

基于 Bi-LSTM 和 CNN 并包含注意力机制的社区问答问句分类方法^①

史梦飞¹, 杨 燕¹, 贺 樑¹, 陈成才²

¹(华东师范大学 计算机科学与软件工程学院, 上海 200062)

²(上海智臻智能网络科技股份有限公司, 上海 201803)

通讯作者: 杨 燕, E-mail: yanyang@cs.ecnu.edu.cn

摘 要: 问句分类的目标是将用户提出的自然语言问句分到预先设定的类别。在社区问答中, 如何准确高效的对问句进行分类是一项重要任务。本文提出了一种基于深度神经网络的问句分类方法, 该方法首先将问句用词向量进行表示, 然后用融合双向长短期记忆网络 (Bi-LSTM) 和卷积神经网络 (CNN) 结构并包含注意力机制的深度学习模型提取问句特征进行分类。该方法的特色在于利用 Bi-LSTM 和 CNN 在句子级文本表示的优点, 充分捕捉问句特征, 并结合问句的对应答案来表示问句, 丰富了问句信息。实验表明, 该问句分类方法准确率较高, 在多个数据集上取得不错结果。

关键词: 问句分类; 答案集; 注意力机制; 神经网络

引用格式: 史梦飞, 杨燕, 贺樑, 陈成才. 基于 Bi-LSTM 和 CNN 并包含注意力机制的社区问答问句分类方法. 计算机系统应用, 2018, 27(9): 157-162. <http://www.c-s-a.org.cn/1003-3254/6536.html>

Question Categorization of Community Question Answering by Combining Bi-LSTM and CNN with Attention Mechanism

SHI Meng-Fei¹, YANG Yan¹, HE Liang¹, CHEN Cheng-Cai²

¹(School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China)

²(Xiaoi Robot Technology Co. Ltd., Shanghai 201803, China)

Abstract: The goal of question categorization is to classify natural language questions that user raised into predefined categories. How to classify question sentences accurately and efficiently is an important task in community question answering. In this study, we propose a question categorization method based on deep neural network. Firstly, the words of the question are transformed to vectors. Then, we use a novel Bidirectional Long Short-Term Memory (Bi-LSTM) based Convolutional Neural Network (CNN) model with attention mechanism to capture the most important features in a question. Finally, the features are fed into the classifier to predict the category of the question. We use the Bi-LSTM and CNN to capture the features of question because of their benefits in representing sentence level documents. We also use the answer set to enrich the information of the question. The experimental results on several datasets demonstrate the effectiveness of the proposed approach.

Key words: question classification; answer set; attention mechanism; deep neural network

① 基金项目: 上海市经济和信息化委员会项目 (201602024); 上海市科学技术委员会项目 (14DZ2260800)

Foundation item: Project of Economic and Information Committee of Shanghai Municipality (201602024); Project of Municipal Science and Technology Committee of Shanghai Municipality (14DZ2260800)

收稿时间: 2018-01-20; 修改时间: 2018-02-09; 采用时间: 2018-03-06; csa 在线出版时间: 2018-08-16

社区问答系统 (Community Question Answering, CQA) 通过提供问题和答案的形式将需要获取有效信息的用户和热心分享知识信息的用户联系起来, 用户既能提出新的问题, 也能回答问题或者丰富已有的问答信息, 这是一种高效便捷的信息获取的方式。目前比较有名的 CQA, 像百度知道、360 问答、Yahoo Answers 等, 都有着非常庞大的用户基数。CQA 以其灵活独特的用户交互特性满足了用户获取和分享知识信息的诉求, 从而受到越来越多的关注。问句分类是其中的一个重要步骤, 可以辅助系统理解用户的询问意图。例如, 对于问句“梅西效力于哪个俱乐部?”应该被分到“体育运动”这个类别。问句的类别信息可以极大的缩小答案搜索空间。除了在问答的过程中, 问句分类在构建社区问答语料库时也发挥着比较重要的作用。准确高效的对问句进行分类将极大提高社区问答语料库的构建速度和质量。

在对问句分类之前, 首先要确定问句有哪几种类型, 具体的类别将由问句分类体系决定。目前的问句分类体系还没有完全统一的标准, 在国际上比较权威的是 UIUC 问句分类体系^[1], 这是基于答案类型的层次分类体系。UIUC 分类体系是针对英文分类的, 哈工大在该分类体系基础上根据汉语的固有特点定义了一套中文分类体系, 主要包含 7 个大类 (人物, 地点, 数字, 时间, 实体, 描述, 未知), 每个大类下又细分了一些小类, 总共 60 个小类^[2]。在一些具体的应用场景会有更有针对性的细致分类, 比如说在社区问答系统“百度知道”中, 问题就被分为“电脑网络”, “体育运动”, “经济金融”等多个类别。

问句分类属于文本分类人任务中的一种, 但与一般的文本分类有很大的不同, 主要原因是问句都是由用户随机提出的自然语言问题, 而不是传统的规范性文本。该任务主要面临两大挑战。第一个挑战是用户提出的问句通常比较简短, 包含的词汇量有限, 存在信息量缺乏的问题^[3,4]。例如, 对于问句“007 是什么?”应该被分到实体这个类别, 但是如何判断“007”是代表的一个电影系列还是一串数字是一个难点。第二个挑战是有些问句比较冗长, 很难捕捉最关键有效的信息。例如对于问句“用西班牙内战作为训练的 6500 名德国空军士兵的名字是什么?”, 包含了一系列的实体名词, 给问句的正确分类造成干扰。所以, 无论问句包含的信息是多是少, 都存在各自分类的困难因素。因此, 如何更好的

利用已有信息或补充信息来理解问句, 找出问句中最有效的信息是问句分类任务急需解决的重要问题。

针对以上的问题, 本文提出了一种基于深度学习的分类方法, 该方法融合双向长短时记忆网络 (Bi-LSTM) 和卷积神经网络 (CNN) 并带有注意力机制。

一方面, 本文提出的方法引入了深度学习模型, 并结合了 Bi-LSTM 和 CNN 两种神经网络结构, 充分利用其各自的优势, 通过学习分布式词向量来表示每个词的特征, 在保留问句时序信息的同时捕捉最主要的信息特征。通过这样的方式有效的缓解了传统词袋模型存在的数据稀疏性和语义敏感性问题。同时, 缓解了传统的基于 CNN 的分类方法只是对由连接词向量组成的 n-gram 向量进行了简单的线性计算, 然而与连接一起的线性操作不能很好的对 n-grams 中的非连续性依赖和交互性进行建模的问题。例如, 在一个问句中包含“not a total loss”这个短语, 显然, 非连续性的依赖“not loss”在这句话中是非常关键的信息, 但是通过简单连接的线性操作很难精确的定位这些信息。在社区问答的问句分类任务中, 由于问句的内容和类型都十分的多样和广泛, 因此结合双向长短时记忆网络和卷积神经网络更能精确捕捉问句特征, 为准确分类提供重要信息。

另一方面, 本文提出的方法考虑到在问句分类时结合问句已有的答案信息, 较好的解决了简短问句信息量少的难点。如表一所示, 在没有利用答案文本的情况下, 仅利用问题文本学习到的词向量作为特征表示, “007 是什么?”很难判断其属于哪个类别。显然这个问句实在太短, 没有包含特别明显的可作为主题判断的信息词, 最关键的词“007”, 机器更可能理解为单纯的数字。然而加入了答案文本后, 由于答案文本中出现“电影”这类词, 这样问题特征所学得的词向量中也就包含“电影”等信息。因此上述问题就可以被准确的识别为“实体类”问题。

具体来说, 本文提出的基于深度学习的问句分类方法使用 Bi-LSTM 和 CNN 结合的方式来更好的加强问句特征表示, 很好的利用了 Bi-LSTM 能够捕捉时序信息的特点及 CNN 捕捉局部特征的优势, 并利用注意力机制, 引入问句的答案内容来增强问句信息量。实验表明, 本文提出的问句分类方法有助于提升问句分类的准确度。

本文的其他部分组织如下: 第 1 节介绍问题分类

的一些相关研究工作;第2节描述本文提出的基于深度学习的问题分类方法;第3节介绍本文的数据集及实验设置与结果;第4节简述结论及未来工作展望。

1 相关工作

目前,问题分类研究主要分为两个大的方向.第一个大方向是基于特征统计的机器学习分类方法,也是目前用的相对较多的方法.第二种是基于深度学习的方法,主要用到各种神经网络模型。

Hui等人^[5]在进行问句分类时考虑到问题文本中词序和词间距的因素,提出了一种扩展类规则模型;Mishra等人^[6]根据从问题文本中抽取的词特征、语义特征和句法特征来训练不同的分类器(朴素贝叶斯、最近邻、支持向量机)进行问题的分类;Aikawa等人^[7]根据用户的主观和客观臆想,将问题分为主客观两类并利用平滑的朴素贝叶斯方法进行问题分类.Liu等人^[8]在SVM的基础上提出了一种依赖句法关系和词性特征的核函数方法.杨思春等人^[9]为了解决问句分类研究中特征提取开销过大的问题,提出了一种包含基本特征和词袋绑定特征的问句特征模型,以此来获取更加有效的问句特征集。

目前,神经网络在自然语言处理领域已经得到广泛应用^[10-12],在图像处理、语音识别和文本分类等任务上都取得了不错的效果.在文本分类领域中, Kim^[13]提出的卷积神经网络分类模型,该方法使用了由word2vec预先训练好的一批词向量,并适当调整了CNN的一些超参,在包含问句集在内的多个英文文本分类数据集上都取得了不错的效果. Shi等人^[14]提出了基于深度长短时记忆网络的非线性不连续特征映射分类模型。

与上述研究内容有所区别的是,本文使用了Bi-LSTM和CNN结合的深度学习模型,并考虑到利用问句的答案信息来增强问句表示,以此提高问句分类的效果。

2 问句分类方法

图1为本文所提出的基于深度神经网络的问句分类方法的架构图.首先,所有的问句都将以词向量来表示并输入网络;接下来将进入卷积层和记忆层,充分发挥CNN和Bi-LSTM各自的优势,保持问句信息并提取特征;随后利用注意力机制来识别问句最主要的特

征;最后经过分类器得出分类结果。

2.1 词向量层

首先,对输入层输入的问句进行分词,并通过Word2Vec得到问句中每个词所对应的词向量表示.这些词向量表示保持了问句最原始的信息,对接下来的步骤影响重大.假设问句 Q 包含 n 个单词 $Q=\{x_1, x_2, \dots, x_n\}$, x_i 代表问句中第 i 个词.如果问句带有答案信息,则加入答案词向量.例如,对于问题 Q “科比曾经效力哪只球队?”,对应答案 A “洛杉矶湖人队.”,则用 Q 和 A 一起表示该问题.如图一所示,首先利用词向量矩阵 E_w 来获得词向量.在这里 d 代表向量的维度, v_w 表示词汇大小.通过如下公式所示操作,可以将一个词 x_i 转变为词向量 e_i :

$$e_i = E_w v^i \quad (1)$$

其中, v^i 是向量 v^m 的大小.本文采用随机初始化词向量的方法,并在训练的过程中不断更新.经过这个步骤,问句将以词向量 $embeddings_q = \{e_1, e_2, \dots, e_n\}$ 的形式进入下一层网络。

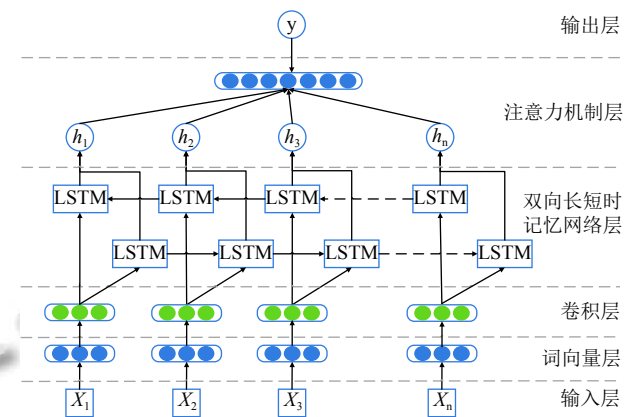


图1 基于深度神经网络的问句分类方法架构图

2.2 卷积层

在经过词向量层后,一个包含 n 个词的问句可以表示成如下形式:

$$h_{0:n-1} = h_0 \oplus h_1 \oplus \dots \oplus h_{n-1} \quad (2)$$

这里符号 \oplus 是两个相邻词之间的连接符.通常,向量 h_i : $i+j$ 表示一系列词向量 $h_i, h_{i+1}, \dots, h_{i+j}$.每一次卷积操作都包含一个过滤器 $w \in R^{md}$,它可以通过一个包含 m 个词的窗口来产生一个新的特征.例如,一个特征 c_i 可以由窗口 $h_{i:i+m-1}$ 产生:

$$c_i = f(w \cdot h_{i:i+m-1} + b) \quad (3)$$

这里的 $b \in R$ 是一个偏置项, f 是一个类似双曲正切的非线性函数. 问句最后可以被表示为:

$$c^* = [c_0, c_1, \dots, c_{n-m}] \quad (4)$$

2.3 双向长短时记忆网络层

长短时记忆网络最初被用来解决梯度消失问题, 随后许多基于长短时记忆的变体网络结构被提出. 本文采用了由 Graves^[15]提出的一种变体结构, 其能够在相同的记忆模块上增加窥视孔连接的权重.

特别地, 基于长短时记忆的循环神经网络有 4 个主要组成部分: 一个带有权重矩阵 W_{xi} , W_{hi} , W_{ci} , b_i 的输入门 i_t ; 一个带有权重矩阵 W_{xf} , W_{hf} , W_{cf} , b_f 的遗忘门 f_t ; 一个带有权重矩阵 W_{xo} , W_{ho} , W_{co} , b_o 的输出门 o_t ; 所有的这些门都将产生一定的影响, 使用当前的输入 x_t , 状态 h_{t-1} 在前一步就已经生成, 单元 c_{t-1} 的当前的状态决定是否使用这个输入, 遗忘之前存储的记忆, 最后输出生成的状态. 这些部分由以下这些公式来证明:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (6)$$

$$g_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + W_{cc}c_{t-1} + b_c) \quad (7)$$

$$c_t = i_t g_t + f_t c_{t-1} \quad (8)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (9)$$

$$h_t = o_t \tanh(c_t) \quad (10)$$

因此, 当前单元状态 c_t 的生成是由通过计算之前单元状态的权重和由这个单元所生成的当前信息决定.

对于很多句子级的处理任务, 考虑上下文信息是十分有必要的. 然而标准的 LSTM 网络在对句子进行建模的时候只是考虑了时序信息而忽略了下文信息. Bi-LSTM 网络通过引入第二层网络结构来扩展单向的 LSTM 网络, 而隐藏的连接在相反的时间顺序流动. 所以, Bi-LSTM 可以利用前后文的信息, 保证了在时间序列上过去和未来的信息都能考虑到.

本文提出的方法就使用了 Bi-LSTM 对问句进行建模. 如图一所示, 这个网络包含两个子网络分别对问句进行前后建模. 输出的第 i 个词如下面的式子所示:

$$h_i = [\vec{h}_i \oplus \overleftarrow{h}_i] \quad (11)$$

向前向后的输出最后是一个融合的结果.

2.4 注意力机制层

为了更好的捕捉问句中的有效信息, 抓住语义重

点, 本文在分类方法中加入了注意力机制^[16]. H 表示由上层 Bi-LSTM 网络输出向量所组成的矩阵, n 表示句子的长度. 问句的表示 r 就由这些向量的加权和构成, 如下公式所示:

$$M = \tanh(H) \quad (12)$$

$$\alpha = \text{softmax}(w^n M) \quad (13)$$

$$r = H\alpha^n \quad (14)$$

其中, $H \in R^{dn}$, d 代表词向量的维度, w 是一个训练的参数向量, w^n 是一个转置. 维度 w, α, r 与 d, n, d 分别对应. 所以最后用于分类的问句表示如下:

$$c^* = \tanh(r) \quad (15)$$

2.5 分类器

在这一层网络结构中, 我们使用一个 softmax 分类器从一组离散的类别 Y 来预测问句 Q 的标签 y . 分类器利用隐藏状态 c^* 作为输入:

$$\hat{p}(y|Q) = \text{softmax}(W^{(Q)}c^* + b^{(Q)}) \quad (16)$$

$$\hat{y} = \text{argmax}_y \hat{p}(y|Q) \quad (17)$$

损失函数如下:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m t_i \log(y_i) + \lambda \|\theta\|_F^2 \quad (18)$$

其中, $t \in R^m$ 是 one-hot 表示, $y \in R^m$ 代表估计每个类别的概率 (m 是目标类别的数目), 表示一个 L2 正则化参数.

2.6 正则化

本文在词向量层、Bi-LSTM 层和倒数第二层引入 dropout. 使用 L2 范式来约束权重向量, 通过重新调节 w , 使得 $\|w\| = s$, 每当梯度下降的时候 $\|w\| > s$, 如公式 (18) 所示.

3 实验分析

3.1 数据集

本实验主要使用了 3 个数据集, 包括: TREC、YahooAns、CQA dataset. 其中 TREC 和 YahooAns 是公共英文问句数据集, 但前者不包含答案集, 后者包含答案集. CQA dataset 是从两大中文社区问答社区: 百度知道和 360 问答中抓取的问句构成. 关于 3 个数据集的一些简要数据统计如表 1 所示, 对于每个数据集的详细介绍如下:

(1) TREC: TREC 问句集包含一系列事实类问句, 遵循广泛应用的 UIUC 英文问句分类体系, 问句分为 6 个大类 (ABBR, DESC, ENTY, HUM, LOC, NUM)^[1], 50 个小类, 每个大类会包含不同的小类. 选择这个数据集是因为该数据集比较经典, 适用广泛, 能较好的证明方法的性能.

(2) YahooAns: YahooAns 数据集是从雅虎问答社区上搜集下来的一批问句集并带有答案信息, 并且通过人工审核校验. 该数据集主要包含如下 4 个类别: “information”、“advice”、“opinion”和“polling”.

(3) CQA dataset: CQA dataset 是从百度知道和 360 问答中抓取的问句组成的数据集并带有答案信息. 所有选取的问句被分为 3 类, 分别为: 电脑网络、体育运动、地区.

3.2 实验设置

3.2.1 参数设置

在所有的 3 个数据集上, 算法模型所使用的参数都是一样的. 为了能够与 Kim^[13]之前的工作进行对比, 所以采用了其实验中的一些基本参数设置. 窗口值的大小为 3, dropout rate 设置为 0.5, l_2 constraint(s) 设置为 3. 在训练的时候, 我们采用小批量随机梯度下降法, 以减小训练损失. 本文把 mini-batch size 的大小设置为 50.

现在许多研究实验习惯利用由无监督的神经语言模型预先训练好的词向量, 因为这可以弥补大型监督训练集的一些缺陷, 从而提高实验效果^[15,16]. 本文在数据集 TREC 和 YahooAns 上使用的词向量是由 Word2Vec 预先训练好的包含 1000 亿词汇量的谷歌新闻语料, 在 CQA dataset 上使用的是由百度新闻上爬取的 10 亿词汇量新闻语料训练的词向量.

3.2.2 评价指标

实验的评价指标为准确率 (Accuracy, Acc), 均方根误差 (Root Mean Squared Error, RMSE), 具体计算公式如下所示:

$$Acc = \frac{AccNum}{TNum} \quad (19)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (p_i - g_i)^2}{TNum}} \quad (20)$$

其中, $AccNum$ 表示分类正确的测试集问句数目, $TNum$ 表示所有测试集的问句数目, $RMSE$ 是用来消除

预测类别 p_i 和真实类别 g_i 之间的差异.

3.3 实验结果

表 2 列出了在 3 个数据集上的分类准确度值和均方根误差值. 考虑到我们的方法是基于经典的 CNN 分类模型的改进, 所以本文的对比标准就是 Kim 提出的 CNN 文本分类模型, 并将该模型实验结果作为基准值.

从实验结果可以看出, 总体上在数据集 CQA dataset 和 YahooAns 上的实验结果没有在数据集 TREC 上好, 主要因为前两个数据集的问句复杂度明显要高于后者, 尤其是 CQA dataset.

模型 LSTM+CNNs 在 TREC 数据集上准确度提升的效果不明显, 但在 YahooAns 和 CQA dataset 上的准确度有较大提升, 分别提升了 2.5% 和 3.1%, 同时, 在 3 个数据集上的均方根误差都有明显降低, 在 3 个数据集上分别降低 0.7%, 1.2%, 1.4%. 实验结果验证了本文提出模型的有效性, 融合 LSTM 和 CNN 能更有效的表示问句.

模型 Att+LSTM+CNNs 取得了预期效果, 在每个数据集上都是取得最优结果, 这也说明考虑问句的答案信息是非常有必要的, 同时显示出注意力机制的有效性, 其确实能够更好的捕捉所要表示的文本特征. 特别在 CQA dataset 上, 取得了 16.1% 的准确度提高, 同时均方根误差下降了 3.6%. 无论是在英文数据集还是中文数据集, 本文提出的方法都取得了明显的效果提升, 不仅验证了该方法的有效性, 也从侧面体现出该方法的泛化能力.

在实验过程中, 除了使用了上文所提的三个问句数据集, 同时测试了两个情感分类的短文本数据集, 对比单一的模型, 本文提出的融合双向长短时记忆网络和卷积神经网络并包含注意力机制的模型也取得了一些准确率的提升. 因此, 本文提出的模型具有一定的通用性, 在作适当修改后应该能应用到其他文本分类任务中去. 另外, 在实验运行的过程中发现, 本文提出的融合模型虽然取得了精度上的提高, 但同时会带来一些运算复杂度的提升, 会稍微增加些运算时间, 但不会造成特别严重的性能下降.

表 2 在各个数据集上的实验结果对比

Model	TREC		YahooAns		CQA	
	Acc	RMSE	Acc	RMSE	Acc	RMSE
CNNs	0.936	0.484	0.731	0.634	0.682	0.655
LSTM+CNNs	0.941	0.477	0.756	0.622	0.713	0.641
Att+LSTM+CNNs	-	-	0.793	0.613	0.789	0.619

4 结论与展望

本文提出了一种基于深度学习的分类方法,该方法融合双向长短时记忆网络(Bi-LSTM)和卷积神经网络(CNN)并带有注意力机制.其特色在于:一方面利用Bi-LSTM和CNN结合的方式来同时获取问句的时序特征和本质特征,以此来最大化提取待分类问句中的有效信息.另一方面,在模型中加入注意力机制,充分利用问句的答案信息来增强问句表示.实验结果表明本文提出的问句分类方法与传统的机器学习方法和单一的神经网络结构相比具有更高的准确率,在多个数据集上都取得了不错的效果.

在接下来的工作中,将考虑对待分问句进行更好的预处理操作,减少噪声数据.同时,尝试优化不同的神经网络模型并进行有效的融合,看能否更好的对问句进行向量化表示,获取问句中最有效的信息.

参考文献

- 1 Li X, Roth D. Learning question classifiers. Proceedings of the 19th International Conference on Computational Linguistics. Taipei, China. 2002. 1-7. [doi: [10.3115/1072228.1072378](https://doi.org/10.3115/1072228.1072378)]
- 2 镇丽华, 王小林, 杨思春. 自动问答系统中问句分类研究综述. 安徽工业大学学报(自然科学版), 2015, 32(1): 48-54, 66. [doi: [10.3969/j.issn.1671-7872.2015.01.010](https://doi.org/10.3969/j.issn.1671-7872.2015.01.010)]
- 3 Shen D, Pan R, Sun JT, *et al.* Query enrichment for web-query classification. ACM Transactions on Information Systems, 2006, 24(3): 320-352. [doi: [10.1145/1165774](https://doi.org/10.1145/1165774)]
- 4 Broder A, Fontoura M, Gabrilovich E, *et al.* Robust classification of rare queries using Web knowledge. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07). Amsterdam, Holland. 2007. 231-238. [doi: [10.1145/1277741.1277783](https://doi.org/10.1145/1277741.1277783)]
- 5 Hui ZJ, Liu J, Ouyang LM. Question classification based on an extend class sequential rule model. Proceedings of the 5th International Joint Conference on Natural Language Processing. Chiang Mai, Thailand. 2011. 938-946.
- 6 Mishra M, Mishra VK, Sharma HR. Question classification using semantic, syntactic and lexical features. International Journal of Web & Semantic Technology, 2013, 4(3): 39-47.
- 7 Aikawa N, Sakai T, Yamana H. Community QA question classification: Is the asker looking for subjective answers or not? IPSJ Online Transactions, 2011, (4): 160-168. [doi: [10.2197/ipsjtrans.4.160](https://doi.org/10.2197/ipsjtrans.4.160)]
- 8 Liu L, Yu ZT, Guo JY, *et al.* Chinese question classification based on question property kernel. International Journal of Machine Learning and Cybernetics, 2014, 5(5): 713-720. [doi: [10.1007/s13042-013-0216-y](https://doi.org/10.1007/s13042-013-0216-y)]
- 9 杨思春, 高超, 秦锋, 等. 融合基本特征和词袋绑定特征的问句特征模型. 中文信息学报, 2012, 26(5): 46-52. [doi: [10.3969/j.issn.1003-0077.2012.05.008](https://doi.org/10.3969/j.issn.1003-0077.2012.05.008)]
- 10 王艳娜, 孙丙宇. 基于卷积神经网络的烟瘾渴求脑电分类. 计算机系统应用, 2017, 26(6): 254-258.
- 11 Wei YC, Zhao Y, Lu CY, *et al.* Cross-modal retrieval with CNN Visual features: A new baseline. IEEE Transactions on Cybernetics, 2017, 47(2): 449-460. [doi: [10.1109/TCYB.2016.2519449](https://doi.org/10.1109/TCYB.2016.2519449)]
- 12 Hao YC, Zhang YZ, Liu K, *et al.* An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada. 2017. 221-231. [doi: [10.18653/v1/P17-1021](https://doi.org/10.18653/v1/P17-1021)]
- 13 Kim Y. Convolutional neural networks for sentence classification. arXiv eprint arXiv: 1408.5882.
- 14 Shi YY, Yao KS, Tian L, *et al.* Deep LSTM based feature mapping for query classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, CA. 2016. 1501-1511. [doi: [10.18653/v1/N16-1176](https://doi.org/10.18653/v1/N16-1176)]
- 15 Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada. 2013. 6645-6649. [doi: [10.1109/ICASSP.2013.6638947](https://doi.org/10.1109/ICASSP.2013.6638947)]
- 16 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv eprint arXiv:1409.0473, 2014.