



Enhancement of Medical Repositories System: Using Clustering, Classical Genetic Algorithm and Fuzzy logic Approaches

Muhammad Shahid*, Noriah Bidin, Yacoob mat Daud, M.Inayat Ullah And Muhammad Talha

Department of Physics

University Technology Malaysia (UTM)

81300 Skudai Johor Malaysia

bluefiber08@gmail.com

Abstract: A repository of Medical diagnostic system extracts the list of diseases and its treatments from data warehouse of the system and helps to detect the disease and treatments according to input symptoms. But repository is unable to detect those diseases whose symptoms are unmatched. In this paper, authors proposed an approach to remove this anomaly from the medical diagnostic system by using three approaches such as fuzzy logic, clustering and classical genetic algorithm. A physician takes input symptoms from patient and yield results. Moreover, physician compares his results with the result of medical diagnostic system.

Keywords: Fuzzy logic, Clustering and Genetic Algorithm

I. INTRODUCTION

In medical environment, data mining is playing a very important role. Medical data of patient via laboratory tests, results, symptoms and diagnoses are growing with the passage of time. Due to increase in medical information, the size of the databases is increasing. Extracting knowledge from the large databases is more difficult and complex. Following three approaches are used to solve this problem.

- [a] Conventional Approach
- [b] Statistical Approach
- [c] Artificial Neural Networks (Non-Linear) Approach

Conventional approach of medical expert systems, medical expert system works on fuzzy sets and fuzzy relations [1]. Definite relationships as well as uncertain relationships were used to verify diagnostic supposition. Conventional approach requires a lot of calculations and formulation of rules if size of input data becomes very large [2]. A complete diagnostic system was developed by Adlassing which works on fuzzy theory [3]. In statistical approach of medical expert system, author [4] presents almost totally identical results of expert systems founded on analysis of patient's data compare to probabilities considered normative according to the Bayesian theory that is statistical approach. The [4] does not differentiate between the kinds of evaluated probability [4]. On the other hand [5] strongly criticizes the use of the Bayesian theory in probabilities evaluation because most routine discussions deal with the extent of sensitivity, specificity and prior probability as fixed parameter serving to update the probability [5]. In 1993 Berry proposed to use the Bayesian theory not for medical diagnoses but for planning, forming and analyzing clinical tests he said that Bayesian theory is flexible for the calculation of the probabilities and these probabilities do not depend on the planning of the experiment design [5]. Another approach, based on statistics, was to construct and evaluate hypotheses by matching a patient's characteristics with stored profiles in a given disease [6]. This approach represents a paradigm shift

from structured reasoning modeled on human expertise, which accesses a knowledge base consisting of operational rules, to one based on empirical evidence, relying on representative databases of historical data. A well known program using this approach was INTERNIST-QMR [7].

This program mirrors hypothetic-deductive reasoning [8] by assigning scores to the clinical findings, listed by disease profiles, such that the accumulation of scores corresponded to build the evidence base for a clinical diagnosis. It could handle coexistent diseases, and was remarkably accurate from the start, given that it made strong assumptions [9]. It is difficult to develop Medical expert system using statistical approach due to problem of uncertainty with numeric values [10] system is another example of statistical approach [11]. Glasgow's system which was developed for the detection of Dyspepsia [12]. Neal introduced "Structured belief net works" for the first time and he implemented it with the Bayes's theorem [13]. All these developments are based on statistical approach but this approach is assumption based approach.

Artificial neural network approach is a non-linear approach, this approach is for handling the noisily data and is suitable for the disease classification [14]. Their role is to find new method for performance, evaluation and classification of new disease [15]. In 1993 Reggia had found out the similarity between empirical and statistical data model. Statistical approach is assumption based while artificial neural network is knowledge driven approach. Statistical modeling is suitable for data distribution, if data is assumption based. Neural networks are used to solve non-linear problems, it is just like black box, input and output variable conversion is automated, minimize the computational time. Artificial neural network is applicable when the class of the object is known in advance but in clustering approach it is not known in advance. Clustering is an important approach for discovering knowledge from the large databases. Algorithms of clustering are more significant and it is the advantage over previous approaches, existing approaches have some limitations application of

these approaches in the biome dical domain is very difficult. In the study of ADAGIO model the author explains the clustering algorithm. That model was able to switch three types of data.

- [a] Qualitative
- [b] Quantitative
- [c] Mixed Type Data
(Folly, 2005)

Fuzzy logic and clustering approach are coupled to form a model. This model extracts approximate values rather than precised. Similarly classical genetic algorithm is used to find exact or approximate solutions to optimize and search problems.

A. Statement of the Problem

The gradually increment in medical information of patients cause to increase in the size of databases of medical softwares. Various approaches have used to handle the large medical databases such as.

- [a] Conventional Approach
- [b] Statistical Approach
- [c] Artificial Neural Network Approach.

All these approaches have used to extract knowledge from the medical databases. The problem of above approaches was that extraction of information of unmatched disease profile would not be diagnosed. When unknown symptoms of the patient were entering in the medical expert system, it was difficult to classify the unknown disease. In all medical expert system there was no specific model which helps to classify the unknown disease. The aim of this research paper is to study the non linear techniques in medical data analysis. The terms, soft computing or computational intelligence is used to cover the non linear techniques such as neural networks, fuzzy logic and neuro fuzzy tools. Recently, it has been shown that these techniques are important for image recognition and image processing. But it is not easy to diagnose such diseases, for example when a doctor entered the symptoms of the patient in the medical expert system and disease profile was not present in the database then expert system became halted. In neural networks, class of the disease profile is known in advance, it is not easy to diagnose the unknown diseases. In case of clustering the class of the disease profile is not known in advance. Clustering technique is suitable to diagnose the unknown diseases. In the proposed techniques fuzzy logic, classical Genetic Algorithm (GA), fuzzy C-mean Algorithm and clustering approach are used to build medical model to extract the treatment of the unknown disease.

B. Scope of the Study

This study based on the household level information about the relationship between tree growing patients data with two sets of questions.

- [a] A doctor enters input symptoms to the medical diagnostic system.
- [b] The diagnostic system diagnosis either the treatment of the related disease in the database is present or not.

C. Objectives

- [a] In case, when the treatment of the related disease is present in the database then doctor match the results with the diagnostic system.

- [b] In case, when treatment of the related disease is not present in the database, then the proposed diagnosing system automatically diagnoses.

II. RESEARCH METHODOLOGY

The working of medical repository system by using fuzzy logic, clustering and Genetic Algorithms are shown in Figure: 1. A Doctor enters input symptoms to the medical diagnostic system. The diagnostic system diagnose, either the treatment of the related disease in the database is present or not. In case, when the treatment of the related disease is present in the database then doctor match the results with the diagnostic system.

In case, when treatment of the related disease is not presently in the database then the proposed diagnosing system will automatically diagnose and create the disease and propose its treatment.

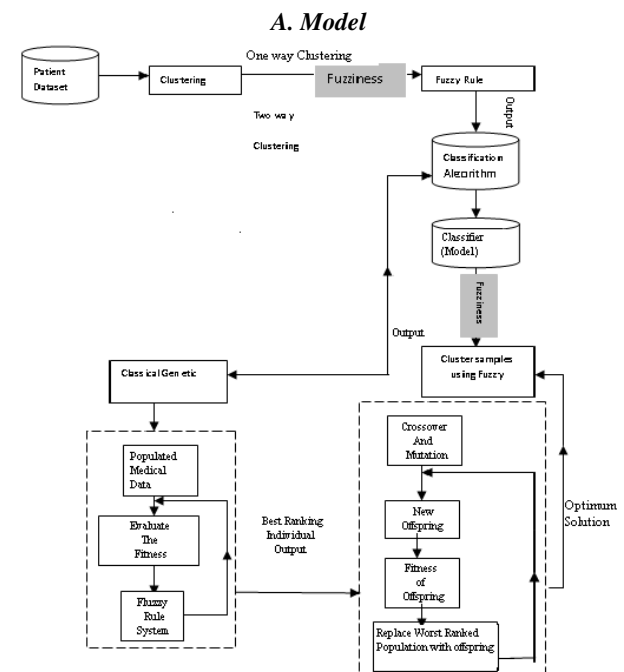


Figure: 1. Model for extracting treatment of unknown disease

B. Algorithm

The description about the work of proposed model Figure: 1 is divided into steps which are following.

- [a] Medical datasets are clustered by using one way or two ways clustering for local view or global view of medical data set.
- [b] By applying Fuzzy rules if the specified cluster is matched then extraction of medical knowledge is successful otherwise next steps will be followed.
- [c] Construct the classification model to find out the class of unknown disease. Each cluster of model points to a disease.
- [d] After this, apply Fuzzy C-Mean Clustering method for clustering the sample clusters.
- [e] To get optimal solution use classical Genetic Algorithm.
- [f] The function of classical Genetic Algorithm has shown in following steps.
- [g] Select the initial datasets.
- [h] Find the class of the selected datasets.

- [i] Repeat
 - a. Pick best ranking solutions to mimic.
 - b. Sort new solutions through intersection and transformation (genetic operations) and give birth to child datasets.
 - c. Calculate the proper class of the child datasets.
 - d. Put back worst ranked part of class with child datasets.
- This classical Genetic Algorithm will repeat its function till optimal solution will be achieved.

III. CONCLUSION

In this paper, fuzzy logic, clustering and genetic algorithms were introduced for the treatment of the unknown diseases from the repository of medical diagnostic system. This system showed successful results, when a patient input symptoms in the medical diagnostic system was present. But if the disease and its treatment were absent in the medical repository then said approaches generated a list of diseases. The proposed model selects the optimal disease and its treatment.

IV. REFERENCES

- [1] Seising, R. 1965. A History of Medical Diagnosis Using Fuzzy Relations
- [2] Aruna. P. and N. Puviarasan and B. Palaniappan 2005. An investigation of neuro-fuzzy systems in psychosomatic disorders, *Expert Systems with Applications* 28, 673–679.
- [3] Adlassing, K. P. (1986). Fuzzy set theory in medical diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-16(12), 260–2265.
- [4] Berry, D. 1993. A case for Bayesianism in clinical trials”, *Statistics in Medicine*, Vol.12, pp. 1377-1393.
- [5] Baron, J. 1994. Uncertainty in Bayes”, *Medical-Decision-Making*, Vol-14, pp, 46-51.
- [6] Schwartz, W.B., Patil, R.S. and Szolovits, P. 1987. Sounding board: artificial intelligence where do we stand? *N. Engl. J. Med.* 1987, 316; 685-688. Dombal, F.T., Leaper, D.J., Tanis land, J.R., McCann, A.P. and Horrocks, J.C. 1972. Computer aided diagnosis of acute abdominal pain, *BMJ*, 2:9-13.
- [7] Miller, R.A., Pople, H.E. Jr. and Myers, J.D. 1982. INTERNIST-1, an experimental computer-based diagnostic consultant for general internal medicine. *N. Engl. J. Med.* 307: 468-676.
- [8] Shortliffe, E.H. Clinical decision-support systems. 1990. In Shortliffe, E.H., Perreault, L.E., Wiederhold, G. and Fagan, L.M. *Medical informatics - computer applications in health care*, Addison-Wesley, Reading, M.A.
- [9] Kulikowski, C.A. Artificial intelligence in medical consultation systems: a review. *IEEE-Eng. In Med. and Biology Mag.* 34-39.
- [10] Dombal, F.T., Clamp, S.E. and Wardle, K.S. 1997. Measuring surgical performance in acute abdominal pain: some reflections from international studies. *Europ. J. Surg.* 163(5):323-329
- [11] Davies, M. and Owen, K. 1990. Complex uncertain decisions: medical diagnosis, Case Study 10 in *Expert System Opportunities from the DTI's Research Technology Initiative*, HMSO.
- [12] McAdam, W.A.F., Brock, B.M., Armitage, T., Davenport, P., Chan, M. and de Dombal, F.T. 1990. Twelve year's experience of computer-aided diagnosis in a district general hospital. *Ann. Roy. Coll. Surg. Engl.* 72:140-146.
- [13] Neal, R.M. 1996. Bayesian learning for neural networks' Springer-Verlag, New York. Pal, S. K., & Mitra, S. 1992. Multilayer perception, fuzzy sets and classification. *IEEE, Transactions on Neural Networks*, 3, 683–697.
- [14] Reggia, J.A. 1993. Neural computation in medicine. *Artificial Intelligence in Medicine*, 5:143-157.
- [15] Ezquerra, N. and Pazos. 1993. A. Neural computing in medicine. *Editorial inartificial intelligence in Medicine.* 6:355-357.