

WYSZUKIWANIE I AKTYWOWANIE REGUŁ W SYSTEMACH WSPOMAGANIA DECYZJI Z WIEDZĄ NIEPEŁNĄ

TOMASZ JACH

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

Streszczenie

W artykule przedstawione są rozważania na temat możliwości użycia metod analizy skupień (grupowania) w celu szybkiego wyszukiwania, aktywowania reguł oraz wnioskowania w złożonych bazach wiedzy w warunkach wiedzy niepełnej. Autor przedstawia dalszy ciąg badań skupiając się na różnicy pomiędzy wynikami osiąganymi przez algorytmy AHC i mAHC. Przedstawiona zostaje również metoda do wyznaczania optymalnej liczby skupień reguł w bazach wiedzy.

1. Wstęp

Skomplikowane systemy informatyczne od wielu lat rozpracowują złożone problemy natury algorytmicznej. Obok nich istnieje grupa problemów, która nie ulega łatwej algorytmizacji. W tym polu pomocne są systemy wspomaganie decyzji (SWD). Działanie ich [1] opiera się na zdolności do wyciągania logicznych i poprawnych wniosków dysponując zbiorem faktów oraz reguł. Klasyczne wnioskowanie polega na uaktywnianiu reguł, których wszystkie przesłanki są spełnione (inaczej: przesłanki są faktami znanymi w systemie).

Przez r_i oznaczmy i -tą regułę w systemie odpowiadającą przyjętej w klasycznych SWD postaci klauzuli Horna, gdzie każdy literał z części przesłankowej i decyzyjnej tworzony jest w oparciu o zbiór atrybutów A oraz zbiory 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: $r_i = d_1 \wedge d_2 \wedge \dots \wedge d_m \Rightarrow DEC_i$.

Jednym z głównych parametrów charakteryzujących efektywność SWD jest czas wnioskowania. Istnieje prosta zależność mówiąca o wzroście czasu wyszukiwania reguł wraz ze wzrostem ich liczby w bazie wiedzy. Długość samej reguły, a ściślej liczba przesłanek, również wpływa na długość czasu wnioskowania. Szczególnym przypadkiem są tzw. reguły złożone – czyli takie, gdzie para (atrybut, wartość) występująca w konkluzji jednej z reguł, występuje w części przesłankowej innej reguły. Reguły złożone powodują znaczący spadek szybkości wyszukiwania reguł. W procesie wnioskowania w przód, po wyszukaniu reguły następuje jej aktywacja. W klasycznych SWD wyszukana reguła musi mieć pełne pokrycie wszystkich przesłanek w zbiorze faktów. Konkluzje uaktywnionych reguł zostają dopisane do tego zbioru i dzięki temu inne reguły (których przesłanki nie znajdowały się wcześniej w zbiorze faktów) mogą zostać uaktywnione. Generowanie zbyt dużej liczby faktów powoduje lawinowy przyrost wiedzy, często trudnej do interpretacji. Przykładowo:

Baza wiedzy:

R1: IF (attr4=8600) AND (attr8=177) AND (attr1=152) => (class=2)
R2: IF (attr4=8600) AND (attr1=151) =>(class=2)
R3: IF (attr4=8600) AND (attr7=30) =>(class=2)

Baza faktów:

(*attr4=8600*), (*attr7=40*), (*attr1=152*)

Aby aktywować którąkolwiek regułę, wszystkie jej przesłanki muszą być spełnione [2]. W przedstawianym przypadku jednak żadna z reguł nie może zostać aktywowana. Zbiór faktów jest nie zawiera wszystkich przesłanek żadnej z reguł. W takim przypadku pojawia się impas: system nie może podjąć decyzji co do aktywacji jakiegokolwiek reguły, użytkownikowi nie zostają przedstawione żadne nowe informacje związane z wykonaniem wniosku. Zastosowanie algorytmów analizy skupień [3] spowoduje zgrupowanie reguł w bazie wiedzy używając wybranego kryterium podobieństwa. Dzięki temu nie tylko przyspieszone zostanie wyszukiwanie reguł, ale również możliwe będzie aktywowanie reguł, których większość przesłanek jest spełniona. Proponowany system po odnalezieniu grupy reguł najbardziej podobnych do aktualnego zbioru faktów umożliwi ich aktywowanie, nawet pomimo niespełnienia warunku o wszystkich przesłankach reguły zawartych w zbiorze faktów. Taka reguła oznaczona zostanie przez system jako niepewna, a wiedza przez nią generowana jako nie całkowicie pewna w sensie logicznym. Znaczne przyspieszenie procesu wyszukiwania jest istotne zwłaszcza w skomplikowanych i dużych bazach wiedzy ze względów praktycznych. Klasyczne algorytmy wnioskowania, zarówno wpród (sterowane faktami) jak i wstecz (sterowane celem) stają się nieefektywne pod względem złożoności czasowej, gdy muszą przeszukać całą bazę wiedzy, reguła po regule. Proponowane podejście wykorzystujące wiedzę o reprezentantach grup reguł podobnych do siebie, znajduje w bardzo krótkim czasie grupę najbardziej podobną do szukanych informacji (podanych faktów) i tylko tę grupę analizuje w procesie wnioskowania.

Zaproponowane podejście jest podobne do teorii współczynników CF [4], rozszerza ją o możliwość szybkiego wyszukania reguł oraz braku konieczności określania wartości współczynników CF dla wszystkich przesłanek. Alternatywne podejście spotkać można w systemach bazujących na zbiorach rozmytych [5] oraz przybliżonych [6].

Autor proponuje użycie pojęcia niepewności wiedzy dla reguł, których nie wszystkie przesłanki są spełnione. W proponowanym systemie takie reguły będą mogły być uaktywnione, przy czym zostanie określony stopień jej pokrycia pozwalający ocenić na ile możemy ufać danej regule.

2. Proponowane rozwiązanie

Autor zakłada, że mechanizmy analizy skupień mogą zostać skutecznie zastosowane do grupowania reguł w bazie wiedzy. Podejście to wraz z wyszukiwaniem reguł metodą pnia najbardziej obiecującego [7] oraz aktywacją tych reguł, których nie wszystkie przesłanki są spełnione pozwoli na ulepszenie dotychczasowych rozwiązań z dziedziny wnioskowania w systemach z wiedzą niepewną.

2.1. Algorytmy grupujące reguły w bazie wiedzy

Grupowanie reguł realizowane jest dzięki algorytmom hierarchicznym analizy skupień AHC oraz Agnes [3]. Szczegółowy sposób działania algorytmów prezentowany jest w poprzedniej pracy autorów [8], poniżej przedstawiony zostaje jedynie zarys i ogólne założenia.

W pierwszym kroku algorytmu analizowane jest podobieństwo reguł. W tym celu generowana jest kwadratowa macierz podobieństwa. Na przecięciu *i*-tego wiersza oraz *j*-tej kolumny znajduje się wartość określająca wzajemne podobieństwo dwóch reguł mierzonych przy użyciu miar podobieństwa szczegółowo omówionych w pracy [8]. Reguły najbardziej podobne tworzą skupienie, dzięki czemu w kolejnych krokach analizowane jest podobieństwo reguł, ale i skupień reguł do innych reguł i skupień. Proces powtarzany jest do otrzymania pełnej hierarchii grup, którą można przedstawić graficznie w formie dendrogramu [9].

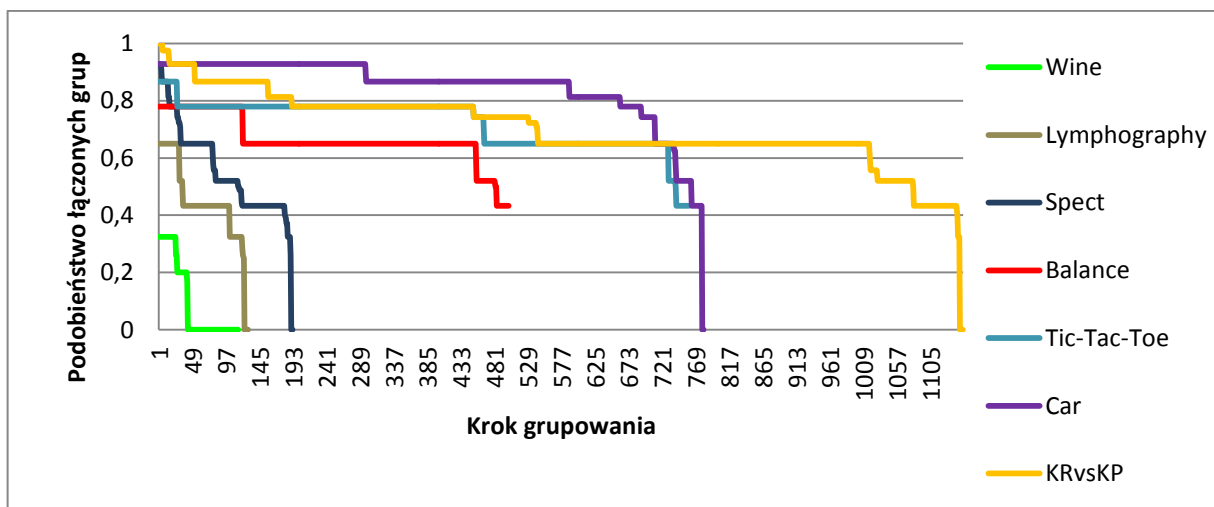
2.2. Wyszukiwanie reguł w bazie wiedzy

Wnioskowanie poprzedzone jest zawsze wyszukiwaniem reguły do aktywowania. We wcześniejszych pracach postawiono problem wyboru optymalnej liczby skupień. Liczba ta była wyznaczana metodą eksperymentalną. W tym opracowaniu przedstawiona zostanie próba automatycznego wyznaczania tej liczby za pomocą metod analizy skupień.

2.3. Wyznaczanie liczby skupień w danych

Proponowanym rozwiązaniem jest obserwacja wzajemnego podobieństwa grup łączonych w konkretnym kroku grupowania. Algorytm przedstawia się następująco:

1. Ustalenie parametrów grupowania: wybór miary podobieństwa, metody łączenia skupień, liczby skupień.
2. Rozpoczęcie algorytmu grupowania.
3. Zapamiętanie maksymalnej wartości podobieństwa dwóch reguł z pierwszego kroku algorytmu grupowania (simMax).
4. Wyznaczenie wartości progowej współczynnika podobieństwa dwóch skupień będącą iloczynem simMax oraz wartości podanej przez użytkownika określającej kiedy zakończyć grupowanie.
5. Wykonywanie algorytmu grupowania dopóki łączone skupienia mają wartość podobieństwa większą od wartości progowej



Rysunek 1: Podobieństwo grup łączonych w poszczególnych krokach algorytmu grupowania.

Rysunek 1 przedstawia wartość podobieństwa grupowanych reguł w kolejnych krokach grupowania. Widać wyraźnie, że w pewnym momencie jakość grupowania drastycznie spada. Autor sądzi, że jest to punkt, w którym łączone ze sobą skupienia reguł są już dość mało do siebie podobne. Jest to sygnał do zaprzestania grupowania. Eksperymenty przeprowadzono dla baz pobranych z Machine Learning Repository [10]. Autor wybrał bazy o różnym stopniu skomplikowania (różna liczba reguł, czy atrybutów opisujących reguły) aby ustalić optymalną wartość parametru współczynnika progowego.

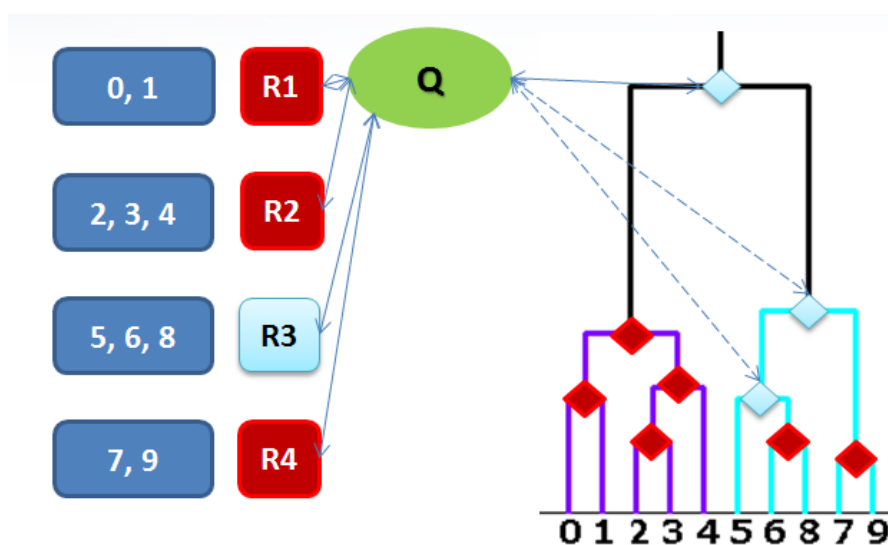
W wyniku eksperymentów, najlepsze rezultaty osiąga się w przypadku gdy współczynnik progowy ma wartość $0,85 * \text{simMax}$ (patrz Rysunek 1).

3. Wykorzystanie algorytmów AHC i mAHC do generowania skupień

Algorytmy AHC i mAHC stanowią przykład hierarchicznych aglomeracyjnych algorytmów grupujących [11]. Największą różnicą pomiędzy nimi jest kryterium stopu. AHC kończy działanie, gdy otrzymana zostanie jedna grupa złożona ze wszystkich reguł poddawanych grupowaniu. Algorytm mAHC za kryterium stopu przyjmuje pewien poziom, ustalany jako parametr. W każdym przypadku możliwym jest wykorzystanie metody ścieżki najbardziej obiecującej w celu szybkiego wyszukiwania relewantnych grup: reprezentanci lewego i prawego poddrzewa zostają porównani z aktualnym zbiorem faktów. Algorytm podąża tą ścieżką, której reprezentant ma większą wartość podobieństwa wobec aktualnego zbioru faktów.

Formalnie, jeśli przez D będziemy rozumieć zbiór deskryptorów (par atrybut-wartość), f to funkcja podobieństwa, która dwóm regułom (skupieniom reguł) przyporządkowuje wartość podobieństwa, a przez k_i, l_i – węzły łączone, wtedy każda grupa w_i będzie definiowana jako $w_i = (D_i, f, k_i, l_i)$, gdzie $D_i = \{d_1, \dots, d_m\}, f: X \times X \rightarrow R[0 \dots 1]$.

Idea działania systemu została zaczerpnięta z systemu SMART Saltona [7]. Analogicznie, wyszukiwanie reguł rozpoczyna się od korzenia drzewa. W każdym kolejnym kroku zbiór faktów porównywany jest z reprezentantami prawego i lewego poddrzewa aktualnie rozpatrywanego węzła (Rysunek 2). Dalsza analiza przeprowadzana jest dla ścieżki o większej wartości podobieństwa zbioru faktów do reprezentanta grupy reguł. Algorytm kończy się w momencie osiągnięcia poziomu liści oznaczającego konkretną regułę w bazie wiedzy.



Rysunek 2: Wyszukiwanie strukturalne (z lewej) i wyszukiwanie hierarchicznej (z prawej)

Metoda ścieżki najbardziej obiecującej opiera się na wyznaczaniu odległości pomiędzy dwoma skupieniami. W tym celu proponuje się trzy różne podejścia: metodę pokrycia deskryptorowego, pokrycia atrybutowego oraz podejście hybrydowe.

Najbardziej intuicyjną miarą jest wyznaczenie liczby deskryptorów występujących zarówno w zbiorze faktów, jak i w poszczególnych węzłach zgodnie ze wzorem:

$$f_d(k, l) = \text{card}(d_k \cap d_l) \quad (1)$$

gdzie d_l oraz d_k to zbiory deskryptorów węzłów l i k odpowiednio. Takie podejście, nazwane metodą pokrycia deskryptorowego, faworyzuje jednak węzły zawierające dużą liczbę

powtarzających się, częstych deskryptorów w systemie. Niepełne bazy wiedzy mogą nie zawierać informacji o wartościach deskryptorów, stąd już informacja o wspólnych atrybutach występujących w obu grupach powinna być brana pod uwagę w wyliczaniu podobieństw. Bez szczegółowej znajomości dziedziny bazy wiedzy niemożliwym jest ustalenie relacji pomiędzy wartościami poszczególnych atrybutów. Dzięki temu niemożliwym jest wyprowadzenie uniwersalnej funkcji podobieństwa uwzględniającej dystans pomiędzy cechami jakościowymi, stąd proponuje się ogólne podejście, właściwe niezależnie od rodzaju danych.

Drugim ze sposobów określania miary podobieństwa jest metoda pokrycia atrybutowego, która przy wyznaczaniu podobieństwa reguł (bądź ich skupień) bierze pod uwagę tylko informacje o wspólnych atrybutach. Przy zachowaniu konwencji oznaczeń omówionych w poprzednim rozdziale oraz gdy zbiory a_k oraz a_l oznaczać będą zbiory atrybutów reguł występujących w poszczególnych grupach, miara pokrycia atrybutowego będzie wyznaczana w sposób następujący:

$$f_a(k, l) = \text{card}(a_k \cap a_l) \quad (2)$$

Zauważono, że reguły są wzajemnie do siebie podobne, a liczność zbiorów V_{a_i} jest duża. Z tego powodu postanowiono wykorzystywać tylko informacje o wspólnych atrybutach w obliczaniu podobieństwa dwóch węzłów.

Ze względu na duże podobieństwo reguł do siebie oraz stosunkowo liczne zbiory wartości poszczególnych atrybutów, autorzy postanowili wykorzystywać tylko informacje o wspólnych atrybutach w obliczaniu podobieństwa dwóch węzłów. Dzięki temu można będzie wyróżnić, być może spójne, grupy reguł.

Trzecią metodą obliczania podobieństwa reguł do siebie będzie liczenie ważonej średniej z obu miar pokrycia przedstawionych wcześniej:

$$f_h(k, l) = \text{card}(d_k \cap d_l) \cdot C_1 + \text{card}(a_k \cap a_l) \cdot C_2 \quad (3)$$

gdzie C_1 oraz C_2 to dodatnie czynniki stopniujące sumujące się do wartości 1.

Autorzy sądzą, że podejście to wykorzysta zalety zarówno dokładności metody z pokryciem deskryptorów, jak również dodatkowych informacji rozróżniających reguły pomiędzy sobą. Współczynniki stopniujące służą zwiększaniu lub zmniejszaniu ważności części deskryptorowej i atrybutowej. W rozważaniach sprawdzono dwa przypadki, w jednym znacznie większą wagę otrzymuje część deskryptorowa, w drugim – część atrybutowa.

Przykładowo, dla dwóch węzłów $k: d_k = \{(A, 1), (A, 1), (A, 2), (B, 1), (B, 1), (C, 1)\}$ oraz $l: d_l = \{(A, 2), (A, 2), (B, 1), (B, 1), (B, 1), (C, 1)\}$ i zbioru faktów $Q = \{(A, 2)(C, 1)\}$ odpowiednie wartości podobieństw przedstawiają się następująco:

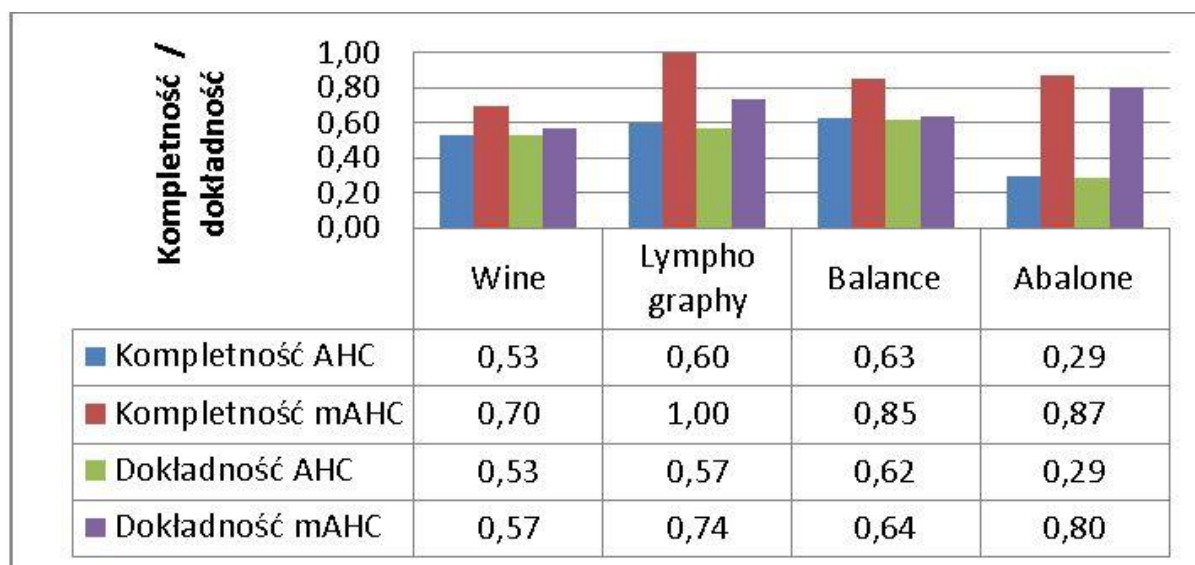
- $f_a(k, Q) = 2$; $f_a(l, Q) = 3$
- $f_a(k, Q) = 4$; $f_a(l, Q) = 3$
- dla $C_1=0,75$ oraz $C_2=0,25$ $f_h(k, Q) = 2,5$; $f_h(l, Q) = 3$
- dla $C_1=0,25$ oraz $C_2=0,75$ $f_h(k, Q) = 3,5$; $f_h(l, Q) = 3$

Widać wyraźnie, że podejście hybrydowe pozwala na uwzględnienie również współwystępowania atrybutów w zbiorze faktów oraz analizowanej regule i być może przyczyni się do wyboru optymalnej ścieżki w warunkach dużego podobieństwa reprezentantów węzłów reprezentujących obie ścieżki do siebie.

4. Eksperymenty obliczeniowe

W celu porównania przedstawionych podejść, autor zaimplementował dwa hierarchiczne algorytmy grupowania: omówiony wcześniej AHC korzystający z pełnego drzewa hierarchicznego reguł i wyszukiwania oraz mAHC wykorzystujący wcześniej omówioną technikę wyznaczania optymalnej liczby skupień i wyznaczający reprezentantów każdej z grup. Wyniki przedstawia Rysunek 3.

Oczywistym jest, że podejście hierarchiczne będzie znacznie szybsze i hipoteza ta znalazła swoje potwierdzenie w wynikach eksperymentalnych. Co więcej – im baza reguł większa, tym widoczny większy zysk czasowy dla podejścia hierarchicznego.



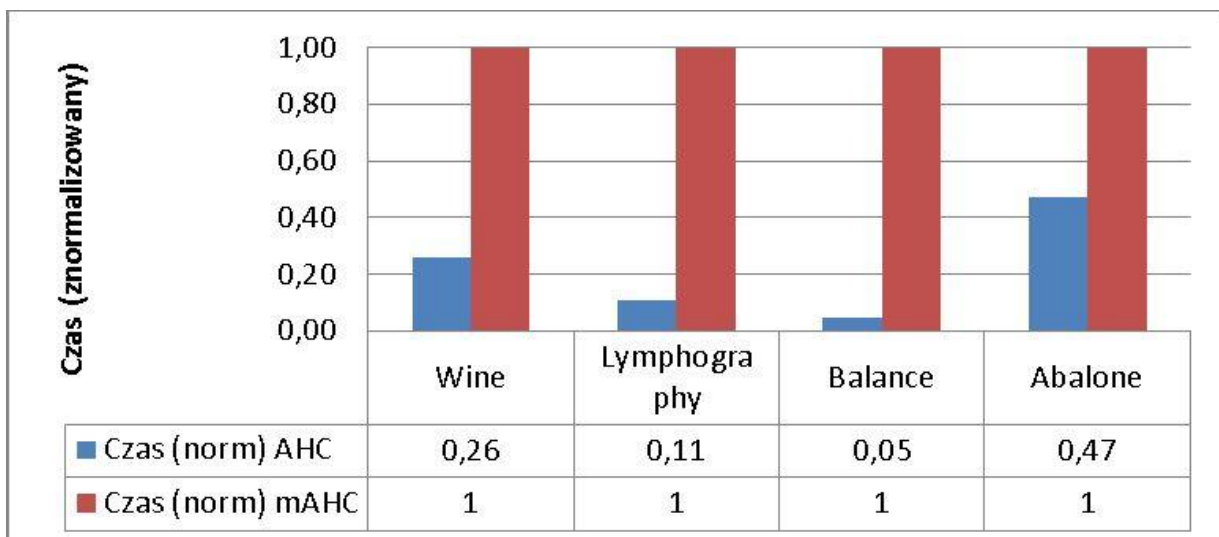
Rysunek 3: Jakość wyszukiwania hierarchicznego i strukturalnego

Oba algorytmy grupowania testowano dla czterech wybranych baz danych z repozytorium MLR [10]. Przygotowano dla każdej bazy wiedzy osobno po 10 zestawów faktów losowo wybranych spośród przesłanek i konkluzji reguł faktycznie zapisanych w tych bazach. Obliczono wartości kompletności [rozumianej jako stosunek liczby wspólnych deskryptorów występujących zarówno w zbiorze faktów jak i w reprezentancie (mAHC) lub otrzymanej grupie (AHC) do liczby wszystkich deskryptorów opisujących zbiór faktów i reprezentanta (mAHC) lub grupę (AHC)].

Niestety wyszukiwanie z użyciem hierarchii daje stosunkowo gorsze rezultaty. Autor sugeruje dalsze prace nad udoskonaleniem wyszukiwania, w szczególności omówione w dalszej części artykułu.

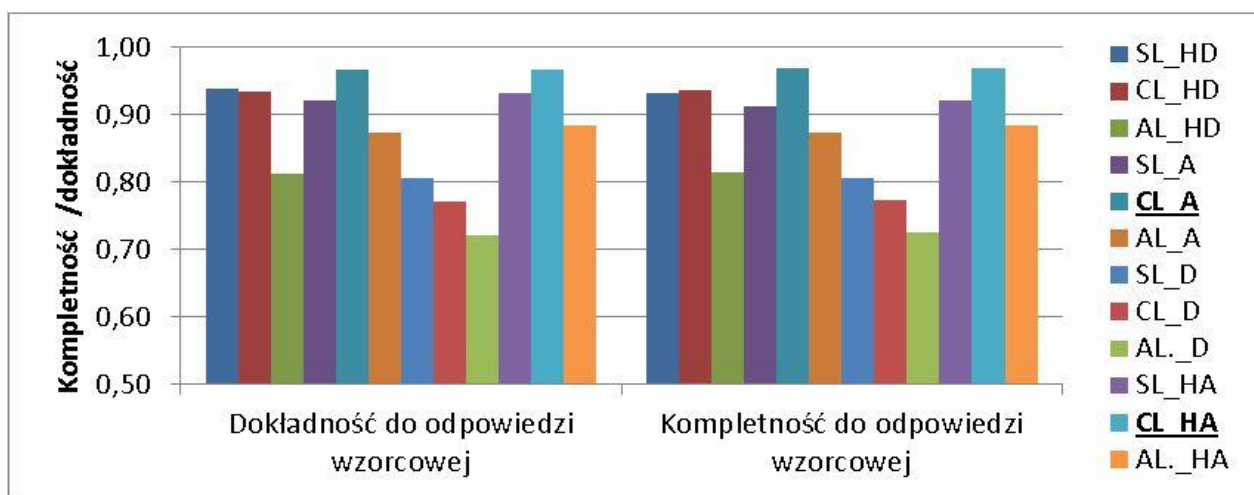
Autor przeanalizował również czas wyszukiwania reguł do uaktywnienia dla obu algorytmów. Wyniki prezentuje Rysunek 4.

W celu sprawdzenia, która z przedstawionych metod da optymalne rezultaty, przeprowadzono eksperymenty obliczeniowe. Na początku, przyjmowano, że deskryptory opisujące aktualnie analizowaną regułę stanowią jednocześnie zbiór faktów. Do pełnego systemu wyznaczonego za pomocą różnej kombinacji metod ścieżki najbardziej obiecującej oraz metody łączenia skupień, zadawano ów zbiór faktów. Odpowiedź systemu traktowano jako odpowiedź wzorcową.



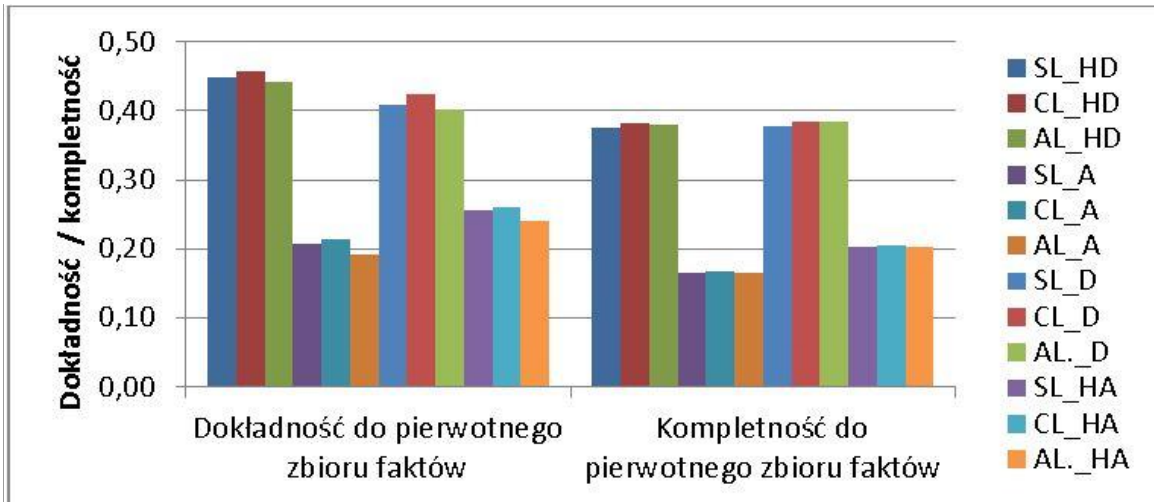
Rysunek 4: Czas wyszukiwania z użyciem algorytmów AHC i mAHC

Następnie, z bazy wiedzy usuwano tę konkretną regułę i powtarzano proces wyszukiwania reguł dla tego samego zbioru faktów. Sprawdzano następnie wartości kompletności i dokładności wyszukiwania dla takiego przypadku. Na wszystkich wykresach przyjęto następujące oznaczenia: SL – Single Linkage, CL – Complete Linkage, AL – Average Linkage, HD – metoda hybrydowa ścieżki najbardziej obiecującej ze zwiększoną wartością współczynnika dla wspólnych deskryptorów, HA - metoda hybrydowa ścieżki najbardziej obiecującej ze zwiększoną wartością współczynnika dla wspólnych atrybutów, A – metoda pokrycia atrybutowego, D – metoda pokrycia deskryptorowego.



Rysunek 5: Eksperymenty dla metody ścieżki najbardziej obiecującej

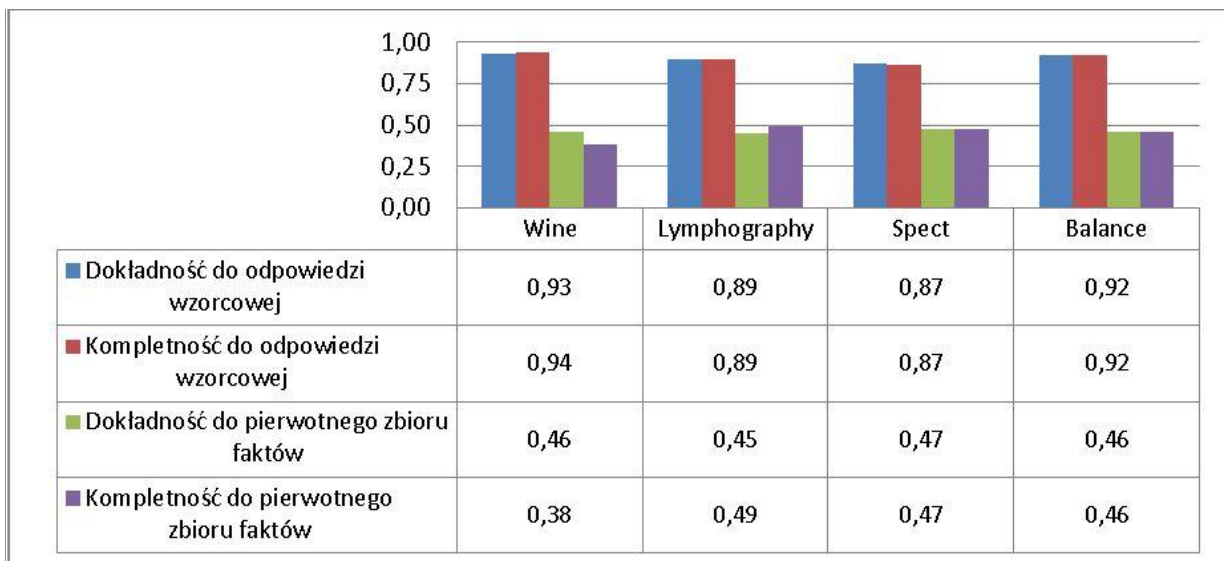
Najlepsze rezultaty uzyskano gdy algorytm korzystał z metody CL do wiązania skupień. W przypadku oceny dokładności i kompletności w stosunku do najlepszej grupy, wyniki wszystkich zaprezentowanych podejść prezentują zbliżone rezultaty (najlepszą jest metoda hybrydowa z większą wagą dla wartości wspólnych atrybutów oraz metody pokrycia atrybutowego). Wydaje się, że podejście rozróżniające grupy daje lepsze rezultaty w stosunku do tradycyjnego podejścia pokrycia deskryptorowego.



Rysunek 6: Wyniki eksperymentów obliczeniowych

W drugiej części eksperymentu, obliczono kompletność i dokładność odpowiedzi dla ograniczonego systemu w stosunku do zadawanego zbioru faktów. W ten sposób autorzy pragnęli zbadać zdolność stworzonego systemu do kompensowania niepełnej wiedzy. Należy jednakże zauważyć, że wartości kompletności i dokładności bliskie maksymalnym nie mogą wystąpić, ze względu na fakt usunięcia reguły, która jest optymalną odpowiedzią na zbiór faktów z systemu (jego ograniczenie).

Rysunek 6 jednoznacznie pokazuje, że autorska metoda hybrydowa ze wzmocnieniem części deskryptorowej sprawdza się lepiej, niezależnie nawet od sposobu łączenia skupień. Wartości parametrów efektywności uzyskiwane dla tej metody są blisko dwukrotnie większe od pozostałych rozwiązań. Do dalszych eksperymentów przyjęto metodę całkowitego wiązania (CL) oraz ścieżki najbardziej obiecującej ze wzmocnieniem wag dla wspólnych deskryptorów. Ustaliwszy je, autor przeprowadził testy dla większej liczby baz wiedzy. Wyniki prezentuje Rysunek 7. Bliskie maksymalnym wartości kompletności i dokładności w stosunku do grupy optymalnej potwierdzają słuszość zaproponowanego podejścia w problemie wnioskowania w systemach z wiedzą niepełną.



Rysunek 7: Wyniki eksperymentów metody hybrydowej dla wybranych baz wiedzy

5. Wnioski oraz kierunki dalszych badań

Autor przedstawił metodę wyznaczania optymalnej liczby skupień w danych złożonych z reguł baz wiedzy. Dzięki temu udało się poprawić rezultaty osiągnięte wcześniej [8]. Metoda ta pozwoliła na wyeliminowanie wielokrotnego powtarzania eksperymentów w celu dobrania właściwej liczby skupień. Zaproponowany został również wstępny model wyszukiwania reguł możliwych do aktywowania. Z eksperymentów wynika, iż model ten sprawdza się również w warunkach wiedzy niepełnej. Autor przedstawił również klasyczny algorytm ścieżki najbardziej obiecującej służący do szybkiego wyszukiwania skupień reguł. Dokonano jego modyfikacji i ulepszenia w celu dostosowania do specyficznych danych, jakim są reguły w bazach wiedzy. Dzięki temu dokonano dalszego skrócenia czasu wyszukiwania, zwiększenia odporności na niepełność wiedzy oraz zwiększenia jakości skupień reguł zwracanych przez system na zadany zbiór faktów. Utworzone skupienia oraz szybki sposób ich wyszukiwania umożliwiają odnajdowanie reguł o minimalnej liczbie niespełnionych przesłanek w krótkim czasie. Po ich aktywacji, zwiększona zostaje liczba faktów, co spowoduje wyprowadzenie obszerniejszej wiedzy z systemu.

W wyniku licznych eksperymentów sądzi się, że współczynnik progowy dla optymalnej liczby grup powinien mieć wartość około 85% wartości największego podobieństwa dwóch reguł między sobą. Powoduje to utworzenie dużej liczby małolicznych grup. Narzut obliczeniowy związany z wyznaczaniem reprezentantów oraz porównywaniem zbioru faktów z regułami lub skupieniami jest rekompensowany przez wysoką jakość odnajdowanych skupień.

Zaproponowane podejście wymaga dalszych badań ze względu na tendencję do łańcuchowania grup reguł. Krótki opis każdej reguły oraz ich mała rozróżnialność względem siebie mogą przyczynić się do zaburzenia równomierności dendrogramu (zdarzało się i tak, że w jednym z poddrzew na każdym poziomie mamy jedną regułę tylko, a w drugim – pozostałe). Sytuacja ta ma miejsce ze względu na małą różnorodność wartości wewnątrz macierzy podobieństwa budowanej na początku działania algorytmu. Przykładowo, baza Abalone posiada 7138531 komórek w macierzy podobieństwa, a tylko 43 unikalne wartości podobieństwa. Z racji tego, kolejność tworzenia skupień determinowana jest poprzez kolejność napływu danych do bazy wiedzy. Trwają prace mające na celu wyeliminowanie tej niedogodności.

Optymistyczną obserwacją jest znacząca poprawa algorytmu wyszukiwania reguł metodą ścieżki najbardziej obiecującej za pomocą autorskiej metody hybrydowej z większą wagą dla wspólnych deskryptorów.

W kolejnych pracach autor zamierza wykorzystać współczynniki pewności CF w celu poprawnego zamodelowania niepewności generowanej przez wiedzę niepełną. Zaproponowane zostaną również metody służące do przeprowadzania pełnego procesu wnioskowania. System zostanie wzbogacony o moduł informowania użytkownika o możliwościach aktywowania reguł niepewnych i tym samym - wnioskowania w warunkach wiedzy niepełnej.

Bibliografia

- [1] Nowak-Brzezińska, A., Simiński, R., Jach, T., Xięski, T.: Towards a practical approach to discover internal dependencies in rule-based knowledge bases. *Rough Sets and Knowledge Technology* (2011)
- [2] Nowak-Brzezińska, A., Wakulicz-Deja, A.: Analiza efektywności wnioskowania w złożonych bazach wiedzy. *Systemy Wspomagania Decyzji* (2007)
- [3] Jain, A., Dubes, R.: *Algorithms for clustering data*. Prentice Hall, New Jersey (1988)
- [4] Chandru, V., Hooker, J.: *Optimization methods for logical inference*. John Wiley & Sons, New York (1999)
- [5] Zadeh, L., Kacprzyk, J.: *Fuzzy logic for the management of uncertainty*. John Wiley & Sons, New York (1992)
- [6] Pawlak, Z.: Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 48-57 (1997)
- [7] Salton, G.: *Automatic Information Organization and Retrieval*. McGraw-Hill, New York, USA (1975)
- [8] Jach, T., Nowak-Brzezińska, A.: Wnioskowanie w systemach z wiedzą niepełną. *Studia Informatica, Zeszyty naukowe Politechniki Śląskiej* Vol. 32, 377-391 (2011)

- [9] Koronacki, J., Ćwik, J.: Statystyczne systemy uczące się. Exit, Warszawa (2008)
- [10] Frank, A., Asuncion, A.: UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. , Irvine, CA: University of California, School of Information and Computer Science. (2010)
- [11] Kaufman, L., Rousseeuw, P. J.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York (1990)
- [12] Latkowski, R.: Wnioskowanie w oparciu o niekompletny opis obiektów (praca magisterska). Wydział Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego, Warszawa (2001)
- [13] Myatt, G.: Making Sense of Data A Practical Guide to Exploratory Data Analysis and Data Mining. John Wiley and Sons, Inc., New Jersey (2007)
- [14] Kumar, V., Tan, P., Steinbach, M.: Introduction to Data Mining. Addison-Wesley (2006)
- [15] Geiger, D., Heckerman, D.: Knowledge representation and inference in similarity networks and Bayesian multinets. Artificial Intelligence, 45-74 (1996)
- [16] Towell, G., Shavlika, J.: Knowledge-based artificial neural networks. Artificial Intelligence, 119-165 (1994)
- [17] Bazan, Jan; Nguyen, Hung Son; Nguyen, Sinh Hoa; Synak, Piotr; Wróblewski, Jakub: Rough set algorithms in classification problems. Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems, 49-88 (2000)
- [18] Bazan, J., Szczuka, M., Wróblewski, J.: A new version of rough set exploration system. Third International Conference - RSCTC, 397-404 (2002)