

Piotr PASZEK*

NIEDETERMINISTYCZNE REGUŁY DECYZYJNE

Streszczenie

Klasyfikatory regułowe pełnią istotną rolę wśród systemów decyzyjnych. W artykule opisujemy wykorzystanie reguł niedeterministycznych w klasyfikatorach regułowych. Proponujemy algorytm klasyfikacji wykorzystujący reguły niedeterministyczne oraz reguły minimalne w sensie zbiorów przybliżonych.

Klasyfikator ten porównano z klasyfikatorem wykorzystującym tylko reguły deterministyczne na danych z UCI Machine Learning Repository. Przedstawione wyniki badań wskazują, że wykorzystanie reguł niedeterministycznych daje możliwość poprawy jakości klasyfikacji, ale wymaga dostrojenia do danych.

Słowa kluczowe: reguły niedeterministyczne, klasyfikatory regułowe, zbiory przybliżone

Wprowadzenie

Klasyfikatory regułowe już od wielu lat zajmują istotne miejsce w uczeniu maszynowym [8], [17]. Mogą być budowane w oparciu o różne metody np. drzewa decyzyjne [4], zbiory przybliżone [3], [7], [13], [14], [18], czy z wykorzystaniem rozkładów prawdopodobieństwa [6].

Istotną cechą, którą powinny się one charakteryzować jest ich efektywność. Efektywność klasyfikatorów oceniamy poprzez zdolność do klasyfikacji nowych przypadków, tzw. predykcji. Okazuje się, że konstruowanie klasyfikatorów regułowych jako hybrydy paru metod, w wielu przypadkach przynosi poprawę ich jakości. Oczywiście przy wyborze metod należy zawsze uwzględniać specyfikę danych, dla których konstruowany jest klasyfikator.

W artykule pokazano, że istnieje możliwość rozbudowania klasyfikatorów regułowych o reguły niedeterministyczne. Wskazówkę do takiego podejścia stanowiły prace Skowrona i Suraja [15], [16], w których wykazano, że istnieją systemy decyzyjne nie dające się

* Dr Piotr Paszek, Wyższa Szkoła Finansów i Prawa, Katedry Informatyki i Metod Ilościowych.

w pełni opisać z wykorzystaniem reguł deterministycznych. Natomiast w pracy [9] udowodniono, że dowolny system decyzyjny (zbiór jego obiektów) może być opisany z wykorzystaniem reguł niedeterministycznych. Miarą oceny przydatności reguł niedeterministycznych dla danego modelu klasyfikacji, wyrażającą jakość klasyfikacji tego modelu, będą standardowe współczynniki dokładności i pokrycia klasyfikatora.

Sam problem klasyfikacji w pracy jest rozumiany w następujący sposób: dla danego systemu decyzyjnego i nowego obiektu, na podstawie atrybutów warunkowych opisujących ten obiekt, należy określić właściwą mu klasę decyzyjną.

1. Reguły w systemach decyzyjnych

Niech $T = (U, A, \{d\})$ będzie systemem decyzyjnym (tablicą decyzyjną), gdzie $U = \{u_1, u_2, \dots, u_n\}$ to skończony, niepusty zbiór obiektów, $A = \{a_1, a_2, \dots, a_m\}$ to niepusty, skończony zbiór atrybutów warunkowych, $d \notin A$ to atrybut decyzyjny [11].

Dla danego T przez V_d oznaczamy zbiór wartości atrybutu decyzyjnego, przez V_A zbiór wartości wszystkich atrybutów warunkowych w systemie decyzyjnym a przez V_a zbiór wartości dla pojedynczego atrybutu ze zbioru A .

Deskryptor (formuła elementarna) to wyrażenie postaci $(a = v)$, gdzie $a \in A, v \in V_a$. Uogólniony deskryptor to formuła postaci $(a \in V)$, gdzie $a \in A, V \subseteq V_a$.

Zwykle klasyfikatory regułowe wykorzystują reguły deterministyczne. Regułę deterministyczną dla systemu decyzyjnego T można przedstawić jako:

$$(a_{j_1} \in V_{j_1}) \wedge \dots \wedge (a_{j_t} \in V_{j_t}) \Rightarrow (d = c) \quad (1)$$

gdzie $a_{j_i} \in A_{j_i}, V_{j_i} \subseteq V_A$ dla $i = 1, \dots, t$.

W części warunkowej reguły deterministycznej mamy koniunkcję skończonej liczby deskryptorów uogólnionych, natomiast w części decyzyjnej deskryptor oparty na atrybucie decyzyjnym.

Ponieważ wiemy ([15], [16]), że nie wszystkie systemy decyzyjne udaje się opisać z wykorzystaniem reguł deterministycznych wprowadza

się w celu poprawy tego opisu reguły niedeterministyczne. W sposób jak najbardziej ogólny, regułę niedeterministyczną [15] dla T możemy przedstawić, jako:

$$(a_{j_1} \in V_{j_1}) \wedge \dots \wedge (a_{j_t} \in V_{j_t}) \Rightarrow (d \in \{c_1, \dots, c_s\}), \quad (2)$$

gdzie $a_{j_i} \in A_{j_i}, V_{j_i} \subseteq V_A$ dla $i = 1, \dots, t$ oraz $\emptyset \neq \{c_1, \dots, c_s\} \subseteq V_d$.

Ze względów praktycznych reguły niedeterministyczne w niniejszej pracy ograniczono do postaci, w której w części warunkowej występują deskrytory a nie deskrytory uogólnione, tzn. do reguł postaci:

$$(a_{j_1} = b_1) \wedge \dots \wedge (a_{j_t} = b_t) \Rightarrow (d \in \{c_1, \dots, c_s\}), \quad (3),$$

gdzie $a_{j_i} \in A_{j_i}, b_i \in V_A$ dla $i = 1, \dots, t$ oraz $\emptyset \neq \{c_1, \dots, c_s\} \subseteq V_d$.

Zgodnie z [5], przyjmujemy następujące oznaczenia. Jeśli r jest regułą niedeterministyczną postaci (3), to:

- $L(r)$ oznacza część warunkową reguły (lewa strona reguły),
- $R(r)$ oznacza część decyzyjną reguły (prawa strona reguły),
- $card(S)$ oznacza moc (liczbę elementów) zbioru S ,
- $\|w\|$ oznacza zbiór wszystkich obiektów z tablicy decyzyjnej T pasujących do formuły w .

Celem oceny jakości reguł niedeterministycznych użyjemy następujących współczynników:

- zaufania (ang. confidence) reguły [1],
- wsparcia (ang. support) reguły,
- wsparcia w postaci znormalizowanej [5].

Dla reguły decyzyjnej r i systemu decyzyjnego T , wsparcie definiowane jest jako

$$supp_T(r) = \frac{card(\|L(r)\| \cap \|R(r)\|)}{card(U)}, \quad (4)$$

zaufanie, jako

$$conf_T(r) = \frac{card(\|L(r)\| \cap \|R(r)\|)}{card(\|L(r)\|)}, \quad (5),$$

a znormalizowane wsparcie jako

$$nsupp_T(r) = \frac{supp_T(r)}{\sqrt{card(V(r))}}, \quad (6),$$

gdzie $V(r) = \{c_1, \dots, c_s\}$ to zbiór wartości atrybutu decyzyjnego dla reguły r .

2. Skrócone reguły niedeterministyczne

W dalszej części pracy wykorzystywać będziemy reguły niedeterministyczne nazwane przez nas skróconymi regułami niedeterministycznymi. Reguły te powstają poprzez usuwanie deskryptorów z części warunkowej reguł deterministycznych, zgodnie z zasadą minimalnego opisu (ang. *minimum description length*) [12]. W dalszej części pracy zostaną one użyte do poprawy jakości klasyfikatorów regułowych.

Skrócone reguły niedeterministyczne – TNDR (ang. *Truncated Nondeterministic Decision Rules*) lub krócej reguły skrócone to zbiór reguł niedeterministycznych r (nad T) postaci

$$(a_{j_1} = b_1) \wedge \dots \wedge (a_{j_t} = b_{j_t}) \Rightarrow (d \in \{c_1, \dots, c_s\}),$$

takich, że:

1. $conf(r) \geq \alpha$, gdzie $\alpha \in [0.5, 1]$, tzn. zaufanie reguł jest co najmniej równe α ,
2. $|V(r)| \leq k < |V_d(T)|$, gdzie k (parametr), to górna granica liczby wartości decyzyj reguły r .

3. Algorytm tworzący reguły niedeterministyczne

W pracy [10] opisaliśmy algorytm zachłanny, który generuje skrócone reguły niedeterministyczne dla danego systemu decyzyjnego. Jego czasowa złożoność obliczeniowa była nie zadawalająca i mieściła się w klasie wykładniczej. Ponieważ reguły niedeterministyczne powstawały poprzez eliminację pewnej liczby deskryptorów z części warunkowej reguł deterministycznych a dokładniej reguł minimalnych tworzonych z użyciem systemu RSES [1], postanowiono ten warunek zmodyfikować, tak aby algorytm mógł być przypisany do wielomianowej klasy złożoności. Modyfikacja była możliwa dzięki zastąpieniu reguł

minimalnych regułami pełnymi tzn. całymi wierszami opisującymi obiekty z systemu decyzyjnego.

Przyjęliśmy następujące założenie: algorytm tworzy reguły niedeterministyczne o dużym wsparciu oraz małej liczbie wartości, w stosunku do liczby wszystkich wartości atrybutu decyzyjnego w części decyzyjnej reguły. Parametrem algorytmu jest współczynnik α , który oznacza minimalne zaufanie dla reguł niedeterministycznych. Główne kroki algorytmu Trunder – w postaci pseudokodu – przedstawione są poniżej w tabeli 1.

Liczba operacji w algorytmie (czasowa złożoność obliczeniowa algorytmu) jest wielomianowa. Zależy od liczby obiektów i liczby atrybutów i jest rzędu $O(kn^2)$, gdzie k to liczba atrybutów a n to liczba obiektów w tablicy decyzyjnej T .

Tabela 1. Pseudokod algorytmu wyznaczania skróconych reguł niedeterministycznych

Trunder – algorytm wyznaczania reguł niedeterministycznych

Wejście: T – system decyzyjny, $\alpha \in [0.5, 1]$ – parametr minimalnego zaufania;

Wyjście: Zbiór reguł niedeterministycznych $NDR(T, \alpha)$

1. $NDR(T, \alpha) = \emptyset$;
2. Dla każdego obiektu x z T :
3. $r = x$; //potraktuj x jak regułę decyzyjną
4. Usuń atrybut z r (deskryptor z reguły), który maksymalizuje nośnik części warunkowej;
5. Dobierz (zachłannie) klasy decyzyjne (do części warunkowej reguły), tak aby utworzona reguła miała zaufanie nie mniejsze niż α ;
6. $r' =$ utworzona w punkcie 5. reguła;
7. Jeśli $nsupp_T(r') \geq nsupp_T(r)$, to
 $r = r'$ i wróć do 4.
8. w przeciwnym wypadku
dodaj r do $NDR(T, \alpha)$.

4. Klasyfikator wykorzystujący reguły niedeterministyczne

Wygenerowane algorytmem Trunder, reguły skrócone użyjemy do budowy klasyfikatora. Klasyfikator wykorzystuje reguły deterministyczne (DR) oraz skrócone reguły niedeterministyczne (TNDR).

Sam proces klasyfikacji dla nowego obiektu v i systemu decyzyjnego T może być przeprowadzony w następujący sposób:

1. Wyznacz DR (zbiór reguł deterministycznych) dla T .
2. Wyznacz $TNDR$ (zbiór reguł niedeterministycznych) dla T .
3. Wybierz wartości atrybutu decyzyjnego dla v na podstawie zbiorów reguł wyznaczonych w punkcie 1 i 2 w następujący sposób:
 - a. Korzystając z standardowego głosowania [3] dla obiektu v i reguł DR , wybierz pojedynczą wartość decyzji $c \in V_d$.
 - b. Korzystając z głosowania na wzorce [7] dla obiektu v i reguł $TNDR$, wybierz zbiór wartości atrybutu decyzyjnego $V \subseteq V_d$.
 - c. Rozstrzygnij konflikty między wyborem decyzji dla reguł deterministycznych i reguł niedeterministycznych za pomocą strategii wykluczającej.

W strategii wykluczającej dla obiektu v , decyzji $c \in V_d(T)$ (wyznaczonej przez standardowe głosowanie na reguły deterministyczne DT) oraz zbioru decyzji $V(r) \subseteq V_d(T)$ (wyznaczonych przez głosowanie na skrócone reguły niedeterministyczne TNDR) wybór wartości atrybutu decyzyjnego dla v odbywa się w następujący sposób:

1. Jeżeli $c \in V(r)$ (brak konfliktu), wtedy jako wynik klasyfikacji wybierz decyzję c .
2. Jeżeli $c \notin V(r)$ (konflikt), wtedy jeśli nośnik dla decyzji c jest większy niż nośnik dla reguły niedeterministycznej r , wybierz decyzję c ; w przeciwnym wypadku wybierz, z reguły niedeterministycznej r , pojedynczą decyzję o maksymalnym nośniku.
3. Jeśli $V(r) = \emptyset$, wtedy jako wynik klasyfikacji wybierz decyzję c .
4. Jeśli $V(r) \neq \emptyset$, wtedy jako wynik klasyfikacji wybierz, z reguły niedeterministycznej r , pojedynczą decyzję o maksymalnym nośniku.
5. W pozostałym przypadku obiekt v nie jest sklasyfikowany.

5. Porównanie klasyfikatorów regułowych

Opisany w poprzednim punkcie klasyfikator regułowy wykorzystujący reguły niedeterministyczne i deterministyczne porównaliśmy z klasyfikatorem regułowym wykorzystującym reguły deterministyczne. Ocenialiśmy jakość klasyfikacji, rozumianą jako procent poprawnie zaklasyfikowanych obiektów.

Badanie jakości klasyfikatorów przeprowadziliśmy na danych z UCI Machine Learning Repository [2]. Do badań – eksperymentów wykorzystaliśmy następujące bazy: Balance scale, Dermatology, Ecoli, Iris, Lymphography, Post operative oraz Primary tumor.

Badania miały potwierdzić następującą hipotezę: wzbogacenie klasyfikatora dodatkową wiedzą w postaci reguł niedeterministycznych przyczynia się do poprawy jakości klasyfikacji tak skonstruowanego modelu.

Zbudowaliśmy 2 klasyfikatory, wykorzystujące te same podziały na zbiory uczące i testowe.

Pierwszy klasyfikator, oznaczony jako ADT, wykorzystywał tylko reguły deterministyczne, dokładniej reguły minimalne, w sensie zbiorów przybliżonych, tworzone z wykorzystaniem biblioteki *RSESLib* (*Rough Set Exploration System library*) [3]. Klasyfikator wykorzystuje standardową procedurę głosowania reguł.

Drugi algorytm klasyfikacji ANDT wykorzystuje skrócone reguły niedeterministyczne, generowane przez algorytm Trunder oraz reguły minimalne. Do rozstrzygnięcia konfliktów używana jest strategia wykluczająca.

Porównywaliśmy jakość klasyfikacji ADT oraz ANDT, metodą krzyżowej walidacji, dla $n=5$. Dla algorytmu ANDT używaliśmy różnych wartości parametru α . W trakcie klasyfikacji wyznaczaliśmy dokładność klasyfikacji (acc) i pokrycie ($cover$) oraz iloczyn tych wartości ($acc \times cover$), czyli procent poprawnie sklasyfikowanych obiektów – jakość klasyfikacji. Wyniki przeprowadzonych eksperymentów zostały zawarte w tabeli 2.

W wierszach oznaczonych jako $acc \times cover$ zapisana jest średnia arytmetyczna jakości klasyfikacji z 20 uruchomień 5-krotnej krzyżowej walidacji (procentowo). W wierszach oznaczonych jako mrd zapisano maksymalne odchylenie względne. Można tę wartość interpretować jako stabilność klasyfikatora. W każdym wierszu pogrubiono wynik najlepszego klasyfikatora.

W kolumnie oznaczonej jako ADT wyniki dotyczą klasyfikatora wykorzystującego reguły minimalne. W kolumnach oznaczonych jako ANDT wpisane są wyniki dotyczące klasyfikatora wykorzystującego reguły niedeterministyczne.

Dla pięciu spośród siedmiu tablic decyzyjnych, tj. Balance Scale, Dermatology, Ecoli, Post operative oraz Primary tumor, klasyfikator ANDT, wykorzystujący skrócone reguły niedeterministyczne, był lepszy (w sensie jakości klasyfikacji) niż klasyfikator DR, wykorzystujący reguły deterministyczne. Dla jednej tablicy decyzyjnej Lymphography wyniki obu klasyfikatorów były identyczne.

Dla tablicy decyzyjnej Iris zastosowanie skróconych reguł niedeterministycznych nie polepszyło jakości klasyfikatora. Klasyfikator deterministyczny był lepszy, ale jedynie o 0,3%.

Maksymalne odchylenie względne obu klasyfikatorów było bardzo zbliżone.

Tabela 2. Jakość klasyfikacji dwóch algorytmów klasyfikacji – metoda krzyżowej walidacji

Dane (T)	wspól. klasyf.	Algorytm klasyfikacji						
		ADT	ANDT, $\alpha =$					
			1.0	0.9	0.8	0.7	0.6	0.5
Balance Scale	acc×cover	78.30	80.74	81.86	81.83	80.79	79.78	76.83
	mrd	0.020	0.032	0.024	0.024	0.023	0.026	0.024
Dermatology	acc×cover	84.62	85.04	84.97	85.27	85.21	84.62	84.59
	mrd	0.012	0.008	0.014	0.011	0.014	0.012	0.009
Ecoli	acc×cover	54.76	55.45	56.01	54.52	54.40	50.63	50.63
	mrd	0.036	0.040	0.033	0.026	0.027	0.020	0.020
Lymphography	acc×cover	37.47	37.47	37.47	37.47	37.50	37.47	37.47
	mrd	0.038	0.038	0.038	0.038	0.037	0.038	0.038
Post-Operative	acc×cover	65.00	65.00	65.00	64.72	66.39	68.28	68.83
	mrd	0.061	0.061	0.061	0.058	0.064	0.072	0.033
Primary Tumor	acc×cover	59.71	60.09	60.09	60.09	60.09	60.09	60.09
	mrd	0.016	0.020	0.020	0.020	0.020	0.020	0.020
Iris	acc×cover	90.47	90.10	89.93	88.87	87.16	87.17	86.17
	mrd	0.018	0.028	0.026	0.031	0.048	0.078	0.058

Wnioski

Przeprowadzając badania porównawcze na bazach z UCI Machine Learning Repository wykazaliśmy, że dodanie do klasyfikatora regułowego, wykorzystującego reguły deterministyczne, dodatkowej wiedzy w postaci skróconych reguł niedeterministycznych (TNR) poprawia jakość klasyfikatora. Aby otrzymać poprawę jakości klasyfikacji, należy zawsze dobrać parametr α , określający pokrycie klas decyzyjnych (zaufanie reguły), do danych.

Literatura

- [1]. Agrawal R., Imielinski T., Swami A., *Mining Association Rules Between Sets of Items in Large Databases*, [w:] *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, (red.) Buneman P., Jajodia S., Washington, D.C., ACM Press, New York 1993, s. 207-216.
- [2]. Asuncion A., Newman D. J., *UCI Machine Learning Repository*. University of California, Irvine, School of Information and Computer Sciences, 2007, <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [3]. Bazan J. G., Szczuka M. S., Wojna A., Wojnarski M., *On the Evolution of Rough Set Exploration System*, *LNAI*, vol. 3066, Springer, Heidelberg 2004, s. 592-601.
- [4]. Cichosz P., *Systemy uczące się*, WNT, Warszawa 2000.
- [5]. Delimata P., Marszał-Paszek B., Moshkov M., Paszek P., Skowron A., Suraj Z., *Comparison of Some Classification Algorithms Based on Deterministic and Nondeterministic Decision Rules*, *Transactions on Rough Sets XII*, LNCS 6190, Springer, Heidelberg 2010, s. 90-105.
- [6]. Koronacki J., Ćwik J., *Statystyczne systemy uczące się*, EXIT, Warszawa 2008.
- [7]. Marszał-Paszek B., Paszek P., *Minimal Templates and Knowledge Discovery*, *LNAI*, vol. 4585, Springer, Heidelberg 2007, s. 411-416.
- [8]. Michalski R., <http://www.mli.gmu.edu/michalski/>
- [9]. Moshkov M., Skowron A., Suraj Z., *Maximal consistent extensions of information systems relative to their theories*, „Information Sciences” 178 (12)/2008, s. 2600-2620.

- [10]. Paszek P., Marszał-Paszek B., *Niedeterministyczne reguły skrócone*, [w:] *Systemy Wspomagania Decyzji*, (red.) Wakulicz-Deja A., Katowice 2010, s. 67-72, ISBN 978-83-62462-01-8.
- [11]. Pawlak Z., *Rough Sets: Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Boston 1991.
- [12]. Rissanen J., *Modeling by Shortest Data Description*, „Automatica” 14/1978, 465-471.
- [13]. Rosetta: <http://www.lcb.uu.se/tools/rosetta/>
- [14]. Rough Set Exploration System: <http://logic.mimuw.edu.pl/rses>
- [15]. Skowron A., Suraj Z., *Rough sets and concurrency*, [w:] „Bulletin of the Polish Academy of Sciences”, 41 (3)/1993, s. 237-254.
- [16]. Suraj Z., *Some Remarks on Extensions and Restrictions of Information Systems*, LNCS, vol. 2005, Springer, Heidelberg 2001, s. 204-211.
- [17]. Triantaphyllou E., Felici G., *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques*, Springer Science and Business Media, LLC, New York 2006.
- [18]. Tsumoto S., *Modelling Medical Diagnostic Rules Based on Rough Sets*. LNCS, vol. 1424, Springer-Verlag, Berlin 1998, s. 475-482.

NONDETERMINISTIC DECISION RULES

Summary

In the paper the authors discuss classifiers based on deterministic decision rules and non-deterministic decision rules and prove that non-deterministic decision rules can be used for improving the quality of classification. The authors propose classifications algorithms based on non-deterministic rules and minimal rules in the sense of rough sets. The classifiers in question are tested on the group of decision tables from the UCI Machine Learning Repository and the results are compared. The reported results of experiments show that proposed classifiers based on non-deterministic rules give the possibility to improve the classification quality but with the requirement of tuning to the data.

Key words: *non-deterministic decision rules, rule-based classifiers, rough sets*