Proceedings of the American Society for Engineering Management 2017 International Annual Conference
E-H. Ng, B. Nepal, and E. Schott eds.

K-MEAN CLUSTERING IN TRANSPORTATION: A WORK ZONE SIMULATOR CASE STUDY

Samareh Moradpour
Suzanna Long, PhD, PEM*
Missouri University of Science and Technology

***longsuz@mst.edu**

Abstract
Transportation engineering considers many different categories such as accident management, infrastructure management, driver behavior, and traffic management and generates large amount of data. Data mining methods are a common engineering approach to understand data-intensive scenarios and use techniques to extract patterns, correlations, and information from large amounts of data. This research uses a simulator to compare driver patterns and behaviors when comparing reactions to the Missouri Department of Transportation (MoDOT) alternate sign with the Manual on Uniform Traffic Control Devices (MUTCD) current sign. K-mean clustering method is used to cluster driver response to work zone sign configurations presented in the simulator environment and uncover patterns that can assist engineers with usability of work zone signage. Key findings of this research will help the transportation engineering manager make data-driven decisions regarding work zone safety and design.

Keywords
Data mining, Transportation, Pattern recognition, Decision analytics

Introduction
Due to increased volumes of data in all engineering fields, utilizing methods that can help decision makers best determine how to use that data are essential. It is important to apply methods to analyze data and to extract knowledge. Data mining methods can assist managers and decision makers to extract patterns and identify decision points suggested by the data. Data mining methods are a new generation of methods that are widely used in fields such as engineering, industry, and management. Transportation engineering is one of the areas that commonly uses data mining for data analysis. This paper presents the results of a simulator-based study comparing two different sign configurations. Driver performance and categorization of behaviors is determined through the use of k-mean clustering.

Literature Review
Data mining methods use statistical algorithms and techniques to determine relationship between the data. In other words, data mining methods are useful for extracting meaning and hidden relationships among data (Rygielski et al., 2002). Data mining methods are divided into cluster analyses, classification and prediction, as well as association rules (Liao et al., 2012). For this project, the clustering method is used. A short discussion of this technique follows below.

Clustering:
The Clustering method is one of the common data mining methods. Based on findings from the literature, clustering methods are effective and reliable for partitioning traffic data into different regimes such as free-flow and congested-flow regimes. In a traffic study, the most common clustering method estimates annual average daily traffic (AADT); this method is considered more acceptable than the other methods (Gecchele et al., 2011) in determining traffic flows. The clustering method is frequently applied to partitioning traffic flow data. Clustering methods divide data into groups (clusters) based on similarities and dissimilarities between groups and are useful for finding patterns between large amounts of data. K-mean clustering is one of the most common clustering methods and is used to determine the center of a cluster.