TITLE:
Simple and Effective Text Matching with Richer Alignment Features

GET:

论文技术核心在Alignment Features

论文任务是Simple and Effective Text Matching

✓ 作者

- Department of Computer Science and Technology, Nanjing University, China
- 南京大学计算机科学与技术系主任、人工智能学院院长 周志华
- Alibaba Group, Hangzhou, China
- 阿里巴巴人工智能实验室

✓ 摘要

- 作者探索简单高效的神经网络结构，所以：

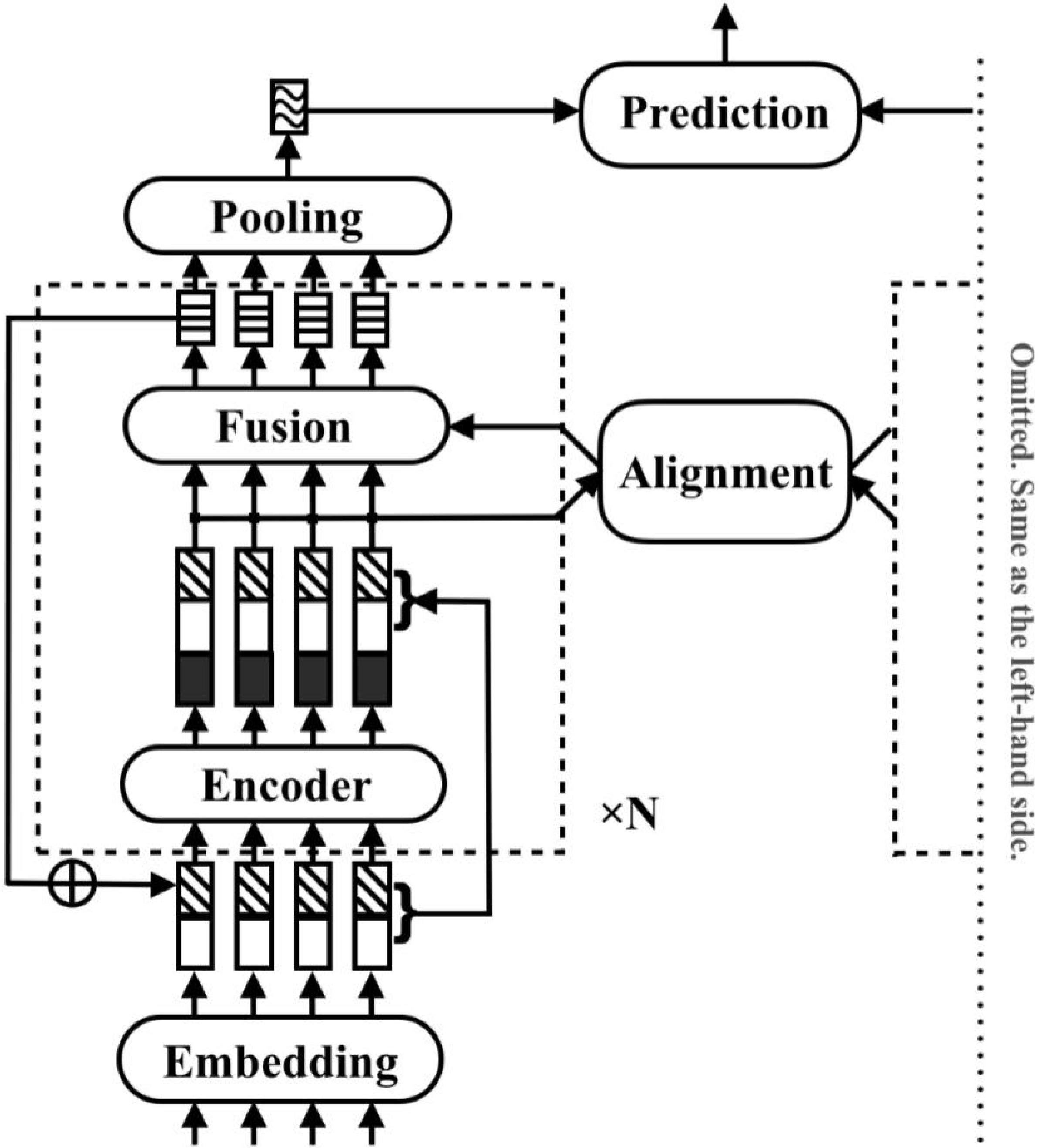
We explore what is sufficient to build a fast and well-performed **text matching model** and propose to keep **three** key features available for **inter-sequence alignment**: original point-wise features, previous aligned features, and contextual features while simplifying all the remaining components.

✓ 介绍

- 提出问题：以前的文本匹配模型很重，因为：

We question the necessity of many slow components in text matching approaches presented in previous literature, including complicated multi-way alignment mechanisms, heavy distillations of alignment results, external syntactic features, or dense connections to connect stacked blocks when the model is going deep.

以上是摘要的铺垫，该论文模型结构见下页



RE2:

观察图中间位置， alignment and fusion layers的输入向量由三部分组成：

① previous aligned features(**R**esidual vectors)

② Original point-wise features(**E**mbedding vectors)

③ contextual features(**E**ncoded vectors)

Several same-structured **blocks** consisting of encoding, alignment and fusion layers then process the sequences consecutively

✓ Our Approach

- 提出两点:

1. 左右对称网络，输入是两段文本
2. 所有参数共享，除了prediction layer

✓ Our Approach

- 2.1 Augmented Residual Connections
- 目的是得到richer features for alignment processes，所以：

$$x_i^{(n)} = [x_i^{(1)}; o_i^{(n-1)} + o_i^{(n-2)}], \qquad (1)$$

- 即：embedding layer的输出(空心矩形)拼接上前面两个blocks的输出之和(斜线矩形)。

- 遇到疑问：

contextual features 是怎么计算的，作用是什么

✓ Our Approach

- 2.2 Alignment Layer
- 相似度得分

eij between ai and bj is computed as the dot product

- 互相表示（from线性代数)

The output vectors a’ and b’ are computed by weighted summation of representations

$$e_{ij} = F(a_i)^T F(b_j). \tag{2}$$

$$\begin{aligned} a'_i &= \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ik})} b_j, \\ b'_j &= \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{kj})} a_i. \end{aligned} \tag{3}$$

✓ Our Approach

- 2.3 Fusion Layer
- 概要：

The fusion layer compares local and aligned representations in three perspectives and then fuse them together.

思考：构造这些丰富的特征方式，值得借鉴

$$
\begin{aligned}
\bar{a}_i^1 &= G_1([a_i; a_i']), \\
\bar{a}_i^2 &= G_2([a_i; a_i - a_i']), \\
\bar{a}_i^3 &= G_3([a_i; a_i \circ a_i']), \\
\bar{a}_i &= G([\bar{a}_i^1; \bar{a}_i^2; \bar{a}_i^3]),
\end{aligned} \tag{4}
$$

✓ Our Approach

- 2.4 Prediction Layer
- 概述：

左右两边text经过pooling池化再输入到prediction层。公式7是公式6的简化版。

$$\hat{\mathbf{y}} = H([v_1; v_2; |v_1 - v_2|; v_1 \circ v_2]). \tag{6}$$

$$\hat{\mathbf{y}} = H([v_1; v_2]). \tag{7}$$

✓ Experiments

- 数据集介绍

SciTail(Khotetal.,2018)(Science Entailment) is an entailment classification dataset constructed from science questions and answers. This dataset contains 27k examples in total. 10k examples are with entailment labels and the remaining 17k are labeled as neutral. Accuracy is used as the evaluation metric for this dataset.

也就是label=[0,1]

Model	Params	Acc.(%)
DecAtt (Parikh et al., 2016)	0.6M	86.8
BiMPM (Wang et al., 2017)	1.6M	86.9
ESIM (Chen et al., 2017)	4.3M	88.0
DIIN (Gong et al., 2018)	4.4M	88.0
MwAN (Tan et al., 2018)	14M	88.3
CAFE (Tay et al., 2018b)	4.7M	88.5
HIM (Chen et al., 2017)	7.7M	88.6
SAN (Liu et al., 2018)	3.5M	88.6
CSRAN (Tay et al., 2018a)	13.9M	88.7
DRCN (Kim et al., 2018)	6.7M	**88.9**
RE2 (ours)	2.8M	**88.9±0.1**
BiMPM (ensemble)	6.4M	88.8
DIIN (ensemble)	17M	88.9
CAFE (ensemble)	17.5M	89.3
MwAN (ensemble)	58M	89.4
DRCN (ensemble)	53.3M	**90.1**
RE2 (ensemble)	22.4M	89.9

Results on Natural Language Inference on SNLI dataset 经过对比single models and ensemble models,

Model	Acc(%)
ESIM (Chen et al., 2017)	70.6
DecompAtt (Parikh et al., 2016)	72.3
DGEM (Khot et al., 2018)	77.3
HCRN (Tay et al., 2018c)	80.0
CAFE (Tay et al., 2018b)	83.3
CSRAN (Tay et al., 2018a)	**86.7**
RE2 (ours)	**86.0±0.6**

- SciTail test set： This dataset is considered much more difficult with fewer training data available and generally low accuracy as a binary classification problem.

Model	time(s/batch)
BiMPM (Wang et al., 2017)	0.05 ± 0.00
CAFE† (Tay et al., 2018b)	0.07 ± 0.01
DIIN† (Gong et al., 2018)	0.85 ± 0.11
DIIN with EM feature†	1.79 ± 0.22
CSRAN† (Tay et al., 2018a)	0.28 ± 0.02
RE2 (1 block)	0.03 ± 0.00
RE2 (2 blocks)	0.04 ± 0.00
RE2 (3 blocks)	0.05 ± 0.00

- Inference time when batch size = 8 on Intel Core i7 CPUs. Models with † marks use POS tags as external syntactic features and the computation time of POS tagging is not included.

With the highly efficient design, our method can perform well without any strong but slow building blocks like recurrent neural networks, dense connections or any syntactic features.

	SNLI	Quora	Scitail	WikiQA
original	88.9	89.4	88.9	0.7740
w/o enc-in	87.2	85.7	78.1	0.7146
residual conn.	88.9	89.2	87.4	0.7640
simple fusion	88.8	88.3	87.5	0.7345
alignment alt.	88.7	89.3	88.2	0.7702
prediction alt.	88.9	89.2	88.8	0.7558
parallel blocks	88.8	88.6	87.6	0.7607

补充pointwise 、 pairwise 、 listwise

- Who established the Nobel Prize?
  1. The Nobel Prize was established more than 100 years ago.
  2. The Fields Medal, established in 1936, is often described as the Nobel Prize of mathematics.
  3. The Nobel Prize was established in the will of Alfred Nobel.

- 打卡的baseline模型
- 打卡的任务场景和数据集
  a. 相似度计算&复述识别
  b. 问答匹配
  c. 对话匹配
  d. 自然语言推理/文本蕴含识别
  e. 信息检索中的匹配
  f. 机器阅读理解问题
- 打卡的Siamese结构（基于表示）
- 打卡的花式attention结构（基于交互）
- 打卡的ranking学习与评估方法
- 打卡的预训练模型
- 打卡的开源工具

短文本匹配中的一个大杀器---ESIM

- ESIM，简称 “Enhanced LSTM for Natural Language Inference“。顾名思义，一种专为自然语言推断而生的加强版 LSTM。
- 精细的设计序列式的推断结构。
- 考虑局部推断和全局推断。
- 作者主要是用句子间的注意力机制(intra-sentence attention)，来实现局部的推断，进一步实现全局的推断

短文本匹配中的一个大杀器---ESIM

- ESIM 牛逼在它的 inter-sentence attention，就是上面代码中的 soft_align_attention，这一步中让要比较的两句话产生了交互。以前我见到的类似 Siamese 网络的结构，往往中间都没有交互，只是在最后一层求个余弦距离或者其他距离。