计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目：基于 capsule-BiGRU 的文本相似度分析算法
作者：赵琪，杜彦辉，芦天亮，沈少禹
网络首发日期：2020-08-26
引用格式：赵琪,杜彦辉,芦天亮,沈少禹.基于 capsule-BiGRU 的文本相似度分析算法.计算机工程与应用.
https://kns.cnki.net/kcms/detail/11.2127.TP.20200826.1635.010.html

网络首发时间：2020-08-26 17:21:37
网络首发地址：https://kns.cnki.net/kcms/detail/11.2127.TP.20200826.1635.010.html

基于 capsule-BiGRU 的文本相似度分析算法

赵　琪，杜彦辉，芦天亮，沈少禹

中国人民公安大学 警务信息工程与网络安全学院，北京 100038

摘　要：针对传统神经网络模型不能很好的提取文本特征的问题，提出基于 capsule-BiGRU 的文本相似度分析方法，该方法将胶囊网络（capsule）提取的文本的局部特征矩阵和双向门控循环单元网络（BiGRU）提取的文本的全局特征矩阵分别进行相似度分析，得到文本的相似度矩阵，将相似度矩阵融合，得到两个文本的多层次相似度向量，从而进行文本相似度的判定。将传统的胶囊网络进行改进，将与文本语义无关的单词视为噪声胶囊，赋予较小权值，从而减轻对后续任务的影响。针对文本相似度的任务，在文本特征矩阵提取前加入互注意力机制，对于待分析的两个文本，通过计算一个文本中单词与另一文本中所有单词的相似度来对词向量赋予权值，从而能更准确地判断文本的相似度。Quora Questions Pairs 数据集进行实验，实验结果表明所提出的方法准确率为 86.16%，F1 值为 88.77%，结果优于其他方法。

关键词：文本相似度；胶囊网络；双向门控循环单元网络；注意力机制
文献标志码: A　　中图分类号:TP391.1　　doi：10.3778/j.issn.1002-8331.2004-0253

赵琪，杜彦辉，芦天亮，等. 基于 capsule-BiGRU 的文本相似度分析算法.计算机工程与应用

ZHAO Qi, DU Yanhui, LU Tianliang, et al. Algorithm of text similarity analysis based on capsule-BiGRU. Computer Engineering and Applications

Algorithm of text similarity analysis based on capsule-BiGRU

ZHAO Qi, DU Yanhui, LU Tianliang, SHEN Shaoyu

School of Police Information Engineering and Cyber Security, People's Public Security University of China, Beijing 100038, China

Abstract: Aiming at the problem that the traditional neural network model cannot extract the features of the text well, a text similarity analysis method based on capsule-BiGRU is proposed. The local features matrix of the text extracted by the capsule network and the global features matrix of the text extracted by the BiGRU are analyzed for similarity separately to obtain the similarity matrix of the text, to judge the similarity of text. The traditional capsule network is improved, words that have nothing to do with text semantics are regarded as noise capsules, and smaller weights are assigned to reduce the impact on subsequent tasks. For the task of text similarity, a co-attention mechanism is added before feature extraction. For two texts to be analyzed, weights are given by calculating the similarity between words in one text and all words in another text, so that determine the similarity of text more accurately.

基金项目：国家重点研发计划（No.20190178）；中国人民公安大学本科教学研究项目（No.2019JY10）；中国人民公安大学基本科研业务费重大项目(No.2019JKF108)。

作者简介：赵琪（1996-），男，硕士研究生，研究领域为自然语言处理；杜彦辉（1969-），通讯作者，男，博士，教授，博导，主要研究方向为网络攻防、人工智能，E-mail：duyanhui@ppsuc.edu.cn；芦天亮（1985-），男，博士，副教授，硕导，研究领域为网络安全、恶意代码分析与检测；沈少禹（1995-），男，硕士研究生，研究领域为网络安全。

Experiment with the Quora Questions Pairs dataset. The experimental results show that the proposed method has an accuracy rate of 86.16% and an F1 value of 88.77%, which is better than other methods.

Key words：text similarity; capsule network; BiGRU; attention mechanism

1 引言

文本相似度在自然语言处理中有着重要的地位。其旨在对给定的两个文本进行特征提取，计算文本特征向量相似度，以此来量化两个文本之间的相似程度。文本相似度在自动问答系统、信息检索、自动文本摘要、文本分类等自然语言处理的任务中都有着广泛的应用[1]。

近年来，随着深度学习的发展，深度学习在文本相似度的任务中得到广泛的应用。由于卷积神经网络[2]与循环神经网络[3]在各个领域的任务中表现出了良好的性能，从而成为如今主要的两种神经网络模型结构。卷积神经网络通过对词向量矩阵进行处理，从而有效地提取出文本的局部特征，但缺点在于不能考虑文本的上下文信息，有时不能表达文本真正的含义。循环神经网络将文本视为一个序列，它可以将上一神经元的输出作用于下一神经元，因此这种网络结构具有记忆性，利用循环神经网络完成文本特征向量提取，可以考虑词语的顺序信息，利用文本的上下文信息提取文本的全局特征，但对于长距离的依赖关系，循环神经网络不能很好地提取文本特征。针对两种网络结构的特点，本文提出基于 capsule-BiGRU 的文本相似度分析方法，该方法将两个文本通过两种神经网络结构处理之后得到的文本特征向量进行相似度分析，得到局部相似度矩阵和全局相似度矩阵，将两个层次的相似度矩阵进行融合，以此完成文本相似度分析。

本文提出的方法首先利用互注意力机制赋予单词不同的权重，针对两个文本，对两个文本的词向量距离进行计算，对于更接近另一个文本的单词给更高的权重。其次结合胶囊网络与 BiGRU 网络构建集成模型，将胶囊网络提取的文本局部特征和 BiGRU 网络提取的文本全局特征分别进行相似度分析，将两个层次的相似度矩阵进行融合。最后根据两个句子的相似度向量判断文本是否相似。

2 相关工作

传统的文本相似度研究的方法主要是以 one-hot、词袋模型、N-gram，TF-IDF 等作为文本的特征向量[4-5]，利用余弦相似度等方法作为量化文本相似程度的指标。但这些方法单纯的以文本的统计信息作为文本的特征向量，没能考虑词语的上下文信息，同时在特征提取时存在特征稀疏和维度爆炸的问题[6]。

随着深度学习的发展，利用深度学习的方法研究文本相似度任务成为了如今的主流方法。

Mikolov 等人[7]在文中提出 word2vec 词向量嵌入方法，作为一种神经网络语言模型，该方法将单词转化为多维向量表示，极大的方便了后续工作。Pennington 等人[8]在文中提出 Glove 词向量嵌入方法，该方法融合了全局矩阵分解方法和局部文本框捕捉方法的优点，词向量嵌入考虑了上下文的信息，更准确地表达了文本的上下文信息，在多个自然语言处理任务中有良好的表现。

Yoon Kim 等人[9]在 2014 年发表的论文中提出了 TextCNN 模型，将 CNN 应用到自然语言处理领域中，进行文本分类任务。使用预先训练好的词向量作为模型的输入，使用多个不同尺寸的卷积核来提取句子的文本特征向量，经过最大池化层筛选句子的显著特征，将筛选出的特征连接，最后进入全连接层输出每个类别的概率。Sabour 等人[10]在 2017 年在文中提出了胶囊网络，胶囊网络是卷积神经网络的一种变体，使用神经元向量代替传统卷积神经网络中的单个神经元节点，以向量的形式保存更多地信息。同时以动态路由机制训练胶囊网络，减少了网络的参数，在手写数字识别数据集上有很好的效果。Zhao 等人[11]将胶囊网络引入自然语言处理中，做文本分类的任务，胶囊网络可以对文本信息进行有效的编码，保存了文本多层次的特征，提取出的特征向量更准确地表达了文本的取得了很好的效果。Matthew 等人[12]在文中使用 CNN 完成实体关系抽取任务，该模型使用的多个粒度的卷积网

络，具有良好的表现。

Mikolov 等人[13]将循环神经网络引入自然语言处理领域，利用循环神经网络完成了机器翻译的任务，使用循环结构遍历整个文本，得到文本的全局特征。Sundermeyer 等人[14]在文献中将 LSTM 应用于自然语言处理领域，LSTM 解决了传统循环神经网络对于输入序列长距离信息依赖关系的问题。Paul 等人[15]提出基于孪生网络结构的双向长短期记忆神经网络用于文本相似度，该网络通过两个 LSTM 网络遍历整个文本，综合考虑每个单词的上下文信息，提取句子的特征，完成文本相似度的判别。Bengio 等人[16]在将注意力机制应用到了自然语言处理领域，通过注意力机制，神经网络具备专注于某些特征的能力，对于重要的特征分配较多的注意力。

Wieting 等人[17]在文中提出基于注意力机制的卷积神经网络模型用于文本相似度分析，该模型利用注意力机制对文本单词赋予权重并使用卷积网络提取特征矩阵。方炯焜等人[18]将 GloVe 词向量嵌入方法与 GRU 网络结合起来做文本分类的任务，该方法利用 GloVe 方法完成单词表示，并利用 GRU 网络作为分类器在文本分类中有较好的表现。Pontes 等人[19]提出利用 CNN 与 LSTM 网络提取文本特征完成文本相似度分析任务。郭浩等人[20]提出基于 CNN-BiLSTM 的文本相似度计算方法，使用 CNN 和 BiLSTM 网络提取文本特征，完成相似度计算。唐庄等人[21]提出一种 transformer-capsule 集成模型，分别利用胶囊网络和 transformer 来提取文本的局部短语特征和全局特征，得到文本序列的多层次特征表示，完成文本分类的任务。尹春勇等人[22]对传统胶囊网络改进，将卷积神经网络和胶囊网络融合做文本分类的任务。

针对 CNN 网络与 RNN 网络在特征提取阶段的优势与不足，本文提出的方法将 CNN 网络变体 capsule 与 RNN 网络变体 BiGRU 结合，同时引入互注意力机制，解决了传统神经网络模型不能很好地提取文本的特征向量的问题。本文提出的方法在文本相似度任务中有较好的表现

3 模型

基于 capsule-BiGRU 的文本相似度分析模型框架如图 1 所示。本文提出的模型主要包括：词向量嵌入模块、特征矩阵提取模块、特征矩阵分析判断模块。本文提出的方法首先针对文本相似度任务，使用互注意力机制分析重要的单词赋予较高的权重。其次将胶囊网络（capsule）和双向门控循环单元网络（BiGRU）相结合，使用胶囊网络提取文本的局部特征，使用 BiGRU 网络提取文本的全局特征，将提取出的两个层次的特征融合得到文本的多层次特征。同时对将传统的胶囊网络进行改进，将与文本语义无关的单词视为噪声胶囊，赋予较小权值，从而减轻对后续任务的影响。

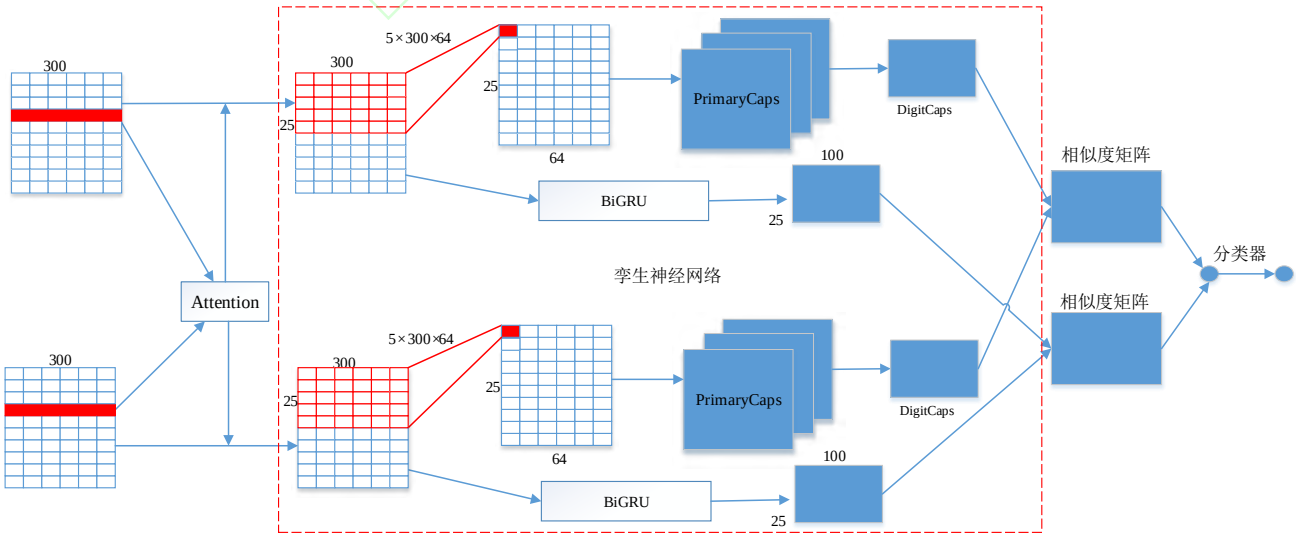

图 1　基于 capsule-BiGRU 的文本相似度分析模型框架

Fig.1　Framework of text similarity analysis model based on capsule-BiGRU

本文提出的方法首先使用预训练的 GloVe 模型，将两个文本分别映射为 300 维的词向量矩阵。将词向量矩阵作为模型的输入，经过注意力机制模块赋予权重，然后将其结果分别输入到 BiGRU 网络和胶囊网络模型中。在胶囊网络中，首先进行卷积运算，经过主胶囊层做胶囊卷积运算，经挤压函数运算后作为主胶囊层的输出，经过动态路由协议机制运算后连接到分类胶囊层，将分类胶囊层的输出结果展开作为文本的局部特征向量。在 BiGRU 网络中，使用双向的 GRU 网络从两个方向提取文本的信息，得到文本的全局特征向量。同时在特征向量提取阶段，使用孪生神经网络结构，即处理两个词向量矩阵使用完全相同的网络结构，这样将两个词向量矩阵编码到同一矢量空间。最后将两个文本各自的局部特征和全局特征分别进行相似度分析，得到两个文本的相似度矩阵，将相似度矩阵作为全连接网络的输入，全连接网络最后一层使用 sigmoid 函数作为分类器，判断两个文本是否相似。

3.1 词向量嵌入模块

在词向量嵌入模块首先对文本进行预处理，主要包括去停用词、特殊符号等，通过分析所有文本，本实验选择句子最大长度为 25 个字符，对不足 25 个字符的句子进行补齐，超过 25 个字符的句子截取前 25 个字符作为句子表示。使用了斯坦福大学自然语言处理小组预训练的 GloVe 模型将文本中每个单词映射为 300 维词向量。

GloVe 模型对单词进行向量化表示过程如下：首先计算语料库的共现矩阵X，其中X_{ij}为在语料库中单词i与单词j共同出现在同一窗口中的次数。

$$X_i = \sum_{j=1}^{N} X_{ij} \qquad (1)$$

$$P_{ij} = \frac{X_{ij}}{X_i} \qquad (2)$$

X_i表示单词i在语料库中出现的次数，P_{ij}表示单词j在单词i的语境中出现的概率。假设已经知道单词i和j的词向量分别是v_i和v_j，计算v_i与v_j的相似度与P_{ij}进行比较，当差值较小的时候证明词向量与共现矩阵一致性较高，词向量对上下文信息把握准确。

$$J = \sum_{i,j}^{N} f(X_{ij})\left(v_i^T v_j + b_i + b_j - log(X_{ij})\right)^2 \qquad (3)$$

使用代价值J表示两项的差值，b_i与b_j为偏差项。通过迭代的更改所有单词的词向量使得代价值J在整个语料库中最小，即得到了语料库中所有单词最优的词向量，这样通过上下文信息计算出单词的词向量。斯坦福大学自然语言处理小组收集了维基百科网站上的数据集作为语料库进行了词向量训练，该数据集包含大量的英文文本，预训练得到的词向量中包含更准确的上下文信息，发布了 50 维，100 维 200 维与 300 维词向量训练结果。本文选用斯坦福大学自然语言处理小组发布的 300 维词向量作为词向量表示。

3.2 特征矩阵提取模块

3.2.1 注意力

在自然语言处理中，传统的注意力模型主要是分析文本中与任务更相关的单词，从而赋予较高的注意力，这样的注意力模型在处理单个句子的任务时会有较好的表现。但针对本文的任务——文本相似度而言，主要关注两个文本是否相似，对于两个输入的文本t_1和t_2而言，更应该关注的是t_1与t_2相似的部分，对于相似的部分给予更高的注意力[23]。计算t_1中任意一个单词与t_2中所有单词之间的相似度并求和，相似度计算方法使用余弦相似度，将余弦相似度的和作为描述该单词的权重的值。假设文本t_1和t_2经过词向量嵌入层得到的词向量矩阵为：

$$v_{t1} = \left(w_1^1, w_2^1, w_3^1, w_4^1, w_5^1\right) \qquad (4)$$

$$v_{t2} = \left(w_1^2, w_2^2, w_3^2, w_4^2, w_5^2\right) \qquad (5)$$

其中w_1^j表示文本t_1中第 j 个单词的词向量。余弦相似度计算公式如下。

$$\cos(a,b) = \frac{a \cdot b}{|a| * |b|} \qquad (6)$$

根据前式文本t_1和t_2的词向量矩阵为v_{t1}, v_{t2}，

利用余弦相似度计算公式计算两个输入文本所有单词与另一文本的相似程度。

$$k_{t1}[i]=\sum_{j}\cos\left(w_i^1,w_j^2\right) \quad (7)$$

$$k_{t2}[i]=\sum_{j}\cos\left(w_j^1,w_i^2\right) \quad (8)$$

式中 $k_{t1}[i]$ 为文本 t_1 中第 i 个单词与文本 t_2 各个单词的余弦相似度和，通过计算得到文本 t_1 与 t_2 中各个单词的余弦相似度 k_{t1}、k_{t2} 并作为计算各个单词权重的值。使用 k_{t1}、k_{t2} 以及 $SoftMax$ 函数完成单词权重的计算。

$$A_{t1}=SoftMax\left(k_{t1}\right) \quad (9)$$

$$A_{t2}=SoftMax\left(k_{t2}\right) \quad (10)$$

A_{t1}、A_{t2} 为文本 t_1 和 t_2 各个单词对应的权值，将单词的词向量与对应权值相乘得到文本的特征矩阵，作为后续网络的输入。

3.2.2 胶囊网络

在文本中存在大量的冠词，连词，感叹词等与文本语义无关的单词，这些单词在两个文本中有极高的概率同时存在，经过注意力模块运算后这些单词可能得到较高的权重，但这些单词对文本的语义没有较大的影响，赋予较大的权重会对最后的结果有一定的影响。在胶囊网络模块中称这些无关单词为噪声胶囊。使用 NLTK 工具对句子中的单词进行词性标注，在胶囊网络中首先根据单词词性对限定词、连词、感叹词、代词赋予较低权重，以减轻噪声胶囊对后续任务的影响，解决上述问题。将经过注意力机制的特征矩阵输入胶囊网络，使用动态路由算法计算上层胶囊输出，计算步骤如下。

a) $A_i=attention\left(u_i\right)$

b) $b_{ij}=0$

c)迭代 r 次：

$$c_{ij}=\frac{e^{b_{ij}}}{\sum_{k}e^{b_{ik}}}$$

$$u_{(j|i)}=w_{ij}A_i$$

$$s_j=\sum_{i}c_{ij}u_{(j|i)}$$

$$squash\left(k\right)=\frac{\left\|k\right\|^2}{1+\left\|k\right\|^2}\frac{k}{\left\|k\right\|}$$

$$v_j=squash\left(s_j\right)$$

$$b_{ij}=b_{ij}+u_{(j|i)}v_j$$

d)返回 v_j

其中 u_i 为互注意力模块得到的特征向量，A_i 为降低噪声胶囊权重之后的特征向量，r 为动态路由算法迭代次数，w_{ij} 为两层胶囊之间的权值矩阵，c_{ij} 为耦合系数，表示下层胶囊 i 激活上层胶囊 j 的可能性，$u_{(j|i)}$ 为上层胶囊的输入，$squash$ 为激活函数，v_j 为上层胶囊的输出。动态路由算法将 b_{ij} 的初始值设为 0，这样 v_j 的初始值为 $u_{(j|i)}$ 的均值，通过迭代更新 b_{ij}，从而更新 c_{ij} 与 v_j 的值。w_{ij} 为神经网络模型的参数，模型通过大量的训练数据学习 w_{ij} 的值。Sabour 在文中提出的胶囊网络包括三层结构，分别为：卷积层、PrimaryCaps 层、DigitCaps 层。在本文提出的方法中，使用 DigitCaps 层的输出作为文本的局部特征矩阵。

3.2.3 BiGRU

双向门控循环单元网络（BiGRU）是一种双向的基于门控的循环神经网络，由前向 GRU 与后向 GRU 组合而成。通过两个方向的网络遍历文本，得到包含文本上下文的信息，解决了 GRU 模型只能包含上文信息的问题。GRU 模型是长短期记忆网络（LSTM）的变体。相较于 LSTM，GRU 模型网络结构较简单，但效果与 LSTM 基本相同，大大减少了网络训练所需的时间。循环神经网络当前时间步的输出与前面时间步的输出有关，这使循环神经网络具有记忆性，适合处理序列数据。但传统的神经网络只具有短期记忆，对于长距离的依赖关系效果不好，同时存在梯度爆炸或梯度消失的问题。LSTM 通过门控机制解决了上述问题，可以学习跨度较长的依赖关系。LSTM 神经元结构如图 2 所示。

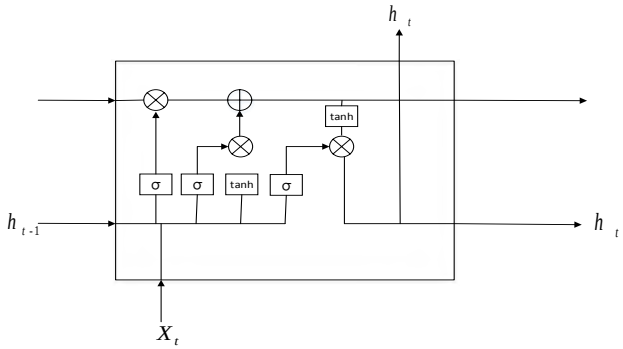

图 2　LSTM 神经元结构图

Fig.2　LSTM neuron structure diagram

GRU 网络将 LSTM 中输入门和遗忘门合并，称为更新门，这使训练网络所需的时间大大减少。GRU 神经元结构如图 3 所示。

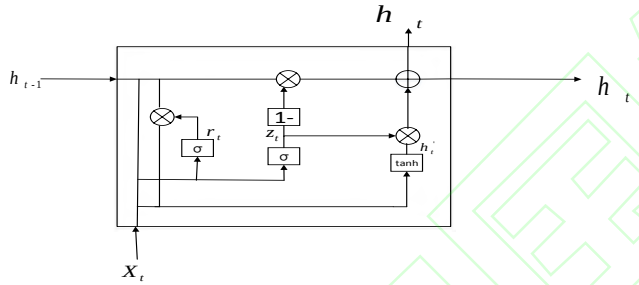

图 3　GRU 神经元结构图

Fig.3　GRU neuron structure diagram

在 GRU 网络中，更新门控制当前时刻的隐状态 h_t 中保留多少历史时刻的隐状态和当前时刻的候选状态。重置门的作用是决定当前时刻的候选状态 $h_t^{'}$ 与上一时刻的隐状态之间的依赖程度。

$$z_t = \sigma\left(w_z x_t + u_z h_{t-1} + b_z\right) \quad (11)$$

$$r_t = \sigma\left(w_r x_t + u_r h_{t-1} + b_r\right) \quad (12)$$

$$h_t^{'} = \tanh\left(w_c x_t + u_c\left(r_t \odot h_{t-1}\right) + b_c\right) \quad (13)$$

$$h_t = z_t \odot h_{t-1} + \left(1 - z_t\right) \odot h_t^{'} \quad (14)$$

$$y_t = \sigma\left(W_0 \cdot h_t\right) \quad (15)$$

x_t 为当前时刻的输入，h_{t-1} 为上一时刻的隐状态，$h_t^{'}$ 为当前时刻的候选状态，h_t 为当前时刻的隐状态，y_t 为当前时刻输出。公式 8 为更新门的计算公式，公式 9 为重置门的计算公式。

在 GRU 网络中信息只能单向传递，但在实际中每个单词可能与上下文中的单词都有依赖关系，使用 BiGRU 网络通过两个方向的网络训练文本，使得模型的效果更好，BiGRU 网络结构如图 4 所示。本文提出的方法使用 BiGRU 网络的输出作为文本的全局特征矩阵[24]。

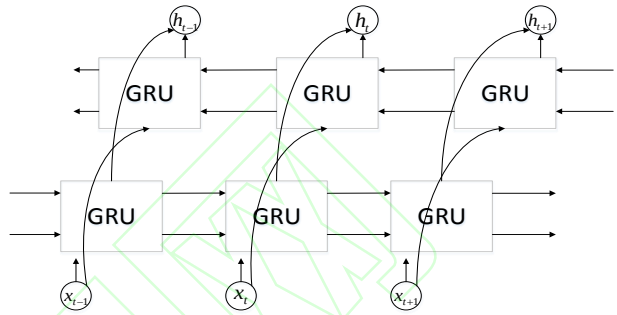

图 4　BiGRU 网络结构图

Fig.4　BiGRU network structure diagram

3.3　特征矩阵分析判断模块

将两个文本的局部特征矩阵与全局特征矩阵分别进行相似度计算，得到局部特征的相似度矩阵 E_1 和全局的相似度矩阵 E_2。其中 E_1 和 E_2 的计算方法相同，这里介绍 E_1 的计算方法。假设两个文本的局部特征分别为 S_1 和 S_2，E_1 计算公式如下。

$$E_1^{ij} = \cos\left(S_1^i, S_2^j\right) \quad (16)$$

E_1^{ij} 为相似度矩阵第 i 行第 j 列元素，S_1^i 为 S_1 的第 i 行，S_2^j 为 S_2 的第 j 行。在得到相似度矩阵后将两个相似度矩阵展平并连接。将融合后的相似度向量作为全连接层的输入，将全连接网络输出与 sigmoid 分类器连接。使用 sigmoid 分类器判别两个文本是否相似。

4　实验结果与分析

4.1　数据集

为了评估模型在文本相似度任务上的表现，本文使用了 Quora Question Pairs 数据集和 MRPC (Microsoft Research Paraphrase Corpus)数据集进行实验。

Quora Question Pairs 数据集包含 404000 个句子对，相似的句子对标签为 1，否则为 0。在本文的实验中将数据集进行分割，80%作为训练集，10%

作为测试集，10%作为验证集。MRPC(Microsoft Research Paraphrase Corpus)数据集包括 4076 个训练样本和 1725 个测试样本,相似的句子对标签为 1，否则为 0。

4.2 实验设置

本文进行的实验基于 keras 框架实现，使用 Adam 优化器，在 Quora Question Pairs 数据集进行的实验模型参数设置如表 1 所示。

表 1 实验参数设置

Table 1 Experimental parameter settings

参数名	参数值
Epoch	25
Batchsiz	512
Dropout	0.3
胶囊维度	64
动态路由迭代次数	3
BiGRU 神经元	100

4.3 评价指标

本文实验的性能评价指标主要包括：准确率、精确率、召回率、F1 值。设 TP 为将正确类预测为正确类的个数；TN 为将错误类预测为错误类的个数；FP 为将错误类预测为正确类的个数；FN 为将正确类预测为错误类的个数。评价指标的计算公式如下。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{17}$$

4.4 实验结果分析

验证本文提出方法的有效性，本文进行了三个实验。实验（1）：与深度学习领域主流模型进行对比实验；实验（2）：与其他论文中提出方法进行对比实验；实验（3）：改变 capsule 网络迭代次数进行对比实验；实验（4）在两个数据集上测试模型表现。

实验（1）中选取深度学习领域主流的模型进行比较实验，包括：LSTM、BiLSTM、capsule、GRU、BiGRU、Siamese-capsule、Siamese-BiGRU、capsule-BiGRU,使用上述模型进行实验，实验结果如表 2 所示。

表 2 实验（1）数据结果

Table 2 Experiment (1) Data results

模型	准确率(%)	精确率(%)	召回率(%)	F1 值(%)
LSTM	80.08	82.41	86.24	84.28
BiLSTM	81.95	81.97	88.12	84.93
GRU	80.11	83.81	84.58	84.19
BiGRU	81.95	83.08	87.71	85.33
Siamese-BiGRU	84.47	86.07	89.02	87.52
capsule	81.91	84.74	86.38	85.55
Siamese-capsule	83.79	88.37	86.31	87.33
capsule-BiGRU	**86.16**	**86.56**	**91.11**	**88.77**

从表 2 中可以看出，与传统的 CNN、LSTM 网络相比，本文提出的模型在文本相似度任务中表现更好。GRU 网络与 LSTM 网络在任务中的表现基本相同，但在相同的网络规模，训练 GRU 网络所需时间远小于训练 LSTM 网络。通过对比 capsule 与 Siamese-capsule，BiGRU 与 Siamese-BiGRU 的表现发现，Siamese-BiGRU 网络相较于 BiGRU 网络，准确率提升了 2.52%，精确率提升了 2.99%，召回率提升了 1.31%，F1 值提升了 2.19%。Siamese-capsule 网络相较于 capsule 网络，准确率提升了 1.88%，精确率提升了 3.63%，召回率提升了 1.93%，F1 值提升了 1.78%。从中可以发现孪生神经网络结构可以有效的提高模型的表现。本文提出的方法在准确率、精确率、召回率、F1 值上的表现都优于传统的神经网络模型。相较于传统的 LSTM 模型准确率提高了 6.08%，F1 值提高了 4.49%。

实验（2）中将本文提出方法与其他论文所提方法进行比较，对比结果如表 3 所示。

表 3 实验（2）数据结果

Table 3 Experiment (2) Data results

模型	准确率(%)	F1 值(%)
capsule-BiGRU	**86.16**	**88.77**
CNN-BiLSTM[20]	84.58	85.02
BiLSTM-DenseNet[22]	85.50	87.10

通过比较可以发现，本文提出的方法相较于文献[19]提出的模型准确率提高了 1.58%、F1 值提高了 3.75%。相较于文献[21]提出的模型准确率提高了 0.66%，F1 值提高了 1.67%，该模型使用了 6 层堆叠的 BiLSTM 网络，模型较为复杂，训练所需时间较长。

实验（3）中改变 capsule 网络中动态路由算法迭代次数做对比实验，实验结果如表 4 所示。

表 4　实验（3）数据结果

Table 4　Experiment (3) Data results

迭代次数	准确率(%)	精确率(%)	召回率(%)	F1 值(%)
1	83.37	87.60	86.30	86.94
2	83.68	86.09	87.85	86.96
3	**83.79**	**88.37**	**86.31**	**87.33**
4	83.58	85.72	87.38	86.54
5	82.34	87.97	84.69	86.30
6	79.85	85.58	83.25	84.30
7	78.57	86.31	81.02	83.58
8	77.97	83.08	82.24	82.26

基于上述实验结果可知，动态路由算法迭代次数对 capsule 网络有一定影响。随着迭代次数的增加，训练模型所需时间不断增加。当动态路由算法迭代次数设置为 3 时，模型有较好的表现且训练时间为 198min，迭代次数超过 3 次后，模型表现逐渐下降。在本文的其他实验中胶囊网络动态路由迭代次数设置为 3，以获得更好的表现。

表 5　实验（4）数据结果

Table 5　Experiment (4) Data results

数据集	准确率(%)	精确率(%)	召回率(%)	F1 值(%)
Quora Question Pairs	86.16	86.56	91.11	88.77
MRPC	87.88	83.04	81.21	82.12

由于 MRPC 数据集样本较少，所以调整了 Dropout 参数为 0.1，其他模型参数不做调整。在表 5 中可以看出，模型在 Quora Question Pairs 数据集上表现更加出色，主要是因为 Quora Question Pairs 数据集中的样本数量更多，模型训练更加完善，说明本文提出的模型的表现比较依赖于数据集中样本的数量。

5　结束语

针对文本相似度任务，本文提出基于 capsule-BiGRU 的文本相似度分析方法。capsule 网络可以有效地提取文本的局部特征向量，BiGRU 网络使用双向的循环网络结构从两个方向遍历整个文本，从而有效地提取上下文信息得到文本的全局特征矩阵，对两个文本的特征矩阵进行相似度分析，判断文本是否相似。实验表明本文提出的方法对文本相似度任务而言有更好的效果。

参考文献:

[1] 王春柳,杨永辉,邓霏,赖辉源.文本相似度计算方法研究综述[J].情报科学,2019,37(3):158-168.

WANG Chunliu,YANG Yonghui,DENG Fei,LAI Huiyuan. A Review of Text Similarity Approaches[J]. Information Science,2019,37(3):158-168.

[2] Bouvrie J. Notes on convolutional neural networks[J]. 2006.

[3] Boden M. A guide to recurrent neural networks and backpropagation[J]. the Dallas project, 2002.

[4] Zhang Y, Jin R, Zhou Z H. Understanding bag-of-words model: a statistical framework[J]. International Journal of Machine Learning and Cybernetics, 2010, 1(1-4): 43-52.

[5] Salton G, Wong A, Yang C S. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.

[6] 张璐,芦天亮,杜彦辉.基于 WMF_LDA 主题模型的文本相似度计算[J].计算机应用研究, 2019, 36(10): 2916-2919+2951.

ZHANG Lu,LU Tianliang,DU Yanhui. Text similarity calculation based on WMF_LDA topic model[J].Application Research of Computers,2019,36(10):2916-2919+2951.

[7] MikolovT , Chen K , Corrado G , et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer ence, 2013.

[8] 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 (EMNLP), 2014.

[9] Kim Y. Convolutional Neural Networks for SentenceClassification[J]. EprintArxiv, 2014.

[10] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Advances in neural information processing systems. 2017: 3856-3866.

[11] Zhao W, Ye J, Yang M, et al. Investigating capsule networks with dynamic routing for text classification[J]. arXiv preprint arXiv:1804.00538, 2018.

[12] Francis-Landau M, Durrett G, Klein D. Capturing semantic similarity for entity linking with convolutional neural networks[J]. arXiv preprint arXiv:1604.00734, 2016.

[13] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]//Eleventh annual conference of the international speech communication association. 2010.

[14] Sundermeyer M, Schlüter R, Ney H. LSTM neural networks for language modeling[C]//Thirteenth annual conference of the international speech communication association. 2012.
[15] Neculoiu P, Versteegh M, Rotaru M. Learning text similarity with siamese recurrent networks[C]//Proceedings of the 1st Workshop on Representation Learning for NLP. 2016: 148-157.
[16] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
[17] He H, Wieting J, Gimpel K, et al. UMD-TTIC-UW at SemEval-2016 Task 1: Attention-based multi-perspective convolutional neural networks for textual similarity measurement[C]//Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 2016: 1103-1108.
[18] 方炯焜,陈平华,廖文雄.结合 GloVe 和 GRU 的文本分类模型[J/OL].计算机工程与应用：1-9 [2020-04-17]. http://kns.cnki.net/kcms/detail/11.2127.tp.20200331.1742.006.html.
FANG Jiongkun,CHEN Pinghua,LIAO Wenxiong. Text classification model based on GloVe and GRU[J/OL].Computer Engineering and Applications: 1-9 [2020-04-17]. http://kns.cnki.
net/kcms/detail/11.2127.tp.20200331.1742.006.html.
[19] Pontes E L, Huet S, Linhares A C, et al. Predicting the semantic textual similarity with siamese CNN and LSTM[J]. arXiv preprint arXiv:1810.10641, 2018.
[20] 郭浩,许伟,卢凯,唐球.基于 CNN 和 BiLSTM 的短文本相似度计算方法[J].信息技术与网络安全，2019，38(6)：61-64+68.
GUO Hao,XU Wei,LU Kai,TANG Qiu. Short text similarity computation method based on hybrid CNN and BiLSTM[J]. Information Technology and Network Security,2019,38(6):61-64+68.
[21] 唐庄,王志舒,周爱,冯美姗,屈雯,鲁明羽.面向文本分类的 transformer-capsule 集成模型[J/OL].计算机工程与应用:1-7[2020-04-11].http://kns.cnki.net/kcms/detail/11.2127.TP.20191219.0859.002.html.
TANG Zhuang, WANG Zhishu, ZHOU Ai, FENG Meishan, QU Wen, LU Mingyu.Transformer-capsule Integrated Model for Text Classification[J/OL].Computer Engineering and Applications:1-7 [2020-04-11]. http://kns.cnki.net/kcms/detail/11.2127.TP.20191219.0859.002.html.
[22] 尹春勇,何苗.基于改进胶囊网络的文本分类[J/OL].计算机应用: 1-7 [2020-08-14]. http://kns.cnki.net/kcms/detail/51.1307.TP.20200706.1614.018.html.
YIN Chunyong, HE Miao. Text classification based on improved capsule network[J/OL].Journal of Computer Applications:1-7
[2020-08-14].http://kns.cnki.net/kcms/detail/51.1307.TP.20200706.1614.018.html.
[23] Chen Z, Wang X, Xie X, et al. Co-Attentive Multi-Task Learning for Explainable Recommendation[C]//IJCAI. 2019: 2137-2143.
[24] 刘继明,于敏敏,袁野.基于句向量的文本相似度计算方法[J].科学技术与工程,2020,20(17):6950-6955.
LIU Jiming,YU Minmin,YUAN Ye. Computing Method of Text Similarity Based on Sentence Vector[J]. Science Technology and Engineering,2020,20(17):6950-6955.
[25] 靳丽．结合神经网络的文本语义相似度研究[D].山东大学,2019.
JIN Li. Text Semantic Similarity Research Based on Neural Network[D]. Shandong University, 2019.