International Journal of Medical Informatics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Technical Notes

Integrating natural language processing expertise with patient safety event review committees to improve the analysis of medication events

Allan Fong[a,*], Nicole Harriott[b], Donna M. Walters[b], Hanan Foley[b], Richard Morrissey[b], Raj R. Ratwani[a,c]

[a] National Center for Human Factors in Healthcare, MedStar Health, 3007 Tilden St. NW, Suite 7M, Washington, D.C. 20008, USA
[b] Georgetown University Medical Center, 3800 Reservoir Rd NW, Washington, D.C. 20007, USA
[c] Department of Emergency Medicine, Georgetown University School of Medicine, 3900 Reservoir Rd. NW Washington D.C. 20007, USA

ARTICLE INFO

Keywords:
Patient safety events
Natural language processing
Medication
Machine learning
Visualization

ABSTRACT

Objectives: Many healthcare providers have implemented patient safety event reporting systems to better understand and improve patient safety. Reviewing and analyzing these reports is often time consuming and resource intensive because of both the quantity of reports and length of free-text descriptions in the reports.
Methods: Natural language processing (NLP) experts collaborated with clinical experts on a patient safety committee to assist in the identification and analysis of medication related patient safety events. Different NLP algorithmic approaches were developed to identify four types of medication related patient safety events and the models were compared.
Results: Well performing NLP models were generated to categorize medication related events into pharmacy delivery delays, dispensing errors, Pyxis discrepancies, and prescriber errors with receiver operating characteristic areas under the curve of 0.96, 0.87, 0.96, and 0.81 respectively. We also found that modeling the brief without the resolution text generally improved model performance. These models were integrated into a dashboard visualization to support the patient safety committee review process.
Conclusions: We demonstrate the capabilities of various NLP models and the use of two text inclusion strategies at categorizing medication related patient safety events. The NLP models and visualization could be used to improve the efficiency of patient safety event data review and analysis.

1. Introduction

Adverse drug events are a leading cause of preventable patient harm [1–3]. In an effort to reduce patient harm events associated with medications many healthcare systems have implemented patient safety event reporting systems to better identify safety hazards associated with pharmacy and medication administration, as well as other types of events [4,5]. The reporting systems generally provide a method for provider staff to submit a description of a safety hazard ranging from a near miss, where no patient harm occurred, to a serious safety event that resulted in patient harm.

Many patient safety event reporting systems contain hundreds to thousands of medication related events and have the potential to dramatically improve care and reduce adverse drug events [6,7]. However, there are several challenges associated with the data from these reporting systems [8]. Often, the data are difficult to interpret and act on because of the large number of reports, amount of free-text, and variability in category assignment by reporters.

In order to utilize the patient safety event data more rigorously many hospitals have created review committees, composed of clinicians focused on safety and quality, to review each event, categorize them appropriately to better understand trends, and develop solutions once trends are recognized. The committee review of the events is an incredibly labor intensive process given the large volume of reports generated each week. This difficulty is compounded in large healthcare systems where data from multiple hospitals need to be efficiently analyzed to understand overall patterns and trends across the system. Each report can take several minutes to initially review and then additional time during the committee meeting to further discuss.

Our goal is to develop a more efficient and streamlined method for categorizing patient safety event reports based on modeling the free-text of event reports to reduce the review time of the committee. We describe a collaborative effort in which informatics and safety science experts joined a clinical safety committee to develop an algorithmic

* Corresponding author.
E-mail addresses: allan.fong@medstar.net, af2012@gmail.com (A. Fong).

http://dx.doi.org/10.1016/j.ijmedinf.2017.05.005
Received 27 February 2016; Received in revised form 24 March 2017; Accepted 8 May 2017

approach to more automatically review and categorize medication events. The intent is to eventually develop a computational system that can categorize events in near real-time, hence reducing the time for committee review and expediting the process of identifying meaningful trends that can then be acted on to reduce adverse drug events. There are three main contributions of this case report. First, we develop and evaluate the performance of different modeling techniques to categorize four medication safety issues. Second, we evaluate model performance of two text inclusion conditions. The first condition includes only the brief factual description from medication related event reports as provided by the frontline staff member entering the report. The second combines both the brief factual description and resolution text, which is a short description typically provided by a manager that has reviewed the event report. Lastly we deploy the best models in an interactive visualization which categorizes reports in near real-time and allows users to provide feedback to the algorithm allowing for continued model training.

2. Background

2.1. Data elements in patient safety event reports

Patient safety event reporting systems are generally composed of structured and unstructured data [9,10]. When entering a report, the frontline staff selects a general category from a predefined list of categories (e.g. medication, fall, surgery) and a specific event type category. The reporter then enters a free-text description (brief factual description) of the safety hazards which can vary in length. Lastly, reports can sometimes be accompanied with additional free-text about how the event was resolved or addressed (resolution).

A major challenge with patient safety event reports is that the categories selected by reporters are often inaccurate and to fully understand the safety event, one has to read the free-text description. These category types are often ambiguous to the reporter and the reporter generally does not have the time to determine which category is the best fit for often complex events [11].

2.2. Clinical committee review

At MedStar Georgetown University Hospital a committee composed of physicians, nurses, pharmacists, and patient safety experts review each patient safety event report. The committee discusses each event, recategorizes the event if necessary, examines whether there are trends in the reports, and develops and implements potential solutions. Each meeting lasts an average of one to two hours, but committee members spend an average of two to four hours prior to the meeting manually reviewing events, categorizing, and identifying trends. Our goal is to develop a more efficient method for categorizing patient safety event reports to reduce the time investment of the committee. To do this, two data analytics experts (AF and RR), who have worked extensively with patient safety event data, joined the committee to learn about their classification process, develop natural language processing (NLP) algorithms, and work with the committee to validate and implement the algorithms [10–12]. Our focus was on medication events because these events are frequently reported, pose tremendous risk to patients, and require extensive time to review by the pharmacist and committee relative to other event types.

2.3. Natural language processing

Natural language processing (NLP) techniques have been previously used to explore and mine patient safety event reports. Examples include identifying latent themes and topics in reports, serious safety events, and health information technology related events [11,13,14]. Various statistical methods, each with different advantages and limitations, have been used to train and classify text [13]. However, previous research has primarily focused on assigning events to general categories such as computer related events or harm events [13,15]. Our focus is on developing algorithms to classify events into specific categories that are more actionable by the patient safety committee, such as medication workflow. For this application we evaluated support vector machines (SVM), decision trees (DT), and cosine similarity (COS) models to classify specific medication related patient safety events. In addition to the difference in specificity, previous work has generally considered reports as a single document either only considering the brief text or concatenating the brief and resolution text. It is unclear from previous work which strategy is more accurate for categorizing events in specific categories. We present an evaluation of these two different text inclusion strategies.

3. Method

3.1. Data sources

To train and validate our models, we started with 774 medication safety events that have been manually annotated and reviewed by the safety and quality committee (2 MDs, 1 PharmD, 3 RNs). Every report has a free-text brief factual description ranging from 9 to 424 words (77.9 mean, 59 median, 63.3 std). Six hundred ninety-five reports (90%) have resolution free-text averaging 50.6 words (29 median 29, 61.7 std) and were used for the model development efforts, Fig. 1. This study was approved by the MedStar Health Research Institute Institutional Review Board (protocol #2014-101).

3.2. Medication categories

We selected four medication safety event categories to model, described in Table 1. These categories tend to focus on workflow and decision making processes around medication safety events and were identified by the committee as promising categories for eventually introducing interventions to reduce the identified safety hazard. Of the 695 reports with brief and resolution text, 56 reports were categorized as pharmacy delivery delays, 68 were categorized as pharmacy dispensing errors, 108 reports were categorized as prescriber errors, and 64 were categorized as Pyxis discrepancy errors. The remaining 399 reports were categorized into other categories and included as negative cases in the model development.

3.3. Approach

We developed classification models for each of the four categories (pharmacy delivery delays, pharmacy dispensing errors, prescriber errors, and Pyxis discrepancy errors) in Table 1. This was done by using the identified events for the error type being modeled as positive cases and using all the remaining reports as negative cases, including the other three error types and the 399 “other” categories. As an example, for the prescriber error model the 108 prescriber error reports served as positive instances and the remaining reports (587) served as negative prescriber error reports for training and testing of the prescriber error model. For each category, we first set aside 20% of the annotated reports for testing, Fig. 1. For each category, 20% of the test set was randomly selected from the corresponding positive instances of the respective category and the remaining 80% of the test set was randomly selected from the respective negative instances. This semi-random approach was to ensure that the proportion of positive reports in the training sets were the same as in the test sets. For prescriber error, 87 positive instances and 470 negative instances were used for all training models and 21 positive instances and 117 negative instances were used for testing all models. For pharmacy dispensing error, 55 positive and 502 negative instances were used for training and 13 positive and 125 negative instances were used for testing. For Pyxis discrepancy, 52 positive and 505 negative instances were used for