A Brief Review of Relation Classification

Jiaming Shen, Yuning Mao, Sha Li, Xiaotao Gu

Department of Computer Science
University of Illinois at Urbana-Champaign

April 26, 2018
Problem Description

Problem Definition
Given a sentence containing entities $e_1$ and $e_2$, identify the most probable semantic relation $R$ between the entities.

Example
Sentence: Bill Gates founded Microsoft.
Entities: (Bill Gates, Microsoft)
Relation: found
## Methods Overview

<table>
<thead>
<tr>
<th>Pattern</th>
<th>kernel</th>
<th>feature</th>
<th>RecursiveNN</th>
<th>CNN</th>
<th>RNN</th>
<th>LSTM</th>
<th>attention</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Brin, 1998]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Agichtein and Gravano, 2000]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Zelenko et al., 2003]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Culotta and Sorensen, 2004]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Kambhatla, 2004]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Bunescu and Mooney, 2005]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Zhao and Grishman, 2005]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[GuoDong et al., 2005]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Mooney and Bunescu, 2006]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Mintz et al., 2009]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Socher et al., 2012]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Hashimoto et al., 2013]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Riedel et al., 2013]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Liu et al., 2013]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Zeng et al., 2014]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Zeng et al., 2015]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Zhang et al., 2015]</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Liu et al., 2015]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xu et al., 2015a]</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xu et al., 2015b]</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Nguyen and Grishman, 2015]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Santos et al., 2015]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Jiang et al., 2016]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Cai et al., 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Xiao and Liu, 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Shen and Huang, 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Miwa and Bansal, 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Lin et al., 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ren et al., 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Wang et al., 2016]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ji et al., 2017]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Liu et al., 2017]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Pattern-based: Snowball [Agichtein and Gravano, 2000]

Example (S: seed, P: pattern, C: corpus)

$\rightarrow$ S: (Bill Gates, Microsoft)
$\Rightarrow$ C: Bill Gates founded Microsoft.
$\Rightarrow$ P: \{Arg1\} founded \{Arg2\}.
$\Rightarrow$ C: Steve Jobs founded Apple.

$\rightarrow$ S: (Steve Jobs, Apple)
$\Rightarrow$ C: Apple was founded by Steve Jobs.
$\Rightarrow$ P: \{Arg1\} was founded by \{Arg2\}.
Pattern-based: MetaPAD [Jiang et al., 2017]

A meta pattern refers to a frequent, informative, and precise subsequence pattern of entity types (e.g., $Person, $Politician, $Country) or data types (e.g., $Digit,$Month, $Year), words (e.g., “politician”, “age”), phrases (e.g., “primeminister”), and punctuation marks, which serves as an integral semantic unit in certain context.

Pipeline

- Incorporate entity typing information.
- Pattern discovery as frequent pattern mining.
- Pattern quality assessment with a few training samples.
- Synonymous pattern grouping.
Supervised Model and Distant Supervision

Supervised Relation Classification

- Extract features from the target sentence (POS tags, dependency parsing, words between entities...)
- Train a supervised classification model (SVM, CRF) based on large annotated data. Determine: whether this sentence indicates some relation between the entities. [Zhao and Grishman, 2005, Culotta et al., 2006]
- Pro: significantly outperforms pattern-based methods!
- Con: requires great human efforts to collect labeled data.

Distant Supervision [Mintz et al., 2009]

- For any $R(e_1, e_2)$ in a knowledge base (e.g. Freebase), sentences containing such an entity pair is selected as a positive instance for $R$.
- Reduce noise: multi-instance learning [Surdeanu et al., 2012]
DS+Pattern: Universal Schema [Riedel et al., 2013]
Deep Learning Models: CNN [Zeng et al., 2014]

Automatic feature extraction

▶ Incorporate word embeddings to capture semantic meanings.
▶ Use Convolutional Neural Networks (CNN) to capture local features (turns out to be very useful in relation inference).
▶ Use position feature to incorporate word position information.

Figure 1: Architecture of the neural network used for relation classification.

Figure 2: The framework used for extracting sentence level features.
Deep Learning Models: RNN [Zhang et al., 2015]

Shortest Dependency Path

SDP-LSTM

(a) Softmax

Hidden layer

LSTM for word embeddings
LSTM for POS embeddings
LSTM for GR embeddings
LSTM for WordNet embeddings

Dependency paths

(b) Hidden layer

Pool

LSTM

Left sub-path

LSTM

Right sub-path

poured
gallons
have
been
into
tillion
of
[region]_{e_2}
A
[water]_{e_1}
an
empty
of
space
Development of Relation Classification

Pattern ⇒ Feature-based ⇒ Distant Supervision ⇒ Deep Learning

Figure: Percentage of recent research works on four main categories of methods for relation classification.
References I

Snowball: Extracting relations from large plain-text collections.
In *Proceedings of the fifth ACM conference on Digital libraries*, pages 85–94. ACM.

Extracting patterns and relations from the world wide web.

A shortest path dependency kernel for relation extraction.


References IV


References VI


References


