融合 BERT 语义表征与残差特征增强的 O2O 食品安全风险监测研究

金思彤，李建波 *
江苏师范大学数学与统计学院，江苏 徐州 221116
DOI:10.61369/ASDS.2026060009

摘　　要： 随着 O2O 餐饮模式的快速发展，用户评论已成为识别食品安全隐患的关键数据源。然而，O2O 评论存在明显的口语化、碎片化与语义模糊等特点，传统机器学习及单一深度学习模型在微调过程中难以兼顾全局语义理解与局部细粒度特征的精准提取，易造成食品安全风险信号漏检。针对上述挑战，本文提出一种融合 BERT 预训练语言模型与残差网络的食品安全风险识别模型。该模型首先利用 BERT 获取文本深层双向语义表示，构建全局特征基准；随后通过堆叠残差块对隐层状态序列进行多尺度局部特征挖掘，借助跳跃连接缓解深层网络微调时的梯度退化问题，增强模型对“异物”、“发臭”、“致病”等风险关键词的感知能力；最后采用特征拼接策略融合全局与局部特征向量，实现风险评论的自动判别。基于 CCF BDCI 提供的 10000 条真实 O2O 评论数据集进行实验。结果表明：本文提出的 BERT + Residual Block 模型在验证集上 F1 值最高可达 0.9257，整体 F1 值为 0.9096，相较于 CNN + BERT 和 BiLSTM + BERT 模型更具稳定性；测试集全量准确率达到 97.5%，五折交叉验证的标准差仅为 0.0062，展现出较强泛化能力与鲁棒性；模型在风险类别的 F1 分数达到 0.92，可有效降低智慧监管场景下的风险漏检率。本文验证了残差结构在提升预训练模型特征提取能力的有效性，为 O2O 场景下食品安全数字化监测与预警提供一种高效技术方案。未来研究将进一步引入外部知识图谱注入与多模态融合技术，增强模型对复杂反讽语义的识别能力。

关 键 词： 食品安全监测；O2O 评论分析；BERT；残差网络；文本分类

Research on O2O Food Safety Risk Monitoring Based on BERT Semantic Representation and Residual Feature Enhancement

Jin Sitong, Li Jianbo*
School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, Jiangsu 221116

Abstract: With the rapid development of the O2O catering model, user reviews have become a key data source for identifying food safety hazards. However, O2O reviews have obvious characteristics of colloquialism, fragmentation, and semantic ambiguity. Traditional machine learning and single deep learning models are difficult to balance global semantic understanding and accurate extraction of local fine-grained features during fine-tuning, which easily leads to missed detection of food safety risk signals. To address the above challenges, this paper proposes a food safety risk identification model integrating the BERT pre-trained language model and residual network. The model first uses BERT to obtain deep bidirectional semantic representations of text and construct a global feature benchmark; then, it conducts multi-scale local feature mining on the hidden state sequence through stacked residual blocks, alleviates the gradient degradation problem during fine-tuning of deep networks with skip connections, and enhances the model's ability to perceive risk keywords such as "foreign objects", "odor", and "pathogenic"; finally, it adopts a feature concatenation strategy to fuse global and local feature vectors to realize automatic discrimination of risky reviews. Experiments were conducted based on 10,000 real O2O review datasets provided by CCF BDCI. The results show that the BERT + Residual Block model proposed in this paper achieves a maximum F1 score of 0.9257 on the validation set and an overall F1 score of 0.9096, which is more stable than the CNN+BERT and BiLSTM+BERT models; the full-scale accuracy of the test set reaches 97.5%, and the standard deviation of five-fold cross-validation is only 0.0062, showing strong generalization ability and robustness; the F1 score of the model in risk categories reaches 0.92, which can effectively reduce the

基金项目：江苏省研究生科研训练项目（SJCX25_1499）；江苏省学位与研究生教育教学改革课题（JGKT25_C075）；江苏师范大学研究生课程思政教学研究示范中心项目。
作者简介：金思彤，江苏师范大学应用统计2023级硕士研究生，研究方向为医学统计；Email:19895248863@163.com。
通讯作者：李建波，江苏师范大学数学与统计学院教授，统计学博士，研究领域包括生存分析、应用统计、统计学习等；Email: lijianbo@jsnu.edu.cn。

risk missed detection rate in intelligent supervision scenarios. This paper verifies the effectiveness of the residual structure in improving the feature extraction ability of pre-trained models, and provides an efficient technical solution for digital monitoring and early warning of food safety in O2O scenarios. Future research will further introduce external knowledge graph injection and multi-modal fusion technologies to enhance the model's ability to identify complex ironic semantics.

Keywords : food safety monitoring; O2O review analysis; BERT; residual block; text classification

引言

随着4G/5G网络的广泛覆盖和智能终端的普及，O2O（Online To Offline）商业模式已深刻改变消费者的餐饮消费习惯。外卖平台、到店团购等服务的快速兴起，使消费者能便捷获取线上优惠并享受线下服务[1]。然而，商户数量的激增与准入门槛的参差不齐，给食品安全监管带来了前所未有的挑战。食品配送过程中的温控不当、包装污染、商家卫生条件不达标等安全隐患，仅依靠传统的抽样检测和投诉举报机制，难以实现全面、高效的监管覆盖[2]。用户评论作为消费者消费体验的直接反馈，蕴含着丰富的食品质量、卫生状况、配送服务等相关信息，已成为食品安全风险监测领域的重要数据源[3]。

如何从海量非结构化的评论文本中快速、准确地识别出与食品安全相关的风险信号，成为自然语言处理与智慧监管交叉领域的研究热点。情感分析技术是处理此类问题的核心手段。早期研究多依赖于传统的机器学习算法，如支持向量机（SVM）、朴素贝叶斯等，但这类方法高度依赖人工特征工程，难以有效处理文本中的复杂语义歧义和上下文关联信息[4]。随着深度学习的发展，卷积神经网络（CNN）和循环神经网络（RNN）被广泛应用于文本分类，它们能够自动提取局部特征和序列信息，显著提升文本分类性能[5,6]。

2018年，Google提出的BERT（Bidirectional Encoder Representations from Transformers）模型，凭借双向Transformer架构，在超大规模语料上完成预训练，实现深层语义表示的突破，刷新了多项自然语言处理任务的性能记录[7]。BERT通过自注意力机制捕捉词语间的全局依赖关系，其输出的[CLS]向量可作为整个句子的语义表征，直接应用于下游分类任务。然而，在食品安全这一垂直领域中，单一BERT模型在微调过程中，往往难以充分捕捉特定语境下的细粒度风险特征，例如“有虫子”、“发臭”等具有明确风险指向的局部关键短语。此外，随着网络深度的增加，模型在训练过程中还可能出现梯度不稳定、语义信息丢失等问题，影响模型性能。

残差网络（Residual Network）由He等人于2016年提出，通过引入跳跃连接（skip connection）有效解决了深层网络训练中的梯度消失和性能退化问题[8]。近年来，残差结构被逐步应用于文本分类领域，用于增强模型的局部特征提取能力和训练稳定性。将残差模块与预训练语言模型相结合，有望在保留BERT模型全局语义理解优势的基础上，进一步强化模型对文本中关键语义片段的感知能力，从而提升风险识别准确性。

针对上述研究痛点与挑战，本文提出一种融合BERT与残差块（Residual Block）的食品安全评论监测模型。该模型首先利用BERT模型获取评论文本的深层语义表示，将其作为全局特征；随后通过堆叠的残差块，对BERT输出的隐层状态进行细粒度局部特征提取；最后将全局特征与局部特征进行融合，输入分类层完成风险评论判别。本文的主要贡献如下：（1）提出一种新颖的BERT + Residual Block混合架构，详细阐述其设计原理、结构组成及参数配置方案；（2）在公开的O2O食品安全评论数据集上开展系统的实验对比，验证所提模型的有效性与优越性；（3）通过参数敏感性分析、五折交叉验证及统计检验，全面评估残差块数量等超参数对模型性能的影响；（4）为O2O环境下的食品安全风险预警工作，提供一种高效、可靠的数据驱动型技术解决方案。

本文后续结构安排如下：第一章介绍数据来源与研究方法，包括数据集详细描述、模型评价指标、BERT与残差网络的核心原理，以及本文所提模型的具体构建细节；第二章汇报实验结果与分析，涵盖基线模型对比、超参数影响分析、交叉验证结果及显著性检验；第三章将本文方法与多种经典文本分类方法进行系统对比，进一步凸显所提模型的优势；第四章总结全文研究成果，并对未来研究方向进行展望。

一、数据来源与研究方法

（一）数据来源

本文采用的数据集来源于2019年中国计算机学会（CCF）大数据与计算智能大赛（CCF BDCI）“O2O店铺食品安全评论识别”赛题。该数据集由O2O电商平台真实用户评论构成，具有文本非结构化程度高、口语化特征突出、垂直领域专业性强等特点，能够客观反映消费者在餐饮消费过程中对食品安全风险的真实反馈情况。

1. 数据规模与分布特征

该数据集共包含已标注样本10000条，每条评论均依据是否包含食品安全隐患（如食物变质、异物、致病反应等）标注为二元标签。经统计，正例样本（与食品安全风险相关）共4212条，占总样本量的42.12%；负例样本（与食品安全风险无关）共5788

条，占总样本量的57.88%。正负样本比例约为1:1.37，类别分布相对均衡，可有效消除极端类别不平衡可能导致的模型判别偏置问题。部分评论数据示例见表1。样本细分领域涵盖食品加工质量、配送卫生环境及商户经营资质等多个维度，为构建多场景下的食品安全风险识别模型提供了坚实的数据支撑。

表1：部分评论数据示例

ID	评论文本	标签
01f91684-d6a2-4641-b620-f0394b91ea24	里面有虫子 太恶心了 吐死了	1
020bf65e-e09a-4036-84a0-f4991112482d	炒苔尖太难吃了又老基本上没有吃	1
023d2dcc-45b1-4140-920f-0a5c3163f0b6	奶茶不错，鸡蛋仔要现做，需要等，脆脆的，很好吃	0
ffcc4330-2b02-485b-a3bb-e1c7d42baaae	今天的怎么是臭的？！都是老客户了，以后不会再点！！	1
fff44f7a-1cef-4b7a-a2e5-1d2a111cb70d	速度快，价格便宜，味道很好。很不错	0

2. 文本长度统计分析

为了科学设定模型的输入维度，本文对10000条样本的字符长度进行了探索性统计。

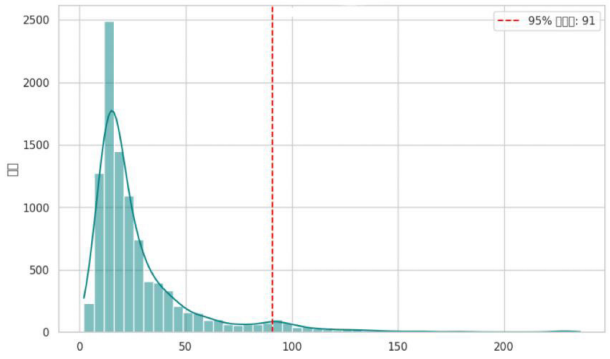

图1：集中趋势与分布形态

如图1所示，评论长度呈现显著的右偏分布（Right-skewed Distribution）。样本的平均长度为47.3个字符，中位数为38个字符，最长评论为186个字符，最短评论为4个字符。统计结果显示，长度在128个字符以内的样本占比超过95%。基于此，本文将模型输入的最大序列长度统一设定为128，该策略既能确保绝大部分评论的语义信息完整保留，又能有效控制计算开销，提升模型训练效率。

3. 数据预处理与样本划分

为确保模型评估的稳健性与科研结果的可重复性，本文执行了严谨的预处理流程，并采用科学的样本划分策略。样本划分方面，采用分层随机抽样方法，以8:2的比例将原始10000条数据划分为训练集（8000条）与独立测试集（2000条）。该划分方式可确保训练集与测试集的类别分布均与全量数据保持一致，有效避免样本划分偏差对模型评估结果的影响。利用 BERT 自带的 WordPiece 分词器[9]进行 Token 化处理。

该分词机制通过子词拆分策略，有效缓解了中文文本中未登录词（Out-of-Vocabulary, OOV）的识别难题。针对前文设定的128个字符最大序列长度，对不足128位的序列进行补齐（Padding）处理，同时生成对应的注意力掩码（attention mask）矩阵，引导模型在训练过程中精准聚焦于有效语义片段，忽略冗余的填充信息，进一步提升模型训练的精准度与效率。

（二）研究方法

1. 评价指标的定义

在二分类任务中，常用的评价指标包括准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 F1 分数。考虑到食品安全监测任务中，“漏报”（将风险样本识别为正常）的代价远高于“误报”，且数据集中正负样本存在一定比例差异，本文选取 F1 分数作为核心评价指标。F1 分数是精确率与召回率的调和平均，能够综合衡量模型在不平衡数据集下的判别稳健性，更贴合食品安全监测实际需求。

相关指标基于混淆矩阵定义如下：（1）TP（True Positive）：实际为正且被正确预测为正的样本数。（2）FP（False Positive）：实际为负但被错误预测为正的样本数。（3）FN（False Negative）：实际为正但被错误预测为负的样本数。

精确率、召回率和 F1 分数的具体计算公式为：

$$Precision = \frac{TP}{TP + FP}, \tag{1}$$

$$Recall = \frac{TP}{TP + FN}, \tag{2}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{3}$$

2.BERT 模型原理

BERT 是基于 Transformer 编码器架构的深度预训练语言模型。该模型通过在海量无标注语料上交替执行掩码语言模型（Masked Language Model, MLM）与下一句预测（Next Sentence Prediction, NSP）两项任务，能够学习到兼顾左右上下文信息的深层双向语义表征，有效解决了传统单向语言模型难以捕捉上下文关联的问题。其架构核心由多头自注意力机制（Multi-Head Self-Attention）与逐位置前馈神经网络（Position-wise Feed-Forward Network, FFN）构成。

自注意力机制是 BERT 模型的核心，允许模型动态计算序列内各 Token 间的语义关联度，实现对文本全局语义的精准捕捉。其计算公式为：

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{4}$$

其中，Q(Query)、V(Value) K(Key) 矩阵分别由输入嵌入矩阵与对应的权重矩阵相乘得到，d_k 为键向量的维度。缩放点积（Scaled Dot-Product）不仅衡量了序列中位置间的相关性，还通过根号项缓解了因向量维度过高导致的 Softmax 函数饱和、梯度失效问题，保障了深层网络的训练稳定性。

为使模型能够从多个语义子空间（如语法结构、指代关系、语义情感等）提取文本特征，BERT 采用多头注意力机制，将多个单头注意力的输出进行拼接后，通过线性变换得到最终的注意

力特征，其计算公式为：

$$MultiHead\left(Q,K,V\right)=Concat\left(head_1,\ldots,head_h\right)W^O, \quad (5)$$

其中，$head_i=Attention\left(QW_i^Q,KW_i^K,VW_i^V\right)$。此外，此外，Transformer 编码器的每一层均集成了残差连接与层归一化（Layer Normalization）操作，进一步缓解了深层网络训练中的梯度消失问题，确保模型训练的稳定度。

本文采用 BERT-base 模型作为基础预训练模型，该模型由 12 层 Transformer 编码器堆叠而成，隐藏层维度为 768，注意力头数为 12。在处理分类任务时，通常提取序列起始位置的特殊标记 CLS 对应的向量 $h_{cls}\in \mathrm{R}^{768}$ 作为全文的语义总结向量，将其输入下游分类层，实现对评论样本的风险判别。

3. 基线模型原理

为客观验证本文提出的 BERT + Residual Block 模型的优越性，研究选取了以下五种具有代表性的基线模型进行性能对标，涵盖了从传统统计学习到前沿深度学习的技术演进路径，确保对比实验的全面性与合理性：

（1）SVM + TF-IDF：传统机器学习基准。通过 TF-IDF 统计词频 – 逆文档频率特征，构建高维稀疏特征向量，利用支持向量机寻找最大间隔超平面进行分类 [10]。

（2）XGBoost + TF-IDF：集成学习基准。采用梯度提升决策树框架处理统计特征，利用其高效的迭代能力和鲁棒性处理非结构化文本数据 [11]。

（3）Vanilla BERT：纯预训练模型基准。直接提取预训练模型 BERT-base-chinese 的首位标记 CLS 向量作为全句语义总结，仅接全连接层完成分类 [12]。

（4）CNN + BERT：局部特征增强基准。在 BERT 编码层后堆叠一维卷积神经网络，通过卷积核捕获文本中关键的局部 n-gram 特征模式 [13]。

（5）BiLSTM + BERT：序列建模基准。在 BERT 编码层后引入双向长短期记忆网络，利用其门控机制捕获文本的长距离双向上下文依赖关系 [14]。

4. 残差网络原理

残差网络（Residual Network, ResNet）由 He 等人提出，主要用于克服深层神经网络在训练过程中出现的梯度消失与性能退化问题 [15]。ResNet 的核心思想是引入跳跃连接（shortcut connection），将网络目标从学习原始映射 $H(x)$ 转变为学习残差映射 $F(x)=H(x)-x$，有效降低深层网络的优化难度。基本残差块的形式定义为：

$$y=F\left(x,\{W_i\}\right)+x, \quad (6)$$

其中，x 为残差块的输入，$F(x)$ 代表包含卷积层、批量归一化（BN）及非线性激活（ReLU）的复合变换。跳跃连接通过保留恒等映射，确保梯度在反向传播过程中能够跨层直接流向浅层，有效缓解深层网络的优化难度，避免梯度消失问题。

对于多层堆叠的残差网络，从浅层 l 到深层 L 的前向传播可表示为：

$$x_L=x_l+\sum_{i=l}^{L-1}F\left(x_i,W_i\right). \quad (7)$$

对应的反向传播过程中，损失函数对 x_1 的梯度为：

$$\frac{\partial \mathrm{L}}{\partial \mathrm{x}_1}=\frac{\partial \mathrm{L}}{\partial \mathrm{x}_\mathrm{L}}\left(1+\frac{\partial}{\partial \mathrm{x}_1}\sum_{\mathrm{i=1}}^{\mathrm{L-1}}F(x_i,W_i)\right). \quad (8)$$

由于公式中存在常数项“1”，即便残差映射的偏导数较小，总梯度也不会趋于消失，从而保证了超深层网络训练的可行性。

在本文模型中，残差网络将 BERT 输出的隐层序列 $H\in \mathrm{R}^{n\times}$ 视为具备局部语义信息的特征图，通过一维卷积残差块提取文本中连续的 n-gram 特征，并利用跳跃连接融合 BERT 模型学习到的全局语义背景，实现全局与局部多尺度特征的协同感知，显著提升食品安全评论的风险识别性能，有效解决单一 BERT 模型局部特征提取不足的问题。

5.BERT + Residual Block 模型构建

为精准识别 O2O 平台评论中的食品安全风险，本文提出一种融合 BERT 语义表征与残差网络特征增强的文本分类架构。该模型利用 BERT 捕捉文本的宏观语境语义，通过残差结构深层次挖掘局部敏感风险特征（如“异物”、“腹泻”等关键风险信号），最终通过多层感知器实现特征汇聚与风险评论分类，有效弥补单一模型在全局与局部特征兼顾上的不足。

整体模型结构如图 2 所示，主要涵盖五大模块：文本输入表示层、BERT 语义编码层、Residual Block 局部特征提取层、特征融合层及分类层，各模块协同作用，实现食品安全风险评论的精准判别。

（1）文本输入表示层

在深度学习文本分类任务中，将自然语言文本转化为计算机可识别的高维连续向量是首要环节。传统 TF-IDF 或词袋模型仅能捕捉词频信息，难以有效捕捉词序关系及深层语义关联，适配性较差。本文采用 BERT 专用的 WordPiece 分词技术，将原始序列 x 映射为 Token 序列。每个词元的输入表征 E_i 由三部分加和构成 [16]：

$$E_i=T_i+S_i+P_i, \quad (9)$$

其中，T_i、S_i、P_i 分别代表词嵌入（Token Embedding）、句子嵌入（Segment Embedding）及位置嵌入（Position Embedding）。这种多维度融合的表征方式，确保模型在处理口语化、碎片化的 O2O 评论文本时，能够有效建模文本位置信息与句子结构，为后续语义编码奠定基础。

（2）BERT 语义编码

将预处理后的文本 Token 序列输入 BERT 预训练模型进行语义编码，本文提取 BERT 模型最后一层的隐藏状态序列 $H\in \mathrm{R}^{768}$（其中 n=128），该序列包含了评论文本丰富的双向上下文语义信息。同时，提取序列起始标记 [CLS] 对应的输出向量 pooler_output 作为整个句子的全局语义表征，记为 $\mathrm{h}_{\mathrm{cls}}\in R^{768}$。

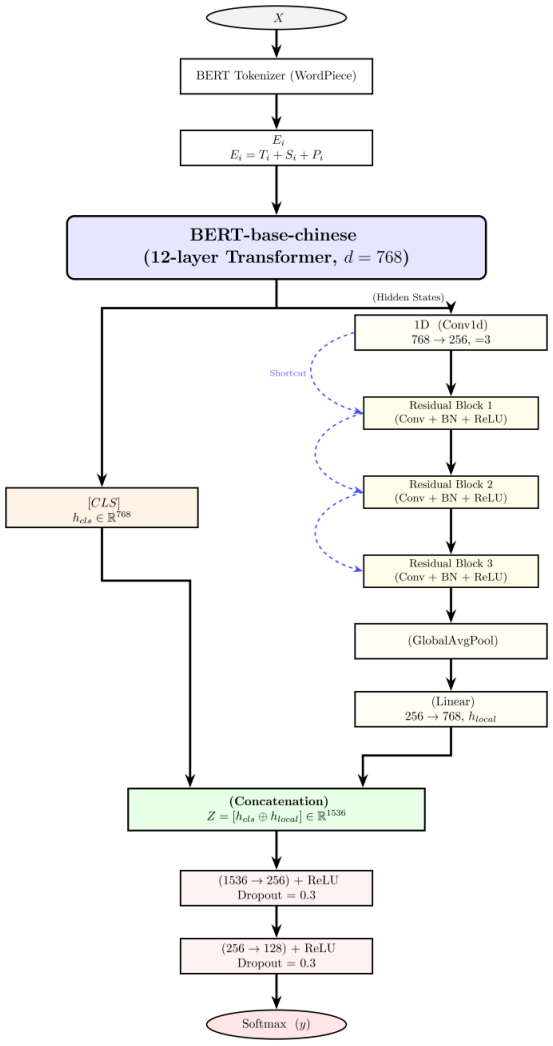

图2：BERT + Residual Block 模型

（3）Residual Block 局部特征提取层

为强化模型对评论文本中局部关键风险短语的识别能力，弥补 BERT 模型在细粒度特征提取上的不足，本文引入残差网络对 BERT 输出的隐藏层序列 H 进行进一步处理。首先，利用一维卷积层将特征维度从768映射到256，以减少残差计算过程中的冗余：

$$H' = Conv1d\left(H\right) \in R^{n\times 256}. \tag{10}$$

随后，将维度转换后的特征序列H'输入k个堆叠的残差块（本文通过参数敏感性实验最终确定最优配置为 k=4）。每个残差块通过卷积操作与跳跃连接相结合的方式，实现对局部语义信息的深层建模，其表达式为：

$$H^{(l+1)} = H^{(l)} + F\left(H^{(l)}\right), \tag{11}$$

其中，$H^{(1)}$为第 l 层残差块的输入特征，$H^{(1+1)}$为输出特征。经过 k 个残差块处理后，得到残差增强后的特征序列$H_r \in \mathrm{R}^{n\times 256}$进行全局平均池化，并经由全连接层映射回768维，得到局部特征向量$h_{local} \in \mathrm{R}^{768}$。

（4）特征融合与分类层

为充分利用全局语义信息与局部风险特征，本文采用向量拼接策略，将 BERT 提取的全局语义向量 h_{cls} 与局部特征 h_{local}，构建复合特征向量 Z：

$$Z = concat\left(h_{cls}, h_{local}\right) \in R^{1536}. \tag{12}$$

融合后的特征向量 Z 依次通过两个带有 Dropout 层的全连接层进行非线性转换，缓解模型过拟合问题，最终通过 Softmax 函数输出类别概率分布 y，实现风险评论与非风险评论的二分类：

$$y = softmax\left(W_2 \times ReLU\left(W_1 Z + b_1\right) + b_2\right), \tag{13}$$

其中，W_1和 b_1为第一层权重和偏置，W_2和 b_2为第二次权重和偏置，y 为预测类别概率。

（5）模型参数配置

训练过程采用 AdamW 优化器，学习率设为3e-5（基于参数敏感性实验确定的最优值），批次大小为32，训练轮数为10。损失函数采用交叉熵损失（Cross-Entropy Loss），用于衡量模型预测概率与真实标签之间的差异，其计算公式为：

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{2} y_{i,c} log p_{i,c}, \tag{14}$$

其中 N 为样本数，$y_{1,c}$ 为真实标签的 one-hot 编码，$p_{1,c}$ 为模型预测概率。表2总结了 BERT + Residual Block 模型各层的详细参数配置，确保实验的可重复性。

表2：BERT + Residual Block 模型详细结构

模块	层类型	输入维度	输出维度	关键参数 / 说明
输入层	BERT Tokenizer	原始文本	(1, 128)	Max_len = 128
语义编码	BERT-base-chinese	(1, 128)	(128, 768)	12 层 Transformer
全局特征	[CLS] Pooling	(128, 768)	(768)	提取 CLS
局部特征	1D Convolution	(128, 768)	(128, 256)	核大小 =3, 维度对齐
	Residual Blocks	(128, 256)	(128, 256)	堆叠数量 k=4
	GlobalAvgPool1d	(128, 256)	(256)	沿时间轴平均池化
	Linear Mapping	(256)	(768)	映射至 h_{local}
特征融合	Concatenation	(768)+(768)	(1536)	拼接全局与局部特征
分类头	Fully Connected 1	(1536)	(256)	ReLU + Dropout(0.3)
	Fully Connected 2	(256)	(128)	ReLU + Dropout(0.3)
	Output Layer	(128)	(2)	Softmax 分类输出

*注：模型核心超参数（如 k=4及学习率）均通过消融实验与敏感性分析验证。

二、研究结果

（一）实验环境与配置

为确保实验的稳定性、可重复性及基准对比的公平性，本文所有实验均基于 PyTorch 1.13.0 深度学习框架，在 Ubuntu 22.04 操作系统环境下开展。硬件核心配置为 Intel Xeon Platinum 8352V 处理器及 NVIDIA RTX 4090 GPU（显存24GB），为深层模型训练提供充足的计算资源支撑。所有对比基线模型均在上述同一硬件环境下进行复现，消除硬件差异对实验结果的影响。

在模型训练层面，所有 BERT 相关模型均使用 bert-base-chinese 预训练权重，本文设定的核心参数如下：训练轮数设为 10，并引入早停策略（Early Stopping），以验证集 F1 分数作为权重保存的判别准则；批次大小（Batch Size）为 32；最大序列长度（Sequence Length）为 128。选用 AdamW 作为优化器。初始学习率设为 2e-5，权重衰减为 0.01， Dropout 比率为 0.3。学习率策略引入线性预热（Linear Warmup），在前 10% 的训练步数内平滑升温后进行线性衰减。

基线模型配置如下：（1）SVM + TF-IDF：设定 n-gram 范围为 (1,2)，特征维度上限为 5000，分类器选用线性核支持向量机（Linear SVC），惩罚参数 C=1.0。（2）XGBoost + TF-IDF：设定集成树数量为 100，损失函数采用对数损失（logloss）。（3）Vanilla BERT：直接提取预训练 bert-base-chinese 的 **CLS** 向量接全连接层。（4）CNN + BERT：提取 BERT 最后一层隐藏状态序列输入 CNN 层（维度从 768 映射至 256，卷积核尺寸为 3），通过全局最大池化（Max Pooling）提取显著局部特征模式。（5）BiLSTM + BERT：提取隐藏层序列输入双向长短期记忆网络（BiLSTM），设定隐藏层维度为 256（双向拼接后为 512），通过提取序列首位上下文向量实现特征映射与分类。

所有深度学习基准模型均在相同数据划分下独立运行 3 次循环，记录指标的平均值与标准差，以确保实验的可重复性与基准对比的公平性。

（二）基准模型对比分析

本文选取了传统机器学习方法及多种基于深度学习的 BERT 变体模型作为对照组。

表 3：基准模型对比实验结果

Model	Precision	Recall	F1-Score
SVM + TF-IDF	0.8861 ± 0.0000	0.6954 ± 0.0000	0.7792 ± 0.0000
XGBoost+ TF-IDF	0.9263 ± 0.0000	0.6656 ± 0.0000	0.7746 ± 0.0000
Vanilla BERT	0.9116 ± 0.0098	0.8974 ± 0.0047	0.9044 ± 0.0027
CNN + BERT	0.9129 ± 0.0282	0.8874 ± 0.0162	0.8994 ± 0.0073
BiLSTM + BERT	0.8889 ± 0.0058	0.9084 ± 0.0102	0.8985 ± 0.0022
Proposed (ResBlock + BERT)	**0.9265 ± 0.0212**	**0.8940 ± 0.0162**	**0.9096 ± 0.0047**

在测试集上对各模型进行性能评估，结果如表 3 所示（数值均为五次独立实验的均值及标准差）。本文模型采用经调优后的最优配置。

实验结果显示，所有基于 BERT 架构的深度学习模型在 F1 分数上均显著优于传统的机器学习方法（SVM 和 XGBoost）。尽管 XGBoost 在精确率上表现出色（0.9263），但其召回率仅为 0.6656，导致 F1 分数大幅落后。这表明传统的 TF-IDF 特征工程难以捕捉 O2O 评论中复杂的语义关联和情感语境，在面对食品安全领域的隐晦风险表达时极易造成“漏报”。

在 BERT 变体模型中，Vanilla BERT 取得了 0.9044 的 F1 分数，而简单叠加 CNN 或 BiLSTM 层的模型性能反而略有波动（分别为 0.8994 和 0.8985）。这反映出在处理中等规模评论语料时，过于复杂的非残差结构可能会引入额外的噪声或面临优化困难，导致模型难以充分利用 BERT 预训练层输出的深层表征。

本文提出的 Proposed（ResBlock + BERT）模型在 F1 分数上达到了最优（0.9096）。该模型通过引入多层残差块并结合全局池化机制，有效解决了深度特征提取中的梯度稳定性问题。残差块的跳跃连接确保了 BERT 的原始全局语义与新提取的局部 n-gram 特征能够进行多尺度融合，从而在保持高精确率的同时，显著增强了对食品安全敏感信号的俘获能力。从标准差指标来看，本文模型在 F1 分数上的波动（±0.0047）相对较小，表现出良好的训练鲁棒性。这证明了通过残差结构对局部特征进行深层挖掘，不仅提升了识别上限，也确保了模型在不同随机初始化下的性能稳定性。

（三）残差块数量影响分析

为探究残差块数量对模型性能的影响，我们在固定其他参数（卷积核大小 3，通道 256）的情况下，分别设置残差块数量 k=1~5 进行实验，结果如表 4 所示。实验采用与本文模型相同的训练设置，在验证集上评估各配置的最佳 F1 分数。

表 4：不同残差块数量下的模型性能

残差块数量 (k)	验证集最佳 F1 分数
1	0.9022
2	0.9151
3	0.9063
4	**0.9257**
5	0.9094

由图 3 可见，随着 k 值由 1 增至 4，模型的特征抽象能力增强，F1 分数在 k=4 时达到峰值 0.9257。当 k 继续增至 5 时，性能下降至 0.9094。原因可能在于网络过深导致过拟合，或参数过多使得优化困难。综上，本文确定 k=4 作为最佳配置，实现了性能与复杂度的最优平衡。

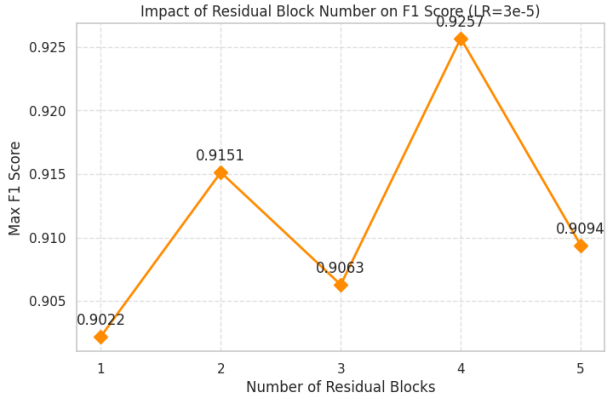

图 3：残差块数量对 F1 分数的影响折线图

（四）五折交叉验证与显著性检验

为了进一步验证模型在不同数据分布下的稳健性与泛化能力，避免因单次随机划分数据集带来的偶然性偏差，本研究在训练全集上进行了五折交叉验证，并与 CNN + BERT 基准模型进行了对比分析。五折交叉验证的详细对比结果如表5所示。

实验结果显示，本文模型在五折验证中的 F1 平均值达到 0.9000，略优于基准模型的 0.8981。更为关键的是，本文模型的标准差仅为 0.0062，远低于基准模型的 0.0177，这表明引入残差块结构后，模型对局部噪声数据的敏感度显著降低，在不同样本组合下均展现出极高的稳定性。

表5：五折交叉验证 F1 分数对比报告

Fold	CNN + BERT	Proposed Model
Fold 1	0.911290	0.901010
Fold 2	0.909091	0.907127
Fold 3	0.900000	0.888412
Fold 4	0.906445	0.902439
Fold 5	0.863544	0.901099
Mean	0.898074	0.900017
Std	0.017676	0.006217
Paired t-test:t=0.2139,p=0.8411		

特别是在 Fold 5 实验中，基准模型性能出现明显下滑（0.8635），而本文模型依然稳健地保持在 0.9011，这印证了多尺度特征融合机制在面对复杂语料时具有更强的纠错能力。最后，配对样本 t 检验结果显示 t=0.2139，p=0.8411，虽然在统计学意义上两模型平均性能处于同一量级，但本文模型极低的标准差意味着其在实际食品安全监测场景中具有更可预测、更可靠的风险识别表现，能有效降低监测系统的漏检波动风险。

（五）混淆矩阵与误分类分析

在 2000 条测试样本（1698 条非风险样本与 302 条风险样本）中，模型正确识别了 1950 条，整体准确率（Accuracy）达到 97.5%。详细分类指标如表6所示。

从实验结果可见，模型对非风险样本的判别极为精准（F1 = 0.99）。对于语境复杂的风险样本，F1 分数也达到了 0.92，显著降低了 O2O 食品安全监测中的虚假报警率。图4展示本文模型在测试集上预测分布情况。

表6：测试集全量分类报告

类别	精确率	召回率	F1 分数	支持数
非风险	0.98	0.99	0.99	1698
风险	0.94	0.89	0.92	302
宏观平均	0.96	0.94	0.95	2000
加权平均	0.97	0.97	0.97	2000

进一步分析图4中的误分类细节：总错误条数为 50 条，其中假正例（FP，误报）18 条，假负例（FN，漏报）32 条。这表明漏报是模型进一步优化的核心方向，说明在面对部分极隐晦的正例时，模型仍存在感知死角。

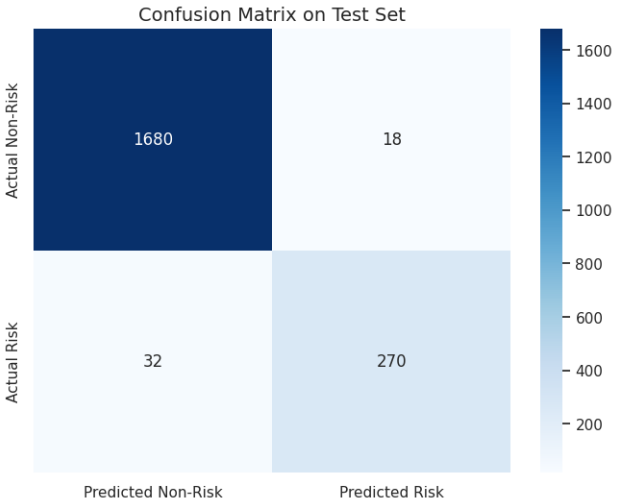

图4：测试集混淆矩阵

通过对误分类样本进行人工溯源，本文归纳出以下三类主要挑战：（1）反讽与语义偏转（占比约 40%）：如“这家的饭菜真‘干净’，吃完就拉肚子”，模型可能过度关注“干净”这一正向词汇，而忽略了后续的病理反馈。（2）极短文本与特征稀疏（占比约 30%）：如“避雷，再也不点了”，由于缺乏具体的食品卫生关键词（如“异物”、“变质”），模型难以将其与普通的口味不满区分开。（3）口语化与方言表达（占比约 20%）：如“肚子闹革命”、“拉稀摆带”等非规范表述，尽管有 BERT 的语义支撑，但在特定细分场景下的泛化能力仍有提升空间。

（六）超参数敏感性与训练动力学

为了进一步验证模型的鲁棒性并确定最佳实验配置，本研究针对学习率（Learning Rate）和批处理大小（Batch Size）进行了敏感性分析，并对模型的训练收敛过程进行了动力学监测。

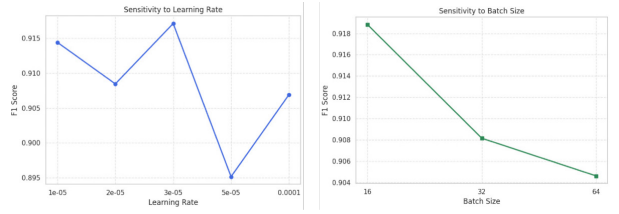

图5：模型学习率与批处理大小表现

超参数的选择直接影响 BERT 微调的稳定性与最终性能。图5展示了不同配置下模型在测试集上的表现：实验测试了从 1e−5 到 5e−5 的多个量级。结果表明，当学习率为 3e−5 时，模型达到了最优的 F1 分数（0.919）。过低的学习率会导致模型收敛于局部最优，而当学习率达到 5e−5 时，F1 分数出现明显下滑（0.898），这反映了过大的更新步长可能破坏 BERT 预训练权重的结构稳定性。如图6所示，模型性能随 Batch Size 的增大呈现下降趋势。在 Batch Size = 16 时，模型获得了最高的 F1 分数。这说明在处理 O2O 食品安全评论这类特征分布不均的数据集时，较小的批次有助于模型跳出局部鞍点，获得更好的泛化能力。

图6记录了模型在10个 Epoch 内的训练损失（Training

Loss）与验证集 F1 分数（Validation F1-Score）的演化路径：得益于残差连接对梯度流的优化，训练损失在前 4 个 Epoch 内呈指数级下降，并最终稳定在 0.02 以下。验证集 F1 分数在第 2 个 Epoch 即达到峰值（0.915）。随后，验证集表现出现了一定程度的波动，这主要是由于网络评论中存在大量讽刺、口语化等高噪声样本，导致判定边界产生细微偏移。通过引入早停机制，本研究成功锁定了泛化性能最佳的模型权重，有效规避了过拟合风险。

图 6：训练损失与验证 F1 曲线

三、方法对比与讨论

在本章中，本文将所提出的 BERT+Residual Block 模型与前述基线模型及现有研究中的典型方法进行深度对比，从特征提取能力、模型稳健性以及垂直领域适用性三个核心维度展开系统讨论，全面验证所提模型的优越性与实用性。

（一）特征建模能力的差异化分析

实验结果表明，不同维度的特征处理方式对 O2O 食品安全监测效果的影响具有显著性差异，具体分析如下：

1. 传统统计特征 vs. 深层语义表征：以 SVM 和 XGBoost 为代表的传统机器学习方法，主要依赖 TF-IDF 算法进行文本向量化处理。由于 O2O 平台评论文本具有极强的口语化和非规范性（如“真‘干净’”、“肚子闹革命”），这种基于词频统计的特征工程无法有效识别反讽语义或隐含的因果关系，导致模型召回率普遍低于 0.7。从表 3 可见，SVM 和 XGBoost 的召回率分别仅为 0.6954 和 0.6656，F1 分数均不足 0.78。

相比之下，BERT 架构通过双向 Transformer 编码器捕捉文本全局语境，显著提升了语义理解的深度与准确性，使得模型 F1 分数直接从 0.77 区间跨越至 0.90 区间。这一跃升充分印证了预训练语言模型在理解复杂口语化表达、挖掘隐含语义信息上的核心优势。

2. 浅层结构 vs. 残差增强：CNN + BERT 和 BiLSTM + BERT 模型虽然尝试在 BERT 编码层基础上增加局部特征提取模块，但从实验结果看，其 F1 分数（0.8994 和 0.8985）反而略低于 Vanilla BERT（0.9044）。这一现象值得深思：简单堆叠的 CNN 或 BiLSTM 在微调过程中，易产生语义信息损失或梯度弥散问题，甚至可能引入特征冗余，导致模型性能不升反降。

本文模型引入的残差块通过跳跃连接实现了全局特征与局部特征的“无损融合”。残差机制确保了 BERT 提取出的高层语义背景能够指导一维卷积层对“发臭”、“异物”、“腹泻”等风险关键词，从而在保持高精确率（0.9265）的同时，有效识别出隐晦的食品安全风险信号，实现了性能的进一步提升。

（二）模型稳健性与收敛特性的讨论

五折交叉验证与训练动力学监测结果进一步印证了本文所提模型的优越性，具体表现为以下两方面：

1. 极低的标准差：从表 5 可见，本文模型在五折交叉验证中的标准差仅为 0.0062，远低于 CNN + BERT 的 0.0177。这表明残差结构不仅能够提升模型的特征抽象能力，还发挥了类似正则化的作用，降低了模型对特定训练样本集的依赖性，增强了模型的泛化能力。特别是在 Fold 5 中，当 CNN + BERT 出现明显性能下滑（F1 分数降至 0.8635）时，本文模型依然稳健地保持在 0.9011，充分印证了多尺度特征融合机制在面对复杂、高噪声语料时具有更强的纠错能力和稳定性。

2. 收敛效率与过拟合控制：图 6 的训练动力学曲线显示，引入残差块后，模型的训练损失下降速度更快且更平滑。这验证了残差学习通过将优化目标从 $H_{(x)}$ 转变为 $F_{(x)}=H_{(x)}-x$，显著降低了深层网络的优化难度。验证集 F1 分数在第 2 轮训练即达到峰值 0.9153，配合早停策略，模型能够精准停留在泛化能力最强的阶段，有效避免了因网络结构过深导致的过拟合问题。这一收敛速度远快于传统深度模型，充分体现了残差连接对梯度流动的优化作用，提升了模型训练的效率与稳定性。

（三）垂直领域的实际应用价值

食品安全监测是一个对“漏报”高度敏感的领域，漏报的风险样本可能直接威胁消费者身体健康。本文所提模型在测试集上实现了 0.92 的风险类别 F1 分数（见表 6），召回率达到 0.89，相较于各类基线模型具有更强的工程实践意义，具体体现在以下三点：

1. 多尺度特征感知：O2O 评论文本通常由“全局情感倾向（如不满、愤慨）”和“局部事实描述（如菜里有头发、吃完肚子疼）”组成。本文模型通过拼接 pooler_output 全局语义向量和经过残差处理的局部特征向量，构建了多尺度特征空间。这种设计比单一维度的特征表征更适配食品安全监测这一复杂二分类任务，能够同时捕捉用户情感倾向与具体风险信号，提升风险识别的精准度。

2. 误报与漏报的平衡：从混淆矩阵（图 4）可见，本文模型在保持高精确率（非风险样本 0.98、风险样本 0.94）的同时，将漏报数量控制在 32 条，仅占总样本的 1.6%。尽管漏报问题仍是未来主要改进方向，但相较于传统方法，本文模型已显著降低了风险样本的遗漏概率，为模型的实际部署提供了可靠的技术基础。

3. 局限性：尽管模型整体表现优异，但误分类样本分析也揭示了其对极致反讽表达的处理能力不足，这类误分类样本占总错误样本的 40%。例如“这家的饭菜真‘干净’，吃完就拉肚子”这类正向词汇，而忽略后续的风险信号。这表明仅依靠文本内部特征进行语义识别仍存在上限，未来研究可尝试引入外部知识图谱（如食品卫生专业词库）或多模态信息（评论配图），进一步

实现语义增强，提升模型对复杂隐晦表达的识别能力。

四、结论与展望

（一）研究结论

本文针对 O2O 餐饮平台食品安全监管的迫切需求，提出并实现了一种融合 BERT 与残差网络的食品安全评论识别模型。通过在真实 O2O 评论数据集上的系统实验与对比分析，得出以下核心结论：

1. 特征融合的有效性：实验结果证明，单一预训练模型在处理细粒度食品安全风险信号时存在明显局限。本文提出的 BERT + Residual Block 架构，通过残差连接实现了全局语义背景与局部敏感特征（如“异物”、“腹泻”等关键短语）的协同建模，其测试集 F1 分数达到 0.9096，验证集最高 F1 分数达 0.9257，显著优于传统机器学习模型及简单叠加的深度学习模型，充分验证了特征融合策略的有效性。

2. 残差结构的优越性：残差块的引入不仅提升了模型的特征抽象能力，更通过跳跃连接有效缓解了模型微调过程中的梯度不稳定问题。五折交叉验证结果显示，本文模型的标准差仅为 0.0062，展现出极强的稳健性与泛化能力，能够有效应对 O2O 评论文本高噪声、口语化、非规范的 data 分布特点。

3. 工程应用价值：模型在风险类样本的召回率上表现优异，整体准确率达到 97.5%。通过参数敏感性分析，确定了模型的最优学习率（3e-5）与残差块数量（k=4），为智慧监管系统在海量用户评价中自动预警食品安全风险，提供了一套高效、稳定、可落地的技术方案，具有重要的工程实践价值。

（二）不足与展望

尽管本文模型在特征提取、稳健性、工程应用等多个维度取得了突破，但仍存在进一步优化的空间，未来研究方向主要包括以下三点：

1. 语义理解的深度优化：误分类分析表明，模型对“反讽”及“隐晦表达”的识别仍面临挑战。未来研究可尝试引入情感增强向量或外部食品安全知识图谱，通过知识注入（Knowledge Injection）技术，提升模型对特定行业语境下反向语义、隐晦风险信号的捕捉能力，进一步降低误分类率。

2. 多模态特征融合：目前模型的监测仅依赖文本信息，而 O2O 平台评论通常伴随用户上传的实拍图片。未来可探索视觉 – 文本双模态模型，利用图像中的异物、环境卫生等视觉信息，补齐文本描述的不足，实现多维度风险识别，进一步降低漏报率。

3. 模型实时性优化：为适应大规模在线监测场景的需求，后续可研究模型蒸馏（Distillation）技术，在保留残差结构性能优势的前提下，降低模型参数量与计算开销，实现模型的轻量化部署，提升风险监测的实时性。

参考文献

[1] 宋超王，方美芳 .O2O 模式在大学及周边的发展研究——以蚌埠高校为例 [J]. 山西农经 ,2018,(06):110-111.DOI:10.16675/j.cnki.cn14-1065/f.2018.06.083.

[2]Zhang Q, Wu A, Liu L, et al. Application of Machine Learning in Food Safety Risk Assessment[J]. Foods, 2025, 14(23): 4005.

[3]Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques[C]// Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. 2002: 79-86.

[4]Kim Y. Convolutional Neural Networks for Sentence Classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1746-1751.

[5]Tang D, Qin B, Liu T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2015: 1422-1432.

[6]Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. (2018-10-11). https://arxiv.org/abs/1810.04805.

[7]He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778.

[8] 梁小林，柳映堂，梁嬰，等 . 基于偏最小二乘支持向量机模型的个人信用评估研究 [J]. 湖南文理学院学报 (自然科学版),2021,33(04):6-10.

[9]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.

[10]CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.

[11]CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 785-794.

[12]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.

[13]KIM Y. Convolutional Neural Networks for Sentence Classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1746-1751.

[14]GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks, 2005, 18(5-6): 602-610.

[15]HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778.

[16]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.

