Can Syntactic Indicators Help Relation Extraction ?

陶琼星

2019.04.08

Contents

Introduction

- **Background:**

  关系抽取是对句子中实体对之间的语义关系进行分类，是自然语言处理中的一项重要任务，被广泛应用于知识图谱构建、问答系统和网络搜索等。

- **Relation Extraction:**

  给定一个句子和已经标注的两个entity，判断在这个句子中两个entity的关系。

- **Example:**
  - “The magician shows a <e1>**bottle**</e1> full of <e2>**soda**</e2>.”

    → *Content-Container(e2,e1)*
  - "The ice <e1>**chunks**</e1> fell from an <e2>**airplane**</e2>."

    → *Entity-Origin(e1,e2)*

Related Work

	Model	Contributions
CNN based	CNN(+WordNet) (Zeng et al., 2014)	Sentence Level Feature+Lexical Level Feature(5个词汇特征) “a [psittacosis] of green [parrots] doing a couple of loops.” WordNet:parrot-**bird**
	CR-CNN (dos Santos et al., 2015)	1.Ranking Loss: $L=\log(1+\exp(\gamma(m^{+}-s_{\theta}(x)_{y^{+}})))+\log(1+\exp(\gamma(m^{-}+s_{\theta}(x)_{c^{-}})))$ 2.Only use the text between entities**(使用两个实体间的文本能达到相近结果)**
	Attention CNN (+POS) (Huang and Shen, 2016)	"[Outbreaks] **caused by** the oral vaccine's live [virus] have happened before."
	Multi-Attention CNN (Wang et al. 2016)	Word Attention+Attention-Based Pooling. (在Max Pooling层使用Attention,不同窗口的重要性不同)
RNN based	Bi-LSTM (Zhang et al., 2015)	Bi-LSTM+WordNet,+Dependency Parser,+Entity Recognizers (NER)
	Attention Bi-LSTM (Zhou et al., 2016)	Bi-LSTM+Attention
	Hierarchical Attention Bi-LSTM (Xiao and Liu, 2016)	$[x_1,...,x_{i-1}]$ $[x_i]$ $[x_{i+1},...,x_{j-1}]$ $[x_j]$ $[x_{j+1},...,x_n]$
	Entity Attention Bi-LSTM (Lee et al., 2019)	利用两个entity和**entities潜在类型（LET）**作为特征

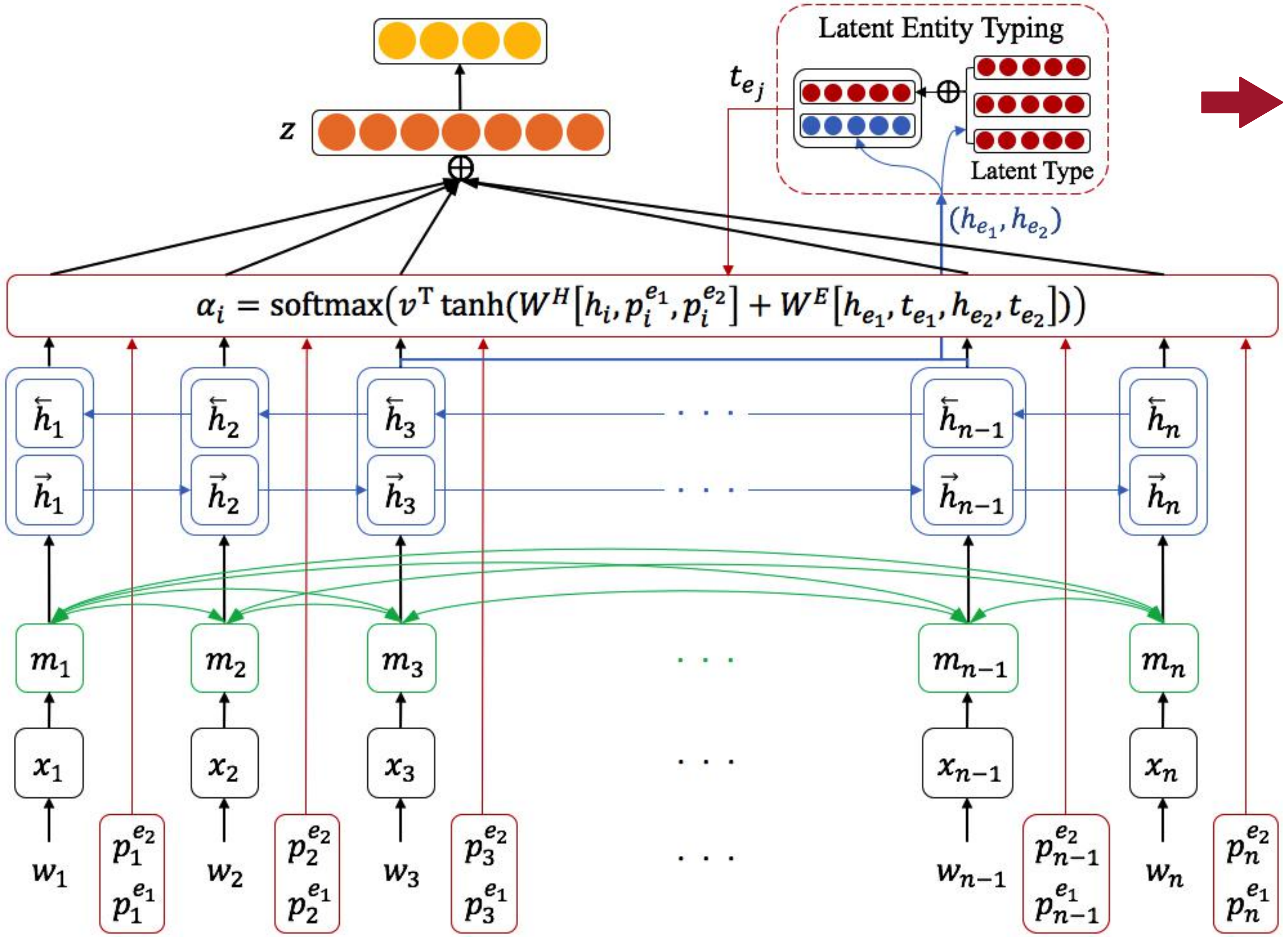
Related Work

目前从论文结果看，在不使用High-Level Lexical Features的情况下，SemEval 2010 Task 8数据集上**最高的F1-Score为：88%，但目前没人复现**，其次是2019年的论文：**85.2%**。

- **已有的模型：**
  - 输入层的标配：word embedding + position embedding
  - 损失函数Ranking Loss 比 softmax 效果好
  - Attention效果好（主要为：word 与 relation，word 与 entity）

- **Our Model:**
  - Give more attention to Syntactic Indicators(caused by,inside,of,focus on,filled with...)
  - New Loss Function(整个句子序列中有较多噪音，Syntactic Indicators序列中内容缺失)

Baseline Model



通过两个entity，两个entity的潜在类型(LET)和position来计算每个隐藏状态的attention权重，得到关系的表示

- Entity-aware Attention
- Bi-LSTM
- Self Attention
- Word Representation
- Input Sentence

Semantic Relation Classification via Bidirectional LSTM Networks with Entity-aware Attention using Latent Entity Typing(Lee et al., 2019)

Syntactic Indicators Selection

两个假设：

- 两个实体之间的词，涵盖了足够的信息用于实体关系分类
  - The most common <e1>**audits**</e1> **were about** <e2>**waste**</e2> and recycling.
  - The <e1>**room**</e1> **was filled with huge Jack-the-Dripper** <e2>**canvases**</e2>.
  - When I'm walking alone, I count the <e1>**steps**</e1> **of the** <e2>**stairs**</e2>, and I also guess approximately how many steps are left till I reach the next street.

- 两实体间的动词、介词或短语等这些关系的触发词对关系分类起决定性作用
  - **audits were about waste** → *Message-Topic(e1,e2)*
  - **room was filled with** huge Jack-the-Dripper **canvases** → *Content-Container(e2,e1)*
  - **steps of** the **stairs** → *Component-Whole(e2,e1)*

Syntactic Indicators Selection

- 用nltk对句子进行词性标注，根据两个实体中间词的POS，设计5个规则提取关系触发词：
    - 消除并列词

      drawings ~~and videos~~ that have been shown in **museums**
    - 只留VB,NN,IN,TO,PRP词性的词

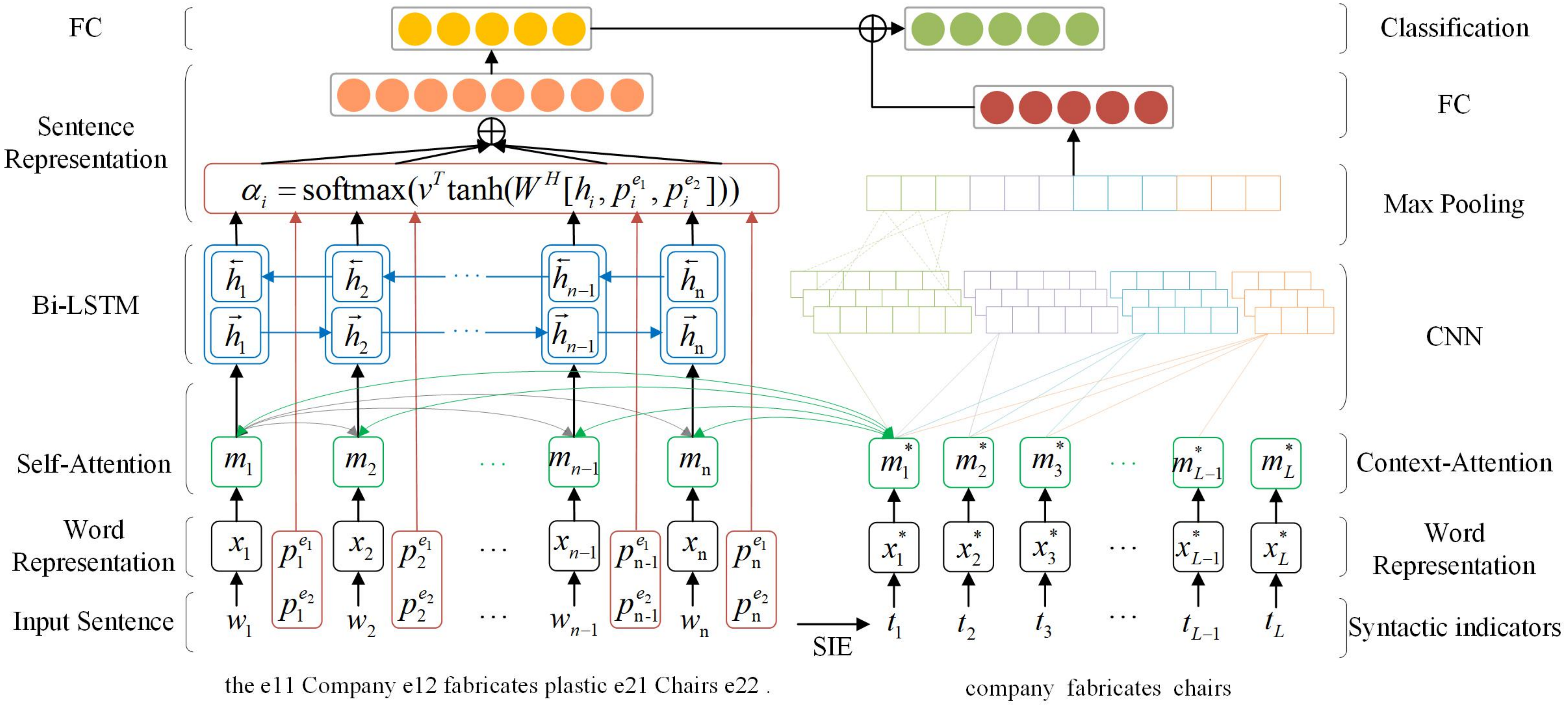
      coins are enclosed in ~~a clear hard plastic~~ **case**
    - 简化多个名词组合词

      treaty establishes ~~majority~~ **rule**
    - 简化两个实体的描述

      colors ~~of rainbow~~ are caused by **dispersion**
    - 保留靠近第二个entity的动作

      analyzer ~~identifies paths~~ using **method**

Our Model



FC
Sentence Representation
Bi-LSTM
Self-Attention
Word Representation
Input Sentence
$\alpha_i = \mathrm{softmax}(v^T \tanh(W^H[h_i, p_i^{e_1}, p_i^{e_2}]))$
SIE
the e11 Company e12 fabricates plastic e21 Chairs e22 .
company fabricates chairs
Classification
FC
Max Pooling
CNN
Context-Attention
Word Representation
Syntactic indicators

Experiments and Results

- **DataSet：** Semeval-2010 task 8，定义了9 种有方向的关系和other 类(不属于9 种类别)，是评测有监督方法的数据集。

Relation	Freq	Pos	IAA
Cause-Effect	1331 (12.4%)	91.2%	79.0%
Component-Whole	1253 (11.7%)	84.3%	70.0%
Entity-Destination	1137 (10.6%)	80.1%	75.2%
Entity-Origin	974 (9.1%)	69.2%	58.2%
Product-Producer	948 (8.8%)	66.3%	84.8%
Member-Collection	923 (8.6%)	74.7%	68.2%
Message-Topic	895 (8.4%)	74.4%	72.4%
Content-Container	732 (6.8%)	59.3%	95.8%
Instrument-Agency	660 (6.2%)	60.8%	65.0%
Other	1864 (17.4%)	N/A[4]	N/A[4]
Total	10717 (100%)		

《Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals》

Experiments and Results

我们的模型获得了**85.9的F1-score**，在不使用High-Level Lexical Features的方法中，超过了目前最好结果（除Multi-Attention CNN）。

与Entity Attention Bi-LSTM(Lee et al., 2019)相比，**我们的结果高出了1.2**，而且比它加入了LET特征后的结果还**高出了0.7**。

与其他加入WordNet、DPT、DEP、NLP tags、NER tags等High-Level Lexical Features的方法相比，我们的模型也取得了最好结果。

Model	F1
CNN (Zeng et al., 2014)	78.9
+WN	82.7
CR-CNN (Dos Santos et al., 2015)	84.1
Attention CNN (Shen and Huang, 2016)	84.3
+POS,WN,WAN	85.9
BLSTM (Zhang et al., 2015)	82.7
+POS,NER,DEP,WN	84.3
Attention BLSTM (Zhou et al., 2016)	84.0
H-Att BLSTM (Xiao and Liu, 2016)	84.3
Attention BLSTM (Lee et al., 2019)	84.7
+LET	85.2
Our Model	**85.9**

Analysis and Discussions

Syntactic Indicators Sequence

① 整个句子，Self Attention+Bi-LSTM+Attention

② Syntactic Indicators Sequence，Context Attention+CNN(size=1,2,3,4)+Max Pooling

③ 整个句子，分别进行①②操作

④ Our model（将①②的输出相加后用于关系分类）

Method	F1
Whole Sentence	84.4
Indicators Sequence	82.5
Whole Sentence + Whole Sentence	84.0
Whole Sentence + Indicators Sequence	**85.9**

Analysis and Discussions

- **Parallel Loss Function**

$$L = -\sum_{i=1}^{|D|} \log p\left(y^{(i)} \middle| S^{(i)}, \theta\right) - \lambda \|\theta\|_2^2$$

$$L' = -\sum_{i=1}^{|D|} \log p_1\left({y_1}^{(i)} \middle| S^{(i)}, \theta\right) - \lambda' \sum_{i=1}^{|D|} \log p_2\left({y_2}^{(i)} \middle| S^{(i)}, \theta\right) + \lambda \|\theta\|_2^2$$

Loss	F1
l	84.5
l'	**85.9**

Conclusion

- **Additional Syntactic Indicators improves Relation Extraction.**
- **Give more attention to Syntactic Indicators can make full use of imformative words and reduce noise.**
- **Parallel Loss Function outperforms single loss function.**
- **Future Work**

Solve the problem of rare Syntactic Indicators by Synonymy phrase using WordNet.

Example :A [film]e1 revolves around a [cadaver]e2 who seems to bring misfortune on those who come in contact with it. → *Message-Topic(e1,e2)*

Reference

[1] Wang L, Cao Z, De Melo G, et al. Relation classification via multi-level attention cnns[J]. 2016.

[2] Huang X. Attention-based convolutional neural network for semantic relation extraction[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 2526-2536.

[3] Santos C N, Xiang B, Zhou B. Classifying relations by ranking with convolutional neural networks[J]. arXiv preprint arXiv:1504.06580, 2015.

[4] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[J]. 2014.

[5] Xiao M, Liu C. Semantic relation classification via hierarchical recurrent neural network with attention[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 1254-1263.

[6] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2016, 2: 207-212.

[7]Zhang S, Zheng D, Hu X, et al. Bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 29th Pacific Asia conference on language, information and computation. 2015: 73-78.

Thank you for your listening!