

文章编号 : 1003-0077 (2017) 05-0156-07

基于上下文的深度语义句子检索模型

范意兴<sup>1,2</sup>, 郭嘉丰<sup>1</sup>, 兰艳艳<sup>1</sup>, 徐 君<sup>1</sup>, 程学旗<sup>1</sup>

(1. 中国科学院网络数据科学与技术重点实验室 中国科学院 计算技术研究所, 北京 100190;  
2. 中国科学院大学, 北京 100190)

**摘 要 :** 传统的信息检索的研究多集中在文档级的检索场景中, 然而, 句子级的检索在如移动应用以及信息需求更加明确的检索场景下具有非常重要的意义。在句子级的检索场景下, 我们认为句子的上下文能够提供更加丰富的语义信息来支撑句子与查询的匹配, 基于此, 该文提出了一个基于句子上下文的深度语义句子检索模型 (context-aware deep sentence matching model, CDSMM)。具体的, 我们使用双向循环神经网络来建模句子内部以及句子上下文的语义信息, 基于句子和查询的语义信息得到它们的匹配程度, 在 WebAP 句子检索数据集上的实验表明, 我们的模型性能显著地优于其他的方法, 并取得了目前最好的效果。

**关键词 :** 信息检索; 文本匹配; 循环神经网络

**中图分类号 :** TP391      **文献标识码 :** A

A Context-aware Deep Sentence Matching Model

FAN Yixing<sup>1,2</sup>, GUO Jiafeng<sup>1</sup>, LAN Yanyan<sup>1</sup>, XU Jun<sup>1</sup>, CHENG Xueqi<sup>1</sup>

1. CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;  
2. University of Chinese Academy of Sciences, Beijing 100190, China)

**Abstract :** Traditional researches on information retrieval are focuse on document-level retrieval, neglecting, sentence-level information retrieval which is of great importance in such applications, as searching in mobile phone. Assuming that the context sentence could provide richer evidence for matching, this paper proposes a context-aware deep sentence matching model (CDSMM). Specifically, the model employs bi-directional LSTM to capture the interior and exterior information of the sentence; Then, a matching matrix is constructed based on the sentence representation and query representation; Finally, we get the matching score after a feed forward neural network. Experiment results on the WebAP dataset show that out model can significantly out-perform the state-of-the-art models.

**Key words :** information retrieval; text matching; RNN

1 引言

网页信息检索作为人们获取信息的一个重要渠道, 已经得到了研究人员深入且广泛的研究。传统的信息检索系统通常对用户提交的查询返回多个相关的网页, 根据网页内容与查询的相近程度将网页排序并呈现给用户。近年来, 随着移动终端如智能

手机的大量普及, 人们通过移动设备进行信息查询的需求也不断加大, 相对于电脑而言, 移动设备屏幕大小的限制使得传统以网页作为结果的检索方式变得越来越不能满足人们的需求。同时, 不同的查询结果可以是不同粒度大小的文本内容, 比如, 事实型的查询结果可能是一个简单命名实体, 而导航类的查询结果则可能是一个完整的网页。在实际的检索场景中, 存在大量的语义丰富的长查询(查询词个数

超过 5 个), 相比其他关键词类型的短查询, 这类查询具有更加明确的信息需求, 能够通过一个相对短的句子或者段落来回答。基于句子级别的检索不仅能解决移动设备屏幕大小的限制问题, 同时针对语义丰富的长查询能返回更加精确的答案。

近年来, 深度学习技术在文本匹配相关领域取得了显著的进步, 例如自动问答<sup>[1]</sup>, 复述问题<sup>[2]</sup>, 信息检索<sup>[3]</sup>等。在自动问答与复述问题等任务中, 深度学习技术建模句子间匹配的能力已经得到研究人员的验证<sup>[1,3]</sup>; 在信息检索中, 已有的模型大都是对整个文档内容与查询的相关性进行建模, 对于信息需求不是十分明确的短查询而言, 完整的文档内容能提供给用户更加满意的答案, 基于文档级的检索是比较合适的; 然而对于信息需求更加明确的长查询, 相关的信息经常出现在一个或是多个简短的句子中。本文中, 我们认为对于信息需求更加明确的查询, 一个文档段落或者是多个连续的句子比整个文档作为查询结果更加合适。在此假设的基础上, 我们提出了一个基于上下文句子的深度句子匹配的模型, 具体的, 我们使用双向循环神经网络来建模每个句子内部以及上下文句子的语义信息, 对于查询同样使用循环神经网络 (recurrent neural network, RNN) 得到其语义信息, 在此基础上, 我们构建了一个查询与句子的匹配交互矩阵, 经过多层感知机网络得到该句子与查询的相关程度。我们在 WebAP 数据集上进行了实验, 结果表明, 使用 RNN 能够很好地捕捉句子上下文的语义信息, 我们的方法取得了更高的准确度和稳定性。

本文的主要贡献是我们提出了一个基于上下文句子的深度句子检索模型, 基于双向循环神经网络有效地将上下文句子的语义信息整合到当前句子中, 从而更好地建模了句子与查询的匹配。同时, 我们在基于句子级的检索任务上进行了详细的实验, 我们的模型比当前最佳模型的效果有显著提升, 我们也对此进行了详细的分析。

## 2 相关工作

本文相关工作主要包括社区问答任务、事实型问答任务以及基于段落的信息检索。接下来将对这几个相关的工作进行详细的介绍。

社区问答任务是针对一个给定的问题, 将所有候选的答案进行排序, 返回相关的答案作为结果。Surdeanu<sup>[4]</sup>等人从 Yahoo! Answers 提取了一个较

大的问答任务的数据集, 同时, 他们综合了一些相似特征以及翻译特征并用学习排序模型来解决这个问题, 取得了不错的结果。Molino<sup>[5]</sup>等人发现基于词向量构建的具有丰富语义信息的特征能够显著提高问答任务的准确度。Jansen<sup>[6]</sup>等人整合了词汇的语义信息以及两类语篇信息作为特征, 提出了一个对结果进行重排的模型。社区问答任务与基于句子的检索任务还是存在一定的差异: 句子检索任务的候选答案通常分布在文档的不同位置甚至多个文档中, 且候选的文档集数量很大, 而社区问答任务的候选答案集合相对较小。

事实型问答任务的特点是其答案往往是某个具体的对象, 如人名或者地名等命名实体。事实型问答任务已经吸引了大量的研究人员关注<sup>[7-9]</sup>, Yao<sup>[8]</sup>等人基于问题与候选答案构建了一个线性链条件随机场模型来学习问题与答案的关联, 将答案的抽取问题变成一个序列标注问题。Yu<sup>[9]</sup>等人基于分布式表达的思想, 将问题与答案都映射到一个分布式的语义表达中, 然后基于二者的表达来学习问题与答案的匹配程度。本文的基于句子的检索任务与事实型问答任务最大的不同是本文聚焦的是非事实型的查询, 其答案往往是一个或者多个相关的句子。

基于段落的检索最早被用来提高基于关键词检索的性能<sup>[10-11]</sup>, Liu<sup>[11]</sup>等人将段落的信息用来进行查询扩展, 从而提高检索的精度。Lv<sup>[12]</sup>等人提出一种位置语言模型, 在文档的每个位置构建了一个语言模型, 每个位置的得分以某种衰减的方式传播到其他的位置, 该模型可以看作是一种软段落的检索。基于段落的检索能有效的提高检索的性能, 相比基于文档的检索, 基于段落的检索还能给出文档中具体匹配的信息。一个段落一般是由多个连续的句子组成, 因此基于段落的检索是介于基于文档的检索与基于句子检索的中间状态。基于句子级的检索近年来也得到了研究人员的关注<sup>[13-15]</sup>, Keikha<sup>[14]</sup>等人 Trec Gov2 的数据集的基础上, 标注了一份非事实型的查询数据, 对每一篇文档的每一个句子进行了相关性的标注, 同时他们还对已有的方法在该数据集上的效果进行了检验, 发现已有的检索模型在该数据集上的效果并不是很好。Yang<sup>[13]</sup>等人提出新的基于语义以及上下文的特征, 利用学习排序模型, 在 Keikha 提出的数据集上取得了比以往更好的实验结果。本文的工作是上述两个工作在基于句子检索任务的延续, 利用循环神经网络更好的建模问题与句子的匹配信息, 避免了人工构建特征, 并且取

得了更好的实验结果。

3 模型

循环神经网络在句子等序列数据的建模上具有独特的优势,近期研究人员发现使用双向的 RNN 比仅使用一个方向的 RNN 能捕捉到更多的序列模式信息<sup>[1]</sup>、更好的建模句子的语义信息。因此,我们使用了双向的循环神经网络来建模查询和句子的语义信息,基于此,我们提出了一个基于上下文的深度句子

匹配模型 CDSMM (context-aware deep sentence match model)。如图 1 所示,模型包含了两个部分:首先是查询与句子的建模部分,对于查询直接使用双向的神经网络模型进行建模,对于句子则使用的是基于上下文句子的双向循环神经网络模型进行建模,基于得到的查询和句子中每个位置的元素的交互,构建了一个交互矩阵  $M$ ;然后在  $M$  中,我们使用了位置相关的动态  $k$  最大池化(dynamic k-max pooling)方法得到显著的  $k$  个信号,并通过一个多层感知机将上一步得到的  $k$  个信号综合成最终的匹配得分。

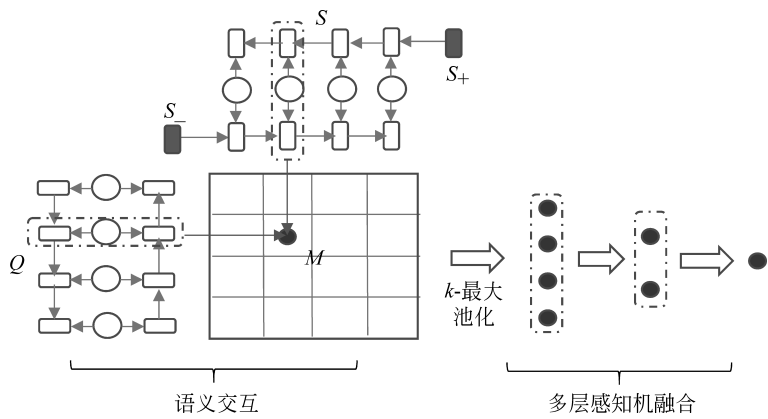


图 1 基于上下文的深度句子匹配模型

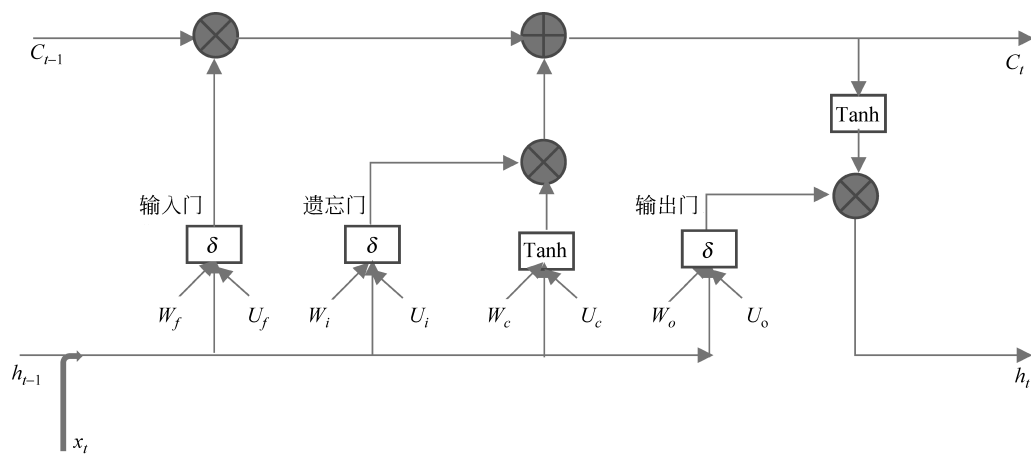


图 2 LSTM 神经元

3.1 第一步: 语义交互

由于查询和句子自然存在序列依赖关系,因此我们选用了双向循环神经网络(Bi-RNN))来建模,但 RNN 的训练存在梯度消失与爆炸的问题,为了解决这个问题,研究人员提出了许多解决方案,最常用的就是使用长短时记忆神经元(long short term memory, LSTM),如图 2<sup>①</sup> 所示,LSTM 特殊的门

机制能够捕捉长距离和短距离的依赖,并且能动态地学习不同位置的不同的依赖模式,充分地建模各个位置的语义信息。具体的,给定一个序列  $S = \{x_0, x_1 \cdots, x_T\}$ ,其中  $x_t$  是位置  $t$  上的单词的表达,LSTM 通过式(1)~(5)来生成每个位置上的表达  $h_t$ :

$$i_t = \delta(W_{xi} x_t + U_{hi} h_{t-1} + b_i) \tag{1}$$

$$f_t = \delta(W_{xf} x_t + U_{hf} h_{t-1} + b_f) \tag{2}$$

① 该 LSTM 神经元的图引自 [www.colah.github.io/posts/2015-08-Understanding-LSTMs/](http://www.colah.github.io/posts/2015-08-Understanding-LSTMs/)

$$c_t = f_t \odot c_{t-1} + i_t \odot u_t \quad (3)$$

$$o_t = \delta(W_{xo} x_t + U_{xo} h_{t-1} + b_o) \quad (4)$$

$$h_t = o_t \odot \tanh(c_t) \quad (5)$$

其中,  $W, U, b$  是参数。  $i$  是输入门,  $f$  是遗忘门,  $o$  是输出门,  $\sigma$  是 sigmoid 激活函数,  $\tanh$  是双曲正切激活函数,  $\odot$  是向量的对应元素相乘的操作。  $h_{t-1}$  为上一个位置的输出,  $c$  为内部存储信息的神经元。对于每个查询  $Q$ , 直接使用双向的循环神经网络分别从左到右和从右到左读入序列, 在每个查询词位置得到两个表达 (分别为  $\overleftarrow{h}_t$  和  $\overrightarrow{h}_t$ )。对于每个候选的句子, 通常情况下, 该句子本身包含的信息不足以回答一个查询, 因此我们选用句子的上下文的句子作为补充信息。具体的, 我们在对当前句子使用双向的循环神经网络建模时, 当从左往右对句子扫描时, 我们将前一个句子的右向输出作为初始的输入; 而当从右往左进行扫描时, 将下一个句子的左向输出作为初始的输入。这样, 我们就能自然地将句子上下文的信息整合到当前的句子表达中来。同样, 在该句子的每个位置我们得到了两个不同的表达。最终, 对于查询和句子中的每个位置, 我们将两个新的表达与原始表达拼接后得到该位置的最终的表达, 具体的, 每个位置  $t$  的表达  $p_t = [\overleftarrow{h}_t, x_t, \overrightarrow{h}_t]^T$ 。

在每个位置表达的基础上, 我们基于每个位置的交互构建整个查询与句子之间的交互。具体的, 我们使用了 4 种不同的交互方式, 它们分别是:

Dot: 计算两个向量的点积,  $s(w, v) = w^T v$ 。

Cosine: 计算两个向量的夹角,  $s(w, v) = \frac{w^T v}{\|w\| * \|v\|}$  其中  $\|\cdot\|$  是  $L2$  范数。

Bilinear: 建模了两个向量不同维度的交互,  $s(w, v) = w^T M v + b$  其中  $M$  是交互矩阵,  $b$  是偏置项。

Tensor: 张量神经网络是一种更加一般性的交互函数<sup>[16]</sup>, 和其他三个不一样的是, 它的输出是一个向量, 如式(6)所示。

$$s(w, v) = f(w^T M^{[1:c]} v + W_w w + W_v v + b) \quad (6)$$

其中,  $M^i, i \in [1, \dots, c]$  是一个矩阵,  $W_w$  和  $W_v$  是线性部分参数,  $b$  为偏置项,  $f$  是非线性函数。

最终, 我们得到了查询  $Q$  与句子的交互矩阵 (或张量)。

### 3.2 第二步: 多层感知机融合

在第一步中, 我们得到了查询  $Q$  与句子  $S$  的交互矩阵 (或张量)。我们使用了动态的  $K$  最大池化<sup>[16]</sup>技术, 从交互矩阵 (或张量) 中得到  $k$  个显著的信号。具体的, 我们根据交互矩阵 (或张量) 的大小,

动态的划分成  $k$  个互不重叠的区域, 然后在每个区域上选取最大的信号组合成  $k$  个显著的信号, 动态  $k$  最大池化是位置相关的, 能更好捕捉查询  $Q$  与句子  $S$  中不同位置的匹配程度。最终, 我们将得到的  $k$  个信号通过多层感知机融合成最终的匹配得分, 如式(7)~(8)所示。

$$r = f(W_r * q + b_r) \quad (7)$$

$$s = W_s + b_s \quad (8)$$

其中,  $W_r$  和  $W_s$  是多层感知机的参数,  $b_r$  和  $b_s$  是对应的偏置项,  $f$  是 Tanh 激活函数。

### 3.3 模型参数及训练

由于本文的实验是一个排序的任务, 我们使用的是基于 pairwise 的 hinge loss 来训练。给定一个三元组  $(q, s^+, s^-)$ , 其中  $q$  是查询,  $s^+$  是比  $s^-$  排序更高的句子。损失函数的定义, 如式(9)所示。

$$L(q, s^+, s^-; \Theta) = \max(0, 1 - S(q, s^+) + S(q, s^-)) \quad (9)$$

其中  $S(q, s)$  是查询  $q$  与句子  $s$  的匹配得分;  $\Theta$  是模型的参数, 包括 LSTM 的参数以及多层感知机的参数。优化方法采用的是标准的后向传播, 使用小批量 (mini-batch) 随机梯度下降方法 Ada-Grad<sup>[17]</sup>, 每个批量的大小设置为 100。同时, 使用了早停止的策略来避免过拟合。LSTM 与多层感知机中的矩阵  $W$  的初始化是以 0 为均值, 0.02 为方差的高斯分布。单词的表达我们选用的是在维基百科语料上预训练的 50 维的词向量, 训练的模型用的是 Mikolov<sup>[18]</sup> 的 Skip-Gram 的模型, 训练参数是作者推荐的默认的参数<sup>①</sup>。

## 4 实验及分析

首先给出本文实验的简单定义, 给定一系列的非事实型的查询  $\{Q, Q_1, \dots, Q_n\}$  以及一系列的文档集合  $\{D_1, D_2, \dots, D_n\}$ , 其中这些文档  $D$  中可能包含某个查询的答案。我们的目标是给定一个查询  $Q$ , 对每个候选的文档  $D$  中的每一个句子进行相关性的判断, 判断句子是否构成该查询答案的一个部分。

### 4.1 数据集与评价指标

本文使用的 WebAP<sup>②</sup> 数据集<sup>[15]</sup>。该数据集是基于 TREC GOV2 的文档以及查询构建的。GOV2

① <https://code.google.com/archive/p/word2vec/>

② 这个数据集是公开可下载的: <http://ciir.cs.umass.edu/downloads/WebAP/>

是从 .gov 的网站上抓取的网页,最早于 2004 年被用到 TREC Terabyte Track 的任务中,包含了 25 000 000 个文档。WebAP 数据集具体的构建过程如下:对于那些存在段落级别的查询(总共是 82 个查询),使用已有的具有最好效果的检索模型<sup>[19-20]</sup>从 GOV2 的文档集中检索出最相关的 50 个文档,在这些文档中,在 TREC 的标注中被标记为相关的文档被挑选出来进行段落级别的标记,每个段落被标记为“完美”,“很好”,“好”,“一般”。最后被标记的查询有 82 个,相应的被标记的段落有 8 027 个,每个查询平均有 97 个段落被标记。在这些被标记的段落中,43% 的段落被标记为完美,44% 的段落被标记为很好,10% 段落被标记为好,剩下的则是一般,其余未被标记的段落则是完全无关。和 Liu<sup>[11]</sup>的工作一样,我们用每个段落的标记信息作为该段落中每个句子的标记,并将原始的“完美”,“很好”,“好”,“一般”映射成[4, 3, 2, 1]的分数,这样的句子的标记存在一个问题,即可能存在一个句子在某一个段落中标记为好,而在另一个段落中标记为一般;针对这个问题,我们通过多数投票的方式来决定这些句子的最终标记。最终我们得到 991 233 个句子,句子的平均长度为 17.58。在这些被标记的句子中,有 99.02% (981 510)个句子标记是 0(也就是不相关),只有少于 1% 的句子是正的标记(149 个句子标记为 1 783 个句子标记是 2 4 283 个句子的标记是 3 4 508 个句子的标记是 4)。

由于数据集样本数量的局限性,我们使用了 5 折交叉验证的方法避免模型的过拟合,具体的,我们将查询切分成 5 个大小相等的集合,然后选择其中的 4 个用来做训练,剩下的一个做测试,如此重复 5 次,最后将 5 个不同集合的结果的平均值作为最终的评价。评价指标方面,我们使用的是信息检索领域常用的评价指标:NDCG@ 10、P@ 10 以及 MRR。

4.2 Baseline

我们选择了三类的模型做对比实验:传统信息检索最好的模型、问答任务类中最新的模型及在 WebAP 数据集上最新的模型。

① 语言模型(LM)<sup>[21]</sup>:语言模型计算的是查询能被句子生成的 log 似然概率,由于句子中词的稀疏性,我们使用了 Dirichlet 分布对单词的似然进行平滑,该方法又被称为查询似然函数,如式(10)所示。

$$f_{LM}(Q, S) = \sum_{w \in Q} \log \frac{tf(w, S) + \mu \frac{P(w|C)}{|S| + \mu}}{|S| + \mu} \quad (10)$$

② 卷积神经网络模型(CNN)<sup>[9]</sup>:卷积神经网络模型被用在 TREC QA Track 的数据集上,并取

得了很好的效果。在本实验中,作者使用了预先训练好的词向量,并加入了词频的特征信息,利用监督学习的方法训练了一个卷积神经网络模型。

③ 多视角深度文本匹配模型(MV-LSTM)<sup>[1]</sup>:该方法使用双向循环神经网络将一段文本表达成多个不同位置的文本表达,然后基于所有位置的文本表达之间进行匹配,并得到最终的匹配得分。我们同样使用了双向循环神经网络来建模句子,不同的是,我们使用了基于上下文句子的信息来增强当前句子的表达,同时在匹配矩阵中我们选用的是基于位置相关的 k 最大池化,而 MV-LSTM 使用的是位置无关的 k 最大池化。

④ 学习排序方法(L2R+CSFeatures)<sup>[13]</sup>:该方法使用了 27 个人工构建的特征,它们分别是:是否精确匹配、重叠词个数、同义词重叠比例、语言模型得分、候选句子长度、句子在原始文档的位置、基于维基百科抽取的语义相关特征、基于词向量的语义特征、实体关联特征,以及前一个句子的所有特征和后一个句子的所有特征,然后使用 MART(multiple regressive trees)将各个特征得分整合成最终的得分。该方法在 WebAP 数据集上取得了最好的结果。

4.3 实验结果

表 1 WebAP 数据集实验结果

模型	NDCG@ 10	P@ 10	MRR
LM	0.134 0	0.145 1	0.339 5
CNN <sup>[13]</sup>	0.059 6	0.064 6	0.125 4
MV-LSTM	0.077 7	0.105 0	0.222 2
L2R+CSFeatures <sup>[13]</sup>	0.186 4	0.202 4	0.451 2
CDSMM	0.256 3	0.230 0	0.580 1

实验结果如表 1 所示。我们的模型(CDSMM)在各个评价指标上都比其他模型的结果要好,具体的,在 Top-1 的评价指标 MRR 上,我们的模型较当前最好的模型提高了 28.57%,而在 Top-10 的评价指标 NDCG@ 10 和 P@ 10 上,我们分别提高了 37.5% 和 13.64%。同时,从表中可以看到,使用了上下文句子信息的模型(L2R+CSFeature 和 CDSMM)效果要远高于没有使用上下文句子信息的模型,这说明对于非事实型的查询,单个句子提供的信息往往不足以给出完整的答案,在借助了上下文句子信息的情况下能给出更好的结果。同时,我们也注意到在没有使用上下文句子信息的情况下,LM 比另外两个基于神经网络的 baseline 方法效果更好,可能的一个原因是在 WebAP 数据集上的查

询都包含丰富的关键词信息,如查询“describe history oil industry”,在这类查询中,精确匹配的信息要远比近似匹配的效用更大,在精确匹配与近似匹配的区分能力上,LM 比 CNN 和 MV-LSTM 的能力更强,因此取得更好的效果。

为了验证我们的模型的鲁棒性,我们对动态  $k$  最大化中的  $k$  取值的结果进行分析,从图 3 中可

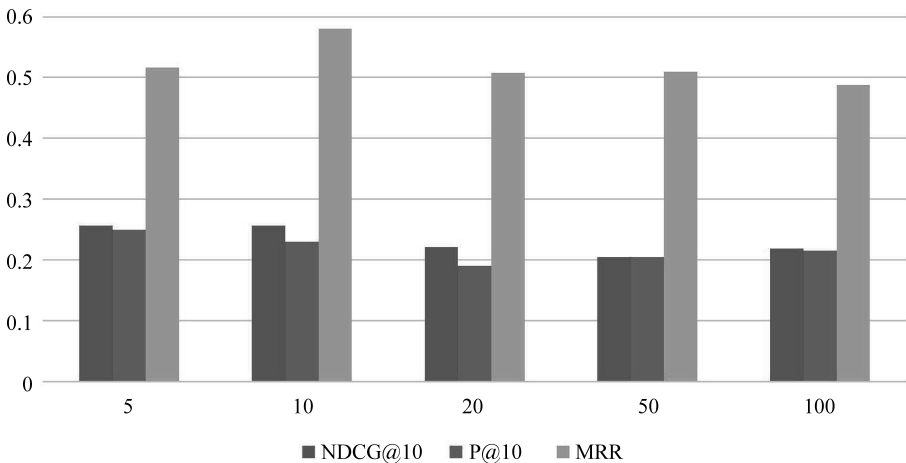


图 3 CDSMM 中不同  $k$  值的结果比较

在查询与句子中词的交互中,我们使用了 4 种不同的交互方式,实验结果如表 2 所示。在表 2 中,我们对这四种不同的交互性能进行了验证。从表中可以看出, dot 交互对应的结果非常不好,主要原因是在 dot 交互时,词之间的相似度值受词向量的长度影响太大,在词与词之间的相似度度量上存在太大偏差。同时,我们观察到基于 cos 的交互方式比 bilinear 以及 tensor 的效果要好。理论上, bilinear 比 cos 捕捉向量交互的能力更强,而 tensor 则更具有一般性,然而在 WebAP 的数据集上结果并没表现更好,主要是因为数据集大小的限制(只有 82 个查询),而 bilinear 和 tensor 的参数数量远大于 cos,导致它们出现了过拟合现象。

表 2 不同交互方式的实验结果

	NDCG@ 10	P@ 10	MRR
CDSMM-dot	0.083 1	0.060 0	0.174 7
CDSMM-cos	0.256 3	0.230 0	0.580 1
CDSMM-bilinear	0.174 6	0.160 0	0.500 0
CDSMM-tensor	0.144 4	0.120 0	0.440 7

5 总结与未来工作

本文提出了一种基于句子上下文的深度句子检索模型,使用了双向循环神经网络来建模句子内部

以看出,不同  $k$  值的选择对结果的影响并不是很大,当  $k$  值等于 100 时,结果略微有所下降,主要是因为数据集中查询  $n$  的平均长度为 6,而平均的句子长度为 15,因此,对于大部分的查询而言,太大的  $k$  值将会引入额外的噪声信息,从而会影响实验的性能。同时,我们观察到当  $k$  取值为 10 时取得最好的结果。

以及句子上下文的语义信息,更加充分地捕捉了查询与候选句子的匹配信息,我们的模型在句子级的检索任务上取得了目前最好的结果。然而,由于标注语料大小的限制,更加复杂的交互模式如双线性及张量神经网络并没取得更好的效果,如何在缺乏足够标注数据集的情况下学习一个足够强大的模型将是我们后续需要继续研究的工作。同时,目前我们的模型在非事实型的查询数据上取得很好结果,在其他类型的查询如事实型的查询数据上的能力需要进一步验证。

参考文献

[1] Shengxian Wan, Y Lan J. Guo, et al. A Deep Architecture for Semantic Matching with Multiple Positional Sentence Representations[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Palo Alto, 2016.

[2] Dolan W B, Brockett C. Automatically constructing a corpus of sentential paraphrases [C]//Proceedings of the 3rd International Workshop on Paraphrasing, Sydney, 2005.

[3] Huang, Po-Sen, et al. Learning deep structured semantic models for web search using click through data [C]//Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management. San Francisco, 2013.

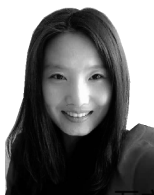
- [4] M. Surdeanu, M. Ciaramita, and H. Zaragoza. Learning to Rank Answers on Large Online QA Collections [C]//Proceedings of the 46th Annual Meeting of the ACL. Columbus, 2008;719-727.
- [5] P Molino, L M Aiello. Distributed Representations for Semantic Matching in Non-factoid Question Answering [C]//Proceedings of the 37th Annual ACM SIGIR Conference. Queensland, 2014;38-45.
- [6] P Jansen, M Surdeanu, P Clark. Discourse Complements Lexical Semantics for Non-factoid Answering Reranking[C]//Proceedings of the 52nd Annual Meeting of the ACL. Baltimore, 2014;977-986.
- [7] W. tau Yih, M W Chang, C Meek, et al. Question answering using enhanced lexical semantic models [C]//Proceedings of the 51st Annual Meeting of the ACL. Sofia, 2013.
- [8] X Yao, B V Durme, C Callison-burch, et al. Answer Extraction as Sequence Tagging With Tree Edit Distance[C]//Proceedings of the 11th Conference of the North American Chapter of the Association for Computational Linguistics. Atlanta, 2013.
- [9] Yu Lei, Karl Moritz Hermann, Phil Blunsom, et al. Deep learning for answer sentence selection[J].Computer Science, arXiv preprint arXiv:1412.1632, 2014.
- [10] M Bendersky, O Kurland. Utilizing passage-based language models for document retrieval[C]//Proceedings of the 30th European Conference on Information Retrieval, Glasgow, 2008;162-174.
- [11] X Liu, W Croft. Passage Retrieval Based on Language Models [C]//Proceedings of the 11th ACM Conference on Information and Knowledge Management. McLean, 2002;375-382.
- [12] Lv Yuanhua, Chengxiang Zhai. Positional language models for information retrieval [C]//Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. Boston, 2009.
- [13] Yang Liu, et al. Beyond Factoid QA: Effective Methods for Non-factoid Answer Sentence Retrieval [M]. European Conference on Information Retrieval. Berlin:Springer, 2016.
- [14] M Keikha, J H Park, W B Croft. Evaluating Answer Passages Using Summarization Measures [C]//Proceedings of the 37th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval. Gold Coast, 2014.
- [15] M Keikha, J H Park, W B Croft, et al. Retrieving Passages and Finding Answers [C]//Proceedings of the 19th Australasian Document Computing Symposium. Melbourne, 2014;81-84.
- [16] Qiu Xipeng, Xuanjing Huang. Convolutional neural tensor network architecture for community-based question answering [C]//Proceedings of the 24th International Joint Conference on Artificial Intelligence. Buenos Aires, 2015.
- [17] Duchi John, Elad Hazan, Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization [J]. Journal of Machine Learning Research 2011,12; 2121-2159.
- [18] Omas Mikolov, Ilya Sutskever, Kai Chen, et al. Distributed Representations of Words and Phrases and their Compositionality [C]//Proceedings of the 27th Neural Information Processing Systems. Lake Tahoe, 2013.
- [19] S Huston, W B Croft. A comparison of Retrieval Models Using Term Dependences [C]//Proceedings of the 25th ACM Conference on Information and Knowledge Management. Shanghai, 2014.
- [20] D Metzler, W B Croft. A Markov Random Field Model for Term Dependencies [C]//Proceedings of 28th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval. Salvador, 2005.
- [21] C Zhai, J Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval [C]//Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New Orleans, 2001;334-342.



范意兴(1990—),博士研究生,主要研究领域为信息检索,机器学习,深度学习。  
E-mail: fanyixing@software.ict.ac.cn



郭嘉丰(1980—),博士,研究员,主要研究领域为信息检索,文本挖掘和机器学习。  
E-mail: guojiafeng@ict.ac.cn



兰艳艳(1982—),博士,副研究员,主要研究领域为深度学习,文本挖掘和机器学习。  
E-mail: lanyanyan@ict.ac.cn