

基于双重预训练的商品属性分类方法

赵哲宇 王中卿 王红玲

苏州大学计算机科学与技术学院 江苏 苏州 215006

摘要 商品属性分类任务是指对一段商品的描述文字进行属性分析并进而对多个属性进行分类的过程,其有助于人们从多个角度了解商品,为市场营销、产品管理等提供帮助。当前大语言模型的使用也愈加广泛,但在商品属性分类问题上,通用大模型由于缺乏领域知识和属性关联等信息,性能不够理想。为此,提出了一个基于双重预训练的商品属性分类方法,旨在通过使用特定的预训练方式提高大语言模型在商品属性分类任务中的性能。在 T5 模型的基础上,引入了领域内文本预训练和基于属性间关联性的预训练两种方法。在 Clothing Fit Data 数据集上的实验结果显示,使用了双重预训练的 T5 模型较未经过预训练的模型以及其他基准模型,在各个属性上的分类效果都取得了一定提升。实验结果证明了所提方法的有效性。

关键词: 双重预训练;多属性分类;大语言模型;T5;商品属性分类

中图分类号 TP391

Commodity Attribute Classification Method Based on Dual Pre-training

ZHAO Zheyu, WANG Zhongqing and WANG Hongling

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract The commodity attribute classification task refers to the process of analyzing the attributes of a piece of merchandise based on its descriptive text and subsequently categorizing multiple attributes. This process aids in providing insights into merchandise from various perspectives, thereby assisting in marketing and product management. While the utilization of large language models is increasingly prevalent, their performance in commodity attribute classification tasks remains suboptimal due to the lack of domain knowledge and attribute correlations. To address this issue, this paper proposes a dual pre-training-based method for commodity attribute classification, aiming to enhance the performance of large language models in such tasks by employing specific pre-training techniques. Building upon the T5 model, this paper introduces two methods: domain-specific text pre-training and attribute correlation-based pre-training. These methods enhance the model's understanding of the specific task from both input and output text perspectives, facilitating the classification of multiple attributes of merchandise. Experimental results on the Clothing Fit Data dataset demonstrate that the dual pre-trained T5 model outperforms both non-pre-trained models and other baseline models in attribute classification, validating the effectiveness of the proposed approach.

Keywords Dual pretraining, Multi-attribute classification, Large language model, T5, Commodity attribute classification

1 引言

商品属性是描述商品特征或特性的一组信息,这些属性不仅包括商品本身的信息,还涵盖用户信息以及与用户商品相关联的信息^[1]。图 1 展示了服装商品属性的示例。服装本身属性包括服装编号、尺寸和种类等;用户信息包括身体、体重和年龄等;购买服装时产生的关联信息包括相称度、用户评价以及购买用途。

商家为提升服务水平,会对商品进行详细分类,并分门别类地进行有针对性的介绍和推广。因此,商品属性分类变得愈发重要。商品属性分类是指按照商品的各个属性进行自动分类,属于多标签分类。它能快速处理商品的描述或评论文本,高效准确地对商品属性进行分类,从而帮助用户更快地找到感兴趣的商品,或者是找到最适合自己的尺码^[2]以提升用户体验。此外,商品属性分类还能生成大量的结构化数据,有

助于数据分析和挖掘,发现商品属性的相关性,为商家决策提供有价值的洞察。本文拟使用服装评论文本作为输入,借助预训练大语言模型对服装种类、用户身材、评价、用途以及服装对用户的合身状况进行分类。示例如图 2 所示。

目前而言,基于自然语言处理的商品属性分类方法大多针对一个属性对文本内容进行多分类(包括二分类)或多标签分类。例如,通过一段商品评价文本对商品的种类进行分类或是对用户情感进行判断^[3]。然而,在实际应用过程中,一段文字往往可以从多个维度进行解读,即可以针对不同属性对文本进行分类。这样的分类方式可以帮助人们更全面地了解一段文本,对人们的日常生产生活也有着更多的指导意义。举例而言,在对一段关于购买服装的评论文本进行分类的任务中,若只对购买者的评价高低进行分类,则无法得知消费者评价对象的具体种类,以及与消费者相关的一些信息。然而,若能进一步提取消费者以及评价对象的相关信息,我们便可

基金项目:国家自然科学基金(62076175,61976146);江苏省双创博士计划

This work was supported by the National Natural Science Foundation of China(62076175,61976146) and Jiangsu Innovation Doctor Plan.

通信作者:赵哲宇(yzha00104@hotmail.com)

以更容易地针对具体商品进行适当的调整,或排除一些由于顾客自身因素而造成的负面评价。但是,商品多属性分类任务存在属性间关系复杂、特征不易提取、难以平衡模型的复杂性与泛化能力的困难。因此,如何提高模型在商品多属性分类任务中的表现,成为我们研究的重点。

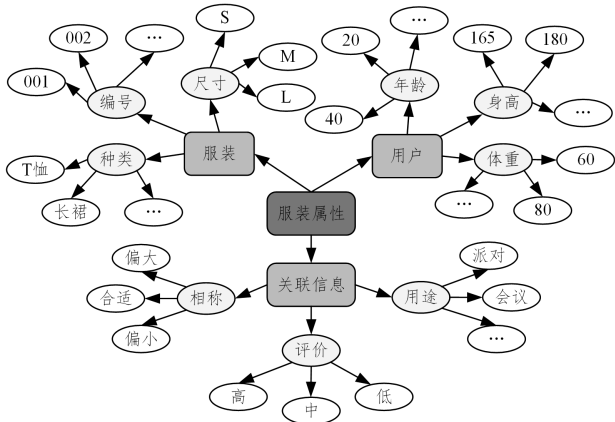


图1 服装商品属性示例

Fig. 1 Example of product attribute

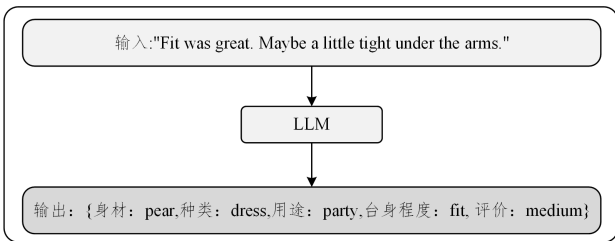


图2 任务示例

Fig. 2 Task example

目前,大语言模型凭借其在信息检索和文本理解上的优异性能,在各种自然语言处理任务中被广泛运用,其也成为我们研究商品属性分类问题的主要工具。大语言模型基于大规模无标签文本进行预训练,其中包含了来自各种领域的丰富语言知识。然而,微调任务所涉及的特定领域信息仅占预训练数据的一小部分,这导致了模型可能对特定领域的语境和特征表示学习不足,影响其在该领域内的性能。针对该问题,本文提出了领域内文本预训练方法,以扩展模型在特定领域的知识范围,让模型更好地理解特定领域内的信息及数据分布,从而提高模型在目标领域内的性能和泛化能力。这一方法有助于减轻领域适应不足的问题,提高模型在特定任务上的表现。

此外,当前大量研究都旨在直接借助模型从输入文本中提出输出信息,而对输出结果之间的关联性关注较少。这种纵向预测的方法虽然有效,但在处理复杂的多属性分类任务时可能无法充分挖掘属性之间的潜在关系。以服装属性分类任务为例,服装的种类往往与其使用场合有着紧密的关联性,如礼服往往与派对相搭配。为了进一步发掘属性间的关联,本文提出基于属性间关联性的预训练,即屏蔽输出属性中的部分内容,让输出属性的其他部分推测该屏蔽部分的内容。通过这种方式,模型被迫考虑输出中不同属性之间的联系和逻辑,从而更好地理解属性之间的关联性。采用该方法对模型进行预训练可以让模型在进行微调任务之前事先了解输出属性中数据之间的关联性,提高

模型在多属性分类任务中的性能。

基于以上分析,本文提出了双重预训练的商品属性分类方法。本文的主要贡献如下:

- 1) 利用领域内的文本对模型进行进一步预训练,以提高对商品评论文本的理解能力;
- 2) 利用基于属性间关联性的预训练方法使模型发掘输出属性间的关联性,以提高模型在多属性分类任务中的准确性;
- 3) 通过在预训练大语言模型上有效结合领域内文本预训练和基于属性间关联性的预训练,使得模型对商品属性分类任务的理解能力显著提高,取得了良好的分类效果。

2 相关工作

早期对商品评论分类的研究多采用传统机器学习进行。例如,Zhang 等^[4]使用 SVM^[5]算法对商品进行自动化评分;Luo 等^[6]基于隐朴素贝叶斯^[7]对商品评论进行情感分类;Fayaz 等^[8]结合了多层感知器(MLP)^[9]、k 近邻(KNN)^[10]和随机森林(RF)^[11]的预测结果,并根据模型的贡献程度投票判断评论的有用性。但是,传统机器学习方法需要手动设计特征,无法有效捕捉复杂的数据表示,且在处理非结构化数据和大规模数据时的表现相对较弱。

随后,越来越多深度学习算法,特别是循环神经网络相关算法的提出,凭借其循环结构缓解了机器学习算法的这些痛点。Peng^[12]基于 RNN^[13]和 LDA 模型^[14]对商品评论情感分类进行研究;Chen 等^[15]使用多层 LSTM^[16]网络来生成语句和文本的表示,并将该表示通过注意力机制与用户信息、商品信息相融合,最后实现对用户情感的分类任务。相比传统机器学习算法,深度学习展现出了更出色的特征提取能力,对复杂内容的理解能力也更加出色。但是,深度学习模型的训练依赖大量标注数据,且存在计算成本较高的问题。

近年来,随着人工智能领域的发展,利用预训练大语言模型(如 BERT^[17])进行文本属性抽取的方法取得了显著进展。Xu^[18]通过微调 BERT 和 TextCNN^[19]相结合的模型对商品的类别进行分类,证明了其提出的 BERT-TextCNN 在商品分类任务中的强泛化能力。Zhang 等^[20]利用 ELECTRA^[21]等预训练语言模型对商品属性进行抽取,其实验结果表明,预训练语言模型较传统基线模型在领域内和跨领域数据的实验中都表现出了更好的性能。Eshel 等^[22]提出了 PreSzeE 模型,并基于诸如品牌、类别等属性以及买家购买历史中对尺码的偏好进行建模,实现了对用户购买服装的尺寸的精准预测。Chatterjee 等^[23]以 Bert 模型为基础,将评论文本、用户信息、商品信息的编码相融合,对服装和顾客身材的相称度进行了准确的预测。相较于一般的深度学习模型,大语言模型由于经过大量数据的预训练,有着更强大的语言理解和生成能力,同时仅需要经过少量数据的微调就可以获取对特定任务的适配能力。

最新研究表明,对所使用的大语言模型进行进一步的预训练有助于更好地进行下游微调任务。Sun 等^[24]研究如何在文本分类问题中更好地微调 BERT 模型时指出,基于任务和领域内信息对模型进行进一步微调能够显著提升模型的表现。Bao 等^[25]在属性级情感分析任务中通过使用元素级图预训练范式和任务级图预训练范式两种预训练范式,提高生成模型的结构意识。

目前,较少有研究在商品属性分类任务中采用进一步预训练的方法,且已有的预训练方法主要针对输入内容进行预训练而缺少对输出内容之间潜在关联性的关注。为此,本文提出了一种基于双重预训练的文本多属性分类方法,以提高多属性分类任务的准确度。模型选用 T5^[26] 模型,拟对输入文本以及输出文本进行针对性的预训练。在实验中,本文选择 Clothing Fit Data^[27] 作为数据集,采用双重预训练加微调的方式,基于购买者的评价文本对购买者评价、购买对象种类等属性进行预测。

3 基于多重预训练的商品属性分类任务

一个句子包含 n 个单词 w_i , 用集合 S 表示为 $S = \{w_1, w_2, \dots, w_n\}$, 同时,有一个包含 k 个属性的属性集合 $C = \{c_1,$

$c_2, \dots, c_k\}$, 假设第 i 个属性中有 j 个类别, 即对应的属性 c_i 有类别集合 $A_i = \{a_{i1}, a_{i2}, \dots, a_{ij}\}$ 。针对上述的数据形式, 属性分类任务定义如下: 给定一个句子 S , 针对属性集合 C 中的每一个属性 c_i 对该句子进行分类。

本研究的任务架构如图 3 所示。任务基于 T5 模型进行实验, 针对输入以及输出设计了两种预训练任务(领域内文本预训练和基于属性间关联性的预训练)。在预训练的过程中, 本文将根据实验结果选择两种预训练任务的最佳先后顺序, 亦或是将两种预训练任务的数据集相混合, 在一次训练中同时完成两种任务。在完成预训练后, 用所选数据集对模型进行进一步的微调。由于 T5 型是文本-文本的模型, 因此, 需要基于正则表达式进行文本模式匹配和提取, 最终得到我们所需要的属性以及相应的属性值。

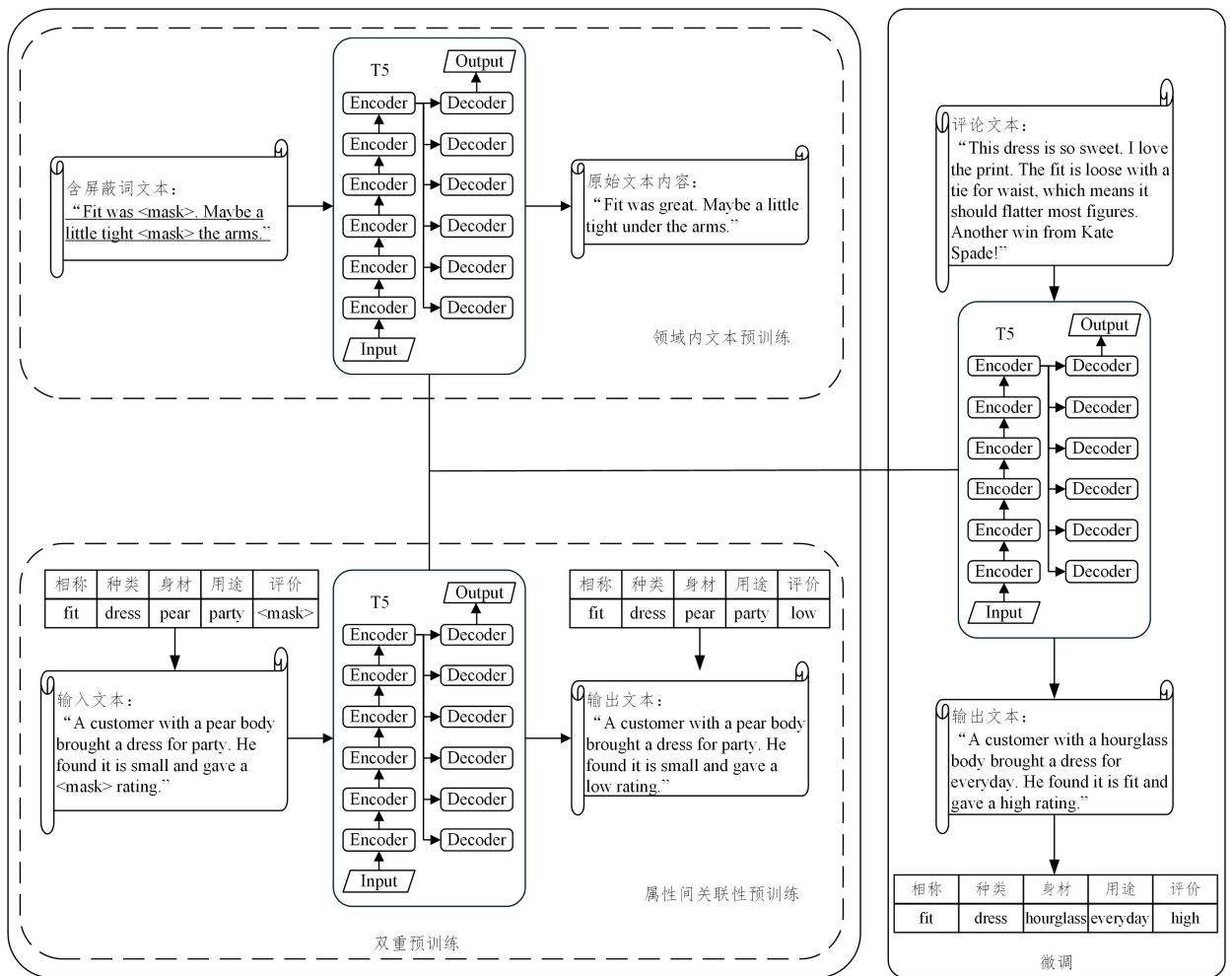


图 3 任务架构图

Fig. 3 Task architecture

3.1 模型

本文采用 T5 作为基础模型开展实验。相比于传统的语言模型, T5 采用了一种统一的框架, 将所有的自然语言处理任务都转化为文本到文本的转换任务, 因此被称为 "Text-To-Text"。这种一致性的框架简化了模型的设计和训练, 同时提高了模型的通用性和泛化能力。模型输入表示包括标记化、词嵌入和位置编码。解码器根据任务描述和输入文本编码预测输出序列。

3.2 基于双重预训练的商品属性分类方法

与 BERT 类似, T5 模型在预训练过程中也采用了

遮蔽语言建模 (Masked Language Modeling, MLM) 方法。在 MLM 任务中, 模型会随机地将输入文本中的一部分标记进行遮蔽, 并试图预测这些被遮蔽的标记。考虑到这种预训练方法基于大量不同领域的的数据, 可能不足以充分解决我们特定分类任务所关注的领域问题, 因此本文设计了特定领域的预训练实验, 以增强模型对目标领域文本的理解。

此外, 为了让模型探索输出结果中各个属性之间的相关性, 本文引入了基于属性间关联性的预训练。在完成这两种预训练任务后, 模型将进行目标任务的微调。两个预处理任

务对数据的处理方式如图 4 所示。

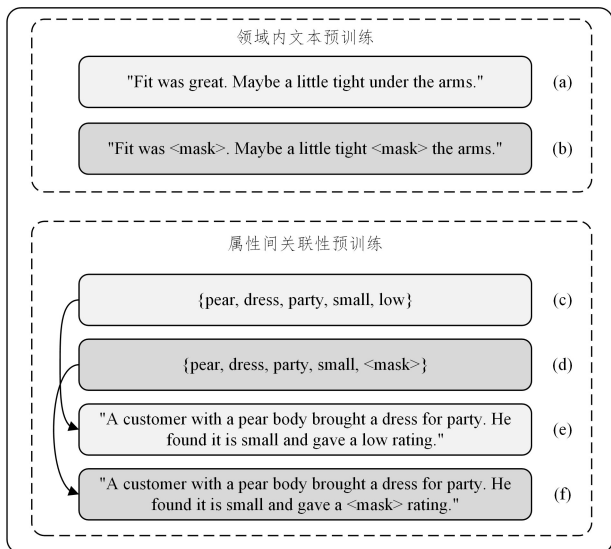


图 4 预训练任务数据处理示例

Fig. 4 Examples of data processing for pre-training task

领域内文本预训练:本模块旨在促使模型更好地理解我们需要对其进行分类的文本所在的领域。该预训练采用了与 MLM 非常相似的方法,即选取一定数量与我们关注的领域相关的文本,然后在其中随机选择输入文本的一部分标记,并用特殊标记进行掩盖。通常,这个掩盖的标记为“<mask>”。模型的目标是根据上下文预测被标记掩盖的文本。例如,原文图 4(a)的内容在掩盖部分文本后,变成图 4(b)中的内容。我们期望预训练后的模型能够将这两个<mask>标记分别预测为“great”和“under”。

基于属性间关联性的预训练:该模块旨在帮助模型理解本研究所需要分类的属性间的关系。相较于基于领域内文本的预训练,该预训练任务对具体多属性分类任务的针对性更强。在设计该预训练任务时,本文借鉴了 MLM 的预训练形式,但相比于对随机文本进行标记和掩盖,该任务只对部分属性值进行掩盖。

例如,分类属性值如图 4(c)所示。它们分别属于体型(body type)、服装类型(category)、购买目的(rented for)、合身程度(fit)和评价(rating)属性。以掩盖评价属性为例,其属性值如图 4(d)所示。

由于 T5 模型被设计用于处理完整的自然语言文本,因此提供完整句子可以更好地满足其预期的输入和输出格式,从而有助于模型更好地理解语境。因此,本研究采取将属性值转换为结构化文本的方式展开研究。基于上述属性值的示例,其结构化文本如图 4(e)所示,经过掩盖后的文本如图 4(f)所示。

其中 low 是属性 rating 的分类结果,我们将其进行掩盖,并期望其他 4 个属性的值能够对该值进行准确预测。经过大量此类预训练,模型将能够通过实现属性间预测的方式发掘不同属性之间的关系。

微调:微调的目的是通过使用预训练模型,并根据特定任务的数据对其参数进行调整,以适应特定任务并提高模型在该任务上的性能。由于 T5 模型是文本到文本的模型,在其预训练的过程中问题和回答都采用了自然语言的形式,因此本研究在输出时将属性值以自然语言的形式呈现,并用正则

表达式的方法对属性值进行提取比较。该任务旨在通过输入文本得到多属性分类结果。

4 实验

4.1 实验数据和预处理

本文采用 Clothing Fit Data^[27]数据集,该数据集总共包含 105508 条数据,每条数据包含了用户评价文本、合身、用户评分等 15 个属性。为便于实验,本文选择评价文本作为模型的输入,并选取了典型的 5 个属性作为分类目标,分别是:

是否合身(fit);评价(rating);体型(body type);购买目的(rent for);服装类型(category)。其中,服装类型属于服装本身的属性;体型属于用于的属性;而是否合身、评价以及购买目的属于用户和服装之间的关联属性。

原始数据中各个类别的数据数量分布较为不均衡,这可能导致模型对少数类别的预测性能下降。因此,本实验对数据进行均衡处理,包括:去除所含数据量极少的类别;为实验所用数据集每一个类别的数据量设置上限。特别地,在 rating 属性中,大多数评价为 10 分和 8 分,2~6 分的评分较少。为此,我们对数据进行改造,将 10 分评价设置为好评(high),将 8 分设置为中评(media),将 2~6 分的数据进行合并,设置为差评(low),以尽可能地实现数据的均衡分布。基于上述数据筛选,本文采用了原始数据集中的 10000 条数据进行实验,所用数据的情况如表 1 所列。

表 1 实验数据分布情况

Table 1 Distribution of experimental data

Attribution	Category_num	Max_num	Min_num	Avg_num	std
Body type	6	1869	1285	1666	277
fit	3	7308	1184	3333	3445
Rent for	5	2999	1162	2000	717
rating	3	6128	958	3333	2610
Category	16	856	304	625	163

其中,fit 属性和 rating 属性由于整体数据较为集中(fit 属性值大多为 fit,rating 属性值大多为 high),因此在尽可能平衡数据分布后标准差仍然较大。

训练集、验证集和测试集的划分如表 2 所列。

表 2 训练集、验证集、测试集分布

Table 2 Distribution of training set, validation set and test set

train	validation	test
8000	1000	1000

4.2 实验参数设置

本研究所使用的实验参数如表 3 所列。

表 3 实验参数

Table 3 Experimental parameters

parameter	value
Transformer blocks	12
Hidden size	512
Self-attention heads	8
Total parameters	1.1×10^8
Learning rate	1×10^{-4}
Max length(input)	256
Max length(output)	256
Batch size	16

此外,实验优化器采用 Adam 优化器,使用 Cross-Entropy Loss 作为损失函数。

4.3 评估准则

在多属性分类任务中,采用 Macro-F1 作为评价指标。Macro-F1 对每个类别的性能平等地进行评估,因此对于类别不平衡的情况较为合适。Macro-F1 的计算如式(1)所示:

$$Macro-F1 = \frac{1}{N} \sum_{i=1}^N F1_i \quad (1)$$

其中, N 表示类别数量, $F1_i$ 表示第 i 个类别的 F1 分数。F1 的计算如式(2)所示:

$$F1 = 2 * \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

在预训练任务中,需要比较生成的句子与目标句子的相似度,故采用 BLEU^[28] 作为评价指标。其计算如式(3)所示:

$$bleu_n = \frac{\sum_{c \in candidates} \sum_{n\text{-gram} \in c} Count_{dip}(n\text{-gram})}{\sum_{c' \in references} \sum_{n\text{-gram}' \in c'} Count_{dip}(n\text{-gram}')} \quad (3)$$

其中,模型生成的句子是 candidates,目标文本是 references。 n -gram 表示文本中连续 n 个单词的组合。 c 和 c' 分别表示生成内容和目标文本中的句子。 $Count_{dip}(n\text{-gram})$ 表示某一个 n -gram 在 reference 中的个数。 $Count(n\text{-gram}')$ 表示 $n\text{-gram}'$ 在 candidate 中的个数。整个分式的含义为同时出现在 candidates 和 references 中的 n -gram 个数和 candidates 中所有 n -gram 个数的比值。在本文的预训练任务中,评价指标设置为 $bleu_{1-4}$ 的平均值。

4.4 基准模型

BERT: BERT^[18] (Bidirectional Encoder Representations from Transformers) 是一种由 Google 研发的自然语言处理模型,它采用了 Transformer 架构并进行了改进。BERT 模型的主要特点是双向编码器,它能够同时考虑句子中的上下文信息,而不是像传统的语言模型那样只能从左到右或者从右到左单向编码。这使得 BERT 在理解句子语境、处理语言任务时具有更好的性能。

XLNet: XLNet^[29] 是由 Google 研究团队提出的一种基于 Transformer 架构的预训练语言模型。与传统的语言模型不同, XLNet 采用了一种全新的自回归性质和一个额外的排列语言建模 (Permutation Language Modeling, PLM) 任务来预训练模型。这种设计使得 XLNet 在理解文本语境和处理自然语言处理任务时具有更好的效果,并在多个基准数据集上取得了 state-of-the-art 的性能。

ChatGPT3.5: ChatGPT-3.5^[30] 是基于 OpenAI 的 GPT-3.5 架构改进的聊天型人工智能模型,相比之前的版本,它在语言理解和生成方面更为出色,能够更准确地理解上下文,并生成更具逻辑性和连贯性的文本。在本文的实验中,首先向模型输入实验需求,包括实验目的、分类属性和属性值集。接着,通过提供示例来帮助模型更好地理解实验任务和输出格

式。最后,通过输入评论文本来获取输出的属性及其对应的属性值。

Llama 2: Llama 2^[31] 是一款基于 Transformer 架构的自回归因果语言模型,属于先进的自然语言处理技术。该模型通过自监督式预训练,从大规模未标注数据集中学习,其训练涉及递归地预测语句中的下一个单词,从而在不断减少预测误差的过程中掌握语言和逻辑模式。研究指出, Llama 2 利用了来自公开来源的超过 2 万亿个令牌进行训练,这些令牌代表了语言中的各种语义单位,如单词、词组等。此外, Llama 2 的基础模型主要用于开发特定应用场景下的模型,而非直接回答用户提示。

此外,本文也选取微调 BERT 和 TextCNN^[20] 相结合的模型进行对比实验。

4.5 实验结果与分析

为了验证我们选用的基础模型及双重预训练方法较现有方法的有效性和优越性,主要进行了以下几个部分的实验。1) 对比使用 T5 模型和现有经典基准模型在本任务中的表现; 2) 比较了使用 T5 模型对多属性进行联合输出与单独输出的结果; 3) 对比加入领域内预训练任务后模型的表现; 4) 对比加入基于属性间关联性的预训练后模型的表现; 5) 结合两种预训练任务后模型的性能对比。

4.5.1 与基准模型的对比分析

与基准模型的对比结果如表 4 所列。从表中可以得出以下结论: 对比 4 个基准模型 (ChatGPT, Llama 2, BERT 和 XLNet), Llama 2, BERT 和 XLNet 的结果比较接近, 体现出它们在该任务中的性能接近。特别地, BERT 和 XLNet 在架构上类似, 只是在与训练任务中分别采用了单项掩码预训练和双向上下文预训练的方式。相对来说, XLNet 在长文本中相对于 BERT 有更强的上下文建模和语义理解能力, 而本研究所使用的输入文本长度较短, 因此这个优势在该任务中较难体现。

Llama 2, BERT 和 XLNet 都是双向模型, 能够同时考虑文本的上下文信息, 因此在一些需要全局上下文信息的任务上表现更好。然而, ChatGPT 是一个单向模型, 在生成文本时只能考虑前面的文本信息, 可能无法很好地利用后文的信息, 因此在本任务中效果较差。

本文采用的基于双重预训练的商品属性分类方法较诸多基础模型以及与该任务相关的方法取得了较高的性能提升, 表明了该方法的有效性。

4.5.2 消融实验

为了进一步探究采用双重预训练方法对实验结果带来的提升, 本文设计了一系列消融实验, 如表 5 所列。

表 4 实验结果

Table 4 Experiment results

Model	Fit	Rating	Rented for	Body type	Category	Avg
ChatGPT	43.0	27.7	30.1	18.0	35.3	30.8
Llama 2	50.2	64.0	55.0	20.0	34.7	44.8
BERT	66.3	50.3	50.5	19.8	40.9	45.6
XLNet	66.9	56.9	47.2	11.4	37.8	44.0
BERT+TextCNN	65.5	52.2	48.5	23.7	43.1	46.6
双重预训练	73.7	67.2	60.2	24.7	53.9	56.0

1) T5 进行多属性分类比分别进行单属性分类的结果

更好, 表明联合输出有助于提高所有属性的预测结果一致性,

提高了模型的稳定性和可靠性,同时也有助于理解不同属性之间的关系和影响,提供更深入的洞察。此外,联合输出可以

通过一次模型推断同时预测多个属性,相比多次单一输出推断,具有更高的效率。

表 5 消融实验

Table 5 Ablation experiment

Model	Fit	Rating	Rented for	Body type	Category	Avg
T5 单属性	69.3	62.8	53.5	16.5	43.5	49.1
T5 多属性	71.9	65.3	56.8	24.1	52.6	54.2
领域内文本预训练	72.8	65.5	58.8	21.8	53.0	54.4
属性关联性预训练	72.3	66.6	58.8	23.6	53.7	55.0
双重预训练	73.7	67.2	60.2	24.7	53.9	56.0

2)单独采用两种预训练方式对模型进行预训练,效果较直接采用基础模型有一定的提升,证明了本文采用的预训练方式的有效性。但单任务预训练带来的性能提升并不明显,可能是因为单个任务的预训练缺乏足够的复杂度和多样性,无法充分激发模型学习更深层次的语言特征和任务相关的知识。

3)相比于单任务进行预训练,采取双重预训练的方式增加了模型的预训练的复杂性和多样性,且对特定的任务具有更强的针对性,有效地提升了模型的整体性能。

4.5.3 领域内预训练的结果分析

为了验证使用领域内的数据对模型在任务中性能的影响,本文参考 T5 在其原本预训练任务上的预训练方式,即随机选取 15% 的文本进行掩盖,让模型预测被掩盖文本的内容。由于实验数据的差异,本文在原本覆盖 15% 文本的基础上又分别进行了覆盖 10% 和 20% 文本的预训练,以获取最佳结果。实验结果如表 6 所列。

表 6 领域内预训练实验结果的对比

Table 6 Comparison of in-domain pre-training experimental results

Method	Avg_score
10% 覆盖	54.3
15% 覆盖	54.4
20% 覆盖	54.1
T5 无预训练	54.2

该结果表明,覆盖 15% 的文本进行领域内数据预训练取得的结果是最好的,这与 T5 原始预训练时设置的参数一致。此外,预训练后的模型较没有通过预训练的模型在平均指标上取得了一定的提升。具体属性中,除了 Body type 这一属性出现指标上的下降,其他属性的分类效果都有所提升,一定程度上说明了预训练任务的有效性。

4.5.4 基于属性间关联性预训练的结果分析

实验中需要分类的属性总共有 5 条,为了能够让剩余的属性尽可能准确地预测被掩盖的属性,我们在实验中设定了覆盖属性的数量不超过 2 条。即可能的覆盖方式有:覆盖 1 条属性;覆盖 2 条属性;覆盖 1 条属性和 2 条属性相结合。文本对上述 3 组覆盖方式分别进行了实验,结果如表 7 所列。

表 7 基于属性间关联性预训练实验结果对比

Table 7 Comparison of pre-training experimental results based on correlation between attributes

Method	Avg_score
覆盖 1 条属性	55.0
覆盖 2 条属性	54.6
覆盖 1,2 条属性	54.3
T5 无预训练	54.2

结果表明,在采用覆盖 1 条属性进行预训练后,该任务取得了最佳的效果,平均 F1 值为 0.550,较原始模型性能有所提高。该任务表明,模型在通过其他属性预测被覆盖属性的过程中更多地发掘了属性间的相对关系,提高了模型在微调过程中通过文本预测多属性类别时的性能。

4.5.5 基于双重预训练的结果分析

上述两种预训练方法证明了,更具针对性的预训练任务有助于加强模型对特定数据集的理解,并提高模型在目标任务中的性能。该实验尝试将两种预训练任务相结合的方式。具体实现中,分为先进行领域内数据预训练再进行基于属性间关联性的预训练,以及先进行基于属性间关联性的预训练再进行领域内数据的预训练。由于两种预训练任务的评价方式相同,因此亦可以采用将两种预训练数据结合共同训练的方式。该实验的结果如表 8 所列。

表 8 基于双重预训练的实验结果对比

Table 8 Comparison of experimental results based on dual pre-training

Method	Avg_score
task1 - > task2	55.2
task2 - > task1	56.0
task1 + task2	54.2
T5 无预训练	54.2

其中,task1 为领域内预训练任务,task2 为基于属性间关联性的预训练任务;->表示实验进行流程,+表示将两种方法所采用的数据集相混合。结果表明,采用先基于属性间关联性的预训练方法进行预训练再进行领域内数据的预训练取得了最好的效果。由表 1 可知,采用该种预训练方法所得到的测试结果无论在平均指标还是在单个属性的指标上都较未进行进一步预训练的 T5 模型以及只进行了一种预训练的模型效果更好,证明了在该任务中,双重预训练模型在多属性分类中有着更好的性能表现。

在该实验中,采用基于属性间关联性的预训练(task2)后接领域内数据预训练(task1)的方法(task2->task1)取得了最佳效果,优于先进行领域内预训练再进行属性关联性预训练(task1->task2)和两种预训练混合(task1+task2)的方法。猜想这一结果的原因可能在于,基于属性间关联性的预训练首先强化了模型对属性间复杂关系的捕捉能力,使得模型在随后的领域内预训练中能更有效地理解和整合领域特定的文本信息。这种顺序优势表明,模型在获取了对属性关联的深刻见解后,更能够利用这些关系来优化对领域内数据的学习,从而在整体分类任务上实现更优的性能表现。

4.5.6 样例分析

本节在预测结果中选取了两个典型实例并进行分析,如

表 9 所列。其中,真实值是数据集中商品的实际属性,基准模型是指未经过双重预训练的基本 T5 模型。

表 9 实例分析
Table 9 Cases study

评论	属性
I passed over this dress a few times when browsing because I wasn't sure if it would work with my body shape. I normally wear a 10 or 12 and in this dress ended up wearing the 12. I wore it to a family wedding and got tons of compliments. It was a little snug around the bust, but it actually ended up being good thing because I didn't wear a bra. The pockets were great! I loved it so much I didn't want to send it back. Please speculate on the details of the purchase.	真实值: fit; high; wedding; full bust; dress 基准模型: fit; high; wedding; full bust; skirt(×) 双重预训练模型: fit; high; wedding; full bust; dress
I hated sending this skirt back! It was dressy enough for a rehearsal dinner(as a guest, not the bride). It was perfect with a statement necklace, well-worn denim shirt, and strappy heels. It has three layers: a top mesh layer, lace layer, and under slip. I was worried it would get snagged on something, but I seemed to avoid doing that. (Just be careful when sitting that your heel doesn't get caught!). I'm 5'4", and this skirt was super long. It came to somewhere between ankle and mid-calf. I thought it would be a little overwhelming, but it was perfect. Really wish I could have kept it, and would rent it again in a heartbeat! Please speculate on the details of the purchase.	真实值: fit; high; party; pear; skirt 基准模型: fit; high; wedding(×); straight & narrow(×); skirt 双重预训练模型: fit; high; party; pear; skirt

从实验结果可以看出,经过双重预训练的模型较基准模型有着更好的分类效果。具体来看,在第二条实例中,用户在评论中提到了自己的身高时,说明了自己的身高(5英尺4英寸=162.56cm),并认为裙子较长,这说明这个顾客应该不属于细长型的身材。以上信息属于服装的领域信息,在基础的预训练任务中可能较难获取相关知识,故认为预测准确度的提升来源于领域内知识预训练。

在第一条实例中,基准模型在分类服装品类时出现错误,认为模型在同是裙类的长裙和短裙中在实际情况进行了错误的判断。在一般情况下,婚礼场景往往与长裙搭配更加合适,这体现了让模型了解属性间关联性的重要性。类似地,第二条实例中服装种类为短裙,而基准模型认为购买目的是婚礼,而非实际的派对。在一般情况下,派对与短裙往往是更常见的搭配,基准模型的错误判断同样体现出了对属性间关联性认知的缺乏。

基于上述实例,本文提出的基于双重预训练的商品属性分类方法能够在领域知识和属性间关联性知识的角度提高模型对任务的理解力,从而达到更好的分类效果。

结束语 本文提出了一种基于双重预训练的商品属性分类方法,通过结合领域内文本预训练和基于属性间关联性的预训练来增强大型语言模型 T5 的分类能力。通过这种创新的双重预训练策略,模型能够更准确地理解和处理商品描述文本,从而有效提高在复杂多属性分类任务中的性能。实验结果表明,相较于未经过预训练的模型以及其他基准模型,采用本文方法的 T5 模型在 Clothing Fit Data 数据集上的分类效果有显著提升。此外,本研究所采用的预训练方法不仅提高了模型在单一属性上的分类准确性,还促进了模型对不同属性间关系的理解,有助于揭示商品属性之间的内在联系。这种方法的成功应用展示了双重预训练策略在提高特定领域分类任务中的潜力,为未来相关领域的研究和应用提供了宝贵的参考。

参考文献

[1] WU Z, DAI X Y, YINC, et al. Improving review representations with user attention and product attention for sentiment classification[C]// Proceedings of the AAAI Conference on Artificial

Intelligence, 2018.
[2] ABDULLA G M, BORAR S. Size recommendation system for fashion e-commerce[C]// KDD Workshop on Machine Learning Meets Fashion, 2017: 1-7.
[3] BHATT A, PATEL A, CHHEDA H, et al. Amazon review classification and sentiment analysis[J]. International Journal of Computer Science and Information Technologies, 2015, 6(6): 5107-5110.
[4] CHENG Z J, ZHU Y H. SVM-based commodity scoring system[J]. Computer Knowledge and Technology, 2018, 14(30): 223-225.
[5] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18-28.
[6] LUO H Q, LU X Y, ZHANG X B et al. Hidden parsimonious Bayes-based sentiment classification method for commodity reviews[J]. Computer Engineering and Design, 2017, 38(1): 203-208.
[7] JIANG L, ZHANG H, CAI Z. A novel bayes model: Hidden naive bayes[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 21(10): 1361-1371.
[8] FAYAZ M, KHAN A, RAHMAN J U, et al. Ensemble machine learning model for classification of spam product reviews[J]. Complexity, 2020, 2020: 1-10.
[9] TAUD H, MAS J F. Multilayer perceptron(MLP)[M]// Geomatic Approaches for Modeling LAND Change Scenarios. Springer, 2018: 451-455.
[10] KRISHNA K, MURTY M N. Genetic K-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29(3): 433-439.
[11] RIGATTIS J. Random forest [J]. Journal of Insurance Medicine, 2017, 47(1): 31-39.
[12] PENG S C. Research on Sentiment Classification of Commodity Reviews Based on RNN and LDA Models [D]. Hangzhou: Zhejiang University of Technology, 2018.
[13] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[J]. arXiv: 1409. 2329, 2014.
[14] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(Jan): 993-

1022.

- [15] CHEN H, SUN M, TUC, et al. Neural sentiment classification with user and product attention[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:1650-1659.
- [16] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [J]. Advances in Neural Information Processing Systems, 2015, 28.
- [17] KENTON J D M W C, TOUTANOVA L K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]//Proceedings of NAACL-HLT. 2019:4171-4186.
- [18] XU M M. Research on commodity classification method based on BERT [D]. Nanchang: Nanchang University, 2023.
- [19] YOON K. Convolutional Neural Networks for Sentence Classifications[C]//Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistic, 2014:1746-1751
- [20] ZHANG S Q, MA J, ZHOU X B, et al. Commodity attribute extraction based on pre-trained language models[J]. Journal of Chinese Information, 2022, 36(1):56-64.
- [21] CLARK K, LUONG M T, LEQ V, et al. Electra: Pre-training text encoders as discriminators rather than generators[J]. arXiv:2003.10555, 2020.
- [22] ESHEL Y, LEVI O, ROITMANH, et al. Presize: predicting size in e-commerce using transformers[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021:255-264.
- [23] CHATTERJEE O, TEJ J R, DASARAJUN V. Incorporating customer reviews in size and fit recommendation systems for fashion e-commerce[J]. arXiv:2208.06261, 2022.
- [24] SUN C, QIU X, XU Y, et al. How to fine-tune bert for text classification? [C]//Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18-20, 2019, Proceedings 18. Springer International Publishing, 2019:194-206.
- [25] BAO X, WANG Z, ZHOU G. Exploring graph pre-training for aspect-based sentiment analysis[C]// Findings of the Association for Computational Linguistics: EMNLP 2023. 2023:3623-3634.
- [26] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. The Journal of Machine Learning Research, 2020, 21(1):5485-5551.
- [27] MISRA R, WAN M, MCAULEY J. Decomposing fit semantics for product size recommendation in metric spaces[C]//Proceedings of the 12th ACM Conference on Recommender Systems. 2018:422-426.
- [28] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002:311-318.
- [29] YANG Z, DAI Z, YANG Y, et al. Xlnet: Generalized autoregressive pretraining for language understanding [J]. Advances in Neural Information Processing Systems, 2019, 32.
- [30] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback [J]. Advances in Neural Information Processing Systems, 2022, 35:27730-27744.
- [31] TOUVRON H, MARTIN L, STONE K, et al. Llama 2: Open Foundation and Fine-Tuned Chat Models [J]. arXiv: 2307.09288 2023.



ZHAO Zheyu, born in 2001, postgraduate. His main research interests include natural language processing and text classification.



WANG Hongling, born in 1975, Ph.D., associate professor, is a member of CCF (No. 14272M). Her main research interests include natural language processing and information retrieval.