www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

RECOMMENDER ALGORITHMS BASED ON BOOSTING ENSEMBLE LEARNING

Cheng Lili

School of electrical and information engineering, Taizhou Vocational & Technical College, Taizhou, Zhejiang, 318000, China

Submitted: Oct. 22, 2014 ***Accepted: Jan. 15, 2015*** ***Published: Mar. 1, 2015***

Abstract—This article introduces ensemble learning algorithms in recommender systems, and in boosting algorithm framework of this article, shows how to filter the basic recommendation algorithm according to the characteristics of boosting algorithm. By comparing the rational choice of the two recommended boosting algorithm is applied to the frame. And then it determines the main parameters of the algorithm through the experiments, ultimately to obtain a more effective integration of the recommendation algorithm. Experimental results on Netflix validate the effectiveness of the proposed algorithm.

Index Terms—Recommended System, Ensemble learning, Collaborative filtering, boosting.

I. INTRODUCTION

Online shoppers can be roughly divided into two categories. One kind is rigid demand customers, or that have a clear purpose for the customer. This type of customers is likely to be badly in need of something. Before buying they have idea to buy goods of the type and brand, and then search for product information online. Once found, they buy immediately. But the second type of customers has no definite purchase purpose. In reality, they have a habit to browse online shops. Every day they regular or irregular look on e-commerce sites. They are not for shopping, but to find the product of interest, so that only the satisfied goods appear, they will purchase. Thus, for the business, it is difficult to increase the turnover by raising purchases of first class customer, just as it is hard to change people's consumption of these requirements on food and clothes, because of the factors affecting customer's rigid demand is very complicated, such as social and economic development, people's standard of living and consumption habits, so should primarily be recommended for the second type of crowd, fundamental purpose is how to improve the customer retention rate and trading. Because it is hard to influence the first kind of people's consumption, now for the second category of people shopping habits were analyzed. When the site is filled with vast amounts of product information, the second class is very difficult to find the commodity which they like and satisfaction, the recommendation system was born. It can help customers quickly find the commodity which might be interested in, this is not a simple question, because it requires recommendation system must understand customers' preferences. Recommend [1]as if with a magic mind-reading ability, can discern customers' minds, in order to capture the hearts and minds, just like with a friend who understand your be fond of, can carry on the close guide for your online shopping, patience for you recommend that you may be interested in goods, from cars to shoes. Literature data show that [2], more than a third of the users will follow the recommended shopping e-commerce site, which is not possible with any mass advertising, and now the impact on consumers of media advertising has been getting lower and lower, it was noted that "personality recommendation technology will become the ultimate form of advertising".

In the 1970 s, the prototype for collaborative filtering algorithm has been formed, in the 1990 s, the recommendation algorithm theoretical framework is already very mature, but too few examples of real application recommendation algorithm, then even fewer commercial operation. Because e-commerce development in recent years has greatly promoted the commercialization of recommendation algorithms, personalized recommendation systems to e-commerce has brought great commercial interests, and even exceeded the previous market bottlenecks, such as Venture Beat data show [3] Amazon's sales come from 35% of the recommended system.

A. The recommender algorithm

On the recommendation of the current mainstream technology in the macro has three mechanisms: collaborative filtering, and hybrid recommendation algorithm based on content, their purpose is to explore the interests of users to provide users with satisfactory products, but they are the theory starting point and adopts the method of each are not identical.

1. Collaborative Filtering

Collaborative filtering algorithm is the most widely used a recommendation algorithm [4]. Its main steps can be divided into two steps: the first is the users to buy records of historical information is used to calculate the similarity between the user; Then, using the high similarity neighbors with the target users for the evaluation of other products to predict target user preferences for a particular product, algorithm according to the degree of be fond of target users to recommend. The biggest advantages of collaborative filtering recommendation system to recommend objects without special requirement, can deal with music, movies and other difficult text structured representation of the object. The algorithm of collaborative filtering recommendation system can be divided into two categories: based on memory, and based on the model. Memory-based algorithms to predict based on the system all play too much product, Set $U=\{u_1,u_2,\dots u_N\}$ collection for the user, $I=\{i_1,i_2,\dots i_M\}$ for the product collection, $r_{a,b}$, u_a for users to product i_b score, the score is don't know, need through the algorithm to predict. In the collaborative filtering, the user u_a is the rate of the product i_b obtained by other users to rate the calculation of i_b. Generally, the user's purchase history converted into matrix form: rows represent users, the columns represent the items. The user was said by the said items vector. He (she) bought items with 1, said didn't buy any of the items with 0, said that the user is as a weighted vector which is composed of elements 1 and 0. Purchasing records directly into the matrix referred to as the original matrix, the result of a run on the matrix recommendation algorithm is called the target matrix or prediction matrix.

Adopted the way of representation, collaborative filtering can be formulated into two steps: First of all, according to the original matrix Y, according to the similarity of the formula to calculate similarity between users, get the user similarity matrix S; And calculate S and Y as the product of prediction matrix Y, Y = S * Y. Variable Y, when the possibility of a value of 0 corresponds to the customer when the Y value is an algorithm for prediction score, the greater the value of the article to the system, the greater is recommended. The main factors that determine the performance of collaborative filtering is the similarity matrix. The final result recommended depends on the similarity formula. Zhou Tao, according to the diffusion theory of matter proposed Diffuse algorithm, which is a web-based bipartite graph recommendation algorithm, although its principles and collaborative filtering are quite different, but very similar to their approach, the equivalent of a relatively new the similarity calculation method.

Diffuse material (diffusion) method [5]

$$sim(a,b)=\frac{1}{k_b}\sum_{s\in S}\frac{r_{a,s}r_{b,s}}{k_s} \tag{1}$$

K indicates the degree of users or items.

2. Based on the matrix factorization algorithm

Matrix factorization algorithm is very popular now, even in some of the sparse matrix which has good performance. This is a great innovation to some methods of matrix theory applied in the recommendation algorithm [6]. Often it can be achieved better effect than recommended by similar users, but the amount of time the price is bigger. Based on the matrix factorization algorithm was produced at Netflix contest, because the goal of collaborative filtering is to make the target matrix Y as far as possible close to the test matrix, the researchers from the perspective of global optimization, depending on the matrix theory method, find the MAE (target matrix by subtracting 2 norm) test and the smallest goal matrix Y. MAE is a kind of commonly used evaluation index. Recommendation algorithm researchers often use optimization theory, by optimizing an evaluation indicator to obtain the target matrix. At Netflix contest based on the matrix factorization algorithm is very popular and obvious effect, the method based on matrix factorization of the performance of 6% than Netflix your recommendation system effect is good, but this is not the result of a single matrix factorization algorithm, but incorporates a variety of results matrix factorization algorithm.

Low-dimensional factor model [7]is one of the most popular models for collaborative filtering method. Its principle is as follows: assume that only a few factors will affect the user's preferences, user interest in items depends on users and items of these a few factors match case, this is a matrix factorization problem, such as in a k-dimensional factor model, given the score matrix, $Y \in R^{mn}$, we need to find two matrix, $U \in R^{mk}, V \in R^{kn}$, makes the $Y \approx UV$ matrix factorization based collaborative filtering is a factorization model of Y, user i is represented by $u_i \in R^k$ and items j is represented by $v_j \in R^k$. Such as in A system, each element of v_j is representative items j characteristics, such as whether it is an action film, music in it is melodious, and the each element in the u_i represents the characteristics of the user, such as whether he (she) like the film characteristics. The result is a linear combination of the final score of user i and j items of these factors. A linear combination of the user i and items j is the final score results

$$y_{ij} = \sum_{l=1}^{k} u_{il} v_{jl} = u_i^T v_j \tag{2}$$

In order to find the optimal U, V, we can solve the following optimization problem:

$$\min_{U \in R^{mk}, V \in R^{kn}} \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) + \sum_{ij \in S}\left(y_{ij} - u_i^T v_j\right)^2 \tag{3}$$

Of which, $\lambda > 0$ are regular parameter, $S = \{ij \mid y_{ij} > 0\}$.

3. Content-based recommendation algorithm

Content-based recommendation algorithm [8] will build configuration file information of users and products. It is the basic requirement of recommended products must be able to attribute analysis, because only can attribute analysis to set up items of configuration information files, such as the label for the items; and user configuration information files can be created according to the configuration file contents users to buy goods over. A piece of music, artificial hard to judge it is sad or joy, even be able to distinguish, the precision of analysis is difficult to guarantee. If the task to the computer, for example, rely on computer semantic sentiment analysis is one can imagine the difficulty, it is difficult to establish the configuration information of the music file, and configuration information file of course buy the user is also very difficult to rely on this piece of music to provide information to build. So content-based recommendation methods are established to analysis based on the items of information. Collaborative filtering recommendation is based on similar user purchase behavior can predict the future behavior of the target user heuristic rules, so you must first make similarity between users of computing, which requires the user to have enough to buy a record, or recommend accurate performance will be seriously reduce. Recommendation algorithm based on content is different, because the configuration information file once to establish user and item similarity, can be directly calculated the user and the product, if higher user and an item similarity, can put the goods is recommended to the user, of course the user purchases the more the better, so that the user profile information may be more detailed and accurate, but not as high as collaborative filtering requirements, perhaps the user only buy a small amount of the typical characteristics of the commodity, the typical characteristic can clearly show the user preferences, users don't have to buy a lot of other typical characteristics of the goods, the recommendation accuracy can reach higher.

4. Hybrid recommendation algorithm and Paper integrates recommendation algorithm

Recommendation system in practice mostly ensemble several recommendation algorithm, so it is by a single recommendation algorithm effect is good: many of the recommended system will be two or more of the recommendation algorithm to combine the results, by using the linear combination forecast rating recommendation, or recommend a better algorithm in an evaluation index results; the typical is adding content based algorithm in collaborative filtering recommender systems, many are based on collaborative filtering algorithm based on content, namely the use of user profile is the traditional collaborative filtering calculation, the user similarity which is calculated through the configuration file content, is no longer a fight too product information also, it can overcome the problem in Collaborative filtering systems, another advantage is not only to be the user can products can be recommended, if the product and user profile very similar will be directly recommended.

B. Boosting ensemble learning

Boosting [9]is firstly used for classification problems. First of all samples with a sampling weights (usually start weights equal that uniform distribution), the sample training a classifier to classify the samples, the error so you can get the classification rate, according to the error rate of a classifier assigns the weight, gross error is more weight is more and more small, the error of the samples we increase the sampling weights of it, will be the next classifier such training focuses on these wrong samples, then according to the error rate and weight calculation, so the iteration, the strong classifier we get is a weak classifier weighting and. We can see that the performance of classifier right major, which embodies the essence of boosting. Boosting a two classification problems in the following [10]:

1) Given data set:
$(x_1, t_1), ..., (x_m, t_m)\ where\ x_i \in X, t_i \in T = \{1, -1\}$

2) Initialized data weights: $Q_{1(i)} = \frac{1}{m}$

3) For t=1,…,T:
4) Data based on the weight training weak classifiers.
5) Get weak classifier: $\boldsymbol{C}_1$: X→R
6) A weight w_t assigned to the weak classifiers $\boldsymbol{C}_t$, Generally w_t use the formula:

$$w_t = \frac{1}{2}\ln(\frac{1-\varepsilon_t}{\varepsilon_t})$$

Of which, ε_t is a classification error rate for $\boldsymbol{C}_t$;

7) Update sample weights according to the classification of the results of $\boldsymbol{C}_t$.
8) $Q_{t+1}(i) = \frac{Q_t(i)\exp(-w_t t_i c_t(x_i))}{Z_t}$, where $\boldsymbol{Z}_t$ is a regular factor.
9) $Z_t = \sum_{i=1}^{m} Q_t(i)\exp(-w_t t_i c_t(x_i))$;

The final result of the strong classifier:

$$F(x) = sign(\sum_{t=1}^{T} w_t c_t(x)) \tag{4}$$

II. RECOMMENDER ALGORITHM BASED ON BOOSTING

A single recommendation algorithm has many limitations, the recommendation accuracy can be very high, while in the other indicators of performance may not be very good, there may even be a data set performance cannot adapt to the new data set, many recommendation algorithm performance is susceptible to data set effect, if the integration of multiple recommendation algorithm can avoid these problems and risks, this method has been applied in many fields, results show that the ensemble performance generally improved. A single recommendation algorithm only from the point of view of a principle or a performance indicators, such as collaborative filtering is the starting point of the target users can predict future purchase behavior heuristic rules to a certain extent based on similar user is different from the target user purchase behavior; matrix factorization algorithm based on MAE index is optimized to obtain prediction score matrix Y the one sidedness of single algorithm, so hard to avoid. In the practical application of [11] the main disadvantage of this single recommendation algorithm is the result of poor generalization ability (performance recommendations to new users and items), recommendation accuracy is not high. At present, an important way to solve the problem is to ensemble learning (ensemble learning). The main idea of ensemble learning [12] is the use of multiple recommendation algorithm in the system, such as: collaborative filtering, recommendation algorithm based on matrix factorization, the content-based recommendation algorithm based on recommended, in real time, a plurality of recommendation algorithm is recommended according to the results of a joint strategy to obtain the final results, the simplest is to vote algorithm, multiple voters to vote, final conclusion.

A. Weak recommender algorithm selection of boosting frame

If the boosting is applied to the recommendation algorithm, it is necessary to conduct the necessary improvements, mainly to the classification problem is transformed into a score: the same boosting will first of all samples (user purchase records) with a sampling weights, the same starting weight are all the same that the uniform distribution, the sample according to the weight of the sample form a data set, a recommendation algorithm on this data set, according to the error of the recommendation results can be obtained in this recommendation algorithm, we according to its error rate is assigned a weight (this is the recommended algorithm weights not sample weight), error more weight is smaller, the recommended the wrong sample we increase the sampling weights it, so will the next operation of the recommendation algorithm focuses on these last recommend error samples, then according to its error rate calculation the weight algorithm, so the iteration, the strong recommendation algorithm, finally we get is a weak recommendation algorithm weighted and, as we can see the performance a good recommendation algorithm right big some, it was the realization of boosting ensemble recommendation algorithm.

How to choose a weak recommendation algorithm is applied to the boosting framework, which is a crucial issue, because it directly affects the performance and efficiency of ensemble recommendation algorithm. We have said not all of the ensemble learning can improve the recommendation accuracy of [13, 17-19], firstly each weak recommendation algorithm recommendation accuracy cannot be less than fifty percent, otherwise the accuracy of recommendation algorithm is lower than that of random guessing (whether items will be recommended to the user) accuracy without any value, the integration does not improve the recommendation performance; the accuracy and the recommendation algorithm should not be too big difference, and if the gap is too large, may result as the accuracy of recommendation algorithm is the best, this is the most basic screening weak recommendation algorithm. Also must pay attention to the boosting algorithm characteristics, only in this way can we combine boosting integration algorithm and the algorithm of recommendation, ensemble recommendation can also play the greatest effect of final algorithm. The first characteristic of Boosting is calculated according to the weight recommendation algorithm rate error of the algorithm, of course we can experiment to find the best effect of recommendation algorithm in the experimental data sets. The second algorithm feature of Boosting is that it focuses on the error recommended sample [14, 16].

B. Rank boosting

The framework of Boosting algorithm includes the above Ada-boosting, and now popular LP-boosting and is used to solve the problem of asymmetric score Rank-boosting [15]. They have their own characteristic: Ada-boosting boosting algorithm is the most basic, is also the most widely used, the most mature; application of LP-boosting optimization principle can use fewer iterations to achieve higher accuracy, but it is in the training weak classifier requires and previously obtained classifier feature comparison, so the time efficiency is not high; and Rank-boosting mainly for scoring, can quickly achieve good promotion effect, very suitable for recommendation systems, especially for some high precision real-time recommendation system.

The specific practices are as follows:

1) And collaborative filtering, they will gene as the user, traits as objects, so that can be associated information genetic traits into gene character matrix of representative genes, columns represent characters. If there is an association between genes and traits, then the corresponding value is 1 in the gene character matrix, whereas there is no correlation is 0, so the gene character matrix is a matrix of 0 or 1 elements.

2) Using collaborative filtering based on kNN, collaborative filtering with different K values in the gene character matrix of prediction, we get many results, their prediction accuracy for each are not identical.

3) In a number of predicted results obtained on the basis that they want to further improve the prediction accuracy, on the choice of boosting ensemble learning algorithm to integrate multiple k-NN predictions, based on the actual situation on the Ada-boosting they made improvements that Rank-boosting.

4) The main improvement of Rank-boosting is: will the prediction problem as a non-symmetric scoring problems in gene, Ada-boosting is initialized to 0 and 1 of the samples treated equal, namely the weight is evenly distributed; and the increase of Rank-boosting values for the sample weight 1, is more focused on the known association training.

This paper also uses Rank-boosting to integrate various weak recommendation algorithm results, while the front Ada-boosting get weak recommendation algorithm in each iteration training weak recommendation model, in fact, the Rank-boosting does not affect the previous selections: collaborative filtering algorithm using the initial iteration, later iterations using matrix factorization algorithm based on. Just to make the following adjustments, algorithm of weak recommended optimal choice in front of the time, only in the collaborative filtering recommendation in the space of selection, recommended results behind iteration is the decomposition algorithm on matrix factorization space to choose.

C. The proposed algorithm

In this paper, the ensemble learning algorithm to recommend boosting the field, and then integrate the algorithm according to the characteristics of reasonable recommendation algorithm selected under the framework of the weak, and ultimately get ensemble recommendation algorithm in this paper. Algorithm as follows:

The users and items as sample set, is denoted by X, T number of Boosting iterations, where m is the number of weak classifiers. Operation of collaborative filtering algorithm, based on k-NN in the user item matrix SVD algorithm, for different values of two parameters k, get a series of results. The prediction of space H. $h_i(x)$ is the score I recommendation algorithm to sample x, especially with $h_i(x)$ is the first I collaborative filtering algorithm based on k-NN score for the sample x, $s_i(x)$ is the first I SVD (singular value decomposition) algorithm to sample the score of X.

Initialize: $X_0 = \{x \in X \mid x = 0\}$, X_0 represents the set of values for the sample of 0; $X_1 = \{x \in X \mid x = 1\}$, X1 represents the set of values for the sample of 1.

$$a_1(x) = \frac{1}{|x_0|} \cdot I_{x \in X_0} + \frac{1}{|x_1|} \cdot I_{x \in X_1} \tag{5}$$

The sample weights, $|X_0|$ represents the value for the number of 0 samples, |X1| represents the value for the number of 1 samples, the weights for 0 of the sample is $\frac{1}{|X_0|}$, weight is 1 samples for $\frac{1}{|X_1|}$.

1) For t=1 to T
2) For i=1 to m
3) $\lambda_i := \sum_{x_0, x_1} a_t(x_0) a_t(x_1)(h_i(x_1) - h_i(x_0))$

4) End.

5) if (p_t 〈 θ) , E_t represents the first t round when the accuracy of the strong recommendation algorithm.

6) θ is a constant, by the back of the experiment are given.

7) $c_t := \arg\max_{k_i \in H} |\lambda_i|$;

8) Else.

9) $c_t := \arg\max_{s_i \in H} |\lambda_i|$;

10) $w_t := \frac{1}{2}\ln(\frac{1+r_t}{1-r_t})$

11) $$a_{t+1}(x) = \frac{a_t(x)\exp(w_t \cdot c_t(x))}{G_t^0} \cdot I_{x \in X_0} + \frac{a_t(x)\exp(-w_t \cdot c_t(x))}{G_t^0} \cdot I_{x \in X_1}; \quad (6)$$

Among them:

$$G_t^1 = \sum_{x \in X_1} a_t(x)\exp(-w_t \cdot c_t(x)) \quad (7)$$

$$G_t^0 = \sum_{x \in X_0} a_t(x)\exp(w_t \cdot c_t(x)) \quad (8)$$

The final score: $H(x) := \sum_{t=1}^{T} w_t c_t(X)$ (9)

III. EXPERIMENTS

Experiment on the Netflix data set, verified the performance of the algorithm on the MATLAB platform. The experimental procedure is as follows:

1) TXT file data set into the MATLAB, the MATLAB is transformed into the user item matrix format.

The algorithm is designed based on the matrix form.

2) To run the diffuse algorithm in the data set (algorithm parameters from various value can produce multiple the recommendation result), and their results as basic recommendation result integration algorithm.

3) The above diffuse recommended results as the experimental data set, the algorithm is ensemble to get the final the recommended results.

4) To run the SVD algorithm user item matrix format in the first step of the data sets (algorithm parameters also provided a number of different value), these results will be the same as the basic recommendation result integration algorithm.

5) The SVD recommendation results as an experimental data set, the algorithm can get the final ensemble the result of integration.

6) Comparative analysis (4) and (5) ensemble in the experimental results, the algorithm is ensemble.

The basic recommendation algorithm boosting pre iteration using collaborative filtering in the Diffuse (diffusion) method, now need to run Diffuse on the training set (diffusion) method, must also be multiple prediction results in the formation of the choice of Rank-boosting space. Diffuse (diffusion) method is: applying the method of graph theory, the relationship between users and items into two parts graph, if there are buying relationship, between users and items with an edge, so that we can establish a recommendation network. In order to remove the impact of information transmission in repeated factors, can proceed as follows: running the diffuse algorithm in the original matrix Y, get the similarity matrix S, we can use the S^2 effect of repeated information transmission. Same S^3 can be expressed on the basis of S^2 repeat transmission. So we can use $S-\alpha S^2$ said to weight after the similarity matrix, further can use $S-\alpha S^2-\beta S^3$.Just change the parameter value, it can be run on the training set, and a series of prediction results are obtained, then these results can be as a choice. The recommendation result integration process in each iteration in the spatial prediction of selecting an optimal results as the basic recommendation algorithm produces, equivalent to not need special training a recommendation algorithm. The final integration results is not the recommended algorithm weighted summation, but directly to their recommendation results are weighted summation.

Figure 1 records the basic recommendation results of 20 iterations, each iteration of the selected, can be seen from the above process, the single highest accuracy is α =0.25 results, and in the first 20 iterations without once selected, which fully embodies the essence of the ensemble learning theory: the basic recommendation algorithm performance results into excellent results, is not necessarily the combination produced, which fully demonstrates the weak recommendation algorithm and the strong recommendation algorithm is equivalent performance, as long as the performance is better than the random guess algorithm can be transformed into more excellent performance of the algorithm, is ensemble learning. So you can see, Rank-boosting is mainly in the prediction of space and how to choose the basic recommendation algorithm and how to determine the weighted coefficient of each of them, in fact, have proved the equivalence of Ada-boosting and Rank-boosting, and they cannot achieve the best ensemble effect, this effect refers to the integration of the margin value after, namely cannot reach the maximum margin value.

Figure 2 records the process of the 20 iteration,when k=200 SVD due to the highest accuracy, the ensemble results far exceeded the recommended effect of single SVD, and we can see that the SVD k=250 recommendation effect was the best when the only to be selected again, k=450 when the SVD effect is not ideal, but in the integration process has been selected many times, or like the Diffuse method with the best recommendation effect (α =0.25) basic recommendation algorithm not in integration was selected as the best choice, but not to choose the most appropriate, ensemble learning effect once again proved.

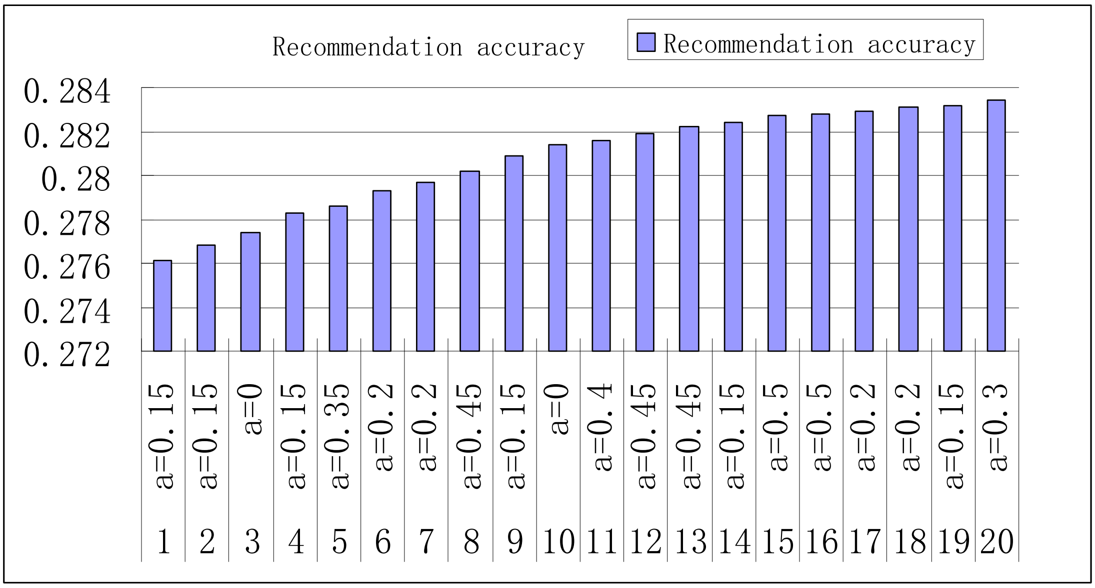

Figure 1. Ensemble resultsof diffuse algorithm

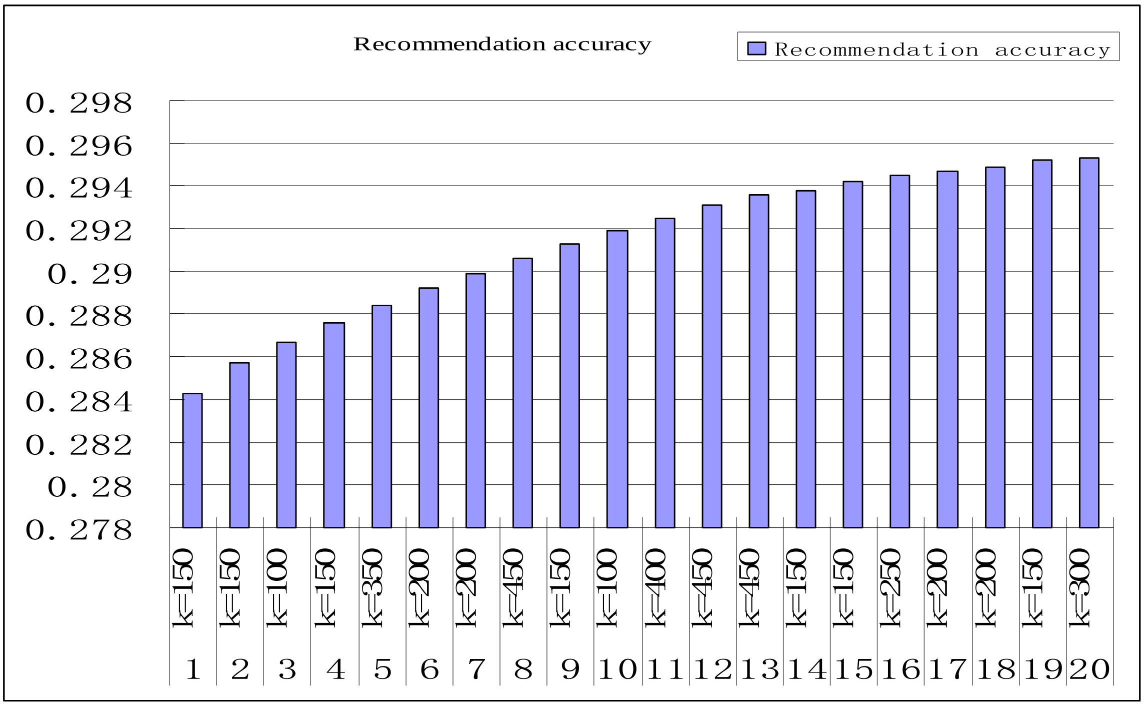

Figure 2 Ensemble results of SVD algorithm

The final experiment is the collaborative filtering (Diffuse) algorithm and SVD algorithm is ensemble. The values of θ were selected Diffuse integration process the accuracy of recommendation 0.2804 and after twelfth iterations of the precision after eighth iterations of 0.2818. Figure 3 and Figure 4 the experimental results are different θ values taken.

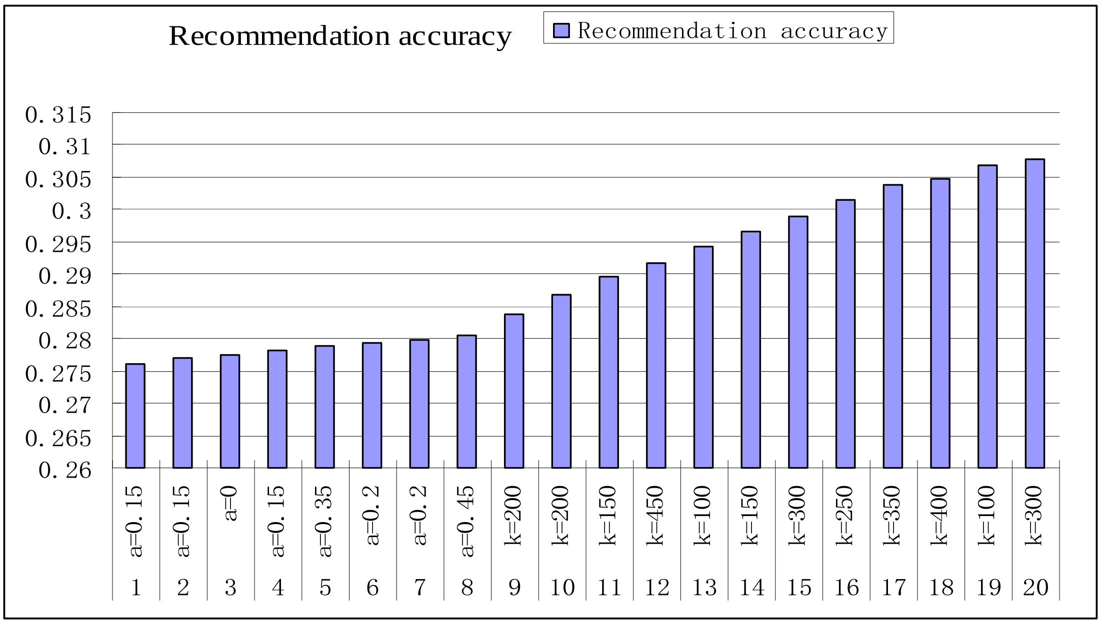

Figure 3 Ensemble results of diffuse and SVD algorithm (θ=0.2804)

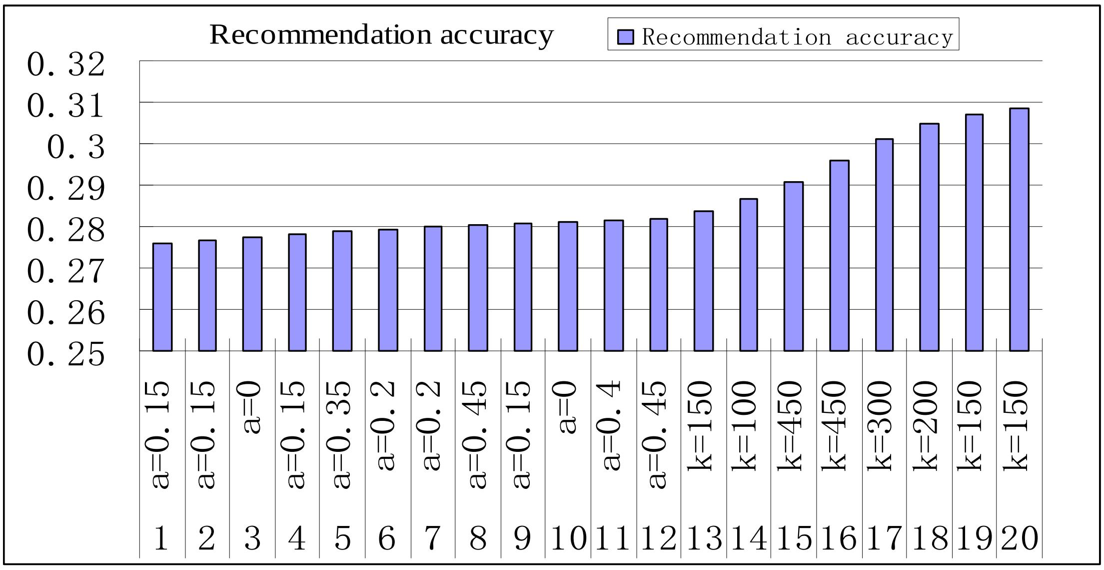

Figure 4 Ensemble results of diffuse and SVD algorithm (θ=0.2818)

The comparison shows that, in the twelfth iteration of the basic algorithm of recommendation from collaborative filtering (Diffuse method) to matrix factorization method of SVD (singular value decomposition) algorithm is better, because the basic recommendation diffuse results of 11, while the basic recommendation results only 8 of the SVD.

In order to prove that this algorithm innovation and advantage, combined with the experimental results in this paper do the following contrast shown in Table 1:

Table 1 Experiment results of various ensemble algorithms

Ensemble of recommender	Diffuse ensemble	SVD ensemble	proposed algorithm
Recommendation accuracy	0.2835	0.2907	0.3089

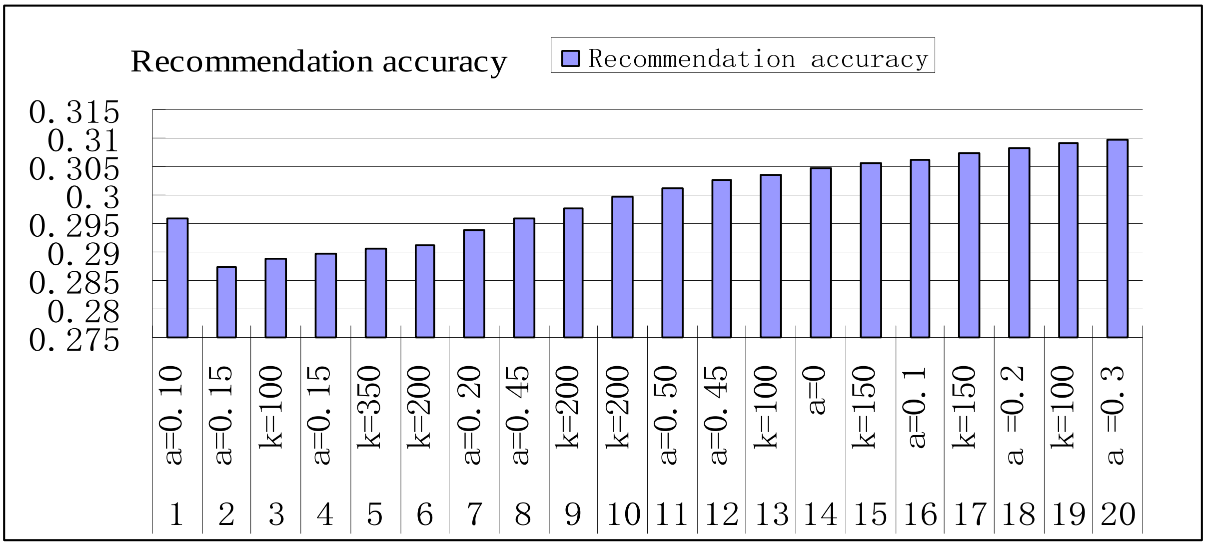

Figure 5 Ensemble experiment results in hybrid space

Figure 5 is the collaborative filtering (Diffuse) and matrix factorization algorithm (SVD) are combined to form a hybrid prediction space, then mix in combination of spatial prediction on integration, 0.3097 accuracy rate can be seen the final ensemble recommendation results, slightly higher than the algorithm 0.3087 and 0.3072 results. But the need to pay attention to the algorithm of this paper the value of θ, only from the theoretical simply two options, the result is not the optimal, can be sure of one point if the more the value of θ experiment, the final results should be better, but never more than in mixed ensemble forecast effect space. No more than 0.3097, that is to say, this algorithm through theoretical inference and not for a large number of experimental attempts to reach such a high degree of accuracy, at this point has been made great improvement.

IV. CONCLUSIONS

This paper introduces the principle of mainstream recommendation system recommended, a single recommendation algorithm accuracy is not high, affecting the performance of instability vulnerable to data set, and boosting integration algorithm only need to increase the basic recommendation algorithm without changing the algorithm previously recommended in iteration and its weight, but also can increase the strong recommendation algorithm to improve the ability to adapt to the new data boundary set which is to enhance the generalization ability, and the boosting algorithm from the collaborative filtering algorithm selected two as the framework of boosting weak recommendation algorithm, finally also combines the Rank-boosting algorithm's advantage, not at each iteration training weak recommendation algorithm, but the operation of a variety of recommendation algorithm recommended results in the formation of spatial prediction in the iteration before, the results are weighted and achieve the integration of purpose is the strong recommendation results.

ACKNOWLEDGMENT

The work was supported by the Zhejiang Provincial Department of education research project << multidimensional nested rural electronic commerce application research under the regional environment>> (Project Number:Y201329505), and The national education information technology research of Twelfth Five Year Plan<< multidimensional nested rural electronic commerce application personnel training research under the information technology environment >>(Project Number: 136231133).

REFERENCES

[1] Koren Y, Factorization meets the neighborhood: A multifaceted collaborative filtering model, Proceedings of the 14 the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 426-434, 2008.

[2] Tao Zhou, Jian-Guo Liu, Joseph Rushton Wakeling, Yi-Cheng Zhang, Solving the apparent diversity-accuracy dilemma of recommender systems. Applied Physical Sciences, pp, 4511-4515, 2010

[3] Andreas Toscher, Michael Jahrer, The Bigchaos Solution to the Netflix Prize 2008. Journal of Machine Learning Research, pp.1-17, 2008

[4] Paterek A, Improving regularized singular value decomposition for collaborative filtering, Proceedings of KDD Cup and Workshop, pp.39-42, 2007.

[5] Koren Y, Factor in the neighbors: Scalable and accurate collaborative filtering, ACM Transactions on Knowledge Discovery from Data, vol.4, no.1, pp.1-24, 2010. .

[6] Mingrui Wu, Collaborative Filtering via Ensembles of Matrix Factorizations Journal of Machine Learning Research, vol.3, pp. 43-47, 2009.

[7] Thomas G. Dietterich, Ensemble learning, In the Handbook of Brain Theory and Neural Networks, Second Edition, 2002.

[8] Sill J, Takacs G, Mackey L, Shawe-Taylor J, Linear programming boosting via column Generation, Machine Learning, vol.46, no.1-3, pp. 225-254, 2002.

[9] T.G. Dietterich. Machine Learning Research: Four Current Directions. AI Magazine, vol.18, no.4, pp. 97-136, 1997.

[10] G. Sen Gupta, S.C.Mukhopadhyay, S. Demidenko and C.H.Messom, “Master-slave Control of a Teleoperated Anthropomorphic Robotic Arm with Gripping Force Sensing”, IEEE Transactions on Instrumentation and Measurement, Vol. 55, No. 6, pp. 2136-2145, December 2006.

[11] Fang Yuke, Fu Yan, Zhou Junlin, Li Siyu, improved neighborhood-based algorithms for large-scale collaborative-filtering, Journal of Computational Information Systems,vol.7, pp.2695-2702, 2011.

[12] G. Sen Gupta, S.C. Mukhopadhyay, Michael Sutherland and Serge Demidenko, Wireless Sensor Network for Selective Activity Monitoring in a home for the Elderly, Proceedings of 2007 IEEE IMTC conference, Warsaw, Poland, (6 pages).

[13] Rada Mihalcea, Classifier Stacking and Voting for Text Filtering, 11th Text retrieval, Information technology, Conference, pp. 775-780, 2002.

[14] N.K.Suryadevara, A. Gaddam, R.K.Rayudu and S.C. Mukhopadhyay, “Wireless Sensors Network based safe Home to care Elderly People: Behaviour Detection”, Sens. Actuators A: Phys. (2012), doi:10.1016/j.sna.2012.03.020, Volume 186, 2012, pp. 277 – 283.

[15] Yoav Freund, Raj Iyer, Robert E. Schapire, Yoram Singer, An Efficient Boosting Algorithm for Combining Preferences, Journal of Machine Learning Research, vol.4, pp.933-969, 2003.

[16] Warmuth M, Glocer K, Vishwanathan, Entropy regularized LP boosting, Proceedings of International Conference on Algorithmic Learning Theory, pp. 256-271, 2008.

[17] N. K. Suryadevara and S. C. Mukhopadhyay, “Determining Wellness Through An Ambient Assisted Living Environment”, IEEE Intelligent Systems, May/June 2014, pp. 30-37.

[18] Hsuan-Tien Lin, Ling Li, Large-Margin Thresholded Ensembles for Ordinal Regression: Theory and Practice, Journal of Machine Learning Research, pp. 319-333, 2010.

[19] Dwivedi R., Kumar, A., Ghosh S.K., Visualization of uncertainty using entropy on noise clustering with entropy classifier,2013 IEEE 3rd International Advance Computing Conference (IACC), pp.1293 - 1299, 2013.

[20] Jing Peng, Seetharaman G, Combining the advice of experts with randomized boosting for robust pattern recognition, 2013 IEEE Applied Imagery Pattern Recognition Workshop: Sensing for Control and Augmentation, pp.1-7, 2013.

[21] Leong Wai Yie, Joel Than Chia Ming, Features of Sleep Apnea Recognition and Analysis, International Journal on Smart Sensing and Intelligent System, vol.7, no.2, pp.481 – 497, 2014.

[22] Khairul Anam, Adel Al Jumaily, Yashar Maali, Index Finger Motion Recognition using self-advise support vector machine, International Journal on Smart Sensing and Intelligent System, vol.7, no.2, pp. 644 – 657, 2014.