Attribute Portfolio Distance: A Dynamic Time Warping based approach to comparing and detecting common spatiotemporal patterns among multi-attribute data portfolios

Jesse Piburn[1], Robert Stewart[1,2], April Morton[1]

[1]Oak Ridge National Laboratory, 1 Bethel Valley Road Oak Ridge, TN 37830,
Email:{piburnjo; stewartrn; mortonam} @ornl.gov

[2]Dept. of Geography, University of Tennessee, Burchfiel Geography Building Room 304 Knoxville, TN 37996

Abstract

In this paper, we propose a novel extension to dynamic time warping (DTW), a data mining algorithm used to find similarities between two time-series vectors, called *Attribute Portfolio Distance* (APD) that estimates the spatiotemporal similarity of a *portfolio* (group) of attribute time series for geographic entities. APD represents a novel response to highly generalized questions such as what countries are most similar economically where "economics" is defined by portfolio selection. We develop APD and a close variation known as Trend Only Attribute Portfolio Distance (TO-APD). We apply both to a portfolio of World Bank economic data for European nations from 1994 to 2013 to show how APD can illuminate which nations are economically similar to Ukraine when considering multiple attributes in space and time.

Keywords: Dynamic Time Warping, Time-Series Clustering, Attribute Portfolio Distance.

1. Introduction

Dynamic time warping (DTW) is a data mining algorithm used to measure the similarity between two time series by computing the minimum cumulative distance between them. DTW has already seen wide application in fields such as automatic speech recognition where detecting similar audio signals is a key capability (Sakoe & Chiba, 1978). In this paper, we propose a novel extension to DTW called *Attribute Portfolio Distance* (APD) that estimates the spatiotemporal similarity of a *portfolio* (group) of attribute time series for geographic entities. APD therefore facilitates a holistic comparison of multi-attributed temporal behaviors. For geocomputation, APD represents a novel response to highly generalized questions such as what countries are most similar economically where "economics" is defined by portfolio selection. We develop APD and a close variation known as Trend Only Attribute Portfolio Distance (TO-APD). We apply both to a portfolio of World Bank economic data for European nations from 1994 to 2013 to show how APD can illuminate which nations are economically similar to Ukraine when considering multiple attributes in space and time.

2. Dynamic Time Warping

Given two vectors $\mathbf{x} = (x_1, x_2, \dots, x_i, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_j, \dots, y_m)$ they can be arranged to form an n x m matrix, **C**, referred to as a cost matrix where each (i, j) pair is the distance between x_i and y_j. Distance here is defined by a distance function, $\delta(.)$. Often times this distance is Euclidean, but any appropriate measure of distance can be used. The goal is to find the warping path, $\mathbf{w} = (w_1, w_2, \dots, w_k)$, which minimizes the distance between **x** and **y** equation 1. Finding the warping path that minimizes the distance between **x** and **y** is equivalent to finding the least cost path across the cost matrix, **C**, where $DTW(\mathbf{x}, \mathbf{y})$ is equal to the cumulative cost of the warping path fig. 1. For a more detailed description of DTW see Berndt & Clifford (1994).

$$DTW(\mathbf{x}, \mathbf{y}) = min\ \mathbf{w}\,[\sum_{k=1}^{p} \delta(w_k)] \qquad (1)$$

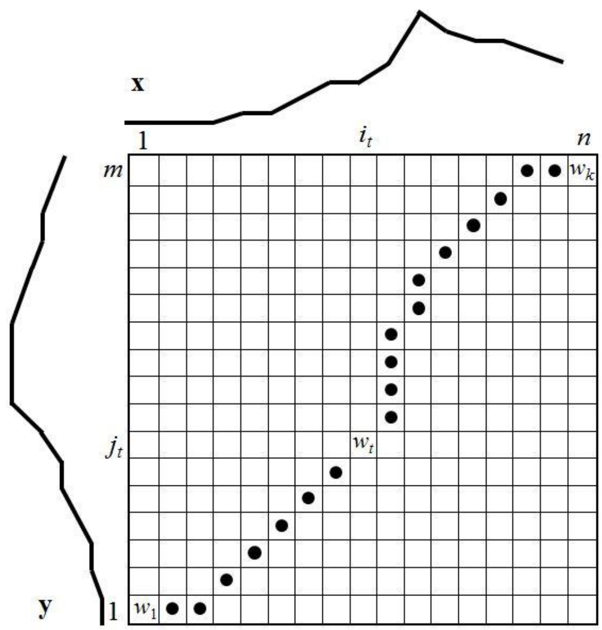

Figure 1. Dynamic Time Warping

3. Attribute Portfolio Distance

Now that we have defined $DTW(.)$ we can extend it to the application of attribute portfolios. Given n countries and m attributes, an n x m distance matrix **D** can be formed where each element (i, j) is equal to $DTW(\boldsymbol{\tau}_j, \mathbf{c}_{ij})$. Where $\boldsymbol{\tau}_j$ is a vector corresponding to the values of attribute j for the chosen target country τ and where $\mathbf{c}_{ij}$ is a vector corresponding to the values of comparison country i for attribute j.

The resulting matrix, **D**, has country centric rows and attribute centric columns. At this point, each column is still in the units of the corresponding attribute thus to allow for equal comparison across attributes we now divide every element in a column by that columns corresponding root mean square (RMS) $\sqrt{\sum_{i=1}^{n}(x^2/n-1)}$. The RMS is a measure of the magnitude of a varying quantity and by dividing each element in a column by its RMS this effectively removes the attribute's original units leaving only a measure of its variation, allowing each attribute to be equally compared regardless of their original unit of measurement.

To reduce our m measures of distance for each country to one summary measure we apply a summary function, $\varphi(.)$ over each row. The specific summary function can be chosen based on the application of the analysis. A simple summation could be appropriate when one is interested in an overall measure of similarity across the entire attribute portfolio, while other summary functions such as variance would allow one to see how consistent the similarities are across attributes in the portfolio. Applying the summary function $\varphi(.)$ row wise on $\mathbf{D}$ results in a $1 \text{ x } n$ column vector, $\mathbf{d}$, where $\mathbf{d}_i = \varphi(\mathbf{D}_{i,*})$ for all indices i from 1 to n.

We now have a single value for each comparison country called the APD, that represents how similar that country is to the target country for the given attribute portfolio. Note that the APD of the target country will always be equal to 0, the lower bound of APD measurement. These APD values can now be explored with the same techniques as any standard spatially referenced attribute, such as Moran's I, LISA, or K-means clustering.

4. Trend Only Attribute Portfolio Distance

The APD values described above considers not only the temporal trend but the magnitude of the trends as well. However we can "level the playing field" and consider only the temporal trend by using Trend Only APD (TO-APD). TO-APD is not necessarily a modification to the APD methodology, but rather a modification of the data that we use as inputs to APD. To only consider the temporal trend of each attribute we transform each country attribute vector, including the target country attributes, into standardized z-scores and perform APD on these transformed data. This transformation removes the differences in magnitude and allows for comparison of just the temporal trends.

5. Application and Results

We apply these two methods by exploring the relationship between European countries and Ukraine using a portfolio of selected economic indicators for the 20 year period of 1994 - 2013. The 34 comparison countries are listed in table 1.

Comparison Countries		
Albania	Germany	Portugal
Austria	Greece	Romania
Belarus	Hungary	Russian Federation
Belgium	Ireland	Serbia
Bosnia and Herzegovina	Italy	Slovak Republic
Bulgaria	Latvia	Slovenia
Croatia	Lithuania	Spain
Czech Republic	Macedonia, FYR	Sweden

Denmark	Moldova	Switzerland
Estonia	Netherlands	United Kingdom
Finland	Norway	
France	Poland	

Table 1. Countries being compared to the target country Ukraine

The selected attributes are a profile of each country's economic and population trends. These 23 attributes are listed in table 2.

Attribute Name	Source
Agriculture, value added (% of GDP)	World Bank
Birth rate, crude (per 1,000 people)	World Bank
Exports of goods and services (% of GDP)	World Bank
GDP (current US$)	World Bank
GDP growth (annual %)	World Bank
GDP per capita (current US$)	World Bank
GNI per capita, Atlas method (current US$)	World Bank
GNI per capita, PPP (current international $)	World Bank
GNI, Atlas method (current US$)	World Bank
GNI, PPP (current international $)	World Bank
Gross capital formation (% of GDP)	World Bank
Imports of goods and services (% of GDP)	World Bank
Industry, value added (% of GDP)	World Bank
Inflation, GDP deflator (annual %)	World Bank
Labor force, total	World Bank
Population ages 0-14 (% of total)	World Bank
Population ages 15-64 (% of total)	World Bank
Population, female (% of total)	World Bank
Population, total	World Bank
Services, etc., value added (% of GDP)	World Bank
Total reserves (includes gold, current US$)	World Bank
Unemployment, total (% of total labor force) (modeled ILO estimate)	World Bank
Urban population	World Bank

Table 2. Names and sources of attributes included in the analysis

For our distance function $\delta(.)$ we use Euclidean distance and our summary function $\varphi(.)$ is a simple summation.

The results for our APD analysis in fig. 2 show that Romania has the closest APD to the target country Ukraine for our attribute portfolio followed closely by Belarus and Lithuania. This means that out of all the countries in our analysis these countries were the most similar in magnitude and temporal trend to Ukraine for our attribute portfolio.

Intuitively this result makes sense given that these countries are similar in both the size and structure of their economies, similar in both the magnitude and the temporal trend of the attributes in our portfolio. Also distinct is the division between east and west with Russia being a large exception. This exception is likely due to Russia being much larger in key attribute measures such as GDP and Population. Even if Ukraine and Russia had the exact temporal trends for every attribute the fact that Russia's magnitude is much larger results in a large APD. To explore if this is the explanation, we use Trend Only APD to investigate the similarities of only the temporal trends. The results are shown in fig. 3.

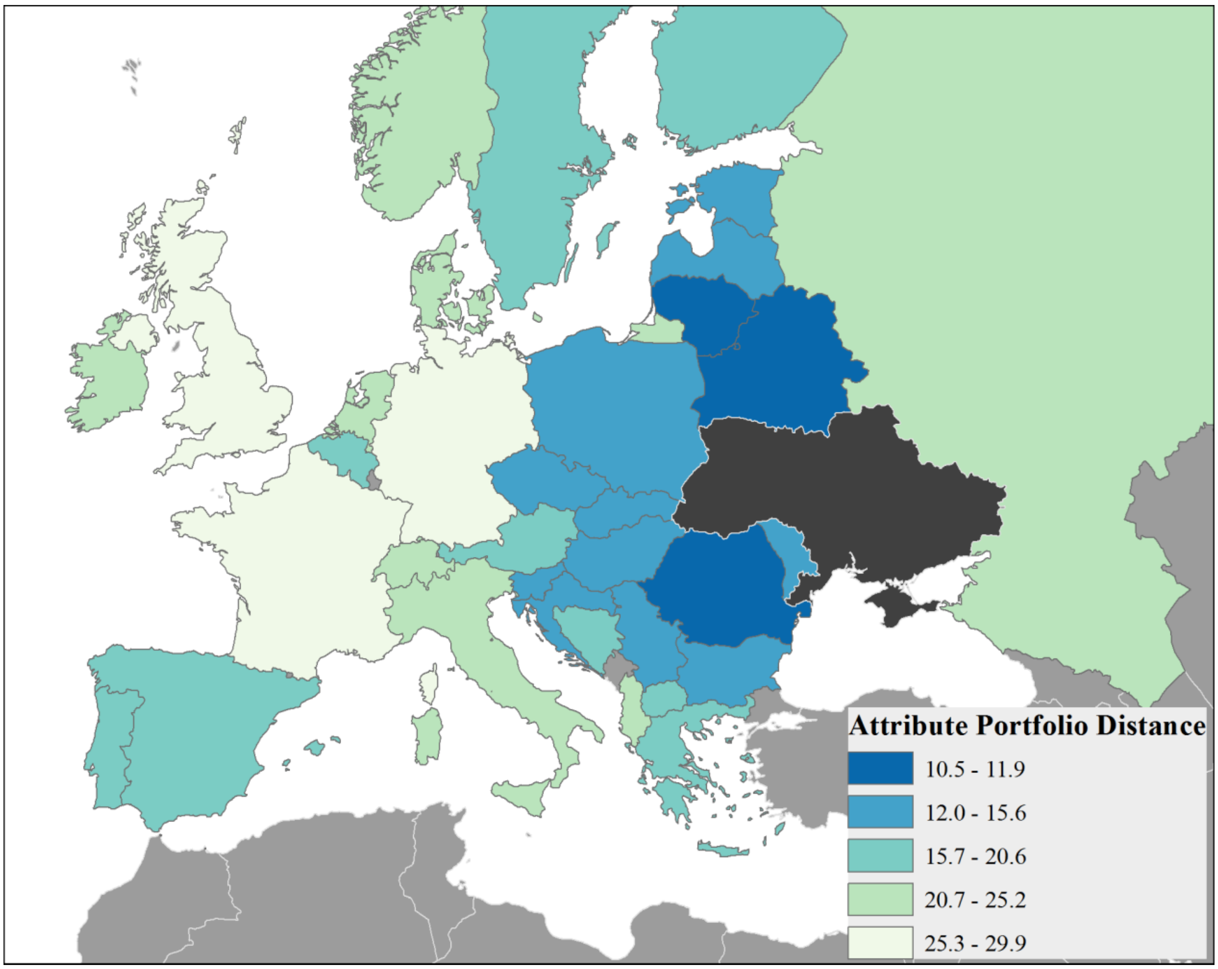

Figure 2. Results of APD analysis on a Ukrainian economic attribute portfolio

When Trend Only APD analysis is applied to our attribute portfolio, with only similarity of temporal trends being considered, Russia and Moldova emerge as the closest in TO-APD to Ukraine. This means that out of all the countries in our analysis Russia and Moldova were the most similar in temporal trends to Ukraine for our attribute portfolio. This highlights an important difference in APD and TO-APD in that although neither Moldova nor Russia are similar to Ukraine in the magnitude of their attributes, they are very close in their temporal trends.

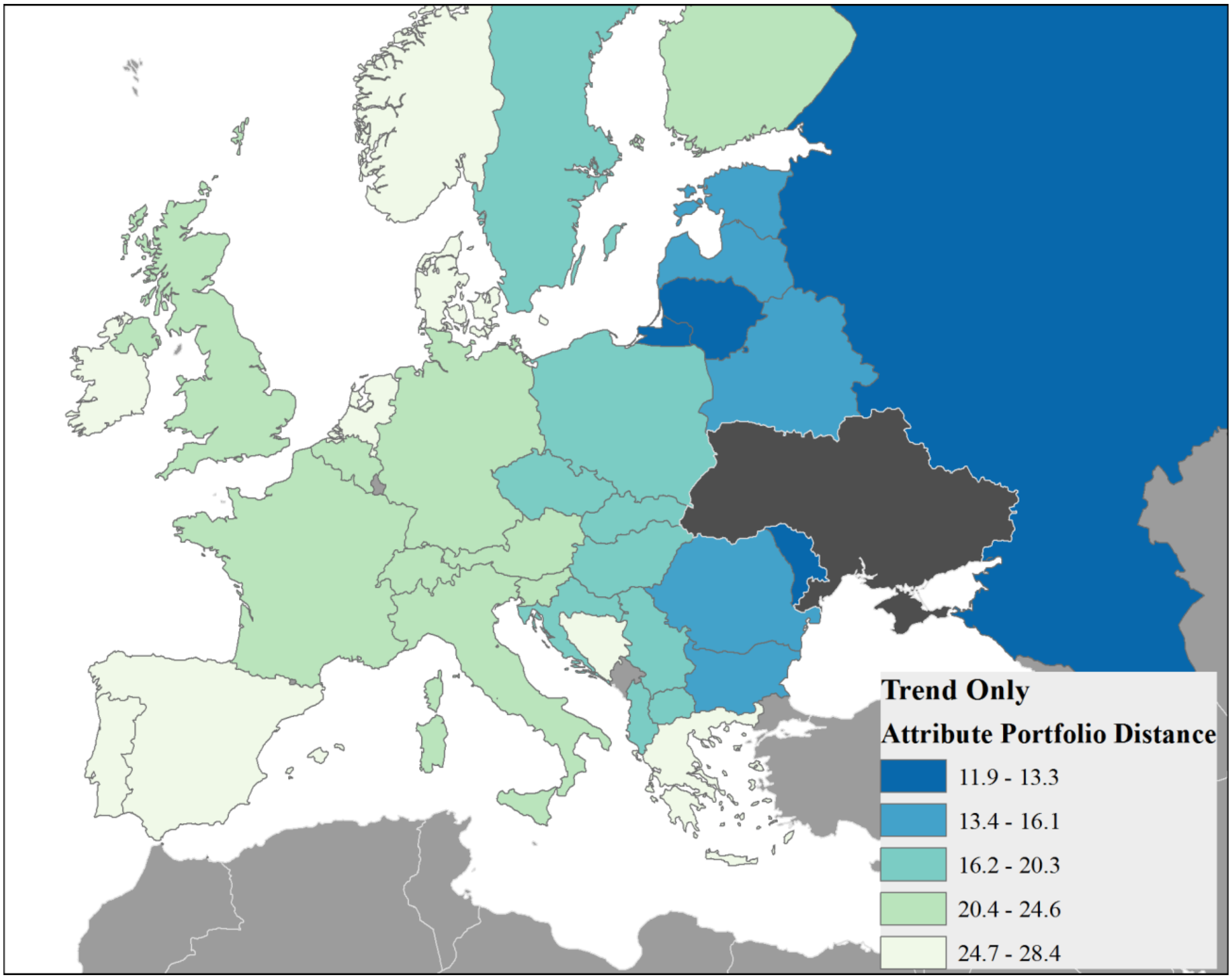

Figure 3. Results of Trend Only APD analysis on a Ukrainian economic attribute portfolio

6. Summary

In this paper we introduce a spatiotemporal extension of DTW known as Attribute Portfolio Distance and its close variant trend only attribute portfolio distance. Both enable modelers to explore and summarize high dimensional relationships in spatiotemporal data. We apply both to World Bank data to demonstrate how both are useful in detecting similar national behaviors for multi-attributed portfolios.

7. Acknowledgements

This manuscript has been authored by employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up,

irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

8. References

Berndt D J, and Clifford J, 1994, Using Dynamic Time Warping to Find Patterns in Time Series, *KDD workshop*, 10(16): 359-370.

Sakoe H, and Chiba S, 1978, Dynamic programming algorithm optimization for spoken word recognition, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 26(1): 43-49.