基于注意力机制的上下文相关的问答配对方法*

王路，张璐，李寿山，周国栋
（苏州大学 计算机科学与技术学院，江苏省 苏州市 215006）

摘 要 目前，关于问答的大部分研究都是面向正式文本的问答对。然而，与以往研究不同的是，本文关注于社会媒体上存在的非正式文本问答对。非正式文本会存在问题文本里包含多个问题以及回答文本里包含多个回答的情况。针对该情况，我们提出了一个新的任务：问答配对，即对问题文本的每个问题，从答案文本中找到和该问题相关的句子。首先，我们从产品问答网站上收集了大规模非正式文本问答对，并在此基础上创建了一个产品问答配对语料库。其次，为了解决非正式文本中存在的噪声问题，提出了一种基于注意力机制的上下文相关的问答配对方法。实验结果表明，本文提出的方法能有效地提升非正式文本的问答配对的性能。

关键词 非正式文本，问答配对，上下文相关，注意力机制
中图分类号：TP391 **文献标识码：**A

Contextual QA Pairing Method Based on Attention Network

Lu Wang, Lu Zhang, Shoushan Li, Guodong Zhou
(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract Nowadays, most studies on QA are based on formal text while there is a large number of QA pairs with informal text in social media. Because of informal text may have the situation that question text has several questions and answer text has several answers, we propose a new task: QA pairing, which means for each question, to find sentences in answer that are relevant with the question. First, we build a novel QA pairing corpus with informal text which is collected from a product reviewing website. Then, in order to solve the nosiy problem in informal text, we propose a novel QA pairing approach, namely contextual QA pairing method based on attention network. Empirical studied demonstrate the effectiveness of the proposed approach to QA pairing.

Keywords informal text, QA pairing, contextual, attention mechanism

1 引言

问答相关的研究是自然语言处理领域的一个热点领域。以往的研究主要集中于问答匹配，并且主要是针对正式文本的。然而，社会媒体上存在大量非正式文本的问答对。非正式文本存在问题文本里包含多个问题以及回答文本里包含多个回答的情况。针对该情况，我们提出了问答配对任务。问答配对是针对问题文本中的每个问题，从答案中找到和该问

例1：问题：更新系统后卡吗？屏幕怎么样？续航怎么样？
答案：一点都不卡，内存是4G的。不一直玩游戏的话，可以用一天。
标注：1-1，2-0，3-3+4

图 1 非正式文本问答对
Fig.1 Informal QA pair

题相关句子的任务。例如：图 1 所示，例 1 中的问题文本"系统更新后卡吗？屏幕怎么样？续航怎么样？"可以被切分为三个问题，分别是"系统更新后卡吗？"、"屏幕怎么样？"以及

*收稿日期： 定稿日期：
基金项目：国家自然科学基金重点项目（No.61331011）；国家自然科学基金（No.61672366）
作者简介：王路（1994——），男，硕士研究生，主要研究方向为问答；张璐（1994——），女，硕士研究生，主要研究方向为情感分析；李寿山（1980——），男，教授，主要研究方向为情感分析；周国栋（1967——），男，教授，主要研究方向为自然语言处理。

“续航怎么样？”。答案文本“一点都不卡，内存是 4G 的。不一直玩游戏的话，可以用一天。”根据空格、逗号、顿号、问号、感叹号和句号可以被切分为四个句子，分别是“一点都不卡”、“内存是 4G 的”、“不一直玩游戏的话”以及“可以用一天”。对于第一个问题“系统更新后卡吗？”，答案文本中只有第一个句子回答了该问题，对于第二个问题“屏幕怎么样？”，答案文本中第三个和第四个句子共同回答了该问题。对于第三个问题“电池怎么样”，答案文本中并没有针对该问题的回答。因此，非正式文本的问答配对是一个崭新并富有挑战性的任务。为了简化问题，我们假设问题已经被切分成多个子问题。本文关注于从答案文本中找到针对每一个问题的具体回答。总体而言，针对该问题的研究主要存在以下难点：

答案文本中句子较多，换言之，噪声较多，挑选出正确的句子较有难度。并且，有的问题的回答是由答案文本中多个句子组合而成的，例如：图 1 所示，问题文本中的第三个问题“续航怎么样？”是由答案文本中“不玩游戏的话” 以及“可以用一天”这两个句子共同回答的。基于此，将答案文本中每个句子和问题进行单独分类就显得不合理。

在本文中，我们关注于非正式文本的问答配对。首先，我们从产品评论网站上收集了大规模非正式文本的问答对，并标注这些问答对从而创建了一个非正式文本的问答配对语料库。然后，针对上述难点，我们提出了一种上下文相关的输入架构来处理答案中多个句子共同回答一个问题的情况，为了处理答案文本中的噪声问题，我们引入了注意力机制来准确地挑选出正确的答案句子。实验结果表明本文提出的基于注意力机制的上下文相关的问答配对方法能够在非正式文本的问答配对任务中获得较好的性能。

本文的结构如下安排：第二节介绍与本文相关的一些工作；第三节介绍语料收集和标注工作；第四节介绍本文提出的关于问答配对的模型；第五节给出实验结果和分析；第六节对本文进行总结，并对下一步工作进行展望。

2 相关工作

目前尚无问答配对的相关工作，与之接近的是问答匹配的相关工作，即给定一个问题和一个答案，判断该答案和问题是否匹配。

总体而言，问答匹配方法可以分为两大类：浅层学习方法和深度学习方法。

浅层学习方法：Wang[3]等采用一些浅层学习算法，如条件随机场、支持向量机和最大熵等来训练模型。除了浅层学习算法，浅层学习方法的相关研究还关注于特征工程、语言学工具和外部资源等，如 Yih[4]等采用了语义资源 Word-Net，Yao[5]等采用了树的编辑距离，Severyn[6]等采用了命名实体等。

深度学习方法：采用神经网络来进行模型训练。主要分为三类：单纯网络（siamense networks），注意力网络（attentive networks）以及聚合比较网络（compare-aggregate networks）。

1. 单纯网络：相关研究采用的是经典的神经网络。Feng[7]等采用卷积神经网络（CNN）来获得问题和答案的融合表示，并利用这个表示进行分类。Yang[2]等采用卷积神经网络(RNN)来分别表示问题和答案，并连接它们进行分类。

2. 注意力网络：不同于用 LSTM 的最终时间步来表示一个句子，一般采用注意力机制来得到每个时间步的权重，然后用这些权重来表示句子。Tan[8]等通过计算出问题的权重来附加到答案中，然后利用问题和附加了问题信息的回答来进行分类。Yin[9]等将问题和答案通过卷积神经网络计算出注意力矩阵，利用注意力矩阵来进行分类。

3. 聚合比较网络：采用不同的匹配策略来获得单词间的关系，从而进行分类。Wang[10]等将问题和答案间的时间步分别进行了矩阵乘法、元素级别乘法、元素级别除法和欧几里得距离等计算，从而进行分类。Wang[11]等提出了一种新的计算方式，相较于传统的单角度计算，他采用了多角度的计算方式，获得了较好的效果。He[12]等利用卷积神经网络计算出一个配对的词的相互作用矩阵，并利用该矩阵进行分类。

然而，上面所有的方法都是一对一（一个问题句子对应一个答案句子）匹配模型。本文为了充分利用上下文信息，采用的是上下文相关的结构，并结合注意力机制来进行问答配对。

3 语料收集和标注

淘宝[1]是中国最大的电子商务平台，我们从淘宝的“问大家”板块收集了 2000 个问答对。这些问答对主要来自于数码领域。需要注意的是，如果一个问题包含多个子问题，我们把问题切分为多个子问题，并且把每个子问题和答案所有的句子看作是一个问答对。例如，如图 1 所示，例 1 包含了三个问题，每一个问题的答案都是通过该问题的答案句子在所有答案句子中的位次来标注的。其中“X-Y”表示的是第 X 个问题的回答是答案文本中的第 Y 个句子。

为了确保较高的一致性，在多次标注少量语料后，我们提出了一些标注规范。然后，我们让更多人根据该标注规范来标注整个数据集。下面具体展示一些标注的例子。

例 3：问题：怎么查是不是正品？另外是新机吗？
答案：是的，手机设置理由
标注：1-2,2-1

例 4：问题：有没有买红米 4 高配的？手机是不是正品？卡不卡？
答案：手机正品 还可以！送的耳机不能用
标注：1-0,2-1,3-2

例 5：问题：可以玩王者吗？卡不？
答案：可以，不开高特效的话，不卡
标注：1-1,2-2+3

对于每个问答对，我们安排了两位标注人员来标注类别，一致性检验 kappa 值为 0.83。为了应对两位标注人员标注不一致的情况，我们安排了一位熟练的标注人员来检查，确保标注的一致性。表 1 展示了语料的类别分布（把每一个问题和答案的每一个句子进行分类，看该答案是否匹配该问题）。

表 1 语料类别分布

Table 1 Category distribution of the corpus

类别	数量
匹配	2387
不匹配	3309

4 基于注意力机制的上下文相关的问答配对

在本章中，我们把我们问答配对的方法分为两步。首先，我们提出了充分利用上下文信息的输入架构。然后，我们提出了计算问题文本和融合了上下文信息的答案文本间的匹配度的基于注意力机制的配对模型。

4.1 基于注意力机制的上下文相关的输入架构

图 2 给出上下文相关的输入架构的整体框架图。该架构的主要目的是充分将当前文本与上下文信息相结合，从而提升信息量。图中的相关层将在下面进行详细介绍。

词编码层：该层有三个输入：当前答案句子前面所有的句子，当前答案句子和当前答案句子后面所有的句子。首先，用 word2vec[2]来预训练词向量，并用这些词向量来表示这三个输入。然后，利用三个双向长短期记忆网络（Bi-directional Long Short Term Memory, Bi-LSTM）来同时把三个句子的时间步编码为上下文相关的词向量。

$$H_{Fro.} = \overrightarrow{LSTM}(W_{Fro.}) \oplus \overleftarrow{LSTM}(W_{Fro.}) \quad (1)$$

$$H_{a} = \overrightarrow{LSTM}(W_{a}) \oplus \overleftarrow{LSTM}(W_{a}) \quad (2)$$

$$H_{Beh.} = \overrightarrow{LSTM}(W_{Beh.}) \oplus \overleftarrow{LSTM}(W_{Beh.}) \quad (3)$$

注意力层：该层的目的是给词向量矩阵分配权重，从而获得当前句子最有信息量的表示。

通过下面的句子计算注意力权重，并归一化：

$$a_{Fro.} = softmax(tanh(W_{Fro.} H_{Fro.} + b_{Fro.})) \quad (4)$$

$$a_{Beh.} = softmax(tanh(W_{Beh.} H_{Beh.} + b_{Beh.})) \quad (5)$$

其中，$W_{Fro.}$ 和 $W_{Beh.}$ 是权重矩阵，$b_{Fro.}$ 和 $b_{Beh.}$ 是偏置矩阵。

[1] https://www.taobao.com/

[2] https://radimrehurek.com/gensim/models/word2vec.html

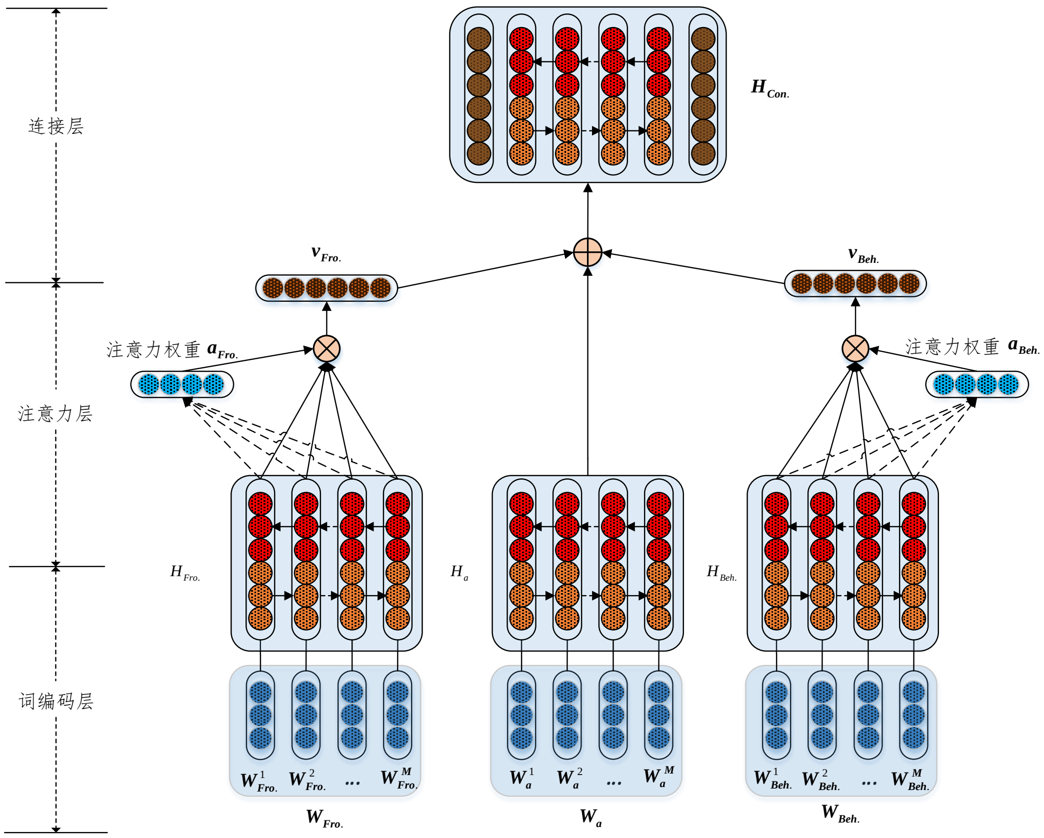

图 2 基于注意力机制的上下文相关的输入架构

Fig. 2 Structure of contextual input based on attention mechanism

然后，我们分别计算出句子向量 $v_{Fro.}$ 和 $v_{Beh.}$：

$$v_{Fro.} = a_{Fro.} H_{Fro.} \tag{6}$$

$$v_{Beh.} = a_{Beh.} H_{Beh.} \tag{7}$$

连接层：该层的目的是连接当前句子的词向量矩阵和上下文句子向量 $v_{Fro.}$ 以及 $v_{Beh.}$，从而获得上下文相关的答案表示。

$$H_{Con.} = v_{Fro.} \oplus H_a \oplus v_{Beh.} \tag{8}$$

4.2 基于注意力机制的上下文相关的问答配对

图 3 给出基于注意力机制的上下文相关的问答配对的整体框架图。该模型的主要目的是利用问答匹配机制来进行分类。图中的相关层将在下面进行详细介绍。

词编码层：该层有两个输入：问题文本和答案文本。答案文本参照上文提出的基于注意力机制的上下文相关的输入架构进行编码，再按前向和后向两个方向分别拆分为 $H_{Con.}$。将问题文本通过一个双向长短期记忆网络来把问题的时间步编码为上下文相关的词向量。

$$\overrightarrow{H_q} = \overrightarrow{LSTM}(W_q) \tag{9}$$

$$\overleftarrow{H_q} = \overleftarrow{LSTM}(W_q) \tag{10}$$

匹配层：该层的目的是在每个方向上分别比较问题的最终时间步向量和问题的每个时间步向量，反之亦然。

通常，我们通过余弦函数来计算两个向量间的相似度：

$$\overrightarrow{S_{q_N - a_i}} = cosine(\overrightarrow{h_q^N}, \overrightarrow{h_a^i}) \tag{11}$$

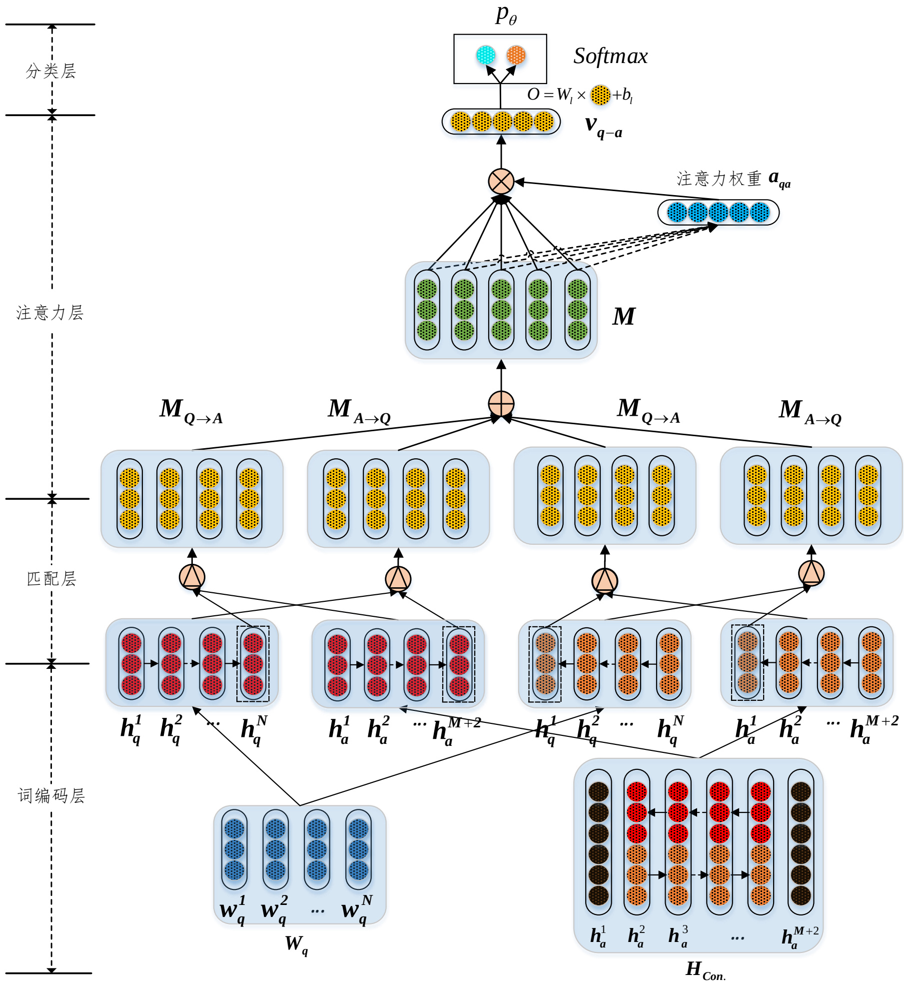

图 3 基于注意力机制的上下文相关的问答配对

Fig.3 Contextual QA Pairing Method Based on Attention Network

受到 Wang[11]的影响，我们采用了多角度的全匹配策略：

$$\overrightarrow{s}^{k}_{q_N-a_i} = cosine(W^p \circ \overrightarrow{h}^N_q, W^p \circ \overrightarrow{h}^i_a) \qquad (12)$$

其中，$\circ$代表元素级别的乘法，W^p代表
W 的第 p 行，它控制着第 p 个角度并且给维度空间中的不同维度分配不同的权重。

然后，我们可以得到相似度矩阵：

$$\overrightarrow{M}_{Q\to A} = \begin{bmatrix} \overrightarrow{s}^{1}_{q_N-a_1} & \cdots & \overrightarrow{s}^{1}_{q_N-a_M} \\ \vdots & \ddots & \vdots \\ \overrightarrow{s}^{P}_{q_N-a_1} & \cdots & \overrightarrow{s}^{P}_{q_N-a_M} \end{bmatrix} \qquad (13)$$

其中，P表示角度的数量。

同样地，我们可以同样地计算出$\overleftarrow{M}_{Q\to A}$，$\overrightarrow{M}_{A\to Q}$ 和$\overleftarrow{M}_{A\to Q}$。

注意力层：该层的目的是给匹配矩阵分配权重，从而获得当前问答对最有信息量的表示。

首先，我们连接$Q\to A$和$A\to Q$相似度矩阵，从而得到问答匹配矩阵。

$$M = \overrightarrow{M}_{Q\to A} \oplus \overleftarrow{M}_{Q\to A} \oplus \overrightarrow{M}_{A\to Q} \oplus \overleftarrow{M}_{A\to Q} \qquad (14)$$

然后，我们计算出注意力权重并归一化。

$$a_{qa} = softmax(tanh(W_{qa}H_{qa} + b_{qa})) \quad (15)$$

其中，W_{qa} 是权重矩阵，b_{qa} 是偏置矩阵。

最后，将 a_{qa} 和 M 相乘，从而获得问答向量 v_{qa}。

$$v_{qa} = a_{qa}M^T \quad (16)$$

分类层：该层的目的是利用问答向量 v_{qa} 来进行分类。我们将 v_{qa} 传入 softmax 分类器：

$$o = W_l v_{qa} + b_l \quad (17)$$

其中，$o \in R^K$ 是输出，W_l 是权重矩阵，b_l 是 偏置矩阵。K 表示类别的数目，那么对于类别 $k \in [1, K]$ 的概率计算如下：

$$p_\theta = \frac{exp(o_k)}{\sum_{t=1}^{K} exp(o_t)} \quad (18)$$

最后，具有最高概率的标签作为最终的分类结果。

4.3 模型训练

我们采用交叉熵损失函数来端到端训练模型。具体而言，输入训练数据集合 x_t, y_t，x_t 是第 t 个要被预测的问答对，y_t 是 x_t 的真实标签。模型表示为一个黑盒函数 $\sigma(x)$。模型的输出是一个代表了每个类别概率的向量。训练的目标是最小化下面的损失函数：

$$J(\theta) = -\sum_{t=1}^{N}\sum_{k=1}^{K} y_t^k \cdot \log\sigma(x_t) + \frac{l^{'}}{2}\|\theta\|_2^2 \quad (19)$$

其中，N 代表训练样例的数目，$l^{'}$ 是对偏置参数的 L_2 正则化。

我们采用 Kingma[13] 等提出的 Adam 算法作为优化算法。数据在 $\left[-\sqrt{6/(r+c)}, \sqrt{6/(r+c)}\right]$ 的范围内进行初始化，其中 r 表示行的数目，c 代表列的数目。

5 实验

5.1 实验设置

数据设置：采用第二节内容介绍的数据集作为实验数据。该数据集包含 5698 个问答对。我们随机将数据分为训练集（每个类别的 80%）和测试集（每个类别剩下的 20%）。此外，将来自训练集的 10%数据作为开发集，用于学习算法中的参数调整。

词切分和词向量训练：采用 Jieba[3]分词来进行分词，并采用 word2vec 来训练词向量。训练词向量的数据来自于数码领域，共计 200000 条问答对。词向量的维度设置为 100。

句子切分：采用斯坦福大学自然语言处理实验室公布的 CoreNLP[4]工具来进行句子切分。

模型参数：实验中所有模型使用深度学习开源框架 Tensorflow[5]搭建。模型中所有的超参数都是通过开发集的性能来进行调整。LSTM 的单元数设置为 128。Batch size 设置为 64。优化器的学习率为 0.001。迭代次数为 30。

评价标准和显著性测试：我们用标准的精确率（P），召回率（R），F1 值和准确率（Accuracy）来评价性能。我们采用 t 测试来评价两个方法间的显著性程度。

5.2 基线模型

所有基线模型均采用和采用和 4.1 节和 4.2 节提出的切分方式。

- **最大熵（MaxEnt）：**基于浅层学习方法最大熵训练的分类器。
- LSTM：Bowman[14]等提出的一种基于单纯网络的问答匹配方法。
- SCNN：Zhang[15]等提出的一种基于单纯网络的隐含话题识别的最佳方法。

---

3 https://pypi.python.org/pypi/jieba/
4 https://stanfordnlp.github.io/CoreNLP/download.html
5 https://www.tensorflow.org/

表 2 不同问答配对方法的整体性能
Fig.2 Overall performances of different approaches to QA pairing

	Macro-F	*Accuracy*
MaxEnt	0.592	0.618
LSTM(Bowman[14],2015)	0.628	0.639
SCNN(Zhang[15],2015)	0.614	0.635
Attentive LSTM(Tan[8],2016)	0.625	0.649
MULT(Wang[10],2016)	0.652	0.677
BIMPM(Wang[11],2017)	0.660	0.685
CIS	0.678	0.704
AM	0.671	0.697
CPMAN	**0.710**	**0.728**

表 3 不同问答配对方法在每个类别中的性能
Fig.3 Performances of different approaches to QA pairing in each category

	匹配			不匹配		
	P	*R*	*F*	*P*	*R*	*F*
MaxEnt	0.487	0.494	0.491	0.696	0.690	0.693
LSTM(Bowman[14],2015)	0.514	0.626	0.564	0.743	0.647	0.692
SCNN(Zhang[15],2015)	0.510	0.541	0.525	0.717	0.690	0.703
Attentive LSTM(Tan[8],2016)	0.529	0.534	0.532	0.721	0.717	0.719
MULT(Wang[10],2016)	0.569	0.551	0.560	0.738	0.752	0.745
BIMPM(Wang[11],2017)	0.582	0.553	0.567	0.741	0.763	0.752
CIS	0.611	0.565	0.587	0.752	**0.786**	0.768
AM	0.599	0.560	0.579	0.748	0.777	0.762
CPMAN	**0.632**	**0.647**	**0.640**	**0.785**	0.776	**0.781**

- Attentive LSTM：Tan[8]等提出的一种属于注意力网络的性能最佳的问答匹配方法。
- MULT：Wang[10]等提出的一种基于聚合比较网络的性能最佳的问答匹配方法。
- BIMPM：Wang[11]等提出的另一种基于聚合比较网络的性能最佳的问答匹配方法。我们实现了集成了四种匹配方式的 BIMPM（Ensemble）。

5.3 我们的方法

- **上下文相关的输入架构（Contextual Input Structure, CIS）**：具体实现中，只采用了上下文相关的输入架构，并没有采用注意力机制，所有通过注意力机制挑选的向量均用 LSTM 输出的所有时间步的词向量的平均来代替。
- **注意力机制（Attention Mechanism, AM）**：具体实现中，只采用了注意力机制，并没有采用上下文相关的输入架构，即输入只有当前句子，不再需要当前句子的上下文信息。
- **基于注意力机制的上下文相关的问答配对方法（Contextual QA Pairing Method based on Attention Network, CPMAN）**：具体实现中，同时利用了上下文相关的输入架构以及注意力机制，整体结果如图 2 以及图 3 所示。

5.4 实验结果

表 2 和表 3 给出了不同问答匹配方法的总体性能及每个类别上面的性能。从表中结果来看，所有深层学习方法的性能均优于浅层学习方法。我们方法的 3 种实现的性能均优于所有基线方法。*t* 测试结果表明我们方法带来的性能提升具有统计显著性（p-value<0.05）。

此外，我们的方法在“匹配”这个类别上表现更出色。我们的 CPMAN 的实现比基线方法中最好的 BIMPM 在 P 值上高了 0.05，在 *R* 值上高了 0.094 并且在 *F* 值上高了 0.073。在所有的实现中，CPMAN 的整体性能最好。该结果表明在问答配对中，上下文相关的输入架构以及注意力机制均可以显著提高配对性能，并且两者相结合的性能最佳。

5.5 注意力可视化

为了更好地理解注意力机制并检验模型是否可以获取句子中的关键信息，我们对注意力权重进行了可视化。

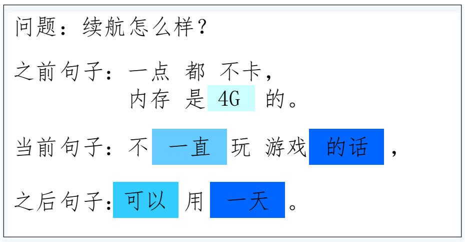

图 4 问答对的注意力可视化

Fig.4 Attention visualization for a QA pair

图 4 给出了某个问答对的注意力可视化结果。图中蓝色表示词的权重，颜色的深度表示了重要程度，颜色越深表示重要性越高。从图中结果可以看出，注意力机制缺失可以挑选出和匹配相关的词。例如：当前句子中的词“的话”，隐含地表示了条件型回答，模型从之后的句子中挑选出了“可以”和“一天”两个词，最终分类问题和当前句子匹配。

6 结束语

本文构造了一个问答配对语料库，用于研究非正式文本的问答配对方法。在此基础上，本文提出了一种基于注意力机制的上下文相关的问答配对方法。该方法提出了一种上下文相关的输入架构并结合了注意力机制。实验结果表明本文提出的方法能够显著提升问答配对的性能。

下一步工作中，我们准备进一步扩大语料的标注规模，尝试标注其他领域的数据。此外，我们将测试本文的方法在其他领域或其他语言的问答配对任务中的有效性。

参考文献

[1] Wang Mengqiu, Smith N A and Mitamura T. What is the Jeopardy Model? A Quasi-Synchronous Grammar for QA// Proceedings of EMNLP, 2007:22-32.

[2] Yang Yi, Yih W and Meek C. WikiQA: A Challenge Dataset for Open-domain Question Answering// Proceedings of EMNLP, 2015:2013-2018.

[3] Wang Mengqiu and Manning C D. Probabilistic Tree-Edit Models with Structured Latent Variables for Textual Entailment and Question Answering// Proceedings of ACL, 2010:1164-1172.

[4] Yih W, Chang Ming-wei, Meek C, et al. Question Answering Using Enhanced Lexical Semantic Models// Proceedings of ACL, 2013:1744-1753.

[5] Yao Xuchen and Dueme B B. Answer Extraction as Sequence Tagging with Tree Edit Distance// Proceedings of NAACL-HLT, 2013:858-867.

[6] Severyn A and Moschitti A. Automatic Feature Engineering for Answer Selection and Extraction// Proceedings of EMNLP, 2013:458-467.

[7] Feng Minwei, Xiang Bing, Glass M R, et al. Applying Deep Learning to Answer Selection: A Study and An Open Task// Proceedings of IEEE ASRU, 2015:813-820.

[8] Tan Ming, Santos C D, Xiang Bing, et al. Improved Representation Learning for Question Answer Match-

ing// Proceedings of ACL, 2016:464-473.

[9] Yin Wenpeng, Schütze H, Xiang Bing, et al. ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs// Computer Science, 2015.

[10] Shuohang Wang and Jing Jiang. A Compare-Aggregate Model for Matching Text Sequences// arXiv preprint arXiv: 1611.01747, 2016.

[11] Wang Zhiguo, Hamza W and Florian R. Bilateral Multi-Perspective Matching for Natural Language Sentences// Proceedings of IJCAI, 2017:4144-4150.

[12] He Hua and Lin J. Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement// Proceedings of NAACL-HLT, 2016:937–948.

[13] Kingma D P and Ba J. Adam: A Method for Stochastic Optimization// Computer Science, 2014.

[14] Bowman S R, Angeli G, Potts C, et al. A large annotated corpus for learning natural language inference// Proceedings of EMNLP, 2015:632-642.

[15] Zhang Biao, Su Jinsong, Xiong Deyi, et al. Shallow Convolutional Neural Network for Implicit Discourse Relation Recognition// Proceedings of EMNLP, 2015:2230-2235.

作者联系方式：王路 江苏省苏州市苏州大学本部理工楼 416 实验室 215006 18896584235 wanglu_1994@126.com