A HYBRID APPROACH FOR PERSONALIZED NEWS RECOMMENDATION WITH ORDERED CLUSTERING ALGORITHM, RICH USER AND NEWS METADATA

ASGHAR DARVISHY

FSKTM 2019 60

A HYBRID APPROACH FOR PERSONALIZED NEWS RECOMMENDATION WITH ORDERED CLUSTERING ALGORITHM, RICH USER AND NEWS METADATA

By

ASGHAR DARVISHY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

September 2019

DEDICATION

This thesis is especially dedicated to my lovely family.

A HYBRID APPROACH FOR PERSONALIZED NEWS RECOMMENDATION WITH ORDERED CLUSTERING ALGORITHM, RICH USER AND NEWS METADATA

By

ASGHAR DARVISHY

September 2019

Chairman : **Professor Hamidah Ibrahim, PhD**
Faculty : **Computer Science and Information Technology**

One of the most commonly used of online services is news reading. A key challenge is selecting news articles from millions of sources considering user behavior and actual nature of news articles to recommend accurately. A personalized news recommendation system provides a news set that is extracted from multiple news press releases in order to handle such a high number of news articles and releases these news items to the user. A news item has a specific nature and it is different from the other items to recommend. In recent years, there has been much focus on the design and development of personalized news recommendation systems that observe and learn user behavior and generate news set based on this behavior. Commonly, the current news recommendation systems employ the collaborative filtering-based (CF-based), Content-based filtering (Content-based) or hybrid methods.

Scalability is one of the issues in news recommendation that requires effective algorithms to deal with large news corpus. One of the common strategies used for solving scalability is clustering. However, the existing clustering algorithms do not take into consideration the news nature in clustering the news items. The early recommender systems use popularity or recency, or both as properties to demonstrate the interestingness of the news items which do not reflect the actual nature of the news items. On the other hand, the existing personalized news recommendation systems do not make an attempt to filter the number of news items to recommend based on the reading rate behavior of a user. Moreover, early researches only consider explicit profile, short-term profile, and long-term profile of a user but none of them use all of the above user profiles in a single solution. News selection is another issue that requires new solution to effectively select news items to recommend.

In this research work, we have proposed a personalized news recommendation framework named Hybrid Personalized NEws Recommendation (HYPNER) which aims to recommend a personalized news set to the users. HYPNER combines both the collaborative filtering-based and the Content-based filtering methods in its framework. To address the above issues, the following have been proposed and incorporated into HYPNER. We have proposed a clustering algorithm named Ordered Clustering with a specific characteristic that allows multiple membership in clusters which reflects the news nature and user behavior. Furthermore, new models for user profile and news metadata construction are proposed where new properties have been incorporated, namely: *ReadingRate*, *HotnessRate*, and *Hotness*. A new model in news selection is proposed based on sub-modularity model. Our proposed framework has been validated through extensive experiments on real dataset. Results exhibit that HYPNER achieved 81.56% improvement in $F1$-score and 5.33% in diversity compared to the existing work, SCENE.

Abstrak tesis dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENDEKATAN HIBRID UNTUK PENGESYOREN BERITA PERIBADI DENGAN ALGORITMA PENGGUGUSAN TERTIB, PROFIL PENGGUNA DAN METADATA BERITA YANG KAYA

Oleh

ASGHAR DARVISHY

September 2019

Pengerusi : **Hamidah Ibrahim, PhD**
Fakulti : **Sains Komputer dan Teknologi Maklumat**

Salah satu perkhidmatan taliany ang paling biasa digunakan adalah membaca berita. Cabaran utama adalah untuk memilih artikel berita daripada berjuta-juta sumber berdasarkan tingkah laku pengguna dan sifat sebenar artikel berita untuk mengesyorkan secara tepat. Sistem cadangan berita peribadi menyediakan satu set berita yang diekstrak daripada pelbagai siaran akhbar berita untuk mengendali sejumlah artikel berita yang begitu banyak dan menyampaikan item-item berita ini kepada pengguna. Satu item berita mempunyai ciri yang khusus dan ianya berbeza dari item lain untuk disyorkan. Kebelakangan ini, terdapat banyak tumpuan terhadap reka bentuk dan pembangunan sistem cadangan berita peribadi yang memerhati dan mempelajari tingkah laku pengguna dan menghasilkan set berita berdasarkan tingkah laku ini. Pada kebiasaannya, sistem cadangan berita semasa menggunakan kaedah berasaskan penapisan usahasama (berasaskan CF), penapisan berasaskan kandungan (berasaskan kandungan), atau kaedah hibrid.

Skalabiliti adalah salah satu isu dalam cadangan berita yang memerlukan algoritma yang berkesan untuk menangani korpus berita besar. Salah satu strategi biasa yang digunakan untuk menyelesaikan skalabiliti adalah pengelompokan. Walau bagaimanapun, algoritma pengelompokan sedia ada tidak mengambil kira sifat semula jadi berita dalam pengelompokan item-item berita. Sistem cadangan awal menggunakan populariti atau kebaruan, atau kedua-duanya sebagai sifat untuk menunjukkan kemenarikan item-item berita yang mana ia tidak mencerminkan sifat semula jadi sebenar sesuatu berita. Sebaliknya, sistem cadangan berita peribadi yang sedia ada tidak membuat percubaan untuk menapis jumlah item-item berita untuk dicadangkan berdasarkan tingkah laku kadar pembacaan seseorang pengguna. Lebih-lebih lagi, penyelidikan awal hanya mempertimbangkan profil yang jelas, profil jangka pendek, dan profil jangka panjang seseorang pengguna tetapi tiada satu pun daripadanya menggunakan kesemua profil pengguna di dalam satu penyelesaian

iii

tunggal. Pilihan berita adalah isu lain yang memerlukan penyelesaian baharu untuk memilih item berita secara berkesan untuk disyorkan.

Dalam kajian ini, satu rangka kerja cadangan berita peribadi bernama *Hybrid Personalized NEws Recommendation* (HYPNER) dicadangkan yang bertujuan untuk mengesyorkan set berita peribadi kepada pengguna. HYPNER mengabungkan kedua-dua kaedah berasaskan penapisan usahasama dan kaedah penapison berasaskan kandungan di dalam rangka kerjanya. Bagi menanggani isu di atas, yang berikut telah dicadangkan dan diimplementasikan di dalam HYPNER. Kami telah mencadangkan satu algoritma pengelompokan dinamakan *Ordered Clustering* dengan ciri khusus yang membolehkan keahlian pelbagai di dalam kelompok yang mencerminkan sifat semula jadi berita dan tingkah laku pengguna. Tambahan pula, model baharu untuk profil pengguna dan pembinaan metadata berita juga dicadangkan di mana sifat-sifat baharu telah diimplementasikan iaitu: *ReadingRate*, *HotnessRate*, dan *Hotness*. Model baharu dalam pemilihan berita dicadangkan berdasarkan model sub-kemodularan. Rangka kerja yang kami cadangkan telah disahkan melalui eksperimen yang menyeluruh ke atas set data sebenar. Keputusan menunjukkan bahawa HYPNER mencapai peningkatan 81.56% dalam skor F1 dan 5.33% dalam kepelbagaian berbanding dengan kerja sedia ada, SCENE.

ACKNOWLEDGEMENTS

APPROVAL

I certify that an Examination Committee has met on Septamber 2018 to conduct the final examination of **Asghar Darvishy** on his PhD thesis entitled "**a hybrid approach personalized news recommendation with ordered clustering algorithm, rich user profile and news meta data**" in accordance with 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 degree of Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Norwati Mustapha, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Chairman)

Azreen Azman, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Internal Examiner)

Lilly Suriani Affendey, Ph.D.
Associate Professor
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Internal Examiner)

Juliana Sutanto, Ph.D.
Professor
Department of Management Science
Lancaster University

(External Examiner)

ROBIAH BINTI YUNUS, PhD.
Associate Professor and 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 Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Hamidah Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Fatimah Sidi, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Aida Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Tun Hussain Onn
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

vii

Declaration by graduate student

I hereby confirm that:
- this thesis is my original work;
- quotation, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other instructions;
- 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, journal, modules, proceeding, popular writing, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in Uneversiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to 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: Asghar Darvishy (GS33367)

viii

Declaration by Member of Supervisory Committee

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

Signature:
Name of Chairman
of Supervisory
Committee: Professor Dr. Hamidah Ibrahim

Signature:
Name of Member
of Supervisory
Committee: Associate Professor Dr. Fatimah Sidi

Signature:
Name of Member
of Supervisory
Committee: Associate Professor Dr. Aida Mustapha

ix

TABLE OF CONTENTS

LIST OF TABLES

xiii

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CB	Content Based
CBR	Case-based Reasoning
CF	Collaborative Filtering
CM	Cluster Matrix
EM	Expectation Maximization
HYPNER	HYbrid Personalized News Recommendation
IF	Information Filtering
IR	Information Retrieval
KB	Knowledge Based
KNGR	Knowledge-based Geographical News Recommender
LDA	Latent Dirichlet Allocation
LSH	Latent Semantic Hashing
MFR	Maximal Forward Reference
NR	News Rating
OC	Ordered Clustering
PLSI	Probabilistic Latent Semantic Indexing
RSS	Rich Site Summary
SM	Similarity Matrix
TF-IDF	Term Frequency-Inverse Document Frequency
TMS	Text Metadata Services

CHAPTER 1

INTRODUCTION

1.1   Overview

One of the most powerful and general social media platforms is online news release that makes easy news accessing, sharing, and commenting. An efficient media for publishing news articles on the World Wide Web requires that it does not have traditional print-based publishing limitations. The number of news articles published per hour grows exponentially; so, multiple news sources around the world and the variety of news categories make it difficult for the user to find preferred news to read. Users prefer a filtered view of relevant news articles, allowing them to focus on news items that contain rich contextual information to their behaviors and interests. News agencies have Rich Site Summary (RSS) news feed. They prepare and recommend a fixed number of news items to users without considering the users preferences and interests in news reading.

A personalized news recommendation system provides a news set that is extracted from multiple news press releases in order to handle such a high number of news articles and release this news set to the user. A typical personalized news recommendation system is made up of the following components: (i) user profile construction which constructs the user profile based on the user historical reading behavior, (ii) news metadata construction which summarizes the news items into terms and their weights, (iii) news and user clustering which groups similar users as well as similar news items into clusters, (iv) news selection which makes selection of news items from a collection of news items based on their similarities, and (v) news representation which presents news items to the user.

A news item has a specific nature and it is different from the other items to recommend (Li et al., 2011b). It is evident that a news item may belong to more than one news category. Apart from that, a news item has a short lifetime and it may expire in a few duration of time. Recency is a commonly used property to determine a news lifetime; it determines how long news item is newly-published. Popularity shows how many times a news item is read by the users throughout its lifetime. It is possible, a hot news item is read millions of times in few minutes while an uninteresting news item is read less than hundred times throughout its lifetime. On the other hand, the popularity and recency of news articles change dramatically over time, which differentiates news items from other items, such as products, books, and movies, rendering the traditional recommendation methods ineffective (Li et al., 2011b).

In recent years, there has been much focus on the design and development of personalized news recommendation systems that monitor and learn users' reading behaviors and generate news set based on these behaviors. The current news recommendation systems employ the collaborative filtering-based (CF-based),

Content-based filtering (Content-based) or hybrid methods. The CF-based news recommendation systems generate personalized recommendation for users based on their behaviors in news reading. In this method clustering similar users in a group is performed based on their news access pattern similarities. Such behaviors are expressed in a form of binary votes or numerical ratings on each news item. The CF algorithms have difficulty in generating reliable recommendation when data are sparse, and they cannot recommend news items that have no rating from the users. They suffer from the problems of scalability, data sparseness, and cold-start recommendation (Leung, 2009). On the other hand, the Content-based news recommendation system clusters the news items based on the content similarities and recommends news items from the selected clusters based on the content similarities between the news items and user profile. It considers a given user's reading behavior and analyzes the content of the newly-published news before presenting it to the users. The recommender system computes similarity between newly-published news items and the user Content-based profile and rates them. The news items with high rates are recommended to the users. The Content-based methods cannot recommend accurately to the new user with low access in news reading (Lu et al., 2015). Meanwhile, the hybrid recommendation systems combine two or more recommendation techniques to gain better performance with fewer drawbacks of any individual one. Most commonly, the CF-based is combined with other techniques in an attempt to avoid the CF-based problems (Burke, 2002). To get more reasonable results, the feasibility of combining the CF-based and the Content-based is investigated.

1.2 Problem Statement

One of the popular and most frequently used recommendation systems are news recommendation systems (Zheng et al., 2012). Scalability is one of the issues in news recommendation that requires elegant algorithms to effectively deal with large news corpus (Li et al., 2011b). Several strategies can be used to address the scalability issue such as MinHash (Liu et al., 2010) and clustering algorithms. The most commonly used clustering algorithms in news recommendation systems are hierarchical clustering (Li et al., 2011b; Zheng et al., 2012) and *k*-means (Jiang and Hong, 2014; Li et al., 2010). Nevertheless, these clustering algorithms do not take into consideration the news nature in clustering the news items. Consequently, a news item will only belong to a single cluster while in reality a news item might have high similarities to other news items belonging to different clusters and within different categories. Moreover, it is obvious that users' interests are not limited to one news category. Hence, the clustering algorithm to be employed in user and news clustering should be able to cluster users and news items without limiting their membership to a single cluster. *Therefore, this work attempts to propose a new clustering algorithm that considers both the news nature and user behavior in news reading.*

One of the issues in news recommendation is to recommend news accurately and effectively by considering the news nature and its unique characteristics (Li et al., 2011b). Thus, constructing a proper model to capture the news nature as well as its characteristics and identifying an appropriate property that reflects the actual nature

2

of news items is inevitable. The early recommender systems use popularity (Das et al., 2007; Jiang and Hong, 2014; Li et al., 2010; Lin et al., 2014; Liu et al., 2010) or recency (Wang et al., 2015; Lak et al. 2016), or both (Li et al., 2011b; Zheng et al., 2012) as properties to demonstrate the interestingness of the news items to recommend to the users. Popularity which means how many times a news item is read by the users throughout its lifetime does not reflect the actual nature of a news item. This is because old news may have high popularity but might no longer be the interesting news items to read. On the other hand, newly-published news might have low popularity but it does not mean that it is not an interesting item to read. Hence, popularity alone cannot represent the sole characteristics of a news item. Meanwhile, recency which determines how long news item is newly-published also does not reflect the actual nature of a news item. A newly-published news item does not always indicate that it is more pleasing to read than those that were published earlier. Even combining both popularity and recency will not determine the interestingness of the news item. *Thus, identifying a proper model for new metadata and new properties that can represent the actual nature of news items are crucial.*

Several news recommendation frameworks (Jiang and Hong, 2014; Li et al., 2011b; Li et al., 2010; Lin et al., 2014; Liu et al., 2010; Zheng et al., 2012) have been proposed that attempt to increase accuracy, overcome the large volume of data, and recommend diverse of news items. A rich model in user profiling is required to represent the user's behavior, e.g. a property is required to determine the user preference in reading a recent news then a popular one, or contrariwise and this behavior should be considered in the recommendation system. On the other hand, even the personalized news recommendation systems such as Google News (Das et al., 2007; Liu et al., 2010), YahooNews (Li et al., 2010), SCENE (Li et al., 2011b) and PENETRATE (Zheng et al., 2012) do not make an attempt to filter the number of news items to recommend. These systems recommend the same number of news items to the users, i.e. they are unable to recommend the appropriate number of news items to each user based on the individual user behavior in news reading. Moreover, early researches consider explicit profile (Jiang and Hong, 2014; Liu et al., 2010; Jugovac & Jannach, 2017), short-term profile (Jiang and Hong, 2014; Li et al., 2010; Zheng et al., 2012), and long-term profile (Das et al., 2007; Li et al., 2011b; Li et al., 2011c) but none of them uses all the above user profiles in a single solution. *Thus, a rich user profile needs to be constructed to ensure the accuracy of news recommendation with respect to the number of news items as well as the news set.*

To the best of our knowledge, SCENE (Li et al., 2011b) is the study that addresses the issue of news selection. In this study, sub-modularity modeling is utilized in news selection and experiments show a news set that matches to the users' interests as much as possible while highest possible diversity is achieved. Rich news metadata and user profile are construed to affect news selection, and consequently news recommendation accuracy is affected (Li et al., 2011b). News selection requires new strategy in utilizing rich user profile and news metadata to assist news recommendation system in achieving accurate and diverse recommendation of news items. *Hence, identifying a strategy to effectively select news items for recommendation considering both the user profile and news metadata is vital.*

1.3   Objective of the Research

The main aim of this research work is to propose a personalized news recommendation framework that incorporates both the CF-based and Content-based methods. In order to achieve this aim, we present the following objectives with respect to the issues posed in Section 1.2:

1. To propose an effective clustering algorithm that considers both the news nature and user behavior in news reading. The proposed clustering algorithm should not limit the membership of a news item to a single cluster, also it should not add any additional complexity.
2. To propose a new model of news metadata construction including a new property to determine the interestingness of a news item.
3. To propose a new model in user profiling that captures rich user profile based on the user's reading behaviors.
4. To propose a model in news selection by considering both the user profile and news metadata to achieve accurate and diverse news recommendation.

1.4   Research Scope

This research work focuses on the proposing of a personalized news recommendation framework to improve news recommendation, where users are served with news recommendation based on their preference and behavior in news reading. In other words, news application or other issues in operational environment such as infrastructure, network, user interface, etc. are not considered in our research. As the focus of this research study is on personalized news recommendation where a news item has a specific nature and different from the other items to recommend, thus recommending other items like music, research articles, hotel, apartment, etc. are not considered in this research work.

1.5   Contributions

In this thesis, we have proposed a personalized news recommendation framework named HYbrid Personalized NEws Recommendation (HYPNER). HYPNER incorporates both the CF-based and Content-based methods in news recommendation. The following contributions are made to answer the issues that are addressed in this thesis.

i. A new clustering algorithm named *Ordered Clustering* is proposed and utilized in our proposed personalized news recommendation system. Ordered Clustering (OC) is a new unsupervised linear clustering algorithm. This clustering algorithm has a specific characteristic and it is designed based on the news nature and user behavior in news reading. This algorithm allows multiple memberships in clusters which means a news item can be clustered in more than one cluster and a given user may have diverse interests in news reading.

ii. A new model in user profiling is proposed, by designing a user profile in user behavior acquisition, including user's explicit profile, user's long-term (also known as CF-based) profile, and user's short-term (also known as Content-based) profile. To enrich the user profile, we have defined two properties, namely: *ReadingRate* and *HotnessRate*.

4

iii. A new model in news metadata construction is proposed. Entered news article is tokenized into name entities and relevance tags by utilizing OpenCalias (Reuters, 2008). To enrich the news metadata a new property is defined called *Hotness*.

iv. A new approach in clustering is proposed that clusters news metadata and short-term user profile based on their content similarities. The content similarity between the news metadata and the short-term user profile is computed using a Jaccard Similarity. A chain of similar news and users is created which leads to more similar news items in news selection.

v. A new model in news selection is proposed based on the sub-modularity model that attempts to achieve accuracy and diversity in news recommendation.

1.6 Organization of the Thesis

This thesis is organized as follows:

Chapter 1 is an introductory chapter that discusses the problem statement of this research study, the objectives, the research scopes, and the contributions of the research.

Chapter 2 presents the different types of news recommendation systems and also explains the user profiling and clustering algorithms that are related to this research. Related research works and review of news personalized recommendation systems are presented in this chapter. Also news recommendation issues are discussed.

Chapter 3 is the research methodology chapter. It describes how this research was conducted. Also, the chapter discusses the different phases in this research and the methodology followed during each phase. The proposed framework is also presented. The measurement metrics and the dataset that has been used in the experiments of this research are presented as well.

Chapter 4 presents the detail description of the proposed CF-based method of HYPNER. This chapter also explains the preliminaries, discusses the proposed Ordered Clustering algorithm, and finally an example is given to further clarify the CF-based method of HYPNER.

Chapter 5 presents the detailed description of the proposed Content-based method of HYPNER. Also, OpenCalais as a web service for news content analyzing is introduced. Short-term user profiling is explained. Also, Ordered Clustering on news metadata and short-term user profile are presented in this chapter.

Chapter 6 proposes the *Personalized News Recommendation* component of HYPNER and describes the proposed news selection and ranking strategies. Finally, an example is presented to clarify the processes involved in the component.

Chapter 7 presents the results of the experiments conducted to evaluate the performance of the proposed approaches in terms of accuracy and diversity. The results are compared to the relevant existing system.

Chapter 8 reflects the conclusions and the contributions of this research. Besides, the recommendations for future works are presented in this chapter.

REFERENCES

Aamodt Agnar and Plaza Enric. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Journal of AI Communications, 7(1)*, 39-59.

Abel Fabian, Gao Qi, Houben Geert-Jan, and Tao Ke. (2011). Analyzing user modeling on twitter for personalized news recommendations. *Proceedings of the International Conference on User Modeling, Adaption, and Personalization*, 1-12.

Adomavicius Gediminas and Kwon YoungOk. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *Journal of IEEE Transactions on Knowledge and Data Engineering, 24(5)*, 896-911.

Adomavicius Gediminas and Tuzhilin Alexander. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Journal of IEEE Transactions on Knowledge and Data Engineering, 17(6)*, 734-749.

Baeza-Yates Ricardo and Ribeiro-Neto Berthier. (1999). *Modern information retrieval 463*, ACM press New York.

Bagher, R. C., Hassanpour, H., and Mashayekhi, H. (2017). User trends modeling for a content-based recommender system. *Journal of Expert Systems with Applications, 87*, 209-219.

Basu Chumki, Hirsh Haym, and Cohen William. (1998). Recommendation as classification: Using social and content-based information in recommendation. *Proceedings of the 10th. International Conference on Artificial Intelligence*, 714-720.

Beel, J., Gipp, B., Langer, S., and Breitinger, C. (2016). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries. 17(4)*, 305–338.

Belkin Nicholas J. and Croft W. Bruce. (1992). Information filtering and information retrieval: Two sides of the same coin? *Journal of Communications of the ACM, 35(12)*, 29-38.

Ben- Gal Irad. (2007). Bayesian networks. *Encyclopedia of Statistics in Quality and Reliability*.

Billsus Daniel and Pazzani Michael J. (1998). Learning Collaborative Information Filters. *Proceedings of the 15th. International Conference on Machine Learning*, 46-54.

Blei David M., Ng Andrew Y., and Jordan Michael I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Bobadilla Jesús, Ortega Fernando, Hernando Antonio, and Gutiérrez Abraham. (2013). Recommender systems survey. *Journal of Knowledge-Based Systems, 46*, 109-132.

Boukhechba Mehdi, Bouzouane Abdenour, Bouchard Bruno, Gouin-Vallerand Charles, and Giroux Sylvain. (2015). Online prediction of people's next point-of-interest: Concept drift support. *Proceedings of the 6th. International Workshop on Human Behavior Understanding, 9277*, 97-116.

Bradley Keith and Smyth Barry. (2001). Improving recommendation diversity. *Proceedings of the 12th. Irish Conference on Artificial Intelligence and Cognitive Science*, 85-94.

Brusilovsky Peter. (1996). Methods and techniques of adaptive hypermedia. *Journal of Personalization Research, User Modeling and User-adapted Interaction, 6(2-3)*, 87-129.

Brusilovsky Peter. (2000). Adaptive hypermedia: From intelligent tutoring systems to web-based education. *Proceedings of the International Conference on Intelligent Tutoring Systems.* 1-7.

Brynjolfsson Erik, Hu Yu, and Simester Duncan. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Journal of Management Science, 57(8)*, 1373-1386.

Burke Robin. (2000). Knowledge-based Recommender Systems. *Encyclopedia of - Library and Information Science, 69*, 175-186.

Burke Robin. (2002). Hybrid recommender systems: Survey and experiments. *Journal of Personalization Research - User Modeling and User-adapted Interaction.* 12(4), 331-370.

Chang AiDen, Liao Jhen-Fu, Chang Pei-Chann, Teng Chin-Hung, and Chen Meng-Hui. (2014). Application of artificial immune systems combines collaborative filtering in movie recommendation system. *Proceedings of the 2014 IEEE 18th. International Conference on Computer Supported Cooperative Work in Design (CSCWD),* 277-282.

Chaturvedi, A. K., Peleja, F., and Freire, A. (2017). Recommender system for news articles using supervised learning. *arXiv preprint arXiv:1707.00506, Master Thesis, Universitat Pompeu Fabra, Barcelona.*

Chu Cheng, Kim Sang Kyun, Lin Yi-An, Yu YuanYuan, Bradski Gary, Ng Andrew Y., and Olukotun Kunle. (2007). Map-reduce for machine learning on multicore. *Journal of Advances in Neural Information Processing Systems, 19*, 281-288.

Chu Wei and Park Seung-Taek. (2009). Personalized recommendation on dynamic content using predictive bilinear models. *Proceedings of the 18th. International Conference on World Wide Web*. 691-700.

Claypool Mark, Brown David, Le Phong, and Waseda Makoto. (2001). Inferring user interest. *Journal of IEEE Internet Computing, 5*(6), 32-39.

Claypool Mark, Gokhale Anuja, Miranda Tim, Murnikov Pavel, Netes Dmitry, and Sartin Matthew. (1999). Combining content-based and collaborative filters in an online newspaper. *Proceedings of the ACM SIGIR Workshop on Recommender Systems – Implementation and Evaluation.*

Das Abhinandan, Datar Mayur, Garg Ashutosh, and Rajaram Shyam. (2007). Google news personalization: Scalable online collaborative filtering. *Proceedings of the 16th. International Conference on World Wide Web*. 271-280.

de Gemmis Marco, Iaquinta Leo, Lops Pasquale, Musto Cataldo, Narducci Fedelucio, and Semeraro Giovanni. (2011). Learning preference models in recommender systems. *Preference Learning*, 387-407: Springer.

de Souza Pereira Moreira, G. (2018). CHAMELEON: A deep learning meta-architecture for news recommender systems. *Proceedings of the 12th. ACM Conference on Recommender Systems*. 578-583.

Dean Jeffrey and Ghemawat Sanjay. (2008). MapReduce: Simplified data processing on large clusters. *Proceedings of the 6th. Conference on Symposium on Operating Systems Design & Implementation, 51(1)*, 107-113.

Erik Shefsky and Roy Sutton. (2015). Stop-outs a new variable in online student retention. *Proceedings of the International Conference in Society for Information Technology & Teacher Education,* 480-487.

Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. (2011). Analyzing user modeling on twitter for personalized news recommendations. *Proceedings of the International Conference on User Modeling, Adaption and Personalization,* 1-12.

Fan Weiguo, Wallace Linda, Rich Stephanie, and Zhang Zhongju. (2006). Tapping into the power of text mining. *Journal of Communications of the ACM, 49(9),* 76-82.

Fesenmaier Daniel R., Wöber Karl W., and Werthner Hannes. (2006). Destination recommendation systems: Behavioral foundations and applications: *Cabi Publication*.

Fleder Daniel and Hosanagar Kartik. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Journal of Management Science, 55(5)*, 697-712.

Foltz Peter W. and Dumais Susan T. (1992). Personalized information delivery: An analysis of information filtering methods. *Journal of Communications of the ACM, 35(12)*, 51-60.

Garcin Florent, Dimitrakakis Christos, and Faltings Boi. (2013). Personalized news recommendation with context trees. *Proceedings of the 7th. ACM Conference on Recommender Systems*, 105-112.

Garrido Angel Luis, Buey María G., Ilarri Sergio, Fũrstner Igor, and Szedmina Livia. (2015). KGNR: A knowledge-based geographical news recommender. *Proceedings of IEEE 13th. International Symposium on Intelligent Systems and Informatics (SISY)*, 195-198.

Goldberg David, Nichols David, Oki Brian M., and Terry Douglas. (1992). Using collaborative filtering to weave an information tapestry. *Journal of Communications of the ACM - Special Issue on Information Filtering, 35(12)*, 61-70.

Gu, W., Dong, S., and Chen, M. (2016). Personalized news recommendation based on articles chain building. *Journal of Neural Computing and Applications, 27(5)*, 1263–1272.

Han Jiawei, Kamber Micheline, and Pei Jian. (2011). *Data mining: concepts and techniques*: Elsevier.

Herlocker Jonathan L., Konstan Joseph A., Terveen Loren G., and Riedl John T. (2004). Evaluating collaborative filtering recommender systems. *Journal of ACM Transactions on Information Systems (TOIS), 22(1)*, 5-53.

Hill William C., Hollan James D., Wroblewski Dave, and McCandless Tim. (1992). Edit wear and read wear. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3-9.

Hofmann Thomas. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50-57.

Holtshouse DK, Borghoff Uwe M., and Pareschi Remo. (2013). Information technology for knowledge management: *Springer Science & Business Media*.

Hopfgartner Frank. (2010). Personalised video retrieval: Application of implicit feedback and semantic user profiles. *ACM SIGIR Forum, 44(2)*, 84-85.

Huang Zan, Zeng Daniel, and Chen Hsinchun. (2007). A comparative study of recommendation algorithms in e-commerce applications. *Journal of IEEE Intelligent Systems, 22(5)*, 68-78.

Huy Nguyen Thac, Chuan Do The, and Nguyen Viet Anh (2016). Implicit feedback mechanism to manage user profile applied in vietnamese news recommender

system. *International Journal of Computer and Communication Engineering*, *5(4)*, 276.

Inés Peña de Carrillo Clara. (2004). Intelligent agents to improve adaptivity in a web-based learning environment: *Knowledge-based Virtual Education*.

Jiang Shan and Hong Wenxing. (2014). A vertical news recommendation system: CCNS—An example from Chinese campus news reading system. *Proceedings of the 9th. International Conference on Computer Science & Education (ICCSE),* 1105-1114.

Jugovac, M. and Jannach, D. (2017). Interacting with recommenders overview and research directions. *Journal of ACM Transactions on Interactive Intelligent Systems, 7(3)*, 10.

Kang U., Tsourakakis Charalampos E., and Faloutsos Christos. (2009). Pegasus: A peta-scale graph mining system implementation and observations. *Proceedings of the 9th. IEEE International Conference on Data Mining,* 229-238.

Karimi, M., Jannach, D., and Jugovac, M. (2018). News recommender systems – Survey and roads ahead. *Journal of Information Processing & Management, 54(6)*, 1203-1227.

Kelly Diane and Belkin Nicholas J. (2001). Reading time, scrolling and interaction: Exploring implicit sources of user preferences for relevance feedback. *Proceedings of the 24th. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* 408-409.

Khattar, D., Kumar, V., Varma, V., and Gupta, M. (2018). Weave&Rec: A word embedding based 3-D convolutional network for news recommendation. *Proceedings of the 27th. ACM International Conference on Information and Knowledge Management*, 1855-1858.

Kim Bongjun and Pardo Bryan. (2014). Adapting collaborative filtering to personalized audio production. *Proceedings of the Second AAAI Conference on Human Computation and Crowdsourcing.* 5.

Kim Jinmook, Oard Douglas W., and Romanik Kathleen. (2001). User modeling for information filtering based on implicit feedback. *Proceedings of the ISKO-France Conference*. 25.

Konstan Joseph A., Miller Bradley N., Maltz David, Herlocker Jonathan L., Gordon Lee R., and Riedl John. (1997). GroupLens: Applying collaborative filtering to Usenet news. *Journal of Communications of the ACM, 40(3)*, 77-87.

Lak, P., Babaoglu, C., Bener, A. B., and Pralat, P. (2016). News article position recommendation based on the analysis of article's content - time matters. *Proceedings of the 3rd. Workshop on New Trends in Content-Based Recommender Systems (CBRecSys'16)*, 11–14.

Lau Cher Han. (2014). Detecting news topics from microblogs using sequential pattern mining. *PhD. Thesis, Queensland University of Technology*.

Lee Hong Joo and Park Sung Joo. (2007). MONERS: A news recommender for the mobile web. *Journal of Expert Systems with Applications, 32(1)*, 143-150.

Leung Wing-ki Cane. (2009). Enriching user and item profiles for collaborative filtering: From concept hierarchies to user-generated reviews. *PhD. Thesis, The Hong Kong Polytechnic University*.

Li Lei, Wang Ding-Ding, Zhu Shun-Zhi, and Li Tao. (2011a). Personalized news recommendation: A review and an experimental investigation. *Journal of Computer Science and Technology, 26(5)*, 754-766.

Li Lei, Wang Ding-Ding, Li Tao, Knox Daniel, and Padmanabhan Balaji. (2011b). Scene: A scalable two-stage personalized news recommendation system. *Proceedings of the 34th. International ACM SIGIR Conference on Research and Development in Information Retrieval*, 125-134.

Li Lei, Zheng Li, and Li Tao. (2011c). Logo: A long-short user interest integration in personalized news recommendation. *Proceedings of the 5th. ACM Conference on Recommender Systems,* 317-320.

Li Lihong, Chu Wei, Langford John, and Schapire Robert E. (2010). A contextual-bandit approach to personalized news article recommendation. *Proceedings of the 19th. International Conference on World Wide Web,* 661-670.

Lin Chen, Xie Runquan, Guan Xinjun, Li Lei, and Li Tao. (2014). Premise: Personalized news recommendation via implicit social experts. *Proceedings of the 21st. ACM International Conference on Information and Knowledge Management*, 1607-1611.

Liu Jiahui, Dolan Peter, and Pedersen Elin Rønby. (2010). Personalized news recommendation based on click behavior. *Proceedings of the 2010 International Conference on Intelligent User Interfaces*, 31-40.

Lops Pasquale, De Gemmis Marco, and Semeraro Giovanni. (2011). Content-based recommender systems: State of the art and trends. *Recommender systems handbook,* 73-105.

Lorenzi Fabiana and Ricci Francesco. (2005). Case-based recommender systems: A unifying view, *Proceedings of the IJCAI Workshop on Intelligent Techniques for Web Personalization,* 89-113.

Lu Zhongqi, Dou Zhicheng, Lian Jianxun, Xie Xing, and Yang Qiang. (2015). Content-based collaborative filtering for news topic recommendation. *Proceedings of the 29th. AAAI Conference on Artificial Intelligence*, 217-223.

Marir Farhi. (2012). Case-based reasoning for an adaptive web user interface. *Proceedings of the International Conference on Computing, Networking and Digital Technologies (ICCNDT2012)*, 306-315.

Matthews Tara, Whittaker Steve, Badenes Hernan, and Smith Barton. (2014). Beyond end user content to collaborative knowledge mapping: Interrelations among community social tools. *Proceedings of the 17th. ACM Conference on Computer Supported Cooperative Work & Social Computing*, 900-910.

McSherry David. (2002). Diversity-conscious retrieval. *Proceedings of the 6th. European Conference on Advances in Case-Based Reasoning*, 219-233.

Melville Prem, Mooney Raymond J., and Nagarajan Ramadass. (2002). Content-boosted collaborative filtering for improved recommendations. *Proceedings of the 18th. National Conference on Artificial Intelligence*, 187-192.

Middleton Stuart E., Alani Harith, and De Roure David C. (2002). Exploiting synergy between ontologies and recommender systems. *arXiv preprint cs/0204012*.

Middleton Stuart E., Shadbolt Nigel R., and De Roure David C. (2004). Ontological user profiling in recommender systems. *Journal of ACM Transactions on Information Systems (TOIS), 22(1)*, 54-88.

Miller Bradley N., Albert Istvan, Lam Shyong K., Konstan Joseph A., and Riedl John. (2003). MovieLens unplugged: Experiences with an occasionally connected recommender system. *Proceedings of the 8th. International Conference on Intelligent User Interfaces*, 263-266.

Mobasher Bamshad, Jin Xin, and Zhou Yanzan. (2004). Semantically enhanced collaborative filtering on the web, *Proceedings of the 1st European Web Mining Forum*, 57-76.

Montaner Miquel, López Beatriz, and De La Rosa Josep Lluís. (2003). A taxonomy of recommender agents on the internet. *Journal of Artificial Intelligence Review, 19(4)*, 285-330.

Mooney Raymond J. and Roy Loriene. (2000). Content-based book recommending using learning for text categorization. *Proceedings of the 5th. ACM Conference on Digital Libraries*, 195-204.

Morita Masahiro and Shinoda Yoichi. (1994). Information filtering based on user behavior analysis and best match text retrieval. *Proceedings of the 17th. Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 272-281.

Nandi, RN., Zaman, MA., Al Muntasir, T., Sumit, SH., Sourov, T., and Rahman, M. J. U. (2018). Bangla news recommendation using doc2vec. *Proceedings of the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. 1-5.

Niwattanakul Suphakit, Singthongchai Jatsada, Naenudorn Ekkachai, and Wanapu Supachanun. (2013). Using of Jaccard coefficient for keywords similarity. *Proceedings of the International Multi Conference of Engineers and Computer Scientists, 1(6)*, 380-384.

Núñez-Valdez Edward Rolando, Lovelle Juan Manuel Cueva, Hernández Guillermo Infante, Fuente Aquilino Juan, and Labra-Gayo José Emilio. (2015). Creating recommendations on electronic books: A collaborative learning implicit approach. *Journal of Computers in Human Behavior, 51*, 1320-1330.

Ott R. Lyman and Longnecker Micheal T. (2015). An Introduction to Statistical Methods and Data Analysis: *Nelson Education*.

Papadimitriou Spiros and Sun Jimeng. (2008). Disco: Distributed co-clustering with map-reduce: A case study towards petabyte-scale end-to-end mining. *Proceedings of the 8th. IEEE International Conference on Data Mining*, 512-521.

Pazzani Michael J. (1999). A framework for collaborative, content-based and demographic filtering. *Journal of Artificial Intelligence Review, 13(5-6)*, 393-408.

Pirolli Peter, Pitkow James, and Rao Ramana. (1996). Silk from a sow's ear: Extracting usable structures from the Web. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 118-125.

Resnick Paul, Iacovou Neophytos, Suchak Mitesh, Bergstrom Peter, and Riedl John. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94)*, 175-186.

Reuters Thomson. (2008). OpenCalais. *Retrieved June, 16*.

Ricci Francesco. (2002). Travel recommender systems. *Journal of IEEE Intelligent Systems, 17*(6), 55-57.

Ricci Francesco and Del Missier Fabio. (2004). Supporting travel decision making through personalized recommendation: *Designing Personalized User Experiences in eCommerce*.

Sarwar Badrul M., Karypis George, Konstan Joseph, and Riedl John. (2000). Application of dimensionality reduction in recommender system - a case study, *Technical Report*, *DTIC Document*.

Schafer J. Ben, Frankowski Dan, Herlocker Jon, and Sen Shilad. (2007). Collaborative filtering recommender systems. *The Adaptive Web Notes in Computer Science, 4321,* 291-324.

Sevli, O. and Küçüksille, EU. (2017). Advertising recommendation system based on dynamic data analysis on Turkish speaking Twitter users. *Journal of Technical Gazette, 24(2)*, 571-578.

Smyth Barry and Cotter Paul. (2000). A personalized television listings service. *Journal of Communications of the ACM, 43(8)*, 107-111.

Su Xiaoyuan and Khoshgoftaar Taghi M. (2009). A survey of collaborative filtering techniques. *Journal of Advances in Artificial Intelligence, 2009*, 4.

Tran Thomas and Cohen Robin. (2000). Hybrid recommender systems for electronic commerce. *Proceedings of the Knowledge-Based Electronic Markets*.

Trewin Shari. (2000). Knowledge-based recommender systems. *Encyclopedia of Library and Information Science, 69*, 180.

Van Asch Vincent. (2013). Macro-and micro-averaged evaluation measures. *Belgium: University of Antwerp. <https://pdfs.semanticscholar.org/1d10/6a2730801b6210a67f7622e4d192bb 309303.pdf >* (09/09/2013).

Van Rijsbergen Cornelis Joost. (1974). Foundation of evaluation. *Journal of Documentation, 30(4)*, 365-373.

Wanas, N., Farouk, A., Said, D., Khodeir, N., and Fayek, M. (2019). Detection and handling of different types of concept drift in news recommendation systems. *International Journal of Computer Science & Information Technology (IJCSIT), 11(1)*, 87.

Wang Dingding, Zhu Shenghuo, and Li Tao. (2009). Comparative document summarization with discriminative sentence selection. *Proceedings of the 18th. ACM Conference on Information and Knowledge Management*, 1963-1966.

Wang Shaoqing, Zou Benyou, Li Cuiping, Zhao Kankan, Liu Qiang, and Chen Hong. (2015). *CROWN:* A context-aware recommender for web news. *Proceedings of the 31st. International Conference on Data Engineering (ICDE)*, 1420-1423.

Watson Ian. (1998). Applying case-based reasoning: Techniques for enterprise systems. *Morgan Kaufmann Publishers Inc*.

Webb Geoffrey I., Pazzani Michael J., and Billsus Daniel. (2001). Machine learning for user modeling. *Journal of User Modeling and User-adapted Interaction, 11(1-2)*, 19-29.

Weber G., Brusilovsky MS., and Steinle F. (2014). ELM-PE: An intelligent learning environment for programming, 1996. *Obtain in: http://www. psychologie. uni-trier.de, 8000*.

Yoshitomi Yasunari, Asada Taro, Kato Ryota, Yoshimitsu Yuuki, Tabuse Masayoshi, Kuwahara Noriaki, and Narumoto Jin. (2015). Music recommendation system through Internet for improving recognition ability using collaborative filtering and impression words. *Journal of Robotics, Networking and Artificial Life, 2(1)*, 54-59.

Zemirli Nesrine. (2012). WebCap: Inferring the user's interests based on a real-time implicit feedback. *Proceedings of IEEE, 7th. International Conference on Digital Information Management (ICDIM),* 62-67.

Zhang, K., Xin, X., Luo, P., and Guot, P. (2017). Fine-grained news recommendation by fusing matrix factorization, topic analysis and knowledge graph representation. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 918-923.

Zhang Mi and Hurley Neil. (2008). Avoiding monotony: Improving the diversity of recommendation lists. *Proceedings of the 2008 ACM Conference on Recommender Systems*, 123-130.

Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., and Li, Z. (2018). DRN: A deep reinforcement learning framework for news recommendation. *Proceedings of the 2018 International Conference on World Wide Web,* 167-176.

Zheng Huanyang and Wu Jie. (2016). Which, when, and how: Hierarchical clustering with human–machine cooperation. *Journal of Algorithms, 9(4)*, 88.

Zheng Li, Li Lei, Hong Wenxing, and Li Tao. (2012). PENETRATE: Personalized news recommendation using ensemble hierarchical clustering. *Journal of Expert Systems with Applications, 40(6)*, 2127-2136.

Zhou Zhurong and Yang Dengwu. (2014). Personalized recommendation of preferred paths based on web log. *Journal of Software, 9(3)*, 684-688.

Ziegler Cai-Nicolas, McNee Sean M., Konstan Joseph A., and Lausen Georg. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th. International Conference on World Wide Web*, 22-32.