating this procedure for all pairwise similarities between users, we obtain the power law adjusted user similarity matrix $S = (s_{uv}^{\beta})_{n \times n}$. We further treat users as vertices and user relationships with non-zero similarities in this matrix as edges to obtain a user network. Similarly, we obtain power law adjusted object similarity matrix $T = (t_{uv}^{\beta})_{m \times m}$ and further construct an object network.
دانلود مقاله

http://daneshyari.com/article/554655