Incremental Learning Researches on Rough Set Theory: Status and Future

Dun Liu, School of Economics and Management, Southwest Jiaotong University, Chengdu, China

Decui Liang, School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China

ABSTRACT

Rough set theory is an effective tool to deal with information with uncertainty, and has been successfully applied in many fields. Incremental learning as an efficient strategy for data analysis in dynamic environment enables acquiring additional knowledge from new information by using prior knowledge and has drawn the widespread attentions of many scholars. In this paper, the authors discuss the status of incremental learning researches on rough sets and give potential future research directions. The authors first review basic concepts of rough sets and list three variations of information system in the dynamic decision procedures. Then, the authors investigate and summarize the corresponding incremental learning strategies for the three variations with different research viewpoints, respectively. Finally, the authors further tease out the research framework of our work and identify some future possible research directions.

Keywords: Data Analysis, Dynamic Decision Procedures, Dynamics Information System, Incremental Learning, Rough Sets

INTRODUCTION

As an effective mathematical tool, rough set theory (RST) proposed by Pawlak (1982), plays an important role in many fields of data analysis. In RST, it utilizes the lower and upper approximations to deal with an imprecise and uncertainty information (concept), and has been successfully applied in many domains, such as water demand prediction (An et al., 1996), business failure prediction (Dimitras et al., 1999), cluster analysis (Lingras et al., 2010), airline service strategies (Liou et al., 2010), etc. Based on the two approximations, the main researches of RST focus on three aspects, the algebraic structure with two approximation operators (Yao, 1996, 1998); attribute reduction (feature selection) (Dey et al., 2011; Liang et al., 2013; Wang et al., 2013a; Wang et al., 2013b; Xu et al., 2011) and decision rule induction (Blaszczyński & Slowinski, 2003; Fan et al., 2009; Greco et al., 2004; Liang, 2010; Shan & Ziarko, 1992).

DOI: 10.4018/ijrsda.2014010107
Unfortunately, the existing researches on rough sets mainly involve a static environment. In the real life, we may face to a dynamic decision environment (Michalski, 1985). For example, a company always encounters some employees leave or recruits this company. In order to keep the pace of produce process, the employer (manager) need decide how to employ in real time. In addition, accompanied by cheaper storage, evolution of digital data and information collection devices, such as cell phones, laptops, and sensors, the data in today’s is getting increasingly larger and need real-time processing (Demirkan & Delen, 2013). With an overwhelming amount of data arriving at a terabyte and even exabyte scale, the big data (or the massive data) becomes one of the biggest challenges in knowledge discovery. In the context of big data in dynamic environments, the existing information processing technologies based on rough sets should be extended to suit for the updating data (Chen et al., 2012a).

Nowadays, the volume of data is growing at an unprecedented rate. The information system evolves over time (Chan, 1998; Li et al., 2007a; Liu et al., 2009a). As an effective strategy of data analysis in rough sets, incremental learning becomes a very important technology for the updating data in dynamic information systems (Bang & Zeungnam, 1999; Chan, 1998; Michalski, 1985; Li et al., 2007). The advantage of incremental learning is not analyze from scratch but use the original information as much as possible (Zhang et al., 2012a; 2012b). The incremental learning can improve the performance of data analysis (Li et al., 2007; Zheng & Wang, 2004), which provides an effective way to deal with the variations in dynamic environments, and has attracted the attention of more and more researchers in nearly two decades (Luo et al., 2013; Zhang et al., 2014a). Observed by the different research viewpoints in literatures, the incremental learning researches on rough sets are generally categorized into three types: (i) incremental learning of the lower and upper approximations (Li et al., 2007; Chen et al., 2013); (ii) incremental learning on attribute reduction (Dey et al., 2011; Liang et al., 2013; Wang et al., 2013a; Wang et al., 2013b; Xu et al., 2011); (iii) incremental learning on decision rules (Blaszczyński & Slowinski, 2003; Fan et al., 2009; Greco et al., 2004; Liu et al., 2009a; Shan & Ziarko, 1992). In addition, since an information system is consisted of three elements, namely, the objects (instances), the attributes (features), and the domain of attributes’ values, the incremental updating approaches with respect to rough sets can be divided into three parts: (i) the variation of objects; (ii) the variation of attributes; (iii) the variation of attribute values. The variations of objects include the immigration and emigration of objects (Liu et al., 2009a, 2011); the variations of attributes include the deleting outdated attributes or adding new available attributes (Zhang et al., 2012b; Li et al., 2013a); the variations of attribute values mainly reflect in the coarsening and refining of the attribute values (Chen et al., 2012b; Liu et al., 2009b). Specially, Zhang et al. proposed a parallel method for computing rough set approximations for massive data analysis (Zhang et al., 2012c, 2014b), and they further investigated the dynamic variation of the composite information system (Zhang et al., 2014a).

In this paper, we elaborately review the existing incremental learning researches on rough sets according to three types of variations with different research viewpoints, respectively. In the light of the review results, we further point out some future possible directions. Our purpose is to summarize the existing results and help the researchers deeply understand the status and further promote its developments.

THE BASIC CONCEPTS OF ROUGH SETS

In this section, the basic concepts of RST are briefly reviewed. In RST, an information system is utilized to store the related information of objects. An information system is a quadruple $S = (U, A, V, f)$, where the universe $U$ is a finite nonempty set of objects and $A$ is a finite nonempty set of attributes, $V = \bigcup_{a \in A} V_a$, $V_a$ is the domain of attribute $a$, and