

Clinical Natural Language Processing in 2014: Foundational Methods Supporting Efficient Healthcare

A. Névéol, P. Zweigenbaum, Section Editors for the IMIA Yearbook Section on Clinical Natural Language Processing
LIMSI CNRS UPR 3251, Orsay, France

Summary

Objective: To summarize recent research and present a selection of the best papers published in 2014 in the field of clinical Natural Language Processing (NLP).

Method: A systematic review of the literature was performed by the two section editors of the IMIA Yearbook NLP section by searching bibliographic databases with a focus on NLP efforts applied to clinical texts or aimed at a clinical outcome. A shortlist of candidate best papers was first selected by the section editors before being peer-reviewed by independent external reviewers.

Results: The clinical NLP best paper selection shows that the field is tackling text analysis methods of increasing depth. The full review process highlighted five papers addressing foundational methods in clinical NLP using clinically relevant texts from online forums or encyclopedias, clinical texts from Electronic Health Records, and included studies specifically aiming at a practical clinical outcome. The increased access to clinical data that was made possible with the recent progress of de-identification paved the way for the scientific community to address complex NLP problems such as word sense disambiguation, negation, temporal analysis and specific information nugget extraction. These advances in turn allowed for efficient application of NLP to clinical problems such as cancer patient triage. Another line of research investigates online clinically relevant texts and brings interesting insight on communication strategies to convey health-related information.

Conclusions: The field of clinical NLP is thriving through the contributions of both NLP researchers and healthcare professionals interested in applying NLP techniques for concrete healthcare purposes. Clinical NLP is becoming mature for practical applications with a significant clinical impact.

Keywords

Awards and Prizes; Decision Making, Computer-Assisted; Medical Informatics/trends; Natural Language Processing; Semantics

Yearb Med Inform 2015;10:194-8
<http://dx.doi.org/10.15265/Y-2015-035>
Published online August 13, 2015

Introduction

In 2015, the IMIA Yearbook starts a new section entitled clinical Natural Language Processing (NLP). For the purpose of this section, we define clinical NLP as Natural Language Processing applied to clinical texts or aimed at a clinical outcome. This clearly encompasses NLP applied to texts in Electronic Health Records (EHRs), which is the case of the bulk of information extraction for decision support or clinical research. We also considered as clinically relevant applications the automatic analysis of patient-authored texts or speeches for the purpose of diagnosis, or the analysis of patient-oriented communication, which leads to delivering better health information to the public.

Selection Process

Our pre-selection of papers followed a three-step process. First, queries were sent to two bibliographic databases relevant for the topic of clinical Natural Language Processing: MEDLINE and ACL Anthology. We used the PubMed interface (<http://www.ncbi.nlm.nih.gov/pubmed/>) and the ACL Anthology Searchbench (<http://aclasb.dfki.de/>). Additionally, we examined the full list of papers published in the major Medical Informatics journals. Second, we examined the retrieved titles and abstracts to select articles which met the following selection criteria: (1) NLP was performed on clinically relevant texts, (2) the work aimed at a clinical outcome, and (3) the contribution was significant from both the NLP and clinical standpoints. Third, selected articles

were grouped by broad topics and ranked. We further refined this selection in order to cover each of the topics while ensuring that the final selection comprised a variety of topics, authors and venues. We then browsed the full text of the final 15 papers across all categories. In the list of references provided at the end of the synopsis, a star indicates papers that were in the selection of candidate best papers.

The MEDLINE queries (see Figure 1) used the metadata recommendations from Lamy et al. [1] combined with (1) the MeSH keyword Natural Language Processing, and (2) free text keywords that were derived from introspection of the section editors and the authors of the survey paper published in the NLP section of the 2015 edition of the Yearbook. The ACL Anthology query restricted our selection to the most selective conferences (ACL, EMNLP, NAACL, EACL, COLING, CONLL) and workshops (ACL BioNLP). It used the free text keywords medical, clinical and health.

The systematic collection of papers through queries brought back 122 titles and abstracts. Additionally, the manual perusal of the table of contents of three major Medical Informatics journals in which most NLP research is published: JAMIA, JBI and BMC Bioinformatics (section on Knowledge-based analysis) identified 32 more papers, resulting in a total selection of 154 papers. Review papers and correspondence were then discarded in order to keep only original research contributions. A significant number of papers were also discarded because they focused more on biological vs clinical aspects, or on knowledge representation vs NLP.

1. MEDLINE query with MeSH keyword

```
(( English [LA] AND journal article [PT] AND 2014 [dp] AND hasabstract[text] NOT
pubstatusaheadofprint )
AND Natural Language Processing[mh] )
```

2. MEDLINE query with free-text keywords

```
(( English [LA] AND journal article [PT] AND 2014 [dp] AND hasabstract[text] ) NOT
pubstatusaheadofprint )
AND
((medical OR clinical OR natural) AND "language processing")
AND
(
("Patient-Centered" OR care OR coordination)
OR
(
("clinical text" OR ((medical or health or clinical) AND (narratives or notes)))
AND
(((semantic or syntactic) AND annotation AND (model or scheme or guidelines))
OR
((annotation or annotator) AND agreement)
OR
(corpus or dataset)
OR
(tools or tool or system or software or application)
))
)
```

Fig. 1 MEDLINE search queries used to retrieve papers relevant to the clinical NLP section

Topics in Clinical NLP

Based on the review of the papers published in 2014, we noticed two main lines of research: foundational methods in clinical NLP, and clinical applications of NLP.

Foundational Methods in Clinical NLP

A number of papers aim at designing or optimizing NLP methods and resources, which contributes to progress in clinical text processing. Annotated text corpora are a key resource to develop NLP systems. However, creating an annotated clinical corpus that can be used for research, whether inside or outside a medical center, raises issues of privacy. This is why automatic de-identification [3-7] remains a very active area of research. While de-identification aims at masking identifying information, while keeping a high sensitivity in its detection, the risk that it also removes information useful for subsequent research must

be assessed [5]. Many de-identification efforts target entities that can link a clinical document to an individual patient. However, protected health information (PHI) covers actual clinical information, which can by itself link a record to a patient. Automatically identifying such information was addressed by Sánchez et al. [7]. Progress continues in parallel in the design of text annotation methods, which results in more reliable datasets [3]. Automatically pre-annotating texts before they are provided to human annotators contributes to faster and more consistent annotation [4, 6, 8, 9].

After these steps, core NLP methods can be investigated on the de-identified and/or annotated texts. We organize the rest of this section according to the usual order of components in a typical text-processing pipeline.

- At the lexical level (i.e. word-level semantics), Henriksson et al. [10] investigate distributional semantics methods for synonym extraction and abbreviation expansion. Chasin et al. [11] address the disambiguation of clinical terms found in EHRs ranging from highly frequent to mildly frequent

by assigning a specific UMLS concept denoting the sense associated to terms in context. They assess the contribution of unsupervised methods, compared to rich domain knowledge and find topic models to perform well. Stenetorp et al. [12] address Semantic Category Disambiguation, which is positioned at a coarser granularity and seeks to assign semantic types to terms in context. They rely primarily on lexical domain knowledge and allow multiple category attributions for a given term.

- At the syntactic level, Laippala et al. [13] study syntactic parsing using statistical methods for domain adaptation on clinical Finnish.
- At the semantic level (i.e. sentence level semantics), a rich body of work addresses the detection of specific types of information, often expressed in named entities: body site, severity modifiers [14], medications [15-16]; problems, procedures, lab tests, and medications [17], tumor-related information, e.g., size and body site [18]. One step further in the analysis is the task of further characterizing entities of interest. For instance, Sohn et al. address entity normalization for medications by automatically mapping entities into the concept unique identifiers of RxNorm [15]. Other work also addresses the tasks of negation [19] and context [20-21] detection, which are found to be difficult to generalize across languages or even datasets in the same language. Within-sentence analysis can then be used to perform sentence classification in online health communities to detect the presence of adverse drug reactions [32] or to categorize a sentence as conveying emotional or informational support to other users [23].
- Finally, at the discourse level, an emerging line of research addresses the extraction of complete clinical timelines across documents [22, 33].

Applications of NLP with a Clinical Purpose

A number of papers primarily aim at a clinical outcome, while using NLP methods with significant depth. The clinical impact of these contributions is sometimes notable by

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2015 in the section 'Clinical Natural Language Processing'. The articles are listed in alphabetical order of the first author's surname.

Section
Clinical Natural Language Processing
<ul style="list-style-type: none"> ▪ Biyani P, Caragea C, Mitra P, Yen J. Identifying emotional and informational support in online health communities. <i>Proc COLING</i> 2014 Sept:827-36. ▪ Carrell DS, Halgrim S, Tran DT, Buist DS, Chubak J, Chapman WW, Savova G. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. <i>Am J Epidemiol</i> 2014 Mar 15;179(6):749-58. ▪ Chasin R, Rumshisky A, Uzuner Ö, Szolovits P. Word sense disambiguation in the clinical domain: a comparison of knowledge-rich and knowledge-poor unsupervised methods. <i>J Am Med Inform Assoc</i> 2014 Sep-Oct;21(5):842-9. ▪ Raghavan P, Fosler-Lussier E, Elhadad N, Lai A. Cross-narrative temporal ordering of medical events. <i>Proc ACL</i> 2014:998-1008. ▪ Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C. Negation's not solved: generalizability versus optimizability in clinical natural language processing. <i>PLoS One</i> 2014 Nov 13;9(11):e112774.

the publication of the research in journals outside the NLP, and more generally the medical informatics community, such as *Am J Epidemiol*, *J Card Fail* or *World J Urol*.

The clinical applications of NLP cover a great variety of medical problems, such as the automatic extraction of heart failure diagnostic criteria from EHRs [24-25], the automatic classification of patient records for breast cancer recurrence [26], binge eating [27], prostatic adenocarcinoma [28], patient acuity [29] and the analysis of patients' transcribed speech to detect agrammatic aphasia [30].

Some work relies on multi-modal data including speech and eye movements as doctors make their own diagnosis and state their confidence level in the diagnosis. Information on the confidence in one's diagnosis is hypothesized to reduce diagnosis errors and to improve communication with patients [31]. Note that both [30] and [31] use transcribed speech as the source for NLP analysis. In both studies, speech transcription is performed manually, which is a limitation that points out the strong need to make the process fully automated.

Some work exploits online texts such as those found in social media for the detection of adverse drug reactions [32]. EHRs are also found useful for the detection of drug-drug interactions [33]. Both sources are known to yield information complementary to that reported through FDA channels.

Finally, an emerging line of research provides insight on the current state of clinical practice. The evaluation of EHR content

quality is an issue of specific interest to the community. It is addressed by Spickard et al. [34] who performed a simple study comparing the content of clinical notes written by interns to the coding of the corresponding patient records. Another study proposed an automatic scoring instrument validated by interns and GPs on a set of type 2 diabetes patient records [35]. The work of Biyani et al. [23] addresses the automatic characterization of contributions in online communities, including those of opinion leaders in order to yield insight on how best to communicate with patients. This could have implications for physicians' understanding on how to behave to convey influential advice to patients. In the context of e-patient Dave's experience [2], this work also illustrates how patients can benefit from medical help from sources outside health professionals.

Concluding Remarks

Interestingly, while most research uses clinical notes found in EHRs, other research uses other texts with a clinical purpose, such as patient-authored texts [7, 9, 23, 30, 32]. In addition, some research compares or combines clinical texts with other types of texts [10, 12, 32]. Also of note, some research addresses a wide range of languages other than English, including Chinese [17, 18], Dutch [20], Finnish [13, 29], French [4], and Swedish [10, 16, 21].

The field of clinical NLP is thriving through the contributions of both NLP researchers and healthcare professionals interested in applying NLP techniques for concrete healthcare purposes. Clinical NLP research is very diverse in terms of the types of texts and languages studied. The field is becoming mature for practical applications with a significant clinical impact.

References

In the reference list below, we order publications according to the two main topics identified in this review. Papers that were shortlisted as best paper candidates are marked with a *. These fifteen papers were submitted to evaluation by Yearbook reviewers, and the five papers in Table 1 were finally selected as 2014 best papers.

1. Lamy JB, Séroussi B, Griffon N, Kerdelhué G, Jaulent MC, Bouaud J. Toward a Formalization of the Process to Select IMIA Yearbook Best Papers. *Methods Inf Med* 2014 Nov 14;54(1).
2. deBronkart D. How the e-patient community helped save my life: an essay by Dave deBronkart. *BMJ* 2013 Apr 2;346:f1990.

A. Methods in NLP

3. Deléger L, Lingren T, Ni Y, Kaiser M, Stoutenborough L, Marsolo K, et al. Preparing an annotated gold standard corpus to share with extramural investigators for de-identification research. *J Biomed Inform* 2014 Aug;50:173-83.
4. Grouin C, Névéol A. De-identification of clinical notes in French: towards a protocol for reference corpus development. *J Biomed Inform*. 2014 Aug;50:151-61.
5. Meystre SM, Ferrández Ó, Friedlin FJ, South BR, Shen S, Samore MH. Text de-identification for privacy protection: a study of its impact on clinical text information content. *J Biomed Inform* 2014 Aug;50:142-50.
6. Li M, Carrell D, Aberdeen J, Hirschman L, Malin BA. De-identification of clinical narratives through writing complexity measures. *Int J Med Inform* 2014 Oct;83(10):750-67.
7. *Sánchez D, Batet M, Viejo A. Utility-preserving privacy protection of textual healthcare documents. *J Biomed Inform* 2014 Dec;52:189-98.
8. *Gobbel GT, Garvin J, Reeves R, Cronin RM, Heavirland J, Williams J, et al. Assisted annotation of medical free text using RapTAT. *J Am Med Inform Assoc* 2014 Sep-Oct;21(5):833-41.
9. Lingren T, Deléger L, Molnar K, Zhai H, Meinen-Derr J, Kaiser M, et al. Evaluating the impact of pre-annotation on annotation speed and potential bias: natural language processing gold standard development for clinical named entity

- recognition in clinical trial announcements. *J Am Med Inform Assoc* 2014 May-Jun;21(3):406-13.
10. Henriksson A, Moen H, Skeppstedt M, Daudaravicius V, Duneld M. Synonym extraction and abbreviation expansion with ensembles of semantic spaces. *J Biomed Semantics* 2014 Feb 5;5(1):6.
 11. *Chasin R, Rumshisky A, Uzuner Ö, Szolovits P. Word sense disambiguation in the clinical domain: a comparison of knowledge-rich and knowledge-poor unsupervised methods. *J Am Med Inform Assoc* 2014 Sep-Oct;21(5):842-9.
 12. Stenetorp P, Pyysalo S, Ananiadou S, Tsujii J. Generalising semantic category disambiguation with large lexical resources for fun and profit. *J Biomed Semantics* 2014 Jun 2;5:26.
 13. *Laippala V, Viljanen T, Airola A, Kanerva J, Salanterä S, Salakoski T, et al. Statistical parsing of varieties of clinical Finnish. *Artif Intell Med* 2014 Jul;61(3):131-6.
 14. *Dligach D, Bethard S, Becker L, Miller T, Savova GK. Discovering body site and severity modifiers in clinical texts. *J Am Med Inform Assoc* 2014 May-Jun;21(3):448-54.
 15. *Sohn S, Clark C, Halgrim SR, Murphy SP, Chute CG, Liu H. MedXN: an open source medication extraction and normalization tool for clinical text. *J Am Med Inform Assoc* 2014 Sep-Oct;21(5):858-65.
 16. Skeppstedt M, Kvist M, Nilsson GH, Dalianis H. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. *J Biomed Inform* 2014 Jun;49:148-58.
 17. *Lei J, Tang B, Lu X, Gao K, Jiang M, Xu H. A comprehensive study of named entity recognition in Chinese clinical text. *J Am Med Inform Assoc* 2014 Sep-Oct;21(5):808-14.
 18. Wang H, Zhang W, Zeng Q, Li Z, Feng K, Liu L. Extracting important information from Chinese Operation Notes with natural language processing methods. *J Biomed Inform* 2014 Apr;48:130-6.
 19. *Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, et al. Negation's not solved: generalizability versus optimizability in clinical natural language processing. *PLoS One* 2014 Nov 13;9(11):e112774.
 20. *Afzal Z, Pons E, Kang N, Sturkenboom M, Schuemie MJ, Kors JA. ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus. *BMC Bioinformatics* 2014 Nov 29;15(1):373.
 21. Velupillai S, Skeppstedt M, Kvist M, Mowery D, Chapman BE, Dalianis H, et al. Cue-based assertion classification for Swedish clinical text—developing a lexicon for pyConTextSwe. *Artif Intell Med* 2014 Jul;61(3):137-44.
 22. *Raghavan P, Fosler-Lussier E, Elhadad N, Lai A. Cross-narrative Temporal Ordering of Medical Events. *Proc ACL* 2014:998-1008.
 23. *Biyani P, Caragea C, Mitra P, Yen J. Identifying Emotional and Informational Support in Online Health Communities. *Proc COLING* 2014. 2014:827-836.

B. Applications

24. Vijayakrishnan R, Steinhubl SR, Ng K, Sun J, Byrd RJ, Daar Z, et al. Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. *J Card Fail* 2014 Jul;20(7):459-64.
25. *Byrd RJ, Steinhubl SR, Sun J, Ebadollahi S, Stewart WF. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. *Int J Med Inform* 2014 Dec;83(12):983-92.
26. *Carrell DS, Halgrim S, Tran DT, Buist DS, Chubak J, Chapman WW, et al. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. *Am J Epidemiol* 2014 Mar 15;179(6):749-58.
27. Bellows BK, LaFleur J, Kamaau AW, Ginter T, Forbush TB, Agbor S, et al. Automated identification of patients with a diagnosis of binge eating disorder from narrative electronic health records. *J Am Med Inform Assoc* 2014 Feb;21(e1):e163-8.
28. Thomas AA, Zheng C, Jung H, Chang A, Kim B, Gelfond J, et al. Extracting data from electronic medical records: validation of a natural language processing program to assess prostate biopsy results. *World J Urol* 2014 Feb;32(1):99-103.
29. Kontio E, Airola A, Pahikkala T, Lundgren-Laine H, Junttila K, Korvenranta H, et al. Predicting patient acuity from electronic patient records. *J Biomed Inform* 2014 Oct;51:35-40.
30. *Fraser KC, Hirst G, Meltzer JA, Mack JE, Thompson CK. Using statistical parsing to detect agrammatic aphasia. *Proceedings of BioNLP* 2014, ACL, 2014 June;134-42.
31. Bullard J, Ovesdotter Alm C, Yu Q, Shi P, Haake A. Towards multimodal modeling of physicians' diagnostic confidence and self-awareness using medical narratives. *Proc COLING* 2014:1718-1727
32. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform* 2015 Feb;53:196-207.
33. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc* 2014 Mar-Apr;21(2):353-62.
34. *Spickard A 3rd, Ridinger H, Wrenn J, O'Brien N, Shpigel A, Wolf M, et al. Automatic scoring of medical students' clinical notes to monitor learning in the workplace. *Med Teach* 2014 Jan;36(1):68-72.
35. Burke HB, Hoang A, Becher D, Fontelo P, Liu F, Stephens M, et al. QNOTE: an instrument for measuring the quality of EHR clinical notes. *J Am Med Inform Assoc* 2014 Sep-Oct;21(5):910-6.

Correspondence to:

Aurélien Névéol, Pierre Zweigenbaum
LIMSI CNRS UPR 3251
Rue John von Neumann
Campus Universitaire d'Orsay
91405 Orsay cedex, France
E-mail: {neveol,pz}@limsi.fr

Appendix: Content Summaries of Selected Best Papers for the 2015 IMIA Yearbook, Section Clinical Natural Language Processing

Biyani P, Caragea C, Mitra P, Yen J
Identifying emotional and informational support in online health communities
Proc COLING 2014 Sept:827-36

This paper presents an analysis of the interactions in an online cancer support community. It draws from sentiment analysis methods to identify emotional messages (which express concern, understanding or sympathy and provide encouragement, affirmation and validation) and informational messages (which focus on providing advice, knowledge and referrals). The authors perform automatic sentence classification using a variety of syntactic, lexical and other linguistic features. The authors provide an annotated corpus of 1,066 forum sentences. Even though the clinical applications of this work can seem remote, it can be seen as putting forth methods that can help healthcare providers to yield insight on how best to communicate with patients by characterizing the contributions of opinion leaders in online communities. In this work, emotional support was shown to prevail over informational support in the contributions of opinion leaders. This could have implications for physicians' understanding on how to behave to convey influential advice to patients. This work also illustrates how patients can benefit from medical help from sources outside health professionals.

Carrell DS, Halgrim S, Tran DT, Buist DS, Chubak J, Chapman WW, Savova G
Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence
Am J Epidemiol 2014 Mar 15;179(6):749-58

This paper describes the state of the art of clinical NLP in a very clear and educational

way. It presents a study aiming to address a major bottleneck of clinical research: the costs and time associated with the manual analysis of clinical data. The authors report on the successful use of state-of-the-art natural language processing techniques to correctly identify cases of cancer recurrence among women with previous breast cancer. Specifically, they perform automatic classification of breast cancer recurrence based on EHR content. Open source cTAKES is used to obtain document sections, as well as concepts and assertion features for the classifier. The overall classification module was evaluated on a large dataset of 1,472 patients. The practical impact of this work is a 10-fold decrease in the need for chart abstraction, though with an 8% loss in case detection. This study is remarkable for illustrating the potential of advanced NLP in clinical practice.

Chasin R, Rumshisky A, Uzuner Ö, Szolovits P
Word sense disambiguation in the clinical domain: a comparison of knowledge-rich and knowledge-poor unsupervised methods

J Am Med Inform Assoc 2014 Sep-Oct;21(5):842-9

This paper addresses an important topic in clinical NLP, namely Word Sense Disambiguation (WSD). It focuses on unsupervised WSD methods, using Bayesian topic modeling and evaluating with a freely available clinical dataset where 50 ambiguous target clinical terms are tagged with senses encoded by Unified Medical Language System (UMLS) concepts. A much larger unlabeled version of this dataset is also used to create the topic

models. The authors evaluate a variety of approaches and find that Latent Dirichlet Analysis (LDA) with a simple bag-of-words representation performs better than knowledge-rich methods, including graph-based semantic similarity, which use UMLS concepts directly. They also observe that the addition of bags of “clinical phrases” (phrases matching UMLS terms after normalization) and syntactic dependency information brings further improvements to the LDA methods.

Raghavan P, Fosler-Lussier E, Elhadad N, Lai A

Cross-narrative temporal ordering of medical events

Proc ACL 2014:998-1008

This paper shows how a timeline of medical events documented across multiple clinical narratives in a single patient record can be inferred. The paper builds on previous work identifying medical events and temporal expressions within a single document. It describes a general framework for aligning sequences of events drawn from multiple documents. To obtain the timelines, a weighted finite state representation is used that allows the consideration of temporal and co-reference relations between medical events described in the clinical narratives. The paper compares the weighted finite state method to pairwise alignment using dynamic programming, and finds the finite state method to be superior. Even though the test corpus used for the evaluation is somewhat small (7 patients, 80 documents), this work is remarkable for addressing the complex

NLP task of temporal analysis in multiple documents. Furthermore, it also evaluates the contribution of co-reference and intra-narrative relations in a practical setting.

Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C

Negation's not solved: generalizability versus optimizability in clinical natural language processing

PLoS One 2014 Nov 13;9(11):e112774

This paper asserts that a review of clinical NLP work addressing negation detection shows excellent results for this task with F-measure scores often in the high 90s. Although this could lead to the conclusion that negation is a solved problem, the authors point out that a more accurate conclusion is that negation detection can be optimized for a given dataset but not necessarily solved across the board. The paper supports this idea by a systematic comparison of different techniques for negation detection in clinical documents using four publicly available datasets. The authors note that the four datasets provide a different picture of negation through significant variation in the annotation guidelines for the types of entities affected by negation as well as how the boundaries of these entities are defined. The results are interesting and the conflict between optimizability and generalizability likely applies to other tasks in clinical NLP. Even though the authors do not provide either a general linguistically motivated definition of negation, or a solution to make progress towards generalizability, this is a nice study articulating the difficulty of negation detection.