

Agnieszka NOWAK-BRZEZIŃSKA, Tomasz JACH
Uniwersytet Śląski, Instytut Informatyki

METODA WSPÓLCZYNNIKÓW NIEPEŁNOŚCI WIEDZY W SYSTEMACH WSPOMAGANIA DECYZJI

Streszczenie. W artykule zaproponowano metodę współczynników niepełności (IF) służących do modelowania niepełności wiedzy w systemach wspomaganie decyzji. Przedstawiono miarę współczynnika niepełności wiedzy oraz wyniki przeprowadzonych eksperymentów i ich interpretację.

Słowa kluczowe: systemy wspomaganie decyzji, data mining, grupowanie, wnioskowanie, wiedza niepełna

THE METHODS OF COMPUTING THE KNOWLEDGE INCOMPLETENESS FACTOR IN DECISION SUPPORT SYSTEMS

Summary. The paper consists of the proposition of using the method of incompleteness factors (IF) in order to model the incompleteness of knowledge in decision support systems. The authors are using cluster analysis methods along with the incompleteness factors to reason in systems with incomplete knowledge.

Keywords: decision support systems, data mining, clustering, inference, incomplete knowledge

1. Wprowadzenie

Systemy wspomaganie decyzji (SWD) są nieodłączną i ważną gałęzią informatyki. Pozwalają na zautomatyzowanie pracy i przekazanie niektórych obowiązków maszynom. Poza tym są w stanie zgromadzić znaczną wiedzę i efektywnie z niej korzystać w celu wspomaganie człowieka-eksperta w procesie wnioskowania i podejmowania decyzji.

Główną cechą SWD jest zdolność do wyciągania logicznych i poprawnych wniosków na podstawie zbioru faktów oraz reguł. Formalnie system wspomaganie decyzji jest zdefiniowany następująco [6]:

$$SWD = \langle U, A, V, f \rangle$$

U – Skończony, niepusty zbiór reguł (uniwersum)

A – Skończony, niepusty zbiór atrybutów

$$C \cup D = A; C \cap D = \emptyset$$

C – zbiór atrybutów warunkowych

D – zbiór atrybutów decyzyjnych

V – zbiór wartości atrybutów

$$V = \bigcup_{a \in A} V_a$$

V_a – zbiór wartości atrybutu V

$f : U \times A \rightarrow V$ – funkcja informacji

Ponadto przez r_i oznaczmy i -tą regułę w systemie odpowiadającą przyjętej w klasycznych SWD postaci klauzuli Horna [1], gdzie każdy literał z części przesłankowej i decyzyjnej tworzony jest na podstawie zbioru atrybutów A oraz zbiorów wartości każdego atrybutu $V_a, a \in A$. Parę (a, v_a) budującą przesłanki i konkluzje reguł będziemy dalej nazywać deskryptorem $(d_i = (a_j, v_{a_j}))$, dzięki czemu regułę r_i możemy przedstawić następująco: $r_i = d_1 \wedge d_2 \wedge \dots \wedge d_m \Rightarrow DEC_i$.

Klasyczne wnioskowanie polega na uaktywnianiu reguł, których wszystkie przesłanki są spełnione (inaczej: przesłanki są faktami znanymi w systemie). Niestety, taka sytuacja jest w rzeczywistych zastosowaniach bardzo pożądana, jednakże często nieosiągalna. W przypadku gdy w systemie nie znajdziemy żadnej reguły, której wszystkie przesłanki znajdują się w zbiorze faktów, wnioskowanie nie może zostać wykonane, a system nie dostarczy użytkownikowi żadnej nowej wiedzy.

Autorzy w swych poprzednich pracach [12-17] proponują użycie mechanizmów analizy skupień w celu stworzenia grup reguł najbardziej podobnych do siebie. Dzięki temu skrócony zostanie czas wyszukiwania reguł w dużych systemach. Proponuje się również wprowadzenie wnioskowania w warunkach niekompletnej wiedzy opartego na uaktywnianiu reguł, których nie wszystkie przesłanki są spełnione.

1.1. Efektywność wnioskowania

Efektywność wnioskowania będzie przez autorów poprawiana za pomocą dwóch czynników. Pierwszy z nich to zwiększenie szybkości odnajdowania reguł relewantnych w stosunku do aktualnego zbioru faktów w systemie przez użycie mechanizmów analizy skupień. Grupowanie reguł w skupienia za pomocą algorytmów hierarchicznych pozwala na stworzenie struktury drzewiastej reguł (dendrogramu) znacznie zwiększającej szybkość wyszukiwania konkretnej reguły w stosunku do wyszukiwania liniowego w płaskiej bazie wiedzy.

Drugim aspektem badanym przez autorów jest możliwość uaktywniania tych reguł, których nie wszystkie przesłanki są w pełni spełnione. W tym celu w stworzonej już strukturze hierarchicznej reguł wyszukiwana jest grupa najbardziej podobna do aktualnie rozpatrywanego zbioru faktów. Jeśli wewnątrz tej grupy znajdują się reguły, których wszystkie przesłanki są spełnione, zostają one uaktywnione i system zachowuje się jak klasyczny system wspomaganie decyzji (bez skupień reguł).

W przeciwnym wypadku jednak, gdy w grupie najbardziej podobnej do zbioru faktów znajdują się reguły, których tylko część przesłanek jest spełniona, proponowane będzie uaktywnienie tychże reguł z określeniem ich współczynnika niepełności. Otrzymane w ten sposób konkluzje będą oznaczone jako niepewne, a ich stopień niepełności (*IF*, ang. *incompleteness factor*) będzie wyznaczany w sposób opisany przez autorów w dalszej części artykułu. Formalnie autorzy rozpatrują trzy rodzaje reguł wchodzących w skład bazy wiedzy:

- reguły pewne
 - $IF(\text{reguła})=1$
- reguły niepewne
 - $IF(\text{reguła}) \in [\text{wsp. jakości}; 1]$
- reguły nieuaktywniane
 - $IF(\text{reguła}) \in [0; \text{wsp. jakości})$

Każdorazowo określony zostanie współczynnik jakości, który będzie oddzielał reguły niebrane pod uwagę we wnioskowaniu.

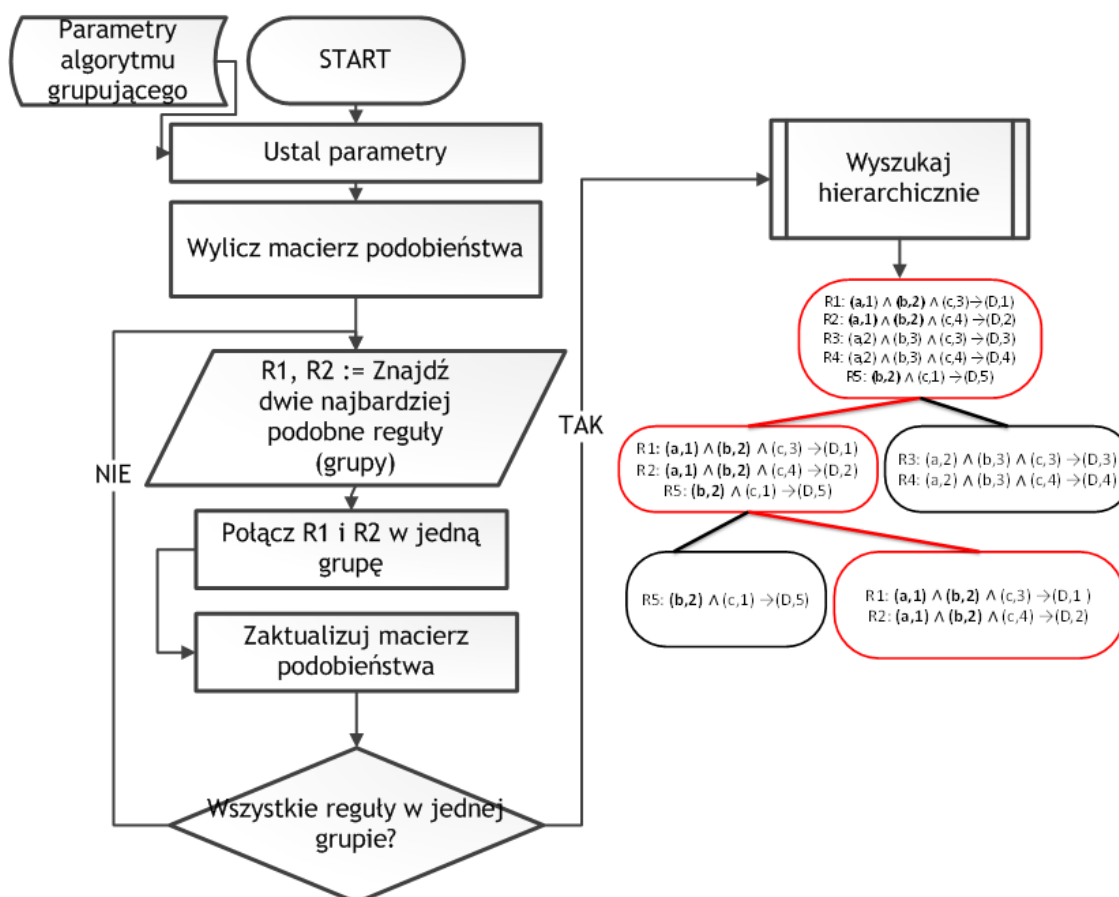
Autorzy rozumieją niepełność wiedzy jako sytuację, gdy dla danej reguły nie wszystkie przesłanki są faktami w bazie wiedzy, z kolei niepewność wiedzy przypisywana jest konkluzji reguły, której nie wszystkie przesłanki były faktami w bazie wiedzy.

1.2. Dotychczasowe badania

W swych dotychczasowych pracach [12-17] autorzy skupili się na wypracowaniu optymalnej metody grupującej reguły w bazie wiedzy. W tym celu przeanalizowano algorytmy hierarchiczne AHC oraz mAHC [2-3]. Dokonano również analizy parametrów wpływających na efektywność wyszukiwania reguł: kryteriów łączenia skupień w grupy, tworzenia reprezentantów skupień reguł. Zaproponowano modyfikację metody ścieżki najbardziej obiecującej, wywodzącej się z prac Saltona [4], oraz przedstawiono propozycję wyznaczania poprawnej liczby skupień w danych. Przykłady reguł oraz konkretne rozwiązania użyte przez autorów zostały przedstawione w poprzednich pracach [12-17].

2. Proponowane rozwiązanie

Autorzy proponują użycie algorytmu AHC do grupowania reguł wchodzących w skład bazy wiedzy. Schematyczny proces działania systemu przedstawiony jest na rysunkach 1 oraz 2.



Rys. 1. Ogólny schemat blokowy algorytmu

Fig. 1. Algorithm's basic flowchart

Algorytm rozpoczyna swoją pracę od pobrania od użytkownika parametrów. Następnie wyznaczana jest kwadratowa macierz podobieństwa, gdzie na przecięciu i -tej kolumny oraz j -tego wiersza wyliczana jest wartość podobieństwa i -tej oraz j -tej reguły z bazy wiedzy. W kolejnym kroku budowana jest struktura hierarchiczna (dendrogram) reguł przez odnajdywanie dwóch najbardziej podobnych skupień oraz łączenie ich w jedno. Proces ten powtarzany jest do momentu połączenia wszystkich skupień w jedną dużą grupę.

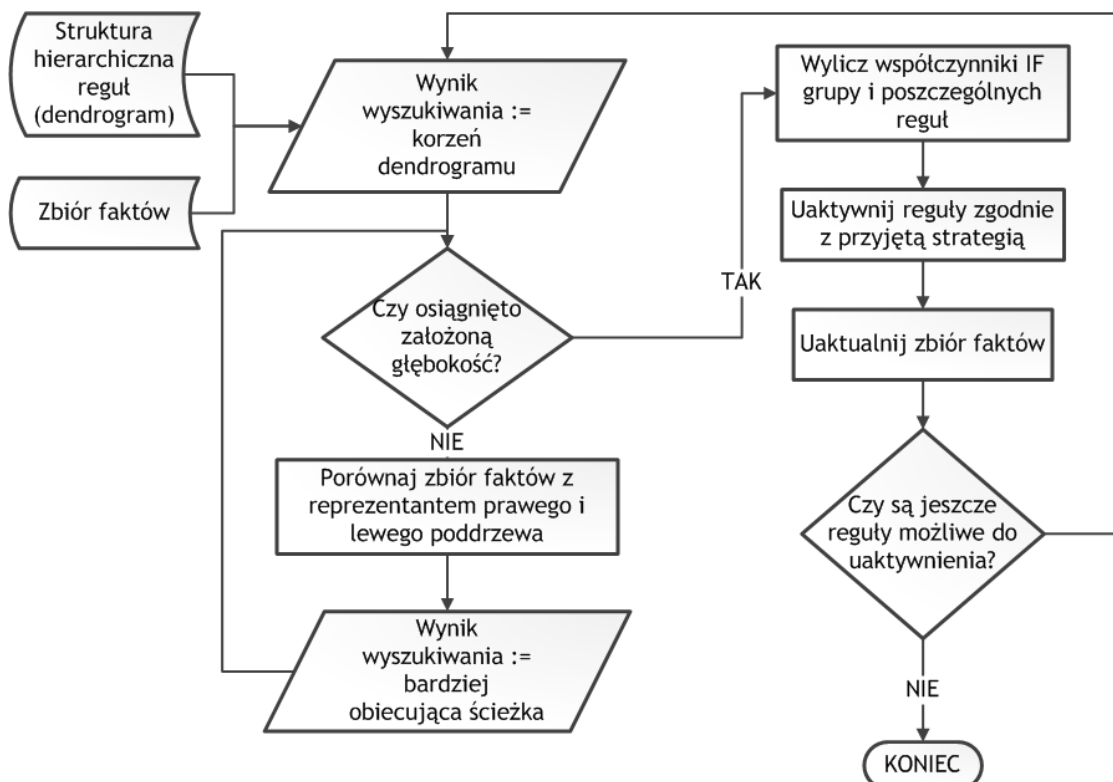
Po stworzeniu dendrogramu wykonywane jest właściwe wnioskowanie w strukturze hierarchicznej. Przy rozpoczęciu od korzenia drzewa aktualny zbiór faktów jest porównywany z reprezentantem lewego i prawego poddrzewa. Do dalszej analizy wybierana jest ścieżka bardziej obiecująca. Krok ten powtarzany jest do momentu uzyskania zakładanej głębokości¹.

¹ Rozumianej przez osiągnięcie odpowiedniego poziomu w dendrogramie; ściślej – możliwym jest uzyskanie w wyniku grupy reguł o zmiennej liczności składającej się z kilku podgrup. W skrajnym przypadku – algorytm w wyniku zwróci dokładnie jedną regułę będącą liściem w drzewie.

Uzyskana grupa analizowana jest pod kątem podobieństwa do aktualnego zbioru faktów. Reguły możliwe do uaktywnienia (pewne oraz niepewne) są uaktywniane, a ich konkluzje zostają dopisane do zbioru faktów wraz z odpowiednimi wartościami współczynników IF.

W przypadku kilku reguł możliwych do uaktywnienia algorytm działa zgodnie z przyjętą strategią rozstrzygania konfliktów omówioną w dalszej części artykułu.

System działa do momentu, w którym uaktywnianie dalszych reguł spowodowałoby dopisanie do bazy faktów wiedzy o zbyt niskiej jakości w stosunku do przyjętego minimum.



Rys. 2. Schemat wyszukiwania hierarchicznego

Fig. 2. Hierarchical search flowchart

2.1. Wyznaczanie stopnia niepełności wiedzy (IF)

Po odnalezieniu grupy najbardziej podobnej do aktualnego zbioru faktów system przystępuje do analizy reguł wchodzących w skład tej grupy. Jeśli znajdują się tam reguły mające pełne pokrycie w zbiorze faktów, te zostają uaktywniane. W przypadku gdy nie wszystkie przesłanki są prawdziwe, uaktywniane są reguły mające tylko częściowe pokrycie w zbiorze faktów. W celu odróżnienia wiedzy wyznaczonej przez reguły pewne od reguł niepewnych wprowadza się współczynnik pełności IF (ang. *incompleteness factor*). Współczynnik ten będzie wyznaczany w następujący sposób:

$$IF(DEC_i) = \frac{\sum_j IF(f_j)}{\text{card}(D_i)}; f_j \in (F \cap D_i),$$

gdzie: $IF(d_i)$ to współczynnik niepełności i -tej przesłanki, $IF(f_j)$ to współczynnik niepełności j -tego faktu będącego częścią wspólną zbioru deskryptorów D_i oraz zbioru faktów (F).

Fakty pewne, znane wcześniej lub też dopisane do zbioru faktów po uaktywnieniu reguł pewnych, mają z definicji wartość współczynnika IF równą 1. Współczynnik ten ma wartości z przedziału $[0;1]$, co pozwala na łatwą interpretację uzyskanych wyników. Przykładowo, dla zbioru faktów $\{(a,1), (b,2), (c,5)\}$ oraz następujących reguł: $R1 : (a,1) \wedge (b,2) \wedge (c,3) \rightarrow (D,1)$; $R2 : (a,1) \wedge (b,3) \wedge (c,4) \rightarrow (D,2)$ współczynniki IF tychże reguł wynoszą odpowiednio $IF(R1) = 0,67$ oraz $IF(R2) = 0,33$. Autorzy proponują również wprowadzenie średniej wartości współczynnika IF dla całej odnalezionej grupy. W przedstawianym przykładzie wartość tego współczynnika dla grupy złożonej z reguł $R1$ oraz $R2$ wynosiłaby 0,5.

Współczynnik IF powstał dzięki inspiracji współczynnikami CF zaproponowanymi w systemie MYCIN [8]. Był on odpowiedzią na trudne do zaimplementowania metody probabilistyczne. Formalnie wartość współczynnika CF to różnica pomiędzy miarą wiarygodności (MB, ang. *measure of belief*) a miarą wątpliwości (MD, ang. *measure of disbelief*) [9-10]. MB opisuje wiedzę systemu o zajściu danej hipotezy h przy znajomości przesłanki e , podczas gdy MD opisuje niewiedzę.

Niestety, pomimo obiecującego systemu obliczania drastycznie spada wartość współczynnika CF konkluzji w trakcie wnioskowania na długich łańcuchach przesłanek i konkluzji. Spadek ten utrudnia prawidłową interpretację końcowych wyników oraz powoduje problemy związane z niewystarczającą precyzją obliczeń zmiennoprzecinkowych.

Zaproponowany przez autorów współczynnik IF wykorzystuje jeszcze prostszy obliczeniowo sposób określania niepełności wiedzy. Stanowi łatwą miarę do modelowania niepewności i niepełności wiedzy, zależną tylko od stopnia pełności wiedzy zapisanej w zbiorze faktów. Umożliwia korzystanie z zalet mechanizmów grupowania (m.in. wybór grupy o największym współczynniku IF). Umożliwia również sterowanie „jakością” wnioskowania. Do użytkownika systemu należy decyzja, czy liczy się z możliwością otrzymania wiedzy niższej jakości, czy woli nie otrzymać żadnych nowych informacji w wyniku procesu wnioskowania.

2.2. Strategie rozwiązywania konfliktów

W przedstawionym systemie istnieje możliwość odnalezienia skupienia, w którym kilka reguł będzie równie relewantnych w stosunku do aktualnego zbioru faktów². W takiej sytu-

² Innymi słowy, współczynniki IF kilku reguł będą identycznie wysokie.

acji autorzy proponują sięgnięcie do znanych strategii uaktywniania reguł [11]. W literaturze wyróżnia się najczęściej 5 strategii stosowanych w systemach wspomaganie decyzji; są to:

1. strategia świeżości (ang. *recency*) – uaktywnianie tych reguł, które dodane były najpóźniej,
2. strategia specyficzności (ang. *specificity*) – uaktywnianie tych reguł, które mają największą liczbę przesłanek,
3. strategia blokowania (ang. *refractoriness*) – każdą regułę uaktywniamy tylko raz,
4. strategia „pierwsza reguła na liście” (ang. *textual order*) – uaktywniamy reguły zgodnie z kolejnością ich zapisu w bazie wiedzy,
5. strategia przypadkowości (ang. *random order*) – uaktywniamy reguły w losowej kolejności.

Jednak strategia przypadkowości bywa bardzo rzadko używana w rzeczywistych zastosowaniach, strategia blokowania jest zaś naturalnym i intuicyjnym sposobem na warunek stopu danego systemu.

Autorzy proponują również użycie strategii świeżości lub specyficzności jako metod rozwiązywania konfliktów. Przykładowo, jeśli baza faktów składa się z deskryptorów $\{(a,2), (b,2)\}$, to przykładowa baza wiedzy wraz z wyznaczonymi współczynnikami IF wygląda następująco:

$$R1: (a,2) \wedge (b,2) \wedge (c,6) \wedge (d,1) \rightarrow (D,4), \quad IF(R1)=0,5$$

$$R2: (a,1) \wedge (b,2) \rightarrow (D,1), \quad IF(R2)=0,5$$

$$R3: (a,1) \wedge (c,5) \rightarrow (D,3), \quad IF(R3)=0$$

$$R4: (a,1) \wedge (b,2) \wedge (c,4) \rightarrow (D,2), \quad IF(R4)=0,33$$

Jak widać, reguły $R1$ oraz $R2$ mają taką samą wartość współczynnika IF, więc trzeba rozstrzygnąć, którą z nich należy uaktywnić w pierwszej kolejności. Strategia świeżości nakaze uaktywnienie najpierw reguły $R2$ (jako dodanej najpóźniej), strategia specyficzności z kolei – reguły $R1$ jako tej, która ma większą liczbę przesłanek.

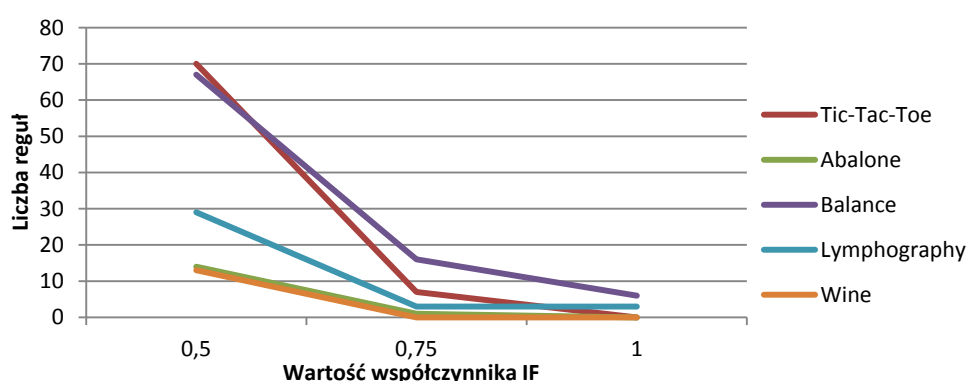
Wybór danej strategii dokonywany jest już na podstawie konkretnej dziedziny, w której zostanie zastosowany system wspomaganie decyzji.

3. Eksperymenty obliczeniowe

Do wykonania eksperymentów obliczeniowych wykorzystane zostały ogólnodostępne bazy danych zamieszczone w Machine Learning Repository [7]. Na ich podstawie za pomocą systemu RSES [5] wygenerowane zostały reguły minimalne składające się na analizowaną przez autorów bazę wiedzy. Wykorzystano bazy Tic-Tac-Toe, Abalone, Balance, Lymphography oraz Wine.

3.1. Minimalny progowy współczynnik IF

Po wygenerowaniu reguł minimalnych w pierwszym zaproponowanym eksperymencie wylosowano około 10% deskryptorów występujących w całym systemie. Następnie deskryptory te oznaczono jako fakty. Sprawdzone pokrycie reguł w bazie wiedzy takim zbiorem faktów, czyli dokonano analizy liczby reguł możliwych do uaktywnienia przy losowo wybranym zbiorze faktów. Analizę tę przeprowadzono w celu rozróżnienia w rozróżnieniu minimalnego współczynnika IF pozwalającego na uaktywnienie danej reguły (innymi słowy: dla $IF=0,75$ tylko 75% przesłanek musiało być spełnionych, aby uznać regułę za możliwą do uaktywnienia). Wyniki przedstawia rysunek 3.



	0,5	0,75	1
Tic-Tac-Toe	70	7	0
Abalone	14	1	0
Balance	67	16	6
Lymphography	29	3	3
Wine	13	0	0

Rys. 3. Liczba możliwych do uaktywnienia reguł a minimalny współczynnik IF

Fig. 3. The number of rules which could be fired compared to the minimal IF value

Jak widać, jeśli dopuszczone zostanie wnioskowanie tylko na regułach, których wszystkie przesłanki są spełnione (a więc wartość współczynnika IF wynosi 1), to aż w trzech przypadkach nie otrzymamy żadnej nowej wiedzy. Jeśli tylko obniżymy minimalny próg współczynnika IF do wartości 0,75, system będzie w stanie zwrócić informacje, które mogą posłużyć do uszczegółowienia zapytania, a co za tym idzie – do skutecznego wspomaganie decyzji podejmowanej przez człowieka-eksperta.

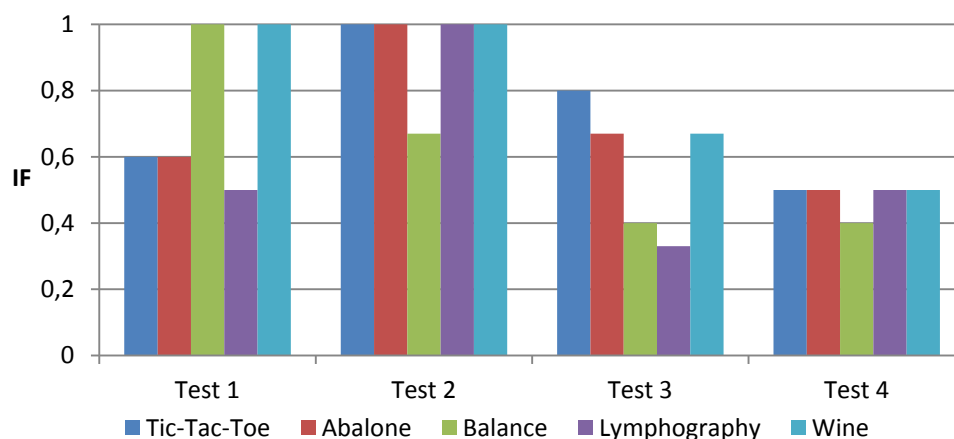
3.2. Maksymalny współczynnik IF wewnątrz odnalezionej grupy

Przeprowadzono również testy efektywności wnioskowania na podstawie powyższych baz wiedzy. W pierwszym z eksperymentów systemowi zawierającego hierarchicznie uporządkowaną bazę wiedzy zadano pytanie składające się z różnej liczby deskryptorów tworzących całą bazę faktów. W odpowiedzi system zwracał grupę reguł uznawanych za najbardziej

relewantne spośród wszystkich reguł obecnych w systemie. Wewnątrz tej grupy reguł odszukiwano taką, której współczynnik IF w stosunku do zadanego pytania był maksymalny.

Test został podzielony na cztery przypadki testowe, z których każdy był powtórzony czterokrotnie, a uzyskane wyniki zostały uśrednione. Przypadki testowe można zaklasyfikować następująco:

1. T1 oraz T2 – baza faktów składała się ze wszystkich przesłanek losowo wybranej reguły z bazy.
2. T3 – baza faktów to około 80% przesłanek losowo wybranej reguły.
3. T4 – baza faktów to około 50% przesłanek losowo wybranej reguły.



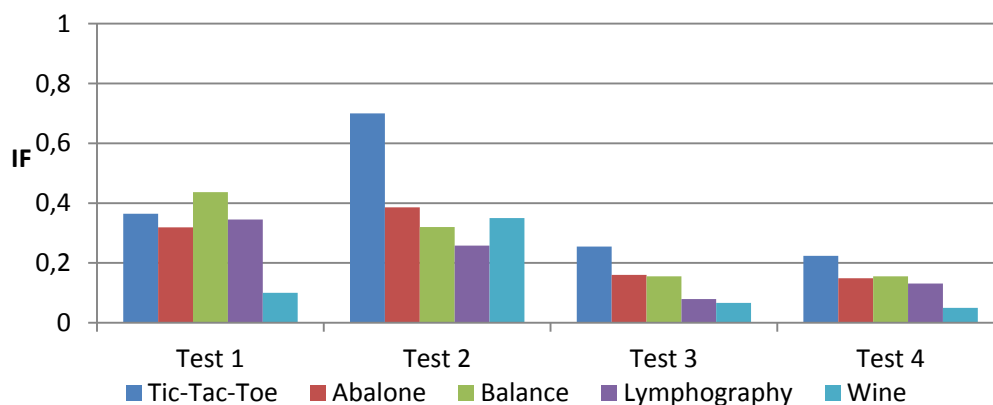
Rys. 4. Wyniki pierwszego eksperymentu
Fig. 4. The results of the first experiment

Wyniki eksperymentu pozwalają sądzić, iż proponowany system dobrze radzi sobie z odnajdowaniem grup reguł relewantnych w stosunku do zadawanego zbioru faktów. Należy tutaj podkreślić, że w teście nr 3 maksymalna możliwa wartość współczynnika IF wynosi około 0,8, a w teście 4 – około 0,5.

Jak widać, algorytm daje rezultaty zbliżone do optimum w większości przypadków. Dla większości baz udało się znaleźć reguły najbardziej odpowiednie do uaktywnienia.

3.3. Średni współczynnik IF wewnątrz odnalezionnej grupy

Kolejny z eksperymentów miał na celu zbadanie średniej wartości współczynnika IF wewnątrz odnalezionnej grupy. Sposób przeprowadzenia eksperymentu był analogiczny do poprzedniego, z tą różnicą, iż badano tutaj średnią wartość współczynnika IF wewnątrz odnalezionnej grupy. Dzięki temu możliwe jest zbadanie, czy system generuje skupienia o wysokiej jakości, w których reguły są spójne i mają dużą część wspólnych przesłanek. Wyniki przedstawione są na rysunku 5.



Rys. 5. Wyniki drugiego eksperymentu
 Fig. 5. The results of the second experiment

Otrzymane wyniki pokazują, iż system zwraca stosunkowo liczne skupienia, dzięki czemu średnia wartość współczynnika IF wewnątrz skupienia jest niska. Autorzy sugerują tutaj dalsze badania pozwalające na większe odróżnienie skupień od siebie, a co za tym idzie – na wzrost średniej wartości współczynnika IF wewnątrz odnalezionej grupy.

4. Wnioski i kierunki dalszych badań

W artykule zaproponowano nowe podejście do problemu wnioskowania w systemach wspomaganie decyzji z wiedzą niepełną. Dzięki zastosowaniu analizy skupień możliwe jest uzyskanie dodatkowej wiedzy z systemu w przypadku wystąpienia impasu, czyli braku reguł możliwych do uaktywnienia w klasycznym przypadku. Proponowane rozwiązanie pozwala na uaktywnianie reguł pewnych (przy wiedzy pełnej) oraz reguł niepewnych (przy wiedzy niepełnej) dzięki określeniu współczynnika niepełności wiedzy (IF).

Współczynnik IF stanowi także alternatywę do przeszukiwania struktury skupień reguł, które wcześniej realizowano metodą węzła najbardziej obiecującego. Trwają aktualnie prace porównujące te dwa podejścia.

Autorzy w swych dalszych badaniach pragną skupić się na optymalnym doborze parametrów algorytmu grupującego do zastosowania wraz z metodą współczynników IF.

BIBLIOGRAFIA

1. Chandru V., Hooker J.: Optimization methods for logical inference. John Wiley & Sons, New York 1999.

2. Kaufman L., Rousseeuw P. J.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York 1990.
3. Myatt G.: Making Sense of Data a Practical Guide to Exploratory Data Analysis and Data Mining. John Wiley and Sons, Inc., New Jersey 2007.
4. Salton G.: Automatic Information Organization and Retrieval. McGraw-Hill, 1975.
5. Bazan J., Szczuka M., Wróblewski J.: A new version of rough set exploration system, [in:] Third International Conference on RSCTC. Springer-Verlag, Malvern, PA 2002.
6. Pawlak Z.: Rough set approach to knowledge-based decision suport. European Journal of Operational Research, 1997.
7. Frank A., Asuncion A.: UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>], School of Information and Computer Science, Irvine, University of California, CA 2010.
8. Buchanan B. G., Shortliffe E. H.: Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, Addison-Wesley, MA 1984.
9. Dempster A. P.: A generalization of Bayesian inference. Journal of the Royal Statistical Society, Series B 30, 1968.
10. Shafer G.: A Mathematical Theory of Evidence. Princeton University Press, 1976.
11. Reichgelt H.: Knowledge Representation: An AI Perspective. Ablex Publishing Corporation, New Jersey, USA 1991.
12. Wakulicz-Deja A., Nowak-Brzezińska A., Jach T.: Inference processes using incomplete knowledge in Decision Support Systems – chosen aspects. Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science, 2012.
13. Jach T., Nowak-Brzezińska A.: Wybrane aspekty wnioskowania w systemach z wiedzą niepełną. Studia Informatica, Vol. 33, No. 2A, Gliwice 2012.
14. Jach T., Nowak-Brzezińska A.: Wnioskowanie w systemach z wiedzą niepełną. Studia Informatica, Vol. 32, No. 2A, Gliwice 2011.
15. Wakulicz-Deja A., Nowak-Brzezińska A., Jach T.: Inference processes in decision support systems with incomplete knowledge. Rough Sets and Knowledge Technology, Lecture Notes in Computer Science, Springer, Berlin/Heidelberg 2011.
16. Nowak-Brzezińska A., Jach T., Xięski T.: Wybór algorytmu grupowania a efektywność wyszukiwania dokumentów. Studia Informatica, Vol. 31, No. 2A, 2010.
17. Nowak-Brzezińska A., Jach T., Xięski T.: Analiza hierarchicznych i niehierarchicznych algorytmów grupowania dla dokumentów tekstowych. Studia Informatica, Vol. 30, No. 2A, Gliwice 2009.

Abstract

The paper presents further research regarding the incompleteness problem during inference in decision support systems. The authors introduce the incompleteness factor (IF), which is partially based on certainty factors developed by Salton, which along with cluster analysis methods lead to a better modeling of the uncertainty in decision support systems. Proposed approach clusters the rules' in knowledge base, chooses the best cluster and makes it possible to fire rules within it. When there are no rules which have all the premises belonging to the facts set, the user is able to fire the most relevant rules. The conclusion drawn from them are added to the fact set, but with the proper value of incompleteness factor.

The detailed way of how the algorithm works is shown on fig. 1 and 2. Furthermore, the authors give the scientific background for the proposed research covering among the others the Dempster-Schaffer theorem and Salton's works.

On figure 3 the advantage of proposed research is being shown by giving the number of rules, which are possible to fire under the different degrees of uncertainty. Afterwards, the experiments are conducted in order to show the maximal IF value found within the relevant cluster (fig. 4) and the average IF value within the same one (fig. 5).

The whole paper is concluded and summarized, where further research subjects are being shown.

Adresy

Agnieszka NOWAK-BRZEZIŃSKA: Uniwersytet Śląski, Instytut Informatyki,
ul. Będzińska 39, 41-200 Sosnowiec, Polska, agnieszka.nowak@us.edu.pl.

Tomasz JACH: Uniwersytet Śląski, Instytut Informatyki, ul. Będzińska 39,
41-200 Sosnowiec, Polska, tomasz.jach@us.edu.pl.