A Hybrid Intelligent System Design for Diabetes Risk Classification

Thirumalaimuthu Thirumalaiappan Ramanathan, B.E (AU)

Faculty of Education, Science, Technology and Mathematics

University of Canberra

November 2015

A Dissertation Submitted in partial fulfillment of the requirements for the degree of Master of Information Sciences (Research)

Abstract

Risk classification is a major technical challenge in medical diagnosis and chronic illness management. Various computational techniques have been developed for risk classification in recent years with improvements in outcomes. This thesis investigates a novel approach combining support vector machine (SVM) and fuzzy modelling. The proposed hybrid model (called SVM-Fuzzy) is designed, implemented and evaluated on an available benchmark dataset.

Diagnosis and management of diabetes mellitus (also known as type 2 diabetes) are the motivating problem for the current investigation of risk classification. Type 2 diabetes is a chronic condition marked by elevated levels of blood glucose. The prevalence of diabetes is increasing at a fast pace due to obesity, in particular, central obesity, physical inactivity, and unhealthy dietary habits. In type 2 diabetes patients, a range of different organs are under stress and are influenced by the altered metabolic condition. Hence the ailment has to be properly managed within stipulated boundaries to minimize other health complications caused by type 2 diabetes and to assure longevity of life. In the proposed model, fuzzy reasoning is used to classify the level of risks from data. SVM is used to design the fuzzy rules. The well-known Pima Indian diabetes dataset (Frank, 2010) is used to train the SVM and for testing the fuzzy system. The goal is to evaluate the proposed design for better accuracy in risk classification and to investigate training the machine learning algorithm using sample real world data. Another goal is also to investigate efficiency in classification by optimizing selection of right sized datasets through appropriate dataset size selection from experiments. The experiments from the SVM-Fuzzy model show promising results on the benchmark Pima Indian diabetes dataset.

Early in the thesis the various soft computing approaches, such as artificial neural network, SVM, Bayesian network and fuzzy logic models which have been previously reported as relevant to diabetes risk classification are reviewed. The drawbacks in these models are analyzed and identified gaps from the previous computational models on risk classification are targeted in the research questions.

The main objective in risk classification is to handle uncertain and incomplete input data for risk classification and to make use of sample datasets using fuzzy reasoning. The experimental results from the proposed model (SVM-Fuzzy) from the type 2 diabetes risk classification problem demonstrates that it is able to handle uncertain input data by making use of sample datasets for fuzzy reasoning. The outcomes from the proposed

model on type 2 diabetes risk classification are summarized and compared with the previous computational models. These experimental results are encouraging. The SVM-Fuzzy model can be further extended to manage type 2 diabetes risks using its output risk values in planning lifestyle and diet for responsive management of type 2 diabetes conditions. The model can further be extended to a multi-agent learning and planning environment for solving classification problems in complex applications, such as extracting patterns in structure of proteins in bioinformatics.

Acknowledgement

I thank my God for giving me strength and knowledge to conduct the research and write this thesis. I would like to convey my gratitude to my supervisor Professor Dharmendra Sharma for his valuable contribution and expertise. I discussed all my ideas with him and he gave me specific direction to follow on. His guidance, motivation and encouragement helped in every step of this research work. My thanks to Professor Xu Huang, Associate Professor Dat Tran and Dr Chris Barnes provided for their guidance and support. I would also like to extend my appreciation to the Faculty of Education, Sciences, Technology and Mathematics for their support for my research. My special thanks to my parents and my uncle Professor Kaliappa Kalirajan (from the Australian National University) for keeping me motivated and for all their love and understanding throughout the journey.

CONTENTS

List of Tables and Figures