

Damian Gąska

Uniwersytet Ekonomiczny we Wrocławiu

ISSN 1644-6739
e-ISSN 2449-9765

DOI: 10.15611/sps.2016.14.08

Streszczenie: Celem pracy jest weryfikacja metod uczenia struktury sieci bayesowskich, wykorzystywanych do konstrukcji klasyfikatorów stosowanych do prognozowania bankructwa przedsiębiorstw. Empirycznej weryfikacji poddano kilka metod uczenia struktury sieci bayesowskich, wykorzystywanych pośrednio do konstrukcji klasyfikatorów. Analizy dotyczyły spółek notowanych na GPW w Warszawie. W pracy pokrótce omówiono podstawy teoretyczne rozpatrywanych zagadnień oraz zaprezentowano wyniki badań.

Słowa kluczowe: prognozowanie bankructwa, sieci bayesowskie, uczenie struktury.

1. Wstęp

Artykuł poświęcony jest tematowi prognozowanie bankructwa firm z zastosowaniem metod klasyfikacji, rozumianej tutaj jako szczególny przypadek uczenia pod nadzorem (z nauczycielem), przy wykorzystaniu wskaźników finansowych jako cech. Pośrednio do klasyfikacji zastosowano sieci bayesowskie.

Załóżmy, że danych jest n obserwacji $L_n = \{x_i, y_i\}: i = 1, 2, \dots, n\}$ będących realizacjami pary (\mathbf{X}, Y) , gdzie $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$ jest wektorem losowym reprezentującym cechy, a Y jest kategoryjną zmienną losową kodującą klasę, do której należy obserwacja.

W odniesieniu do prognozowania bankructwa współrzędne wektora \mathbf{X} odpowiadają cechom charakteryzującym firmę. Przyjmijmy umownie, że $Y = -1$ oznacza spółkę w dobrej kondycji finansowej, $Y = 1$ zaś koduje klasę bankrutów.

Zadanie polega na określeniu funkcji d przyporządkowującej nową obserwację do klasy spółek zdrowych lub zagrożonych upadłością. Funkcję d taką, że

$$d : \mathcal{X} \rightarrow \{-1, 1\}, \quad (1)$$

gdzie $\mathcal{X} \subseteq \mathbb{R}^m$ jest przestrzenią obserwacji $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$, nazywamy *klasyfikatorem* lub funkcją klasyfikacyjną.

Celem artykułu jest empiryczne porównanie kilku metod konstrukcji klasyfikatorów, wykorzystujących pośrednio różne warianty uczenia struktury sieci bayesowskich. Sieci bayesowskie należą do tzw. graficznych modeli probabilistycznych i umożliwiają wnioskowanie probabilistyczne (niekiedy nazywane są sieciami wnioskowania probabilistycznego – zob. [Cichosz 2007]). Mogą być również wykorzystane do konstrukcji klasyfikatorów postaci (1).

W części 2 zwięźle nakreślone zostały podstawy teoretyczne omawianych zagadnień, ilustrujące wykorzystanie sieci bayesowskich do rozwiązania problemu klasyfikacji. Część 3 opisuje analizy empiryczne dotyczące wykorzystania rozpatrywanych metod do prognozowania bankructwa polskich przedsiębiorstw notowanych na Giełdzie Papierów Wartościowych. Porównano skuteczność klasyfikacyjną metod. W ostatniej części krótko podsumowano otrzymane wyniki.

2. Wykorzystanie sieci bayesowskich w klasyfikacji

Zamieszczone w tej części pracy rozważania teoretyczne mają ogólny charakter, swoim zakresem wykraczają ponad omawiany tutaj problem prognozowania bankructwa i można je odnieść do dowolnego zagadnienia, w którym wykorzystuje się dychotomiczną klasyfikację.

Praktyczna konstrukcja klasyfikatorów (1) podporządkowana jest intencji minimalizacji błędu klasyfikacji. Optymalna pod tym względem (zob. np. [Ćwik, Koronacki 2008; Hastie i in. 2009]) jest funkcja klasyfikacyjna d^* określona wzorem

$$d^*(\mathbf{x}) = \begin{cases} 1, & \text{gd}y P(Y = 1 | \mathbf{X} = \mathbf{x}) > P(Y = -1 | \mathbf{X} = \mathbf{x}) \\ -1, & \text{w przeciwnym wypadku} \end{cases} \quad (2)$$

nazywana *klasyfikatorem bayesowskim*.

Prawdopodobieństwa warunkowe występujące we wzorze (2) zależą od rozkładu pary (\mathbf{X}, Y) , który nie jest przeważnie znany dla spotykanych w praktyce problemów. Z tego względu konstruuje się funkcje klasyfikacyjne jak najbardziej zbliżone do klasyfikatora bayesowskiego, wykorzystując do tego zaobserwowane dane tworzące próbę uczącą $L_n = \{(\mathbf{x}_i, y_i) : i = 1, 2, \dots, n\}$. Można w tym celu zastosować sieć bayesowską.

Przyjmijmy, dla ułatwienia notacji, że $X_{m+1} := Y$. Załóżmy, że każdej współrzędnej wektora losowego $\mathbf{X} = (X_1, X_2, \dots, X_{m+1})^T$ odpowiada węzeł pewnego acyklicznego grafu skierowanego G .

Graf skierowany definiuje jako uporządkowaną parę (V, E) , gdzie $V = \{X_1, \dots, X_{m+1}\}$ jest zbiorem węzłów (wierzchołków), a $E \subseteq \{X_i, X_j\}$:

$i \neq j\}$ jest zbiorem krawędzi (skierowanych) pomiędzy wierzchołkami, które na rysunkach przedstawiane są jako strzałki. Acykliczność grafu oznacza spełnianie warunku zakazującego istnienia ścieżki (ciągu bezpośrednio połączonych wierzchołków) mającej za początek i koniec ten sam węzeł.

Graf G reprezentuje *sieć bayesowską* dla $\tilde{\mathbf{X}}$, jeżeli rozkład tego wektora losowego dany jest funkcją

$$p(\tilde{\mathbf{x}}) = \prod_{i=1}^{m+1} p(x_i | \mathbf{X}_{Par(i)} = \mathbf{x}_{Par(i)}). \quad (3)$$

Przy tym litera p oznacza gęstość lub funkcję masy prawdopodobieństwa, w zależności od typu rozkładu. Wyrażenie $Par(i)$ we wzorze (3) oznacza zbiór indeksów rodziców węzła X_i w grafie G . Rodzicami węzła X_i są wszystkie te wierzchołki, z których krawędź prowadzi bezpośrednio do X_i . Przez $\mathbf{x}_{Par(i)}$ należy rozumieć wartości odpowiadające zmiennym z $Par(i)$. Przykładowo, jeśli rodzicami węzła (zmiennej) X_3 są węzły X_1 i X_2 , to $Par(3) = \{1, 2\}$, a zapis $\mathbf{X}_{Par(3)} = \mathbf{x}_{Par(3)}$ należy rozumieć jako $X_1 = x_1 \wedge X_2 = x_2$.

Równanie (3) oznacza warunkową niezależność zmiennych od wszystkich innych węzłów (zmiennych), które nie są ich potomkami, przy znanych wartościach rodziców. Jest to tzw. *własność Markova* sieci bayesowskiej.

Znając zarówno strukturę grafu sieci, jak i odpowiednie rozkłady warunkowe można – wykorzystując wzór Bayesa i twierdzenie o prawdopodobieństwie całkowitym – obliczać najbardziej prawdopodobne konfiguracje stanów pewnych zmiennych w sieci, przy znanych wartościach innych zmiennych. Rozumowania tego typu to tzw. *wnioskowanie w oparciu o sieć bayesowską (bayesian network inference)*.

Do rozwiązania problemu klasyfikacji konieczne jest wykorzystanie wnioskowania w oparciu o sieć bayesowską do obliczenia prawdopodobieństw warunkowych $P(Y = y | \mathbf{X} = \mathbf{x}) = P(X_{m+1} = x_{m+1} | \mathbf{X} = \mathbf{x})$ występujących we wzorze (2) na klasyfikator bayesowski. Stosując równanie (3), można pokazać, że

$$P(Y = y | \mathbf{X} = \mathbf{x}) = \frac{\prod_{i=1}^{m+1} p(x_i | \mathbf{X}_{Par(i)} = \mathbf{x}_{Par(i)})}{\sum_{x_{m+1} \in \{-1, 1\}} \prod_{i=1}^{m+1} p(x_i | \mathbf{X}_{Par(i)} = \mathbf{x}_{Par(i)})}. \quad (4)$$

Klasyfikator $d_G(\mathbf{X})$, powstały przez podstawienie (4) do wzoru (2), to klasyfikator skonstruowany w oparciu o sieć bayesowską daną grafem G .

Dotychczasowe rozważania odnoszą się do wyidealizowanej sytuacji, gdy znana jest struktura sieci (tzn. graf reprezentujący sieć) oraz rozkłady warunkowe występujące w równaniu (3). Nie jest to jednak typowe dla problemów napotykanych w praktyce. Gdy znana jest struktura sieci, to zachodzi zwykle konieczność specyfikacji odpowiednich rozkładów warunkowych. Wnioskowanie na temat tych rozkładów opiera się na próbie uczącej i wykorzystaniu metody największej wiarygodności (lub innej metody, np. algorytmu EM).

Z perspektywy rozpatrywanego tutaj zagadnienia najistotniejszym (i najtrudniejszym) typem wnioskowania dotyczącego sieci bayesowskich jest uczenie struktury (postaci grafu) sieci bayesowskiej z danych (*structure learning*).

Każdą metodę uczenia struktury, jakie zostały zastosowane na potrzeby analiz empirycznych, można przyporządkować do jednej z trzech grup odpowiadających trzem ogólnym podejściom do rozwiązania tego problemu (zob. [Scutari 2010]).

Pierwszą z nich stanowią metody uczenia struktury oparte na ograniczeniach (*constraint-based algorithms*). Sieć buduje się wówczas, tworząc zestaw warunków (zwanymi ograniczeniami co do struktury) typu $(X_1 \perp\!\!\!\perp X_2 | X_3)$, co oznacza „ X_1 jest niezależne od X_2 przy danym X_3 ”. Wykorzystuje się w tym celu testy warunkowej niezależności. Procedura przebiega przy tym dwuetapowo. W pierwszym etapie, zwanym identyfikacją ograniczeń (*constraint identification phase*), określa się, które zmienne powinny być połączone krawędzią. Wynikiem tego kroku jest tzw. szkielet sieci, tzn. graf o krawędziach nieskierowanych. W kolejnym etapie – nazywanym orientowaniem krawędzi (*edge orientation phase*), determinuje się ostateczną postać grafu G , nadając krawędziom odpowiedni kierunek.

Alternatywną grupę metod stanowią tzw. algorytmy wartościujące (*score-based algorithms*). W tym podejściu każdej „kandydującej” strukturze sieci przypisuje się pewną ocenę na bazie funkcji wartościującej i dąży się do jej maksymalizacji z wykorzystaniem algorytmów zachłanych (*greedy algorithms*).

Trzecia grupa to metody hybrydowe, łączące cechy obu poprzednich podejść.

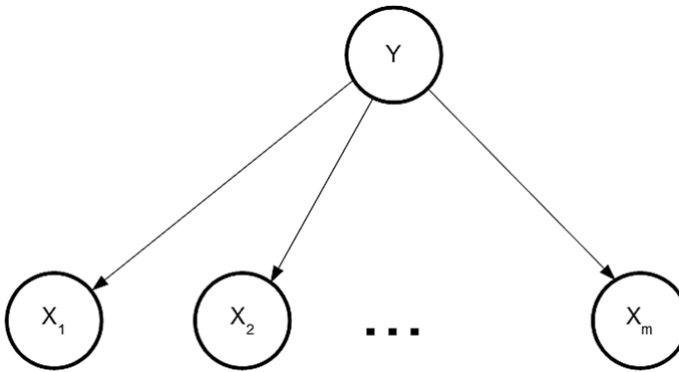
Rozważany w tej pracy sposób wykorzystania sieci bayesowskich w klasyfikacji można podsumować trzema krokami:

1. Wykorzystaj próbę $L_n = \{(\mathbf{x}_i, y_i), i = 1, 2, \dots, n\}$ do uczenia struktury sieci bayesowskiej – tzn. wyznaczenia grafu G , dla wektora (\mathbf{X}, Y) , oraz estymacji odpowiednich rozkładów warunkowych zmiennych X_i , $|\mathbf{X}_{Par(i)}$ występujących w (3).

2. Zastosuj wnioskowanie w oparciu o sieć bayesowską otrzymaną w kroku 1 do oszacowania prawdopodobieństw warunkowych $\hat{P}(Y = 1 | \mathbf{X} = \mathbf{x})$ i $\hat{P}(Y = -1 | \mathbf{X} = \mathbf{x})$, wykorzystując (4).

3. Do klasyfikacji nowych obserwacji wykorzystaj klasyfikator $d_G(\mathbf{x})$ zdefiniowany poprzez zastąpienie prawdopodobieństw we wzorze (2) na klasyfikator bayesowski oszacowaniami uzyskanymi w kroku 2.

Jeżeli nie stosuje się uczenia struktury sieci bayesowskiej, to przyjmuje się najprostsz y wariant sieci, jaki jest wykorzystywany do klasyfikacji (zob. [Ong 2011]), tj. sieć, której wszystkie krawędzie prowadzą od zmiennej kodującej etykietę klasy do zmiennych reprezentujących cechy klasyfikowanego obiektu. Sieć ta została przedstawiona na rys. 1.



Rys. 1. Struktura sieci bayesowskiej odpowiadająca naiwnemu klasyfikatorowi bayesowskiemu

Źródło: opracowanie własne.

Równanie (3) przyjmuje dla sieci z rys. 1 postać

$$p(x_1, x_2, \dots, x_m, y) = p(y) \prod_{i=1}^m p(x_i | y), \quad (5)$$

co oznacza warunkową niezależność zmiennych $X_i (i = 1, 2, \dots, m)$ w obrębie klas $Y = y$.

Klasyfikator oparty na sieci bayesowskiej z rys. 1 to tzw. *naiwny klasyfikator bayesowski*. Charakteryzuje się on dużą prostotą oraz niewielkimi wymaganiami obliczeniowymi, dzięki czemu jest popularnie stosowany.

Naiwny klasyfikator bayesowski może być użyteczny w zastosowaniach praktycznych (tzn. odzwierciedlać się akceptowalnym poziomem błędu klasyfikacji), nawet gdy założenie (5) nie jest spełnione. W [Ha-

stie i in. 2009] uzasadniono to faktem, że niedokładność aproksymacji

$p(\mathbf{x} | Y = y) \approx \prod_{i=1}^m \hat{p}(x_i | Y = y)$ wynikająca z nieprawdziwości założenia o niezależności jest równoważona mniejszą wariancją estymatorów w porównaniu do występującej w przypadku bezpośredniej estymacji wielowymiarowej funkcji $p(\mathbf{x} | Y = y)$.

Jednym z celów analiz empirycznych, których wyniki zrelacjonowano w dalszej części pracy, było zbadanie, czy uczenie struktury sieci na podstawie próby uczącej pozwoli na redukcję błędu klasyfikacji w stosunku do otrzymanego przy zastosowaniu naiwnego klasyfikatora bayesowskiego, dla którego struktura sieci jest z góry ustalona.

W analizach empirycznych – oprócz sieci bayesowskich – wykorzystano również klasyfikator oparty na modelu regresji logistycznej. Powodem uwzględnienia tej metody w analizie porównawczej jest jej duża popularność przy wykorzystaniu do prognozowania bankructwa, co uzasadnia traktowanie jej jako pewnego punktu odniesienia wobec metod proponowanych w treści pracy.

Klasyfikator oparty na modelu regresji logistycznej dany jest wzorem

$$d(\mathbf{x}) = \begin{cases} -1, & \text{gdy } \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m)} \leq \frac{1}{2} \\ 1, & \text{w przeciwnym wypadku} \end{cases} \quad (6)$$

gdzie β_i są parametrami modelu regresji logistycznej (zob. [Ostasiewicz 2012]).

3. Analizy empiryczne

Na potrzeby analiz przyjęto, że firmę reprezentuje zestaw wskaźników finansowych przedstawionych w tab. 1.

Etykiety klas zostały określone następująco: $Y = -1$ oznacza spółkę finansowo zdrową, $Y = 1$ zaś koduje klasę bankrutów.

Materiałem źródłowym do obliczenia wskaźników były roczne sprawozdania finansowe spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie w latach 2008–2015 (do maja), za rok poprzedzający o dwa lata datę upadłości.

Z analiz wyłączone zostały firmy z sektora finansowego, ze względu na ich odmienną charakterystykę. Ostatecznie wyłoniono 108 spółek, w tym 42, które zbankrutowały we wspomnianym okresie. Próbkę konstruowano w ten sposób, by – w miarę możliwości – spółki „zdro-

Tabela 1. Wybrane wskaźniki finansowe

Wskaźniki rentowności	X_1	zysk brutto ze sprzedaży/aktywa ogółem
	X_2	zysk netto/aktywa ogółem
	X_3	zysk brutto/aktywa ogółem
	X_4	zysk z działalności operacyjnej/przychody netto ze sprzedaży
Wskaźniki płynności	X_5	aktywa obrotowe bez krótkoterminowych rozliczeń międzyokresowe/zobowiązania krótkoterminowe
	X_6	aktywa obrotowe bez krótkoterminowych rozliczeń międzyokresowych-zapasy/zobowiązania krótkoterminowe
	X_7	kapitał obrotowy/aktywa ogółem
	X_8	inwestycje krótkoterminowe/zobowiązania krótkoterminowe
Wskaźniki zadłużenia	X_9	zobowiązania krótkoterminowe/aktywa ogółem
	X_{10}	zobowiązania ogółem/aktywa ogółem
	X_{11}	kapitał własny/zobowiązania ogółem
	X_{12}	(kapitał własny + zobowiązania długoterminowe)/aktywa trwałe
	X_{13}	(zysk netto + amortyzacja)/zobowiązania ogółem
	X_{14}	zysk brutto/zobowiązania krótkoterminowe
Wskaźniki sprawności	X_{15}	koszty operacyjne(bez pozostałych kosztów operacyjnych)/zobowiązania krótkoterminowe
	X_{16}	przychody ze sprzedaży/suma bilansowa
	X_{17}	przychody ze sprzedaży/należności krótkoterminowe

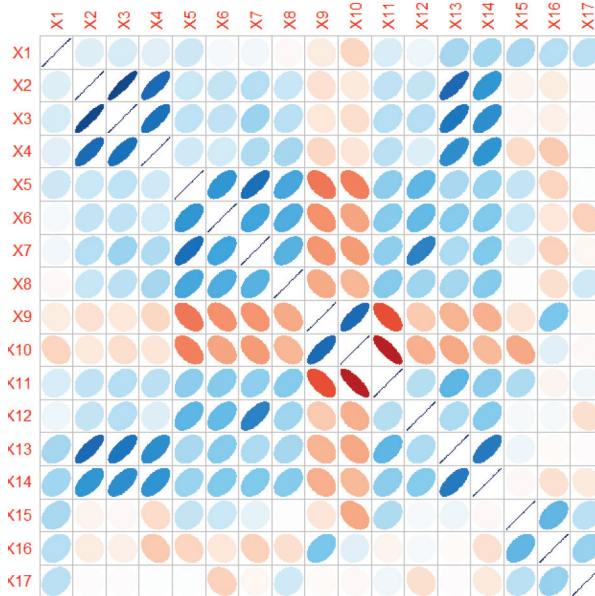
Źródło: opracowanie własne.

we” były zbliżone do poszczególnych bankrutów pod względem sektora działalności oraz wielkości przedsiębiorstwa.

Fakt, że przyjmowane w naiwnej klasyfikacji bayesowskiej założenie o niezależności zmiennych objaśniających w grupach określonych kategoriami zmiennej Y nie jest tutaj spełnione, ilustruje rys. 2 przedstawiający współczynniki korelacji Pearsona pomiędzy cechami w grupie bankrutów ($Y = 1$). Wartościom wszystkich korelacji odpowiadają elipsy. Im elipsa jest bardziej spłaszczona, tym wyższa – co do modułu – jest wartość odpowiedniego współczynnika korelacji. Ułożenie głównej półosi elipsy od lewego dolnego rogu do górnego prawego świadczy o korelacji dodatniej, odwrotne ułożenie odpowiada korelacji ujemnej. Ponadto siłę korelacji odzwierciedla natężenie koloru, jakim elipsy oznaczone są na rysunku.

Widać, że występują cechy mocno skorelowane liniowo, co nie jest niczym zaskakującym wzięwszy pod uwagę definicje cech.

W badaniach przeanalizowano metody uczenia sieci bayesowskich dostępne w bibliotece `bnlearn` (zob. [Scutari 2010]) pakietu statystycznego R. Do wnioskowania w oparciu o skonstruowane sieci wyko-



Rys. 2. Ilustracja macierzy korelacji pomiędzy cechami w grupie bankrutów

Źródło: opracowanie własne.

rzystano z kolei R-pakiet `gRain` [Hojsgaard 2012]. Zgodnie z obowiązującą praktyką (zob. [Ong 2011]) przyjęto, że zmienna reprezentująca klasę jest obligatoryjnie rodzicem wszystkich pozostałych zmiennych (pozostałe związki podlegają uczeniu). Listę przetestowanych metod prezentuje tab. 2.

Tabela 2. Metody uczenia struktury sieci bayesowskich dostępne w R-pakiecie `bnlearn`

Metoda	Oznaczenie	Typ metody	Odnosićnik
Grow-Shrink	gs	<i>constrained-based</i>	[Margaritis 2003]
Incremental Association	iamb	<i>constrained-based</i>	[Tsamardinos i in. 2003]
Fast Incremental Association	fast.iamb	<i>constrained-based</i>	[Yaramakala, Margaritis 2005]
Interleaved Incremental Association	inter.iamb	<i>constrained-based</i>	[Yaramakala, Margaritis 2005]
Hill-Climbing	hc	<i>score-based</i>	[Russell, Norvig 2009]
Tabu Search	tabu	<i>score-based</i>	[Russell, Norvig 2009]
Max-Min Hill-Climbing	mmhc	hybrydowa	[Tsamardinos i in. 2006]
2-phase Restricted Maximization	rsmax2	hybrydowa	[Tsamardinos i in. 2006]

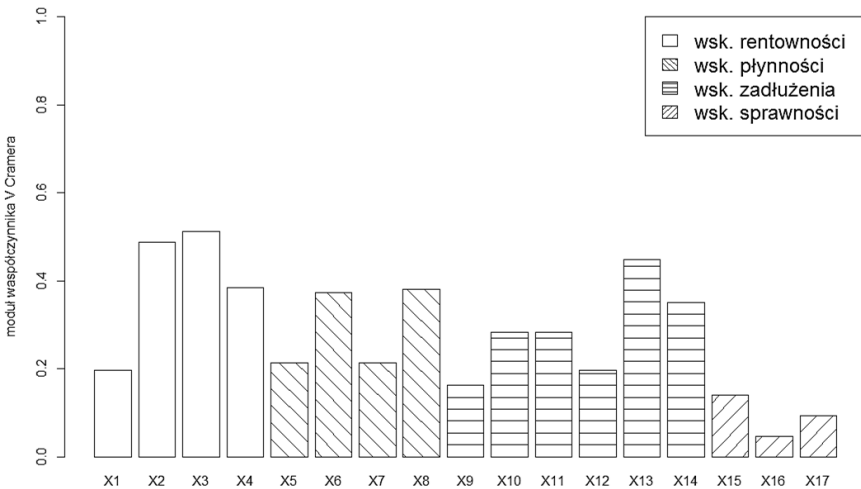
Źródło: opracowanie własne na podstawie dokumentacji pakietu `bnlearn`.

Pewnym ograniczeniem oprogramowania zastosowanego do wnioskowania w sieciach bayesowskich (chodzi mianowicie o pakiet gRain) jest obsługa wyłącznie cech typu nominalnego lub porządkowego o skończonej i stosunkowo małej liczbie wartości.

W celu umożliwienia zastosowania wskaźników finansowych w zadaniu klasyfikacji przy użyciu sieci bayesowskich, cechy te (mające charakter atrybutów ciągłych) poddano kategoryzacji przedziałowej w oparciu o kwantyle empiryczne. W praktyce każdą wartość cechy w próbie zastąpiono odpowiednim przedziałem $(-\infty, q_{33\%}]$, $(q_{33\%}, q_{67\%}]$, $(q_{67\%}, \infty)$. Gdzie q_α oznacza kwantyl (empiryczny) rzędu α z wartości cechy w próbie. Przesłanką dla takiego podziału była chęć uzyskania możliwie niewielu kategorii, o rozsądnej liczbie obserwacji przypadającej na każdą, bez tracenia przy tym siły dyskryminacyjnej indywidualnych cech, mierzonej tutaj współczynnikiem V Cramera (zob. [Acock, Stavig 1979]) pomiędzy cechą a etykietą klasy. Rysunek 3 przedstawia wykres słupkowy wartości współczynnika V dla wszystkich cech (im wartość jest bliższa 1, tym silniejszy związek między cechą a Y).

Zauważmy, że pod względem indywidualnej siły dyskryminacyjnej wyróżniają się wskaźniki rentowności i zadłużenia.

W charakterze metod referencyjnych dla klasyfikacji uwzględniającej każdy z wariantów uczenia sieci bayesowskiej wymienionych w tab. 2, wykorzystano naiwny klasyfikator bayesowski (oznaczony tu



Rys. 3. Wykres wartości współczynników V Cramera między skategoryzowanymi cechami a etykietą klasy

Źródło: opracowanie własne.

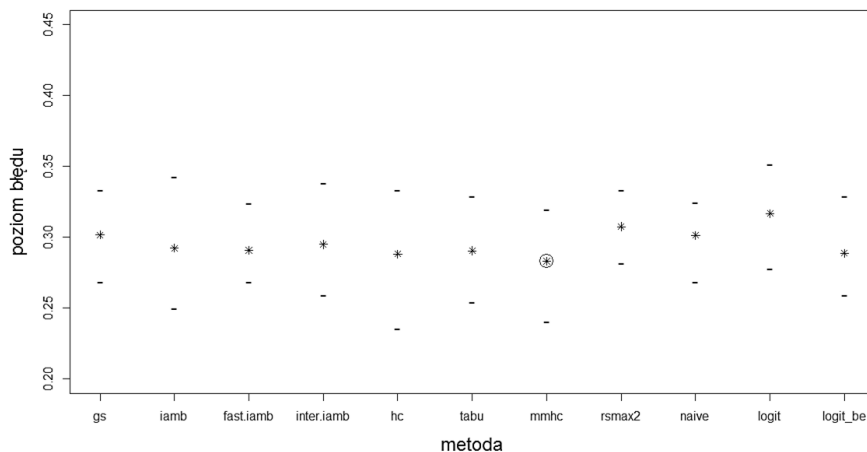
jako *naive*) oraz popularnie stosowaną w prognozowaniu upadłości (zob. [Prusak 2005]) metodę bazującą pośrednio na modelu regresji logistycznej, tj. metodę konstrukcji klasyfikatorów postaci (6). Tę ostatnią uwzględniono w wariancie ze wszystkimi cechami oraz z wykorzystaniem wyboru cech metodą krokową eliminacji wstecznej (*backward elimination*). Listę metod referencyjnych przedstawiono w tab. 3.

Tabela 3. Metody referencyjne uwzględnione w przeprowadzonych analizach

Metoda	Oznaczenie	Odnosnik
Naiwny klasyfikator bayesowski	naive	[Ćwik, Koronacki 2008; Krzyśko i in. 2008]
Klasyfikator bazujący na modelu regresji logistycznej	logit	[Krzyśko i in. 2008; Ostasiewicz 2012]
Jw. z uwzględnieniem wyboru cech metodą <i>backward elimination</i>	logit_be	[Agresti 2002]

Źródło: opracowanie własne.

Jako wskaźnik skuteczności metod przyjęto błąd klasyfikacji wyznaczany w oparciu o metodę pięciokrotnego sprawdzania krzyżowego (*5-fold-Cross-Validation*, zob. np. [Krzyśko i in. 2008]) zwaną dalej *CV*. W zamiarze redukcji wariancji estymatora *CV*, procedurę sprawdzania krzyżowego powtórzono 1000-krotnie. Liczba iteracji umożli-

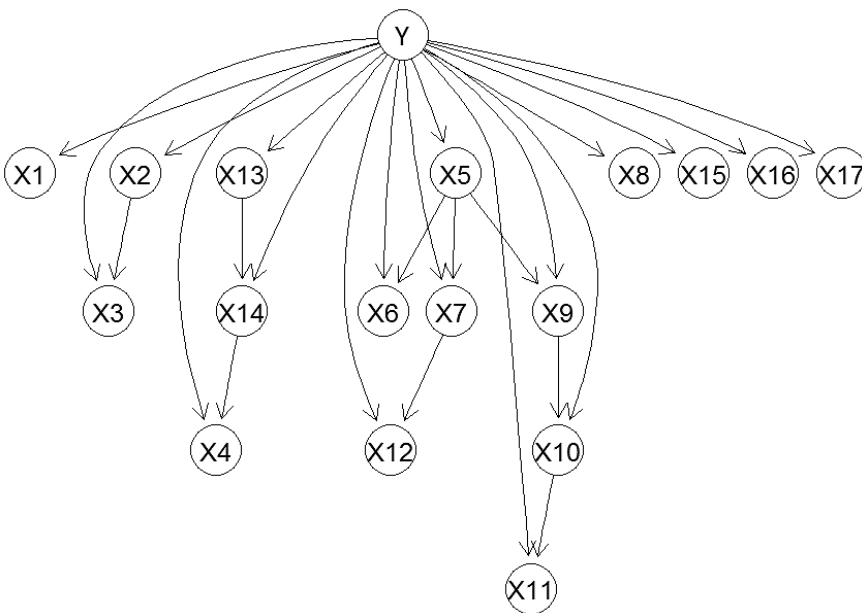


Rys. 4. Wskaźnik błędów klasyfikacji uzyskany metodą *CV*. Gwiazdką oznaczono średni błąd uzyskany z 1000 powtórzeń przebiegu metody. Kreski przedstawiają wartości 5. i 95. percentyla

Źródło: opracowanie własne.

wiła podanie – obok uśrednionego wyniku – kwantyli empirycznych rzędu α , wynoszące 5 i 95%. Wyniki prezentuje rys. 4.

Widoczne jest zróżnicowanie wyników w zależności od przyjętej metody uczenia struktury sieci bayesowskich. Chociaż skala różnic jest względnie nieduża, to widać, że nie jest obojętne, jakiej metody uczenia struktury użyjemy. Najniższy średni błąd (na poziomie 28,3%) uzyskano dla metody Max-Min Hill-Climbing (mmhc, wartość na rysunku wyróżniona obwódką). Graf sieci, dla której uzyskano najlepszy wynik, zaprezentowano na rys. 5.



Rys. 5. Struktura sieci bayesowskiej uzyskana metodą mmhc, dla której uzyskano najmniejszy błąd klasyfikacji

Źródło: opracowanie własne.

Na tle skuteczności klasyfikacyjnej traktowanego jako referencyjny naiwnego klasyfikatora bayesowskiego (ok. 30%) poprawa jest stosunkowo nieznaczna, chociaż zauważalna.

Mimo tego, że uwzględnienie uczenia struktury sieci bayesowskich nie przyczyniło się do znaczącej, jakościowej poprawy skuteczności klasyfikacji, najlepszy wynik uzyskany metodą mmhc (lepszy od wyników wszystkich metod referencyjnych: naive – 30,1%, logit – 31,6%, logit_be – 28,9%) sugeruje jednak pewien potencjał tkwiący w rozwijaniu metod uczenia struktury sieci bayesowskiej do klasyfikacji.

4. Podsumowanie

Nr 14(20)

Przeprowadzone analizy empiryczne dotyczące prognozowania upadłości nie dały jednoznacznego potwierdzenia przydatności metod uczenia sieci bayesowskich w tym obszarze zagadnień, ze względu na zbyt małą skalę różnic w osiągniętych wynikach.

Punktem odniesienia w porównywaniu i ocenie metod bazujących na uczeniu struktury sieci bayesowskiej był naiwny klasyfikator bayesowski, któremu odpowiada sieć o z góry ustalonej strukturze (zob. rys. 1). Nie okazał się on znacząco gorszy od metod dopuszczających złożoną postać sieci. Zróżnicowanie wyników w zależności od przyjętej metody uczenia struktury sugeruje jednak, że właśnie w rozwoju i poprawie skuteczności tych metod może tkwić klucz do sukcesu w ich zastosowaniu do klasyfikacji.

Wymagana w modelach sieci bayesowskich warunkowa niezależność zmiennych (równanie (3)) jest wciąż mocnym założeniem, ale dopuszczenie uczenia struktury sieci z danych, a nie wymuszenia jej konkretnego i to bardzo szczególnego kształtu – tak jak w naiwnym klasyfikatorze bayesowskim, daje dużo większą elastyczność.

Uzyskany ostatecznie model będzie zwykle pewnym uproszczeniem rzeczywistości, ale skala tego uproszczenia może być mniejsza niż w przypadku naiwnej klasyfikacji bayesowskiej.

Utworzenie struktury sieci bayesowskiej umożliwi ponadto graficzną prezentację schematu zależności między cechami, co samo w sobie jest pożyteczne, gdyż może przyczynić się do lepszego zrozumienia analizowanych danych. Może być zatem pomocniczo stosowane na etapie poprzedzającym właściwą konstrukcję klasyfikatora – np. w fazie wyboru cech.

Literatura

- Acock A.C., Stavig G.R., 1979, *A measure of association for nonparametric statistics*, Social Forces vol. 57, no. 4, s. 1381–1386.
- Agresti A., 2002, *Categorical Data Analysis*, Wiley & Sons, Hoboken, NJ.
- Cichosz P., 2007, *Systemy uczące się*, WNT, Warszawa.
- Ćwik J., Koronacki J., 2008, *Statystyczne systemy uczące się*, EXIT, Warszawa.
- Hastie T., Tibshirani R., Friedman J., 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer.
- Hojsgaard S., 2012, *Graphical independence networks with the gRain package for R*, Journal of Statistical Software, vol. 46, no. 10, s. 1–26.
- Krzyżko M., Wołyński W., Górecki T., Skorzybut M., 2008, *Systemy uczące się*, WNT, Warszawa.

- Margaritis D., 2003, *Learning Bayesian Network Model Structure from Data*, School of Computer Science, Carnegie-Mellon University, Pittsburgh.
- Ong H.C., 2011, *Improving classification in Bayesian Networks using structural learning*, World Academy of Science, Engineering and Technology, vol. 5, no. 3, s. 1073–1077.
- Ostasiewicz W., 2012, *Myślenie statystyczne*, Wolters Kluwer Polska, Warszawa.
- Prusak B., 2005, *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa.
- Russell S.J., Norvig P., 2009, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.
- Scutari M., 2010, *Learning Bayesian Networks with the bnlearn R Package*, Journal of Statistical Software, vol. 35, no. 3, s. 1–22.
- Tsamardinos I., Aliferis C.F., Statnikov A., 2003, *Algorithms for large scale Markov blanket discovery*, [w:] *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*, AAAI Press, s. 376–381.
- Tsamardinos I., Brown L.E., Aliferis C.F., 2006, *The max-min hill climbing Bayesian Network structure learning algorithm*, Machine Learning, vol. 65, no. 1, s. 31–78.
- Yaramakala S., Margaritis D., 2005, *Speculative Markov blanket discovery for optimal feature selection*, [w:] *Proceedings of the Fifth IEEE International Conference on Data Mining*, IEEE Computer Society.

BANKRUPTCY PREDICTION WITH BAYESIAN NETWORKS

Summary: The aim of the paper is to compare accuracy of some bankruptcy prediction models based on Bayesian networks. Some network structure learning algorithms were analyzed as a tool for classifiers construction. Empirical analysis was applied to companies listed on Warsaw Stock Exchange. The paper gives short overview of theoretical background behind discussed issues and presents results of empirical analysis.

Keywords: bankruptcy prediction, Bayesian network, structure learning.

