An Efficient Expert System For Diabetes By Naïve Bayesian Classifier

A.Ambica¹, Satyanarayana Gundi², Amarendra Kothalanka³
M.Tech scholar¹, Associate Professor², Professor³
1,² Department of Computer Science & Engineering, ³ Department of Information Technology,
1,2,³ Dadi Institute of Engineering and Technology (Affiliated to JNTUK), Anakapalle, Andhra Pradesh

Abstract: In this paper we are proposing an efficient decision support system for Diabetes Disease, apart from the traditional simple support vector machine. We are proposing an efficient two level approach for classifying data. In initial phase we extract optimal feature set from the training data by analyzing the optimality in the dataset, then new dataset is formed as optimal training dataset, now we apply our classification mechanism on the optimal feature set.

I. INTRODUCTION

Researchers developed a fuzzy-based controller that incorporates expert knowledge to regulate the blood glucose level. Magni and Bellazzi [2] devised a stochastic model to extract variability from a self monitoring blood sugar level time series. Polat and Gunes [3] designed an expert system to diagnose the diabetes disease based on principal component analysis. Polat et al. [4] also developed a cascade learning system to diagnose the diabetes. Chang and Lilly [5] developed an evolutionary approach to derive a compact fuzzy classification system. Goncalves et al. [6] introduced an inverted hierarchical neuro-fuzzy BSP system for pattern classification and rule extraction in databases. Kahramanli and Allahverdi [7] designed a hybrid neural network system for classification of the diabetes database. Chang-Shing Lee [8] designed as fuzzy expert system for diabetes decision support application based on the fuzzy ontology with five layer fuzzy ontology. Ismail Saritas et al. [9] developed a fuzzy expert system to determine drug dose in treatment of chronic intestine inflammation using the concept of fuzzification. Mehdi Fasanghari et al. [10] developed a fuzzy expert system for Tehran stock exchange using the concept of fuzzification. Diabetes treatment focuses on controlling blood sugar levels to prevent various symptoms and complications through diet and exercise. The American Diabetes Association [11] categorizes diabetes into type-1 diabetes, which is normally diagnosed in children and young adults, and type-2 diabetes, i.e., the most common form of diabetes that originates from a progressive insulin secretnary defect so that the body does not produce adequate insulin or the insulin does not affect the cells. The Bayesian classification easing number of diabetics worldwide has drawn the attention of a diverse array of fields, including artificial intelligence and biomedical engineering, explaining why related technologies such as fuzzy inference mechanisms and fuzzy expert systems have been adopted for diabetes research.

More number of the studies have shown that patients suffering from Diabetes can significantly delay the onset and slow down the progression of diabetes micro- and macro-angiopathic complications through intensive treatment and monitoring as in general intensive treatments imply a careful blood case of glucose level (BGL) self-monitoring process of analysis of BGL measurements is one of the most important tasks in order to assess the glucose metabolic control and to revise the therapeutic model in Recent clinical studies have shown the correlation between the glucose variability and the long-term diabetes related complications.

II. RELATED WORK

Expert knowledge system has interesting research work during these years, specifically in the medical field of Diabetes mellitus. This illness requires continuous and regular treatment for the patient who are suffering with Diabetes mellitus, researchers in a way to find an optimal solution of expert knowledge system. Most of the Traditional knowledge systems and classifications works with probability densities density variation between the training and testing datasets. These mechanisms suffering with so many drawbacks like new attribute recovery and mismatch of the the attribute in the training data set and testing dataset, because in there time environment end user analyst cannot expect the semantic so the attributes of the training and testing datasets. So, if analyst can reduce the computational complexity regarding mismatched attributes and We are rectifying one more drawback in the classification approach with elimination features, we will ignore the mismatched unavailable features from the training and testing datasets and calculates the posterior probability with the attributes of the datasets, the following procedure shows the document wise filtering and elimination. We integrated optimal extraction for the optimal diabetes results by eliminating the unavailable attributes from the training datasets and testing datasets.
III. PROPOSED SYSTEM

In this proposed approach we are introducing two level approaches. Initially we Extract the optimal feature set from the existing training data and calculate the positive and negative probability, until a new dataset is formed with same size and forwards the current generated dataset for the classification; there it classifies the testing data features with the new Dataset.

A) Optimal Feature Extraction

Traditional approaches of knowledge expert system works with static measures and it may contains unnecessary information, which means the attributes doesn’t satisfies the minimum threshold value. In our proposed approach Dataset is gathered for the decision support system, with relevant diabetes characteristics or feature sets and before forwarding the dataset to the classification ,forward the Dataset to the Optimal feature set selection process with the specified threshold values of the Dataset.

The objects satisfies the minimum threshold value can be treated as positive attributes and other can be treated as negative attributes, for Optimal extraction of the datasets, remove the records which contains the negative attributes, forward the remaining dataset as optimal dataset to further classification process.

B) Baye’s Theorem

Bayes theorem is a simple calculation of finding probability factors over existing and new attribute possible values over the samples .This approach shows the how the probability works with theoretical values of the samples, Bayesian can be applied in many ways while there is a possibility of existing attribute values and new samples and can be measures in terms of probability class labels. There are so many real time application which are using bayesian approach bank relation, corporate analyzes, educational analysis, science and technology etc..

Notation of the bayes approach is given as follow with sample notation of the probability and with its conditional probability with respect to event

\[
P(T|E) = \frac{P(E|T) \times P(T)}{P(E|T) \times P(T) + P(E|\neg T) \times P(\neg T)}
\]
C) Classification

For the Classification mechanism, we are using Bayesian classifier for classify the testing dataset with newly formed optimal feature set based Dataset for diabetes. This approach works with corresponding posterior probability of the individual features with respect to the original dataset.

For the classification process we are using Bayesian classifier for analyzing the testing data with the training information. Bayesian classifier is defined by a set C of classes and a set A of attributes. A generic class from C is denoted by c and a generic attribute belonging to A as A. Consider a database D with a set of attribute values and the class label of the case. The training of the Naïve Bayes Classifier consists of the estimation of the conditional probability distribution of each attribute given in the class.

Here P(X) is prior probability =

P is the data sample from our set of fruits is red and round) P(X), P(H), and P(X/H) may be estimated from given data . Use of Bayes Theorem in Naïve Bayesian Classifier

1. Each data sample is of the type
   X=(x_i) i =1(1)m, where x_i is the values of X for attribute A_i

2. Consider we have m classes C, i=1(1)m.
   X ∈ C_i iff
   P(C_i|X) > P(C_j|X) for 1 ≤ j ≤ m, j≠i

i.e BC assigns X to class C_i having highest posterior probability conditioned on X . The class for which P(C_i|X) is maximized is called the maximum posterior hypothesis. From Bayes Theorem

3. P(X) is constant. Only need be maximized.
   ♦ If class prior probabilities not known and assume all classes to be equally likely
   ♦ else maximize
   P(C_i) = S_i/S

Problem: computing P(X|C_i) is unfeasible!

(find that how you find it and why it is infeasible)

4. Naïve consideration is: attribute independence
   = P(x_1,…,x_n|C) = ΠP(x_i|C)

5. In order to classify an unknown sample X, evaluate for each class C_i. Sample X is included to the class C_i iff
   P(X|C_i)P(C_i) > P(X|C_j)P(C_j) for 1 ≤ j ≤ m, j≠i.

Experimental Analysis:

Our implementation purpose we have used language java and some synthetic datasets for analysis, the following representation shows the complete implementation of the architecture.
The above figure shows the synthetic dataset before optimal extraction process and it will be forwarded to optimal extraction process as follows.

Now we will forward the optimal dataset to our naïve classification approach for classifying the testing data with training data.

In the above screen we will specify the threshold values of the attributes, then optimal extraction can be done (i.e., retrieves the record which satisfies the threshold values).

Performance representation of the traditional naïve Bayesian and our optimal approach shown as below.
IV. CONCLUSION

We conclude that the project our system provides an efficient knowledge expert system by the naïve classification, in our proposed approach instead of classifying the traditional testing data with training data, we forward the initial training data to the optimal process, to extract the optimal data set, on that optimal dataset we apply classification with Bayesian classifier.

REFERENCES


[6] S. Andreassen, J. Benn, R. Hovorka, K. G. Olesen, and E. R. Carson, A probabilistic approach to glucose prediction and insulin dose adjustment: description of metabolic model and pilot evaluation study,


BIOGRAPHIES

A. Ambica completed her MSC (Computer Science), and currently she is currently pursuing M.Tech in Department of CSE in Dadi institute of Engineering and Technology. Her interested areas are data mining and data warehousing.

Satyanarayana Gandhi is an Associate Professor & Head of the Department of Information Technology, Dadi Institute of Engineering and Technology (Affiliated to JNTUK), Anakapalle, Andhra Pradesh, India. He obtained his M.Tech. in Computer Science & Engineering from Andhra University. He is pursuing Ph.D. in Computer Science & Engineering from GITAM University, Visakhapatnam. His main research interests are Safety critical systems, computer Networks, Service computing. He is the Student Branch Counselor of the DIET CSI Student Branch.

Amarendra Kothalanka is a Professor & Head of the Department of Computer Science & Engineering, Dadi Institute of Engineering and Technology (Affiliated to JNTUK), Anakapalle, Andhra Pradesh, India. He obtained his M.Tech. in Computer Science & Technology from Andhra University. He is pursuing his Ph.D in Computer Science & Engineering from GITAM University,Visakhapatnam. His main research interests are Safety Critical Computer Systems, Software Engineering and Mobile Computing. He is the Sponsor of DIET ACM Student, Women in Computing and Professional Chapters.