공학박사학위논문

Probabilistic Prediction of Traffic States Using Bayesian Network

베이지안 네트워크를 활용한 교통상태의 확률론적 예측

2017년 8월

서울대학교 대학원

공과대학 건설환경공학부

박 호 철

Probabilistic Prediction of Traffic States Using Bayesian Network

지도교수 고 승 영

이 논문을 공학박사 학위논문으로 제출함.
2017년 5월

서울대학교 대학원
공과대학 건설환경공학부
박 호 철

박호철의 박사 학위논문을 인준함.
2017년 6월

위 원 장 이 청 원 (인)

부위원장 고 승 영 (인)

위 원 송 준 호 (인)

위 원 김 동 규 (인)

위 원 장 기 태 (인)

Abstract

Traffic state prediction is an important issue in traffic operations. One of the main purposes of traffic operations is to prevent flow breakdown. Therefore, it is necessary to predict the traffic state in such a way as to reflect the stochastic process of traffic flow. However, the traffic state transition is affected complexly and simultaneously by many factors, which lead to a lack of understanding and accurate prediction. Meanwhile, the Bayesian network is a methodology that not only is suitable for a problem with uncertainty but also can improve the understanding of a problem. Also, it is possible to derive fair probability with incomplete information, which allows the analysis of various situations. In this study, we developed a traffic state prediction model using the Bayesian network to reflect dynamic and stochastic traffic flow characteristics. In order to improve the structure of the Bayesian network, which has been used simply in transportation problems, we proposed a modeling procedure using mixture of Gaussians (MOGs). Also, spatially extended variables were used to consider the spatiotemporal evolution of traffic flow patterns. In particular, traffic state identification was performed by estimating the link speed in order to consider the spatial propagation of congestion. In the performance evaluation, the Bayesian network had better performance than logistic regression and had the same level of performance as artificial neural network based on machine learning. Also, by performing sensitivity analyses, we provided the understanding of the traffic state prediction and the guidelines for model

improvement. Therefore, the Bayesian network developed in this study can be considered as a traffic state prediction model with good prediction accuracy and provides insights for traffic state prediction.

Keywords: Bayesian network, traffic state prediction, flow breakdown, probabilistic model, stochastic process

Student Number: 2011-20980

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research Background and Purpose

1.1.1 Importance of traffic state prediction

Traffic state prediction is an important issue in traffic information systems. Much research has developed a traffic state prediction model and tried to obtain specific values of traffic variables, such as flow, speed, and density, which can describe the traffic state. These values are useful for drivers and operators of a freeway in term of traffic information.

Traffic state prediction is also important in terms of traffic operations. One of the main purposes of traffic operations is to prevent and impede flow breakdown (transition from free-flow traffic state to congested traffic state) because it takes a long time to recover if flow breakdown occurs. Therefore, it is necessary to predict whether the future traffic state is congested or not and to respond to the future traffic state in advance. However, the predicted specific values of the traffic variables mentioned above have limitations in clearly predicting the traffic state. This is because traffic flow is intrinsic with stochastic and dynamic characteristics. In particular, the traffic state around the capacity is highly uncertain since the flow breakdown, as abrupt traffic state transition, is a stochastic process or a probabilistic event. Therefore, a model that directly predicts the traffic state is needed for actual traffic operations.

1.1.2 Characteristics of traffic state

Traffic state is affected by many factors such as the traffic load from upstream, back of queue caused by breakdown and perturbations from downstream, merging and diverging behaviors of vehicles entering from a ramp or exiting from a mainline, and behaviors of individual drivers in a target area. These factors are complex and simultaneous for changes in the traffic state. Thus, the traffic state transition (e.g., flow breakdown) is a stochastic process and only predictable in term of probabilities. Therefore, it is necessary to adopt a methodology that reflects this stochastic characteristic as probabilities for traffic state prediction. Also, the model developed in the study methodology should not only be effective for computation performance but also to enhance the understanding of the traffic state transition mechanism.

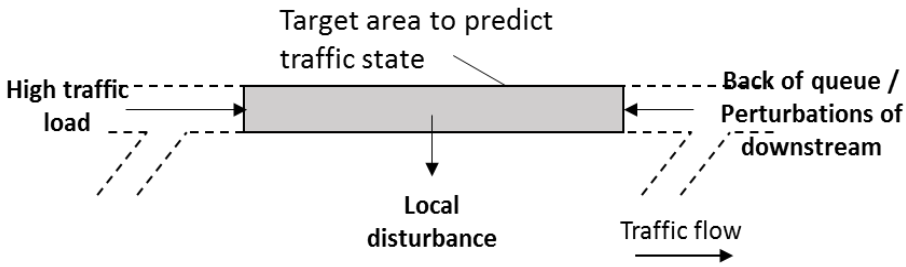


Figure 1.1 Complex and simultaneous situation of traffic state

1.1.3 Limitations of stationary loop detector

Traffic congestion pattern can be considered as both a temporal and a spatial phenomenon. The traffic variables speed (μ), flow (q), and density (k) (or occupancy (o)) vary along a section. In homogeneous traffic situations such as free-flow and severe congestion, these variables on both ends of the section do not have a significant difference. However, in transition periods from free-flow to congestion or vice versa, these variables show significant differences. However, stationary loop detectors used in traffic management systems have limitations in detecting spatiotemporal change of the traffic state. Due to the fixed position of the detector, only a part of the traffic stream can be sampled. In addition, there are technical limitations including measurement errors and some missing values. Therefore, we need a prediction methodology that obtains robust results in the limitations of the existing detectors.

1.1.4 Bayesian network for traffic state prediction model

We adopted a Bayesian network as a probabilistic model in order to develop a traffic prediction model that reflects the stochastic characteristic of traffic flow and copes with the limitation of stationary loop detectors. The Bayesian network is a graphical and probabilistic model and is known as a methodology suitable for uncertain knowledge or stochastic process problems. In addition, the structure of the model can be constructed based on existing knowledge and

has the advantage of being able to infer various situations (e.g., detector data with missing values are available). In spite of these advantages, there have been few studies that use the Bayesian network for traffic state prediction. This study can be a starting point for traffic state prediction using a Bayesian network.

1.1.5 Research purpose

The purpose of this study is to develop a probabilistic prediction model that reflects the complex and simultaneous situation of the traffic state using the Bayesian network. We propose a Bayesian network modeling method for traffic state prediction, which can improve naive Bayesian network structures used in previous studies. The final objective of this study is to evaluate the performance of the developed model and to analyze the developed model for insights in understanding the traffic state transition mechanism. Therefore, the Bayesian network developed in this study can be considered as a traffic state prediction model with good prediction accuracy and interpretability.

1.2 Research Scope and Procedure

1.2.1 Research scope

This study focuses on uninterrupted flow (e.g., freeway) and includes both on- and off-ramps as well as the mainline of a freeway or expressway. The Bayesian network, the methodology of this study, is a data-driven prediction model based

on machine-learning. Field data used for training are limited to stationary loop detector data. Stationary loop detector data, section detector data, and prove car data are available in the field data. However, the stationary loop detector data is the most universal in the world and can be collected continuously for 24 hours. Future research using various data will be possible. In this study, the loop detector data (30 seconds aggregated flow, speed, and occupancy) were provided by the California Department of Transportation (Caltrans) Performance Measurement System (PeMS). The collected data was used in the model by removing and correcting the anomaly through a validity check.

In addition, the model developed in this study covers normal traffic conditions excluding the effects of weather, incidents, and construction. Normal traffic conditions are divided into congestion and non-congestion, and include both spontaneous and induced breakdown.

1.2.2 Research composition and procedure

As shown in Figure 1.2, we conduct a literature review, data collection and preparation, Bayesian network modeling, evaluation, and conclusions. The paper is divided into six chapters and is organized according to the research procedure.

Chapter 2 summarizes the literature review. First, we briefly review characteristics of the traffic state in previous studies. Second, we review traffic state estimations and prediction models based on these characteristics. Third,

we describe the theoretical background of a Bayesian network, which is the methodology of this study, and studies using a Bayesian network in the transportation field. Based on this review, we derive the originality and contributions of this study.

Chapter 3 describes the data collection and preparation. First, we describe data specifications such as the area and the time of data collection, and we describe a validity check process for outlier correction of data. Second, we propose a link speed estimation method for traffic state identification. Third, we summarize the description of the final data for the analysis.

Chapter 4 provides Bayesian network modeling. We divide Bayesian network modeling into structure construction and parameter learning. In the structure learning, we define nodes for traffic state prediction and we describe the process of establishing links between the nodes based on existing traffic engineering knowledge. In the parameter learning, the conditional probability distribution of each node is trained when given the constructed structure.

Chapter 5 describes the model evaluation. We evaluate the performance of the Bayesian network and compare the performance with other methodologies based on parametric and nonparametric approach methods. Also, we perform a sensitivity analysis using the developed Bayesian network in order to provide insights on traffic state prediction and we discuss future improvement directions.

Chapter 6 summarizes the contents of this study and presents guidelines for traffic state prediction, limitations of the study, further

applications, and future research.

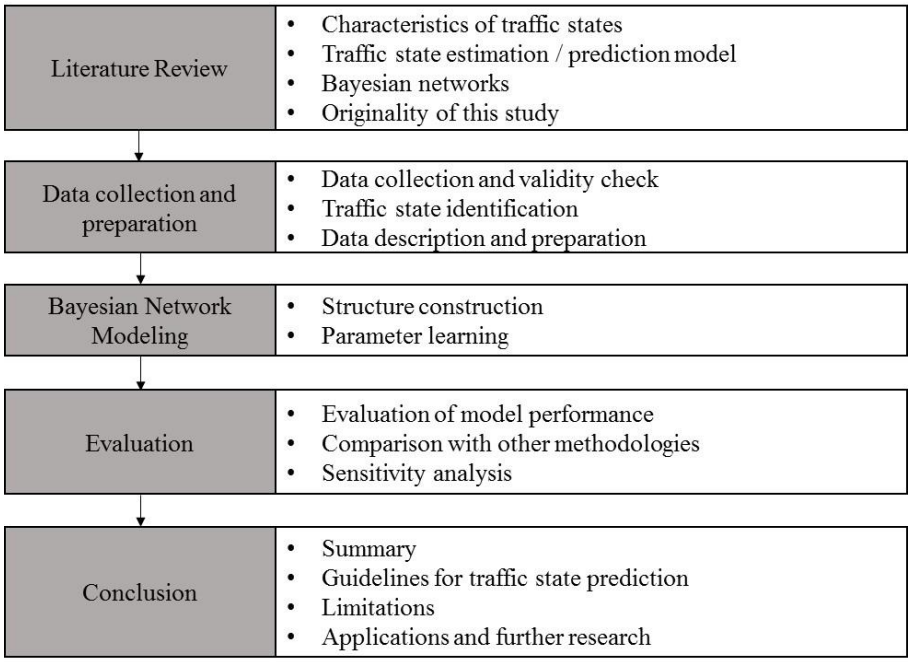


Figure 1.2 Research Procedure

Chapter 2. Literature Review

2.1 Characteristics of Traffic State

2.1.1 Definition of traffic state

A traffic state represents a homogeneous group with similar characteristics and can be described by traffic variables such as flow, density, and speed (Antoniou et al., 2013). Various studies have been conducted to explain and classify traffic states.

May (1990) discussed single-regime and multi-regime models while explaining traffic flow fundamentals. Edie (1961) proposed a two-regime model, i.e., free-flow and congested-flow, to achieve higher model performance, and the Northwestern University research team developed a three-regime model, i.e., free-flow, transitional-flow, and congested-flow (May and Keller, 1967; Drake et al., 1967). The multi-regime models have strengths and weaknesses compared to single-regime models. Some studies have continued to overcome the limitations (Sun and Zhou 2005; Wang et al., 2011).

In empirical data study, two-phase traffic theory (phase transition due to flow breakdown and recovery) was widely used for the sake of simplicity (Elfteriadou et al. 1995; Persaud et al., 1998). Kerner (2004) proposed three-phase traffic theory, i.e., free-flow phase, wide moving jams, and synchronized flow, to improve empirical accuracy. Vocal criticism to the three-phase traffic

theory is being raised (Antonious et al., 2013).

Various studies have classified traffic states into three or more according to the purpose of the study. Antonious et al. (2013) classified traffic states into from three to eight traffic states to predict the traffic states for improvement of the speed prediction accuracy. Noroozi and Hellinga (2014) classified the traffic states into four traffic states and applied a Markov model to predict the traffic states. Wang et al. (2010) classified the traffic states into 12 and 14 traffic states and made the Markov models for calculating of traffic flow breakdown probability

In this study, the future traffic state, which is a prediction target, was assumed to be two-phase, and the current traffic state of each area was classified based on the performance of the Bayesian network.

2.1.2 Traffic flow breakdown: stochastic event to determine the traffic state

If the traffic state is classified into free-flow and congested traffic state, a breakdown is considered to be an important phenomenon as a criterion for distinguishing the traffic states. When the breakdown occurs, a queue is formed, and its effect is propagated. In particular, the breakdown is important because it has a direct relationship with the capacity of a freeway.

The term “capacity” in the freeway traffic operations is a key aspect and a description of the vehicle-carrying ability of a freeway (Lorenz and Elefteriadou, 2001; Kondyli et al., 2013). When traffic demand exceeds the

capacity, “breakdown,” which is the transition of traffic flow from free-flow into congested conditions, has generally been thought to occur (Kondyli et al., 2013). The capacity has been traditionally believed to be a known constant. However, much research in recent years has shown that the breakdown can occur at various traffic flows. It indicates that a breakdown is a stochastic event and the capacity is changed into a random variable. Therefore, understanding the stochastic breakdown phenomena is important to prevent the traffic congestion.

Flow breakdowns are caused by the simultaneous action of three factors (Treiber and Kesting, 2013):

1. High traffic load (temporal aspect)
2. A bottleneck (spatial aspect, macroscopic flow instability)
3. Disturbance caused by individual drivers (the trigger)

The high traffic load is the most obvious factor. If the traffic load is sufficiently small, the traffic flow can be recovered to the stable condition without growing and propagating of a queue even if a disturbance caused by individual drivers occurs. In addition, the bottleneck can be defined as a local reduction of the freeway capacity, which is a weak point where the flow breakdown occurs when there are the high traffic load and the disturbance. Finally, the third factor is necessary for the flow breakdown. The disturbances, which are caused by individual drivers, includes abruptly braking, lane changing, and so on. Since

this is a single-vehicle nature, it is difficult to identify directly in aggregate detector data. However, the disturbance caused by individual drivers can be indirectly reflected in the model by using a density. Distance gaps of vehicles can be directly attributed to the density. Therefore, the distance gap becomes smaller as the density increases, and the probability of the breakdown due to the driver disturbance can increase as the density increases.

The occurrence of the flow breakdowns could be due to a local perturbation in the traffic flow at the considered freeway section (i.e., spontaneous breakdown) or caused by an external disturbance associated with a queue spillback from a downstream location (i.e., induced breakdown) (Dong and Mahmassani, 2012). On the other hand, most studies on the breakdown considered only the spontaneous breakdown for clear estimation, prediction, and interpretation. However, a universal model, which can take into account all the causes of the breakdown in the real world, is needed to be utilized in practical traffic flow managements.

We cannot make predictions about the location and time of individual flow breakdowns (Treiber and Kesting, 2013). Nevertheless, we can make statements about the probability that a breakdown will occur on a given freeway section in a certain period.

2.1.3 Spatiotemporal evolution of congested traffic patterns

Freeway bottleneck activation is difficult to identify but important. The

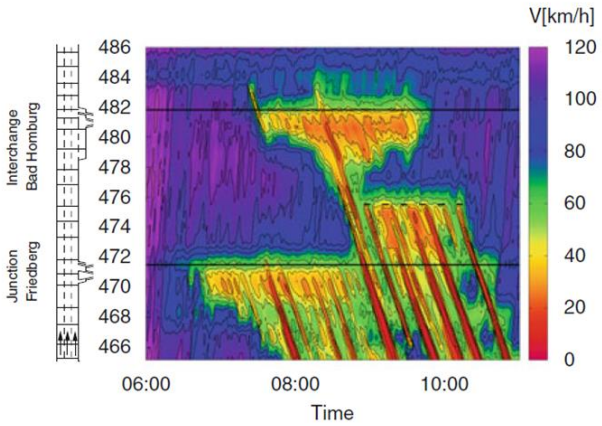
spatiotemporal evolution of congestion from downstream to upstream is extremely difficult to quantify, and it varies from one day to another (Elhenawy and Rakha, 2015). Treiber et al. (2010, 2013) discussed the congested traffic patterns using empirical data and simulated data. In this part, we reviewed the characteristics of the congested traffic patterns based on the work of Treiber et al. (2010, 2013).

The congestion patterns are most often caused by the breakdown in the bottleneck, which is affected by various reasons. The beginning of congestion is a bottleneck caused by a variety of causes, and even in large distances on a driver under congestion. It is likely that the congestion is due to downstream bottlenecks. These bottlenecks can also vary in location and extent depending on the effects of on- and off-ramp, local narrowings, lane reduction, curvature, gradients, and accidents.

Spatiotemporal evolution of multiple bottleneck situations in reality also makes the congested traffic pattern more complex. The congested traffic patterns are rarely isolated from each other and interact with each other. First, moving jams may activate upstream bottlenecks and the moving traffic wave varied by the interaction with the bottlenecks. For example, moving jam caused by accident can activate a recurrent bottleneck in the upstream and create a strong recurrent bottleneck due to the accident. Secondly, the bottleneck activation may alleviate the downstream congested traffic state. A bottleneck in the upstream causes a capacity drop, which can reduce the downstream flow. However, in multiple bottleneck situations, both upstream and downstream

bottlenecks can be affected and it may be difficult to interpret them clearly.

Next, the propagation velocities of downstream and upstream fronts of moving jams have different characteristics as shown in Figure 2.1. The downstream front is either stationary or moves upstream at a constant velocity which is between -20 km/h and -15 km/h. The "downstream front" is the transition zone where drivers leave the congested zone, and it is stationary or moves upstream according to the condition of the bottleneck. On the other hand, the upstream front does not have a characteristic speed. The upstream front propagates upstream or downstream depending on traffic demand of upstream and bottleneck capacity. When the upstream front and the downstream front meet, the traffic jam is resolved. This is called “recovery” in which the traffic state transition occurs in the opposite direction of the breakdown (congested traffic state→ free-flow traffic state).



Source: Treiber et al. (2013)

Figure 2.1 Spatiotemporal local speed on the German highway A5

However, more empirical studies are still required for some contents. Despite the understandings of the congested traffic patterns mentioned above, insights for the congested traffic patterns (breakdown activation and queue formation and propagation) are still insufficient.

2.2 Traffic State Estimation and Prediction

In this study, the research on traffic state estimation and prediction is divided into three parts, i.e., traffic flow and speed estimation and prediction, traffic state estimation and prediction (directly estimating and predicting the traffic state), and flow breakdown probability estimation and prediction. Also, each study was distinguished by a parametric approach or a nonparametric approach.

Various modeling approaches have been used to estimate or predict the traffic state. Parametric approach models have a finite number of parameters, while nonparametric approach models have the number of parameters is (potentially) infinite. In other words, the complexity of the model grows with the number of training data in nonparametric models. For example, logistic regression, and linear SVM are considered as the parametric models. In contrast, K-nearest neighbor, neural network, or RBF kernel-based SVMs are considered as the nonparametric models because the number of parameters grows with the data size. Table 2.1 shows the literature review of traffic state estimation and prediction.

Table 2.1 Literature review of traffic state estimation and prediction

	Parametric approach	**Nonparametric approach**
Traffic flow & speed estimation/prediction	• ARIMA family of models • Kalman filtering models	• Neural networks • Nearest neighbors • Support vector regressions • Bayesian networks
Traffic state estimation/prediction	• Logistic regression	• Clustering–based models • Markov models • Neural networks • Support vector machine
Flow breakdown probability estimation/prediction	• Empirical Analysis • Survival analysis • Car-following-based models • Platoon-based models	• Markov models • Bayesian networks

2.2.1 Traffic flow and speed estimation and prediction

Much research estimated and predicted traffic flow or speed to know the traffic state. Some research estimated and predicted occupancy or density to know the traffic state. The research focused on the calculation of specific value, and the models were divided into the parametric and nonparametric approach.

The parametric approach includes Auto-regressive Integrated Moving Average (ARIMA) family of models and Kalman filtering models.

2.2.1.1 ARIMA family of models

Auto-regressive Integrated Moving Average (ARIMA) is one of the most widely used time series models and has main goal to minimize the noises (Oh et al., 2015). To improve the performance of the ARIMA, also, the extension of ARIMA class has been developed (e.g., Seasonal ARIMA).

Ahmed and Cook (1979) developed the ARIMA models by using the Box and Jenkins technique to predict the freeway traffic volume and occupancy. The paper suggested that the traffic forecasting models help to determine the control strategies for ramp metering, incident detection and VMS. Hamed et al. (1995) developed the ARIMA models by using the Box and Jenkins technique to forecast traffic volume in urban arterials. Williams et al. (1998) used seasonal ARIMA and Winters exponential smoothing methods to predict urban freeway traffic flow. They compared proposed models with existing models to evaluate the accuracy performance. The results show that seasonal ARIMAs outperform the other models. Chandra and Al-Deek (2009) developed a vector autoregressive (VAR) model to consider the effect of upstream and downstream location information in the case of freeway traffic prediction. They found that the past information of upstream and downstream affected the future traffic flow, so the VAR outperformed the traditional ARIMA model for traffic prediction. Min and Wynter (2011) adopted a multivariate spatial-temporal autoregressive (MSTAR) model to develop a highly accurate and scalable method.

2.2.1.2 Kalman filtering models

Kalman filtering is a prediction method using continuous updates of state variables. Therefore, it uses both smoothed historical data and current measurements. Okutani and Stephanedes (1984) employed the Kalman filtering to predict short-term traffic volume. The average prediction error was less than 9%, and it indicated that Kalman filtering models performed substantially better than the previous model. Sun et al. (2003) used mixture Kalman filtering in which switching state-spate model and Kalman filtering were combined. Also, they used a cell transmission model based switching state-spate model to estimate unobserved densities and traffic states on a highway section. The results show the improved performance. Wang and Papageorgiou (2005) used the extended Kalman filtering to estimated freeway traffic state and conducted a number of simulation investigation to evaluate the estimator performance. They concluded that the estimator can track real-time changes of the model parameters.

The nonparametric approach includes neural network, nearest neighbors, support vector regression, and Bayesian network.

2.2.1.3 Neural network

Florio and Mussone (1996) used the neural network models to forecast the traffic flow. The models include various variables, i.e., traffic flow, speed, density, the percentage of heavy goods vehicles, brightness, weather conditions,

visibility, and the presence of VMS. They concluded that this method was easy to implement for prediction of traffic variables. Vlahogianni et al. (2005) predicted the arterial traffic flows using multilayer feed-forward perceptrons (MLPs) which is one of the main categories of ANNs. Also, this study suggested multilayered structural optimization strategy and considered spatiotemporal characteristics of traffic flow by selecting detector data located properly. The proposed models in this study showed quite robust results. Zheng et al. (2006) proposed combined neural network model with an adaptive single neural predictors and heuristic credit assignment algorithm based on Bayes' rule. They found that the combined model outperforms the singular predictors.

2.2.1.4 Nearest neighbors

Davis and Nihan (1991) used the k-nearest neighbor (k-NN) approach to forecast short-term traffic flow. In this study, the k-NN approach was compared with simple univariate linear time-series forecasts. However, the results did not show that k-NN is better than simple univariate linear time-series forecasts. In addition, to improve the performance of the k-NN, larger databases were emphasized. Smith and Demetsky (1994) compared the neural network approach and the nearest neighbors approach for prediction of short-term traffic flow. As a result, the nearest neighbor model performed better than the neural network model. This study suggested that nearest neighbor models have portable and sufficiently accurate for the application. Lam et al. (2006)

developed not only the k-nearest neighbor, but also ARIMA, neural network, and Gaussian maximum likelihood to forecast short-term hourly traffic flow in Hong Kong. They pointed out that the k-nearest neighbor model, which has a good performance in dynamic nature of the common traffic patterns, is recommended.

2.2.1.5 Support vector regression

Wang and Shi (2013) developed a support vector regression (SVR) model to predict the short-term traffic speed. To improve the SVR model, they proposed new kernel function using a wavelet function and used phase space reconstruction theory to identify the model structure parameters. The model was evaluated using the real traffic speed data and results were encouraging.

2.2.1.6 Bayesian network

Sun et al. (2006) used Bayesian networks as a new approach to predict the traffic flow. The network in Bayesian networks modeled using the traffic flows among adjacent road links in the networks. They pointed out that Bayesian networks can work when incomplete data exist.

2.2.2 Traffic state estimation and prediction

Unlike traffic flow and speed estimation and prediction, there are some studies that directly estimate or predict the traffic state. These studies mainly used

models based on clustering and classification techniques.

The parametric approach involves logistic regression.

2.2.2.1 Logistic regression

Li et al. (2010) used ordered logistic regression to prediction traffic state. This study classified traffic states based on the average vehicle speed. In addition, this study considered the time and weather condition, which has a great impact on traffic state. Huili et al. (2011) developed a logistic regression model to predict traffic state probability. They concluded that the model could predict the accurate and objective traffic state to meet the demand of traffic guidance.

The nonparametric approach includes clustering-based models, Markov models, neural networks, and support vector machines.

2.2.2.2 Clustering-based models

Xia and Chen (2007) proposed a nested clustering method to classify traffic state. This method was flexible to determine the number of traffic state, and it was effective in identifying traffic state from historical data. However, it had difficulty implementing the method due to its computation time. Xia et al. (2012), therefore, proposed real-time clustering procedure to identify traffic state. This procedure included the cluster evolving step after offline clustering for traffic state initialization. The results showed that the proposed procedure was promising.

2.2.2.3 Markov models

Antoniou et al. (2007, 2013) developed a traffic state prediction model by combining model-based clustering and variable-length Markov chain. They pointed out that traffic state identification could result in the more accurate prediction model. Noroozi and Hellinga (2014) developed Markov model to predict traffic state by characterizing the transition between traffic states. Also, the proposed model considered time-varying covariates. The prediction of travel speed is improved by using the transition probabilities of this model.

2.2.2.4 Neural networks

Florio and Mussone (1996) used the neural network models to forecast the traffic state (i.e., stable, critical, and unstable). The models include various variables: traffic flow and density, percentage of heavy goods vehicles, brightness, visibility, and presence of VMS. They concluded this method was easy to implement for prediction of traffic variables.

2.2.2.5 Support vector machines

Deng et al. (2009) proposed a pattern-based approach that combines clustering and classification method to estimate the traffic state. They adopted fuzzy-set clustering method and multiclass support vector machine, and the result shows the proposed method is promising for dynamic estimation of traffic state. Sun et al. (2012) applied parallel support vector machine to improve the training

speed. General SVM have difficulty using in practical application due to the expensive training computation cost. In example analysis, the computation speed was improved without reducing the accuracy performance.

2.2.3 Flow breakdown probability estimation and prediction

As noted, because the traffic state transition (e.g., flow breakdown) is a stochastic process, it is only predictable on the probabilities in principle (Treiber and Kesting, 2013). If the traffic states are divided into two-phase as free-flow and congested, the transition from free-flow to congested is caused by flow breakdown, and then the capacity of the freeway decreases. Much research has attempted to establish the occurrence of flow breakdown as a probabilistic model, which is different from the methodology mentioned above.

The parametric approach in the flow breakdown probability model includes empirical analysis, survival analysis, car-following-based models and platoon-based models.

2.2.3.1 Empirical analysis

Elefteriadou et al. (1995) developed the probabilistic model for breakdown process at ramp-freeway junction using visual examination. They concluded that breakdown is a probabilistic event and is a function of ramp-vehicle cluster occurrence. Persaud et al. (1998) aimed at exploring a breakdown and a capacity drop in traffic flow. Therefore, they developed the probabilistic

breakdown model at various traffic levels, which has practical significance and are useful for improving the metering system. Lorenz and Elefteriadou (2001) examined the freeway breakdown process in detail and defined the breakdown occurrence and the breakdown flow rate. Also, they developed the probabilistic breakdown model at various data intervals and proposed a probabilistic capacity definition. Persaud et al. (2001) quantified the breakdown probability as flow function of the critical location. They used a logistic model in which the breakdown probability was estimated.

2.2.3.2 Markov Model

Evans et al. (2001) applied Markov chain model to develop breakdown probability model. This model based on the zonal merging probabilities considering freeway flow, available gaps, and driver's actions can predict the breakdown. The results showed higher arrival rates lead to higher breakdown probabilities. Dong and Mahmassani (2009) proposed breakdown probability model using a discrete time Markov chain considering the stochasticity in traffic flow patterns. Using the transition probability matrix, the breakdown and recovery probabilities given traffic states were estimated. Also, they analyzed travel time reliability using the breakdown probability. Noroozi and Hellinga (2014) developed Markov model to predict traffic state by characterizing the transition between traffic states. Also, the proposed model considered time-varying covariates. Using the transition probabilities of this model, the prediction of travel speed is improved.

2.2.3.3 Survival analysis

Brilon et al. (2005) developed capacity distribution function considering breakdown phenomena to understand freeway capacity. The function was estimated by using Product Limit Method (PLM) based on the survival analysis. Also, they assumed that the capacity is Weibull-distributed. The function is expanded into reliabilities of freeway networks. Kim et al. (2010) studied the effect of rain on traffic flow breakdown. To achieve this goal, they developed the flow breakdown probability models for rain and no rain (clear) weather conditions using the survival analysis. They found that flow drop is not much difference between rain and no rain conditions. Elefteriadou et al. (2011) applied Product Limit Method (PLM) to develop breakdown probability model and proposed a new method for ramp metering based on the breakdown probability model. The results showed that the method could postpone the beginning of the breakdown and reduced the duration of congested periods.

2.2.3.4 Car-following-based model

Son et al. (2004) developed a probabilistic breakdown model based on wave propagation model which explains the stochastic movements of traffic waves and disturbances with the levels of gap time. This model includes four variables (i.e., spacing, gap distance, gap time, and headway) describing the spatial and temporal separations between two vehicles. Xu et al. (2013) developed an evolution model of speed perturbation for prediction of traffic flow breakdown

on congested traffic state. They pointed out that most important variables are time headway, driver behavior, and stochastic nature of state transitions for the flow breakdown prediction. The simulation results showed the proposed model is promising for field application.

2.2.3.5 Platoon-based model

Shiomi et al. (2011) proposed stochastic processes models based on traffic flow dynamics. The models included platoon formation behind a bottleneck and speed transitions within a platoon. The models were applied to calculate breakdown probability depending on traffic flow rate. Kühne and Lüdtke (2013) proposed the first passage time probability distribution, which is the probability when firstly hitting the critical platoon length, and calculated the breakdown probability with this probability distribution.

2.2.3.6 Bayesian network

Armstrong (2011) applied Bayesian network to predict the traffic state. The model included the mainline flow and ramp flow in the target area. The probability was calculated deterministically based on a logit formulation. It was a simple and limited probabilistic model.

2.2.4 Literature review results of traffic state estimation and prediction

In this study, the research on traffic state estimation and prediction was divided

into three parts. There are many studies to estimate or predict traffic flow variables such as the traffic flow, speed, and density. The results are useful information for both the user and the operator of the freeway, but specific values of those cannot clearly identify the traffic state. Therefore, some research directly estimated or predicted traffic state (e.g., free-flow state or congested state) instead of using traffic flow variables. However, there are a few studies addressing the stochastic process of traffic state transition and predicting the traffic state on the term of probability. The probabilistic approach was mainly attempted in the studies on breakdown estimation or prediction because the stochastic process is particularly prominent in the phenomenon of flow breakdown. The breakdown estimation or prediction studies mainly focused on a capacity and considered simply a freeway mainline or ramps. In addition, among the flow breakdowns, only the spontaneous breakdown was considered in most cases, so the model performance was guaranteed only in the limited situation. Therefore, the usability in the field where various situations occur can be reduced. In order to apply traffic state prediction model in actual field, a probabilistic model is needed to cover various traffic conditions.

The studies on traffic state estimation and prediction were classified as a parametric approach or a nonparametric approach. The parametric approach models have generally well-established theoretical backgrounds and have been validated by transportation researchers. In parametric approach models, we know the functional relationship between the explanatory variable and the dependent variable in the model. Also, we can interpret the coefficients

of variables and provide insights on a problem. On the other hand, the parametric approach cannot explain all the various situations due to the relatively simple formation with limited assumptions. It is also difficult to implement large-scale networks or large datasets. The nonparametric approach models can effectively deal with relatively large-scale networks or large datasets. It can also handle problems with complex and nonlinear properties, and its performance has been validated in many studies. In particular, the nonparametric approach is necessary for transportation problems with complex and nonlinear characteristics. However, black-box procedures are the main demerits of the nonparametric models. Also, since most of the parameters in the model are no physical meaning about the targeted problem, it is difficult to obtain insights of the traffic mechanisms through the interpretation. In addition, excellent performance can be guaranteed when the integrity of data storage and sufficient size of data are available. Therefore, we need a traffic state prediction model that can be analyzed and provide insights on the traffic mechanisms like as a parametric approach and is with good performance like as a nonparametric approach.

Finally, fixed loop detector data used in most studies have limitations. In most studies, the spatial range of traffic state estimation and prediction is set to fixed loop detectors. This limited range cannot represent a congested traffic pattern that varies spatiotemporally and is not suitable for a freeway operator needs. The spatial range of the traffic state estimation and prediction should be extended beyond the loop detector point to account for the spatiotemporal

variation of the congested traffic pattern. Also, fixed loop detector data can be incomplete due to measurement error or missing value. In most methods, applying incomplete data results in either malfunction or biased outcome. However, a researcher using actual field data requires a model that can give robust results even in the incomplete data.

Table 2.2 Literature review of estimation and prediction of traffic variables

	Method	**Road type**	**Variables and Agg. Level**	**Equipment**	**Prediction horizon**
Ahmed and Cook (1979)	ARIMA (Box-Jenkins)	Three freeways	Flow (20s, 30s, 60s) Occupancy (20s, 30s, 60s)	VDS	20s, 30s, 60s
Okutani and Stephanedes (1984)	Kalman filtering	One highway	flow (15min)	VDS	5min
Davis and Nihan (1991)	K-nearest neighbor	A freeway section with three on-ramps	Freeway flow (1min) and occ (1min) On-ramp flow (1min)	VDS	2min
Smith and Demetsky (1994)	Neural Network Nearest Neighbor	Two freeways	Freeway flow (1min)	VDS	15min
Hamed et al. (1995)	ARIMA (Box-Jenkins)	Five arterials	Flow (1min)	-	1min
Florio and Mussone (1996)	Neural Network	Twenty stations in a freeway	Freeway flow, speed, density (30-120s) Weather conditions, visibility, accident heavy vehicle percentage, VMS message	VDS	1-2min

Williams et al. (1998)	ARIMA	A freeway	Flow (15min)	VDS	15min
Sun et al. (2003)	Kalman filtering	A freeway section with two on-ramp and two off-ramp	Freeway flow, On- and off- ramp flow	VDS	-
Vlahogianni et al. (2005)	Neural Network	An urban signalized arterial	Flow (3min)	VDS	3, 6, 9, 12, 15min
Wang and Papageorgiou (2005)	Kalman filtering	A freeway with one on-ramp and one off-ramp	Freeway flow, speed, density (10s) Ramp flow (10s)	Simulation data	-
Lam et al. (2006)	ARIMA Neural Network Nearest Neighbor	A urban road	Flow (1hour)	VDS	1hour
Sun et al. (2006)	Bayesian network	An urban signalized arterial	flow (15min) Upstream and downstream flow (15min)	VDS	15min
Zheng et al. (2006)	Neural Network	An expressway corridor	flow (15min) Upstream and downstream Flow (15min)	VDS	15min

Chandra and Al-Deek (2009)	Vector Autoregressive model (VAR)	A freeway five stations	Flow and speed (5min) Upstream and downstream flow and speed (5min)	VDS	5min
Min and Wynter (2011)	ARIMA	Ten highway links	Flow and speed (5min) Upstream and downstream flow and speed (5min)	VDS	5, 10, 15, 30, 45, 60min
Wang and Shi (2013)	Support vector regression	A freeway	Flow, speed, occupancy (5min)	VDS	5min

Table 2.3 Literature review of traffic state estimation/prediction

	Method	**Road type**	**Variables and Agg. Level**	**Equipment**	**Prediction horizon**	**Traffic state**
Florio and Mussone (1996)	Neural Network	20 stations in a freeway	Freeway flow, speed, density (30-120s) Weather conditions, Visibility, Accident Heavy vehicle percentage, VMS message	VDS	1-2min	3 states (stable, critical, and unstable)
Antoniou et al. (2007)	Markov chain	A freeway	Flow, speed, and occupancy (2min)	VDS	-	3 and 5 states
Xia et al. (2007)	Clustering–based	A freeway	Flow, speed, and occupancy (5min)	VDS	-	5 states
Deng et al. (2009)	Support vector machine	A freeway	Flow, speed, and occupancy (5min)	VDS (loop and video data)	-	4 states

Li et al. (2010)	Logistic Regression	A highway	Speed, time, weather, pavement materials, and slope	VDS	-	3 states (free, crowd, and block)
Sun et al. (2012)	Support vector machine	An expressway	Flow, speed, and occupancy (5min) Upstream and downstream flow, speed, and occupancy (5min)	VDS (loop and video data)	-	2 states (free flow and congested flow)
Xia et al. (2012)	Clustering–based	A freeway	Flow, speed, and occupancy (5min)	VDS	-	5 states
Antoniou et al. (2013)	Markov chain	A freeway	Flow, speed, and occupancy (5min)	VDS	5min	3, 4, 5, 8 states
Noroozi and Hellinga (2014)	Markov chain	An expressway	Flow, speed, and occupancy (5min)	VDS	5min	4 states (two free flow, a transitional, and a congested)

Table 2.4 Literature review of prediction of flow breakdown probability

	Method	**Road type**	**Variables and Agg. Level**	**Equipment**	**Traffic state**
Elefteriadou et al. (1995)	Empirical analysis	A freeway with a ramp (junction)	Mainline flow Ramp flow (1min)	Camera data	2 states (free flow and congested flow)
Persaud et al. (1998)	Empirical analysis	A freeway with a ramp (junction)	Mainline flow (median-lane) (1, 3, 4, 10, 15min)	VDS	2 states (free flow and congested flow)
Lorenz and Elefteriadou (2001)	Empirical analysis	A freeway with a ramp (junction)	Mainline flow (20sec, 1, 2, 5, 15 min)	VDS	2 states (free flow and congested flow)
Persaud et al. (2001)	Empirical analysis (logistic distribution)	A freeway with a ramp (junction)	Mainline flow (1min)	VDS	2 states (free flow and congested flow)
Evans et al. (2001)	Markov Model (Zonal merging probabilities)	A freeway with a ramp (junction)	Mainline flow (2sec)	Camera data	2 states (good state / bad state)
Son et al. (2004)	Car-following based	A freeway with a ramp (junction)	Time Gap	-	2 states (free flow and congested flow)

Brilon et al. (2005)	Survival analysis (Weibull)	A freeway with lane reduction	Mainline flow (5min)	VDS	2 states (free flow and congested flow)
Dong and Mahmassani (2009)	Markov Model (Weibull)	A freeway	Mainline flow (5min)	VDS	2 states (free flow and congested flow)
Kim et al. (2010)	Survival analysis (Weibull, gausiaan, logistic)	Freeway 5 location	Mainline flow (5min)	VDS	2 states (free flow and congested flow)
Wang et al. (2010)	Markov Model	Freeway 100 location	Mainline speed, flow, density (5min)	VDS (loop and video data)	Depending on the density
Armstrong (2011)	Bayesian network	A freeway with ramp (junction)	Mainline flow, Ramp flow (40sec)	VDS	2 states (free flow and congested flow)
Elefteriadou et al. (2011)	Survival analysis (Product limit method)	A freeway with ramp (junction)	Mainline flow, Ramp flow (1min)	VDS	2 states (free flow and congested flow)
Shiomi et al. (2011)	Platoon-based traffic flow model	A freeway	Time headway, Mainline flow Mainline speed, # of platoon (5min)	Camera data	2 states (free flow and congested flow)

Kühne and Lüdtke (2013)	Platoon-based traffic flow model	A freeway	Time headway, Mainline flow Mainline speed, # of platoon (1min)	VDS	2 states (free flow and congested flow)
Xu et al. (2013)	Car-following based	A freeway with ramp (junction)	Driver's acceleration and deceleration wave	-	2 states (free flow and congested flow)
Noroozi and Hellinga (2014)	Markov Model	A freeway	Mainline speed, occupancy (5min)	VDS	4 phase (two free flow, one transitional, one congested

2.3 Bayesian Network

2.3.1 Introduction to Bayesian network

In this section, we briefly explain the specification of Bayesian network following Jenson and Nielsen's (2009) convention of notation. Bayesian network, also called Bayesian belief net or brief network, can be classified as a probabilistic graphical model. Bayesian network is well established as practical representations of knowledge for reasoning under uncertainty (Druzdel and Van Der Gaag, 2000). Also, this model can be used to understand complex and stochastic problems and predict probabilistic events.

A Bayesian network consists of following (Jenson and Nielsen, 2009):

- There is a set of variables and a set of directed edges between variables.
- Each variable has a finite set of mutually exclusive states.
- The variables together with the directed edges form an acyclic directed graph (DAG).
- To each variable A with parents $B_1, \dots, B_n$ there is attached a conditional probability table $P(A|B_1, \dots, B_N)$.

The variables are presented by the nodes, and the directed edges are presented by the links of the DAG. The variables can be discrete or continuous. The direction of the link can be determined by cause-effect relationships, but it is not a strict requirement for modeling the structure of the Bayesian network. The conditional probability tables (or distributions) are considered as parameters of

Bayesian network. To calculate the parameters, we can define a Bayesian network over a universe of variables $U = \{A_1, \dots, A_n\}$. Then the Bayesian network specifies a unique joint probability distribution $P(U)$ given by the product of all conditional probability tables specified in the Bayesian network (Jenson and Nielsen, 2009):

$$P(U) = \prod_{i=1}^{n} P(A_i | pa(A_i))$$

Where $P(U)$: Joint probability distribution

A_i: set of variables indexed by i

$pa(A_i)$: set of parent variables of variable A_i

Now, if the joint probability distribution $P(U)$ is known and new multiple findings $e = \{e_1, \dots, e_m\}$ are provided, we can calculate joint probability distribution given evidence e:

$$P(U, e) = \prod_{i=1}^{n} P(A_i | pa(A_i)) \prod_{j=1}^{m} e_j$$

By marginalizing $P(U, e)$, the probability of any variable A within U can be calculated as follows:

$$P(A|e) = \frac{\sum_{U/A} P(U, e)}{P(e)} = \frac{\sum_{U/A} P(U, e)}{\sum_{A} P(A, e)}$$

Using this type of two-way evidential reasoning, all the probabilistic queries can be calculated (Zhang and Taylor, 2006).

2.3.2 Bayesian networks in transportation

Bayesian networks have not been widely used for transportation problems. A few studies are as follows: traffic state prediction (Sun et al., 2006; Armstrong, 2011), traffic incident detection (Zhang and Taylor, 2006), and crash prediction (Hossain and Muromachi, 2012; Lin et al., 2015; Sun and Sun, 2015).

Sun et al. (2006) developed a new approach based on the Bayesian network to predict traffic flow. In the study, Bayesian network models included traffic flows of adjacent road links as cause nodes. The joint probability distribution was modeled as a Gaussian mixture model (GGM). They pointed out that the Bayesian network can include adjacent road links to analysis the trends of the current link and predict the traffic flow in case of incomplete data.

Zhang and Taylor (2006) proposed incident detection algorithm based Bayesian network to capture general expert knowledge and to perform consistent reasoning. The results show that the proposed algorithm has comparable performance to neural network algorithm which was developed and tested in previous studies. Also, they found that the proposed algorithm is not sensitive to the incident decision threshold and has transferability.

Armstrong (2011) applied Bayesian network to predict the traffic state. The model includes the mainline flow and ramp flow in the target area. The probability is calculated deterministically based on logit formulation. It is a simple and limited probabilistic model.

Hossain and Muromachi (2012) applied Bayesian network to develop

the real-time crash prediction model. The predictors in the model were identified using random multinomial logit model. In the evaluation, the results showed that the accuracy of the model was 66% with a false alarm rate of less than 20%.

Lin et al. (2015) built Bayesian network models to predict accident risk for comparison. They discretized all continuous variables and modeled Bayesian network with a simple structure.

Sun and Sun (2015) applied dynamic Bayesian network with time series data of speed and different state combinations to predict the likelihood of crashes. In comparative analysis between dynamic and static Bayesian networks, they concluded that the dynamic Bayesian network model is more suitable to predict the likelihood of crashes.

In the studies, the problems with stochastic process in transportation mechanism were modeled by taking advantages of the Bayesian network. There has been an attempt to model the network structure in accordance with the characteristics of the transportation problem such as the dynamic Bayesian network (Sun and Sun, 2015). In most studies, however, a simple cause-effect relationship between symptom variables and problem variables is only considered to construct a structure of the Bayesian network. This structures can guarantee model performance, but there is a limit to interpret model parameters and provide insights on transportation problems through Bayesian network reasoning. In previous studies, therefore, Bayesian networks were developed and evaluated, but there was almost no discussion through model analysis.

2.4 Originality of This Research

This study has the following originalities that are different from previous research.

- **Traffic state prediction model with good performance and interpretability**

The existing statistical methods (i.e., parametric approach) built an analytical model based on a well-established background and obtained insights for a given problem, but it was difficult to obtain good performance for dealing with complex and highly nonlinear data. On the other hand, the machine learning based methods (i.e., nonparametric approach) were suitable for utilizing nonlinear data and produced generic, accurate and convenient models. However, the results which were derived from the black-box process were not interpretable. Therefore, even if high performance is obtained, there is a limit to increasing understanding of the problem.

In this study, we used the Bayesian network to construct an interpretable model with good performance. The Bayesian network is a probabilistic model based on machine learning with high accuracy in predicting the phenomenon with uncertainty. In addition, we can identify factors affecting probability variation of problem variables by analyzing the Bayesian network based on a graphical representation, which provides valuable insights concerning the traffic state prediction. The analyses are important contributions

of this study to find insights on the traffic state transition and the congested traffic pattern.

- **Traffic state prediction model including both spontaneous and induced breakdown**

The existing breakdown probability models considered only spontaneous breakdown. In other words, those were limited models that considered only the situation where the flow breakdown occurred due to the excess of capacity by the upstream demand. In this study, we developed a probabilistic model that reflects not only the upstream demand but also the complex and simultaneous situation of the traffic flow. Therefore, it is possible to increase the applicability in the field and to analyze various congested traffic patterns.

- **Traffic state prediction model with spatiotemporally extended variables**

Sun et al. (2006) claimed that most research on the traffic state prediction could not make good use of information on nearby freeways or did not even use the information to analyze the trend of the target freeway section. However, in order to perform an accurate analysis, both the information of target area and the adjacent area should be considered. Previous studies tried to reflect the effects of upstream and downstream through multivariate analysis, and some machine learning-based research using spatiotemporally extended variables has

been conducted. In this study, we used spatiotemporally extended variables to improve the accuracy of the prediction, as well as analyzing the effect of upstream, downstream, and ramp on congestion.

- **Bayesian network modeling procedure and structure for traffic state prediction**

In medicine field, Bayesian networks are widely used for studies such as examining diseases or judging the treatment of diseases according to prescriptions. The variables in medicine can be divided into diseases (problem variables) that are the target of treatment and symptoms (symptom variables) caused by the diseases. The symptom variables can be medical examinations or changes in the body, lifestyle, etc. However, these symptom variables are not directly affected by the problem variables. The symptoms can be identified by changes in the state of the blood, hormones, or muscles (background variables). Therefore, it is important to set background variables that have causal relationships with the problem variables. In order to make in-depth analysis using the Bayesian network, the mechanism for the problem should be correctly constructed as a structure.

However, in transportation problem, the Bayesian networks have used with naïve structures that had a patent node as an effect variable to all the other nodes and simple structures that considered a simple cause-effect relationship between the variables (including problem variables and symptom variables, i.e.,

mostly fixed loop data, without background variables). Strictly speaking, the values observed from detectors do not have a direct causal relationship with the problem variables (i.e., traffic state or crash potential). The value from a detector, which is a sampling of the traffic state that changes in time and space, can be compared to a medical test. Therefore, it is necessary to define the background variables of the transportation problem and construct a structure that can analyze the mechanism.

In this study, out of naïve and simple Bayesian network, we defined measurement nodes and traffic state nodes and proposed an interpretable structure of the Bayesian network. Also, the Bayesian networks structure was elicited from a combination of prior knowledge and statistical data, and continuous improvement may be possible in further studies.

- **Traffic state prediction model applicable to incomplete loop data**

Sun et al. (2006) claimed that previous studies have difficulty predicting the traffic state when the data are partially missing or unavailable. In addition, Chen et al. (2012) mentioned that missing value of the detector data also degrades the prediction performance. Some studies addressed this issue by often adopting the historical average method, but the prediction accuracy is quite limited. Also, a general well-defined method for prediction with incomplete data is not yet known. Therefore, a methodology to deal with incomplete data is needed.

In this study using the Bayesian network, robust results can be

obtained even if some information is lost. In addition, by using the advantage of the Bayesian network, we evaluated the effect of variables on the problem variable by setting various scenarios.

- **Traffic state identification considering evolution of spatiotemporal traffic congestion pattern**

The fixed loop detector data used in most studies have limitations. In most studies, the spatial range of traffic state estimation and prediction was set to fixed loop detectors. This limited range cannot represent congested traffic patterns that vary spatiotemporally, and the range is not suitable for freeway operator's needs. The spatial range of the traffic state estimation and prediction should be extended beyond the loop detector point to account for the spatiotemporal variation of the congested traffic patterns.

In this study, we set the target area and estimated the measures to represent the area for traffic state identification considering the evolution of spatiotemporal traffic congestion patterns in the area. As a result, the Bayesian network proposed in this study can be practically used reflecting the stochastic process of traffic state transition in traffic operations.

Chapter 3. Data Collection and Preparation

3.1 Data Collection and Validity Check

The study site was determined by considering the quality of the detector data, the geometry of the freeway, including the on and off ramps, and the congestion patterns. The site of the study is Interstate highway I80-W in the city of Berkeley, Alameda County, California, and the selected area had both spontaneous breakdown and congestion due to a downstream queue. The main data sources were aggregated stationary detectors on the freeway.

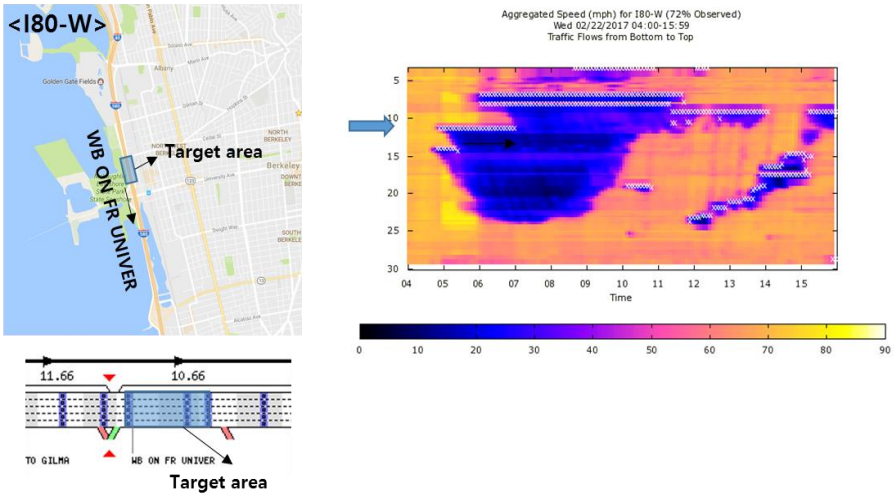


Figure 3.1 Data collection site

The data were obtained from the Caltrans Performance Measurement System (PeMS) provided by the California Department of Transportation. The

data were aggregated at 30-second intervals, and 57,600 samples were collected over a period of eight weeks, i.e., 1/2/2017 -2/24/2017, on non-holiday weekdays only. We collected data of traffic flow, speed, and occupancy in the target area as well as upstream and downstream. We also collected data on just flow on the ramps. We defined the traffic variables of each site as follows: we calculated the average flow across all lanes; speed was the flow-weighted average across all lanes; and we averaged occupancy across all lanes. We removed the data taken during adverse weather conditions. The criterion for adverse weather is precipitation of 0.2 inches or more (National Weather Service websites). Missing data and erroneous observations were also corrected. We used 70% of the data for learning and 30% for testing.

3.2 Traffic State Identification

Traffic state identification is needed to construct the dataset for supervised learning of the traffic state prediction model. However, there is no consistent method or guidelines for detector location and type of performance measures to identify the traffic state (Kondyli et al. 2013). In addition, the time lag was generated according to the distance between the fixed detector and the flow breakdown point. Therefore, it was necessary to select a measure that reflected the spatial transition of congestion in order to perform reliable traffic state identification. We investigated the appropriate methodology for traffic state identification based on the fixed detector data.

3.2.1 Traffic state identification considering spatial transition of congestion

Most previous studies were performed on a single point in determining a traffic state or whether a flow breakdown occurred or not. In this study, it is necessary to identify the traffic state in the entire target area rather than a single point in the target area. Due to the nature of the congestion, spatial transitions of congestions must be considered to determine the traffic state within the area.

When the traffic state identification is performed on a single point, a time lag may occur in detecting the congestion according to the spatial variation of the congestion. Figure 3.2 shows this phenomenon. As seen in the figure, the speed at the start point detectors of the target area is drastically decreased due to the bottleneck activation at the target area, but the speed at the end point detectors of the target area is gradually decreased. That is, due to a queue activation, the start point of the target area is affected by the congestion, but the end point is not. Therefore, the time lag may occur when the traffic state identification is performed based on the end point detector. On the other hand, when the induced congestion by queue spillback occurs at the target area, the opposite phenomenon will occur as described above. The end point detectors first encounter the queue, and a difference in time to detect congestion occurs between the end point and the start point of the target area. This phenomenon is repeated in the recovery of congestion, and the traffic state identification with a single point cannot reflect the spatial transition of the congestion.

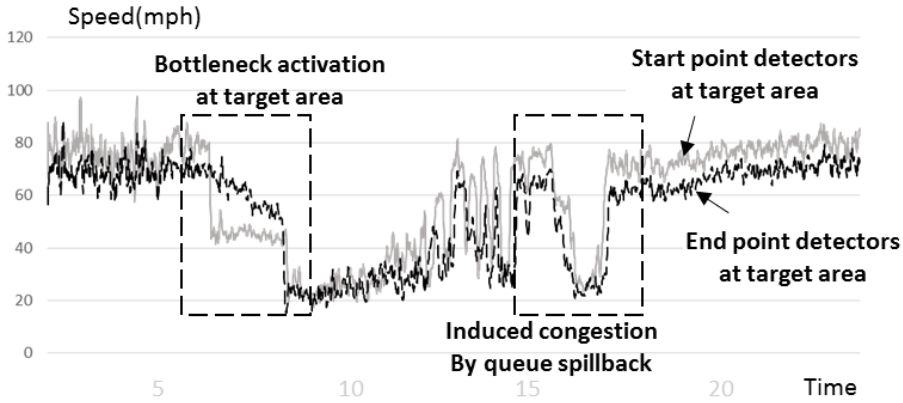


Figure 3.2 Speed difference of start point detectors and end point detectors at target area

Therefore, in this study, we proposed methodologies for the traffic state identification by calculating a measure that represents the entire target area rather than the point where the detector is installed. The methodologies are divided into three categories according to the conditions of given data.

1) Directly density observation
2) Density estimation (detection system with high accuracy)
3) Link speed estimation (detection system with low accuracy)

The best measure of the spatial characteristics of traffic flow is density. The traffic density is a fundamental macroscopic indicator and is widely used in assessing traffic performance (May, 1990). If it is possible to measure the density in the target area directly, it can be the best indicator for the traffic state identification. Without the measured density, you can estimate the density using data from the detection system with high accuracy. If the given data is measured

from the detection system with low accuracy, it is generally possible to perform the traffic state identification by estimating the link speed using the spot speed which is the most reliable among traffic variables from a detector. The details are described as follows.

In this study, the traffic state identification is performed through link speed estimation due to the low accuracy of the given data.

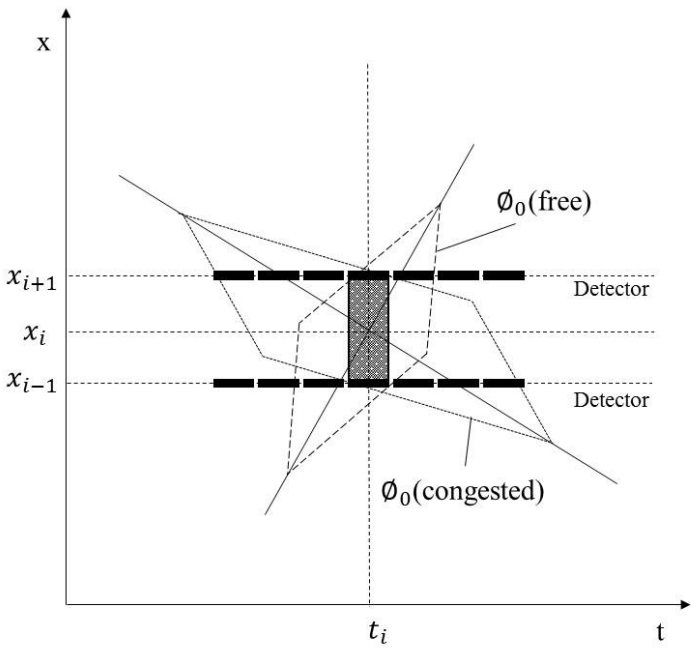
3.2.2 Link-speed-based traffic state identification

In general, the flow data collected by the detector are difficult to have high accuracy due to the limitation of the detector itself. Since the density estimated from the detection system with low accuracy can have large errors, other alternatives are needed.

Among the flow, speed, and occupancy, which are collected from the loop detectors, speed is the indicator with the smallest error from the actual value. This is, speed is a robust aggregated data compared to flow and occupancy even if there are some non-detectable vehicles. Therefore, it is possible to estimate the link speed representing the target area using the speed information, which can be relatively accurate at the detection system with low accuracy, and we performed the traffic state identification using the link speed. The link speed was calculated as follow Treiber and Kesting's (2013) work.

As shown in Figure 3.3, the link speed was estimated based on two-dimensional spatiotemporal interpolation algorithm. The basis of this algorithm

is a discrete convolution with two kernels $\emptyset_0(free)$ and $\emptyset_0(congested)$. To consider propagation velocities of macroscopic speed varying depending on traffic state (i.e., free-flow and congested traffic state), two different weighting kernels was used. After $V_{free}(x,t)$ and $V_{cong}(x,t)$ were estimated using the kernels, the final link speed $V(x,t)$ was estimated according to the weight $w(x,t)$.



Source: Treiber et al. (2013)

Figure 3.3 Concept of traffic-adaptive averaging method

As noted above, the traffic-adaptive averaging method estimates the link speeds in free-flow traffic and in congested traffic to reflect the traffic flow pattern in the spatiotemporal diagram. In the free-flow traffic, perturbations propagate

downstream along the traffic flow. This speed is also slightly lower than the local speed of the vehicles. On the other hand, in the congested traffic, perturbations propagate in the opposite direction of the traffic flow. $V_{free}(x,t)$ and $V_{cong}(x,t)$ were calculated as follow formulas given location x and time t.

$$V_{free}(t) = \frac{1}{N_{free}(x,t)} \sum_i \emptyset_0(x - x_i, t - t_i - \frac{x - x_i}{c_{free}}) v_i$$

$$V_{cong}(t) = \frac{1}{N_{cong}(x,t)} \sum_i \emptyset_0(x - x_i, t - t_i - \frac{x - x_i}{c_{cong}}) v_i$$

Where $V_{free}(t)$*:* smoothed speed in free flow traffic

$V_{cong}(t)$*:* smoothed speed in congested traffic

$\emptyset_0(x,t) = exp[-(\frac{|x|}{\sigma} + \frac{|t|}{\tau})]$*:* weighing kernel

$N_{free}(x,t)$, $N_{cong}(x,t)$ *:* normalization constants $N(x,t) = \sum_i \emptyset_0(x,t)$

c_{free}, c_{cong}*:* propagation velocities of perturbations in free flow and congested traffic

v_i*:* measured speed from detectors.

The final link speed $V(x,t)$ is calculated as follow formula by using $V_{free}(x,t)$ and $V_{cong}(x,t)$

$$w(x,t) = \frac{1}{2}[1 + tanh(\frac{V_c - V^*}{\Delta V})$$

$$V(x,t) = w(x,t)V_{cong}(x,t) + [1 - w(x,t)]V_{free}(x,t)$$

Where V_c: threshold between free and congested traffic

$V^* = min[V_{free}, V_{cong}]$

ΔV: width of the transition between free and congested traffic.

Using the Next Generation Simulation (NGSim) data, we validated the estimated link-speed based on the traffic-adaptive averaging method.

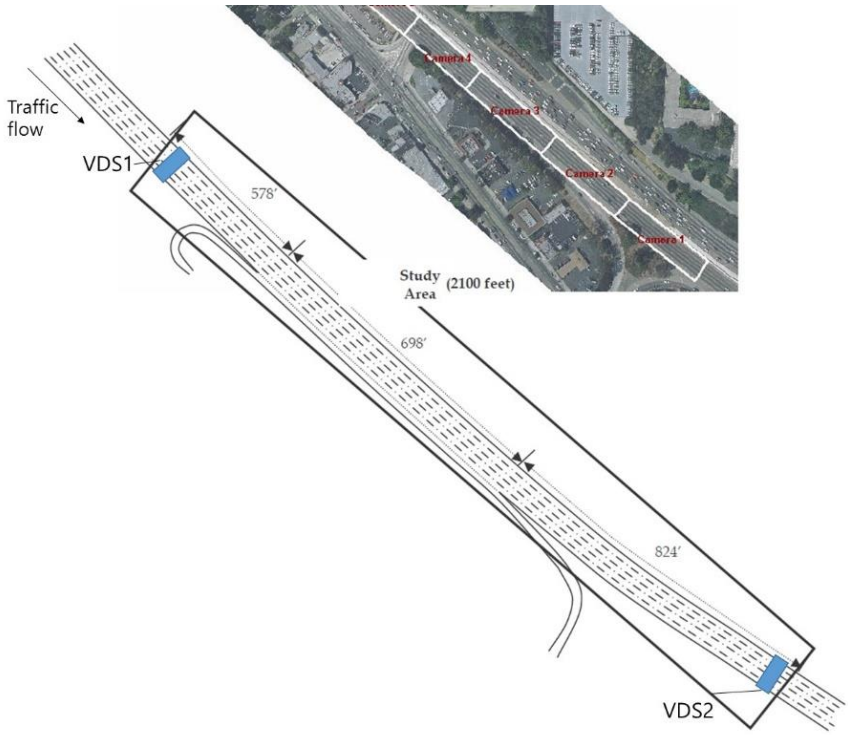


Figure 3.4 Location for the validation of traffic-adaptive averaging method

As shown in Figure 3.4, the analysis location is the southbound direction of U.S. Highway 101 (Hollywood Freeway) in Los Angeles, California. The analysis time is June 15, 2005, from 7:50 am to 8:35 am and is divided into three parts for 45 minutes (each for 15 minutes). The section of

study has a 2,100 feet long and five-lane one-way. The trajectory data in this section were extracted by the vehicle detection and tracking process.

In this study, two different speed data were extracted from NGSim trajectory data. One is the 30-second aggregated speed data from the virtual detectors at both ends of the study area (VDS1 and VDS2 in Figure 3.4), and another is the 30-second aggregated link speed data in the study area.

First, we set up the VDS1 (start point of the study area) and the VDS2 (end point of the study area) on the study area to make the speed data collected from the detectors. By using the vehicle speed only observed at the virtual detectors, the 30-second aggregated harmonic mean was calculated. The calculated data were used for the analysis through the smoothing step. Next, in order to calculate the link speed in the study area, the vehicle information (i.e., vehicle ID, location coordinates, and time) was classified in 30-second intervals. Total traveled distance and total travel time of the vehicles were calculated within each divided data, and the final link speed was calculated by dividing the total traveled distance into the total travel time.

Figure 3.5, 3.6 and 3.7 show the calculated speed according to three datasets. In this data, this section is continuously congested for 45 minutes and shows a congestion pattern of which the spatial distribution changes. In dataset 1, congestion is detected at the link start point, but the congestion is not detected at the link end point, which indicates that a queue appears to be propagating upstream due to a breakdown in the study area. In the dataset 2 and 3, the congestion is detected at the link end point, and the speed of the two detectors

is fluctuated caused by the congestion.

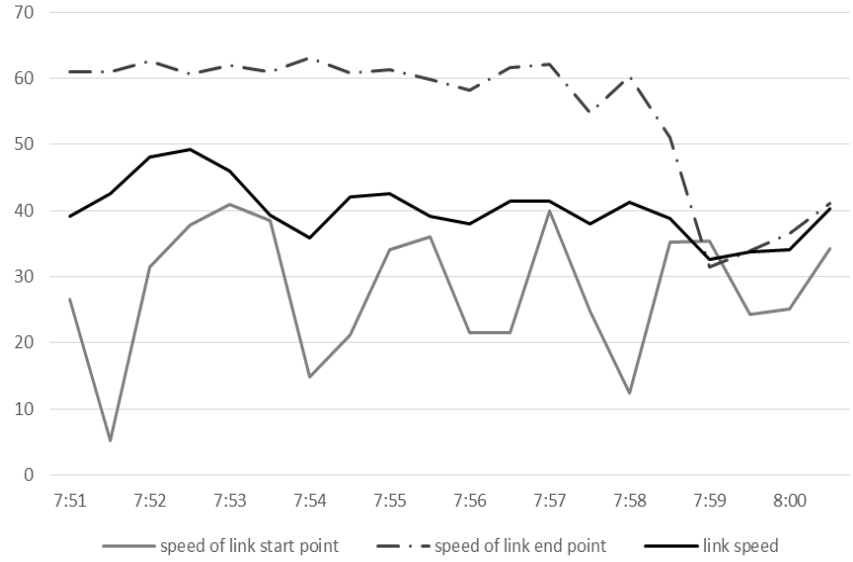


Figure 3.5 Time series of speed in the dataset 1

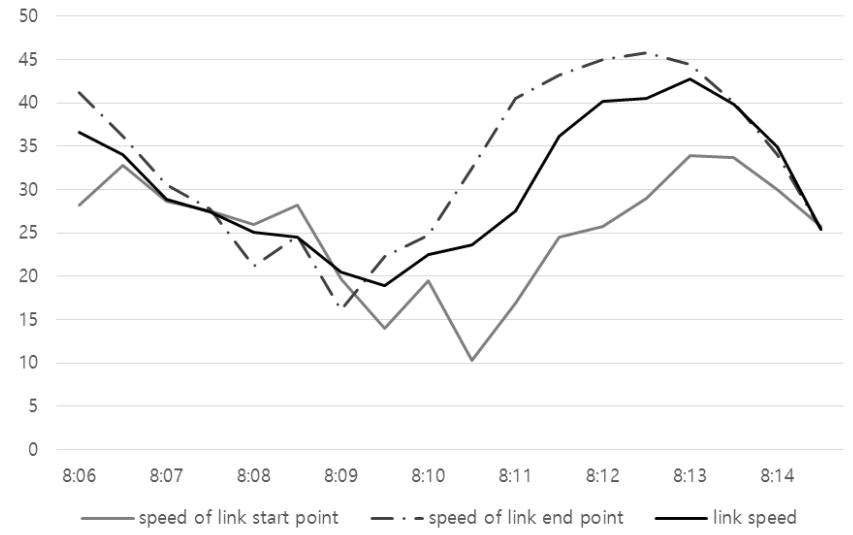


Figure 3.6 Time series of speed in the dataset 2

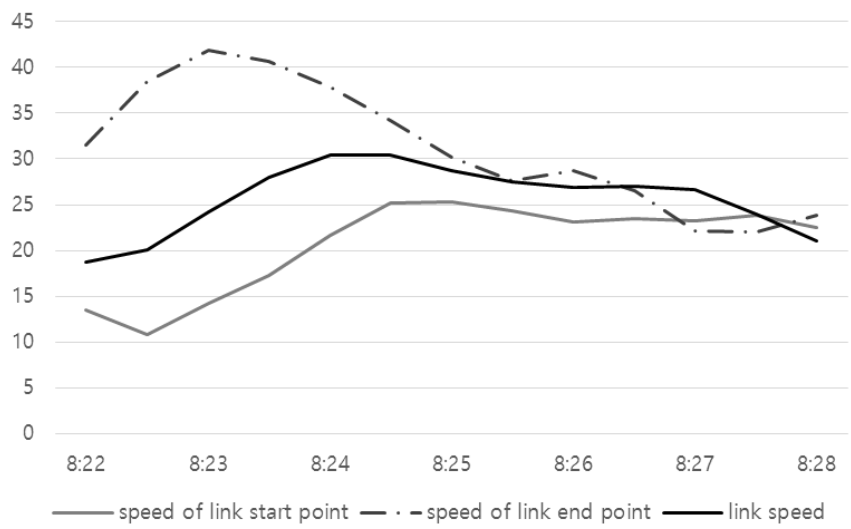


Figure 3.7 Time series of speed in dataset 3

We applied the traffic-adaptive averaging method to this datasets and evaluated the accuracy of estimated link speed. For the comparison of the results, we used simple moving averages (SMAs). The SMA is the unweighted mean, which is usually taken from an equal number of data on either side of a central value. The link speed at a time t and a location x is equally weighted mean for n detector speeds. In this study, n was set to 3, 5, and 7 periods. Also, MARE (mean absolute relative error) and MAPE (mean absolute percentage error) were used to evaluate the accuracy of the link speed.

$$MARE = \frac{1}{n}\sum_{t=1}^{n} |\hat{V}(t) - V(t)|$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|\hat{V}(t) - V(t)|}{V(t)} \times 100$$

Where $\hat{V}(t)$ denotes the estimated link speeds; $V(t)$ denotes the actual link speeds; n denotes the number of times compared.

As a result, the MARE of the traffic-adaptive averaging method was 1.78-2.88 and the MAPE was 5.3-7.5%. It is the best result compared to other methods. According to the graphs (Figure 3.8, 3.9, and 3.10), the traffic-adaptive averaging method is more sensitive to the changes in speed than the other methods and is most similar to the actual link speed. In cases of the SMA, it does not react sensitively to the speed as the period becomes longer. Therefore, the error increases in the section where the speed changes.

Table 3.1 MARE of link speed estimation method

MARE	Traffic-adaptive averaging method	SMA (3-periond)	SMA (5-periond)	SMA (7-periond)
Dataset1	2.88	3.17	3.23	3.80
Dataset2	1.78	1.90	2.46	3.28
Dataset3	1.85	2.11	2.06	2.24

Table 3.2 MAPE of link speed estimation method

MARE	Traffic-adaptive averaging method	SMA (3-periond)	SMA (5-periond)	SMA (7-periond)
Dataset1	7.3%	8.1%	8.2%	9.7%
Dataset2	5.3%	5.7%	7.4%	10.1%
Dataset3	7.5%	9.0%	8.9%	9.6%

As a result of the validation, we determined that the link speed calculated through the traffic-adaptive averaging method was highly reliable.

Using this method, the link speed of the target area was estimated and used as input data for traffic state identification.

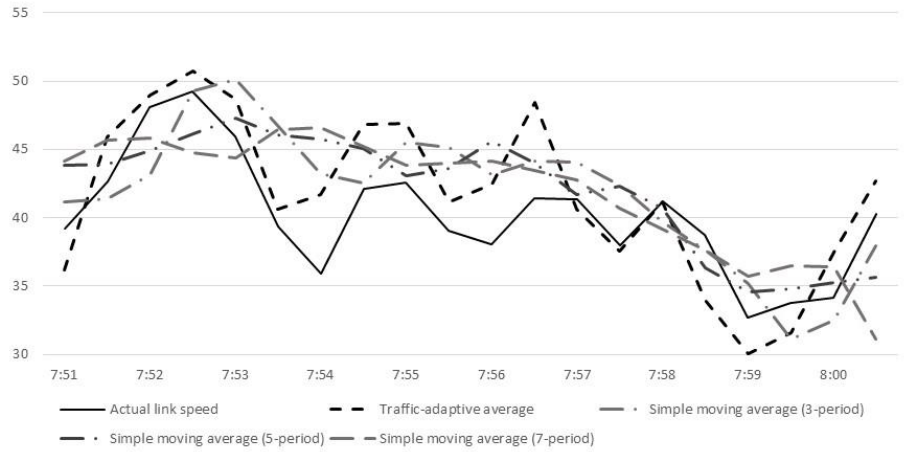


Figure 3.8 Actual and estimated link speed in dataset 1

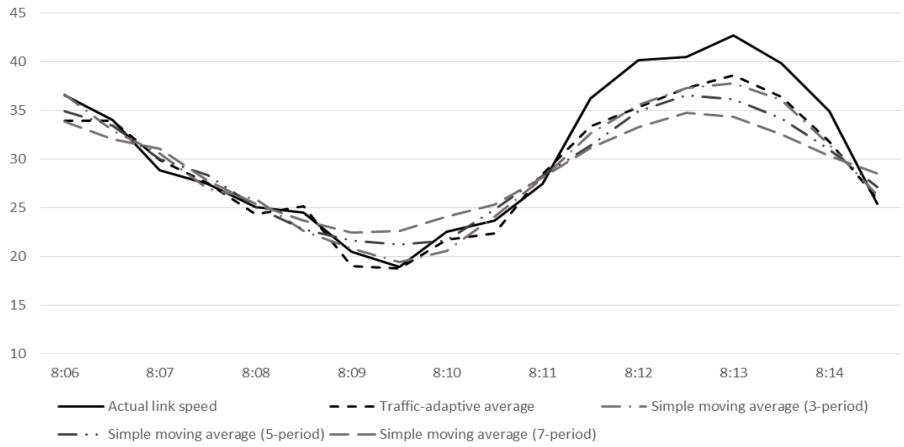


Figure 3.9 Actual and estimated link speed in dataset 2

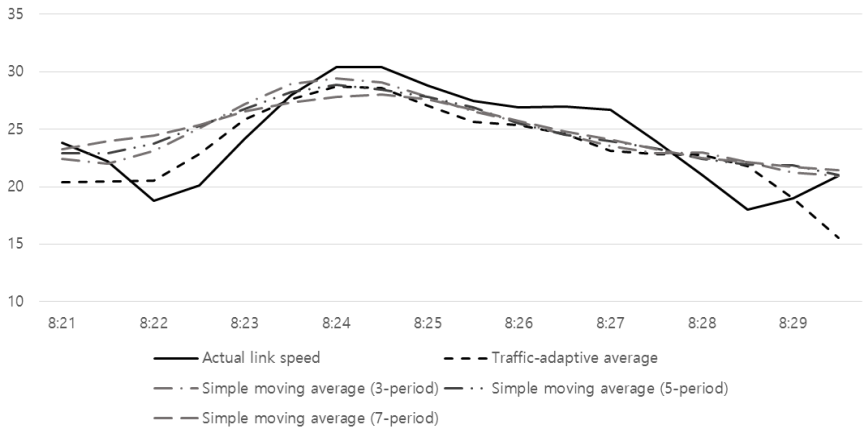


Figure 3.10 Actual and estimated link speed in dataset 3

3.2.3 Traffic state identification algorithms

Most studies have used speed as a primary performance measure to identify the traffic state. In general, speed responds instantaneously to changes in traffic state relative to other indicators (Kondyli el al., 2013). In previous research, the basic logic of the identification algorithm was to set the speed threshold at which the traffic state changes from the free-flow traffic state to the congested traffic state or vice versa (Lorenz and Elefteriadou, 2001; Brilon et al., 2005; Dong and Mahmassani, 2009; Shiomi et al. 2011). In addition, some researchers considered a time duration (Lorenz and Elefteriadou, 2001; Dong and Mahmassani, 2009; Kim et al., 2010). If the time duration of the speed below the threshold was greater than a certain time, it was determined that the traffic state changed from the free-flow state to the congested state or vice versa.

In this study, we use the same traffic state identification algorithm as

previous studies based on the link speed. The traffic state identification algorithm is divided into two steps, i.e., identification step and post-processing step. In the identification step, the traffic state is determined according to preset thresholds. After the identification step is completed, the post-processing step is performed to eliminate noise when the duration of the same traffic state is less than a certain time.

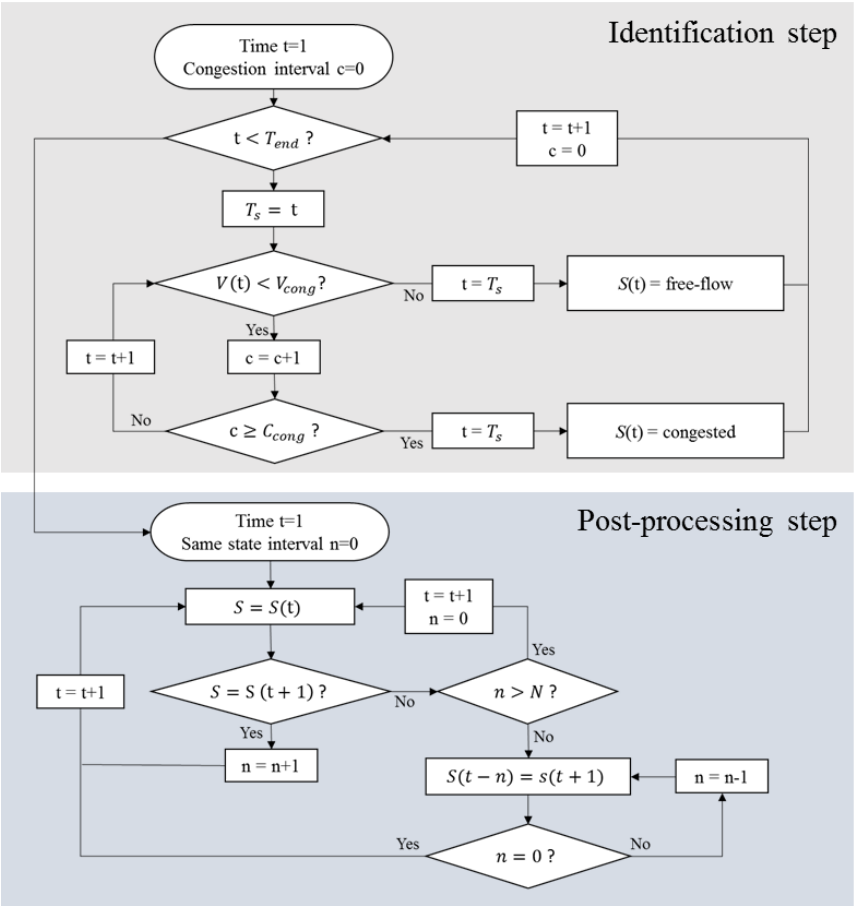


Figure 3.11 Algorithm of traffic state identification

In the identification step, if the time duration of the speed ($V(t)$) below the threshold ($\mathrm{V_{cong}}$) was greater than a certain time ($\mathrm{C_{cong}}$), it was determined that the traffic state changed from the free-flow state to the congested state or vice versa. After the identification step was performed all the time, the post-processing step was carried out. In the post-processing step, if the duration of which the traffic state remains the same was less than a certain time step N, it was determined as noise and then removed. In this algorithm, $\mathrm{V_{cong}}$, $\mathrm{C_{cong}}$, and N must be determined in advance. In this study, fundamental diagrams were used to determine the thresholds of the algorithm (Figure 3.12 and 3.13). The thresholds were set as follows: $V_{cong} =$ 80 km/h, $C_{cong} = 10\ time\ steps$ (5 minutes), $N = 10\ time\ steps$ (5 minutes)

Figure 3.14 shows the results of the traffic state identification. On the given day, the traffic state was continuously changing with congested traffic state or free-flow traffic state, and the algorithm identifies a total of 6 congested traffic state intervals. The results indicated that the traffic state transitions due to instantaneous speed changes (i.e., perturbation) were reduced and the recovery time could be detected in advance.

In this study, the traffic state identification is performed by using this algorithm, and this derived data are used in Bayesian network modeling.

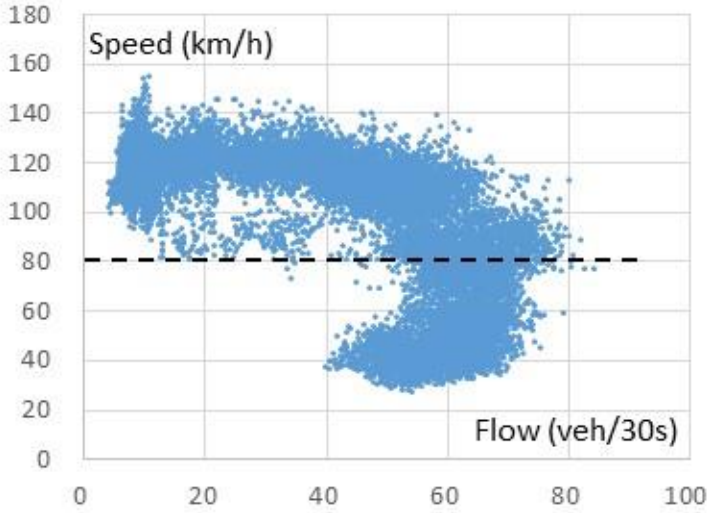


Figure 3.12 Selection of V_{cong} using the flow-speed diagram

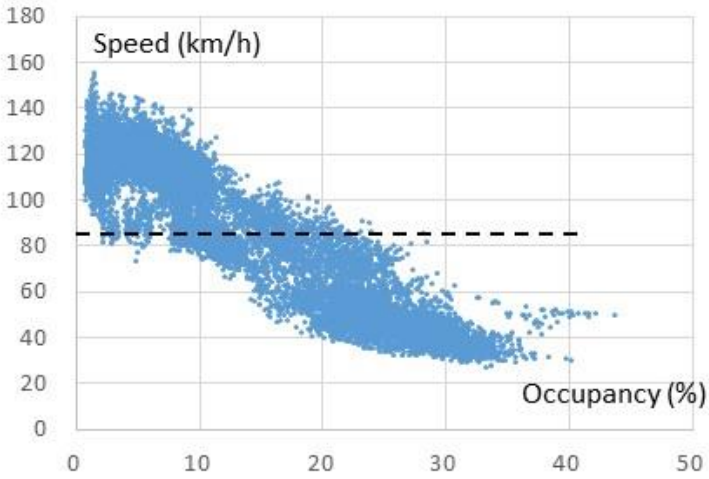


Figure 3.13 Selection of V_{cong} using the occupancy-speed diagram

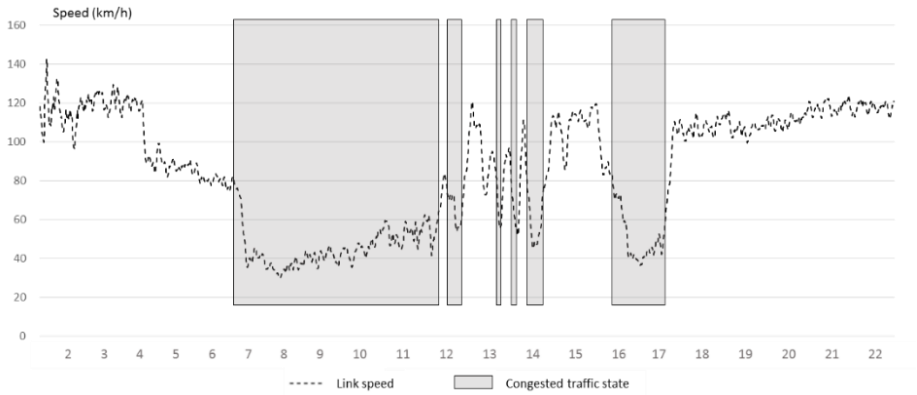


Figure 3.14 Results of traffic state identification

3.3 Data Description

In this section, the data that were collected and prepared are described to model Bayesian network for traffic state prediction. As shown in Figure 3.15, the data consisted of 12 variables, i.e., one prediction variable, which is the result of traffic state identification in the target area, and 11 observed variables in each area.

Prediction variable

$s_{target}(t+10)$ = traffic state in target area during interval $t+10$

Observed variables

$q_{target}(t)$ = traffic flow in the target area during interval t

$\mu_{target}(t)$ = speed in the target area during interval t

$o_{target}(t)$ = occupancy in the target area during interval t

$q_{up}(t)$ = traffic flow at upstream detector station during interval t

$\mu_{up}(t)$ = speed at upstream detector station during interval t

$o_{up}(t)$ = occupancy at upstream detector station during interval t

$q_{down}(t)$ = traffic flow at downstream detector station during interval t

$\mu_{down}(t)$ = speed at downstream detector station during interval t

$o_{down}(t)$ = occupancy at downstream detector station during interval t

$q_{onramp}(t)$ = traffic flow at on-ramp detector station during interval t

$q_{offramp}(t)$ = traffic flow at off-ramp detector station during interval t

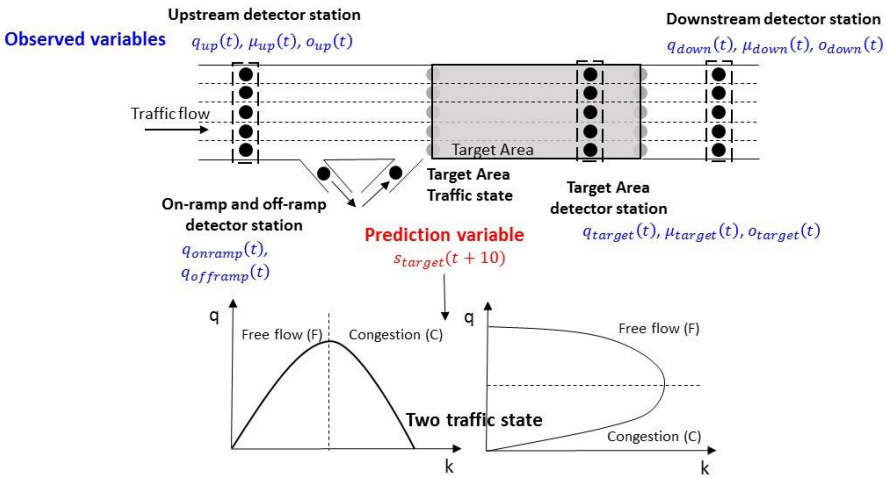


Figure 3.15 Data description in this study

In this study, the aggregate interval time of the data was 30 seconds, and we predicted the traffic state after 5 minutes. Also, the prediction variable, as the target of this study, was assumed to be a binary traffic state, i.e., a free flow state or a congested state. We performed Bayesian network modeling

using this dataset.

Before the modeling part, we clarified the meaning of the traffic state derived by the traffic state identification and the model prediction. In the traffic state identification, the duration condition is applied to reduce the effect of perturbations on the traffic state. Therefore, the congested traffic state is an outcome including information that the congested traffic state has been maintained for a predetermined duration. Likewise, the predicted traffic state by the supervised learning model with the data also implies this duration. That is, we can predict the probability including that a congested traffic state will be maintained for a predetermined duration at a time $t + \alpha$ using information at time t in the model.

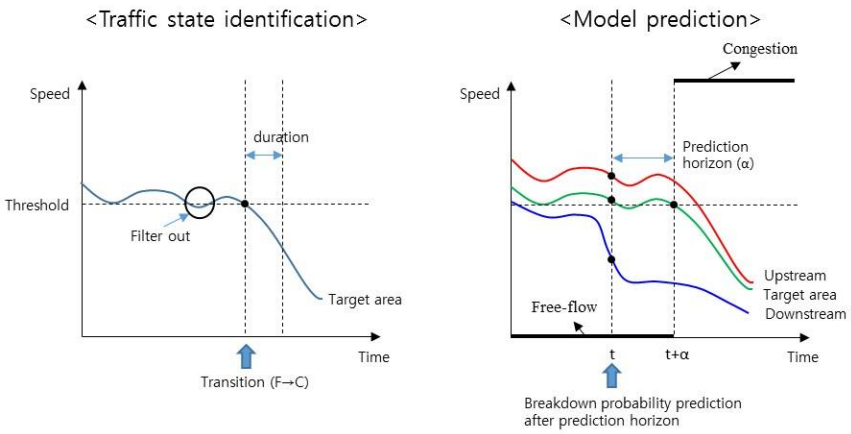


Figure 3.16 Concepts of traffic state identification and model prediction

Chapter 4. Bayesian Network Modeling

In this chapter, we proposed the Bayesian network modeling procedure and elicited the structure of the Bayesian network for the traffic state prediction. To construct the structure based on expert knowledge, we used an object-oriented modeling approach (Kjærulff and Madsen, 2008). First, the modeling procedure was briefly described, and then the interface containing each module and the design of the modules was described. Finally, we performed the verification of the structure and parameter learning of the elicited structure.

4.1 Modeling Procedure

The Bayesian network consists of two main components. One component is a structure that consists of nodes that represent variables and directed links that represent the relationships between the variables. The other component is a parameter, the conditional probability distribution of each node for a given structure. Two components are called qualitative and quantitative components. The structure (qualitative component) is expressed as a graphical language, and the parameter (quantitative component) is expressed as a numerical language. Therefore, the modeling procedure consists of constructing the structure and learning the parameters. First, we constructed a structure and then estimated the conditional probability distribution of each node through parameter learning.

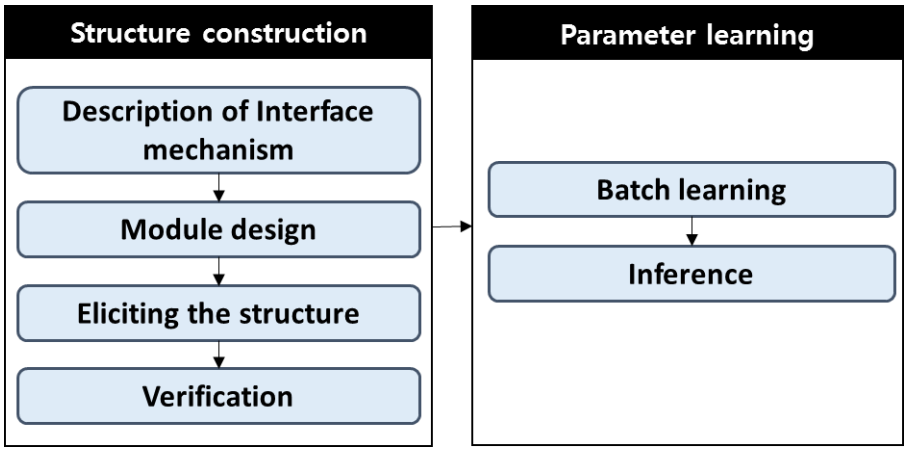


Figure 4.1 Bayesian network modeling procedure

The ways to build a structure of the Bayesian network are divided into an expert knowledge-based method and data-driven method. The manual construction of the structure based on expert knowledge requires close communication with expert groups. On the other hand, the data-driven method based on machine learning requires the complete data. If you have flawless data that include all cases, the utility of the data-driven method can be maximized. However, if the data-driven method is used based on incomplete data that does not include all cases, a distorted structure can be obtained. In this study, therefore, the expert knowledge-based method was applied to construct the structure as a qualitative component.

To manually construct the structure, we adopted an object-oriented modeling approach (Kjærulff and Madsen, 2008). This approach can be used when the structure of Bayesian network includes copies of almost-identical

network fragments. In this study, there are almost-identical network fragments in different locations, because the data are collected in the same way (i.e., stationary loop detectors). Therefore, instead of repeatedly constructing same network fragment, a generic network fragment was constructed once and it was instantiated to other objects. In this approach, a generic network fragment is called a module.

In addition, this approach has a top-down process for the model construction. We first constructed the interface of modules (abstraction level) without specifying their internal details. The interface is constructed by considering the spatiotemporal relationships between modules based on the information for prediction of traffic state. After setting the interface of modules, module design was conducted. To design the module, we used Mixture of Gaussian (MoGs) distribution because the modules include both the continuous variables (loop detector measured data) and discrete (traffic states) variables. The traffic states for MoGs modeling were classified by K-mean cluster analysis. The entire structure was obtained by using the repetitive modules.

After eliciting the structure of the Bayesian network, we conducted verification analysis based on the dependency and independency relation of nodes. And then, parameter learning was performed to estimate the conditional probability distribution of the Bayesian network. Finally, the completed Bayesian network which included the structure and parameters enabled probabilistic inference.

4.2 Description of Interface Mechanism

The interface of modules was designed based on the spatiotemporal relationships between the modules. The variables in the modules can be divided into problem variables and information variables, and we added the traffic state at time t of each area as a mediating variable.

4.2.1 Definition of modules in the interface

In this study, the Bayesian network for traffic state prediction has almost-identical modules. This is because the data in each area are observed from the fixed loop detectors in the same way. The data observed from the fixed loop detector are samples of the installed area (spatially-constrained) and cannot represent the entirety. In addition, the observed data of the detector are only the measured symptom variables and cannot logically have a causal influence to other variables (e.g., problem variables). Considering the actual traffic flow patterns, the future traffic state in the target area is causally influenced by the current traffic states of upstream, downstream as well as the target area. The current traffic states are described by the loop detector data in each area. Based on this logical relationship, we defined each module according to the spatial division. The modules are divided into five areas, i.e., target area, upstream, downstream, on-ramp, and off-ramp. The traffic state of each module is put into the structure as mediating variables and acts as background information.

4.2.2 Spatiotemporal relationships between modules

In order to design the interface of the modules, we defined the spatiotemporal relationships based on information for the traffic state prediction. Figure 4.2 shows that the meaning of each module's information for the traffic state prediction, i.e., how each module affects the future traffic state. The upstream can give information about the future demand. If the demand of upstream exceeds the capacity of the target area, the probability of the congested traffic state in the future increased. Also, the information of the target area (e.g., densities of vehicles) and ramp (e.g., the number of vehicles entering) can be directly related to the potential to the flow breakdown. The downstream can give information about the influence of the queue due to downstream bottlenecks and the perturbation. The target area information also means the current traffic state before the transition of the future traffic state.

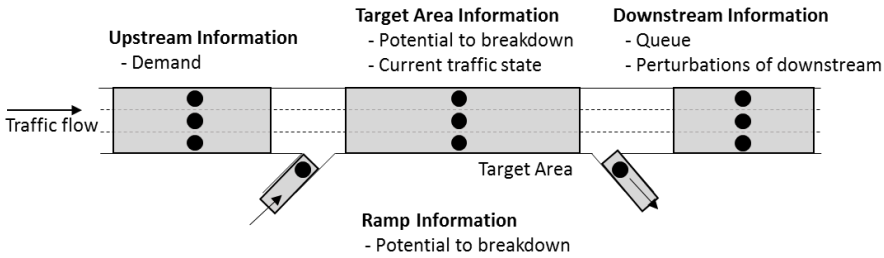


Figure 4.2 Information for traffic state prediction

As shown in Figure 4.3, the diagram of the spatiotemporal relationships between the modules was drawn to determine the directions of links between the modules. The dependencies among the current traffic states (in modules) at time t may be low, because they are spatially-exclusive and at the same point in time. Therefore, for simplicity of the model, the modules were assumed to be independent of each other. These modules at time t had a direct effect on $t + 10$ traffic state (cause-effect relationship).

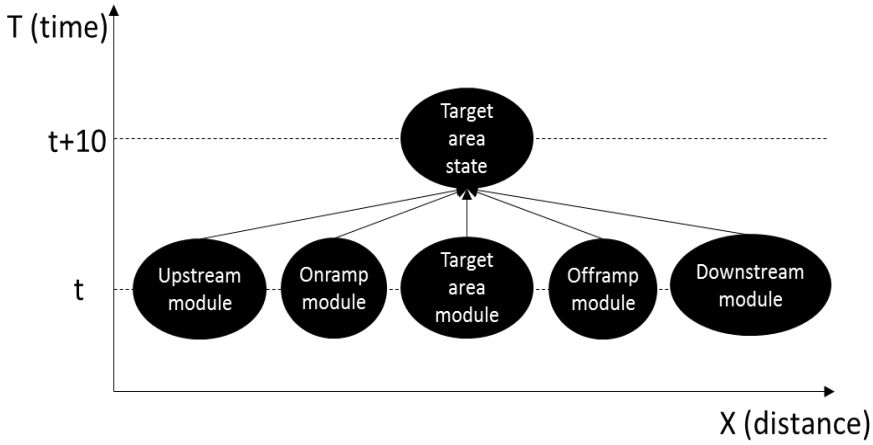


Figure 4.3 Spatiotemporal relationships between the modules

Based on the spatiotemporal relationships between the modules, the directions of links between modules were determined. All modules at time t had causal impact on target area state at $t + 10$ (future traffic state) and were mutually independent.

4.2.3 Type of variables

In this section, the variables of the Bayesian network are categorized according to the characteristics. The variables include the observed traffic variables listed in chapter 3 and the mediating variables (current traffic states). The variables are divided into problem variable and information variables, and the information variables are divided into background information variables and symptom information variables as follows:

- Problem variable: $s_{target}(t+10)$
- Information variables
  - Background information: $s_{up}(t)$, $s_{down}(t)$, $s_{target}(t)$, $s_{onramp}(t)$, $s_{offramp}(t)$
  - Symptom information: $q_{up}(t)$, $\mu_{up}(t)$, $o_{up}(t)$, $q_{down}(t)$, $\mu_{down}(t)$, $o_{down}(t)$, $q_{target}(t)$, $\mu_{target}(t)$, $o_{target}(t)$, $q_{onramp}(t)$, $q_{offramp}(t)$

The prediction variables of this study, $s_{target}(t+10)$, is set as a problem variable, and the current traffic states and traffic variables from detectors are set as background information variables and symptom information variables, respectively.

In the variables, the detector data are continuous and the traffic states are discrete. The characteristics of the variables are summarized as follows:

- Discrete variables: $s_{target}(t+1)$, $s_{up}(t)$, $s_{down}(t)$, $s_{target}(t)$, $s_{onramp}(t)$, $s_{offramp}(t)$
- Continuous variables: $q_{up}(t)$, $\mu_{up}(t)$, $o_{up}(t)$, $q_{down}(t)$, $\mu_{down}(t)$, $o_{down}(t)$, $q_{target}(t)$, $\mu_{target}(t)$, $o_{target}(t)$, $q_{onramp}(t)$, $q_{offramp}(t)$

The different variable types have typical causal dependency relations. As shown in Table 4.1, background variables (the current traffic states) give the causal impact on problem variables and symptom variables as root variables. The future traffic state in the target area (problem variable) is causally influenced by the current traffic states, and the observed traffic variables (symptom variables) are influenced by the current traffic state because they partially describe the traffic state.

Table 4.1 Typical causal dependency relations for different variable classes

Type	**Causally influenced by**
Background variables	None
Problem variables	Background variables
Symptom variables	Background and problem variables

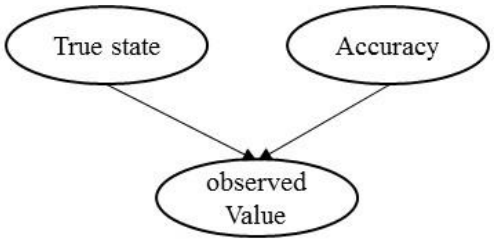
Source: Kjærulff and Madsen (2008)

The interface of modules was configured as above, and we specified their internal details based on the interface.

4.3 Module Design

Each module consists of the node of the current traffic state and the values observed from detectors. As mentioned earlier, the detector values may be samples or test results that represent the traffic state of a defined area, and cannot have direct causality to other variables. Therefore, we set the current traffic states (mediating variables) as background variables.

Neil et al. (2000) developed five idioms, semantics of commonly occurring substructures, to construct the structure of the Bayesian network. In the case of measurements among the idioms, a true state, an observed value, and an accuracy are included in the substructure shown in Figure 4.4. In the idiom, arrow direction is from the traffic state to the observed value (values from a detector) and the accuracy information is excluded in this study shown in Figure 4.5. This substructure was applied to all modules except for the future traffic state.



Source: Kjærulff and Madsen (2008)

Figure 4.4 Idiom for measurement

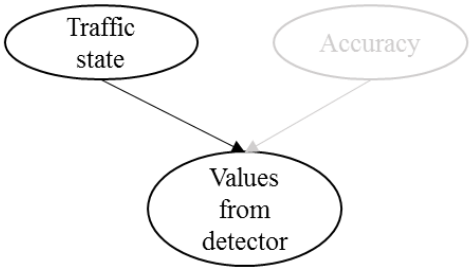


Figure 4.5 Substructure in each modules

Based on the measurement idiom, we modeled the relationship between a traffic state and values from a detector. The values observed from a detector are continuous data such as flow, speed, and occupancy, and the traffic state is unobserved and discrete data. Due to the heterogeneity of data, a module design method is needed to model continuous variables with discrete variables as a parent node. One option is to discretize the continuous variables like as the parent node. It may be a good alternative considering network computational efficiency. However, the performance of the Bayesian network may vary depending on the discretized result and it may be difficult to perform rigorous analysis using the model. In other cases, we can use a Mixture of Gaussians (MoGs) distribution to approximate the continuous distribution of the variable. It is well known that any probability distribution can be approximated by the MoGs (Kjærulff and Madsen, 2008).

An MoGs is a sum of the finite number of Gaussian distributions with weights of components. Let's assume X is a continuous variable and $S = \{s_1, \dots, s_n\}$ is a parent node of X. the MoGs of X is as follows:

$$f(x) = \sum_{i=1}^{n} p_i N(m_i, \sigma_i^2)$$

Where $m_i, \sigma_i^2 \in R$ and $0 \leq p_i \leq 1$ such that $\sum_i p_i = 1$ are the mean, variance, weight of the i^{th}component in the mixture. (Kjærulff and Madsen, 2008). By this formula, the continuous variable X is represented with a conditional linear Gaussian distribution. As shown in Figure 4.6, values of the detector (traffic flow, speed, and occupancy) are represented with the traffic state as a parent node using the MoGs.

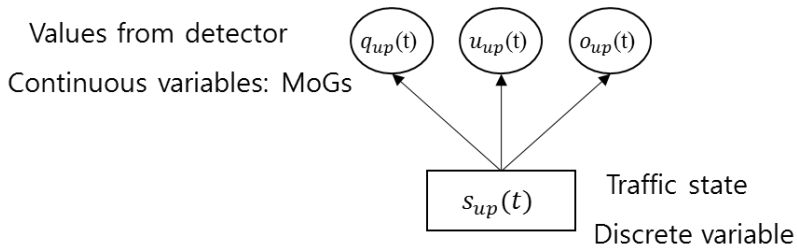


Figure 4.6 MoGs in each module

To approximate the values from detectors using MoGs, we should classify the traffic state s_i and obtain the mean m_i and variance σ_i^2. The traffic state classification was performed by K-mean clustering with the traffic flow, speed, and occupancy. K-mean clustering is an unsupervised and nonhierarchical clustering method and one of the most popular clustering algorithms due to its simplicity. The K-mean clustering divides the n observations into k clusters of which the number is an input parameter. The

objective function of the K-mean clustering is the minimization of the within groups sum of squared errors (WGSS). The formulation of the K-mean clustering is as the follows (Selim and Ismail, 1984; Huang, 1998):

$$minimize\ f(W,Z) = \sum_{j=1}^{k}\sum_{i=1}^{n} w_{ij} D\left(x_i, z_j\right)$$

$$subject\ to \sum_{j=1}^{k} w_{ij} = 1\,,\ \ 1 \le i \le n$$

$$w_{ij} \in \{0,1\}, \quad 1 \le i \le n, 1 \le j \le k$$

Where W is an $n \times k$ partition matrix, $Z = \{z_1, \ldots., z_k\}$ is a set of the centers of clusters, and $D(\cdot,\cdot)$ is the squared Euclidean distance between two objects.

As mentioned earlier, the K-mean clustering requires the number of clusters to be entered in advance. In the structure modeling part, several alternatives were evaluated and an optimal alternative was selected, unlike the future traffic state with the two phases. This means that we focused on the model performance rather than the physical meaning of the traffic state. We gave physical meaning to the results derived by using the K-mean clustering. The alternatives of the number of traffic states were 2, 3, 4, 5, and 6, and the overall accuracy and computation time were evaluated for each case. The optimal number of traffic states was selected based on the evaluation results.

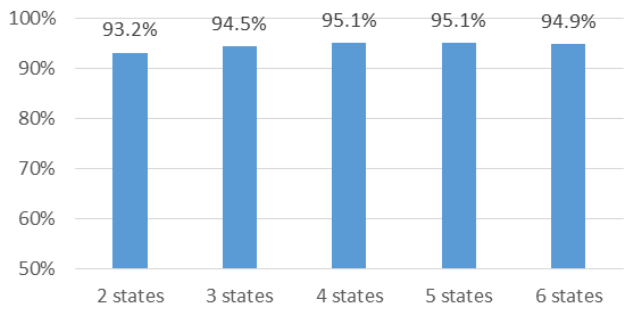


Figure 4.7 Overall accuracy of the BN depending on the number of traffic states

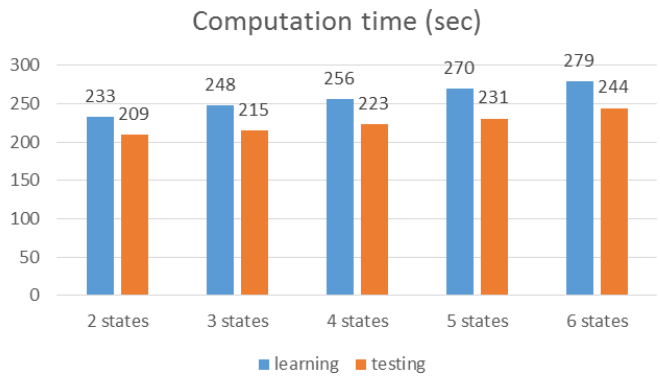


Figure 4.8 Computation time of the BN depending on the number of traffic states

As shown in Figure 4.7, the change of overall accuracy depending on the number of traffic states was not significant. Among them, when the number of traffic states was 4 or 5, the highest accuracy (95.1%) was achieved. In the case of the computation time, as shown in Figure 4.8, learning and testing time became longer as the number of traffic states increased. Based on the results,

four traffic states with the highest overall accuracy and small computation time were selected. The fundamental diagrams of the four traffic states are shown in Figure 4.9. As shown in Figure 4.9, the traffic states were classified into one free flow (state 1), two transitional (state 2 and state 3), and one congested (state 4).

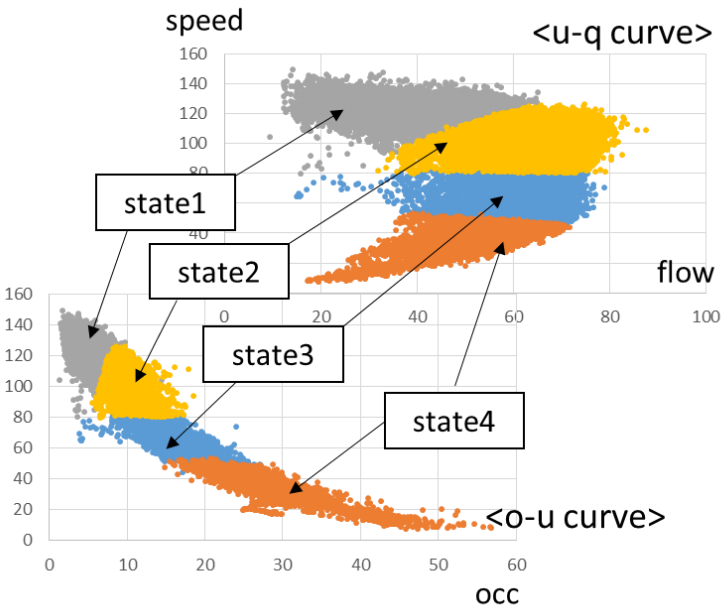


Figure 4.9 Fundamental diagrams of the four traffic states

The traffic states with four states, as a parent node, were set through the classification process. To construct the MoGs of the traffic variables nodes (traffic flow, speed, and occupancy), the means and variances of them corresponding to each state of the parent node were needed. The means and variances were calculated from the traffic states classified by the K-mean

clustering. Figure 4.10 shows the distributions of the flow and speed depending on the traffic states and Gaussian distributions based on the calculated means and variances depending on the traffic states of the target area. In the case of the traffic flow, the distributions were relatively similar to the Gaussian distribution, while the distributions of speed were skewed to the left or right. However, we judged that the MoGs were able to provide usable performance since most values were concentrated in the middle and the values at both ends were small.

The modules of the target area, upstream, downstream, on-ramp, and off-ramp were all designed using the MoGs. These designed modules were arranged according to the preset interface to obtain the final structure of the Bayesian network.

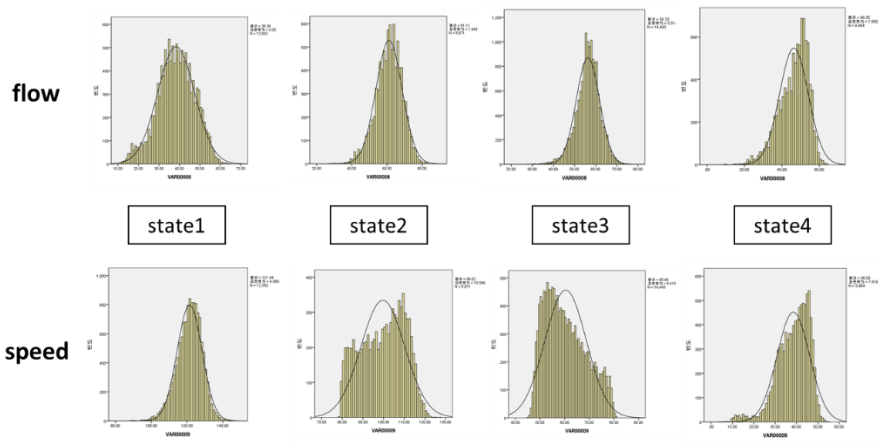


Figure 4.10 Traffic flow and speed distribution in the target area

4.4 Eliciting the Structure

By performing the design of the interface and modules, we elicited the structure of the Bayesian network as shown in Figure 4.11. Also, Table 4.2 shows all variables in the structure. In the future, it may include additional modules such as upstream, downstream, or ramp detector stations, and may include temporal information such as historical data $t-1, t-2$ ....

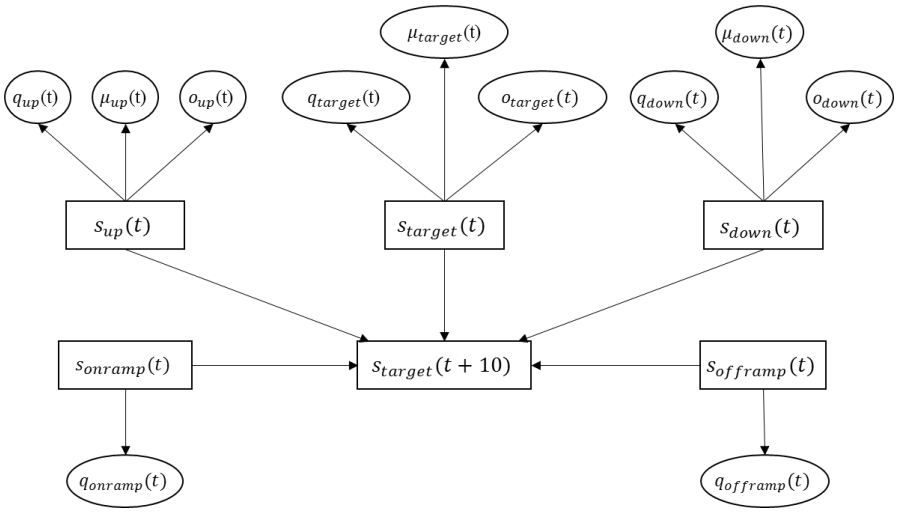


Figure 4.11 Structure of the Bayesian network

4.5 Verification

Once the structure of the Bayesian network has been elicited, model verification process should be performed. Verifying the structure which is a qualitative component of the Bayesian network requires logical judgments by collaborating between a problem-domain expert and a knowledge engineer

(Kjærulff and Madsen, 2008). A method to perform the verification is an inspection of the dependency and independency relation among variables (d-separation).

Table 4.2 Description of variables in the Bayesian network

		Variables			
		Target area	**Upstream**	**Downstream**	**Ramp**
Problem variables		$s_{target}(t+10)$			
Information variables	**Background information**	$s_{target}(t)$	$s_{up}(t)$	$s_{down}(t)$	$s_{onramp}(t)$ $s_{offramp}(t)$
	Symptom information	$q_{target}(t)$ $\mu_{target}(t)$ $o_{target}(t)$	$q_{up}(t)$ $\mu_{up}(t)$ $o_{up}(t)$	$q_{down}(t)$ $\mu_{down}(t)$ $o_{down}(t)$	$q_{onramp}(t)$ $q_{offramp}(t)$

In this study, the structure consists of almost-identical modules by using the object-oriented modeling approach. Also, the variables can be briefly divided into the traffic variables observed from loop detectors, the current traffic states of different areas, and the future traffic state of the target area. Therefore, with three check points, the inspection of the dependency and independency relation among variables can be completed. Also, model verification can be used to modify reversed links or to introduce additional (mediating) variables.

- Check points

① Relation among observed variables from a detector ($q_{\mathrm{up}}(t)$, $u_{up}(t)$, and $o_{up}(\mathrm{t})$) and the current traffic state ($s_{up}(t)$) in the same module.

② Relation among the current traffic states ($s_{up}(t)$, $s_{target}(t)$, $s_{down}(t)$, $s_{onramp}(t)$, and $s_{offramp}(t)$) and the future traffic state at time $t+10$ ($s_{target}(t+10)$).

③ Relation among observed variables from a detector ($q_{\mathrm{up}}(t)$, $u_{up}(t)$, and $o_{up}(\mathrm{t})$) and the future traffic state at time $t+10$ ($s_{target}(t+10)$).

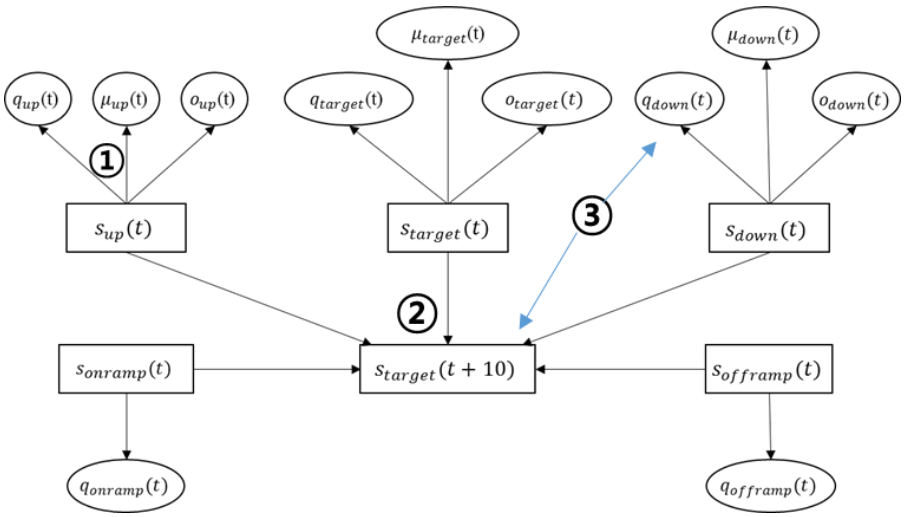


Figure 4.12 Verification of the BN structure

In the check point ①, according to the conditional independence, the observed variables are dependent when the current traffic state is unknown: Observing low speed will increase our belief that the occupancy will be high. In the other hands, the observed variables are independent when the current

traffic state is known: if we know that the current traffic state is congested, then observing low speed will not change our belief about the value of occupancy which is directly determined by the current traffic state.

In check point ②, the current traffic states are independent when the future traffic state is unknown: Even if we know the current traffic state of upstream, no information about current traffic state of downstream is available. This is because the spatial pattern of the traffic flow cannot be known. In the other hands, the current traffic states are dependent when the future traffic state is known: if we know that the future traffic state is congested, then observing the current traffic state of the target area is free-flow makes it possible to infer that upstream or downstream traffic states are likely to be congested. This is because it can make the target area congested after 5 minutes.

In check point ③, the future traffic state and the observed variables are dependent when the current traffic state is unknown: Observing low speed will increase our belief that the future traffic state is congested. It is a mechanism in which the probability of future traffic state is inferred by observed variables without the current traffic states in actual application. In the other hands, the future traffic state and the observed variables are independent when the current traffic state is known: if we know that current traffic state of target area is congested, then observing low speed of target area will not change our belief about the future traffic state which is directly determined by the current traffic state.

Model verification was performed by reviewing the three check points.

As a result, we could not find a logical error of dependency and independency relation among variables in this structure. Therefore, parameter learning was performed using the given structure.

4.6 Parameter Learning

After eliciting the structure of a Bayesian network, we can now estimate the parameters of the model (conditional probability distributions) using an available database.

There are two approaches to the parameter learning, i.e., batch learning and on-line learning. The number of updates of two approach is different for the same dataset. The batch learning is used to estimate the parameters once and for all, while on-line learning does update each data. Also, the batch learning is slightly more efficient in computation time. In this study, we performed parameter learning with the batch learning approach.

We used Maximum Likelihood Estimation (MLE) to learn the parameters of the Bayesian network. MLE is the simplest approach and is only available when there is complete and sufficient data. The data used in this study was complete and had 57,600 samples (70% was used for learning and 30% was used for testing). Therefore, we selected MLE for efficient learning. For each case $d \in D$, the probability $P(d|\theta)$ is called the likelihood of θ, a set of model parameters, given d. If it is assumed that the cases in D are independent given the model, then the likelihood of θ given D is as follows

(Kjærulff and Madsen, 2008):

$$L(\theta|D) = \prod_{d \in D} P(d|\theta)$$

Where, $P(d|\theta)$ is the likelihood of θ given d, for each case $d \in D$;

$L(\theta|D)$ is the likelihood of θ given D.

Next, the parameter estimate $\hat{\theta}$ that maximizes the likelihood is estimated.

$$\hat{\theta} = arg\, max\, L(\theta|D)$$

In this study, we used the Bayes Net Toolbox (BNT) MATLAB package (Murphy, 2001) to construct the Bayesian network including a structure and parameters and to infer this model in given evidence.

Chapter 5. Model Evaluation

In this chapter, we evaluated the Bayesian network model consisting of the structure and the parameters which are learned from the training data. The model evaluation consists of three parts: evaluation results, comparison with other methodologies, and sensitivity analysis. In the evaluation results, the model performance was evaluated based on the testing data. Also, the possibility of using this model was evaluated by comparison with the parametric approach (logistic regression) and the nonparametric approach (Artificial Neural Network (ANN)). In the sensitivity analysis, we evaluated the model performance depending on the change of prediction horizon and decision threshold. Also, we provided the understanding on traffic state prediction by conducting evidence sensitivity analysis.

5.1 Evaluation Results

5.1.1 Performance measurements

For the model evaluation, we used a classification table in this study. As shown in Table 5.1, the classification table is used to evaluate the prediction results for binary data. In Table 5.1, predicted values are compared with actual values, and the correctly predicted counts and false counts are expressed as a, b, c, and d.

Table 5.1 Classification table

		Predicted		Summary
		0 (free-flow)	1 (congested)	
Actual	0 (free-flow)	a	b	a+b
	1 (congested)	c	d	c+d

- $specificity = a\ /\ (a + b)$
- $sensitivity = d\ /\ (c + d)$
- $false\ positive\ rate = b\ /\ (b + c) = 1 - specificity$
- $false\ negative\ rate = c\ /\ (c + d)\ = 1 - sensitivity$
- Overall accuracy $= (\mathrm{a} + \mathrm{d})/\ (a + b + c + d)$

In this study, five performance measurements were calculated using the classification table. The accuracy of the model was measured by its sensitivity (the ability to predict an event correctly) and specificity (the ability to predict a nonevent correctly). The sensitivity is the proportion of event responses that were predicted to be events. The specificity is the proportion of nonevent responses that were predicted to be nonevents. We also computed three other conditional probabilities, i.e., false positive rate, false negative rate, and overall accuracy. The false positive rate is the proportion of predicted event responses that were observed as nonevents. The false negative rate is the proportion of predicted nonevent responses that were observed as events. Finally, overall accuracy is a rate of corrected classification that considers both event (congested) and non-event (free-flow) in all data.

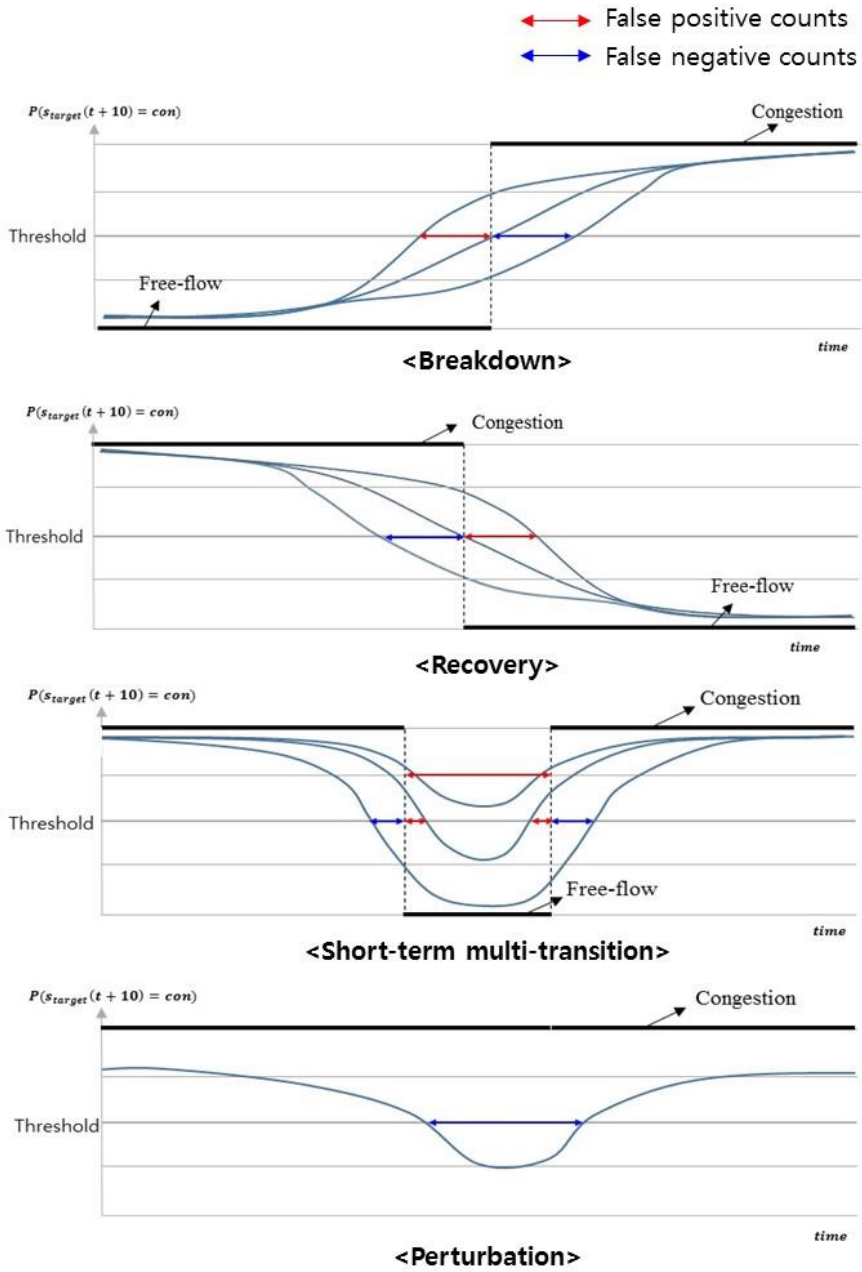


Figure 5.1 Cases where false counts occur

We also compared false counts directly to show the difference of performance between models because the rate was not clear to present the difference. The false counts were separated by false positive counts and false negative counts. The false counts were determined depending on how accurately the transition was predicted. Depending on the model, the false positive counts and the false negative counts could vary. Figure 5.1 shows cases where the false counts occurred.

5.1.2 Results of Bayesian Network

In order to evaluate the developed Bayesian network, 17,280 testing data were used. This data includes 8,712 actual free-flow states and 8,568 actual congested states. The evaluation results are shown in Table 5.2.

Table 5.2 Evaluation results of Bayesian network

BN		Predicted		Summary
		0 (free-flow)	1 (congested)	
Actual	0 (free-flow)	8,132	580	8,712
	1 (congested)	272	8,296	8,568

The performance measures were estimated using the classification table in Table 5.2.

- $Specificity = 8{,}132 \,/\, (8{,}132 + 580) \;=\; 93.3\%$
- $Sensitivity = 8{,}296 / \,(8{,}296 + 272) \;=\; 96.8\%$
- $False\ positive\ rate = 1 - \;93.3\% = 6.7\%$
- $False\ negative\ rate = 1 - 96.8\% = 3.2\%$

- Overall accuracy $= (8{,}132 + 8{,}296) / (8{,}712 + 8{,}568) = 95.1\%$

As a result, the specificity, sensitivity, and overall accuracy were 93.3%, 96.8%, and 95.1%, respectively. The false positive rate (6.7%) was greater than the false negative rate (3.2%). This indicated that there were more errors that a predicted value was a congested traffic state when an actual value was not a congested traffic state.

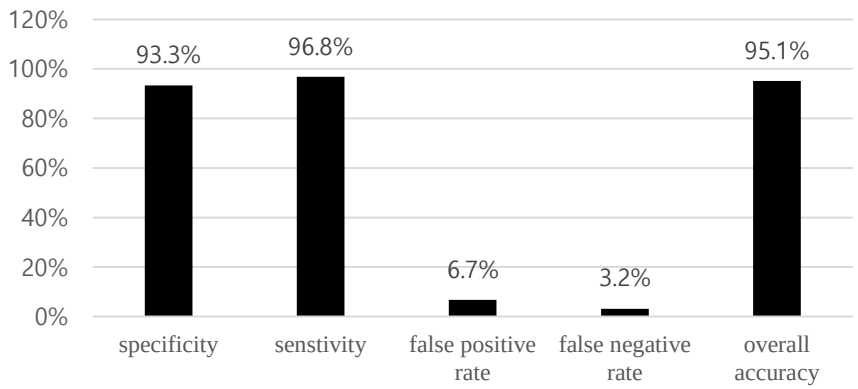


Figure 5.2 Evaluation results of the Bayesian network

By using the Bayesian network, the purpose of this study was to develop a model that provided insights on the traffic state prediction as well as had high prediction accuracy. Therefore, to evaluate the level of performance, we compared the Bayesian network with a logistic regression model, as a parametric approach model, and Artificial Neural Network (ANN), as a nonparametric approach model.

5.2 Comparison with Other Methodologies

In order to evaluate the performance of the traffic state prediction model based on the Bayesian network, a comparison was conducted with other methodologies, i.e., logistic regression and artificial neural network (ANN). The models were constructed and evaluated using the same training and testing data. In the following, we explained the models briefly and compared the model performances.

5.2.1 Parametric approach: Logistic regression

A logistic regression has been used in various fields as a classifier. The logistic regression for classification or prediction not only has well-established theoretical backgrounds but also has the advantage that it can be used more easily than other methodologies. In addition, the logistic regression is a methodology that can be used to interpret the estimated parameters and thereby improve understanding of the modeled problem.

A logistic regression is a regression model with categorical dependent variables. In this study, the binary logistic regression was used to predict the future traffic state. The predicted probability of binary dependent variable varies depending on the predictors. Unlike linear regression, logistic regression employs a logistic function as a link function between explanatory variables and dependent variable. The logistic regression has the following expression:

$$prob(y = 1) = \frac{e^{\sum_{k=1}^{K} \beta_k x_k}}{1 + e^{\sum_{k=1}^{K} \beta_k x_k}}$$

$$prob(y = 0) = 1 - prob(y = 1) = \frac{1}{1 + e^{\sum_{k=1}^{K} \beta_k x_k}}$$

Where y: binary dependent variable (predicted output, 0 or 1)

x_k: predictors $(x_1, x_2, \dots, x_k)$

β_k: coefficients of predictors $(\beta_1, \beta_2, \dots, \beta_k)$

To model the logistic regression for accurate prediction, the input data should be satisfied by several assumptions, which is similar to the assumptions made in the linear regression. The estimate of β_k can be derived by maximizing the likelihood function.

The logistic regression for the traffic state prediction included traffic flow and speed in the target area as input variables. The occupancy in the target area and the traffic variables of other areas were excluded from the model due to the multicollinearity problem. The constructed logistic regression model was as follows:

$$P\big(s_{target}(t + 10) = congested\ traffic\ state\big) = \frac{exp(\beta_0 + \beta_1 q_{target}(t) + \beta_2 \mu_{target}(t))}{1 + exp(\beta_0 + \beta_1 q_{target}(t) + \beta_2 \mu_{target}(t))}$$

Table 5.3 Estimation results of logistic regression

	$\widehat{\boldsymbol{\beta}}$	$\boldsymbol{se}(\widehat{\boldsymbol{\beta}})$	$\boldsymbol{p-value}$	$\boldsymbol{exp}(\widehat{\boldsymbol{\beta}})$
$\mu_{target}(t)$	-0.110	0.001	0.000	0.896
$q_{target}(t)$	-0.047	0.002	0.000	0.954
$Intercept$	11.863	0.184	0.000	141887.069
$-2LogL$	16833.591			

As a result, both speed and flow in target area had negative coefficients. That is, the probability, which the future traffic state would be a congested traffic state, increased as the speed and flow in the target area became smaller. Also, the coefficient of $\mu_{target}(t)$ was smaller than that of $q_{target}(t)$. The probability of congestion was more sensitive to changes in speed than flow. This was consistent with the results of the Bayesian networks (Chapter 6. Discussion).

5.2.2 Nonparametric approach: Artificial Neural Networks (ANN)

ANN, which is a very popular machine learning method, has been used extensively in various transportation problems. A significant amount of research in which ANN has been used has shown that ANN has a good predictive ability and modeling flexibility with large, multidimensional data. In this study, we used three-layer, feed-forward ANN as one of type of ANN. The input layer had 11 input (variables) neurons and one bias neuron. The hidden

layer was composed of 10 hidden neurons and one bias neuron, and the output layer had two output neurons. The input variables were values that were observed only from stationary loop detectors without current traffic state variables as mediating nodes in the Bayesian network. Also, the number of hidden neurons was determined by evaluating the performance of the model, and these neurons had no physical meaning. The output neurons were free-flow and congested traffic state as binary data.

We used MATLAB's NN toolbox to efficiently implement a three-hidden layer feedforward NN. All nodes had a tan-sigmoid transfer function, except the final output node, which was a linear transfer function. The weights and biases of the network were initialized with random values, taken from -1 to 1. We chose to use batch training, where the weights between the nodes were updated after all the training examples had been exposed to the network. Supervised training of the ANN was performed using Levenberg-Marquardt algorithm back-propagation. In this case, the errors associated with each input-output neuron pair were computed and back-propagated, and the synaptic weights were adjusted to reduce the total errors. This procedure was performed until the algorithm converges (Collazo et al., 2016).

The nonparametric approach, including ANN based on machine learning, is suitable for large data and produces a generic, accurate and convenient model. However, the results derived from the black box process are not interpretable. Therefore, even if high performance is obtained, there is a limit to providing an understanding of the problem. In contrast, the Bayesian

network not only is a good method for predicting problems with uncertainty but also provides the understanding of the problem by using the developed model. Based on this model, the direction of improvement of the model can be derived by itself.

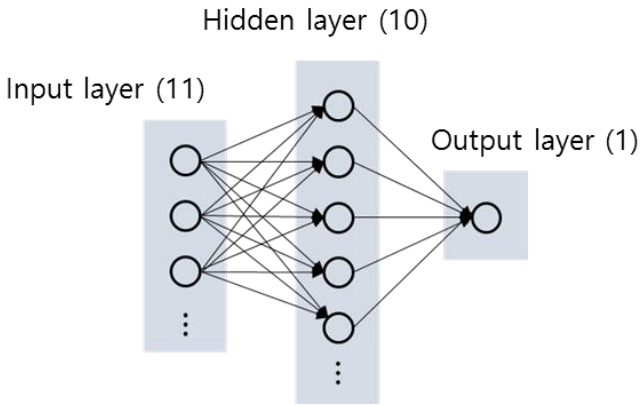


Figure 5.3 Artificial Neural Network (ANN) in this study

By comparing the models (the parametric or the nonparametric approach), we tried to show the model performance of the Bayesian network. In addition, we tried to suggest the direction of the model improvement by confirming the pros and cons compared with the existing models.

5.2.3 Results comparison

Logistic regression and ANN were constructed using the same training dataset and evaluated using the same testing dataset. The results are as follows (Table 5.4 and 5.5). As shown in Figure 5.4, the overall accuracy of the ANN was

significantly high (95.2%), but logistic regression was 85.1% which was lower than other models. The specificity, sensitivity, false positive rate, and false negative rate of ANN were 94.2%, 96.3%, 5.8%, and 3.7%, respectively. The specificity, sensitivity, false positive rate, and false negative rate of the logistic regression were 80.7%, 89.6%, 19.3%, and 10.4%, respectively. Comparing the Bayesian network and the ANN, both showed a high level of performance with over 95% overall accuracy. The Bayesian network had a higher false positive rate and a lower false negative rate than ANN. The Bayesian network constructed in this study had a similar performance to the methodology based on existing machine learning. Consequently, the Bayesian network had a higher level of improvement compared to the parametric approach model and had the same level of performance as the nonparametric approach model.

Table 5.4 Evaluation results of logistic regression

Logistic regression		Predicted		Summary
		0 (free-flow)	1 (congested)	
Actual	0 (free-flow)	7,027	1,685	8,712
	1 (congested)	895	7,673	8,568

Table 5.5 Evaluation results of Artificial Neural Network (ANN)

ANN		Predicted		Summary
		0 (free-flow)	1 (congested)	
Actual	0 (free-flow)	8,207	505	8,712
	1 (congested)	317	8,251	8,568

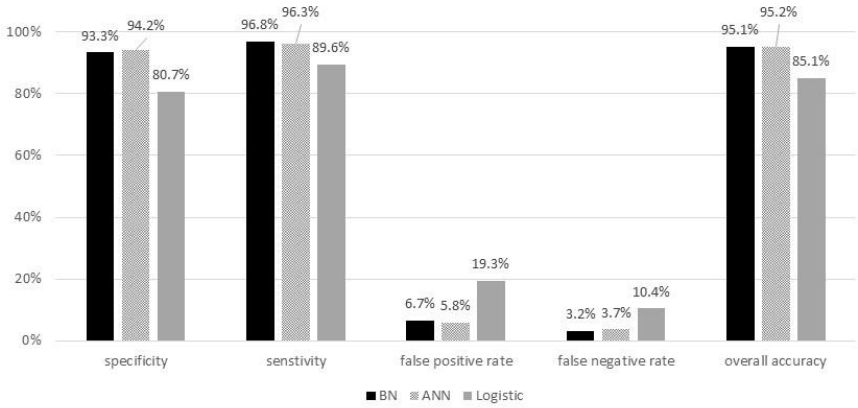


Figure 5.4 Performance comparison of BN, ANN, and logistic regression

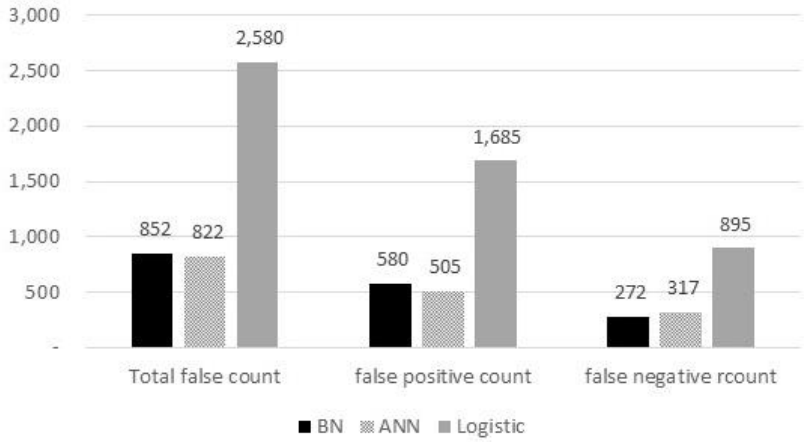


Figure 5.5 False counts of BN, ANN, and logistic regression

As shown in Figure 5.5, the false counts for each model were calculated for a more clear comparison. The results showed that Bayesian network reduced total false counts by 67.0% (2,580 → 852) compared to the logistic regression. To find specific improvements, individual cases were analyzed.

5.2.4 Individual case analysis

Individual case analysis was performed to find the main cause of the prediction errors. By comparing the probabilities of the models in various conditions, we tried to find the cases where errors were reduced due to the Bayesian network and the cases where errors occurred even though the Bayesian network was used. In addition, the results of the individual case analysis was used as information for future model improvements.

5.2.4.1 Comparison with the parametric approach model

Errors occurred mainly when the target area was near capacity of the freeway. In case of the logistic regression, the probability of the future traffic state was determined by the information in the target area. Thus, when the target area was near capacity, the probability could be sensitive to small changes in observed variables in the target area. In other words, the prediction errors might occur due to temporary fluctuation in the target area.

Figure 5.6 shows the predicted probabilities of the future traffic states from the Bayesian network and the logistic regression. Figure 5.6 (a)-(d) shows that the Bayesian network improved the prediction error in the logistic regression. As shown in Figures 5.6 (a) and (b), the Bayesian network reduced the false negative counts in the logistic regression. Also, the Bayesian network reduced the false positive counts in the logistic regression (Figures 5.6 (c) and

(d)). In the flow breakdown, the logistic regression with only target area information could cause late prediction, but the Bayesian network with information about the target area and the adjacent areas could prevent late prediction. Conversely, in the recovery, the errors caused by the logistic regression were improved by using the Bayesian network. Also, in the perturbation in the target area, the logistic regression reacted sensitively to the perturbation which leaded to the false counts, but the Bayesian network yielded stable prediction probabilities. As a result, the Bayesian network derives the probability to predict the future traffic state proactively and stably with adjacent area information.

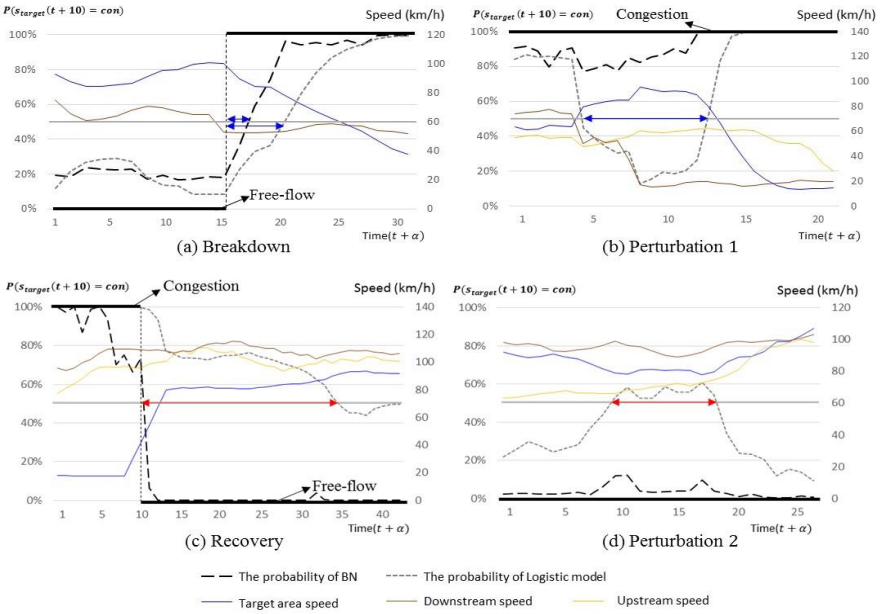


Figure 5.6 Individual case analysis: the parametric approach (1)

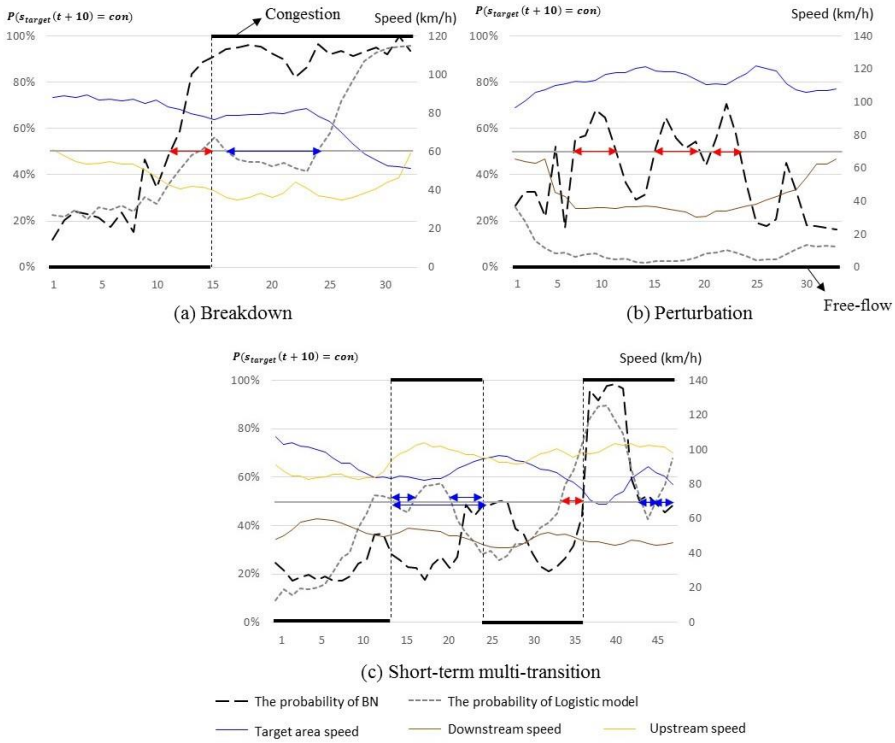


Figure 5.7 Individual case analysis: the parametric approach (2)

However, the Bayesian network still had many errors (Figure 5.7). First, in the flow breakdown, the false positive counts occurred due to the influence of the adjacent area information (especially, upstream information) (Figure 5.7 (a)). In addition, when the downstream or upstream was very congested, the effect resulted in the false positive counts (Figure 5.7 (b)). Here, we could suspect that there were correlations between the traffic states. In this study, the current traffic states in each area were assumed to be independent from each other. As shown in Figure 5.6, when the target area was in the transitional or congested traffic state, the influence of upstream or downstream

information was appropriate. However, when the target area was in the free-flow state, the adjacent information could have an excessive effect. In other words, it means that the relationships were different depending on the traffic states of the target area and the adjacent area. Therefore, it is necessary to clarify the correlations between the traffic states and to improve them. As shown in Figure 5.7 (d), the errors, when the transition occurred consecutively in the short-time interval, were still, even though the Bayesian network was used.

5.2.4.2 Comparison with the nonparametric approach model

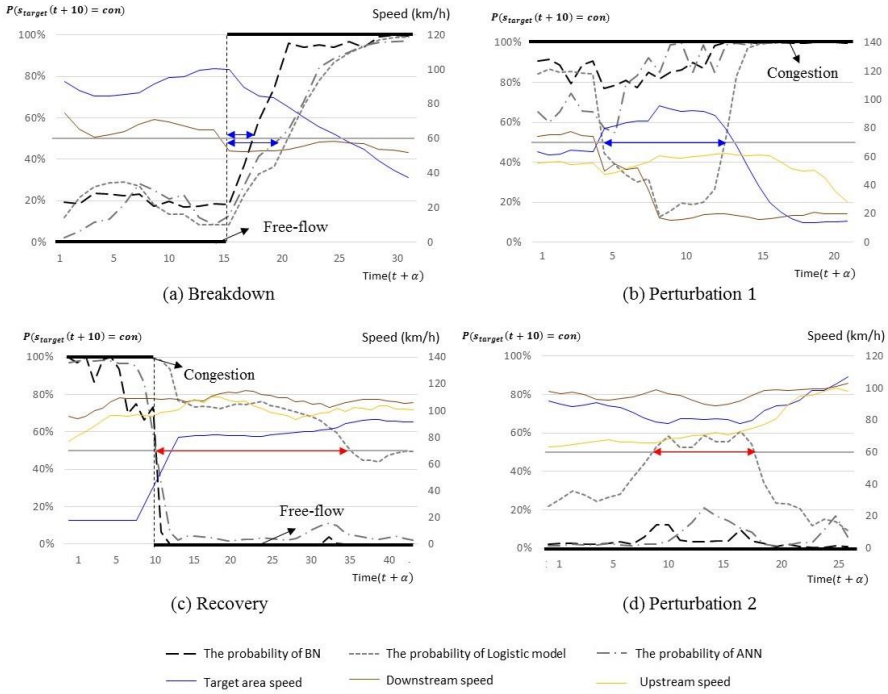


Figure 5.8 Individual case analysis: the nonparametric approach (1)

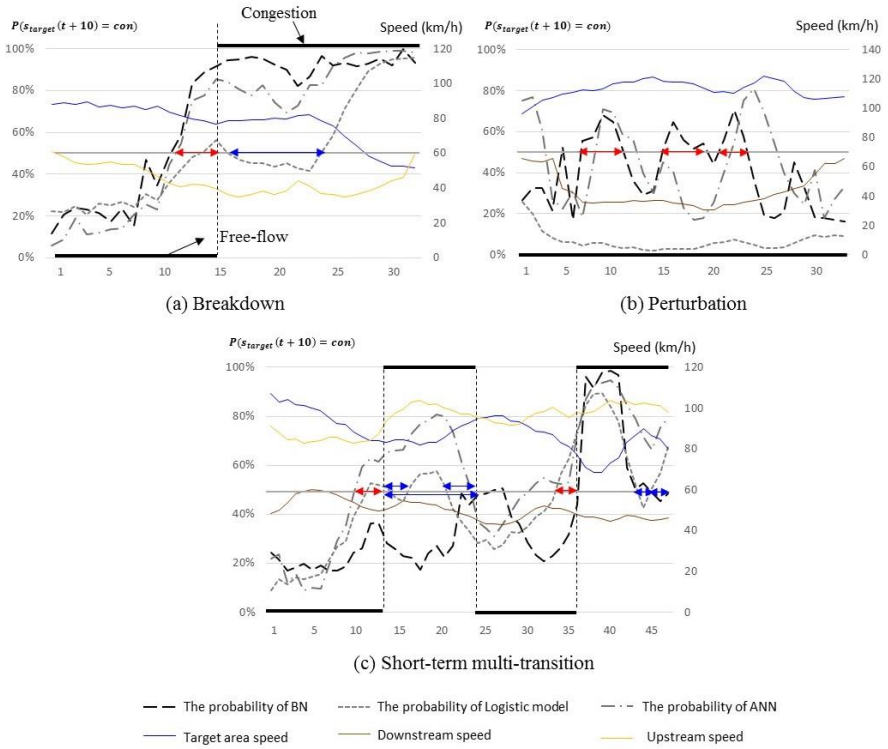


Figure 5.9 Individual case analysis: the nonparametric approach (2)

The individual case analysis was conducted with the Bayesian network and the ANN (nonparametric approach model). As shown in Figure 5.8, the probability predicted by the ANN over time was similar to the probability predicted by the Bayesian network. Although some of the results of the probabilities were different, we could not find a clear tendency to explain them. Figure 5.9, also, shows that ANN could not improve the errors yielded in the Bayesian network. It is found that the error patterns of the ANN and the Bayesian network were similar, but additional analysis is needed for the difference.

5.3 Sensitivity Analysis

In this chapter, we performed the sensitivity analysis using the developed Bayesian network. Through the sensitivity analysis, we found the effect of prediction horizon and decision threshold. Also, the understanding of the mechanism of the traffic state transition was improved by conducting evidence sensitivity analysis. Therefore, based on the results, we provided insights on the traffic state prediction and discussed future improvement directions.

5.3.1 Prediction horizon

In this study, the prediction horizon (α) was set to ten, which meant to predict the traffic state after five minutes, to consider the aggregation time (30 seconds) and efficient use in traffic operations. It is necessary to confirm whether the prediction horizon, which was set to ten, guarantees the best performance because the performance of the model may change greatly according to the change of the prediction horizon. In the sensitivity analysis, therefore, the effect of the prediction horizon was analyzed by changing it from 1 to 60.

As a result, the model performance was changed according to the prediction horizon. As shown in Figure 5.10, the overall accuracy of the model increased as the prediction horizon increased when the prediction horizon was between 1 and 10. On the other hand, when the prediction horizon was greater than 10, the overall accuracy decreased as the prediction horizon increased.

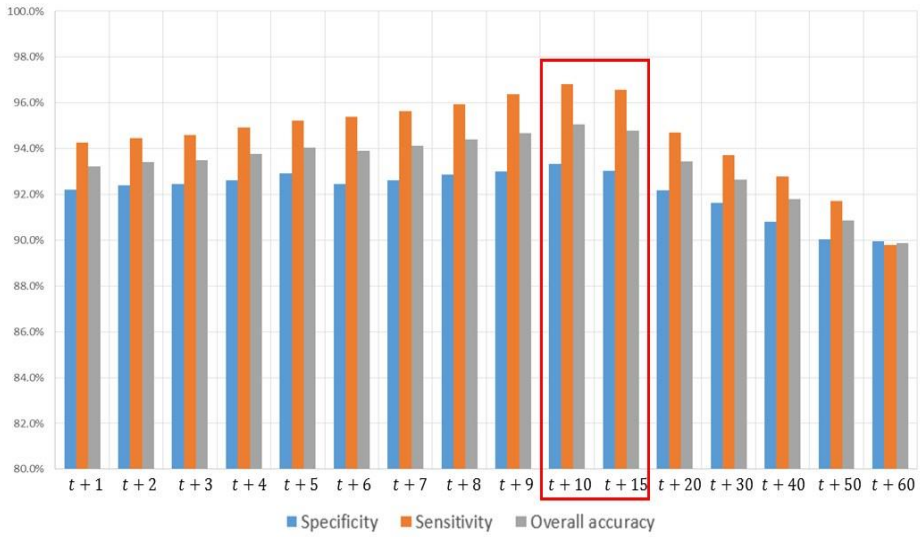


Figure 5.10 Model performance depending on prediction horizon

The results can be explained by the characteristics of the propagation of the traffic flow. The propagation velocities vary depending on the traffic state. Congestion patterns at downstream move upstream generally at a constant speed which is between -20 km/ h and -15 km/h (Treiber et al., 2010). Therefore, as shown in Figure 5.11, the congestion patterns at downstream take 10-13 time steps, about 5-7 minutes, to pass through the target area. Conversely, in the free-flow traffic state, the traffic patterns at upstream take 2-3 time steps, about 1-2 minutes, to pass through the target area. However, the prediction horizon for the best model performance can be determined by the downstream propagation velocities, because the propagation velocity of traffic patterns at upstream may vary depending on traffic state of the target area.

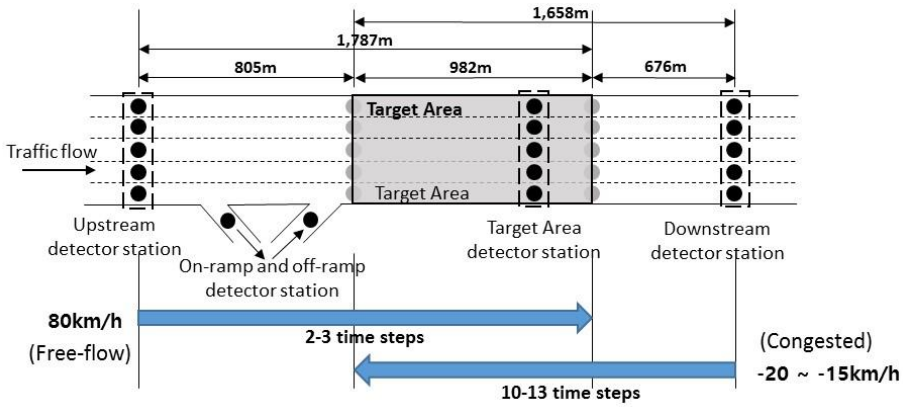


Figure 5.11 Propagation velocities of macroscopic speed in this study

In previous studies comparing multiple prediction horizons for the traffic state prediction, we found similar results. The longer the length of the prediction horizon and the lower the model accuracy (Chang et al., 2012, Vlahogianni et al., 2005; Min and Wynter, 2011). This is in agreement with the results of this study, in which the model accuracy continuously decreased after t + 10. However, no studies have been found to analyze multiple prediction horizons of less than 5 minutes. Nevertheless, we could confirm that the prediction horizon setting was necessary considering propagation velocity and detector locations.

As an additional analysis, the performances of models (Bayesian network, ANN, and logistic regression) depending on the prediction horizon were compared. As a result, regardless of the models (BN and ANN), the best model performance was achieved at $t + 10$. Meanwhile, the performance of the logistic regression model decreased after $t + 3$. The increase in the logistic

regression model might be caused by the difference between the link speed, which was a measure identifying the traffic state of the entire area, and the detector data, which was a sample of the area. The results indicated that the prediction horizon was affected by data collection situation. Therefore, in order to set the prediction horizon, we need to consider propagation velocities and detector locations. Also, it is a consideration when the traffic state prediction is transferred.

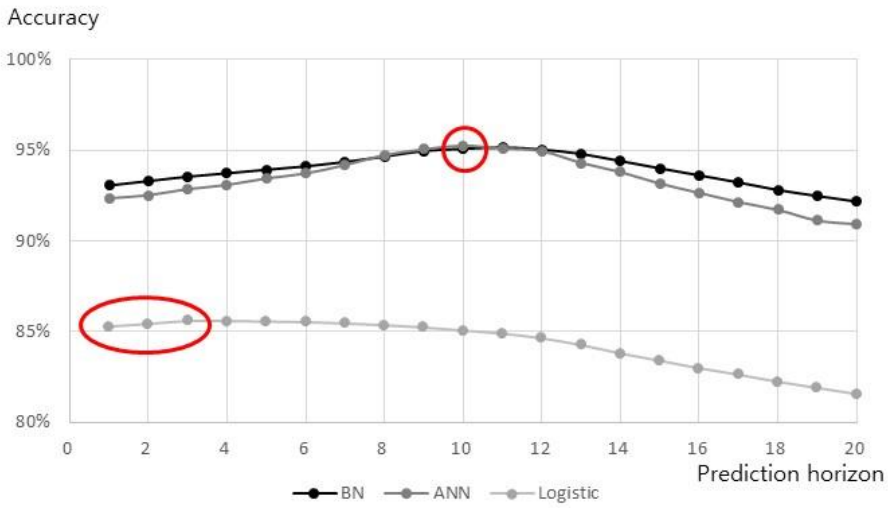


Figure 5.12 Performance comparison depending on prediction horizon

5.3.2 Decision threshold

To confirm how robust the probability estimated from the Bayesian network, we analyzed the change of the false counts depending on the decision threshold

for determining the traffic state. In the model evaluation, the threshold was set at 50%, i.e., if the predicted probability was less than 50%, the traffic state was determined to be free-flow, and conversely, if the predicted probability was equal to or more than 50%, the traffic state was determined to be congested. In this analysis, the decision threshold was changed from 10% to 90% and the change of the false counts was analyzed.

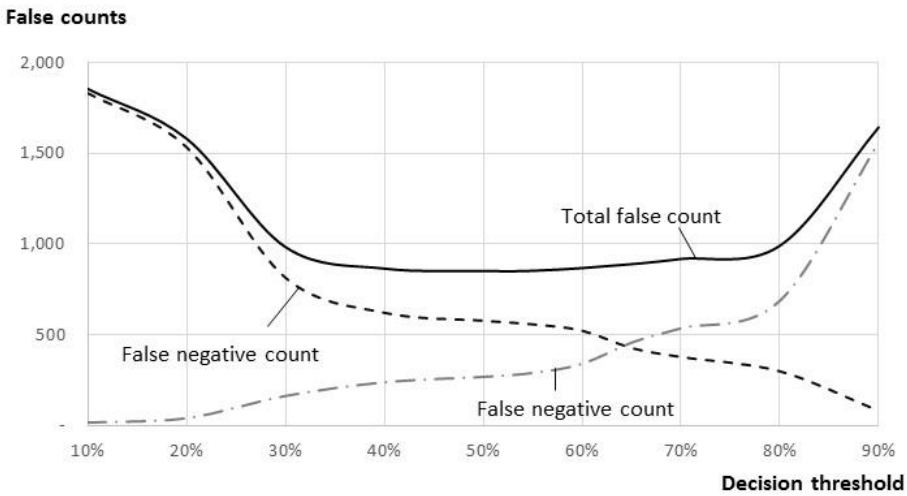


Figure 5.13 Change of false counts depending on decision threshold

As shown in Figure 5.13, the false counts were not sensitive to the decision threshold in 40-60%. The result showed that we could obtain low uncertainty results using the Bayesian network. Therefore, an application using the predicted probability can obtain a robust result. Also, this advantage can increase the availability of the Bayesian network.

5.3.3 Evidence sensitivity analysis

Based on the structure and the conditional probability distributions, the Bayesian network can derive an unbiased probability using pieces of evidence. The function enables to perform various scenario analysis and provides valuable insights on the modeled problem. In this paper, we constructed the traffic state prediction model through the Bayesian network and could improve the understanding of the mechanism of the traffic state transition. In this section, we analyzed the probability changes in traffic state prediction according to the given evidence and identified the effect of each evidence. This is called evidence sensitivity analysis. The analysis was performed by separating the background information (traffic states) and symptom information (detector data).

5.3.3.1 Performance measurements

In this study, we used three performance measurements to perform the evidence sensitivity analysis. The background information consisted of the variables that represented the traffic state of each area, and it had discrete forms, but the symptom information had continuous forms as measured values from the detectors. Therefore, considering these two types of data, the performance measurements were selected. Also, an additional analysis was conducted to study the mechanism of traffic state transition.

Table 5.6 Performance measurements of evidence sensitivity analysis

	Background information (traffic states)	Symptom information (detector data)
Evidence sensitivity analysis	• Cost-of-omission (Kullback-Leibler distance)	• Identification of minimum and maximum beliefs • Normalized likelihood (NL)
Additional analysis	• Performance comparison by omitting each information	

The cost-of-omission, which is a performance measurement based on the Kullback-Leibler distance, was used to analyze the background information. The Kullback-Leibler distance, or Kullback-Leibler divergence, comes from the field of information theory and is given as (Kjaerulff and Madsen, 2008):

$$K(P,Q) = H(P,Q) - H(P) = -\sum_{i=1}^{n} p_i \, log_2 \, q_i + \sum_{i=1}^{n} p_i \, log_2 \, p_i = \sum_{i=1}^{n} p_i \, log_2 \frac{p_i}{q_i}$$

Where $H(P,Q)$ is the cross-entropy of P and Q which expresses the overall difference between two distributions, and $\mathrm{H(P)}$ is the entropy of P, which is a measure of how much information P carries. The value of this measure ranges from 0 to ∞. Here, the cost-of-omission is the measurement that replaced P with $P(X|\varepsilon)$ and Q with $P(X|\varepsilon\backslash\{\varepsilon_i\})$. Using this measurement, we can analyze the impact of the omitted evidence, ε_i, on the problem variable X. Let $X = \{x_1, \dots x_n\}$ be a problem variable and let $\varepsilon = \{\varepsilon_1, \dots, \varepsilon_n\}$ be a set of evidence. Given that, the cost-of-omission of ε_i is as follows (Kjaerulff and Madsen, 2008):

$$c(\varepsilon_i) = c(P(X|\varepsilon), P(X|\varepsilon\backslash\{\varepsilon_i\})) = \sum_{x \in dom(X)} P(x|\varepsilon) log(\frac{P(x|\varepsilon)}{P(X|\varepsilon\backslash\{\varepsilon_i\})})$$

An evidence is considered to be unimportant if the calculated probability is the same with and without the evidence. Conversely, if the calculated probability is affected significantly by the absence of that evidence, it is considered to be important.

In the cost-of-omission, the problem variable X is the future traffic state at time t + 10, and evidence, ε, means the traffic state of each area at time t. The problem variable X has two alternatives (the free-flow and the congested traffic state), so the probability of the congested traffic state increases when the probability of the free-flow traffic state decreases, i.e., the probability changes that occur in the two alternatives are symmetric and opposite when evidence is given. However, cost-of-omission shows an asymmetrical result. For example, when evidence ε_i is omitted and then the probability of the free-flow traffic state is increased, the cost-of-omission of ε_i in the congested traffic state is larger than that in the free-flow traffic state. This means that a higher cost-of-omission is obtained in the traffic state in which the probability is increased by inputting omitted evidence. Therefore, it is possible to analyze how each piece of evidence increases or decreases the probability of the traffic state.

For further analysis of background information, we also performed performance comparison by omitting each information. We analyzed the

accuracy of the future traffic state at time t + 10 without information of each area in relation to the results of the cost-of-omission in the previous section.

Next, to perform evidence sensitivity analysis on symptom information, identification of minimum and maximum beliefs and normalized likelihood (NL) were calculated. Identification of minimum and maximum beliefs is a performance measure to find the maximum and minimum probability of the problem variable X given all possible observations of a variable Y. Here, Y is the variable of interest, and it is not included in evidence ε (Kjaerulff and Madsen, 2008).

$$\underset{y \in Y}{mix}\, P(x|\varepsilon, y),\ \underset{y \in Y}{max}\, P(x|\varepsilon, y)$$

This measurement identifies the range of the posterior belief in a problem variable X according to the variable Y, which indicates the impact of Y on the probability of X. In the study, the measurement of Y was calculated given no evidence ε.

Normalized likelihood (NL) is a useful performance measure that can be used to analyze the impact of different subsets of the evidence on the problem variable x_i $(\in X)$. The magnitude and direction of the probability change of problem variable x_i $(\in X)$ are analyzed according to an evidence. The normalized likelihood (NL) of a problem variable x_i given an evidence ε' is as follows (Kjaerulff and Madsen, 2008):

$$NL = \frac{P(\varepsilon'|x)}{P(\varepsilon')} = \frac{P(\varepsilon', x)/P(x)}{P(\varepsilon')} = \frac{P(x|\varepsilon')P(\varepsilon')/P(x)}{P(\varepsilon')} = \frac{P(x|\varepsilon')}{P(x)}$$

where it is assumed that $P(\varepsilon') > 0$ and $P(x) > 0$. If an NL is 1, it can be concluded that the evidence had no impact on the change of the probability where no evidence is given. If an NL is greater than 1, the evidence, as a specific value of the symptom information variable, has an impact on increasing the probability and vice versa.

5.3.3.2 Background information variables

Using the cost-of-omission, we analyzed the impact of the current traffic state of each area at time t on the future traffic state of the target area at time t+10. As noted above, the cost-of-omission has asymmetrical results in the probability of the binary alternative. Therefore, it is possible to determine which variables increase the probability of a certain traffic state. We divided the cost-of-omission results into the free-flow traffic state, the congested traffic state, and the sum of the two states.

As shown in Figure 5.14, the cost-of-omission was divided by the target area, upstream, downstream, on-ramp, and off-ramp. In the case of the sum, it was determined that each current traffic state had a great impact on the future traffic state in the following order, i.e., target area, downstream, upstream, on-ramp, and off-ramp. Also, the effects of the current traffic state acted in favor of the congested traffic state in the future (except for on-ramp). In other words, the current traffic state information usually had a greater effect on increasing the probability of the congested traffic state.

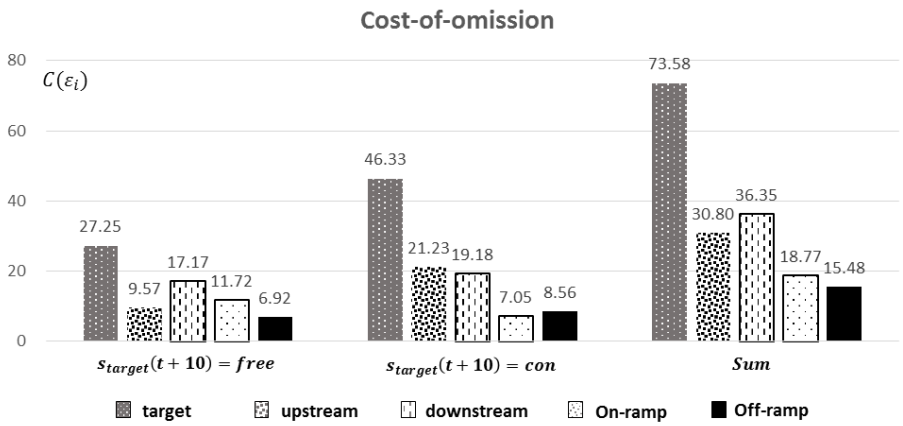


Figure 5.14 Results of cost-of-omission in each area

In addition, it should be noted that the difference in the cost-of-omission between $s_{target}(t+10) = free - flow$ and $s_{target}(t+10) = congested$ in the upstream traffic state. In the case of upstream, the cost-of-omission of the congested traffic state was more than twice as much as that of the free-flow traffic state, whereas these values were similar in the case of downstream. This shows that the current traffic state of the upstream acts more in the direction of increasing the probability of the congested traffic state in the future. In other words, the result indicated that there were many cases that the probability of congested traffic state in the future was increased by the upstream traffic state, or the probability of the congested traffic state in the future was increased greatly by the upstream traffic state.

In order to conduct a detailed analysis, the cost-of-omissions in each area was divided according to the current traffic state of the target area,

$s_{target}(t)$, which has the greatest influence on the prediction of the traffic state. As shown in Figure 5.15, in the sum of each area, the highest cost-of-omission was in transitional state 2 or 3. This is because there is a lot of uncertainty in the transitional states about a future traffic state, which can be lowered by knowing the information of each area. That is, in the transitional state, all information of the areas is important in the traffic state prediction.

When the current traffic state of the target area was free-flow ($s_{target}(t) = 1$), the probability of a future traffic state was almost certain, i.e., cost-of-omissions was very small. Due to the impact of the information of the target area, the probability change by the information of the other area was insignificant. However, one exception was that the future traffic state could become a congested state due to downstream traffic state (The cost-of-omission was 4.29 when $s_{target}(t) = 1$ and $s_{target}(t + 10) = \text{con}$). This could have been caused by an induced breakdown.

It should be noted that the patterns of cost-of-omission of upstream and downstream according to $s_{target}(t)$ were different. The upstream information had a greater impact on the probability of the congested traffic state in all current traffic states of the target area. However, when the current traffic state of the target area was congested ($s_{target}(t) = 3$ or $s_{target}(t) = 4$), downstream information had a greater impact on increasing the probability of a free-flow traffic state.

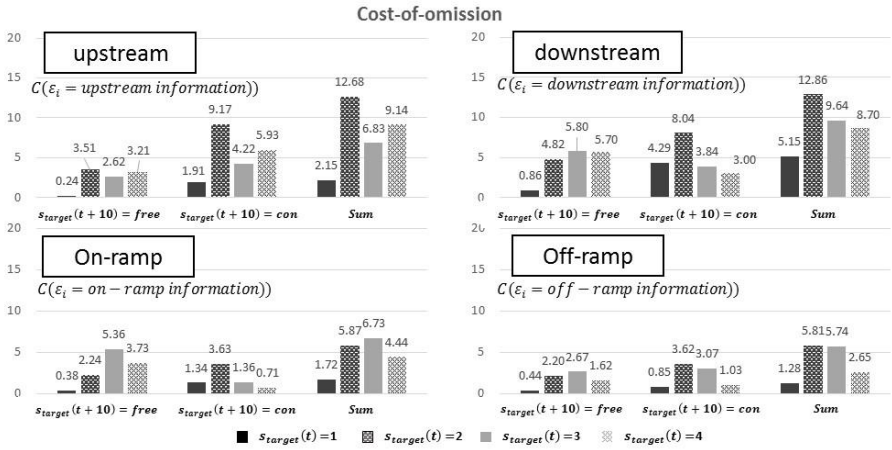


Figure 5.15 Results of cost-of-omission in each area according to the traffic state of the target area

Additional analysis was performed to confirm the impact of the current traffic state information of each area on the future traffic state. Performance comparison by omitting each area information showed how the performance changes if there was no information of each area. After that, the performance comparison results were examined for consistency with the results of the evidence sensitivity analysis. As shown in Figure 5.16, it was confirmed that the overall accuracy decreased with the absence of information of each area. Also, the lowest prediction accuracy was obtained without the target area information. In addition, the false positive rate was significantly increased when there was no target area and downstream information, and the false negative rate was significantly increased without upstream information.

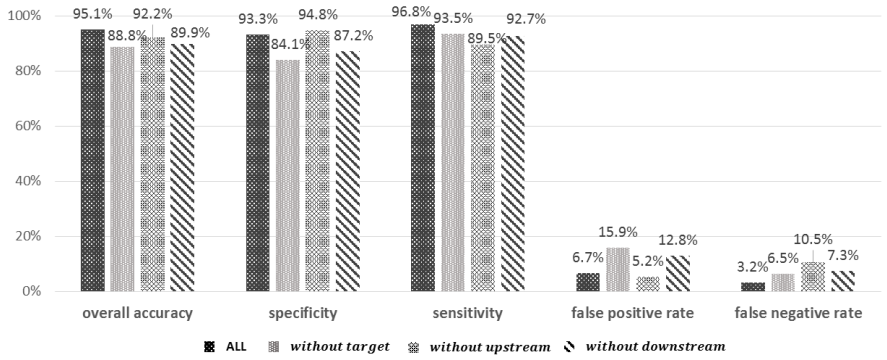


Figure 5.16 Performance comparison by omitting each information

The result which the false negative rate increased due to the absence of upstream information indicated that the future traffic state was actually congested but mispredicted as free-flow in many cases. In other words, due to the absence of upstream information, which had a greater impact on the probability of the congested traffic state in the future, the uncertainty about the probability of the congested traffic state in the future was increased and consequently, the errors in the congested traffic state were increased. On the contrary, the result which the false positive rate increased due to the absence of downstream information indicated that the future traffic state was actually free-flow but mispredicted as congested in many cases. In other words, due to the absence of downstream information, which had a greater impact on the probability of the free-flow traffic state in the future, the uncertainty about the probability of the free-flow traffic state in the future was increased and consequently the errors in the free-flow traffic state were increased. This was consistent with the cost-of-omission result previously reported.

Here, a peculiar point was that the false positive rate in the model without upstream information (5.2%) was lower than that in the model including all information (6.7%). To interpret the result, $P(s_{target}(t+10) = congested|\epsilon)$ of each model was plotted over time in Figure 5.17. As shown in Figure 5.17, when the actual traffic state was free-flow, the probability was denoted as 0, and when it was congested, it was denoted as 1. Also, Figure 5.17 shows that the probability of the models including the upstream information, i.e., the model with all information, the model without the target area information, and the model without downstream information, tended to be higher than 0.5 before the actual spontaneous breakdown occurs. In other words, if there was no upstream information (the blue line in Figure 5.17), the false positive rate could be reduced, but the false negative rate could be much larger. Although the upstream information might cause the false positive rate, it prevented a prediction delay caused by insensitivity to transition from the free-flow traffic state to the congested traffic state. In addition, the spontaneous breakdowns could be greatly affected by upstream information, but the induced breakdowns caused by the downstream queue could be well predicted without upstream information.

If there is no information of the target area and downstream, it can be confirmed that the probability was rarely close to 0 or 1 because uncertainty exists in given information. Especially, when there is no information of the target area, it was confirmed that the variation of the probability was relatively large due to the effects of the surrounding situation.

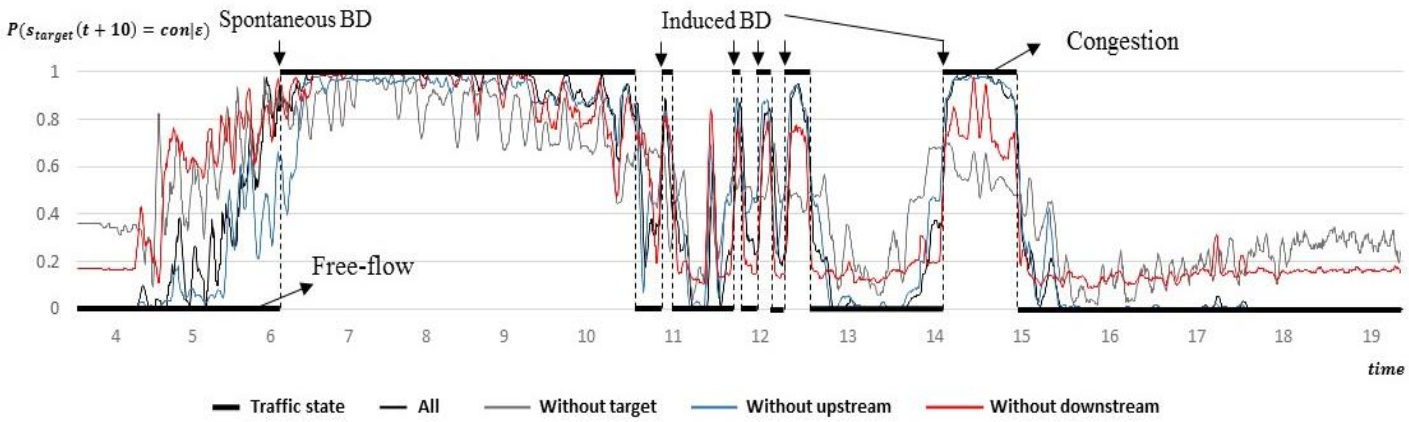


Figure 5.17 Changes of the probability $P\left(s_{target}(t+10) = congested|\varepsilon\right)$ by omitting each information

5.3.3.3 Symptom information variables

To analyze the impact of the traffic variables (symptom information variables) on the traffic state prediction (problem variable), we calculated the identification of the minimum and maximum beliefs and normalized likelihood. The identification of minimum and maximum beliefs is a measure of how much the input evidence ε_i can change the probability of the problem variable. Figure 5.18 shows the result of identification of minimum and maximum beliefs of the traffic variables. The probability $P\left(s_{target}(t+1) = congested \middle| no\ evidence\right)$ is 44.3%.

In the terms of the area, the traffic variables in the target area had a greater impact on $P(s_{target}(t+1) = congested|\mathrm{y})$ than the upstream or downstream variables. Also, upstream information and downstream information had different directions of the impacts. Upstream information could increase $P(s_{target}(t+1) = congested|\mathrm{y})$ to a larger extent than decrease the probability. This is, upstream information mainly contributes to increasing the probability of the congested traffic state. In the other hands, downstream information could decrease $(s_{target}(t+1) = congested|\mathrm{y})$ to a larger extent than increase the probability. This is, downstream information mainly contributes to decreasing the probability of the congested traffic state. The impacts of ramp information on $P(s_{target}(t+1) = congested|\mathrm{y})$ were relatively small.

The impacts on $P(s_{target}(t+1) = congested|\mathrm{y})$ were different,

according to the traffic variables. The impact of traffic flow on $P(s_{target}(t+1) = congested|\mathrm{y})$ was relatively small compared to other variables. In particular, the variation of the probability based on the upstream flow was very small. The traffic state prediction with only flows may have limitations, while speed and occupancy had similar results. Therefore, speed and occupancy in the same area may be regarded as substitution variables in traffic state prediction.

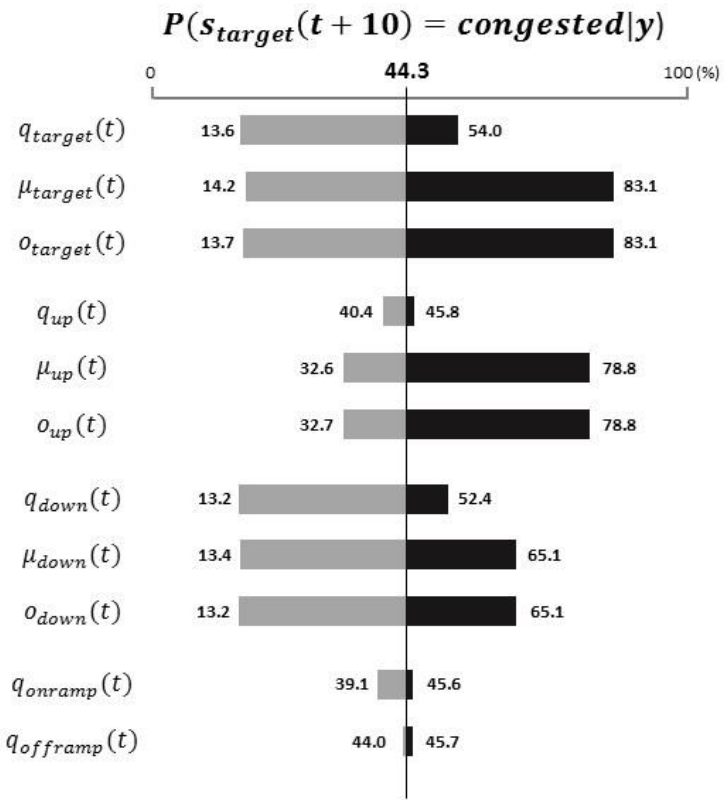


Figure 5.18 Results of identification of minimum and maximum beliefs

Next, normalized likelihood (NL) was calculated to analyze how the probability $P(s_{target}(t+1) = congested|\varepsilon_i)$ changes as the continuous value of symptom information changes. If an NL is 1, it can be concluded that the evidence had no impact on the change of the probability where no evidence is given. If an NL is greater than 1, the evidence, as a specific value of the symptom information variable, has an impact on increasing the probability and vice versa.

Figure 5.19 shows the results of the normalized likelihood of the congested traffic state according to the traffic variables of each area. In the results of flow, it has a significant impact on lowering $P(s_{target}(t+1) = congested|\varepsilon_i)$ when the flows in downstream and target area were smaller than about 40 vehicles per 30 seconds. However, it was close to 1 when the flow was above the value. It also showed a peak at about 60 vehicles per 30 seconds. Consistent with previous results, the upstream flow had little impact on $P(s_{target}(t+1) = congested|\varepsilon_i)$. In the results of speed and occupancy, NL of speed usually decreased as the speed increase, and reversely NL of occupancy usually increased as the occupancy increase. Also, there were abrupt changes occurred in the NLs of speed and occupancy. The NL of speed abruptly decreased at about 80 km/h, which indicated that $P(s_{target}(t+1) = congested|\varepsilon_i)$ was sharply increased at this value and the probability was consistently higher after this value. Like this, the NL of occupancy abruptly increased at about 18%. The flows of ramps had less impact than other traffic variables of the mainline.

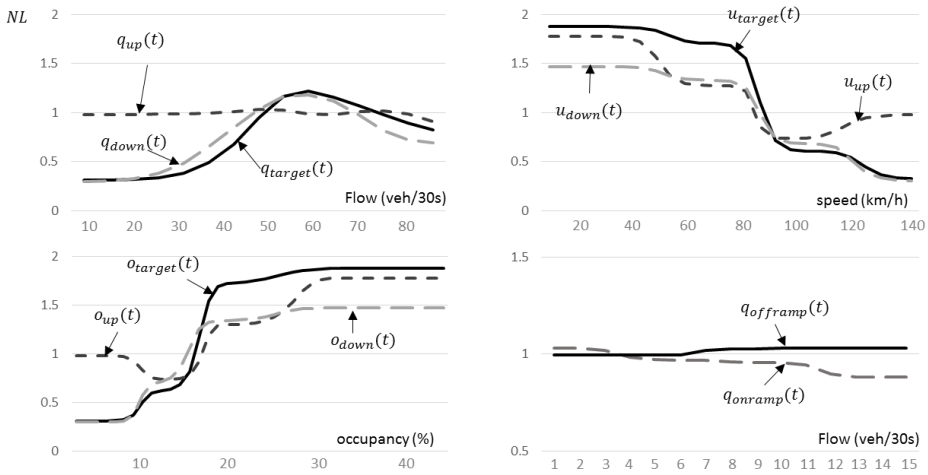


Figure 5.19 Results of normalized likelihood (NL)

It has been confirmed that traffic flow, speed, and occupancy have different impacts on predicting future traffic state. Traffic flow is a good contributor to predict the free-flow traffic state, which was consistent with a previous study (Xia et al. 2012). The results also showed that the speed and occupancy in the same area had almost similar effects. These two variables may have a relationship as substation variables for traffic state prediction. This suggests that the impact of different traffic variables and different areas on predicting future traffic state should be analyzed carefully when selecting traffic variables for a traffic state prediction model.

5.4 Transferability

According to Anney (2014), “transferability refers to the degree to which the results of qualitative research can be generalized or transferred to other contexts or settings”. The contexts and settings can be defined spatially, temporally, or any other manner. However, it is very difficult to develop a model that can satisfy all situations. In addition, the transferability of a model may be high only in a certain context. Therefore, it is important that researchers thickly describe the study contexts or settings to ensure transferability of developed models.

In the traffic state prediction, transferable elements are distinguished by prediction logic and trained and calibrated parameters. Depending on the situation, both or only one can be transferable to other settings. Some studies assessed the transferability of models, which were divided into the parametric and nonparametric approach. Hartig and Dormann (2013) claimed that the parametric approach offers the advantage of the transferability by using previous knowledge (e.g., prior distribution) on the problem. Also, some studies (Brenning, 2005; Elith et al., 2006) claimed that the nonparametric approach might be susceptible to fitting the complex or nonlinear problem unrealistically, which can reduce the transferability of the models. Only a simple model based on the nonparametric approach may have transferability (Hjort et al., 2014). However, even with a parametric approach model, it is difficult to transfer the parameters of a model. Oh et al. (2015) explained that coefficients are site-specific in the parametric approach and, also, the

parameters trained are site-specific in the nonparametric approach. Xu et al. (2014), also, claimed that parameters in the parametric model do not remain stable over time or space. Based on the previous research, we concluded that general knowledge-based and simple logics could transfer to other contexts. For a thorough review of the transferability, the following factors should be considered, i.e., uncertainty in modeled problem, an accuracy of data, a causality of the predictors, scale of analysis, and modeling method.

Generally, a Bayesian network model includes a structure (nodes and links, qualitative element) and parameters (conditional probability distributions, quantitative element). In incident detection, the logic and parameters in a Bayesian network could transfer to another site by converting specific values to states of traffic parameters (Zhang and Taylor, 2006). In crash prediction, also, a dynamic Bayesian network had good transferability with the logic and parameters (Sun and Sun, 2015). In river velocity estimation, the logic and parameters in a Bayesian network could transfer to another site which was in similar conditions of related variables (Palmsten et al. 2013). Therefore, the structure based on general knowledge can be transferable, and the parameters based on data can be partially transferable. Also, to improve transferability, the discretized states instead of absolute values of traffic measurements can be used as a node variable. The states can be classified based on general knowledge.

In this study, the Bayesian network for the traffic state prediction includes the knowledge-based structure and data-driven parameters. Therefore, the structure can be transferable not depending on the external factors, and the

parameters can be partially transferable. Among the parameters, the conditional probabilities between traffic state nodes as discrete data can be transferable in similar geometric sites. If a fundamental diagram is changed in other sites, retraining is needed.

To ensure the transferability of developed models, our study contexts and settings are described. In addition, as shown in Table 5.7, influential factors for the transferability are listed. In this study, the freeway (Alameda County, City of Berkeley, I80-W) is merging and diverging section (with on- and off-ramp) with five-lane, and there are no lane reduction, slope, and curved section. Also, the freeway has 2-5% heavy vehicle percentage and 65 mph speed limit. Construction, incident, and adverse weather are not included.

Table 5.7 Influential factors for transferability

Category	**Factors**
Geometric factors	• Base section / merging and diverging section (on-ramp, off-ramp, on- and off-ramp) / weaving section • Lane reduction / uphill and downhill slope / curved section • Number of lane
Traffic factors	• Heavy vehicle percentage/ speed limit / control strategy • Construction / incident • Driver behaviors
Weather factors	• Clear / rain / snow / fog

Chapter 6. Conclusions

6.1 Summary

The purpose of this study was to develop a traffic state prediction model that reflects the complexity and stochastic process of the traffic state. Therefore, we used Bayesian network suited for a prediction problem with uncertainty and nonlinearity. The greatest advantage of a Bayesian network was interpretability as well as good performance.

First, we conducted a literature review of traffic state estimation and prediction in order to present the originality of this study. Second, we collected the data and we prepared the Bayesian network modeling. Third, we divided the data into two types: observed traffic variables from detectors and future traffic states identified by the traffic state identification algorithm. Using this data, we developed a Bayesian network for traffic state prediction by following the proposed modeling procedure in this study. Fourth, we evaluated model performance and we provided insights to understand the congested traffic pattern. In the evaluation results, this model has high accuracy as a machine learning-based model. Fifth, we conducted sensitivity analyses and we proposed guidelines for traffic state prediction. Finally, we confirmed that the developed model has good performance and interpretability.

6.2 Guidelines for Traffic State Prediction

As noted above, a Bayesian network can be interpreted based on information theory and can derive the probabilities for various scenarios given pieces of evidence. In addition, we conducted sensitivity analyses in the model evaluation. By summarizing the results, we suggested the guidelines for traffic state prediction. The guidelines are applicable for all models of traffic state prediction as well as the Bayesian network.

The impacts on the future traffic state of the target area are different depending on the location of information (e.g., target area, upstream, downstream, and ramps). Therefore, traffic state prediction modeling considering the impacts is needed. The information of the target area has the greatest impact on the prediction. Upstream information has a greater impact on increasing the congestion probability of the target area relatively, and downstream has a greater impact on increasing the probability of free-flow traffic state of the target area when the target area is congested.

The effects of the traffic variables observed from detectors on the future traffic state of the target area are also different. Traffic flow has a minimal impact on the prediction of traffic state in comparison to other variables. However, small traffic flow has a significant effect on lowering the probability of congested traffic state. A consideration is that speed and occupancy are substitution variables. If necessary (e.g., for simplification of the model), only one of the two may be used to reflect the effect.

The following is a guideline for setting the prediction horizon. In the results of the sensitivity analysis for prediction horizon, propagation velocity of traffic patterns and the data aggregated interval, and the detector location should be considered to determine the optimal prediction horizon. The propagation of traffic flow patterns may also be influenced by a variety of external factors. The expansion of spatiotemporal traffic variables can be applied to increase the prediction accuracy and to lengthen the prediction horizon.

Finally, this is a guideline for adapting the Bayesian network to other sites. The structure and modeling procedure can be transferred and the parameter can be partially transferred. Among the parameters, the conditional probabilities between traffic state nodes as discrete data can be transferable in similar geometric sites. If a fundamental diagram is changed on another site, retraining is needed.

6.3 Limitations of the Study

This study has some limitations. The limitations are described in terms of modeling and data. First, in terms of the simplicity of the model, the current traffic states of different areas at the same time is assumed to be independent. However, from the results of individual case analysis, there may be correlations between the current traffic states. The structure considering the correlations of the traffic state can improve the prediction performance of the Bayesian network. Also, this model includes only current traffic variables. Additional

nodes, which can lead to achieving further improvement, are needed to improve this limitation. However, there may be problems caused by the complexity of the modified model, such as increased computation time and correlation with other nodes.

Second, the mixture of Gaussian (MoG) used in this study assumes that the distribution of variables of each state is Gaussian. Actual variables may not satisfy the Gaussian distribution. Thus, it is necessary to examine the effect of the distribution on model performance. In addition, the current traffic states are divided into four states based on model performance. A theoretical review of the classified traffic states may be required.

Third, in terms of the data, there is a lack of applications for various conditions. In this study, the site is a merging and diverging section with an on- and off-ramp and is a serial bottleneck section, where both spontaneous and induced breakdowns occur. If case studies according to various geometry and traffic conditions are performed, we can provide deeper insights for traffic state prediction. Also, traffic state identification is performed based on the link speed estimated in this study. If spatial measurements, such as density, can be measured, the accuracy of the model identification can be improved.

6.4 Applications and Future Research

The Bayesian network for traffic state prediction can have various applications. First, by expanding the model by a corridor, we can obtain more valuable

information for the traffic state. Using a predicted congested traffic state of the corridor, the critical bottleneck of a freeway can be identified and spatiotemporal evolution of congestion probabilistically can be quantified. The identification time and location of potential congestion can be useful information in terms of traffic operation and information. Second, when a connected vehicle environment system is deployed, a freeway operator can communicate with the vehicles approaching a congested section based on quantified traffic state prediction information. In other words, the operator can provide a potential for the congestion that the driver will face in the near future. The driver can then change the path and the operator can alleviate the congestion of a target area. In traffic management, the predicted congested traffic state is also useful to determine the type of control strategy and operation time.

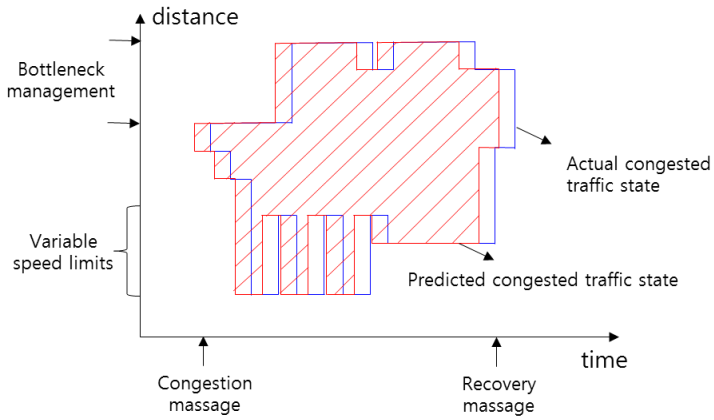


Figure 6.1 predicted congested traffic state in traffic operations

The following is using the developed Bayesian network for traffic operation strategy. The Bayesian network provides a probability using a prior conditional probability distribution of variables with all or some evidence. The estimated probability can be used in various ways, such as in a decision-making problem. In this study, the Bayesian network is designed to predict congestion probability at each time interval. We can estimate an expectation using the probability. The expectation can be expressed as a function of the probability and cost by breakdown:

$$(Expectation) = f(Probability, Cost\ by\ breakdown)$$

The cost by breakdown can be a delay, capacity, or accident cost. Therefore, the expectation can be used for mobility and safety in traffic operations. Also, this function can be helpful in decision making in conjunction with the Bayesian network. As shown in Figure 6.1, the control measure in traffic operations can be evaluated in the Bayesian network by developing the utility function. This strategy can be regarded as an algorithm that considers a dynamic and stochastic traffic flow in comparison with previous algorithms that are deterministic by constant thresholds.

Future research for extended prediction horizon is important. The longer the prediction horizon with acceptable performance, the higher the utilization of the model. Therefore, it is necessary to develop an idea to increase prediction horizon with acceptable performance. Kwon et al. (2000) claimed

that current traffic state is a good predictor for the near future, and that historical data is good for longer range prediction. Thus, the prediction horizon in the Bayesian network can be improved by inputting additional traffic state modules, which can have multiple historical traffic data or multiple spatial data. In addition, the most appropriate times and locations can be determined depending on the detector locations and propagation pattern or traffic flow.

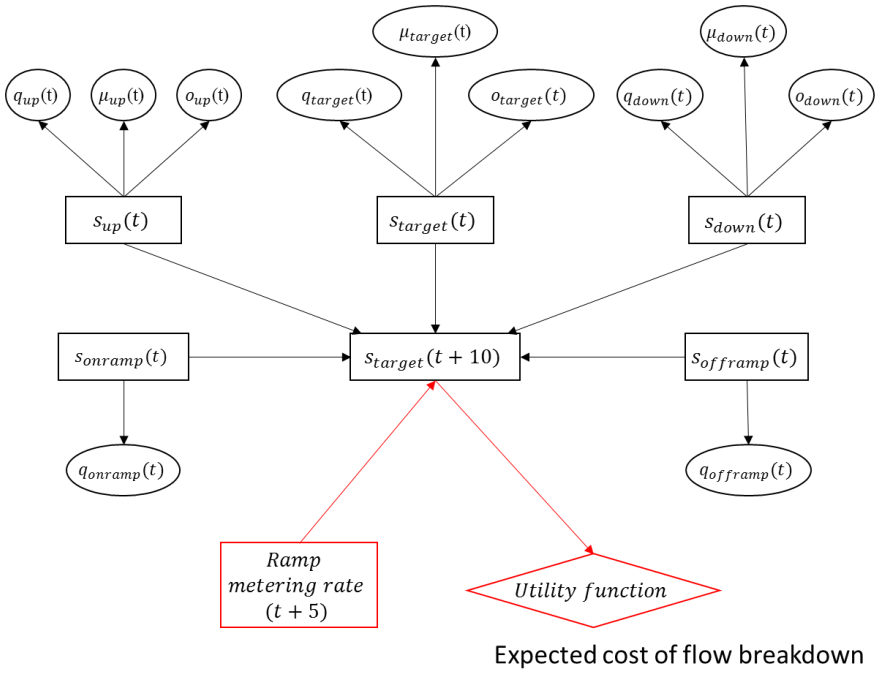


Figure 6.2 Applications of the Bayesian network for decision making

As noted in the limitations of this study, the structure considering the correlations of the traffic state can improve the prediction performance of the

Bayesian network. Therefore, parent nodes are needed that can accurately reflect the correlations. The impacts of the correlations are different depending on the difference of traffic states between the target area and other areas. Therefore, parent nodes, which can adjust the impacts according to the difference level of the traffic states, can be introduced. The parent nodes may have classes that are separated based on traffic states or other indicators. Also, it is necessary to examine the relationship between parent nodes and the future traffic state node.

However, adding additional nodes in the Bayesian network may increase the complexity of the model. Therefore, the correlations in the model, the complicated structure of the Bayesian network, computation problem, overfitting, and transferability are needed for analysis in future research. Continuous variables can also be discretized for simplicity of the model.

References

1. Ahmed, M. S. and Cook, A. R. (1979). Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. Transportation Research Board, 722, pp. 1-9.
2. Anney, V. N. (2014). Ensuring the Quality of the Findings of Qualitative Research: Looking at Trustworthiness Criteria. Journal of Emerging Trends in Educational Research and Policy Studies, 5(2), pp. 272-281.
3. Antoniou, C., Koutsopoulos, H. N., and Yannis, G. (2007). Traffic State Prediction Using Markov Chain Models. Proceedings of the European Control Conference, 2007, 2-5 July, Greece.
4. Antoniou, C., Koutsopoulos, H. N., and Yannis, G. (2013). Dynamic Data-Driven Local Traffic State Estimation and Prediction. Transportation Research Part C: Emerging Technologies, 34, pp. 89-107.
5. Armstrong, J. (2011). Optimizing the Efficiency and Equity of Traffic Flow. PhD thesis, Carleton University Ottawa.
6. Brenning, A. (2005). Spatial Prediction Models for Landslide Hazards: Review, Comparison and Evaluation. Natural Hazards and Earth System Science, 5(6), pp. 853-862.
7. Brilon, W., Geistefeldt, J., and Regler, M. (2005). Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. Proceedings of the 16th International Symposium on Transportation and Traffic Theory: Transportation and Traffic Theory, Flow, Dynamics and Human Interaction. Elsevier Ltd., Oxford, United Kingdom, 2005, pp. 125–144.
8. Cassidy, M. J. and Windover, J. R. (1995). Methodology for Assessing Dynamics of Freeway Traffic Flow. Transportation Research Record, 1484,

pp. 73-79.

9. Chandra, S. R. and Al-Deek, H. (2009). Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. Journal of Intelligent Transportation Systems, 13(2), pp. 53-72.
10. Chang, H., Lee, Y., Yoon, B., and Baek, S. (2012). Dynamic Near-Term Traffic Flow Prediction: System-Oriented Approach Based on Past Experiences. IET Intelligent Transport Systems, 6(3), pp.292-305.
11. Chen, C., Wang, Y., Li, L., Hu, J., and Zhang, Z. (2012). The Retrieval of Intra-Day Trend and Its Influence on Traffic Prediction. Transportation Research Part C: Emerging Technologies, 22, pp. 103-118.
12. Chung, K., Rudjanakanoknad, J., and Cassidy, M. J. (2007). Relation between Traffic Density and Capacity Drop at Three Freeway Bottlenecks. Transportation Research Part B: Methodological, 41(1), pp. 82-95.
13. Collazo, R. A., Pessôa, L. A. M., Bahiense, L., Pereira, B. d. B., and Reis, A. F. D. (2016). A Comparative Study between Artificial Neural Network and Support Vector Machine for Acute Coronary Syndrome Prognosis. Pesquisa Operacional, 36(2), pp. 321-343.
14. Davis, G. A. and Nihan, N. L. (1991). Nonparametric Regression and Short-Term Freeway Traffic Forecasting. Journal of Transportation Engineering, 117(2), pp. 178-188.
15. Deng, C., Wang, F., Shi, H., and Tan, G. (2009). Real-Time Freeway Traffic State Estimation Based on Cluster Analysis and Multiclass Support Vector Machine. Paper presented at the Intelligent Systems and Applications, ISA 2009. International Workshop on.
16. Dong, J. and Mahmassani, H. S. (2009). Flow Breakdown and Travel Time Reliability. Transportation Research Record, 2124, pp. 203-212.
17. Dong, J. and Mahmassani, H. S. (2012). Stochastic Modeling of Traffic Flow Breakdown Phenomenon: Application to Predicting Travel Time Reliability. IEEE Transactions on Intelligent Transportation Systems, 13(4),

pp. 1803-1809.

18. Drake, J. S., Schofer, J. L., and May, A. D. (1967). A Statistical Analysis of Speed Density Hypothesis. Proceedings of Third International Symposium on Theory of Traffic Flow, Elsevier North Holland, New York, 1967.
19. Druzdel, M. and Van Der Gaag, L. C. (2000). Building Probabilistic Networks:" Where Do the Numbers Come from?". IEEE Transactions on Knowledge and Data Engineering, 12(4), pp. 481-486.
20. Edie L. C. (1961). Car-Following and Steady-State Theory for Non-Congested Traffic. Operation Research, 9, pp. 66-76.
21. Elefteriadou, L., Kondyli, A., Washburn, S., Brilon, W., Lohoff, J., Jacobson, L., and Persaud, B. (2011). Proactive Ramp Management under the Threat of Freeway-Flow Breakdown. Procedia-Social and Behavioral Sciences, 16, pp. 4-14.
22. Elefteriadou, L., Roess, R. P., and McShane, W. R. (1995). Probabilistic Nature of Breakdown at Freeway Merge Junctions. Transportation Research Record, 1484, pp. 80-89.
23. Elhenawy, M. and Rakha, H. A. (2015). Automatic Congestion Identification with Two-Component Mixture Models. Transportation Research Record, 2489, pp. 11-19.
24. Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., Zimmermann, N. E., and Araujo, M. (2006). Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. Ecography, 29(2), pp. 129-151.
25. Ecography Evans, J. L., Elefteriadou, L., and Gautam, N. (2001). Probability of Breakdown at Freeway Merges Using Markov Chains.

Transportation Research Part B: Methodological, 35(3), pp. 237-254.

26. Florio, L. and Mussone, L. (1996). Neural-Network Models for Classification and Forecasting of Freeway Traffic Flow Stability. Control Engineering Practice, 4(2), pp. 153-164.
27. Hamed, M. M., Al-Masaeid, H. R., and Said, Z. M. B. (1995). Short-Term Prediction of Traffic Volume in Urban Arterials. Journal of Transportation Engineering, 121(3), pp. 249-254.
28. Hartig, F. and Dormann, C. F. (2013). Does Model-Free Forecasting Really Outperform the True Model?. Proceedings of the National Academy of Sciences, 110(42), E3975.
29. Hjort, J., Ujanen, J., Parviainen, M., Tolgensbakk, J., and Etzelmüller, B. (2014). Transferability of Geomorphological Distribution Models: Evaluation Using Solifluction Features in Subarctic and Arctic Regions. Geomorphology, 204, pp. 165-176.
30. Hossain, M. and Muromachi, Y. (2012). A Bayesian Network Based Framework for Real-Time Crash Prediction on the Basic Freeway Segments of Urban Expressways. Accident Analysis & Prevention, 45, pp. 373-381.
31. Huang, Z. (1998). Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery, 2(3), pp. 283-304.
32. Huili D., Guohua, W., and Min, G. (2011). Traffic Guidance Oriented Model of Traffic State Probability Forecast. Journal of Transportation Systems Engineering and Information Technology, 11(2), pp. 27-32.
33. Jensen, F. V. and Nielsen, T. D. (2009). Bayesian Networks and Decision Graphs. Springer Science + Business Media, New York.
34. Kerner, B. S. (2004). The Physics and Traffic. Springer, Heidelberg.
35. Kim, J., Mahmassani, H., and Dong, J. (2010). Likelihood and Duration of Flow Breakdown: Modeling the Effect of Weather. Transportation Research

Record, 2188, pp. 19-28.
36. Kjaerulff, U. B. and Madsen, A. L. (2008). Bayesian Networks and Influence Diagrams. Springer Science + Business Media, New York.
37. Kondyli, A., Elefteriadou, L., Brilon, W., Hall, F. L., Persaud, B., and Washburn, S. (2013). Development and Evaluation of Methods for Constructing Breakdown Probability Models. Journal of Transportation Engineering, 139(9), pp. 931-940.
38. Kühne, R. and Lüdtke, A. (2013). Traffic Breakdowns and Freeway Capacity as Extreme Value Statistics. Transportation Research Part C: Emerging Technologies, 27, pp. 159-168.
39. Kwon, J., Coifman, B., and Bickel, P. (2000). Day-to-Day Travel-Time Trends and Travel-Time Prediction from Loop-Detector Data. Transportation Research Record, 1717, pp. 120-129.
40. Lam, W. H., Tang, Y., Chan, K., and Tam, M.-L. (2006). Short-Term Hourly Traffic Forecasts Using Hong Kong Annual Traffic Census. Transportation, 33(3), pp. 291-310.
41. Li, W., Li, C., Du, X., Qian, K., Zhang, H., and Hou, D. (2010). A Traffic Flow Prediction Model Based on Ordered Logistic Regression. Paper presented at the Digital Content, Multimedia Technology and its Applications (IDC), 2010 6th International Conference on.
42. Lin, L., Wang, Q., and Sadek, A. W. (2015). A Novel Variable Selection Method Based On Frequent Pattern Tree for Real-Time Traffic Accident Risk Prediction. Transportation Research Part C: Emerging Technologies, 55, pp. 444-459.
43. Lorenz, M. and Elefteriadou, L. (2001). Defining Freeway Capacity as Function of Breakdown Probability. Transportation Research Record, 1776, pp. 43-51.
44. May A. D. (1990). Traffic Flow Fundamentals. Prentice-Hall, New Jersey.
45. May A. D. and Keller, E. M. (1967). Non-integer Car-following Models.

Highway Research Record, 199, pp. 19-32
46. Min, W. and Wynter, L. (2011). Real-Time Road Traffic Prediction with Spatio-Temporal Correlations. Transportation Research Part C: Emerging Technologies, 19(4), pp. 606-616.
47. Murphy K. (2001). The Bayes Net Toolbox for Matlab. Computing Science and Statistics, 33(2), pp. 1024-1034
48. Neil, M., Fenton, N., and Nielson, L. (2000). Building Large-Scale Bayesian Networks. The Knowledge Engineering Review, 15(3), pp. 257-284.
49. Noroozi, R. and Hellinga, B. (2014). Real-Time Prediction of Near-Future Traffic States on Freeways Using a Markov Model. Transportation Research Record, 2421, pp. 115-124.
50. Oh, S., Byon, Y.-J., Jang, K., and Yeo, H. (2015). Short-Term Travel-Time Prediction on Highway: A Review of the Data-Driven Approach. Transport Reviews, 35(1), pp. 4-32.
51. Okutani, I. and Stephanedes, Y. J. (1984). Dynamic Prediction of Traffic Volume through Kalman Filtering Theory. Transportation Research Part B: Methodological, 18(1), pp. 1-11.
52. Palmsten, M. L., Holland, K. T., and Plant, N. G. (2013). Velocity estimation using a Bayesian network in a critical-habitat reach of the Kootenai River, Idaho. Water Resources Research, 49(9), pp. 5865-5879.
53. PeMS. The Freeway Performance Measurement System. pems.eecs. berkeley.edu/
54. Persaud, B., Yagar, S., and Brownlee, R. (1998). Exploration of the Breakdown Phenomenon in Freeway Traffic. Transportation Research Record, 1634, pp. 64-69.
55. Persaud, B., Yagar, S., Tsui, D., and Look, H. (2001). Breakdown-Related Capacity for Freeway with Ramp Metering. Transportation Research Record, 1748, pp. 110-115.
56. Selim, S. Z. and Ismail, M. A. (1984). K-Means-Type Algorithms: A

Generalized Convergence Theorem and Characterization of Local Optimality. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1, pp. 81-87.

57. Shiomi, Y., Yoshii, T., and Kitamura, R. (2011). Platoon-Based Traffic Flow Model for Estimating Breakdown Probability at Single-Lane Expressway Bottlenecks. Transportation Research Part B: Methodological, 45(9), pp. 1314-1330.
58. Smith, B. L. and Demetsky, M. J. (1994). Short-Term Traffic Flow Prediction Models- A Comparison of Neural Network and Nonparametric Regression Approaches. Paper presented at the Systems, Man, and Cybernetics, 1994. Humans, Information and Technology., 1994 IEEE International Conference on.
59. Son, B., Kim, T., Kim, H. J., and Lee, S. (2004). Probabilistic Model of Traffic Breakdown with Random Propagation of Disturbance for Its Application. Paper presented at the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems.
60. Sun, J. and Sun, J. (2015). A Dynamic Bayesian Network Model for Real-Time Crash Prediction Using Traffic Speed Conditions Data. Transportation Research Part C: Emerging Technologies, 54, pp. 176-186.
61. Sun, L. and Zhou, J. (2005). Development of Multiregime Speed-Density Relationships by Cluster Analysis. Transportation Research Record, 1934, pp. 64-71.
62. Sun, S., Zhang, C., and Yu, G. (2006). A Bayesian Network Approach to Traffic Flow Forecasting. IEEE Transactions on Intelligent Transportation Systems, 7(1), pp. 124-132.
63. Sun, X., Muñoz, L., and Horowitz, R. (2003). Highway Traffic State Estimation Using Improved Mixture Kalman Filters for Effective Ramp Metering Control. Paper presented at the Decision and Control, 2003. Proceedings. 42nd IEEE Conference on.
64. Sun, Z.-q., Feng, J.-q., Liu, W., and Zhu, X.-m. (2012). Traffic Congestion

Identification Based on Parallel SVM. Paper presented at the Natural Computation (ICNC), 2012 Eighth International Conference on.

65. Treiber, M. and Helbing, D. (2003). An Adaptive Smoothing Method for Traffic State Identification from Incomplete Information. Interface and Transport Dynamics. Springer, pp 343-360.
66. Treiber, M. and Kesting, A. (2013). Traffic flow dynamics. Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg.
67. Treiber, M., Kesting, A., and Helbing, D. (2010). Three-Phase Traffic Theory and Two-Phase Models with a Fundamental Diagram in the Light of Empirical Stylized Facts. Transportation Research Part B: Methodological, 44(8), pp. 983-1000.
68. Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2005). Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction: A Genetic Approach. Transportation Research Part C: Emerging Technologies, 13(3), pp. 211-234.
69. Wang, H., Li, J., Chen, Q.-Y., and Ni, D. (2011). Logistic Modeling of the Equilibrium Speed–Density Relationship. Transportation Research Part A: Policy and Practice, 45(6), pp. 554-566.
70. Wang, H., Rudy, K., Li, J., & Ni, D. (2010). Calculation of Traffic Flow Breakdown Probability to Optimize Link Throughput. Applied Mathematical Modelling, 34(11), pp. 3376-3389.
71. Wang, J. and Shi, Q. (2013). Short-Term Traffic Speed Forecasting Hybrid Model Based on Chaos–Wavelet Analysis-Support Vector Machine Theory. Transportation Research Part C: Emerging Technologies, 27, pp. 219-232.
72. Wang, Y. and Papageorgiou, M. (2005). Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A General Approach. Transportation Research Part B: Methodological, 39(2), pp. 141-167.
73. Williams, B., Durvasula, P., and Brown, D. (1998). Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated

Moving Average and Exponential Smoothing Models. Transportation Research Record, 1644, pp. 132-141.

74. Xia, J. and Chen, M. (2007). A Nested Clustering Technique for Freeway Operating Condition Classification. Computer-Aided Civil and Infrastructure Engineering, 22(6), pp. 430-437.
75. Xia, J., Huang, W., and Guo, J. (2012). A Clustering Approach to Online Freeway Traffic State Identification Using Its Data. KSCE Journal of Civil Engineering, 16(3), pp. 426-432.
76. Xu, C., Wang, W., Liu, P., Guo, R., and Li, Z. (2014). Using the Bayesian Updating Approach to Improve the Spatial and Temporal Transferability of Real-Time Crash Risk Prediction Models. Transportation Research Part C: Emerging Technologies, 38, pp. 167-176.
77. Xu, T.-D., Hao, Y., Peng, Z.-R., and Sun, L.-J. (2013). Modeling Probabilistic Traffic Breakdown on Congested Freeway Flow. Canadian Journal of Civil Engineering, 40(10), pp. 999-1008.
78. Zhang, K. and Taylor, M. A. (2006). Towards Universal Freeway Incident Detection Algorithms. Transportation Research Part C: Emerging Technologies, 14(2), pp. 68-80.
79. Zheng, W., Lee, D.-H., and Shi, Q. (2006). Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. Journal of Transportation Engineering, 132(2), pp. 114-121.

초 록

베이지안 네트워크를 활용한 교통상태의 확률론적 예측

서울대학교 대학원
공과대학 건설환경공학부
박 호 철

교통운영 측면에서 교통상태 예측은 중요한 이슈이다. 교통운영의 중요한 목적 중 하나는 교통류 와해현상을 예방하는 것으로 이를 위해서는 확률적으로 변화하는 교통류 특성을 반영한 교통상태 예측이 필요하다. 그러나 교통류 상태 전이는 다양한 요인에 의해 복잡하고 동시다발적인 영향을 받으며, 정확한 예측과 이해가 부족한 현실이다. 한편, 베이지안 네트워크는 불확실성이 높은 문제에 대한 예측력이 높을 뿐 아니라 연구자에 의해 구축된 모형을 해석하여 문제에 대한 이해를 높여줄 수 있는 방법론이다. 또한, 불완전한 정보에도 편의되지 않은 확률을 도출할 수 있어 다양한 상황에 대한 분석이 가능하다.

본 연구에서는 이러한 동적이고 확률적인 교통류 특성을 반영하기 위해 베이지안 네트워크라는 확률론적 모형을 활용하여 교통상태 예측 모형을 개발하였다. 기존에 교통문제에서 단순하게 활용되던 베이지안 네트워크의 구조를 개선하기 위해 Mixture of Gaussians (MOGs)을 활용한 모형 구축 방안을 제시하였고, 교통류의 공간적 전파를 고려하기 위해 상류, 하류, 램프의 정보를 모두

활용하였다. 특히, 혼잡의 공간적 전파를 고려하기 위해 구간 속도를 산정하여 교통상태 식별을 수행하였다. 구축된 모형의 성능을 평가한 결과, 로지스틱 회귀분석 보다는 크게 개선되고 기계학습 기반의 인공신경망 모형과 유사한 수준의 성능을 가지는 것으로 분석되었다. 또한, 모형 기반의 민감도 분석을 통해 교통상태 예측에 대한 이해를 높이고 향후 모형 개선을 위한 방향을 제시하였다. 따라서 본 연구에서 구축된 베이지안 네트워크는 높은 예측력뿐만 아니라 해석을 통한 통찰력을 얻을 수 있는 교통상태 예측 모형이라고 할 수 있다.

주요어: 베이지안 네트워크, 교통상태 예측, 교통류 와해현상, 확률 모형, 확률 과정

학 번: 2011-20980