EKONOMETRIA ECONOMETRICS 4(54) • 2016

ISSN 1507-3866
e-ISSN 2449-9994

Anna Myrda
Cracow University of Economics
e-mail: ann.m.myrda@gmail.com

THE MEANS-END APPROACH IN MARKET SEGMENTATION – CLUSTERING OF LADDERING DATA*

PODEJŚCIE ŚRODKÓW-CELÓW W SEGMENTACJI RYNKU – ANALIZA SKUPIEŃ DANYCH DRABINKOWYCH

DOI: 10.15611/ekt.2016.4.05
JEL Classification: C14, C38, M31, M39

Summary: The objective of this paper was to present results of research on two methodological issues related to the clustering of laddering data. The first was the methods of aggregation of information from ladders generated by one respondent and the second was the measurement of ladders' dissimilarity. Two methods of aggregation of information from ladders were proposed, three sequence dissimilarity measures were presented and all combinations of them were used in the analysis. The clustered data originate from a research on Polish adolescents' online consumer behaviour wherein the means-end approach was used. 1004 high school students participated in the research. Data were clustered from 2 to 10 groups, six modes of analysis were used, thus 54 solutions were built. Solutions with the same number of groups were compared with the adjusted Rand index. Analysis indicates the influence of sequences dissimilarity measures and methods of aggregation of information from ladders on clustering results.

Keywords: means-end approach, laddering, clustering, sequences dissimilarity measures.

Streszczenie: Celem artykułu było zaprezentowanie wyników badania dotyczącego dwóch metodologicznych zagadnień związanych z grupowaniem danych drabinkowych: sposobów agregowania informacji pochodzących ze wszystkich drabinek wygenerowanych przez jednego respondenta oraz metod pomiaru niepodobieństwa drabinek. Zaproponowano dwie metody agregacji informacji z drabinek, przedstawiono trzy miary niepodobieństwa sekwencji, a następnie wszystkie ich kombinacje wykorzystano w trakcie grupowania. Analizowane dane pochodziły z badania dotyczącego zachowań konsumenckich młodzieży w Internecie. W badaniu uczestniczyło 1004 uczniów ostatnich klas publicznych szkół średnich z Małopolski. Dane były dzielone na od 2 do 10 grup. Wykorzystano sześć sposobów analizy powstałych z kombinacji dwóch metod agregacji danych z drabinek i trzech miar niepodobieństwa sekwencji; zbudowano 54 rozwiązania. Modele mające tę samą liczbę grup porównano za

* Project was financed by the National Science Center, Poland, decision no. DEC-2013/11/N/HS4/02891.

pomocą skorygowanego indeksu Randa. Zarówno użyte miary niepodobieństwa sekwencji, jak i metody agregacji danych znacznie wpływały na ostateczne wyniki grupowania.

Słowa kluczowe: teoria środków-celów, analiza skupień, laddering, miary niepodobieństwa sekwencji.

1. Introduction

Market segmentation is an important step in the process of creating marketing strategies [Kotler 2005], and cluster analysis is one of the most popular statistical tools for segmentation [Wedel, Kakamura 1998]. The effectiveness of segmentation strongly depends on the features of the consumers chosen to distinguish the segments [Walesiak 2000]. The means-end approach [Gutman 1982], has delivered a theoretical background and research tools that could be used for collecting complex data about consumers' motives and values. However, laddering data are rarely used as the basis of segmentation. Laddering data are qualitative and sequential and therefore they are more often used for describing segments than for identifying them.

In this paper two issues related to the clustering of laddering data are presented: the first is the aggregation of information from ladders generated by one respondent and the second is dissimilarity measurement of ladders. Although these two issues are key steps in the laddering data clustering process, little research on this subject has been carried out so far.

2. Laddering data in segmentation

2.1. The means-end approach and laddering in market segmentation

The means-end theory describes consumers' cognitive-motivational structures that affect their market choices [Gutman 1982; 1984]. Consumers' knowledge about products is ordered in three levels: attributes – the product's features and properties, consequences – effects of the product's usage and a wide range of personal values that can be achieved through the use of the product. These three levels of consumers' knowledge form the hierarchy: *attributes (A) ->consequences (C) ->values (V)* called the *means-end ladder* or the *means-end chain*[1], because consumers perceive products and their properties as a means to achieve their own goals [Olson, Reynolds 2001].

Consumers associate the characteristics of a product with the positive or negative consequences of its usage. These consequences are important for consumers because

[1] A means-end ladder is an *A-C-V* sequence generated by the respondent; a means-end chain is an *A-C-V* sequence constructed by the researcher on the basis of co-occurrence of attributes, consequences and values in the collected data [Olson, Reynolds 2001].

they are consistent (or inconsistent) with personal values which control market behaviour [Gutman 1982; Reynolds, Gutman 2001, p. 62]. The product's attributes are not significant in themselves, they become important in the context of the perceived consequences of the product's usage and these consequences are evaluated positively or negatively through the system of personal values, which is the core of the human cognitive-motivational structure [Gutman 1982].

Means-end ladders can be measured by laddering, which is a set of research techniques developed on the basis of the means-end theory. Two types of ladder interviews are distinguished: soft and hard [Grunert, Grunert 1995]. Soft laddering is similar to an individual in-depth interview with a standardized list of seeking information, and hard ladder interviews are paper and pencil methods. Data from ladder interviews are qualitative and sequential: every *A-C-V* ladder created by the respondent is a sequence of categorical variables and has to be taken into account during the analysis – the omission of this results in the loss of some relevant information [Domurat 2004].

The means-end approach has four main applications in marketing: it is used for preparing market segmentation, for product planning and product developing, for creating promotion strategies and for researching consumer behaviour [Gutman 1982; Reynolds, Gutman 2001]. Despite the fact that market segmentation is one of the four main fields where the means-end approach is used, laddering data are hardly ever used as the basis of segmentation [Kąciak 2011]. In most of the surveys described in the literature, laddering data were not used for identifying groups of consumers but for discovering the most frequent means-end chains (e.g. [Grunert 1995; Mentzer et al. 1997; Gruber et al. 2008; Reppel et al. 2006]). Some authors distinguished segments on the basis of other consumers' characteristics and after that used laddering data for describing groups (e.g. [Bottshen et al. 1999; Reynolds, Rohon 2001; Fotopoulos 2003; Baker et. al 2004]). Single variables from *A-C-V* ladders or fragments of *A-C-V* ladders were used for identifying segments in some approaches [Bottshen et al. 1999], however full ladders composed of at least three connected elements were much less used [Kąciak 2011]. In this paper a method of laddering data usage as a basis of segmentation obtained by clustering is presented and two issues related to the clustering of laddering data are discussed.

2.2. Ladders' aggregation and ladders' dissimilarity measurement

Ladders collected by laddering interviews are sequences of qualitative variables and two problems need to be solved before clustering them. The first is that during the ladder interview one respondent can generate more than one *A-C-V* ladder and most of them usually do so. Information from all ladders constructed by one respondent need to be aggregated before starting the analysis. A configuration of variables that will be used as the basis of clustering must be considered and there are at least two different methods to choose from:

- *Merging*: all ladders generated by one respondent may be merged into one sequence A_1-C_1-V_1- A_2-C_2-V_2-. . . – A_n-C_n-V_n. In this case, one object is characterized by one string of variables treated as one feature of this object.
- *Averaging*: every single sequence may be treated as one of the characteristics of an object. In this case, the object is described by a set of *n A-C-V* sequences and each string of variables is treated as one feature of the object.

When the first method of aggregation is used, the distance between two objects is equal to the distance calculated between two sequences that describe these objects. When the second aggregation method is used, the distance between two objects might be calculated as an average of the distances computed between all the pairs of ladders linked with each object:

$$D(X,Y) = \frac{\sum d(x_i, y_j)}{i * j},$$

where: $D(X,Y)$ – distance between the object X and the object Y, $d(x_i, y_j)$ – distance between i^{th} ladder related to the object X and j^{th} ladder related to the object Y, i – number of ladders related to object X, j – number of ladders related to object Y.

Although the presented methods of information aggregation require operations on different steps of analysis, they are solutions to the same problem. For this reason they are presented and discussed together.

The second problem that needs to be solved before laddering data clustering is that the measurement of the dissimilarities between *A-C-V* ladders requires taking into account the dissimilarity and the order of variables that built these sequences. The distance measures often used for clustering are not appropriate for laddering data, sequence dissimilarity measures seem to provide better results [Elzinga 2003; Gbadhinio 2011]. Some measures that can be used are: the longest common prefix [Elzinga 2007], the longest common suffix [Elzinga 2007], the longest common subsequence [Elzinga 2007], the Hamming distance [Hamming 1950], the dynamic Hamming distance [Lesnard 2006], and a set of dissimilarity measures called optimal matching [Levensthein 1966].

The longest common prefix, the longest common suffix and the longest common subsequence measures are based on the number of the matching positions of two sequences: Table 1 contains the formulas[2]. The difference between them is in the method of counting the common elements of two sequences. The longest common prefix (suffix) of two sequences is k first (last) identical elements of them which occurred in the same order one after another in both the compared sequences. The longest common subsequence is k common elements occurring in the same order, but not necessarily one after another in two sequences.

[2] For more details see [Elzinga 2007].

Table 1. Sequence dissimilarity measures

Sequence dissimilarity measure	Formula
The Longest Common Prefix	$LLCP = \lvert x \rvert + \lvert y \rvert - 2LCP(x, y)$
The Longest Common Subsequence	$LLCS = \lvert x \rvert + \lvert y \rvert - 2LCS(x, y)$
Longest Common Suffix/ Reversed RLCS	$RLCS = \lvert x \rvert + \lvert y \rvert - 2RCS(x, y)$ 1

*$|x|$, $|y|$ – length of the sequences x, y (number of elements of the sequences x, y), $LCP(x, y)$ – length of the longest common prefix of sequences x and y, $LCS(x, y)$ – length of the longest common subsequence of sequences x and y, $RCS(x, y)$ – length of the longest common suffix of sequences x and y.

Source: [Elzinga 2007].

Optimal matching is a set of distances based on the minimal cost of transformation one sequence into another and the operation that can be used in this process are insertions, deletions and substitutions [Levensthein 1960]. The cost of insertions and deletions is a single value specified by analytic means (usually 1), the cost of substitution might be constant or might be equal to the transition rates between elements observed in the data. In the second case the substitution cost is equal to

$$s_{i,j} = 2 - p(i \mid j) - p(j \mid i),$$

where $p(i|j)$ ($p(j|i)$) is a probability of observing element s_i (s_j) at a position t + 1 given that element s_j (s_i) has been observed at position t [Ghabadinio et al. 2011].

Full *A-C-V* ladders are hardly ever used for segmentation – they are rarely clustered, thus the impact of methods of aggregation of information and sequence dissimilarity measures on clustering results has not been fully discussed yet in the literature. The purpose of this paper is to present and to compare the results of laddering data clustering obtained when different methods of aggregation and different sequence dissimilarity measures were used.

3. Clustering of laddering data

Clustered data originate from a research on Polish adolescents' online consumer behaviour wherein the means-end approach was used. The purpose of the study was to reach the cognitive-motivational structures that influence adolescents' online consumer behaviour and to build a segmentation model on the basis of them. 1004 people aged 18 to 20 participated in the research, 53% of them were women and 47% were men. 40 out of 228 public high schools from the Malopolska province were randomly selected and one final class was chosen in every school. All the students from the selected classes who were present at school on the research day participated in the study. An interviewer, who came into classroom during a lesson, gave out questionnaires and instructions on how to fill them in to the respondents. Data were collected from February to April 2015.

The questionnaire used in the research had two parts: first the respondents answered the question whether they buy online, then there were some questions about their e-shopping habits. The second part of the research tool was a hard ladder interview. Respondents were given four three-columns tables: there were 14 attributes of online shopping in the first column, 14 consequences of e-shopping in the second column and 14 values in the last column[3]. The respondents' task was to construct *A-C-V* ladders from the proposed elements. The respondents answered the question *Why are you buying online?* by choosing one attribute from the first column, then they finished the sentence: *It's important for me because. . .* by choosing a consequence from the second column and after that they once more finished the sentence *It's important for me because. . .* by choosing a value from the last column. Each respondent could construct a maximum of four three-element ladders.

908 respondents who had ever bought something online participated in the hard laddering interview. Each of them generated at least one three-element *A-C-V* ladder, 96% of them constructed two ladders, 88% three and 84% four ladders. The program R packages TraMiner [Gabadinho et al. 2011] and clusterSim [Walesiak, Dudek 2016] were used for data analysing.

Table 2. The Adjusted Rand Index

Compared method	Number of groups								
	2	3	4	5	6	7	8	9	10
Lcs – Om	1	1	1	1	1	1	1	1	1
Lcp – Lcs(Om)	1.000	0.216	0.257	0.247	0.292	0.285	0.231	0.223	0.247
LCS12 – OM12	0.988	0.941	0.795	0.820	0.624	0.807	0.766	0.588	0.543
LCP12 – LCS12	0.888	0.268	0.173	0.130	0.110	0.075	0.082	0.073	0.066
LCP12 – OM12	0.900	0.267	0.173	0.134	0.092	0.075	0.089	0.078	0.069
Lcp – LCP12	1.000	0.312	0.172	0.186	0.142	0.121	0.091	0.084	0.070
Lcs – LCS12	0.888	0.331	0.405	0.357	0.197	0.217	0.228	0.195	0.257
Om – OM12	0.900	0.350	0.430	0.380	0.207	0.218	0.257	0.230	0.286

Lcs – aggregation method: averaging, sequence dissimilarity measures: LLCS; Lcp – aggregation method: averaging, sequence dissimilarity measures: LLCP; Om – aggregation method: averaging; sequence dissimilarity measures: OM; LCS12 – aggregation method: merging; sequence dissimilarity measures: LLCS; LCP12 – aggregation method: merging; sequence dissimilarity measures: LLCP; OM12 – aggregation method: merging; sequence dissimilarity measures: OM.

Source: own elaboration.

[3] The attributes, consequences and values of online shopping were collected during the qualitative study that preceded the research presented in this paper.

Both (presented in the previous section) methods of information from ladders aggregation and three sequence dissimilarity measures (the longest common prefix, the longest common subsequence, the optimal matching with substitution cost equal to transition rates) were used in the analysis and they were combined, thus there were six modes of analysis. Every mode was used for data clustering from 2 to 10 groups, 54 solutions were built. Models with the same number of groups were compared with the adjusted Rand index (aRand)[4] [Hubert, Arabie 1985].

Irrespective of the aggregation method and sequence dissimilarity measure used in the analysis, the clustering effects were the most similar when data were divided into two parts and for these solutions the adjusted Rand index was between 0.888 and 1. When the groups' number was increased, the values of aRand decreased but these changes were not monotonic.

When the LLCS and the OM distances were used and the distances between ladders were averaged, the clustering results were identical no matter how many groups there were in the compared solutions (adjusted Rand index=1). When the ladders were merged into one sequence, the clustering effect obtained with the OM and the LLCS distances were varied: the solutions with two groups were nearly the same (aRand=0.988), for the rest of models the adjusted Rand index contained between 0.588 (model with 9 groups) and 0.941 (model with 3 groups).

When the solutions built with the LLCP were compared to the solutions built with the LLSC and the OM distances, the adjusted Rand index was the highest for models with two groups and its values were low for the rest of the collated solutions. When the method of information aggregation used was averaging, the aRand values were between 0.216 and 0.285 for models with three or more groups and 1 for the first solution. When the ladders were merged into one sequence and data was divided into two parts, the adjusted Rand index was 0.888 for the model built with the LLCP and the LLCS and 0.9 for the model built with the LLCP and the OM. For models with 3 groups the adjusted Rand index was 0.268 (LLCP-LLCS) and 0.267 (LLCP-OM) and for the rest of the compared solutions the values of aRand were lower than 0.2.

---

[4] Given a set of n objects $S = \{o_1, o_2, \ldots, o_n\}$ and two partitions of it: $U=\{u_1, u_2, \ldots, u_r\}$ and $V = \{v_1, v_2, \ldots, v_c\}$ define the following:

a – the number of pairs of objects in S that are in the same subset in U and in the same subset in V;
b – the number of pairs of objects in S that are in the same subset in U and in different subsets in V;
c – the number of pairs of objects in S that are in different subsets in U and in the same subset in V;
d – the number of pairs of objects in S that are in different subsets in U and in different subsets in V.

Adjusted Rand Index can be computed by [Santos, Embrechts 2009]:

$$Adjusted\ Rand\ Index = \frac{\binom{n}{2}(a+d)-[(a+b)(a+c)+(c+d)(b+d)]}{\binom{n}{2}^2-[(a+b)(a+c)+(c+d)(b+d)]}.$$

The closer the adjusted Rand index is to 1, the more similar the two compared partitions are [Hubert, Arabie 1985].

Models built with the same distance measure and different methods of information aggregation differed significantly. When LLCS was used, the values of the adjusted Rand index for models with 3 or more groups obtained with averaging and ladder merging were between 0.405 and 0.197. When the OM distance was used, the adjusted Rand index for the compared models with more than two groups was from 0.207 to 0.403 and when the LLCP measure was used, the values of aRand were between 0.07 to 0.312.

4. Concluding remarks

Laddering data are a valuable source of information about the motives and values that guided consumers' choices which could be used during market segmentation. Although clustering is one of the most popular statistical methods used for segmentation, full ladders built of at least three elements *A-C-V* are hardly ever clustered. In this paper two important aspects of cluster analysis of laddering data were presented: the first was the aggregation of information from ladders created by one respondent and the second was the measurement of ladders' dissimilarity.

The results of the analysis presented in this paper indicated that methods of information aggregation and methods of ladders' dissimilarity measurement strongly affect clustering effects. The obtained solutions were very similar only when data was divided into two parts, but it was the effect of data features rather than the properties of the used analytical methods. In all the models with two groups there was one large cluster with more than 800 observations and one small group with about 100 objects and some of them were probably outliers.

Two methods of aggregation of information from the ladders compared in this paper strongly influenced the clustering results, and before one of them is chosen the properties of clustering data should be considered. The question about the importance of differences in the number of ladders generated by one respondent must be answered. When ladders are merged, the obtained sequences differ in length and it effects the values of sequence dissimilarity measures [Elzinga 2007]. When the distances between the ladders are averaged, the influence of ladders' number on dissimilarity measures is eliminated.

Sequence dissimilarity measures affect the results of clustering. The effects of analysis obtained with the longest common subsequence distance and the optimal matching distance were more similar to each other than to the solution built with the longest common prefix distance. The variety of the clustered results obtained with different sequence dissimilarity measures was bigger with merging of ladders than with their averaging.

The theoretical background and characteristics of the collected data should be taken into account before one of the compared distance measures and aggregation methods is chosen. It seems to be appropriate to use more than one distance measure and aggregations method, compare the obtained models and choose one of them on the basis of all the information gathered during analysis. The results of the analysis presented in this paper indicate the influence of dissimilarity measures and the methods of aggregation on the clustering results, but they do not answer the question about their effectiveness, thus more research on this issue is needed.

Bibliography

Baker S., Thompson K., Engelken J., 2004, *Mapping the values driving organic food choice: Germany vs UK*, European Journal of Marketing Communication, 8(1), pp. 19-30.

Bottschen G., Thelen E., Pieters R., 1999, *Using means-end structures for benefit segmentation*, European Journal of Marketing, 33(1/2), pp. 38-58.

Domurat A., 2009, *Identyfikacja wartości osobowych w badaniach psychologicznych. Wartości jako cele działań i wyborów*, Wydawnictwo UW, Warszawa.

Elzinga C.H., 2003, *Sequence similarity: a nonaligning technique*, Sociological Methods and Research, 32(1), pp. 3-29.

Elzinga C.H., 2007, *Sequence analysis: Metric representations of categorical time series*, Technical Report, Department of Social Science Research Methods. Vrije Universiteit, Amsterdam.

Fotopoulos C., Krystallis A., Ness M., 2003, *Wine produced by organic grapes in greece: using means-end chains analysis to reveal organic buyers purchasing motives in comparison to the non-buyers*, Food Quality and Preferences, 14, pp. 549-566.

Gabadinho A., Ritschard G., Müller N.S., Studer M., 2011, *Analyzing and visualizing state sequences in R with TraMineR*, Journal of Statistical Software, 40(4), pp. 1-37.

Gabadinho A., Ritschard G., Studer M., Muller N.,2011*, Mining sequence data in R with the TraMineR package: A user's guide*, http://mephisto.unige.ch/pub/TraMineR/doc/1.2/TraMineR-1.2-Users--Guide.pdf (23.05.2015).

Gruber T., Szmigin I., Reppel A., Voss R., 2008, *Designing and conducting online interviews to investigate interesting consumer phenomena*, Qualitative Marketing Research: An International Journal, 1193, pp. 256-274.

Grunert K., 1995, *Food quality: a means-end perspective*, Food Quality and Preferences, 6, pp. 171-176.

Grunert K., Grunert S.C., 1995, *Measuring subjective meaning structures by the laddering method: theoretical considerations and methodological problems*, International Journal of Research in Marketing, 12(3), pp. 209-225.

Gutman J., 1982, *A means-end chain model based on consumer categorization process*, Journal of Marketing, 46(2), pp. 60-72.

Gutman J., 1984, *Analyzing Consumer Orientation towards Beverages through Means-End Chain Analysis*, Psychology and Marketing, 1(3/4), pp. 23-43.

Hamming R.W., 1950, *Error-detecting and error-correcting codes*, Bell System Technical Journal, 26, pp. 147-160.

Hubert L., Arabie P., 1985, *Comparing partitions*, Journal of Classification, 2, pp. 193-218.

Kąciak E., 2011, *Teoria środków – celów w segmentacji rynku. Studium metodologiczno-empiryczne*, Oficyna a Wolters Kluwer business, Warszawa.

Kotler P., 2005, *Marketing,* wydanie 11, Rebis, Poznań.

Lesnard L., 2006, *Optimal Matching and Social Sciences*, Série des Documents de Travail du CREST, Institut National de la Statistique et des Etudes Economiques, Paris.

Levenshtein V.I., 1966, *Binary codes capable of correcting deletions, insertions and reversals*, Soviet Physics Doklady, 10(8), pp. 707-710.

Mentzer J., Rutner S., Matsuno K., 1997, *Application of the means-end value hierarchy model of understanding logistics services quality*, International Journal of Psychical Distribution and Logistics Management, 27(9/10), p. 230-243.

Olson J., Reynolds T., 2001, *The Means-End Approach to Understanding Consumer Decision Making*, [in:] Reynolds T., Olson J. (eds.), *Understanding Consumers Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, Lawrence Earlbaum Associates, Mahwah, NJ, pp. 3-20.

Reppel A., Szmigin I., Gruber T., 2006, *The iPod phenomenon: identifying a market leaders secrets through qualitative marketing research*, Journal of Product and Brand Management, 15(4), pp. 239-249.

Reynolds T., Gutman J., 2001, *Laddering Theory, Method, Analysis and Interpretation*, [in:] Reynolds T., Olson J. (eds.), *Understanding Consumers Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, Lawrence Earlbaum Associates, Mahwah, NJ, pp. 25-62.

Reynolds T., Rohon J., 2001, *Consumer Segmentation Based on Cognitive Orientations: the Chem--Lawn Case*, [in:] Reynolds T., Olson J. (eds.), *Understanding Consumers Decision Making: The Means-End Approach to Marketing and Advertising Strategy*, Lawrence Earlbaum Associates, Mahwah, NJ, pp. 283-298.

Santos J.M., Embrechts M., 2009, *On the Use of the Adjusted rand Index as a Metric for Evaluating Supervised Classification*, [in:] *Artificial Neutral Networks – ICANN 2009*, C. Alippi, M. Poly carpou, Ch. Panayiotou, G. Ellinas (eds.), Springer, pp. 175-184.

Walesiak M., 2000, *Segmentacja rynku. Kryteria i metody*, [in:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, A. Zaliś (ed.), Wydawnictwo AE, Kraków, pp. 191-201.

Walesiak M., Dudek, A., 2015, *Searching for Optimal Clustering Procedure for a Data Set*, https://cran.r-project.org/web/packages/clusterSim/clusterSim.pdf. (20.01.2015).

Wedel M., Kakamura W., 1998, *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishing, Norwell.