



文章编号: 1001-9081(2016) 01-0000-10

# 基于深度学习的问答匹配方法

荣光辉, 黄震华\*

(1. 同济大学 计算机科学与技术系, 上海 201800)

(2. 同济大学 计算机科学与技术系, 上海 201800)

(\*通信作者电子邮箱 huangzhenhua@tongji.edu.cn)

**摘要:** 面向中文问答匹配任务, 提出了基于深度学习的问答匹配方法, 以解决机器学习模型因人工构造特征而导致特征不足, 准确率偏低的问题。在该方法中, 主要有三种不同的模型。首先是组合式的循环神经网络(RNN)与卷积神经网络(CNN)模型去学习句子中的深层语义特征, 并计算特征向量的相似度距离。在此模型的基础上, 加入了两种不同的注意力机制, 根据问题构造答案的特征表示, 去学习问答对中细致的语义匹配关系。实验表明, 基于组合式的深度神经网络模型的实验效果要明显优于基于特征构造的机器学习方法, 而基于注意力机制的混合模型可以进一步提高匹配准确率, 其结果最高在平均倒数排序(MRR)和 Top-1 accuracy 评测指标上分别可以达到 80.05% 和 68.73%。

**关键词:** 问答匹配; 深度学习; 循环神经网络; 卷积神经网络; 注意力机制; 机器学习

**中图分类号:** TP183 神经网络与计算

**文献标志码:** A

## Question answer matching method based on deep learning

RONG Guanghui, HUANG Zhenhua\*

(1. Department of Computer Science and Technology, Tongji University, Shanghai, China 201800;

2. Department of Computer Science and Technology, Tongji University, Shanghai, China 201800)

**Abstract:** Aimed at Chinese question answer matching task, a method of question answer matching based on deep learning is proposed, to solve the problem of lack of features and low accuracy that due to artificial structural feature in machine learning. This method mainly includes 3 different models. The first model is the combination of Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) to learn the deep semantic features in the sentence, and calculate the similarity distance of feature vectors. Moreover adding two different attention mechanism on this model, the feature representation of answer is constructed according to the question to learn the detailed semantic matching relation of them. Experiments show that the results of combined deep neural network model is superior to the method of feature construction based on machine learning, and the hybrid model based on attention mechanism can further improve the matching accuracy where the best results can reach 80.05% and 68.73% in the standard evaluation of mean reciprocal rank (MRR) and Top-1 accuracy respectively.

**Key words:** question answer matching; deep learning; Recurrent Neural Network (RNN); Convolution Neural Network (CNN); attention mechanism; machine learning

## 0 引言

随着互联网技术的快速发展, 大量的文本数据不断产生, 人们能够通过搜索引擎检索到丰富的信息。然而传统的搜索引擎仅仅返回与结果相关的网页, 导致难以让用户获取准确

收稿日期: 2016-00-00; 修回日期: 2016-00-00。

基金项目: 中央高校基本科研业务费专项资金; 上海市青年科技启明星计划 (15QA1403900); 上海市自然科学基金 (17ZR1445900); 霍英东教育基金会高等院校青年教师基金 (142002)。

作者简介: 荣光辉 (1992-), 男, 安徽六安人, 硕士研究生, 主要研究方向: 深度学习、自然语言处理。黄震华 (1980-), 男, 福建泉州人, 教授, 博士, CCF 会员, 主要研究方向: 数据分析、数据挖掘、机器学习。



的关键信息。相比之下,智能问答系统能够以一问一答的自然语言形式,通过与用户进行交互,精确的定位用户意图,从而能够高质量的快速满足用户的信息需求,降低网站运行成本等重大实际问题。目前,智能问答系统已经是自然语言处理领域中一个非常重要的研究热点<sup>[1-2]</sup>,具有重大的应用价值。

对于问答匹配任务,早期的方法主要是基于机器学习模型,其需要依赖人工构造的规则特征。这些基于统计机器学习的模型通常会定义包括词法,句法,语法等相关特征(例如 BM25, Edit Distance, TFIDF 等),并借助机器学习分类器(例如 KNN, 贝叶斯分类, SVM 等)来得到问答对之间的匹配关系。这种基于特征构造的方法是一种浅层学习模型,其结果往往依赖特征提取的质量高低,对于不同领域的数据泛化能力较差,缺乏对数据深层语义信息的学习能力。

近年来,深度学习模型在图像处理,语音识别等领域的相关任务上取得了优异的表现,并逐渐在自然语言处理任务上取得了一些不错的研究成果。深度学习主要是以深度神经网络为主,主要包括卷积神经网络和循环神经网络。在 2014 年,Johnson<sup>[3]</sup>等人提出基于卷积神经网络的短文本分类模型,将文本分类的准确率提高到了新的层次。Zhang<sup>[4]</sup>等人应用卷积深度信念网络来学习词汇和句子两个层面的潜在特征。在机器翻译上,Le<sup>[5]</sup>等人利用两个循环神经网络结构分别作为编码器和解码器,构造出从序列到序列(Sequence to Sequence, Seq2Seq)的语言翻译框架,提升了机器翻译模型的质量。因此,应用深度学习模型可以对问答对中细粒度的深层语义匹配关系进行有效的表示学习建模。

由于深度学习模型在自然语言任务中具有强大的表示学习能力<sup>[6]</sup>,为了能够学习到问答对中细致的语义匹配关系,在不依赖人工构造特征的情况下,本文改进了一般的深度神经网络模型,提出了 3 种不同的实验模型,使其更加适合于问答匹配任务。通过基于深度神经网络的模型,并结合注意力机制的方法,将文本信息向量化,可以学习到句子中更加抽象的语义特征以及问答对之间关键的语义匹配信息,从而选择出正确的候选答案。

## 1 相关工作

问答匹配任务的目标是给定一个问题和候选答案集,从中选择出语义关系最为匹配的答案,其关键在于让相关模型能够学习到文本中深层的语义特征和细粒度的匹配关系。以前的相关模型多是基于机器学习的浅层模型,例如 IBM Watson 部门的 Gondek 等人使用了答案融合和特征排序的方法<sup>[7]</sup>,Wang 等人<sup>[8]</sup>提出了基于关系主题空间特征的多核 SVM 分类模型。为了得到更丰富的特征,这些相关模型中会引入如语义知识库(Freebase)和语义词典(WordNet)等外部资源信息<sup>[9-10]</sup>,以此来提高问答句之间的语义匹配精度。然而这些模

型都缺少强大的表示学习能力且往往需要依赖大量的特征工程。

近年来,深度学习模型逐渐开始应用到问答匹配任务中,Nyberg 等人提出了基于多层长短期记忆网络(Long Short-Term Memory,LSTM)模型训练问答对的联合特征向量<sup>[11]</sup>,把问答匹配问题转换为分类或排序学习问题。Feng<sup>[12]</sup>等人提出了基于共享卷积神经网络来训练问答对,并在此基础上进行语义相似度计算,该模型在 InsuranceQA 英文数据集上取得了突破性的实验效果。另外,基于注意力机制的模型方法在一系列自然语言处理相关任务中也取得了显著的效果提升,其主要是仿照人脑的注意思维,通过自动加权的方式对整体信息进行局部重点关注。因此在问答匹配任务上可以尝试使问题更加关注到答案中某些语义特性与其更加相关的词语,重点将这些关键的词语参与到答案的特征表示计算。例如在文本蕴含领域上,Tim 等人使用一种 word-by-word attention<sup>[13]</sup>方法,根据假设自适应地学习前提的表示,能够更好地判断假设与前提的蕴含关系。Yin 等人提出一种基于注意力机制的多层卷积神经网络模型<sup>[14]</sup>来对两个文本语义表示建模,并在答案匹配,释义识别和文本蕴含任务上都达到了很好的实验效果。注意力机制主要分为 Soft Attention 与 Hard Attention 两种形式。其中 Soft Attention 是指对每一个特征都分配一个注意力权值,认为是个概率分布。而 Hard Attention 在选取特征组合时,并不是对所有的特征都生成权值,可能只选取一个或者多个。本文所提出的注意力模型是将门限循环单元(Gated Recurrent Unit,GRU)<sup>[15]</sup>与 Soft Attention 机制相结合的混合学习模型。

## 2 基于深度学习的问答匹配模型

### 2.1 基于循环卷积神经网络的问答匹配模型

循环神经网络被广泛应用于处理可变长的文本序列输入,因为可以学习到句子的词序特征,其关键的结构是一个记忆单元(Memory Unit)。记忆单元可以记忆某时间段的信息,对于一个句子则可以选择性记住前一时刻词语的信息。列如一个问句:“腾讯在线教育由哪几个部分组成?”,其中“在线教育”这个词所对应的隐藏单元输入可以更多地保留来自词语“腾讯”所传递的上文信息。当前循环神经网络主要有两种不同的变体 LSTM 和 GRU,它们都能解决传统 RNN 存在的长期距离依赖与梯度消失问题。相比 LSTM 的网络结构,GRU 内部的隐藏单元少一个控制门,参数少,收敛较快,在保证模型效果的同时,模型的结构也得到了有效简化,所以本文使用 GRU 深度层次神经网络来代替 LSTM。给定句子输入  $x = \{x_1, x_2, \dots, x_T\}$ ,  $x_i \in R^d$ , 其中  $x_i$  表示  $t$  时刻句子的预训练的词向量,前一个时刻的隐藏层输出表示为  $h_{t-1}$ , 则  $t$  时刻的 GRU 内部记忆单元状态如下。



遗忘门:  $f_t = \sigma(W_{xf}x_t + U_{hf}h_{t-1})$

更新门:  $z_t = \sigma(W_{xz}x_t + U_{hz}h_{t-1})$

$t$  时刻内部状态:  $\tilde{h}_t = \tanh(W_{xh}x_t + U_{hh}(f_t \odot h_{t-1}))$

$t$  时刻输出:  $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$

其中  $\sigma$  是 *sigmoid* 激活函数,  $\odot$  代表逐元素 (element-wise)

相乘,  $W$  与  $U$  为参数矩阵。由于单向 GRU 在  $t$  时刻无法学习

到在它之前和之后的上下文语义信息, 所以本文模型中使用

双向门限循环单元 (Bidirectional Gated Recurrent Unit, BGRU), 即将输入序列从正向和反向两个 GRU 层分别输入,

把两个方向在  $t$  时刻的隐层输出串接为一个新向量  $h_t \in R^H$ 。

相对于单向的 GRU 层, BGRU 模型能够使一个词向量表示

包含更丰富的信息, 从而构造出更好的句子特征表示, 其模型

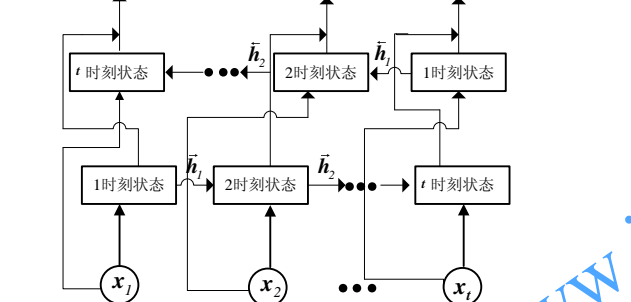


图1 双向门限循环单元神经网络结构

Fig. 1 The BGRU neural network structure

为了能够学习到句子中更深层次的潜在局部特征, 本文将 BGRU 层的输出接入到一个卷积层。卷积层的操作就是将句子中连续的词向量串接为一个向量, 再把这个向量通过某种函数映射为一个新的特征向量。模型中设置卷积窗口大小为  $k$ , 拼接相邻的  $k$  个向量得到矩阵  $S \in R^{k|H| \times T}$ , 卷积核 *filter* 的数目为  $c$ , 整个卷积操作计算过程如公式 (1) 所示。

$$G = f(W_{gc}S + b) \quad (1)$$

其中  $f$  是 *relu* 激活函数, 可以加速模型的收敛, 更易于优化, 权重  $w_{gc}$  和偏差  $b$  随机初始化为均匀分布  $U(-0.05, +0.05)$ 。

不同于传统单一粒度的卷积核, 在实验中本文使用多粒度的卷积核, 即有不同大小的卷积核, 每一种粒度的卷积核都可以学习到句子的一种  $n$ -gram 特征。卷积操作后经 *max-pooling* 得到固定大小的句子特征向量, 串接所有的向量得到句子的最终表示。另外为了防止模型出现过拟合现象, 训练过程中使用了 *Dropout*<sup>[16]</sup>, 整个 BGRU-CNN 模型结构如图 2 所示。

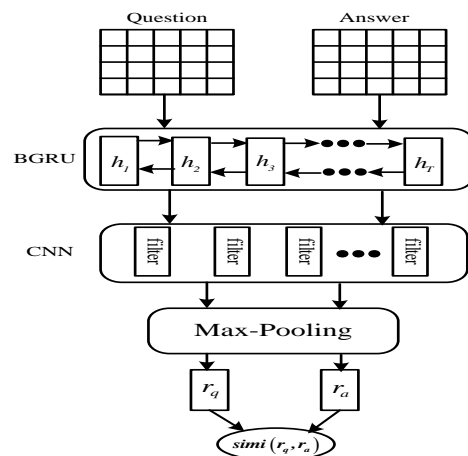


图2 基于循环卷积神经网络的问答匹配

Fig. 2 Question answer matching based the BGRU-CNN model

## 2.2 基于注意力机制的问答匹配混合模型

在 BGRU-CNN 模型中, 对于答案的每个词向量其实都是同等对待。问题和答案在特征学习的过程中并没有任何语义上的交互影响, 也没有考虑到一些具有关键特征的词语在答案构造表示中对句子语义匹配的重要性。另外一般句子中都存在很多无关的干扰词, 它们对语义匹配的贡献度很小, 甚至会降低匹配的准确度。而基于注意力机制的混合深度学习模型能够使问题关注到答案语句中更为重要的特征词语, 让这些关键词语更多地参与答案的特征表示计算中。

*Soft Attention* 机制就是对句子的局部重要特征进行重点关注, 在模型中对每个词向量特征设置不同大小的注意力权重系数, 不同的权重反映出语义重要性的大小, 如以下的的一个问答匹配对。

Q: 你知道在上海, 有哪些非常好玩的著名景点?

A: 上海是中国最大的经济城市, 这里有交通发达, 是国际化的大城市, 旅游景点也很发达, 主要有东方明珠, 外滩, 南京路步行街等游玩景点。

根据这个问题去匹配候选答案的时候, 则“东方明珠”, “外滩”, “南京路步行街”这些词语的语义与问题表述更为相关, 所以在模型中应当增强这些词语的特征表示作用。为了能够重点关注到答案句子中与问题语义最相关的有效词语, 本文提出的第一个基于注意力机制的问答匹配混合模型叫做 AR-BGRU, 其主要借鉴 Hermann<sup>[17]</sup>等人在阅读理解相关任务中所提出的一种注意力机制方法称为 *Attentive Reader*, 相比 Tan<sup>[18]</sup>等人提出的 *Attention-LSTM* 模型, 本文基于 BGRU 网络的注意力模型在复杂度上和训练代价上都较为有优势, AR-BGRU 模型的网络结构如图 3 所示。

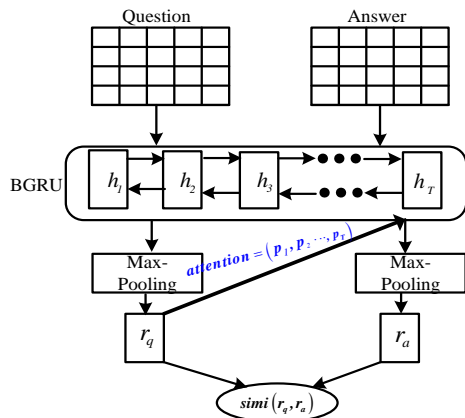


图 3 基于 AR-BGRU 注意力模型的问答匹配

Fig. 3 Question answer matching based the AR-BGRU attention model

答案中每一个词向量的所对应的注意力权值是根据问题所决定。从 BGRU 层输出的问题词向量矩阵经 max-pooling 得到问题特征向量  $r_q$ ，答案句子  $t$  时刻的输出  $h_t$  所对应的权值计算方式如公式 (2) 所示。

$$P_{a,(t)} \propto \exp(v^T \tanh(W_s h_{(t)} + U_s r_q)) \quad (2)$$

$$h_{a,(t)} = P_{a,(t)} h_{(t)}$$

$P_{a,(t)}$  就是  $h_{(t)}$  对应的注意力权值，通过  $r_q$  对  $h_{(t)}$  加入注意力使其更新为新的特征向量  $h_{a,(t)}$ 。 $h_{(t)}$  与  $r_q$  的在语义匹配层面越相关， $P_{a,(t)}$  所表示的重要性程度系数也就越大。这种方式可以突出答案中那些与问题相关的词语，使具有关键特征的词语在答案表示中发挥着更大的作用。

AR-BGRU 模型训练注意力重要性系数的方式存在着参数较多，结构更加复杂化的问题，导致模型训练的代价增大。而且 max-pooling 往往选取最重要的特征，从而会丢失一些其他较为重要的信息。因此本文提出另外一种基于注意力机制的混合模型叫做 AP-BRGU，不是简单使用 max-pooling 去得到最终的答案表示，而是依据不同的注意力权值系数通过加权求和的方式去表示答案。与 AR-BGRU 相比，模型的结构得到简化，实验的效果相比较也有所提升，可以更好地在语义层面上匹配到正确的候选答案，AP-BGRU 网络结构如图 4 所示。

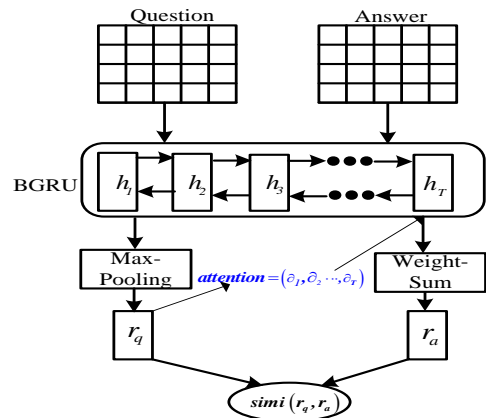


图 4 基于 AP-BGRU 注意力模型的问答匹配

Fig. 4 Question answer matching based the AP-BGRU attention model

从 BGRU 隐藏层输出得到答案词向量矩阵  $A=(a^1, a^2, \dots, a^T)$ ，根据问题向量  $r_q$  计算  $t$  时刻答案词向量  $a^t$  的注意力权值系数  $\delta_t$ ，其计算方式如公式 (3) 所示。

$$\delta_t = \frac{\exp(P_t)}{\sum_{t=1}^T \exp(P_t)} \quad (3)$$

其中  $P_t$  表示  $a^t$  对问题  $r_q$  的重要性程度函数，计算方式如公式 (4) 所示：

$$P_t = (r_q^T W_{qa} a^t) \quad (4)$$

其中  $W_{qa}$  是参数矩阵，所以答案的特征表示为每个更新后的词向量做加权求和运算，其计算方式公式 (5) 所示。

$$r_a = \sum_{t=1}^T \delta_t a^t \quad (5)$$

经过给不同的词语设置不同的权值系数，使得与问题无关的词向量  $a^t$  在最终的答案特征表示中发挥较小的作用。

### 3 实验与分析

#### 3.1 词向量

深度学习在自然语言处理任务中普遍应用基于分布式的  $K$  维稠密实数向量来表示词语的特征，通过词语在向量空间上的距离来判断它们之间的语义相似度。相比于 One-hot 语言模型，这种方法不会出现维数灾难。Google 在 2013 年发布强大开源工具 word2vec<sup>[19]</sup>就是一种分布式的词表征方式。本文使用 word2vec 来预训练词向量，训练语料主要来自爬取的百度百科相关网页数据，去除特殊字符和标点符号后，语料库大小大约有 22GB，所训练的词向量维数  $d$  设置为 300，句子最大的长度设置为 50。

#### 3.2 实验设置





模型的目标函数定义为：

$$L(q, a^+, a^-) = \max\{0, m - \cosine(q, a^+) + \cosine(q, a^-)\}$$

其中  $a^+$  为正确答案,  $a^-$  为随机选取的负向答案,  $m$  是为了优化目标函数而设置的阈值参数, 目标函数的意义就是让正确答案和问题之间的  $\cosine$  值要大于负向答案和问题的  $\cosine$  值。模型的优化器设置为 Adam<sup>[20]</sup> 来更新神经网络模型中的参数, Dropout rate 设置为 0.2, 卷积核数  $c$  设置为 200。实验环境是基于 Spark 集群的 CentOS 6.5 服务器, 每个 CPU 有 16 个核, 4 块型号为 TITAN X 的显卡, 每个显存大小为 12GB, 支持 Nvidia Cuda 深度学习运算平台。

### 3.3 数据集

数据集为 NLPCC-ICCPOL 2016 评测任务 DBQA 所提供的开放域问答数据集, 数据集分为训练集与测试集。训练集有 8772 个问题, 共 181882 问答对, 测试集有 5997 个问题, 共 122530 个问答对。本文在预处理阶段去除了一些常见无意义的停用词, 如“啊”, “吧”, “的”等。

### 3.4 实验结果

本文实验采用了两种评测指标: 准确率 (Top-1 Accuracy) 和 MRR, 其中 MRR 的实验结果如表 1 所示。

表 1 本文多个模型的实验评测结果

Tab.1 The proposed multi-models experimental evaluation results

方法	MRR(%)
BGRU-CNN	75.32
AR-BGRU	77.67
AP-BGRU	80.05

以及本文所提出的各个模型在 Top-1 accuracy 上的实验结果如图 5 所示:

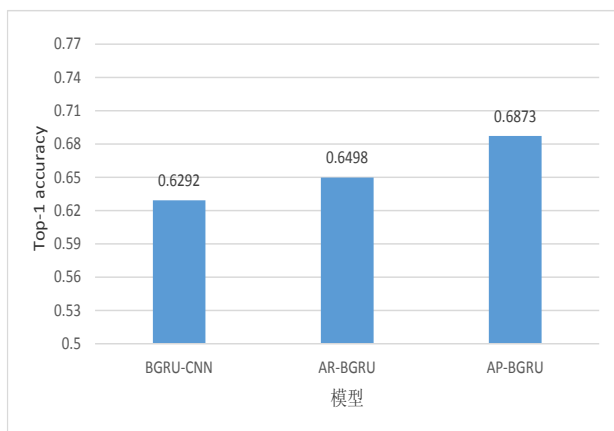


图 5 实验评测标准 Top-1 的准确率

Fig. 5 The Top-1 Accuracy experimental results

实验数据集主要是 NLPCC-ICCPOL 2016 DBQA Task 所提供的中文开放域问答数据, 为了验证本文模型的有效性, 在相同数据集下将其他模型的实验对比结果如表 2 所示。

表 2 与其他相关模型的实验对比结果

Tab.2 The experimental evaluation results compared with other relevant methods

来源	方法	MRR(%)
Wu <sup>[21]</sup>	TFISF sum	45.31
	Re-estimate query likelihood model	69.36
	Edit distance	20.98
	BLSTM	66.34
Wang <sup>[22]</sup>	Word overlap	51.54
	Average word embedding	46.10
	Embedding-based features	74.70
	All features	80.08
本文模型	AR-BGRU	77.67
	AP-BGRU	80.05

从表 2 中的实验结果对比可以看出, 就 Wu 等基于传统的人工构造特征 (如: TF, Edit distance 等) 而言, 基于深度学习的双向长短期记忆网络 (BLSTM) 模型在问答匹配上的实验效果要明显优于基于机器学习的 SVM 分类排序方法。而本文 BGRU-CNN 模型与 BLSTM 相比, MRR 指标提高了将近 9.0% 左右, 可以推断出在双向门限循环神经网络加卷积层能够进一步学习到句子中更深层次的语义特征, 从而可以提高匹配的准确率。在与 Wang 等模型的实验对比结果中可以看出, BGRU-CNN 实验效果要比基于 Word-overlap 和 Average word embedding 特征构造的模型要好很多。这两个方法存在的共同点就是无法像深度神经网络那样可以学习到句子中抽象的语义信息。相比 Embedding-based feature 方法, BGRU-CNN 模型的效果也有一定的提升, MRR 指标大约提高了 0.7% 左右。从表 1 和表 2 可以看出, 引入注意力机制的 AR-BGRU 与 AP-BGRU 模型的实验效果都要优于 BGRU-CNN 模型, MRR 指标分别提高了 2.35% 和 4.73%, 从而可以说明基于注意力机制的深度学习混合模型在问答匹配任务上的有效性, 它们都可以增强答案特征中与问题最相关的表示部分。另外可以发现 AP-BGRU 比 AR-BGRU 的匹配效果提高了 2.4% 左右, 间接地说明基于注意力机制的加权特征构造方式与 max-pooling 相比, 可以更好地表示答案。与 Wang 等组合多种构造特征 (All feature) 并通过回归或者排序学习模型去选取候选答案的方法相比, 本文 AP-BGRU 的效果要稍微差一点, 推测这是整个训练样本受限的原因。

## 4 结语

针对问答匹配任务, 本文提出了三种基于深度学习的模型去改善机器学习方法中存在的不足, 实验结果验证了本文



模型在中文开放域问答数据集上的有效性。在不依赖人工构造复杂特征的情况下, 本文的模型取得了良好的实验效果, 尤其是在引入注意力机制后, 模型的准确率能够得到进一步的提升。目前中文问答系统的语料相对来说还是欠缺, 除了收集和标记更多的语料之外, 还可以尝试如何在深度学习模型中有效地融合传统的特征提取方法。另外目前基于注意力机制的模型大多数都是在词汇级别 (word-level) 上, 而短语级别 (phrase-level) 的文本可能包含更丰富的语义信息, 如何将注意力机制应用到短语级别, 更加有效地学习句子的潜在抽象特征, 将是未来的重点研究方向。

#### 参考文献

- [1] 王元卓, 贾岩涛, 刘大伟, 等. 基于开放网络知识的信息检索与数据挖掘[J]. 计算机研究与发展, 2015, 52(2):456-474.( Wang Y, Jia Y, Liu D W, et al. Open web knowledge aided information search and data mining[J]. Journal of Computer Research & Development, 2015, 52(2):456-474.)
- [2] ZHOU T C, LYU M R, KING I. A classification-based approach to question routing in community question answering[C]//Proceedings of the 21st International Conference on World Wide Web. ACM, 2012: 783-790...
- [3] JOHNSON R, Zhang T. Effective use of word order for text categorization with convolutional neural networks[J]. arXiv preprint arXiv:1412.1058, 2014.
- [4] ZHANG D, WANG D. Relation classification via recurrent neural network[J]. arXiv preprint arXiv:1508.01006, 2015..
- [5] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 2014: 3104-3112..
- [6] HU B, LU Z, LI H, et al. Convolutional neural network architectures for matching natural language sentences[C]//Advances in neural information processing systems. 2014: 2042-2050.
- [7] GONDEK D C, LALLY A, KALYANPUR A, et al. A framework for merging and ranking of answers in DeepQA[J]. IBM Journal of Research and Development, 2012, 56(3.4): 14: 1-14: 12.
- [8] WANG C, KALYANPUR A, FAN J, et al. Relation extraction and scoring in DeepQA[J]. IBM Journal of Research and Development, 2012, 56(3.4): 9: 1-9: 12..
- [9] KASNECI G, SUCHANEK F M, IFRIM G, et al. NAGA: harvesting, searching and ranking knowledge[C]//Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008: 1285-1288..
- [10] YIH W T, CHANG M W, MEEK C, et al. Question answering using enhanced lexical semantic models[C]// Meeting of the Association for Computational Linguistics. 2013:1744-1753.
- [11] WANG D, NYBERG E. A long short-term memory model for answer sentence selection in question answering[C]// Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015:707-712.
- [12] FENG M, XIANG B, GLASS M R, et al. Applying deep learning to answer selection: A study and an open task[C]//Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop on. IEEE, 2015: 813-820..
- [13] ROCKTASCH T, GREFFENSTETTE E, HERMANN K M, et al. Reasoning about entailment with neural attention[J]. arXiv preprint arXiv:1509.06664, 2015..
- [14] YIN W, SCHUTZE H, XIANG B, et al. Abcnn: Attention-based convolutional neural network for modeling sentence pairs[J]. arXiv preprint arXiv:1512.05193, 2015..
- [15] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014..
- [16] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958..
- [17] HERMANN K M, KOCISKY T, GREFFENSTETTE E, et al. Teaching machines to read and comprehend[C]//Advances in Neural Information Processing Systems. 2015: 1693-1701.
- [18] TAN M, SANTOS C, XIANG B, et al. Lstm-based deep learning models for non-factoid answer selection[J]. arXiv preprint arXiv:1511.04108, 2015.
- [19] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [20] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [21] WU F, YANG M, ZHAO T, et al. A hybrid approach to DBQA[C]//International Conference on Computer Processing of Oriental Languages. Springer International Publishing, 2016: 926-933.
- [22] WANG B, NIU J, MA L, et al. A chinese question answering approach integrating count-based and embedding-based features[C]//International Conference on Computer Processing of Oriental Languages. Springer International Publishing, 2016: 934-941.

中央高校基本科研业务费专项资金; 上海市青年科技启明星计划(15QA1403900); 上海市自然科学基金(17ZR1445900); 霍英东教育基金会高等院校青年教师基金(142002)。

荣光辉(1992-), 男, 安徽六安人, 硕士研究生, 主要研究方向: 深度学习、自然语言处理;

黄震华(1980-), 男, 福建泉州人, 教授, 博士, CCF 会员, 主要研究方向: 数据分析、数据挖掘、机器学习。

This work is partially supported by the Fundamental Research Funds for the Central Universities of Ministry of Education of China, Shanghai Youth Science and Technology Star Project (15QA1403900), Shanghai Science and Technology Committee Rising-Star Program(17ZR1445900), Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China (142002).

**RONG Guanghui**, born in 1992, M. S. candidate. His research interests include deep learning, natural language processing;  
**HUANG Zhenhua**, born in 1980, Ph. D., professor, CCF member. His research interests include data analysis, data mining, machine learning.



[www.joca.cn](http://www.joca.cn)