Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.8-10
ISSN: XXX-XXXX

Classification for Disorder Forecast from Epigenetic Data Statistics

T.Sukumar[1], V.BabyDeepa[2]
[1]Research Scholar, [2]Assistant Professor,
[1, 2] Department of Computer Science, Government Arts College, Karur, Tamilnadu, India
Email: sukumarmsg@gmail.com, deepamct@gmail.com

ABSTRACT - Epigenetic is the field of study arose at the moment when it was discovered that some changes in the genotype of living organisms are not directly related to the DNA structure and its changes. Epigenetic is the science that studies changes in genes without changing the inherited structure of the DNA (during reproduction) as well as the influence of external factors on the level expression of genes. At present times, it is required to develop an effective technique for proper prediction of diseases from epigenetic information. This paper presents a proficient feature selection based classification model to predict the disease by the use of highly related genes (i.e. considered as features) linked to it. For feature selection, oppositional based social spider optimization (OBSSO) algorithm is applied, which increases the search space exploration and the important regions in a determined neighbourhood. Then, logistic regression (LR) makes use of extracted features and performs classification.

I. INTRODUCTION

The cluster analysis is one of the part of data mining, and is used very often, not only in computer science, but also in many other fields. The problem of cluster analysis may be used in many areas such as economics, life science, and many other fields where is needed division analysed objects into group. In the life sciences cluster analysis can be used, for example to divide patients according to some specific factors, such as blood tests, gene expression levels, etc. The goal of cluster analysis is to classify objects from given data set into groups (called clusters). Finally each object from data set should be appropriate assigned to exactly one cluster. Unlike a classification [1], [2] which is also one of data mining methods, groups obtained during clustering proces are unknown before the grouping task is performed. The problem of cluster analysis is a relatively complex problem, because division into clusters is not always strict. To get expected classification appropriate selection of methods is necessary. Due to the fact, that cluster analysis is an often-used in data analysis and exploration, there are a lot of methods for clustering.

II. GENETIC ALGORITHMS

Genetic algorithms [14][15][16][17] are group of algorithms inspired by mechanisms occurring in inheritance and evolution processes. A genetic algorithm operates on a specific population of possible solutions. The population consist of group of individuals. Each individual of the population is one of possible solutions of the problem solved by using genetic algorithm and should be created randomly. Apart from that, it is necessary to decide how to represent (encoding) individuals - as it will have a significant influence on whether we will succeed in finding the best solution. There are different ways of encoding individuals, and the choice of how to encode individuals depends on the problem that has to be solved. Each individual is characterized by its own genotype (which is also a specific solution of the problem). Due to their specificity and aspiration to obtain the well-adjusted individuals, GAs are usually applied for optimisation purposes.

III.EPIGENETICS

The term Epigenetic first surfaced on the printed page, researchers, physicians, and others poked around in the dark crevices of the gene, trying to untangle the clues that suggested gene function could be altered by more than just changes in sequence. Today, a wide variety of illnesses, behaviours, and other health indicators already have some level of evidence linking them with epigenetic mechanisms, including cancers of almost all types,

Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.8-10
ISSN: XXX-XXXX

cognitive dysfunction, and respiratory, cardiovascular, reproductive, autoimmune, and neurobehavioral illnesses. Known or suspected drivers behind epigenetic processes include many agents, including heavy metals, pesticides, diesel exhaust, tobacco smoke, polycyclic aromatic hydrocarbons, hormones, radioactivity, viruses, bacteria, and basic nutrients.

IV.EXPERIMENTS

This section presents the results of experiments evaluating the effectiveness of the presented genetic algorithm. Six datasets were used to test the effectiveness of the proposed algorithm. The description of the data is presented in TABLE I. TABLE I contains name of data set, size of data set (number of rows), dimension (number of column taken to clustering), and number of clusters (in all data sets number of clusters and division into groups are known). The visualisations of data with the division into the correct groups are shown in Fig. 1.

Table 1: Description of Data Sets Used in Experiments

Data Set	Description	Size of data set	Number of dimensions	Number of clusters	Number of objects in clusters
Dataset 1	Mouse Data Set	490	2	3	290
					100
					100
Dataset 2	Artificially Generated Data Set	500	3	3	170
					171
					159
Dataset 3	Iris Data Set	150	4	3	50
					50
					50
Dataset 4	Flame Data Set	238	2	2	150
					88
Dataset 5	Jain Data Set	373	2	2	97
					276
Dataset 6	Cassini Data Set	700	2	3	280
					280
					140

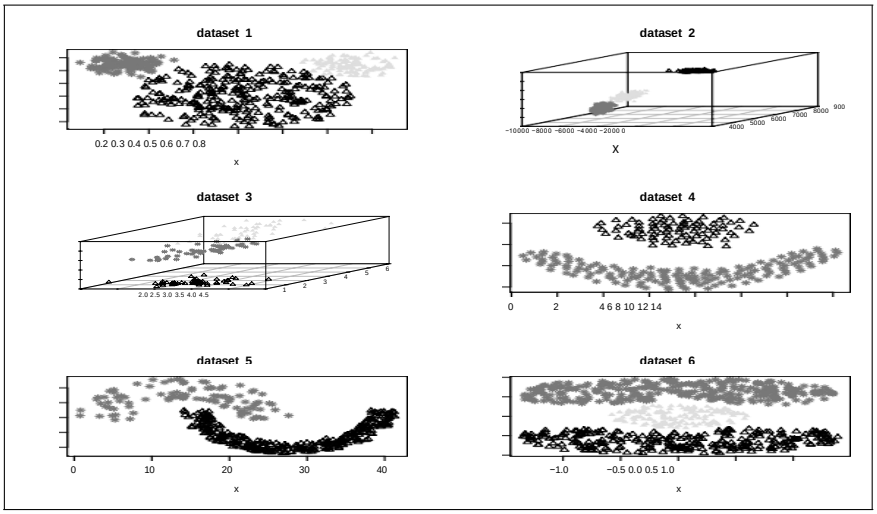

Figure 1: Visualisation of Data Sets Used in Experiments

Journal of Information Technology and Society
Volume: 01 Issue: 01 June 2021 Page No.8-10
ISSN: XXX-XXXX

V. CONCLUSION

In this article, the Epigenetic Grouping Genetic Algorithm (Epi GGA) for clustering data was presented. Based on the presented experiments, it was proved that the proposed algorithm achieves a very good results in the problem of data clustering. For the data sets used in the experiments, the quality of classification was higher for the proposed algorithm than the commonly used methods. Thus, the proposed algorithm can be an effective alternative to the commonly used methods for data clustering.

REFERENCES

[1] T. Amin and Igor C. and M. Moshkov and B. Zielosko, Classifiers Based on Optimal Decision Rules, Fundam. Inform., Vol. 127, p. 151-160

[2] I. Chikalov and M. Moshkov and B. Zielosko, Online Learning Algorithm for Ensemble of Decision Rules, Rough Sets, Fuzzy Sets, Data Mining and Granular Computing - 13th International Conference, Springer, 2011, p. 310-313

[3] A. Hartigan and M. A. Wong ”A K-Means Clustering Algorithm”, Applied Statistics, Vol. 28, No. 1, 1979, p. 100-108

[4] D. Pelleg, A. Moore, X-means: Extending K-means with Efficient Estimation of the Number of Clusters, Proceedings of the 17th International Conf. on Machine Learning, 2000, p. 727-734

[5] M. Ester, H. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996, p. 226-231

[6] L. Rokach, O. Maimon, ”Clustering methods.” Data mining and knowledge discovery handbook, Springer US, 2005, p. 321-352

[7] J. Vesanto, E. Alhoniemi, Clustering of the Self-Organizing Map, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 11, NO. 3, 2000, p. 586-600 [8] U. Maulik, S. Bandyopadhyay, Genetic algorithm-based clustering technique, Pattern Recognition, ISSN: 0031-3203, Vol: 33, Issue: 9, p: 1455- 1465 Publication Year: 2000

[9] Z. Feng, Data Clustering using Genetic Algorithms, Evolutionary Computation: Project Report, CSE484, 2012

[10] M. Mor, P. Gupta, P. Sharma, A Genetic Algorithm Approach for Clustering, International Journal Of Engineering And Computer Science, ISSN:2319-7242, Volume 3 Issue 6 June, 2014, p. 6442-6447

[11] P. Kudova, Clustering Genetic Algorithm, 18th International Workshop on Database and Expert Systems Applications (DEXA 2007), Regensburg, 2007, p. 138-142. doi: 10.1109/DEXA.2007.65

[12] S. Gajawada, D. Toshniwal, N. Patil, K. Garg, Optimal Clustering Method Based on Genetic Algorithm, Advances in Intelligent and Soft Computing, vol 131. Springer, India