Clinical Text Analysis Using Interactive Natural Language Processing

Gaurav Trivedi
Intelligent Systems Program
University of Pittsburgh
4200 Fifth Avenue, Pittsburgh,
PA 15260
trivedigaurav@pitt.edu

IUI'15 Companion, Mar 29–Apr 01, 2015, Atlanta, GA, USA.
ACM 978-1-4503-3308-5/15/03.
http://dx.doi.org/10.1145/2732158.2732162

Abstract
Natural Language Processing (NLP) systems are typically developed by informaticists skilled in machine learning techniques that are unfamiliar to end-users. Although NLP has been widely used in extracting information from clinical text, current systems generally do not provide any provisions for incorporating feedback and revising models based on input from domain experts. The goal of this research is to close this gap by building highly-usable tools suitable for the analysis of free text reports.

Author Keywords
Electronic medical records; visualization; interactive machine learning.

ACM Classification Keywords
H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous; I.2.7 [Natural Language Processing]: Text analysis.

Context and Motivation
Electronic Health Records (EHRs) are organized collections of information about individual patients. Although present-day EHRs serve as invaluable data sources to the clinicians in their day-to-day work, research communities are still working towards exploiting the wealth of data contained in them for clinical research and

data-driven quality measures. NLP methods have been widely applied to clinical text [6, 12, 8]. However, due to the complexity of clinical text, the accuracy of these techniques can vary [8]. There is a need to focus on development of NLP systems that are not only generalizable for use in different tasks but are also usable with reduced dependence on NLP developers.

My master's thesis work aims to explore the design of an interactive tool to review NLP outcomes and build learning models iteratively. I believe that by leveraging visualizations to help the users understand the NLP models and allowing them to make necessary corrections, such a tool can provide a feedback loop that may result in improved accuracy of the NLP models. The two main bottlenecks from a user-interface building perspective are in receiving annotations and labeling examples from the user, and allowing them to make changes to the models. I have proposed novel interaction methods and visualizations techniques to address these challenges.

Related Work

Visualization and Sensemaking

Visualization tools such as WordTree [10] and Tiara [11] help in providing a visual summary of large amount of text data. While Tiara focuses on content evolution of each topic over time, WordTree provides a keyword in context method of exploring the text. Other tools such as Jigsaw [9] help users interpret document collections by visualizing documents in multiple graph, cluster and list -type views. Our task in reviewing clinical documents is somewhat different, in that our goals are to understand common textual patterns and to use those patterns to improve NLP models. We have adapted elements of these views - in particular, WordTree's phrase view and Jigsaw's document view to support our goals.

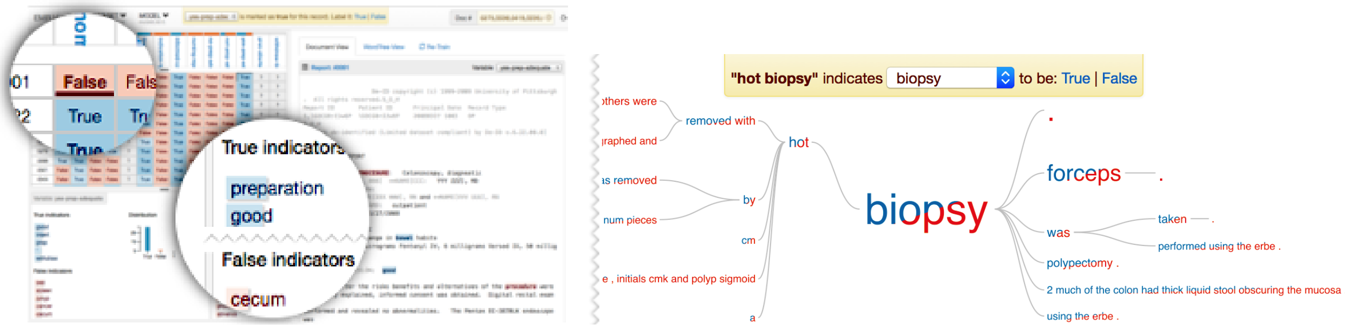
Interactive Machine Learning

There have been many efforts to develop user-centric tools for machine learning and NLP making it easier for the end users to use them. D'Avolio et. al. [4] have described a prototype tool with a user interface to configure the machine learning algorithms for NLP and export their results, thus solving a part of our problem.

Other efforts have explored the development of interactive machine learning systems that learn iteratively from their end-users. These efforts describe learning systems whose output is used by the end-user to further inform the system, creating a closed loop that can be used to build continuously improving predictive models. Examples include applications in interactive document retrieval [7], image segmentation [5], bug triaging [1] etc.

Research Methods and Progress

I have implemented an interactive web-based tool that facilitates both the review of structured values extracted from clinical reports and also has a provision for expert feedback that can be used to improve the accuracy of NLP models. The tool consists of three main views: (a) The *grid view* (Figure 1(a)) shows boolean variables extracted from the text in columns and individual documents in rows. It provides an overview of NLP results and also has links to view the full-text of the document represented in each cell; (b) The *WordTree view* (Figure 1(b)) provides the ability to search for and explore word sequence patterns found across the documents in the corpus and provide feedback that will be used to retrain NLP models; (c) The *retrain view* (not shown) lists user-provided feedback, including any potential inconsistencies, and specifies changes in variable assignments due to retraining.



(a) *Grid view* - showing the NLP results. Below it, we have a list of words that serve as indicators for the prediction values, which are also highlighted in the document text on the right side.

(b) *WordTree view* - All sentences containing the word in the center are shown through a WordTree [10]. Our tree design also shows the class split using color gradients. The top yellow bar is used for providing feedback.

Figure 1: Screenshots of two of the main views from the prototype tool. Magnified portions are shown inside the circles.

To demonstrate the use of the tool, I have used a dataset of colonoscopy reports by building on the work done in [6]. The user-interface has been built using the Angular (`angularjs.org`) , D3 [2] and jQuery (`jquery.com`) Javascript frameworks and libraries. Our NLP pipeline uses a bag of words feature-set and a support vector machine (SVM) learning model, but it can also be extended for use with different models and complement other existing tools.

To gain insight into the usability factors for the prototype, I conducted a formative user study with 5 clinical researchers, who reviewed colonoscopy reports using it. The protocol followed the think aloud method. The System Usability Scale [3] was used to assess subjective reactions to the tool.

After revising the tool based on feedback from this user study, I will conduct an empirical evaluation of the tool with the objective of testing the following hypotheses:

- H1: The interactive tool will facilitate the review of clinical text and the building of NLP models.
- H2: Visual presentation and interactive feedback components (word tree, interactive grid and retrain views) will allow quicker and more accurate completion of the task of building NLP models.
- H3: Manual review supported by this tool may enable the use of small training sets to quickly converge on highly accurate models.

Selection of comparison tools and recruitment present challenges in the design of the empirical evaluation. As participants in our think-aloud studies were not aware of alternative tools for the NLP review task, comparison to current systems is not possible. Comparison of design alternatives (such as with or without the Word Tree) might be an option. Snowball sampling may not be sufficient for identifying the 20-30 domain experts that we expect would be needed for statistically compelling results. Given these constraints, we are considering a descriptive study that would explore the efficacy of the system for building highly accurate models using small training sets.

Acknowledgment

I thank my adviser, Dr. Harry Hochheiser, and collaborators on the project, Phuong Pham, Dr. Wendy Chapman, Dr. Rebecca Hwa and Dr. Janyce Wiebe. This research is supported by NIH grant 5R01LM010964.

References

[1] Amershi, S., Lee, B., Kapoor, A., Mahajan, R., and Christian, B. Cuet: Human-guided fast and accurate network alarm triage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 157–166.

[2] Bostock, M., Ogievetsky, V., and Heer, J. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2011).

[3] Brooke, J. SUS: a quick and dirty usability scale. In *Usability evaluation in industry*, P. W. Jordan, B. Weerdmeester, A. Thomas, and I. L. Mclelland, Eds. Taylor and Francis, London.

[4] D'Avolio, L. W., Nguyen, T. M., Goryachev, S., and Fiore, L. D. Automated concept-level information extraction to reduce the need for custom software and rules development. *Journal of the American Medical Informatics Association 18*, 5 (2011), 607–613.

[5] Fails, J., and Olsen Jr, D. Interactive machine learning. In *Proceedings of the 8th international conference on Intelligent User Interfaces*, ACM (2003), 39–45.

[6] Harkema, H., Chapman, W. W., Saul, M., Dellon, E. S., Schoen, R. E., and Mehrotra, A. Developing a natural language processing application for measuring the quality of colonoscopy procedures. *JAMIA 18*, Supplement (2011), 150–156.

[7] Heimerl, F., Koch, S., Bosch, H., and Ertl, T. Visual classifier training for text document retrieval. *Visualization and Computer Graphics, IEEE Transactions on 18*, 12 (2012), 2839–2848.

[8] Nadkarni, P. M., Ohno-Machado, L., and Chapman, W. W. Natural language processing: an introduction. *Journal of the American Medical Informatics Association 18*, 5 (2011), 544–551.

[9] Stasko, J., Görg, C., and Liu, Z. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization 7*, 2 (Apr. 2008), 118–132.

[10] Wattenberg, M., and Viegas, F. B. The word tree, an interactive visual concordance. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1221–1228.

[11] Wei, F., Liu, S., Song, Y., Pan, S., Zhou, M. X., Qian, W., Shi, L., Tan, L., and Zhang, Q. Tiara: A visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, ACM (New York, NY, USA, 2010), 153–162.

[12] Ye, Y., Tsui, F. R., Wagner, M., Espino, J. U., and Li, Q. Influenza detection from emergency department reports using natural language processing and bayesian network classifiers. *Journal of the American Medical Informatics Association* (2014).