Emerging Technologies and Future of Work, Vol. 117, 2023, 49–56
https://doi.org/10.54941/ahfe1004404

Uncertainty-Aware AI-Facilitated Decision Support System for Emergency Call Takers

Minjung Lee and Sungwon Byon

Electronics and Telecommunications Research Institute, Daejeon, 34129, Republic of Korea

ABSTRACT

This paper presents an artificial intelligence (AI)-based decision support system designed to assist call takers sitting behind emergency numbers, such as 1-1-2 or 9-1-1. Accurate situational awareness is crucial during the intake of emergency reports for effective initial response. However, in emergency situations, callers' utterances can be uncertain, leading to potential false predictions by the AI model, which may delay the appropriate response. To address this issue, we propose an uncertainty-aware AI model for the decision support system. This system displays a set of predicted emergency situational candidates, based on text transcribed in real-time from the voice report intake, thereby aiding call takers in efficiently handling emergency calls. The type and number of these candidates are determined by considering the uncertainty inferred from the AI model. We provide a detailed explanation of the proposed system and evaluate its performance using actual domestic 1-1-2 police emergency report data.

Keywords: Uncertainty-awareness, Artificial intelligence, Decision support system, Emergency call, Emergency report

INTRODUCTION

In recent times, there has been a growing trend in both the private and public sectors to leverage various technologies through the integrated application of Information and Communication Technology (ICT) and Artificial Intelligence (AI). The deployment of these technologies across public systems, including safety, defense, education, law enforcement, and culture, can potentially lead to numerous innovations. These innovations include the enhancement of public service efficiency and cost minimization. In this context, the present study introduces an AI-facilitated system specifically designed to assist call takers operating behind emergency numbers, such as 1-1-2 or 9-1-1(Baek et al., 2021 & Blomberg et al., 2021).

Emergency call centers play a pivotal role in the overall emergency response process. The initial interaction between the caller and the call taker often determines the effectiveness of the subsequent response. During this interaction, the call taker must quickly and accurately identify the type of emergency based on the information provided by the caller. Accurate situational awareness during the emergency report intake is crucial for an effective initial response. However, recognizing the type of emergency, while simultaneously asking a series of questions to the caller and recording information to be relayed to first responders (e.g., police officers, firefighters, and ambulance crews), is a challenging task. Consequently, there has been a growing demand for AI models to alleviate the workload of call takers in emergency report intake systems.

A representative example of this technology is the speech recognition AI model that automatically transcribes voice reports into text. In various countries, including the United States and South Korea, speech recognition technology for emergency calls has reached considerable maturity and has been implemented in actual intake systems. This technology help to prevent the waste of crucial initial response time, often lost to repeated questions when call takers struggle to understand the caller's voice. Additionally, it enables the collection of a large volume of spoken text data in emergency situations, disasters, and crime-related fields. The text data, alongside with images, is an area where recent AI models have made significant advancements, paving the way for the development of various AI-facilitated systems. This study also explores a text classification AI model that predicts the type of emergency based on the real-time transcribed voice of the caller and the call taker. Such a model could be instrumental in developing a system that supports the real-time decision-making process of call takers.

Despite the convenience of the AI-facilitated emergency call-taking system mentioned above, there are concerns about the direct integration of AI models into systems that are crucial to public safety. Callers are non-experts in emergency reporting, may not know what information is vital for the call taker. Furthermore, in emergency situations, callers can be flustered and agitated, driven by urgency and fear, which may result in unclear and inaccurate speech. This raises a critical question: Can we trust the results obtained from the AI model when it processes such uncertain and potentially unreliable utterance data? If the AI model's unreliable predictions lead to incorrect emergency responses, it could pose serious threats to citizen safety. To mitigate these risks, we have incorporated an uncertainty-aware AI model into the emergency call-taking system. The proposed system is designed to display a broader range of emergency situational candidates in cases of high uncertainty, and a more refined set when the uncertainty is low, thereby minimizing the risk of unreliable predictions.

CRIME SITUATIONAL RECOGNITION IN EMERGENCY CALL INTAKE

In Korea, there are two primary emergency call numbers: 1-1-2 and 1-1-9. The number 1-1-2 is designated for crime-related emergencies. According to the 1-1-2 manual of the Korean National Police Agency, Korean police officers utilize the 1-1-2 system, which incorporates advanced IT technology and the police communication network, for handling cases reported through this emergency number. When receiving calls, the police officer in charge of reception is responsible for leading the conversation with the caller.

This involves asking a series of questions to ascertain critical information such as the caller's location, perpetrator details, the urgency of the situation, and the type of crime. This information is then recorded to the 1-1-2 system by typing, enabling police officers dispatched to the scene to access and verify it through the system.

To enhance crime situational awareness, the receptionist should categorize each report into one of the 58 crime categories, such as murder, robbery, violence, fraud, and traffic accidents, and enter this information into the 1-1-2 system. However, relying solely on the receptionist's subjective judgment for situational awareness can be problematic, as it may be influenced by their competence and condition. Incorrectly perceived crime situations at the reception stage can adversely affect subsequent stages (such as dispatch, command, and response), leading to wasted time and police resources. To address this issue, an AI model has been developed to automatically recognize crime situations based on the report summaries entered by the receptionist into the 1-1-2 system, as illustrated in Table 1. However, this existing model, which determines the crime situation based on the receptionist's summarized report, has limitations in early detection and may be biased due to the subjectivity of the receptionist.

This study develops an AI model that recognizes crime situational candidates based on the dialogue between the receptionist and the caller, which is automatically transcribed by a voice recognition AI model. By employing an AI model capable of quantifying uncertainty, this system can display a range of possible crime situations to the receptionist, depending on the level of uncertainty. This prompts the receptionist to ask more detailed questions to the caller about the situation, thereby enhancing the accuracy and efficiency of the emergency response process.

Table 1. Differences between the existing and the proposed AI model.

Feature	**Existing AI model**	**Proposed AI model**
Recognition Timing	After entering report	During the emergency call
Input	Report summary	Transcribed dialogue
Uncertainty Awareness	X	O

DISSONANCE-AWARE NEURAL NETWORKS

Deep neural networks have significantly advanced the predictive accuracy in text classification tasks (Balkus & Yan., 2022, Sun et al., 2019). However, traditional deep neural networks often face issues of high-confidence misclassifications without considering the inherent uncertainty of specific task (Gall & Ghahramani. 2016). To address this, recent studies have conducted to provide approaches in deep neural networks that not only measure predicted values but also quantify the associated uncertainty, aiming to circumvent the issue of misclassification (Kendall & Gal 2017, Chen et al., 2020).

In our decision support system, we focus on 'dissonance (Josang et al., 2018)', a type of uncertainty obtainable from evidential neural networks

(ENNs) (Sensory et al., 2018, Hu & Khan., 2021). We hypothesize that the information made by callers describing an event may contain words and sentences that could be confused with various emergency situations. Therefore, our proposed model is designed to measure dissonance, relates to topic mismatch (conflicting evidence). Consequently, we have named our proposed model Dissonance-aware neural networks (DANNs).

The output of a standard neural network classifier is a probability assignment over possible classes, but DANNs assume a Dirichlet distribution for the probabilities of categories (opinions), treating the model's predictions as subjective opinions and learning a function to gather evidence that elicits such opinions form the data. A multinomial opinion in given $x_{dialogue}$ and x_{extra} is represented by $\omega_Y = \left(\boldsymbol{b}_Y, u_Y, \boldsymbol{a}_Y\right)$ where a domain is $\mathbb{Y} \equiv \{1, \ldots, K\}$, a random variable Y takes value in $\mathbb{Y}$, $K = |\mathbb{Y}| \geq 2$ and ω_Y is given as $\sum_{y \in \mathbb{Y}} \boldsymbol{b}_Y(y) + u_Y = 1$. $\boldsymbol{b}_Y$ denotes belief mass function over $\mathbb{Y}$. u_Y denotes uncertainty mass representing vacuity of evidence. $\boldsymbol{a}_Y$ represents base rate distribution over $\mathbb{Y}$, with $\sum_y \boldsymbol{a}_Y\left(y\right) = 1$. Then the projected probability distribution of a multinomial opinion is given by:

$$\boldsymbol{P}_Y(y) = \boldsymbol{b}_Y(y) + \boldsymbol{a}_Y(y) u_Y, \ \forall_y \in \mathbb{Y}. \tag{1}$$

Multinomial probability density over a domain of cardinality K is represented by the K-dimensional Dirichlet PDF where the special case with $K = 2$ is the Beta PDF as a binomial opinion. It denotes a domain of K mutually disjoint elements in $\mathbb{Y}$ and α_Y the strength vector over $y \in \mathbb{Y}$ and $\mathbf{P}_Y$ the probability distribution over $\mathbb{Y}$.

$$\boldsymbol{Dir}\left(\boldsymbol{p}_Y; \boldsymbol{\alpha}_Y\right) = \frac{1}{B\left(\boldsymbol{\alpha}_Y\right)} \prod_{y \in \mathbb{Y}} \boldsymbol{p}_Y(y)^{(\boldsymbol{\alpha}_Y(y)-1)}, \tag{2}$$

where $B\left(\boldsymbol{\alpha}_Y\right)$ is a multivariate beta function as the normalizing constant, $\boldsymbol{\alpha}_Y \geq 0$, and $\boldsymbol{p}_Y(y) \neq 0$ if $\boldsymbol{\alpha}_Y < 1$.

Given a sample with the input $\boldsymbol{x}_{dialogue_i}$ and $\boldsymbol{x}_{extra i}$ and the ground-truth label $\boldsymbol{y}_i$, let $f(\boldsymbol{x}_{dialogue_i}, \boldsymbol{x}_{extra i}|\theta)$ represents the predicted evidence vector predicted by the classifier with parameters θ. Then the corresponding Dirichlet distribution has parameters $\alpha_i = f(\boldsymbol{x}_{dialogue_i}, \boldsymbol{x}_{extra i}\,|\theta) + 1$. The Dirichlet density $Dir\left(\boldsymbol{p}_i; \alpha\right)$ is the prior on the Multinomial distribution $Multi\left(\boldsymbol{y}_i; \boldsymbol{p}_i\right)$. Then the following sum of squared loss is optimized for training:

$$\begin{aligned}\mathcal{L}\left(f\left(\boldsymbol{x}_{dialogue_i}, \boldsymbol{x}_{extra i}\middle|\theta\right), \boldsymbol{y}_i\right) &= \int \frac{\left\|\boldsymbol{y}_i - \boldsymbol{p}_i\right\|_2^2}{B\left(\boldsymbol{\alpha}_i\right)} \prod_{j=1}^{K} p_{ij}^{(\alpha_{ij}-1)} d\boldsymbol{p}_i \\ &= \sum_{j=1}^{K} \left(y_{ij}^2 - 2y_{ij}\mathbb{E}\left[p_{ij}\right] + \mathbb{E}\left[p_{ij}^2\right]\right).\end{aligned} \tag{3}$$

Dissonance is determined to be high when the Dirichlet distribution parameter values assigned to multiple categories are large, indicating data with conflicting evidence. In the emergency call reception support system, the candidate set to be presented to the call taker is determined using the predicted probability vector and topic dissonance obtained from the trained model. Given a multinomial opinion with non-zero belief masses, the measure of dissonance is:

$$Diss\left(\boldsymbol{\alpha}_{\mathrm{Y}}\right) = \sum_{y_i \in \mathbb{Y}} \left(\frac{\boldsymbol{b}_{\mathrm{Y}}(y_i) \sum_{y_j \in \mathbb{Y} \setminus y_i} \boldsymbol{b}_{\mathrm{Y}}(y_j) Bal(y_j, y_i)}{\sum_{y_j \in \mathbb{Y} \setminus y_i} \boldsymbol{b}_{\mathrm{Y}}(y_j)} \right), \tag{4}$$

where the relative mass balance between a pair of belief masses $\boldsymbol{b}_{\mathrm{Y}}(y_j)$ and $\boldsymbol{b}_{\mathrm{Y}}(y_i)$ is expressed by:

$$Bal\left(y_j, y_i\right) = \begin{cases} 1 - \frac{|\boldsymbol{b}_{\mathrm{Y}}(y_j) - \boldsymbol{b}_{\mathrm{Y}}(y_i)|}{\boldsymbol{b}_{\mathrm{Y}}(y_j) + \boldsymbol{b}_{\mathrm{Y}}(y_i)}, & \text{if } \boldsymbol{b}_{\mathrm{Y}}\left(y_j\right) \boldsymbol{b}_{\mathrm{Y}}\left(y_i\right) \neq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

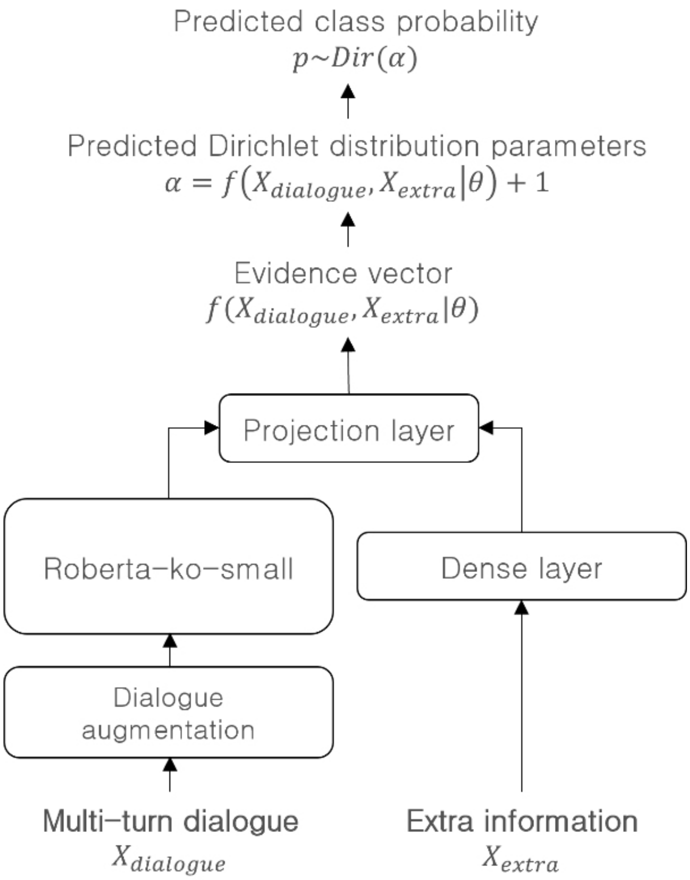

Figure 1: Dissonance-aware neural network structure.

Figure 1 illustrates the structure of DANNs. To effectively extract features from the multi-turn dialogues between the call taker and the caller, we employed a pre-trained model specifically designed for the Korean language, named 'roberta-ko-small' (https://huggingface.co/lassl/roberta-ko-small), guided by the LASSL (Language framework for Self-Supervised Learning) (https://github.com/lassl/lassl). For training, multi-turn dialogues were augmented by sampling dialogue turns of arbitrary lengths. Additionally, we incorporated incident-specific information, such as the reporting time (ranging from 1 to 24 hours). This was used for the model's training as a one-hot encoded vector.

EXPERIMENTS AND RESULTS

The dataset used for training the model consists of 17,218 reports from real cases recorded in the Korean police's 112 system. This dataset encompasses a wide range of crime categories, including 10 major crimes (such as homicide, robbery, sexual assault), 15 miscellaneous crimes (such as violence, fraud, threats, gambling), 8 public order-related crimes (such as quarrels, disturbances), 6 traffic-related crimes (such as traffic accidents, traffic inconveniences), 9 other police-related tasks (such as inquiries, suicide) and 10 other agency-related crimes (fire, noise). Consequently, the proposed deep neural network model is designed to classifies 58 types of emergency situations.

To evaluate out models, we consider Accuracy and macro F1 score. Accuracy is a straightforward metric that quantifies the proportion of correctly predicted instances out of the total instances. The F1 score is the harmonic mean of precision and recall, where precision is the ratio of correctly predicted positive observations to the total predicted positives, and recall (or sensitivity) is the ratio of correctly predicted positive observations to all the observations in the actual class.

Table 2 presents the experimental results, which compare different combinations related to the type of input text data, the inclusion of uncertainty estimation, the use of additional information, and the presence of multi-turn conversation augmentation. The proposed model demonstrated the best performance with a classification accuracy of 83.04% and an F1 score of 77.96%.

Table 2. Experimental results.

					(Proposed)
Input text	Report summary	Transcribed dialogue	Transcribed dialogue	Transcribed dialogue	**Transcribed dialogue**
Uncertainty	X	X	O	O	**O**
Extra information	X	X	X	O	**O**
Data augmentation	X	X	X	X	**O**
Accuracy	63.13	65.40	71.92	80.42	**83.04**
F1 score	60.77	63.70	67.98	73.97	**77.96**

CONCLUSION

This paper introduces an artificial intelligence-based system designed to assist emergency call recipients. The system displays a set of potential emergency situation candidates, which are determined based on the dialogue transcribed in real-time between the caller and the receiver through a speech recognition model during an emergency call. A key feature of the proposed classification model is its ability to estimate uncertainty about its predictions. When the dialogue contains a mix of various emergency situation contents, the model estimates high uncertainty. In response to this, the system presents a broader set of candidate to the receiver. This approach is expected to prompt the receiver to ask more specific questions to the caller, thereby obtaining information that can help narrow down the set of candidates to the correct emergency situation. This method aims to address the issue of incorrect initial responses caused by the AI model's misjudgement of the emergency type. The proposed system is anticipated to enhance the work efficiency of receptionists and enable faster, more accurate emergency responses. In the future, this system has the potential to be utilized as a tool in various fields that require multi-turn dialogue-based situation (or topic) recognition.

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

Baek, M. S., Park, W., Park, J., Jang, K. H., & Lee, Y. T. (2021). Smart policing technique with crime type and risk score prediction based on machine learning for early awareness of risk situation. IEEE Access, 9, 131906–131915.

Balkus, S. V., & Yan, D. (2022). Improving short text classification with augmented data using GPT-3. Natural Language Engineering, 1–30.

Blomberg, S. N., Christensen, H. C., Lippert, F., Ersbøll, A. K., Torp-Petersen, C., Sayre, M. R.,... & Folke, F. (2021). Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: a randomized clinical trial. JAMA network open, 4(1), e2032320–e2032320.

Chen, W., Zhang, B., & Lu, M. (2020). Uncertainty quantification for multilabel text classification. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(6), e1384.

Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050–1059). PMLR.

Hu, Y., & Khan, L. (2021, August). Uncertainty-aware reliable text classification. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (pp. 628–636).

Josang, A., Cho, J. H., & Chen, F. (2018, July). Uncertainty characteristics of subjective opinions. In 2018 21st International Conference on Information Fusion (FUSION) (pp. 1998–2005). IEEE.

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision?. Advances in neural information processing systems, 30.

Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. Advances in neural information processing systems, 31.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune bert for text classification?. In Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18 (pp. 194–206). Springer International Publishing.