Polska Akademia Nauk • Instytut Badań Systemowych

AUTOMATYKA
STEROWANIE
ZARZĄDZANIE

Książka jubileuszowa
z okazji
70-lecia urodzin

PROFESORA KAZIMIERZA MAŃCZAKA

pod redakcją
Jakuba Gutenbauma

Polska Akademia Nauk • Instytut Badań Systemowych

AUTOMATYKA STEROWANIE ZARZĄDZANIE

Książka jubileuszowa
z okazji
70-lecia urodzin

PROFESORA KAZIMIERZA MAŃCZAKA

pod redakcją
Jakuba Gutenbauma

Warszawa 2002

Książka jubileuszowa z okazji
70-lecia urodzin
Profesora Kazimierza MAŃCZAKA

Redaktor
prof. dr hab. inż. Jakub Gutenbaum

Warszawa 2002

ISBN 83-85847-78-2

Wydawca: Instytut Badań Systemowych PAN
ul. Newelska 6 01-447 Warszawa
http://www.ibspan.waw.pl

Opracowanie składopisu: Anna Gostyńska, Jadwiga Hartman

Druk: KOMO-GRAF, Warszawa
nakład 200 egz., 34 ark. wyd., 31 ark. druk.

ZASTOSOWANIE ZMODYFIKOWANEGO ZADANIA POKRYCIA W UCZENIU MASZYNOWYM

Grażyna Szkatuła
Instytut Badań Systemowych PAN

Abstract: *In this paper we propose an improved inductive learning method, derive classification rules that correctly describe most of the examples belonging to a class and do not describe most of the examples not belonging to this class. The learning problem is represented as a modification of the set covering problems. The results are very encouraging.*

Keywords: *learning from examples, classification rules, set covering problem.*

1. Wprowadzenie

Rozwój systemów informatycznych ułatwił gromadzenie dużych zbiorów danych. Ważne stało się nie tylko efektywne przechowywanie takich danych, ale także ich analiza, zdolność interpretacji i wyciągania potencjalnie użytecznych i zrozumiałych dla człowieka wniosków. Wykorzystanie pełnej informacji zawartej w tych zbiorach oraz sformułowanie ukrytej w nich wiedzy zwykle przekracza możliwości percepcji człowieka. W związku z tym powstało zapotrzebowanie na nowe metody i narzędzia informatyczne, które mogą pomóc człowiekowi w odkrywaniu wiedzy z danych. Duże znaczenie zyskały metody uczenia maszynowego wykorzystujące informacje zawarte w dostarczonym zbiorze przykładów. W pracy ograniczono się do zagadnień uczenia nadzorowanego, w którym tworzone są opisy klas zawierających przykłady posiadające pewne wspólne właściwości, które odróżniają je od innych klas. Zadanie uczenia maszynowego na podstawie przykładów rozumiane jest w pracy w sposób przedstawiony poniżej.

Przyjmijmy, że dysponujemy dużym zbiorem danych, np. zbiorem historii choroby pacjentów. Dane w nich zawarte dotyczą wyników badań laboratoryjnych, badań lekarskich, rozpoznania, zastosowanego leczenia i jego wyników. Każdego pacjenta można opisać przy pomocy wybranego zbioru cech, takich jak np. „lepkość plazmy", „pH krwi", „występowanie choroby wieńcowej", itp. Warunek związany z daną cechą można zapisać w postaci („cecha" = „wartość cechy"), np. („fibrynogen" = "większy od 298"), („występowanie choroby wieńcowej" = "nie występuje"). Przykładem nazywa się koniunkcję warunków związanych ze wszystkimi cechami określonymi dla danego pacjenta.

Zbiór takich przykładów, ze względu na wartości wybranej cechy (zwanej cechą decyzyjną), można podzielić na rozłączne klasy. Przez klasę rozumie się zbiór przykładów mających tę samą wartość cechy decyzyjnej. Np. dla cechy decyzyjnej „występowanie choroby wieńcowej" z wartościami {„występuje", „nie występuje"} można wyróżnić dwie klasy. Pierwszą klasę stanowią pacjenci u których „występuje" choroba wieńcowa, a drugą pacjenci u których „nie występuje" choroba wieńcowa.

Na podstawie zbioru tak opisanych przykładów (zwanego zbiorem uczącym) metody uczenia maszynowego tworzą opisy rozpatrywanych klas. Opis klasy wyrażany jest na ogół w postaci *reguł* mających postać wyrażeń logicznych typu *"JEŻELI spełnione są określone warunki związane z cechami TO zachodzi przynależność do danej klasy"*; *drzew decyzyjnych*; lub też w postaci odpowiednio dobranych wag połączeń w *sieciach neuronowych* i ich struktury.

Reprezentacja opisu klasy w postaci reguł uznawana jest za bardziej czytelną dla człowieka niż inne reprezentacje i dlatego będzie stosowana w pracy. Część warunkowa reguły będzie określana w postaci koniunkcji warunków określonych dla wybranego podzbioru cech opisujących przykłady, a część decyzyjna będzie określała przynależność przykładu spełniającego te warunki do danej klasy. Przykład takiej reguły przedstawiono poniżej.

JEŻELI („lepkość plazmy" = "większa od 1.9") ∧
            („fibrynogen" = "większy od 298")
TO („występowanie choroby wieńcowej" = "występuje")

Utworzone reguły mogą być zastosowane do klasyfikowania nowych przykładów, dla których nie jest znana przynależność do klasy. Mogą być również użyte do uzyskania nie znanych wcześniej w sposób jawny informacji o istniejących w zbiorze danych zależnościach.

W pracy przyjęto, że tworzone reguły powinny spełniać pewne dodatkowe wymagania. Ze względu na fakt, że w rzeczywistych zbiorach danych pozyskanych w wyniku pomiarów, testów, wywiadów z ekspertami z różnych dziedzin itp. mogą istnieć błędy, nie należy wymagać, aby reguły były dokładnie dopasowane do przykładów uczących, ale można przyjąć, że powinny poprawnie opisywać "większość" przykładów należących do rozpatrywanej klasy i nie opisywać "prawie wszystkich" przykładów do tej klasy nie należących. Powinny również mieć minimalną długość, np. w sensie liczby warunków tworzących regułę lub łącznej ważności warunków występujących w regule (z punktu widzenia przyjętego kryterium).

Założono, że cechy nie są tak samo ważne i można wprowadzić współczynniki ważności dla każdej cechy i dla każdej wartości którą może dana cecha przyjmować. Przyjęto, że większe/mniejsze wartości współczynników są odpowiednio korzystne/niekorzystne z punktu widzenia rozpatrywanego kryterium. Współczynniki te określają pewną preferencję cech, która będzie uwzględniana przy tworzeniu reguł. Reguły, zawierające więcej warunków związanych z cechami preferowanymi, powinny lepiej opisywać zbiór przykładów, z punktu widzenia rozpatrywanego kryterium.

Można uwzględniać różne kryteria, przykładowo odporność na błędy pomiaru, większą wiarygodność czy też istotność w procesie diagnozy danego przypadku i przy współpracy z ekspertami zastosować dla określenia współczynników ważności np. podejście AHP (ang. Analytical Hierarchy Process) (Saaty 1980). Przykładowo, bardziej wiarygodna będzie reguła zawierająca warunki związane z cechami mniej podatnymi na błędne określenie ich wartości i wygodniej będzie stosować w praktyce regułę, w której występuje łatwa do określenia temperatura, niż wynik skomplikowanego badania krwi (Szkatuła 1995).

Współczynniki można również tworzyć w oparciu o analizę całego zbioru przykładów uczących, w celu preferowania w procesie uczenia tych wartości cech, które są bardziej „typowe" dla rozpatrywanej klasy. Utworzone reguły zawierające preferowane wartości cech powinny zawierać mniejszą liczbę warunków i lepiej opisywać zbiory danych (Szkatuła 1995 oraz Kacprzyk, Szkatuła 1998, 1999).

Reguły, o których była mowa wyżej, można tworzyć stosując różne metody uczenia maszynowego. W pracy, wprowadzono pojęcie długości ważonej reguły (uwzględniającej współczynniki ważności), której minimalizacja jest zgodna z naszymi oczekiwaniami znalezienia reguły krótkiej, zawierającej małą liczbę mniej preferowanych cech. Zadanie tworzenia reguł dla danej klasy, przedstawiono w postaci ciągu zadań cząstkowych, sformułowanych w postaci pewnych modyfikacji zadania pokrycia (Szkatuła 1995). Zaproponowano metodę heurystyczną tworzącą reguły typu *JEŻELI ...TO ...*, o minimalnej długości ważonej prezentujące istniejące regularności w przykładach, z możliwością uwzględniania założenia o poprawnym opisywaniu przez nie „większości” przykładów uczących.

W rozdziale drugim przedstawiono rozpatrywane zadanie oraz przyjęte założenia. Rozdział trzeci zawiera podstawy teoretyczne zaproponowanej metody i krótki opis algorytmu.

2. Sformułowanie problemu

Niech U będzie niepustym, skończonym zbiorem przykładów, $A=\{a_1,\dots,a_K\}$ niepustym, skończonym zbiorem cech. $V_{a_j}=\{v_{j_1},\dots,v_{j_j}\}$, $V_{a_j}\neq\emptyset$, jest zbiorem wartości cechy a_j, dla $j=1,\dots,K$ oraz $V=\bigcup_{j=1,\dots,K} V_{a_j}$. Funkcja $f\colon U\times A\to V$ taka, że $\forall e\in U$, $\forall a_j\in A$ zachodzi $f(e,a_j)\in V_{a_j}$ określa jaką wartość w przykładzie e przyjmuje cecha a_j. Niech $g(v_{t(j)})\in(0,1)$ oznacza współczynnik ważności dla cechy a_j, $j=1,\dots,K$ i wartości $v_{t(j)}\in V_{a_j}$ którą dana cecha przyjmuje. Indeks $t(j)$ określa jaką wartość przyjmuje j-ta cecha, $t(j)\in\{j_1,\dots,j_j\}$.

Warunkiem elementarnym dla cechy a_j, $j=1,\dots,K$, i przykładu $e\in U$ nazywamy wyrażenie postaci $(a_j=v_{t(j)};\ g(v_{t(j)}))$, gdzie $v_{t(j)}\in V_{a_j}$, $v_{t(j)}=f(e,a_j)$, $g(v_{t(j)})\in(0,1)$. Każdy przykład $e\in U$ można opisać w postaci koniunkcji K warunków elementarnych w sposób następujący

$$e=\bigwedge_{j=1}^{K}(a_j=v_{t(j)};\ g(v_{t(j)})) \qquad (1)$$

dla $a_j\in A$, $v_{t(j)}\in V_{a_j}$, $g(v_{t(j)})\in(0,1)$. *Koniunkcję* l warunków elementarnych, $l\le K$, dla wszystkich cech należących do pewnego podzbioru cech $\{a_{j_1},\dots,a_{j_l}\}\subseteq A$ można zapisać w sposób następujący

$$C^l=\bigwedge_{j\in I}(a_j=v_{t(j)};\ g(v_{t(j)})) \qquad (2)$$

gdzie $I=\{j_1,\dots,j_l\}\subseteq\{1,\dots,K\}$, $card(I)=l$, $v_{t(j)}\in V_{a_j}$, $g(v_{t(j)})\in(0,1)$.

Mówimy, że przykład $e\in U$ jest *opisany* przez koniunkcję C^l jeżeli $\forall j\in I$ zachodzi warunek $f(C^l,a_j)=f(e,a_j)$.

Przykład. Koniunkcja (a_1="kobieta"; 0.6) $\wedge$ (a_3="około 35 lat"; 0.3) opisuje przykład (a_1="kobieta";0.6) $\wedge$ (a_2="wysoka";0.6) $\wedge$ (a_3="około 35 lat";0.3) ale nie opisuje przykładu (a_1="mężczyzna"; 0.4) $\wedge$ (a_2="niski"; 0.2) $\wedge$ (a_3="około 35 lat"; 0.3).

Niech przykład e będzie opisany w postaci koniunkcji K warunków elementarnych $(a_l = v_{t(l)}; g(v_{t(l)})) \wedge ... \wedge (a_K = v_{t(K)}; g(v_{t(K)}))$, wzór (1), z której wybrano l - warunków, $l \leq K$, tworząc koniunkcję postaci C^l = $(a_{j_l} = v_{t(j_l)}; g(v_{t(j_l)})) \wedge ... \wedge (a_{j_l} = v_{t(j_l)}; g(v_{t(j_l)}))$, wzór (2), odpowiadającą zbiorowi indeksów $I = \{j_l, ..., j_l\} \subseteq \{1, ..., K\}$. Mówimy, że *koniunkcja* C^l *jest równoważna wektorowi* zero-jedynkowemu $x = [x_j]^T$, $j = 1, ..., K$, takiemu, że $x_j = 1$ jeżeli warunek $(a_j = v_{t(j)}; g(v_{t(j)}))$ z przykładu e występuje w koniunkcji C^l, a $x_j = 0$ w przeciwnym przypadku.

Przykład. Niech przykład e = (a_1="wysoki";0.9) $\wedge$ (a_2="blond";0.8) $\wedge$ (a_3="niebieski";0.3). Wektor $[0,1,1]^T$ jest równoważny koniunkcji (a_2 = "blond"; 0.8) $\wedge$ (a_3="niebieski"; 0.3); a koniunkcja (a_1 = "wysoki"; 0.9) jest równoważna wektorowi $[1,0,0]^T$.

Jeżeli ze zbioru cech A wybierzemy jedną cechę a_d, to możemy ze względu na wartości które ona przyjmuje dokonać podziału zbioru przykładów na rozłączne klasy. Elementy zbioru $A \setminus \{a_d\}$ nazywamy *cechami warunkowymi*, a cechę a_d nazywamy *cechą decyzyjną*.

Podziałem zbioru przykładów U ze względu na cechę decyzyjną $a_d \in A$ mającą dziedzinę $V_{a_d} = \{v_{d_l}, ..., v_{d_d}\}$ nazywamy niepuste podzbiory przykładów $\{U_{v_{d_l}}, U_{v_{d_2}}, ..., U_{v_{d_d}}\}$, gdzie $\forall\, v_{t(d)} \in V_{a_d}$

$$U_{v_{t(d)}} = \{e \in U \colon f(e, a_d) = v_{t(d)}\} \tag{3}$$

oraz $U_{v_{d_l}} \cup ... \cup U_{v_{d_d}} = U$, $U_{v_{d_i}} \cap U_{v_{d_j}} = \emptyset$ dla $i \neq j$. Tak więc cecha decyzyjna dzieli zbiór przykładów na niepuste, rozłączne i w sumie tworzące cały zbiór podzbiory, które nazywamy *klasami*, i których opisy w postaci reguł będziemy tworzyć. Przykłady $e \in U$ należące do klasy $U_{v_{t(d)}}$ nazywamy *pozytywnymi* dla tej klasy i oznaczamy

$$POS_A(U_{v_{t(d)}}) = \left\{e \in U;\ f(e,a_d) = v_{t(d)}\right\} \tag{4}$$

a pozostałe przykłady nie należące do tej klasy, dla których nie istnieje w zbiorze przykładów pozytywnych przykład o takich samych wartościach wszystkich cech ze zbioru $A \setminus \{d_d\}$, nazywamy przykładami *negatywnymi*

$$NEG_A(U_{v_{t(d)}}) = \{e \in U : f(e,a_d) \neq v_{t(d)}\ i\ \forall e' \in POS_A(U_{v_{t(d)}})$$

$$\exists a_j \in A \setminus \{a_d\},\ f(e,a_j) \neq f(e',a_j)\}. \tag{5}$$

Należy zauważyć, że zachodzą następujące zależności $POS_A(U_{v_{t(d)}}) \neq \emptyset$, $NEG_A(U_{v_{t(d)}}) \neq \emptyset$, $POS_A(U_{v_{t(d)}}) \cap NEG_A(U_{v_{t(d)}}) = \emptyset$.

Regułą elementarną dla klasy $U_{v_{t(d)}}$ nazywamy wyrażenie postaci

$$JE\dot{Z}ELI\ C^I\ TO\ (a_d = v_{t(d)}) \tag{6}$$

gdzie koniunkcję C^I określa wzór (2). *Reguła* dla klasy $U_{v_{t(d)}}$ tworzona jest w postaci dysjunkcji (oznaczanej "$\cup$") reguł elementarnych do tej klasy

$$JE\dot{Z}ELI\ C^{I_1} \cup \ldots C^{I_L}\ TO\ (a_d = v_{t(d)}) \tag{7}$$

$I_l \subseteq \{1,\ldots,K\}$, $l = 1,\ldots,L$, $C^{I_l} = \underset{j \in I_l}{\wedge} (a_j = v_{t(j,l)};\ g(v_{t(j,l)}))$, $v_{t(j,l)} \in V_{a_j}$.

Indeks $t(j,l)$ określa wartość jaką przyjmuje cecha a_j w koniunkcji I_l.

Można zauważyć, że dla koniunkcji C^I utworzonej z $l \leq K$ warunków wybranych z przykładu $e \in U$, równoważny jej K-elementowy wektor x ma elementy $x_j = 1$, $\forall j \in I$, a dla $j \in \{1,2,\ldots,K\} \setminus I$ zachodzi warunek $x_j = 0$.

Długość ważoną koniunkcji $C^I = \underset{j \in I}{\wedge} (a_j = v_{t(j)};\ g(v_{t(j)}))$ można zdefiniować w sposób następujący

$$d(C^I) = \sum_{j \in I} (1 - g(v_{t(j)})) \cdot x_j + \sum_{j \in \{1,2,\dots,K\} \setminus I} (1 - g(v_{t(j)})) \cdot x_j =$$
$$= \sum_{j=1}^{K} (1 - g(v_{t(j)})) \cdot x_j \tag{8}$$

Długość ważoną reguły, której część warunkowa złożona jest z L koniunkcji, $R = C^{I_1} \cup \dots \cup C^{I_L}$ zgodnie ze wzorem (7), można określić jako największą długość ważoną tworzących ją koniunkcji:

$$d_R(C^{I_1} \cup \dots \cup C^{I_L}) = \max_{l=1,\dots,L} d(C^{I_l}) \tag{9}$$

Zadanie uczenia maszynowego można sformułować następująco: na podstawie zbioru przykładów pozytywnych i negatywnych dla danej klasy należy utworzyć regułę o minimalnej długości ważonej

$$\min_{I_1,\dots,I_L} d_R(C^{I_1} \cup \dots \cup C^{I_L}) \tag{10}$$

która opisuje „prawie wszystkie" przykłady pozytywne oraz nie opisuje „prawie wszystkich" przykładów negatywnych dla rozpatrywanej klasy. Powyższą minimalizację należy rozumieć jako zadanie znalezienia cech, które występują we wszystkich koniunkcjach i określenie wartości, które te cechy przyjmują. Zadanie (10) można uprościć do ciągu zadań cząstkowych:

$$\min_{I_1} d(C^{I_1}), \dots, \min_{I_L} d(C^{I_L}) \tag{11}$$

gdzie minimalizacja w każdym zadaniu cząstkowym jest rozumiana jako zadanie znalezienia cech, które występują w danej koniunkcji i określenie wartości, które te cechy przyjmują.

3. Opis zastosowanej metody

Załóżmy, że mamy niepusty, skończony zbiór przykładów U i niepusty, skończony zbiór cech $A = \{a_1, \dots, a_K\} \cup \{a_d\}$. $V_{a_j} = \{v_{j_1}, \dots, v_{j_{l_j}}\}$, $V_{a_j} \neq \emptyset$, jest zbiorem wartości cechy a_j, $j = 1, \dots, K$ oraz $V = \bigcup_{j=1,\dots,K} V_{a_j}$.

Cecha decyzyjna a_d ze zbiorem wartości $V_{a_d} = \{v_{d_1}, \dots, v_{d_d}\}$ dzieli zbiór U na rozłączne klasy $\{U_{v_{d_1}}, U_{v_{d_2}}, \dots, U_{v_{d_d}}\}$. Należy utworzyć reguły klasyfikacji dla każdej klasy $U_{v_{t(d)}}$ tak, aby poprawnie opisywały

„większość" przykładów uczących. Przyjmijmy, że mamy P przykładów pozytywnych dla klasy $U_{v_{t(d)}}$, wzór (4), zapisanych w postaci $e^p = \bigwedge_{j=1}^{K}(a_j = v_{t(j,p)}; g(v_{t(j,p)}))$, $p = 1,...,P$, oraz N przykładów negatywnych, wzór (5), w postaci $e^n = \bigwedge_{j=1}^{K}(a_j = v_{t(j,n)}; g(v_{t(j,n)}))$, $n = 1,...,N$. Indeks $t(j,p)$ określa wartość, jaką przyjmuje cecha a_j w przykładzie e^p, analogicznie należy rozumieć indeks $t(j,n)$.

Dla danego przykładu e^p i wszystkich przykładów negatywnych e^n tworzymy macierz zero-jedynkową $Z_{N*K}=[z_{nj}]$, $n = 1,...,N$, $j = 1,...,K$,

$$z_{nj} = \begin{cases} 1 & jeżeli\ v_{t(j,p)} \neq v_{t(j,n)} \\ 0 & jeżeli\ v_{t(j,p)} = v_{t(j,n)} \end{cases} \tag{12}$$

Wiersze tak określonej macierzy Z_{N*K} odpowiadają przykładom negatywnym e^n, $n = 1,...,N$, a kolumny cechom $a_1,...,a_K$. Zachodzi warunek $z_{nj} = 1$ jeżeli cecha a_j przyjmuje w przykładzie pozytywnym e^p i w przykładzie negatywnym e^n różne wartości; $z_{nj} = 0$ jeżeli wartości te są sobie równe. W macierzy nie występują wiersze zerowe, ponieważ dla każdego przykładu pozytywnego i każdego negatywnego zawsze istnieje co najmniej jedna cecha przyjmująca różne wartości w obu przykładach, zgodnie z (4) i (5).

Rozważmy przykład e^p. Każdy zero-jedynkowy i K-elementowy wektor x określa koniunkcję C^I, $I \subseteq \{1,...,K\}$, złożoną z wybranych warunków z przykładu e^p w sposób opisany poprzednio.

Niech wektor zero-jedynkowy $x = [x_1,...,x_K]^T$ określający koniunkcję C^I spełnia dodatkowo następujące ograniczenia

$$\sum_{j=1}^{K} z_{nj}x_j \geq \gamma_n, \qquad n = 1,...,N \tag{13}$$

gdzie $\gamma = [\gamma_1,...,\gamma_N]^T$ jest wektorem jednostkowym. Wtedy $\forall n$, istnieje co najmniej jedno $j \in I$ dla którego $z_{nj}x_j \geq 1$. Wnioskując, możemy stwierdzić, że koniunkcja określona przez taki wektor x nie opisuje wszystkich przykładów negatywnych (ponieważ zawiera co najmniej jeden warunek który różni ją od każdego przykładu negatywnego). Oczywiście, zawsze opisuje ona co najmniej jeden przykład pozytywny (na pewno opisuje przykład e^p).

Niech wektor zero-jedynkowy $x = [x_1, ..., x_K]^T$ określający koniunkcję C^I spełnia ograniczenia (13) gdzie $\gamma = [\gamma_1, ..., \gamma_N]^T$ jest tym razem wektorem zero-jedynkowym. Analogicznie rozumując można stwierdzić, że koniunkcja C^I na pewno nie opisuje tych przykładów negatywnych e^n, dla których $\gamma_n = 1$; może natomiast opisywać te przykłady negatywne, dla których $\gamma_n = 0$, ponieważ $\forall j \in I$ może być spełniony warunek $z_{nj} x_j = 0$.

Minimalizację w zadaniu cząstkowym, wzór (11), można zapisać

$$\min_{x,\gamma:\, Zx \geq \lambda} \sum_{j=1}^{K} (1 - g(v_{t(j,p)})) \cdot x_j \tag{14}$$

gdzie minimalizacja w sensie szukania zbiorów indeksów $I_1, ..., I_L$, wzór (11), została zastąpiona minimalizacją ze względu na wektor x i wektor γ. Znaleziony zero-jedynkowy wektor x^* minimalizujący zadanie (14) określa koniunkcję o najmniejszej długości ważonej, wzór (8), a wektor γ^* określa, które przykłady mogą być opisywane przez koniunkcję, tzn. koniunkcja nie opisuje co najmniej $(\frac{100}{N} \sum_{n=1}^{N} \gamma_n)$% przykładów negatywnych, wzór (13).

Dlatego też zadanie szukania reguły o minimalnej długości ważonej, wzór (11), można teraz zapisać w postaci ciągu zadań cząstkowych

$$\min_{x,\gamma:\, Z^1 x \geq \gamma_1} d(C^{I_1}), ..., \min_{x,\gamma:\, Z^L x \geq \gamma_L} d(C^{I_L}) \tag{15}$$

z których każde można zapisać jako pewną modyfikację zadania pokrycia

$$\min_{x,\gamma} \sum_{j=1}^{K} c_j x_j \tag{16}$$

$$\sum_{j=1}^{K} z_{nj} x_j \geq \gamma_n, \qquad n = 1, ..., N \tag{17}$$

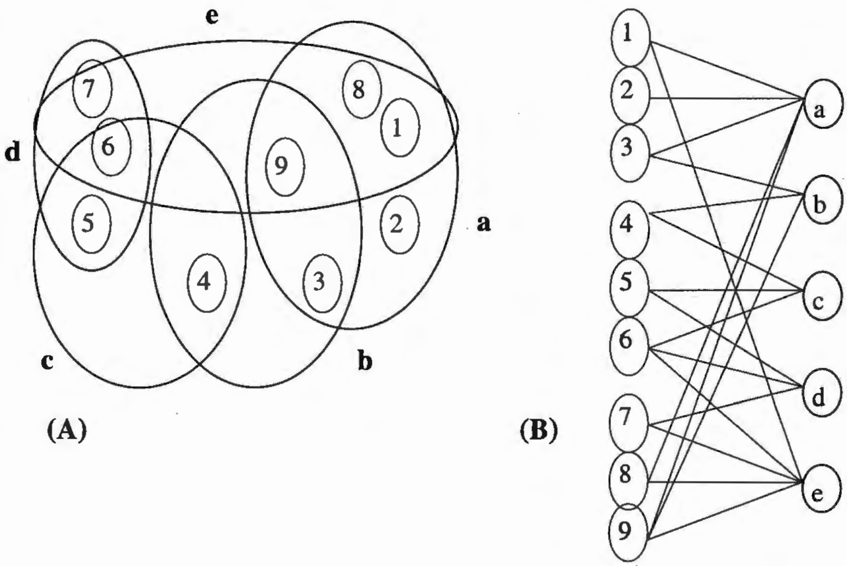
$$\sum_{n=1}^{N} \gamma_n \geq N - rel \qquad (18)$$

gdzie $c_j = 1 - g(v_{t(j,p)})$, $z_{nj} \in \{0,1\}$, $x_j \in \{0,1\}$, $j = 1,\dots,K$, $\gamma_n \in \{0,1\}$, $rel \geq 0$. Parametr rel narzuca warunek, aby tworzona koniunkcja nie opisywała co najmniej $(\frac{100}{N}(N - rel))$% przykładów negatywnych dla rozpatrywanej klasy. Przyjęcie $rel = 0$ gwarantuje, że utworzona koniunkcja nie będzie opisywała wszystkich przykładów negatywnych (warunek spójności), przyjęcie $rel > 0$ dopuszcza, że nie wszystkie przykłady muszą być nie opisane (warunek częściowej spójności).

Zadanie pokrycia jest dobrze znanym *NP*-trudnym zadaniem kombinatorycznym (Garey, Johnson 1979). Niech $I = \{1,\dots,N\}$ będzie danym zbiorem i niech P będzie rodziną „akceptowalnych" podzbiorów I. Z każdym elementem P_j rodziny P, $j = 1,\dots,K$ jest związany koszt c_j. *Zagadnienie pokrycia* zbioru I polega na wyborze takich elementów rodziny P, $P^* \subseteq P$, dla których łączny koszt jest minimalny i spełniony jest warunek, aby każdy $n \in I$ był zawarty w co najmniej jednym elemencie rodziny P. Na rysunku 1 przedstawiono prosty przypadek zadania pokrycia (Grossman, Wool 1996), w którym elementy zbioru oznaczono cyframi a podzbiory oznaczono literami.

Zadanie określone przez (16), (17), (18) tym różni się od klasycznego zadania pokrycia, że określana jest w nim dodatkowo prawa strona ograniczeń (17), co odpowiada spełnieniu wymagania, aby „większość" elementów $n \in I$ była zawarta w co najmniej jednym elemencie rodziny P (a nie wszystkie elementy). Zadanie polega na pokryciu co najmniej $N - rel$ wierszy macierzy Z o wymiarach $N \times K$ podzbiorem kolumn o minimalnym koszcie. Każdej kolumnie $1 \leq j \leq K$ jest przyporządkowany koszt $c_j = 1 - g(v_{t(j,p)})$. Określamy $x_j = 1$ jeśli j-ta kolumna (z kosztem $c_j > 0$) należy do rozwiązania oraz $x_j = 0$ gdy nie należy. Spełnienie ograniczenia (18) gwarantuje, że „większość" wierszy (co najmniej $N - rel$ wierszy) będzie opisana przez co najmniej jedną kolumnę. Ze względu na sposób konstrukcji macierzy Z i wymaganej rozłączności zbiorów przykładów pozytywnych i negatywnych zadanie powyższe ma zawsze rozwiązanie dopuszczalne, jest nim wektor jednostkowy o K elementach.

W zadaniu (16), (17), (18) szukamy zero-jedynkowego wektora x^* o minimalnym koszcie oraz zero-jedynkowego wektora γ^* który określa opisane wiersze macierzy Z; $\gamma_n^* = 1$ jeżeli n-ty wiersz jest opisany przez rozwiązanie x^* i $\gamma_n^* = 0$ jeżeli n-ty wiersz może nie być opisany. Z założenia, co najmniej $N\text{-}rel$ wierszy (których liczba jest zadana przez parametr $rel \geq 0$) musi być opisanych przez rozwiązanie x^*.



Rys. 1. Zadanie pokrycia (A) i odpowiadający mu graf incydencji (B).

W literaturze zamieszczono wiele optymalnych i heurystycznych algorytmów dla zadań pokrycia. Pierwszymi praktycznymi algorytmami były heurystyki typu zachłannego (Johnson 1974, Lovasz 1975, Chvatal 1979). Obiecujące wyniki uzyskano stosując metody oparte na algorytmach genetycznych (Beasley 1996), jak również oparte na sieciach neuronowych (Croall, Mason 1991). Ciekawe porównanie dziewięciu przybliżonych algorytmów zawarto w pracy (Grossman, Wool 1995).

Przy tworzeniu reguł o minimalnej długości ważonej kolejno dla każdej klasy $U_{v_{t(d)}}$, $\forall v_{t(d)} \in \{v_{d_1}, ..., v_{d_d}\}$, można zastosować zmodyfikowaną metodę *IP* (Szkatuła 1995), tworzącą reguły poprawnie opisujące „większość" przykładów uczących. W skrócie, kolejne kroki metody *IP* przedstawiono poniżej.

Krok 1. Interacja $s = 0$. Dany jest $NEG_A(U_{v_{t(d)}})$, $S := POS_A(U_{v_{t(d)}})$, $R^* = \varnothing$, $rel \geq 0$. Wyznacza się współczynniki ważności (Szkatuła 1995).

Krok 2. $s = s + 1$. Można modyfikować współczynniki ważności analizując zbiór przykładów uczących (Kacprzyk, Szkatuła 1998, 1999).

Krok 3. Znajduje się najbardziej „typowy" przykład $e^p \in S$ dla danej klasy (Szkatuła 1995).

Krok 4. Dla przykładu e^p i zbioru przykładów negatywnych e^n, $n = 1,...,N$ tworzy się macierz $Z_{N*K}=[z_{nj}]$, $n = 1,...,N$, $j = 1,...,K$, wzór (12). Określa się s-te zadanie pokrycia (16), (17), (18).

Krok 5. Do rozwiązania s-tego zadania pokrycia stosuje się np. algorytm oparty na algorytmie zaproponowanym przez Chvatala (1979), który buduje pokrycie w sposób sekwencyjny (Szkatuła 1995), można również zastosować algorytm genetyczny (Kacprzyk, Szkatuła 2002).

Krok 6. Utworzoną w Kroku 5 koniunkcję $C^{I_s^*}$ dołącza się do części warunkowej reguły, $R^* = R^* \cup C^{I_s^*}$. Z S usuwa się przykłady już opisane.

Krok 7. Jeżeli dokładność klasyfikacji przykładów z zastosowaniem utworzonej reguły nie jest jeszcze wystarczająca, powtarza się Krok 2.

4. Podsumowanie

W pracy, zadanie uczenia maszynowego na podstawie przykładów sformułowano w postaci ciągu zadań cząstkowych, zapisanych w postaci pewnych modyfikacji zadania pokrycia. Zaproponowano metodę tworzącą reguły typu *JEŻELI* ... *TO* ..., prezentujące istniejące regularności w przykładach, z możliwością uwzględniania założenia o poprawnym opisywaniu przez tworzone reguły „większości" przykładów uczących. Przeprowadzone eksperymenty pokazały, że przedstawiona metoda z powodzeniem może być zastosowana w zagadnieniach praktycznych, np. dla zadań z dziedziny medycyny (Kacprzyk, Szkatuła 1998, 2002), przy modelowaniu preferencji wyborców w wyborach do Sejmu (Szkatuła, Hołubiec, Wagner 2000).

Literatura

Beasley J.E. (1996) A genetic algorithm for the set covering problem. *European Journal of Operational Research*, **94**, 392-404.

Chvatal V. (1979) A greedy heuristic for the set-covering problem. *Math. of Oper. Res*, **4** (3) 233-235.

Croall I.F., Mason J.P. (eds.) (1991), *Industrial Applications of Neural Networks.* Springer-Verlag, Berlin.

Garey M.R., Johnson D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H.Freeman and Co., San Francisco.

Garfinkel R. S., Nemhauser G.L. (1978) *Integer programming.* John Wiley & Sons, New York-London-Sydney-Toronto.

Grossman T., Wool A. (1996) *Computational experience with approximation algorithms for the set covering problem.* Working paper, Theoretical Division and CNLS, Los Alamos National Laboratory.

Hołubiec J., Szkatuła G., Wagner D. (w druku) Modelowanie preferencji wyborców w postaci reguł decyzyjnych. *Modelowanie Preferencji a Ryzyko '01.* Akademia Ekonomiczna w Katowicach.

Johnson D.A. (1974) Approximation algorithms for combinatorial problems. *J. Computer System Sci.* **9**, 256-278.

Kacprzyk J., Szkatuła G. (1998) IP1 - *An Improved Inductive Learning Procedure with a Preprocessing of Data.* Proceedings of IDEAL'98 (Hong Kong), Springer-Verlag.

Kacprzyk J., Szkatuła G. (1999) An inductive learning algorithm with a preanalysis od data. *International Journal of Knowledge-Based Intelligent Engineering Systems.* **3**, 135-146.

Kacprzyk J., Szkatuła G. (2002) An integer programming approach to inductive learning using genetic and greedy algorithms. W: L.C. Jain and J.Kacprzyk (eds.) *New learning paradigms in soft computing.* Studies in Fuzziness and Soft Computing. Springer-Verlag, 323-367.

Lovasz L. (1975) On ratio of optimal integral and fractional covers. *Disc. Math.* **13**, 383-390.

Szkatuła G. (1995) *Uczenie maszynowe na podstawie przykładów w przypadku błędów w danych.* Rozprawa doktorska, IBS PAN Warszawa.

Szkatuła, G, Hołubiec J., Wagner D. (2000) Forecasting voting behaviour using machine learning-Poland in transition. *Annals of Operation Research.* **97**, 31-41

ISBN 83-85847-78-2