STUDIA INFORMATICA 2011
Volume 32 Number 2A (96)

Beata ZIELOSKO
Uniwersytet Śląski, Instytut Informatyki
Mikhail MOSHKOV
KAUST, MCSE Division

APPROXIMATE ALGORITHM FOR β-DECISION RULE OPTIMIZATION

Summary. In the paper greedy algorithm for construction of β-decision rules and algorithm for construction of β-complete systems of decision rules are studied. Obtined bounds on accuracy of the considered algorithms are presented.

Keywords: decision rules, greedy algorithm

ALGORYTM ZACHŁANNY DLA KONSTRUOWANIA β–REGUŁ DECYZYJNYCH

Streszczenie. W artykule został przedstawiony algorytm zachłanny dla konstruowania β-reguł decyzyjnych oraz dla konstruowania β-kompletnych systemów reguł decyzyjnych. Zostały zaprezentowane granice dokładności wyników uzyskiwanych za pomocą rozważanych algorytmów.

Słowa kluczowe: reguły decyzyjne, algorytm zachłanny

1. Introduction

When we use decision rules for knowledge representation, we would like to have relatively short rules. If exact decision rules are long, we can consider approximate rules. If we use decision rules in classifiers, then exact rules can be overfitted, i.e., dependent essentially on the noise or adjusted too much to the existing examples. In this case, it is more appropriate to work with approximate rules.

This idea is not new. For years, in rough set theory partial decision rules are studied intensively [2, 3, 4, 5].

In this paper, we consider one more approach to definition of the notion of approximate decision rules. We study so-called β-decision rules and algorithms for such rule construction. Our aim is to try to minimize the length of β-decision rules.

This paper consists of six sections. In Section 2 we discuss main notions. In Section 3, we study greedy algorithm for construction of approximate cover. Section 4 is devoted to the consideration of algorithms for construction of β-decision rules and β-complete systems of decision rules. In Section 5, we discuss complexity of the problem of minimization of β-decision rule length and complexity of the problem of optimization of β-complete system of decision rules. Section 6 contains conclusions.

2. Main Notions

A *binary decision table* is a rectangular table which elements belong to the set {0,1}. Columns of this table are labeled with attributes $f_1,\ldots,f_n$. Rows of the table are pairwise different and each row is labeled with a natural number (a decision). A decision which is attached to the maximum number of rows in T is called the *most common decision for T*. If we have more then one such decisions we choose the minimum one. If T is empty then 1 is the most common decision for T.

Let T be a binary decision table with n columns which are labeled with attributes $f_1,\ldots,f_n$. A *subtable* of the table T is a table obtained from T by removal some rows. Let $f_{i(1)},\ldots,f_{i(m)} \in \{f_1,\ldots,f_n\}$ and $\delta_1,\ldots,\delta_m \in \{0,1\}$. We denote by $T(f_{i(1)}, \delta_1)\ldots T(f_{i(m)}, \delta_m)$ the subtable of the table T which consists of rows that at the intersection with columns $f_{i(1)},\ldots,f_{i(m)}$ have numbers $\delta_1,\ldots,\delta_m$. We will say that T is a *degenerate* table if T does not have rows or all rows of T are labeled with the same decision. By $P(T)$ we denote the number of unordered pairs of rows from T labeled with different decisions. This parameter can be considered as an *uncertainty* of the table T.

Let β be a real number such that $0 \le \beta < 1$.

A *decision* rule over T is an expression of the kind

$$f_{i(1)}=b_1\wedge\ldots\wedge f_{i(m)}=b_m\rightarrow d$$

where $f_{i(1)},\ldots,f_{i(m)} \in \{f_1,\ldots,f_n\}$, $b_1,\ldots,b_m \in \{0,1\}$ and d is a natural number. The number m is called the *length* of the rule. Let $r=(\delta_1,\ldots,\delta_n)$ be a row of T. The considered rule is called a β-*decision rule for T and r* if $b_1=\delta_{i(1)},\ldots,b_m=\delta_{i(m)}$, d is the most common decision for the table $T'=T(f_{i(1)},b_1)\ldots(f_{i(m)},b_m)$ and $P(T')\leq\beta P(T)$. We denote by $L_\beta(T,r)$ the minimum length of β-decision rule for T and r.

The considered decision rule is called *realizable for r* if $b_1=\delta_{i(1)},\ldots,b_m=\delta_{i(m)}$. The considered rule is β-*true for T* if d is the most common decision for T' and $P(T')\leq\beta P(T)$.

A system S of decision rules over T is called a β-*complete system of decision rules for T*, if each rule from S is β-true for T and for each row r of T there exists a rule from S which is realizable for r. We denote $L(S)$ the maximum length of a rule from S and by $L_\beta(T)$ we denote the minimum value of $L(S)$ where minimum is considered among all β-complete systems of decision rules for T.

We consider two optimization problems: the problem of minimization of β-decision rule length and the problem of optimization of β-complete system of decision rules (the problem of minimization of the parameter $L(S)$).

3. Construction of Approximate Covers

We begin from an approximate algorithm for the minimization of cardinality of β-cover. Let β be a real number such that $0\leq\beta<1$.

Let A be a set containing $N>0$ elements, and $F=\{S_1,\ldots,S_p\}$ be a family of subsets of the set A such that $A=\cup^p_{i=1} S_i$. A subfamily $\{S_{i(1)},\ldots,S_{i(t)}\}$ of the family F will be called B-cover for A, F if $|\cup^t_{j=1} S_{i(j)}|\geq(1-\beta)|A|$. The problem of searching for a β-cover with minimum cardinality is NP-hard [6].

We consider a greedy algorithm for construction of β-cover. During each step this algorithm chooses a subset from F which covers maximum number of uncovered elements from A. This algorithm stops when the constructed subfamily is an β-cover for A,F. We denote by $C_{\text{greedy}}(\beta)$ the cardinality of constructed β-cover, and by $C_{\min}(\beta)$ we denote the minimum cardinality of β-cover for A,F. For the completeness, we consider our own proof of the following statement from [1].

Theorem 1 [1] *Let* $0<\beta<1$. *Then* $C_{\text{greedy}}(\beta)<C_{\min}(0)\ln(1/\beta)+1$.

Proof Denote $m=C_{\min}(0)$. If $m=1$ then, as it is not difficult to show, $C_{\text{greedy}}(\beta)=1$ and the considered inequality holds. Let $m\geq 2$ and S_i be a subset of maximum cardinality in F. It is clear that $|S_i|\geq N/m$. So after the first step we will have at most $N-N/m=N(1-1/m)$ uncovered elements in the set A. After the first step we have the following set cover problem: the set $A\backslash S_i$ and the family $\{S_1\backslash S_i,\ldots,S_p\backslash S_i\}$. For this problem, the minimum cardinality of a cover is at most m. So after the second step, when we choose a set $S_j\backslash S_i$ with maximum cardinality, the number of uncovered elements in the set A will be at most $N(1-1/m)^2$, etc.

Let the greedy algorithm in the process of β-cover construction make g steps and construct a β-cover of cardinality g. Then after the step number $g-1$ more then βN elements in A are uncovered. Therefore $N(1-1/m)^{g-1} > \beta N$ and $1/\beta > (1+1/(m-1))^{g-1}$. If we take the natural logarithm of both sides of this inequality we obtain $\ln(1/\beta) > (g-1)\ln(1+1/(m-1))$. It is known that for any natural r, the inequality $\ln(1+1/r) > 1/(r+1)$ holds. Therefore $\ln(1/\beta) > (g-1)/m$ and $g < m\ln(1/\beta)+1$. Taking into account that $m=C_{\min}(0)$ and $g=C_{\text{greedy}}(\beta)$, we obtain $C_{\text{greedy}}(\beta) < C_{\min}(0)\ln(1/\beta)+1$.

4. Construction of Approximate Decision Rules and Systems of Rules

We can apply the greedy algorithm for construction of β-cover to construct β-decision rules.

Let T be a nondegenerate decision table containing n columns labeled with attributes $f_1,\ldots,f_n$, $r=(\delta_1,\ldots,\delta_n)$ be a row of T, and β be a real number such that $0 < \beta < 1$. We consider a set cover problem $A(T,r)$, $F(T,r)=\{S_1,\ldots,S_n\}$ where $A(T,r)$ is the set of all unordered pairs of rows from T with different decisions. For $i=1,\ldots,n$, the set S_i coincides with the set of all pairs from $A(T,r)$ such that at least one row from the pair has at the intersection with the column f_i a number different from b_i. One can show that the decision rule

$$f_{i(1)}=\delta_{i(1)}\wedge\ldots\wedge f_{i(m)}=\delta_{i(m)}\rightarrow d$$

is β-true for T (it is clear that this rule is realizable for r) if and only if d is the most common decision for the table $T(f_{i(1)},\delta_{i(1)})\ldots(f_{i(m)},\delta_{i(m)})$ and $\{S_{i(1)},\ldots,S_{i(m)}\}$ is a β-cover for the set cover problem $A(T,r)$, $F(T,r)$. Evidently, for the considered set cover problem $C_{\min}(0)=L_0(T,r)$.

Let us apply the greedy algorithm to the considered set cover problem. This algorithm constructs a cover which corresponds to an β-decision rule for T and r. From Theorem 1 it follows that the length of this rule is at most

$$L_0(T,r)\ln(1/\beta)+1$$

We denote by $L_{\text{greedy}}(T,r,\beta)$ the length of the rule constructed by the following polynomial algorithm: for a given decision table T, row r of T and β, $0 < \beta < 1$, we construct the set cover problem $A(T,r)$, $F(T,r)$ and then apply to this problem the greedy algorithm for construction of β-cover. We transform the obtained β-cover to a β-decision rule for T and r. According to what has been said above we have the following statement.

Theorem 2 Let T be a nondegenerate decision table, r be a row of T and β be a real number such that $0 < \beta < 1$. Then

$$L_{greedy}(T,r,\beta)\leq L_0(T,r)\ln(1/\beta)+1.$$

We can use the considered algorithm to construct a β-complete decision rule system for T. To this end, we apply the algorithm sequentially for the table T, number β and each row r of T. As a result, we obtain a system of rules S in which each rule is β-true for T and for every row of T there exists a rule from S which is realizable for this row. We denote $L_{\text{greedy}}(T,\beta)=L(S)$. From Theorem 1 it follows

Theorem 3 Let T be a nondegenerate decision table and β be a real number such that $0 < \beta < 1$. Then

$$L_{greedy}(T,\beta)\leq L_0(T)\ln(1/\beta)+1.$$

5. Complexity of Problems

Let us consider a set cover problem A,F where $A=\{a_1,\ldots,a_N\}$ and $F=\{S_1,\ldots,S_m\}$. We define a decision table $T(A,F)$. This table has m columns corresponding to sets $S_1,\ldots,S_m$ respectively, and N+1 rows. For j=1,…,N the j-th row corresponds to the element a_j. The last (N+1)-th row is filled by 0. For j=1,…,N and i=1,…,m at the intersection of j-th row and i-th column 1 stays if and only if $a_j \in S_i$. The decision corresponding to the last row is equal to 2. All other rows are labeled with the decision 1.

One can show that a subfamily $\{S_{i(1)},\ldots,S_{i(t)}\}$ is a β-cover for A,F, $0 \leq \beta < 1$, if and only if the decision rule

$$f_{i(1)}=0\wedge\ldots\wedge f_{i(t)}=0\rightarrow d$$

is a β-decision rule for $T(A,F)$ and the last row of $T(A,F)$ for some $d \in \{1,2\}$.

So we have a polynomial time reduction of the problem of minimization of β-cover cardinality to the problem of minimization of β-decision rule length. Since the first problem is NP-hard [6], we have

Proposition 1 For any β, $0 \leq \beta < 1$, the *problem* of minimization of β-decision rule length is NP-hard.

Let β be a real number such that $0 < \beta < 1$. Let us consider the decision table $T(A,F)$. For j=1,…,$N+1$, we denote by r_j the j-th row of $T(A,F)$. Let $j \in \{1,\ldots,N\}$. We know that there exists a subset $S_i \in F$ such that $a_j \in S_i$. Therefore the decision rule

$$f_i=1\rightarrow 1$$

is a β-decision rule for $T(A,F)$ and r_j. It is clear that $L_\beta(T(A,F),r_j) \geq 1$. Hence, $L_\beta(T(A,F),r_j)=1$. From here it follows that $L_\beta(T(A,F))=L_\beta(T(A,F),r)$ where $r=r_{N+1}$. So if we find a β-complete decision rule system S for $T(A,F)$ such that $L(S)=L_\beta(T(A,F))$ then in this system we will find a β-decision rule of the kind

$$f_{i(1)}=0\wedge\ldots\wedge f_{i(t)}=0\rightarrow d$$

for which $t=L_\beta(T(A,F),r)$. We know that $\{S_{i(1)},\ldots,S_{i(t)}\}$ is a β-cover for A,F with minimum cardinality. Since the problem of minimization of cardinality of β-cover is *NP*-hard, we have.

Proposition 2 The problem of optimization of β-decision rule system is *NP*-hard for any β, $0 < \beta < 1$.

6. Conclusions

The paper is devoted to the consideration of approximate β-decision rules. We studied algorithms for construction of β-decision rules and β-complete systems of decision rules that can be used as classifiers. We proved that the problem of minimization of β-decision rule length and the problem of optimization of β-complete systems of decision rules are *NP*-hard.

In the book [2], α-decision rules are studied which are defined in an another way.

Let T have n columns labeled with attributes $f_1,\ldots,f_n$, and $r=(\delta_1,\ldots,\delta_n)$ be a row of T labeled with the decision d. We denote by $P(T,r)$ the number of rows from T with decision different from d. A decision rule

$$f_{i(1)}=\delta_{i(1)}\wedge\ldots\wedge f_{i(m)}=\delta_{i(m)}\rightarrow d$$

is called an α-decision rule for T and r if attributes $f_{i1},\ldots,f_{im}$ separate from r at least $(1-\alpha)P(T,r)$ rows with decisions different from d.

The book [2] contains bounds on complexity and algorithms for construction of α-decision rules. In particular, it is proven that a simple greedy algorithm (under some assumptions on the class *NP*) is close to the best polynomial approximate algorithms for minimization of α-decision rule length.

BIBLIOGRAPHY

1. Cheriyan J., Ravi R.: Lecture Notes on Approximation Algorithms for Network Problems. 1998, http://www.math.uwaterloo.ca/˜jcheriya/lecnotes.html.
2. Moshkov M., Piliszczuk M., Zielosko B.: Partial Covers, Reducts and Decision Rules in Rough Sets: Theory and Applications. Springer book series Studies in Computational Intelligence, vol. 145, Springer, Berlin, Heidelberg 2008.
3. Nguen H. S.: Approximate Boolean reasoning. [in:] Peters J. F., Skowron A. (eds.): LNCS Transactions on Rough Sets V. LNCS, vol. 4100. Springer, Heidelberg 2006.
4. Nguen H. S., Slezak D.: Approximate reducts and association rules – correspondence and complexity results. [in:] Zhong N., Skowron A., Ohsuga S. (eds.): RSFDGrC 1999, LNCS (LNAI), Vol. 1711, Springer, Heidelberg 1999.

5. Pawlak Z.: Rough Sets – Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht 1991.
6. Slezak D.: Normalized decision functions and measures for inconsistent decision tables analysis. Fundamenta Informaticae 44(3), 2000, p. 291÷319.

Recenzent: Dr inż. Marek Sikora

Wpłynęło do Redakcji 16 stycznia 2011 r.

Omówienie

Reguły decyzyjne mogą być traktowane jako sposób reprezentacji wiedzy i w tym przypadku długość reguły ma znaczenie. Zamiast długich, dokładnych reguł decyzyjnych możemy stosować reguły przybliżone (aproksymacyjne), zawierające mniejszą liczbę atrybutów. Jeśli reguły decyzyjne stosowane są do budowy klasyfikatorów, wówczas dokładne reguły, z dużą liczbą atrybutów mogą być przeuczone, tj. zbyt mocno dopasowane do istniejących przykładów lub zależne od szumu informacyjnego. W tym przypadku bardziej odpowiednie są reguły aproksymacyjne. Ta idea nie jest nowa, od wielu lat jest badana m.in w teorii zbiorów przybliżonych.

W artykule autorzy przedstawiają kolejne podejście do definiowania pojęcia przybliżonych reguł decyzyjnych. Został przedstawiony algorytm zachłanny dla konstruowania β-reguł decyzyjnych oraz dla konstruowania β-kompletnych systemów reguł decyzyjnych oraz granice dokładności wyników uzyskiwanych za pomocą tych algorytmów. Ich celem jest minimalizacja długości β-reguł decyzyjnych.

Addresses

Beata ZIELOSKO: Uniwersytet Śląski, Instytut Informatyki, ul. Będzińska 39, Sosnowiec 41-200, Polska, beata.zielosko@us.edu.pl.

Mikhail MOSHKOV: King Abdullah University of Science and Technology, Mathematical and Computer Sciences &Engineering Division, Thuwal, 23955-6900, Saudi Arabia, mikhail.moshkov@kaust.edu.sa.