Methods on Defining Spatio-Temporal Neighbourhoods

Tao Cheng Berk Anbaroğlu

Department of Civil, Environmental and Geomatic Engineering
University College London, WC1E 6BT London, the United Kingdom
Telephone: +44 (0) 20 7679 2738
Email: {tao.cheng; b.anbaroglu}@ucl.ac.uk

1. Introduction

Prediction, classification, clustering and outlier detection are among the major tasks of spatio-temporal data mining and analysis. These tasks involve analyzing spatio-temporal neighbourhoods (STNs) so that the spatio-temporal associations and correlations can be incorporated in the analysis. Since STNs of an instance will give important clues on the evolution of the instance itself, better defining and identifying the neighbouring instances will lead to better modelling of the phenomena under investigation. It will also improve accuracy of prediction, classification, clustering and outlier detection in space-time series analysis.

STNs are defined and used in various research areas. Since there is no formally accepted and standardized method to define the STN, it seems that each research proposed its own spatial or spatio-temporal neighbourhood definition intuitively. We believe that a systematic and quantitative formulation of STNs should be derived in order to prevent the ambiguity of choosing STNs intuitively and therefore have better modelling of spatio-temporal data.

The aim of this paper is to review the methods available on defining STNs and to propose a quantitative approach to identify the STNs.

2. Existing Methods of STN Definitions

STNs exhibit different forms in point, grid, polygon and graph/network data types with different applications which range from epidemics to traffic analysis, and from image/video analysis to cellular automata and outlier detection. The following sub-sections provide overview of STNs defined under these different data types. Most STNs are defined based upon spatial neighbourhoods, especially for point data.

2.1 Point Data

Defining spatial neighbourhood on point data can be achieved in two ways. The first way is to use a distance metric (e.g. Euclidean or Mahalanobis) and then define a threshold which will indicate whether two instances will be neighbours or not. For example, Lu et al. (2003) and Chen et al. (2008) used k-Nearest Neighbour (k-NN) based on Mahalanobis distance to find the spatial neighbourhood of an instance. If the Mahalanobis distance between the instance and its spatial neighbourhood exceeds a threshold, the instance is detected as spatial outlier. Huang et al. (2006) and Celik et al. (2006) used Euclidean distance to create the spatial neighbourhood of each instance to detect the co-occurrent patterns.

Second approach in defining the spatial neighbourhood is to transform the point data into polygons by Delaunay Triangulation. Instances which share a common edge will be in their spatial neighbourhoods (Min-qi et al., 2008). In this approach, longer Delaunay edges which lie at the borders are pruned to have a more meaningful neighbourhood.

In all these researches, the parameters such as k of k-NN, threshold of Mahalanobis/Euclidean distance and pruning threshold in Delaunay Triangulation need to be set intuitively or based on trial and error which is time consuming. In addition, the temporal dimension has not been considered in the definition of STNs for point data.

2.2 Grid Data

Several neighbourhood strategies have been developed for grid data which have regular structure. These neighbourhood strategies can be considered as metaphor of chess pieces and are shown in Figure 1. For example, the Rook neighbourhood has neighbouring grids in its vertical and horizontal directions, and the Queen neighbourhood has neighbouring grids in its vertical, horizontal and diagonal directions. The Within-row and the Across-row neighbourhoods are special cases which are used if there is prior information on the data.

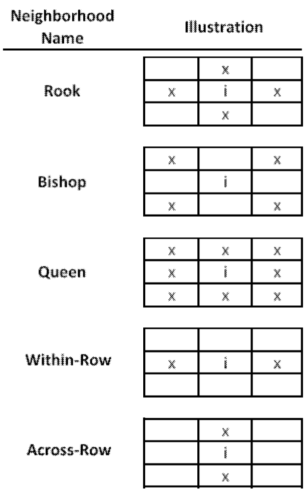

Fig 1. Instance *i*'s possible neighbourhoods

The Queen-based spatial neighbourhoods are widely used in image/video analysis for various applications. Ng and Solo (1998) conducted research on optical flow estimation and the Queen-based spatial neighbourhoods are defined by optimising a performance function. Zhang et al. (2008) modelled the dynamic changing background in video-sequence data and the STNs are defined by combining previous time stamp and the Queen neighbourhood. Similarly, Zhao et al. (2008) defined STNs by combining five time stamps within 4×4 pixels spatial neighbourhood. They modelled background of a video on night conditions where changes in lighting conditions were taken into account.

Another commonly used spatial neighbourhood for grid data is the Rook neighbourhood. Yin and Collins (2007) defined STNs as Rook neighbourhood combined with temporal neighbourhoods which cover the previous and future time stamps. By using STN of the pixel, they detected and tracked moving objects. Reynolds and Madden (1988) used Rook neighbourhood and suggested spatio-temporal autocorrelation metric could be used unless there is prior information on the data. In addition, importance of weighting the neighbours' influence on the instance was mentioned. They conducted these experiments on a plant disease epidemic. Rook neighbourhood is also commonly used in digital terrain modelling analysis. Kidner (2003) stressed on the importance of choosing the correct neighbourhood size which will lead to better interpolation in digital terrain modelling. However, heterogeneity in the terrain is not considered.

Not all spatial neighbourhoods need to have a regular structure. Zhao and Billings (2006) used spatial neighbourhood of a cell to extract cellular automata rules from observed patterns in which spatial neighbourhood can be in any arbitrary shape. Since the cells of the cellular automata evolve according to the rules defined on that neighbourhood, it is important to find the correct neighbourhood.

2.3 Polygon Data

In almost all of the researches, if two polygons share a common edge then they are thought to be spatial neighbours. The first order spatial neighbours cover the polygons which have a common edge with the query polygon, and the second-order spatial neighbours share a common edge with the first-order spatial neighbours of the query instance. Billard et al. (2007) predicted an epidemic (i.e. mumps) in time across 12 states of US using first- and second- order spatial neighbourhood relationships by using space-time bilinear model. Chawla and Sun (2006) and Lu et al. (2003) used the first-order spatial neighbours to detect spatial outliers by comparing the non-spatial attributes of the query polygon and its spatial neighbourhood. Janeva and Atluri (2009) proposed a three step procedure in defining the first-order spatial neighbourhoods, which were refined by using the interrelations between attributes and their distributions.

2.4 Graph/Network Data

Network data are treated as graphs where nodes in the graph are connected through edges. Spatial neighbourhoods are defined by graph connectivity: if two instances are connected via an edge, those two instances will be in their spatial neighbourhood. Also, temporal neighbourhood is constructed as previous and future time stamps. Spatio-temporal outliers of traffic data were detected by comparing the value of a traffic sensor and those of the sensors as its STNs (Shekhar et al. 2001). Wang et al. (2007) took a similar approach to cluster traffic sensors by defining spatial neighbourhood as spatial connectivity of the nodes and temporal neighbourhood as time-series similarity between the two nodes. In both approaches the graph connectivity is thought as first-order connectivity and all connected edges are weighted equally. However, the direction of the traffic flow or capacity of the road which affect the definition of the STN has not been considered.

3. Defining STN based upon STARIMA model

Choosing STNs of an instance intuitively may lead to a wrong understanding and modelling of the space-time data. As discussed above, a quantitative and a formal approach is needed, which should reflect the correlation among space and time in the data precisely. This section introduces a space-time model - STARIMA (Spatio-Temporal Autoregressive Integrated Moving Average) which is adopted to define the STNs.

The STARIMA model captures the linear space-time autocorrelation structure from space-time series data. In STARIMA model, each observation $z_i(t)$ at current time t and location i is not only influenced by the previous time series at the location, but also affected by the previous time series of its spatial neighbours (Box et al, 1984), which is expressed as a linear combination as follows:

$$z_i(t) = \sum_{k=1}^{p} \sum_{h=0}^{m_k} \phi_{kh} W^{(h)} z_i(t-k) - \sum_{l=1}^{q} \sum_{h=0}^{n_l} \theta_{lh} W^{(h)} z_i(t-l) + \varepsilon_i(t) \qquad (1)$$

where p is the autoregressive order, q is the moving average order, m_k is the spatial order of the k^{th} autoregressive term, n_l is the spatial order of the l^{th} moving average term. Φ_{kh} is the autoregressive parameter at temporal lag k and spatial lag h, θ_{lh} is the moving average parameter at temporal lag l and spatial lag h, $W^{(h)}$ is the $N \times N$ matrix of weights for spatial order h, and $\varepsilon_i(t)$ is a normally distributed random error at time t and location i. After defining the model specifications, the question is to identify these spatial/temporal lag parameters (Pfeifer and Deutsch, 1980).

STARIMA model has been widely used as a model to predict the future changes of the phenomena under investigation such as prediction of traffic flow (Kamarianakis, 2005) and timber prices (Zhou and Buongiorno, 2006). Here the principle of STARIMA is used to define the STN of an instance because the spatial order at different autoregressive order (or at different moving average order) can be considered as spatial neighbours defined at different time. Finding p and q and their corresponding spatial orders will define the range of the STNs.

The parameters (p, q) depend on the space-time dependence which is measured by the space-time autocorrelation function (ST-ACF) and partial space-time autocorrelation function (ST-PACF). From the calculation of ST-ACF and ST-PACF, time lag (temporal neighbour) and the space lag (spatial neighbour at particular time lag) could be defined. This method has been tested in a real case of outlier detection (Cheng and Anbaroglu, 2009).

4. Summary and Conclusions

This paper addresses the importance of defining STN by providing the broad range of application areas that benefit from STN. Almost all of the researches define STN intuitively. In particular, most STN are actually based upon spatial neighbours and temporal dimension are added on top of the spatial dimension, i.e. no formal definition of integrated STNs is available yet.

Quantitative formulation of STNs by the means of STARIMA is proposed in this paper. To achieve this, spatio-temporal autocorrelations (ST-ACF and ST-PACF) are used.

Capturing the spatio-temporal autocorrelation to define the STNs of network data will be the direction for future research since network data seems most complicated compared with other data types. In addition, how to incorporate additional information in refining the STNs will be an open research area which will increase the effectiveness of the STNs, but also increase the computational complexity.

Acknowledgements

This research is jointly supported by UK EPSRC (EP/G023212/1), Chinese NSF (40830530) and 863 Programme (2009AA12Z206).

References

Billard L et al., 2007, Modeling Spatial-Temporal Epidemics Using STBL Model, *International Conference on Machine Learning and Applications,* 629-633.

Box G E P et al., 2007, Time Series Analysis: Forecasting and Control. Prentice-Hall.

Chawla S and Sun P, 2006, SLOM: a new measure for local spatial outliers, *Knowledge and Information Systems*, 9(4): 412-429.

Celik M et al, 2006, Sustained Emerging Spatio-Temporal Co-occurrence Pattern Mining: A Summary of Results, *18th IEEE International Conference on Tools with Artificial Intelligence,* 106-115.

Chen D et al., 2008, On Detecting Spatial Outliers. *GeoInformatica*, 12(4): 455-475.

Cheng T and Anbaroglu B., 2009, Spatio-Temporal Outlier Detection in Temperature Data, COSIT-09 Workshop on Spatial and Temporal Reasoning for Ambient Intelligence Systems, 21 September 2009, France

Huang Y and Pei J and Xiong H, 2006, Mining Co-Location Patterns with Rare Events from Spatial Data Sets, *GeoInformatica*, 10(3): 239-260.

Janeja V P and Atluri V, 2009, Spatial outlier detection in heterogeneous neighborhoods, *Intelligent Data Analysis*, 13(1): 85-107.

Kamarianakis Y and Prastacos P, 2005, Space-time modeling of traffic flow, Computers and Geosciences, 31:119-133

Kidner D B, 2003, Higher-order interpolation of regular grid digital elevation models, *International Journal of Remote Sensing*, 24(14), 2981-2487.

Kim D and Choi J and Kim J, 1998, Adaptive motion estimation based on spatio-temporal correlation, *Signal Processing: Image Communication*, 13(2): 161-170.

Lu C-T and Chen D and Kou Y, 2003, Detecting spatial outliers with multiple attributes, *15th IEEE International Conference on Tools with Artificial Intelligence,* 122-128.

Min-qi Z et al., 2008, An Algorithm for Spatial Outlier Detection Based on Delaunay Triangulation, *International Conference on Computational Intelligence and Security,* 102-107.

Ng L and Solo V, 1998, Choosing the optimal neighbourhood size in optical flow problems with errors-in-variables modelling, *International Conference on Image Processing,* (2):186-190.

Pfeifer P E and Deutsch S J, 1980, A three-stage iterative procedure for space-time modeling, *Technometrics,* 22: 35-47

Reynolds K M and Madden LV, 1988, Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology*, 78(2): 240-246.

Shekhar S and Lu C and Zhang P, 2001, Detecting graph-based spatial outliers: algorithms and applications (a summary of results), *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining*, San Francisco, 371-376.

Wang Y et al., 2007, SPANBRE: An Efficient Hierarchical Clustering Algorithm for Spatial Data with Neighborhood Relations, *4th International Conference on Fuzzy Systems and Knowledge Discovery,* 665-669.

Yin Z and Collins R, 2007, Belief Propagation in a 3D Spatio-temporal MRF for Moving Object Detection, *IEEE Conference on Computer Vision and Pattern Recognition.*

Zhang S and Yao H and Liu S, 2008, Dynamic background modeling and subtraction using spatio-temporal local binary patterns, *ICIP 2008,* 1556-1559.

Zhao Y and Billings S, 2006, Neighborhood detection using mutual information for the identification of cellular automata, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* , 36(2): 473-479.

Zhao Y et al., 2008, Spatio-temporal patches for night background modeling by subspace learning, *ICPR 2008.*

Zhou M and Buongiorno J, 2006, Space-Time Modeling of Timber Prices, *Journal of Agricultural and Resource Economics*, 31(01):40-56.