ONTOLOGY BASED RECOMMENDER SYSTEM WITH USING DISSIMILAR USERS

Musa MILLI

Dept. of Computer Engineering, Ege University, Bornova, Izmir, 35100, Turkey
Dept. of Computer Engineering, Bartın University, Bartin 74100, Turkey
musa.milli@ege.edu.tr

Mehmet MILLI
Dept. of Computer Engineering, Dokuz Eylül University, Buca, Izmir, 35390, Turkey
mehmetmilli@cs.deu.edu.tr

Abstract: Rapid development of the e-commerce, increase in the product range and customers in the past years makes it difficult to find the products that customers are looking for, in terms of customer made confusion and led to time losses. These developments have necessitated the development of new customer-oriented marketing strategy. Recommender system directs interests of customers to the product which they can like and buy. So the companies that use these systems increase profits by providing strategic advantages while helping customers. Collaborative Filtering (CF) is a successful technique used in recommendation system. They try to find out costumer's interest in a new product, based on similarities of customers' previous ratings. In the literature, these calculations are performed with the principle which is: similar users have similar tastes. In this paper these calculations are performed with thoughts that dissimilar users have dissimilar tastes, too. Currently algorithm of collaborative filtering suffers from some problems such as cold start and sparse data. At the same time this paper aims to propose ontology based solution for cold start problem.

Keywords: Ontology, Recommender System, Dissimilar User

Introduction

Rapid development of the e-commerce, increase in the product range and customers in the past years makes it difficult to find the products that customers are looking for, in terms of customer made confusion and led to time losses. These developments have necessitated the development of new customer-oriented marketing strategy. At this point recommender systems have become modern marketing tools for new generation e-stores by providing personalized service to users while using past user preferences, product database and some other parameters. The main aim of the recommender systems find products that the user is interested in and, offer meaningful advice among millions of products (Melville, 2010, p.829). Today recommender system has been increasingly used in companies in particular e-commerce applications and these companies provide strategic advantage. Recommender systems are important part of some e-commerce sites such as Amazon.com and Netflix.com. They are real world examples of these systems.

In the past a wide variety of recommender systems are proposed such as Content Base Filtering (Chen, 2011, p.1371; Barranco, 2010, p.409; Pazzani, 2007, p.325), Collaborative Filtering (CF) (Herlocker, 1999, p.230; Herlocker, 2002, p.287; Gong, 2010, p.745; Al Mamunur Rashid, 2006), Knowledge Based Systems (Burke, 2000, p.180), Hybrid Systems (Burke, 2002, p.331; Salter, 2006, p.35), etc. Although there are wide variety of recommender system discussed any of them satisfy some problem like cold start and sparse data without any additional mechanism. In this paper we propose a naive ontology based collaborative method to deal with cold start problem.

The rest of the paper organized as follows: Section 2 provides background on ontologies and CF algorithms and defines some problems which CF suffers from them. In section 3 we define our ontology CF algorithms using dissimilar users. Section 4 provides explanation of dataset and evaluation metric. Section 5 represents the result and the last section concludes the paper.

Background and Motivation

Recommender systems are classified according to their prediction approach (Adomavicius, 2005, p.734). The recommender systems can be divided into four main categories; Content based filtering, collaborative filtering, knowledge based systems and hybrid systems. Collaborative filtering has achieved most success in real world application area and our study is focused on collaborative filtering.

Collaborative Filtering (CF)
The word collaborative filtering was mentioned the first commercial recommenders system called Tapestry (Goldberg, 1992, p.61). CF works by collecting user opinions about items as ratings, they use this ratings to calculate similarities between users. Then CF finds users interest about items that they have never seen and taste before, using this ratings. If calculated interest is positive then CF recommend items to users.

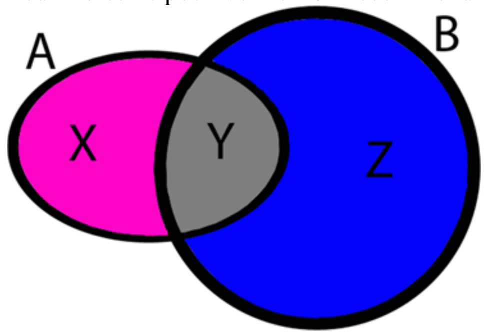

Figure 1: The representation of principles of collaborative filtering.

Where A is the set of items rated by Alice and B is the set of items rated by Bob. Region Y is the set of items rated by both users. Region X is the set of items rated by Alice but Bob has never seen or tried before. Region Z is the set of items rated by Bob but Alice has never seen or tried before. The CF says that the similarity calculation made through the region Y. At the result of calculation there is positive correlation between Alice and Bob we define they are similar users, so most probably they have similar tastes. According to this inference Alice most probably likes the items which are in region Z.

Generally we examine CF algorithms in three following steps: first step similarity calculation, second step neighborhood selection and the last step prediction computation.

1. Similarity calculation

In the first step of CF algorithms similarities between active user and the other users were calculated. In CF algorithms there are several similarity methods have been used such as cosine vector similarity, adjusted cosine vector similarity and Pearson correlation coefficient, etc (Gong, 2010, p.745). According to our experiments Pearson correlation coefficient gives better results; because Pearson correlation coefficient takes into account the user average.

$$sim(a,u) = \frac{\sum_{i \in I} (R_{a,i} - \overline{R}_a).(R_{u,i} - \overline{R}_u)}{\sqrt{\sum_{i \in I} (R_{a,i} - \overline{R}_a)^2} . \sqrt{\sum_{i \in I} (R_{u,i} - \overline{R}_u)^2}} \tag{1}$$

where $R_{a,i}$ denotes the ratings of user a on item i, $\overline{R}_a$ is the average rating of user a. I is the set of items rated by both user a an u.

2. Neighborhood selection

Second step of CF algorithms is the step to find nearest user to active user. In this step the results which obtained from the similarity step have been used to find active user's nearest neighbor. User based CF algorithms works this principle that similar users have similar tastes. According to this definition each user is not included in the calculation process. So CF selects similar users and discards dissimilar users. This process is called neighborhood selection in literature. There are two different methods for neighborhood selection. One of them is threshold method and the latter is k-nearest neighbor method.

3. Prediction calculation

The users selected in step two acts to the extent of their similarities to the prediction. In this stage the user similarities between the active user utilize as weight vector. There are variety of prediction method while the most accurate results are obtained Adjusted Weighted Average (AWA); because AWA takes into account how users perceive the rating scales. In this study we also observed that this method gave the best results.

$$R_{a,i} = \overline{R}_a + \frac{\sum_{u \in U}^{n} (R_{u,i} - \overline{R}_u).sim(a,u)}{\sum_{u \in U}^{n} sim(a,u)} \quad (2)$$

where $\overline{R}_a$ is the average rating of user a, $\overline{R}_u$ is the average rating of user u. U is the set of active user's nearest neighbors that rated the item i.

Semantic Web and Ontologies

According to Tim Berners-Lee who invertor of the world wide web, semantic web is defined extension of current web (Berners-Lee, 2001, p.28). The semantic web provides a common language that allows data to be shared between and reuse some applications. Most information on the web environment is designed for human understandable but semantic web assures information that can be understood by humans and computers. The main structure of semantic web is ontologies. Ontologies most common use definition is an explicit specification of a conceptualization (Gruber, 1995, p.907). Ontologies have key duty in technology by integrating interoperability and data, information and process (Grobelnik, 2009, p.59). Recent years ontologies have been used with recommender systems in academia.

Semantic Recommendation

In semantic web recommendation approach the recommendation process is generally based on concept diagram or an ontology describing acknowledge based and uses semantic web Technologies (Peis, 2008). Semantic recommendation systems used cold start and data sparsity problems of collaborative filtering system (Wang, 2007, p.4069). In this study semantic recommendation approach is used work out cold start problem in recommendation systems. Cold start problem in recommendation system can be divided into three categories; new system, new user, new item.

New system

When establishing new recommender system there is no data about user preferences, so it is difficult to give the good advice. The users rates items over time and the input data of the system increases. Thus allows recommendation system to give better advice. We think that semantic web cope with this problem until the system collects enough data.

New user

When a new user registration there is no history of this user, so the system couldn't predict what the new user interested in. To deal with this problem some recommender systems want to the user to rate a set of item when registering. We think that semantic web cope with this problem until the new user rate some items.

New item

Like new user problem when an item is added the system, there is no past information of this item, so the system can't recommend this item to the user. This problem refers to new item problem in literature. We think that semantic web approach work out new item problem until the item is rated by some users.

We can solve these cold start problems by using knowledge based structure of semantic web technologies. Because ontologies are knowledge based technologies, so they don't take into account of users and items past information.

Proposed System

Pure CF algorithms works this principle that similar users have similar tastes. From this principle the CF find user similarities by using past rates of user. Then CF use this similarities to predict interest level of user to the item that the user has never seen before. So pure CF uses similarity between users, but we use dissimilarity between users and we obtain good results like (Bulut, 2014). We thought that if the similar users have similar tastes and interests, the dissimilar users have dissimilar tastes and interests. If the prediction can calculated by using positive similarity correlation, its reverse is also possible. So the prediction can calculated by using negative similarity correlation.

In our approach we calculated similarities between active user and the other steps just like the traditional CF algorithm. In the neighborhood selection step, we choose dissimilar users (not nearest neighbors) unlike traditional CF. We use AWA method to find prediction in the last step.

In the ontology side of the system (figure 2), it was identified two main classes which is called person and movie. Person class was divided into subclasses (director, screenwriter, actor, etc.). The film class was categorized by genre (anime, documentary, action, sci-fi, etc.). In the next step, it was identified the necessary relationships between classes. Data properties of these classes were identified such as production year, actor and budget. Last step in this creating ontology, we generate individuals (instance).

Dataset and Evaluation Metric

This section provides dataset introduction and evaluation metric.

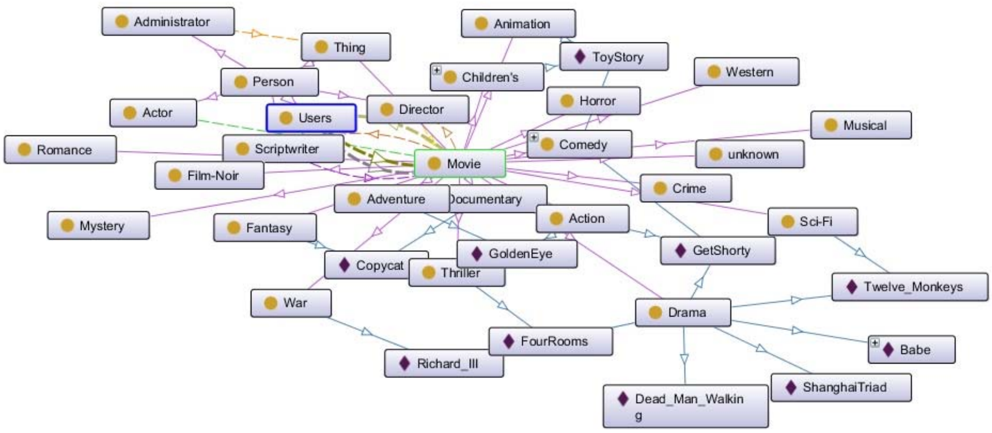

Figure 2: Visualization of the system generated by OntoGraf which is plug-in of the Protégé 4.3.

Dataset

In order to compare the results of pure CF algorithm we use the movielens dataset that is collected by Grouplens research Project from Minnesota University. The dataset was collected through movielens web site from September 1997 to April 1998. This data set consists of 100,000 ratings (1-5) from 943 users on 1682 movies and Each user has rated at least 20 movies.

We divided the database into a train-dataset and a test-dataset. %20 percent of the all ratings were randomly selected and used test-dataset and the others were used train-dataset.

Evaluation Metric

Statistics accuracy metrics measure the prediction accuracy that found by recommender systems. They show us how to prediction success is. It is the numerical distance of the actual value.

$$MAE = \frac{1}{N} . \sum_{i=1}^{N} \left| Ru,i - \hat{R}u,i \right| \qquad (3)$$

where N is the number of test data, Ru,i is the actual value of rating, and $\hat{R}u,i$ is the prediction of CF algorithm.

Experimental Results and Discussions

This section is provide our experiment result. In steps of CF algorithm we keep the parameter constant and we execute the traditional and proposed CF algorithm. We used movielens dataset which was mentioned in previous section.

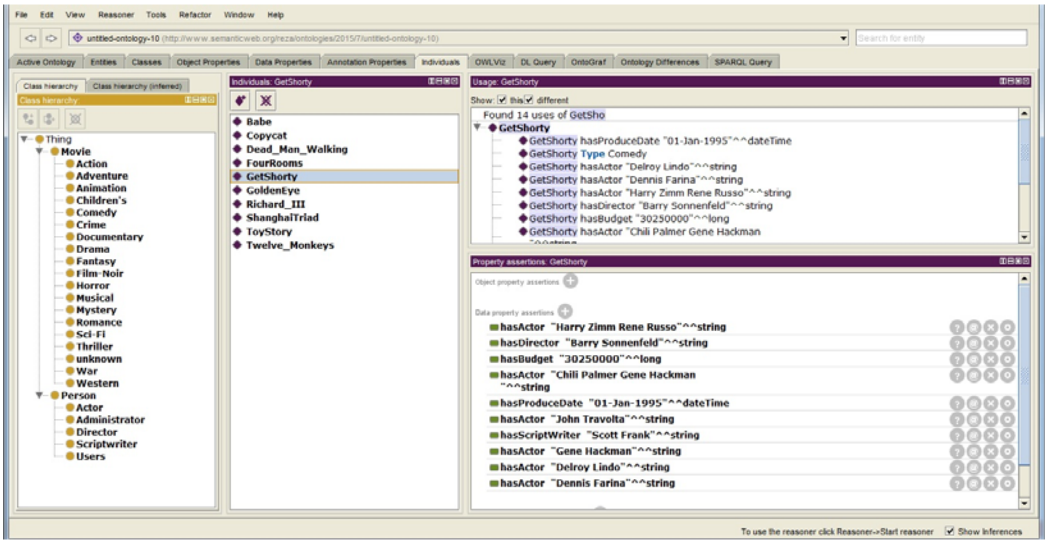

Figure 3: creating individuals by using protégé ontology tool.

Table1: Comparison of proposed CF and traditional CF.

Algorithm	MAE
Traditional CF	0,817
Dissimilar Approach	0,885

These parameters used for this experiment.

- Similarity method : Pearson Correlation Coefficient
- Neighborhood selection method : Threshold (0)
- Prediction method : AWA
- Evaluation metric : MAE

Table 1 shows the performance of the traditional and the proposed CF. As we can see, the proposed system compete with the traditional one. If we can study a bit more on proposed system maybe it will be able to give better results.

We create individuals of classes (figure 3). The class properties and data properties are used for user that newly registered and used for item that newly added. We used this ontology for recommend item to new users until the new users rated enough item.

Conclusion and Future Work

This paper represents a new CF algorithm which uses dissimilarity and our experiments show that the proposed algorithm can compete with traditional algorithm. Maybe in our future work if we are setting parameters again in accordance with the proposed algorithm or we perform minor changes in step one or step three, the proposed CF will be able to give better results. Or we can combine the proposed system and the traditional one it will be able to give the better results.

In our study we propose ontology based recommender system to solve cold start problem in this area. We have achieved good tips and future work we can improved our algorithm and will be able to come out this problem.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749.

Al Mamunur Rashid, S. K. L., Karypis, G., & Riedl, J. (2006). ClustKNN: a highly scalable hybrid model-& memory-based CF algorithm. Proceeding of WebKDD.

Barranco, M. J., & Martínez, L. (2010). A method for weighting multi-valued features in content-based filtering. In Trends in Applied Intelligent Systems (pp. 409-418). Springer Berlin Heidelberg.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 28-37.

Bulut, H and Milli, M. (2014) New prediction methods for collaborative filtering, Pamukkale University Journal of Engineering Sciences.

Burke, R., (2000), Knowledge-based Recommender Systems, Encyclopaedia of Library and Information Systems, 69(32):180-200.
Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4), 331-370.
Chen, Z. S., Jang, J. S. R., & Lee, C. H. (2011). A kernel framework for content-based artist recommendation system in music. Multimedia, IEEE Transactions on, 13(6), 1371-1380.
Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61-70.
Gong, S. (2010). A collaborative filtering recommendation algorithm based on user clustering and item clustering. Journal of Software, 5(7), 745-752.
Grobelnik, M., Mladenić, D., & Fortuna, B. (2009). Semantic technology for capturing communication inside an organization. Internet Computing, IEEE, 13(4), 59-67.
Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing?. International journal of human-computer studies, 43(5), 907-928.
Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999, August). An algorithmic framework for performing collaborative filtering. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 230-237). ACM.
Herlocker, J., Konstan, J. A., & Riedl, J. (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Information retrieval, 5(4), 287-310.
Melville, P., & Sindhwani, V. (2010). Recommender systems. In Encyclopedia of machine learning (pp. 829-838). Springer US.
Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In The adaptive web (pp. 325-341). Springer Berlin Heidelberg.
Peis, E., del Castillo, J. M., & Delgado-López, J. A. (2008). Semantic recommender systems. analysis of the state of the topic. Hipertext. net, 6, 1-5.
Salter, J., & Antonopoulos, N. (2006). CinemaScreen recommender agent: combining collaborative and content-based filtering. Intelligent Systems, IEEE, 21(1), 35-41.
Wang, R. Q., & Kong, F. S. (2007). Semantic-enhanced personalized recommender system. In Machine Learning and Cybernetics, 2007 International Conference on (Vol. 7, pp. 4069-4074). IEEE.

