Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Recurrent neural networks for classifying relations in clinical notes

Yuan Luo

Department of Preventive Medicine, Division of Health and Biomedical Informatics, Northwestern University, Chicago, IL, United States

ARTICLE INFO

Article history:
Received 15 March 2017
Revised 13 June 2017
Accepted 6 July 2017
Available online 8 July 2017

Keywords:
Natural language processing
Medical relation classification
Recurrent neural network
Long Short-Term Memory
Machine learning

ABSTRACT

We proposed the first models based on recurrent neural networks (more specifically Long Short-Term Memory - LSTM) for classifying relations from clinical notes. We tested our models on the i2b2/VA relation classification challenge dataset. We showed that our segment LSTM model, with only word embedding feature and no manual feature engineering, achieved a micro-averaged f-measure of 0.661 for classifying medical problem-treatment relations, 0.800 for medical problem-test relations, and 0.683 for medical problem-medical problem relations. These results are comparable to those of the state-of-the-art systems on the i2b2/VA relation classification challenge. We compared the segment LSTM model with the sentence LSTM model, and demonstrated the benefits of exploring the difference between concept text and context text, and between different contextual parts in the sentence. We also evaluated the impact of word embedding on the performance of LSTM models and showed that medical domain word embedding help improve the relation classification. These results support the use of LSTM models for classifying relations between medical concepts, as they show comparable performance to previously published systems while requiring no manual feature engineering.

1. Introduction

In knowledge representation, identifying relations from text documents is important for creating or augmenting structured knowledge bases and in turn supporting question answering, inference reasoning and decision making. The task usually breaks down to annotating unstructured text with named entities and identifying the relations between these annotated entities. State-of-the-art named entity recognizers can now recognize concept with high accuracy [1], but relation extraction is not as straightforward. In the biomedical and clinical domain, extracting relations from scientific publications and clinical narratives has also been an important focus over the past decade with numerous challenges due to the complexity of language and domain specific knowledge involved [2].

Biomedical relation extraction is critical in understanding clinical notes, facilitating automated diagnostic reasoning and clinical decision making. In pathology reports, immunophenotypic features are often written as relations among medical concepts. For example, in "Studies performed at MGH reveal that the [lymphoid cells] are [CD10] *positive*, [BCL6] *positive*, and [BCL2] *negative*.", "lymphoid cells", "CD10", "BCL6" and "BCL2" are medical concepts; "CD10", "BCL6" and "BCL2" are biomarkers of the cell. If one only captures bag-of-words or bag-of-concepts features and do not account for how concepts are interrelated, one would fail to encode in such feature representation whether "lymphoid cells" are positive or negative for "CD10", "BCL6" and "BCL2". In this and many other similar situations, the relations between the biomedical concepts need to be understood in the context of syntactic and/or semantic cues in order to resolve possible ambiguities.

In a broad sense, one can define a relation as a tuple $r(c_1, c_2, \ldots, c_n)$, $n \geq 2$, where c_i's are biomedical concepts (e.g., cells, biomarkers, etc.), and the c_i's are semantically and/or syntactically interconnected by an overarching relation r, as expressed in text. Note that such a definition requires a relation to at least involve two concepts and precludes either a single concept or an assertion of a single concept from being regarded as a relation. Specifically, if n is two, we call the relation a two-concept relation. In the previous sentence example, one may treat the sentence as encoding a relation between four medical concepts that are of interest. One may also use the term relation to specifically refer to two-concept relations, for example

```
positive-expression(lymphoid cells, CD10)
positive-expression(lymphoid cells, BCL6)
negative-expression(lymphoid cells, BCL2)
```

From the perspective of composite relations, one may be able to decompose a multi-concept relations using certain logics over a list of two-concept relations, for example

E-mail address: Yuan.luo@northwestern.edu

http://dx.doi.org/10.1016/j.jbi.2017.07.006
1532-0464/© 2017 Elsevier Inc. All rights reserved.

```
and (positive-expression (lymphoid cells, CD10),
positive-expression (lymphoid cells, BCL6),
negative-expression (lymphoid cells, BCL2))
```

In some cases, logics can become more complex than the Boolean logic when we need to understand what are often referred to as events, which are defined as grammatical objects that combine lexical elements, logical semantics and syntax [3]. For example, the ternary relation `treated_by (patient, Harvoni, 8-week course)` as expressed in "[the patient] was *administered* [Harvoni] for an [8-week course]" can be understood as an event, where the event trigger is "administered", the theme is the Hepatitis C medication "Harvoni" and the target argument is "patient". Clearly, with a variety of logics such as temporal logic one can represent increasingly flexible events and relations. Two-concept relations are building blocks of such compositions and the most frequent forms of relations; correctly classifying two-concept relations will produce fundamental insights on how to devise better natural language processing (NLP) algorithms for elucidating the interactions between biomedical concepts.

2. Background and related work

Some of the critical clinical information contained in clinical narratives can be represented by relations of concepts. Biomedical relations are critical in facilitating applications such as clinical decision making, clinical trial screening, pharmacovigilance https://www.ncbi.nlm.nih.gov/pubmed/28643174 [4–12]. Determining the exact relation between the two concepts requires an understanding of the context in which the two concepts are discussed.

Part of the advances in the state-of-the-art specialized clinical NLP systems for identifying medical problems have been documented in challenge workshops such as the yearly i2b2 (Informatics for Integrating Biology to the Bedside) Workshops, which have attracted international teams to address successive shared classification tasks. One such challenge focused in part on identifying the relations that may hold between medical problems and treatments, between medical problems and tests, as well as between pairs of medical problems [13]. Many systems applied Support Vector Machines (SVMs) to tackle the relation extraction task by combining lexical, syntactic, and semantic features. Some systems adopted a two-step approach by first determining the candidate pairs that did not relate to each other, and then classifying the specific relation type for the rest of the candidate pairs [14–16]. Some teams added annotated and/or unannotated external data to complement their machine learning system [15,17]. Other teams complemented their machine learning systems with rules that capture simple linguistic patterns of relations [18].

All challenge participating systems involved heavy feature engineering; they explored lexical, semantic, syntactic, general domain and medical domain ontology features [13]. Many systems also harvested features from existing NLP pipelines such as cTakes [19] and MetaMap [20]. Systems that use many human engineered features often do not generalize well to new datasets [21]. In general domain NLP, a growing number of studies have successfully used recurrent neural networks (RNNs) combined with word embedding [22] on tasks including language modeling [23], text classification [24–27], question answering [25,26,28,29], machine translation [25,30–32], named entity recognition [33–36], and relation classification [37,38]. Inspired by general domain successes, recent progress on applying RNNs to clinical datasets also aims to reducing the amount of engineered features and has achieved some success on modeling both structured and unstructured clinical data. For structured clinical data, Choi et al. [39] applied Gated Recurrent Unit networks (GRUs) for early detection of heart failure onset using time-stamped medical events (diagnosis, medications and procedures). They showed RNNs outperformed multiple statistical learning models including logistic regression, support vector machine (SVM), k-nearest neighbor (kNN), and multi-layer perceptron (MLP). Che et al. [40] applied GRUs to perform mortality and diagnosis code prediction using time series data consisting of physiologic measurements, lab-tests values, and prescriptions. Their GRU-based model showed better AUC than logistic regression, SVM, and random forests (RF). Lipton et al. [41] trained Long Short-Term Memory networks (LSTMs) to classify 128 diagnoses from 13 frequently but irregularly sampled clinical measurements from patients in pediatric ICU. Their model showed significant improvements with respect to several strong baselines, including multilayer perceptron trained on hand-engineered features. Razavian et al. [42] used LSTMs to predict onset of 133 diseases and conditions simultaneously based on 18 common lab tests measured over time. They showed that the LSTM learned representations outperformed a logistic regression baseline with hand engineered features. Pham et al. [43] used LSTMs to model the longitudinal records of diagnoses, medications and procedures and made dynamic predictions of future diagnoses, medications and procedures. They showed improved performance over competitive models including SVM and RF. For unstructured clinical data, Dernoncourt et al. [44] applied bi-directional LSTMs to de-identifying patient notes. They adopted two bi-directional LSTM layers, one at character level and the other at word level. Their character level embedding and LSTM aim to address data sparsity due to out-of-vocabulary tokens, misspellings, and different noun forms or verb endings. The two-layer bi-directional LSTMs showed improved de-identification performance from state-of-the-art Conditional Random Field (CRF) models. Jagannatha et al. [45] applied bidirectional RNNs using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) to recognize named entities or concepts such as medications, diseases and their associated attributes (e.g. frequency of medications). Their bi-directional LSTMs showed significant improvement from state-of-the-art CRF models. We refer the reader to Miotto et al. [46] for a comprehensive review of other related deep learning approaches for healthcare applications. In general, there have been fewer studies on applying RNNs to unstructured data than those to structured data in the clinical domain. This is likely due to the lack of large clinical corpus available to train word or phrase embeddings. To address this issue, Jagannatha et al. [45] combined an EHR corpus of 99,700 clinical notes with English Wikipedia and PubMed Open Access articles to train word embedding. The recent release of 2 million clinical notes from MIMIC-III database [47] has at least partially alleviated the corpus issue. In fact Dernoncourt et al. [44] used the MIMIC-III corpus as the embedding training corpus for de-identification. We used MIMIC-III trained word-embedding to enable the clinical relation classification. Our models differ from general domain relation classification models [37,38], in that we do not use syntactic/ semantic resources (compared to Yan et al. [37]), and we explicitly distinguish the words within and surrounding the two concepts (compared to Zhou et al. [38]). To the best of our knowledge, this work is the first attempt on using recurrent neural networks to classify the semantic relations between candidate concepts in the clinical notes.

3. Data

In this work, we used the relation classification data from the 2010 i2b2/VA challenge, which includes relations between medical problems and treatments (TrP), relations between medical problems and tests (TeP), as well as relations between medical