
NLPCC 2017 DBQA

语义计算与知识检索

韩 云

yunhan@pku.edu.cn

May 27, 2018

1 DBQA

问答任务 (QA) 的目标是构建可以回答自然语言问题的系统, 基于文档的回答 (Document-based Question Answering, DBQA) 是 QA 的子任务, 它旨在衡量问题和答案之间的匹配关系, 这可以视为句子匹配问题。

在本次实验中, 我们针对 NLPCC 2017 中的 Task 5: Open Domain Question Answering 中的 Document-based QA (DBQA) task。

2 小组成员

- 姓名: 韩云
- 学号: 1701214083

3 实验方法

QA 属于 NLP 中一类比较重要的问题, 我们通常将基于文档的问答看做两段文本的语义匹配的问题。语义匹配也是 NLP 中许多应用程序的关键任务, 旨在对句子建模, 然后计算相似性或者相关性。最近, 深度学习的很多方法取得了很大的进步, 因为它能自动学习给定的任务的最佳特征表示, 包括对句子建模。在神经网络模型中, LSTM, CNN 是对句子和句子对建模的常用模型。其中, CNN 擅长提取特征, 并且能够进行特征选择。

1. CNN with Attention1

在此模型中, 分别对问题和回答使用 CNN 提取特征, 这里借鉴了 Siamese 结构, 使用两个共享变量的 CNN 来分别对问题和答案建模。另外, 在这个模型结构中, 我们使用了两个 Attention 机制, 第一个位于卷积前, 为了影响卷积在问题表示和回答表示中相似的部分; 第二个位于卷积之后, 为了影响池化层在卷积结果中相似的部分。我们实验采用的是两层的卷积层, 每一层卷积的前后都有 Attention 机制。具体的网络结构 (只显示一个卷积层) 如下:

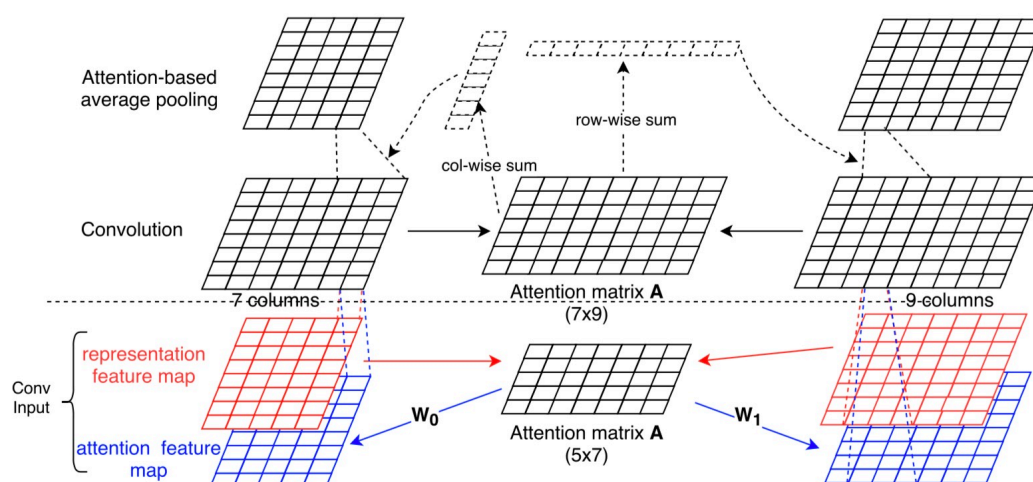


图 1: CNN&Attention

在使用 CNN 提取问题和回答的特征后，我们将其和问题长度、回答长度、问题与回答重复词的个数（除去停用词）、问题与回答重复词的 IDF 四个特征一起用于 SVM 分类，最终得到我们的实验结果。

2. CNN with Attention2

此模型主要由四个部分组成：首先使用 Soft-Attention 来度量问题和答案的关系，得到问题和答案的 Attention 矩阵；然后输出与答案中每个词对应的相关向量；接着是比较层，使用基于元素的矩阵相乘的结果作为比较结果，并将结果作为下一层 CNN 的输入；最后一层全连接层将 CNN 的输出映射为每个答案的得分。具体的网络结构如下：

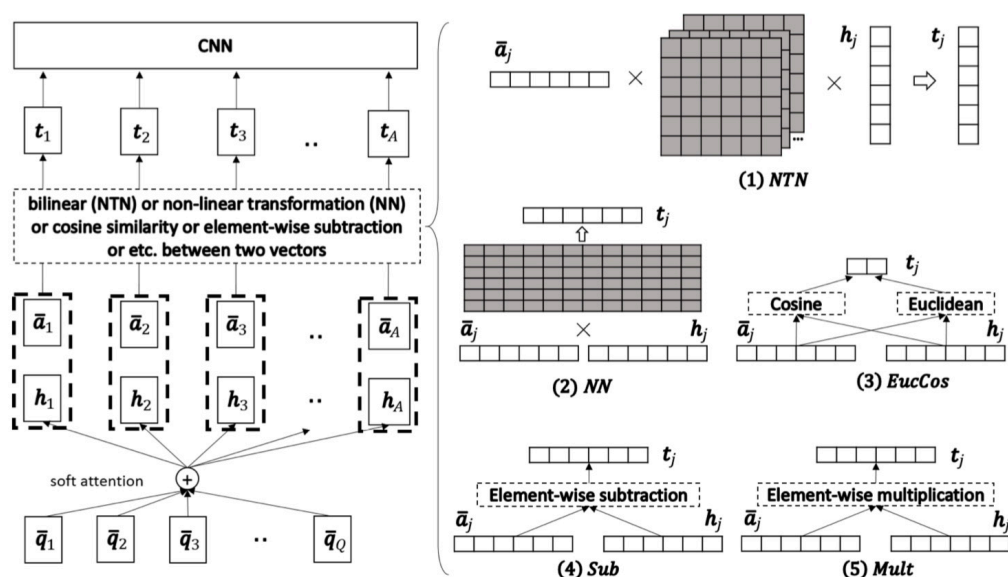


图 2: CNN&Attention2

另外，受到上面的模型的启发，我们认为问题长度、回答长度、问题与回答重复词的个数（除去停用词）、问题与回答重复词的 IDF 这四个特征对于语义匹配分类是比较重要的特征，因此我们仍然将模型提取的特征于这四个合并后输入到 SVM 进行分类。

4 实验数据和环境

· 实验数据

实验提供了训练集、验证集和测试集。每个数据集的格式相同，下面以训练集为例子给出具体的数据格式：

俄罗斯贝加尔湖的面积有多大?	\t	贝加尔湖, 中国古代称为北海, 位于俄罗斯西伯利亚的南部。	\t	0
俄罗斯贝加尔湖的面积有多大?	\t	贝加尔湖是世界上最深, 容量最大的淡水湖。	\t	0
俄罗斯贝加尔湖的面积有多大?	\t	贝加尔湖贝加尔湖是世界上最深和蓄水量最大的淡水湖。	\t	0
俄罗斯贝加尔湖的面积有多大?	\t	它位于布里亚特共和国 (Buryatiya) 和伊尔库茨克州 (Irkutsk) 境内。	\t	0
俄罗斯贝加尔湖的面积有多大?	\t	湖型狭长弯曲, 宛如一弯新月, 所以又有“月亮湖”之称。	\t	0
俄罗斯贝加尔湖的面积有多大?	\t	贝加尔湖长 636 公里, 平均宽 48 公里, 最宽 79.4 公里, 面积 3.15 万平方公里。	\t	1
俄罗斯贝加尔湖的面积有多大?	\t	贝加尔湖湖水澄澈清冽, 且稳定透明 (透明度达 40.8 米), 为世界第二。	\t	0

图 3: 数据格式

在训练集和验证集中, 将提供问题 (第一列), 文档句子 (第二列) 和答案注释 (第三列)。如果文档句子是问题的正确答案, 则其答案注释将为 1, 否则其答案注释将为 0。三列将用符号 '\t' 分隔。在测试集中, 只提供问题及其文档句子。

- 实验环境 (Python3.6)
 - HanLP 预训练的 300 维的词向量以及停用词表。
 - 中文分词工具 jieba。
 - sklearn 中的 linear_model 和 svm 等线性分类器。

- 评测方法

- MRR

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

$|Q|$ 表示问题的总数量, $rank_i$ 表示问题中正确答案的在结果中的排序, 如果该正确答案没有被预测, 则 $\frac{1}{rank_i}$ 为 0。

- MAP

$$MAP = \frac{1}{|Q|} \sum_{i=1}^{|Q|} AveP(C_i, A_i)$$

$AveP(C_i, A_i)$ 表示平均准确率, 表示的是每个问题的平均准确度。

5 实验结果和分析

5.1 实验结果

使用两个模型搭建的网络, 分别执行 20 个 epoch 和 40 个 epoch, 每个计算每个 epoch 中模型对于验证集的 MAP 和 MRR, 具体的实验结果如下:

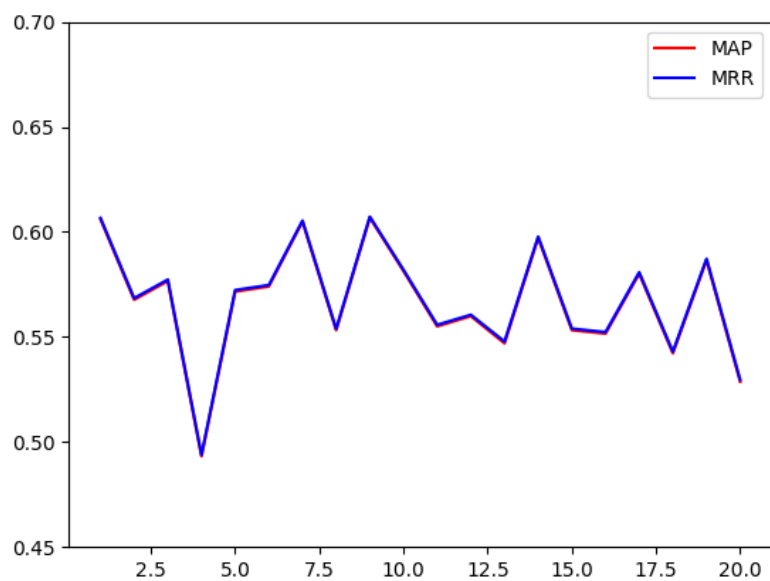


图 4: CNN with Attention1 运行结果

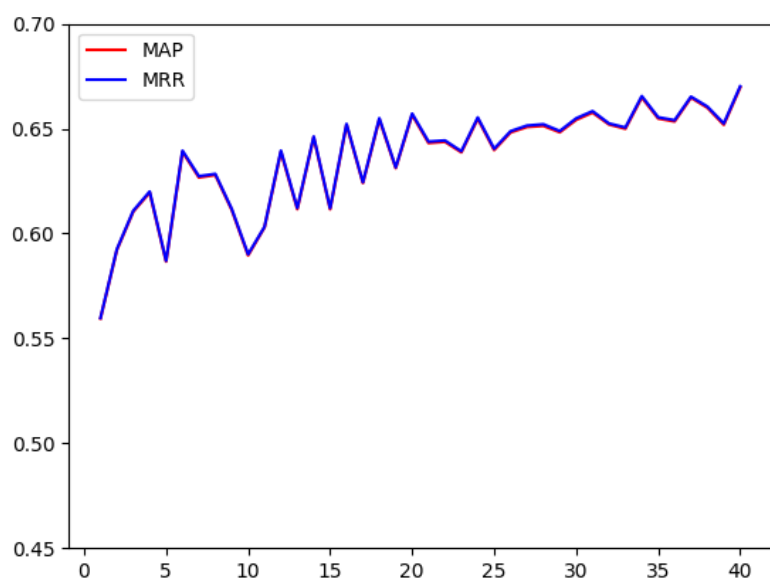


图 5: CNN with Attention2 运行结果

5.2 结果分析

根据结果曲线图我们可以看出，最终模型在验证集上的 MAP 和 MRR 最优可以分别是 66.96%, 67.01%。两个模型在验证集的准确度随着 epoch 的增加而逐渐增高，但是都在 60%, 65% 左右，没有很大的提升。

6 参考文献

- [1] <http://tcci.ccf.org.cn/conference/2017/taskdata.php> .
- [2] <https://github.com/fxsjy/jieba> .
- [3] <https://github.com/hankcs/HanLP> .
- [4] Yin, Wenpeng, et al. " Abcnn: Attention-based convolutional neural network for modeling sentence pairs." arXiv preprint arXiv:1512.05193 (2015).
- [5] Fu, Jian, Xipeng Qiu, and Xuanjing Huang. " Convolutional deep neural networks for document-based question answering." Natural Language Understanding and Intelligent Applications. Springer, Cham, 2016. 790-797.
- [6] Wang, Shuohang, and Jing Jiang. " A compare-aggregate model for matching text sequences." arXiv preprint arXiv:1611.01747 (2016).
- [7] Wang, Ziliang, et al. " Improved Compare-Aggregate Model for Chinese Document-Based Question Answering." National CCF Conference on Natural Language Processing and Chinese Computing. Springer, Cham, 2017.