

Classifying Supporting, Refuting, or Uncertain Evidence for Pneumonia Case Review

Brett R. South^{*1,2}, Heidi S. Kramer^{1,2}, Barbara Jones^{1,2}, Melissa Tharp¹ and Wendy Chapman^{1,2}

¹Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; ²VA Salt Lake City Health Care System, Salt Lake City, UT, USA

Objective

We sought to identify relevant evidence that supports, refutes or contributes uncertainty when reviewing cases of suspected pneumonia and characterize their interaction with uncertainty phenomena found in clinical texts.

Introduction

Characterizing mentions found in clinical texts that support, refute, or represent uncertainty for suspected pneumonia is one area where automated Natural Language Processing (NLP) screening algorithms could be improved. Mentions of uncertainty and negation commonly occur in clinical texts, and opportunities exist to extend existing algorithms [1] and taxonomies [2]. In general there are three main sources of uncertainty found in healthcare: 1) probability or risk; 2) ambiguity – lack of reliability, credibility or adequacy of the information; and, 3) complexity – aspects of the phenomenon that make it difficult to comprehend [3].

Methods

We conducted an automated screening of all outpatient encounters occurring at the VA Salt Lake City Health Care system before 01/01/2012 to identify a cohort of suspected cases of pneumonia. Screening criteria included: a) presence of ICD-9 code for pneumonia and; b) presence of an electronic physician note and/or same day chest imaging report. From this larger cohort, we selected a random sample of 25 cases containing 58 documents.

All cases were reviewed by a pulmonologist, an internist and five allied health professionals. Using criteria based on the CDC pneumonia case definition, and the available clinical documentation, each case was classified as “suspected”, “unlikely”, or “cannot be determined”. Reviewers classified evidence into three semantic classes: a) words or phrases that *support*; b) *refute*; or c) are *uncertain for pneumonia diagnosis*. To accomplish this task we used an open source annotation tool called eHOST [4] and an annotation approach that focused on identifying and characterizing relevant spans of clinical text that support, refute or represent uncertainty for pneumonia. We report entire ranges of pair-wise inter-annotator agreement and the prevalence of annotations in each semantic class. For those annotations marked as uncertain we categorize the information according to the three general sources of uncertainty.

Results

Seven annotators generated a total of 2,042 annotations for *supports* (1,302, 63%), *refutes* (470, 23%), and *uncertain* (268, 13%). Average agreement for case level classification was 0.60. Range for pair-wise inter-annotator agreement across all semantic classes was (0.34-0.61) and individually for *supports* (0.25-0.67), *refutes* (0.37-0.47), *uncertain* (0.36-0.45). Errors where one or more reviewer identified a span of text and others did not were more common than classification errors. The majority (70%) of annotations reviewers marked as uncertain were found in chest imaging reports. For annotated mentions marked as uncertain, (159 59%) represented information where linguistic cues implied ambiguity, (29 11%),

where data was unavailable, and only (10 4%) where the data quality was questionable. Opportunities exist to incorporate more formal linguistic analyses and extend uncertainty taxonomies.

Conclusions

We found substantial annotator variability in identifying supporting, refuting, or uncertain evidence for the diagnosis of pneumonia in clinical text. Future work will expand these methods to a larger case sample and incorporate a more formal linguistic analysis to identify specific lexical cues thereby extending existing taxonomies of uncertainty and improving automated NLP algorithms

Keywords

Natural Language Processing; Chart Review; Annotation

Acknowledgments

This study was carried out using resources and support from the VA Informatics and Computing Infrastructure (VINCI) Project ID: HIR 08-204.

References

1. Chapman, W.W., Chu, D., Dowling, J.N. ConText: An algorithm for identifying contextual features from clinical text. In: ACL-07 2007.
2. Mowery, D., Velupillai, S., Chapman, W.W. Medical diagnosis lost in translation – Analysis of uncertainty and negation expressions in English and Swedish clinical texts. Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. Association for Computational Linguistics. 2012.
3. Han, Paul K.J., Klein, William M.P., Arora, N.K. Varieties of Uncertainty in Health Care A Conceptual Taxonomy. Medical Decision Making 31.6 (2011): 828-838.
4. South, B., Shen, S., Leng, J., Forbush, T., DuVall, S., Chapman, W.W. A prototype tool set to support machine-assisted annotation. Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. BioNLP '12, Stroudsburg, PA, USA, Association for Computational Linguistics. 2012. 130-139.

*Brett R. South

E-mail: brett.south@hsc.utah.edu

