Chapter 7.16
Analysing Clinical Notes for Translation Research:
Back to the Future

Jon Patrick
The University of Sydney, Australia

Pooyan Asgari
The University of Sydney, Australia

ABSTRACT

There have been few studies of large corpora of narrative notes collected from the health clinicians working at the point of care. This chapter describes the principle issues in analysing a corpus of 44 million words of clinical notes drawn from the Intensive Care Service of a Sydney hospital. The study identifies many of the processing difficulties in dealing with written materials that have a high degree of informality, written in circumstances where the authors are under significant time pressures, and containing a large technical lexicon, in contrast to formally published material. Recommendations on the processing tasks needed to turn such materials into a more usable form are provided. The chapter argues that these problems require a return to issues of 30 years ago that have been mostly solved for computational linguists but need to be revisited for this entirely new genre of materials. In returning to the past and studying the contents of these materials in retrospective studies we can plan to go forward to a future that provides technologies that better support clinicians. They need to produce both lexically and grammatically higher quality texts that can then be leveraged successfully for advanced translational research thereby bolstering its momentum.

INTRODUCTION

This volume gives a picture of the current progress in using Natural Language Processing to provide better access to written materials that are on the whole published under formal assessments and processes. Such materials are more amenable to different NLP tasks than the genre of content produced by

clinicians working at the point of care with very sick and dying patients. The point-of-care texts are distinctly different in two important ways - *weaker adherence to grammatical structure*, and *the complexity and veracity of* lexical form. A further difference is in the objectives of the processing tasks themselves. The papers of this volume have defined the necessary task as that of information retrieval in and of itself. Processing clinical notes has a much larger number of objectives which makes it both more complex but also provides a greater number of opportunities to be creative and contribute to enhancements to the field of health care. Translational research, or more colloquially known as "bench to bed" research, is looming to be by far the most important field for improving our health care. This is the process of linking together the research performed in the laboratory to the processes and treatments conducted at the bedside. Much of this work is seen as personalized medicine that will produce drugs targeted at an individual's personal genome, however research into this ideal healthcare is highly restricted without technology that enables linking of the patient records created at the point-of-care with the models of drug-genome interaction built in the laboratory. NLP is a most important bridge between these two end points in the search for better healthcare. Without effective and accurate processing of the point-of-care records to provide evidence of the effectiveness of drugs and the course of diseases in patients under drug treatment there can be no development of evidence based drug treatment, and translational research will progress more slowly than it would otherwise.

This paper presents some of the models we have developed for representing the different processing needs of clinical staff and describes the lessons we have learned from problems encountered in analysing a large corpus of 44 million words of clinical notes.

BACKGROUND

The task of performing natural language processing over clinical notes goes back to 1983 Chi, E. C., Sager, N., Tick, L. J., & Lyman, M. S.; Gabrieli, E. R., Speth, D. J., 1986) with the work Chi, Sager, Tick and Lyman, and it is only gradually increased in activity to this date. However, in 2008 with have had the first conference specifically targeted at the "Text and Data Mining of Clinical Documents" with a conference organized by the Turku Centre for Computer Science, in Finland (Karsten, H., Back, B., Salakoski, T., Salanterä, S., & Suominen, H., 2008). Much of the work prior to the 1990s has been superceded by later shifts in processing power and new ideas on the software development for this task. The review of the literature in this paper is restricted to later topics that are particularly relevant to automated processing of the language of clinical notes.

In 2001, Taira and Soderland published their approach to information extraction from clinical notes in the radiology domain. They proposed a general structure for such applications which was later used by many other research groups (Huang, Y., Lowe, H., Klein, D., & Cucina, R., 2005; Arnold, C. W., Bui, A. A. T., Morioka, C., El-Saden, S., & Kangarloo, H., 2007; Thomas, B. J., Ouellette, H., Halpern, E. F., & Rosenthal, D. I., 2005; Sinha, U., & Kangarloo, H., 2002). Their proposal had five steps of processing for a complete data retrieval system: *Structural analyzer,* Lexical analyzer, Parser, Semantic interpreter and Frame constructor. They performed an evaluation on this structuring and reported obstacles in: deep understanding of the domain, ability to deal with ungrammatical writing styles, shorthand and telegraphic writings, finding solutions for assumed knowledge between the writer and reader, and handling a large vocabulary. Following their work, different studies have expanded their proposed system to other clinical domains or addressed the reported issues (Sun, J. Y., & Sun, Y., 2006) and in some cases reported new obstacles.

Related Content

Overview of Biomedical Engineering
Ahmad Taher Azar (2012). *Telemedicine and E-Health Services, Policies, and Applications: Advancements and Developments (pp. 112-139).*
www.irma-international.org/chapter/overview-biomedical-engineering/64986

RETRA: Web Based Resource Allocation Tool for Emergency Management
Venkata S. Inampudi, Russell Kondavetiand Aura Ganz (2013). *International Journal of E-Health and Medical Communications (pp. 80-93).*
www.irma-international.org/article/retra/94635

Assistive Technology for Deaf and Hard of Hearing Students
Michael Fitzpatrickand Raschelle Theoharis (2010). *Handbook of Research on Human Cognition and Assistive Technology: Design, Accessibility and Transdisciplinary Perspectives (pp. 179-191).*
www.irma-international.org/chapter/assistive-technology-deaf-hard-hearing/42836

Challenges with Adoption of Electronic Medical Record Systems
Abirami Radhakrishnan, Dessa Davidand Jigish Zaveri (2008). *Encyclopedia of Healthcare Information Systems (pp. 194-200).*
www.irma-international.org/chapter/challenges-adoption-electronic-medical-record/12941

M-Health Telemedicine and Telepresence in Oral and Maxillofacial Surgery: An Innovative Prehospital Healthcare Concept in Structurally Weak Areas
Katharina Witzkeand Olaf Specht (2017). *International Journal of Reliable and Quality E-Healthcare (pp. 37-48).*
www.irma-international.org/article/m-health-telemedicine-and-telepresence-in-oral-and-maxillofacial-surgery/187035