UNIVERSITI PUTRA MALAYSIA

AN IMPROVED RECOMMENDER SYSTEM BASED ON NORMALIZATION OF MATRIX FACTORIZATION AND COLLABORATIVE FILTERING ALGORITHMS

AAFAQ ZAHID

FSKTM 2015 25

AN IMPROVED RECOMMENDER SYSTEM BASED ON NORMALIZATION OF MATRIX FACTORIZATION AND COLLABORATIVE FILTERING ALGORITHMS

By

AAFAQ ZAHID

Thesis submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Master of Sciences (MS)

February 2015

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Sciences.

AN IMPROVED RECOMMENDER SYSTEM BASED ON NORMALIZATION OF MATRIX FACTORIZATION AND COLLABORATIVE ALGORITHMS

By

AAFAQ ZAHID

February, 2015

Chair: Nurfadhlina Mohd. Sharef, PhD
Faculty: Computer Science and Information Technology

Recommendation System (RS) came to lime light when the information on the internet started growing to the extent that it became time consuming to get the required information. There are different techniques used in RS. Some works are based on user past knowledge known as Content Based (CB) while more popular techniques referred to as neighborhood models (CF and MF) are based on finding similar users for recommendation. Existing techniques have certain drawbacks such as user getting the same information. This problem is known as stability versus plasticity (in CB). Another problem called cold start gives wrong recommendations amongst new users as data of new users is not enough for recommendation. Other limitations include too much dependence on other users or no consideration of user personal preferences (CF and MF). There is a technique known as normalization which develops models like user involvement in subject matter or user likeness according to the details of item to predict ratings to user. Normalization shows good results but it is truly personalized from single user perspective and lacks other user's opinion for the recommendation. Some researchers combine different techniques into hybrid to overcome the problems in RS, but there is very limited work that has investigated the effect of hybridizing normalization technique on neighborhood models. Therefore, this research is dedicated to combining the normalization technique with neighborhood models (CF and MF) to produce CF+N (collaborative filtering and normalization) and MF+N (matrix factorization and normalization). The hypothesis is that the tendency of normalization technique to simplify the data combined with the accuracy of the neighborhood models can improve the accuracy of the RS. This hybrid technique rates user personal preferences more than other user's

recommendation towards the final recommendation, while still considering user's personal recommendation as important input in the process. Several experiments have been conducted on the movielens dataset where 80% of data is used as training set while 20% is used as test set. The experiments are designed to perform the comparisons with the existing works that target to solve the existing problems in RS. There are three categories of evaluation of RS predictive accuracy metrics, classification accuracy metrics and rank accuracy metrics. This study follows MAE and RMSE from predictive accuracy metrics for evaluation of results since the main focus of the study is to reduce errors in RS. Results show that MF+N unite well as hybrid technique where the gray sheep is handled by MF and normalization manages cold start, mood changes, stability versus plasticity and difference of opinion. On the contrary, CF+N technique requires some enhancements as the results were below expectations because of the tendency of CF to produce big differences in the prediction of raw data. It is concluded that the resultant hybrid techniques can perform well if the variables provided to normalization by neighborhood model (MF and CF) do not have big differences in order for the hybrid normalization model to outperform every algorithm in comparison.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
Sebagai memenuhi keperluan untuk ijazah

SISTEM CADANGAN YANG DIPERTINGKATKAN BERDASARKAN NORMALISASI MATRIKS PEMFAKTORAN DAN ALGORITMA PENAPISAN KERJASAMA

Oleh

AAFAQ ZAHID

Februari 2015

Pengerusi: Nurfadhlina Mohd. Sharef, PhD
Fakulti: Sains Komputer dan Teknologi Maklumat

Sistem cadangan telah mendapat tumpuan apabila maklumat di dalam internet semakin bertambah, kerana dengan bertambahnya maklumat maka jangkamasa untuk mendapatkan maklumat yang berkaitan menjadi lebih panjang. Terdapat beberapa teknik yang digunakan di dalam sistem cadangan. Beberapa karya berkaitan pengetahuan pengguna yang lepas dikenali sebagai berasaskan kandungan (CB) manakala kaedah yang lebih popular adalah berdasarkan pencarian pengguna yang sama untuk dicadangkan kadang-kala dirujuk sebagai model kejiranan. Contoh teknik-teknik di dalam model kejiranan adalah Penapisan Bekerjasama (CF) dan Pemfaktoran Matriks (MF). Kaedah-kaedah ini mempunya beberapa kelemahan seperti pengguna selalu mendapat maklumat yang sama dikenali sebagai kestabilan lawan *plasticity* (di dalam CB). Mereka juga mempunya keterhadan di kalangan pengguna baru yang mana data tidak mencukupi untuk cadangan lalu memberikan cadangan salah dan maklumat ini dikenali sebagai masalah mula sejuk. Kekangan-kekangan lain termasuk kebergantungan kepada pengguna secara berlebihan atau kadang-kala tidak menimbangkan pilihan pengguna (CF dan MF). Terdapat juga teknik-teknik seperti penormalan yang membangunkan model seperti penglibatan pengguna di dalam perihalan atau kesukaan pengguna terhadap perincian perkara untuk meramal penilaian pengguna. Penormalan menunjukkan keputusan yang baik tetapi ia adalah terlalu personalisasi bagi perspektif pengguna tunggal dan kekurangan dalam pengalaman pengguna yang sama bagi cadangan. Beberapa penyelidik telah menggabungkan teknik-teknik ini secara gabungan bagi mengatasi beberapa masalah seperti mula sejuk. Teknik gabungan masih mempunyai cadangan pengguna-pengguna lain dengan menetapkan lebih pemberat tetapi dengan ketepatan yang lebih rendah. Penyelidikan ini menggabungkan penormalan dengan model kejiranan bagi mengatasi kekurangan personalisasi dan mengeksploitasi

pilihan pengguna di dalam model kejiranan. MF+N (Pemfaktoran Matriks dan Penormalan) dan CF+N (Penapisan Bekerjasama dan Penormalan) adalah digabungkan di dalam teknik cadangan hybrid yang mana telah meminimakan masalah seperti lambat mula dan kestabilan lawan *plasticity*. Teknik gabungan ini menilai pilihan pengguna peribadi lebih daripada cadangan pengguna lain kepada cadangan akhir, manakala masih menimbangkan cadangan pengguna lain sebagai input penting di dalam proses keseluruhan. Eksperimen ke atas dataset *movielens* telah dijalankan dan keputusan menunjukkan bahawa MF+N dan CF+N menaikkan ketepatan sistem cadangan berbanding model kejiranan. Dapat disimpulkan bahawa jika parameter yang diberikan kepada normalisasi oleh model kejiranan (MF dan CF) tidak mempunyai perbezaan yang besar sepertimana MF, model gabungan normalisasi dapat mengatasi algoritma lain yang dibandingkan.

ACKNOWLEDGEMENTS

I want to thank my supervisor whose tireless efforts make this research a reality. I also want to thank my brother in law and sister who help me when it matters the most, my wife to keep me moving when I give up, my friend Kow and Ken who help me whenever required putting their commitments behind and last but not the least my MOM for her endless prayers.

My special thanks to UPM for the grant as this research is partly sponsored by the Research University Grant Scheme under Universiti Putra Malaysia in project titled Conversational System for Museum Tour Planning Personalization and Recommendation.

APPROVAL

Approval Sheet 1

I certify that a Thesis Examination Committee has met on 10th Feburary, 2015 to conduct the final examination of Aafaq Zahid on his thesis entitled "An Improved Recommender System Based On Normalization Of Matrix Factorization and Collaborative Algorithms" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science (MS).

Members of the Thesis Examination Committee were as follows:

Norwati Mustapha, PhD
Associate Professor;
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Rusli bin Hj Abdullah, PhD
Professor;
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Ali Selamat, PhD
Professor;
Department of Software Engineering, Faculty of Computing
Universiti Teknologi Malaysia
Malaysia
(External Examiner)

_____
ZULKARNAIN ZAINAL, PhD
Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia
Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Nurfadhlina Mohd. Sharef, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Aida Mustapha, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Alicia Tang Yee Chong, PhD
Senior Lecturer
Centre of Agent Technology, College of Information Technology
University of Tenaga Nasional
(Member)

_____

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No: _____

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of
Chairman of
Supervisory
Committee: _____

Signature: _____
Name of
Member of
Supervisory
Committee: _____

Signature: _____
Name of
Member of
Supervisory
Committee: _____

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CB	Content Based
CF	Collaborative Filtering
MF	Matrix Factorization
RS	Recommender System
ED	Euclidean Distance
HD	Hellinger Distance
SVMReg	Support Vector Machines Regression
CCF	Clustering-based Collaborative Filtering
CME	Cluster with minimum Error
DC	Dynamic Classification
KBS	Knowledge Based Systems
NMAE	Normalized Mean Absolute Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
Tf	Term Frequency
Idf	Inverse Document Frequency
AI	Artificial Intelligence

CHAPTER I

INTRODUCTION

1.1. Background

The information overload in recent times has formulated recommender systems (RS) as an important tool to handle this problem. The RS has been an active research area since 90's and a lot of methods and techniques have been used in the process of making it more reliable. The RS usually work by storing the information of the user in a user profile. The user profile contains the interests of the user which help the system to recommend the user about the appropriate places of interest.

There is a wide range of RS like news recommendation (Frasincar, IJntema, Goossen and Hogenboom, 2011), e-commerce (Schafer, Konstan, and Riedl, 1999), restaurant recommender (Burke, 2002a), and many others. These recommenders use different techniques for their recommendation process such as Content-based (CB), Collaborative Filtering (CF), Matrix Factorization (MF), demographic, semantic, etc. (Montaner, López, and Rosa, 2003).

These techniques provide different ways to do recommendation, but there are lots of hitches. The RS mostly works on saving user profile which updates over time as the user uses different items and provides the feedback to the system. This ever updating profile helps the system to update the user's preference and provide suitable feedback to the user. User profile way of recommendation is accurate and mostly techniques use profiles in altered forms. However, the problem arises if the user has just joined the system and his profile is empty. Then these recommendation techniques cannot recommend items to the user. This problem is called cold start problem (Son, 2015).

In order to deal with cold start problem the literature divides different techniques in 3 groups: (i) using user additional data; (ii) putting the new user in most prominent groups; (iii) combining hybrid methods to give predictions. The first group requires additional data about the user. This additional data can contain user's demographic information, social information or some starting questionnaire to place user in appropriate cluster. The data is then mapped using fuzzy algorithms such as suggested by Son and Thong (2015). The second group uses the global preferences of the new users to combine

them in the clusters (Liu et al., 2014). The third group combines different methods for calculation of the ratings. Leung, Chan, and Chung (2008) infuse fuzzy sets in to association rule mining to overcome the cold start problem.

The algorithms discussed to solve the cold start problem have their drawbacks. They require demographic data or ratings to solve the cold start problem and if that data is not given then algorithm behavior is incorrect. Secondly, they all rely on pearson distance which has its own limitations in terms of large noisy data in order to transform users into clusters (Toledo, Mota, and Martínez, 2015).

From the analysis of the cold start problem, the idea behind this study was to consider the effect of cold start problem in the proposed algorithms. Variables in the algorithms are introduced which are not directly involved with user profile and have enough significance towards the final prediction that the new user problem can be minimized. The variables are as follows

a) Average ratings of the items
b) Average ratings of the global user for this item
c) Average rating of the cluster user is assigned after initial few ratings.

Cold start problem is not the only problem that occurs because of profiling. Some recommender techniques rely more on the user profile than the others. Although cold start has been there constantly, other scenarios also can occur, e.g. a user went to a western restaurant and enjoyed the food and thus, rated the restaurant very high. The RS implies that if user likes western restaurant, he likes western food. It is somehow true, but the problem is that currently there is one item in the user profile so system would always recommend western food from which he eventually gets tired of. This problem is discussed in literature as stability versus plasticity (Wu, Chang, and Liu, 2014).

To avoid these problems, some techniques have divided users into groups. The flow goes like this. If two users namely A and B have gone to see a movie and liked it, then they are placed in the same group. A recommendation is given to B if A likes some other movie and vice versa. The stability versus plasticity problem is thought to have been completely overcome with the use of groups. However, if nobody watches the movie of specific group in the cluster, the movie would still not be suggested to the cluster users.

The proposed algorithm overcomes this deficiency by reducing the effect of cluster over the proposed ratings.

The groups give rise to dependency factor, and the division of the groups which is known as clusters but that also has few drawbacks. Those groups are made on similar items rated with almost similar ratings. Consider two groups who rated an item. There are three members in each group. Group A members rated some item 1, 1 and 5 while group B members rated same item 6, 10, 10. Here the users who rated that item 5 and 6 are on the border. Although they come in these groups, but their opinion in their respective group is not as consistent. So they might not get good recommendation and might be shifted into different groups because they always exist on the margin. These users are called gray-sheep (Ghazanfar and Prügel-Bennett, 2014).

The gray sheep issue has a direct impact on the accuracy of the RS (Ghazanfar and Prügel-Bennett, 2014). Ghazanfar and Prügel-Bennett (2014) tried to solve the gray sheep problem with switching hybrid RS (Burke, 2002b). The existing approaches try to divide each user in the clusters and then try the error detection on the clusters to solve the gray sheep problem. There is a deficiency in this solution if the user is divided while he is new or has rated fewer items to be placed in appropriate group or his new ratings are different than the training rating. In such cases, the users get stuck into the wrong group.

The algorithm proposed in this research tries to solve the gray-sheep problem by following methods.

a) Every time a user is inserted into the cluster the centroid of the cluster changes accordingly and the average distance of each user is changed to centroid.
b) After each rating entered by the user, its distance with the centroids is recalculated and is adjusted to the other clusters if necessary.
c) The cluster rating given by the CF is not taken as final rating as they are being adjusted by other variables since it only has the 20% weightage in the final ratings predicted by the proposed algorithm.

Apart from the factors described above, there is another very strong factor in the recommendation process. This factor is user mood (Winoto and Tang, 2010). To analyze this factor, consider an example. How often this happens that sometimes people are in a mood of slow music while desiring to listen to rock or jazz on other times? The 'mood change' factor is inside every human being.

So if RS always depends on other users' recommendation and even if they say that they have the same taste, RS cannot predict when the mood will swing.

The literature shows the impact of the user mood is high in the final outcome of the rating it provides for a certain item (Shan et al., 2009). The effect of mood changes can be more complicated in terms of RS. The proposed algorithm only tries to handle certain aspect of the mood changes in RS instead of going deep in the topic. The mood changes are handled in certain aspects by using user personal rankings in the final predicted ratings.

1.2. Problem Statement

The techniques used to solve the issues in RS including difference of opinion, mood changes, gray sheep, cold start and stability versus plasticity (Toledo, Mota, and Martínez, 2015) are CB, CF, MF and Hybrid. CB (Pera and Ng, 2013) is the very old technique which was used in late 80's and early 90's as a standalone system in RS. It uses the user's previous ratings to give the prediction for the items. CB mainly depends on the previous ratings given by the users to predict the next items. It has few prominent drawbacks such as cold-start problem (Vizine, Luiz and Hruschka, 2015) and stability versus plasticity.

CF is most commonly used technique in RS which implements the word of mouth (Shardanand and Maes, 1995). CF is more reliable as it divides the users in groups. CF solves the problem of stability versus plasticity, which is common in CB. However by solving those problems it gives rise to problems like gray sheep (Ghazanfar and Prügel-Bennett, 2014).

MF is another RS technique which follows neighborhood models as CF (Luo, Yunni, and Qingsheng, 2012). MF is more compact than CF but the problem in MF is that it rates the neighbors more while giving the ratings for the system which resultantly ignores the mood change factor in final ratings.

Some researchers have combined different techniques in hybrid. The combination of CB and CF is powerful in itself (Kant and Bharadwaj, 2012). Most of the literature combines these methods in hybrid to overcome the problems like stability versus plasticity and cold start problems. The combination still holds its drawbacks like gray sheep problems.

A relatively less used technique in RS is normalization (Bell and Koren, 2007b). Normalization is a data correction technique but typically requires additional information for the achievement of the results. The tendency of normalization to reduce the huge data into specific range helps to control the gray-sheep problem because normalization rescales the values (Bell and Koren, 2007a). Literature mostly focuses on the sophisticated algorithms but the effects of normalization are as significant as any algorithm (Bell et al., 2007).

Existing works that combine normalization with neighborhood models (CF and MF) performed the normalization of data before the process of neighborhood models. The process used is very general and is not targeting any problem in the RS. These methods use the normalization to avoid the big difference in the final rating of RS. These methods also overlook the effect of the ratings of the clusters in which user is. While these techniques improve the predictions to some extent but still rely mainly on the neighborhood models (CF and MF) and in turn also bring their own short comings as discussed above.

The normalization which is well known in standard data mining tasks on averaging user's ratings can be used more than just to avoid big difference in the final rating of RS. Therefore, this research addresses the effect of employing normalization technique in combination with CF and MF to solve the existing problems in RS. While the average ratings of items and users are important in normalization, the research also considers the cluster averaging and item group averages in order to further normalize the ratings to target cold start problem, gray sheep problem, stability versus plasticity problem and mood changes.

1.3. Hypothesis

The proposed algorithms in this research combine the normalization with neighborhood models (CF and MF). The data scaling simplification by the normalization technique and its tendency to have lesser errors combined with the better accuracy of neighborhood models (CF and MF) can produce more accurate results by removing the errors in the RS.

1.4. Objectives

The objectives of this research are

a) To develop normalizations of MF algorithm for improving the RS.

b) To develop normalizations of CF algorithms for improving the RS.

1.5. Scope and Relevance

The thesis has proposed solutions to the problems in RS for gray sheep and cold start besides solving the mood changes factor. These problems have caused past research solutions to reach low average and mean average error (MAE). The devised method is based on the hybrid of normalization and neighborhood models. The performance has been tested on a dataset sourced from the MovieLens which is validated based on the reduced MAE of the predicted preferred items. The scope of this thesis is limited in the problems of RS discussed in Section 1.1.

1.6. Contribution

The problems discussed in RS in Section 1.1 are affecting the accuracy of the RS. Two novel algorithms which utilize the distance between user average ratings and the global average ratings, combined with neighborhood model's cluster's average ratings and the difference between global average ratings are devised. The main focus of these algorithms is to solve the problems explained in Section 1.1 in neighborhood models with the help of normalization. Based on the results of the CF (neighborhood models, and MF), a prediction component is derived which helps to find the final rating for the user. The result shows that MF+N and CF+N improve the accuracy of the RS.

1.7. Organization of Thesis

The rest of the thesis is organized as follows.

Chapter II describes the history of RS and various techniques used in the recommendation process. The comparison across the techniques are also presented and concluded by highlighting the limitations in the techniques. These explanations led to the justification of the new approach based on normalization application on CF and MF.

Chapter III describes the methodology used during this research. The design of the algorithms is given in this chapter including the combination process.

Chapter IV describes the proposed RS algorithms called CF+N and MF+N. The framework and the flowchart of both algorithms are described.

Chapter V discusses the results of proposed algorithms as compared to other methods described in the literature to solve the problems in RS.

Chapter VI discusses the contributions of the research towards RS and the suggestion of future works.

1.8. Summary

This chapter includes the introduction of RS and common problems which are affecting the accuracy of the RS. The discussion of proposed algorithms is given along with the ways these algorithms can help improve the accuracy of the RS by solving these problems. Different techniques used in the RS and their related problems are specified followed by how normalization can affect the neighborhood models in the improvement of the results. The main objectives of the research are described as developing normalizations of MF and CF. The scope of the research is limited to solving problems of gray sheep, cold start, and stability versus plasticity, mood changes and difference of opinion. Towards the end of the chapter the contributions have been highlighted as two algorithms to solve existing problems in RS. Chapter ends by describing the organization of thesis.

Table 1: RS Challenges and Existing Solutions

		Techniques				
		CF	**MF**	**CB**	**Hybrid**	**Normalization**
P R O B L E M S	Cold start	✓		✓	✓	
	Gray sheep	✓				
	Plasticity versus Stability			✓		
	Difference of Opinion	✓	✓	✓	✓	
	Mood Changes	✓	✓	✓	✓	✓

REFERENCES

Al-Shamri, M. Y. H., & Bharadwaj, K. K. (2008). Fuzzy-Genetic Approach to Recommender Systems Based on a Novel Hybrid User Model. *Expert Systems with Applications,* 35(3): 1386–99.

Anand, D., & Bharadwaj, K. K. (2011). Utilizing Various Sparsity Measures for Enhancing Accuracy of Collaborative Recommender Systems Based on Local and Global Similarities. *Expert Systems with Applications*, 38(5): 5101–9.

Bell, R. M, Koren, Y., Ave, P., & Park, F. (2007). Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights. ICDM 2007: *In Seventh IEEE International Conference On Data Mining*. IEEE Computer Society.

Bell, R. M., & Koren, Y. (2007a). Improved Neighborhood-Based Collaborative Filtering. *KDDCup and Workshop*, 5(2): 7–14.

Bell, R. M & Koren, Y. (2007b). Lessons from the Netflix Prize Challenge. *ACM SIGKDD Explorations Newsletter*, 9(2): 75.

Bilge, A., & Polat, H. (2013). A Comparison of Clustering-Based Privacy-Preserving Collaborative Filtering Schemes. *Applied Soft Computing Journal,* 13(5): 2478–89.

Bobadilla, J., Ortega, F., Hernando, A. & Gutiérrez, A. (2013). Recommender Systems Survey. *Knowledge-Based Systems,* 46: 109–32.

Bogdanova, G., & Georgieva, T. (2008). Using Error-Correcting Dependencies for Collaborative Filtering. *Data and Knowledge Engineering,* 66(3): 402–13.

Burke, R. (2002a). Hybrid Recommender Systems: Survey and Experiments. *User modeling and user-adapted interaction,* 12(4): 331–70.

Burke, R. (2002b). Interactive Critiquing for Catalog Navigation in E-Commerce. *Artificial Intelligence Review*, 18(3): 245–67.

Chandak, M., Girase, S., & Mukhopadhyay, D. (2015). Introducing Hybrid Technique for Optimization of Book Recommender System. *Procedia Computer Science*, 45: 23–31.

Feng, H., Tian, J., Wang, H. J., & Li, M. (2015). Personalized Recommendations Based on Time-Weighted Overlapping Community Detection. *Information & Management*. Doi: 10.1016/j.im.2015.02.004

Frasincar, F., IJntema, W., Goossen, F., & Hogenboom, F. (2011). A semantic approach for news recommendation. Business Intelligence Applications and the Web: *Models, Systems and Technologies.* IGI Global.

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(7): 3261–72.

Goldberg, D., Nichols, Oki & Terry (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35(12): 61–70.

Hammond, K. (1989). *Case-Based Planning: Viewing Planning as a Memory Task.* Boston,MA: Academic Press.

Holmstrom, J. E. (1948). Classification of Classifications. *Royal Society Scientific Information Conference*, paper 33. London: Royal Society.

Hu, R., & Pu., P. (2010). Using Personality Information in Collaborative Filtering for New Users. *Recommender Systems and the Social Web*, 17.

Jin, R., & Si, L. (2004). A study of methods for normalizing user ratings in collaborative filtering. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 568-569). ACM.

Jin, R., Chai, J. Y., & Si, L. (2004). An automatic weighting scheme for collaborative filtering. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 337-344). ACM.

Karen S. J. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of documentation,* 28(1): 11–21.

Kaleli, C. (2014). An Entropy-Based Neighbor Selection Approach for Collaborative Filtering. *Knowledge-Based Systems,* 56: 273–80.

Kant, V., & Bharadwaj, K. K. (2012). Enhancing Recommendation Quality of Content-Based Filtering through Collaborative Predictions and Fuzzy Similarity Measures. *Procedia Engineering,* 38(Icmoc): 939–44.

Kolodner, J. (1993). *Case-Based Reasoning*. San Mateo, CA: Morgan Kaufmann.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques For Recommender Systems. *Computer,* 42(8): 30–37.

Kurata, Y. (2010). CT-Planner2 : More Flexible and Interactive Assistance for Day Tour Planning. *Enter,* 2011(2011): 25–37.

Lee, S. K., Cho, Y. H., & Kim, S. H. (2010). Collaborative Filtering with Ordinal Scale-Based Implicit Ratings for Mobile Music Recommendations. *Information Sciences,* 180(11): 2142–55.

Leung, C. W., Chan, S. C., & Chung, F. (2008). An Empirical Study of a Cross-Level Association Rule Mining Approach to Cold-Start Recommendations. *Knowledge-Based Systems,* 21(7): 515–29.

Liu, H., Hu, Z., Mian, A., Tian, H. & Zhu, X. (2014). A New User Similarity Model to Improve the Accuracy of Collaborative Filtering. *Knowledge-Based Systems*, 56: 156–66.

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender System Application Developments: A Survey. *Decision Support Systems*. Doi: 10.1016/j.dss.2015.03.008

Luo, X., Yunni, X., & Qingsheng, Z. (2012). Incremental Collaborative Filtering Recommender Based on Regularized Matrix Factorization. *Knowledge-Based Systems,* 27: 271–80.

Luo, X., Yunni, X., & Qingsheng Z. (2013). Applying the Learning Rate Adaptation to the Matrix Factorization Based Collaborative Filtering. *Knowledge-Based Systems*, 37: 154–64.

Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Advances in Neural Information Processing Systems*, 7(7): 281–97.

Maes, P. (1994). Agents That Reduce Work and Information Overload. *Communications of the ACM*, 37(7): 30–40.

Montaner, M., López, B., & Rosa., J.L. D. L. (2003). A Taxonomy of Recommender Agents on the Internet. *Artificial intelligence review*, 19(4): 285–330.

Moreno, A., Valls, A., Isern, D., Marin, L., & Borràs, J. (2013). Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. *Engineering Applications of Artificial Intelligence*, 26(1): 633-651.

Nie, D., An, Y., Dong, Q., Fu, Y. & Zhu, T. (2015). Information Filtering via Balanced Diffusion on Bipartite Networks. *Physica A: Statistical Mechanics and its Application*, s(421): 44–53.

Park, Y., Park, S., Jung, W., & Lee, S. (2015). Reversed CF: A Fast Collaborative Filtering Algorithm Using a K-Nearest Neighbor Graph. *Expert Systems with Applications*, 42(8): 4022–28.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. Lecture Notes in Computer Science: *The adaptive web*: (pp. 325-341). Springer Berlin Heidelberg.

Pera, M. S., & Ng, Y. K. (2013). A Group Recommender for Movies Based on Content Similarity and Popularity. *Information Processing and Management*, 49(3): 673–87.

Pinho Lucas, J., Segrera, S. & Moreno, M. N. (2012). Making Use of Associative Classifiers in Order to Alleviate Typical Drawbacks in Recommender Systems. *Expert Systems with Applications*, 39(1): 1273–83.

Pirasteh, P., Hwang, D., & Jung, J. J. (2015). Exploiting Matrix Factorization to Asymmetric User Similarities in Recommendation Systems. *Knowledge-Based Systems*. Doi: 10.1016/j.knosys.2015.03.00

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994, October). GroupLens: an open architecture for collaborative filtering of netnews. CSCW '94: *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 175-186). ACM.

Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3): 56-58.

Rich, E. (1979). User Modeling via Stereotypes. *Cognitive science*, 3(4): 329–54.

Rocchio, J. J. Jr. (1966). Document retrieval systems-optimization and evaluation. Doctoral Dissertation, Harvard University. *Report ISR-10, to the National Science Founa'ation, Harvard Computational Laboratory*, Cambridge, MA.

Salton. G. (1968). *Automatic Information Organization and Retrieval*. McGraw Hill Text, 1968.

Sanderson, M., & Croft, W. B. (2012). The history of information retrieval research. 100 Special Centennial Issue: *Proceedings of the IEEE* (pp. 1444-1451). IEEE.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. WWW '01: *Proceedings of the 10th international conference on World Wide Web* (pp. 285-295). ACM.

Schafer, J. B., Konstan, J., & Riedl, J. (1999, November). Recommender systems in e-commerce. EC'99: *Proceedings of the 1st ACM conference on Electronic commerce* (pp. 158-166). ACM.

Schiaffino, S., & Amandi, A. (2009). Building an Expert Travel Agent as a Software Agent. *Expert Systems with Applications*, 36(2): 1291–99.

Shan, M. K., Kuo, F. F., Chiang, M. F. & Lee, S. Y. (2009). Emotion-Based Music Recommendation by Affinity Discovery from Film Music. *Expert Systems with Applications*, 36(4): 7666–74.

Shardanand, U., & Maes, P. (1995, May). Social information filtering: algorithms for automating "word of mouth". CHI '95: *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 210-217). ACM Press/Addison-Wesley Publishing Co.

Son, L. H. (2015). HU-FCF++: A Novel Hybrid Method for the New User Cold-Start Problem in Recommender Systems. *Engineering Applications of Artificial Intelligence*, 41: 207–22.

Son, L. H., & Thong, N. T. (2015). Intuitionistic Fuzzy Recommender Systems: An Effective Tool for Medical Diagnosis. *Knowledge-Based Systems*, 74: 133–50.

Switzer, P. (1965). Vector images in document retrieval. *Statistical association methods for mechanized documentation*, 163-171.

Toledo, R. Y., Mota, Y. C., & Martínez, L. (2015). Correcting Noisy Ratings in Collaborative Recommender Systems. *Knowledge-Based Systems*, 76: 96–108.

Vizine P., Luiz, A., & Hruschka, E. R. (2015). 'Simultaneous Co-Clustering and Learning to Address the Cold Start Problem in Recommender Systems.' *Knowledge-Based Systems*. DOI: 10.1016/j.knosys.2015.02.016

Vozalis, M., & Margaritis, K. (2007). Using SVD and Demographic Data for the Enhancement of Generalized Collaborative Filtering. *Information Sciences*, 177(15): 3017–37.

Winoto, P., & Tang, T. Y. (2010). The Role of User Mood in Movie Recommendations. *Expert Systems with Applications*, 37(8): 6086–92.

Wu, M. L., Chang, C. H. & Liu, R. Z. (2014). Integrating Content-Based Filtering with Collaborative Filtering Using Co-Clustering with Augmented Matrices. *Expert Systems with Applications*, 41(6): 2754–61.

Xia, H., Fang, B., Gao, M., Ma, H., Tang, Y., & Wen, J. (2015). A Novel Item Anomaly Detection Approach against Shilling Attacks in Collaborative Recommendation Systems Using the Dynamic Time Interval Segmentation Technique. *Information Sciences*, 306: 150–65.

Xie, F., Chen, Z., Shang, J., Feng, X. & Li, J. (2015). A Link Prediction Approach for Item Recommendation with Complex Number. *Knowledge-Based Systems*, 81: 148–58.

Zhou, X., He, J., Huang, G., & Zhang, Y. (2015). SVD-Based Incremental Approaches for Recommender Systems. *Journal of Computer and System Sciences*, 81(4): 717–33.