International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Market-Basket Analysis Using Agglomerative Hierarchical Approach for Clustering a Retail Items

Rujata Saraf[1], Prof. Sonal Patil[2]

[1]North Maharashtra University, G.H.Raisoni Institute of Engineering and Management, Shirsoli Road, Jalgaon 421001, India

[2]G.H.Rsisoni Institute of Engineering and management, North Maharashtra University, Shirsoli Road, Jalgaon 421001, India

Abstract: ***With the advent of data mining technology, cluster analysis of items is frequently done in supermarkets and in other large-scale retail sectors. Clustering of items has been a popular tool for identification of different groups of items where appropriate programs and techniques in data mining like Market-Basket analysis have been defined for each group separately with maximum effectiveness and return. For example, items frequently purchased together are placed in one place in the shelf of a retail store. There are various algorithms used for clustering. Hierarchical algorithms find successive clusters using previously established clusters. These algorithms usually are either agglomerative ("bottom-up") or divisive ("top-down"). The paper presents the Market-Basket Analysis using Agglomerative ("Bottom-up") hierarchical approach for clustering a retail items. Agglomerative hierarchical clustering creates a hierarchy of clusters which are represented in a tree structure called a Dendorogram. In agglomerative hierarchical clustering, dendrograms are developed based on the concept of 'distance' between the entities or, groups of entities.. The clustering will done in such a way that the Purpose of Market-Basket Analysis will achieve.***

Keywords: Retail Sectors, Market-Basket analysis, Hierarchical Clustering, Agglomerative Hierarchical Clustering, Dendrogram etc

1. Introduction

Customer relationship management (CRM) is the most important application in today's business world. The basic task of any type of business is to integrate 'relationship technology' with 'loyalty schemes'. By providing a different loyalty schemes, CRM is expected to enhance value to customers through raising satisfaction levels on transactions. If customers appreciate the value provided by a scheme in CRM, they are expected to continuously enhance the relationship with the firm through loyalty to the products/brands, purchasing more, advocating the firm to others, etc. In today's world, concept of Data Mining i.e Market-Basket analysis is widely used in retail stores to made the task of Planning such a schemes much easier and efficient. Retail stores hence consider to be a Best place for achieving market-basket analysis as it's a place where an ample number of different quality products, in different quantities with different rates are made available to customers. The items in retail stores are organized in proper and systematic manner so that an Individual can select the product from many options available and buy according to individual needs. And hence choosing the correct location for a particular product in retail store is a challenging task.

One such a illustration have been presented in this paper with the help of clustering of retail items for the purpose of Market-basket analysis. In today's world the places like Super-market, Malls are at the center point for any type of shopping. Super-market is a large form of the traditional grocery store and a self-service shop offering a wide variety of food and household products, organized into large shelfs. These places are consider to be most crowdie places where a huge mob of customers are find out in order to purchase different items as per their needs. Customers are giving preference to such a supermarket as the items in supermarket are arranged in proper and systematic manner on shelf. People can easily find whatever they want to purchase because of such a systematic arrangement of product on shelf. The shelf contains different varieties of single product on single shelf or the shelves have been partitioned into different section where each section can contain a particular product along with its varieties [6].

Generally in super-market such arrangements of product is done manually, i.e. lots of human resources are require to make such arrangement with which the customers can easily get whatever they want more efficiently. The products are arranged in such a way that the items which are purchased together are placed in one shelf beside to each other and by providing the different schemes on such a items, the total sales of the product have been increased [5]. But while doing so it's very difficult to predict which products should be kept beside to each other and in which product the customer will shown their interest. For this purpose it is necessary to find out which products are frequently purchased by customers from the total sale and by using this we can easily achieve the market-basket analysis by placing the most frequently purchased items besides to those items which are necessary along with the purchased item but not compulsory to purchased.

For this purpose the Data Mining Techniques like Mining algorithms and Clustering techniques are useful. With the help of mining algorithms, we can easily find out in which items the users are interested and which items are frequently purchased by customers. Similarly after finding such a items, Clustering techniques are useful to achieve the purpose of Market-Basket analysis. Several aspects of market basket analysis have been studied, such as using customer interest profile and interests on particular products for one-to-one marketing purchasing patterns in a multi-store environment to improve the sales[2][4]. But the existing technique related to clustering of such a retail items which can directly affect or

increase the total revenues of the market has some pitfalls while working. The proposed work in this paper will try to overcome the encountered drawbacks in existing system by using the Agglomerative Hierarchical clustering.

The complete paper is organized as per follows: Section 1 introduces the paper title in short concepts. Section 2 describes background for the topic including the details about Clustering and Various clustering Techniques in Data Mining with their respective advantages and disadvantages. Section 3 gives details regarding Agglomerative Hierarchical Clustering and the Dendrogram concept used in Agglomerative hierarchical clustering with its importance have been described. Section 4 gives the detail regarding complete working with cluster analysis, Section 5 is about some evaluation done while proposed work execution. Section 6 describes General result analysis for the proposed work. Section 7 finally conclude the complete topic with final result as clustering of retail items is necessary in retail stores as its directly proportional to the total sales and profit. And the Agglomerative Hierarchical Clustering is Best technique to achieve Market-basket analysis in retail stores**.**

2. Literature Survey

Data mining is vast area where a hidden knowledge from large databases have been extracted. Data mining analyses data in different perspective. It classifies the data and summarizing it into useful information. The results of the data mining can be used to increase the effectiveness of the performance of the user. For this type of analyzing purpose data mining uses a number of techniques[5]. Clustering is one the important functionality of the data mining. Clustering is an adaptive methodology in which objects are grouped together, based on the principle of optimizing the inside class similarity and minimizing the class-class similarity. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Cluster analysis groups objects (observations, events) based on the information found in the data describing the objects or their relationships. The aim is the objects in a group should be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the better the clustering. Cluster analysis can be used as a only data mining tool to gain insight into the data distribution, or as a preprocessing step for other data mining algorithms operating on the detected clusters.

A. Clustering Techniques:

While forming successive clusters, similarities and dissimilarities are based on the attribute values of an objects which describes it. It is an unsupervised learning and faces many challenges such as a high dimension of the dataset, arbitrary shapes of clusters, scalability, domain knowledge, ability to deal with noisy data and insensitivity to the order of input records. Large number of clustering methods [8] had been proposed till to address these challenges.

1) Partition Based Clustering: As the name suggest, in Partitioning clustering data gets divide into several subsets. The reason behind it is that to check all possible subset systems is computationally not feasible; there are certain greedy heuristics are used in the form of iterative optimization. Specifically, this means different relocation schemes that iteratively reassign points between the k clusters. For this purpose Relocation algorithms are used which gradually improve clusters.

2) Hierarchical Clustering: A hierarchical method creates a hierarchy of the given set of data objects. This hierarchical structure is represented by a tree structure where every cluster node contains child clusters, sibling clusters and the points generated after the partitioning of sibling clusters covered by their common parent. In hierarchical clustering each item is assigned to a particular cluster in such a way that if we have N items then we have N clusters[14]. It Finds the closest pair of clusters and merge them into single cluster. Compute distance between new cluster and each of old clusters. This method works on both bottom-up and top-down approaches. Based on the approach hierarchical clustering is further subdivided into agglomerative and divisive.

   The agglomerative hierarchical technique follows bottom-up Approach and begins with each element as a separate cluster and merge them into successively larger clusters. and hence these clustering also called as “Bottom-Up clustering”. Whereas Divisive hierarchical clustering follows the top-down approach. This method starts with a single cluster containing all objects, and then successively splits resulting clusters until only clusters of individual objects remain [11].

3) Density based Clustering: The density based method has been developed based on the true meaning of density that is the no of objects in the given cluster. The density based clustering basically developed to discover clusters with arbitrary shape,. The general idea is to continue growing the given cluster as long as the density in the neighborhood cluster exceeds some threshold; that is for each data point within a given cluster; the neighborhood of a given radius has to contain at least a minimum number of points[12].

4) Grid Based Clustering: As the name suggest, grid based clustering methods uses a multidimensional grid data structure. It divides the object space into a finite number of cells that form a grid structure on which all of the operations for clustering are performed. Grid based clustering maps the infinite number of data records in data streams to finite numbers of grids[10]. In this clustering the single uniform grid mesh is used to partition the entire problem domain into cells and the a set of statistical attributes will used to locate data objects within a cell. One of the distinct features of this method is the fast processing time, as it depends not on the number of data objects but only on the number of cells. The grid-based clustering algorithms are STING, Wave Cluster, and CLIQUE [12].

3. Proposed Work

3.1 Agglomerative Hierarchical Clustering

Agglomerative hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a Dendrogram[17]. A Dendrogram is a branching diagram that represents the relationships of similarity among a group of entities. The root of tree consists of a single cluster containing all observations, and the leaves correspond to individual observations. Such hierarchical algorithms may be conveniently broken down into two groups of methods. The first group is that of linkage methods – the single, complete, weighted and un-weighted average linkage methods [.These are methods for which a graph representation can be used. The second groups of hierarchical clustering methods are methods which allow the cluster centers to be specified.

The proposed work is based on the second type of agglomerative hierarchical clustering methods where cluster centers are specified to consider dissimilarities between two different points. It can be said as stored Dissimilarity based approach as it is an alternative to the general dissimilarity based algorithm for clustering. The stored data approach for the agglomerative hierarchical algorithm is as follows:

Step 1: Examine all inter point dissimilarities &form cluster from two closest points.
Step 2: Replace two points clustered by representative point or by cluster fragment
Step 3: Return to step 1, treating clusters as well as remaining objects, until all objects are in one cluster.

In steps 1 and 2, "point" refers either to objects or clusters, both of which are defined as vectors in the case of cluster center methods. This algorithm is justified by storage considerations, since we have O(n) storage required for n initial objects and O(n) storage for the n−1 (at most) clusters. While agglomerating the two closest point in one cluster, the inversion criteria of arbitrary points [5] must be taken into consideration. the inversion situation while hierarchy construction is explain as follows with an example.

For example: Consider the five arbitrary points as shown in figure 1

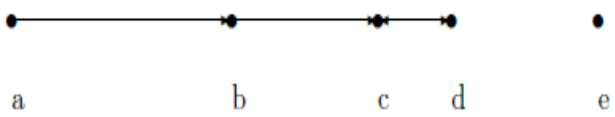

Figure 1: Five arbitrary Points with their respective nearest neighbors

A nearest Neighbor chain consists of an arbitrary point a in Fig 1 followed by its nearest neighbor b which is followed by the nearest neighbor from among the remaining points c, d, and e in Fig. 1 of this second point; and so on until we necessarily have some pair of points which can be termed reciprocal or mutual nearest neighbors. Such a pair of reciprocal nearest neighbors may be the first two points in the chain; and with assumption that no two dissimilarities are equal. While constructing a Nearest Neighbor chain, irrespective of the starting point, we may agglomerate a pair of Reciprocal nearest neighbors as soon as they are found as shown in figure 3(B). because there is no guarantees that whether we can arrive at the same hierarchy as if we used traditional "stored dissimilarities" or "stored data" algorithms

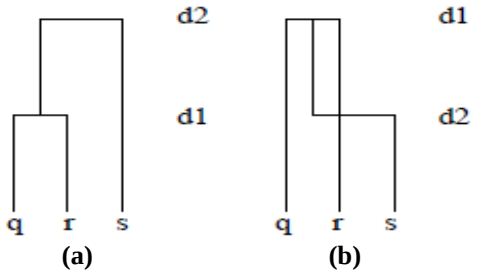

Figure 2: Hierarchy representation (a) without inversion and (b) with an inversion

Essentially this is the same condition as that under which no inversions (figure 2(a)) or reversals are produced by the clustering method. Fig.2 gives an example of this, where s is agglomerated at a lower criterion value (i.e. dissimilarity) than was the case at the previous agglomeration between q and r. Our ambient space has thus contracted because of the agglomeration. This is due to the algorithm used – in particular the agglomeration criterion – and it is something we would normally wish to avoid.

3.2. Dendrogram in Agglomerative Hierarchical Clustering:

The agglomerative hierarchical clustering build a cluster hierarchy that is commonly displayed as a tree diagram called a Dendrogram. They begin with each object in a separate cluster. At each step, the two clusters that are most similar are joined into a single new cluster. A Dendrogram is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to Individual observations. Any valid metric may be used as a measure of similarity between pairs of observations. The choice of which clusters to merge or split is determined by a linkage criterion [17] which is a function of the pair-wise distances between observations. Different clusters are obtained at different levels of the tree diagram of Dendrogram. This gives an opportunity to understand how multiple Levels of particular Dendrogram have been read to make sure that every cluster of that tree different from another one.

The process of Agglomerative Hierarchical Clustering (AHC) starts with the single observation clusters and progressively combines pairs of clusters, forming smaller numbers of clusters that contain more observations [17]. Then clusters successively merged until the desired cluster structure is obtained.

For Example: in fig 3., six elements a, b, c, d, e and f are shown in the Euclidean field. Clustering is to be done on the basis of Euclidean distance of similarity distance The most important distinction one should be consider while clustering is whether the clustering uses symmetric or asymmetric. The

distance functions used in the work have the property that distances are symmetric i.e. the distance from object A to B is the same as the distance from B to A. The first step is to determine which elements to merge in a cluster. Usually, we want to take the two closest elements, according to the chosen distance metric.

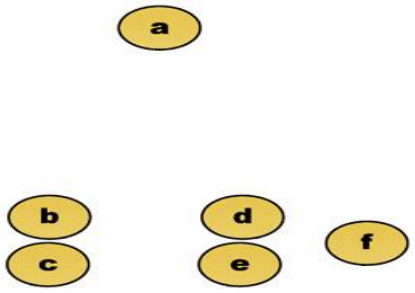

Figure 3: Objects for clustering in A Field

Based on Euclidean distance metric, the Dendrogram will be constructed. Dendrogram is a hierarchy of clusters represented in a tree like structure as shown in following figure. Dendrograms are made up of sub trees, and those sub trees, in turn, have sub trees nested within them. Each cluster of execution profiles in a dendrograms comprises a sub tree of the Dendrogram, and each sub tree or cluster has several attributes that can be examined and used in the refinement technique[2]. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to individual observations. Any valid metric may be used as a measure of similarity between pairs of observations. The choice of which clusters to merge or split is determined by a linkage criterion, which is a function of the pair-wise distances between observations. Different clusters are obtained at different levels of the tree diagram of Dendrogram. This gives an opportunity to compare the performance of various clustering in different levels with respect to a selected performance criterion. By examining the way in which executions are arranged in clusters and sub trees, their similarity to each other and to other clusters may be evaluated. The height of any sub tree in a Dendrogram indicates its similarity to other sub trees - the more similar two Executions or clusters are to each other, the further from the root their first common ancestor is.

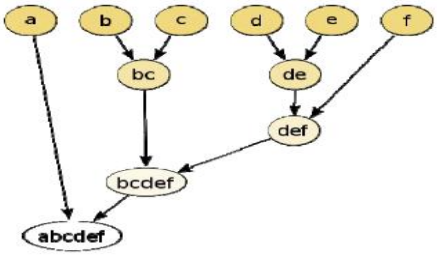

Figure 4: Hierarchical Clustering – Dendrogram

4. Implementation

4.1. Algorithm for Clustering:

For n samples, agglomerative algorithms begin with n clusters and each cluster contains a single sample or a point. Then two clusters will merge so that the similarity between them is the closest until the number of clusters becomes 1 or as specified by the user. The algorithm works as per following steps:

1. Start with n clusters, and a single sample indicates one cluster.
2. Find the most similar clusters Ci and Cj then merge them into one cluster.
3. Repeat step 2 until the number of cluster becomes one or as specified by the user.

The distances between each pair of clusters are computed to choose two clusters that have more similarity between them and hence to merge. There are several ways to calculate the distances between the clusters Ci and Cj, the Linkage method is one of such a method which is basically used to calculate the distance between two clusters which we have to merge together. All such a Linkage methods are described in table 3.

Table 1: Linkage Methods to calculate the association between two clusters

Single Linkage	$d_{12} = {}^{min}_{ij} d(X_i, Y_i)$	This is the distance between the closest members of the two clusters.
Complete Linkage	$d_{12} = {}^{max}_{ij} d(X_i, Y_i)$	This is the distance between the farthest apart members.
Average Linkage	$d_{12} = \frac{1}{kl} \sum_{i=1}^{k} \sum_{j=1}^{l} d(X_i, Y_j)$	This method involves looking at the distances between all pairs and averages all of these distances

Notation:
X1, X2,..., Xk = Observations from cluster 1
Y1, Y2,..., Yl = Observations from cluster 2
d (x,y) = Distance between a subject with observation vector x and a subject with observation vector y.

4.2. Working:

The proposed work complements the existing literature on the technique of association clustering. The proposed technique in this work considers only the positive associations for clustering. The detail working is explained as follows:

There are two steps in this technique. First step is mining association rules with threshold value of '1.00' for lift. Suitable threshold values for support and confidence are also chosen. In the second step, these association rules along with the support, confidence and lift are taken as input for clustering. Hence, this step gives an algorithm for developing the Dendrogram. The proposed algorithm for clustering makes use of 'all strength measures of association rule/s' instead of only 'support of item set'.

As discussed, there are two phases involved in the proposed clustering technique as mentioned below.

Phase I: Mine association rules from the transaction data with some threshold values of rule support and confidence with lift more than 1.00. The threshold values of rule support and confidence are chosen with low values so that all the items under consideration find place at least in one rule. Amongst

these rules, only those rules will mined which has lift value greater than 1 as main target is to mine positive association rule. The input and output in this step are as given below.

Input:
I = {I1, I2,……,Im} //set o items
D = {t1, t2 …tn} //a transaction database with ti as one transaction
Threshold rule support = s
Threshold rule confidence = c & Threshold lift = 1.00
Output:
A = { {r1, s1, c1, l1}, {r2, s2, c2, l2},……,{ rn, sn, cn, ln}}
// set of positive association rules (ri) with rule support (si), confidence (ci) and lift (li)
Phase-II: Make use of the output in phase-1 for developing the Dendrogram. Key steps involved in this phase are described below.
Step 1: Input to this phase is the output of phase-I, i.e., the set of association rules With the values of rule support, confidence and lift, given by set A.
Step 2: Obtain the set of individual items present in at least one of the association rules in A by
I = {I_1, I_2,……,Im}.
Step 3: Start with tree level s initiated as 1, the item set similarity is defined as very high value (tending to infinity), and number of clusters (item sets) is m (the total number of items in I). Hence, the set of clusters at level 1, L{1}, contains all 1-item clusters in I.
That is, $L_{\{1\}}$ = {{I_1}, {I_2},……,{Im}}.
Step 4: To generate a set of candidate item sets for next level ($C^{(s+1))}$ each pair of item sets in the previous level are joined.
Step 5: To evaluate item set similarity, i.e., similarity amongst the items in a cluster, each of the association rules is checked if all the items in the candidate set exist in the rule (either in the antecedent or in the consequent). If all the items exist in a rule and no other item is present in the rule, then sum up rule support, confidence and lift for the rule. Similarly, sums are obtained for all other rules where all the items are present in a rule. Sum of all such sums is taken as the measure of similarity.
Step 6: To generate $L^{(s+1)}$ (i.e., the set of item sets in level (s+1)), the two item sets are merged if their similarity is the highest value among all item sets in $c^{(s+1).}$
Hence, $L^{\{s+1\}}$ = { $L^{\{s\}}$ - $La^{\{s\}}$ - $Lb^{\{s\}}$} U {$La^{\{s\}}$ U $Lb^{\{s\}}$ }.
Step 7: The steps 4-6 are iterated with updating the Dendrogram (DE) by adding the topple < s, sim, k, $L^{\{s\}}$ > into DE.
Where s =s+1, sim = sim{$La^{\{s\},}$ $L_b^{\{s\}}$, k = k-1, $L^{\{s\}} = L^{\{s+1\}}$.
Iteration stops when there is no association rule with all items of any pair of combined clusters/item sets in a level and this level is the last level of clustering. Hence, all items may not be merged in one cluster as per the proposed algorithm in most of the cases.

5. Evaluation in Proposed Work

The strategy for using dendrograms to form a hierarchy of cluster has three phases, which are as follows:

1) Select the number of clusters into which the Dendrogram will be divided.
2) Examine the individual clusters for homogeneity by choosing the two executions in each cluster with maximum similarity according to the chosen similarity metric, and determine whether these two executions have the same cause. If the selected executions have the same or related causes, it is possible that all of the other failures in the cluster do as well.
3) Choose appropriate candidates for merging of clusters, according to the properties both the candidate having.

a) Split a cluster if it is found to be non-similar, and one or both of the resulting clusters would be a non-singleton or contain a largest homogeneous sub tree.
b) Merge two clusters if they are similar, which will be siblings, and if their failures have the same cause.

While doing so we have to examine the clusters having largest amount of internal dissimilarity first. Internal dissimilarity of particular cluster may be measured by calculating the average similarity between all individual execution profiles in the cluster. If it is found that the clusters having high internal dissimilarity are similar, then it is countable to assume that the others are as well, though it is still useful or necessary to examine clusters with more internal similarity. If there will be possible chances to split the clusters which are resulting from a splitting operation again, then in that case it may be advantageous to split the new cluster as well.

5.1. Evaluating Desired Characteristics of clustering:

While conducting many experiments during the implementation work, the desired characteristics of a clustering algorithm have been examines at different places. These characteristics are depends on a particular problem under consideration. The following is a list of characteristics

1) Scalability: Clustering techniques for large sets of data must be scalable, both in terms of speed and space. It is not unusual for a database to contain millions of records, and thus, any clustering algorithm used should have linear. Furthermore, clustering techniques for databases cannot assume that all the data will fit in main memory or that data elements can be randomly accessed. These algorithms are, likewise, infeasible for large data sets. Accessing data points sequentially and not being dependent on having all the data in main memory at once are important characteristics for scalability. This property have been used and tested with the data set used for the proposed work.
2) Effective means of evaluating the validity of clusters that are produced: It is common for clustering algorithms to produce clusters that are not “good” clusters when evaluated later. To check this whether the generated clusters are good or not, Validation for those clusters have been tested regularly.
3) The ability to find clusters in subspaces of the original space: Clusters often occupy a subspace of the full data space. Hence, the popularity of dimensionality reduction techniques is there. Many algorithms have difficulty in finding items to be keep in particular cluster in particular space, for example, a 5 dimensional cluster in a 10 dimensional space. The proposed work supports to the fact

that the entire cluster must be find in subspaces of the original space.

4) Ability to function in an incremental manner: In certain cases, e.g., data warehouses, the underlying data used for the original clustering which can change over time. If the clustering algorithm can incrementally handle the addition of new data or the deletion of old data, then this is usually much more efficient than re-running the algorithm on the new data set. The proposed works have an ability to handle the new data.

6. Result Analysis

The works regarding the topic have been proposed by using the online shopping scenario. And the expected results according to the proposed work are as follows:

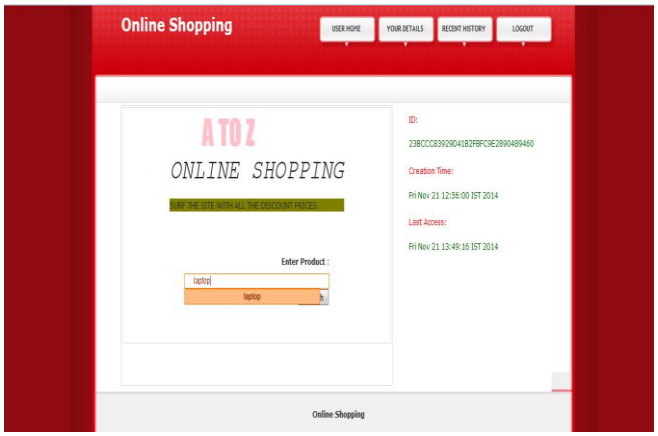

Figure 5: Searching a product (Laptop here)

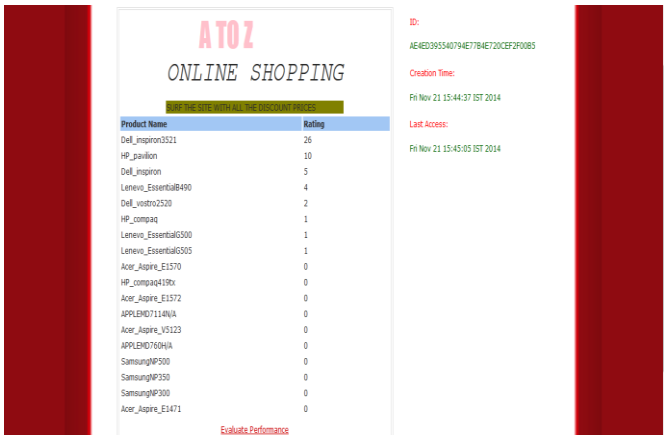

Figure 6: Suggetion Window for searching Product

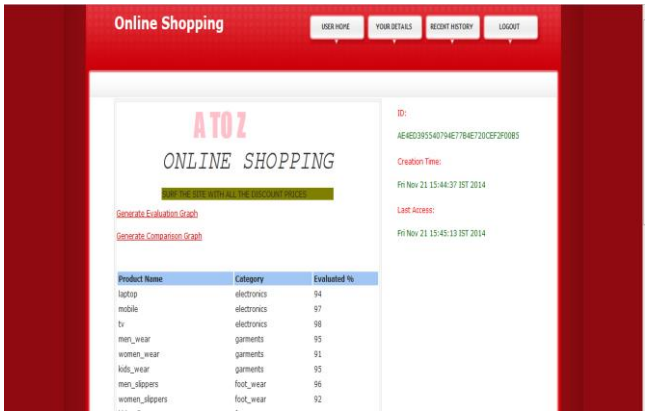

Figure 7: Performance Evaluation in%

Figure 8: General Result Analysis: Evaluation Graph

As shown in figure 5 very first the expected product will searched. As a result of this we will get all the product related to the searchable product along with their rating factors. While searching a particular product individual items in a database are consider as a separate cluster and then based on the distance criteria the most closest product related to the searching one are again put together for clustering purpose and hence we will get all the related products which are most closest to the searching one will get in next window i.e. in suggestion window as shown in fig 6. The performance of all the searched product is then evaluated in percentage value to achieve the market-basket analysis task based on the rating factor got in suggestion window as shown in fig 7. Figure 8 shows the graphical representation of general evaluation of the proposed work. In this diagram, the X-axis of a graph belongs to the no of items on which clustering have to be perform while Y-axis belongs to the efficiency of the proposed work in terms of some threshold values. The red line in graph represents the complete efficiency of a proposed work with respect to the items present in the data set. As shown in graph, for the 30 items in a data set the complete efficiency of a product is almost more than 90%. i.e. we can get at least 90% exact cluster by using the proposed work to achieve the Purpose of Market-Basket Analysis.

7. Conclusion

Market-basket analysis is an integral part of today's business world. Customer satisfaction is at the center point in Market-Basket Analysis and to achieve this it is necessary to find the main interest of customer in a particular product. The

agglomerative hierarchical clustering used in the proposed work creates the clusters by considering each item or product as a individual cluster from its starting with which the retailers can easily identify which products are frequently purchased by the customer from a huge dataset. The clustering of retail items with this technique gives more efficient and reliable result than other techniques as here clustering of item start from a individual element The placement of product in retail with the help of such clustering will not only effective and impressive but also helpful to achieve the goal of market-Basket Analysis. The technique presented is useful in the area of failure classification in retail stores or in supermarket, since the current failure classification methods do not have a definitive way to determine the number of clusters into which a set of program executions should be divided.

References

[1] Ashok Kumar D and Loraine Charlet Annie M.C, "Market Basket Analysis for a Supermarket Based on Frequent Itemset Minng", IJCSI, Vol. 9. Issue 5, No.3,September 2012.
[2] Aastha Joshi, Rajneet Kaur, "Comparative study of Clustering Techniques in Data mining", IJARCSSE, 2012.
[3] Berry, M.J.A., Linoff, G.S.: Data Mining Techniques: for Marketing, Sales and Customer Relationship Management (second edition), Hungry Minds Inc., 2004.
[4] "Cluster analysis" in http://en.wikipedia.org/ wiki/Cluster_Analysis
[5] Chen, Y.-L., Tang, K., Shen, R.-J., Hu, Y.-H.: "Market basket analysis in a multiple store environment, Decision Support Systems", 2004.
[6] Erik Buchmann, Leonardo Weiss Ferreira Chaves and Klemens Bohm, "Finding misplaced items in retail by clustering RFID data", EDBT 2010,March 22-26,2010, Lausanne, Switzerland.
[7] Er. Arpit Gupta 1,Er.Ankit Gupta 2,Er. Amit Mishra 3, " Research Paper On Cluster Techniques Of Data Variations", International Journal of Advance Technology & Engineering Research (IJATER).
[8] Fionn Murtagh and Pedro Contreras, "Methods of Hierarchical Clustering",arXiv:1105.021v1 [CS:IR], 30th April 2011.
[9] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.
[10] Kalyani M Raval, " Data Mining Techniques", IJARCSE, Volume 2, Issue 10, October 2012
[11] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta, "A Comparative Study of Various Clustering Algorithms in Data Mining",International Journal of Engineering Reserch and Applications (IJERA), Vol. 2, Issue 3, pp.1379-1384, 2012
[12] MR ILANGO, Dr V MOHAN, "A Survey of Grid Based Clustering Algorithms", International Journal of Engineering Science and Technology, pp.3441-3446, 2010.
[13] "Measuring Association d12 Between Clusters 1 and 2" in http://www.stat.psu.edu/online/courses/stat505/18_cluste r/0 5_cluster_between.html
[14] Neha Soni, Amit Ganatra, — "Categorization of Several Clustering Algorithms from Different Perspective: A Reviewl", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2,no.8,pp.63-68,Aug. 2012.
[15] P. Berkhin. (2001) — "Survey of Clustering Data Mining Techniques" [Online]. Available:http://www.accure.com/products/rp_cluster_re view.pdf
[16] Rui Xu, Donald C. Wunsch II, — "Survey of Clustering Algorithmsl", IEEE Transactions on neural Networks, vol. 16, pp. 645-678, May 2005.
[17] Rahmat Widia Sembiring, JAsni Mohamad Zain, Abdullah Embong, "A Comparative Agglomeartive Hierarchical Clustering Method to Cluster Implemented Course", journal of Computing, volume 2,Issue 12, December 2010.
[18] S.M. Savaresi and D. Boley, "On the performance of bisecting K-means and PDDP",. Proc. SIAM Data Mining Conf, 2001.
[19] W. Bishop, "Documenting the value of merchandising. Technical report, National Association for Retail Merchandising Service", 2000.

Author Profile

Ms. Rujata Saraf pursuing M.E. Degree in Computer Science and Engineering from North Maharashtra University, Jalgaon and received the B.E., degree from Mumbai Univ. in 2011. After degree worked as an assistant professor (from January 2012) in the Dept. of Information Technology, the G.H.Raisoni Institute of Engineering and Management, Jalgaon., and now working as a Lecturer (from January 2015) at ST. Francis Institute of Technology, Borivali in Mumbai University

Prof. Sonal P. Patil received the Diploma and B.E.. degrees, from MSBTE and NMU Univ. in 2005 and 2008, respectively. She received the M.Tech degree from Bhopal Univ. in 2013. She have been Appreciated as Best Outgoing Student during Diploma & Degree College. Also Appreciated by North Maharashtra University, Jalgaon for active involvement and management of YUVARANG 2009. She worked as an assistant professor (from 2009 to 2014) in the Dept. of Information Technology, in G.H.Raisoni Institute of Engineering and Management, Jalgaon and now (from July 2014) working as a HOD of Information Technology department in Same college. She has successfully published book having a title "Computer Organization" for Second Year CSE & IT Students of Engineering in January 2014. Her research interest includes Data Mining and its Application. She is a member of CSI, CAP process of North Maharashtra University from Dec 2010 Exam, Syllabus setting committee member for the Computer & I.T. department subjects and Certified & lifetime member of ISTE.