INFORMATICA, 2001, Vol. 12, No. 2, 239–262

Visualisation of Multidimensional Objects and the Socio-Economical Impact to Activity in EC RTD Databases

Gintautas DZEMYDA, Vytautas TIEŠIS
Institute of Mathematics and Informatics
Akademijos 4, 2600 Vilnius, Lithuania
e-mail: dzemyda@ktl.mii.lt, tiesis@ktl.mii.lt

Received: Spetember 2000

Abstract. The paper deals with the analysis of Research and Technology Development (RTD) in the Central European countries and the relation of RTD with economic and social parameters of countries in this region. A methodology has been developed for quantitative and qualitative ranking and estimates of relationship among multidimensional objects on the base of such analysis. The knowledge has been discovered in four databases: two databases of European Commission (EC) containing data on the RTD activities, databases of USA CIA and The World bank containing economic and social data. Data mining has been performed by means of visual cluster analysis (using the non-linear Sammon's mapping and Kohonen's artificial neural network – the self-organising map), regression analysis and non-linear ranking (using graphs of domination). The results on clustering of the Central European countries and on the relations among RTD parameters with economic and social parameters are obtained. In addition, the data served for testing various features of realisation of the self-organising map. The integration of non-classical methods (the self-organising map and graphs of domination) with classical ones (regress analysis and Sammon' mapping) increases the capacity of visual analysis and allows making more complete conclusions.

Key words: research and technology development, knowledge discovery, data mining, regression analysis, neural network, self-organising map, non-linear ranking, graphs of domination, visual analysis, cluster analysis.

1. Introduction

Science and technologies are global for a long time. The globalisation accelerates the technological progress. Recently, Internet and data published on it about Research and Technology Development (RTD) projects and results are the significant stimulator of the globalisation. The main objective of the paper is to develop a methodology for quantitative and qualitative ranking and estimates of relationship among multidimensional objects on the base of analysis on the RTD activities in the Central European (CE) countries and the economical and social characteristics of these countries. This objective is led by the following tasks:

- to get evaluations of CE countries with respect to RTD as well as to socio-economic development;

- to get evaluations of relations between socio-economic conditions and RTD;
- to propose and investigate the methodology for such evaluations and their visualisation;
- to examine data analysis methods by using RTD and socio-economical data.

The analysis has been carried out on the activities of the Central European countries in the EC RTD databases *CORDIS RTD-Projects* and *CORDIS RTD-Results* (Cordis, 1999) and the relation of RTD with economic and social parameters of these countries. The research covers 10 Central European countries. Their list is presented in Table 1. The necessary economic and social data are extracted from the database of the USA CIA *The World Factbook* (1999) and the database of the World Bank *Country Data* (1999). We try to extract the implicit, unknown and potentially useful information from the databases above. Such an action is defined by Fayyad (1997) as knowledge discovery. Data mining has been performed in this research by means of

- regression analysis,
- non-linear ranking using graphs of domination,
- visual cluster analysis using the classic non-linear mapping and artificial neural networks.

The integrated application of the means above makes a basis of the methodology that is presented in Sections 2 and 3 (and partially – Section 4.1, where the scaling is performed). It covers

- integration of a proper set of methods that are oriented to the visual analysis of multidimensional objects.
- selection of a set of parameters that represent the RTD activities and socio-economical state of the countries.

The research was stimulated by the 4-th IEEE Baltic Workshop on Databases and Information Systems (Čaplinskas, 2000). This paper extends the results of Dzemyda and Tiešis (2000) by new methods of analysis and by more generalised view to the problem. In addition, various features of realisation of the self-organising map were examined using the data extracted from databases.

2. Data Sources

A unique initiative for strengthening the competitiveness of European organisations, *CORDIS*, the Community Research and Development Information Service, represents the RTD starting point. As an integral part of the European Commission's INNOVATION Programme, *CORDIS* provides information on a vast range of research, development and innovation activities undertaken on a European level. We use two parameters that characterise the activities of each country in *CORDIS* (1999) databases (here we assume that these parameters represent the RTD activities of the countries):

- x_1 is a number of projects carried out by institutions of the country and put in the database *CORDIS RTD-Projects* per million of population of the country;

Table 1

The data extracted from databases

Country	Label	Records in *CORDIS RTD-Projects* per million of population (x_1)	Records in *CORDIS RTD-Results* per million of population (x_2)	(-1)*infant mortality rate (x_3)	GDP per capita in USD (x_4)	GDP in industry and services (%) (x_5)	Exports per capita in thousands of USD (x_6)	Telephones per capita (x_7)	International economic aid per capita in USD (x_8)
Hungary	1	71.8627	95.0980	-9.460	7400	97.00	2.0294	0.2118	12.03
Czech Republic	2	60.6796	30.5825	-6.670	11300	95.00	2.3107	0.3252	34.14
Lithuania	3	31.9444	4.1667	-14.710	4900	87.00	1.1667	0.3000	63.47
Latvia	4	42.0833	3.3333	-17.190	4100	93.00	0.7917	0.2958	40.08
Slovakia	5	50.9259	2.4074	-9.480	8300	95.20	1.9815	0.2519	78.13
Poland	6	16.5803	1.2953	-12.760	6800	94.90	0.7047	0.2124	111.71
Romania	7	15.8744	0.7623	-18.120	4050	81.00	0.3677	0.1166	22.87
Estonia	8	107.1429	0.0000	-13.830	5500	93.80	1.8571	0.3786	98.07
Bulgaria	9	37.5610	5.1220	-12.370	4100	74.00	0.5488	0.3378	15.73
Slovenia	10	122.5000	10.0000	-5.280	10300	95.00	4.6000	0.3450	2.50
The average values of parameters	AVE	55.7155	15.2768	-11.987	6675	90.59	1.6358	0.2775	–
The worst values of parameters	MIN	15.8744	0.0000	-18.120	4050	74.00	0.3677	0.1166	–
The best values of parameters	MAX	122.5000	95.0980	-5.280	11300	97.00	4.6000	0.3786	–

- x_2 is a number of records by institutions of the country in the database *CORDIS RTD-Results* per million of population of the country.

The database *CORDIS RTD-Results* is oriented to the users of RTD results (customers, partners, etc.). The data may be put into the database without any restrictions by investigators from any country having some RTD results and seeking some form of active collaboration for the exploitation of their results. However, the query form in this database has no field of a record supplier country. Therefore, seeking to find the values of x_1 and to avoid other referring to the country, we had to use the contextual search by the name and phone code of the country. The database *CORDIS RTD-Projects* contains information about all the EC projects. Data from *CORDIS* databases are extracted at 1999.11.24.

Data on the economic and social conditions of Central European countries are taken from the USA CIA *The World Factbook* (1999) database, which is presented on the Internet as a well-structured text. Therefore, a special filter has been developed to extract the necessary data. The wide used essential economic and social parameters are selected:

- x_3 is the infant mortality rate (deaths / 1000 live births); the parameter is used instead of the life expectancy at birth as the later accumulates the life quality in the very long period;
- x_4 is the Gross Domestic Product (GDP) per capita in US dollars obtained taking into account the purchasing power parity of the national currency but not the exchange rate;
- x_5 is the percentage of GDP developed in the industry and services (not in the agriculture);
- x_6 is the export per capita in thousands of US dollars;
- x_7 is the number of telephones per capita; it may influence on the activity in Internet.

These parameters may influence RTD, determine the potential of country for a longer period, and vary essentially through the countries of different level of development. Moreover, the data about the international aid in US dollars per capita (x_8) were used in the investigation. *The World Factbook* (1999) has no this data for Bulgaria. Therefore, this data was taken from the World Bank *Country Data* (1999) database.

The scheme of data extraction for the analysis is presented in Fig. 1. The data extracted from four databases are presented in Table 1. Each country is described by two parameters of activity in the *CORDIS* databases (parameters x_1 and x_2), five economic and social parameters ($x_3, \ldots, x_7$), and one parameter of the international aid (x_8). We use negative values of x_3 in Table 1 and in the further analysis because it is more convenient that the greater values of parameters $x_1, \ldots, x_7$ be the better. This rule was not applied to the last parameter x_8 because it is impossible to give an exact answer what extent of the international aid is good, and what is bad.

Denote the data matrix obtained via analysis of the four databases by $X = \{x_{ij},\ i = \overline{1,10},\ j = \overline{1,8}\}$. Here index i is the number of analysed object (country), and index j is the number of parameter. AVE, MIN and MAX indicate the objects derived from the data on 10 countries. Denote all the objects by $X_1, X_2, \ldots, X_{10}, X_{\mathrm{AVE}}, X_{\mathrm{MIN}}, X_{\mathrm{MAX}}$. The values of parameters characterising object X_{AVE} are obtained by averaging values of

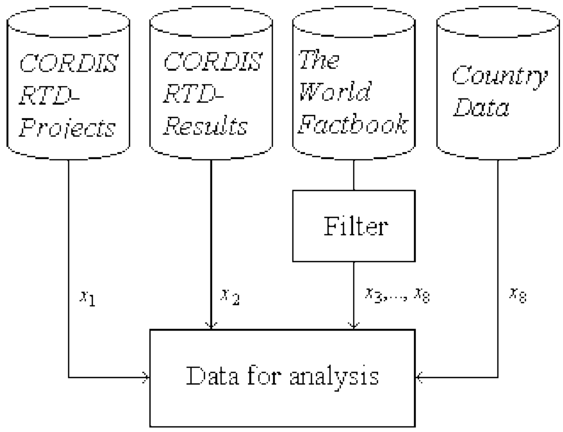

Fig. 1. The scheme of data extraction.

respective parameters over all 10 countries: $x_{\mathrm{AVE}j} = \frac{1}{10}\sum_{i=1}^{10} x_{ij}$, $j = \overline{1,7}$. The values of parameters characterising object X_{MIN} are the worst values of respective parameters over all 10 countries: $x_{\mathrm{MIN}j} = \min_{i=\overline{1,10}} x_{ij}$, $j = \overline{1,7}$. The values of parameters characterising object X_{MAX} are the best values of respective parameters over all 10 countries: $x_{\mathrm{MAX}j} = \max_{i=\overline{1,10}} x_{ij}$, $j = \overline{1,7}$. Such an introducing of three artificial objects X_{AVE}, X_{MIN} and X_{MAX} is useful for our analysis because of possibility to have basic points describing the imaginary average, worst and the best countries.

3. Visualisation as a Tool for Analysis

In its simplest form, visualisation is concerned with the visual representation of data. As the volume of data increases, the visual representation of data quickly becomes the most practical method of understanding the data. Over the past few years, the role of visualisation in aiding the analysis of data has increased dramatically. From its initial role as simply a tool for presenting results, it has developed into a valuable tool for understanding data.

The proposed methodology for visual analysis of RTD and socio-economical data integrates five methods for visualisation of n-dimensional data:

- Sammon's mapping,
- the self-organising map (SOM),
- combination of the methods above,
- graphs of domination,
- regress analysis.

The first method seeks to keep the Euclidean metrics of the resulting reduced dimensionality data. This makes the projected data easier understandable by human. Results of application of the self-organising map have a partial metrics character, only. The first three methods are oriented to the reducing of dimensionality of data – for mapping a high dimensional space onto a low dimensional one and for further visual presentation to the investigator. Graphs of domination give a non-metric data presentation. The visual presentation of results of regress analysis (regress curve together with the source data points) serves as a source for additional knowledge for investigator, too. Therefore, the integration of the set of methods of different nature would facilitate the knowledge discovery in the given data set and the evaluation of the data set from different standpoints.

We describe the methods below in brief. Novelties are in the realisation of SOM and graphs of domination. However, we present all the methods seeking to make the reading of the paper easier.

Denote objects in Sections 3.1–3.3 by Y_i, $i = \overline{1, s}$, because either the set of countries $X_1, X_2, \dots, X_{10}, X_{\mathrm{AVE}}, X_{\mathrm{MIN}}, X_{\mathrm{MAX}}$ or the set of parameters $x_1, \dots, x_7$ are assumed as objects in this paper when the analysis by Sammon's mapping and SOM is carried out. Here s is the number of objects, and n is their dimensionality. Therefore, values of s and n will depend on the analysed data set (countries or parameters).

3.1. *Sammon's Mapping*

There exist a lot of methods that can be used for reducing the dimensionality of data. The analysis by Bezdek and Pal (1995) and Flexer (1997) of relative performance of different algorithms in reducing the dimensionality of multidimensional objects (vectors) indicates Sammon's (1969) projection to be still one of the best methods of this class. Sammon's projection is a non-linear projection method of mapping a high dimensional space onto a space of lower dimensionality.

Denote by

- d^*_{ij} the distance between objects Y_i and Y_j in the feature space which dimensionality is n;
- d_{ij} the distance between the same objects Y_i and Y_j in the projected space (in our case the dimensionality of projected space is 2).

Sammon's algorithm tries to minimize the distortion of projection:

$$E = \frac{1}{\sum\limits_{\substack{i,j=1\\ i<j}}^{s} d^*_{ij}} \sum_{\substack{i,j=1\\ i<j}}^{s} \frac{(d^*_{ij} - d_{ij})^2}{d^*_{ij}}.$$

In fact, Sammon's mapping is closely related to the metric multidimensional scaling (MDS) (see Doctor's Thesis of Kaski (1997) for details on the metric and non-metric MDS). It, too, tries to optimise a cost function that describes how well the pairwise distances in a data set are preserved. The only difference between Sammon's mapping and the metric MDS is that the errors in distance preservation are normalised with the distance in the original space (see the distortion of projection E above). For the experiments below we modified a realisation of Sammon's mapping from Murtagh's (2000) Internet site.

However, the realisation from, e.g., SOM_PAK by Kohonen *et al.* (1996) may be used, too. It is noted by Kohonen (2001) that Sammon's projection is always recommended for a preliminary test of the data to be used for self-organising maps.

3.2. *The Self-Organising Map*

The self-organising map (SOM) (see, e.g., Fausett, 1994; Kohonen, 1989, 2001; Ritter, Schulten, and Martinetz, 1991) is a well-known method for mapping a high dimensional space onto a low dimensional one. We consider here a mapping onto a two-dimensional grid of neurons. Let $Y_1, \ldots, Y_s \in R^n$ be a set of n-dimensional vectors (objects) for mapping.

Usually, the neurons are connected to each other via rectangular or hexagonal topology. In this paper, we consider the rectangular case, only (see Fig. 2 for example of SOM of size 4×4: circles denote neurons; indices of neurons are given inside the circles). The rectangular SOM is a two-dimensional array of neurons $M = \{m_{ij},\ i = \overline{1, k_x},\ j = \overline{1, k_y}\}$. Here k_x is the number of rows, and k_y is the number of columns (in Fig. 2, both k_x and k_y are equal to 4). The total number of neurons is equal to $k_x \times k_y$. All neurons adjacent to a given neuron can be defined as its neighbours of a first order, then the neurons adjacent to a first order neighbour, excluding those already considered, as neighbours of a second order, etc. For example, the first order neighbours of m_{23} are m_{12}, m_{13}, m_{14}, m_{22}, m_{24}, m_{32}, m_{33}, m_{34}; the remaining neurons are the second order neighbours. The dimension of the vectors, which will be presented as inputs to train the network, is n. Each component of the input vector is connected to every individual neuron. Thus, there is a connection between the neuron of the network and every component of the input vector. The weights of these connections form an $n-$dimensional synaptic weight vector (the codebook vector, also called reference, model, or parameter vector – see Kohonen (2001)). Thus, any neuron is entirely defined by its location on the grid (number of row i and column j) and by the codebook vector, i.e., we can consider a neuron as an n-dimensional vector $m_{ij} = (m_{ij}^1, m_{ij}^2, \ldots, m_{ij}^n) \in R^n$. In this way, each vector (neuron) m_{ij} represents a part of R^n because $Y_1, \ldots, Y_s \in R^n$.

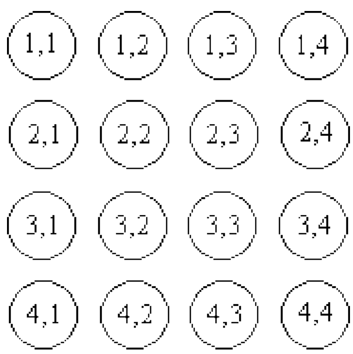

Fig. 2. The rectangular SOM.

The map is trained in an unsupervised manner using competitive learning. Learning starts from the vectors m_{ij} initialised randomly (other possible ways to initialise the vectors m_{ij} are given, e.g., by Fausett (1994)). At each learning step, an input vector Y is drawn from the training set $\{Y_1, \dots, Y_s\}$ and passed to the neural network. The Euclidean distance from this input vector to each vector m_{ij} is calculated and the vector (neuron) $m_c \in \{m_{ij},\ i = \overline{1, k_x},\ j = \overline{1, k_y}\}$ with the minimal Euclidean distance to Y is designated as a winner. Denote the row, where m_c is located, by i_c, and the column by j_c, i.e., c is a combination of two numbers – i_c and j_c. The components of the vector m_{ij} are adapted according to the rule

$$m_{ij} \leftarrow m_{ij} + h_{ij}^c (Y - m_{ij}), \tag{1}$$

where h_{ij}^c is the learning rate, which is maximal for the winning neuron, and decreases with the neighbourhood order and the learning steps.

After a large number of learning steps, the network has been organised and n-dimensional input vectors have been mapped – each input vector is related to the nearest neuron (vector) m_{ij}, i.e., the objects are distributed among the elements of the map during training. Some elements of the map may remain unrelated with any object, but there may occur elements related with some objects.

Using the SOM-based approach above we can draw a table with cells corresponding to the neurons. The cells corresponding to the neurons-winners are filled with the numbers of vectors $Y_1, \dots, Y_s$. Some cells may remain empty. One can decide visually on the distribution of vectors $Y_1, \dots, Y_s$ in the $n-$dimensional space R^n in accordance to their distribution among the cells of the table. The essential property of such visualisation is clustering of the objects in accordance to their similarities.

A variety of realisations of SOM have been developed (see, e.g., Kohonen, 2001; SOM_PAK by Kohonen *et al.*, 1996; Murtagh, 2000; Murtagh and Hernindez-Pajares, 1995; Zell *et al.*, 1995; Kleiweg, 1996). All of them produce different results to some extent. Therefore, we present below additional details on our realisation of SOM used in the experiments. Some additional details on the realisation and application of SOM used in our research are given in the paper by Dzemyda (2001) devoted to the visual analysis and clustering of parameters.

Let us introduce a term "learning iteration". The learning iteration consists of s learning steps: the input vectors from Y_1 to Y_s are passed to the neural network in consecutive or random order. The random order eliminates some influence of numeration of the input vectors on the learning process. The whole learning process consists of v iterations. Both consecutive and random orders and $v = 200$ were used in the experiments. In our case, such a partition of the learning process into learning iterations is sensible because of a small number s of input vectors $Y_1, \dots, Y_s$.

$$h_{ij}^c = \frac{\alpha}{\alpha \eta_{ij}^c + 1}, \quad \alpha = \max\left(\frac{v + 1 - e}{v}, 0.01\right),$$

where η_{ij}^c is the neighbourhood order between the neurons m_c and m_{ij}, e is the number of current iteration ($e \in [1, v]$).

The set of possible neighbours of m_c in (1) is restricted: we recalculate the vector m_{ij} if

$$\eta_{ij}^c \leqslant \max\left[\alpha \max(k_x, k_y), 1\right].$$

Note that $0 < h_{ij}^c \leqslant 1$, $h_{ij}^c = 1$ in the first learning iteration when $i = i_c$ and $j = j_c$, only.

In the paper by Dzemyda (2001), where correlation matrices were analysed, the vectors m_{ij} were initialised by random vectors of unit length, i.e., initially $\|m_{ij}\| = 1$, $i = \overline{1, k_x}$, $j = \overline{1, k_y}$. In addition the cosines between any pair of the initial vectors m_{ij} were set to be positive (this has been achieved by generating vectors of unit length that contain only positive components). In this paper, we examine two cases of the initialisation: with restriction on the positiveness of the components and without it.

3.3. *Combination of the Self-Organising Map and Sammon's Mapping (Combined Mapping)*

The table, filled in accordance to the distribution of objects $Y_1, \ldots, Y_s$ among cells of the SOM, doesn't answer the question, how much the objects of the neighbouring cells are close in the n-dimensional space. Following Sammon (1969) who suggested that clustering could be used as a front-end to his mapping algorithm, Kaski (1997) made an assumption (without its theoretical or experimental background) that an especially useful combination seems to be first to reduce the amount of data by SOM, and then to analyse the codebook vectors corresponding to non-empty cells by using Sammon's mapping, i.e., to visualise the relative distances between these codebook vectors of the SOM (Kohonen (2001) notes too, that Sammon's projection is a useful tool in preliminary data analysis as well as for the monitoring of the codebook vectors). In this case the neural network performs some sorting (clustering) of data, and Sammon's algorithm presents the results visually to gain an additional insight. Such combination of mapping algorithms is examined and grounded experimentally by Dzemyda (2001) in solving the problem of clustering of parameters characterised by their correlation matrix. SOM_PAK software by Kohonen *et al.* (1996) allows such the combination, too.

3.4. *Regress Analysis*

The application of the first three methods is a proper mean to observe the general two-dimensional view on the distribution of objects. However, the information is compressed and therefore distorted by the visualisation methods. So, the regression analysis is used for more detailed investigation of relationships among parameters. Because of a small number of observations, in this paper we consider case of relations between the couples of parameters, only. If we have s observations of parameters x_k and x_j, they may be interpreted as two random variables, and the table (x_{ik}, x_{ij}), $i = \overline{1, s}$ may be interpreted as a random sample. The usual method to approximate the regression (the conditional expectation) $x_k = \bar{y}(x_j) = M\{x_k|x_j\}$ is to fit a representative of some family of curves

y to the sample using the least squares method. One can try to choose a family of the curves y whose optimal representative visually approximates the sample in the best way. Usually the best regression produced the best coefficient of determination R^2. $R^2 = 1-\frac{\sigma_E^2}{\sigma_T^2}$, where $\sigma_E^2 = \frac{1}{s}\sum_{i=1}^{s}(x_{ik} - y(x_{ij}))^2$ is the estimation of variation of approximating curve $y(x_j)$ error (the unexplained variation); $\sigma_T^2 = \frac{1}{s}\sum_{i=1}^{s}(x_{ik} - m_k)^2$ is the estimation of variation of parameter x_k; $m_k = \frac{1}{s}\sum_{i=1}^{s} x_{ik}$ is the estimation of average of x_k.

3.5. *Graphs of Domination*

Here we suggest the ranking of multiparametric (n-dimensional) objects by the relation of domination. The object $Y_l = (y_{l1}, \dots, y_{ln})$ is dominated by Y_k if $\forall j\ y_{kj} \geqslant y_{lj}$. Denote such relation by $Y_k \succ Y_l$. If the majority of objects may be compared by this relation then the corresponding direct graph that visualise the non-linear ranking of objects may be drawn. The arc points from an object Y to successive object Y' if $Y \prec Y'$ and there is no object Y'' such that $Y \prec Y'' \prec Y'$.

In addition, we introduce the levels (nonnegative numbers) of an object as its quantitative characteristics. The objects at the bottom of graph are of zero level. Several consequent levels (interval of levels $[a, b]$) may be allocated to one object. The least level a of an object equals to the length (number of arcs) of the longest direct paths from the bottom of a graph to the object. So, the least level of an object is a number of dominated objects in a succession of domination. The maximal level of an object $b = b_{\max} - d$ where d equals to the length of the longest direct paths from the object to the top of a graph, the number of top level $b_{\max}$ equals to the length of the longest direct paths in a graph. So, the maximal level of an object differs from $b_{\max}$ in a number of predominant objects in a succession of domination. If $b > a + 1$, the intermediate levels between a and b are also allocated to this object. The objects of the same level are of equal value from a viewpoint of the relation of domination.

The matrix of domination for s objects D is formed directly by the use of domination relation: $D = \|d_{ij},\ i,j = 1,\dots,s\|$ where $d_{ij} = 1$, if $Y_i \succ Y_j$, $d_{ij} = 0$ otherwise. Let us describe the algorithm that generates a graph of domination $G(N,U)$ where $N = \{1,\dots,s\}$ is a set of nodes corresponding to a set of objects $\{Y_1,\dots,Y_s\}$ and $U = \{(i,j),\ i,j \in N\}$ is a set of edges. The algorithm may be divided into 3 stages.

Stage 1. The allocation of least levels to the objects.

Step 1.1. A current level $l = 0$, $D' = D$, matrix' dimension $s' = s$.

Step 1.2. As the relation of domination is transitive some (say k) i_l exist such that $\sum_{j}^{s'} d'_{i_l j} = 0$. The level l is allocated to the nodes i_l, and all i_l-th rows and columns are eliminated from D'. Update $l = l + 1$ and $s' = s' - k$. The Step 1.2 is repeated while the condition $s' > 0$ is true.

Stage 2. The allocation of maximal levels to the objects.

The calculation follows the Stage's 1 steps with transposed D and considering levels down from the top of the graph.

Step 2.1. The top level $l = l - 1$, $D = D^T$, $s' = s$.

Step 2.2 repeats Step 1.2 with exception that l is decreasing: $l = l - 1$.

Stage 3. Forming the set U.

Step 3.1. A number of current node $i = 1$.

Step 3.2. Consider j where $d_{ji} = 1$. If there is no such j, then i is one of top nodes. The edge $(i, j) \in U$ if $d_{ki}d_{jk} = 0$ for all $k \neq j, i$. Repeat the step with $i = i + 1$ while $i \leqslant s$.

The algorithm automatically produces the graph $G(N, U)$ that allows to rank multidimensional objects and to evaluate the levels of objects. As some objects may be incomparable the level of the object may be determine ambiguous.

4. Visual Analysis by Using Mapping and SOM

Here we apply Sammon's mapping and the self-organising map to the analysis both of 10 Central European countries and of parameters characterising the activities in *CORDIS* databases and economic and social parameters.

In addition, in this paper we examine the realisation of SOM presented in Section 3.2 on the basis of the socio-economic data from Table 1. The investigation covers:

1. Two cases of the initialisation:
   - with restriction on the positiveness of the components (*Posit*),
   - without the restriction (*Any*).
2. Two types of orders to pass the objects (input vectors) to the neural network during its training:
   - consecutive (*Consec*),
   - random (*Random*).

These additional investigations were oriented to evaluate the efficiency of four modifications of the realisation of SOM: (*Posit*, *Consec*), (*Any*, *Consec*), (*Posit*, *Random*), (*Any*, *Random*).

4.1. *Results of Analysis of 10 Central European Countries*

Parameters $x_3, \ldots, x_7$ characterise 13 objects $X_1, X_2, \ldots, X_{10}, X_{\mathrm{AVE}}, X_{\mathrm{MIN}}, X_{\mathrm{MAX}}$ in various economic and social aspects. Their units of measurement are different, and their nominal values differ in some orders. Therefore, it is necessary to unify the scales of parameters before the analysis. This has been done for each parameter x_j as follows:

- the mean value $\overline{x_j}$ and variance σ_j^2 was computed on the basis of 13 values of the parameter,

- each value x_{ij} of the parameter was transformed in accordance with the formula $(x_{ij} - \overline{x_j})/\sigma_j$.

The results of analysis of the objects $X_1, X_2, \ldots, X_{10}, X_{\mathrm{AVE}}, X_{\mathrm{MIN}}, X_{\mathrm{MAX}}$ by using neural networks from the view of economic and social parameters $x_3, \ldots, x_7$ are presented in Tables 2–7. SOM was trained using five components $x_3, \ldots, x_7$ (economic and social parameters) of $v = 13$ objects $X_1, X_2, \ldots, X_{10}, X_{\mathrm{AVE}}, X_{\mathrm{MIN}}, X_{\mathrm{MAX}}$, i.e., $n = 5$. There were used three variations of SOM: 3×3-dimensional, 4×4-dimensional, and 5×5-dimensional. The corresponding Sammon's projections of results of the neural-analysis are presented in Figs. 3–8. Labels $1, \ldots, 10$, AVE, MIN and MAX in the tables and figures indicate the objects.

The following modifications of realisation of SOM were used: (*Posit*, *Consec*), (*Any*, *Consec*), (*Posit*, *Random*), (*Any*, *Random*). All the four modifications were examined using 4×4-dimensional SOM. The results are presented in Tables 2–5 and Figs. 3–6. In addition, the realisation (*Any*, *Random*) of 3×3-dimensional and 5×5-dimensional SOM was applied to the data. The results are presented in Table 6 and Fig. 7 for 3×3-dimensional SOM and in Table 7 and Fig. 8 for 5×5-dimensional SOM.

Comments and conclusions on the relative efficiency of different realisations of SOM are presented in Section 4.3, and these on countries in general are presented in Section 4.4.

The distribution of objects from the view of parameters x_1 and x_2 characterising the activities of each country in *CORDIS* databases is presented in Fig. 9. Fig. 10 shows the pure Sammon's projection from the view of economic and social parameters $x_3, \ldots, x_7$. See comments in Section 4.4.

Table 2
Analysis of countries:
Distribution on SOM 4×4, $Pos + Consec$

MAX, 10			MIN, 7
2			9
5			3
1	6	AVE	4, 8

Table 3
Analysis of countries:
Distribution on SOM 4×4, $Any + Consec$

4	9		MIN, 7
8	3		6
		AVE	
MAX, 10	2	5	1

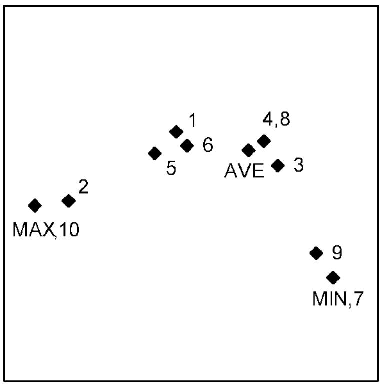

Fig. 3. Analysis of countries: Combined mapping (SOM 4×4, $Pos + Consec$).

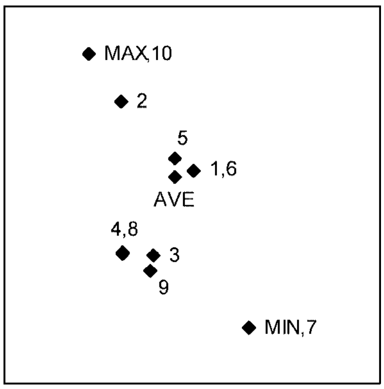

Fig. 4. Analysis of countries: Combined mapping (SOM 4×4, $Any + Consec$).

Table 4
Analysis of countries:
Distribution on SOM 4×4, $Pos + Random$

1	6		MIN, 7
5	AVE		9
2		3	
MAX,10		8	4

Table 5
Analysis of countries:
Distribution on SOM 4×4, $Any + Random$

MIN, 7		6	1
9		AVE	5
	3		2
4	8		MAX,10

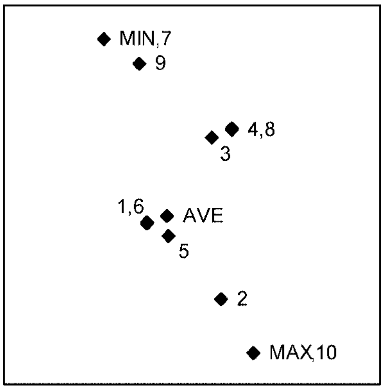

Fig. 5. Analysis of countries: Combined mapping (SOM 4×4, $Pos + Random$).

Fig. 6. Analysis of countries: Combined mapping (SOM 4×4, $Any + Random$).

Table 6
Analysis of countries:
Distribution on SOM 3×3, $Any + Random$

2, 10, MAX		1, 5, 6
	AVE	
7, MIN	9	3, 4, 8

Table 7
Analysis of countries:
Distribution on SOM 5×5, $Any + Random$

1		6		MIN,7
	5			
2		AVE		9
			3	
MAX,10		8		4

4.2. *Results of Analysis of Parameters*

The methodology of experiments is similar to that above. The goal of analysis here was to determine the relationship of the parameters x_1 and x_2 characterising the activities in *CORDIS* databases with economic and social parameters $x_3, \ldots, x_7$. The training of SOM has been performed on a basis of $v = 7$ objects, because there were analysed 7 parameters $x_1, \ldots, x_7$. $n = 10$, because the analysis of parameters has been performed from the view-point of 10 countries $X_1, \ldots, X_{10}$. Data for the analysis are presented in 1–10 rows of Table 1, columns denoted by $x_1, \ldots, x_7$. Like in Section 4.1, scales of parameters are unified before the analysis. The unification has been applied to all the 7 parameters $x_1, \ldots, x_7$ on the basis of 10 their values.

The results of analysis of the parameters $x_1, \ldots, x_7$ by using neural networks are presented in Tables 8–13. There were used three variations of SOM: 2×2-dimensional,

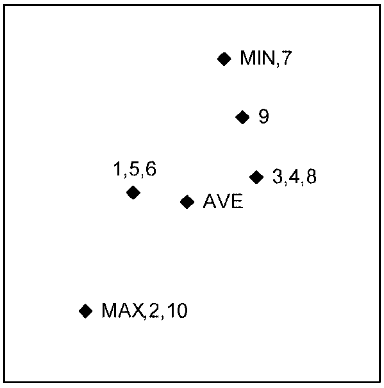

Fig. 7. Analysis of countries: Combined mapping (SOM 3×3, $Any + Random$).

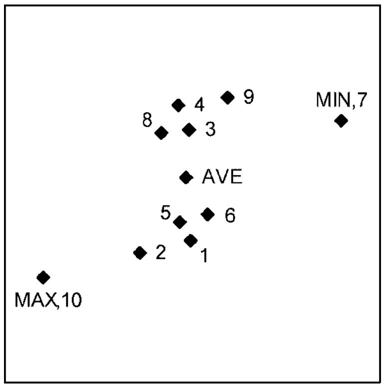

Fig. 8. Analysis of countries: Combined mapping (SOM 5×5, $Any + Random$).

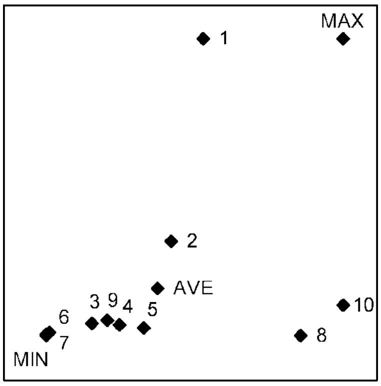

Fig. 9. Analysis of countries: the distribution of countries from the view of parameters x_1 and x_2 characterising the activities in *CORDIS* databases.

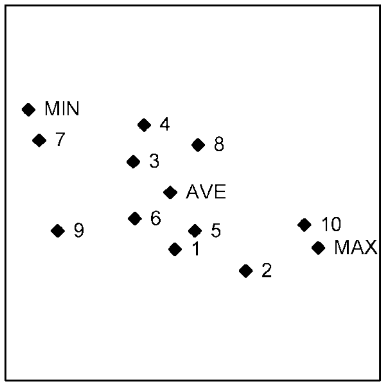

Fig. 10. Analysis of countries: Sammon's projection of countries from the view of economic and social parameters $x_3, \ldots, x_7$.

3×3-dimensional, and 4×4-dimensional. The corresponding Sammon's projections of results of the neural-analysis are presented in Figs. 11–16. Labels $1, \ldots, 7$ in the tables and figures indicate the parameters.

Like in Section 4.1, the following modifications of realisation of SOM were used: (*Posit*, *Consec*), (*Any*, *Consec*), (*Posit*, *Random*), (*Any*, *Random*). All the four modifications were examined using 4×4-dimensional SOM. The results are presented in Tables 8–11 and Figs. 11–14.

In addition, the realisation (*Any*, *Random*) of 2×2-dimensional and 3×3-dimensional SOM was applied to the data. The results are presented in Table 12 and Fig. 15 for 2×2-dimensional SOM and in Table 13 and Fig. 16 for 3×3-dimensional SOM. The smaller additional dimensions of SOM (2×2-dimensional and 3×3-dimensional) were selected here because of the smaller number of analysed objects (here we analyse 7 parameters, and the number of objects analysed in Section 4.1 is 13).

Table 8
Analysis of parameters:
Distribution on SOM 4×4, $Pos + Consec$

7		5	2
3			
4		6	1

Table 9
Analysis of parameters:
Distribution on SOM 4×4, $Any + Consec$

1			2
6			
3			
4	5		7

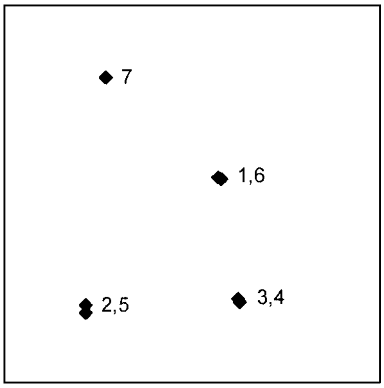

Fig. 11. Analysis of parameters: Combined mapping (SOM 4×4, $Pos + Consec$).

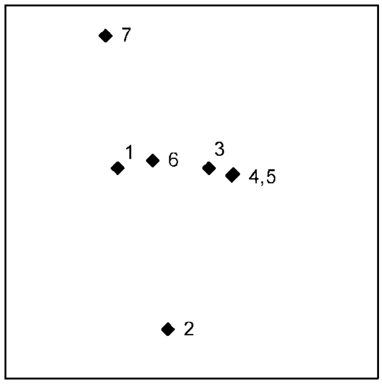

Fig. 12. Analysis of parameters: Combined mapping (SOM 4×4, $Any + Consec$).

Table 10
Analysis of parameters:
Distribution on SOM 4×4, $Pos + Random$

5			2
	1		3
7		6	4

Table 11
Analysis of parameters:
Distribution on SOM 4×4, $Any + Random$

7		6	4
	1		3
5			2

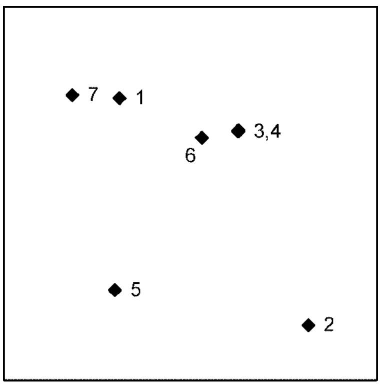

Fig. 13. Analysis of parameters: Combined mapping (SOM 4×4, $Pos + Random$).

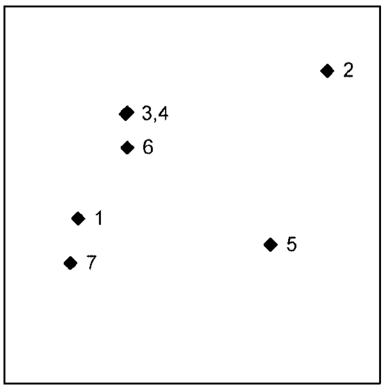

Fig. 14. Analysis of parameters: Combined mapping (SOM 4×4, $Any + Random$).

Table 12
Analysis of parameters:
Distribution on SOM 2×2, $Any + Random$

2, 5	
1, 6, 7	3, 4

Table 13
Analysis of parameters:
Distribution on SOM 3×3, $Any + Random$

2		5
7, 1	6	3, 4

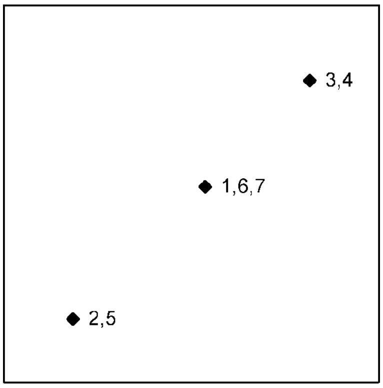

Fig. 15. Analysis of parameters: Combined mapping (SOM 2×2, $Any + Random$).

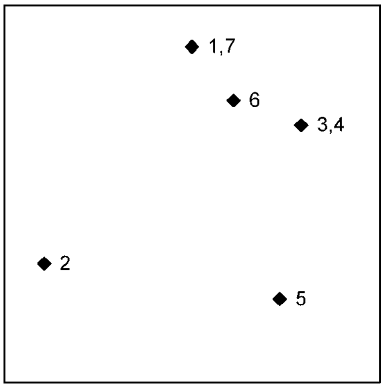

Fig. 16. Analysis of parameters: Combined mapping (SOM 3×3, $Any + Random$).

Comments and conclusions on the relative efficiency of different realisations of SOM are presented in Section 4.3, and these on parameters in general are presented in Section 4.4.

The pure Sammon's projection of parameters $x_1, \ldots, x_7$ on a plane is presented in Fig. 17. See comments in Section 4.4.

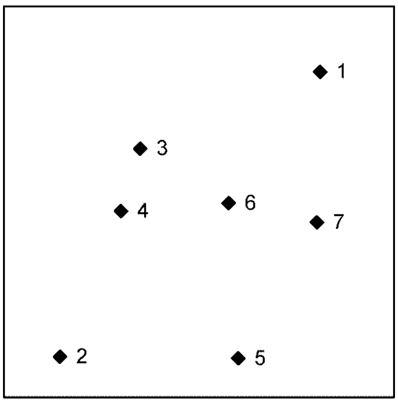

Fig. 17. Analysis of parameters: Sammon's projection of parameters $x_1, \ldots, x_7$.

4.3. *Conclusions on the Efficiency of Realisations of SOM*

Here we will draw conclusions on the comparative efficiency of the following modifications of realisation of SOM: (*Posit*, *Consec*), (*Any*, *Consec*), (*Posit*, *Random*), (*Any*, *Random*). All the four modifications were examined using 4×4-dimensional SOM on the basis of two data sets: data on countries (see results in Tables 2–5 and Figs. 3–6) and parameters (see results in Tables 8–11 and Figs. 11–14).

The worst modification of realisation of SOM is (*Posit*, *Consec*). Why? The answer follows from Figs. 3–6 and Tables 2–5. The distribution of objects in Fig. 3 differs significantly from these in Figs. 4–6: the distance between the opposite objects X_{MIN} and X_{MAX} in Fig. 3 is much less than in Figs. 4–6. Moreover, in Table 2 we also observe significant differences compared with Tables 3–5: in Table 2 the objects are located on the border-cells, and remaining six neighbouring cells are empty; meanwhile we do not see such empty zone in Tables 3–5. All these observations allow us to assume that (*Posit*, *Consec*) is the worst modification of realisation of SOM. The reason is that here we restrict the initial values of components of the codebook vectors m_{ij} (only positive values are used), and chose the consecutive order to pass the objects (input vectors) to the neural network during its training (the numeration of objects makes an influence on the training result).

For both the data sets, the results of two modifications (*Posit*, *Random*) and (*Any*, *Random*) coincide (Fig. 5 coincide with Fig. 6, and Table 5 is a "mirror-reflection" of Table 4; Fig. 14 is rotation of Fig. 13 by 90 degrees, and Table 11 is a "mirror-reflection" of Table 10 – rotations and mirror-reflections don't change the inter-location of mapped objects). This shows the advantage of the random order: if we chose the random order to pass the objects (input vectors) to the neural network during its training, the result is independent on the type of initialisation of the codebook vectors m_{ij} (with restriction on the positivity of the components or without the restriction).

If compare (*Posit*, *Consec*) with (*Posit*, *Random*) or (*Any*, *Random*), we observe a lot of essential differences between them. The obvious reason is the different order of passing the objects to the neural network. This allows us to refuse the consecutive order when analysing data of general nature (not obtained from correlation matrices).

We observe some similarities of (*Any*, *Consec*) with (*Any*, *Random*) (and (*Posit*, *Random*) that results coincide with (*Any*, *Random*)) and with (*Posit*, *Consec*). This means that (*Any*, *Consec*) is some intermediate between (*Posit*, *Consec*) and (*Any*, *Random*). The obvious reason is that the elimination of restriction on the positiveness of the initial values of components of the codebook vectors causes similarities between (*Any*, *Consec*) and (*Any*, *Random*), and the consecutive order causes similarities between (*Any*, *Consec*) and (*Posit*, *Consec*).

As a result of the discussion above we can draw a conclusion that, when analysing data of general nature, the most reasonable modification of realisation of SOM among the discussed above is (*Any*, *Random*). This is the reason why the additional experiments with SOM of various dimensions were carried out in Sections 4.1 and 4.2.

4.4. *Conclusions from the Visual Analysis by Using Mapping and SOM*

The conclusions have been made here on the basis of modification (*Any*, *Random*) of realisation of SOM. The choice of this modification is motivated in Section 4.3. Results for the conclusions are presented in Tables 5, 6, 7, 11, 12, 13 and Figs. 6, 7, 8, 9, 10, 14, 15, 16, 17.

When analysing the countries in view of the economic and social parameters $x_3, \ldots, x_7$ (Tables 5, 6, 7 and Figs. 6, 7, 8, 9, 10), we can see three groups (clusters) of the countries. These groups may be defined when analyse at the same time Figs. 6, 7 and 8 where (in comparison with Figs. 9 and 10) the countries are clustered additionally using the neural network. These three figures are also much more informative than the corresponding Tables 5, 6 and 7.

The first (the best) group consists of Slovenia (10) and Czech Republic (2). The worst group consists of Romania (7) and Bulgaria (9). The rest countries are in rather similar positions. They take places around the average AVE in Figs. 3, 4 and 6. However, in Figs. 6 and 7 we can visually group these countries into two subgroups:

- Hungary (1), Slovakia (5) and Poland (6);
- The Baltic States: Lithuania (3), Latvia (4) and Estonia (8).

From Fig. 8 we can draw a conclusion that Czech Republic (2) has some similarities with the first group, and Bulgaria (9) has these with the second one.

In view of activity of the countries in the CORDIS databases (parameters x_1 and x_2), Romania (7) and Poland (6) are in the worst position (see Fig. 9). The separate group that is slightly worse than the average consists of Lithuania (3), Bulgaria (9), Latvia (4) and Slovakia (5). In view of the parameter x_1, Estonia (8) and Slovenia (10) are in a good position. In view of x_2, Hungary (1) shows good results, Czech Republic (2) is slightly better than the average.

The investigation shows that it is impossible to make a strict conclusion that the country with better economic and social state would also be more active in CORDIS databases. We can see such trends only partially. It implies a conclusion that there exist some other factors facilitating activities of the country. Parameters x_1 and x_2 depend on these factors. Some of the economic and social parameters $x_3, \ldots, x_7$ also depend on these additional factors. Which of them? The visual analysis of the parameters can give an indirect answer.

The results on distribution of parameters $x_1, \ldots, x_7$ are drawn on the basis of Tables 11, 12, 13 and Figs. 14, 15, 16, 17.

When analysing jointly Figs. 14, 15 and 16 we see two pairs of very close parameters:

1. x_1 (number of projects) and x_7 (number of telephones per capita);
2. x_3 (infant mortality rate) and x_4 (gross domestic product per capita).

Parameter x_6 (export per capita) defines some intermediate state between the pairs above. Two parameters x_2 (published RTD results) and x_5 (level of industrialisation) are located far from the other five parameters. They may be considered as a separate (third) group. The first group of parameters characterises the level of communications (here we assume that the number of projects reflects extend of contacts among people). The second group characterises economical state of country. The third group characterises the level of technological development.

We can see in the Figs. 14, 15, 16, 17 (the best view in Fig. 17) that the parameters x_1 and x_2 describing activities in CORDIS databases are very different (most distinct) in the context of all parameters. This fact may be explained in this way: the activities of search of consumers of existing RTD results x_2 are in rather weak dependence from the number x_1 of projects supported by the European Commission.

5. Application of Regression Analysis

We have investigated here the relationships between the RTD activity, evaluated by numbers x_1 and x_2 of records in the CORDIS databases, and the economic and social parameters $x_3, \dots, x_8$ in corresponding countries. There are only a few data: $s = 10$.

The MS Excel was used for the analysis. The most interesting results are presented in Figs. 18–25. Rhombuses in the figures mark the data from the table (x_k, x_j). The results of analysis led to the conclusion that there is a tendency of growth in the number of records in the *CORDIS* databases (parameters x_1 and x_2) together with the growth of the economic and social parameters $x_3, \dots, x_7$. There is only exception that deals with the relationship between the industrialisation level x_5 and the number x_2 of records in CORDIS *RTD-Results* (see Fig. 24). Countries that have only a few records in the database *CORDIS RTD-Results* may be of any industrialisation level.

There is an interesting relationship between records in the databases *CORDIS RTD-Results* x_2 and *CORDIS RTD-Projects* x_1 (see Fig. 20). We may guess that there are an optimal number of projects to be in action. The sense of this conclusion is corroborated by the relationship between the international economic aid and the RTD results: more aid does not mean more RTD results, rather contrary (see Fig. 21).

There is exponential relation between GDP per capita x_4 and RTD results per million inhabitants x_2 (see Fig. 18). The interpretation may be this:

- There are only a few resources for RTD in the case of low GDP. On the contrary, sufficiently rich society may significantly facilitate RTD.
- Only very hard efforts in RTD and wide spread of results enable the growth of GDP.

It seems that the interpretations are not contradictory as both factors x_2 and x_4 intensify each other. Fig. 19 shows that the number x_1 of projects supported by EC is slightly weaker related with GDP than x_2. The parameters x_1 and x_2 have the similar regression relationship with the other economic and social parameters like the relationship with GDP discussed above.

Slightly different situation is presented in Figs. 22 and 23. The export per capita x_6 is strongly correlated with the number x_1 of EC projects but has week relation with the number x_2 of RTD results. The fact may be interpreted in the way that those who succeeded to gain the aid from EC are more successful in marketing then those who are successful in producing RTD results.

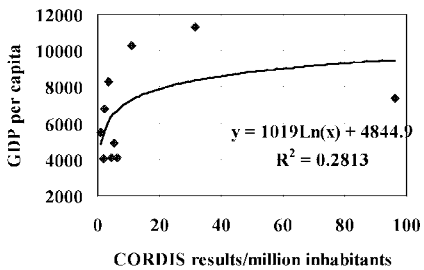

Fig. 18. Relationship $x_4 = y(x_2)$.

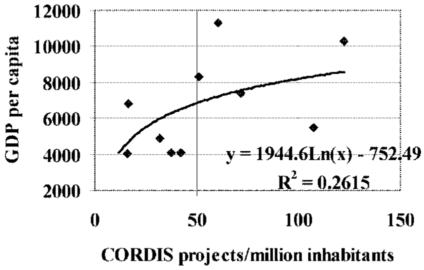

Fig. 19. Relationship $x_4 = y(x_1)$.

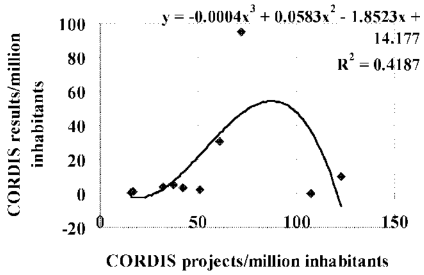

Fig. 20. Relationship $x_2 = y(x_1)$.

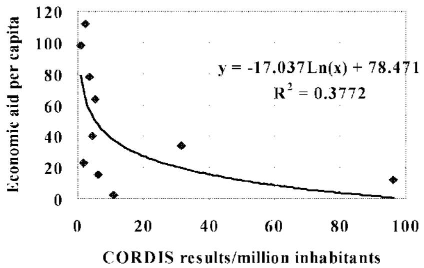

Fig. 21. Relationship $x_8 = y(x_2)$.

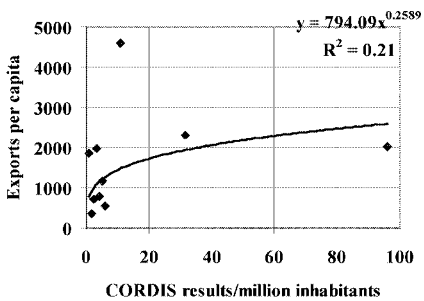

Fig. 22. Relationship $x_6 = y(x_2)$.

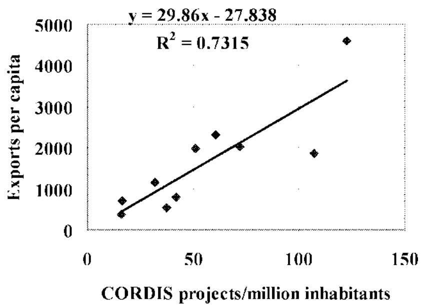

Fig. 23. Relationship $x_6 = y(x_1)$.

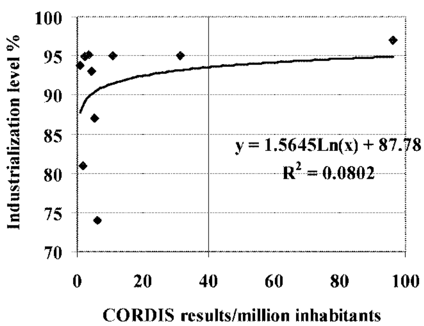

Fig. 24. Relationship $x_5 = y(x_2)$.

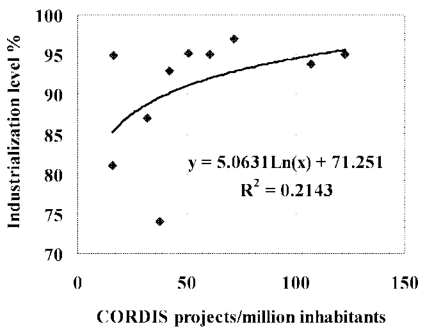

Fig. 25. Relationship $x_5 = y(x_1)$.

Country	Label
Hungary	1
Czech Republic	2
Lithuania	3
Latvia	4
Slovakia	5
Poland	6
Romania	7
Estonia	8
Bulgaria	9
Slovenia	10

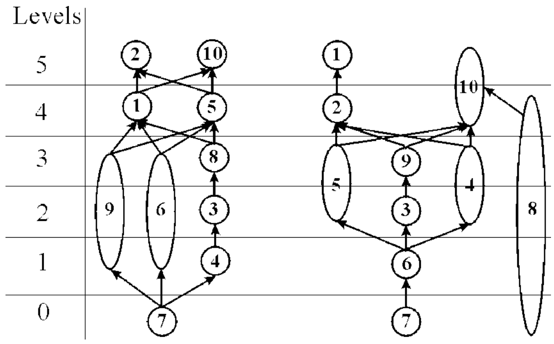

Fig. 26. Ranking of countries by domination.

6. Visual Analysis by Using Graphs of Domination

In Fig. 26 the countries are ranked by domination. On the right side of Fig. 26, the ranking is based on the RTD parameters x_1 and x_2; on the left the countries are ranked according to the core socio-economic parameters x_3, x_4 and x_6. Both rankings are similar, therefore they show the strong relation between RTD and socio-economic conditions.

7. Conclusions

The methodology, proposed in this paper, made it possible to get quantitative and qualitative estimates of relationship between the RTD activities in the Central European countries and the economical and social characteristics of these countries. We can observe a tendency of improvement of the economical and social state of the country when its RTD activity grows. Moreover, the valuable results are obtained on the clustering of countries both from the standpoint of the RTD activities and of the economical and social parameters. Data on 10 Central European countries were analysed. These countries are of different level of development, and neither essential social nor political cataclysms have been damaged these countries. The detailed conclusions presented in Sections 5 and 6 allow us to determine a position of any country among the other Central European countries from the standpoint of both the RTD activities and of the whole economical and social plane. An important result is the quantitative estimate of the place of Lithuania among the Central European countries: Lithuania appears a bit below the average from the point of view of a set of selected parameters.

Experiments on the real data set indicate that the integration of non-classical methods (SOM and graphs of domination) with classical ones (regress analysis and Sammon mapping) increases quality of analysis and allows making more exact conclusions.

Acknowledgement

The research in this paper was carried out in accordance with the European Commission's (EC) program *INCO-COPERNICUS*, project *INTACCOMP* (1999). The objective of the project is to develop an integrated system of national databases containing data on achievements in selected RTD fields. Using the Internet, the collected data would make it possible to get quantitative and qualitative estimates of the RTD results and activities in the Central Europe. The institutions of Czech Republic, France, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Spain participate in the project (see references in the WWW sites of *INTACCOMP* (1999)).

References

Bezdek, J.C., N.R. Pal (1995). An index of topological preservation for feature extraction. *Pattern Recognition*, **28**, 381–391.

CORDIS (1999). *The Community Research and Development Information Service CORDIS.* `http://cordis.lu/src/g_005_en.htm`

Country Data (1999). *The World Bank Group-Development Data*. `http://www.worldbank.org/data/`.

Čaplinskas, A. (Ed.) (2000). *Databases & Information Systems*, *Proceedings of the 4-th IEEE International Baltic Workshop*, Vol. 1–2. Technika, Vilnius.

Dzemyda, G. (2001). Visualization of a set of parameters characterized by their correlation matrix. *Computational Statistics and Data Analysis*, **36**(1), 15–30.

Dzemyda, G., V. Tiešis (2000). Analysis of economic and social impact to the activity in EC RTD databases. In A. Čaplinskas (Ed.), *Databases&Information Systems*, *Proceedings of the 4-th IEEE International Baltic Workshop*, Vol. 2. Technika, Vilnius. pp. 220–230.

Fausett, L. (1994). *Fundamentals of Neural Networks*. Prentice Hall.

Fayyad, U.M. (1997). Editorial. *Data Mining and Knowledge Discovery*, **1**, 5–10.

Flexer, A. (1997). Limitations of self-organizing maps for vector quantization and multidimensional scaling. In M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*. MIT Press/Bradford Books, Cambridge, MA. pp.445–451.

INTACCOMP (1999). *Integrated Network of RTD Accomplishments INTACCOMP*. `http://www.science.mii.lt/intaccomp/`, `http://neris.mii.lt/INTACCOMP/`

Kaski, S. (1997). *Data Exploration Using Self-Organizing Maps*. *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series*, **82.** Espoo, Finish Academy of Technology. `http://www.cis.hut.fi/~sami/thesis/thesis.html`

Kohonen, T. (1989). *Self-Organization and Associative Memory*, 3rd ed. Springer-Verlag, Heidelberg.

Kohonen, T. (2001). *Self-Organizing Maps*, 3nd ed. *Springer Series in Information Sciences*, Vol. 30. Springer-Verlag, Heidelberg.

Kohonen, T., J. Hynninen, J. Kangas, J. Laaksonen (1996). *SOM_PAK: The Self-Organizing Map Program Package*. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland. `http://www.cis.hut.fi/research/som_lvq_pak.shtml`

Kleiweg, P. (1996). *Neurale Netwerken: Een Inleidende Cursus Met Practica Voor de Studie Alfa-Informatica, Master's Thesis.* Rijksuniversiteit Groningen. `http://odur.let.rug.nl/~kleiweg/kohonen/kohonen.html`

Murtagh, F. (2000). *Fionn Murtagh's Multivariate Data Analysis Software and Resources Page*. `http://newb6.u-strasbg.fr/~fmurtagh/mda-sw/`

Murtagh, F., M. Hernindez-Pajares (1995). The Kohonen self-organizing map method: an assessment. *Journal of Classification*, **12**, 165–190.

Ritter, H., K. Schulten, T. Martinetz (1991). *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley.

Sammon, J.W., Jr. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, **C-18**, 401–409.

The World Factbook (1999).
`http://www.odci.gov/cia/publications/factbook/country.html`

Zell, A., G. Mamier, M. Vogt, N. Mache, R. Hübner, S. Döring, K.-U. Herrmann, T. Soyez, M. Schmalzl, T. Sommer, A. Hatzigeorgiou, D. Posselt, T. Schreiner, B. Kett, G. Clemente, J. Wieland, M. Reczko, M. Riedmiller, M. Seemann, M. Ritt, J. DeCoster, J. Biedermann, J. Danz, C. Wehrfritz, R. Werner, M. Berthold, B. Orsier (1995). *SNNS. Stuttgart Neural Network Simulator*, *User Manual*, *Version 4.1*. Report No. 6/95, Institute for Parallel and Distributed High Performance Systems, University of Stuttgart.

G. Dzemyda graduated from Kaunas University of Technology, Lithuania, in 1980, and in 1984 received there the doctoral degree in technical sciences (Ph.D.) after post-graduate studies at the Institute of Mathematics and Informatics, Vilnius, Lithuania. In 1997 he received the degree of Doctor Habilius from Kaunas University of Technology. He was conferred the title of Professor (1998) at Kaunas University of Technology. Recent employment is at the Optimization Department of the Institute of Mathematics and Informatics as a Principal Researcher. The main field of scientific interests is the interaction of optimization and data analysis. The interests include optimization theory and applications, multiple criteria decisions, neural networks, and data analysis.

V. Tiešis graduated from the Kaunas University of Technology. He occupies the position of a researcher at the Institute of Mathematics and Informatics. He has published more than 70 scientific papers. His research interests include non-linear and integer optimisation algorithms, conditions of convergence, and efficiency of rough algorithms. His present interests include modelling and control in the epidemiology, economics, military operation and in the technology.

Daugiamačių objektų vizualizacija ir socio-ekonominių faktorių įtaka aktyvumui Europos komisijos mokslo ir technologijų duomenų bazėse

Gintautas DZEMYDA, Vytautas TIEŠIS

Straipsnyje pasiūlyti ir ištirti daugiamačių objektų vizualizacijos metodai, naudojant netiesinę Sammon projekciją, klasifikuojantį Kohoneno neuroninį tinklą, regresinę analizę bei dominavimo grafus. Šie metodai taikyti analizuojant Vidurio Europos šalių mokslinių ir technologinių tyrimų (MTT) aktyvumą ir jo ryšį su ekonominiais-socialiniais parametrais. Duomenys apie MTT projektus ir rezultatus išskirti iš Europos komisijos duomenų bazių CORDIS, o duomenys apie ekonominius-socialinius parametrus iš JAV CŽV ir Pasaulio banko duomenų bazių. Gauti rezultatai apie šalių tarpusavio grupavimąsi bei MTT ryšį su ekonominiais-socialiniais parametrais.