Bulletin of the *Transilvania* University of Braşov • Vol. 6 (55) •No. 1 - 2013
Series V: Economic Sciences

THE TOURISM AMONG THE RIPARIAN BLACK SEA COUNTRIES

Nicoleta PETCU[1]

Abstract: *The industry of tourism and travelling is nowadays the most dynamic field worldwide, thus being the most important job generator. From an economic point of view, tourism is also a main source of recovery of the national economies of those countries that have important tourist resources and exploit them accordingly. Its action develops on more plans, starting with stimulating economic development to improving social structure, with the superior capitalization of resources to the improvement of life conditions. This work aims at the implementation of the multi-dimensional analysis methods: main components analysis, cluster analysis, relative distances method to establish Romanian's place among the countries with harbour to the Black Sea, on the basis of the tourism indicators.*

Key words: *relative distances method, main components analysis, cluster analysis.*

1. Introduction

Data Mining (the analysis step of the **Knowledge Discovery in Databases** process, or KDD) Fayyad, Usama; Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996), a relatively young and interdisciplinary field of computer science, is the process of discovering new patterns from large data sets involving methods from statistics and artificial intelligence but also database management. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data in order to extract previously unknown interesting patterns such as groups of data records: **cluster analysis** Agrawal, R.; Gehrke, J.; Gunopulos, D.; Raghavan, P. (2005) unusual records - **anomaly detection**, Hans-Peter Kriegel, Peer Kröger, Arthur Zimek (2009, dependencies - **association rule mining**, Varun Chandola, Arindam Banerjee, and Vipin Kumar(2009)) and **decision support system**, Keen (1978).

Methods: Cluster analysis or clustering is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters.

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is a way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities.

In statistics, regression analysis includes any techniques for modelling and

[1] Dept. of Management and Economic Informatics, *Transilvania* University of Braşov.

analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, **regression analysis** helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Factor analysis is a statistical method used to describe variability among observed variables in terms of a potentially lower number of unobserved variables called **factors**. In other words, it is possible, for example, that variations in three or four observed variables mainly reflect the variations in fewer such unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modelled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.

2.The multi-criterion hierarchy of the territorial units

The comparisons at the territorial level and the classifications of the territorial units are of great importance both at national level, for a harmonious and balanced development of all the regions of the country, and at international level, to ensure the measurement of the discrepancies and the drawing up of the best development strategy [1].

To achieve a multi-criterion hierarchy of the territorial units, several methods have been formulated in the specialized literature, out of which the relative distances method will be exemplified.

The relative distances method belongs to the category of methods that calculate the distance between the elements (territorial units) of an *m*-space, when m represents the number of features included in the analysis.

The method implies:

- choosing a fictitious unit whose features present the minimum (or maximum) levels noticed in the real community;
- choosing a method of measuring the distance between the real units and the fictitious unit (for each feature analysed);
- settling a procedure of aggregating the information got for each real unit.

The aggregation of the information in an average index with separate levels for each unit of the community allows the measurement of the real discrepancies between units, as well as the use of the multicriterial hierarchy results in a subsequent statistical research based on parameter procedures.

For the example chosen, the relative distances method between units is applied, the distances being calculated as co-ordinating relative sizes against the unit with a maximum performance:

$$d_{ij}^{\%} = \frac{x_{ij}}{\max\limits_{1\le i\le n}(x_{ij})} \cdot 100 \; for \quad i = \overline{1,n} \quad j = \overline{1,m} \tag{1}$$

To aggregate the co-ordinating relative sizes in an average index, the geometric mean is used:

$$\bar{d}_i = \sqrt[m]{\prod_{j=1}^{m} d_{ij}} \; for \quad i = \overline{1,n} \tag{2}$$

The option for a multiplying-type aggregation is determined by the following reasons:

- a product of indices leads to the calculation of a geometric mean, which has the advantage of being less exposed to the influence of extreme values and therefore is a more precise value than the arithmetic mean;
- the probability of getting the same product for two or more territorial units is less than that of the equal sums, thus diminishing the subjective intervention in establishing the final hierarchy.

Amongst the Central and Eastern Europe

countries, Romania is considered to be endowed with the richest and most varied natural and anthropic tourist resources, which results in a great availability to tourism. The potential of the Romanian tourism could compete with the tourist offer of any other country in the world, contributing to the increasing revenues coming from the external exchanges.

Nowadays, the Romanian hospitality industry faces problems, such as: the decline of the internal and external tourist demand; an outdated tourist product; the low standard of the tourist services provided, which do not meet the tourist demand; insufficient promotion etc.

The prospect that the area of the Black Sea countries will become an important pole in the future, receiving the international tourist flows, has to convince and also stimulate Romania.

It is important to establish Romanian's place within the tourist framework of these countries, considering a range of general and specific indicators. Among the general indicators, allowing the achievement of some pertinent analogies at the level of the group of the analyzed countries, we will analyze:

V1- GDP per capita, (PPP USD), **V2** - Exports (mill US $) FOB, **V3** - Imports (mill US $) CIF, **V4** - Unemployment (% of total labour force), **V5** - Inflation, consumer prices (annual %), **V6** - Investment, gross fixed (% of GDP), **V7** - Public spending on education (% of GDP), **V8** - Health expenditure, public (% of GDP, Tourism indicators: **V9** - International tourism, receipts (million US $), **V10** - International tourism, expenditures (mill US $), **V11** - International tourism, number of arrivals(thou), **V12** - International tourism, number of departures (thou), considering the year 2011.

Economic indicators Table 1

Geo	V1	V2	V3	V4	V5	V6	V7	V8
Bulgaria	14603	26080	28390	9.6	4.2	22.4	4.4	3.7
Romania	15163	62500	70820	5.1	5.8	23.9	4.3	4.4
Turkey	17499	133000	212200	9.8	6.5	21.8	2.9	5.1
Ukraine	7251	60670	72080	7	8	19	5.3	4.4
Georgia	5503	3083	5960	16.3	8.5	16.3	3.2	2.4

The data source: http://www.factfish.com/catalog/economy

Table 2 Relative distances method Table 2

Geo	V1	V2	V3	V4	V5	V6	V7	V8	Avg	Final score
Bulgaria	83.45	19.61	13.38	53.13	100	93.72	83.02	72.55	53.35	4
Romania	86.65	46.99	33.37	100.00	72.41	100.00	81.13	86.27	71.58	2
Turkey	100.00	100.00	100.00	52.04	64.62	91.21	54.72	100.00	80.00	1
Ukraine	41.44	45.62	33.97	72.86	52.5	79.50	100.00	86.27	60.02	3
Georgia	31.45	2.32	2.81	31.29	49.41	68.20	60.38	47.06	22.31	5

Applying the relative distances method, the classification of the countries by the eight characteristics is: first place - Turkey, second place Romania, Ukraine comes third 3, Bulgaria comes forth and Georgia is fifth.

Indicators specific to tourism Table 3

Geo	**International tourism, receipts (million US $)**	**International tourism, expenditures (mill US $)**	**International tourism, number of arrivals (thou)**	**International tourism, number of departures (thou)**
Bulgaria	4035	1382	6047	3676
Romania	1653	1897	7575	11723
Turkey	24784	5451	27000	11002
Ukraine	4696	4134	21203	17180
Georgia	738	328	2033	2089

The data source: http://www.factfish.com/catalog/economy

Relative distances method Table 4

Geo	**V9**	**V10**	**V11**	**V12**	**Avg**	**Final score**
Bulgaria	16.28	25.35	22.40	21.40	21.09	4
Romania	6.67	34.80	28.06	68.24	25.82	3
Turkey	100.00	100.00	100.00	64.04	89.46	1
Ukraine	18.95	75.84	78.53	100.00	57.96	2
Georgia	2.98	6.02	7.53	12.16	6.36	5

3. Principal Component Analysis

The factorial methods have a double objective [3]:

- *to simplify* a table with raw data by passing from a great number of variables to a smaller number of new variables got by grouping the initial ones;
- *to select* from an important multitude of variables those that appear more frequently in the description of the phenomenon considered;
- *to structure and interpret* the input data, owing to the small number of new variables – hidden components – that simplify the interpretation of the less legible data at the beginning.

The existence of the correlations between variables makes it possible for the reduction of the dimension of their representation space, not by their reduction, but by building new aggregative variables with the following features:

- The new variables are linear combinations of the initial combinations and they each contribute with a descending part to the variance of the data. They are called principal components, and each Pi component is of the following form:

$$P_i = a_1 V_1 + a_2 V_2 + ... + a_k V_k \quad (3)$$

- The *p* principal components are independent of each other, i.e. not correlated.

The factorial analysis methods are data reduction methods, replacing the initial scatter by one of more restricted dimensions for a convenient graphical representation.

The reduction is possible if the data chart can be represented by two scatters: that of **individuals-points** in the space of variables and that of **variables-points** in the space of individuals. The simultaneous representation in the same reduced space is thoroughly justified and the link between them can be analyzed.

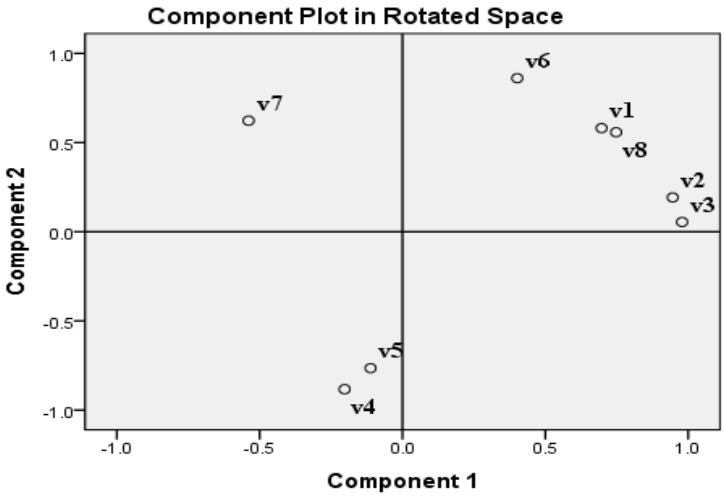

Fig. 1. *Variables-points in the space of individuals*

Regarding the chart, it becomes obvious that factor no. 1 is close to the variables: V2, V3, V8, V1, Exports, Imports, Health expenditure, public, Gross domestic product per capita, while factor no. 2 is close to the variables V4, V5, V6, Unemployment, Inflation, and Investment.

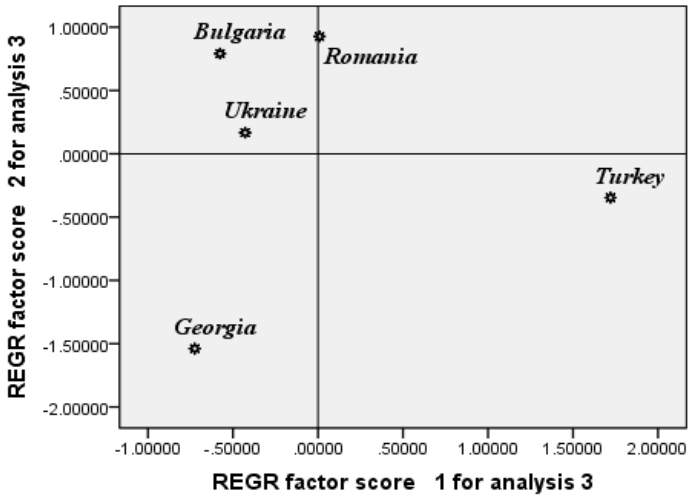

Fig. 2. *Individuals-points in the space of variables*

Turkey registers the highest values of the indicators that define the first factor. Ukraine, Bulgaria and Romania register similar values regarding Public spending on education (% of GDP). Georgia registers the highest values of V4 - Unemployment (% of total labour force), V5 - Inflation, consumer prices (annual %), while for the other economic indicators, the values are lower.

Data processing was made using the SPSS product.

4. Cluster analysis, kMeans method

The cluster analysis aims at describing a group of individuals or of objects characterized by a group of attributes by means of their regrouping in classes. These classes are established in such a way that the objects belonging to the same class should be the most similar possible and the objects belonging to two different classes should be the most different.

The input data are organized in an individual-variable table. The groups are established according to two big categories of procedures which resort to the hierarchical or non-hierarchical methods using the rectangular or Euclidean distances.

$$d_{ij} = \left\| X_i - C_j \right\|^2 = \sum_{q=1}^{Q} (x_{qi} - c_{qj})^2 \quad (4)$$

where X_i is the vector of encoded input fields for record i, C_j is the cluster centre vector for cluster j, Q is the number of encoded input fields, x_{qi} is the value of the q_{th} encoded input field for the i_{th} record, and c_{qj} is the value of the q_{th} encoded input field for the j_{th} record.

For each record, the distance between the record and each cluster centre is calculated, and the cluster centre whose distance from the record is smallest is assigned as the record's new cluster.

When all records have been assigned, the cluster centres are updated.

Using the cluster analysis, you can see that, from the tourist indicators point of view, Romania, Bulgaria and Georgia are included in a category (the smallest values), followed by Ukraine (medium values), out of which Turkey is standing out (the highest values).

5. Conclusions

The integration of the Romanian tourism within the European trends proves to be a hard, long-term and still insecure process. Even if tourism had not been a distinctive negotiation chapter in view of the adherence to the European Union, there are many other chapters directly or indirectly connected to tourism and which had been negotiated and closed. Among the most important ones, we mention: the free circulation of persons, services and capital, the transportation policy, the small and medium-size companies, the regional politics and the co-ordination of the structural instruments, the environment etc. But we are supposed to know our place and role that we play or we want to play in the European tourism.

Romania will face a challenge which requires a great financial, professional and educational effort, considering the strong points of the Romanian tourism, but being aware and disposing of the weak points characterizing this field today.

References

1. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: *From Data Mining to Knowledge Discovery in Databases"*. AI Magasine, 1996, p. 41.
2. Fernandez, G.: *Data Mining Using SAS Applications*. Charman & Hall, 2003, p. 187.
3. Härdle, W., Simar, L.: *Applied Multivariate Statistical Analysis*. TECH, 2002, p. 202.
4. Petcu, N.: *Statistică – teorie şi aplicaţii în SPSS*. Editura Infomarket, Braşov. 2003, p. 164.
5. Petcu, N.: *Tehnici de Data Mining rezolvate în SPSS Clementine*. Editura Albastră, Cluj-Napoca. 2010.
6. Tenenhaus, M.: *Methodes Statistiques en gestion*. Dunod, Paris. 1994, p. 243.
7. Voineagu, V., Furtună, F., Voineagu, E., Ştefănescu, C.: *Analiza factorială a fenomenelor social-economice în profil territorial*. Editura Aramis Print s.r.l., Bucureşti. 2002, p. 175.
8. Chandola, V., Banerjee, A., Kumar, V.: *Anomaly Detection: A Survey*. ACM Computing Surveys, Vol. 41(3), Article 15, July 2009.