Research Journal of Applied Sciences, Engineering and Technology 5(23): 5474-5477, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: November 27,2012 Accepted: January 17, 2013 Published: May 28, 2013

A Collaborative Filtering Recommendation Mechanism for Peer-to-Peer Video Sharing

[1, 2]Chun-Xia Yin, [1]Hui-Ying Zhang and [3]Jian Liu
[1]College of Information Science and Engineering, Yanshan University, China
[2]Ocean College of Hebei Agricultural University, China
[3]Hebei Vocational College of Foreign Languages, China

Abstract: Almost all collaborative filtering recommendation systems based on C/S mode have to face the problems of one-point-failure and unscalable. This study proposes a scalable collaborative filtering recommendation mechanism for video sharing in unstructured Peer-to-Peer (P2P) networks. The mechanism is named as CFRPV, which can recommend videos in distributed way. The CFRPV mechanism includes four parts: peer model definition, neighbor peer set construction, CF-based recommendation for videos and neighbor peer set update. In CFRPV, peer users rank all the videos that they had watched. Then a video can be represented as a point in video vector space and its rank is the value of this point. One peer's preference also can be represented by a vector of the ranked videos in the video vector space. All the peers construct and dynamic reconstruct neighbor peer set in real time through calculating preference similarity between each other. From their neighbor peer sets, peers receive video recommendations that had been filtered. Finally, simulation results are discussed.

Keywords: Collaborative filtering, neighbor peer set, unstructured P2P networks, video recommendation

INTRODUCTION

A video recommendation system can generate personalized video recommendation list for users according to their preference. Collaborative Filtering (CF) is one of the most successful recommendation techniques. It identifies users whose tastes are similar to those of a given user and it recommends items those users have liked in the past (Yehuda, 2010). In general, video recommendation service is supplied by big Website systems, which collect information of users' preferences on ranked video items. However, most C/S mode video recommendation systems have to face some problems, if recommendation system servers were down or not reachable, Website users can't use the service. In addition, the centralized systems are unscalable.

Deploying recommendation systems in P2P networks with a decentralized architecture is an effective way to solve the upper problems (Kim *et al*., 2008; Huiying *et al*., 2009). P2P networks have emerged as a successful way to exchange resources and services across a large number of peers. Now, an increasing number of P2P users and shared files also raise a serious complexity for the users searching and selecting their desired contents (Fuyong *et al*., 2011). P2P users need an individualized, autonomous system that can learn a user's preference and search for relevant information.

In this study, a video recommendation mechanism, CFRPV, is proposed based on collaborative filtering in unstructured P2P network. In CFRPV, the characteristic of video is described using a point with its weight value in the video vector space, which has been watched by user in recent time period. Then a query or peer's preference can be represented by a vector, which consists of many example video items.

To learn about the peer's true intention, the peer's current preference on the watched video needs to be feedback so that CFRPV can learn from this preference to retrieve video more similar to the one the peer really wants. So, each peer constructs a neighbor peer set with similar preference and updates its members in real time. Simulation results showed that the CFRPV could achieve highest hit ratio and low root mean square error compared with other two relative recommendation mechanisms.

LITERATURE REVIEW

Vector space model: In traditional information retrieval, queries are normally expressed as a set of keywords that is quite convenient according to the Vector Space Model (VSM) or term vector model. VSM is an algebraic model for representing text documents and any objects as vectors of identifiers. It is used in information filtering, information retrieving, indexing and relevancy rankings. In VSM, queries and responses are all modeled as term vectors in a multidimensional information space. If we want to calculate the similarity of two persons' preferences, we can calculate the cosine value of the angle of their preference vectors.

Collaborative filtering: Generically, CF is an algorithm that filters information for a user based on a

Corresponding Author: Jian Liu, Hebei Vocational College of Foreign Languages, China

collection of user profiles. One common characteristic of these algorithms is that they require a centralized user-item rating matrix as the input source. These approaches can be divided into memory-based method and model-based method (Alan *et al*., 2012). In memory-based method, the task of CF is to predict the votes of active users from the user database, which consists of a set of votes corresponding to the vote of user i on item j. The memory-based CF method calculates this prediction as a weighted average of other users' votes on that item through the following formula:

$$P_{a,j} = \overline{v}_a + \kappa \sum_{i=1}^{n} \varpi(a, j)(v_{i,j} - \overline{v}_i) \quad (1)$$

$P_{a,j}$ denotes the prediction of the vote for active user a on item j and n is the number of users in user database. $\overline{vi}$ is the mean vote for user i as:

$$\overline{v}_i = \frac{\sum_{j \in I_i} v_{i,j}}{|I_i|} \quad (2)$$

I_i is the set of items on which user i has voted. The weights $\varpi(a, j)$ reflect the similarity between active user and users in the user database. κ is a normalizing factor to make the absolute values of the weights sum to unity.

Recommendation mechanism based on collaborative filtering in P2P networks: In recent years, with the pervasive deployment of computers, P2P is increasingly receiving attention in research and more and more P2P systems have been deployed in the Internet. A few attempts towards decentralized CF in P2P networks have been introduced (Yehuda, 2010; Kim *et al*., 2008; Huiying *et al*., 2009; Zhaobin *et al*., 2010). In some pure text retrieval mechanisms, text and user's preference are expressed by vectors. Each document of text was viewed as a vector of word frequency and the similarity between two documents was computed as the cosine value of the angle between them. The CF technique treats each user's preference as a document and the votes as the frequency for items.

VIDEO RECOMMENDATION MECHANISM

Based on the study of Yehuda (2010), Kim *et al*. (2008), Huiying *et al*. (2009), Soboroff and Nicholas (2000), Bracewell *et al*. (2008) and Du and Fang (2008), this study presents a CFRPV mechanism as follows. Firstly, the model of the peer in the unstructured P2P networks is defined. Secondly, all the peers construct and initialize their neighbor peer set. Thirdly, each peer receives video recommendations that is high ranked by the neighbor peer set and generate top-k recommendation list. At last, the peers can update their neighbor peer sets according to the change of their preference.

Peer model definition: In CFRPV, peer is defined as h_i (i =1, 2, … , n). h_i has its individual peer model, which composed of three parts, peer profile, neighbor peer set and target peer set. A peer profile includes its preference information, which is used to find similar peers as neighbor peer set to receive recommendations. Target peer set is composed of the peers that request to the host peer for forwarding video recommendations. Video items, which have been watched and ranked, include user's preference information. These video items are saved and updated automatically. One preferred video item set composed of saved video items is used to create the peer profile. The preferred video item set is defined as $P = \{p_1, p_2, ..., p_j, ..., p_L\}$, which denotes that peer h_i has L saved video items. p_j is composed of video name vn and its rank value vr. Whenever h_i ranks a video, a new item p_j is added to P.

Each peer calculates neighbor similarity between itself and other peers. Then it can select neighbor peers, which have higher similarity than others. Neighbor peer set of h_i is defined as $N = \{n_1, n_2, ..., n_j, ..., n_S\}$, where S is the maximum number of neighbors. Additionally, candidate neighbor peer set is defined as $CN = \{cn_1, cn_2, ..., cn_j, ..., cn_S\}$. Each peer in N has to be exchanged dynamically with a more similar peer in CN. Target peer set is defined as $T = \{t_1, t_2, ..., t_j, ..., t_T\}$, where T is the maximum number of target peers. T is organized by the request of other peers with similar preference.

Neighbor peer set construction: In the whole P2P network, the number of all the videos is M and one video item is related to one dimension in the multidimensional space. So the video vector space is set to M-dimensions vector space. In this vector space, all the point values of a new profile vector P (h_i) should be initialized by 0. Then, we get all rank values from P of hi and set point values of the P (h_i). For example, after the rank value vr of p_j being get from P of h_i, the point value of P (h_i) in the sth dimension, which is related to p_j, is set to vr.

In this part, multi random walk search method is used to discover neighbor peers. The peer h_i sends out several query-vectors P (h_i) to an equal number of randomly chosen connected peers. Each query-vector is forwarded to a randomly chosen peer at each step by intermediate peers. Then, peer h_j that received a query-vector calculates neighbor similarity, NS (h_i, h_j), according to cosine distance function of the VSM. If the value of NS (h_i, h_j) is higher than a threshold, ns, that can be set by user, h_j would be a member of the N of h_i. Given P(h_i) and P(h_j), which are profile vectors of h_i and h_j respectively, NS(h_i, h_j) is defined as follows:

$$NS(h_i, h_j) = \frac{P(h_i) \bullet P(h_j)}{\|P(h_i)\| \|P(h_j)\|} \quad (3)$$

For initializing N of h_i with an empty P, multi random walk method is also employed to search neighbors who save video items more frequently.

CF-based recommendation for videos: In this part, the method of computing text similarity is used to compare each recommended video and peer's preference. After receiving high ranked recommendations from N, top-k video list should be generated for hi. Using the text information of each video name, the distance between each recommended video item and P is calculated and k videos with the shortest distance are added to the top-k video list. User skims through the list to see if there are any video of interest to watch and rank.

Vx is a word frequency vector that can be generated from a recommend video name. It is represented a point of the vector space. VP is a set of vector space points, which can be constructed using all the vn of a P. The distance function between a Vx and a VP aggregates multiple distance components from Vx to the space points set. An aggregate function Distance (Vx, VP), which is derived from (Lee *et al*., 2000), is used to calculate the distance:

$$Distance(Vx, VP) = \sqrt{\frac{L}{\sum_{i=1}^{L} \frac{1}{d^2(Vx, vp_i)}}} \quad (4)$$

L = The number of space points in a VP

vp_i = The ith point of VP

d (x, p_i) = A distance function between a video vector Vx and a query point vp_i. A video vector with the shortest distance component to any one of query points is treated as it with the shortest aggregate distance. The d (Vx, vp_i) in (5) is defined as follow:

$$d(Vx, p_i) = \sqrt{\sum_{m=1}^{M} \frac{(vx_m - p_{im})^2}{\sigma_m}} \quad (5)$$

M is the number of dimensions of feature space, vx_m and pim are coordinates of x and p_i on the mth dimension respectively. $1/\sigma_m$ is the weight of the mth dimension in the feature space and σ_m is standard deviation of coordinates of the mth dimension.

Neighbor peer set update: Each peer has to update its N according to the values of neighbor similarity. In recommendation process, the higher the value of one peer's neighbor similarity in an N, the more possible its recommended video would be watched. The peer with the highest neighbor similarity makes the most contribution to recommendation process. Hence, each peer's CN is composed of S peers, which must be the most similar to the host peer in the N of each host peer's neighbors respectively. So, the host peer explores the CN. If a more similar peer cn is discovered in CN than the peer n that has the lowest similarity to the peer in N, the cn is included to the N and the n is discarded.

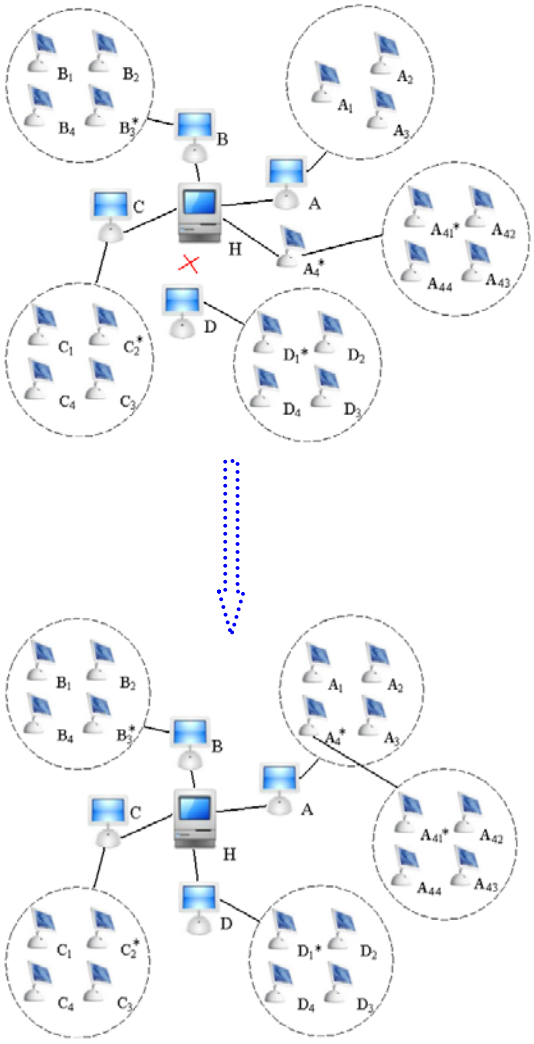

Fig. 1: Update of neighbor peer set

For example, as shown in Fig. 1, the *N* of peer h_i are *A*, *B*, *C* and *D*. *A*, *B*, *C* and *D* have their own *N*. A_1, A_2, A_3 and A_4* make up of the *N* of *A*. A_4*, B_3*, C_2* and D_1* are the most similar peers to h_i in *N* of *A*, *B*, *C* and *D* respectively and A_{41}* is the most similar peer to h_i in *N* of A_4*. *CN* of h_i includes A_4*, B_3*, C_2* and D_1*. If *A_4 has higher neighbor similarity than *D*, A_4* becomes new neighbor of h_i. *D* and *$D_1$are excluded from *N* and *CN* of h_i respectively and A_{41}* becomes new candidate neighbor peer of h_i.

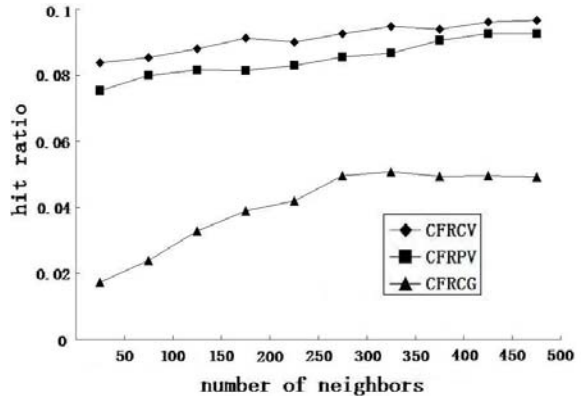

Fig. 2: hit ratio comparison

Table 1: Root mean square error comparison

Mechanism	CFRPV	CFRCV	CFRCG
20% data set	1.194±0.005	1.153±0.005	1.412±0.005
80% data set	1.093±0.005	1.087±0.005	1.302±0.005

SIMULATION

In simulation, hit ratio and root mean square error are performed. Hit ratio is defined as the ratio of the success number of recommendations to the number of connections. The root mean square error or root mean square deviation is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. The performance of CFRPV is compared with that of two relative centralized mechanisms named as CFRCV and CFRCG. CFRCV is similar to CFRPV, but neighbor construction and update process use all peers' *P* information. CFRCG is a classical centralized recommendation mechanism based on GVSM (Soboroff and Nicholas, 2000.). Simulation was performed for Movielens data set (Huy and Dinh, 2012), which consists of 1 million ratings from 6000 users on 4000 movies. The Values of parameters *k* and *ns* are set to 20 and 0.5, respectively.

As shown in Fig. 2 and Table 1, CFRCV achieves the highest hit ratio and lowest root mean square error. CFRCG get the lowest hit ratio and the biggest root mean square error. In both hit ratio aspect and root mean square error aspect, the performance of CFRPV is very close to CFRCV that generates recommendation lists using all users' profile information. CFRPV and CFRCV, which use the dynamic method to construct and update neighbor peer set, can represent the users' preference more accurate than CFRCG. So CFRPV can achieve high forecast accuracy with P2P architecture, which is naturally scalable.

CONCLUSION

In this study, a video recommendation mechanism in unstructured P2P network is presented. As deploying the CF in pure P2P network, one-point-failure and unscalable problems can be solved. Comparisons in hit ratio and root mean square error showed that the CFRPV mechanism achieves a high performance with great scalability. The video recommendation can be more efficient by dramatically updating neighbor peer set.

ACKNOWLEDGMENT

This study was supported by the Hebei provincial department of science and technology, CHINA, under the Grants: Provincial Science and Technology Research and Development Program No.11213597.

REFERENCES

Alan, S., B.J. Jain and S. Albayrak, 2012. Analyzing weighting schemes in collaborative filtering: Cold start, post cold start and power users. Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC '12), pp: 2035-2040.

Bracewell, D.B., J.N. Yan and F. Ren, 2008. Single document keyword extraction for internet news articles. Int. J. Innov. Comput. I., 4(4): 905-913.

Du, A. and B. Fang, 2008. Novel approach for web filtering based on user interest focusing degree. Int. J. Innov. Comput. I., 4(6): 1325-1334.

Fuyong, Y., J. Liu and C. Yin, 2011. A collaborative filtering recommendation mechanism based on user profile in unstructured P2P networks. J. Shandong Univ., Nat. Sci., 46(5): 28-33.

Huiying, Z., C. Yin and J. Liu, 2009. PNRC: Personalized news recommendation mechanism based on collaborative filtering in unstructured P2P networks. ICIC Express Lett., 3(3B): 561-566.

Huy, N. and T. Dinh, 2012. A modified regularized non-negative matrix factorization for movie lens. Proceedings of the IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation and Vision for the Future (RIVF), pp: 1-5.

Kim, J.K., H.K. Kim and Y.H. Cho, 2008. A user-oriented contents recommendation system in peer-to-peer architecture. Expert Syst. Appl., 34(1): 300-312.

Lee, J., C. Wu, F. Christos, S. Katia and R.P. Terry, 2000. FALCON: Feedback adaptive loop for content-based retrieval. Proceedings of the 26th International Conference on Very Large Data Bases, pp: 297-306.

Soboroff, I. and C. Nicholas, 2000. Collaborative filtering and the generalized vector space model. Proceeding of the 23rd ACM SIGIR. Athens, Greece, pp: 351-353.

Yehuda, K., 2010. Collaborative filtering with temporal dynamics. Commun. ACM, 53(4): 89-97.

Zhaobin, L., W. Qu, H. Li and C. Xie, 2010. A hybrid collaborative filtering recommendation mechanism for P2P networks. Future Gener. Comput. Syst., 26(8): 1409-1417.