The Application of Social Tagging Based Collaborative Filtering Personal Recommender Strategy in Electricity Market

J.H. Yang
Central China Grid Company Limited
China

H.B. Wang
Purchasing Corporation of HeBei Iron & Steel Group
China

C.H. Gao, Y. Dai
Nari Group Corporation/State Grid Electric Power Research Institute
China

Z.L. Lv
Computing Center
Northeastern University
China

Abstract—In the Internet world, when people access to the information, they are also providing information to others. Therefore, how to find valuable information from the vast amounts of information in order to meet the user's needs, and how to find and enjoy the valuable information by the required users, have been a hot issue which is concerned by academia and the business. Collaborative filtering (CF) and social tagging are the most widely recommendation techniques. In this paper, tag-based collaborative filtering algorithm is proposed to the electricity market. The individual requirement can be satisfied according to different power consumers. This new algorithm can mine the potential preferences of users, and then recommend items in the user's preferences scope. This method can improve the traditional collaborative filtering methods, and can solve the single interest model problem of traditional methods. The experiments based on electricity consumer data set shows that the tag-based collaborative filtering method is significantly better than the traditional collaborative filtering methods in recommendation effects.

Keywords-collaborative filtering; recommender system; social tagging; electricity market

I. INTRODUCTION

With the development of Web2.0 technology, social tag has been widely used in many personalized sites. Using tags to describe resources is more accurate, and it can reflect the real preferences of the individual users. Therefore, social tag is more suitable for the recommendation technique[1]. Collaborative filtering is novel because it involves the relationship between two or more documents, namely a message and its reply, or a document and its annotations[2]. Unlike current filtering systems, Collaborative filters coupled with social tagging, cannot be computed by simply examining a document when it arrives, but rather require repeatedly issuing queries over the entire database of previously received documents [3]. This is because sometime after a document arrives, a human may read that document and decide it is interesting [4].

Power demand forecasting is an important problem in power marketing, power system programming and decision-making of power utilities. Aiming at the practical demand of power marketing programming in the real electricity market, In this paper, based on the current forecasting techniques and the requirement of different power consumers, the collaborative filters coupled with social tagging algorithm based on the theory of times series and linear regression is used. As power of big data accumulation of huge amounts of data geometric growth index, Some new methods are needed to analyze the big data and the development suited to power large data, a data mining software for electricity market users increase the power of big data value-added services.

The current research focuses on the forecasting model and theory method, and is immature for practical application. Considering the practical demand, this paper classifies the power demand into industry customers, agriculture customers, traffic customers, city customers and dynamical customers, analyzes various factors influencing the power demand using

II. THE COLLABORATIVE FILTERING RECOMMENDER ALGORITHM WITH SOCIAL TAGGING

In the social label system , the user can according to their interests and preferences of cognitive resources the project label tag, a project can be several users several label tag, a label can also be used to mark different users use different items, so the variables are defined as follows:

U is denoted the set of users, I denotes the set of projects and T denotes the set of labels. (u,i,t) denotes three-dimensional matrice, and $u \in U$, $i \in I$, $t \in T$.

(u,i,t) means that user u is labeled the project i with the set of t.

For $u \in U$, I_u represents user u annotated collection of items, T_u represents user u set of used tags.

For $t \in T$, I_t represents a collection of tags t annotated all items, U_t represents a set of users used the tag t.

For $i \in I$, T_i represents all projects that mined (u, i, t) with the set of tags i; The number of label t signed projects i, is

countTI(t, i); The number of users u with label t is countUT(u, t).

The correlation of project with the label can be calculated by the TF-IDF(Term Frequency-Inverse Document Frequency), based on the countTI(t, i). TF-IDF algorithm is a weighted technology commonly used for information retrieval and text mining [5, 6]. It is used to Measure a keyword to represent the degree of the project, and to represent the degree of a project. In this paper, TF-IDF algorithm is changed to build the tags - project correlation matrix.

The steps based on the user - label correlation method are as follows:

Input: user - label correlation matrix, tag - project correlation matrix.

Output: The set of projects recommended for the current user u_c.

Step 1: for $t_a \in Tu_c$, the tag-user correlation can be calculated in the current label set Tu_c used by the current users, the equation is as follows:

$$relveance(u_c, t_a) = \frac{countUT(u_c, t_a)}{\sum_{p \in Tu_c} countUT(u_c, p)} \tag{1}$$

$u_c \in U$, relevance(uc) is the tag-user correlation set.

Step 2: the best tag correlation set Tq is chosen in the set relevance(uc).

Step 3: the relevant project corresponding to $t_q \epsilon T_q$ is selected in the tag - project correlation matrix.

Step 4: repeat step3, project collection I_q is recommended to the users uc.

III. THE EXPERIMENTAL RESULT AND DISCUSSION

In this experiment, the data is based on the actual quantity sold by a power supply enterprise last 5 years which are considered the influence of GDP. The different users in this experiment are chosen as commerce, big industry and so on.

TABLE I .THE STATISTICAL RESULT OF ELECTRICITY CONSUMPTION EACH YEAR.

Project	Big industry	Uncommon industry	Commerce	The second industry	The third industry
2009	32954	5898	2980	41.5682	54.1236
2010	28246	5312	3955	44.2754	63.1458
2011	30401	5698	4015	55.3214	79.2561
2012	33619	6985	4998	59.2846	86.6485
2013	38982	7646	5000	70.15698	93.4876

TABLE II .EXPERIMENTAL ENVIRONMENT.

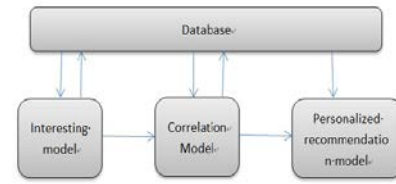|	
The language of the experimental platform	JDK 1.7, Eclipse 4.3
Database	My SQL 5.2
Operating system	Window 8
Hardware platform	Intel CORE i5 GHz CPU, 4GRAm

FIGURE I. PERSONALIZED RECOMMENDER SYSTEM STRUCTURE.

Figure.1 shows the personalized recommender system structure which used in this experiment. Then the experiment results are compared using collaborative filtering algorithm based on user, collaborative filtering algorithm based on the project corresponding to two types of appraisal indicators: MAE and RMSE.

Mean absolute error(MAE) is the most widely used recommendation system of evaluation methods[7]. The accuracy of recommendation system prediction is calculated by the difference between predicted values and the actual value. The equation is as follows:

$$MAE = \frac{\sum_{\{i,j\}} \left| p_{i,j} - r_{i,j} \right|}{n} \tag{2}$$

n denotes the total number of ratings, $p_{i,j}$ denotes user i forecast the score of project j, $r_{i,j}$ denotes user i really score the project j. The MAE value is smaller, the higher the recommendation 's accuracy is.

Root mean square error (RMSE) reflect the discrete degree of the rating data, In the Netflix prize [8, 9], the evaluation index is RMSE, the equation is as follows

$$RMSE = \sqrt{\frac{1}{n} \sum_{\{i,j\}} \left(p_{i,j} - r_{i,j} \right)^2} \tag{3}$$

RMSE value is smaller, score data of discrete degree is smaller, the greater the stability of the algorithm is.
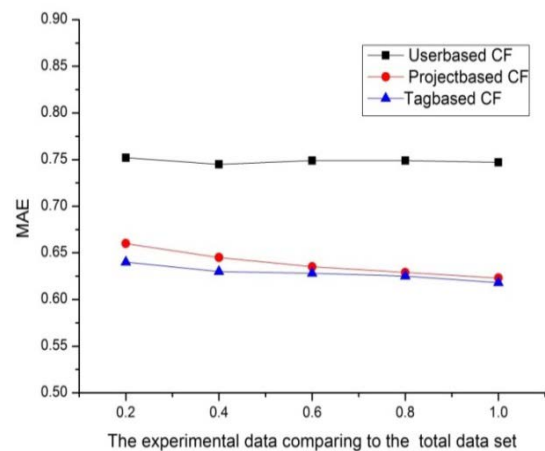

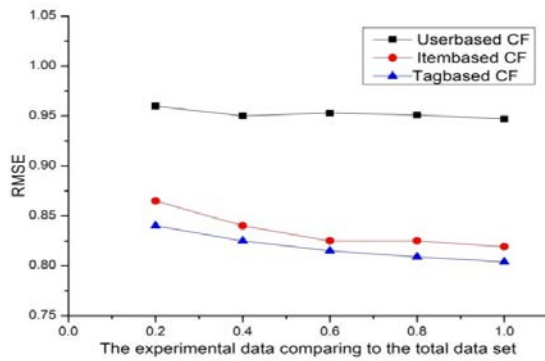FIGURE II. CONTRAST OF MAE BETWEEN THREE TYPES OF ALGORITHM IN THE SITUATION OF NON-INTEGER PREDICTION.

FIGURE III. CONTRAST OF RMSE BETWEEN THREE TYPES OF ALGORITHM IN THE SITUATION OF NON-INTEGER PREDICTION.

In Figure. 2, during the MAE indexes of evaluation, the algorithm is higher because of the lower value of MAE, collaborative Filtering based on the social tag is superior too other collaborative algorithm. However due to the collaborative filtering algorithm based on social tagging for the calculation of the nearest neighbor classification method are adopted to decrease the data needed to calculate the size, the operation speed is superior to collaborative filtering algorithm based on the project. Because the forecast scoring criteria calculated by the algorithm is adopted the floating-point number, which is different to the data set, So there exists score measure of deviation. After artificially adjust prediction score, prediction score of discrete degree becomes higher, algorithm stability is affected. In Figure. 3, with the expansion of the experimental data set, the algorithm performance is better. It shows that the collaborative filtering algorithm based on social tagging on the stability is better than two traditional collaborative filtering algorithm.

IV. CONCLUSION

This experiment using three algorithm similarity calculation was carried out on the training data, respectively, to forecast score value of the test data. Collaborative filtering algorithm based on the user, and the collaborative filtering algorithm based on project do similarity calculation to forecast score by single interest model on the basis of training data. Collaborative filtering algorithm based on social tag generate the predicted rating and similarity calculation according to the user interest model of the user-projects-tag triples data to establish, and each model on the training set. According to the real value in the predicted values and the test data, MAE and RMSE index are assessed. It is shown that the collaborative filtering recommendation algorithm based on social tagging is superior to the traditional collaborative filtering algorithm based on social tagging from MAE and RMSE two evaluation indexes and the operation efficiency.

REFERENCES

[1] Resnick P, Iakovou N, Sushak M, et al. GroupLens: An open architecture for collaborative filtering of netnews. Proc 1994 Computer Supported Cooperative Work Conf., Chapel Hill, 1994 : 175 -186

[2] Grouplens Research, University of Minnesota. Tagsplanations: Explaining recommendations using tags. Proceedings of the 13th International Conference on Intelligent User lnterfaces, IUT09

[3] Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry. Communications of the ACM. 1992, 35(12):61-70.

[4] Lessons on recommendation systems. http://weigend.com/teaching/stanford/

[5] Salton G, Automatic Text Processing. Addison-Wesley , 1989

[6] Vig, Sen, Riedl. Tagsplanations: Explaining Recommendations Using Tags. In Proceedings of the International Conference on Intelligent User Interfaces. 2009. 47-56

[7] Goldberg K, Roeder T, Gupta D, et al. Eigentaste: A Constant Time Collaborative Filtering Algorithm [J]. Information Retrieval, 2001, 4(2): 133-151.

[8] Netflix prize, http://www.netflixprize.com/

[9] R. Sinha and K. Swearingen, Comparing recommendations made by online systems and friends, in Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, 2001