

# 나이브 베이즈 분류기를 이용한 돌발상황 검지 알고리즘 개발

## Development of Incident Detection Algorithm Using Naive Bayes Classification

강 성 관\* · 권 봉 경\*\* · 권 철 우\*\*\* · 박 상 민\*\*\*\* · 윤 일 수\*\*\*\*\*

\* 주저자 : 한국도로공사 건설처 차장  
 \*\* 교신저자 : 한국도로공사 해외사업처 차장  
 \*\*\* 공저자 : 아주대학교 건설교통공학과 석사과정  
 \*\*\*\* 공저자 : 아주대학교 건설교통공학과 박사과정  
 \*\*\*\*\* 공저자 : 아주대학교 교통시스템공학과 부교수

Sunggwon Kang\* · Bongkyung Kwon\*\* · Cheolwoo Kwon\*\*\* ·  
 Sangmin Park\*\*\*\* · Ilsoo Yun\*\*\*\*\*

\* Construction Division, Korea Expressway Corporation  
 \*\* Overseas Project Div. Korea Expressway Corporation  
 \*\*\* Dept. of Transportation Eng., Ajou University  
 \*\*\*\* Dept. of Transportation Eng., Ajou University  
 \*\*\*\*\* Dept. of Transportation Eng., Ajou University  
 † Corresponding author : Bongkyung Kwon, bk.kwon528@gmail.com

Vol.17 No.6(2018)

December, 2018

pp.25~39

pISSN 1738-0774

eISSN 2384-1729

<https://doi.org/10.12815/kits.2018.17.6.25>

2018.17.6.25

Received 28 August 2018

Revised 19 September 2018

Accepted 7 November 2018

© 2018. The Korea Institute of  
 Intelligent Transport Systems. All  
 rights reserved.

### 요 약

본 연구에서는 최근 활발하게 활용되고 있는 머신러닝 기법을 교통분야에 적용하여 효율적인 돌발상황 검지 알고리즘을 개발하는 것을 목적으로 하였다. 미시교통시뮬레이션 모형을 통하여 대상지의 네트워크를 구축하였고 돌발상황에 영향을 줄 것으로 예상되는 변수의 여러 조합을 통해 시나리오를 설정하여 가상의 돌발상황 데이터를 수집하였다. 다음으로 대표적인 돌발상황 검지 알고리즘인 McMaster 알고리즘과 본 연구에서 개발한 나이브 베이즈 분류기를 구현하여 비교·평가하였다. 비교 결과, 나이브 베이즈 분류기가 McMaster 알고리즘에 비해 돌발상황 검지 간격에 따른 부정적인 영향이 적었고 더 우수한 검지율을 보였다. 하지만 검지율이 증가하는 만큼 오검지율 또한 증가하는 것을 확인할 수 있었다. McMaster 알고리즘은 4주기를 통해 검지가 가능하지만 나이브 베이즈 분류기는 1주기(30초)만으로 돌발상황을 판단할 수 있다. 본 연구를 통해 개발한 나이브 베이즈 분류기가 효율적으로 돌발을 파악할 수 있다는 것을 확인할 수 있었다.

핵심어 : 돌발 검지 알고리즘, 머신러닝, McMaster 알고리즘, 나이브 베이즈 분류기

### ABSTRACT

The purpose of this study is to develop an efficient incident detection algorithm by applying machine learning, which is being widely used in the transport sector. As a first step, network of the target site was constructed with micro-simulation model. Secondly, data has been collected under various incident scenarios produced with combination of variables that are expected to affect the incident situation. And, detection results from both McMaster algorithm, a well known incident detection algorithm, and the Naive Bayes algorithm, developed in this study, were compared. As a result of comparison, Naive Bayes algorithm showed less negative effect and better detect rate (DR)

than the McMaster algorithm. However, as DR increases, so did false alarm rate (FAR). Also, while McMaster algorithm detected in four cycles, Naive Bayes algorithm determine the situation with just one cycle, which increases DR but also seems to have increased FAR. Consequently it has been identified that the Naive Bayes algorithm has a great potential in traffic incident detection.

Key words : Incident detection algorithm, Machine learning, McMaster algorithm, Naive Bayes classification

## I. 서 론

### 1. 연구의 배경 및 목적

우리나라의 교통사고 사망률은 꾸준히 감소하고 있지만, 여전히 다른 선진국과 비교하면 높은 수준이다. 2015년 기준 우리나라 인구 10만 명 당 사망자는 9.1명으로 OECD 국가 중 3위를 기록하였다(Korea Road Traffic Authority, 2018). 특히 교통사고나 차량의 고장이 일어났으나 뒤따르던 차량이 그 사실을 미처 알지 못해 다시 사고를 일으키는 '2차사고'로 인한 고속도로 사망자는 연 평균 37명(최근 5년 기준)으로 전체 사망자의 15.3%를 차지하고, 치사율은 52.7%로 일반사고 평균(9.1%)의 약 6배 수준으로 매우 높았다. 지난 2012년 이후 2차사고의 발생은 지속적으로 감소 추세지만, 2017년에는 전년도 31명 대비 29% 급증한 바 있다(Gailbo, 2018). 이처럼 사전에 예측할 수 없는 교통사고, 차량고장 등의 돌발상황(incident)이 발생하였을 때 즉각적으로 검지하여 운전자가 안전하고 효율적으로 대응할 수 있도록 교통정보 서비스를 제공하는 것이 매우 중요할 것으로 판단된다.

이러한 목적을 위하여 고속도로 교통관리센터에서는 고속도로에 교통정보 수집장치(Vehicle Detection System, 이하 VDS)를 설치하여 차로별 차량의 속도, 교통량, 점유율, 차량 길이 등의 정보를 수집하고 있다. VDS로부터 수집된 차로별 데이터는 가공을 통해 단위구간의 소통정보로 생성되어 운전자에게 전광판, Web, App 등의 교통정보 제공 장비를 통해서 서비스되고 있다. 특히 돌발상황을 판단하는 돌발 검지 시스템은 고속도로와 같은 연속류 도로에서 매우 중요한 하부 시스템이며, C-ITS 등 고속도로가 지능화될수록 더욱 중요한 역할을 수행할 것으로 기대된다.

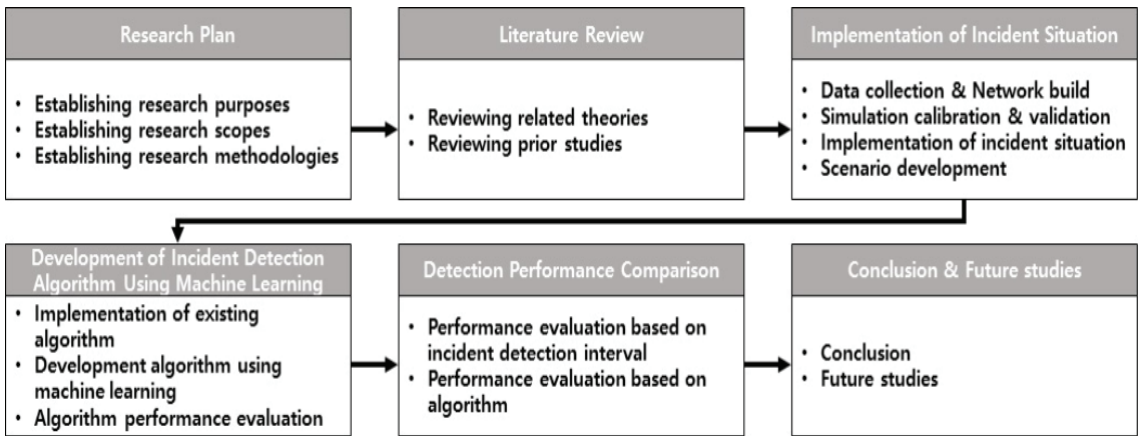
기존 돌발 검지 알고리즘은 패턴 인식, 통계적 기법, 시계열 분석 등의 다양한 방법을 기반으로 개발되었지만 알고리즘의 높은 오검지율과 낮은 검지율 등의 성능 문제로 인해 실질적으로 현장에서 적용되기에는 한계가 있다. 따라서 기존 돌발 검지 알고리즘의 방법이 아닌 새로운 방법을 적용할 필요가 있다.

이에 본 연구에서는 최근 머신러닝 기법을 활용하여 기존 문제를 해결하고 시스템의 성능을 높이기 위한 연구 사례가 급격히 증가하는 추세에 맞춰, 머신러닝 기법을 교통 분야에 적용하여 효율적인 돌발 검지 알고리즘에 대한 방향을 제시하고자 한다. 이를 위해 머신러닝 기법의 일종인 나이브 베이즈 분류기(Naive Bayes Classification)를 이용하여 돌발 검지 알고리즘을 개발하고 기존 알고리즘과 비교하여 그 성능을 검증하고자 한다.

### 2. 연구의 범위 및 방법

본 연구의 시간적 범위는 2017년이며, 공간적 범위는 인천-김포민자고속도로의 청라지하차도로 설정하였고, 고속도로의 본선 구간에 중점을 두어 연구를 진행하였다. 인천-김포민자고속도로는 기존에 사용되는 루프 검지기 외에 루프 검지기 보다 설치 간격이 좁은 지자기 검지기를 설치하는 등 신속한 돌발 검지를 위하

여 많은 노력을 기울이고 있다. 이러한 노력의 일환으로 돌발 검지 알고리즘에 대한 고도화 연구 등이 진행된 바 있다. 청라지하차도는 2017년 2월 개통되었으며, 2017년 총 사고 건수가 3건으로 충분한 돌발 데이터를 수집하기에 어려움이 있었다. 따라서 본 연구는 미시교통시물레이션인 VISSIM ver.10.0을 이용하여 네트워크를 구축한 후 돌발상황에 영향을 줄 수 있는 변수들을 선정하고 조합하여 가상의 돌발 시나리오를 산출하였다. 다음으로 Python 3.6을 이용하여 대표적인 기존 돌발 검지 알고리즘으로서 McMaster 알고리즘과 머신러닝 기법인 나이브 베이지 분류기를 활용하여 돌발 검지 알고리즘을 개발하였다. 알고리즘의 성능평가 지표로는 검지율과 오검지율을 선정하였고, 기존 알고리즘과 본 연구에서 개발한 나이브 베이지 분류기 알고리즘을 돌발검지 간격 조정에 따른 성능을 비교·평가하였다. 본 연구의 절차는 <Fig. 1>에 제시되었다.



<Fig. 1> Research process

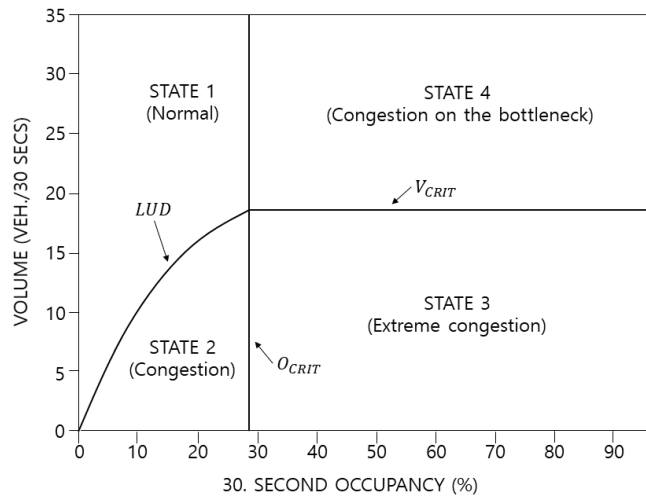
## II. 관련 이론 및 연구 고찰

### 1. 관련 이론 고찰

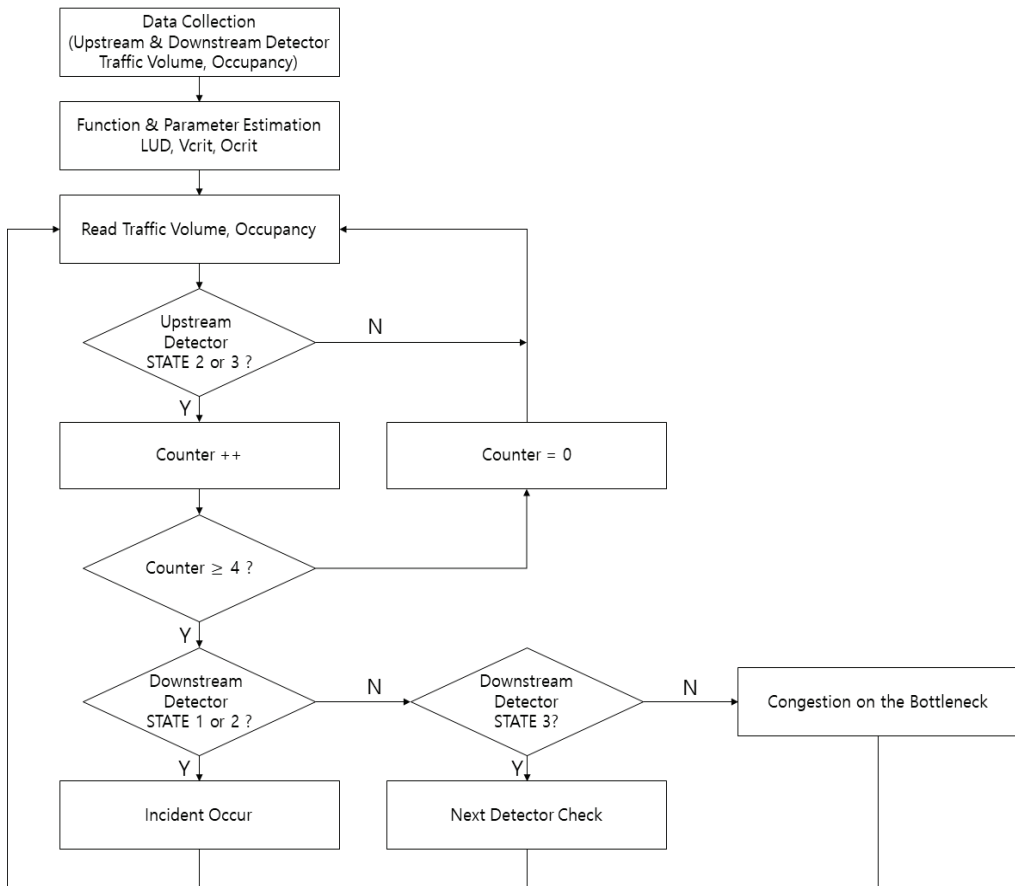
#### 1) McMaster 알고리즘

일반적으로 돌발이 발생하면 발생지점의 상류부에서는 점유율 증가, 속도 감소 등과 같은 정체 상태가 발생하게 되고 발생지점의 하류부에서는 점유율 감소, 속도 유지 또는 증가 등과 같은 소통양호 상태를 띄게 된다. 이러한 교통류 이론에 기반을 둔 대표적인 돌발 검지 알고리즘이 McMaster 알고리즘이다(Hall et al., 1991).

McMaster 알고리즘은 복잡한 교통류 이론을 이용하여 돌발상황 시 교통행태를 묘사하고 예측한 후, 실제 교통 변수와 모형에서 예측한 변수들을 비교하는 방법을 사용한다. 먼저 각 검지기의 1주일 정도의 안정적인 점유율과 교통량 자료를 이용하여 표준치차함수인 lower bound of uncongested data(LUD) 곡선을 추출하고 대기행렬 풀림 교통류의 최소교통량(critical volume, 이하 Vcrit), 임계점유율(critical occupancy, 이하 Ocrit)을 이용하여 아래 <Fig. 2>와 같이 정체 상태를 네 가지로 구분한다(Sun et al., 2017). 다음으로 상류부 검지기를 이용하여 혼잡 여부를 확인하고 하류부 검지기 자료의 교통상태를 통하여 혼잡의 원인을 결정하게 된다(Kim et al., 2016). <Fig. 3>은 McMaster 알고리즘의 작동 절차를 보여주고 있다.



<Fig. 2> Traffic volume-occupancy graph for McMaster algorithm



<Fig. 3> McMaster algorithm flow chart

## 2) 나이브 베이즈 분류기

머신러닝(machine learning)은 인공지능의 한 분야로 데이터를 통해 얻어지는 경험으로부터 특정한 목표 작업에 대한 성능을 향상시키는 일련의 과정으로 정의된다. 수많은 머신러닝 기법 중 적절한 알고리즘을 선정하기 위해서는 목적에 대한 정의와 데이터에 대한 이해가 필요하다(Cho et al., 2016). 머신러닝 기법 중 하나인 나이브 베이즈 분류기는 주로 스팸 필터나 키워드 검색을 활용한 문서 분류에 사용된다. 나이브 베이즈 분류기의 기본 원리는 조건부 확률에 베이즈 정리(Bayes's theorem)를 적용하고, 문서나 데이터를 구성하는 각각의 요소들이 등장할 확률에 대한 독립성을 가정하여 입력 벡터를 확률적으로 분류하는 것이다(Jang, 2016). 식 (1)은 베이즈 정리를 활용하여 새로운 자료를 각 분류에 속할 것인지에 대한 확률을 계산하는 식이다.

$$p(C_k|X) = \frac{p(X|C_k)p(C_k)}{p(X)} \quad (1)$$

만약 데이터 간의 모든 변수의 독립성 가정을 하지 않고 이를 계산하기 위해서는 식 (2)와 같이 변수간의 영향을 고려하기 때문에 매우 복잡한 연산을 수행해야한다.

$$\begin{aligned} \cdot p(C_k, x_1, \dots, x_n) &= p(C_k)p(x_1, \dots, x_n|C_k) \\ &= p(C_k)p(x_1|C_k)p(x_2, \dots, x_n|C_k, x_1) \\ &= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1)p(x_3, \dots, x_n|C_k, x_1, x_2) \\ &= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1) \cdots p(x_3, \dots, x_n|C_k, x_1, x_2, x_3, \dots, x_{n-1}) \end{aligned} \quad (2)$$

하지만 나이브 베이즈 분류기는 독립성 가정을 통해 식 (3)과 같이 각 특성의 확률에 대한 곱으로 표현가능하며, 간단하게 계산할 수 있다(Jeong, 2018).

$$\begin{aligned} \cdot p(C_k, x_1, \dots, x_n) &\propto p(C_k, \dots, x_n) \\ &= p(C_k)p(x_1|C_k)p(x_2|C_k)p(x_3|C_k) \cdots \\ &= p(C_k) \prod_{i=1}^n p(x_i|C_k) \end{aligned} \quad (3)$$

1950년대 이후 광범위하게 연구되고 있으며, 적절한 전처리를 거치면 서포트 벡터 머신(Support Vector Machine)과도 경쟁할 만큼 우수한 분류 성능을 보여준다. 나이브 베이즈 분류기의 가장 큰 장점은 소량의 데이터를 가지고 작업이 이루어지며 여러 개의 분류 항목을 다룰 수 있다는 점에 있다. 특히 독립변수의 차원이 증가할 경우 모형의 설명력을 유지하기 위해 필요한 데이터의 수가 기하급수적으로 많아지는 차원의 저주 문제를 완화 할 수 있다. 하지만 모든 독립변수는 동등하게 중요하고 독립적이라는 가정에 의존하는 단점을 가지고 있다(Jeong, 2018).

## 2. 관련 연구 고찰

Rakha et al.(1999)는 돌발검지 알고리즘 테스트베드 구축을 위해 실제 현장에서 수집된 168시간의 26건 돌발 데이터와 시뮬레이션을 통해 수집된 60시간의 75건 돌발 데이터를 수집하였다. 수집된 데이터의 테스트를 위해 McMaster 알고리즘을 적용한 후 테스트한 결과, 실제 현장 데이터와 시뮬레이션 데이터에서 검지율 98%의 성능을 보였다. 검지 성능에 영향을 주는 변수는 피크 시간의 유무, 돌발의 지속 시간, 돌발의 위치, 램프의 위치 등으로 나타났다.

Kim et al.(2008)은 McMaster 알고리즘의 파라메타의 민감도를 분석하여 일정 수준의 신뢰도를 갖는 파라메타의 오차범위를 결정하여 제시하였다. 파라메타의 민감도 분석을 위해서, 시뮬레이션 모형을 사용하여 교통류 변수자료를 수집하였고, 컴퓨터 언어를 사용하여 McMaster 알고리즘을 구축하였다. 구축된 알고리즘 수행결과, 오검지율, 검지시간 등 성능지표의 값들의 변화에 따라 다르게 나타났으며, 통계적인 민감도 분석 결과 성능지표들 중 오검지율이 가장 민감하게 반응하는 것으로 나타났다.

Lee et al.(2003)는 연속류 도로의 돌발상황을 검지하기 위한 기법으로서 반복정체의 확률을 예측하는 Ferrari\_A1 알고리즘과 발생한 돌발상황을 검지하는 VKDIFF 알고리즘의 개념을 제안하였다. 이론상의 한계점을 넘어설 경우 실제로 정체상황이 발생함을 보여주어 우회 안내 등 교통류 관리에 충분한 이용 가능성을 보여주었다. Transition indicator는 가장 단순하면서도 혼잡 상황과 돌발 상황의 정체 특징을 잘 보였다. 또한 연구의 심화를 통해 교통량-Indicator와 속도-Indicator의 데이터 값을 종합 분석하면 교통량이나 속도 등의 하나의 속성으로 돌발상태를 예측하는 것보다 더 정확한 예측이 가능할 것으로 기대된다.

Lee(2017)는 머신러닝 기법을 이용하여 교통사고가 잦은 지점 선정 방법론을 개발하였다. 실제 5년간 발생한 교통사고 특성을 반영한 시뮬레이션 데이터를 생성하고 데이터 분석을 진행하여 타당성을 평가하였다. 분석 결과, 로지스틱 회귀분석의 결과는 선형회귀 모형 알고리즘보다 성능이 좋았는데 그 이유는 로지스틱 회귀분석 방법이 훨씬 더 최적화 되어 있기 때문으로 판단하였다. 머신러닝 기반의 알고리즘은 분석과정이 단순해 적용이 용이할 뿐만 아니라, 분석 결과 또한 우수해 교통사고 잦은 지점을 선정하는데 강력한 이점을 지니는 것으로 나타났다.

Jung(2018)은 순서형 로짓 모형과 나이브 베이즈 분류기를 이용하여 렌터카 교통사고의 심각도를 예측하는 모형을 개발하였다. 모형 구축에 사용한 렌터카 교통사고 자료는 총 12개의 변수로 구성되었으며, 종속변수로 사용할 교통사고 심각도와 독립변수로 사용할 인적 특성, 사고 특성, 기하구조 특성으로 구분하였다. 분석 결과, 순서형 로짓 모형과 나이브 베이즈 분류기 모두 기존 연구에서 유의한 변수만 사용할 경우 예측 성능이 우수하게 나왔다. 예측 성능은 순서형 로짓 모형에 비해 성능이 5.3% 높은 것으로 나타났지만 정분류 개수가 9.3건으로 모형간의 성능 차이는 미미한 것으로 판단하였다.

### 3. 기존 연구와의 차별성

돌발 검지 알고리즘 관련 연구는 기존에 존재하는 돌발 검지 알고리즘의 낮은 검지율과 높은 오검지율의 단점을 극복하고자 다양한 방법을 활용하였다. 또한 머신러닝 기법을 활용하여 기존의 다양한 교통 분야 알고리즘 문제점을 극복하는 연구도 살펴볼 수 있었다. 하지만 돌발 검지 분야에서의 적용은 활발하지 않다. 이에 본 연구는 데이터의 신속하고 효율적인 처리가 가능하게 됨에 따라 발전하게 된 머신러닝 기법을 활용하여 새로운 돌발 검지 알고리즘을 개발하였다. 대표적인 기존 돌발 검지 알고리즘인 McMaster 알고리즘과 검지 성능을 비교 평가함으로써 연구의 차별성을 제시하고 앞으로의 돌발 검지 알고리즘이 나아갈 방향성을 제시하고자 한다.

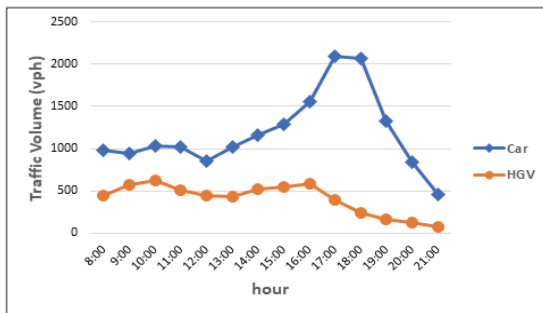
### Ⅲ. 미시교통시물레이션 모형을 이용한 돌발상황 구현

#### 1. 분석지점 선정 및 교통 현황

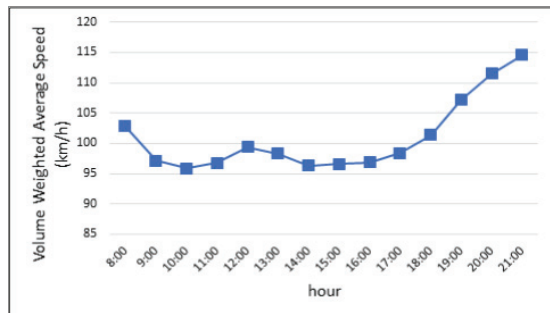
새로운 돌발 검지 알고리즘을 개발하기 위한 분석 지점으로는 인천-김포민자고속도로의 청라지하차도로 선정하였다. 현재 청라지하차도는 편도 3차선 도로로서 진입, 진출 램프가 없는 본선 구간으로 구성되어 있다. 총 연장은 2.54km이며 2017년 2월 개통하여 2017년 총 사고건수는 3건으로 충분한 돌발 데이터가 부족하다. 따라서 본 연구에서는 미시교통시물레이션 모형을 이용하여 가상의 돌발데이터를 산출하여 연구를 진행하였다. 대신에, 기존 검지기 간격 등은 청라지하차도에 설치된 현황을 그대로 반영하였다.

#### 2. 자료수집

자료수집은 미시교통시물레이션 프로그램인 VISSIM의 입력 값들을 구하기 위해 수행되었다. VISSIM의 네트워크 구축 시 필요한 기하구조 자료, 교통량 및 속도 자료를 수집하였다. 기하구조 자료는 대상지의 교통관리 개황도를 활용하였고 교통량 및 속도 자료는 2018년 1월 5일 08시 00분부터 22시 00분까지 총 14시간의 실제 검지기 자료를 이용하였다. 해당시간의 소형차와 중·대형차(HGV)의 교통량 분포 그래프와 교통량 가중평균 속도 그래프는 아래 각각 <Fig. 4>와 <Fig. 5>와 같다. <Fig. 4>는 중·대형차의 교통량은 시간에 따라 차이가 크지 않았지만 소형차는 17시에서 19시 사이에 급격히 많은 교통량을 가졌다. 참고로 <Fig. 4>의 교통량은 편도 3차로의 교통량이며, 전체적으로 교통량 수준은 매우 낮은 것을 확인할 수 있었다. <Fig. 5>에서 교통량 가중 평균 속도는 일정한 속도를 떠다 퇴근 시간 이후부터 증가하는 것을 확인할 수 있었다.



<Fig. 4> Traffic volume graph



<Fig. 5> Traffic volume weighted average speed graph

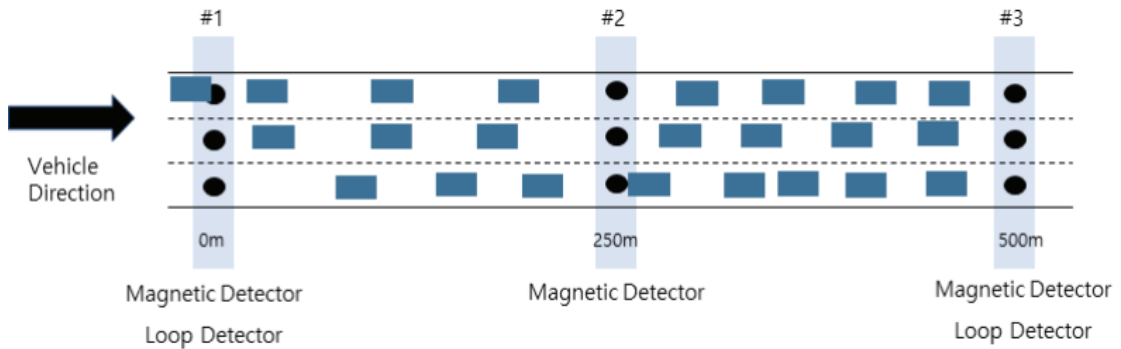
#### 3. 네트워크 구축

대표적인 미시교통시물레이션 프로그램인 VISSIM은 실제 교통상황을 가상에서 현실과 유사하게 구현할 수 있기 때문에 국내를 비롯한 70여 국가에서 활용되고 있다. 본 연구에서는 VISSIM ver. 10.0을 이용하여 네트워크를 구축하였다. 구축에 앞서 수집한 기하구조, 교통량, 속도 등의 데이터를 입력하고 <Table 1>과 같이 파라메타 값을 조정하였다.

<Table 1> Simulation parameter adjustment

Parameters	Values
Desired Speed Distribution(car)	80 ~ 130 km/h
Desired Speed Distribution(HGV)	80 ~ 90 km/h
Link Car-following Model	Weidmann 99
Vehicle Composition	Car(75%) HGV(25%)
Temporary Lack of Attention	5.00s, 5.00%
Desired Position at Free Flow	any
Observe Adjacent Lane	admit
Warmup Time	900s

구축한 네트워크를 통해 자료를 얻기 위해서 가상의 검지기로서 VISSIM에서 제공하는 기능인 ‘data collection point’를 활용하였고, 청라지하차도 개황도를 따라 250m 간격으로 3개 지점에 가상의 검지기를 설치하였다. 청라지하차도의 검지 지점은 지자기 검지기와 루프 검지기가 같이 매설된 지점과 지자기 검지기만 매설된 지점으로 나뉜다. 실제 검지 지점을 도식화한 <Fig. 6>를 살펴보면 지자기 검지기는 검지 지점 #1, #2, #3 모든 지점에 매설되어 있는 반면, 루프 검지기는 검지 지점 #1과 #3에만 매설되어있다. 돌발 검지 알고리즘은 대체적으로 상류부 검지기와 하류부 검지기의 수치적 차이를 이용하여 판단하기 때문에 지자기 검지기를 활용하였을 경우 250m 검지 간격으로, 루프 검지기를 활용하였을 경우 500m 검지 간격으로 돌발 검지가 가능하다. 본 연구는 청라지하차도의 지자기 검지기와 루프 검지기를 비교하여 지자기 검지기의 대체 유용성을 평가한 선행연구에서 청라지하차도의 지자기 검지기와 루프 검지기의 성능이 동일하다는 내용을 바탕으로 진행하였다(Ko et al., 2017).



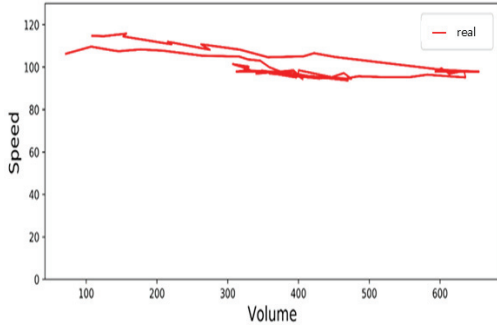
<Fig. 6> Simulation detector position

#### 4. 시뮬레이션 정산 및 검증

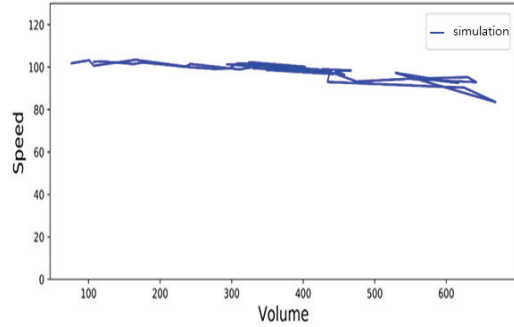
정산(calibration)이란 구축된 네트워크를 실제 네트워크와 가깝게 조정하는 과정을 말한다(Yun et al., 2013). 본 연구에서는 Desired Speed 및 차량들의 가속속도를 조정하여 정산 작업을 진행하였다. 다음으로 시뮬레이션 네트워크에 구축한 가상의 검지기의 30초 간격 속도, 점유율, 교통량 자료를 통해 실제 검지기와 비교하여 검증(validation)하였다. <Fig. 7>과 <Fig. 8>에서 보인 바와 같이 실 데이터와 시뮬레이션 데이터의



교통량-속도 그래프를 이용하여 조정된 시뮬레이션 네트워크가 유의한 정산 결과를 가지는지 파악하였다. 검증 결과 시뮬레이션과 실 데이터 모두 교통량이 적어 강제류까지 가지 않았으며 교통량에 따라 속도가 떨어지는 추세가 비슷하였다.



<Fig. 7> Real data traffic volume-speed graph



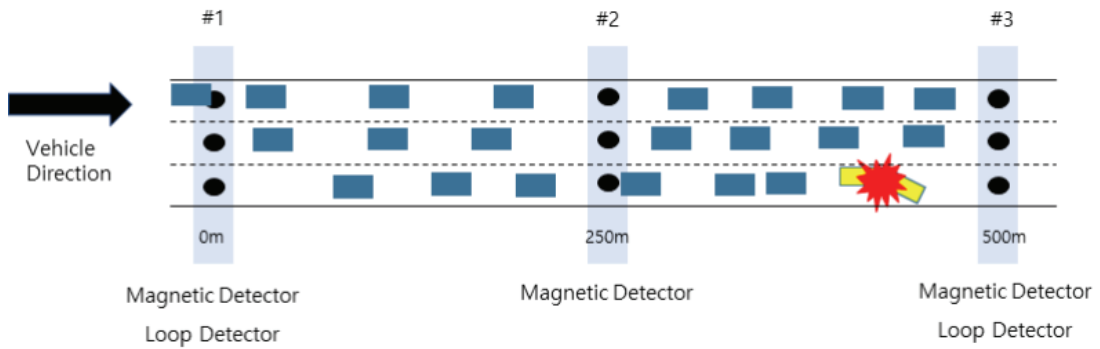
<Fig. 8> Simulation data traffic volume-speed graph

## 5. 돌발상황 구현

현실과 유사하게 구현할 수 있는 VISSIM은 돌발상황과 같은 비이상적인 교통 현상을 발생시키는 기능을 가진 옵션이 존재하지 않는다. 따라서 정체 및 차로 차단 등을 구현하기 위하여 링크의 중간에 VISSIM에서 제공하는 기능인 ‘parking lot’을 생성하거나 신호를 추가하는 방법을 통하여 가상의 돌발상황을 구현할 수 있다. 본 연구에서는 신호를 추가하여 돌발상황을 구현하였으며, 차단 차로, 차단 지점, 차단 시간을 설정하여 시나리오를 개발하였다.

## 6. 돌발상황 시나리오 구축

정산된 네트워크를 이용해 가상의 돌발상황을 포함한 다양한 시나리오를 설정하였다. 기존 논문(Kim et al.,)을 참고하여 돌발상황에 영향을 미칠 것으로 예상되는 네 가지 변수인 교통량, 차단된 차로 수, 돌발상황 지점, 돌발상황 지속시간에 따라 개별 시나리오를 개발하였다. 교통량은 실제 검지기에서 수집한 자료를 바탕으로 2,600대/시/방향을 기준으로 50%인 1,300대/시/방향, 150%인 3,900대/시/방향 총 세 가지 경우에 돌발상황을 고려하였다. 차단 차로는 1차로 또는 연속된 2차로 두 가지를 설정하였다. 돌발상황 지점은 상류부 검지기로부터 300m, 400m, 500m 총 세 가지 지점을 선정하였다. 돌발상황 지속시간은 15분, 30분, 45분 세 가지 경우를 고려하였다. <Table 2>에서 제시된 바와 같이 네 가지 변수의 조합을 통해 돌발상황이 포함된 54개와 돌발상황이 포함되지 않은 3개, 총 57개의 시나리오를 설정하였다. 설정된 시나리오를 바탕으로 연구를 진행하였으며 3개의 가상의 검지기를 통하여 30초 단위의 교통량, 속도, 점유율 데이터를 생성할 수 있으며, <Fig. 9>은 한 예시로서 400m 지점, 한 개 차로에서 발생한 시나리오를 보여주고 있다. 모든 시나리오는 루프 검지기가 매설된 #1과 #3을 활용하는 500m 검지 간격 분석과 지자기 검지기가 매설된 #2와 #3을 활용하는 250m 검지 간격 분석을 실시하여 알고리즘의 성능 비교가 가능하다. 기존 연구에서 검지기 설치 간격이 돌발상황 검지 성능에 밀접한 관계가 있음을 확인하였으며, 이를 바탕으로 알고리즘의 성능을 비교하기 위해 상류부 검지기와 하류부 검지기 간격을 250m와 500m일 때로 나누어 설정된 시나리오를 통해 결과를 도출하고자 하였다(KOTI., 2010).



<Fig. 9> Simulation incident scenario example (section 2)

<Table 2> Simulation incident situation scenario

Incident Location	Traffic Volume	Lane	Time (minute)	Incident Location	Traffic Volume	Lane	Time (minute)	Incident Location	Traffic Volume	Lane	Time (minute)
Section 1 (300m)	1,300 vph	1	15	Section 2 (400m)	1,300 vph	1	15	Section 3 (500m)	1,300 vph	1	15
			30				30				30
			45				45				45
		2	15			2	15			2	15
			30				30				30
			45				45				45
	2,600 vph	1	15		2,600 vph	1	15		2,600 vph	1	15
			30				30				30
			45				45				45
		2	15			2	15			2	15
			30				30				30
			45				45				45
3,900 vph	1	15	3,900 vph	1	15	3,900 vph	1	15			
		30			30			30			
		45			45			45			
	2	15		2	15		2	15			
		30			30			30			
		45			45			45			

#### IV. 나이프 베이스 분류기를 활용한 돌발 검지 알고리즘 개발

##### 1. 기존 돌발 검지 알고리즘 선정 및 구현

많은 돌발 검지 알고리즘 기법들이 있으나 실제 시스템으로 운영되어 활용되고 있는 알고리즘으로는

APID, DES, DELOS, McMaster 알고리즘 등이 있다(Korea Road Traffic Authority, 2018). 서울시에서 발간한 보고서에 따르면 2002년 5월 서울 내부순환도로에 적용한 McMaster, APID, AIDA 알고리즘의 사고 및 고장차량에 의한 돌발상황에 대한 검지율 및 오검지율 평가 결과 가장 우수하게 평가되었다. 본 연구에서는 교통류 모형과 수학적 모델에 근거한 알고리즘인 McMaster 알고리즘을 기존 돌발상황 검지 알고리즘으로 선정하고 구현하였다(Seoul, 2002). McMaster 알고리즘을 구현하기에 앞서 각 검지기에서 사용될 파라메타를 산정해야 한다. 실제 1주일 정도의 off-line 자료를 수집하고, 독립변수는 점유율이고 종속변수는 교통량으로 회귀분석을 실시한 LUD 함수를 산출하였다. 표준이차함수인 LUD 함수를 통하여 상류부 검지기의 혼잡을 판단할 수 있는데 혼잡이 4주기 이상 반복된다면 하류부 검지기를 통해 해당 구간의 혼잡의 원인을 찾을 수 있다. 알고리즘은 Python 3.6을 이용하여 구현하였으며 돌발검지 간격 조정에 따라 알고리즘의 성능을 비교하기 위해 상류부 검지기와 하류부 검지기 간격을 250m와 500m일 때로 나누어 설정된 시나리오를 통해 결과를 도출하였다.

## 2. 머신러닝 알고리즘 선정 및 구현

돌발상황 검지는 교통변수를 이용하여 돌발의 유무를 분류하는 것이기 때문에 머신러닝의 지도학습 중 분류 기법을 사용하는 것이 바람직하다. 분류 기법 중 활발히 이용되고 있는 나이브 베이스 분류기는 계산 과정이 간단하지만 성능이 매우 우수하여 다양한 분야에서 연구되고 있다. 따라서 본 연구는 나이브 베이스 분류기를 활용하여 새로운 돌발 검지 알고리즘을 개발하였다. 나이브 베이스 분류기를 활용한 돌발 검지 알고리즘에 사용되는 독립변수는 상류부 및 하류부 검지기의 30초 단위 속도, 교통량, 점유율 자료를 사용하였고 종속변수로는 사고발생 여부를 사용하였다. 머신러닝을 활용하기 위해서는 학습을 위한 데이터와 검증을 위한 데이터로 구분하는 것이 필요하다. 랜덤 샘플링 기법을 활용하여 총 57가지의 시나리오 중 46개(약 80%)의 시나리오를 학습 데이터로 선정하였고, 나머지 11개(약 20%)를 검증 데이터로 분류하였다. 기존 알고리즘의 성능 평가와 마찬가지로 돌발검지 간격 조정에 따라 성능을 비교하고자 검지 간격이 250m일 때와 500m일 때로 구분하여 나이브 베이스 분류기 알고리즘을 통해 각각 학습하였다. 나이브 베이스 분류기 기법은 가장 대표적인 가우시안 정규 분포 나이브 베이스 분류기를 이용하였다. 학습 결과 검지 간격이 250m일 때는 98.53%, 500m일 때는 97.66%의 우수한 나이브 베이스 분류기 모형이 구축되었다.

## 3. 알고리즘 성능 지표 선정

구현한 돌발 검지 알고리즘을 평가하기 위해서는 성능평가 산출이 필요하다. 일반적으로 사용되는 성능평가 지표로 검지율(detection rate, DR), 오검지율(false alarm rate, FAR), 평균검지시간(mean detection time, MDT)을 사용한다(Lee, 2010). 본 연구는 시뮬레이션 기반이기 때문에 정확한 돌발 발생 시간을 정의할 수 없기에 돌발 발생 후 검지하는데 소요되는 평균적인 시간을 정의하는 평균검지시간을 제외한 검지율과 오검지율을 통하여 알고리즘의 성능을 평가하였다.

$$\cdot \text{검지율}(DR) = \frac{\text{검지된 유고 건수}}{\text{전체 유고 건수}} \times 100 \quad (4)$$

$$\cdot \text{오검지율}(FAR) = \frac{\text{유고라고 판단하였으나, 실제로는 유고가 아닌 건수}}{\text{유고라고 판단한 총 건수}} \times 100 \quad (5)$$

## V. 알고리즘 성능 비교

### 1. 돌발검지 간격에 따른 성능 평가

각 알고리즘의 성능을 평가하기 위해 상류부 검지기로부터 하류부 검지기 간의 거리 간격 조정을 통해 성능을 평가하였다. 성능 평가 결과를 나타낸 <Table 3>을 살펴보면, 기존 McMaster 알고리즘과 개발한 나이브 베이즈 분류기 알고리즘 모두 돌발검지 간격이 넓어질수록 검지력이 떨어지는 것을 확인할 수 있었지만 개발한 나이브 베이즈 분류기 알고리즘이 기존 McMaster 알고리즘에 비해 돌발검지 간격에 따른 부정적인 영향을 적게 받았음을 확인할 수 있다. 특히, 기존 McMaster 알고리즘을 활용하면 교통량이 가장 적은 1,300대/시이며 돌발검지 간격이 500m일 때의 돌발을 검지 할 수 없었지만, 개발한 나이브 베이즈 분류기 알고리즘은 우수한 검지가 가능한 것을 확인할 수 있었다. 기존 McMaster 알고리즘은 상류부 검지기의 혼잡이 판단되어야 프로세스가 진행이 된다. 그렇기 때문에 돌발상황과 상류부 검지기의 거리가 멀어질수록 검지율이 떨어지는 것으로 판단된다. 그에 비해 개발한 나이브 베이즈 분류기 알고리즘은 상류부 검지기와 하류부 검지기의 교통량, 속도, 점유율 데이터인 총 6개의 변수를 활용하기 때문에 상류부 검지기와 하류부 검지기의 교통 상태를 동시에 고려하는 알고리즘이다. 따라서 McMaster 알고리즘보다 돌발검지 간격에 따른 성능평가의 저하가 줄어드는 것으로 판단된다.

### 2. 알고리즘에 따른 성능 평가

마찬가지로 <Table 3>을 살펴보면 돌발검지 간격에 관계없이, 개발한 나이브 베이즈 분류기를 이용한 알고리즘이 기존 McMaster 알고리즘에 비해 매우 우수한 검지율을 가진 것을 보였다. 기존 McMaster 알고리즘은 4주기에 거쳐 혼잡이 있다고 판단이 되면 돌발상황이라고 판단을 내리지만 그에 비해 나이브 베이즈 분류기 알고리즘은 1주기만으로 돌발상황을 판단할 수 있기 때문에 검지율이 높고 빠른 검지가 가능하다. 하지만 4주기를 통해 신중한 검지가 가능한 기존 McMaster 알고리즘에 비해 개발한 나이브 베이즈 분류기 알고리즘의 오검지율이 다소 높아지는 것을 확인할 수 있었다.

<Table 3> Algorithm performance comparison

Incident Location	Traffic Volume (vehicle/hour)	Lane	Time (minute)	VDS distance : 250m				VDS distance : 500m			
				DR(%)		FAR(%)		DR(%)		FAR(%)	
				McMaster	Naive Bayes	McMaster	Naive Bayes	McMaster	Naive Bayes	McMaster	Naive Bayes
Section 1 (300m)	1,300	1	15	90.00	100	3.57	3.23	0	83.33	-	3.85
	1,300	2	30	96.67	100	2.25	3.23	4.44	75.56	0	5.56
	2,600	2	45	95.00	100	3.39	7.69	85.00	96.67	7.27	6.85
Section 2 (400m)	1,300	1	15	95.00	100	0	1.64	0	98.33	-	3.28
	2,600	2	30	90.00	100	10.00	14.29	66.67	96.67	16.67	14.71
	3,900	1	45	94.44	100	1.16	5.26	78.89	98.89	3.61	8.25
Section 3 (500m)	1,300	2	15	63.33	98.33	5.00	3.28	0	98.33	100	4.84
	2,600	2	30	92.22	100	3.49	6.25	85.56	100	6.10	6.25
	3,900	2	45	86.67	100	10.34	11.76	76.67	100	17.86	18.92
No Incident	3,900	2	15	95.56	100	3.37	12.62	92.22	100	6.74	5.26
No Incident	2,600	-	30	100	100	0	0	100	100	0	0

## VI. 결 론

### 1. 결 론

본 연구에서는 실질적인 현장 적용에 어려움을 겪고 있는 기존의 돌발 검지 알고리즘의 문제점을 해결할 수 있는 머신러닝을 활용한 새로운 돌발 검지 알고리즘을 개발하였다.

머신러닝 기법을 활용한 새로운 돌발 검지 알고리즘을 개발하기 위해서 본 연구는 돌발상황에 영향을 줄 수 있는 변수를 선정하여 여러 조합을 통해 57개의 시나리오를 설정하였고 VISSIM ver.10.0 프로그램을 사용하여 교통류의 변수자료(교통량, 속도, 점유율)를 수집하였다. 다음으로 대표적인 기존 돌발 검지 알고리즘인 McMaster 알고리즘을 Python 3.6을 이용해 구현하고 돌발검지 간격을 250m, 500m로 나누어 성능지표를 산출하였다. 산출 결과 McMaster 알고리즘은 교통량이 적을수록, 차단시간이 짧을수록, 차단차로가 적을수록, 돌발상황이 상류부 검지기와 멀어질수록 검지 성능이 떨어졌으며, 돌발 검지 간격은 넓을수록 검지율이 떨어지는 것을 볼 수 있었다. 특히 1,300대/시이고 돌발 검지 간격이 500m인 시나리오에서 McMaster 알고리즘은 돌발 상황을 검지하지 못 하는 것이 확인되어졌다. 따라서 본 연구는 이러한 기존 돌발 검지 알고리즘의 문제점을 해결하고자 머신러닝을 활용한 돌발 검지 알고리즘을 개발하였다. 머신러닝 기법으로는 단순하지만 우수한 성능을 지닌 대표적인 지도학습인 나이브 베이스 분류기를 선정하였다. 설정된 시나리오를 바탕으로 학습 데이터와 검증 데이터로 구분하고 기존 알고리즘과 마찬가지로 돌발검지 간격을 250m와 500m로 나누어 나이브 베이스 분류기 알고리즘의 학습을 수행하였다. 학습결과 250m 돌발검지 간격은 98.53%, 500m는 97.66%의 우수한 정확도를 가진 알고리즘이 구축되었다. 개발한 알고리즘을 검증 데이터로 검증하였고 기존 McMaster 알고리즘에서 활용한 성능평가지표를 통해 McMaster 알고리즘과 나이브 베이스 분류기 알고리즘을 비교 분석하였다. 결과적으로 개발한 나이브 베이스 분류기 알고리즘은 기존 McMaster 알고리즘보다 우수한 검지율을 가진 것을 확인하였다. McMaster 알고리즘은 4주기에 거쳐 혼잡의 원인을 찾는 프로세스를 진행하는 반면 개발한 나이브 베이스 분류기 알고리즘은 1주기만으로 돌발상황을 판단할 수 있기 때문에 검지율이 높아지는 것으로 판단된다. 하지만 개발한 나이브 베이스 분류기 알고리즘의 검지율이 우수해지는 만큼 오검지율 또한 높아지는 것을 알 수 있었다. 분석하는 상류부 검지기로부터 하류부 검지기간의 거리가 늘어날수록 검지율의 차이가 발생하였는데, 기존 McMaster 알고리즘에 비해 개발한 나이브 베이스 분류기 알고리즘은 검지율의 하락폭이 작았다. 기존 McMaster 알고리즘은 상류부 검지기의 혼잡이 판단되어야 프로세스가 진행이 된다. 그렇기 때문에 돌발상황과 상류부 검지기의 거리가 멀어질수록 검지율이 떨어지는 것으로 판단된다. 그에 비해 개발한 나이브 베이스 분류기 알고리즘은 상류부 검지기와 하류부 검지기의 교통량, 속도, 점유율 데이터인 총 6개의 변수를 활용하기 때문에 상류부 검지기와 하류부 검지기의 교통 상태를 동시에 고려하는 알고리즘이다. 따라서 McMaster 알고리즘보다 돌발검지 간격에 따른 성능평가의 저하가 줄어드는 것으로 판단된다.

본 연구를 통해서 개발한 나이브 베이스 분류기 알고리즘은 기존 McMaster 알고리즘보다 우수한 검지율을 가지므로 기존의 알고리즘보다 효율적으로 돌발을 파악할 수 있는 알고리즘이 될 수 있다는 가능성을 확인할 수 있었다. 하지만 오검지율 또한 증가하였기에 알고리즘의 보완을 거쳐야 할 필요가 있을 것으로 판단된다. 또한 본 연구는 시뮬레이션 자료를 활용한 분석이기 때문에 현장에 적용하기 위해서는 실측 자료를 이용한 다양한 돌발 데이터의 확보를 통해 알고리즘의 검증이 추가로 이루어져야 할 것이다.

## 2. 향후 연구과제

본 연구에서 개발한 모형을 실제 돌발 검지 시스템에 적용하여 현장에 활용하기 위해서는 다음과 같은 연구가 추진되어야 할 것이다. 먼저 본 연구는 고속도로의 본선구간만을 실험대상으로 적용하였기 때문에 고속도로 진입, 진출 램프 등 본선 구간 외의 구간을 고려한 알고리즘의 개발이 요구되며 돌발상황에 영향을 줄 것이라 판단되는 변수들이 유의한지에 대한 연구가 필요할 것으로 보인다. 또한 본 연구에서는 머신러닝 기법을 활용할 때 상류부 및 하류부 검지기의 교통량, 속도, 점유율 자료를 독립변수로서 사용하였지만 교통량, 속도, 점유율의 상관관계를 반영할 수 있는 머신러닝 알고리즘이 개발된다면 본 연구의 한계를 극복할 수 있을 것이라 판단된다.

## ACKNOWLEDGEMENTS

이 논문은 2015년도 정부(교육과학기술부)의 재원으로 한국연구재단(NRF)의 기초연구사업 지원을 받아 수행된 것임(2015R1A1A1A05028008).

## REFERENCES

- Cho S. J. and Kang S. H.(2016), "Industrial Applications of Machine Learning," *Industrial Engineering Magazine*, vol. 23, no. 2, pp.34-38.
- Cho Y. T.(2016), *Development of an Estimating Method for the Parameters of Volume Delay Functions Using a Genetic Algorithm*, Doctoral thesis, Ajou University.
- Deniz O. and Celikoglu H. B.(2011), "Overview to some existing incident detection algorithms: a comparative evaluation," *Procedia-Social and Behavioral Sciences*, vol. 2, pp.153-168.
- Gailbo, <http://www.gailbo.com>, 2018.07.21.
- Hall F. L., Shi Y. and Atala G.(1991), "On-Line Testing of the McMaster Incident Detection Algorithm Under Recurrent Congestion," *Transportation Research Record*, no. 1394, pp.1-7.
- Jang M. S.(2018), *Classifier Integration Model with Naive Bayes Classifier*, Master thesis, Myongji University.
- Jeong H. R.(2018), *Study on Prediction of Severities of Rental Car Traffic Accidents Using Ordinal Logit Model and Naive Bayes Classifier*, Master thesis, Ajou University.
- Kim S. G. and Kim Y. C.(2006), "Development of Incident Detection Algorithm Using Speed-Occupancy Relationship," *Seoul City Studies*, vol. 7 no. 3, pp.235-249.
- Kim S. G., Kim C. H. and Kim Y. C.(2008), "A Sensitivity Test of McMaster Algorithm Depending on the Changes in Parameter Values," *Seoul City Studies*, vol. 9, no. 2, pp.121-131.
- Kim S. H.(1999), "A Development of Neural Networks-based Incident Detection Model," *Journal of Engineering & Technology, Hanyang University*, vol. 8, no. 1, pp.187-196.
- Kim S. H., Kim J. H., Doo M. S., Byun W. H., Sin A. G., Yun Y. H., Lee Y. T., Lee J. B. and Chun S. H.(2016), *Road Traffic ITS Theory and Design*, Cheongmoongak, pp.108-109.
- Ko S. G., Kwon C. W., Kim T. S., Cho W. S. and Yun I. S.(2017), "Study on Performance Evaluation

- of Magnetometer Detectors,” *Journal of The Korea Institute of Intelligent Transport Systems*, vol. 2017, no. 10, pp.63-67.
- Korea Road Traffic Authority(2018), *Traffic accident statistics*.
- Korea Transport Institute(2010), *Study on the Collection and Provision of Traffic Information on Highway*.
- KOROAD(2014), “Central Traffic Information Center Traffic Information,” *KOROAD Traffic Science Institute*, pp.8-17.
- Lee D. H.(2017), *A study on Hotspot Identification using Machine Learning*, Master thesis, University of Seoul.
- Lee J., Kim B. K. and Kim S. H.(2000), “A Study on the Application of the Interrupted Traffic Flow Incident Detection Algorithm using Fixed Detector,” *Journal of Korean society of Transportation*, vol. 4, pp.33-36.
- Lee S. H. and Kang H. C.(2003), “Development and Evaluation of New Incident Detection Algorithm,” *Journal of Korean society of Transportation*, vol. 44, pp.1-6.
- Mak C. L. and Fan H. S.(2005), “Transferability of expressway incident detection algorithms to Singapore and Melbourne,” *Journal of transportation engineering*, vol. 131, no. 2, pp.101-111.
- Philip H. M., Joseph K. L. and Kam W.(1991), “Incident Detection Algorithms for COMPASS An Advanced Traffic Management System,” *Vehicle Navigation and Information Systems Conference*, vol. 2, pp.295-310.
- Rakha H., Hellinga B. and Van Aerde M.(1999), “Testbed for Evaluating Automatic Incident Detection Algorithms,” In *Intelligent Transportation System Safety and Security Conference*.
- Sun D., Zhang C., Zhao M., Zheng L., Liu W.(2017), “Traffic Congestion Pattern Detection Using an Improved McMaster Algorithm,” *2017 29th Chinese Control And Decision Conference*, pp.2814-2819.
- Yang C., Miao Z. and Dongyuan Y.(2015), “Automatic Incident Detection for Urban Expressways Based on Segment Traffic Flow Density,” *Journal of Intelligent Transportation Systems*, vol. 19, pp.205-213.
- Yun I. S., Han E., Lee C. K., Rho J. H., Lee S. J. and Kim S. B.(2013), “Mobility and Safety Evaluation Methodology for the Locations of Hi-PASS Lanes Using a Microscopic Traffic Simulation Tool,” *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 12, no. 1, pp.98-108.