

A Collaborative Filtering Recommendation Algorithm Based On Conditional Entropy Trust Model

Yitao Wu^a, Xingming Zhang, Xiaofeng Qi, Erning Xiao, Liang Jin

National Digital Switching System Engineering and Technological R&D Center

Zhengzhou 450002, China

^a wu_yitao@126.com

Keywords: non-linear dependence, conditional entropy, similarity calculation, collaborative filtering.

Abstract. The traditional collaborative recommendation algorithm doesn't take the non-linear dependence between users into consider, which is not accurate enough in prediction. The collaborative filtering algorithm based on entropy can measure the nonlinear characteristics of users, but it can't reasonably describe the relationship between users and is subject to sparsity. To address this problem, conditional entropy trust model is proposed, which uses the conditional entropy to describe the non-linear dependence between users, and Laplace estimation is introduced to alleviate sparsity. A collaborative filtering algorithm based on the conditional entropy trust model (CECF) is designed. The experiments show that this algorithm doesn't increase the time complexity and significantly improve the degree of accuracy.

Introduction

With the rapid development of Internet, users can't find their favorite items in the enormous data resources. Recommender systems can help users select theirs' preference by filtering out useless information [1]. The collaborative filtering (CF) assume that similar users have similar interests. The CF is easy to establish and has a good prediction performance, which is the most successful recommender system [2].

The user-based CF algorithm is predicted by three steps: first, the algorithm calculates the similarities of all users to the target user. Next, the algorithm selects the top k similarity users as neighbors. Finally, the prediction for the target user is calculated by the neighbors' ratings. The key of CF is selecting a reasonable similarity calculation method. The traditional collaborative recommendation algorithm usually uses cosine similarity, Pearson correlation coefficient and so on to compute the similarities.

A superior similarity calculation method should gain the positive factors and decay the negative factors between uses. Therefore the superior method should consider the non-linear dependence between users. However, the traditional similarity calculation method only considers the linear dependence between users, and assumes the positive factors is the same as the negative factors, which neglects the non-linear dependence between users [3].

Entropy $H(X)$ measures the order and regularity of the random variable X . $H(X)$ will be higher if the distribution of X has less order, $H(X)$ measures the non-linear dependence [4]. Piao[5] proposed a new entropy-based recommendation algorithm, which contributes to solving the cold-start problem and discovering users' hidden interests; Zhang[6] used the theory of information entropy to calculate the information entropy of both user level and item level to solve the cold-start problem; Kaleli[7] introduced a novel entropy-based neighbor selection approach which focuses on measuring uncertainty of entity vectors, which significantly improved recommendation accuracy.

The papers above all use entropy to measure the non-linear dependence between users. However, there are two weaknesses of these papers. Entropy can only measure the order and regularity of one user or item, it would not be accurate that use entropy to measure the relationship between uses. Moreover, the calculation of entropy will have a large error due to the sparsity.

To solve above weaknesses, this paper designs a new conditional entropy trust model, which uses the conditional entropy to describe the non-linear dependence between users, and Laplace estimation is introduced to alleviate sparsity. The model calculates the direct conditional entropy and the indirect conditional entropy based on the relationship between users respectively, finally gets the conditional entropy on the combination of direct and indirect conditional entropy. A collaborative filtering algorithm based on the conditional entropy trust model (CECF) is designed. CECF considers the non-linear and the linear dependence between users. The experiments show that this algorithm doesn't increase the time complexity and significantly improves the degree of accuracy.

Conditional entropy model

To solve the weaknesses, this paper proposes a new conditional entropy trust model, which can reasonably describe the non-linear relationship between users. This section introduces the characteristics of conditional entropy, and analyzes the relationship between conditional entropy and trust. Finally, we propose the conditional entropy trust model.

Conditional Entropy. Conditional entropy $H(Y|X)$ measures the order and regularity of the random variable Y when the random variable X is known. $H(Y|X)$ is calculated as follows [8]:

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) \log_2 p(Y|X=x) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 p(y|x) \quad (1)$$

where $p(x,y)$ and $p(y|x)$ are the joint distribution and the conditional probability distribution of random variable (X,Y) respectively. And \mathcal{X} and \mathcal{Y} are the range of x and y . If X is known, $H(Y|X)$ will be 0 when Y is the function of X . Conditional entropy measures the non-linear relationship between variables, and there is an obvious relationship between conditional entropy and prediction accuracy.

We assume that user A's ratings follow the distribution of the random variable X , and user B's ratings follow the distribution of the random variable Y . We can use user B's ratings to predict user A's ratings, and the prediction results is the random variable Z . It's obviously that the distribution of Z is only relevant to Y , is conditional independence with X . Thus the sequence X, Y, Z is a Markov chain. P_e is the estimation error probability, namely $P_e = P\{X \neq Z\}$, and the range of X and Y are same. We can get the formula based on Fano inequality as follows [8]:

$$P_e \geq \frac{H(X|Y) - 1}{\log_2(|\mathcal{X}| - 1)} \quad (2)$$

where $|\mathcal{X}|$ represents the range of X , we can calculate the estimation error probability's lower bounded by $H(X|Y)$. $H(X|Y)$ is directly related to the prediction results, a smaller $H(X|Y)$ indicates a better prediction results. If user AB's conditional entropy is smaller, user B is more familiar with user A, the use B's prediction for use A will be more accurate. This prove that we can enhance the prediction accuracy by introduce the conditional entropy into recommender systems.

Some studies find that the trust of two users will be higher if the users' ratings are closer [9]. We can know that if the users' ratings are closer, the conditional entropy will be smaller. There is an inverse correlation between trust and conditional entropy. Conditional entropy is satisfied with the asymmetry of trust ($H(Y|X) \neq H(X|Y)$), which can reasonably describe the incompleteness transitivity of trust. Moreover, conditional entropy dissatisfies the triangle inequality. The conditional entropy is determined by own characteristics not transfer. Therefore users' conditional entropy describes the users' trust actually.

Conditional Entropy Trust Model. Based on the relevant definitions above, this paper proposes the conditional entropy trust model, which is shown as Fig.1. This model divided user A's and B's m -dimensional feature vector into three parts: x -dimensional direct feature vector of user B related to user A, which mean the number of common rating items by user A and B is x , y -dimensional indirect feature vector of user B related to user A, which mean the number of items rated by user A but unrated by user B is y , and $(m-x-y)$ -dimensional non-interacting feature vector, which mean the number of items unrated by user B is $m-x-y$. The $(m-x-y)$ -dimensional non-interacting feature vector is not impact

$H_{A,B}$. This model calculate the $H_{A,B}$ by combination of $HD_{A,B}$ and $HI_{A,B}$. $H_{A,B}$ can represent use A's trust to user B, so we can calculate $T_{A,B}$ by $H_{A,B}$.

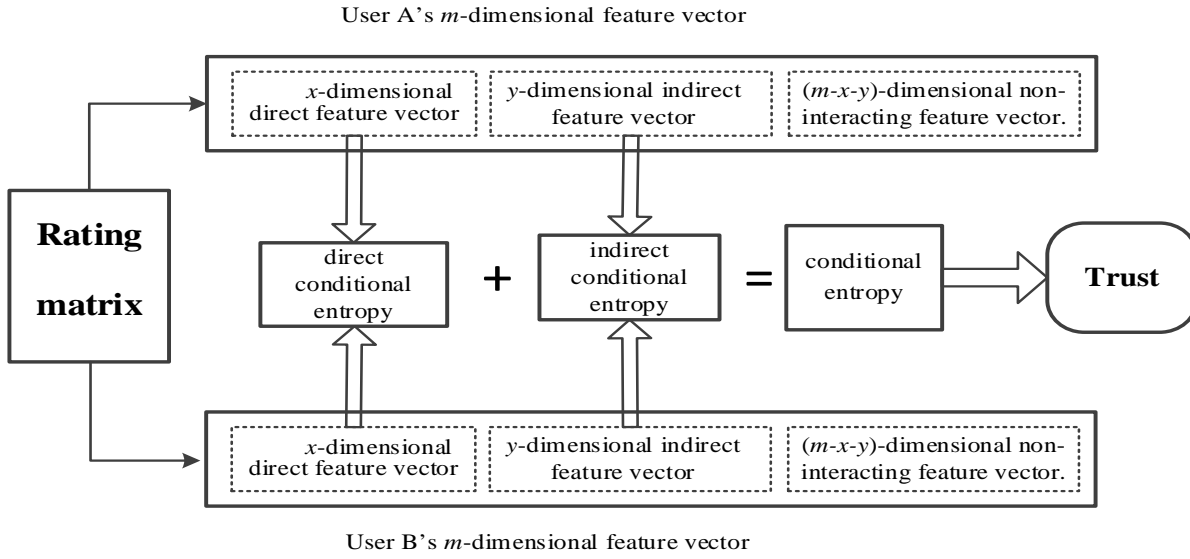


Fig. 1 Conditional entropy trust model.

The collaborative filtering algorithm based on the conditional entropy trust model

This section designs CECF based on the conditional entropy trust model. The algorithm considers the non-linear and the linear dependence between users to make predictions. This section firstly put forward the related calculation of CECF, which mainly includes the calculation of direct and indirect conditional entropy, the calculation of trust and the fitting between trust and similarity, and analyzes the time complexity of CECF.

The Calculation of Conditional Entropy. $H(A|B)$ represents the known user B's familiarity with user A. The smaller $H(A|B)$, the less uncertainty user B associated with user A has, namely user A trusts user B more. We can calculate the direct conditional entropy based on common rating items, which are very few. The direct conditional entropy will have a large error due to the sparsity. B. Marlin [10] pointed out that it doesn't a random processes in user select items. This show that user's unrated items can reflect user's characteristic. According to this idea, we propose indirect conditional entropy, which use indirect feature vector to represent uses' characteristics. We can calculate the conditional entropy by combination of direct conditional entropy and indirect conditional entropy. The direct conditional entropy describes the direct trust relationship between users, and the indirect conditional entropy describes the indirect trust relationship between users. Actually, the conditional entropy describes the trust between users.

We can calculate the direct conditional entropy by the x -dimensional direct feature vector of user B related to user A. However, the conditional entropy $H(A|B=s)$ will have a large error due to the sparsity. We use Laplace estimation to alleviate the influence of the sparsity, $p(A=k|B=s)$ is estimated as:

$$p(A = k | B = S) = \frac{n(r_A = k, r_B = s) + l * p_B(S) * p_A(k)}{n(r_B = s) + l * p_B(S)} \quad (3)$$

where $n(r_B=s)$ represents the number of items that user A rates as s points, $n(r_A=k, r_B=s)$ represents the number of items that user A rates as k and user B rates as s , l is a tuning parameter, which assume that user A and B have been rated l items, $l * p_B(s)$ represents the number of user A rates l items as s points, and $l * p_B(s) * p_A(k)$ represents the number of user A rates one item as k points when user B rates the item as s points. In this way can avoid the zero probability problems, and can alleviate the sparsity problem. When $l=20$, we can get the optimal results. $p_A(k)$ is the priori probability of user A rates as k . We also use Laplace estimation to adjust $p_A(k)$, which is shown as below:

$$p_A(k) = \frac{n(r_A = k) + 1}{n(r_A) + |s|} \quad (4)$$

where $n(r_A)$ represents the number of items that user A rates and $|s|$ denotes the strength of the ratings, $|s|=5$.

The calculation of conditional entropy doesn't consider the rating difference between user A and user B. Thus we introduce the square of the difference between user A's ratings and user B's ratings. $H(A|B=s)$ and $HD_{A,B}$ are calculated as follows:

$$H(A|B=s) = -\sum_{k=1}^5 (k-s)^2 p(A=k|B=s) \log_2 p(A=k|B=s) \quad (5)$$

$$HD_{A,B} = \sum_{s=1}^5 p_B(s) * H(A|B=s) \quad (6)$$

We can calculate the indirect conditional entropy $HI_{A,B}$ by the y-dimensional indirect feature vector of user B related to user A. $HI_{A,B}$ is estimated as:

$$p(A=k|B=\emptyset) = \frac{n^*(r_A=k) + |m| * p_A(k)}{n(r_B \setminus r_A) + |m|} \quad (7)$$

where $n(r_A \setminus r_B)$ represents the number of elements of user A and B's indirect feature vector, $n^*(r_A=k)$ represents the number of elements of user A rates as k points in indirect feature vector, and $|m|$ denotes the strength of the ratings, $|m|=5$. $HI_{A,B}$ is calculated as follows:

$$HI_{A,B} = -\sum_{k=1}^5 p(A=k|B=\emptyset) \log_2 p(A=k|B=\emptyset) \quad (8)$$

Both $HD_{A,B}$ and $HI_{A,B}$ can reflect the user B's familiarity with user A. $HD_{A,B}$ considers the condition user A and user B both rate an item. Therefore $HD_{A,B}$ can reflect the difference between user A and user B better. The weight of $HD_{A,B}$ should be bigger than $HI_{A,B}$, and the weight of $HD_{A,B}$ have a relationship with the number of elements of user A and B's indirect feature vector. Thus $H_{A,B}$ is calculated as follows, where $q \in [0,1]$.

$$H_{A,B} = HD_{A,B} + q * \frac{x}{x+y} HI_{A,B} \quad (9)$$

The Calculation of Trust. There is an inverse correlation between $H_{A,B}$ and $T_{A,B}$. A smaller $H_{A,B}$ indicates that the less uncertainty user B associated with user A has, namely user A trusts user B more. $T_{A,B}$ is calculated as follows:

$$T_{A,B} = \frac{1}{H_{A,B} \times \sum_{u \in U} \frac{1}{H_{A,u}}} \quad (10)$$

where U represents the set of all users. Based on Eq.10, $T_{A,B}$ is between 0 and 1.

The Fitting between Trust and Similarity. The trust mainly considers the non-linear dependence between users based on the relationship between users' rating patterns, while neglects the linear dependence between users. For the better optimization of recommendation results, this paper combines trust with cosine similarity, which is calculated as follows:

$$sim_{A,B} = \cos(A, B) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \times \|\mathbf{B}\|} \quad (11)$$

\mathbf{A} and \mathbf{B} represent the rating vector of user A and B respectively, while $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ represent the length of the rating vector of user A and B. In $T_{A,B}$'s normalization processing, we don't break the multiple relationship between the trust. Moreover, trust and cosine similarity are different dimensions of users' relationships, we can't fit trust and cosine similarity with the plus computation. Thus we calculate the final relationship between user A and user B, $W_{A,B}$, as follows:

$$W_{A,B} = T_{A,B} * sim_{A,B} \quad (12)$$

The Time Complexity Analysis. The time overhead of CECF is the calculation of conditional entropy, trust and similarity.

1. The calculation of conditional entropy: run $n(n+1)$ times;
2. The calculation of trust: run n times;
3. The calculation of similarity: run $n(n+1)$ times;
4. The prediction: run k times;

The total running times of CECF is $f(n) = 2n^2 + n + k + 1$. Thus the time complexity of CECF is $T(n) = T(n^2) + T(n) + T(k) + T(1)$, namely, $T(n) = O(n^2)$. Through above analysis, the time complexity of CECF is $O(n^2)$, which is the same as traditional collaborative recommendation algorithm.

Experiment

Data Set and Evaluation Metric. In this paper, we use three benchmark data sets: MovieLens 100k, and Netflix 3m1k datasets, where each user votes movies in five crisp ratings 1-5. In this paper, we use Mean Absolute Error (MAE) to evaluate the performance of the proposed methods. MAE is defined as follows [11]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - r_i| \quad (14)$$

where p_i is the predicted rating, r_i is the true rating, and N denotes the number of all items. The lower the MAE is, the more accurate the predictions are.

The Compared Algorithms and Find Optimal Parameter. This experiment compares other two algorithms with CECF. The cosine similarity collaborative recommendation algorithm (CF) and the collaborative recommendation algorithm rating information entropy method (RIEM)[6].

To find the optimal parameter q , this experiment fixedly divide MovieLens 100K dataset into two parts: the training set contains 80% of data, and the remaining 20% constitutes the testing set. After setting the number of neighbors to 20, 30 or 40, we carry out the experiment to find the optimal parameter q . the results is show as Fig.2.

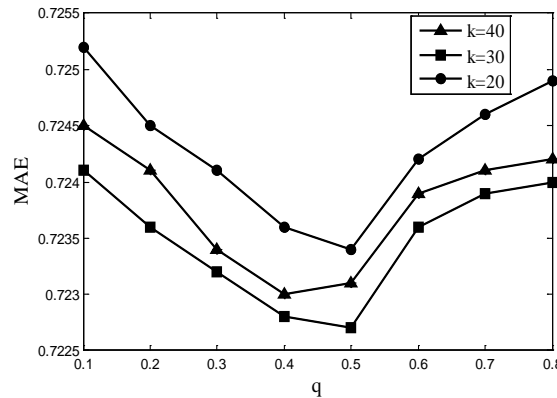


Fig. 2 The relationship between q and MAE.

According to the outcomes in Fig.2, we can find MAEs drop when $q < 0.4$, while MAEs fall when $q > 0.5$. After many attempts, we find the optimal parameter $q = 0.47$. Consequently, we use the optimal q to carry out the following experiments.

Experiment Results. To test the recommendation algorithms, in this section, we conduct two experiments on two data sets and compare the proposed method with other methods. Each data set is randomly divided into two parts: the training set contains 80% of data, and the remaining 20% constitutes the testing set.

Effect of Neighborhood Size. In this experiment, the number of the neighborhood will vary monotonically from 5 to 50. We compare the accuracy of the three algorithms based on two kinds of datasets. The results are shown in Fig.3 and Fig.4:

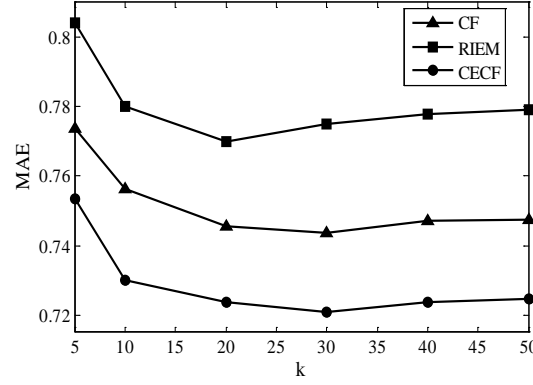


Fig. 3 The effect of neighborhoods in MovieLens 100K.

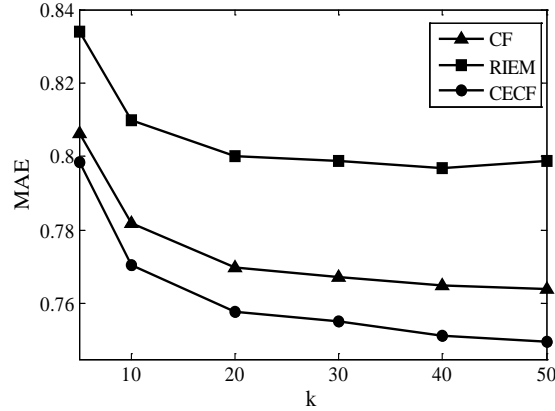


Fig. 4 The effect of neighborhoods in Netflix 3m1k.

In MovieLens 100k dataset, the MAE of all algorithms firstly decrease then increase with the increasing of the number of neighbors. While in Netflix 3m1k dataset, the MAE of all algorithms decrease with the increasing of the number of neighbors, when the number of neighbors is more than 20, the MAE has very small change. We can get that our method (CECF) can obtain better performance with different numbers of neighbors than other methods, and the performance of the CF is better than that of the RIEM.

The results indicate that CECF can obtain the best performance, which is because CECF not only consider the non-linear dependence between users, but also consider the linear dependence between users. In this way can CECF find the appropriate neighbors. RIEM has a worse performance, which is because entropy can't be accurate that use entropy to measure the relationship between uses and the calculation of entropy will have a large error due to the sparsity.

Comparison of Elapsed Time. We can get the time complexity of CECF is $O(n^2)$, which is the same as traditional collaborative recommendation algorithm. Therefore the time cost of CECF is acceptable.

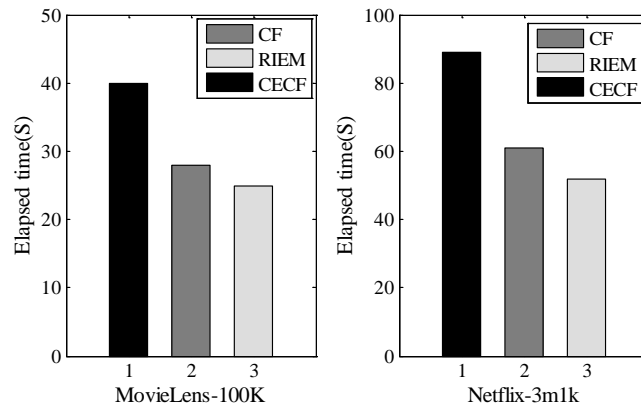


Fig. 5 The Comparison of elapsed time.

Summary

This paper design a new conditional entropy trust model, which uses the conditional entropy to describe the non-linear dependence between users, and Laplace estimation is introduced to alleviate sparsity. This model can reasonably describe the relationship between users and alleviate the sparsity problem. A collaborative filtering algorithm based on the conditional entropy trust model (CECF) is designed, CECF consider the non-linear and the linear dependence between users, can find the appropriate neighbors. The experiments show that this algorithm doesn't increase the time complexity and significantly improve the degree of accuracy. Our future work is aim to explore the relationship between the direct conditional entropy and the indirect conditional entropy, and find a more appropriate formula of conditional entropy, which can decrease the complexity of CECF.

Acknowledgment

This work was jointly supported by the National High Technology Research and Development Program ("863"Program) of China (grant numbers 2014AA01A704), the National Natural Science Foundation for Creative Research Groups of China (grant numbers 61521003), the National Natural Science Foundation of China (grant No. 61572520).

References

- [1] J.M. Yang, and K. F. Li, Recommendation based on rational inferences in collaborative filtering, *Knowledge-Based Systems* 22.1(2009):105–114.
- [2] J. Herlocker, et al, Evaluating collaborative filtering recommender systems, *Acm Transactions on Information Systems* 22.1(2004):5-53.
- [3] J.F. Huete, et al, Using past-prediction accuracy in recommender systems, *Information Sciences* 199.15(2012):78–92.
- [4] J. Ni, et al, Ceiling effect of online user interests for the movies, *Physica A Statistical Mechanics & Its Applications* 402.2(2014):134-140.
- [5] C.H. Piao, J. Zhao, and L. J. Zheng, Research on entropy-based collaborative filtering algorithm and personalized recommendation in e-commerce, *Service Oriented Computing & Applications* 3.2(2009):147-157.
- [6] F. Zhang, H. Liu, Y. Cui, Rating information entropy for cold-start recommendation, *Journal of Information & Computational Science*. 2011, 8(1) 16–22.
- [7] C. Kaleli, An entropy-based neighbor selection approach for collaborative filtering, *Knowledge-Based Systems*, 2014, 56(3):273–280.
- [8] H. S. Witsenhausen, and A. D. Wyner, A conditional entropy bound for a pair of discrete random variables, *IEEE Transactions on Information Theory* 21.5(1975):493-501.
- [9] C N. Ziegler, Golbeck J., Investigating interactions of trust and interest similarity, *Decision Support Systems*, 2006, 43(2):460-475.
- [10] B. Marlin, and R. S. Zemel, Collaborative prediction and ranking with non-random missing data, *Conference on Recommender Systems* 2009:5-12.
- [11] Y.X. Zhu, Evaluation Metrics for Recommender Systems, *Journal of University of Electronic Science & Technology of China* (2012).