A Logical Approach to Fuzzy Data Analysis

Churn-Jung Liau[1] and Duen-Ren Liu[2]

[1] Institute of Information Science
Academia Sinica, Taipei, Taiwan
liaucj@iis.sinica.edu.tw
[2] Institute of Information Management
National Chiao-Tung University, Hsinchu, Taiwan
dliu@iim.nctu.edu.tw

Abstract. In this paper, we investigate the extraction of fuzzy rules from data tables based on possibility theory. A possibilistic decision language is used to represent the extracted fuzzy rules. The algorithm for rule extraction is presented and the complexity analysis is carried out. Since the results of the rule induction process strongly depend on the representation language, we also discuss some approach for dynamic adjustment of the language based on the data.

1 Introduction

The results reported here are originally motivated by the quantization problem in rough set-based data mining methods[5,4]. In those methods, the induced rules are represented by the so-called decision language(DL). The basic building blocks of DL are descriptors of the form (a, v), where a is an attribute and v is a value. If the domain of attribute values is continuous, a quantization process is usually necessary to replace the value v by an interval containing it. However, to make the induced rules more robust, replacing a crisp interval by fuzzy linguistic terms may be an interesting alternative. In this paper, we propose a possibilistic decision logic which facilitates the representation of fuzzy rules and so solve the quantization problem to some extent.

The possibilistic decision logic is based on possibility theory, which is developed by Zadeh from fuzzy set theory[8]. Given a universe W, a *possibility distribution* on W is a function $\pi : W \to [0, 1]$. Obviously, π is a characteristic function of a fuzzy subset of W. Let $\mathcal{F}(W)$ denote the class of all fuzzy subsets of W, then for $X, Y \in \mathcal{F}(W)$, two measures may be defined

$$CON(X, Y) = \sup_{w \in W} \mu_X(w) \otimes \mu_Y(w),$$

$$INC(Y, X) = \inf_{w \in W} \mu_Y(w) \to_{\otimes} \mu_X(w),$$

where $\otimes : [0, 1] \times [0, 1] \to [0, 1]$ is a t-norm[1] and $\to_{\otimes}$ is the implication function defined as $a \to_{\otimes} b = 1 - (a \otimes (1 - b))$ for all $a, b \in [0, 1]$. Hence, $CON(X, Y)$

[1] A binary operation $\otimes$ is a t-norm iff it is associative, commutative, and increasing in both places, and $1 \otimes a = a$ and $0 \otimes a = 0$ for all $a \in [0, 1]$.

J.M. Żytkow and J. Rauch (Eds.): PKDD'99, LNAI 1704, pp. 412–417, 1999.

denotes the degree of intersection between X and Y, whereas $INC(Y, X)$ is the degree of inclusion of Y in X.

2 Possibilistic Decision Logic

To represent the rules extracted from fuzzy data tables, we propose a possibilistic decision logic(PDCL) here. The linguistic terms used in the logic are fixed in advance and their meaning is given by a context. Once the context is determined, the semantics of wffs of the logic can be defined via possibility theory.

2.1 Syntax and Context

Let A be a set of attributes, L be a set of linguistic terms such that a function $type : L \to A$ assigning each linguistic term with its type, and H be a set of linguistic hedges, then the atomic formulas of PDCL are in one of the forms, $(a, \pi l), (a, \nu l), (a, \tau\pi l)$, and $(a, \tau\nu l)$, where $a \in A$, $l \in L$, $\tau \in H$ and $type(l) = a$. The set of well-formed formulas of PDCL is the smallest set containing atomic ones and closed under Boolean connectives.

For example, $(t, \nu\mathbf{high})$, $(t, \pi\mathbf{high})$, $(t, \mathbf{very}\nu\mathbf{high})$, and $(t, \mathbf{very}\pi\mathbf{high})$ denote respectively "the temperature is certainly high", "the temperature is possibly high", "it is very certain the temperature is high", and "it is very possible the temperature is high". Here, the term "very" is the so-called linguistic hedge.

It is well-known that many natural language terms are highly context-dependent. For example, the term "tall" may have quite different meanings between "a tall basketball player" and "a tall child". To model the context-dependency, we associate a context with each PDCL. The context determines the domain of values of each attributes and assigns appropriate meaning to each linguistic term and hedge. Formally, a context associated with a PDCL is a triple $(\{V_a\}_{a\in A}, m_1, m_2)$, where V_a is a domain of values for each $a \in A$, m_1 is a function on L such that $m_1(l) \in \mathcal{F}(V_a)$ if $type(l) = a$, and $m_2 : H \to ([0, 1] \to [0, 1])$ is a function mapping each hedge to a function from $[0, 1]$ to $[0, 1]$. While the domains V_a and m_1 are totally determined by the users or linguistic experts to reflect the intended meaning of these attributes and linguistic terms, there exist some common definitions for the linguistic hedges in the literature[1].

2.2 Semantics

Given a PDCL with set of attributes A, set of linguistic terms L, set of linguistic hedges H, and a context $(\{V_a\}_{a\in A}, m_1, m_2)$, a fuzzy data table (FDT) is a pair $S = (U, F(A))$, where U is a finite set of objects and $F(A) = \{f_a : U \to \mathcal{F}(V_a) \mid a \in A\}$. Intuitively, $f_a(x)$ denotes the uncertain value of attribute a for object x. Thus $f_a(x) = V_a$ when the value is missing and $f_a(x)$ is a singleton when the value is precise. This means FDT's can represent both precise and imprecise data in a uniform framework. Let $\mathcal{L}$ denote the set of wffs of the PDCL, then for an FDT $S = (U, F(A))$, we can define the truth valuation function $E_S : U \times \mathcal{L} \to [0, 1]$ as follows:

1. $E_S(x,(a,\pi l)) = \sup_{v\in V_a} \mu_{m_1(l)}(v) \otimes \mu_{f_a(x)}(v)$
2. $E_S(x,(a,\nu l)) = \inf_{v\in V_a} \mu_{f_a(x)}(v) \rightarrow_{\otimes} \mu_{m_1(l)}(v)$
3. $E_S(x,(a,\tau\pi l)) = m_2(\tau)(E_S(x,(a,\pi l)))$
4. $E_S(x,(a,\tau\nu l)) = m_2(\tau)(E_S(x,(a,\nu l)))$
5. $E_S(x,\neg\varphi) = 1 - E_S(x,\varphi)$
6. $E_S(x,\varphi\wedge\psi) = E_S(x,\varphi)\otimes E_S(x,\psi)$
7. $E_S(x,\varphi\vee\psi) = E_S(x,\varphi)\oplus E_S(x,\psi)$, where $\oplus$ is a t-conorm defined by $a\oplus b = 1-(1-a)\otimes(1-b)$
8. $E_S(x,\varphi\longrightarrow\psi) = E_S(x,\varphi)\rightarrow_{\otimes} E_S(x,\psi)$
9. $E_S(x,\varphi\equiv\psi) = E_S(x,\varphi\longrightarrow\psi)\otimes E_S(x,\psi\longrightarrow\varphi)$

We will define $[\![\varphi]\!]_S = \bigotimes_{x\in U} E_S(x,\varphi)$ as the truth degree of φ with respect to an FDT S. Let $p_a = (a,\pi l),(a,\nu l),(a,\tau\pi l)$, or $(a,\tau\nu l)$ be an atomic formula, called a-basic formula, then a CD-decision rule for $C, D \subseteq A$ is a wff of the form

$$\bigwedge_{a\in C} p_a \longrightarrow \bigwedge_{a\in D} p_a.$$

When φ is a CD-decision rule, $[\![\varphi]\!]_S$ will be the strength of the rule according to our semantics. In the next section, we will present the approach to discover this kind of rules from an FDT.

3 Rule Induction Process

In traditional rough set based approach to data analysis, for a decsriptor (a,v), v appears somewhere in the table, so we do not have to fix the atomic formulas of the decision logic in advance. However, in an FDT, it is possible that for some numerical attribute a, $f_a(x)$ has precise value, and to avoid the quantization problem, we would still like to use some linguistic terms to represent the induced rules. For example, we may have a data table with precise temperature values and want to discover rules of the form "If temperature is *high*, then ...". Thus it is necessary to fix the set of linguistic terms L of our PDCL in advance. On the other hand, if the linguistic terms and the context is given independent of the FDT, it is possible that the data values are not completely covered by these terms. To resolve the dilemma, we will first describe the rule induction algorithm by assuming a fixed set of linguistic terms and its associated context is given by the domain experts, and then consider the process of setting up or adjusting the language and context. For simplicity, we temporarily assume $H = \emptyset$ and omit the m_2 component of a context. Without loss of generality, we can also assume the set of decision attributes is a singleton. The algorithm is described in Figure 1.

In the first step of the procedure, we test whether x is a support of the a-basic formulas $(a,\pi l)$ and $(a,\nu l)$ for each $x\in U$, $a\in C$, and $l\in L_a$. If the the degree of consistence between $f_a(x)$ and the linguistic term l is equal to 1, then x supports the statement $(a,\pi l)$, and if $f_a(x)$ implies l to the degree 1, then x supports the statement $(a,\nu l)$. Since P_x is the Cartesian product of P_x^a for all

Procedure Rule Induction
Input: A FDCL $\mathcal{L}(A, L)$, a context $(\{V_a\}_{a\in A}, m_1)$, and a FDT $S = (U, F(A))$. Assume $L = \bigcup_{a\in A} L_a$ and $A = C \cup \{d\}$, where $L_a = \{l \in L \mid type(l) = a\}$, C is the set of condition attributes, and d is the decision attribute.
Output: A set of $C\{d\}$-decision rules with strength.
Steps:

1 Let $P_x^a = \{\pi l \mid l \in L_a, CON(m_1(l), f_a(x)) = 1\} \cup \{\nu l \mid l \in L_a, INC(f_a(x), m_1(l)) = 1\}$.
2 Let $P_x = \times_{a\in C} P_x^a$ and $P = \bigcup_{x\in U} P_x$.
3 For each tuple $\mathbf{t} \in P$ and $l \in L_d$, let

$$\varphi_\pi(\mathbf{t}, l) = \bigwedge_{a\in C} (a, \mathbf{t}(a)) \longrightarrow (d, \pi l),$$

$$\varphi_\nu(\mathbf{t}, l) = \bigwedge_{a\in C} (a, \mathbf{t}(a)) \longrightarrow (d, \nu l).$$

4 Return the set $\{(\varphi_\pi(\mathbf{t}, l), [\![\varphi_\pi(\mathbf{t}, l)]\!]_S) \mid \mathbf{t} \in P, l \in L_d\} \cup \{(\varphi_\nu(\mathbf{t}, l), [\![\varphi_\nu(\mathbf{t}, l)]\!]_S) \mid \mathbf{t} \in P, l \in L_d\}$

End

Fig. 1. The procedure to discover fuzzy rules

$a \in C$, for any tuple $\mathbf{t} \in P_x$, x will be the support of the wff $\bigwedge_{a\in C}(a, \mathbf{t}(a))$. Thus each tuple in P corresponds to a conjunctive wff which has at least a support in the FDT S, and our extracted rules are those with at least a support. The number of supports for a tuple $\mathbf{t}$ in P (and its corresponding rules) is equal to the number of x's such that $\mathbf{t} \in P_x$. For further refinement, we can eliminate rules with number of supports less than some predefined threshold value. Analogously, we return rules with arbitrary strength at the last step, however, we can also set a threshold and drop out any rules with strength less than the threshold.

To carry out the complexity analysis of the rule induction process, let us define $|U| = m$, $|L_a| = n_a$ for all $a \in A$, and $|L| = n = \sum_{a\in A} n_a$. Then step 1 of the procedure will need $O(mn)$ time for all $x \in U$ and $a \in A$ since it takes at most $2m(n - n_d)$ times of computation for measures CON and INC. In step 2, the cardinality of P is at most $\prod_{a\in C} 2n_a$, so step 3 will need $O(\prod_{a\in A} 2n_a)$ time, and finally step 4 will take $O(m \cdot \prod_{a\in A} 2n_a)$ time since the computation of $[\![\varphi]\!]_S$ will go through each element of U for any φ. If there exists constant N such that $n_a \leq N$ for all $a \in A$, then the overall time complexity of the procedure is $O(m(n + (2N)^{|A|}))$. Thus the time complexity of the procedure is linear in the number of training cases, though it is exponential in the number of attributes. In other words, the procedure is efficient for problems with small numbers of attributes.

3.1 Context Construction and Adjustment

In the procedure given above, we assume the set of linguistic terms and their meaning is fixed in advance. For example, for the temperature attribute, we can assume the available linguistic terms are only "HIGH", "MEDIUM", and "LOW" and the associated fuzzy sets are given by domain experts. This assumption has the implication that the possible candidates of the induced rules are limited. However, it also has the defect that the linguistic terms may not adequately describe the data. In other words, the set P_x^a may be empty for all $x \in U$. To resolve the problem, we may adjust the context and the set of linguistic terms dynamically according to the FDT. According to the types of the attributes, two cases are considered.

First, if the attribute a is nominal, then we can simply let $L_a = V_a$ and for each $v \in L_a$, its meaning is the singleton set $\{v\}$. In this case, the semantics of atomic formulas $(a, \pi v)$ and $(a, \nu v)$ collapse into the same one, i.e., the one for (a, v) in original decision logic, when the data $f_a(x)$ is precise for all $x \in U$.

Second, when the attribute a is numerical, we assume a metric $\delta : V_a \times V_a \to [0, \infty)$ exists. Then we define $\mathcal{V}_1 = \{f_a(x) \mid x \in U, f_a(x) \text{ is a crisp set}\}$, $\mathcal{V}_2 = \{f_a(x) \mid x \in U\} - \mathcal{V}_1 - \{V_a\}$, and $V = \bigcup \mathcal{V}_1$. Note that all these three sets are finite and $V \subset V_a$. Now, some classical clustering techniques can be applied to partition V into crisp clusters $V_1, V_2, \ldots, V_k$[2,3]. For example, by a linkage method for hierarchical clustering[3], starting from $\mathcal{V}_1$, we merge two singleton sets with shortest distance into one. The new created set is used to replace its two original components in $\mathcal{V}_1$ and so we have a coarser partition. The process continues by merging two closest subsets each time until a predefined limit k is achieved. Here, the distance between two crisp subsets of V, X and Y, is defined as $\delta(X, Y) = \max_{x \in X, y \in Y} \delta(x, y)$. After the clustering process, we can find the center of each V_i, v_i^* as

$$v_i^* = \arg\min_{v \in V_i} \delta(\{v\}, V_i - \{v\}).$$

Let $d_i^* = \delta(\{v_i^*\}, V)$, then we can define a fuzzy set $X_i \in \mathcal{F}(V_a)$ with the membership function $\mu_{X_i}(x) = \max(0, 1 - \frac{\delta(v_i^*, x)}{d_i^*})$. Let $\mathcal{V}_3 = \mathcal{V}_2 \cup \{X_1, \ldots, X_k\}$, then $\mathcal{V}_3$ is the set of candidates for meanings of our linguistic terms. If the number of fuzzy sets in $\mathcal{V}_3$ is still too many, then we can apply clustering techniques to $\mathcal{V}_3$ again, but using the similarity measure between fuzzy sets to determine distance between them. Then the center of each cluster is collected into a set $\mathcal{V}$. Finally, we can associate each element in $\mathcal{V}$ with a appropriate linguistic label and the set of linguistic labels are our L_a and their meaning are naturally defined as the corresponding elements in $\mathcal{V}$.

4 Conclusion

In a recent article, Pawlak, the founder of rough set theory, point out that discretization of quantitative attribute values are badly needed for rough set-based data analysis[6]. In this regard, many discretization methods have been

explored[4]. On the other hand, the management of uncertainty has been a long-standing requirement in intelligent data analysis. In this paper, we present a uniform logical framework for handling both uncertain and quantitative data. In the framework, uncertain attribute values are represented as fuzzy subsets of the domain and quantitative values may belong to some fuzzy sets to some different degrees. Linguistic terms correspond to the fuzzy subsets are taken as the basic building blocks of a PDCL. Then, for each item of data, the information contained in it decides the degree of truth of wffs of the language. A rule in our framework is an implication formula of the language and the aggregated degree of truth of the formula on all data items is taken as the strength of the rule. The formulas of decision logic are called information pre-granules in [7], so our wffs of PDCL can be analogously called fuzzy information pre-granules. Therefore, our logical approach to fuzzy data analysis can be seen as a formal instance of fuzzy granular information processing.

References

1. R. López de Mántaras and L. Godo. "From fuzzy logic to fuzzy truth-valued logic for expert systems: A survey". In *Proceedings of the 2nd IEEE International Conference on Fuzzy Systems*, pages 75–755, San Francisco, CA, 1993. IEEE.
2. A. Kandel. *Fuzzy Mathematical Techniques with Applications.* Addison-Wesley Publishing Co., 1986.
3. S. Miyamoto. *Fuzzy Sets in Information Retrieval and Cluster Analysis.* Kluwer Academic Publishers, 1990.
4. H.S. Nguyen and S.H. Nguyen. "Discretization methods in data mining". In L. Polkowski and A. Skowron, editors, *Rough Sets in Knowledge Discovery*, pages 451–482. Physica-Verlag, 1998.
5. Z. Pawlak. *Rough Sets–Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, 1991.
6. Z. Pawlak. "Rough sets: Present state and perspectives". In *Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems(IPMU)*, pages 1137–1145, 1996.
7. L. Polkowski and A. Skowron. "Towards Adaptive Calculus of Granules". In *Proceedings of the 7th IEEE International Conference on IFuzzy Systems*, pages 111–116, 1998.
8. L.A. Zadeh. "Fuzzy sets as a basis for a theory of possibility". *Fuzzy Sets and Systems*, 1(1):3–28, 1978.