Distributed Fuzzy Rule Miner (DFRM)

Samane Sharif  
Center of Excellence on Soft Computing and Intelligent Information Processing, Ferdowsi University of Mashhad, Iran  
Sharif.samane@gmail.com

Mohammad R. Akbarzadeh-T.  
Center of Excellence on Soft Computing and Intelligent Information Processing, Ferdowsi University of Mashhad, Iran  
Akbarzadeh@ieee.org

Abstract— Nowadays scalability and capability of parallel execution are the most important characteristics for data mining algorithms due to the growing size of data sets. In this paper, a new distributed framework called DFRM is proposed to extract fuzzy rules from numerical data using a multi-agent approach. These extracted rules can be used for classification and decision making tasks. Scalability, self-organization and uncertainty handling are important characteristics of the proposed system. Scalability and self-organization are provided by autonomous agents in the learning process. Interaction among agents can lead to a more compact fuzzy rule base for decision making. Moreover the training samples are split equally between the agents randomly. Therefore each agent has a partial view of data set. Four UCI data sets are used to evaluate the proposed framework based on accuracy and rule base size. Experimental results show that the resulting distributed classification model maintains acceptable accuracy with fewer rules. In addition, this model is robust against non-availability of training data.

Keywords— Agent; Distribute; Fuzzy Rules extraction; Self-Organization.

I. INTRODUCTION

Today’s datasets are rapidly growing in size, complexity and geographical spread. Efficient mining of such large datasets and maintaining accurate classification models can be difficult as these objectives may be at odds [1]. In other words, a model should enhance system accuracy without increasing its complexity or becoming incomprehensible [1],[2]. In particular, interpretability of the model can be very important in some application domains such as medical decision systems [3], [4]. Fuzzy systems have an advantage in this regard due to their linguistic rule base structure and proximity to human knowledge, but they are unable to learn alone. Simplest and most common choice for adding learning ability is to combine fuzzy systems with neural networks such as [5].

Many researchers have been involved in designing fuzzy systems from databases [6],[7],[8]. But all of these structures use centralized learning. Furthermore, with increased training data, these structures extract more fuzzy rules. In other words, the size of the dataset may reduce performance of data mining algorithms in terms of both efficiency and accuracy. There are three main approaches to deal with this problem [1], [9]:

- Use a priori knowledge to search in small subspaces instead of the complete feasible space
- Use scalable algorithms
- Use data reduction methods

Scalability helps the algorithm succeed even when the problem size becomes larger [1], [10]. Distributed architecture and distributed computing is a solution for this task. The main challenge when using distributed computing is the need for new algorithms that take the architecture into account and have parallel execution ability [11], [12].

In [1] a parallel implementation of genetic algorithms is introduced. This method can extract classification rules in distributed architecture. Each node has a local population and GA runs on local population. Finally, this method extracts some classification rules but these rules are not fuzzy. Therefore the model is not interpretable enough and ability to deal with uncertainties can not be observed. In addition, the genetic algorithm may be time consuming.

Multi-agent approach has potential for parallel execution therefore it is suitable for distributed architecture [13], [14]. Moreover, this approach provides a high learning ability due to the interaction and information sharing between agents. In this paper we try to integrate multi-agent approach and fuzzy systems for building a proper framework to extract fuzzy rules from data sets. This framework can take into account uncertainty, also provide interpretation and suitable accuracy in classification and decision making process. The most important characteristic of this framework is scalability in large data set. The proposed system uses a set of intelligent and autonomous agents with self-organization capability in learning and knowledge acquisition from data. A fuzzy rule base emerges as a result of this self-organized process. Finally, decision making is done based on this fuzzy rule base.

The paper is structured as follows. The next section provides a short definition on multi-agent system and fuzzy rules. Section III describes our proposed framework. The experimental results are explained in Section IV. Finally, the conclusion is presented.

II. HELPFUL HINT

A. Multi Agent System

Multi agent system is one of most attractive paradigms of distributed artificial intelligence. A multi agent system consists of a number of autonomous agents who cooperate and