Chapter 7

On Sharp Boundary Problem in Rule Based Expert Systems in the Medical Domain

Oladipupo O. Olufunke  
Covenant University, Nigeria

Uwadia O. Charles  
University of Lagos, Nigeria

Ayo K. Charles  
Covenant University, Nigeria

ABSTRACT

Recently, the application of the conventional rule based expert system for disease risk determination in medical domains has increased. However, a major limitation to the effectiveness of the rule based expert system approach is the sharp boundary problem that leads to underestimation or overestimation of boundary cases, which ultimately affects the accuracy of their recommendation. In this paper, an expert driven approach is used to investigate the viability of a fuzzy expert system in the determination of risk associated with coronary heart disease with regards to the sharp boundary problem in rule based expert system.

1. INTRODUCTION

The use of human expert knowledge in form of rules to solve real-world problems that normally would require human intelligence, known as rule based expert system, has played an important role in modern intelligent systems (Harleen & Siri, 2006). However, a major limitation to the effectiveness of this class of expert systems is the sharp boundary problem (SBP) which leads to underestimation or overestimation of boundary cases as a result of the quantitative attributes partitioning strategy; which consequently affects the accuracy of the expert system (Verlinde, Cook, & Boute, 2006).

In the medical domain, the use of rule based expert system has increased greatly because of the scarcity of human experts in the domain and

DOI: 10.4018/978-1-4666-1755-1.ch007
the availability of fast growing databases which could be used to model inferences and discover patterns in form of rules. In real live application, medical databases contain different kinds of attributes such as binary and quantitative attributes (Delgado, Marin, Sanchez, & Vila, 2001). Binary takes values from 0 or 1; for instance, a patient’s smoking status could be ‘yes’ or ‘no’. Quantitative attributes that are categorical, numerical, or non-fractional in nature, take values from an ordered numerical scale, often a subset of the real number (Kuok, Fu, & Wong, 1998). Quantitative attributes are very common in medical databases. For example heart disease patients can take age values between 20-79 years, result from laboratory test for systolic blood pressure level could take values within <120 to >= 160mm/Hg, cholesterol measures could be within the range of <160 to >= 280mg/dL.

In building an expert system, quantitative attributes need to be partitioned into ranges because of the very wide range of values defining their domain. There are several approaches to partitioning quantitative attributes as discussed in literature (Han & Kamber, 2001). The partitioning process is referred to as binning, that is an interval is considered as a “bin”. The common binning strategies are: 1) Equiwidth binning, where the interval size of each bin is the same; 2) Equidepth binning, where each bin has approximately the same number of tuples assigned to it; and 3) Homogeneity-based binning, where bin size is determined so that the tuples in each bin are uniformly distributed. Also, there is the Distance based partitioning strategy, which seems most intuitive since it groups quantitative values that are closed together within the same interval (Han & Kamber, 2001). All of these partitioning strategies are subject to sharp boundary problem because of the classical set theory (Kuok, Fu, & Wong, 1998). However, to prevent this problem, in (Navruz & Serhat, 2007) fuzzy logic concept was introduced into a rule based expert system to determine coronary heart disease risk. The design gives the user the risk ratio and most of the experimented test data risk ratio from the fuzzy approach was reported to give relatively the same percentage risk as Adult Treatment Panel III (ATP III) calculation, which reflect the extent to which fuzzy concept was able prevent sharp boundary problem. In our work a comparative study was undergone to investigate the effect of SBP on quantitative binary partition strategy and fuzzy partition strategy in building a rule base expert system.

The starting point for fuzzy set theory (Zadeh, 1965) is that it is against intuition to model vague concepts such as young and high by crisp intervals. For why would a person be considered as young while he is younger than 40, and on his 40th birthday suddenly loses this status? In the real sense, the transition between being youngAge and middleAge is not abrupt but gradual (Verlinde, Cook, & Boute, 2006). In fuzzy set theory, an element can belong to neighbouring sets each with set membership value in [0,1] depending on the type of membership function used. This value is assigned by each membership function associated with each fuzzy set. For attribute age and its domain $D_{age}$, the mapping of the membership function is $\mu_{age}(x): D_{age} \rightarrow [0,1]$. Fuzzy set is said to provide a smooth change between the boundaries (Kuok, Fu, & Wong, 1998). This is a very good argument for modeling vague concepts by fuzzy sets instead of crisp sets, as many researchers have already used this for the introduction of fuzzy logic to rule discovery (Cock, De Cornelis, & Kerre, 2003; Delgado, Marin, Sanchez, & Vila, 2003; Gyenesei, 2001). In Verlinde, Cook, and Boute (2006) an argument was actually raised against this in favour of binary partition strategy in association rule mining process. The argument was experimentally investigated using data driven approach. However, this argument could not be generalised since expert driven approach is yet to be considered (Verlinde, Cook, & Boute, 2006).

In this paper, Subject Matter Experts (SME), that is, medical doctors’ knowledge (Ajith, 2005)