Classification of Police Reports and Non-Police Reports with Data Length Normalization Learning

Hyunho Park
Police Science & Public Safety ICT Research Center
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
hyunhopark@etri.re.kr

Eungyeol Lee
Electronics and Telecommunications Research Institute, Daejeon
Gwangju Institute of Science and Technology, Gwangju
Republic of Korea
doryeon514@gm.gist.ac.kr

Sungwon Byon
Police Science & Public Safety ICT Research Center
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
swbyon@etri.re.kr

Eunjung Kwon
Police Science & Public Safety ICT Research Center
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
ejkwon@etri.re.kr

Minjung Lee
Police Science & Public Safety ICT Research Center
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
minjunglee@etri.re.kr

Eui-Suk Jung
Police Science & Public Safety ICT Research Center
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
esjung@etri.re.kr

***Abstract*—The classification of police reports and non-police reports is important when emergency police calls (e.g., 112 calls in South Korea) are received at a police station. Non-urgent calls should be promptly classified as non-police reports to reduce the burden on police operations. When conducting training for the classification of police and non-police reports, there is an issue of misclassification due to the differing lengths of training text data between police reports and non-police reports. This paper proposes data length normalization learning (DLNL), which normalizes data for training and learns the normalized data, to improve the classification of police reports and non-police reports. This paper also explains the performance of the improved classification with an F1-score of 0.99. The enhanced classification of police and non-police reports with the DLNL will help to reduce the burden on police operations and, in turn, enhance police response capabilities.**

***Keywords*—police reports, non-police reports, emergency calls, data dength normalization learning, police response capabilities**

I. Introduction

It is important to distinguish between urgent and non-urgent tasks when a police emergency calls are received at a police station. By promptly assigning non-urgent tasks, which do not require police intervention, to other agencies, the burden on police response can be reduced and the congestion of emergency calls can decrease, thereby increasing the efficiency of police operations. According to statistics from the Korean National Police Agency, about 7.8 million out of approximately 19 million emergency police calls (e.g., 112 calls in South Korea) were allocated to non-emergency tasks. This means that approximately 40% of all 112 calls were non-urgent [1].

Research on using deep learning to analyze police emergency calls has been ongoing [2-5]. Study [2] suggests an architecture of an emergency dispatch support system for police officers to respond swiftly and accurately to incidents by analyzing police emergency calls. Study [3] focused on analyzing emergency calls to identify characteristics for better dispatch efficiency. Study [4] introduced a crime response system that calculates danger levels based on emergency calls. Study [5] discussed a system implementation that provides context-aware information by analyzing emergency calls. However, there has been little research to distinguish between urgent and non-urgent tasks from emergency police calls.

This paper proposes data length normalization learning (DLNL) to classify police reports (PRs), which indicate urgent tasks and require police intervention, and non-police reports (NPRs), which indicate non-urgent tasks and do not require police intervention. In South Korea, a significant number of non-police reports from 112 calls (i.e., police emergency calls in South Korea) are redirected to an 182 call center, which handles complaints related to the police [6], or to the call center of national human rights commission (NHRC), which provides consulting services related to human rights [7]. We developed a classification model that distinguishes between PRs and NPRs by learning from examples related to PRs as well as examples from the websites [6-7] of the 182 call center and NHRC. However, we found that the model misclassified 112 calls depending on the text length of the 112 calls. The misclassification occurred because the examples used for training from PRs, the 182 call center, and the NHRC had different text lengths. DLNL normalizes text lengths for training data and thus reduces misclassification depending on text lengths of 112 calls.

II. Misclassification Problems Depeding on Text Lengths of 112 Calls without DLNL

Before demonstrating DLNL, this section explains misclassification problems depending on text lengths of 112 calls. We gathered data to train and test for classifying police reports (PRs) and non-police reports (NPRs) from 112 calls. Data for PRs were received by our partner company that has experiences to use data of police emergency calls. The partner company generated the data for PRs based on the experiences to use data of police emergency calls. Data for NPRs were created by referring to call examples from the websites of the 182 call center and NHRC, as indicated in [6] and [7].

Table I shows examples of data for PR, 182, and NHRC. PR data include reports that require police intervention, such as assault, robbery, and murder. 182 data include reports that do not require police intervention, such as inquiries about traffic regulations. NHRC data include reports that involve requests for consultations on laws or regulations related to human rights. However, the NHRC data have longer texts than

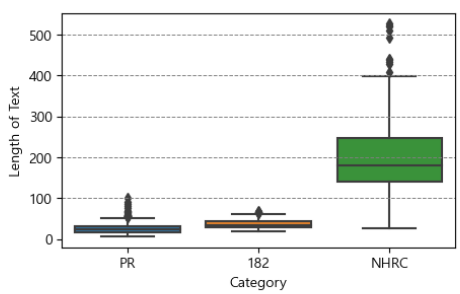
PR data and 182 data because the examples of NHRC reports in [7] also include explanations of the background of the human rights violations.

TABLE I. DATA EXAMPLES FOR PR, 182, AND NHRC.

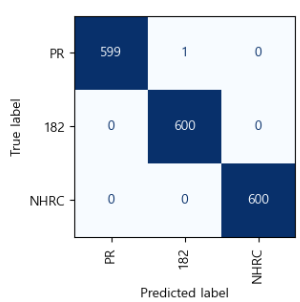
Categories	Examples depending on the Categories
PR data	Someone is wielding a weapon and threatening the people around them.
182 data for NPRs	What is the safe speed to drive in the Hi-Pass section?
NHRC data for NPRs	In relation to the designation of a basic livelihood security recipient, I discovered that a public official had accessed my property and occupation records without my consent. …… I am still dissatisfied with the situation.

We created a classification model by training the PR, 182, and NHRC data with SNUNLP sentence-BERT (SBERT) embeddings [8] and a support vector machine (SVM) with a regularization margin of 0.5. The number of PR data, 182 data, and NHRC data for training is 2,800 each. The number of PR data, 182 data, and NHRC data for testing is 600 each.

Figure 1 shows the distributions of data lengths by categories (i.e., PR, 182, and NHRC) and a confusion matrix indicating the performance to classifying testing data without the proposed DLNL. Figure 1(a) shows text length distributions of testing data by categories. Average text lengths of PR data, 182 data, and NHRC are respectively about 26, 37, and 196. Maximum text lengths of PR data, 182 data, and NHRC are respectively about 102, 68, and 527. Minimum text lengths of PR data, 182 data, and NHRC are respectively about 6, 18, and 27. Figure 1(b) shows that the classification model misclassified only one PR datum as an 182 datum. The F1-score of the classification model with the testing data was 0.9994.



(a) Text length distributions of testing data

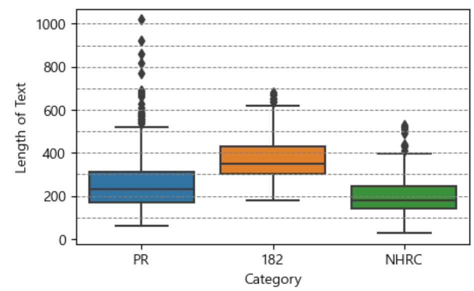


(b) Confusion matrix with testing data

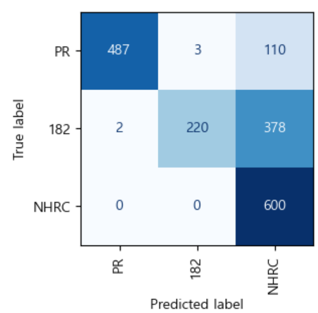
Fig. 1. Text length distribution and classification results by using test data without DLNL.

However, we found that data length affects misclassification. If we input the long texts to the classification model without DLNL, the classification model classified the text as NHRC data. By inputting the PR and 182 data with increased lengths by repeating 10 times, we verified that this issue exists.

Figure 2 shows the distributions of data lengths by category and a confusion matrix for classifying extended testing data with increased lengths in PR and 182 data without DLNL. Figure 2(a) shows text length distributions of extended testing data by categories. Average text lengths of PR data, 182 data, and NHRC are respectively about 260, 368, and 196. Maximum text lengths of PR data, 182 data, and NHRC are respectively about 1020, 680, and 527. Figure 2(b) shows that the classification model misclassified about one-sixth of PR data and one-fifth of 182 data as NHRC data. The F1-score of the classification model with the extended testing data was 0.7261. This result indicates that the performance of the classification model can decrease by more than 20% in terms of F1 score as the data length increases.



(a) Text length distributions of extended testing data



(b) Confusion matrix with extended testing data

Fig. 2. Text length distribution and classification results by using extended test data without DLNL.

III. DATA LENGTH NORMALIZATION LEARNING TO IMPROVE THE PERFORMANCE OF CLASSIFYING PRS AND NPRS

We propose data length normalization learning (DLNL) to improve the performance of classifying PRs and NPRs. DLNL is a training method that normalizes training data to the same text length. Figure 3 explains the process of DLNL. In the first step, DLNL extends the training data by repeating up to the maximum length for data extension (MLDE). The ML of DLNL can be determined by referring to the maximum length of the training data. In this paper, we set the MLDE to 600 because the maximum data length of the NHRC data is 527, as shown in Fig. 1 (a). In the second step, the DLNL splits the extended data into basic segments of basic segment length

(BSL). In this paper, we set the BSL to 10 because the minimum text length of testing data is 6. In the third step, the DLNL trains the basic segments with categories of PR, 182, and NHRC. In this paper, the DLNL learns the segments by using with SNUNLP SBERT embeddings and a SVM with a regularization margin of 0.5.

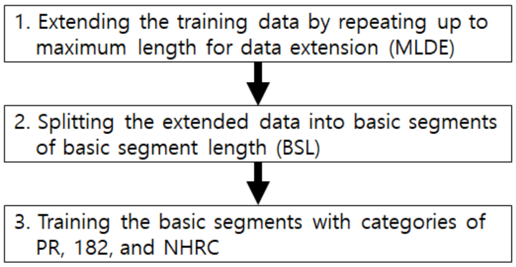


Fig. 3. Process of DLNL.

Figure 4 shows the confusion matrix using DLNL and testing data that have the text length distributions shown in Fig. 1(a). Three PR data points are misclassified as 182 data, and six PR data points are misclassified as NHRC data. Compared to Fig. 1(b), the classification model with DLNL has lower performance than the classification model without DLNL, as shown in Fig. 1(b). However, the F1-score of the classification model using DLNL and testing data is 0.995, indicating that the classification model with DLNL also has good performance.

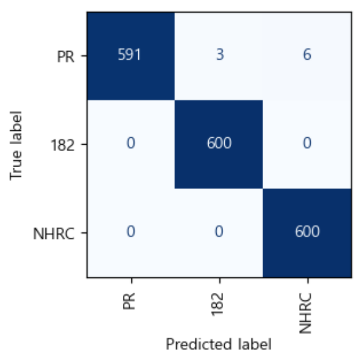


Fig. 4. Confusion matrix using DLNL and testing data.

Figure 5 shows the confusion matrix using DLNL and extended testing data that have the text length distributions shown in Fig. 2(a). Figure 5 shows that the classification model misclassified only one PR datum as an 182 datum. Compared to Fig. 2(b), the classification model with DLNL has much higher performance than the classification model without DLNL, as shown in Fig. 2(b). The F1-score of the classification model using DLNL and extended testing data is 0.9994. As shown in Figures 4 and 5, it demonstrates that DLNL can prevent performance degradation due to the length of the input data.

IV. Conclusion

This study addressed the critical need to efficiently classify police emergency calls into PRs and NPRs using DLNL. We found and verified misclassification problems depending on text lengths. We proposed and implemented DLNL, which extends training data, splits the extended data into basic segments, and then trains the basic segments with SBERT and SVM. DLNL demonstrated an F1-score of over 0.99 in classifying PR, 182, and NHRC data regardless of input sentence length. DLNL solves the misclassification problems depending on text lengths. Therefore, DLNL has the potential to enhance police operations by streamlining the handling of emergency calls, ultimately contributing to better resource allocation and reduced congestion in emergency call centers.

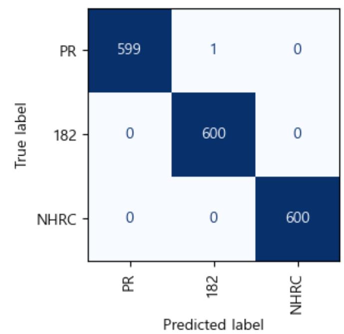


Fig. 5. Confusion matrix using DLNL and extended testing data.

Acknowledgment

This research was supported and funded by the Korean National Police Agency. [Project Name: 112 Emergency Dispatch Decision Support System/ Project Number:PR08-03-000-21]

References

[1] Statistics of Korean National Police Agency, Retrieved from https://www.police.go.kr/www/open/public/public0210.jsp

[2] H. Park et al., “Emergency Dispatch Support System for Police Officers in Incident Scenes,” *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1542-1544, October 2021.

[3] S. Byon et al., “An Analysis of Incident Report Data for Emergency Dispatch,” *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1425-1427, October 2021.

[4] H. Park et al., "Crime Response System based on Danger Degree Calculations of Crime Scenes," *2022 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 2221-2224, October 2022.

[5] S. Byon et al., "An Implementation of Retrieving Situation-Aware Information for Emergency Incident." *2022 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1932-1934, October 2022.

[6] Police complaint portal, Retrieved from https://minwon.police.go.kr

[7] National human rights commission of Korea, Retrieved from https://www.humanrights.go.kr/eng/contents/view?contentsNo=124&menuLevel=3&menuNo=123

[8] snunlp/KR-SBERT-V40K-klueNLI-augSTS, Retrieved from https://huggingface.co/snunlp/KR-SBERT-V40K-klueNLI