Traffic Framework for Causal Graph with Causal Density and Mutual Information

Soobin Yim* — Sejong University
Chanyoung Jung† — Sejong University
Giwoong Park‡ — Sejong University
Yun Jang§ — Sejong University

Abstract

In this work, we propose a novel approach to examine spatiotemporal causality using a visual analytics framework. Our framework objective is to expand the traffic causal tree by applying the techniques of Mutual information and Granger causality. To analyze traffic causality, the visual analytics framework computes the mutual information and causality of traffic data across the entire road network, allowing users to scrutinize the relationship between them. The study's key contributions are a new visual analytics framework for examining spatiotemporal traffic causality and an evaluation of the framework's practicality using real-world data.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual analytics; Information systems—Information systems applications—Decision support systems—Data analytics;

1 Introduction

Recurring traffic congestion causes delays and significant social costs. To alleviate these problems, several studies have been proposed, including traffic jam type analysis [4], traffic flow pattern analysis [7], and signal time optimization [6]. In this paper we aim to analyze the root causes of traffic congestion and identify structural issues within the urban traffic network. We introduce a novel visual analytics framework to examine spatiotemporal causality. We concentrate on identifying spatiotemporal traffic causality within small areas using the Mutual Information (MI) and Granger causality. The proposed framework allows users to analyze the correlation between MI and causality of traffic data across the entire road network. The framework is intended to analyze the spatiotemporal causality of small areas while expanding the traffic causal tree. Our framework calculates MI and causal density using traffic volume and speed data to measure the mutual dependence of two random variables with entropy and to determine the strength of traffic causality using the Granger causality test, respectively. MI quantifies the mutual dependence of two random variables with entropy. Causal density measurement aids in the selection of significant causality among numerous ones. The distance of each road's travel is related to causality, and the spatiotemporal characteristics of the road network must be considered to avoid deriving inaccurate causality in the traffic causality analysis. The study's contributions are a novel visual analytics framework for analyzing spatiotemporal traffic causality and an evalution of the framework's practicality using real-world data.

2 Mutual Information

Entropy is a representative method of measuring the amount of information in information theory. The entropy of X is represented as follows.

$$H(X) = -\sum_{x\in X} p(x)log\left(p(x)\right)$$

The joint entropy of X and Y is computed as follows.

$$H(X,Y) = -\sum_{x,y\in X,Y} p(x,y)log\left(p(x,y)\right),$$

where statistically $H(X,Y) = H(X)+H(Y)$ when the two variables X and Y are independent. MI quantifies the mutual dependence of two random variables with entropy.

$$I(X,Y) = \sum_{x,y} p(x,y)log\left(\frac{p(x)p(y)}{p(x,y)}\right),$$

where $I(X,Y) = H(X)+H(Y)-H(X,Y) = H(X)-H(X \mid Y) = H(Y)-H(Y \mid X)$. At this time, MI has 0 if the two variables X and Y are independent. When $MI \neq 0$, MI measures the mutual dependence of the two variables X and Y, which means how much the mutual uncertainty is reduced [1].

3 Causal Calculation

Causality refers to the idea that an event can cause another event to occur. The Granger causality test has been proposed to analyze the causal relationship between two variables [2]. However, this test only determines the existence of a causal relationship without measuring its strength. To address this limitation, Seth [5] proposed a method that assigns a numerical value to the causal relationship called causal density. Causal density is the average of G-causality within a framework, and it measures the ratio of two variances when the variable Y has G-causality to the variable X, such as Equation 1.

$$F_{\{Y\rightarrow X\}} =: ln\frac{var(\varepsilon'_x(t))}{var(\varepsilon_x(t))} \tag{1}$$

$$CD(X) =: \frac{1}{n(n-1)}\sum_{i\neq j} F_{\{X_i\rightarrow X_j\}} \tag{2}$$

The calculation of the causal density involves a comparison of variances across all the elements in the framework, as expressed in Equation 2. This value serves as an indicator of the dynamical complexity, with a CD of 0 suggesting that the elements are completely independent of one another.

4 Visual Analytics Framework

We describe our visual analytics framework used to analyze traffic network models and causal traffic patterns in Manchester speed data [3], recorded at 5-minute intervals illustrated in Fig. 1. The causal relationship analysis begins when a user selects a node pair according to the MI value in (a). Note that the darker the color of the heatmap, the higher MI. The cause node X and effect node Y are the nodes corresponding to the row and column of the cell selected in (a). In (b), a heatmap shows the causal density by the time lag between the pair of nodes selected in (a). (b-1) and (b-2) are the causal density heatmap from the cause node X to the other node and effect node Y to the other node, respectively. Similar to (a), for (b),

*e-mail: tn12qls@gmail.com
†e-mail: cyjung6@gmail.com
‡e-mail: rct3232@gmail.com
§e-mail: jangy@sejong.edu

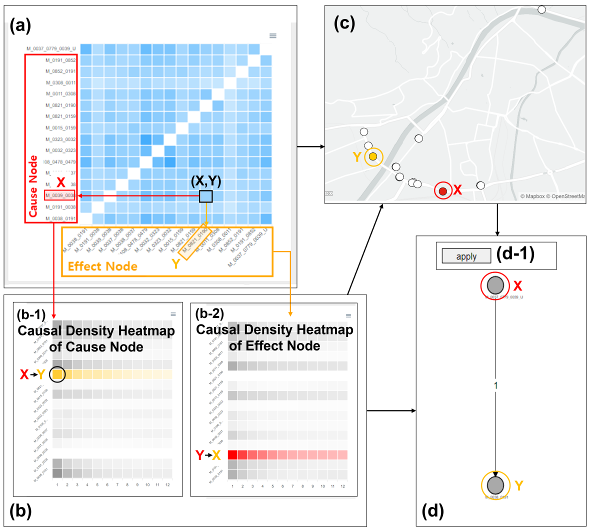

Figure 1: A visual analytics framework used to analyze traffic network models and traffic flow propagation patterns in Manchester. (a) is a heat map for the MI between each node of traffic data. (b) is a causal density heat map, where (b-1) and (b-2) are causal density heat maps for cause and effect nodes, respectively. (c) is a visualization of nodes on a map, and (d) is a causal graph that the user generates.

the darker the color of the heatmap, the higher the causal density value of the node pair. (c) shows the positions of the nodes on the map. Users can also select other node pairs and time lag through (b). The framework is updated accordingly by any cause node, effect node, or time lag is changed. Finally, the user clicks the apply button (d-1) to create or expands a causal relationship tree in (d).

5 Case Study

We evaluate the framework's practicality using real-world data. This case study is conducted for experts who worked traffic analysis research for 13 years to discover causal traffic relationships within small areas. We used traffic data [3] from Manchester, England, recorded at 5-minute intervals to showcase the functionality of their proposed framework and evaluate its usefulness with domain experts. Fig. 2 (a) and (b) display the causal density heatmaps for nodes A and B, respectively. In (b), node B shows a high causal density for different time lags of nodes C and D, as demonstrated in the causal tree graph in (c). The expert noted that the strong causal density from node B to node D might have been influenced by external nodes excluding nodes B and D. Therefore, expert analyzed node C, which showed strong causal density in node B. (d) shows the causal density heatmap for node C, which was added to the causal relationship tree. Node B was added as the child of node C again for time lag 9 with the highest causal density, as shown in (e). While the influence of the traffic flow between nodes A, B, and D proceeds in one direction after time lag 12, the expert found that during time lags 4 and 9 at nodes B and C, the effect of the traffic flow was circular. As a result, the expert deduced that the traffic flows from node B to C and then back to node B affect node D. The expert noted that the circulation of these traffic flows requires geographical information and identified node B as a road connecting the intersections of nodes C and D. The expert mentioned that the flow circulations between nodes B and C may be caused by congestion factors such as signal time at the intersection of node C or road capacity, and that node B affects node D again because it is a road connecting the two intersections. Finally, the expert emphasized that resolving the congestion problem at node C is essential if traffic congestion occurs in the area comprising nodes A, B, C, and D.

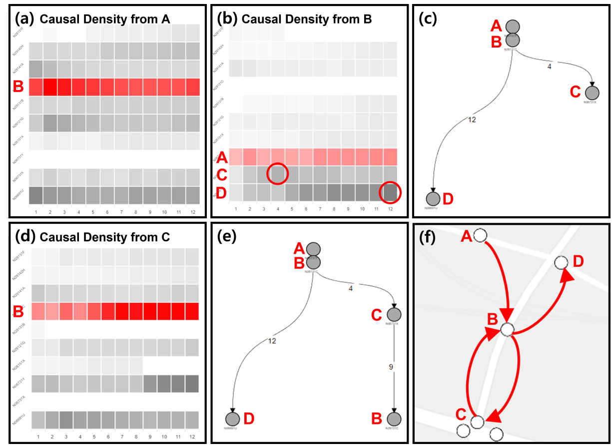

Figure 2: This is a case study for Manchester data. (a), (b) is a causal density heatmap for nodes A and B, and (c) is a causal graph for A, B, C, and D. (d) is a causal density heatmap for node C, and (e) is a graph with the addition of the direction of the causal relationship from C to B. (f) is the visualization of A, B, C and D on the map according to the direction of the causal density.

6 Conclusion

In this paper, we introduced a novel framework for analyzing traffic flow to assist in making decisions about traffic policies and tracking congestion. The framework involves constructing a tree of causal relationships between different traffic nodes, which is achieved by calculating the mutual information and causal density of traffic data. The framework's usefulness has been demonstrated using a real dataset, and we propose will be improved through future work by using more diverse data and methods, and developing new visualization methods.

Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-02076, Developing Reasoning AI Engine in Complex Systems (REX) and its Applications) and (No.2022-0-00305, Development of automatic data semantic information composition/expression technology based on augmented analysis for diagnosing industrial data status and maximizing improvement). Yun Jang is the corresponding author.

References

[1] T. M. Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[2] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

[3] A. Loder, L. Ambühl, M. Menendez, and K. W. Axhausen. Understanding traffic capacity of urban networks. *Scientific reports*, 9(1):1–10, 2019.

[4] M. Pi, H. Yeon, H. Son, and Y. Jang. Visual cause analytics for traffic congestion. *To appear in IEEE Transactions on Visualization and Computer Graphics*, 2019.

[5] A. K. Seth. Causal connectivity of evolved neural networks during behavior. *Network: Computation in Neural Systems*, 16(1):35–54, 2005.

[6] K. Wada, K. Usui, T. Takigawa, and M. Kuwahara. An optimization modeling of coordinated traffic signal control based on the variational theory and its stochastic extension. *Transportation research procedia*, 23:624–644, 2017.

[7] L. Xin, D. Yang, Y. Chen, and Z. Li. Traffic flow characteristic analysis at intersections from multi-layer spectral clustering of motion patterns using raw vehicle trajectory. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 513–519. IEEE, 2011.