A Review on User Recommendation System Based Upon Semantic Analysis

Lovedeep Kaur, Naveen Kumari
Punjabi University Regional Centre for Information Technology and Management, Mohali (Sector 61), Punjab, India

Abstract— Recommender system applied various techniques and prediction algorithm to predict user interest on information, items and services from the tremendous amount of available data on the internet. Recommender systems are now becoming increasingly important to individual users, businesses and specially e-commerce for providing personalized recommendations. Recommender systems have been evaluated and improved in many, often incomparable, ways. In this paper, we review the evaluation and improvement techniques for improving overall performance of recommendation systems and proposing a semantic analysis based approach for clustering based collaborative filtering to improve the coverage of recommendation. The basic algorithm or predictive model we use are – simple linear regression, k-nearest neighbours (kNN), naive bayes, support vector machine. We also review the pearson correlation coefficient algorithm and an associative analysis-based heuristic. The algorithms themselves were implemented from abstract class recommender, which was extended from weka distribution classifier class. The abstract class adds prediction method to the classifier.

Keywords— Recommendation, KNN, Content based Filtering, Data Mining

I. DATA MINING

Information overloaded on the searching and recommendation systems has created huge challenges for users to get more accurate searched and recommended results at the same time under same one system. Several recommender systems employing different algorithms, approaches and methods are used to address some challenges. This research paper emphasizes on graphical model that offers a standard data demonstration and support different recommendation methods and algorithms. Furthermore models and algorithms needed for Bayesian Recommender Systems are discussed that describes how Bayesian methods can be applied to recommendation systems to make optimal recommendations. A challenging task for recommendation and searching system is to improve the accuracy for the new items for new users. In recommendation system usually data is analyzed about particular item and interactions between users and items are found as a result. This paper focuses on algorithms that are used by researchers and engineers to improve the behavior of recommendation systems.

II. RECOMMENDATION SYSTEM

Recommender systems, recommendation engines, recommendation frameworks, recommendation platforms or simply recommender form or work from a specific type of information filtering system technique that attempts to recommend information items (movies, TV program/show/episode, video on demand, music, books, news, images, web pages, scientific literature etc.) or social elements (e.g. people, events or groups) that are likely to be of interest to the user. Typically, a recommender system compares a user profile to some reference characteristics, and seeks to predict the 'rating' or 'preference' that a user would give to an item they had not yet considered. These characteristics may be from the information item (the content-based approach) or the user's social environment (the collaborative filtering).[6] Researchers at Xerox PARC developed Tapestry, the first recommendation support system.[7] Tapestry was an electronic messaging system that allowed users to either rate messages ("good" or "bad"). Although Tapestry provided good recommendations, it had one major drawback; the user was required to write complicated queries [4]. The first system to generate automated recommendations was the Group Lens system. The Group Lens system provided users with personalized recommendation on Usenet postings. It recommended articles found interesting by users similar to the target user. The technique we use for recommendation is data mining. Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. Data mining is the analysis of data and the use of software techniques for finding patterns and regularities in sets of data. Data mining provides a number of algorithms to obtain profiles of users based on historical data, which are used to predict the...
preferences of new users. The process of applying data mining techniques on web data in order to obtain customer usage patterns is known as web mining. The process of data mining typically consists of 3 steps, carried out in succession: Data Preprocessing, Data Analysis, and Result Interpretation. The two basic entities which appear in any Recommender System are the user (sometimes also referred to as customer) and the item (also referred to as product). The input to a Recommender System depends on the type of the employed filtering algorithm. Generally, the input belongs to one of the following categories: Ratings (also called votes), which express the opinion of users on items; Demographic data, which refer to information such as the age, the gender and the education of the users; Content data, which are based on a textual analysis of documents related to the items rated by the user. The goal of Recommender Systems is to generate suggestions about new items or to predict the utility of a specific item for a particular user. The output of a Recommender System can be either a Prediction or a Recommendation.

- A Prediction is expressed as a numerical value, \( r_{a,j} \), which represents the anticipated opinion of active user \( u_a \) for item \( i_j \). This predicted value should necessarily be within the same numerical scale (example: 1-bad to 5-excellent) as the input referring to the opinions provided initially by active user \( u_a \).
- A Recommendation is expressed as a list of \( N \) items, where \( N \leq n \), which the active user is expected to like the most. The usual approach in that case requires this list to include only items that the active user has not already purchased, viewed or rated.

Various approaches of Collaborative Filtering are:

1) User-based approach: This approach was proposed in the end of 1990s by the professor of University of Minnesota Jonathan L. Herlocker. In the user-based approach, the users perform the main role. If certain majority of the customers has the same taste then they join into one group. Recommendations are given to user based on evaluation of items by other users form the same group, with whom he/she shares common preferences. If the item was positively rated by the community, it will be recommended to the user.

2) Item-based approach: This approach was proposed by the researchers of University of Minnesota in 2001. Referring to the fact that the taste of users remains constant or change very slightly similar items build neighborhoods based on appreciations of users. Afterwards the system generates recommendations with items in the neighborhood that a user would prefer.

3) Hybrid recommendation approaches: For better results some recommender systems combine different techniques of collaborative approaches and content-based approaches. The combination of approaches can proceed in different ways.

Collaborative-Filtering: A significant role is play by a Collaborative Filtering (CF) methods in the recommendation process and because of that Collaborative filtering is most extensively used approach to design recommender system [1, 2]. In this approach recommendation for each active user is received by comparing with the preferences of other users who have rated the product in similar way to the active user.

Content-Based filtering: In content-based filtering recommendations depends on users former choices. Item description and a profile of the user’s orientation play an important role in Content-based filtering. Content-based filtering algorithms try to recommend items based on similarity count.

Demographic Filtering: In demographic filtering recommendations is established on a demographic profile of the user. Here recommendation is based on the information provided by the user is considered to be similar according to demographic parameter such as nationality, age, gender etc.

Hybrid filtering: The hybrid filtering is a combination of more than one filtering approach. The hybrid filtering approach is introduced to overcome some common problem that are associated with above filtering approaches such as cold start problem, overspecialization problem and sparsity problem. Another motive behind the implementation of hybrid filtering is to improve the accuracy and efficiency of recommendation process.

III. SEMANTIC ANALYSIS BASED RECOMMENDATION SYSTEM

(SERSs) are a relatively recent approach for overcoming the limitations due to lack of transparency, the cold-start problem and data sparsity through the incorporation of domain semantics. This approach assumes that these limitations are strongly related with the lack of understanding (and exploitation) of domain semantics of traditional recommenders, and relies on ontology-based representations of knowledge to take advantage of this domain semantics. Taking into account that each recommendation approach is useful to overcome part of the limitations mentioned above, it is plausible to foresee that combining these two approaches (i.e. incorporating both semantics and contextual information) will provide better results. To date, little work has been carried out to study this kind of hybrid systems and additional research is needed to understand under which conditions they can outperform current recommenders. The objective of
this work is to provide the basis for this research by proposing different semantically-enhanced, context-aware recommendation algorithms that incorporate semantics and contextual information into well-known recommendation strategies such as CB and CF.

IV. HYBRID RECOMMENDATION SYSTEM

Recent research has proved that a hybrid approach could be more effective in some cases. Basically Collaborative filtering and Content-based filtering approaches most extensively used in information filtering application. As we know that every coin has two side similarly each approach has its own reward and weaknesses. Basically the main motive of hybrid approach is to aggregate collaborative filtering and content-based filtering to improve recommendation accuracy. Hybrid approaches can be implemented in various ways:

1. Implement collaborative and content-based methods individually and aggregate their predictions.
2. Integrate some content-based characteristics into a collaborative approach,
3. Comprise some collaborative characteristics into a content-based approach, and
4. Construct a general consolidative model that integrate both content-based and collaborative characteristics.

Cold start and the sparsity are common problems in recommender systems which are resolved by using these methods. Good example of hybrid recommender systems is Netflix. They make recommendations by comparing the looking out and exploring habits of similar exploiters (collaborative filtering) as well as by providing movies that share features with films that an exploiters has rated highly (content-based filtering). The online DVD rental company Netflix released a data set containing approximately 100 million anonymous movie ratings in October 2006 and challenged investigators and practitioners to beat the accuracy of the company’s recommendation system, Cinematch [23]. Although the released data set represented only a small fraction of the company’s rating data, thanks to its size and quality it fast became a standard in the data mining and machine learning community. The data set contained ratings in the integer scale from 1 to 5 which were accompanied by dates. The year of release were provided for every movie and title. No information about users was given. Submitted predictions were evaluated by their root mean squared error (RMSE) on a qualifying data set containing over 2817,131 unknown ratings. Total 20,000 are registered teams out of that 2000 teams submitted at least one answer set. The grand prize of $1000,000 was awarded to a team on 21 September 2009 that performed better over the Cinematch’s and also increases accuracy by 10%. In this competition we learned several lessons [24]. Firstly, the company acquired a superior recommendation system that improve users satisfaction and also company gained lot of publicity. Secondly, ensemble methods play an important role for improving the accuracy of predictions. Thirdly, we discovered that when RMSE drops below a certain level that time accuracy improvements are increasingly demanding. Finally, despite the company’s effort, namelessness of its users was not sufficiently assured [25].

Figure 2 shows the methods that estimate CBF and CF recommendations individually and subsequently combine them to yield better recommendations across the board.

![Figure 2: Collaborative and Content Based Filtering based recommendation](image)

CBF and CF can be aggregated in different ways [1]. Following figures shows the different choices for aggregating CB and CBF. Figure 2 shows the methods that estimate CBF and CF recommendations individually and subsequently combine them to yield better recommendations across the board.

![Figure 2: Collaborative and Content Based Filtering based recommendation](image)

Figure 3 shows the methods that integrate CBF characteristics into the CF approach. So that it will overcome the cold start problem in collaborative filtering and overspecialization problem of content-based filtering.

![Figure 3: integration of CBF characteristics into the CF approach.](image)

Figure 4 illustrates the methods for construction of a unified utility system with both CBF and CF characteristics. In this method by combining some features of CBF and CF one unified model is constructed that can improve effectiveness of recommendation process.
Figure 4: construction of a unified utility system with both CBF and CF characteristics

Figure 5 shows the methods that incorporate CF characteristics into a CBF approach.

Content-based filtering systems can allow recommendations for "cold-start" items for which no training information is available, but it suffers by lower accuracy than collaborative filtering systems. Conversely, collaborative filtering approach frequently provide accurate recommendations, but go wrong for cold start items. Hybrid schemes try to aggregate these different kinds of information to get efficient recommendation result.

Following is taxonomy for the hybrid recommendation systems he classifies Hybrid Recommendation Systems into following seven classes:

**Weighted:**
- Different recommendation components scores are combined statistically. This class aggregates scores from each factor using additive formula.

**Switching:**
- From available recommendation components system chooses particular component and applies the picked out one.

**Mixed:**
- Different recommender provides their recommendation that will be introduced together. This class is based on merging and presentation of multiple rated list into single rated list.

**Feature Combination:**
- Contributing and actual recommender are two different recommendation components are exist for this class. The working of actual recommender is depends on the data modified by the contributing one. The contributing one throws features of one source on to the other components source.

**Feature Augmentation:**
- This class is similar to the feature combination hybrids but only difference is that the contributor gives novel characteristic. It is more elastic than feature combination method.

**Cascade:**
- This class play an role of tie breaker. Here for every recommender assign some priority and according to that assign priority, lower priority recommenders play an tie breakers role over higher priority.

**Meta-level:**
- Their exist contributing and actual recommenders but the early one completely substitutes the data for the latter one.

V. EXTENDING CAPABILITIES OF RECOMMENDER SYSTEMS

Recommender system can be extended in several ways that include the improving the understanding of users and items, incorporating the contextual information into the recommendation process, supporting multi-criteria ratings, and providing more flexible and less intrusive types of recommendations. This section describe about the proposed extensions and also identify various research opportunities for developing them.

<table>
<thead>
<tr>
<th>Recommendation Approach</th>
<th>Recommendation Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic-based</td>
<td>Model-based</td>
</tr>
<tr>
<td>Content-based</td>
<td>Commonly Used:</td>
</tr>
<tr>
<td></td>
<td>• TF-IDF</td>
</tr>
<tr>
<td></td>
<td>• Clustering</td>
</tr>
<tr>
<td></td>
<td>Representative research examples:</td>
</tr>
<tr>
<td></td>
<td>• Lang 1995</td>
</tr>
<tr>
<td></td>
<td>• Balabanovic &amp; Shoham 1997</td>
</tr>
<tr>
<td></td>
<td>• Pazzani &amp; Billsus 1997</td>
</tr>
<tr>
<td></td>
<td>Commonly used techniques:</td>
</tr>
<tr>
<td></td>
<td>• Bayesian classifiers</td>
</tr>
<tr>
<td></td>
<td>• Clustering</td>
</tr>
<tr>
<td></td>
<td>• Decision trees</td>
</tr>
<tr>
<td></td>
<td>• Artificial neural networks</td>
</tr>
<tr>
<td></td>
<td>Representative research examples:</td>
</tr>
<tr>
<td></td>
<td>• Pazzani &amp; Billsus 1997</td>
</tr>
</tbody>
</table>
**Table: Improvements for Recommendation System**

<table>
<thead>
<tr>
<th>Collaborative</th>
<th>Commonly used techniques:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nearest neighbor (cosine, correlation)</td>
</tr>
<tr>
<td></td>
<td>Clustering</td>
</tr>
<tr>
<td></td>
<td>Graph theory</td>
</tr>
<tr>
<td>Representative research examples:</td>
<td></td>
</tr>
<tr>
<td>Resnick et al. 1994</td>
<td></td>
</tr>
<tr>
<td>Hill et al. 1995</td>
<td></td>
</tr>
<tr>
<td>Shardanand &amp; Maes 1995</td>
<td></td>
</tr>
<tr>
<td>Breese et al. 1998</td>
<td></td>
</tr>
<tr>
<td>Nakamura &amp; Abe 1998</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hybrid</th>
<th>Combining content-based and collaborative:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear combination of predicted ratings</td>
</tr>
<tr>
<td></td>
<td>Various voting schemes Incorporating one component as a part of the heuristic for the other</td>
</tr>
<tr>
<td>Representative research examples:</td>
<td></td>
</tr>
<tr>
<td>Balabanovic &amp; Shoham 1997</td>
<td></td>
</tr>
<tr>
<td>Claypool et al. 1999</td>
<td></td>
</tr>
<tr>
<td>Good et al. 1999</td>
<td></td>
</tr>
<tr>
<td>Melville et al. 2002</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Commonly used techniques:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bayesian networks</td>
</tr>
<tr>
<td></td>
<td>Clustering</td>
</tr>
<tr>
<td></td>
<td>Artificial neural networks</td>
</tr>
<tr>
<td></td>
<td>Linear regression</td>
</tr>
<tr>
<td></td>
<td>Probabilistic models</td>
</tr>
<tr>
<td>Representative research examples:</td>
<td></td>
</tr>
<tr>
<td>Billsus &amp; Pazzani 1998</td>
<td></td>
</tr>
<tr>
<td>Breese et al. 1998</td>
<td></td>
</tr>
<tr>
<td>Ungar &amp; Foster 1998</td>
<td></td>
</tr>
</tbody>
</table>

**VI. IMPROVEMENTS FOR RECOMMENDATION SYSTEM**

In [7], they conclude that, the first RS were focused on improving recommendation accuracy through filtering. Most memory-based methods and algorithms were developed and optimized in this context (e.g., kNN metrics, aggregation approaches, singular value decomposition, diffusion-based methods, etc.). At this stage, hybrid approaches (primarily collaborative–demographic and collaborative content filtering) improved the quality of the recommendations. In the second stage, algorithms that included social information with previous hybrid approaches were adapted and developed (e.g., trust-aware algorithms, social adaptive approaches, social networks analysis, etc.). Currently, the hybrid ensemble algorithms incorporate location information into existing recommendation algorithms. Evaluation of the predictions and recommendations has evolved since the origins of RS, which weighted prediction errors (accuracy) heavily. They also recognized the convenience of evaluating the quality of the top n recommendations as a set; evaluation of the top n recommendations as a ranked list was then incorporated. Currently, there is a tendency to assess new evaluation measures, such as diversity and novelty. The authors in [8], describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make RS applicable to an even broader range of applications. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multicriteria ratings, and a provision of more flexible and less intrusive types of recommendations. In [9], Social tagging systems pose new challenges to developers of RS. As observed by recent research, traditional implementations of classic recommender approaches, such as collaborative filtering, are not working well in this new context. To address these challenges, a number of research groups worldwide work on adapting these approaches to the specific nature of social tagging systems. In joining this stream of research, they have developed and evaluated two enhancements of user-based collaborative filtering algorithms to provide recommendations of articles on CiteULike, a social tagging service for scientific articles. In [10], they showed that a CF framework can be used to combine personal Information Filtering (IF) agents and the opinions of a community of users to produce better recommendations than either agents or users can produce alone. It also shows that using CF to create a personal Combination of a set of agents produces better results than either individual agents or other combination mechanisms. One key implication of these results is that users can avoid having to select among agents; they can use them all and let the CF framework select the best ones for them. Following is the metrics wise literature survey to improve the RS:

**A. Accuracy**

In [11], this paper presents a metric to measure similarity between users, which is applicable in collaborative filtering processes carried out in RS. The proposed metric is formulated via a simple linear combination of values and
weights. Values are calculated for each pair of users between which the similarity is obtained, whilst weights are only calculated once, making use of a prior stage in which a genetic algorithm extracts weightings from the RS which depend on the specific nature of the data from each RS. The results obtained present significant improvements in prediction quality, recommendation quality and performance. Thus the improvements can be seen in the system’s accuracy. In [12], they examine an advanced collaborative filtering method that uses similarity transitivity concepts. By propagating “similarity” between users, in a similar way as with “trust”, we can significantly expand the space of potential recommenders and also improves the recommendation’s accuracy. In [13], they propose several new approaches to improve the accuracy of recommendations by using rating variance (which, as we show, is inversely related to the recommendation accuracy) to gauge the confidence of recommendations. They then empirically show how these approaches work with different recommendation techniques. We also show how these approaches can generate more personalized recommendations, as measured by the coverage metric (described later in the paper in more detail). As a result, users can be given a better control to choose whether to receive recommendations with higher coverage or accuracy. In [14], despite its success, similarity-based collaborative filtering suffers from some significant limitations, such as scalability and sparsity. This paper introduces trust to the domain of collaborative filtering to overcome these limitations. Compared with the similarity-based CF, introduction of trust does improve the performance of CF in terms of coverage, prediction accuracy, and robustness in the presence of attacks. Experimental results based on a real dataset are illustrated as evidences to support their claim. In [15], they have presented two contributions to the RS field, both of them based on a semantic approach. The common goal of their work has been to improve collaborative filtering recommendations in e-commerce, in terms of accuracy and reliability. To this aim, their strategies rely on an ontology that formalizes the semantic descriptions of commercial products.

\[
\text{ACCURACY} = \frac{(TP+TN)}{(TP+TN+FP+FN)}
\]

TP: True Positive
TN: True Negative
FP: False Positive
FN: False Negative

B. Coverage

In [11], they proposed a use of genetic algorithms that can be applied for obtaining optimal similarity functions. Those similarity functions obtained provide better quality and faster results than the ones provided by the traditional metrics. Improvements can be seen in the system’s accuracy (MAE), in the coverage and in the precision and recall recommendation quality measures. The proposed use of GAs applied to the RS is a novel approach and has the main advantage that it can be used in all CF-based RS, without the need to use hybrid models which often cannot be applied, as in many cases no reliable demographic information or content-based filtering information is available. In [12], they proposed a novel similarity propagation scheme to confront the data sparsity problem in RS and evaluate their method over two datasets with different characteristics, exhibiting a much higher recommendation coverage and better accuracy than classical collaborative filtering methods even under very sparse data conditions. In [13], using a simple filtering approach they have demonstrated that prediction accuracy can be significantly improved by filtering out recommendations above a minimum rating standard deviation threshold. However, there was also a corresponding decrease in the coverage of recommendations. They then proposed the smart and safe approaches which generate recommendations of greater value by providing a good balance of prediction accuracy and coverage. New approaches are especially useful, since they can confidently improve the accuracy of recommendation, and in addition to that, a user can control the balance between the accuracy and coverage of recommendations. In [14], it has been shown by the experimental results that the trust metrics and corresponding prediction making approach do improve the performance of traditional similarity-based CF in terms of coverage, prediction accuracy and robustness.

C. Diversity

There is increasing awareness in the RS field that diversity is a key property that enhances the usefulness of recommendations. In [16], they argued that as new types of recommendation domains and tasks emerge, this blind faith in the similarity assumption begins to seem flawed. They showed that very often recommendation diversity is important and that traditional recommendation systems are marred by poor diversity characteristics. They evaluate a new class of diversity-preserving algorithm capable of addressing this without compromising similarity or efficiency. In [17], they introduce and explore a number of item ranking techniques that can generate substantially more diverse recommendations across all users while maintaining comparable levels of recommendation accuracy. Comprehensive empirical evaluation consistently shows the diversity gains of the proposed techniques using several real-world rating
datasets and different rating prediction algorithms. In [18], Genre information can serve as a means to measure and enhance the diversity of recommendations and is readily available in domains such as movies, music or books. In this work they propose a new Binomial framework for defining genre diversity in RS that takes into account three key properties: genre coverage, genre redundancy and recommendation list size-awareness. They show that methods previously proposed for measuring and enhancing recommendation diversification fail to address adequately these three properties. They also propose an efficient greedy optimization technique to optimize Binomial diversity. In [19], they present topic diversification, an algorithmic framework to increase the diversity of a top-N list of recommended products. In order to show its efficiency in diversifying, they also introduced their new intra-list similarity metric. Contrasting precision and recall metrics, computed both for user-based and item-based CF and featuring different levels of diversification, with results obtained from a large scale user survey, they showed that the user’s overall liking of recommendation lists goes beyond accuracy and involves other factors, e.g., the users’ perceived list diversity. They were thus able to provide empirical evidence that lists are more than mere aggregations of single recommendations, but bear an intrinsic, added value. In [20], in most cases, new techniques are designed to improve the accuracy of recommendations, whereas the recommendation diversity has often been overlooked. In particular, they showed that, while ranking recommendations according to the predicted rating values provides good predictive accuracy, it tends to perform poorly with respect to recommendation diversity. Therefore, in this paper, they proposed a number of recommendation ranking techniques that can provide significant improvements in recommendation diversity with only a small amount of accuracy loss. In addition, these ranking techniques offer flexibility. In [21], they propose a neighbor diversification collaborative filtering algorithm to improve the recommendation lists. By using Movielens dataset for empirical analysis, they investigated the influence of neighbor diversity to the recommendation accuracy, diversity, novelty and coverage. Intensive experimental results proved the efficiency of their proposed algorithm for improving recommendation lists.

D. Quality

In [22], Information Filtering and Collaborative Filtering techniques have been used for selecting information based on the user’s previous preference tendency and the opinion of other people who have similar tastes with the user. Combining both Information Filtering and Collaborative Filtering, or hybrid systems, they have also been proposed to get better recommendation results. In this paper, they present an improved recommendation method that copes with the sparsity problem of the hybrid systems and increases the quality of recommendation results. In [23], this paper involves a prefiltering process that eliminates the least representative users from the k-neighbour selection process and retains the most promising ones. The improvements obtained are always positive with respect to both the prediction quality measures and the recommendation quality measures. This demonstrates that certain users should not be included among the active user’s neighbours and that the traditional similarity measures are not capable of detecting them. The favourable results obtained here can be considered generally applicable due to the broad margin of improvement observed and their testing on the two most representative databases of collaborative filtering RS. In [11], they have presented a genetic algorithm method for obtaining optimal similarity functions. The similarity functions obtained provide better quality and quicker results than the ones provided by the traditional metrics.

E. Scalability:

In [24], they presented and experimentally evaluated a new approach in improving the scalability of RS by using clustering techniques. Their experiments suggest that clustering based neighborhood provides comparable prediction quality as the basic CF approach and at the same time significantly improves the online performance.

F. User Preferences

In [25], many RS employed in commercial web sites use collaborative filtering. The main goal of traditional collaborative filtering techniques is improvement of the accuracy of recommendation. Nevertheless, such techniques present the problem that they include many items that the user already knows. These recommendations appear to be good when they consider accuracy alone. On the other hand, when they consider users’ satisfaction, they are not necessarily good because of the lack of discovery. In their work, they infer items that a user does not know by calculating the similarity of users or items based on information about what items users already know. They seek to recommend items that the user would probably like and does not know by combining the above method and the most popular method of collaborative filtering. In [26], they have presented an approach to improve traditional RS which are special types of expert systems able to select automatically the most relevant information for each individual by exploiting the knowledge of an expert in a particular domain and the users’ preferences. Specifically, their strategy prevents from selecting fake neighborhoods in collaborative RS. This problem appears in domains such as e-commerce where there are a wide range
of products, and the different categories contain items of very different nature (such as books, music, clothes or food). In these contexts, taking into account all the preferences registered in user’s profiles when estimating their similarity of interests can lead to the selection of fake neighbors, that is, neighbors that have dissimilar interests with the target user with respect to the target product, but with similar preferences to him/her regarding lots of items of other categories.

G. Reliability Collaborative

RS select potentially interesting items for each user based on the preferences of like-minded individuals. Particularly, e-commerce has become a major domain in this research field due to its business interest, since identifying the products the users may like or find useful can boost consumption. During the last years, a great number of works in the literature have focused in the improvement of these tools. Expertise, trust and reputation models are incorporated in collaborative RS to increase their accuracy and reliability. However, current approaches require extra data from the users that is not often available. In [15], they present two contributions that apply a semantic approach to improve recommendation results transparently to the users. On the one hand, they automatically build implicit trust networks in order to incorporate trust and reputation in the selection of the set of like-minded users that will drive the recommendation. The common goal of their work has been to improve collaborative filtering recommendations in e-commerce, in terms of accuracy and reliability. To this aim, their strategies rely on an ontology that formalizes the semantic descriptions of commercial products. The exploitation of semantics enables reasoning about the data stored in the users’ personal profiles and inferring new knowledge.

VII. CONCLUSION

Evaluation and Improvement techniques used for various metrics of Recommendation Systems has discussed in detail. There are two methods for evaluating RS, first is the system oriented evaluation that is also known as offline evaluation and second is the user oriented evaluation that is also known as the online evaluation. RS can be improved with the help of improving various metrics of RS such as accuracy, coverage, diversity, quality, scalability, user preferences, reliability, etc. We reviewed various algorithms such as nearest neighbor, support vector machine, naïve bayes. We also presented the Pearson correlation coefficient algorithm and an associative analysis-based heuristic. We also introduce various modern recommendation approaches such as context-aware approaches, Semantic-based approaches, cross-domain based approaches, peer-to-peer approaches and cross-lingual approaches. We have also uncovered areas that are open to many further improvements, and where there is still much exciting and relevant research to be done in coming years.

REFERENCES

[14] Jianshu Weng, Chunyan Miao, Angela Goh, “Improving Collaborative Filtering with Trust Based Metrics”
[26] Manuela I. Martin-Vicente, Alberto Gil-Solla, Manuel Ramos-Cabrér, Jose J. Pazos-Arias, “A semantic appr