Using the Fuzzy Logic to Find Optimal Centers of Clusters of K-means

1 author:

Wed Kadhim
University of Babylon
7 PUBLICATIONS   **13** CITATIONS

Some of the authors of this publication are also working on these related projects:

expert system View project

❒　　3068

Using the Fuzzy Logic to Find Optimal Centers of Clusters of K-means

Wed Kadhim Oleiwi

Departement of Computer, College of Scince for Weman, Babylon University, Iraq

Article Info	ABSTRACT
	Techniques of data mining that used in the medical diagnosis a number of diseases like cancer, diabetes, stroke, and heart disease. The great importance emerging fields for providing diagnosis and a profounder understanding of medical data, its coms from Data mining in medical field .researcher attempts to solve real world health problems in the prognosis and treatment of diseases, by using Healthcare data mining. In this research, the algorithm of k-means is used for grouping medical data, the problem of k-means is to find optimal centers of clusters so, and fuzzy logic is used to get optimal centers of clusters.

Corresponding Author:

Wed Kadhim Oleiwi,
Departement of Computer,
College of Scince for Weman,
Babylon University, Iraq.
Email: wd_dd_ww@yahoo.com

1.　INTRODUCTION

Altered types of heart based problems are analyzer by Major medical data mining techniques .to discovery the efficient data retrieval in data mining the data about the mixtures of heart disease should be used. The important stage in knowledge discovery, is data mining. Some difficult was appear with customary statistical techniques to detect and extract hidden and formerly unidentified relationships, patterns, and knowledge, where data mining used to investigation of huge datasets. Heart disease growing by the world-wide, data mining techniques used by researchers to help health care specialists in the prognosis and treatment of heart disease. Dependent on the convenience of enormous amount of patients' data that could be used to excerpt valuable knowledge [1-2].

One of the technique that commonly used to authorize the accuracy of medical data, and it the elementary medical data mining techniques, is K-Means Clustering algorithm [3].

Clustering is the manner of discovery set that object belong to, such that the object in a set will be similar to one another and different from the objects in other set, in cluster analysis there is no prior knowledge which elements belong to which clusters. Based on the attribute value(s) that best defines the object, Similarities are evaluated. Often distance the grouping or clusters are defined through an analysis of the data [3].

Narendra S. Chaudhari and Avishek Ghosh [4] they have applied an addition of Sammon's algorithm by using fuzzy logic approach. Have used a projection algorithm with an objective to map p-dimensional patterns to q-dimensional space such that the structure of the data is preserved. Finally they have presented that their result was excellent.

Han J. and Kamer M. [3] proposed way for clustering. Representing data by less clusters necessarily misses certain many data object by little clusters and therefore, it representations data by its clusters.

An Improved K-means Clustering Algorithm proposed by Chunfei Zhang, Zhiyi Fang [5]. The idea of the K-means clustering algorithm analysis the advantages and disadvantages of the traditional K-means clustering algorithm, the traditional K-means algorithm is a widely used clustering algorithm, with a wide range of applications. Elaborates the method of improving the K-means clustering algorithm based on improve the initial focal point and determine the K value.

Laurence Morissette and Sylvain Chartier [6] they use three different algorithms: the Forgy/Lloyd algorithm, the Macqueen algorithm and the Hartigan & Wong algorithm. To present powerful of the k-means clustering technique, through then present an implementation in Mathematical and several examples of the different choices available to show the application of the technique.

2.   RESEARCH METHOD

This section describes the methodology of work where, fuzzy logic at the first applied on the heart data to compute the fuzzy value for each patient .at the second step the K-means determined the optimal center for each class depending on fuzzy value.

The process is passed through two steps:

1.   Application of fuzzy logic for obtaining optimal cluster centers.
2.   Application of K _means clustering on the data using the optimal cluster centers. As it explain in the Figure 1.
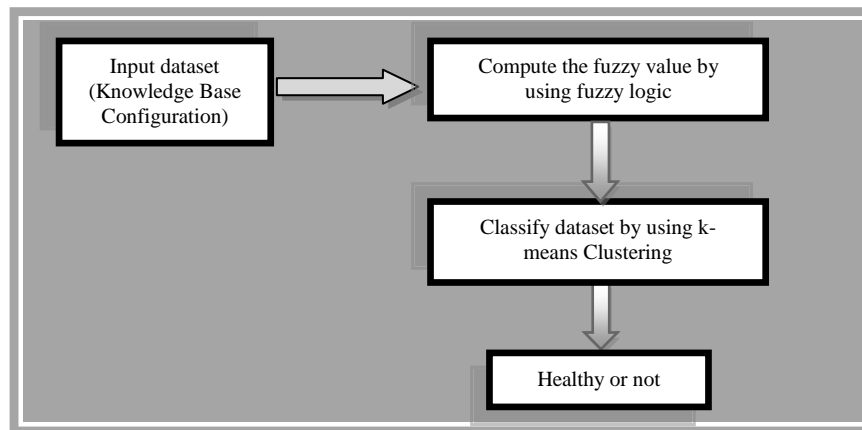

Figure 1. The methodology of the proposed system

3.   RESULTS AND ANALYSIS
3.1.   Knowledge Base Configuration

The main components of the system are the data base and the knowledge base. In this work ,Heart disease diagnosis system's dataset is taken from University of California, Irvin (UCI) [7], stat log dataset for heart disease for 270 data vector (patient) this dataset has 13 attributes (which have been take out from a more set of 75). The proposed system used 11 attributes from the 13 as is shown in Table 1.

Table 1. Attributes of Heart Disease

No	Name	Description
1	Chest pain	1-typical angina; 2-atypical angina; 3-non anginal pain; 4-asymptomatic
2	Age	Real No.
3	Sex	Female (0); male (1)
4	Resting blood pressure	Real No.
5	EGC	0=normal; 1-ST-T wave; 2-hypertrophy
6	Old peak	ST depression induced by exercise relative rest
7	Maximum heart rate	Real No.
8	Thal	3-normal; 6-fixed defect; 7-reversible defect
9	Exercise	1-yes; 0-no
10	Fasting blood sugar	1-true (>120mg/dl); 0-false
11	cholestroal	Real No.

3.2. Fuzzy logic

Use of soft computing,fuzzy logic, genetic algorithm, neural network, supper vector machine, fuzzy Petri nets, in the fields of medicine space diagnosis and treatment has highly increased [1], [8]. Rule based system has been successfully used to sampling human problem-solving where much of human reasoning deals with imprecise, incomplete, or vague information [9-10], the important aspect of design any information system are sound theoretical , capability, accuracy and ease of use [11].

Fuzzy system can take a decisions according to its crisp input (data vector), transformed to the form linguistic variable, that emerge from membership functions, which can be used to determine the fuzzy set that input belong to and it's truth degree, then the linguistic variables are processed with Fuzzy rule IF-THEN rule, the response of each rule is weighted, truth degree of its inputs, that deduce from Fuzzy implication [11-12].

In this work, this truth degree (rule is weighted) used in determined the center of each class at the first cycle of the K-means.

Knowledge base in the form of fuzzy rule, depending on the attribute of heart disease there is 44 rules, each of these rules has a value represent the strength of the belief in the fuzzy rule, certainty factor, and the traditional certainty factor is a constant determined by an expert. In this work, Gaussian membership was used to compute the fuzzy value for each data vector, The Gaussian membership function μA of vector x have been represented by:

$$\mu(x: p, q, s) = exp\left[-\frac{1}{2}\left|\frac{x-p}{q}\right|^{s}\right]$$

In Figure 2 sample of the data after compute Fuzzy value.

age	sex	chest	resting_bl	serum_ch	fasting_bl	resting_el	maximum	oldpeak	slope	thal	class	Fuzzy valu
70	1	4	130	322	0	2	109	2.4	2		3 present	0.218
67	0	3	115	564	0	2	160	1.6	2		7 absent	0.018
57	1	2	124	261	0	0	141	0.3	1		7 present	0.573
64	1	4	128	263	0	0	105	0.2	2		7 absent	0.142
74	0	2	120	269	0	2	121	0.2	1		3 absent	0.099
65	1	4	120	177	0	0	140	0.4	1		7 absent	0.109
56	1	3	130	256	1	2	142	0.6	2		6 present	0.543
59	1	4	110	239	0	2	142	1.2	2		7 present	0.612
60	1	4	140	293	0	2	170	1.2	2		7 present	0.524
63	0	4	150	407	0	2	154	4	2		7 present	0.561
59	1	4	135	234	0	0	161	0.5	2		7 absent	0.231

Figure 2. Sample of the Data After Compute Fuzzy Value

3.2. K-means Clustering Algorithm

In this work, the tool used in classification system is hard clustering or partition clustering. Particularly, a popular partition clustering method called k-means clustering [13].

The method work of the algorithm abstract by take the data features form a vector space and work on find natural group for it [14]. The points are clustered about centroids, the center of a cluster for the *k*-means algorithm is the mean point of all points in the cluster, which are found by minimizing the objective, and Steps of the algorithm are as follows [15-16]:

1. The objects Partition into *k* not empty subgroups
2. Calculate centers of the each group; which at first time chose randomly.
3. Each object (point) assign to the group according to the adjacent center.
4. Stop when no more change in the center, otherwise, go back to step 2.

While in this work the center of the each class not chose randomly at the first time; depending on the fuzzy value that compute for each data vector at the first process of proposed system,as it show in Figure 3. Where there are k =2; healthy or not healthy. Steps of the algorithm are as follows:

Algorithm of K-means clustering.
•••

Input: Heart database I, number of cluster (K)
Output: classify heart database

Begin

Step1: chose the data vector with maximum fuzzy value as the first center of healthy class   .

Step2: chose the data vector with minimum fuzzy value as the first center of not healthy class

Step3: For ith=1 to 270

Step4: For jw=1 to k

Step4-1: Compute the distance between each state of the heart database with the center of each cluster.

Step4-2: State go to the cluster with minim distance.

Next jw

Next ith

Step5: Update the center of each cluster by divide the summation of states in each cluster on the number of the states in that cluster

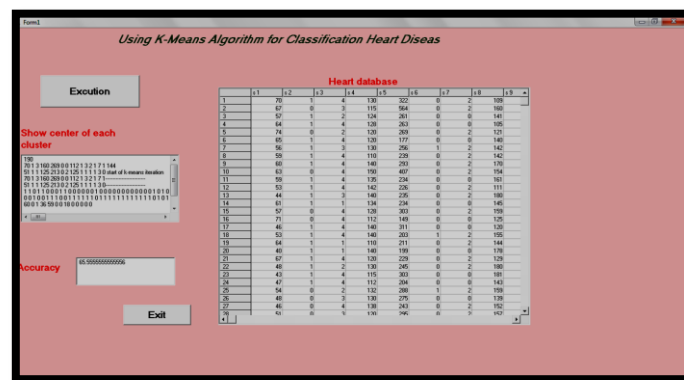Step6: If the center of each cluster do not change stop else go to step 3.

End

Figure 3. K-Means Classification for Heart Disease

4.   CONCLUSION

K-means is a common clustering algorithm that needs an enormous initial set to start the clustering. Experimental result show that, k-means is humble and easy to understand but the main drawback is randomly in first step (selection initial centers of clusters). In this work, we used the fuzzy logic to get out this drawback. Where the first step in this search was compute the Fuzzy value for each data record of heart database, where this value help in chose the seed center of each cluster. The accuracy of this work is 95.48%, while the accuracy of the K-means without fuzzy logic on the same data was 90.78%.

REFERENCES

[1]   Amit K., "Artificial Intelligence and Soft Computing," CRC Press LLC, pp. 2, 2000.
[2]   G. Karthiga, *et al*., "Heart Disease Analysis System Using Data Mining Techniques," *International Journal of Innovative Research in Science, Engineering and Technology*, vol/issue: 3(3), 2014.
[3]   Han J. and Kamer M, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publisher, 2001.
[4]   N. S. Chaudhuri and A. Ghosh, "Feature Extraction using fuzzy rule base system," *International Journal of Computer Science and Applications*, vol/issue: 5(3), pp. 1–8, 2008.
[5]   C. Zhang and Z. Fang, "An Improved K-means Clustering Algorithm," *Journal of Information & Computational Science*, vol/issue: 10(1), 2013.
[6]   L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica," *Tutorials in Quantitative Methods for Psychology*, vol/issue: 9(1), pp. 15-24, 2013.
[7]   R. Detrano, "Medical Center," Long Each and Cleveland Clinic Foundation.
[8]   V. Sundarapandian and E. P. Ephzibah, "Framing Fuzzy Rules using support sets for Effective Heart Disease Diagnosis," *International Journal of Fuzzy Logic Systems (IJFLS)*, vol/issue: 2(1), pp. 11-16, 2012.
[9]   O. O. Oladipupo, *et al.*, "A Fuzzy Association Rule Mining Expert-Driven (FARME-D) approachto Knowledge Acquisition," *African Journal of Computing & ICT*, vol/issue: 5(5), pp. 53-60, 2012. ISSN 2006-1781.
[10] W. Siler, *et al.*, "Fuzzy Expert Systems and Fuzzy Reasoning," *Willey Interscience*, pp. 1-424, 2005.
[11] R. Das, *et al.*, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, Elsevier, vol/issue: 36(2009), pp. 7675–7680, 2009.
[12] Tiruchengode and Namakkal, "Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods," *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, vol/issue: 2(2), pp. 2319–2526, 2013.

[13] P. M. Patel, *et al.*, "Image segmentation using K-mean clustering for finding tumor in medical application," *International Journal of Computer Trends and Technology (IJCTT)*, vol/issue: 4(5), pp. 1239-1242, 2013.

[14] O. A. Abbas, "Comparison between Data clustering Algorithm," *The International Arab Journal Information Technology*, vol/issue: 5(3), 2008.

[15] R. Harikumar, *et al.*, "Performance Analysis for Quality Measures Using K-means Clustering and EM Models in Segmentation of Medical Images," *Int. J. of Soft Computing and Engineering*, vol/issue: 1(6), 2012.

[16] Moslem M. K., *et al.*, "Applying New Method for Computing Initial Centers of K-means clustering with Color Image Segmentation," *J.Thi-Qar Sci.*, vol/issue: 3(1), 2011.

BIOGRAPHY OF AUTHOR

Wed Kadhim Oleiwi: I got a degree Bachelor of Computer Science from the University of Babylon \ College of Sciences \ Department of Computer 2005 High-a good grade, then earned a master's degree from the University of Babylon \ College of Sciences \ Department of Computer Year 20112 High-a good grade in the field of artificial intelligence, and do, Assistant Lecturer at the University of Babylon \ College of Science for women \ Computing Department since 2007 till now.