ELSEVIER

Fuzzy Sets and Systems 106 (1999) 349–356

FUZZY sets and systems

Learning optimization in simplifying fuzzy rules

Xizhao Wang[a,*], Jiarong Hong[b]

[a] Department of Mathematics, Hebei University, Baoding 071002, Hebei, People's Republic of China
[b] Department of Computer science, Harbin Institute of Technology, Harbin 150001, People's Republic of China

Received January 1997; received in revised form June 1997

Abstract

It is important that an optimal learning problem is proved to be NP-hard and the heuristic algorithm for solving the problem has to be given. This paper deals with a learning problem appearing in the process of simplifying fuzzy rules, proves that the solution optimization is NP-hard and gives its heuristic algorithm. This heuristic, regarded as a new, fuzzy learning algorithm, has many significant advantages. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Knowledge acquisition and learning; Learning from examples; Artificial intelligence; Fuzzy algorithms; NP-hardness

1. Introduction

Learning is an essential way with which human beings acquire wisdom. Machine learning is a basic way with which a computer system can possess intelligence. Learning from examples is one of the ripe branches of machine learning. The main objective of learning from examples is to extract a family of rules from examples which are divided into several classes. This family of rules must cover these examples explicitly. Therefore, learning from examples is also an important way to acquire knowledge. Because knowledge acquisition has been universally regarded as the bottleneck of the development of expert systems, increasingly great importance is attached to learning from examples. Several systems of learning from examples appear in succession such as ID3 [9] and AQ15 [7].

*Corresponding author. E-mail: wangxz@hbu.edu.cn.

As there exist cognitive uncertainties such as vagueness and ambiguity, examples used in learning are generally considered to be fuzzy data. There fuzzy data preserved in a database are regarded as a type of knowledge. The learning from examples with handling uncertainties is fuzzy learning. The objective of fuzzy learning is also to generate a family of rules, especially fuzzy rules. Several algorithms of fuzzy learning from examples such as induction of fuzzy decision trees and fuzzy ID3 have been developed in [1, 2, 11, 13–15].

For a given problem of fuzzy learning from examples there exist many algorithms, each one of which can generate a family of fuzzy rules covering these examples. As the fact that the cover capability of a simple rule is stronger than that of a complicated rule has generally been acknowledged, a natural problem is how to search for an algorithm which can generate the simplest family of fuzzy rules. This is an essential optimization problem in learning

PII: S0165-0114(97)00300-X

theory. We deal with this optimal learning problem in this paper. Starting from a family of initial fuzzy rules, this paper proves that the optimization of simplifying fuzzy rules is NP-hard, and it also gives a rather effective and intuitive heuristic algorithm. This heuristic, which is regarded as a new fuzzy learning algorithm, possesses many advantages.

The fuzzy learning algorithm presented in this paper is also suitable for the crisp case where the true degree of initial fuzzy rules is regarded as 1.

2. An optimal learning problem in simplifying fuzzy rules

In this section, fuzzy rules and their true degree are defined. A fuzzy knowledge base is considered and the core and the reduction of a initial fuzzy rule are discussed. An optimal learning problem of simplifying fuzzy rules is introduced.

2.1. Fuzzy rules and the true degree of fuzzy rules

Definition 1 (Yuan and Shaw [15]). A fuzzy rule takes a form: IF A THEN B which defines a fuzzy relation from condition fuzzy set A to conclusion fuzzy set B.

A rule IF A THEN B is true means that A implies B, denoted by $A \Rightarrow B$. The implication operator can be interpreted in many ways [10, 15]. As the interpretation of [15], the implication $A \Rightarrow B$ in this paper is understood to be $A \subset B$.

Definition 2 (Yuan and Shaw [15]). The true degree of a fuzzy rule $A \Rightarrow B$ is defined to be $\sum_{u\in U} \min(\mu_A(u), \mu_B(u))/\sum_{u\in U} \mu_A(u)$ where A and B are two fuzzy sets defined on the same universe U.

As an instance, we consider the following family of training examples (Table 1, adopted from [9, 15] with some modification).

Table 1 shows a small set of training examples that uses the 'Saturday morning' attributes, Positive class and Negative class are two unspecified activities. There are 12 fuzzy sets defined on the same universe $U = \{e_1, e_2, \dots, e_{14}\}$. They are sunny, overcast, rain, hot, mild, cool, high, normal, false, true, positive and negative. It is easy to verify the following fuzzy rules.

sunny $\Rightarrow$ negative with true degree 0.66;
hot$\cap$*sunny* $\Rightarrow$ negative with true degree 0.86;
high$\cap$*hot*$\cap$*sunny* $\Rightarrow$ negative with true degree 0.90;
false$\cap$*high*$\cap$*hot*$\cap$*sunny* $\Rightarrow$ negative with true degree 1.00.

If fuzzy sets sunny, hot, high, and false are considered to be fuzzy evidence, the above four fuzzy rules can give an explanation, i.e., the true degree of fuzzy rules will be becoming big as the evidence accumulates.

2.2. Fuzzy knowledge base

Table 2 is said to be a fuzzy knowledge base. Where there are n rows and m attributes $Attrj$ ($j = 1, 2, \dots, m$). A_{ij} ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$) are all fuzzy sets defined on the same universe $U = \{1, 2, \dots, n\}$. For each i and each j ($1 \leqslant i \leqslant n$; $1 \leqslant j \leqslant m$) A_{ij} is regarded as the value of the ith example for the jth attribute, C_i is the classification result of the ith example, the ith row is explained to be an initial fuzzy rule taking a form $\bigcap_{p=1}^{m} A_{ip} \Rightarrow C_i$ with true degree α_i and inconsistent degree β_i (See Definition 4.)

From the ith initial fuzzy rule, many fuzzy rules can be generated such as $\bigcap_{t=1}^{k} A_{ij_t} \Rightarrow C_i$ with a true degree and an inconsistent degree, where $\{j_1, j_2, \dots, j_k\} \subset \{1, 2, \dots, m\}$. Let $S = \{Attrj_1, Attrj_2, \dots, Attrj_k\}$ be a subset of attributes ($k \leqslant m$), we denote the fuzzy rule $\bigcap_{t=1}^{k} A_{ij_t} \Rightarrow C_i$ with a true degree α_i and an inconsistent degree β_i, in short, by $Attr|_S^i \Rightarrow C_i\ [\alpha_i, \beta_i]$.

Definition 3. A fuzzy knowledge base is said to have true evidence if $S_1 \subset S_2$ implies $\alpha_1 \leqslant \alpha_2$ for an arbitrary, given ith row where α_1 and α_2 are true degrees of fuzzy rules $Attr|_{S_1}^i \Rightarrow C_i$ and and $Attr|_{S_2}^i \Rightarrow C_i$, respectively.

Definition 4. For a given fuzzy rule $Attr|_S^i \Rightarrow C_i$ with true degree α_i, the inconsistent degree β_i is defined by $|E|$ where $|E| = \{j \mid Attr|_S^i = Attr|_S^j,\ C_i \neq C_j\}$, $|E|$ denotes the number of elements of the set E.

As an example, the fuzzy knowledge base, Table 3, can be generated by selecting the maximal membership of each attribute over its range of nonfuzzy label values in Table 1.

Table 1
A family of training examples

Case	Outlook			Temperature			Humidity		Windy		Class	
e_j	Sunny	Overcast	Rain	Hot	Mild	Cool	High	Normal	False	True	Positive	Negative
$j=1$	0.9	0.1	0.0	0.9	0.1	0.0	0.8	0.2	0.7	0.4	0.4	0.7
$j=2$	0.9	0.1	0.1	0.8	0.2	0.1	0.9	0.2	0.1	0.8	0.3	0.7
$j=3$	0.1	0.9	0.2	0.9	0.1	0.1	0.9	0.1	0.9	0.1	0.8	0.3
$j=4$	0.0	0.1	0.9	0.1	0.9	0.0	0.6	0.5	0.8	0.3	0.6	0.5
$j=5$	0.1	0.0	0.9	0.0	0.1	0.9	0.0	1.0	0.8	0.2	0.9	0.2
$j=6$	0.1	0.1	0.9	0.0	0.2	0.9	0.1	0.9	0.1	0.9	0.3	0.8
$j=7$	0.0	1.0	0.0	0.0	0.1	0.9	0.1	0.9	0.2	0.9	0.9	0.3
$j=8$	0.9	0.1	0.0	0.3	0.9	0.1	0.9	0.1	1.0	0.0	0.2	0.9
$j=9$	0.8	0.2	0.0	0.0	0.4	0.6	0.0	1.0	1.0	0.0	0.9	0.2
$j=10$	0.0	0.1	0.9	0.0	1.0	0.0	0.0	1.0	0.9	0.1	0.6	0.5
$j=11$	0.9	0.1	0.0	0.0	0.9	0.1	0.1	0.9	0.0	1.0	0.8	0.3
$j=12$	0.0	1.0	0.0	0.1	0.9	0.0	1.0	0.0	0.0	1.0	0.7	0.4
$j=13$	0.0	0.9	0.1	1.0	0.0	0.0	0.0	1.0	0.9	0.1	0.7	0.2
$j=14$	0.0	0.1	0.9	0.0	0.9	0.1	0.9	0.1	0.0	1.0	0.1	0.9

Table 2
Fuzzy knowledge base

No.	Attr1	Attr2	...	Attrm	Class	True degree	Inconsistency
r_1	A_{11}	A_{12}	...	A_{1m}	C_1	α_1	β_1
r_2	A_{21}	A_{22}	...	A_{2m}	C_2	α_2	β_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r_n	A_{n1}	A_{n2}	...	A_{nm}	C_n	α_n	β_n

It is easy to prove that the inconsistent degree of the fuzzy rule $Attr|_S^i \Rightarrow C_i$ will become small as the set S increases monotonically. For instance, the inconsistent degrees of fuzzy rules

false $\Rightarrow N$ ($S = \{Windy\}$),
high$\cap$*false* $\Rightarrow N$ ($S = \{Humidity, Windy\}$),
hot$\cap$*high*$\cap$*false* $\Rightarrow N$
($S = \{Temperature, Humidity, Windy\}$),
and
sunny$\cap$*hot*$\cap$*high*$\cap$*false* $\Rightarrow N$ ($S = \{Outlook, Temperature, Humidity, Windy\}$)

are 6, 2, 1 and 0, respectively. Obviously, the inconsistent degree of every initial fuzzy rule in Table 3 is 0 (the last column of Table 3).

2.3. Core and reduction of initial fuzzy rule

Consider a fuzzy knowledge base where the set of attributes is supposed to be C.

Definition 5. For a given fuzzy rule, $Attr|_S^i \Rightarrow C_i$ with a true degree α and an inconsistent degree β, an attribute A ($A \in C$) is said to be dispensable in the fuzzy rule if $Attr|_{S-\{A\}}^i \Rightarrow C_i$ has a true degree greater than or equal to δ (a given threshold) and an inconsistent degree less than or equal to β, otherwise, attribute A is indispensable in the rule.

In the following, we always regard δ as a given threshold and do not repeat the meaning of the Greek letter δ.

As an example, we consider the first initial fuzzy rule listed in Table 3. Because the inconsistent degree of the initial fuzzy rule *sunny*$\cap$*hot*$\cap$*high*$\cap$*false* $\Rightarrow N$ is 0 but that of the fuzzy rule *hot*$\cap$*high*$\cap$*false* $\Rightarrow N$ is 1, the attribute Outlook is indispensable in the first initial fuzzy rule. Similarly, we have attributes *Temperature*, *Humidity* and *Windy* that are dispensable in the first initial fuzzy rule.

Table 3
A fuzzy knowledge base generated by Table 1

No. r_j	*Outlook*	*Temperature*	*Humidity*	*Windy*	*Class*	True degree	Inconsistency degree
$j = 1$	Sunny	Hot	High	False	Negative	1.00	0
$j = 2$	Sunny	Hot	High	True	Negative	0.92	0
$j = 3$	Overcast	Hot	High	False	Positive	0.92	0
$j = 4$	Rain	Mild	High	False	Positive	1.00	0
$j = 5$	Rain	Cool	Normal	False	Positive	1.00	0
$j = 6$	Rain	Cool	Normal	True	Negative	0.93	0
$j = 7$	Overcast	Cool	Normal	True	Positive	1.00	0
$j = 8$	Sunny	Mild	High	False	Negative	1.00	0
$j = 9$	Sunny	Cool	Normal	False	Positive	1.00	0
$j = 10$	Rain	Mild	Normal	False	Positive	0.83	0
$j = 11$	Sunny	Mild	Normal	True	Positive	0.93	0
$j = 12$	Overcast	Mild	High	True	Positive	0.89	0
$j = 13$	Overcast	Hot	Normal	False	Positive	0.86	0
$j = 14$	Rain	Mild	High	True	Negative	1.00	0

Definition 6. For a given fuzzy rule, $Attr|_S^i \Rightarrow C_i$ with a true degree α and an inconsistent degree β, if all attributes in S are indispensable, this rule is called independent. A subset of attributes R ($R \subset S$) is called a reduct of the rule $Attr|_S^i \Rightarrow C_i$ if $Attr|_R^i \Rightarrow C_i$ is independent and has an inconsistent degree less than or equal to β and a true degree greater than or equal to δ. (For convenience, we also call $Attr|_R^i \Rightarrow C_i$ a reduct of $Attr|_S^i \Rightarrow C_i$). The set of attributes which are indispensable in the initial rule $Attr|_C^i \Rightarrow C_i$ is called the core of the initial fuzzy rule.

For a given initial fuzzy rule, its core is unique but its reducts are not. We continue considering, for example, the first initial rule listed in Table 3. The core is $\{Outlook\}$ while $\{Outlook, Temperature\}$ and $\{Outlook, Humidity\}$ are two reducts of the rule.

Theorem 1. *For a given fuzzy knowledge base with true evidence, the core of an initial fuzzy rule is equal to the intersection of all reducts of the rule. That is*

$$Core(Attr|_C^i \Rightarrow C_i) = \cap\{\mathrm{Re}duct(Attr|_C^i \Rightarrow C_i)\}$$

Proof. If A is a dispensable attribute in the rule $Attr|_C^i \Rightarrow C_i$ with an inconsistent degree β (i.e. A does not belong to the core), the rule $Attr|_{C-\{A\}}^i \Rightarrow C_i$ will have a true degree greater than or equal to δ and an inconsistent degree less than or equal to β. A reduct R, $R \subset C - \{A\}$, can be obtained by removing superfluous attributes in the set $C - \{A\}$. Obviously, A does not belong to R, i.e., attribute A does not belong to some reduct of the rule $Attr|_C^i \Rightarrow C_i$.

On the other hand, if attribute A does not belong to some reduct of the rule $Attr|_C^i \Rightarrow C_i$, there exists a subset of C, R, such that A does not belong to R and the rule $Attr|_R^i \Rightarrow C_i$ has a true degree greater than or equal to δ and an inconsistent degree less than or equal to β. As the fuzzy knowledge base has true evidence, $R \subset C - \{A\}$implies that the rule $Attr|_{C-\{A\}}^i \Rightarrow C_i$ has a true degree greater than or equal to δ and an inconsistent degree less than or equal to β. Hence, attribute A is dispensable in the rule $Attr|_C^i \Rightarrow C_i$. This completes the proof. □

Definition 7. A reduct of an initial fuzzy rule $Attr|_C^i \Rightarrow C_i$, R, is said to be minimal, if S is not a reduct of the initial fuzzy rule for each set S with $S \subset R$ and $S \neq R$.

For convenience, the rule $Attr|_R^i \Rightarrow C_i$ is also said to be a minimal reduct of the initial fuzzy rule.

Theorem 1 gives us a method of obtaining minimal reducts for fuzzy knowledge base with true evidence. Starting from the core, a minimal reduct of an initial fuzzy rule can be obtained by adding progressively attributes to the core and verifying the true degree and the inconsistent degree. For a fuzzy knowledge base without true evidence, the process of obtaining a minimal reduct must start from empty set.

2.4. An optimal learning problem

A fuzzy example is considered to be a fuzzy set defined on the nonfuzzy label space consisting of all values of attributes. For instance, the first row of Table 1 $e_1 = (0.9, 0.1, 0.0, 0.9, 0.1, 0.0, 0.8, 0.2, 0.7, 0.4, 0.4, 0.7)$ is a fuzzy example of the fuzzy knowledge base Table 3, where the nonfuzzy label space is (*sunny, overcast, rain, hot, mild, cool, high, normal, false, true, positive, negative*).

Definition 8. A fuzzy rule $Attr|_S^i \Rightarrow C_i$ is said to cover a fuzzy example if the membership of attributes and the membership of classification for the example are all greater than or equal to η (a threshold).

Let $\eta = 0.6$, for instance, the fuzzy rule "*Temperature is hot*"$\cap$"*Outlook is Sunny*" $\Rightarrow N$ covers the example e_1 because the membership of attributes is $0.9 > \eta$ and the membership of classification is $0.7 > \eta$.

The ith initial fuzzy rule covers the ith fuzzy example by means of selecting a feasible η. Obviously, a reduct of the ith initial fuzzy rule covers the i-th fuzzy example too ($1 \leqslant i \leqslant n$).

The main task of inductive learning is to extract rules from given examples. When cognitive uncertainties of a given example are handled, rules extracted are generally fuzzy [15]. One optimal learning from examples is how to extract rules such that these extracted rules cover the given examples and these rules are "simplest". Now, we introduce a problem of optimal learning from examples appearing in the process of simplifying fuzzy rules.

Let E be a family of fuzzy examples (e.g. Table 1), T be a fuzzy knowledge base generated by selecting the maximal membership of each attribute over its range of nonfuzzy label values (e.g. Table 3). Every row of the fuzzy knowledge base is considered to be an initial fuzzy rule with some true degree and with some inconsistent degree (e.g. the last two columns of Table 3). An optimal learning problem is in search of a family of fuzzy rules covering all given fuzzy examples such that (1) each fuzzy rule of this family is a minimal reduct of some initial fuzzy rule, (2) the number of fuzzy rules of this family is least.

3. NP-hard problem and heuristic algorithm

A computational problem is said to be a P-problem if there exists an algorithm such that the exact solution to the computational problem can be obtained within time of polynomial, if not, it is called NP-hard. For a NP-hard problem, only the approximate heuristic algorithm can be given. It is the kernel of a computational problem that the exact algorithm for the solution to the problem is given or the problem is proved to be NP-hard and the approximate, heuristic algorithm is given. Details of NP-hard problems can be found in [3, 12]

In this section, the optimal learning problem of simplifying fuzzy rules mentioned in Section 2.4 is proved to be NP-hard and a rather intuitive, effective, heuristic algorithm is given. Suppose the true degree of each initial fuzzy rule is greater than or equal to a given threshold, we consider the optimal learning problem mentioned in Section 2.4. Obviously, all initial fuzzy rules constitute a family which covers all given fuzzy examples (e.g. Table 3 covers Table 1, the threshold is taken to be 0.6). The optimal learning problem is divided into three tasks to be fulfilled.

The first task is in search of a minimal reduct for each initial fuzzy rule. Theorem 1 provides a method of obtaining a minimal reduct, which can be divided into six steps.

Step 1: For the ith initial fuzzy rule ($1 \leqslant i \leqslant n$), the core, K, can be given by verifying whether an attribute is dispensable in the condition set (if the fuzzy knowledge base has not true evidence, K is supposed to be empty set), $\Gamma := 1$.

Step 2: Take Γ attributes $A_1, A_2, \ldots, A_\Gamma$ from $C - K$.

Step 3: Add $A_1, A_2, \ldots, A_\Gamma$ to K ($K := K \cup \{A_1, A_2, \ldots, A_\Gamma\}$).

Step 4: Compute the true degree and the inconsistent degree of the fuzzy rule $Attr|_K^i \Rightarrow C_i$.

Step 5: *If* K is a reduct then exit successfully, else new Γ attributes $A_1, A_2, \ldots, A_\Gamma$ are taken from $C - K$, goto Step 3.

Step 6: If all combinations of elements of $C - K$, have been used and a reduct does not appear, $\Gamma := \Gamma + 1$, goto Step 2.

This process of determining a minimal reduct of an initial fuzzy rule is illustrated by considering the first row of Table 3. Let δ be 0.8, it is easy to verify that the attribute Outlook is uniquely indispensable in the first initial fuzzy rule of Table 3. Therefore, *Outlook* is the core of the first initial fuzzy rule. Take the second attribute Temperature and add it to the core. Notice

$$\begin{aligned} sunny \cap hot &\Rightarrow N \text{ with true degree } 0.86 \\ &\quad \text{and inconsistent degree } 0 \\ sunny &\Rightarrow N \text{ with true degree } 0.66 (< \delta) \\ &\quad \text{and inconsistent degree } 2 (> 0) \\ hot &\Rightarrow N \text{ with true degree } 0.59 (< \delta) \\ &\quad \text{and inconsistent degree } 2 (> 0) \end{aligned}$$

we have the conclusion that sunny and hot are indispensable in the fuzzy rule $sunny \cap hot \Rightarrow N$, and $\{Outlook, Temperature\}$ is a minimal reduct of the initial fuzzy rule.

This method of determining a minimal reduct is only suitable for the case that the number of attributes is not much. Generally, the number of attributes of a learning problem is far less than that of examples.

The second task is in search of a family of minimal reducts for the ith fuzzy example ($1 \leqslant i \leqslant n$) such that each reduct inside of this family covers the ith fuzzy example. Let the set of fuzzy examples be $E = \{e_1, e_2, \ldots, e_n\}$ and the set of minimal reducts be $R = \{r_1, r_2, \ldots, r_n\}$ where r_i is the minimal reduct of the ith initial rule ($1 \leqslant i \leqslant n$). For each i ($1 \leqslant i \leqslant n$), R_i, a subset of R, can be determined by checking whether the rule covers e_i, i.e.

$$R_i = \{r_j | r_j \in R,\ r_j \text{ covers } e_i\} \quad (i = 1, 2, \ldots, n).$$

The last task is in search of a subset of R, denoted by R^*, such that R^* covers all fuzzy examples and $|R^*| = Minimum$. R^* can be constructed by selecting one rule from each R_i ($1 \leqslant i \leqslant n$) and putting them together. Obviously, R^* generated by using this way is not unique. The main difficulty is how to select a suitable one from each R_i such that number of elements of the selected subset attains minimum.

In order to fulfill our last task, we give the following Theorem 2.

Theorem 2. *The last task of the optimal learning problem is NP-hard.*

The validity of Theorem 2 can be given by the following Lemma 1.

Lemma 1. *Let F be a finite set, $X_1, X_2, \ldots, X_m$ be m nonempty subsets of F. A problem of optimal selection of elements is in search of a subset Y ($Y \subset F$) generated by selecting one element from each X_i ($1 \leqslant i \leqslant m$), such that the number of elements of the subset Y, $|Y|$, attains minimum. The problem of optimal selection of elements is NP-hard.*

To prove Lemma 1, we briefly recall some concepts of the set cover.

Definition 9. Let T be a finite set, $F = \{S_1, S_2, \ldots, S_p\}$ be a family of subsets of T. We say F is a cover of T if $\bigcup_{i=1}^{p} S_i \supset T$. We say $|F^*|$ is a optimal cover of T if F^* is a cover of T and $|F^*| \leqslant |F|$ for an arbitrary cover of T, F, where $|\ |$ denotes the number of elements of a set.

The problem of optimal cover of a set mentioned in Definition 9 has been proved to be NP-hard in the literature [4].

Proof of Lemma 1. Without losing generality, we explicitly give the process of proof via examples. Let $F = \{1, 2, \ldots, 8\}$, $X_1 = \{1, 3, 4\}$, $X_2 = \{1, 2, 3\}$, $X_3 = \{3, 4, 5, 6\}$, $X_4 = \{3, 7, 8\}$, $X_5 = \{4, 6, 7\}$ and $X_6 = \{2, 5, 7, 8\}$. We consider the problem of optimal selection of elements in this case. The following Table 4 can be constructed by arranging X_i ($1 \leqslant i \leqslant 6$). The ith row of Table 4 is regarded as the set X_i ($1 \leqslant i \leqslant 6$) where elements which do not appear in the set X_i ($1 \leqslant i \leqslant 6$) are replenished by $*$ called dead element.

A path of Table 4 can be obtained by selecting one nondead element from each row. For instance, $1 \to 1 \to 3 \to 3 \to 4 \to 5$ and $3 \to 3 \to 3 \to 3 \to 7 \to 7$ are two paths of Table 4. The first path involves the columns 1, 3–5 of Table 4 while the second path involves only the third column and the seventh column of Table 4. Obviously, the problem of optimal selection of elements is equivalent to looking for a path involving least columns in Table 4.

Construct Table 5 via replacing every nondead element of Table 4 by the row label of the element.

It is obvious that searching for a path involving least columns in Table 4 is as same as in Table 5. We denote by $S = \{1, 2, 3, 4, 5, 6\}$, $S_1 = \{1, 2\}$, $S_2 = \{2, 6\}$, $S_3 = \{1, 2, 3, 4\}$, $S_4 = \{1, 3, 5\}$, $S_5 = \{3, 6\}$, $X_6 = \{3, 5\}$, $S_7 = \{4, 5, 6\}$, and $S_8 = \{4, 6\}$. Obviously, $\{S_1, S_2, \ldots, S_8\}$ constitutes a cover of S. The problem of searching for a path involving least columns in Table 5 is equivalent to the problem of optimal cover of the set S (An optimal cover of the set S is $\{S_3, S_7\}$). The problem of optimal cover of a set has been proved to be NP-hard in [4], therefore, the proof of this lemma is completed.

Because of the NP-hard, it is unrealistic to fulfill the last task by searching for an exact algorithm. The following is a heuristic algorithm for our last task, which is rather effective and intuitive.

Let $\Omega = \{R_1, R_2, \ldots, R_n\}$ be a family of nonempty subsets of set R. The initial value of R^* is supposed to be empty set (R and R^* have been indicated in the second task).

Repeat the following three steps:

Step 1. For each $r \in R$, compute the number of times with that r appears in the family Ω (i.e. compute $T(r) = \sum_{j=1}^{n} \lambda_j$ where $\lambda_j = 1$ if $r \in R_j$ and $\lambda_j = 0$ if $r \notin R_j$).

Step 2. Select r^* such that $T(r^*) = \mathrm{Max}_{1 \leqslant r \leqslant n} T(r)$.

Step 3. For $j = 1, 2, \ldots, n$, remove R_j from Ω if $r^* \in R_j$ and replace R^* with $\{r^*\} \cup R^*$.
Until Ω becomes empty.

Example 1 illustrates the computational process of the algorithm above.

Example 1. The initial state: $R_1 = \{6, 7\}$, $R_2 = \{1, 2, 3, 4\}$, $R_3 = \{1, 6\}$, $R_4 = \{1, 2, 4, 5\}$, $R_5 = \{2, 8\}$, $R_6 = \{3, 4, 5, 7\}$, $R_7 = \{2, 4\}$, $R_8 = \{4, 5\}$,

Table 4

	1	2	3	4	5	6	7	8
1	1	*	3	4	*	*	*	*
2	1	2	3	*	*	*	*	*
3	*	*	3	4	5	6	*	*
4	*	*	3	*	*	*	7	8
5	*	*	*	4	*	6	7	*
6	*	2	*	*	5	*	7	8

Table 5

	1	2	3	4	5	6	7	8
1	1	*	1	1	*	*	*	*
2	2	2	2	*	*	*	*	*
3	*	*	3	3	3	3	*	*
4	*	*	4	*	*	*	4	4
5	*	*	*	5	*	5	5	*
6	*	6	*	*	6	*	6	6

$\Omega = \{R_1, \ldots, R_8\}$, $R^* = empty$.

State 1. Max $T(r) = 5$, $r^* = 4$, $\Omega = \{R_1, R_3, R_5\}$, $R^* = \{4\}$

State 2. Max $T(r) = 2$, $r^* = 6$, $\Omega = \{R_5\}$, $R^* = \{4, 6\}$

State 3. Max $T(r) = 1$, $r^* = 2$, $\Omega = empty$, $R^* = \{2, 4, 6\}$ [*Stop*].

Example 2. Consider the optimal learning concerning Tables 1 and 3. Using the heuristic algorithm mentioned above, we give the following result of the learning (see Table 6).

Table 6 shows five fuzzy rules (the inconsistent degree of each fuzzy rule is 0), that is *overcast* $\Rightarrow P$ ($T = 0.85$, $I = 0$), *sunny*$\cap$*high* $\Rightarrow N$ ($T = 0.93$, $I = 0$), *rain*$\cap$*false* $\Rightarrow P$ ($T = 0.83$, $I = 0$), *rain*$\cap$*true* $\Rightarrow P$ ($T = 0.96$, $I = 0$) and *sunny*$\cap$*normal* $\Rightarrow P$ ($T = 0.96$, $I = 0$) where T is the true degree and I is the inconsistent degree.

4. Conclusions

This paper deals with the learning process of simplifying fuzzy rules, proves the main optimization problem is NP-hard, and gives a rather

Table 6
A learning result concerning Tables 1 and 3

Outlook	Temperature	Humidity	Windy	Class	True degree
Overcast	*	*	*	P	0.85
Sunny	*	High	*	N	0.93
Rain	*	*	False	P	0.83
Rain	*	*	True	N	0.96
Sunny	*	Normal	*	P	0.96

effective and intuitive heuristic algorithm. This heuristic regarded as a new fuzzy learning algorithm has the following advantages.

(1) It allows handling fuzziness existing in the process of learning.
(2) It possesses a rather rapid speed of training and matching.
(3) It is especially suitable for the learning problem on a large scale.
(4) It directly generates fuzzy rules without the aid of decision trees.
(5) It generates a family of fuzzy rules, which is approximately simplest.
(6) It is also fit for the learning problem of crisp case where the true degree is regarded as 1.

These advantages are verified through a typical case of learning from examples, knowledge acquisition for sleep states of human, which comes from Illinois University [7].

Acknowledgements

We wish to thank the anonymous referees for their very careful reading and their useful comments and suggestions.

References

[1] R.L.P. Chang, T. Pavlidis, Fuzzy decision tree algorithms, IEEE Trans. Systems Man Cybernet. SMC-7 (1997) 28–35.
[2] K.J. Cios, L.M. Sztandera, Continuous ID3 algorithm with fuzzy entropy measures, in: Proc. IEEE Internet. Conf. on Fuzzy Systems, San Diego, CA, 8–12 March 1992, pp. 469–476.
[3] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness.
[4] D.S. Johnson, Approximate algorithm for combinational problems, J. Comput. Systems Sci. 9 (1973) 3.
[5] G.J. Klir, Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like?, Fuzzy Sets and Systems 24 (1987) 141–160.
[6] B. Kosko, Neural Networks and Fuzzy Systems, Prentice-Hall, Englewood Cliffs, NJ, 1992.
[7] R.S. Michalski, I. Mozetic and J.R. Hong, The multipurpose incremental learning system AQ15, in: Proc. 5th AAAI (1986) 1041–1045.
[8] Z. Pawlak, Rough Sets, Kluwer, The Netherlands, 1991.
[9] J.R. Quinlan, Induction of decision trees, Mach. Learning 1(1) (1986) 81–106.
[10] D. Ruan, E.E. Kerre, Fuzzy implication operators and generalized fuzzy method of cases, Fuzzy Sets and System 54 (1993) 23–37.
[11] T. Tani, M. Sakoda, Fuzzy modeling by ID3 algorithm and its application of heater outlet temperature, in: Proc. IEEE Internat. Conf. on Fuzzy Systems, San Diego, CA, 8–12 March 1992, pp. 923–930.
[12] J.F. Traub, Algorithms and Complexity: New Directions and Recent Results, Academic Press, New York, 1976.
[13] R. Weber, Fuzzy ID3: a class of methods for automatic knowledge acquisition, in: Proc. 2nd Internat. Conf. on Fuzzy Logic & Neutral Networks, Iizuka, Japan, 17–22 July 1992, pp. 265–268.
[14] R. Weber, Automatic knowledge acquisition of fuzzy control applications, in: Proc. Internat. Symp. on Fuzzy Systems, Iizuka, Japan, 12–15 July 1992, pp. 9–12.
[15] Y. Yuan, M.J. Shaw, Induction of fuzzy decision trees, Fuzzy Sets and Systems 69 (1995) 125–139.