UC Santa Barbara

Title
A Taxonomic Classification Approach for Global Spatio-temporal Data

Authors
Stewart, Robert
Piburn, Jesse
Walters, Sarah
et al.

Peer reviewed

A Taxonomic Classification Approach for Global Spatio-temporal Data

Robert Stewart, Jesse Piburn, Sarah Walters, Jason Kaufman, Evan Ezell, David Anderson, Eric Axley, Josh Grant, Bryan Eaton, Alex Sorokine, and Greg Simpson

Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831
stewartrn@ornl.gov

Abstract. The World Bank, World Health Organization, and other major vendors provide thousands of country level, time series data sets concerning the environment, health, economics, violence, education, and national security. Practitioners consuming these time series data are informed by and interested in the detection of trends and patterns, keen on knowing how those trends spatially arrange. Trends that indicate economic rebounds, flattened pandemic curves, or steady increases in violence are just a few examples. Given the high dimensionality and noisy nature of the data, it is often difficult to extract and reason about semantic patterns. We propose Categorize Trends, an exploratory analysis tool which labels high dimensional time series data into a manageable set of key behavioral classes and provides the functionality to examine the results spatially using behavioral maps. We apply our method to an important use case exploring the interaction between food scarcity and fertility.

Keywords: spatio-temporal · data mining · dynamic time warping · exploratory analysis · visualization · World Bank

1 Introduction

The World Bank, World Health Organization (WHO), and other major vendors provide thousands of country level, time series (TS) data sets concerning the environment, health, economics, violence, education, and national security. Practitioners consuming these data are often informed by and interested in the detection of specific trends relevant to the study of socioeconomic development. Finding trends that exhibit specific temporal patterns within a collection of large and noisy data sets is an important task for practitioners in multiple domains; which patterns are worth finding however is often domain specific and can take place over different time horizons. Constructing exemplar templates of domain specific patterns and using them to search for individual trends of interest is an

approach to this problem that has seen successful applications in both the financial time series literature [5,9], where templates of well-known chart patterns are constructed and used to search for potentially profitable trades, as well as within remote sensing [4,6], where sequential NDVI images have been compared to reference profiles for the identification of agricultural practices.

In this paper we adapt this approach to the domain of socioeconomic development and demonstrate its usefulness for working with global spatio-temporal socioeconomic data sets. We propose Categorize Trends (CT), an exploratory analysis tool which classifies high dimensional TS data into a taxonomy of behavior types. CT is fundamentally built on Dynamic Time Warping (DTW) [1], employing it to compare TS data to mathematically generated time series representing behaviors. Developed as an exploratory capability, CT a) assigns each trend to its closest behavior type, b) creates behavioral maps that show *spatial patterns* of *temporal behavior*, c) produces scatter plots that show relationships between trend and magnitude, and d) supports *behavioral queries* that filter univariate and multivariate cross tabulations allowing discovery of anomalies, bivariate behaviors, and specialized trends of interest. To illustrate the features of CT, we develop a use case exploring relationships between food scarcity and fertility.

2   Methods

First, we develop an initial set of 10 behavior classes with the help of geospatial analysts that represent practical patterns of interest (Fig. 1). For example, common patterns such as constant, down, up, and oscillating are joined by patterns that might indicate anomalies (e.g., spike-down, spike-up), sudden dramatic shifts (e.g., shift-up, shift-down), recovery (e.g., smile), and so forth. These are only exemplar interpretations and scholars can easily add their own particular patterns of interest. Patterns are mathematical functions such as linear models (down, up), parabolas (smile, frown), sine functions (wave), or step functions (spikes, shifts) that generate synthetic taxonomic trends. For time-variant patterns (spikes, shifts, wave), multiple trends can be generated for a behavior class with each trend having the constructed feature occur at a different point in time.

The next step is to associate time series trends of interest with one or more of these 10 behavior types. We calculate the distance between each normalized trend and the taxonomic types using DTW, a non-linear similarity measure that is robust for comparing limited or noisy trends [1]. This yields a similarity vector of normalized distances to the taxonomic types. For the time-variant patterns with multiple synthetic trends, we select the closest match DTW distance.

Finally, we provide a tool for exploring the taxonomic classification results. We display the results spatially, producing behavioral maps where cartographic colors indicate semantic behavior using the strongest trend associations for each country (Fig. 2). This allows investigators to see how the temporal behaviors of their data spatially arrange. Spatially isolated anomalous behaviors, clusters of
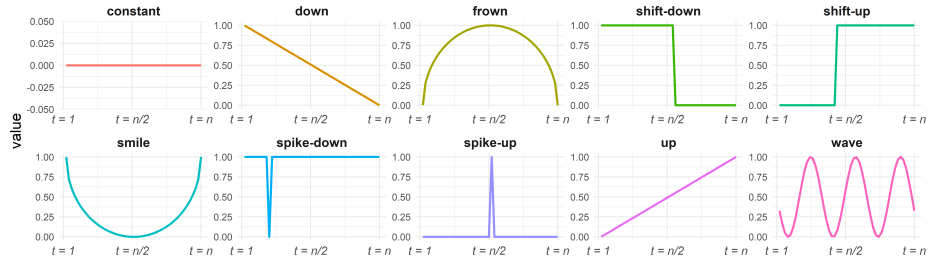
Fig. 1. Taxonomic trend behaviors used in pattern classification

common behaviors, or hypotheses about spatial randomness can be motivated by observing these maps. We also generate a scatter plot of trend magnitude (e.g., mean, maximum, last value) versus trend taxonomy (Fig. 2) to help the practitioner understand what behavioral types may be preferentially low or high in magnitude. Finally, we provide the ability to further query results by country, attribute, and trend to explore themes (e.g., concurrence, divergence, etc.) revealing multivariate relationships of importance.

3   Use Case

The interaction between food scarcity and fertility is of concern to policy makers, health officials, and organizations that focus on human health and welfare, to include the long-term impacts of climate change [3]. Insufficient caloric intake due to regular food scarcity renders women prone to fertility issues such as amenorrhea or temporary loss of fecundity [2, 3, 8]. Using World Bank data, we explore this problem analyzing *Fertility Rate, Total (births per woman)* and *Depth of the Food Deficit (kilo-calories per person per day)* across Africa and Asia from 2001 through 2016. We apply CT in this context to understand major regional trends, uncover joint fertility-deficit behaviors, discover anomalies, and assess relationships between trend shape and magnitude. Investigating these topics allow practitioners to become familiar with the data, narrow the frame of study, and pose new hypotheses.

4   Results

We used the case of fertility and food scarcity as a vehicle for demonstrating the exploratory power of CT, particularly how it can support spatio-temporal reasoning in the opening stages of analysis. Fig. 2 presents spatial behavioral maps and scatter plots from applying CT to the fertility and food deficit data. This uncovers two substantial and spatially clustered trends of *up* and *down* taxonomies for fertility (Fig. 2a). The majority of countries present downward trends which visually appear spatially contiguous throughout most of Africa and

Southern Asia. Upward trend clusters appear largely in north Africa and central and eastern Asia. The remaining taxonomic classifications are spatially disconnected and depart from the primarily upward and downward patterns observed. These include smile (Iran, Fiji, Brunei, Lebanon, and Solomon Islands), wave (Uzbekistan, Azerbaijan, and South Korea), and shift-down (Oman, Tajikistan, and Thailand) serving as interesting candidates for deeper exploration.
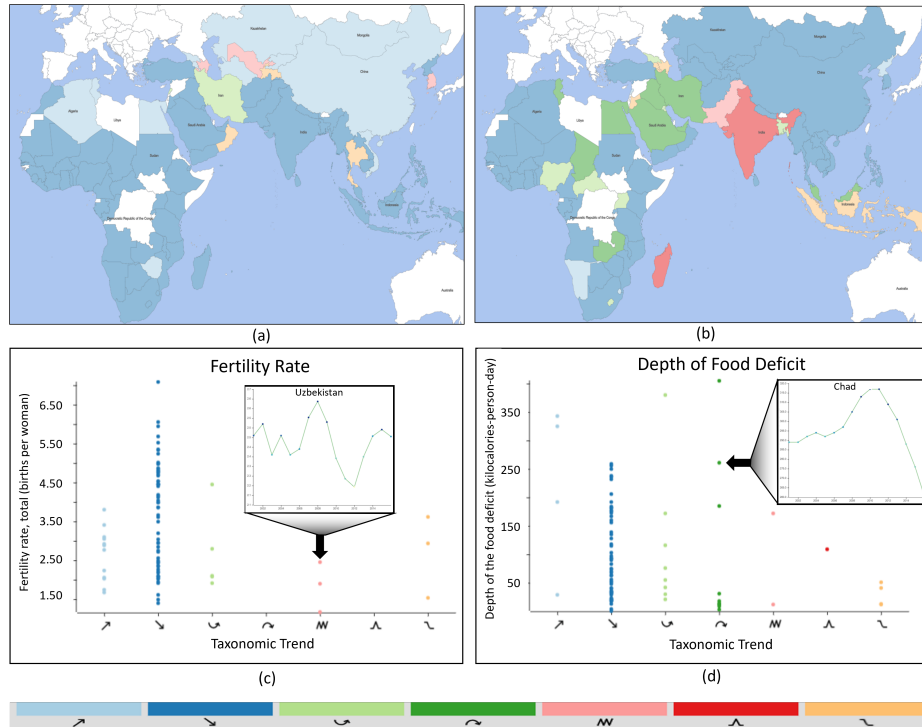

Fig. 2. Spatial mapping of temporal behaviors for (a) fertility and (b) food deficit allows discovery of trends in space and in time on a static map. Scatter plots of taxonomy verses magnitude for (c) fertility and (d) food deficit reveal relationship between taxonomy and magnitude.

Fig. 2b presents results for a slightly more complex food deficit story. Many countries still exhibit a downward trend but other behavioral types, such as smile and frown, are more prominent and appear to spatially cluster. For example, Egypt, Saudi Arabia, Iraq, and Iran form a smile cluster. In central Africa, we notice a concentration of smiles and frowns. Notable outliers both in terms of frequency and spatial isolation are the spike-up countries of India and Madagascar. The four shift-down nations of Azerbaijan, Armenia, Jordan, and Indonesia are spatially scattered.

A quick look at the scatter plots for both fertility (Fig. 2c) and food deficit (Fig. 2d) indicate that the less common and more spatially distributed fertility trends tend to associate with lower fertility rates while other major trends, such as *down*, tend to occur across the spectrum of fertility values. Again, the story is different for food deficits. The scatter plot shows a broader distribution of taxonomic classes compared to fertility, however still contains a large shift-down class. Included in Fig. 2c and Fig. 2d are inset time series graphs for Uzbekistan (wave) and Chad (frown), respectively, displaying the similarity between raw trend and assigned taxonomic classification.

In Table 1, we present the cross tabulation results of fertility and food deficit taxonomic classifications revealing deeper bivariate relationships. About half the countries experienced a simultaneous down trend for both variables. The remaining taxonomic combinations are small and scattered across remaining categories. From here scholars can zero in on uncommon patterns or patterns of theoretical interest. For example, in all cases where fertility rates are up, food deficits are decreasing in one way or another (down, frown, shift-down) with exception of one country. Lebanon is experiencing an up growth in food deficit along with a smile pattern for fertility. This means that a recent (and possibly troubling pattern) of increasing fertility and increasing food deficit is happening and is unique to Lebanon.

Table 1. Cross Tabulation of Trend Classifications

		Depth of Food Deficit							
		down	frown	shift-down	smile	spike-up	up	wave	*Totals*
	down	43	6	2	6	2	3	1	*63*
Fertility Rate	shift-down	3	0	0	0	0	0	0	*3*
	smile	1	1	0	1	0	1	1	*5*
	up	8	2	1	1	0	0	0	*12*
	wave	2	0	1	0	0	0	0	*3*
	Totals	*57*	*9*	*4*	*8*	*2*	*4*	*2*	*86*

5   Conclusion

We have developed an exploratory approach for behaviorally classifying, mapping, and analyzing global time series data. CT allows practitioners to link time series to domain specific behaviors of interest, and then create maps that convey those temporal behaviors spatially. These allow analysts to see major behavioral trends, detect anomalies, and query for particular joint behaviors of interest. As an illustration, we applied CT to the study of fertility and food scarcity. CT was able to effectively sort and map 172 time series into 10 categories quickly bringing into view major trends and interesting anomalies.

In the authors' experience (e.g., [7]) over 25,000 spatio-temporal attributes are known to exist across the holdings of global vendors such as the World Bank and WHO. Our use case study represents just one example of many possible analysis from these large panel sets. Future work will focus on surfacing subsequence taxonomic classifications and creating behavioral profiles for time series containing several distinct patterns. This will allow us to develop additional analytic capabilities aimed at answering more complex questions.

References

1. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: KDD workshop. vol. 10, pp. 359–370. Seattle, WA, USA: (1994)
2. Boutari, C., Pappas, P.D., Mintziori, G., Nigdelis, M.P., Athanasiadis, L., Goulis, D.G., Mantzoros, C.S.: The effect of underweight on female and male reproduction. Metabolism **107**, 154229 (2020)
3. Grace, K., Nagle, N.N.: Using high-resolution remotely sensed data to examine the relationship between agriculture and fertility in mali. The Professional Geographer **67**(4), 641–654 (2015)
4. Guan, X., Huang, C., Liu, G., Meng, X., Liu, Q.: Mapping rice cropping systems in vietnam using an ndvi-based time-series similarity measurement based on dtw distance. Remote Sensing **8**(1), 19 (2016). https://doi.org/10.3390/rs8010019
5. Kim, S., Lee, H., Ko, H., Jeong, S., Byun, H., Oh, K.: Pattern matching trading system based on the dynamic time warping algorithm. Sustainability **10**(12), 4641 (2018). https://doi.org/10.3390/su10124641
6. Mondal, S., Jeganathan, C.: Mountain agriculture extraction from time-series modis ndvi using dynamic time warping technique. International Journal of Remote Sensing **39**(11), 3679–3704 (2018). https://doi.org/10.1080/01431161.2018.1444289
7. Piburn, J., Stewart, R., Myers, A., Sorokine, A., Axley, E., Anderson, D., Burdette, J., Biddle, C., Hohl, A., Eberle, R., et al.: The world spatiotemporal analytics and mapping project (wstamp): Further progress in discovering, exploring, and mapping spatiotemporal patterns across the world's largest open source data sets. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences **4** (2017)
8. Silvestris, E., Lovero, D., Palmirotta, R.: Nutrition and female fertility: an interdependent correlation. Frontiers in endocrinology **10**, 346 (2019)
9. Tsinaslanidis, P., Guijarro, F.: What makes trading strategies based on chart pattern recognition profitable? Expert Systems **38**(5) (2021). https://doi.org/10.1111/exsy.12596