

## A Review and Classification of Recommender Systems Research

Deuk Hee Park  
School of Management  
KyungHee University  
Seoul, Korea  
e-mail: parkdeukhee@khu.ac.kr

Hyea Kyeong Kim  
School of Management  
KyungHee University  
Seoul, Korea  
e-mail: kimhk@khu.ac.kr

Il Young Choi  
School of Management  
KyungHee University  
Seoul, Korea  
e-mail: choice102@khu.ac.kr

Jae Kyeong Kim  
School of Management  
KyungHee University  
Seoul, Korea  
e-mail : jaek@khu.ac.kr

**Abstract**— To understand the trend of recommender system researches by examining the published literature, and to provide practitioners and researchers with insight and future direction on recommender systems, we reviewed 164 articles on recommender systems from 31 journals which were published from 2001 to 2009. The literature search on recommender systems was conducted from top 125 journals of the MIS Journal Rankings. Using our classification framework, we classify the review literature into eight categories each of recommendation fields and eight categories into data mining techniques. The review and classification framework was independently verified. Findings of this paper indicate that the research field of movie field received most research field. We hope that this paper helps anyone who is interested in recommender systems research to suggest insight for future research

**Keywords**- *Recommender systems; Literature review; Data mining technique; Classification*

### I. INTRODUCTION

Recommender systems have become an important research field since emergence of the first paper on collaborative filtering in the mid-1990s [7] [15] [19].

In general, recommender systems are defined as the supporting systems which help users to find information, products, or services (such as books, movies, music, digital products, web sites, and TV programs) by aggregating and analyzing suggestions from other users, which means reviews from various authorities, and user attributes[5][6][9]. Recommender systems are broadly classified into collaborative filtering (CF) and content-based filtering (CB). CF is an information filtering technique based on user's evaluation of items or previous purchases records. However, this has exposed two major issues: sparsity problem and scalability problem [3][16][17][18]. CB analyzes a set of

items rated by an individual user and uses the content of these items, as well as the provided ratings, to infer profile that can be used to recommend additional item of interest [1]. However, syntactic nature of CB to detect similarity between items that share the same attributes or features causes overspecialized recommendations that only include very similar items to those the user already knows [22].

Over the last decade, lots of researchers have studied new approaches of recommender systems to solve problems of CF and CB, and to apply to real world. Especially, applications of data mining technique to recommender systems have been effective to offer personalized information to the user through analyzing his/her preference.

### II. RESEARCH METHODOLOGY

The aim of study is to understand the trend of recommender system researches by examining the published literature, and to provide practitioners and researchers with insight and future direction on recommender systems.

Thus, we will identify distribution of articles on recommender systems by year of publication, and classify the articles by data mining technique used for recommendation and by the recommendation field. However, considering the nature of the research on recommender systems, it would be difficult to confine to specific disciplines. Moreover evidence of this can be seen from the fact that articles on recommender systems are scattered across various journal in disciplines such as marketing, information technology, information science, computer science, management and business. So, we compile the increasing literature on recommender systems systematically. Consequently, following various online journal databases were searched to provide a comprehensive bibliography of the academic literature on recommender systems:

- EBSCO Academic Search Premier;
- EBSCO Business Source Premier;

- Science Direct;
- IEEE/IEE Library;
- ABI/INFORM Database;
- ACM Portal

The literature search on recommender systems was conducted from top 125 journals of the MIS Journal Rankings. The literature search was based on the descriptors “Recommender system”, “Recommendation system”, “Personalization system”, “Collaborative filtering” and “Contents filtering”. The full text of each article was reviewed to eliminate the article that was not actually related to recommender system. Many of articles were excluded because those criteria were unfit for our research, described as below:

- Conference papers, master’s and doctoral dissertations, textbook, unpublished working papers, non-English publication papers and news were excluded because the majority of practitioners and academics often use journals to gain information and disseminate release new finding.
- As researches on recommender systems in this area is relatively current, we have only searched research articles published from 2001 and up to the end of 2009. This 9-year period is considered to be typical of the recommender systems.
- Only those articles that obviously described how the mentioned recommender systems could be applied in the field were selected.

We extracted 164 articles on recommender systems from 31 journals. Each article was prudently reviewed and classified into one of the eight categories in the recommendation field and data mining technique. Although the search was not exhaustive, it serves as a comprehensive basis for an understanding recommender systems research.

### III. CLASSIFICATION METHOD

The framework includes a recommendation field and data mining techniques. In this paper, we classify the review literature into eight categories each of recommendation fields and eight categories into data mining techniques. Graphical classification framework for recommender systems articles are shown in Figure 1.

#### A. Classification framework for Recommendation fields

Many recommender systems have used to suggest consumers with information to help them determine which products to purchase [2]. However, the existing literatures have not classified systematically thought recommender systems have been applied to the various business, thus it is meaningful to investigating recommendation field. We classify articles by recommendation fields such as book, document, image, movie, music, shopping, TV program and others. Through in-depth reviews of literature, shopping field includes online, offline and mobile shopping product and document field include paper, web blog and web page. Also others field includes a minority of recommendation field such as tour, food, news and so on.

#### B. Classification framework for Data mining techniques

In general, data mining techniques are defined as extracting or mining knowledge from amount of data. It is used for the exploration and analysis of large quantities of data in order to discover meaningful pattern and rules [11]. It can be used to lead decision making and predict the effect of decision. Especially, a lot of researchers have used data mining technique for improving performance of recommender systems. Accordingly, it is meaningful to classify the articles by data mining techniques used for the recommend articles. We broadly classified data mining techniques into the following eight categories: Association rule, Clustering, Decision tree, k-nearest neighbor, Link analysis, Neural network, Regression, and other heuristic methods.

- Association rule: Given a set of transactions where each transaction set of items, an association rule applies the form  $X \Rightarrow Y$ , where X and Y are two sets of items [23].
- Clustering: Clustering method is identifying a finite set of categories or clusters to describe the data. Among the clustering methods, most popular clustering methods are K-means and SOM. K-means takes the input parameter, K, and partitioning a set of n objects into K clusters [12]. The self-organizing map (SOM) is a method for an unsupervised learning, based on artificial neurons clustering technique [21].
- Decision tree: Most of popular classification method is the decision tree induction. The top node in a tree is the root node. A decision tree is a tree that each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each terminal node (leaf node) denotes a class prediction [10].
- k-nearest neighbor: The k-NN (k-nearest neighbor) model make a user profile using the user’s preference ratings that are gained directly from explicit ratings of items or provided indirectly from the purchase or usage information [8].
- Neural network: Neural network build a class of very pliable model that can be used for a diversity of different applications. e.g. prediction or non-linear regression, classification. Unfortunately, the term ‘neural network’ is not distinct defined [20].
- Link analysis: Link Analysis has presented great potential in improving the accomplishment of web search.



Figure 1. Classification framework.

Link analysis consists of PageRank and HITS algorithms. Most link analysis algorithms handle a web page as a single node in the web graph [4].

- Regression: Regression analysis is a powerful and diversity process for analyzing associative relationships between dependent variable and one or more independent variables. Uses of regression contain curve fitting, prediction, testing systematic hypotheses about relationships between variables [14].
- Other heuristic methods: Heuristic methods developed by researcher, include mixture model, ontology method and so on.

### C. Classification process

Each of selected articles was reviewed and classified according to the suggested classification framework by two of the four researchers in the team. The classification process is composed of the following four steps:

- Online data base search
- Initial classification by one of the two researchers in the first team
- Independent verification of classification results by one of the two researchers in the second team; and
- Final verification of classification results discussed by two of the four researchers in the team

In case of disagreement in the classification of the article, we exclude the article. The selected criteria and evaluation framework is shown in Figure 2. The set of articles was analyzed in accordance with recommender systems, by year of publication, by journal in which the article were published, and by recommendation field and data mining technique

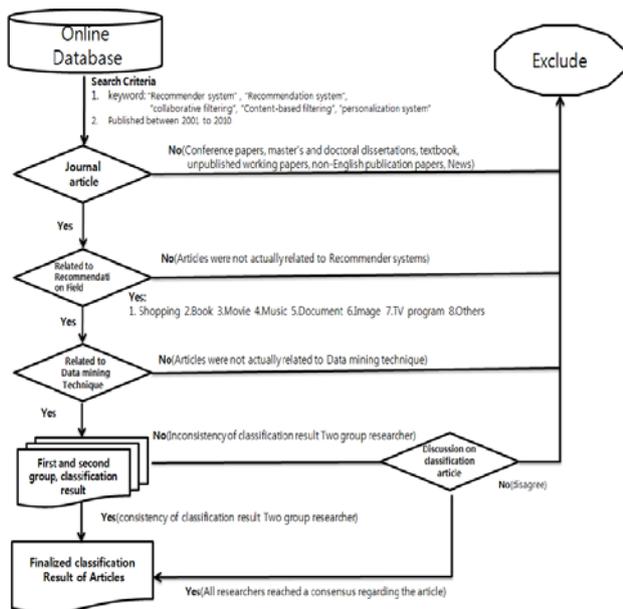


Figure 2. Selection criteria and evaluation framework.

## IV. CLASSIFICATION OF THE ARTICLES

We selected a total of 164 articles from 31 journals. According as classification framework, the selected articles were classified. The result of analysis will offer guideline for future research on recommender systems. The details are expressed below.

### A. Distribution of article by year of publication

The distribution of articles by year of publication from 2001 to 2009 is presented in Figure 3. It is apparent that publications which are related to recommender systems have steadily increased from 2000 to 2004, while these researches have rapidly increased from 2007 to 2009. Such a rapid increase is guessed as an issue happened in conjunction with recommender systems.

### B. Distribution of articles by journal in which the article was published

Our results contain a total of 31 different journals. Distribution of articles by journal is shown Table 2. The majority of articles are published in the Expert Systems with Applications, which focus on the knowledge of the application of expert and intelligent system industry, government and university worldwide, contains more than 35% (61 out of 164 articles, 37.20%) of the total number articles published [13]. The Intelligent system, IEEE (19 out of 164 articles, 11.59%), and the Decision Support System (11 out of 164 articles, 6.71%) had the second and third biggest percentage of recommender systems articles among the journals.

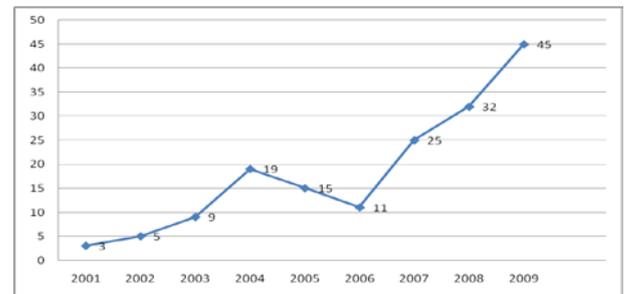


Figure 3. Distribution of articles by year of publication.

TABLE I. DISTRIBUTION OF ARTICLES BY JOURNALS IN WHICH THE ARTICLES WERE PUBLISHED

Journal	No. of articles
Expert Systems with Applications	61
IEEE, Intelligent Systems	19
Decision Support Systems	12
ACM Transactions on Information Systems	11
Knowledge-Based Systems	7
IEEE Transactions on Consumer Electronics,	7
International Journal of Electronic Commerce	7
IEEE, Internet Computing,	6
IEEE Transactions on Knowledge and Data Engineering,	4
Electronic Commerce Research & Applications	3

C. Distribution of articles by journal in which the article was published

Distribution of articles by recommendation field is shown in Figure 4. The most of the articles are related to movie field (46 out of 164 articles, 28.04%) and shopping field (33 out of 164 articles, 20.9%). Because movie and shopping recommender systems have abundance of practical applications than others, it is interpreted that many articles

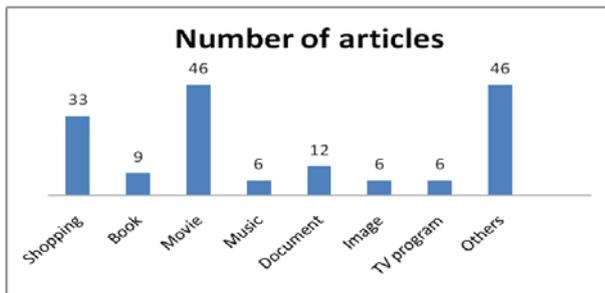


Figure 4. Distribution of articles by recommendation fields.

TABLE II. DISTRIBUTION OF ARTICLES BY RECOMMENDATION FIELDS AND JOURNALS

Field	Journal	No. of articles
Book	ACM Transactions on Information Systems	2
	Decision Support Systems	2
	IEEE, Internet Computing,	2
Document	Expert Systems with Applications	3
	IEEE, Intelligent Systems	3
	Decision Support Systems	2
Image	Expert Systems with Applications	4
Movie	Expert Systems with Applications	18
	ACM Transactions on Information Systems	6
	International Journal of Electronic Commerce	4
Music	IEEE Transactions on Audio, Speech, and Language Processing,	2
	Expert Systems with Applications	1
Shopping	Expert Systems with Applications	13
	IEEE ,Intelligent Systems	5
	Decision Support Systems	3
TV program	IEEE Transactions on Consumer Electronics	3
Others	Expert Systems with Applications	20
	IEEE, Intelligent Systems,	6

Note: The selected journals contain at least two articles by each recommendation field

were published while the fewest of the articles are related to Image, Music and TV program field (6 out of 164 articles, 3.67%).

Distribution of articles by recommendation field and journal is shown in Table 2. Among the recommendation field and journal, the Expert Systems with Applications contained most of recommendation field.

Distribution of articles by recommendation field and data mining technique is shown in Table 3. Among the data mining technique, k-NN (k-nearest neighbor) is used in the most of recommendation field. Because CF system is the one of the most successful recommender systems, k-NN is applied in the most of the recommendation field.

V. CONCLUSION AND FUTURE WORK

Recommender systems have attracted the attention of between academics and practitioners. In this paper, we have identified 164 articles on recommender systems, which are published from 2001 to 2009 to understand the trend of recommender systems research and to provide practitioners and researchers with insight and future direction on recommender systems. The results represented in this paper have several significant implications.

- Based on previous publication rates and the growing interest in the field research on recommender systems will grow significantly in the future.
- Of the 46 articles related to movie recommendation whereas image, music and TV program recommendation are identified only 6 articles. Therefore, more researches are required to be studied for image and TV program recommendation.

TABLE III. DISTRIBUTION OF ARTICLES BY RECOMMENDATION FIELDS AND DATA MINING TECHNIQUES

Field	Data mining technique	No. of techniques
Book	k-NN	1
	Regression	1
Document	Link analysis	1
	k-NN	3
	Clustering	4
Image	Neural network	2
	k-NN	3
	Other heuristic methods	1
Movie	Link analysis	1
	k-NN	20
	Clustering	7
Music	Association rule	4
	k-NN	2
	Clustering	1
Shopping	Clustering	1
	k-NN	10
	Association rule	10
TV program	Clustering	3
	Other heuristic methods	3
	Decision tree	1
Others	Link analysis	1
	Clustering	4
	k-NN	3
	Regression	3

- Recently, social network analysis has been used in the various applications. However studies on recommender systems using social network analysis are deficient. Henceforth, we expect that new approaches using social network analysis will be developed in the recommender systems.
- Our classification model will provide the practitioner and academic with guideline for future research on recommender systems.

However our research might have some limitations. First, we only surveyed articles published from 2001 to 2009, in which searching are based on 125 journals of the MIS Journal Rankings. Second, our finding is based on articles in which were selected from only academic journals. Third, our study was conducted based on a keyword search of “Recommender system”, “Recommendation system”, “Personalization system”, “Collaborative filtering” and “Contents filtering”. Besides these 5 keywords, we do not search all the keywords including “Hybrid Filtering” and so on. Finally, we do not include non-English paper in study.

We have not completed classifying the articles by data mining techniques. Accordingly, we will progress them continuously. Moreover, it is also necessary for us to include conference papers for extending our classification model.

#### ACKNOWLEDGMENT

This work was supported by the Knowledge service Ubiquitous Sensing Network (USN) Industrial Strategic technology development program, 10035426, Personalization marketer for an intelligent exhibit marketing funded by the Ministry of Knowledge Economy (MKE, Korea).

#### REFERENCES

- [1] Basu, C., Hirsh, H. and Cohen, W. "Recommendation as classification: using social and content-based information in recommendation", In Proceedings of the 15th National Conference on Artificial Intelligence, 1998, pp.714-720
- [2] Ben Schafer J., Joseph A. and John Riedl. "E-commerce recommendation Application", Data Mining and Knowledge Discovery, Vol. 5, 2001, pp.115-153
- [3] Claypool, M., Gokhale, A., Miranda, T., Murmikov, P., Netes, D., and Sartin M. "Combining content-based and collaborative filters in an online newspaper", Proceedings of the ACM SIGIR'99 Workshop on Recommender Systems, 1999
- [4] Deng C., Xiaofei H., Ji-Rong W., and Wei-Ying M. "Block-level Link Analysis", Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, 2004, pp.440 - 447
- [5] Frias-Martinez, E., Magoulas, G., Chen, S. Y., and Macredie, R., "Automated user modeling for personalized digital libraries", International Journal of Information Management, Vol. 26, 2006, pp.234-248.
- [6] Frias-Martinez, E., Chen, S. Y., and Liu, X. "Evaluation of a personalized digital library based on cognitive styles: Adaptivity vs. Adaptability", International Journal of Information Management, Vol. 49, 2009, pp.48-56
- [7] Herlocker, J.L., and Konstan, J.A. "content-Independent Task-Focused Recommendation", IEEE Internet Computing, Vol. 5, 2001, pp. 40-47
- [8] Hyea Kyeong Kim, Jae Kyeong Kim, and Young U Ryu. "Personalized Recommendation over a Customer Network for Ubiquitous Shopping", IEEE Transaction on service computing, Vol.2, NO.2, 2009
- [9] Jae Kyeong Kim, Hyea Kyeong Kim and Hee Young Oh and Young U. Ryu. "A group recommendation system for online communities", International Journal of Information Management, Vol. 30, 2010, pp.212-219
- [10] Jae Kyeong Kim , Yoon Ho Cho, Woo Ju Kim, Je Ran Kim and Ji Hae Suh. "A personalized recommendation procedure for Internet shopping", Electronic Commerce Research and Applications, Vol.1, 2002, pp.301-313
- [11] Michael, J.A Berry A., Gordon S. Linoff. "Data Mining Techniques Second Edition", wiley publishigng, Inc, 2004
- [12] Michael, J.A Berry A., Gordon S. Linoff. "Data Mining Techniques Second Edition", wiley publishigng, Inc, 2004
- [13] Ngai, E.W.T., Li Xiu and Chau, D.C.K. "Application of data mining techniques in customer relationship management : a literature review and classification", Expert System Applications, vol. 36, 2009, pp.2592-2602
- [14] Naresh k., Malhotra. "Marketing research An applied orientation fifth edition", PearsonEducation Inc, 2007
- [15] Resnick, P., Iakovou, N., Sushak, M., Bergstrom, P. and Riedl, J. "GroupLens An Open Architecture for Collaborative Filtering of Netnews," Computer Supported Cooperative Work Conf, 1994
- [16] Sarwar, B. "Sparsity, scalability, and distribution in recommender systems", PhD thesis, University of Minnesota, 2001
- [17] Sarwar, B., Karypis, G., Konstan, J. A., and Riedl, J. "Application of dimensionality reduction in recommender system—A case study", Proceedings of the ACM WebKDD-2000 Workshop, 2000a
- [18] Sarwar, B., Karypis, G., Konstan, J. A., and Riedl, J. "Analysis of recommendation algorithms for e-commerce", Proceedings of the ACM E-Commerce, 2000b, pp.158-167
- [19] Shardanand, U., Maes, P. "Social Information Filtering: Algorithms for Automating 'Word of Mouth'", Conf. Human Factors in Computing Systems, 1995
- [20] Ulrich Anders, Olaf Korn. "Model selection in neural networks ", Neural Networks, Vol.12, 1999, pp.309-323
- [21] Wu Lihua, Liu Lu, Li Jing and Li Zongyong. "Modeling user multiple interests by an improved GCS approach", Expert Systems with Applications, Vol.29, 2005, pp.757-767
- [22] Yolanda Blanco-Ferna'ndez, Jose J. Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, Martn Lopez-Nores, Jorge GarcDuque, Ana Fernandez-Vilas, Rebeca P. Daz-Redondo, Jesus Bermejo-Munoz. "A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems", Knowledge-Based Systems, Vol.21, 2008, pp.305-320
- [23] Yoon Ho Cho, Jae Kyeong Kim and Soung Hie Kim. "A personalized recommender system based on web usage mining and decision tree induction", Expert System Applications, Vol.23, 2002, pp.329-342