Formal Concept Analysis and Rough Set Theory in Clustering

Ho Tu Bao
Japan Advanced Institute of Science and Technology, Japan
National Institute of Information Technology, Vietnam

Abstract. This paper is concerned with the fundamental role of two mathematical theories in some clustering problems. Formal concept analysis provides the algebraic structure and properties of possible concepts from a given context, and rough set theory provides a mathematical tool to deal with imprecise and incomplete data. Based on these theories, we developed models and algorithms for solving three clustering problems: conceptual clustering, approximate conceptual clustering, and text clustering.

1 Formal Concept Analysis and Rough Set Theory

A theory of concept lattices has been studied under the name *formal concept analysis* (FCA) by Wille and his colleagues [1, 11]. Considers a *context* as a triple $(\mathcal{O}, \mathcal{D}, \mathcal{R})$ where $\mathcal{O}$ be a set of objects, $\mathcal{D}$ be a set of primitive descriptors and $\mathcal{R}$ be a binary relation between $\mathcal{O}$ and $\mathcal{D}$, i.e., $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{D}$ and $(o, d) \in \mathcal{R}$ is understood as the fact that object o has the descriptor d. For any object subset $X \subseteq \mathcal{O}$, the largest tuple common to all objects in X is denoted by $\lambda(X)$. For any tuple $S \in \mathcal{T}$, the set of all objects satisfying S is denoted by $\rho(S)$. A tuple S is *closed* if $\lambda(\rho(S)) = S$. Formally, a *concept* C in the classical view is a pair (X, S), $X \subseteq \mathcal{O}$ and $S \subseteq \mathcal{T}$, satisfying $\rho(S) = X$ and $\lambda(X) = S$. X and S are called *extent* and *intent* of C, respectively. Concept (X_2, S_2) is a *subconcept* of concept (X_1, S_1) if $X_2 \subseteq X_1$ which is equivalent to $S_2 \supseteq S_1$, and (X_1, S_1) is then a *superconcept* of (X_2, S_2).

It was shown that λ and ρ define a Galois connection between the power sets $\wp(\mathcal{O})$ and $\wp(\mathcal{D})$, i.e., they are two order-reversing one-to-one operators. As a consequence, the following properties hold which will be exploited in the learning process:

$$\text{if} \quad S_1 \subseteq S_2 \quad \text{then} \quad \rho(S_1) \supseteq \rho(S_2) \quad \text{and} \quad \lambda\rho(S_1) \subseteq \lambda\rho(S_2)$$
$$\text{if} \quad X_1 \subseteq X_2 \quad \text{then} \quad \lambda(X_1) \supseteq \lambda(X_2) \quad \text{and} \quad \rho\lambda(X_1) \subseteq \rho\lambda(X_2)$$
$$S \subseteq \lambda\rho(S), \quad X \subseteq \rho\lambda(X)$$
$$\rho\lambda\rho = \rho, \quad \lambda\rho\lambda = \lambda, \quad \lambda\rho(\lambda\rho(S)) = \lambda\rho(S)$$
$$\rho(\bigcup_j S_j) = \bigcap_j \rho(S_j), \quad \lambda(\bigcup_j X_j) = \bigcap_j \lambda(X_j)$$

The basic theorem in formal concept analysis [11] states that the set of all possible concepts from a context $(\mathcal{O}, \mathcal{D}, \mathcal{R})$ is a *complete lattice*[1] $\mathcal{L}$, called Galois lattice, in which infimum and supremum can be described as follows:

$$\bigwedge_{t \in T} (X_t, S_t) = (\bigcap_{t \in T} X_t, \lambda\rho(\bigcup_{t \in T} S_t)) \tag{1}$$

$$\bigvee_{t \in T} (X_t, S_t) = (\rho\lambda(\bigcup_{t \in T} X_t), \bigcap_{t \in T} S_t) \tag{2}$$

Rough set theory, a mathematical tool to deal with uncertainty introduced by Pawlak in early 1980s [10]. The starting point of this theory is the assumption that our "view" on elements of a set of objects $\mathcal{O}$ depends on some equivalence relation E on $\mathcal{O}$. An *approximation space* is a pair $(\mathcal{O}, E)$ consisting of $\mathcal{O}$ and an equivalence relation $E \subseteq \mathcal{O} \times \mathcal{O}$.

The key notion of the rough set theory is the *lower* and *upper approximations* of any subset $X \subseteq \mathcal{O}$ which consist of all objects *surely* and *possibly* belonging to X, respectively. The lower approximation $E_*(X)$ and the upper approximation $E^*(X)$ are defined by

$$E_*(X) = \{o \in \mathcal{O} : [o]_E \subseteq X\} \tag{3}$$

$$E^*(X) = \{o \in \mathcal{O} : [o]_E \cap X \neq \emptyset\} \tag{4}$$

where $[o]_E$ denotes the equivalence class of objects indiscernible with o with respect to the equivalence relation E.

2 FCA-based Conceptual Clustering

Conceptual clustering concerns mainly with symbolic data [9]. It does simultaneously two tasks: (i) *hierarchical clustering* (i.e., finding a hierarchy of useful subsets of unlabelled instances); and (ii) *characterization* (i.e., finding an intensional definition for each of these instance subsets). An important feature of conceptual clustering is that a partitioning of data is viewed as

[1] A lattice $\mathcal{L}$ is complete when each of its subsetf X has a least upper bound and a greatest lower bound in $\mathcal{L}$.

Table 1: Scheme of OSHAM conceptual clustering

Input	concept hierarchy H and an existing splittable concept C_k.
Result	H formed gradually.
Top-level	call OSHAM(root concept, $\emptyset$).

1. While C_k is still splittable, find a new subconcept of it that corresponds to the hypothesis minimizing the quality function $q(C_k)$ among η hypotheses generated by the following steps
   (a) Find a "good" attribute-value pair concerning the best cover of C_k.
   (b) Find a closed attribute-value subset S containing this attribute-value pair.
   (c) Form a subconcept C_{k_i} with the intent is S.
   (d) Evaluate the quality function with the new hypothesized subconcept.
   Form intersecting concepts corresponding to intersections of the extent of the new concept with the extent of existing concepts excluding its superconcepts.
2. If one of the following conditions holds then C_k is considered as unsplittable
   (a) There exist not any closed proper feature subset.
   (b) The local instances set C_k^r is too small.
   (c) The local instances set C_k^r is homogeneous enough.
3. Apply recursively the procedure to concepts generated in step 1.

'good' if and only if each cluster has a 'good' conceptual interpretation. In this sense, FCA is a good tool for conceptual clustering as it formalizes the duality between objects and their properties by Galois connections. Based on FCA, we have developed a conceptual clustering method OSHAM with some additional components to the concept representation by extent and intent. The key idea here to enrich the concept representation in FCA by adding several components based on the probabilistic and exemplar views on concepts that allow dealing better with typical or unclear cases in the region boundaries. The conceptual clustering method OSHAM to form a concept hierarchy in the framework of concept lattices is introduced and described in [2]. OSHAM searches to extract a good concept hierarchy by exploiting the structure of Galois lattice of concepts as the hypothesis space.

Instead of characterizing a concept only by its intent and extent, OSHAM represents each concept C_k in a concept hierarchy $\mathcal{H}$ by a 10-tuple

$$< l(C_k), f(C_k), s(C_k), i(C_k), e(C_k), d(C_k), p(C_k), d(C_k^r), p(C_k^r|C_k), q(C_k) > \quad (5)$$

where

- $l(C_k)$ is the level of C_k in $\mathcal{H}$;
- $f(C_k)$ is the list of direct superconcepts of C_k;
- $s(C_k)$ is the list of direct subconcepts of C_k;
- $i(C_k)$ is the intent of C_k (set of all common properties of instances of C_k);
- $e(C_k)$ is the extent of C_k (set of all instances satisfying properties of $i(C_k)$);
- $d(C_k)$ is the dispersion between instances of C_k;
- $p(C_k)$ is the occurrence probability of C_k;
- $d(C_k^r)$ is the dispersion of local instances of C_k which are not classified into subconcepts of C_k;
- $p(C_k^r|C_k)$ is the conditional probability of these unclassified instances of C_k;
- $q(C_k)$ is the quality estimation of splitting C_k into subconcepts C_{k_i}.

Table 1 represents the essential idea of algorithm OSHAM that allows discovering both disjoint and overlapping concepts depending on the user's interests by refining the condition 1.(a) and the intersection operation. In short, OSHAM combines the concept intent, hierarchical structure information, probabilistic estimations and the nearest neighbors of unknown instances. A experimental comparative evaluation of OSHAM is given in [2].

3 Approximate Conceptual Clustering

Kent[7] has pointed out common features between formal concept analysis and rough set theory, and formulated the *rough concept analysis* (RCA). For the sake of simplicity, we restrict ourselves here to present the basic idea of presenting approximate concepts in case of binary attributes where $\mathcal{D}$ is identical to the set $\mathcal{A}$ of all attributes a. Saying that a given formal context $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ is not obtained completely and precisely means that the relation $\mathcal{R}$ is incomplete and imprecise. Let $(\mathcal{O}, E)$ be any approximation space on objects $\mathcal{O}$, we wish to approximate $\mathcal{R}$ in terms of E. The lower approximation $\mathcal{R}_{*E}$ and the upper approximation $\mathcal{R}^{*E}$ of $\mathcal{R}$ w.r.t. E can be defined element-wise as

equal concepts). Note that approximate contexts of $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ in $(\mathcal{O}, E)$ vary according to the equivalence relation E. In [3] we introduce algorithm A-OSHAM for learning approximate concepts in the framework of rough concept analysis. Essentially, A-OSHAM induces a concept hierarchy in which each induced concept is associated with a pair of its lower and upper approximations. A-OSHAM generates concepts with their approximations recursively and gradually, once a level of the hierarchy is formed the procedure is repeated for each class.

4 Document clustering based on a Tolerance Rough Set Model

Given a set $\mathcal{D}$ of M full text documents. Our method of generating a hierarchical structure of this document collection consists of two phases. The first