

## 面向海关进出口商品税率检测的归纳交互网络模型

吴安奇, 车超, 张强, 周东生

### 引用本文

吴安奇, 车超, 张强, 周东生. 面向海关进出口商品税率检测的归纳交互网络模型[J]. 计算机科学, 2023, 50(11A): 230200086-7.

WU Anqi, CHE Chao, ZHANG Qiang, ZHOU Dongsheng. Inductive Interactive Network Model for Customs Import and Export Commodity Tax Rate Detection [J]. Computer Science, 2023, 50(11A): 230200086-7.

---

### 相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

#### Similar articles recommended (Please use Firefox or IE to view the article)

##### [融合迭代式关系图匹配和属性语义嵌入的实体对齐方法](#)

Entity Alignment Method Combining Iterative Relationship Graph Matching and Attribute Semantic Embedding

计算机科学, 2023, 50(11A): 230200041-6. <https://doi.org/10.11896/jsjcx.230200041>

##### [基于深度跨模态信息融合网络的股票走势预测](#)

Deep Cross-modal Information Fusion Network for Stock Trend Prediction

计算机科学, 2023, 50(5): 128-136. <https://doi.org/10.11896/jsjcx.220400089>

##### [一种基于三维卷积的声学事件联合估计方法](#)

Sound Event Joint Estimation Method Based on Three-dimension Convolution

计算机科学, 2023, 50(3): 191-198. <https://doi.org/10.11896/jsjcx.220500259>

##### [有监督相似性保持的深度二阶哈希方法](#)

Supervised Similarity Preserving Deep Second-order Hashing

计算机科学, 2022, 49(11A): 210900021-8. <https://doi.org/10.11896/jsjcx.210900021>

##### [融合多层次信息的海关同义词识别方法](#)

Customs Synonym Recognition Fusing Multi-level Information

计算机科学, 2022, 49(11A): 210800197-5. <https://doi.org/10.11896/jsjcx.210800197>

# 面向海关进出口商品税率检测的归纳交互网络模型

吴安奇<sup>1</sup> 车超<sup>1</sup> 张强<sup>1,2</sup> 周东生<sup>1</sup>

<sup>1</sup> 大连大学先进设计与智能计算省部共建教育部重点实验室 大连 116622

<sup>2</sup> 大连理工大学计算机科学与技术学院 大连 116086

(wuanqiyes@163.com)

**摘要** 中国海关传统的人工商品税率审查方式存在效率低、判断依据不一致、精度不高等问题,使用文本分类方法对商品分类自动确定税率可以有效降低海关税率风险。但面向海关商品数据进行分类时,商品类别具有层次性,同一大类下的许多子类别的商品描述具有高度相似性,给商品分类带来了很大的挑战。因此,提出了一种归纳交互网络模型,在BERT和CNN基础上加入归纳和交互指导模块。在归纳模块中采用动态路由算法对CNN提取的特征进行迭代运算,可以有效解决相邻特征融合和冗余问题。同时,为了解决不同子类别之间特征相似问题,提高分类性能,引入交互指导模块,该模块主要是将归纳模块提取的特征信息同[CLS]分类向量进行交互。在真实的海关数据集上进行实验,实验结果表明,该方法能达到较好的效果,其中准确率高达92.98%,且性能明显优于各基线模型。

**关键词:** 中国海关;税率检测;归纳交互;动态路由;交互指导

中图法分类号 TP391

## Inductive Interactive Network Model for Customs Import and Export Commodity Tax Rate Detection

WU Anqi<sup>1</sup>, CHE Chao<sup>1</sup>, ZHANG Qiang<sup>1,2</sup> and ZHOU Dongsheng<sup>1</sup>

<sup>1</sup> Key Laboratory of Advanced Design and Intelligent Computing, Ministry of Education, Dalian University, Dalian 116622, China

<sup>2</sup> School of Computer Science and Technology, Dalian University of Technology, Dalian 116086, China

**Abstract** The traditional way of examining the tax rate of manual goods in China Customs has problems such as low efficiency, inconsistent judgment basis and low precision. Using text classification method to automatically determine the tax rate of commodity classification can effectively reduce the risk of customs tax rate. However, when classifying customs commodity data, commodity categories are hierarchical. Many sub-categories under the same category have highly similar commodity descriptions, which brings great challenges to commodity classification. Therefore, an inductive interactive network model is proposed, and inductive and interactive guidance modules are added on the basis of BERT and CNN. In the induction module, the dynamic routing algorithm is used to perform iterative operation on the features extracted by CNN, which can effectively solve the problem of adjacent feature fusion and redundancy. At the same time, in order to solve the feature similarity problem between different subcategories and improve the classification performance, the interactive guidance module is introduced, which is mainly to interact the feature information extracted by the induction module with [CLS] classification vector. Experiment is carried out on the real customs data set, and the results show that the method can achieve good results, the accuracy is up to 92.98%, and the performance is obviously better than that of each baseline model.

**Keywords** China Customs, Tax rate detection, Inductive interaction, Dynamic routing, Interactive guidance

### 1 引言

中国是世界上最大的贸易国之一,据中国海关统计,2022年前三季度,中国进出口总值达到了31.11万亿元人民币,同比增长9.9%。目前海关主要采用传统人工方式对进出口商品税率进行审查,效率低、判断依据不一致、精度不高是其主要存在的问题。随着自然语言处理技术的迅速发展,将人工

智能技术应用在海关税率检测中,可以有效提高海关税率检测的准确性和效率。海关商品税率检测指根据商品申报的文本信息,来判断商品的类别。本文主要依据世界海关组织协调制度(HS)编码作为分类标准,完成海关进出口商品的税率检测。

海关商品数据以海关特定的格式和商品专业术语的形式呈现,每个申报文本中商品信息都是由多个要素组成,他们之

基金项目:高等学校学科创新引智计划(D23006);辽宁省“兴辽英才计划”(XLYC2008017);国家自然科学基金(62076045,62102058);辽宁省教育厅服务地方项目(揭榜挂帅)(LJKFZ20220290)

This work was supported by the 111 Project(D23006),LiaoNing Revitalization Talents Program(XLYC2008017),National Natural Science Foundation(62076045,62102058) and Educational Department of Liaoning Provincial(LJKFZ20220290).

通信作者:张强(liubin@dlut.edu.cn)

间通过明确的分隔符“|”进行分割,没有明确的语义联系,分割的元素之间没有连续的语义。如图 1 所示,商品的文本申报信息由商品编号、商品序号、HS 编码、商品名称、规格型号多个部分组成,其中商品的规格型号又由品牌类型、出口享惠情况等多个申报要素组成。海关进出口商品税率检测任务便是通过商品文本申报信息来预测出商品的 HS 编码,从而判断商品的类别,进而协助海关进行商品税率查验,判别该商品是否存在偷税漏税现象。

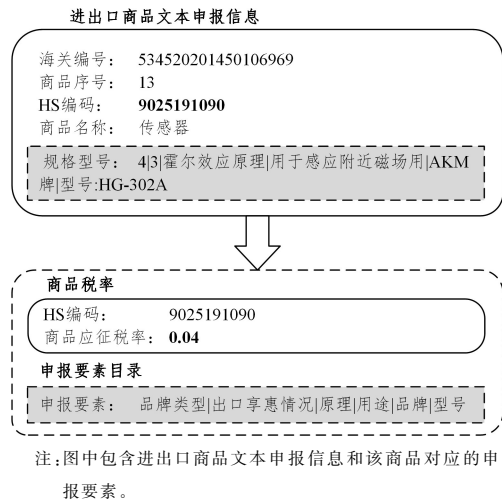


图 1 商品征税示例

Fig. 1 Examples of commodity taxation

当前文本分类的方法主要以深度学习算法为主<sup>[1-2]</sup>。其中基于预训练模型进行本文分类已经相对成熟,如 ELMO<sup>[3]</sup>,Transformer<sup>[4]</sup>,BERT<sup>[5]</sup>等模型近年来不仅在文本分类中取得优秀成绩,在自然语言处理的实体识别、问答系统、文本生成等领域也取得了优秀的业绩。BERT 是基于双层 Transformer 编码器的双向结构,其掩码词预测任务 (MLM) 可以有利于增强模型对上下文的记忆,下一句预测任务 (NSP) 可以让模型更好地理解句子间的关系<sup>[5]</sup>。但是海关不具有上下文语义信息,无法很好地学习到要素的语义信息。且常用的文本分类模型大多是针对文本进行序列处理,通过对文本序列的学习进行文本分类。其次海关商品数据以海关特定的格式和商品专业术语的形式呈现,每个申报文本中商品信息都是由多个要素组成,同时商品类别具有层次性,同一大类下的许多子类别的商品描述不可避免地具有相似性。因此采用常规分类模型进行海关商品分类的不足之处主要表现在两个方面:(1)容易造成特征信息重叠和冗余,降低分类精度;(2)针对不同类别中具有相似性特征的分类,容易造成分类误差。

卷积神经网络 (CNN)<sup>[6]</sup>在融合多层次特征时性能优越,让人留下深刻的影响,但对于海关商品数据这种高度结构化文本,CNN 可能会过度融合特征,导致错误的分类<sup>[7]</sup>。例如图 1 中,“霍尔效应原理”和“AKM 牌”属于不同类型和规格的要素,在卷积层中会被嵌入融合在一起,导致特征过度融合。而堆积层又会使这种融合变得更加复杂,因此会导致错误的分类。为了有效避免这种特征的过度融合,我们引入了归纳模块,利用动态路由算法<sup>[8]</sup>对 CNN 提取的数据特征进行迭代运算,迭代运算会根据数据信息决定信息流出输出向量的种类和数量,做到特征信息的唯一提取,避免相邻特征

信息的过度融合和特征信息冗余问题,捕捉到更加清晰的特征信息,提高分类的性能。

其次,海关商品类别具有层次性,同一大类下的许多子类别的商品描述具有相似性,对这种特征相似的商品进行分类存在有困难,因此如何提高具有相似特征文本的分类准确性是需要解决的问题。BERT 在第一句前会加一个 [CLS] 标志,[CLS] 本身没有语义,主要是用于分类输出的特殊符号,单纯地依赖其对海关商品数据进行分类,效果不够理想。我们设计了一个辅助分类任务——交互指导模块,该模块将动态路由提取到的特征向量同 [CLS] 利用 Transformer 的自注意力计算进行交互,有效地对商品分类进行指导,提高了分类的精准度。因此,我们针对海关进出口商品税率检测任务提出了一种归纳交互网络模型。本文的贡献主要在于以下 3 个方面:

1) 引入了归纳模块,利用动态路由算法对 CNN 提取的数据特征进行迭代运算,可以避免相邻特征信息的过度融合和特征信息冗余问题,捕捉到更加清晰的特征信息,提高分类的性能。

2) 设计了一个辅助分类任务——交互指导模块,该模块将动态路由提取到的特征向量同 [CLS] 进行交互,可以有效地对商品分类进行指导,提高海关商品类别之间特征相似背景下分类的精准度。

3) 在两个真实的海关商品数据集上进行实验,实验结果表明,本文模型相比其他基线模型能够取得优秀的实验结果。

## 2 相关工作

目前同本文的研究联系较为密切的文本分类方法,大致可以分为基于卷积神经网络、基于注意力算法、基于预训练语言模型的文本分类 3 个部分。

### 2.1 基于卷积神经网络的文本分类

卷积神经网络 (CNN) 最初被提出用于图像分类,其原理主要是利用卷积滤波器提取图像的特征。由于 CNN 可以将不同卷积核应用于同一个序列的多个块上,因此它也被广泛应用于 NLP 任务,包括文本分类、机器翻译等任务。在文本分类任务中,通过将文本中的单词向量拼接成矩阵,将矩阵送入包含不同维度滤波器的卷积层中,将卷积层的结果经过池化层进行池化,得到最终的文本向量表示。

在此基础上, Kim 提出的 TextCNN 模型<sup>[1]</sup>可以更好地应用于本文分类任务,它可以更好地确定池化层中的判别短语,并通过静态的词向量来学习除词向量以外的超参数。Johnson<sup>[9]</sup>为了更好地处理未标记的数据,提出了基于双视图半监督学习的 CNN 文本分类模型,它可以获得更好的性能,同时也增加了计算复杂度。基于此,深度金字塔卷积神经网络 (DPCNN)<sup>[10]</sup>被提出,通过提高网络深度,不增加过多的计算开销,来进一步提高计算的精度。此外,针对文本嵌入单位的不同,Nguyen 等<sup>[11]</sup>以及 Adams 等<sup>[12]</sup>分别提出基于字典的深度学习方法及基于字符级的 CNN 模型,均取得了不错的效果。

### 2.2 基于注意力的文本分类

虽然基于 CNN 的神经网络模型在文本分类任务上取得了不错的效果,但是仍然存在可解释性差、分类误差较大等缺陷,因此基于注意力的方法被应用于文本分类任务。层次注

注意力 HAN 网络<sup>[13]</sup> 在 Bahdanau 等<sup>[14]</sup> 的基础上被提出,该模型包括两个编码器和两个不同的注意力层,注意力算法可以对特定的输入分配不同的注意力,以学习每个词句对分类判断的贡献。此外,LSTMN<sup>[15]</sup> 被提出,它从左到右逐步处理文本,并通过记忆和注意进行浅层的推理。为了解决跨文本分类和少文本分类的问题,双注意模型和 HAPN<sup>[16]</sup> 被提出。

自注意力机制 Self-attention<sup>[4]</sup> 通过构建句子之间的  $K, Q$  和  $V$  矩阵来捕获句子中单词的权重分布,这些矩阵可以捕获对文本分类的长期依赖关系。在基于方面的情绪分析(ABSA)<sup>[17-18]</sup> 中,MGAN<sup>[19]</sup> 提出了一种细粒度的注意机制和粗粒度的注意机制,用于学习上下文和方面之间的单词级交互。抽取式问答可以被看作是文本分类任务,它输入一个问题和多个候选答案,并对每个候选答案进行分类,以识别正确的答案。为了捕捉问答任务中每个问题和候选答案之间复杂的语义关系,Tan 等<sup>[20]</sup> 引入了 CNN 和 RNN,并通过使用受问题上下文影响的简单单向注意算法来生成答案嵌入。

### 2.3 基于预训练语言模型的文本分类

基于预训练语言模型的文本分类大大推动了自然语言领域的快速发展,其中就包括文本分类任务<sup>[21]</sup>。2018年,Peter 提出了 ELMO 语言模型<sup>[3]</sup>,它是一种上下文相关的文本表示方法,能够有效处理一词多义问题。ELMO 通过使用针对语言模型训练好的双向 LSTM<sup>[22]</sup> 来构建文本表示,由此捕捉上下文相关的词义信息。ELMO 模型不仅简单,而且表现出众,在自然语言处理领域的 6 个典型下游任务的数据集上全面刷新了最优成绩,尤其是在阅读理解任务上效果提升显著<sup>[23]</sup>。其主要贡献是提供了一种新的文本表征的思路。Transformer 在处理长期依赖性方面比 LSTM 有更好的表现,它在机器翻译等任务上取得的成果也使一些业内人士开始认为其是 LSTM 的替代品。在此背景下,Open AI 使用生成式方法的 GPT<sup>[24]</sup> 预训练模型也应运而生。

GPT 模型虽然达到了很好的效果,但本质上仍是一种单向语言模型,对语义信息的建模能力有限。因此,建立一个基于 Transformer 的双向预训练语言模型是一种重要的研究思路。BERT 使用了一种特别的预训练任务来解决这个问题,它通过堆叠 Transformer 子结构来构建基础模型,并通过 Masked-LM 这个特殊的预训练方式达到了真双向语言模型的效果。

BERT 在 11 项基本任务中都有了显著的提升效果,其卓越的表现对自然语言处理领域的发展有着重大的推动作用。自然语言处理领域的许多后续研究一般也以 BERT 为基础进行改进,学术界认为 BERT 模型的出现使得自然语言处理领域可以像计算机视觉那样进行迁移学习<sup>[25]</sup>。

## 3 问题描述

本节将介绍海关进出口商品税率检测任务以及该任务的一些符号和术语。

### 3.1 数据预处理

本文所采用的原始实验数据来自于海关真实商品数据的集合,为了更好地验证模型的效果,我们对数据进行了预处理。首先对商品数据进行去噪和规范化处理,然后对商品文本信息中每一个要素分配独立的属性,并使用“|”分隔符对其进行规范化分割。最后如图 2 所示,将商品要素名称和要素

内容进行一一对应产生映射关系,并对该商品数据集进行随机打乱,这样就构建了一份可机器学习的标准数据集。

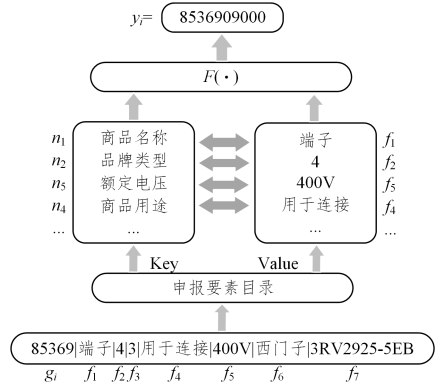


图 2 海关商品税率检测任务抽象模型

Fig. 2 Abstract model of customs commodity tax rate detection task

### 3.2 商品标准化申报目录

为了规范申报行为,提高申报数据的质量,海关总署制定了《中华人民共和国海关进出口商品规范申报目录》,对海关商品文本申报信息的要素信息进行了规范化,通过这一申报目录,我们建立要素名称和要素内容的对应关系。如图 2 所示,根据该商品类别以及要素名称( $n_1, \dots, n_{l-1}$ )和要素内容( $f_1, \dots, f_l$ )的对应关系,该商品文本申报信息可以表示为  $x = \{(f_1, \dots, f_l), (n_1, \dots, n_{l-1}), g\}$ 。

### 3.3 海关进出口商品税率检测任务

海关总署会根据商品的申报文本信息对每一种进出口商品进行检验,如图 1 所示,商品的申报文本信息包含了商品编号、商品序号、HS 编码、商品名称、规格型号等多个部分,完整地代表了商品的基本信息,这也是海关进行征税的依据。假设第  $i$  个要素内容被定义为  $f_i$ ,那么一条具有  $l$  个要素的商品文本便可以用  $(f_1, \dots, f_l)$  来表示。

海关进出口商品税率检测任务,归根结底就是根据商品的文本申报信息进行商品分类。具体描述可以表示为,给定一个具有  $N$  个商品数量的训练数据集  $D_{train} = (x_1, \dots, x_n)$ ,其中每一个商品数据可以表示为  $x_i = \{(f_1, \dots, f_l)_i, (n_1, \dots, n_{l-1})_i, g_i\}$ ,通过训练一个模型  $M$  来对商品类别  $y_i$  进行预测,具体任务可以表示为  $M(x_i) = y_i$ 。

## 4 本文方法

### 4.1 模型架构

本文模型的架构如图 3 所示,首先利用 BERT 对数据进行 embedding 处理,将处理完之后的文本向量输入到多个 CNN 的卷积层中得到特征矩阵,虽然 CNN 具有很强的特征提取能力,但 CNN 提取到的特征可能会导致相邻特征信息融合以及特征信息冗余,为了解决这种问题,我们引入了归纳模块对特征信息进行迭代处理。归纳模块利用动态路由算法,通过权重分配不断的迭代,迭代运算会根据数据信息决定信息流出输出向量的种类和数量,输出唯一的商品特征信息。然后在交互指导模块的作用下,将提取到的特征信息同 [CLS] 特征信息利用自注意力计算方式进行交互,进而对 [CLS] 的分类进行指导,提高海关商品分类的准确性。

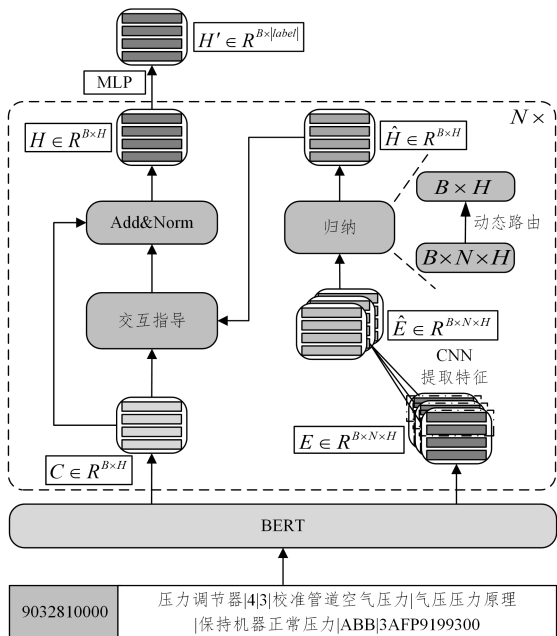


图3 归纳交互网络模型架构

Fig. 3 Summarize the interactive network model architecture

## 4.2 卷积神经网络

CNN 具有良好的特征提取能力,本文首先利用 BERT 对文本进行 Embedding 处理,然后送入 CNN 中,将输入文本中的单词向量拼接成一个矩阵,然后将矩阵输入到具有不同维度滤波器的卷积层。最后,将卷积层的结果经过池化层,并将池化结果连接起来,得到文本的特征向量表示  $\hat{\mathbf{E}}_i$  ( $\hat{\mathbf{E}}_i$  是第  $i$  个特征向量的输出),特征向量的维度为  $R^{B \times N \times H}$ 。但是类似于 CNN 这样的神经网络模型,应用于海关这种结构化文本时,很容易造成特征信息过度融合以及特征信息冗余这种现象。为了更好地对海关商品进行特征提取,降低分类错误率,我们在归纳模块中引入了动态路由算法,利用迭代运算对特征信息进行权重分配,得到清晰有效的特征信息。

## 4.3 归纳模块

本节主要介绍归纳模块中的动态路由算法。针对 CNN 提取的数据特征信息,动态路由算法通过迭代更新计算,每次迭代都会根据数据信息决定信息流向输出向量的种类和数量,因此通过多次迭代,可以做到特征信息的唯一提取,这样可以解决特征的过度融合和重复问题,捕捉到更加清晰的特征信息,提高分类的性能。具体原理如下,将 CNN 提取的文本特征向量  $\hat{\mathbf{E}}_i$ ,通过归纳模块归纳出商品特征向量  $\hat{\mathbf{H}}_j$  ( $\hat{\mathbf{H}}_j$  是类别  $j$  的向量输出),主要目的是设计一个从样本向量到类向量的非线性映射:

$$\{\hat{\mathbf{E}}_i \in R^{B \times N \times H}\} \mapsto \{\hat{\mathbf{H}}_j \in R^{B \times H}\} \quad (1)$$

挤压函数(Squash)是一个非线性压缩函数,它使向量的方向保持不变,将文本特征向量长度缩减为 1,特征输出向量  $\hat{\mathbf{H}}_j$  的计算式如下:

$$\text{squash} = \frac{\|x\|^2}{1 + \|x\|^2} \frac{x}{\|x\|} \quad (2)$$

$$\hat{\mathbf{H}}_j = \text{squash}(\tilde{\mathbf{H}}_j) \quad (3)$$

给定每个文本预测向量  $\hat{\mathbf{E}}_{j|i}$ ,每个特征候选向量  $\tilde{\mathbf{H}}_i$  是第  $i$  类中所有文本预测向量  $\tilde{\mathbf{E}}_{ij}$  的加权和,其中  $d_{ij}$  是由迭代动态

路由过程确定的耦合系数,表示该类中所有文本之间的连接概率。

$$\tilde{\mathbf{H}}_j = \sum_i d_{ij} \cdot \tilde{\mathbf{E}}_{j|i} \quad (4)$$

特征向量  $\hat{\mathbf{E}}_i$  乘以共享权重矩阵  $\mathbf{W}_j$  运算得到对应的预测向量  $\hat{\mathbf{E}}_{j|i}$ 。

$$\tilde{\mathbf{E}}_{j|i} = \mathbf{W}_j \hat{\mathbf{E}}_i \quad (5)$$

为了确保类向量总结提取文本特征向量,应用动态路由进行迭代。在每次迭代中,该过程动态地调整连接强度,并通过计算使该类中所有文本之间的耦合系数  $d_i$  之和为 1,  $b_i$  在第一次迭代中初始化为 0。

$$d_{ij} = \frac{\exp(b_{ij})}{\sum_j \exp(b_{ij})} \quad (6)$$

通过每次迭代更新  $b_{ij}$  的值,进行自上而下的反馈。如果产生的类候选向量与一个文本预测向量有一个大的标量输出,增加该文本的耦合系数,降低其他文本的耦合系数。每个  $b_{ij}$  更新为:

$$b_{ij} = b_{ij} + \tilde{\mathbf{E}}_{ij} \cdot \hat{\mathbf{H}}_j \quad (7)$$

通过归纳模块利用动态路由算法对文本向量进行迭代更新,对商品特征信息进行有效的归纳,减少特征信息重复和冗余,得到特征信息的唯一提取,提高分类的性能。

## 4.4 交互指导模块

BERT 利用 [CLS] 在分类任务上表现出良好的性能,但是针对特殊场景,如海关商品语境下的文本分类,并不能很好地体现优势。因此,本文利用交互指导模块对通过归纳模块提取到的特征信息同 [CLS] 进行交互处理,提高分类的准确性。交互指导的基本原理如图 4 所示,将两个不同注意层的输出作为自注意力层的输入,这样利用相互的输出信息进行交互处理,具体的过程和原理如下。

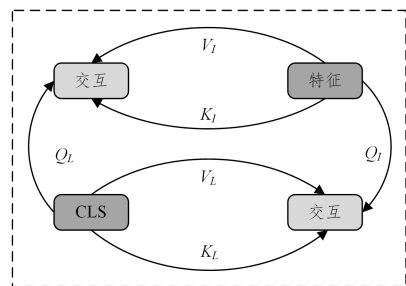


图4 交互指导示意图

Fig. 4 Interactive guidance diagram

$\mathbf{H}_I \in R^{B \times H}$  和  $\mathbf{H}_L \in R^{B \times H}$  分别代表归纳模块的输出特征信息和分类向量 [CLS] 的特征信息,通过自注意力机制进行交互指导,实现两个任务的双向连接。与普通的 Transformer 相同,我们使用不同的线性投影将矩阵  $\mathbf{H}_I$  和  $\mathbf{H}_L$  映射到查询 ( $\mathbf{Q}_I, \mathbf{Q}_L$ )、键 ( $\mathbf{K}_I, \mathbf{K}_L$ ) 和值 ( $\mathbf{V}_I, \mathbf{V}_L$ ) 矩阵,  $\mathbf{C}_L$  为输出的加权和。然后经过残差连接和层归一化处理得到特征向量  $\mathbf{H}$ 。

$$\mathbf{C}_L = \text{softmax} \left( \frac{\mathbf{Q}_L \mathbf{K}_I^T}{\sqrt{d_I}} \right) \mathbf{V}_I \quad (8)$$

$$\mathbf{H} = \text{LN}(\mathbf{H}_L + \mathbf{C}_L) \quad (9)$$

不同于自注意力机制的隐式交互,交互模块可以明确地考虑两个不同任务之间的交叉影响,通过双向连接进行交互。在海关商品分类任务中,商品类别具有层次性,不同子类别之

间特征相似,通过交互处理建立动态路由的输出特征和BERT的分类向量[CLS]之间的双向连接,对[CLS]分类进行指导,能够有效提高分类的性能。

## 5 实验

### 5.1 数据集介绍

本文根据海关真实进出口商品数据集自定义了数据集1和数据集2两个数据集,通过对两个数据集的实验来验证模型的性能。数据集的详细情况如表1所列。

表1 海关数据集统计

Table 1 Customs data set statistics

	数据集	训练集	验证集	测试集
数据集1	商品条数	39871	5000	5000
	商品类别数	966	764	751
数据集2	商品条数	16000	2000	2000
	商品类别数	266	254	238

数据集1:该数据集的数据情况如表1所列,其共包含49871条数据,总体分布同中国海关真实进出口商品申报文本相似,通过对数据预处理之后形成格式和内容标准的数据集,然后采用随机分布的方式构建数据集,并按照标准比例划分训练集、验证集和测试集。

数据集2:通过表1可以发现,该数据集总体数据量远远少于数据集1,共包含20000条数据。因为该数据集是根据海关商品税率检测中错误的商品税率申报信息所构成的,所以在此数据集上进行实验,能够更好地验证模型在实际应用中的性能。

### 5.2 实验设置

本文实验是在Pytorch环境下进行的,电脑的处理器的Core i9-10900F @ 2.80 GHz,内存大小为64 GB,显卡为NVIDIA GeForce RTX 3090。部分超参数取值如表2所列。

表2 本文模型的超参数取值

Table 2 Value of hyperparameters of the proposed model

	参数名	参数值
通用超参数	Embedding	300
	Max_length	128
	Epoch	10
	Early_stop	3000
	Optimizer	Adam
	Batch_size	128
本文方法	Learning_rate	$2 \times 10^{-5}$
	Drop_out	0.1
	Batch_size	32
	Epoch	30
	Iteration	3

### 5.3 实验结果

#### 5.3.1 评价指标

本文的评价指标选择了4种广泛应用于文本多分类任务的指标,包括准确度(Acc)、加权精度(WP)、加权召回率(WR)和F1值(F1),用于比较各个模型的性能。其中涉及到的符号含义如表3所列。 $TP$ 和 $FP$ 分别为预测为正类的正样本和负样本, $TN$ 和 $FN$ 分别是预测为负类的正样本和负样本。

Acc准确度的计算式为:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Precision的计算式为:

$$P = \frac{TP}{TP + FP} \quad (11)$$

Weighted Precision的计算式如下:

$$WP = P_{class_1} * (num_{class_1} / num_{all\_class}) + \dots + P_{class_n} * (num_{class_n} / num_{all\_class}) \quad (12)$$

其中, $num_{class_n}$ 表示第 $n$ 个类别的数量, $num_{all\_class}$ 表示所有类别的数量。

Recall的计算式为:

$$R = \frac{TP}{TP + FN} \quad (13)$$

Weighted Recall的计算式为:

$$WR = R_{class_1} * (num_{class_1} / num_{all\_class}) + \dots + R_{class_n} * (num_{class_n} / num_{all\_class}) \quad (14)$$

F1的计算式为:

$$F1 = \frac{2P * R}{P + R} \quad (15)$$

表3 混淆表

Table 3 Confusion

	正例	负例	总计
正例	$TP$	$TN$	$P$
负例	$FP$	$FN$	$N$
总计	$P'$	$N'$	$P + N$

#### 5.3.2 实验结果分析

将本文方法与7种基线方法进行比较,以更好地验证混合文本的实际性能,其中包括一般的文本分类算法,如TextCNN和BERT,以及最新的一些方法。详细的基线概述如下。

1)TextCNN<sup>[1]</sup>:一个经典的卷积模型。它使用多个卷积内核来捕获文本的局部语义信息,并使用最大池化层来捕获最重要的特征。

2)Transformer<sup>[4]</sup>:使用注意力算法提取输入与输出的依赖关系,相比RNN的顺序结构,改善了训练慢的缺陷,使得模型并行训练,并且具有全局信息。

3)BERT<sup>[5]</sup>:BERT的全称是来自变压器的双向编码器表示,一个以双向多层变压器编码器为主要模型结构的预训练语言模型,在11项任务中取得了显著的效果。

4)RoBERTa<sup>[26]</sup>:在BERT的基础上进行改进,主要是使用更一般的训练语料,优化预训练策略,在一些任务上能够取得比BERT更好的效果。

5)BART<sup>[27]</sup>:一种预训练语言模型,使用编码器加解码器架构,在文本生成领域取得了领先的成果。

6)ERNIE3.0<sup>[28]</sup>:通过融合自回归网络和自动编码网络,在大型语料库上训练的大规模知识增强模型。

本文在两个海关真实数据集上进行实验,通过不同的评价指标进行衡量,本文模型和其他基线模型的实验结果如表4所列。在数据集1中,本文模型的分分类准确率可以达到92.98%,超过了其他所有基线的分数,可见本文方法针对海关商品分类任务是十分有效的。与TextCNN和Transformer相比,本文模型的准确率大概提高了5%左右。相比于BERT,准确率提高了3%,在与RoBERTa进行对比时,准确率提升了2%。由此可见本文模型在海关进出口商品税率检测任务上还是更有优势的。

同时也在数据集 2 上进行了对比,正如前文的介绍,该数据集是根据海关商品税率检测中错误的商品税率申报信息所构成的。在此数据集上进行实验,能够更好地验证模型在实际应用中的性能。与 BERT 和 RoBERTa 进行对比,准确率能够提高 2.5% 左右。这说明本文模型在实际海关商品分类应用中,依旧能够取得良好的效果,具有很强的实用性和适用性。

表 4 不同模型在海关数据集上的实验结果

Table 4 Experimental results of different models on customs

datasets

(单位:%)

数据集	实验模型	Acc	WP	WR	F1
数据集 1	TextCNN	88.08	87.91	88.08	87.74
	Transformer	88.33	88.68	88.33	87.81
	BERT	89.83	89.42	89.83	89.14
	BART	89.13	88.70	89.13	88.28
	ERNIE3.0	90.94	91.27	90.94	90.52
	RoBERTa	90.76	90.95	90.76	90.63
	本文方法	<b>92.98</b>	<b>93.04</b>	<b>92.98</b>	<b>92.66</b>
数据集 2	82.36	82.75	82.36	81.49	
	ERNIE3.0	83.99	83.76	83.99	83.53
	BERT	82.84	79.62	82.84	80.21
	RoBERTa	82.97	83.14	82.97	82.76
	本文方法	<b>85.29</b>	<b>84.15</b>	<b>85.29</b>	<b>83.91</b>

#### 5.4 消融实验

为了展示本文方法的有效性,我们在真实的海关数据集 2 上进行了消融实验,实验结果如表 5 所列。在本文实验的基础上分别去除归纳模块和交互指导模块,得到的实验结果分别为 84.33% 和 84.88%。归纳模块主要是利用动态路由算法对 CNN 提取到的特征信息进行迭代更新计算,这样可以解决特征的过度融合和冗余问题,捕捉到更加清晰的特征信息,提高特征信息提取的有效性,降低分类的错误率,通过实验可以看出该模块能够提高准确率 0.96%,说明该模块能够很好地对特征信息进行处理。

表 5 在海关数据集 2 上进行的消融实验结果

Table 5 Results of ablation experiments on customs data set 2

模型	Acc/%
本文方法	85.29
去除归纳模块	84.33
去除交互指导模块	84.88

交互指导模块主要用于解决同类别商品之间特征相似问题,提高了分类准确性,主要是将归纳模块提取的特征同 [CLS] 分类特征向量进行交互,对 [CLS] 分类进行指导。通过实验证明该模块能够表现出不错的作用,分类准确率能够提高大约 0.4%。由此可见,这两个模块针对海关税率检测任务能够起到有效的作用,提升分类准确率,同时归纳模块比交互模块的提升效果更为明显。

#### 5.5 案例分析

为了直观验证归纳交互网络模型在商品分类任务中的有效性,本文进行了案例研究。表 6 列出了同属于大类 90328 的两种商品的申报文本。由于是同一大类,这两种商品的申报文本非常相似,因此使用归纳模块来对这种特征相似的商品进行分类十分有用。

然而,使用普通卷积网络进行商品分类时,由于特征信息高度相似可能引起特征融合和冗余问题,导致模型无法学习

到同一类别下的两种商品的有效特征,造成这两种商品被错误分类。而通过使用归纳交互网络模型,归纳模块会对特征信息进行迭代处理,防止两种商品的特征信息重复冗余,能够更加准确地提取特征信息。同时交互模块也会根据提取到的差异性要素“原理”和“用途”的信息来更好地辅助 [CLS] 对商品进行分类。

表 6 一个分类错误的例子

Table 6 Example of wrong prediction of HS Code

商品大类	要素名称
90328	商品名称 品牌类型 出口享惠情况 驱动方式 原理 用途 品牌 型号
海关编码	要素内容
9032810000	压力调节器 4 3 校准管道空气压力 气压压力原理 保持机器正常压力 ABB 3AFP9199300
9032899090	调节器 4 3 控制电路电流 利用晶管体的开关作用 自动调节开关通断 COSEL牌 LDA15F-12-Y

**结束语** 本文提出了一种面向海关进出口商品税率检测的归纳交互网络模型。该模型主要是利用归纳模块中的动态路由算法对 CNN 特征信息进行迭代运算,以提高特征提取的准确性,捕捉到更加清晰的特征信息,提高特征信息提取的有效性,降低分类的错误率;利用交互指导模块将归纳模块提取到的特征信息与 [CLS] 进行交互,对分类进行辅助指导,提高最终的分分类精确度。在真实的海关商品数据集上进行实验,实验结果表明,本文方法较其他基线模型、分类准确率能够提升 2%。同时我们观察到,本文方法在样本不足时,实验结果有所变差,后续我们会在此基础上进一步探究在少样本的海关商品数据集上,如何有效地提升分类的准确性。

#### 参考文献

- [1] KIM Y. Convolutional Neural Networks for Sentence Classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1746-1751.
- [2] JOHNSON R, ZHANG T. Deep Pyramid Convolutional Neural Networks for Text Categorization[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017:562-570.
- [3] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]// Annual Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2018:2227-2237.
- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017:6000-6010.
- [5] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding [C]// The North American Chapter of the Association for Computational Linguistics. 2018:4171-4186.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks[C]// Conference and Workshop on Neural Information Processing Systems. 2012:1097-1105.
- [7] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for natural language processing [J]. KI-Künstliche Intell, 2016, 26(6):180-189.

- [8] SABOUR S, FROSST N, HINTON G E. Dynamic Routing Between Capsules[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;3856-3866.
- [9] JOHNSON R, ZHANG T. Semi-supervised Convolutional Neural Networks for Text Categorization via Region Embedding [J]. Advances in neural information processing systems, 2015, 28(5):919-927.
- [10] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017;562-570.
- [11] NGUYEN H, NGUYEN M. A Deep Neural Architecture for Sentence-Level Sentiment Classification in Twitter Social Networking [C] // Computational Linguistics: 15th International Conference of the Pacific Association for Computational Linguistics, PACLING, 2018;15-27.
- [12] ADAMS B, MCKENZIE G. Crowdsourcing the character of a place: Character-level convolutional networks for multilingual geographic text classification [J]. Transactions in GIS, 2018, 22(2):394-408.
- [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] BAHDANAU D, CHO K H, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//3rd International Conference on Learning Representation. ICLR, 2015.
- [15] CHENG J, DONG L, LAPATA M. Long short-term memory networks for machine reading [J]. EMNLP, 2016;551-561.
- [16] SUN S, SUN Q, ZHOU K, et al. Hierarchical attention prototypical networks for few-shot text classification[C]//Proceedings of the 019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019;476-485.
- [17] MUBAROK M S, ADIWIJAYA, ALDHI M D. Aspect-based sentiment analysis to review products using Naïve Bayes[C]//AIP conference proceedings. AIP Publishing LLC. 2017; 020060.
- [18] MA Y, PENG H, CAMBRIA E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018;5876-5883.
- [19] FAN F, FENG Y, ZHAO D. Multi-grained attention network for aspect-level sentiment classification [C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;3433-3442.
- [20] TAN M, DOS SANTOS C, XIANG B, et al. Improved representation learning for question answer matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016;464-473.
- [21] HOWARD J, RUDER S. Universal Language Model Fine-tuning for Text Classification[C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018;328-339.
- [22] YU Y, SI X, HU C, et al. A review of recurrent neural networks; LSTM cells and network architectures [J]. Neural Computation, 2019, 31(7):1235-1270.
- [23] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014;1717-1724.
- [24] FLORIDI L, CHIRIATTI M. GPT-3: Its nature, scope, limits, and consequences [J]. Minds and Machines, 2020, 30(4):681-694.
- [25] 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. 2016;770-778.
- [26] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach [J]. arXiv:1907.11692, 2019.
- [27] LEWIS M, LIU Y, GOYAL N, et al. BART: Denoising sequence-to-sequence pretraining for natural language generation, translation, and comprehension [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020;7871-7880.
- [28] SUN Y, WANG S, FENG S, et al. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation [J]. arXiv:2107.02137, 2021.



**WU Anqi**, born in 1997, postgraduate. His main research interests include natural language processing and so on.



**ZHANG Qiang**, born in 1971, Ph.D., professor. His main research interests include biological computing and artificial intelligence, big data analysis and processing.