An Efficient User Centric Clustering Approach for Product Recommendation Based on Majority Voting: A Case Study on Wine Data Set

Y. Subba Reddy and Prof. P. Govindarajulu

S.V. University, Tirupati, Andhrapradesh, India

Abstract
Analysis, design and implementation of software systems for online services, are a tedious and challenging. Amazon software provides product recommendations, Yahoo! dynamically recommends WebPages, afflux creates recommendations for movies, and Google creates advertisements on the Internet. Items are recommended based on the preferences, needs, characteristics and circumstances of users. The wine data set has been in use in research for several years and still it remains as the benchmark data set. Quality of wines is difficult to define as there are many factors that influence the perceived quality. This paper presents a critical review of research trends on Wine quality and user-centric similarity measures as well. A novel user centric similarity measure in product clustering is proposed to evaluate the popular Wine data set named Red Wine dataset. The experimental results obtained in this work are able to provide better recommendations to product buyers than the existing systems. The proposed approach is competent to group the Red wine dataset into ordered groups of preferred wine variants and can judge the wine quality based on these user preference groups.

Keywords:
User-centric, clustering, preference/voting/ranking, wine dataset,

1. Introduction:

1.1 About the wine datasets

The intrinsic characteristics (visual, taste, smell), environmental characteristics (climate, region, site) and management practices (viticulture practice), as well as physicochemical ingredients (acid, pH, etc.) are the factors of interest in assessing the quality of Wine. Data mining techniques in predicting wine quality are in progress, with some promising results in the domain. Physicochemical and sensory tests are crucial in Wine certification. It is the routine practice in physicochemical laboratory tests, to characterize wine by determination of density, alcohol or pH values, but sensory tests rely mainly on human experts. Wine classification is a difficult task as taste is the least understood of the human senses. The relationship between the physicochemical and sensory analysis are complex to understand. In the food industry, in addition to the food quality research, machine learning techniques have also been applied in classification of wine quality. Machine learning methods provide the way to build models from data of known class labels to predict the quality of a wine. In old days Wine was considered as a luxury item. Today, it is popular and enjoyed by a wide variety of people. Professional wine reviews offer insights on wines available in large quantity in each year. A systematic way is needed to utilize those large number reviews to benefit wine consumers, distributors, and makers. No two persons judge the wine alike even they taste the wine simultaneously while being able to share and detect all the same attributes. Experience helps a lot and hinders the taster. So assessing the quality of wine depending only on the taster's experience and sensing is a big process

1.2 About the user-centric approaches

The task of any recommended system is to provide the user's information about finding the preferred items from the very large set of items. The preference of a user on a particular or selected item is obtained by Voting/rating response of the user to the recommender system.
Nowadays the information in the world is generating more than thousands of times faster than the actual data process capacity. Collaborative filtering is a statement-of-the-art technique used for efficient data processing. Two main problems of collaborative filtering are scalability and quality. Existing collaborative filtering algorithms will be able to see lakhs of neighboring products but the actual demands of modern computerized systems are in hundreds of lakhs. Many of the existing algorithms are suffering from performance problems. Another important point is that there is a need to improve the quality of the recommendations for the customers. Scalability and quality generally contradict each other. A good balance between these two factories is needed. The relationship between products is most important than the relationship between customers. Recommendations for the customers are evaluated by finding products that are similar to other products according to customer likes. Recommender systems for very large communities should not expect that

Manuscript received October 5, 2017
Manuscript revised October 20, 2017

liking of each customer is known to all other customers. Sometimes there is a need to select the objects based on preferences of values of attributes. One way to select best products from the set of products is by using the linear weighted function of preferences and attribute values. In this method, the linear function computes the score for each product Using scores products are ranked in the order that makes each for selecting top score products. Present web applications are able to process multi parametric ranked queries only on small datasets. Many databases do not support efficient evaluation of customer preference queries. Here for response and throughput is retrieval first and then scoring function is evaluated for each tuple. Finally, a few products are ranked and displayed. The main disadvantage of the traditional database queries is that these query evaluation techniques need retrieval and ordering of the entire data preference-based queries can be augmented with features for scalability purpose.

Special Indexing techniques are needed for efficient processing weighted preference based query execution. These indexing techniques are useful for obtaining linear function preference based optimized query results. Linear function optimization queries are called preference selection queries because such queries retrieve tuples maximizing a linear function defined over the attribute values of the tuples in a relation. Linear preference function also needs to modify for obtaining better query results

2. Related work

A good number of research papers have been published on wine quality that is mostly based on the empirical studies in the wine industry. Most of the research was endorsed to assess wine quality using physicochemical data is based on small sample sizes. In [10] pattern recognition approaches that include clustering, principle component analysis, nearest neighbors, etc. were applied to classify wines from Galicia (northwestern Spain) among several different brands. The dataset used consists of 42 white wines. Principle component analysis (PCA) for wine classification according to the geographical region was reported in [13]. The authors used the data set that contains 33 Greek wines with physicochemical variables. The details of a 2-stage classification done (principle component and clustering) from 24 industrial fermentations of a particular type of wine were given in [1. This study tried to detect undesirable fermentation behavior.

In [6] authors proposed an improved KNN method with weights, which considerably improves the performance of KNN method. The authors employed a kind of preprocessing on train data. They introduced a new value named Validity to train samples which cause to more information about the situation of training data samples in the attribute space. This new value takes into accounts the value of stability and robustness of any train samples regarding with its neighbors. KNN with applied weights employs validity as the multiplication factor yields to more robust classification rather than simple KNN method, efficiently.

The Wine dataset is the result of a chemical analysis of wines grown in the same region in Italy that derived from three different cultivars. The analysis determined the quantities of 13 attributes found in each of the three types of wines. This data set has been in use with many others for comparing various classifiers. In the context of classification, this is a well-posed problem with well-behaved class structures.

Data mining techniques to classify the quality of wines using a larger physicochemical data set were used in more recent works. Cortez and his colleagues [12] built models using support vector machine, multiple regression, and neural networks. A dataset with a large number of records is considered (vinho Verde samples from the Minho region of Portugal). A computational procedure was developed that performs simultaneous variable and model selection. Support vector machine achieved desirable results, “outperforming the multiple regression and neural network methods”. This model is vital in supporting the oenologist wine tasting evaluations and to improve wine production. The results of this research are relevant to the wine science domain, helping in the understanding of physicochemical characterization and the things that affect the final quality. In[7] authors enforced unsupervised neural network (NN) based on Adaptive Resonance Theory (ART1) as an alternative to statistical classifier so as to discriminate among the 178 samples of wine possessing 13 numbers of attributes. The dimensionality of the feature variables was reduced to 5 by principal component analysis (PCA).Out of 13, the first 2 numbers of principal components captured over 55.4 % of the variance of the wine dataset. Nonhierarchical K-means clustering algorithm was used to choose the classes available among the samples of wine. Appalasamy et al [2] applied two classification algorithms, decision tree and Naive Bayes and compared results with the recent work results.

In [3] authors proposed a brand new data science area named Wine informatics. In order to automatically retrieve wines' flavors and characteristics from reviews, which are stored in the human language format, authors proposed a novel "Computational Wine Wheel" to extract keywords. Two completely different public-available datasets are produced based on the new technique in their paper. The hierarchical clustering algorithm is applied to the primary dataset and got purposeful clustering results. Association rules mining is performed on the second dataset to predict whether or not a wine is scored higher than 90 points or

not supported on the wine savory reviews. Fivefold cross-validation experiments were executed based on different parameters and results with a range of 73% to 82% accuracy were generated. This new domain will bring huge benefits to fields as diverse as computer science, statistics, business, and agriculture.

In [5] authors proposed a data analysis approach to classify wine into different quality categories. A data set of white wines of 4898 records was used in the analysis. As the data set was imbalanced with about 93% of the observations are from one category with respect to the occurrence of events in it, Synthetic Minority Over-Sampling Technique (SMOTE) was applied to oversample the minority class. A balanced data was considered to model a classifier that categorizes a wine into three categories. These categories include high quality, normal quality, and poor quality. Three classification techniques used in this work include decision tree, adaptive boosting and random forest. Among the techniques, random forest produced to produce the desired results with the minimal error. Based a Wine dataset of 4898 instances the authors attempt to build models that classify different wines into quality categories. With this model, the test data is tested. The quality variable is assessed by many factors The authors concluded that the analysis would give a clearer idea to winemakers as to which variables influence the quality the most and what steps could be an attempt to attain more desirable outcomes.

In [8] authors used Analytical Hierarchy process (AHP) classification algorithm. This algorithm provides the way to recommend wine on the basis of the components of the wine. Wine selection on the basis of its attributes is a different approach. The Machine Learning Techniques used here helped in finding the component accuracy of wine attributes. The Analytical Hierarchy Process (AHP) is used for arranging and examining complicated problems by mathematical calculations. AHP defines multiple attribute issue to advise a particular commodity to an individual. The process of AHP is mainly used to calculate weights. The inputs for AHP are relative preferences and attributes. The authors have taken red wine dataset and weights were allotted to them based on AHP. The obtained results after analysis of the data were used for recommending a wine to individuals.

Users need reasonably good recommendations in finding products according to their likes and dislikes. In [14] authors compare data-centric evaluation with user-centric evaluation and obtain remarkable results in favor of user-centric approach. In [4] authors used a new clustering algorithm based on the mutual vote, which adjusts itself automatically to the given dataset, needs minimum overhead in terms of parameters, and also able to detect clusters with different densities in the same dataset. Currently, many Voting/Rating machine searing based automated recommender software systems are developing continuously by many companies for voting based selection of products. A customer would be interested in purchasing products that are similar to the products that he/she liked earlier. In [15] authors proposed an algorithm that combines user-based approach, item-based and Bhattacharyya approach. The main advantage of this hybrid approach is its capability to find more reliable items for recommendation. Collaborative filtering technology works by creating a database of preferences for products by their customers. Collaborative technology is becoming popular in the latest research areas such as E-business, Banking, Space, and Share Market and so on. In [11] authors proposed an improved collaborative filtering algorithm that combines k-means algorithm with CHARM algorithm. This hybrid approach improved the prediction quality of recommendation system. In [9] authors tried to improve the learning speed by splitting the cluster tree into sub-clusters and by using exploration and exploitation phases and aggregates as well. From user-centric sensor data, Friendbook discovers lifestyles of users and measures the similarity of lifestyles between users. It recommends friends to users if their lifestyles have high similarity. In [16] authors model the daily life of users as life documents. Using these documents lifestyles are extracted by using the Latent Dirichlet Allocation algorithm.

2.1 The evolution summary:

The evolution in the field of wine quality research is threefold.

i) With respect to the dataset

Most of the research about Wine has been carrying out based on two popular data sets named Wine and Wine quality which are publicly available on UCI machine learning platform. Two datasets are available of which one dataset is in red wine and have 1599 completely different varieties and therefore the alternative is on white wine and has 4898 varieties. All wines are produced in a particular area of Portugal. Data are collected on 12 completely different properties of the wines one amongst that is Quality, supported on sensory knowledge, and therefore the rest are on chemical properties of the wines together with density, acidity, alcohol content etc. All chemical properties of wines are continuous variables. Quality is an ordinal variable with the possible ranking from 1 (worst) to 10 (best). Every form of wine is tasted by three independent tasters and therefore the final rank assigned is that the median rank given by the tasters. Knowledge regarding reviews given by users is additionally out there.

ii) With respect to the objectives

The common objective of most of the researchers is to assess/predict wine quality based on the wine

characteristics and user reviewers and use the results to recommend winemakers, marketers, and users as well.
iii) With respect to the methodology
The majority of the methods and tools are designed based on machine learning and data mining techniques. Classification, clustering, and association rule mining are the common approaches observed. Wine attributes and user reviews are the main sources of input. These datasets contain real-valued attributes. The recent works in the field account for the adoption of advanced machine learning techniques. To maintain the uniformity the original values are normalized in some papers. Some researchers converted the normalized values to binary values to get the ease of computations. Techniques like principle component analysis are used to reduce the dimensionality of the datasets. K-means clustering and its modified variants are used for clustering. Bayesian networks, neural networks, and other classification techniques are adopted. A few papers address the weighted measures and hybrid deep learning techniques to improve the usability of the results.

3. Objectives and methodology

3.1 The objectives of the proposed work:

The objectives of the present work include:

a) To introduce novel query approaches and efficient algorithms for their execution.
b) To introduce a novel user-centric similarity measures in product clustering.
c) To apply the proposed clustering to evaluate the popular Wine dataset
   i) To test whether the wine quality is well supported by its chemical properties.
   ii) To assess Wine quality in a novel way.
d) To compare the user-centric metrics with conventional metrics.
e) To analyze the results and present the findings.

3.2 Proposed methodology:

The product or service recommendation needs a critical evaluation of product features and the user preferences for the product. A hybrid algorithm is proposed in this work that comprises the following subtasks:

a) Top k and reverse top k queries approaches to rank the products based on the customer preferences.
b) Incorporating the weighted ranks in similarity calculations, to cluster the products.
c) Nearest neighbor similarity search using Jaccard coefficient and modified Jaccard coefficient.

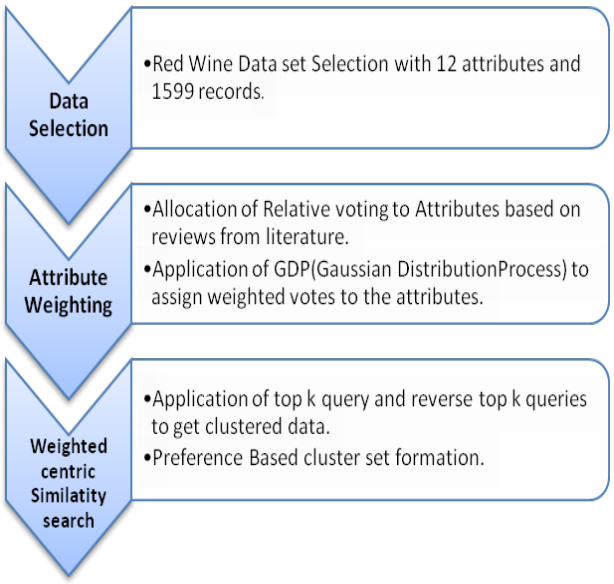

Figure 1: Process flow

3.3 Rationality/Justification of the proposed Work:

The proposed process groups the wine dataset records into priority based clusters. The clustered data using classification form a model to assign the test data records with a recommended voting label. Most of the previous research on wine data limited to normal clustering and classification approaches depending on the taster sensing data whereas, the proposed novel hybrid approach can recommend the user a better wine combination without depending on the taster sensing data.

4. Algorithm and Description

The clustering algorithm using Jaccard coefficient similarity measure, weighted query method with top k and reverse top k approaches for clustering Red wine dataset is as follows:

4.1 Algorithm

1. Read the data set of ‘n’ tuples into the array data structure.
2. Produce voting/rating/preferences details of' a number of customers for all the products using "***Weighted Attribute Collection”*** method.
3. Prepare top-k query results for all the products for all the' m' number of customer voting/rating/preferences
4. Compute reverse top-k queries for all the products obtained in the step-3.

5. s = get Reverse Set Products Count.
6. threshold = get Threshold Value
7. minimumcount = s * threshold
8. While(s > minimumCount) do
9. {
10. StartCluster = first cluster of the present list of products
11. For cluster i = 2 to last in the current list compute similarity measure, Sim(StartCluster, i) and store the result.
12. Combine all the groups whose similarity measure value > than the specified threshold value into one cluster.
13. presentCount = number of groups combined in the step11.
14. s = s - presentCount value 'n' is stored efficiently in the memory. Each customer specifies voting/rating/performance details of products. Traditional
15. }

4.1.1 Method: Weighted Attribute Collection

This method collects votes for individual physiochemical attributes of wine variants based on the user preferences.This collection is generated synthetically by Gaussian distribution method to proceed with the present work.

4.2 Algorithm Description

Traditional similarity finding methods use distance metrics on values of attributes for finding similarity measures between products (objects). The proposed method considers attributes and their respective voting/rating/performance details as well for computing similarity between two products. Weighted query method with top k and reverse top k approaches is considered.

A linear scoring function is used for finding similarity measure between two products. This linear function uses both values of attributes and the corresponding voting/rating/performance values. A Top–k query provides the list of a top-k number of the best products based on the preferences with the help of a linear function. Reverse top–k query gives all customers who have included in top-k products lists. It was assumed that the value of the variable s represents the total number of products included in the voting/rating/performance lists of customers. The control structure while loop iterates to groups products into clusters. The variable present count indicates the total number of products clustered in the current iteration. The iteration process ends with updating the total number of products to be clustered.

5. Dataset

The dataset considered is the "Red Wine Quality Data Set with 1599 records and 11 attributes" from the UCI Machine Learning Repository. The data were recorded with the help of a computerized system (iLab). Each entry denotes an analytical or sensory test and the database was exported into a single sheet in (.csv) format.

Input variables (based on physicochemical tests): 1. Fixed acidity 2.Volatile acidity 3. Citric acid 4 .residual sugar 5.Chlorides 6. Free sulfur dioxide 7. Total sulphur dioxide 8.Density 9.PH 10.Sulphates and 11.Alcohol. sulphur

Tartaric acid, citric acid, and malic acid are the important ingredients of wine. Ascorbic, sorbic and sulfurous acids are added during winemaking. The residual sugar determines the sweetness of a wine and plays a major role in determining the taste of a wine. In wine, a by-product of yeast metabolism is Alcohol. The attribute preferences of the Wine variants are collected from the literature review.

6. Results

Case wise Execution and analysis is done by fixing some algorithmic variants.

CASE 1:

Fixed things: Number of reviews/voting points considered: 100. Threshold value 0.1.

Varying things:

Sub case 1a:

The value of k for top k query = 50

Number of clusters formed: 13

Largest cluster size: 82

Smallest cluster size: 1

Total time: 2 seconds

FINAL CLUSTERS ARE:

Cluster No	Cluster Elements
1	[8, 26, 39, 52, 90, 130, 134, 304, 659, 661, 743, 962, 963, 1237, 1253, 1337, 1338, 1339, 1357, 1382, 1416, 1501, 1522]
2	[19, 105, 172, 173, 257, 294, 323, 627, 628, 811, 1333, 1356]
3	[27, 45, 49, 53, 57, 65, 81, 98, 99, 123, 128, 144, 151, 171, 223, 689, 728, 729, 807, 809, 814, 827, 831, 908, 913, 916, 940, 980, 985, 987, 997, 998, 999, 1004, 1006, 1015, 1021, 1022, 1023, 1024, 1061, 1068, 1069, 1077, 1078, 1079, 1081, 1087, 1114, 1125, 1127, 1131, 1198, 1201, 1202, 1215, 1254, 1278, 1280, 1288, 1294, 1300, 1335, 1348, 1349, 1370, 1393, 1396, 1397, 1404, 1419, 1421, 1433, 1439, 1462, 1481, 1483, 1485, 1489, 1490, 1492, 1506]
4	[37, 51, 127, 129, 531, 536, 705, 774, 775, 805, 810, 912, 1109, 1415, 1417, 1479]
5	[66, 1277, 1347]

6	[266, 672, 674, 874, 954]
7	[282]
8	[518, 1098, 1100]
9	[603, 1192, 1194]
10	[806, 808, 914, 1057, 1060]
11	[910, 915, 986, 1000, 1065]
12	[911]
13	[953]

From the above table, it is observed that the target cluster formed was 82 items. These top 82 items represent the highest preferred wine types by the customer from this set. We can make recommendations from a new customer based upon made the attributes preferences of wine.
As this clustering consider attributes preferences but not consider the total quality values, this approach saves the time as well as the effort in data clustering based on user preferences. Similarly, for all other cases, the same clustering approach is followed and the results obtained are presented in summary format

Table 1:

Case No.	Fixed things		Varying things				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 a)	100	0.1	50	13	82	1	2
b)			100	14	160	1	2
c)			150	11	190	1	3
d)			200	12	160	1	2
2 a)	200	0.1	50	16	80	1	4
b)			100	13	159	1	4
c)			150	14	195	1	5
d)			200	15	242	1	4
3 a)	200	0.1	500	19	214	1	12
b)			1000	24	204	1	29
c)			2000	31	139	1	41
4 a)	200	0.3(varying)	2000	75	230	1	16
b)		0.5	2000	158	97	1	6min 31 sec
c)		0.7	2000	247	174	1	10min 44 sec

(1) No. of reviews /voting points considered
(2) Threshold value
(3) Value of k for top k query
(4) No. of clusters formed
(5) Largest cluster size
(6) Smallest cluster size
(7) Total time

7. Comparisons and findings

7.1 The comparison between the user-centric metrics and conventional metrics.

To evaluate the similarity between two data items, many authors have proposed different similarity metrics like the Euclidean distance and the cosine similarity. Using these metrics the similarity among data items is computed based on their attribute values. These metrics do not consider users' opinions. As traditional similarity measures for clustering do not consider weights and preferences, the clusters formed to represent the sets of objects (Wines in this study) which are closed to one another in terms of their attribute values. The clusters are formed based on the attribute values only. This cluster information conveys a little about the object (product) quality and provides a little or nothing to the customer in decision making towards a product selection.
The user-centric approach for similarity computation takes into consideration not only the attribute values but also users' preferences. Generally, a business manager would like to know the customer views about the company products. They also want to know the comparison between their products and the competitors existing products. It is quite interesting to know which of the products belong to the favorite list of most of the customers.
The proposed user-centric approach uses top k query and reverse top k query approaches and considers user preferences/votes. In this approach, the clusters formation considers top k list of preferred products. The obtained knowledge helps to focus on products, having similar groups of customers that rank them in high positions. This information provides the way for better and efficient marketing policy establishment and to create clusters of products that are preferable to particular customer groups. So based on the cluster information, it is possible to recommend products or assess the quality of a product (Wine).
Now, a business manager is able to perform a query that returns similar products (new products also) which are based on the product characteristics and users' preferences as well.

7.2 The comparison of results in various contexts

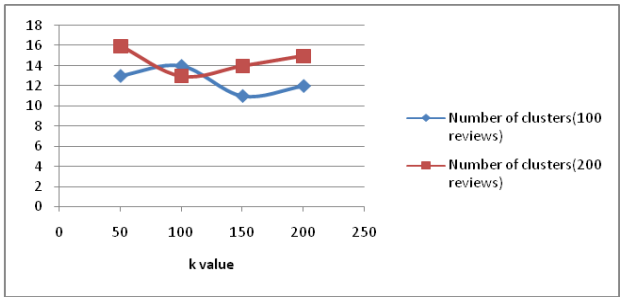

Figure 2: Number of clusters for 100 and 200 reviews with varying k value

From Figure 2 it was observed that the top k query value influencing the number of clusters well for large values of k and as the number of reviews increased the number of clusters is finer.

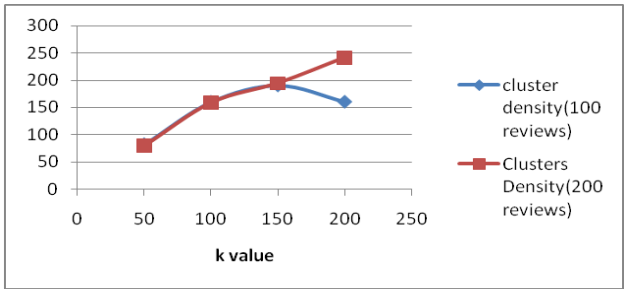

Figure 3: Cluster density for 100 and 200 reviews with varying k value

From Figure 3 it was observed that the top k query value influencing the cluster density well, as the number of reviews increased.

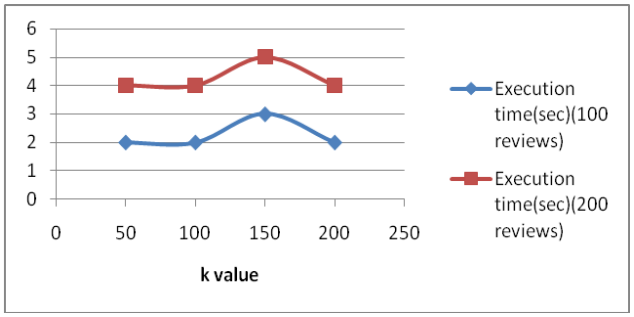

Figure 4: Execution time for 100 and 200 reviews with varying k value

From Figure 4 it was observed that with the increase in top k query value the execution time increased a little and falls back for larger k values. More reviews need more execution time

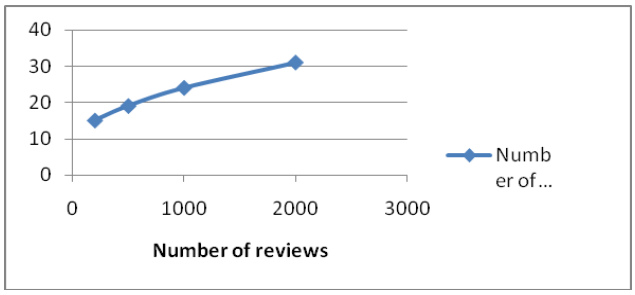

Figure 5: Number of reviews Vs number of clusters

From Figure 5 it is observed that with the increase in a number of reviews the proportional increase is observed in the number of clusters. Finer clustering needs larger reviews.

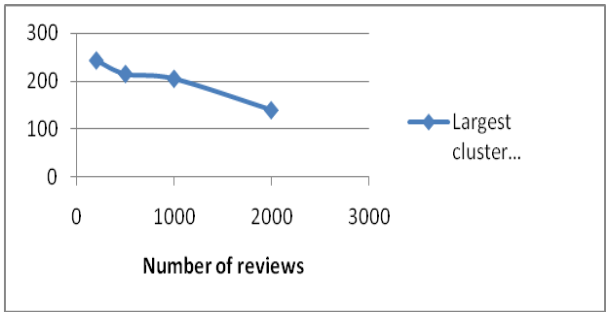

Figure 6: Number of reviews Vs cluster density

From Figure 6 it is observed that with the increase in a number of reviews the proportional decrease is observed in the density of clusters. Sparser clusters are resulted by larger reviews.

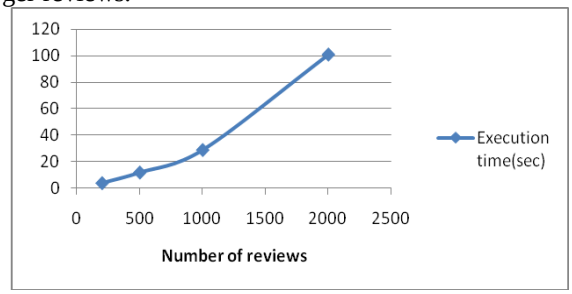

Figure 7: Number of reviews Vs execution time

From Figure 7 it was observed that with the increase in a number of reviews the proportional increase was observed in the execution time.

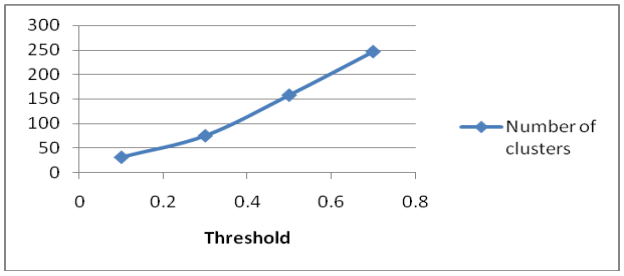

Figure 8: Threshold Vs number of clusters

From Figure 8 it was observed that with the increase in threshold value the proportional increased was observed in the number of clusters. More clusters are resulted by higher threshold values.

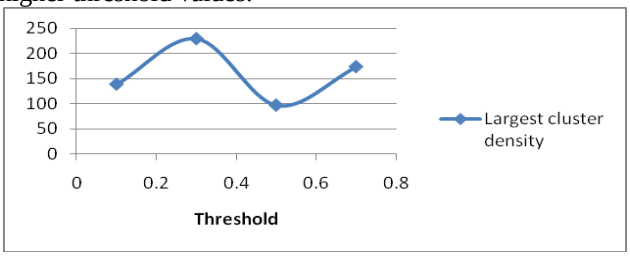

Figure 9: Threshold Vs cluster density

From Figure 9 it is observed that with the increase in threshold value the oscillatory nature is observed in the density of clusters.

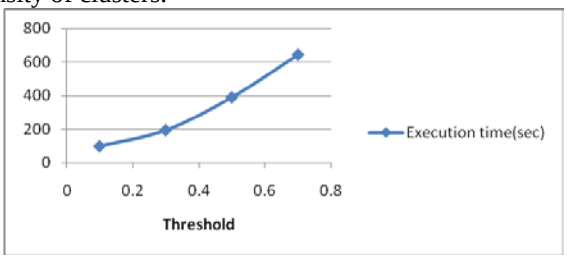

Figure10: Threshold Vs Execution time

From Figure 10 it is observed that with the increase in the threshold value the proportional increase is observed in the execution time.

From the sets of cases executed it is found that the more the number of customer reviews, the clearer the information from the clusters about the customers' taste. As the number of reviews increased, the number of clusters also increased and the density of individual clusters came down. For more reviews to be incorporated in computing weighs, the execution time increased accordingly.

The kinds of influence discussed above are useful for market analysis, and it is directly correlated with the number of customers that value a particular product. Although the techniques of reverse top-k queries are good alternatives for traditional clustering methods, they are acknowledged to sustain significant processing and I/O overhead. The reason is that a query typically requires finer execution of multiple top-k queries when computing the customers that prefer the queried product. It is observed with the cases experiments that that when an increase in k occurred the query execution time increases.

When combating with the Item-based collaborative filtering techniques used in the literature for product recommendations, particularly with the case of Wine quality study they share a similar insight, but in contrary to the methods used in this study. They suggest that customers have a taste of some products and thus rate them. In the cases of a recently launched product in the market or a product with initial stages of it designing during its manufacturing process the recommendation system is quite difficult to apply. In addition, the proposed framework needs no previous knowledge about users' opinions for the products. Alternatively, the preferences can be expressed in a more general way with the help of a weighting factor for each attribute of products, which is different from rating the individual products.

8. Conclusion

In this paper, a user-centric similarity framework is introduced in which the similarity of products is assessed by user preferences. A popular dataset named "Red wine quality "is considered in this work to assess the quality of Wine by grouping the individual products into clusters and then grade the groups based on preferences. The user-centric approach provided quite different and interesting results than the conventional approaches have, that do not consider the preferences that the customers have expressed. It is observed that the query types introduced in this work need higher execution times as the number of users and the preferences increased. This may lead to a scalability problem of the proposed framework. This can be smoothened by introducing R-tree like data structures for searching and indexing purpose that can optimize the execution time of the proposed framework. The purpose of developing this kind of a system is to support and advise wine users for better selection and winemakers for providing a better quality.

References

[1] Alejandra Urtubia, J Ricardo Perez-Correa, Alvaro Soto, and Philippe Pszcz´olkowski. "Using data mining techniques to predict industrial wine problem fermentations". Food Control, 18(12):1512–1517, 2007.

[2] Appalasamy P, A Mustapha, N. D. Rizal, F Johari, and A. F. Mansor. "Classification-based data mining approach for quality control in wine production". Journal of Applied Sciences, 12(6):598–601, 2012.

[3] Bernard Chen, Christopher Rhodes and Aaron Crawford, "Wineinformatics: Applying Data Mining on Wine Sensory Reviews Processed by the Computational Wine Wheel", 2014 IEEE International Conference on Data Mining Workshop.
[4] Charif Haydar, 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.
[5] Gongzhu Hu, Tan Xi, Faraz Mohammed, "Classification of Wine Quality with Imbalanced Data", 2016 IEEE.
[6] Hamid Parvin, Hoseinali Alizadeh, Behrouz Minati," A Modification on K-Nearest Neighbor Classifier" GJCST, Vol.10 Issue 14 (Ver.1.0) November 2010.
[7] Kavuri N. C. and Madhusree Kundu. "ART1 Network: Application in Wine Classification", International Journal of Chemical Engineering and Applications, Vol. 2, No. 3, June 2011.
[8] Kunal Thakkar et al, (IJCSIT),"AHP and Machine Learning Techniques for Wine Recommendation" International Journal of Computer Science and Information Technologies, Vol. 7 (5), 2016, 2349-2352.
[9] Linqi Song, Cem Tekin, Mihaela van der Schaar "Clustering Based Online Learning in Recommender Systems: A Bandit Approach", SongTekin ICASSP2014.
[10] Maria J Latorre, Carmen Garcia-Jares, Bernard Medina, and Carlos Herrero, "Pattern recognition analysis applied to classification of wines from Galicia (northwestern Spain) with the certified brand of origin", Journal of Agricultural and Food Chemistry, 42(7):1451–1455, 1994.
[11] Paritosh Nagarnaik, Prof. A. Thomas, "Survey on Recommendation System Methods", IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEM (ICECS 2015).
[12] Paulo Cortez, Antonio Cerdeira, Fernando Almeida, Telmo Matos, and Jos´e Reis. "Modeling wine preferences by data mining from physicochemical properties". Decision Support Systems, 47(4):547–553, 2009.
[13] S Kallithraka, IS Arvanitoyannis, P Kefalas, An El-Zajouli, E Soufleros, and E Psarra. "Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin". Food Chemistry, 73(4):501–514, 2001.
[14] Soude Fazeli, Hendrik Drachsler, Marlies Bitter-Rijpkema, Francis Brouns, Wim van der Vegt, and Peter B. Sloep, "User-centric Evaluation of Recommender Systems in Social Learning Platforms: Accuracy is Just the Tip of the Iceberg", IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES.
[15] 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).
[16] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, Zhi Wang, "Friendbook: A Semantic-based Friend Recommendation System for Social Networks", IEEE TRANSACTIONS ON MOBILE COMPUTING 2015.

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, 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 is Databases, Data Mining, Image processing, Intelligent Systems and Software Engineering