HSAP: A Human-in-the-loop Social Media-based Situation Awareness Platform

Xiangmin Zhou
RMIT University, Australia
xiangmin.zhou@rmit.edu.au

Chengkun He
RMIT University/ Data61, Australia
chengkun.he@student.rmit.edu.au

Xi Chen
Soochow University, China
chenxi10@suda.edu.cn

Yanchun Zhang
Victoria University, Australia
yanchun.zhang@vu.edu.au

ABSTRACT

Situation-awareness (SA) has been important for natural disaster management and smart decision making. Traditionally, security officers recognize disaster situations through emergency reporting with phone calls. However, due to the busy phone lines or power outages caused by disasters, traditional SA has been limited in terms of time and mitigation response, which may cause high loss in life and properties in disaster areas. Social media-based SA has been studied recently. However, existing systems are designed for events without nonconsecutive migrations over location and time. In this demo, we design HSAP, the first human-in-the-loop social media-based SA platform for effective and efficient management of disasters with complex event migrations. HSAP is designed with a number of novel techniques, including complex social event detection, summarization, and human-in-the-loop result filtering. We demonstrate the usage of HSAP via Nepal earthquake 2015.

PVLDB Reference Format:
Xiangmin Zhou, Chengkun He, Xi Chen, and Yanchun Zhang. HSAP: A Human-in-the-loop Social Media-based Situation Awareness Platform. PVLDB, 17(12): 4493 - 4496, 2024.
doi:10.14778/3685800.3685908

1 INTRODUCTION

Situation awareness (SA) helps understand the elements in the environment, the current situation, and project future actions. Real applications like crisis management require real-time awareness of critical situations. However, traditional services like phone calls can be easily delayed due to busy lines, transfer delays or limited communication ability in the disaster area. Social media are vital in applications like disaster management since critical events that cause great loss in life and properties are commonly identified in them. In natural disasters, events usually involve significant evolution over time and space. For instance, during the Nepal earthquake in 2015, the epicentre occurred at Gorkha District on 25 April, and finally moved to Chinese border between the capital of Kathmandu and Mount Everest on 12 May. Over 9,000 people were killed, and over 600,000 structures were damaged or destroyed in this earthquake. Though social media-based systems enhanced the SA using "citizens as sensors" [5], they cannot recognize the complex events with nonconsecutive time and space movement.

Proceedings of the VLDB Endowment, Vol. 17, No. 12 ISSN 2150-8097.
doi:10.14778/3685800.3685908

To overcome the limitations of existing SA solutions, our recent studies have focused on the social events with complex evolution over time [2] and locations [1, 7]. Effectively detecting complex events is difficult due to the uncertainty and complex evolution of social media. Recently, human-in-the-loop (HITL) has become an effective way in enhancing the systems to recover information from data. Unlike traditional methods that rely on automatic systems, the knowledge of domain experts can contribute to the verification of automatic event information discovery. With the help of HITL technique, the process of SA well utilizes the power of human to enhance the quality of awareness. For being aware of natural disasters with complex event migrations, it is necessary to develop effective HITL situation awareness platform to well harness the power of human and the advanced automatic systems.

In this paper, we design a HITL social media SA platform, called HSAP. HSAP consists of three modules: User Interface, Functional Manager and Data Management. The user interface module provides users, security officers and model developers, with event enquiry and human feedback submitting. By this interface, people can select an event they will check and submit their feedback on the results returned by the system to recursively refine the results. The function manager provides the main functionalities of HSAP, including event detection, event summarization and tracking event movement temporally and spatially. The data management module store data in MongoDB, and provide API for the data access of other modules. The Novelty and contributions are as below.

- We design a novel platform HSAP that incorporates our recently proposed event detection, summarization and tracking techniques for enhanced disaster situation awareness. HSAP enables smart human participation in disaster identification with a novel HITL filter, while visualizes the disaster situation for better event understanding.
- We demonstrate the usage of HSAP from a user perspective, and compare the detection results with HITL filter and those without it. We showcase HSAP using the Nepal earthquake 2015 [7] disaster awareness on Twitter.

2 HSAP OVERVIEW

2.1 Architecture

HSAP is a 'Search & Discovery' web application. It supports three functions: the event detection, the event summarization, and the event tracking. Its user interface runs on the client web browser, and supports users three functions: event enquiry, crisis situation and human feedback. Its data are stored in a MongoDB on the server. The three functions run on the server to generate the event enquiry results that are delivered to client side and shown on interfaces. Figure 1 shows the architecture of HSAP.

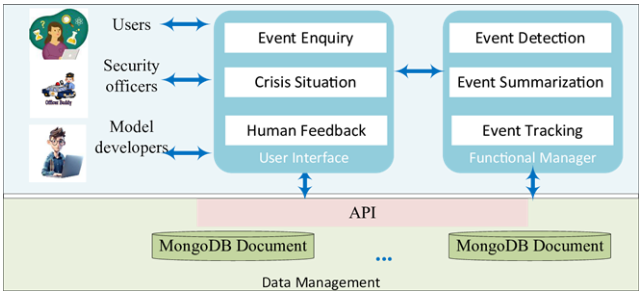


Figure 1: HSAP Framework.

2.2 Workflow

When the HSAP is started, the main page interface shows a brief description on this platform. Given a user or security officer or model developer, s/he can select an event from the list of datasets on the event detection interface, and conduct event detection. Meanwhile, the user can navigate to another webpage for summarizing the selected event. The event enquiry request is passed to functional manager on server that search the MongoDB document of the selected event, and return the crisis situation results to the user. HSAP allows the user to provide his/her feedback by submitting the resulting messages that they label as true event posts, by which the event detection results are filtered by HITL technique.

3 HSAP DESIGN

3.1 Data Management

To achieve fast response time, we design a MongoDB database on the backend to support the data management of HSAP. Each tweet's information is stored as a document that includes its id, hashtags, location, text, timestamp, user id, ConTF/IDF feature [7], and detection result. The designed MongoDB database supports two operations, detection and update. Given a candidate number k, the tweets are selected by MongoDB statement: documents = collection.aggregate(["$match": "IsEvent": "1", "$sample": "size": k]). Given a tweet *id* with a new detection result r, the document is updated by: collection.update_one({"id": *id*}, {"$set": {"IsEvent": r}}).

3.2 User Interface

The user interface module enables users to select their interested event, discover the crisis situation, and interact to this event.

3.2.1 Interfaces. Five interfaces are created for the HSAP.
Homepage interface provides a brief introduction of HSAP, including the function of each component in HSAP. The content on this page is customized for the heading, logo, and welcome, which can be updated as the HSAP upgrade in the future.
Event Detection Interface supports the functionality of HSAP to detect the crisis events. Given a selected disaster, the interface displays the event detection results obtained by ConTF/IDF with contexts [7]. When the current user interacts with the detected results by manually labelling a number of identified posts, the HITL filter is invoked to refine the event detection results and return the updated results to the interface. For each detection process, the accuracy is automatically shown on the interface.
Event Summarization Interface displays the most representative and concise social event messages/words for an event selected on event detection interface. Meanwhile, it shows the ground truth summarization results for the comparison.
Event Movement Interface shares the same input with event detection interface on what event to be analyzed, and dynamically visualizes the event moving trend over locations on the world map.
Event Evolution Interface visualizes the temporal evolving summaries for a selected event. We display this evolvement using two graphs. One shows the topic change over time, while the other shows the summary representative distributions.

3.2.2 Event Enquiry. With the event enquiry function, users may retrieve the brief descriptions and some statistics of a selected event, and show them on the left frame of the current interface. Before starting to detect an event, the user can specify the directory to store the file keeping the detection results. The user can submit the detection task by simply clicking the "Run" button, whose progress is monitored by a progress bar. The detection process can be stopped anytime by pressing the 'Stop' button. To monitor event movement over locations, the user can specify the start time and ending time of his/her interested event time period. Similarly, for event summarization, the user can specify the target time period of the event, and find the representative words of the event messages by pressing the "Run" button on the event summarization page.

3.2.3 Crisis Situation. After the detection or summarization is finished, the crisis situation described using detected social posts or representative summaries will be returned and displayed on the interfaces. In addition to the information on crisis situations, the accuracy and the time cost of the online processing are also computed and shown on the interfaces.

3.2.4 Human Feedback. After the event detection is completed using ConTF/IDF with contexts, the HITL filter enables the user to label the detected posts with the least relevance scores which guide for the HITL filter to label the rest of posts returned by automatic event detection. The user-labelled social posts are submitted to the server, and the refined system labelling results are returned and displayed on the interface. This human interaction can be conducted by multiple loops until the returned results are satisfied by the user. In each loop, the evaluation results are reported on screen.

3.3 Functional Manager

3.3.1 Event Detection. Event detection is one of the core components in HSAP. To achieve high-quality detection, HSAP is equipped with an event detector and an event post filter.
Event Detector Using ConTF/IDF With Contexts. We design an event detector to discover the social messages on an event. When a user starts the application and switch to event detection page, HSAP provides a default accessed dataset, and shows the brief description on its contained event. If the user wants to understand a crisis situation, s/he can just select the corresponding dataset and click the button "Run", which invokes HSAP to automatically search the corresponding MongoDB document containing the message features of this event, and identify the potential event messages. The sub-event detection in our paper [7] is considered as the basic framework to find the event messages, due to its incorporating

database-oriented techniques with ConTF/IDF and context modelling robust to attribute uncertainty without costly learning. For more details, please refer to our previous work [7].

Event Post Filter Using HITL. We apply a binary neural network model to design an event post filter. The filter model maps the ConTF/IDF vector into 0,1 space to determine whether a message is relevant to the event. As in [6], the filter model consists of an input layer, five hidden layers, and an output layer. The input layer reshapes the matrix of n 50-dimensional ConTF/IDF vectors into its transpose to fit the input format of the model. Each hidden layer contains a fully connected layer with an activation function (ReLU) to learn the representation of messages. The output layer maps hidden features into the probabilities by a sigmoid function.

To Train the initial model, we collect messages which are detected by the event migration detection, and divide the dataset formed by these messages into 20% and 80% temporally. As in existing work [3], the former 20% is further divided into training, validation, and testing by 6:2:2. Then, we train the filter model based on their ground truth labels labelled by domain experts.

For user interaction, we adopt bi-directional labelling in HITL user interactions. HSAP provides the top k most relevant messages and the top k least relevant messages detected by automatic system. The user labels the messages with least relevance scores and submits the labelled results to the event post filter model for continuously learning. The updated filter model labels the rest of event message candidates returned by the event detector. This continuous learning is conducted to gradually improve the detection accuracy of HSAP. According to the Max Entropy strategy [4], labelling these messages brings more information to the model training.

We adopt incremental learning for filter maintenance. When new data is labeled, the system updates the filter model by incremental learning that is implemented keras library.

3.3.2 Event summarization. This component enhances user understanding in disaster situations by returning a set of keywords to describe a given event compactly and informatively. On the event summarization page, by default, the server considers the whole time period of a given dataset for summarization. When a user presses "Run" button, HSAP analyzes the event detection results returned by SOMA [1]. This process includes matrix factorization-based location estimation, L-LSTM-based summary candidate generation and summarization performance-based summary generator. For more details, please refer to our previous work [1]. To enable a more flexible analysis on crisis situation, we allow users to select their interested time period and the number of returned words for event summarization. Given the start and end dates of the user's interest and the word number, by pressing "Run", the user can find the event summary on the part within the selected dates.

3.3.3 Event Tracking. This component visualizes the event movement over space as shown on event movement interface and over time as shown on event evolution interface. The event movement over space shows how the event's route changes on the map within a selected time period. The event evolution searches the event summaries over different time periods over the whole timeline, which shows the event topic evolvement continuously. For details on event evolvement, please refer to our previous work [1].

4 DEMONSTRATION

We describe three scenarios of HSAP on Nepal earthquake 2015.

4.1 Event Detection

This subsection exhibit the steps of event detection in HSAP system. **Step 1: Event Detection Interaction Interface.** When a user starts the HSAP, s/he can choose to shift to the event detection interface as directed by the introduction on homepage, and choose the dataset to analyze (Figure 2(a)). Suppose that the user selected the dataset "Nepal Earthquake 2015". The messages with top relevant scores are listed in the "Event Posts" box in Figure 2(a).
Step 2: Detection Result labelling. Once the event is detected, the messages with the least relevant scores are output to message labelling page as shown in Figure 2(b). The user can navigate to message labelling page by clicking "Labelling Messages" button, and click "Relevant" or "Irrelevant" for the displayed messages.
Step 3: HITL Event Filtering. Then the user feedback from message labelling is passed to HITL event filter. The user can run HITL filter multiple loops to generate a refined event message filtering result list, and show the improvement of each loop compared with the results generated in the previous loop (Figure 2(c)).

4.2 Event Summarization

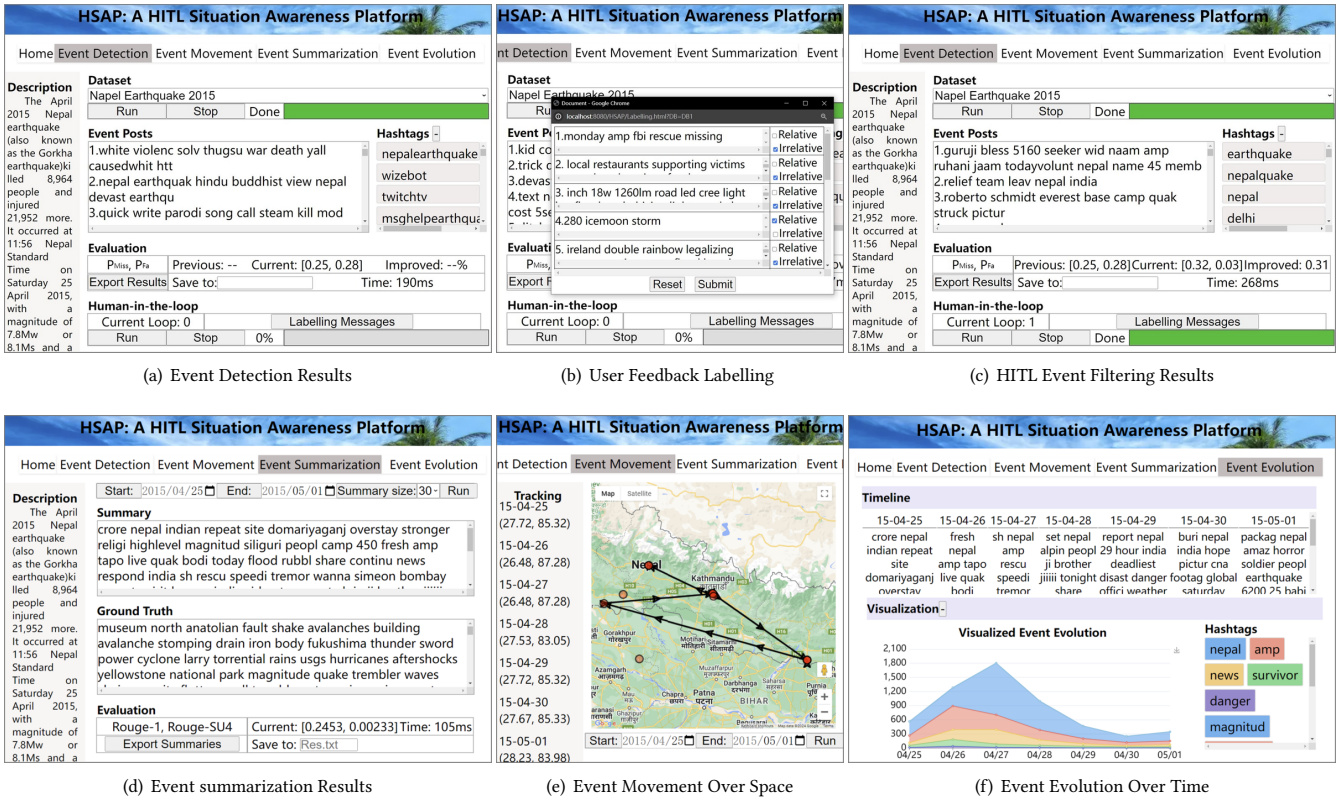
HSAP has the following steps on the event summarization.
Step 1: Interface of Event Summarization. When a user shifts to this interface, s/he can get the brief description of the event to be analyzed, and select the "Start" date and "End" date for the event part to be summarized and the number of words in summary to be returned (Figure 2(d)), and run the event summarization process.
Step 2: Display Event Summaries Then, the user can get the summary of the selected event that happened in a time period from the selected "Start Date" and "End Date". This summary has been generated by SOMA [1]. Meanwhile, HSAP displays the ground truth summary of the corresponding event part for comparison. Figure 2(d) shows the event summary of NepalEQuake 2015 generated by SOMA and its summary ground truth.

4.3 Event Tracking

We show the event movement over space and evolution over time.
Step 1: Event Movement Over Space. The user can navigate to Event Movement page to see the event spreading over different locations on a world map. The location movement of the selected Nepal Earthquake 2015 event is visualized in Figure 2(e).
Step 2: Event Evolution Over Time. The user can check the event topic changes over time on Event Evolution page. For the selected event, the Event Evolution page can visualize its evolution over the whole timeline using two ways as shown in Figure 2(f): the topic content change over different time periods as shown in the top part of the page, and the event hashtag frequency distribution over the timeline as shown by the visualized event evolution in the bottom part of the page.

5 EXPERIMENTAL EVALUATION

Evaluation Criteria. We evaluate the effectiveness of event detection using the probability of missed detection (P_{Miss}) and the probability of false alarm (P_{Fa}) [7]. For event summarization, we



(a) Event Detection Results (b) User Feedback Labelling (c) HITL Event Filtering Results

(d) Event summarization Results (e) Event Movement Over Space (f) Event Evolution Over Time

Figure 2: Situation awareness scenario for Nepal earthquake 2015.

evaluate effectiveness based on Rouge-1 and Rouge-SU4 [1]. To evaluate efficiency, we report the time cost. The system is tested over two real datasets [7]: (1) NepalEQuake contains 27,960 tweets generated from April 25 to May 1, 2015; (2) TexasFlood contains 14,874 tweets generated from May 22 to May 28, 2015.

Experimental Result. The evaluation results are shown in Table 1 and 2. With HITL filter, HSAP produces better effectiveness. In addition, HSAP can conduct these functions in real time.

	Detection		Detection+HITL		Summarization	
	P_{Miss}	P_{Fa}	P_{Miss}	P_{Fa}	Rouge-1	Rouge-SU4
NepalEQuake	0.25	0.28	0.32	0.03	0.25	0.002
TexasFlood	0.21	0.23	0.21	0.17	0.24	0.002

Table 1: Effectiveness evaluation results.

Time (ms)	Detection	Detection+HITL	Summarization
NepalEQuake	354.51	373.28	105.06
TexasFlood	272.96	141.86	98.26

Table 2: Efficiency evaluation results.

6 CONCLUSION

In this demo, we present a novel HSAP that provides a general platform for conducting various situation awareness tasks in natural disasters based on social media content and contexts. Several novel techniques have been implemented in the system, such as migrating event detection [7] and contextual event summarization [1]. In addition, HSAP further integrates the HITL technique that acts as an event filter for harnessing the human power in loops and enables the tracking of events over time and space.

ACKNOWLEDGMENTS

This work was supported by ARC Discovery Project DP200101175.

REFERENCES

[1] Xi Chen, Xiangmin Zhou, Jeffrey Chan, Lei Chen, Timos Sellis, and Yanchun Zhang. 2023. Complex Event Summarization Using Multi-Social Attribute Correlation. *IEEE Trans. Knowl. Data Eng.* 35, 11 (2023), 11180–11195.

[2] Xi Chen, Xiangmin Zhou, Timos Sellis, and Xue Li. 2018. Social event detection with retweeting behavior correlation. *Expert Syst. Appl.* 114 (2018), 516–523.

[3] Bradley Efron. 1992. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*. Springer, 569–593.

[4] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. 2017. Deep Bayesian Active Learning with Image Data. In *ICML*. 1183–1192.

[5] Rabindra Lamsal, Aaron Harwood, and Maria Rodriguez Read. 2023. Socially Enhanced Situation Awareness from Microblogs Using Artificial Intelligence: A Survey. *ACM Comput. Surv.* 55, 4 (2023), 77:1–77:38.

[6] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[7] Xiangmin Zhou and Lei Chen. 2022. Migrating Social Event Recommendation Over Microblogs. *Proc. VLDB Endow.* 15, 11 (2022), 3213–3225.