

# 2014\_ARC\_论文阅读

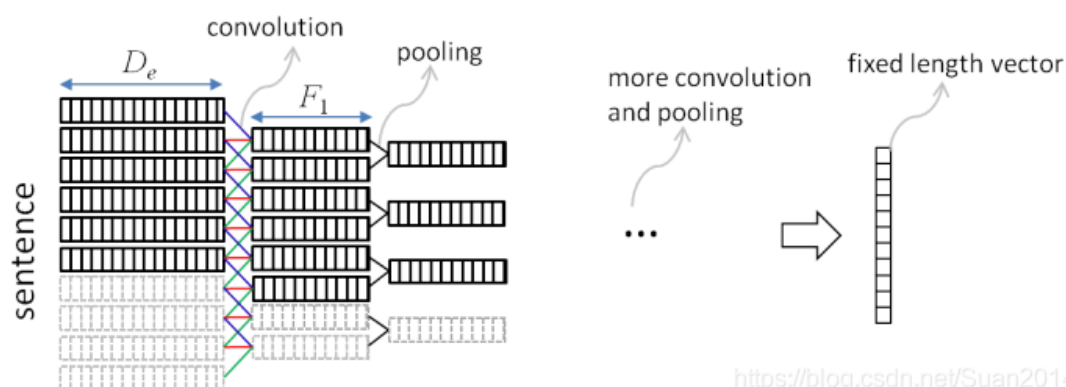
- 论文题目：Convolutional Neural Network Architectures for Matching Natural Language Sentences
- 论文链接：<https://arxiv.org/abs/1503.03244v1>
- 代码链接：<http://nlp.stanford.edu/~socherr/classifyParaphrases.zip>
- 代码语言：matlab 【没看代码】

## 摘要

较好的匹配算法能够同时为句子内部结构和句子间的交互信息建模。本文提出的模型通过逐层构图和合并充分表达句子的层次信息，而且能够在不同的层次捕捉到丰富的匹配模式。该算法无需先验知识，而且具有通用性。

## 句子建模

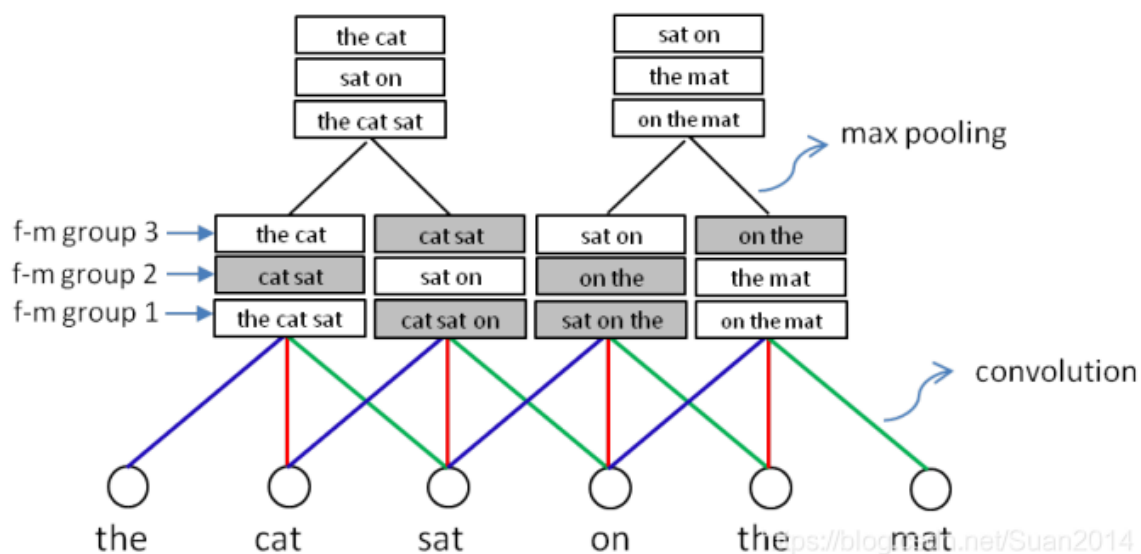
句子匹配任务的核心是能够充分地句子建模（即提取sentence-level的特征），本文基于CNN为句子建模，结构如图1所示：



流程：规定句子最大长度为 $M$ ，将句子中的词通过词嵌入表达成一个向量，假设向量维度为 $d$ ，则 $M \times d$ 的矩阵为该模型的输入（未达到最大长度的部分采用zero-padding），通过卷积层+max\_pooling层将输入矩阵转化为一个固定长度的向量。

### 这样为什么有效呢？

如图2所示，采用卷积能够从句子中提取出局部的语义组合，如图2中的 the cat sat和cat sat on等这样的短语组合。卷积多个filter能够提取多个feature map，这样就能够从多个角度提取语义组合，如图2中的第一个滑动窗涵盖的feature map1 是the cat sat的特征，feature map2 为cat sat的特征，feature map3为the cat的特征。采用max pooling层会对多种语义组合进行选择，过滤掉一些置信度低的特征组合，图2中亮的表示特征值较大。

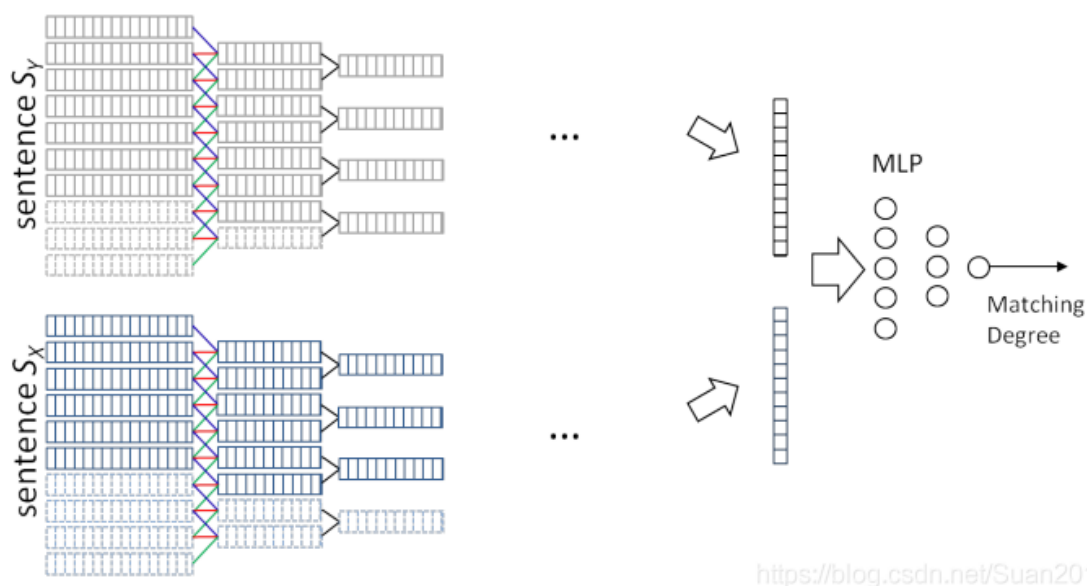


## 匹配模型

句子建模后能够提取出一个句子的特征向量，接下来需要利用特征向量对句子对进行匹配建模。

### ARC-I

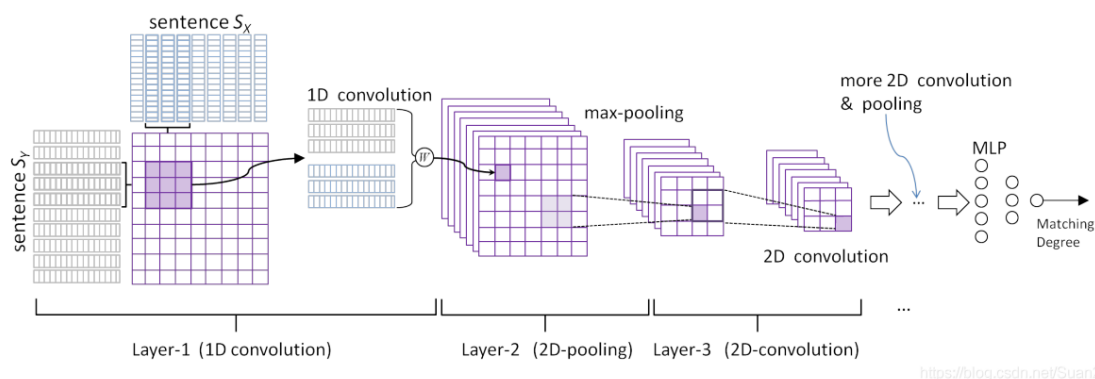
模型结构：



这种结构比较直观，对两个句子分别进行特征提取，提取后的两个特征做拼接之类的操作合成一个向量，再喂给MLP进行2分类。缺点是：**两个句子是单独提取特征的，没有交互**。使得句子在提取特征时丢失了很多重要的匹配信息，提取特征过程中没有体现句子间的交互性。

### ARC-II

- 模型结构：

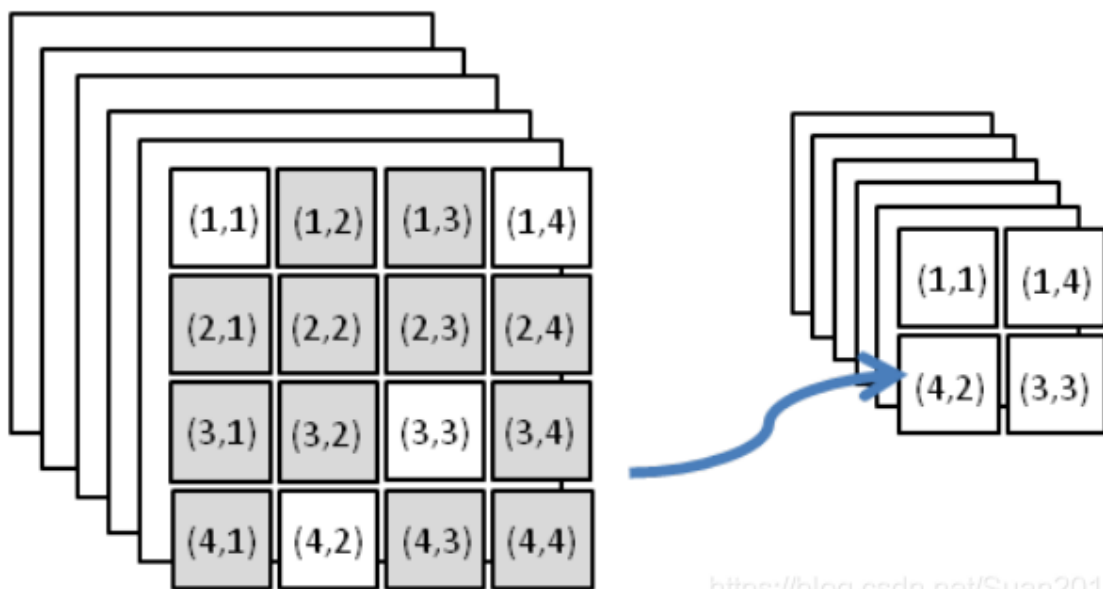


### 做法：

将X与Y中的向量两两拼接在一起【例如： $x_1-y_1, x_1-y_2, x_2-y_1, x_2-y_2$ 】，然后做1D卷积，得到得结果再加权求和，构成一个新的矩阵（相当于XY交互的产物），然后再通过正常的卷积、池化等操作。

这样做之所以有效：

- 1.对句子建模时，就利用了句子间的交互信息
- 2.保留了词序信息，如下图所示。



【上图类似一个卷积池化后的结果，可以看到保留了一些词序信息】

## 训练

- loss函数：排序损失函数 ( ranking-based loss )：

$$e(x, y^+, y^-) = \max(0, 1 + s(x, y^-) - s(x, y^+))$$

合页损失，句子x与**与其匹配的句子**间的相似度得分要高于**与其不匹配的句子**间相似度得分。

- batch\_size: 100~200
- word embedding维度：50，算法Word2Vec
- 数据：英文数据：Wikipedia(~1B Words) 中文：微博数据(~300M Words)

## 实验结果

- 三个匹配任务：句子自动填充、推文匹配、同义词匹配
- 效果：结构 1 和结构 2 远好于其他算法，结构 2 稍好于结构 1

Model	P@1(%)
Random Guess	20.00
DEEPMATCH	32.50
WORDEMBED	37.63
SENMLP	36.14
SENNA+MLP	41.56
URAE+MLP	25.76
ARC-I	47.51
ARC-II	<b>49.62</b>

Table 1: Sentence Completion.

Model	P@1(%)
Random Guess	20.00
DEEPMATCH	49.85
WORDEMBED	54.31
SENMLP	52.22
SENNA+MLP	56.48
ARC-I	59.18
ARC-II	<b>61.95</b>

Table 2: Tweet Matching.

Model	Acc. (%)	F1(%)
Baseline	66.5	79.90
Rus et al. (2008)	70.6	80.50
WORDEMBED	68.7	80.49
SENNA+MLP	68.4	79.70
SENMLP	68.4	79.50
ARC-I	69.6	80.27
ARC-II	69.9	80.91

Table 3: The results on Paraphrase.