GRADE ANALYSIS TO STUDY ICT USAGE IN POLISH ENTERPRISES

DOMINIK SIERADZKI, DOMINIKA URBAŃCZYK

Department of Econometrics and Statistics, Warsaw University of Life Sciences (SGGW)

The aim of this work is to study the structure of information and telecommunications technologies (ICT) usage in Polish enterprises in 2015. In this paper the overrepresentation maps were considered as a tool of data mining. These maps are crated as the results of GCA algorithm. The analysis of data about ICT in Polish enterprises were shown as an example of GCA application. The advantage of the method is the visualization possibility and legibility of the obtained maps. In this article we studied the level of usage of ICT in individual voivodeships, the expenditures for various sectors of ICT for small, medium-size and big enterprises.

Keywords: grade correspondence analysis, data mining, GCA, ICT, enterprises

1. Introduction

Nowadays the worldwide rapid development of information and communications technologies (ICT) and its significant impact on economics, business, industry, science and many sectors of everyday life are observed. This transforms the approach to thinking about innovation and becomes an important subject of research in recent years.

In the subject literature we can find a huge spectrum of definitions and information about ICT. The term "information and communication technologies" has been used in scientific research since the end of the 90s of twentieth century. The concept of ICT includes any products, tools and solutions of information technology (like computers, robots and other digital devices), communication (e.g. tele-

communication lines, wireless signals, audio-visual and telephone networks, fast and efficient internet connections) and information manipulation (software, specialist systems, algorithms and models) with especially emphasis on applications of these technologies for an easier and more efficient process of information transformation and transmission [9].

It is not surprising that the development and popularity of ICT have also been reflected in the approach to business management, especially also in strategies for enterprises. The necessity of introducing modern solutions is additionally supported by the processes of globalization and integration occurring in the worldwide economy. The ICT solutions actually help to improve effectiveness and they are the support for overcoming various barriers in business. Today it is necessary to accelerate the implementation of ICT solutions to maintain business performance and competitiveness [2, 5, 6, 13].

Information and communication technologies also may open new opportunities for polish enterprises [1, 12]. The results of scientific research indicate that in Poland the degree of business adaptation of modern ICT solutions is lower than in the world [10]. The situation is particularly unfavourable for small and medium-sized enterprises (SMEs). The undoubted significance of ICT for dynamic growth is evidenced by programs realized by Polish Agency for Enterprise Development including advanced support services, education and other activities stimulating innovation. A key issue during process of such programs planning is to reach precisely the sectors and regions that are the most in need of support. This makes the widely understood researches of this subject required.

The aim of this work is empirical study of the structure of information and telecommunications technologies adoption in Polish enterprises in 2015. In order to do this, the Internet connection speed and the level of expenses for various types of ICT solutions were chosen to analysis. The differentiation of these factors was studied versus the size of enterprises as well as their geographical localization (considering individual voivodeships). The implementation of this type of research allowed to know the degree of advancement in application of the information and telecommunications technologies in individual regions of the country. In addition, the sectors of ICT, for which the expenses are the especially high or low, are indicated, in case of small, medium-size and big enterprises.

The research method applied in this work was the grade correspondence analysis (GCA) and overrepresentation maps. This approach of data exploration is appropriate to visualize the data and to assess similarities and differences of the structure of implementation of ICT solutions in Polish enterprises. The main advantage of the method is the transparency and legibility of the obtained results. This allows to discover new, unknown, significant and understandable regularities in large data volumes. The goal of this work is not so much to know precisely the individual situation in each of the voivodeships, but to indicate the diversity in ICT adaptation in enterprises across the whole country. The applied GCA method allows to carry out this type of research.

2. Data

The data employed in this article refer to ICT usage in Polish enterprises and relate to the period of 2015 year. The database comes from the Central Statistical Office of Poland. The data of this kind are the source of a great deal of information and they deal with a lot of aspects of ICT application. This article focuses on two of them: speed of internet connection and expenditures for various sector of ICT in enterprises (like IT equipment, telecommunications equipment, financial leasing of ICT devices, own development or modification of the purchased software). In addition, the information refer to individual voivodeships and size of the enterprises.

The chosen factors allow to analyze important aspects of ICT application in Polish enterprises. The data about speed of internet connection provide knowledge about advancement in information and telecommunications technologies used in enterprises. On the other hand, the study of the structure of expenditures for various ICT solutions shows which of them are the most popular and how much money the enterprises are able to spend on various kinds of modern technologies. The applied data are presented in the tables 1 - 4.

Table 1. The number of enterprises using a certain speed of internet connection in individual voivodeships

Voivodeship	<2 Mbit/s	[2;10) Mbit/s	[10;30) Mbit/s	[30;100) Mbit/s	>=100 Mbit/s
Mazowieckie	414	3491	4628	3025	2337
Podlaskie	46	561	617	433	296
Dolnoslaskie	163	1729	2213	1095	780
Zachodniopomorskie	130	909	1316	596	396
Pomorskie	309	1589	1642	1151	641
Lubelskie	137	838	1242	477	390
Slaskie	329	3577	3532	2227	1342
Wielkopolskie	389	3101	2695	1657	1068
Kujawsko-pomorskie	117	1447	1591	720	445
Swietokrzyskie	132	645	653	303	236
Malopolskie	169	2722	2412	1373	818
Lubuskie	51	740	826	320	235
Lodzkie	237	2088	1428	993	588
Warminsko-mazurskie	82	972	786	499	195
Opolskie	74	656	708	339	107
Podkarpackie	213	1545	1555	574	326

Source: own preparation on the basis of the Central Statistical Office of Poland materials

Table 2. The average expenditures (in Polish Zlotys) for various kind of ICT sectors in individual voivodeships

Voivodeship	Financial leasing of ICT devices	IT equipment	Telecommunications equipment	Own development or modification of the purchased software
Slaskie	650 141,89	72 364,53	73 042,84	102 600,80
Dolnoslaskie	475 767,86	125 291,98	54 424,83	109 266,06
Malopolskie	254 340,43	83 354,72	64 613,09	63 629,39
Lodzkie	239 464,29	61 467,47	47 851,85	74 282,69
Zachodniopomorskie	133 806,45	42 589,47	32 267,41	43 271,43
Podkarpackie	300 827,59	63 819,90	20 799,19	186 571,43
Wielkopolskie	730 990,20	156 615,51	77 417,64	507 002,39
Lubuskie	243 620,69	34 724,14	14 349,69	188 093,33
Opolskie	66 666,67	43 232,46	15 232,76	58 037,50
Kujawskie	207 935,48	96 460,08	25 261,64	193 042,55
Swietokrzyskie	69 666,67	41 572,52	10 257,53	113 049,38
Mazowieckie	341 402,83	195 700,95	221 645,37	555 200,44
Pomorskie	115 462,37	147 779,02	157 216,97	204 589,04
Lubelskie	17 038,46	59 214,59	20 886,40	58 071,43
Podlaskie	38 444,44	86 682,01	30 587,91	116 480,77
Warminsko-mazurskie	8 611,11	32 584,80	25 901,59	48 760,00

Source: own preparation on the basis of the Central Statistical Office of Poland materials

Table 3. The number of enterprises using a certain speed of internet connection versus the size of the enterprises

Size	<2 Mbit/s	[2;10) Mbit/s	[10;30) Mbit/s	[30;100) Mbit/s	>=100 Mbit/s
Small	2693	23002	22582	11414	7118
Medium	305	3495	4675	3598	2357
Big	13	385	891	1043	915

Source: own preparation on the basis of the Central Statistical Office of Poland materials

Table 4. The average expenditures (in Polish Zlotys) for various kind of ICT sectors versus the size of the enterprises

Size	Financial leasing of ICT devices	IT equipment	Telecommunications equipment	Own development or modification of the purchased software
Small	33 267,71	15 773,01	98 959,75	63 530,22
Medium	84 836,81	64 507,31	171 407,02	93 604,97
Big	1 063 879,05	742 832,28	1 149 200,00	1 984 663,43

Source: own preparation on the basis of the Central Statistical Office of Poland materials

3. Research methods

In order to analyze ICT usage in Polish enterprises, the grade correspondence analysis (GCA) was applied. It is a data mining tool, which allows to analyze dissimilarity structures. This method was used to analyze for instance expenditure structure in Poland [7, 8], attractiveness of internet packages [14], export structure in European Union [3]. The computations were made using GradeStat. This software was developed at the Institute of Computer Science Polish Academy of Sciences (see http://gradestat.ipipan.waw.pl).

A structure of given objects could be considered as a vector of non-negative values, which sum equals unity. The number of such vector depends on a number of groups, in case of this study concerning speed of internet connection or group of ICT expenditure. To compare two structures (two distributions), a concentration curve and an associated index are used [11].

Let $\mathbf{x} = (x_1,...,x_k)$ and $\mathbf{y} = (y_1,...,y_k)$ denote some two distributions. In the unit square joining points $(0,0)$, (x_1,y_1), (x_1+x_2, y_1+y_2), …, $(1,1)$ a polygonal chain is obtained. This polygonal chain is the concentration curve of $\mathbf{x}$ distribution in relation to $\mathbf{y}$ distribution. The quotient $\dfrac{x_i}{y_i}$ denotes a slope of further sections of the polygonal chain to $\mathbf{x}$-axis. After ordering coordinates of vectors $\mathbf{x}$ and $\mathbf{y}$ in such a way, that $\dfrac{x_i}{y_i}$ would be ascending, the concentration curve is called the maximum concentration curve. The maximum concentration index corresponding to the maximum concentration curve is called index ar_{max}. To be more formal, let concentration index be given as follows:

$$ar = 1 - 2 \cdot \int_0^1 C(t)dt \,, \tag{1}$$

194

where C denotes conentration curve of the distribution **x** in relation to **y**. These two tools – the maximum concetration and the concentration index – are used in GCA algorithm.

Next, consider a matrix P which contains values of p_{ij}, where

$$\sum_{i=1}^{m}\sum_{j=1}^{k} p_{ij} = 1, p_{ij} \geq 0.$$

The main purpose of the GCA algorithm is to set rows and columns in such a way, that all ar indices equal ar_{max}. GCA changes arrangement of rows and columns at every step. The detailed description of criterion of rows and columns relocation could be found in the literature [4]. The important advantage of GCA is setting similar columns and rows next to each other. Another result of this method is the possiblity to create some kind of plots, called overrepresentation maps. For this purpose it is needed to calculate overrepresentation indices as follows:

$$h_{ij} = \frac{p_{ij}}{p_{i+}p_{+j}}, \qquad (2)$$

where p_{i+} and p_{+j} denote sum of columns and sum of rows, respectively. In the case when the overrepresentation index equals 1, it means that distribution is of ideal proportionality. Otherwise i.e. when h_{ij} is greater or smaller than 1, it means that overrepresentation or underrepresentation, respectively, is observed. In practice it means there exists a relationship between rows and columns.

4. Results

The ICT expenditures of all enterprises in Poland in 2015 reached the amount of 6 859 971 Polish Zlotys. During this period there were 95 969 enterprises and 37 498 of them spent part of their funds on ICT. The structure of these expenditures depends on certain factors. In Figure 1 the first map was presented. It shows structure of average ICT expenditures in each of Polish voivodeships. On the basis of this overrepresentation map it can be stated that the biggest share of the ICT expenditures takes financial leasing of ICT devices (the widest column denotes the biggest share of expenditures). Next columns respectively refer to own development or modification of the purchased software, IT equipment and telecommunications equipment. The purple color on the all presented maps represents overrepresentation, hence it could be noticed that for example in Slaskie voivodeship the expenditures for financial leasing of ICT devices set over the average level of ICT expenditures. On the diagonal of this overrepresentation map very similar hue is observed, hence there is a dependence between ICT expenditures and a region where the economic activity is carried. In can be concluded that some voivodships have very similar structure of ICT expenditure. Based on the analysis of the

overrepresentation map regarding the level of ICT expenditures, the following three groups of voivodeships could be distinguished: the first group, which consists of Podlaskie, Pomorskie, Mazowieckie, Lubelskie, Opolskie, Kujawsko-Pomorskie and Warminsko-Mazurskie voivodeship (where the smallest role is played by leasing of ICT devices), the second group – Lubuskie, Wielkopolskie, Podkarpackie (where you can observe lower than average expenditure on the purchase of telecommunications and IT equipment) and the third group – Zachodniopomorskie, Lodzkie, Malopolskie, Dolnoslaskie and Slaskie (where the leasing of equipment is popular, but the expenditures for modifying the purchased software and own development are significantly below average).
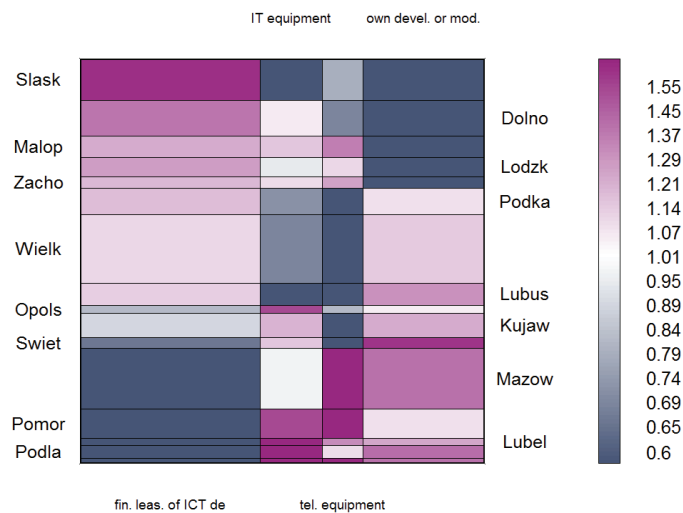

Figure 1. The overrepresentation map illustrating the expenditures structure in each of Polish voivodeships (purple represents overrepresentation, gray underrepresentation and white areas on the map refer expenses on average level)

When analyzing the next overrepresentation map (Figure 2), it can be noticed, that the structure of average ICT expenditures in Polish enterprises varies depending on a size of an enterprise. For big enterprises almost ideal representation of every category of expenditures is observed. Small and medium enterprises have very similar structure of ICT expenditures. Furthermore, quite strong overrepresentation in case of financial leasing of ICT devices and underrepresentation for own development or modification of the purchased software can be observed.
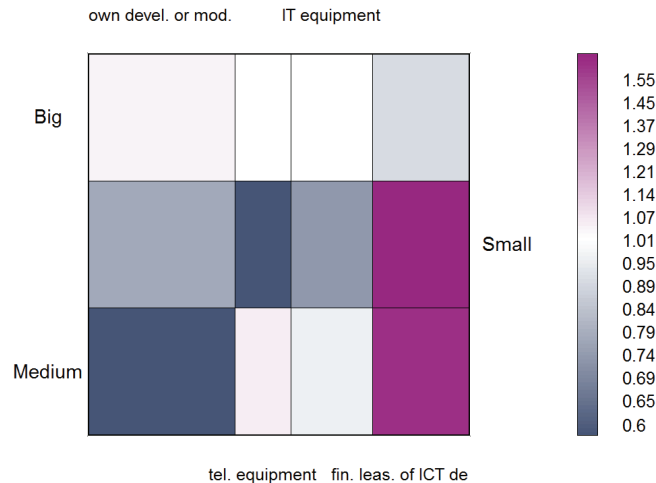
Figure 2. The overrepresentation map presenting the expenditures structure in each of enterprise size groups (purple represents overrepresentation, gray underrepresentation and white areas on the map refer expenses on average level)

In the next stage of this research the analysis for speed of internet connection in each of voivodeships and size groups of enterprises were made. The overrepresentation map (Figure 3) shows subtle overrepresentation of small enterprises with low connection speed (below 10 Mbit/s). On the other hand, underrepresentation of high speed of internet connection (over 30 Mbit/s) could be observed. The medium and big enterprises have very similar structure of using speed of internet connection. In both of them speed over 30 Mbit/s is overrepresented and speed below 10 Mbit/s is underrepresented.

In the case of the study of the ICT usage in each of enterprise size group the overrepresentation maps based on GCA version for equal rows were used. In this way, the effects that complicate the interpretation of results were eliminated. They were related only with the nature of the data, not the occurring phenomena. There were such as the obvious higher level of expenditures in large enterprises (Figure 2) and the greater number of entities (separate contracts for the provision of the Internet) for small enterprises (Figure 3).

On the basis of overrepresentation map in Figure 4 it can be stated that there is no dependence between speed of Internet connection and voivodeship. It can be noticed that in Pomorskie and Swietokrzyskie voivodeship speed below 2 Mbit/s is overrepresented, the highest speed over 100 Mbit/s is overrepresented in Mazowieckie and Podlaskie, whereas in Warminsko-mazurskie, Opolskie and Podkarpackie speed over 100 Mbit/s is underrepresented. It can be concluded that the biggest share of enterprises uses the speed of internet connection between 10 and 30 Mbit/s and between 2 and 10 Mbit/s.
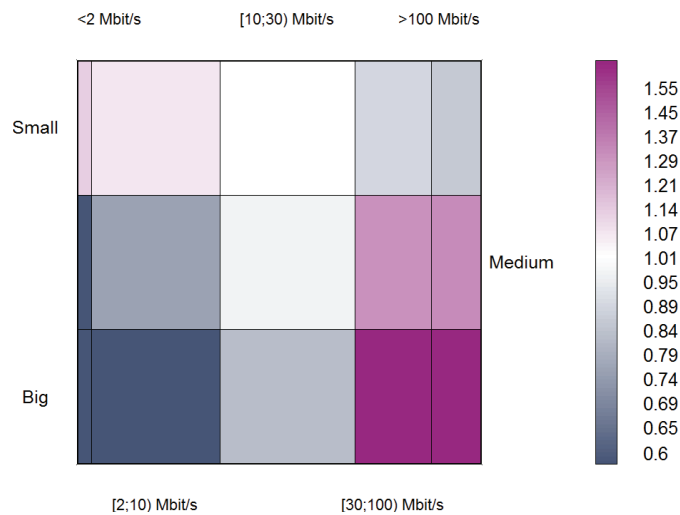
Figure 3. The overrepresentation map visualizing the usage of Internet connection speed in each of enterprise size groups (purple represents overrepresentation, gray underrepresentation and white areas on the map refer expenses on average level)
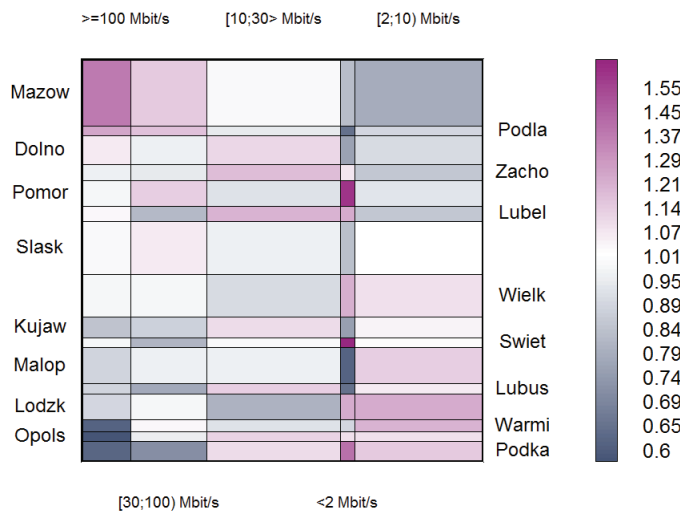

Figure 4. The overrepresentation map showing the usage of Internet connection speed in each of Polish voivodeships (purple represents overrepresentation, gray underrepresentation and white areas on the map refer expenses on average level)

5. Conclusion

The aim of this study was realized by analyses of two important aspects of ICT application in Polish enterprises. The obtained results allow to formulate conclusions regarding both the applied data mining method as well as the investigated issue.

The GCA and overrepresentation maps approach, by the easily and transparent identification of similarities and dissimilarities between variables, is a useful data mining tool, especially in case of multidimensional data and large data volumes. The important advantage is the possibility of visualization, which makes comparison, evaluation the similarities and discovering unknown regularities in the data easier.

In this study a variation in the application of ICT solutions in Polish enterprises was noticed. The identified dissimilarities concern the location of enterprises as well as their size. It can be noticed that the structure of expenditures in case of small and medium enterprises are significantly different from those for big ones. In addition, for various voivodeships the sectors with particularly good or much worse usage of ICT can be identified. This kind of results can be used to indicate precisely the needs during planning of the enterprises development support programs.

It is worth considering to perform analogical analysis for more points in time. These results would allow to know the changes and trends in the use of ICT in Poland over longer time horizon. It could lead to conclusions regarding relations between the economic development and the interest in modern technologies. The addition analysis to study the diversity of the structure of ICT usage in particular groups of enterprises separately for each voivodeship is also the future research plan. This would complement the results of this work, the aim of which is to investigate the structure of ICT adoption across the whole country.

REFERENCES

[1] Adamczewski P. (2011) *E-business applications in Polish SME sector – condition and development*, STUDIA INFORMATICA, Vol. 32, No 2B (97)

[2] Barba-Sánchez V., del Pilar Martínez-Ruiz M., Jiménez-Zarco A. I. (2007), *Drivers, Benefits and Challenges of ICT Adoption by Small and Medium Sized Enterprises (SMEs): A Literature Review*, Problems and Perspectives in Management / Volume 5, Issue 1, 2007, 103-114.

[3] Binderman Z., Borkowski B., Szczesny W., Shachmurove Y. (2012) *Zmiany struktury eksportu produktów rolnych w wybranych krajach UE w okresie 1980—2010*, Metody Ilościowe w Badaniach Ekonomicznych, Tom 13, nr 1, Wydawnictwo SGGW, Warszawa.

[4] Ciok A., Kowalczyk T., Pleszczyńska E., Szczesny W. (1995), *Algorithms of grade correspondence-cluster analysis,* The Coll. Papers on Theoretical and Applied Computer Science.

[5] Consoli D. (2012) *Literature analysis on determinant factors and the impact of ICT In SMEs*, Procedia - Social and Behavioral Sciences 62 (2012), 93 – 97.

[6] Cuevas-Vargas H., Estrada S., Larios-Gómez E. (2016) *The effects of ICTs as innovation facilitators for a greater business performance. Evidence from Mexico.*, Procedia Computer Science 91 (2016), 47 – 56.

[7] Gostkowski M., Jałowiecki P., Tekień A. (2015) *Regional Diversification of Expenditure Structure in Poland: A GCA Approach*, Zeszyty Naukowe SGGW w Warszawie – Problemy Rolnictwa Światowego, XV(4), 71-79

[8] Gostkowski M., Sieradzki D. (2018) *Overrepresentation maps as a tool to analysis of expenditure structure*, to appear

[9] Januszewska M., Jaremen D. E., Nawrocka E. (2015) The effects of the use of ICT by tourism enterprises, Service Management 2/2015, Vol. 16, 65-73.

[10] Kowalski A., Nasierowski W. (2007) *Bariery wdrażania innowacji: perspektywa ICT w polskich MSP*, Organizacja i Kierowanie, Nr 1(2007), p. 87-99

[11] Koszela G. (2016) *Wykorzystanie gradacyjnej analizy danych do klasyfikacji podregionów pod względem struktury agralnej*, Wiadomości Statystyczne 6/2016, 10 – 30.

[12] Leoński W. (2014) *Technologie informacyjno-komunikacyjne jako czynnik poprawy konkurencyjności polskich przedsiębiorstw*, STUDIA I PRACE WYDZIAŁU NAUK EKONOMICZNYCH I ZARZĄDZANIA NR 38, t. 1, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, 181-192.

[13] Tarutė A., Gatautis R. (2014) *ICT impact on SMEs performance*, Procedia - Social and Behavioral Sciences 110 (2014), 1218 – 1225.

[14] Ząbkowski T., Szczesny W. (2012) *Badanie atrakcyjności oferty dostępu do Internetu za pomocą analizy gradacyjnej*, Metody Ilościowe w Badaniach Ekonomicznych, Tom XIII/3, 2012, 276 – 287.