Formal Rough Concept Analysis*

Jamil Saquer Jitender S. Deogun
deogun(jsaquer)@cse.unl.edu
The Department of Computer Science,
University of Nebraska,
Lincoln, NE 68588, USA

abstract

In this paper, we present a novel approach for approximating concepts in the framework of formal concept analysis. Two main problems are investigated. The first, given a set A of objects (or a set B of features), we want to find a formal concept that approximates A (or B). The second, given a pair (A, B), where A is a set of objects and B is a set of features, the objective is to find formal concepts that approximate (A, B). The techniques developed in this paper use ideas from rough set theory. The approach we present is different and more general than existing approaches.

1 Introduction

Formal concept analysis (FCA) is a mathematical framework developed by Rudolf Wille and his colleagues at Darmstadt/Germany that is useful for representation and analysis of data [8]. A pair consisting of a set of objects and a set of features common to these objects is called a concept. Using the framework of FCA, concepts are structured in the form of a lattice called the concept lattice. The concept lattice is a useful tool for knowledge representation and knowledge discovery [2]. Formal concept analysis has also been applied in the area of conceptual modeling that deals with the acquisition, representation and organization of knowledge [4]. Several concept learning methods have been implemented in [1, 2, 3] using ideas from formal concept analysis.

Not every pair of a set of objects and a set of features defines a concept [8]. Furthermore, we might be faced with a situation where we have a set of features (or a set of objects) and need to find the best concept that approximates these features (or objects). For example, when a physician diagnosis a patient, he finds a disease whose symptoms are the closest to the symptoms that the patient has.

*This research was supported in part by the Army Research Office, Grant No. DAAH04-96-1-0325, under DEPSCoR program of Advanced Research Projects Agency, Department of Defense and by the U.S. Department of Energy, Grant No. DE-FG02-9 7ER1220.

N. Zhong, A. Skowron, S. Ohsuga (Eds.): RSFDGrC'99, LNAI 1711, pp. 91–99, 1999.

In this case we can think of the symptoms as features and diseases as objects. It is therefore of fundamental importance to be able to find concept approximations regardless how little information is available.

In this paper we present a general approach for approximating concepts. We first show how a set of objects (or features) can be approximated by a concept. We prove that our approximations are the best that can be achieved using rough sets. We then extend our approach to approximate a pair of a set of objects and a set of features.

2 Background

Relationships between objects and features in FCA is given in a context which is defined as a triple (G, M, I), where G and M are sets of objects and features (also called attributes), respectively, and $I \subseteq G \times M$. If object g possesses feature m, then $(g, m) \in I$ which is also written as gIm. The set of all common features to a set of objects A is denoted by $\beta(A)$ and defined as $\{m \in M \mid \ gIm \ \forall g \in A\}$. Similarly, the maximal set of objects possessing all the features in a set of features B is denoted by $\alpha(B)$ and given by $\{g \in G \mid \ gIm \ \forall m \in B\}$. A formal concept is defined as a pair (A, B) where $A \subseteq G$, $B \subseteq M$, $\beta(A) = B$ and $\alpha(B) = A$. A is called the extent of the concept and B is called its intent.

Using the above definitions of α and β, it is easy to verify that $A_1 \subseteq A_2$ implies that $\beta(A_1) \supseteq \beta(A_2)$, and $B_1 \subseteq B_2$ implies that $\alpha(B_1) \supseteq \alpha(B_2)$ for every $A_1, A_2 \subseteq G$, and $B_1, B_2 \subseteq M$ [8]. Let $\mathcal{C}(G, M, I)$ denote the set of all concepts of the context (G, M, I) and (A_1, B_1) and (A_2, B_2) be two concepts in $\mathcal{C}(G, M, I)$. (A_1, B_1) is called a subconcept of (A_2, B_2) which is denoted by $(A_1, B_1) \leq (A_2, B_2)$ whenever A_1 is a subset of A_2 (or equivalently B_1 contains B_2). The relation $\leq$ is an order relation on $\mathcal{C}(G, M, I)$.

In the sequel we give an overview of few basic rough set theory terms. Let U be a nonempty finite set of objects called the Universe. Let A be a set of attributes. Associate with each $a \in A$ a set V_a of all possible values of a called its domain. Let $a(x)$ denote the value of the attribute a for element x. Let B be a subset of A (B can be equal to A). A binary relation R^B on U is defined as $xR^By \iff a(x) = a(y) \forall a \in B$. Clearly, R^B is an equivalence relation and thus forms a partition on U. Let $[x]_B$ denote the equivalence class of x with respect to R^B. When B is clear from context, we will write $[x]$ instead of $[x]_B$. Let U/R^B denote the set of all equivalence classes determined by R^B. Equivalence classes of the relation R^B are called B-elementary sets (or just elementary sets). Any finite union of elementary sets is called a definable set.

Given a set $X \subseteq U$, X may not be definable. The relation R^B can be used to characterize X by a pair of definable sets called its lower and upper approximations. The lower and upper approximations of X with respect to R^B (or set of attributes B) are defined as $\underline{B(X)} = \{m \in U \mid [m]_B \subseteq X\}$ and $\overline{B(X)} = \{m \in U \mid [m]_B \cap X \neq \emptyset\}$, respectively. Clearly, the lower approximation of X is the greatest definable set contained in X and the upper approximation