

# Challenges in clinical natural language processing for automated disorder normalization



Robert Leaman, Ritu Khare, Zhiyong Lu \*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, United States

## ARTICLE INFO

### Article history:

Received 11 August 2014

Revised 18 June 2015

Accepted 8 July 2015

Available online 14 July 2015

### Keywords:

Natural language processing

Electronic health records

Information extraction

## ABSTRACT

**Background:** Identifying key variables such as disorders within the clinical narratives in electronic health records has wide-ranging applications within clinical practice and biomedical research. Previous research has demonstrated reduced performance of disorder named entity recognition (NER) and normalization (or grounding) in clinical narratives than in biomedical publications. In this work, we aim to identify the cause for this performance difference and introduce general solutions.

**Methods:** We use closure properties to compare the richness of the vocabulary in clinical narrative text to biomedical publications. We approach both disorder NER and normalization using machine learning methodologies. Our NER methodology is based on linear-chain conditional random fields with a rich feature approach, and we introduce several improvements to enhance the lexical knowledge of the NER system. Our normalization method – never previously applied to clinical data – uses pairwise learning to rank to automatically learn term variation directly from the training data.

**Results:** We find that while the size of the overall vocabulary is similar between clinical narrative and biomedical publications, clinical narrative uses a richer terminology to describe disorders than publications. We apply our system, DNORM-C, to locate disorder mentions and in the clinical narratives from the recent ShARE/CLEF eHealth Task. For NER (strict span-only), our system achieves precision = 0.797, recall = 0.713, f-score = 0.753. For the normalization task (strict span + concept) it achieves precision = 0.712, recall = 0.637, f-score = 0.672. The improvements described in this article increase the NER f-score by 0.039 and the normalization f-score by 0.036. We also describe a high recall version of the NER, which increases the normalization recall to as high as 0.744, albeit with reduced precision.

**Discussion:** We perform an error analysis, demonstrating that NER errors outnumber normalization errors by more than 4-to-1. Abbreviations and acronyms are found to be frequent causes of error, in addition to the mentions the annotators were not able to identify within the scope of the controlled vocabulary.

**Conclusion:** Disorder mentions in text from clinical narratives use a rich vocabulary that results in high term variation, which we believe to be one of the primary causes of reduced performance in clinical narrative. We show that pairwise learning to rank offers high performance in this context, and introduce several lexical enhancements – generalizable to other clinical NER tasks – that improve the ability of the NER system to handle this variation. DNORM-C is a high performing, open source system for disorders in clinical text, and a promising step toward NER and normalization methods that are trainable to a wide variety of domains and entities. (DNORM-C is open source software, and is available with a trained model at the DNORM demonstration website: <http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/tmTools/#DNORM>.)

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The application of clinical natural language processing to the clinical narratives in electronic health records has the potential to make a significant impact in many aspects of healthcare and biomedical research. Recent work has demonstrated this potential at the point of care (such as with clinical decision support [1,2]), for

\* Corresponding author at: MSC3825 NCBI/NLM/NIH, Bldg 38A, Rm 1003A, 8600 Rockville Pike, Bethesda, MD 20894, United States. Tel.: +1 301 594 7089.

E-mail addresses: [robert.leaman@nih.gov](mailto:robert.leaman@nih.gov) (R. Leaman), [ritu.khare@nih.gov](mailto:ritu.khare@nih.gov) (R. Khare), [zhiyong.lu@nih.gov](mailto:zhiyong.lu@nih.gov) (Z. Lu).

patients understanding their own records [3], for public health (e.g. biosurveillance [4]), and in biomedical research (e.g. cohort identification [5,6], identifying novel potential clinical associations [7,8] or pharmacovigilance [9,10]).

A common task in clinical natural language processing is the identification of key clinical variables mentioned in the text, such as disorders or treatments. This subtask includes both locating mentions of key entities (named entity recognition) and identifying mentions found (normalization) with respect to a controlled vocabulary such as SNOMED-CT [11]. An example sentence from a clinical narrative, together with annotations for disorder mentions and their respective concepts can be seen in Fig. 1. The results of these subtasks are then used by downstream components to provide the higher level processing required by the end task, making the result highly dependent on the quality of the normalization results obtained.

While clinical natural language processing has been an area of increasing attention, progress still trails the general and biomedical domains [12]. While this is partially due to the relative scarcity of clinical narrative corpora due to privacy concerns, it is also partially due to the additional difficulties encountered in clinical narrative text. Biomedical text is written to communicate research results to a wide audience and is edited for clarity. Clinical narrative text, on the other hand, is written by health care professionals to communicate the status and history of a single patient to other health care professionals or themselves. These notes record the reason the patient was seen (i.e. the chief complaint), document treatments given, the findings of any tests administered, and other information needed to make informed decisions regarding care for a single patient. Typical notes include progress notes documenting treatment provided, consult reports communicating the results of diagnostic tests, and discharge summaries recording the entire course of a hospital stay. As these notes each have different purposes, they are highly heterogeneous in their content and level of detail. Moreover their format is typically unstructured, resulting in a wide variety of ad-hoc structural components.

We illustrate many of the problems clinical narrative text presents for natural language processing with examples in Table 1. In general, systems which process clinical narratives cannot expect the clear structure and communication common in published text. Instead, clinical narrative is prepared under considerable time pressure, using a combination of ad-hoc formatting, eliding words which could be inferred from context, and with liberal use of parenthetical expressions, jargon and acronyms to increase the information density.

In this article, we approach the task of normalization of key entities in clinical narrative using disorders as a case study. While disease normalization in biomedical publications is a difficult task due to the wide variety of naming patterns for diseases, ambiguity, and term variation, the idiosyncrasies of clinical text cause additional difficulties. Disorder normalization in clinical text has been previously attempted using both lexical and rule-based techniques [13–17]. In this line of research, the task has often been to identify diagnoses, as in ICD-9 coding [18], or only a specific subset of problems [19]. In this study, we instead describe a system to identify and normalize all disorders mentioned in a clinical narrative. We employ DNorm, a machine learning method for learning the similarity between mentions and concept names from a specified controlled vocabulary directly from training data. Previous work with DNorm demonstrated state-of-the-art performance in biomedical publications [20]. Our participation in the recent ShARE/CLEF eHealth shared task resulted in the highest normalization performance of any participant [3], but also demonstrated performance significantly reduced compared to biomedical publications. In this manuscript we demonstrate that the vocabulary used to describe disorders in clinical text is richer than

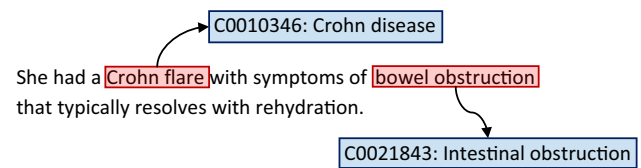


Fig. 1. Example of normalization in clinical text, showing two disorder mentions, with their respective spans and SNOMED-CT concept identifiers.

in publications – and hence more difficult to model – while the variability in overall language use is similar (see Methods Section). We use this insight to make improvements to the named entity recognition component to handle the richer disorder vocabulary.

## 2. Related work

Natural language processing of clinical text has a long history, going back to the Linguistic String Project–Medical Language Processor (LSP–MLP) project in 1986 [21]. The field was thoroughly surveyed by Meystre et al. [22] in 2008.

A recurring theme is the scarcity of annotated corpora, or datasets which can be used to develop and evaluate natural language processing systems [12]. One consequence of the relative lack of annotated data is a longstanding emphasis on knowledge intensive approaches. For example, two of the most widely used tools (MetaMap [23,24] and MedLEE [25]) both emphasize a hybrid of natural language processing and lexical approaches using the UMLS Metathesaurus, rather than machine learning methods trained directly on clinical text. MetaMap, created by the National Library of Medicine, has been successfully applied to clinical narratives for a wide variety of purposes, including biosurveillance [4], cohort identification [5], and to find potentially novel clinical associations [7]. MedLEE has also been applied to clinical text for a wide variety of purposes, including pharmacovigilance [9], SNOMED-CT coding [25], and determining comorbidities [26].

The field has moved forward in large part due to the efforts of several groups to provide annotated data through the context of a shared task. The 2007 Medical NLP challenge involved assigning ICD-9 codes to radiology reports [18]. The 2010 i2b2 obesity challenge consisted of predicting the status of obesity and 15 related comorbidities, requiring participants to determine whether the disorders were mentioned [19]. This dataset was recently used by Tang et al. [27] to compare several techniques for named entity recognition of disorders in hospital discharge summaries. The 2012 i2b2 challenge required the identification of clinically significant events (including clinical problems) and their temporal relationships [28].

One recent shared task specifically assessed the state-of-the-art on automatic disorder recognition in clinical narrative text. This task, called ShARE/CLEF eHealth Task 1 [3], required participants to locate the span of all disorders mentioned in the clinical narrative (Task 1a) – the task of named entity recognition (NER) – and normalize (or ground) the mention to a concept within the Disorder semantic group of the controlled vocabulary SNOMED-CT (Task 1b).

A notable trend that can be seen in the shared tasks is a shift in from lexical and natural language processing based techniques for (NER) to techniques based primarily on machine learning. In the 2010 i2b2 challenge, for example, the highest-performing systems used a lexical approach, while in the 2012 challenge, the highest-performing systems were hybrids employing machine learning for NER and rule-based approaches for normalization (or grounding). Most methods for automatic normalization of diseases

Download English Version:

<https://daneshyari.com/en/article/6928001>

Download Persian Version:

<https://daneshyari.com/article/6928001>

[Daneshyari.com](https://daneshyari.com)