Human Error, Reliability, Resilience, and Performance, Vol. 132, 2024, 98–102
https://doi.org/10.54941/ahfe1004885

Data Collection and Analysis of Inter-Area Communication During a Disaster Exercise at a Large Hospital

Hiroki Obara[1], Taro Kanno[1], Kazumi Kjiyama[2], Haruka Yoshida[3], Misumi Yamazaki[4], Michihiro Tsubaki[5], and Sachika Sharikura[6]

[1]The University of Tokyo, Bunkyo, Tokyo 1138656, Japan
[2]Kitasato University Hospital, Sagamihara, Kanagawa 2520375, Japan
[3]Nihon University, Narration, Chiba 2740072, Japan
[4]Japanese Redcross College of Nursing, Shibuya, Tokyo 1500012, Japan
[5]Kitasato University, Sagamihara, Kanagawa 2520373, Japan
[6]Showa University Hospital, Shinagawa, Tokyo 1420064, Japan

ABSTRACT

This study collected inter-area communications using PHS(personal handy-phone system) during a hospital disaster response exercise and analyzed the data using a large language model. We attached a voice recorder to a PHS to record conversations on both sides of the communication. After transcribing the recorded verbal data, we employed a large language model (GPT-4.0) to automatically classify the utterances into nine performative functions and calculate the anticipation ratio for the communication between different areas in the hospital. The results confirmed that the GPT-4.0 classification was sufficiently reliable, with a high agreement (Cohen's kappa = 0.73) against the gold standard created by human analysts. Additionally, the anticipation ratio of the command post to other areas was higher in the 2023 exercise than in the 2022 exercise, suggesting that information-sharing initiated from the command post improved in the 2023 exercise.

Keywords: Hospital disaster response, Disaster exercise and drills, Business continuity plan, Communication evaluation, Large language model

INTRODUCTION

Hospitals, especially regional disaster base hospitals, play a critical role in saving lives during disasters. Therefore, it is important for them to conduct disaster response exercises and thoroughly evaluate the results to understand the current level of response capability and to identify potential problems in disaster response and hospital business continuity. However, because data collection and disaster exercise analysis requires substantial manpower and time, such evaluations have not been well conducted. Thus, aiming to develop evaluation indices of exercise performance, this study collected data on patient and document flow, inter-area communication, and exercise participant behavior using video cameras, voice recorders, and NFC (Near Field Communication) tags.

Focusing on inter-area communication, this study describes the data collection process using voice recorders attached to exercise players and reports

on the analysis of the inter-area communication, including an attempt to use a large-scale language model to automatically classify the verbal data with several performative verbs.

DATA COLLECTION

The data used in our analysis were collected during disaster response exercises conducted in 2022 and 2023 at a disaster base hospital in Kanagawa Prefecture and (Ideguchi et al., 2023; Kanno et al., 2023). This hospital, designated as a regional disaster base hospital, is expected to play a central role in regional disaster medicine. These two exercises were both designed to deal with mass casualties caused by a major earthquake.

To analyze the communication among different areas and departments in the hospital during the disaster response and business continuity, we attached a voice recorder to the key players in different areas who use the PHS, including the command post, the headquarters, and each first-aid station. Figure 1 shows the voice recorder with an earphone, through which we recorded voices on both ends. The voice data were transcribed and used for the analysis. Table 1 shows the summary of the recorded data.

We conducted the data collection with the approval of the Ethics Committee of the School of Engineering at the University of Tokyo (KE22-75).

Figure 1: Voice recorder with a microphone for PHS.

Table 1. Summary of the voice data collected.

Player	Total Call	Word Counts*
Chief of the command post	0:05:56	41,908
Chief of surgical scheduling at the command post	0:15:03	33,301
Red area (leader doctor)	0:04:43	46,750
Operating room (leader nurse)	0:10:14	30,776
Headquarters	0:04:43	35,125

*In Japanese

ANALYSIS METHOD

We analyzed the conversation content from the viewpoint of performative verbs to calculate the anticipation ratio of inter-area communication. Usually, this kind of analysis is done manually, which requires many manhours. In this study, we applied the GPT-4.0 language model (OpenAI, 2023) to automatically classify the utterances into nine performative verbs: greet, inform, acknowledge, request, query, accept, declare, confirm, and suggest. Table 2 shows the definitions of these nine performative verbs. For the analysis, we first validated the accuracy of the classification by the GPT-4.0. Then we calculated the anticipation ratio of the communication between the command post and other areas/departments to evaluate the efficiency of the communication.

Table 2. Performative verbs used for the analysis.

Performative Verb	Definition
Greet	Refers to the exchange of ceremonial words when making a call, such as the initial "hello," the concluding "excuse me," or stating one's own name to the other party.
Inform	Refers to the act of conveying new information to the recipient.
Acknowledge	Involves recognizing and accepting the other party's statements or actions, with responses such as "yes," "understood," or "I see" being typical examples.
Request	Refers to utterances that involve asking the other party for some action or provision of information.
Query	Refers to the act of asking questions or confirming pros and cons in order to seek information.
Accept	Refers to utterances that show agreement with and readiness to act according to the other party's suggestions or opinions.
Declare	Refers to utterances that convey a clear intention or decision to the other party.
Confirm	Refers to utterances that involve repeating what the other party has said or verifying information that has already been provided.
Suggest	Pertains to the act of expressing suggestions or opinions to the other party.

Validation of the Classification of Performative Verbs Using GPT-4.0

To evaluate the accuracy of the classification by GPT-4.0, we manually categorized 296 utterances during the 2022 exercise into the nine performative verbs and used the results as a gold standard. In this manual classification, two analysts first performed the classification individually and then determined the final classification through discussion, referring to their classifications. Cohen's kappa was 0.83 in the first analysis, indicating that the classification by the two analysts has high agreement and the gold standard is fairly reliable.

In the GPT-4.0 classification, as prerequisite knowledge, we first entered an overview of the hospital disaster response exercise, the definitions of the performative verbs, and the format for input/output of the sentences. We then asked GPT-4.0 to classify each sentence into the nine performative verbs. Since GPT's classifications were somewhat inconsistent, we performed this automatic classification five times and determined the final categorization on a majority basis. The Cohen's kappa against the gold standard was 0.73, indicating that there was substantial agreement between the automatic and the manual classifications. In the following analysis, we used the GPT-4.0 classification with minor manual corrections.

RESULTS AND DISCUSSIONS

We analyzed 666 utterances of PHS communication between the command post and other areas. Figure 2 shows the frequency of calls between various areas in the 2022 and 2023 exercises. Table 3 provides a summary of the classification by performative functions.

Next, we calculated the anticipation ratio—a ratio of "push" to "pull" communication often used as an indicator of efficient information-sharing of the verbal communication—between different areas (Johnson et al., 2023). The higher the anticipation ratio, the more an actor pushes more information than pulls, suggesting that they share information efficiently. Among the nine performative verbs used in this study, "inform" corresponded to "push" and "query" corresponded to "pull." We calculated the anticipation ratio using the counts of utterances categorized to these performative verbs.

Table 4 shows the anticipation ratios in the 2022 and 2023 disaster exercises. A comparison of the ratios between 2022 and 2023 revealed that the ratio was higher for the command post in 2023, in particular from the command post to the red area (1.84-> 3.00) and headquarters (2.21-> 6.99), indicating that the command post in 2023 proactively provided information to these areas in advance before they were asked for the information.

Table 3. Summary of the GTP-4.0 classification by performative verbs.

Performative Verb	CP -> OR	OR -> CP	CP -> Red	Red -> CP	CP -> HQ	HQ -> CP
Greet	14	10	19	23	12	10
Inform	59	28	18	40	16	11
Acknowledge	24	8	25	14	7	10
Request	7	12	10	13	3	2
Query	16	44	35	14	9	8
Accept	3	11	7	13	0	2
Declare	4	12	6	8	3	0
Confirm	15	23	4	8	7	7
Suggest	3	4	3	2	0	0

* CP: command post, OR: operating room, Red: red area, HQ: headquarters

Table 4. Anticipation ratios in the 2002 and 2023 exercises.

Year	CP -> OR	OR -> CP	CP -> Red	Red -> CP	CP -> HQ	HQ -> CP
2022	1.00	1.50	1.00	1.20	1.40	0.89
2023	1.45	2.17	1.36	1.11	4.00	NA

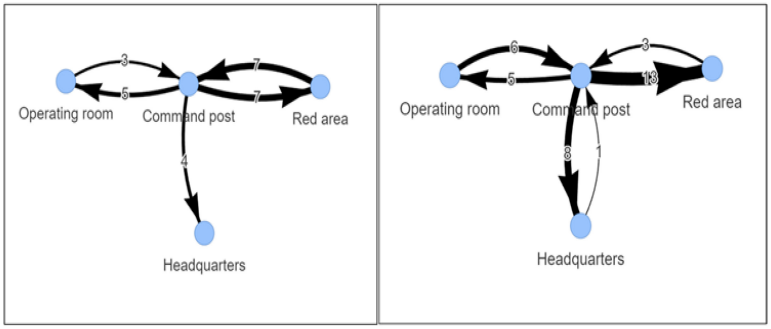

Figure 2: Frequency of calls in 2022 (left) and 2023 (right).

CONCLUSION

This study confirmed that inter-area conversations in a disaster exercise can be clearly recorded with a voice recorder with a microphone attached to an exercise participant. We also demonstrated that a large language model can be used to classify data according to performative verbs, thus saving man-hours in calculating the anticipation ratio. The analysis of the anticipation ratio suggests that the communication in the command post improved in the 2023 exercise compared to the 2022 exercise. The major contribution of this study is it provides a reliable automated classification method to conduct a task that was previously performed manually, thus speeding up the analysis of exercise recordings to find clues for better communication and ultimately a more disaster resilient hospital.

REFERENCES

Ideguchi, T., Kanno, T., Umemoto, M., Kajiyama, K., Ikari, R., Yamazaki, M. and Sharikura, S. (2023). Extensive Data Collection in an In-Hospital Disaster Response Exercise for Evaluating Disaster Resilience, Safety Management and Human Factors, Volume 105, pp. 42—46.

Johson, C. J., Demir, M., McNeese, N. J., Gorman, J. C., Wolff, A. T., and Cooke, N. J. (2023). The Impact of Training on Human-Autonomy Team Communications and Trust Calibration, Volume 65, No. 7, pp. 1554–1570.

Kanno, T., Harada, J., Mitsuhashi, D., Kajiyama, K., and Sharikura, S. (2023). Work-as-Imagined vs. Work-as-Done during a Hospital Response to a Disaster, Proc. Int'l Symposium on Resilience Engineering (RE10).

Open AI. (2023), GPT-4 Technical Report.arXiv.