

Efficient Classification of Clinical Reports Utilizing Natural Language Processing

Efsun Sarioglu, MS

Department of Computer Science
The George Washington University
801 22nd Street NW
Washington, DC, USA
efsun@gwu.edu

Kabir Yadav, MDCM MS

Department of Emergency Medicine
The George Washington University
2150 Pennsylvania Ave NW, Suite 2B-408
Washington, DC, USA
kyadav@gwu.edu

Hyeong-Ah Choi, PhD

Department of Computer Science
The George Washington University
801 22nd Street NW, Room 713
Washington, DC, USA
hchoi@gwu.edu

Abstract

The recent emphasis on health information technology has highlighted the importance of leveraging the large amount of electronic clinical data to help guide medical decision-making. Developing such clinical decision aids requires manual review of many past patient reports in order to generate a good predictive model. In this research, we investigate classification of clinical reports using natural language processing (NLP). The proposed system uses NLP to generate structured output from computed tomography (CT) reports and then machine learning techniques to code for the presence of clinically important injuries for traumatic orbital fracture victims. Our results show that NLP improves upon raw text classification results.

Introduction

Medical Language Extraction and Encoding(MedLEE) is the one of the most widely used NLP software in the medical research community(Meystre et al. 2008), and has successfully interpreted findings from raw text procedure reports such as head CT imaging for stroke and chest radiography for pneumonia (Elkins et al. 2000; Hripcsak et al. 1999). A strength of the structured output of MedLEE that it codes its findings to Unified Medical Language System (UMLS) (U.S. National Library of Medicine 2012) Concept Unique Identifiers (CUIs). These CUIs add more semantics and fix the problems that could be caused by using different synonyms.

The primary approach of our system, as seen in Figure 1, is to take these patient reports as input to the MedLEE to tag them with UMLS CUIs and modifiers that show the probability and temporal status. After this tagging process, the output is filtered to exclude findings with low certainties or findings linked with patient's history or future modifiers. These raw text files and NLP findings are combined with their associated outcomes, and passed to the data mining tool Weka 3.7.5 (Hall et al. 2009) for classification. Classification

Copyright © 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

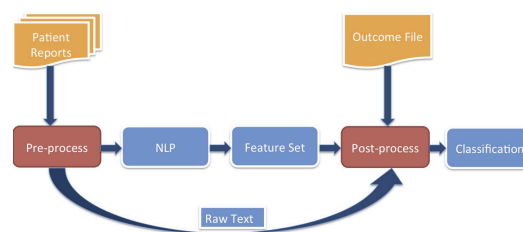


Figure 1: System Overview

of patient reports using MedLEE has previously been shown to be promising (Sarioglu, Yadav, and Choi 2012) .

Table 1: Classification using Text and NLP features

	Decision Tree			SVM		
	Text	NLP		Text	NLP	
		All	Filtered		All	Filtered
Precision	0.947	0.955	0.97	0.959	0.968	0.971
Recall	0.948	0.956	0.97	0.960	0.969	0.971
F-Score	0.948	0.955	0.97	0.959	0.968	0.971

Methodology

A detailed description of how MedLEE works has been previously described (Friedman 2000). Briefly, for a given a block of raw text, MedLEE preprocessor splits the text into sentences and does a lexical lookup to identify words, phrases, sentences, and abbreviations. Its parser then utilizes a grammar to recognize syntactic and semantic patterns and generates intermediate forms consisting of primary findings and different types of modifiers. If the initial parsing effort fails, error recovery procedure is followed so that a partial analysis is attempted by using alternative strategies. A composer module then composes individual words into phrases using a table. Finally, an encoder module maps words/phrases into codes using a table (Friedman et al. 1994). To adapt MedLEE for other clinical investigations, its lexicon, abbreviations and section names can be changed to reflect the terms and organization seen in the documents to

Raw Text

Impression: Right lamina papyracea fracture. No evidence of entrapment.

MedLEE Output

```
<sectname v = "report impression item"></sectname>
<sid idref = "s7"></sid>
<code v = "UMLS:C0016658_Fracture"></code>
<problem v = "entrapment" code = "UMLS:C1285497_Entrapment (morphologic abnormality)">
  <certainty v = "no"></certainty>
</problem>
```

Figure 2: MedLEE output

be interpreted. Figure 2 shows a sample input and its corresponding output from MedLEE. In this example, MedLEE successfully identifies the presence of a fracture and absence of entrapment with their corresponding CUIs.

After considering many different classification algorithms, decision tree and Support Vector Machines (SVM) were chosen. Decision tree is preferred due to its explicit rule based output that can be easily evaluated for content validity, whereas SVM is known to perform well in text classification tasks (Joachims 1998). SVM is also known to be robust to overfitting (Sebastiani 2002) and they are much faster than decision trees.

Experiments

Retrospective chart review of consecutive CT imaging reports for patients suffering traumatic orbital injury was obtained over 26 months from two urban hospitals (Yadav et al. 2012). Using conventional methods, the investigators prospectively collected clinical data and outcomes on over 3,000 patients, including CT imaging reports. Staff radiologists dictated each CT report and the outcome of acute orbital fracture was extracted by a trained data abstractor. A random subset of 511 CT reports were double-coded, and inter-rater analysis revealed excellent agreement with Cohen's kappa of 0.97 (95% CI 0.94-0.99). Among the 3,705 reports, 3,242 had negative outcome while 463 had positive.

During preprocessing, all protected health information were removed to meet Institutional Review Board requirements. Medical record numbers from each report were replaced by observation numbers, which are sequence numbers automatically assigned to each report.

After MedLEE is run, the output was further processed to include only the relevant information. Several approaches for extracting features from MedLEE output were considered. The first method extracts the problems with body locations with their corresponding UMLS codes that MedLEE had found. In the second approach, only the valid findings are used where status and certainty modifiers are verified.

The raw text of the reports and feature sets from NLP output were compiled into individual files in attribute relation file format (arff), where each line represents one report with its associated outcome. This file can then be loaded into Weka, where it is converted into a word vector representation and classified using the algorithms mentioned above. Classification results using raw text and NLP features are compared in Table 1. Using features from NLP output gets better performance than using the raw text of the reports. Between the two post-processing approaches to creat-

ing NLP feature sets, filtered codes produce slightly better results. NLP classification results are similarly excellent using either decision trees or SVM.

Conclusion

In this research, NLP is used to improve raw text classification of Emergency Department CT reports. Using NLP features improves classification results compared to the use of raw text. Within NLP features, filtering the codes using modifiers produces the best performance.

References

- Elkins, J. S.; Friedman, C.; Boden-Albala, B.; Sacco, R. L.; and Hripcsak, G. 2000. Coding neuroradiology reports for the Northern Manhattan Stroke Study: a comparison of natural language processing and manual review. *Comput Biomed Res* 33(1):1–10.
- Friedman, C.; Alderson, O. P.; Austin, H. J.; Cimino, J. J.; and Johnson, B. S. 1994. A General Natural-Language Text Processor for Clinical Radiology. *Journal of the American Medical Informatics Association* 1(2):161–174.
- Friedman, C. 2000. A broad-coverage natural language processing system. *Proc AMIA Symp* 270–274.
- Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* 11(1):10–18.
- Hripcsak, G.; Kuperman, G.; Friedman, C.; and Heitjan, D. 1999. A reliability study for evaluating information extraction from radiology reports. *Journal of the American Medical Informatics Association* 6(2):143.
- Joachims, T. 1998. Text categorization with support vector machines: Learning with many relevant features.
- Meystre, S. M.; Savova, G. K.; Kipper-Schuler, K. C.; and Hurdle, J. F. 2008. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform* 128–144.
- Sarioglu, E.; Yadav, K.; and Choi, H.-A. 2012. Classification of Emergency Department CT Imaging Reports using Natural Language Processing and Machine Learning. In *AMIA Proceedings: Summits on Clinical Research Informatics*, 150.
- Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Comput. Surv.* 34(1):1–47.
- U.S. National Library of Medicine. 2012. UMLS Home Page. <http://www.nlm.nih.gov/research/umls/>.
- Yadav, K.; E, C.; JS, H.; Z, A.; V, N.; and et al., G. P. 2012. Derivation of a clinical risk score for traumatic orbital fracture. In Press.