Self-Organizing Neural Networks for the Analysis of Country Development

Ildefonso Padrón Peña[1], Oto Mráz[2]

[1]Department of Theoretical Computer Science and Mathematical Logic,
Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
Malostranské náměstí 25, 118 00 Prague 1
ildepadronhosna@gmail.com
[2]Department of Informatics, Faculty of Natural & Mathematical Sciences,
King´s College, London, UK
Bush House, 30 Aldwych, WC2B 4BG London
oto.mraz@gmail.com

Abstract. National economies can be influenced in various ways. Our research is focused on post-Soviet and Spanish speaking countries. To understand the actual state of the countries and to assess possible trends for their development, we use the SOM-networks. Based on development indicators provided by the World Bank, our approach is shown to explain both major trends as well as changes in the development of the countries. Stable economies are characterized by minor changes and sustained development. Crises, on the other hand, tend to manifest themselves rather by turbulent movements of the economy.

This paper demonstrates various changes in the country's economy, such as economic and moral crises, or socio-political turmoils, but also positive events like joining economical unions or organizing international sports events. Further, we will provide possible reasons for the causes of these trends and crises.

Key words: self-organizing maps, macroeconomic analysis with neural networks, data analysis, data visualization.

1 Introduction

Crises may take different forms such as war, economic downturn, natural disasters, etc. Conversely, events like hosting the football championships can bring tangible improvements, in Russia, it increased the Gross Domestic Product (GDP) by an extra 0.2% [7]. For different countries, however, their impact on society may differ. The main goal of our work is therefore to propose a new methodology capable of detecting regularities or anomalies in country development represented in the form of time-series data. The amount and character of the found anomalies might, namely, help to asses and visualise the actual state of the investigated country. This paper is focused on the evolution of post-Soviet and Spanish speaking countries based on data from the World Bank. We will consider a large range of indicators, i.e. GDP per capita, inflation or literacy rates, processed using self-organizing Kohonen maps (SOMs).

2 Related Work

Various research papers demonstrate applicability of neural networks and other machine learning techniques to the analysis of economic data. Recent models for time-series forecasting [2] and data clustering methods [1], [3], [5], [8] have been studied with emphasis on different aspects, e.g., time and space complexity, dependencies between variables, or correlation between data size and the model's architecture. Unfortunately, most of the known clustering approaches do not adequately reflect the topological structure of the data, which might provide better insight into the details of the data set. The capability of topology-preserving mappings represents, however an advantage of SOMs [4] with a huge potential impact on the result interpretation.

Kohonen maps provide a powerful tool allowing to process high dimensional data by reducing the dimensionality of the input space. The arrangement of the neurons on a 2D-grid should reflect the topology of the input structure. After a large number of training iterations, the neurons will be located accordingly on the map.

3 Data and its Pre-Processing

Our data was obtained from the World Bank database [9]. Altogether, 19 different indicators showing economic and social aspects from 2000 until 2012 were used, each carrying equal weight (**Tab. 1**). In our research we investigated the development of 16 post-Soviet states and 17 Spanish speaking countries in 2000-2012[1]. In total, 14.66% and 13.41% of the data was missing from the post-Soviet and the Spanish speaking countries, respectively. For the approximation, the missing ends were substituted by the closest known value. Missing intervals with known bounding values were linearly approximated. If for a given indicator, a country had no data at all, the mean of the indicator was used. Then, the values of the *i*-th indicator, denoted val_i^{old}, were transformed into val_i^{new}, such that the transformed values retained the original distribution and remained in the range $[0,1]$. Equation (1) was used for pre-processing:

$$val_i^{new} = \frac{val_i^{old} - val_i^{min}}{val_i^{max} - val_i^{min}} \quad (1)$$

val_i^{min} and val_i^{max} correspond to the min and max values of the original val_i^{old}.

4 Supporting Experiments

For the prepared data, two SOMs with the same initial parameters were trained (one for the post-Soviet and one for the Spanish speaking countries). The SOM model used had a 2D hexagonal grid with 20×30 neurons in the Kohonen layer as this provides

[1] The post-Soviet countries are Russia, Estonia, Latvia, Lithuania, Belarus, Ukraine, Moldova, Georgia, Armenia, Azerbaijan, Kazakhstan, Uzbekistan, Turkmenistan, Tajikistan and Kyrgyzstan, Czech Republic was added for reference. The Spanish speaking states include Argentina, Bolivia, Chile, Colombia, Cuba, Dominican Republic, Ecuador, El Salvador, Honduras, Mexico, Panama, Paraguay, Peru, Puerto Rico, Spain, Uruguay and Venezuela.

a reasonably large space to obtain convincing results given the data size. Both SOMs were trained for 50000 iterations. Next, the data for each country was projected onto the respective SOM. The evolution of the economy over time is shown by a line connecting data points representing two consecutive years. Lastly, the initial (2000) and final (2012) data points were marked in green and red, respectively.

Among the post-Soviet countries two clusters were formed. The larger one on the right contains mainly the poorer regions, like Central Asia or the Caucasus. The smaller one on the left includes rather richer countries, e.g., the Baltic States. In the case of the Spanish speaking countries, geographically close regions tend to be adjacent to each other on the SOM. The main cluster on the right contains most Latin American countries. The two smaller clusters on the top left correspond to Spain and Puerto Rico (due to its political relationship with the US, an unincorporated territory).

Tab. 1. The used macroeconomic indicators

Development indicator	Units
1.GDP growth	annual %
2.GDP per capita	current US$
3.GNI per capita, Atlas method	current US$
4. GNI per capita, PPP	current international $
5. Exports of goods and services	% of GDP
6. Imports of goods and services	% of GDP
7. GINI index	World Bank estimate
8. Foreign direct investment per capita	Balance of Payments, current US$
9. Unemployment rate	% of total labour force
10. Poverty headcount ratio at national poverty lines	% of population
11. Inflation, consumer prices	annual %
12. Inflation, GDP deflator	annual %
13. Internet users	per 100 people
14. Life expectancy at birth, total	years
15. Adult literacy rate, population 15+ years, both sexes	%
16. Agriculture, value added	% of GDP
17. CO_2 emissions	metric tons per capita
18. Central government debt, total	% of GDP
19. Population density	per km^2

5 Analysis of Country Development

To find regularities or anomalies, we used the Manhattan distance, $d_M(\boldsymbol{c}^t, \boldsymbol{c}^{t+1})$:

$$d_M(\boldsymbol{c}^t, \boldsymbol{c}^{t+1}) = |c_i^t - c_i^{t+1}| + |c_j^t - c_j^{t+1}|$$

of Kohonen neurons $\boldsymbol{c}^t$ with the coordinates $\boldsymbol{c}^t = (c_i^t, c_j^t)$ and $\boldsymbol{c}^{t+1}$ with the coordinates $\boldsymbol{c}^{t+1} = (c_i^{t+1}, c_j^{t+1})$ representing the considered country in the respective years

t and $t+1$. Then we compared the obtained values $d_M(\boldsymbol{c}^t, \boldsymbol{c}^{t+1})$ with a threshold distance experimentally chosen to be 10% of the biggest possible displacement in the map. Since we used a 20 by 30 grid of neurons, the max Manhattan distance between any two neurons is $|30-1| + |20-1| = 29 + 19 = 48$ giving a threshold value of 4.8. If the obtained distance was higher, that year was marked as a significant change. This metric proved great in identifying the most important changes, which correlated very well with the socio-economic development in those years. **Tab. 2** contains the most interesting countries, with the significant changes in highlighted in grey.

Tab. 2. Positions of 6 example countries on the Kohonen maps

Country	2000	01	02	03	04	05	06	07	08	09	10	11	12
Spain	24	24	24	2	2	2	1	1	0	0	0	0	0
	0	0	0	10	10	10	9	9	10	7	7	7	8
Estonia	16	16	2	2	3	3	3	3	4	2	3	3	4
	16	16	5	5	3	3	5	5	4	4	4	4	5
Azerbaijan	8	8	3	3	4	6	3	5	5	5	15	12	14
	24	24	19	20	19	19	17	18	17	17	23	24	24
Argentina	17	17	14	16	16	24	24	24	24	22	27	27	23
	0	0	5	9	9	2	2	2	2	0	2	2	1
Armenia	3	4	4	4	4	4	3	3	4	4	5	5	5
	15	14	14	14	14	14	14	14	15	16	14	15	16
Bolivia	19	20	20	20	19	21	21	21	22	21	20	20	21
	6	5	5	5	5	7	8	8	8	11	11	11	10

Considering post-Soviet countries, Estonia, e.g., started in the main cluster, but moved to the smaller one in 2002, most likely due to its EU accession. Conversely, Azerbaijan changed its position quite frequently but lacked sustained progress. The first change in 2002 could be due to the constitutional amendments. In 2006 GDP growth peaked at 41.5%. Lastly in 2010 the price of oil increased heavily, which Azerbaijan relies on unlike its neighbours Georgia and Armenia.

Looking at the Spanish speaking countries, the isolation of Spain in a separate cluster since 2003 could be explained by the Euro adoption in 2002. Considering Argentina, big position changes happened in 2002, 2003, 2005, 2010, and 2012. The first one can be related to the Argentine Great Depression (1998-2002, the economy shrank by 28%). The economic recovery started with 2003 presidential elections. In 2005 the GDP exceeded pre-crisis levels. In 2009 Argentina was affected by the global recession. First GDP growth decreased to nearly zero, and then expanded to around 9% in 2010 falling back to 1,9% in 2012, one the lowest in Latin America.

Analysing the evolution of the countries in the Kohonen maps, we can thus distinguish the following three main types of countries (a complete set of the Kohonen maps as well as the data used can be accessed online [6]):

1. Countries with a very local movement on the grid. This can be understood as steady economic development, with no considerable socio-economic changes, e.g. Armenia (**Fig. 1**), Tajikistan, Bolivia or El Salvador.

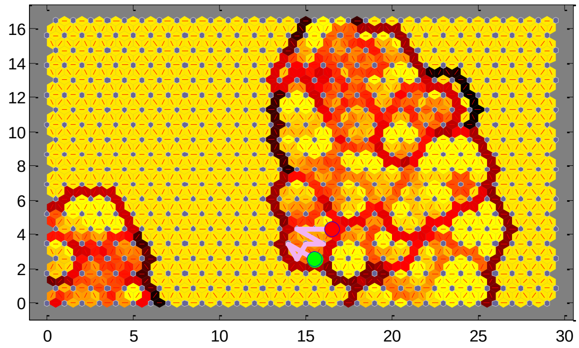

Fig. 1. Projection of data from Armenia

2. Regions which underwent a notable shift such as Spain (**Fig. 2**), Estonia, Latvia or Lithuania. Major changes in the country's position in the lattice often relate to significant social, political, or economic changes.

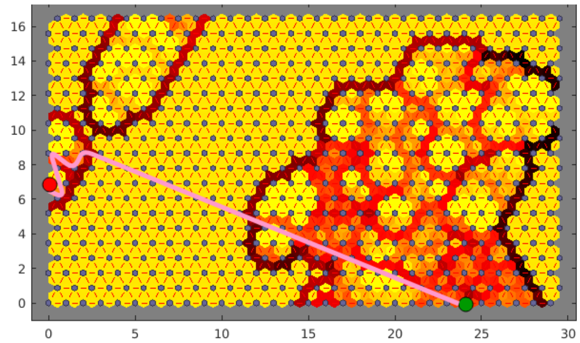

Fig. 2. Projection of data from Spain

3. Countries behaving in an unstable manner, making several significant movements back and forth. These are often unstable economies with high levels of corruption and social insecurity, e.g., Azerbaijan (**Fig. 3**), Moldova, Argentina or Colombia.

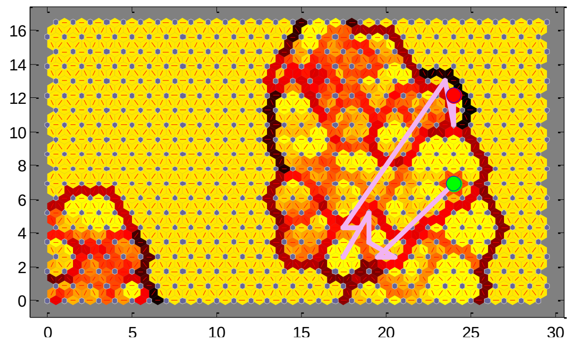

Fig. 3. Projection of data from Azerbaijan

6 Conclusion

Our research has shown the usefulness of Kohonen maps for the analysis of high-dimensional macroeconomical data. The processed 19 indicators for post-Soviet and Spanish speaking countries were projected onto a 2D SOM-map, which helped us group together countries with similar development. Our original contribution to the area consists in introducing a new methodology based on the analysis of pre-trained Kohonen maps to identify anomalies in the development of time-series data.

Major shifts in the positions of the countries on the Kohonen maps were often accompanied by socio-political or economical reforms, like EU accession. Conversely, countries rich in oil shift a lot, yet without any steady improvements. Other reasons for instability may include also significant economic downturns, natural disasters or wars. Within the framework of our further research we would like extend our experiments to more countries from different parts of the world and to investigate additional indicators, identifying the crucial ones for a country's stable development and economic growth. This knowledge would help us predict which country will develop stably, and which one is likely to undergo a considerable change in the near future.

Literature

1. Collan, M., Eklund, T., Back, B.: Using the self-organizing map to visualize and explore socio-economic development. EBS Review 22 (2007) 6–15.
2. Cook, T.R., Hall, A.S.: Macroeconomic Indicator Forecasting with Deep Neural Networks. Research Working Paper No. 17-11. Federal Reserve Bank of Kansas City, USA (2017) 1–38, https://doi.org/10.18651/RWP2017-11 [online accessed on 17. 6. 2018].
3. Focardi, S.M.: Clustering economic and financial time series: Exploring the existence of stable correlation conditions. Discussion Paper No. 04, The Intertek Group, Paris, France (2001) 1–15, https://www.researchgate.net/profile/Sergio_Focardi2/publication/2547390_Clustering_economic_and_financial_time_series_Exploring_the_existence_of_stable_correlation_conditions/links/56b379c608ae5deb2657d5ba.pdf [online accessed on 17. 6. 2018].
4. Kohonen, T.: Self-organizing maps. Springer-Verlag, Berlin, 2001.
5. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. JMLR 9 (2008) 2579–2605.
6. Mráz, O., Peña, I. P.: Kohonen map research (2018), https://otomraz.wixsite.com/kohonenmapresearch
7. Palata, L.: Russia was different. For a while. (Rusko bylo jiné. Na chvíli.) Mladá Fronta Dnes, Praha, 18. 7. 2018 (2018) p.7 (in Czech).
8. Řezanková, H.: Cluster Analysis of Economic Data. Statistika: Statistics and Economy Journal 94 (2014) 73–86.
9. World Bank: World Bank Open Data. World Bank, Washington, USA (2018), https://data.worldbank.org [Online accessed on 17. 6. 2018]