International Journal of Research
e-ISSN: 2348-6848 & p-ISSN 2348-795X Vol-5, Special Issue-10
International Conference on Innovations in Information and Communication Technology
Held on 9th and 10th March 2018 organized by The Department of Information Technology,Meenakshi Sundararajan Engineering College Kodambakkam,Chennai, India

Application of Density Based Clustering Algorithm in Pharmacy

Veeramani.S AP/CSE[1], Raj Kumar[2], Narendra Prasanth[3]
[1,2,3] *Dr.Mahalingam College of Engineering and Technology*

Abstract:

At present in the pharmacies may or may not have stock details and also due to current hike in the prices of medicines because of GST, people are unaware of the uses and also the prices of medicines. So people has to go to the pharmacy directly to know the status of the medicines. Time, money are being wasted. This applications enable users to view the pricelist, medicines in stock and location of pharmacies. If users search required medicines, this application returns pharmacies with optimal total cost (possibly low) and distance are returned as output.

The Scheme "Application of Density Based Clustering algorithm in pharmacy" using dynamic clustering of data with DBSCAN algorithm. The algorithm can be used in such a way that it takes the input and first searches in the available datasets. But it cannot be applies to large datasets. So we go for DBSCAN algorithm which is known density based spatial clustering algorithm.

Here the similar available datasets have been grouped and pre-processed before searching and the remaining datasets have been termed as noise. Pharmacies the location and range are taken as input and also the cost of each medicine available in it. These grouped datasets will be searched for the feasible solution.

The application of DBSCAN algorithm works for the overall feasible solutions in a particular area and if not enough, it searches the next cluster which is termed as the noise. Here the optimal solution based on distance of shop and cost of medicine have been obtained by using this algorithm. The DBSCAN algorithm ranks pharmacies based on distance and the additional filter is applied to rank pharmacies based on price constraint. Hence, the people could not be exploited as they know the best price medicine available in the nearest shop

.

1. Introduction:

Data clustering is defined as the group of similar information or datasets from a whole database. This data clustering technique is only suitable for small datasets. But basic level of data clustering techniques are not at all suitable for the current growing technologies. So advanced data clustering techniques have been introduced which includes k-means, k-medioids, hierarchical and also partial clustering algorithms. Clustering techniques are used in different fields like image analysis, pattern recognition, knowledge discovery and bio-informatics. Application of clustering techniques in spatial databases possess the challenge to discover clusters with input parameters of algorithm with minimal requirements of domain knowledge and a good efficiency on large databases.

If we take k-means clustering algorithm it requires a large number of datasets as input and also it uses its special kernel function using which the datasets have been grouped and searched. But this kernel function is not much efficient. It has time and space complexity of O (n^2) and hence not suitable for large datasets. Hierarchical clustering algorithms have been used so that the clusters have been represented in hierarchical structure named dendogram. CURE is an improved version of single-link which selects a random sample of points and shrinks them towards the required input. BIRCH agglomerative hierarchical clustering algorithm which uses a tree based representation for reducing time complexity but it can find only spherical shaped compact clusters and also clustering result is effected by input order of data. Multiview based algorithm obtains improvement in clustering and also it yields high accuracy. The Hierarchical methods have a time complexity of O (n^3). The user provided data may be incorporated into clustering algorithm towards a better solution. Semi-supervised learning methods includes labelled data which obtains improved clusters.

2. Literature Review:

G. Nagy, State of the art in pattern recognition[1], it is a branch of machine learning that focuses on the recognition of patterns and regularities in data, although it is in some cases considered to be nearly synonymous with machine learning. Algorithmic learning and adaptation is facilitated by accurate statistics gleaned from large samples in the case of symbolic patterns, and by skilled human judgment in the case of natural patterns. Recent technological advances like pocket computers, camera phones and wireless networks will have greater influence on mobile, distributed, interactive recognition of natural patterns than on conventional high-volume applications like mail sorting , check reading or forms process.

Pattern recognition S.Theodoridis, K.Koutroumbas**[**2**],** it is based on identifying the correlated patterns which are similar to one another. Pattern recognition is in the centre of a number of application areas, including image analysis, speech and audio recognition, biometrics, bioinformatics, data mining, and information retrieval. Despite their differences, these areas share, to a large extent, a corpus of techniques that can be used in extracting, from the available data, information related to data categories, important "hidden" patterns, and trends.

Efficient Clustering Algorithm by S. Guha, R. Rastogi, K. Shim[3], Clustering in data mining is used for discovering groups and identifying interesting distributions in underlying data. Traditional clustering algorithms either favour clusters with spherical shapes and similar sizes or are very fragile in the presence of outliers .We propose a new clustering algorithm called CURE. That is more robust outliers and identifies clusters having non spherical shapes and wide variances and size.

BIRCH: Efficient Clustering Algorithm by Z. Tian, R. Raghu[4], Finding useful patterns in large data sets has attracted considerable interest recently and one of the most widely studied problems in this area is the identification of clusters. This paper presents the data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchy) and demonstrates especially suitable for very large database.

Semisupervised Learning By O.Chapelle, B.Schölkopf ,A.Zien,[5] We examine data clustering which is a particular type of data mining problem. Large set of multi dimensional data points the data space is usually not uniformly occupied. Data clustering identifies the sparse and crowded places and hence discovers the overall distribution patterns of the data set.

3. Proposed Work:

Density based clustering algorithm with noise is known as the proposed system DBSCAN algorithm. In Improved DBSCAN algorithm the cluster is taken as input along with the noise, so that the output is obtained as the optimal solution and not as the feasible solutions. Here the first step is that based on the epsilon distance the required medicine is searched in the cluster, if the solution is not optimal it searches on to the next cluster which is regarded as the noise. This algorithm is very efficient when regarded to all other clustering algorithm both in space and time complexity. Density Based Spatial Clustering of Applications with Noise (DBSCAN) can discover clusters of arbitrary shape and also handles outliers effectively. DBSCAN obtains clusters by finding the number of points within the specified distance from a given point. It involves computing distances from given point to all other points in the dataset.

The performance of DBSCAN degrades considerably with noise due to unnecessary distance computations introduced by noise points while the proposed method is robust to noise by pruning out noise points early and eliminating the unnecessary distance computations. The cluster results produced by our method are exactly similar to that of DBSCAN but executed at a much faster pace. Density Based Spatial Clustering of Applications with Noise (DBSCAN) is the pioneer of density based clustering techniques which can discover clusters of arbitrary shape and also handles noise or outliers effectively. DBSCAN algorithm has a quadratic time complexity with dataset size. The algorithm can be extended to large datasets by reducing its time complexity using spatial index structures. But, they cannot be applied for high dimensional datasets. In this paper we propose an algorithm which has groups to

accelerate the neighbour search queries. Groups method is efficient in handling large amounts of noise present in the data.

3.1 DBSCAN – A Density Based Approach:

DBSCAN algorithm defines cluster as a region of densely connected points separated by regions of non-dense points. If similarity measure is taken as Euclidean distance the region is a hyper sphere of radius eps at the given point p as centre.

1. Eps-neighbourhood: for a point x AD, the Eps-neighbourhood denotes the set of points whose distance from x is less than or equal to eps. The cardinality of Eps-neighbourhood defines the threshold density of x.

2. Esp.-connected: for a pair of points x; y AD, if x->y reps from the view of a DBSCAN method every point in the dataset will fall into either core point or border point. Further a border point can be either noise point or density connected point.

3. Core point: A point with threshold density greater than or equal to minpts.

4. Border point: A point with threshold density less than minpts.

5. Noise point: A point p is a noise point if the threshold density of p is less than minpts and all points in the Eps-neighbourhood of p are border points.

6. Density connected point: A Border point with at least one core point in its Eps neighbourhood.

This method ensures that always for a given pattern the neighbour searching does not need to move points farther than distance of 5. Eps, which is an advantage over conventional. DBSCAN that requires searching all patterns in the dataset. It is observed from experimental analysis that inappropriate parameter values are neglected and the similar elements which are available in the particular cluster and also some solutions from noise have been obtained. Group's method is more stable than index-based structures as it do not any require specific input parameters from user to build the groups index structure. Also, Groups method is robust to noise by pruning outliers early with zero or few distance computations.

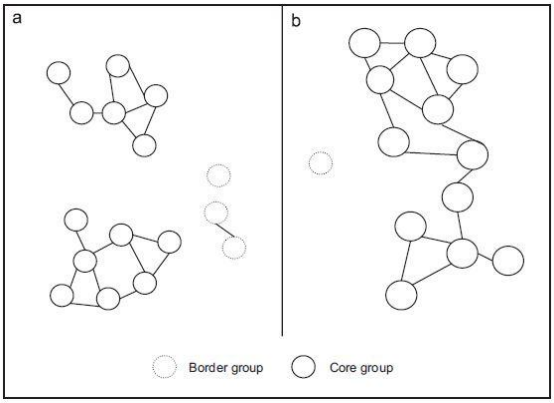


Figure 1. Core group formation using DBSCAN Algorithm

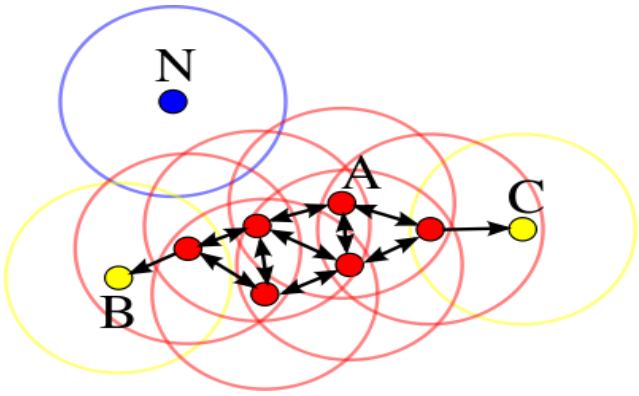
3.2 DBSCAN Algorithm:

Recommender systems are effective to identify items that could interest clients on e-commerce web sites or predict evaluations that people could give to items such as movies. In this context, clustering can be used to improve predictions or to reduce computational time. In this paper, we present a cluster- ing approach based on item metadata information's. Evaluations are clustered according to item genre. As items can have several genres, evaluations can be placed in several clusters. Each cluster provides its own rating prediction and weighting strategies are then used to combine these results in one evaluation.

$$R(i,j) = k\ K\ w\ k\ R\ k\ (i,j)/k\ K\ w\ k\ .$$

Average: local predictions $R\ k\ (i, j)$ are computed for each active genre and are then used to compute a standard average (N is the number of active genres):

$R\ (i, j) = 1\ N\ \ k\ \ K/R\ k\ (i, j)$.



1. Figure 2. Pattern Matching based on DBSCAN Algorithm.

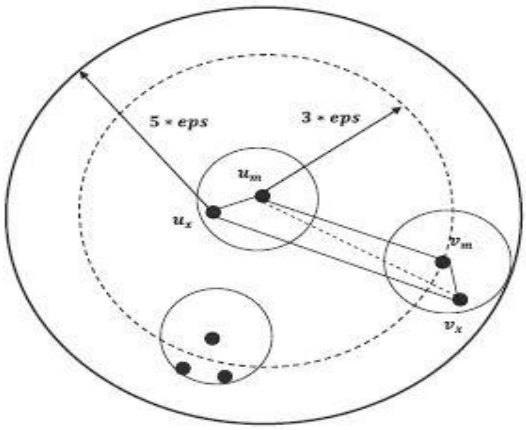


Figure 3. Minimum Distance pattern in DBSCAN Algorithm.

3.3 Advantages:

1. Groups method ensures that always for a given pattern the neighbour searching does not need to move points farther than distance of 5. Eps, which is an advantage over conventional.

2. DBSCAN that requires searching all patterns in the dataset.

3. It is observed from analysis that inappropriate parameter values for hierarchical index construction gives a performance degradation of up to two to three fold magnitude of actual running time.

4. Group's method is more stable than index-based structures as it do not any require specific input parameters from user to build the groups index structure. Also, Groups method is robust to noise by pruning outliers early with zero or few distance computations.

Figure 4. Block Diagram of DBSCAN Algorithm in Pharmacy

4. Conclusion:

Application based on Density Based Clustering algorithm (DBSCAN) was implemented. It was implemented in Java and each data are stored in Sqlyog database. Each data stored in database are gathered from various pharmacies. The DBSCAN provides a mechanism by which each shops located in optimal distance will be

formed into a cluster where user can get optimal distance, price of each medicine through the application we developed. If the shop or medicine doesn't fit into a cluster the improved DBSCAN will increases the range of search. Each range or distance can be provided by the user based on their requirement. Each distance (latitude and longitude about shop location). It helps to reduce the time required for user to search medicine. This method helps rural people where they can easily search for medicines without travelling for long distances.

5. References:

[I] R.C. Gonzalez, R.E. Woods, Digital Image Processing, third ed., Pearson Prentice-Hall, Upper Saddle River, NJ, 2015

[II] S. Theodoridis, K. Koutroumbas, Pattern Recognition, second ed., Academic Press, New York, 2016.

[III] U.M. Fayyad, G.P. Shapiro, P. Smyth, R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, MIT Press, Boston, MA, 2015.

[IV] S. Madeira, A. Oliveira, Bi-clustering algorithms for biological data analysis: a survey, IEEE/ACM Trans. Comp. Biol. Bioinformatics. 1 (1) 24–45, 2015

[V] R.H. Gueting, An introduction to spatial database systems, VLDB J. 3 (4) 357–399 2015

[VI] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognition. Lett. 31 (8) 651–666, 2016

[VII] M. Girolami, Mercer kernel-based clustering in feature space, IEEE Trans. Neural Network. 13 (3) 780–784, 2015

[VII] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Computer. Survey. 31 (3) 264–323, 2015

[IX] B. King, Step-wise clustering procedures, J. Am. Stat. Assoc. 62 (317) 86–101, 2016

[X] G. Nagy, State of the art in pattern recognition, in: Proceedings of IEEE 56, 2015, pp. 836–862.