A survey on data mining and machine learning techniques for internet voting and product/service selection

Mr. Y. Subba Reddy and Prof. P. Govindarajulu

Department of Computer Science
S.V. University, Tirupati, AP, INDIA

Abstract
Internet voting is the process of collection of opinions on a specific, defined issue for the purpose of collecting information about objects like people, products, and services and so on. A voting method can be used as a rating process by adding a new dimension to it in terms of the group definition of ratable objects. Social networks like Twitter, LinkedIn, Facebook, and Google+ have gained remarkable attention in recent days. People started relying more on a social network for manifold information requirements. The results in large volumes of data, and this data is very complex to analyze manually. Data mining facilitates the extraction of useful knowledge from diverse aspects of the social network, to support decision making. Voting assistance applications are basically used to advise voters in electing the right alternative. Vote recommendation systems usually exploited during elections, may be extended to the selection of suitable products and services based on user preferences, ratings, reviews, and profiles. Recommended System exploits association among users by the way of item recommendation. Mining the constructive reviews from the user comments, votes, and preferences is an interesting area of research in recent times. In this paper, a detailed study and analysis are done on the existing techniques for a recommendation of rated /voted products, services policies, and users as well. It is explored about the role of classification, clustering, and other data mining and machine learning techniques in meeting the current data analysis and information needs. The modern trends of data and the applicability of the recommendation techniques to satisfy the current information needs is pointed. The extensibility of the voting advising techniques/recommendation techniques in various contexts is discussed along with the proposals for new procedures that suit the current information needs.

Key words:
Internet voting, Data mining, Clustering, Recommender system, Collaborative filtering, similarity measures.

1. Introduction

1.1 Recommender systems

Product recommendation is a critical business activity, which attracts customers towards products of their interest. The quality of a recommendation system is vital as the fulfillment of customers' needs is crucial in competitive business environment. A recommended system is an information system which recommends items to users based on their interests and profiles. It helps users to come across interesting items. These systems apply data analysis techniques to help users in finding the items they would like to purchase online based on the calculated likeliness score or the frequency list of recommended items for a given user. One of the popular and successful recommendation techniques is collaborative filtering. This technique works on the process of recommendations based on the opinions of other like-minded users. One more important technique is content-based recommendation where users are recommended items interconnected with their past buying history. By assisting the user in their attempt to search for a product and saving the customer' time the recommendation system is able to increase customers' loyalty and retention.

A collaborative filtering (CF) system collects information about the views of the users from the web servers or databases. This information is used to compute the similarity levels among the users with respect to their interests. Users with similar characteristics are grouped. This method had two challenges in terms of sparsely and capability. Recommendation systems that use collaborative filtering approach need long computation time that grows proportionally with an increase in a number of customers and the number of products or web pages.

There are some limitations with collaborative filtering (CF). The CF approaches depend mainly on rating information, which is normally provided by the user. A sufficient number of ratings needs to be provided before the system can return accurate recommendation. The performance is predictably impaired when the user is new, which is known as the "new user" phenomenon. Another kind of limitation is the "new item" problem, the system can recommend a new item provided the item is rated by a certain number of users.

Manuscript received September 5, 2017
Manuscript revised September 20, 2017

1.2 Opinion mining (OM) or Sentiment analysis (SA)

Opinions reflect people's view or sentiment about any object like products, people, services or events. Online opinions about products or topics constitute a part of web content. Web content mining is the subfield of data mining. Therefore, opinion mining or sentiment analysis is the process of analyzing user's views about a product or topic. People express feelings by giving their views in terms of opinions, feedback, suggestions or ideas about any object. These expressions are generally found on twitter, facebook reviews, blogs and other means of social media. People started using online sources to buy products or services. For this purpose, people are relying on product ratings and opinion polls available online. The growth in modern technology with faster rates committing people to look at the blogs, social networking sites or shopping websites for taking decisions such as voting for a politician, buying a product etc.

Some of the application areas of Opinion mining or sentiment analysis are as follows:

i) **Detecting malicious reviews:**
Someone may write reviews about product or service with negative intention. Opinion mining can aggregate these spam attempts.

ii) **Product or Service recommendation:**
We can easily evaluate other's views about any product or service. The comparative evaluation can be made by the competing brands. The evaluation results assist the user to select a right product or service.

iii) **Improvements in Product or service:**
The providers of a service or the manufacturers of a product make use of the positive and negative feedback of their product/service to improve the quality of their products/services.

iv) **As a marketing strategy:**
The analysis about a product/service with respect to the buyer's mindset or government policy can be conducted to catch the higher levels of marketing.

v) **Policy Making:**
Sentiment analysis assists the policymakers in creating and implementing finer policies for citizens. Sentiment Analysis is an evergreen field of research. It is the computational treatment of views, ratings, opinions, sentiments, and subjectivity of objects or entities under study. SA can be considered as a classification process generally at levels of abstraction. The data sets are the key elements in SA. The fundamental sources of data are from the product reviews. The analysis of these reviews is the vital source for the business organizations to go with the business decisions about their products. The review sites provide the reviews of the users. SA is not limited to products and can be applied to political debates, stock markets, news articles, movie reviews and booking systems. Sentiment classification uses two main approaches categorized as Machine learning approach and Lexicon based approach. Machine learning approach uses supervised and unsupervised learning. Lexicon based approach uses dictionary-based and corpus-based approaches. Feature Selection (FC) is the prerequisite task before SA. There is lack of benchmark data sets in the field of sentimental analysis [18]. IMDB and Amazon.com are popular sources of review data. The amazon.com is a source of many product reviews while IMDB is a source of movie reviews. SA and SC tasks can be successfully applied to these data sources. One more source of review data is twitter which becomes popular in recent past, where its tweets express people's opinions [21]. Review rating of the product given by the reviewer is generally retrieved along with the other characteristics. The rating to products is denoted by the number of stars that a reviewer allocates to the product, showing on a scale of 1 to 5. One star is the lower rating and the rating increases with the number of stars, where five stars are the best rating.

2. Problem Context

The interest in the area of recommender systems still remains demanding because it constitutes a problem-rich research area and a lot of practical applications. Recommendation systems help users to deal with information overload and provide personalized recommendations, contents, and services. For example, recommendations for books, CDs, and other products are done at e-business sites (like Amazon.com). Recommendations for a movie can be observed at sites like BookMyShow. Moreover, some vendors have incorporated recommendation capabilities into their commerce servers. However, despite of these advances, the present generation of recommended systems still needs enhancements to create recommendation strategies more effective and applicable to a broader vary of real-life applications, together with recommending vacations, certain types of financial services to investors, and products to purchase in a store made by a "smart "shopping cart. These enhancements include higher strategies for representing user behavior, provision of data regarding the items to be recommended, advanced recommendation modeling strategies, incorporation of various contextual information into the recommendation process and utilization of multi-criteria ratings. Development of less intrusive and additional flexible recommendation methods is that the current demand of recommendation systems. Opinion

aggregation and contradiction analysis is still missing in the current research and this is the hot topic of the present research (1)

3. Related Work

Ioannis Katakis et al. [9] introduced a collaborating filtering approach for voting aid applications. Voting advice applications (VAA) are gaining popularity with the developments in web-based electing and rating applications. The use of internet voting for people, policies, and products is very common today. In order to aid users in deciding what to vote, it needs an assistant system on which users (voters) can rely. The voting assistant applications are becoming popular recently. The system presents every user with a set of important issues and the user is asked to submit their opinion by assigning the ratings to issues. The application gathers the information for all candidates that are about to compete in the elections. Hence, it can provide recommendations to users about the candidates that agree with the user on these selected issues. The voting advice was approximated as a recommendation problem. Clustering was done for Like-minded voters according to their profiles using the k-means algorithm. The recommendation towards a party/candidate is indicated by the percentage of voting intention of the members of a cluster. The proposed method outperforms the other approaches like the user-candidate and user-average party voter similarity. This approach produced better predictive results with insight about voter opinion. Sung-Shun weng et al. [18] analyze customers' transaction records for preference analysis and then grouped customers with collaborative filtering technique, to recommend potential needs or interests of target customers based on customer's preference. The experiments of this research are designed based on movie rentals. The research is based on movie data, and the result shows that the combination of customer profile module and customer cluster module works better than just one module. The result also demonstrates that two-stage clustering can cluster similar customers effectively. Bashed et al. [1] have proposed an algorithm based on web usage mining called Profile Aggregations. Here Clustering is done on a database with respect to similarity of transactions. Page view clustering is also applied to predict the similar pages in each transaction. Magdalini Eirinaki et al. [11] proposed a clustering algorithm that provides quick recommendations to the user. Semantically coherent clusters are used for clustering. The authors used Domain ontology for a recommendation based on the keywords extracted from the web contents. Sikha Bagui et al. [15] presented techniques to study or mine voting patterns in the US House of Representatives. They have made use of whole data mining processing – from processing input data to advanced data mining techniques like association rule mining and decision tree generation. The exploratory data mining techniques, t-weights, gave the authors a picture of what percentage of each party voted on a particular issue. The attribute relevance analysis provided the information about the most discriminating issues. They also classify the political party of a member based on what they voted for. Through this study, the authors are able to see that there is quite a bit of difference in how the Democrats and Republicans vote, and they got some interesting results. Feng Hsu Wanga et al. [4] proposed clustering and association rule mining using web usage mining that provides better recommendation. Hierarchical Bisecting Mediods are used for clustering. J. Meghana Ramya et al. [9] proposed a method that calculates the review relevance value using votes for each review together with similarity and correlation. The user-generated content may be with non-relevant or less relevant data. Users can post their own review, which may or may not be related to the article. They can also post some irrelevant links or can reply to others reviews. This collection can have diverse importance with the article. Content which is not relevant can be treated as noise. Thus, it is needed to arrange all the reviews, as per the relevance. Better accuracy and efficiency are achieved with the capability of retrieving useful reviews from social networks. Based on the evaluation, it is found that that combination of SimRank, CorrelationRank, and VoteRank showed better performs by segregating the related reviews in the top with unrelated reviews at the end. Tuzhilin et al. [14] suggested five stages of personalization whereas customer responses are measured to guide other stages.The stages are: (1) Collection of customer information, (2) consolidating the profile of customers, (3) profile comparison, (4) presentation of personalized information, and (5) measurement of customer responses. Soude Fazeli et al. [16] proposed five quality metrics for recommendations. They are usefulness, accuracy, novelty, diversity, and serendipity. The authors conducted a user-centric evaluation to measure the quality of the recommendations supposed and the results are quite interesting and useful. Charif Haydar and Anne Boyer [2] proposed a density-based clustering algorithm called mutual vote (MV) based on a statistical model.It adapts itself to each vector's perception in its neighborhood. The aggregation is done on the perception of neighboring vectors. Cluster fitness is checked here. It is proved for lesser efforts needed and higher efficiency provided. They concluded that "the quality of labels might be slightly better in Mutual vote". Rishabh Soni, K. James Mathai [13] used cluster-then – predict approach. They cluster the tweets using a k-means algorithm and then perform classification using Classification Trees. This clustering operation makes the

data domain-specific which leads to a better classification of sentiments of a recently launched product. Effective results are obtained in sentiment analysis of a recently launched product 'iPhone 6s' of Apple using the 'Cluster-then-Predict' approach. The K-means clustering technique for clustering followed by CART algorithm to each of the test clusters were used to classify the sentiment. The proposed approach outperformed compared to other machine learning classification techniques such as Support Vector Machines, Random Forests, Naïve Bayes, etc. Compared with different existing algorithms, the proposed approach is superior in critically evaluating the parameters of accuracy, AUC and F Score. The results obtained from the proposed approaches are more interpretable. Glenn Pietila et al. [7] performed the study of product sound quality using paired-comparison methods. They consider the demographic characters if the jury members as preferences. They presented an unsupervised clustering technique that may be applied to jury-paired preferences to infer the quantity of subgroups in an exceedingly jury pool. Two succeed in American and European jurors. Soundarya.V et al. [17] use the approach of filtered reviews and business intelligence extraction. Here instead of finding the overall sentiment of each review, the authors found the sentiment score associated with each feature of the product. They concluded that the system works for any product by dynamically creating the feature term list from the given input. They developed a comprehensive system that can extract and analyze reviews at multiple levels. A. Razia Sulthana et al. [1] proposed an integrated framework for classifying the reviews of the customers based on the association rules using ontology. The authors first collected the reviews of active users in Amazon, who give more than five reviews for different books and also check the origin of the user. Further, they extend the same to the sub-networks of Amazon for the friends of the users. The dataset consists of ratings from users who rated a total of nearly 1, 00,000 books. Crawl is stored in the database and then preprocessed data is used to get frequent words using association mining. The refinement is done using ontology. Finally, hierarchical clustering is done to generate review summary. The accuracy of the results is improved because of the usage of ontology in the process. The authors concluded that the integrated approach with K-Means clustering and ontology provides improved performance for classifying Customer Reviews. Gilda Moradi Dakhel et al. [6] consider user versus product data where a user rates items. The items are usually books, articles, movies, music etc. A rate is a numerical score or a letter grade that is assigned by the users to the items. These datasets are worked out to predict some items that the active user has not seen and are likely to be liked by them. The authors pointed some challenges of the traditional collaborative filtering algorithm. The first one is the scalability problem, in which the computation increases nonlinearly with the increase in users /products. The second one is the sparsity problem, in which the item Vs user matrix has more number of cells with unseen /unrated information. This sparsity leads to the problems in measuring similarities. The third problem is the cold-start problem, which starts when a new user or a new item is entered into the list, and recommendations to new entries cannot be made suddenly. The authors proposed a new collaborative filtering algorithm based on user clustering and voting. Movie Lens collaborative filtering dataset: (http://www.grouplens.org/) is used to evaluate the proposed algorithm with 943 users that rated on 1682 movies. The users rated the movies between 1 and 5. 80% of the dataset is used for the training.5-fold Cross-validation is used for comparison. The results showed that the new approach is more accurate and less time-consuming than the traditional algorithm. Duen-Ren Liu et al. [3] developed a novel product recommendation methodology to fulfill the quality of recommendation. This approach combines group decision-making and data mining techniques. Lifetime value of a customer to a product or firm is generally evaluated based on recency, frequency, monetary (RFM) variables to determine the relative weights of RFM variables in evaluating customer lifetime value or loyalty the analytic hierarchy process (AHP) was applied. Considering the lifetime value also, Clustering techniques were worked out. An association rule mining approach was used to provide product recommendations. The authors found that the experimental results from this approach outperformed compared to other existing methods. Utkarsh Gupta et al. [22] proposed a novel recommender system based on a hierarchical clustering algorithm. The Item specific or user-specific information is grouped into a set of clusters using Hierarchical clustering algorithm called Chameleon. Following this, a voting system is used to predict the rating of a particular item given by users. The process started with the set of users with their features, based on which clustering is done using hierarchical cluster algorithm. Then for a given item and a user, the mapping is done to predict rating The prediction is done by mapping a user into a particular cluster and then voting scheme is applied for all user present in that cluster for the specific item. The performance of Chameleon based recommender system is evaluated by comparing it with an existing technique based on K-means clustering algorithm. The results showed that Chameleon based recommender system reduce errors significantly as compared to K-means based Recommender System. The dataset used is a Movie rating dataset with a sample of 80k ratings with information about users and items. A number of users are 943, with feature set (age, gender, occupation, pin code). The number of items is 1682 with feature set (release year, movie type). The proposed approach is better than the existing K-Means

based approach in terms of low Mean Absolute Error. Rakesh Kumar et al. [13] proposed a ranking mechanism that uses the numerous ratings of a review and calculates the aggregate score of the product. The ranking of various products is done by means of their reviews rating through rank voting method. The proposed product-ranking approach using reviews rating identify the top list of products and help the customer in choosing the best products. In this framework, the collected data is preprocessed and transformed for feature selection. After omitting the unimportant features the classification process train the data set to get the final model. Now the ranking approach picks the top k-products. The proposed approach significantly reduces the user time in selecting the right product. Umutoni Nadine et al. [21] presented a new hybrid competitive recommendation approach to improving the effectiveness through the competition process among a series of algorithms. Authors state that "combination of different algorithms will provide more accurate and effective recommendations than a single algorithm". The drawbacks of one algorithm are smoothened by some other algorithm. The collective approach of multiple recommendation techniques can reduce the weaknesses of an individual technique in a combined model. Keeping these facts in mind, a new hybrid method is suggested. The hybrid Recommendation Algorithm achieves more efficient and stable performance through the competition among a series of algorithms. The ranking is computed using individual algorithms and finally, the combined rank for each product is calculated using a variety of ranking functions. Techniques like sorting, averaging, SD are used for this integrated ranking approach. Joy deep Das et al. [11] present a Recommender System based on data clustering techniques. This approach deal with the scalability problem associated with the recommendation task. Different voting systems' algorithms are used to combine opinions from multiple users for recommending items of interest to the new user. In this work authors used DBSCAN clustering algorithm for clustering the users. Depending on the cluster to which the item belongs voting algorithms recommend items to the user. The idea behind this approach is "clusters –then-apply voting" which partitions the users of the RS into groups and then apply the Recommendation Algorithm separately to each group. The proposed system recommends an item to a user of a cluster based on rating statistics of the other users of that cluster. This approach avoids computations over the entire data, limits it to the targeted data and reduces the running time of the algorithm. The algorithm is tested on the Netflix prize dataset. Netflix with 17770 rating files such that one per movie is considered. The movie rating file consists of the rating information with the attribute set (movie id, year of release, title, average rating, genre) given by the customers to that movie. The rating of each movie given by all the customers is used to calculate an average rating. The system recommends, according to the user's preference of movie genres. For selecting the most popular items in a cluster, a voting based algorithm is applied individually to the clusters.

4. The Comparative Analysis of the Relative Approaches

Recommendation systems deal with voters and contestants or users and items. The association between the components of the pairs is offered by the information stored in a utility matrix about the degree to which a user likes an item or a voter intends to vote for a contestant. Normally, most of the entries of a utility matrix are unfilled, and the challenge of the system is to recommend items to users. The information required to recommend items is gathered based on the values of the known entries.

Content-based recommendation systems measure similarity by looking for common features of the items. Collaborative filtering recommendation system uses similarity of users by their item preferences and measures similarity of items based on the users who like them. The item profiles include features/attributes of items or tags given by the interested users. User profiles include the likes, number of visits, buying frequency etc. Classification of rows /columns of a utility matrix provide the interesting profiles about a user/item. Decision trees are used for this purpose.

Similarity measures are applied to rows/columns of the utility matrix to proceed with the collaborative filtering approach. A similarity measure is a numerical measure that reflects the degree to which two items are alike.

The similarity measures frequently used in the literature are:

a) Jaccard distance is a common measure which is appropriate when the entries of the matrix are only 1's and blanks.

b) Cosine distance is appropriate for more general values in the utility matrix. Before using this measure, the values of the utility matrix are normalized by subtracting the average value from each entry to get the cosine measure. It is a simple measure and gives the values in the range [0, 1].It has a drawback that, the variation in rating among users is not considered in the computation.

c) Adjusted Cosine similarity is overcoming the drawbacks of cosine similarity measure by eliminating the corresponding user average.

d) Extended Jaccard distance is one more measure which is suitable for non-binary values also.

Every measure has its own merits and shortfalls. Often the utility matrix is with majority number of blank cells. This sparsity creates difficulty to apply the similarity measures. To overcome this majority of the authors used the clustering of items and users as well to smooth the computation of similarity. The clustering approach provides the small number of user/item groups with which it is easier to deal.

UV decomposition is a technique to get two thin utility matrices, whose product is an approximation to the original utility matrix. Here U and V represent a relatively small number of issues compared to the original matrix. To measure the closeness of the product UV to the given utility matrix, RMSE (root-mean-square error) measure is used. This is computed by averaging the square of the differences between UV and the utility matrix, in those cells where the utility matrix is filled. RMSE is the square root of this average.

An important motivator of research in the field of recommendation systems was the NetFlix challenge. According to this, a prize amount of $1,000,000 was offered for a contestant/researcher who could produce a better algorithm that was 10% better than Netflix's own algorithm for predicting movie ratings by users. They award the prize in Sept. 2009.

The efforts made by various authors, tools and techniques along with datasets used and the gains achieved, are listed in the following comparative table.

Table 1: Authors and Techniques used for Recommendation systems.

S. No	Authors & Title	Approach	Tools	Experiments/Data	Remarks
1	Bamshad et al, "Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data"	Collaborative Filtering(CF)	Profile aggregation Clustering	Page view clustering on web content data	Scalability problem reduced
2	Ioannis Katakis et al , " Clustering Online Poll Data: Towards a Voting Assistance System"	Collaborative Filtering(CF)	k-means clustering	Voter profile clustering	produced better predictive results
3	Charif Haydar and Anne Boyer, "A New Statistical Density Clustering Algorithm based on Mutual Vote and Subjective Logic Applied to Recommender Systems"	Recommender Systems (density based clustering)	mutual vote (MV) method	Neighbor voting data	quality of labels improved
4	Rishabh Soni, K. James Mathai, " Effective Sentiment Analysis of a Launched Product using Clustering and Decision Trees",	Cluster-then-Predict	feature extraction, *K-means clustering, CART algorithm*	User opinion on the product 'iPhone 6s'from Twitter.	More accurate and interpretable results.
5	Sikha Bagui et al, "Data Mining Techniques To Study Voting Patterns In The US".	Whole data mining processing,	t-weights	Roll Call Votes data	*interesting results to interpret voting patterns of US House of Representatives.*
6	Glenn Pietila and Teik C. Lim, "Identifying Preferences in a Jury Study Using Clustering Techniques", University of Cincinnati, Cincinnati, Ohio. SOUND & VIBRATION/AUGUST 2014	visual screen test	K-means clustering and Ward's agglomerative clustering algorithms	40 American and more than 40 European jurors	Sub grouping of preference data.
7	Soundarya.V et al, "Extracting business intelligence from online product reviews"	Review filtering	Filtering and sentiment analysis	Mobile Phones, Tablets and Television Sets for the Samsung manufacturer with various models.	Comprehensive system with extracting and analyzing reviews at multiple levels.
8	J. Meghana Ramya et al, "An Effective Approach to Rank Reviews Based on Relevance by Weighting Method"	Review relevance, weighted vote	Similarity and correlation	Data from http://www.androidpolice.com/, http://www.phonearena.com/, http://moneycontrol.com/.	Relevant review analysis provides better and quicker outlook.

9	A. Razia Sulthana and Ramasamy Subburaj, "An Improvised Ontology based K-Means Clustering Approach for Classification of Customer Reviews"	Integrated approach	Association rules using ontology. K-means algorithm	Ratings for books from December 2010 to March 2015,from Amazon users	The usage of ontology and modified K-means increased the accuracy.
10	Gilda Moradi Dakhel, Mehregan Mahdavi, "A New Collaborative Filtering Algorithm Using K-means Clustering and Neighbors' Voting"	collaborative filtering algorithm based on user clustering	k-means clustering algorithm to categorize users	MovieLens collaborative filtering data set: *(http://www.grouplens.org/) is used*	New CF approach is more accurate and less time-consuming than the traditional algorithm.
11	Duen-Ren Liu, Ya-Yueh Shih, " Integrating AHP and data mining for product recommendation Based on customer lifetime value".	Integrated AHP, clustering and association rule mining	k-means clustering With weighted RFM.	Product data	Experimental results from this approach outperformed compared to other existing methods.
12	Utkarsh Gupta and Dr Nagamma Patil," Recommender System Based on Hierarchical Clustering Algorithm Chameleon"	Chameleon Clustering based recommender system (RC)	Clustering and voting	Movie rating dataset	Low Mean Absolute Error.
13	Rakesh Kumar et al, "A Framework for Ranking Products Using Ranked Voting Method".	Rank voting	Ranking and classification	Sample data	Reduce the analysis time for user.
14	Umutoni Nadine et al," *Competitive Recommendation Algorithm for E-commerce*".	Integrated ranking approach	Competitive recommendation algorithm	MovieLens data set	More efficient and stable recommendation through integrated approach than individual approaches.
15	Joydeep Das et al," Clustering-Based Recommender System Using Principles of Voting Theory"	Clustering based collaborative filtering(CF)	DBSCAN clustering algorithm	Netflix prize dataset	Avoids unnecessary computations and reduce computational time.

Summarizing 50 previous works on the basis of approaches used, data considered and application areas the following facts are elucidated.

Commonly used techniques for recommendation systems include:

a) Bayesian classifiers
b) Clustering
c) Decision trees
d) Artificial neural networks
e) Random Forest algorithms.
f) Bayesian networks
g) Clustering
h) Artificial neural networks
i) Linear regression
j) Probabilistic models

The recommender systems applications identified using the study include but not limited to:

People and policies: Internet voting, voting on policies of business organizations.

a) Entertainment: movies, music, pubs, and resorts.
b) Content: personalized newspapers, documents, Web pages and Q&A sites.
c) E-education: e-learning and e-mail filters.
d) Electronic commerce: products, books, cameras, TVs, PCs etc.
e) Services: travel services, expert's advice, rent houses, matchmaking.

Table 2: Categorization of Recommender system approaches

Recommender system approach	Percentage of use
Content-based	36
Collaborative	40
Hybrid	16
Others	8

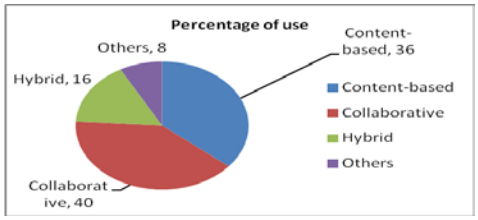

Fig. 1 Percentages of usage of different approaches.

Table3: Categorization of Type of data considered

Type of data considered	**Percentage of use**
Opinion polls, voting	**15**
Product reviews	**60**
Service reviews	**25**

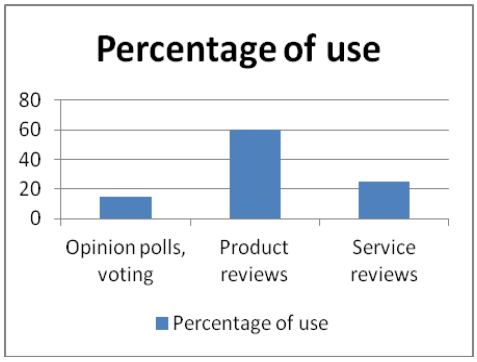

Fig. 2 Percentages of usage of different data sets

Table 4: Categorization of Type of Data mining and machine learning techniques used

Data mining and machine learning techniques used	Percentage of use
Classification	25
Clustering	45
Association rule mining	10
Cluster-then-classify	20

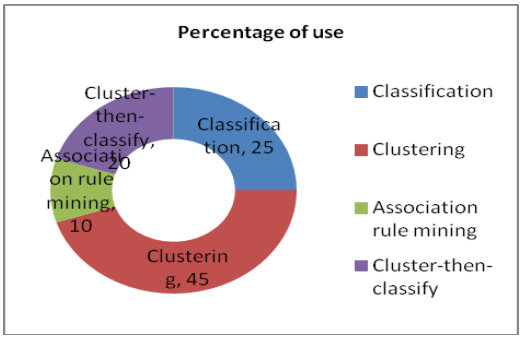

Fig. 3 Percentages of usage of different data mining techniques

5. Improvable Capabilities of Recommender Systems in Future

Recommender systems can be extended to several user-centric and data-centric contexts. The required betterments include improvements in the understanding of opinion data and capability in dealing with multi-criteria ratings. The recommended system's framework can be mould to support multiple dimensions. The present dimensions are user and items (objects). The third dimension called priority /preference can be incorporated. With this added dimension the weights for the objects under study can be calculated by aggregating the ratings and reviews from multiple dimensions. The application space of the recommendation systems can be expanded to many other areas of user interest.

6. Conclusion

Today people are relying on web content and information systems for electing the right person, for buying the right product and for getting the best services. The recommendation systems gather the related information regarding item preferences and profiles; analyze the same and advise the user back to make decisions about products, people, policies, and services. From the day of the availability of electronic and web content, researchers started analyzing the content to extract the vital information for better recommendations. In this way, recommendation systems become popular in assisting numerous problem contexts. Data mining and machine learning techniques provide the way for the fast evolution of recommendation techniques. The research in this area is ever trending. To overcome the present limitations and to improve the reliability and dependability of recommendation systems a new dimension called user-centric preferences is proposed in this paper. This new dimension can be incorporated into the system by computing the weights as a function of users' views, preferences and rating/voting. The recommendation systems with present techniques embedded with the proposed weights will provide better results.

References

[1] Razia Sulthana and Ramasamy Subburaj, "An Improvised Ontology-based K-Means Clustering Approach for Classification of Customer Reviews", Indian Journal of Science and Technology Vol 9(15), 10.17485/ijst/2016/v9i15/87328, April 2016.

[2] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa. "Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data", 2002.

[3] Charif Haydar and Anne Boyer, "A New Statistical Density Clustering Algorithm based on Mutual Vote and Subjective Logic Applied to Recommender Systems", UMAP'17, July 9-12, 2017, Bratislava, Slovakia.

[4] Duen-Ren Liu, Ya-Yueh Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value", Information & Management 42 (2005) 387–400 @2004 Elsevier B.V.

[5] Feng Hsu Wanga, Hsiu-Mei Shao. 2004, "Effective personalized recommendation based on time-framed navigation clustering and association mining", Expert Systems with Applications, 27, 365–377.

[6] G. Adomavicius, and A. Tuzhilin, "Using data mining methods to build customer profiles", IEEE Computer, 34(2), 001, pp. 74-82.

[7] Gilda Moradi Dakhel, Mehregan Mahdavi, "A New Collaborative Filtering Algorithm Using K-means Clustering and Neighbors Voting", 978-1-4577-2152-6/11 @2011 IEEE.
[8] Glenn Pietila and Teik C. Lim, "Identifying Preferences in a Jury Study Using Clustering Techniques", University of Cincinnati, Cincinnati, Ohio, SOUND & VIBRATION/AUGUST 2014.
[9] Ioannis Katakis, Nicolas Tsapatsoulis, Vasiliki Triga, Constantinos Tziouvas, "Clustering Online Poll Data: Towards a Voting Assistance System", 2012 Seventh International Workshop on Semantic and Social Media Adaptation and Personalization.
[10] J. Meghana Ramya Shri and V. Subramaniyaswamy 2015. "An Effective Approach to Rank Reviews Based on Relevance by Weighting Method", Indian Journal of Science and Technology, Vol 8(11).
[11] Joydeep Das, Partha Mukherjee, Subhashis Majumder, and Prosenjit Gupta, "Clustering-Based Recommender System Using Principles of Voting Theory", 2014 International Conference on Contemporary Computing and Informatics (IC3I) @2014 IEEE.
[12] Magdalini Eirinaki, Charalampos Lampos, Stratos Paulakis, Michalis Vazirgiannis. 2004, "Web Personalization Integrating Content Semantics and Navigational Patterns", WIDM'04, November 12-13. Washington, DC, USA. Copyright 2004 ACM 1-58113-978-0/04/0011.
[13] Rakesh Kumar, Aditi Sharan, Payal Biswas, "Framework for Ranking Products Using Ranked Voting Method", 2016, Second International Conference on Computational Intelligence & Communication Technology© 2016 IEEE DOI 10.1109/CICT.2016.138.
[14] Rishabh Soni, K. James Mathai," Effective Sentiment Analysis of a Launched Product using Clustering and Decision Trees", IJIRCCE, Vol. 4, Issue 1, January 2016, ISSN(Online): 2320-9801.
[15] Tuzhilin and G. Adomavicius, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6, 2005, pp.734-749.
[16] Sikha Bagui, Dustin Mink, and Patrick Cash, "Data Mining Techniques to Study Voting Patterns in the Us", Data Science Journal, Volume 6, 20 April 2007.
[17] Soude Fazeli et al, "User-centric Evaluation of Recommender Systems in Social Learning Platforms: Accuracy is Just the Tip of the Iceberg", IEEE Transactions on Learning Technologies 2017.
[18] Soundarya.V et al, "Extracting business intelligence from online product reviews", International Journal on Soft Computing (IJSC) Vol.4, No.3, August 2013.
[19] Sung-Shun Weng, Mei-Ju Liu, "Personalized Product Recommendation in E-Commerce", Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04) 0-7695-2073-1/04.
[20] Tsytsarau Mikalai, Palpanas Themis. "Survey on mining subjective data on the web", Data Min Knowl Discov 2012; 24:478–514.
[21] Umutoni Nadine, Huiying Cao, Jiangzhou Deng, "Competitive Recommendation Algorithm for E-commerce", 2016 12th International Conference on Natural Computation, Fuzzy Systems, and Knowledge Discovery, (ICNC-FSKD) 978-1-5090-4093-3/16 ©2016 IEEE.
[22] Utkarsh Gupta and Dr. Nagamma Patil, "Recommender System Based on Hierarchical Clustering Algorithm Chameleon", 2015 IEEE International Advance Computing Conference (IACC), 978-1-4799-8047-5/15/$31.00 c_2015 IEEE.
[23] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal (2014) 5, 1093–1113.

Y.Subba Reddy received M.Sc (Computer Science) degree from Bharathidasan University, Tiruchirapalli, TN and M.E degree in Computer Science & Engineering from Sathyabama University, Chennai, TN. He is a research scholar in the Department of Computer Science, Sri Venkateswara University, Tirupati, AP, India. His research focus is on Data Mining in Clustering and Similarity measures.

P.Govindarajulu, Professor, Department of Computer Science, Sri Venkateswara University, Tirupathi, AP, India. He received his M. Tech., from IIT Madras (Chennai), Ph. D from IIT Bombay (Mumbai), His area of research are Databases, Data Mining, Image processing, Intelligent Systems and Software Engineering