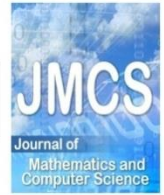




Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com



Designing an Intelligent System for Diagnosing Diabetes with the Help of the XCSLA System

Ehsan Sadeghipour^{1,*}, Ahmad Hatam², Farzad hosseinzadeh^{1,+}

¹Sama technical and vocational training college, Islamic Azad University,
Bandar Abbas Branch, Bandar Abbas, Iran,

*ehsan61.ai@gmail.com – (the corresponding author)

+fhosseinzadeh05@gmail.com

²University of Hormozgan, Faculty of Power and Computer Engineering,
University Hormozgan, Bandar Abbas, Iran,
a.hatam@hormozgan.ac.ir

Article history:

Received July 2014

Accepted October 2014

Available online November 2014

Abstract

An intelligent method for diagnosing diabetes is introduced in this article. One of the main problems involved in this disease is that it is not diagnosed correctly and in time and, due to the destructive effects of the progression of the disease on the human body, the need for its timely prediction and diagnosis is felt more than ever before. At present, doctors diagnose diabetes based on documents, scientific tests, and their own experience. However, considering the huge number of patients, a decision support system for recognizing the disease pattern in diabetics can be used. Results of Program Implementation Document (PID) on databases indicated the higher efficiency of the proposed method in diagnosing diabetes compared to the classic XCS system, the ELMAN neural network, SVM clustering, KNN, C4.5, and AD Tree.

Keywords: Decision support system, blood sugar, processing, intelligent system, artificial intelligence

1. Introduction

The daily increase in the number of diabetics all over the world was one of the problems threatening public health in 2010, and has continued to be so ever since. Diabetes is a chronic metabolic disorder that reduces the ability of the human body to use and completely metabolize sugars (and it decreases

the rate of this consumption and metabolism too). In recent decades, the prevalence of both types of diabetes (Type I and Type II) has increased considerably in most societies. In Type I diabetes (the insulin dependent type), the pancreas of the diabetic person cannot secrete insulin, and in Type II diabetes (the insulin independent type), the pancreas secretes insulin but the level of insulin absorption by body tissues is very low. Based on predictions, by 2010 the number of people with diabetes Type II reached about 200 million [1].

At present, three out of every 20 Iranians suffer from diabetes, more than half of them unaware of their disease [2], and the number of diabetics is still rising. Most diabetics are unaware of their disorder when it begins (because of its mild symptoms at that time), and progression of the disease can lead to serious complications such as eye and skin disorders, kidney diseases (nephropathy), neurological disorders (diabetic neuropathies), cardiovascular disorders, sleepiness, fatigue that occur more than others[3]. Therefore, early diagnosis will help considerably in its successful treatment.

Although it is the doctors who observe the symptoms and use tests to make the final diagnosis, rapid detection and diagnosis is of great importance in modern medicine, and that is why the need is felt for systems to help doctors make decisions. For such decision making, there are decision support systems with various approaches most of which are based on employing machine learning and techniques that can find relationships between the data and, to a large extent, perform correctly in diagnosing the disease. Theoretically, it has been proved that a group of classifiers gives more accurate results compared to using just the best of them [4], [10].

Machine learning refers to a wide range of supervised and unsupervised learning algorithms the purpose of which in data mining is to avoid exhaustive search of data and to replace this kind of time-consuming search with intelligent methods that make clustering data or modeling data behavior very simple through finding the pattern that exists in the data.

During the past two decades, many methods have been introduced in the realm of data mining in which various supervised, unsupervised, or reinforcement-learning algorithms are used for purposes such as recognition and allocation of patterns. Rules-based methods are among the most popular machine learning methods. They use a limited set of ACTION and CONDITION rules to display a small part of the total solution space. The CONDITION rules specify a part of the problem domain and the ACTION rules represent decision based on sub-problems specified by the CONDITION rules.

In recent years, many techniques have been developed for diagnosing and predicting diabetes [11], [12]. Tahami et al. [2] used two methods to predict this diseases and reported accuracies of 78.17 percent. Jing et al. [13] used the nearest neighbor method and could diagnose diabetes with 75.5%

accuracy. Haddadnia et al. used the combination of decision tree and fuzzy clustering algorithm and raised diagnosis accuracy to 84.33%.

In this research, application of the XCSLA method in diagnosing diabetes was studied and results were compared with those of other algorithms such as neural network, KNN, SVM, ELMANC4.5, and AD Tree algorithms. These results indicated the percentage error in the XCSLA algorithm was much less compared the other algorithms.

2. Classification methods for diagnosing diabetes

Automating tasks, and pattern recognition in realizing it, is one of the most important reasons why pattern recognition enjoys a special status in new research for diagnosing various diseases. In designing pattern recognition systems, data related to the system is needed that describes the system well. For this purpose, those features of the diseases are considered that can be used to distinguish diseases belonging to different classes.

2.1. The KNN algorithm

The KNN algorithm, also called the nearest neighbor algorithm, is one of the instance-based learning algorithms. It stores only the training instances during the learning phase. To characterize the class an instance belongs to, the KNN algorithm calculates the distance between the instance and the other training instances. The most common criterion for calculating such distances is the Euclidean criterion, although other criteria such as Manhattan Minovsky are also used for this purpose. After calculating the distance, a majority vote is taken of the k nearest training instances and the label of the majority is allocated to the test instance.

2.2. The SVM algorithm

SVM is a method for classifying linear and nonlinear data. In this method, a nonlinear mapping is used first to convert the initial data to data with higher dimensions, and a search is then made in the new dimensions to find the best separating hyperplane. This hyperplane is a decision boundary separating records of one class from those of others. Data belonging to different classes is easily separated by a hyperplane through a nonlinear mapping onto sufficiently high dimensions. The SVM algorithm finds this hyperplane by using support vectors and the margins defined by support vectors. There are many decision boundaries for separating data belonging to various classes, but those boundary borders or separating hyperplanes with maximum margins, or MMH, are always sought that separate data with greater accuracy and fewer errors.

2.3. The C4.5 algorithm

This algorithm acts by generating a decision tree. It uses a criterion based on entropy and at the same time employs pruning techniques to remove extra branches [15]. The process of building the decision table includes finding a list of the best characteristics for the pattern. This list is such that it has the least amount of error in the population of the records. The fatigue criterion is used to select the optimal features, but other criteria such as information gain and interest rate can also be used.

2.4. The AD Tree algorithm

Ad Trees are data structures used to accelerate conjunctive counting queries on a data set: they return those records with values that apply for a particular combination of features. To build an AD Tree, each feature is given an index. The root node has index zero and stores the number of records present in the data set. All features expand from the root node, and there are the following two restrictions for the expansion of other nodes from each given node:

- (a). For each node, expansion is limited to the maximum number of records
- (b). Expansions with value zero are omitted

This tree does not store all possible values in real numbers, but other combinations can be evaluated by rather the rest of the combinations can be increased in the form of correct changes from the present nodes [16].

2.5. The Elman neural network

The Elman network is a three-layer network with feedback from the output of the hidden layer to the input. Its advantage over maximum return networks is that backpropagation can be used to train the network because connections to context units are fixed so that every context unit acts as a sampling cycle delay. The context layer makes a copy of the output of the hidden neurons and stores it and the value of the neurons in the context layer is applied as an extra input signal on the hidden layer. In the Elman network, weights from the hidden layer to the context layer are fixed with a weight of one because the values of the context neurons must certainly be copied from the output of the hidden layer neurons. Moreover, the weights of the first context neurons are equal to half the output range of the other neurons in the network. One of the changes made in this network is that the outputs of the hidden layer represent network states. Network outputs are functions of the current state, the previous state, and the current input. This means that when a set of inputs is shown to the network, it can learn to present the appropriate output in the context of the previous states of the network. The Elman network has tansig neurons in the hidden layer and purelin neurons in the output layer.

3. The XCS learning system

Machine learning refers to a wide range of supervised and unsupervised learning algorithms the purpose of which in data mining is to avoid exhaustive search of data and to replace this kind of time-consuming search with intelligent methods that make clustering data or modeling data behavior very simple through finding the pattern that exists in the data. During the past two decades, many methods have been introduced in the realm of data mining in which various supervised, unsupervised, or reinforcement-learning algorithms are used for purposes such as recognition and allocation of patterns. Classifier systems are among the most successful of these methods. Generally, classifier systems include a set of rules with “if-then” format each of which presents a potential solution for the target problem. A reinforcement-learning mechanism gradually evaluates this set of rules and updates it at specific time intervals with the help of a genetic algorithm. During the course of this gradual evolution, the system learns the behavior of the environment and then, in the application phase, presents suitable answers to the queries raised by the user. The first classifier system called Learning Classifying Systems (LCS) was introduced by Holland in 1976. In this system, the value of each rule was evaluated by an index called “strength.” The strength of a rule increased in proportional to the degree of its correct answering to training instances in the framework of standards for reinforcement-learning and, at specific time intervals, an evolutionary algorithm in search (usually a genetic algorithm) was responsible for producing new rules and for omitting inefficient ones. At the end of the training stage, this set of rules had the relative ability to present acceptable solutions when faced with new queries. At the same time, successful performance of the LCS was contingent on selecting appropriate values for control parameters of the system, which depended on the type of experience the system designer had. Since the introduction of LCS, other types of classifier systems have been proposed including the Extended Classifier Systems (XCS). Prior to the introduction of XCS in 1995, these systems had very limited ability in obtaining appropriate answers but, since their introduction, classifier systems have gradually developed into more intelligent and more accurate factors and it is now believed that XCS and their improved versions can solve complex problems without requiring parameter adjustment. With the introduction of the classifier system having continuously varying variables (XCSR), some of the innate weaknesses of binary classification systems such as the inability to introduce specific intervals of variable values were mostly resolved. Now these systems have been accepted as one of the most successful learning agents for solving data mining problems in environments that are only partially observable [17]. According to the common strategy used in training XCSR, the fitting of only those rules continues that give correct answers to the training data. This means that the chances of a rule for not being omitted and for participating in the production of

new rules directly depend on the way it answers to the training data, and that the realistic determination of these chances requires making use of a large amount of training data. Since in real-life problems there is a limited amount of training data that cannot be easily increased, use of the XCSR in such applications is usually not justified considering the computational time and cost involved.

A new method for improving performance and for increasing convergence speed of the XCSLA through using limited training data is presented in the remainder of the article

3.1. The learning system (XCSLA)

In this method, all the randomly produced rules can be used as memories to store the state by first adding an LA field to the end of each one. It is known that in the cellular learning automaton the package of reinforcement signals is only a function of the action selected by each cell and of the actions selected by the neighboring cells. In this method, the set of training data is used for modifying features of the rules including “prediction”, “prediction error”, “fitness”, and “the value of LA.” To do this, the following relationships are used:

Updating prediction and prediction error:

$$\text{If } \exp_i < 1/\beta \text{ then } P_i = P_i + (R - P_i) / \exp_i, \quad \varepsilon_i = \varepsilon_i + (|R - P_i| - \varepsilon_i) / \exp_i, \quad L_a = L_a + (R - P_i) / 1 - \exp_i$$

$$\text{If } \exp_i \geq 1/\beta \text{ then } P_i = P_i + \beta (R - P_i), \quad \varepsilon_i = \varepsilon_i + \beta (|R - P_i| - \varepsilon_i), \quad L_a = L_a + (R - P_i) / 1 - \exp_i$$

Updating fitness:

$$\text{If } \varepsilon_i < \varepsilon_0 \text{ then } k_i = 1$$

$$\text{If } \varepsilon_i \geq \varepsilon_0 \text{ then } k_i = \beta \left(\frac{\varepsilon_i}{\varepsilon_0} \right)^{-\gamma} + \beta (L_{a_i} - f_i)$$

$$F_i = f_i + \beta [L_{a_i} / \sum k_j] - f_i$$

In the above relations, β is the learning rate, γ the power of the accuracy rule, ε the prediction error, \exp the experience rule, P the prediction rule, R the reward received from the environment, k the accuracy rule, L_a the value of the automaton field added, f its fitness, and the i index the number of the rule in the rule set. The action sets of all training instances are then randomly used and parents are selected through a roulette wheel selection process, and the genetic operations are performed. Rules that are placed in the match set and are of very low fitness then select the value of the field that has the highest fitness in the match set in place of their own action values and update their LA field values, and the possibility of reduction in min action size is thus created. If this happens, the new rule that is

produced is replaces one of the rules that had the lowest fitness value in the entire rule set so that the more powerful rules produced replace weaker ones that may not be used until the end of the process. The new rules are selected in this way: the “random selection of residuals” method is used and fields that make up the CONDITION part of the data are stored in a table called the mating pool, the method of selection of the various pairs by using the learning automaton field is added, and the crossover method is employed. The strings that represent the CONDITION part are ordered according to the values of their La fields, the medial crossover method is used to replace the initial string pair by a new pair of children (new data), and the ACTION part of the new data is also produced by employing the available data. These operations continue until the learning cessation condition (in this example, 20 runs of the training instances on the produced rules) is satisfied with the help of the completed data [18].

4. Methodology

Data is the first thing required in using various methods. Hospital systems are among those that deal with large volumes of data and can provide the needed information. The data employed in this research was from the PID set of data belonging to reference [19], and included eight features (defined by the World Health Organization) of 768 women with the minimum age of 21. Five hundred of the tested 768 women were healthy and 268 were diabetic. The recorded features are presented in Table 1

Table 1: The set of data related to diabetics

Number of times pregnant
Plasma glucose concentration 2 hours in an oral glucose tolerance test
Diastolic blood pressure (MMHG)
Triceps skin fold thickness
2 hours serum insulin
Body mass index
Age
Class variable

We used 528 random samples for training the described methods and 240 other random samples for the trained system. Table 2 shows the efficiencies of the various classifier systems. Results indicate the superiority of XCSLA system.

Table 2: Efficiencies of the five algorithms in diagnosing diabetes

XCSLA	XCS	AD Tree	SVM	C4.5	KNN	ELMAN
95.6	87.19	73.18	77.84	71.32	70.26	79.42

5. Conclusions

In this article, the XCSLA classification system was introduced for data mining purposes in diagnosing diabetes. The data mining technique and machine learning can be used in discovering rules and in extracting useful information from medical data. Comparison of some of the machine learning algorithms indicated the proposed method is superior to the others mentioned. This successful application gives the promise that this system can also be used in diagnosing other diseases and in helping doctors as a decision support system.

References

- [1] Report of WHO: “*Prevention and control of diabetes mellitus*”.(1993), 3-8
- [2] Seyyed Ehsan Tahami, S., M. Bamshaki; M., A. Khalilzadeh, “*Diagnosing Diabetes Type I by using the ANFIS, GA-NN algorithm*”, The First Joint Conference of Intelligent Systems and Fuzzy Systems, Ferdowsi University of Mashhad (2007).
- [3] Iran Diabets Society, <http://www.ir-diabetes-society.com / Avarez %20diabet.htm>
- [4] L.I. Kuncheva, J.C. Bezdek, R.P.W. Duin, “*Decision templates for multiple classifier fusion: an experimental comparison*”, Pattern Recognition, 34(2) (2001), 299–314.
- [5] Bi. Yaxin, Bell. David, Wang. Hui, Guo. Gongde, Guan. Jiwen, “*Combining Multiple Classifiers Using Dempster’s rule for text Caregorization*”, Applied Artificial Intelligence, 21:3 (2007), 211- 239.
- [6] S. Tulyakov, S. Jaeger, V. Govindaraju, D. Doermann, “*Review of Classifier combination Methods*”, Studies in Computational Inteligence(SCI) 90, (2008), 361-386.
- [7] M. Sugeno, “*Industrial Applications of Fuzzy Control*”, Elsevier,Book, New York, (1985).
- [8] F. Bergh, A. Engelbrecht, “*A new locally convergent particle swarm optimizer*”, IEEE, Fundamenta Informatic Journal (2002).
- [9] L.I. Kuncheva, “*Combining Pattern Classifiers, Methods and Algorithms*”, New York, NY: Wiley Interscience, (2005).
- [10] J. Kittler, M. Hatef, R. Duin P. W, J. Matas, “*OnCombining Classifiers*”, IEEE Transactions on Pattern Analysis and Machine Intel ligenge, 20(3) (1998), 226–239.
- [11] M. Shanker, “*Using neural networks to predict the onset of diabetes mellitus*”, J Chem Inform Computer Science, 36 (1996) 35–41.

- [12] D. Dazzi, F. Taddei, A. Gavarini, E. Uggeri, R. Negro, A. Pezzarossa, “*The control of blood glucose in the critical diabetic patient: a neuro-fuzzy method*”, *Journal of Diabetes Complications*, 15(2) (2001) 80–87.
- [13] Y. Jiang, Z. Zhou, “*Editing Training Data for kNN Classifiers with Neural Network Ensemble*”, in *Proc. ISNN (1)*, 2224, pp.396-361.
- [14] Javad Haddadnia, J. Vahidi, A. Gharah-khani, V. M. Fayyouzi, “*Fuzzy diagnosis of Diabetes Based on Rules and Optimal Characteristics Based on a Combination of data mining Systems and Artificial Intelligent Algorithms ICNMO*”, (2012).
- [15] Mohammadreza Keyvanpour, L. Khalatbari, “*Comparison of Classifier Algorithms in Diagnosing Diabetes and heart Failure*”.
- [16] I. H. Witten, E. Frank, “*Data Mining (Practical Machine Learning Tools and Techniques)*”, _nded, San Francisco: Morgan. .Kaufmann, (2005).
- [17] M. Shariat Panahi, N. Moshtaghi Yazdani, “*An Improved XCSR Classifier System for Data Mining with Limited Training Samples*”, *Global Journal of Science, Engineering and Technology*, (ISSN: 2322-2441) Issue 2, (2012), pp. 52-57.
- [18] Navid Moshtaghi Yazdani, A. Yazdani, “*Improving the XCS algorithm by Using Learning Automata*”, *The International Computer, Information Technology and Digital Media Conference*, (2013).
- [19] Blake C.L., Merz C.J., “*UCI Repository of Machine Learning Databases*”, (1998), <http://www.ics.uci.edu/mllearn/MLRepository.html>.