

Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection (NIPS2011)

Zhe Han

1401214342

iampkuhz@gmail.com

2015 年 1 月 9 日

- Paraphrase identification(复述检测)
 - Definition
 - Common methods
- 本文的方法
 - Recursive Autoencoder
 - Dynamic Pooling
- 实验效果
 - 实验效果
 - 分析, 对比其他任务

Paraphrase identification

- definition
 - 给定一组句子, 判断其是否是复述
 - binary classification
- Microsoft Research Paraphrase Corpus (MSRP)
 - train: 4,076 sentence pairs (2,753 positive: 67.5 %)
 - test: 1,725 sentence pairs (1,147 positive: 66.5 %)
 - 2 个标注者, 83% 的一致性, 第三个人更正

Sample data

- Sentence 1: Amrozi accused his brother, whom he called "the witness", of deliberately distorting his evidence.
- Sentence 2: Referring to him as only "the witness", Amrozi accused his brother of deliberately distorting his evidence.
- Class: 1 (true paraphrase)

Paraphrase identification

- Common methods
 - lexical features
 - n-gram features, skip-gram features, ...
 - semantic features
 - POS tag, wordnet similarity, dependency tree relation, ...
 - classification
 - SVM, voted classifications
- Challenge
 - 没有提取句子的全局信息 (dependency features 利用不足)
 - 对句子涵义的特征提取不足 (没有真正理解句子)

This paper

Main method

- 利用 NYT 新闻训练每个单词的向量 (100 维)
 - 对于每个句子 (多个单词向量) 采用训练一个递归的自动编码器, 得到一个句子级别的语义向量.
 - 通过判断两个句子的语义向量的相似性得到语义相似性特征
-
- 递归的自动编码器 (Unfolding Recursive Autoencoder)
 - 抽取句子的语义向量, 得到语法数上每个节点 (单词, 短语) 的向量
 - Dynamic Pooling
 - 对于长度变化的两个句子, 抽取固定维数的特征

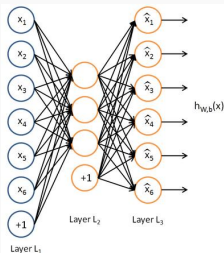
Unfolding Recursive Autoencoder

● Autoencoder

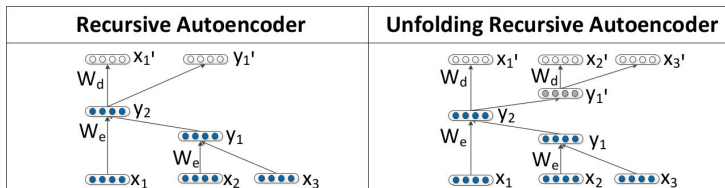
- 不对所有层的参数进行一次性的优化, 而是一层层的优化
- 对于每一层的输入, 跑一层神经网络, 得到输入的特征表示 (认知过程)
- 对于之前得到的特征再跑一层神经网络, 我们希望输入和输出尽可能相似 (生成过程)

● Recursive Autoencoder

- 进一步的, 对于深层的网络 (语法树), 递归使用同一个简单的 Autoencoder

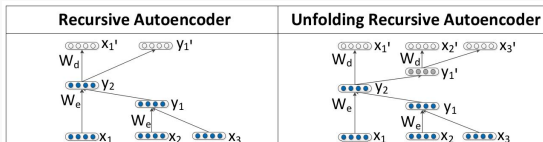


Unfolding Recursive Autoencoder



- Recursive Autoencoder
- Comparison (on $y_2 - x_1 y_1$)
 - Neural network
 - minimum $\| y'_2 - y_2 \|$
 - Recursive Autoencoder
 - minimum $\| [x'_1; y'_1] - [x_1; y_1] \|$
 - Unfolding Recursive Autoencoder
 - minimum $\| [x'_1; x'_2; \dots; x'_j] - [x_1; x_2; \dots; x_j] \|$

Unfolding Recursive Autoencoder(RAE)



- 初始化每个单词的向量
 - 100 维，可以通过 word2vec 或 glove 实现
- 得到句子的 constituency tree
 - 二叉

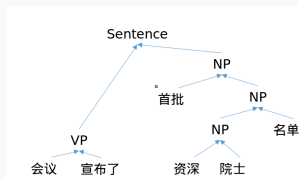


Figure : constituency tree

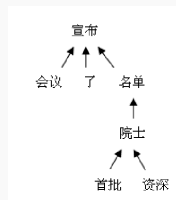
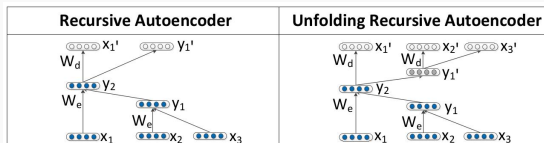


Figure : dependency tree

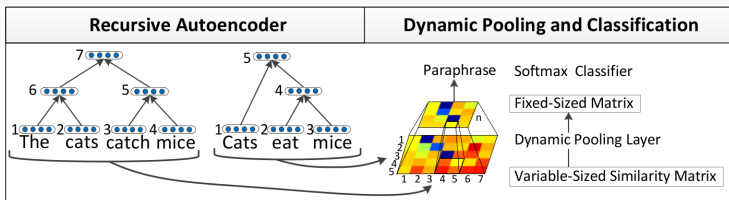
Unfolding Recursive Autoencoder(uRAE)



损失函数

- 对于左图的关系: $y_1 \rightarrow x_2 x_3, y_2 \rightarrow x_1 y_1$
 - 先正向 (自底向上): 对于规则 $p \rightarrow c_1 c_2$, 有 $p = f(W_e[c_1, c_2] + b)$, 带入两个关系 y_1, y_2 依次实例
 - 后逆向 (自顶向下): 对于上面的规则 $[c'_1; c'_2] = f(W_d p + b_d)$, 如果 $c'_1; c'_2$ 不是叶节点, 递归做
- 损失函数为 $E_{rec}(y_{(i,j)}) = \| [x_i; \dots; x_j] - [x'_i; \dots; x'_j] \|^2$
 - 对比 RAE: 逆向时只做一层, 损失函数为
$$E_{rec}(p) = \| [c_1; c_2] - [c'_1; c'_2] \|^2$$
- 梯度下降求解 W_e, b, W_d, b_d

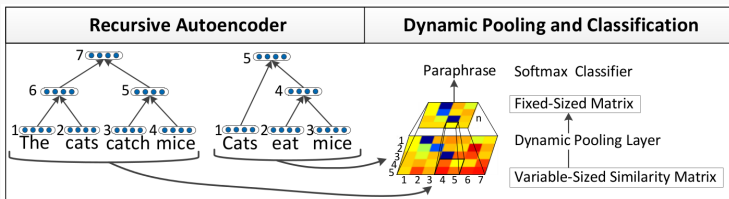
Dynamic Pooling



● motivation

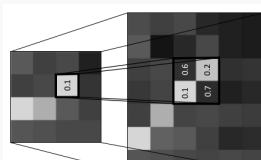
- 如何对两个长度变化的句子（而且很可能不一样）抽取固定维数的特征？
 - 长度为 n 的句子，cTree 有 $(2n-2)$ 个节点，不同的句子对应的维数不同
- polling! 把不同长度的句子压缩（扩张）到相同的维数
 - (实验验证，15 维最好，略低于平均句子长度)

Dynamic Pooling



method

- 将向量“等”分成 k ($=15$) 份，原来的 $(2n-1) \times (2m-1)$ 维向量分成 k^2 块，从每块中提取最小值作为该块的值
 - 平均值掩盖了特异的特征，最小值 (or 最大值?) 更容易体现其中一个的特异特征



Experiment result(2011)

- features

- Dynamic pollinged matrix $S(\mathbb{R}^{15 \times 15})$, three number features, sentence length, string mathes

Model	Acc.	F1
All Paraphrase Baseline	66.5	79.9
Rus et al. (2008) [16]	70.6	80.5
Mihalcea et al. (2006) [17]	70.3	81.3
Islam and Inkpen (2007) [18]	72.6	81.3
Qiu et al. (2006) [19]	72.0	81.6
Fernando and Stevenson (2008) [20]	74.1	82.4
Wan et al. (2006) [21]	75.6	83.0
Das and Smith (2009) [15]	73.9	82.3
Das and Smith (2009) + 18 Features	76.1	82.7
Unfolding RAE + Dynamic Pooling	76.8	83.6

Das and Smith (2009)	product of experts	supervised	76.1%	82.7%
Wan et al. (2006)	dependency-based features	supervised	75.6%	83.0%
Socher et al. (2011)	recursive autoencoder with dynamic pooling	supervised	76.8%	83.6%
Madhani et al. (2012)	combination of eight machine translation metrics	supervised	77.4%	84.1%

Analysis

● QA

● Why use uRAE instead of RAE or Recursive.avg?

- 多个单词组成的句子（高层节点），需要更多的单词信息，RAE 只关心最近的 2 个儿子节点
- Recursive.avg：两个儿子向量的平均忽视了结构关系
- 实验证明，Recursive.avg 找不出来；RAE 对 2 个单词组成的短语，识别其近义词效果很好；uRAE 对于 2-3 个单词组成的短语的效果很好，甚至 5 个单词组成的短语有些也可以正确找到。

Center Phrase	Recursive Average	RAE	Unfolding RAE
the U.S.	the U.S. and German	the Swiss	the former U.S.
suffering low morale	suffering a 1.9 billion baht UNK 76 million	suffering due to no fault of my own	suffering heavy casualties
to watch hockey	to watch one Jordanian border policeman stamp the Israeli passports	to watch television	to watch a video
advance to the next round	advance to final qualifying round in Argentina	advance to the final of the UNK 1.1 million Kremlin Cup	advance to the semis
a prominent political figure	such a high-profile figure	the second high-profile opposition figure	a powerful business figure
Seventeen people were killed	"Seventeen people were killed, including a prominent politician "	Fourteen people were killed	Fourteen people were killed
conditions of his release	"conditions of peace, social stability and political harmony "	conditions of peace, social stability and political harmony	negotiations for their release

Analysis & Summary

- QA

- Does deep RAE improve the acc?
 - No. Slow and worse. 过拟合，和两个儿子非常像，忽视了更深的叶节点，长的短语 (>2) 效果非常差

- Summary

- 非常漂亮
 - 特征的类别很少 (4 类)，语义信息起到了很大的作用
 - (我自己) 21 种 feature, 77.2% (单一 SVM: 76.8%)
 - 之前的文章，鲜有使用 semantic info。而且一般只用了 wordNet 的同义词/上位词，结合 dependency pair 进行判断，局限于单个 label，没有全局信息，对整体的效果提升没有这么显著
- 利用文章的句子语义表述方法
 - 维基百科谓词归一：根据描述句子的相似性判断谓词是否表达同一意思