Detection of CAN by Ensemble Classifiers Based on Ripple Down Rules

Andrei Kelarev[1], Richard Dazeley[1], Andrew Stranieri[1], John Yearwood[1], and Herbert Jelinek[2]

[1] Centre for Informatics and Applied Optimization
School of SITE, University of Ballarat
P.O. Box 663, Ballarat, Victoria 3353, Australia
{a.kelarev,r.dazeley,a.stranieri,j.yearwood}@ballarat.edu.au
[2] Centre for Research in Complex Systems and School of Community Health
Charles Sturt University, P.O. Box 789, Albury, NSW 2640, Australia
hjelinek@csu.edu.au

Abstract. It is well known that classification models produced by the Ripple Down Rules are easier to maintain and update. They are compact and can provide an explanation of their reasoning making them easy to understand for medical practitioners. This article is devoted to an empirical investigation and comparison of several ensemble methods based on Ripple Down Rules in a novel application for the detection of cardiovascular autonomic neuropathy (CAN) from an extensive data set collected by the Diabetes Complications Screening Research Initiative at Charles Sturt University. Our experiments included essential ensemble methods, several more recent state-of-the-art techniques, and a novel consensus function based on graph partitioning. The results show that our novel application of Ripple Down Rules in ensemble classifiers for the detection of CAN achieved better performance parameters compared with the outcomes obtained previously in the literature.

1 Introduction

Ripple Down Rules produce models which are easier to maintain and update than other alternatives, [5], [37]. In addition they have the most compact representations of the models, which can be better explained to and understood by medical practitioners, see Section 4 below for more details. The present article deals with an experimental investigation and comparison of several ensemble methods based on Ripple Down Rules in a novel application for the detection of cardiovascular autonomic neuropathy (CAN) in diabetes patients. Our experiments included several essential ensemble methods, a few more recent state-of-the-art techniques, a novel consensus function based on graph partitioning, and used the Diabetes Screening Complications Research Initiative (DiScRi) data set collected at Charles Sturt University, Albury, Australia.

DiScRi is a very large and unique data set containing a comprehensive collection of tests related to CAN. It has been previously considered in [20], where

D. Richards and B.H. Kang (Eds.): PKAW 2012, LNAI 7457, pp. 147–159, 2012.

decision trees were used. Our new results based on Ripple Down Rules and presented in this paper have achieved substantially higher accuracies compared with the previous outcomes in [20].

The paper is organised as follows. Section 2 deals with cardiovascular autonomic neuropathy. Section 3 describes the Diabetes Complications Screening Research Initiative (DiScRi) organised at Charles Sturt University, and the corresponding data set. Section 4 contains brief background information on Ripple Down Rules (RDR). Section 5 describes ensemble methods investigated in this paper. Section 6 presents the experimental results comparing the efficiencies of several ensemble methods based on RDR for this application domain. These outcomes are discussed in Section 7, where the main conclusions are also provided.

2 Cardiovascular Autonomic Neuropathy

Cardiovascular autonomic neuropathy (CAN) is a condition associated with damage to the autonomic nervous system innervating the heart and highly prevalent in people with diabetes, [13], [14], [29]. It is known as one of the causes of mortality among type 2 diabetes patients. The classification of disease progression associated with CAN is important, because it has implications for planning of timely treatment, which can lead to an improved well-being of the patients and a reduction in morbidity and mortality associated with cardiac arrhythmias in diabetes.

The most important tests required for identification of CAN rely on assessing responses in heart rate and blood pressure to various activities, usually consisting of five tests described in [13] and [14]. It is often difficult for clinicians to collect all test data from patients, since they are likely to be suffering from other illnesses affecting their general fitness and making it hard to follow correct procedures for all tests. More details on various other associated tests for CAN are given in the next section.

3 Diabetes Complications Screening Research Initiative

This paper used the data set of test results and health-related parameters collected at the Diabetes Complications Screening Research Initiative, DiScRi, organised at Charles Sturt University, [8], [20], [34]. There are no other alternative data sets containing comparable collections of test outcomes. The collection and analysis of data in the project has been approved by the Ethics in Human Research Committee of the university before investigations started. People participating in the project were attracted via advertisements in the media. The participants were instructed not to smoke and refrain from consuming caffeine containing drinks and alcohol for 24 hours preceding the tests as well as to fast from midnight of the previous day until tests were complete. The measurements were conducted from 9:00am until 12midday and were recorded in the DiScRi data base along with various other health background data including age, sex