ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

AdaP-Boost: An Iterative Personalized Adaptive Boosting Algorithm for Improving Recommender Stability

R.Carolene Praveena[1], S.Periyasamy[2]

Assist Professor, Dept. of Computer Science, Sri Jayendra Saraswathy College of Arts and science, Coimbatore, TamilNadu, India [1]

M.Phil Scholar, Dept. of Computer Science, Sri Jayendra Saraswathy College of Arts and Science, Coimbatore, TamilNadu, India[2]

ABSTRACT: E-shopping is one of the major emerging trends among busy people. These days' social networks have improved a lot and many users' desires to share their opinions about their experience. Many Recommender System (RS) have followed the above factors for optimal suggestion to shop. Even though, the results are optimal and reliable, the system suffers from classification and over-fitting problem. Some of the social factors have been used in RS; still they have not been fully measured. The personalization can't be determined with only social similarity, this also need to be identified by their personalized attributes. While considering the other attributes and factors, this created a dimensionality problem. To overcome the challenges in the RS, the current work proposed a superfine model named as Iterative Recommender System (IRS), which combines user's profile, their intrapersonal and interpersonal interest similarity along with the interpersonal influence. This system developed based on the traditional boosting technique and the proposed iterative Recommender System, which can be used to improve accuracy and adaptive feature of RS. The AdaP-boosting (adaptive Personalized boosting) approach extracts numerous training samples from the original dataset and combine predictions made based on these samples to form an aggregate final prediction for each user. The AdaP-boosting algorithm uses numerous iterations to repeatedly and explicitly adjust predictions of items for recommendations are based on its other predictions in order to make them more consistent with each other.

KEYWORDS: Recommender systems, collaborative filtering, recommendation stability, iterative smoothing, bagging, boosting.

I. INTRODUCTION

In the recent scenario, web data analysis on e cart domains has tremendous growth. There are more and more applications and companies provide online services nowadays. However, the rapid growth of online shopping information imposes an increasing challenge for users who have to choose from a large number of available products from shopping networks for satisfying their personalized needs. Additionally to increase the reliability of product selection has to understand the preferences from different users and serve more attractive products. Due to its enormous growth, product recommendation is very tedious. In traditional recommendation systems (RS), RS uses rating of products for recommendation, i.e. products with higher rating will be recommended to the user. Hence the demand for intelligent shopping services is expected to increase dramatically. This insists the system to create an accurate online product recommendation and alert system. In research perspective, there is only few researchers' developed solution for huge over-fitting and NP-complete problem in RS. However several studies failed to perform the personalized product selection with optimal performance improvement.

II. LITERATURE REVIEW

Product and service recommendations in social networks are very tedious and difficult to find individual interest factor. For this, recommender systems have been proposed with effective data mining techniques [1]. However the recommender system implemented successfully, several problems arises in the recommender systems, which are

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

classification and over-fitting problems. In Classification scenarios only ranking and ratings are available in service recommendation. And in over-fitting problem, new item are added regularly to RS (Recommender System) and the items are dynamically changing, this will cause random error. Later CF (Collaborative Filter) systems [2] rely solely on users' preferences to make recommendations. Therefore the recommender system would not be able to recommend the item until the new item is rated by a substantial number of users. When considering any recommender system in this criterion, the number of ratings already obtained is usually very small compared to the number of ratings that need to be guessed. Hence, effective and stable rating prediction from a small number of examples is very significant in such recommender systems. The success of the collaborative recommender system [3] depends on the availability of a critical collection of users and their ratings. Consider the product recommendation system where, there may be many products that have been rated by only few people and these products would be recommended very rarely in the products recommender, if those few users gave high ratings to those items. Also, for the user whose interests are unusual compared to the rest of the users, there will not be any other users who are particularly similar, leading to weak recommendations. In existing approaches [4][5] they didn't recommend the product user. In existing system, most techniques are based on the parameters such as interest, rating and usage behaviors. If the user use the alternative system means recommendation is not suitable for them.

The existing system discovered two general approaches, one is bootstrap aggregation and another one is iterative smoothing. The both techniques improved the stability of any given recommendation technique. bootstrap aggregation (or bagging) is an already widely used approach in various data mining and machine learning applications, its potential usefulness with respect to the stability improvement has not been established or discussed before. Even though bootstrap aggregation is useful, that does not provide general finite-sample guarantees. Due to the "aggregating" nature of bagging, the existing system proposed bagging as a natural initial meta-algorithmic approach toward stability improvement, even though bagging represents an indirect approach to improving stability. The iterative smoothing algorithm helps to predict the ratings of a recommendation based on several iterations, this degrades the performance. The followings are the disadvantages of using existing approaches, the techniques are only suitable if the items and algorithms are similar, if new item arrive then the system failed to predict. The paper [6] doesn't consider the feedback of RS. i.e.," how many users have responded to the RS". And the existing system failed to perform user behavior studies to investigate the value of stable recommendations.

III.PROBLEM DEFINITION

A. CHALLENGES:

There are some potential problems with the CF Recommendation System. One is the scalability, which is how quickly a recommender system can generate Product recommendation; another one is over-fitting and also classification Problem.

i. Data Over-fitting:

The possibility of over fitting exists because the criterion used for training the prototype is not the same as the criterion used to judge the efficacy of a prototype. In specific, a model or prototype is typically trained by maximizing its performance on some set of training data. In order to avoid over fitting problems in RS, it is necessary to use additional techniques such as regularization and early pruning process etc. The number of products specified on major shopping sites is extremely large such as flipkart, snap deal like shopping sites. The majority of users will only have rated a small subset of the overall databases. Thus even the most popular products have very few rankings in the real time, so there is a need to switch the parameter according to the result.

ii. Classification Problem:

The *classification* problem occurs when a new user or item has just entered the system; it is difficult to find similar ones because there is not enough information. In some literature there is a *classification* problem which is also called the *new user problem* or *new item problem*. Because supervised learning process doesn't always suitable if the training data is insufficient. New products cannot be recommended until some user's rate it, and new users are unlikely given good Product recommendations because of the lack of their ranking or shopping history.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

iii. Scalability:

In many of the environment that these systems make product recommendation in interest and ranking based. There are millions of users and products available in the real-time. Thus a large amount of computation power is often necessary to calculate product recommendation. For example, with tens of millions of customers (*M*) and millions of distinct catalog products (*N*) a *CF* algorithm with the complexity of $O(n)$ is already too big. Many systems need to react immediately to online requirements and make Product recommendations for all users regardless of their interest and rankings log which demands a high scalability of a *CF* system.

IV. PROPOSED SYSTEM

An alternative approach for Recommender system with similarity computation of the user interest based on different factors has been introduced. Our proposed scheme improves upon that in [7] and [8] in terms of reliability, efficiency and accuracy. In particular, the proposed scheme depends on the existing CF and baseline based recommender algorithms, which significantly reduces communication and computational costs.

Our proposed system explores two general-purpose and practical meta-algorithmic approaches—based on the traditional boosting technique and the proposed iterative recommendation system technique—that can be used to improve stability of a wide variety of state-of-the-art recommendation algorithms using different parameters. The Ada-P-Boosting approach extracts multiple training sub-samples from the original dataset and combine predictions made based on these samples to form an aggregate final prediction for each unknown rating. Our system also reduces the over fitting problem along with prediction issues. The iterative smoothing approach uses multiple iterations to repeatedly and explicitly adjust predictions of a recommendation algorithm based on its other predictions in order to make them more consistent with each other.

This proposes a personalized iterative recommendation system which combining user usage, user profile, intrapersonal and interpersonal interest similarity. The thing of user individual interest makes associations between user and product with unseen parameters and unknown ratings of products. The followings are the contributions of the proposed work.

- Our proposed work introduces a new approach named as Iterative Recommendation system (IRS), which is based on Adaptive personalized Boosting technique.
- Our system predicts unknown ratings based on personalized factors.
- Identifying and analyzing user behavior to investigate the value of stable recommendations.
- The proposed RS utilizes interpersonal and intra personal similarities from the user log.
- Creates a memory based collaborative filtering for online shopping product recommendation.
- Identifying user's usage patterns and how many users accepted the recommendations after the suggestion. Appling a new recommender system with computational overhead reduction.
- This has be applied an ensemble approach, which involves training a learning algorithm to combine the predictions of several other learning algorithms.

The main contributions of our proposal are summarized above. The influence of the additional factors in the recommendation system with performance metrics will improves the stability of recommender systems.

V. ITERATIVE RECOMMENDER SYSTEM

In AdaP-boosting other user's interest and log will be taken as training data for another user. After the successful training phase the system performs the test phase of individual interest measures, all suggestions and recommendations are taken in this phase. Finally this predicts the rating and user interested products with high correlation. In order to perform the stable prediction and recommendation, our system follows the following algorithm initially.

Algorithm 1: Initial rating prediction
Input: Known ratings D, user usage records U
Number of training samples T
Output: predicted ratings and product suggestion algorithm
Steps:

1. Extract training samples' and find the average of T using Avg(T)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

2. For each Ds s € {1,2….T}
3. Train *f*s on T with several base learners L on the first part s[1].
4. Test the base learners ti=L(T) on the second part.
5. Using the predictions from step 3 as the inputs, and the correct responses as the outputs, train a higher level learner.
6. Find known rating k, and predicted rating k1 using T.
7. Return K and classifier C

The basic principle of **AdaP-boosting is** to pursue iterative Solutions, which has been gathered from iteration. The iterative solution performs the following steps to evaluate the classifier weightage.

Algorithm 1: Ada-P_Boosting
Input: predicted ratings along with its classifier
Number of iterations I.
Output: final recommendation (R)
Steps:

0) Set $\tilde{W}_i^{(0)} = 1/n$ for $i = 1, \ldots, n$

1) At the m^{th} iteration we find (any) classifier $h(\mathbf{x}; \hat{\theta}_m)$ for which the *weighted classification error* ϵ_m

$$\epsilon_m = 0.5 - \frac{1}{2}\left(\sum_{i=1}^{n} \tilde{W}_i^{(m-1)} y_i h(\mathbf{x}_i; \hat{\theta}_m)\right)$$

is better than chance.

2) The new component is assigned votes based on its error:

$$\hat{\alpha}_m = 0.5 \log((1 - \epsilon_m)/\epsilon_m)$$

3) The weights are updated according to (Z_m is chosen so that the new weights $\tilde{W}_i^{(m)}$ sum to one):

$$\tilde{W}_i^{(m)} = \frac{1}{Z_m} \cdot \tilde{W}_i^{(m-1)} \cdot \exp\{ -y_i \hat{\alpha}_m h(\mathbf{x}_i; \hat{\theta}_m) \}$$

The similarity value calculation between users in the same category by means of interpersonal and intra personal as well as product based is important task in IRS. Our system proposed the effectiveness of IRS model with consideration of individual preference, interpersonal inspiration and intra personal inspiration. The system considers the independence of user interest in the e shopping domain. It means this can recommend items based on user interest at a certain extent this also utilizes user's association with the items to train the hidden feature vectors in boosting algorithm, especially for the existing and more expecting users. The system also considers the Interest circle inference technique. As per the boosting algorithm this segments the social network into several sub-networks and each of them correspond to particular item collection. To overcome the classification users who has a few rating records from the ulog then the ratings of their associated user's interest in the same category to link user interest products. Due to the individuality especially users with huge rating records in e shopping domain, users usually choose products all by themselves with little inspiration others. To provide the product recommendation without affecting individuality for experienced users, the system proposed an optimal personalized recommendation system. The significance of user and item depends on the relevance of user interest *Tu* and item topic *Ti* to a certain domain this takes several attributes such as product category, company name, price and offers. This denoted the relevance of user T's personal interest to the category of item *i* in the IRS model by *IRSu,i*

IRSu,i = Sim(Tu,Ti,Fsi), where Fsi is the feedback of previous suggested item

This performs the similarity measure by fine filtered attributes. The derived products should satisfy the personal interest and as well as social inspiration without affecting their attribute consideration.

(Pi) € sim (user1, user2, Pn)

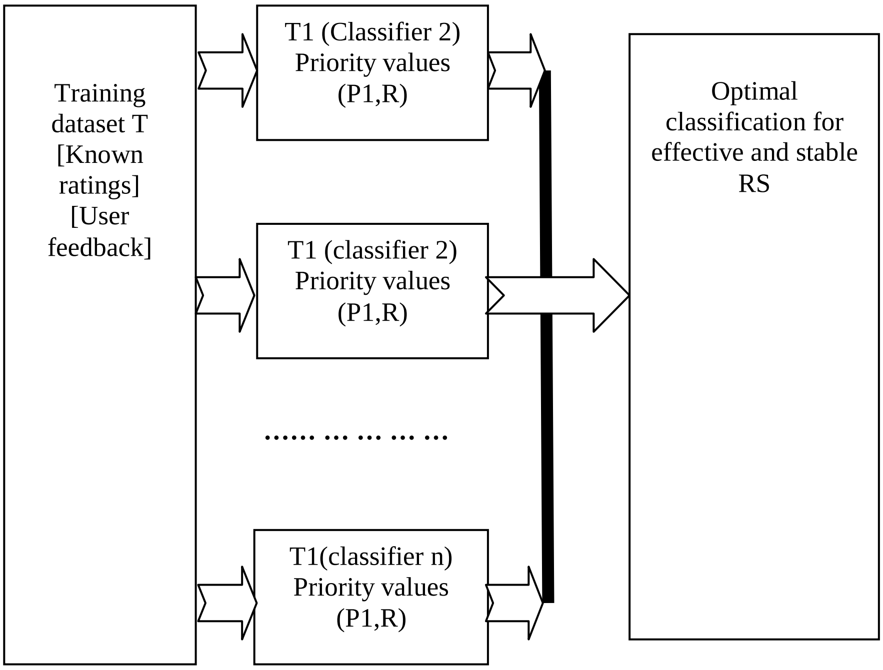
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Where, Pi is the priority list. From the above equation the similarity has been calculated between two users with the consideration of any particular attribute such as price, ie this measures whether the average price spent by user 1 is similar to user2, based on this the system suggests products to user 1 from user 2 log. This also considers the feedback of previous recommendations.



Evaluates and prunes at every iteration

Fig 1.0 : Ada-P Boosting process

From the above fig 1.0, the Ada-P-Boosting process has been identified. Initially all known rating and other details from shopping log are taken as training dataset T. At every iteration, the classifier find the priority product list (P) based on the rate (R) and other factors. Finally after all iterations, the system finds the stable classifier and re evaluates the product priority list according to the user's interest and need.

VI. IMPLEMENTATION AND RESULTS

Our proposed system explores the use of objective features to model the subjective perception of n number of products which have been received from every user. This presents a social circle and location-based similarity measure for product Retrieval system which evolves and uses different product similarity measures for individual users. In particular, user supplied training allows the system to determine which subset of features approximates more efficiently the subjective product similarity of a specific user in different iterations. This section describes the implementation process. Implementation is the realization of an application, or execution of plan, idea, model, design of a research. This section explains the software, datasets and modules which are used to develop the research.

A. Software

The experiments are performed on an Intel Dual Core with a RAM capacity 2GB. The algorithms are implemented in ASP.Net for shopping site creation and C#.NET as coding language and are run under Windows family.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

B. dataset

The system mainly deals with the problem of classification and over fitting problem. So this need a set of data related to the shopping products with respective user profile and shopping logs. The data collected from real time websites such as amazon, flipkart and snapdeal shopping, but some attributes are not available in the website, so the experiments taken the synthetic dataset which is described below.

Table 1.0 Comparison table

Techniques	**Number of positive responders**	**Number of negative responders**
Iterative smoothing	32	98
Proposed(AdaP-boosting)	91	39

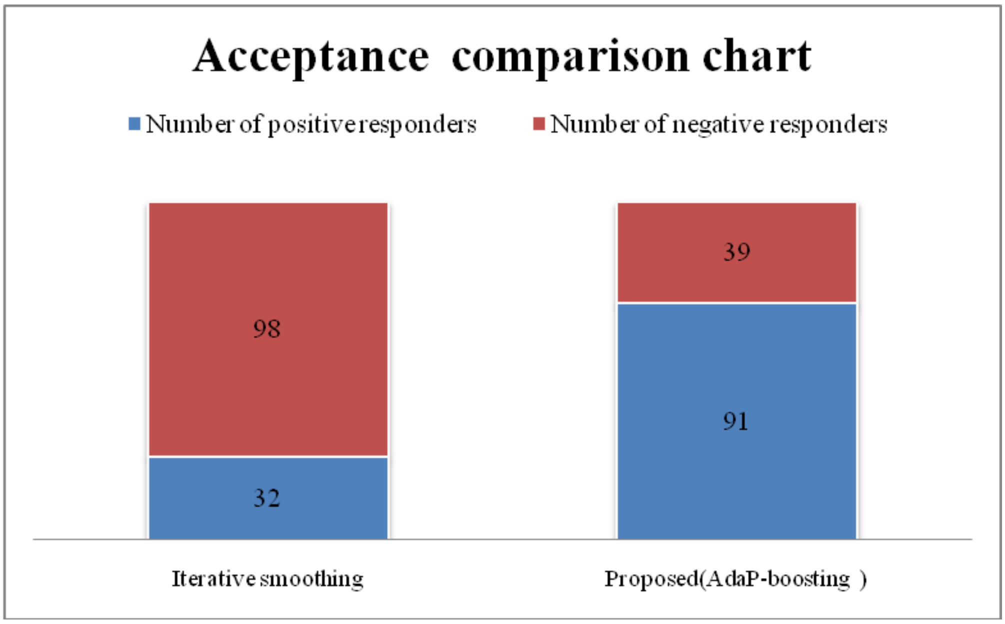


Fig 2.0 : Comparison chart

To evaluate the performance of the proposed schemes, user feedbacks have been taken as the main parameter for recommendation accuracy evaluation. Without loss of generality, this defines by the number of positive and negative responders from 130 users totally. From the fig 2.0 iterative smoothing techniques have 32 positive responders from total 130 users. And our proposed work has 91 positive responders from 130 users. This shows, the proposed recommendation improves the stability of RS gradually with the user behaviour log.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

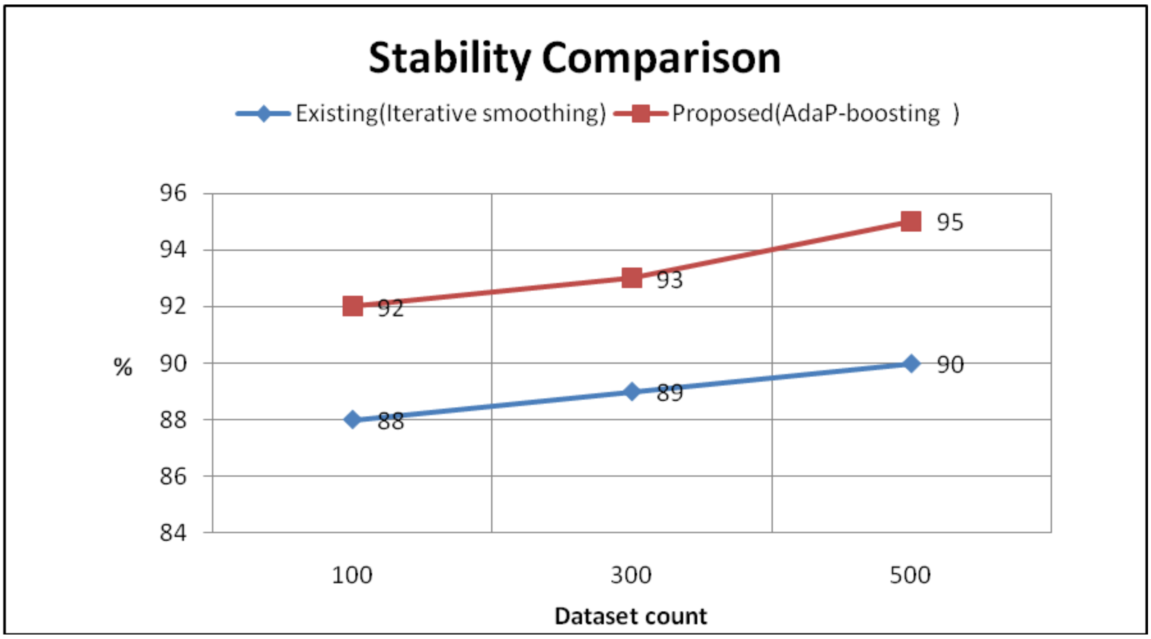


Fig 3.0: Stability comparison chart

To evaluate the stability accuracy of the proposed schemes using various factors such as user profile, interests, inter personal similarity along with user feedbacks have been take. The above fig 2.0 shows the proposed system yielded better stability in every experiment than the existing bagging technique.

Parameters	Existing system	Proposed system
Algorithm used	bootstrap aggregation Iterative smoothing.	IRS(iterative Recommender system) AdaP-Boosting
advantages	Helps to predict the ratings	Helps to predict the ratings and suggests product with personal interest measure Priority based iteration Suitable for all
Handling missing values	No	Yes
user behavior studies	No	Yes
Time consumption	High	Medium
Accuracy	Medium	High
Error rate	High	low

The performance of this proposed work AdaP-boosting and CF Scheme is compared with the existing approach Iterative smoothing algorithm with various factors. The table 2.0 shows the performance comparison of the proposed method with other existing approaches based on different parameters.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

VII. CONCLUSION

A personalized recommendation and alert approach has been proposed by combining several real time factors known as profile based, personal interest based, interpersonal interest similarity, intra personal similarity and product influence with various attributes. This overcomes the classification and recommendation problem, which is common in recommender system. To overcome the above issues, the system implemented improved boosting ensemble algorithm with memory based collaborative filtering techniques. This approach is implemented in an e-shopping dataset with relevant similarity measurement phases. This measures individuality of rating items with the reference of experienced users with various factors. At present the personalized recommendation model in the literature only takes interpersonal relationship and user's historical rating records. In the proposed AdaP-boosting, the system takes the area information and other attributes to recommend more personalized and real-time items to the users.

REFERENCES

[1] X.-W. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proc. 18th ACM SIGKDD Int. Conf. KDD*, New York, NY, USA, Aug. 2012, pp. 1267–1275.

[2] M. Jiang *et al.*, "Social contextual recommendation," in *Proc. 21st ACM Int. CIKM*, New York, NY, USA, 2012, pp. 45–54.

[3] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. ACM Conf. RecSys*, Barcelona, Spain, 2010, pp. 135–142.

[4] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proc. NIPS*, 2008.

[5] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[6] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis

[7] M. Jahrer, A. Toscher, and R. Legenstein. "*Combining predictions for accurate recommender systems*". KDD'10, pp. 693-702, 2010.

[8] H. Ma, I. King, and M. R. Lyu. "*Learning to recommend with social trust ensemble*". SIGIR'09, 2009.

[9] X. Yang, Y. Guo and Y. Liu. "*Bayesian-inference based recommendation in online social networks*". The 30th Annual IEEE International Conference on Computer Communications (INFOCOM), 2011.

[10] J. Huang, X. Cheng, J. Guo, H. Shen, and K. Yang. "*Social recommendation with interpersonal influence*". In Proceedings of the 19th European Conference on Artificial Intelligence (ECAI), pp.601–606, 2010.