A Hybrid Method for Normalization of Medical Concepts in Clinical Narrative

Yen-Fu Luo
Computer Science Department
University of Massachusetts Lowell
MA, USA
yenfu_luo@uml.edu

Weiyi Sun
Nuance Communication Inc.
MA, USA
weiyi_sun@yahoo.com

Anna Rumshisky
Computer Science Department
University of Massachusetts Lowell
MA, USA
arum@cs.uml.edu

Abstract—Normalization maps clinical terms in medical notes to concepts in standardized medical vocabularies. To complement the traditional lexical transformation based approach, we propose a hybrid normalization system which incorporates a deep learning model to capture semantic similarity between different surface expressions of the same concept. When evaluating our system against the mentions which may be normalized to existing concepts in the ShARe/CLEF eHealth 2013 dataset, our hybrid system achieves 90.6% in accuracy and outperforms a strong exact match + edit distance baseline by 2.6%. The results suggest the potential of the deep learning model to further improve the performance of normalization by mapping concept mentions to concepts using semantic similarity.

Keywords—Normalization; Clinical Terms; Deep Learning;

I. INTRODUCTION

Clinical findings, diagnoses, procedures, and medications recorded in the medical narratives are invaluable resources for clinical decision making, mortality prediction, adverse drug effect analysis, and etc. There are two major steps which include “Named Entity Recognition” (NER) and “Named Entity Normalization” (NEN) in order for effective utilization and exchange of the clinical information. NER for clinical notes which recognizes clinically-relevant mention spans has been well explored. The well-developed NER systems include MetaMap[1], cTAKES[2], and CLAMP[3]. NEN links named entities to concepts in standardized medical vocabularies. For example, “heart attack”, “MI”, and “myocardial infarction” refer to the same medical concept. The NEN task was explored in ShARe/CLEF eHealth 2013[4], SemEval-2014 Task 7[5] and SemEval-2015 Task 14[6] challenges. In general, most of the existing NEN systems normalize a concept mention to a concept using lexical transformation in combination with similarity ranking. However, these approaches may not normalize a concept mention which is semantically similar but typographically different from a concept. In this work, we focus specifically on the NEN task and propose a hybrid method which contains an additional deep learning model to measure semantic similarity on top of the current lexical transformation based system.

II. DATASET

ShARe/CLEF eHealth 2013 dataset[4], containing de-identified clinical free-text notes from the MIMIC II[7], includes 199 notes as training set and 99 notes as test set. The annotated disorder mention is defined as any span of text which may be mapped to a Concept Unique Identifier (CUI) in the SNOMED CT[8] and which belongs to one of the 11 UMLS semantic types including “Congenital Abnormality”, “Acquired Abnormality”, “Injury or Poisoning”, “Pathologic Function”, “Disease or Syndrome”, “Mental or Behavioral Dysfunction”, “Cell or Molecular Dysfunction”, “Experimental Model of Disease”, “Anatomical Abnormality”, “Neoplastic Process”, and “Signs and Symptoms”. If a disorder mention may not be mapped to a CUI, CUI-less is assigned.

III. METHODS

Automatically learned edit distance method at character level by Ghiasvand and Kate[9] was top-ranked team in SemEval 2015 clinical task. We therefore use Kate’s[10] open-source edit distance system as baseline and build our hybrid deep learning system on top of it. Our deep learning model contains four layers: 1) an embedding layer, 2) three Bi-LSTMs for the concept mention, the left context, and the right context, 3) a dense layer, and 4) a softmax layer. The deep learning model architecture is illustrated in Figure 1. Because the number of classes (CUIs) is large, sampled softmax loss is used during the training. Considering 30% of the data are diverse CUI-less mentions, CUI-less is assigned to a mention only when the CUI with the highest probability is below the threshold.

The concept mention and the left and right contexts are represented using word2vec embeddings[11]. Gensim[12] is used to train 200-dimensional word embeddings over all MIMIC III notes[13] with skip-gram, hierarchical softmax and the minimum frequency 3 settings. We pre-process the notes using the following steps: 1) using regular expressions to replace Protected Health Information (PHI) patterns with unique tokens; 2) removing punctuation; 3) replacing floating numbers with a unique token; 4) lowercasing.
IV. RESULTS

We evaluate our system against ShARe/CLEF test set under two settings: 1) on the overall mentions, including the CUI-less mentions, and 2) on the mentions which may be normalized to existing UMLS concepts, excluding the CUI-less mentions. Table I shows the system performance under two settings. In the first setting, the exact match + edit distance and the hybrid methods show the comparable performance. In the second setting, the hybrid method achieves 90.6% in accuracy and outperforms the exact match + edit distance baseline by 2.6%. In addition, our deep learning model may normalize the mention, “unarousable” to the concept, “lethargy” considering semantic similarity.

<table>
<thead>
<tr>
<th>Setting</th>
<th>EM+ED</th>
<th>EM+ED+DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting 1</td>
<td>89.8</td>
<td>89.6</td>
</tr>
<tr>
<td>Setting 2</td>
<td>88.0</td>
<td>90.6</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this work, we proposed a novel hybrid method for linking clinically relevant entity mentions in free-text medical notes to concepts in standardized medical vocabularies. The deep learning model which complements the lexical transformation based approach shows the potential to capture the semantic similarity between medical concepts and mentions.

ACKNOWLEDGMENT

This work was supported in part by a research grant from Philips HealthCare.

REFERENCES


