

# A survey on data mining techniques in recommender systems

Maryam Khanian Najafabadi<sup>1</sup> · Azlinah Hj. Mohamed<sup>1</sup> · Mohd Naz'ri Mahrin<sup>2</sup>

Published online: 7 November 2017  
© Springer-Verlag GmbH Germany, part of Springer Nature 2017

**Abstract** Recommender systems have been regarded as gaining a more significant role with the emergence of the first research article on collaborative filtering (CF) in the mid-1990s. CF predicts the interests of an active user based on the opinions of users with similar interests. To extract information on the preference of users for a set of items and evaluate the performance of the recommender system's techniques and algorithms, a critical analysis can be conducted. This study therefore employs a critical analysis on 131 articles in CF area from 36 journals published between the years 2010 and 2016. This analysis seems to be the exclusive survey which supports and motivates the community of researchers and practitioners. It is done by using the applications of users' activities and intelligence computing and data mining techniques on CF recommendation systems. In addition, it provides a classification of the literature on academic database according to the benchmark recommendation databases, two users' feedbacks (explicit and implicit feedbacks) which reflect their activities and categories of intelligence computing and data mining techniques. Eventually, this study provides a road map to guide future direction on recommender systems research and facilitates the accumulated and derived knowledge on the application

of intelligence computing and data mining techniques in CF recommendation systems.

**Keywords** Collaborative recommendation systems · Public datasets · Survey · User feedback · Intelligence computing · Data mining techniques

## 1 Introduction

The recommender system is a useful tool that assists users by providing them with choices of appropriate items according to their interests. This system has been used in various genres like music, books, movies, jokes, news articles and others. It is used to deduce information about users' interactions with items and preference between users and items so that recommendations can be done personally according to users' wants and needs (Feng et al. 2015; Choi and Suh 2013). In the world of consumer recommender systems, researchers require benchmark recommendation datasets to evaluate their technique and demonstrate experimental results. In requiring this, a number of public recommendation datasets that have already been released are employed to help researchers to extract information on the preference of users for a set of items and evaluate the performance of different techniques and algorithms proposed in recommender systems. The users' ratings are monitored, and their activities such as the kind of the music they listened to, books that they read and Web sites they visit are analyzed implicitly in the datasets. These are done in order to develop new techniques and to evaluate them against the existing recommendation techniques (Li and Chen 2013; Langseth and Nielsen 2015).

Collaborative filtering (CF) is one of the most popular recommendation techniques which achieve efficiency from the similar measurement of users and items. The following

Communicated by V. Loia.

✉ Maryam Khanian Najafabadi  
maryam64266@yahoo.com; maryamkn@uitm.edu.my  
Azlinah Hj. Mohamed  
azlinah@tmsk.uitm.edu.my

<sup>1</sup> Advanced Analytics Engineering Centre (AAEC), Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia

<sup>2</sup> Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia

four steps describe the trend of CF technique (Bauer and Nanopoulos 2014; Najafabadi and Mahrin 2016):

- i. Collect user's ratings of available items (e.g., movies, CDs or books) in user profile (user rating database) in order to show the users' preferences in the corresponding domain.
- ii. Identify a set of users (known as neighbors) who are similar to the active users. CF evaluates the similarity between users based on their ratings on common items in the user profile. For example, either they have given similar rating on available items or they have used similar items.
- iii. Predict the rating on products that active users would give by observing the ratings of neighbors of active users. Notice that, when trying to predict the rating on a specific item, there will be many vacant ratings of the product in neighborhood of active users. In other words, a significant number of neighbors might have not rated the product; therefore, mechanisms should be developed that enable ratings to be predicted on products based on minimum number of ratings.
- iv. Find products that the active user is interested in based on interests of like-minded users.

Unfortunately, CF may lead to poor recommendation when users' ratings on items are very sparse in comparison with the huge number of users and items in user-item matrix (data sparsity problem). In the case of a lack of user rating on items, implicit feedback based on users' activities is used to profile a user's item preferences. Implicit feedback can indicate users' preferences by providing more evidences and information through observations made on users' behaviors. So, this paper is aimed at giving a thorough review on the CF-based recommender systems that are assessed using the public databases. It summarizes the techniques used to improve the user feedback (explicit and implicit feedbacks) and also summarizes the existing recommender techniques. Therefore, this paper helps researchers who are interested in conducting further research in CF-based recommender systems.

There are literature reviews on recommender systems that have been published. However, these articles focus only on a specific domain of recommendation systems development or recommendation approaches. None of these articles concentrate on the comprehensive analysis of recommendation datasets with their specific characteristics as common benchmarks that form the base for researchers. For example, the study done by Adomavicius and Tuzhilin (2005) provided an overview of recommendation approaches including collaborative filtering, content-based and hybrid approaches. They presented the limitations of these recommendation approaches and described possible extensions that could enhance performance of recommender systems. Bobadilla

et al. (2013a,b) presented an overview of group recommendation techniques, fundamental recommendation, social filtering and recently developed techniques including the bio-inspired and location-aware recommendation techniques. Park et al. (2012) cluster 210 articles on recommender system areas by their application fields, year of publication, the journal and data mining techniques. Burke (2002) evaluated the landscape of actual hybrid recommendation techniques. Burke also provided a review on hybridization methods in recommender systems.

It has been identified that there is no study so far that has been published or written that reviews collaborative recommender systems from the requirements of current public databases. It is hoped that this research provides the knowledge on recommender system datasets for researchers in order to select the most suitable datasets according to the research purpose. Hence, this study describes the artificial intelligence and data mining techniques that could improve CF recommendation capabilities which focus on improving users' feedback. It distributes the selected studies by current public databases, by year of publication, by assorted recommendation approaches/techniques in a range of application fields and by historical record of users' activities used. This study provides valuable insights and acts as a guide for industrial practitioners and researchers. The main contributions of this study are as follows:

- (1) This research analyzes the studies on collaborative recommender systems from the requirements of current public databases, which show the distribution of articles in public application domains including movies, Web pages, books, music and jokes;
- (2) Most importantly, this research systematically examines and classifies the studies conducted by current public databases, which provide a list of the historical records of users' activities referenced most often in the literature;
- (3) This research classifies the articles by their year of publication, by assorted recommendation approaches/techniques in a range of current recommendation dataset and by historical records of users' activities used.
- (4) For each application domain, it analyzes the research achievements on CF recommendation systems and effectively classifies the studies conducted by techniques used in the domain. This will directly support and motivate the practitioners and researchers with the application of recommendation techniques in different domains and provide them with a scheme of recommendation techniques.

The report of this review is organized according to the following sections: A summary of recommender system datasets is depicted in Sect. 2. Section 3 presents the research method-

ology involved for conducting critical review on existing research articles which have employed CF recommender systems. Section 4 provides a method for classifying articles based on recommendation techniques and public databases used. The findings of this classification will be discussed in this section. Section 5 provides significant implications of the results. Finally, the conclusion is provided in Sect. 6.

## 2 Summary of public recommendation datasets

This paper thoroughly reviews the collaborative filtering-based recommender systems evaluated using public databases. A number of publicly available recommendation datasets have been employed, which help researchers to extract information on the preference of users for a set of items and evaluate the performance of their proposed recommendation techniques. At the beginning of testing newly developed recommendation techniques, the following preprocesses are conducted. This is a widely used preprocessing technique in the area of evaluating the recommender systems (Horsburgh et al. 2015; Briguez et al. 2014; Najafabadi et al. 2017).

- (1) The first task is to select the datasets. The experiment will be conducted using a public dataset which is a freely accessed database. This experiment includes separating the dataset into training and test portions. The training portion of the dataset is used to learn different experimental parameters (80% of the dataset). The test portion (20% of the dataset) is used to evaluate the training technique. The researchers can replicate experiments to investigate and improve their techniques.
- (2) The second task is to identify a suitable evaluation metric. An evaluation metric is important for the experimental verification process. In order to claim how well our technique works (improved predictive utility), we employ a metric that is well understood, widely used and reproducible. The evaluation metric measures the quality of a recommender system by measuring how close a recommender system's predicted ranking of items for a user differs from the user's true ranking of preference. Also, it should measure how well a system can predict an exact rating value for a specific item or evaluate the accuracy of the proposed technique by measuring ratio of recommended items that are relevant.

Table 1 depicts the information in work proposed by Najafabadi and Mahrin (2016) and lists the characteristics of each public dataset. These datasets have diverse meta-information and various types of users' preferences according to five domains which are movies, jokes, books, Web pages and music tracks. Delicious and MSD incorporate implicit feedback and social information (grasp users'

**Table 1** Description of the public recommender system datasets

Data	Without social information				With social information		
	MovieLens 100 k	MovieLens 1 M	MovieLens 10M	Netflix	Jester	BookCrossing	Delicious MSD
Ratings	100,000	1 million	10 million	100 million	4.1 million	1.1 million	104,833
Users	943	6,040	71,567	480,000	73,496	278,858	1867
Items	1,682	3,592	10,681	17,770	100	271,379	69,226
Range	{1, ..., 5}	{1, ..., 5}	{1, ..., 5}	{1, ..., 5}	-10, +10	{1, ..., 10}	Implicit
Tags	N/A	N/A	13,222	N/A	N/A	N/A	Implicit
Items	Movies	Movies	Movies	Movies	Jokes	Books	URLs
							Music

interests from social networking activities such as tagging, making social comments and music listening information of the users to boost the recommendation (Huang et al. 2014)). Social tags on items are valuable implicit sources of information about the contents associated with the items to represent the users' interests and preferences. Therefore, tagging information can be used to enrich item profiles and user profiles in improving the recommendations. In MSD, songs can associate or link with other data in sibling datasets such as audio features, artist data, song tags, play count, lyrics, etc. The MSD is a cluster of complementary dataset contributed by Second Hand Songs dataset, MusicXmatch dataset, Last.fm dataset and Taste profile subset.

### 3 Research methodology

The purpose of this report is to figure out applications of user feedbacks, artificial intelligence and data mining techniques on CF recommendation systems. This is done by examining the published articles and affording the community of researchers and practitioners with insight and future direction on CF recommender systems. Hence, this study provides an academic database of the literature between the periods of 2010–2016 by covering 36 journals and proposes a classification scheme according to recommendation databases, user feedbacks which reflected users' activities and intelligence computing and data mining algorithms to classify the published articles.

In this study, the ways to improve the user feedbacks and an overview of state-of-the-art techniques in recommender systems, particularly the scarcity of historical record of users' activities applied from public datasets, will be discussed. Research articles associated with the descriptors "collaborative filtering," "sparsity problem" and "available recommendation data-sets" were carefully selected as follows:

#### i. Step 1: Identification of academic databases

In order to present a comprehensive bibliography of articles on collaborative recommendation systems, these articles were searched from academic databases including IEEE Xplore, Sage, ScienceDirect and ACM Library.

#### ii. Step 2: Preliminary screening of research articles

The search was first conducted based on eight descriptors: "collaborative filtering," "sparsity problem" and available recommendation datasets including "Jester," "MovieLens," "Netflix," "Delicious," "Book-Crossing" and "Million Song datasets." These datasets have been chosen, because they are available datasets on Web and the most used datasets by developers and researchers in CF technique domain.

#### iii. Step 3: Result filtering for articles

The research papers were selected as references if they had satisfied the following criteria: (1) Publication time: Only the papers that had been published between 2010 and 2016 were selected. This was to ensure that the data gathered were still fresh and up to date. (2) Published in high-quality publication: To gain the highest level of related articles, only articles published by academic journals were selected because they were reliable and worthy of comment, since they were peer-reviewed publication. Thus, conference articles, textbooks, unpublished articles, master and doctoral dissertations, non-English papers, and notes were excluded in this study. Following this filtering process, selected research papers were used as the preliminary references for this study.

#### iv Step 4: Research paper selection

Lastly, the full text of each publication and articles that were not related to collaborative recommender systems in solving the sparsity problem were omitted. Ultimately, a total of 131 articles from 36 journals were selected as the final reference list for this research. Figure 1 depicts the research methodology of this research review and the defined criteria for conducting the review.

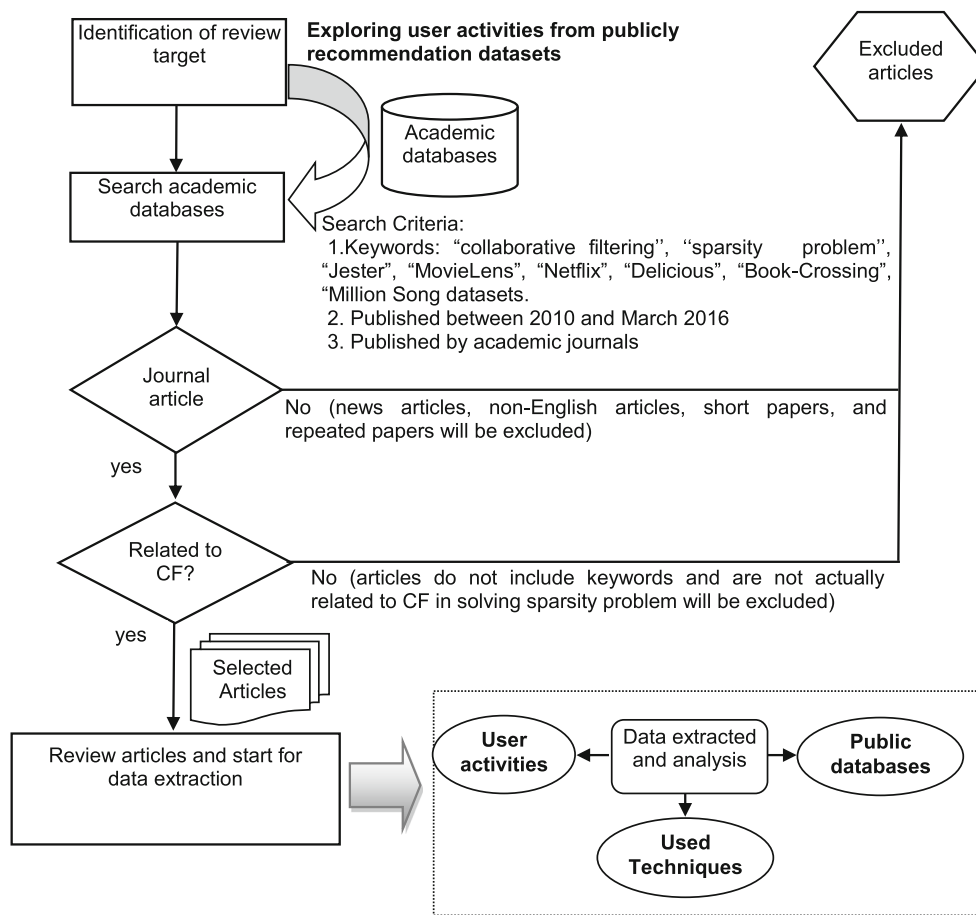
This study provides a road map for future work direction on recommender systems research by classifying the academic database of literature. It facilitates accumulated and derived knowledge on the application of user feedbacks, artificial intelligence and data mining techniques in CF recommendation systems.

### 4 Proposing a classification method

The reviewed articles were categorized into six categories of recommendation databases, two users' feedbacks which reflected users' activities and three main categories of techniques used in improving CF. The overall classification scheme for collaborative recommendation articles is presented in Fig. 2. To the best of our knowledge, there has been no research work conducted to comprehensively review collaborative recommendation articles by considering employed public datasets, users' feedbacks and techniques used in recommender systems.

#### 4.1 Analyzed users' activities from each public database

The selection of input data (such as user and item characteristics, implicit browsing or buying or clicking activities and explicit ratings) for improvement technique in a recommender system plays an important role in ensuring the quality of recommendations. This section aims to investi-



**Fig. 1** Research methodology of research review

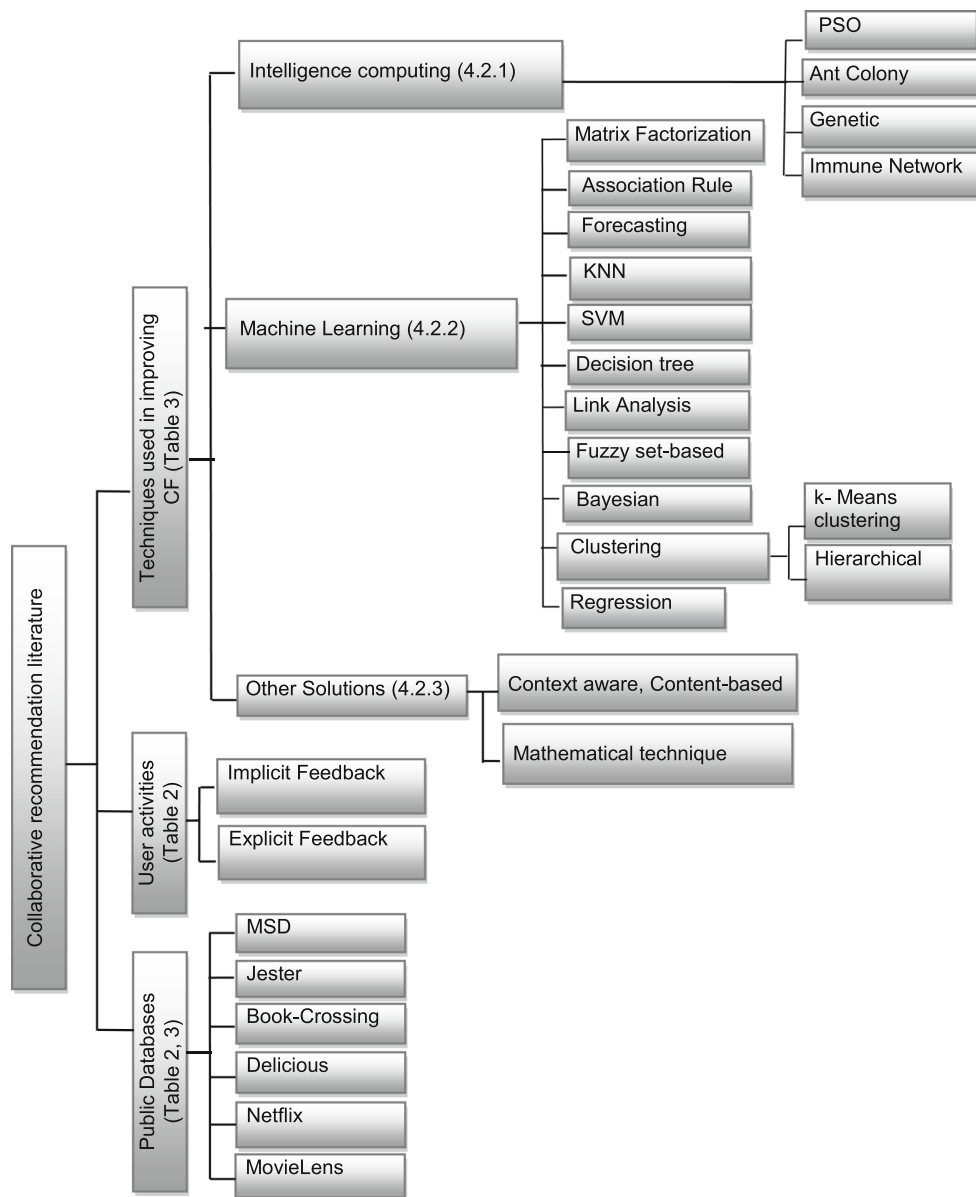
gate how information on user's interests can be gained by collecting the user activities including the implicit data and explicit ratings. Hence, it is meaningful that each of selected articles is reviewed and classified by considering the public datasets and the user feedbacks with product attributes and user attributes from public datasets are explored. In other words, recommendation techniques are usually compared on the base of publicly available dataset of movies, songs, jokes, Web pages, books that contain the description of user interests on a set of items and input data for recommender systems (Kim and El Saddik 2015; Liu et al. 2013; Hsiao et al. 2014).

Table 2 lists the publicly available datasets (first column of Table), user activities (second column) and user feedback applied (third column) to recommender systems. The public databases have found real-world applications in movie recommendations in MovieLens and Netflix (Feng et al. 2015; Pirasteh et al. 2015), book recommendations in BookCrossing (Li and Chen 2013; Langseth and Nielsen 2015), jokes recommendations in Jester (Berkovsky et al. 2012; Yan et al. 2013; Casino et al. 2015), Web pages recommendations in Delicious (Huang et al. 2014), and music recommendations in Last.fm and Million Song Dataset (Kim and El Saddik

2015; Liu et al. 2013; Hsiao et al. 2014). In fact, improvement of users' feedback (data captured about the interactions of users with an item) in recommender systems can have a profound impact on improving CF technique, since it helps in understanding and predicting users' interests (Hsiao et al. 2014). Accordingly, the research papers with attention to questioning how explicit and implicit user feedbacks studied can be an efficient way to improve recommendation technologies as described in the fourth column of Table 2. Through observing user' behaviors (explicit (user rating) or implicit), recommender systems can infer users' preferences to indicate which products they probably like and thus select to purchase.

It is noteworthy that the algorithms are the main components of recommendation technologies in employing various types of input data such as demographic information (age, salary, gender, education, etc.), production data (actor, topic, release time, etc.), and user-item interactions (such as explicit ratings, scores, and implicit comments, search, click times, purchasing data, etc.) as input in recommender systems to predict user interests. As shown in Table 2, many research papers rely on explicit feedback, which are the





**Fig. 2** Classification scheme

most convenient in modeling users' interest for items and improve user feedback with incorporation of the additional information to the rating values. Examples of this additional information include content-based information (Wu et al. 2014; Kaššák et al. 2015), demographic information (Bakshi et al. 2014; Mehta and Banati 2014), explicit trust information (Bellogín et al. 2014), semantic information (Hawalah and Fasli 2014; Moreno et al. 2016) and social information (Mehta and Banati 2014).

The incorporation of such additional information to ratings has proved to be successful in dealing with sparsity problem. However, explicit ratings and additional external information are not always available and require user effort and additional manual labor. For example, Delicious collects

bookmarks for each URL as there are no explicit ratings. Therefore, in such a case, the recommender system can derive user' interests from the implicit user feedback, which indirectly indicates interests of user by observing user behavior (Yakut and Polat 2012; Tyagi and Bharadwaj 2013; Peng et al. 2016).

In many research papers (Hawalah and Fasli 2014; Geng et al. 2015), the explicit values are simplified as 1 and 0 which is a typical technique for implicit data to reflect whether the user like the item or not. Peng et al. (2016) and Liu et al. (2013) have applied a factorization machine technique to incorporate explicit user ratings to implicit data by converting ratings in auxiliary data via removing the ratings to improve user-item matrix. However, some research works

**Table 2** Distribution of articles by public databases and extracted users' activities

Database	User activities	User feedback	Improve user feedback	References
MovieLens, Netflix	User ratings	Implicit	Coding the ratings with scores of 3 or above to 1 (like) and rest to 0 (dislike)	<a href="#">Feng et al. (2015)</a>
			Taking (user, item) pairs with ratings equal to 5 to simulate the transaction activity as “like” preferences and other users' ratings are used to simulate examination activity (“browsed” activities)	<a href="#">Pan et al. (2015b)</a>
			Converting rating values to 0 (as dislike) and 1 (as like) and items with the least increase in the connectedness of bipartite graph is recommended	<a href="#">Gharibshah and Jalili (2014)</a>
			Applying transfer learning solution to transfer ratings from auxiliary data in binary form as like and dislike	<a href="#">Pan and Yang (2013)</a>
			Expanding the original user profile just based on user ratings by local or global techniques	<a href="#">Formoso et al. (2013)</a>
		Explicit	Using binary encoding as rating of 3 or higher is specified as item selected and lesser than 3 is seen as not selected	<a href="#">Geng et al. (2015)</a>
			Proposing a new similarity function in order to weight the rating of a user on an item based on finding the similarity between the item and the target item	<a href="#">Choi and Suh (2013)</a>
			Using a novel user similarity measure to provide a similarity measurement between users with very few ratings	<a href="#">Pirasteh et al. (2015)</a>
			Using negative ratings to extract the interests of users when the ratings are very sparse	<a href="#">Zeng et al. (2011)</a>
			Developing a similarity measure using local ratings to be calculated by correlation of the users' ratings and global ratings to be extracted from similarity between a pair of items	<a href="#">Patra et al. (2015)</a>
			Proposing a new metric to find similar users by the values of their ratings and proportions of their common items rated	<a href="#">Bobadilla et al. (2010)</a>
			Improving the neighborhood formation in CF by weighting the user ratings according to their importance	<a href="#">Bobadilla et al. (2012a)</a>
			Proposing similarity measure in which group of similar users and similar items with active user and active item are obtained.	<a href="#">Bobadilla et al. (2012b)</a>
			Identifying the relationship among users groups to each item group based on characterizing the movies	<a href="#">Cai et al. (2014b)</a>
			Proposing similarity functions using genetic algorithm on the user ratings	<a href="#">Bobadilla et al. (2011)</a>
			Predicting ratings by identifying the item ratings from k-nearest neighbors	<a href="#">Zhu et al. (2014)</a>
			Presenting a reliability measurement that provides prediction ratings based on a wider number of users	<a href="#">Hernando et al. (2013)</a>
			Presenting a similarity measure in which the number of items rated between the 2 users with the same significance is computed	<a href="#">Bobadilla et al. (2013a, b)</a>
			Improving neighborhood formation by considering the uncertainty values of user's preference or rating on diversification of items	<a href="#">Kaleli (2014)</a>

**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References		
MovieLens, Netflix	User ratings	Explicit	Proposing a new matrix factorization in which the corresponding latent features by existing ratings of each user on items are defined	<a href="#">Luo et al. (2012)</a>		
			Providing fuzzy profiles for items, users and ratings to mange noise detection in rating database	<a href="#">Yera et al. (2016)</a>		
			Providing a modified random walk model to contain the relations between user nodes and item nodes based on user ratings on items	<a href="#">Chen et al. (2013a)</a>		
		Explicit/implicit	Applying rating elicitation strategy to enlarge the set of available ratings by identifying useful ratings provided from each user	<a href="#">Elahi et al. (2013)</a>		
			Measuring the users’ similarity based on users’ typicality degrees in all user groups corresponding to each item group	<a href="#">Cai et al. (2014b)</a>		
			Proposing a power function to adjust user similarity scores derived from ratings converted to binary (assigning 1 to ratings more than 3) to find relevant links between users and items	<a href="#">Gan and Jiang (2013)</a>		
			Extending the user profiles based on contextual user preferences and providing semantic relatedness between concepts via properties such as directors, actors, country and release year	<a href="#">Hawalah and Fasli (2014)</a>		
Netflix	User ratings + tagging	Explicit/implicit	Presenting a novel algorithm that looks at tendencies or variations between items and users, instead of looking for their similarities	<a href="#">Cacheda et al. (2011)</a>		
			Using the aggregate rating information and the movie watching experience to improve the individual ratings estimations in the individual users rating datasets	<a href="#">Umyarov and Tuzhilin (2011)</a>		
			Employing tags and movie information using kernel-based method to build user-based and item-based recommendations	<a href="#">Ghazanfar et al. (2012)</a>		
Netflix	User ratings	Implicit	Presenting a new neighborhood model based on global optimization function and implicit user feedback (rating matrix into a binary matrix)	<a href="#">Koren (2010)</a>		
			Explicit	Proposing filtering algorithms based on analyzing item popularity that is specified as the number of votes received in a certain time interval	<a href="#">Javari and Jalili (2015)</a>	
		Last.fm		Play counts on artists	Implicit	Computing the similarity among users using global and local ratings. The overall user preferences on items are the combination of preferences level of the items attribute values
			Using information from social networks by applying a social graph on listened artists by users to compute the implicit similarities between users			<a href="#">Hsiao et al. (2014)</a>
Last.fm	Play counts on artists + tagging	Explicit/implicit	Proposing a matrix factorization framework on users’ implicit interaction with items to incorporate different types of recommendation (recommending items, groups and recommending friends)	<a href="#">Chen et al. (2013b)</a>		
			Cooperating item contents and social relations with user ratings into Bayesian probabilistic matrix factorization	<a href="#">Liu et al. (2013)</a>		



**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References
MSD	Play count on track + tagging	Implicit	Retrieving the information of the items and finding similarity between items descriptions users' profiles Converting play count to ratings and predicting the unknown user's rating on items by computing similarities between items	<a href="#">Kolomvatsos et al. (2014)</a> <a href="#">Ghazarian and Nematbakhsh (2015)</a>
	Play count on track	Implicit	Focus on quantitative implicit feedback including play count records of songs by users	<a href="#">Bauer and Nanopoulos (2014)</a>
MovieLens	User ratings	Implicit	Converting the rating given by user to binary ratings to show the used item or not and defines the direct inter-item relations to reduce sparsity of matrix and generate the ranking-based recommendations	<a href="#">Nikolakopoulos et al. (2015)</a>
			Applying the implicit trust-based approaches to improve the neighborhood formation utilizing information derived from existing historical ratings	<a href="#">Shambour and Lu (2015)</a>
			Utilizing the raw user-item rating matrix and item taxonomy as inputs of trust-semantic approach to create user-item prediction matrix	<a href="#">Shambour and Lu (2012)</a>
			Mapping user profiles onto categorical information of items as if the user rated an item (means previously purchased); each category that item belongs to is presented by 1	<a href="#">Bilge and Polat (2013)</a>
			Calculating implicit trust statements from rating matrix and applying graph clustering algorithm to make recommendations	<a href="#">Moradi et al. (2015)</a>
			Taking implicit feedback from ratings as user ratings larger than his/her average rating shows the preference movie of the user	<a href="#">Kim et al. (2011a)</a>
		Explicit	Providing item recommendations via identifying relevant tags when user annotates tags to items	<a href="#">Kim and Kim (2014)</a>
			Predicting the unrated items in the rating matrix based on the k most similar items to the items that have been rated by the active user	<a href="#">Xie et al. (2014)</a>
			Learning the user feature subspace from the rating data matrix in the auxiliary domain to predict missing user ratings in the target domain	<a href="#">Wang and Ke (2014)</a>
			Utilizing the user rating data and information about the items (movie types) and users (gender, age, career) as the basic data of antigens in artificial immune network	<a href="#">Chen et al. (2015)</a>
			Smoothing meager rating matrix in CF and then using kernel fuzzy c-means method for clustering the rating matrix	<a href="#">Devi and Venkatesh (2013)</a>
			Employing data reduction technique in dimensional reduction of user rating matrix	<a href="#">Wang et al. (2014a)</a>
			Using the user demographic information and user ratings by using associative classification to data annotated with semantic metadata according to a domain-specific ontology	<a href="#">Moreno et al. (2016)</a>
			Using the current ratings in the dataset without needing any additional information to manage the inconsistent ratings	<a href="#">Toledo et al. (2015)</a>

**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References
MovieLens	User ratings	Explicit	Using the user's ratings on a small number of products for effective personalized recommendation	<a href="#">Ahn et al. (2010)</a>
			Asking from user to rate the 30 randomly selected movies from the top most frequently rated films in order to find a similar user group	<a href="#">Hostler et al. (2012)</a>
			Using a framework that considers both rating and content information	<a href="#">Wu et al. (2014)</a>
			Using explicit rating data and user demographic information for improving accuracy of CF	<a href="#">Gogna and Majumdar (2015a, b)</a>
			Converting user ratings to like or dislike cases as when rating is greater than or equal to three, it represents that user likes item, if the rating is less than three, it shows dislike	<a href="#">Xie et al. (2015)</a>
			Exploring the relationships among users, items and features of items using Bayesian network	<a href="#">De Campos et al. (2010)</a>
			Clustering movies based on their genre and defining "category experts" in each category to use their ratings for prediction	<a href="#">Hwang et al. (2016)</a>
			Using the user demographic information, ratings and movie information by applying particle swarm optimization to find similar users	<a href="#">Bakshi et al. (2014)</a>
			Converting ratings into binary values as the item is represented by 1 if the rating of item is higher than the average rating of all items in a particular transaction otherwise represented by 0	<a href="#">Tyagi and Bharadwaj (2013)</a>
			Using cosine similarity between users for CF and items features for content-based to generate top-N recommended items	<a href="#">Kaššák et al. (2015)</a>
			Using SVD with an incremental approach in reducing the size of the original rating matrix based on sampling some of their columns as representative items	<a href="#">Zhou et al. (2015)</a>
			Constructing a virtual user for a group of users when the individual user profiles are in terms of set of items and their precedence relations	<a href="#">Kagita et al. (2015)</a>
			Applying cluster ensembles composed of different clustering algorithms to generate the multiple partitions of the original dataset	<a href="#">Tsai and Hung (2012)</a>
			Utilizing the conditional preferences by sorting the average ratings of users on the genres of movie	<a href="#">Liu et al. (2015)</a>

**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References
MovieLens	User ratings	Explicit	Applying fuzzy techniques to use the user ratings described in the linguistic variables	<a href="#">Zhang et al. (2013)</a>
			Maximizing log-posterior over the predicted preference order by sorting the observed rating matrix	<a href="#">Liu et al. (2014a)</a>
			Presenting an incremental CF based on user weights into the clustering process to affect users with few ratings	<a href="#">Salah et al. (2016)</a>
			Extracting latent variables on domain information from rating matrix to represents users' preferences	<a href="#">Braida et al. (2015)</a>
			Presenting a building model to identify prediction errors of pre-predicted ratings for users.	<a href="#">Kim et al. (2011b)</a>
			Using the classification methods to group users based on their demographic data and find the user's neighbors to predict ratings for items	<a href="#">Lika et al. (2014)</a>
			Combining qualitative and quantitative aspects of movie domain such as genre, cast or movie ratings to compute argumentation process into recommendation	<a href="#">Briguez et al. (2014)</a>
			Calculating difference of user ratings in the evolutionary process and providing clusters of similar users at each timestamp.	<a href="#">Rana and Jain (2014)</a>
			Casting the rating prediction task by defining the trust measures in neighbor selection	<a href="#">Bellogín et al. (2014)</a>
			Finding the preference similarity among users and weighing their relative preferences to make recommendations	<a href="#">Wang et al. (2014b)</a>
			Utilizing singular value decomposition to cluster users and items based on their latent relationships to predict interests of users	<a href="#">Hoseini et al. (2012)</a>
			Calculating the similarity between users by considering time information on users' rating	<a href="#">Rafeh and Bahrehmand (2012)</a>
			Visualizing the similarity relationships between the movies by considering their theme, genre, sequel, director, etc., as displaying the reliability between pair of items and which items are rated in a more similar way by users	<a href="#">Hernando et al. (2014)</a>
			Identifying a set of key ratings by taking both item popularity and user rating activity to impute the missing values according to each individual user's rating in the set of key ratings	<a href="#">Ren et al. (2013)</a>
			Extracting the user' ratings on item features to enhance the similarity measurement and calculate the weights of various item features to confirm item feature	<a href="#">Hu et al. (2014)</a>
			Using item metadata and user demographic data in the latent factor model to match the user-item matrices to the available auxiliary information and improve ratings	<a href="#">Gogna and Majumdar (2015a, b)</a>
		Explicit/implicit	Converting the collected ratings (explicitly or implicitly) to binary ones (as rating value greater than or equal to 4 is assigned to 1 and 0 otherwise)	<a href="#">Yakut and Polat (2012)</a>

**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References
MovieLens	User ratings + tagging	Explicit/implicit	Using bipartite graph to predict the links between users and items as lesser rating than 3 on item specifies that user has collected this item	<a href="#">Wen et al. (2014)</a>
			Using user opinions on items' metadata based on ratings and observations done by users	<a href="#">Colace et al. (2015)</a>
			Proposing a new similarity measure which is based on local context of available ratings and the global preference of these user ratings	<a href="#">Liu et al. (2014b)</a>
			Using explicit ratings and implicit data via converting user ratings in auxiliary data by removing the rating values to expand user-item matrix in factorization machine	<a href="#">Pan et al. (2015a)</a>
			Proposing a new matrix factorization-based user and item attributes (e.g., occupation, gender and age of users, release year and genre of items) to learn the latent profile of both users and items	<a href="#">Peng et al. (2016)</a>
			Using auxiliary data such as user-produced text data such as tags, ratings and media content (or its description text to learn topic distributions for medias and user interests	<a href="#">Tan et al. (2014)</a>
			Using rating information about movies in addition to their content information for making recommendations	<a href="#">Ghazanfar and Prügel-Bennett (2014)</a>
			Constructing user preference matrix based on item domain features and detecting the implicit relationships among users	<a href="#">Zhang et al. (2014)</a>
			Extracting subjective assessments (SAs) from tags to group them by referring Word.Net and compute the user similarity from item classes and SAs	<a href="#">Nakatsuji and Fujiwara (2014)</a>
			Improving recommendations via item profile enriched by mining tags from genre of movies and user preferences on features of movie such as actors and directors	<a href="#">Anand and Mampilli (2014)</a>
MovieLens	User ratings + tagging	Explicit/implicit	Using probabilistic factor analysis to integrate users' social trust network and social tags with user-item rating matrix to improve the recommendation quality	<a href="#">Ma et al. (2011)</a>
	User tagging	Implicit	Proposing a new user similarity measure with the help of tag information by tripartite graphs	<a href="#">Shang et al. (2010)</a>
			Using item visited by user (based on their tags created) and semantic features of items in ontology with the relational characteristics thereof	<a href="#">Lv et al. (2015)</a>
			Generating semantic user and item profiles by filtering social tags and discovering tags that are frequently generated by users	<a href="#">Movahedian and Khayyambashi (2014)</a>
Jester	User ratings + expert comments	Implicit	Modeling user's preferences based on linguistic terms ("like", "dislike") and expert opinions rather than numbers to represent the degree of user's preferences	<a href="#">Cheng and Wang (2014)</a>
	User ratings	Implicit	Normalizing user-item rating matrix in the scale of 0,1 as 0 shows that item is not rated by corresponding user and takes implicit users rating for making recommendation	<a href="#">Shinde and Kulkarni (2012)</a>

**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References
MovieLens, Jester	User ratings	Explicit	Improving the meager local neighborhood set in the sparsity rating matrix by the globally similar neighbors	<a href="#">Anand and Bharadwaj (2011)</a>
			Using a hybrid model-based CF in which the estimated trust and user–item similarity improve the quality of suggestions	<a href="#">Birtolo and Ronca (2013)</a>
			Representing all items and users as a relational probabilistic model to combine a user perspective and an item perspective	<a href="#">Langseth and Nielsen (2012)</a>
			Using the perturbed data for providing recommendations while guaranteeing user k-anonymity	<a href="#">Casino et al. (2015)</a>
			Searching the optimal trust path between users from rating matrix to find trustworthy friends by applying trust graph and ant colony algorithm	<a href="#">Bedi and Sharma (2012)</a>
			Enhancing Pearson correlation coefficient (PCC) similarity measure by taking the number of co-rated items	<a href="#">Polatidis and Georgiadis (2016)</a>
			Proposing a nonnegative matrix factorization to map users and items into involved feature of users and items rather than on the whole feature matrices	<a href="#">Luo et al. (2014)</a>
			Decentralizing the user profiles in which the users store and maintain their profiles	<a href="#">Berkovsky et al. (2012)</a>
			Applying a novel regularization technique for matrix factorization in CF in which pre-estimation ratings for a set of unknown ratings is created	<a href="#">Zhang et al. (2012)</a>
			Reducing the sparsity problem through employing the imputation of unknown rating with the iterative solutions	<a href="#">Ranjbar et al. (2015)</a>
MovieLens, Delicious	User ratings + tagging	Implicit	Setting higher ratings (ratings 4 and 5) in the rating matrix to one as interesting items and other rating values are set to zero	<a href="#">Ramezani et al. (2014)</a>
		Explicit/implicit	Combining user's trust network based on item ratings and social links into neighbor selection in CF to generate recommendations	<a href="#">Yan et al. (2013)</a>
MovieLens, Last.fm	User ratings + play count on artists + tagging	Implicit	Converting all the items rated by user to 1 and uses user_item_tag tripartite graph to find relationship between user_item and item_tag .	<a href="#">Zhang et al. (2010)</a>
MovieLens, Last.fm	User ratings + play count on artists	Explicit/implicit	Inferring user interests from user-generated content such as rating records, listening counts of users to artists and tags to enhance the recommendation accuracy	<a href="#">Xu and Yin (2015)</a>
			Providing a hierarchical structure and classification of items or tags by linking them to vocabulary (taxonomy) classes	<a href="#">Nakatsuji et al. (2016)</a>
			Using binary scale from listening count and exploits users with maximum number of ratings to cluster user–item matrix	<a href="#">Zahra et al. (2015)</a>
			Clustering items and users into multiple clusters by using user ratings, item features and user social networks to extract user–item groups for making recommendation	<a href="#">Huang et al. (2015)</a>

**Table 2** continued

Database	User activities	User feedback	Improve user feedback	References
	User ratings + tagging	Explicit/implicit	Using explicit data from user ratings and implicit data from tagging activities and user ratings and history of visited items by Boolean values (visited or not)	<a href="#">Da Costa and Manzato (2016)</a>
MovieLens, Last.fm, Jester, BookCrossing	User ratings + play count on artist	Explicit/implicit	Converting play counts on artists in the ratings and positioning the proposed learning algorithm in MapReduce architecture	<a href="#">Langseth and Nielsen (2015)</a>
BookCrossing	User rating	Implicit	Modeling the transition information (assume each rating as a transaction of buy) to a bipartite user–item interaction graph to infer a link prediction between users and items of previous transactions	<a href="#">Li and Chen (2013)</a>
MovieLens, Jester, Netflix	User ratings	Explicit	Using linear regression-based models to discover the relationships between rating data characteristics based on three groups: rating space, rating value and rating frequency distribution	<a href="#">Adomavicius and Zhang (2012)</a>
Delicious, Last.fm	Bookmark + user tagging	Implicit	Integrating the user feedbacks, social networks and item content to learning the latent factors both of items and users and the favors of trusted friends for the item recommendation	<a href="#">Wu et al. (2016)</a>
Delicious	User tagging	Implicit	Combining a probabilistic model of tags from the items with tags from the user	<a href="#">Krestel and Fankhauser (2012)</a>
			Incorporating social relationships among users to identify the most similar users and the correlation between a user and an item from tagging records	<a href="#">Sun et al. (2015)</a>
			Presenting a hybrid of CF to find similar users and content-based filtering to recommend similar items by incorporating recency, duration and frequency of tagging information	<a href="#">Huang et al. (2014)</a>
Last.fm	User tagging	Implicit	Modeling the relationships in tagging data as encoding the user's historical tags to recommend tags which receive highest ranking scores	<a href="#">Zhao et al. (2015)</a>
			Presenting pseudo-tags by aggregating tags from similar tracks to represent a track (pseudo-tags learned from similar tracks)	<a href="#">Horsburgh et al. (2015)</a>
			Proposing new graph model that profits from implicit relationships between items, users and between users and items based on social tagging	<a href="#">Kim and El Saddik (2015)</a>

([Peng et al. 2016](#); [Huang et al. 2015](#); [Yan et al. 2013](#); [Liu et al. 2013](#); [Yakut and Polat 2012](#)) have considered implicit and explicit feedback jointly to achieve useful recommendations so that they suppose users' ratings as implicit feedback on user purchase. For example, [Yakut and Polat \(2012\)](#) converted the user ratings to binary ones (as rating value greater than or equal to 4 is assigned to 1 and 0 otherwise) to show users have watched movie or not. [Huang et al. \(2015\)](#) have employed both explicit ratings and implicit play counts on artists to cluster items and users into multiple clusters in order to extract user–item groups for making recommendation. As a result, information about user activities and interactions between users and items (see [Table 2](#)) is analyzed as input for recommendation algorithms in order to compare the ways

of enhancing the extraction and application of user feedback in recommender systems.

A recommender system is based on the relationship between users and items from users' interaction records on items (such as explicit ratings, scores, and implicit comments, search, click times, purchasing data, the link of Webs, etc.). To determine the patterns captured and understand the relationship between users and items from users' interaction records, many models and ways have been proposed during the past 5 years (see [Table 2](#)). According to our classification method, these models and ways have been classified into explicit and implicit user feedback and indicate the user preferences evidence (tags, link relations, personal user data, social contacts, user-contributed content and user–item inter-



action data). [Shang et al. \(2010\)](#) and [Liu et al. \(2013\)](#) have suggested that implicit feedback may be more accurate than explicit feedback in reflecting the user's preferences. Other researchers ([Cheng and Wang 2014](#); [Hsiao et al. 2014](#)) have suggested that in certain systems, implicit user feedback can be more reliable than explicit user feedback.

## 4.2 Overview of algorithms used in improving CF and classification of articles

Several techniques to improve CF have been developed, which are clustering ([Bilge and Polat 2013](#); [Moradi et al. 2015](#)), classification ([Da Costa and Manzato 2016](#); [Kim and El Saddik 2015](#)), artificial neural network (ANN) ([Devi and Venkatesh 2013](#)), particle swarm optimization (PSO) ([Bakshi et al. 2014](#); [Tyagi and Bharadwaj 2013](#)), support vector machine (SVM) ([Ghazarian and Nematbakhsh 2015](#)), evolutionary computing techniques ([Mehta and Banati 2014](#); [Lu et al. 2015](#)), link analysis ([Feng et al. 2015](#); [Zeng et al. 2011](#)), regression, matrix factorization ([Pirasteh et al. 2015](#); [Pan and Yang 2013](#)), and they are listed in Table 3.

The distribution of one hundred and thirty-one (131) papers classified according to public datasets and used techniques in recommender systems is represented in Table 3, and then techniques employed in making recommendations are summarized. It is significant to note that classifying the research works into techniques according to proposed classification scheme has been comprehensively conducted by doing the review of recommendation system papers published by [Park et al. \(2012\)](#), [Lu et al. \(2015\)](#) and [Bobadilla et al. \(2013a,b\)](#). It is meaningful to analyze and investigate published works to employ public recommendation databases through techniques used in improving CF. The aim of Table 3 is to support and motivate practitioners and researchers by providing the state-of-the-art knowledge on public datasets and providing guidelines on how to implement and validate recommender systems under different domains to support users in various decision activities. Table 3 helps to understand how to improve CF with benchmarks and standard datasets.

Table 3 classifies 131 published works according to artificial intelligence and data mining techniques including classification, link analysis, association rule mining, evolutionary computing, regression, matrix factorization, context awareness-based and content-based, mathematical technique, clustering and fuzzy set. The descriptions of aforementioned techniques are as follows:

### 4.2.1 Intelligence computing algorithms in improving CF

Evolutionary computation technique is a sub-area of artificial intelligence that can be defined by the type of algorithms it

deals with. The most employed intelligence computing algorithms in solving sparsity rating data are as follows:

- i. *Particle swarm optimization (PSO)*: Particle swarm optimization (PSO) is initialized with a group of random particles or solutions and then searches for optimal solution by updating these particles. PSO finds the global optimum with rough initial condition setting. Since it requires only primitive mathematical operators, it is computationally inexpensive by memory and speed requirement ([Bakshi et al. 2014](#); [Tyagi and Bharadwaj 2013](#)). PSO technique finds local neighbors (users who have co-rated items) and global neighbors (users are connected via local neighbors and have transitive similarities) for making prediction on unrated items ([Bakshi et al. 2014](#)).
- ii. *Ant colony*: Ant colony is a probabilistic technique to solve the computational problems and reduce computational problems by finding good paths through graphs. Metaphors in ant colony help to select the most optimal path in the user interface graph. In CF-based recommender systems, the best neighborhood is selected based on biological metaphor of ant colonies to make recommendations for active user ([Bedi and Sharma 2012](#)).
- iii. *Genetic*: Genetic algorithms are stochastic search techniques for solving optimization problems with an objective function that is subjected to soft and hard constraints ([Lu et al. 2015](#)). They have mainly been employed in two aspects of collaborative recommender systems: clustering and hybrid user models ([Mehta and Banati 2014](#)). Genetic algorithm uses parameters to represent chromosomes. Each parameter is encoded as a genetic variable. Therefore, a parameter is optimized together with the chromosome containing many genes. The population is thus a structure established with a certain number of genes together with the variables associated with these genes. A fitness function is used to assess the goodness of an individual solution, and the next generation is formed by the progeny cells of the chromo generated during the crossover process. The chromo with the highest fitness is selected as the parent to provide offspring in the next generation. This process is repeated until certain fitness has been achieved or until a particular number of offsprings have been produced ([Lu et al. 2015](#); [Lv et al. 2015](#)). [Bobadilla et al. \(2011\)](#) improve the calculation of similarity between users in CF using a genetic algorithm. They have presented a metric to utilize the combination of values computed between users in similarity stage and weights calculated by genetic algorithm.
- iv. *Immune Network*: Immune network technique simulates the mechanism of a biological immune system fighting foreign pathogens. It can be provided as com-

**Table 3** Distribution of articles by public databases and techniques used in improving CF

Database	Techniques used in improving CF	References
MovieLens, Netflix	Link analysis + association rule mining (ARM)	<a href="#">Feng et al. (2015)</a>
	New similarity measure for finding similar items	<a href="#">Choi and Suh (2013)</a>
	Bayesian classifier	<a href="#">Pan et al. (2015b)</a>
	Matrix factorization	<a href="#">Pirasteh et al. (2015)</a>
	Link analysis	<a href="#">Zeng et al. (2011)</a>
	New similarity measure for finding similar users	<a href="#">Patra et al. (2015)</a>
	New similarity measure for finding similar users	<a href="#">Bobadilla et al. (2010)</a>
	New similarity measure to measure the significance of items and users	<a href="#">Bobadilla et al. (2012a)</a>
	New similarity measure for groups of items and groups of users	<a href="#">Bobadilla et al. (2012b)</a>
	New similarity measure for finding similar users	<a href="#">Gan and Jiang (2013)</a>
	Clustering technique	<a href="#">Cai et al. (2014b)</a>
	Link analysis	<a href="#">Gharibshah and Jalili (2014)</a>
	Genetic algorithm	<a href="#">Bobadilla et al. (2011)</a>
	Matrix factorization	<a href="#">Pan and Yang (2013)</a>
	K-NN	<a href="#">Zhu et al. (2014)</a>
	Regression	<a href="#">Ghazanfar et al. (2012)</a>
	New measurement in predictions of items to users	<a href="#">Hernando et al. (2013)</a>
	New similarity measure for finding similar users by use of low-cost hardware	<a href="#">Bobadilla et al. (2013a,b)</a>
	New similarity measure for finding similar users	<a href="#">Kaleli (2014)</a>
	Context awareness-based	<a href="#">Formoso et al. (2013)</a>
	Immune network	<a href="#">Geng et al. (2015)</a>
	Matrix factorization	<a href="#">Luo et al. (2012)</a>
	Context awareness-based	<a href="#">Hawalsh and Fasli (2014)</a>
	Fuzzy set-based	<a href="#">Yera et al. (2016)</a>
	Context awareness-based	<a href="#">Cacheda et al. (2011)</a>
	Context awareness-based	<a href="#">Umyarov and Tuzhilin (2011)</a>
	Link analysis	<a href="#">Chen et al. (2013a)</a>
	Context awareness-based	<a href="#">Elahi et al. (2013)</a>
	Clustering technique	<a href="#">Cai et al. (2014a)</a>
Netflix	New similarity measure for finding similar users	<a href="#">Koren (2010)</a>
	Context awareness-based	<a href="#">Javari and Jalili (2015)</a>
	Context awareness-based	<a href="#">Eckhardt (2012)</a>
MovieLens	Forecasting	<a href="#">Xie et al. (2014)</a>
	Context awareness-based	<a href="#">Wang and Ke (2014)</a>
	Immune network	<a href="#">Chen et al. (2015)</a>
	Artificial neural network (ANN)+ fuzzy set-based	<a href="#">Devi and Venkatesh (2013)</a>
	Clustering+ genetic algorithm	<a href="#">Wang et al. (2014a)</a>
	Bayesian classifier	<a href="#">Tan et al. (2014)</a>
	Clustering technique	<a href="#">Nikolakopoulos et al. (2015)</a>
	Context awareness-based (semantic web + ontology-based)	<a href="#">Moreno et al. (2016)</a>
	Context awareness-based	<a href="#">Toledo et al. (2015)</a>
	Context awareness-based	<a href="#">Ahn et al. (2010)</a>
MovieLens	K-NN	<a href="#">Hostler et al. (2012)</a>
	Content-based + clustering	<a href="#">Wu et al. (2014)</a>
	Matrix factorization	<a href="#">Gogna and Majumdar (2015a,b)</a>
	Link analysis	<a href="#">Xie et al. (2015)</a>
	Link analysis	<a href="#">Shambour and Lu (2015)</a>

**Table 3** continued

Database	Techniques used in improving CF	References
	Link analysis	Shambour and Lu (2012)
	Bayesian classifier + content-based	De Campos et al. (2010)
	Context awareness-based + clustering	Hwang et al. (2016)
	Clustering + PSO	Bakshi et al. (2014)
	Bayesian classifier	Yakut and Polat (2012)
	ARM + PSO	Tyagi and Bharadwaj (2013)
	Clustering technique	Ghazanfar and Prügel-Bennett (2014)
	Content-based	Kaššák et al. (2015)
	Matrix factorization	Zhou et al. (2015)
	Context awareness-based	Kagita et al. (2015)
	Clustering technique + content-based	Bilge and Polat (2013)
	Fuzzy set-based	Cheng and Wang (2014)
	Clustering technique + link analysis	Moradi et al. (2015)
	Clustering technique	Tsai and Hung (2012)
	Context awareness-based	Zhang et al. (2014)
	Matrix factorization	Liu et al. (2015)
	PSO + genetic algorithm	Mehta and Banati (2014)
	Context awareness-based	Nakatsuji and Fujiwara (2014)
	Regression+ matrix factorization	Liu et al. (2016)
	Genetic algorithm	Lv et al. (2015)
	Link analysis + clustering technique	Liu et al. (2010)
	Clustering technique	Boratto et al. (2015)
	K-NN + fuzzy set-based	Zhang et al. (2013)
	Matrix factorization	Liu et al. (2014a)
	Clustering technique	Salah et al. (2016)
	Reducing dimensionality	Braida et al. (2015)
	Content-based	Kim et al. (2011a)
	New prediction measure to users on items	Kim et al. (2011b)
	New similarity measure for finding similar users based on tagging information	Shang et al. (2010)
	New similarity measure + link analysis	Wen et al. (2014)
	Context awareness-based	Colace et al. (2015)
	New similarity measure for finding similar uses	Liu et al. (2014b)
	Context awareness-based	Pan et al. (2015a)
	Bayesian classifier	Lika et al. (2014)
	Fuzzy set-based	Anand and Mampilli (2014)
	ARM + link analysis	Kim and Kim (2014)
	Context awareness-based	Briguez et al. (2014)
	Clustering technique	Rana and Jain (2014)
	Context awareness-based	Ma et al. (2011)
MovieLens	Link analysis	Bellogín et al. (2014)
	Context awareness-based	Wang et al. (2014b)
	Matrix factorization	Hoseini et al. (2012)
	Context awareness-based	Movahedian and Khayyambashi (2014)
	Link analysis	Rafeh and Bahrehmand (2012)
	Link analysis	Hernando et al. (2014)
	Context awareness-based	Ren et al. (2013)
	New similarity measure based on item features	Hu et al. (2014)

**Table 3** continued

Database	Techniques used in improving CF	References
Jester MovieLens, Jester	Matrix factorization	Gogna and Majumdar (2015a, b)
	Matrix factorization	Peng et al. (2016)
	Clustering technique	Shinde and Kulkarni (2012)
	Link analysis	Yan et al. (2013)
	New similarity measure based on global and local user similarity	Anand and Bharadwaj (2011)
	Link analysis + clustering technique	Birtolo and Ronca (2013)
	Bayesian classifier	Langseth and Nielsen (2012)
	Context awareness-based	Casino et al. (2015)
	Decision tree	Ramezani et al. (2014)
	Link analysis + ant colony algorithm	Bedi and Sharma (2012)
	New similarity measure for finding similar users	Polatidis and Georgiadis (2016)
	Matrix factorization	Luo et al. (2014)
	Context awareness-based	Berkovsky et al. (2012)
	Matrix factorization	Zhang et al. (2012)
MovieLens, Delicious	Matrix factorization	Ranjbar et al. (2015)
MovieLens, Last.fm	Link analysis	Zhang et al. (2010)
	Matrix factorization	Xu and Yin (2015)
	Clustering technique	Nakatsuji et al. (2016)
	Clustering technique	Zahra et al. (2015)
MovieLens, Last.fm, Jester, BookCrossing	Matrix factorization + Bayesian classifier	Da Costa and Manzato (2016)
	Clustering technique + link analysis	Huang et al. (2015)
	Regression	Langseth and Nielsen (2015)
	Regression	Adomavicius and Zhang (2012)
Delicious	Context awareness-based	Krestel and Fankhauser (2012)
	Content-based	Huang et al. (2014)
	Clustering technique	Sun et al. (2015)
	Link analysis	Zhao et al. (2015)
Delicious, Last.fm	Matrix factorization (MF) + link analysis	Wu et al. (2016)
Last.fm	Content-based	Horsburgh et al. (2015)
	Link analysis	Hsiao et al. (2014)
	Matrix factorization	Chen et al. (2013b)
	Matrix factorization	Liu et al. (2013)
	Link analysis	Kim and El Saddik (2015)
	Support vector machine (SVM)	Ghazarian and Nematbakhsh (2015)
Million Song Dataset	Matrix factorization	Bauer and Nanopoulos (2014)
	Content-based	Kolomvatsos et al. (2014)
	Link analysis	Li and Chen (2013)
BookCrossing	Link analysis	Li and Chen (2013)

putational systems inspired by theoretical immunology and observed immune functions, models and principles which are applied in problem solving. [Chen et al. \(2015\)](#) combine CF with immune network to solve sparsity problem via treating the rating data as antigens. Then, antigens are copied as the antibodies of the immune networks to generate number of immune networks for finding the nearest neighbors in CF for an active user or item ([Chen et al. 2015](#); [Geng et al. 2015](#)).

#### 4.2.2 Machine learning algorithms in improving CF

To overcome sparsity problem, some recommender system researches adopt data mining techniques to provide possible connections between users and items based on retrieving either explicitly or implicitly user preferences and obtain the most efficient results. In this subsection, a brief review of literature which is related to data mining techniques in solving sparsity problem is presented. Data mining is known as

knowledge discovery tools to find hidden, new knowledge or unexpected patterns from databases or Web sites.

- i. *Matrix factorization*: Matrix factorization is one of the most successful methods which are highly scalable and accurate in reducing sparsity problem in CF. It transforms both items and users into the same latent factor space. Moreover, each entry is specified by a feature vector to be inferred from the existing ratings. Then, the unknown rates are predicted using the inner products of the corresponding vector pairs (Xu and Yin 2015; Pirasteh et al. 2015; Pan and Yang 2013). Xu and Yin (2015) apply matrix factorization to the user similarity matrix in which the corresponding latent features by existing ratings of each user on items are defined to predict unrated items and improve accuracy of CF technique.
- ii. *Association rule mining*: The association rules mining technique has also been applied to represent users' interests in various fields for providing recommendation models. This is due to its ability to scale to large datasets and achieve high accurate recommendations (Tyagi and Bharadwaj 2013). Association rule mining discovers the interesting association relationships (known as rules) hidden in databases which are above user-specified minimum confidence and minimum support levels. These specified minimum confidence and support define how strong the association rules are and how likely the rules are to occur again. Selected rules can form a model for predicting the future interests of a user. In other words, the extracted association rules predict the presence of an item via the occurrences of other items in a transaction where each transaction is user actions to include a set of items. Association rules can eliminate the dependency of CF on the users' co-rated items by discovering the hidden connections between users and items from users' past behaviors (Tyagi and Bharadwaj 2013). Support and confidence measures are two important concepts for evaluating the rules in the association rules mining technique. The support and confidence of an association rule  $X \rightarrow Y$  is defined by Eqs. (1) and (2), and only the rules having the following support and confidence values are selected as the useful rules (Tyagi and Bharadwaj 2013)

Support ( $X \rightarrow Y$ )

$$= \frac{\text{number of transactions which contain } X \text{ and } Y}{\text{number of all transactions in the database}}$$

Confidence ( $X \rightarrow Y$ )

$$= \frac{\text{number of transactions which contain } X \text{ and } Y}{\text{number of transactions which contain } X}$$

- iii. *Forecasting*: Forecasting technique predicts the future behaviors of users based on their past record pat-

terns. It deals with continuously valued outcomes to shape the logical relationships among users in predicting their interests. Artificial neural network (ANN) is a known model for forecasting (a parallel distributed information processing system). This system learns and self-organizes a large number of uncomplicated processing entities which are interconnected to form a network that conducts complex computational tasks (Devi and Venkatesh 2013; Xie et al. 2014). This technique focuses on insufficient ratings to obtain the required rating prediction in CF for decision making to users (Ramezani et al. 2014).

- iv. *K-nearest neighbor (K-NN)*: A common basic CF technique used in recommender systems to predict future behavior of active user based on interests of users who share similar interests with the active user. This technique ascertains similar users who have previously exhibited similar preferences in order to provide recommendations (Zhu et al. 2011; Hostler et al. 2012).
- v. *Support vector machine (SVM)*: SVM is an intelligent data analysis technique in classification to find a linear hyperplane (decision boundary) that splits the data in such a way that the margin is maximized. For instance, there are many possible boundary lines that can separate a two-class separation problem in two dimensions where each boundary has an associated margin. The rationale behind SVM is that if the one that maximizes the margin is chosen, it is less likely that the unknown items can be categorized. Hence, SVM is a pattern analysis to find and analyze the general types of relations (for example, rankings, classifications, clusters, correlations, principal components) in datasets to compute similarities between pairs of users (Ghazarian and Nematbakhsh 2015). SVM is a function learning algorithm which learns the provided function from input data in the best manner. It tries to find  $f(x)$  function that approximates the relations between data points. The pairs of input data are as follows:  $\{(x_1, y_1), \dots, (x_i, y_i)\}$  (Yu and Kim 2012). In linear function, the relationships between input and output data are linear. The function is computed as follows:

$$f(x) = wx + b,$$

where  $w \in X$ ,  $X$  shows the input space and  $b$  shows a real value.

- vi. *Regression*: Regression analysis technique uses a linear relationship to provide a connection among two or more variables systematically. It is a diversified process and powerful in analyzing associative relationships between dependent variable and one or more independent variables. Uses of regression involve making prediction, curve fitting and testing systematic hypotheses about relationships between variables. The curve can be use-



ful to identify a trend within dataset, whether it is linear, parabolic or of other forms (Adomavicius and Zhang 2012). Adomavicius and Zhang (2012) use the linear regression-based models to find the connections among rating data characteristics based on three groups: rating space, rating value and rating frequency distribution to improve recommendation quality in CF. The equation is linear as shown below (Liu et al. 2016):

$$Y = a + bX,$$

where  $Y$  represents the dependent variable (variable that goes on the  $Y$ -axis),  $a$  represents the  $y$ -intercept,  $X$  is the independent variable (i.e., it is plotted on the  $X$ -axis), and  $b$  is the slope of the line.

- vii. *Link analysis*: Link analysis can effectively deal with sparsity problem by exploring trends and pattern by building up networks of interconnected objects (users or items). Link analysis finds the associations between objects in a database. It has presented great potentials in enhancing the performance of Web search. Social network analysis is one type of link analysis technique to discover a fundamental social structure via analyzing the patterns relationships and interactions between social actors. Social network analysis provides recommendation by employing users' social interaction (such as making social comments, social tags, online friending, etc.). Link analysis can also employ graph-based techniques or graph structure to obtain information on relations between users and items, as nodes in graph are users and items and the links between nodes are user–item interactions to show interest of user. Most of link analysis algorithms utilize a single node in the Web graph to handle a Web page (Li and Chen 2013; Park et al. 2012). The feedbacks and transactions are formed as links connecting the nodes between the two sets. The intuitions behind bipartite graphs are explained by following example. Suppose the user–item interaction matrix is modeled as a bipartite graph shown in Fig. 3 where two sets of nodes are items and users, and an edge connects user  $X$  to item  $P_4$  if there is transaction done by  $X$  on item  $P_4$  (for example, item  $P_4$  has been purchased by user  $X$ ) (Feng et al. 2015).

In addition, trust-based recommender systems (trust network) utilize a social network augmented with trust ratings, known as link analysis, to provide recommendations for users based on people they trust. A trust network is initialized with a directed graph in which the nodes are users and the edges are weighted according to the degree of trust assigned by one user to

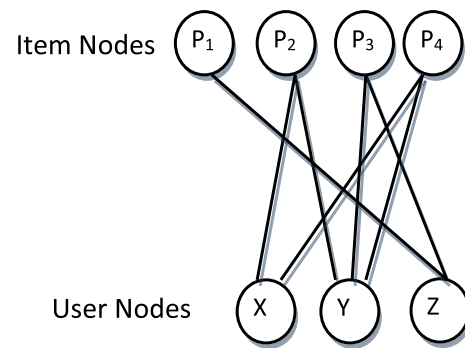


Fig. 3 Example for bipartite graph (Feng et al. 2015)

- another. Semantic-based recommender systems exploit the underlying semantic properties and attributes associated with users and items to provide recommendations (Lu et al. 2015).
- viii. *Fuzzy set-based*: Fuzzy set theory offers a broad spectrum of techniques for the management of non-stochastic ambiguity. Fuzzy set theory is efficient in handling imprecise information, the unsharpness of classes of situations or objects, and the steadiness of users' profiles (Zhang et al. 2013; Lu et al. 2015). In paper published by Zhang et al. (2013), an item in a collaborative recommender system was represented as a fuzzy set over an assertion set. The value of an attribute for an item is a fuzzy set over the subset of the assertions relevant to the feature. The user's interests are represented as the basic interest module that can evaluate items. The user's extensional interests are expressed as a fuzzy set over the user's experienced items whose membership degrees are the ratings. Based on the representation, the user's interests for an item can then be deduced (Lu et al. 2015; Cheng and Wang 2014; Anand and Mampilli 2014). Anand and Mampilli (2014) have proposed a fuzzy set theory in which item profile enriched by mining tags from genre of movies and user preferences on features of movie such as actors and directors are combined to improve recommendations in CF.
- ix. *Bayesian networks*: Bayesian networks are probabilistic graphical models which use probability to represent uncertainty about the relationships gathered from the data. These models are based on the definition of conditional probability and Bayes theorem. In addition, the concept of prior probability which is used in these networks serves very crucial aspect in classifying. This is because prior probability represents our expectations or our acquired knowledge about what the true relationship might be. In particular, the probability of a model given the data (posterior) is proportional to the product of the likelihood times the prior probability (the prior). The probability component includes the



effect of the data, while the prior specifies the belief in the model before the data are observed. When a Bayesian network is implemented in a collaborative recommender system, each node shows an item and the states correspond to each possible vote value. In the network, there will be a set of parent items for each item which show its best predictors (Lu et al. 2015).

Bayesian network also known as probabilistic networks and belief networks to explore the relationships among users and items for predicting the user preferences based on these relationships to improve CF (Langseth and Nielsen 2012). Probability is the most common task to be solved using Bayesian networks. Suppose that grass can be wet by two events: by raining or sprinkler is on. A Bayesian network can model the mentioned situation for computing the posterior probability of each explanation (where 1 (for true) and 0 (for false) (De Campos et al. 2010; Liu et al. 2013)).

$$\begin{aligned}
 P_r(S = 1|W = 1) &= \frac{P_r(S = 1, W = 1)}{P_r(W = 1)} \\
 &= \frac{\sum_{c,r} P_r(C = c, S = 1, R = r, W = 1)}{P_r(W = 1)} \\
 &= \frac{0.2781}{0.6471} = 0.43 \\
 P_r(R = 1|W = 1) &= \frac{P_r(R = 1, W = 1)}{P_r(W = 1)} \\
 &= \frac{\sum_{c,s} P_r(C = c, S = s, R = 1, W = 1)}{P_r(W = 1)} \\
 &= \frac{0.4581}{0.6471} = 0.708 \\
 P_r(W = 1) &= \sum P_r(C = c, S = s, R = r, W = 1) \\
 &= 0.6471.
 \end{aligned}$$

So it can be shown that it is more likely that the grass is wet because it is raining: The likelihood ratio is  $0.7079/0.4298 = 1.647$

- x. *Clustering*: Clustering technique classifies a set of data into a set of sub-clusters in order to find the meaningful groups that exist within them (Park et al. 2012). Once clusters have been formed, the opinions of other users in a cluster can be averaged and used to generate recommendations for individual users. A good clustering technique will provide high-quality clusters where the intra-cluster similarity is high and the inter-cluster similarity is low. Ghazanfar and Prügel-Bennett (2014) have proposed a clustering technique in which a user has partial participation in different clusters and recommendations are based on the average across the clusters of participation that is weighted by degree of participation (Ghazan-

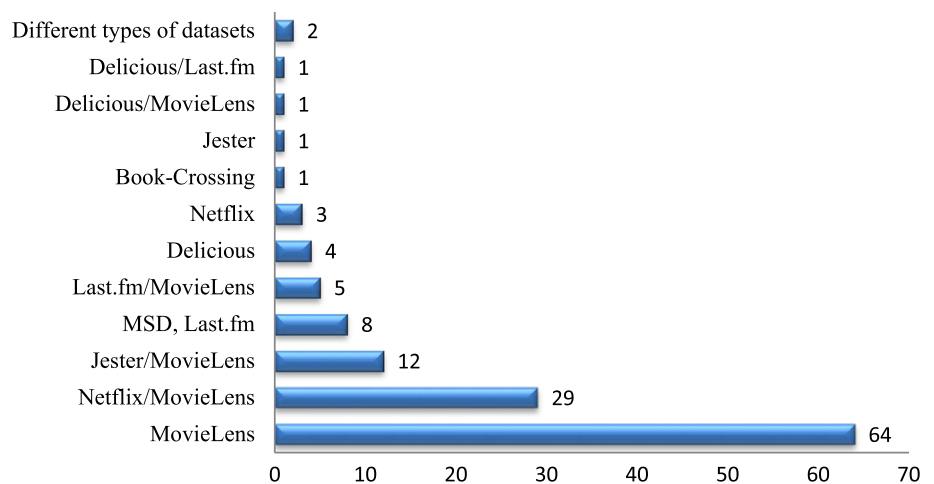
far and Prügel-Bennett 2014). K-means and hierarchical clustering are two kinds of clustering techniques used in recommender systems. K-means takes an input parameter and then partitions a set of items into clusters (Ghazanfar and Prügel-Bennett 2014; Shinde and Kulkarni 2012). Hierarchical clustering generates a set of nested cluster organized as a hierarchical tree (Park et al. 2012).

- xi. *Decision tree*: This technique classifies specific entities into a set of known classes in the form of a tree structure based on the features of the entities: A root node is the top node, followed by the leaf nodes. Each node is labeled with a question (a single attribute value) to determine which branch of the sub-tree applies and a curve associated with each node or leaf node covers all possible responses (indicate the value of the target attribute) (Ramezani et al. 2014; Park et al. 2012). This technique can improve the calculation of similarities between users or items in CF for improving accuracy of CF recommender systems. For example, Ramezani et al. (2014) create different subspaces of users' interests on items in order to remove the redundant item subspaces for each user. Then, users who share the same interest patterns on each subspace are defined as neighbor users based on a user's tree structure.

#### 4.2.3 Other solutions in improving CF

- i. *Context awareness-based, content-based*: Recently, many studies have been done in CF that focused on extracting the context information that can be used to characterize the situation of an entity (Formoso et al. 2013; Lu et al. 2015). An entity could be a place, a person or an object that is considered relevant to the interaction between a user and an application, including the user and the application themselves. The contextual information captures the additional information for recommendation making, especially for some applications in which it is not sufficient to consider only users and items, such as recommending the user's preferences under different conditions. For example, we like hot drinks in winter, but we prefer cold drinks in summer. This kind of preference relations is called conditional preference which exists in real-life context or situation.

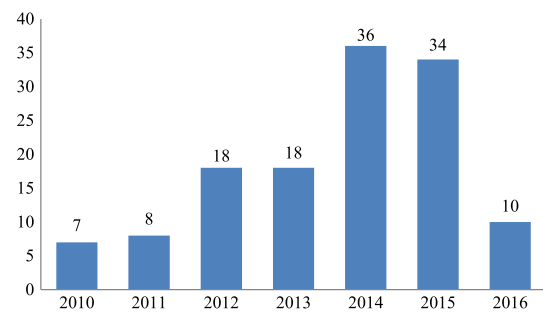
Movahedian and Khayyambashi (2014) improve the accuracy of recommendations by employing subjective assessments assigned to items by users as users assign a low or high rating to the item with reason. The context information such as time, contextual user preferences, geometrical information, ratings knowledge by means of user, the company of other people (friends, families or colleagues for example) or expert opinions has been recently considered in existing recommender systems

**Fig. 4** Distribution of research articles by public datasets

(Movahedian and Khayyambashi 2014; Ren et al. 2013). Context awareness-based includes semantic models, item and rating classifications to find out noisy preferences (inconsistent among users when they elicit ratings for items), local and global techniques and popularity-based models (Hawalab and Fasli 2014).

Various techniques for solving the sparsity problem have been developed to combine content-based technique with CF (Wu et al. 2014; De Campos et al. 2010). Content-based matches the user profiles with the attributes of items in order to generate predictions on users' interests and unlike CF ignores the contributions from other users (Wu et al. 2014).

- ii. *Mathematical techniques: proposing similarity measure or prediction measure:* Neighborhood formation is an important part of CF, which forms neighborhood of an entity (finding similar users to an active user or similar items to the candidate item) by using the traditional similarity measure, i.e., Pearson's correlation coefficient or cosine-based similarity measure (Najafabadi et al. 2017). When user preferences are very sparse, traditional similarity measures that utilize ratings of only co-rated items might end up with unreliable neighborhoods of an entity. To date, researchers propose a similarity measure for neighborhood formation or predicting the preference of the active user in order to improve basic CF. Some research papers (Gan and Jiang 2013; Patra et al. 2015) propose a new similarity function instead of basic similarity function in CF to calculate similarity among users. This similarity function can improve neighbor formation in CF (specify neighbors of an active user) or enhance the accuracy of predictions by proposing a new prediction measurement (Hernando et al. 2013; Kim et al. 2011) to provide value about how much active user will like item.

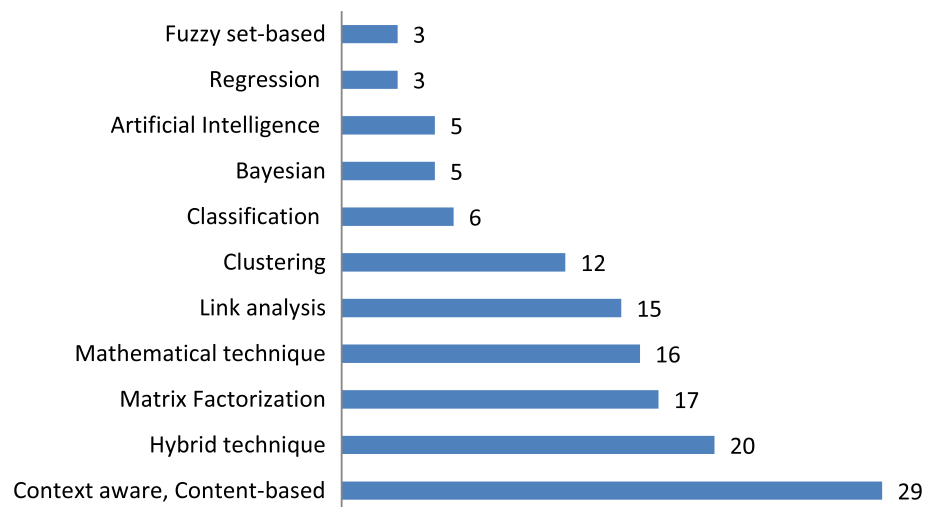
**Fig. 5** Distribution of research articles by publication year

#### 4.3 Distribution of articles by public databases and publication years

One hundred and thirty-one articles from 36 journals were selected and classified according to proposed classification scheme (see Fig. 2). The details are shown in Figs. 4 and 5. Distribution of articles by public databases is represented in Fig. 4. It is apparent that the majority of the recommender system researches employed the MovieLens (64 out of 131 articles, or 49%) to construct movie recommendations. MovieLens dataset is the standard benchmark dataset that has less sparse data provided by GroupLens Research. Thus, the preprocessing data in this dataset are easier to use and the subsequent analysis and evaluation for assessing the quality and efficiency of recommendation technique proposed can be done without much problem. One of the other popular datasets used in research works is Netflix; this database has been frequently used with MovieLens dataset as the experimental data (29 out of 131 articles, or 22%). Some recommendation techniques have been proposed and evaluated by using both application platforms including Jester and MovieLens datasets (12 out of 131 articles, or 9%).

Figure 4 shows that although many articles were published in recommendation fields, only several of them used Delicious (4 out of 131 articles, or 3%), BookCrossing (1 out

**Fig. 6** Distribution of articles by techniques used in recommender systems



of 131 articles or 0.76%) and Million Song Dataset (MSD) and Last.fm (8 out of 131 articles, or 6%) as the experimental datasets. Less than 4% (only 5 out of 131 articles) of the research works has taken the experimental data from Last.fm along with MovieLens dataset. Therefore, it looks to be necessary to implement and evaluate new techniques proposed in other fields except movie. It is noted that research works which use more than two benchmark recommender system datasets for developing new techniques and comparing them with other recommendation techniques have been classified in different types of datasets.

The distribution of published research articles between 2010 and 2016 (the first quarter of 2016) is shown in Fig. 5. It is clear that publications which are related to CF research have steadily increased between 2010 and 2013. They continue to rapidly increase between 2013 and 2015. The decrease in articles in 2016 is probably due to the progress of the research and the write-ups. Based on the results of previous publication rates, it expects that interest in CF research will grow significantly in the future.

#### 4.4 Distribution of articles by techniques used in improving CF

The details about techniques used in improving CF are described in Sect. 4.2. It is noted that hybrid recommendation technique has been developed by combining more than one technique presented in classification scheme (see Fig. 2). Articles that employ more than one technique have been classified in hybrid techniques. In recent years, it has been proven that a single algorithm is not generally able to minimize the shortcomings in using basic CF and optimize recommendation accuracy. This explains why researchers have conducted a number of successful recommender systems in which to employ hybrid techniques (20 out of 131 research articles) as shown in Fig. 6.

As shown in Fig. 6, most of researches conducted in CF have focused on context awareness-based in improving accuracy of CF (29 out of 131 research article). This is due to the fact that contextual information and content-based capture the additional information sources beyond the user-item matrix (such as time, tags, comments) to enrich user profiles in providing the appropriate recommendation. Researchers have also conducted several studies on proposing a new similar function to identify users who are similar to the active users or present a new prediction measurement to provide value about how well active users will like the item that has been classified in mathematical techniques (16 out of 131 research articles). It is noted that matrix factorization is the main research focus of current CF research (17 out of 131 research articles). This has been proven that that matrix factorization technique can address CF problems with highly accurate and scalable result in most of the application fields. It is apparent that recently the developed advanced techniques such as link analysis, regression and fuzzy set-based are successful techniques that are widely used in recommender systems today.

Association rule mining is one of the most successful techniques that researchers combine with other techniques including particle swarm optimization (PSO) and link analysis to alleviate the sparsity problem in CF. However, few researches have considered developing recommender systems by association rule mining; thus, there are still some issues on association rule mining that need to be addressed in the light of the emerging recommendation systems. In addition, several efforts have been proposed that can facilitate better handling of the challenges in CF by employing clustering (12 out of 131 research articles). Heuristic methods have been developed by adding new method to existing methods. Classification algorithms include K-NN, SVM, forecasting and decision tree that have been developed by the few numbers of referred papers (6 out of 131 research articles).

As a consequence, a recommendation system is based on three basic kinds of entities including items (e.g., music, news, books and movies), users and user–item historical records (e.g., tags, comments, scores). The main task is to determine the useful patterns in describing the association among users and items which are from the user–item historical records. Then, predictions are made for possible user–item links based on patterns. To accomplish this task, many techniques or algorithms were developed during the past years. In general, the existing research works can be further divided into three categories which can effectively deal with data sparseness problem in CF. One emphasizes the usage of item-specific contents, such as link relations, comments, tags; the other research works focus exploiting user-specific information such as the trust relationship between users. The third category uses mechanism to employ pure rating data to find “neighbors” of users and make predictions. However, these research papers have used either user’s social information or item contents and few of them have considered them jointly.

## 5 Research implications

The findings represented in this paper have several significant implications as follows:

- The findings have showed that even though research done in CF has obtained great development in different application fields, there are still some issues on music, book, joke and document recommendation systems that require further research especially with the emerging of new recommender system applications. MovieLens has been subjected to significant research on CF, since this dataset is known as the best known example for common users and the use of this dataset is easy. Therefore, in order to fill this gap, more researchers are required to use datasets in other application fields instead of movie.
- Based on reviews on academic research papers and issues gathered on CF research, it is obvious that a good mechanism to improve user’ preference matrix and select a set of “neighbors” of each user is very significant. Thus, instead of using Pearson and cosine metrics, researchers have employed a better way by considering data mining and artificial intelligence techniques to select “neighbors” of users for CF in order to facilitate better handling of the challenging problem of CF (data sparseness problem).
- Two important features of this research which clearly distinguish it from other review articles in CF area are: (1) It targets and focuses on the public application platform of recommender systems and (2) it systematically inves-

tigates the research articles through three dimensions: (i) techniques used (including classification, association rule, link analysis, evolutionary computing, regression, matrix factorization, context awareness-based, content-based, mathematical technique, clustering, fuzzy set-based), (ii) benchmark recommendation databases and (iii) two user feedbacks which reflected users’ activities (implicit and explicit feedbacks).

- Research works using practical solutions to derive user’ interests from their implicit behavior are growing every year. Unfortunately, few research works have been published to grasp user interests from social networking activities such as tagging and music listening information of the users to boost the recommendation. Hence, researchers are driven to develop effective techniques for dealing with such implicit data.

## 6 Conclusion and future work

This paper is aimed at providing the descriptions and comparison on public recommendation datasets from different domains in choosing the suitable dataset to analyze and investigate users’ activities which can influence the recommender system developed based on CF technique. Sources of users’ interest evidence on user–item interaction data such as tags, implicit comments, users’ clicking, explicit ratings, interaction record and social contacts are more effective to be used in achieving useful recommendations. A critical analysis was conducted on existing research articles which have employed CF. A hundred and thirty-one research articles have been selected, and they were published between 2010 and 2016. These articles were employed to analyze public recommendation datasets with various types of user preferences about items or resources belonging to domains including Web pages, movies, jokes, books and music tracks. The purpose of this study was to understand applications of user feedbacks, artificial intelligence and data mining techniques on CF recommendation systems by examining the published articles and to afford the community of researchers and practitioners with insight and future direction on CF recommender systems. Hence, this study provides an academic database of the literature between the periods of 2010–2016 covering 36 journals and proposes a classification scheme according to recommendation databases, user feedbacks which reflected users’ activities and artificial intelligence and data mining techniques to classify the published articles.

**Acknowledgements** The authors would like to thank the Research Management Centre of—Universiti Teknologi MARA (UiTM) and the Malaysian Ministry of Education for their support and cooperation



including researches and other individuals who are either directly or indirectly involved in this study.

### Compliance with ethical standards

**Conflicts of interest** Maryam Khanian Najafabadi, Azlinah Hj. Mohamed and Mohd Naz'ri Mahrin declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

- Adomavicius G, Zhang J (2012) Impact of data characteristics on recommender systems performance. *ACM Trans Manag Inf Syst* 3(1):3
- Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 17(2005):734–749
- Ahn HJ, Kang H, Lee J (2010) Selecting a small number of products for effective user profiling in collaborative filtering. *Expert Syst Appl* 37(4):3055–3062
- Anand D, Mampilli BS (2014) Folksonomy-based fuzzy user profiling for improved recommendations. *Expert Syst Appl* 41(5):2424–2436
- Anand D, Bharadwaj KK (2011) Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities. *Expert Syst Appl* 38(5):5101–5109
- Bakshi S, Jagadev AK, Dehuri S, Wang GN (2014) Enhancing scalability and accuracy of recommendation systems using unsupervised learning and particle swarm optimization. *Appl Soft Comput* 15:21–29
- Bauer J, Nanopoulos A (2014) Recommender systems based on quantitative implicit customer feedback. *Decis Support Syst* 68:77–88
- Bellogín A, Castells P, Cantador I (2014) Neighbor selection and weighting in user-based collaborative filtering: a performance prediction approach. *ACM Trans Web* 8(2):12
- Berkovsky S, Kuflik T, Ricci F (2012) The impact of data obfuscation on the accuracy of collaborative filtering. *Expert Syst Appl* 39(5):5033–5042
- Bedi P, Sharma R (2012) Trust based recommender system using ant colony for trust computation. *Expert Syst Appl* 39(1):1183–1190
- Bilge A, Polat H (2013) A comparison of clustering-based privacy-preserving collaborative filtering schemes. *Appl Soft Comput* 13(5):2478–2489
- Birtolo C, Ronca D (2013) Advances in clustering collaborative filtering by means of fuzzy C-means and trust. *Expert Syst Appl* 40(17):6997–7009
- Boratto L, Carta S, Fenu G (2015) Discovery and representation of the preferences of automatically detected groups: exploiting the link between group modeling and clustering. *Future Gener Comput Syst* 64:165–174
- Bobadilla J, Ortega F, Hernando A, Glez-de-Rivera G (2013a) A similarity metric designed to speed up, using hardware, the recommender systems k-nearest neighbors algorithm. *Knowl Based Syst* 51:27–34
- Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013b) Recommender systems survey. *Knowl Based Syst* 46(2013):109–132
- Bobadilla J, Hernando A, Ortega F, Gutiérrez A (2012a) Collaborative filtering based on significances. *Inf Sci* 185(1):1–17
- Bobadilla J, Ortega F, Hernando A, Alcalá J (2011) Improving collaborative filtering recommender system results and performance using genetic algorithms. *Knowl Based Syst* 24(8):1310–1316
- Bobadilla J, Ortega F, Hernando A, Bernal J (2012b) Generalization of recommender systems: collaborative filtering extended to groups of users and restricted to groups of items. *Expert Syst Appl* 39(1):172–186
- Bobadilla J, Serradilla F, Bernal J (2010) A new collaborative filtering metric that improves the behavior of recommender systems. *Knowl Based Syst* 23(6):520–528
- Braida F, Mello CE, Pasinato MB, Zimbrão G (2015) Transforming collaborative filtering into supervised learning. *Expert Syst Appl* 42(10):4733–4742
- Briguez CE, Budan MC, Deagustini CA, Maguitman AG, Capobianco M, Simari GR (2014) Argument-based mixed recommenders and their application to movie suggestion. *Expert Syst Appl* 41(14):6467–6482
- Burke R (2002) Hybrid recommender systems: survey and experiments. *User Model User Adapt Interact* 12(4):331–370
- Casino F, Domingo-Ferrer J, Patsakis C, Puig D, Solanas A (2015) A k-anonymous approach to privacy preserving collaborative filtering. *J Comput Syst Sci* 81(6):1000–1011
- Cai Y, Leung HF, Li Q, Min H, Tang J, Li J (2014a) Typicality-based collaborative filtering recommendation. *IEEE Trans Knowl Data Eng* 26(3):766–779
- Cai Y, Lau RY, Liao SS, Li C, Leung HF, Ma LC (2014b) Object typicality for effective web of things recommendations. *Decis Support Syst* 63:52–63
- Cacheda F, Carneiro V, Fernández D, Formoso V (2011) Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Trans Web* 5(1):2
- Chen MH, Teng CH, Chang PC (2015) Applying artificial immune systems to collaborative filtering for movie recommendation. *Adv Eng Inform* 29(4):830–839
- Cheng LC, Wang HA (2014) A fuzzy recommender system based on the integration of subjective preferences and objective information. *Appl Soft Comput* 18:290–301
- Chen YC, Lin YS, Shen YC, Lin SD (2013a) A modified random walk framework for handling negative ratings and generating explanations. *ACM Trans Intell Syst Technol* 4(1):12
- Chen L, Zeng W, Yuan Q (2013b) A unified framework for recommending items, groups and friends in social media environment via mutual resource fusion. *Expert Syst Appl* 40(8):2889–2903
- Choi K, Suh Y (2013) A new similarity function for selecting neighbors for each target item in collaborative filtering. *Knowl Based Syst* 37:146–153
- Colace F, De Santo M, Greco L, Moscato V, Picariello A (2015) A collaborative user-centered framework for recommending items in online social networks. *Comput Hum Behav* 51:694–704
- Da Costa AF, Manzato MG (2016) Exploiting multimodal interactions in recommender systems with ensemble algorithms. *Inf Syst* 56:120–132
- De Campos LM, Fernández-Luna JM, Huete JF, Rueda-Morales MA (2010) Combining content-based and collaborative recommendations: a hybrid approach based on Bayesian networks. *Int J Approx Reason* 51(7):785–799
- Devi MK, Venkatesh P (2013) Smoothing approach to alleviate the meager rating problem in collaborative recommender systems. *Future Gener Comput Syst* 29(1):262–270
- Elahi M, Ricci F, Rubens N (2013) Active learning strategies for rating elicitation in collaborative filtering: a system-wide perspective. *ACM Trans Intell Syst Technol* 5(1):13
- Eckhardt A (2012) Similarity of users' (content-based) preference models for collaborative filtering in few ratings scenario. *Expert Syst Appl* 39(14):11511–11516
- Feng H, Tian J, Wang HJ, Li M (2015) Personalized recommendations based on time-weighted overlapping community detection. *Inf Manag* 52(7):789–800

- Formoso V, Fernández D, Cacheda F, Carneiro V (2013) Using profile expansion techniques to alleviate the new user problem. *Inf Process Manag* 49(3):659–672
- Gan M, Jiang R (2013) Improving accuracy and diversity of personalized recommendation through power law adjustments of user similarities. *Decis Support Syst* 55(3):811–821
- Geng B, Li L, Jiao L, Gong M, Cai Q, Wu Y (2015) NNIA-RS: a multi-objective optimization based recommender system. *Physica A* 424:383–397
- Gogna A, Majumdar A (2015a) Matrix completion incorporating auxiliary information for recommender system design. *Expert Syst Appl* 42(14):5789–5799
- Ghazarian S, Nematbakhsh MA (2015) Enhancing memory-based collaborative filtering for group recommender systems. *Expert Syst Appl* 42(7):3801–3812
- Gharibshah J, Jalili M (2014) Connectedness of users-items networks and recommender systems. *Appl Math Comput* 243:578–584
- Ghazanfar MA, Prügel-Bennett A (2014) Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems. *Expert Syst Appl* 41(7):3261–3275
- Ghazanfar MA, Prügel-Bennett A, Szedmak S (2012) Kernel-mapping recommender system algorithms. *Inf Sci* 208:81–104
- Gogna A, Majumdar A (2015b) A comprehensive recommender system model: improving accuracy for both warm and cold start users. *IEEE Access* 3:2803–2813
- Hawalsh A, Fasli M (2014) Utilizing contextual ontological user profiles for personalized recommendations. *Expert Syst Appl* 41(10):4777–4797
- Hernando A, Moya R, Ortega F, Bobadilla J (2014) Hierarchical graph maps for visualization of collaborative recommender systems. *J Inf Sci* 40(1):97–106
- Hernando A, Bobadilla J, Ortega F, Tejedor J (2013) Incorporating reliability measurements into the predictions of a recommender system. *Inf Sci* 218:1–16
- Horsburgh B, Craw S, Massie S (2015) Learning pseudo-tags to augment sparse tagging in hybrid music recommender systems. *Artif Intell* 219:25–39
- Hoseini E, Hashemi S, Hamzeh A (2012) SPCF: a stepwise partitioning for collaborative filtering to alleviate sparsity problems. *J Inf Sci* 38(6):578–592
- Hostler RE, Yoon VY, Guimaraes T (2012) Recommendation agent impact on consumer online shopping: the movie magic case study. *Expert Syst Appl* 39(3):2989–2999
- Hsiao KJ, Kulesza A, Hero AO (2014) Social collaborative retrieval. *IEEE J Sel Top Signal Process* 8(4):680–689
- Hwang WS, Lee HJ, Kim SW, Won Y, Lee MS (2016) Efficient recommendation methods using category experts for a large dataset. *Inf Fusion* 28:75–82
- Huang S, Ma J, Cheng P, Wang S (2015) A hybrid multigroup coclustering recommendation framework based on information fusion. *ACM Trans Intell Syst Technol* 6(2):27
- Huang CL, Yeh PH, Lin CW, Wu DC (2014) Utilizing user tag-based interests in recommender systems for social resource sharing websites. *Knowl Based Syst* 56:86–96
- Hu L, Song G, Xie Z, Zhao K (2014) Personalized recommendation algorithm based on preference features. *Tsinghua Sci Technol* 19(3):293–299
- Javari A, Jalili M (2015) Accurate and novel recommendations: an algorithm based on popularity forecasting. *ACM Trans Intell Syst Technol* 5(4):56
- Kaššák O, Kompan M, Bieliková M (2015) Personalized hybrid recommendation for group of users: top-N multimedia recommender. *Inf Process Manag*
- Kagita VR, Pujari AK, Padmanabhan V (2015) Virtual user approach for group recommender systems using precedence relations. *Inf Sci* 294:15–30
- Kaleli C (2014) An entropy-based neighbor selection approach for collaborative filtering. *Knowl Based Syst* 56:273–280
- Kolomvatsos K, Anagnostopoulos C, Hadjiefthymiades S (2014) An efficient recommendation system based on the optimal stopping theory. *Expert Syst Appl* 41(15):6796–6806
- Krestel R, Fankhauser P (2012) Personalized topic-based tag recommendation. *Neurocomputing* 76(1):61–70
- Kim HN, El Saddik A (2015) A stochastic approach to group recommendations in social media systems. *Inf Syst* 50:76–93
- Kim H, Kim HJ (2014) A framework for tag-aware recommender systems. *Expert Syst Appl* 41(8):4000–4009
- Kim HN, Ha I, Lee KS, Jo GS, El-Saddik A (2011a) Collaborative user modeling for enhanced content filtering in recommender systems. *Decis Support Syst* 51(4):772–781
- Kim HN, El-Saddik A, Jo GS (2011b) Collaborative error-reflected models for cold-start recommender systems. *Decis Support Syst* 51(3):519–531
- Koren Y (2010) Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Trans Knowl Discov Data* 4(1):1
- Langseth H, Nielsen TD (2015) Scalable learning of probabilistic latent models for collaborative filtering. *Decis Support Syst* 74:1–11
- Last.fm dataset, the official song tags and song similarity collection for the million song dataset. <http://labrosa.ee.columbia.edu/millionsong/lastfm>, (June 2014)
- Langseth H, Nielsen TD (2012) A latent model for collaborative filtering. *Int J Approx Reason* 53(4):447–466
- Liu J, Sui C, Deng D, Wang J, Feng B, Liu W, Wu C (2016) Representing conditional preference by boosted regression trees for recommendation. *Inf Sci* 327:1–20
- Liu W, Wu C, Feng B, Liu J (2015) Conditional preference in recommender systems. *Expert Syst Appl* 42(2):774–788
- Liu J, Wu C, Xiong Y, Liu W (2014a) List-wise probabilistic matrix factorization for recommendation. *Inf Sci* 278:434–447
- Liu H, Hu Z, Mian A, Tian H, Zhu X (2014b) A new user similarity model to improve the accuracy of collaborative filtering. *Knowl Based Syst* 56:156–166
- Liu J, Wu C, Liu W (2013) Bayesian probabilistic matrix factorization with social relations and item contents for recommendation. *Decis Support Syst* 55(3):838–850
- Liu Z, Qu W, Li H, Xie C (2010) A hybrid collaborative filtering recommendation mechanism for P2P networks. *Future Gener Comput Syst* 26(8):1409–1417
- Lika B, Kolomvatsos K, Hadjiefthymiades S (2014) Facing the cold start problem in recommender systems. *Expert Syst Appl* 41(4):2065–2073
- Li X, Chen H (2013) Recommendation as link prediction in bipartite graphs: a graph kernel-based machine learning approach. *Decis Support Syst* 54(2):880–890
- Lv G, Hu C, Chen S (2015) Research on recommender system based on ontology and genetic algorithm. *Neurocomputing*
- Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. *Decis Support Syst* 74:12–32
- Luo X, Zhou M, Xia Y, Zhu Q (2014) An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Trans Industr Inf* 10(2):1273–1284
- Luo X, Xia Y, Zhu Q (2012) Incremental collaborative filtering recommender based on regularized matrix factorization. *Knowl Based Syst* 27:271–280
- Ma H, Zhou TC, Lyu MR, King I (2011) Improving recommender systems by incorporating social contextual information. *ACM Trans Inf Syst* 29(2):9
- Mehta S, Banati H (2014) Context aware filtering using social behavior of frogs. *Swarm Evol Comput* 17:25–36
- Moreno MN, Segrera S, López VF, Muñoz MD, Sánchez ÁL (2016) Web mining based framework for solving usual problems in rec-



- ommender systems. A case study for movies' recommendation. *Neurocomputing* 176:72–80
- Moradi P, Ahmadian S, Akhlaghian F (2015) An effective trust-based recommendation method using a novel graph clustering algorithm. *Physica A* 436:462–481
- Movahedian H, Khayyambashi MR (2014) Folksonomy-based user interest and disinterest profiling for improved recommendations: an ontological approach. *J Inf Sci* 40(5):594–610
- Najafabadi MK, Mahrin MNR (2016) A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback. *Artif Intell Rev* 45(2):167–201
- Najafabadi MK, Mahrin MNR, Chuprat S, Sarkan HM (2017) Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Comput Hum Behav* 67:113–128
- Nakatsuji M, Toda H, Sawada H, Zheng JG, Hendler JA (2016) Semantic sensitive tensor factorization. *Artif Intell* 230:224–245
- Nakatsuji M, Fujiwara Y (2014) Linked taxonomies to capture users' subjective assessments of items to facilitate accurate collaborative filtering. *Artif Intell* 207:52–68
- Nikolakopoulos AN, Kouneli MA, Garofalakis JD (2015) Hierarchical itemspace rank: exploiting hierarchy to alleviate sparsity in ranking-based recommendation. *Neurocomputing* 163:126–136
- Pan W, Liu Z, Ming Z, Zhong H, Wang X, Xu C (2015a) Compressed knowledge transfer via factorization machine for heterogeneous collaborative recommendation. *Knowl Based Syst* 85:234–244
- Pan W, Zhong H, Xu C, Ming Z (2015b) Adaptive bayesian personalized ranking for heterogeneous implicit feedbacks. *Knowl Based Syst* 73:173–180
- Pan W, Yang Q (2013) Transfer learning in heterogeneous collaborative filtering domains. *Artif Intell* 197:39–55
- Park DH, Kim HK, Choi IY, Kim JK (2012) A literature review and classification of recommender systems research. *Expert Syst Appl* 39(11):10059–10072
- Peng F, Lu J, Wang Y, Yi-Da Xu R, Ma C, Yang J (2016) N-dimensional Markov random field prior for cold-start recommendation. *Neurocomputing* 191:187–199
- Patra BK, Launonen R, Ollikainen V, Nandi S (2015) A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data. *Knowl Based Syst* 82:163–177
- Pirasteh P, Hwang D, Jung JJ (2015) Exploiting matrix factorization to asymmetric user similarities in recommendation systems. *Knowl Based Syst* 83:51–57
- Polatidis N, Georgiadis CK (2016) A multi-level collaborative filtering method that improves recommendations. *Expert Syst Appl* 48:100–110
- Ranjbar M, Moradi P, Azami M, Jalili M (2015) An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems. *Eng Appl Artif Intell* 46:58–66
- Ramezani M, Moradi P, Akhlaghian F (2014) A pattern mining approach to enhance the accuracy of collaborative filtering in sparse data domains. *Physica A* 408:72–84
- Rana C, Jain SK (2014) An evolutionary clustering algorithm based on temporal features for dynamic recommender systems. *Swarm Evol Comput* 14:21–30
- Rafeh R, Bahrehmand A (2012) An adaptive approach to dealing with unstable behaviour of users in collaborative filtering systems. *J Inf Sci* 38(3):205–221
- Ren Y, Li G, Zhang J, Zhou W (2013) Lazy collaborative filtering for data sets with missing values. *IEEE Trans Cybern* 43(6):1822–1834
- Salah A, Rogovschi N, Nadif M (2016) A dynamic collaborative filtering system via a weighted clustering approach. *Neurocomputing* 175:206–215
- Shambour Q, Lu J (2015) An effective recommender system by unifying user and item trust information for B2B applications. *J Comput Syst Sci* 81(7):1110–1126
- Shambour Q, Lu J (2012) A trust-semantic fusion-based recommendation approach for e-business applications. *Decis Support Syst* 54(1):768–780
- Shang MS, Zhang ZK, Zhou T, Zhang YC (2010) Collaborative filtering with diffusion-based similarity on tripartite graphs. *Physica A* 389(6):1259–1264
- Shinde SK, Kulkarni U (2012) Hybrid personalized recommender system using centering-bunching based clustering algorithm. *Expert Syst Appl* 39(1):1381–1387
- Sun Z, Han L, Huang W, Wang X, Zeng X, Wang M, Yan H (2015) Recommender systems based on social networks. *J Syst Softw* 99:109–119
- Tan S, Bu J, Qin X, Chen C, Cai D (2014) Cross domain recommendation based on multi-type media fusion. *Neurocomputing* 127:124–134
- Toledo RY, Mota YC, Martínez L (2015) Correcting noisy ratings in collaborative recommender systems. *Knowl-Based Syst* 76:96–108
- Tyagi S, Bharadwaj KK (2013) Enhancing collaborative filtering recommendations by utilizing multi-objective particle swarm optimization embedded association rule mining. *Swarm Evol Comput* 13:1–12
- Tsai CF, Hung C (2012) Cluster ensembles in collaborative filtering recommendation. *Appl Soft Comput* 12(4):1417–1425
- Umyarov A, Tuzhilin A (2011) Using external aggregate ratings for improving individual recommendations. *ACM Trans Web* 5(1):3
- Wang Z, Yu X, Feng N, Wang Z (2014a) An improved collaborative movie recommendation system using computational intelligence. *J VisLang Comput* 25(6):667–675
- Wang S, Sun J, Gao BJ, Ma J (2014b) VSRank: a novel framework for ranking-based collaborative filtering. *ACM Trans Intell Syst Technol* 5(3):51
- Wang J, Ke L (2014) Feature subspace transfer for collaborative filtering. *Neurocomputing* 136:1–6
- Wen Y, Liu Y, Zhang ZJ, Xiong F, Cao W (2014) Compare two community-based personalized information recommendation algorithms. *Physica A* 398:199–209
- Wu H, Yue K, Pei Y, Li B, Zhao Y, Dong F (2016) Collaborative topic regression with social trust ensemble for recommendation in social media systems. *Knowl Based Syst*
- Wu ML, Chang CH, Liu RZ (2014) Integrating content-based filtering with collaborative filtering using co-clustering with augmented matrices. *Expert Syst Appl* 41(6):2754–2761
- Xie F, Chen Z, Shang J, Feng X, Li J (2015) A link prediction approach for item recommendation with complex number. *Knowl Based Syst* 81:148–158
- Xie F, Chen Z, Shang J, Fox GC (2014) Grey forecast model for accurate recommendation in presence of data sparsity and correlation. *Knowl Based Syst* 69:179–190
- Xu Y, Yin J (2015) Collaborative recommendation with user generated content. *Eng Appl Artif Intell* 45:281–294
- Yakut I, Polat H (2012) Estimating NBC-based recommendations on arbitrarily partitioned data with privacy. *Knowl Based Syst* 36:353–362
- Yan S, Zheng X, Chen D, Wang Y (2013) Exploiting two-faceted web of trust for enhanced-quality recommendations. *Expert Syst Appl* 40(17):7080–7095
- Yera R, Castro J, Martínez L (2016) A fuzzy model for managing natural noise in recommender systems. *Appl Soft Comput* 40:187–198
- Yu H, Kim S (2012) SVM tutorial—classification, regression and ranking handbook of natural computing. Springer, Berlin, pp 479–506
- Zahra S, Ghazanfar MA, Khalid A, Azam MA, Naeem U, Prugel-Bennett A (2015) Novel centroid selection approaches for

- KMeans-clustering based recommender systems. *Inf Sci* 320:156–189
- Zeng W, Zhu YX, Lü L, Zhou T (2011) Negative ratings play a positive role in information filtering. *Physica A* 390(23):4486–4493
- Zhao W, Guan Z, Liu Z (2015) Ranking on heterogeneous manifolds for tag recommendation in social tagging services. *Neurocomputing* 148:521–534
- Zhou X, He J, Huang G, Zhang Y (2015) SVD-based incremental approaches for recommender systems. *J Comput Syst Sci* 81(4):717–733
- Zhang J, Peng Q, Sun S, Liu C (2014) Collaborative filtering recommendation algorithm based on user preference derived from item domain features. *Physica A* 396:66–76
- Zhang Z, Lin H, Liu K, Wu D, Zhang G, Lu J (2013) A hybrid fuzzy-based personalized recommender system for telecom products/services. *Inf Sci* 235:117–129
- Zhang Z, Zhao K, Zha H (2012) Inducible regularization for low-rank matrix factorizations for collaborative filtering. *Neurocomputing* 97:52–62
- Zhang ZK, Zhou T, Zhang YC (2010) Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs. *Physica A* 389(1):179–186
- Zhu T, Ren Y, Zhou W, Rong J, Xiong P (2014) An effective privacy preserving algorithm for neighborhood-based collaborative filtering. *Future Gener Comput Syst* 36:142–155