Rough Set Approximations in Formal Concept Analysis

Yiyu Yao and Yaohua Chen

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
{yyao, chen115y}@cs.uregina.ca

Abstract. A basic notion shared by rough set analysis and formal concept analysis is the definability of a set of objects based on a set of properties. The two theories can be compared, combined and applied to each other based on definability. In this paper, the notion of rough set approximations is introduced into formal concept analysis. Rough set approximations are defined by using a system of definable sets. The similar idea can be used in formal concept analysis. The families of the sets of objects and the sets of properties established in formal concept analysis are viewed as two systems of definable sets. The approximation operators are then formulated with respect to the systems. Two types of approximation operators, with respect to lattice-theoretic and set-theoretic interpretations, are studied. The results provide a better understanding of data analysis using rough set analysis and formal concept analysis.

1 Introduction

Definability deals with whether and how a set can be defined in order to be analyzed and computed [38]. A comparative examination of rough set analysis and formal concept analysis shows that each of them deals with a particular type of definability. While formal concept analysis focuses on sets of objects that can be defined by conjunctions of properties, rough set analysis focuses on disjunction of properties [33]. The common notion of definability links the two theories together. One can immediately adopt ideas from one to the other [33,34]. On the one hand, the notions of formal concepts and formal concept lattices can be introduced into rough set analysis by considering different types of formal concepts [34]. On the other hand, rough set approximation operators can be introduced into formal concept analysis by considering a different type of definability [8,35]. The combination of the two theories would produce new tools for data analysis.

An underlying notion of rough set analysis is the indiscernibility of objects [12,13]. By modelling indiscernibility as an equivalence relation, one can partition a finite universe of objects into a family of pair-wise disjoint subsets called a partition. The partition provides a granulated view of the universe. An equivalence class is considered as a whole, instead of many individuals, and is viewed as an elementary definable subset. In other words, one can only observe, measure, or characterize the equivalence classes.

J.F. Peters and A. Skowron (Eds.): Transactions on Rough Sets V, LNCS 4100, pp. 285–305, 2006.

The empty set and unions of equivalence classes are also treated as definable subsets. In general, the system of such definable subsets is only a proper subset of the power set of the universe. Consequently, an arbitrary subset of universe may not necessarily be definable. It can be approximated from below and above by a pair of maximal and minimal definable subsets.

Under the rough set approximation, there is a close connection between definability and approximation. A definable set of the universe of objects must have the same approximations [2]. That is, a set of objects is definable if and only if its lower approximation equals to its upper approximation.

Formal concept analysis is developed based on a formal context given by a binary relation between a set of objects and a set of properties. From a formal context, one can construct (objects, properties) pairs known as the formal concepts [6,22]. The set of objects of a formal concept is referred to as the extension, and the set of properties as the intension. They uniquely determine each other. The family of all formal concepts is a complete lattice. The extension of a formal concept can be viewed as a definable set of objects, although in a sense different from that of rough set analysis [33,34]. In fact, the extension of a formal concept is a set of indiscernible objects with respect to the intension. Based on the properties in the intension, all objects in the extension cannot be distinguished. Furthermore, all objects in the extension share all the properties in the intension. The collection of all the extensions, sets of objects, can be considered as a different system of definable sets [35]. An arbitrary set of objects may not be an extension of a formal concept. The sets of objects that are not extensions of formal concepts are regarded as undefinable sets. Therefore, in formal concept analysis, a different type of definability is proposed.

Saquer and Deogun proposed to approximate a set of objects, a set of properties, and a pair of a set of objects and a set of properties, based on a formal concept lattice [16,17]. Hu *et al.* proposed a method to approximate a set of objects and a set of properties by using join- and meet-irreducible formal concepts with respect to set-theoretic operations [8]. However, their formulations are slightly flawed and fail to achieve such a goal. It stems from a mixed-up of the lattice-theoretic operators and set-theoretic operators. To avoid their limitation, a clear separation of two types of approximations is needed. In this paper, we propose a framework to examine the issues of rough set approximations within formal concept analysis. We concentrate on the interpretations and formulations of various notions. Two systems are examined for the definitions of approximations, the formal concept lattice and the system of extensions of all formal concepts.

The rest of the paper is organized as follows. In Section 2, we discuss three formulations of rough set approximations, subsystem based formulation, granule based formulation and element based formulation. In Section 3, formal concept analysis is reviewed. In Section 4, we apply the notion of rough set approximations into formal concept analysis. Two systems of definable sets are established. Based on each system, different definitions of approximations are examined. Section 5 discusses the existing studies and investigates their differences and connections from the viewpoint of approximations.