

结合多粒度信息的文本匹配融合模型



吕乐宾 刘群 彭露 邓维斌 王崇宇

重庆邮电大学计算智能重庆市重点实验室 重庆 400065

(lebinlv@foxmail.com)

摘要 常规的文本匹配模型大致分为基于表示的文本匹配模型和基于交互的文本匹配模型。由于基于表示的文本匹配模型容易失去语义焦点,而基于交互的文本匹配模型会忽视全局信息,文中提出了结合多粒度信息的文本匹配融合模型。该模型通过交互注意力和表示注意力将两种文本匹配模型进行了融合,然后利用卷积神经网络提取了文本中存在的多个不同粒度的粒度信息,使得模型既能抓住局部的重要信息又能获取全局的语义信息。在3组不同的文本匹配任务上的实验结果表明,所提出的模型在NDCG@5评价指标上分别优于其他最优模型5.3%,0.4%,1.5%。通过提取文本中的多个粒度信息并结合交互注意力和表示注意力,提出的模型能够有效地关注不同级别的文本信息,解决了传统模型在文本匹配过程中易失去语义焦点和忽视全局信息的问题。

关键词 文本匹配;交互注意力;表示注意力;粒度网络;多粒度信息

中图法分类号 TP391

Text Matching Fusion Model Combining Multi-granularity Information

LYU Le-bin, LIU Qun, PENG Lu, DENG Wei-bin and WANG Chong-yu

Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract Conventional text matching methods are basically divided into representational text matching models and interaction-based text matching models. Since the representation-based text matching model is easy to lose semantic focus and the interaction-based text matching model ignores global information, a text matching fusion model combining multi-granularity information is proposed in this paper. This model fuses two text matching models through interactive attention and expressing attention, and then uses convolutional neural networks to extract multiple different levels of granularity information presented in the text. Then the local important information and global semantic information can be captured. The experimental results on three different text matching tasks show that the proposed model outperform other optimal models by 5.3%, 0.4%, 1.5% on the NDCG@5 evaluation index respectively. By extracting multiple granularity information of the text and combining interactive attention and expressed attention, the proposed model can effectively pay attention to the text information of different levels, and solve the problem of losing semantics and ignoring global information during the text matching process in the traditional models.

Keywords Text matching, Interactive attention, Expressive attention, Granular network, Multi-granularity information

1 引言

文本匹配任务是自然语言处理(Natural Language Processing, NLP)中重要的研究方向,无论是在信息检索(Information Retrieval, IR)、问题回答(Question Answer, QA)还是复述识别(Paraphrase Recognition, PR)等任务中都扮演着重要角色。传统的文本匹配方法依赖于预定义的模板和人工提取的规则。随着深度学习的发展,神经网络已经普遍应用于自然语言处理任务中,以节省人工提取特征所耗费的成本和时间。文本匹配任务旨在给定两段文本 Q 和 D , 通过提

取文本中存在的语义信息和相似度特征来给出两段文本的相似度值,由最终的相似度值可以得知两段文本的内容是否属于相似的描述。

基于深度学习的文本匹配模型大致可以分为两类:基于表示的文本匹配模型^[1]和基于交互的文本匹配模型^[2]。基于表示的文本匹配模型通过对文本进行预处理以及构建索引,能够更好地提取每段文本里面的信息,但是在信息表示的过程中容易失去语义焦点,难以衡量词在上下文中的重要性。基于交互的文本匹配模型虽然可以较好地把握语义焦点,针对上下文进行更好的建模,但是却忽视了全局信息,会造成无

到稿日期:2020-07-15 返修日期:2020-08-20 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发计划资助项目(2018YFC0832100, 2018YFC0832102);国家自然科学基金重点项目(61936001)

This work was supported by the National Key Research and Development Program of China(2018YFC0832100, 2018YFC0832102) and Key Program of National Natural Science Foundation of China(61936001).

通信作者:刘群(liuqun@cqupt.edu.cn)

法刻画出全局匹配信息的后果。显然,上述两种不同的模型并非单纯的包含关系,而是一种相互补充的关系。文本表示中往往存在多个粒度的信息,简单的单词匹配会丢失词组信息,抽象的全局表示又存在语义丢失的情况,所以能否捕获句子中多个粒度之间的信息表示是文本匹配的关键。本文提出了一种结合多粒度信息的文本匹配融合模型(Text Matching Fusion Model Combining Multi-granularity Information, FM-MI),该模型首先通过构建特征提取的粒度网络,来提取文本中存在的多个粒度信息,进而利用两种注意力机制将基于交互和基于表示的文本匹配模型进行融合,其通过加强模型对细粒度和粗粒度文本信息的关注,有效地提升了多粒度文本信息匹配的效果。

本文第2节简要介绍了相关工作;第3节详细描述了本文提出的FMMI模型架构;第4节介绍了3个实验数据集并将实验结果进行了对比分析;最后总结全文并展望未来。

2 相关工作

在文本匹配模型中,基于表示的文本匹配模型是利用神经网络将每段文本转换为对应的张量表示,然后将表示的张量进行结合后提取内部的交互信息。而基于交互的文本匹配模型则是利用两段文本直接匹配得到交互矩阵,然后从交互矩阵当中提取有用的特征信息。本节将主要介绍这两类方法的主流模型。

2.1 基于表示的文本匹配模型

基于表示的文本匹配模型的主要研究点包括两个方面:一方面是如何对两段文本进行更好的表示从而获取文本中所蕴含的重要语义信息;另一方面是如何利用得到的两个张量来计算相似度。为了更好地获取张量表示,近几年的学者开始研究利用不同的神经网络来对输入的文本进行转换,其中包括卷积神经网络(Convolutional Neural Network, CNN)和循环神经网络(Recurrent Neural Network, RNN),这两种模型都是将输入文本利用神经网络转换为特定的表示张量,主要区别在于构造表示的过程和计算相似度的方式。

Huang等^[1]提出的DSSM(Deep Structured Semantic Models)模型最早将神经网络应用到文本匹配任务中。该模型利用深度神经网络把Query和Document的词袋模型生成的向量表示为低维度的语义向量,再通过余弦公式来计算两个向量的相似度。与传统的文本匹配模型相比,DSSM模型在NDCG^[3](Normalized Discounted Cumulative Gain)度量上有显著的改进。由于词袋模型生成的向量在进行表示时会丢失来自文本的上下文结构信息,因此,Shen等^[4]利用卷积神经网络代替DSSM模型中的多层感知机得到CDSSM模型。其采用单词序列作为模型的输入,通过词的n-gram和卷积池化操作来获得上下文关系,与DSSM相比,CDSSM的匹配精度得到了进一步的提高。为了使模型能提取长距离上的句子特征,Wan等^[5]提出了一种基于双向长短时记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)的多语义模型MV_LSTM。该模型有效地提取了文本中的重要特征,并在实验中得到了不错的效果。通过考查每个单词在不同语境下的含义,该模型实现了从多粒度角度对句子进行分析。Gong

等^[6]提出了一种基于推理的通用表示模型DIIN。该模型与其他的文本匹配模型结构类似,也是采用卷积神经网络和循环神经网络对文本进行特征提取,但是在模型的输入层添加了词向量和字向量,并加入了一些能够表示句法信息的额外特征。同时,在最后的特征提取层还采用了DenseNet网络,故保留了较完整的原始特征信息,提升了模型的匹配效果。

2.2 基于交互的文本匹配模型

与基于表示的文本匹配模型不同,基于交互的文本匹配模型旨在提取两段文本直接匹配的交互特征信息,其关键在于如何构造两段文本的交互矩阵,以及如何从构造的交互矩阵中有效地提取交互信息。该模型首先要考虑对应关键词的匹配,其次考虑每个词的相对位置,最后综合利用这两个特征来计算原文本的匹配分数。

Hu等^[7]提出的ARC-II模型,首先对两段文本分别利用一维卷积来关注相邻的单词向量,其次将卷积后得到的两个张量进行结合,以便获得描述两个句子之间关联的抽象表示,最后使用多层感知机(Multilayer Perceptrone, MLP)将抽象化的表示转换为匹配分数输出。与之前的DSSM和CDSSM模型相比,ARC-II的表现更好。Pang等^[2]提出了MatchPyramid模型,该模型直接对查询和文档的单词向量进行点积运算来构建交互特征,再通过卷积池化进行特征提取,最后利用全连接网络得出匹配分数。Xiong等^[8]提出的K-NRM模型对两段文本作词嵌入之后,计算两段文本的余弦相似度来得到相似性矩阵,然后利用径向基函数核(Radial Basis Function, RBF kernel)对得到的相似性矩阵进行池化,最后利用全连接层来得到最终的匹配分数。Dai等^[9]改进了K-NRM模型,提出了Conv-KNRM模型。该模型通过卷积神经网络提取两段文本的n-gram嵌入,利用嵌入后的多个特征获取交互张量,再将结果进行拼接,最后利用一个全连接层计算出最终的相似性结果。

3 结合多粒度信息的文本匹配融合模型

本节将详细介绍结合多粒度信息的文本匹配融合模型(FMMI),该模型的详细架构如图1所示。

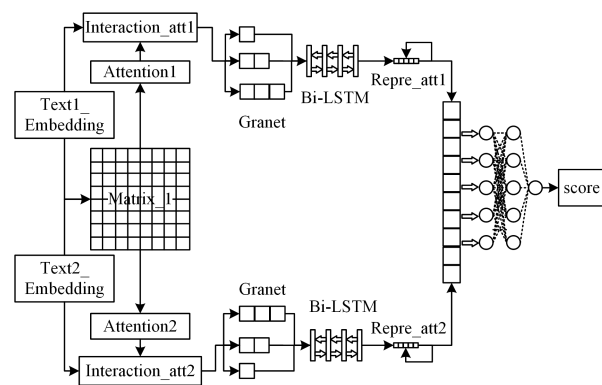


图1 结合多粒度信息的文本匹配融合模型

Fig.1 Matching fusion model combining multi-granularity information

该模型不仅利用交互注意力和表示注意力将基于交互和基于表示的模型进行融合,而且充分利用和提取了文本中存

在的多个粒度信息,使得模型可以在更多地捕获原文中存在的不同层次的信息表示,提高文本匹配的精度。下面将对模型的每个部分进行详细介绍。

3.1 词嵌入

本文提出的 FMMI 模型的输入层是一个嵌入查找函数,该函数利用预先训练好的 GloVe^[10] 词向量得到查找矩阵 $E \in \mathbf{R}^{V \times K}$,将查询和文档的每个单词映射到高维向量空间,其中 V 表示需要嵌入的单词数量, K 表示每个单词嵌入向量的维度。

3.2 交互注意力

模型利用查询和文档中嵌入后的单词向量计算句子中对每对单词之间的相似性,得到相似矩阵 $Matrix_1$,对相似矩阵分别按行和列进行归一化加和处理,分别得到查询和文档的注意力权重,利用注意力权重对原文本进行加权处理得到新的文本表示。加权后的文本在信息提取和文本表示的过程中更容易抓住一段文本的重点部分。该交互注意力机制由 Seo 等^[11] 提出,并且证实了交互注意力在阅读理解任务中的表现优异。在文本匹配过程中,句子中的单词对于匹配任务的贡献度不同,为了在文本表示过程中更好地发挥重要单词的作用,需要对不同的单词赋予不同的权重信息,因此本文将这种注意力机制引入到文本匹配模型中,分别对两段输入文本加入来自另一方的注意力权重。该交互注意力由查询到文档的注意力(Q2D)和文档到查询的注意力(D2Q)组成,向量中的每个权重值间接表示了该文本中的每个单词对另一段文本中全部单词的整体重要程度。交互矩阵和交互注意力的计算方式如下。

假设分别给定文本长度大小为 X 的查询段落 $Q^x = \{Q^1, \dots, Q^x\}$ 和文本长度大小为 Y 的文档段落 $D^y = \{D^1, \dots, D^y\}$,令 M_{xy} 表示查询和文档交互后的相似性矩阵,通过相似矩阵得到注意力张量的过程可以计算如下:

$$M_{xy} = Linear(Q^x \cdot D^y + bias) \quad (1)$$

$$A_{Q2D} = \text{sum}_{col}(\sigma_{col}(M_{xy}) \cdot D^y) \quad (2)$$

$$A_{D2Q} = \text{sum}_{row}(\sigma_{row}(M_{xy}) \cdot Q^x) \quad (3)$$

其中, $bias$ 表示在线性函数后面加的偏置, \cdot 表示张量之间的点积运算, σ 表示 $softmax$ 激活函数, $\text{sum}(\cdot)$ 表示沿着指定的轴计算张量之和, A_{D2Q} 表示得到的查询注意力向量, A_{Q2D} 表示得到的文档注意力向量。将两个注意力向量与原文本进行结合,可以得到新的加权文本表示,计算如下:

$$Q_{ATT} = Q \cdot A_{Q2D} \quad (4)$$

$$D_{ATT} = D \cdot A_{D2Q} \quad (5)$$

其中, Q_{ATT} 和 D_{ATT} 分别表示加权处理后的文本表示。

3.3 多粒度信息提取

ResNeXt 网络^[12] 在图像特征提取中不仅可以解决深度神经网络的退化问题,而且有利于网络的收敛,如图 2(a) 所示。本文将这种残差网络引入到文本匹配任务中对文本信息进行特征提取,如图 2(b) 所示。网络中每一层的三元组信息分别表示输入特征维度、卷积核尺寸及输出特征维度。FMMI 模型利用不同卷积窗口的一维卷积代替残差网络中的二维卷积来对文本进行多粒度特征提取,本文将此特征提取网络称为粒度网络(Granular Network, Granet)。该网络利用 3

种不同的卷积窗口对原文本进行分组卷积,每组卷积窗口会提取到不同粒度的特征表示,使模型可以捕获更多的文本特征信息。多粒度网络的具体计算公式如下:

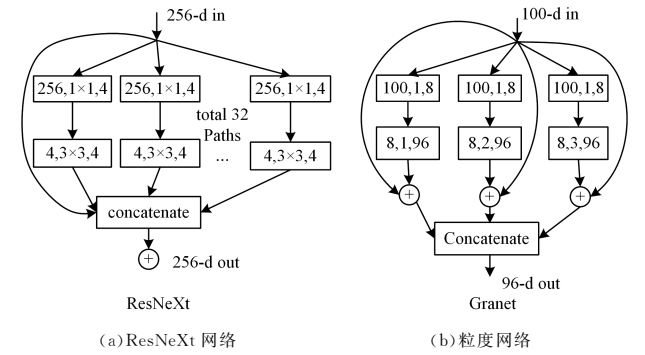


图 2 ResNeXt 网络和粒度网络

Fig. 2 ResNeXt network and Granular network

$$Q_{reduct}^i = \sigma(W_{reduct}^i Q_{ATT} + b^i) \quad (6)$$

$$D_{reduct}^j = \sigma(W_{reduct}^j D_{ATT} + b^j) \quad (7)$$

$$Q_{multi}^i = \sigma(W_{multi}^i Q_{reduct}^i + b^i) \quad (8)$$

$$D_{multi}^j = \sigma(W_{multi}^j D_{reduct}^j + b^j) \quad (9)$$

其中, $i, j \in \{1, 2, 3\}$; $Q_{reduct}^i, D_{reduct}^j$ 表示 Granet 网络第一层的原文本经过信息提取和降维后的张量表示; Q_{multi}^i, D_{multi}^j 分别表示第二层中粒度信息提取和扩大维度后的结果; W_{multi} 表示粒度滑动窗口,随着窗口的滑动,逐步提取文本序列中的粒度信息; σ 表示 $RELU$ 激活函数。模型的最后一步采用了 ResNeXt 网络中的残差连接操作,将特征提取后的信息与原始信息进行连接,计算方式如下:

$$Q_{res}^i = \text{concat}(Q_{multi}^i, Q_{ATT}, axis = -1) \quad (10)$$

$$D_{res}^j = \text{concat}(D_{multi}^j, D_{ATT}, axis = -1) \quad (11)$$

其中, $i, j \in \{1, 2, 3\}$; $\text{concat}(\cdot)$ 表示张量的拼接操作; Q_{res}^i, D_{res}^j 分别表示查询和文档中对每个粒度的文本信息进行残差连接后的表示。最后模型将所有粒度的信息按行进行拼接,得到新的查询和文档的表示张量 Q_{all}, D_{all} ,其公式计算如下:

$$Q_{all} = \text{concat}(Q_{res}^1, Q_{res}^2, Q_{res}^3, axis = -2) \quad (12)$$

$$D_{all} = \text{concat}(D_{res}^1, D_{res}^2, D_{res}^3, axis = -2) \quad (13)$$

3.4 表示层和表示注意力

模型对拼接后的两段表示张量分别采用 Bi-LSTM 对多个粒度的文本特征进行学习表示和维度压缩。Bi-LSTM 能够在高层更抽象地实现对序列的特征表达,使模型的表示可以更好地捕获序列的全局信息,而不是只局限于提取单词或词组之间的相似性特征。因此模型对特征提取后的信息进行表示学习,既可以获取每个粒度内的全局信息,又可以获取粒度之间的全局信息。文本表示的计算公式如下:

$$Q_{rep} = \text{Bi-LSTM}(Q_{all}) \quad (14)$$

$$D_{rep} = \text{Bi-LSTM}(D_{all}) \quad (15)$$

粒度网络提取到的词组信息对匹配任务的重要度不同,重要度低的词组信息对于语义理解来说可能是噪声,所以 FMMI 模型选择在 Bi-LSTM 表示后面加入表示注意力机制,通过全连接层来计算表示信息每个时间步对于全局信息的重要程度,然后将这个重要度作为表示信息的权重值;并且为表示信息加入权重约束,让两段句子交互时可以更好地发挥重

要度高的词组信息的作用,抑制重要度低的词组信息的作用。表示注意力的计算方式如下:

$$Q_{rep_att}^i = \sigma(W^i Q_{rep}^i + b^i) \quad (16)$$

$$D_{rep_att}^j = \sigma(W^j D_{rep}^j + b^j) \quad (17)$$

其中, $i \in \{1, 2, \dots, X\}$, $j \in \{1, 2, \dots, Y\}$; X 和 Y 分别表示查询和文档的长度; Q_{rep}^i , D_{rep}^j 分别表示输入的两段文本的第 i 个和第 j 个时间步的张量表示; W^i , W^j 代表可学习的参数; $\sigma(\cdot)$ 表示 *sigmoid* 激活函数; $Q_{rep_out}^i$, $D_{rep_out}^j$ 分别表示查询和文档第 i 个, 第 j 个时间步的权重值。然后利用权重值对原文本进行加权处理, 得到最终的输出 Q_{out} , D_{out} , 其公式如下所示:

$$Q_{out} = Q_{rep_att} \cdot Q_{rep} \quad (18)$$

$$D_{out} = D_{rep_att} \cdot D_{rep} \quad (19)$$

3.5 交互输出层

模型将表示后的两段文本平铺并在第一维度上进行连接, 然后输入全连接层来计算两段文本的匹配得分。

$$Z_{rep} = \text{concat}(Q_{out}, D_{out}, \text{axis} = -1) \quad (20)$$

$$Score = \sigma(WZ_{rep} + b) \quad (21)$$

其中, Q_{out} , D_{out} 分别是查询和文档加权表示后的序列; $\text{concat}(\cdot)$ 表示拼接函数; Z_{rep} 表示拼接之后的张量; W 表示可学习的参数, $\sigma(\cdot)$ 表示 *Linear* 激活函数; $Score$ 是最终输出的两段文本的匹配分数。

3.6 训练

在训练过程中, 损失函数的计算如下:

$$Loss = \max(0, \text{margin} + y' - y) \quad (22)$$

针对给定的两段输入序列, 最终的正确预测得分 y 和错误预测得分 y' 的差值可用来表示两个预测结果的相似关系, margin 是我们根据具体情况设定的系数。 y 越高, y' 越低, 即 $y - y'$ 的值越大, 代表模型的效果越好, 但二者得分之差最多为 margin , 差距更大并不会获得更多奖励。

4 实验

4.1 实验数据

本文使用了 Microsoft 的 WikiQA 数据集^[13]、Stanford 的 SNLI 数据集^[14] 和基于 tweet 的 SemEval2016-Task 3 数据集^[15] 分别在问题回答、文本蕴涵识别及姿态检测 3 种文本匹配任务上进行实验对比。

WikiQA 是开放域问题解答的公开可用数据集, 数据集包含从 Bing 的查询日志中抽取的 3047 个问题和 29258 个答案, 其中 1473 个句子被标记为对应问题的答案。根据用户的点击行为, 每个问题与 Wikipedia 的多个答案相关联, 问题和答案的总数为 29258。然后, 采用人工标记的正确问题的答案作为候选答案, 因此将 1473 个句子标记为正确答案。训练集包含 20 k 对句子, 测试集包含 6.1 k 对句子, 验证集包含 2.7 k 对句子, 查询句平均包含 6.89 个词, 文档句平均包含 22.73 个词。

Stanford 的 SNLI 数据集是斯坦福大学发布的用于自然语言处理中文本蕴涵识别任务的数据集。SNLI 数据集是由人工标注的, 一共包含 570 k 个文本对, 其中训练集有 550 k, 验证集有 10 k, 测试集有 10 k; 其一共包含 3 类文本对, 分别是 entailment, contradiction, neutral。SNLI 数据集中的查询句

平均包含 12.85 个词, 文档句平均包含 7.41 个词。

SemEval-2016 Task 3 包含两个子任务, 即“问题-答案相似度”和“问题-问题相似度”, 实验在“问题-答案相似度”数据上进行了实验结果对比; 在“问题-答案相似度”任务中, 给出特定的问题, 然后根据与问题的相关性对答案进行排名。在姿态检测任务中, 目标是确定对给定(预先选择)目标的偏好性, 目标可能不是原文本中的意见目标, 所以需要模型对句子有更深入的理解和推理能力。实验选用“external answer”作为此次任务的备选答案, 标签分为 Good, Potentially Useful, Bad 3 种。数据集查询句平均包含 39.29 个词, 文档句平均包含 36.85 个词。

4.2 评价指标

本文实验利用归一化折现累积收益 (Normalized Discounted Cumulative Gain, NDCG) 和平均精度 (Mean Average Precision, MAP) 进行算法比较。

NDCG@ p 是一种计算模型排名质量的评价指标, 广泛应用于评估搜索引擎性能上。NDCG@ p 的计算方式是在给定特定查询的情况下, 估计前 p 个搜索结果中的相关信息量, 如果高度相关的文档稍后出现在搜索结果中, 则会给最终分数增加惩罚。其最终得分在 0 到 1 之间, 分数越高说明模型的性能越好。本文实验选择使用 NDCG@3, NDCG@5 和 MAP 作为模型的最终评价指标。

MAP 是信息检索任务的另一个常用指标, 其建立在平均精度 (AP) 度量标准的基础上, 其中 AP 衡量的是特定查询的相关文档与所有检索到的文档之比。现在假设有 Q 个查询, 则 MAP 分数只是其 AP 分数的平均值, 也介于 0 和 1 之间, 直观上, MAP 的分数越高表明模型的性能越好。

4.3 参数设置

针对 WikiQA 数据集, FMMI 模型采用的词向量维度大小为 100, Bi-LSTM 输出单元大小为 30, 批处理大小设置为 50, 查询输入的字符串长度为 10, 文档输入的字符串长度为 40, 模型训练使用的是 Adam 优化器, 初始学习率设置为 0.0005, 全连接神经网络和 Bi-LSTM 丢弃率都设置为 0.3。

针对于 SNLI 数据集, FMMI 模型采用的词向量维度大小为 100, Bi-LSTM 输出单元大小为 30, 批处理大小为 50, 查询输入的字符串长度为 20, 文档输入的字符串长度为 15, 使用 Adam 优化器, 初始学习率设置为 0.0005, 神经网络和 Bi-LSTM 丢弃率都设置为 0.3。

针对于 SemEval-2016 Task 3 数据集, FMMI 模型采用的词向量维度大小为 100, Bi-LSTM 输出单元大小为 30, 批处理大小为 30, 查询输入的字符串长度为 35, 文档输入的字符串长度为 35, 使用 Adam 优化器, 初始学习率设置为 0.0005, 神经网络和 Bi-LSTM 丢弃率都设置为 0.3。

实验根据每个数据集的实际文本长度选择了不同的查询和文档的输入字符长度。

4.4 实验对比

本节将本文模型 FMMI 与其他文本匹配模型进行了对比, 其他模型的实现均基于开源项目 MatchZoo^[16]。对比模型如下: ARC-I^[7], ARC-II^[7], MatchPyramid^[2], MV-LSTM^[5],

CONV_KNRM^[9], DUET^[17], DRMM^[18] 和 DRMMTKS^[19]。

表 1 列出了各个主流模型与本文的 FMMI 模型在 WikiQA 数据集上的实验结果对比,图 3 给出了每个模型在 WikiQA 数据集上的训练过程折线图。问答数据集相比于文本蕴含识别数据集更依赖于模型对文本数据多个级别上语义的理解,所以能够更好地提取文本语义的模型在该数据集上会表现出更优的效果。本文提出的 FMMI 模型在 NDCG@3, NDCG@5, MAP 3 个指标中的测试结果均优于其他模型。因此,相比于直接捕获单词之间单个粒度的相似性信息,FM-MI 通过对多粒度信息的提取和加权表示可以捕获更高层次上的语义信息。

表 1 在 WikiQA 数据集上的测试结果

Table 1 Results of testing on WikiQA dataset

Model	NDCG@3	NDCG@5	MAP
ARC-I	0.5785	0.6276	0.5909
ARC-II	0.6112	0.6674	0.6197
MatchPyramid	0.5725	0.6268	0.5914
DRMM	0.6404	0.6808	0.6384
Mv_LSTM	0.6708	0.7183	0.6753
DUET	0.5491	0.6203	0.5791
DRMMTKS	0.6545	0.7072	0.6639
CONV-KNRM	0.6200	0.6851	0.6360
FMMI	0.7252	0.7718	0.7308

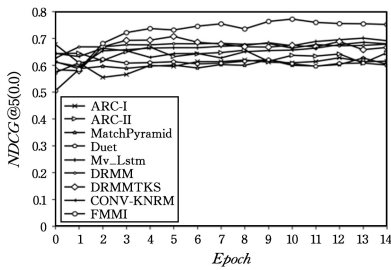


图 3 模型在 WikiQA 数据集上的训练过程

Fig. 3 Model training process on WikiQA dataset

表 2 列出了各个主流模型与本文的 FMMI 模型在 SNLI 数据集上的实验结果对比,图 4 给出了各个模型在 SNLI 数据集上的训练过程。与问答数据集相比,文本蕴含识别任务除了依赖于模型对文本整体语义信息的提取之外,还依赖于模型对文本局部特征信息的获取,所以能够更好地进行特征提取的模型虽然在 WikiQA 数据集上表现欠佳,但是在 SNLI 数据集上却表现出较好的实验效果,如 MatchPyramid, DUET, CONV_KNRM 等模型。由于本文提出的 FMMI 模型不仅可以捕获到高层次的文本语义信息,还可以较好地关注到局部特征信息,因此其在 SNLI 数据集上也呈现出了最优效果。

表 2 在 SNLI 数据集上的测试结果

Table 2 Results of testing on SNLI dataset

Model	NDCG@3	NDCG@5	MAP
ARC-I	0.8492	0.8496	0.8205
ARC-II	0.8876	0.8879	0.8726
MatchPyramid	0.8972	0.8973	0.8854
DRMM	0.8983	0.8985	0.8870
Mv_LSTM	0.8562	0.8565	0.8300
DUET	0.8853	0.8854	0.8690
DRMMTKS	0.8948	0.8950	0.8822
CONV-KNRM	0.9023	0.9026	0.8925
FMMI	0.9064	0.9066	0.8979

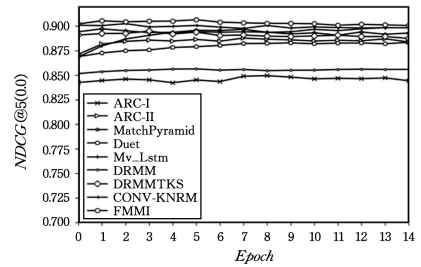


图 4 模型在 SNLI 数据集上的训练过程

Fig. 4 Model training process on SNLI dataset

表 3 列出了各个模型与 FMMI 模型在 SemEval2016-Task 3 数据集上的实验对比结果,图 5 给出了各个模型在 semEval2016-Task 3 数据集上的训练过程。该数据集选用外部答案(即待匹配文本没有在原文本中明确提及)作为候选目标,使得模型很难通过比较两段句子单纯的特征来判断其是否具有相关性,会更加考验模型对原文本的理解和推理能力,所以基于表示的模型可能更胜任这个任务,如 ARC-II, Mv_LSTM 等。本文提出的 FMMI 模型在该数据集上的结果依然优于其他匹配模型,因此可以证明 FMMI 在姿态检测任务中较其他模型而言可以更好地确定原文本中的陈述对于给定目标的偏好性。

表 3 在 SemEval2016-Task 3 数据集上的测试结果

Table 3 Results of testing on SemEval2016-Task 3 dataset

Model	NDCG@3	NDCG@5	MAP
ARC-I	0.1668	0.1732	0.1833
ARC-II	0.1843	0.2003	0.1876
MatchPyramid	0.1495	0.1584	0.1700
DRMM	0.1196	0.1245	0.1490
Mv_LSTM	0.1810	0.1808	0.2004
DUET	0.1811	0.1746	0.1725
DRMMTKS	0.1843	0.2003	0.1876
CONV-KNRM	0.1814	0.1819	0.1765
FMMI	0.2007	0.2158	0.2135

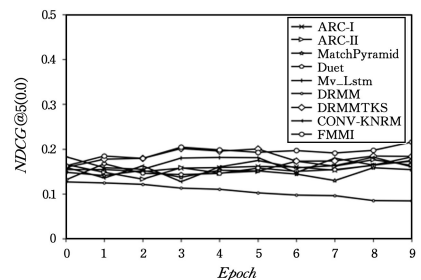


图 5 模型在 SemEval2016-Task 3 数据集上的训练过程

Fig. 5 Model training process on SemEval2016-Task 3 dataset

表 4 列出了本文提出的 FMMI 模型去掉一个必要的处理单元后在 WikiQA 数据集上的表现效果, FMMI-IA, FMMI-LSTM, FMMI-RA, FMMI-Granet 分别表示初始模型在移除交互注意力单元、Bi-LSTM 表示单元、表示注意力单元和粒度网络单元后的模型。从实验结果可以看出各个单元对整体模型的贡献程度。其中贡献度最大的是模型前期引用的交互注意力机制,其首先对原文本的每个单词进行的加权处理有效地增强了模型后期信息提取的效果和文本表示的能力;其次是用于文本表示的 Bi-LSTM 单元,由此可知 Bi-LSTM

单元在语义表示提取的过程中发挥了重要的作用;再次是表示注意力,其对粒度提取后的信息进行加权处理,使得模型对有用的词组信息进行关注;最后是多粒度网络,该网络通过获取更多的粒度信息在一定程度上提升了文本匹配的精度。

表4 FMMI消融实验

Table 4 FMMI ablation experiment

Model	NDCG@3	NDCG@5	MAP
FMMI-IA	0.6059	0.6711	0.6262
FMMI-LSTM	0.6625	0.7095	0.6657
FMMI-RA	0.6968	0.7524	0.7076
FMMI-Granet	0.7192	0.7561	0.7173
FMMI	0.7252	0.7718	0.7308

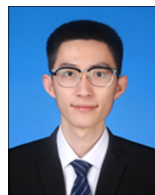
结束语 本文提出了一种结合多粒度信息的文本匹配融合模型(FMMD),该模型首先提取了文章中存在的多个粒度信息,进而利用了两种注意力机制将基于交互和基于表示的文本匹配模型进行了融合。通过加强模型对细粒度和粗粒度文本信息的关注,有效地提升了多粒度文本信息匹配的效果。在问答数据集、复述识别数据集和姿态检测数据集上将FM-MI模型和其他主流模型进行测试对比,FMMI在NDCG@3、NDCG@5和MAP 3个评价指标上均表现良好。在下一步的工作中,将针对如何更好地处理冗余的粒度信息和如何更好地计算两段文本的匹配分数展开进一步的研究工作。

参考文献

- [1] HUANG P S, HE X, GAO J, 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. California: ACM Press, 2013; 2333-2338.
- [2] PANG L, LAN Y, GUO J, et al. Text matching as image recognition[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Arizona: AAAI Press, 2016; 2793-2799.
- [3] JÄRVELIN K, KEKÄLÄINEN J. Cumulated gain-based evaluation of IR techniques[J]. ACM Transactions on Information Systems (TOIS), 2002, 20(4), 422-446.
- [4] SHEN Y, HE X, GAO J, et al. Learning semantic representations using convolutional neural networks for web search[C]// International Conference on World Wide Web. Seoul: ACM Press, 2014; 373-374.
- [5] WAN S, LAN Y, GUO J, et al. A deep architecture for semantic matching with multiple positional sentence representations [C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Arizona: AAAI Press, 2016; 2835-2841.
- [6] GONG Y, LUO H, ZHANG J. Natural Language Inference over Interaction Space. International Conference on Learning Representations [EB/OL]. [2020-7-11]. <https://arxiv.org/pdf/1709.04348.pdf>.
- [7] HU B, LU Z, LI H, et al. Convolutional Neural Network Architectures for Matching Natural Language Sentences[C]// Advances in Neural Information Processing Systems. Montreal: MIT Press, 2015; 2042-2050.
- [8] XIONG C, DAI Z, CALLAN J, et al. End-to-End Neural Ad-hoc Ranking with Kernel Pooling[J]. In Proceedings of the 40th International ACM SIGIR Conference on Research and Develop-

ment in Information Retrieval. Tokyo: ACM Press, 2017, 51(cd); 55-64.

- [9] DAI Z, XIONG C, CALLAN J, et al. Convolutional Neural Networks for Soft-Matching N-Grams in Ad-hoc Search[C]// Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. California: ACM Press, 2018; 126-134.
- [10] PENNINGTON J, SOCHER R, MANNING C. Glove: Global Vectors for Word Representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: ACL Press, 2014; 1532-1543.
- [11] SEO M, KEMBHAVI A, FARHADI A, et al. Bidirectional Attention Flow for Machine Comprehension. arXiv: Computation and Language [EB/OL]. [2020-7-11]. <https://arxiv.org/pdf/1611.01603.pdf>.
- [12] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated Residual Transformations for Deep Neural Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii: IEEE Press, 2017; 5987-5995.
- [13] YANG Y, YIH W, MEEK C. WikiQA: A Challenge Dataset for Open-Domain Question Answering[C]// In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: ACL Press, 2015; 2013-2018.
- [14] BOWMAN S, ANGELI G, POTTS C, et al. A large annotated corpus for learning natural language inference[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: ACL Press, 2015; 632-642.
- [15] NAKOV P, MÁRQUEZ L, MOSCHITTI A, et al. SemEval-2016 Task 3: Community Question Answering. ArXiv: Information Retrieval [EB/OL]. [2020-07-11]. <http://arxiv.org/abs/1912.00730.pdf>.
- [16] FAN Y, PANG L, HOU J, et al. MatchZoo: A Toolkit for Deep Text Matching. ArXiv: Information Retrieval [EB/OL]. [2020-07-11]. <https://arxiv.org/pdf/1707.07270v1.pdf>.
- [17] MITRA B, DIAZ F, CRASWELL N. Learning to match using local and distributed representations of text for web search [C]// Proceedings of the 26th International Conference on World Wide Web. Australia: ACM Press, 2017; 1291-1299.
- [18] GUO J, FAN Y, AI Q, et al. A Deep Relevance Matching Model for Ad-hoc Retrieval[C]// Conference on Information and Knowledge Management. Indianapolis: ACM Press, 2016; 55-64.
- [19] YANG Z, LAN Q, GUO J, et al. A deep top-k relevance matching model for ad-hoc retrieval[C]// China Conference on Information Retrieval. Guilin: Springer Press, 2018; 16-27.



LYU Le-bin, born in 1994, master. His main research interests include natural language processing and so on.



LIU Qun, born in 1969, Ph. D, professor, is a member of China Computer Federation. Her main research interests include data mining, complex network and so on.