Habitat International 45 (2015) 92–98

Contents lists available at ScienceDirect

Habitat International

journal homepage: www.elsevier.com/locate/habitatint

Visualizing urban social change with Self-Organizing Maps: Toronto neighbourhoods, 1996–2006

Andrew C.-D. Lee, Claus Rinner*

Department of Geography, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B2K3, Canada

ARTICLE INFO

Article history:
Available online 26 July 2014

Keywords:
Geographic information system
Neighbourhood trajectory
Self-organizing map
Socio-economic status
Space-time-attribute data
Urban change

ABSTRACT

Change in the socio-economic status of urban neighbourhoods is a complex phenomenon with multiple space, time, and attribute dimensions. The objective of this research was to explore the use of a Self-Organizing Map (SOM) to visualize patterns of urban social change. In a case study, we collected, organized, and joined data from the 1996, 2001, and 2006 Canadian Census for the City of Toronto. Urban neighbourhoods were represented by Census tracts. The SOM translates multi-dimensional data into two-dimensional graphical patterns of neighbourhood socio-economic status. These were associated with patterns in geographic space. Spatio-temporal change was represented by trajectories in the SOM. The study identified trends of decreasing neighbourhood diversity and shifts in the dynamics of urban social change in Toronto. The proposed methodology could assist with strategic planning of urban development and efficient resource allocation that fits with local needs.

Introduction

Change in the socio-economic fabric of urban neighbourhoods results from the interactions between social, political, and economic activities (Hulchanski, 2007). Observing, analyzing, and understanding urban change can be difficult due to the complexity of the phenomenon and the data suitable to represent it. Census data contain information regarding the socio-economic status of the population in small geographic areas at selected times. Geographic visualization of Census data has become an important analytical tool for stakeholders such as policy-makers and demographers (Skupin & Hagelman, 2005). A common approach to examining urban change is to analyze space-time-attribute data using Geographic Information Systems (GIS) (Kauko, 2009; Spielman & Thill, 2008; Timmermans, Arentze, & Joh, 2002). Gorricha and Lobo (2012) noted that the reduction of high-dimensional geographic data to groups with similar characteristics (clustering) can be effectively supported by visual search for patterns.

In this research, our objective was to explore the use of a Self-Organizing Maps (SOM) to visualize the changes in urban social patterns over time. The following sections outline the research context in terms of visualizing spatio-temporal data using the SOM neural network approach. We then describe the case study data and software used. The results are organized by the elements of SOM output, including the representation of individual neighbourhoods by neurons on the SOM plane and the characteristics of their socio-economic trajectories. The paper concludes with a discussion of results and outlook on future work.

Research context

Approaches to visualizing spatio-temporal data

Since the 1970s, space-time measures have received attention due to the need to construct empirical spatial models in the context of urban and regional analyses (Martin & Oeppen, 1975). Some social scientists have implemented mathematical models and geographic visualization for analyzing space-time patterns (Chen, MacEachren, & Guo, 2008; Kyriakidis & Journel, 1999; Spielman & Thill, 2008). Time is an irreversible variable that linearly moves forward and inevitably influences everything. Within a period of time, events may be predictably recurring with multiple cycles and forming hierarchical structures, which can overlap and interact with each other (Andrienko et al., 2010). In fact, analysts often need to consider the interactions between space, time, and attributes when they attempt to understand variations of an area over time. Andrienko et al. (2010) propose two approaches to visualizing such data: space-in-time and time-in-space displays, the choice among which depends on the focus of analysis.

* Corresponding author. Tel.: +1 416 979 5000x2686.
E-mail address: crinner@ryerson.ca (C. Rinner).

http://dx.doi.org/10.1016/j.habitatint.2014.06.027
0197-3975/

Due to the complexity of spatio-temporal interactions, the sense of space and time often requires the experience and knowledge of the analysts (Andrienko, Andrienko, Dykes, Fabrikant, & Wachowicz, 2008). Existing approaches to pattern identification and interpretation can be distinguished as computationally and visually led methods (Chen et al., 2008; Guo, Gahegan, MacEachren, & Zhou, 2005). Computational methods can find specific patterns within large volumes of data, but the interpretation ability is limited. In contrast, visualization methods can propose explanations and generate hypotheses for further analysis, and also recognize and present patterns in an easy-to-understand form (Guo et al., 2005; Nöllenburg, 2007). Visualization can be difficult though, when the size and complexity of datasets are beyond the cognitive abilities of the analyst (Koua, 2003). The traditional geographic visualization method is limited in presenting multi-dimensional datasets. Hence, complex geographic data are often visually examined by creating numerous maps for comparison (Skupin & Hagelman, 2005).

Developing new visualization methods that combine the strengths of computational and visual methods is required (Guo et al., 2005), and such development is one of the major research areas in the spatial analysis discipline (Koua & Kraak, 2004). The developed approaches usually include a combination of different methods (Guo et al., 2005). Recently, combinations of cluster analysis and visualization were developed for effectively extracting meaningful information from large geospatial datasets, and explicitly representing underlying structures and processes (Gahegan, 2000; Miller & Han, 2001; Schreck, Bernard, Landesberger, & Kohlhammer, 2009). The fundamental rationale behind the cluster visualization process is to organize objects into groups depending on certain shared characteristics (Skupin & Agarwal, 2008), and to visualize the result in a two-dimensional space (Schreck et al., 2009; Skupin & Agarwal, 2008). For instance, dendrograms are one of the most widely known clustering techniques that visually present clustering results on a graph (Skupin & Agarwal, 2008). Such clustering results implicitly synthesize visual representations (Schreck et al., 2009). A visual cluster output that can explicitly present spatio-temporal changes is essential.

Self-Organizing Maps for space-time-attribute analysis

Artificial neural networks are one of many emerging computing techniques that have been actively studied over the last three decades (Takatsuka, 2001). They are inspired by ideas from neuroscience, where a sophisticated "computing" system can be constructed from a network of simple processing units (Takatsuka, 2001). Most neural networks exploit their learning capabilities to obtain appropriate connections from input data. Such specialized processes are based on the interconnectivity between processing units.

The SOM is a neural network that uses an unsupervised training algorithm and passes through a process of self-organization, which is a competitive learning method that reduces data dimensions (Kohonen, 1998). The process combines clustering and projection steps in order to present amenable visual analysis (Vesanto, 1999). The SOM output is organized to map units that represent the vectorial topology of the original data (Kohonen, 1998). These map units are called neurons.

The SOM process iteratively compares the output neurons against the input data, and finds the best match. The best match is defined as the neuron that has the shortest Euclidean distance to the input observation. The best matched neuron is also called the best matched unit. Once the best matched unit is recognized, the adjacent neurons of the best matched unit will also be modified by a predefined neighbourhood function (e.g. Gaussian function) in order to present the distance between the input observations in vector space (Spielman & Thill, 2008). This property allows the SOM to bond similar input data vectors on the neurons. Similar neurons are mapped close together on a pre-defined grid layer, which is called the SOM plane (Andrienko et al., 2010; Kohonen, 1998; Koua, 2003; Yan & Thill, 2008). Therefore, SOM viewers are able to judge the similarity of data based on their proximity (Spielman & Thill, 2008).

The SOM method is applicable to different data types that can be represented by vectors (Andrienko et al., 2010). For instance, Deboeck (1998) applied a SOM for exploring global financial data and understanding their trends and patterns, while Páez, Scott, Potoglou, Kanaroglou, and Newbold (2007) utilized a SOM to investigate the spatial and demographic variability in elderly travel behaviours in the Hamilton, Ontario, area. Gorricha and Lobo (2012) utilize colour coding to combine the SOM output with cartographic representation of the underlying geospatial input data in order to enhance their clustering results for Census data for the Lisbon metropolitan area. Another urban social analysis using SOM was conducted by Ju, Lam, and Chen (2006) for ten socio-economic variables by Census block groups of New Orleans.

The SOM method has also been applied for space-time-attribute analysis. Because of the ability to deal with complex multi-dimensional data, there is an increase in the number of space-time-attribute approaches implementing the SOM method. For example, Hsu and Li (2010) explored long-term spatiotemporal characteristics of precipitation data to improve water resource management and reduce impacts of climate-induced disasters. Crime analysts also showed an interest in the space-time-attribute approach in order to understand crime patterns within policing boundaries (Gorr, Olligschlaeger, & Thompson, 2003). Andrienko et al. (2010) demonstrated applying the SOM method to identify patterns in a 41-year time series of seven crime rate attributes in the states of the USA. According to the result, Andrienko and colleagues found numerous spatiotemporal patterns of crime rates throughout the USA, including an increase in crime rates (except murders) in the southeastern states during the period of 1960–2000 with a peak between 2000 and 2010.

Skupin and Hagelman (2005) extended the SOM technique to visualize spatiotemporal change in the State of Texas, based on Census data between 1980 and 2000. The SOM neurons were connected based on each temporal vertex on the SOM plane to form trajectories. With the trajectories, the result allows analysts to visualize the temporal changes of the selected attributes. The trajectories assisted analysts to visually identify the socio-economic transitions of different counties in Texas between 1980 and 2000. A similar approach was chosen by Delmelle, Thill, Furuseth, and Ludden (2013), who analyzed neighbourhood quality of life in the city of Charlotte, North Carolina. They analyzed 17 quality of life indicators from a biennial survey between 2000 and 2010, using clustering of SOM output nodes and trajectories connecting the output nodes for the same neighbourhood across time.

Data and software

In this study, data from the Canadian Census of the population at the Census tract (CT) level were used. A CT is a small and relatively stable area that contains a population of 2500 to 8000. In Canada, socio-economic data for CTs are available in large urban centres with an urban core population of 50,000 or more, and are often used to represent city neighbourhoods (Statistics Canada, 2011; Larsen & Gilliland, 2008).

To measure socio-economic change in Toronto, 15 variables from the Census categories of population, household/family,