ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2015
Seria: ORGANIZACJA I ZARZĄDZANIE z. 82 Nr kol. 1940

Miki SIROLA
Aalto University, PO Box 15400, FI-00076 Aalto, Finland,
Miki.Sirola@aalto.fi

PERSPECTIVES IN SCIENCE – INTERDISCIPLINARY EXAMPLES IN DATA-ANALYSIS VISUALIZATION AND POLITICAL DECISION MAKING

Summary. This article discusses perspectives in science with a somewhat limited scope. The areas of science included reflect the interest and experience of the author with an interdisciplinary view. Scientific domains, applications and methodologies constitute a characteristic entity.

The three main pillars of the model are psychology, civics and energy production. The psychology here is mainly adapted in work environment. The psychology has two main branches one concentrating on individual behaviour and the other on group behaviour. The latter connected to civics converges towards sociology. The energy production represents the industrial view in the model, and it can be broaden wider in the general model, although here we concentrate mostly on the mentioned application area.

From psychology we get assistance in user interfaces and visualizations. One branch in civics is politics. From the energy production and civics we get into production models. From psychology and politics we derive a path to decision making. Also from user interfaces, visualizations and even production models we get into decision making. System technics has a connection to decision making as well as to the other fields mentioned here. Data analysis can be used as a tool in all these fields. In addition scientific visualization is discussed, and visualization examples are presented.

The use of data analysis in political decision making is presented as a more detailed example. For instance in predicting the voting behaviour in a society data from pre-election gallup poll reviews and voting advice applications can be used. This political example shows the opportunity to cross fields in science to derive interesting results.

Keywords: science, system technics, data analysis, visulaization, political decision making.

PERSPEKTYWY W NAUCE – INTERDYSCYPLINARNY PRZYKŁAD ANALIZY DANYCH W WIZUALIZACJI I PODEJMOWANIA DECYZJI W POLITYCE

Streszczenie. W tym artykule omówiono perspektywy zastosowania różnych podejść w nauce, jeśli chodzi o analizę danych i ich wizualizację. Dziedziny nauki omówione w publikacji odzwierciedlają zainteresowania i doświadczenia autora.

Omówione dziedziny nauki, programy i metody mają charakter interdyscyplinarny i wynikają z badań autora.

Trzy główne filary modelu to: psychologia, społeczeństwo i produkcja energii. Psychologia tutaj jest analizowana głównie w środowisku pracy. Ma ona dwa główne działy: jeden koncentruje się na indywidualnych zachowaniach, a drugi na zachowaniach grupowych. Produkcja energii reprezentuje przemysłowy element modelu.

Od psychologii możemy uzyskać pomoc w tworzeniu interfejsów użytkownika i przy wizualizacji danych. Z kwestii dotyczących produkcji energii i społeczeństwa uzyskujemy dane do tworzenia modeli produkcyjnych. Na podstawie psychologii tworzymy ścieżkę podejmowania decyzji. Również z wizualizacji możemy nawet przejść do procesu podejmowania decyzji. Analiza danych może być stosowana jako narzędzie w tych obszarach. W artykule przedstawiono przykłady wizualizacji.

Wykorzystanie analizy danych w podejmowaniu decyzji politycznych jest przedstawione w pracy w sposób najbardziej szczegółowy. Rezultaty przedstawiono na przykładzie zachowań wyborców. Wykorzystano w tym celu wyniki sondażu Gallupa i na tej podstawie opracowano aplikację, która może pozwolić na prognozowanie wyników głosowania.

Słowa kluczowe: nauka, urządzenia systemowe, analiza danych, wizualizacja, podejmowanie decyzji.

1. Introduction

The scientific domain is manifold, and the various branches, fields and perspectives constitute a huge entity, which is very difficult to follow and be familiar with completely. Therefore most scientists have specialized themselves to one rather narrow topic or area. Also the different research methodologies divide scientists into various alignments. However, an interdisciplinary view is sometimes very welcome perspective to integrate different pieces together.

This article discusses perspectives in science in limited and partly subjective scope. The areas of science here reflect the interest and experience of the author in an interdisciplinary view. Scientific domains, applications and methodologies constitute a schematic entity, see Fig. 1.

The three main pillars in the model are psychology, civics and energy production, see Fig. 1. The psychology here concerns mainly work environment. The psychology has two main branches: individual behaviour and group behaviour. The latter connected to civics converges towards sociology. The energy production represents the industrial view in the model, and it can be broaden wider in the general model, although here we concentrate on the mentioned application area.

From psychology we get assistance in user interfaces and visualizations, see e.g. [9] and [13]. One branch in civics is politics. From the energy production and civics we get into production models. From psychology and politics we derive a path to decision making. Also from user interfaces, visualizations and even production models we get into decision making.

System technics has a connection to decision making as well as to the other fields mentioned here. Data analysis can be used as a tool in all of these fields.

The use of data analysis in political decision making is presented in an example including technical details of the analysis. For instance in predicting the voting behaviour in a society data from pre-election Gallup poll reviews and voting advice applications is used. With this political example we show the opportunity to cross fields in science to derive interesting results. This combination of methods used is an example of a new perspective.

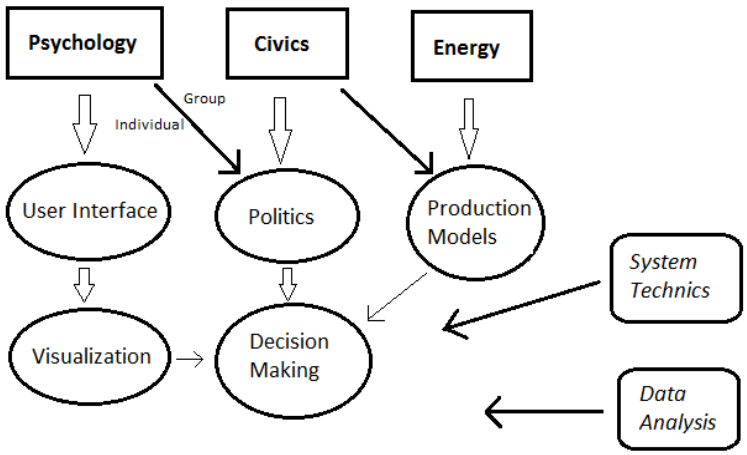

Fig. 1. Perspectives in science in a schematic graph
Rys. 1. Trzy perspektywy procesu podejmowania decyzji

2. Related work and basis background

The perspectives of sciences are largely discussed in literature from many different perspectives. Related articles are found e.g. in [3, 5, 7, 14, 25].

The topic is covered in very many views. A division in four categories according to learning issues is introduced in [3]. Article [14] discuss about constructivist perspectives where the constructivism is the centered concept. A philosophical bridge between language and mathematics is built, and the communication issue is addressed.

In [14] sociocultural perspectives are analyzed. The basic concepts come from cognitive psychology. Knowledge, politics and cultural beliefs are challenged. Metaphors and practices have scientific influence taking into account the economic and political forces, and also education. The entity is broad and diversified.

Psychological basis in user interfaces and visualization have an important role. A large analysis of the requirements in the control rooms is in [9]. Expanding the control room

concept with a work psychologist view is described in [13]. Adding a perspective from risk analysis adds in an interdisciplinary view [15]. Political participation is analyzed in [10].

Energy production is a civic issue solved in various ways in different nationalities. From nuclear energy production [16] to coal, oil, turf as well as renewable sources such as solar power and wind power. All different sources of energy are widely used.

The control theory and system technics add an interdisciplinary view to the figure [6]. Decision theory and value theory together constitute decision analysis method [2, 4], which is one important basis in decision making. From data analysis methodologies Self-Organizing Map (SOM) [12] is one of our interest areas. Combining these methodologies we achieve mentioned goals. The control theory and system technics constitute a general view of relations and therefore suits well to draw interdisciplinary figures.

In our work we have used all these mentioned perspectives in the model in some phase. Several years project in energy production is summarized in [18]. Methodologies from system technics and data-analysis are used in constructing production models and support decision making. From psychological side we have got support to user interfaces and visualizations. One interesting area not covered yet is user modeling. On civics side one of our very recent studies is about using data-analysis methods in political decision making [20].

We provide a summary of results that deals with perspectives in science that has influenced the development of user environments and the application domain of data visualization.

Visualization provides a bridge that links data analysis to different disciplines such as psychology, civics and energy production. Psychological aspect has played a significant role in the development of work environments and in particular visualization environments. The contributions from psychology are documented in recent studies, but have historically been a cornerstone in the development of human factors research.

With the application of new technologies in visual data analysis, examples of decision making capabilities illustrates the power of visualization. These examples include the visual quantification of customer rejection classes and a regional and political demographic breakdown of the public voting trends during the Finnish presidential election. These examples also demonstrate the visual perspectives that are enhanced by the application of different analysis techniques.

3. Visualization of data

In very many scientific fields there are needs to visualize various concepts and experimental results. The visualization gives to the numbers a figure and characterizes wished features. In our projects we have used visualizations for various purposes, and developed them as well.

In Fig. 2 we have an example visualization of rejection reason classes from car inspection data [21]. The rejection reason classes are the red constant size balls in the figure. Old cars are represented by blue balls and newest by green balls. Edges exist between rejection reasons and cars. Sizes of the car balls and edges between cars and rejection reasons are proportional with the rejection rates. The rejection reason classes are not seen in detail in this figure. The rejection reasons can be such as brakes, tyres, shock absortion, parking brake, lights, etc. or the rejection reason can be unclassified.

The data is got from Finnish A-katsastus [27] and the reliability rating reports such as the German TÜV reports [28] are reflected in the study. The visualizations here are based on Principal Component Analysis (PCA) method [8]. Principal Component Analysis is a method for linear transformation. The dimension of the data is reduced by transforming it to a new coordinate system such that the great variance lies on the first component.

The network layout here is based on the ForceAtlas2 (FA2) method [11]. ForceAtlas2 is a continuous algorithm and the model is based on attraction and repulsion proportional to distance between nodes. Various layouts are achieved with different initial coordinates and parameter settings. The main goal is to produce a readable spatialization and devise an energy model that could be easily understood by users. A clear visualization where nodes are separated and not overlapped is reached by various parameter settings. This feature can be forced in Gephi software [1]. In our model default values were increased and gravity decreased to obtain sparser graph.

Visualization is certainly important in user interfaces, and the modern user interfaces have also become more and more interactive. But also other fields and scientific perspectives utilize visual information and forms of visualization in various ways. Actually there are only quite few scientific branches that do not use visual information at all.

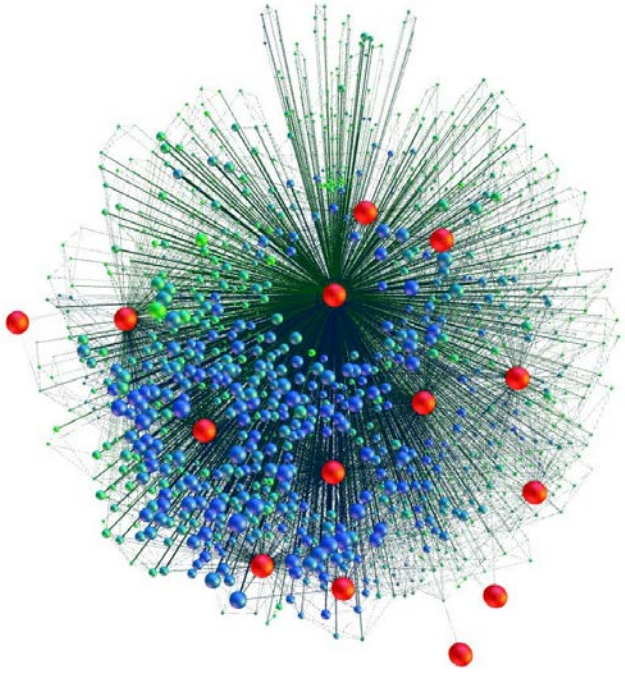

Fig. 2. An example visualization from analysis of technical data
Rys. 2. Przykład wizualizacji z zakresu analizy danych technicznych

As another visualization example we have a Self-Organizing Map based on SOM method from a similar application area as in Fig. 2. In Fig. 3 a process visualization tool [22] is presenting the same rejection reason classes from car inspection data as before now in a form of a SOM map, and also concentrating on presenting various features of this visualization tool.

Multidimensional data is used to draw conclusions on the structure of car inspection data. The goal here is to visualize filtered data on a low dimension to preserve as much information as needed. An add-in for the SOM Toolbox [24] was developed. SOM Toolbox is a flexible general-purpose software library developed in the Helsinki University of Technology.

Collaborative Filtering (CF) [26] method is used here together with the Self-Organizing Map method [12]. Collaborative Filtering is used in the data preprosessing before training the SOM map. Missing value imputation and data filtering are done with the CF method. The method was able to filter both zero and one values reliably with a suitable regularization parameter selection.

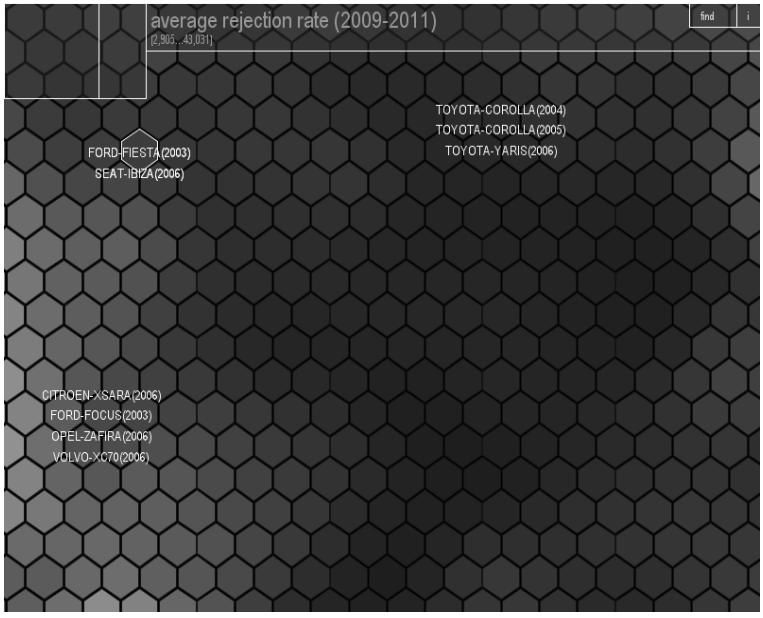

Fig. 3. A SOM visualization in grey scale colouring produced by a special visualization tool
Rys. 3. Wizualizacja SOM, przygotowana w szarej kolorystyce, wykonana przez specjalne narzędzie do wizualizacji

High dimensional data can be visualized in low dimensional views by using self-Organizing Map (SOM) method. The method has two phases, training and mapping, as if in most artificial neural networks. The SOM consists of neurons which are usually initialized with small random numbers. In the iterative training phase, one sample vector from the input data is chosen randomly.

The distance measurement between it and the weight vectors of the SOM is derived. The weight vectors are updated so that the closest neuron weight vector with input is moved closer

to the input vector in the input space. The topological neighbours of the so called best matching unit are moved in a similar way weighted by the neighbourhood function.

A visualization tool was developed. The tool was implemented using the Processing language on the Android platform. Some of the functionalities such as zooming and component plane selection are based on the mouse touch positions. The tool is most useful in cases where the data consists of a large amount of label information or dozens of komponent planes.

The SOM map can clarify for instance the cluster structure in the data. We have used the SOM visualizations in many different applications in our research, and shown the usefulness of the method also in new application branches. Visualization is an important part of our model as well.

4. An example of political decision making

In this chapter we present how to support political decision making with data analysis methods keeping in mind all forms of visualizations. This example is presented more in detail in [20], and here we outline only the main features. This paper uses data from the last Finnish presidential election.

The data sources in this example are repeated Gallup results and a voting advice application in internet. This data is compared with election results. The analysis has the following steps: Gallup preprosessing, sample size estimation with the data information, and candidate support in time. The methods are partly based on rationality estimation with parameter optimization. The experiments support the basic figure and the hypothesis. The methods combine two temporal data sets.

Another political example analyzes parliamentary elections based on voting advice application data, and is discussed in detail in [23]. The main goal in this paper is to model values of Finnish citizens and the members of the parliament. Two databases are combined: voting advice application data and the results of the parliamentary elections in 2011.

The data is converted to a high-dimensional space. It is projected to two principal components. The projection allows visualize the main differences with the parties. The value grids are produced with a kernel density estimation method without explicitly using the questions of the voting advice application.

Meaningful interpretations for the axes in the visualizations with the analyzed data were found. All candidate value grids are weighted by the results of the parliamentary elections. The result can be interpreted as a distribution grid for Finnish voters' values.

Technical details of this application will follow. Different phases of this analysis are described beginning from the data preprosessing and ending in results and discussion.

Questions with a large amount of factual information were selected for the Voting Advice Application (VAA). The questions were divided into such subgroups according to the topic as general questions, pensions, economy, taxes, defense, foreign countries, domestic, and localities. It can be assumed that the candidates used more time with the questions than the citizens on average.

In the Voting Advice Application it was possible to give weights by the user for the questions depending on their importance. Three different stages could be selected. In traditional analysis the question importance and multiple choices of questions are usually omitted.

In data preprosessing the qualitative data was first converted to matrix form. This is a kind of transfer from qualitative analysis into quantitative analysis including careful assessment. Part of the important information is lost, but this bias cannot be completely avoided in moving into numerical analysis.

The methods used here are the Principal Component Analysis (PCA) [8] and kernel density estimation [17]. To high dimensionality data first dimensionality reduction is done. Feature extraction transforms the data on the high-dimensional space to a space of fewer dimensions. The transform is linear in Principal Component Analysis method, but also nonlinear dimensionality techniques exist [8].

Principal component analysis is a useful tool for finding relevant variables for the system and model. It is a linear transformation to a new lower dimensional coordinate system while retaining as much as possible of the variation, especially with highly uncorrelated data. In this application the PCA was selected as the main method, because results are taken into consideration. The PCA method does not necessarily give the best results, but it is much more intuitive and less complicated than multi-variant data analysis. In addition orthogonal axes make visualizations easier to read, especially if the data is not highly correlated.

Kernel density estimation is a non-parametric fundamental data smoothing technique where inferences about the population are made based on a finite data sample [17]. In our experiments several matrices (grids) were defined for densities. Here missing value imputation is done for the data.

In Fig. 4 the six stages of the parliamentary data analysis are seen. The concept of data mining process varies depending on the current problem [23]. Many decisions in e.g. parameter selections are based on intermediate results. Different candidate classifications and visualizations by histograms, scatter plots, factor analysis, Principal Component Analysis (PCA), Self-Organizing Map (SOM) and Multidimensional scaling (MDS) have been published after the data release.

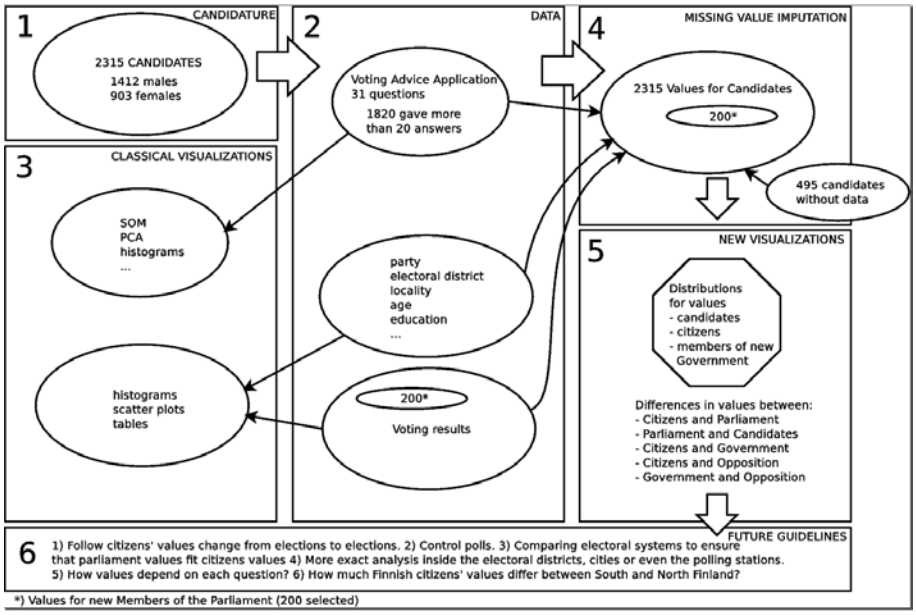

Fig. 4. Parliamentary data analysis process including six different stages
Rys. 4. Proces analizy danych dotyczących wyborów parlamentarnych, zawierający sześć faz przygotowywania wizualizacji

The experiment setup was done according to the schematic diagram in Fig. 4. The six stages are seen as boxes in the figure: candidates, data, classical visualization, missing value imputation, new visualizations and future guidelines.

After preprocessing and the analysis three different results were visualized as plots. They are values of candidates and parties, values of Finnish voters, and values of the parliament.

The acronyms in the three figures are the Finnish political parties. In order according to the current magnitude as members in parliament (or voting percentage with the smallest ones) they are national coalition party (KOK), the Finnish social democratic party (SDP), true Finns (PS), centre party of Finland (KESK), left-wing alliance (VAS), green league (VIHR), Swedish people's party in Finland (RKP), Cristian democrats in Finland (KD), communist party of Finland (SKP), Finnish seniors party (SEN), communist workers' party (KTP), Finnish labour party (STP), independence party (IPU), for the poor (KA), pirate party of Finland (PIR), change 2011 (M2011) and liberty party – future of Finland (VP).

In Fig. 5 the visualization can be used to get an estimate how the median values of candidates in each party (labels) have relations with each other. The numbers in contour lines describe the densities of candidates.

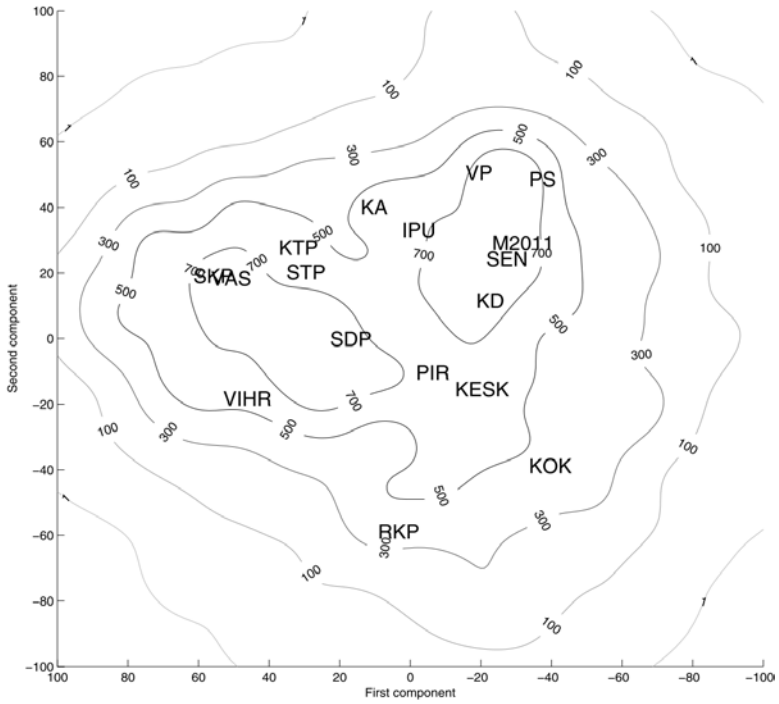

Fig. 5. The political values of the candidates
Rys. 5. „Wartości" poszczególnych kandydatów

In Fig. 6 the visualization can be used to get an estimate for one party how the opinions of the supporters are related to the supporters of other parties. Each party label is a mode of its supporters and those are in smaller font, if the density of the party supporters (in the mode) is less than some other party's density. The numbers in contour lines describe the density of citizens in thousands.

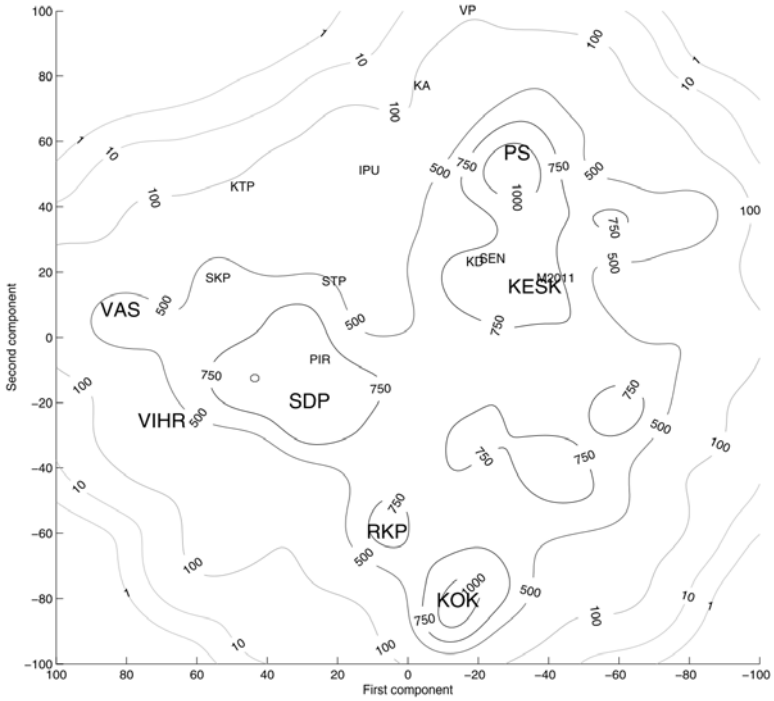

Fig. 6. The values of the Finnish voters
Rys. 6. „Wartości" fińskich wyborców

In Fig. 7 each party label corresponds to one politician. The president of the party is visualized by larger font. Numbers in contour lines describe the density of politicians in each party. Members of parliament whose values are approximated (labeled in parenthesis) are close to the maximum point.

Areal information about value distribution can be achieved by limiting voting application data by electoral districts. The results for each candidate by cities or even the polling station are available. Citizens' values can be visualized on the map for example. Difference of value distributions between some polling stations, such as different parts of a town, can be used to explain some other measures, such as housing prices, criminal statistics, unemployment, etc.

Preprosessing and analysis methods for voting advice application data were presented as well as visualizations of value distribution of the political field that seemed to be relatively insensitive small changes in the data. More detailed analysis of this application is in [23]. The analysis opens up new possibilities on this field. New questions are again waiting for answers.

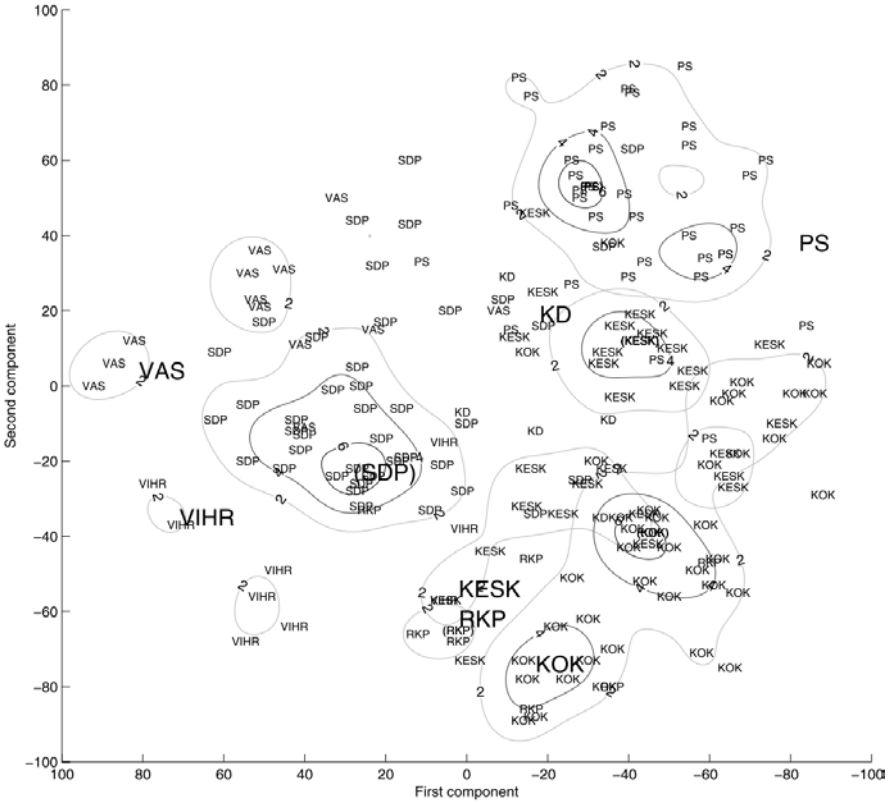

Fig. 7. Members of parliament values
Rys. 7. „Wartości" poszczególnych członków parlamentu

5. Conclusion

We have drawn an interdisciplinary figure of the scientific domain with somewhat subjective emphasis. In other words we have described how interdisciplinary relationships contribute to the scientific domain even when the emphasis is partly subjective. Our model is

based on three main pillars: psychology, civics and energy. From the main pillars there leads paths to more applied branches such as user interfaces, visualizations, politics, production models and finally to decision making. The two main supporting methodologies in the view are system technics and data-analysis.

From the applied branches we have concentrated more on visualizations and political decision making. A detailed example of political decision making with data-analysis methods is described. We have compared our perspective to literature with similarities, and reflected some importance to the basis built and described. On the background there are also several our studies from previous years.

The same topic is studied in [19] including less technical details in the visualization examples and in the example analysis of political decision making. The analysis in this article is wider and it is technically more detailed and accurate.

We have shown that interdisciplinary scientific view can also produce new topics of interest to study further later on. In a more and more focused scientific world a larger perspective needs sometimes to be drawn to better identify and formulate the entity as a whole.

To emphasize and outline synergetic effect of interdisciplinary research we have derived examples to show how different methodologies and different application areas meet with each other and are able to produce results that a narrower view hardly ever can. Visualizing the effects also outline the reader the ideas and emphasize the new reality achieved.

A distinction should be made between exact and non-exact research fields. For instance social sciences are less exact than research where the measures are easier to define. Also the interdisciplinary approach adds often less exact features to the whole figure.

From man-machine interface point of view user modelling is a new field with growing interest. Better measurement equipment also enables better user modelling. How a human interpret different advanced visualizations is another important branch to go forward.

Our background experience concentrate on industrial energy production on application view, and on the methodological side knowledge-based and neural methods in decision support. In future we plan to concentrate also on completely new application areas even not mentioned here in our broad analysis.

Bibliography

1. Bastian M., Heymann S., Jacomy M.: Gephi: an open source software for exploring and manipulating networks. Proceedings of the International AAAI Conference on Weblogs and Social Media, 2009.
2. Bunn D.: Applied Decision Analysis. McGraw-Hill, Inc., USA 1984.

3. Eylon B-S., Linn M.C.: Learning and Instruction: an Examination of Four Research Perspectives in Science Education. „Review of Educational Research", Vol. 58, No. 3, 1988, p. 251-301.
4. French S.: Decision Theory: an Introduction to the Mathematics of Rationality. Ellis Horwood Limited, England 1986.
5. Gilbert J.K., Watts D.M.: Concepts, Misconceptions and Alternative Conceptions: Changing Perspectives in Science Education. „Studies in Science Education", Vol. 10, No. 1, 1983, p. 61-98.
6. Glad T., Ljung L.: Control Technics – Basic Theory, Studentlitteratur. Lund, Sweden 1981, 200.
7. Gobern W.W. (ed.): Socio-Cultural Perspectives on Science Education: an International Dialogue. Kluwer Academic Publisher, Norwell, USA 1998.
8. Hair J., Anderson R., Tatham R., Black W.: Multivariate Data Analysis. Prentice Hall, 5th edition, 1998.
9. Heimbürger H. et al.: Control Room – Design Principles and Practices. Suomen Automaatioseura ry, Helsinki 2010.
10. Hirzalla F., Van Zoonen L., De Ridder J.: Internet Use and Political Participation: Reflections on the Mobilization/ Normalization Controversy. „The Information Society", Vol. 27, No. 1, 2010, p. 1-15.
11. Jacomy M., Heymann S., Venturini T., Bastian M.: ForceAtlas2, a graph layout algorithm for handy network visualization, http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf.
12. Kohonen T.: The Self-Orgainizing Map. Springer, 1995.
13. Koskinen H., Norros L.: Expanding Control Room – a New Frame for Designing Spatial Affordances of Control Places. VTT-R-05555-10, 2010.
14. Lemke J.L.: Articulating Communities: Sociocultural Perspectives on Science Education. „Journal of Research in Science Teaching", Vol. 38, No. 3, 2001, p. 296-316.
15. Paulsen J.: Design Process Displays Based on Risk Analysis Techniques. PhD thesis, Technical University of Denmark and Risø National Laboratory, Roskilde 2004, p. 143.
16. Pershagen B.: Light Water Reactor Safety. Pergamon Press, Studsvik, Sweden 1989.
17. Scott D.W.: Data Preparation for Data Mining. Morgan Kaufmann, 1999.
18. Sirola M., Talonen J., Parviainen J., Lampi G.: Decision Support with Data-Analysis Methods in a Nuclear Power Plant. TKK Reports in Information and Computer Science, TKK-ICS-R29, Espoo 2010.
19. Sirola M.: Perpectives in Science. Proceedings of the IEEE International Conference on Intelligent Data Acquisition and Advanced Coputing Systems: Technology and Applications, Berlin, Germany, 12-14 September 2013, p. 120-123.
20. Talonen J., Sirola M., Sulkava M.: Data fusion of pre-election gallups and polls for improved support estimates in time. Unpublished.

21. Talonen J., Sirola M., Sulkava M.: Network visulaization of car inspection data using graph layout. Proceedings of the International Conference on Data Analytics, Barcelona, Spain 2012.
22. Talonen J., Sulkava M., Sirola M.: The Finnish car rejection reasons shown in an interactive SOM visualization tool. Proceedings of the Workshop on Self-Orgaizing Maps (WSOM 2012), Santiago, Chile 2012.
23. Talonen J., Sulkava M.: Analyzing Parliamentary Elections Based on Voting Advice Application Data. Proceedings of the International Symposium on Intelligent Data Analysis (IDA 2011), Porto, Portugal 2011.
24. Vesanto J., Himberg J., Alhoniemi E., Parhankangas J.: Self-organizing map in matlab: the SOM toolbox. Proceedings of the Matlab DSP Conference, 1999.
25. Wheatley G. H.: Constructivist Perspectives on Science and Mathematics Learning. „Science Education", Vol. 75, No. 1, 1991, p. 9-21.
26. Xiaoyuan S., M. Taghi K.: Survey of Collaborative Filtering Techniques. „Advances in Artificial Intelligence", Hindawi Publishing Corporation, 2009, Article ID 421425.
27. A-Katsastus Group´s web site, http://www.a-katsastus.com/, retrieved on June 2012.
28. TÜV reports, http://www.anusedcar.com/, retrieved on June 2012.

Omówienie

W artykule omówiono perspektywy zastosowania różnych podejść w nauce w ramach analizy danych i ich wizualizacji. Dziedziny nauki zawarte w artykule odzwierciedlają zainteresowania i doświadczenia autora w obszarze interdyscyplinarnego podejścia do zagadnienia. Trzy główne filary przedstawionego modelu to: psychologia, wychowanie obywatelskie i produkcja energii. W publikacji zaprezentowano wykorzystanie analizy danych w podejmowaniu decyzji politycznych. Rezultaty przedstawiono na przykładzie zachowań wyborców. Wykorzystano w tym celu wyniki sondażu Gallupa i na tej podstawie opracowano aplikację, która może pozwolić na prognozowanie wyników głosowania.