

Normalization of Input Vectors in Deep Belief Networks (DBNs) for Automatic Incident Detection

Daehyon Kim¹⁾

Abstract

Traffic incidents have a serious negative impact on safety and traffic flow, and fast accurate automatic incident detection on freeways is a major theme in transportation engineering. Therefore, various types of AID (Automated Incident Detection) algorithms have been proposed for more accurate and rapid incident detection, and Artificial Neural Network models have provided significantly improved performance in terms of detection and false alarm rates. Recently, Deep Neural Networks (DNNs) has received much attention due to its excellent performance and was also used for automatic incident detection on highways. However, in learning algorithms such as Backpropagation and SVMs(Support Vector Machines), the prediction performance is known to be highly depend on the input vector characteristics. The purpose of this study is to examine whether the input detection performance of DNNs differs according to the normalization method of the input vector and to verify how sensitive it is to the method. Furthermore, the best way to normalize the input vector of the DNNs model has been proposed in order to obtain the best performance in terms of DR (Detection Rate) and FAR (False Alarm Rate) in AID (Automatic Incident Detection).

Keywords : AID (Automatic Incident Detection), Artificial Neural Network, Deep Neural Networks (DNNs), Normalization, Input Vectors

1. Introduction

Traffic incidents on the highway can interfere with normal traffic flow and cause serious traffic incidents. Therefore, the main topic of the Intelligent Transportation Systems (ITS) has been to detect and eliminate incidents on highways quickly and accurately.

There have been many studies for automatic incident detection, and a number of algorithms have been developed over the last 30 years[1-3]. Recently, ANNs, which are easy to implement a real-time detection system, are the most attractive algorithms for detecting incidents on freeways, because they can provide fast, reliable and fault tolerant operations[4-7].

While there are many learning algorithms, the Backpropagation model, the most common

Received(October 09, 2018), Review Result(1st: October 26, 2018, 2nd: November 26, 2018), Accepted(December 10, 2018)

1) (Professor) 59626 Division of Ocean Civil Engineering, Chonnam National University, Yeosu-Si, Jeollanam-Do, Korea

email: daehyon@chonnam.ac.kr

neural network model, has been used for highway auto-incident detection[8][9]. Recently, however, DNNs (Deep Neural Networks) has demonstrated remarkable performance in extracting and expressing high levels of patterns in various pattern recognition classifications, and has proved an excellent detection power in freeway incident detection[10].

However, there are a few factors to decide before using the DNNs model, such as network architecture and normalization methods. The input vector properties of DNNs model control the capability of the network to provide an accurate learning of the input-output relationships and are related to the normalization method for input vectors.

In this paper, two different normalization methods are studied to find the optimal normalization method. Experimental results show that the performance learning algorithm related to incident detection relies heavily on the normalization method. The best normalization method of DNNs model for freeway incident detection is proposed through experiments to obtain the best performance in terms of DR (Detection Rate) and FAR (False Alarm Rate).

2. Deep Neural Networks (DNNs) Model

Kim[10] compared the performance of DNNs and Backpropagation model on test data sets via experiments with real world incident and incident-free data. Experimental results showed that Deep Belief Networks (DBNs), which is a type of deep neural networks, is better than Backpropagation in freeway incident detection.

Deep Belief Networks (DBNs) algorithm is described in detail in Kim[10], and the main contents of the following algorithm are described in the paper. The Deep Belief Networks (DBNs) require multiple hidden layers with high number of hidden units to learn and extract features from raw data. Compared with Backpropagation, DBNs can use unlabeled data to pre-train a multi-layer generative model in unsupervised learning with Restricted Boltzmann Machines (RBMs)[11]. The training of the DBNs is defined by two steps: pre-training and fine-tuning. The pre-training will train each Restricted Boltzmann Machine (RBM) independently, such that the output of the lower RBM is provided as input for the next higher-level RBM and so forth. The fine-tuning process will be performed by Backpropagation[10]. The algorithm of Restricted Boltzmann Machines (RBMs) can be summarized as follows[10]:

Step 1: Conduct following process for $n = 1 \cdots N$ (number of data samples)

A. set visible states to n^{th} data sample, i.e., $v^{(n)} = x^{(n)}$ where $x^{(n)}$ is data sample

B. compute hidden probability $q_j^{(n,0)}$ by using Eq. (1) for all j (number of hidden units)

$$q_j: p(h_j = 1|v) = \sigma(b_j + \sum_i v_i w_{ij}), \text{ , where } \sigma(x) = 1/(1 + \exp(-x)) \quad (1)$$

C. sample hidden state $h_j^{(n,0)} \in \{0,1\}$ from $q_j^{(n,0)}$ for all j .

Step 2: Calculate reconstructed visible probability for all (number of visible units) by using Eq. (2)

$$p_i : p(v_i = 1|h) = \sigma(a_i + \sum_j h_j w_{ij}), \text{ , where } \sigma(x) = 1/(1 + \exp(-x)) \quad (2)$$

Step 3: Compute hidden probability by using Eq. (1) for all (number of hidden units)

Step 4: Compute expectations over data distribution for all i and j by using Eq. (3)

$$\langle v_i q_j \rangle_{data} = \frac{1}{N} \sum_n v_i^{(n,0)} q_j^{(n,0)}, \langle v_i \rangle_{data} = \frac{1}{N} \sum_n v_i^{(n,0)}, \langle q_j \rangle_{data} = \frac{1}{N} \sum_n q_j^{(n,0)} \quad (3)$$

Step 5: Compute expectations over reconstructions for all i and j by using Eq. (4)

$$\langle v_i q_j \rangle_{recon} = \frac{1}{N} \sum_n v_i^{(n,1)} q_j^{(n,1)}, \langle v_i \rangle_{recon} = \frac{1}{N} \sum_n v_i^{(n,1)}, \langle q_j \rangle_{recon} = \frac{1}{N} \sum_n q_j^{(n,1)} \quad (4)$$

Step 6: Compute changes in weights and biases for all i and j by using Eq. (5)

$$\Delta b_j = \varepsilon (\langle v_i q_j \rangle_{data} - \langle v_i q_j \rangle_{recon}), \Delta a_i = \varepsilon (\langle v_i \rangle_{data} - \langle v_i \rangle_{recon}), \Delta b_j = \varepsilon (\langle q_j \rangle_{data} - \langle q_j \rangle_{recon}) \quad (5)$$

, where ε is learning rate, a_i is bias of i^{th} visible unit and b_j is bias of j^{th} hidden unit.

Step 7: Apply changes in weights and biases for all i and j by using Eq. (6)

$$w_{ij} = w_{ij} + \Delta w_{ij}, a_i = a_i + \Delta a_i, b_j = b_j + \Delta b_j \quad (6)$$

3. Normalization Methods for Input and Output Vectors

For the application of Deep Belief Networks (DBNs), each input data should be normalized between 0 and 1. Kim and Kim et al.[12-14] have shown that it is very important to find appropriate normalization methods to obtain the best learning performance in learning models such as Backpropagation and Support Vector Machines(SVM). However, the best normalization method is highly dependent on the learning model. Therefore, the best normalization should be determined for incident detection in the Deep Belief Networks (DBNs).

A detector-based incident detection algorithm typically uses three parameters, such as traffic

volume, occupancy and speed measured at upstream and downstream detectors to determine incidents on freeways[13][14]. The values of these three parameters are higher than 1.0 and the input vectors must be normalized for the application of Deep Neural Networks. In the case of automatic incident detection using a neural network model, the input vector is the normalized value of the above three parameters obtained from the detectors during multiple time fragments. However, other types of normalization method may be proposed for possible encoding and decoding formats of input parameters.

In this paper, two methods are considered to compare the detection performance of DR (Detection Rate) and FAL (False Alarm Rate), and a better normalization method for automatic incident detection problem is proposed. The first method, called FORM 1, uses the minimum and maximum values of the entire input vector for each input parameter data. Input parameter data means input with three parameters; traffic volume, speed and occupancy, from upstream to downstream detectors during given time. This normalization method is the most common way to normalize an input vector when applying a neural network model for automatic incident detection. The first normalization formula is given by Eq. (7).

$$y_{pi} = \frac{x_{pi} - x_{p \min}}{x_{p \max} - x_{p \min}} \quad (7)$$

where $x_{p \max} = \max(x_{pi}; i = 1, \dots, m)$, $x_{p \min} = \min(x_{pi}; i = 1, K, m)$ and $p(p = 1, \dots, P)$ are the input parameter data. In Equation(7), y_{pi} denotes the normalized value of the unit i of input vector $x_p = (x_{p1}, x_{p2}, x_{p3}, \dots, x_{pm})$, and x_{pi} denotes the original value of the input unit i in the pattern p . By this normalization method, each value of the input vectors is the range $[0, 1]$.

On the other hand, the second method called FORM 2 is based on the Detector Station (D.S.) proposed by Kim et al.[14]. Unlike previous normal incident detection models, this study added location information. The normalization formula for each given detector with location information is given by the following equation (8).

$${}^{ld}y_{pi} = \frac{{}^{ld}x_{pi} - {}^{ld}x_{p \min}}{{}^{ld}x_{p \max} - {}^{ld}x_{p \min}} \quad (8)$$

, where ${}^{ld}x_{p \max} = \max({}^{ld}x_{pi}; l = 1, 2, 3, \dots, L, d = 1, 2, 3 \text{ and } p = 1, K, P)$,

${}^{ld}x_{p \min} = \min({}^{ld}x_{pi}; l = 1, 2, 3, \dots, L, d = 1, 2, 3 \text{ and } p = 1, K, P)$, $p(p = 1, \dots, P)$ are the input patterns, l is location information, and d is the detector station. To apply the DNNs

model, each data obtained from three detector stations was normalized and used as input data. Three stations can be used to experiment with different characteristics of the road and detector, such as the distance between two adjacent sensor stations, the type of the road construction, and detector faults. In this study, the output vector is defined as $y = [y_1 \ y_2]$, where y_1 is an incident pattern and y_2 is a non-incident pattern.

4. Experiments and Results

4.1 Input and Output Data Sets

This study used the same experimental data that Kim and Kim et al.[10][13][14] used to study for automated incident detection. Traffic information required for the experiment was collected using the image processing device installed on the Seoul Expressway. The image processors installed on the expressway collect three traffic parameters, such as traffic volume, speed and occupancy for all lanes on the expressway. In case of an incident, the volume of traffic decreases, the speed decreases, and the occupancy increases. On the other hand, as traffic decreases in case of a non-incident, the speed increases and the occupancy decreases. This study identifies these patterns through learning with three parameters, traffic volume(vehicles per hour), speed(expressed in km/h) and occupancy(expressed in sec).

A total of 231 incident and non-event data were used for the experiment: 31 incident data and 200 incident-free data. To overcome the lack of incident data, the following methods were used. A total of 31 incident data and 200 non-incident data sets were randomly split into two subsets, data sets I and II. One is for training and the other is for testing. The first attempt was to learn about data set I to evaluate network performance for dataset II. In the second trial, learning was performed in data set II and model performance was evaluated in data set I. This cross-validation can improve the reliability of the results of the experiment.

Experimental data includes traffic volume, speed, and occupancy for each of the three lanes at three locations every 30 seconds. Traffic volume, speed and occupancy collected from three detector stations, including upstream and downstream detector stations where incidents occurred, were used in this study. Also, the input vector uses each parameter data of each lane during four time cycles, i.e., the current cycle, t , and three previous cycles, $(t-1)$, $(t-2)$ and $(t-3)$. Therefore, the input vector x is composed of 72 units for three lanes, three places, three traffic parameters, and four intervals. As described above, in the incident detection problem,

the output vector is defined as two different cases as follows. That is, $y_1 = [0 \ 1]$, $y_2 = [1 \ 0]$, where y_1 is defined as an incident and y_2 is defined as a non-incident.

4.2 Network Topology

To build Deep Network Networks (DNNs), the network topology must be determined before experimentation. Deep Neural Networks in general use many hidden layers because they are distinguished from single hidden layer neural networks by depth. As a result, DBNs used three or more hidden layers in many studies[15][16].

This study used three network topologies[10]; i.e. 72(input units) - 144 (hidden units) - 2(output units), 72(input units) - 144(hidden units) - 288(hidden units) - 2 (output units), and 72(input units) - 144(hidden units) - 288(hidden units) - 576(hidden units) - 2(output units). The number of hidden units was determined with the following rule; $2x$ (1st hidden layer) - $4x$ (2nd hidden layer) - $8x$ (3rd hidden layer), where x is the number of input units.

4.3 Implementation and Results

In order to evaluate and compare the performance of two normalization models, FORM 1 and FORM 2, 30 trials were implemented with the same values for the parameters in the same network, but the initial weights of each trial were initialized to different random values each time. Test results were statistically analyzed by ANOVA using a post-hoc test - Tukey HSD (honest significant difference) - at a significance level of $P = 0.05$, in order to assess the statistical significance.

In this paper, two MOEs (measures of effectiveness), DR (Detection Rate) and FAR (False Alarm Rate), were used to evaluate the performance of incident detection. The DR (Detection Rate) is the number of incidents detected by the algorithm divided by the total number of incidents during the recording time. The FAR (False Alarm Rate) is defined as the number of incident-free intervals that give false alarms divided by the total number of incident-free intervals.

[Table 1] and [Fig. 1] show the results of the two different normalization methods, FORM 1 and FORM 2, in various topologies of the Deep Brief Networks(DBNs). The predicted error rate in the table represents the average recognition error rate for different experiments using a new random initial weight (a total of 30 trials) in the test set(the number of error/the number of test sets). As shown in the table, there was no learning failure in the Deep Brief Network

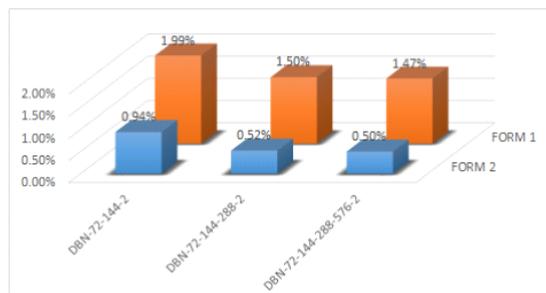
model. This means that the Deep Brief Network model is more effective for complex network architectures than the Backpropagation model.

Experimental results showed that the detection rate ranged from 73.97% to 88.48% and the FAR from 0.50% to 1.99%. The second normalization method, FORM 2, has a higher detection rate and lower false alarm rate than FORM 1, the first normalization method. The results of this experiment are similar to those of the SVM for highway incident detection[14]. In general, the optimal method for the normalization of the input vectors may vary according to the learning model. However, the experimental results of this study show that the performance of DBNs-based automatic incident detection system greatly improves by using the normalization method of D.S.-by-D.S. with location information.

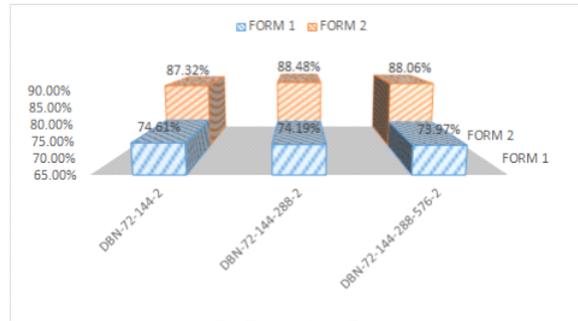
Also, as shown in the table, there is a statistically significant difference in the performance of the two methods FORM 1 and FORM 2 according to the ANOVA test results. This means that the performance of FORM 2 is better than FORM 1 and that FORM 2 is suitable for the normalization of input vectors in freeway incident detection.

[Table 1] Predictive Performance of Normalization Methods in Different Architectures of DBNs

Network Topology		72-144-2		72-144-288-2		72-144-288-576-2	
Neural Network model		FORM 1	FORM 2	FORM 1	FORM 2	FORM 1	FORM 2
Incident Data	Prediction Error Rate	7.87/31	3.93/31	8.0/31	3.57/31	8.07/31	3.70/31
	Variance	0.326437	0.064368	0.0	0.254023	0.064368	0.217241
Incident - Free Data	Prediction Error Rate	3.97/200	1.87/200	3.0/200	1.03/200	2.93/200	1.00/200
	Variance	0.033333	0.11954	0.068966	0.033333	0.064368	0.0
Total	Prediction Error Rate	11.84/231	5.80/231	11.0/231	4.60/231	11.0/231	4.70/231
	Variance	0.281609	0.303448	0.068966	0.317241	0.0	0.217241
	Groupings	B	A	B	A	B	A
False Alarm Rate (%)		1.99	0.94	1.50	0.52	1.47	0.50
Detection Rate (%)		74.61	87.32	74.19	88.48	73.97	88.06



(A) False Alarm Rate



(B) Detection Rate

[Fig. 1] Performance of DBNs Using Two Normalization Methods

5. Conclusions

Accurate and rapid detection of incidents on the road is very important in traffic research. The main purpose of this study is to improve the traffic incident detection system on highway which is currently used by the traffic control center and results of this study show that the research objective can be fully achieved. It is possible to increase the detection rate for traffic incidents and to minimize the false detection rate, thereby improving the reliability of the currently operating system in Traffic management centers (TMCs).

In operation center, DR (Detection Rate) and FAR(False Alarm Rate) are used for system reliability test. In the automatic incident detection system, when a traffic incident is detected, the light is on and the alarm is activated. If it is not an incident, it will be called FAR (False Alarm Rate). Currently, various algorithms are currently being used for automatic traffic incident detection in the traffic control center and DNNs(Deep Neural Networks) can be used to improve the detection rate accuracy.

However, in Deep Neural Networks(DNNs), it is important to determine input data normalization method for learning. This study examined two different normalization methods, FORM 1 and FORM 2, of input data. According to the results of this study, FORM 2 method is superior to FORM 1. That is, FORM 2 provides higher DR(Detection Rate) and lower FAR (False Alarm Rate) than FORM 1. The normalization of Form 2 is the result of reflecting the position characteristics of the road with the location information of the detectors. The results of this study indicate that information of road and information of sensor location is very important for normalization of input variables. So far, we have used only normalized traffic volume, speed, and occupancy value for learning in ANN(Artificial Neural Networks), but this

study proposes to add various information such as sensor location information and road information in order to improve system reliability in the future. In particular, the normalization method proposed in this study can be used not only in the traffic operation field but also in various ANN applications.

References

- [1] E. Nathanaila, P. Kourosa, P. Kopeliasa, Traffic volume responsive incident detect, *Transportation Research Procedia*, (2017), Vol.25, pp.1755-1768.
- [2] R. Rossi, M. Gastaldi, G. Gecchele, V. Barbaro, Fuzzy logic-based incident detection system using loop detectors data, *Transportation Research Procedia*, (2015), Vol.10, pp.266-275.
- [3] Y. J. Stephanedes, A. P. Chassiakos, Freeway incident detection through filtering, *Transportation Research Part C*, (1993), Vol.1, pp.219-233.
- [4] C. E. Hatri, J. Boumhidi, Fuzzy deep learning based urban traffic incident detection, *Cognitive Systems Research*, (2018), Vol.50, pp.206-213.
- [5] J. Xiao, SVM and KNN ensemble learning for traffic incident detection, *Physica A*, (2019), Vol.517, pp.29-35.
- [6] Q. Wang, Traffic incident detection based on Artificial Neural Network, *IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*, (2011), May 27-29; Xi'an, China
- [7] J. Lu, S. Chen, W. Wang, B. Ran, Automatic traffic incident detection based on nFOIL, *Expert Systems with Applications*, (2012), Vol.39, pp.6547-6556.
- [8] L. Yu, L. Yu, J. Wang, L. Yu, Y. Qi, H. Wen, Back-Propagation Neural Network for Traffic Incident Detection Based on Fusion of Loop Detector and Probe Vehicle Data, *ICNC Fourth International Conference on Natural Computation*, (2008), October 15; Jinan, China
- [9] Z. Zhang, X. Lin, B. Hu, Algorithm design of traffic incident automatic detection based on mobile detection, *IEEE International Conference on Service Operations, Logistics, and Informatics (SOLI)*, (2011), July 10-12; Beijing, China
- [10] D. Kim, Deep Learning Neural Networks for Automatic Vehicle Incident Detection, *Asia-pacific Journal of Convergent Research Interchange*, (2018), Vol.4 No.3 pp.119-128.
- [11] A. Mohamed, G. Dahl, G. Hinton, Acoustic modeling using deep belief networks, *IEEE Transactions on Audio, Speech, and Language Processing*, (2012), Vol.20, No.1, pp.14-22.
- [12] D. Kim, Normalization Methods for Input and Output Vectors in Backpropagation Neural Networks, *International Journal of Computer Mathematics*, (1999), Vol.71, No.2, pp.199-211.
- [13] D. Kim, Automatic Incident Detection Using Machine Learning, *Seoul Studies*, (2005), Vol.6, No.3,

pp.71-80.

- [14] D. Kim, S. Lee, S. Cho, Input Vector Normalization Methods in Support Vector Machines for Automatic Incident Detection, *Transportation Planning and Technology*, (2007), Vol.30, No.6, pp.593-608.
- [15] G. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, (2006), Vol.313, No.5786, pp.504-507.
- [16] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature*, (2015), Vol.521, No.7553, pp.436-444.