Entity disambiguation with memory network

Yaming Sun a,∗, Zhenzhen Ji a, Lei Lin a, Xiaolong Wang a, Duyu Tang b

a Computer Science Department, Harbin Institute of Technology, Harbin, China
b Microsoft Research, Beijing, China

A R T I C L E   I N F O

Article history:
Received 16 March 2017
Revised 26 October 2017
Accepted 6 November 2017
Available online xxx
Communicated by Guan Ziyu

Keywords:
Entity disambiguation
Memory network
Natural language processing
Deep learning

A B S T R A C T

We develop a computational approach based on memory network for entity disambiguation. The approach automatically finds important clues of a mention from surrounding contexts with attention mechanism, and leverages these clues to facilitate entity disambiguation. Unlike existing feature-based methods, this approach does not rely on any manually designed features. Unlike existing neural models such as recurrent or convolutional neural network, this approach leverages the importance of context words in an explicit way. The model could be easily trained with back-propagation. To effectively learn the model parameters, we automatically collect large-scale mention-entity pairs from Wikipedia as training data. We verify the effectiveness of the proposed approach on a benchmark dataset from TAC-KBP evaluation in 2010. Experimental results demonstrate that our approach empirically surpasses strong feature based and neural network based methods. Model analysis further reveals that our approach has the capacity to discover important clues from contexts.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Entity disambiguation is a fundamental problem in natural language processing. The task aims at linking a mention, which is a name string from a document, to an entity in the reference knowledge base such as Wikipedia [1]. For instance, given a sentence “Apple will launch a trio of new iPads this spring, according to Barclays research analysts.” and the mention “Apple”, the goal of entity disambiguation is to understand that the mention “Apple” in this specific context refers to Apple Inc. 1 rather than Apple 2. Entity disambiguation helps us to better understand an entity through gathering their information from huge number of documents on the web. Moreover, it could help us to better comprehend a document with the augmented entity information from knowledge base.

In natural language processing community, entity disambiguation is typically considered as a ranking problem. The goal is to rank candidate entities based on their semantic similarities with the mention [2–4]. Towards this goal, a major challenge is how to get an effective mention representation. It is widely accepted that context plays a crucial role in entity disambiguation, since context words carry important evidences about a mention. Taking the aforementioned example, without knowing the mention Apple is related to the electronic product iPads, it is hard for an intelligent system or even a person to comprehend which Apple is talked about. Existing studies typically use manually designed features or neural network architectures to model the context information. Neural network approaches achieve promising performances in recent years, which have the capacity to learn powerful text representation from data without using manually designed features. In this direction, denoising autoencoder [3] and convolutional neural networks (CNN) [4,5] have been explored to model the context information. However, existing neural network approaches have the same weakness: they could not explicitly reveal the importance of context words. For example, in the aforementioned example, the context word “iPads” is more important than the context word “spring” for understanding the mention “Apple”. Accordingly, “iPads” and “spring” should have different degrees of importance when they are used to represent the mention. A desirable system should have the ability to automatically discover important context evidences and effectively utilize them for entity disambiguation.

In this work, we develop a memory network based approach for entity disambiguation. Our approach includes two external memories. One memory represents the context words of a mention, and the other one stands for the descriptive words that occur in an entity’s Wikipedia page. We use the same strategy to deal with the mention part and the entity part. Taking the mention part as an example, our approach regards context words as the memory, and regards the mention as the query to find important evidences from the memory. The result is a context-sensitive mention representa-

1 https://en.wikipedia.org/wiki/Apple_Inc.
tion, which could be viewed as a new query to repeat this process. In this way, the model has the ability to interact with the memory for multiple times, and learn complex functions with multiple levels of composition. As each component is differentiable, the entire model could be trained with standard back propagation. We leverage Wikipedia and use the anchor text to collect large-scale training data without using manual annotation.

We evaluate the performance of our approach on a benchmark dataset from TAC-KBP evaluation [1]. We compare with both feature-based and neural network based methods. Empirical results demonstrate that our approach performs better than strong baseline methods. Model analysis further shows that the important evidences discovered by the model are consistent with our intuition.

This paper is organized as follows. Section 2 describes related research works in literature. Section 3 introduces the background of memory network. Section 4 presents the details of the proposed approach. Experimental study is given in Section 5, followed by the summary and future plans.

2. Related work

Entity disambiguation is typically regarded as a ranking task, which calls for measuring the similarity between the context of a mention and the text associated with a candidate entity (e.g. the text in the corresponding page in KB). Existing algorithms for entity disambiguation can be generally divided into local approaches and global (collective) approaches. The former [16,7] uses local statistics of a mention \( m \) and an entity \( t \). The latter [8–11] takes consideration of all mentions in a given document simultaneously. Both directions require the semantic relatedness between a mention \( m \) and an entity \( t \). Representative mention features include document surface features such as lexical and part-of-speech tags of context words, entropy based representations [7], and structured text representations such as dependency paths and topic feature representation [12]. Typical entity features include name tagging, KB infoboxes, synonyms, and semantic categories [13,14]. Since feature engineering is time-consuming and requires expert knowledge, many neural network approaches are developed to learn text representation automatically from data. For example, He et al. [3] used denoising autoencoder to encode mention document and entity document. Sun et al. [4] embedded mention, context and entity separately, and used a neural network framework to integrate these representations for entity disambiguation. Fang et al. [15] paid attention to word and entity embeddings and proposed to jointly embedding knowledge base and text in the same vector space. Ganea and Hofmann [16] focused on learning entity embedding and selectively leveraging contexts through a local attention mechanism. The intuition of the local attention module is analogous to our work. The method developed in this paper differs from previous work (except for [16]) in that the contexts of mention and descriptive words of entity are not treated equally.

LSTM [17,18] has been proven effective in many NLP applications, and could be used to implicitly learn the importance of context words through adaptively remembering the input and forgetting the history. However, LSTM based approach could not explicitly reveal the importance of context words, and it is not non-trivial to do multiple step reasoning over the contexts. In our approach, important contexts are automatically extracted and leveraged in an explicit way. Our approach is based on memory network, which reasons over memory cell through multiple computational layers (hops), each of which explicitly conducts weighted average over the memory cells. We implement our model in a local setting without capturing all the mentions involved in a document. We believe that implementing our model in a collective setting [11,19] could obtain better performance, and we leave this as a future plan.

The backbone of our approach is memory network [20], which is a general machine learning framework that has the ability to read from and write to external memory. Memory network has been successfully applied to many natural language processing tasks including question answering [21], language modeling [22], machine comprehension [23], aspect-based sentiment classification [24], etc. For example, in the question answering task, the memory comes from supporting sentences, and the query is the input question which iteratively interacts with the memory. Sukhbaatar et al. [22] showed that using multiple hops (computational layers) could learn more abstractive evidences from memory and thus achieve better performance. To the best of our knowledge, our work is the first attempt that applies memory network to entity disambiguation.

3. Background: memory network

Our approach is inspired by the recent success of memory network in natural language processing [20,22–24]. We briefly describe the background of memory network in this section.

Memory network is a general machine learning paradigm [20]. The central idea is reasoning over external memory. The memory component could be read, written to, and jointly learned with the goal of using it for prediction. Formally, a memory network is composed of a memory \( m \), and four components \( I, G, O, R \). The memory \( m \) is typically implemented with an array of objects, such as an array of vectors. The \( I \) component maps the input to internal feature representation. The \( G \) component updates old memories with the input. The \( O \) component generates an output representation based on the input and the current memory state. The \( R \) component outputs a response based on the output representation.

Let us take question answering as an example to explain how memory network works. Given a list of sentences and a question, the task aims to find evidences from those sentences and generate an answer. In the inference process, the \( I \) component reads one sentence \( s_t \) each time and maps it into a vector representation. Then the \( G \) component updates a memory \( m_t \) based on the current sentence representation. We finally get a memory matrix \( m \) which stores the semantic meanings of all sentences. Given a question \( q \), the memory network also maps it into a vector representation \( e_q \). Afterwards, the \( O \) component uses \( e_q \) to find question related evidences from the memory \( m \) and generates an output vector \( o \). Finally, the \( R \) component takes \( o \) as the input and outputs the final response.

It is worth noting that the \( O \) component could consist of multiple computational layers (hops). The intuition is that more abstractive representation could be learned if the model could recursively leverage previously extracted evidences. Sukhbaatar et al. [22] demonstrated that multiple hops could uncover more abstractive evidences than single hop on both question answering and language modeling.

4. Methodology

We present the proposed approach in this section. The goal is to calculate the semantic similarity between a given mention and every candidate entity. The top ranked candidate entity will be regarded as the answer.

4.1. Approach overview

The input of our approach includes four parts: a mention, the contexts that surround the mention, a candidate entity, and the descriptive words that occur in the candidate entity’s KB article. The first two information are used to get context-sensitive mention
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات