ABSTRACT

Title of Thesis:	DETECTING OUTLIERS FOR IMPROVING THE QUALITY OF INCIDENT DURATION PREDICTION
	Yi-Chi Lu, Master of Science, 2021
Thesis Directed By:	Professor Gang-Len Chang, Department of Civil Environmental Engineering

To circumvent the needs of domain expertise and the excessive data for developing a knowledge-based prediction system, such as the I-95 incident duration estimation model, this study has developed an efficient transferability analysis method to assess the applicability of adopting the prediction rules from an existing well-developed model to a different highway. The proposed analysis method has considered the common nature of incident response operations and local-specific incident characteristics in assessing the transferability of available knowledge-based rules for estimating the required clearance duration of different types of incidents. Evaluation of the proposed method with the I-695 incident records clearly shows that the prediction model developed with such an effective transferring method can achieve the same level of performance as ones with the original rule-searching and refinement method.

Since most incident records for model development are collected on-line during the emergency incident response process, some of the key data are likely to be mis-

recorded, which inevitably causes many existing models to yield undesirable performance, especially with respect to those incidents with insufficient records or excessive long duration. As such, this study has also developed a two-phase outlier detection process for identifying outliers and removing those viewed as faulty records from the dataset for model calibration and model evaluation. Using the I-695 incident records for a case study, the resulting performance of the proposed two-phase outlier detection process has proved its promising property for filtering faculty data from the incident records prior to the use for model development.

DETECTING OUTLIERS FOR IMPROVING THE QUALITY OF INCIDENT DURATION PREDICTION

by

Yi-Chi Lu

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2021

Advisory Committee:
Professor Gang-Len Chang, Chair
Professor Ali Haghani
Professor Cinzia Cirillo

Acknowledgements

These months, this Master's thesis was composed with assistance from so many people who spent their precious time and provided kindly help; thus, I want to state my deep appreciation to them.

First, I am sincerely thankful to my academic advisor, Professor Gang-Len Chang. It is hard to complete this study and the thesis without his professional directions and continuous support. These years, his teaching and insightful suggestions motivate my work and improve the comprehensiveness of my thoughts. Whenever I have difficulties in my research, his thoughtful and wise advice always helps me overcome them and keep moving forward.

Additionally, I am grateful to Professor Ali Haghani and Professor Cinzia Cirillo, who served as my thesis committee members. I appreciate their precious time and expertise to enhance my study.

Also, I want to show my gratitude to my dear colleagues and friends for their help in these years. I have special acknowledgment to Yao Chen and Yen-Lin Huang, who are always willing to give me useful guidance and recommendations on my research. My study and thesis could not be finished without their selfless assistance. I also want to show my gratitude to other members of our group where the cooperative atmosphere encourages me to learn and work hard every day. I will always remember and appreciate their selfless assistance.

Last but not the least, I am thankful to my dear family, who always stay with me. Their love always supports me overcome obstacles to achieve a better life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Research Background

Efficient incident management has long been a priority task of traffic agencies because non-recurrent congestion, due to the incidents on the roadway, is one of the main contributors to freeway traffic delays, often resulting in excessive fuel consumption, emissions, and secondary incidents. According to the Urban Mobility Report (Schrank et al., 2019), it is noticeable that traffic congestion from 1982 to 2017 is a persistently growing problem, especially in 2017 during which its congestion cost amounts to $166 billion. Among all traffic delays encountered by U.S. travelers, it has been reported that approximately half of congestion delays are non-recurring in nature, and such non-recurrent congestion delay may cause much more negative impacts on the efficiency of a freeway's operations and management.

Hence, to cope with the non-recurrent congestion caused by the incidents, most responsible highway agencies have attempted to deploy various response plans and coordinated management systems to manage the incident impacts on the highway. Traffic Incident Management System (TIMS) is one of such plans most widely adopted in practice, which relies on the best-estimated clearance duration for a detected incident to further predict critical information essential for incident responses and traffic management, including the length of traffic queues, the possible maximum affected area, the resulting delay times, and the selection of traffic control plans (Owens et al., 2010).

Considering the critical role of accurate incident duration information in various incident management strategies, the traffic community has devoted considerable efforts and resources to developing a tool for such needs (Garib et al., 1997; Peeta et al. 2000; Yu and Xia, 2012; Khattak et al., 2012; Qi and Teng, 2008; Chung, 2010; Hu et al., 2011; Kang and Fang, 2011; Araghi et al., 2014; Ji et al., 2014; Li, Pereira and Ben-Akiva, 2015; Zou et al., 2016).

Among those, the piloting knowledge-based model for Maryland ‘s I-95 freeway segment by Won et al. (2018), developed with the Association Rule Mining method (Agrawal and Ramakrishnan, 2015), has been reported to yield sufficiently reliable results for use by the field incident response operators. In addition, more of such a model for different highways are under developments for covering the entire Maryland’s freeway networks.

However, despite the effectiveness of the model by Won et al (2018) for Maryland’s I-95, its development method demands the researchers to have sufficient knowledge of the target highway’s incident characteristics and response agencies’ operational strategies, as well as adequate incident records of various natures for reliable model calibration. As such, it is expected that the development of such models for other highways shall take advantage of knowledge and prediction rules embedded in the well-developed existing systems, because incidents on different highways as long as responded by the same agency shall share some common characteristics and constraints aside from some local-specific factors such as driving patterns and geometric features. Hence, the development of an efficient method for reliably transferring applicable knowledge and prediction rules from an existing model to a new one for different highways is an imperative task and constitutes one of this study’s primary research objectives.

Another critical issue that impacts the accuracy of the incident duration estimation is the existence of missing/faulty data and outliers, which renders most existing studies to yield unacceptable estimates for the incident with an excessively short (e.g., less than 15 minutes) or long duration (e.g., more than 120 minutes). In other words, most prediction models can yield satisfactory prediction results for incidents with duration in a normal range, but not for those that should probably be considered as outliers in incident datasets (Valenti et al., 2010; Khattak et al., 2012; Li, 2015; Won et al., 2018).

According to Li et al. (2018). This may be attributable to the fact that most statistical methods and prediction algorithms tend to capture the central tendency in incident data instead of characteristics of the outliers, thus letting such outliers bias the model calibration and degrade the resulting accuracy. In addition, it should be mentioned that most critical data related to incident response and operations are recorded on-line during the emergency response process. As such, unless the data reporting system is well designed with an error-prevention function, it is likely that some faculty data may have been wrongly recorded into the incident database. The model development efforts and resulting accuracy can certainly benefit from the dataset without plaguing by the randomly distributed faculty data.

1.2. Research Objectives

In view of the above concerns, this research will focus on the following two issues:

1) To capture the individual and collective key factors associated with incident durations among various highways, this study performs a model transferability analysis where the knowledge and prediction rules in the existing models are utilized to develop new incident duration estimation models for other highways with less effort and fewer incident records; and
2) To reliably identify the outliers and further improve the accuracy of incident duration estimates, this study will develop an outlier detection model with the most effective outlier detection methods based on the distinct natures of incident records and the distributions as well as interrelations between their recorded key factors.

1.3. *Organization of the Thesis*

The rest of this thesis is organized as follows:

- **Chapter 2**: It introduces the definition of incident duration as well as outliers, and briefly reviews related studies and commonly used models in the fields of incident duration prediction and outlier detection;
- **Chapter 3**: It presents the core logic of the I-95 system, and then provides a detailed explanation of the proposed methodology for model transferability analysis;
- **Chapter 4**: It reports the analysis results with respect to the natures of the incident duration data and then discusses the review results of available methods for outlier detection. The rationales for the selection of three detection methods and the process for their collective use in outlier detection also constitute the core of this chapter;
- **Chapter 5**: It highlights the results of a case study with the I-695 incident data, including the effectiveness of the proposed outlier detection methods and their contribution to better the accuracy of the developed incident duration prediction model with the knowledge-based methodology; and
- **Chapter 6**: Conclusions of the research findings and future research works to ensure reliable deployment of such a vital system for incident management constitute the core of this chapter.

Chapter 2: Literature Review

Over the past decades, both incident duration prediction and outlier detection have been extensively studied in the literature. A brief review of related research and commonly used models in these two fields is presented below.

2.1. Incident Duration Prediction

2.1.1 Definition of Incident Duration

According to Highway Capacity Manual (2016), the entire incident duration as shown in Figure 2-1 can be divided into three stages: 1) the **incident detection time** indicates the period from the onset of an incident to its detection by response teams, 2) the **incident response time** is the duration from the detection of an incident to the arrival of the first response unit at the incident scene, and 3) the **incident clearance time** is the duration from the arrival of the first response unit to the clearance of all incident-related activities. Additionally, the so-called recovery time is measured from the completion of incident clearance work until the full recovery of the traffic conditions.

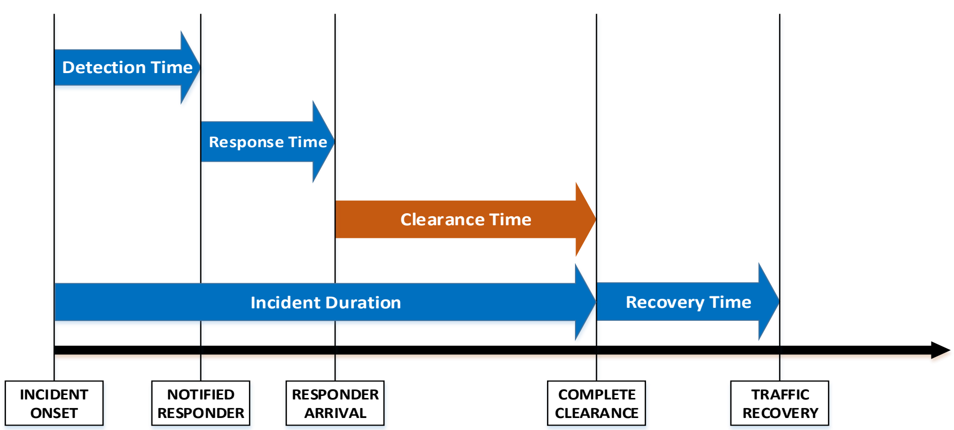


Figure 2-1. The Timeline of an Incident Duration

Due to the limitation of data availability for the incident detection, most studies in the literature mainly focused on the incident response or the clearance duration (Lee, et al., 2008; Alkaabi, et al., 2011; Ghosh, et al., 2012; Hou et al., 2014). For example, Hojati et al. (2014)

developed a set of models to predict the sum of the incident duration and the recovery time of an incident. Kaabi et al. (2012) and Hou et al. (2013) developed a prediction model on the response time of an incident. Nam and Mannering (2000) and Li (2015) focused on the incident response and management process (i.e., detection time, response time, and clearance time), and demonstrated that the required time for each phase of the incident response and management process may differ under different conditions. In this study, since one of the objectives is to build an extension model based on the model developed in Won et al. (2018), the review is focused on the related literature associated with the estimation of the incident clearance duration.

2.1.2 Characteristics of Incident Duration Related Factors

A large body of studies in the literature has already demonstrated that the incident duration would be influenced by various factors. Generally, those related factors can be briefly categorized into incident natures, temporal factors, environmental conditions, traffic flow conditions, responders and operational factors, and vehicle characteristics (Li et al. 2018). The impact of such factors on the incident duration varies with the incident nature and the stage of response and clearance operations. For example, some researchers concluded that the durations of different incident types, such as collisions and disabled vehicles, are determined by different factors (Hojati et al., 2013). Nam and Mannering (2000) and Li (2015) reported that the duration of various incident stages (e.g., detection, response, or clearance stage) are related to different influential factors.

In addition to those observable factors associated with the incident duration, there are some potentially critical factors that are not available due to the limitation of data collection methods. Thus, some studies investigated the possibility of using observable factors to infer those unobserved ones, but such latent influencing factors would result in heterogeneity of the incident duration by nature. To contend with the issue of heterogeneity, several studies have proposed the use of the gamma distribution with its parameters varying across observations depending on a pre-specified distribution, known as the random-parameter distribution model (Nam and Mannering, 2000; Hojati et al., 2013; Hojati et al., 2014; Li, 2015; Li et al., 2015; Chung et al., 2015).

The significant right-skewed distribution of the incident duration data is another issue that often causes the difficulty developing effective prediction models. Some prior studies have shown that the incident duration from different resources and datasets may exhibit different distribution patterns. For example, some studies revealed that the incident duration follows a log-normal distribution (Golob et al., 1987; Giuliano, 1989; Chung and Yoon, 2012), but other studies reported that their incident data follow the log-logistic distribution, or Weibull distribution (Jones et al., 1991; Nam and Mannering, 2000; Zhang and Khattak, 2010; Hojati et al., 2012; Wang et al., 2013; Chimba et al., 2014).

Among all types of distributions, the generalized F distribution was reported to best fit the incident duration data in some studies (Ghosh et al., 2012). In addition, some researchers studied the incident duration distribution under different natures and surrounding environmental conditions, concluding that the distributional assumptions vary with incident duration stages (Nam and Mannering, 2000; Li, 2015) and incident types (Hojati et al., 2012; Hojati et al., 2013; Hojati et al., 2014). Overall, the selection of an appropriate distribution is well recognized as one of the critical tasks in analyzing the incident data and developing prediction models (Smith and Smith, 2002). Note that since different types of distributions are suitable for different incident datasets, a recent study suggested the employment of mixture models to circumvent the difficulty selecting an appropriate incident duration distribution (Zou et al., 2016).

2.1.3 Traffic Incident Duration Prediction Models

As discussed above, the variable of incident duration, influenced by various complex factors, exhibits the nature of heterogeneity in its distribution. Thus, developing a reliable incident duration prediction model is a challenging task. Even so, many different methodologies have been proposed over the past two decades, using data from various sources to contend with the aforementioned issues. More specifically, most such methodologies reported in the literature can be classified into four categories: statistical approaches, classification techniques, Machine Learning algorithms, and multi-technique algorithms. Key studies in each category are briefly reviewed below.

Statistical Method

Among those studies in the first category, Garib et al. (1997) developed a regression model to analyze the incident duration data and identified several significant factors (e.g., response time) contributing to the duration of an incident. Peeta et al. (2000) developed two different models for different incident types (i.e., crashed and debris) and used the estimated results to provide timely traffic advisory and route guidance.

Following the same logic, Yu and Xia (2012) firstly classified the incident data based on incident types into traffic accidents and vehicle assistances, and then developed a regression model for estimating the incident duration of both types. Khattak et al. (2012) constructed a dynamic incident duration model with the quantile regression to enhance the accuracy of the incident duration estimation. Their prediction results indicate that their model can yield more accurate estimates for the incidents of longer duration, compared to the Ordinary Least Squares (OLS) regression model.

In addition to standard regression methods, hazard-based models were also widely adopted in developing models for the incident duration prediction to cope with the non-normal distribution issue of the incident data and the complex relations between contributing factors and the incident duration (Qi and Teng, 2008; Chung, 2010; Hu et al., 2011; Kang and Fang, 2011; Araghi et al., 2014; Ji et al., 2014; Li, Pereira and Ben-Akiva, 2015; Zou et al., 2016). For example, Chung (2010) proposed a log-logic accelerated failure time metric model to estimate the incident duration for the Korean Freeway Systems and identified associated contributing factors for the incident duration. Qi and Teng (2008), on the other hand, exploited available information to develop a time-sequential procedure that is a different hazard-based regression model with various factors at each stage.

Classification Method

Due to the difficulty determining proper distributions for incident data and the complex correlations among various contributing factors, several researchers have attempted to apply either the classification or tree-structured methods for development of a reliable incident duration prediction model. One of the commonly used approaches in this category is the Classification Tree Method (CTM). Ozbay and Kachroo (1999) first applied the linear regression technique to generate a hybrid decision tree to estimate the incident duration, which follows neither log-normal nor log-logistic distributions. Similarly, Smith and Smith (2002) utilized the Classification and Regression Tree (CART) algorithm to develop a decision tree that classifies the estimated incident duration into three categories: "<15 minutes", "15-30 minutes", and "30 minutes".

To contend with the stochastic natures of incident data and the issue of missing data in incident datasets, Ozbay and Noyan (2006) employed Bayesian Networks (BNs) to construct a dynamic decision tree by using the data that may contain partially incomplete information. Different from typical tree models, the authors demonstrated the dominance of utilizing BNs for the incident duration estimation due to its capability of generating stochastic nodes with the probabilities of each factor to impact the predicted incident duration that is jointly determined by various contributing factors. Moreover, Boyles et al. (2007) developed the naïve Bayesian classifier to tackle incomplete information collected from various time points. Their model was then calibrated by using the historic incident data from the Georgia Department of Transportation, showing that it can outperform the standard linear regression.

Machine Learning Method

Along with the advance in computing technologies, many recent studies utilized Machine Learning (ML) and Artificial Intelligence (AI) algorithms to investigate complex incident datasets and develop robust incident duration prediction models. Along this line, Wei and Lee (2007) exploited data fusion techniques and the Artificial Neural Network (ANN) to produce a time-sequential procedure to estimate incident duration, aiming to minimize the impact of data noises. Following a similar logic, Lee and Wei (2010) sequentially utilized ANNs and genetic algorithms to develop two models that produce the predicted incident duration from the time of the incident notification to its clearance, and the evaluation results indicate that their models well fit the actual incident duration data and effectively alleviate the impact of data noises. In addition, Park et al. (2015) employed the Bayesian ANN models and the pedagogical rule extraction algorithm (TREPAN) to generate an incident duration prediction model with comprehensive rules.

Multi-Technique Method

Most aforementioned studies, using one method to construct the incident duration models, often cannot yield a satisfactory level of accuracy, due to the complex natures of incident data. In response to the limitations of a single technique algorithm, some researchers proposed to employ more than one methodology to better estimate incident duration. Lin et al. (2004) explored the ordered probit model to estimate the incident duration lasting shorter than

60 minutes, while applied a rule-based supplemental module to manage the incidents with their durations longer than one hour. Similarly, Kim et al. (2008) utilized CTM, Rule-Based Tree Model (RBTM), and Discrete Choice Model (DCM) collectively for the estimation of the incident duration. Kim and Chang (2011) also developed similar hybrid models that consist of RBTM, Multi-Nomial Logit model (MNL), and Naïve Bayesian Classifier (NBC) to predict the incident duration. Lin et al. (2016) developed a hazard-based duration model (M5P-HBDM) and utilized HBDMs as the leaves of the M5P tree for enhancement of the M5P to provide a more reliable incident duration estimation.

Association Rule Mining Method

Won et al. (2018) proposed a knowledge-based system to estimate the range of the incident clearance duration. In their study, all incidents of I-95 in Maryland were divided into two categories: 1) shoulder-only blockage and 2) travel-lane blockage. The cases with travel-lane blockage were then further categorized by incident types (i.e., Collision with Fatality (CF), Collision with Personal Injury (CPI), and Collision with Property Damage (CPD)) and the number of travel lanes blocked (i.e., one travel lane blocked, two travel lanes blocked, and three and more than three travel lanes blocked). Then, the data in each category are first assigned into two categories of "<30 minutes" or "≥30 minutes", depending on the classification rules derived from historical incident data with the Association Rule Mining method. With the same approach, those incidents in the category of "≥30 minutes" are further classified into two sub-categories of "<60 minutes" and "≥60 minutes". Similarly, the cases predicted to be in the category of "≥60 minutes" are then classified into two sub-categories of "<120 minutes" and "≥120 minutes". Finally, according to the distribution of these categories, this system generates an interval-based estimate based on the historical data and produces several estimated duration estimations under different confidence levels for a detected incident. The evaluation results indicate that the proposed prediction method can achieve the accuracy of higher than 75% in both training and testing datasets, and the knowledge-based system has shown its promise to best use the invaluable historical information to circumvent many data quality and availability issues.

2.1.4 Outlier Issues

Note that the accuracy of those prediction models highly depends on data quality. For instance, Khattak et al. (2012) employed an OLS regression model to predict incident duration with around 37% of Mean Absolute Percentage Error, but the model generated unacceptable estimation results for the incidents with duration either shorter than 10 minutes or longer than 2 hours.

Additionally, Li (2015) utilized the hazard-based model to estimate most incidents with acceptable accuracy but produced unreasonable predictions for those incidents lasting less than 15 minutes or more than 2 hours. Without exception, the knowledge-based system, developed with the Association Rule Mining method proposed by Won et al. (2018) to circumvent data quality issues, remains adversely affected by outliers, especially for those incidents with the duration exceeding one hour.

To summarize, most existing studies failed to yield satisfactory prediction results for those incidents with extremely long or short durations, or those with unusual characteristics (Li et al., 2018). Hence, these viewed as outliers by the employed models should be detected from the incident datasets and separately analyzed with proper methods to capture their data patterns for a better incident duration prediction. In addition, information on these outliers can also be used to enhance data quality (Won, 2020). Therefore, identifying those outliers in the incident datasets is a critical issue in developing robust incident duration estimation models, which, unfortunately, has been tackled by only a few studies. For instance, Won (2020) exploited several outlier detection models, including Principle Component Analysis (PCA), Partitioning

Around Medoids (PAM), and Isolation Forest to compute the outlier score of each data point in the incident dataset, and then applied a hybrid association rule mining method to classify these outliers into anomalies and noises to improve the performance of an incident duration estimation model.

2.2. Outlier Analysis

2.2.1 Definition of Outliers

The existence of anomalies would inevitably reduce the quality of the incident data and degrade the performance of any model developed for estimating incident duration. According to Hawkins (1980), such anomalies can be defined as observations that deviate so much from other observations to arouse suspicion that these were generated by a different mechanism. Additionally, these observations are called outliers when the number them is significantly smaller than the proportion of normal cases, typically lower than 5% (Domingues et al., 2018). As shown in Figure 2-2, the referred dataset has two normal clusters, N1 and N2, and most cases are located in these two clusters, while those points being significantly far away from the regions (e.g., O1, O2, and the points in O3) are the outliers (Bansal et al., 2016). According to Won (2020), the outliers in the incident dataset mostly are those incident records correctly recorded but having unusual incident duration, due to their much different data patterns, while the others are likely from recording errors.

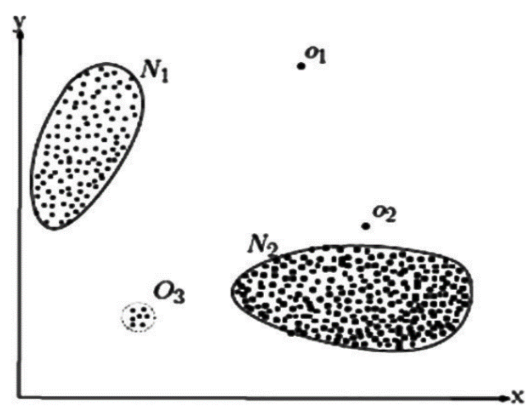


Figure 2-2. An Illustration of Outliers in a Two-dimensional Dataset (Bansal et al., 2016)

2.2.2 Outlier Detection Algorithm

Outlier detection is a notoriously hard task especially when these abnormal observations overlap with nominal clusters in which these clusters are densely distributed. Also, contamination problems (e.g., utilizing the dataset contaminated by outliers as input) make outlier detection even more critical because such anomalies may degrade the produced model's quality if its training algorithm lacks robustness. Hence, a large body of outlier detection algorithms has been reported in the literature, and some of those commonly used can be categorized into the following categories: distance-based methods, probabilistic methods, neighbor-based methods, domain-based methods, neural networks, and isolation methods.

Distance-Based Method

Among the methodologies for detecting outliers, utilizing the distance space to find out outliers is one of the most intuitive approaches, and this class of methods is called the distance-based algorithms. The K-Nearest Neighbor algorithm (k-NN) (Byers and Raftery, 1998), one of the commonly used methods, utilizes the distance from a data point to its *k*-nearest neighbor to indicate how anomalous the point is. However, the computational complexity could be prohibitive as the dimension and the size of data increase. Another popular method is based on Mahalanobis Distance, which can be applied to multivariate datasets constructed with a single normal-distributed cluster (Ben-Gal, 2005). However, identifying anomalies by solely using Mahalanobis Distance is not practically applicable for complex real-world data because it needs to work through the entire large and high dimensional dataset to identify those embedded attributes' co-relations.

Probabilistic Method

In response to such limitations embedded in Mahalanobis Distance methods, some studies adopted the concept of the likelihood to identify the outliers in a dataset. Given the model parameters θ, the probabilistic algorithms will be used to determine the probability density function of each parameter in a dataset where the observations having the smallest probability, $P(X|\theta)$, are detected as outliers. In review of the literature on this subject, it is noticeable that some studies proposed to first train the Gaussian Mixture Model (GMM) with the Expectation-Maximization algorithm (EM) (Dempster et al., 1977), and then to fit the target dataset with the given Gaussian distributions with different parameters to identify outliers.

To cope with the complexity of determining the number of parameters in the Gaussian distribution, Blei and Jordan (2006) proposed a nonparametric Bayesian algorithm method, called the Dirichlet Process Mixture Model (DPMM) that optimizes the model parameters and verifies the convergence by tracking a non-decreasing lower bound of the log-marginal likelihood. By doing so, the outlier detection method can be applied without any preset parameter settings.

Along the same line, Parzen (1962) developed kernel density estimators (KDE) to assign a kernel function to every observation in a dataset and then computed the sum of the local contributions of the kernels for approximating the density function of the dataset. However, KDE is so sensitive to the outliers that the performance of the model would be impacted adversely by the data contamination problem (e.g., input data contaminated by outliers). Hence, Kim and Scott (2008) constructed a Robust Kernel Density Estimator (RKDE) to overcome the defects of the standard KDE by utilizing M-estimation methods (i.e., robust loss functions).

For the same objective of outlier detection but with a different approach, Tipping and Bishop (1999) employed a latent variable model, called Probabilistic Principal Component Analysis (PPCA), to assess the principal components in a dataset and utilized the log-likelihood function to show the degree of outlier-ness for a new observation.

Neighbor-Based Method

Some other studies, making no prior assumption about the data, model the outliers as the data points that are isolated from their surrounding neighborhood. These studies are categorized as the set of neighbor-based algorithms. Along this line, Kriegel et al., (2009b) proposed the Subspace Outlier Detection (SOD) to find the set of neighbors, which are shared by those data points and their *k*-nearest neighbors, and then take the standard deviation of a data point from the mean of a given subspace (i.e., the set of its neighbors) as its outlier score. Following the same logic, Breunig et al. (2000) defined the Local Outlier Factor (LOF) to estimate the outlier score, based on the distance between a data point and its locally reachable neighborhood assigned to each data point. The LOF value of a data point demonstrates the contrast between its density and those of its neighborhoods.

Note that both k-NN and LOF suffer from the parameter-setting problem. To be more specific, it is quite difficult to select appropriate parameters for a real-world dataset without labeling real outliers, especially for the scattered datasets. To assess the outlier-ness of an object in a scattered dataset, Zhang et al. (2009) defined a Local Distance-based Outlier Factor (LDOF) to determine the degree of an object deviating from its neighborhood by using the relative location of the object to its neighbors. Compared to k-NN and LOF, LDOF is relatively sensitive to the outliers in scattered datasets and performs more stable under a large range of parameter values. In other words, compared to k-NN, it is relatively convenient to identify a set of proper parameters for LDOF to ensure its effectiveness. Moreover, intending to enhance the effectiveness of LOF (Breunig et al., 2000), Kriegel et al. (2009) proposed the method of Local Outlier Probabilities (LoOP) to interpret the value of a “factor” as the probability of a data point to be an outlier. Also, in response to the parameter setting issue, Kriegel et al. (2008) constructed Angle-Based Outlier Detection (ABOD) that utilizes the radius and variance of angles computed at each input vector to detect outliers.

Domain-Based Method

In addition to the aforementioned algorithms, some novel techniques have been adopted by some researchers for outlier detection. For example, Shyu et al. (2003) employed the Robust Principal Component Classifier, which measures the distance of an outlier from the normal data points in the principal component space, composed of the major and the minor principal components of normal instances, to estimate its outlier-ness. However, the Robust Principal Component Classifier can only be built under the assumption of having the normally distributed data points. Schölkopf et al. (2000) applied the method of support vector machine (SVM) algorithms to one-class problems (One-class SVM). The core of this algorithm is to compute a separate hyperplane and the boundaries that can best fit the input data and utilize the principle of maximized the margin from the origin to define those observations outside the boundaries as outliers.

Neural Network

For detecting the outliers, Muñoz and Muruzábal (1998) employed the algorithm of Self-Organization Map (SOM) which is a kind of neural networks and a dimension reduction method. It maps the original data points onto a certain number of points (Kohonen, 1990), and computes the SOM quantization errors as outlier scores for outlier detection.

Isolation Method

For the same purpose of detecting outliers, Liu et al. (2008) proposed the method of Isolation Forest, which utilized random forests to measure an isolation score for each observation in a dataset. Their evaluation with respect to the performance of Isolation Forest only focused on the dataset without categorical variables.

2.3. Summary

This chapter has reviewed several popular models on incident duration prediction and outlier detection algorithms, and also identified some areas for further improvement. Regarding the incident duration prediction, it is noticeable that most existing incident prediction models overlook some common key factors associated with incident response and operations on different highways/freeways. Effectively capturing the individual and collective impacts of those key factors on the incident response and operations may offer an avenue for developing a generalized, rather than roadway-specific, model for estimating the required incident duration for the highway network within the same region.

With respect to the potential impacts of data outliers, although many of those models in the literature can produce an acceptable prediction for those incidents with a normal range of clearance duration, they are not able to yield sufficiently reliable estimates for incidents with extremely long or short clearance durations, because of the presence of the outliers. However, only a few studies in the literature were proposed to address the outlier detection issue in developing a robust system for estimating the duration needed to clear various types of highway incidents.

As such, this study, in view of the aforementioned issues, intends to build an extended model, grounded in the work by Won et al. (2018), that can circumvent the impacts by some outlier data, and achieve better prediction accuracy with less data-imputation and classification efforts.

Chapter 3: Model Transferability Analysis: Extending from the I-95 Incident Duration Model to the I-695 System

In view of the large number of factors contributing to the incident duration and the need of a sufficient sample size for model development, this chapter presents a methodology for transferring the information from the knowledge-based I-95 model, developed by Won et al. (2018), to new models of same purposes for different highways sharing similar roadway and operational features. This chapter first reviews the core logic embedded in the I-95 system, followed by a detailed illustration of the proposed methodology for model transferability analysis.

3.1. *Core Logic for the I-95 Model*

As shown in Figure 3-1, the process of developing a knowledge-based prediction model, such as the I-95 system, comprises the following stages: 1) **incident data pre-processing**, 2) **incident categorization**, 3) **classification rule mining process**, and 4) **assessment of estimation accuracy for incident clearance time**.

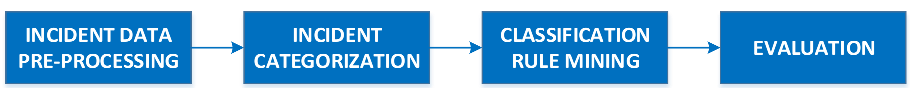


Figure 3-1. Flowchart of the Model Development Process by Won et al. (2018)

3.1.1 Incident Data Pre-Processing

The objective of the first stage is to remove the obvious data errors made by system operators during the real-time incident response and management process. The pre-processing for data quality analysis can be done with the following steps:

- **Step-1**: Remove those incident records showing unrealistically short duration (e.g., 5 minutes) between their event-cleared and all-blocked-lane-reopened times (including shoulders) from the dataset;

- **Step-2**: Remove those incident records exhibiting inconsistency between their incident clear times and all-blocked-lane-reopened times; and
- **Step-3**: Remove those incident records with unreasonably short incident clearance times for any type of lane blockage events.

3.1.2 Incident Categorization

After pre-processing the dataset, the available incident records are separated into several categories based on the incident nature and the number of blocked lanes. As shown in Figure 3-2, all selected incident records are firstly classified into two classes: with travel-lane blockage and with only shoulder-lane blockage. Those with travel-lane blockage are then further partitioned into Collision with Fatality (CF), Collision with Personal Injury (CPI), and Collision with Property Damage (CPD). Following the same partition logic, those in the subsets of CPI and CPD are further categorized into six categories, depending on the number of travel lanes blocked (e.g., CPI1 is a subset that contains the collisions with personal injury and one-travel-lane blockage). With such a sequential partition process, the entire set of incident records will be divided into the following seven categories, CF, CPI1, CPI2, CPI3+, CPD1, CPD2, and CPD3+.

Notably, due to the small sample size and different CT patterns, those records in CF are not further divided. So are those incidents with only shoulder blockage.

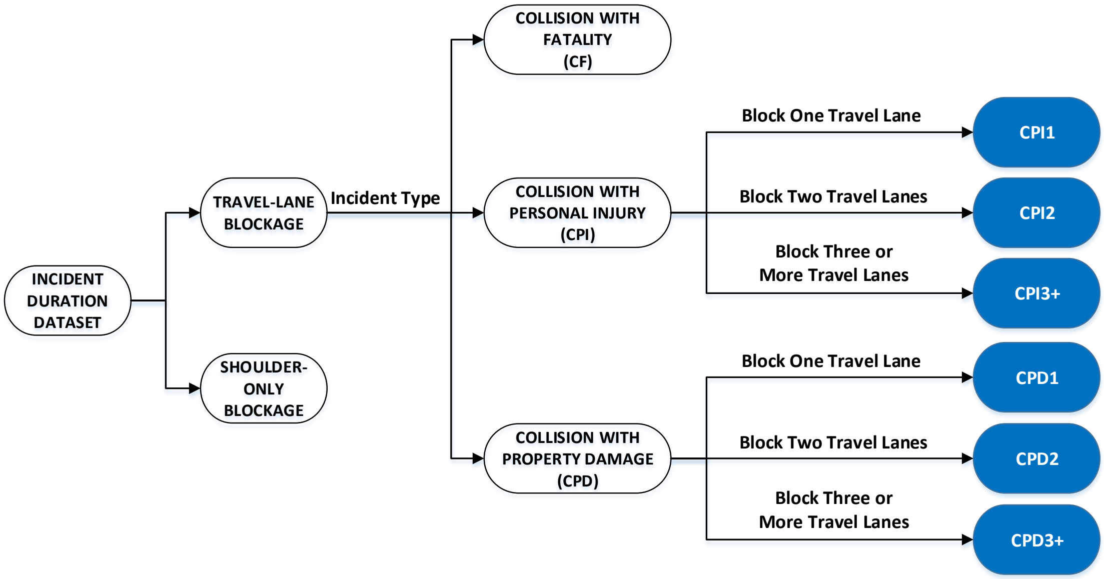


Figure 3-2. Incident Categorization Tree

3.1.3 Classification Rule Mining Process

After completing the categorization task, Won et al. (2018) employed the Association Rule Mining method for each pre-classified category to identify the common characteristics from those included incident records with a sequence of IF-THEN rules for approximating the resulting incident clearance times. As demonstrated in Figure 3-3, the procedures for searching new classification rules are presented as follows:

- **Step-1**: Construct the classification rules to classify the data into two distinct groups of "<30 minutes" and "≥30 minutes" by using the Association Rule Mining method;
- **Step-2**: Select a rule that achieves a confidence level of higher than 75% and has the highest support level;
- **Step-3**: Filter out the incident records associated with the selected rule from the dataset; and
- **Step-4**: Stop the process if no further rule can be excavated to classify the remaining incident records; otherwise, go to Step-1.

With such aforementioned procedures, for each pre-classified category, incident records are firstly divided into two groups of "<30 minutes" or "≥30 minutes", and for those in the group of "≥30 minutes" subset are then further classified into two subgroups of "<60 minutes" and "≥60 minutes". Along with the same sequential logic, those in the "≥60 minutes" subset are further classified into two smaller subgroups of "<120 minutes" and "≥120 minutes".

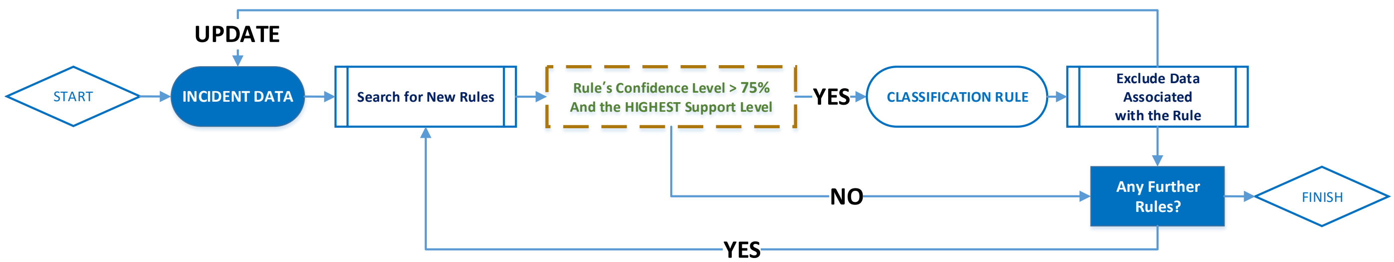


Figure 3-3. Classification Rule Mining Process

3.1.4 Assessment and Estimation of Incident Clearance Time

Finally, the model with a tree structure can be constructed for performing the prediction of a detected incident's estimated clearance duration under 60%, 70%, and 80% confidence levels. Notably, due to the small sample size and much longer clearance time (CT), the rules for those incidents involving collisions with fatality (CF) are generated with a different procedure. Similarly, the events with only shoulder blockage are not further classified because most of their CTs are shorter than 30 minutes.

Although the above knowledge-based model developed by Won et al. (2018) has been used in practice and reported to yield acceptable performance. Its development not only is quite time-consuming but also demands a sufficiently large sample of incident records to well capture the complex interrelations between the resulting clearance times of a detected incident and all contributing factors, including traffic and environmental conditions, as well as the resources and efficiency of responsible agencies. As such, it may suffer from both the model accuracy and coverage issues if one intends to directly apply the same development methodology to some other highways with insufficient incident records. Hence, to circumvent the data limitation issue and to take advantage of valuable information embedded in the I-95 system's prediction rules, the following section will present a methodology for the model transferability analysis. The proposed methodology allows a developer to evaluate and select applicable prediction rules from the I-95 system for the new highway system, and then apply rule-search method improved from Won's work to tackle the remaining incident records that reflect local-specific traffic and response features.

3.2. *Extended Model Development with the Prediction Rule-Transferring Process*

The proposed process intends to take advantage of the common characteristics between different roadways so as to transfer applicable prediction rules from the well-developed I-95 system to any other highway. Figure 3-4 illustrates the process for constructing an extended model for a new target highway. The entire process consists of the following phases: 1) **incident data pre-processing**, 2) **incident categorization**, 3) **classification rule partitioning and ordering**, 4) **automatic rule evaluation and transferring process**, 5) **new classification rule mining**, and 6) **interval-based estimation of incident clearance time**. A brief description of these activities included in each phase is presented below.

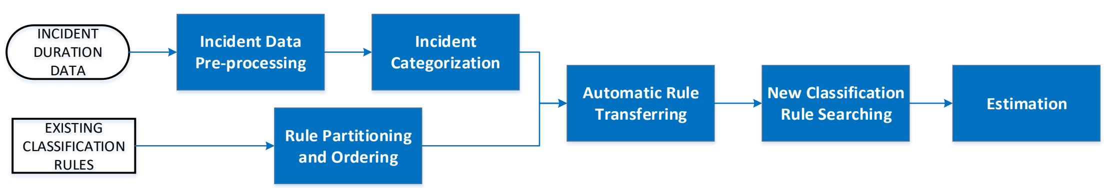


Figure 3-4. Flowchart of the Extended Incident Clearance Time Estimation Model

3.2.1 Incident Data Pre-Processing and Incident Categorization

To remove obvious data errors, the same incident data pre-processing criteria proposed by Won et al. (2018) can also be adopted in this phase. This task will then be followed by categorizing all target incident data based on the incident type and status of lane blockage, as used by Won et al. (2018).

3.2.2 Classification Rule Partitioning and Ordering

Since the classification rules from the previously developed model are used to classify incident data in each pre-classified category, one of the most critical tasks is to well sort out the classification rules from previously developed models based on their predicted results. As shown in Figure 3-5, the classification rules from the previously developed model for each pre-classified category can be partitioned into three groups: 1) a group of rules determining whether incident CT would be shorter or longer than 30 minutes, 2) a group of rules determining whether incident CT would be shorter or longer than 60 minutes, and 3) a group of rules determining whether incident CT would be shorter or longer than 120 minutes.

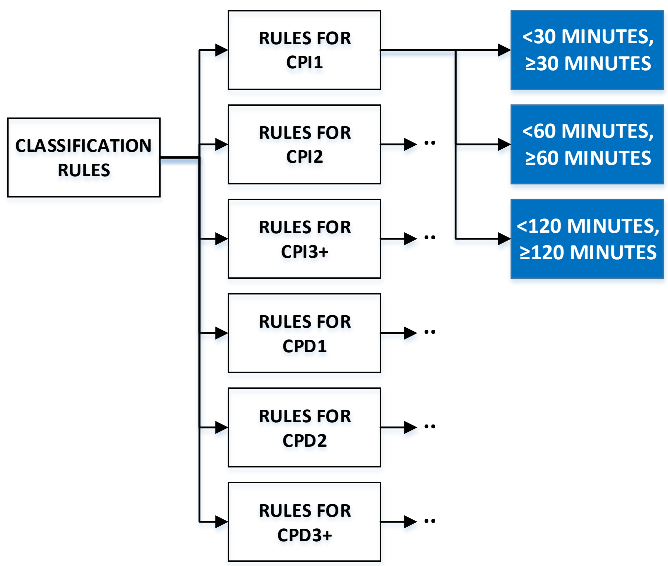


Figure 3-5. The Rule Partitioning Tree

Additionally, to best transfer each classification rules, it is essential to have an effective process to determine the priority of transferring analysis from those available rules based on the data characteristics of the target roadway because the final system's prediction accuracy may vary with not only the selected rules for model transfer but also their sequence in the prediction tree structure. To do so, the relative importance of contributing factors to the target category of incidents needs to be explored, and their resulting rankings of impacts will be adopted as the basis for sequentially evaluating and transferring rules from the previously developed model to a new target system.

All contributing factors to the incident CT are initially classified into 7 major categories, "**the Number of Responded Units**", "**the First Arrival of the Responded Units**", "**Vehicles Involved**", "**Pavement Condition**", "**Lanes**", "**Response Center**", and "**Time of a Day**". Then, the permutation-based variable-importance measure is utilized to evaluate the importance of each with the Random Forest method (Breiman, 2001; Fisher et al., 2018). With the relative importance of each contributing factor measured by its estimated impact on the resulting duration of incidents on the target roadway, one can then determine the rank of these seven pre-classified categories, and then perform the rule evaluation and transferability analysis in descending order with the following scoring procedures:

- Take the ranking of each pre-classified category as its assigned score (e.g., score equals 1 for the 1st rank) for each individual rule;
- If a rule is a combination of several individual rules with "AND", then sum up the score of each individual rule as the final score of this integrated rule;
- If a rule is a combination of several individual rules with "OR", then take the minimum score among those individual rules and add a number of 100 for defining its role in the new system; and
- If the estimated result with one particular rule is shorter than those specified bounds for categorization (e.g., <30 minutes, <60 minutes, and <120 minutes), then take the minimum score among those individual rules and add a number of 200 for defining its role in the new system.

Taking I-695 incident data as an example, with the aforementioned procedure, the importance of each contributing factor is obtained and shown in the left part of Figure 3-6. It shows that the most important contributing factor is "*the total number of the responded units arrived*" which belongs to "**the Number of Responded Units**" category, so the rank of "**the Number of Responded Units**" is 1. Excluding those factors related to "**the Number of Responded Units**", the second important factor is "*the number of trucks involved*" which belongs to "**Vehicles Involved**" category, thus this category of factors is ranked 2. Following the same logic, one can rank all remaining categories, as shown in the right part of Figure 3-6. With the computed rank of each pre-specified category, one can use the scoring system to determine the sequence of the rule transferring process. For instance, in Table 3-1, the rule, "*More than 8 responded units arrived*", belongs to "**the Number of Responded Units**" category, so it is given one point. For the rule, *"[Weekend] AND [Fireboard first arrived]"*, its "*Weekend*" belonging to "**Time of a Day**" category is given 5 points, and its "*Fireboard first arrived*" residing in "**the First Arrival of the Responded Units**" category is given 6 points. Thus, the score of this integrated rule is 11, which is the summation of the scores of these two individual rules. Since the rule, "*[Weekend] OR [Night]*", is combined with "OR" and the smallest score between these two individual rules is 5 points, the score of such an integrated rule is set to be 105. Besides, although the rule, "*No tow service arrived*", belongs to "**the Number of Responded Units**" category, its estimation result is shorter than 60 minutes, thus the score of this rule is set to be 201. Based on the same logic, the scores and transferring orders of all available rules in the candidate model for transferability analysis can be obtained, as shown in Table 3-1.

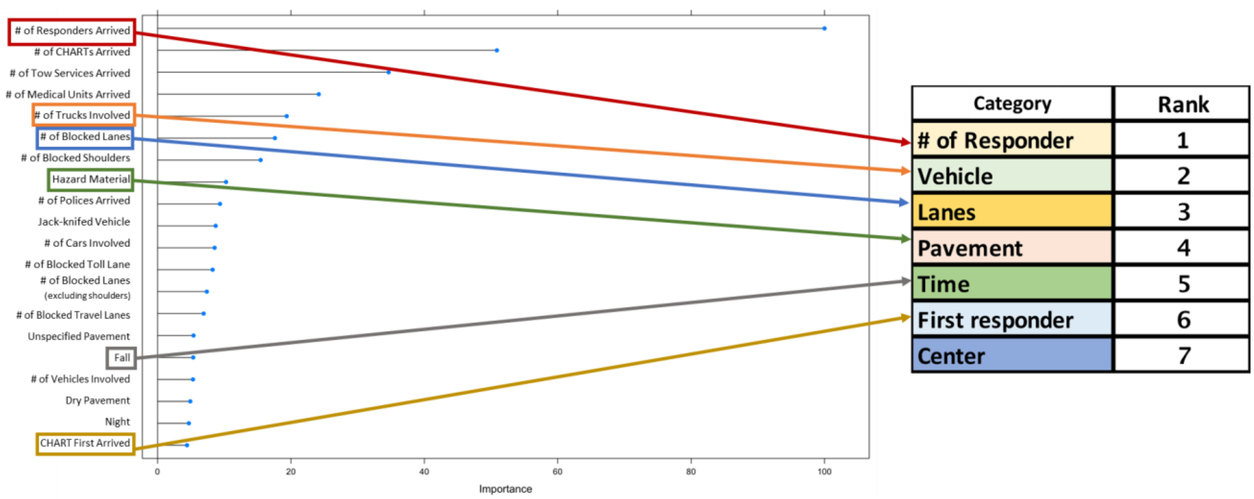


Figure 3-6. Ranking for the Contributing Factor Categories

Table 3-1. An Example of the Assigned Score and Transferring Order for Rule Classification

Rules for CPI3	Estimation	Score	Transferring Order
[More than 8 responded units arrived]	>60	1	1
[Medical service arrived]	>60	1	2
[No tow service arrived]	<60	201	9
[More than 3 travel lane blocked]	>60	3	3
[Weekend] OR [Night]	>60	105	8
[Dry pavement]	<60	204	10
([More than 5 responded units arrived] OR [Winter] OR [More than 1 tow service arrived]) AND [More than 2 vehicles involved]	>60	103	7
[Weekend] AND [Fireboard first arrived]	>60	11	5
[Holiday] AND [Truck involved]	>60	7	4
([More than 5 responded units arrived] OR [More than 1 CHART arrived]) AND [More than 1 tow service arrived]	>60	102	6

Conceivably, those compound rules embedded with "OR" should be transferred in a lower priority than those individual rules, because they can be adopted for transferring under fewer constraints than the others but may not yield the expected accuracy. Besides, those rules for estimating those incidents with their durations shorter than the specified bound for each category (e.g., <30 minutes, <60 minutes, and <120 minutes) will also be in a lower priority of the transferability analysis than those rules used to extending the tree-classification structure for identifying incidents of longer duration.

Note that the transferring order of each rule (See Table 3-1) is used to set the sequence of those rules to be examined in the next phase of automatic rule evaluation and transferring.

3.2.3 Automatic Rule Evaluation and Transferring Process

After determining the transferring order, each rule in the previously developed model will be examined in the sequence defined in the last section to verify if it can be transferred to yield the prespecified level of accuracy for the new target highway. Such transferability assessment is based on 1) the **confidence level** that demonstrates the predicting accuracy of a classification rule, and 2) the **support level** that shows the percentage of incident records covered by the candidate rule. Those rules yielding a sufficiently high confidence level and covering a reasonable number of cases will be deemed transferable and included in the model for the target highway. One can follow the same process until all existing rules have been examined in their assigned evaluation sequence. As shown in Figure 3-7, the proposed automatic rule transferring process can be illustrated in the following steps:

- **Step-1**: Pre-determine the minimum confidence level (X%), and the lower bound (S_L%), and upper bound (S_U%) of the support level for transferability analysis for each candidate rule;
- **Step-2**: Utilize the data in each pre-classified category in the target highway's incident data to verify the transferability of each candidate classification rule based on their assigned sequence for transferability analysis;
- **Step-3**: Transfer the candidate classification rule if it has a confidence level higher than the acceptable percentage and the support level within the specified interval;
- **Step-4**: Update the target highway's dataset for further transferability analysis by filtering out those incident records associated with the rules with the acceptable transferability; and

- **Step-5**: Stop the transferring process if no more classification rules for transferability analysis. Otherwise, go to Step-2.

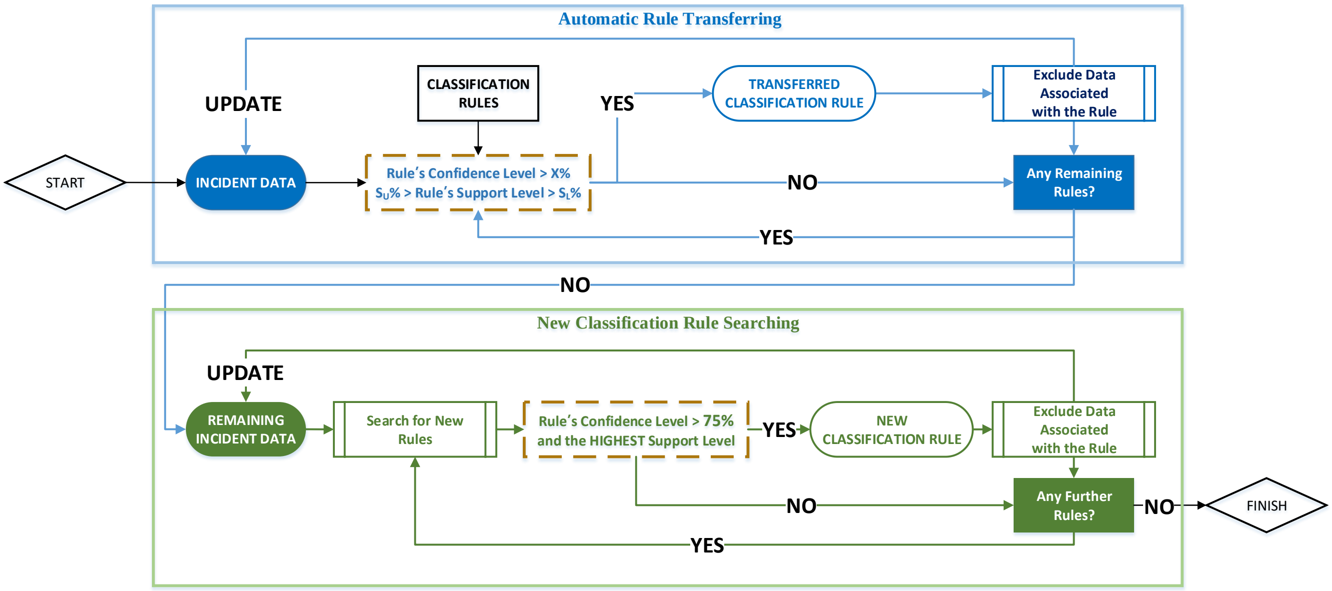


Figure 3-7. Automatic Rule Transferring and New Rule Searching Processes

For each group of classification rules, the acceptable confidence level and the range of support level can be determined based on the target incident data size and features. Such an automatic rule transferring process with adjustable criteria enables users to best balance the number of transferred rules and their effectiveness to develop a new incident duration prediction model for another highway in a cost-efficient way. Additionally, it not only can best reflect the common features of key factors contributing to the clearance time among different highways, but also circumvent the difficulty in the model development for those highways with insufficient incident records.

3.2.4 New Classification Rule Searching Process

Conceivably, after the transferring process, some of those incidents' resulting durations on the new target highway may not be predicted at the desirable level of accuracy with any of those transferred rules, due likely to the impacts of local specific factors on the response and clearance operations. Hence, one shall employ the same rule-generation process used by Won et al. (2018) based on the Association Rule Mining method presented in the first section, to develop some local-specific prediction rules for those remaining unclassified incident records. Figure 3-7 illustrates the whole process integrating the automatic rule transferring process with the new classification rule searching process.

Notably, to further enhance the developed incident duration prediction model's applicability in practice, in addition to further design of local-specific prediction rules, this study has adopted those rules of likely resulting in overestimate than underestimate in the tradeoff scenarios during the model transferring and development process. This is proposed in response to the concerns of the incident response agencies because the message of conveying an overestimated incident clearance time is less likely to be complained by the motorists than the underreported one.

In brief, to better illustrate the aforementioned model development and transferring process, the incident records in the CPI1 subset is used hereafter as an example to demonstrate its application. As shown in Figure 3-8, the first step is to test the applicability of the existing classification rules from the existing prediction models to the new target roadway with the evaluation sequence calibrated from the target incident data set. Then, those classification rules that meet the criteria can be transferred to the candidate model, and the remaining incident records are further explored with the Association Rule Mining method to generate the new set of supplemental local-specific classification rules. For instance, after incident records in the CPI1 category are classified into "< 30 minutes" and "≥ 30 minutes" groups with the transferred and new prediction rules, those classified as "≥ 30 minutes" are further separated into "< 60 minutes" and "≥ 60 minutes" groups by employing the same steps. With the same procedures, the incidents classified as "≥ 60 minutes" are then finally divided into "< 120 minutes" and "≥ 120 minutes" groups.

By employing the same procedures to analyze incident records in other pre-classified categories, all incident events in each category can be assigned into the following four groups: 1) "CT < 30 minutes", 2) "30 minutes ≤ CT < 60 minutes", 3) "60 minutes ≤ CT <120 minutes", and 4) "120 minutes ≤ CT". For each group, the average CT and three different expected incident clearance times with the 60%, 70%, and 80% of probability would be computed for further estimation.

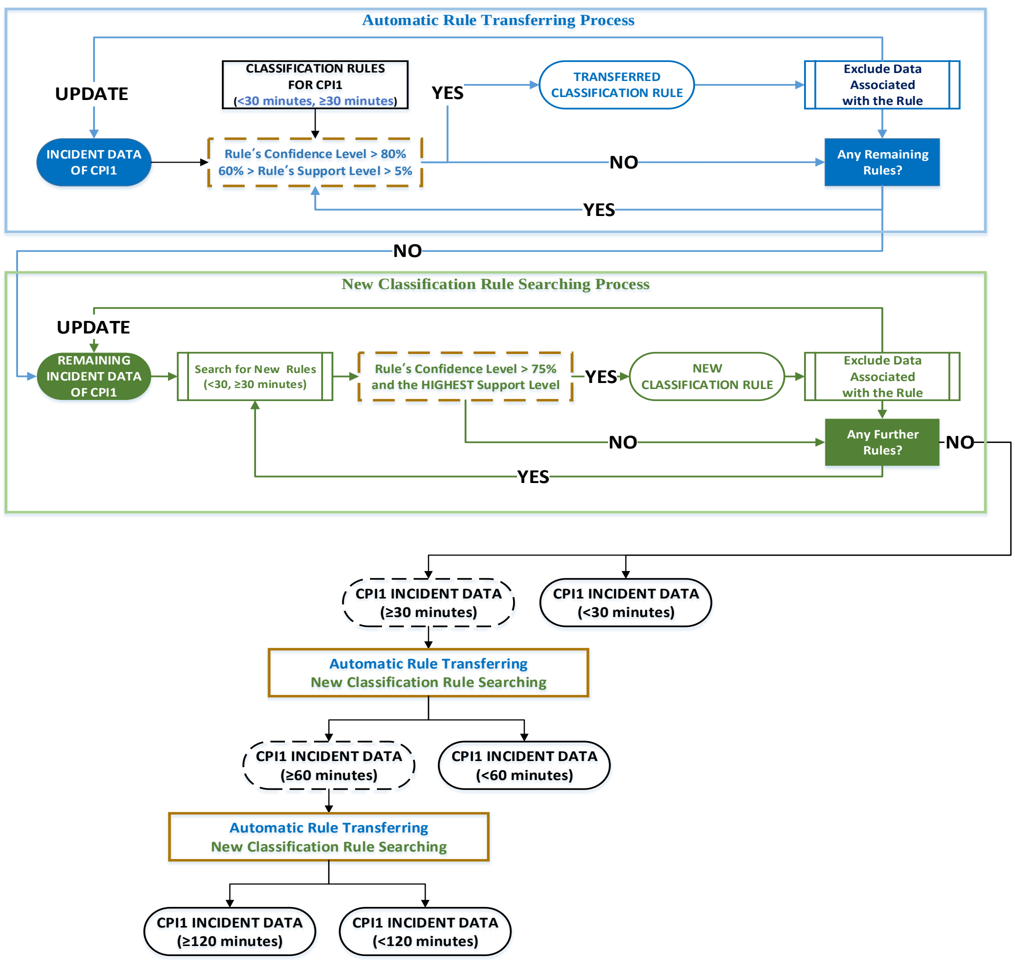


Figure 3-8. An Example of the Automatic Rule Transferring Process and the New Classification Rule Searching Process for CPI1 Incident Data with One Travel Lane Blocked

Such sequential processes for the automatic rule transferring and the new classification rule searching can efficiently capture the individual and collective impacts of those key factors on the incident response and operations, offering an avenue for developing a generalized, rather than roadway-specific, incident duration model for other highway networks within the same region.

3.3. Discussion

This study, grounded in the logic of the knowledge-based model proposed by Won et al. (2018), has developed an effective process for model transferability analysis that allows the users to reduce much effort for new model development and for circumventing the data insufficient issue. However, as with most existing prediction models in the literature, the proposed rule-based model often suffers from the impact of those data anomalies in the incident records which are either outliers or the results of recording errors. Such an impact on the model's prediction accuracy is especially pronounced for those incident categories with relatively small sample records. Hence, this study has further adopted several outlier identification methods and reported them in the next chapter to refine model development and improve its prediction accuracy.

Chapter 4: Outlier Detection Methodology

4.1. <u>Modeling Concept</u>

The transferring procedures proposed in Chapter 3 offer an effective way for traffic engineers to develop an incident duration prediction model for a new highway with limited incident records. This chapter presents one more set of development procedures to improve the accuracy of a newly developed model, that is, to identify possible outliers from the available data records.

Outliers are referred to those incident records showing quite different patterns from others, which constitute mostly a small portion of the data but can adversely affect the developed model's prediction accuracy (Won, 2020). Hence, this chapter hereafter presents the core logic of two outlier detection methods adopted to complement the development of a new incident duration prediction model.

4.2. <u>Data Characteristics of Incident Records</u>

As noted in the work by Won et al. (2018), the clearance time in most incident records lasts shorter than 60 minutes, and only less than 10% of incidents' clearance times exceed 120 minutes. Taking the incident records of I-695 as an example, it is noticeable from Figure 4-1 that incident records with longer durations are scattered in several mini-clusters in a sharply skewed distribution, as marked in the red circle in Figure 4-1. Due to the unique characteristics and distribution of those long-duration incident records, the commonly used outlier detection approaches (e.g., k-NN, LOF) with the assumption of having most data points distributed in a few main clusters are not applicable.

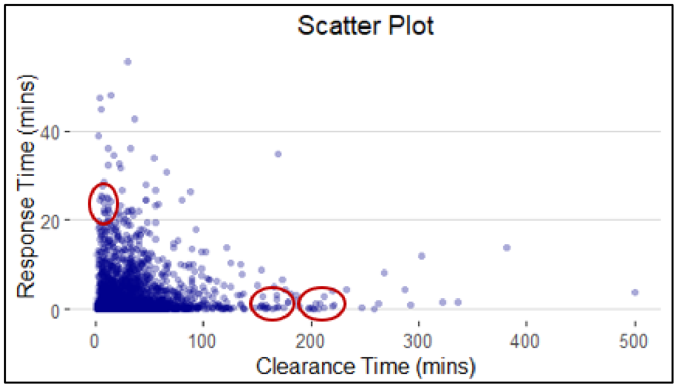


Figure 4-1. Scatter Plot of Response Time and Incident Clearance Time

Moreover, Figure 4-2 reveals that the density curves of key variables in the incident dataset of I-695 in MD are not normally distributed. Thus, some robust algorithms (e.g., Robust PCA), conditioned on the assumption of most key factors being normally distributed in the dataset, are not suitable for use to detect the outliers for the incident data set.

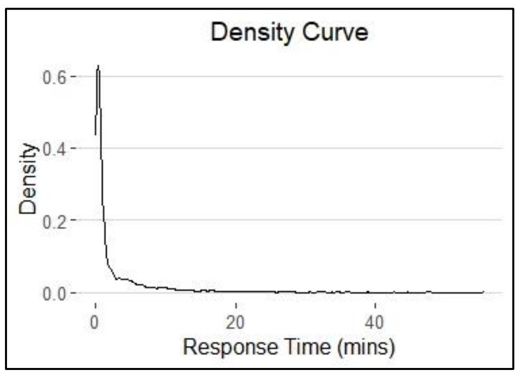


Figure 4-2. Density Curve of Response Time

In addition, the fact that contributing factors associated with incident duration include both various continuous (e.g., response time) and categorical factors (e.g., incident type, pavement condition) (Won et al., 2018) have excluded some algorithms, such as the Isolation Forest algorithm (Liu et al., 2008) from the selection list.

Aside from the characteristics of the incident data that contribute to the difficulties in outlier detection, the lack of prior knowledge about genuine anomalies has rendered outlier detection for incident data as an unsupervised learning problem. Hence, it is difficult to set correct parameters for most outlier detection algorithms from the incident datasets, and hard to evaluate their outlier detection results without the confirmation by the domain experts (Zhang et al., 2009). Additionally, it is also challenging to interpret the outlier score computed from most outlier detection approaches, and to determine whether the observation in the dataset is indeed an outlier (Kriegel et al., 2009a).

Note that the variation of key incident factors associated with the incident duration over time also plays an important role in the selection and design of the effective method for outlier detection. For instance, as shown in Figure 4-3, the density of "*the number of responded units arrived*" in the training dataset (the data from 2016 to 2018) compared to the testing dataset (data in 2019) exhibits some differences in their patterns. Such natural temporal variation embedded in the key variables of the incident records needs to be addressed in the design of thresholds for identifying outliers.

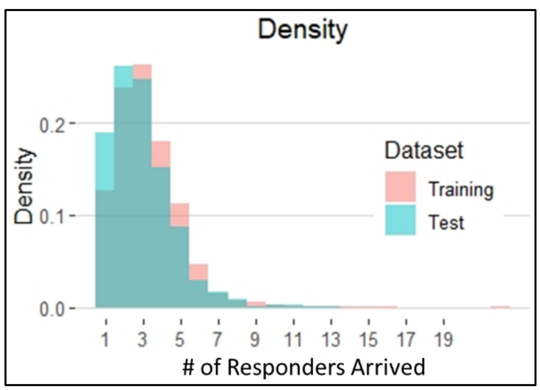


Figure 4-3. Density of "the Number of the Responded Units Arrived"

In brief, considering the unique characteristics and distributions of key factors associated with the incident clearance time, the selection of outlier detection method for incident records should take into account the following critical issues:

1) Those incident records are scattered in a sharp-skewed distribution and into several mini-clusters, especially for cases with long duration;

2) Most key factors associated with the incident clearance time in the incident dataset are not normally distributed;
3) The required clearance duration of a detected incident is affected by both continuous and categorical factors;
4) No prior knowledge associated with the characteristics and nature of true outliers in the incident records exist in the literature and state of the practices;
5) The difficulty in selecting the proper thresholds from the incident dataset to identify its anomalies; and
6) The inevitable year-to-year variation embedded in most key factors in the incident records often poses the challenge for an algorithm to correctly identify the data outliers.

4.3. Modeling Methodology

To contend with those issues identified in the last section for improving a model's prediction accuracy, this study develops an outlier detection model consisting of two phases as shown in Figure 4-4. Considering the unique characteristics of incident datasets, Phase I aims to properly identify the outliers with the training dataset. For mitigating the impact of the yearly data variation and concurrently accounting for other critical issues, Phase II targets to train the model for effectively capturing outliers in the test dataset by using the cleansed dataset from Phase I. The dataset processed through Phase I is then used to refine the original model that will be evaluated with the test dataset after it has been processed through the outlier detection mechanism in Phase II.

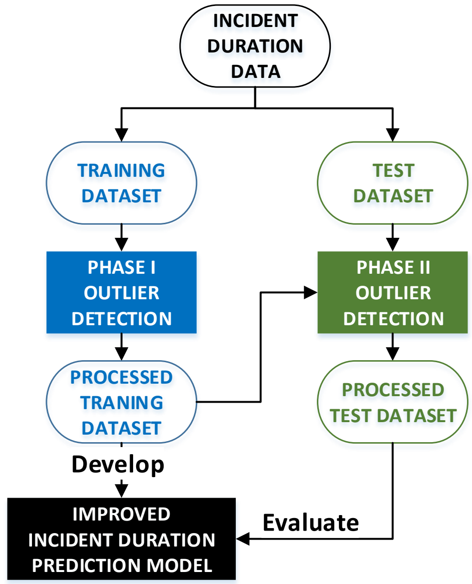


Figure 4-4. Structure of the Proposed Outlier Detection Model

4.3.1 Outlier Detection Methods in Phase I

Although various types of outlier detection approaches are reviewed in Chapter 2, each method is best only for those datasets with characteristics consistent with its underlying assumptions. Hence, in view of various technical issues identified in Section 4.2, this study proposes to use the following two complementary detection algorithms to perform the outlier detection: The Local Distance-based Outlier Factor (LDOF) and the Local Outlier Probability (LoOP). A summary of their strengths to tackle various identified data issues is shown in Table 4-1.

Table 4-1. Advantages of Selected Methods

Advantage	**LDOF**	**LoOP**
Be sensitive to outliers in the scattered dataset	V	
Without previous assumptions	V	V
Have only one parameter and perform stably with a wide range of parameter values	V	V
Can work with both categorical and continuous features	V	V
Provide interpretable outlier scores		V

The LDOF method is a local density-based outlier detection approach proposed by Zhang et al. (2009) that utilizes the relative location of a data point to its neighbors to determine the degree to which the point deviates from its neighborhood. The procedures for computing the LDOF value consist of the following steps:

- **Step-1:** Given a parameter $\boldsymbol{k}$, calculate the **$\boldsymbol{k}$-nearest neighbor distance** of object $\boldsymbol{x_p}$, defined as $\overline{\boldsymbol{d}_{\boldsymbol{x_p}}}$ by using Eq. (1), which equals the average distance from object $\boldsymbol{x_p}$ to $\mathcal{N}_{\boldsymbol{p}}$ defined as the set of the $\boldsymbol{k}$-nearest neighbors of object $\boldsymbol{x_p}$ (excluding object $\boldsymbol{x_p}$)

$$\overline{\boldsymbol{d}_{\boldsymbol{x_p}}} \coloneqq \frac{1}{k} \sum_{x_i \in \mathcal{N}_p} dist(x_i, x_p) \tag{1}$$

- **Step-2:** Given $\mathcal{N}_{\boldsymbol{p}}$ of object $\boldsymbol{x_p}$, compute the **$\boldsymbol{k}$-nearest neighbor inner distance** of object $\boldsymbol{x_p}$, defined as $\overline{\boldsymbol{D}_{\boldsymbol{x_p}}}$ by using Eq. (2), which is the average distance among object in $\mathcal{N}_{\boldsymbol{p}}$

$$\overline{\boldsymbol{D}_{\boldsymbol{x_p}}} \coloneqq \frac{1}{k(k-1)} \sum_{x_i, x_{i'} \in \mathcal{N}_p, i \neq i'} dist(x_i, x_{i'}) \tag{2}$$

- **Step-3:** Compute the local distance-based outlier factor (LDOF) of object $\boldsymbol{x_p}$ by using Eq. (3)

$$LDOF_k(\boldsymbol{x_p}) \coloneqq \frac{\overline{\boldsymbol{d}_{\boldsymbol{x_p}}}}{\overline{\boldsymbol{D}_{\boldsymbol{x_p}}}} \tag{3}$$

In the LDOF algorithm, there is only one tuning parameter, $\boldsymbol{k}$, which is used for each data point to compute the LDOF value that indicates how far the data point lies outside its neighborhood system.

Notably, different from other outlier detection methods, the LDOF method can identify outliers in the scattered dataset. Also, it does not require any preset assumption and can work effectively with the dataset having both categorical and continuous variables. In addition, the parameter setting in the LDOF method is relatively robust, as it can be applicable in a wide range. However, despite the robust nature and effectiveness of the LDOF method, no physical or statistical implications can be attached with its computed outliers, which poses the difficulty in setting a numeric threshold for identifying outliers. As such, this study proposes the concurrent use of the LoOP method to set the thresholds for outlier detection because its computed outlier score can be interpreted as the probability for the target point to be an outlier.

The LoOP method is presented by Kriegel et al. (2009a) in which its scoring of outlier-ness is associated with a probability. The procedures for computing the LDOF value for the target dataset consist of the following steps:

- **Step-1:** Given a parameter ***k***, compute a **standard distance** of object ***o*** to the object (***s***) in the set of its ***k*-**nearest neighbors (***S(o)***) by using Eq. (4), which is similar to the standard deviation

$$\sigma\big(\mathrm{o}, \mathrm{S}(\mathrm{o})\big) \coloneqq \sqrt{\frac{\sum_{s \in S} distance(o, s)^2}{|S(o)|}} \tag{4}$$

- **Step-2:** Under 99.7% confidence interval, compute the **probabilistic set distance** of object ***o*** to ***S(o)*** by using Eq. (5)

$$pdist\big(o, S(o)\big) \coloneqq 3 \cdot \sigma\big(\mathrm{o}, \mathrm{S}(\mathrm{o})\big) \tag{5}$$

- **Step-3:** Compute the **probabilistic local outlier factor (PLOF)** of object ***o*** with respect to S(o) by using Eq. (6) to estimate the density around object ***o***

$$PLOF_{S(o)} \coloneqq \frac{pdist\big(o, S(o)\big)}{E_{s \in S(o)}\big[pdist\big(s, S(s)\big)\big]} - 1 \tag{6}$$

- **Step-4:** Under 99.7% confidence interval, compute the aggregate value **nPLOF** by using Eq. (7) as a kind of standard deviation of PLOF values to prepare for the normalization of PLOF

$$nPLOF \coloneqq 3 \cdot \sqrt{E[(PLOF)^2]} \tag{7}$$

- **Step-5:** Under 99.7% confidence interval, apply the Gaussian Error Function, defined as ***erf()*** in Eq. (8) to obtain the **Local Outlier Probability (LoOP)**

$$LoOP_{S(o)} \coloneqq \max\left\{0, \operatorname{erf}\left(\frac{PLOF_{S(o)}}{nPLOF \cdot \sqrt{2}}\right)\right\} \tag{8}$$

Note that the core of the LoOP method is to computes the outlier factor with respect to its surrounding neighborhood (i.e., k-nearest neighbors) based on the local reachability of each data point and the local reachability of its k-nearest neighbor points, and then normalizes the outlier score into the range of [0, 1], known as Local Outlier Probability (LoOP).

Although collaborating the LDOF method with the LoOP method can cope with the main issues and perform the training of the proposed detection algorithms with the training dataset, it remains necessary to tackle the yearly data variation embedded in the incident dataset when applying the detection algorithm with the well-trained parameters to the new dataset for model evaluation. Hence, an additional detection method is employed in Phase-II of the proposed detection process.

4.3.2 Outlier Detection Methods in Phase II

Note that the outlier detection used in Phase II for the target or test dataset should be effective in not only detecting general outliers, which have similar characteristics as those in the training dataset, but also identifying the unique local outliers existing in the test dataset. To achieve such an objective, Phase II will adopt the Self-Organizing Map (SOM) along with the LDOF method to detect local outliers uniquely embedded in the test dataset.

The Self-Organizing Map algorithm developed by Kohonen (1990) is an unsupervised artificial neural network that is able to perform data clustering (Tan and George, 2004) and provide a topological ordering where the relationships between data points are made apparent. In addition, it is also one of the dimension reduction approaches that map the data of high dimension (i.e., incident datasets) onto lower-dimensional subspaces. In the SOM, the outlier score of a data point, showing the degree of the difference between the subject data point and others, is computed with the distance between this data point and the mapped neuron point (i.e., the represented point to map the input data pattern).

In brief, Phase I aims to identify the outliers from the dataset for better model development with concurrent use of the LDOF and the LoOP methods. Phase II focuses on detecting the outliers from the test dataset that may exhibit year-to-year variation in their key variables. To mitigate the impacts of such temporal variation on the detection of true data outliers, this study proposed the use of the SOM algorithm along with the LDOF method to identify both general outliers and local outliers that exist uniquely in the test dataset.

4.4. Implementation Procedures for the Outlier-Detection Model

Figure 4-5 illustrates the structure of the proposed outlier detection model. A brief description of each step in the development and application process is presented in the ensuing sections.

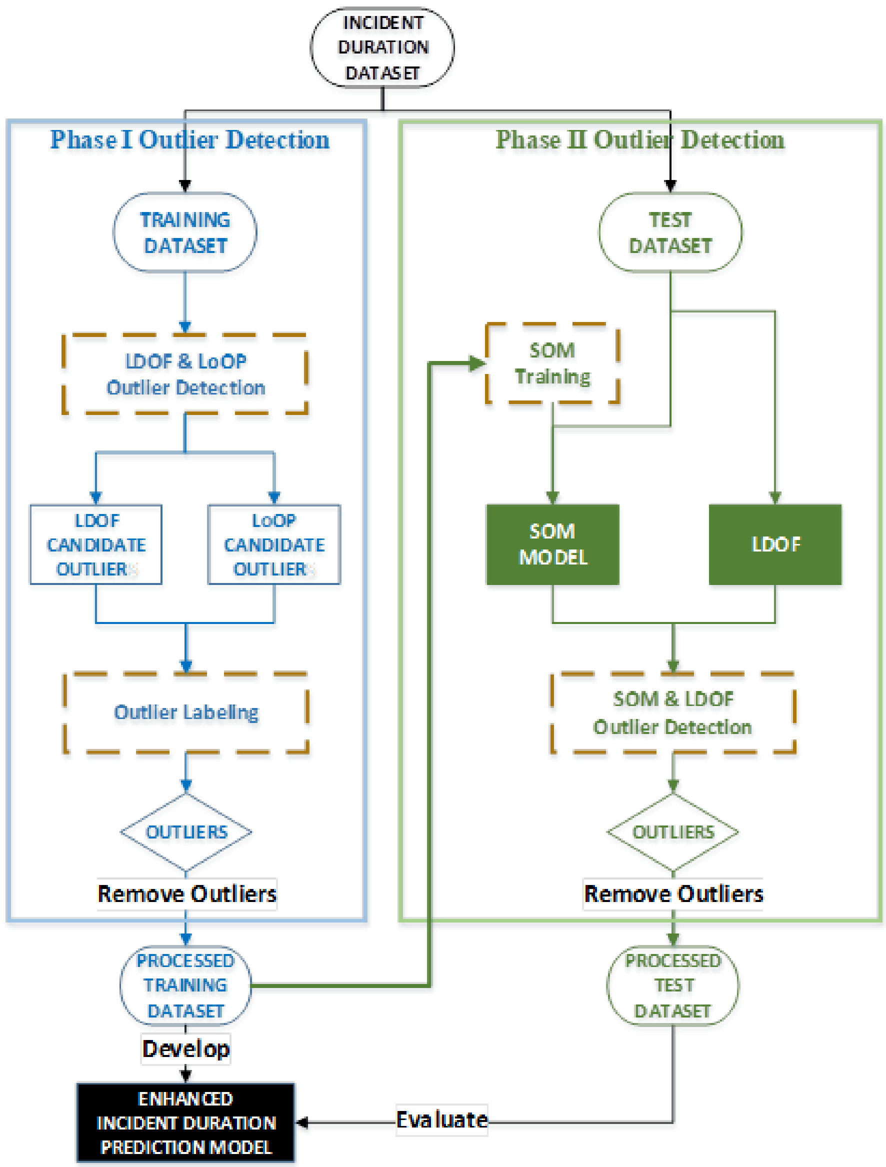


Figure 4-5. Flowchart of the Proposed Outlier Detection Model

4.4.1 Phase I: Outlier Detection for the Training Dataset

To perform the outlier detection for the training dataset, Phase I includes the execution of both **the LDOF and the LoOP methods** to find out a preliminary set of potential outliers and identify the real outliers with an **outlier labeling** procedure as illustrated in Figure 4-5. The training dataset processed with Phase I Outlier Detection algorithm would be used to develop an incident duration prediction model.

Stage-1: Outlier Detection of the LDOF Method and the LoOP Method

To compute outlier scores of each data point by using the LDOF method and the LoOP method, their parameters, k_{LDOF} and k_{LoOP}, need to be well set in advance. According to the literature, as k_{LDOF} and k_{LoOP} increase and close to the dimension of the input dataset, both methods perform progressively better and stably for a wide range (Zhang et al., 2009; Kriegel et al., 2009a). On this basis, the range of parameters can be first calibrated based on the input dataset. Then, due to no domain knowledge of the real outliers, one has to explore all possible outlier detection results by using various combinations of k_{LDOF} and k_{LoOP} within the pre-calibrated range to verify the effectiveness of their outlier detection results.

For all data points in the training dataset, data points having **top-*n* LDOF value** are taken as potential LDOF outliers. Similarly, ones **with top-*n* LoOP value** are regarded as potential LoOP outliers. To determine the proper thresholds for outlier detection, one can adopt the LoOP outlier scores, interpreted as the probability of being an outlier, to determine the value of *n*. For instance, given a LoOP outlier score of each data point, if eight data points in a dataset have LoOP values higher than a predetermined threshold (i.e., probability of being an outlier), *n* should then be set to 8.

To investigate all parameters thoroughly for identifying a proper combination of the threshold and *n* value, the threshold of LoOP value has been tested from 0.001 to 1 at an increment of 0.001 (i.e., 0.001, 0.002, 0.003, and so on). By doing so, various combinations of $\boldsymbol{k_{LDOF}}$, $\boldsymbol{k_{LoOP}}$, and $\boldsymbol{n}$ can be evaluated to find out the most effective set for use in outlier detection.

To select the best set of parameters (i.e., $\boldsymbol{k_{LDOF}}$, $\boldsymbol{k_{LoOP}}$, and $\boldsymbol{n}$) for the Phase-I application, one can apply the overlapping rate as shown in Figure 4-6 and Eq. (9) to evaluate each set, and the one producing the most overlapped outlier points (e.g., the highest overlapping rate) will be the most effective one for use in the Phase-I detection and computation. Conceivably, a higher overlapping rate represents a higher similarity between the LDOF candidate outliers and the LoOP candidate outliers, which indicates that the result under such a parameter setting is more reliable. That is to say, the most reliable outlier detection result produced by the parameter setting combination of $\boldsymbol{k_{LDOF}}$, $\boldsymbol{k_{LoOP}}$, and $\boldsymbol{n}$ should have the best overlapping rate.

$$\text{Overlapping Rate} \coloneqq \frac{\#\ of\ objects\ being\ both\ LDOF\ and\ LoOP\ outliers}{\#\ of\ objects\ being\ LDOF\ outliers} \tag{9}$$

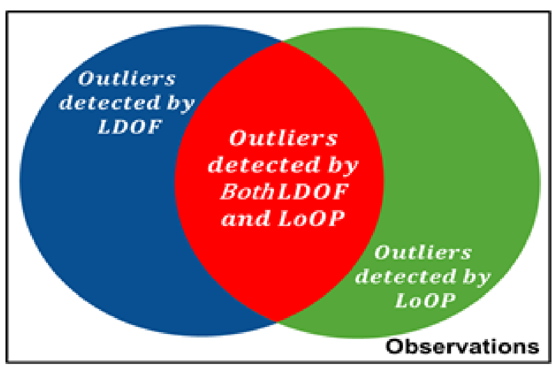


Figure 4-6. Overlapping Part of LDOF and LoOP Potential Outliers

Stage-2: Outlier Labeling

With the LDOF candidate outliers and the LoOP candidate outliers identified from Stage-1, Stage-2 will use a series of conditions to identify the real outliers from the candidate outliers in the output of Stage-1. As stated in the literature, the LDOF value of an outlier is usually much greater than one (Zhang et al., 2009). According to such information, as Figure 4-7 illustrates, outlier labeling can be conducted with the following steps:

- **Step-1**: Verify whether the LDOF of a data point is greater than one or not. If so, go to Step-2, otherwise, it is considered as a normal data point; and
- **Step-2**: Inspect whether such a data point is an outlier based on both the LDOF and the LoOP methods. (i.e., examine whether it is in the overlapping part as shown in Figure 4-6.) If so, it is labeled as an outlier; otherwise, it is considered as a normal data point.

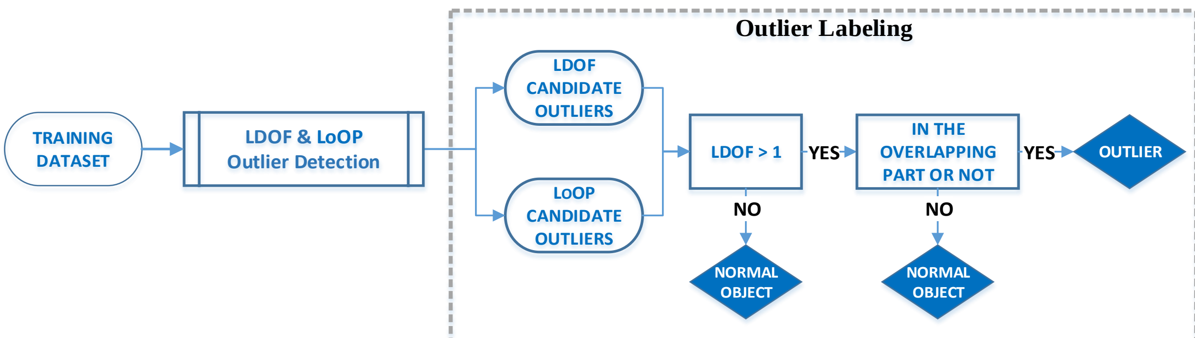


Figure 4-7. Flowchart of Outlier Labeling

4.4.2 Phase II: Detection of Outliers in the Test Dataset

As stated previously, Phase II aims to concurrently utilize the SOM algorithm and the LDOF method to detect local outliers uniquely embedded in the test dataset. As illustrated in Figure 4-5, there are two stages in phase II, including the adoption of **SOM training** to well map the patterns of the dataset from Phase I, and the concurrent use of **the SOM algorithm and the LDOF method** to identify general and local outliers.

Stage-1: SOM Training

To generate acceptable outlier scores, Stage-1 aims to train the proposed SOM to well capture the patterns of the dataset from Phase-I for detecting general outliers. According to Tan and George (2004), to well train a SOM, it is important to set appropriate learning parameters since applying different parameters to the same input will lead to different maps. To measure the reliability of different parameter settings, Tan and George (2014) provided two methods for assessing the quality of SOM maps: 1) the **quantization error** (E_q) and 2) the **topographic error** (E_t). The quantization error measures how well the inputs are represented in the output space of maps, which can be computed by using Eq. (10).

$$E_q = \frac{1}{n}\sum_{i=1}^{n}\|x_i - m_c(x_i)\| \tag{10}$$

By default, n is the number of inputs, and $m_c(x_i)$ is the output that has the closest distance from the input x_i. The topographic error, proposed by Kiviluoto (1996), represents whether the map preserves topographical order in the input space (i.e., how well the relationships among input data are represented in the output space of maps). In general, the global topographic error (E_t) can be calculated by using Eq. (11), where $u(x_i)$ is a binary parameter which equals 1 if the best and the second-best neuron points are non-adjacent, and 0 otherwise.

$$E_t = \frac{1}{n}\sum_{i=1}^{n}u(x_i) \tag{11}$$

However, there is a trade-off between the topographic and the quantization errors. In this regard, considering that the core task is to well reflect the characteristics of the input data, capturing the relationships among the input data is much more important than ensuring the accuracy. Namely, the priority target of parameter setting should be to minimize the topographic error, followed by seeking the lowest quantization error under such an identified condition.

The guidelines for training parameter setting for a SOM, as suggested by Tan and George (2014), are shown as follows:

1) The number of neurons $\approx 5\sqrt{\#\ of\ instances}$
2) Map size = $i \times i \approx$ # of neurons
3) Training cycle $\geq 500 \times (i \times i)$
4) Neighborhood radius = from $\frac{1}{2}i\ to\ 0$
5) Learning rate leading to the minimum topographic error and the low quantization error

This well-trained trained SOM based on the dataset processed from Phase-I can be applied to compute the outlier score of each data point in the test dataset.

Stage-2: Outlier Detection of the SOM Algorithm and the LDOF Method

In addition to general outliers, the LDOF method is further employed to detect local outliers in the test dataset. To verify the effectiveness of the outlier detection algorithm, all possible outlier detection results would be evaluated. Hence, the final stage intends to inspect various outlier detection results produced by different combinations of parameters, including k_{LDOF}, top-$n_{LDOF,k}$ (k indicates k_{LDOF}), and n_{SOM}.

Note that k_{LDOF} within a pre-specified range is used to compute each data point's LDOF value where data points having top-$n_{LDOF,k}$ LDOF value are taken as local outliers, whereas data points having top-n_{SOM} SOM outlier score are taken as general outliers. Figure

4-8 shows the procedure to find out the best parameter setting for the SOM algorithm and the LDOF method.

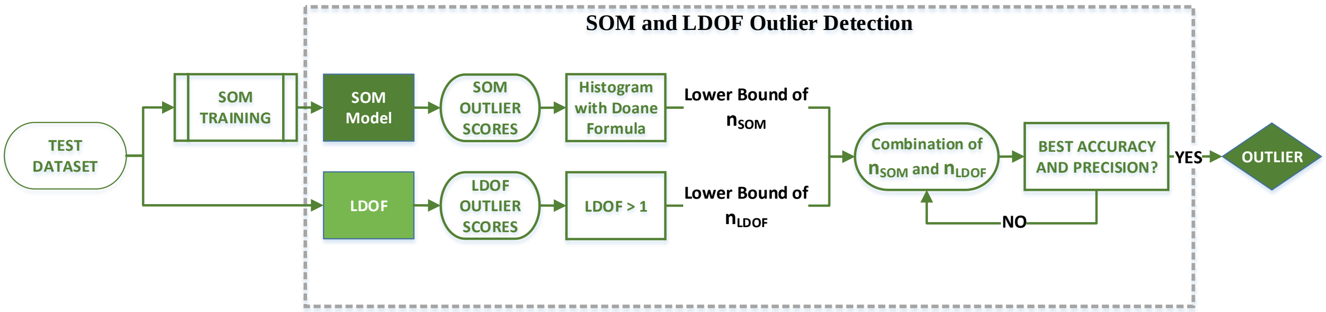
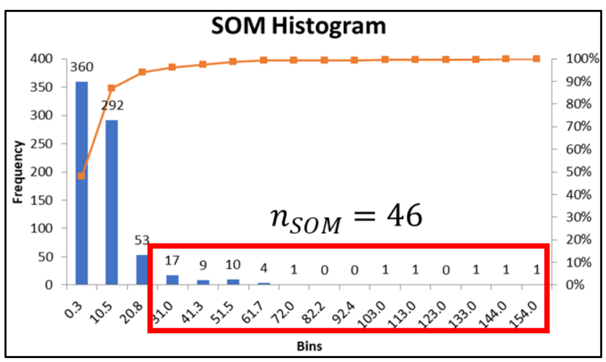


Figure 4-8. Flowchart for Outlier Detection of the SOM Algorithm and the LDOF Method

In brief, the entire procedures consist of the following steps:

- **Step-1**: Use Doane's formula (Doane, 1976) to build a histogram for outlier scores of the SOM, where the number of data points having outlier score significantly different from others is viewed as the lower bound of n_{SOM} ($\hat{n}_{SOM}$), is shown below:



- **Step-2**: Take the number of data points having LDOF value greater than 1 as the lower boundary of $n_{LDOF,k}$ ($\hat{n}_{LDOF,k}$) for a certain k_{LDOF};
- **Step-3**: For each k_{LDOF}, label objects that have the top-$n_{LDOF,k}$ LDOF values or the top-n_{SOM} SOM outlier scores as outliers; and
- **Step-4**: Explore the results generated by various combination of n_{SOM}, and $n_{LDOF,k}$ (i.e., $\hat{n}_{SOM} \times \hat{n}_{LDOF,k}$ kinds of parameter settings are explored.) and find out the

combination of n_{SOM} and $n_{LDOF,k}$ yielding the best outcome for each k_{LDOF}, which has the best **accuracy** and **precision**.

Note that the best outlier detection result with a combination of $n_{LDOF,k}$ and n_{SOM} would be found first for each k_{LDOF} by using the aforementioned procedures, and then the most reliable one would be selected from the results of various k_{LDOF}.

For the sake of efficiency of exploration, the lower bounds for $n_{LDOF,k}$ ($\hat{n}_{LDOF,k}$) and n_{SOM} ($\hat{n}_{SOM}$) are identified first, respectively, since there are numerous possible parameter settings. For the LDOF method, the indicator for the lower bound is the number of objects whose LDOF value is greater than one. On the other hand, due to the lack of criteria to identify outliers when using the SOM algorithm, $\hat{n}_{SOM}$ is determined by using a type of descriptive statistical histogram where the bin size is estimated by the Doane's formula, which features to alleviate the impact of non-normal distribution, as shown in Eq. (12) (Doane, 1976), to ensure its effectiveness since outlier scores of SOM are not normally distributed.

$$k = 1 + \log_2(n) + \log_2\left(1 + \frac{|g_1|}{\sigma_{g_1}}\right), \text{ where } \sigma_{g_1} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \qquad (12)$$

To best improve the evaluation of the incident duration model, the incident duration prediction model developed with the training dataset of no outliers from Phase I can provide useful information to assess outlier detection results in Phase II.

More specifically, if those misestimated incident cases in the test dataset with the model developed by using the Phase I dataset are detected as outliers with Phase II outlier detection, then they can be labeled as true positives, otherwise, they are false negatives. Such information makes it possible to evaluate various sets of parameters by using **accuracy** and **precision** as shown in Eq. (13) and Eq. (14) to find out the best set of parameters for outlier detection. The reason for using these two measures is that this study not only focuses on how accurate the outlier detection result is, but also intends to prevent false-positive that is, to wrongly view a normal data point as an outlier.

$$\text{Accuracy} = \frac{TP + TN}{TN + FP + FN + TP} \tag{13}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

, where $TN = True\ Negative,\ FP = False\ Positive,$
$FN = False\ Negative,\quad TP = True\ Positive$

Via Phase II Outlier Detection, both general and local outliers can be effectively detected and removed to obtain the cleansed dataset for model evaluation.

4.5. <u>Discussion</u>

This chapter has presented a two-phase model for use in detecting outliers from the incident datasets prior to its use for model development and evaluation. Phase-I aims to cleanse the training dataset by jointly using the LDOF and the LoOP methods, while Phase-II is developed to detect general and local outliers in the dataset for evaluation with the well-trained SOM and the LDOF method. In Phase I, without previous knowledge of the real outliers and considering the unique characteristics of incident datasets, the candidate outlier would be found by jointly using the LoOP and the LDOF methods, followed by using the outlier labeling procedure to identify the true outliers in the training dataset. The dataset without the outliers then can be utilized to develop an incident duration estimation model with higher accuracy. Phase II aims to identify the outliers in the test data set for a reliable model evaluation with the well-trained SOM and the LDOF method.

Chapter 5: Case Study

This chapter presents a case study to demonstrate the model transferability procedures proposed in Chapter 3 and to evaluate the effectiveness as well as benefits of the outlier detection algorithms illustrated in Chapter 4 by using the incident records of I-695 in Maryland.

This chapter is organized as follows: the first section is to demonstrate how to utilize the proposed transferring process to develop an extended model by using incident records of I-695 and show its performance. The second section is to illustrate the application of the proposed outlier detection model for improving the quality of incident datasets for model development and refinement.

Data Description

The data for the case study are from the incident records of I-695 in Maryland derived from the Coordinated Highways Action Response Team (CHART) II Database. The incident data from 2016 to 2018 are used as the training dataset to develop an extended model by using the proposed transferring procedures, and the data in 2019 are used as the test set for model performance evaluation. Figure 5-1 shows the spatial distribution of I-95, I-495, and I-695, where the last is to be developed with the transferability analysis procedures from the incident duration prediction systems developed for the first two highways. Also, Table 5-1 shows the available factors in each incident record associated with the incident clearance time in the dataset. All such factors can be categorized into seven categories: **incident type**, **time**, **location**, **environmental condition**, **operation center**, **lane blockage**, **involved vehicle**, and **response unit**.

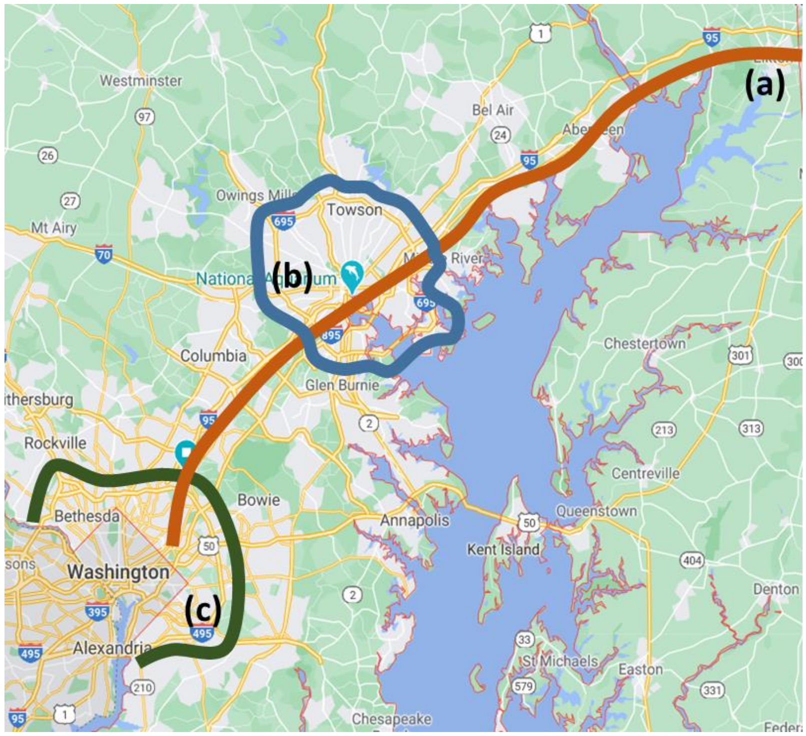


Figure 5-1. Spatial Distribution of (a) I-95 in MD, (b) I-695, (c) I-495 in MD

Table 5-1. List of Key Factors Associated with the Response/ Clearance of a Detected Incident

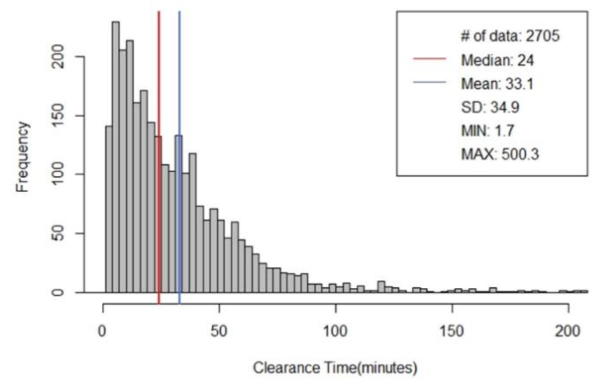
Category	Variable	Classification
Incident Type	Incident Type	Collision with Fatality (CF), Collision with Personal Injury (CPI), Collision with Property Damage (CPD)
Time	Hour Indicator	AM-Peak (7 a.m. ~10 a.m.), Day Time (10 a.m. ~4 p.m.), PM-Peak (4 p.m. ~7 p.m.), Nighttime (7 p.m. ~7 a.m.)
	Weekend Indicator	Weekend, Weekday
	Holiday Indicator	Holiday, Non-Holiday
	Season Indicator	Spring, Summer, Fall, Winter
Location	County Indicator	Baltimore City, Baltimore
	Direction Indicator	Inner Bound, Outrebound
	Exit Number Indicator	Exit 1, …, Exit 44
Environmental Condition	Pavement Condition Indicator	Dry, Wet, Snow/Ice, Chemical Wet, Unspecified
	Hazard Material Related	Yes, No
Operation Center	Center Indicator	AOC, SOC, TOC4, TOC5, TOC7, Others
Lane Blockage	# of Blocked Lanes[1]	1, 2, 3, 4, …
	# of Blocked Shoulders	0, 1, 2, 3, …
	# of Blocked Travel Lanes[2]	0, 1, 2, 3, …
	# of Blocked Traffic Lanes	0, 1, 2, 3, …
	# of Blocked Auxiliary Lanes[3]	0, 1, 2, 3, …
	Travel Lane Blocked in Tunnel	Yes, No
	Travel Lane Blocked in Toll	Yes, No
Involved Vehicle	Vehicle States	Jack-knife, Over-turned, Lost load
	# of Total Involved Vehicles[4]	1, 2, 3, 4, …
	# of Involved Passenger Cars	0, 1, 2, 3, …
	# of Involved Trucks	0, 1, 2, 3, …
	# of Involved Motorcycles	0, 1, 2, 3, …
Response Unit	# of Total Response Units	0, 1, 2, 3, …
	# of Arrived CHART	0, 1, 2, 3, …
	# of Arrived Police	0, 1, 2, 3, …
	# of Arrived Medical Service	0, 1, 2, 3, …
	# of Arrived Tow Service	0, 1, 2, 3, …
	First Response Unit	CHART, Police, Fireboard, Medical, Tow

1. Lanes = Shoulders + Travel Lanes
2. Travel Lanes = Traffic Lanes + Auxiliary Lanes
3. Auxiliary lane includes on-ramp, off-ramp, acceleration lane, deceleration lane, and collector/ distributor lane.
4. Vehicle includes passenger car, truck, bus, cyclist, pedestrian, and motorcycle.

5.1. *Model Development and Evaluation with the Transferring Methodology*

5.1.1 Data Preprocessing and Incident Categorization

Figure 5-2 shows that the distribution of the pre-processed incident clearance times is right skewed with a long tail, where the maximum clearance time is over 500 minutes, but approximately 90% of incidents' durations are less than 66 minutes. To remove the obvious data errors in the incident dataset of I-695, this study uses the procedures proposed by Won et al. (2018).



Quantile	10%	30%	50%	70%	80%	90%
Time (mins)	5.7	13.6	24	38.3	48.9	66.1

- Incident Type: Collision
- Data Period: January of Year 2016 ~ December of Year 2019

Figure 5-2. Distribution of Preprocessed Incident Duration Dataset of I-695

The pre-processed data of I-695 are then assigned into several categories according to the incident type and the status of travel lanes. Figure 5-3 shows the average incident clearance time and its estimated ranges with 60%, 70%, and 80% of probability for each pre-classified category.

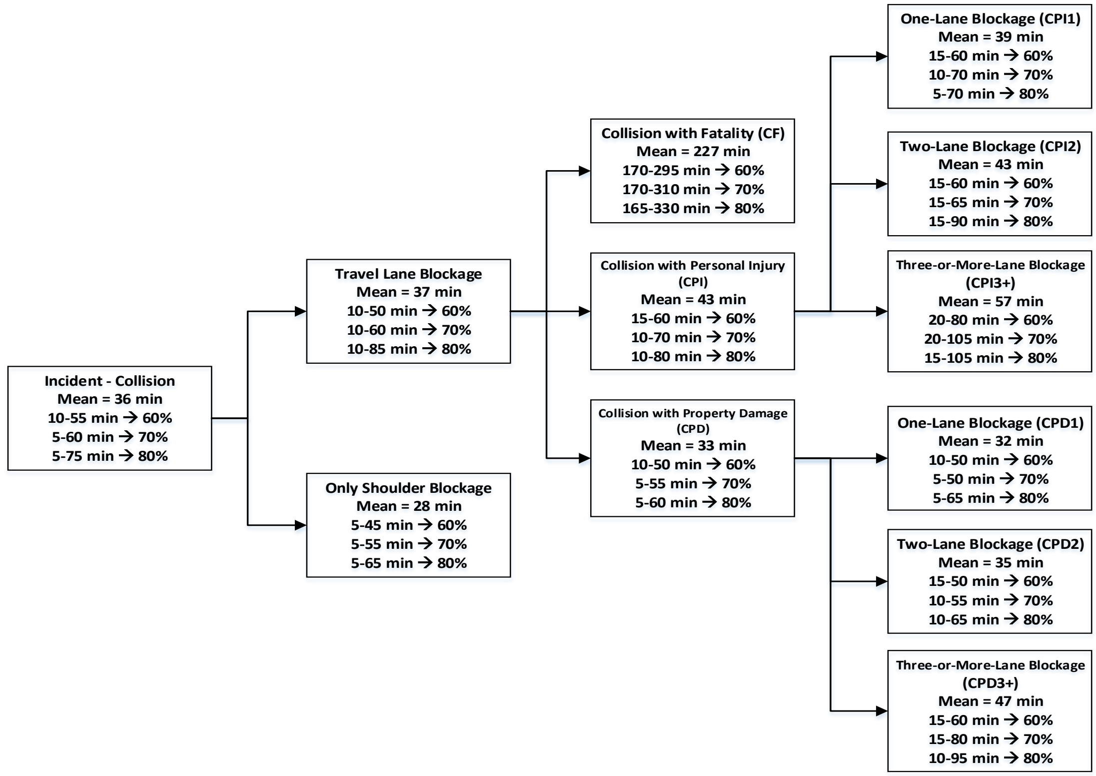


Figure 5-3. Initial Incident Categorization and Estimated Clearance Duration

5.1.2 Classification Rule Partitioning and Ordering

In addition to the incident records of I-695, the existing classification rules from the models of I-95 and I-495 are partitioned into several groups based on their estimated clearance time, as shown in Table 5-3. According to Table 5-3, there are only a few existing rules in each group, indicating that the transferring order would not be a critical issue in this case study.

Table 5-2. The Partitioning of Classification Rules from the Models of I-95 and I-495

Estimated CT	**"≥30 minutes" or "<30 minutes"**		**"≥60 minutes" or "<60 minutes"**		**"≥120 minutes" or "<120 minutes"**	
Source	**I-95**	**I-495**	**I-95**	**I-495**	**I-95**	**I-495**
CPI 1	9	7	5	4	3	1
CPI 2	5	4	5	3	1	1
CPI 3+	5	4	6	3	4	2
CPD 1	8	6	3	4	1	1
CPD 2	5	3	3	2	2	1
CPD 3+	3	2	2	2	3	2
Total	35	26	24	18	14	8

5.1.3 Automatic Rule Transferring and New Classification Rule Searching

To present the process of the automatic rule transferring and new classification rule searching, the incident records of collisions with personal injury and three blocked travel lanes (CPI3) are taken as an example for illustration and shown in Figure 5-4. Based on those incident records, the existing predicting rules, which are categorized into CPI3 and conform to the pre-determined criteria as shown in Table 5-3, would be transferred from the systems for I-95 and I-495 into the new extended model for I-695, as marked in green in Figure 5-4.

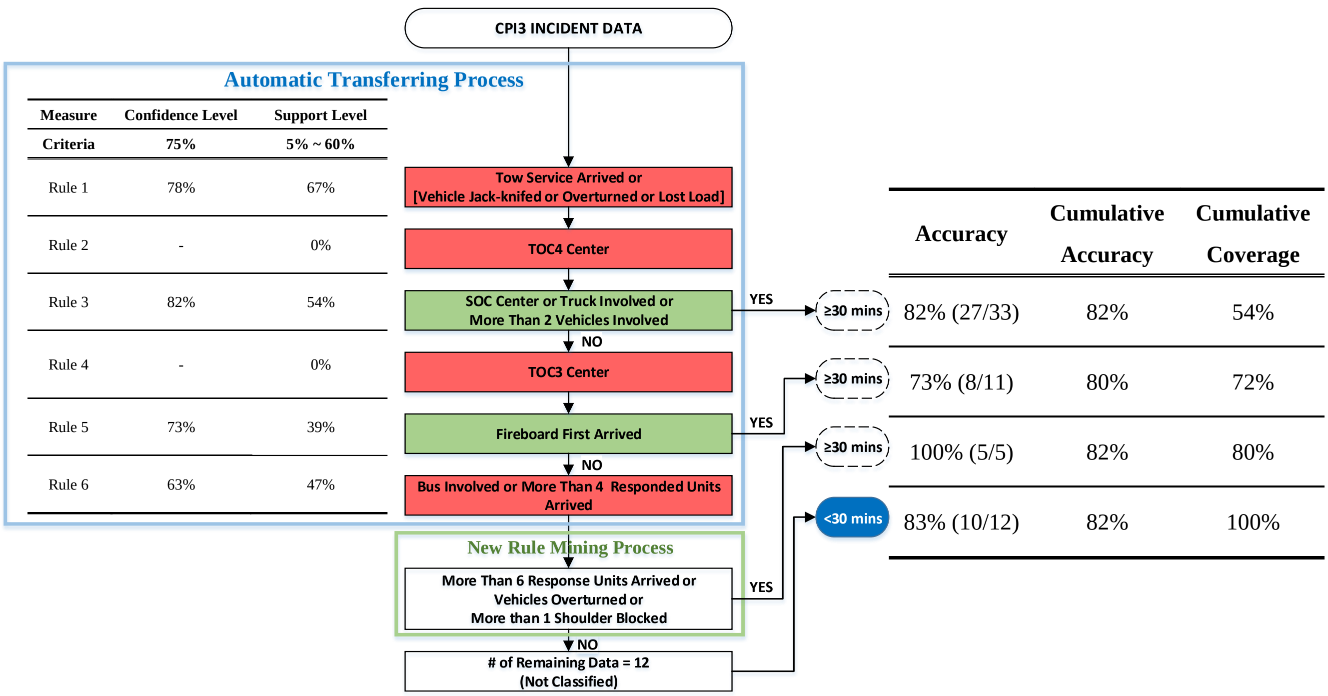


Figure 5-4. An Example of the Classification Rules for “≥30 Minutes” and “<30 Minutes” in CPI 3

Table 5-3. Standard Confidence Level and Boundaries of Support Level for Each Group of Classification Rules

Estimated CT	Confidence Level	Support Level Upper Boundary	Support Level Lower Boundary
"≥30 minutes" or "<30 minutes"	75%	60%	5%
"≥60 minutes" or "<60 minutes"	80%	80%	0%
"≥120 minutes" or "<120 minutes"	80%	–	–

After completion of the transferring process with the criteria shown in Table 5-3, for the remaining I-695 incident records that cannot be classified by the transferred rules, one can then adopt the Association Rule Mining method to identify new prediction rules to account for the local-specific factors. Those newly developed rules with the confidence level greater than 75% and having the highest support level are taken as the new set of classification rules in the extended model for I-695.

Via collaborating automatic transferring with the new rule mining process, the incident records in CPI3 can be classified into the groups of "<30 minutes", "30-60 minutes", "60-120 minutes", and "≥120 minutes"; and the clearance time estimation with 60%, 70%, and 80% probability would also be computed as shown in Figure 5-5.

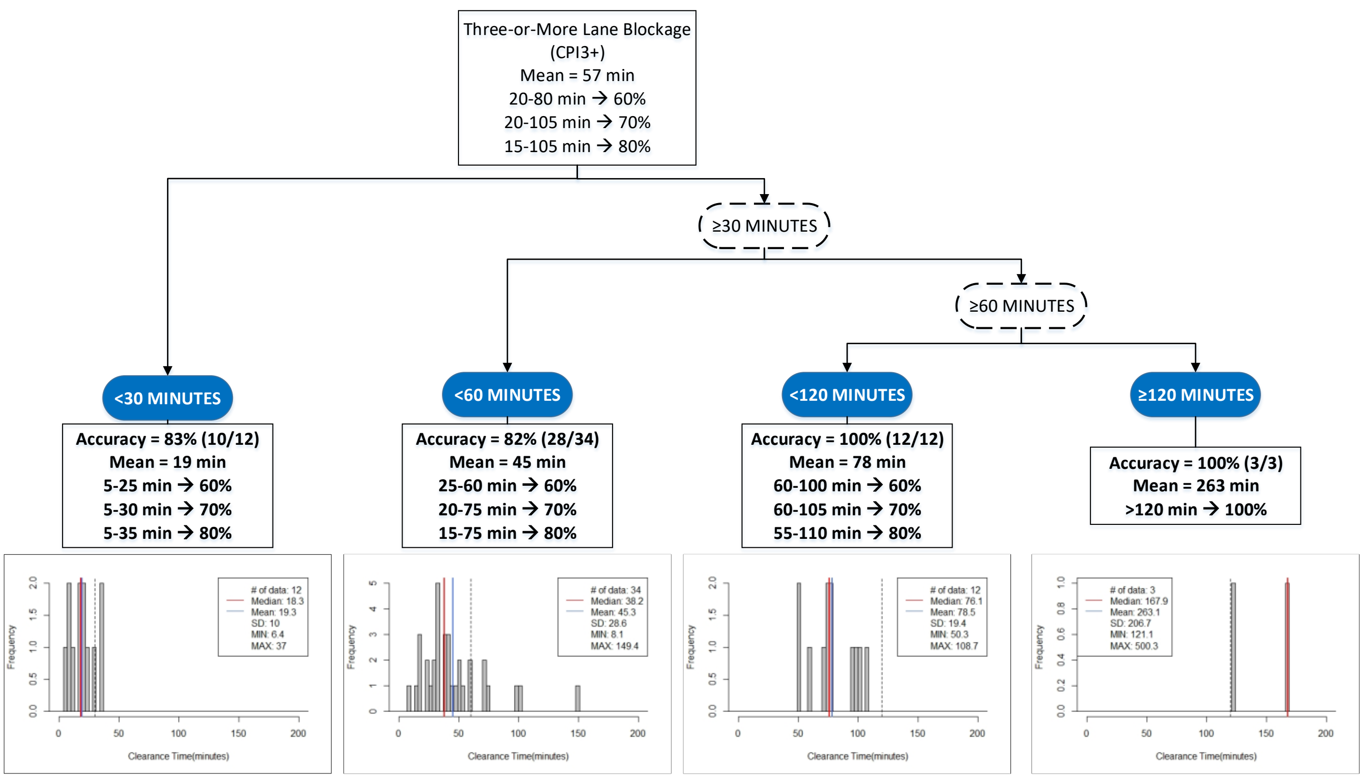


Figure 5-5. Distributions and Ranges of Incident Clearance Time with Different Confidence Levels

With the same procedures for all pre-classified categories of incident data, it has been concluded that 16 rules are eligible for transfer from the previously developed models for I-95 and I-495, and 43 rules are identified as new rules to the extended model for I-695. The entire set of rules, including both the transferred and local rules for the I-695 is shown in the appendix.

5.1.4 Model Evaluation and Discussion

This case study has first utilized the incident data of I-695 from 2016-2018 to train the extended model and the incident records in 2019 for model performance evaluation. Table 5-4 shows the evaluation results of the extended model by each pre-classified incident category, compared to the previously developed models for I-95 and I-495. In addition, the estimated results of the extended model based on different clearance time intervals, and their comparisons with other models are shown in Table 5-5. More specifically, Table 5-6 summarizes the distribution of estimation errors with the extended model.

Table 5-4. Accuracy Comparison by Incident Types and the Number of Blocked Travel Lanes

Highway	Collision with Travel Lane Blockage						
	CPI 1	**CPI 2**	**CPI 3+**	**CPD 1**	**CPD 2**	**CPD 3+**	**Total**
I-95 (2012-2017)	77.2% (446/578)	84.6% (203/240)	78.8% (82/104)	74.3% (795/1070)	80.5% (177/220)	83.7% (41/49)	**77.1% (1744/2261)**
I-495 (2015-2018)	78.7% (392/498)	78.7% (295/375)	61.7% (113/183)	79.8% (631/791)	81.6% (301/369)	79.2% (95/120)	**80.0% (2018/2523)**
I-695 (2016-2019)	85.6% (297/347)	82.4% (150/182)	78.7% (59/75)	87.0% (842/968)	87.6% (219/250)	82.7% (43/52)	**85.9% (1610/1874)**

Table 5-5. Accuracy Comparison by Incident Clearance Time Categories

Highway	Actual CT (Minutes)				
	<30	**30~60**	**60~120**	**≥120**	**Overall**
I-95 (2012-2017)	96.3% (880/914)	81.9% (678/828)	35.3% (149/422)	50.0% (61/122)	**77.3% (1768/2286)**
I-495 (2015-2018)	88.2% (1560/1769)	72.9% (380/521)	26.8% (48/179)	50.7% (35/69)	**79.7% (2023/2538)**
I-695 (2016-2019)	90.8% (896/987)	89.0% (561/630)	60.3% (129/214)	55.8% (24/43)	**85.9% (1610/1874)**

Table 5-6. Distribution of Estimation Errors by Incident Clearance Time Categories of I-695

Errors		Actual CT (minutes)				Overall
		< 30	30 ~ 60	60 ~ 120	≥ 120	
Over-estimated	> 120	0	0	0	0	0
	60~120	2	0	0	0	4
	30~60	2	0	2	0	4
	10~30	6	4	1	0	11
	0~10	81	9	1	0	91
Within boundaries		**896**	**561**	**129**	**24**	**1610**
Under-estimated	-10~0	0	41	**23**	**2**	66
	-30~-10	0	13	**43**	**0**	56
	-60~-30	0	0	**13**	**3**	16
	-120~-60	0	0	**2**	**9**	11
	< -120	0	0	**0**	**5**	5
Total # of cases		**987**	**630**	**214**	**43**	**1874**
TP rate and Accuracy		**90.8%**	**89.0%**	**60.3%**	**55.8%**	**85.9%**

As shown in Table 5-4, the I-695 system with the transferred and local rules performs as well as the previously developed models for I-95 and I-495 with an overall accuracy of over 80%. Such performance shows that the extended model developed by the proposed transferring procedures can effectively and efficiently capture both the individual and collective impacts of all variables associated with the incident clearance time with fewer development efforts and data needs.

However, it is observable from the results in Table 5-5 that the model's prediction accuracy for incidents in the categories of "60~120 minutes" and "≥120 minutes" are approximately 55~60%, not achieving the desired level of accuracy. Additionally, Table 5-6 shows that 104 incidents with clearance durations exceeding 60 minutes are underestimated. One contributing reason for such results is likely to be the right-skewed distribution of the incident data with a long tail. The existence of data outliers due to either recording errors or special incident response scenarios may also cause the model's undesirable level of performance with respect to incidents of long duration. Hence, the outlier detection model presented in Chapter 4 is applied hereafter to the I-695 dataset for model refinement.

5.2. Outlier Detection Module Development and Evaluation Results

5.2.1 Phase I Outlier Detection

To ensure the stable performance of the LDOF and the LoOP methods, it is well recognized that appropriate parameters need to be well set first. To do so, the ranges of parameters for the LDOF (k_{LDOF}) and the LoOP (k_{LoOP}) methods should be calibrated based on the incident records of I-695. With the I-695 incident dataset (data from 2016 to 2018) as the input, the study has conducted extensive experimental analysis and concluded that the performance of the LDOF method is relatively stable when k_{LDOF} is from 31 to 40, and the performance of the LoOP method is relatively stable when k_{LoOP} is from 24 to 55. The calibration procedures for k_{LDOF} and k_{LoOP} are available in the appendix.

After assessing all possible outlier detection results generated by various combinations of k_{LDOF} and k_{LoOP} within the pre-calibrated ranges, the most reliable outlier detection performance is found by using the defined overlapping rate. The results show that the best parameter for $\boldsymbol{k_{LDOF}}$ is **34**, and that for $\boldsymbol{k_{LoOP}}$ is **30**. Also, a data point with either **top-21 LDOF outlier scores** or **top-21 LoOP outlier scores** is taken as a potential outlier, as shown in Figure 5-6.

Figure 5-7 shows that there are **20 outliers** in the training dataset of 2523 incident records and the detected outliers are marked as green points in Figure 5-7 (a) and Figure 5-7 (b). With such information, one can then remove those outliers and develop another I-695 model with better quality data.

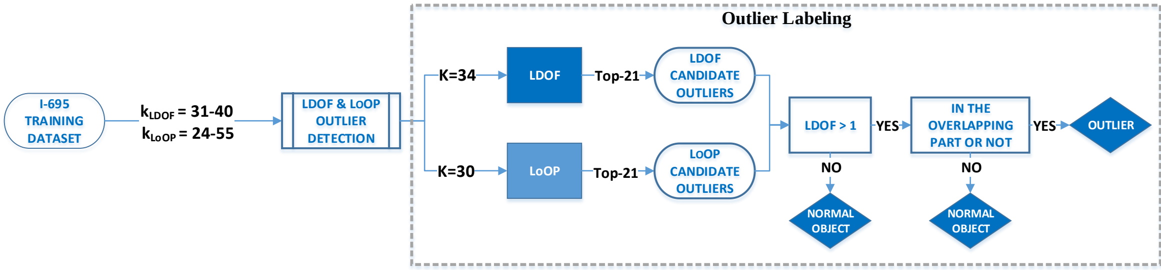
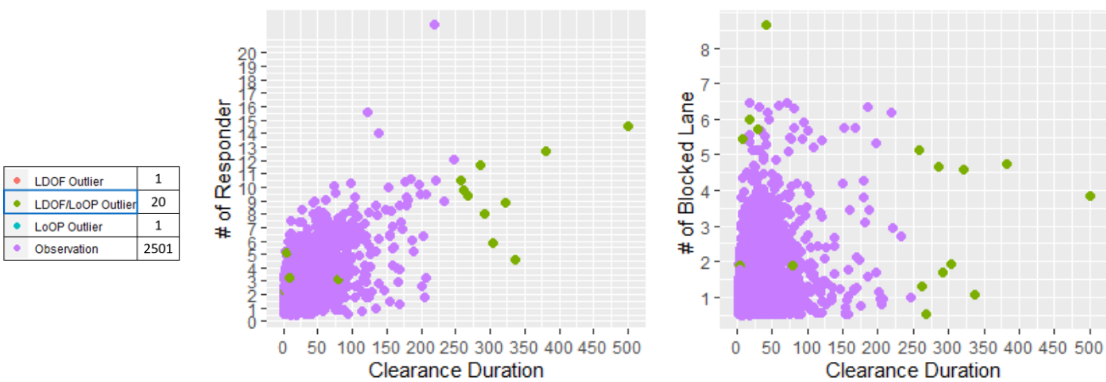


Figure 5-6. Flowchart of Phase I Outlier Detection



(a) (b)

Figure 5-7. Scatter Plots of (a) the Number of Responded Units Arrived and Incident Clearance Duration (minutes), (b) the Number of Blocked Lanes and Incident Clearance Duration (minutes)

5.2.2 Phase II Outlier Detection

Since the likely presence of outliers in the test dataset may also bias the resulting performance of the candidate model for evaluation, the detection procedures in Phase II of the outlier detection model are applied in the test dataset of I-695 incident records in 2019.

To detect the general outliers in the test dataset, the first step is to well train SOM to capture the data characteristics embedded in the training dataset. Note that even though effective guidelines for parameter settings are available in the literature, its respective learning rate needs to be calibrated from the primary incident data for model development by using the following two measures: **topographic error** and **quantization error**. To well preserve the relationships between key incident-duration related variables, the acceptable SOM should have the minimum topographic error but relatively lower quantization error. With such a principle, the learning rate has been calibrated to be 0.2. The whole parameters used to train the SOM to map the training dataset are shown in Table 5-7. With this well-trained SOM, the outlier score of each incident event in the test dataset can be computed based on the primary dataset and the criteria consistent with Phase I.

Table 5-7. SOM Training Parameter Setting

Parameter	**Value**
Map Size	16×16 Rectangular Map (256 Neurons)
Training Cycle	128000
Neighborhood Radius	From 8 to 0
Learning Rate	Start From 0.2 and Decay to 0

On the other hand, the LDOF method is used to identify the local outliers in the test dataset. Using the test dataset of I-695 as the input, the LDOF method performs stably when its key parameter, $\boldsymbol{k_{LDOF}}$, lies in the range of **31 to 38**.

For the sake of efficiency for exploration, as discussed in section 4.4, the lower bounds of $\boldsymbol{n_{SOM}}$ ($\boldsymbol{\hat{n}_{SOM}}$) and $\boldsymbol{n_{LDOF,k}}$ ($\boldsymbol{\hat{n}_{LDOF,k}}$) for each k need to be identified first. Figure 5-8 shows that $\boldsymbol{\hat{n}_{SOM}}$ is **46**, and Table 5-8 reports the lower bound of $\boldsymbol{n_{LDOF,k}}$ ($\boldsymbol{\hat{n}_{LDOF,k}}$) for each $\boldsymbol{k}$.

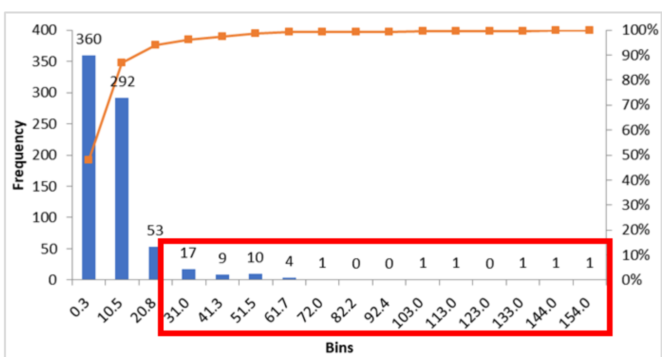


Figure 5-8. Distribution of Outlier Scores Computed by SOM

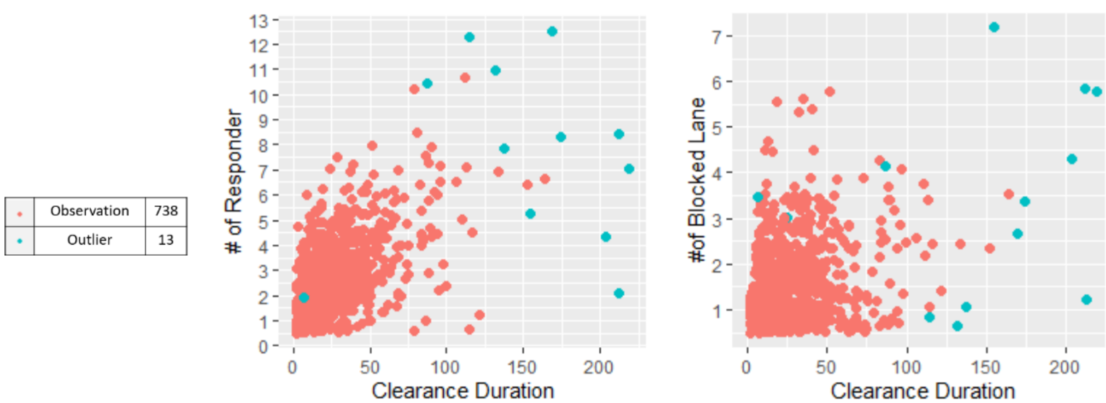
Table 5-8. The Lower Bound for Each *k*

$\boldsymbol{k_{LDOF}}$	31	32	33	34	35	36	37	38
Lower Bound	74	75	76	73	73	72	73	73

The next step is to investigate all possible outlier detection results under various combinations of $n_{LDOF,k}$ and n_{SOM} for each k. Among these, the outlier detection results with the best accuracy and precision of each k are listed in Table 5-9, where $\boldsymbol{k}$ either **37** or **38** can yield the **best accuracy and precision** as well as the **smallest number of outliers**. Also, the incident events having **top-6 LDOF values** or **top-8 SOM outlier scores** are taken as outliers. As shown in Figure 5-9, there are **13 outliers** among 751 incident records in the test dataset, and the outliers are marked as green-blue points in Figure 5-9 (a) and (b).

Table 5-9. The Most Reliable Outlier Detection Result of Each k

k	n_{LDOF}	n_{SOM}	Accuracy	Precision
31	12	8	82.13%	66.67%
32	11	8	82.13%	66.67%
33	10	8	82.13%	66.67%
34	9	8	82.13%	66.67%
35	8	8	81.89%	63.64%
36	7	8	82.13%	70.00%
37	6	8	82.13%	70.00%
38	6	8	82.13%	70.00%



(a) (b)

Figure 5-9. Scatter Plots of (a) the Number of Responded Units Arrived and Incident Clearance Duration (minutes), (b) the Number of Blocked Travel Lanes and Incident Clearance Duration (minutes)

Finally, the test dataset without those outliers can be used to reliably evaluate the performance of the I-695 model developed with the training dataset without outliers.

5.2.3 Model Evaluation and Discussion

To examine the effectiveness of the proposed outlier detection module, this study has compared the performance of the I-695 models developed with the dataset with and without the outliers. Table 5-10 shows the comparison results by the incident type and the number of blocked travel lanes. Additionally, Table 5-11 shows the same performance comparison by the categorized incident clearance time, and Table 5-12 summarizes the distribution of the estimation errors.

Table 5-10. Accuracy Comparison by Incident Type and the Number of Blocked Lanes for the I-695 Model Using the Dataset with and without Outliers

Model	**Collision with Travel Lane Blockage**						
	CPI 1	**CPI 2**	**CPI 3+**	**CPD 1**	**CPD 2**	**CPD 3+**	**Total**
Original (2016-2019)	85.6% (297/347)	82.4% (150/182)	78.7% (59/75)	87.0% (842/968)	87.6% (219/250)	82.7% (43/52)	**85.9% (1610/1874)**
Enhanced (2016-2019)	91.3% (314/344)	88.6% (156/176)	85.1% (57/67)	87.0% (841/967)	90.7% (224/247)	94.1% (48/51)	**88.6% (1640/1852)**
Improvement	**6.7%**	**7.5%**	**8.1%**	**0.0%**	**3.5%**	**13.8%**	**3.1%**

Table 5-11. Accuracy Comparison by Incident Clearance Time for I-695 Using the Dataset with and without Outliers

Model	**Actual CT (Minutes)**				
	<30	**30~60**	**60~120**	**≥120**	**Overall**
Original (2016-2019)	90.8% (896/987)	89.0% (561/630)	60.3% (129/214)	55.8% (24/43)	**85.9% (1610/1874)**
Enhanced (2016-2019)	91.7% (899/980)	91.4% (575/629)	67.5% (143/212)	74.2% (23/31)	**88.6% (1640/1852)**
Improvement	**1.0%**	**2.7%**	**12.0%**	**32.9%**	**3.1%**

Table 5-12. Distribution of Estimation Errors by Incident Clearance Time for I-695 Using the Dataset with and without Outliers

Errors		**Actual CT (Minutes) and Modeling Stages**								**Overall**	
		<30		**30~60**		**60~120**		**≥120**			
		Original	**Enhanced**	**Original**	**Enhanced**	**Original**	**Enhanced**	**Original**	**Enhanced**	**Original**	**Enhanced**
Over-Estimated	>120	0	0	0	0	0	0	0	0	0	0
	60~120	2	0	2	0	0	0	0	0	4	0
	30~60	2	3	0	0	2	0	0	0	4	3
	10~30	6	13	4	7	1	1	0	0	11	21
	0~10	81	65	9	10	1	0	0	0	91	75
Within Boundaries		**896**	**899**	**561**	**575**	**129**	**143**	**24**	**23**	**1610**	**1640**
Under-Estimated	-10~0	0	0	41	27	23	16	2	0	66	43
	-30~-10	0	0	13	10	43	36	0	0	56	46
	-60~-30	0	0	0	0	13	15	3	2	16	17
	-120~-60	0	0	0	0	2	1	9	4	11	5
	<-120	0	0	0	0	0	0	5	2	5	2
Total # of cases		**987**	**980**	**630**	**629**	**214**	**212**	**43**	**31**	**1874**	**1852**
TP rate and Accuracy		**90.8%**	**91.7%**	**89.0%**	**91.4%**	**60.3%**	**67.5%**	**55.8%**	**74.2%**	**85.9%**	**88.6%**

Intuitively, an effective outlier detection module should be able to properly detect outliers in the dataset to improve the incident duration estimation. Table 5-10 demonstrates that the improvement rates are over 5% across the different lane blockages of CPI incidents, and the maximum improvement of more than 13% lies in the category of CPD3+. Similarly, Table 5-11 shows that the improvement rate for the category of "≥120 minutes" is most pronounced and exceeds 30%. As for the category of "60~120 minutes", the performance is also improved by 12%, despite that the improvement to the categories of "<30 minutes" and "30~60 minutes" is limited because the model developed with the original dataset has achieved high accuracy for incidents in those two categories. Moreover, as shown in Table 5-12, the number of under-estimated cases in the categories of "60~120 minutes" and "≥120 minutes" decrease by 27.

Note that Table 5-12 shows that the category of "60~120 minutes" has the lowest accuracy but approximately 92% (195/212) of them have less than 30 minutes time difference between the actual clearance time and the estimated clearance time. Thus, the overall accuracies are still robust enough to provide a reliable estimated clearance time.

In brief, these evaluation results indicate that the proposed outlier detection module can properly identify the outliers in the incident dataset and provide a better-quality dataset for the best model development. As expected, the procedures in Phase I of Outlier Detection can effectively detect outliers in the training dataset to improve data quality for the model development for clearance time estimation. The algorithm adopted in Phase II of Outlier Detection also offers an effective tool for producing a quality dataset without outliers for a reliable model performance evaluation.

Chapter 6: Conclusions

6.1. Summary of Research Works

Traffic incidents often not only result in significant delays to roadway users, but also decrease the reliability of a highway network. To alleviate the adverse influences caused by those incidents, it is essential for the responsible highway agencies to perform efficient responses and clearance operations, including providing proper detour instructions, implementing integrated controls, and displaying incident-related information (e.g., queue length, travel time). The effectiveness and reliability of such operations inevitably rely on the accurate estimation of the incident clearance duration. Hence, over the past decades, a large body of studies has proposed various methods for addressing this imperative issue, but the development of a cost-effective model for reliable use in practice remains a challenging task for the traffic community. Some of the critical issues contributing to difficulties in developing a reliable model include (1) insufficient incident records, especially for those major incidents with long duration; (2) inadequate recording of all factors affecting the resulting incident duration; and (3) unreliable data quality due to either recording errors or missing of some key factors associated with the incident duration.

To cope with such issues, this study has first proposed a rule-transferring process to efficiently develop a traffic incident duration estimation model for a new highway from the well-established system, i.e., the work by Won et al. (2018). Such a proposed process for model development takes advantage of the knowledge from previously developed models to effectively capture the individual and collective impacts of key contributing factors on the incident clearance duration. It can maintain the consistency among the rules adopted for new incident duration estimation models for different highways, and is especially cost-effective for those highways with limited incident records for model development. The proposed transferring procedures have been applied to develop the incident duration prediction model with the I-695 incident records, yielding the overall accuracy of 85.9% that is as good as the performance of other previously developed models.

To further increase the accuracy of the incident duration estimation models, this study has developed a two-phase outlier detection model by considering the distinct features of incident data and their unique local-specific characteristics. Within the outlier detection model, Phase I jointly uses the LDOF and the LoOP methods to first find out potential outliers, and then further identify the true outliers with outlier labeling to the training dataset for model development. Phase II is to use the well-trained SOM to detect general outliers and the LDOF method to identify the local outliers in the test dataset for model performance evaluation. Such an outlier detection model can effectively detect outliers to enhance the incident data quality and further improve the accuracy of the estimated incident clearance time. The effectiveness of the proposed outlier detection model has also been evaluated by using incident records of I-695. It has been found that the model developed with the dataset without those identified outliers can significantly improve its prediction accuracy, especially with respect to those incidents difficult to predict at the desirable level of reliability.

In brief, this research has made the following contributions toward the development of a reliable incident duration estimation model:

- Developed an efficient knowledge transferring process to best utilize the information from the previously developed model to a new model for a target highway, offering a new avenue for cost-effectively developing a generalized model for all highways within the same region;
- Designed a two-phase outlier detection model for the development of the incident duration model to free from the contamination of outliers or data errors; and

- Allow the responsible highway agencies with the proposed outlier detection model to evaluate the quality of their available incident records and to review as well as to perform potential improvement on the incident reporting and recording process.

6.2. Future Research

Although this study has proposed an effective outlier detection model to cleanse the incident dataset, and an efficient process for developing an incident duration prediction model with the cleansed incident data, some vital issues remain to be studied to ensure the reliability of such a system for use in practice under various traffic conditions:

- **Development of a Generalized Model**: The extended incident duration prediction model in this study was developed by using the data of I-695 in Maryland, based on the previously developed models of I-95 and I-495 in Maryland. Such a development process can be further improved when more models for various highways become available and a larger number of effective rules have been established so that a generalized model can be constructed to estimate the incident clearance time for different highways within the same region; and

- **Utilization of the Outlier Information**: With the outlier detection model, one can identify the outliers in the incident dataset to not only improve the quality of the model for incident duration prediction, but also apply the outlier detection results to assess the entire incident reporting and recording process which often suffers from either poor interface design for data recording or the lack of data filtering process to ensure the quality of either on-line or off-line recorded data.

Appendices

The Set of Rules for I-695

(a) Rules for collisions with personal injury and one-travel-lane blockage (CPI1)

	Description		CT
If	Tow service arrived	Then	>30
	Fireboard first arrived		<30
	TOC4 center & No truck involved		<30
	More than 1 CHARTs arrived & Police first arrived		>30
	More than 1 trucks involved & More than 3 response units arrived		>30
	Pickup involved & [More than 2 response units arrived OR Police arrived]		>30
	Else then		<30
If	[Night & More than 6 response units] OR More than 4 vehicles	Then	>60
	Snow-iced pavement condition OR More than 1 trucks involved OR More than 7 response units arrived OR AOC center		>60
	Pickup involved & [Auxiliary lane blocked OR Winter]		>60
	Weekend & Vehicle overturned		>60
	More than 1 polices arrived & More than 1 fireboards arrived		>60
	Else then		<60
If	Truck involved & More than 5 response units arrived	Then	>120
	Lane blocked in toll line OR More than 1 shoulders blocked		>120
	Else then		<120

(b) Rules for collisions with personal injury and two-travel-lane blockage (CPI2)

	Description		CT
If	Tow service arrived	Then	>30
	Peak hour & More than 4 response units arrived		>30
	Weekend & SOC center		>30
	More than 5 response units arrived		>30
	Else then		<30
If	Night & [More than 1 polices arrived OR More than 5 response units arrived]	Then	>60
	Snow-iced pavement condition OR [More than 7 response units arrived & Auxiliary lane blocked]		>60
	Else then		<60
If	Wet pavement condition OR More than 4 vehicles involved	Then	>120
	Else then		<120

(c) Rules for collisions with personal injury and three-and-more-travel-lane blockage (CPI3+)

	Description		CT
If	SOC center OR Truck involved OR More than 2 Vehicles involved	Then	>30
	Fireboard first arrived		>30
	More than 5 response units arrived OR Vehicle overturned OR More than 1 shoulders blocked		>30
	Else then		<30
If	Medical service arrived	Then	>60
	More than 8 response units arrived		>60
	No tow service		<60
	More than 1 tow service arrived & [More than 5 response units arrived OR More than 1 CHARTs arrived]		>60
	Else then		<60
If	More than 4 lanes blocked & More than 2 CHARTs arrived	Then	>120
	Wet pavement condition & More than 2 tow service arrived		>120
	Else then		<120

(d) Rules for collisions with property damage and one-travel-lane blockage (CPD1)

	Description		CT
If	Tow service arrived OR Hazard materials related	Then	>30
	No police OR [No fireboard & Peak hour]		<30
	[Daytime & More than 4 response units arrived] OR [Truck involved & More than 1 polices arrived]		>30
	[Snow-iced pavement condition & [Truck involved OR More than 3 response units arrived]] OR [More than 3 vehicles involved & Fireboard first arrived]		>30
	TOC4 center & No auxiliary lane blocked		<30
	Winter & Pickup involved		>30
	Truck involved & [TOC4 center OR More than 1 CHARTs arrived]		>30
	More than 2 CHARTs arrived OR [More than 4 response units arrived & Wet pavement condition]		>30
	More than 1 CHARTs arrived & Pickup involved		>30
	Else then		<30
If	Night & More than 5 response units arrived	Then	>60
	More than 6 response units arrived OR Truck overturned OR Bus involved OR Vehicle lost load		>60
	[Snow-iced pavement condition & Weekend] OR More than 1 trucks involved		>60
	More than 4 response units arrived & [Holiday OR Pickup involved OR More than 3 vehicles involved]		>60
	Else then		<60
If	Truck involved & [Vehicle overturned OR Wet pavement condition OR Snow-iced pavement condition]	Then	>120
	More than 2 vehicles involved & Fireboard first arrived		>120
	Else then		<120

(e) Rules for collisions with property damage and two-travel-lane blockage (CPD2)

	Description		CT
If	Tow service arrived & Fireboard arrived	Then	>30
	Tow service arrived & [More than 2 vehicles involved OR More than 1 auxiliary lanes blocked]		>30
	Truck overturned OR More than 1 shoulder lane blocked OR Truck involved OR Pickup involved		>30
	Snow-iced pavement condition OR Chemical wet pavement condition OR Truck Jack-knifed OR More than 6 response units arrived		>30
	Else then		<30
If	Weekend & Night & More than 2 vehicles involved & Tow service arrived	Then	>60
	Truck involved & More than 5 response units arrived		>60
	More than 4 response units arrived & Wet pavement condition		>60
	Weekend & Vehicle overturned		>60
	Else then		<60
If	More than 9 response units arrived	Then	>120
	More than 5 response units arrived & Pickup involved		>120
	Else then		<120

(f) Rules for collisions with property damage and three-and-more-travel-lane blockage (CPD3+)

	Description		CT
If	Tow service arrived	Then	>30
	Night & [More than 1 CHARTs arrived OR Truck involved]		>30
	Else then		<30
If	More than 1 shoulder blocked & More than 3 response units	Then	>60
	More than 4 vehicles involved OR Holiday		>60
	Else then		<60
If	More than 3 CHARTs arrived OR More than 8 response units arrived	Then	>120
	Else then		<120

Parameter Calibration for k

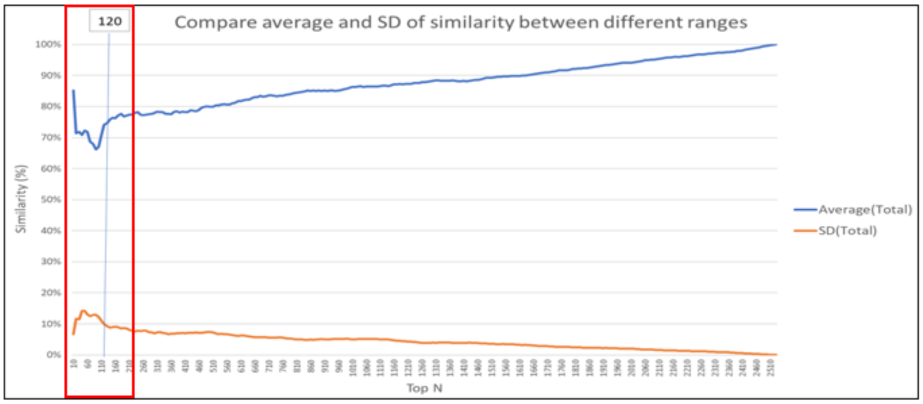
To ensure the stable performance of the LDOF and the LoOP methods, appropriate ranges of parameters for the LDOF ($\boldsymbol{k_{LDOF}}$) and the LoOP ($\boldsymbol{k_{LoOP}}$) methods to compute outlier scores should be found first. Based on the literature (Zhang et al. 2009; Hans-Peter Kriegel et al. 2009), when $\boldsymbol{k_{LDOF}}$/ $\boldsymbol{k_{LoOP}}$ increases to the dimension of the dataset (i.e., the number of variables in the dataset, and for I-695, it is 32), the detection performance rises and remains stable for a wide range of $\boldsymbol{k_{LDOF}}$/ $\boldsymbol{k_{LoOP}}$ values. Accordingly, the calibration consists of the following steps:

- **Step-1**: Take the training dataset as the input and compute the outlier scores for each object using various k_{LDOF} (or k_{LoOP}) (i.e., from a number smaller than 32 to a relatively large number) and then rank these scores in a descending order to get outlier score ranking of each k_{LDOF} (or k_{LoOP});
- **Step-2**: Take the outlier score ranking of k_{LDOF} (or k_{LoOP}) being 32 as the baseline, compute the similarity of outlier score ranking generated under each k_{LDOF} (or k_{LoOP}) in different top-***n*** groups (e.g., the top-10 group indicates top-10 high outlier score data points) by using below equation. For example, given a certain k_{LDOF}, if five data points in the top-10 group of this k_{LDOF} are also in the baseline's top-10 group, then this k_{LDOF}'s similarity of the top-10 group is 50%;

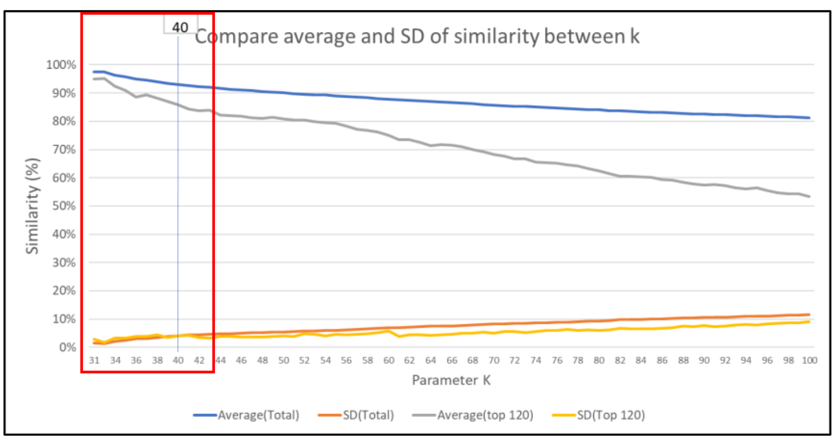
$$Similarity_{n,k} \coloneqq \frac{\#\, data\ points|\ (top\ n\ outlier\ score\ of\ k) \cap (top-n\ outlier\ score\ of\ k_{32})}{n} \times 100\%$$

$$\text{, where } n = \{10, 20, 30, \ldots,\ \#\, data\ points\ in\ the\ dataset\}$$

- **Step-3**: Compute and compare the average and the standard deviation of similarity of different k_{LDOF} (or k_{LoOP}) for each top-***n*** groups as below, and then find out the top-***n*** groups where the average and the standard deviation of similarity fluctuate the most that are marked in the below figure:



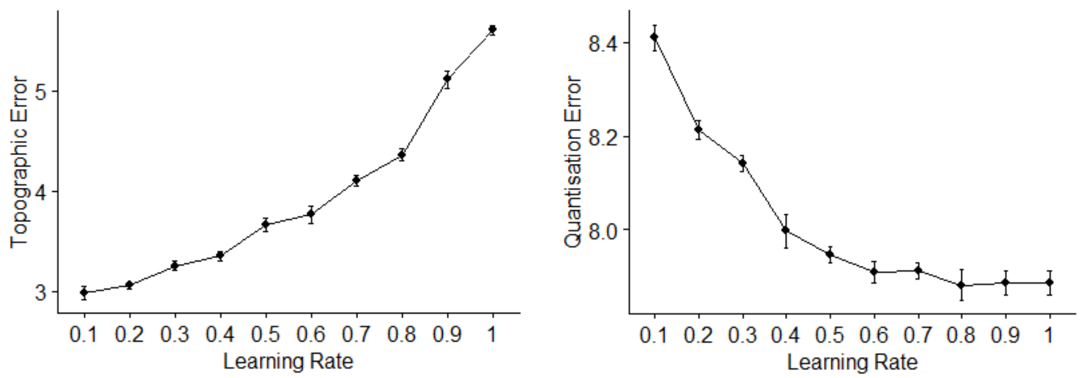
- **Step-4**: Compute and compare the average and the standard deviation of similarity of the meaningful top-***n*** groups and ones of the total data points for each k_{LDOF} (or k_{LoOP}) as below, and finally find out a certain range of k_{LDOF} (or k_{LoOP}) where the similarity is relatively stable.



With such procedures, one can find the performance of the LDOF method is relatively stable when $\boldsymbol{k_{LDOF}}$ is from **31 to 40**. With the same searching procedures, the range of $\boldsymbol{k_{LoOP}}$ is from **24 to 55**.

SOM Learning Rate

The learning rate needs to be calibrated based on the incident data of I-695 by using two measures, topographic and quantization errors. As shown in Figure 1, as the learning rate rises, the topographic error is also increasing, while the quantization error is decreasing, thus it is necessary to trade off these two measures. To select the best learning parameter leading to minimum topographic error and low quantization error, the learning rate is determined as 0.2 since it does not generate a significantly different topographic error with the learning rate being 0.1 but has a lower quantization error by using the Pairwise Wilcoxon Rank Sum test with a 95% confidence level.



(a) (b)

Figure 1. (a) Topographic Errors under Different Learning Rate, (b) Quantization Errors under Different Learning Rate

Bibliography

1. Adler, M.W., Ommeren, J. van and Rietveld, P. (2013). Road congestion and incident duration. *Economics of Transportation*, 2(4), pp.109–118.

2. Agrawal, R. and Ramakrishnan, R. (2015). Fast algorithms for mining association rules in large databases. *IBM Almaden Research Center*.

3. Alkaabi, A.M.S., Dissanayake, D. and Bird, R. (2011). Analyzing Clearance Time of Urban Traffic Accidents in Abu Dhabi, United Arab Emirates, with Hazard-Based Duration Modeling Method. *Transportation Research Record: Journal of the Transportation Research Board*, 2229(1), pp.46–54.

4. Araghi, B.N., Hu, S., Krishnan, R., Bell, M. and Ochieng, W. (2014). A Comparative Study of k-NN and Hazard-based Models for Incident Duration Prediction. *17th International IEEE Conference on Intelligent Transportation Systems*.

5. Bansal, R., Gaur, N. and Singh, S.N. (2016). Outlier Detection: Applications and techniques in Data Mining. *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*.

6. Ben-Gal, I. (2005). Outlier Detection. *Data Mining and Knowledge Discovery Handbook*, pp.131–146.

7. Blei, D.M. and Jordan, M.I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1), pp.121–143.

8. Boyles, S., Fajardo, D. and Waller, S.T. (2007). Naive Bayesian Classifier for Incident Duration Prediction. *trid.trb.org*. Paper presented at the TRB 86th Annual Meeting Compendium of Papers CD-ROM, Washington DC, United States.

9. Breunig, M.M., Kriegel, H.-P., Ng, R.T. and Sander, J. (2000). LOF: Identifying density-based local outliers. *ACM SIGMOD Record*, 29(2), pp.93–104.

10. Byers, S. and Raftery, A.E. (1998). Nearest-Neighbor Clutter Removal for Estimating Features in Spatial Point Processes. *Journal of the American Statistical Association*, 93(442), pp.577–584.

11. Chang, H.-L. and Chang, T.-P. (2013). Prediction of Freeway Incident Duration based on Classification Tree Analysis. *Journal of the Eastern Asia Society for Transportation Studies*, 10, pp.1964–1977.

12. Chimba, D., Kutela, B., Ogletree, G., Horne, F. and Tugwell, M. (2014). Impact of Abandoned and Disabled Vehicles on Freeway Incident Duration. *Journal of Transportation Engineering*, 140(3), p.04013013.

13. Chung, Y. (2010). Development of an accident duration prediction model on the Korean Freeway Systems. *Accident Analysis & Prevention*, 42(1), pp.282–289.

14. Chung, Y. and Yoon, B.-J. (2012). Analytical method to estimate accident duration using archived speed profile and its statistical analysis. *KSCE Journal of Civil Engineering*, 16(6), pp.1064–1070.

15. Chung, Y.-S., Chiou, Y.-C. and Lin, C.-H. (2015). Simultaneous equation modeling of freeway accident duration and lanes blocked. *Analytic Methods in Accident Research*, 7, pp.16–28.

16. Dawei, L. and Cheng, L. (2011). Bayesian Network Classifiers for Incident Duration Prediction. Paper presented at the Transportation Research Board 90thAnnual Meeting, Washington DC.

17. Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), pp.1–38.

18. Dimitriou, L. and Vlahogianni, E.I. (2015). Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions. *Analytic Methods in Accident Research*, 5–6, pp.59–71.

19. Doane, D.P. (1976). Aesthetic Frequency Classifications. *The American Statistician*, 30(4), pp.181–183.

20. Domingues, R., Filippone, M., Michiardi, P. and Zouaoui, J. (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition*, 74, pp.406–421.

21. Fan, W., Bouguila, N. and Ziou, D. (2011). Unsupervised Anomaly Intrusion Detection via Localized Bayesian Feature Selection. *2011 IEEE 11th International Conference on Data Mining*.

22. Garib, A., Radwan, A.E. and Al-Deek, H. (1997). Estimating Magnitude and Duration of Incident Delays. *Journal of Transportation Engineering*, 123(6), pp.459–466.

23. Ghosh, I., Savolainen, P.T. and Gates, T.J. (2012). Examination of factors affecting freeway incident clearance times: a comparison of the generalized F model and several alternative nested models. *Journal of Advanced Transportation*, 48(6), pp.471–485.

24. Giuliano, G. (1989). Incident characteristics, frequency, and duration on a high volume urban freeway. *Transportation Research Part A: General*, 23(5), pp.387–396.

25. Golob, T.F., Recker, W.W. and Leonard, J.D. (1987). An analysis of the severity and incident duration of truck-involved freeway accidents. *Accident Analysis & Prevention*, 19(5), pp.375–395.

26. Hawkins, D.M. (1980). *Identification of Outliers*. Dordrecht: Springer Netherlands.

27. He, Q., Kamarianakis, Y., Jintanakul, K. and Wynter, L. (2013). Incident Duration Prediction with Hybrid Tree-based Quantile Regression. *Complex Networks and Dynamic Systems*, pp.287–305.

28. Hodge, V. and Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2), pp.85–126.

29. Hojati, A., Ferreira, L., Charles, P. and Kabit, M. (2012). Analyzing freeway traffic incident duration using an Australian data set. *Road and Transport Research*, 21, pp.16–28.

30. Hojati, T.A., Ferreira, L., Washington, S. and Charles, P. (2013). Hazard based models for freeway traffic incident duration. *Accident Analysis & Prevention*, 52, pp.171–181.

31. Hojati, T.A., Ferreira, L., Washington, S., Charles, P. and Shobeirinejad, A. (2014). Modelling total duration of traffic incidents including incident detection and recovery time. *Accident Analysis & Prevention*, 71(ISSN 0001-4575), pp.296–305.

32. Hou, L., Lao, Y., Wang, Y., Zhang, Z., Zhang, Y. and Li, Z. (2013). Modeling freeway incident response time: A mechanism-based approach. *Transportation Research Part C: Emerging Technologies*, 28, pp.87–100.

33. Hou, L., Lao, Y., Wang, Y., Zhang, Z., Zhang, Y. and Li, Z. (2014). Time-varying effects of influential factors on incident clearance time using a non-proportional hazard-based model. *Transportation Research Part A: Policy and Practice*, 63, pp.12–24.

34. Hu, J., Krishnan, R. and Bell, M.G.H. (2011). Incident Duration Prediction for In-vehicle Navigation System. Paper presented at the Transportation Research Board annual meeting, Washington DC.

35. Ji, Y. Beibei, Jiang, R., Qu, M. and Chung, E. (2014). Traffic Incident Clearance Time and Arrival Time Prediction Based on Hazard Models. *Mathematical Problems in Engineering*, 2014, pp.1–11.

36. Ji Yang, B., Zhang, X. and Sun, L.J. (2008). Traffic Incident Duration Prediction Based on the Bayesian Decision Tree Method. *Transportation and Development Innovative Best Practices 2008*.

37. Jones, B., Janssen, L. and Mannering, F. (1991). Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis & Prevention*, 23(4), pp.239–255.

38. Kaabi, A.A., Dissanayake, D. and Bird, R. (2012). Response Time of Highway Traffic Accidents in Abu Dhabi. *Transportation Research Record: Journal of the Transportation Research Board*, 2278(1), pp.95–103.

39. Kang, G. and Fang, S. (2011). Applying Survival Analysis Approach to Traffic Incident Duration Prediction. *ICTIS 2011*.

40. Khattak, A., Wang, X. and Zhang, H. (2012). Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays. *IET Intelligent Transport Systems*, 6(2), p.204.

41. Khattak, A.J., Liu, J., Wali, B., Li, X. and Ng, M. (2016). Modeling Traffic Incident Duration Using Quantile Regression. *Transportation Research Record: Journal of the Transportation Research Board*, 2554(1), pp.139–148.

42. Khattak, A.J., Schofer, J.L. and Wang, M.-H. (1995). A SIMPLE TIME SEQUENTIAL PROCEDURE FOR PREDICTING FREEWAY INCIDENT DURATION. *I V H S Journal*, 2(2), pp.113–138.

43. Kim, H.J. and Choi, H.-K. (2001). A COMPARATIVE ANALYSIS OF INCIDENT SERVICE TIME ON URBAN FREEWAYS* *This research was supported by the Korean Research Institute for Human Settlements (KRIHS). *IATSS Research*, 25(1), pp.62–72.

44. Kim, J.S. and Scott, C. (2008). Robust kernel density estimation. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*.

45. Kim, W. and Chang, G.-L. (2011). Development of a Hybrid Prediction Model for Freeway Incident Duration: A Case Study in Maryland. *International Journal of Intelligent Transportation Systems Research*, 10(1), pp.22–33.

46. Kim, W., Chang, G.-L. and Rochon, S. (2008). Analysis of Freeway Incident Duration for ATIS Applications. In: 15th World Congress on Intelligent Transport Systems and ITS America’s 2008 Annual Meeting, New York NY.

47. Kiviluoto, K. (1996). Topology preservation in self-organizing maps. *Proceedings of International Conference on Neural Networks (ICNN’96)*, 1.

48. Knibbe, W.J.J., Alkim, T.P., Otten, J.F.W. and Aidoo, M.Y. (2006). Automated estimation of incident duration on Dutch highways. *IEEE Xplore*, pp.870–874.

49. Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), pp.1464–1480.

50. Kriegel, H.-P., Kröger, P., Schubert, E. and Zimek, A. (2009a). LoOP: local outlier probabilities. *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM ’09*.

51. Kriegel, H.-P., Kröger, P., Schubert, E. and Zimek, A. (2009b). Outlier Detection in Axis-Parallel Subspaces of High Dimensional Data. *Advances in Knowledge Discovery and Data Mining*, pp.831–838.

52. Kriegel, H.-P., S Hubert, M. and Zimek, A. (2008). Angle-based outlier detection in high-dimensional data. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*.

53. Lee, J.-Y., Chung, J.-H. and Son, B. (2008). Analysis of traffic accident size for Korean highway using structural equation models. *Accident Analysis & Prevention*, 40(6), pp.1955–1963.

54. Lee, Y. and Wei, C.-H. (2010). A Computerized Feature Selection Method Using Genetic Algorithms to Forecast Freeway Accident Duration Times. *Computer-Aided Civil and Infrastructure Engineering*, 25(2), pp.132–148.

55. Li, R. (2015). Traffic incident duration analysis and prediction models based on the survival analysis approach. *IET Intelligent Transport Systems*, 9(4), pp.351–358.

56. Li, R., Pereira, F.C. and Ben-Akiva, M.E. (2015). Competing risks mixture model for traffic incident duration prediction. *Accident Analysis & Prevention*, 75, pp.192–201.

57. Li, R., Pereira, F.C. and Ben-Akiva, M.E. (2018). Overview of traffic incident duration analysis and prediction. *European Transport Research Review*, 10(2).

58. Lin, L., Wang, Q. and Sadek, A.W. (2016). A combined M5P tree and hazard-based duration model for predicting urban freeway traffic accident durations. *Accident Analysis & Prevention*, 91, pp.114–126.

59. Lin, P.-W., Zou, N. and Chang, G.-L. (2004). Integration of a Discrete Choice Model and a Rule-Based System for Estimation of Incident Duration: a Case Study in Maryland. In: CD-ROM of Proceedings of the 83rd TRB Annual Meeting, Washington, D.C.

60. Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2008). Isolation Forest. *2008 Eighth IEEE International Conference on Data Mining*.

61. Liu, H., Shah, S. and Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & Chemical Engineering*, 28(9), pp.1635–1647.

62. Lopes, J., Bento, J., Pereira, F.C. and Ben-Akiva, M. (2013). Dynamic Forecast of Incident Clearance Time Using Adaptive Artificial Neural Network Models. Paper presented at the Transportation Research Board 92nd annual meeting Washington DC, 2013-1-13 to 2013-1-17.

63. Malik, K., Sadawarti, H. and G. S, K. (2014). Comparative Analysis of Outlier Detection Techniques. *International Journal of Computer Applications*, 97(8), pp.12–21.

64. Muñoz, A. and Muruzábal, J. (1998). Self-organizing maps for outlier detection. *Neurocomputing*, 18(1–3), pp.33–60.

65. Nam, D. and Mannering, F. (2000). An exploratory hazard-based analysis of highway incident duration. *Transportation Research Part A: Policy and Practice*, 34(2), pp.85–102.

66. National Research Council (U.S.). Transportation Research Board (2016). *Highway capacity manual: a guide for multimodal mobility analysis*. Transportation Research Board, pp.29–13, 17–19.

67. Owens, N., Science Applications International Corporation, American Transportation Research Institute and United States. Federal Highway Administration. Office Of Transportation Operations (2010). *2010 Traffic Incident Management Handbook Update*. Washington, Dc: Federal Highway Administration.

68. Ozbay, K. and Kachroo, P. (1999). Incident Management in Intelligent Transportation Systems. *Electrical and Computer Engineering Faculty Publications*, pp.1–248.

69. Ozbay, K. and Noyan, N. (2006). Estimation of incident clearance times using Bayesian Networks approach. *Accident Analysis & Prevention*, 38(3), pp.542–555.

70. Park, H., Haghani, A. and Zhang, X. (2015). Interpretation of Bayesian neural networks for predicting the duration of detected incidents. *Journal of Intelligent Transportation Systems*, 20(4), pp.385–400.

71. Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3), pp.1065–1076.

72. Peeta, S., Ramos, J. and Gedela, S. (2000). Providing Real-Time Traffic Advisory and Route Guidance to Manage Borman Incidents On-Line Using the Hoosier Helper Program. *Joint Transportation Research Program, 1284 Civil Engineering Building, Purdue University, West Lafayette, Indiana 47907-1284*.

73. Pereira, F.C., Rodrigues, F. and Ben-Akiva, M. (2013). Text analysis in incident duration prediction. *Transportation Research Part C: Emerging Technologies*, 37, pp.177–192.

74. Qi, Y. and Teng, H. (2008). An Information-Based Time Sequential Approach to Online Incident Duration Prediction. *Journal of Intelligent Transportation Systems*, 12(1), pp.1–12.

75. Quinn, J.A. and Sugiyama, M. (2014). A least-squares approach to anomaly detection in static and sequential data. *Pattern Recognition Letters*, 40, pp.36–40.

76. Schölkopf, B., Williamson, R.C., Smola, A.J., Shawe-Taylor, J. and Platt, J.C. (2000). Support vector method for novelty detection. pp.582–588. Advances in Neural Information Processing Systems 12, MIT Press.

77. Schrank, D., Eisele, B. and Lomax, T. (2019). *Urban Mobility Report*. The Texas A&M Transportation Institute with cooperation from INRIX.

78. Shen, L. and Huang, M. (2011). Data Mining Method for Incident duration Prediction. *Communications in Computer and Information Science*, pp.484–492.

79. Shyu, M.-L., Chen, S.-C., Sarinnapakorn, K. and Chang, LW. (2003). A Novel Anomaly Detection Scheme Based on Principal Component Classifier. In: *ICDM Foundation and New Direction of Data Mining workshop*. pp.172–179.

80. Smith, K. and Smith, B. (2002). Forecasting the Clearance Time of Freeway Accidents. *Center for Transportation Studies, University of Virginia, Charlottesville*.

81. Tan, H.S. and George, S.E. (2004). Investigating Learning Parameters in a Standard 2-D SOM Model to Select Good Maps and Avoid Poor Ones. *Lecture Notes in Computer Science*, 3339, pp.425–437.

82. Tarassenko, L. (1995). Novelty detection for the identification of masses in mammograms. *4th International Conference on Artificial Neural Networks*.

83. Tipping, M.E. and Bishop, C.M. (1999). Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), pp.611–622.

84. Valenti, G., Lelli, M. and Cucina, D. (2010). A comparative study of models for the incident duration prediction. *European Transport Research Review*, 2(2), pp.103–111.

85. Vlahogianni, E.I. and Karlaftis, M.G. (2013). Fuzzy-Entropy Neural Network Freeway Incident Duration Modeling with Single and Competing Uncertainties. *Computer-Aided Civil and Infrastructure Engineering*, 28(6), pp.420–433.

86. Wang, J., Cong, H. and Qiao, S. (2013). Estimating freeway incident duration using accelerated failure time modeling. *Safety Science*, 54, pp.43–50.

87. Wang, W., Chen, H. and Bell, M. (2002). A Study of the Characteristics of Traffic Incident Duration on Motorways. *Traffic And Transportation Studies (2002)*.

88. Wang, W., Chen, H. and Bell, M.C. (2005). Vehicle Breakdown Duration Modelling. *Journal of Transportation and Statistics*, 8(1), pp.75–84.

89. Wei, C.-H. and Lee, Y. (2005). Applying Data Fusion Techniques to Traveler Information Services in Highway Network. *Journal of the Eastern Asia Society for Transportation Studies*, 6, pp.2457–2472.

90. Wei, C.-H. and Lee, Y. (2007). Sequential forecast of incident duration using Artificial Neural Network models. *Accident Analysis & Prevention*, 39(5), pp.944–954.

91. Weng, J., Qiao, W., Qu, X. and Yan, X. (2015). Cluster-based lognormal distribution model for accident duration. *Transportmetrica A: Transport Science*, 11(4), pp.345–363.

92. Won, M. (2020). Outlier Analysis to Improve the Performance of an Incident Duration Estimation and Incident Management System. *Transportation Research Record: Journal of the Transportation Research Board*, p.036119812091647.

93. Won, M. and Chang, G.-L. (2019). Ensemble Model to Estimate Incident Clearance Durations using Sequential Partitioning Process and Robust Regression. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(8), pp.554–566.

94. Won, M., Kim, H. and Chang, G.-L. (2018). Knowledge-Based System for Estimating Incident Clearance Duration for Maryland I-95. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(14), pp.61–72.

95. Yang, P. and Huang, B. (2008). KNN Based Outlier Detection Algorithm in Large Dataset. *2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing*.

96. Yu, B. and Xia, Z. (2012). A Methodology for Freeway Incident Duration Prediction Using Computerized Historical Database. *CICTP 2012*.

97. Zhan, C., Gan, A. and Hadi, M. (2011). Prediction of Lane Clearance Time of Freeway Incidents Using the M5P Tree Algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), pp.1549–1557.

98. Zhang, H. and Khattak, A.J. (2010). Analysis of Cascading Incident Event Durations on Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, 2178(1), pp.30–39.

99. Zhang, K., Hutter, M. and Jin, H. (2009). A New Local Distance-Based Outlier Detection Approach for Scattered Real-World Data. *Advances in Knowledge Discovery and Data Mining*, pp.813–822.

100. Zou, Y., Henrickson, K., Lord, D., Wang, Y. and Xu, K. (2016). Application of finite mixture models for analyzing freeway incident clearance time. *Transportmetrica A: Transport Science*, 12(2), pp.99–115.