

KIWIEL



POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

WSPOMAGANIE DECYZJI

SYSTEMY EKSPERCKIE

pod redakcją

Romana Kulikowskiego i Lucyny Bogdan

Warszawa 1995

WSPOMAGANIE DECYZJI

SYSTEMY EKSPERCKIE

pod redakcją

Romana Kulikowskiego i Lucyny Bogdan

Warszawa 1995

Wydano z wykorzystaniem dotacji
KOMITETU BADAŃ NAUKOWYCH

Materiały konferencji: "Analiza Decyzyjna, Systemy Ekspertyczne, Zastosowania Systemów Komputerowych",
Warszawa, 25-27 maja 1994r.

Komitet Programowy Konferencji:

Andrzej Ameljańczyk, Zdzisław Bubnicki, Wiesław Grudzewski, Olgierd Hryniewicz, Janusz Kacprzyk, Lech Kruś, Roman Kulikowski (przewodniczący), Kazimierz Mańczak, Ireneusz Nykowski, Zdzisław Pawlak, Roman Słowiński, Andrzej Straszak, Andrzej Weryński, Andrzej Wierzbicki.

Wykonano z oryginałów tekstowych dostarczonych przez autorów

© Instytut Badań Systemowych PAN, Warszawa 1995

ISBN 83-85847-85-5

UCZENIE INDUKCYJNE Z UŻYCIEM DODATKOWEJ INFORMACJI UZYSKANEJ OD EKSPERTÓW DZIEDZINOWYCH

Grażyna Szkatuła

Instytut Badań Systemowych Polskiej Akademii Nauk

ul. Newelska 6, 01-447 Warszawa

E-mail: szkatulg@ibspan.waw.pl

Problem pozyskiwania wiedzy i umiejętność jej praktycznej implementacji jest wstępnym i jednym z najważniejszych kroków przy tworzeniu systemów opartych na wiedzy.

Znane i stosowane od dawna tradycyjne metody pozyskiwania wiedzy w bezpośrednim kontakcie z ekspertem, np. wywiad z ekspertem, sortowanie kart, techniki analizy protokołów i inne, wiązały się ze znacznymi kosztami wynikającymi z zatrudniania specjalistów oraz dużą czasochłonnością samego procesu.

W latach 80-tych metody tradycyjne zostały uzupełnione metodami automatycznego pozyskiwania wiedzy, stosującymi algorytmy uczenia maszynowego i stały się one bardzo pomocne w sytuacjach braku możliwości pracy z ekspertem. Metody te potrafiły praktycznie wykorzystywać w procesie nauki istniejące duże bazy danych (metody uczenia z przykładów) lub też przetwarzać wiedzę już zgromadzoną w pamięci systemów komputerowych.

W literaturze można zauważyć wyraźny podział metod uczenia maszynowego z przykładów na kilka grup, a mianowicie, na metody stosujące sztuczne sieci neuronowe, metody oparte na algorytmach genetycznych, metody empiryczne oraz metody hybrydowe łączące kilka podejść. Wśród metod empirycznych wyróżnia się metody tworzące reguły, tworzące drzewa decyzyjne, metody grupowania.

Jak do tej pory, większość podejść w uczeniu maszynowym z przykładów jest istotnie indukcyjna, podkreślają one rolę podobieństw w przykładach i dlatego często nazywane są uczeniem na podstawie podobieństw, uczeniem indukcyjnym. Tym niemniej, uczenie może być również procesem dedukcyjnym, skierowanym na ujawnienie użytecznej wiedzy o problemie nie widocznej przed uczeniem

Przykłady mogą być reprezentowane np. w postaci listy binarnych cech, z których każda odpowiada istnieniu lub nie pewnej cechy. Np. dla danego pacjenta

możemy określać istnienie żółtaczk lub kaszlu. Na takiej reprezentacji danych wejściowych bazują przeważnie metody oparte na sztucznych sieciach neuronowych i algorytmach genetycznych.

Przykłady mogą też być opisywane za pomocą takiej samej liczby cech o wykluczających się wartościach. Np. cecha włosy może przyjmować wartość czarna lub ruda. Taki sposób reprezentacji jest stosowany przeważnie w metodach empirycznych przy konstrukcji drzew decyzyjnych lub reguł.

Relacje pomiędzy przykładami mogą być opisywane za pomocą relacyjnych lub strukturalnych reprezentacji. Np. można opisywać, że obiekt A znajduje się ponad obiektem B. Taka strukturalna informacja może zostać przedstawiona w wielu formalizmach, włączając w to logikę predykatów i sieci semantyczne a w niektórych przypadkach może zostać przetransformowana na niestructuralną poprzez dodanie nowych cech.

Uczenie maszynowe na podstawie przykładów, zwane uczeniem indukcyjnym, będziemy traktowali jako proces tworzenia reguł klasyfikujących do pewnej klasy na podstawie danych a priori przykładów do tej klasy należących, zwanych przykładami pozytywnymi. Ponadto, w procesie tworzenia reguł mogą również być używane przykłady spoza klasy, zwane przykładami negatywnymi, głównie dla zmniejszenia przestrzeni możliwych rozwiązań i podniesienia efektywności stosowanych algorytmów. Utworzone reguły pozwalają decydować o przynależności lub nie danego nowego przykładu do rozpatrywanej klasy.

Procedury tworzenia takich reguł klasyfikacji powinny spełniać pewne dodatkowe wymagania, z których najbardziej istotne w naszych rozważaniach są:

- * reguła poprawnie opisuje wszystkie przykłady pozytywne z danego a priori zbioru przykładów, tzw. warunek *zupełności*, (ang. completeness),
- * reguła nie opisuje żadnego przykładu negatywnego z danego a priori zbioru przykładów, tzw. warunek *zwartości*, (ang. consistency),
- * reguła musi być określana w skończonej liczbie kroków,
- * szukana jest reguła o minimalnej długości, rozumianej np. jako minimalna liczba użytych do jej określenia cech lub warunków.

W rzeczywistych zagadnieniach, takie wymagania, ze względu na błędy występujące w zbiorze danych, należy nieco osłabić i mówimy wtedy o częściowej zupełności i częściowej zwartości. Również cechy używane do opisu przykładów nie są zazwyczaj równoważne, w sensie np.

- * łatwości dokonywania ich pomiarów,
- * odporności na błędne określanie ich wartości,
- * istniejącej preferencji pewnych cech ze względu na specyfikę zadania, itp.

Wydaje się korzystne uwzględnianie w procesie uczenia takiej dodatkowej informacji dotyczącej pewnej preferencji cech, możliwej do uzyskania od ekspertów.

Proponowana metoda jest modyfikacją metody typu 'star' zaproponowanej przez Michalskiego i współpracowników (1973, 1983) i tworzy reguły klasyfikacji w postaci dysjunkcji warunków. W szczególności proponuje się wykorzystanie w procesie tworzenia reguł dodatkowej informacji uzyskanej od eksperta z dziedziny, np. medycyny, poprzez określenie wag związanych z cechami, i uwzględnianie ich w procesie tworzenia reguły klasyfikacji.

Mogą być uwzględniane uzyskiwane od ekspertów dodatkowe informacje np. na temat:

- * preferencji cech ze względu na specyfikę zadania, np. informacja o cechach, które są uważane za decydujące z medycznego punktu widzenia i powinny być w pierwszym rzędzie brane pod uwagę,
- * większej istotności pewnych cech w procesie diagnozy danego przypadku, co wynika z doświadczeń lekarzy i pewnej tradycji w uwzględnianiu niektórych czynników.
- * informacji o odporności cech na błędne określanie ich wartości, co związane jest dokładnością aparatury, metod stosowanych do ich określenia,
- * wiarogodności określonych cech, związanej z interpretacją, zwłaszcza danych jakościowych, przez różnych specjalistów,
- * łatwości dokonywania pomiarów wartości poszczególnych cech.

Zauważmy, że informacje te są zgodne w takim sensie, że większe/mniejsze wartości dla każdego z nich są korzystne/niekorzystne z naszego punktu widzenia.

Taka dodatkowa wiedza o problemie pozwala nam określać cechy, które powinny być preferowane w procesie tworzenia reguły poprzez przypisanie im pewnych wag W . Do określania tych wag w przypadku istnienia kilku kryteriów, proponuje się zastosowanie podejścia AHP (ang. analytical hierarchy process) [Saaty, 1980], które wymaga współpracy z ekspertami. Podejście to jest dobrze znane i zostało sprawdzone w wielu zastosowaniach praktycznych. Eksperti dokonują porównania wszystkich cech parami, kolejno z punktu widzenia każdego z uwzględnianych kryteriów. Opinie te są następnie łączone i ostatecznie otrzymujemy wagi W .

Wprowadzamy pojęcie *uogólnionej długości reguły klasyfikacji* i ją minimalizujemy, co jest zgodne z naszymi oczekiwaniami znalezienia reguły krótkiej, zawierającej małą liczbę mniej preferowanych cech.

W zaproponowanej metodzie, dzięki wprowadzeniu współczynników ważności dla cech wybranych do opisu przykładów, istnieje możliwość uwzględniania zadawanych warunków w procesie tworzenia reguły klasyfikacji. Na przykład, wygodniej jest w praktyce stosować regułę w której występuje łatwa do określenia temperatura, niż wynik skomplikowanego badania krwi.

Postawienie zadania

Rozpatrywane zadanie uczenia maszynowego na podstawie przykładów można postawić w następujący sposób. Dany jest zbiór przykładów zwany przykładami

pozytywnymi ozn. S_p , $p = 1, \dots, P$ oraz zbiór przykładów zwany przykładami negatywnymi ozn. S_n , $n = 1, \dots, N$ oraz $S_p \cap S_n = \emptyset$. Zakładamy, że liczba i rodzaj cech wybranych do opisu przykładów w sposób dostatecznie dokładny opisują charakter rozpatrywanego zjawiska. Każdy przykład zostaje więc opisany za pomocą skończonej liczby tych samych K cech ozn. a_1, \dots, a_K i zgodnie z VL formalizmem [Michalski 1983], może zostać zapisany jako

$$e = \bigwedge_{i=1}^K [a_i = v_i] \quad (1)$$

gdzie a_i oznacza i -tą cechę przyjmującą wartość $v_i \in d_i$, d_i jest skończonym zbiorem wartości dla cechy a_i .

Ogólnie, warunek $s_i = [a_i \# v_i]$, gdzie $\#$ jest relacją typu $=, <, >, \approx, \geq$, itp. będziemy nazywali *selektorem*, a koniunkcję selektorów zapisaną w postaci

$$\bigwedge_{i \in I \subseteq \{1, \dots, K\}} s_i = C_I \quad (2)$$

będziemy nazywali *kompleksem*.

Mówimy, że kompleks opisuje (ang. *cover*) dany przykład, jeśli wszystkie warunki w kompleksie są spełniane przez wartości jakie przyjmują odpowiadające im cechy w przykładzie.

Na przykład kompleks postaci $[a_1 = \text{"kobieta"}] \wedge [a_3 = \text{"35 lat"}]$ opisuje przykład $[a_1 = \text{"kobieta"}] \wedge [a_2 = \text{"wysoka"}] \wedge [a_3 = \text{"35 lat"}]$, nie opisuje przykładu $[a_1 = \text{"kobieta"}] \wedge [a_2 = \text{"niska"}] \wedge [a_3 = \text{"20 lat"}]$.

Rozważmy regułę klasyfikacji w postaci dysjunkcji kompleksów

$$(s_{i_{11}} \wedge \dots \wedge s_{i_{1k}}) \cup \dots \cup (s_{i_{m1}} \wedge \dots \wedge s_{i_{mk}}) \rightarrow [\text{przykład pozytywny}] \quad (5)$$

gdzie zbiory indeksów $\{i_{11}, \dots, i_{1k}\}, \{i_{m1}, \dots, i_{mk}\} \subseteq I = \{1, \dots, K\}$, a " \cup " odpowiada łącznikowi "lub".

Przyjmijmy, że dla ustalonego pozytywnego przykładu e^p , zapisanego w postaci

$$e^p = [a_1 = v_1^p] \wedge \dots \wedge [a_K = v_K^p] \quad (6)$$

rozważamy kompleks C^p

$$C^p = [a_{i_1} = v_{i_1}^p] \wedge \dots \wedge [a_{i_l} = v_{i_l}^p] \quad (7)$$

ze zbiorem indeksów $I = \{i_1, \dots, i_l\}$, $l \leq K$.

Kompleks ten jest równoważny z wektorem zerojedynkowym $x = [x_j]^T$ takim, że $x_j = 1$ jeśli selektor $[a_j = v_j^p]$ występuje w kompleksie (7) oraz $x_j = 0$ gdy nie występuje.

Długość kompleksu (7) definiujemy w postaci

$$d(C^p) = \frac{1}{K} \sum_{j=1}^K x_j \quad (8)$$

Jeśli dla każdego selektora występującego w kompleksie (7), tzn. dla którego $x_j = 1$, przypiszemy pewną wagę $w_j^p \in [0, 1]$, to możemy określić kompleks ważony ozn. C_W^p , i możemy zdefiniować długość ważoną kompleksu (ang. *weighted length*) w postaci

$$d'_W(C_W^p) = \frac{1}{K} \sum_{j=1}^K w_j^p x_j \quad (9)$$

Wygodnie jest wprowadzić uogólnioną długość ważoną (ang. *augmented weighted length*) kompleksu C_W^p w postaci

$$d_W(C_W^p) = \frac{1}{K-1} \left(\sum_{j=1}^K x_j - \sum_{j=1}^K w_j^p x_j \right) = \frac{1}{K-1} \sum_{j=1}^K (1 - w_j^p) x_j \quad (10)$$

Taki sposób określenia uogólnionej długości ważonej jest motywowany przez nasze oczekiwanie znalezienia reguły krótkiej z małą liczbą tworzących ją selektorów, zawierającej dużą liczbę selektorów o wyższej wadze, bardziej istotnych.

Upraszczając, można przyjmować taką samą wagę w_j dla wszystkich wartości przyjmowanych przez daną cechę a_j .

Przykład 1.

Załóżmy, że liczba cech wynosi 5 oraz wagi dla każdej z nich są równe odpowiednio $w_1 = 1/5$, $w_2 = 1/10$, $w_3 = 3/5$, $w_4 = 1/20$, $w_5 = 1/20$. Wtedy uogólnione ważne długości czterech kompleksów złożonych z jednego, trzech lub pięciu selektorów wynoszą odpowiednio

$$\begin{aligned} d_W(s_1) &= 1/4 \cdot 4/5 = 1/5 \\ d_W(s_3) &= 1/10 \\ d_W(s_2 \wedge s_3 \wedge s_5) &= 9/16 \\ d_W(s_1 \wedge s_2 \wedge s_3 \wedge s_4 \wedge s_5) &= 1 \end{aligned}$$

Długość reguły klasyfikacji R złożonej z L kompleksów, $R = (C^1 \cup \dots \cup C^L)$, definiujemy jako

$$d(C^1 \cup \dots \cup C^L) = \max_{i \in \{1, \dots, L\}} d(C^i) \quad (11)$$

oraz uogólnioną długość reguły klasyfikacji (ang. *augmented weighted classification rule*) $R_w = (C_w^1 \cup \dots \cup C_w^L)$ definiujemy jako

$$d_{R_w}(C_w^1 \cup \dots \cup C_w^L) = \max_{i=1, \dots, L} d_w(C_w^i) \quad (12)$$

Dla reguły utworzonej z kompleksu złożonego z pięciu selektorów z przykładu 1 mamy oczywiście $d_{R_w} = 1$.

Zauważmy, że definicje długości reguły (11) i (12) są wyraźnie pesymistyczne, ale z powodu tego, że nasza wiedza jest obciążona błędami, wydają się one odpowiednie w rozważanym przypadku.

Zadanie uczenia maszynowego na podstawie przykładów uwzględniając określoną preferencję cech możemy teraz sformułować, jako tworzenie optymalnej reguły klasyfikacji $R_w^* = (C_w^1 \cup \dots \cup C_w^L)^*$ takiej, że

$$R_w^* \rightarrow \min_{R_w} d_{R_w}(C_w^1 \cup \dots \cup C_w^L) = \min_{R_w} \max_{i \in \{1, \dots, L\}} d_w(C_w^i) \quad (13)$$

tzn. takiej, która minimalizuje uogólnioną długość reguły klasyfikacji, przy dodatkowych wymaganiach dotyczących zupełności i zwartości.

Warunek zwartości tzn. nie opisywania przykładów negatywnych, zostaje uwzględniony poprzez uwzględnienie warunku $Zx \geq \Lambda$, gdzie Z jest macierzą zerowej jedynką o wierszach odpowiadających kolejnym przykładom negatywnym i o kolumnach odpowiadających kolejnym cechom użytym do opisu przykładów. Macierz ta zawiera jedynkę w przypadku przyjmowania przez daną cechę różnych wartości w przykładzie startowym i danym przykładzie negatywnym. Wektor Λ jest wektorem jednostkowym o wymiarze N . X jest szukanym K wymiarowym wektorem zerowej jedynką.

Taka minimalizacja przy spełnionym warunku $Zx \geq \Lambda$ gwarantuje, że znalezione kompleksy nie będą opisywały przykładów negatywnych. Wystąpienie zera w wektorze Zx na i -tej pozycji oznaczałoby, że i -ty przykład negatywny jest opisywany przez kompleks określony znalezionym wektorem zerowej jedynką x .

Zadanie takie jest znanym zadaniem pokrycia w programowaniu całkowitoliczbowym [Garfinkel, Nemhauser 1978]. Rozwiązanie dopuszczalne zawsze istnieje, ponieważ z warunku rozłączności zbiorów przykładów pozytywnych i negatywnych wynika, że nie występują w macierzy Z wiersze zerowe.

Do jego rozwiązania można zastosować algorytm heurystyczny typu greedy, prosty w implementacji i efektywny w zastosowaniach praktycznych.

Opis algorytmu

Znając wagi W związane z cechami, możemy przedstawić heurystyczny algorytm, rozwiązujący problem (13).

Krok 1: Przez S oznaczamy początkowo zbiór wszystkich przykładów pozytywnych S_p . Zakładamy, że szukana reguła R_w^* jest pusta.

Krok 2: Określamy wagi W .

Krok 3: Określamy typoid, tzn. nieistniejący przykład zawierający selektory z najczęściej występującymi wartościami cech w rzeczywistym zbiorze przykładów pozytywnych. Pojęcie typoidu określającego punkt startowy każdej iteracji jest bardzo ważne dla efektywności przedstawionego algorytmu.

Krok 4: Określamy punkt startowy tak, aby wybrany rzeczywisty przykład był bliski typoidowi, w sensie pewnej miary podobieństwa.

Krok 5: Znajdujemy kompleks C_w^* o minimalnej uogólnionej długości ważonej, startując z przykładu określonego w kroku 4, używając do tego celu algorytmu opartego na elementach programowania całkowitoliczbowego.

Krok 6: Dołączamy kompleks C_w^* znaleziony w kroku 5 do szukanej reguły klasyfikacji R_w^* , $R_w^* := R_w^* \cup C_w^*$ i, ze zbioru przykładów pozytywnych S usuwamy przykłady opisywane przez znaleziony kompleks C_w^* .

Krok 7: Jeśli zbiór przykładów pozytywnych S jest pusty, kończymy obliczenia i reguła R_w^* została określona, w przeciwnym przypadku powtarzamy krok 3.

Przykład medyczny

Zaproponowaną metodę przetestowano na rzeczywistych danych onkologicznych dotyczących mięsaków tkanek miękkich, które zostały skompletowane w 1986 r. przez doc. Włodzimierza Rukę z Kliniki Chirurgicznej Centrum Onkologii im. Marii Skłodowskiej-Curie w Warszawie. Wybrano 22 cechy (czynniki prognostyczne) które potencjalnie mogłyby mieć największy związek z rokowaniem, takie jak np.: płeć, wiek, stosowanie radioterapii i chemioterapii przed przyjęciem do instytutu, względny przyrost guza w czasie, powikłania węzów chłonnych, typ, umiejscowienie i maksymalny rozmiar guza, występowanie nacieczenia, wieloogniskowość, ocena

zaawansowania, rodzaj wykonanego zabiegu, margines chirurgiczny podczas zabiegu i ewentualne uszkodzenie guza, ocena gojenia rany, wystąpienie gorączki, typ mięsaka oraz stopień jego złośliwości i stopień zaawansowania klinicznego określany po zabiegu.

Obliczenia wykonano dla zbioru zawierającego pacjentów, którzy przeżyli ponad dwa lata po zabiegu chirurgicznym (zbiór przykładów pozytywnych), oraz dla zbioru zawierającego pacjentów, którzy nie dożyli dwóch lat (zbiór przykładów negatywnych).

Uzyskane reguły zostały ocenione pozytywnie przez onkologa w sensie zgodności z jego doświadczeniem praktycznym,

Został również wykonany dla danego zestawu danych medycznych eksperyment obliczeniowy, pokazujący w jakim stopniu zaburzenie danych wpływa na postać reguł określanych przez algorytm dla tego zestawu danych.

Założono w nim, że zbiór rzeczywistych danych medycznych którym dysponujemy jest pozbawiony błędów i przyjęto go za zbiór testowy do obliczeń. Następnie został on poddawany zaburzeniu i dla takich zaburzonych danych wyznaczano postać reguł klasyfikujących. Poprawność klasyfikacji tak wyznaczonych reguł sprawdzano dla wszystkich przykładów ze zbioru testowego.

Postępowanie takie powtarzano po kilkaset razy dla różnych stopni zaburzenia. Przyjęto *oszacowanie błędu klasyfikacji* w postaci sumy średniego błędu pominięcia i średniego błędu dołączenia:

$$E_{pd} = E_p + E_d.$$

Średni błąd pominięcia $E_p = (E_p^1 + E_p^2)/2$, oznacza błąd tego, że przykład należący do pewnej klasy nie został do tej klasy zaklasyfikowany. E_p^1 jest średnią liczbą przykładów pozytywnych ze zbioru testowego nie zaklasyfikowanych jako pozytywne, podzieloną przez liczbę testowych przykładów pozytywnych. E_p^2 jest średnią liczbą przykładów negatywnych ze zbioru testowego nie zaklasyfikowanych jako negatywne, podzieloną przez liczbę testowych przykładów negatywnych.

Średni błąd dołączenia $E_d = (E_d^1 + E_d^2)/2$ oznacza błąd tego, że przykład nie należący do pewnej klasy został do tej klasy zaklasyfikowany. E_d^1 jest średnią liczbą przykładów negatywnych ze zbioru testowego zaklasyfikowanych jako pozytywne, podzieloną przez liczbę testowych przykładów negatywnych. E_d^2 jest średnią liczbą przykładów pozytywnych ze zbioru testowego błędnie zaklasyfikowanych jako negatywne, podzieloną przez liczbę testowych przykładów pozytywnych

Przykładowo, zaburzane były m. in. wartości przyjmowane przez wybraną cechę, w zestawie losowo wybieranych 10%, 30% przykładów pozytywnych i odpowiednio 10%, 30% przykładów negatywnych ze zbioru testowego.

Przy zaburzaniu 10% wartości kolejno dla każdej cechy w zbiorze danych, oszacowanie błędu klasyfikacji dla uzyskanych reguł nie przekroczyło 6.1 %. Przy zaburzaniu 30% wartości, oszacowanie było mniejsze niż 15.7 %.

Zakończenie

W pracy przedstawiono metodę uczenia maszynowego z przykładów, która dzięki wprowadzeniu współczynników ważności dla każdej cechy wybranej do opisu przykładów, uwzględniła dodatkowe informacje z dziedziny możliwe do uzyskania od eksperta.

LITERATURA

- Garfinkel R. S., Nemhauser G. L.: "Programowanie całkowitoliczbowe". PWN 1978.
- Iwański C., Szkatuła G.: "Inductive learning of the minimum length classification rules without risking combinatorial explosion". *Control and Cybernetics*, vol. 18, no. 3-4, pp. 117-126, 1989.
- Iwański C., Szkatuła G.: "Inductive Learning Supported by Integer Programming". *Computers and Artificial Intelligence*, vol. 10, no. 1, pp. 57-65, 1991.
- Michalski R.: "A Theory and Methodology of Inductive Learning". In: R. Michalski, J. Carbonell, T.M. Mitchell (eds.): *Machine Learning*. Tioga, Palo Alto 1983.
- Michalski R., Mozetic J., Hong J., Lavrac.: "The Multi-Purpose Incremental Learning System AQ15 and Its Testing Application to Three Medical Domains". In: *Proceedings of Fifth National Conference on Artificial Intelligence AAAT-86*, Philadelphia 1986.
- Saaty T. L.: "The analytic hierarchy process". McGraw-Hill, New York, 1980.
- Quinlan J., R.: "Discovering rules from large collections of examples: a case study". In: Michie D. (ed.): *Expert Systems in the Micro Electronic Age*. Edinburgh University Press, Edinburgh 1979.
- Quinlan J., R.: "Learning efficient classification procedures and their applications to chess end games". In: R. Michalski, J. Carbonell, T.M. Mitchell (eds.): *Machine Learning*. ss 463-482, Tioga, Palo Alto 1983.
- Kacprzyk J., Szkatuła G.: "Machine learning from examples under errors in data". *Proceedings of 5th International IPMU Conference, Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Paris, pp. 1047-1051, July 4-8, 1994.

ISBN 83-85847-85-5

**W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy
prosimy o kontakt
z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 36-19-01 w. 241 e-mail: kotuszew@ibspan.waw.pl**