2018 9th ACM/IEEE International Conference on Cyber-Physical Systems

Towards a Cognitive Assistant System for Emergency Response

Sarah Preum*, Sile Shu†, Jonathan Ting*, Vincent Lin*, Ronald Williams†, John Stankovic*, Homa Alemzadeh†
*Computer Science, University of Virginia
†Electrical and Computer Engineering, University of Virginia
{preum, ss5de, jt4ue, vcl4fb, rdw, jas9f, ha4d}@virginia.edu

***Abstract*—This abstract presents our preliminary results on development of a cognitive assistant system for emergency response that aims to improve situational awareness and safety of first responders. This system integrates a suite of smart wearable sensors, devices, and analytics for real-time collection and analysis of in-situ data from incident scene and providing dynamic data-driven insights to responders on the most effective response actions to take.**

***Index Terms*—Cognitive assistant system, Medical emergency, Signal processing, Natural language processing, EMS**

I. Introduction

In an emergency situation, the first responders need to collect, aggregate, filter, and interpret information from different static and real time sources within a short time. Processing such a huge information load at the incident scene requires significant amount of human cognitive effort. Hence, we aim to develop a cognitive assistant for emergency response, that will improve situational awareness and safety of first responders by real-time collection and analysis of data from incident scene and providing dynamic data-driven feedback to them.The system will leverage the responder-worn devices and smart sensors to monitor activities and communications at the incident scene and aggregate this data with static data sources such as emergency response protocol guidelines to generate insights that can assist first responders with effective decision making and taking safe response actions.

Cognitive assistant systems have been applied to many different applications, including transportation and health [1]. Specifically, in [1] a Google glass based assistive system is developed to perform context-aware real-time scene interpretation by identifying objects, faces, and activities for people suffering from cognitive decline. Another relevant research direction is electronic reporting from incident scenes [2], [3]. ImageTrend [2] provides computer interfaces for data entry by EMS personnel. A mobile entry solution [3] facilitates data collection by dynamic customization of data fields. But these systems still rely on touch screen and messaging interfaces that are hard to manipulate in the midst of an incident. Instead, automatically extracting data from responders' speech can reduce the cognitive burden.

Hence, we propose a system that enables resilient data collection and analysis in a distributed environment consisting of responder-worn devices and an emergency response secured cloud platform. Figure 1 shows the overall architecture of the

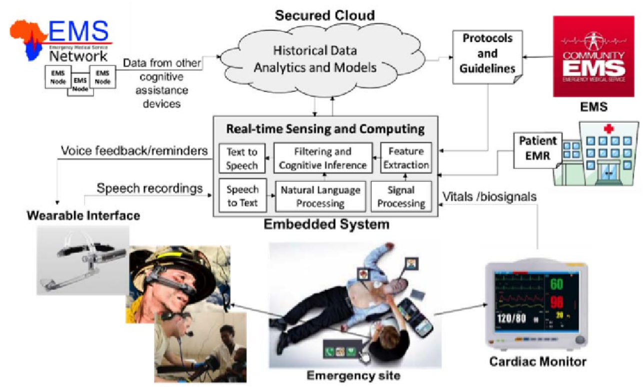

Fig. 1. Architecture of the Cognitive Assistant System

system. The system collects speech data from first responders in a noisy environment and processes the data to extract relevant information for decision making. The overall pipeline for processing data consists of speech-to-text conversion, information extraction (named entity recognition, temporal expression extraction, and medical concept extraction), and protocol modeling and execution.

In this abstract, we present the overall system architecture and following preliminary results on the implementation and performance evaluation of two modules in the pipeline. This is critical as errors in any module will propagate to the rest of the pipeline and affect the overall performance of the system and responders' decision making process.

- We quantitatively compare the performance of four off-the-shelf, state-of-the-art speech-to-text conversion tools under noise in terms of word error rate and computation time using data collected from real emergency scenes. These tools are Google Online API, Microsoft speech API, PocketSphinx, and IBM BlueMix API.
- We evaluate the performance of multiple state-of-the-art open-source natural language processing tools in extracting relevant named entities (e.g., location, organization, person), and temporal expressions from the text generated by the speech-to-text conversion step.

II. System Architecture

As shown in Figure 1 the proposed system consists of a suite of real-time sensing and computing modules for (i)

0-7695-6378-3/18/$31.00 ©2018 IEEE
DOI 10.1109/ICCPS.2018.00047

	Recall			Precision		
	Core NLP	Open NLP	Illinois	Core NLP	Open NLP	Illinois
Loc.	**0.368**	0.132	0.066	**0.903**	0.667	0.833
Per.	**0.934**	0.43	0.702	**0.831**	0.8125	0.685
Org.	**0.105**	0.053	0.092	**0.178**	0.16	0.079
Time	**0.769**	0.115	0.308	**0.241**	0.231	0.186

TABLE I
STANFORD CORENLP TOOL OUTPERFORMS THE OTHER TOOLS IN TERMS OF PRECISION AND RECALL IN EXTRACTING THE FOUR NAMED ENTITIES.

collecting and analyzing real-time data from heterogeneous sources, (ii) extracting knowledge from the data and (iii) providing feedback to the first responders to increase their situational awareness, decision making capacity and safety.

Firstly, real-time data will be collected from various data streams from incident scene, including voice recordings of observations made by the first responders (e.g., patient signs and symptoms) and communications with other responders and medical command, biosignal sensor data from cardiac monitors or other medical devices, and geolocation data. The collected data will be then converted into structured format and filtered by extraction of important features using speech-to-text, natural language processing, and biomedical signal processing algorithms. Next, the real-time data will be aggregated with the models learned from the historical data to infer actionable insights and provide suggestions to first responders on important actions to consider. Finally, the time-stamped structured data collected from incident scene will be stored into the secured cloud where multi-sensor fusion and learning algorithms will run to infer models based on historical data on past incidents, protocols and response guidelines, as well as data from other emergency response communities/agencies.

III. PRELIMINARY RESULTS

This section presents preliminary results from multiple stages of our information extraction pipeline, namely, speech-to-text conversion and keyword extraction.

A. Speech-to-Text Conversion

While some previous works compare different off-the-shelf speech-to-text conversion tools for dialogue systems [4], we evaluate the comparative performance of such tools under noise. Four off-the-shelf speech-to-text APIs are compared using transcripted audio data collected from emergency scenes in both noisy and noise-free environments. The performance is measured in terms of run time and word error rate (WER). Execution time is crucial in our application, as delay in converting audio data to text data can cause delay in the subsequent phases of the pipeline. Word error rate (WER) is the most widely used unit of measurement in determining the performance of speech to text programs. WER is calculated using the following formula: $WER = (I + D + S)/N$

Here, I, D, S, and N indicate the number of inserted words, the number of deleted words, the number of substituted words, and the total number of words in the text, respectively. The results of this analysis are presented Table II. The Google cloud API outperforms the other three APIs in terms of both WER and runtime.

Scenario	Metrics	PocketSphinx	Google	Microsoft	IBM
Noise-free	WER	0.80	**0.19**	0.24	0.45
	Runtime	2.48	**2.72**	3.42	5.34
Noisy	WER	1.05	**0.39**	0.62	0.89
	Runtime	3.41	**3.00**	3.38	9.84

TABLE II
COMPARING DIFFERENT SPEECH-TO-TEXT CONVERSION APIS IN TERMS OF WORD ERROR RATE (WER) AND RUNTIME

B. Named Entity and Temporal Expression Extraction

The following three named entities are extracted as potentially important keywords that can affect decision making of the first responders, (i) person (e.g., name of the victims and responders), (ii) location (e.g., incident location), and (iii) organization (e.g., hospital name). We also extract temporal expressions, including, timestamp, time of the day, duration of a medical procedure as they can aid inference and decision making. We compare three state-of-the-art open-source named entity recognizers, namely, Stanford CoreNLP, Apache OpenNLP, and Illinois tagger in terms of precision and recall.

We used the original noise-free text transcripts for named entity extraction that are used to evaluate the speech-to-text conversion stage for consistency. These transcripts are annotated by two human annotators for ground truth on words or phrases representing person name, location name, organization name, and temporal expression. The results are presented in Table I. The Stanford CoreNLP tool yields higher precision and recall for all four different types of entities than both OpenNLP and Illinois NER tool.

IV. FUTURE WORK

In future, subsequent stages of the pipeline will be implemented, such as, extracting relevant medical concepts and emergency terms from scene description, and radio calls and conversations of the responders. Later, these extracted concepts will be mapped to emergency protocol guidelines using logic programming based rule engine. The medical concepts identified in the previous steps will be fed as the inputs to the rule engine in real-time and the inferred actions will be provided to the responder to aid in decision making.

V. ACKNOWLEDGEMENT

This work was supported by the award 60NANB17D162 from U.S. Department of Commerce, National Institute of Standards and Technology (NIST).

REFERENCES

[1] K. Ha, Z. Chen, W. Hu, W. Richter, P. Pillai, and M. Satyanarayanan, "Towards wearable cognitive assistance," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 2014, pp. 68–81.

[2] ImageTrend, "Critical Care Solutions, ImageTrend," http://www.imagetrend.com/solutions-ems-critical-care/, 2017, [Online; accessed 10-Jan-2018].

[3] M. Fleshman, I. Argueta, C. Austin, H. Lee, E. Moyer, and G. Gerling, "Facilitating the collection and dissemination of patient care information for emergency medical personnel," in *Systems and Information Engineering Design Symposium (SIEDS), 2016 IEEE*, 2016, pp. 239–244.

[4] F. Morbini, K. Audhkhasi, K. Sagae, R. Artstein, D. Can, P. Georgiou, S. Narayanan, A. Leuski, and D. Traum, "Which asr should i choose for my dialogue system?" in *Proceedings of the SIGDIAL 2013 Conference*, 2013, pp. 394–403.