

Sentence Matching With Deep Self-Attention and Co-Attention Features

1. Introduction

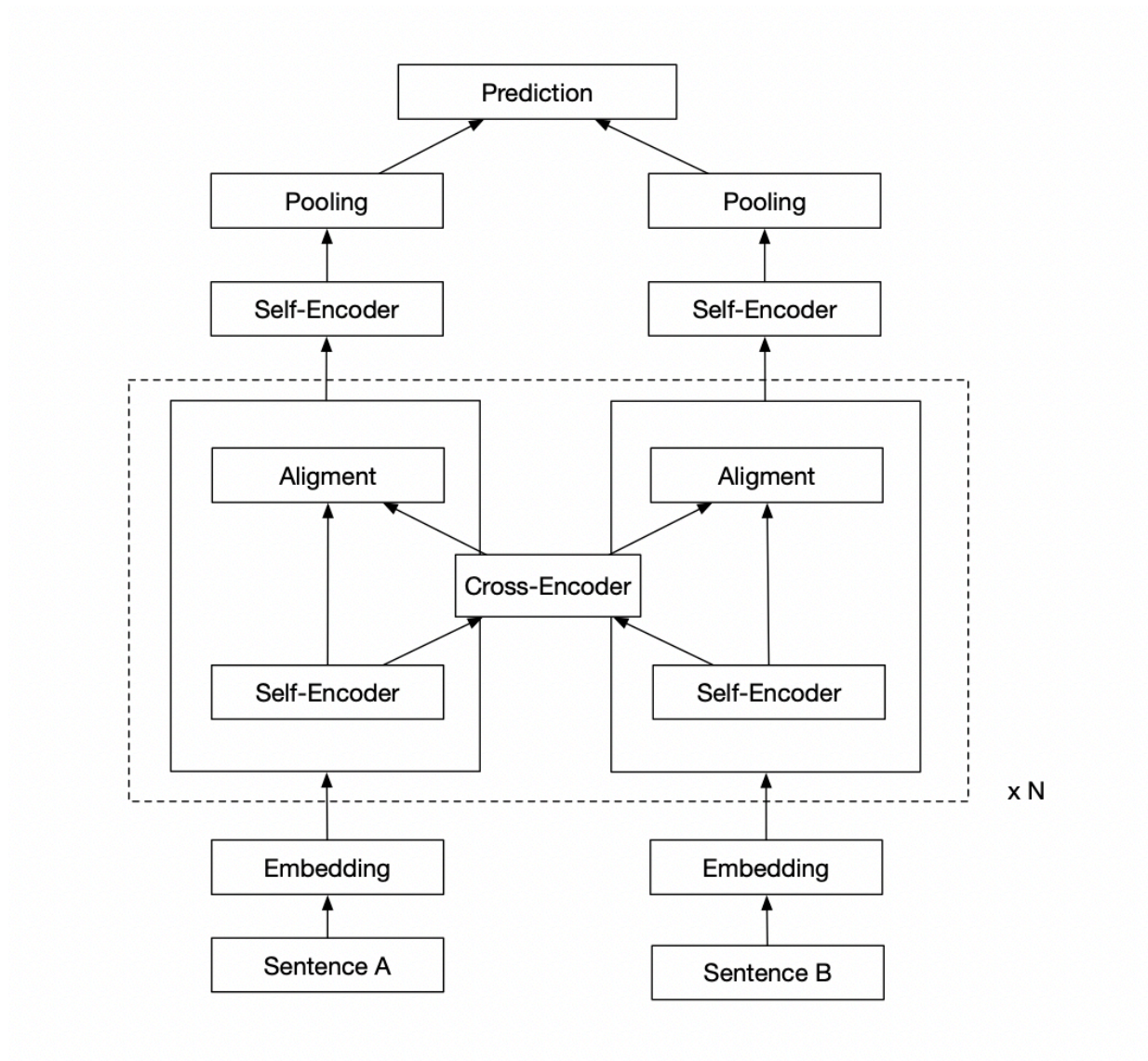
1.1 Motivation

- Attention机制 强大编码能力
- Attention机制 运行速度优势
- Attention机制 提取交互特征

1.2 Contribution

- 目前来看，第一个在编码阶段仅使用Attention机制，其他论文均是要用CNN或者LSTM提取特征
- 保持领先的准确率的同时，减少模型推理时间（暂时未测具体数值）

2. Framework



2.1 Embedding

使用预训练好的开源词向量，glove-800B-300D

2.2 Self-Encoder

使用Self-Attention对句子进行编码

2.3 Cross-Encoder

使用Cross-Attention对两个句子进行交互编码

2.4 Pooling

Pooling是指使用TextCNN及Max-Pooling对输出进行Pooling操作

2.5 Prediction

Prediction是一个两层的全联接网络，输入是[a, b, a-b, abs(a-b), a * b]

3. Experiments

3.1 Experiments on SNLI

Num	Model	Acc on SNLI test
1	ESIM	88.6
2	RE2	88.9
3	DRCN	88.9
4	Our Model	88.8

3.2 Experiments on Quora

Num	Model	Acc on Quora test
1	ESIM	88.6
2	BiMPM	89.2
3	DIIN	89.1
4	Our Model	89.4

3.3 Experiments on Scitail

Num	Model	Acc on Scitail test
1	ESIM	70.6
2	CAFE	83.3
3	RE2	86.0
4	Our Model	85.7

3.4 Experiments on WikiQA

Todo

3.5 Experiments on Multi-NLI

Todo

3.6 Experiments on Other-Datasets

Todo

4. Conclusion

- 模型能达到State-of-the-art, 验证了模型的有效性

5. Future Work

- 在其他数据集上进行测试
- Embedding层加入更多特征（如字符特征）
- 测试推理时间