Computers, Environment and Urban Systems 61 (2017) 1–12

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Contextual mapping: Visualization of high-dimensional spatial patterns in a single geo-map

Vahid Moosavi

Chair for Computer Aided Architectural Design, Institute of Technology in Architecture, Faculty of Architecture, ETH Zurich, HPZ F/John-von-Neumann-Weg 9, CH-8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form 17 August 2016
Accepted 17 August 2016
Available online 14 October 2016

Keywords:
Multivariate geo-mapping
Spatial clustering
Pattern recognition
Self-organizing map
Urban planning

ABSTRACT

In this study, we proposed a generic methodology for combining high-dimensional spatial data to identify and visualize the hidden spatial patterns in a single-layer geo-map. By using the less explored one-dimensional self-organizing maps, we showed how the high-dimensional data can be transformed into a spectrum of one-dimensional ordered numbers. These numbers (codes) can index a high-dimensional space with the important property that similar indices refer to similar high-dimensional contexts. Thus, the high-dimensional vectors will be attributed to single numbers, and this one-dimensional output can be easily rendered as a new single data layer in the original geographic map. As a result, it simultaneously identifies the main spatial clusters and visualizes the high-dimensional correlations (if any) in a single geographic map. Further, because the output of the proposed method is a set of ordered indices, there is no need to define a fixed number of clusters in advance. Because these composite spatial layers are identified on the basis of the selected context (i.e., the selected features or aspects of the spatial phenomena), they are called *contextual maps*.
Finally, we showed the results of applying the proposed methodology to several synthetic and real-world data sets.

1. Introduction

With the current rapid growth in the amount of digital data, we must address the challenge of finding appropriate techniques to harness the power of these data streams. For example, in many cities across the world, no longer does anyone lack access to digital spatial maps; instead, the current challenge is, considering the amount and diversity of these digital data regarding different aspects of the cities, how one can picture his/her own map of the space as a combination of several factors of interest.

Toward this direction, there have been several interesting cases such as peoplemaps[1] or Livehood projects (Cranshaw, Schwartz, Hong, & Sadeh, 2012), which are explorations and mapping of activities within cities based on data available from online social networks. One of the cases most similar to our work is a project called Whereabout,[2] where by applying the K-means data-clustering algorithm to a collection of spatial data consisting of >200 different aspects of each ward in the city of London, a fixed number of groups were created by grouping based on informational similarities (not physical locations). Then, on top of the classical map of London, people get an impression of different regions on the basis of their similarities in all of these categories. In a similar manner, but only based on demographic information, a new coding system of London called LOAC was developed (Longley & Singleton, 2014).

The classical clustering algorithms divide the high-dimensional data space into a predetermined number of groups, where each will be given a label (usually an arbitrary number). Then, these cluster labels attributed to each spatial data point can be visualized on the geographic map with a specified color code. However, despite the fact that standard clustering methods such as K-means are easy to use, they have some limitations in the domain of spatial pattern recognition. One of the main problems is that they divide the space into a small number of categories. Instead, it would be preferred to have a continuous and smooth changing pattern on top of the high-dimensional data. Further, one needs to select the number of clusters in advance, which is a critical decision (Tibshirani, Walther, & Hastie, 2001). In addition, in the context of spatial clustering, because the cluster labels are not ordered according to their high-dimensional similarities, the colored visualization of clusters in the geographic map is not directly helpful. Therefore, similar colors in a clustered geo-map do not necessarily refer to similar high-dimensional patterns. As a result, increasing the number of clusters with different colors may result in final spatial visualizations that are not helpful, but having too few clusters produces results that are too aggregated. One current solution to this problem is to create an RGB (red, green, blue) pattern after data clustering by reducing the high-dimensional vectors of the cluster centers to their first three principal components (Mahinthakumar, Hoffman, Hargrove, & Karonis, 1999). However, in

E-mail address: svm@arch.ethz.ch.
[1] http://peoplemaps.org/.
[2] http://whereaboutslondon.org/#/about.

http://dx.doi.org/10.1016/j.compenvurbsys.2016.08.005

addition to losing some information (by selecting only three principal components), the color interpretations will need an additional step.

The main hypothesis of this study is that if we find a method to sort the clusters in a way such that similar cluster indices refer to similar contexts (i.e., similar high-dimensional patterns), we can make a direct projection from high-dimensional spatial data to a one-dimensional vector and visualize the high-dimensional patterns in the geographical maps using a simple color spectrum. In this manner, by having many indices instead of dividing the high-dimensional data into a few distinct groups, one can create a spectrum of high-dimensional patterns that are visualized with a colored spectrum on spatial maps. Because the high-dimensional patterns would change gradually, this would also solve the problem of distinct cluster borders and the fixed number of clusters. As we show in Section 2, our proposed approach can be discussed from the viewpoint of dimensionality reduction and manifold learning (Bengio, Courville, & Vincent, 2013), where one of the best methods that satisfies these requirements is self-organizing maps (SOMs) (Kohonen, 2013).

2. SOMs in the domain of spatial analysis

SOM is a general-purpose machine-learning method that offers interesting solutions to different data-driven modeling tasks (Kohonen, 2013).

SOM is a nonlinear space transformation method that tries to preserve the topology of high-dimensional data, while transforming them into a low-dimensional space. This means that SOM projects the high-dimensional data points to a lower-dimensional space (normally a two-dimensional grid) in a manner such that neighboring objects in high-dimensional space remain neighbors in low-dimensional space. This topology-preserving transformation unfolds the nonlinear and high-dimensional patterns into a low-dimensional space that can be easily visualized.

Nevertheless, a major difference between SOM and other data-unfolding and dimensionality reduction methods such as locally linear embedding (Roweis & Saul, 2000), complete isometric feature mapping or ISOMAP (Tenenbaum, De Silva, & Langford, 2000) and *t*-distributed stochastic neighbor embedding, known as *t*-SNE (Van der Maaten & Hinton, 2008) is that it creates an abstraction of the data into new prototypes, while in typical dimensionality reduction methods, there is always a one-to-one relationship between all the observed points in the high- and low-dimensional space. In SOM algorithm, these identified abstract prototypes (usually called nodes or codebooks) are essential elements for the pattern recognition and data reduction tasks such as clustering. These nodes have a dual representation, including a low-dimensional vector, showing the location of the node in the lower-dimensional space, and a high-dimensional weight vector, which has the same dimensionality as the original high-dimensional data. Therefore, these nodes, as distributed models of the training data, can be used separately in different modeling problems. This property of SOM makes it very attractive for many tasks such as data visualization, function approximation, and data clustering in general. In the domain of spatial analysis, SOM has been used in several applications (Delmelle, Thill, Furuseth, & Ludden, 2013; Frenkel, Bendit, & Kaplan, 2013; Agarwal & Skupin, 2008; Skupin & Esperbé, 2011; Wang, Biggs, & Skupin, 2013 and Arribas-Bel, Nijkamp, & Scholten, 2011; Spielman & Thill, 2008) and is well-known as a tool for visual data mining and exploration of high dimensional spatial interactions (Yan & Thill, 2009).

Because data clustering is an exploratory activity, the high-dimensional maps resulting from the clustering of high-dimensional vectors using SOM are commonly visualized on two-dimensional colored maps known as component planes (see Fig. 3 for an example). However, in the context of spatial clustering, there is normally an extra constraint on projecting the final outputs of the pattern recognition algorithms onto the original spatial map. Considering this requirement, the main problem of classical SOM is that it loses the spatial index of data that are not part of the training data (Bação, Lobo, & Painho, 2005).

Therefore, one of the main concerns of spatial clustering is how to consider the effect of spatial coordinates of data points alongside the other attributes (Bação et al., 2005 and Hagenauer & Helbich, 2013).

Spatial autocorrelation is one of the underlying concepts in spatial data modeling, which states that physically nearby objects are more likely to exhibit similar properties (Tobler, 1970).

To address this issue, there are two modifications to the original SOM. The first approach is to consider the similarity of spatial objects as a weighted sum of similarity between high-dimensional attributes and their physical proximity. However, since spatial coordinates are not semantically comparable with other attributes, this approach is not widely accepted (Bação et al., 2005). The second approach leads to a class of spatial variants of SOM such as GeoSOM (Bação et al., 2005; Henriques, Bacao, & Lobo, 2012), where, the algorithm forces the training data that spatially similar observations are placed in similar regions of the low-dimensional map of SOM. Therefore, spatial coordinates and spatial attributes are contributing next to each other, but not at the same time in one single distance measure.

A more recent method in this approach is contextual neural gas (CNG) (Hagenauer & Helbich, 2013), which is based on similar idea to GeoSOM, but implemented in the context of the neural gas (NG) algorithm (Martinetz & Schulten, 1991). The NG algorithm is a modified version of SOM that unlike SOM there is no defined low dimensional grid and the nodes are dynamically distributed in the high-dimensional input space. In the case of CNG, the geographical map is used as the lower-dimensional representation.

The main contribution of these two spatial variants of SOM (i.e., GeoSOM and CNG) is that, to an extent, they replace the original synthetic topology of SOM with a spatial map with the cost of n based method discussed above, the final two-dimensional SOM grid can be used as a bivariate color code (Guo, Gahegan, MacEachren, & Zhou, 2005). Although a two-dimensional SOM performs better than the PCA-based map coloring method, which is a linear dimensionality reduction approach, because we are dealing with high-dimensional data, we need another diagram to connect these bivariate color codes of clusters on top of the SOM grid to show the characteristics of the clusters in terms of their high-dimensional vectors. As a result, one needs to select a small number of clusters for better visualization in most of these applications.

As an alternative approach to those mentioned above, the principle idea of this study is to view the problem of spatial clustering from the perspective of manifold learning and dimensionality reduction (Bengio et al., 2013). In the context of spatial pattern recognition, this implies that if there exist some underlying spatial similarities in high-dimensional data that are not easy to track in the original spatial maps, there should be an appropriate manner of encoding the data from high-dimensional space to lower-dimensional codes (specifically to a one-dimensional vector) while preserving the patterns in the encoded data. These low-dimensional codes should index a high-dimensional state space with the important property that similar regions in the high-dimensional state space receive similar codes. Specifically, if we could encode the high-dimensional vectors as a single-dimensional code, we can treat these codes as a type of numerical value, and they can be treated as a single layer of spatial data in the same way that, for example, we can render the surface temperature in a geographic map. Therefore, if there are spatial patterns in the high-dimensional data, one can quickly see them visualized in the geographical maps. This will solve the abovementioned problems of the current spatial clustering approaches.

In addition, because we transform the high-dimensional space into single-dimensional numbers, these numbers can be seen as abstractions of those high-dimensional spaces that they refer to. Therefore, as we will show in the following sections, one can combine the results of several clustering steps in a systematic and hierarchical manner.

Toward this goal, by using the less explored one-dimensional SOM in a different way than usual, we show in this study how the high-dimensional data can be transformed into a spectrum of one-dimensional ordered numbers (codes) called contextual numbers (Moosavi, 2014).

In Section 3, we present the proposed methodology; this is followed by various experimental results, including a synthetic data set and two real-world data sets from London and New York City. Further, we discuss the possibility of combining the results of different clustering steps using their contextual numbers, which is beyond the mere visualization function of SOM in spatial applications.

3. One-dimensional SOMs and spatial clustering

In this section, we assume that the reader is familiar with the original SOM algorithm. Therefore, we skip its re-explanation here and refer the reader to Kohonen (2001) for details regarding the training process.

We instead present how one can project high-dimensional spatial data onto geographical maps while preserving the high-dimensional correlations by using the less explored one-dimensional SOM.

We consider the training data set $X=\{x_i, \ldots, x_M\}$ as a set of M points in an n-dimensional space $x_i \in R^n, i=1,n,M$.

Assuming a low-dimensional grid with K nodes (or cells), we have a set of indices $y_j, j=\{1, \ldots, K\}$, each with an attached high-dimensional weight vector with the same dimensionality as the input data, $w_j \in R^n$.

The goal of the algorithm is for similar data points in the high-dimensional space to be given similar indices.

The final output of a trained SOM can be visualized in different ways. One of the most common ways to represent the nonlinear patterns among different dimensions is through a two-dimensional map of the trained SOM called a component plane (Vesanto, 1999). A component plane is similar to a geographical map with fixed coordinates. Because each node in the trained SOM has a fixed coordinate and a high-dimensional weight vector, one can render different dimensions of the weight vectors simultaneously in this new coordinate system. As a result, this map unfolds the interrelationships between high-dimensional features. For example, suppose that we have a four-dimensional data set with the following interrelations between different dimensions: x_1 and x_4 are independent random variables with a uniform distribution from 0 to 1, $x_2 = -2 \cdot x_1$, and $x_3 = (x_1)^4 + (x_1)^2$.

Fig. 1 shows how a SOM with a two-dimensional grid can visualize these relationships between four dimensions. Here, the colors indicate the values of each variable. As we expected, there is a negative and linear correlation between the first two variables and a nonlinear relationship between the first and third dimensions. Further, as we also expected, the pattern of change in the values of dimension four is orthogonal to those of the other dimensions, which indicates its independence of the other dimensions.

This is the generic use of SOM, and the same methodology can be applied to spatial data points. However, as discussed in Section 2, although these visualized nonlinear patterns between several dimensions are made possible by the SOM algorithm, what is missing in the domain of spatial analysis is *a sense of space*. To highlight this issue, we use an example from a real spatial data set. The data set is taken from a collection of spatial data in the city of London. In this specific example, each data point shows the average percentage of 16 different age groups within each spatial region, which is called a ward. Fig. 2 shows the spatial distribution of these aspects in London.

After training a classical SOM on this data set, Fig. 3 shows the correlations between different age groups. As shown, there are three specific and logical patterns in these 16 dimensions. The first is related to the age groups < 17 years and that of 45–59 years. This group can be families living together with their children. The second group consists of regions with highlighted age groups from 19 to 44 years. This group, which has an opposite pattern from that of the previous group, can be called individual young workers with no children. The last group consists of older groups of age 60 years and above.

An important factor in these high-dimensional graphs (shown in Fig. 3) is that these three identified age groups are, to a large extent, nonoverlapping. Therefore, in terms of spatial analysis, this indicates that the regions corresponding to these three groups might be spatially segregated as well. However, neither of these two types of visualization is completely helpful for spatial analysts and urban planners. On the one hand, in classical geographical maps, it is not easy to find high-dimensional patters. On the other hand, SOM visualization shows the high-dimensional patterns with the cost of losing the spatial distributions of each pattern.

The principal idea of this study is to find a way to take advantage of both spaces (i.e., the lower-dimensional map of SOM and geographical mapping at the same time) by inverting the usual way of rendering the SOM using a one-dimensional SOM.

Traditionally, in applications of the SOM algorithm, the final index of the trained SOM will not be used directly as a numerical value, but instead through its assigned weight vector.

However, in contrast to the current method of using SOM indices, if we design these indices such that they form a new one-dimensional space of indices, they can be used directly as some sort of numerical values (Moosavi, 2014).

If we use a one-dimensional SOM, although we cannot visualize the component planes as usual (such as Figs. 1 and 3), the final index of an SOM with K nodes creates an ordered set such that numerically similar codes are more similar based on their high-dimensional weight vectors. More formally, we will have the following condition:

If $\|y_i - y_j\| < \|y_i - y_k\|$, then $\|w_i - w_j\| < \|w_i - w_k\|$, where $1 \le y_i \le K$ is a scalar value. While this topology preservation is not mathematically proven for an SOM with two-dimensional or higher topologies, it is proved to be valid for one-dimensional SOMs (Erwin, Obermayer, & Schulten, 1992; Cheng, 1997; Flanagan, 2001).

However, as a limiting factor to this change in the topology of SOM it should be noted that in general, having higher grid dimensions or a more-connected neighborhood topology in the SOM network can improve the performance and quality of the trained SOM in terms of quantization error and topology preservation. Therefore, because we are strictly using a one-dimensional SOM, we could, in principle, assume a trade-off between dimensionality reduction and detail preservation. We will discuss this issue further in Section 6.

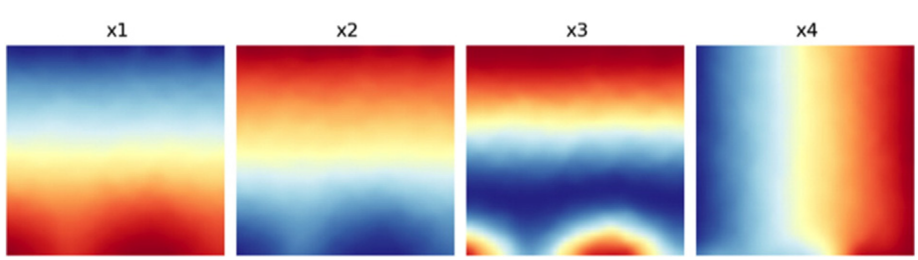

Fig. 1. Typical component plane visualization of SOM for synthetic four-dimensional data.

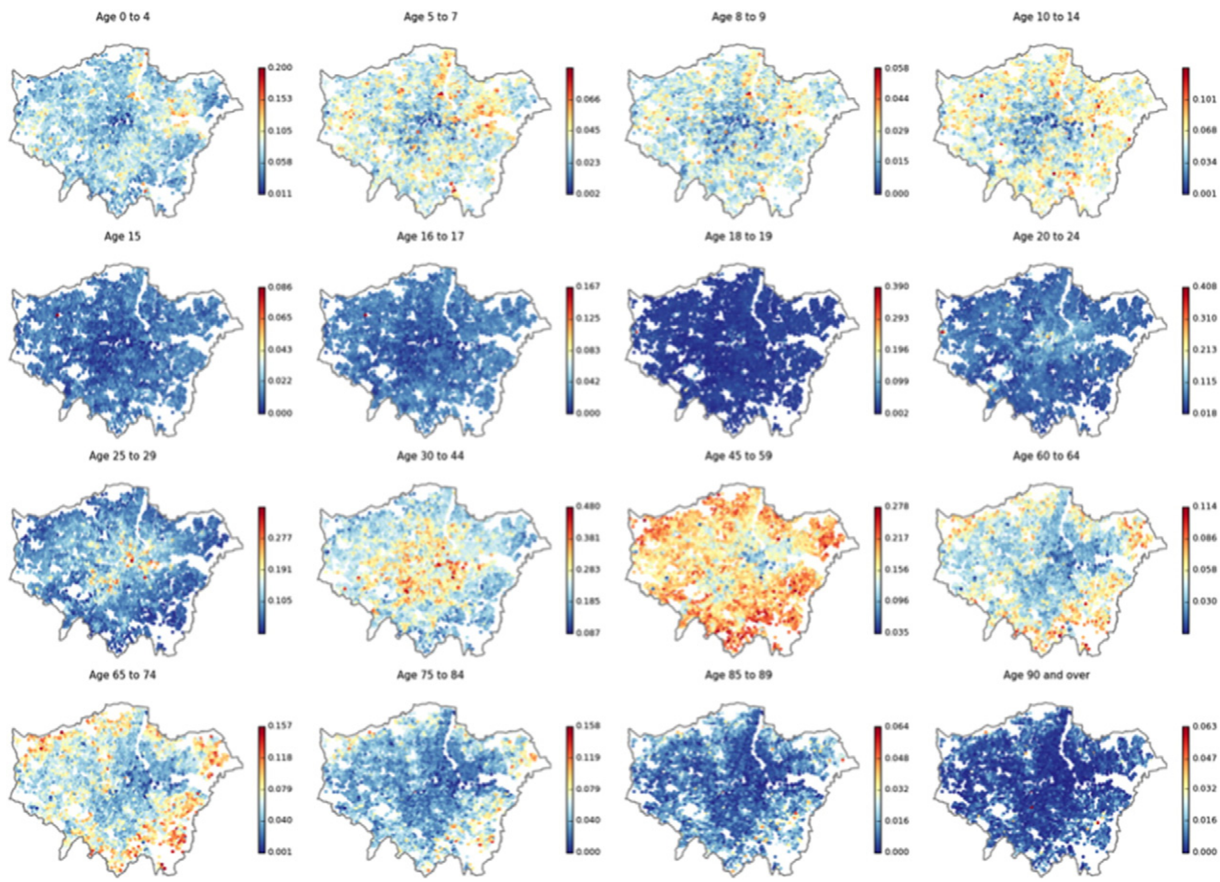

Fig. 2. Spatial percentages of the population in different age groups in London.

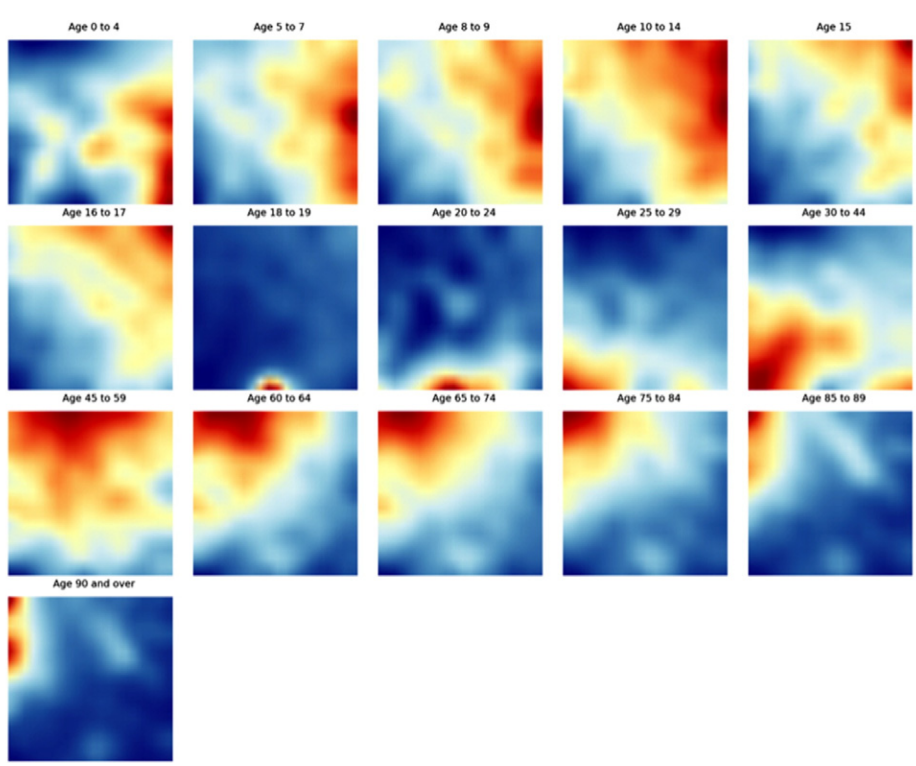

Fig. 3. Interrelationship between the percentages of population in different age groups in London.

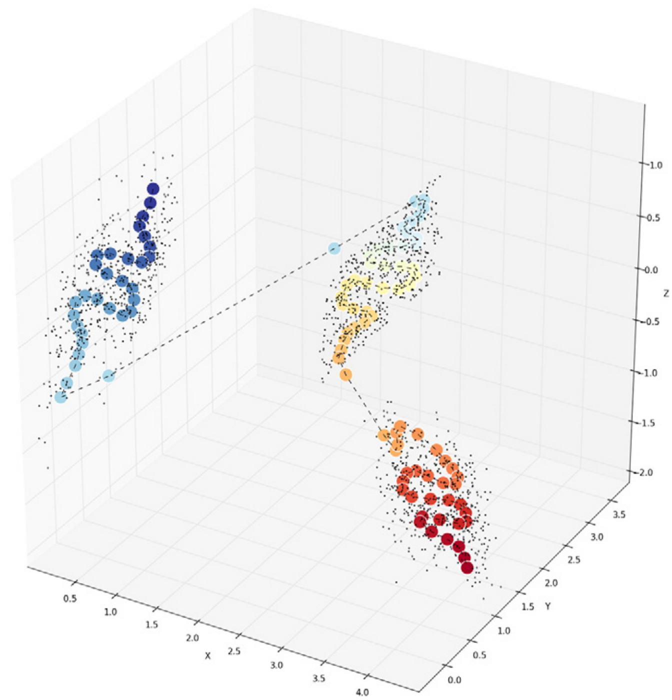

Fig. 4. One-dimensional SOM with 100 nodes, trained with a three-dimensional data set: the final weight vectors of the trained SOM are color-coded based on their index numbers from 1 to 100.

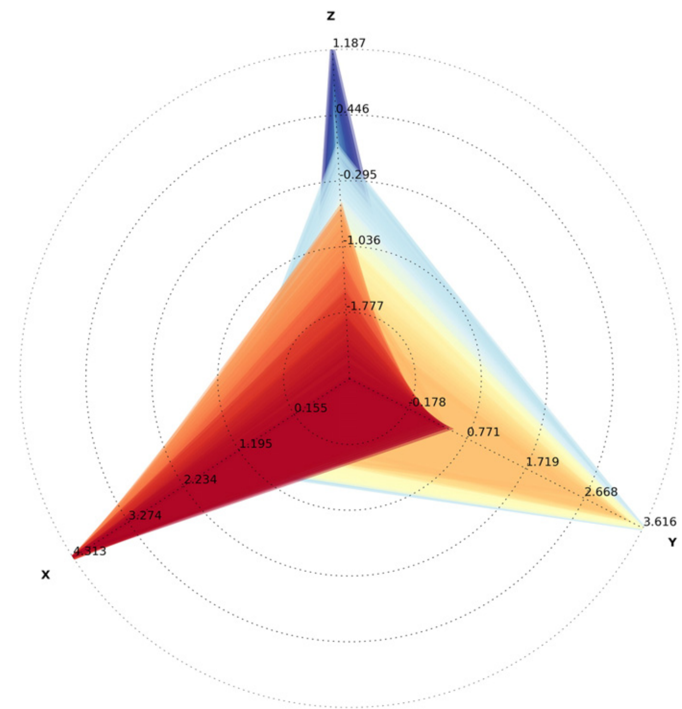

Fig. 5. Radar diagram showing the weight vectors corresponding to the trained SOM in Fig. 4.

As an example, Fig. 4 shows a case with synthetic three-dimensional data (shown with small dots), where there are 3 distinct clusters. A SOM with 100 nodes in a one-dimensional topology folds itself in a way to mimic the training data vectors. In addition to the identification of three main clusters, there are local distributions of the nodes within each cluster that account for the internal varieties within each cluster.

Considering the index of these nodes from 1 to K, which are colored in a spectrum, one can encode the high-dimensional data (in this example, the three-dimensional space) as a set of ordered numbers.

In this regard, a one-dimensional SOM can be seen as a sequence of ordered numbers pointing to a high-dimensional space, but compared to classical numbers such as natural numbers, the main difference here is that these numbers are ordered according to their similarities within the selected high-dimensional state space (context). As a result, they are called *contextual numbers* (Moosavi, 2014). Regarding the problem of multivariate mapping, with this transformation, the high-dimensional spatial vectors will be converted into single contextual numbers along with their geographic coordinates. Because we expect to have a smooth changing pattern from node to node, we can color these spatial maps with one-to-one relationships with standard color spectrums. From this viewpoint, one can claim that this new single data layer visualizes the emergent and composite spatial properties (i.e., clusters) in a quantitative manner. Therefore, in a similar argument, we call them *contextual maps*. Further, we can visualize the high-dimensional weight vectors corresponding to each contextual number by classical approaches such as a radar diagram. Fig. 5 shows a selected sample of weight vectors trained for the synthetic three-dimensional data set shown in Fig. 4. There are one-to-one relationships between the color-coded circles in Fig. 4 and each layer in Fig. 5. As we expected, there are three main clusters, and within each cluster, we have a spectrum of different combinations with similar color codes.

In Section 4, we will show the result of this approach for two different real-world spatial data sets.

4. Experiments with real-world spatial data

In this section, we show the results of the proposed method using two real-world spatial data sets. One is a collection of 235 attributes of the so-called wards in London (Fig. 2). The data set is provided by Future Cities Catapult from the abovementioned project Whereabout. The second data set is obtained from US census 2000 and 2010, including the distribution of different race groups at the census block level, corresponding to five boroughs of New York City.

In the data set from London, there are 23 different categories, each of which consists of several subcategories, for 235 spatial attributes. Table 1 shows the list of main categories.

The selection of each category or combination of different subcategories depends on the purpose of the study, and therefore, each selection will result in a different contextual map. However, as we are mainly interested in showing the potential uses of the methodology

Table 1
Different categories of data used for the case of London.

No.	Category	No.	Category
1	Age	**12**	Passports held
2	Cars	**13**	Tenure
3	Central heating	**14**	Qualification
4	Crime	**15**	Mode of travel to work
5	Distance travelled to work	**16**	Flickr photo count
6	Dwelling type	**17**	Mean of Medians Weighted by Sale Counts (2009)
7	Establishment type	**18**	Main language
8	General health	**19**	Food agency
9	Hours worked	**20**	Cycle hire locations count
10	Household composition	**21**	Pubs per square kilometer
11	Occupation	**22**	NS-SeC (different social classes)
		23	Green space

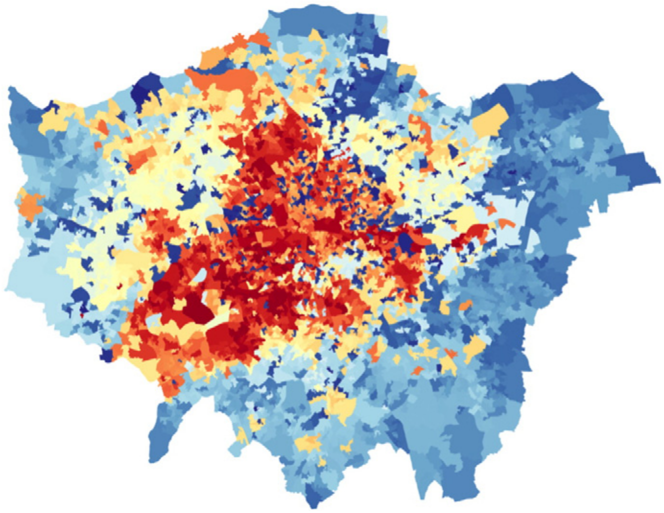

Fig. 6. Contextual map of London based on the two categories of qualification and general health.

proposed in this study, we show a few specific cases with and without distinct spatial patterns.

We implemented the proposed methodology in the Python programming environment, and the source code is openly available in an online repository.[3]

For the first example of London, we considered the combination of categories qualification and general health. Both variables are categorical, and the values of each subcategory show the percentage of people in that specific region (i.e., ward) who belong to that subcategory.

We trained a one-dimensional SOM with 200 nodes on this data set. After assigning the contextual numbers to each spatial point (i.e., values from 1 to 200), we project each spatial point to the geographic map of London using a simple color spectrum. As we expected, Fig. 6 shows that there is a large and distinct area with high values of red, while the distribution of other extremes with very low values (blue) are spread in smaller but distinct patches along the eastern side of London. Further, there is a clear non-overlapping region in the center, where we have completely opposite patterns of lifestyles. Compared to classical spatial clustering methods, we can here infer that the opposite colors indicate very opposite lifestyles, while in the classical cluster coloring, the colors are not comparable at all.

To interpret these colors, Fig. 7 shows a radar diagram. For each contextual number, we have a high-dimensional weight vector. Therefore, each axis shows the values of each dimension. As shown, the clusters have expanded between two orthogonal extremes. In red clusters, we have those regions where people have high levels of qualifications and very good health conditions, and on the other extreme in blue clusters, we have regions with lower levels of qualification and relatively poor health conditions. At the same time, there are other clusters between these two extremes.

An important property of this approach is that because those nodes with similar values refer to similar high-dimensional patterns, the number of nodes in the one-dimensional SOM acts as a level of resolution in the color spectrum used for the geo-map. By adding more nodes, we will have more individual distinctions, while the general patterns will remain stable. Fig. 8 shows the effect of the number of nodes on the configuration of clusters in the example shown in Fig. 7. As shown, after almost 200 nodes, the final categorization becomes very stable.

In another example, we used the category of crime as the input for the contextual maps. As shown in Fig. 9, although there is no evident spatial pattern across London (except in the central part), Fig. 10 shows that different types of crime have formed four main clusters.

Further, as another example with no categorical data, Fig. 11 shows the results of the proposed method for 4 dimensions of mean values of working hours for females and males plus the average distance to work and average age within each ward.

Further, Fig. 12 shows a radar diagram explaining the configuration of its corresponding contextual map.

In the previous examples, we used data from a single cross-section in time, but it is possible to implement the same method on the spatiotemporal data to examine places that have changed similarly with time. In this regard, as another experiment, we applied this method to the distribution of different races in New York City for the years of 2000 and 2010. The data are obtained from USA census 2000 and 2010, and the resolution of the data is at the census block level. In this case, after data preprocessing, we had approximately 30,000 spatial points with 5 dimensions, each representing the percentage of people in any of the groups of White, Asian, Black/African American, Latino, and others. Here, we have two goals. First, we are interested to know how integrated or segregated different regions are and how different regions have changed from 2000 to 2010. We trained a one-dimensional SOM with 1500 nodes using both data sets from 2000 and 2010. As a result, we have the same referencing system to show levels of aggregation/segregation. Figs. 13 and 14 show how segregated or integrated NYC was in different regions in 2000 and 2010. While there are mainly regions with unique races, some mixed regions also exist in transition from one cluster to another.

Further, Fig. 15 depicts the distribution of each race within each contextual color.

The additional benefit that we obtain from using contextual numbers is that we can treat them like normal numbers even though they refer to high-dimensional vectors. For example, in the case of New York City's racial mixture, we can visually observe the degrees of change in the racial mixtures from 2000 to 2010. To do so, we normalized the contextual numbers to a range of 0–1, and we then simply mapped the absolute degrees of change in the corresponding contextual numbers for 2000 and 2010. Fig. 16 shows that while most blocks have been stable (regardless of their mixture), there are some regions that have changed radically, either from being integrated to very segregated or vice versa.

While this case was limited to only 10 years, it would be interesting if one could examine this methodology over several decades to see, for example, how those regions that were segregated (integrated) have changed gradually over time.

As a final note in this section, it is worth mentioning that since in this paper we only use static figures to visualize the radar diagrams, sometimes it becomes hard to see all the underlying layers in a diagram. Therefore, the ideal visualization for these radar diagrams would be to have web based interactive diagrams, where the user can easily explore the geo-maps in parallel.

5. Discussions and future research

In this section, we will discuss two main technical issues related to the proposed methodology plus one potential application in the field of urban planning and zoning.

The first point is about the chosen one-dimensional topology of SOM. As we briefly mentioned before, it is known that having higher grid dimensions or a more-connected neighborhood topology in the SOM network can improve the performance and quality of the trained SOM in terms of quantization error and topology preservation. Therefore, since in the proposed method we are strictly using a SOM with one-dimensional grid structure, we should, assume a trade-off between dimensionality reduction and detail preservation.

On the basis of our experiments in this study, we observed that by appropriate selection of the number of training cycles both topologies

[3] https://github.com/sevamoo/SOMPY.

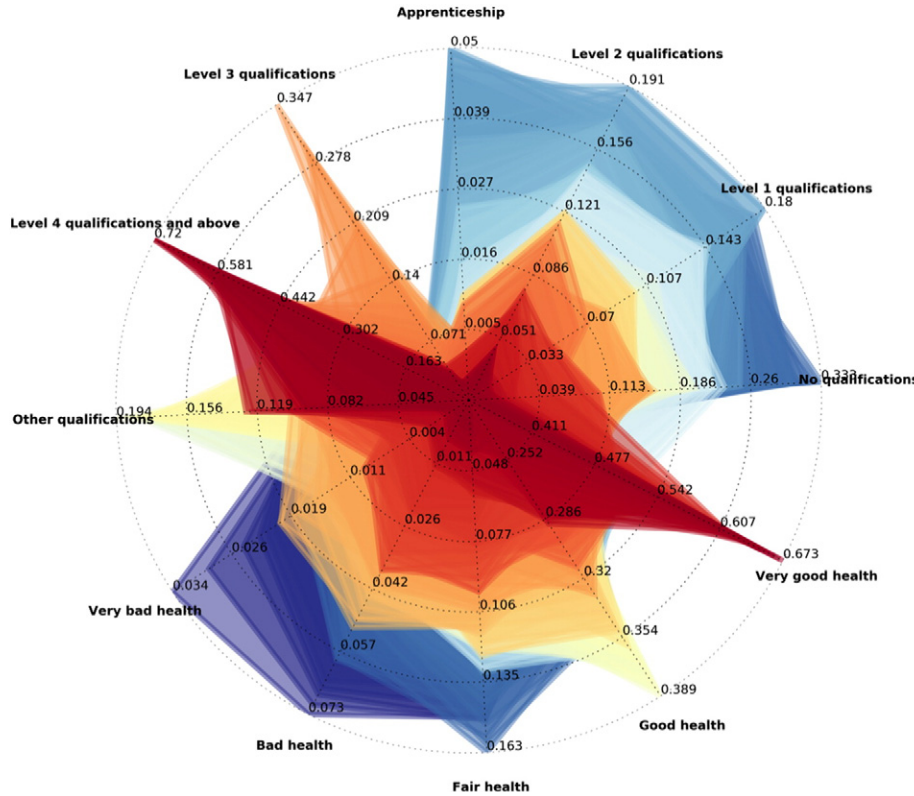

Fig. 7. Radar diagram for the contextual map of London shown in Fig. 6.

(i.e. one-dimensional and two-dimensional grid structures) can end up to very similar quantization errors. In order to test this issue, we compared the quantization errors in SOMs with the same size of nodes, but different topologies of 1×400 grid and 20×20 grid respectively. Except the number of iterations for rough training and fine-tuning, all the other factors such as learning rates and initialization methods were always chosen the same based on the recommended setting in Vesanto, Himberg, Alhoniemi, and Parhankangas (2000). Further, in

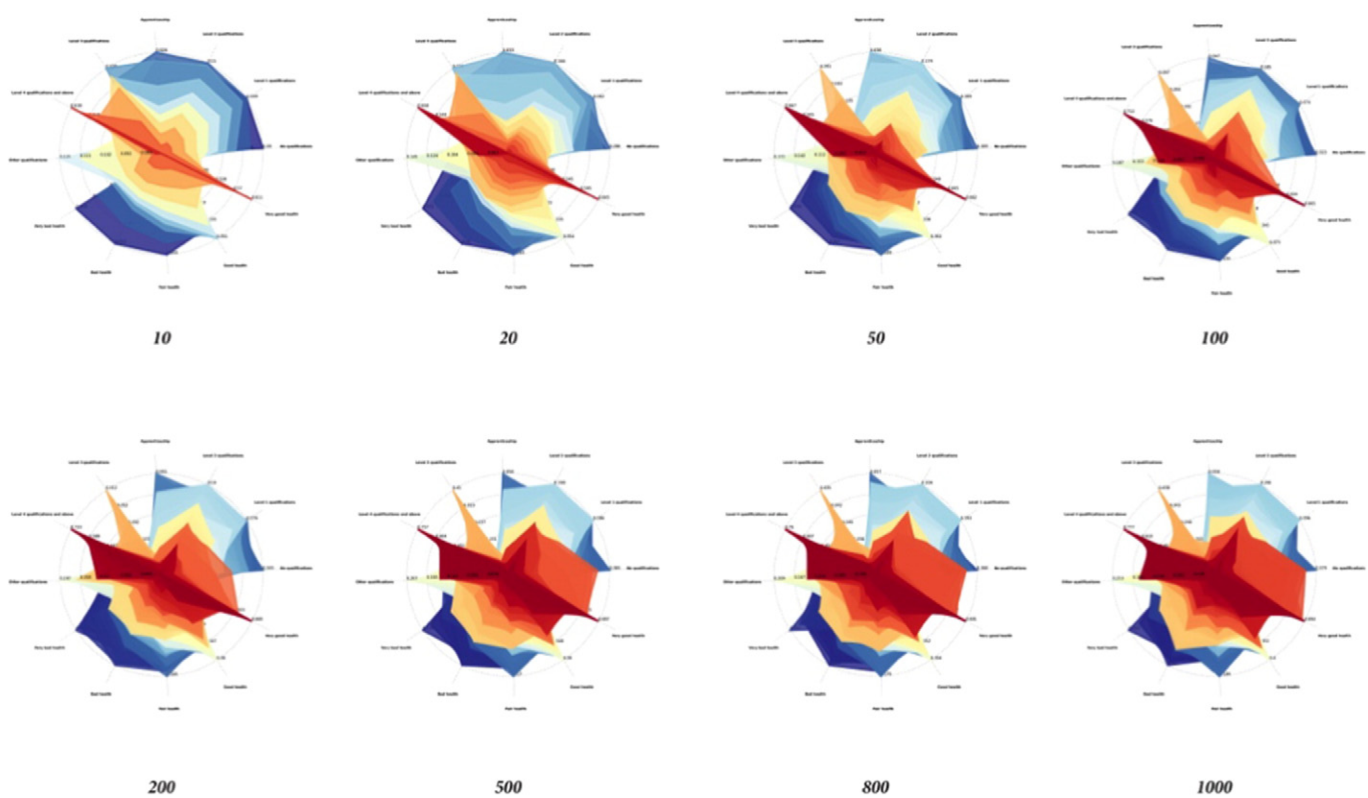

Fig. 8. Effect of the number of nodes in a one-dimensional SOM on the configuration of identified patterns.

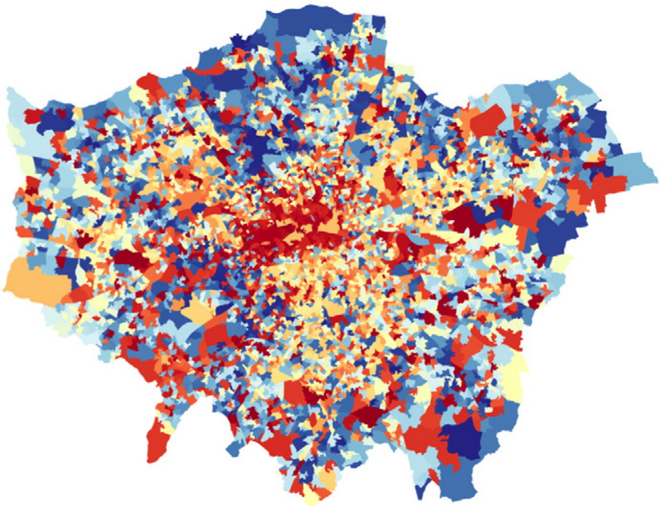

Fig. 9. Contextual map of London based on the category of crime.

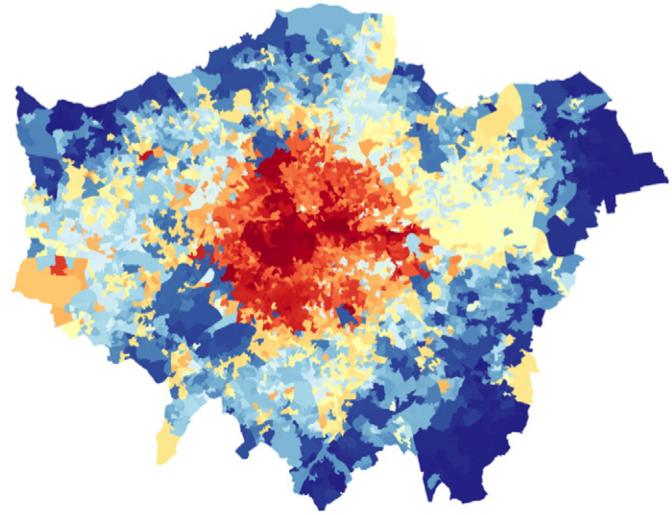

Fig. 11. Contextual map of London based on the average working hours, the average distance travelled to work, and the average age within each ward.

order to see the effect of dimensionality of training data on the quantization errors, we ran several tests with different data dimensions (from 2 to 20 dimensions). The data sets were randomly chosen from London data set, which was introduced in Section 4. As expected Fig. 17 shows that by increasing the data dimension the quantization errors will increase regardless of the chosen grid topology. Further, in comparison a one-dimensional SOM with the same training iterations as a two-dimensional SOM has always relatively higher quantization errors. However, it can be seen that by doubling the training iterations, both topologies achieve very similar quantization errors.

In addition to quantitative measures, on the basis of our experiments with spatial data sets, the radar diagrams visualizing the identified high-dimensional patterns always show logical patterns from one cluster to another, with no twisted or disordered clusters.

The second important technical issue of the proposed methodology is that while the distances between indices of the trained one-dimensional SOM are equal (i.e., from 1 to K equal to the number of nodes), their similarity does not change linearly or in another words, their

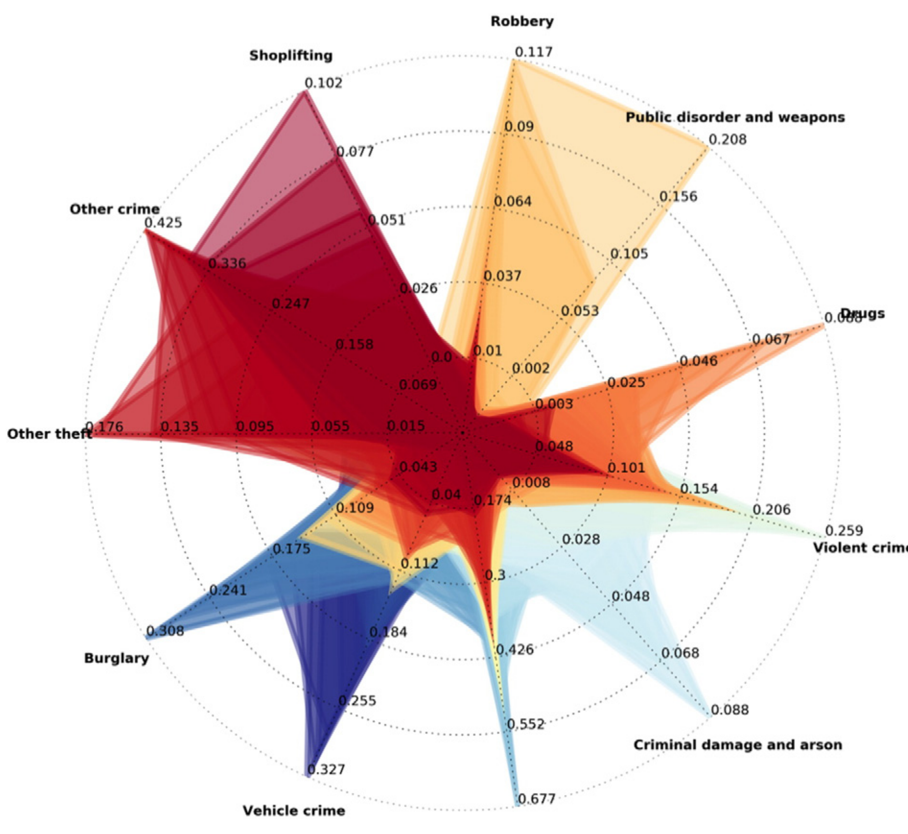

Fig. 10. Radar diagram for the contextual map of London shown in Fig. 9.

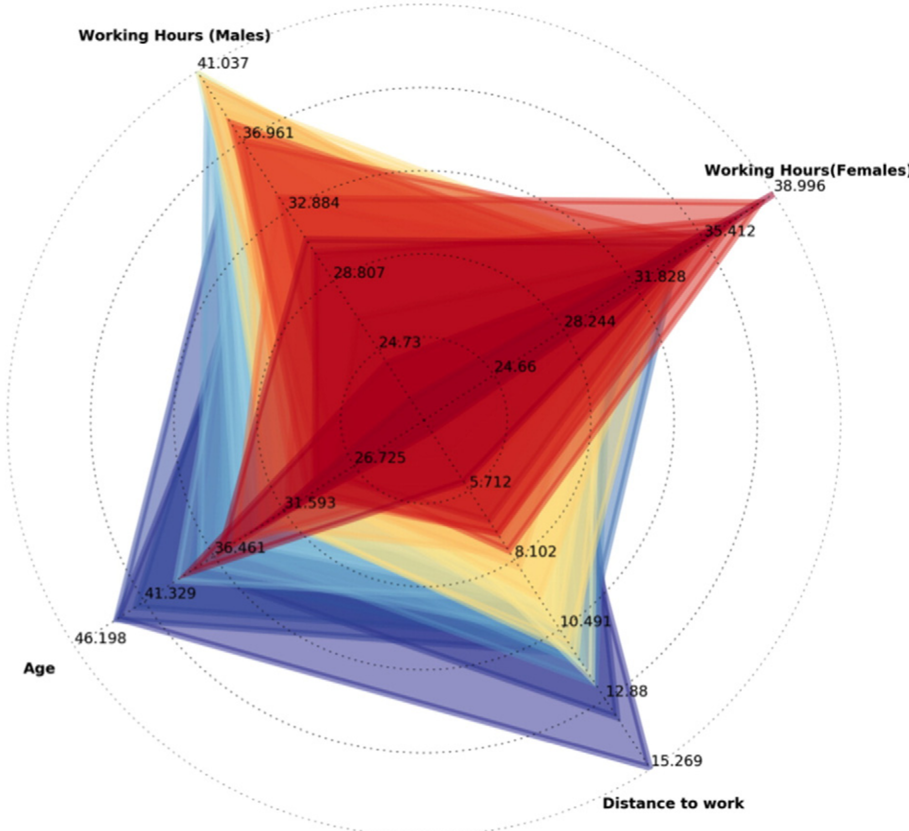

Fig. 12. Radar diagram for the contextual map of London shown in Fig. 11.

distance in high dimensional space is uneven. However, as shown in Fig. 4, if there is a gap between high-dimensional clusters, they will be usually identified in the corresponding radar diagrams (e.g. Fig. 5) with a clear gap between the identified clusters. Nevertheless, we should highlight that the color-coded geo-maps should not be interpreted individually and only with a reference to the radar diagrams. The same problem exists in those cases such as the example in NYC (Figs. 13, 14 and 15) that contextual numbers have been used computationally.

As an initial solution we propose the following steps in order to solve this problem.

Assume $CN_j = j$, for $j = 1, \ldots, K$ is the originally proposed contextual numbers. If we define $HDist_j = w_j - w_{j-1}$ for $j = 2, \ldots, K$ as the sequential distances between high dimensional vectors of the original contextual numbers, we introduce an adjusted contextual number ACN_j for $j = 1, \ldots, K$ that can be calculated as follows:

$$\mathrm{ACN}_1 = 0$$
$$\text{for } j \text{ in } \{2, \ldots K\}$$
$$ACN_j = ACN_{j-1} + HDist_j / \min_j HDist_j$$

With this simple procedure, we define the minimum distance as a unit of contextual numbers and shift the indices to the right in a way that the adjusted values preserve the distances between nodes in high

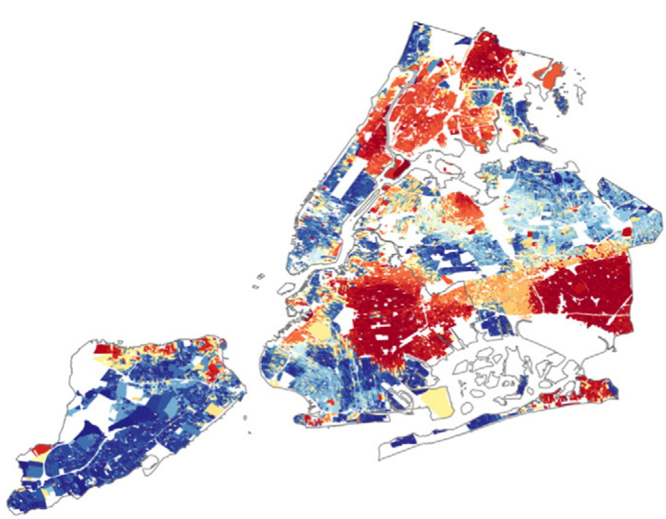

Fig. 13. Contextual map of the racial mixture in New York City in 2000.

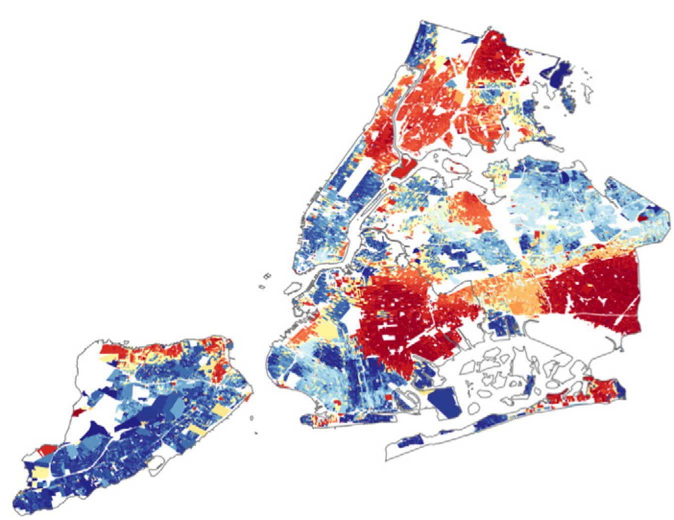

Fig. 14. Contextual map of the racial mixture in New York City in 2010.

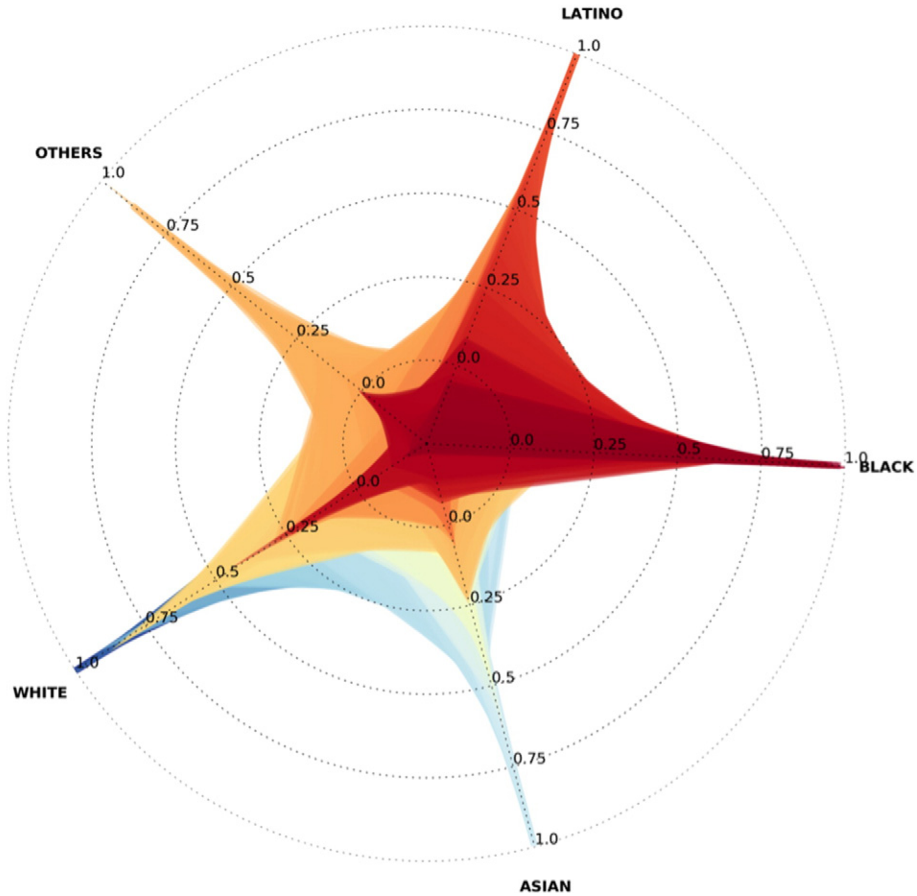

Fig. 15. Radar diagram for the contextual map of racial mixture in New York City.

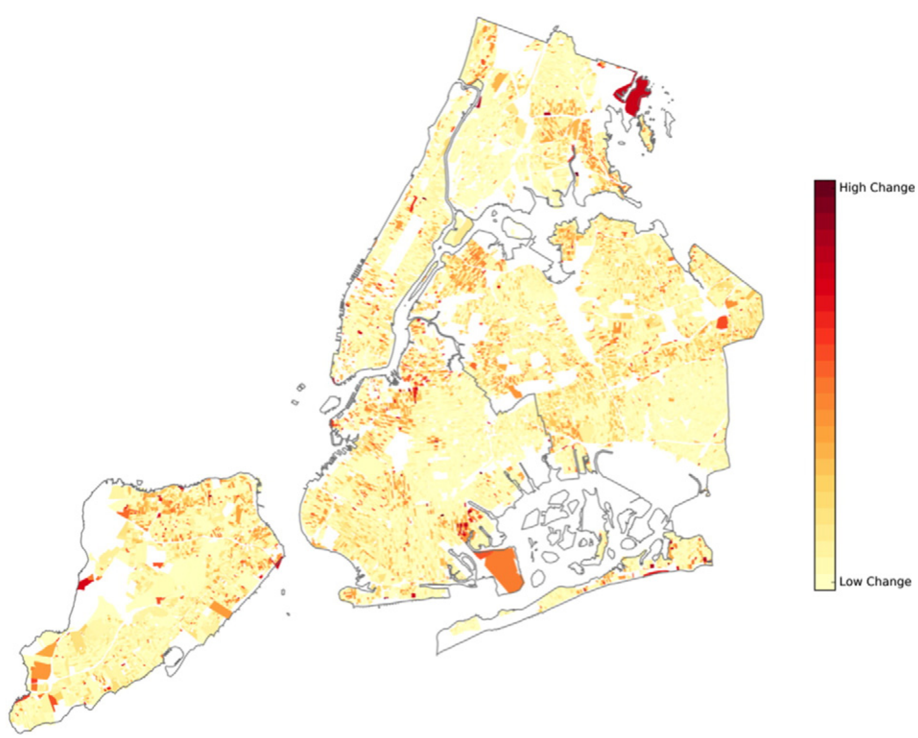

Fig. 16. Qualitative degrees of change in the racial mixture of NYC from 2000 to 2010.

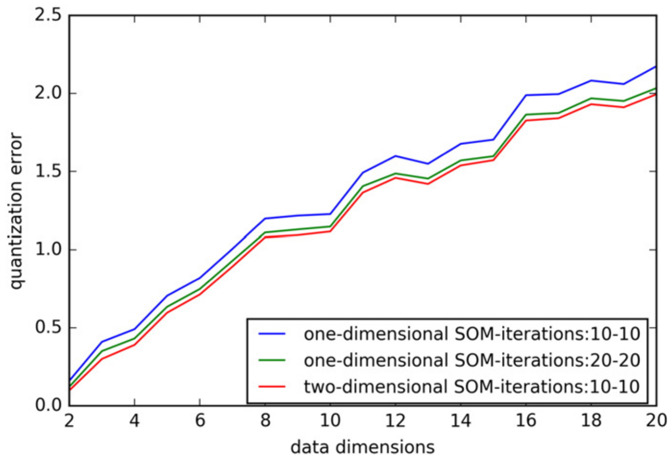

Fig. 17. Comparison of the quantization errors of two different SOM grid topologies for the same data sets with different dimensions.

dimensional space. Fig. 18 shows the result of this procedure on the example shown in Fig. 4. As expected, there are two clear leaps between the adjusted contextual numbers that indicate the discontinuity in the higher dimensional space. Finally, we should mention that although the proposed adjustment works well in this example, we would like to leave further investigations in this problem to the future research.

In this paper, we only showed some basic applications of the proposed method. In addition to these examples, we showed in Moosavi (2014) that the contextual numbers could be used in combination with other contextual numbers or other features in a hierarchical manner.

As a possible application of this idea, we think that in the domain of urban planning and zoning, as an alternative to the traditional zoning approaches based on a fixed set of aspects of a city, it is possible to develop *collaborative images of the city* by using the proposed contextual mapping in a participative manner, where the final map is produced from many individualized maps (each of which is based on different aspects of a city).

This problem can be considered a type of spatial clustering problem, where the entire region of a city is normally divided into a set of homogenous regions because of administrative requirements such as elections, public transportation, policy analysis, or similar requirements. The criteria for this spatial clustering, such as the one recently done for London (known as LOAC), are typically combinations of census variables (Longley & Singleton, 2014). Nevertheless, the output of this clustering problem, which is only based on a few categorical aspects, will later have political power in further decisions about that city.

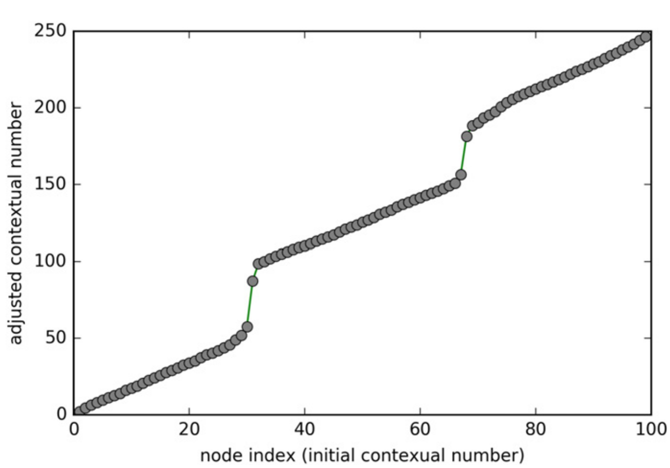

Fig. 18. The adjusted contextual numbers in comparison to node indices (initial contextual numbers) for the example shown in Fig. 4.

Alternatively, in a collaborative planning scenario, by providing a user-friendly web-based platform, one can develop a type of digital survey, where citizens (including experts and non-experts) are asked to select their main aspects of interest through which they see the city they live in. Considering the selected aspects for each user, there will be a corresponding contextual map of that city (i.e., a unique emergent image of the city per person). Next, on another level, it is possible to take each of these emergent images of the city as a new spatial dimension and produce another contextual map that can be considered a collaborative image of the city. The problem of identifying the image of a city has been an important issue for urban planners both in the past (Lynch, 1960) and present through the use of social media (Salesses, Schechtner, & Hidalgo, 2013). Nevertheless, we think that none of the previous works have developed a method to combine these high-dimensional individualized images of the city in a collaborative manner. Therefore, as a future work, we would like to test this idea in a realistic scenario.

6. Conclusions

With the ever-growing availability of digital data in many spatial domains, we need to develop appropriate methods to explore high-dimensional and complex spatial patterns. Compared to classical data clustering problems, one of the main issues of spatial pattern recognition and spatial clustering is that in spatial clustering, in addition to finding high-dimensional patterns, one needs to keep the spatial coordinates in parallel to other features. Finally, it is always desired to project the high-dimensional patterns onto the geographic maps.

In this study, we developed a methodology for mapping high-dimensional spatial data onto a single geographical map, through which we can visualize the high-dimensional spatial patterns in an intuitive way.

We also showed how the less explored one-dimensional SOM could transform high-dimensional data points into a set of one-dimensional ordered indexes called contextual numbers. The main property of these contextual numbers is that similar numbers refer to similar high-dimensional contexts. As a result, the high-dimensional spatial data points that are normally considered several spatial layers in parallel can be visualized based on their corresponding contextual numbers in a single-layer geographical map. Because these output maps can be interpreted in relation to the selected context (i.e., the selected high-dimensional space), they are called contextual maps.

We showed applications of the proposed method based on one synthetic and two real-world data sets. Finally, we discussed two main technical issues related to the proposed method and one potential application in the field of urban planning that need to be investigated in future research.

Acknowledgments

This research was supported by the National Research Foundation Singapore (NRFS) through the Singapore-ETH Centre for Global Environmental Sustainability (SEC) and the Chair for Computer Aided Architectural Design (CAAD) at ETH Zurich. Further, the author would like to thank the reviewers of the paper as their comments on the initial submission significantly improved the quality of the final paper.

References

Agarwal, P., & Skupin, A. (Eds.). (2008). *Self-organising maps: Applications in geographic information science*. John Wiley & Sons.

Arribas-Bel, D., Nijkamp, P., & Scholten, H. (2011). Multidimensional urban sprawl in Europe: A self-organizing map approach. *Computers, Environment and Urban Systems*, *35*(4), 263–275.

Bação, F., Lobo, V., & Painho, M. (2005). The self-organizing map, the geo-SOM, and relevant variants for geosciences. *Computers & Geosciences*, *31*(2), 155–163.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(8), 1798–1828.

Cheng, Y. (1997). Convergence and ordering of Kohonen's batch map. *Neural Computation*, *9*(8), 1667–1676.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. (2012, June). The livehoods project: Utilizing social media to understand the dynamics of a city. *In International AAAI Conference on Weblogs and Social Media* (pp. 58).

Delmelle, E., Thill, J. C., Furuseth, O., & Ludden, T. (2013). Trajectories of multidimensional neighbourhood quality of life change. *Urban Studies*, *50*(5), 923–941.

Erwin, E., Obermayer, K., & Schulten, K. (1992). Self-organizing maps: Ordering, convergence properties and energy functions. *Biological Cybernetics*, *67*(1), 47–55.

Flanagan, J. A. (2001). Self-organization in the one-dimensional SOM with a decreasing neighborhood. *Neural Networks*, *14*(10), 1405–1417.

Frenkel, A., Bendit, E., & Kaplan, S. (2013). The linkage between the lifestyle of knowledge-workers and their intra-metropolitan residential choice: A clustering approach based on self-organizing maps. *Computers, Environment and Urban Systems*, *39*, 151–161.

Guo, D., Gahegan, M., MacEachren, A. M., & Zhou, B. (2005). Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science*, *32*(2), 113–132.

Hagenauer, J., & Helbich, M. (2013). Contextual neural gas for spatial clustering and analysis. *International Journal of Geographical Information Science*, *27*(2), 251–266.

Henriques, R., Bacao, F., & Lobo, V. (2012). Exploratory geospatial data analysis using the GeoSOM suite. *Computers, Environment and Urban Systems*, *36*(3), 218–232.

Kohonen, T. (2001). *Self-organizing maps. Vol. 30 of Springer Series in Information Sciences.*

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65.

Longley, P. A., & Singleton, A. D. (2014). *London output area classification: Final report.*

Lynch, K. (1960). *The image of the city. Vol. 11*, MIT Press.

Mahinthakumar, G., Hoffman, F. M., Hargrove, W. W., & Karonis, N. T. (1999, January). Multivariate geographic clustering in a metacomputing environment using globus. *Proceedings of the 1999 ACM/IEEE conference on supercomputing* (pp. 5). ACM.

Martinetz, T., & Schulten, K. (1991). *A "neural-gas" network learns topologies.* University of Illinois at Urbana-Champaign, 397–402.

Moosavi, V. (2014). *Computing with contextual numbers. arXiv preprint arXiv:1408.0889.*

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. *PloS One*, *8*(7), e68400.

Skupin, A., & Esperbé, A. (2011). An alternative map of the United States based on an n-dimensional model of geographic space. *Journal of Visual Languages and Computing*, *22*(4), 290–304.

Spielman, S. E., & Thill, J. C. (2008). Social area analysis, data mining, and GIS. *Computers, Environment and Urban Systems*, *32*(2), 110–122.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319–2323.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *63*(2), 411–423.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using *t*-SNE. *Journal of Machine Learning Research*, *9*(2579–2605), 85.

Vesanto, J. (1999). SOM-based data visualization methodsl. *Intelligent Data Analysis*, *3*(2), 111–126.

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). *SOM toolbox for Matlab 5.* Finland: Helsinki University of Technology.

Wang, N., Biggs, T. W., & Skupin, A. (2013). Visualizing gridded time series data with self organizing maps: An application to multi-year snow dynamics in the Northern Hemisphere. *Computers, Environment and Urban Systems*, *39*, 107–120.

Yan, J., & Thill, J. C. (2009). Visual data mining in spatial interaction analysis with self-organizing maps. *Environment and Planning B: Planning and Design*, *36*(3), 466–486.