Survey on Patient Analysis using Incremental Semi-Supervised Algorithm

R.HariShankar[1], G.Keerthana[2], A.Kirithika[3], V.Krishna Kumar[4]
[1, 2, 3, 4] Department of Computer Science and Engineering
[1, 2, 3, 4] Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India

Abstract-*Diabetes, which is referred as lifestyle disease, has become very common nowadays. Nearly 29.1 million people have diabetes today. There are type 1 and type 2 diabetes and these are caused due to various reasons like insulin level, body weight, genes, environmental conditions. To reduce the rate of diabetic patients we collect the dataset of patients , analyses the dataset and predict the chances of diabetics.*

To make this as efficient, incremental semi-supervised algorithm is implemented. This has a major advantage of using the random subspace technique and also uses the incremental member selection process. This random subspace technique is effectively used for handling high dimensional data and the incremental ensemble member selection is used to reduce redundant data based on the local cost function and global cost function. Incremental semi-supervised algorithm compares each patient dataset with normal level and also with other patients and predicts the chance for diabetics. Various parameters that causes diabetes are taken and they are analyzed, then the algorithm is implemented to predict the chance for diabetes.

Keywords-Random Subspace, Incremental Semi-Supervised algorithm.

I. INTRODUCTION

Predictive analytics is a part of data mining, which forecasts on extracting information from the dataset to predict the patterns and future outcomes. It also practices on what might happen in the future with acceptable level of reliability and also risk assessments.

Cluster ensemble approaches are gaining more and more attention recently. Due to its useful applications in the areas of , bioinformatics, pattern recognition, data mining, and also in other areas. When cluster ensemble approaches are compared with traditional single clustering algorithms, they are able to integrate multiple clustering solutions obtained from different data sources into a unified solution, and provide an accurate result which is more robust and stable. However, conventional cluster ensemble approaches have several limitations [1] They do not consider how to make use of prior knowledge by experts, which is given as pair wise constraints [2] All the ensemble members do not contribute to the final result [3] Not all the ensemble members achieve satisfactory results on high dimensional datasets. Inorder to reduce all these limitations , we go with the random subspace based semi-supervised clustering ensemble framework. This is used to integrate the random subspace technique and to perform data clustering. The incremental semi-supervised clustering ensemble is used to remove redundant members.

II. METHODS AND MATERIAL

Literature Survey:

Alexander Strehl and Joydeep Ghoshet al[6] , Clustering ensembles have emerged as a powerful method for improving both the robustness as well as the stability of unsupervised classification solutions. However, finding a consensus clustering from multiple partitions is a difficult problem that can be approached from graph-based, combinatorial, or statistical perspectives. This study extends previous research on clustering ensembles in several respects. First, we introduce a unified representation for multiple clustering and formulate the corresponding categorical clustering problem. Second, we propose a probabilistic model of consensus using a finite mixture of multinomial distributions in a space of clustering.

Natthakan Iam-On, Tossapon Boongoen, Simon Garrett, and Chris Price et al[7] , This paper explains the much of previous attention on decision trees focuses on the splitting criteria and optimization of tree sizes. The dilemma between over fitting and achieving maximum accuracy is seldom resolved. A method to construct a decision tree based classifier is proposed that maintains highest accuracy on training data and improves on generalization accuracy as it grows in complexity.

Albert T .Corbett and John R .Anderson et al[8],This paper presents a novel pair wise constraint propagation approach by decomposing the challenging constraint propagation problem into a set of independent semi-supervised classification subproblems which can be solved in quadratic time using label propagation based on k-nearest neighbor graphs. Considering that this time cost is

proportional to the number of all possible pairwise constraints, our approach actually provides an efficient solution for exhaustively propagating pairwise constraints throughout the entire dataset .

Javad Azimi, Monireh Abdoos, and Morteza Analoui et al[9] , We propose a novel approach for solving the perceptual grouping problem in vision. Rather than focusing on local features and their consistencies in the image data, our approach aims at extracting the global impression of an image. We treat image segmentation as a graph partitioning problem and propose a novel global criterion, the normalized cut, for segmenting the graph. The normalized cut criterion measures both the total dissimilarity between the different groups as well as the total similarity within the groups.

The image segmentation problem is concerned with partitioning an image into multiple regions according to some homogeneity criterion. This article is primarily concerned with graph theoretic approaches to image segmentation. Segmentation-based object categorization can be viewed as a specific case of spectral clustering applied to image segmentation. We show that an efficient computational technique based on a generalized eigenvalue problem can be used to optimize this criterion. We applied this approach to segmenting static images, as well as motion sequences, and found the results to be very encouraging.

III. RESULTS AND DISCUSSION

From this paper, The image segmentation problem is concerned with partitioning an image into multiple regions according to some homogeneity criterion. This article is primarily concerned with graph theoretic approaches to image segmentation. Segmentation-based object categorization can be viewed as a specific case of spectral clustering applied to image segmentation. We show that an efficient computational technique based on a generalized eigenvalue problem can be used to optimize this criterion. We applied this approach to segmenting static images, as well as motion sequences, and found the results to be very encouraging.

III. CONCLUSION

It aims to predict the chances of diabetes by implementing the incremental semi-supervised algorithm by cost effective and thus providing accurate results.

REFERENCES

[1] Zhiwen Yu, Senior member,IEEE, Peinan Luo, Jane You, Hau-San Wong, Jun Zhang, Senior member, IEEE "Incremental Semi-Supervised Clustering Ensemble for High Dimensional Data Clustering"

[2] Simon Tong Daphne Koller,Computer Science and Department of Camford University Camphell N Christianini, A.Simona "Support vector machine active learning with application to text classification" International Conference of machine learning 2000

[3] Natthakan Iam-On, Tossapon Boongeon, Simon Gareet and Chris Price "A Link-Baesed Cluster ensemble Approach for Categorial Data Clustering" IEEE Transactions on knowledge and Data Engineering, VOL 24,NO 3,March 2012

[4] Shashikant Ghumbre, CChetan Patil and Ashok Ghatol "Heart Disease Diagnosis using Support Vector machine" International Conference on Computer Science and Information Technology(ICCSIT'2011)Pattaya Dec. 2011

[5] P.Yasodha, M.Kannan "Analysis of a population of Diabetic Patients Databases in Weka tool" International Journal of Science And Engineering May 2011

[6] Alexander Strehl and Joydeep Ghosh "Cluster Ensembles–A Knowledge Reuse Framework for Combining Partitionings " Department of Electrical and Computer Engineering The University of Texas at Austin

[7] Natthakan Iam-On, Tossapon Boongoen, Simon Garrett, and Chris Price " A Link-Based Cluster Ensemble Approach for Categorical Data Clustering " IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 3, MARCH 2012

[8] Albert T. Corbett and John R. Anderson"Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge "Human Computer Interaction Institute,Carnegie Mellon University, USA

[9] Javad Azimi, Monireh Abdoos, and Morteza Analoui " A New Efficient Approach in Clustering Ensembles" Computer Engineering Department-Iran University of Science and Technology, Tehran, Iran