International Journal of Computer Science and Software Engineering (IJCSSE), Volume 5, Issue 2, February 2016
ISSN (Online): 2409-4285 www.IJCSSE.org Page: 14-18

An Introduction on Separating Gray-Sheep Users in Personalized Recommender Systems Using Clustering Solution

Shamsi Ghorbani[1] and Ahmad Habibizadeh Novin[2]

[1] Department of Computer Engineering, East Azerbaijan Science and Research Branch, Islamic Azad University, Tabriz, Iran

[2] Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

[1]ghorbani_shamsi@yahoo.com, [2]manhabibi@yahoo.com

ABSTRACT

Nowadays, regarding the development of the Internet and the uncontrolled increase in the amount of information, users have faced difficulties for retrieving information for webs. To solve retrieving information, Personalized Recommender Systems appeared. These systems provide information appropriate to users' needs regarding information obtained from them and their behaviors in searching texts. But, one of the commonest difficulties in Personalized Recommender Systems is the existence of users called grey-sheep users. These users have little similarity with other users; therefore, their presence in these systems along with normal users causes the reduction in the precision of predictions and suggestions for the two groups. An appropriate method for reducing the effect of the presence of grey-sheep users is the use of clustering methods. Therefore, in the present study, clustering methods for separating users and reducing the mean absolute error as well as increasing the precision of clustering methods. To separate grey-sheep users, some suggestions for future research were presented.

Keywords: *Personalized Recommender Systems, Grey-sheep Users, Clustering Methods.*

1. INTRODUCTION

Personalized Recommender Systems are defied as computer-based smart methods for facilitating transactions with information overload. Personalized Recommender Systems contribute to customers via guiding them towards their desired alternatives and realizing their needs. In addition, by increasing the references of customers to these systems and their satisfaction, organizations providing services will attain their aims (an introduction to Recommender Systems, 2012).

In recent years, the use of Recommender Systems has become prevalent. With the expansion of information, the use these system has been an important issue in the world of today. Using the knowledge obtained from information can be used in different functions and regarding the large number of customers and goods, Recommender Systems are to use methods which can provide better recommendations for users.

Todays, we are living in the age called the Communication Age. In spite of technologies and instruments at the hand of contemporary humanity, the production and transmission of information has been conducted with an increasing speed. With the development of the Internet, data and information available on this network has been increasingly developed. This development has been so great that currently, one of the problems of using the Internet is the large volume of information.

A diverse range of methods has been presented for solving the problem of information overload each of which has its own problems and limitations. For example, for retrieving information from the web, are a combination of using search engines and manual search on the web is one of the most common methods, but they do not have the capacity of providing information desired by users without the inquiry process (Trestian et al., 2010). In addition, the content of most of websites such as news ones, pages of information about products (such as films, music, different goods, etc.) and personal weblogs are continuously updated. Mostly, regular investigation of these websites are exhausting and time-consuming for seeing updated information.

Recommender Systems use different information resources for creating predictions and suggestions of items for users. These systems try to establish a balance in criteria such as precision, innovation, and sustainability in recommendations. Collaborative filtering methods has a significant role in these recommendations and these methods mostly are used with other filtering methods such as content-based and social types (Bobadilla & Ortega, 2013).

Recommender Systems can be classified based on how recommendations can be divided (Adomavicius & Schafer et al, 2007):

1. Collaborative filtering
2. Statistical filtering,
3. Content-based filtering,
4. Hybrid filtering

In the statistical filtering, it is hypothesized that individuals with common personality characteristics (gender, age, nationality, etc.) have common interests (Porcel et al.; Krulwich, 1997; Pazzani, 1999). Collaborative filtering allows users to rank a set of elements in such a way that when sufficient information are stored in the system, it can provide recommendations for each user based on information obtained from users having the highest familiarity with that user (Herlocker; Antonopoulus & Salter, 2006). The content-based filtering tries to provide items for active users which are similar to items which have obtained positive ranks previously, but this filtering method depends on the issue that items with the same characteristics are similarly ranked (for example, in Recommender Systems of web-based e-commerce, if the user have purchased a number of sci-fi films, these systems recommend the newest sci-fi films which the user has not been purchased (Meteren & Someren.2000; Lang, 1995). The content-based filtering includes the following stages:

1. Retrieving characterizes of items for recommendation
2. Comparing characteristics of items for active users' preferences,
3. Recommending items with characteristics which are compatible with users' interests (Bobadilla et al., 2013).

Hybrid filtering (Porcel et al., 2012) usually is a combination of collaborative filtering with statistical filtering or collaborative filtering with content-based filtering (Barraganschoi et al., Martinez et al., 2010) for using strengths of each of these methods. Recommender Systems are divided into two memory-based and model-based sets: memory-based methods are defied as methods which act only with the User-Item Matrix and each produced rank is used before the reference process (i.e. results are always updated). Memory-based methods usually use similarity degree for obtaining the distance between two users or items (Kong & Sun, 2005; Antonopoulus & Salter, 2006). Model-based methods use information of Recommender Systems for creating a model producing recommendations (Antonopoulus & Salter, 2006; Su & Khoshgoftaar, 2009).

Recommender Systems are faced with limitations for providing precise recommendations compatible with users' needs. Some of the problems are as follows: desertedness of databases, cold start, and grey-sheep.

The problem of desertedness is presented when databases are used in large dimensions. In this case, the User-Item Matrix gets very large and deserted creating challenges for the performance of Recommender Systems.

The problem of cold start occurs when the probability of creating trustful recommendations is not available due to the lack of initial ranks (Antonopoulus & Salter, 2006). This problem can be divided into three types: new society, new items, and new users. The existence of new user is the most important type in Recommender Systems.

But one of the problems which has received less attention is the existence of users called grey-sheep. These users have the least level of similarity with other users. Therefore, when we want use collaborative filtering for recommendations, precise recommendations cannot be created. The existence of this type of users have two negative effects on Recommender Systems: 1) these users cannot receive precise recommendations themselves; 2) they have negative recommendations on other users' recommendations (Ghazanfar & Prugel-Bennet, 2011). An appropriate solution for reducing the effect of the presence of grey-sheep users on the performance of Recommender Systems is separating these users from other users. One of the separating methods us to use clustering method. In clustering method, each user, regarding the degree of similarity with other users is placed into one cluster in such a way that those users placed in a cluster have the most familiarity with users of in the cluster and have the least familiarity with users in other clusters. The rest of the present study are as follows: in section 2, clustering methods are introduced. In section 3, these methods are compared and contrasted.

2. A REVIEW ON PREVIOUS LITERATURE

The diversity of Recommender Systems is very high and are applied in a lot of domains such as recommendations of film, webpages, books, books, and news. But the problem is that Recommender Systems are faced with them and pay less attention to the existence of grey-sheep users. These users have negative effects on individuals' recommendations. The main concentration of the present study is on clustering users and separating these grey-sheep users from normal users so that mean absolute error and increase the precision of recommendations.

2.1 Hierarchical Clustering Method

This method has been presented for recommending news. Users' preferences is not limited only to literature

review on users, but it acts according to group literature of users with similar preferences and with this hypothesis that each group of users have unique preferences to news topics. In addition, users' personal pages is hierarchically indicated with news which is a mixture of several news topics which via combining news hierarchy related to users' groups in which users use adaptive hierarchical clustering, users' interests can be easily determined (Zheng et al., 2013). In this part, users' personal profiles become enriched regarding similar users' profiles. Firstly, users are divided into some groups in terms of similarity and then, users' personal page in each group is created using a weighting method in such a way that users having more interest to a topic share more pieces of information in group profiles (Zheng et al., 2013).

2.2 Personalized Method for News in He Wen Using Hybrid Method

The architecture of hybrid Recommender Systems developed by Wen (wen & Fang, 2012) for recommending news in the web, by diagnosing interests and models, their preferences are separated from each other. The data required by users are analyzed in order that it can present user model. An automatic clustering method for clustering contents searched by users is used. In this method, words are diagnosed by a method called Word Net (Miller, 2009) and the weight of words is calculated using the Term Frequency Inverse Document Frequency. Three parts of a webpage, i.e. information cloud, effective net area, and web address, are separately extracted and classified. The total weight of words are used for categorizing information cloud and effective net area.

It is hypothesized that if a user continuously observe a particular content and is interested in that particular type, in the suggested model, the user modeling method includes two stages: diagnosing the content of web-pages using the categorizing page method and Naïve Bayes model for updating interests and preferences model of users. In this section, the user preferences model rates each website based on the rank from which a user desires to retrieve information (Wen & Fang, 2012).

2.3 Centroid-based Clustering Method Based

Centroid-based clustering algorithm developed by Shinde in are used Personalized Recommender Systems. This method does the process of recommendation in the two phases; in the first phase, users' comments are collected in the form of the User-Item Ranking Matrix, and then, they are clustered offline and using the suggested algorithm in the predetermined number of clusters, and after that, they are stored in databases for future recommendations. In the second phase, recommendations are online for active users using similarly criteria and are produced by selecting clusters with the best rating quality. This issue causes that the efficiency and quality of Recommender Systems be improved for active users. In old Recommender Systems, similarity was usually used explorations in the process of recommendation while in this method, similarity is mixed with the density of clusters. This issue helps the exploration of other clusters which have closer similarities with active users and provide better recommendations for them. The advantage of this method is its better performance compared to ARS, k-means, and new k-medodis algorithms. In addition, this method reduces problems such as cold start, the first rater, and desertedness (Shinde & Kulkarni, 2012).

2.4 Clustering Method Using the Classification Algorithm

One of the most important problems in Recommender Systems is the cold-start problem. This problem is related to recommendations for new users and items. With the advent of new users, Recommender Systems do not have sufficient information about their preferences for creating recommendations. In Lika method which is known as classification algorithm, is combined with similarity methods and recommendation methods and provides instruments required for producing recommendations. In addition, classification methods are mixed with the collaborative filtering system in order to identify users with similar behaviors. This system uses a three-stage method for creating recommendations for new users. This method is adopts a mechanism based on which similarity methods find neighboring users. Those users with the highest level of similarity with new users are called neighbors. In fact, individuals which similar background or characteristics, most likely have similar preferences. Therefore, each new user in a group is classified and is the rating prediction method and responsible for producing rates for items. The advantage of this system is that produces lower mean absolute error values and increases the precision of rating prediction (Lika et al., 2014).

2.5 Evolutionary Clustering Algorithms Method

Rana evaluates the efficiency of the evolutionary clustering algorithms for producing evolutionary clustering and qualified clustering as well as recommendations based on a set of data in real-life situations. Clusters related are used for illustrating users' preferences and creating an adaptive Recommender System. The algorithm uses rates called variance rates and calculate rate differences of users using calculation of their variations in the evolutionary process in each

time. the results indicate that in this method, recommendations with higher quality and lower calculation costs are produced in comparison with other traditional clustering methods of production, but the disadvantages f this method is low speed of learning it (Rana & Jain, 2013).

2.6 Clustering Method Based on Collaborative Filtering

In the Lopez-Nores method, feature-based collaborative filtering has been presented as a new strategy for Recommender Systems and this issue will be based on calculating similarities of available items with users who have compatibilities with them in particular features. While other strategies are relied on the direct relationship between users and items, feature-based collaborative filtering are based on users' features and their separation form items and their features. In this method, there is the possibility of constructing a matrix of values indicting that how features of an item are effective on its appropriateness for an individual with particular user (negative or positive) feature. This issue contributes to solving available problems in filtering methods such as desertedness, hiddenness, and grey-sheep users (Lopez-Nores et al, 2012).

2.7 Neighborhood-Based Clustering Method

Guo has presented a new method for combing trustful neighborhoods with traditional collaborative filtering method in order to solve cold start and desertedness problems from which old Recommender Systems suffer. In this method, rates of trustful neighborhoods are combined with each other in order to complete preferences of active users and illustrate them. The quality of combined rates are calculated regarding the trustfulness of applied rates and incompatibilities between positive and negative beliefs (i.e. rates). Trustful rates are used for calculating user similarity. Prediction for items depends on calculation of mean rates of similar users produced regarding their importance (Guo & Zhang, 2014).

3. CONCLUSION AND FUTURE RESEARCH

In the present study, some of the works conducted on the field of Recommender Systems were studied. In addition, a number of clustering methods for clustering users in Recommender Systems were explained and separating grey-sheep users in the Recommender System was explicated. In addition, all methods tried to improve the quality of clustering users in Recommender Systems and also the reduction of mean absolute error.
Table 1 compares investigated methods and some of their features.
As suggestions for future research on clustering users in Recommender Systems, the following cases are illustrated:

1. The size of clusters has been considered fixed, in the future research, the size of clusters can be considered as one of the variables and can improve the precision of clustering.
2. Revolutionary algorithms such as genetic algorithm-based clustering and Imperialist Competitive Algorithm (ICA) were used for clustering users. These types of clustering causes that the possibility of the presence of normal users' errors in the cluster of grey-sheep users reduces.

Table 1: a comparison of investigated methods and some of their features

Disadvantages	Advantages	Comparison methods
Both accuracy and the efficiency can be improved, it has temporal costs.	User preferences are not confined only to the literature of the study, it has learning power	Hierarchical clustering method
	It has no temporal limitation	clustering method of web-based news
It only used the collaborative filtering method in recommendations	The problem of cold start, this method reduces raters and desertedness	Centroid-based clustering method based
	It produces lower values of mean absolute error, it increases the precision of prediction	Clustering method using the classification algorithm
Lowness of the speed of learning	Lower calculation costs, it has suggestions with higher quality than traditional clustering methods	Evolutionary clustering algorithms method
	It eliminates the problem of desertedness, hiddenness, and grey-sheep users	Clustering method based on collaborative filtering
Incompatibility between negative and positive beliefs	It solves the problem of cold start and desertedness	Neighborhood-based clustering method

REFERENCES

[1] Antonopoulus N., Salter J. (2006). Cinema screen recommender agent: combining collaborative and content-based filtering. IEEE Intelligent Systems. 35–41.

[2] Barragáns-Martinez A.B., Costa-Montenegro E., Burguillo J.C., Rey-López M., Mikic-Fonte F.A., Peleteiro A. (2010). A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. Information Sciences. 180 (22), 4290–4311.

[3] Bobadilla J., Ortega F., Hernando A., Gutiérrez A. (2013). Recommender systems survey. Knowledge-Based Systems. 46, 109–132.

[4] Ghazanfar M.A., Prügel-Bennett A. (2014). Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems. Expert Systems with Applications. 41, 3261–3275.

[5] Ghazanfar M.A., Prügel-Bennett A. (2011). Fulfilling the Needs of Gray-Sheep Users in Recommender Systems, a Clustering Solution. . In 2011 International conference on information systems and computational intelligence. http:// eprints.ecs.soton.ac.uk/21770/.

[6] Herlocker J.L., Konstan J.A., Borchers A.L., Riedl J.T. (1999). 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.

[7] Kong F., Sun X., Ye S. (2005). A comparison of several algorithms for collaborative filtering in startup stage. IEEE Transactions on Networks, Sensing and Control. 25–28.

[8] Lika B., Kolomvatsos K., Hadjiefth-ymiades S. (2014). Facing the cold start problem in recommender systems. Expert Systems with Applications. 41, 2065–2073.

[9] López-Nores M., Blanco-Fernández Y., Pazos-Arias J.J., Gil-Solla A. (2012). Property-based collaborative filtering for health-aware recommender systems. Expert Systems with Applications. 39, 7451–7457.

[10] Miller, G.A. (2009). WordNet – About us. WordNet, Princeton University, <http:// wordnet.princeton.edu>

[11] Porcel C., Tejeda-Lorente A., Martinez M.A., Herrera-Viedma E. (2012). A hybrid Recommender system for the selective dissemination of research resources in a technology transfer office. Information Sciences. 184 (1), 1–19.

[12] Rana C., Jain S.K. (2013). An evolutionary clustering algorithm based on temporal features for dynamic recommender systems. Swarm and Evolutionary Computation. 14, 21-30.