Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 31 (2014) 468 – 473

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2014)

Analysis of global data education and patent activity using new methods of pattern analysis

Alexey Myachin*

National Research University High School of Economics, 26/5 Shabolovka Street, 5422, Moscow, 101000, Russia

Abstract

This paper study the world of education data and patent activity for the period of 1979-2006 years using the latest methods of pattern analysis: a linear pattern-classification and ordinal-invariant pattern clustering. Attempt is made to reflect the situation regarding primary, secondary and higher education in 37 countries.

Keywords: pattern analysis; linear pattern classification; data analysis

1. Introduction

Today very urgent problem is the development of science and education, which is very complicated, but interesting for the study. Regard to this, the work is devoted to aggregate data analysis of primary, secondary and higher education, as well as patent activity in 37 countries. Data considered to a relatively large time interval, more precisely, 28 years.

To carry out the necessary analysis here used a relatively new and prospective way of analysis - a method of pattern analysis. This method is based on the principle of the search of relationships between objects that reflect their attributive description, classification and construction of the dynamic analysis. In other words, the initial set of objects is partitioned into a number of subsets called patterns, with the proviso that each pattern objects have similar characteristics but different patterns differed significantly.

This method was chosen due to a number of its advantages over classical methods (such as ranking), which

* Corresponding author.
E-mail address: a_miachin@mail.ru

doi:10.1016/j.procs.2014.05.291

is important in the joint analysis of the education data and patenting activity in different countries, such as the ratio of one pattern of objects that have substantially different quantitative characteristics, but similar structure as internal indicators themselves and their relationships. Dynamic part of the analysis, which are formed at different dynamic groups can not only find a certain trend of development, but also identify the relationship between implicit indicators and atypical dynamics of the objects, which is a great interest for expert analysis.

In this paper we also applied the modified algorithms of pattern analysis: linear pattern classification and ordinal-invariant pattern clustering. The latest algorithms are selected due to their distinctive features, which allow to identify sustainable patterns that do not depend on the order of consideration of selected data that allows the study not only by peer review, but also to consider all possible combinations of the basic indicators.

The objects of study are from 37 countries: Bulgaria, Belgium, Czech Republic, Denmark, Germany, Ireland, Greece, Spain, France, Italy, Cyprus, Latvia, Liechtenstein, Luxembourg, Malta, Norway, Austria, Portugal, Romania, Slovenia, Finland, Sweden, United Kingdom, Ireland, Norway, Croatia, Turkey, Russia, Canada, USA, Mexico, China, India, Singapore, Australia and New Zealand. Purpose is to reflect not only relationships, but also the situation regarding primary, secondary and higher education, as well as patent activity in the countries.

2. Description of the research methodology

For the beginning, we describe the method of pattern analysis in general, and then present the latest research algorithms. The analysis process consists of 3 stages, each of which is divided into several sub-steps:

1. The initial stage;
1.1 Modelling of initial system of indicators;
1.2 Data Collection;
1.3 Check of quality of the data found;
1.4 The correlation analysis of the data used;
1.5 Aggregation;
1.6 Preparation of the base system of indicators.
2. Analytical stage;
2.1 Construction of objects in the form specific functions (in this paper used piecewise linear functions);
2.2 Conduct of clustering;
2.3 Interpretation of the results of clustering.
3. Dynamic phase of analysis;
3.1 Construction of trajectories expressing development of objects in time;
3.2 Identification of dynamic groups.

It should be noted that the method of pattern analysis previously been successfully applied to solve a number of applied character problems such as policy[5], management and personnel management[3,4,5], macroeconomics[1,2] and the analysis of the banking sector[5].

For holding the whole analysis, we describe a formal methodology for the study. There are aggregate of objects X, and a set of clusters Y. You must specify a function p (x, x') to refer to a measure of proximity of objects x and x', belonging to the set of objects X. This set is divided into disjoint subsets of similar by metric objects p. Each subset of objects is attributed to the number (label name) y_i. This pattern is understood as a set of values that describe and define the group of objects, as well as a set of objects with similar values of the indicators.

A system of parallel coordinate[6] is used in this work to present pattern as a piecewise linear function which consists of n parallel lines, each reflecting one of the indicators. Combining the i-th vertex, we obtain a piecewise linear function, reflecting the structural relationship between the parameters of each examined object.

Clustering in this work is done using sequential execution of algorithms of linear pattern classification and ordinal-invariant pattern clustering.

3. Data description

To solve the problem of development of science and education we formulate a necessary condition for parameters of the basic system, reflecting, on the one hand, the educational potential of the countries at the primary, secondary and higher education, on the other hand, the patent activity in the countries studied. The source data must satisfy a number of criteria:

1. Reliability of both the data and the source;
2. The indicator should be sensitive to changes in the situation associated with the formation and study of patent activity in the country;
3. Unambiguous definition of factors such as the method of computing and semantic definition of the indicator;
4. Exception of emissions while correlation analysis;
5. Certain regularity when collecting data;
6. Minimum error of data measurement.

In the study was supposed to use an advanced system of indicators, however, due to the fact that have had selected a sufficiently long period of analysis, the real problem was the lack of some of the data therefore a system of indicators has been changed. As a result, baseline characterized with 2 units:

1. Educational potential of the country, characterized by quantitative ratios of students in primary and secondary education, higher education students, as well as the number of teachers and faculty members;
2. Patenting activity in countries characterized by quantitative ratios filed and issued patent applications in countries.

As a result, were selected seven indicators for which the study was held. All data is collected from the sources[7,8] for 37 countries around the world in the period 1979 - 2006 and are shown in Table 1.

Table 1. Initial data

№	Index	Source
	Block 1. Education data	
1	General admission to primary education. Public and private schools. All Levels	Official site of the United Nations Educational, Scientific and Cultural Organization "UNESCO"
2	General admission into secondary education. Public and private schools. All Levels	Official site of the United Nations Educational, Scientific and Cultural Organization "UNESCO"
3	Total number of teachers in secondary education. Public and private schools. Complete and incomplete working days. All programs	Official site of the United Nations Educational, Scientific and Cultural Organization "UNESCO"
4	General admission to the higher education system. Public and private schools. All Levels	Official site of the United Nations Educational, Scientific and Cultural Organization "UNESCO"
5	Total number of teaching staff in higher education. Public and private schools. Complete and incomplete working days. All programs	Official site of the United Nations Educational, Scientific and Cultural Organization "UNESCO"

Block 2. The data relating to patent activity in the countries		
6	Patents granted by priority year at the national level	The official website of the «Eurostat»
7	Patent applications by priority year at the national level	The official website of the «Eurostat»

4. Data analysis

The study used data for the period of 1979 - 2006 years within 37 different countries, i.e. turned out 1036 piecewise linear function. Since the analysis is a dynamic, multiple patterns may include the same country but in a different time periods. Here is an example of some of the resulting patterns.

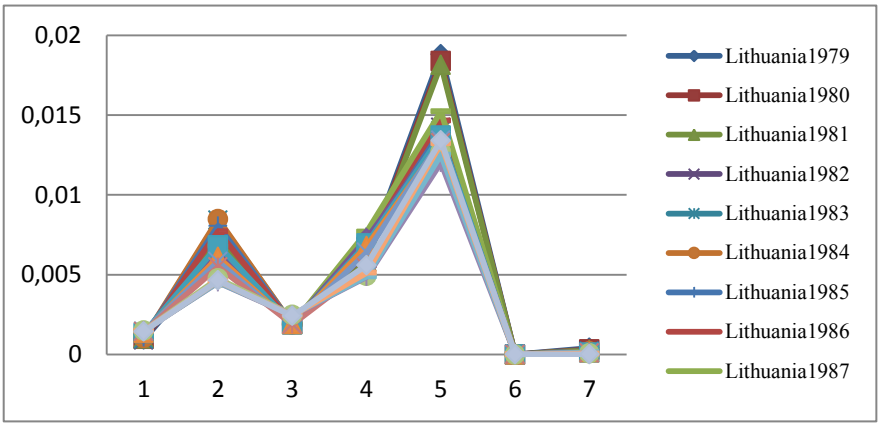

Fig. 1. First pattern

The first pattern is characterized solely for Lithuania in the period 1979 - 1997 years. It is worth noting that when aggregated, values of the 6 (i.e., the number of patents granted) was zero in the period 1979 - 1994. This fact means the lowest activity of this indicator compared to all the countries studied in the period. In 1995-1997 this indicator is also very close to 0. In 1997 we have a relatively small number of teachers in high schools (13,136), but the number of teachers exceeds more than 3 times. Level of patent activity is very low.

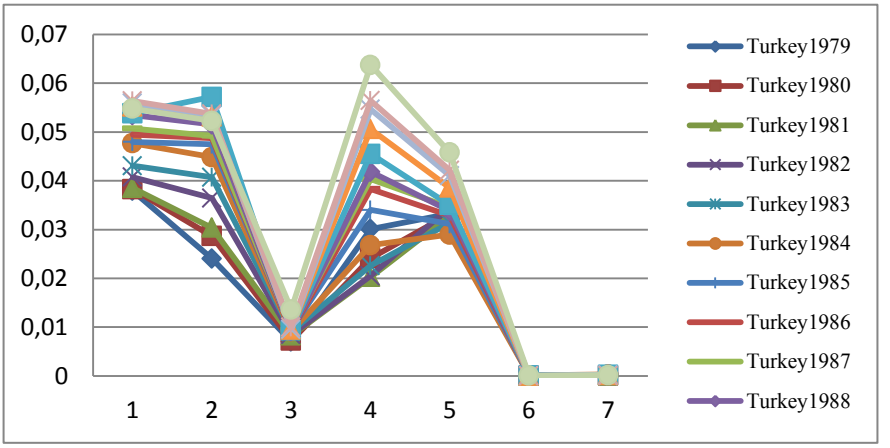

Fig. 2. Second pattern

The second pattern characterize Turkey 1979 - 1993 period, i.e. includes 15 piecewise linear functions. Here we see a fairly low patent activity, while there is a relatively large number of students (1993 admission to secondary education is 3,216,418 students), and the gap between the number of teachers in schools and lecturers in universities is 4.6 times.

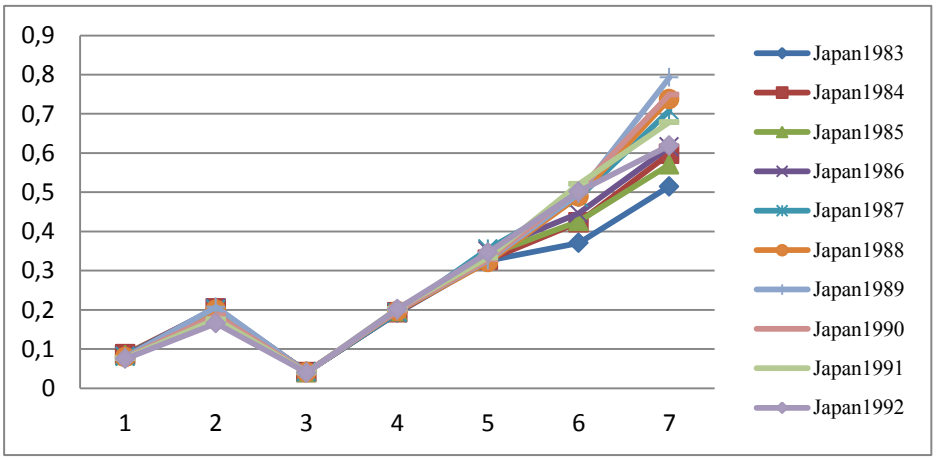

Fig. 3. Third pattern

The third includes exclusively Japan, during the period 1983 - 1992, i.e. it includes 10 kinked curve. At this period traced a high patenting activity, the ratio between the number enrolled in primary and secondary education is almost equal to 1, and between enrolled in secondary and higher differs almost 3.5 times (Japan 1991).

As an example, we present a dynamic analysis of the trajectory of Russia, India and China in Figure 4.

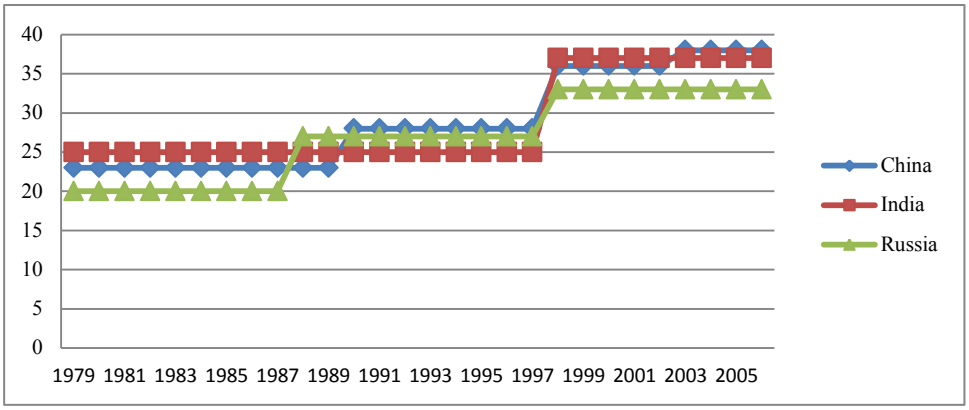

Fig. 4. Dynamic analysis

References

1. Aleskerov F., Alper C.E. Inflation, Money, and Output Growths: Some Observations // Bogazici University Research Paper, #SBE 9606, 1996.
2. Aleskerov F., Alper C.E. A clustering approach to some monetary facts: a long-run analysis of cross-country data // The Japanese Economic Review, v.51, no.4, 2000, 555-567.
3. Aleskerov F., Ersel H., Gundes C., Minibas A., Yolalan R. Environmental Grouping of Bank Branches and their Performances // Yapi Kredi Discussion Paper Series, No: 97-03, 1997, Istanbul, Turkey
4. Aleskerov F., Ersel H., Gundes C., Yolalan R. A Multicriterial Method for Personnel Allocation among Bank Branches // Yapi Kredi Discussion Paper Series, No:98-01, 1998, Istanbul, Turkey.
5. Aleskerov F., Ersel H., Yolalan R. Multicriterial Ranking Approach for Evaluating Bank Branch Performance // International Journal of Information Technology and Decision Making, v.3, no.2, 2004, 321-335.
6. Few S. Multivariate Analysis Using Parallel Coordinates, 2006. Internet Source: http://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf
7. The official website for European statistics «Eurostat». Internet Source: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/
8. The official website of « UNESCO ». Internet Source: http://www.unesco.org