Centrum Rzeczoznawstwa Budowlanego

Open Access Journal
Modern Engineering 1 (2017) 11-16
journal homepage: http://mengineering.eu

MODERN ENGINEERING

Badanie wiarygodności procedur detekcji zagrożeń epidemiologicznych
Reliability Assessment of Epidemiological Detection Procedures

Andrzej AMELJAŃCZYK * [1]

[1]Wojskowa Akademia Techniczna, Warszawa, Polska

Streszczenie

Praca jest kontynuacją pracy „Modelowanie i optymalizacja procedur detekcji zagrożeń epidemiologicznych". Przydatność modelu detekcji zagrożeń epidemiologicznych i zastosowanej metody detekcyjnej w praktyce w dużym stopniu zależy od wiarygodności uzyskanych wyników. W pracy zdefiniowano charakterystyki jakościowe komputerowej procedury detekcyjnej takie jak : informacyjność wyników badań sanitarnych, wyrazistość i niejednoznaczność rozpoznania oraz kompleksowy wskaźnik wiarygodności procedury detekcyjnej. Głównym wynikiem realizacji algorytmu ustalenia wstępnego rozpoznania zagrożenia jest zbiór zagrożeń od których nie ma bardziej prawdopodobnych oraz jego ranking. Wiarygodność procedury detekcyjnej a zatem i jej wyniku, jest funkcją wielu czynników między innymi takich jak „zawartość" zbioru M (wzorców zagrożeń) oraz wiarygodności danych o symptomach i czynnikach ryzyka zagrożenia w rejonie.

Słowa kluczowe: informacyjność wyników, wyrazistość i niejednoznaczność rozpoznania wiarygodność procedury detekcyjnej

Abstract

The work is a continuation of the work "Modeling and optimization of procedures for detecting epidemiological threats". The usefulness of the epidemiological threat detection model and the detection method used in practice largely depends on the reliability of the results obtained. The work defines the qualitative characteristics of the computer detection procedure, such as: information on the results of sanitary tests, clarity and ambiguity of the diagnosis and a comprehensive indicator of the reliability of the detection procedure. The main result of the implementation of the algorithm for the preliminary identification of a threat is the set of threats from which there are no more probable ones and its ranking. The credibility of the detection procedure and, hence, its result, is a function of many factors, including such as the "content" of the set (hazard patterns) and the reliability of data on symptoms and risk factors in the region.

Keywords: informative results, clarity and ambiguity of recognition, reliability of the detection procedure

1 Wprowadzenie

Jakość uzyskanego rozpoznania wstępnego rozumianego jako jego przydatność detekcyjna zależy od bardzo wielu czynników. W pracy przedstawiono główne aspekty decydujące o jakości generowanych przez algorytm wstępnego rozpoznania zagrożeń. Zbadano między innymi takie charakterystyki jakościowe jak : informacyjność wyników badan sanitarnych, wyrazistość rozpoznania zagrożenia, niejednoznaczność rozpoznania oraz kompleksowy wskaźnik jakościowy zwany wiarygodnością rozpoznania zagrożenia. Dodatkowym rezultatem prezentowanej pracy jest symulacyjna metoda badania charakterystyk jakościowych procedury detekcyjnej opartej na rozpoznawaniu wzorców zagrożeń epidemiologicznych. Metoda ta pozwala na generowanie danych symulacyjnych dotyczących symptomów zagrożeń, aktualnych czynników ryzyka jak również całego repozytorium wzorców zagrożeń epidemiologicznych w dowolnych scenariuszach sytuacyjnych. Dla każdego przebiegu detekcyjnego określane są wartości poszczególnych charakterystyk jakościowych , świadczących o praktycznej przydatności uzyskanego rozpoznania zagrożenia epidemiologicznego. Wartości tych charakterystyk niosą dodatkową informację analityczną możliwą do wykorzystania przez organ nadzoru sanitarnego. Duża podatność symulacyjna algorytmu wynikająca z prostoty generowania danych symulacyjnych przy

*Corresponding author: E-mail address: andrzej.ameljanczyk@wat.edu.pl (Andrzej AMELJAŃCZYK)

Received 13 February 2017
Available online 16 September 2017
ISSN 2450-5501
Published by Centrum Rzeczoznawstwa Budowlanego

różnych scenariuszach umożliwia łatwe badanie poszczególnych charakterystyk jakościowych oraz określanie ich wartości progowych skutkujących odrzuceniem bądź przyjęciem rozpoznania o faktycznym zagrożeniu epidemiologicznym. Z punktu widzenia istotności wspomagania komputerowego w procesie wsparcia procedury detekcyjnej szczególnie ważna jest informacja o stopniu wiarygodności otrzymanych wyników , gdyż to decyduje o możliwości ich wykorzystania.

Z punktu widzenia przydatności wspomagania komputerowego w procesie detekcji zagrożeń szczególnie ważna jest informacja o stopniu wiarygodności otrzymanych wyników , gdyż to decyduje o możliwości ich wykorzystania. [1–4, 8, 10–16, 18–22]

2 Wiarygodność procedury detekcyjnej

Algorytm detekcyjny jest tak skonstruowany, że niezależnie od „jakości informacyjnej" uzyskanych wyników z wywiadów, badań laboratoryjnych jak również niezależnie od zawartości repozytorium (zbiór wzorców zagrożeń) prowadzi do wyniku w postaci zbioru Pareto i jego rankingu. Jeśli jednak w zbiorze M nie ma wzorca występującego w praktyce zagrożenia to oczywiście nie jest możliwe jego wyznaczenie jako rozpoznania zagrożenia.

Obok więc samego rozpoznania ważna jest informacja o stopniu wiarygodności całej procedury jak również „istotności informacyjnej" uzyskanych wyników badań sanitarnych. Taką informację i wiele innych charakterystyk można uzyskać analizując otrzymane wyniki wspomagania komputerowego. Badanie testowe komputerowej aplikacji algorytmu [6] w pełni potwierdziły taką możliwość. Odpowiednie „wykalibrowanie" modelu i algorytmu pozwalają kwalifikować uzyskane wyniki w oparciu o np. stopnie wiarygodności rozpoznania. Położenie podzbioru obrazów odległościowych zbioru rozpoznań wstępnego oszacowania d (M_0(x)) bardzo zależy od jakości informacyjnej wyników badań (na ile są one „wyraziste") i oczywiście od liczności zbioru M zawierającego wzorce zagrożeń epidemiologicznych (od zawartość repozytorium wzorców). Elementem charakteryzującym położenie tego zbioru jest tzw. punkt idealny $y(\dot{x})$ który jest kresem dolnym podzbioru d (M_0(x)) przy przyjętym modelu relacji podobieństwa R (model Pareto) [9, 17].

Jego odległość od obrazu wzorca wirtualnego zagrożenia najbardziej prawdopodobnego (czyli od punktu (0,0)) jest miarą wiarygodności uzyskanego wyniku. Symbolem *w(x, M)* oznaczymy stopień wiarygodności uzyskanego rozpoznania w oparciu o wyniki badań *f(x)* [6], w rejonie $x \in X$ oraz liczność repozytorium wzorców zagrożeń, reprezentowanych przez zbiór *M*.

$$w(x, M) = 1 - q\left((0,0)\dot{y}(x)\right) = 1 - \|\dot{y}(x)\|_p, p \geq 1 \tag{1}$$

Rola parametru p w wyznaczaniu konkretnej postaci normy (3.1) jest omówiona dokładnie w [9, 17].W praktyce stosujemy najczęściej *p=2* , otrzymując zwykłą geometryczną odległość.

Na rys.1 pokazane są dwie sytuacje detekcyjne otrzymane w wyniku działania algorytmu przy tym samym repozytorium M a różnych zestawach wyników badań w różnych rejonach (rejon $x \in X$ oraz rejon $\overline{x} \in X$). Detekcyjna przestrzeń podobieństwa, w której badane jest położenie obrazów zagrożeń uzyskanych w wyniku badań w rejonach jest znormalizowana do [0,1] x [0,1]. W zależności od wyników badań (bardziej „wyraziste" lub „mniej wyraziste") dla poszczególnych rejonów otrzymuje się odpowiednie „skupienia" punktów. Tworzą one pewne charakterystyczne ,wydzielone podzbiory (chmury punktów).

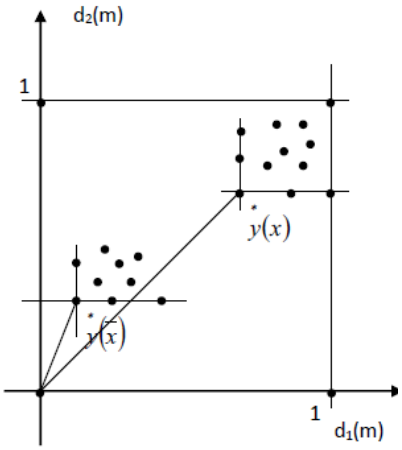
Odległość $q((0,0)\dot{y}(x))$ w przypadku rejonu $x \in X$ jest znacznie większa od odległości (liczby) $q((0,0)\dot{y}(\overline{x}))$ przy tej samej zawartości repozytorium wzorców zagrożeń epidemiologicznych. Liczby te charakteryzują wiarygodnośćrozpoznań zagrożeń w poszczególnych rejonach przy tym samym repozytorium wzorców zagrożeń. Mamy zatem sytuację:

$$q((0,0), \dot{y}(x)) >> q((0,0,\dot{y}(\overline{x}))) \rightarrow w(\overline{x}, M) >> w(x, M) \tag{2}$$

A to oznacza, że uzyskane rozpoznanie dotyczące rejonu x jest znacznie mniej wiarygodne. Wyniki badań w rejonie $\overline{x} \in X$ posiadają zatem znaczenie większą „informacyjność detekcyjną" (patrz (18),(19)) niż wyniki badań uzyskane w rejonie $x \in X$.

3 Informacyjność wyników badań

Stopień informacyjności wyników badań (w aspekcie symptomów i czynników ryzyka) dla poszczególnych zagrożeń można określić następującymi liczbami [6, 7]



Rysunek 1. Detekcyjna przestrzeń podobieństwa

$$\begin{aligned} i_S(x) &= \sum_{s_k^m \in s_o^m(x)} w\left(x, s_k^m\right) \alpha(s_k^m) \\ i_R(x) &= \sum_{r_l^m \in R_o^m(x)} w\left(x, r_l^m\right) \beta(s_l^m) \end{aligned} \tag{3}$$

Możemy zatem napisać iż stopień wiarygodności rozpoznania zagrożenia jest funkcją zawartości repozytorium M (dokładnie liczności tego zbioru) oraz informacyjności $i_s(x)$ oraz $i_r(x)$

$$w(x, M) = f(i_s(x), i_r(x), |M|) \tag{4}$$

4 Niejednoznaczność i wyrazistość rozpoznania

Istnieje wiele innych charakterystyk omawianych mechanizmów wnioskowania detekcyjnego, mających duże znaczenie informacyjne dla inspektora sanitarnego. Wynikają one głównie z analizy własności zbioru Pareto otrzymanych wyników detekcyjnych . Są to między innymi :

- wskaźnik niejednoznaczności rozpoznania zagrożenia epidemiologicznego $n(x)$ rejonie $x \epsilon X$

$$n(x) = \frac{\max_{y \in Y_N^R} \left(\left|F^{-1}(y)\right|\right)}{\left|F^{-1}(Y_N^R)\right|} \tag{5}$$

- wyrazistość rozpoznania $s(x)$

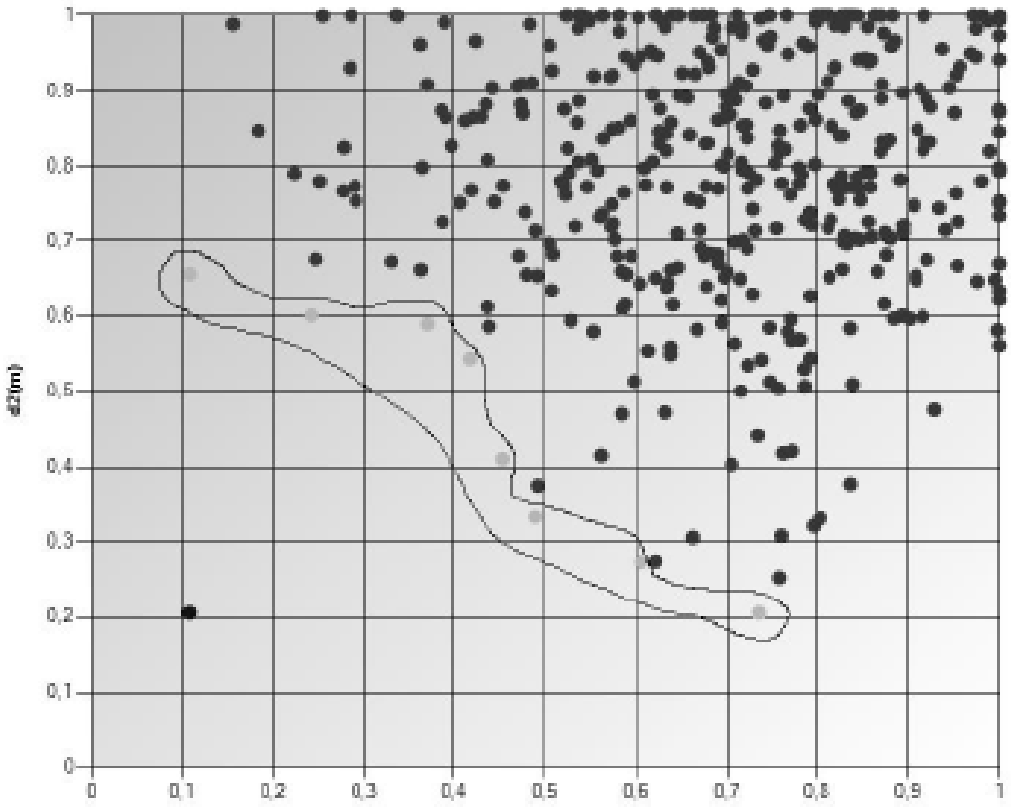
$$s(x) = 1 - \frac{\left|Y_N^R\right|}{|M|} \tag{6}$$

Wszystkie te charakterystyki łącznie ze znajomością zawartości zbioru rozpoznań od których nie ma bardziej prawdopodobnych oraz jego rankingu tworzą zestaw informacji wspomagających proces rozpoznania wstępnego. Dokładna analiza wyników procedury detekcyjnej a przede wszystkim wartości parametrów jakościowych zrealizowanej procedury są podstawą decyzji sanitarnych co do dalszych czynności związanych z zagrożeniem epidemiologicznym.

5 Symulacyjne badania jakości procedur detekcyjnych

Opracowany algorytm wyznaczania rozpoznania wstępnego jak i jego najważniejsze własności zostały poddane badaniom symulacyjnym z wykorzystaniem zaprojektowanego symulatora [7]. Aplikacja symulatora została napisana w oparciu o technologie Microsoft .NET Framework. W procesie implementacji został zastosowany język C. Środowisko projektowe, którego użyto do wykonania oprogramowania to Microsoft Visual Studio 2010. Repozytorium symulatora jest generowane w sposób dynamiczny (przy określeniu odpowiednich parametrów podanych przez użytkownika). Istnieje także możliwość wykorzystania bazy danych MSSQL w celu gromadzenia danych dotyczących zagrożeń i wykorzystywania ich w procesie symulacyjnym. Przedstawione oprogramowanie ma postać aplikacji okienkowej wykorzystującej graficzny interfejs Windows Forms. W celu skorzystania z aplikacji wymaga się uprzedniego zainstalowania .Net Framework w wersji 3 lub wyższej. Wyniki poszczególnych symulacji obrazowane są dzięki wykorzystaniu dodatkowej biblioteki Microsoft Charts, pozwalającej na dynamiczne generowanie różnych typów wykresów. W ramach każdego przebiegu symulacyjnego możliwe jest badanie :

- zawartości zbioru Pareto - wiarygodności rozpoznania - wyrazistości rozpoznania - niejednoznaczności rozpoznania Po zakończeniu symulacji generowane są wykresy przedstawiające przestrzeń detekcji podobieństwa a w szczególności zbiór zagrożeń od których nie ma bardziej prawdopodobnych. Na rys.2 przedstawiony jest typowy wykres komputerowy sytuacji detekcyjnej zagrożeń epidemiologicznych w pewnym rejonie x uzyskany w wyniku symulacji na symulatorze opisanym w pracy [4].Obraz jest uzyskany dla pewnego (symulowanego) zestawu wartości stwierdzonych symptomów zagrożeń epidemiologicznych oraz występujących (symulowanych) w tym rejonie czynników ryzyka w określonym momencie czasu. Wprowadzając w kolejnych momentach czasu (na przykład w kolejnych dniach) odpowiednie dane dotyczące symptomów zagrożenia jak i odpowiednich zmian w czynnikach ryzyka otrzymujemy kolejne „obrazy" detekcyjne zagrożenia epidemiologicznego w tym rejonie. Każde zagrożenie (a dokładnie obraz jego podobieństwa do stanu sanitarnego w rejonie x) przedstawiona jest na wykresie jako punkt o ciemnym zabarwieniu. „Zagrożenia od których nie ma bardziej prawdopodobnych" (tzw. front Pareto") oznaczane są jasnym odcieniem.



Rysunek 2. Typowy obraz komputerowy przestrzeni detekcyjnej oraz zbioru $M_N{}^{\overline{R}}$ zagrożeń, od których nie ma bardziej prawdopodobnych

Wyznaczony przez algorytm punkt utopijny (punkt idealny) [5, 17] (znajdujący się w lewym, dolnym rogu wykresu) to obraz „idealnego podobieństwa" stanu sanitarnego rejonu do „idealnie pasującego" zagrożenia utopijnego" , którego niestety nie ma w repozytorium. Użytkownik ma możliwość zdefiniowania, na jakich etapach symulacji generowane są wykresy z wynikami (ile ma ich być). Istnieje także możliwość zapisu otrzymanych wyników do pliku tekstowego. Numer (nazwę) zagrożenia epidemiologicznego reprezentowanego na wykresie przez określony punkt można uzyskać klikając kursorem.

6 Wnioski

Problem wiarygodności procedur detekcyjnych jest jak widać niezwykle ważny. Profesjonalny System Komputerowego Wsparcia Procedur Detekcyjnych powinien generowane propozycje decyzji opatrzyć wyczerpującą informacją dotyczącą jakości przedstawionych decyzji oraz informacyjności wprowadzanych danych o symptomach i czynnikach ryzyka. Interpretacja komputerowego rozpoznania zagrożenia epidemiologicznego bez pełnego opisu jakościowego zrealizowanej procedury nie ma w praktyce większego sensu. Uzyskane informacje dodatkowe mogą mieć dla inspektora sanitarnego niekiedy nawet większą wartość niż sama propozycja rozpoznania, uczulając go na „pułapki" detekcyjne w które łatwo wpaść ulegając powierzchownym sugestiom. Stąd też wydaje się niezwykle ważne by obok algorytmów wnioskowania , Komputerowy System Wsparcia Detekcyjnego wyposażony był w podsystem generujący stosowne charakterystyki jakościowe przeprowadzonej procedury detekcyjnej. Ważnym elementem użytkowym takiego podsystemu jest oczywiście ustalenie wszelkich wartości progowych (decyzyjnych) przedstawianych wskaźników jakościowych możliwych do uzyskania w trakcie procesu testowania i kalibracji. Progi te byłyby podstawą suwerennych decyzji inspektora w zakresie stopnia wykorzystania przedstawionych propozycji detekcyjnych jak również ustalania kolejnych działań sanitarnych. System nie ma więc wyręczać czy zastępować inspektora sanitarnego w procesie wykrywania i identyfikacji zagrożeń epidemiologicznych. System ma być nowoczesnym ,wiarygodnym narzędziem informatycznym wspierającym pracę inspektora. Komputerowe rozpoznanie zagrożenia (np. w postaci frontu Pareto) łącznie z jego graficzną interpretacją na ekranie komputera oraz dodatkowym opisem dotyczącym jego wiarygodności i jakości powinny pełnić w pracy inspektora sanitarnego podobną rolę jak w pracy lekarza diagnosty pełni wykres EKG , zdjęcie rentgenowskie, obraz USG czy też obraz tomograficzny, które to tylko czasami dają lekarzowi pełne przekonanie diagnostyczne a czasami jedynie sugestię lub nawet pogłębiają jego wątpliwości. Uzyskana komputerowa propozycja rozpoznania zagrożenia może więc być pewną sugestią detekcyjną, sugestią traktowaną tym bardziej poważnie im towarzyszące jej informacje dotyczące wiarygodności zastosowanej procedury, informacyjności wykorzystanych wyników badań , niejednoznaczności rozpoznania, jego wyrazistości czy też „dyspersji" frontu Pareto będą na odpowiednio wysokim poziomie. Wyposażenie wszystkich placówek nadzoru sanitarnego w kraju w podobny system komputerowego wsparcia detekcyjnego z możliwością systematycznego (okresowego) wprowadzania danych o symptomach i pojawiających się czynnikach ryzyka zagrożenia epidemiologicznego dawałyby możliwość obserwowania wszystkich parametrów detekcyjnych jak i stopni zagrożeń rejonów w czasie. Odpowiednie zestawianie „szeregów czasowych" przedstawiających dane o zagrożeniach w różnych rejonach dawałyby niewątpliwie nowe możliwości w zakresie predykcji zagrożeń i ich rozprzestrzeniania się.

Bibliografia

1. Allan, M. & Spencer, J. *Crash Course - wywiad i badanie przedmiotowe* (red. Kokot, F.) (Elsevier Urban & Partner, 2005).
2. Ameliańczyk, A. Matematyczne aspekty modelowania pajęczynowego obiektów. *Biuletyn Instytutu Systemów Informatycznych* **4,** 9–17 (2009).
3. Ameliańczyk, A. Wielokryterialne mechanizmy wspomagania podejmowania decyzji medycznych w modelu repozytorium w oparciu o wzorce. *Biuletyn Instytutu Systemów Informatycznych* **5,** 1–6 (2010).
4. Ameljańczyk, A., Długosz, P. & Strawa, M. *Komputerowa implementacja algorytmu wyznaczania wstępnej diagnozy medycznej* w *VII Konferencja Naukowa Modelowanie Cybernetyczne Systemów Biologicznych* (Kraków, 2010).
5. Ameljańczyk, A. *Optymalizacja wielokryterialna w problemach sterowania i zarządzania* (Zakład Narodowy im. Ossolińskich, Wrocław, 1984).
6. Ameljańczyk, A. *Matematyczny model przestrzeni życia w komputerowym systemie wspomagania decyzji medycznych* w *I Krajowa Konferencja Systemy Komputerowe i Teleinformatyczne w Służbie Zdrowia* (Warszawa, 2009).

7. Ameljańczyk, A. *Matematyczny model wspomagania diagnozowania medycznego na podstawie symptomów chorobowych i czynników ryzyka* w *VII Konferencja Naukowa Modelowanie Cybernetyczne Systemów Biologicznych* (Kraków, 2010).
8. Anderson, R. & May, R. Directly transmitted infectious diseases: control by vaccination. *Science* **215,** 1053–1060 (1982).
9. Anderson, R. & May, R. *Infectious Diseases of Humans, Dynamics and Control* ISBN: 019854040X (Oxford University Press, 1992).
10. Beynon, H., van den Bogaerde, J. & Davies, K. *Interpretacja danych klinicznych w pytaniach i opisach przypadków* (red. Kokot, F.) (Elsevier Urban & Partner, 2007).
11. Douglas, R. & Collins. *Algorytmy interpretacji objawów klinicznych* ISBN: 978-83-61104-21-6 (Medipage, Warszawa, 2010).
12. Kermack, W. & McKendrick, A. A Contributions to the mathematical theory of epidemics. *Procceedings of the Royal Society A* **115** (772 1927).
13. Kokot, F. & Kokot, S. *Badania laboratoryjne – zakres norm i interpretacja* ISBN: 83-200-2711-X (PZWL, Warszawa, 2002).
14. *Diagnostyka różnicowa objawów chorobowych* (red. Kokot, F.) (Państwowy Zakład Wydawców Lekarskich, 1990).
15. Murray, J. *Mathematical Biology II: Spatial Models and Biomedical Applications* ISBN: 0-387-95228-4 (Springer-Verlag Berlin Heidelberg, 2003).
16. Pawlak, Z. Rough sets. *International Journal of Computer and Information Sciences* **11,** 341–356 (1965).
17. *The Merck Manual. Objawy kliniczne. Praktyczny przewodnik diagnostyki i terapii* (red. Porter, R., Kaplan, J. & B.P., H.) ISBN: 978-83-7609-173-0 (Elsevier Urban & Partner, 2010).
18. Sanchez, K. Medical diagnosis and composite fuzzy relations. *Advances in fuzzy sets theory and applications* (1979).
19. Siegenthalerz, W. *Rozpoznanie różnicowe w medycynie wewnętrznej: Tom 1 - 2* ISBN: 978-83-61104-09-4 (Medipage, Warszawa, 2009).
20. Smets, P. Medical diagnosis fuzzy sets and degrees of belief. *Fuzzy sets and Systems* **5** (1981).
21. Wechsler, H. Applications of fuzzy logic to medical diagnosis. *Proc. Symp. on Multiple* (1975).
22. Yu, P. & Leitmann, G. Compromise solutions, domination structures and Salukwadześ solution. *JOTA* **13** (1974).