

Prognozowanie kondycji ekonomiczno – finansowej przedsiębiorstw z wykorzystaniem sieci przekonań Bayesa

B. Kuczmowska¹

1. Wprowadzenie

W dzisiejszym, dynamicznym środowisku gospodarczym wszelkiego rodzaju oceny, analizy i prognozy stają się coraz bardziej skomplikowane i wymagają do swojego opisu zaawansowanych metod i technik. Stąd też w ostatnich latach pojawiły się propozycje zastosowania do oceny i prognozowania kondycji finansowej przedsiębiorstw sieci przekonań Bayesa.

W 2001 roku S. Sarkar i R. S. Sriram [9] opracowali modele sieci przekonań dla wczesnego ostrzegania o upadłościach banków. Stwierdzili, że zarówno naiwny jak i złożony model sieci Bayesa daje wyniki porównywalne z dobrze znanym algorytmem klasyfikacji drzewa wymuszonych decyzji (Induced Tree Classification).

W 2004 roku L. Sun, P.P. Shenoy [10] wykorzystali naiwny kaskadowy model sieci Bayesa w procesie prognozowania upadłości firm. Do budowy modelu wykorzystali dane z 890 firm upadłych zapisane w bazach danych Compustat Reserch i Lewis-Nexis Bankruptcy Report i 7000 jednostek nie zagrożonych upadłością z różnych branż w okresie 1989-2002. Informacje finansowe dla upadających firm pochodziły z najaktualniejszych dostępnych sprawozdań rocznych złożonych przed wszczęciem postępowania upadłościowego. Informacje finansowe dla firm nie zagrożonych upadkiem pochodziły z końca ostatniego roku fiskalnego przed rokiem badanym. Do wyboru zmiennych objaśniających z puli potencjalnych zmiennych zaproponowali metodę heurystyczną wykorzystującą informacje o korelacjach i częściowych korelacjach. Wyliczyli korelacje pomiędzy wszystkimi zmiennymi objaśniającymi i zmienną decyzyjną (stan upadłości). Zmienne, dla których wartość bezwzględna współczynnika korelacji Person'a jest większa od 0,20 zostały włączone do modelu jako zmienne objaśniające. Do wyliczenia współczynnika korelacji wykorzystana została cała próba firm (7000 nie zagrożonych upadłością i 890 upadłych). Dokładność klasyfikacji zbudowanego modelu wahała się w granicach 76,65% - 81,24%.

2. Metodologia badań

2.1. Podstawy bayesowskich sieci przekonań

Sieci bayesowskie to jedna z metod reprezentowania i odkrywania wiedzy oparta na rachunku prawdopodobieństwa.

Prawdopodobieństwo bezwarunkowe (a priori) określa liczbowo szansę wystąpienia jakiegoś zjawiska, gdy nie są znane żadne okoliczności związane z tym zjawiskiem. Prawdopodobieństwo warunkowe (a posteriori) $P(A|B)$ jest to prawdopodobieństwo zdarzenia A obliczone tylko w sytuacjach, w których zaszło zdarzenie B. Prawdopodobieństwo wyraża się wzorem:

$$(1) \quad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Prawdopodobieństwem warunkowym musimy posługiwać się zawsze, kiedy chcemy wyliczyć prawdopodobieństwo jakiegoś zdarzenia i posiadamy wiedzę o innych zdarzeniach zależnych. $P(A)$ jest poprawnym prawdopodobieństwem zdarzenia A o ile nie posiadamy żadnej wiedzy. Jeżeli wiemy że zachodzi B, to poprawnym prawdopodobieństwem zdarzenia A jest $P(A|B)$, jeżeli wiemy że zachodzi jeszcze C to poprawnym prawdopodobieństwem jest $P(A|B \cap C)$.

Z dwukrotnego zastosowania wzoru (1) można uzyskać prostą zależność zwaną regułą Bayesa.

$$(2) \quad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

W kontekście przewidywania kondycji ekonomiczno-finansowej przedsiębiorstwa można to zinterpretować następująco. Załóżmy, że interesuje nas zdarzenie A, które reprezentuje złą kondycję finansową przedsiębiorstwa. Zaczynamy od prawdopodobieństwa a priori $P(A)$ reprezentującego nasze przekonanie co do A, przed zaobserwowaniem jakichkolwiek stosownych dowodów. Dla przykładu, $P(A)$ może być mierzone jako średni procent firm w całej populacji, które wykazują złą kondycję finansową w ostatnich latach. $P(B)$ niech będzie prawdopodobieństwem, że wskaźnik rentowności sprzedaży brutto jest mały, $P(B|A)$ przedstawia prawdopodobieństwo, że firmy o złej kondycji finansowej mają mały wskaźnik rentowności sprzedaży brutto. Załóżmy, że następnie obserwujemy zdarzenie B. Jeżeli mamy obserwację zdarzenia B to prawdopodobieństwo $P(A|B)$ jest otrzymane z reguły Bayesa: tj z pomnożenia $P(A)$ przez iloraz $P(B|A)/P(B)$. Przy większej liczbie zmiennych objaśniających, zależności pomiędzy nimi można przedstawić w postaci grafu – sieci przekonań. Do wyliczenia prawdopodobieństw warunkowych może posłużyć zbiór uczący (baza informacyjna) zawierająca wszystkie obserwacje zdarzeń (przypadki).

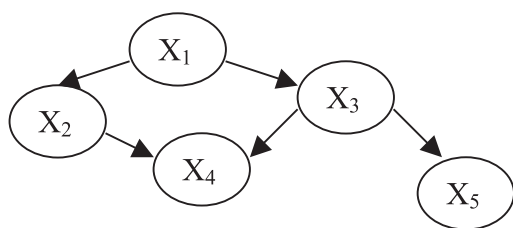
Bayesowska sieć przekonań [1] jest acyklicznym grafem skierowanym, złożonym z węzłów reprezentujących zmienne objaśniające (atrybuty bazy danych) i łączących je krawędzie. Krawędzie określają związki przyczynowo – skutkowe pomiędzy węzłami. Węzeł X_j jest bezpośrednim następnikiem węzła X_i jeżeli istnieje krawędź skierowana od węzła X_i do węzła X_j , a węzeł X_i jest poprzednikiem węzła X_j . Węzeł który nie ma poprzedni-

¹ Wyższa Szkoła Zarządzania i Administracji w Zamościu, Katedra Informatyki i Inżynierii Wiedzy.

ków jest korzeniem sieci. Z każdym węzłem związana jest tablica prawdopodobieństw warunkowych, dla wszystkich kombinacji wejściowych jego bezpośrednich poprzedników. Przy założeniu że węzły są zależne tylko od swoich bezpośrednich poprzedników łączny rozkład prawdopodobieństwa wyraża się wzorem.

$$(3) \quad P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \Pi(X_i))$$

gdzie $\Pi(X_i)$ oznacza zbiór zmiennych objaśniających (atrybutów), które są bezpośrednimi poprzednikami (rodzicami) wierzchołka X_i



Rysunek 1 przedstawia prostą sieć Bayesa o łącznym rozkładzie prawdopodobieństwa $P(X_1, X_2, X_3, X_4, X_5)$. Zależności zdefiniowane w sieci wyrażają ten rozkład poprzez prawdopodobieństwa warunkowe w następujący sposób:

$$(4) \quad \begin{aligned} P(X_1, X_2, X_3, X_4, X_5) &= \\ &= P(X_1) * P(X_2 | X_1) * P(X_3 | X_1) * P(X_4 | X_2, X_3) * P(X_5 | X_3) \end{aligned}$$

Warunkiem użyteczności sieci jest poprawna struktura i poprawne wartości przechowywane w tablicach prawdopodobieństwa. Struktura sieci może być zbudowana przez eksperta z danej dziedziny lub automatycznie na podstawie danych (zbiór uczący). Możliwość automatycznego konstruowania sieci Bayesa zwiększa szanse na ich zastosowanie w wielu dziedzinach. Wyróżnić można cztery warianty uczenia sieci zależne od posiadanej wiedzy:

1. nieznana struktura sieci i pełna obserwowalność atrybutów zbioru uczącego,
2. znana struktura sieci i pełna obserwowalność atrybutów zbioru uczącego,
3. znana struktura sieci i częściowa obserwowalność,
4. nieznana struktura sieci i częściowa obserwowalność.

W praktyce najczęściej występuje wariant 1. W tym przypadku zakłada się że wszystkie struktury sieci są jednakowo prawdopodobne. Należy przeszukać przestrzeń możliwych struktur w poszukiwaniu takiej, która będzie najbardziej zgodna z danymi uczącymi. Do konstrukcji sieci można wykorzystywać różne algorytmy: SGS, Pearl, PC, K2, FCI. [1,11]

2.2. Konstrukcja sieci przekonań w zastosowanym programie

Do przeprowadzenia badań wybrano program Belief-SEEKER [2] opracowany w Katedrze Systemów Ekspertowych i Sztucznej Inteligencji Wyższej Szkoły Informatyki i Zarządzania w Rzeszowie. W programie tym do konstrukcji sieci zastosowano algorytm K2 [2,11]. Do uzyskania największej zgodności zastosowano maksymalizację miary jakości (funkcji dopasowania) - parametru informującego o zależnościach pomiędzy węzłami sieci.

Parametr ten wyraża się następującą zależnością:

$$(5) \quad ML = \prod_{i=1}^v \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + n_{ij})} \prod_{k=1}^{c_i} \frac{\Gamma(\alpha_{ijk} + n_{ijk})}{\Gamma(\alpha_{ijk})}$$

w której:

$i=1, \dots, v$, gdzie v jest liczbą atrybutów sieci Bayes'a,
 $j=1, \dots, q_i$, gdzie q_i jest liczbą możliwych konfiguracji rodziców atrybutu X_i , (jeżeli dany atrybut nie ma rodziców to q_i przyjmuje wartość 1),

$k=1, \dots, c_i$ gdzie c_i jest liczbą kategorii atrybutu X_i ,

n_{ijk} - jest liczbą takich wierszy w bazie danych, których rodzice atrybutu X_i przyjmują wartość j , a ten atrybut przyjmuje wartość k ,

a_{ijk}, α_{ij} - są parametrami rozkładu Dirichlet'a [3], G - jest funkcją, której wartość dla liczb naturalnych można obliczyć ze wzoru:

$$\Gamma(n) = (n-1)!$$

Do konstrukcji sieci można wykorzystywać różne algorytmy. Jednym z nich jest heurystyczny algorytm K2. Poszukiwanie optymalnego rozwiązania – najlepszego modelu uczenia w postaci sieci przekonań – rozpoczyna się od ustalenia wartości miary jakości dla sieci bez połączeń, a kończy się w momencie, gdy miara jakości osiągnie maksimum. Oznacza to, że w kolejnych krokach budowy modelu uczenia, łuki łączące węzły są tworzone wyłącznie wtedy, gdy ich dodanie maksymalizuje funkcję dopasowania.

W pierwszym kroku testowana jest możliwość utworzenia łuku od pierwszego węzła w kierunku węzłów występujących po nim w bazie informacyjnej. Wybierany jest taki łuk, którego dodanie spowoduje maksymalne zwiększenie wartości miary jakości. Jeżeli dodanie żadnego z łuków nie spowodowałoby zwiększenia wartości funkcji dopasowania, to nastąpi przejście do drugiego kroku, w którym w procesie tworzenia sieci brałby udział kolejny węzeł. Następnie testowana jest możliwość dodania do tej fragmentarycznej struktury kolejnego łuku łączącego pierwszy węzeł z jednym z pozostałych węzłów. Postępując analogicznie, pierwszy krok algorytmu kończy się w momencie, gdy wszystkie możliwości zostaną wyczerpane lub gdy dodanie żadnego z pozostałych łuków nie zwiększa wartości funkcji dopasowania. Algorytm kończy działanie po wykonaniu opi-

sanej procedury kolejno dla każdego z węzłów. Sieć można wygenerować dla różnych wartości parametru Dirichlet'a [3,11] występującego we wzorze 5.

Po wygenerowaniu sieci należy wyliczyć wartości w tablicach prawdopodobieństwa dla każdego węzła sieci. Do ich obliczenia konieczna jest znajomość rozkładu prawdopodobieństwa a priori oraz rozkładu prawdopodobieństwa dla sieci bez połączeń.

Rozkład prawdopodobieństwa a priori wylicza się według wzoru;

$$(6) \quad \pi = \frac{1}{c_i}$$

gdzie c_i – jest liczbą kategorii (stanów) danego atrybutu,

Rozkład prawdopodobieństwa dla sieci bez połączeń wylicza się według wzoru:

$$(7) \quad P(X_i) = \frac{\pi \cdot \alpha + x_i}{\alpha + n}$$

gdzie:

- π - prawdopodobieństwo a priori atrybutu X_i ,
- α - jest przyjętą z góry wartością (najczęściej 1), im większa jest to liczba tym rozkład prawdopodobieństwa jest bardziej zbliżony do równomiernego,
- x_i - liczba wystąpień instancji atrybutu X_i ,
- n - liczba wszystkich przypadków w bazie

Rozkłady prawdopodobieństwa wyrażające zależność przyczynowo-skutkową, czyli rozkłady prawdopodobieństwa węzłów posiadających rodziców wylicza się według wzoru:

$$(8) \quad P(X_i) = \frac{\pi \cdot \alpha_i + n_i}{\alpha_i + n}$$

gdzie:

- π - prawdopodobieństwo a priori atrybutu X_i ,
- $\alpha_i = \frac{\alpha}{q_i}$ gdzie q_i to liczba możliwych konfiguracji wartości rodziców,
- n_i – liczba takich przypadków w bazie, że dana kombinacja wartości atrybutów występuje i przyjmuje oczekiwaną wartość,
- n – liczba wszystkich wystąpień oczekiwanych wartości rodzica

Ostatnim etapem działania jest obliczenie wartości rozkładu brzegowego. W tym celu korzysta się ze wzoru na prawdopodobieństwo całkowite sieci (3) i przeprowadza obliczenia dla ustalonej wartości atrybutu.

Program BeliefSEEKER umożliwia przekształcenie wygenerowanego modelu uczenia – w postaci tradycyjnej sieci przekonań – w zbiór reguł składniowych typu JEŻE-LI...TO. [7] Konwersji sieci w zbiór reguł dokonuje się przy użyciu parametru przedstawiającego wpływ najbardziej znaczących atrybutów na zmienną zależną (poziom akceptacji). Proces generowania reguł rozpoczyna się od utworzenia zbioru reguł zawierającego najistotniejsze zmienne, będące bezpośrednimi rodzicami zmiennej decyzyjnej. W kolejnym kroku zbiór ten może zostać uzupełniony o zmienne z poprzedniego pokolenia.

Klasyfikacja nieznanych przypadków może odbywać się przy pomocy wygenerowanych reguł. Wymaga to opracowania edytora, który po wprowadzeniu wartości atrybutów wykorzystanych w regułach sklasyfikuje dany przypadek jako 0 lub 1.

3. Przykład zastosowania programu do oceny kondycji ekonomiczno-finansowej małych przedsiębiorstw transportowych z województwa lubelskiego

Badania przeprowadzono dla przedsiębiorstw sekcji Transport, gospodarka magazynowa i łączność (601_642). Zbiór informacyjny zawiera 185 przypadków: 29 sklasyfikowanych jako 0 i 156 sklasyfikowanych jako 1.

Tablica 1. Struktura zbioru informacyjnego

	Y		
rok	0	1	Suma całkowita
1	3	19	22
2	8	33	41
3	7	36	43
4	7	31	38
5	4	37	41
Suma całkowita	29	156	185

Na podstawie tego zbioru skonstruowano dwa zbiory: uczący i testujący. Aby zapewnić proporcjonalny udział przypadków obu klas ($y = 0$ i $y = 1$) w zbiorze uczącym do zbioru tego włączono wszystkie przypadki z $y = 0$ i taką samą liczbę przypadków z $y = 1$ dla każdego roku. Do zbioru testującego włączono wszystkie przypadki z $y = 0$ i pozostałe przypadki z $y = 1$.

Zbiór informacyjny zawiera 43 atrybuty i atrybut decyzyjny y . Atrybut decyzyjny przyjmuje dwie wartości: 0 lub 1. W celu ograniczenia liczby atrybutów użytych do budowy modelu policzono korelacje pomiędzy każdym atrybutem a atrybutem decyzyjnym. Do wyliczenia tych korelacji wykorzystano cały zbiór informacyjny. Do budowy modelu włączono atrybuty, których współczynnik korelacji r_i z atrybutem decyzyjnym spełnia warunek $|r_i| > 0,1^i$

Ponieważ wszystkie atrybuty mają wartości ciągłe należało je skategoryzować. W programie BeliefSEEKER wygenerowano i przetestowano modele sieci dla różnej liczby kategorii atrybutów i różnej wartości parametru Dirichleta. Najlepsze efekty klasyfikacyjne uzyskano przy 10 kategoriach wszystkich atrybutów i parametrze Dirichleta $\alpha=30$.

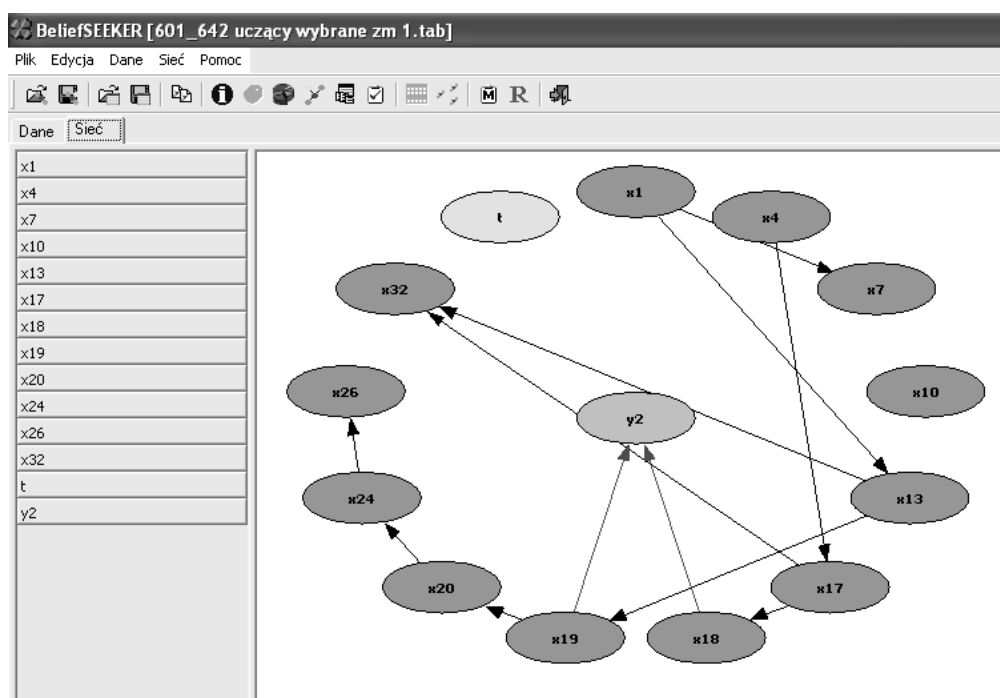
Dla tej sieci najbardziej znaczące (mające bezpośredni wpływ na decyzję) są atrybuty X_{18} (Wskaźnik globalnego obrotu aktywami) i X_{19} (wydajność pracy)

Tablica 2. Definicje atrybutów użytych do budowy modelu

Nazwa atrybutu	Definicja
X_1	Udział rzeczowych składników majątku w aktywach ogółem
X_4	Udział środków pieniężnych w aktywach ogółem
X_7	Obciążenie zobowiązaniami bieżącymi
X_{10}	Udział kapitału obrotowego w finansowaniu majątku ogółem
X_{13}	Zapotrzebowanie na kapitał obrotowy
X_{17}	Wskaźnik poziomu kosztów finansowych
X_{18}	Wskaźnik globalnego obrotu aktywami (produktywności majątku)
X_{19}	Wydajność pracy
X_{20}	Wskaźnik intelektualnej wartości dodanej (VAIC)
X_{24}	Wskaźnik rentowności sprzedaży brutto
X_{26}	Wskaźnik rentowności majątku
X_{32}	Udział sprzedaży na eksport w sprzedaży ogółem
T	rok

W oparciu o te atrybuty zostały wygenerowane reguły klasyfikacyjne. Następnie zastosowano program RuleSEKER, który dla niesklasyfikowanych przypadków wygenerował kolejne reguły w oparciu o pozostałe atrybuty. W rezultacie otrzymano następujący zbiór reguł klasyfikacyjnych:

1. JEZELI $X_{18} \geq 1,024$ ORAZ $X_{18} < 3,84$ ORAZ $X_{19} \geq 0,015$ ORAZ $X_{19} < 0,075$
TO Y JEST 0
2. JEZELI $X_{18} \geq 3,136$ ORAZ $X_{18} < 3,84$ ORAZ $X_{19} \geq 0,134$ ORAZ $X_{19} < 0,193$
TO Y JEST 0
3. JEZELI $X_{19} \geq 0,372$ ORAZ $X_{19} < 0,431$
TO Y JEST 0
4. JEZELI $X_{18} \geq 0,32$ ORAZ $X_{18} < 1,728$ ORAZ $X_{19} \geq 0,075$ ORAZ $X_{19} < 0,134$
TO Y JEST 1
5. JEZELI $X_{32} \geq 0,0034$ ORAZ $X_{32} < 0,0737$
TO Y JEST 0
6. JEZELI $X_{18} < 3,3661$ ORAZ $X_{26} < -0,157$
TO Y JEST 0
7. JEZELI $X_7 < 0,1$
TO Y JEST 1
8. JEZELI $X_{32} \geq 0,0737$ ORAZ $X_{32} < 0,1694$
TO Y JEST 1
9. JEZELI $X_{10} \geq -0,0004$ ORAZ $X_{20} \geq 0,2192$ ORAZ $X_{32} \geq 0,5634$
TO Y JEST 1



Rys. 2. Wygenerowana sieć przekonań

10. JEZELI $X_{19} \geq 0,0619$ ORAZ $X_{24} \geq -0,0912$ ORAZ $X_{26} < -0,157$
TO Y JEST 1
11. JEZELI $X_1 \geq 80,0242$ ORAZ $X_{13} \geq 0,0757$ ORAZ $X_{24} \geq -0,0912$
TO Y JEST 1
12. JEZELI $X_{18} \geq 3,3661$ ORAZ $X_{26} \geq -0,0005$ ORAZ $X_{32} < 0,0034$
TO Y JEST 1

Uzyskany model przetestowano przy pomocy zbioru testującego opisanego powyżej i przy pomocy zbiorów testujących zawierających wszystkie przypadki z działu Transport lądowy i rurociągowy (601_603) i z grupy Transport lądowy pozostały (602_602). Wyniki klasyfikacji nieznanych przypadków przy pomocy wygenerowanych reguł przedstawiono poniżej.

Transport, gospodarka magazynowa i łączność (601_642)

Właściwość	Wartość	Prog.\Popr.	0	1
Liczba przypadków	156	0	29	33
Liczba błędnych klasyfikacji	33	1	0	94
Liczba niesklasyfikowanych przypadków	0			
Błąd klasyfikacji	21.15 [%]			

Transport lądowy pozostały (602_602)

Właściwość	Wartość	Prog.\Popr.	0	1
Liczba przypadków	117	0	24	24
Liczba błędnych klasyfikacji	24	1	0	69
Liczba niesklasyfikowanych przypadków	0			
Błąd klasyfikacji	20.51 [%]			

Transport lądowy, transport rurociągowy (601_603)

Właściwość	Wartość	Prog.\Popr.	0	1
Liczba przypadków	117	0	24	24
Liczba błędnych klasyfikacji	24	1	0	69
Liczba niesklasyfikowanych przypadków	0			
Błąd klasyfikacji	20.51 [%]			

Rys. 3. Wyniki klasyfikacji nieznanych przypadków.

4. Podsumowanie

Model sieci przekonań wygenerowany na podstawie danych sekcji został przetestowany przy pomocy danych sekcji (601_642), działu (601_603) i grupy (602_602). Wyniki klasyfikacji są bardzo zbliżone, stąd nasuwa się wniosek że model wygenerowany dla sekcji może być stosowany do klasyfikacji nieznanych przypadków w dziale i grupie.

Wyniki klasyfikacji w tym modelu nie są najlepsze. Przyczyną tego jest duże niezrównoważenie zbioru uczącego. Zapewnienie proporcjonalnej liczby obu klas zmiennej decyzyjnej w zbiorze uczącym spowodowało ograniczenie zbioru uczącego do 58 przypadków. Powiększenie zbioru uczącego o informacje z kolejnego roku powinno poprawić sprawność modelu.

Bibliografia

1. Cichosz P.: Systemy uczące się. WNT Warszawa 2000;
2. Grzymała-Busse J.W., Hippe Z.S., Mroczek T.: System BeliefSEEKER ? A New Approach to Induction of Belief Networks and Belief Rules w: Burczyński T., Cholewa W., Moczulski W. (Eds.) Artificial Intelligence Methods (AI-METH), Silesian University of Technology Edit. Office, Gliwice (Poland) 2005, pp. 59-60.
3. Heckerman D.: Atutorial on Learning Bayesian Networks. Technical report MSR-TR-95-06.
4. Kłopotek M.A, Wierzchoń S. T.: Discovery of Bayesian Networks from Data with Maintainance of Partially Oriented Graphs. Intelligent Information Systems. Advances in Soft Computing Series of Physica-Verlag/Springer Verlag, Heidelberg/New York 2000, s. 277- 288
5. Kononenko, I.: Comparison of Inductive and Naive Bayesian Learning Approaches to Automatic Knowledge Acquisition. In B. Wielinga (Ed.), Current Trends in Knowledge Acquisition. Amsterdam, The Netherlands: IOS Press. 1990.
6. Kuczmowska B. Zastosowanie sieci Bayesa do prognozowania bankructwa, Zamojskie studia i materiały , Zamość 2006
7. Mroczek T., Grzymała-Busse J.W., Hippe Z.S.: Rules from Belief Networks: A Rough Set Approach In: Tsutomoto S., Słowiński R., Komorowski J., Grzymała-Busse J.W. (Eds.) Rough Sets and Current Trends in Computing, Springer-Verlag, Berlin-Heidelberg 2004, pp. 483-487.
8. Neuman Ł. Kościów Sz.: Wyszukiwanie informacji z wykorzystaniem sieci Bayesa. Prace naukowe Akademii Ekonomicznej we Wrocławiu nr 931 Wrocław 2002
9. Sarkar, S., R. S. Sriram.: Bayesian Models for Early Warning of Bank Failures. Management Science, 47. 2001
10. Sun L., Shenoy P.P.: Using Bayesion Networks for Bankruptcy Prediction, School of Business, Working Paper no.295 University of Kansas
11. Wąsowski A. Zdalna generacja sieci bayesowskich z baz danych. Praca dyplomowa pod kierunkiem dr had. inż. Mieczysława A. Kłopotka Politechnika Warszawska, Wydział Matematyki i Nauk Informacyjnych . Warszawa 2000