Towards Fuzzification of Attribute-Oriented Generalization

Shusaku Tsumoto and T.Y. Lin

Shusaku Tsumoto
Department of Medicine Informatics,
Shimane Medical University, School of Medicine
89-1 Enya-cho Izumo City,
Shimane 693-8501 Japan
tsumoto@computer.org

T.Y. Lin
Department of Mathematics and Computer Science
San Jose State University
129S 10th St, San Jose,
California 95192-0103 USA
tylin@mathcs.sjsu.edu

Abstract

Conventional studies on rule discovery and rough set methods mainly focus on acquisition of rules, the targets of which have mutually exclusive supporting sets. However, mutual exclusiveness does not always hold in real-world databases, where conventional probabilstic approaches cannot be applied. In this paper, first, we show that these phenomena are easily found in data mining contexts: when we apply attribute-oriented generalization to attributes in databases, generalized attributes will have fuzziness for classification. Then, we show that in general these contexts should be solved by using context-free fuzzy sets.

1 Introduction

Conventional studies on machine learning[7], rule discovery[2] and rough set methods[4, 9, 10] mainly focus on acquisition of rules, the targets of which have mutually exclusive supporting sets. Supporting sets of target concepts form a partition of the universe, and each method search for sets which covers this partition. Especially, Pawlak's rough set theory shows the family of sets can form an approximation of the partition of the universe. These ideas can easily extend into probabilistic contexts, such as shown in Ziarko's variable precision rough set model[12]. However, mutual exclusiveness of the target does not always hold in real-world databases, where conventional probabilstic approaches cannot be applied.

In this paper, first, we show that these phenomena are easily found in data mining contexts: when we apply attribute-oriented generalization to attributes in databases, generalized attributes will have fuzziness for classification. In this case, we have to take care about the conflicts between each attributes, which can be viewed as a problem with multiple membership functions. Then, we show that in general these contexts should be analyzed by using context-free fuzzy sets introduced by Lin[3].

2 Attribute-Oriented Generalization and Fuzziness

In this section, first, a probabilistic rule is defined by using two probabilistic measures. Then, attribute-oriented generalization is introduced as tranforming rules.

2.1 Probabilistic Rules

2.1.1 Accuracy and Coverage

In the subsequent sections, we adopt the following notations, which is introduced in [6].

Let U denote a nonempty, finite set called the universe and A denote a nonempty, finite set of attributes, i.e., $a : U \rightarrow V_a$ for $a \in A$, where V_a is called the domain of a, respectively.Then, a decision table is defined as an information system, $A = (U, A \cup \{d\})$.

The atomic formulas over $B \subseteq A \cup \{d\}$ and V are expressions of the form $[a = v]$, called descriptors over B, where $a \in B$ and $v \in V_a$. The set $F(B, V)$ of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation.

For each $f \in F(B, V)$, f_A denote the meaning of f in A, i.e., the set of all objects in U with property f, defined inductively as follows.

1. If f is of the form $[a = v]$ then, $f_A = \{s \in U | a(s) = v\}$
2. $(f \wedge g)_A = f_A \cap g_A$; $(f \vee g)_A = f_A \vee g_A$; $(\neg f)_A = U - f_a$

By the use of this framework, classification accuracy and coverage, or true positive rate is defined as follows.

Definition 1
Let R and D denote a formula in $F(B, V)$ and a set of objects which belong to a decision d. Classification accuracy and coverage(true positive rate) for $R \rightarrow d$ is defined as:

$$\begin{aligned} \alpha_R(D) &= \frac{|R_A \cap D|}{|R_A|}(= P(D|R)), \text{ and} \\ \kappa_R(D) &= \frac{|R_A \cap D|}{|D|}(= P(R|D)), \end{aligned}$$

where $|A|$ denotes the cardinality of a set A, $\alpha_R(D)$ denotes a classification accuracy of R as to classification of D, and $\kappa_R(D)$ denotes a coverage, or a true positive rate of R to D, respectively. [1]

2.1.2 Definition of Rules

By the use of accuracy and coverage, a probabilistic rule is defined as:

$R \stackrel{\alpha,\kappa}{\rightarrow} d \;\; s.t. \;\; R = \wedge_j \vee_k [a_j = v_k], \; \alpha_R(D) \geq \delta_\alpha, \; \kappa_R(D) \geq \delta_\kappa.$

This rule is a kind of probabilistic proposition with two statistical measures, which is an extension of Ziarko's variable precision model(VPRS) [12].[2]

It is also notable that both a positive rule and a negative rule are defined as special cases of this rule, as shown in the next subsections.

2.2 Attribute-Oriented Generalization

Rule induction methods regard a database as a decision table[4] and induce rules, which can be viewed as reduced decision tables. However, those rules extracted from tables do not include information about attributes and they are too simple. In practical situation, domain knowledge of attributes is very important to gain the comprehensability of induced knowledge, which is one of the reasons why databases are implemented as relational-databases[1]. Thus, reinterpretation of induced rules by using information about attributes is needed to acquire comprehensive rules. For example, terolism, cornea, antimongoloid slanting of palpebral fissures, iris defects and long eyelashes are symptoms around eyes. Thus, those symptoms can be gathered into a category "eye symptoms" when the location of symptoms should be focused on. symptoms should be focused on. The relations among those attributes are hierarchical as shown in Figure 1. This process, grouping of attributes, is called attribute-oriented generalization[1].

Attribute-oriented generalization can be viewed as transformation of variables in the context of rule induction. For example, an attribute "iris defects" should be transformed into an attribute "eye symptoms=yes".It is notable that the transformation of attributes in rules correspond to that of a database because a set of rules is equivalent to a reduced decision table. In this case, the case when eyes are normal is defined as "eye symptoms=no". Thus, the tranformation rule for iris defects is defined as:

$$[\textit{iris-defects} = yes] \rightarrow [\textit{eye-symptoms} = yes] \qquad (1)$$

$$Location \begin{cases} \text{Head} \cdots \\ \text{Face} \cdots \\ \text{Eye} : \begin{cases} terolism \begin{cases} hyper \\ normal \\ hypo \end{cases} \\ cornea \begin{cases} megalo \\ large \\ normal \end{cases} \\ antimongoloid\ slanting\ of \\ \quad palpebral\ fissures \begin{cases} yes \\ no \end{cases} \\ iris\ defects \begin{cases} yes \\ no \end{cases} \\ eyelashes \begin{cases} long \\ normal \end{cases} \\ \cdots \end{cases} \\ Noses \cdots \\ \cdots\cdots \end{cases}$$

Figure 1: An Example of Attribute Hierarchy

In general, when $[A_k = V_l]$ is a upper-level concept of $[a_i = v_j]$, a transforming rule is defined as:

$$[a_i = v_j] \rightarrow [A_k = V_l],$$

and the supporting set of $[A_k = V_l]$ is:

$$[A_i = V_l]_A = \bigcup_{i,j} [a_i = v_j]_a,$$

where A and a is a set of attributes for upper-level and lower level concepts, respectively.

2.3 Examples

Let us illustrate how fuzzy contexts is observed when attribute-oriented generalization is applied by using a small table (Table 1). Then, it is easy to see that a rule of "Aarskog",

$$[\textit{iris-defects} = yes] \rightarrow Aarskog \quad \alpha = 1.0, \kappa = 1.0$$

is obtained from Table 1.

When we apply trasforming rules shown in Figure 1 to the dataset of Table 1, the table is tranformed into Table 2. Then, by using transformation rule 1, the above rule

---

[1] Pawlak recently reports a Bayesian relation between accuracy and coverage[5]:

$$\begin{aligned} \alpha_R(D)P(D) &= P(R|D)P(D) = P(R,D) \\ &= P(R)P(D|R) = \kappa_R(D)P(R) \end{aligned}$$

This relation also suggests that *a priori* and *a posteriori* probabilities should be easily and automatically calculated from database.

[2] This probabilistic rule is also a kind of *Rough Modus Ponens*[5].

Table 1: A Small Database on Congenital Disorders

U	round	telorism	cornea	slanting	iris-defects	eyelashes	class
1	no	normal	megalo	yes	yes	long	Aarskog
2	yes	hyper	megalo	yes	yes	long	Aarskog
3	yes	hypo	normal	no	no	normal	Down
4	yes	hyper	normal	no	no	normal	Down
5	yes	hyper	large	yes	yes	long	Aarskog
6	no	hyper	megalo	yes	no	long	Cat-cry

DEFINITIONS: round: round face, slanting: antimongoloid slanting of palpebral fissures, Aarskog: Aarskog Syndrome, Down: Down Syndrome, Cat-cry: Cat Cry Syndrome.

Table 2: A Small Database on Congenital Disorders (Transformed)

U	eye	eye	eye	eye	eye	eye	class
1	no	no	yes	yes	yes	yes	Aarskog
2	yes	yes	yes	yes	yes	yes	Aarskog
3	yes	yes	no	no	no	no	Down
4	yes	yes	no	no	no	no	Down
5	yes	yes	yes	yes	yes	yes	Aarskog
6	no	yes	yes	yes	no	yes	Cat-cry

DEFINITIONS: eye: eye-symptoms

is trasformed into:

$$[\textit{eye-symptoms} = \textit{yes}] \rightarrow \textit{Aarskog}.$$

It is notable that mutual exclusiveness of attributes has been lost by tranformation. Since five attributes (telorism, cornea, slanting, iris-defects and eyelashes) are generalized into *eye-symptoms*, the candiates for accuracy and coverage will be (5/6, 2/3), (3/4, 3/3), (3/4, 3/3), (3/3, 3/3), and (3/4, 3/3), respectively. Then, we have to select which value is suitable for the context of this analysis.

In [8], one of the authors selected the mimimum value in medical context: accuracy is equal to 3/4 and coverage is equal to 2/3.

Thus, the rewritten rule becomes the following probabilistic rule:

$$[\textit{eye-symptoms} = \textit{yes}] \rightarrow \textit{Aarskog}$$
$$\alpha = 3/4 = 0.75, \kappa = 2/3 = 0.67.$$

This examples show that the loss of mutual exclusiveness is directly connected to the emergence of fuziness in a dataset. It it notable that the rule used for transformation is a deterministic one. When this kind of transformation is applied, whether applied rule is deterministic or not, fuzziness will be observed. However, no researchers has pointed out this problem with combination of rule induction and tranformation.

It is also notable that the conflicts between attributes with respect to accuarcy and coverage correponds to the vector representation of membership functions shown in Lin's context-free fuzzy sets[3].

3 Functional Representation of Context-Free Fuzzy Sets

Lin has pointed out problems with multiple membership functions and introduced relations between context-free fuzzy sets and information tables[3]. The main contribution of context-free fuzzy sets to data mining is that information tables can be used to represent multiple fuzzy membership functions. Usually when we meet multiple membership functions, we have to resolve the conflicts between functions. Lin discusses that this resolution is bounded by the context: min, maximum and other fuzzy operators can be viewed as a *context*. The discussion in Section 2 illustrates Lin's assertion. Especially, when we analyze relational-database, tranformation will be indispensable to data mining of multi-tables. However, tranformation may violate mutual exclusiveness of the target information table. Then, multiple fuzzy membership functions will be observed.

Lin's context-free fuzzy sets shows such analyzing procedures as a simple function as shown in Figure 3. The important parts in this algorithm are the way to construct a list of membership functions and the way to determine whether this algorithm outputs a metalist of a list of membership functions or a list of numerical values obtained by application of fuzzy operators to a list of membership functions.

4 Conclusions

This paper shows that mutual exclusiveness of conditional and decision attributes does not always hold in real-world databases, where conventional probabilstic approaches cannot be applied.

procedure *Resolution of Multiple Memberships*;
var
i : *integer*; L_a, L_i : *List*;
A: a list of Attribute-value pairs (multisets:bag);
F: a list of fuzzy operators;
begin
$L_i := A$;
while $(A \neq \{\})$ **do**
begin
$[a_i = v_j](k) = first(A)$;
Applend $\mu([a_i = v_j](k))$ to $L_{[a_i=v_j]}$
/* $L_{[a_i=v_j]}$: a list of membership function for attribute-value pairs */
$A := A - [a_i = v_j](k)$;
end.
if $(F = \{\})$ **then**
/* Context- Free */
return all of the lists $L_{[a_i=v_j]}$;
else
/* Resolution with Contexts*/
while $(F \neq \{\})$ **do**
begin
$f = first(F)$;
Apply f to each $L_{[a_i=v_j]}$;
$\mu_f([a_i = v_j]) = f(L_{[a_i=v_j]})$
Output all of the membership functions
$\mu_f([a_i = v_j])$
$F := F - f$;
end.
end {*Resolution of Multiple Memberships*};

Figure 2: Resolution of Multiple Fuzzy Memberships

It is surprising that tranformation will easily generate this situation in data mining from relation databases: when we apply attribute-oriented generalization to attributes in databases, generalized attributes will have fuzziness for classification. In this case, we have to take care about the conflicts between each attributes, which can be viewed as a problem with multiple membership functions. Then, the authors pointed out that these contexts should be analyzed by using fuzzy techniques, where context-free fuzzy sets will be a key idea to solve this problem. It will be our future work to induce fuzzy $if-then$ rules from this database and to compare these fuzzy rules with other conventional approaches.

References

[1] Y.D. Cai, N. Cercone and J. Han, Attribute-oriented induction in relational databases. in: Shapiro, G. P. and Frawley, W. J. (eds), *Knowledge Discovery in Databases*, AAAI press, Palo Alto, CA, pp.213-228 (1991).

[2] Fayyad, U.M., et al.(eds.): *Advances in Knoweledge Discovery and Data Mining*, AAAI Press (1996).

[3] Lin, T.Y. Context Free Fuzzy Sets and Information Tables *Proceedings of EUFIT'98*, Aachen, pp.76-80, 1998.

[4] Pawlak, Z., *Rough Sets.* Kluwer Academic Publishers, Dordrecht, 1991.

[5] Pawlak, Z. Rough Sets and Decision Analysis, *Fifth IIASA workshop on Decision Analysis and Support*, Laxenburg, 1998.

[6] Skowron, A. and Grzymala-Busse, J. From rough set theory to evidence theory. In: Yager, R., Fedrizzi, M. and Kacprzyk, J.(eds.) *Advances in the Dempster-Shafer Theory of Evidence*, pp.193-236, John Wiley & Sons, New York, 1994.

[7] *Readings in Machine Learning*, (Shavlik, J. W. and Dietterich, T.G., eds.) Morgan Kaufmann, Palo Alto, 1990.

[8] Tsumoto, S. Knowledge Discovery in Medical Databases based on Rough Sets and Attribute-Oriented Generalization. *Proceedings of IEEE-FUZZ99*, IEEE Press, Anchorage, 1998.

[9] Tsumoto, S. Automated Induction of Medical Expert System Rules from Clinical Databases based on Rough Set Theory *Information Sciences* **112**, 67-84, 1998.

[10] Tsumoto, S., Automated Discovery of Plausible Rules based on Rough Sets and Rough Inclusion, *Proceedings of PAKDD'99*, (in press), LNAI, Springer-Verlag.

[11] Zadeh, L.A., Toward a theory of fuzzy information granulation and its certainty in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* **90**, 111-127, 1997.

[12] Ziarko, W., Variable Precision Rough Set Model. *Journal of Computer and System Sciences.* 46, 39-59, 1993.