
Recommendation System Issues, Approaches and Challenges Based on User Reviews

Khalid Benabbes^{1,*}, Khalid Housni¹, Ali El Mezouary²
and Ahmed Zellou³

¹*MISC Laboratory, Faculty of Sciences, Ibn Tofail University, Kénitra, Morocco*

²*IRF-SIC Laboratory, EST, Ibn Zohr University, Agadir, Morocco*

³*SPM Research Team, ENSIAS, Mohammed V University, Rabat, Morocco*

E-mail: khalid.benabbes@uit.ac.ma; housni.khalid@uit.ac.ma;

a.elmezouary@uiz.ac.ma; ahmed.zellou@um5.ac.ma

**Corresponding Author*

Received 29 August 2021; Accepted 19 January 2022;
Publication 12 April 2022

Abstract

With the ever-increasing volume of online information, recommender systems have been effective as a strategy to overcome information overload. They have a wide range of applications in many fields, including e-learning, e-commerce, e-government and scientific research. Recommender systems are search engines that are based on the user's browsing history to suggest a product that expresses their interests. Being usually in the form of textual comments and ratings, such reviews are a valuable source of information about users' perceptions. Recommender systems (RSs) apply various approaches to predict users' interest on information, products and services among a huge amount of available items. In this paper, we will describe the recommender system, discuss ongoing research in this field, and address the challenges, limitations and the techniques adopted. This paper also discusses

Journal of Web Engineering, Vol. 21.4, 1017–1054.

doi: 10.13052/jwe1540-9589.2143

© 2022 River Publishers

how review texts are interpreted to solve some of the major problems with traditional recommendation techniques. To assess the value of a recommender system, qualitative evaluation measures are discussed as well in this research. Based on a series of selected articles published between 2008 and 2020, the study allowed us to conclude that the efficiency of RSs is strongly centered on the control of information context, the operated exploration algorithm, the method, and the type of processed data in addition to the information on users' trust.

Keywords: Recommender system, collaborative filtering, content filtering, user reviews.

1 Introduction

Currently, the World Wide Web has become a powerful platform for storing and retrieving information as well as for extracting useful knowledge to predict people's interests. Web mining for instance, plays a crucial role in finding data used frequently on the web [1]. It is certainly challenging for users to screen all this information and retain the essential features. Thus, many online e-commerce websites recommend products to users, and provide millions of items in a single platform. Due to the titanic amount of items and their dynamic nature, the use of an additional tool, called a recommendation system (RS) is necessary [2–4]. RSs solve the problem of information overload, while recommending specific items tailored to the preferences of users [4]. Recommender systems predict user interest by analyzing search history, the behavior of similar users and the various data processing techniques from eBay, Yelp, Netflix and Amazon [1]. The recommender system is considered as an independent search space where suggestions for a target user are based on several criteria including ratings, reviews, location, time, etc. [5]. Additionally, the quality of a RS has a significant impact on the decisions, the organizations' profits, and the user satisfaction. For instance, in the e-learning field and as a consequence of a reliable recommendation system, learners can benefit from interesting courses presented to them [6].

Currently, RSs are widely used in e-commerce and unit-based product sales. They are also used in other domains such as education, health, translation, advertising and agriculture [7, 8]. In recent years, RSs have attracted many researchers, thus several studies have been carried out to highlight their characteristics, techniques and problems [1–19]. Nevertheless, none of these

studies has covered all the technical aspects of a RS in a comprehensive manner.

RS models are often categorized into three broad categories: (1) classical class [8], (2) class of Su et al. [9] and (3) class of Nageswara Rao and Talwar [10]. These classes use one or more algorithms under these main approaches: (1) content-based (CB) [20], (2) collaborative-filtering (CF) [16] and (3) hybrid approach [11].

In this paper, the following objectives are sought: (1) develop a comprehensive catalog of RSs, (2) Discuss the different approaches of RSs and issues associated with current recommender systems, (3) evaluate the performance of a RS, (4) discuss the application of RSs in different domains and (5) discuss the handling of textual, contextual, emotional and semantic data.

2 Recommender System

The objective of recommender systems is to filter and adapt the information to users, in order to maximize individual users' satisfaction. For this purpose, several recommendation methods have been proposed and classified into several categories: content-based [3], collaborative-based [24], semantic-based [42], knowledge-based [21], demographic-based [12], utility-based [22], context-based [23] and hybrid-based [11]. Since the birth of RSs in the mid-1990s, information overload are addressed to provide users with specific recommendations that respond exactly to their interests [19]. The most frequently used classifications are based on content-based and collaborative filtering [13]. Recommendation systems help users to make their best decisions through a number of steps as shown in Figure 1. It starts with collecting data then processing and filtering it using one of the three techniques, namely, content filtering, collaborative filtering and hybrid filtering.

In addition, information contained in reviews such as historical data, demographics, requirements individual and emotion data, as shown in Table 1, are collected implicitly or explicitly and taken into account to minimize the problems of traditional evaluation systems, such as the prediction, the quality and the sparsity problems.

This study aims to evaluate the state of the art text-based reviews recommendation systems and published works to provide analysts and researchers with some perspectives and insights on evaluation-based approaches. To refine our approaches, methodologies, and our research roadmap, a number of Research Questions (RQ) are being raised and summarized as follows:

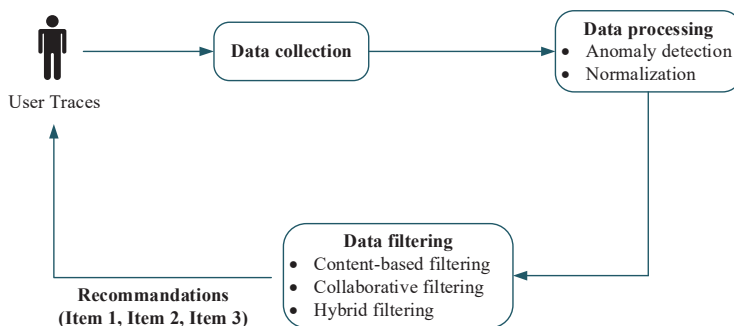


Figure 1 Recommender system process.

Table 1 Taxonomy of knowledge source

Historical Data	Behavior data [25]	Duration of surfing, click on times, links of webs; shop, print, scroll, delete, open, close, refresh of URLs, choice, edition, search, reproduction, paste, bookmark and even download of internet content material.
	Trade data [42]	Purchase amount and date, rates, discounting amount.
	Opinion data [16]	Rating, tags, reviews, comment, a latent comment such as bad, worse, best, good.
Demographic data [12]		Gender, age, name, surname, telephone, address, hobbies, salary, information experience, birth date, profession, income.
Requirements individual data [26]		Query, constraints, references, and context.
Context meta data [29]		A device, alone or in a group; moving or stationary, time of day/week /month/year.
Emotion data [25]		Agreement or disagreement, positive or negative, good or fine, busy or tired, angry, uncertainty, anxiety, discomfort or displeasure.

In RQ1, we will present the different categories of recommender systems using traditional approaches to overcome the main related issues, as well as their context of application and their advantages and disadvantages.

In RQ2, we will review the standard techniques and algorithms of recommendation systems based on text-reviews and their main issues as reported in the literature.

In RQ3, we will cover potential solutions outlined by recent studies based on text-reviews, which can be seen as reliable for solving the main challenges raised in most application domains.

In the following section, each category and the subgroups with the traditional techniques used to carry out each approach studied as a summary of RQ1 will be briefly explained.

RQ1 summary

2.1 Content-based Filtering (CB)

Content-inspired techniques leverage the domain of search and natural language. They associate items' characteristics with users' profiles. To recommend items, the RS must match the attributes of items with the interests and preferences of users. When an item is newly created, the recommendation model must assign a degree of preference to the user for the given item [20]. Content-based techniques are limited to features given by the item's content, a source of the over-specialization problem (only items similar to consumed items are predicted) [3]. Also, content-based techniques suffer from a cold-start problem (new-item, new-user) [2] and a diversity problem (user can see new items) [4].

Two well-known types of recommendations use content-based approach and these are:

Semantic-based approach [42] in which a semantic module uses the domain of ontology to interpret items. This is characterized by:

- Type of knowledge used, such as lexicon, ontology, etc.
- Techniques used for the annotation of items.
- Types of characteristics describing the user's profile.
- Way of associating the given items with the user's profile.

Keyword-based approach [30], when a user wants to access particular resources, RS recommends items which are related, as these items have shared keywords.

The techniques used to carry out content-based filtering are summarized in Table 2, along with their advantages and drawbacks.

2.2 Collaborative Filtering (CF)

Unlike content-based approaches, which rely on the content of items previously evaluated by a user, collaborative approaches recommend resources that have been highly scored by users with the same preferences as the current user. This approach, which is based on a rating structure, is usually represented by a rating matrix [31]. Breese et al. [24] classified collaborative approaches into two categories as shown in Figure 2.

Table 2 Comparison between recommender systems approaches

Class/Approach	Techniques and Algorithms Used	Advantages	Drawbacks
Content-based [3]	<ul style="list-style-type: none"> - Content similarity analysis (TF/IDF) [18] - Clustering [3] - Decision tree 	<ul style="list-style-type: none"> - Improvement of the quality of recommendation [42] - Reduction of data sparsity [42] 	<ul style="list-style-type: none"> - Lack of recommendation diversity - Content indexing (extraction of representative attributes) - Problems of indexing multimedia documents
Collaborative-based [24]	<ul style="list-style-type: none"> - Clustering [3] - Dimensionality reduction (SVD, PCA) [42] - Association rule learning, sequential pattern, Markov models: Web Usage Mining (WUM) [18] 	<ul style="list-style-type: none"> - Improvement of the quality of recommendation - Prediction of future behavior 	<ul style="list-style-type: none"> - Costly model construction - Loss risk of pertinent information due to dimensionality reduction - Problem of calculating pattern rules when the system lacks sufficient data/relevance due to dimensionality reduction - Does not take into account the user profile for WUM models
CF Memory-based [36]	<ul style="list-style-type: none"> - CF using KNN¹ (user-based, item-based) [43] 	<ul style="list-style-type: none"> - Simple implementation - Easy integration of new data - Great accuracy of recommendation 	<ul style="list-style-type: none"> - Dependency on grade data - Deterioration of recommendation quality due to sparsity - Scalability problem

Demographic-based [12]	<ul style="list-style-type: none"> - Collaborative recommendation techniques [12] - Considering three main components: - K-means² clustering [43] - Matrix-factorization [42] - Singular value decomposition (SVD) [29] 	<ul style="list-style-type: none"> - The demographic data used - Classify users 	<ul style="list-style-type: none"> - Requires a database of homogeneous users to classify users - Low performance
Knowledge-based [21]	<ul style="list-style-type: none"> - Always inspired by the explicit knowledge of users [36] - Rule-based systems make it easier to represent knowledge [72] 	<ul style="list-style-type: none"> - Production of new preference results - Relationships between the selected preferences - No cold start problem 	<ul style="list-style-type: none"> - Knowledge depiction of users is considered - Problem with the accuracy of predictions - Users with unrestricted reasoning
Utility-based [22]	<ul style="list-style-type: none"> - Classification of item data based on a core attribute of the items - Genetic algorithm (GA) used for the weight value of certain attributes - Multi-attribute utility (MAU) as a function to measure the usefulness of each item, including the value of the user's preferred attribute - Top-N utility element [22] 	<ul style="list-style-type: none"> - Same functionality as content-based recommendation 	<ul style="list-style-type: none"> - Evolution issue - Full re-computation for each new one - Users with unrestricted reasoning

¹ K-nearest neighbors

² One of the most popular clustering algorithms using unsupervised machine learning technique, and that takes only a data input vector without considering known or labeled results.

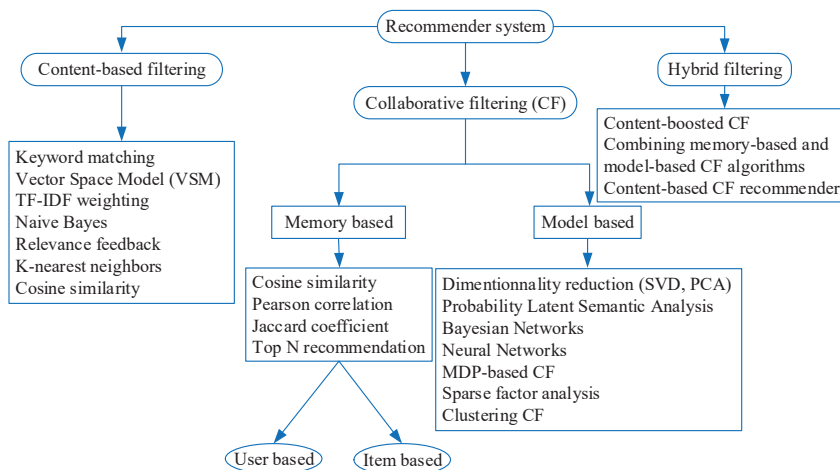


Figure 2 Recommender system approaches.

Memory-based CF uses available data input (clicks, votes, likes, etc.) to establish a similarity relationship between users (collaborative-filtering), as user-based or items (content-based recommendation) as item-based. In the case of collaborative filtering, recommendations are suggested based on items viewed by users close to us, hence the term collaborative.

In another way, content-based recommendation compares items using their characteristics (movie genre, actors, publisher or author of a book, etc.) to recommend new similar items.

In short, memory-based techniques depend heavily on simple similarity measures, namely Jaccard index, cosine similarity, Pearson correlation, etc., as shown in Figure 2, to match similar objects or users. For instance, in a huge matrix with users in one dimension and elements in the other, and with cells containing similar votes or elements, memory-based techniques use similarity measures on both matrix vectors to produce a score indicating the similarity.

The memory-based filtering techniques along with their advantages and drawbacks are presented in Table 2. As an indication KNN (user-based, item-based), Pearson or cosine measurements are the most commonly used techniques for this type of CF.

Model-based CF is a technique that attempts to fill out the matrix better. With such technique one can guess how much a user will like an item he has never seen before. To achieve this goal, numerous algorithms are used to train the vector of items for a specific user, and then construct a model that can predict the user's score for a newly added item to the system [33]. Thus,

a RS develops its intelligence to detect interesting items even before the users see and evaluate them.

The techniques used to implement model-based filtering as well as the other main approaches in various recommender systems, with their advantages and limitations are presented in Table 2. For instance, The Probabilistic Latent Semantic Indexing, Bayesian Networks, Singular Value Decomposition, Clustering CF, Neural networks, Sparse factor analysis or Markov decision-making processes can all be used in data sparsity reduction to allow better recommendation quality (Figure 2).

These techniques seem much more interesting for new items that appear and disappear quickly. They also present the reactions of users to items using the concept of latent factors of items and users [31].

2.3 Hybrid-based Filtering

Hybridization is the consolidation of multiple methods and techniques of recommendation to address the deficiencies of each method and capitalize on their strengths to enhance the efficiency of the system [34]. Hybrid RSs are used to address difficult problems such as serendipity, sparsity and cold-start [12].

Hybridization can be classified into many categories:

- (a) **Weighted** in which the simulated or predicted results from each technique are aggregated into a unique result.
- (b) **Mixed category** where predictions of numerous techniques of this hybrid are presented as ranks, then the main algorithm merges them into a single ranked list.
- (c) **Switching** category where the system toggles between the two types of recommendation techniques depending on the situation.
- (d) **Feature combination** where the data generated by the two techniques are merged and fed into a unified recommendation algorithm.
- (e) **Feature augmentation** uses the outcome of one technique as an entry for the other technique.
- (f) **Cascade** uses a sequence of two types of interventions, one to produce a preliminary ranking and the second to refine the recommended list.
- (g) **Meta-Level** uses the recovered model as an entry for the next technique, but not for the prediction list of results.

Recommender systems can also use one or a combination of the following techniques depending on the specific issue at hand, namely, Utility-based,

Table 3 Comparison between recommendation approaches

Dimension	Content-based Recommendation	Collaborative Filtering Recommendation
Input [15]	Content or Feature of items/users	User/Item Rating Matrix
Dependency [12]	User independence	User rating dependence
Novelty [3]	Over-specialization	Diversity
Serendipity [12]	Trend to recommend monotonous and obvious items	Ability to recommend unexpected items
Explanation [4]	Access to ancillary information about items	No access to ancillary information about items
Popularity [4]	Ability to recommend unpopular items	Trend to recommend popularly items
Other problems/ Issues [14]	User Cold-Start	Sparsity, User Cold-Start, Shilling Attack, Item Cold-Start

Semantic-based, Demographic-based, Psychological knowledge-based recommendation, Group recommendation techniques.

RQ2 summary

As a summary of RQ2, the following section explores the most recurrent challenges and problems in the context of text-review based recommender systems, considered as important aspects in research, also the standard techniques and algorithms manipulated by each article as well as their main contributions applied in a specific application domain to mitigate a certain number of issues. Each experiment is applied to the dataset (e.g., MovieLens) in either offline or online mode. Each mode measures the performance of the approach in terms of a given evaluation metric (i.e., MAE, MSE, RMSE, Table 8), in order to produce an enhanced recommender system with a lower evaluation metric than the other methods.

3 Main Issues and Research Challenges on Standard RS Techniques

Despite the growth in popularity of the recommendation systems, there have been serious drawbacks. As a follow-up to RQ3, most researchers are looking to develop a relevant recommendation system that meets the following criteria: Area Independence, Adaptability, Serendipity, Shilling attack problem, Data sparsity, Cold-start, Grey-sheep, Scalability and Diversity.

Area Independence is the view and realization of a recommender system to be applied across different fields and scopes.

Adaptability is the ability to adjust to a variety of new users.

Serendipity is the possibility to give the user the most unexpected, diverse and relevant recommendations. As content-based recommender systems recommend only items fitting the user's profile, the user will solely get suggestions that are similar to the items he or she has previously seen or evaluated. There is no way to obtain an unexpected recommendation and receive products never seen before. This may lead the user to get annoyed with the recommendations [27]. An effective recommendation algorithm should not only recommend what we are likely to like, but also suggest random and objective elements to help us discover other fields and areas [12].

Shilling attack problem arises when the scam use fake profiles with fake ratings to affect the recommendation process and influence the users' willingness to purchase items. There are two types of attacks, push and nuke. In push attacks, the items are recommended by giving them false ratings, while nuke attacks are carried out in order to negatively affect an item by reducing the probability that the item will be found [35]. The recommendations are reached according to the attacker's objective.

Data sparsity is the problem that occurs when the number of items evaluated by a user is very low compared to the total number of items available in the system. This leads to a very low density of items/user matrix. As a result, the recommendations become irrelevant, as the items are not all processed during this operation. This is a situation where the number of valid and relevant operations is fewer to establish useful similarities for users and target meaningful items [19]. For an individual who has evaluated items that have not elicited comments, it is difficult to identify similar users; therefore it will be complex to offer him suitable recommendations. The great challenge is therefore to overcome this gap by computing adequate preventions despite the low level of existing evaluations [35]. In this context, some solutions related to sparsity are proposed and use deep methods designed to examine and study the issue, as well as the predefined terms, search criteria, inclusion and exclusion criteria, together with the impact and compliance of existing research results [19]. To this end, these techniques (Factorization Machines (FM) [36], Social Balance Theory (SBT) [37] have been applied to better predict under sparsity and inform users of opposite tastes of similar users. Deep learning has also been in operation, in particular the vibrational auto-encoder (VA) [38] and the generated model using multi-layer perception (MLP) [39]. Restricted Boltzmann Machine (RBM) [40] also adopted Multi-layer deep structure to strengthen the effect of the recommendation where each element is perceived as that of others in the form of RBM that exchanges the same

weight, bias as that of other users. Hierarchical Bayesian approach [41] for deep collective learning is also suggested. Regarding the latent factor model, Matrix factorization (MF) remains the most popular technique to define new underlying factors associated with already targeted variables [42]. The Review-based recommendation studies, described in section IV, also alleviate this issue in many ways.

Cold-start problem is due to the fact that a user has not rated any item. Such a problem often occurs when a new user or article is added. It also puts the user back on the border independently of the others. Two types of cold-start problem can be distinguished:

- **Item-cold-start**, this is often one of the major problems of collaborative filtering techniques (Tables 2 and 3). When a new item is added to the system, it might have some content information but no evaluation [26, 42, 29 and 30]. Commonly, if there is no reaction, the collaborative filtering technique cannot recommend the item in any case. Meanwhile when a little interaction is out there, the algorithm will predict with very low quality. Consequently, this will lead to another problem called unpopularity items [27].
- **User-cold-start**, this is frequently one of the major problems of collaborative filtering techniques (Tables 2 and 3). When a new user is registered in the system, the RS provides him/her with items without counting on the history of interactions of the user that have not yet taken place [27]. The moment the user is subjected to poor quality recommendations, he may decide to leave the system even before the RS understands his interests [12].

Grey-sheep problem is an issue where users have unique or exotic tastes. This makes it hard to identify their similar interests with other system users [48]. In the context of e-commerce, gray sheep users have a real impact on the ability to carry out effective shopper marketing. To address some specific groups of customers with special offers or advertisements, the organization must identify both products and customers that would interest them. The question is how to affect a “grey-sheep” client to a particular group of users’ profiles [8]. To identify them, the author proposes a simpler approach in terms of complexity by using a mechanism of identification of the values based on the distribution of similarity between different users (KNN) or the analysis of the correlations between one user and the other [43].

Scalability is the problem that happens when the number of items and users increases, and the cost of recommendation computation increases as

well. RSs with a large number of items and users tend to increase the computation time rather than the prediction relevance [32, 44]. Bondi [45] has defined two types of scalability: structural scalability and load scalability. Structural scalability is in what way, the system can scale up without making significant changes to its architecture. While Load scalability is the ability of the system to adapt to the growth of data flows offered. Indeed, systems with low load scalability are likely to face this problem because they operate inefficiently, have inadequate algorithms, or cannot take advantage of the parallelism principle. Different techniques such as Matrix Factorization (MF) and Slowly Changing Dimensions (SCD) [19], Probabilistic Matrix Factorization (PMF) [26], Recurrent Neural Network(RNN), Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) [27] have been used to address this kind of problem, by using some information features like contextual information, social hashtag, demographic data, semantic data, on service logs and browsing and session history.

Diversity is the variety of recommended products that allows covering the interests of the user and can allow him to see new items. Diversification is now one of the main solutions to most effectively address the problem of overfitting in recommender systems. Several researches address this aspect, which has become one of the key issues in recommender systems [28]. There are still many challenges to diversification and the most significant, in our view, is while they did question the users about the perceived diversity of recommendations, they did not question how the users actually define diversity and use this knowledge to develop an appropriate diversity measure. Further users studies with expanded questions could therefore provide some benefit when developing new definitions of diversity [42]. The problem that arises is how to handle systems that work with many different item types (collaborative RS for example) or with items that do not have all the meta-data available (user created content for example). A sort of generalization should therefore be proposed with such measures in order to simplify their use in different systems [48]. Diversity should be considered during the recommendation process instead of being applied during the post-recommendation process, as is the case with most of the algorithms presented in this review. If a system suffers from overfitting, one can come across a case where all of the recommendation items will be more or less the same, which will cause the diversification process to fail. Diversification should therefore be present from the start of the recommendation procedure and should be included in the process of ranking/calculation of predicted ratings [3]. Future research on diversification of recommender systems should therefore focus not only on

developing new algorithms, but also on finding an evaluation measure that would reflect the average user's perception of diversity [28].

Classical recommender system methods suffer from a recurring problem because they are based on users' numerical ratings as the main source of information about users' interests [44]. Unfortunately, scalar ratings are often not supported by a good semantic interpretation to illustrate the user's true expectations, leading to considerable degradation of the recommendation's relevance [14]. To overcome this setback, some recommendation approaches associate both user reviews and ratings.

In this section, as a summary of RQ3, we have dealt with potential solutions based on text-reviews, which can be considered as reliable to solve the main challenges raised. Finally, we have made contributions to the new future tendencies of this field.

RQ3 summary

4 User Review and Ratings

The important information gathered from reviews has been summarized and utilized to improve the standard recommendation approaches discussed. Habitually, a large evaluation catalog called user-item matrix is filled with the ratings given by a user for any item used as a single source of information to predict the missing ratings in the matrix based on the available ratings and suggest a relevant item to a user by referring to the choice of its neighbor [47].

In case of the sparsity of the user-item rating matrix due to explicit ratings of few items, the quality of recommendation strongly decreases especially in the diversity of the suggested items. Therefore, many researchers consider review content information to solve this issue [48]. To address this matter, various recommendation approaches incorporate user-review and ratings. Typically, the evaluations are expressed as texts of reviews and sample words, reflecting the different actions or opinions of a user towards a specific item [42]. Moreover, even though the raw review information is in unstructured textual form that cannot be easily interpreted by the system, developments in the field of opinion mining (sentiment analysis) and topic modeling make it possible to process the reviews and derive useful information from them [49]. In the past, many elements that could be extracted by the recommendation system have been represented [19, 42, 44]. For example, Opinion by sentiment (negative or positive) [33], Keywords, Semantic data or No-demographic data, Data knowledge, Contextual information, Acoustic

features are some of these elements. In the following subsections, some of these elements are briefly discussed as to how they can be used in a recommender system.

Review terms: A user review is expressed in natural language; therefore the easiest way to examine it is to select the most significant elements. A weight measure as TF-IDF [18] can be used to determine the relevance of each term. The extracted terms may then be used to measure the similarity between different users instead of numerical review ratings in a user-item matrix of the CF approach [19].

Review topics: Topics are defined as the aspects of an item that the author describes in his review. Thus in the statement “the laptop’s storage capacity is huge”, the topics mentioned include the laptop and its storage capacity. There are numerous techniques for recognizing topics in reviews, such as frequency-based techniques for identifying topics in reviews. The first one is the frequency-based approach, which firstly identifies the most frequent terms based on a set of words and then manually classifies them into topics or automatically, based on a predefined dictionary [18]. The second one is conditional random fields [35], syntax and topic modeling technique and Latent Dirichlet Allocation (LDA) [16] to automatically detect hidden topics in review texts, Probabilistic Latent Semantic Analysis (PLSA) [31] or Latent Semantic Analysis (LSA) [50]. Review topics can be associated with the similarity measure in the neighborhood-based CF [47] and latent factors in the model-based CF approach [14]. Detected review topics can also be used to enhance the actual evaluations in classical CF approaches [25].

Contextual opinions: The statement “first of the day sports by the sea” provides contextual information. This other statement “the battery life of the laptop is bad when I execute several tasks at the same time and heat up in the summer” provides contextual information about the review. The “in the summer” quote is the context; “the battery life of the laptop” quote is the feature and the adjective “bad” is a negative opinion. These opinions can be joined with ratings to infer the relevance of a user’s selection in various contexts [30], to represent the interests of the feature concerning the context or the latent factors regarding the context of an individual [2]. As aspects, reviews contexts are typically fixed or learned contexts that are retrieved automatically. They can be selected using rule-keyword matching, based reasoning or using a classifier like LDA classifier, clustering via k-means & cuckoo techniques [3]. Review context has also proven its benefit in enhancing recommendation performance, by either joining it with explicit rating to

Table 4 Primitive issues being addressed by researchers based on review terms

Issues Articles	Approach Used	Review Element User/Item Profile	Main Contribution	Tested Dataset	Accuracy Performance	
					Achieved Accuracy	Baselines Approach
Salakhutdinov and Mnih [18] (Linear learning, Support Vector Machines (SVM), Naive Bayes (NB), (KNN), Supervised Learning Models)	Hybrid approaches	Aspect Sentiment words, tweets, TF-IDF weighting words	Enhance accuracy of tweet sentiment analysis. Reduce language ambiguity noises. Evaluation metric: ROC curve	25,000 IMDB Movie Reviews and 25,000 tweets, MDB Movie Review Datasets, Twitter	IMDB movie review: 0.908 Twitter dataset: 0.769	Perceptron: 0.890
Suglia et al. [16] (RNN, LSTM, top-N recommendation, logistic regression layer)	Content-based, Deep neural networks	Latent factor ratings, latent topics and item review words (Genres of both movies and books, textual descriptions of the items, the plot of the book)	Cover data sparsity. Evaluation metric: F1@10	DBbook data (Books), ML1M (Movies)	AMAR+: 0.551 on ML1M data AMAR+: 0.662 on DBbook data	Popularity: 0.643 U2U: 0.630 I2I: 0.641 BPRMF: 0.634 WRMF: 0.63 TF-IDF: 0.640 Word2Vec (W2V): 0.655 TTM
Gong and Zhang [50] (Local attention channel and global channel (CNN+Attention-5))	Attention-based, CNN, Deep neural networks	Latent feature rating and latent factors from keywords.	Solve the hashtag recommendation task as a classification problem. Evaluation metric: F1, Recall, Precision	Dataset collected from areal microblogging service (Social networks, Semantic analysis, Machine translation, Speech recognition)	F1 = 0.388 Recall = 0.357 Precision = 0.439	F1 = 0.363 Recall = 0.351 Precision = 0.377

Table 5 Primitive issues being addressed by researchers based on reviews topics

Issues Articles	Approach Used	Review Element User/Item Profile	Main Contribution	Tested Dataset	Accuracy Performance	
					Achieved Accuracy	Baselines Approach
McAuley and Leskovec [51] Hidden Factors as Topics (HFT)	Collaborative filtering	Latent topics by LDA	Enhance the rating prediction accuracy Handle the sparsity problem Evaluation metric: MAE	Amazon (movies, books), Yelp (restaurants), Citysearch, Beeradvocate and Ratebeer (wines, beers)	1.253	HFT: 1.178 Latent factor model: 1.262
Chin et al. [52] (ANR)	Collaborative filtering	Aspect importance and latent aspect ratings by Aspect-based Neural	Increase prediction accuracy Evaluation metric: MSE	Yelp, Amazon (businesses)	1.009	ALFM: 1.075 DeepCoNN: 1.178
Cheng et al. [53] (ALFM)	Collaborative filtering	Rating factors and latent topics by Matrix Factorization	Alleviate data sparsity and Interpretability problem Enhance prediction accuracy Evaluation metric: RMSE	Yelp, Amazon (businesses)	0.893	BMF: 1.004

Table 6 Primitive issues being addressed by researchers based on reviews context

Issues Articles	Approach Used	Review Element User/Item Profile	Main Contribution	Tested Dataset	Accuracy Performance	
					Achieved Accuracy	Baselines Approach
Rhul and Om [3] (K-means & cuckoo)	Clustering-based Collaborative filtering.	Latent aspect ratings with time, location features	Alleviate ranking, Overfitting and Sparseness problem. Improve prediction speed and accuracy. Search optimization Evaluation metric: MAE	MovieLens 100k (Movies)	0.682	PCA-CAKM: 0.976 GAKMCLUSTER: 0.763
Zhang et al. [54] A hybrid of NSMF and AdatNSMF	Collaborative filtering	Ratings and position features	Overcome ranking Overfitting and sparseness problem Evaluation metric: RMSE	MovieLens 100K (Movies), MovieLens-1M (Movies)	0.937	MF: 0.958 NS: 3.218 NSMF: 0.928

provide a user's rating for an item in a particular context [54], or by applying it in user modeling as proposed by Rhul and Om [3] using context-sensitive aspect preferences or context-independent aspect preferences.

Review emotions: Emotion is the mental state of a person (for example, happiness, sadness, distress, joy, etc.). It is difficult to detect it in reviews than in opinions. On the other hand, an emotion classifier can be designed to label a text with one or more emotions [55]. Emotions gathered from reviews can be used to define the probability that a user likes an article [57]. For example, emoticons (emotional and distressed faces), which represent emotions symbolically, can also be combined with opinions.

4.1 Techniques Considering User Review Terms

These techniques are based on review terms such as keywords gathered from user-reviews where each word k_j is assigned a weight $U_{i,j}$ by TF-IDF algorithm that indicates its interest for the user U_i [56]. In the same way, the term-based item index T_{P_i} is defined using terms taken from the reviews posted on item P_i . In fact, the index-based approach that is an upgrade of the content-based approach, makes the correspondence between the user index and the indexes of items, and finds the most similar items [56] during the recommendation process. Following the evaluation conducted on the Flixter dataset, which consists of 43,173 reviews expressed by 2,157 users towards 763 films [44], the result indicates that the accuracy of the index-based approach is lower than CF classification methods based on ratings, but more effective in terms of diversity (average dissimilarity of products in the recommendation list) [42] and novelty (amount of different or new products recommended to an individual) [57].

In the same way, the improved version of the user-based method is also used, but this time to calculate the similarity between individuals by referring to textual reviews similarities, instead of the ratings for each evaluated item. These similarity scores are thus included as weights during the prediction of ratings [58]. For this purpose, the convolutional matrix factorization model (ConvMF), which uses word embedding and convolutional processing to retrieve latent information of articles from their comments is proposed. Then, the induced latent characters are integrated into a matrix factorization model to obtain the users' ratings of targeted articles [59].

The NARRE model [60] that uses convolutional neural networks (CNN) has also been proposed to extract latent incorporations of items and users from text reviews, and to estimate missing scores. The latent factors of users

Table 7 Primitive issues being addressed by researchers based on reviews emotions

Issues Articles	Approach Used	Review Element User/Item Profile	Main Contribution	Tested Dataset	Accuracy Performance	
					Achieved Accuracy	Baselines Approach
Ebesu and Fang [42] Neural Semantic Personalized Ranking (NSPR) with two variants: Logistic (NSPR-L) and Probit (NSPR-P)	Deep neural network Deep learning in collaborative filtering Item-based CF	Inferred ratings from opinion emotional classification	Alleviate Cold-start, Sparsity and Implicit user feedback problem. Evaluation metric: NDCG@10, R@300	Citeulike-a (share citations to papers), Yahoo! Movies (Movies)	NDCG@10: 0.038 R@300: 0.526	SVDFeature NDCG@10: 0.003 R@300: 0.125 CMF NDCG@10: 0.001 R@300: 0.155 CTR NDCG@10: 0.004 R@300: 0.090
Nassar et al. [29] (DNN)	Deep learning Collaborative filtering User-based CF	Ratings with aspects' opinion ratings	Overcoming the Cold start problem. Evaluation metric: MAE	TripAdvisor (Hotels)	0.755 ± 0.0050	Single DNN 0.785 ± 0.015 KNN with baseline 0.811 ± 0.004 SVD++ 0.822 ± 0.007 Hermes: 1.687
Nozari et al. [21] (IBGR) Fuzzy C-Means (FCM) clustering method and the Pearson Correlation Coefficient (PCC) similarity	Group recommender systems	Latent rating factors inferred from trust, partnership and members influence words merged with aspect emotions scores	Establishing compatible groups. Improve the efficiency of the influence process and leader identification. Help people find their favorite items, in the shortest possible time, out of a wide range of existing items. Evaluation metric: RME	MovieLens 100k (Movies)	0.981	

with attention scores are integrated into a factorization machine (FM) [4], in contrast, ratings information and reviews were merged into a common model. The so called model uses an attention mechanism and CNNs to identify suitable latent properties by referring to their reviews. Finally, the model establishes latent ratings by rating-based component for elements and users from the interaction matrix. To infer the final rating, the selected content features and latent ratings are added to an MF.

4.2 Techniques Considering User Review Topics

A new recommendation technique using aspects extracted from reviews combined with ratings to generate predictions, such as the Hidden Factor and Topic (HFT) technique, which associates ratings with review topics [51]. It starts with Matrix Factorization (MF) to model the ratings and LDA-based topic model to present the text of reviews. Then the items are designated as a topical list, and the topics with high ratings are repeated to increase their impact. In addition, users are distributed in a similar topical field through their ratings. To predict ratings, the observations of users and items are incorporated into a latent factorization model [42]. Similarly, the Hidden Factor and Topic (HFT) technique uses classical MF to model ratings and LDA-based topics for reviews, of which Softmax transformation function is exploited to incorporate latent topics during learning the latent feature model. Ratings are then obtained from the built model.

Moreover, Aspect-Aware Latent Factor Model (ALFM) that is based on Aspect-aware Topic Model (ATM) has been also proposed to represent the items/users according to aspect as composite topic repartitions, each one is displayed as a set of terms. The observations obtained from ATM are combined with latent evaluation factors in ALFM to generate missing ratings using the MF method [52].

Moreover, since most users put only one rating and one comment, two algorithms (ICFTM and UCFTM) have been proposed to produce recommendations by taking into account user reviews and user ratings. Thus, the sparsity problem is treated by calculating the similarity of items or users through a topics model [Article topics]. With this regard, the comments are examined using the topic model (ASUM) in order to produce the appropriate topic allocation. Then the topic allocation of the element (or user) is inferred from the topic allocation of the review. Afterward, the similarity of items (or users) is calculated on the basis of the most appreciated characters of the item (or user).

In the same way, the Aspect-based Neural Recommender (ANR) technique is developed using a neural network to determine the importance of ratings of latent aspects. The latent aspect ratings are obtained by calculating the weighted sum of all the terms included in the reviews. The importance of latent aspects is also deduced based on the common similarity of each pair of latent aspect ratings of users' items. As a matter of fact, the synthetic rating of every user-article pair is calculated by combining aspects' importance with aspect ratings within the Modified Latent Factor Model (MLF) [53].

4.3 Techniques Considering User Review Context

Authors often use the notion of rating (User * item => rating) to express user preferences. The advent of smart devices such as mobiles and the ability to use them anytime and anywhere, has given more opportunity to recommend relevant items to users (User * Context => RatingContext) [38]. According to [42], a context represents any information that describes the state of an object (time, location, product, person, etc.) and is designated as an essential element for processing user preferences that have not been taken into consideration by previous evaluations in accordance with laws on the books (GDPR) and other anti-tracking issues. Imen et al. [61] used contextual data to propose guidelines to help the user make a better decision from a large number of possible actions, such as the movie to watch, the diet for a specific disease, the place to visit and when to go on a trip, and the friend to add to the social network. On the other hand, the long-term tastes of users are expressed not only in the form of items and users, but through three main elements: user, item, and context [62]. Clearly unlike traditional systems, contextual recommender systems integrate contextual information beyond the user and item's frame, to estimate ratings on items not ever seen by the user [23]. In practice, the integration of contextual meaning in a recommender system can be fulfilled using Contextual pre-filtering approach, Contextual modeling approach and Contextual post-filtering approach [63].

Contextual pre-filtering approach uses the contextual information as a label to remove measures that are not so adequate with the adopted context. This is done before the traditional recommendation procedure is run on the selected data set. If a context C is considered, the approach selects all appreciations related to the specified C context from the basic set, and builds the "item * user" matrix including only the data related to the C context. Then, the recommender system approach, namely collaborative filtering, is applied to the reduced dataset to remove the items related to the C context.

Contextual modeling approach involves the contextual information within the algorithm operated in the multi-dimensional recommender system framework during the recommendation phase (Users * Items * Context).

Contextual post-filtering approach uses contextual information after the main two-dimensional recommendation approach (User * Item) is launched. Once the unknown ratings are assessed and recommendations are produced, the system analyzes the data for a particular user in a precise context to find models of usage of specific items and uses them to contextualize the recommendations (Item * User * Context) resulting from the classical recommendation approach (Item * user), like collaborative filtering.

Yue et al. [30] developed a new context-sensitive movie recommendation algorithm based on joint matrix factorization (JMF) to recommend movies based on mood tags of movies and keywords of their stories. It is used to exploit contextual similarities between movies (mood-specific, movie similarity) and similarity between movies based on PK-based movie similarity. This algorithm can be adapted to other types of recommendations, such as movies with a specific actor or music with a specific genre.

Another method of review rating prediction (RRP) based on user context (UC) and product context (PC) has been proposed. It begins by modeling the correlation between the sentiment word and the user context and presents a prediction approach of review rating that associates content review and user context. Then, it models the correlation between the sentiment word and the product context on which it models the review rating prediction approach based on review content and product context. Lastly, it offers a rating review prediction approach based on the user and product context [64].

4.4 Techniques Considering User Review Emotions

Sociologists and psychologists have shown that information that evokes emotions and feelings is a strong indicator of users' interests (agreement or disagreement) [57]. The emotional aspects of reviews (i.e., happiness, sadness, surprise, fear, shame, etc.), which represent the attitude and state of the mood of users during the draft of reviews, can also be used to learn about the interests of individuals.

Ultimately, the emotions described in reviews have been used to evaluate the possibility a user will like a product [49, 65]. For example, each movie is evaluated according to three types of spaces: the movie, the emotion and the semantic space. To identify emotional features, Ortony et al. [55] used the

OCC model and the emotion classifier developed by Moshfeghi et al. [65] to check whether a type of emotion is expressed in the content.

As another example shown in Table 7, Ebesu and Fang [42] transform reviews into global emotions scores using a machine learning method. To do this, the review vectors are combined with the actual user ratings to learn a Naive Bayes model of the positive and negative classes. This model is then used to deduce the new evaluations from the reviews retrieved using the deep neural network technique. To better predict the evaluations given by the reviews and used to build an evaluation matrix that is incorporated into the traditional neighborhood-based CF techniques.

On the other side, Nassar et al. [29] introduced a user preference-based FC that integrates aspect-level information to give a reflection of users' interests from reviews. Particularly, two measures for aspect interests have been suggested, namely, aspect importance and aspect need to express the difference in opinions on aspects and the relationship between explicit rating and aspect. Using these measures, the authors computed the similarity within users, which is later incorporated into the memory-based CF for making further recommendations.

As a feature field (c) and a user with his evaluation history (u), the probability that a user appreciates a movie (m) is $P(m, u, c)$, they elaborated the Latent Dirichlet Allocation (LDA) model to designate the evaluator as a probability distribution in some latent groups and items that are appreciated by individuals of the same group described as a probability distribution. The probability $P(m, u, c)$ is obtained by marginalization over these latent sets. Finally, Ortony et al. [55] used a classical machine learning method (gradient boosted-tree) [66] to combine the predictions of the three kinds of feature spaces to obtain the ultimate prediction.

Moreover, emotions as symbols of emoticons can be combined with opinion words to deduce the average of each reviewer. As the approaches in the literature have not been able to overcome the problems of data sparsity, overloading and cold-start, the integration of emotions in reviews can help diminish the risks of cold-start, data sparsity and overloading [67]. In the same way, WIRROR framework, which uses positive or negative emotions in reviews and neutralizes emotion in local or global contexts, has been proposed to improve the relevance of recommendations. In fact, emotion signals can be employed to infer user satisfaction with a product even if the user has not reviewed the product. The integration of this information is based on matrix factorization. Thus, the identification of emotional signatures of spam and irrelevant data can help establish a solid model for appropriate RS [67].

Furthermore, Personal Assistance Agents (PAA) are used to extract, filter, and provide only relevant information to users [59]. Having that in mind, there is no research currently in the field of real-time textual processing that answers the gap-based emotions. The implementation of real-time RS based on emotions would be crucial, requires to take part the theory of the emotion desired as well as approaches of natural language process, elaborated in the fields of artificial intelligence and the MAS (Multi-agent system) [65].

In order to evaluate the accuracy and the performance of the approaches, the following section describes the evaluation measures techniques adopted, and their drawbacks in terms of the use context.

5 Approach for Evaluating Recommendation Systems

Various measures are used by RS researchers to evaluate the performance of the approaches adopted and to determine its effectiveness in a given recommender system [33]. On this basis, two types of evaluation are performed; one is used to train the system, while the other is for testing the user's preferences produced. System relevance is not limited to the possibility of producing useful recommendations for a user [19], neither to the ability that these recommended products are consistent with the preferences of each user even in a sparse database.

The question that often arises is what value of relevance measure should be provided when evaluating a recommendation system following the approach being developed. To respond to this question, the most commonly used evaluation metrics in RSs have been in Table 8 grouped along with their characteristics, definitions, limitations and formulas.

We have focused on the evaluation results generated by each evaluation metric to select the best values for the contributions processed, the approaches, and the recommendation techniques used in a specific domain as shown in Table 8. Evaluation methods can be categorized into two distinct approaches [68]: The first evaluation approach is online; it recommends products to users and asks them how they perceive the received items. The second approach is offline, the user interaction in this case is not required, but the approach uses the history of user/item interactions to recommend (top-1 items) or a set of items (state-based recommendation) [68]. Currently, the offline approach [32] is still the most used because of its ease to handle and it only takes little time to process even though the online approach provides information that is more relevant for a user [71].

Table 8 Evaluation metrics considered in the papers reviewed

Characteristics	Metrics	Formulas	Explanations
Classification problem	Recall [44, 50]	$\frac{TP}{(TP + FN)}$	(1) – Percent of truly positive instances that were classified such as Sensitivity, coverage, true positive rate
Accuracy	Precision [50]	$\frac{TP}{(TP + FP)}$	(2) – It computes how often are we correct in our positive prediction – Positive Predicted Value (PPV)
	Root Mean Squared Error : RMSE [19, 54]	$\sqrt{\frac{1}{n} \sum (y_j - \bar{y}_j)^2}$	(3) – Root Mean Squared Error is the square root of the Mean Squared error. It measures the standard deviation of Residuals.
	Normalized mean absolute error: F1-measure [16, 50]	$2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$	(4) – Harmonic mean of precision and recall – Utilize the precision and recall creating a test's accuracy through the "harmonic mean". It focuses on the left-bottom to right-top diagonal in the Confusion Matrix.
Cold-start	Click-Through Rate : CTR [69]	$\frac{Total\ Clicks}{Total\ Impressions} \times 100$	(5) – It computes the proportion of recommendations ultimately clicked
User satisfaction	receiver operating characteristic curve : ROC curve [18]	Sensitivity: $\frac{TP}{(TP + FN)}$ Specificity: $\frac{TN}{(TN + FP)}$	(6) – It measures the rate of items that the user likes in the recommendation list (7)

(Continued)

Table 8 Continued

Characteristics	Metrics	Formulas	Explications
Diversity	Normalized Discounted Cumulative Gain: NDCG [42]	$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)} = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2(i+1)}$ $IDCG_p = \sum_{i=1}^{ REL } \frac{2^{rel_i} - 1}{\log_2(i+1)}$ $nDCG_p = \frac{DCG_p}{IDCG_p} \quad (8)$	- It is a measure of ranking quality
Novelty, Serendipity	Novelty [57, 43]	$\sum_{i \in L_{rec}} \min_{j \in L_{his}} dis(class(i), class(j)) \quad (9)$ <p>Where L_{his} is the list of the user's history, dis is a distance measure, class (i) and class (j) indicate the classes of items i and j, respectively.</p>	- It calculates the novelty of the recommendations offered
	Others [23] /AUC & [29]/MRR		

TP (True positive): outcome where the model properly predicts the positive class.
 TN (True negative): outcome where the model properly predicts the negative class.
 FP (False positive): outcome where the model mispredicts the positive class.
 FN (False negative): outcome where the model mispredicts the negative class.

6 Discussion and Feedback

Most studies have been conducted on user review text (Section 4) to optimize classical recommendation techniques (Section 2). For instance, in the first group, text-based ratings from Lexicon Sentiment reviews can be used to override explicit ratings in memory-based CF approaches [36]. Latent incorporations of items and users obtained by attention-based CNN and deep neural networks from reviews can be employed as properties in latent matrix factorization (LMF) to carry out rating predictions. KNN (user-based) similarity can also be improved by using reviews and Lexicon Sentiment [18]. Furthermore, neural network techniques [50] have confirmed their effectiveness in comparison with techniques based on modeling topics with a negative impact on deep textual properties through coarse-grained text mining algorithms. Besides, they are effective in the presentation of neural networks for mining relevant semantic elements of review texts to identify items and users.

The second group focuses on the use of explicit ratings adjusted by implicit ratings inferred from review-text to get reliable and accurate ratings. As an illustration, through integrated CNN structure, the studies as shown in [21] & [50] have developed numerical ratings in standard latent factor model in coordination with latent properties vector coming from review terms. Different classical latent factor models have been modified, such as ALFM, JMARS, HFT, TopicMF and RBLT to enhance real ratings adjusted with latent topics in the text reviews [51, 53]. In addition, traditional probabilistic MF has been optimized by tuning real ratings with sentiment notations extracted from text reviews [70].

The third group explores specific supplementary information from review texts to improve the effectiveness of the system according to the evolution of the user's temporal preference and the benefit from a large amount of data. As an illustration, Semantic (non-demographic) and demographic data derived from text reviews are often combined to fill in the gaps due to the cold-start and sparsity problem [4]. Social tweets, social tags and social hashtags gathered from reviews are also used as side sources inferred by social-based recommender systems through RSboSN (Recommender Systems based on Social Networks) technique [44]. In fact, the use of a recurrent neural network is more efficient to extract the temporal information constituting the latent factors of users in collaborative filtering [36]. Additionally, the merging of different types of contextual information in recommendation such as text, image, music and the integration of other deep learning methods like the

attention mechanism certainly help to improve the quality of the recommendation according to the user's changing behavior, interests, preferences and profile as a result [17].

This research contributes not only in terms of evaluations, but also helps to propose model and strategies to overcome the inability of these existing systems to make an appropriate recommendation to the users. This covers the systems that focus only on single criteria recommendation, this is one of the future trends that can be explored to enhance the effectiveness of recommender systems.

Firstly, all studies that take up combinations of two features from reviews have proven that accuracy of RS is enhanced over any other baseline. Thus, it is necessary to build up multi-criteria recommendations that look at more combinations of features to enhance the reliability of recommendations [29].

Another forthcoming tendency is the precise profiling for items and users based on the features retrieved from the user's reviews. Accordingly, precise profiling of the items and users is essential to get an accurate recommendation. Although, there is a limited number of studies that focus on organizing and using review features to build or optimize profiles of items and users [20].

Another future trend is the enhancement of the current review-based multi-criteria RS assessment process, as it has some restrictions, particularly in terms of the type of baselines with which it is mapped. For instance, the CB approach that handles reviews to construct a user's profile has interfaced with profiles that are constructed from static item details [48] and not with others that are built using reviews, and approaches that employ reviews to render item rankings have matched their approach against the popularity-based approach, not with the approaches that implement traditional preference ranking [29].

Finally, the ample information gained from user evaluations can be employed to generate some explanation to users throughout the recommendation process. This explanation will raise the user's confidence to use the recommender system because it represents and gives reasons why these recommendations are made to them. Currently, there is a huge gap in the literature on this issue.

In the long run, three major problems (sparsity, cold-start and serendipity) used as references by 70% of the studies to improve the accuracy and efficiency of prediction systems ([43]) performed in less sparse datasets such as *incarMusic* (Figure 3) using improved similarity measures compared to traditional similarity measures (e.g., fusion of PCC and Jaccard index, JacUOD, JMSD, etc.) to address their main limitations [19].

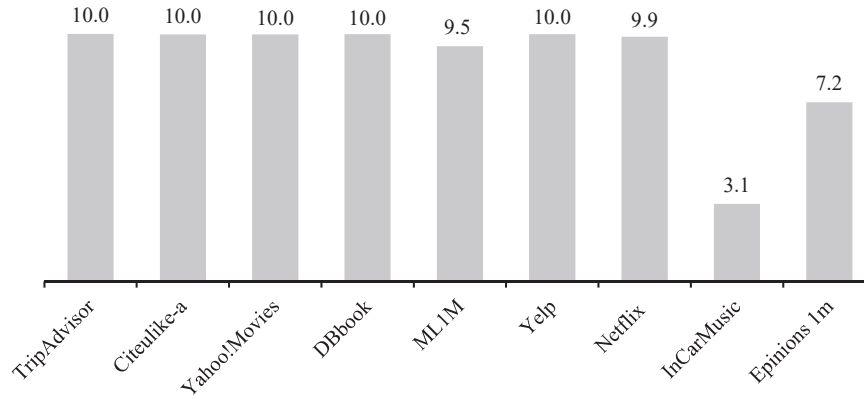


Figure 3 Degree of sparsity (%) of the datasets.

7 Conclusion and Future Work

Nowadays, due to the emergence of text mining, context, semantic data, filtering, normalization, and relevant data recommendation techniques, tremendous efforts have been devoted to addressing the major challenges affecting the performance of recommendation and incorporation of review texts. Various review texts are used to fill gaps in the classical rating-based model and offer relevant items in conformity with users' interests (e.g., review terms, review emotions, review topics and review context opinions). In this paper, a detailed analysis of the main challenges impacting the performance of a recommender system, cold-start, serendipity, sparsity, overfitting, grey-sheep, accuracy and evaluation metrics, has been provided. The techniques used to overcome each challenge are also described, as well as the different user review text manipulated and how they are exploited to enhance the standard user profile rating features. At last, review-based recommender systems in terms of improving prediction accuracy and overcoming the sparsity and cold-start problems are as well discussed. The discussion and findings of these studies have led to various recommendations and perspectives for the future.

Regardless of the progress and development of new algorithms in the RS area, we noticed through our review of different RS techniques, that further work is needed. For instance, the combination of several RS review-based methods and RS classification could be more effective in determining the best user preferences. Additionally, another potential research area could be the implementation of intelligent text exploration approaches to better manage

the combination of ratings and review texts using multi-criteria and multi-attribute systems.

All of the chosen articles have been published during the last ten years and the selected articles have been cited by others, which makes them more credible.

We hope that this study will allow researchers gain a better idea on recommendation systems using multi-criteria reviews and persuade them to exploit the implicit ratings of a given review in their future studies.

Acknowledgements

I am grateful to the Hassan II Institute of Agronomy and Veterinary Medicine for its extensive support. I also wish to express my thanks to Mustapha NAIMI, Professor at the Hassan II Institute of Agronomy and Veterinary Medicine and Mr. Abderrahman El HADDI, CTO and Founder of Enduradata Corporation, and Mr. Brahim HMEDNA, Professor at the University of Ibn Zohr, for their valuable comments and suggestions.

References

- [1] P. Nagarnaik and A. Thomas, "Survey on recommendation system methods," in 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 2015, pp. 1603b–1608. <https://doi.org/10.1109/ECS.2015.7124857>.
- [2] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities," *Appl. Sci. (Basel)*, vol. 10, no. 21, p. 7748, 2020.
- [3] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search," *Egyptian Informatics Journal*, vol. 18, no. 2, pp. 105–112, 2017.
- [4] N. Idrissi, A. Zellou, O. Hourrane, Z. Bakkoury, and E. H. Benlahmar, "A New Hybrid-Enhanced Recommender System for Mitigating Cold Start Issues," in *Proceedings of the 2019 11th International Conference on Information Management and Engineering*, London United Kingdom, 2019, pp. 10–14.
- [5] P. Jariha and S. K. Jain, "A state-of-the-art Recommender Systems: An overview on Concepts, Methodology and Challenges," in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1769–1774.

- [6] J. Valverde-Berrocoso, M. C. Garrido-Arroyo, C. Burgos-Videla, and M. B. Morales-Cevallos, "M. del C. Garrido-Arroyo, C. Burgos-Videla, and M. B. Morales-Cevallos, "Trends in Educational Research about e-Learning: A Systematic Literature Review (2009–2018)," *Sustainability* (Basel), vol. 12, no. 12, p. 5153, 2020.
- [7] J. Konaté, A. G. Diarra, S. O. Diarra, and A. Diallo, "SyrAgri: A Recommender System for Agriculture in Mali," *Information* (Basel), vol. 11, no. 12, p. 561, 2020.
- [8] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [9] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Adv. Artif. Intell.* vol. 2009, pp. 1–19, 2009.
- [10] K. Nageswara Rao, "Application Domain and Functional Classification of Recommender Systems - A Survey," *DESIDOC J. Libr. Inf. Technol.*, vol. 28, no. 3, pp. 17–35, 2008.
- [11] R. G. Lumauag, A. M. Sison, and R. P. Medina, "An Enhanced Recommendation Algorithm Based on Modified User-Based Collaborative Filtering," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, Singapore, 2019, pp. 198–202. <https://doi.org/10.1109/CCOMS.2019.8821741>.
- [12] N. Idrissi, A. Zellou, O. Hourrane, Z. Bakkoury, and E. H. Benlahmar, "Addressing Cold Start Challenges In Recommender Systems: Towards A New Hybrid Approach," in *2019 International Conference on Smart Applications, Communications and Networking (SmartNets)*, Sharm El Sheik, Egypt, 2019, pp. 1–6.
- [13] A. A. Kardan and M. Ebrahimi, "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups," *Inf. Sci.*, vol. 219, pp. 93–110, 2013.
- [14] L. Liao and F. Köttig, "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction," *Appl. Soft Comput.*, vol. 44, pp. 191–199, 2016.
- [15] S. Gong, H. Ye, and H. Tan, "Combining Memory-Based and Model-Based Collaborative Filtering in Recommender System," in *2009 Pacific-Asia Conference on Circuits, Communications and Systems*, Chengdu, China, 2009, pp. 690–693.

- [16] A. Suglia, C. Greco, C. Musto, M. de Gemmis, P. Lops, and G. Semeraro, "A Deep Architecture for Content-based Recommendations Exploiting Recurrent Neural Networks," in Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava Slovakia, 2017, pp. 202–211.
- [17] Z. Huang, C. Yu, J. Ni, H. Liu, C. Zeng, and Y. Tang, "An Efficient Hybrid Recommendation Model With Deep Neural Networks," *IEEE Access*, vol. 7, 137900–137912, 2019.
- [18] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in Proceedings of the 25th international conference on Machine learning - ICML '08, Helsinki, Finland, 2008, pp. 880–887. <https://doi.org/10.1145/1390156.1390267>.
- [19] N. Idrissi and A. Zellou, "A systematic literature review of sparsity issues in recommender systems," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, p. 15, 2020.
- [20] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items," *Expert Syst. Appl.*, vol. 69, pp. 29–39, 2017.
- [21] R. Barzegar Nozari and H. Koochi, "A novel group recommender system based on members' influence and leader impact," *Knowl. Base. Syst.*, vol. 205, p. 106296, 2020.
- [22] F. Deng, "Utility-based Recommender Systems Using Implicit Utility and Genetic Algorithm," presented at the 2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering, Shenyang, China, 2015. <https://doi.org/10.2991/meic-15.2015.197>.
- [23] Z. Gantner, S. Rendle, and L. Schmidt-Thieme, "Factorization models for context-/time-aware movie recommendations," in Proceedings of the Workshop on Context-Aware Movie Recommendation – CAMRa '10, Barcelona, Spain, 2010, pp. 14–19. <https://doi.org/10.1145/1869652.1869654>.
- [24] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," p. 10.
- [25] M. de Gemmis, P. Lops, G. Semeraro, and C. Musto, "An investigation on the serendipity problem in recommender systems," *Inf. Process. Manage.* vol. 51, no. 5, pp. 695–717, September 2015.
- [26] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: social recommendation using probabilistic matrix factorization," in Proceeding of the 17th ACM conference on Information and knowledge mining – CIKM '08, Napa Valley, California, USA, 2008, p. 931. <https://doi.org/10.1145/1458082.1458205>.

- [27] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2065–2073, 2014.
- [28] T. Di Noia, V. C. Ostuni, J. Rosati, P. Tomeo, and E. Di Sciascio, "An analysis of users' propensity toward diversity in recommendations," in *Proceedings of the 8th ACM Conference on Recommender systems – RecSys '14*, Foster City, Silicon Valley, California, USA, 2014, pp. 285–288. <https://doi.org/10.1145/2645710.2645774>.
- [29] N. Nassar, A. Jafar, and Y. Rahhal, "A novel deep multi-criteria collaborative filtering model for recommendation system," *Knowl. Base. Syst.*, vol. 187, p. 104811, 2020.
- [30] Y. Shi, M. Larson, and A. Hanjalic, "Mining contextual movie similarity with matrix factorization for context-aware recommendation," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 1, pp. 1–19, 2013.
- [31] Y. Koren and R. Bell, "Advances in Collaborative Filtering," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds. Boston, MA: Springer US, 2015, pp. 77–118.
- [32] T. Silveira, M. Zhang, X. Lin, Y. Liu, and S. Ma, "How good your recommender system is? A survey on evaluations in recommendation," *Int. J. Mach. Learn. Cybern.* vol. 10, no. 5, pp. 813–831, 2019.
- [33] M. Srifi, A. Oussous, A. Ait Lahcen, and S. Mouline, "Recommender Systems Based on Collaborative Filtering Using Review Texts—A Survey," *Information (Basel)*, vol. 11, no. 6, p. 317, 2020.
- [34] R. Burke, *User Model. User-adapt. Interact.* vol. 12, no. 4, pp. 331–370, 2002.
- [35] M. K. Najafabadi, A. Mohamed, and C. W. Onn, "An impact of time and item influencer in collaborative filtering recommendations using graph-based model," *Inf. Process. Manage.* vol. 56, no. 3, pp. 526–540, 2019.
- [36] V. Leksin and A. Ostapets, "Job recommendation based on factorization machine and topic modelling," in *Proceedings of the Recommender Systems Challenge on - RecSys Challenge '16*, Boston, Massachusetts, 2016, pp. 1–4. <https://doi.org/10.1145/2987538.2987542>.
- [37] P. Kouki, J. Schaffer, J. Pujara, J. O'Donovan, and L. Getoor, "Personalized explanations for hybrid recommender systems," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Marina del Ray California, 2019, pp. 379–390.
- [38] Z. Yang, P. Baraldi, and E. Zio, "Automatic Extraction of a Health Indicator from Vibrational Data by Sparse Autoencoders," in *2018 3rd International Conference on System Reliability and Safety (ICSRS)*,

- Barcelona, Spain, Nov. 2018, pp. 328–332. <https://doi.org/10.1109/ICSR.S.2018.8688720>.
- [39] H. Taud and J. F. Mas, “Multilayer Perceptron (MLP),” in *Geomatic Approaches for Modeling Land Change Scenarios*, M. T. Camacho Olmedo, M. Paegelow, J.-F. Mas, and F. Escobar, Eds. Cham: Springer International Publishing, 2018, pp. 451–455.
- [40] A. Fischer and C. Igel, “An Introduction to Restricted Boltzmann Machines,” in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 7441. L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 14–36.
- [41] J. V. Schaaf, M. Jepma, I. Visser, and H. M. Huizenga, “A hierarchical Bayesian approach to assess learning and guessing strategies in reinforcement learning,” *J. Math. Psychol.*, vol. 93, p. 102276, 2019.
- [42] Z. Lin and H. Chen, “A Probabilistic Model for Collaborative Filtering,” in *Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics – WIMS2019*, Seoul, Republic of Korea, 2019, pp. 1–8. <https://doi.org/10.1145/3326467.3326472>.
- [43] Y. Zheng, M. Agnani, and M. Singh, “Identifying Grey Sheep Users By The Distribution of User Similarities In Collaborative Filtering,” in *Proceedings of the 6th Annual Conference on Research in Information Technology – RIIT ’17*, Rochester, New York, USA, 2017, pp. 1–6. <http://doi.org/10.1145/3125649.3125651>.
- [44] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, et al., “Recommender systems based on social networks,” *J. Syst. Softw.*, vol. 99, pp. 109–119, 2015.
- [45] A. B. Bondi, “Characteristics of scalability and their impact on performance,” in *Proceedings of the second international workshop on Software and performance - WOSP ’00*, Ottawa, Ontario, Canada, 2000, pp. 195–203. <https://doi.org/10.1145/350391.350432>.
- [46] T. Ebesu and Y. Fang, “Neural Semantic Personalized Ranking for item cold-start recommendation,” *Inf. Retrieval*, vol. 20, no. 2, pp. 109–131, 2017.
- [47] S. Khatwani and M. B. Chandak, “Building Personalized and Non Personalized recommendation systems,” in *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, Pune, India, 2016, pp. 623–628.

- [48] A. K. Sahu and P. Dwivedi, “Knowledge transfer by domain-independent user latent factor for cross-domain recommender systems,” *Future Gener. Comput. Syst.*, vol. 108, pp. 320–333, 2020.
- [49] L. Chen, G. Chen, and F. Wang, “Recommender systems based on user reviews: The state of the art,” *User Model. User-adapt. Interact.* vol. 25, no. 2, pp. 99–154, 2015.
- [50] Y. Gong and Q. Zhang, “Hashtag Recommendation Using Attention-Based Convolutional,” *Neural Netw.* vol., p. 7, 2016.
- [51] J. McAuley and J. Leskovec, “Hidden factors and hidden topics: understanding rating dimensions with review text,” in *Proceedings of the 7th ACM conference on Recommender systems*, Hong Kong China, 2013, pp. 165–172.
- [52] J. Y. Chin, K. Zhao, S. Joty, and G. Cong, “ANR: Aspect-based Neural Recommender,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino Italy, Oct. 2018, pp. 147–156.
- [53] Z. Cheng, Y. Ding, L. Zhu, and M. Kankanhalli, “Aspect-Aware Latent Factor Model: Rating Prediction with Ratings and Reviews,” in *Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW ’18*, Lyon, France, 2018, pp. 639–648. <https://doi.org/10.1145/3178876.3186145>.
- [54] R. Zhang, H. Bao, H. Sun, Y. Wang, and X. Liu, “Recommender systems based on ranking performance optimization,” *Front. Comput. Sci.*, vol. 10, no. 2, pp. 270–280, 2016.
- [55] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, 1st ed., Cambridge University Press, 1988, <https://doi.org/10.1017/CBO9780511571299>.
- [56] M. Paolanti and E. Frontoni, “Multidisciplinary Pattern Recognition applications: A review,” *Comput. Sci. Rev.*, vol. 37, p. 100276, 2020.
- [57] S. Wang, C. Aggarwal, J. Tang, and H. Liu, “Attributed Signed Network Embedding,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore, 2017, pp. 137–146.
- [58] M. Terzi, M. Rowe, M.-A. Ferrario, and J. Whittle, “Text-Based User-kNN: Measuring User Similarity Based on Text Reviews,” in *User Modeling, Adaptation, and Personalization*, vol. 8538, V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog, and G.-J. Houben, Eds. Cham: Springer International Publishing, 2014, pp. 195–206. https://doi.org/10.1007/978-3-319-08786-3_17.

- [59] D. Kim, C. Park, J. Oh, S. Lee, and H. Yu, “Convolutional Matrix Factorization for Document Context-Aware Recommendation,” in Proceedings of the 10th ACM Conference on Recommender Systems, Boston Massachusetts USA, 2016, pp. 233–240.
- [60] C. Chen, M. Zhang, Y. Liu, and S. Ma, “Neural Attentional Rating Regression with Review-level Explanations,” in Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW ’18, Lyon, France, 2018, pp. 1583–1592. <https://doi.org/10.1145/3178876.3186070>.
- [61] I. Ben Sassi, S. Mellouli, and S. Ben Yahia, “Context-aware recommender systems in mobile environment: On the road of future research,” *Inf. Syst.*, vol. 72, pp. 27–61, 2017.
- [62] S. Kulkarni and S. F. Rodd, “Context Aware Recommendation Systems: A review of the state of the art techniques,” *Comput. Sci. Rev.*, vol. 37, p. 100255, 2020.
- [63] E. Negre, “Systèmes de recommandation contextuels?: Vers une typologie de contexte,” *ROISI*, vol. 1, no. 4, 2020. <https://doi.org/10.21494/I STE.OP.2020.0584>.
- [64] B. Wang, S. Xiong, Y. Huang, and X. Li, “Review Rating Prediction Based on User Context and Product Context,” *Appl. Sci. (Basel)*, vol. 8, no. 10, p. 1849, 2018.
- [65] Y. Moshfeghi, B. Piwowarski, and J. M. Jose, “Handling data sparsity in collaborative filtering using emotion and semantic based features,” in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information – SIGIR ’11, Beijing, China, 2011, p. 625. <https://doi.org/10.1145/2009916.2010001>.
- [66] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Stat.*, vol. 29, no. 5, 2001. <https://doi.org/10.1214/aos/1013203451>.
- [67] X. Meng, S. Wang, H. Liu, and Y. Zhang, “Exploiting Emotion on Reviews for Recommender Systems,” p. 8.
- [68] A. Mogenet, T. A. N. Pham, M. Kazama, and J. Kong, “Predicting online performance of job recommender systems with offline evaluation,” in Proceedings of the 13th ACM Conference on Recommender Systems, Copenhagen Denmark, Sep. 2019, pp. 477–480.
- [69] C. Wang and D. M. Blei, “Collaborative topic modeling for recommending scientific articles,” in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining – KDD ’11, San Diego, California, USA, 2011, p. 448. <https://doi.org/10.1145/2020408.2020480>.

- [70] R.-P. Shen, H.-R. Zhang, H. Yu, and F. Min, "Sentiment based matrix factorization with reliability for recommendation," *Expert Syst. Appl.*, vol. 135, pp. 249–258, 2019.
- [71] K. Li, X. Zhou, F. Lin, W. Zeng, and G. Alterovitz, "Deep Probabilistic Matrix Factorization Framework for Online Collaborative Filtering," *IEEE Access*, vol. 7, pp. 56117–56128, 2019.

Biographies



Khalid Benabbes is a PhD student at the MISC Laboratory, Faculty of Sciences, Ibn Tofail University, Kénitra, Morocco. He is currently a software engineer at Hassan II Institute of Agronomy & Veterinary Medicine in Rabat. He holds an engineering degree in Computer Sciences from ENSA, Agadir. His research interests include MOOC, Recommender system, Machine Learning, and Data Science.



Khalid Housni received his PhD in computer sciences from the University of Ibn Zohr Agadir, Morocco. In 2014, he joined the Department of Computer Science, Faculty of Sciences, Ibn Tofail University, Kénitra, Morocco. His research interests include image/video processing and networks reliability.



Ali El Mezouary received his PhD in computer sciences from the University of Ibn Zohr, Agadir, Morocco. He is a full professor at ESTA, University of Ibn Zohr. His research interests include Web Semantic and Adaptive Learning Systems.



Ahmed Zellou received his PhD in Computer Sciences from Mohammedia School of Engineering, Rabat, Morocco. He is currently a professor at the National School of Computer Science and Systems Analysis (ENSIAS). His research interests include Information Integration, Hybrid, Semantic and Fuzzy, Content Integration, Semantic P2P, Cloud Integration.