MANAGING RELATED VARIETY IN SMART SPECIALISATION INDUSTRIES: COMPANY-SPECIFIC DEVELOPMENT PATHS

Korneliusz Pylak – Dariusz Majerek

Abstract

The aim of this paper is to create a new, structured definition of different industries' related varieties involved in regional smart specialisations. To define related variety in each industry, we used machine-learning-method decision trees. The input and target variables are the number of companies from 86 industries located in 2,531 communities in Poland. Decision trees allow us to predict how many companies from each industry exist in communities, given the precise number of companies from related industries and the number of communities in which these relationships occur. The trees indicate the most common structures of related industries. Our findings confirm that related variety differs in size and scope for every industry and includes companies both within and outside the natural value chain (suppliers and clients). The findings prove the utility of the new definition of related variety, as defining related variety properly and precisely may facilitate the process of implementing and developing smart specialisations in regions.

Key words: related variety, entrepreneurial discovery, regional innovation strategy, development path

JEL Code: O21, O31, R11

Introduction

Smart specialisations, introduced by Foray, David, and Hall (2009), have been a crucial issue in European Union regional policy since 2008 and are currently a basis for managing European Structural and Investment Funds in the programme period from 2014 to 2020 (Benner, 2014). Implementing European policies on research, technology, and development (Foray, 2009) may contribute to the growth of regions, including the convergence of weaker regions, by revealing through entrepreneurial discovery one or several areas of science and technology in which one region may have an advantage over others (Foray et al., 2009). These areas are called smart specialisations. Obviously, growth in smart specialisation industries will result in changes in related industries and lead to the transformation or reinvention of the whole related economy. The relatedness of industries may reflect the impact of changes triggered by supported smart specialisations. However, the economic literature lacks a consensus regarding how to measure related variety (Boschma & Iammarino, 2009; Frenken, Van Oort, & Verburg, 2007).

Thus, the aim of this paper is to create a new, structured definition of different industries' related varieties involved in regional smart specialisations. The paper is structured as follows: The first section explores and clarifies the theoretical background of related variety and smart specialisations. The second section lays out the empirical design of the research, including methods, data, and measurement. The third section presents and discusses the empirical results of the analysis of related variety in different smart specialisation industries. The fourth section offers a conclusion regarding the scale and scope of revealed related varieties.

1 Managing related variety in smart specialisation industries

Related variety can be defined as a group of industries that have shared or complementary competences (consisting of knowledge, skills, and attitudes) and involves knowledge networks that extend beyond business networks (Boschma & Iammarino, 2009). Related variety can go beyond the generally accepted value chains and competitive forces in the marketplace (Porter, 1979) as well as clusters connecting manufacturers, suppliers, and specialised service providers (Porter, 2003) and cover entire value chains of products and services based on complementary competences and belonging to different industries (Boschma & Iammarino, 2009). Innovativeness and niches leading to the renewal of the economy and boosts in growth occur mostly between different, purportedly unrelated industries (Foray, 2014). Thus, we cannot measure related variety only through comprehensive and common value chains such as this: agriculture → food production → food services.

There is no consensus in the literature on the definition and measurement of related varieties of industries. Frenken et al. (2007) claim that related industries belong to the five-digit Nomenclature statistique des Activités économiques dans la Communauté Européenne (NACE) codes (e.g. A.01.11) within the two-digit NACE codes (e.g. 01). This is not a full-fledged approach to measuring related variety, as there are many examples of similar knowledge-based industries that belong to different sectors of the economy (e.g. C.10 Manufacture of food products and I.56 Food and beverage service activities). Hidalgo, Barabási, Winger, and Hausmann (2007) presented another approach, creating a product space theory based on the belief that similar knowledge is needed to create the products and services individual countries most commonly export. This assumption also applies at the regional level, which means that one can identify related industries by analysing their co-existence in different regions. This is also relevant because the products and services produced in a region do not necessarily reflect the region's export profile (Boschma, Minondo, & Navarro, 2012). We can therefore distinguish three approaches to measuring related variety: 1) exclusively within the sector, 2) within clusters and value chains, and 3) within industries with shared knowledge and expertise. The latter approach is the most versatile and flexible and covers industries that appear unrelated at first glance. Such an understanding of related variety is also the basis of the concept of smart specialisation and thus will be used and elaborated on in this study.

If we support a smart specialisation industry, we automatically support all of its related industries because of localisation and urbanisation economies. Localisation economies include (1) the easy flow of skills and knowledge between companies, (2) easy access to a skilled workforce or public goods within related industries, (3) financial savings thanks to cooperation or subcontracting some parts of the production process, and (4) lower transportation costs of materials and intermediate products (Henderson, 2003). Urbanisation economies are related to concentrated demand and high density of economic activity, which cause an easy flow of people and knowledge between formally unrelated industries within close neighbourhood (Jacobs, 1969, 97–98). Thus, the more extensive the related variety, the greater the effects on the economy may be. We need to analyse related varieties of smart specialisation industries supported in Poland to find out which specialisations will affect the economy to the greatest extent.

At present, all Polish voivodships have discovered smart specialisations. Although the process of discovering smart specialisations was often cursory and intuitive, regional authorities officially support these specialisations (Nazarko, 2014), and thus we can analyse them in our research. Smart specialisations cover mostly manufacturing industries, including high-technology industries (food production, chemical and pharmaceutical industries, biotechnologies, nanotechnologies, production of machinery and equipment, manufacture of wood and furniture, etc.) and to a lesser extent, services (mainly tourist services, health services, information and communication technologies, business services, creative services, etc.) as well as energy and construction. Thus, we will focus on these industries when discussing the empirical results.

2 Research design

The aim of the research is to define related variety in each industry. The most common method for that purpose is multiple regression, but in our case, the relationships among predictors are complex, and endogenous variables are collinear and cause redundancy. Therefore, we cannot use multiple regression to define related variety. Instead, we used a machine learning method, decision trees, which do not require the structure of relationships between predictors to be known, and redundancy is not an issue (Kuhn, 2008). Decision trees allow us to predict how many companies from each industry exist in communities, given the precise number of companies from related industries and the number of communities in which these relationships occur. Thus, the trees describe the most common structures of related industries. Decision trees also allow us to generalise the related variety structure by decreasing the complexity parameter, although doing so lowers the predictive performance of the method.

We validated the predictions by dividing the sample randomly into two groups: a training group used to train the model and a test group used to validate the model's prediction performance, accounting for 75% and 25% of the records, respectively. The validation of the model was based on the analysis of mean-square errors *(mse)* given by (1)

$$mse = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad (1)$$

where N is the sample size, y_i is the observed value, and $\hat{y}_i$ is the fitted value from the model. To define the related variety of different industries, we used the number of companies from 86 industries located in 2,531 communities in Poland in 2009 as input and target variables.

3 Empirical results

Decision trees created for Polish industries revealed that related variety differs significantly across smart specialisation industries in size and type of relationships (see Tab. 1). Although the predictive performance of the trees is not satisfactory (the square root of *mse* is higher than the number of companies most expected in the analysed industry in every model), related industries can still be indicated without the precise numbers of companies.

First, the number of related industries in every model is not high; it does not exceed six industries (and only C_16 and F_42 have six). Such a low number of industries is a result of the high generalisation abilities of decision trees; the trees contain only crucial predictors. As shown in Tab. 1, most of the related industries belong to the value chain of the analysed industry (suppliers, clients, or partners), especially in the service sector. Moreover, related industries are mostly client industries.

Tab. 1: Characteristics of related varieties of smart specialisation industries based on the decision tree machine learning method

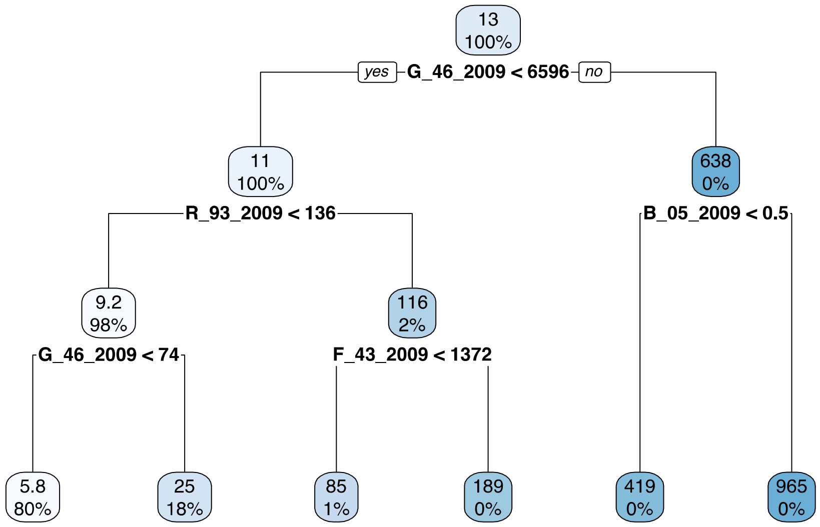
Industry	The most likely average number of companies from the industry in the model	The square root of *mse*	Related industries (the order reflects the importance of the industry as a predictor in the model; industries in the value chain are in bold)	Observations on the related industries and their roles in the value chain
C_10	5.80	9.56	**G_46**; **R_93**; F_43; B_05	G_46 & R_93 are clients.
C_16	8.40	15.05	**F_41**; **C_32**; **A_02**; C_10; S_96; A_01	F_41 is a client industry; C_32 & A_02 are suppliers.
C_17	0.75	2.68	C_21; C_22; **M_69**; **A_02**	A_02 is a supplier; M_69 is a service provider.
C_20	0.85	2.48	**C_13**; M_71; **C_18**	C_13 & C_18 are clients.
C_21	0.10	0.56	J_63; S_94	*The industries are outside the value chain.*
C_26	0.95	22.61	**U_99**; **H_53**; **J_62**	All industries are possible clients.
C_28	1.10	11.52	**E_37**; **R_90**; **G_46**; **A_01**	All industries are possible clients.
C_31	3.60	11.95	**G_45**; **S_95**; **D_35**; **Q_86**; **B_09**	All industries are possible clients.
C_32	3.90	25.14	**G_46**; J_63; P_85; B_05	G_46 is a client.
D_35	0.84	8.38	**C_30**; **O_84**; **K_64**; **C_33**	C_30 & O_84 are possible clients, and K_64 & C_33 are suppliers.
F_41	34.00	67.29	**G_47**; **E_38**; **K_66**; **A_02**	E_38 & A_02 are suppliers; G_47 & K_66 are clients.
F_42	5.00	19.58	G_46; G_47; **F_43**; N_82; B_06; A_02	F_43 is a supplier.
F_43	49.00	111.68	**S_96**; C_33; A_01	S_96 is a potential client.
J_62	6.80	21.08	L_68; **J_60**; **M_69**	J_60 is a supplier; M_69 is a client.
J_63	2.40	8.11	**B_09**; **C_33**; A_01	B_09 & C_33 are clients.
K_64	3.70	12.43	**H_51**; **R_92**; **P_85**; **A_01**	All industries are possible clients.
M_69	10.00	41.57	**F_43**; **Q_86**; **S_96**; **A_01**	All industries are possible clients.
M_70	8.50	25.52	**J_62**; **I_56**	All industries are possible clients.
M_72	0.37	1.68	**B_09**; **H_52**; **A_01**	All industries are possible clients.
M_73	6.20	28.69	**C_28**; **J_60**; **M_69**	All industries are possible clients.
M_74	12.00	34.40	**E_39**; **H_49**; **K_66**	E_39 & H_49 are clients; K_66 is a supplier.
N_79	2.00	10.98	**I_56**; **H_53**	All industries are suppliers.
Q_86	24.00	90.19	**J_60**; **H_49**; **P_85**; **E_38**; **B_09**	All industries are suppliers.
R_90	3.60	10.93	B_09; R_92; P_85; A_01	*The industries are outside the value chain.*

Source: Our own estimation based on data from the Central Statistical Office of Poland. Abbreviations are NACE codes: A_01: Crop and animal production, hunting and related service activities; A_02: Forestry and logging; B_05: Mining of coal and lignite; B_06: Extraction of crude petroleum and natural gas; B_09: Mining support service activities; C_10: Manufacture of food products; C_13: Manufacture of textiles; C_16: Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials; C_17: Manufacture of paper and paper products; C_18: Printing and reproduction of recorded media; C_20: Manufacture of chemicals and chemical products; C_21: Manufacture of basic pharmaceutical products and pharmaceutical preparations; C_22: Manufacture of rubber and plastic products; C_26: Manufacture of computer, electronic and optical products; C_28: Manufacture of machinery and equipment; C_30: Manufacture of other transport equipment; C_31: Manufacture of furniture; C_32: Other manufacturing; C_33: Repair and installation of machinery and equipment; D_35: Electricity, gas, steam and air conditioning supply; E_37: Sewerage; E_38:

Waste collection, treatment and disposal activities; materials recovery; E_39: Remediation activities and other waste management services; F_41: Construction of buildings; F_42: Civil engineering; F_43: Specialised construction activities; G_45: Wholesale and retail trade and repair of motor vehicles and motorcycles; G_46: Wholesale trade, except of motor vehicles and motorcycles; G_47: Retail trade, except of motor vehicles and motorcycles; H_49: Land transport and transport via pipelines; H_51: Air transport; H_52: Warehousing and support activities for transportation; H_53: Postal and courier activities; I_56: Food and beverage service activities; J_60: Programming and broadcasting activities; J_62: Computer programming, consultancy and related activities; J_63: Information service activities; K_64: Financial service activities, except insurance and pension funding; K_66: Activities auxiliary to financial services and insurance activities; L_68: Real estate activities; M_69: Legal and accounting activities; M_70: Activities of head offices; management consultancy activities; M_71: Architectural and engineering activities; technical testing and analysis; M_72: Scientific research and development; M_73: Advertising and market research; M_74: Other professional, scientific and technical activities; N_79: Travel agency, tour operator and other reservation service and related activities; N_82: Office administrative, office support and other business support activities; O_84: Public administration and defence; compulsory social security; P_85: Education; Q_86: Human health activities; R_90: Creative, arts and entertainment activities; R_92: Gambling and betting activities; R_93: Sports activities and amusement and recreation activities; S_94: Activities of membership organisations; S_95: Repair of computers and personal and household goods; S_96: Other personal service activities; Activities of households as employers of domestic personnel; Undifferentiated goods- and services-producing activities of private households for own use; U_99: Activities of extraterritorial organisations and bodies.

Surprisingly, although related variety consists of industries from the value chain, it often does not contain the key and most intuitive client or supplier. The food producing industry (C_10) is just one example (see Fig. 1); we do not see industry A_01 (crop and animal production, hunting and related service activities) playing a crucial role as a supplier. However, industry G_46 (wholesale trade, except of motor vehicles and motorcycles) plays a crucial role as a client industry in the real economy and in the model because the industry is located on top of the tree. This industry also appears at the bottom of the tree in the most probable leaf. According to the leaf, in 80% of communities in which industry C_10 exists, if there are fewer than 136 companies providing sports, amusement, and recreation activities (R_93) and fewer than 74 wholesale trade companies, the number of food manufacturers will be approximately 5.8.

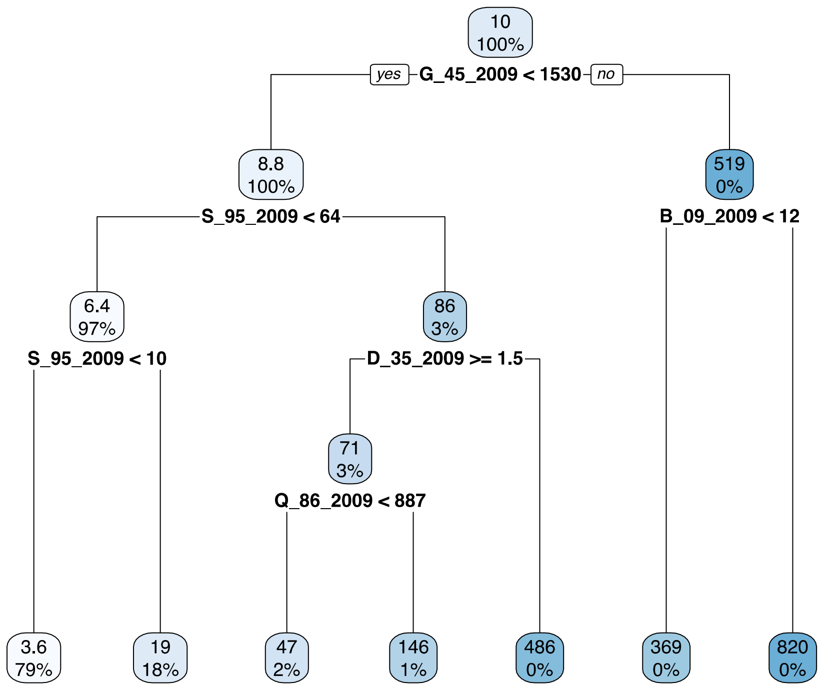
Fig. 1: The related variety decision tree of industry C_10



Source: Our own estimation based on data from the Central Statistical Office of Poland.

The same situation occurs in the related variety of the furniture manufacturing industry (C_31). Industry C_31 should be strongly associated with wood manufacturing (C_16) as a supplier; however, C_16 is not represented in the tree (see Fig. 2). All the industries in the related variety are potential customers.

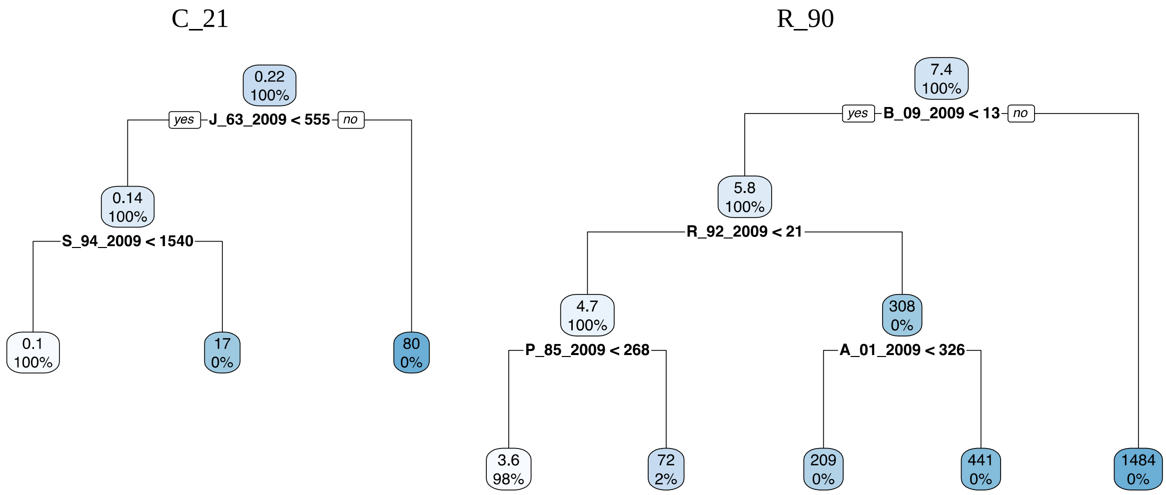
Fig. 2: The related variety decision tree of industry C_31



Source: Our own estimation based on data from the Central Statistical Office of Poland.

We can also indicate two industries in which the related varieties do not contain any industries from their value chains (see Fig. 3). These industries are the manufacture of pharmaceutical products (C_21) and creative, arts, and entertainment activities (R_90). It is unlikely that the related industries indicated in the trees belong to the same knowledge pool, as, for example, the manufacture of pharmaceuticals and (C_21) and activities of membership organisations (S_94) or creative, arts, and entertainment activities (R_90) and mining support service activities (B_09) have totally different knowledge bases. In these two cases, it is most likely that related industries create urbanisation economies that make up the foundations of related varieties.

Fig. 3: The related variety decision trees of industries C_21 and R_90



Source: Our own estimation based on data from the Central Statistical Office of Poland.

Conclusion

The aim of the paper was to create a new, structured definition of different industries' related varieties involved in regional smart specialisations. We confirmed that related variety is a group of industries that have shared or complementary competences (consisting of knowledge, skills, and attitudes) and involves knowledge networks that extend beyond business networks. However, we discovered that a shared knowledge pool is not all that links related industries. Related industries very often co-exist in the same community because these industries commonly create urbanisation economies. This finding should be elaborated on in future research.

Our findings confirm that related variety differs in size and scope for every industry and includes companies both within and outside the natural value chain (suppliers and clients). There are more client than supplier industries in the related varieties of the analysed industries. However, we discovered that related variety often does not contain the main or the most intuitive client or supplier. This finding may suggest that spatial proximity is not very important for such industries, as we focused on communities in our analysis. Thus, in future research we suggest analysing related variety in different spatial areas (such as in Pylak & Majerek, 2017).

The findings prove the utility of the new definition of related variety, as defining related variety properly and precisely may facilitate the process of implementing and developing smart specialisations in regions. Thanks to our findings, innovative entrepreneurs may focus their development strategies more on discovering new possibilities within related industries, and regional decision-makers can more adequately shape policies to meet entrepreneurs' needs.

References

Benner, M. (2014). From smart specialisation to smart experimentation. Building a new theoretical framework for regional policy of the European Union. *Zeitschrift Fur Wirtschaftsgeographie, 58*(1), 33–49.

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography, 85*(3), 289–311. doi:10.1111/j.1944-8287.2009.01034.x

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. *Papers in Regional Science, 91*(2), 241–256. doi:10.1111/j.1435-5957.2011.00387.x

Foray, D. (2009). Understanding smart specialisation. In D. Pontikakis, D. Kyriakou, & R. van Bavel (Eds.), *The question of R&D specialisation: Perspectives and policy implications* (pp. 14–26). Brussels: European Commission, Joint Research Centre – Institute for Prospective Technological Studies, Directorate General Research.

Foray, D. (2014). From smart specialisation to smart specialisation policy. *European Journal of Innovation Management, 17*(4), 492–507. doi:10.1108/EJIM-09-2014-0096

Foray, D., David, P. A., & Hall, B. (2009). Smart specialisation—the concept. *Knowledge Economists Policy Brief, 9*, 1–5.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies, 41*(5), 685–697. doi:10.1080/00343400601120296

Henderson, V. J. (2003). Marshall's scale economies. *Journal of Urban Economics, 53*(1), 1–28. doi:10.1016/S0094-1190(02)00505-3

Hidalgo, C. A., Barabási, A. L., Winger, B., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317*(5837), 482–487. doi:10.1126/science.1144581

Jacobs, J. (1969). *The economy of cities*. New York: Random House.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26.

Nazarko, Ł. (2014). Inteligentne specjalizacje polskich regionów – przyczynek do ewaluacji. *Przedsiębiorczość i zarządzanie, XV*(8), 247–262.

Porter, M. E. (1979). How competitive forces shape strategy. *Harvard Business Review, 57*(2), 137–145.

Porter, M. E. (2003). The economic performance of regions. *Regional Studies, 37*(6–7), 549–578. doi:10.1080/0034340032000108688

Pylak, K., & Majerek, D. (2017). Identifying industries requiring spatial and sectoral relationships to develop: Do smart specializations always work? *Procedia Engineering, 174*, 1119–1127. doi:10.1016/j.proeng.2017.01.265

Contact

Korneliusz Pylak
Lublin University of Technology
Nadbystrzycka 38D St., 20–618 Lublin, Poland
korneliusz.pylak@pollub.pl

Dariusz Majerek
Lublin University of Technology
Nadbystrzycka 38D St., 20–618 Lublin, Poland
d.majerek@pollub.pl