HHS Public Access

Author manuscript

Stud Health Technol Inform. Author manuscript; available in PMC 2020 July 14.

Published in final edited form as:

Stud Health Technol Inform. 2019 August 21; 264: 1586–1587. doi:10.3233/SHTI190547.

Named Entity Recognition in Prehospital Trauma Care

Greg M. Silverman[a,b], **Elizabeth A. Lindemann**[c], **Geetanjali Rajamani**[d], **Raymond L. Finzel**[e], **Reed McEwan**[a], **Benjamin C. Knoll**[b], **Serguei Pakhomov**[e], **Genevieve B. Melton**[b,c], **Christopher J. Tignanelli**[b,c,f]

[a]Academic Health Center – Information Systems, University of Minnesota, Minneapolis, Minnesota, USA

[b]Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, USA

[c]Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA

[d]Stanford University, Stanford, CA, USA

[e]College of Pharmacy, University of Minnesota, Minneapolis, Minnesota, USA

[f]Department of Surgery, North Memorial Health Hospital, Robbinsdale, Minnesota, USA

Abstract

Natural language processing (NLP) methods would improve outcomes in the area of prehospital Emergency Medical Services (EMS) data collection and abstraction. This study evaluated off-the-shelf solutions for automating labelling of clinically relevant data from EMS reports. A qualitative approach for choosing the best possible ensemble of pretrained NLP systems was developed and validated along with a feature using word embeddings to test phrase synonymy. The ensemble showed increased performance over individual systems.

Keywords

Emergency medical services; Labeling; Natural language processing

Introduction

More people die each year from trauma (10% of all deaths) than from malaria, tuberculosis and HIV/AIDs combined [1]. Inappropriate prehospital care is one of the largest contributors to preventable trauma mortality in the United States [2]. The use of natural language processing (NLP) named entity (NE) recognition (NER) for performance monitoring and quality improvement is a novel approach to bridge this gap. Early studies published in 2019 have begun utilizing NLP for prehospital stroke notes [3]. However, NLP for prehospital trauma notes represents a new domain of inquiry.

Address for correspondence: Christopher Tignanelli, M.D., ctignane@umn.edu.

NLP techniques hold promise for circumventing current limitations of discrete element documentation of EMS reports by reducing the effort of manual data abstraction [4]. The efficacy of how existing systems can best be utilized for notes with specific sublanguage characteristics of prehospital trauma is not well characterized and is the subject of this study.

We leveraged a small corpus of EMS reports of motor vehicle collisions (MVC) to develop an objective measure for creating an optimized NLP ensemble as discussed by Finley, et al. [5]. We also used word embeddings as a feature, as discussed by Turian, et al. [6]; and applied methods as discussed by Pakhomov, et al. [7], and Meng and Morioka [8] to use this feature as a test for phrase synonymy.

Methods

This study was a pilot to classify clinically relevant phrases related to prehospital trauma care using NER. Results were derived through an evaluation of four clinical NLP annotation systems, on their own, and as an optimized ensemble.

Data Sources

This study utilized de-identified North Memorial Health Hospital prehospital EMS MVC reports.

Creation of a Gold Standard Corpus—The development of a gold standard corpus consisted of two main parts: (1) schema creation, and (2) manual text annotation. 37 entities were identified based on clinical guidelines and the NEMSIS 3 standards [9] and iteratively incorporated into an annotation schema. Three trained annotators individually annotated 25 reports to establish inter-rater agreement (0.89 kappa, 99% agreement). Following this step, the remaining reports used for this study were manually annotated.

Creation of System Generated Annotations—Given the paucity of manual annotations for use in supervised training of a statistical model, we utilized the Artifact Discovery and Preparation Toolkit (NLP-ADAPT) [10], which included the clinical NLP annotator systems: cTAKES, CLAMP, BioMediCUS and MetaMap [11:14], to annotate our corpus of EMS MVC reports.

Evaluation Methods

We partitioned the gold standard annotations into a set of 10 reports to determine the best-at-task system annotation types. The remaining 112 reports were used for evaluation of the selected best-at-task system annotation types.

Matches between manual and system annotated reports were determined using a relaxed rule as noted by Finley, et al. [7]. Precision, recall, and F_1 score for each entity and system annotation pairing were calculated based on matches.

Best-at-task Evaluation—To compare NLP systems with respect to NE capture, we ranked each system annotation types for each entity using the three measures of NER performance shown in Figure 1. The geometric mean of the rankings was calculated to

classify the best-at-task system annotation type. The annotation type with the lowest geometric mean (GM) was deemed best-at-task.

For validation of best-at-task systems, we analyzed the entity Procedure Indication using the remaining 112 EMS reports by comparing them to their corresponding manually annotated reports. We then combined best-at-task systems as an ensemble to evaluate performance and further test how a word2phrase model as an additional feature affected recall [6:8].

A threshold of cosine distance of 0.5 was chosen after qualitatively evaluating several terms (e.g., “unresponsive,” “unconscious,” “agonal,” and “tachycardic”) and their resultant set when processed through the word2vec distance function [7]. Using string alignment methods [8], we used the Levenshtein edit distance (LD) to estimate degree of synonymy by identifying system and manual annotation match pairs on the resultant set with the lowest LD value.

Results

42 manual annotations pertained to the Procedure Indication entity in the set of 10 notes. Three best-at-task system annotation types were identified for this entity (Table 1).

931 manual annotations pertaining to Procedure Indication were noted in the 112 notes. Individually, the three systems performed similarly with respect to recall and precision compared to the gold standard. The union ensemble resulted in significant improved coverage (87%) (Table 2).

The top 2 best-at-task rankings were consistent during validation (Rank). As anticipated, the ensemble performed very well with respect to recall, but precision was still poor due to a high false positive rate. Also, the ensemble did not account for matching on synonymous phrases. For example, patients that are unconscious and have agonal respirations meet the procedural indication for intubation. Using our word2phrase resultant set, phrases identified in system and manual annotations were matched for synonymy (Table 3). We were able to identify 93% (104 of 112 notes) coverage (match) with mean LD value of 2.5 (range 0-19).

Conclusions

The present work represents one of the earliest NLP studies conducted in the prehospital trauma domain. Here we describe an approach to create an optimized NLP ensemble that when supplemented with a word2phrase model allows for over 90% NE capture. While these results are encouraging, the development of an extensive prehospital trauma corpus is paramount to facilitate development of models with improved precision, and further validation and extension of our methods.

Acknowledgements

NIH NCATS UL1TR002494 and U01TR002062, NIGMS R01GM120079, and AHRQ R01HS022085 and R01HS024532.

References

[1]. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. https://www.cdc.gov/injury/wisqars/overview/key_data.html. Accessed on December 10, 2017.

[2]. NHTSA Fatality Analysis Reporting System (FARS).

[3]. Garg R, Richards CT, Naidech A, Prabhakaran S, Predicting Cincinnati Prehospital Stroke Scale Components in Emergency Medical Services Patient Care Reports Using NLP and Machine Learning, Stroke 50 (2019), 296.

[4]. American College of Surgeons Committee on Trauma. Resources for optimal care of the injured patient. The Committee, Chicago, IL, 2014.

[5]. Finley GP, Liu H, Xu H, Melton GB, Pakhomov SVS, Using ensembles of NLP engines without a common type system to improve abbreviation disambiguation, AMIA Podium Abstract, 2017.

[6]. Turian J, Ratinov L, Bengio Y, Word representations: A simple and general method for semi-supervised learning, ACL, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010.

[7]. Pakhomov SVS, Finley GP, McEwan R, Wang Y, and Melton GB, Corpus domain effects on distributional semantic modeling of medical terms, Bioinformatics 32 (2016), 3635–3644. [PubMed: 27531100]

[8]. Meng F, Morioka C, Automating the generation of lexical patterns for processing free text in clinical documents, J Am Med Inform Assoc 5 (2015), 980–986.

[9]. NEMSIS Data Dictionary. NHTSA v3.4.0 https://nemsis.org/technical-resources/version-3/version-3-data-dictionaries/ Accessed on May 1, 2019.

[10]. Finzel RL, NLP-ADAPT. https://github.com/nlpie/nlp-adapt. Accessed on October 21, 2018.

[11]. BioMedICUS. https://github.com/nlpie/biomedicus Accessed on November 1, 2018.

[12]. MetaMap. https://metamap.nlm.nih.gov. Accessed on November 1, 2018.

[13]. cTAKES. http://ctakes.apache.org. Accessed on November 1, 2018.

[14]. CLAMP. https://clamp.uth.edu. Accessed on November 1, 2018.

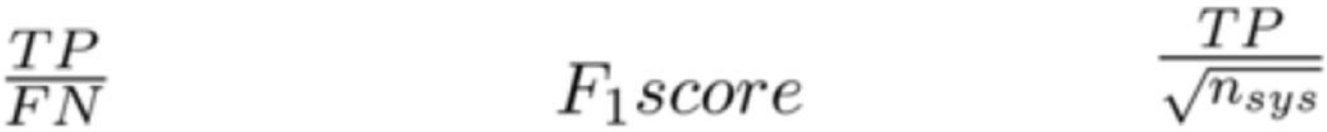

Figure 1 –.
NER Performance Measures; Abbreviations: TP, true positive; FN, false negative; n_{sys}, total system annotations

Table 1-

Best-at-task Annotation Types for Procedure Indication; Abbreviation: GM, Geometric Mean

System/Type	F_1	Precision	Recall	GM
CLAMP/Sentence	0.03	0.01	0.50	1.59
cTAKES/Sentence	0.02	0.01	0.52	2.15
MetaMap/Phrase	0.01	0.00	0.60	2.30

Table 2-

Best-at-task Validation for Procedure Indication Compared to Gold Standard with Final Rank

System/Type	F_1	Precision	Recall	Rank
CLAMP/Sentence	0.04	0.02	0.53	1
cTAKES/Sentence	0.03	0.02	0.54	2
MetaMap/Phrase	0.02	0.01	0.54	4
Ensemble	0.11	0.06	0.87	

Table 3-

Word2phrase and System Annotations; Abbreviation: w2p, word2phrase

Note	w2p phrases	w2p common token	Synonymous best-at-task annotation
1	agonal	shallow	tachypnic shallow
2	tachycardic	sbp	80’s sbp.
3	tachycardic	lungs	lungs clear bilat
4	unconscious	scene	alert on-scene