

Public Abstract

First Name:Vanessa

Middle Name:Amado

Last Name:Gonzalez

Adviser's First Name:Mark

Adviser's Last Name:Virkler

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:SP 2008

Department:Civil Engineering

Degree:PhD

Title:Knowledge Discovery and Data Mining From Freeway Section Traffic Data

The rapid development of intelligent transportation systems (ITS) has generated large amounts of data for transportation professionals. Currently, operators, planners, researchers, air quality analysts, transit providers, consultants, media, and others are using archived data. Benefits generated from these systems have already been accounted for in many large cities in the United States. However, analysis of archived ITS generated data can provide additional benefits for highway users. Additional research of the data generated from ITS will provide transportation professionals the ability to make better decisions.

Nonetheless, the better utilization of archived data will take time, but the more experimentation with data will allow greater benefits. In this research, a set of archived traffic generated data from Las Américas Expressway (PR-18) in the San Juan Metropolitan Region in Puerto Rico was examined. Traffic flow data, accident data, and work zone data from this facility were studied by means of data mining.

The methodology used was a combination of association mining and the knowledge discovery in databases (KDD) process. Association mining was used to learn about the hidden patterns within each model created and the KDD process was used as the framework that guided the entire process. The KDD process used consisted of seven steps: building the data mining database, examining and preparing the data, evaluating the data mining application, building the models, evaluating the models, preparing a list of the conclusions/knowledge gained, and the decisions. A total of six specific studies were developed in which different variables were studied using the association mining tools of the IBM Intelligent Miner for Data software package. The objective was to gain knowledge from the data about interrelationship between the variables. Thus the results obtained from the mining tool were excellent for the purpose of this research.

The approach was found to be a source of valuable information that could not have been detected by the use of traditional statistical analysis alone. The approach allowed the identification of: "red flags" during work zone operations; similar patterns in levels of service (LOS) between Tuesdays and Wednesdays and similar patterns in LOS between Mondays, Thursdays, and Fridays; and it allowed the analysis of LOS over time.

In addition, the analysis allowed the identification of temporary traffic control devices being impacted by vehicles, most common accidents, and the day of the week with the worst LOS. The major benefit learned from applying data mining to ITS generated data was that it allowed the analysis of numerous variables from multiple levels of information. The information learned could be used by the agency to improve safety on the segment of highway that was analyzed through the use of variable message boards, additional warning/regulatory signs, and maintenance and management decisions. New regulations could also arise from the information learned that could be implemented for work zone operations. Thus, the agency could improve work zone operations for the benefit of the safety of drivers and the construction workers.

The new knowledge provides the basis for more advanced studies to be developed. In addition, the methodology could be used at other locations to increase the quality of information available for decision-making on similar facilities.