Middle-East Journal of Scientific Research 23 (12): 2878-2887, 2015
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2015.23.12.22775

Intelligent Decisive Support System for Disease Prediction with Multi Attribute Relational Deapthness Clustering of Breast Cancer Data Set Using Data Mining

[1]D. Banumathy and [2]S. Selvarajan

[1]CSE at Paavai Engineering College, Namakkal, Tamilnadu, India
[2]Principal at Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

Abstract: The problem of high dimensional clustering has been studied in different situations, but the approaches suffer with the problem of false indexing and the implication of clustering can be applied in many problems. We propose a novel approach to support decisive systems and perform disease prediction based on relational depthness measure. The proposed approach uses a multi attribute relational depthness clustering, where the relation between the data points of any cluster is computed based on the relative similarity between various attributes of high dimensional data set. We compute the multi attribute relational depthness measure which represents the relative measure and the bond the data point has with the data points of the cluster. Based on computed multi attribute relational depthness measure used to identify the class of any data point. The proposed method improves the quality of cluster being generated and reduces the false indexing ratio. To perform disease prediction, the same multi attribute depthness similarity is computed and a set of probability values is computed for each of the disease class available. The proposed method improves the performance of decisive support system and reduces the time complexity also.

Key words: Decisive Support System · Disease Prediction · Multi Attribute Relational Depthness Measure · Data Mining · Clustering

INTRODUCTION

Data Mining is the process of identifying the related information according to the input query from a large set of data sets. The process of data mining can be applied in medical solution where the problem is to group the related medical records into different categories. The identified categories and the medical records can be used to perform many tasks like identifying similar medical records which could be used for analysis.

Intelligent decisive support systems are one which supports the medical practitioners to identify or predict the possible disease according to the medical record provided. To provide support for the decisive systems the related medical records have to be kept in a proper manner. The decisive system may use any measure to perform disease prediction, but the efficiency of the decisive system is highly based on the clustering of records.

Clustering is the process of grouping related or similar records of medical information based on some similarity. The similarity of the records may be computed based on any measure like Euclidean distance. The application of Euclidean distance can be applied only for the limited dimensional data points. But when the dimension of the data point grows, the application of Euclidean distance like measure cannot be applied and needs some other quality measure. Such quality measure has to be identified and formed which supports clustering to be performed in the most effective manner.

The relational measure is one, which represents the bonding or similarity of any two data points according to one to one correspondence. The data point may contain an N number of attributes, but among them a few attribute values are more closure which represent that those data points are related according to that particular property. While grouping high dimensional large data set, it is not necessary that all the properties are

Corresponding Author: D. Banumathy, CSE at Paavai Engineering College, Namakkal, Tamilnadu, India

similar but the most has to be similar because some of the properties may be common between different group of data clusters.

For example, when we consider about the symptoms like temperature, esinophile, body pain and cold are common symptoms in the diseases like normal fever and typhoid, but some may be different for the typhoid and it will have stomach pain also. Like this in order to group the data points into a category, there must be most similarity of data properties but not all. The relational measure is one which represents such closure, can be used to cluster high dimensional data sets.

Related Works: There are many approaches has been discussed for the development of disease prediction and decisive support systems with clustering and data mining. We discuss a few of them here and discuss them in detail.

Clustering, high dimensional data using density measures has been discussed in [1], where the clusters are considered as dense regions. The method represents the noisy values present in the borders. The method also provides a detailed review of earlier methods which has worked based on density measures and compared their efficiency in various aspects in this paper.

A graph based subset selection approach has been discussed in [2], to cluster high dimensional data. The method identifies the most impact features of each class based on which the subset of features is formed. The method describes that the features belong to any class is independent. The evaluation results shows that the method has a higher probability in identifying the subset and produces high quality results.

The problem of high dimensional clustering has been handled using the concept lattice and concept feature vector in [3]. The efficiency of the concept lattice tool which has more impact in data analysis is enforced in this paper to solve the high dimensional problem. The method uses the intent and extent of concepts and represent them in a hierarchical manner in the form of tree nodes. Also the method improves the efficiency of high dimensional sparse data clustering.

Enhanced Mining of High Dimensional Data using Efficient Clustering Algorithm [4] proposes a novel approach called supervised attribute clustering algorithm is proposed to improve the accuracy and check the probability of the patterns. In this method, faster retrieval of relevant data is made more efficient and accurate. By using this method, users can get precise results and negligible data loss. This method displays results based on the high probability density, thereby providing privacy for data and reducing the dimensionality of the data.

The problem of curse dimensionality has been studied in [5], which considers the low dimensional features and uses the popular k-means clustering approach and assigns various data points in neighbor list and computes the distance between them. To improve the running time of clustering methods and to improve the consistency of the clustering methods the method proposes an unsupervised parameter selection approach to choose the number of clusters and the dimension of the data points and many more. The method has produced promising results, but suffers with the efficient selection of dimension.

Identifying the subset features which are relevant in clustering has been reviewed and proposed in [7], to support high dimensional clustering. In this, the method removes the irrelevant features by mapping from available features which are already listed. In this method the feature selection is performed based on the threshold value mentioned in the particular feature. Based on selected features, a graph has been generated which is clustered using a minimum span tree algorithm at the later stage. The subset feature decides the location of the node in the tree and to assign a class to the data point.

In Relevant Clustering Algorithm for High-Dimensional Data [11], the method eliminates irrelevant features from the dataset; the relevant features are selected by the features having the value greater than the predefined threshold. In the second step selected relevant features are used to generate the graph, divide the features using graph theoretic method and then clusters are formed by using Minimum Spanning Tree. In the third step find the subsets features that are more related to the target class is selected.

An agent based approach for disease prediction has been discussed in [19], which uses multiple agents to monitor the patients and for surveillance. Similarly, the agents are used to provide various services like educational to perform chronic disease management. This approach is more applicable where the patient monitoring is more essential and the data management is required in huge importance. The agents are useful in collecting the medical data and maintaining them in an efficient manner at all the sessions. The agents were helpful in generating alerts to the medical peoples about the conditions of the patient through alert messages, notifications. Also the same can be used in prediction of diseases in many ways.

Application of an artificial immune system based intelligent multi agent model to a mine detection problem has been discussed in [20]. The method uses two approaches; one to monitor the behavior of the human system and produces alerts to the controlling system and the second one is acts as an evolutionary system to control the human anatomy. The method has produced efficient results in all the factors of artificial immune systems.

In [21], the author discussed a intelligent heart disease prediction method called IHDPS which works over data mining techniques. The method adapted decision tree, neural networks and naïve bayes approaches. The method uses various user profiles like sex, age, pressure, sugar to perform disease prediction. The method produces significant results in the heart disease prediction.

All the above discussed approaches has the problem of false prediction and less prediction accuracy which affects the efficiency of the system.

Multi Attribute Relational Deapthness Clustering Based Decisive Support System: The need of support decisive systems has been solved with more stragic solution, an multi attribute relational depthness measure based clustering approach. The method computes relational measure according to the depthness of relation a data point has with other data points of the cluster. The proposed approach has various stages namely, preprocessing, multi attribute relational depthness measure computation, relational depthness clustering and disease prediction.

Need for DP-MARD Clustering: To support intelligent decision making system which performs disease prediction from the set of symptoms and their values provided, the researchers have to consider many factors according to the dimension of data point given. Let the data set Ds has N data points, each has H dimension, to perform decision making the data points has to grouped into a G number of groups where G specifies the number of clusters. The first challenge to the researcher is to perform efficient clustering and on the other hand the disease prediction has to be performed efficient manner. While considering the first problem, the researcher has to formulate exact measure which measures the similarity of data points of any cluster Ci. On the other way, the process of disease prediction must be met with a small set of features or symptoms, because there may be an S number of symptoms the user may have but the researcher has to find out the most important symptom to perform disease prediction.

Also, to perform prediction the number of data points and the number of symptoms with their values is more important. The method cannot conclude that the presence of a small set of symptom is prone for any disease, so it requires a more strategic approach. There are few methods for the prediction of disease which works based on a small set of dimension discussed in agent based disease prediction []. The problem of dimensionality is not handled well and the prediction accuracy is poor in those approaches. This motivates us to design and develop an efficient methodology to perform clustering, high dimensional data set and to develop a supportive system for disease prediction.

- The better method has to use small set of symptoms to perform disease prediction.
- The methods must produce efficient clustering accuracy and should handle high dimensional problem.
- The false positive and false negative ratio has to be reduced.

Overview of DP-MARD: The above discussed problem of high dimensional clustering is handled in proposed disease prediction using multi attribute relational depthness (DP-MARD) based clustering. The functional architecture of the DP-MARD is illustrated in Figure 1. The DP-MARD has different functional components and each has different dedicated jobs to be performed. The input data set is being preprocessed to remove the noisy records which have no value or missing value or missing attribute by identifying the total dimension of the data set. At the next stage, the preprocessed data points are converted into a feature vector. The relational depthness clustering is performed by identifying the number of relations gets matched with the data points of each cluster. Based on the depthness measure the data points are assigned to any of the categories being identified. Finally, using the same multi attribute depthness similarity a set of symptom is identified as influencing the breast cancer.

Introduction to DP-MARD Clustering: In this section, we first introduce the structure of data points and the cluster.

Then we define a set of relations considered and discuss how it has been visualized to the problem of high dimensional clustering. Also, we discuss about how this can be adapted for the problem of disease prediction.

Structure of Data Point: The data points consider has various symptoms and each has a different type of values and their values also in different range. For example, Consider a data point Dp which has K number of dimensions and has N number of attribute value the we can formulate the data point as follows:

Dp={{F1,A1,V1, {F2,A1,V2}, {F3,A2,V3}, {F4,A3,V4}, ,…….,{Fk,An,Vk}}.

From the above data point Dp, the value of F represents the feature name, A represent the type of attribute and V represent the value of the attribute A.

As discussed above the dimension of the data point has no limit and the size of data set also has no limit. If the value of K is unlimited then the problem of clustering has to consider and the clustering approach has to use the K features to perform clustering.

If we consider there exist C clusters, that the data points are clustered into an M number of clusters then the data points of any cluster Ci has similar features and values in the most. For example the data point has the dimension value 10 then the data points located in any cluster will have similarity between them in a minimum of M-1 attribute(i.e 9 attribute). It is not necessary that the attribute similarity and the data point similarity should be less than the value of M but is can be equal. Also the deviation cannot be more than certain value which named as threshold value. The threshold value decides the level of false ratio and also decides the accuracy of clustering.

$$Similarity = \frac{\sum_{i-1}^{k} \sum Distance(Attribute(Ai(dp)) - Attribute(Ai(dp1))) < Th}{Number\ of\ data\ points\ of\ cluster} \quad (1)$$

The equation (1) shows how the similarity value of data point towards any cluster can be computed.

Based on computed multi attribute relational similarity the problem of disease prediction can be computed. We propose such strategic method to perform high dimensional clustering and support disease prediction.

In the proposed approach, at the training phase, the input data is preprocessed where the features of the data point are extracted. Using the extracted features, the method computes the multi attribute relational depthness measure and cluster the data point using relation depthness clustering. At the testing phase, the features of the data point are extracted and compute the multi attribute relational depthness measure. Based on the relational depthness measure , the data point is assigned with a class.

The Figure 1, shows the architecture of the proposed system and the functional components of the proposed approach.

Feature Extraction: At the feature extraction stage, the input data set is read and the dimension of the data set is being identified. We identify the attributes of the data points and compute the number of data points present in the data set also identified. The features of each data point are extracted to form as a feature vector and will be used to perform clustering.

Feature Extraction Algorithm:
Input: Data Set Ds.
Output: Feature Set Fs.

Step 1: Start
Step 2: Initialize Attribute Set As.
Step 3: for each data point Dp of Ds

As = $\Sigma Asi(As) \exists As$
End

Step 4: Compute size of data set Sds = $\Sigma\ dp \in Ds$
Step 5: For each data point Dpi from Ds

Extract Features from Dpi.

$$Fs = \sum_{i=1}^{size(Dp)} \sum Fs + \sum_{j=1}^{size(Dpi)} DPi(j)$$

End

Step 6: Stop.

The above discussed preprocessing algorithm identifies the attributes set of data points and computes the total number of points present in the data set. Then the set of features is identified and added to the feature set.

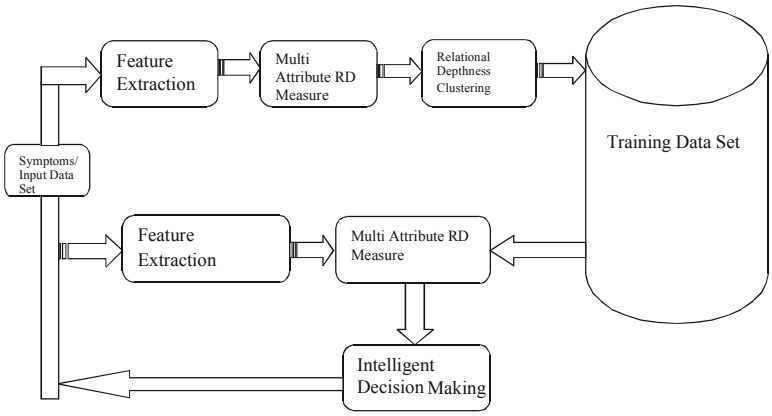


Fig. 1: Architecture of the proposed approach

Relational Depthness Clustering: The proposed method performs clustering of data points based on the relational depthness measure. The method maintains set of clusters and initialize them with random seed values and using the input data set, for each data point, with the available data points of the cluster, we compute the relational depthness similarity measure which is computed towards various attributes of the data points. For each attribute the similarity is computed based on the tolerance value maintained by the proposed relational depthness clustering.

Relational Depthness Clustering Algorithm:
Input: Feature Set Fs, Cluster Cs.
Output: Cluster Cs.

Step 1: Start
Step 2: For each cluster Cs_i from Cs
Initialize Cs_i with random points.
Cs_i = Random(Fs).
End.
For each feature vector Fv_i from Fs
compute multi attribute relational depthness measure.
for each cluster Cs_i
MARD = MultiAttributeDepthnessMeasure(Fvi)
End
End
Choose the most valued MARD measure and assign the class label.
Cs_i(fvi) = Max(MARD).

Step 3: Stop.

The relational depthness clustering algorithm compute the multi attribute relational depthness measure for each of the cluster and the data points present in them. Based on computed measure the cluster with maximum depthness measure is identified.

Multi Attribute Relational Depthness Measure: The multi attribute relational depthness clustering represents the depthness of relation or bond exist between any two or the data points of the cluster. For each data point of the cluster and for each of the attribute value of data points of the cluster, we compute the depthness of relation of different attributes with the input feature point. IF the feature value of the data point and the feature vector are within the closure tolerance then we can say they have a relation in that particular dimension. This will be performed for each of the dimensions and the depthness of relation is measured.
Multi Attribute Relational Depthness Measure Algorithm:
Input: Feature Vector Fvi
Output: MARD

Step 1: Start
Step 2: Initialize relational counter rc, relational depthness set Rds.
Step 2: For each cluster Ci from Cs
For each data point Dpi
For each feature Fi
Compute feature closure $Fc = \int_{i=1}^{size(Fi)} Dist(Fi(Ai) - Dpi(Ai))$
If Fc< FCThreshold then
Rc= Rc+1;
End
End
Add to relational depthness set Rds.
Rds = Σrdi (Rds) + Rc
End

Compute multi Attribute relational depthness measure MARD.

$$\text{MARD} = \frac{\sum RD_S}{size(Rds)}$$

End

Step 3: stop.

The above discussed algorithm computes the multi attribute relational depthness measure for the given feature vector to the each of the data points present in all the clusters identified.

Intelligent Decision Making: At this stage, the disease prediction is performed by the proposed approach, given N number of symptoms and their values, the probability of each disease consider being computed. The probability value shows how much the patent is close to the disease. To perform such a prediction with the given input set of symptoms, the method identifies the set of input symptoms given and the value mentioned for each of them. The given feature is formed as feature vector and used to compute the multi attribute relational depthness measure. The MARD measure shows the closeness of the feature vector with the feature set present in each class of disease, which is computed at all the dimensions. Once the multi attribute depthness measure has been computed, then the disease probability could be computed to identify the disease class.

Intelligent Decision Making Algorithm:
Input: Symptom Set SS.
Output: Disease Probability

Step 1: Start
Step 2: Read symptoms SS.
Step 3: Convert SS into feature vector Fv.
Step 3: For each disease class Dc
compute MARD = Multi Attribute Relational Depthness (Fv).

End
Step 4: For each disease class compute probability
Disease probability Dp = N×Log(MARD).
N- number of data points of class
End

Step 5: Choose most probable disease class Dc.
Step 6: Stop.

The disease prediction algorithm computes multi attribute relational depthness measure for each of the disease class identified. Using the computed MARD value, with the number of data points present in the cluster, the probability of disease for each class is computed. From the computed probability value the maximum possible disease class is identified.

RESULTS AND DISCUSSION

The proposed Multi attribute relational depthness measure based high dimensional clustering for decisive support systems has been implemented using Matlab with various data sets. The approach has been validated for its clustering efficiency using various data sets. The method has been tested with training set of 80 percent data and the remaining is used as a test set. The proposed method produced efficient results with various data sets. We have evaluated the algorithm with the following data sets.

The Wisconsin data set has been published by UCI machine leaning repository and contains 569 samples of each with 32 dimensions. The attribute values are multivariate Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. The samples of the data set are classified into two classes like malignant and benign.

Similarly the “Broad”, data set published by “Broad Institute of Cancer”, has 75 samples with 38 dimensions or attributes. It has two classes of samples and the data attributes are of two types namely alphanumeric and numeric. The proposed method has been evaluated with the Broad data set where 70 percent of them has been used as training and the rest is taken as test samples.

The BCSC dataset has 2,39,2998 samples of mammogram, which are named as index mammograms. The data set is being provided by the “Breast cancer surveillance consortium”. The index mammogram is performed for the woman who does not undergo any screening in the past nine months. The data set provides samples of patients who has been undergoing screening in the recent years are used.

Experimental analysis is performed using all the three data sets and the clustering efficiency has been measured by splitting the data set into training and testing set. The analysis is performed with 70 percent of training set with 30 percent as testing set. Also the analysis is performed by varying the size of test set and training set. The details of clustering accuracy produced by the analysis has been furnished below.

Table1: Comparative analysis of test cases

	Hubness	Relavant	Fast Clustering	DP-MARD	Hubness	Relavant	Fast Clustering	DP-MARD
Data set	Clustering accuracy at 80/20 Test Case				Clustering accuracy at 90/10 Test case			
Wisconsin	78.67	82.32	86.67	98.76	80.76	84.12	88.97	99.23
BROAD	81.12	84.34	87.97	98.56	83.43	86.76	89.06	99.43
BCSC	84.34	86.45	87.56	99.26	85.76	87.89	89.97	99.76

The Table 1 , shows the result of various test cases performed on measuring clustering accuracy using different data sets. The result shows that therac proposed approach has produced efficient clustering accuracy at all the test suite provided.

Clustering Efficiency: The clustering efficiency is computed based on the number of samples has been indexed exactly to the same class and the number of samples which has not been indexed correctly.

$$\text{Clustering Efficiency } CF = \frac{True\ Positive}{Total\ Number\ of\ points} \times 100$$

True Positive Tp = Number of data points have been indexed exactly to the same class which is considered for training and testing.

The method has been evaluated for the efficiency of the cluster being produced and compared with the results produced by different methods. The analysis informed that the proposed multi attribute relational depthness based high dimensional clustering has produced a higher efficient cluster than the other methods considered.

Frequency of Overlapping: The overlapping frequency of any clustering algorithm can be computed, by computing the number of samples which has been assigned more than one label from given input sample N.

$$\text{Frequency of overlapping } Fo = \frac{Number\ of\ samples\ assigned\ with\ multiiple\ classes}{Total\ Number\ of\ input\ samples} \times 100$$

The quality of clustering produced by any clustering algorithm can be measured according to the overlapping factor. The overlapping is the factor that the same data point gets two names which is meaningless and the algorithm specifies that the data point is belong to different clusters. Such overlapping factor has been computed to evaluate the performance of different algorithms on the data set considered earlier. The results strike that the proposed multi attribute relational depthness based clustering approach has produced less overlapping ratio than the other methods.

False Indexing Ratio: The False indexing ratio is computed by identifying the number of test samples which has been assigned a wrong class name at the testing phase.

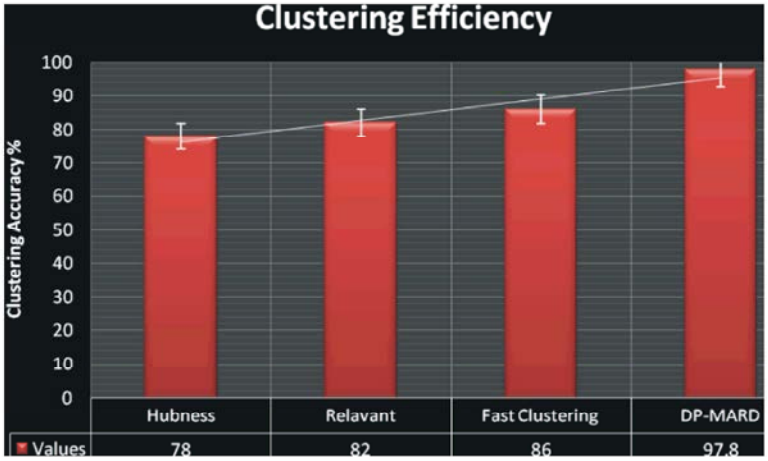
False Positive value Fp = Number of input samples assigned to false class.

$$\text{False positive ratio Fpr} = \frac{Fp}{Total\ number\ of\ sample} \times 100$$

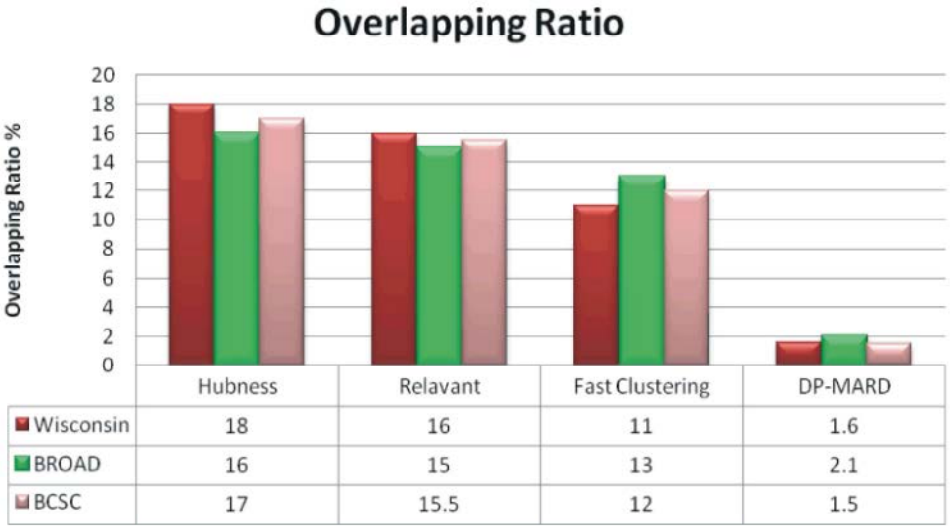
Another quality factor which implicates the efficiency of any clustering algorithm is a false Indexing ratio. Here we perform such analysis of the false indexing ratio produced by different methods and it shows clearly that the proposed method has produced efficient clustering without producing any false indexing.

Disease Prediction accuracy can be compute by providing an N number of symptoms set and computing the number of times the algorithm classifies the symptom set to a specific disease class.

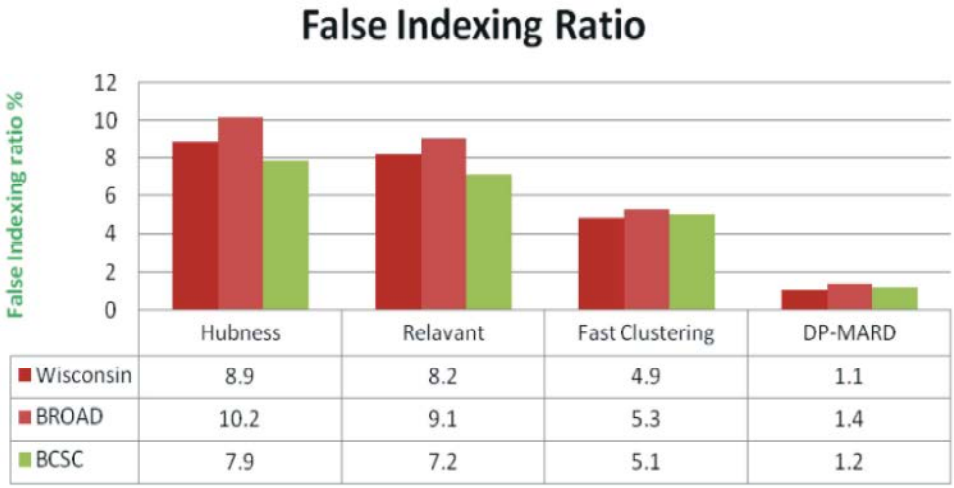
$$\text{Disease Prediction Accuracy DPA} = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ symptom\ set} \times 100$$



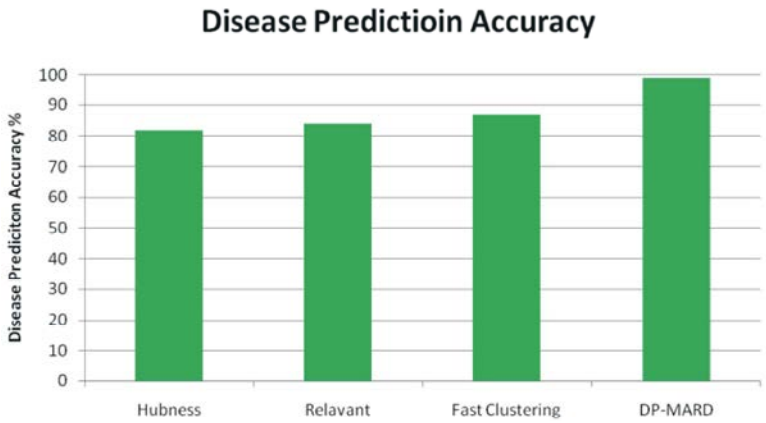
Graph 1: Comparative Analysis of Clustering Efficiency



Graph 2: Comparative analysis of overlapping



Graph 3: Comparative study on false indexing ratio



Graph 4: Comparison of disease prediction accuracy

Table 2: Comparative results of disease prediction accuracy

Method Name	Wisconsin	BROAD	BCSC
Hubness	72.6	74.2	78.3
Relevant	76	78	79
Fast Clustering	77.3	78.6	79.4
DP-MARD	94	95	99.7

The Graph 4, shows the comparison result of disease prediction accuracy produced by different methods and it shows clearly that the proposed method has produced more prediction accuracy than other methods.

The Table 2, shows the comparative analysis performed on disease prediction produced by different methods on all the data sets considered. The proposed multi attribute relational depthness based clustering and disease prediction approach has produced efficient results than other methods.

The efficiency of disease prediction is directly proportional to the number of dimensions and the number of samples. The proposed method could produce more efficiency with the high dimensional data set.

CONCLUSION

An multi attribute relational depthness clustering based intelligent decisive support system has been proposed. The method identifies the set of features of the data points and identifies the dimension using the data file and the data points. With the given data set, for each data point we compute multi attribute relational depthness closure with all the data points of each cluster. Based on a computed MARD measure, we compute the disease probability of each class and then a single class of disease is selected and its probability has been used. The proposed method has produced efficient results compared to others and has reduced the time complexity also. The false indexing ratio is hugely reduced using the proposed approach. Further the clustering accuracy can be improved by adapting different similarity measure which consideres more number of dimensions to measure the relevancy of data points.

REFERENCES

1. Sunita, Jahirabadkar and Parag Kulkarni, 2013. Article: Clustering for High Dimensional Data: Density based Subspace Clustering Algorithms. International Journal of Computer Applications, 63(20): 29-35.
2. Song, Q., J. Ni and G. Wang, 2011. A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data, IEEE Transactions on Knowledge and Data Engineering.
3. High dimensional sparse data Clustering Algorithm Based on Concept Feature Vector (CABOCFV), IEEE Transaction on Service Operations and Logistics and Informatics, 1, Page(s): 202-206, 2008.
4. Enhanced Mining of High Dimensional Data using Efficient Clustering Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering, 3(12), December 2013.
5. The Role of Hubness in Clustering High-Dimensional Data, IEEE Transactions on Knowledge and Data Engineering, Revised January 2013.
6. Muller, E., S. Gunnemann, I. Assent and T. Seidl, 2009. “Evaluating clustering in subspace projections of high dimensional data,”Proceedings of the VLDB Endowment, 2: 1270-1281.
7. Toma¡sev, N. and D. Mladeni´c, 2012. “Nearest neighbor voting in high dimensional data: Learning from past occurrences,” Computer Science and Information Systems, 9(2): 691-712.

8. Toma¡sev, N., R. Brehar, D. Mladeni´c and S. Nedevschi, 2011. "The influence of hubness on nearest-neighbor methods in object recognition," in Proc. 7th IEEE Int. Conf. on Intelligent Computer Communication and Processing (ICCP), pp: 367-374.
9. Buza, K., A. Nanopoulos and L. Schmidt-Thieme, 2011. "INSIGHT: Efficient and effective instance selection for time-series classification," in Proc. 15th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD), Part II, pp: 149-160.
10. Mehta, R.G., N.J. Mistry and M. Raghuwanshi, 2014. Article: Towards Unsupervised and Consistent High Dimensional Data Clustering. International Journal of Computer Applications, 87(2): 40-44.
11. Bini Tofflin, R., 2014. A Relevant Clustering Algorithm for High- Dimensional Data , International Journal of Innovative Research in Computer and Communication Engineering, 2(1).
12. Kab´an, A., 2012. "Non-parametric detection of meaningless distances in high dimensional data," Statistics and Computing, 22(2): 375-385.
13. Ning, K., H. Ng, S. Srihari, H. Leong and A. Nesvizhskii, 2010. "Examination of the relationship between essential genes in PPI network and hub proteins in reverse nearest neighbor topology," BMC Bioinformatics, 11: 1-14.
14. Chang, C.T., J.Z.C. Lai and M.D. Jeng, 2010. "Fast agglomerative clustering using information of k-nearest neighbors," Pattern Recognition, 43(12): 3958-3968.
15. Toma¡sev, N., M. Radovanovi´c, D. Mladeni´c and M. Ivanovi´c, 2011. "Hubness-based fuzzy measures for high-dimensional knearest neighbor classification," in Proc. 7th Int. Conf. on Machine Learning and Data Mining (MLDM), pp: 16-30.
16. "A probabilistic approach to nearest-neighbor classification: Naive hubness bayesian kNN," in Proc. 20th ACM Int. Conf. on Information and Knowledge Management (CIKM), pp: 2173-2176.
17. Radovanovi´c, M., A. Nanopoulos and M. Ivanovi´c, 2010. "Timeseries classification in many intrinsic dimensions," in Proc. 10th SIAM Int. Conf. on Data Mining (SDM), pp: 677-688.
18. Toma¡sev, N., 2010. "Hubs in space: Popular nearest neighbors in highdimensional data," Journal of Machine Learning Research, 11: 2487-2531.
19. Koutkias, V.G., 2005. A multiagent system enhancing home-care health services for chronic disease Management.
20. Sathyananth, S., 2002. Application of artificial immune system based intelligent multi agent model to a mine detection problem.
21. Palaniappan, Intelligent heart disease prediction system using data mining techniques, AICCSA, 2008.
22. Bennett, K.P., 1992. "Decision Tree Construction Via Linear Programming." Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp: 97-101.