A LITERATURE REVIEW AND CLASSIFICATION OF BOOK RECOMMENDATION RESEARCH

Feng Wang[1], Lingling Zhang[2], Xin Xu[3]

[1]   School of Economics and Management, University of Chinese Academy of Sciences, Department of Management and Marketing, Hong Kong Polytechnic University
    Email: wangfeng173@mails.ucas.ac.cn
[2]   School of Economics and Management, University of Chinese Academy of Sciences
    Email: zhanglingling@ucas.ac.cn
[3]   Department of Management and Marketing, Hong Kong Polytechnic University
    Email: xin.xu@polyu.edu.hk

Article Info:

Abstract:

The act of reading has benefits for individuals and societies, which can be a long-term commitment. While the overload of books information and readers' specific needs make book recommendation (BR) in demand, BR is receiving great attention from the research community with different perspectives. The increasing amount of research conducted with BR calls for a classification methodology regarding trends and distribution in this field. This paper presents a study of recommender systems in the domain of BR. The main goal of this work is to provide authors with insights on the trends of academic literature reviews in the proposed context and to present a comparison of different research approaches. The authors searched for up-to-date research papers related to recommender systems for BR within a time period of eighteen years, from 2000 to 2018. Starting from 2000, a significant amount of research related to the subject field of recommender systems was conducted, which led to the first ACM Conference on Recommender Systems. After the filtering process, 39 papers were finally selected from journals, conferences and theses in five different academic databases (i.e. IEEE, ACM, Science Direct, Springer and ProQuest). The general classification is presented in this work, in order to describe the recommendation approaches for BR. This work can be extended in the future to include novel methodologies and trends of recommender systems for BR or other fields.

Keywords:

Classification, Recommender systems, Book recommendation

Introduction

Reading is the basic way for humans to obtain, transmit, exchange and accumulate information and knowledge. As an essential skill for everyone to develop, reading is the driving force behind a productive, successful life, and is an integral part of the educational system (Allington & Gabriel, 2012; Pera & Ng, 2015). From traditional paper-based books and newspapers, to TV commercials, webpage information, e-books, and social networks like Tweeter or Weibo, the objects and ways for people to obtain information are constantly changing, but the need for more efficient and accurate information is ever increasing rather than decreasing.

Reading activity is indispensable to people's daily lives, especially in this information age, when reading is not only restricted to paper-based books but also anything that appears on interface screens. New technologies are transforming the ways Internet operates and the ways people work. They are also reshaping the expectations, needs, and opportunities in reading and learning. This evolving world is boosting demand for new ideas, knowledge and skills. As a result, the amount of instructive information that offers to people is growing rapidly, so does the need to search out the most suitable information become demanding even more quickly. (Hämäläinen et al., 1996). While, the prosperity of Internet and e-commerce leading to information overload, people are overwhelmed by massive amount of information and commodities. Recommender systems designed as information retrieval tool to make personalized suggestions for users meet urgent need and become thriving (Ricci et al., 2015).

While the deployment of recommender systems gain success in many fields, especially in e-commerce selling movie, music, book or any products to consumers personally. The benefits of recommender systems can reach further scenarios. In this paper, we focus on book recommendation. The enormous number of available books makes the use of book recommender system a necessity. In 2010, Google counted 129,864,880, and the count goes up – about one million new and revise titles were published in UK, China and USA in 2013 alone (Alharthi et al., 2018). It is an ambitious objective to find the best match for the target reader in this flood of publications, and to select interested readers for target book. It goes that with the rapid increases of publications, how to find out relevant books for target reader remains a problem.

While systematic research dedicating to the analysis of book recommendation is quite rare. We think that there is a need to have an overlook on book recommendation research, to explore the contribution that dedicated to recommender systems for book recommendation. Over thirty published papers addressed the specific problem of book recommendation. Their work ought to be summarized in order to highlight the trends, methods and potential research opportunities in this area. For future research, the problems and open questions should be identified. In addition, book recommendation owns a set of features different from those of other commodities, which generate specific techniques, algorithms, types, data resources and evaluation metrics used in book recommendation. Our survey aggregates these different feathers into five categories and seven types, according to different mechanisms and objectives of book recommendation. We classify the datasets and evaluation methods and describe how to employ each of them.

The paper is structured as follows: The first section presents the introduction to the research of book recommendation. The second section is related work regarding previous research work on book recommendation and its classification. Then, the next section describes the data collection and methodology used to classify the different research papers. The fourth section shows the classification framework proposed for the different approaches presented in our

selected academic literature, from perspectives of methods, types, algorithms, data resources and evaluation metrics. Concluding remarks and discussion are presented in the final section.

Literature Review

In this section, different methodologies classifying recommender system in the context of book recommendation, are introduced. There are a handful of research works from different perspectives, contributing to the classification of recommender systems. To trace back the origin, the first research paper in recommender systems came out in the mid-1990s (Deshpande & Karypis, 2004). Based on different types of algorithms used to perform information filtering, recommender systems basically can be classified into three types, namely collaborative filtering (CF), content-based filtering (CB) and hybrid filtering (HF) (Pazzani, 1999). Majority of the works in recommender systems are concentrated on techniques based on collaborative filtering (Akhil & Joseph, 2017). CF is an information filtering technique depending on user's evaluation of items or previous purchases records. CF is used in book recommendation frequently. However, it has exposed two major issues: sparsity problem and scalability problem (Miranda et al., 1999; Sarwar, 2001; Sarwar et al., 2000), and other shortcomings, like cold start problem, gray sheep problem, shilling attack problem and so forth (Su & Khoshgoftaar, 2009).

In Park et al. (2012), the authors presented a review and classification of different approaches of recommender systems, grouping them based on their application fields and the types of data mining techniques that were used (Park et al., 2012). The authors identified 164 articles on recommender systems, which are published from 2001 to 2009 conducted from top 125 journals of the MIS Journal Rankings. Using the classification framework, the recommender systems were classified into eight categories of recommendation fields (e.g. shopping, book, movie, music, document, image, TV program and others), and eight categories into data mining techniques (e.g. association rule, clustering, decision tree, k-nearest, neural network, link analysis, regression and other heuristic methods).

In Haifa Alharthi et al. (2018), according to the techniques used, the authors conducted a survey on the book recommender systems and put them into six categories, namely traditional recommendation methods, recommendations based on library loans, stylometry-based recommendations, e-book recommendations, recommendations based on social media. The authors then classified the evaluation methods of recommender systems, including evaluation metrics and datasets. There were challenges remaining in book recommendation. Recommendations of textual items, the effect of mood on book selection, books for non-readers, the long tail in book purchases, audiobook recommendation and recommendation in light of reading modes and trouble with translated books, as challenges and opportunities of book recommendation were put forward. Dedication to the classification of recommender systems on techniques used in microblogs was also presented in Terán et al. (2018), which provided a comparison of different research approaches based on papers selected from four different datasets.

While compared with the significant amount of work done on classification of recommender systems, systematic research dedicating to the analysis of book recommendation is quite few. The objective of this work is to analyze the overall research situation and research trends, and to provide researchers a systematically classification framework in order to assist and inspire further extended research.

Data Collection and Classification Framework

To understand the trends on the use of book recommendation, an analysis of academic literature was conducted. Additionally, a classification framework based on the results of the initial analysis was introduced.

Data Collection

The selected papers were collected from five scientific databases: ACM Digital Library, Elsevier Science Direct, IEEE Library, ProQuest Dissertations & Theses, and Springer Link. The initial keywords used in these databases included the following: Recommender systems, recommendation system, book recommendation, readers recommendation, collaborative filtering, content-based filtering, hybrid-based filtering, reader, author and book. Depending on the results, the keywords are then decomposed and recombined to find the subset that gave the best results. The final list of keywords included: recommendation system, recommender system, book recommendation, readers recommendation.

Table 1: Number of Papers Returned From Scientific Databases with Variation of Keywords

Library	Selection of words	Papers found	Papers selected (Initial/Final)
ACM Portal	Recommender systems, recommendation system, book recommendation	239	50/9
	Recommender systems, recommendation system, book recommendation, readers recommendation	50	
IEEE Xplore	Recommender systems, recommendation system, book recommendation	218	14/8
	Recommender systems, recommendation system, book recommendation, readers recommendation	14	
ProQuest	Recommender systems, recommendation system, book recommendation	2190	15/3
	Recommender systems, recommendation system, book recommendation, readers recommendation	15	
Science Direct	Recommender systems, recommendation system, book recommendation	2574	28/7
	Recommender systems, recommendation system, book recommendation, readers recommendation	28	
Springer Link	Recommender systems, recommendation system, book recommendation	3331	53/12
	Recommender systems, recommendation system, book recommendation, readers recommendation	53	
Total			160/39

Source: drawn by the authors.

Papers were selected if their title contained a combination of at least two keywords from the listed proposed for this study. The contents of the papers were skimmed to check whether they were relevant to be selected for deeper review. During the review process, the articles selected were studied in more detail, and further unrelated papers were discarded. Those selected papers were dedicated to improving the effect of recommendation techniques used for book recommendation, with data resources and evaluation metrics to demonstrate the experiment effects. The final selection of papers relevant to our classification are presented in table 1.

Classification Framework

After the analysis of the selected papers, a classification framework was introduced to better understand the methodologies and trends used in the development of book recommendation. Based on the analysis of the selected papers, five categories were proposed for the classification: methods used, recommendation types, techniques used, dataset used, and evaluation metrics used. Table 2 shows the classification framework to provide a view of the evolution of different technologies used in book recommendation.

Table 2: Classification Framework of Book Recommendation

Basis		Categories		Papers
Book recommendation	Method	Collaborative Filtering (CF)		26
		Content-Based (CB)		7
		Hybrid-Based (HB)		6
	Type	General Books		14
		Libraries		13
		Audiobooks		0
		E-books		5
		Specific Groups		7
		Specific Domains		1
		Others		1
	Technique	Text Representation		1
		Feature Identification		19
		Feature Extraction		25
		Classification		31
		Linkage		18
	Dataset	Open Datasets		12
		Collected Data		18
		Private Data		7
		Others		6
	Evaluation Metric	Usage		35
			Rating	8
			Ranking	17
			Others	4

Source: drawn by the authors.

Distribution Using the Classification Framework

In this section, the distribution of the five categories used in the classification framework proposed in this work, is presented in Figure 1 and Figure 2. Figure 1 shows the distribution of methods used along the 18 years that correspond to this study. It shows that most of the studies were based on collaboration filtering method; nevertheless, the other two methods (content-based and hybrid-based) cannot be neglected as well. Figure 2a shows the trend of all techniques found during the analysis of research papers. It shows a clear tendency of the use of classification, while feature extraction & identification, and linkage are significant. Figure 2b shows the tendency for studies regarding recommendation types. The tendency for studies on general book and library is presented. On the other hand, Figure 2c presents the collected data from online resource is widely used. Finally, Figure 2d shows the obvious tendency of evaluation metrics is the quality of usage, including accuracy, coverage and recall rate.

Classification of Book Recommendation

This section presents the basis of classifying recommendation techniques used for book recommendation and explains the reasons of constructing the classification framework. Details of this framework, from five dependent perspectives, are showed in section 4.1-4.5. Each category gives the description of research status on book recommendation.
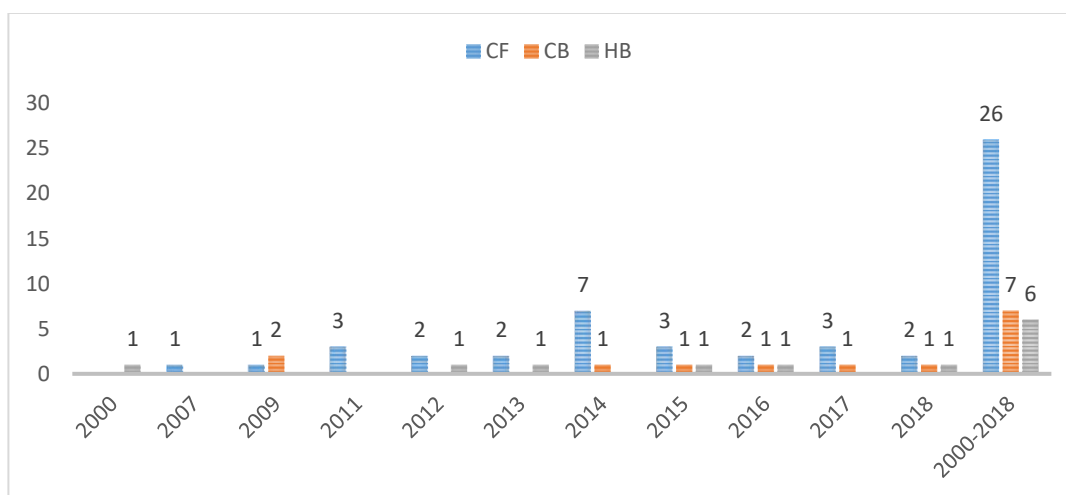
Recommendation Methods

Based on the methods of recommender systems used, book recommendation follows the broad classification of recommender systems, which includes three categories, namely, Collaborative filtering (CF), Content-based filtering (CB) and Hybrid-filtering (HF). The distribution of selected papers on recommendation methods is showed in Table 6.
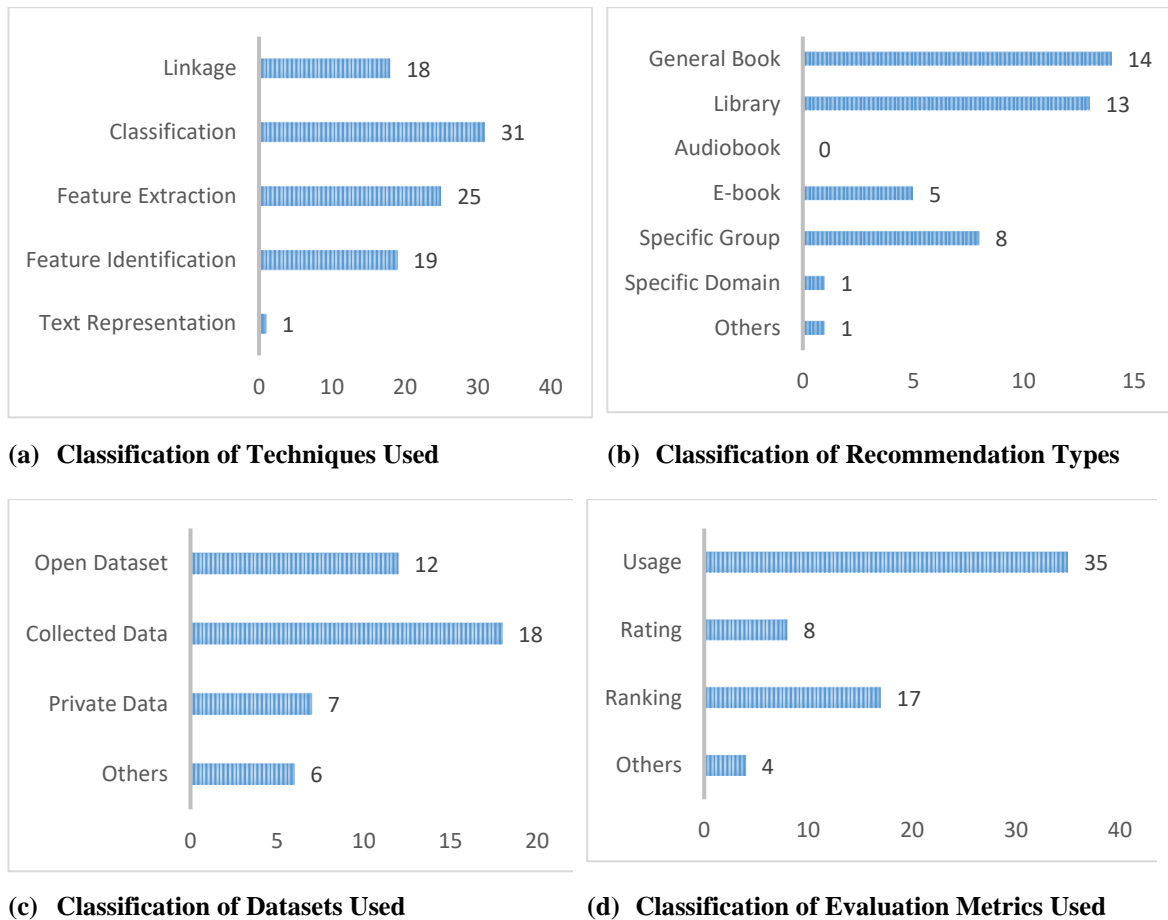
Collaborative Filtering Approaches

Collaborative filtering is widely used in book recommendation, nearly 67% papers (26/39) in our research adopting this method (Pera & Ng (2013); Cao (2011); Yuan et al.(2011); Liu et al.(2014); Dey & Fahrnberger (2017); Wu et al.(2012); et al.). CF uses the available ratings of active users to predict other users' preferences. It can be divided into two forms: user-based and item-based CF. User-based CF finds the similarity between users. A user will receive recommendation of items liked by similar users. Item-based CF computes the similarity between two co-rated items, rated by common users (Sarwar et al., 2001). A user will receive recommendations of items that are similar to those he or she was interested in the past. In this case, an item neighborhood is exploited.

To make recommendations, CF only requires an item-user rating matrix, so it is simple to develop. The rating matrix, however, can be sparse, which may lead to poor recommendations. The shortcoming of CF is obvious as well. Sparsity of rating matrix causes cold start problem, in the case of new items or new users. Grey sheep problem and shilling attack problem are making the CF facing more challenges, which also could be the new research field for book recommendation and recommender systems.

Figure 1: Classification of Methods Used in Book Recommendation

Source: drawn by the authors.

(a) Classification of Techniques Used

(b) Classification of Recommendation Types

(c) Classification of Datasets Used

(d) Classification of Evaluation Metrics Used

Figure 2: Distribution Using the Classification Framework.

Source: drawn by the authors.

Content-Based Filtering Approaches

Out of the tremendous amount of information in a book, CB is introduced in order to identify the most appropriate book to users. In content-based recommendation, the content of items — e.g., the genre and author of a book—is represented as a bag of words, a vector space model, or an ontology (i.e., a class of domain knowledge connected by relations) (Lops et al., 2011). A CB recommendation is a classifier that learns the patterns and similarities in the purchase history of a user to predict his or her future interests. A book's content might refer to its title, summary, outline, whole text or metadata, including author, year of publication, publisher, genre, page number, and so on.

CB recommendation is useful for book recommendation, for it based solely on the content of books instead of users' raring information. It avoids the cold start problem to a great extent. However, when it comes to inadequate descriptions of books, it experiences difficulties. More information processing techniques are required to improve the combative effects of book recommendation. Different approaches are explored to better describe books and build connection between books and users (Pera & Ng (2000); Shi et al. (2009); Nishioka & Ogata (2018); Cui & Chen (2009); et al.).

Hybrid-Based Filtering Approaches

For the sake of improving recommendation effect, i.e. more personalizing and more accurate, different recommendation methods are combined and employed, forming a hybrid model. HF

combines the advantages of both collaborative, content-based filtering and other methods, which can have an integrated outcome and avoid their individual limitations at the same time (Balabanovic & Shoham, 1997). There are different possible ways of combining CF and CB methods into a hybrid model/system (Akhil & Joseph, 2017).

In Woodruff et al. (2000), the authors explored the application of spreading activation over text and citation data to generate hybrid recommendation. Vaz et al. (2012) studied book and author recommendations in a hybrid recommendation combining two item-based collaborative filtering algorithms to predict books and authors that the user will like. To make hybrid recommendation now becomes the trend for enhancing the final recommendation effect. More techniques and methods are introduced to explore the optimized combination effect.

Recommendation Types

In the context of recommender systems, we category the book recommendation, according to the different types of books recommended, into seven categories, i.e. general books, libraries, audiobooks, e-books, books aimed for specific group, books of specific domain and others, showed in Table 6. Book recommendation developed for different types, out of difference in presentation and dissemination of content, are described in following sections.

General Books

Though items that recommended to users can be various, for example movies, news, music, books and so forth, the recommender system is usually generic, not restricted to one particular type. Amazon employs the same recommendation technique to work with all their products (Linden et al. 2003). The situation is identical to the book recommendation, though there are seven kinds of books listed in this paper, different types share the similar book recommendation method and techniques. Most types of the book recommendation are developed for general use, which is not limited to one specific type of books. There is a clear concentration on the general books according to our research, 14 out of 39 papers designing for it. Techniques developed for book recommendation can be extended to different kind of products, such as movies, music and so on.

Libraries

Been next only to general books, libraries employ recommender systems widely quite early. Owning abundant resources, a great amount of circulation records and readers' background information, libraries, including public, school and online digital libraries, keeps an extensive research field. Libraries avoid the data scarcity problem, which is a great advantage for developing and applying book recommendation system. According to our research, nearly a third of papers focused on book recommendation in library (Vaz et al.(2012); Tsuji et al.(2014); Jomsri (2018); Wallace (2016); Akbulut (2017); Yuan (2011); Liu (2014); Dey and Fahrnberger (2017); et al.).

In Tsuji et al. (2014) used library loan records and information about book contents, based on support vector machine (SVM), to recommend books to subjects, and asked them for evaluations of the recommendations that were given. In Jomsri (2018) presented a book recommendation system for university libraries to support user interests which are related in the same topic and faculty.

Audiobooks

Audiobooks are the fastest growing segment in the digital publishing industry. The United States continues to be the biggest market for the audio format and in 2017; there was over $2.5

billion dollars in sales, which is a slight increase from the $2.1 billion generated in 2016. Michelle Cobb of the Audiobook Publishers Association stated, "26% of the US population has listened to an audiobook in the last 12 months[1]. The boom in audiobook production has provided a huge rate of return and has put personal audiobook recommendation to users in great demand. Though it could have been a research area of great meaning and potentiality, there are few papers focusing on it yet. None of our selected papers involved audiobooks, neither.

E-books

Electronics and mobility influence every step of a consumer's life, including media consumption as well. The book industry is no exception, which is reflected in the e-book usage being on the rise. The forecasted 13-percent growth in sales between 2013 and 2018 is also expected to result in revenue gains. Within the same period, sources predict revenues from e-book sales in the United States alone to increase by more than four billion U.S. dollars, reaching nearly 8.7 billion by the end of 2018[2]. In the past decade, millions of books have been digitized in collaborations between libraries and companies like Amazon and Google, e.g., through the Million Book project, and made available online, e.g., through the Internet Archive. At the same time, e-books and eReaders have gained wide acceptance and popularity, their sales skyrocketing (Kazai et al., 2012).

In Nishioka & Hiroaki (2018), a research paper recommender system for students on the e-book system was introduced, which connected research and education in universities. In Nez-Valdez et al. (2015), the authors proposed an architecture to build a content recommendation platform based on e-book reading user behavior, allowing users to learn about the digital content collaboratively. E-books enable researchers an access to process and analyze the full text of a book, and E-readers' reading records. It keeps a rising direction for book recommendation.

Books Aimed for Specific Groups

Recommending books for specific group of people is the same way as setting target audience in publication. Based on our research, there has formed some specific group in book recommendation, for example K-3 and K-12. In Pera & Ng (2012, 2013, 2014, 2015), the authors have developed a serial of book recommenders, which applied approaches to suggest books that simultaneously match the interests/preferences and reading abilities of K-3 or K-12 readers. In Pera (2014), besides assisting K-12 readers, the recommender can be used by parents/teachers/librarians in locating reading materials to be suggested to their (K-12) children/students/patrons.

Books for Specific Domains

Besides aimed at specific reader groups, book recommendation also has its specific domains to target. In Shi et al. (2009), the authors presented a recommender framework in the knowledge domain of Chinese Medicine, to give the diverse semantic recommended medicine terminologies and book pages when a reader searching for medicine information in digital library.

Recommendation designed for specific domain is facing opportunity. While little academic dedication has made on it, the demand of constructing book recommendation is on rise.

---

[1] https://goodereader.com/blog/audiobooks/global-audiobook-trends-and-statistics-for-2018
[2] https://www.statista.com/statistics/234106/e-book-market-share-worldwide/

Others

Recommendation for book usually comes with recommendation of other objects, e.g. papers, related information, and study materials.

Techniques Used in Book Recommendation

Technique is critical to book recommendation. Five categories of classification of algorithms are employed for book recommendation. Techniques used in book recommendation are identical with those used in recommender systems, following the process of data collection, text representation, feature identification & extraction, classification and linkage. Distribution of algorithms used in book recommendation is displayed in Table 3 and Table 6.

Word2Vector is commonly used for text representation. There also are other algorithms, like Tree2Vector showed in Zhang (2017). Word2vec is a group of related models that are used to produce word embeddings. Word2vec and Tree2Vector take as their input a large corpus of text and produces a vector space, typically of several hundred dimensions.

Feature Identification is the foundation for feature extraction and analysis. Frequently used algorithms include: Citation Analysis, Knowledge Graph, Self-organizing Map, Semantic Analysis, Sentiment Analysis, Citation Analysis, Spatio-Textual Analysis, Spreading Activation, Tagging, Word Inbedding and so on. According to our research, semantic analysis is mostly used to realize the function of feature identification from book texts.

After features of text are identified, the next is feature extraction. Algorithms of feature extraction are composed of Topic Modeling, Latent Dirichlet Allocation, Latent Factor Model, Matrix Factorization, TF-IDF, Word Embedding, Regression, Principle Component Analysis et al. Given the widely usage of CF, which is based on rating matrix, matrix factorization is ranked the first as the algorithm used for feature and topic extraction.

Classification is the central part of a recommender system. Algorithms of classification include Decision Tree, Naïve Bayes, Neural Network, K-Nearest, SVM, Association Rule, Boosting and so forth. Each of them has its advantages for specific objective, as a result there does not exist a common concentration of one algorithm

Table 3: Distribution of Algorithms Used in Book Recommendation

Techniques	Algorithms	Abbreviation	Papers
Text Representation	Tree2Vector,	T2V	1
	Word2Vector	W2V	0
Feature Identification	Citation Analysis	CA	1
	Knowledge Graph	KG	3
	Self-organizing Map	SM	1
	Semantic Analysis	SeA	9
	Spreading Activation	SpA	1
	Tagging	Ta.	3
	Word Enbedding	WE	1
Feature Extraction	Clustering	Cl.	5
	Topic Modeling	TM	3
	Latent Dirichlet Allocation	LDA	2
	Latent Factor Model	LFM	2
	Matrix Factorization	MF	8

	TF-IDF	T-I	2
	Word Embedding	WE	1
	Regression	Re.	2
	Principle Component Analysis	CPA	3
Classification	Naïve Bayes,	NB	3
	Neural Network	NN	1
	K-Nearest	KN	1
	Support Vector Machine	SVM	2
	Association Rule	AR	2
	Boosting	Bo.	4
	Similarity Computation	SC	28
	Link Prediction	LP	1
	Ranking	Ra.	17

Source: drawn by the authors.

Linkage is the critical process for matching recommendation. Common algorithms are Similarity Computation, Link Prediction, Complex Network, Social Network, Ranking and so on. Similarity computation and ranking are frequently employed among selected papers.

Datasets Used in Book Recommendation
Without data, any experiment on book recommendation is impossible. Data becomes more and more important in the research area of book recommendation. The use of dataset is presented in Table 1. It shows that nearly a quarter of papers employed open datasets, which is free to download for academic use. Free data resources embody the open spirit of Internet, encouraging sharing and promoting the development of recommender systems together. Open datasets are composed of Book-Crossing dataset, LitRec, LibraryThing, Amazon Reviews, INEX Book Track and so forth. These open datasets provide both the information of items and review matrix made by users. Open dataset is useful in context of recommendation traced from Netfilix Competition. Features of notable datasets are compared and displayed in Table 4 and Table 7.

Table 4: Open Datasets of Book Recommendation

Datasets	Category	Number of users	Number of books	Number of rating matrix
BookCrossing	Free	278,858	271,379	1,149,780
LibraryThing	Free	2,350,000	130 million	-
LitRec	Free	1,927	3,710	38,591
Amazon Reviews	Free	20.98 million	9.35 million	82.83 million
INEX Book Track	Free	94,000	2.8 million	-

Source: drawn by the authors.

As to collected data, which are the most commonly employed resources among research works. Our research shows that 46% (18/39) papers deployed the collected data from Internet, for the sake of pursuing their particular research objectives. Data collected from Internet resources are not as easy as that downloaded from open datasets. Researchers have to crawl and process raw data from target websites--Gutenberg, Goodreads, BiblioNasium, OpenLibrary, CLCD,

Penguin Group, eInkPlusPlus Reader, Douban, ISBNdb and so on. Data used for book recommendation are composed of book information, users' information, and users' review matrix, can be collected from not only one website according to different research direction. Book information includes both public information (e.g. author, title, date, press) and text information (e.g. abstract or full content).

Expect for crawling from open websites, private data comes from researchers' personal collection, or institution owned dataset. In Woodruff (2000), the authors extracted text from 653 of the 719 documents. Jomsri (2018) collected documents consist of 126,521 transactions during January 2012 to February 2016. Each record in the historical loaning corpus contains: Book ID, book name, category number, return date, loan date, barcode, type of user and user id, and bibliographic information. In Xin (2013,2014), real world dataset is from some academic library, which contains 82, 987 records form 2011-06-07 to 2012-06-27, including five different operations (borrowing, returning, renewing, booking and removing the booking with the time stamp).

Data obtained neither from open datasets, websites, nor private owned data, is assorted to dataset of others. For example, data that is confidential or collected from secret resource, belongs to the category of others.

Evaluation Metrics Used in Book Recommendation

Evaluation is an indispensable component after conducting experiment on new book recommendation algorithms. In our research, we put the evaluation metrics used in review works into four categories, based on the different aims of evaluation. For metrics that cannot identify the evaluation objective, we assort them into the kind of others. The classification result is showed in Table 5 and Table 7.

In Shani & Gunawardana (2011) and Alharthi (2018), authors achieved a common view on the division of evaluation metrics, to measuring the prediction quality of usage, ratings and ranking. We follow the idea of classification of evaluation metrics on book recommendation.

The quality of prediction usage is to find out whether a reader likes the book recommended to him/her or not. Accuracy stands for the success rate of recommendation. Recall rate and coverage measure the diversity of recommendation. F-measure tests the combined effect of recommendation, making a balance of precision and diversity.

For collaborative filtering, to measure rating effect is extremely necessary. Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used for testing the difference between predicted and actual numerical rating.

Ranking is to give a list of recommended books. Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain (NDCG) and Expected Reciprocal Rank (ERR) are frequently employed to view the effect of ranking.

Our reviewed papers show the preference for accuracy, coverage, recall rate, and F-measure, to evaluating the quality of book recommendation. Prediction of recommendation usage is the fundamental evaluation metrics, which gives an overall review on a recommendation. To further evaluate the book recommender system, MAE, RMSE, MRR, NDCG and other metrics are introduced.

Discussion

The results showed in Figure 1 and Table 2 illustrate the tendencies regarding the research on book recommendation. In order to make personal book recommendation, previous readers' review is indispensable reference resource, so CF is the uppermost recommender method for book recommendation. Given the rich features of books, tremendous text information of books requires nature language process (NLP) and more complex techniques, content-based and hybrid-based are also widely used to make personal book recommendation. These two methods are forming their research fields and cannot be neglected.

Based on papers we reviewed, it is also observed that contrary to the rise sale of e-book and audiobook, research works on them are still rare, especially to audiobook. It may be limited by the audio processing technology, which would be a future focus worthy of dedication.

Data collecting is the main task for testing the effect of book recommendation, different data resources often result in different evaluation ways, which may lead to difficulty of measuring the real quality of a recommender system. With more researchers employing self-collected or private data, the issue of evaluation will become more noticeable.

The limits of our study are inevitable. First of all, we conduct the analysis mainly on 39 papers selected of book recommendation, the sample size of the study may affect the results. Secondly, though we try to be as carefully as we can during the reviewing process, there would still be some mistakes which may impact the final analysis as well.

Table 5: Classification of Evaluation Metrics of Book Recommendation

Objective	Metrics	Abbreviation	Papers
Usage	Accuracy	Acc.	12
	Coverage	Cov.	4
	Recall	Rec.	8
	F-measure	F	9
	Success Ratio	SR	1
	Hit Ratio	HR	1
Rating	Mean Absolute Error	MAE	2
	Root Mean Squared Error	RMSE	4
	Hamming Loss	HL	1
Ranking	Mean Reciprocal Rank	MRR	7
	Expected reciprocal rank	ERR	1
	Normalized Discounted Cumulative Gain	NDCG	7
	Ranking Loss	RL	1
	Success at rank	S@R	1
Others	Quality of Recommendation	QR	1
	Average Effective Rate	AER	2
	Receiver Operating Characteristic	AUC	1
	One Error	OE	1

Source: drawn by the authors.

Conclusions

This paper presents a literature review of the overall state of book recommendation, with systematically methodology to give a classification framework from five perspectives. Book recommendation is not a new issue, but to realize personal and professional recommendation

to present or potential readers is remaining an urgent problem, which is also a hot area of recommender systems.

Former research works from five academic presses, from 2000 to 2018, are reviewed to identify the latest trend and the integrated situation of book recommendation. A classification framework is constructed in order to comprehensively describe the various methods, techniques, types, datasets and evaluation metrics, which are adopted in book recommendation to improve the combined effect and to meet readers' personal needs. This framework is the outcome of 39 research papers, referring to literature on classification of recommender systems appeared recently. Section 3 gives the distribution of the overall classification framework, showing the trends and direction in this research area. In Section 4, details of this framework, limits and future direction of this research on book recommendation are discussed. The classification framework can also be used as an analysis and study guide for future researchers.

From the distribution based on the classification framework, it shows clearly the aspects of book recommendation as follows:

The most frequently used method is collaborative filtering, and content-base & hybrid-based filtering cannot be neglected. General book and library are recommendation types of most frequency, while books for specific domain and group have significance as well, and e-book and audiobook worth owning more academic attention in the future. Classification algorithms are most commonly used techniques in book recommendation, next to it are techniques of feature identification & extraction, resulting from the multiple texts of characteristics in books. More researchers prefer collecting data from open Internet resources for their particular research objectives, while there are still other resources to obtain data as well, for example open datasets or specific institution. Evaluation metrics of usage precision (e.g. accuracy, recall, F-measure) are most elementary and widely used, however metrics to evaluate the rating and ranking are of equal importance as well.

Finally, this work is intended to make contribution to insisting future researchers by using the proposed classification framework, to providing direction in finding relevant work on the field. Further, based on the classification, a clearer research map of book recommendation is described, which could help researchers to locate their academic position and figure out next research focus.

Table 6: Selected Papers of Book Recommendation on Methods, Types and Algorithms

Author & Year	Methods			Types						Algorithms																								
										Rep.	Feature Identification										Feature Extraction					Classification							Linkage	
	CF	CB	HB	GB	EB	Lib	SD	SG	Ot.	T2V	SeA	CA	Ta.	WE	KG	SpA	SM	Cl.	T-I	TM	LDA	LFM	MF	Re.	PCA	NB	NN	KN	SVM	AR	Bo.	SC	LP	Ra.
Woodruff et al. (2000)	✓				✓						✓	✓				✓			✓							✓						✓		
Basile et al. (2007)			✓																							✓						✓		✓
Cui & Chen (2009)		✓		✓			✓				✓																					✓		✓
Shaomin et al. (2009)		✓			✓						✓																							
Zhen et al. (2009)	✓			✓																												✓		✓
Cao (2011)	✓			✓														✓													✓	✓	✓	
Jin et al. (2011)	✓					✓							✓																			✓		
Yuan et al. (2011)	✓				✓																							✓						
Calado et al. (2012)	✓				✓																	✓				✓					✓			
Pera & Ng (2012)	✓					✓		✓															✓											
Wu et al. (2012)			✓			✓		✓														✓												
Carrillo et al. (2013)	✓					✓															✓											✓		
Pera & Ng (2013)	✓			✓				✓														✓												✓
Xin et al. (2013)			✓	✓										✓													✓							
Garrido et al. (2014)	✓			✓				✓			✓																							
Pera & Ng (2014a)	✓			✓																	✓											✓		
Pera & Ng (2014b)	✓			✓		✓		✓															✓											✓
Soledad (2014)		✓				✓							✓											✓					✓					
Tsuji et al. (2014)	✓					✓			✓									✓				✓										✓		✓
Xin et al. (2014a)	✓					✓		✓													✓		✓									✓		✓
Xin et al. (2014b)	✓			✓																			✓											
Yada (2014)	✓					✓						✓								✓														
Kurmashov et al. (2015)	✓			✓								✓																			✓	✓		✓
Lu & Zhang (2015)	✓		✓		✓																													
Pera & Ng (2015)	✓		✓	✓													✓	✓														✓		
Toine et al. (2015)	✓			✓													✓	✓							✓									✓
Valdez (2015)	✓				✓				✓		✓			✓										✓								✓		
Kumar et al. (2016)		✓		✓																														
Ng (2016)	✓					✓		✓				✓										✓				✓						✓		
Sarquib et al. (2016)	✓			✓											✓																			
Stephen (2016)	✓		✓		✓																		✓											
Akbulut et al. (2017)	✓			✓		✓												✓																
Dey & Fahrmberger (2017)		✓		✓	✓													✓				✓		✓										
Zhang et al. (2017)		✓				✓																	✓	✓										
Zhang et al. (2017)	✓				✓																													
Divanshu et al. (2018)	✓		✓		✓				✓		✓				✓																			
Jomsri (2018)	✓					✓										✓																		
Musto et al. (2018)	✓											✓																			✓			
Nishioka & Ogata (2018)	✓											✓																		✓				✓

Table 7: Selected Papers of Book Recommendation on Databases and Evaluation Metrics

Author & Year	Datasets				Evaluation Metrics																	
	Open Datasets	Collected Data	Private Data	Others	Usage						Rating				Ranking					Others		
					Acc.	Cov.	Rec.	F	SR	HR	MAE	RMSE	HL	AUC	MRR	NDCG	FRR	RL	S@R	QR	AER	OE
Woodruff et al. (2000)			✓			✓																
Basile et al. (2007)		✓																				
Cui & Chen (2009)		✓																			✓	
Shaomin et al. (2009)			✓	✓																		
Zhen et al. (2009)			✓		✓		✓															
Cao (2011)	✓				✓		✓							✓								
Jin et al. (2011)	✓						✓															
Yuan et al. (2011)				✓													✓					
Calado et al. (2012)	✓		✓		✓		✓	✓														
Pera & Ng (2012)		✓																				
Wu et al. (2012)	✓														✓							
Carrillo et al. (2013)		✓			✓			✓							✓	✓						
Pera & Ng (2013)			✓									✓										
Xin et al. (2013)	✓	✓			✓	✓		✓					✓					✓				✓
Garrido et al. (2014)		✓													✓	✓						
Pera & Ng (2014a)			✓	✓					✓													
Pera & Ng (2014b)			✓		✓	✓	✓	✓														
Soledad (2014)		✓			✓	✓										✓						
Tsuji et al. (2014)		✓			✓			✓														
Xin et al. (2014a)		✓			✓		✓	✓							✓	✓						
Xin et al. (2014b)			✓		✓	✓																
Yada (2014)	✓											✓										
Kurmashov et al. (2015)		✓																		✓	✓	
Lu & Zhang (2015)		✓																				
Pera & Ng (2015)	✓										✓				✓	✓						
Toine et al. (2015)	✓														✓	✓			✓			
Valdez (2015)		✓			✓		✓	✓														
Kumar et al. (2016)				✓	✓																	
Ng (2016)	✓														✓							
Saquib et al. (2016)				✓								✓										
Stephen (2016)	✓									✓												
Akbulut et al. (2017)	✓	✓			✓						✓											
Dey & Fahrnberger (2017)				✓											✓							
Zhang et al. (2017)		✓			✓		✓	✓				✓										
Divanshu et al. (2018)	✓	✓			✓		✓				✓											
Jomsri (2018)		✓			✓																	
Musto et al. (2018)			✓		✓			✓				✓										
Nishioka & Ogata (2018)		✓																				

References

Akbulut, A., Catal, C., & Akbulut, F. P. (2017). A cloud-based recommendation service using principle component analysis–scale-invariant feature transform algorithm. Neural Computing and Applications, 28(10), 2859-2868.

Akhil, P. V., & Joseph, S. (2017). A SURVEY OF RECOMMENDER SYSTEM TYPES AND ITS CLASSIFICATION. International Journal of Advanced Research in Computer Science, 8(9).

Alharthi, H., Inkpen, D., & Szpakowicz, S. (2018). A survey of book recommender systems. Journal of Intelligent Information Systems, 51(1), 139-160.

Allington, R. L., & Gabriel, R. E. (2012). Every child, every day. Educational Leadership, 69(6), 10-15.

Balabanović, M., & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3), 66-72.

Basile, P., Tinelli, E., Degemmis, M., Di Noia, T., Semeraro, G., & Di Sciascio, E. (2007, September). Semantic Bayesian profiling services for information recommendation. In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (pp. 711-719). Springer, Berlin, Heidelberg.

Cao, B. (2011). Collaborative and transfer learning in recommendations. Hong Kong University of Science and Technology (People's Republic of China).

Carrillo, D., López, V. F., & Moreno, M. N. (2013). Multi-label classification for recommender systems. In Trends in Practical Applications of Agents and Multiagent Systems (pp. 181-188). Springer, Cham.

Cui, B., & Chen, X. (2009, August). An online book recommendation system based on web service. In Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09. Sixth International Conference on (Vol. 7, pp. 520-524). IEEE.

Deshpande, M., & Karypis, G. (2004). Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS), 22(1), 143-177.

Dey, S. K., & Fahrnberger, G. (2017, March). Reader's Choice. In International Conference on Database Systems for Advanced Applications (pp. 243-255). Springer, Cham.

Garrido, A. L., Pera, M. S., & Ilarri, S. (2014, July). SOLE-R: A semantic and linguistic approach for book recommendations. In 2014 IEEE 14th international conference on advanced learning technologies (ICALT) (pp. 524-528). IEEE.

Goel, A., Khandelwal, D., Mundhra, J., & Tiwari, R. (2018). Intelligent and Integrated Book Recommendation and Best Price Identifier System Using Machine Learning. In Intelligent Engineering Informatics (pp. 397-412). Springer, Singapore.

Hämäläinen, M., Whinston, A. B., & Vishik, S. (1996). Electronic markets for learning: education brokerages on the Internet. Communications of the ACM, 39(6), 51-58.

Jin, S., Fan, C., Meng, Y., Wu, X., & Chen, Q. (2011, September). The design of a new book auto recommendation system based on readers' interest. In Electrical and Control Engineering (ICECE), 2011 International Conference on (pp. 3224-3227). IEEE.

Jomsri, P. (2018). FUCL mining technique for book recommender system in library service. Procedia Manufacturing, 22, 550-557.

Kazai, G., Landoni, M., Eickhoff, C., & Brusilovsky, P. (2012, October). Booksonline'12: 5th workshop on online books, complementary social media and their impact. In Proceedings of the 21st ACM international conference on Information and knowledge management (pp. 2764-2765). ACM.

Koolen, M., Bogers, T., Van Den Bosch, A., & Kamps, J. (2015, March). Looking for books in social media: An analysis of complex search requests. In European Conference on Information Retrieval (pp. 184-196). Springer, Cham.

Kurmashov, N., Latuta, K., & Nussipbekov, A. (2015, September). Online book recommendation system. In Electronics Computer and Computation (ICECCO), 2015 Twelve International Conference on (pp. 1-4). IEEE.

Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet computing, (1), 76-80.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In Recommender systems handbook (pp. 73-105). Springer, Boston, MA.

Lu, L., & Zhang, H. (2015, July). A tree-structured representation for book author and its recommendation using multilayer SOM. In Neural Networks (IJCNN), 2015 International Joint Conference on (pp. 1-8). IEEE.

Miranda, T., Claypool, M., Gokhale, A., Mir, T., Murnikov, P., Netes, D., & Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. In In Proceedings of ACM SIGIR Workshop on Recommender Systems.

Musto, C., Narducci, F., Lops, P., de Gemmis, M., & Semeraro, G. (2019). Linked open data-based explanations for transparent recommender systems. International Journal of Human-Computer Studies, 121, 93-107.

Ng, Y. K. (2016, July). Recommending books for children based on the collaborative and content-based filtering approaches. In International Conference on Computational Science and Its Applications (pp. 302-317). Springer, Cham.

Ng, Y. K. . (2016). Recommending Books for Children Based on the Collaborative and Content-Based Filtering Approaches. Computational Science and Its Applications -- ICCSA 2016. Springer International Publishing.

Nishioka, C., & Ogata, H. (2018). Research Paper Recommender System for University Students on the E-Book System. In proceedings of the 18th acm/ieee on joint conference on digital libraries (pp. 369-370). ACM.

Núñez-Valdez, E. R., Lovelle, J. M. C., Hernández, G. I., Fuente, A. J., & Labra-Gayo, J. E. (2015). Creating recommendations on electronic books: A collaborative learning implicit approach. Computers in Human Behavior, 51, 1320-1330.

Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. Expert Systems with Applications, 39(11), 10059-10072.

Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. Artificial intelligence review, 13(5-6), 393-408.

Pera, M. S. (2014). Using online data sources to make recommendations on reading materials for k-12 and advanced readers. Dissertations & Theses - Gradworks.

Pera, M. S., & Ng, Y. K. (2012, October). Personalized recommendations on books for K-12 readers. In Proceedings of the fifth ACM workshop on Research advances in large digital book repositories and complementary media (pp. 11-12). ACM.

Pera, M. S., & Ng, Y. K. (2013, October). What to read next?: making personalized book recommendations for K-12 users. In Proceedings of the 7th ACM conference on Recommender systems (pp. 113-120). ACM.

Pera, M. S., & Ng, Y. K. (2014, August). How can we help our K-12 teachers?: using a recommender to make personalized book suggestions. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02 (pp. 335-342). IEEE Computer Society.

Pera, M. S., & Ng, Y. K. (2014, October). Automating readers' advisory to make book recommendations for k-12 readers. In Proceedings of the 8th ACM Conference on Recommender systems (pp. 9-16). ACM.

Pera, M. S., & Ng, Y. K. (2015, August). Analyzing book-related features to recommend books for emergent readers. In Proceedings of the 26th ACM Conference on Hypertext & Social Media (pp. 221-230). ACM.

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: introduction and challenges. In Recommender systems handbook (pp. 1-34). Springer, Boston, MA.

Sarwar, B. M. (2001). Sparsity, scalability, and distribution in recommender systems. University of Minnesota.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Application of dimensionality reduction in recommender system-a case study (No. TR-00-043). Minnesota Univ Minneapolis Dept of Computer Science.

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

Shi, S., Wei, B., & Yan, Y. (2009, November). Msuggest: a semantic recommender framework for traditional chinese medicine book search engine. In Proceedings of the 18th ACM conference on Information and knowledge management (pp. 533-542). ACM.

Shrivastva, K. M. P., & Singh, S. (2016). Cross domain recommendation using semantic similarity and tensor decomposition. Procedia Computer Science, 85, 317-324.

Sohail, S. S., Siddiqui, J., & Ali, R. (2016, September). Book recommender system using fuzzy linguistic quantifier and opinion mining. In The International Symposium on Intelligent Systems Technologies and Applications (pp. 573-583). Springer, Cham.

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009.

Terán, L., Mensah, A. O., & Estorelli, A. (2018). A literature review for recommender systems techniques used in microblogs. Expert Systems with Applications, 103, 63-73.

Tsuji, K., Takizawa, N., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2014). Book recommendation based on library loan records and bibliographic information. Procedia-social and behavioral sciences, 147, 478-486.

Vaz, P. C., Martins de Matos, D., Martins, B., & Calado, P. (2012, June). Improving a hybrid literary book recommendation system through author ranking. In Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries (pp. 387-388). ACM.

Wallace, S. (2016). Extending a Business Performance Improvement Framework with a Semantic Enhanced Hybrid Recommender System Design. Pace University.

Woodruff, A., Gossweiler, R., Pitkow, J., Chi, E. H., & Card, S. K. (2000, April). Enhancing a digital book with a reading recommender. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 153-160). ACM.

Wu, D., Yuan, Z., Yu, K., & Pan, H. (2012, November). Temporal social tagging based collaborative filtering recommender for digital library. In International Conference on Asian Digital Libraries (pp. 199-208). Springer, Berlin, Heidelberg.

Xin, L., Haihong, E., Junde, S., Meina, S., & Junjie, T. (2013, December). Collaborative Book Recommendation Based on Readers' Borrowing Records. In Advanced Cloud and Big Data (CBD), 2013 International Conference on (pp. 159-163). IEEE.

Xin, L., Haihong, E., Junjie, T., Meina, S., & Yi, L. (2014, May). Enhancing Book Recommendation with Side Information. In Service Sciences (ICSS), 2014 International Conference on (pp. 142-146). IEEE.

Xin, L., Haihong, E., Song, J., Song, M., & Tong, J. (2013, December). Book recommendation based on community detection. In Joint International Conference on Pervasive Computing and the Networked World (pp. 364-373). Springer, Cham.

Yada, S. (2014, November). Development of a book recommendation system to inspire "infrequent readers". In International Conference on Asian Digital Libraries (pp. 399-404). Springer, Cham.

Yuan, Z., Yu, T., & Zhang, J. (2011, October). A social tagging based collaborative filtering recommendation algorithm for digital library. In International Conference on Asian Digital Libraries (pp. 192-201). Springer, Berlin, Heidelberg.

Zhang, H., Wang, S., Wang, E. K., Li, Y., Zhang, Y., & Chu, D. (2017, July). Recommending e-books by multi-layer clustering and locality reconstruction. In Industrial Informatics (INDIN), 2017 IEEE 15th International Conference on (pp. 1056-1061). IEEE.

Zhang, Q., Wu, D., Lu, J., Liu, F., & Zhang, G. (2017). A cross-domain recommender system with consistent information transfer. Decision Support Systems, 104, 49-63.

Zhen, L., Huang, G. Q., & Jiang, Z. (2009). Collaborative filtering based on workflow space. Expert Systems with Applications, 36(4), 7873-7881.