

基于多特征融合的智能客服模型^①



李美玲, 任亚伟, 孙军梅, 李秀梅, 何鑫睿

(杭州师范大学 信息科学与工程学院, 杭州 311121)

通讯作者: 李秀梅, E-mail: xiumei_li@hotmail.com

摘要: 智能客服利用人工智能技术准确回答用户的咨询问题, 良好的句子相似度算法可以提高智能客服中问答的准确度. 本文针对金融证券领域客服, 提出了基于多特征融合的句子相似度算法模型, 提高了客服的智能性. 通过矩阵拼接的方式, 融合用户提问语句和知识库语句的词形特征和语义特征, 其中词形特征考虑 N-gram 相似度、编辑距离、Jaccard 相似度三种词形信息, 并针对语义特征提取, 提出了基于多头注意力机制 (multi-head attention) 的神经网络模型 LBMA. 利用上述融合的特征, 运用机器学习分类器判断两个语句是否相似, 并将分类器分类结果作为多特征融合模型的计算结果. 在尽量不改变语义信息的前提下, 通过数据增强 (Data Augmentation, DA) 技术扩充数据集, 提升了模型泛化能力. 实验结果表明, 与已有方法相比, 该模型在智能客服数据集上能够有效提升相似度计算的准确性, 准确率达到 94.69%.

关键词: 智能客服; 句子相似度; 多头注意力机制; 机器学习; 数据增强

引用格式: 李美玲, 任亚伟, 孙军梅, 李秀梅, 何鑫睿. 基于多特征融合的智能客服模型. 计算机系统应用, 2021, 30(7): 239-245. <http://www.c-s-a.org.cn/1003-3254/8068.html>

Intelligent Customer Service Model Based on Multi-Feature Fusion

LI Mei-Ling, REN Ya-Wei, SUN Jun-Mei, LI Xiu-Mei, HE Xin-Rui

(School of Information Science and Engineering, Hangzhou Normal University, Hangzhou 311121, China)

Abstract: Intelligent customer service accurately answers users' inquiries with artificial intelligence technology, and a good sentence similarity algorithm can improve the accuracy of questions and answers. In this study, we focus on the customer service in financial securities and propose a multi-feature fusion-based sentence similarity algorithm model, which improves the intelligence of customer service. Matrix splicing is adopted to fuse the morphological and semantic features of user question sentences and knowledge base sentences. The morphological features include N-gram similarity, edit distance, and Jaccard similarity. To extract semantic features, we propose a multi-head attention based neural network model named LBMA. With these fused features, a machine learning classifier is used to determine whether two sentences are similar, and the classification result of the classifier is regarded as the calculation result of the multi-feature fusion model. On the premise of not changing the semantic information as much as possible, the data set is expanded through data augmentation to improve the generalization of the model. Experimental results show that the proposed model can calculate the similarity of customer service data sets at an accuracy of 94.69%, higher than that of existing methods.

Key words: intelligent customer service; sentence similarity; multi-head attention mechanism; machine learning; data enhancement

① 基金项目: 国家自然科学基金 (61571174); 杭州师范大学星光计划项目

Foundation item: National Natural Science Foundation of China (61571174); Starlight Project of Hangzhou Normal University

收稿时间: 2020-11-06; 修改时间: 2020-12-12; 采用时间: 2021-01-18; csa 在线出版时间: 2021-06-30

1 引言

随着电子商务的快速兴起,智能客服成为金融领域的一大研究热点.智能客服的核心是快速、准确地回答用户的咨询问题.传统的智能客服模型主要是计算用户提问语句和知识库语句的相似度,通过匹配的方式检索出答案.尽管用户提问语句和知识库语句包含相似的关键词,但语义逻辑却可能截然不同^[1],如表1所示.其中,Q表示用户提问语句,T表示语义相同的语句一,F表示语义不同的语句二.因此,智能客服模型中语句相似度的计算需要综合考虑关键词信息和语义逻辑信息.

表1 语句例子

标签	句子内容
Q	今天买进的股票可以卖出取钱吗?
T	同一股票可以当天买入当天卖吗?
F	今天把买进的股票卖了能取钱吗?

句子相似度是实现智能客服的关键技术,是文本相似度识别的子集^[2].基于深度学习的方法是进行文本相似度计算的重要方法,它从不同角度提取文本中的信息,可以更大限度利用上下文信息,从而提高输入表征的信息量,克服文本语义表示上的鸿沟问题,使计算结果更加准确.基于深度学习的方法利用 Word2Vec^[3]等词向量方法将词语转化为词向量后,输入到神经网络模型中,得出句子的语义特征表示,再将语义特征表示送入全连接层或使用距离公式进行语义相似度计算. Hu 等将卷积神经网络 (Convolutional Neural Network, CNN)^[4]用于问题文本相似度计算中,分别对两个句子用 CNN 进行编码,然后将编码后的文本向量进行拼接,送入全连接网络^[5]. Yin 等使用两个独立的带有注意力最大值池化的字符级和单词级 CNN 模型进行相似度计算^[6]. CNN 在卷积层使用滤波器抽取文本特征,只能获取滤波器窗口内单词间的依赖关系,无法获取句子中所有单词间的依赖信息,无法获取整体结构信息.与 CNN 不同,循环神经网络 (Recurrent Neural Network, RNN) 不仅考虑前一时刻的输入,而且赋予网络对前面内容的记忆功能. Mueller 在词向量的基础上提出了使用长短时记忆 (Long-Short Term Memory, LSTM)^[7]神经网络提取句子的语义特征,并结合余弦距离计算获得的句子向量的相似程度^[8].由于单向长短时记忆神经网络只能捕获前一时刻的信息, Neculoiu 提出利用双向长短时记忆神经网络 (Bidirectional Long-

short Term Memory, Bi-LSTM)^[9]同时获取上下文的信息,加强句子特征的提取^[10]. Bowman 等人使用时间递归神经网络在数据集 SICK (Sentence Involving Compositional Knowledge) 进行语义推理^[11]. 纪明宇等人将门控循环网络 (Gated Neural Network, GRU) 应用在智能客服数据的相似度计算上取得了不错的效果^[12].

循环神经网络如 LSTM、GRU 编码句子时,句子间的依赖程度会随着距离增加而减弱.为弥补循环神经网络这一缺点, Hu 引入了注意力机制^[13],其核心思想是通过给文本词向量和其他隐藏表示向量分配不同的权重系数来突出对结果影响较大的特征.杨飞以单词和句子为研究对象,提出 Siamese-LSTM 模型融入注意力机制可以在高维的结构空间中学习文本的丰富语义特征表示^[14]. Lin 等提出自注意力机制能编码远距离的依赖关系^[15]. Vaswani 等提出的基于自注意力的翻译模型 (Transformer) 中采用的多头注意力 (multi-head attention) 能集成不同子空间的信息提高编码依赖关系的能力^[16]. 胡艳霞等基于多头注意力机制和 Tree-LSTM 的句子相似度计算,实现句子对间相关语义结构特征的充分利用^[17].

本文面向金融证券领域智能客服,在文献 [8] 所研究的 Siamese 网络的基础上,引入多头注意力机制提取不同语义空间中字词间的依赖关系,提出多特征融合的句子相似度模型,更好地捕捉两个句子语义中的上下文信息,进而更好地计算用户提问语句和知识库语句之间的相似度.本文主要贡献有以下两点:

1) 提出一种面向智能客服的句子相似度模型,该模型融合语义特征和词形特征,其中词形特征基于传统统计特征方法实现,极大保留了关键词和词频相似度特征;语义特征基于深度学习实现,能够抽取文本全局结构信息和深层语义信息,并且加入多头注意力机制集成不同子空间的信息以理解字词之间的依赖关系,更好地计算金融证券领域中用户提问语句与知识库语句之间的相似度.相比于传统单一神经网络模型和基于注意力机制的神经网络模型,本文提出的多特征融合模型可以取得更好的相似度计算效果.

2) 利用网络爬虫技术抓取数据,得到金融证券领域的用户询问数据集,并在该数据集基础上运用数据增强技术^[18]进行不同倍数的数据增强,利用数据增强后的数据集进一步验证了本文提出的多特征融合模型的有效性和泛化能力.

2 多特征融合模型

本文面向金融证券领域,提出多特征融合模型用于计算智能客服中用户提问语句和语料库语句相似度.该模型综合考虑语句的词形特征和语义特征,利用关键词信息和词频信息得到语句的词形特征,利用语义信息得到语句的语义特征.模型具体结构如图1所示.

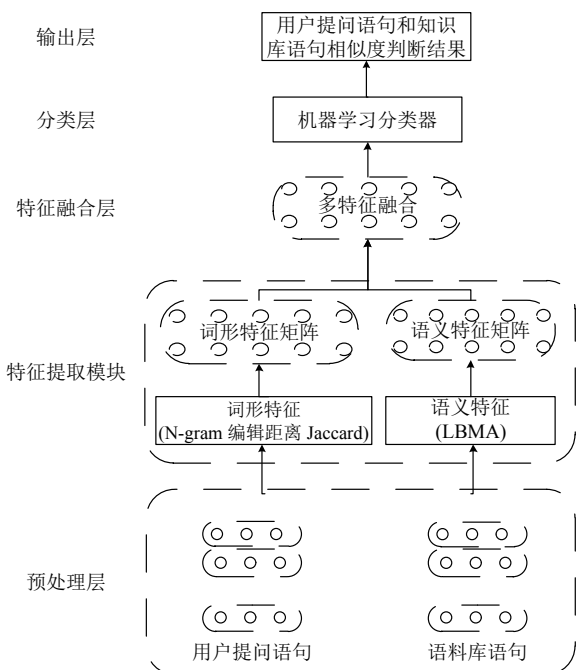


图1 多特征融合模型图

首先在预处理层,对用户提问语句 S_1 与语料库语句 S_2 进行去无用字符、去停用词、中文纠错、中文分词等预处理操作,然后通过特征提取模块分别获得词形特征矩阵和语义特征矩阵,并在特征融合层将词形特征矩阵和语义特征矩阵按列拼接,之后输入到分类层进行相似度判断.最终将机器学习分类器分类结果作为两个语句间多特征融合相似度判断结果.多特征融合模型充分利用关键词和词频方面信息提取两个语句的词形特征,以及深度学习方法提取两个语句的语义特征.

2.1 获取语句词形特征模块

该模块由N-gram算法^[19]、编辑距离^[20]和Jaccard^[21]相似度3种传统统计特征方法构成,用于获取预处理层输出语句的词形特征矩阵.首先采用N-gram算法,该算法是基于统计语言模型的经典算法,但其只考察了语句对中词出现的信息,没有完全考虑词之外的其他信息,例如语句间的编辑距离、相同词的

数量等.因此,本文在N-gram算法基础上,引入语句编辑距离差异和Jaccard相似度作为影响因素.获取词形特征模块分以下4个步骤进行:

1) 利用N-gram算法得到语句 S_1 和语句 S_2 在词序列方面的相似度特征,具体计算如下式所示:

$$SimS(S_1, S_2, n) = \frac{1}{ANC(n)} \sum_{i=1}^{\min(len_1(n), len_2(n))} SimN(i-gram_1, i-gram_2) \quad (1)$$

其中, $SimS(S_1, S_2, n)$ 表示 S_1 和 S_2 基于不同阶N-gram的句子相似度值, $len_1(n)$ 和 $len_2(n)$ 分别是 S_1 和 S_2 的 n 元N-gram的个数, $ANC(n) = \frac{len_1(n) + len_2(n)}{2}$ 意义在于计算两个长度不同句子的相似度时引入一定的惩罚措施. $SimN$ 指计算两个句子的对应N-gram集合之间的相似度,通过将文本里面的内容按照字节进行大小为 N 的滑动窗口操作,形成长度为 N 的字节片段序列,然后通过计算公共子串^[22]的比例来衡量相似度,这里 N 取3.

2) 通过动态规划求 S_1 转换到 S_2 的操作数(插入、删除和替换)得到语句 S_1 和语句 S_2 在编辑距离方面的相似度特征.操作数越少则编辑距离越小,相似性越高.

3) 通过Jaccard算法统计得到语句 S_1 和语句 S_2 在所含相同词汇数量方面的相似度特征.通常,两个句子中共同出现的词语较多,即反应了两个句子间相同成分占的比例越大,则相似性越高,具体计算如下式所示:

$$Jaccard = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{|S_1 \cap S_2|}{|S_1| + |S_2| - |S_1 \cap S_2|} \quad (2)$$

4) 将获得的N-gram相似度矩阵、编辑距离相似度矩阵和Jaccard相似度矩阵按列拼接获得三特征融合的词形特征矩阵.

2.2 获取语句语义特征模块

该模块通过深度学习方法获取预处理层输出语句的语义特征矩阵.我们提出基于多头注意力机制的神经网络模型LBMA(LSTM Based on Multi-head Attention),主要包括词向量嵌入层、网络编码层、注意力交互层、语义特征融合层.如图2所示,语句 S_1 和语句 S_2 经过词嵌入层将句子中的词映射为对应的词向量,这里我们将位置信息映射为与词向量相同维度的向量,位置编码的目的是区分句子中词的位置

信息, 计算公式如式 (3) 和式 (4) 所示, 并将词向量和位置编码合并输入网络编码层。

图 3 为 LBMA 模型中网络编码层、注意力交互层和语义特征融合层的详细结构图。包含位置编码的词向量经过 LSTM 网络和 Bi-LSTM 网络进行全局语义特征编码表示, 将 LSTM 网络输出经过点乘的方法送入 CNN 网络层充分提取句子间的相似特征信息, 同时运用多头注意力机制获取 LSTM 网络输出不同子空间的字词依赖特征, 其中多头注意力计算过程如式 (5)~式 (7) 所示, 并将经过多头注意力机制重新语义编码的向量全局平均池化后经过余弦相似度和曼哈顿相似度两个特征提取方法获取深层语义特征, 将 Bi-LSTM

网络输出经过乘法、减法和最大值 3 个特征提取方法获取差异特征, 最终将以上特征通过矩阵拼接的方式进行语义特征融合。

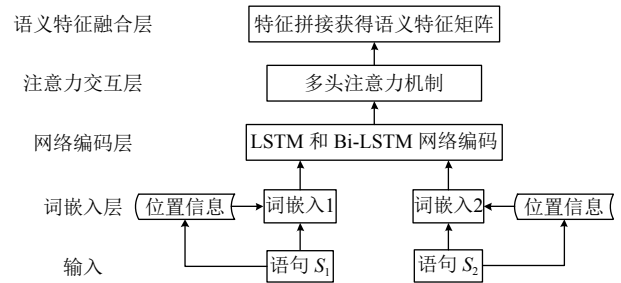


图 2 LBMA 模型框架

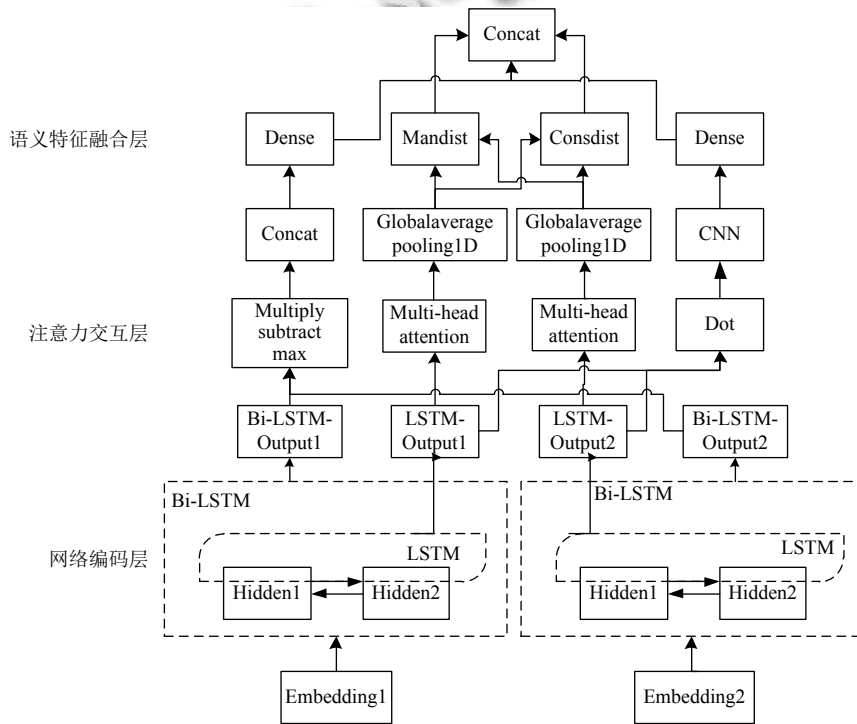


图 3 LBMA 中网络编码层、注意力交互层和语义特征融合层的详细结构图

$$PE(pos, 2i) = \sin\left(\frac{pos}{1000 \frac{i-p\%2}{d_{model}}}\right) \quad (3)$$

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{1000 \frac{i-p\%2}{d_{model}}}\right) \quad (4)$$

其中, pos 表示词语在句子中的位置, i 表示词向量的位置, $\%$ 表示取余操作, d_{model} 表示词向量维度。

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (6)$$

$$MultiHead(Q, K, V) = Concat(head_1 \dots head_h)W^o \quad (7)$$

其中, $W_i^Q, W_i^K, W_i^V \in R^{d_{model} \times d_k}$, $W^o \in R^{d_{model} \times d_k}$, $i = 1, \dots, h$, Q, K, V 均为上文中经过网络编码层得到的输入向量矩阵, QW_i^Q, KW_i^K, VW_i^V 分别表示对 Q, K, V 进行线性变换, h 表示注意力的个数, 每个注意力捕捉文本中一个子空间的信息, 将 h 个注意力头进行拼接, 通过矩阵 W^o 得到多头注意力值。

3 实验

3.1 数据集

本文所采用的基础数据集为2019“海康杯”第八届浙江省大学生服务外包创新应用大赛公共数据集。此外,我们通过Python网络爬虫技术模拟浏览器行为的方式,从百度知道动态抓取数据,对该基础数据集进行扩充,得到金融证券领域用户询问数据集。通过人工标注的方法,将已有数据中每对相似语句标注为标签1,作为正样本数据集,即两个句子同义;将已有数据中每对不相似语句标注为标签0,作为负样本数据集,即两个句子异义。共获取金融证券领域相关数据117588条,其中数据正负样本比为52877:64711,并将数据集中的60%作为训练集,40%作为测试集。为验证模型泛化能力,在尽量不改变语义的前提下,本文利用适合中文语料的4种数据增强方法^[18]扩充数据集。4种方法为同义词替换(Synonym Replacement, SR)、随机插入(Random Insertion, RI)、随机交换(Random Swap, RS)和随机删除(Random Deletion, RD),其中,SR是从句子中随机选取 n 个不属于停用词表中的字词,并随机选择其同义词替换它们;RI是随机找出句中某个不属于停用词表的词,并求出其随机的同义词,将该同义词插入句子的随机位置,重复 n 次;RS是随机的选择句子两个字词并交换它们的位置,重复 n 次;RD是以每一条语料中改动的词所占的比例 p 随机删除句子中的一个字词,重复 n 次,其中, p 和 n 均是人为设置的值。该数据增强技术生成类似于原始数据并引入一定程度噪声的增强数据,该方法有助于防止过拟合,其中同义词替换和随机插入操作会引入新的词汇,允许模型泛化到那些在测试集中但不在训练集中的字词。

3.2 评价指标

实验根据召回率(Recall, R)实际识别出的正确结果与数据库中总的正确结果的百分比,如式(8);准确率(Accuracy, A)返回结果中所有预测正确结果占总样本数的百分比,如式(9); $F1$ 分数($F1$ -score)值为召回率(R)和精确率(P)的调和平均值,如式(11)。3个评价指标的计算结果来评价算法计算相似度的性能。

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$A = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$F1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2 * P * R}{P + R} \quad (11)$$

其中, TP 为真正类(True Positive)即一个实例是正类且被预测为正类; FN 为假负类(False Negative)即一个实例是正类但被预测为负类; FP 为假正类即一个实例为负类但被预测为正类; TN 为真负类(True Negative)即一个实例是负类且被预测为负类。

3.3 超参数

在本文的实验中,词向量采用Word2Vec词嵌入向量表示方法,其中每个词向量维度为300维,词典大小为2.08 MB, Batch size 设置为512,激活函数采用Sigmoid,学习率设为0.001,代价函数选择二元交叉熵(binary_crossentropy),为了缓解训练中可能出现的过拟合问题,随机失活率(dropout rate)设置为0.5,采用Adam优化方法,模型迭代次数(epoch)最大设为20,使用Earllystopping,多头注意力机制中注意力个数设置为8。

3.4 实验及结果分析

为验证本文提出的词形特征模块有效性和语义特征模块有效性,设置实验1和实验2;为验证提出的多特征融合模型有效性,设置实验3;为验证数据增强方法有效性,设置实验4,具体实验结果如下。

实验1. 基于词形特征的语句相似度计算

为验证三词形特征融合方法的有效性,我们分别用N-gram、编辑距离和Jaccard相似度3种基线方法和三词形特征融合方法对测试集进行相似度计算,根据召回率、准确率和 $F1$ 值进行评价。当三词形特征融合时,我们经过多次对比实验后选择基于RBF核的支持向量机分类器^[23]进行分类判断,设置阈值为0.8,将测试结果中大于阈值的数据归为1即判断两个语句同义,小于阈值的数据归为0即判断两个语句异义,并将分类器分类结果作为两个语句间的词形相似度判断结果。各评价指标结果如表2所示。

表2 词形特征计算相似度的各评价指标结果(%)

词形特征方法	召回率	准确率	$F1$ 值
N-gram	84.42	85.99	84.41
编辑距离	72.32	80.07	76.52
Jaccard	78.70	83.37	80.96
三词形特征融合	85.58	86.80	85.35

从表2可知,三词形特征融合能提升效果,使召回率、准确率为、 $F1$ 值都提升到85%以上。

实验 2. 基于语义特征的句子相似度计算

为验证基于多头注意力机制的神经网络模型 LBMA 的有效性和比较不同深度学习模型的效果, 我们加入了 Siamese-LSTM、Siamese-LSTM-Att 和 Siamese-LSTM-Self Att 三个基线模型进行对比, 实验采用五折交叉验证的方式选取评价指标的平均值作为各评价指标结果, 如表 3 所示, 模型准确率对比曲线如图 4 所示。

表 3 神经网络模型计算相似度的各评价指标结果 (%)

模型	召回率	准确率	F1值
Siamese-LSTM	78.48	84.19	81.69
Siamese-LSTM-Att	81.50	88.77	86.71
Siamese-LSTM-Self Att	88.75	91.01	89.87
LBMA	95.10	94.43	93.92

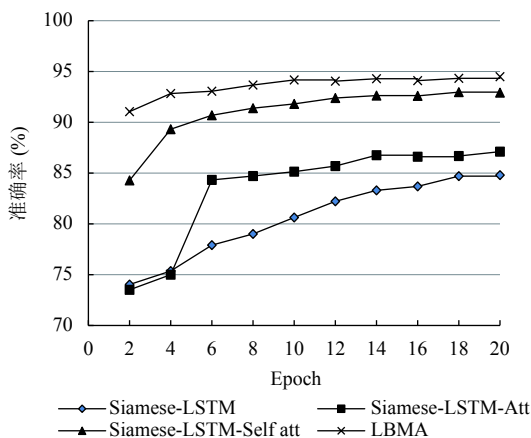


图 4 不同模型准确率的变化曲线

从表 3 可知, Siamese-LSTM 模型准确率为 84.19%, 而加入注意力机制的 Siamese-LSTM-Att 模型相比 Siamese-LSTM 模型准确率提升了 4.58%, 可以看出局部注意力对于模型表现提升是有效的。加入自注意力机制的 Siamese-LSTM-Self Att 相比 Siamese-LSTM 模型提升了 6.82%, 表明自注意力机制更能捕捉句子中的语义特征。此外, LBMA 相比 Siamese-LSTM 模型提升了 10.24%, 可以看出基于多头注意力的神经网络模型 LBMA 提取语义特征能力更强。

从图 4 可知, 我们的模型在第 2 个 epoch 时就达到了 90% 的准确率, 并在第 8 个 epoch 开始缓慢提升, 模型接近收敛点, 证明我们的基于多头注意力机制的神经网络模型收敛速度明显快于其他的模型。

实验 3. 多特征融合模型相似度计算

为验证多特征融合模型的有效性, 进行多特征融合模型相似度计算实验, 该实验结果与三词形特征融

合模型和语义特征模型 LBMA 结果相比较, 各评价指标结果如表 4 所示。

表 4 不同模型计算相似度的各评价指标结果 (%)

模型	召回率	准确率	F1值
三词形特征融合模型	85.58	86.80	85.35
语义特征模型LBMA	95.10	94.43	93.92
多特征融合模型	95.28	94.69	94.16

从表 4 可知, 多特征融合模型计算结果高于三词形特征融合模型和语义特征模型 LBMA 计算结果, 因此多特征融合方法对于金融证券领域句子相似度计算是有效的。

实验 4. 数据增强有效性验证

为验证数据增强方法有效性, 进行数据增强实验, 并将该实验结果与原数据集实验结果相比较。数据增强方法如下, 本文借助外部 Synonyms 知识库引入新的中文词汇, 实现 4 种数据增强方法中同义词替换和随机插入操作, 然后设置对于每一条语料增强的数据量 n 和每一条语料中改动的词所占的比例 p , 并在原始数据的基础上对于每一条语料分别通过 4 种数据增强方法获得 $4n$ 条增强数据, 并将 $4n$ 条数据打乱顺序后选出前 n 条并保留原始数据为最终数据。具体地, 我们两次运用数据增强技术, 其中比例 p 都设置为 0.1, n 分别设置为 1 和 4, 即每条数据分别增强 1 条和 4 条并保留原数据, 分别扩充至 235 176 条 (2 倍) 和 587 940 条 (5 倍), 与原数据集进行对比实验, 多特征融合模型中的机器学习分类器经过多次对比实验后选择基于 RBF 核的支持向量机, 数据增强后各评价指标结果如表 5。

表 5 多特征融合模型数据增强后各评价指标结果 (%)

增强倍数	召回率	准确率	F1值
/	95.28	94.69	94.16
2	99.13	99.26	99.26
5	99.24	99.37	99.37

从表 5 可知, 多特征融合模型在数据增强后召回率、准确率和 F1 值均有所提升, 说明数据增强方法对于该数据集是有效的, 可以提升模型泛化能力, 进一步说明模型对于增强的那一部分数据的特征提取效果更好。

4 总结

本文为提高金融证券领域客服的智能性, 提出融合词形特征和语义特征的句子相似度计算模型, 其中词形特征考虑了 N-gram 相似度、编辑距离、Jaccard

相似度三种词形信息,并提出基于多头注意力机制的神经网络模型 LBMA 提取句子间的语义特征.实验结果表明,与单一神经网络模型和基于注意力的神经网络模型相比, LBMA 的性能有所提升,且多特征融合模型在客服数据中也获得了较好的效果.为验证数据增强方法对于多特征融合模型的有效性,我们在尽量不改变语义信息的前提下使用数据增强方法扩充数据集进行对比实验,进一步提升了多特征融合模型的句子相似度计算效果.

参考文献

- 张涛, 马海群, 易扬. 文本相似度视角下我国大数据政策比较研究. 图书情报工作, 2020, 64(12): 26–37.
- 江长柱, 明晓乐, 张东香. 基于抽象知识点模型的句子相似度计算. 计算机系统应用, 2015, 24(5): 193–197.
- Sharma AK, Chaurasia S, Srivastava DK. Sentimental short sentences classification by using CNN deep learning model with fine tuned Word2Vec. Procedia Computer Science, 2020, 167: 1139–1147. [doi: 10.1016/j.procs.2020.03.416]
- Lawrence S, Giles CL, Tsoi AC, *et al.* Face recognition: A convolutional neural-network approach. IEEE Transactions on Neural Networks, 1997, 8(1): 98–113. [doi: 10.1109/72.554195]
- Hu BT, Lu ZD, Li H, *et al.* Convolutional neural network architectures for matching natural language sentences. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2014. 2042–2050.
- Yin WP, Yu M, Xiang B, *et al.* Simple question answering by attentive convolutional neural network. Proceedings of the 26th International Conference on Computational Linguistics. Osaka, Japan. 2016. 1746–1756.
- Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]
- Mueller J, Thyagarajan A. Siamese recurrent architectures for learning sentence similarity. Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, AZ, USA. 2016. 2782–2792.
- Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 1997, 45(11): 2673–2681. [doi: 10.1109/78.650093]
- Neculoiu P, Versteegh M, Rotaru M. Learning text similarity with Siamese recurrent networks. Proceedings of the 1st Workshop on Representation Learning for NLP. Berlin, Germany. 2016. 148–157.
- Bowman SR, Potts C, Manning CD. Recursive neural networks can learn logical semantics. Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. 2015. 12–21.
- 纪明宇, 王晨龙, 安翔, 等. 面向智能客服的句子相似度计算方法. 计算机工程与应用, 2019, 55(13): 123–128. [doi: 10.3778/j.issn.1002-8331.1812-0142]
- Hu DC. An introductory survey on attention mechanisms in NLP problems. In: Bi YX, Bhatia R, Kapoor S, eds. Intelligent Systems and Applications. Cham: Springer, 2020. 432–448.
- 杨飞. 基于 LSTM 的文本相似度识别方法研究 [硕士学位论文]. 长春: 吉林大学, 2018.
- Lin ZH, Feng MW, Santos CND, *et al.* A structured self-attentive sentence embedding. Proceedings of the 5th International Conference on Learning Representations. arxiv: 1703.03130, 2017.
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA. 2017. 6000–6010.
- 胡艳霞, 王成, 李弼程, 等. 基于多头注意力机制 Tree-LSTM 的句子语义相似度计算. 中文信息学报, 2020, 34(3): 23–33. [doi: 10.3969/j.issn.1003-0077.2020.03.003]
- Wei J, Zou K. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China. 2019. 6382–6388.
- Brown PF, deSouza PV, Mercer RL, *et al.* Class-based *n*-gram models of natural language. Computational Linguistics, 1992, 18(4): 467–479.
- Ristad ES, Yianilos PN. Learning string-edit distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(5): 522–532. [doi: 10.1109/34.682181]
- 俞婷婷, 徐彭娜, 江育娥, 等. 基于改进的 Jaccard 系数文档相似度计算方法. 计算机系统应用, 2017, 26(12): 137–142. [doi: 10.15888/j.cnki.csa.006123]
- Mohamed TM, Elmahdy HN, Onsi HM. Efficient watermark detection by using the longest common substring technique. Egyptian Informatics Journal, 2011, 12(2): 115–123. [doi: 10.1016/j.eij.2011.05.001]
- Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293–300. [doi: 10.1023/A:1018628609742]