

Ewa STRASZECKA

ZAKŁAD ELEKTRONIKI BIOMEDYCZNEJ, INSTYTUT ELEKTRONIKI, POLITECHNIKA ŚLĄSKA W GLIWICACH

Jednoczesna ocena informacji ilościowej i jakościowej podczas wspomaganie diagnostyki medycznej

dr inż. Ewa STRASZECKA

Studia i doktorat na Wydziale Automatyki, Elektroniki i Informatyki Politechniki Śląskiej. Staże zagraniczne na uczelniach w Belgii, Francji i Austrii. Obecnie adiunkt w Zakładzie Elektroniki Biomedycznej Instytutu Elektroniki Politechniki Śląskiej. Autorka ponad 70 publikacji konferencyjnych i artykułów naukowych z zakresu wspomaganie diagnostyki medycznej z zastosowaniem zbiorów rozmytych i teorii Dempstera-Shafera.



e-mail: ewa.straszecka@polsl.pl

Streszczenie

Diagnoza medyczna bazuje na niepełnej i nieprecyzyjnej informacji, dlatego algorytmy wspomaganie wnioskowania medycznego muszą spełniać specyficzne wymagania. Praca koncentruje się na jednoczesnej i równoważnej ocenie parametrów medycznych różnej natury: mierzalnych (np. testy laboratoryjne), sformułowanych ściśle (ciąża), określanych nieprecyzyjnie (przyrost wagi), a czasem definiowanych w umownej skali (ból). Proponuje się modelowanie wnioskowania medycznego z zastosowaniem teorii Dempstera-Shafera rozszerzonej poprzez zdefiniowanie rozmytych elementów ogniskowych. Pozwala to na reprezentację wiedzy w postaci reguł. W przesłankach tych reguł mogą występować zarówno zmienne ilościowe, jak i jakościowe. Każdej regule jest przypisana wartość bazowego prawdopodobieństwa zdefiniowanego zgodnie z teorią Dempstera-Shafera. Funkcje przynależności charakteryzujące zmienne w przesłankach reguł oraz rozkład bazowego prawdopodobieństwa można wyznaczyć na podstawie danych uczących. Wniosek diagnostyczny jest wynikiem porównania wartości miar przekonania (*Bel*) dla kilku hipotez. Przedstawiony model wnioskowania został zweryfikowany się dla 3 niezależnych baz danych dotyczących chorób tarczycy.

Słowa kluczowe: wspomaganie diagnozy medycznej, zbiory rozmyte, teoria Dempstera-Shafera

Simultaneous estimation of quantity and quality information in medical diagnosis support.

Abstract

Medical diagnosis is based on uncertain and imprecise information. Therefore, algorithms that support medical inference comply with specific requirements. This paper is focused on simultaneous and equal estimation of medical parameters of different nature: measurable (like laboratory tests), precisely formulated (pregnancy), described in an imprecise way (putting on weight), or defined on an assumed scale (pain). It is suggested to model a medical inference in the framework of the Dempster-Shafer theory extended for fuzzy focal elements. By means of the proposed algorithm, diagnostic rules can be formulated. Premises of the rules may include both quantity and quality variables. Each rule is assigned with a value of the basic probability assignment that is defined according to the Dempster-Shafer theory. Membership functions of rule predicates as well as the basic probability assignment are found from training data. The diagnostic conclusion is formulated after a comparison of belief values for several hypotheses. The model of inference is verified for 3 independent data bases of thyroid gland diseases.

Keywords: medical diagnosis support, fuzzy sets, the Dempster-Shafer theory of evidence

1. Wstęp

Diagnostyka medyczna wykorzystuje niepełną i nieprecyzyjną informację. Lekarz pierwszego kontaktu podejmuje decyzje na podstawie danych pochodzących z wywiadu, badania przedmiotowego oraz wyników podstawowych testów laboratoryjnych. Dane diagnostyczne mają więc charakter zarówno jakościowy (na przykład występowanie zadyszki wysiłkowej), jak i ilościowy (wyniki pomiarów ciśnienia lub testów laboratoryjnych). Można wyróżnić parametry medyczne sformułowane w sposób precyzyjny (przykładowo: ciąża), jak i opisywane nieprecyzyjnymi określeniami lingwistycznymi (znaczny przyrost wagi). Często należy rozpatrzyć objawy, których intensywność jest określana w umownej skali. Klasycznym przykładem takiego objawu jest ból. Podczas wnioskowania medycznego należy zatem stosować miary precyzyjności. Diagnoza medyczna może być również niepewna. Z jednej strony, objawy mogą wskazywać na chorobę z pewnym prawdopodobieństwem. Z drugiej strony, wniosek diagnostyczny jest sformułowany na podstawie niepełnej informacji, ponieważ wykonuje się zazwyczaj tylko kilka spośród szerokiego wachlarza badań. Konieczne jest więc uwzględnienie miar pewności. Wynika z tego, że podczas wnioskowania informacja ilościowa i jakościowa muszą być rozpatrywane jednocześnie, wraz z rozważeniem pewności diagnozy i precyzyjności symptomów. Stawia to szczególnie trudne zadania przed algorytmami wspomaganie decyzji w medycynie. Miary precyzyjności (na przykład w postaci funkcji przynależności zbiorów rozmytych), jak i pewności (przykładowo: prawdopodobieństwo warunkowe) były dotychczas wykorzystywane w wielu medycznych systemach doradczych, lecz zawsze były osobno interpretowane. W dodatku zastosowanie prawdopodobieństwa pociąga za sobą wiele problemów obliczeniowych. Dla umożliwienia obliczeń często konieczne są istotne uproszczenia modelu diagnozy, najczęściej założenie niezależności objawów, a nawet przyjęcie wartości prawdopodobieństw warunkowych bez dostatecznych danych populacyjnych.

Powyższe problemy były powodem poszukiwania modelu wnioskowania medycznego, w którym przesłanki byłyby sformułowane z wykorzystaniem miary precyzyjności, a wnioski (tj. diagnozy) charakteryzowałyby miara pewności. Jako miarę precyzyjności objawów zastosowano funkcje przynależności zbiorów rozmytych. Miary pewności diagnozy to miara przekonania (*Bel*) oraz domniemania (*Pl*) [1] wyznaczone na podstawie rozkładu bazowego prawdopodobieństwa, definiowanego w teorii Dempstera-Shafera [1]. Proponowana reprezentacja wiedzy umożliwia znalezienie relacji łączących zmienne jakościowe i ilościowe oraz opisanie tych relacji za pomocą reguł. Regułom przypisuje się wagę diagnostyczną w postaci wartości bazowego prawdopodobieństwa. Można zdefiniować poziom precyzyjności jako podobieństwo pomiędzy objawem reprezentowanym w regule diagnostycznej, a obserwacją uzyskaną podczas diagnozowania. Stwarza to warunki do jednolitej interpretacji obserwacji (czyli informacji wejściowej algorytmu), niezależnie od tego, czy ma ona charakter ilościowy, czy też jakościowy.

2. Metody

2.1. Miary pewności diagnozy

W teorii Dempstera-Shafera [1] definiuje się miarę przekonania (oznaczaną przez Bel – ang. „belief” [3]) oraz domniemania (Pl – ang. „plausibility” [3]). Określenie wartości tych miar następuje na podstawie zdefiniowanych uprzednio elementów ogniskowych czyli przesłanek, o których posiadamy informacje. Informacje te wyraża rozkład bazowego prawdopodobieństwa [1], [3]. W teorii tej, inaczej niż dla prawdopodobieństwa, pomija się założenie o niezależności zdarzeń, dlatego każda informacja odnosząca się do pojedynczych elementów ogniskowych, jak i ich zbiorów może być wykorzystana w procesie wnioskowania. Rozkład bazowego prawdopodobieństwa m jest zdefiniowany następująco [3]:

$$m(f) = 0, \quad (1)$$

$$\sum_{a \in T} m(a) = 1,$$

gdzie f oznacza przesłankę fałszywą, a T zbiór elementów ogniskowych.

Jeżeli przesłanka b wskazuje na hipotezę a ($b \Rightarrow a$), wówczas wartość przekonania do hipotezy a na podstawie informacji o b wylicza się jako [1]:

$$Bel(a) = \sum_{(b \Rightarrow a) = t} m(b), \quad (2)$$

gdzie t oznacza prawdę. Podaje się także definicję $Pl(a)$ jako miary dualnej do $Bel(a)$, to jest [1]:

$$Pl(a) = \sum_{(b \Rightarrow a) \neq f} m(b). \quad (3)$$

Elementy ogniskowe a mogą być interpretowane jako pojedyncze symptomy lub ich zbiory. Zbiór T gromadzi wszystkie symptomy brane pod uwagę w wybranej diagnozie. Element ogniskowy f odpowiada brakowi jakichkolwiek symptomów. Rozkład bazowego prawdopodobieństwa można ustalić na podstawie bazy danych lub zaczerpnąć z opracowań literaturowych [7]. W pierwszym przypadku można zbadać częstość występowania objawu u pacjentów z wybraną diagnozą, a następnie wyznaczyć częstości znormalizować tak, aby była spełniona zależność (1). Tak więc, dla reguły diagnostycznej dotyczącej podzbioru symptomów $s_i = \{s_{i1}, \dots, s_{ik}, \dots, s_{in}\}$ i diagnozy d_i :

$$\text{JEŻELI } s_i \text{ TO } d_i, \quad (4)$$

rozkład bazowego prawdopodobieństwa będzie określony jako:

$$m_i(f) = 0, \quad (5)$$

$$\sum_{s_i \in S_i} m_i(s_i) = 1.$$

Podzbiór symptomów będzie nazywany symptomem złożonym. W definicji (5) S_i oznacza zbiór wszystkich symptomów związanych z diagnozą d_i , czyli wszystkich przesłanek reguł postaci (4). Zachodzi: $s_i \subset S_i$ oraz istnieją $s_i \cap s_j \neq \emptyset$, dla $i \neq j$. W przypadku zaobserwowania złożonego symptomu s_i^* , dla którego wszystkie objawy składowe s_{ik} są obecne lub s_i^+ , dla którego niektóre z objawów składowych mogą być nieobecne, miary przekonania i domniemania diagnozy d_i wyznaczonej na podstawie wszystkich reguł (4) jej dotyczących, można obliczyć jako:

$$Bel(d_i) = \sum_{s_i^* \Rightarrow d_i} m_i(s_i^*), \quad (6)$$

$$Pl(d_i) = \sum_{s_i^+ \Rightarrow d_i} m_i(s_i^+). \quad (7)$$

Fakt obecności lub braku objawu stwierdza się na podstawie poziomu precyzyjności. W szczególnym przypadku s_i jest objawem pojedynczym, wówczas: $s_i^* \equiv s_i^+$.

2.2. Funkcje przynależności

Funkcje przynależności mogą być stosowane do definiowania objawów, dla których nie można podać norm, a nawet przedziału zmienności, na przykład „wzmożona potliwość”. W przypadku takich objawów konieczne jest jednak przyjęcie pewnej umownej skali jako nośnika zbioru rozmytego. Skale takie, jak wskazuje przykład skali VAS (ang. „Visual Analog Scale”) dla bólu, mają znaczącą wartość diagnostyczną. Jednakże zasadność zastosowania zbiorów rozmytych nie ogranicza się do tych przypadków. Także zmienne ilościowe, nawet takie, dla których istnieją normy warto charakteryzować za ich pomocą. Normy uniwersalne, jak również zależne od laboratorium, można interpretować za pomocą funkcji przynależności zbiorów rozmytych i weryfikować dla danych z wybranych populacji. W ten sposób można łączyć wiedzę eksperta z charakterystyką populacji. Jest to szczególnie ważne dla niektórych grup chorób. Na przykład występowanie chorób tarczycy łączy się z niedoborem jodu na pewnych terenach. Po zastosowaniu rozmytych elementów ogniskowych ten sam algorytm wnioskowania może być lepiej przystosowany do konkretnych warunków diagnostycznych bez bezpośredniej ingerencji użytkownika w bazę wiedzy, jedynie poprzez uwzględnienie odpowiednich danych uczących. Co więcej, możliwe staje się również wprowadzenie obserwacji diagnostycznej w postaci informacji lingwistycznej.

2.3. Poziom precyzyjności

Poziom precyzyjności określa stopień występowania (nasilenia) objawu. W przypadku parametrów jakościowych jednoznacznie określonych łatwo jest stwierdzić czy objaw występuje (tj. poziom precyzyjności jest równy 1), czy też jest nieobecny (zerowy poziom precyzyjności). Należy się jednak zastanowić nad interpretacją objawów postaci „przybytek wagi” (nazwa objawu oryginalna za tłumaczeniem indeksu Crooka [4]). W przypadku gdy pacjentka przytyła 1 kg w czasie 2 miesięcy objaw „przybytek wagi” nie jest jednoznaczny. Zapewne bardziej prawidłowe byłoby zdefiniowanie przynajmniej trzech funkcji przynależności (dla wartości lingwistycznych: „spadek wagi”, „utrzymanie wagi”, „przybytek wagi”) i przyporządkowanie im indywidualnych diagnoz. Wtedy niewielki przyrost wagi mógłby być scharakteryzowany przez małą wartość funkcji przynależności odpowiadającej „przybytkowi wagi”. Objaw należałoby więc opisać jako „ s_{ik} jest μ_{ik} ” gdzie μ_{ik} oznacza określenie lingwistyczne, dla którego jest zdefiniowana funkcja przynależności $\mu_i(x_{ik})$, a x_{ik} jest zmienną związaną z dziedziną objawu s_{ik} . Obecność pojedynczego objawu rozmytego definiuje poziom precyzyjności η_{ik} :

$$\eta_{ik} = \max_{x_{ik}} \min(\mu_i(x_{ik}), \mu_i^*(x_{ik})), \quad (9)$$

gdzie $\mu_i^*(x_{ik})$ oznacza obserwowaną wartość lingwistyczną symptomu. W najprostszym przypadku (dla mierzalnych parametrów) wartość obserwowana jest singletonem: $\mu_i^*(x_{ik}) = \delta_{x_{ik}, x_{ik}^*}$, czyli wartość lingwistyczna jest równoznaczna z numeryczną. Poziom precyzyjności symptomu złożonego s_i ocenia się na podstawie poziomu precyzyjności symptomów składowych. Symptom s_i^* (patrz (6)) jest obecny, jeżeli minimum poziomów precyzyjności symptomów składowych jest większe od założonego progu η , natomiast s_i^+ obserwuje się, gdy maksimum tych poziomów przewyższa próg:

$$\eta_i^* = \min_k \eta_{ik}, \eta_i^* \geq \eta, \Rightarrow s_i^* \text{ jest obecny,} \quad (10)$$

$$\eta_i^+ = \max_k \eta_{ik}, \eta_i^+ \geq \eta, \Rightarrow s_i^+ \text{ jest obecny.} \quad (11)$$

Wartość progu może zależeć od tego czy jest on używany do formułowania wiedzy, czyli obliczania rozkładu bazowego prawdopodobieństwa, czy też do wnioskowania diagnostycznego. Oznaczmy ten pierwszy jako η_{RBP} , a drugi jako η_W . Wówczas:

$$m_l(f) = 0, \quad \sum_{\substack{s_i \in S_l \\ \eta_i^* \geq \eta_{RBP}}} m_l(s_i) = 1, \quad (12)$$

$$Bel(d_l) = \sum_{\substack{s_i^* \Rightarrow d_l \\ \eta_i^* \geq \eta_W}} m(s_i^*), \quad (13)$$

$$Pl(d_l) = \sum_{\substack{s_i^+ \Rightarrow d_l \\ \eta_i^+ \geq \eta_W}} m(s_i^+). \quad (14)$$

Podobnie jak w klasycznej teorii Dempstera-Shafera $Bel(d_l) \leq Pl(d_l)$. Miary są zawsze równe gdy wszystkie symptomy są pojedyncze, co w praktycznych problemach diagnostycznych nigdy nie ma miejsca.

2.4. Wyznaczanie diagnozy

Na podstawie miary przekonania i domniemania można określić na ile pewna jest analizowana diagnoza. Ważniejsza jest wartość miary przekonania, ponieważ wnioskowanie medyczne powinno być „ostrożne”, a właśnie ta miara wskazuje pewność diagnozy dla w pełni potwierdzonych objawów. Miara domniemania pozwala stwierdzić, jaką część z możliwej do analizy wiedzy stanowią objawy, o których posiadamy jakąkolwiek informację. Jest ona wobec tego przydatna, gdy dysponujemy wartościami niewielu spośród bardzo licznych parametrów medycznych badanych dla wybranej diagnozy. Ta ostatnia sytuacja często ma miejsce z uwagi na koszty (nie tylko finansowe) badań. Dlatego miara domniemania powinna także być uwzględniona podczas wnioskowania. Wniosek diagnostyczny jest formułowany po porównaniu przynajmniej dwóch hipotez diagnostycznych: choroby i zdrowia. Może być równocześnie analizowanych wiele hipotez, ale oczywiście muszą się one odnosić do tej samej grupy chorób. W celu wyznaczenia ostatecznego wniosku należy porównać wartości przekonania dla różnych hipotez. Ta diagnoza, dla której wartość miary przekonania jest największa, stanowi ostateczną konkluzję diagnostyczną. Jeżeli wartości dla dwóch hipotez są równe, nie istnieje możliwość określenia wniosku końcowego. Analizowane wartości miary przekonania, podobnie jak i wartości miary domniemania zależą od progu η_W i generalnie są tym większe, im mniejszy jest próg. Oczywiście, obniżenie progu dopuszcza do analizy coraz bardziej wątpliwe objawy. Miara domniemania jest bardziej „odporna” na podwyższenie progu, ponieważ zwykle poziom przynależności jednego ze złożonych objawów jest wysoki. Dlatego za wartość miary przekonania, która będzie porównywana z przekonaniem do innych hipotez przyjmuje się tę wartość, która jest wyznaczona dla maksymalnego progu, przy którym wartość domniemania jest jeszcze największa. W ten sposób obie miary biorą udział w wyznaczaniu wniosku końcowego, a użytkownikowi prezentowane są pojedyncze wartości przekonania charakteryzujące hipotezy. Ułatwia to intuicyjną ocenę jakości ostatecznego wniosku. Przedstawione metody modelowania wnioskowania oraz sposób interpretacji wyników obliczeń składają się na algorytm wspomaganie diagnozy, którego działanie zostało sprawdzone na podstawie wybranych baz danych.

3. Bazy danych

Proponowany algorytm został zweryfikowany podczas modelowania wspomaganie diagnozy medycznej w chorobach tarczycy. Wybrano trzy bazy danych: dwie internetowe i jedną własną. Zdefiniowano trzy kategorie diagnostyczne: H – eutyreoza (czyli zadowolający stan gruczołu tarczowego), D_1 – nadczynność tarczycy, D_2 – niedoczynność tarczycy. Wszystkie dane zostały podzielone na zbiory uczące i testowe. Pierwsza baza: <ftp.ics.uci.edu/pub/machine-learning-databases/thyroid-disease>, pliki *new-thyr.**, gromadzi dane dla 5 zmiennych ilościowych. Proporcje podziału przypadków na zbiór uczący/testowy były następujące: H : 75/75, D_1 : 15/20, D_2 : 15/15. W drugiej bazie: <ftp.ics.uci.edu/pub/machine-learning-databases/thyroid-disease>, pliki *ann.**, występują dane dla 6 zmiennych ilościowych i 15 zmiennych jakościowych. Podział przypadków narzuca struktura bazy: H : 191/177, D_1 : 93/73, D_2 : 3488/3178. Wreszcie przygotowano własną bazę danych zawierającą 97 parametrów medycznych rozmaitej natury, np. wiek, poronienia, nieregularne miesiączki, indeks Crooksa (jako suma), pojedyncze objawy z indeksu Crooksa (np. zadyszka wysiłkowa), testy laboratoryjne (np. T_3 , T_4 , TSH). Liczność zbiorów uczących/testowych wynosiła w tym przypadku: H : 26/26, D_1 : 52/16, D_2 : 23/23.

4. Obliczenia

Dla wszystkich baz danych przeprowadzono obliczenia zgodnie z proponowanym algorytmem wspomaganie decyzji. W przypadku danych internetowych reguły wnioskowania były tworzone na podstawie kwalifikacji przypadków do poszczególnych diagnoz zamieszczonych w bazach i obliczeń statystycznych (współczynników korelacji dla danych numerycznych oraz testu χ^2 i miary Q Kendalla [8] dla danych jakościowych). Dla pierwszej bazy danych obejmujących 5 zmiennych numerycznych utworzono 9 reguł: 5 reguł dla pojedynczych zmiennych, 3 dla kombinacji dwuelementowych skorelowanych zmiennych i jedną dla trójki skorelowanych zmiennych. Dla drugiej bazy utworzono 135 reguł zawierających pojedyncze zmienne ilościowe i jakościowe, jak również kombinacje tych zmiennych, łącząc w niektórych regułach zmienne jakościowe z ilościowymi. Reguły dla własnej bazy danych były tworzone przez lekarza. Parametry w niej uwzględnione są znacznie liczniejsze niż liczba przypadków, ale stworzona baza danych jest wciąż uzupełniana, a reguły grupują wiele objawów (np. cały indeks Crooksa, czyli 26 objawów). W przypadku tej bazy chodziło również o stwierdzenie czy reguły sformułowane przez lekarza można efektywnie łączyć z funkcjami przynależności i wartościami bazowego rozkładu prawdopodobieństwa obliczonymi na podstawie danych uczących.

Rozkład bazowego prawdopodobieństwa i funkcje przynależności w przypadku wszystkich baz wyznaczano dla danych uczących. Następnie dla każdego przypadku ze zbioru testowego, stosując proponowany algorytm, wyznaczano diagnozę końcową. Z kolei wyznaczano błąd klasyfikacji jako procent nieprawidłowych kwalifikacji lub braku konkluzji algorytmu w stosunku do diagnoz wskazywanych przez bazy danych, dla każdej grupy testowej z osobna. Obliczenia powtarzano dla progów η_{RBP} i η_W zmieniających z krokiem 0.05 w przedziale [0, 1].

5. Wyniki testowania i dyskusja

Dla pierwszej bazy danych najmniejszy błąd, wynoszący 2.67% uzyskano dla progów $\eta_{RBP} \in [0.17, 0.23]$ oraz $\eta_{RBP} \in [0.86, 0.9]$ i $\eta_W \in [0.09, 0.4]$. W tym przypadku błąd był znacznie mniejszy niż w metodach porównawczych proponowanych przez autorów bazy danych [2].

Autorzy drugiej bazy danych wskazują, że prawidłowy algorytm powinien rozpoznawać poprawnie przynajmniej 92% przypadków. W literaturze [5], [6] można znaleźć propozycje sieci neuronowych i algorytmów genetycznych umożliwiających klasyfikację z błędem mniejszym niż 2% dla danych testowych. Jednakże, jak zauważają autorzy tych opracowań, sieć neuronowa nie wykazywała optymalnej struktury ani właściwej generalizacji [6]. W proponowanym algorytmie, dla danych uczących, błąd określony jako przyporządkowanie przypadku do nieprawidłowej diagnozy wynosił 3.7%. Dla niektórych przypadków wartości miary przekonania były równe dla dwu lub trzech diagnoz. Po zaliczeniu tych przypadków do grupy błędnych kwalifikacji błąd wzrasta do 4.7%. Wartość minimalnego błędu praktycznie nie zależy od progów. Jest to spowodowane specyficznym kształtem funkcji przynależności, otrzymanych dla danych uczących. Mają one bardzo strome zbocza, co upodabnia je do funkcji charakterystycznych. Po procesie uczenia (obejmującym wyznaczenie funkcji przynależności i rozkładu bazowego prawdopodobieństwa), dla danych testowych błędy wynosiły odpowiednio: 3.4% i 4.3%. Obniżenie wartości błędu dla danych testowych wskazuje, że dla lepiej dobranych danych uczących błąd mógłby być mniejszy. Rzeczywiście, zaobserwowano to po wymianie pewnej części przypadków pomiędzy zbiorami uczącymi i testowymi, przy zachowaniu rozdzielności zbiorów. Ponieważ jednak w niniejszej pracy odniesiono się do literaturowych kryteriów oceny automatycznej diagnozy, wyniki te nie będą tutaj omawiane. Należy jednak zauważyć, że algorytm spełnia wstępny warunek prawidłowego rozpoznawania (jak wspomniano, błąd mniejszy od 8%). Wydaje się przy tym, że wielkość błędu wskazuje na właściwą generalizację. W wyniku działania algorytmu uzyskuje się nie tylko kwalifikację przypadków chorobowych, ale również wiedzę w postaci reguł rozumiałych dla lekarza. Umożliwia to gromadzenie wiedzy na podstawie danych.

W diagnostyce chorób tarczycy najważniejszą rolę odgrywają zmienne ilościowe, ponieważ to one (głównie testy laboratoryjne) ostatecznie decydują o konieczności leczenia, na przykład poprzez podanie hormonów. W przypadku ograniczenia wnioskowania tylko do zmiennych ilościowych, dla danych uczących, nieprawidłowe diagnozy algorytm wskazał 11.1%, a nieprawidłowe i nieokreślone w 12.0% przypadków. Podobne obliczenia dla danych testowych wykazały błędy odpowiednio – 8.9% i 9.6%. Wynika z tego, że bez zmiennych jakościowych algorytm nie spełnia założonego dopuszczalnego kryterium błędów, tak więc są one konieczne dla poprawnego wnioskowania.

Dla indywidualnie zgromadzonych danych błąd kwalifikacji zaobserwowano jedynie dla diagnozy D_2 , wynosił on 7.69% i był prawdopodobnie spowodowany bardzo dużymi brakami w danych diagnozowanych pacjentów oraz niejednoznacznością ich odpowiedzi podczas zbierania wywiadu. Zgromadzenie wiarygodnych danych dla tej grupy pacjentów jest szczególnie trudne, ponieważ często charakteryzuje ich spowolnienie myślowe. Błąd ten praktycznie nie zależał od progów i był jednakowo dla η_{RBP} , $\eta_W \in [0.1, 1]$.

6. Wnioski

Algorytm zaproponowany w pracy pozwala na wykorzystanie podczas wnioskowania zarówno informacji ilościowej, jak i jakościowej, zgodnie z heurystycznymi regułami, które są wykorzystywane w diagnozowaniu pacjenta przez lekarza. Reguły te może formułować ekspert, można je także uzyskać na podstawie danych uczących. Reguły wyznaczone w oparciu o dane mogą być krytycznie oceniane przez lekarza i w ten sposób mogą stać się nowym źródłem wiedzy. Mogą też sprzyjać

sprecyzowaniu określeń lingwistycznych, poprzez zdefiniowanie właściwych dla nich funkcji przynależności. Działanie algorytmu sprawdzono dla trzech niezależnych baz danych dotyczących chorób tarczycy. Przykłady obliczeń dla baz internetowych <ftp.ics.uci.edu/pub/machine-learning-databases/thyroid-disease>, wykazały, że algorytmu w stopniu zadowalającym (pliki *ann.**) lub lepszym (pliki *new-thyr.**) od metod referencyjnych spełnia zadanie kwalifikacji pacjentów do odpowiednich diagnoz. Wstępne wyniki dla własnej bazy danych są również zadowalające. Należy również zaznaczyć, że w przeciwieństwie do sieci neuronowej, struktura wiedzy sformułowanej w postaci reguł może być przeniesiona do innej populacji, dla której jedynie miary precyzyjności (czyli funkcje przynależności) oraz pewności (czyli rozkład bazowego prawdopodobieństwa) będą indywidualnie wyznaczone. Zastosowanie progów przy obliczaniu rozkładu bazowego prawdopodobieństwa, jak również podczas wnioskowania pozwala na ustalenie żądanej jakości wiedzy i informacji o konsultowanym pacjencie. Ma ono jednak sens tylko wtedy, gdy większość zmiennych ma charakter ilościowy, a zależne od danych uczących funkcje przynależności nie mają zbyt stromych zboczy. W przeciwnym wypadku próg praktycznie nie ma wpływu na błąd klasyfikacji. Gdy błąd wykazuje zależność od progów, nie powinny być one bardzo niskie, ponieważ nie należy uwzględniać najbardziej wątpliwych symptomów. Z kolei, zbyt wysoka ich wartość eliminuje wartościowe dane wejściowe, które chociaż nie w pełni potwierdzają objawy choroby, jednak powinny być uwzględnione w diagnozie.

Algorytm nie jest skomplikowany numerycznie. Nakład obliczeniowy jest również znacznie mniejszy niż w przypadku sieci neuronowej, zwłaszcza o niestabilnej strukturze. Możliwe jest przedstawienie użytkownikowi rozpatrywanych hipotez diagnostycznych wraz z jednoznaczными miarami przekonania, co znacznie upraszcza interpretację wyników wnioskowania. Dlatego wydaje się, że przedstawione metody mogą stanowić istotny przyczynek do komputerowego wspomaganie diagnozy medycznej. Dotychczas, do wstępnego wyznaczenia reguł w algorytmie były stosowane badania statystyczne, wyznaczające zależności między zmiennymi. Autonomiczny algorytm wyznaczania najbardziej istotnych diagnostycznie związków pomiędzy zmiennymi stanowi odrębny temat planowanych badań.

7. Literatura

- [1] L. Bolc, W. Borodziejewicz, M. Wójcik: Podstawy przetwarzania informacji niepewnej niepełnej; PWN Warszawa, 1991.
- [2] D. Coomans, I. Broeckaert, M. Jonckheer, D. L. Massart: Comparison of multivariate discrimination techniques for clinical data-application to the thyroid functional state, *Meth. Inform. Med.* 22, 1983, 93-101.
- [3] J. Gordon, E. H. Shortliffe: The Dempster-Shafer Theory of Evidence, in: *Rule-Based Expert Systems*. Buchanan B.G., Shortliffe E.H. eds, Addison Wesley 1984, 272-292.
- [4] Górnicki T.: Choroby tarczycy, PZWL, Warszawa 1975.
- [5] W. Schiffman, M. Joost, R. Werner: Synthesis and performance analysis of multilayer neural network architectures, <ftp.archive.cis.ohio-state.edu>, technical report 1992
- [6] W. Schiffman, M. Joost, R. Werner: Optimization of Backpropagation Algorithm for training multilayer perceptrons, <ftp.archive.cis.ohio-state.edu>, technical report 1994
- [7] E. Straszeka: An interpretation of focal elements as fuzzy sets, *Int. J. of Intelligent Systems* 18, 2003, 821-835.
- [8] R. Tadeusiewicz, A. Izworski, J. Majewski: *Biometria*, Wydawnictwa AGH, Kraków 1993.