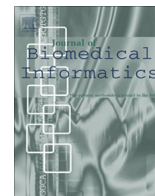




Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study

Maria Skeppstedt ^{a,*}, Maria Kvist ^{a,b,c}, Gunnar H. Nilsson ^d, Hercules Dalianis ^a

^a Department of Computer and Systems Sciences (DSV), Stockholm University, Forum 100, SE-164 40 Kista, Sweden

^b Department of Clinical Immunology and Transfusion Medicine, Karolinska University Hospital, Stockholm, Sweden

^c Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Sweden

^d Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:

Received 10 July 2013

Accepted 23 January 2014

Available online xxxx

Keywords:

Named entity recognition

Corpora development

Clinical text processing

Disorder

Finding

Swedish

ABSTRACT

Automatic recognition of clinical entities in the narrative text of health records is useful for constructing applications for documentation of patient care, as well as for secondary usage in the form of medical knowledge extraction. There are a number of named entity recognition studies on English clinical text, but less work has been carried out on clinical text in other languages.

This study was performed on Swedish health records, and focused on four entities that are highly relevant for constructing a patient overview and for medical hypothesis generation, namely the entities: *Disorder*, *Finding*, *Pharmaceutical Drug* and *Body Structure*. The study had two aims: to explore how well named entity recognition methods previously applied to English clinical text perform on similar texts written in Swedish; and to evaluate whether it is meaningful to divide the more general category Medical Problem, which has been used in a number of previous studies, into the two more granular entities, Disorder and Finding.

Clinical notes from a Swedish internal medicine emergency unit were annotated for the four selected entity categories, and the inter-annotator agreement between two pairs of annotators was measured, resulting in an average *F*-score of 0.79 for Disorder, 0.66 for Finding, 0.90 for Pharmaceutical Drug and 0.80 for Body Structure. A subset of the developed corpus was thereafter used for finding suitable features for training a conditional random fields model. Finally, a new model was trained on this subset, using the best features and settings, and its ability to generalise to held-out data was evaluated. This final model obtained an *F*-score of 0.81 for Disorder, 0.69 for Finding, 0.88 for Pharmaceutical Drug, 0.85 for Body Structure and 0.78 for the combined category Disorder + Finding.

The obtained results, which are in line with or slightly lower than those for similar studies on English clinical text, many of them conducted using a larger training data set, show that the approaches used for English are also suitable for Swedish clinical text. However, a small proportion of the errors made by the model are less likely to occur in English text, showing that results might be improved by further tailoring the system to clinical Swedish. The entity recognition results for the individual entities Disorder and Finding show that it is meaningful to separate the general category Medical Problem into these two more granular entity types, e.g. for knowledge mining of co-morbidity relations and disorder-finding relations.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Electronic health records contain valuable information in the form of symptom descriptions, documentation of examinations, diagnostic reasoning and motivations for treatment decisions. Automatic extraction of this information makes it possible to improve applications for patient care documentation, and enables

secondary usage of the information in the form of medical knowledge mining.

While a subset of the health record information, e.g. medication lists and diagnosis coding, is documented in a structured format, much important information is only available as free text [1]. An automatic summary of the free text part is therefore called for, providing health personnel with the possibility of forming a quick overview of the patient [2,3]. The information contained in health records can also be used for clinical text mining, i.e. to generate new medical knowledge from a large corpus of electronic health

* Corresponding author. Fax: +46 (0)8 703 90 25.

E-mail address: mariask@dsv.su.se (M. Skeppstedt).

<http://dx.doi.org/10.1016/j.jbi.2014.01.012>

1532-0464/© 2014 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Please cite this article in press as: Skeppstedt M et al. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study. J Biomed Inform (2014), <http://dx.doi.org/10.1016/j.jbi.2014.01.012>

records. Syndromic surveillance [4], comorbidity studies [5] and automatic detection of adverse drug reactions [6] are examples of clinical text mining applications.

An important component in information extraction from health record text is named entity recognition (NER) of relevant entities mentioned in the text, i.e. the automatic detection of spans of text referring to entities of certain semantic categories [7]. This study focuses on recognition of four entity categories that are particularly relevant for constructing a patient overview as well as for studies of co-morbidity [5,8], disorder and finding co-occurrences [9] and adverse drug reactions [6], namely the categories: *Disorder*, *Finding*, *Pharmaceutical Drug* and *Body Structure*.

There are previous studies on the recognition of clinical entities in English text, but very few studies have been carried out on clinical text written in other languages. The present study was performed on Swedish clinical text, and although both Swedish and English are Germanic languages, NER in Swedish poses additional challenges, as Swedish is more inflective and compounding of words occurs frequently. In addition, medical terminological resources are less extensive for Swedish than for English.

Moreover, previous annotation and NER studies have typically combined the two more granular entity categories *Disorder* and *Finding* into one more general category, e.g. called *Condition* or *Medical Problem* [10–12], or have focused only on the entity category *Disorder* [13,14]. To the best of our knowledge, there is only one previous corpus [15] in which the categories *Disorder* and *Finding* are annotated as two separate entity categories. The study describing the creation of this corpus does not, however, investigate the effect of this more granular division.

The present study therefore has two specific research questions:

- To what extent is it possible to use the NER methods, which have been successful for English clinical texts, on health record texts written in Swedish?
- To what extent is it possible to separate the more general entity category *Medical Problem* into the two more granular entity categories *Disorder* and *Finding*?

2. Related research

There are several studies that describe the creation of corpora annotated for named entities and that measure inter-annotator agreement scores between annotators (Table 1). There are also a number of studies in which the created corpora are used for training and/or evaluating NER systems (Table 2).

The annotation study by Chapman et al. [16] showed that detailed guidelines for annotating Clinical Conditions resulted in a substantially higher *F*-score than less detailed, but no significant differences in inter-annotator agreement between pairs of physicians and between physicians and lay people were found (but lay people required more training and had a lower ability to retain their annotation skills over time). Within the *Clinical E-Science Framework*, Roberts et al. [17] observed a higher *F*-score for lay people (a biologist/linguist and a computational linguist) than for a clinician, when measuring agreement to a constructed consensus set containing annotations for *Condition* (symptom and diagnosis), *Drug* or *Device* and *Locus* (e.g. anatomical structure or location). Wang [18] measured the inter-annotator agreement between two computational linguists annotating the categories *Finding* (corresponding to *Medical Problem*), *Substance* and *Body*, while Ogren et al. [19] measured the average agreement between four clinical data retrieval experts for annotating identical spans of text denoting the category *Disorder*. For the *i2b2 medication challenge* [20,21], inter-annotator agreement was calculated for annotations of *Medication*

Names on pre-annotated data. No statistically significant differences were observed between pairs of NLP community annotators, pairs of expert annotators, or pairs of experts annotating raw text. One participating group [22] annotated an additional subset of the development data provided for the challenge. Pre-annotation was also applied when annotating the *MiPACQ* corpus [15] for entity categories including *Disorder*, *Anatomy*, *Sign* or *Symptom*, and *Chemical* and *Drug*.

The created corpora have been used as training and evaluation data for machine learning-based NER systems and as evaluation data for rule- and terminology-based systems. An SVM (support vector machine) with uneven margins was trained on a subset of the *Clinical E-Science Framework* corpus [24], and two studies have been performed on the corpus created by Wang; one using the CRF (conditional random fields) package CRF++ [18] and one combining output from CRF++ with an SVM and an ME (maximum entropy) classifier [10]. All but one of the best performing systems in the *i2b2/VA challenge on concepts, assertions, and relations* used CRF for concept recognition [23,25]. The best performing system (by de Bruijn et al. [11]) instead used semi-Markov HMM. The second best (by Jiang et al. [12]) found that CRF (CRF++) outperformed SVM, and also managed to improve the results with a rule-based post-processing module. In the *i2b2 medication challenge*, on the other hand, which included the identification of *Medication Names*, a majority of the ten top-ranked systems were rule-based [20]. The best performing system (by Patrick and Li [22]) did, however, use CRF++, while the second best (by Doan et al. [26]) was built on terminology matching and a spell checker developed for drug names. This rule-based system was later employed by Doan et al. [27] in an ensemble classifier, together with an SVM and a CRF++ system. On the Ogren et al. [19] corpus, a terminology-based method for recognising disorders by matching to SNOMED CT has been evaluated [14,28], and there is also a terminology-based study for recognising diseases and drugs in Swedish discharge summaries, described by Kokkinakis and Thurin [13], in which the MeSH terminology was used.

Typical features used for training the machine learning models were the tokens (sometimes in a stemmed form), orthographics (e.g. number, word, capitalisation), prefixes and suffixes, part-of-speech information, as well as the output of terminology matching, which had a large positive effect in many studies (e.g. [10,18]). Most studies used features extracted from the current and the two preceding and two following tokens, while Roberts et al. [24] used a window size of ± 1 . The best performing system in the *i2b2/VA concepts challenge* used a very large feature set with a window size of ± 4 , also including character *n*-grams, word bi/tri/quad-grams and skip-*n*-grams, as well as sentence, section and document features (e.g. sentence and document length and section headings). In addition, features from semi-supervised learning methods were incorporated, in the form of hierarchical word clusters constructed on unlabelled data [11].

3. Methods

A corpus was first annotated and evaluated. Thereafter, suitable features for the NER task were evaluated and a model was trained using these features and evaluated on held-out data. Finally, an error analysis was carried out.

The IOB-encoding [7, pp. 763–764] of the annotated entities was used, as exemplified in Fig. 1. As machine learning algorithm, the CRF (conditional random fields) [29] implementation CRF++ [30] was chosen, which has been used in many previous clinical NER studies. CRF++ was used as linear chain CRF, in which each output variable is dependent on the previous and subsequent output variable.

Download English Version:

<https://daneshyari.com/en/article/6928423>

Download Persian Version:

<https://daneshyari.com/article/6928423>

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