

# 基于动态网络的简答题答案标注方法

洪旭东, 严梅\*, 莫媛媛

安徽工业大学计算机科学与技术学院, 安徽 马鞍山

收稿日期: 2022年8月23日; 录用日期: 2022年9月21日; 发布日期: 2022年9月28日

## 摘要

阅读理解问答是当前自然语言处理的研究热点之一。针对阅读理解问答中的简答题, 由于其答案往往有多个、分布在阅读材料的不同位置, 现有方法大多难以有效获取。本论文面向简答题的答案标注方法开展研究, 将简答题答案标注问题看作是序列标注问题, 考虑到问句焦点词对于寻找答案具有重要作用, 提出基于动态网络的简答题答案标注方法, 实现了简答题答案的端到端标注, 实验表明提出的方法在F1和EM值上效果都有所提升。

## 关键词

阅读理解, 问答, 答案标注, 动态网络, 关注机制

# Answer Annotation for Short-Answer Questions Based on Dynamic Network

Xudong Hong, Mei Yan\*, Yuanyuan Mo

School of Computer Science and Technology, Anhui University of Technology, Ma'anshan Anhui

Received: Aug. 23<sup>rd</sup>, 2022; accepted: Sep. 21<sup>st</sup>, 2022; published: Sep. 28<sup>th</sup>, 2022

## Abstract

Reading comprehension style question answering is one of the current research hotspots in natural language processing. For the short-answer questions, since the answers are often multiple and distributed in different positions of the reading passages, most of the existing methods are difficult to obtain effectively. This paper studies the answer labeling method for short-answer questions, regards the short-answer question answer labeling problem as a sequence labeling problem, and considers that the focus words of questions play an important role in finding answers, and proposes a dynamic network-based short-answer question answer labeling method. End-to-end

\*通讯作者。

annotation of short-answer questions' answers, experiments show that the proposed method improves both F1 and EM values.

## Keywords

Reading Comprehension, Question Answering, Answer Pointing, Dynamic Network, Attention Mechanism

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

阅读理解问答指的是给定一段阅读材料, 计算机通过对内容的阅读, 自动回答用户相关的提问, 在自动客服、辅助答题、信息检索等诸多领域具有广泛的应用前景, 也是当前的研究热点之一。现有阅读理解问答研究大多针对答案是一个连续的短文本片段的场景, 通过判断答案的开始和结束位置, 从而完成答案抽取。答案的长度往往仅为 2~3 个单词。对于简答题, 由于往往是综合性问题, 其答案通常有多个, 分布在阅读材料的不同位置, 而且答案内容有长有短。例如“这次会议有哪些人参加了?” 它对应的答案往往就散布在阅读材料的不同位置上。现有方法难以有效获取。

论文面向阅读理解问答中简答题的答案标注方法开展研究, 构建了端到端的答案标注模型。在构建模型时, 一是将简单题答案标注看成序列标注问题, 对多个答案片段进行标注。二是考虑到问句中的焦点词对于寻找答案具有重要作用, 例如“……哪些人参加了?” 中的“人”就表明了答案是与“人”相关的概念, 将在答案的识别过程中同时对问句焦点词进行识别, 并通过动态网络, 将识别过程与答案标注过程结合。此外论文针对该研究构建了一个面向新闻事件的问答语料库, 并在此基础上进行了简答题的答案标注实验, 实验结果表明论文提出的方法能够有效提升简答题的答案标注效果。

## 2. 相关工作

现有面向阅读理解的问答方法, 大多为基于深度学习的方法。这些方法所采用的深度学习模型总体架构如图 1 所示。模型主要包括词向量层、编码层、交互层、答案标注层 4 个层次。首先利用词向量层对问题和段落中的词语进行分布式表示。然后阅读层利用循环神经网络对问题和阅读材料的上下文进行阅读, 获得词语在上下文环境下新的语义表示。接着交互层利用关注机制获得阅读材料中每个词在特定问题下的向量表示。答案标注层则是对答案片段的开始和结束位置进行识别。

对于词向量层, [1]采用了词级别词向量 Glove [2]和处理未登录词的字符级词向量[3], 并利用 highway 网络[4]对两个词向量进行整合。[5]提出对词语的语法分析结果, 如词性、实体类别等, 进行向量化表示并将结果加入到词向量中。[6]提出在词向量中增加 2 维, 用来标识阅读材料中的词语是否在问题中出现。对于阅读层, [7]采用了 LSTM [8] [9]采用了 GRU [10] [11]则采用双向循环神经网络。[12]采用多层循环神经网络。[13]对阅读材料中不同长度内容同时进行语义建模, 获得阅读材料不同层次的语义表示。对于交互层, [14]使用阅读材料中词语对问题中词语进行关注, 得到在给定问题下阅读材料的语义表示, [1] [14] [15]提出了双向关注机制, 同时利用问题中词语对阅读材料中词语进行关注, 得到阅读材料中不同词语的重要性。[16]提出了多种关注值的计算方法, 从而获得词语在不同视角下的关联。[11]考虑了阅读材料内部

词语之间的关联，提出在问题和阅读材料的交互后，进一步进行阅读材料的自我关注。对于答案标注，现有方法大多借助于 PointerNet [17]计算答案开始和结束位置的概率。

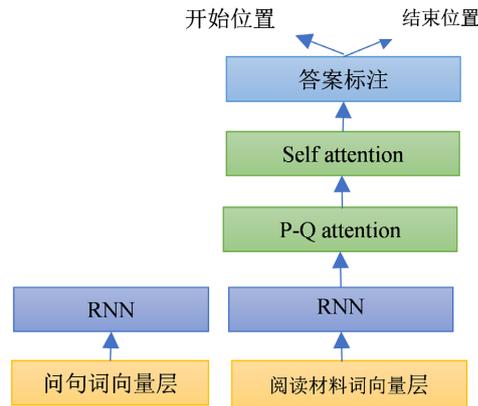


Figure 1. Model architecture of reading comprehension question answering

图 1. 阅读理解问答模型架构

这些阅读理解问答研究，大多基于 SQuAD 2.0 [18]、NarrativeQA [19]、MS MARCO [20]等阅读理解问答数据集，然而这些数据集的答案大多为一个阅读材料中的一个连续的文本片段，或是一段人为撰写的答案，与简答题有多个不同答案片段的场景有所区别。相关方法也无法直接应用于简答题的答案标注。

### 3. 问题定义

假设  $P = \{w_i^p\}_1^n$  表示阅读材料的词语序列， $Q = \{w_i^q\}_1^m$  表示问题的词语序列， $m$  和  $n$  分别表示问题和阅读材料中词语的个数。简答题答案标注的目的是从  $P$  中找到多个关于问题  $Q$  的答案文本片段  $A = \{a_i\}_0^v$ ，即  $f(P, Q) \rightarrow A$ 。其中  $v \geq 0$  表示答案片段的个数， $a_i = \{w_i^p\}_{s_i}^{e_i}$ ， $s_i$  表示第  $i$  个答案在  $P$  中的开始位置和结束位置。论文将简答题的答案标注看作是序列标注问题。对每个  $w_i^p$  进行分类，得到每个词的类别  $L_i^p \in \{B, I, O\}$ 。其中  $B$  表示该词语为一个答案文本片段的开始， $I$  表示该词语为答案文本的中间词， $O$  表示该词语为非答案文本。

## 4. 模型构建

### 4.1. 模型结构

模型结构如图 2 所示。模型首先通过语义编码获得阅读材料及问题的语义表示，然后通过问句焦点词识别判断一个词是否为焦点词，接着通过动态网络，根据判断结果采用不同的关注策略进行问题和阅读材料的语义交互，并通过 Self attention 融合上下文信息，最后通过 CRF 层进行答案标注。

### 4.2. 阅读材料及问题语义编码

首先使用 word2vec 将阅读材料  $P = \{w_i^p\}_1^n$  及问句  $Q = \{w_i^q\}_1^m$  中的词语表示为词向量的形式  $\{e_i^p\}_1^n$  和  $\{e_i^q\}_1^m$ ，然后将这些词向量分别输入  $\text{BiLSTM}_P$  和  $\text{BiLSTM}_Q$  获得融合了上下文语义信息的向量表示  $\{u_i^p\}_1^n$  和  $\{u_i^q\}_1^m$ ， $u_i^p = [\bar{u}_i^p, \bar{u}_i^p]$ ， $u_i^q = [\bar{u}_i^q, \bar{u}_i^q]$ 。  $u_i^p, u_i^q \in R^{2 \times d}$ 。接着将  $[u_s^s, u_e^s]$  作为阅读材料中每个句子的语义表示，其中  $u_s^s$ 、 $u_e^s$  表示阅读材料中第  $i$  个句子的开始和结束词语对应的语义表示。通过  $\text{BiLSTM}_S$  获

得融合上下文的新的句子语义表示  $S_i^P = [\bar{s}_i^P, \bar{s}_i^P] \in R^{2d}$ 。最后将  $\bar{u}_i^P = [u_i^P, S_i^P] \in R^{4d}$  作为阅读材料中第  $t$  个词语的语义表示，其中第  $t$  个词语属于第  $i$  个句子。

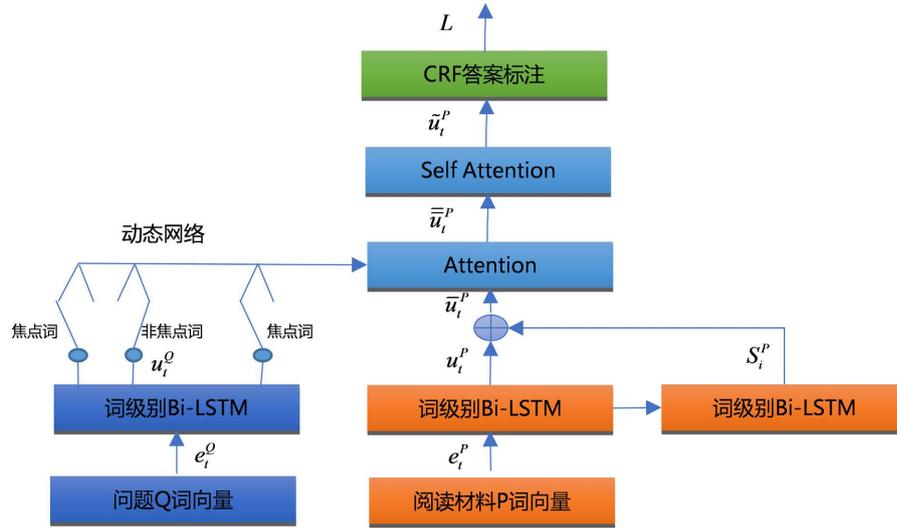


Figure 2. Answer annotation model based on dynamic network  
图 2. 基于动态网络的答案标注模型

### 4.3. 问句焦点词识别

由于问句的焦点词对于寻找答案至关重要，然而相较于其他词语，焦点词往往很少出现在真正的答案片段中。这就导致了在使用关注机制时，如果使用和其它词相同的计算方法，会导致焦点词获得相较于其他非焦点词更低的关注值。将首先通过(1)式判断问句中每个词语是否为焦点词，其中  $w_c \in R^{1 \times d}$ 、 $b_c \in R$  为参数， $\sigma$  为 sigmoid 函数， $p_t$  表示问句中第  $t$  个词为问焦点词的概率。 $p_t \geq 0.5$  被视为焦点词，否则将被视为非焦点词，

$$p_t = \sigma(w_c u_t^O + b_c) \tag{1}$$

### 4.4. 基于动态网络及关注机制的问句信息融合

对于阅读材料中的每个词语，为了准确判断其是否为答案内容，需要在通过语义编码后获得的  $\bar{u}_i^P$  基础上，进一步通过动态网络和关注机制融合问句信息。

首先将通过(2)式计算阅读材料与焦点词的关注值  $\text{Att}(\bar{u}_j^P, u_i^{OF})$ ，与非焦点词的关注值  $\text{Att}(\bar{u}_j^P, u_i^{ON})$  则通过(3)式计算，其中  $W_1^F, W_1^N \in R^{k \times 2d}$ ， $W_2^F, W_2^N \in R^{k \times 4d}$  为参数。

$$A_{ij}^F = \text{Att}(\bar{u}_j^P, u_i^{OF}) = \text{ReLU}(W_1^F u_i^{OF})^T * \text{ReLU}(W_2^F \bar{u}_j^P) \tag{2}$$

$$A_{ij}^N = \text{Att}(\bar{u}_j^P, u_i^{ON}) = \text{ReLU}(W_1^N u_i^{ON})^T * \text{ReLU}(W_2^N \bar{u}_j^P) \tag{3}$$

然后通过(4)和(5)式得到融合焦点词语义表示  $\bar{u}_i^{PF}$ ，通过(6)和(7)得到融合非焦点词信息后的  $\bar{u}_i^{PN}$ ，其中  $m_1$  和  $m_2$  分别为问句中焦点词和非焦点词的个数。 $A_{m_1+1}^F = \text{Att}(\bar{u}_j^P, u_{\emptyset}^{OF})$ ， $A_{m_2+1}^N = \text{Att}(\bar{u}_j^P, u_{\emptyset}^{ON})$ ，其中  $u_{\emptyset}^{OF}, u_{\emptyset}^{ON} \in R^{2d}$  为参数。如果  $\bar{u}_j^P$  与问句中所有焦点词或所有非焦点词都无关， $\bar{u}_j^P$  将更多关注  $u_{\emptyset}^{OF}$  和  $u_{\emptyset}^{ON}$ 。 $[\bar{u}_i^{PF}, \bar{u}_i^{PN}]$  将被送入  $\text{BiLSTM}_A$  得到融合问句信息语义表示  $\bar{u}_i^P = [\bar{\bar{u}}_i^P, \bar{\bar{u}}_i^P] \in R^{2d}$

$$a_j^F = \text{softmax}(A_j^F) \quad (4)$$

$$\bar{u}_i^{PF} = \sum_{i=1}^{m_1+1} a_{it}^F * u_i^{OF} \quad (5)$$

$$a_j^N = \text{softmax}(A_j^N) \quad (6)$$

$$\bar{u}_i^{PN} = \sum_{i=1}^{m_2+1} a_{it}^N * u_i^{QN} \quad (7)$$

接着在得到  $\bar{u}_i^P$  的基础上, 将通过 self-attention 对整个阅读材料中的内容进行进一步融合得到  $\hat{u}_i^P \in R^{2d}$ , 计算过程如(8)式所示。 $\hat{u}_i^P$  将被送入 BiLSTM<sub>A</sub> 得到  $\tilde{u}_i^P = [\tilde{u}_i^P, \tilde{u}_i^P] \in R^{2d}$ 。答案标注模型将根据  $\tilde{u}_i^P$  进行答案标注。

$$\begin{aligned} SA_{ij}^P &= \text{ReLU}(W^P \hat{u}_i^P)^T * \text{ReLU}(W^P \hat{u}_j^P) \\ a_j^P &= \text{softmax}(SA_j^P) \\ \hat{u}_i^P &= \sum_{i=1}^n a_{it}^P * \tilde{u}_i^P \end{aligned} \quad (8)$$

#### 4.5. 答案标注

考虑到词语的答案标注前后依赖, 将采用 CRF 进行答案标注。首先  $\hat{u}_i^P$  通过(9)得到每个词属于  $\{B, I, O\}$  的概率  $P_i \in R^3$ ,  $|P_i|=1$ ,  $W_p \in R^{3 \times d}$  为参数。

$$P_i = \text{softmax}(W_p \hat{u}_i^P) \quad (9)$$

然后通过使得(10)式最大对各个参数进行求解, 其中  $L$  表示训练数据中阅读材料  $P$  和问题  $Q$  对应的答案,  $L(P, Q)$  则表示他们所有可能的答案集合。 $L_t$  表示阅读材料中第  $t$  个词语对应的类别,  $M$  为转移概率矩阵,  $M_{ij}$  表示类别  $i$  到类别  $j$  的转移概率。

$$\begin{aligned} \log P(L | P, Q) &= S(L, P, Q) - \log \left( \sum_{l \in L(P, Q)} e^{S(l, P, Q)} \right) \\ S(L, P, Q) &= \sum_{t=0}^n M_{L_t L_{t+1}} + \sum_{t=0}^n P_{t, L_t} \end{aligned} \quad (10)$$

### 5. 模型训练

由于模型结构受焦点词识别结果的影响, 无法使用传统的梯度下降/上升算法求解, 因此使用强化学习方法, 对(11)式对应的期望值进行优化, 实现参数的求解。其中  $C$  表示  $Q$  的焦点词识别结果,  $C = [c_1, c_2, \dots, c_n]$ ,  $c_i \in \{0, 1\}$ ,  $P(C | Q) = \prod_i P_i$ ,  $\theta$  表示模型的所有参数。对(11)的梯度计算结果如(12)式所示, 其中  $L'$  表示基于焦点词识别结果  $C$  进行的答案标注结果。 $R(L, L' | C)$  为计算结果  $L'$  对应的奖励, 计算过程为(13)式所示。

$$J(\theta) = -E_{C \sim P(C|Q)}(\log P(L | P, Q, C)) \quad (11)$$

$$\begin{aligned} \nabla_{\theta} J(\theta) &= -E_{C \sim P(C|Q)} [\log P(L | P, Q, C) \nabla_{\theta} \log P(C | Q) + \nabla_{\theta} \log P(L | P, Q, C)] \\ &\approx -E_{C \sim P(C|Q)} [R(L, L' | C) \nabla_{\theta} \log P(C | Q) + \nabla_{\theta} \log P(L | P, Q, C)] \end{aligned} \quad (12)$$

$$R(L, L' | C) = \begin{cases} 2, & \text{if } L = L' \\ f1(L, L'), & \text{else if } L \cup L' \neq \emptyset \\ -1, & \text{else} \end{cases} \quad (13)$$

模型的训练过程如下所示：

---

训练算法

---

输入：训练数据集 $\langle L, P, Q \rangle$   
 输出： $\theta$   
 初始化：使用 SQuAD 2.0 预训练  $\theta$   
 for 对训练数据中的每个问题  $Q$  执行：  
 根据  $P(C|Q)$  为问题  $Q$  随机采样一个焦点词识别结果  $C$ ；  
 根据  $C$  以及模型得到答案标注结果  $L'$   
 根据  $L'$  计算  $R(L, L'|C)$   
 通过  $R(L, L'|C) \nabla_{\theta} \log P(C|Q)$  更新焦点词识别对应参数  
 通过  $\nabla_{\theta} \log P(L|P, Q, C)$  更新模型其它部分参数  
 End for

---

## 6. 实验

### 6.1. 数据集及评价标准

由于当前的公开数据集无法直接应用于简答题答案标注，论文自行构建了训练和测试数据集。数据来源于多个英文新闻网站关于在政治、经济、军事等 1000 个新闻片段，并人工为每个片段标注了 3 个问题和对应的答案。数据集划分及答案的个数分布如表 1 所示。对于结果采用 EM 和 F1 作为评价标准。

**Table 1.** Datasets abstract

**表 1.** 数据集概要

	训练数据集	测试数据集
阅读材料个数	800	200
问题个数	2400	600
2~3 个答案的问题数	1153	311
4~5 个答案的问题数	952	256
6 个及以上答案的问题数	285	33

### 6.2. 对比实验

参与对比的方法包括 DCN [14]、BiDAF [1]、R-NET [11]。由于这些方法针对的都是答案只有 1 个的情况，因此论文将这些模型的答案标注层统一换成了 CRF。此外为了保证公平比较，论文将这些模型除 CRF 外部分的参数和结构固定，然后在本数据集上进行了重新训练。实验结果如表 2 所示，结果表明所提出方法在 F1 和 EM 两个指标上效果都有所提升，特别是在 F1 上提升幅度较大。

**Table 2.** Comparison of experimental results

**表 2.** 实验结果对比

	F1	EM
DCN	74.9	65.1
BiDAF	75.3	65.9
R-NET	76.5	66.8
提出的基于动态网络的方法	<b>78.9</b>	<b>68.4</b>

### 6.3. 消融实验

论文于 R-NET 的不同之处主要在于在对阅读材料进行语义编码, 以及在焦点词识别两个方面。为了证明方法的有效性, 针对这两个环节进行了消融实验。在不进行焦点词识别时, 采用了传统的梯度下降算法进行模型求解。实验结果如表 3 所示。实验结果表明两个方面的举措都能够有效提升模型的效果。

**Table 3.** Ablation experimental results

**表 3.** 消融实验结果

	F1	EM
去除句子信息编码	78.4	67.9
去除焦点词识别部分	76.8	67.1

## 7. 下一步工作

在阅读理解中, 简答题的答案往往有多个, 论文针对简答题答案标注方法开展研究, 提出了基于动态网络的简答题答案标注方法, 在现有阅读理解问答模型框架的基础上, 从答案标注、阅读材料语义建模、问句内容解析方面进行了改进。识别效果取得了较大的提升。

由于简答题的答案形式多种多样, 内容有长有段, 有些答案则是阅读材料中内容的总结。目前由于训练语料规模有限, 答案的长度也都比较短, 后续将进一步扩充语料库, 针对多样化的答案抽取生成方法开展研究。

## 基金项目

感谢安徽省高校自然科学研究重点项目“向简答题的阅读理解问答关键技术研究”No. KJ2019A0064, 安徽工业大学校青年科学基金项目“跨语言话题发现与摘要方法研究”, No. QZ201717 资助。

## 参考文献

- [1] Seo, M., Kembhavi, A., Farhadi, A., *et al.* (2016) Bidirectional Attention Flow for Machine Comprehension. *Proceedings of the International Conference on Learning Representations*, Toulon, 24-26 April 2017, 1-13.
- [2] Pennington, J., Socher, R., Manning, C.D., *et al.* (2014) Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, October 2014, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- [3] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, October 2014, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [4] Srivastava, R.K., Greff, K., Schmidhuber, J., *et al.* (2015) Highway Networks.
- [5] Chen, D.Q., Fisch, A., Weston, J., *et al.* (2017) Reading Wikipedia to Answer Open-Domain Questions. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 1870-1879. <https://doi.org/10.18653/v1/P17-1171>
- [6] Weissenborn, D., Wiese, G., Seiffe, L., *et al.* (2016) Making Neural QA as Simple as Possible but Not Simpler. *Proceedings of the 21st Conference on Computational Natural Language Learning*, Vancouver, 3-4 August 2017, 271-280. <https://doi.org/10.18653/v1/K17-1028>
- [7] Wang, H.H. and Jiang, J. (2016) Machine Comprehension Using Match-LSTM and Answer Pointer.
- [8] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [9] Choi, E., Hewlett, D., Uszkoreit, J., *et al.* (2017) Coarse-to-Fine Question Answering for Long Documents. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 209-220. <https://doi.org/10.18653/v1/P17-1020>

- 
- [10] Cho, K., Van Merriënboe, B., Gulcehre, C., *et al.* (2016) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 1724-1734. <https://doi.org/10.3115/v1/D14-1179>
- [11] Wang, W.H., Yang, N., Wei, F.R., *et al.* (2017) Gated Self-Matching Networks for Reading Comprehension and Question Answering. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 189-198. <https://doi.org/10.18653/v1/P17-1018>
- [12] Huang, H.-Y., Zhu, C.G., Shen, Y.L., *et al.* (2017) FusionNet: Fusing via Fully-Aware Attention with Application to Machine Comprehension. *Proceedings of the International Conference on Learning Representations*, Vancouver, 30 April-3 May 2018, 1-20.
- [13] Tay, Y., Tuan, L.A. and Hui, S.C. (2018) Multi-Range Reasoning for Machine Comprehension.
- [14] Xiong, C.M., Zhong, V. and Socher, R. (2016) Dynamic Coattention Networks for Question Answering. *International Conference on Learning Representations 2017*, Toulon, 24-26 April 2017, 1-14.
- [15] Cui, Y., Chen, Z., Wei, S., *et al.* (2017) Attention-over-Attention Neural Networks for Reading Comprehension. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 593-602. <https://doi.org/10.18653/v1/P17-1055>
- [16] Wang, Z.G., Mi, H.T., Hamza, W. and Florian, R. (2016) Multi-Perspective Context Matching for Machine Comprehension.
- [17] Vinyals, O., Fortunato, M. and Jaitly, N. (2015) Pointer Networks. *Proceedings of the Conference on Advances in Neural Information Processing Systems*, Volume 2, 2692-2700.
- [18] Rajpurkar, P., Zhang, J., Lopyrev, K., *et al.* (2016) SQuAD: 100,000+ Questions for Machine Comprehension of Text. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, 1-5 November 2016, 2383-2392. <https://doi.org/10.18653/v1/D16-1264>
- [19] Kocisky, T., Schwarz, J., Blunsom, P., *et al.* (2017) The NarrativeQA Reading Comprehension Challenge.
- [20] Nguyen, T., Rosenberg, M., Song, X., *et al.* (2016) MS MARCO: A Human Generated Machine Reading Comprehension Dataset.