Standard of Living in Poland at Regional Level - Classification with Kohonen Self-Organizing Maps

Marta Kuc[1], Aneta Sobiechowska-Ziegert[2]

Abstract

The standard of living is spatially diversified and its analyzes enable shaping regional policy. Therefore, it is crucial to assess the standard of living and to classify regions due to their standard of living, based on a wide set of determinants. The most common research methods are those based on composite indicators, however, they are not ideal. Among the current critiques moved to the use of composite indicators is the normative nature of the weights, the drawbacks of the linear method of aggregation and the fact that composite indicators flatten alternatives and differences among analyzed objects. That is why more objective alternative solutions are nowadays gaining popularity.

The aim of this article is the regional classification of Polish NUTS-4 regions due to their standard of living using SOM - self-organizing maps - the model of a neural network mapping multidimensional space into a two-dimensional map of neurons. In this study, the set of selected 45 variables describing the standard of living in Polish NUTS-4 regions in 2016 was used.

Keywords: *standard of living, neural networks, Kohonen self-organizing maps, regional analysis*

JEL Classification: C45, I31, R12

DOI: 10.14659/SEMF.2018.01.23

1 Introduction

For many years the standard of living has been the subject of interest not only to scientists but as well to politicians, journalists, and public opinion. As numerous studies show the standard of living is spatially diversified, not only at international level (Wawrzyniak, 2016) but as well at intra-country level (Kuc, 2017). Therefore monitoring the spatial diversity of the standard of living is important from the point of view of sustainable and coherent development, and may be used in shaping the regional policy at both country and the European Union level (Chrzanowska et at., 2017).

The main goal of this research is the implementation of Kohonen self-organizing maps to the regional classification of Polish NUTS-4 regions due to their standard of living in 2016. The analysis covers 314 counties (urban counties were not taken into consideration).

[1]Corresponding author: Gdansk University of Technology, Faculty of Management and Economics, Department of Economic Science, Traugutta 79, 80-233 Gdansk, marta.kuc@zie.pg.gda.pl.

[2]Gdansk University of Technology, Faculty of Management and Economics, Department of Economic Science, Traugutta 79, 80-233 Gdansk, as@zie.pg.gda.pl.

2 Standard of living measurement

The methodology of research on the standard of living has evolved from GDP, through measures 'adjusting' GDP (i.e. MEW, ISEW or Green GDP), measures 'replacing' GDP (i.e. HDI, HPI, The where to be born index or Gallup Index), measures 'supplementing' GDP (i.e. Decoupling Indicators) up to soft modelling, fuzzy analysis and artificial neural networks. Nowadays the most popular way of analyzing standard of living is the one in which author begins by constructing a conceptual scheme of some sort describing his understanding on the standard of living, including its constituents and determinants. There is no one, widely accepted, set of standard of living determinants. Eurofound is currently working on the preparation of a set of indicators that can be used in standard of living and social cohesion analysis, both at international and regional level. Based on the literature review one can prepare the following Mandala of the standard of living (Fig. 1). Where the standard of living is the core of Mandala and three circles represent resources shaping standard of living.

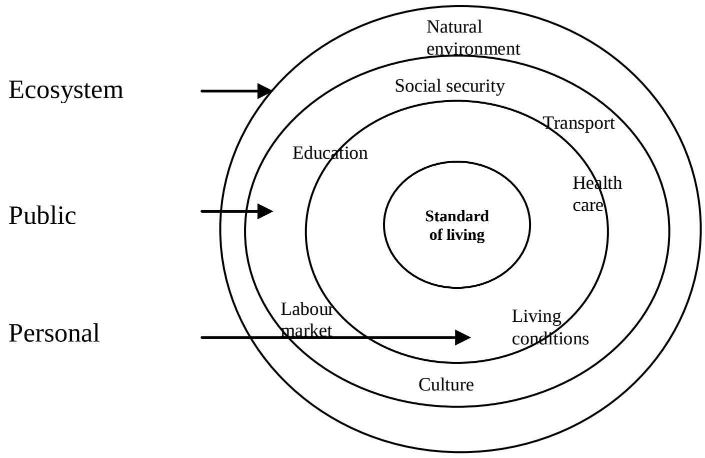

Fig. 1. The Mandala of Standard of Living.

Source: Authors' own study based on Goossens ed. (2007) and Michalos et al. (2011).

3 Pros and cons of composite indicators

Nowadays composite indicators are not only used in the standard of living and well-being analysis (Becker et al., 2017). They are also used in analysis of countries' competitiveness (Kruk and Waśniewska, 2017), socioeconomic development (Sobiechowska-Ziegert and Mikulska, 2013), quality of institutions (Balcerzak and Pietrzak, 2017), sustainable development (Pietrzak et al., 2017) and many others.

In the literature (Saltelli, 2007; Michalos et al., 2011) one can find following advantages of composite indicators: ease of interpretation; possibility to summarise multi-dimensional phenomenon; they are excellent communication tool to be used by media; make it easier to evaluate and present an overall trends of phenomena over time and across geographic regions; help to reduce the visible size of a set of indicators without dropping the underlying information base; are flexible by including desired or excluding undesired variables.

However, composite indicators are not ideal. Among the current critiques of using composite indicators the following disadvantages can be listed (Saltelli, 2007; Michalos et al., 2011; Becker et al., 2017): lack of transparency in construction; subjective selection of indicators; ad hoc selection of weights and aggregation methods; risk of misinterpretation; oversimplification of complex phenomenon; risk of misusing by supporting desire policy; risk of giving misleading policy directions; occurrence of redundant variables and double-counting; normative nature of the weights; drawbacks of the linear method of aggregation; flattening alternatives and differences among analyzed objects; lack of clear meaning; variability of domains and indicators over time; sensitiveness to the choice of normalization and aggregation methods.

4 Kohonen self-organizing maps for the classification problem

Bearing in mind the above-mentioned set of disadvantages, in this research we want to use Kohonen self-organizing maps (SOM) to classify Polish NUTS-4 regions due to their standard of living. SOMs are exploratory data analysis technique projecting multi-dimensional data onto a two-dimensional space. That procedure allows clear visualization of the data and easy identification of groups with similar characteristics. In this sense, Kohonen maps can be thought of as a factor analysis combined with a cluster analysis (Deichmann et al., 2013). The advantage of self-organizing maps over composite indicators in the classification procedure is revealed by excluding such problems as normative nature of weight, sensitivity to used normalization and aggregation method, and also a subjective selection of indicators. Another advantage of Kohonen maps is the self-organizing property of the map which makes estimated components varies in a monotonic way across the map (Deichmann et al., 2007). The utility of SOMs in cluster analysis of multidimensional phenomenon has been presented, among others in Pisati et al. (2010).

The basic idea for constructing Kohonen maps is to reduce the complexity of matrix X consisting of input vectors x_i representing the coordinates of observation i in the d-dimensional input space, by projecting it onto a lower dimensional output space. This space usually takes the form of a two-dimensional grid arranged in a square or hexagonal lattice. Each grid cell called ‘node’ or ‘neuron’ is specialized in attracting observations that possess certain combinations of attributes. It means that each SOM node is characterized by a unique

1xd weight vector that belongs to the same coordinate space as the input vectors x_i. Input vectors are compared in the learning process of the SOM with the weight vectors and each observation *i* can be properly allocated to the best matching neuron. The detailed description of SOM algorithm is widely presented and can be found in Kohonen (2013), Pisati et. al (2010) and others.

5 Empirical analysis

To classify Polish NUTS-4 regions[3] due to their standard of living in 2016 Kohonen self-organizing maps were applied. In this research we used 45 variables describing the following standard of living dimension[4]:

- education: percentage of kids aged 3-5 participating in pre-school education (17), children aged 3-5 years per one place in a pre-school education center (26), students per one branch in primary schools (21), students per one branch in secondary schools (19), students per one branch in upper secondary schools (36), net (30) and gross (32) primary education ratio, net (31) and gross (33) junior high school education ratio;
- social security: percentage of population receiving social benefits (23), crime detection rate (27), number of crimes per 1000 inhabitants (28), divorces per 1000 inhabitants (35), poverty rate (38), percentage of children up to 17year old, for which parents received child benefit (39), old age dependency ratio (15);
- culture and recreation: number of people per 1 library (3), number of books borrowed by one reader (4), number of people per 1 place in cinema (5), number of accommodation places per 100000 inhabitants (12), number of hotels per 100000 inhabitants (11);
- healthcare: number of people per 1 pharmacy (1), number of doctors per 1000 inhabitants (10), number of nurses per 1000 inhabitants (20), deaths on cancer as a % of total deaths (43), deaths on cardiovascular diseases as a % of total deaths (44), infant mortality rate (45);
- labour market: unemployment rate (34), long-term unemployment rate (2), youth unemployment rate (16), number of job offers per 1 unemployed person (18), employment rate (24), average monthly wage (25), number of accidents at work per 1000 workers (42);

[3] All urban counties were excluded from the analysis due to avoid classification problems of other counties due to large differences in the level of neuron activation.
[4] Number in brackets is the same as input ID on Fig. 4.

- natural environment: percentage of protected area (22), industrial water sewage rate (37);
- transport and communication: number of cars per 1000 inhabitants (29), length of paved roads per 100km^2 (40), car accidents per 100000 inhabitants (42);
- living conditions: percentage of households connected to water supply (6), percentage of households connected to sewage system (7), percentage of population using gas installation (8), percentage of population connected to sewage treatment plants (9), average flat area per 1 person (13), number of flats per 1000 inhabitants (14).

The main variable selection criterion was data availability at NUTS-4 level. The sample size was 314 counties. To prepare U-matrix, batch-learning version of the SOM was used. After several trials, the most satisfying results[5] gave the SOM consisting of 49 nodes arranged in a 7x7 hexagonal lattice. From the U-matrix and the previous taxonomic analysis[6], it was possible to point six main groups of regions which show similarities in the scope of selected variables (see Fig. 2).

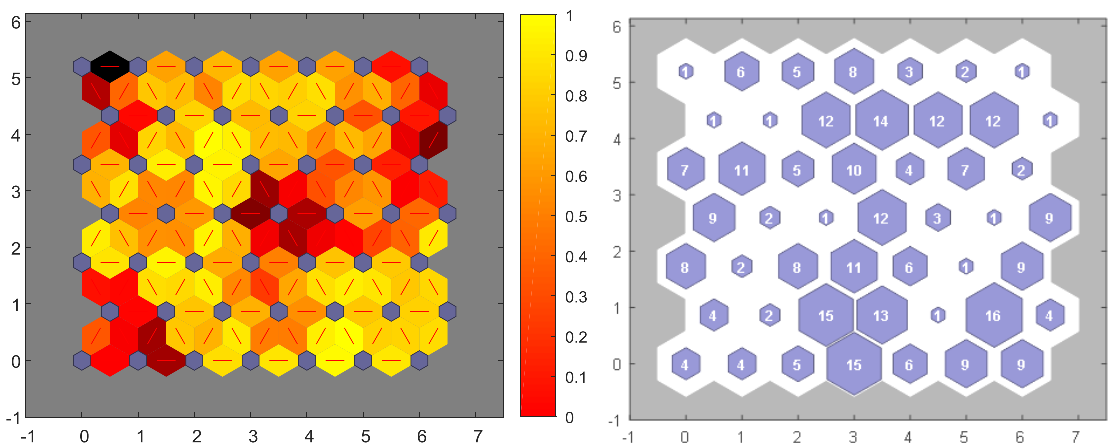

Fig. 2. The U-matrix of 7x7 SOM (left picture). Light shades indicate large similarity whereas dark shades show dissimilarities and cluster borders. The neuron locations in the SOM topology and the number of hits are shown in the right picture.

The light colour on the map represents smaller distances between neurons and hence the high intra-cluster homogeneity. The darker colour on the map, the larger distances between neurons and hence the smaller inter-cluster homogeneity. Based on the cells colours and

[5] There were also constructed other SOMs arranged in a rectangular grid as well as hexagonal lattice from 3x10 up to 10x10. When choosing the model, the aim was to ensure that the network spreads as well as possible on the examined objects and does not contain empty neurons.
[6] Tree Diagram for Ward's method and k-means clustering results indicated the existence of 6 clusters.

number of regions associated with each of the neurons, NUTS-4 regions were grouped into clusters which spatial arrangement is shown in Fig. 3.

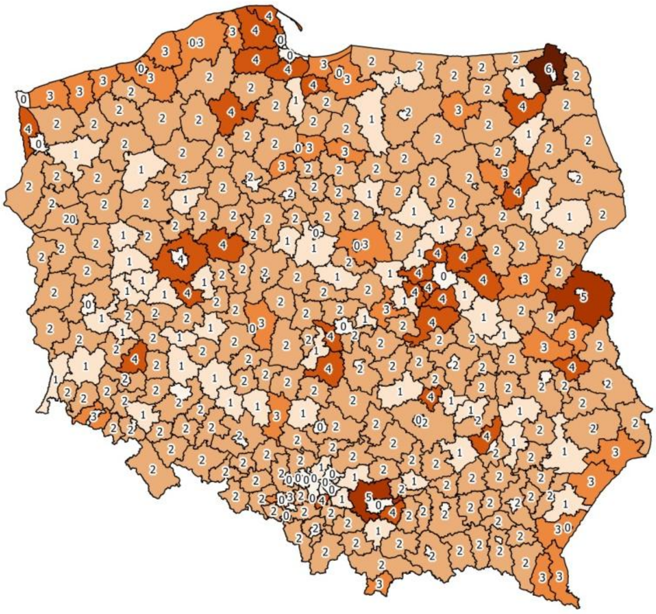

Fig. 3. The spatial arrangement of clusters according to NUTS-4 classification. 0 means urban counties excluded from the research.

According to this arrangement, biggest, suburban counties are mostly described with cluster 4, these are the regions with the highest average values for health care and almost the highest values for education, culture, labour market, transport and living conditions (Table 1). Which allows to define them as counties with the highest standard of living. Counties qualified as cluster 1 are regions further away from the cities, with no clearly developed service and production facilities (relatively low average values for the labour market and living conditions - Table 1). Majority of counties classified in cluster 2 are similar regardless of geographic location. Cluster 3 stretches along the coastline, in the south-eastern border areas and in some other counties which can be regions perceived as, among the others, developing tourism. In cluster 5 there were two counties classified - bialski and krakowski, which according to Table 1 can be perceived as the one with the lowest average values in almost all standard of living dimensions. In contrast to county suwalski which creates separate cluster no. 6 - with the highest values of variables presented in Table 1 (the exception is health care - in this area suwalski turned out to have the lowest value).

Table 1 presents the average values of variables with the strongest influence on neurons (Fig. 4) from each standard of living dimension (Fig. 1).

Table 1. Cluster statistics - group mean values for chosen variables.

Cluster no.	Input 17	Input 15	Input 4	Input 44	Input 2	Input 37	Input 29	Input 6
	S	**D**	**S**	**D**	**D**	**S**	**S**	**S**
1	75.77	31.78	18.80	47.12	39.26	97.56	586.18	89.16
2	72.67	30.57	19.20	47.04	39.36	96.73	574.20	84.16
3	67.45	29.42	18.18	45.78	38.43	97.75	570.99	84.94
4	83.85	28.89	18.65	43.34	36.86	99.83	564.01	87.80
5	74.33	29.85	18.09	51.50	41.65	96.22	608.83	84.00
6	47.13	28.70	21.18	41.53	30.70	100.00	604.11	87.50

The shades of colours and corresponding numbers of clusters on the map should not be associated directly with the standard of living. What they mean is the similarities and differences in the counties due to the 45 variables used for describing the standard of living. According to the results, one can notice that despite visible differentiation, counties do not form clusters solely on the basis of the geographical neighbourhood. The former division into Poland A and B related to the socio-economic development is also not observable here. Instead, the similarities can be sought rather in the proximity to large agglomerations and tourist attractiveness of the regions. SOM weight planes presented in Fig. 4 demonstrate the influence of each variable on a different cluster node. The darker the colour, the smaller the impact of a given variable. It can be noticed for instance that variables describing culture and recreation group, had influence only on the cluster in the left lower corner of the map which represents cluster number 3.

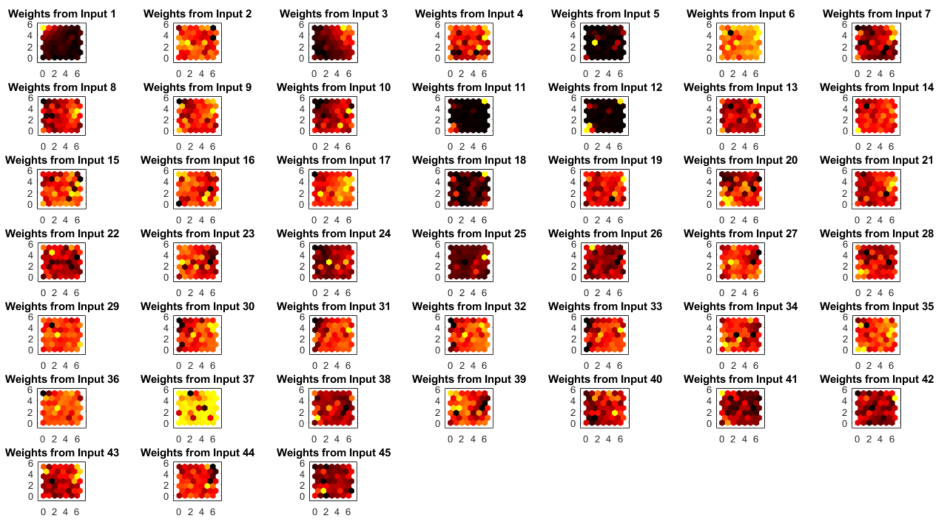

Fig. 4. Component planes for all variables.

To check the comparability of clusters obtained using SOM we also grouped analyzed counties based on composite indicator (CI) value. Three different methods were used to calculate CI (quotient transformation by max was used to normalize all variables):

- without a pattern: $CI_i = \frac{\sum_{i=1}^{m} z_{ij}}{m}$,
- with a pattern: $CI_i = 1 - \frac{d_i^+}{d_-}$,
- TOPSIS: $CI_i = \frac{d_i^-}{d_i^+ + d_i^-}$.

Where: z_{ij} - normalized values of j^th variable in i^th region ($i = 1, .., n; j = 1, ..., m$); d_i^+- Euclidean distance between each region and a pattern (a pattern is an object with maximum values for stimulants and minimum for destimulants); $d_- = \bar{d} - 2s_d$; d_i^-- Euclidean distance between each region and an anti-pattern (an anti-pattern is an object with minimum values for stimulants and maximum for destimulants).

CI values were a basis to group counties into 4 groups (I - $CI_i \geq \overline{CI} + s_{CI}$; II - $\overline{CI} \leq CI_i < \overline{CI} + s_{CI}$; III - $\overline{CI} - s_{CI} \leq CI_i < \overline{CI}$ and IV - $CI_i < \overline{CI} - s_{CI}$. Also k-means algorithm was used as grouping tool[7]. The results of V-Cramer statistics are presented in Table 2.

Table 2. V-Cramer statistics.

Clustering method	**CI without a pattern**	**CI with a pattern**	**CI TOPSIS**	**k-means 3 clusters**	**k-means 4 clusters**	**k-means 6 clusters**
SOM	0.569	0.628	0.603	0.816	0.744	0.763

As can be seen in Table 2 SOM clustering results are much closer to those obtained using k-means algorithm than to those based on composite indicator value.

Conclusions

In this paper, the standard of living in Polish counties was considered based on 45 determinants. The first part of the paper discussed briefly standard of living measurement and pros and cons of composite indicators. At the empirical part, regions were clustered using self-organizing maps approach, which is a more objective clustering method. A number of clusters, however, has been obtained using U-matrix and the taxonomic analysis. The analysis showed that there is visible differentiation among the counties, however, they form clusters neither on the basis of geographical neighbourhood nor historical conditions. Two interesting

[7] $\overline{CI}$ - an average value of CI value, s_{CI} - standard deviation of CI value.

clusters of similar objects can be distinguished: the first one consist of counties adjacent to large agglomerations (cluster no. 4 - which contains counties with the highest standard of living); and cluster no. 3 that consists of counties attractive for tourism. Results presented in Table 1 shows that SOM has good discriminatory abilities not only in terms of structures similarity but as well values similarity, as they maximise between-group dissimilarities and minimalize within-group dissimilarity. They are also a useful tool to visualize high-dimensional data.

Further analysis should include dynamics of the described phenomenon as well as its determinants. Well-constructed SOM model will allow to predict and assess whether changes taking place in particular counties will improve or worsen their standard of living.

References

Balcerzak, A. P. & Pietrzak, M.B., (2017). Human Development and Quality of Institutions in Highly Developed Countries. In: Bilgin, M. H., Danis, H., Demir, E. & Can U. (Eds.). *Financial Environment and Business Development. Proceedings of the 16th Eurasia Business and Economics Society.* Springer International Publishing, 231-241.

Becker W., Saisana, M., Paruolo, P. & Vandecasteele, I. (2017), Weights and Importance in Composite Indicators: Closing the Gap. *Ecological Indicators*, *80*, 12-22.

Chrzanowska, M., Dudek, H. & Landmesser J. (2017), Standard of Living in Selected Countries in Central Europe in the Period 2000-2014. In: Jedlicka, P., Maresova, P. & Soukal, I. (eds.) *Proceedings of the International Scientific Conference Hradec Economic Days 2017*, *7*(1), 325-331.

Deichman, J., Eshghi, A., Haughton, S. & Woolford, S. (2007). Measuring the international digital divide: An application of Kohonen self-organizing maps. *International Journal of Knowledge and Learning*. *3*(6), 552-575.

Deichman, J., Haughton, D., Malgwi, C. & Soremekun, O. (2013). Kohonen Self-Organizing Maps as a Tool for Assessing Progress toward the UN Millennium Development Goals. *Journal of Human Development and Capabilities*, *14* (3), 393-419.

Goossens, Y. (ed.), (2007). *Alternative progress indicators to Gross Domestic Product (GDP) as a means towards sustainable development*. European Parliament Study.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65.

Kruk, H. & Waśniewska, A. (2017). Application of the Perkal method for assessing competitiveness of the countries of Central and Eastern Europe. *Oeconomia Copernicana*, *8*(3), 337–352.

Kuc, M. (2017), Social convergence in Nordic countries at regional level. *Equilibrium Quarterly Journal of Economics and Economic Policy, 12* (1), 25-41.

Michalos, A. C., Smale, B., Labonté, R., Muharjarine, N., Scott, K., Moore, K., Swystun, L., Holden, B., Bernardin, H., Dunning, B., Graham, P., Guhn, M., Gadermann, A.M., Zumbo, B.D., Morgan, A., Brooker, A.-S. & Hyman, I. (2011). *The Canadian Index of Wellbeing. Technical Report 1.0. Waterloo.* On: Canadian Index of Wellbeing and University of Waterloo.

Pietrzak, M. B., Balcerzak, A. P., Gajdos, A. & Arendt, Ł. (2017), Entrepreneurial environment at regional level: the case of Polish path towards sustainable socio-economic development. *Entrepreneurship and Sustainability Issues*, *5*(2). 190-203.

Pisati, M., Whelan, Ch., Lucchini, M. & Maitre, B. (2010). Mapping patterns of multiple deprivation using self-organising maps: An application to EU-SILC data for Ireland. *Social Science Research, 39* (3), 405-418.

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research, 81*(1), 65-77.

Sobiechowska-Ziegert, A. & Mikulska, A. (2013). Measure of the level of socio-economic development in provinces. *Quantitative methods in economics*, *XIV*(2), 200–209.

Wawrzyniak, D. (2016), Standard of Living in The European Union. *Comparative Economic Research - Central and Eastern Europe*, *19* (1).