

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 336

Badania marketingowe – nowe podejścia oraz metody na współczesnym rynku

Redaktorzy naukow
Krystyna Mazurek-Łopacińska
Magdalena Sobocińska



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Joanna Szynal
Redaktor techniczny: Barbara Łopusiewicz
Korektor: K. Halina Kocur
Łamanie: Małgorzata Czupryńska
Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon
http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

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

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2014

ISSN 1899-3192
ISBN 978-83-7695-476-9

Wersja pierwotna: publikacja drukowana

Druk i oprawa:
EXPOL, P. Rybiński, J. Dąbek, sp.j.
ul. Brzeska 4, 87-800 Włocławek

Spis treści

Wstęp	9
Krystyna Mazurek-Łopacińska, Magdalena Sobocińska: Badania kodów kulturowych w tworzeniu marketingowych modeli biznesu	11
Anna Olejniczuk-Merta: Rozwój innowacji społecznych a badania marketingowe	22
Monika Hajdas: Techniki pomiaru kompatybilności marek i idei kulturowych.....	31
Paweł Chlipała: Triangulacja podejść metodologicznych w badaniach naukowych z dziedziny marketingu	39
Adam Sagan: Analiza rzetelności skal w wielopoziomowych modelach pomiaru	49
Adam Sagan, Mariusz Łapczyński: Modele hybrydowe CART-logit w analizie procesu podejmowania decyzji w gospodarstwie domowym.....	60
Anna Myrda: Segmentacja łańcuchów środków-celów: miary podobieństwa sekwencji i ilościowe wskaźniki jakości grupowania a wyniki grupowania	70
Kamila Pilch: Asymetryczne skalowanie wielowymiarowe w wizerunkowych badaniach jednostek terytorialnych.....	79
Alicja Kusińska: Analizy wielowymiarowe jako źródło wiedzy o zachowaniach konsumentów na rynku	89
Krzysztof Błoński: Wykorzystanie metod wielowymiarowych do analizowania związku między emocjami a satysfakcją klienta	99
Sylwester Białowas, Iwona Olejnik: Poziom opiekuńczości państwa a zachowania oszczędnościowe – analiza wielowymiarowa	110
Jadwiga Stobiecka: Interpretacyjne konsekwencje oceny stabilności opinii respondentów w badaniach konsumpcji, oszczędzania i inwestowania.....	118
Piotr Tarka: HOMALS – wielowymiarowa analiza korespondencji jako metoda konstrukcji skali pomiarowej w badaniach marketingowych.....	129
Lukasz Skowron: Zastosowanie modelowania ścieżkowego do wyznaczenia przebiegu procesu lojalnościowego wśród klientów lubelskich centrów handlowych.....	140
Ireneusz P. Rutkowski: Metody CMMI i SGMM oceny dojrzałości procesu innowacji i wprowadzania produktu na rynek.....	152
Hanna Hall: Nowy konsument a zmiany w metodach jego badania.....	163
Tomasz Olejniczak: Techniki badawcze wykorzystywane w badaniu cyklu życia gospodarstwa domowego	174

Anna Dąbrowska, Arkadiusz Wódkowski: Kompetencje konsumentów w świetle badań ilościowych	185
Sylwia Makomaska: Wpływ muzyki tła na reakcje konsumentów w miejscu sprzedaży – problematyka interdyscyplinarności badań	195
Lucyna Witek: Metodyczne aspekty badania postaw konsumentów (na przykładzie rynku produktów ekologicznych)	205
Magdalena Olejniczak: Zróżnicowanie technik badawczych w badaniu motywacji zakupowych konsumentów żywności funkcjonalnej.....	215
Agata Dziakowicz: Metody badań marketingowych na rynku dóbr luksusowych.....	224
Wanda Patrzałek, Aleksandra Perchla-Włosik: Zastosowanie analizy semiologicznej w badaniach wpływu mody na zachowania młodych konsumentów	233
Agata Stolecka-Makowska: Zastosowanie podejścia interpretacyjnego w badaniu zmian zachowań nabywczych konsumentów podlegających akulturacji	244
Arkadiusz Wódkowski: Zmiana paradygmatu w marketingowych badaniach jakościowych?.....	257
Grzegorz Maciejewski: Zogniskowane wywiady grupowe w badaniach zachowań młodych dorosłych na rynku usług finansowych	266
Iga Rudawska: Zastosowanie zogniskowanego wywiadu grupowego do oceny jakości obsługi pacjentów przewlekle chorych	275
Ewa Nowakowska, Adam Sagan: Kontrfaktyczno-porównawcze studium przypadku w marketingu usług zdrowotnych.....	284
Krzysztof Kapera, Mariusz Kuziak: Skuteczność wybranych metod komunikacji z respondentami w badaniach internetowych	296
Iwona Escher: Niejednoznaczność statusu metodologicznego internetowego wywiadu grupowego i jego poszczególnych odmian	310
Magdalena Daszkiewicz, Sylwia Wrona: Zogniskowane wywiady grupowe online jako alternatywa dla tradycyjnych metod gromadzenia danych – szanse rozwoju i wyzwania dla badaczy	321
Olgierd Witczak: Potencjał wykorzystania serwisów społecznościowych w badaniach jakościowych	331
Agnieszka Dejnaka: Facebook jako obszar prowadzenia badań marketingowych.....	339
Robert Wolny: Możliwości wykorzystania obserwacji w Internecie w badaniach rynku e-usług.....	348

Summaries

Krystyna Mazurek-Łopacińska, Magdalena Sobocińska: Research of cultural codes in creating marketing models of business	21
---	----

Anna Olejniczuk-Merta: The development of social innovation and marketing research.....	30
Monika Hajdas: Techniques for measuring the compatibility of brands and cultural ideas.....	38
Paweł Chlipała: Triangulation of methodological approaches in scientific research of marketing field	48
Adam Sagan: Reliability analysis in multilevel measurement models	59
Adam Sagan, Mariusz Łapczyński: CART-logit hybrid models in the analysis of decision-making process in the households	69
Anna Myrda: Segmentation of Means-End Chains: sequence dissimilarity measures and quantitative cluster validity indexes vs. clustering results...	78
Kamila Pilch: Asymmetric multidimensional scaling in the research of territorial units image	88
Alicja Kusińska: Multidimensional analysis as a source of knowledge about consumer behaviour.....	98
Krzysztof Błoński: The use of multidimensional methods to analyze the relationship of emotions and customer satisfaction.....	109
Sylwester Białowąs, Iwona Olejnik: The level of the state's social security and its influence on saving behaviour – multidimensional analysis.....	117
Jadwiga Stobiecka: Interpretative consequences of the assessment of respondents' opinions stability in the studies of consumption, saving and investing.....	128
Piotr Tarka: HOMALS – multiple correspondence analysis as the method for measurement scale construction in marketing research.....	139
Lukasz Skowron: The usage of the Structural Equation Modeling for determining the loyalty building process among the customers of the shopping centers located in Lublin.....	151
Ireneusz P. Rutkowski: CMMI and SGMM methods of maturity evaluation of the product innovation process and introduction of a product on the market	162
Hanna Hall: New consumer and changes in the methods of their research....	173
Tomasz Olejniczak: Research techniques used in the study of the household life cycle.....	184
Anna Dąbrowska, Arkadiusz Wódkowski: Consumer competences in the light of quantitative research	194
Sylwia Makomaska: The effects of background music on consumers response in the place of commerce – the problem of interdisciplinary research	204
Lucyna Witek: Methodical aspects of research of consumers' attitudes (on the example of market of green products).....	214
Magdalena Olejniczak: The diversity of marketing research techniques in the study of purchase motivation of functional food consumers.....	223

Agata Dziakowicz: Methods of marketing research for the luxury goods market	232
Wanda Patrzalek, Aleksandra Perchla-Wlosik: Studies of the impact of fashion on the behavior of young consumers using semiological analysis	243
Agata Stolecka-Makowska: The use of an interpretative approach in a study of purchasing behaviour changes of consumers acculturation	256
Arkadiusz Wódkowski: Change of paradigm in qualitative market research?	265
Grzegorz Maciejewski: The focus group interview in the research of the young-adults behaviors on the financial services market	274
Iga Rudawska: The use of focus group interview to evaluate the service quality of chronically ill patients	283
Ewa Nowakowska, Adam Sagan: Comparative-counterfactual case research in health service marketing	295
Krzysztof Kapera, Mariusz Kuziak: Effectiveness of selected methods of communication with respondents in online surveys	309
Iwona Escher: The ambiguity of the methodological status of <i>online group interview</i> and its particular types	320
Magdalena Daszkiewicz, Sylwia Wrona: Online focus group interviews as an alternative for traditional methods of data collection – opportunities for development and challenges to researchers	330
Olgierd Witczak: The potential of using social network sites in qualitative research	338
Agnieszka Dejnaka: Facebook as a marketing research area	347
Robert Wolny: The use of Internet observations in research of e-services market	357

Anna Myrda

Uniwersytet Ekonomiczny w Krakowie

SEGMENTACJA ŁAŃCUCHÓW ŚRODKÓW-CELÓW: MIARY PODOBIEŃSTWA SEKWENCJI I ILOŚCIOWE WSKAŹNIKI JAKOŚCI GRUPOWANIA A WYNIKI GRUPOWANIA

Streszczenie: Łańcuchy środków-celów zbudowane z jakościowych i sekwencyjnych danych opisujących struktury poznawczo-motywacyjne nabywców można wykorzystać podczas segmentacji rynku. Celem analiz było sprawdzenie, czy i jak miary podobieństw sekwencji wpływają na właściwości wybranych ilościowych wskaźników jakości grupowania oraz na końcowe wyniki eksploracyjnej analizy skupień. W grupowaniu 482 drabinek użyto następujących miar podobieństwa sekwencji: Hamminga, najdłuższej wspólnej podsekwencji, najdłuższego wspólnego przyrostka i optymalnego dopasowania oraz 3 wskaźników jakości grupowania: Silhouette, Bakera, Huberta oraz Huberta i Levina. Na podstawie indeksów jakości grupowania za najlepszy uznano model z 8 grupami, bez względu na sposób wyznaczania podobieństwa między obiektami.

Słowa kluczowe: segmentacja rynku, łańcuchy środków-celów, analiza skupień, miary niepodobieństwa sekwencji, wskaźniki jakości grupowania.

DOI: 10.15611/pn.2014.336.07

1. Wstęp

Dotarcie do motywów stojących za konsumenckimi wyborami jest kluczowe dla opracowywania strategii marketingowych. Teoria środków-celów (*means-end theory*) dostarcza modelu struktur poznawczo-motywacyjnych, który ułatwia zrozumienie zachowań konsumenckich, oraz zestawu narzędzi – wywiadów drabinkowych – pozwalających na docieranie do tych struktur. Dane zebrane za pomocą wywiadów drabinkowych mogą posłużyć do segmentacji rynku. Dane te są jakościowe i sekwencyjne, dlatego opracowywanie ich za pomocą analizy skupień wymaga zastosowania odpowiednich miar odległości oraz rozwiązania problemu wyboru optymalnej liczby grup. Celem analiz zaprezentowanych w głównej części artykułu było sprawdzenie, czy i w jakim stopniu miary podobieństwa sekwencji wpływają na wybrane ilościowe wskaźniki jakości grupowania oraz na końcowe wyniki grupowania.

2. Teoria środków-celów i wywiady drabinkowe

Wiedza konsumentów na temat produktów dotyczy trzech obszarów. Pierwszy z nich to cechy (atrybuty A) produktów, drugi to skutki używania (konsekwencje K) produktów, a trzeci to cele (wartości W), które można realizować dzięki produktowi. Teoria środków-celów zakłada, że istnieją między nimi hierarchiczne powiązania: atrybut \rightarrow konsekwencja \rightarrow wartość [Olson, Reynolds 2001, s. 13], nazywane łańcuchami środków-celów, gdyż konsumenci postrzegają produkty i ich właściwości jako środki do osiągnięcia własnych celów [Reynolds, Dethloff, Westberg 2001, s. 92-93]. Nabywcy uczą się łączyć cechy produktów z wynikającymi z nich pozytywnymi lub negatywnymi konsekwencjami. Konsekwencje te są ważne dla nabywców, ponieważ są zgodne (lub niezgodne) z wartościami osobowymi, jakimi nabywcy się kierują [Kąciak 2011, s. 43].

W teorii środków-celów model struktur poznawczo-motywacyjnych należy do podejść typu „góra – dół”, natomiast wywiady drabinkowe prowadzą „od dołu do góry” [Domurat 2009, s. 189]. W trakcie wywiadu konsument najpierw wskazuje atrybuty produktu, które uznaje za ważne, następnie podaje konsekwencje wynikające z tych atrybutów, a na końcu dociera do powodów (wartości osobowych) sprawiających, że wymienione na etapie 2 konsekwencje są istotne [Reynolds, Gutman 2001 s. 26-28]. W efekcie otrzymuje się drabinki A-K-W: sekwencje zbudowane ze zmiennych jakościowych. W trakcie analiz danych z wywiadów drabinkowych ważne jest uwzględnienie sekwencyjności, a pomijanie tego elementu skutkuje utratą części istotnych informacji.

3. Segmentacja łańcuchów środków-celów

Dzięki danym z wywiadów drabinkowych możliwe jest dokonanie segmentacji rynku nie tylko na podstawie cech klientów czy oferty, ale relacji zachodzącej między jednym a drugim. Kąciak [2011] zwraca uwagę na to, że brak w literaturze przedmiotu procedur segmentacji, które można by zastosować do danych pochodzących z wywiadów drabinkowych, i proponuje metodę składającą się ze skalowania wielowymiarowego i dwustopniowego grupowania. W tej pracy przedstawione zostało podejście, w którym dane sekwencyjne są grupowane za pomocą analizy skupień metodą k-medoidów.

Jednym z problemów, jaki należy rozwiązać, używając analizy skupień, jest wybór miary odległości [Walesiak 2009, s. 409]. Miary odległości (niepodobieństwa) przeznaczone dla danych sekwencyjnych można podzielić na te, których konstrukcja opiera się na zliczaniu pasujących do siebie elementów dwóch sekwencji, oraz na miary bazujące na wyznaczaniu kosztów przekształcania jednej sekwencji w drugą [Gabadinho i in. 2011, s. 24]. Im niższe wartości przyjmuje dana miara, tym bardziej podobne są porównywane sekwencje. W analizach opisanych w dalszej części artykułu skorzystano z 4 miar odległości sekwencji: Hamminga (HAM), najdłuższej

wspólnej podsekwencji (LCS), najdłuższego wspólnego przyrostka (RLCP) oraz optymalnego dopasowania (OM).

Najdłuższa wspólna podsekwencja to wszystkie wspólne elementy dwóch sekwencji następujące w tej samej kolejności, niekoniecznie bezpośrednio po sobie [Elzinga 2008, s. 14-15]. Im więcej takich samych elementów występujących w tej samej kolejności w porównywanych sekwencjach, tym bardziej dwie sekwencje są do siebie podobne. Najdłuższy wspólny przyrostek to wszystkie wspólne elementy następujące bezpośrednio po sobie, liczone od tyłu porównywanych sekwencji [Gabadinho i in. 2011, s. 25]. Aby dwie sekwencje zostały uznane za podobne, muszą kończyć się tym samym elementem: jeśli para sekwencji posiada wspólne elementy, ale ich ostatnie elementy są różne, to zostaną one uznane za niepodobne. Miary LCS i RLCP są wyznaczone według wzoru [Gabadinho i in. 2011, s. 25]:

$$d(x, y) = |x| + |y| - 2A(x, y), \quad (1)$$

gdzie: $d(x, y)$ – miara niepodobieństwa,

$|x|$ – długość sekwencji x (liczba elementów sekwencji x),

$|y|$ – długość sekwencji y (liczba elementów sekwencji y),

$A(x, y)$ – najdłuższa wspólna podsekwencja (LCS), najdłuższy wspólny przyrostek (RLCP).

Optymalne dopasowanie (OM) to zestaw miar niepodobieństwa sekwencji pochodzących z pracy Levenshteina [1966]. W procedurze OM stopień podobieństwa dwóch sekwencji jest równy najmniejszemu kosztowi (najmniejszej liczbie operacji), koniecznemu do przekształcenia jednej z tych sekwencji w drugą. Operacje dopuszczane w trakcie przekształcania sekwencji to wprowadzenie elementu do sekwencji, usunięcie elementu z sekwencji oraz substytucja [Gabadinho i in. 2009, s. 97-98]. Z każdą z tych operacji są związane koszty: Levenshtein [1966, s. 707] każdą podstawową operację obciążał jednakowym kosztem równym 1, koszty można jednak wycenić inaczej. W tej pracy przyjęto, że koszt wprowadzenia oraz usunięcia elementu jest równy 1, natomiast koszt substytucji elementu i elementem j jest równy 2.

Miara Hamminga (HAM) [Hamming 1950] jest wyznaczana jako liczba operacji substytucji koniecznych do przekształcenia jednej sekwencji w drugą. Konstrukcja miary Hamminga sprawia, że można z niej korzystać, jeśli analizuje się zestaw sekwencji o tej samej długości. W pracy Hamminga koszt substytucji był zawsze równy 1, w prezentowanych tutaj analizach też tak wyceniono koszt substytucji.

Ustalenie liczby grup, na jakie należy podzielić zbiór, to kolejny problem związany z analizą skupień [Walesiak 2009, s. 409]. W literaturze przedmiotu można znaleźć wiele ilościowych wskaźników jakości grupowania, stworzonych po to, aby ułatwić wybór najlepszej liczby grup (np. [Milligan, Cooper 1985; Rousseeuw 1987; Migdał-Najman, Najman 2005; Steinley 2006]). W tej pracy użyto indeksu Bakera

i Huberta, indeksu Huberta i Levine'a oraz indeksu Silhouette Rousseeuw. Najlepszy podział zbioru to ten spośród rozważanych, dla którego miara Huberta i Levine'a jest najniższa, a miary Bakera i Huberta i Silhouette – najwyższe [Walesiak 2009, s. 418].

4. Eksploracyjna analiza skupień łańcuchów środków-celów

Analizowane dane zebrano w trakcie projektu badawczego na temat palenia papierosów, współrealizowanego przez Katedrę Analizy Rynku i Badań Marketingowych UEK. Grupowano zbiór składający się z 482 łańcuchów środków-celów. Do analizy wybrano pełne (trzyelementowe) drabinki środków-celów – wylosowano około 40% drabinek spośród ponad 1200 sekwencji A-K-W wygenerowanych przez respondentów. W danych było 8 cech (atrybutów) papierosów, 11 konsekwencji i 9 wartości tworzących 29 różnych sekwencji A-K-W.

Dane opracowano za pomocą pakietów TraMineR, mclus i clusterSim programu R. Stopień podobieństwa wygenerowanych drabinek wyznaczono przy użyciu 4 opisanych we wprowadzeniu miar; dane były grupowane metodą k-centroidów. Przy użyciu każdej z 4 macierzy odległości zbudowano 9 modeli mających od 2 do 10 grup: dane były grupowane 36 razy. Dla każdego modelu wyliczono wartości 3 wskaźników jakości grupowania: Silhouette Rousseeuw, Bakera i Huberta oraz Huberta i Levine'a. Dla modeli o tej samej liczbie grup wyznaczonych w oparciu o różne macierze odległości wyliczono wartości skorygowanego indeksu Randa [Hubert, Arabie 1985]. Wskaźnik ten posłużył do oceny stopnia zgodności podziału zbioru na jednakową liczbę grup, gdy podobieństwo między obiektami było wyznaczane na różne sposoby.

Wyniki analiz, w których do ocenienia stopnia podobieństwa sekwencji użyto miar HAM, LCS i OM, były identyczne. Skorygowany indeks Randa dla modeli o kolejnych liczbach grup wyznaczanych na podstawie tych odległości był równy 1. Wartości użytych wskaźników jakości grupowania były takie same dla modeli z poszczególną liczbą grup, bez względu na to, którą z tych metod wyznaczano podobieństwo między obiektami. Odmienne wyniki dawały analizy oparte na mierze RLCP. Stopień zgodności podziału zbioru na podstawie tej miary i trzech pozostałych, mierzony skorygowanym indeksem Randa, wahał się od $-0,04$ do $0,389$. Najniższy był dla podziału zbioru na dwie części, a najwyższy, kiedy wyodrębniano 3 i 10 grup. Tabela 1 zestawia wartości miar użytych w analizie.

Dla modeli zbudowanych w oparciu o odległości HAM, LCS i OM indeks Silhouette rósł z podziału na podział. Najwyższą wartość osiągnął podczas wyodrębnienia 8 grup, a następnie spadał. Kiedy podstawą analiz była macierz odległości wyznaczona metodą RLSC, indeks *Silhouette* rósł wraz z liczbą grup, na jakie dzielono zbiór. Osiągnął maksimum dla podziału na 9 grup, a następnie spadł.

Tabela 1. Wartości wskaźników jakości grupowania dla modeli różniących się liczbą grup i zbudowanych w oparciu o cztery miary odległości sekwencji oraz skorygowanego indeksu Randa wyznaczonego między grupowaniami z użyciem miary RLCS a pozostałymi miarami

Indeks	Miara odległości sekwencji	Liczba grup								
		2	3	4	5	6	7			
Silhouette	HAM, LSC, OM	0,254	0,315	0,396	0,459	0,591	0,596	0,648	0,63	0,546
	RLCP	0,208	0,285	0,37	0,429	0,485	0,534	0,582	0,636	0,602
Huberta i Levina	HAM, LSC, OM	0,764	0,682	0,563	0,538	0,349	0,314	0,279	0,258	0,244
	RLCP	0,831	0,801	0,668	0,595	0,547	0,511	0,483	0,695	0,646
Bakera i Huberta	HAM, LSC, OM	0,606	0,704	0,867	0,882	0,987	0,985	0,994	0,995	0,99
	RLCP	1	1	1	1	1	1	1	1	1
Skorygowany Randa		-0,04	0,389	0,251	0,261	0,241	0,194	0,22	0,286	0,389

Źródło: opracowanie własne.

Wartości indeksu Huberta i Levine’a spadały podczas dodawania kolejnych grup i najniższy był dla modelu z 10 skupieniami, kiedy modele powstawały w oparciu o odległości HAM, OM i LCS. Największy spadek wartości tej miary zaobserwowano przy przejściu od 5 do 6 grup. Gdy podstawą grupowania była macierz zbudowana metodą RLCS, indeks HL spadał dla kolejnych modeli – najniższy był, gdy zbiór dzielono na 8 grup, następnie wzrósł.

Dla wszystkich modeli wyodrębnionych na podstawie miary RLCS indeks Bakera i Huberta był równy 1. Gdy w analizie korzystano z miar HAM, OM i LCS, wartość tego wskaźnika była najwyższa dla podziału zbioru na 9 grup. Podczas przechodzenia od modelu z 2 do modelu z 6 grupami wystąpiły wyraźne wzrosty wartości indeksu BH, po dodaniu kolejnej grupy indeks nieznacznie spadł, po czym wzrósł, uzyskał maksimum dla podziału danych na 9 grup i spadł po dodaniu kolejnego skupienia.

5. Wybór optymalnej liczby grup

Miary odległości LCS, OM i HAM: na podstawie wartości indeksu Silhouette za najlepszy należałoby uznać podział zbioru na 8 grup. Przyjmuje się, że wartości tej miary z przedziału od 0,51 do 0,7 świadczą o tym, iż podział zbioru jest dobry [Kolenda 2006, s. 121], więc model z 8 grupami należałoby uznać za satysfakcjonujący. Indeks Bakera i Huberta był największy dla modelu z 9 grupami, a dla modelu z 8 był minimalnie niższy (0,001). Na podstawie wartości indeksu Huberta i Levine’a za najlepszy należałoby uznać podział zbioru na 10 grup. Największy spadek

tej miary odnotowano przy przejściu od modelu z 5 do modelu z 6 grupami. Na podstawie ilościowych wskaźników grupowania za optymalny można uznać model z 8 grupami: indeks Silhouette wyraźnie wskazuje, że ten podział jest najlepszy spośród rozważanych, na podstawie miary HB można uznać, że model z 8 grupami jest tylko nieznacznie gorszy od modelu z 9 grupami i sam w sobie dobry: wartość indeksu HB jest bliska 1. Natomiast miara HL jest dla modelu z 8 grupami o 0,485 niższa niż dla modelu z 2 grupami i tylko o 0,035 wyższa niż dla ostatniego podziału.

Miara RLCP: indeks Silhouette był najwyższy dla podziału zbioru na 9 części – na podstawie jego wartości można uznać ten podział zbioru za dobry. Zmienność wskaźnika Huberta i Levine'a sugeruje, że najlepszy jest model z 8 grupami. Natomiast wartości indeksu Bakera i Huberta wskazują, że wszystkie modele są jednako dobre. Indeks Huberta i Levine'a dla modelu z 9 grupami był najwyższy – na podstawie tej miary podział zbioru na 9 grup należy uznać za najgorszy ze wszystkich rozważanych. Natomiast indeks Silhouette wskazuje, że model z 8 grupami jest trzecim spośród 9 rozważanych pod względem optymalności. Jednocześnie jego wartość można interpretować jako wskazówkę, że wyodrębnienie 8 grup jest dobrym rozwiązaniem. Na podstawie tych dwóch wskaźników jakości grupowania można uznać, że model z 8 skupieniami jest najlepszy.

Tabela 2. Podział zbioru na 8 grup w oparciu o miary HMA, OM, LCS oraz RLCP: liczebności skupień, liczba atrybutów, konsekwencji i wartości tworzących drabinki w poszczególnych skupieniach

HAM, LCS, OM				RLCP			
Nr grupy (liczebność)	cecha	korzyść	wartość	Nr grupy (liczebność)	cecha	korzyść	wartość
1 (134)	1	1	6	1 (38)	1	1	1
2 (44)	1	2	1	2 (44)	1	2	1
3 (56)	3	4	2	3 (161)	7	9	3
4 (37)	1	3	1	4 (22)	1	1	1
5 (81)	1	2	3	5 (27)	1	1	1
6 (55)	2	3	2	6 (38)	3	3	1
7 (26)	1	1	2	7 (72)	2	1	1
8 (49)	1	2	2	8 (80)	4	5	1

Źródło: opracowanie własne.

Na podstawie wszystkich zgromadzonych danych można za najlepszy uznać model z 8 grupami. Jednocześnie podział zbioru na osiem części w oparciu o macierz zbudowaną za pomocą RLCP i za pomocą pozostałych miar podobieństwa sekwencji znacząco się różni. Gdy podstawą grupowania była macierz wygenerowana metodą HAM, LCS lub OM, podczas podziału zbioru na 8 części wyodrębniono skupienia o następującej liczebności: 134, 44, 56, 37, 81, 55, 26, 49. Największa grupa zawierała drabinki rozpoczynające się od cechy papierosów „tanie”, wskazy-

waną korzyścią była „oszczędność”, która była istotna dzięki zgodności z jedną z 6 możliwych wartości osobowych (m.in. hedonizm). Natomiast podczas grupowania z użyciem odległości RLCS skupienia miały następujące liczebności: 38, 44, 161, 22, 27, 38, 72, 80. W największej grupie znalazły się zróżnicowane drabinki, mogące rozpoczynać się od jednej z 7 cech, z których wynika jedna z 9 konsekwencji prowadząca do jednej z 3 wartości. Jednocześnie wyodrębnione zostały 3 skupienia składające się z jednorodnych drabinek (jeden atrybut, jedna konsekwencja, jedna wartość), do których trafiło łącznie 87 obserwacji. Żadne tego typu skupienie nie zostało wyodrębnione podczas grupowania opartego na macierzach wygenerowanych za pomocą miary HAM, LCS lub OM. Tabela 2 zestawia liczebności skupień w dwóch modelach z 8 grupami oraz liczbę atrybutów, konsekwencji i wartości, z których zbudowane były drabinki w poszczególnych skupieniach. Z braku miejsca szczegółowy opis tych modeli nie zostanie zamieszczony.

6. Wnioski z analizy

Wskazania co do najlepszej liczby grup były jaśniejsze, gdy podstawą grupowania były macierze odległości wyznaczone metodami HAM, OM i LCS, niż gdy stopień podobieństwa sekwencji wyliczono metodą RLSC. W tym drugim przypadku indeks Bakera i Huberta okazał się nieprzydatny – jego wartość dla wszystkich rozważanych modeli był stała. Ogólnie wartości miary Silhouette były wyższe, a wartości miary Huberta i Levine’a niższe, gdy analizę prowadzono z użyciem miar HAM, OM i LCS niż miary RLCS. Różnice te są szczególnie duże w przypadku indeksu HL: najlepszy model zbudowany na podstawie odległości RLCS można na podstawie wartości tej miary uznać za co najwyżej przeciętny ($HL=0,483$), a najlepszy model zbudowany na podstawie jednej z trzech pozostałych miar podobieństwa sekwencji można uznać za całkiem dobry ($HL=0,244$). Natomiast najwyższe wartości miary Rousseeuw wskazują na dobry podział zbioru bez względu na to, z której miary podobieństwa sekwencji skorzystano.

Bez względu na to, jaka macierz odległości stanowiła podstawę grupowania, za najlepszy można uznać model z 8 grupami. Jednak podziały zbioru na osiem części w oparciu o macierz zbudowaną za pomocą RLCP i za pomocą pozostałych miar podobieństwa sekwencji znacznie się różnią.

7. Zakończenie

Przesłanki dotyczące najlepszej liczby grup płynące z użytych wskaźników jakości grupowania wskazywały, że najlepszy będzie podział zbioru na 8 części bez względu na to, którą z miar odległości wykorzystano w analizie. Jednak dla miar HAM, OM i LCS były one jaśniejsze niż dla miary RLCS. Sam podział zbioru na 8 grup znacznie się różnił, gdy odległości między sekwencjami wyznaczano metodą HA, LCS, OM i RLCS. Można przypuszczać, że różnice te są efektem konstrukcji miar

odległości i tego, że podział zbioru był w pewnym stopniu wymuszony – obiekty nie tworzą 8 wyraźnie odrębnych grup.

Dane uzyskane z wywiadów drabinkowych mają charakter sekwencyjny – uwzględnienie ich sekwencyjności w trakcie analiz jest niezbędne do uzyskania pełnego obrazu prawidłowości dotyczących zbadanych zachowań konsumenckich. Dane tego typu można wykorzystać w trakcie segmentacji rynku za pomocą popularnej analizy skupień – w takim przypadku konieczne jest skorzystanie z miary odległości stworzonej z myślą o danych sekwencyjnych. Przeprowadzone analizy pokazują, że konstrukcja użytej do analizy miary podobieństwa sekwencji może wpływać na ostateczny wynik grupowania oraz modyfikować użyteczność wskaźników jakości grupowania. Zagadnienia te wymagają dalszych badań.

Literatura

- Domurat A., *Identyfikacja wartości osobowych w badaniach psychologicznych. Wartości jako cele działań i wyborów*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2009.
- Elzinga C.H., *Sequence analysis: metric representations of categorical time series*, 2008, <http://home.fsw.vu.nl/ch.elzinga/MetricsRevision.pdf> [dostęp: 22.02.2013].
- Gabardin A., Ritschard G., Müller N.S., Studer M., *Analyzing and Visualizing State Sequences in R with TraMineR*, „Journal of Statistical Software” 2011, no. 40(4), s. 1-37.
- Gabardin A., Ritschard G., Studer M., Muller N., *Mining sequence data in R with the TraMineR package: A user's guide*, 2009. <http://mephisto.unige.ch/pub/TraMineR/doc/1.2/TraMineR-1.2-Users-Guide.pdf> [dostęp: 23.02.2013].
- Hamming R.W., *Error-Detecting and Error-Correcting Codes*, „Bell System Technical Journal” 1950, no. 26, s. 147-160.
- Hubert L., Arabie P., *Comparing partitions*, „Journal of Classification” 1985, no. 2, s. 193-218.
- Kąciak E., *Teoria środków – celów w segmentacji rynku. Studium metodologiczno -empiryczne*, Oficyna a Wolters Kluwer business, Warszawa 2011.
- Kolenda M., *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*, Wydawnictwo AE, Wrocław 2006.
- Lesnard L., *Optimal Matching and Social Sciences*, Série des Documents de Travail du CREST, Institut National de la Statistique et des Etudes Economiques, Paris 2006.
- Levenshtein V. I., *Binary codes capable of correcting deletions, insertions and reversals*, „Soviet Physics Doklady” 1966, no. 10 (8), s. 707-710.
- Migdał-Najman K., Najman K., *Analityczne metody ustalania liczby skupień*, „Taksonomia” 12, *Klasyfikacja i analiza danych – teoria i zastosowanie*, red. K. Jajuga, M. Walesiak, Wydawnictwo AE, Wrocław 2005, s. 264-265.
- Milligan G.W., Cooper M.C., *An examination of procedures for determining the number of clusters in data set*, „Psychometrika” 1985, vol. 50, no. 2, s. 159-179.
- Olson J.C., Reynolds T.J., *The Means-End Approach to Understanding Consumer Decision Making*, [w:] T.J. Reynolds, J.C. Olson (red.), *Understanding Consumer decision Making. The Means-End Approach to Marketing Decision Making and Advertising Strategy*, Mahwah, Lawrence Earlbaum Associates, New York 2001, s. 3-20.
- Reynolds T.J., Dethloff C., Westberg S.J. *Advancements in Laddering*, [w:] T.J. Reynolds, J.C. Olson (red.), *Understanding Consumer decision Making. The Means-End Approach to Marketing Deci-*

- sion Making and Advertising Strategy*, Mahwah, Lawrence Earlbaum Associates, New York 2001, s. 91-118.
- Reynolds T.J., Gutman J., *Laddering Theory, Method, Analysis, and Interpretation*, [w:] T.J. Reynolds, J.C. Olson (red.), *Understanding Consumer decision Making. The Means-End Approach to Marketing Decision Making and Advertising Strategy*, Mahwah, Lawrence Earlbaum Associates, New York 2001, s. 25-62.
- Rousseeuw P.J. *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*, "Journal of Computational and Applied Mathematics" 1987, vol. 20, s. 53-65.
- Steinley D., *K-means clustering: a half-century synthesis*, „British Journal of Mathematical and Statistical Psychology" 2006, no. 59, s. 1-34.
- Walesiak M., *Analiza skupień*, [w:] *Statystyczna analiza danych z wykorzystaniem program R*, M. Walesiak, E. Gatnar (red.), PWN, Warszawa 2009, s. 407-433.

SEGMENTATION OF MEANS-END CHAINS: SEQUENCE DISSIMILARITY MEASURES AND QUANTITATIVE CLUSTER VALIDITY INDEXES VS. CLUSTERING RESULTS

Summary: The Means-End Chains are built with qualitative and sequential data describing the cognitive-motivational structures of consumers, which can be used for market segmentation. The objective of the study is to analyse whether and how sequence dissimilarities measurement affects the properties of the selected quantitative cluster validity indexes and the final results of exploratory cluster analysis. The following measures of sequence dissimilarity were used in the grouping of 482 ladders: Hamming, the longest common subsequence, the longest common suffix and optimum matching. S cluster validity indexes were used to measure clustering goodness: silhouette, index of Baker and Hubert and index of Hubert and Levin. Based on the cluster validity indexes the model with 8 groups was selected as the best irrespective of the sequence dissimilarity measure used in the analysis.

Keywords: market segmentation, Means-End Chains, cluster analysis, sequence dissimilarity measure, cluster validity index.