Collaborative Filtering in Recommender Systems: Technicalities, Challenges, Applications, and Research Trends

Chapter · July 2020

DOI: 10.1201/9781003007210-8

3 authors:

Pradeep Kumar Singh
Kamla Nehru Institute of Technology, Sultanpur

Pijush Kanti Dutta Pramanik
Jodhpur Institute of Engineering and Technology

Prasenjit Choudhury
National Institute of Technology, Durgapur

CHAPTER 8

Collaborative Filtering in Recommender Systems: Technicalities, Challenges, Applications, and Research Trends

PRADEEP KUMAR SINGH, PIJUSH KANTI DUTTA PRAMANIK, and PRASENJIT CHOUDHURY

Department of Computer Science and Engineering, National Institute of Technology Durgapur, India
E-mail: pijushjld@yahoo.co.in (P. K. D. Pramanik)

ABSTRACT

The rapid development and extensive use of recommender systems (RSs) have changed the face of online service experience. The enormous data generated and the complexity involved in analyzing these data for an effective recommendation has attracted researchers from different domains, especially data analytics. In this direction, collaborative filtering (CF) has been the most widely considered approach. The objective of this chapter is to represent a comprehensive study of the CF. The chapter is written in a tutorial fashion so that it can be followed by the readers who are the beginners in this field or unfamiliar with the RS. Different aspects of CF such as classifications, approaches, data extraction methods, similarity metrics, prediction approaches, and performance metrics are studied meticulously. The application of CF in different domains is reviewed. More than 100 research articles are surveyed and categorized according to the application domain of CF they have covered. The challenges involved in the successful adoption of the CF are validly examined. In addition to a brief survey on CF, a systematic survey, considering 277 related papers, on current research trends (2011–2017) on CF is presented. A special discussion of future directions of CF is also stated.

8.1 INTRODUCTION

The recommender system (RS) has become the backbone of e-commerce. In addition to the basic searching facility, every e-commerce portal is opting for RS as an integral part of it. As the e-commerce market is continuously growing, more products and services are made available for online purchasing. Among this sea of online products and services, customers find it very difficult to find the appropriate item for themselves. The e-commerce vendors have come up with the solution for helping the customer to find the appropriate item by recommending the item to the customer which he/she might like or desire. The technical scheme that enables the recommendation process is termed as a RS. RS attempts to predict the items that prospective online buyers may prefer and recommend these anticipated items. Unlike the search tools where people ferret out the online products, recommendation engines aim to consciously catch the attention of the users to the likable products. The overall objective is to bail out users form explicit and tiresome searching and to improve the online shopping experience. The success of e-business largely depends on the intelligence of the algorithm used for a product recommendation. Hence, in the age of digital marketing, it is crucial for online stores to adopt intelligent recommendation techniques in order to sustain in the market competition. Companies like Flipkart, Amazon, eBay, Netflix, MovieLens, IMDb, etc. use RS extensively and innovatively as a core part of their business innovation and exploration.

RS assesses the preference and choice of users by tracking and analyzing their buying and browsing habits and history. The tool used for this purpose is generally known as the filtering approach. Filtering Approach is a method which makes the selective presentation among an array of available commodities using various filtering parameters which makes the filtered products more favorable to the recipient. There are several filtering approaches of an RS in the literature such as: (i) content-based (CB), (ii) CF, (iii) hybrid filtering, (HB), (iv) knowledge-based (KB), and (v) context-aware (CA) as shown in Figure 8.1. CF is more popular filtering approach among these over the past few years (Burke, 2002). CF works on the fact of comparison of user activities, purchases, ratings, preferences, and using this data for comparison and subsequent analysis. The customers prefer products that have been liked or given higher preference by people with a similar taste (Deshpande and Karypis, 2004). Hence, the CF is the most important in this regard.

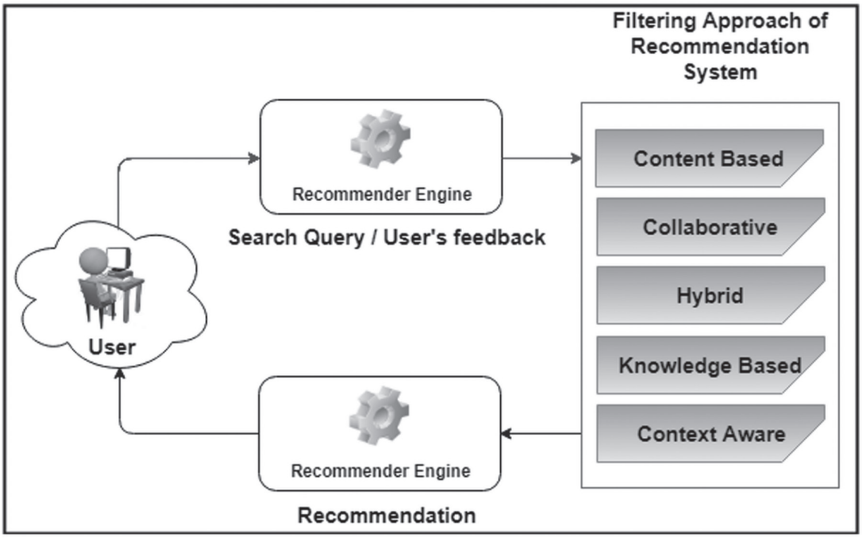

FIGURE 8.1 A general framework for the recommender system.

CF attempts to guess the target user's interest by assessing the top-n similar users' interests on the basis of the assumption that if two persons' choice matches for certain things, it is highly probable that their choices will match for other things as well. The CF suggests that the ratings given by similar users tend to be substantively similar and similar items also tend to receive similar ratings. The CF algorithms exploit this assumption for the recommendation and actually use the similarity value to predict user preferences. Similarity allows the recommender engines to find the user purchase patterns as well as allowing them to understand how those rating patterns are similar to other users. All rating information is stored into memory for prediction in making of the top-n list for recommendation. Similar users or items have a major contribution in the prediction phase of CF-based RS. The top-n list of the recommended items affected if similarity provides the wrong result. CF uses two approaches for considering similarity (Singh, Pramanik, and Choudhury, 2018):

i. **User Similarity-Based Approach (USBA):** Tries to predict the rating based on rating information collected from similar users; and
ii. **Item Similarity-Based Approach (ISBA):** Uses the same idea as USBA but, it uses item similarity instead of user similarity.

After computation of the similarity value of users' and items,' prediction approaches are used to predict the ratings of a target item for a target user. Furthermore, CF-based RS generates a list of top-n items and recommend to the target user. The top-n list of the recommended items affected if similarity provides the wrong result.

The structure of the remaining chapter is as follows. Several application domains of RS are stated in Section 8.2. More than 100 papers are surveyed and categorized according to the application domain they address and the filtering approach they used. Section 8.3 mentions different CF approaches. The working principle of neighborhood-based CF is explained in Section 8.4. Section 8.5 introduces the data extraction methods used in CF. Section 8.6 explains the similarity metrics used in CF algorithms while a comparative study on different similarity metrics has been presented in Section 8.7. Sections 8.8 and 8.9 discuss the prediction approaches and performance metrics used in CF-based RSs, respectively. The challenges in CF-based RSs, as well as the security and trust attacks, are discussed meticulously in Section 8.10. Section 8.11 mentions some of the notable works on CF. Section 8.12 presents the research trends in CF-based RS. 277 related papers are studied for this purpose. The future scope of CF-based Rs is discussed briefly in Section 8.13. And finally, the conclusion of the paper is presented in Section 8.14.

8.2 APPLICATIONS OF RECOMMENDER SYSTEMS (RSS)

RS has found many application domains. Below a few of them mentioned:

1. **E-Government:** It is the medium by which the government makes use of the internet and computers to deliver services to the citizens. It is the most effective modern method which helps the government to connect people across the country.
2. **E-Library and E-Learning:** It is the medium by which the system of education is provided to individuals completely over the internet with the help of electronic devices. It is a formal way of delivering education through electronic resources.
3. **E-Tourism:** It is the digital process which is implemented to achieve the strategy of e-commerce in tourism. It also helps to keep the client connected with the travel partners. E-tourism leads to an excellent medium of marketing and promotions of a company.

4. **E-Resource:** It refers to any resource or collection preserved in electronic format. This type of resource requires an electronic device to access the information. Since the resource is available in the electronic format, huge sets of data can be available for access.
5. **E-Commerce:** Any commercial transactions, exchange, or transfer of data which is carried on via the internet is termed as electronic commerce or e-commerce. It is the fastest method of conducting business in the modern world and thus leads to the digitization of society.

The performance of these applications can be improved using memory-based CF. People can easily provide their opinion about the services of these applications and due to this; they can be received more personalized, diverse, novel, and accurate recommendations. Table 8.1 lists different application domains of RS. It also mentions the notable research works towards these domains and also the filtering approaches used in those works.

TABLE 8.1 Application Domains of Recommender Systems, Notable Works Towards That Domain, and Filtering Approach Used

Application Domain	**Filtering Approach**	**Recommender System**
E-government	Knowledge-based	Meo, Quattrone, and Ursino, 2008; Teran and Meier, 2010; Esteban et al., 2014; Cornelis et al., 2007
	Collaborative	Guo and Lu, 2007
	Collaborative, Hybrid, Knowledge-based	Wu, Zhang, and Lu, 2015; Lu et al., 2010
E-library and e-learning	Content-based, Collaborative, Hybrid	Balabanović and Shoham, 1997; Renda and Straccia, 2005
	Hybrid, Knowledge-based	Porcel, López-Herrera, and Herrera-Viedma, 2009; Porcel, Herrera-Viedma, and Moreno, 2009; Porcel and Herrera-Viedma, 2010; Serrano-Guerrero et al., 2011; Cobos et al., 2013
	Knowledge-based, Content-based	Zaíane, 2002; Chen, Duh, and Liu, 2004; Chen and Duh, 2008; Capuano et al., 2014; Farzan and Brusilovsky, 2006; Santos et al., 2014; Lu, 2004; Biletskiy et al., 2009
E-tourism	Knowledge-based	Burke, Hammond, and Young, 1996; Fesenmaier et al., 2003; García-Crespo et al., 2011

TABLE 8.1 *(Continued)*

Application Domain	Filtering Approach	Recommender System
	Knowledge-based, Collaborative, Context-aware, Hybrid	Avesani, Massa, and Tiella, 2005; Martínez, Rodríguez, and Espinilla, 2009; Ruotsalo et al., 2013; García-Crespo et al., 2009; Console et al., 2003; Moreno et al., 2013
	Content-based, Collaborative, Hybrid, Demographic	Schiaffino and Amandi, 2009; Luz et al., 2013; Baraglia et al., 2012
	Context-aware	Tung and Soo, 2004; Pashtan et al., 2003; Rikitianskii, Harvey, and Crestani, 2014; Xing et al., 2013
	Collaborative, Context-aware	Yanga and Hwang, 2013
E-resource	Content-based	Jinni, 2017; Rotten Tomatoes, 2017; IMDb, 2017; Asnicar and Tasso, 1997; ACRnews, 2017; Chesnevar and Maguitman, 2004; Park, 2013
	Collaborative	Ali and Van Stam, 2004; Konstan et al., 1997; FoxTrit, 2017; Miller, Konstan, and Riedl, 2004; Hauver and French, 2001; Marcel et al., 2003; Lee, Cho, and Kim, 2010; TASTEKiD, 2017; nanoCROWD, 2017; Movielens, 2017
	Context-aware, Collaborative	Braunhofer, Kaminskas, and Ricci, 2013; Baltrunas et al., 2012; Levandoski et al., 2012; Natarajan, Shin, and Dhillon, 2013; Oh et al., 2014
	Collaborative, Knowledge-based	Zhang, Zhou, and Zhang, 2011; Hayes and Cunningham, 2001; Sánchez et al., 2011; Boutet et al., 2013
	Content-based, Collaborative, Hybrid	Smyth and Cotter, 2000; Blanco-Fernández et al., 2006; Salter and Antonopoulos, 2006; Melville, Mooney, and Nagarajan, 2002; Domingues et al., 2013; Christou, Amolochitis, and Tan, 2016; Parra, Brusilovsky, and Trattner, 2014; Amolochitis, Christou, and Tan, 2014
	Knowledge-based, Content-based	Jäschke et al., 2007; Hotho et al., 2006; Celma and Serra, 2008; Bjelica, 2010; Moukas and Maes, 1998; Billsus and Pazzani, 2000; Nguyen, Lu, and Lu, 2014; Martín-Vicente et al., 2012; Zhang et al., 2012

TABLE 8.1 *(Continued)*

Application Domain	Filtering Approach	Recommender System
E-commerce	Knowledge-based, Demographic	Garfinkel et al., 2006; Mccarthy et al., 2004; Cao and Li, 2007; Hu et al., 2012; Zhao et al., 2016; Zhao et al., 2014;
	Knowledge-based, Content-based	Burke, 1999; Nanopoulos et al., 2010; Zhang et al., 2013; Yin et al., 2014
	Collaborative, Hybrid	Pratikshashiv, 2015; Lawrence et al., 2001; Chen and Pu, 2012; Walter et al., 2012; Liu and Karger, 2015

8.3 COLLABORATIVE FILTERING (CF) APPROACHES

The CF technique can be classified into two categories as shown in Figure 8.2 (Su and Khoshgoftaar, 2009):

i. **Model-Based CF:** It uses some algorithms of machine learning (ML) like Bayesian network clustering and rule-based approaches which builds a model on user-item rating dataset and then recommends items to the user.
ii **Neighborhood/Memory-Based CF:** Similarity and prediction computation are the two major steps used in this category of CF.

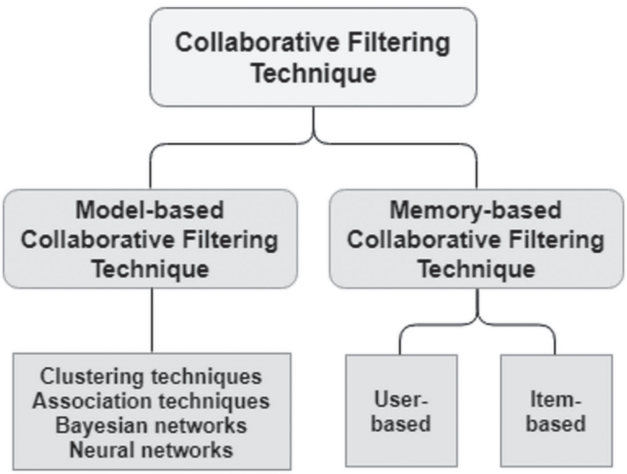

FIGURE 8.2 Collaborative filtering techniques.

8.4 WORKING PRINCIPLE OF NEIGHBORHOOD-BASED COLLABORATIVE FILTERING (CF)

Figure 8.3 shows the conceptual framework of neighborhood-based CF (Yang et al., 2016). Neighborhood CF defines the closest neighbors using the following two algorithms:

- **User-Based CF Algorithm:** User similarity metric is used to find the nearest neighbors. The rating value of these neighbors and their similarity values are utilized in the prediction of unrated items of users for the formation of the Top-n list in the recommendation.
- **Item-Based CF Algorithm:** In the item-based CF algorithm, the nearest neighbors are determined using the similarity values of items, and these similarity values and rating values of these neighbors are used in the formation of the recommendation list to the user.

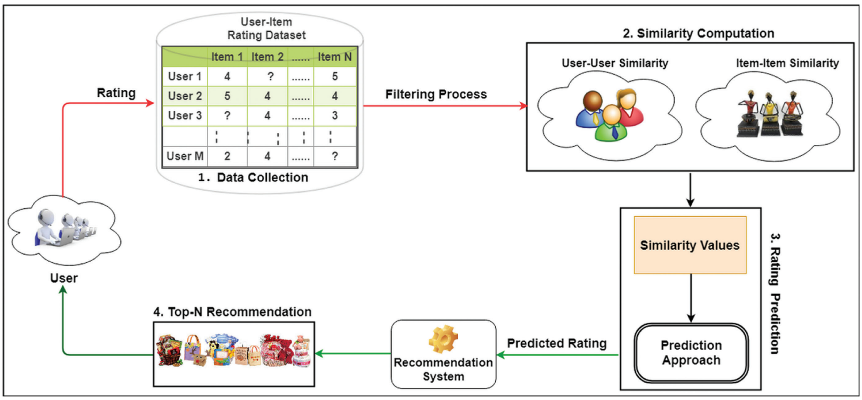

FIGURE 8.3 A conceptual framework for neighborhood-based collaborative filtering.

Table 8.2 shows the descriptions of the notations used in this chapter.

8.5 DATA EXTRACTION METHODS USED IN COLLABORATIVE FILTERING (CF)

CF uses ratings in the recommendation process. Two types of ratings have been used in CF for a recommendation-explicit rating and implicit rating (Li et al., 2018).

TABLE 8.2 Notations and Their Descriptions

Notation	Description
Sim(i,j)	Similarity between two items i and j
$R_{u,i}$	Rating value of user u on item i
$\bar{R}_u$	Average or mean rating value of user u
$\lvert U_{ij} \rvert$	Number of ratings of user u on both items i and j
$\widehat{r_{ui}}$	Predicted rating value of user u on item i
$\bar{r}_i$	Average or mean rating value of item i

1. **Explicit:** These ratings are the specific rating that a user gives to a product (for example, a user rates a book 3 on a scale of 1 to 5). These explicit ratings are directly used in the extractions of users' interest for future recommendation. The disadvantage of explicit data is that it makes user responsible for data collection and future rating prediction who hardly takes interest to give a rating on a particular item.
2. **Implicit:** These ratings are collected by logging the user's data generated while browsing the website. Implicit data are easier to collect as it does not put any pressure on the user to rate the products on the site. However, dealing with an implicit rating is very complicated as it is hard to find the users' preferences from these collected users' browsing data. Using these collected ratings (explicit or implicit); RSs predict the unknown ratings of the user based on different similarity metrics and these predicted ratings used in the recommendation process.

8.6 SIMILARITY METRICS USED IN COLLABORATIVE FILTERING (CF) ALGORITHMS

There are various similarity metrics used in the CF to find the nearest neighbors and similarity values (Sarwar et al., 2001; Bilge and Kaleli, 2014; Bobadilla et al., 2012). The metrics used in the item-based CF are:

1. **Cosine Similarity (CS):** The function of cosine distance finds similarity between two samples by studying the cosine of the angle between them to quantify the similarity. The similarity values are in the range [1, –1], where 1 shows the maximum similarity and –1

depicts no similarity. CS between two items i and j, is calculated using:

$$sim(i, j) = cos(i, j) = \frac{i.j}{||i||^2 * ||j||^2}$$

Here, i and j identifies the dot-product between two items.

2. **Adjusted Cosine Similarity (ACS):** It is similar to cosine distance, also caters to the individual user's rating. To achieve this, it subtracts the average user rating from the individual ratings to get uniformity. It is computed by:

$$sim(i, j) = \frac{\sum_{u \in U} \left(R_{u,i} - \overline{R}_u\right)(R_{u,j} - \overline{R}_u)}{\sqrt[2]{\sum_{u \in U} \left(R_{u,i} - \overline{R}_u\right)^2} \sqrt[2]{\sum_{u \in U} \left(R_{u,j} - \overline{R}_u\right)^2}}$$

Here, $R_{u,i}$ and $R_{u,j}$ are the rating value of user u on two items i and j, respectively. $\overline{R}_{u,}$ shows the average rating value of user u.

3. **Pearson Correlation (PC):** It is the most popular Similarity Metric and is widely used in various experiments. The similarity in it is represented between [1, –1], where 1 shows the maximum similarity and –1 depicts no similarity. Similarity using PC, in Item-based CF algorithm is:

$$sim(i, j) = \frac{\sum_{u \in U} \left(R_{u,i} - \overline{R}_i\right)(R_{u,j} - \overline{R}_j)}{\sqrt[2]{\sum_{u \in U} \left(R_{u,i} - \overline{R}_i\right)^2} \sqrt[2]{\sum_{u \in U} \left(R_{u,j} - \overline{R}_j\right)^2}}$$

Here, $\overline{R}_i$ and $\overline{R}_j$ are the mean rating value of two items i and j, respectively.

4. **Jaccard Similarity (JS):** It considers only all the common ratings between items in spite of the absolute rating value of items. It is calculated by [1]:

$$sim(i, j) = \frac{|R_i| \cap |R_j|}{|R_i| \cup |R_i|}$$

5. **Spearman Correlation (SC):** It is calculated just like PC, but it uses the respective rank of the actual rating value. The equation of calculating similarity value by SC is as follows:

$$sim(i, j) = \frac{\sum_{u \in U} \left(k_{u,i} - \overline{k}_i\right)(k_{u,j} - \overline{k}_j)}{\sqrt[2]{\sum_{u \in U} \left(k_{u,i} - \overline{k}_i\right)^2} \sqrt[2]{\sum_{u \in U} \left(k_{u,j} - \overline{k}_j\right)^2}}$$

Here, $k_{u,i}$ *and* $k_{u,j}$ show the respective rank of items *i* and *j* of rating value of user u. $\overline{k}_i$ and $\overline{k}_j$ denote the average rank of items i and j respectively.

6. **Euclidean Distance (ED):** The Euclidian distance uses the under-root of the squared sum of the difference between individual ratings of the two samples whose similarity we want to find. The distance gives an insight into how different the rating patterns are:

$$sim(i, j) = \sqrt[2]{\frac{\sum_{u \in U_{i,j}} (r_{i,u} - r_{j,u})^2}{|U_{ij}|}}$$

7. **Manhattan Distance (MD):** The equation to find similarity using MD is given below.

$$sim(i, j) = \frac{\sum_{u \in U_{i,j}} (r_{i,u} - r_{j,u})^1}{|U_{ij}|}$$

8. **Mean Squared Distance (MSD):** It is similar to ED only difference is that the whole Euclidian distance is squared, thus removing under-root from the mathematics thus making calculations easier. The equation of MSD for calculating the similarity value is shown by:

$$sim(i, j) = \frac{\sum_{u \in U_{i,j}} (r_{i,u} - r_{j,u})^2}{|U_{ij}|}$$

8.7 COMPARISON OF DIFFERENT SIMILARITY METRICS

Purpose of the RS is to provide optimized and personalized products recommendation to the users. RS has various options to choose similarity metrics (in literature), which gives various lists of top-n recommendation items.

Table 8.3 illustrates the list of top-10 similar movies of target movie id 1, using the traditional similarity measures. It can be observed that every similarity measure has a different top-10 movies list. Hence, there is a need for a comparative study on similarity metrics to enhance the accuracy of CF. On the basis of a comparative study of similarity measures; we can improve the accuracy of RS because each similarity measures have some limitations. For constructing Table 8.3, we collect the MovieLens dataset, i.e., ml–20 m. The filtering criteria have been applied to minimize the sparsity. These filtering criteria are:

i. Select the users who provide ratings to a minimum of 100 numbers of movies.
ii. Select the movies which are received to a minimum of 1000 number of ratings.

TABLE 8.3 List of Top-10 Similar Movies of Target Movie id 1, Using the Traditional Similarity Measures

Similarity Metric	Top-10 Similar Movies									
Pearson Correlation	926	1272	1276	623	730	869	215	95	999	1301
Cosine Distance	1276	1192	1027	956	1079	949	352	1088	215	915
Adjusted Cosine Distance	1276	352	1027	1079	926	1088	580	401	15	729
Mean Squared Distance	1276	352	1088	926	1079	729	1027	1142	1239	354
Euclidean Distance	1276	352	1088	926	1079	729	1027	1142	1239	354
Manhattan Distance	1276	29	15	1088	1079	352	85	1239	1182	119
Spearman Correlation	1276	352	1027	926	1079	869	1088	915	949	1142

8.8 PREDICTION APPROACHES

Different prediction approaches have been utilized in the prediction phase of CF-based RS (Sarwar et al., 2001; Wu et al., 2013; Herlocker et al., 1999). These methods for item-based CF are:

1. **Mean Centering (MC):** In this approach, the mean of the target item's rating is added with the weighted average (WA) of subtraction between all available ratings of top-n similar items with their respective mean is done, using as weights the correlation values computed by the similarity measures. The equation of the MC approach to predict the rating as given below:

$$\widehat{r_{ui}} = \overline{r_i} + \frac{\sum_{j \in N_u(i)} sim(i,j)(r_{ju} - \overline{r_j})}{\sum_{j \in N_u(i)} |sim(i,j)|}$$

2. **Weighted Average (WA):** To predict the rating for a target item, a WA of all available ratings of top-n similar items is calculated using weights as the correlation values computed by the similarity measures. The equation to predict rating using WA is:

$$\widehat{r_{ui}} = \frac{\sum_{j \in N_u(i)} sim(i,j) r_{ju}}{\sum_{j \in N_u(i)} |sim(i,j)|}$$

3. **Z Score (ZS):** Using the standard deviation of rating of the item in MC, the equation of Z-score for item-based CF is as follows:

$$\widehat{r_{ui}} = \overline{r_i} + \sigma_i \frac{\sum_{j \in N_u(i)} sim(i,j)(r_{ju} - \overline{r_j}) / \sigma_i}{\sum_{j \in N_u(i)} |sim(i,j)|}$$

Here, σ_i represents the standard deviation of the rating value of item i.

These prediction approaches have some limitations in the sparse dataset. Hence, for the more personalized and accurate recommendation, there is a need for a comparative study of prediction approaches in CF. By mutually exchanging i and j with u and v respectively, we can get the computational equation of SMs and PAs in user-based CF.

8.9 PERFORMANCE METRICS

Various performance metrics have been used in the literature of CF-based RSs (Singh, Pramanik, and Choudhury, 2018; Samundeeswary and Krishnamurthy, 2017; Zuva and Zuva, 2017; Pampın, Jerbi, and O'Mahony, 2015):

1. **Mean Absolute Error (MAE):** It is the amount of error in the rating prediction. The equation for calculating MAE is:

$$MAE = \frac{\sum_{i=1}^{N} | p_i - \hat{q}_i |}{N}$$

Here, < p_i and $\hat{q}_j$ > denote each original ratings-predicted ratings pair and, N shows the total number pairs that represent original and predicted ratings pair.

2. **Root Mean Square Error (RMSE):** After some modification in the equation of MAE, we get the equation of RMSE as follows:

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{N} (p_i - \hat{q}_i)^2}{N}}$$

3. **Coverage:** Item coverage is the percentage of items included in the recommendation list over the number of potential items:

$$Coverage_{item} = \frac{n}{N} * 100$$

User coverage is the percentage of users for whom the recommender was able to generate a recommendation list over the number of potential users.

$$Coverage_{user} = \frac{u}{U} * 100$$

Catalog coverage is the percentage of recommended user-item pairs over the total number of potential pairs. The number of recommended user-item pairs can be represented by the length of the recommended lists L.

$$Coverage_{user} = \frac{length(L)}{N * U} * 100$$

And finally, user interaction coverage is the percentage of rated predictions over the total number of ratings. Here, n, and u represent the number of items in the recommendation list and the number of users involved in the generation of this recommendation list respectively. N and U denote the number of potential items and the number of potential users, whereas L shows the number of user-item in the recommendation list.

4. **Diversity:** It measures how dissimilar recommended items are for a user. This similarity is often determined using the item's content (e.g., movie genres) but can also be determined using how similar items are rated.

$$Diversity = \frac{2}{N(N-1)} \sum_{i_j \in L(u)} (1 - sim(i_j, i_k))$$

Here, $sim(i_j, i_k)$ denotes the similarity between item j and item k.

5. **Serendipity:** It is the measure of how surprising the successful or relevant recommendations are. The probability of a recommendation is simply a function of its overall rank over *n* items:

$$P_i = \frac{n - rank_i}{n - 1}$$

Here, P_i represents the probability of item i for recommendation and $rank_i$ shows the rank of item i over n items. The equation of findings unexpected recommendation is:

$$UNEXP = \frac{RS}{PM}$$

Here, PM denotes the set of recommendations generated by a primitive prediction model, and RS shows the generated recommendations. UNEXP consists that list which does not belong to RS. We define serendipity as follows:

$$Serendipity = \frac{\sum_{i=1}^{N} u(RS_i)}{N}$$

6. **Novelty:** It can be defined as:

$$Novelty = \sum_{i \in L} \frac{log_2 P_i}{n}$$

Higher novelty values represent that less popular items are being recommended, thus less well-known items are likely being surfaced for users.

The equations of computing precision, recall, F-measure, and accuracy are as follows:

1. **Precision:** It can be calculated by the fraction of the recommended items that are actually relevant to the target user.

$$Precision = \frac{t_p}{t_p + f_p}$$

2. **Recall:** It consists of the relevant items that are part of the set of recommended items. Hence, the equation of calculating recall becomes:

$$Recall = \frac{t_p}{t_p + f_n}$$

3. **F-Measure:** Precision and Recall values have been used **t**o compute the F-measure, and the equation is:

$$F\text{-}measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

4. **Accuracy:** It shows how close a predicted rating is to the actual rating. The equation of computing accuracy as follows:

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n}$$

Here, t_p, f_p, t_n, and f_n denote the true positive, false positive, true negative, and false negative respectively.

8.10 CHALLENGES IN COLLABORATIVE FILTERING (CF)

8.10.1 NEW USER PROBLEM

When users newly register to an RS, they do not have any ratings in their profile ratings denote the taste or preferences of the users. In the absence of a user, since CF is based on user preferences, it is unable to recommend many of the items (Lakshmi and Lakshmi, 2014). Even when users have scanty profiles with very few ratings, CF fails to render a reliable, personalized recommendation to these users. To overcome this problem, an RS used demographic features of the user from the user's profile for the recommendation. But it also has some issues that two users, having the same profile, may not have the same intent towards a particular item.

8.10.2 NEW ITEM PROBLEM

The new item is an additional issue in cold start problems which is based on the items, recurrently added to the list (Lakshmi and Lakshmi, 2014). Firstly, the items are rated then only they can be recommended to users.

8.10.3 SPARSITY PROBLEM

It takes place when the user has used some particular product but didn't bother to rate it, and another possibility can be that the user was completely unfamiliar with the product, so he didn't rate it (Lakshmi and Lakshmi, 2014). To run over this problem one approach of RS is a clustering method. Clustering method refines the data according to the preference of the user, and by doing so; it makes it easy for recommending items. But again, some issues have to be resolved in the case of multi-level clustering.

8.10.4 SCALABILITY PROBLEM

CF works on the database that contains user-item rating, and it has some scalability issue for the users and items set in large numbers. For large item set, the complexity of CF algorithms will be too large. High scalability of CF system is required as many of the systems need to respond immediately to fulfill online requirements which make recommendations for all users based on their purchase and rating history (Alloway, 2018; Poonam, Goudar, and Sunita, 2015).

8.10.5 SYNONYMY PROBLEM

Another problem with the CF approach is the synonymy problem (Xiaoyuan and Taghi, 2009). Most of the CF algorithms are unable to find similar items with various names (synonyms). Due to this, some association problem occurs, for e.g., "kids' movie" and "kids' film" is basically the same items to be searched, but according to memory-based CF, there is no match between above two terms to compute similarity. The next problem is abbreviations which are used a lot nowadays. Sometimes users are shown different results when they search for particular data, inserting abbreviations. Here the work should include these shortened words and categories them in the same list as per their full forms. Then there come the issues which are caused due to symbols or smileys. Some users prefer smiles to give a review of some products. For example, if a user wants to say that, he liked a product; he will simply give a smiley or a thumb up and used, or thumb down for dislike. So, such symbols should also be evaluated because some sites like Amazon do not hold any importance for smileys. Such sites rather ask to write a review in a minimum of 20 words (BBC, 2018). With this, there also comes the problem of reviewing the product in different regional languages. Different users want to give a review to the product in their respective languages for e.g., Hindi ("bahutachha"), Bengali ("khubbhalo"), English ("very nice"), etc. These different languages give out the same meaning that the product is good. But, if only one language will be considered then the reviews of other users, will lack its importance. For the betterment of the RS, it is also very important to take all these issues under examination.

8.10.6 LONG TAIL (LT) PROBLEM

In addition to the above-mentioned problems, one major problem also will arise, namely long tail (LT) problem. This section will discuss: What LT problem is and how to deal with it?

RSs basically use the past records of the user and then it anticipates the possible future likes and dislikes of the user and recommends accordingly. A better RS would propose fewer common options to draw the user's interest. It would not recommend similar kind of items repeatedly. Diversity is related to this aspect. This aspect implies the need for recommending diverse items to the user and how different the item is with respect to each other. But the RS lacks to co-operate with this aspect which leads to LT problem (Lei, 2013). The user will be deprived of many other necessary items just because he did not rate those items or because he did not have any access to those items. This generally leads to an LT problem. LT problem is when many items remain unrated or low rated. To deal with this problem, one idea is to rank the items in different ways. There is a need for segregating the ratings of the users and then rank it. Apart from the highest-rated items, there is also a need for recommending low rated items. The low rated items do not hold any importance. Researchers face the problem in the filtering of important low rated items. There is also a need to rank the items according to the purchase history and then recommend the lowest purchased item. However, it is not important that items that are more prestigious should necessarily be at the top of the list. We can see this aspect in the case of recommending books.

The LT problem can be reduced in an RS by considering (i) accuracy, (ii) similarity, (iii) diversity, and (iv) LT (Oscar, 2010; Daniel and Kartik, 2007; Yoon and Alexander, 2008; Hongzhi et al., 2012).

- **Accuracy:** A good RS should always check the accuracy level of the items recommended. To what extent the item is accurate will make the RS system run more smoothly.
- **Similarity:** This area emphasizes the fact, how much the product is similar to the users' past interest. There are various algorithms used in the RS system to find the similarity between users or items.
- **Diversity:** The same kind of products should never be recommended to the user on a regular basis, as this lesser down the interest of the user. A dynamic RS gives more diversity.

- **Long Tail (LT):** Some good items do not come into the top-n recommended list of items due to the smaller number of ratings. This problem provides the recommendations of more popular items only.

8.10.7 ATTACKS

The open nature of CF-based RS makes them prone to attacks known as Shilling attacks. Every RS identifies an item set favored by a certain user termed as the recommendation list for that user. Unscrupulous people use unethical ways to push their product into the Top-n recommendation list or pull down their competitors' product from that list. Hence, every attack is either a push or a nuke attack. To accomplish this, attackers inject fake profiles into the RS and give biased ratings to the items leading to erroneous recommendations. An attacker creates a fake profile in such a way as to remain effective and undetected at the same time. The rating is represented as the m-dimensional vector, where m represents a total number of items in the system. Every attack profile has four subparts (Mobasher et al., 2007):

- **Target Item (I_T):** A singleton item, which is to be pushed or nuked.
- **Selected Items (I_S):** A set of items whose rating is determined by a function based on the type of attack.
- **Filler Items (I_F):** A set of items chosen and rated randomly to copy the behavior of an authentic user.
- **Unrated Items (I_N):** A set of items not rated by the attacker.

The attackers closely follow certain attack models while designing an attack (Mobasher et al., 2005; Mobasher et al.,2015; Kaur and Goel, 2016). The target item is generally rated with the highest or the lowest rating. But the rating functions of the filler and selected items lead to different attack models. Let $\bar{r}$ denote the average rating of the RS over all items and users and let $\bar{r}_i$ denote the average rating of a certain item i. Similarly, let σ be the standard deviation of all ratings value over all users and items, and σ_i the standard deviation value of ratings of an item i. Let, $N(r, \sigma^2)$ denote the Gaussian distribution having mean r and variance σ^2 and $\rho(i)$ be the function based on which the filler items I_F are rated.

8.10.7.1 RANDOM ATTACK

Except for target item, attack profiles are generated based on randomly selected users' ratings from the database which contains information about the distribution of ratings. This attack was first mentioned by Lam and Riedl (2004). The set I_S contains no items, whereas I_F contains randomly picked up items whose ratings are given by the function $N(r, \sigma^2)$ centerd on the overall average rating in the database. In random attack:

$$I_S = \phi \text{ and } \rho(i) = N(\overline{r}, \sigma^2)$$

8.10.7.2 AVERAGE ATTACK

In average attack, target item's mean rating is used to generate an attack profile across all the users for filler items. Like the random attack, the I_S remains empty. The filler items are rated by the function $N(\overline{r}_i, \sigma_i^2)$ centered on the average rating σ_i of each item i in the database (Burke et al., 2005). Average attack proves to be more effective than a random attack. In average attack:

$$I_S = \phi \text{ and } \rho(i) = N(\overline{r}_i, \sigma_i^2)$$

8.10.7.3 BANDWAGON ATTACK

In bandwagon attack, high ratings are added to generate profiles for the selected item to increase the ratings of popular items. Here, $I_S = \{popular\ items\}$ *and* $\rho(i) = N(\overline{r}_i, \sigma_i^2)$. The items in the I_S are assigned high ratings. This attack needs an additional knowledge about the most popular items in an RS.

8.10.7.4 SEGMENT ATTACK

It is created to increase the recommendations of the set of target items for a certain set of users. Here, I_S is the items similar to target items and $\rho(i) = N(\overline{r}_i, \sigma_i^2)$. The items in I_S are termed as segment items that are well-liked by the target users. Like the target items, the segment items in I_S are given high ratings while the filler items are given low ratings.

8.11 A SURVEY ON COLLABORATIVE FILTERING (CF)

CF is at the heart of the RSs. It has been employed to develop recommendation techniques that suggest the best suitable items for customers. Yang et al. (2014) have presented a survey on the CF-based RSs categorizing them into social recommendation approaches using matrix factorization and neighborhood-based methods. They have also proposed the idea of utilizing the information from social networks as an additional input in CF for better quality recommendations. Elahi et al. (2016) have discussed the two most popular rating prediction algorithms used in a CF-neighborhood-based model and latent factor model while throwing some light on the cold start problem faced by CF techniques. Instead of using the entire data for CF and also introduces the use of active learning which involves obtaining high-quality data that can better represent a user's preferences, as a solution to the problem of a cold start. An excellent comparison of the CF algorithms found in the literature has been made by Cacheda et al. (2011) using several evaluation metrics, presenting the merits and demerits of every technique. To deal with sparse datasets, a new CF algorithm has been introduced in (Cacheda et al., 2011) that focus on the differences between the items or users rather than looking at their similarities. Shi et al. (2014) have presented a brief review of CF explaining the traditional memory-based and model-based CF approaches in detail. In addition, it also surveys some extended CF algorithms that make use of different information sources apart from the user-item matrix and presents the challenges faced by them. From the perspective of e-vendors in the domain of e-commerce, Karimova et al. (2007) have presented a literature review of the various recommendation techniques including the CF approaches. The analysis reveals the limitations of CF such as computational complexity, accuracy, and so on. An excellent survey of CF has been done by Su et al. (Su and Khoshgoftaar, 2009) where the three categories memory-based, model-based and hybrid CF algorithms have been studied in detail. The strengths and weaknesses of these algorithms, as well as their predictive performances, have been analyzed using several evaluation metrics. Finally, the various challenges faced by CF—scalability, sparsity, synonymy, grey sheep, shilling attacks, and privacy protection, etc. have been presented along with their possible solutions. Nagarnaik and Thomas (2015) have surveyed the various recommendation techniques including the CF algorithms explaining its various categories. A literature review has been done on the techniques that have been proposed to overcome the challenges faced by CF algorithms. Also, a hybrid CF technique has been proposed taking a combination of CF techniques and pattern finding algorithms for a better-quality web page recommendation. Yang et al. (2016) have shown the entire framework of a typical CF-based RS. A detailed survey of the working of CF algorithms-similarity metrics, prediction algorithms, and neighbor selection has been done, and case studies have presented to measure the accuracy of various CF algorithms using evaluation metrics.

8.12 CURRENT RESEARCH TRENDS IN COLLABORATIVE FILTERING (CF) (2011–2017)

Journal articles and conference proceedings were collected from four major electronic databases, i.e., ACM Digital Library, IEEE Explore, Springer, and Science Direct (Elsevier) to find the research trends in CF for the recommendation. The following queries were executed in Google Scholar:

a. (CF in RSs OR (issue OR issues OR challenge OR challenges OR problem OR problems)).
b. (Neighborhood CF in recommendation systems OR (issue OR issues OR challenge OR challenges OR problem OR problems)).
c. (CF OR neighborhood CF OR RSs)

Additional specifications about papers were also added in the advanced search of Google Scholar, to get more filtered papers out:

a. 2011–2017 is selected in the field of date section.
b. IEEE OR Elsevier OR ACM OR Springer has-been selected in the "published in" field.

The application of the above queries yielded around 500 research papers. Of the 500 research papers, we selected only 277 papers which are related to the area of computers and its allied fields. The keywords and abstract of each paper were used in the categorization of the collected papers, which led to the following results as shown in Figures 8.4–8.6. Figure 8.4 depicts the papers as well as their distribution among the top publications on selected 277 papers of CF. Out of the 277 papers, 77 are model-based, 162 are neighborhood-based, and the rest apply some other filtering technique in addition to CF as shown in Figure 8.5. And finally, Figure 8.6 states that the 162 papers containing neighborhood-based CF can be further segregated in a number of sub-domains based on the problems they try to rectify.

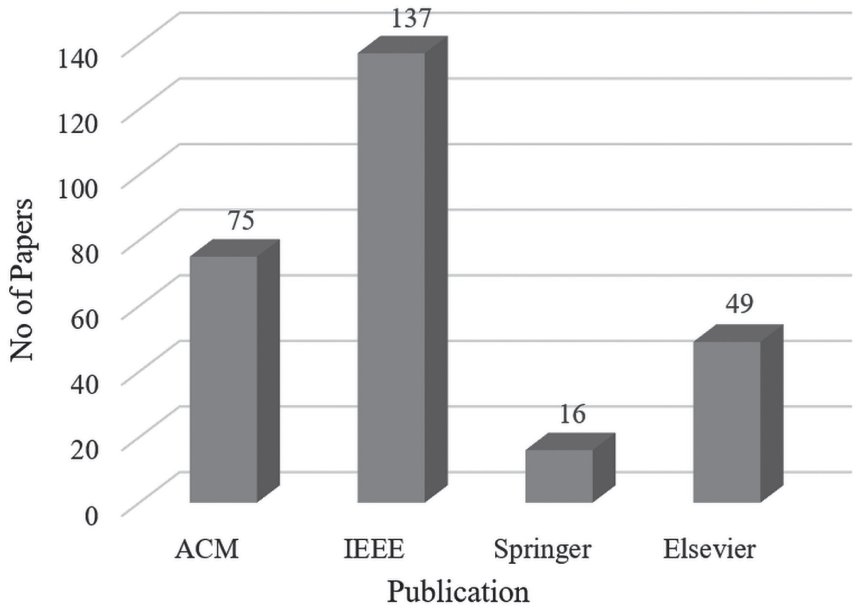

FIGURE 8.4 Number of paper distribution of CF, in each publication.

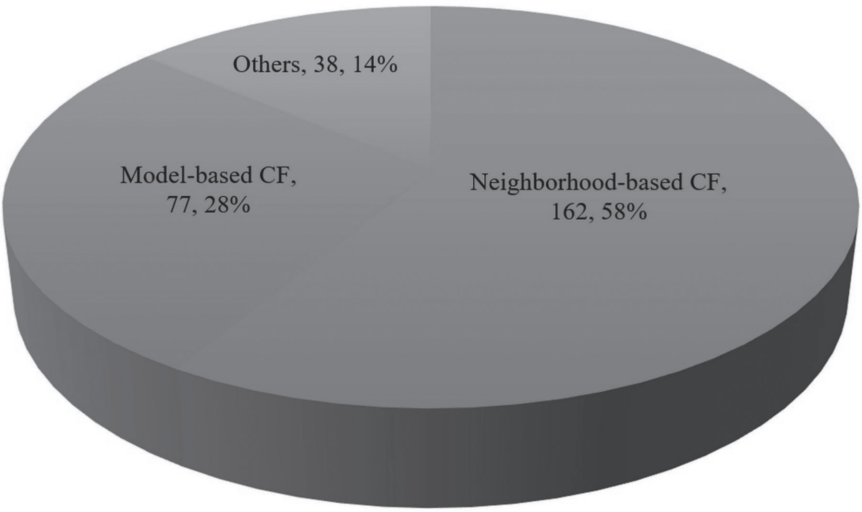

FIGURE 8.5 Number of papers in different categories of CF.

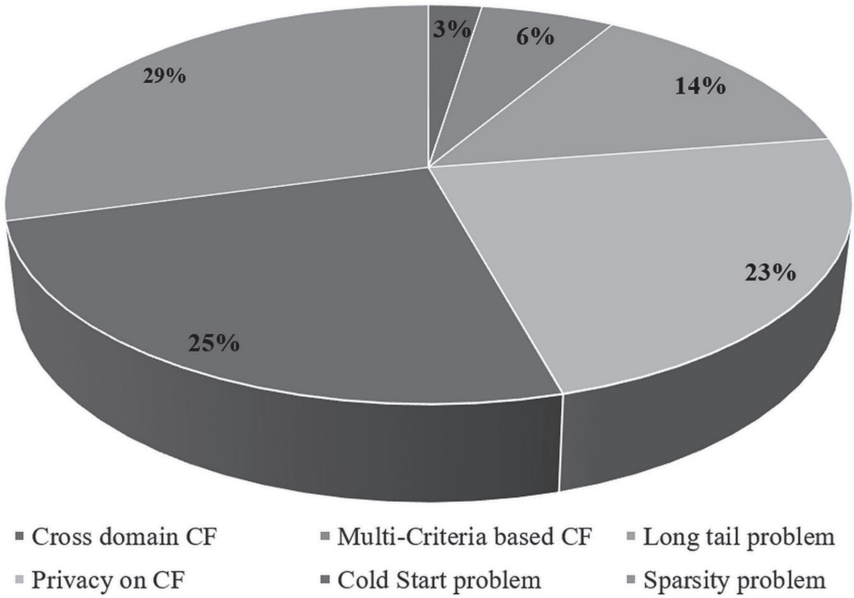

FIGURE 8.6 The percentage share of published papers in different categories of neighborhood-based CF.

8.13 THE FUTURE OF COLLABORATIVE FILTERING (CF) BASED RECOMMENDER SYSTEMS (RSS)

The accuracy of neighborhood-based CF mainly depends upon the top-n list of similar users or items (Yi et al., 2019; Soojung, 2019). Recommendations using these similar users/items tend towards more popular items. But an ideal RS has different properties such as more personalization, more diverse, more serendipity, and more novel. Recommendations using neighborhood-based CF can be improved if the researchers use information from social networks and contextual information of user or item with their rating information (Ambulgekar et al., 2019).

8.14 CONCLUSION

RS has found its usefulness in several fields of e-services. Among several filtering approaches used in RSs, CF is most common and popular. The mainly used perception of CF-based is that the rating of the items given by similar users will be close and similar items have similar rating patterns. On this basis, CF suggests the recommendable items to the users. CF extracts user ratings either implicitly or explicitly. To find the user-user and item-item similarity, several metrics are used to find the similarity of user-user and item-item. The similarity value has been used to predict the recommendable items. Several prediction approaches such as MC, WA, Z score (ZS), etc. are used. Different performance metrics such as MAE, RMSE, Coverage, etc. measure the correctness of recommendation. For effective implementation of CF-based RS, several challenges need to be addressed. New user and a new item, sparsity, scalability, synonymy, and LT problems are among them. Security and trust attacks on RS are also a major concern in people's acceptance of RSs. Due to the utility of CF in RS; it has attracted the attention of the academicians and researchers to make it more effective. The future of CF-based RS will be more personalized and diverse with more serendipity.

KEYWORDS

- **data extraction**
- **e-commerce**
- **future direction**
- **neighborhood-based collaborative filtering**
- **performance metrics**
- **recommendation system**
- **recommender system attacks**
- **research trends**
- **similarity metrics**

REFERENCES

ACRnews, (2017). [Online]. Available: http://www.acr-news.com/ (accessed on 16 February 2020).

Ali, K., & Van Stam, W., (2004). "TiVo: Making show recommendations using a distributed collaborative filtering architecture." In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Alloway, T., (2018). "*Amazon Urges California Referendum on Online Tax*." [Online]. Available: https://ftalphaville.ft.com/2011/07/12/619451/amazon-urges-california-referendum-on-online-tax/ (accessed on 16 February 2020).

Ambulgekar, H. P., Manjiri, K. P., & Kokare, M. B., (2019). "A survey on collaborative filtering: Tasks, approaches and applications." In: *Proceedings of International Ethical Hacking Conference* (pp. 289–300).

Amolochitis, E., Christou, I. T., & Tan, Z. H., (2014). "Implementing a commercial-strength parallel hybrid movie recommendation engine." *IEEE Intelligent Systems*, *29*, 92–96.

Asnicar, F., & Tasso, C., (1997). "ifWeb: A prototype of user model-based intelligent agent for document filtering and navigation in the world wide web." In: *Proceedings of 6th International Conference on User Modeling*.

Avesani, P., Massa, P., & Tiella, R., (2005). "Moleskiing.it: A Trust-aware Recommender System for Ski Mountaineering." In: *Proceedings of the ACM Symposium on Applied Computing*.

Balabanović, M., & Shoham, Y., (1997). "Fab: Content-based, collaborative recommendation." *Communications of the ACM*, *40*, pp. 66–72.

Baltrunas, L., Ludwig, B., Peer, S., & Ricci, F., (2012). "Context relevance assessment and exploitation in mobile recommender systems." *Personal and Ubiquitous Computing*, *16*(5), 507–526.

Baraglia, R., Frattari, C., Muntean, C. I., Nardini, F. M., & Silvestri, F., (2012). "RecTour: A recommender system for tourists." In: *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*.

BBC, (2018). "*Orlando Figes to Pay Fake Amazon Review Damages*." [Online]. Available: https://www.bbc.com/news/uk-10670407 (accessed on 16 February 2020).

Biletskiy, Y., Baghi, H., Keleberda, I., & Fleming, M., (2009). "An adjustable personalization of search and delivery of learning objects to learners." *Expert Systems with Applications, 36*(5), pp. 9113–9121.

Bilge, A., & Kaleli, C., (2014). "A multi-criteria item-based collaborative filtering framework." In: *11th International Joint Conference on Computer Science and Software Engineering.*

Billsus, D., & Pazzani, M. J., (2000). "User modeling for adaptive news access." *User Modeling and User-Adapted Interaction*, *10*, pp. 147–180.

Bjelica, M., (2010). "Towards TV recommender system: Experiments with user modeling." *IEEE Transactions on Consumer Electronics*, *56*(3), 1763–1769.

Blanco-Fernández, Y., Arias, J. J. P., Nores, M. L., Gil-Solla, A., & Cabrer, M. R., (2006). "AVATAR: An improved solution for personalized TV based on semantic inference." *IEEE Transactions on Consumer Electronics*, *52*(1), pp. 223–231.

Bobadilla, J., Hernando, A., Ortega, F., & Abraham, G., (2012). Collaborative filtering based on significances. *Information Sciences*, *185*(1), pp. 1–17.

Boutet, A., Frey, D., Guerraoui, R., Jegou, A., & Kermarrec, A. M., (2013). "WhatsUp: A decentralized instant news recommender." In: *IEEE 27th International Symposium on Parallel Distributed Processing*.

Braunhofer, M., Kaminskas, M., & Ricci, F., (2013). "Location-aware music recommendation." *International Journal of Multimedia Information Retrieval*, *2*(1), 31–44.

Burke, R. D., Hammond, K. J., & Young, B. C., (1996). "Knowledge-based navigation of complex information spaces." In: *Proceedings of the Thirteenth National Conference* *on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference (AAAI)*. Portland, Oregon.

Burke, R., (1999). "The wasabi personal shopper: A case-based recommender system." In: *Proceedings of the 11th National Conference on Innovative Applications of Artificial Intelligence.*

Burke, R., (2002). "Hybrid recommender systems: Survey and experiments." *User Modeling and User-Adapted Interaction, 12*(4), pp. 331–370.

Burke, R., Mobasher, B., Bhaumik, R., & Williams, C., (2005). "Segment-based injection attacks against collaborative filtering recommender systems." In: *Fifth IEEE International Conference on Data Mining (ICDM'05)*.

Cacheda, F., Carneiro, V., Fern'andez, D., & Formoso, V., (2011). "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems." In: *TWEB* (Vol. 5, No. 1/2, p. 33).

Cao, Y., & Li, Y., (2007). "An intelligent fuzzy-based recommendation system for consumer electronic products." *Expert Systems with Applications*, *33*(1), pp. 230–240.

Capuano, N., Gaeta, M., Ritrovato, P., & Salerno, S., (2014). "Elicitation of latent learning needs through learning goals recommendation." *Computers in Human Behavior*, *30*, pp. 663–673.

Celma, Ò., & Serra, X., (2008). "FOAFing the music: Bridging the semantic gap in music recommendation." *Web Semantics: Science, Services and Agents on the World Wide Web*, *6*(4).

Chen, C. M., & Duh, L. J., (2008). "Personalized web-based tutoring system based on fuzzy item response theory." *Expert Systems with Applications*, *34*, pp. 2298–2315.

Chen, C. M., Duh, L. J., & Liu, C. Y., (2004). "A personalized courseware recommendation system based on fuzzy item response theory." In: *IEEE International Conference on e-Technology, e-Commerce and e-Service*.

Chen, L., & Pu, P., (2012). "Critiquing-based recommenders: Survey and emerging trends." *User Modeling and User-Adapted Interaction*, *22*(1), pp. 125–150.

Chesnevar, C. I., & Maguitman, A. G., (2004). "ArgueNet: An argument-based recommender system for solving Web search queries." In: *2nd International IEEE Conference on Intelligent Systems*.

Christou, I. T., Amolochitis, E., & Tan, Z. H., (2016). "AMORE: design and implementation of a commercial-strength parallel hybrid movie recommendation engine." *Knowledge and Information Systems*, *47*(3), 671–696.

Cobos, C., Rodriguez, O., Rivera, J., Betancourt, J., Mendoza, M., Leó, N. E., & Herrera-Viedma, E., (2013). "A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes." *Information Processing and Management: An International Journal*, *49*, 607–625.

Console, L., Torre, I., Lombardi, I., Gioria, S., & Surano, V., (2003). "Personalized and adaptive services on board a car: An application for tourist information." *Journal of Intelligent Information Systems*, *21*(3), pp. 249–284.

Cornelis, C., Lu, J., Guo, X., & Zhang, G., (2007). "One-and-only item recommendation with fuzzy logic techniques." *Information Sciences, 177*, pp. 4906–4921.

Daniel, M. F., & Kartik, H., (2007). "Recommender systems and their impact on sales diversity." In: *Proceedings of the 8th ACM Conference on Electronic Commerce (EC '07)* (pp. 192–199). ACM, New York, NY, USA.

Deshpande, M., & Karypis, G., (2004). "Item-based top-n recommendation Algorithms." *ACM Transactions on Information Systems (TOIS), 22*, pp. 143–177.

Domingues, M., Gouyon, F., Jorge, A., Leal, J., Vinagre, J., Lemos, L., & Sordo, M., (2013). "Combining usage and content in an online recommendation system for music in the long tail." *International Journal of Multimedia Information Retrieval*, *2*(1), 3–13.

Elahi, M., Ricci, F., & Rubens, N., (2016). "A survey of active learning in collaborative filtering recommender systems." In: *Computer Science Review* (Vol. 20, pp. 29–50).

Esteban, B., Tejeda-Lorente, Á., Porcel, C., Arroyo, M., & Herrera-Viedma, E., (2014). "TPLUFIB-WEB: A fuzzy linguistic Web system to help in the treatment of low back pain problems." *Knowledge-Based Systems*, *67*, pp. 429–438.

Farzan, R., & Brusilovsky, P., (2006). "Social navigation support in a course recommendation system." In: *Adaptive Hypermedia and Adaptive Web-Based Systems: 4th International Conference (AH 2006)*. Dublin, Ireland.

Fesenmaier, D. R., Ricci, F., Schaumlechner, E., Wöber, K., & Zanella, C., (2003). "DIETORECS: Travel advisory for multiple decision styles." In: *Proceedings of the International Conference on Information and Communication Technologies in Tourism*. Wien, Austria.

FoxTrit, (2017). [Online]. Available: http://www.foxtrot.com/wp-content/endurance-page-cache/_index.html (accessed on 16 February 2020).

García-Crespo, A., Chamizo, J., Rivera, I., Mencke, M., Colomo-Palacios, R., & Gómez-Berbís, J. M., (2009). "SPETA: Social pervasive e-tourism advisor." *Telematics and Informatics*, *26*, pp. 306–315.

García-Crespo, Á., López-Cuadrado, J. L., Colomo-Palacios, R., González-Carrasco, I., & Ruiz-Mezcua, B., (2011). "Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain." *Expert Systems with Applications*, *38*, pp. 13310–13319.

Garfinkel, R., Gopal, R., Tripathi, A., & Yin, F., (2006). "Design of a shopbot and recommender system for bundle purchases." *Decision Support Systems, 42*(3), pp. 1974–1986.

Guo, X., and Lu, J., (2007). "Intelligent e-government services with personalized recommendation techniques." *International Journal of Intelligent Systems*, *22*, pp. 401–417.

Hauver, D., & French, J., (2001). "Flycasting: using collaborative filtering to generate a playlist for online radio." In: *First International Conference on Web Delivering of Music*.

Hayes, C., & Cunningham, P., (2001). "Smart radio-community based music radio." *Knowledge Based Systems*, *14*, pp. 197–201.

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J., (1999). "An algorithmic framework for performing collaborative filtering." In: *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Hongzhi, Y., Bin, C., Jing, L., Junjie, Y., & Chen, C., (2012). "Challenging the long tail recommendation." In: *Proc. VLDB Endow., 5*(9), 896–907.

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G., (2006). "Information retrieval in folksonomies: Search and ranking." In: *The Semantic Web: Research and Applications: 3rd European Semantic Web Conference*. Budva, Montenegro.

Hu, J., Wang, B., Liu, Y., & Li, D. Y., (2012). "Personalized tag recommendation using social influence." *Journal of Computer Science and Technology*, *27*(3), pp. 527–540.

IMDb, (2017). [Online]. Available: http://www.imdb.com/ (accessed on 16 February 2020).

Jäschke, R., Marinho, L. B., Hotho, A., Schmidt-Thieme, L., & Stumme, G., (2007). "Tag Recommendations in Folksonomies." In: *11th European Conference on Principles and Practice of Knowledge Discovery in Databases*. Warsaw, Poland.

Jinni, (2017). [Online]. Available: http://www.jinni.com/ (accessed on 16 February 2020).

Karimova, F. (2016), "A Survey of e-Commerce Recommender Systems," *European Scientific Journal*, *12*(34), 75–89.

Kaur, P., & Goel, S., (2016). "Shilling attack models in recommender system." In: *International Conference on Inventive Computation Technologies (ICICT)* (Vol. 2, pp. 1–5).

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J., (1997). "GroupLens: Applying collaborative filtering to Usenet news." *Communications of the ACM*, *40*, pp. 77–87.

Lakshmi, S. S., & Lakshmi, T. A., (2014). "Recommendation systems: Issues and challenges." In: *International Journal of Computer Science and Information Technologies*, *5*.

Lam, S. K., & Riedl, J., (2004). "Shilling recommender systems for fun and profit." In: *Proceedings of the 13th International Conference on World Wide Web* (pp. 393–402).

Lawrence, R., Almasi, G., Kotlyar, V., Viveros, M., and Duri, S., (2001). "Personalization of supermarket product recommendations." *Data Mining and Knowledge Discovery*, *5*(1), pp. 11–32.

Lee, S. K., Cho, Y. H., & Kim, S. H., (2010). "Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations." *Information Sciences, 180*(11), pp. 2142–2155.

Lei, S., (2013). "Trading-off among accuracy, similarity, diversity, and long-tail: A graph-based recommendation approach." In: *Proceedings of the 7th ACM Conference on Recommender Systems*.

Levandoski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F., (2012). "LARS: A location-aware recommender system." In: *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*.

Li, D., Miao, C., Chu, S., Mallen, J., Yoshioka, T., & Srivastava, P., (2018). "*Stable Matrix Approximation for Top-n Recommendation on Implicit Feedback Data*." In Hawaii International Conference on System Sciences 2018 (HICSS-51).

Liu, Q., & Karger, D. R., (2015). "Kibitz: End-to-end recommendation system builder." In: *RecSys*.

Lu, J., (2004). "Personalized e-learning material recommender system." In: *Proceedings of International Conference on Information Technology for Application*.

Lu, J., Shambour, Q., Xu, Y., Lin, Q., & Zhang, G., (2010). "BizSeeker: A hybrid semantic recommendation system for personalized government-to-business e-services." *Internet Research*, *20*, pp. 342–365.

Luz, N., Moreno, M., Anacleto, R., Almeida, A., & Martins, C., (2013). "A hybrid recommendation approach for a tourism system." *Expert Systems with Applications*, *9*(40), 3532–3550.

Marcel, M. A., Ball, M., Boley, H., Greene, S., Howse, N., Lemire, D., & Mcgrath, S., (2003). "RACOFI: A rule-applying collaborative filtering system." In: *Proc. IEEE/WIC COLA'03*. Halifax, Canada.

Martín-Vicente, M. I., Gil-Solla, A., Ramos-Cabrer, M., Blanco-Fernández, Y., & Servia-Rodríguez, S., (2012). "Semantics-driven recommendation of coupons through digital TV: Exploiting synergies with social networks." In: *IEEE International Conference on Consumer Electronics*.

Martínez, L., Rodríguez, R. M., & Espinilla, M., (2009). "REJA: A georeferenced hybrid recommender system for restaurants." In: *Proceedings of the 2009 IEEE/WIC/ACM*

International Joint Conference on Web Intelligence and Intelligent Agent Technology (pp. 187–190).

Mccarthy, K., Reilly, J., Mcginty, L., & Smyth, B., (2004). "Thinking positively-explanatory feedback for conversational recommender systems." In: *Proceedings of the ECCBR 2004 Workshops*.

Melville, P., Mooney, R. J., & Nagarajan, R., (2002). "Content-boosted collaborative filtering for improved recommendations." In: *Eighteenth National Conference on Artificial Intelligence*. Edmonton, Alberta, Canada.

Meo, P. D., Quattrone, G., & Ursino, D., (2008). "A decision support system for designing new services tailored to citizen profiles in a complex and distributed e-government scenario." *Data and Knowledge Engineering, 67*, pp. 161–184.

Miller, B. N., Konstan, J. A., & Riedl, J., (2004). "PocketLens: Toward a personal recommender system." *ACM Transactions on Information Systems*, *22*(3), pp. 437–476.

Mobasher, B., Burke, R., Bhaumik, R., & Sandvig, J. J., (2007). "Attacks and remedies in collaborative recommendation." *IEEE Intelligent Systems, 22*(3), 56–63.

Mobasher, B., Burke, R., Bhaumik, R., & Williams, C. (2007). "Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness." *ACM Trans. Internet Technol., 7*(4).

Mobasher, B., Burke, R., Bhaumik, R., & Williams, C., (2005). "Effective attack models for shilling item-based collaborative filtering systems." In: *Proceedings of the WebKDD Workshop, Held in Conjunction with ACM SIGKDD2005*.

Moreno, A., Valls, A., Isern, D., Marin, L., & Borràs, J., (2013). "SigTur/E-destination: Ontology-based personalized recommendation of tourism and leisure activities." *Engineering Applications of Artificial Intelligence*, *26*(1), pp. 633–651.

Moukas, A., & Maes, P., (1998). "Amalthaea: An evolving multi-agent information filtering and discovery system for the WWW." *Autonomous Agents and Multi-Agent Systems*, *1*(1), 59–88.

Movielens, (2017). [Online]. Available: https://movielens.org/ (accessed on 16 February 2020).

Nagarnaik, P., & Thomas, A., (2015). "Survey on recommendation system methods." In: *2nd International Conference on Electronics and Communication Systems (ICECS)* (pp. 1603–1608).

nanoCROWD, (2017). [Online]. Available: http://nanocrowd.com/ (accessed on 16 February 2020).

Nanopoulos, A., Rafailidis, D., Symeonidis, P., & Manolopoulos, Y., (2010). "Music box: Personalized music recommendation based on cubic analysis of social tags." *IEEE Transactions on Audio, Speech and Language Processing*, *18*(2), pp. 407–412.

Natarajan, N., Shin, D., & Dhillon, I. S., (2013). "Which app will you use next?: Collaborative filtering with interactional context." In: *Proceedings of the 7th ACM Conference on Recommender Systems*.

Nguyen, T. T. S., Lu, H. Y., & Lu, J., (2014). "Web-page recommendation based on web usage and domain knowledge." *IEEE Transactions on Knowledge and Data Engineering*, *26*(10), pp. 2574–2587.

Oh, J., Kim, S., Kim, J., & Yu, H., (2014). "When to recommend: A new issue on TV show recommendation." *Information Sciences, 280*, pp. 261–274.

Oscar, C., (2010). "*Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*." Springer Publishing Company, Incorporated.

Pampın, H. J. C., Jerbi, H., & O'Mahony, M. P. (2015), "Evaluating the relative performance of collaborative filtering recommender systems, *Journal of Universal Computer Science*, *21*(13), 1849–1868.

Park, Y. J., (2013). "An adaptive match-making system reflecting the explicit and implicit preferences of users." *Expert Systems with Applications: An International Journal*, *40*, 1196–1204.

Park, Y. J., & Tuzhilin, A. (2008, October), "The long tail of recommender systems and how to leverage it," In *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 11–18).

Parra, D., Brusilovsky, P., & Trattner, C., (2014). "See what you want to see: Visual user-driven approach for hybrid recommendation." In: *Proceedings of the 19th International Conference on Intelligent User Interfaces*.

Pashtan, A., Blattler, R., Andi, A. H., & Scheuermann, P., (2003). "CATIS: A context-aware tourist information system." In: *Proceedings of the 4th International Workshop of Mobile Computing*.

Poonam, T. B., Goudar, R. M., & Sunita, B., (2015). "Article: Survey on collaborative filtering, content-based filtering and hybrid recommendation system." *International Journal of Computer Applications*, 31–36.

Porcel, C., & Herrera-Viedma, E., (2010). "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries." *Knowledge-Based Systems, 23*, pp. 32–39.

Porcel, C., Herrera-Viedma, E., & Moreno, J. M., (2009). "A multi-discipliner recommender system to advice research resources in university digital libraries." *Expert Systems with Applications*, *36*, pp. 12520–12528.

Porcel, C., López-Herrera, A. G., & Herrera-Viedma, E., (2009). "A recommender system for research resources based on fuzzy linguistic modeling." *Expert Systems with Applications: An International Journal*, *36*, pp. 5173–5183.

Pratikshashiv, (2015). "*Flipkart Uses Collaborative Based Filtering*." [Online]. Available: https://pratikshashiv.wordpress.com/ (accessed on 16 February 2020).

Renda, M. E., & Straccia, U., (2005). "A personalized collaborative digital library environment: A model and an application." *Information Processing and Management: An International Journal*, *41*, 5–21.

Rikitianskii, A., Harvey, M., & Crestani, F., (2014). "A personalized recommendation system for context-aware suggestions." In: *Advances in Information Retrieval: 36th European Conference on IR Research*. ECIR.

Rotten Tomatoes, (2017). [Online]. Available: https://www.rottentomatoes.com/ (accessed on 16 February 2020).

Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T., & Hyvönen, E., (2013). "Smart museum: A mobile recommender system for the web of data." *Web Semantics: Science, Services and Agents on the World Wide Web*, *20*.

Samundeeswary, K., & Krishnamurthy, V. (2017, June), "Comparative study of recommender systems built using various methods of collaborative filtering algorithm." In *2017 International Conference on Computational Intelligence in Data Science (ICCIDS)* (pp. 1–6). IEEE.

Salter, J., & Antonopoulos, N., (2006). "Cinema screen recommender agent: Combining collaborative and content-based filtering." *IEEE Intelligent Systems*, *21*(1), pp. 35–41.

Sánchez, L. Q., Recio-García, J. A., & Díaz-Agudo, B., (2011). "Happy movie: A Facebook application for recommending movies to groups." In: *23rd International Conference on Tools with Artificial Intelligence (ICTAI)*.

Santos, O. C., Boticario, J. G., D. Pérez-Marín, Santos, O., Boticario, J., & Perez-Marin, D., (2014). "Extending web-based educational systems with personalized support through user centered designed recommendations along the e-learning life cycle." *Science of Computer Programming*, *88*, pp. 92–109.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J., (2001). "Item-based collaborative filtering recommendation algorithms." In*: 10th International Conference on World Wide Web*.

Schiaffino, S., & Amandi, A., (2009). "Building an expert travel agent as a software agent." *Expert Systems with Applications: An International Journal*, *36*(2), 1291–1299.

Serrano-Guerrero, J., Herrera-Viedma, E., Olivas, J. A., Cerezo, A., & Romero, F. P., (2011). "A google wave-based fuzzy recommender system to disseminate information in university digital libraries 2.0." *Information Sciences: An International Journal, 181*, 1503–1516.

Shi, Y., Larson, M., & Alan, H., (2014). "Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges." In: *ACM Comput. Surv.,* (Vol. 47, No. 1–3, p. 45).

Singh, P. K., Pramanik, P. K. D., & Choudhury, P., (2018). "A comparative study of different similarity metrics in highly sparse rating dataset." In: *Data Management, Analytics and Innovation, Proceedings of ICDMAI* (Vol. 2, pp. 45–60). Springer.

Smyth, B., & Cotter, P., (2000). "A personalized television listings service." *Communications of the ACM*, *4*(8), pp. 107–111.

Soojung, L., (2019). "Using entropy for similarity measures in collaborative filtering." In: *Journal of Ambient Intelligence and Humanized Computing*.

Su, X., & Khoshgoftaar, T. M., (2009). "A survey of collaborative filtering techniques." *Advances in Artificial Intelligence.* Article ID 421425.

TASTEKiD, (2017). [Online]. Available: https://www.tastekid.com/ (accessed on 16 February 2020).

Teran, L., & Meier, A., (2010). "A fuzzy recommender system for elections." In: Electronic *Government and the Information Systems Perspective, First International Conference (EGOVIS)*. Bilbao, Spain.

Tung, H. W., & Soo, V. W., (2004). "A personalized restaurant recommender agent for mobile e-service." In: *Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04)*. Washington, DC, USA.

Walter, F. E., Battiston, S., Yildirim, M., & Schweitzer, F., (2012). "Moving recommender systems from on-line commerce to retail stores." *Information Systems and e-Business Management*, *10*(3), pp. 367–393.

Wei, K., Huang, J., & Fu, S., (2007). "A survey of e-commerce recommender systems." In: *International Conference on Service Systems and Service Management* (pp. 1–5).

Wu, D., Zhang, G., & Lu, J., (2015). "A fuzzy preference tree-based recommender system for personalized business-to-business e-services." *IEEE Transactions on Fuzzy Systems*, *23*, pp. 29–43.

Wu, J., Chen, L., Feng, Z., Zhou, M., & Wu, Z., (2013). "Predicting quality of service for selection by neighborhood-based collaborative filtering." *IEEE Transactions Systems, Man, and Cybernetics: Systems, 43*(2), pp. 428–439.

Xiaoyuan, S., & Taghi, K. M., (2009). "*A Survey of Collaborative Filtering Techniques*." Adv. in Artif. Intell.

Xing, X., Yu, Z., Liuhang, Z., & Yuan, N. J., (2013). "T-finder: A recommender system for finding passengers and vacant taxis." *IEEE Transactions on Knowledge and Data Engineering*, *25*, pp. 2390–2403.

Yang, X., Guo, Y., Liu, Y., & Steck, H., (2014). "A survey of collaborative filtering based social recommender systems." In: *Computer Communications, 41*, 1–10.

Yang, Z., Wu, B., Zheng, K., Wang, X., & Lei, L., (2016). "*A Survey of Collaborative Filtering-Based Recommender Systems for Mobile Internet Applications* (pp. 3273–3287)." IEEE Access 4.

Yanga, W. S., & Hwang, S. Y., (2013). "iTravel: A recommender system in mobile peer-to-peer environment." *Journal of Systems and Software*, *86*(1), 12–20.

Yi, M., Nianhao, X., Ruichun, T., Liang, L., & Xiaohan, Y., (2019). "An efficient similarity measure for collaborative filtering." In: *Procedia Computer Science* (pp. 147, 416–421).

Yin, H., Cui, B., Sun, Y., Hu, Z., & Chen, L., (2014). "LCARS: A spatial item recommender system." *ACM Transactions on Information Systems*, *32*(3), pp. *11*, 1–11, 37.

Zaíane, O. R., (2002). "Building a recommender agent for e-learning systems." In: *Proceedings of the International Conference on Computers in Education*. Washington, DC, USA.

Zhang, H., Gao, Y., Chen, H., & Li, Y., (2012). "TravelHub: A semantics-based mobile recommender for composite services." In: *16th International Conference on Computer Supported Cooperative (CSCWD)*.

Zhang, Z. K., Zhou, T., & Zhang, Y. C., (2011). "Tag-aware recommender systems: A state-of-the-art survey." *Journal of Computer Science and Technology*, *26*.

Zhang, Z., Lin, H., Liu, K., Wu, D., Zhang, G., & Lu, J., (2013). "A hybrid fuzzy-based personalized recommender system for telecom products/services." *Information Sciences*, *235*, pp. 117–129.

Zhao, W. X., Li, S., He, Y., Wang, L., Wen, J. R., & Li, X., (2016). "Exploring demographic information in social media for product recommendation." *Knowledge of Information System*, *49*(1), pp. 61–89.

Zhao, X. W., Guo, Y., He, Y., Jiang, H., Wu, Y., & Li, X., (2014). "We know what you want to buy: A demographic-based system for product recommendation on microblogs." In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Zuva, K., & Zuva, T., (2017). "Diversity and serendipity in recommender systems." In: *Proceedings of the International Conference on Big Data and Internet of Thing*.