

数据分析与知识发现  
*Data Analysis and Knowledge Discovery*  
ISSN 2096-3467, CN 10-1478/G2

## 《数据分析与知识发现》网络首发论文

题目：基于多维度图卷积网络的旅游评论有用性识别  
作者：刘洋, 丁星辰, 马莉莉, 王淳洋, 朱立芳  
网络首发日期：2022-10-26  
引用格式：刘洋, 丁星辰, 马莉莉, 王淳洋, 朱立芳. 基于多维度图卷积网络的旅游评论有用性识别[J/OL]. 数据分析与知识发现.  
<https://kns.cnki.net/kcms/detail/10.1478.g2.20221025.1836.004.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

## 基于多维度图卷积网络的旅游评论有用性识别

刘洋<sup>1,2</sup>, 丁星辰<sup>3</sup>, 马莉莉<sup>4</sup>, 王淳洋<sup>1</sup>, 朱立芳<sup>5</sup>,

<sup>1</sup> (武汉大学信息管理学院 武汉 430072)

<sup>2</sup> (武汉大学大数据研究院 武汉 430072)

<sup>3</sup> (武汉大学国家网络安全学院 武汉 430072)

<sup>4</sup> (武汉大学经济与管理学院 武汉 430072)

<sup>5</sup> (广东财经大学人文与传播学院, 广州, 510320)

**【目的】**在线评论有用性有助于消费者快速获取有效的信息,本研究利用深度学习模型对旅游评论的有用性进行识别,以更好地让消费者和酒店管理者受益。

**【方法】**本文提出多维度图卷积网络和多模态融合的有用性识别方法,使用BERT和MAE分别对文本和图片进行预训练,利用多维度图卷积网络对多模态特征进行建模,再通过注意力机制捕捉多模态间的交互信息,最后融入文本特征进行有用性识别。

**【结果】**在Yelp数据集上进行对比实验,结果表明该方法检测的准确率为73.21%,相较于传统单模态和现有多模态模型,提升5%~15%。

**【局限】**目前仅在Yelp数据上尝试文本和图片两种模态,在其他数据融合以及更多模态后续有待研究。

**【结论】**本文将多维度的图卷积网络和多模态特征融入有用性识别中,有效提升识别的效果。

**关键词:** 多模态特征; 多维度; 图卷积网络; 旅游评论; 有用性识别

## The helpfulness detection of travel reviews based on multi-dimensional graph convolutional network

Liu Yang<sup>1,2</sup>, Ding Xingchen<sup>3</sup>, Ma Lili<sup>4</sup>, Wang Chunyang<sup>1</sup> and Zhu Lifang<sup>5</sup>

<sup>1</sup> (School of Information Management, Wuhan University Wuhan 430072)

<sup>2</sup> (Big Data Research Institute, Wuhan University, Wuhan 430072)

<sup>3</sup> (School of Cyber Science and Engineering, Wuhan University Wuhan 430072)

<sup>4</sup> (Economics and Management School, Wuhan University Wuhan 430072)

<sup>5</sup> (School of Humanities and Communication, Guangdong University of Finance and Economics, Guangzhou, 510320)

### Abstract:

**[Objective]** The helpfulness of online reviews helps consumers, who quickly obtain effective information. This paper utilizes a deep learning model to recognize the helpfulness of travel reviews, which can better benefit consumers and hotel managers.

**[Method]** This paper proposes a helpfulness recognition method based on multi-

<sup>1</sup> 本文系国家自然科学基金青年项目“突发公共卫生事件公众心理应激信息表征及干预机制研究”(72204190)研究成果、教育部人文社科项目青年项目“基于社交机器人的突发公共卫生事件公众心理应激干预研究”(22YJZH114)之一。

dimensional graph convolutional network and multi-modal fusion, this method uses BERT and MAE to pre-train text and pictures, respectively, then adopts a multi-view graph neural network to model multi-modal features, which captures the interaction information between multimodalities through the attention mechanism, and finally integrates text features for helpfulness detection.

**[Results]** Comparative experiments were carried out on the Yelp data set. The experimental results showed that the accuracy of this method reached 73.21%, which was improved by 5%-15% compared with the traditional single-modal and existing multi-modal models.

**[Limitations]** Only two modalities of text and images have been tried on Yelp data. More modalities on other data fusion will research in the future.

**[Conclusions]** In this paper, multi-dimensional graph convolutional network and multi-modal features are integrated into helpfulness recognition, which effectively improves the recognition effect.

**Key words:** multimodal features; multi-dimensional; graph convolutional network; travel reviews; helpfulness detection

## 1. 引言

信息技术的快速发展,使得消费者评论的重要性愈发凸显,并影响着消费者的购买决策和公司的业绩表现<sup>[1]</sup>。同时,由于酒店的在线评论数据呈指数级增长,信息过载导致消费者翻阅信息时耗时耗力,影响其最终的购买决定<sup>[2]</sup>。基于在线评论在客户决策中的重要作用,研究在线评论的有用性就有其现实必要<sup>[3]</sup>。尽管信息内容广泛,但由于非结构化和多模态的性质,在线评论的有用性识别仍然是一个具有挑战性的任务<sup>[4]</sup>,成为学界关注的重要议题之一。研究评论的有用性有以下体现:消费者阅读了评论并在评估后将其确定为有帮助,并促成下一个消费者做出理性的消费决定。

随着人工智能技术的发展,机器学习技术已经广泛应用于旅游评论的有用性识别<sup>[5]</sup>。现阶段的旅游评论的特征提取大多数都是基于文本模态,通常是从相关的文本信息中提取特征进行有用性识别,使用多模态数据特征的研究较少<sup>[6]</sup>。现有模型主要考虑了单一维度的评论特征,而这类特征在特征提取中无法深度挖掘旅游评论的含义,提取特征较片面,缺乏从多种维度综合考虑旅游评论的意义,来帮助模型提高有用性识别的准确度。

本研究从排名第一的旅游网站Yelp上收集酒店数据,基于多维度图卷积网络(Graph convolutional neural networks, GCN)构建消费者有用性评论。本研究通过文献回顾确定可能影响旅游评论有用性的影响因素,并利用多模态深度学习构建模型,以识别有用性的旅游评论。同时,提出多维度图卷积网络模型,结合不同的维度特征,对比不同的基线模型,提高识别的准确性。本文的贡献概括如下:

(1)提出基于多维度图卷积网络和注意力机制相结合的模型,可确定不同维度的重要性权重,通过实验验证了其性能优于当前的主流方法;

(2)设计多模态的特征提取结构,提高模型多维度学习文本与图片间语义相关性的表达能力;

(3)构建消费者评论有用性的识别框架,为消费者和酒店方提供商业决策,也

可将此框架推广到其他相关领域。

## 2. 相关工作

### 2.1 旅游评论的有用性识别

旅游评论的有用性识别可以辅助消费者做出购买决定,提高酒店管理者的决策分析。有用性的识别方法可以分为以下两个方面<sup>[7]</sup>:旅游评论的因素<sup>[8]</sup>和确定评论的有用性的识别方法<sup>[8]</sup>。**Weathers**等以用户体验为导向讨论旅游产品评论的有用性<sup>[9]</sup>,但该研究缺乏文本的语义分析。**Chua**等采用统计方法考虑评论文本的情绪、产品类型和评论质量<sup>[10]</sup>。**Zhang**等应用统计工具并采用了多语言方法。但此类研究没有考虑评论多维度方面<sup>[11]</sup>。**Siering**等,**Gao**等还发现消费者评论由具有更多经验的用户撰写时更有帮助<sup>[12, 13]</sup>。此外,**Zhou**等的提出使用用户照片,他们的评论被认为更有帮助<sup>[14]</sup>。**马超**等融合图文特征向量,认为图文结合的多模态评论能够更好地进行在线评论有用性识别<sup>[15]</sup>。

研究人员还提出了一些其他因素影响消费者评论有用性识别。例如,考虑到分析消费者评论的语义和风格的影响<sup>[16]</sup>或消费者情绪的影响<sup>[17]</sup>。**Kaushik**等发现正面评价对旅游产品的销售产生积极影响<sup>[18]</sup>。**Chen**等基于情感维度的分析并得出结论,负面情绪会影响有用性,但是没有评估情感的导向性<sup>[19]</sup>。因此消费者的评论内容会影响其他消费者如何看待的情感的促进作用<sup>[20]</sup>。

现有研究从不同因素的对消费者评论进行有用性识别,为本研究的开展提供了参考。但它们多是从单个因素切入,忽略了每个因素间的特征联系,因此,本文将消费者的评论作为多模态数据输入,并结合图卷积结构来描述两个节点消费者评论之间的相似性。在图卷积网络框架的帮助下,多维度模型可以有效的表示多模态数据特征,有助于提高有用性识别的准确性。

### 2.2 图卷积网络在旅游领域的应用

如前所述,旅游评论的有用性识别引起了学术界的广泛关注,之前的研究主要使用传统的机器学习<sup>[21]</sup>,相较而言,深度学习模型的优势无需赘言。尤其消费者有用性识别的任务中取得成功<sup>[22]</sup>。例如卷积神经网络 (CNN),该模型从评论中学习编码上下文理解文本的语义特征<sup>[23]</sup>。**Kim**对在用于句子级别的文本分类任务上表现非常出色<sup>[24]</sup>。**Chen**等识别消费者评论的有用性使用两种技术:嵌入门控 CNN 和跨领域关系学习,他们的模型基于单词、字符和主题表示,并且还能够从评论中提取多粒度文本特征<sup>[25]</sup>。**Saumya**等分析评论特征,预测评论的有用性,构建了一个两层的 CNN 模型,采用不同的过滤器和回归方法来找到最小均方差<sup>[26]</sup>。相比之下,长短期记忆 (LSTM) 模型可以更准确地处理更长时间序列数据。**Fan**等采用 Bi-LSTM 模型处理产品的原始数据<sup>[27]</sup>。**Ma**等研究评论质量,将图片视为影响因素,结果证明带有图片的评论更比单一的文本评论更有帮助<sup>[28]</sup>。

随着卷积神经网络的广泛应用,图卷积网络越来越受到关注。这些基于图卷积的网络嵌入方法,不仅利用节点及其邻域的特征信息,而且还在图嵌入中保留全局的结构信息。在图卷积网络的开创性工作中,**Kipf**等提出了一个简化的图卷积网络模型<sup>[29]</sup>,有效地集成了连接边和节点的特征信息。尽管该模型在大量数据上取得了不错的分类结果,它仍然有两个局限性:一方面,图卷积网络需要通过堆叠卷积来整合信息,来计算其特征层;另一方面,图卷积网络通过节点迭代传播邻域特征,即通过较低节点传播较高的特征信息,这使得信息在每一步都损失和变形<sup>[30]</sup>。因此,本文利用丰富的多模态信息,构建多维度的图卷积网络,最



大程度的提取消费者评论中的多模态特征信息,有效的提高消费者识别的准确率。

### 3. 研究设计

本研究提出的基于多维度图卷积网络和多模态融合的旅游评论有用性识别方法,主要包含 3 个模块,分别是输入模块、图卷积网络模块和识别模块。详细模型结构如图 1 所示。

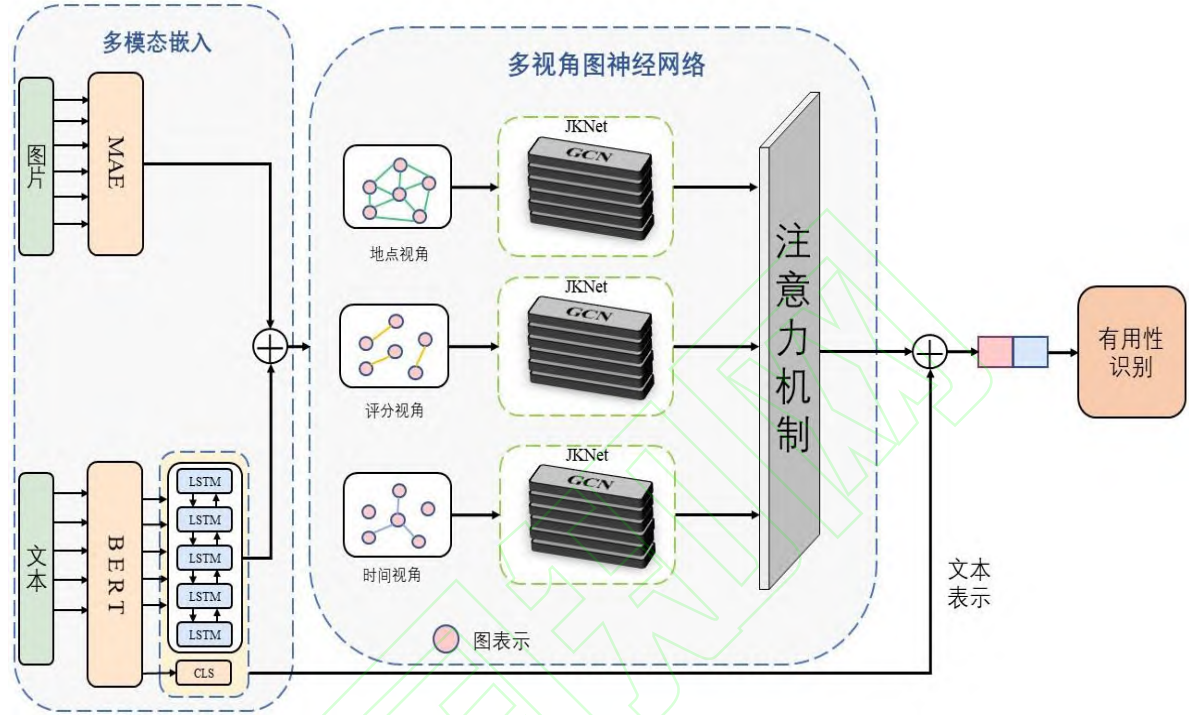


图 1 本研究的模型结构图

Figure 1 The model structure of this paper

#### 3.1 输入模块

##### 3.1.1 文本输入

本研究的训练模型利用 BERT<sup>[31]</sup>模型,它是一种新的语言表示模型,使用一个双向 Transformer 建立一个大型的语料库,并对预训练的模型进行微调。此时本文用的 BERT 的 Transformer 块数为 24,隐藏层大小为 1024, self-attention<sup>[32]</sup>个数为 16,预训练模型参数为 340M。因此,我们使用 BERT 作为本文的预训练模型。

然后用双向 LSTM 对文本进行编码,嵌入矩阵可以表示为:

$$H^T = \{\vec{h}_1^T, \vec{h}_2^T, \dots, \vec{h}_n^T\} = Bi-LSTM(X) \quad (1)$$

##### 3.1.2 图片输入

本研究基于掩码自动编码器(Masked Auto Encoder, MAE)<sup>[33]</sup>具有非对称架构。对比传统的 VGG, 它使用很小的卷积核(3×3×3)构建各种深度的卷积神经网络结构,并对这些网络结构进行了评估,能够取得较好的识别精度。相比传统的 VGG 方法, MAE 可以作为计算机视觉的可扩展自监督学习器使用,遮掩掉输入图像的随机像素并重建它们。其核心概念是一个编码器只对可见的像素子集进行操作,另一个简单解码器可以从潜在表征和被遮掩掉单元重建原始图像。解码器的架构可以是十分轻量化的模型,且具体的架构对模型性能影响很大。遮掩掉

大部分输入图像会产生重要且有意义的自监督任务。相比传统的 VGG 等预训练方法，在图卷积时改进了图片像素单元的重构，有效生成大量的新的图片向量，这样就能高效地训练大型模型。在 ImageNet-1K 数据测试上，MAE 模型的最高精确度为 87.80%<sup>[33]</sup>，明显提升了图片的预训练效果。因此本文采用最新的 MAE 作为图片的预训练方法。

MAE 采用输入图片是小分子集和一个更小的解码器，向量化输入的图谱的概率密度数据， $x$  是通过分解计算得到的概率链法则，可以表示为：

$$p(x) = \prod_{d=1}^D p(x_d | x_{<d}) \quad (2)$$

在自动编码器输出中，每个维度都可以解释为  $D$  条件概率分布之一。如上所示，每个输出单元只依赖于前一个输入单位。我们将每个条件分布参数化为一个混合高斯的  $C$ ，即自动编码器输出均值、方差和混合成分的概率。

### 3.1.3 多模态融合

本研究直接将文本特征和图片特征进行拼接融合，从而对两者进行加权联合，使得各模态特征之间相互补充和关联。对于每一条评论  $R$ ，经过 BERT 特征提取后，得到对应的词向量  $W=\{w_1, w_2, w_3, \dots, w_n\}$ 。同时，图片经过 MAE 的特征提取后，得到对应  $k$  的词向量的图片序列  $P=\{p_1, p_2, p_3, \dots, p_k\}$ 。每一条评论包含  $m$  张图片。同理，每一条评论对应的图片特征向量为  $I_p=\{p_1, p_2, p_3, \dots, p_m\}$ 。最后得到的多模态融合的向量为  $R=\{w_1, w_2, w_3, \dots, w_n, p_1, p_2, p_3, \dots, p_m\}$ 。

## 3.2 图卷积网络模块

该模块根据时间、评分和地点三个维度一起使用图卷积网络，将每个维度分别定义为  $\hat{A}_1$ 、 $\hat{A}_2$  和  $\hat{A}_3$ 。以上三个维度分别对多模特征进行单独输入，再通过图卷积网络提取其特征。

在时间维度中，它侧重于消费者评论中的多模态时间序列信息，主要表现在文本与对应图片的数量，长度。尤其是文本内容中的单词和句子带有重要的语义，评论中上下文的序列信息影响评论的整体质量和有用性。然而，传统的词袋和词嵌入方法无法全面的捕获多模态数据的交互信息。由于图卷积网络可以有效提取其多模态特征，因此本文在时间维度中以文本和图片融合后的数量作为节点，其不同文本间的相似性作为连边，构成图卷积网络，这些作为模型的输入。

在评分维度中，模型主要依赖于消费者的情感特征。然而，传统的机器学习模型只能提取多模态信息中的显性情感信息，一些隐性的情感信息无法有效的提取，比如一些无意义或方面情感信息。这要求模型具备强大的情感分析能力。由于图卷积网络可以借助边和节点深入挖掘多模态数据的情感信息。因此本文在评分维度中以消费在的不同评分数量作为节点，不同文本的相似性作为连边，构成图卷积网络，这些作为模型的输入。

在地点维度上，主要突出其空间分布特征。本文收集了美国的四个城市，每个酒店都有不同的地理信息。单依靠传统的社交网络方法不能全面提取其空间信息。图卷积网络可以更深入提取其酒店分布的空间地理信息，由于本研究输入了酒店品牌和地点，很容易构成了图卷积网络的边与节点，这些作为模型的输入。

然后根据每条文本和对应的图片作为以后节点  $Q^t = \{q_1^t, q_2^t, \dots, q_r^t\}$ , 基于此进行构建邻接矩阵  $A_t$ , 给定图结构  $g = (v, \varepsilon)$ , 节点为  $N = |V| = |R|$ , 边的集合为  $\varepsilon$ ,  $A_t \in \mathbb{R}^{N \times N}$  是邻接矩阵,  $X \in \mathbb{R}^{N \times N}$  代表是特征矩阵。根据图卷积网络节点是否在同一组中构建地点节点之间的边, 稀疏邻接矩阵  $A$  定义为:

$$A_{ij} = \begin{cases} 1 & \text{if } i, j \text{ have the same semantic information} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

其中  $i, j$  是不同的消费者评论实体。

本文基于 Defferrar 等提出的光谱滤波器<sup>[34]</sup>, 构建一个  $K^{th}$  局部化的 ChebNet, 并使用高达  $K$  阶的 Chebyshev 多项式, 因而图卷积可以定义为:

$$g\theta * x \approx \sum_{k=0}^K \theta'_k T_k(L_{sym})x \quad (4)$$

其中  $x \in \mathbb{R}^n$  是图上的信号,  $g\theta$  是频谱滤波器,  $*$  表示卷积核子,  $T_k$  是 Chebyshev 多项式,  $\theta' \in \mathbb{R}^K$  是 Chebyshev 的系数,  $L_{sym}$  是对称的拉普拉斯算子。

本研究将多模态特征归一化邻接矩阵和拉普拉斯矩阵  $L = D - A$  描述一阶接近度, 它模拟矩阵之间的局部成对相似性节点。其中归一化的邻接矩阵  $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$  是随机的转移概率矩阵。此外,  $k$  阶邻近矩阵可以在建模之前计算,  $k$  阶接近度矩阵可以表示为:

$$\hat{A}^k = \underbrace{\hat{A} * \hat{A} \dots \hat{A}}_k \quad (5)$$

每种维度的多模态特征邻近矩阵都包含邻居的多尺度信息。具体来说, 利用图卷积网络提供不同的深度邻近矩阵并遵循更新规则<sup>[35]</sup>。由于图卷积网络可以优化多模态数据输入的卷积参数, 有效的从大量样本中学习到相应地特征, 避免了复杂的特征提取过程。因此, 在时间维度中, 通过图卷积网络学习到了文本与图片的交互语义特征。在评分维度中, 模型学习到了多模态数据的隐性的情感信息。在地点维度中, 模型学习到了多模态数据的空间分布信息。

本文构建图卷积网络的目标是获得多维度特征的相似性, 通过图嵌入方法提高识别的性能。图嵌入的方法旨在将网络中的节点映射到低维向量, 保留网络中结构和节点的特征空间, 但无法捕获超出节点的二阶信息。除了图卷积不停地将邻域特征传播到节点, 这使得每一步的信息都会变形, 即高维度的信息通过较低维度的节点。这样的传播方式使高维的特征信息过度变形, 并造成大量的损失。

为了应对这样的信息损失, 本文提出了多维度的图卷积网络, 包括消费者评论的时间, 情感和空间特征, 并以显式的方式同步保存了这些信息, 这使得来自不同类别的节点变得更容易识别。具体来说, 取而代之的是编码单元重新设置了消费者评论的多模态信息, 将每个酒店的空间特征视为一个节点, 并使用他们的相关多模型信息来构建图卷积网络。在每一个维度中, 每个图呈现一种关系转化的特定特征关系网络。为了利用不同维度的邻近信息。本文扩展使用图卷积网络捕获邻近的多尺度信息, 然后是节点特征和图卷积块的输出由注意力机制整合以

获得最终的表示，主要是把以上三个维度的信息通过注意力机制进行融合。由于注意力机制每一步计算不依赖于上一步的计算结果，因此可以并行处理三个维度的特征。

根据从三个维度获得的多模态复合特征，通过非线性得到相应的注意力分数  $u_i (1 \leq i \leq m)$ ，然后通过 softmax 函数对每个注意力分数进行归一化。

$$u_i = \tanh(W_i z_i + b_i) \quad (6)$$

$$a_i = \frac{\exp(u_i)}{\sum_{l=1}^m \exp(u_l)} \quad (7)$$

$$F_a = \sum_{n=1}^3 a_i z_i^n \quad (8)$$

其中  $a_i$  是每个图模型的权重。对三个图的特征进行加权平均，得到算法的输出向量  $F_a \in \mathbb{R}^{n \times f}$ 。

### 3.3 识别模块

加入文本特征，与以上的输出向量合并。对于有用性的识别任务，将分为有用或没用两个标签。这里用线性层进行分类，识别值  $\hat{y}$  表示为：

$$\hat{y} = \text{softmax}(f_a W_p^n + b_p^n) \quad (9)$$

其中  $W_p^n \in \mathbb{R}^{d \times c}$ ,  $b_p^n \in \mathbb{R}^c$ ,  $c$  是有用性类别的数量，然后使用交叉熵的损失函数训练旅游评论的多模态数据。

$$\text{Loss} = - \sum_{o \in Z_u} \sum_{g=1}^l y_{ig} \ln \hat{y}_{ig} \quad (10)$$

其中  $y_{ig}$  是旅游评论有用性的真实值， $Z_u$  表示每个酒店  $o$  中的所有评论数量。

## 4. 实验与分析

### 4.1 数据收集及预处理

本研究使用基于 Python 的 Selenium 包对 Yelp 网站进行爬取，并选取四个品牌的酒店（Crown Plaza, Hilton, Marriot, Westin）以美国的四个旅游城市，（迈阿密、纽约、洛杉矶和旧金山）构建数据集。数据采集的时间为 2008 年 12 月 21 日到 2021 年 10 月 10 日，共获取 32015 条评论。为了减少评论数据中的噪音，对其进行了如下处理：①去除重复的评论；②去除停用词；③去除不相关的图片。最后获取了有效文本 25908 条，图片 10265 张。

### 4.2 对比实验设计

#### 4.2.1 文本模态

(1) **Text-CNN**<sup>[36]</sup>：采用所提出的文本模型 Text-CNN 作为特征提取，经过一个全连接层来实现评论有用性的识别。



(2) **Bi-LSTM**<sup>[37]</sup>: 采用长期短期记忆(LSTM) 来捕获文本的序列信息。

(3) **BERT**: 该模型是谷歌公司提供的文本预训练模型, 模型的大小为 110M, 同样经过一个 softmax 来实现识别。

#### 4.2.2 图片模态

(1) **Visual**: 采用预先训练的 VGG-19<sup>[38]</sup>和全连接层提取图片特征, 然后进行有用性识别, 全连接层级的维度为 128。

(2) **VAQ**<sup>[39]</sup>: 该视觉识别模型源自于问答系统, 把此模型的最后一层进行修改, 适应有用性识别的任务。

#### 4.2.3 多模态

(1) **Att-RNN**<sup>[40]</sup>: 采用 LSTM 提取文本特征, 同时采用 VGG-19 提取图片特征, 并利用注意力机制融合多模态特征, 实现有用性识别。

(2) **MAVE**<sup>[41]</sup>: 该模型学习文本和图片多模态特征共享表示, 以识别有用性, 利用变分自动编码器对输入数据进行重构得到共享表示, 并利用二值分类进行有效性识别。

(3) **BEANN**: 采用 BERT 提取文本特征, 与图片特征进行连接, 融合多模态信息进行有用性识别。

#### 4.2.4 模型参数设置

本研究将标注后的数据以 80%的比例划分为训练集, 20%的比例为测试集, 训练集为 20726 条, 测试集为 5182 条。在实验中采用 Python 的 NLTK 工具包来分隔单词和标签。利用 BERT 模块将每个单词以 768 维作为词嵌入, 同理图片也是一样的维度, 并利用 MAE 作为预训练, 图片的分辨率为 32\*32, 图卷积网络的层数为 2 层, 隐藏单元的 GCN 中设置为[128, 64], 在 L2 正则化中 $\lambda$ 为 0.01, 以 0.002 为学习率来训练模型, 以 32 的 batch 尺寸和 0.1 的 dropout 防止模型的过拟合。

在结果评估中, 本文使用 Accuracy、Precision、Recall 和 F1-Score 分数来评估模型的性能。本文主要使用准确度作为性能指标, 并以十折交叉的方法验证的准确性, 以减少参数随机性的影响。

### 4.3 实验结果

表 1 本文的实验结果

Table 1 Experimental results of this paper

	模型	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
文本 模态	Text-CNN	58.32	58.48	57.32	56.14
	Bi-LSTM	58.41	59.36	58.16	59.08
	BERT	62.25	61.85	62.07	62.66
图片 模态	Visual	55.68	54.82	55.13	54.76
	VAQ	56.06	55.58	56.79	55.16
	Att-RNN	63.83	62.76	63.14	62.75
多模态	MAVE	66.42	65.95	66.08	66.93
	BEANN	68.18	68.36	69.74	68.40
	本研究	73.21	74.68	74.20	73.85

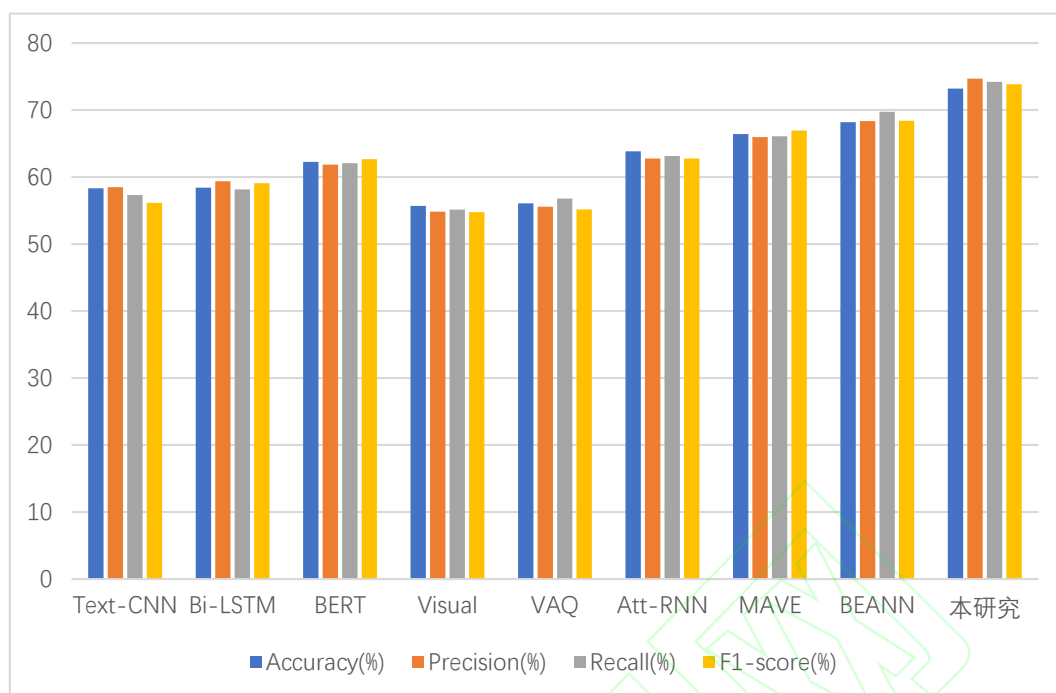


图 2 对比实验结果

Figure 2 Comparison of experimental results

表 2 多维度与图卷积网络的结果

Table 2 Results of multi-dimensional and graph convolutional network

模型	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
Location+ GCN	72.15	71.88	71.46	71.63
Rating + GCN	71.52	72.06	71.84	71.46
Time + GCN	72.03	71.56	72.79	71.24
GCN	68.44	67.69	68.20	67.08

从表 1 和图 2 来看，在单模态模型中，文本模态的有用性的识别明显高于图片模态。因为在消费者有用性识别中，文本模态可以表达消费者的直接情感特征，其语义特征直接可以帮助识别其有用性。在文本模态中，利用 BERT 模型明显优于 Text-CNN 模型，准确率提升了 3.93%，甚至达到多模态模型的效果。这说明文本的预训练模型对于其特征提取有显著的增强效果。

在多模态模型中，MAVE 相较于 VAQ 和 Att-RNN 模型检测效果更好。该模型利用了 Text-CNN 模型进行提取，相较于文本模态的 Text-CNN 准确度提升了 8.1%，证明了图片特征对文本特征上进行了有效的增强。其中，本研究提出的多维度图卷积网络在各项指标均优于其他模型，主要是利用多维度拼接一定程度上提高了多模态特征的权重，注意力机制又能全局捕捉模态之间的交互信息。加上利用了 BERT 和 MAE 预训练模型，使得文本和图片特征能够有效融合，达到最优的效果。

根据表 2 所示，本文对不同维度和图卷积网络进行了对比，可以发现地点维度的准确度最高，高于其他维度 1%。是由于在图卷积网络建模过程中，主要以地点维度为边，不同酒店为节点进行建模，这样形成图网络的稳定结构，有效提升了识别的准确性。本研究如果不用多维度，准确度只有 68.44%，远低于使

用了多维度的图模型，结果也证明多维度很大程度上提升了识别的准确性。

综上，多模态与单模态模型相比，能够提高消费者有用性识别的效果，其中文本特征起了关键作用，图片特征对文本特征进行有效的补充。BERT 模型对文本特征的提取效果显著，同时注意力机制能有效捕捉不同模态间的交互信息，充分挖掘了文本和图片的潜在关系；多维度的图卷积网络可以充分挖掘多模态特征，使得消费者有用性的识别效果具有显著提升。此外，从多维度的结果对比来看，对图卷积网络识别效果的提升也很明显。因此，本研究所提出的基于多维度图卷积网络和多模态特征融合的有用性识别模型，在 Yelp 的数据集上的识别效果显著提升。

4.4 消融实验

表 3 消融实验结果

Table 3 Ablation experiment results

模型	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
Text only	68.13	69.01	68.56	68.44
Image only	64.81	63.25	64.76	63.45
-Location	69.80	70.26	69.14	68.25
-Rating	72.04	71.96	72.05	71.62
-Time	71.73	72.96	71.04	72.86
-Attention	71.54	72.45	71.86	72.05
本研究	73.21	74.85	73.49	75.54

(1) 多模态模型优于单模态模型

为进一步了解模型中各个特征的作用，本研究在 Yelp 数据集中进行了消融实验，结果如表 3 所示。可以观察到单独的文本特征为 68.13%，明显高于图片的特征 64.81%。与单模态相比，多模态特征明显高于文本和图片的单模态。比如地点、评分和时间这三个维度明显高于文本的 68.13% 和图片的 64.81%。有效地证明了多模态的模型性能。以上实验结果也证明多模态特征对有用性识别的重要性。

(2) 多维度角度优于单个维度

本文提出多维度的图卷积网络，从地点、消费者评分和时间来进行图卷积网络的模型构造。可以从表中观察到，准确度去掉地点维度是 69.80%，去掉消费者评分维度是 72.04%，去掉时间维度是 71.73%。对比之下，消费者的评分维度权重更高，主要因为消费者的评论特征与其评分关联度最高。特别是，去掉注意力机制的精确度为 71.54%，与其他的维度相比，比地点维度要高，但比其他两个部分要低，说明注意力机制模型还是在维度后的特征提取中发挥了重要的作用，本文利用多模态包含的丰富的特征，并结合注意力机制，有效的提高了有用性的识别。

5. 讨论与结语

5.1 讨论

旅游评论有用性的多层因素分析获得众多学者的关注，本研究的主要目的在于了解评论的文本内容，以分析其有用性。本研究从三个维度分析了相应的权重，解决了一个关键的方面，即消费者在评论有用性的识别。利用多维度提取消费者的有用性评分，从整体上丰富该评论的内容含义。

本研究从多维度的角度,估计了每个维度的权重,可以有效的提高消费者评论的有用性识别。由于消费者评论的有用性影响其品牌形象,本研究成果可以进一步扩展到酒店品牌推广领域,为相关管理人员提供决策参考。在消费者购买流程中,有用的评论起着重要的作用,尤其是在其服务的替代品的评估上。一个内容丰富、写得好的评论可以减少消费者对其他竞争产品的选择。评估有用性可以作为产品推荐的工具。当有经验的用户给出评论时,它可以被视为消费者购买后的行为特征,它将提供有关产品的服务信息。在这样的场景下,它是帮助消费者了解酒店产品、服务和品牌等信息的主要途径。

## 5.2 结语

本研究深入探讨旅游评论的有用性问题。在本文中,不同于与传统直接编码图卷积网络的方法,本研究获取每个酒店空间节点的表示,具有从消费者多模态信息转换而来的图形结构信息。为了更好的利用邻接的特征信息,扩展原始GCN模型,以捕获多维度的邻接信息,并保留丰富的多粒度信息。在Yelp数据集上进行实验,识别消费者评论的有用性,通过多模态深度学习对消费者的评论进行深度挖掘,对比现有的三种基线模型,证明本研究所提出模型的优越性。

本文从一个新角度解决了有用性的识别问题,但依然存在一些局限。首先,基于深度学习的模型有很多新的技术,后续研究中可以引入。图卷积网络在这项研究中运行良好,但有提示学习和对比学习的新模型会打开技术应用的新思路。其次,后续研究维度可以进一步拓展,例如消费者的类型、发布时间等。

## 参考文献

1. Ye Q, Law R, Gu B, et al. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings[J]. Computers in Human behavior, 2011, 27(2): 634-639.
2. Chen C C, Tseng Y D. Quality evaluation of product reviews using an information quality framework[J]. Decision Support Systems, 2011, 50(4): 755-768.
3. Zhu L, Yin G, He W. Is this opinion leader's review useful? Peripheral cues for online review helpfulness[J]. Journal of Electronic Commerce Research, 2014, 15(4): 267.
4. Zhu W, Cui P, Wang Z, et al. Multimedia big data computing[J]. IEEE multimedia, 2015, 22(3): 96-c3.
5. Jain P K, Pamula R, Srivastava G. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews[J]. Computer Science Review, 2021, 41: 100413.
6. Tsai C F, Chen K, Hu Y H, et al. Improving text summarization of online hotel reviews with review helpfulness and sentiment[J]. Tourism Management, 2020, 80: 104122.
7. Diaz G O, Ng V. Modeling and prediction of online product review helpfulness: a survey[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 698-708.
8. Almagrabi H, Malibari A, McNaught J. A survey of quality prediction of product reviews[J]. International Journal of Advanced Computer Science and Applications, 2015, 6(11): 49-58.



9. Weathers D, Swain S D, Grover V. Can online product reviews be more helpful? Examining characteristics of information content by product type[J]. *Decision Support Systems*, 2015, 79: 12-23.
10. Chua A Y K, Banerjee S. Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality[J]. *Computers in Human Behavior*, 2016, 54: 547-554.
11. Zhang Y, Lin Z. Predicting the helpfulness of online product reviews: A multilingual approach[J]. *Electronic Commerce Research and Applications*, 2018, 27: 1-10.
12. Siering M, Muntermann J, Rajagopalan B. Explaining and predicting online review helpfulness: The role of content and reviewer-related signals[J]. *Decision Support Systems*, 2018, 108: 1-12.
13. Gao B, Hu N, Bose I. Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews[J]. *Decision Support Systems*, 2017, 95: 1-11.
14. Zhou S, Guo B. The order effect on online review helpfulness: A social influence perspective[J]. *Decision Support Systems*, 2017, 93: 77-87.
15. 马超, 李纲, 陈思菁, 等. 基于多模态数据语义融合的旅游在线评论有用性识别研究[J]. *情报学报*, 2020, 39(2): 199-207. (Research on usefulness recognition of tourism online reviews based on multimodal data semantic fusion, *Journal of the China Society for Scientific and Technical Information*, 2020, 39(2): 199-207)
16. Cao Q, Duan W, Gan Q. Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach[J]. *Decision Support Systems*, 2011, 50(2): 511-521.
17. Salehan M, Kim D J. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics[J]. *Decision Support Systems*, 2016, 81: 30-40.
18. Kaushik K, Mishra R, Rana N P, et al. Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon. in[J]. *Journal of retailing and Consumer Services*, 2018, 45: 21-32.
19. Chen M J, Farn C K. Examining the influence of emotional expressions in online consumer reviews on perceived helpfulness[J]. *Information Processing & Management*, 2020, 57(6): 102266.
20. Yang S, Yao J, Qazi A. Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining[J]. *Information Processing & Management*, 2020, 57(2): 102179.
21. Zhang Y, Lin Z. Predicting the helpfulness of online product reviews: A multilingual approach[J]. *Electronic Commerce Research and Applications*, 2018, 27: 1-10.
22. Li J, Liu X, Yin W, et al. Empirical evaluation of multi-task learning in deep neural networks for natural language processing[J]. *Neural Computing and Applications*, 2021, 33(9): 4417-4428.
23. Du J, Rong J, Wang H, et al. Helpfulness prediction for online reviews with explicit content-rating interaction[C]//*International Conference on Web Information*

Systems Engineering. Springer, Cham, 2020: 795-809.

24. Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1510.03820, 2015.
25. Chen C, Qiu M, Yang Y, et al. Multi-domain gated CNN for review helpfulness prediction[C]//The World Wide Web Conference. 2019: 2630-2636.
26. Saumya S, Singh J P, Dwivedi Y K. Predicting the helpfulness score of online reviews using convolutional neural network[J]. Soft Computing, 2020, 24(15): 10989-11005.
27. Fan M, Feng C, Guo L, et al. Product-aware helpfulness prediction of online reviews[C]//The World Wide Web Conference. 2019: 2715-2721.
28. Ma Y, Xiang Z, Du Q, et al. Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning[J]. International Journal of Hospitality Management, 2018, 71: 120-131.
29. Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.
30. Soni U, Bhambhani M, Khapra M M. Network embedding using hierarchical feature aggregation[J]. 2018.
31. 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.
32. Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations[J]. arXiv preprint arXiv:1803.02155, 2018.
33. He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[J]. arXiv preprint arXiv:2111.06377, 2021.
34. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in neural information processing systems, 2016, 29.
35. Cao S, Lu W, Xu Q. Grarep: Learning graph representations with global structural information[C]//Proceedings of the 24th ACM international on conference on information and knowledge management. 2015: 891-900.
36. Luan Y, Lin S. Research on text classification based on CNN and LSTM[C]//2019 IEEE international conference on artificial intelligence and computer applications (ICAICA). IEEE, 2019: 352-355.
37. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.
38. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
39. Antol S, Agrawal A, Lu J, et al. Vqa: Visual question answering[C]//Proceedings of the IEEE international conference on computer vision, 2015: 2425-2433.
40. Jin Z, Cao J, Guo H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]//Proceedings of the 25th ACM international conference on Multimedia., 2017: 795-816.
41. Khattar D, Goud J S, Gupta M, et al. Mvae: Multimodal variational autoencoder for

fake news detection[C]//The World Wide Web Conference,2019:2915-2921.

(通讯作者: 丁星辰, ORCID: 0000-0002-4229-5340, E-mail: xingos@whu.edu.cn)

**作者贡献声明:**

刘洋: 数据处理, 设计研究方案, 论文写作;

丁星辰: 数据处理, 模型构建与实现;

马莉莉: 设计研究方案, 论文修改;

王淳洋: 论文修改;

朱立芳: 论文修改。

**利益冲突声明:**

所有作者声明不存在利益冲突关系。

