Cross Disciplinary Consultancy: Negation Detection Use Case

Mike Conway[1], Howard Burkom[2], Amy Ising[3]

[1]University of Utah, Salt Lake City, Utah, United States, [2]Johns Hopkins Applied Physics Laboratory, Baltimore, Maryland, United States, [3]UNC Chapel Hill, Chapel Hill, North Carolina, United States

Objective

This abstract describes an ISDS initiative to bring together public health practitioners and analytics solution developers from both academia and industry to define a roadmap for the development of algorithms, tools, and datasets to improve the capabilities of current text processing algorithms to identify negated terms (i.e. negation detection).

Introduction

Despite considerable effort since the turn of the century to develop Natural Language Processing (NLP) methods and tools for detecting negated terms in chief complaints, few standardised methods have emerged. Those methods that have emerged (e.g. the NegEx algorithm [1]) are confined to local implementations with customised solutions. Important reasons for this lack of progress include (a) limited shareable datasets for developing and testing methods (b) jurisdictional data silos, and (c) the gap between resource-constrained public health practitioners and technical solution developers, typically university researchers and industry developers.

To address these three problems ISDS, funded by a grant from the Defense Threat Reduction Agency, organized a consultancy meeting at the University of Utah designed to bring together (a) representatives from public health departments, (b) university researchers focused on the development of computational methods for public health surveillance, (c) members of public health oriented non-governmental organisations, and (d) industry representatives, with the goal of developing a roadmap for the development of validated, standardised and portable resources (methods and data sets) for negation detection in clinical text used for public health surveillance.

Methods

Free-text chief complaints remain a vital resource for syndromic surveillance. However, the widespread adoption of Electronic Health Records (and federal Meaningful Use requirements) has brought changes to the syndromic surveillance practice ecosystem. These changes have included the widespread use of EHR-generated chief complaint "pick lists" (i.e. pre-defined chief complaints that are selected by the user, rather than text strings input by the user at a keyboard), triage note templated text, and triage note free-text (typically much more comprehensive than traditional chief complaints). A key requirement for a negation detection algorithm is the ability to successfully and accurately process these new and challenging data streams.

Preparations for the consultancy included an email thread and a shared website for published articles and data samples leading to a structured pre-consultancy call designed to inform participants regarding the purpose of the consultancy and to align expectations. Then, health department users were requested to provide data samples exemplifying negation issues in the classification process. Presenting developers were asked to explain their underlying ideas, details of method implementation, size and composition of corpora used for evaluation, and classification performance results.

Results

The consultancy was held on January 19th & 20th 2017 at the University of Utah's Department of Biomedical Informatics, and consisted of 25 participants. Participants were drawn from various different sectors, with representation from ISDS (2), the Defense Threat Reduction Agency (1), universities and research institutes (10), public health departments (5), the Department of Veterans Affairs (4), non-profit organisations (2), and technology firms (1). Participants were drawn from a variety of different professional backgrounds, including research scientists, software developers, public health executives, epidemiologists, and analysts.

Day 1 of the consultancy was devoted to providing an overview of NLP and current trends in negation detection, including a detailed description of widely used algorithms and tools for the negation detection task. Key questions included: Should our focus

be chief complaints only, or should we widen our scope to emergency department triage notes?, How many other NLP tasks (e.g. reliable concept recognition) is it necessary to address on the road to improved negation detection? With this background established, Day 2 centered on presentations from five different United States local and regional health departments (King County WA, Boston MA, North Carolina, Georgia, and Tennessee) on the various approaches to text processing and negation detection utilized across several jurisdictions.

Several key areas of focus emerged as a result of the consultancy discussion. First, there is a clear need for a large, easily accessible corpus of free-text chief complaints that can form a standardised testbed for negation detection algorithm development and evaluation. Annotated data, in this context, consists of chief complaints annotated for concepts (e.g. vomiting, pain in chest) and the negation status of those concepts. It is important that the annotation include both annotated clinical concepts and negation status to allow for the uniform evaluation and performance comparison of candidate negation detection algorithms. Further, the annotated corpus should consist of several thousand (as opposed to several hundred) distinct and representative chief complaints in order to compare algorithms against a sufficient variety and volume of negation patterns.

Conclusions

The consultancy was stimulating and eye-opening for both public health practitioner and technology developer attendees. Developers unfamiliar with the everyday health-monitoring context gained an appreciation of the difficulty of deriving useful indicators from chief complaints. Also highlighted was the challenge of processing triage notes and other free-text fields that are often unused for surveillance purposes. Practitioners were provided with concise explanations and evaluations of recent NLP approaches applicable to negation processing. The event afforded direct dialogue important for communication across professional cultures.

Please note that a journal paper describing the consultancy has recently been published in the *Online Journal of Public Health Informatics* [2].

Acknowledgement

References

1. Chapman WW, Bridewell W, Henbury P, Cooper G, Buchanan B. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*. 34(5), 301-10. PubMed https://doi.org/10.1006/jbin.2001.1029

2. Conway M, Mowery D, Ising A, Velupillai S, Doan S, et al. 2018. Cross-disciplinary consultance to breige public health technical needs and analytic developers: negation detection use case. *Online J Public Health Inform*. 10(2), e209. PubMed https://doi.org/10.5210/ojphi.v10i2.8944