Classifying accident avoidance maneuvers on heterogeneous road networks using exploratory spatial data analysis

Introduction

Are abnormal maneuvers and other conflicts within the traffic stream directly associated with road classifications, speed limits, congestion, roadway configuration, signaling, weather, and other traffic related factors? Contemporary transportation research is focused on developing a broader understanding of these relationships within the broader context of transportation infrastructure design and driver behavior with traffic congestion and vehicle accidents. With many years of collected data detailing the number of vehicle-miles traveled and the frequency of traffic accidents, a multi-disciplinary research agenda has emerged in which innovative techniques for identifying the complex interactions between road geometry, driver characteristics, and environmental conditions are providing unique insights into the dynamics of human cognitive behavior relative to various exogenous and endogenous user-network factors. As with most investigations, the ability to fully develop a meaningful research strategy is dependent upon techniques that can effectively parameterize the interactions between (a) the driver and the road network, and (b) how these relationships dynamically evolve over time and space. This study proposes to address these issues by utilizing exploratory spatial data analysis to better define a framework for understanding the relationships expressed between individual driving behavior (e.g., sudden vehicle maneuvers, driver demographics, accident frequency) and road infrastructure characteristics (e.g., network topography, geometry, and environmental conditions).

Goals & Objectives

To better understand and predict occurrences of traffic accidents as they relate to various driver characteristics and road network design, contemporary geo-processing techniques associated with geographic information systems (GIS) software will be employed to identify existing and reveal new relationships between various driver-network parameters and other environmental characteristics.

The primary goal defined for this study will be to establish a geospatial framework on which driver-network relationships can be parameterized in a consistent and robust manner. The first objective will be to spatially integrate various behavioral and environmental variables within a high-resolution road network. In doing so, multiple behavioral and demographic attributes sampled from various traffic study and flow models will be combined with accident data to produce a topologically integrated road network. A second objective will apply various exploratory geo-analytical techniques to analyze patterns, reveal clusters, identify associations, measure geographic distributions, and model spatial relationships according to global and local processes observed within the data. Finally, recommendations for additional spatial data analysis will be provided based on the lessons learned from this study.

Methodology

To achieve this project's stated goal and objectives, exploratory spatial data analysis (ESDA) will be used to categorize driver and traffic data relative to the spatial relationships exhibited by the underlying transportation network. These classifications may be further parameterized using various exploratory regression techniques that rely on spatially robust statistical techniques, which can quantify the strength of various explanatory variables that contribute the most to crashes/accidents. Results will be summarized and topologically integrated within the road network in order to support additional research activities.

Tasks

First, multiple behavioral and demographic parameters sampled from a traffic study will be extracted, transformed, and combined with accident data published by the National Highway Traffic Safety Administration (NHTSA). Data will be geocoded and referenced to a topologic road network developed by the Louisiana Department of Transportation and Development (LDOTD). Additional variables derived from socio-demographic, environmental, and geographic sources will be incorporated to provide a geographic framework on which to perform analysis.

With a reference inventory established, the next task will employ various exploratory geo-analytical techniques, including hot spot analysis, spatial pattern recognition, nearest neighbor analysis, and more to classify the various driver profiles and network characteristics relative to the road network topology. Research will be conducted to identify existing and develop new strategies for performing ESDA relative to referenced road networks. Application of these methods is expected to produce a robust classification scheme in which driver profiles can be reproduced and assessed various driving profiles with heterogeneous road geometries.

Additional techniques, including ordinary least squares (OLS) and geographically weighted regression (GWR) may be performed to reveal spatially dependent patterns and identify explanatory variables that address both global and local processes that influence the relationships within the data. Key factors, which may include signal lights and signs, lane count, flow rate, and others, can be assessed for their macro (global) and micro (local) scale influences on network congestion and, more precisely, traffic conflict behavior.

Expected Outcomes

As noted, the goal of this study is to define a consistent framework for effectively classifying traffic accident phenomena relative to driver behavior and network characteristics. In doing so, existing research endeavors may be able to model the variables identified by this study within agent-based regimes. Future research strategies may apply this framework within additional geo-analytical methodologies to more effectively model and potentially forecast the dynamic relationships between drivers and the road network. Accordingly, the methods and results produced from this study may provide a more effectively reveal how these driver-network relationships can evolve over heterogeneous spatio-temporal reference frames.

Deliverables

Results produced by this study will include (a) an extensively attributed GIS vector dataset comprising the topologically integrated road network, (b) multivariate geoprocessing models capable of integrating multiple parameters for driver behavior, road network design, and other data attributes, (c) published results in relevant peer-reviewed journals, (d) presentation and promotion of demonstrated research techniques to various conferences and workshops. A final report composed in the LSU UTC report format will be provided at the end of the project.

In order to facilitate demonstrations, a web-based geoprocessing application will be published and maintained by the P.I. for the duration of this project. Published web applications will remain online while resources remain available.

Schedule & Work Plan

The anticipated duration for this project is six months, and will have a completion date of June 30, 2013.

	Monthly Schedule (End Date: June 30, 2013)					
	1	2	3	4	5	6
Task 1: Data Transformation & Processing	X	X				
Task 2: ESDA		X	X	X		
Task 3: Exploratory Regression			X	X	X	
Task 4: Reports			X			X

Budget & Justification

Staffing Plan:

1. **Principle Investigator:** (approximately 2.5 months funded; 0.5 months matched)
   Research on current geoid models, as detailed in task 2, will be conducted by C4G staff.
   - Manages the tasks and assume responsibility for the conduct and completion of all aspects of this project.
   - Performs the research tasks associated with GIS analysis of network data assets.
   - Serve as a point of contact for all inquiries regarding aspects of the contract.
   - Author data products, map products, and Web-based mapping services.
   - Authors the final report.

Hardware & Facilities Available:

This research will be conducted at Louisiana State University, on the Baton Rouge campus. Hardware, software, and services utilized for this project are provided below.

1. **Network Attached Data Storage Unit:** (100% matched)
   Data resources required to perform this project will require a storage unit capable of supporting

the anticipated terabytes of raw, derivative, and published data layers. Storage units will be acquired and offered as match. Estimated expense for this unit is $12,000.[1]

2. **Geospatial Data Workstation:** (100% matched)
   This project will acquire a workstation that will support data types from various system platforms (e.g., Windows, Linux, Unix, and Mac). Estimated cost for this resource is $3,500.[2]
3. **GIS Workstation Software:** (100% matched)
   As a geoprocessing-intensive project, updated computing resources will be employed for the duration of this project. Software (e.g., Global Mapper and Google Earth Pro) will be acquired to help extract, transform, and load data resources for data analysis and application development. Estimated expense for this resource is $1,500.

Facilities and Administrative (F&A) rate for on-campus research with state agencies (48% of the MTDC).

Implementation:

The data produced by this study will be available to the GCCETR staff. These data will be made available via GIS based services maintained by the PI, and will be provided in the final report. Research findings and recommendations will be complied within the final report.

[1] Dell PowerVault NX3200 – NAS: $11,680.90 BTO (see appendix for quote).
[2] Apple Mac Mini with OS X Server : $3,244.80 BTO (see appendix for quote).