计算机科学与探索
Journal of Frontiers of Computer Science and Technology
ISSN 1673-9418,CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：面向网络文本的 BERT 心理特质预测研究
作者：张晗，贾甜远，骆方，张生，邬霞
网络首发日期：2020-10-23
引用格式：张晗，贾甜远，骆方，张生，邬霞．面向网络文本的 BERT 心理特质预测研究．计算机科学与探索．
https://kns.cnki.net/kcms/detail/11.5602.TP.20201022.1717.002.html

计算机科学与探索
Journal of Frontiers of Computer Science and Technology doi: 10.3778/j.issn.1673-9418.2007009

面向网络文本的 BERT 心理特质预测研究

张 晗[1]，贾甜远[1]，骆 方[2]，张 生[3]，邬 霞[1,4+]

1. 北京师范大学 人工智能学院，北京 100875
2. 北京师范大学 心理学部，北京 100875
3. 北京师范大学 中国基础教育质量监测协同创新中心，北京 100875
4. 智能技术与教育应用教育部工程研究中心，北京 100875
+ 通信作者 E-mail: wuxia@bnu.edu.cn

摘 要: 随着互联网的普及应用，通过网络平台进行表达和交流的用户越来越多，在此过程中不可避免地会留下与个人相关的大量网络文本数据和信息，这些非结构化的文本数据往往体现着不同场景下的真实表达，反映了人们内在的心理特质及人格倾向。利用文本挖掘相关技术基于网络文本数据分析心理特质可以弥补传统心理测量方法易受应试动机等因素影响的缺陷。近年来，BERT（Bidirectional Encoder Representations from Transformers）语言表示模型在文本分类、情感分析等任务上取得了很好的效果。通过针对网络文本数据构建心理特质预测模型，基于 BERT 获取完整的上下文语义特征和长距离的上下文依赖关系；同时考虑到分类器内部结构的差异可能会导致不同的分类效果，在下游分类任务中分别采用 $BERT_{BASE}$ 模型的全连接层和经典的随机森林算法作为两种不同的分类器进行模型效果对比。结果显示，基于 BERT 的文本分类模型能够有效实现心理特质的预测，平均准确率、平均精准率等各项指标都在 97%以上。

关键词: BERT；心理特质；注意力机制；Transformer；文本挖掘
文献标志码：A **中图分类号：**TP391

张晗，贾甜远，骆方，等. 面向网络文本的 BERT 心理特质预测研究[J]. 计算机科学与探索
ZHANG H, JIA T Y, LUO F, et al. A Study on Predicting Psychological Traits of Online Text by BERT [J]. Journal of Frontiers of Computer Science and Technology

A Study on Predicting Psychological Traits of Online Text by BERT

ZHANG Han[1], JIA Tianyuan[1], LUO Fang[2], ZHANG Sheng[3], WU Xia[1,4+]

1. School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China
2. Faculty of Psychology, Beijing Normal University, Beijing 100875, China
3. National Innovation Center for Monitoring Basic Education Quality, Beijing Normal University, Beijing 100875, China
4. Engineering Research Center of Intelligent Technology and Educational Application, Ministry of Education, Beijing 100875, China

*The National Key Research and Development Program of China under Grant No. 2018YFC0810602 (国家重点研发计划课题).

Abstract: With the rapid development and popularity of the Internet, an increasing number of people would like to use online platforms to express themselves and communicate with others. It is inevitable that a large amount of online text data is constantly emerging with personal information, which often indicates individual real expression in different conditions and reflects personal inner psychological traits and personality tendency. Applying text mining techniques to analyzing psychological traits behind the online text is not only helpful for individuals to understand themselves, but also useful to avoid the motivation interfere while using traditional methods for psychological assessment. In recent years, the language model named bidirectional encoder representations from transformers (BERT) has greatly improved the performance of both the text classification task and the sentiment analysis task. In this paper, prediction models for psychological traits were constructed based on online text. Comprehensive semantic features and long dependency in the context were obtained by BERT. Considering that distinct algorithm frameworks of classifiers can lead to different classification results, the fully-connected layer of the $\text{BERT}_{\text{BASE}}$ model and the random forest algorithm were used in the downstream classification task to make comparison. The results showed that psychological traits can be effectively predicted from text classification based on BERT.

Key words: BERT; psychological trait; attention mechanism; Transformer; text mining

1 引言

随着信息时代的飞速发展，利用网络平台进行学习、工作、娱乐、社交等活动已经成为人们日常生活中必不可少的一部分，极大地丰富了人们表达自我、记录思想和交流沟通的方式。在互联网中的海量数据中，非结构化文本与信息数据占据了很大一部分，用户的文本信息往往是其在自然状态而非测试状态下的真实表达。研究表明，个体所使用的语言文本往往在一定程度上反映着其心理状态或某些特质倾向[1,2]，例如羞怯特质（Shyness）、合作性特质(Cooperativeness)、完美主义特质（Perfectionism）、焦虑特质（Anxiety）等。这些心理特质的差异会对一系列的个人行为（如：个人决策、职业能力等）产生多重影响[3,4]，通过对心理特质进行分析评价和持续监测能够帮助深入了解心理状态和人格特质，在发展过程中发现行为变化中所存在的问题和原因并及时加以解决。因此，对网络文本数据进行分析，可以为预测心理特质提供非常有用的参考价值，弥补传统的心理测量方法易受应试动机等因素影响的缺陷。

近年来，深度学习[5]相关技术在计算机视觉、模式识别、自然语言处理等领域得到了广泛应用[6-8]，因其在很大程度避免了传统文本分类方法中存在的诸多问题，所以也越来越多地应用于文本挖掘。具体来说，深度学习能够将文本数据转换为适合神经网络处理的格式来有效进行文本表示，并通过特定的神经网络结构自动获取关键特征进而避免了人工特征工程的复杂过程，从而在大规模文本分类及情感分析领域中表现出了很好的性能，进一步提高了模型分类的精度[9,10]。尤其是 2018 年出现的深度神经网络语言模型 BERT(Bidirectional Encoder Representations from Transformers)[11]，它采用多头注意力机制[12]，不仅能够准确提取字符级别和词级别的信息，而且可以充分捕捉句内关系和句间关系，具有很强的模型泛化能力和鲁棒性，在包括文本分类在内的多种自然语言处理任务中都表现出优异的模型性能[12-15]。

本文针对网络文本数据，提出了基于 BERT 的心理特质预测模型，主要贡献如下：

（1）将 BERT 语言框架应用于心理特质预测，

利用BERT双向训练的模式及Transformer的编码模块挖掘更加完整的上下文语义特征和更长距离的上下文依赖关系，解决了心理特质语义特征的增强向量表示的问题；

（2）在下游分类任务中分别采用 $BERT_{BASE}$ 模型的全连接层和基于集成学习原理的随机森林算法作为两种分类器，避免分类器多样性较差而造成的分类准确率受限、模型性能欠佳等问题。

2 相关工作

在针对大五人格特质预测的研究中，以往采用的文本挖掘方法大多依赖于传统的机器学习算法。例如，Kwantes 等[16]在 2016 年利用潜在语义分析的算法对被试所写的关于在特定场景中自我感受的文章进行内容处理，并结合被试的量表信息，实现了包括开放性、外倾性、神经质性/情绪稳定性在内的三种人格特质预测。近年来，基于深度学习的神经网络结构在文本挖掘领域的应用使得心理特质与人格预测研究有了进一步突破，主要采用的网络结构包括经典的卷积神经网络（Convolutional Neural Networks, CNN）、循环神经网络（Recurrent Neural Network, RNN）及其对应的网络结构变种。例如，Wei 等[17]在 2017 年提出了一个用于预测微博用户个性特征的异构信息集成框架，通过收集用户所的语言文本数据、头像、表情和互动模式等异构信息，采用 CNN 的改进结构 Text-CNN、Responsive-CNN 以及词袋聚类等多种不同的策略来进行语义表示和特征提取，不仅很好地实现了开放性、尽责性、外倾性、宜人性以及神经质性这五种人格的预测，而且表现出了优于其他传统模型的性能。Majumder 等[18]提出了一种基于 CNN 从文章中提取大五人格特征的模型，将文章中的句子输入到卷积滤波器中得到 N-Gram 特征向量形式的句子模型，在大五人格预测实验中也表现出了良好的模型性能。

由此可见，深度学习的神经网络算法在基于文本挖掘的情感分析及心理特质预测研究中具有明显的可行性及有效性。但需要指出的是，上述研究所建立的情感分析及心理特质预测模型，大多是基于经典的神经网络算法，这些网络结构本身的限制在很大程度上会使模型性能受限。而深度学习领域的更新迭代速度之快使得许多广泛应用于文本分类的神经网络算法出现了更加完善的改进结构及变种形式，尤其是近几年提出的注意力机制以及 BERT 语言模型，在包括句子级和词条级在内的 11 项自然语言处理任务中都表现出了超越其他技术的效果[11,19]，对文本分类任务的提升效果也极为显著。但是这些最新的网络结构及算法改进往往仅在自然语言处理领域得到了较大范围的应用，并未引入心理特质预测研究中。因此，本文基于注意力机制的原理，利用 BERT 算法并对其下游结构进行微调，构建基于 BERT 的文本分类模型，用于实现羞怯、合作性、完美主义、焦虑这四种心理特质的预测。

3 基于 BERT 的心理特质预测模型设计

3.1 BERT 算法基本原理

BERT 的全称为 Bidirectional Encoder Representations from Transformers，可译为“来自变换器的双向编码器表示”，由 Google 在 2018 年年底提出，本质上是一种基于 Transformer 架构且能够进行双向深度编码的神经网络语言模型[11,12]。在 BERT 内部组成结构中，最为关键的部分是 Transformer 的解码模块。Transformer 由编码器（Encoder）和解码器（Decoder）两部分组成，Encoder 用于对输入的文本数据进行编码表示，Decoder 用于生成与输入端相对应的预测序列；由于 BERT 作为语言表

示模型使用，所以仅采用了 Transformer 的 Encoder 而未使用 Decoder。BERT 之所以能在文本分类、阅读理解、语言翻译等各类自然语言处理任务中都表现出极强的模型泛化能力和提升效果，关键在于其利用自注意力（Self-Attention）机制的原理，并引入掩蔽语言模型（Masked Language Model, MLM）和下一句预测（Next Sentence Prediction, NSP）两种策略对不同层的上下文联合处理来进行双向深度预训练，以此缓解单向性约束问题，这在以往常用的语言表示模型中是无法实现的。所谓 MLM，简单来说，类似于完形填空的过程，指的是将句子中的词按照随机的形式进行遮蔽（Mask），然后依据上下文语义信息对其进行预测，被遮蔽的词在大多数情况下（80%）将被替换为[MASK]标签，在其他情况下分别替换为随机词（10%）或保留该词不做替换（10%），通过这种 Mask 操作使得每个词的关注度得到提高；NSP 则是根据句子之间的依赖关系对下一个句子进行预测，在句首和句末分别插入[CLS]和[SEP]标签，通过学习句间语义相关性判断某个句子是否为输入句子的下一句，按照 50%的概率进行匹配，适用于问答任务或推理任务的过程。

在 BERT 框架的实现过程中，主要包括预训练和微调两个部分。在预训练过程中，BERT 针对不同的任务对未标记的数据进行训练；在微调阶段，先利用预训练的参数对 BERT 进行初始化之后，再结合下游具体任务的标记数据实现对所有参数的微调过程。正是由于 BERT 的预训练与下游任务结构之间没有太大差别，所以 BERT 具有跨不同任务之间的通用架构。

3.2 模型基本思想

本文基于 BERT 模型，分别采用 $BERT_{BASE}$ 模型的全连接层和经典的随机森林（Random Forest, RF）算法作为分类器，构建了 BERT-B 和 BERT-RF 心理特质分类模型，将四种心理特质预测任务转化为基于 BERT 二元分类的多标签文本分类任务。模型的基本思想是利用BERT本身所具有的MLM对词预测以及 NSP 对两个句子是否有上下文关系进行分类的两种策略使模型能够在更大程度上根据上下文语义实现词的预测，提高纠错能力。且 Transformer 的 Encoder 捕获更长距离的上下文依赖关系，通过双向训练的模式使得模型对语义信息特征的获取更加全面和高效。此外，考虑到心理特质是一种较为复杂的内隐特征，不同的分类器构造可能会影响分类结果的精度，因此，在下游分类任务中分别采用 $BERT_{BASE}$ 模型的全连接层和经典的随机森林算法作为两种不同的分类器，丰富模型的多样性，并对模型结果进行对比分析，寻找预测精度更高的心理特质预测模型。由于四种心理特质所对应的标签之间不存在明显的相互依赖或排斥的关系，故分别针对每一种心理特质训练一个相应的二分类模型。

基于 BERT 的两种模型整体框架结构如图 1 所示。在完成文本数据预处理后并将句子长度裁剪为 512 个字符以内后，首先进入嵌入层（Embedding Layer）将句子按字分割，把单个字符转化为词向量表示的形式；然后将嵌入操作后的结果送入 Transformer 的 Encoder 层，每个 Encoder 层由 Self-Attention 层和 Self-Output 层、Intermediate 层、Output 层组成，其中 Self-Attention 由 Query、Key、Value 三个全连接层组成；接着将最后一层 Encoder 的第一个字符[CLS]的字向量经过池化层（Pooler），经由 tanh 函数处理后得到整个句子的句向量表示；最后将携带强语义表示的句向量送入分类层，其中，BERT-B 模型和 BERT-RF 模型分别采用 $BERT_{BASE}$ 的全连接层和 RF 算法两种方式作为四种心理特质类别的分类器，分别得到每个类别的分类结果，完成对四种心理特质的预测。

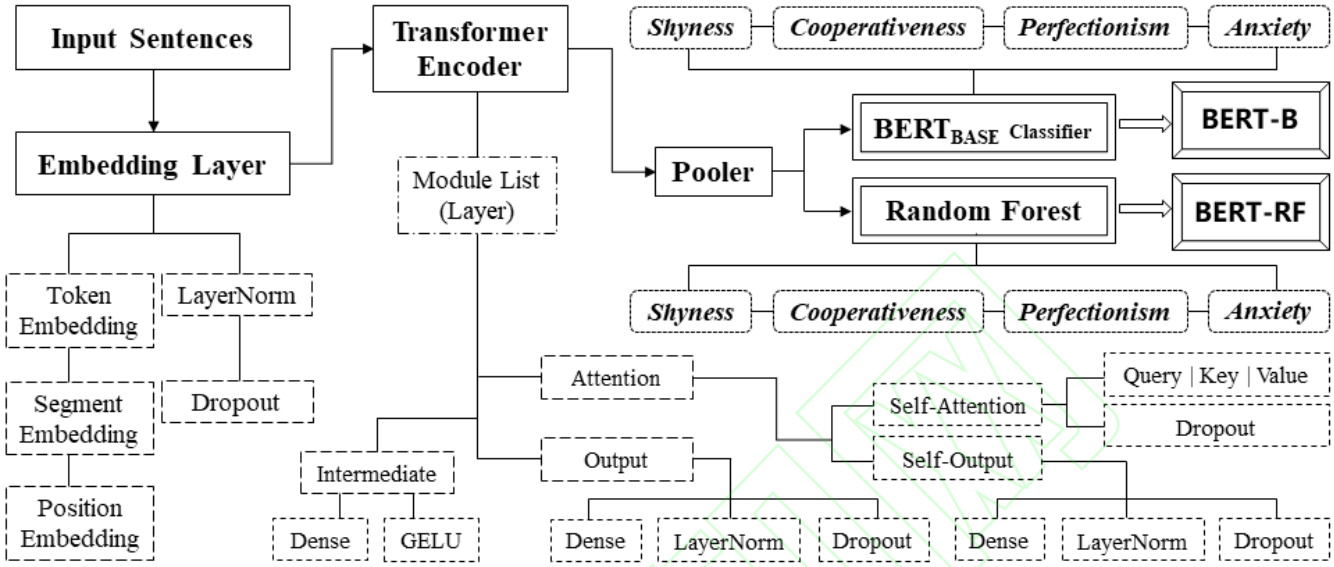


Fig.1 The framework of prediction models for psychological traits based on BERT

图 1 基于 BERT 的心理特质预测模型框架结构

3.3 模型构建

基于 BERT 的 BERT-B 模型与 BERT-RF 模型构建过程主要包括如下步骤：

（1）BERT 输入表示 由于本研究所采用的数据为中文文本，故将单个汉字直接作为细粒度的文本语义单位。每个字的输入表示由其对应的 token embedding、segment embedding 和 position embedding 三个嵌入加和构成。其中，token 表示词/字，token embedding 即为词/字向量表示 word embedding，根据字向量表的查询结果将一个 token 表示为一维向量的形式；segment 表示部分/段，segment embedding 即为区分字/词的语义属于哪个句子的向量表示，判断一个 token 属于左（EA）右（EB）两边的哪一个 segment，此处输入的文本数据为句子而非句对，故只有 EA 没有 EB；position 表示位置，position embedding 即为位置向量表示，根据词/字在文本中所对应的特定位置和顺序，每个 token 都被赋予一个携带了其自身位置信息的向量编码表示。最后将以上 3 个 embedding 相加即得到了 BERT 的线性序列输入表示。图 2 展示了 BERT 输入的嵌入处理可视化表示形式，其中，句首加入的特殊字符[CLS]代表句子的开始，用于下游的文本分类任务；句末的[SEP]表示分割符，使得序列中被打包的句子分隔开来。

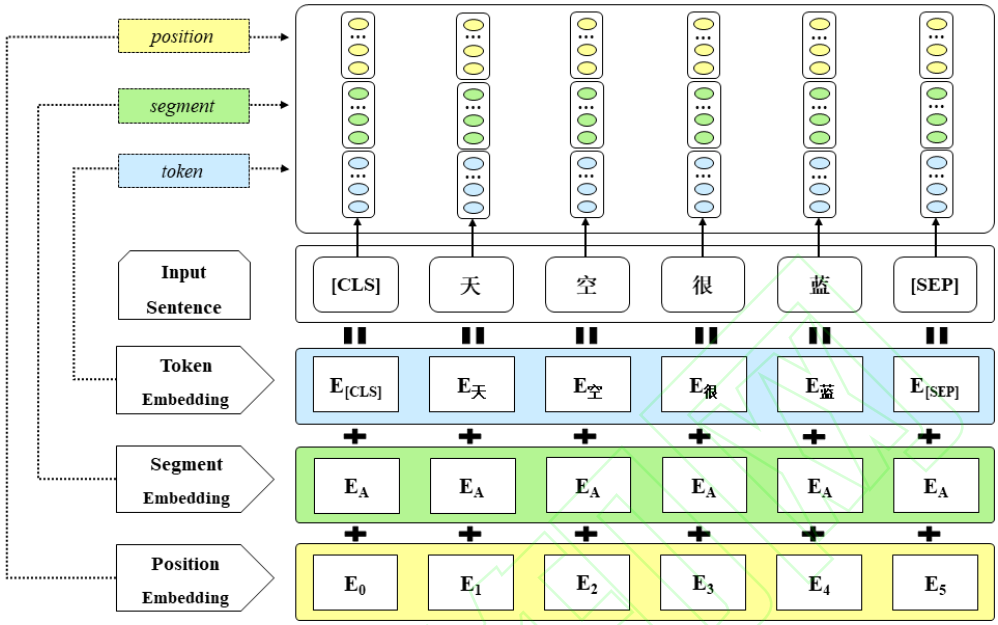


Fig.2 The embedding example of the input

图 2 模型输入表示的嵌入处理示例

（2）**Encoder 注意力计算** 为了挖掘输入文本中句子内部的字词间语义相关性及依赖程度、增强文本序列编码的语义表示，采用 Self-Attention 机制，将文本序列表示为由每个字的查询（Query）以及一系列的键值对<Key, Value>的组合构成，Query=Key=Value，计算每个字 Query 与每个 Key 之间的相关性并求得相应 Value 的权值，完成归一化再进行加权求和操作得出 Attention 的值。为了学习到不同语义场景中的信息表示，使 Attention 的多样性进一步得到扩展，因此在多个语义空间中采用不同的 Self-Attention 构成多头自注意力（Multi-Head Self-Attention）模块，如图 3 所示。Multi-Head Self-Attention 利用多个不同的头来发现不同的关注点，在对 Query、Key、Key 进行线性变换后计算 h 次 Attention，将所有 Attention 进行拼接后再次线性变换得到多头的结果，进而得到与初始输入向量长度相同但包含全文完整语义特征的向量表示，计算公式如下所示[12]：

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_k}})\boldsymbol{V} \quad (1)$$

$$head_i=\mathrm{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}},\boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}},\boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}}) \quad (2)$$

$$MultiHead(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Concat(head_1,head_2,...,head_h)\boldsymbol{W}^{O} \quad (3)$$

其中，$\boldsymbol{W}$ 为线性变换的参数，每进行一次线性变换，$\boldsymbol{W}$ 值也会随之变化。

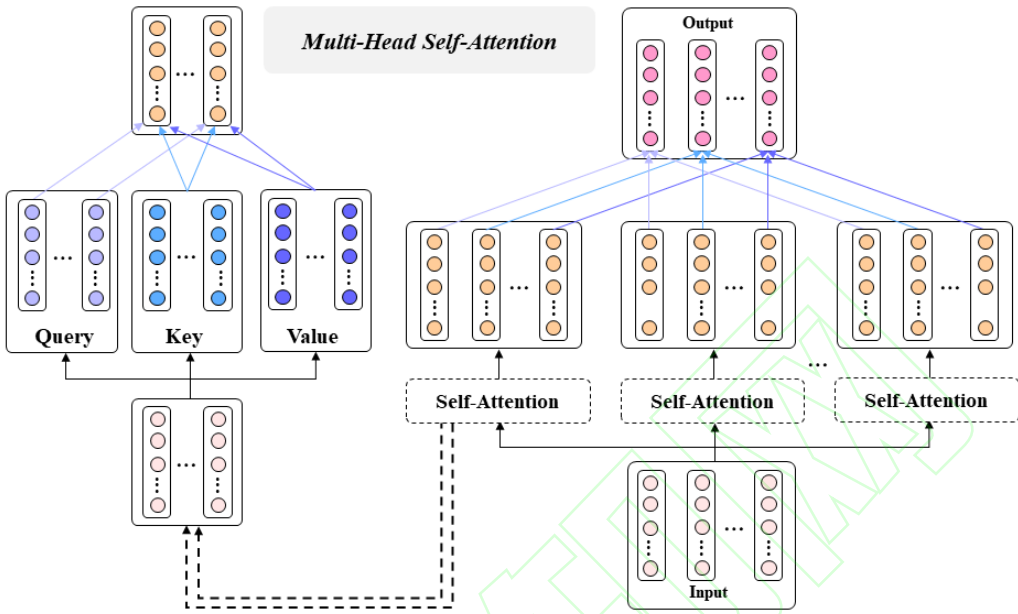


Fig.3 The framework of Multi-Head Self-Attention.

图 3 多头自注意力机制框架

（3） BERT 输出处理 将最后一层 Transformer 的 Encoder 中第一个字符[CLS]的字向量经过池化操作和 tanh 函数处理，即可得到融合了全文字词相关性及全面语义信息的整个句子向量输出表示，将其存储以实现下游的文本分类任务。

（4）分类预测 在分类任务中，考虑到不同分类器的选择可能会导致分类精度的差异，而且有研究针对 179 种分类算法分析后发现基于集成学习的 RF 是分类效果最好的算法之一[20]，因此分别采用经典的机器学习算法 RF 作为一种强分类器，以及深度学习算法中普遍使用的全连接层作为另一种分类器，比较两者基于 BERT 增强语义向量表示的分类结果。

4 实验与结果

4.1 样本数据描述

本研究的被试为教育资源共享与在线互动学习网站“教客网（https://jiaoke.runhemei.com/）”的 2660 名小学生用户（男生 1380 名，女生 1280 名），年龄在 5~14 岁之间（平均年龄为 8.55 岁），来自全国不同省市的 26 所学校，其中，702 名学生来自甘肃省酒泉市的 7 所学校，1844 名学生来自辽宁省大连市的 18 所学校，108 名学生来自四川省成都市的 1 所学校，6 名学生填写的地区及学校信息不详。被试人口学特征如表 1 所示。

对以下小学生用户自 2013 年 6 月 1 日至 2018 年 1 月 18 日在教客网上发布的所有博客日志、评论回复等在线文本进行数据爬取和收集，共得到 160154 篇文本。针对文本数据，按照“正常”“无意”“短无”“短意”“重复”五类标准进行数据清理，各标准的内容范围如表 2 所示。对“正常”“短意”两类标准下的文本予以保留；其他标准下的文本直接剔除。最后得到“正常”“短意”标准下的有效文本为 88659 篇，共计 132323（13 万）条完整语句数据。

针对羞怯、合作性、完美主义以及焦虑四种心理特质，基于北京师范大学心理学部相关专业研究人员确定的对应上述四种心理特质行为表现的常见词语，进一步筛选符合相应心理特质的关键词共 386 个，其中羞怯特质包含 97 个关键词，合作性特质包含 116 个关键词，完美主义特质包含 71 个关键词，焦虑特质包含 102 个关键词。将得到的关键词分别与 13 万条数据进行匹配，若句子中出现某种心理特质所对应的关键词，则该条句子的此类特质标记为正类 1，否则为负类 0。通过标签匹配，使得每条句子带有 4 个标签，即代表四种心理特质。

Table 1 Democratic Characteristics of Subjects

表 1 被试人口学特征

性别		年龄	年级					地区			
男	女	5~14	一	二	三	四	五	甘肃酒泉	辽宁大连	四川成都	其他
1380	1280	平均	1840	9370	1012	4490	780	7020	18440	1080	60
		8.55			2						
总计			2660								

Table 2 Criterions for Data Cleaning

表 2 数据清洗标准

标准	含义	内容范围
正常	正常文本	日常生活、故事创作、课后延伸、主题作文、读后感
无意	无意义文本	无意义文本、乱码等
短无	无意义短文本	小于10个字，无感情色彩文本，例如：已读、已阅
短意	有意义短文本	大于或等于10个字，带有感情色彩的文本，例如：棒、真好
重复	重复文本	与上文重复的文本

4.2 实验设置

通过 Pandas 对语料数据进行处理，将句子长度裁剪为 512 个字符以内，并将句子按字分割后转化为词向量形式，完成词嵌入后进入 12 个 Encoder 层进行多头注意力计算，再将最后一层 Encoder 的第一个字符[CLS]的向量（768 维）经过池化层操作，经由 tanh 函数处理得到整个句子的句向量表示，最后分别利用 BERT-B 模型的全连接层和 BERT-RF 模型的 RF 分类器对携带强语义表示的句向量进行分类，得到每种心理特质的分类结果。由于针对四种心理特质分别训练了 4 个分类器，本质为 4 个二分类任务，故将 softmax 函数直接改为 sigmoid 函数。选取 0.5 作为分类阈值，概率小于 0.5 为负类（标为 0），大于等于 0.5 为正类（标为 1），分别得到每个标签的分类结果进行输出。基于 BERT-B 和 BERT-RF 的心理特质分类预测模的主要参数设置如表 3 所示。

Table 3 Parameters Setting Based on BERT

表 3 基于 BERT 模型的主要参数设置

BERT-B		BERT-RF	
Parameter	**Setting**	**Parameter**	**Setting**
hidden_size	768		
num_hidden_layers	12	min_samples_split	2
num_attention_heads	12		
intermediate_size	3072		
intermediate_act_fn	gelu	min_samples_leaf	1
hidden_dropout_prob	0.1		
attention_probs_dropout_prob	0.1		
initializer_range	0.02	min_weight_fraction_leaf	0
max_position_embeddings	512		

此外，将其他经典的深度学习算法 CNN、RNN 的变种双向长短期记忆（Bidirectional Long Short-Term memory, Bi-LSTM）网络与注意力机制（Attention）结合后的模型 CNN-Attention、Bi-LSTM-Attention 应用于本数据集进行验证，以进一步说明本模型的性能优势。

4.3 模型评价指标

对于分类模型而言，以二分类为例，假设只有 0 和 1 两类，最终的判别结果有四种情况： 真正（True Positive, TP），即被模型预测为正的正样本；假正（False Positive, FP），即被模型预测为正的负样本；假负（False Negative, FN），即被模型预测为负的正样本；真负（True Negative, TN），即被模型预测为负的负样本。本研究在实验中结合预测结果和实际结果，得到 TP、FP、FN 和 TN 的数量，选择如下六个指标对模型性能进行评价。

① 准确率（Accuracy） 分类正确的样本占总样本个数的比例，其计算公式为

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}。 \quad （4）$$

② 精准率（Precision） 模型预测为正的样本中实际也为正的样本占被预测为正的样本的比例，其计算公式为：

$$Precision = \frac{TP}{TP + FP}。 \quad （5）$$

③ 召回率（Recall） 实际为正的样本中被预测为正的样本所占实际为正的样本的比例，其计算公式为：

$$Recall = \frac{TP}{TP + FN}。 \quad （6）$$

④ F1 分数（F1 score） 精确率和召回率的调和平均值，其计算公式为：

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \Rightarrow F_1 = \frac{2TP}{2TP + FP + TN}。 \quad （7）$$

⑤ ROC 曲线（Receiver Operating Characteristic Curve） 通过计算出 TPR 和 FPR 两个值，分别以它们为横、纵坐标作图得到 ROC 曲线。如果 ROC 曲线在坐标系中的位置越接近于左上方，则说明分类模型的性能越好。

⑥ AUC（Area Under Curve） ROC 曲线下的面积，取值范围为[0, 1]。之所以使用 AUC 值作为模型表现的评价指标之一，是因为有些情况下仅凭 ROC 曲线无法清楚判断不同分类模型的优劣。通过 AUC 值使得 ROC 曲线的表现得以量化，AUC 值越大，代表与之对应的分类模型效果越好。

4.4 实验结果及分析

将 13 万条文本数据划分为独立的三个部分，其中，12 万条作为训练集，用于训练分类模型的参数；5 千条作为验证集，用于检验训练过程中模型的状态及收敛情况并调整超参数；5 千条作为测试集，用于评价模型的泛化能力。选取准确率、精准率、召回率、F1 分数、ROC 曲线和 AUC 值六个指标，分别对模型表现做出评价。

对于羞怯、合作性、完美主义、焦虑四种心理特质的分类预测表现，BERT-B 模型和 BERT-RF 模型的准确率、精准率、召回率及 F1 值表现如表 4 所示。从表 4 中可以看出，BERT-RF 模型对于羞怯、合作性、完美主义、焦虑四种心理特质的分类预测都取得了较为理想的效果，每种心理特质所对应的准确率、精准率、召回率及 F_1 值都在 0.97~0.99 之间；其中对于羞怯特质的分类预测效果最为突出，四个指标都高于 0.98。对比发现，相对于 BERT-RF 模型，BERT-B 模型对于四种心理特质的分类预测表现则更加显著，每种心理特质所对应的准确率、精准率、召回率以及 F_1 值全部高于 0.98，尤其是对于合作性特质和羞怯特质的预测所有指标都高达 0.99 以上，模型表现非常优秀。

Table 4 Performance of Predicting Four Psychological Traits

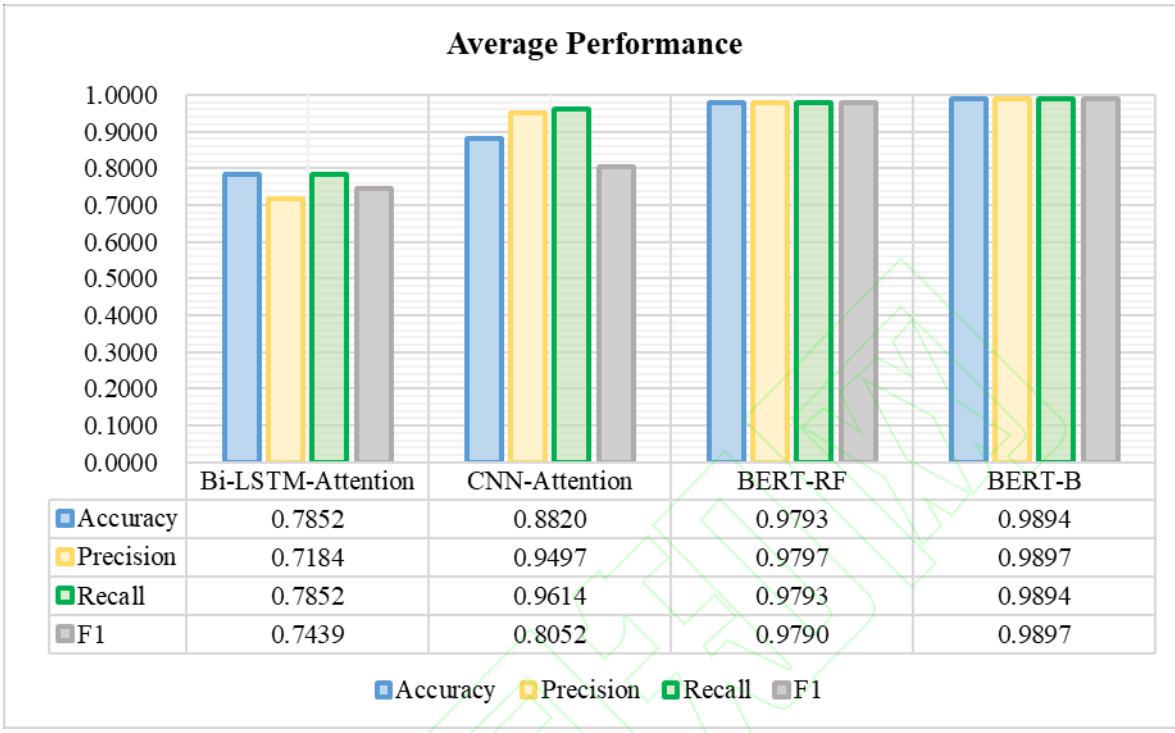
表 4 四种心理特质分类预测的表现

Prediction Model	Performance	Shyness	Cooperativeness	Perfectionism	Anxiety
BERT-B	Accuracy	0.9918	0.9927	0.9861	0.9871
	Precision	0.9918	0.9928	0.9861	0.9882
	Recall	0.9918	0.9927	0.9861	0.9871
	F1	0.9918	0.9927	0.9861	0.9880
BERT-RF	Accuracy	0.9898	0.9722	0.9784	0.9767
	Precision	0.9900	0.9725	0.9789	0.9772
	Recall	0.9898	0.9722	0.9784	0.9767
	F1	0.9896	0.9719	0.9781	0.9764

将 CNN-Attention 及 Bi-LSTM-Attention 两种经典的深度学习模型应用于本数据集，与本文基于 BERT 构建的两种模型在心理特质分类预测的平均表现进行对比，得到结果如图 4 所示。对四种心理特质预测的平均准确率、平均精准率、平均召回率及平均 F_1 值，Bi-LSTM-Attention 模型表现在 0.71~0.79 之间，CNN-Attention 模型表现在 0.80~0.96 之间；而 BERT-B 和 BERT-RF 模型的对应指标均在 0.97~0.99 之间，明显优于其他两种深度学习模型的整体预测效果。由此可见，相较于其他两种经典的深度学习模型，本文基于 BERT 构建的心理特质分类模型的性能表现更胜一筹。

为进一步对基于 BERT 的 BERT-B 和 BERT-RF 心理特质分类模型性能做出对比和评价，图 5 按照从左到右、从上到下的顺序依次展示了 BERT-B 模型和 BERT-RF 模型在羞怯、合作性、完美主义、焦虑四种心理特质分类预测中所对应的 ROC 曲线及其 AUC 值。如图 5 所示，BERT-B 模型和 BERT-RF 模型的 ROC 曲线都位于坐标系的左上方，且 AUC 值基本都在 0.99 左右，说明两个模型的分类效果都不错。通过进一步比较后发现，相对于 BERT-RF 模型来说，BERT-B 模型在羞怯、合作性、完美主义、焦虑分类预测中的 ROC 曲线明显更靠近坐标系的左上角，而且其每种心理特质所对应的 AUC 值也都更大; 尤其是在羞怯特质中的表现, BERT-RF 模型的 AUC 值为 0.9822，BERT-B 模型的 AUC 值为 0.9985，尽管两个模型效果都相当不错，但 BERT-B 模型的表现仍然优于 BERT-RF 模型。

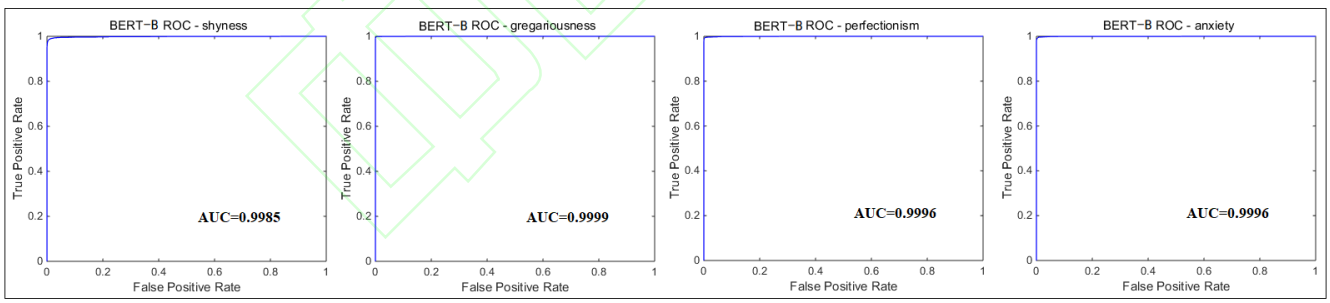
通过对上述结果的分析，初步推断 BERT 下游分类任务由分类器的不同所产生的分类精度差异，可能是由于 BERT 模型框架依赖于深度学习的神经网络结构，而随机森林作为一种传统的机器学习算法与神经网络的学习方式不同，在加入 BERT 下游分类任务层时可能会出现由于内部算法原理不同所导致的模型架构差异较大而影响分类精度。因此，本研究中在 $BERT_{BASE}$ 下游结构中添加全连接层所构建的 BERT-B 模型更加适于网络文本的心理特质预测。



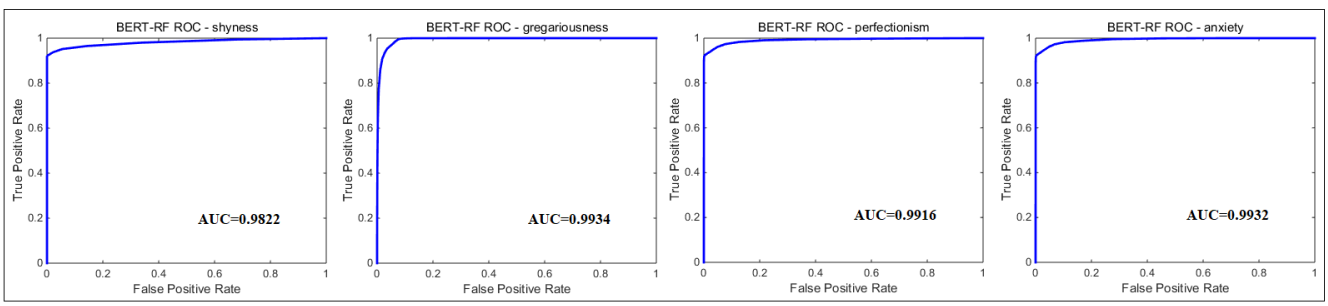
	Bi-LSTM-Attention	CNN-Attention	BERT-RF	BERT-B
Accuracy	0.7852	0.8820	0.9793	0.9894
Precision	0.7184	0.9497	0.9797	0.9897
Recall	0.7852	0.9614	0.9793	0.9894
F1	0.7439	0.8052	0.9790	0.9897

Fig.4 Comparison of average performance with other deep learning models.

图 4 与其他深度学习模型平均表现的对比



(a) BERT-B



(b) BERT-RF

Fig.5 The ROC and AUC of BERT-B and BERT-RF

图 5 BERT-B 模型与 BERT-RF 模型 ROC 曲线及 AUC 值

5 总结

本文基于 BERT 构建了 BERT-B 和 BERT-RF 两种网络文本心理特质预测模型，通过双向训练的模式使模型能够在更大程度上获得完整的上下文语义特征，利用 Transformer 的 Encoder 捕捉更长距离的上下文依赖关系；在下游分类任务中分别采用 $BERT_{BASE}$ 模型的单层神经网络结构和经典的随机森林算法作为两种不同的分类器，都实现出了较为理想的分类预测效果。与经典的深度学习模型 CNN-Attention 和 Bi-LSTM-Attention 的结果对比后进一步验证了本研究所提出模型的有效性。从平均表现来看，BERT-B 和 BERT-RF 两个模型的平均准确率、平均精准率等各项指标都在 0.97 以上，结合多项指标可以看出 BERT-B 模型的平均效果强于 BERT-RF 模型。就部分而言， BERT-RF 模型在羞怯特质中表现最好，准确率和精准率都大于 0.98，在合作性、完美主义及焦虑特质的分类预测中准确率及精准率也都在 0.97~0.99 之间；BERT 模型对于合作性和羞怯两种特质的分类预测准确率和准确率高达 0.99 以上，对完美主义和焦虑两种特质的分类预测准确率均高于 0.98，而且 BERT-B 模型的 ROC 与 AUC 表现也比 BERT-RF 更好。由此可以得出结论，相对于 BERT-RF 模型，BERT-B 模型的平均表现以及在每种心理特质的分类预测表现都更为出色。构建不同的心理特质预测模型，同时对其他相关心理特质或人格倾向的预测模型研究具有一定的参考价值。

从模型算法角度来看，对于四种心理特质的预测任务本质上为文本的多标签分类任务。考虑到心理特质本身是一种极为复杂的内隐特征，尽管目前来看本研究所预测的羞怯、合作性、完美主义、焦虑四种心理特质所对应的四种标签之间并不存在明显的相互依赖关系，但是仍然无法完全排除这一可能性，因此这种做法无法避免不同标签之间相互影响的风险。在后续研究中，可以考虑直接利用多标签输出的方式进行建模。

References:

[1] Gou L, Zhou M X, Yang H. KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014: 955-964.

[2] Tausczik Y R, Pennebaker J W. The psychological meaning of words: LIWC and computerized text analysis methods[J]. Journal of Language and Social Psychology, 2010, 29(1): 24-54.

[3] Barrick M R, Mount M K. The Big Five Personality Dimensions And Job Performance: A Meta-Analysis[J]. Personnel Psychology, 1991, 44(1): 1-26.

[4] Ford J K. Brands laid bare: Using market research for evidence-based[M]. Hoboken: John Wiley & Sons, 2005.

[5] Hinton G E, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.

[6] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] Xing J, Heeger D J. Measurement and modeling of center-surround suppression and enhancement[J]. Vision research, 2001, 41(5): 571-583.

[8] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th international conference on Machine learning, 2008: 160-167.

[9] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.

[10] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]// Twenty-ninth AAAI conference on artificial intelligence, 2015.

[11] Devlin J, Chang M-W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems, 2017: 5998-6008.

[13] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 1480-1489.

[14] Wang B, Zhang X, Zhou X, et al. A gated dilated convolution with attention model for clinical cloze-style

reading comprehension[J]. International Journal of Environmental Research and Public Health, 2020, 17(4): 1323.

[15] Adhikari A, Ram A, Tang R, et al. Docbert: Bert for document classification[J]. arXiv preprint arXiv:1904.08398, 2019.

[16] Kwantes P J, Derbentseva N, Lam Q, et al. Assessing the Big Five personality traits with latent semantic analysis[J]. Personality and Individual Differences, 2016, 102: 229-233.

[17] Wei H, Zhang F, Yuan N J, et al. Beyond the Words: Predicting User Personality from Heterogeneous Information[C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017: 305–314.

[18] Majumder N, Poria S, Gelbukh A, et al. Deep learning-based document modeling for personality detection from text[J]. IEEE Intelligent Systems, 2017, 32(2): 74-79.

[19] Sun C, Qiu X, Xu Y, et al. How to fine-tune BERT for text classification?[C]// China National Conference on Chinese Computational Linguistics, 2019: 194-206.

[20] Fernández-Delgado M, Cernadas E, Barro S, et al. Do we need hundreds of classifiers to solve real world classification problems?[J]. The Journal of Machine Learning Research, 2014, 15(1): 3133-3181.

ZHANG Han was born in 1995. She received the M.S. degree from Beijing Normal University in 2020. Her research interests include text mining, machine learning, and signal processing.

张晗(1995-)，女，山东泰安人，北京师范大学人工智能学院硕士研究生，主要研究领域为文本挖掘，机器学习，信号处理。

JIA Tianyuan was born in 1998. She is a M.S candidate at Beijing Normal University. Her research interests include text mining, machine learning, and signal processing.

贾甜远(1998-)，女，安徽蚌埠人，北京师范大学人工智能学院硕士研究生，主要研究领域为文本挖掘，机器学习，信号处理。

LUO Fang was born in 1979. She is an associate professor and Ph.D. supervisor at Beijing Normal University. Her research interests include psychological statistics and psychological assessment..

骆方(1979-)，女，河南驻马店人，北京师范大学心理学部副教授、博士生导师，主要研究领域为心理测量与统计。

ZHANG Sheng was born in 1979. He is an associate professor at Beijing Normal University. His main research interests include adaptive learning, big data in education, and evaluation criteria for future school.

张生(1979-)，男，四川绵阳人，北京师范大学中国基础教育质量监测协同创新中心副教授，主要研究领域为自适应学习、教育大数据、未来学校评估标准等。

WU Xia was born in 1978. She is a professor and Ph.D. supervisor at Beijing Normal University. Her research interests include machine learning algorithm and application, and intelligent information processing, etc.

邬霞(1978-)，女，湖南郴州人，北京师范大学人工智能学院教授、博士生导师、CCF 会员，主要研究领域为机器学习算法与应用、智能信息处理等。