Induction of Classification Rules by Granular Computing

J.T. Yao Y.Y. Yao

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: {jtyao, yyao}@cs.uregina.ca

Abstract. A granular computing model is used for learning classification rules by considering the two basic issues: concept formation and concept relationships identification. A classification rule induction method is proposed. Instead of focusing on the selection of a suitable partition, i.e., a family of granules defined by values of an attribute, in each step, we concentrate on the selection of a single granule. This leads to finding a covering of the universe, which is more general than partition based methods. For the design of granule selection heuristics, several measures on granules are suggested.

1 Introduction

Classification deals with grouping or clustering of objects based on certain criteria. It is one of the basic learning tasks and is related to concept formation and concept relationship identification. While concept formation involves the construction of classes and description of classes, concept relationship identification involves the connections between classes. These two related issues can be studied formally in a framework that combines formal concept analysis and granular computing (GrC) [9].

There are two aspects of a concept, the intension and extension of the concept [3, 8]. In the granular computing model for knowledge discovery, data mining, and classification, a set of objects are represented using an information table [5, 9]. The intension of a concept is expressed by a formula of the language, while the extension of a concept is represented as the set of objects satisfying the formula. This formulation enables us to study formal concepts in a logic setting in terms of intensions and also in a set-theoretic setting in terms of extensions.

Classification rules obtained from a supervized classification problem capture the relationships between classes defined by a set of attributes and the expert class. In many classical top-down induction methods such as ID3 [6], one attribute is selected in each step [4]. The selected attribute induces a partition that is more informative about the expert classes than other attributes. There are several problems with such attribute centered strategies. Although the selected partition as a whole may be more informative, each equivalence class may not be more informative than equivalence classes produced by another attribute.

Attribute centered strategy may introduce unnecessary attributes in classification rules [2]. In order to resolve such problems, granule centered strategies can be used, in which one granule is defined by an attribute-value pair. An example of granule centered strategies is the PRISM learning algorithm [1, 2].

There has been very little attention paid to granule centered strategies. Based on the granular computing model, we provide a formal and more systematic study of granule centered strategies for the induction of classification rules.

2 A Granular Computing Model

This section presents an overview of the granular computing model [9, 11].

2.1 Information tables

An information table can be formulated as a tuple:

$$S = (U, At, \mathcal{L}, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}),$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, $\mathcal{L}$ is a language defined using attributes in At, V_a is a nonempty set of values for a $\in At$, and $I_a : U \rightarrow V_a$ is an information function. An information table represents all available information and knowledge [5]. In the language $\mathcal{L}$, an atomic formula is given by $a = v$, where $a \in At$ and $v \in V_a$. Formulas can be formed by logical negation, conjunction and disjunction. If a formula ϕ is satisfied by an object x, we write $x \models_S \phi$ or in short $x \models \phi$ if S is understood[9]. If ϕ is a formula, the set $m_S(\phi)$ defined by: $m_S(\phi) = \{x \in U \mid x \models \phi\}$, is called the meaning of ϕ in S. If S is understood, we simply write $m(\phi)$. The meaning of a formula ϕ is the set of all objects having the property expressed by the formula ϕ. A connection between formulas of $\mathcal{L}$ and subsets of U is thus established. With the introduction of language $\mathcal{L}$, we have a formal description of concepts. A concept definable in an information table is a pair $(\phi, m(\phi))$, where $\phi \in \mathcal{L}$. More specifically, ϕ is a description of $m(\phi)$ in S, the intension of concept $(\phi, m(\phi))$, and $m(\phi)$ is the set of objects satisfying ϕ, the extension of concept $(\phi, m(\phi))$. An example information table is given by Table 1, which is adopted from Quinlan [6].

Granulation of a universe involves dividing the universe into subsets or grouping individual objects into clusters. A granule is a subset of the universe. A family of granules that contains every object in the universe is called a granulation of the universe. Partitions and coverings are two simple and commonly used granulations of universe. A partition consists of disjoint subsets of the universe, and a covering consists of possibly overlap subsets. Partitions are a special type of coverings.

Definition 1. *A partition of a finite universe U is a collection of non-empty, and pairwise disjoint subsets of U whose union is U. Each subset in a partition is also called a block or an equivalence granule.*

Object	height	hair	eyes	class
o_1	short	blond	blue	+
o_2	short	blond	brown	-
o_3	tall	red	blue	+
o_4	tall	dark	blue	-
o_5	tall	dark	blue	-
o_6	tall	blond	blue	+
o_7	tall	dark	brown	-
o_8	short	blond	brown	-

Table 1. An information table

Definition 2. *A covering of a finite universe U is a collection of non-empty subsets of U whose union is U. A covering τ of U is said to be a non-redundant covering if any collection of subsets of U derived by deleting one or more granules from τ is not covering.*

By using the language $\mathcal{L}$, we can construct various granules. For an atomic formula $a = v$, we obtain a granule $m(a = v)$. If $m(\phi)$ and $m(\psi)$ are granules corresponding to formulas ϕ and ψ, we obtain granules $m(\phi) \cap m(\psi) = m(\phi \wedge \psi)$ and $m(\phi) \cup m(\psi) = m(\phi \vee \psi)$. In an information table, we are only interested in granules, partitions and coverings that can be described by the language $\mathcal{L}$.

Definition 3. *A subset $X \subseteq U$ is called a definable granule in an information table S if there exists a formula ϕ such that $m(\phi) = X$. A subset $X \subseteq U$ is a conjunctively definable granule in an information table S if there exists a formula ϕ such that ϕ is a conjunction of atomic formulas and $m(\phi) = X$.*

Definition 4. *A partition π is called a conjunctively definable partition if every equivalence class of π is a conjunctively definable granule. A covering τ is called a conjunctively definable covering if every granule of τ is a conjunctively definable granule.*

One can obtain a finer partition by further dividing equivalence classes of a partition. Similarly, one can obtain a finer covering by further decomposing a granule of a covering. This naturally defines a refinement order on the set of all partitions $\Pi(U)$ and the set of all covering $\mathcal{T}(U)$.

Definition 5. *A partition π_1 is refinement of another partition π_2, or equivalently, π_2 is a coarsening of π_1, denoted by $\pi_1 \preceq \pi_2$, if every block of π_1 is contained in some block of π_2. A covering τ_1 is refinement of another covering τ_2, or equivalently, τ_2 is a coarsening of τ_1, denoted by $\tau_1 \preceq \tau_2$, if every granule of τ_1 is contained in some granule of τ_2.*

Since a partition is also a covering, we use the same symbol to denote the refinement relation on partitions and refinement relation on covering. For a covering τ and a partition π, if $\tau \preceq \pi$, we say that τ is a refinement of π. Based on the refinement relation, we can construct multi-level granulations of the universe.

2.2 Measures associated with granules

We introduce and review three types of quantitative measures associated with granules, measures of a single granule, measures of relationships between a pair of granules [9, 10], and measures of relationships between a granule and a family of granules, as well as a pair of family of granules.

The measure of a single granule $m(\phi)$ of a formula ϕ is the *generality* $G(\phi) = |m(\phi)|/|U|$ which indicates the relative size of the granule $m(\phi)$. Given two formulas ϕ and ψ, we introduce a symbol $\Rightarrow$ to connect ϕ and ψ in the form of $\phi \Rightarrow \psi$. The strength of $\phi \Rightarrow \psi$ can be quantified by two related measures [7, 9]. The *confidence* or *absolute support* of ψ provided by ϕ is $AS(\phi \Rightarrow \psi) = |m(\phi \wedge \psi)|/|m(\phi)| = |m(\phi) \cap m(\psi)|/|m(\phi)|$. The *coverage* ψ provided by ϕ is the quantity $CV(\phi \Rightarrow \psi) = |m(\phi \wedge \psi)|/|m(\psi)| = |m(\phi) \cap m(\psi)|/|m(\psi)|$.

Consider now a family of formulas $\Psi = \{\psi_1, \ldots, \psi_n\}$ which induces a partition $\pi(\Psi) = \{m(\psi_1), \ldots, m(\psi_n)\}$ of the universe. Let $\phi \Rightarrow \Psi$ denote the inference relation between ϕ and Ψ. In this case, we obtain the following probability distribution in terms of $\phi \Rightarrow \psi_i$'s:

$$P(\Psi \mid \phi) = \left(P(\psi_1 \mid \phi) = \frac{|m(\phi) \cap m(\psi_1)|}{|m(\phi)|}, \ldots, P(\psi_n \mid \phi) = \frac{|m(\phi) \cap m(\psi_n)|}{|m(\phi)|} \right).$$

The conditional entropy $H(\Psi \mid \phi)$ defined by:

$$H(\Psi \mid \phi) = -\sum_{i=1}^{n} P(\psi_i \mid \phi) \log P(\psi_i \mid \phi), \quad (1)$$

provides a measure that is inversely related to the strength of the inference $\phi \Rightarrow \Psi$. Suppose another family of formulas $\Phi = \{\phi_1, \ldots, \phi_m\}$ define a partition $\pi(\Phi) = \{m(\phi_1), \ldots, m(\phi_m)\}$. The same symbol $\Rightarrow$ is also used to connect two families of formulas that define two partitions of the universe, namely, $\Phi \Rightarrow \Psi$. The strength of this connection can be measured by the conditional entropy:

$$H(\Psi \mid \Phi) = \sum_{j=1}^{m} P(\phi_j) H(\Psi \mid \phi_j) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P(\psi_i \wedge \phi_j) \log P(\psi_i \mid \phi_j), \quad (2)$$

where $P(\phi_j) = G(\phi_j)$. In fact, this is a most commonly used measure for selecting attribute in the construction of decision tree for classification [6].

The measures discussed so far quantified two levels of relationships, i.e., granule level and granulation level. As we will show in the following section, by focusing on different levels, one may obtain different methods for the induction of classification rules.

3 Induction of Classification Rules by Searching Granules

This section first clearly defines the consistent classification problem and then suggests a granule based rule induction method based on the measures discussed in the last section.

3.1 Consistent classification problems

In supervised classification, each object is associated with a unique and predefined class label. Objects are divided into disjoint classes which form a partition of the universe. Suppose an information table is used to describe a set of objects. Without loss of generality, we assume that there is a unique attribute **class** taking class labels as its value. The set of attributes is expressed as $At = F \cup \{\mathbf{class}\}$, where F is the set of attributes used to describe the objects. The goal is to find classification rules of the form, $\phi \Longrightarrow \mathbf{class} = c_i$, where ϕ is a formula over F and c_i is a class label.

Let $\pi_{\mathbf{class}} \in \Pi(U)$ denote the partition induced by the attribute **class**. An information table with a set of attributes $At = F \cup \{\mathbf{class}\}$ is said to provide a consistent classification if all objects with the same description over F have the same class label, namely, if $I_F(x) = I_F(y)$, then $I_{\mathbf{class}}(x) = I_{\mathbf{class}}(y)$.

For a subset $A \subseteq At$, it defines a partition π_A of the universe [5]. The consistent classification problem can be formally defined [11].

Definition 6. *An information table with a set of attributes $At = F \cup \{\mathbf{class}\}$ is a consistent classification problem if and only if $\pi_F \preceq \pi_{\mathbf{class}}$.*

For the induction of classification rules, the partition π_F is not very interesting. In fact, one is interested in finding a subset of attributes from F that also produces the correct classification. It can be easily verified that a problem is a consistent classification problem if and only if there exists a conjunctively definable partition π such that $\pi \preceq \pi_{\mathbf{class}}$. Likewise, the problem is a consistent classification problem if and only if there exists a non-redundant conjunctively definable covering τ such that $\tau \preceq \pi_{\mathbf{class}}$. This leads to different kinds of solutions to the classification problem.

Definition 7. *A partition solution to a consistent classification problem is a conjunctively definable partition π such that $\pi \preceq \pi_{\mathbf{class}}$. A covering solution to a consistent classification problem is a conjunctively definable covering τ such that $\tau \preceq \pi_{\mathbf{class}}$.*

Let X denote a granule in a partition or a covering of the universe, and let $des(X)$ denote its description using language $\mathcal{L}$. If $X \subseteq m(\mathbf{class} = c_i)$, we can construct a classification rule: $des(X) \Rightarrow \mathbf{class} = c_i$. For a partition or a covering, we can construct a family of classification rules. The main difference between a partition solution and a covering solution is that an object is only classified by one rule in a partition based solution, while an object may be classified by more than one rule in a covering based solution.

Consider the consistent classification problem of Table 1. We have the partition by **class**, a conjunctively defined partition π, and a conjunctively non-redundant covering τ:

$$\begin{aligned}
\pi_{\mathbf{class}}: &\quad \{\{o_1, o_3, o_6\}, \{o_2, o_4, o_5, o_7, o_8\}\}, \\
\pi: &\quad \{\{o_1, o_6\}, \{o_2, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\}, \\
\tau: &\quad \{\{o_1, o_6\}, \{o_2, o_7, o_8\}, \{o_3\}, \{o_4, o_5, o_7\}\}.
\end{aligned}$$

Clearly, $\pi \preceq \pi_{\mathbf{class}}$ and $\tau \preceq \pi_{\mathbf{class}}$. A set of classification rules of π may include rules such as "$\mathbf{hair} = \text{blond} \wedge \mathbf{eyes} = \text{blue} \Longrightarrow \mathbf{class} = +$".

3.2 Construction of a granule network

The top-down construction of a decision tree for classification searches for a partition solution to a classification problem. The induction process can be briefly described as follows. Based on a measure of connection between two partitions such as $H(\Psi \mid \Phi)$, one selects an attribute to divide the universe into a partition [6]. If an equivalence class is not a subset of a user defined class, it is further divided by using another attribute. The process continues until one finds a decision tree that correctly classifies all objects. Each node of the decision tree is labelled by an attribute, and each branch is labelled by a value of the parent attribute.

When we search a covering solution, we can not immediately use a decision tree to represent the results. We modify the decision tree method and introduce the concept of granule network. In a granule network, each node is labelled by a subset of objects. The arc leading from a larger granule to a smaller granule is labelled by an atomic formula. In addition, the smaller granule is obtained by selecting those objects of the larger granule that satisfy the atomic formula. The family of the smallest granules thus forms a conjunctively definable covering of the universe.

Atomic formulas define *basic* granules, which serve as the basis for the granule network. The pair $(a = v, m(a = v))$ is called a basic concept. Each node in the granule network is a conjunction of some basic granules, and thus a conjunctively definable granule. The granule network for a classification problem can constructed by a top-down search of granules. Figure 1 outline an algorithm for the construction of a granule network.

The two importance issues of the algorithm is the evaluation of the fitness of each basic concept and the modification of existing partial granule network. The algorithm is basically a heuristic search algorithm. The measures discussed in the last section can be used to define different fitness functions. This will be topics of our future research. In the rest of this section, we will use an example to illustrate the basic ideas.

Table 2 summarizes the measures of basic concepts with respect to the partition $\pi_{\mathbf{class}}$. There are three granules which are subset of one of class values, i.e., $\{o_3\} \subseteq (\mathbf{class} = +)$, $\{o_4, o_5, o_7\} \subseteq (\mathbf{class} = -)$ and $\{o_2, o_7, o_8\} \subseteq (\mathbf{class} = -)$. The values of entropy of these granules are the minimum, i.e., 0. The generality of last two granules are among the highest, so they are chosen first. One of possible orders of selection of these granules is $m(\mathbf{hair} = \text{dark})$, $m(\mathbf{eyes} = \text{brown})$ and then $m(\mathbf{hair} = \text{red})$. These three granules cannot cover the universe, i.e., they are not a covering solution to the classification problem. We will further analyze on other granules in order to find a set of granules that cover the whole universe. With the consideration of non-redundant covering, if adding candidate covering granule cannot form a non-redundant covering, we

(1) Construct the family of basic concept with respect to atomic formulas:

$$BC(U) = \{(a = v, m(a = v)) \mid a \in F, v \in V_a\}.$$

(2) Set the unused basic concepts to the set of basic concepts:

$$UBC(U) = BC(U).$$

(3) Set the granule network to $GN = (\{U\}, \emptyset)$, which is a graph consists of only one node and no arc.

(4) While the set of smallest granules in GN is not a covering solution of the classification problem **do** the following:

- **(4.1) Compute** the fitness of each unused basic concept.
- **(4.2) Select** the basic concept $C = (a = v, m(a = v))$ with maximum value of fitness.
- **(4.3) Set** $UBC(U) = UBC(U) - \{C\}$.
- **(4.4) Modify** the granule network GN by adding new nodes which are the intersection of $m(a = v)$ and the original nodes of GN; **connect** the new nodes by arcs labelled by $a = v$.

Fig. 1. An Algorithm for constructing a granule network

Formula	Granule	Generality	Confidence		Coverage		Entropy
			+	-	+	-	
height = short	$\{o_1, o_2, o_8\}$	3/8	1/3	2/3	1/3	2/5	0.92
height = tall	$\{o_3, o_4, o_5, o_6, o_7\}$	5/8	2/5	3/5	2/3	3/5	0.97
hair = blond	$\{o_1, o_2, o_6, o_8\}$	4/8	2/4	2/4	2/3	2/5	1.00
hair = red	$\{o_3\}$	1/8	1/1	0/1	1/3	0/5	0.00
hair = dark	$\{o_4, o_5, o_7\}$	3/8	0/3	3/3	0/3	3/5	0.00
eyes = blue	$\{o_1, o_3, o_4, o_5, o_6\}$	5/8	3/5	2/5	3/3	2/5	0.97
eyes = brown	$\{o_2, o_7, o_8\}$	3/8	0/3	3/3	0/3	3/5	0.00

Table 2. Basic granules and their measures

will not choose this granule even if other measure are in favor of this granule. If many objets in a candidate granule are already in granule network, this granule will not be chosen. Granule $m($**hair** $=$ blond$)$ is considered the most suitable granule in this example and thus will be chosen. Now we have a covering $\tau = \{\{o_4, o_5, o_7\}, \{o_2, o_7, o_8\}, \{o_3\}, \{o_1, o_2, o_6, o_8\}\}$ which covers the universe. Obliviously, the objects in $m($**hair** $=$ blond$)$ are not belong to the same class, therefore a further granulation to this granule will be conducted in order to find smaller definable granules. Considering the generality and non-redundant covering, granule $m($**hair** $=$ blond$\wedge$**eyes** $=$ blue$) = \{o_1, o_6\}$ became the most suitable granule of a covering solution.

4 Conclusion

A consistent classification problem can be modelled as a search for a partition or a covering defined by a set of attribute values. In this paper, we apply a granular computing model for solving classification problems. The notion of granule network is used to represent the classification knowledge. The set of the smallest granules in the granule network forms a covering of the universe. Although the classification rules may have overlaps with each other, they may be shorter than the rules obtained from classical decision tree methods. This stem from the fact that at each step, only the most suitable granule defined by an attribute-value pair is selected, instead of a partition.

The main contribution of the paper is the formal development of the granule centered strategy for classification. As future research, we will study various heuristics defined using the measures suggested in this paper, the evaluation of the proposed algorithm using real world data sets.

References

1. Bramer, M.A.,"Automatic induction of classification rules from examples using N-PRISM", *Research and Development in Intelligent Systems XVI*, Springer-Verlag, pp.99-121, 2000.
2. Cendrowska, J., "PRISM: an algorithm for inducing modular rules", *International Journal of Man-Machine Studies*, **27**, 349-370, 1987.
3. Demri, S. and Orlowska, E., "Logical analysis of indiscernibility", in: *Incomplete Information: Rough Set Analysis*, Orlowska, E. (Ed.), Physica-Verlag, Heidelberg, pp.347-380, 1998.
4. Ganascia, J.-G., "TDIS: an algebraic formalization", *Proceedings of IJCAI 1993*, pp1008-1015, 1993.
5. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.
6. Quinlan, J.R. "Learning efficient classification procedures and their application to chess end-games", in: *Machine Learning: An Artificial Intelligence Approach*, Vol. 1, Michalski, J.S. *et al.*(Eds.), Morgan Kaufmann, pp.463-482, 1983.
7. Tsumoto, S., "Modelling medical diagnostic rules based on rough sets", *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 1424*, Springer-Verlag, Berlin, pp.475-482, 1998.
8. Wille, R., "Concept lattices and conceptual knowledge systems", *Computers Mathematics with Applications*, **23**, 493-515, 1992.
9. Yao, Y.Y, "On Modeling data mining with granular computing", *Proceedings of COMPSAC 2001*, pp.638-643, 2001.
10. Yao, Y.Y. and Zhong, N. An analysis of quantitative measures associated with rules, *Proceedings of PAKDD'99*, pp.479-488, 1999.
11. Yao, Y.Y. and Yao, J. T., "Granular computing as a basis for consistent classification problems", *Proceedings of PAKDD'02 Workshop on Toward the Foundation of Data Mining*, pp.101-106, 2002.
12. Zadeh, L.A., "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic", *Fuzzy Sets and Systems*, **19**, 111-127, 1997.