2014 IEEE 27th International Symposium on Computer-Based Medical Systems

Automated Classification of Computer-based Medical Device Recalls:
An Application of Natural Language Processing and Statistical Learning

Homa Alemzadeh, Raymond Hoagland, Zbigniew Kalbarczyk, Ravishankar K. Iyer

Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801
{alemzad1, hoaglan2, kalbarcz, rkiyer}@illinois.edu

Abstract— **This paper presents MedSafe, a framework for automated classification of computer-based medical device recalls. The data is collected from the U.S. Food and Drug Administration (FDA) recalls database. We combined techniques in natural language processing and statistical learning to automatically identify the computer-related recalls, by interpreting the natural language semantics of recall descriptions. We evaluated MedSafe on over 16K recall records submitted to the FDA between years 2007-2013.**

Keywords- Medical Devices; FDA Recalls; Computer-related Failures; Natural Language Processing; Statistical Learning.

I. INTRODUCTION

The recalls related to computer-based medical devices constitute nearly 22.9 percent of all the recalls reported to the FDA [1]. The analysis of these recalls provides valuable insights on the causes of computer failures in medical devices and on how the design of future systems could be improved to prevent adverse impacts on patients.

The FDA's "Recalls" database is a public database of medical device recalls, reported by the manufacturers, distributors, or other responsible parties since 2002. Each record in the database contains the information on a recalled device such as the *Product Name*, *Recalling Firm*, *Quantity in the Commerce* (i.e. number of devices on the market), *Reason for Recall,* and recovery *Action* taken to correct the device or remove it from the market [2].

The main challenge in analysis of the recalls data is that important fields in the records, including the *Product Name*, *Reason for Recall,* and *Action,* are entered by the human reporters in an unstructured text format. Therefore, the identification of causes of failures requires semantic interpretation of the natural language text.

Previous work on analysis of the FDA databases used either keyword searching or manual reviewing of recall descriptions to extract the software-[3][4][5], computer- [1], and security-[6] related problems based on subsets of FDA data. These approaches require significant amount of human effort and still may not produce accurate results due to human mistakes or inadequate list of keywords.

We present MedSafe, a framework for automated identification of computer-based medical device recalls. MedSafe uses natural language parsing in conjunction with statistical learning to extract relevant features from the *Reason for Recall* descriptions in order to classify the recalls into computer- vs. non-computer categories. The proposed approach enables automated analysis of larger sets of recalls and provides a way to measure impact of computer failures in different device categories and in terms of number of devices on the market that were affected by the recalls.

II. RECALLS DATA ANALYSIS

MedSafe goes through two main steps for extraction and classification of computer-related recalls:

Recall Data Extraction: First, all the recall records submited for a desired period of time are downloaded from the database by web crawling the FDA online database and parsing the HTML files of the records. Then the number of devices (*Device Quantity*) affected by each recall record are extracted by parsing the *Quantity in Commerce* field, using regular expressions and heuristic rules.

Many of the recall records represent the same failure event reported for different devices manufactured by the same company. So the unique recall events related to the same failures are extracted through coaleasing the recall records with the same *Event ID*. The total number of devices affected by each recall is calculated by summing up the numeric *Device Quantity* values extracted for each record.

Evaluation: The results of this step were evaluated by manually reviewing recalls submitted to the FDA between years 2007-2013. MedSafe achieved 97.3% accuracy in calculating the total number of devices affected by the recalls. We found that during the study period, a total of 16,881 recall records were reported to the FDA, from which 6,864 (40.7%) were unique recall events.

***Recall Classification*:** We define a *computer-related* recall as an event causing a computer-based medical device to function improperly or present harm to patients or users due to failures in device's software, hardware, I/O, or battery [1]. MedSafe uses a set of manually classified computer-related recalls from our previous study [1], as a training set to automatically classify the recalls into computer- vs. non-computer classes.

First, the *Reason for Recall* field of each recall event is normalized by removing the punctuations and English stop words and converting text into lowercase. Next, the text is tokenized into words and a part-of-speech tagger is used to extract nouns, adjectives, and verbs from the tokens. Then

1063-7125/14 $31.00 © 2014 IEEE
DOI 10.1109/CBMS.2014.134

in order to select the most relevant features (set of keywords) for classification of the recalls, we used a mutual information (MI) metric [7] to measure how much the presence or absence of each keyword in the *Reason for Recall* field of a recall event contributes to the classification of that recall into the computer-related class. For each keyword *k*, the MI metric is calculated as follows:

$$MI(R_k, C_r) = \sum_{i,j\in(0,1)} P(R_k = i, C_r = j)\, log_2 \frac{P(R_k = i, C_r = j)}{P(R_k = i).P(C_r = j)} \quad (1)$$

where the R_k and C_r are binary indicator random variables, respectively representing whether the recall event *R* contains the keyword *k* ($R_k = 1$), and whether the recall event *r* is in the computer-related class ($C_r = 1$).

The keywords are then sorted according to their *MI* scores and the top half of the list is used as the set of features for classification of the recalls. MedSafe uses a multinomial Naïve Bayes classifier *[*7] to calculate the probability of a recall *R* being computer-related, as follows:

$$Class(R) \propto \underset{c\in\{0,1\}}{argmax}\left[log\, P(c) + \sum_{1\le i\le M} logP(k_i|c)\right] \quad (2)$$

where $P(c)$ is the prior probability of a recall being in computer- *(c = 1)* or non-computer-related *(c = 0)* class, *M* is the total number of keywords extracted in the feature selection phase, and $P(k_i|c)$ is the conditional probability of keyword k_i appearing in the *Reason for Recall* field of a computer-related recall. All the probabilities in equations (1) and (2) are calculated using maximum likelihood estimation (MLE) from the data.

Evaluation: In order to assess the performance of the classifier, we performed a ten-fold cross-validation, using 4,398 manually labeled recall events from 2007-2011 [1]. Each subset of 439 recalls was used as the test data, and the remaining (3,959) recalls were used for feature selection and training. The standard metrics, including *sensitivity*, *specificity*, and *F-Score* (considers both *precision* and *recall* metrics with a value of 100% indicating perfect results) were used for evaluation of the results and were averaged over the ten runs of experiments. An average sensitivity and specificity of 88.2% and F-Score of 77.9% was achieved.

In order to evaluate the classification results on the new data, we used the manually-labeled data from 2007-2011 as a training set to classify the recalls submitted for the years 2012 and 2013. Our experiments show that of 5,011 recall records submitted during 2012-2013, 2,466 (49.2%) were unique recall events, from which 634 (25.7) were computer-related. By manual review of these results, we found that MedSafe identified the computer-related recalls with a sensitivity of 93.8% and a specificity of 95.8%.

Recall Analysis Results: Figure 1 shows the distribution of computer- vs. non-computer-related recall events for 2007-2013 (after relabeling the recalls that were incorrectly classified by MedSafe). The number of computer-related recalls has been increased since 2007, but has almost stayed constant during 2011-2013. The secondary axis shows the total number of devices affected by the recalls per year.

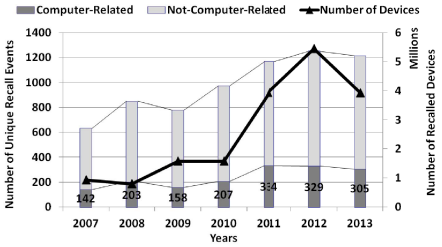

Figure 1. Number of computer-related recalls and recalled devices

Figure 2 shows the 7 medical device categories that had the highest number of computer-related recalls reported during 2007-2013. About 58% of the computer recalls were related to Radiology and Cardiovascular devices, such as automated external defibrillators and linear accelerators.

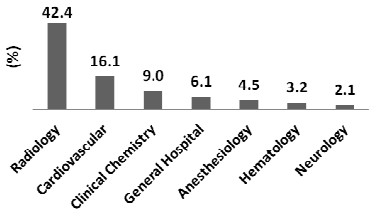

Figure 2. Computer-related recalls across different device categories

III. CONCLUSIONS

The proposed approach for automated identification of computer-related recalls enables faster analysis of larger sets of recalls data by assisting the manual reviewing process. Future work will focus on improving the classification results and automated identification of fault classes, failure modes, and action categories from recall descriptions. More advanced feature extraction techniques such as considering n-grams and including features related to *Product Name* and *Action* fields might improve the results in the future.

REFERENCES

[1] H. Alemzadeh, et al., "Analysis of Safety-Critical Computer Failures in Medical Devices," *IEEE Security and Privacy, vol. 11, no. 4, pp. 14-26, July/August* 2013.

[2] U.S. Food and Drug Administration, "Medical Device Recalls," http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRES/res.cfm.

[3] D. Wallace and D. Kuhn, "Failure modes in medical device software: an analysis of 15 years of recall data," *International Journal of Reliability Quality and Safety Engineering*, vol. 8, pp.351–372, 2001.

[4] Z. Bliznakov, G. Mitalas, and N. Pallikarakis, "Analysis and classification of medical device recalls," in *World Congress on Medical Physics and Biomedical Engineering 2006*. Springer, 2007, pp. 3782–3785.

[5] H. Yang and W. Hyman, "An analysis of software-related recalls of medical devices," *Journal of Clinical Engineering*, vol. 35, no. 3, p. 153, 2010.

[6] D.B. Kramer et al., "Security and Privacy Qualities of Medical Devices: An Analysis of FDA Postmarket Surveillance,"*PLoS ONE*, vol. 7, no. 7, 2012, pp. 1–7.

[7] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. vol. 1. Cambridge, MA: Cambridge University Press, 2008.