

Joanna Olbryś¹

SIEĆ BAYESOWSKA JAKO NARZĘDZIE POZYSKIWANIA WIEDZY Z EKONOMICZNEJ BAZY DANYCH

Streszczenie: Proces decyzyjny w inwestowaniu rozpoczyna się od percepcji i przetwarzania napływających informacji. Podłoże decyzji stanowią przekonania dotyczące prawdopodobieństwa zajścia określonego zdarzenia. Jednostki racjonalne posługują się narzędziami teorii prawdopodobieństwa i statystyki, rozumując zgodnie z prawem Bayesa, czyli aktualizując wyobrażenia o prawdopodobieństwie zdarzenia wraz z ujawnianiem wszelkich nowych informacji, zarówno ilościowych, jak i jakościowych. Wydaje się zatem, że bardzo dobrym narzędziem wspomagającym decyzje inwestycyjne może być odpowiednio skonstruowany model sieci bayesowskiej (*Bayesian Network*). W artykule postawiono za cel główny prezentację możliwości zastosowania modelu sieci bayesowskiej do pozyskiwania wiedzy z ekonomicznej bazy danych, z uwzględnieniem informacji jakościowych oraz preferencji i subiektywnych ocen analityka finansowego, podejmującego decyzje w warunkach niepewności.

Słowa kluczowe: sieć bayesowska, system wspomagający decyzje inwestycyjne, diagram wpływu

1. Wstęp

Proces decyzyjny w inwestowaniu rozpoczyna się od percepcji i przetwarzania napływających informacji. Podłoże decyzji stanowią przekonania dotyczące prawdopodobieństwa zajścia określonego zdarzenia. Przekonania zaś to rezultat procesu wnioskowania o nieznanym, którego przesłankami są napływające informacje [5]. W ekonomii i finansach przyjmuje się, że jednostka racjonalna (inwestor):

- przetwarza wszelkie możliwe informacje, odpowiednio uaktualniając swoje przekonania,

¹ Wydział Informatyki, Politechnika Białostocka, Białystok

- na podstawie przekonań formułuje preferencje i podejmuje decyzje w taki sposób, aby maksymalizować swoją oczekiwaną użyteczność.

Podejście to zakłada, że jednostki racjonalne posługują się narzędziami teorii prawdopodobieństwa i statystyki, rozumując zgodnie z prawem Bayesa, czyli aktualizując wyobrażenia o prawdopodobieństwie zdarzenia wraz z ujawnianiem wszelkich nowych informacji, zarówno ilościowych, jak i jakościowych. Założenie bayesowskiego myślenia nie oznacza, że każdy osąd poprzedzony jest mozolnymi obliczeniami. Chodzi raczej o przeciętny rezultat wnioskowania, który jest zgodny z zasadami prawdopodobieństwa [5].

Biorąc pod uwagę powyższe rozważania można mniemać, że bardzo dobrym narzędziem wspomagającym decyzje inwestycyjne może być odpowiednio skonstruowana sieć bayesowska, która umożliwi uwzględnienie w procesie decyzyjnym informacji o zróżnicowanym charakterze. Pozwala również na wprowadzanie nowych informacji w postaci obserwacji (*evidence*) oraz na modyfikacje modelu zgodnie z preferencjami i subiektywnymi ocenami inwestora w warunkach niepewności.

Sieci bayesowskie (*Bayesian Networks*) [19], nazywane również probabilistycznymi modelami graficznymi, sieciami przekonań lub sieciami przyczynowo – skutkowymi, stały się ostatnio popularnym narzędziem do reprezentacji wiedzy w warunkach niepewności. Sieć bayesowska jest acyklicznym grafem skierowanym i zawiera część jakościową, która stanowi zbiór zmiennych – węzłów grafu wraz z probabilistycznymi zależnościami pomiędzy nimi oraz część ilościową, reprezentującą rozkład prawdopodobieństwa łącznego dla tych zmiennych. Z punktu widzenia inżynierii wiedzy, sieć bayesowska może odzwierciedlać strukturę przyczynowo – skutkową, która pozwala pełniej zrozumieć modelowany problem, zarówno ekspertom jak i użytkownikom systemu [6]. Głównym etapem w budowaniu sieci bayesowskiej jest określenie jej struktury oraz parametryzacja. Podstawową zaletą sieci bayesowskich jest to, że pozwalają one na zintegrowanie wiedzy eksperta z danymi statystycznymi. Jeżeli odpowiednia liczba danych jest dostępna, sieci bayesowskie, zarówno ich struktura jak i parametry, mogą być nauczone z wykorzystaniem danych.

Wnioskowanie w sieci bayesowskiej sprowadza się do wyznaczenia rozkładu prawdopodobieństwa *a posteriori*, pod warunkiem zaobserwowania wartości zmiennych modelu. Rozkład tego prawdopodobieństwa może być bezpośrednio wykorzystany we wspomaganiu decyzji inwestycyjnych.

Głównym celem, jaki w artykule stawia sobie autor, jest prezentacja możliwości zastosowania modelu sieci bayesowskiej do pozyskiwania wiedzy z ekonomicznej bazy danych, przy uwzględnieniu informacji jakościowych oraz preferencji i subiektywnych ocen analityka finansowego, podejmującego decyzje w warunkach niepewności. Dane ekonomiczne wykazują dużą zmienność, zatem ekono-

miczna baza danych powinna być często aktualizowana, aby jak najpełniej odzwierciedlała bieżącą sytuację rynkową.

2. Sieć bayesowska jako narzędzie pozyskiwania wiedzy z ekonomicznej bazy danych

W ekonomii i finansach definicja jednostki racjonalnej (*homo oeconomicus*) obejmuje dwie cechy tej jednostki: konsekwencję działania oraz dążenie do podnoszenia poziomu osobistego dobrobytu [5]. Prawie wszystkie decyzje inwestycyjne podejmowane są na podstawie oszacowanych wartości prawdopodobieństwa realizacji określonego scenariusza zdarzeń [10]. Prawdopodobieństwo jest miarą stopnia przekonania podmiotu co do prawdziwości zajścia określonego zdarzenia. Z formalnego punktu widzenia podejście bayesowskie polega na traktowaniu wszystkich wielkości, których wartości nie są znane przed dokonaniem obserwacji, jako zmiennych losowych i na konsekwentnym stosowaniu we wnioskowaniu prostych reguł rachunku prawdopodobieństwa [18]. W szczególności stosuje się dwie zasady probabilistyki: warunkowanie (względem danych statystycznych) oraz wyznaczanie rozkładów brzegowych. Dane historyczne służą do uzyskania rozkładu *a priori*. Przyjęcie rozkładu *a priori* jest cechą wyróżniającą podejście bayesowskie. W problemach decyzyjnych rozkład *a priori* ma możliwie najpełniej odzwierciedlać całą wstępną wiedzę o zmiennej [18]. Nowe informacje ilościowe, jakościowe, opinie ekspertów są odpowiednie do rozwinięcia rozkładu *a posteriori*. Proces aktualizacji informacji przebiega zgodnie z regułą Bayesa:

$$P(A|I) = \frac{P(A) \cdot P(I|A)}{P(I)} \quad (1)$$

gdzie:

$P(A|I)$ - prawdopodobieństwo *a posteriori* zdarzenia A po uzyskaniu informacji I,

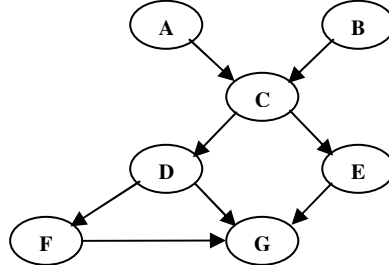
$P(A)$ - prawdopodobieństwo *a priori* (bazowe) zdarzenia A przed pojawieniem się informacji I,

$P(I|A)$ - prawdopodobieństwo pojawienia się informacji I, jeśli A jest prawdą,

$P(I)$ - prawdopodobieństwo całkowite zdarzenia I.

Sieci bayesowskie są szczególnym przypadkiem graficznych modeli probabilistycznych. Zwykle są one wykorzystywane do modelowania złożonych systemów o niepewnych lub niekompletnych danych. Sieć bayesowska jest skierowanym, acyklicznym grafem, którego topologia opisuje zależności (lub ich brak) między

zmiennymi modelu. Pojawienie się łuku łączącego dwa węzły jest interpretowane jako dowód istnienia bezpośredniego związku przyczynowo–skutkowego między węzłami. Zgodnie z definicją sieci bayesowskiej według Jensa [14] graf sieci zawiera zarówno węzły z rodzicami, jak i takie, które nie mają rodziców.



Rys. 1. Acykliczny graf skierowany. Prawdopodobieństwa, które należy wyznaczyć to:
 $P(A), P(B), P(C|A, B), P(E|C), P(D|C), P(F|D), P(G|D, E, F)$

Na rys.1 węzły, które nie mają rodziców, to są węzły A i B. Z nimi związane są brzegowe rozkłady prawdopodobieństwa. Pozostałe węzły C, D, E, F oraz G posiadają rodziców i odpowiadają im warunkowe rozkłady prawdopodobieństwa.

Bayesowska sieć zależności składa się z dwóch podstawowych części: jakościowej, czyli graficznej struktury zależności w modelu, oraz ilościowej, reprezentowanej przez rozkłady prawdopodobieństwa związane z grafem.

Trzy najważniejsze etapy tworzenia sieci bayesowskiej to:

- graficzna prezentacja zależności w modelu;
- specyfikacja numerycznych zależności między zmiennymi;

Rezultaty uzyskane w wyniku działania sieci w wysokim stopniu zależą od postaci graficznej modelu. Można stwierdzić, że najlepsze efekty uzyskujemy na ogół z bardzo dobrej reprezentacji graficznej problemu decyzyjnego, podczas gdy rozkłady prawdopodobieństwa związane z wierzchołkami grafu mogą być oszacowane w przybliżeniu [6]. Struktura sieci może mieć postać drzewa, polidrzewa lub grafu. Na przykład, algorytm Chow-Liu [4] nie buduje sieci przyczynowo-skutkowej, a jedynie nieorientowane drzewo zależności. Jeżeli sieć Bayesa określonego rozkładu ma postać drzewa, to ten algorytm powinien poprawnie odtworzyć jego kształt. Z kolei algorytm Pearl [19] jest rozszerzeniem algorytmu Chow-Liu. Otrzymana struktura nie jest drzewem skierowanym, ale ma postać polidrzewa. Najbardziej kosztowne natomiast jest uczenie grafu o dużej liczbie wierzchołków. Struktura sieci można zbudować z wykorzystaniem wiedzy eksperta lub na podstawie danych statystycznych, w sposób automatyczny. Jest to zada-

nie złożone obliczeniowo, lecz wykonalne za pomocą szerokiej gamy skutecznych algorytmów (m.in. [3], [7], [8], [9], [11], [14]).

Tym, co odróżnia sieci bayesowskie od innych metod reprezentowania wiedzy, jest wielość możliwości wnioskowania. Skupiając się na opisie jakościowym (czyli graficznej strukturze modelu), możemy identyfikować warunkowe zależności między zmiennymi. Uwzględniając opisy ilościowe (parametryczne modele przypisane węzłom) możemy, po wprowadzeniu do modelu nowej obserwacji (*evidence*), uzyskać rozkłady prawdopodobieństwa *a posteriori* pojedynczych zmiennych modelu lub rozkład łączny zbioru zmiennych. Możemy aktualizować, na podstawie opinii ekspertów, prawdopodobieństwa stanów zmiennych lub wartości zmiennych. Możemy też znaleźć najbardziej prawdopodobną (w świetle dostępnych obserwacji) konfigurację zmiennych nieobserwowalnych, jak również oszacować prawdopodobieństwo hipotezy, biorąc pod uwagę konkretne obserwacje. W pracy [15] podano następujące rodzaje wnioskowania w sieciach bayesowskich:

- O wiarygodności poszczególnych hipotez dla zadanych obserwacji.
- Poszukiwanie uzasadnienia dla zadanej hipotezy i obserwacji.
- Znalezienie prawdopodobieństwa prawdziwości wyrażenia logicznego przedstawionego w postaci koniunkcji wyrażeń elementarnych atrybut=wartość.
- Znalezienie odpowiedzi na kwerendę złożoną.
- O brzegowej i warunkowej niezależności zmiennych.

Podsumowując należy stwierdzić, że sieć bayesowska wydaje się bardzo użytecznym narzędziem wspomagającym zarządzanie portfelem papierów wartościowych ([6], [20]). Kombinacja czynników makroekonomicznych i mikroekonomicznych, tworzących odpowiednie węzły sieci, pozwala oszacować wartości oraz ryzyko portfela. Umożliwia też aktualizację wyceny portfela poprzez uwzględnienie informacji jakościowych, napływających z rynku oraz preferencji i subiektywnych ocen analityka finansowego - eksperta, podejmującego decyzje w warunkach niepewności.

3. Model diagnostyczny wspomagający decyzje inwestycyjne w warunkach niepewności – etap wstępny

Tradycyjne modele wyceny dóbr kapitałowych, takie jak model CAPM (*Capital Asset Pricing Model*), czy też model APT (*Arbitrage Pricing Theory*) opisują relacje między zmiennymi makroekonomicznymi, mikroekonomicznymi oraz stopami zwrotu z papierów wartościowych i mogą być wykorzystane do modelowania relacji między zmiennymi w sieci bayesowskiej ([6], [20]).

Przykładowy model sieci bayesowskiej, zaprezentowany w dalszej części pracy, zbudowany został na podstawie ekonomicznej bazy danych z rynku polskiego i w założeniu ma być częścią większego modelu wspomagającego proces wyceny oraz szacowania wartości zagrożonej Value-at-Risk portfeli akcji na Giełdzie Papierów Wartościowych w Warszawie. W modelu zaproponowano następujące zmienne makroekonomiczne (na podstawie [1], [2], [6]): inflacja, stopa bezrobocia, deficyt budżetowy, produkcja przemysłowa sprzedana, saldo handlu zagranicznego, średnia rentowność bonów skarbowych 52-tygodniowych. Substytutem portfela rynkowego jest Warszawski Indeks Giełdowy (WIG). Sektory na giełdzie reprezentują indeksy branżowe: WIG-BANKI, WIG-BUDOW (Budownictwo), WIG-INFO (Informatyka), WIG-SPOZYW (Przemysł Spożywczy), WIG-TELKOM (Telekomunikacja). Pozostałe indeksy branżowe, czyli WIG-MEDIA oraz WIG-PALIWA nie były uwzględnione w modelu z powodu zbyt krótkiego okresu funkcjonowania na giełdzie (WIG-MEDIA od 2004r., natomiast WIG-PALIWA od 2005r.). Próba statystyczna obejmuje dane miesięczne z okresu marzec 1998 – czerwiec 2006 (po 100 obserwacji dla każdej zmiennej). Szeregi czasowe zmiennych makroekonomicznych pochodzą z Biuletynów Statystycznych wydawanych przez Główny Urząd Statystyczny (www.stat.gov.pl) oraz Biuletynów Informacyjnych publikowanych przez Narodowy Bank Polski (www.nbp.pl). Źródłem danych dotyczących indeksów giełdowych były portale finansowe, m.in. <http://bossa.pl>, www.bankier.pl, www.wp.pl, www.gpw.com.pl.

Tabela 1

Symbole i definicje zmiennych makroekonomicznych oraz indeksów

l.p.	Symbol zmiennej	Nazwa zmiennej	Opis zmiennej
1.	ZPI_t	Zmiana poziomu inflacji r / r^2	Indeks cen towarów i usług konsumpcyjnych $r / r (PI_t)$, pomniejszony o 100
2.	ZPB_t	Zmiana poziomu bezrobocia ³	Miesięczna stopa wzrostu poziomu bezrobocia (PB_t) $ZPB_t = \frac{PB_t - PB_{t-1}}{PB_{t-1}} \cdot 100$
3.	ZDB_t	Zmiana poziomu deficytu budżetowego ⁴	Miesięczna stopa wzrostu deficytu budżetowego (DB_t) $ZDB_t = \frac{DB_t - DB_{t-1}}{DB_{t-1}} \cdot 100$

² Analogiczny miesiąc poprzedniego roku = 100³ Stopa bezrobocia jako zarejestrowani bezrobotni w stosunku do cywilnej ludności aktywnej zawodowo (w Polsce: 16-60 lat dla kobiet oraz 16-65 lat dla mężczyzn)⁴ Deficyt budżetowy = dochody budżetu państwa – wydatki budżetu państwa

4.	ZPP_t	Zmiana poziomu produkcji przemysłowej sprzedanej ⁵	Indeks łańcuchowy produkcji przemysłowej sprzedanej m/m (PP_t), pomniejszony o 100
5.	ZSH_t	Zmiana deficytu bilansu handlu zagranicznego ⁶	Miesięczna stopa wzrostu salda handlu zagranicznego (SH_t) $ZSH_t = \frac{SH_t - SH_{t-1}}{SH_{t-1}} \cdot 100$
6.	ZBS_t	Zmiana poziomu rentowności bonów skarbowych 52-tygodniowych	Miesięczna stopa wzrostu średniej rentowności bonów skarbowych 52-tygodniowych (BS_t) $ZBS_t = \frac{BS_t - BS_{t-1}}{BS_{t-1}} \cdot 100$
7.	$ZWIG_t$	Zmiana poziomu Warszawskiego Indeksu Giełdowego	Miesięczna stopa wzrostu indeksu giełdowego (WIG_t) $ZWIG_t = \frac{WIG_t - WIG_{t-1}}{WIG_{t-1}} \cdot 100$
8.	$ZWBANKI_t$	Zmiana poziomu indeksu branżowego $WIG - BANKI$	Miesięczna stopa wzrostu indeksu branżowego sektora bankowego ($WBANKI_t$) $ZWBANKI_t = \frac{WBANKI_t - WBANKI_{t-1}}{WBANKI_{t-1}} \cdot 100$
9.	$ZWBUDOW_t$	Zmiana poziomu indeksu branżowego $WIG - BUDOW$	Miesięczna stopa wzrostu indeksu branżowego sektora budownictwo ($WBUDOW_t$) $ZWBUDOW_t = \frac{WBUDOW_t - WBUDOW_{t-1}}{WBUDOW_{t-1}} \cdot 100$
10.	$ZWINFO_t$	Zmiana poziomu indeksu branżowego $WIG - INFO$	Miesięczna stopa wzrostu indeksu branżowego sektora informatyka ($WINFO_t$) $ZWINFO_t = \frac{WINFO_t - WINFO_{t-1}}{WINFO_{t-1}} \cdot 100$
11.	$ZWSPOZYW_t$	Zmiana poziomu indeksu branżowego $WIG - SPOZYW$	Miesięczna stopa wzrostu indeksu branżowego sektora przemysł spożywczy ($WSPOZYW_t$) $ZWSPOZYW_t = \frac{WSPOZYW_t - WSPOZYW_{t-1}}{WSPOZYW_{t-1}} \cdot 100$

⁵ Dynamika produkcji sprzedanej jest indeksem mierzącym comiesięczne zmiany wartości produkcji sprzedanej w cenach bazowych

⁶ Saldo handlu zagranicznego = eksport - import

12.	$ZWTELKOM_t$	Zmiana poziomu indeksu branżowego WIG-TELKOM	Miesięczna stopa wzrostu indeksu branżowego sektora telekomunikacja ($WTELKOM_t$) $ZWTELKOM_t = \frac{WTELKOM_t - WTELKOM_{t-1}}{WTELKOM_{t-1}} \cdot 100$
-----	--------------	--	---

Źródło: opracowanie własne

Tabele 3 oraz 4 (Dodatek) zawierają wartości podstawowych statystyk zmiennych makroekonomicznych oraz indeksów giełdowych, natomiast tabela 5 – współczynniki korelacji miesięcznych stóp zwrotu indeksów branżowych z indeksem WIG. Należy zauważyć, że stopy zwrotu z indeksu WIG są dość silnie skorelowane dodatnio ze stopami zwrotu wszystkich indeksów branżowych.

Po uporządkowaniu wartości zmiennych w postaci bazy danych zostały one poddane procesowi dyskretyzacji⁷. Utworzono dla każdej zmiennej przedziały o jednakowej liczebności, tworząc trzy stany reprezentujące niskie (*Low*), średnie (*Medium*) oraz wysokie (*High*) wartości, z prawdopodobieństwami odpowiednio: 34%, 33%, 33%. Jedynie dla zmiennej ZPB odpowiednie prawdopodobieństwa wynoszą: 35%, 32%, 33%. Powstałe w ten sposób rozkłady brzegowe przedstawia tabela 2.

Tabela 2

Brzegowe rozkłady prawdopodobieństwa zmiennych makroekonomicznych [%]

Stan	ZPI	ZPB	ZDB	ZPP	ZSH	ZBS
<i>Low</i>	[0,30; 1,75]	[-4,07; -0,52]	[-96,50; 6,89]	[-19; -1,80]	[-1480; -35,72]	[-14,76; -3,46]
<i>Medium</i>	[1,75; 6,45]	[-0,52; 0,94]	[6,89; 22,24]	[-1,80; 2,75]	[-35,72; 14,60]	[-3,46; 0,70]
<i>High</i>	[6,45; 13,90]	[0,94; 14,86]	[22,24; 550,10]	[2,75; 21,80]	[14,60; 3000]	[0,70; 14,57]

Źródło: opracowanie własne

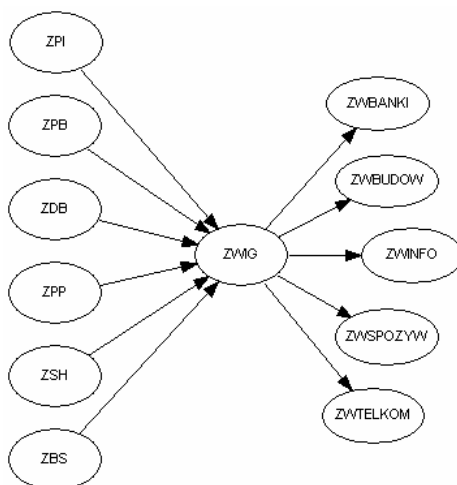
Zakładamy, że w budowanym modelu zmienne makroekonomiczne, tworzące rozkłady brzegowe, są wzajemnie niezależne [6], natomiast wartość zmiennej ZWIG, reprezentującej indeks giełdowy (substytut portfela rynkowego) zależy od wszystkich zmiennych makroekonomicznych. Jest to zgodne z przyjętymi w teorii portfela założeniami, że poziom ważonych wartościami rynkowymi indeksu giełdowego odzwierciedla już efekt oddziaływania zmiennych makroekonomicznych. Oznacza to, że węzły sieci bayesowskiej, odpowiadające zmiennym makroekonomicznym ZPI, ZPB, ZDB, ZPP, ZSH i ZBS są rodzicami węzła

⁷ Z wykorzystaniem narzędzia **GeNIe** (<http://genie.sis.pitt.edu>)

ZWIG, natomiast same nie mają rodziców. Z węzłem ZWIG związany jest warunkowy rozkład prawdopodobieństwa $P(\text{ZWIG} / \text{ZPI}, \text{ZPB}, \text{ZDB}, \text{ZPP}, \text{ZSH}, \text{ZBS})$ w $729 = 3^6$ możliwych stanach. Z kolei wartości indeksów branżowych są uzależnione od głównego indeksu giełdowego WIG (Tabela 5), na który mają wpływ, zdyskontowane przez inwestorów, informacje płynące z makroekonomicznego otoczenia rynku papierów wartościowych. Zatem w tworzonej sieci bayesowskiej węzłom ZWBANKI, ZWBUDOW, ZWINFO, ZWSPOZYW oraz ZWTELKOM powinny odpowiadać warunkowe rozkłady prawdopodobieństwa:

$$P(\text{ZWBANKI} / \text{ZWIG}), P(\text{ZWBUDOW} / \text{ZWIG}), P(\text{ZWINFO} / \text{ZWIG}), \\ P(\text{ZWSPOZYW} / \text{ZWIG}), P(\text{ZWTELKOM} / \text{ZWIG})$$

Pierwsza część modelu sieci bayesowskiej, zawierająca zmienne makroekonomiczne oraz indeksy giełdowe (rys. 2), była utworzona z wykorzystaniem modułu uczącego narzędzia **GeNIe**. Metoda *Greedy Trick Thinning*, kryterium oceny modeli *BDeu* (*Bayesian marginal likelihood with uniform Dirichlet prior*) [3] umożliwiła automatyczne zbudowanie grafu zgodnego z oczekiwaniami odnośnie do związków przyczynowo - skutkowych pomiędzy zmiennymi wchodzącymi w skład modelu (uczenie z wykorzystaniem wiedzy wstępnej). Wybrane tablice warunkowych rozkładów prawdopodobieństwa zmiennych reprezentujących indeksy branżowe WIG-BANKI oraz WIG-INFO przedstawiają rysunki 3 oraz 4.



Rys. 2. Graficzny model zależności między zmiennymi makroekonomicznymi oraz zmiennymi reprezentującymi indeksy główny i indeksy branżowe (opis zmiennych w Tabeli 1)

ZWIG	Low	Medium	High
► Low	0.814103	0.184818	0.00660066
Medium	0.179487	0.481848	0.333333
High	0.00641026	0.333333	0.660066

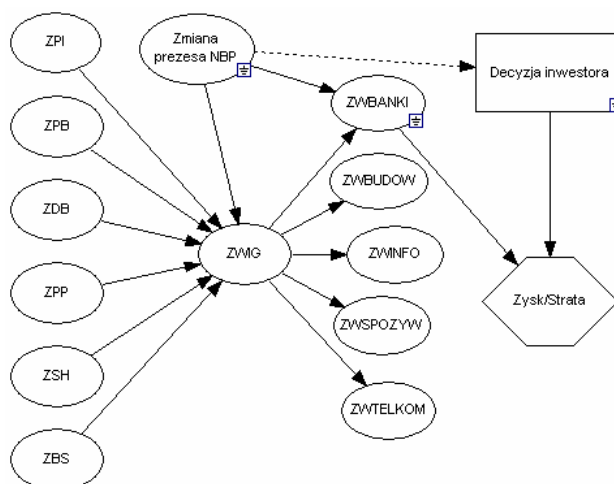
Rys. 3. Tablica warunkowego rozkładu prawdopodobieństwa $P(ZWBANKI / ZWIG)$

ZWIG	Low	Medium	High
► Low	0.785256	0.184818	0.0363036
Medium	0.179487	0.630363	0.184818
High	0.0352564	0.184818	0.778878

Rys. 4. Tablica warunkowego rozkładu prawdopodobieństwa $P(ZWINFO / ZWIG)$

Kolejnym etapem tworzenia systemu wspomagającego decyzje inwestycyjne na polskim rynku będzie rozbudowanie sieci z rys. 2 poprzez włączenie do modelu zmiennych mikroekonomicznych, specyfikujących spółki wchodzące w skład poszczególnych indeksów branżowych. W przyszłości głównym celem wykorzystania sieci będzie wycena portfeli akcyjnych oraz szacowanie wartości zagrożonej Value-at-Risk portfeli [16]. Ze względu na ograniczoną objętość artykułu będzie to tematem kolejnego opracowania.

Równie istotnym zagadnieniem, z punktu widzenia inwestora, jest możliwość rozbudowania sieci i stworzenie diagramu wpływu (*Influence Diagram* [13]), wspomagającego proces podejmowania optymalnych decyzji, z uwzględnieniem informacji jakościowych oraz subiektywnych ocen użytkownika modelu. Uproszczony schemat przykładowego diagramu wpływu przedstawiono na rys. 5. Informacja (*evidence*) o zmianie prezesa Narodowego Banku Polskiego może mieć, w ocenie inwestora, bezpośredni wpływ na wartość indeksów giełdowych oraz pośredni na stopę zwrotu (zysk lub stratę) z portfela akcyjnego, złożonego z akcji banków.



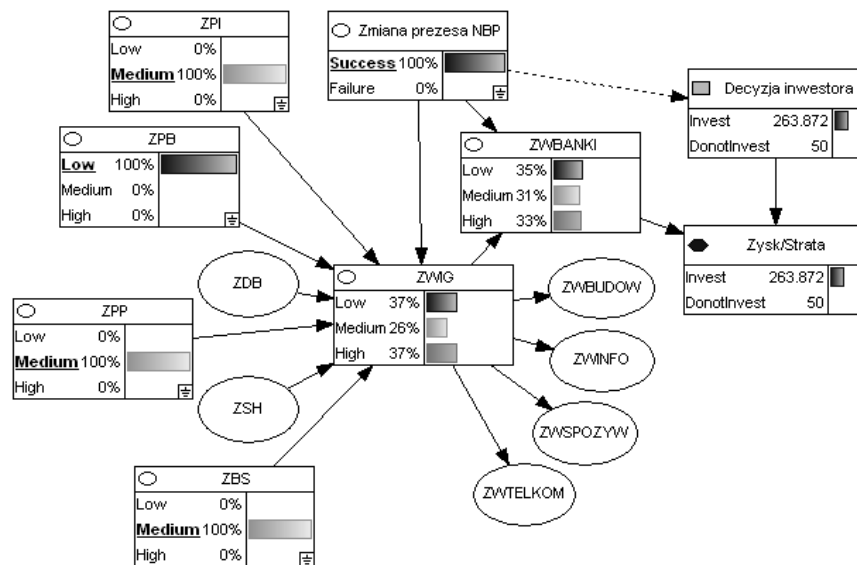
Rys. 5. Diagram wpływu (rozbudowana sieć bayesowska z rys. 2)

Zmienna *Decyzja inwestora* jest reprezentowana przez węzeł typu *Decision node*. Tablicę wartości zmiennej *Zysk/Strata*, reprezentowanej przez węzeł typu *Value node*, przedstawia rys. 6.

Decyzja inwest...	Invest			DonotInvest		
ZwBANKI	Low	Medium	High	Low	Medium	High
Value	100	200	500	50	50	50

Rys. 6. Tablica wartości zmiennej *Zysk/Strata* (Rys. 5.)

Na rys. 7 przedstawiono diagramu wpływu z rys. 5 po uwzględnieniu nowych obserwacji oraz aktualizacji modelu. Po zaobserwowaniu średnich wartości zmiennych makroekonomicznych ZPI, ZPP, ZBS oraz niskiej wartości zmiennej ZPB (tabela 1), jak również pozytywnej oceny (w opinii eksperta - analityka finansowego) wpływu zmiany prezesa Narodowego Banku Polskiego na sytuację rynkową, otrzymujemy rozkłady prawdopodobieństwa *a posteriori* interesujących nas zmiennych ZWIG oraz ZWBANKI. Tablica węzła *Zysk/Strata* przedstawia oczekiwaną wartość zysku z podjętej inwestycji.



Rys. 7. Diagram wpływu po uwzględnieniu obserwacji dotyczących wybranych węzłów sieci.

Zastosowania diagramów wpływu na rynku finansowym wymagają oddzielnego opracowania ze względu na bardzo istotną rolę preferencji i subiektywnych ocen analityków finansowych w procesie podejmowania decyzji w warunkach niepewności [5].

Zaprezentowany model diagnostyczny z rys. 2 był zmodyfikowany i rozbudowany w pracy [17] oraz wykorzystany do wspomagania wyceny akcji na Giełdzie Papierów Wartościowych w Warszawie, na przykładzie akcji Banku BPB. Włączenie do modelu kolejnych papierów wartościowych pozwoli w przyszłości stworzyć większy system wspomagający decyzje inwestycyjne.

4. Dodatek

Tabela 3

Podstawowe statystyki zmiennych makroekonomicznych [%]

l.p.	Statystyka	ZPI	ZPB	ZDB	ZPP	ZSH	ZBS
1.	Średnia	4,94	0,44	26,9	0,8	11,23	-1,55
2.	Odchylenie standardowe	3,86	2,59	86,58	6,82	399,71	5,3
3.	Mediana	4,15	0	11,76	-0,25	-16,97	-1,08
4.	Min	0,3	-4,07	-96,5	-19	-1480	-14,76
5.	Max	13,9	14,86	550,1	21,8	3000	14,57

Zródło: opracowanie własne

Tabela 4

Podstawowe statystyki indeksów giełdowych [%]

l.p.	statystyka	ZWIG	ZWBANKI	ZWBUDOW	ZWINFO	ZWSPOZYW	ZWTELKOM
1.	Średnia	0,75	1,14	1,02	0,27	0,74	0,16
2.	Odchylenie standardowe	7,36	7,49	8,42	12,34	7,22	12,01
3.	Mediana	1,92	2,14	0,98	-0,98	-0,04	-0,09
4.	Min	-27,09	-26,25	-32,63	-25,89	-20,27	-25,88
5.	Max	18,67	19,05	24,01	46,01	28,31	46,01

Źródło: opracowanie własne

Tabela 5

Współczynniki korelacji miesięcznych stóp zwrotu indeksów branżowych z indeksem WIG

	ZWBANKI	ZWBUDOW	ZWINFO	ZWSPOZYW	ZWTELKOM
ZWIG	0,882	0,784	0,796	0,685	0,821

Źródło: opracowanie własne

Podziękowania

Model przedstawiony w pracy został utworzony i przetestowany przy użyciu biblioteki klas C++: **SMILE**[®] i narzędzia **GeNIe**, służącego do tworzenia i wnioskowania w graficznych modelach probabilistycznych. Narzędzia powstały w Laboratorium Systemów Decyzyjnych (*Decision Systems Laboratory*) Uniwersytetu Pittsburgh'skiego i są dostępne na stronie <http://genie.sis.pitt.edu>.

Szczególne podziękowanie za pomoc w przygotowaniu artykułu oraz cenne wskazówki składam dr Agnieszce Onisko z Wydziału Informatyki Politechniki Białostockiej.

Literatura:

- [1] Adamczak, A.: *Empiryczna weryfikacja modelu arbitrażu cenowego w warunkach Giełdy Papierów Wartościowych w Warszawie*, [w:] T.Trzaskalik (red.) *Modelowanie preferencji a ryzyko'99*, Katowice, 1999, str. 31 – 46.
- [2] Berry, M.A., Burmeister, E., McElroy, M.B.: *Sorting Out Risk Using Known APT Factors*, *Financial Analysts Journal*, vol. 44, no. 2, 1988, pp. 29-42.

- [3] de Campos, L.M.: *A Scoring Function for Learning Bayesian Networks based on Mutual Information and Conditional Independence Tests*, Journal of Machine Learning Research 7 (2006), pp. 2149 – 2187.
- [4] Chow, C.K.: *Approximating Discrete probability Distributions with Dependence Trees*, IEEE Transactions on Information Theory, 14(3), 1968, pp. 462-467.
- [5] Cieślak, A.: *Behawioralna ekonomia finansowa. Modyfikacja paradygmatów funkcjonujących w nowoczesnej teorii finansów*, Materiały i Studia Zeszyt nr 165, NBP, Warszawa, 2003.
- [6] Demirer, R., Mau, R.R., Shenoy, C.: *Bayesian Networks: A Decision Tool to Improve Portfolio Risk Analysis*, Working Paper, University of Kansas, School of Business, 2006.
- [7] Druzdzel, M.J., Oniśko, A., Wasyluk, H.: *Uczenie parametrów sieci bayesowskich z danych z wykorzystaniem bramek Noisy-OR* [w:] Z.Bubnicki, O.Hryniewicz, R.Kulikowski (red.) Problemy współczesnej nauki. Teoria i zastosowania. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2002, str. 19 – 26.
- [8] Druzdzel, M.J., Cheng, J.: *AIS-BN: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian Networks*, Journal of Artificial Intelligence Research, 13, 2000, pp. 155-188.
- [9] Druzdzel, M.J., Yuan, Ch.: *An Importance Sampling Algorithm Based on Evidence Pre-propagation*, Proceedings of the Nineteenth Annual Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2003, pp. 624-631.
- [10] Hagstrom, R.G.: *The Warren Buffett Portfolio*, Wiley & Sons, New York, 1999.
- [11] Heckerman, D.: *A Tutorial on Learning with Bayesian Networks*, technical Report, MSR-TR-95-06, 1996.
- [12] Heckerman, D.: *Bayesian Networks for Data Mining*, Data Mining and Knowledge Discovery, 1, 1997, pp. 79 – 119.
- [13] Howard, R.A., Matheson, J.E.: *Influence Diagrams*, [in:] Howard R.A and Matheson J.E. (eds.) Applications of Decision Analysis, vol. 2, 1984, Menlo Park, Calif.: Strategic Decisions Group, pp. 721 – 762.
- [14] Jensen, F.V.: *Bayesian Networks and Decisions Graphs*, Springer-Verlag, 2001.
- [15] Kłopotek, M.A.: *Inteligentne wyszukiwarki internetowe*, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2001.
- [16] Olbryś, J.: *Estymatory miar Expected Shortfall i Value-at-Risk: przykłady zastosowania do pomiaru ryzyka walutowego*, Inwestycje finansowe i ubez-

- pieczenia. Tendencje światowe a rynek polski. Prace Naukowe Akademii Ekonomicznej im. O. Langego, Wrocław, Nr 1088, 2005, str. 65 – 72.
- [17] Olbryś, J.: *Model sieci bayesowskiej wspomagający decyzje inwestycyjne w warunkach niepewności*, Szósta Ogólnopolska Konferencja Naukowa „Modelowanie Preferencji a Ryzyko'07”, Ustroń, 2007, w trakcie recenzji.
- [18] Osiewalski, J.: *Ekonometria bayesowska w zastosowaniach*, Wydawnictwo Akademii Ekonomicznej w Krakowie, 2001.
- [19] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, 1988.
- [20] Shenoy, C., Shenoy, P.P.: *Bayesian Network Models of Portfolio Risk and Return*, [in:] Abu-Mostafa, Y.S., LeBaron, B., Lo, A.W., Weigend, A.S. (eds) *Computational Finance*, MIT Press, 1999, pp. 87-106.

BAYESIAN NETWORK AS A TOOL OF EXTRACTING KNOWLEDGE FROM AN ECONOMIC DATABASE

Abstract: Making a decision in investment starts from perception and analysis of incoming information. Rational investors reason according to Bayes formula and try to develop posterior probabilities after new evidence has been added. Virtually all decisions that investors make are exercises in probability. Bayesian networks have been used in different decision support system contexts that combine qualitative and quantitative information. Main goal of this paper is to present Bayesian network as a tool of extracting knowledge from an economic database, with respect to historical quantitative information, uncertain qualitative information, incomplete knowledge and evidence.

Keywords: bayesian networks, investment decision support system, influence diagram

Artykuł zrealizowano w ramach pracy badawczej statutowej S/II/1/03.

