Applied Soft Computing 20 (2014) 103–111

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system

José Antonio Sanz [a,*], Mikel Galar [a], Aranzazu Jurio [a], Antonio Brugos [b], Miguel Pagola [a], Humberto Bustince [a]

[a] *Dept. of Automática y Computación, Universidad Publica de Navarra, Spain*
[b] *Dept. of Ciencias de la Salud, Universidad Publica de Navarra, Spain*

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form
15 November 2013
Accepted 16 November 2013
Available online 28 November 2013

Keywords:
Linguistic fuzzy rule-based classification systems
Interval-valued fuzzy sets
Genetic fuzzy systems
Cardiovascular diseases

ABSTRACT

Objective: To develop a classifier that tackles the problem of determining the risk of a patient of suffering from a cardiovascular disease within the next 10 years. The system has to provide both a diagnosis and an interpretable model explaining the decision. In this way, doctors are able to analyse the usefulness of the information given by the system.
Methods: Linguistic fuzzy rule-based classification systems are used, since they provide a good classification rate and a highly interpretable model. More specifically, a new methodology to combine fuzzy rule-based classification systems with interval-valued fuzzy sets is proposed, which is composed of three steps: (1) the modelling of the linguistic labels of the classifier using interval-valued fuzzy sets; (2) the use of the K_α operator in the inference process and (3) the application of a genetic tuning to find the best ignorance degree that each interval-valued fuzzy set represents as well as the best value for the parameter α of the K_α operator in each rule.
Results: The suitability of the new proposal to deal with this medical diagnosis classification problem is shown by comparing its performance with respect to the one provided by two classical fuzzy classifiers and a previous interval-valued fuzzy rule-based classification system. The performance of the new method is statistically better than the ones obtained with the methods considered in the comparison. The new proposal enhances both the total number of correctly diagnosed patients, around 3% with respect the classical fuzzy classifiers and around 1% vs. the previous interval-valued fuzzy classifier, and the classifier ability to correctly differentiate patients of the different risk categories.
Conclusion: The proposed methodology is a suitable tool to face the medical diagnosis of cardiovascular diseases, since it obtains a good classification rate and it also provides an interpretable model that can be easily understood by the doctors.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Cardio vascular diseases (CVDs) affect the heart and they are usually caused by some disorder that hinders the blood flow. These diseases imply a high risk of suffering from severe illness like heart attacks or thrombosis among others. They are the main health problem in adult population provoking a high death rate in many developed countries [1]. Therefore, it is important to obtain an early diagnosis of the risk of suffering from such diseases so as to start a proper medical treatment to reduce the chances of developing them.

In order to estimate such risk, Spanish doctors look up specific tables called REGICOR [2]. These tables consider different variables like gender, age, presence or absence of diabetes, systolic and diastolic blood pressure, total cholesterol and HDL cholesterol values, among others. The value provided by this procedure quantifies the risk of the patient of suffering from a CVD during the next ten years. In this manner, different categories of patients according to this value can be established. Hence, the problem of estimating the patients' risk category can be considered as a classification problem.

Fuzzy rule-based classification systems (FRBCSs) [3] are a useful tool to face classification problems. These systems are widely used because of their good performance and their capability to build an interpretable model which uses common linguistic terms for the user in the problem domain. Moreover, they offer the possibility of mixing information coming from different sources, i.e. expert

---

* Corresponding author. Tel.: +34 948166048; fax: +34 948168924.
E-mail addresses: joseantonio.sanz@unavarra.es (J.A. Sanz), mikel.galar@unavarra.es (M. Galar), aranzazu.jurio@unavarra.es (A. Jurio), antonio.brugos@unavarra.es (A. Brugos), miguel.pagola@unavarra.es (M. Pagola), bustince@unavarra.es (H. Bustince).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.11.009

knowledge, mathematical models or empirical measures. For this reason, FRBCSs are suitable to deal with medical diagnosis problems since, besides from providing the patients diagnosis, doctors can know the reasoning behind the decision by looking the rule or set of rules involved in the final classification. FRBCSs use fuzzy logic [4] in order to model the linguistic terms used by the system. A key step for the subsequent success of fuzzy systems is the definition of the membership functions representing the problem information as well as possible. Sometimes, it is really difficult to determine the membership functions because the same concept can be defined in different ways by different persons [5]. This problem led Zadeh to suggest the notion of type-2 fuzzy sets [6] as an extension of fuzzy sets [4]. A particular case of type-2 fuzzy sets are the Interval-Valued Fuzzy Sets (IVFSs) [7] that assign as membership degree of the elements to the set an interval instead of a number. IVFSs allow the system uncertainties to be modelled whereas their computational effort is less than the one demanded by the use of type-2 fuzzy sets.

In Ref. [8], the authors proposed an interval-valued FRBCS (IV-FRBCS), that is, a FRBCS whose linguistic labels are modelled with IVFSs enhancing the performance of classical FRBCSs. In this manner, the inherent ignorance related to the definition of the membership functions was modelled by means of the IVFSs. Then, the shape of every IVFS considered in the system was optimized by using an evolutionary tuning approach. Furthermore, an Interval-valued fuzzy reasoning method (IV-FRM) was proposed, where the first two steps, namely, the computation of the matching and the association degrees for each rule of the FRBCS, used IVFSs. In order to apply the remainder of the method as in the classical FRM [9], a number was given as a result of the association degree. To compute it, the two values associated with the lower and upper bounds of the intervals were averaged, which may cause that the system does not make the most of the interval information.

In this paper, in order to handle the interval information in the IV-FRM, we introduce the K_α operator defined by Atanassov [10] to compute the association degree. In this manner, the information given by the IVFSs is exploited, since other values rather than the average one can be obtained. As a result of introducing the K_α operator, the values for the α parameters need to be found. In order to do so, we propose an evolutionary tuning to compute the best α value for each rule involved in the inference process and therefore, to provide the system with a new mechanism to take advantage of the extra information given by the IVFSs. In the experimental study, we will show that our new IV-FRBCS allows one to improve the behaviour of the previous approaches when predicting the risk of suffering a CVD and hence, it allows helping the primary care doctors. The new FRBCS will only use as inputs the physical values that can be measured directly by the doctor, i.e. gender, age, smoking condition, blood pressure and body mass index. The objective of the system is to provide the doctors with a quick and reliable estimation of the patients' risk category, in such a way that they can make better decisions like deriving the patient to the secondary health centres (hospitals) or starting an appropriate treatment according to the patient's risk category if necessary.

In order to show the validity of our proposal, in the experimental study we will use two well-known FRBCSs, namely the Chi et al.'s method [11] and the Fuzzy Hybrid Genetics-Based Machine Learning (FH-GBML) algorithm [12]. We will study the behaviour of our new methodology with respect to both the initial FRBCSs and the previous IV-FRBCS [8]. To this end, we will consider the standard classification accuracy as well as the classification rate for each one of the three different CVD risk categories in which the patients can be classified. The paper is organized as follows: the problem of the CVDs is presented in Section 2. Next, the basic concepts of IVFSs and FRBCSs along with the description of the previous proposal to combine FRBCSs with IVFSs are given in Section 3. Then, in Section 4, we describe in detail both our new proposal to introduce the K_α operator in the IV-FRM and the genetic tuning of the parameter α. Section 5 shows the experimental framework along with the analysis of the obtained results. Finally, the main conclusions of this paper are drawn in Section 6.

2. Problem description

CVDs affect different parts of the body, mainly the heart and the arteries of the brain, heart and legs. Most of these diseases are induced by the decrease of either the calibre or the diameter of the arteries. The lack of blood supply does not only damage the heart but also the legs and the brain, which can lead to health disorders implying an increase of the risk of suffering from heart attacks, thrombosis or rupture of blood vessels, among others.

Among adult population, CVDs are the main health problem in general, being in the first place of the list of death cause of persons older than forty five years in many countries. As an example, about 100,000 persons per year die in Spain due to these diseases, representing a death rate of 75–150 deaths per 100,000 inhabitants depending on the region. This rate is similar in most of the developed countries [1]. Therefore, it is really important to estimate the patients' risk of developing a CVD in order to obtain a quick diagnosis to try to avoid their consequences.

The main CVD risk factors were identified in the Framingham Heart Study published in 1951 [13]. In Fig. 1 the different factors involved in the CVDs are depicted. Although these factors are known, their epidemiological relevance is different and, in some cases, they need extra studies so as to correctly weight them. In order to estimate the risk of suffering from a CVD, a global evaluation using the Anderson Table [14] should be made. This table is easily applicable, being advised to estimate the total coronary risk. Specifically, the following categories of risk are defined in this table:

- **High CDV risk**: persons having a probability greater than 25% of suffering from a CVD in 10 years.
- **Moderate CDV risk**: the probability of a person to develop a CVD in 10 years is between 10% and 20%.
- **Low CDV risk**: the person have a probability less than 10% of suffering from a CVD in 10 years.

The lower number of CVD events suffered in Spain led to develop the REGICOR tables [2], which are a version of the Anderson tables adapted to the Spanish population. Thereafter, Marrugat et al. published an estimation of the CVD risk in Spain using the calibrated Framingham equation [13]. In this paper, authors showed that in 10 years, the probability of the Spanish population of suffering from a CVD estimated by their method is thirteen times lower than the probability computed using the Framingham approach with the patients considered as high risky when using the Framingham method.[1]

In any case, the use of these tables is only advocated for primary care prevention health centres. That is, a patient derived to secondary prevention (due to previous events of ischemia, cerebral vascular or peripheral arterial accidents) has a high risk independently of the value computed using the aforementioned tables.

The patients having the main priority to be monitored and treated are those who belong to secondary prevention. They are followed by the patients belonging to primary care prevention having

---

[1] We must remark that we have to be careful with this result since the validation period of the hypothesis is still in process.

Download English Version:

https://daneshyari.com/en/article/495325

Download Persian Version:

https://daneshyari.com/article/495325

Daneshyari.com