Using Self-Organizing Maps to analyse spatial temporal diffusion of infectious diseases

Ellen-Wien Augustijn-Beckers[1], Raul Zurita-Milla[1] , Nicoline van der Maas[2]

[1]University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, 7500 AA Enschede, The Netherlands
Telephone: +31 (0)53 4874 414
Fax: +31 (0)53 4874335
Email: p.w.m.augustijn@utwente.nl; r.zurita-milla@utwente.nl

[2] National Institute for Public Health and the Environment, Ministry of Health, Welfare and Sport, Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven
Email: nicoline.van.der.maas@rivm.nl

Abstract

The amount of data available to study spatial temporal patterns in complex phenomena like disease diffusion has vastly increased over the years. Extraction of useful information is often hindered by the complexity of the patterns. In this paper we propose a method of training a SOM on a complete dataset and mapping back subsets of this dataset to allow comparison between a complete dataset and its subcomponents. We illustrate this method using a temporal case study for epidemic waves of Measles in Iceland and a case study comparing spatial patterns of different age groups for Pertussis in the Netherlands.

Keywords: Spatiotemporal analytics, Self-Organizing Maps (SOMs), complex patterns, disease.

1. Introduction

Infectious diseases like influenza, Ebola, measles etc. still form an important thread to human health and pose an economic and social burden on our society. They may occur in the form of a regular outbreak but also as epidemics or pandemics, affecting multiple countries or continents. Infectious diseases that were almost extinguished (like measles and pertussis) are re-emerging because of resistance to vaccines and new genetic change in the pathogen reservoirs. These developments make constant monitoring of disease outbreaks necessary. For instance, to test the effectiveness of interventions and gain understanding of the mechanisms behind the diffusion processes.

Due to better surveillance techniques and global awareness of the importance of constant monitoring for many infectious diseases, datasets based on long-term data collections at suitable spatial and temporal resolutions for spatial-temporal analysis are becoming available. Normally these datasets are based on daily notifications that are spatially aggregated to the level of postal code or health unit.

Questions posed by health authorities about the spread of infectious diseases can be split into a number of categories including temporal trends and fluctuations, spatial-temporal diffusion patterns (expansion versus relocation diffusion, where expansion diffusion can be divided into contagious spread and hierarchical diffusion (Cliff et al., 1981)), effectiveness of interventions (closing schools, limiting travel, isolation of patients) and correlations with other environmental characteristics. Self-Organizing Maps (SOMs) are a good method for studying spatial-temporal diffusion patterns, as they can be used on large datasets, allow for missing observations and can be applied both in space and time.

In this abstract we discuss two examples of applying SOMs for the analysis of infectious diseases. In spatial-epidemiology, SOMs are mostly used to study multivariate patterns (Wang et al., 2011, Basara and Yuan, 2008, Koua and Kraak, 2004) but they also enable the integrated analysis of different spatial-temporal diffusion. The first example will focus on analysing a long time series of measles data for Iceland (Augustijn and Zurita-Milla, 2013), and the second example will discuss the use of SOMs to analyse pertussis data for the Netherlands. For this latter example we will further extend the earlier applied methods by investigating differences between spatial diffusion patterns for individual age categories.

2. Application of SOMs for disease analysis

SOMs are a type of unsupervised artificial neural network introduced by Kohonen (2001). They are typically used to cluster high dimensional data by projecting it onto a low-dimensional lattice, which consists of neurons that are trained iteratively to extract patterns from the input data. These patterns are generalizations of the input data and are referred to as codebook vectors. The application of SOMs follows the following steps: (1) training of the SOMs based on the input data, (2) secondary clustering based on either visual interpretation or using a secondary clustering method, (3) mapping of the data onto the trained SOM.

The extracted patterns depend on the way the data is organized. This applies to both the training data and the data that is mapped back onto the trained SOM. In this study we used the Kohonen R package (Kohonen, 2001, Wehrens and Buydens, 2007) to train several SOMs as discussed in the following subsections.

2.1 Measles in Iceland

For the measles example we used a long-term monthly dataset (1946-1970) reporting measles occurrences for 50 medical districts in Iceland. We compared the eight epidemic outbreaks that were captured in this dataset by identifying similarity in spatial-temporal diffusion patterns among these outbreaks.

The input data for training the SOM was organized in three different ways: Space x Time (SxT), Time x Space (TxS) and Space over Wave (SxW) allowing to identify similarity (synchrony) between spatial location and similarity between epidemics based on temporal diffusion pattern. Although the training of the SOM was based on the complete dataset we used the technique of mapping back only one single epidemic outbreak (wave) to be able to compare the spatial temporal patterns of different outbreaks.

Although the method was successfully applied, the fact that Iceland is a very isolated country, with a limited number of inhabitants, and restricted interactions with neighbouring countries asked for further research.

2.2 Pertussis in the Netherlands

The pertussis dataset of the Netherlands consists of 18 years of daily surveillance data at the spatial aggregation level of the postal code. Before performing the SOM, this data was spatially aggregated to the level of the health unit (58 units) and temporally to weeks.

In this case, we organized the training data as a SxT matrix but we split the data into age categories before mapping it back. The age categories used are 0-4 years old, 4-12 years old, 12-18 years old, 18-24, 25-45 and older than 45. Each of these categories corresponds to different activity groups, for children based on the type of school they attend (pre-school age, elementary school, secondary school, university) and adults were split into a group with, and a group without children of school age.

Compared to Iceland, the Netherlands is a very highly populated country with six urban systems (spatial agglomerations) that consist of a number of tightly interlinked cities and other communities. Because of this high level of integration (home – work commuting), it is of interest to investigate if cities belonging to the same urban system are more synchronised compared to cities that belong to different urban systems.

As a result, the following SOM based analyses were performed:

- Identification of differences in spatial temporal patterns between different age groups
- Identification of the level of synchrony between areas that belong to the same urban system.

3. Acknowledgements

The Dutch pertussis data was made available by the National Institute for Public Health and the Environment (RIVM), Ministry of Health, Welfare and Sport.

4. References

AUGUSTIJN, E.-W. & ZURITA-MILLA, R. 2013. Self-organizing maps as an approach to exploring spatiotemporal diffusion patterns. *International Journal of Health Geographics,* 12**,** 60.

BASARA, H. & YUAN, M. 2008. Community health assessment using self-organizing maps and geographic information systems. *International Journal of Health Geographics,* 7**,** 67.

CLIFF, A. D., HAGGETT, P., ORD, J. K. & VERSEY, G. R. 1981. *Spatial Diffusion, An Historical Geography of Epidemics in an Island Community*, Cambridge, USA, Press Syndicate of the University of Cambridge.

KOHONEN, T. 2001. *Self-Organizing Maps*, New York, USA: Springer-Verlag.

KOUA, E. & KRAAK, M.-J. 2004. Geovisualization to support the exploration of large health and demographic survey data. *International Journal of Health Geographics,* 3**,** 12.

WANG, J.-F., GUO, Y.-S., CHRISTAKOS, G., YANG, W.-Z., LIAO, Y.-L., LI, Z.-J., LI, X.-Z., LIA, S.-J. & CHEN, H.-Y. 2011. Hand, foot and mouth disease: spatiotemporal transmission and climate. *International Journal of Health Geographics,* 10**,** 1-10.

WEHRENS, R. & BUYDENS, L. M. C. 2007. Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software,* 21**,** 1-19.