The Development, Status and Trends of Recommender Systems: A Comprehensive and Critical Literature Review

Jin Xu[1], Karaleise Johnson-Wahrmann[1], Shuliang Li[1,2]

[1] School of Economics & Management, Southwest Jiaotong University, Chengdu, Sichuan, 610031, China
xj_james@126.com
[2] Westminster Business School, University of Westminster, 35 Marylebone Road, London NW1 5LS, UK
shuliangliuk@gmail.com

***Abstract*—**Recommender systems have been used in many fields of research and business applications. In this paper, a comprehensive and critical review of the literature on recommender systems is provided. A classification mechanism of recommender systems is proposed. The review pays attention to and covers the recommender system algorithms, application areas and data mining techniques published in relevant peer-reviewed journals between 2001 and 2013. The development of the field, status and trends are analyzed and discussed in the paper.

***Keywords*-**Recommender Systems; Recommendation Systems; Comprehensive and Critical Literature Review; Research Status and Trends

I. INTRODUCTION

With the increasing availability of large volumes of data, it has become cumbersome to wade through all available options in the hopes of finding what one desires. Even with advancements in searching capabilities, the user may still have many options that do not necessarily fit their personal preferences. This information overload problem as given rise to recommender systems.

Recommender Systems (RS, or Recommendation Systems) track past actions of a group of users to make recommendations to individual members of the group [1]. A recommender system is one of the most interesting research areas for investigating information overload [2]. The recommender system's task is to turn users' current preference data into predictions of future likes and interests [3]. Recommender systems are responsible for providing users with a series of personalized suggestions (recommendations) on particular items. A recommender system extracts the user's relevant characteristics to form user profile, it then determines the set of items that may be of interest based on those characteristics [4]. In most recommender systems, users provide recommendations as input that the system can then aggregate and redirect to other appropriate users [5].

Recommender systems have had extensive application in ecommerce where there are a high volume of products for shoppers to choose from. Recommender systems have become a necessary part of the online shopping experience not just for consumers but retailers as well [6]. Popular websites such as Amazon.com, Ebay.com, Taobao.com among many others, use recommendations to suggest products to their users based on past activities. They are also popular in other domains such as mobile applications, movie recommendations, social networks, document recommendations etc…

However, there is still lack of effort on conducting a comprehensive literature review of the past 10 years' research on recommender systems. Therefore, in this paper, we are going to propose a classification mechanism to figure out the blueprint of recommender systems' research and give an analysis and critical evaluation on development, status and trends .

II. RECOMMENDER SYSTEMS APPROACHES

Recommender systems produce recommendations either through a collaborative filtering approach (CF) or a content-based filtering approach (CBF) [2]. It is a process by which information on the preference and actions of a group of users is tracked by the system [1]. It is an emerging technology to deal with information overload by using customer interests to guide them to products they might like [7].

CF recommender systems work by storing relevant information in a database that contains the ratings of a large number of users on a large number of items (films, books, jokes, study material, holiday destinations, etc.). CF aims to determine similarity between users and then to recommend to the active user the items preferred by users similar to him or her [4].

CBF recommender systems make suggestions to users based on a profile learnt from previously rated items [8]. CBF recommendation systems typically (1) build an item profile from extracted attributes from each item in the set, (2) build a content-based user profile based on the attributes of the items which each user has purchased, (3) calculate the similarity between users and items based on profile information using similarity functions and (4) make recommendations of top *n* items with highest similarity scores [9]. The content-based filtering methods analyzes the content of items and try to understand the similarities between items for generating relevant recommendations.

ISBN: 978-1-61804-247-7

Thus, a user is offered recommendations of those items which have high similarity to the ones the user showed preference for in the past [10].

Both recommender approaches can be augmented with context awareness, that is, the ability for the recommender to be aware of the environment in which the activity is taking place. Context awareness is about capturing a wide range of contextual attributes (such as location and surrounding environments) to better understand what the user is trying to accomplish, and what services the user might prefer [11].

Although both approaches can generate reasonable recommendations, each of them has its own drawbacks. CBF recommendations do not necessarily include preference similarity across individuals while CF do not necessarily incorporate feature information and must face the sparsity and cold-start problems [3] . Cold start problems occur when not enough information is available to make recommendations for new users. CBF relies heavily on textual descriptions, leading to other spinoff problems such as limited information retrieval, new user issues, and overspecialized recommendations [12]. Many hybrid methods that combine the two main approaches has been explored to find solutions to these problems.

The research into recommender systems has strong focus on finding solutions to the above mentioned problems, however a complete solution has not yet been created and the research trend shows continued efforts to find both new applications for recommender systems as well as applying multi-disciplinary methods to solve its existing problems.

III. Review Methodology

A. Review Process

The review methodology employed involved examining articles on recommender systems published in relevant peer-reviewed journals over a 12 year period between the years 2001 and 2013. The purpose of the review is to understand the development, status and trends that exist in the field and to explore the algorithms used that may be of benefit for its application some domains. The following relevance criteria were employed: *Algorithms used in recommender systems, prototypes of recommender systems, validation and evaluation and performance of recommender systems*. The algorithms would give details of the research done in recommender algorithms that can be applied in further researches, the prototypes will give insights in the fields of applications and the evaluation will give information on the research on recommendation system performance.

To accomplish this, a search was done on major research database for the terms *Recommender System, Recommendation System, Personalization System, Collaborative Filtering* and *Content Based Filtering*. Of the papers returned; the titles and abstracts were read to determine relevance. Excluded from consideration were the following results: Conference papers, Masters and PHD Dissertations, Unpublished working papers, Editorials, and Reports. Once a paper was considered of relevance, then it is classified according to the area of focus of the paper. The classifications were then analyzed, with the results collated and reported.

B. Classification Mechanism

An abundance of papers related to recommender systems exists and are scattered across different disciplines and in different journal databases. Therefore it was very difficult to classify papers according to disciplines.

Park et al [13] had previously done a literature review of recommender systems between the years 2001 and 2010. Their study had similar relevance criteria but lacked granularity in categorizing the recommender system approaches and did not specify the focus areas of the papers in their review. Since 2010 some of the data mining techniques have evolved and others have become more popular thus requiring their own categories. New algorithms have been incorporated and researchers have found new application fields for recommender systems.

For these reasons it was not possible to simple extend the study by Park et al to 2013, but rather to revisit the literature from 2001 in order to make the classifications that were of interest for this research. The study by Park et al was therefore very useful as a guide and for validation purposes but the articles presented in this review were individually analyzed and classified to match the criteria described in the classification methodology.

The classification mechanism employed is as follows:

1) Application Areas

The application area refers to the type of items that is being recommended in the articles. These can fall into the following categories:

- E-commerce - recommenders of products and services for purchase online
- Education - recommendations relating to learning such as courses, research papers etc…
- Entertainment - recommendations of multimedia such as movies, music, images etc…
- Book/Documents - recommendations on documents and books
- Tourism/Travel - recommendations for travel destinations and tourist activities
- Health Care - recommendations on medical items
- Social Media - recommendations on social activities
- Web pages/ Online News - recommendations of web sites and new feeds
- Other - all other application areas not listed above

2) Algorithm Type

The algorithms type refers to the recommendation algorithm that the paper discussed. These were group in the categories:

- Collaborative Filtering
- Content Based Filtering
- Context Aware
- Hybrid

The hybrid category is given to those papers that discuss the use of a combination of any of the three previous categories.

3) Information Retrieval Technique

Information Retrieval techniques refer to the data mining technique discussed in the paper that is used in determining the similarity of items or users that generate the recommendations. The categories are taken from the most popular data mining techniques and are as follows:

- K- Nearest Neighbour (KNN) - This algorithm calculates a set of k users whose similarities are most comparable to the user for whom a recommendation is intended (the active user) [14]. The algorithm searches for users similar to the active user in terms of ratings for the previously seen items. Then, ratings predictions are made for the unknown item based on the ratings that were assigned to this item by other similar users [15].
- Association Rule Mining (ARM) - The aim of ARM is to uncover inter-relationships of two or more items included in a transaction. An association rule is expressed as "If U, then V (U)V)", interpreted as "If event U occurs, then event V also occurs". Rules generated by ARM are easily to understand, and easy to deploy in practice [16].
- Clustering - Data clustering determines a group of patterns in a dataset that are homogeneous in nature. It is an unsupervised pattern classification technique that defines a group of n objects into m clusters without any prior knowledge [17]. The objective is to develop an automatic algorithm that is able to accurately classify unleveled datasets into groups [18]. The clustering problem addresses the partitioning of datasets into n patterns in a d-dimensional space into K distinct set of clusters, in such a way that the data within the same cluster have a higher similarity to each other than to data in other clusters [19].
- Fuzzy Set - Fuzzy sets deal with subjective rather than precise reasoning. For example: a buyer might consider an item to be expensive but the linguistic label "expensive" is not precise and may differ depending on the buyer. "The fuzzy linguistic approach is based on the representation of qualitative aspects as linguistic values by means of linguistic variables" [20]. It is also concerned with approximate reasoning under uncertainty with certain level of confidence or a degree of certainty.
- Singular Value Decomposition (SVD) - SVD is used fundamentally in dealing with noisy data [21]. The SVD of an m×n real matrix A is to mathematically transform A to a diagonal matrix, with nonnegative diagonal elements, through a transformation of the form PAQ with an m×m orthogonal matrix P and an n×n orthogonal matrix Q [21]. An important feature of SVD, that is useful for recommender systems, is that it provides the best low-rank approximation of the original matrix [22].
- Latent Semantic Analysis (LSA) - LSA is an automatic mathematical/statistical technique for extracting meaning and inferring relationships of expected contextual usage of words [23]. LSA is originally an information retrieval method, however, it is also widely used in text categorization [24]. "The purpose of LSA is to extract a smaller number of dimensions that are more robust indicators of meaning than the individual terms" [17].
- Naïve Bayes / Bayesian Network (BN) - BN is a probabilistic model that provides a representation of a joint probability distribution [25]. BN nodes are graphically represent attributes and arcs represent attribute dependencies. Attribute dependencies are quantified by conditional probabilities for each node given its parents. Bayesian networks are often used for classification problems. With BN a learner attempts to construct a classifier from a given set of training instances with class labels [26].
- Other – All other techniques.

4) Focus Area

The areas of focus of the papers reviewed will be classified in the following 9 categories:

- Review/Survey - Papers that conduct literature reviews, surveys, and overviews of recommender systems without specific interest in any particular domain application.
- Empirical Studies - Papers that focus mainly on empirical evidence gathered in recommendation system usage studies.
- Performance/Evaluation - Papers that focus on the evaluation of recommender systems algorithms and performance metrics.
- Algorithm Improvement - Papers that focus on improving the main algorithm types and showing the merits of their improved approach.
- Prototype/Simulation - Papers that present a working prototype of a recommender system or a simulation of that system.
- New technique - Papers that focus on applying a new cross discipline techniques to improving recommendations.
- User Behavior - Papers focusing on the human actions and reactions when interacting with recommender systems.
- Mobile Specific - Papers that focus specifically on recommender systems for a mobile channel.
- Data Capture - Papers that focus on the explicit or implicit methods of capturing the preference data used to generate recommendations.

IV. The Findings and Results

The results revealed a total of 403 articles of relevance between that periods. Each of the relevant papers were reviewed and classified according to the methodology and mechanism proposed. The papers were analyzed as follows:

A. Distribution by Year of Publication

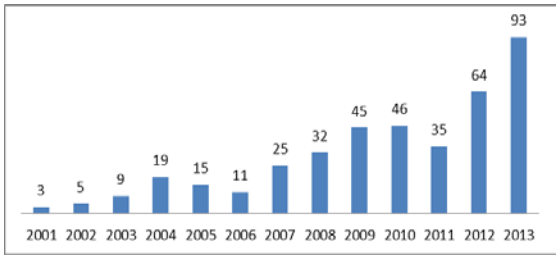

Figure 1. Article distribution by year of publication

The results showed that recommender system is still popular among researchers despite over 20 years since the first recommender system publication. The trends show a steady rise in research publishing since 2001 with 93 papers published in 2013 accounting for the highest number of publishing so far in any one year. This demonstrated that recommender systems is still a hot research topic and has captured the attention of researchers for some time. The full distribution by year is shown in Figure 1.

B. *Distribution by Journals*

The review resulted in 403 articles published between 2001 and 2013 from peer reviewed journals that match the relevance criteria. The results revealed papers from a total of 72 different journals across various disciplines. The top publishing journal for recommender system articles was *Expert Systems with Applications* with a total of 129 articles accounting for 32.8%. *Knowledge Based Systems* journal as well as *Decision Support Systems*, *Information Sciences* and *IEEE Intelligent Systems* were also amongst the top publishers accounting for between 5.3% to 7.6% of the total papers. The gap between the second highest publishing journal, *Knowledge Based Systems* at 7.6% and *Expert Systems with Applications* is very wide thus indicating that the latter plays a leading role in publishing research on recommender systems. Distributions of the research by top 15 journals is shown in Table 1.

Table 1. Paper distribution by Journal

Journal	Amount
Expert Systems with Applications	129
Knowledge-Based Systems	30
Decision Support Systems	25
Information Sciences	24
IEEE Intelligent Systems	21
Information Processing & Management	16
ACM Transactions on Information Systems	12
Electronic Commerce Research and Applications	10
IEEE Transactions on Consumer Electronics	10
IEEE Transactions on Knowledge and Data Engineering	10
IEEE Transactions on Internet Computing	9
International Journal of Electronic Commerce	7
Journal of Systems and Software	7
Computers in Human Behavior	5
International Journal of Human-Computer Studies	5

C. *Distribution by Application Field*

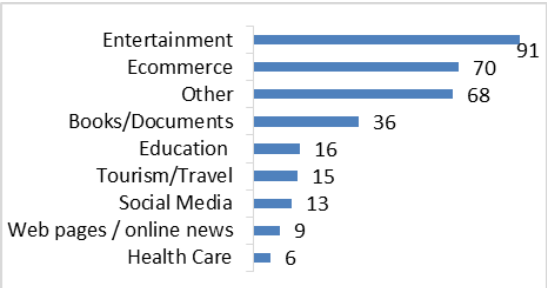

Figure 2. Article distribution by application field

There were many articles that discussed recommender system but did not mention a specific field of application. Of the ones that did specify an application field, the results are shown in Figure 2. As seen the majority of the article discussed application to entertainment particularly movie and music recommendations, however there were also those articles that discussed image and other video recommendations that were also classified in the entertainment category. E-commerce was also very popular among researchers and a total of 70 articles were found that discussed recommending products for online shoppers. The categories of 'other' grouped recommenders of varying items include real estates, restaurants, stocks, digital ecosystems and others.

D. *Distribution by Algorithm*

There were many articles that discussed recommender systems but did not get into a specific algorithm, such articles mainly gave an overview of the recommendations generated. Of the articles that specified the algorithm type, collaborative filtering was by far the most popular among researchers, followed by hybrid algorithms. Context aware algorithms became more popular in later years since 2008. In 2001 there was a complete absence of papers that focused on any of these algorithms. The distribution of papers by algorithm type can be seen in Figure 3.

There was a clear interest by researchers in collaborative filtering algorithms and much effort was expelled in developing new ways to improve such algorithms in various areas of application. Many existing data mining techniques were used including popular techniques such as K-nearest neighbor, association rule and clustering, however many more were discussed but were too numerous to itemize so there were categorized as other. Some of these other data mining techniques explained in the papers included: vector space model, matrix factorization, neural networks, genetic algorithms, product taxonomy and others.

E. *Distribution by Focus Area*

The review results show that the majority of papers written on recommender systems during the years of study were about improving the algorithms making up 35% of the total papers analyzed. Prototypes and simulations were also popular as researchers demonstrated the utility of the recommender systems they had developed, this accounted for 25% of the papers. Evaluations of the performance of recommender systems was also of interest to researcher accounting for 13%. The full distribution of the focus areas can be seen in Figure 4.

V. CONCLUSIONS AND DISCUSSIONS

It is clear from the review that recommender systems have captured the attention of researchers since its inception and is still doing so. The application fields for recommender systems seems to be ever expanding, however e-commerce and entertainment has still remained the most popular. Throughout the years of the review, collaborative filtering algorithms were the main approaches used in building recommender systems especially for their social value and particularly for movie, music and product recommendations. In the later years, researchers have shifted more towards

travel and tourism recommendations. Content based algorithms are mainly applied in situations where collaborative filtering may not be feasible such as in education. Context aware systems are gaining in popularity amongst researchers and during the period studied, papers discussing context aware systems were mainly for the mobile channel.

The review revealed the evolution of algorithms over the period. With each application field, a new method is applied to existing algorithms and in some case new algorithms are invented. It was very difficult to categorize all the algorithms as they were so diverse and dynamic in nature. It also showed that no one technique fits all application fields and businesses often change algorithms to match their specific needs. There was also an increase in survey and review papers addressing user behavior and consumer psychology for users interacting with a recommender system. There were also a few papers that discussed the privacy concerns regarding recommender systems.

Of great importance is the almost complete lack of papers focusing on business processes using recommender systems. From an organizational point of view, very little interest was expressed by researchers in improving business operations with recommenders, the focus has been on consumers' needs, users' needs and in one case business-to-business needs. However, almost no interest in improving internal operations with recommender systems has been found. This may be recognized as one of the future research areas.

Another issue is that each of the methods or techniques mentioned in the literature review has its own weaknesses or limitations. Therefore, a hybrid approach [27, 28, 29] integrating the strengths and advantages of various techniques, technologies and methods for developing and applying recommender systems is required. One topic of the studies in this area will be concerned with the creation and use of hybrid intelligent recommender systems.

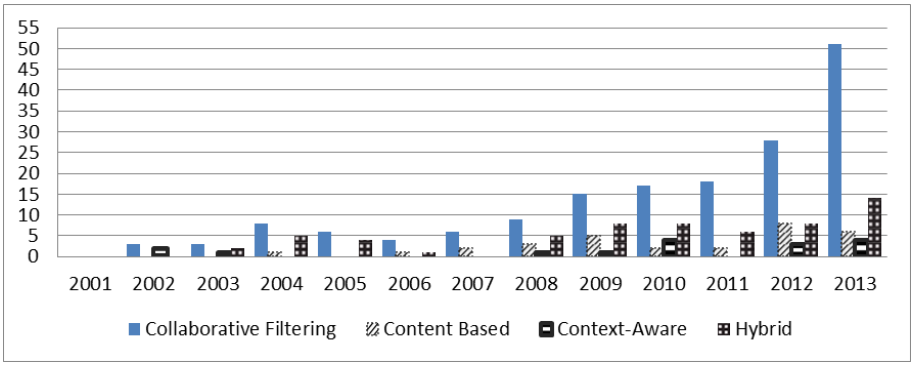

Figure 3. Article distribution by algorithm type

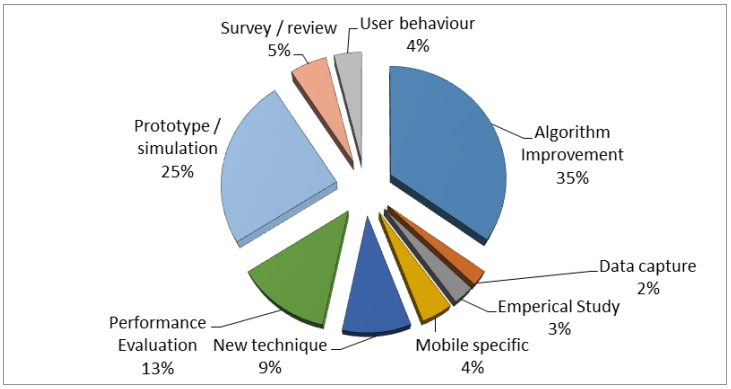

Figure 4. Article distribution by focus area

ACKNOWLEDGEMENT

This study is supported by grants from National Natural Science Foundation of China (71002064 & 71090402), the Fundamental Research Funds for the Central Universities of China (2682013CX075), the Soft Science Research Plan of Sichuan Province (2013ZR0043), the Experimental Teaching and Technique Project of Southwest Jiaotong University (SYJS201324), and Sichuan 100-Talent Scheme (bai ren ji hua) grant.

REFERENCES

[1] R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, "Recommender Systems a probablistic Analysis," Journal of Computer and System Sciences, pp. 42-61, 2001.

[2] X. Z. Hongchen Wu, "Div-clustering:Exploring active users for social collaborative recommendation," Journal ofNetworkandComputerApplications, pp. 1642 - 1650, 2013.

[3] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang and T. Zhou, "Recommender Systems," Physics Reports, pp. 1-49, 2012.

[4] F. Ortega, J.-L. Sánchez, J. Bobadilla and A. Gutiérrez, "Improving collaborative filtering-based recommender systems results using Pareto dominance," Information Science, pp. 50-61, 2013.

[5] P. Resnick and H. R. Varian, "Recommender systems. (Special section: recommender systems)(Cover story)," Association of the ACM, pp. 56-58, 1997.

[6] S. R. L. C. N. Paloma Ochia, "Predictors of use r perceptions of web recommender systems: How the basis for generating experience and search product recommendations affects user responses," International Journal of Human-Computer Studies , pp. 472-482, 2013.

[7] C. Kalel, "An entropy-based neighbor selection approach for collaborative filtering," Knowledge-Based Systems, pp. 273-280, 2014.

[8] R. Mooney, N. P. Bennett and R. Loriene, "Book Recommending Using Text Categorization with Extracted Information," in Workshop on Recommender Systems, Madison, 1998.

[9] Y. S. Keunho Choi, "A new similarity function for selecting neighbors for each target item in collaborative filtering," Knowledge-Based Systems, pp. 146-153, 2013.

[10] K. ShwetaTyagi, "Enhancing collaborative filtering recommendations by utilizing multi-objective particles warm optimization embedded association rule mining," Swarmand Evolutionary Computation, pp. 1-12, 2013.

[11] E.-H. S. J. K. S. K. Jongyi Hong, "Context-aware system for proactive personalized service based on context history," Expert Systems with Applications, pp. 7448-4457, 2009.

[12] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang and J. Lu, "A hybrid fuzzy-based personalized recommender system for," Information Sciences, pp. 117-129, 2013.

[13] H. H. Park, K. H. Kyeong, C. I. Young and K. J. Kyeong, "A literature review and classification of recommender system research," Expert Systems with Applications, pp. 10059-10072, 2012.

[14] J. Bobadilla, F. Ortega, A. Hernando and G. Glez-de-Rivera, "A similarity metric designed to speed up, using hardware, the recommender systems k-nearest neighbors algorithm," Knowledge-Based Systems, pp. 27-34, 2013.

[15] M. Nilashi, O. b. Ibrahim and N. Ithnin, "Hybrid recommendation approaches for multi-criteria collaborative filtering," Expert Systems with Applications, pp. 3879-3900, 2014.

[16] Y. S. Kim and B.-J. Yum, "Recommender system based on click stream data using association rule mining," Expert Systems with Applications, pp. 13320-13327, 2011.

[17] W. Song and S. C. Park, "Genetic algorithm for text clustering based on latent semantic indexing," Computers and Mathematics with Applications, pp. 1901-1907, 2009.

[18] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithm for partitional clustering," Swarm and Evolutionary Computation, pp. 1-18, 2014.

[19] I. Saha and U. Maulik, "Incremental learning based multiobjective fuzzy clustering for categorical data," Information Sciences, pp. 35-57, 2014.

[20] J. Serrano-Guerrero, E. Herrera-Viedma and J. A. Olivas, "A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0," Information Sciences, pp. 1503-1516, 2011.

[21] T. Maehara and K. Murota, "Simultaneous singular value decomposition," Linear Algebra and its Applications, pp. 106-116, 2011.

[22] A. B. Barragáns-Martínez, E. Costa-Montenegro, J. C. Burguillo, M. Rey-López, F. A. Mikic-Fonte and A. Peleteiro, "A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition," Information Sciences, pp. 4290-4311, 2010.

[23] Y. Tonta and H. R. Darvish, "Diffusion of latent semantic analysis as a research tool: A social network analysis approach," Journal of Informetrics, pp. 166-174, 2010.

[24] B. Yu, Z.-b. Xu and C.-h. Li, "Latent semantic analysis for text categorization using neural network," Knowledge-Based Systems, pp. 900-904, 2008.

[25] H. Langseth and T. D. Nielsen, "A latent model for collaborative filtering," International Journal of Approximate Reasoning, pp. 447-466, 2012.

[26] L. Jian, Z. Cai, H. Zhang and D. Wan, "Not so greedy: Randomly Selected Naive Bayes," Expert Systems with Applications, pp. 11022-11028, 2012.

[27] S. Li, “The development of a hybrid intelligent system for developing marketing strategy”, Decision Support Systems, 27(4), 2000, pp. 395-409.

[28] S. Li & J. Z. Li, Li, “AgentsInternational: integration of multiple agents, simulation, knowledge bases and fuzzy logic for international marketing decision making”, Expert Systems with Applications, 37(3), 2010, pp. 2580-2587, 2010.

[29] S. Li & J. Z. Li, WebInternational: combining Web knowledge automation, fuzzy rules and online databases for international marketing planning, Expert Systems with Applications, 37(10), 2010, pp.7094-7100.