

# Learning with Hierarchical Quantitative Attributes by Fuzzy Rough Sets

Tzung-Pei Hong<sup>1</sup> Yan-Liang Liou<sup>1</sup> Shyue-Liang Wang<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, National University of Kaohsiung

<sup>2</sup>Department of Computer Science, New York Institute of Technology

## Abstract

This paper proposes an approach to deal with the problem of producing a set of cross-level fuzzy certain and possible rules from examples with hierarchical and quantitative attributes. The proposed approach combines the rough-set theory and the fuzzy-set theory to learn. Some pruning heuristics are adopted in the proposed algorithm to avoid unnecessary search. A simple example is also given to illustrate the proposed approach.

**Keywords:** machine learning, rough set, hierarchical value, quantitative value.

## 1. Introduction

Recently, the rough-set theory has been used in reasoning and knowledge acquisition for expert systems [1][3][6]. It was proposed by Pawlak in 1982 [7] with the concept of equivalence classes as its basic principle. Several applications and extensions of the rough-set theory have also been proposed. Training data in real-world applications sometimes consist of quantitative values. Fuzzy-set concepts are often used to represent quantitative data expressed in linguistic terms and membership functions in intelligent systems because of its simplicity and similarity to human reasoning. Dubois and Prade combined rough sets and fuzzy sets together in order to get a more accurate account of imperfect information [2]. In the past, we also proposed a method which combined rough-set theory and fuzzy-set theory to deal with the problem of producing a set of certain and possible rules from quantitative data [4].

Attributes are usually organized into hierarchy in real applications. Deriving rules on multiple concept levels may thus lead to the discovery of more general and important knowledge from data. This paper thus extends our previous approach to deal with the problem of producing a set of cross-level maximally general fuzzy certain and possible rules from examples with hierarchical and quantitative attributes [5]. Some pruning heuristics are adopted in the proposed

algorithm to avoid unnecessary search. Rule effectiveness for future data is also derived from these membership values.

## 2. Rough Set

The rough-set theory, proposed by Pawlak in 1982 [7], can serve as a new mathematical tool for dealing with data classification problems. Formally, let  $U$  be a set of training examples (objects),  $A$  be a set of attributes describing the examples,  $C$  be a set of classes, and  $V_j$  be a value domain of an attribute  $A_j$ . Also let  $v_j^{(i)}$  be the value of attribute  $A_j$  for the  $i$ -th object  $Obj^{(i)}$ . When two objects  $Obj^{(i)}$  and  $Obj^{(k)}$  have the same value of attribute  $A_j$ , (that is,  $v_j^{(i)} = v_j^{(k)}$ ),  $Obj^{(i)}$  and  $Obj^{(k)}$  are said to have an indiscernibility relation (or an equivalence relation) on attribute  $A_j$ . Also, if  $Obj^{(i)}$  and  $Obj^{(k)}$  have the same values for each attribute in subset  $B$  of  $A$ ,  $Obj^{(i)}$  and  $Obj^{(k)}$  are also said to have an indiscernibility (equivalence) relation on attribute set  $B$ . These equivalence relations thus partition the object set  $U$  into disjoint subsets, denoted by  $U/B$ , and the partition including  $Obj^{(i)}$  is denoted  $B(Obj^{(i)})$ . The sets of equivalence classes for subset  $B$  are referred to as *B-elementary sets*.

Let  $X$  be an arbitrary subset of the universe  $U$ , and  $B$  be an arbitrary subset of attribute set  $A$ . The lower and the upper approximations for  $B$  on  $X$ , denoted  $B_*(X)$  and  $B^*(X)$  respectively, are defined as follows:

$$B_*(X) = \{x \mid x \in U, B(x) \subseteq X\}, \text{ and}$$

$$B^*(X) = \{x \mid x \in U \text{ and } B(x) \cap X \neq \emptyset\}.$$

Elements in  $B_*(x)$  can be classified as members of set  $X$  with full certainty using attribute set  $B$ , so  $B_*(x)$  is called the lower approximation of  $X$ . Similarly, elements in  $B^*(x)$  can be classified as members of the set  $X$  with only partial certainty using attribute set  $B$ , so  $B^*(x)$  is called the upper approximation of  $X$ . After the lower and the upper approximations have been found, the rough-set theory can then be used to derive both certain and uncertain information and induce certain and possible rules from them.

Since each certain rule derived from the lower approximation will also be derived from the upper approximation. It thus causes redundant derivation and

wastes computational time. The boundary approximation is thus used in this paper. For convenience, the symbol  $B^*(X)$  is used from here on to represent the boundary approximation of attribute subset  $B$  on  $X$ , instead of the upper approximation. It is defined as follows:

$$B^*(X) = \{x | x \in U, B(x) \cap X \neq \emptyset \text{ and } B(x) \not\subseteq X\}.$$

### 3. Hierarchical Attributes

Most of the previous studies on rough sets focused on finding certain rules and possible rules on the single concept level. However, hierarchical attributes are usually predefined in real-world applications and can be represented by hierarchy trees. Terminal nodes on the trees represent actual attribute values appearing in training examples; internal nodes represent value clusters formed from their lower-level nodes. Deriving rules on multiple concept levels may lead to the discovery of more general and important knowledge from data.

The concept of equivalence classes in the rough set theory makes it very suitable for finding cross-level certain and possible rules from training examples with hierarchical values. The equivalence class of a non-terminal-level attribute value for attribute  $A_j$  can be easily found by the union of its underlying terminal-level equivalence classes for  $A_j$ . Also, the equivalence class of a cross-level attribute value combination for more than two attributes can be derived from the intersection of the equivalence classes of its single attribute values. In this paper, we will propose a fuzzy-rough learning algorithm for deriving cross-level certain and possible rules from training examples with hierarchical quantitative attribute values.

### 4. Fuzzy-Rough Sets

In the past, we proposed a method which combined rough-set theory and fuzzy-set theory to deal with the classification problem [4]. It is extended here to find a set of cross-level maximally general fuzzy certain and possible rules from examples with hierarchical and quantitative attributes. Some definitions about fuzzy approximations are introduced below.

When the same linguistic term  $R_{jk}$  of an attribute  $A_j$  exists in two fuzzy objects  $Obj^{(i)}$  and  $Obj^{(r)}$  with membership values  $f_{jk}^{(i)}$  and  $f_{jk}^{(r)}$  equal to or larger than a certain  $\alpha$  value,  $Obj^{(i)}$  and  $Obj^{(r)}$  are said to have a fuzzy  $\alpha$ -indiscernibility relation (or fuzzy  $\alpha$ -equivalence relation) on attribute  $A_j$  with membership value  $\min(f_{jk}^{(i)} \wedge f_{jk}^{(r)})$ . Also, if the same linguistic terms of an attribute subset  $B$  exist in both  $Obj^{(i)}$  and

$Obj^{(r)}$  with membership values equal to or larger than  $\alpha$ ,  $Obj^{(i)}$  and  $Obj^{(r)}$  are said to have a fuzzy  $\alpha$ -indiscernibility relation (or a fuzzy  $\alpha$ -equivalence relation) on attribute subset  $B$  with a membership value equal to the minimum of all the membership values. These fuzzy  $\alpha$ -equivalence relations thus partition the fuzzy object set  $U$  into several fuzzy subsets that may overlap, and the result is denoted by  $U/B$ . The set of partitions, based on  $B$  and including  $Obj^{(i)}$ , is denoted  $B(Obj^{(i)})$ . Thus,  $B(Obj^{(i)}) = \{(B_1(Obj^{(i)}), \mu_{B_1}(Obj^{(i)})), \dots, (B_r(Obj^{(i)}), \mu_{B_r}(Obj^{(i)}))\}$ , where  $r$  is the number of partitions included in  $B(Obj^{(i)})$ ,  $B_j(Obj^{(i)})$  is the  $j$ -th partition in  $B(Obj^{(i)})$ , and  $\mu_{B_j}(Obj^{(i)})$  is the membership value of the  $j$ -th partition.

Fuzzy  $\alpha$ -lower and fuzzy  $\alpha$ -boundary approximations are defined below. Let  $X$  be an arbitrary subset of the universe  $U$ , and  $B$  be an arbitrary subset of the attribute set  $A$ . The fuzzy lower and the fuzzy boundary approximations under the threshold value  $\alpha$  for  $B$  on  $X$ , denoted  $B_*(X)$  and  $B^*(X)$ , respectively, are defined as follows:

$$B_*(X) = \{(B_k(x), \mu_{B_k}(x)) | x \in U, B_k(x) \subseteq X, 1 \leq k \leq |B(x)|\}, \text{ and}$$

$$B^*(X) = \{(B_k(x), \mu_{B_k}(x)) | x \in U, B(x) \cap X \neq \emptyset \text{ and } B(x) \not\subseteq X, 1 \leq k \leq |B(x)|\}.$$

Elements in  $B_*(x)$  can be classified as members of set  $X$  with full certainty using attribute set  $B$ . Also, their membership values may be considered effectiveness measures of fuzzy lower approximations for future data. A low membership value with a fuzzy lower approximation means the lower approximation will have a low tolerance (or effectiveness) on future data.

On the other hand, elements in  $B^*(x)$  can be classified as members of set  $X$  with only partial certainty using attribute set  $B$ , and their certainty degrees can be calculated from the membership values of elements in the boundary approximations.

### 5. The Proposed Algorithm

The details of the proposed learning algorithm are described as follows.

#### ***A learning algorithm processing hierarchical and quantitative attributes by fuzzy rough sets:***

Input: A quantitative data set with  $n$  objects,  $m$  hierarchical attributes, a set of membership functions, and a value  $\alpha$  for  $\alpha$ -cut.

Output: A set of cross-level certain and possible rules.

Step 1: Partition the object set into disjoint subsets according to class labels. Denote each set of

objects belonging to the same class  $C_l$  as  $X_l$ .

Step 2: Transform the quantitative value  $v_j^{(i)}$  of each object  $Obj^{(i)}$ ,  $i = 1$  to  $n$ , for each appearing terminal-level node  $A_j$  in  $Obj^{(i)}$  into a fuzzy set  $f_j^{(i)}$ , represented as:

$$\left( \frac{f_{j_1}^{(i)}}{R_{j_1}} + \frac{f_{j_2}^{(i)}}{R_{j_2}} + \dots + \frac{f_{j_l}^{(i)}}{R_{j_l}} \right),$$

using the given membership functions, where  $R_{jk}$  is the  $k$ -th fuzzy region of a terminal-level attribute node  $A_j$ ,  $f_{jk}^{(i)}$  is  $v_j^{(i)}$ 's fuzzy membership value in region  $R_{jk}$ , and  $l (=|A_j|)$  is the number of fuzzy regions for  $A_j$ .

Step 3: Remove the linguistic term  $R_{jk}$  from a fuzzy set  $f_j^{(i)}$  with its  $f_{jk}^{(i)} < \alpha$ .

Step 4: Find the terminal-level fuzzy elementary sets of single attributes using fuzzy operations based on the definitions in Section 4.

Step 5: Set  $l = 1$ , where  $l$  is used to count the number of classes currently being processed.

Step 6: Compute the fuzzy lower approximation of each single terminal-level attribute node  $A_j^l$  for class  $X_l$  as:

$$A_{j^*}^l(X_l) = \{(A_{jk}^l(x), \mu_{A_{jk}^l}(x)) \mid x \in U, A_{jk}^l(x) \subseteq X_l, \\ 1 \leq k \leq |A_j^l(x)|\},$$

where  $A_{jk}^l(x)$  is the terminal-level fuzzy equivalence class including object  $x$  and derived from the  $k$ -th fuzzy region of attribute node  $A_j^l$ , and  $|A_j^l(x)|$  is the number of fuzzy regions for  $A_j^l$ .

Step 7: Compute the fuzzy boundary approximation of each single terminal-level attribute node  $A_j^l$  for class  $X_l$  as:

$$A_j^*(X_l) = \{(A_{jk}^*(x), \mu_{A_{jk}^*}(x)) \mid x \in U, A_{jk}^*(x) \cap X_l \neq \emptyset, \\ A_{jk}^*(x) \not\subseteq X_l, 1 \leq k \leq |A_j^*(x)|\}.$$

Step 8: Compute the fuzzy lower and fuzzy boundary approximations of each single non-terminal-level attribute node  $A_j^{nt}$  for class  $X_l$  from the terminal level to the root level by the following substeps:

- Derive the fuzzy equivalence class  $A_j^{nt_{ki}}$  of the  $i$ -th fuzzy region for the non-terminal-level attribute node  $A_j^{nt}$  on level  $k$  by the union of the equivalence classes of the same fuzzy regions in its underlying nodes.
- Put the equivalence class of  $A_j^{nt_{ki}}$  into the  $k$ -level lower approximation for non-terminal-level attribute node  $A_j^{nt}$  if all its underlying equivalence classes of the same fuzzy regions are in the  $(k+1)$ -level lower approximation for attribute node  $A_j^{nt}$ .
- Put the equivalence class of  $A_j^{nt_{ki}}$  in the  $k$ -level boundary approximation for attribute node  $A_j^{nt}$

if at least one of its underlying equivalence classes of the same fuzzy regions is in the  $(k+1)$ -level lower or boundary approximation for attribute node  $A_j^{nt}$ .

Step 9: Set  $q = 2$ , where  $q$  is used to count the number of attributes currently being processed.

Step 10: Compute the fuzzy lower and the fuzzy boundary approximations of each attribute set  $B_j$  with  $q$  attributes (on any levels) for class  $X_l$  from the terminal level to the root level by the following substeps. Only a fuzzy region from an attribute can be put in the combination.

- Skip the combinations which have the equivalence class of at least one of its subsets already in the lower approximation for  $X_l$ .
- Derive the equivalence class of each remaining combination by the intersection of the equivalence classes of its corresponding single attribute regions. Set the membership values for the derived equivalence class as the minimum of their membership values.
- Put the equivalence class  $B_j(x)$  of each combination in substep (b) into the lower approximation class  $X_l$  if  $B_j(x) \subseteq X_l$ .
- Put the equivalence class  $B_j(x)$  of each combination in substep (b) into the boundary approximation class if  $B_j(x) \cap X_l \neq \emptyset$  and  $B_j(x) \not\subseteq X_l$ .

Step 11: Set  $q = q + 1$  and repeat Step 10 until  $q > m$ .

Step 12: Derive the certain rules from the fuzzy lower approximation  $B_*(X_l)$  of any subset  $B$ , and set the membership values of elements in the lower approximation as effectiveness measures for future data.

Step 13: Remove certain rules more specific than others and keep the more general ones.

Step 14: Derive the possible rules from the fuzzy upper approximation  $B^*(X_l)$  of any subset  $B$ , set the membership values of elements in the upper approximation as effectiveness measures for future data, and calculate the plausibility measure of each rule for  $B_k(x)$  as:

$$p(B_k(x)) = \frac{\sum_{x \in (B_k(x) \cap X_l)} \mu_{B_k}(x)}{\sum_{x \in B_k(x)} \mu_{B_k}(x)}.$$

Step 15: Remove possible rules with condition parts more specific and plausibility values equal to or smaller than those of some other possible or certain rules.

Step 16: Set  $l = l + 1$  and repeat Steps 6 to 14 until  $l > c$ .

Step 17: Output the fuzzy certain rules and fuzzy possible rules.

## 6. An Example

Assume there are two decision attributes  $A = \{Transport, Residence\}$ , and a class attribute  $C = \{Consumption Style\}$ . Both the attributes have a simple hierarchy as shown in Figures 1. There are two levels of hierarchical attribute values for attributes *Transport* and *Residence*. Only values of terminal nodes can appear in training examples. Assume the class has only two possible values: {High (H), Low (L)}.

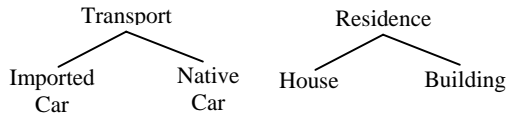


Figure 1: Hierarchy of Transport and Residence

Table 1 contains ten objects  $U = \{Obj^{(1)}, Obj^{(2)}, \dots, Obj^{(10)}\}$  for learning. Each training example has a quantity representing the cost (unit: ten thousands NT dollars) for the attribute value. Also the membership functions of each terminal attribute node are shown in the Figure 2.

Table 1: The training data used in this example

|              | <i>Transport</i> | <i>Residence</i> | <i>Consumption Style</i> |
|--------------|------------------|------------------|--------------------------|
| $Obj^{(1)}$  | Imported car:190 | House:1700       | High                     |
| $Obj^{(2)}$  | Imported car:210 | House:1800       | High                     |
| $Obj^{(3)}$  | Native car:50    | Building:230     | Low                      |
| $Obj^{(4)}$  | Native car:42    | Building:220     | Low                      |
| $Obj^{(5)}$  | Native car:45    | House:350        | Low                      |
| $Obj^{(6)}$  | Native car:40    | House:330        | Low                      |
| $Obj^{(7)}$  | Native car:145   | Building:1300    | High                     |
| $Obj^{(8)}$  | Native car:130   | Building:1400    | High                     |
| $Obj^{(9)}$  | Imported car:50  | Building:250     | Low                      |
| $Obj^{(10)}$ | Imported car:60  | Building:270     | Low                      |

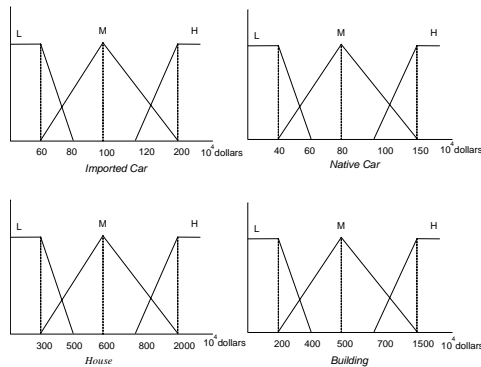


Figure 2: Membership functions of the attributes

In this example, six linguistic certain rules and fifteen linguistic possible rules are derived. For example, the certain rule “If *Transport* is *High\_Cost*

*Imported\_Car* then *Consumption Style* is *High*, with future effectiveness = 0.58” is one of them.

## 7. Conclusion

In this paper, we have proposed a new learning algorithm based on fuzzy rough sets to find fuzzy cross-level certain and possible rules from training data with hierarchical and quantitative attribute values. The proposed method adopts the concept of fuzzy equivalence classes and uses some pruning heuristics. The fuzzy rules derived can be used to infer results from a new event with both terminal and non-terminal attribute nodes.

## Acknowledgement

This research was supported by the National Science Council of the Republic of China under contract NSC 94-2213-E-390-005.

## References

- [1] C. C. Chan, “A rough set approach to attribute generalization in data mining,” *Journal of Information Sciences*, 107, 1998, pp. 169-176.
- [2] D. Dubois and H. Prade, “Putting rough sets and fuzzy sets together,” *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, 1992, pp.203-232.
- [3] J. W. Grzymala-Busse, “Knowledge acquisition under uncertainty: A rough set approach,” *Journal of Intelligent Robotic Systems*, 1, 1988, pp. 3-16.
- [4] T. P. Hong, T. T. Wang and S. L. Wang, “Knowledge acquisition from quantitative data using the rough-set theory,” *Intelligent Data Analysis*, Vol. 4, 2000, pp. 289-304.
- [5] T. P. Hong, C. E. Lin, J. H. Lin and S. L. Wang, “Learning from hierarchical attribute values by rough sets,” *The Third International Conference on Intelligent Systems Design and Applications*, 2003, pp. 559-568.
- [6] E. Orłowska, “Reasoning with incomplete information: rough set based information logics,” in: V. Alagar, S. Bergler and F. Q. Dong (eds.), *Incompleteness and Uncertainty in Information Systems*, Springer, pp. 16-33.
- [7] Z. Pawlak, “Rough set,” *International Journal of Computer and Information Sciences*, pp. 341-356, 1982.