Traffic Causality Analysis for Robust Road Freight

John Krumm
Cyrus Shahabi
University of Southern California
jkrumm@usc.edu

Project Objective

The goal of this project is to develop an algorithm that uses logs of traffic speeds on roads to automatically find sections of road that persistently cause traffic slowdowns on other roads, especially for roads that frequently carry freight traffic. Focusing resources on the causal roads will be an efficient way to relieve congestion on other roads.

Problem Statement

Given traffic speed data from several different road locations as time series, use the data to automatically discover which sections of road likely cause traffic slowdowns on other sections of road. Specifically, find "causal pairs" of roads X and Y, where a slowdown on road X is likely responsible for a slowdown on road Y. Figure 1 shows an example, where $x(t_i)$ represents the time series of road speeds a location X, and $y(t_i)$ represents the time series at location Y.

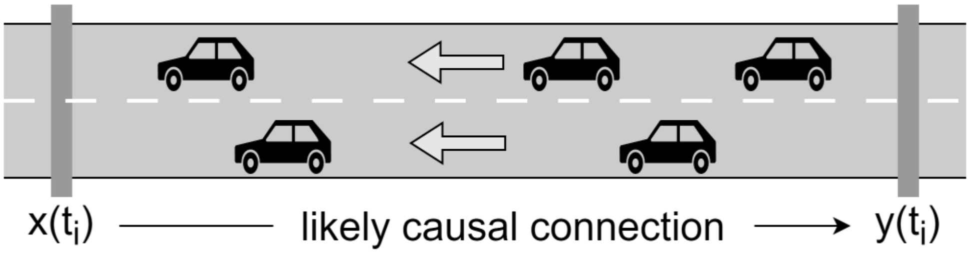

Figure 1: Road speeds at this pair of locations are likely causally connected as a slowdown at X would cause a slowdown at Y after some lag in time. Our algorithm automatically finds pairs of locations like this.

Research Methodology

We downloaded road speed data for Los Angeles highways from PeMS, which gives traffic speed at intervals of five minutes for hundreds of road locations as shown in Figure 2. We developed an algorithm that finds the leading edges of unusual traffic slowdowns at each road location. These slowdowns serve as natural experiments for finding cause-and-effect relationships among the roads. Using the traffic slowdown events, we developed another algorithm that uses machine learning to detect causal pairs, where a slowdown on one road causes a slowdown on another road. We trained and tested our machine learning model on a set of ground truth causal pairs, where the structure of the roads makes obvious the causal or non-causal connection between the two locations.

Figure 2: Locations of “Mainline” sensors in Caltrans District 7. The blue dots represent the 195 stations we used in our causality analysis, because they had sufficient speed data.

Results

We tested causality detection algorithm on thousands of candidate causal pairs. Given a pair of road locations and their associated speed time series, the algorithm makes a binary decision about whether or not there is a causal connection. We can adjust the sensitivity of the algorithm to trade off the number of true positives and false positives. Figure 3 shows the result as we adjust the algorithm’s sensitivity. Our algorithm’s performance is shown as the red curve. From the plot, we find that one operating point is very close to giving 0% false positives and 100% true positives. The red ROC curve in Figure 3 has an area beneath it of 0.9995, which is close to the maximum possible 1.0, meaning that the algorithm was very accurate in discovering causal pairs.

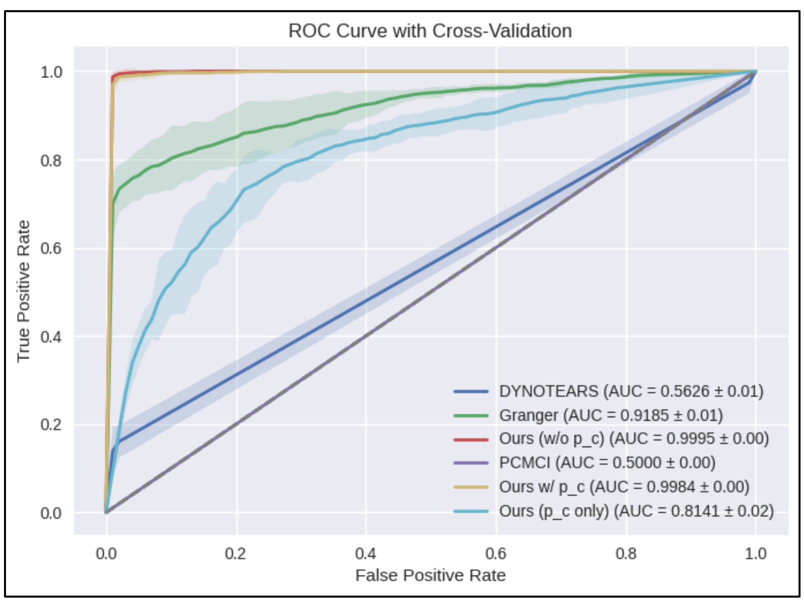

Figure 3: The idea operating point is at the upper left of this ROC curve. Our algorithm’s performance is shown as the red curve, which is nearly perfect compared to competing methods for discovering causal pairs.