

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 384

Taksonomia 24

**Klasyfikacja i analiza danych –
teoria i zastosowania**

Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2015

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp.....	9
Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: 25 lat SKAD.....	15
Beata Basiura, Anna Czapkiewicz: Symulacyjne badanie wykorzystania entropii do badania jakości klasyfikacji.....	25
Andrzej Bąk: Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie <code>pllord</code>	33
Justyna Brzezińska: Analiza klas ukrytych w badaniach sondażowych.....	42
Grażyna Dehnel: Rejestr podatkowy oraz rejestr ZUS jako źródło informacji dodatkowej dla statystyki gospodarczej – możliwości i ograniczenia ..	51
Sabina Denkowska: Wybrane metody oceny jakości dopasowania w <i>Propensity Score Matching</i>	60
Marta Dziechciarz-Duda, Klaudia Przybysz: Zastosowanie teorii zbiorów rozmytych do identyfikacji pozafiskalnych czynników ubóstwa.....	75
Iwona Foryś: Potencjał rynku mieszkaniowego w Polsce w latach dekonjunktury gospodarczej.....	84
Eugeniusz Gatnar: Statystyczna analiza konwergencji krajów Europy Środkowej i Wschodniej po 10 latach członkostwa w Unii Europejskiej.....	93
Ewa Genge: Zaufanie do instytucji publicznych i finansowych w polskim społeczeństwie – analiza empiryczna z wykorzystaniem ukrytych modeli Markowa.....	100
Alicja Grześkowiak: Wielowymiarowa analiza uwarunkowań zaangażowania Polaków w kształcenie ustawiczne o charakterze pozaformalnym.....	108
Monika Hamerska: Wykorzystanie metod porządkowania liniowego do tworzenia rankingu jednostek naukowych.....	117
Bartłomiej Jefmański: Zastosowanie modeli IRT w konstrukcji rozmytego systemu wag dla zmiennych w zagadnieniu porządkowania liniowego – na przykładzie metody TOPSIS.....	126
Tomasz Józefowski, Marcin Szymkowiak: Wykorzystanie uogólnionej miary odległości do porządkowania liniowego powiatów województwa podkarpackiego w świetle funkcjonowania specjalnej strefy ekonomicznej Euro-Park Mielec.....	135
Krzysztof Kompa: Zastosowanie testów parametrycznych i nieparametrycznych do oceny sytuacji na światowym rynku kapitałowym przed kryzysem i po jego wystąpieniu.....	144
Mariusz Kubus: Rekurencyjna eliminacja cech w metodach dyskryminacji....	154

Marta Kuc: Wpływ sposobu definiowania macierzy wag przestrzennych na wynik porządkowania liniowego państw Unii Europejskiej pod względem poziomu życia ludności	163
Paweł Lula: Kontekstowy pomiar podobieństwa semantycznego	171
Iwona Markowicz: Model regresji Feldsteina-Horioki – wyniki badań dla Polski	182
Kamila Migdał-Najman: Ocena wpływu wartości stałej Minkowskiego na możliwość identyfikacji struktury grupowej danych o wysokim wymiarze	191
Małgorzata Misztal: O zastosowaniu kanonicznej analizy korespondencji w badaniach ekonomicznych.....	200
Krzysztof Najman: Zastosowanie przetwarzania równoległego w analizie skupień	209
Edward Nowak: Klasyfikacja danych a rachunkowość. Rozważania o relacjach	218
Marcin Pelka: Adaptacja metody <i>bagging</i> z zastosowaniem klasyfikacji pojęciowej danych symbolicznych.....	227
Józef Pocięcha, Mateusz Baryła, Barbara Pawelek: Porównanie skuteczności klasyfikacyjnej wybranych metod prognozowania bankructwa przedsiębiorstw przy losowym i nielosowym doborze prób	236
Agnieszka Przedborska, Małgorzata Misztal: Wybrane metody statystyki wielowymiarowej w ocenie jakości życia słuchaczy uniwersytetu trzeciego wieku	246
Wojciech Roszka: Konstrukcja syntetycznych zbiorów danych na potrzeby estymacji dla małych domen	254
Aneta Rybicka: Połączenie danych o preferencjach ujawnionych i wyrażonych	262
Elżbieta Sobczak: Poziom specjalizacji w sektorach intensywności technologicznej a efekty zmian liczby pracujących w województwach Polski	271
Andrzej Sokołowski, Grzegorz Harańczyk: Modyfikacja wykresu radarowego	280
Marcin Szymkowiak, Marek Witkowski: Wykorzystanie mediany do klasyfikacji banków spółdzielczych według stanu ich kondycji finansowej ..	287
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: Wpływ wyboru metody klasyfikacji na identyfikację zależności przestrzennych – zastosowanie testu <i>join-count</i>	296
Dorota Witkowska: Wykorzystanie drzew klasyfikacyjnych do analizy zróżnicowania płac w Niemczech	305
Artur Zaborski: Analiza niesymetrycznych danych preferencji z wykorzystaniem modelu punktu dominującego i modelu grawitacji.....	315

Summaries

Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: XXV years of SKAD	24
Beata Basiura, Anna Czapkiewicz: Simulation study of the use of entropy to validation of clustering.....	32
Andrzej Bąk: Problem of choosing the optimal linear ordering procedure in the p_llord package.....	41
Justyna Brzezińska-Grabowska: Latent class analysis in survey research...	50
Grażyna Dehnel: Tax register and social security register as a source of additional information for business statistics – possibilities and limitations.....	59
Sabina Denkowska: Selected methods of assessing the quality of matching in Propensity Score Matching	74
Marta Dziechciarz-Duda, Klaudia Przybysz: Applying the fuzzy set theory to identify the non-monetary factors of poverty.....	83
Iwona Foryś: The potential of the housing market in Poland in the years of economic recessions.....	92
Eugeniusz Gatnar: Statistical analysis of the convergence of CEE countries after 10 years of their membership in the European Union.....	99
Ewa Genge: Trust to the public and financial institutions in the Polish society – an application of latent Markov models.....	107
Alicja Grześkowiak: Multivariate analysis of the determinants of Poles' involvement in non-formal lifelong learning	116
Monika Hamerska: The use of the methods of linear ordering for the creating of scientific units ranking.....	125
Bartłomiej Jefmański: The application of IRT models in the construction of a fuzzy system of weights for variables in the issue of linear ordering – on the basis of TOPSIS method	134
Tomasz Józefowski, Marcin Szymkowiak: GDM as a method of finding a linear ordering of districts of Podkarpackie Voivodeship in the light of the operation of the Euro-Park Mielec special economic zone	143
Krzysztof Kompa: Application of parametric and nonparametric tests to the evaluation of the situation on the world financial market in the pre- and post-crisis period.....	153
Mariusz Kubus: Recursive feature elimination in discrimination methods ...	162
Marta Kuc: The impact of the spatial weights matrix on the final shape of the European Union countries ranking due to the standard of living.....	170
Paweł Lula: The impact of context on semantic similarity.....	181
Iwona Markowicz: Feldstein-Horioka regression model – the results for Poland.....	190

Kamila Migdal-Najman: The assessment of impact value of Minkowski's constant for the possibility of group structure identification in high dimensional data.....	199
Małgorzata Misztal: On the use of canonical correspondence analysis in economic research.....	208
Krzysztof Najman: The application of the parallel computing in cluster analysis.....	217
Edward Nowak: Data classification and accounting. A study of correlations	226
Marcin Pelka: The adaptation of bagging with the application of conceptual clustering of symbolic data.....	235
Józef Pociecha, Mateusz Baryła, Barbara Pawelek: Comparison of classification accuracy of selected bankruptcy prediction methods in the case of random and non-random sampling technique.....	244
Agnieszka Przedborska, Małgorzata Misztal: Selected multivariate statistical analysis methods in the evaluation of the quality of life of the members of the University of the Third Age.....	253
Wojciech Roszka: Construction of synthetic data sets for small area estimation.....	261
Aneta Rybicka: Combining revealed and stated preference data.....	270
Elżbieta Sobczak: Specialization in sectors of technical advancement vs. effects of workforce number changes in Poland's voivodships.....	279
Andrzej Sokółowski, Grzegorz Harańczyk: Modification of radar plot.....	286
Marcin Szymkowiak, Marek Witkowski: Classification of cooperative banks according to their financial situation using the median.....	295
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: The influence of classification method selection on the identification of spatial dependence – an application of join-count test.....	304
Dorota Witkowska: Application of classification trees to analyze wages disparities in Germany.....	314
Artur Zaborski: Asymmetric preference data analysis by using the dominance point model and the gravity model.....	323

Ewa Genge

Uniwersytet Ekonomiczny w Katowicach

e-mail: ewa.genge@ue.katowice.pl

ZAUFANIE DO INSTYTUCJI PUBLICZNYCH I FINANSOWYCH W POLSKIM SPOŁECZEŃSTWIE – ANALIZA EMPIRYCZNA Z WYKORZYSTANIEM UKRYTYCH MODELI MARKOWA

Streszczenie: Ukryte modele Markowa zaliczane są do grupy modeli ze zmiennymi ukrytymi, w których wykorzystywana jest idea mieszanek rozkładów. Modele te stosowane są najczęściej dla danych panelowych czy szeregów czasowych, gdzie celem jest już nie tylko podział obserwacji na homogeniczne grupy, ale również pewna analiza zmian w czasie. W części empirycznej artykułu uwzględnione zostaną dodatkowo tzw. zmienne towarzyszące. Celem badań będzie podział Polaków na klasy o podobnym stopniu zaufania do instytucji publicznych i finansowych, a także zaobserwowanie zmian ich postaw na przestrzeni kilku lat. Badania przeprowadzone będą za pomocą ukrytych modeli Markowa (*latent Markov models*) z wykorzystaniem pakietów *depmixS4* i *Rsolnp* programu R.

Słowa kluczowe: ukryty model Markowa, mieszanka rozkładów, podejście modelowe w taksonomii.

DOI: 10.15611/pn.2015.384.10

1. Wstęp

Modele mieszanek rozkładów dla zmiennych jakościowych (mierzonych na skalach słabych) zwane są modelami (*latent class models*) lub analizą klas ukrytych (*latent class analysis*). W ostatnim czasie coraz bardziej na popularności zyskują modele klas ukrytych dla danych panelowych czy szeregów czasowych, gdzie celem jest już nie tylko podział obserwacji na homogeniczne grupy, ale również pewna analiza zmian w czasie t . W tym przypadku stosowane są ukryte modele Markowa (*LMM*, *latent Markov model*), które bardzo często wykorzystywane są w naukach społecznych. Modele te stosowane są również w psychologii do modelowania procesów uczenia się (zob. np. [Wickens 1982, Schmittmann i in. 2006]). W ekonomii modele te zwane są również modelami o zmiennych reżimach (zob.

np. [Ghysels 1994; Kim 1994]). Wśród innych zastosowań wymienić należy również rozpoznawanie mowy [Rabiner 1989] oraz różnego rodzaju badania genetyczne (zob. np. [Krogh 1998]). W tego rodzaju aplikacjach modele te nazywane są modelami ukrytego łańcucha Markowa (*HMM*, *hidden Markov models*). W modelach LMM analizowane są najczęściej krótkie, wielowymiarowe szeregi czasowe o dużej liczbie obserwacji (dane panelowe). Modele HMM stosowane są głównie do długich, jednowymiarowych szeregów czasowych pojedynczych procesów lub jednostek. Bardziej szczegółowe informacje na temat tego rodzaju modeli można znaleźć w pracach takich jak [Cappe, Moulines, Ryden 2005; Frühwirth-Schnatter 2006; Zucchini, MacDonald 2009].

W literaturze (zob. [Visser, Speekenbrink 2010]) można również spotkać się z terminem zależne modele mieszanek (*dependent mixture model*) na określenie modeli zarówno LMM, jak i HMM.

Impulsem do napisania niniejszego artykułu był poruszony w Raporcie Diagnozy Społecznej 2013 [Szumlicz 2013] problem niskiej wiarygodności instytucji zaufania publicznego wśród Polaków. Celem referatu będzie podział Polaków na klasy o podobnym stopniu zaufania do instytucji publicznych i finansowych, a także zaobserwowanie zmian ich postaw w ciągu kilku lat w zależności od wybranych zmiennych demograficznych.

2. Ukryty model Markowa

W ukrytym modelu Markowa badana jest funkcja gęstości wielowymiarowego szeregu czasowego $f(\mathbf{Y}_t)$, w którym struktura przejścia pomiędzy stanami zdefiniowana jest za pomocą procesu Markowa. W modelu tym dyskretna zmienna losowa X_t nie jest bezpośrednio mierzalna, a stany¹ łańcucha nazywa się ukrytymi.

Ukryty model Markowa można zapisać jako:

$$f(\mathbf{Y}_t) = \sum_{X_1=1}^u \cdots \sum_{X_T=1}^u f(X_1) \prod_{t=2}^T f(X_t | X_{t-1}) \prod_{t=1}^T f(\mathbf{Y}_t | X_t), \quad (1)$$

gdzie $f(X_1)$ – funkcja gęstości rozkładu początkowego, prawdopodobieństwo znalezienia się w ukrytej klasie w czasie $t = 0$.

$f(X_t | X_{t-1})$ – prawdopodobieństwo przejścia, które określa prawdopodobieństwo przynależności do klasy ukrytej w czasie t , pod warunkiem przynależności do tej klasy w czasie $t-1$. Ukryta macierz przejścia \mathbf{A} o elementach a_{sr} ($s, r = 1, \dots, u$),

¹ Ponieważ analizowane obserwacje są zmienne w czasie t , w literaturze z zakresu łańcuchów Markowa często spotykanym określeniem jest stan. Z racji zastosowania teorii łańcuchów Markowa na gruncie taksonomii w dalszej części pracy autor posługiwał się będzie terminem klasa lub grupa. W literaturze polskiej wśród prac z zakresu teorii łańcuchów Markowa wymienić należy [Podgórska i in. 2000; Stawicki 2004].

oznacza prawdopodobieństwo przejścia z ukrytej klasy s do klasy r , tj. $a_{sr} = P(X_t = r | X_{t-1} = s)$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1u} \\ \vdots & \ddots & \vdots \\ a_{u1} & \cdots & a_{uu} \end{bmatrix}. \quad (2)$$

Suma prawdopodobieństw w każdym wierszu macierzy \mathbf{A} jest równa jeden. $f(\mathbf{Y}_t | X_t)$ – funkcja gęstości rozkładu wielowymiarowego.

Ukryty model Markowa może dodatkowo zawierać jeszcze tzw. zmienne towarzyszące \mathbf{Z}_t (*covariates* lub *concomitant variables*). W modelach klas ukrytych ocenia się wpływ tych zmiennych na przynależność obiektów do klas (wpływ na prawdopodobieństwa *a priori*) (zob. np. [Dayton, Macready 1988, s. 173-178; Hagenars, McCutcheon 2002; Genge 2013]). W ukrytych modelach Markowa analizowany jest głównie wpływ zmiennych towarzyszących na prawdopodobieństwa przejścia do poszczególnych klas [Chung i in. 2007; Vermunt i in. 1999; Visser 2011].

Wpływ zmiennych towarzyszących na prawdopodobieństwa przejścia wyrażony jest za pomocą wielomianowej funkcji logitowej [Agresti 2002]. Parametry modeli mieszanek najczęściej szacowane są za pomocą algorytmu EM [Dempster i in. 1977, s. 1-38], a wybór modelu optymalnego dokonywany jest na podstawie kryteriów informacyjnych, tj. BIC oraz AIC [Akaike 1974; Schwarz 1978].

3. Analiza empiryczna

Analizę empiryczną przeprowadzono na podstawie danych panelowych pochodzących z projektu Diagnoza Społeczna [Diagnoza społeczna...2013]². Analizie poddano 6610 osób, które wzięły udział w trzech edycjach badania panelowego, przeprowadzonego w latach 2009, 2011, 2013. Szacowno więc modele dla 19 830 obserwacji łącznie. Analiza została przeprowadzona z uwzględnieniem ośmiu zmiennych.

W nawiasie podano oryginalne nazwy ze zbioru udostępnianego na stronie internetowej³, litera e poprzedzająca symbol zmiennej (np. $ep101_1$) dotyczy badania w roku 2009, litera f – badania w roku 2011, zaś litera g – badania w roku 2013.

1. Y_1 ($ep101_1, fp98_1, gp_104_1$): Zaufanie do banków⁴;
2. Y_2 ($ep101_2, fp98_3, gp_104_3$): Zaufanie do Sejmu;

² W raporcie Diagnozy [Szumlicz 2013] dokonano analizy porównawczej procentowego udziału odpowiedzi na każde z zadanych pytań w roku 2013 z osobna. W pracy podkreślono, że zmiana skali wyrażania opinii w roku 2013 pozwoliła poprawić wiarygodność niektórych instytucji finansowych i instytucji państwa w porównaniu do lat ubiegłych („sztucznie” podwyższyć oceny).

³ www.diagnoza.com.

⁴ Możliwe odpowiedzi na każde z poniższych pytań: 1 – tak; 2 – nie; 3 – nie mam zdania.

3. Y_3 (*ep101_3, fp98_4, gp_104_4*): Zaufanie do prezydenta;
4. Y_4 (*ep101_6, fp98_5, gp_104_5*): Zaufanie do Parlamentu Europejskiego;
5. Y_5 (*ep101_7, fp98_6, gp_104_6*): Zaufanie do policji;
6. Y_6 (*ep101_8, fp98_7, gp_104_7*): Zaufanie do rządu;
7. Y_7 (*ep101_9, fp98_8, gp_104_8*): Zaufanie do ZUS-u;
8. Y_8 (*ep101_5, fp98_10, gp_104_10*): Zaufanie do funduszy emerytalnych (OFE).

W badaniu uwzględniono również następujące zmienne towarzyszące:

- a) Z_1 : wiek – mniej niż 18 lat, 18-25 lat, 26-40 lat, 41-60 lat, więcej niż 60 lat,
- b) Z_2 : status (1 – pracownicy sektora publicznego; 2 – pracownicy sektora prywatnego; 3 – prywatni przedsiębiorcy; 4 – rolnicy; 5 – renciści; 6 – emeryci; 7 – uczniowie i studenci; 8 – bezrobotni; 9 – inni bierni zawodowo);
- c) Z_3 : edukacja (1 – podstawowe i niższe; 2 – zasadnicze zawodowe/gimnazjum; 3 – średnie; 4 – wyższe i policealne).

W badaniach wykorzystano pakiet depmixS4 programu R.

Ukrytą liczbę klas wybrano na podstawie wartości kryteriów informacyjnych AIC oraz BIC. Kryterium BIC oraz AIC wskazało minimalną wartość dla liczby klas równej trzy. Niewiele większą wartość otrzymano dla czterech klas, w przypadku kryterium AIC. W takich sytuacjach często wybierane są modele mniej złożone (zob. np. [Collins, Lanza 2011]), dlatego też w dalszej części pracy analizowano model o trzech klasach ukrytych.

Następnie szacowano ukryty model Markowa dla zmiennych $Y_1 - Y_8$.

W klasie pierwszej, stanowiącej 47% wszystkich respondentów (9332 respondentów), aż 43% w ogóle nie ma zaufania do banków, tylko 31% badanych jest przeciwnego zdania, a 26% respondentów nie chce określić swego nastawienia do tej instytucji. Prawie wszyscy ankietowani w tej klasie nie mają zaufania do sejmu (96%) oraz rządu (94%). Niestety zdecydowana większość badanych (80%) nie ufa również głowie naszego kraju, 15% ankietowanych ma odmienne zdanie. Tylko 13% badanych ma zaufanie do Parlamentu Europejskiego (64% pozostaje nieufne). 73% ankietowanych w tej klasie ma negatywne nastawienie do ZUS, a 65% do OFE. Największym, bo 41-procentowym zaufaniem respondenci darzą policję (połowa badanych ma odmienne zdanie).

Klasa druga jest klasą mniej liczną – należy do niej 29% wszystkich ankietowanych (5714 respondentów). Największy odsetek w tej grupie stanowią osoby, które unikały określenia swej postawy wobec Parlamentu Europejskiego (91%). Prawie 75% badanych nie chce odpowiedzieć na każde z pytań dotyczących Sejmu, rządu i prezydenta. Pozostała część badanych jest do nich nastawiona negatywnie (odpowiednio 23%, 23%, 14%). Podobnie jak w klasie pierwszej, respondenci najbardziej ufni pozostają wobec policji (37%), 50% nie ma zdania na temat tej instytucji państwowej. 15% badanych darzy zaufaniem ZUS, a 7% OFE. Zdecydowana większość respondentów tej klasy to osoby niemające zdania na temat instytucji finansowych tj., ZUS i OFE (odpowiednio 68%, 77,6%).

Do klasy trzeciej zaliczono 24% badanych, tj. 4784 osób. W klasie tej największy odsetek stanowią respondenci ufający policji (90%) oraz głowie państwa (86%). 73% badanych ma zaufanie do Parlamentu Europejskiego (16% jest bez zdania). Niespełna 70% ankietowanych zaufaniem darzy rząd, a 62% sejm (odpowiednio 25% i 29% ankietowanych ma odmienne nastawienie). Jeżeli chodzi o zaufanie do instytucji finansowych, to 58% badanych ma zaufanie do banków, 56% do ZUS i tylko 31% do OFE. Nieufni wobec tych instytucji stanowią kolejno 18% , 30% i 32%.

Oszacowane prawdopodobieństwa przejścia ukrytego modelu Markowa, obrazujące stabilność pozostania w danej klasie w kolejnych okresach, przedstawione zostały w tab. 1.

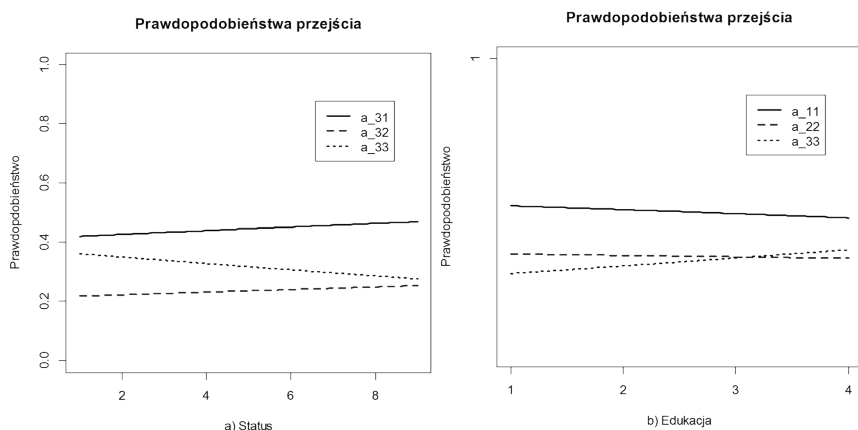
Tabela 1. Prawdopodobieństwa przejścia dla trzech klas

Klasa s/Klasa r	Klasa 1	Klasa 2	Klasa 3
Klasa 1	0,48	0,29	0,23
Klasa 2	0,47	0,33	0,20
Klasa 3	0,44	0,23	0,33

Źródło: opracowanie własne.

Największe prawdopodobieństwo pozostania w tej samej klasie w kolejnym okresie odpowiada klasie pierwszej, kolejno znacznie mniejsze prawdopodobieństwa, równe 0,33, odpowiadają klasie drugiej i trzeciej. Niestety osoby, które cechują się najmniejszym zaufaniem do instytucji publicznych i finansowych (klasa pierwsza), najprawdopodobniej nie zmienią swego podejścia (zdania). Jeżeli jednak uda im się zmienić swoje nastawienie czy pokonać różnego rodzaju bariery, to bardziej prawdopodobne okazuje się przejście do klasy drugiej, tj. $a_{12} = 0,29$, aniżeli do klasy trzeciej ($a_{13} = 0,23$). Mniejsze szanse na pozostanie w danej klasie w następnym okresie mają respondenci należący do klasy drugiej ($a_{22} = 0,33$). Jeżeli zmienią swoje nastawienie, to również dla nich mniej prawdopodobne okazuje się przejście do grupy osób o największym stopniu zaufania (tj. do klasy trzeciej, $a_{23} = 0,20$), aniżeli do klasy pierwszej ($a_{21} = 0,47$). Ankietowani należący do grupy respondentów o największym zaufaniu cechują się taką samą tendencją pozostania w tej grupie ($a_{33} = 0,33$) w następnym okresie jak klasa druga. Dla tej grupy osób istnieje większe prawdopodobieństwo przejścia do stanu o najgorszym nastawieniu (tj. klasy pierwszej), aniżeli do grupy osób niechcących wyrazić swego zdania (tj. klasy drugiej).

W kolejnej części pracy dokonano analizy wpływu zmiennych towarzyszących na prawdopodobieństwa przejścia (zmiany zdania) respondentów do poszczególnych klas. Spośród analizowanych zmiennych towarzyszących istotne okazały się tylko zmienne, tj. status oraz edukacja. Ilustrację graficzną wybranych prawdopodobieństw przejścia do poszczególnych klas dla obu zmiennych towarzyszących pokazano na rys. 1.



Rys. 1. Prawdopodobieństwa: a) przejścia do poszczególnych klas dla zmiennej „status” (lewa); b) pozostania w poszczególnych klasach dla zmiennej „edukacja” (prawa)

Źródło: opracowanie własne.

Dla zmiennej towarzyszącej „status” prawdopodobieństwo pozostania w klasie trzeciej jest tym mniejsze, im niższy jest status respondentów (zob. rys. 1a). Im „gorszy” status społeczny, tym większe prawdopodobieństwo przejścia do klasy pierwszej (zarówno dla a_{31} , jak i a_{21}) i drugiej (zarówno dla a_{32} , jak i a_{12}). Z kolei prawdopodobieństwo przejścia do klasy trzeciej spada wraz z pogarszaniem się statusu respondentów (dla każdej z badanych klas).

Dla zmiennej towarzyszącej „edukacja” prawdopodobieństwo pozostania w klasie pierwszej jest najwyższe, spada ono nieznacznie wraz z wykształceniem. Wraz ze stopniem wykształcenia spada również nieznacznie prawdopodobieństwo pozostania w klasie drugiej. Obserwuje się jednak wzrost prawdopodobieństwa pozostania w klasie trzeciej wraz z lepszym stopniem wykształcenia (zob. rys. 1b). Dokonując analizy prawdopodobieństw przejścia (wykresów, które ze względu na ograniczenia objętościowe nie zostały zamieszczone w pracy), można stwierdzić, że prawdopodobieństwa przejścia do klasy pierwszej są takie same dla osób o różnym stopniu wykształcenia. Wraz z lepszym stopniem wykształcenia wzrasta prawdopodobieństwo przejścia do klasy trzeciej (z klasy zarówno 1, jak i 2), a spada prawdopodobieństwo przejścia do klasy drugiej (z klasy zarówno 3, jak i 1).

4. Zakończenie

W artykule przedstawiono przykład zastosowania ukrytego modelu Markowa do oceny wiarygodności instytucji finansowych oraz instytucji państwa w latach 2007-2013. Wyodrębniono trzy klasy o podobnym stopniu zaufania wśród Polaków. Oszacowano również prawdopodobieństwa przejścia do wyodrębnionych

klas. Zbadano także wpływ zmiennych towarzyszących na oszacowane prawdopodobieństwa przejścia (prawdopodobieństwa zmiany nastawienia do instytucji zaufania publicznego i finansowego). W dalszych badaniach analizie poddanej zostaną również ukryte modele Markowa z uwzględnieniem zmiennych towarzyszących stałych i zmiennych w czasie t (*time-constant, time-varying covariates*) [Vermunt 2010].

Literatura

- Agresti A., 2002, *Categorical Data Analysis*, John Wiley&Sons, Hoboken.
- Akaike H., 1974, *A new look at statistical model identification*, IEEE Transactions on Automatic Control, 19, s. 716-723.
- Cappe O., Moulines E., Ryden T., 2005, *Inference in Hidden Markov Models*, Springer-Verlag, New York.
- Chung H., Walls T., Park Y., 2007, *A latent transition model with logistic regression*, Psychometrika, 72, s. 413-435.
- Collins L.M., Lanza S.T., 2011, *Latent Class and Latent Transition Analysis with Applications in the Social, Behavioral, and Health Sciences*, John Wiley&Sons, Wiley.
- Dayton C.M., Macready G.B., 1988, *Concomitant-variable latent-class models*, Journal of the American Statistical Association, 83(401), s. 173-178.
- Diagnoza społeczna 2013. Warunki i jakość życia Polaków (raport)*, Czapiński J., Panek T. (red.), Warszawa, Rada Monitoringu Społecznego (22.08.2013).
- Dempster A.P., Laird N.P., Rubin D.B., 1977, *Maximum likelihood for incomplete data via the EM algorithm (with discussion)*, Journal of the Royal Statistical Society, no. 39, ser. B, s. 1-38.
- Frühwirth-Schnatter S., 2006, *Finite Mixture and Markov Switching Model*, Springer-Verlag, New York.
- Genge E., 2013, *A latent class analysis of the public attitude towards the euro adoption in Poland*, Advances in Data Analysis and Classification, Springer 2013, s. 1-16.
- Ghysels E., 1994, *On the periodic structure of the business cycle*, Journal of Business and Economic Statistics, no. 12 (3), s. 289-298.
- Hagenaars A.J., McCutcheon A.L., 2002, *Applied Latent Class Analysis*, Cambridge University Press, Cambridge.
- Kim C.J., 1994, *Dynamic linear models with Markov-switching*, Journal of Econometrics, vol. 60, s. 1-22.
- Krogh A., 1998, *An introduction to hidden Markov models for biological sequences*, [w:] Salzberg S.L., Searls D.B., Kasif S. (red.), *Computational Methods in Molecular Biology*, Elsevier, Amsterdam, s. 45-63.
- Podgórska M., Śliwka P., Topolewski M., Wrzosek M., 2000, *Łańcuchy Markowa w teorii i zastosowaniach*, SGH, Warszawa.
- Rabiner L.R., 1989, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of IEEE, vol. 77, no 2, s. 267-295.
- Schmittmann V.D., Visser I., Raijmakers M.E.J., 2006, *Multiple learning modes in the development of rule-based category-learning task performance*, Neuropsychologia, 44 (11), s. 2079-2091.
- Schwarz G., 1978, *Estimating the dimension of a model*, Annals of Statistics, 6, s. 461-464.
- Stawicki J., 2004, *Wykorzystanie łańcuchów Markowa w analizie rynku kapitałowego*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

- Szumlicz T., 2013, *Stan społeczeństwa obywatelskiego. Stosunek do instytucji finansowych. Diagnoza Społeczna 2013. Warunki i Jakość Życia Polaków – Raport* [Special issue], Contemporary Economics, 7, 310-316.
- Vermunt J.K., Langeheine R., Böckenholt U., 1999, *Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates*, Journal of Educational and Behavioral Statistics, 24, s. 178-205.
- Vermunt J.K., 2010, *Longitudinal Research Using Mixture Models*, [w:] K. van Montfort, J.H.L. Oud, A. Satorra (red.), *Longitudinal research with latent variables*, Springer, Berlin Heidelberg, 119-152.
- Visser I., Speekenbrink M., 2010, *depmixS4: An R Package for Hidden Markov Models*, Journal of Statistical Software, vol. 36, no. 7, s. 1-21.
- Visser I., 2011, *Seven things to remember about hidden Markov models: a tutorial on Markovian models for time series*, Journal of Mathematical Psychology, 55(6), s. 403-415.
- Wickens T.D., 1982, *Models for Behavior: Stochastic Processes in Psychology*, W.H. Freeman and Company, San Francisco.
- Zucchini W., MacDonald I., 2010, *Hidden Markov Models for Time Series: An Introduction Using R*, Monographs on Statistics and Applied Probability, Boca Raton, CRC Press.

TRUST TO THE PUBLIC AND FINANCIAL INSTITUTIONS IN THE POLISH SOCIETY – AN APPLICATION OF LATENT MARKOV MODELS

Summary: The latent Markov model is a variation of the latent class model that is applied to estimate not only the prevalence of latent class membership, but the incidence of transitions over time in latent class membership. We used model-based clustering approach for grouping and detecting inhomogeneities of the Polish attitudes to public and financial institutions. We focused especially on the latent Markov models with covariates which additionally allow us to investigate the dynamic pattern of trust to the Polish institutions for different demographic features. We analyzed data collected as part of the Polish Social Diagnosis using depmixS4 package of R.

Keywords: latent Markov model, mixture distributions, model-based clustering.