

GRADUATION THESIS

基于弱监督学习的电商图像 关键信息识别标注

答辩学生：张世雨

指导老师：王薇 王小宁

中国传媒大学广告学院
2021级广告学（计算广告双学士学位复合型人才培养项目）

目录

Contents

01

选题的背景意义

Background significance of the topic

02

研究问题与方法

Research problems and methods

03

实验过程与结果

Experimental process and results

04

论文结论与不足

Conclusion and deficiency of this paper

研究背景

行业需求与技术挑战



电子商务市场扩张

2024 年中国网上零售额达 15.5 万亿元，连续12年成为全球最大网络零售市场。全球市场规模预计 2027 年达 8 万亿美元，商品快速上架对图像标注效率提出极高要求。



传统标注瓶颈

监督学习方法人工标注成本高（单图耗时 > 3 分钟）、噪声大（标注一致性问题），难以应对日均百万级新品上架。



弱监督学习价值

利用图像级粗粒度标签（如商品标题、行业分类），挖掘图像和文本的隐含关联，降低对精确标注的依赖。但仍存在跨模态特征对齐不充分等技术问题。

研究目标

图像中的关键信息对电商营销至关重要，但传统的全监督模型依赖大量精准标注样本，弱监督方法存在跨模态对齐不充分问题。

针对文本数据中隐含的粗粒度监督信号，探究如何通过语义解析等技术，实现弱监督信号的高效提取与噪声鲁棒性筛选，突破传统人工标注的语义局限性。

围绕文本语义空间与图像视觉特征空间的异构性鸿沟，研究跨模态特征对齐的理论方法，实现文本弱监督信号与图像局部的深度融合与互补增强。

研究意义

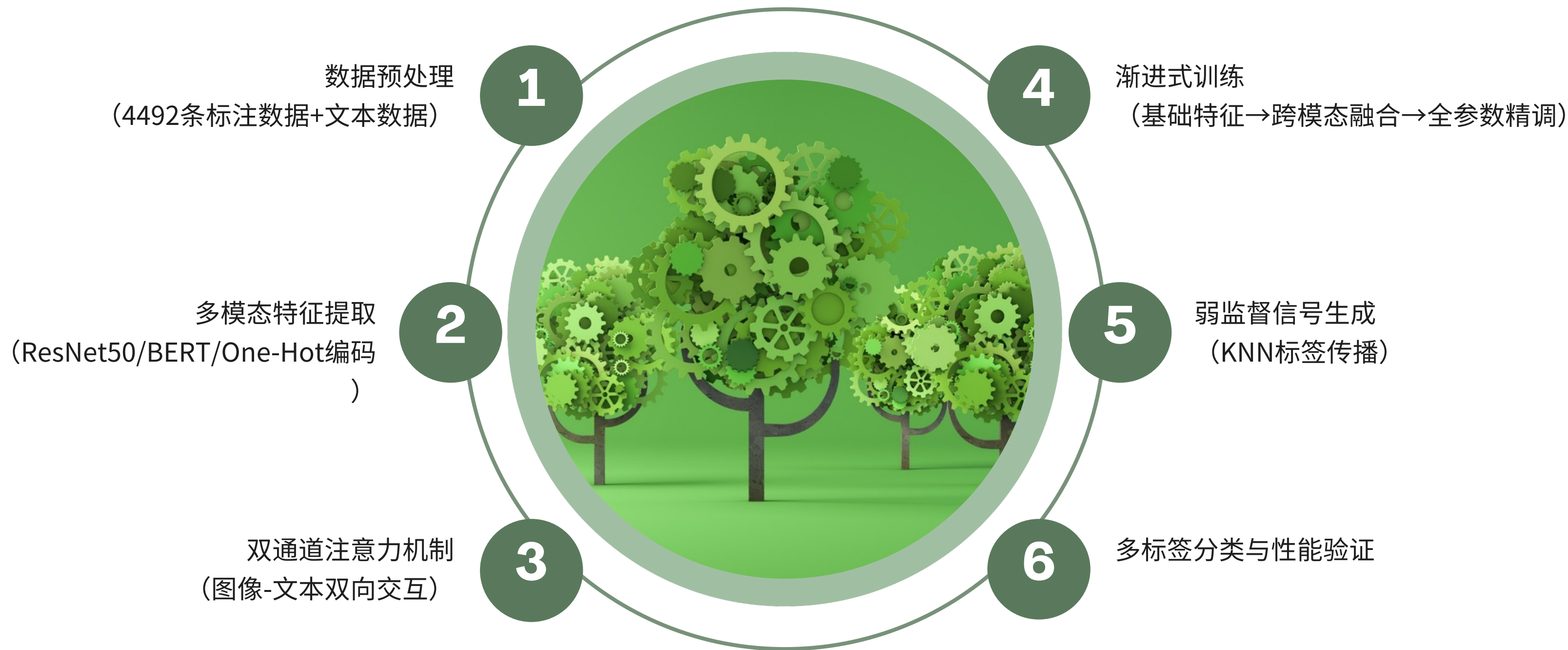
理论意义

- 构建多模态弱监督融合框架，推动跨模态语义对齐理论创新。
- 提出双通道注意力机制，拓展注意力机制在弱监督场景的应用边界。

实践意义

- 减少对人工标注的依赖，降低电商平台图像数据标注成本。
- 支持精准搜索与广告投放，提升用户体验与运营效率。

技术路线



研究方法

多模态特征交互框架

三模态输入：

- 视觉特征：ResNet50 提取 2048 维图像特征。
- 文本特征：BERT 生成 768 维语义向量（CLS token）。
- 行业特征：One-Hot 编码 91 维行业属性（一级 + 二级分类）。

特征融合：通过线性变换与跨模态注意力，动态捕捉图像区域与文本关键词的语义关联。

01

双通道注意力机制

双向注意力设计：

- 图像→文本注意力：定位图像关键区域（如 LOGO、赠品）对文本关键词的响应。
- 文本→图像注意力：强化文本语义（如 “限时折扣”）对视觉特征的引导。

优势：解决传统单向注意力的模态错位问题，提升特征表示的语义丰富性。

02

渐进式训练策略

三阶段优化：

- 基础学习（Epoch 1-5）：冻结注意力层，仅优化分类头，快速建立基础分类能力。
- 跨模态融合（Epoch 6-15）：解冻注意力模块，学习模态间交互权重，提升复杂元素识别精度。
- 精调阶段（Epoch 16-30）：降低学习率，全参数优化，增强噪声鲁棒性。

03

实验与结果



01

数据集

阿里瓴羊提供 4492 条电商图像，涵盖“主体”“赠品”“LOGO”等 5 类标签，平均每图 3.17 个标注框。



02

评估指标

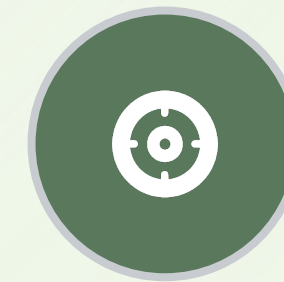
- 宏平均 F1 值、加权平均 F1 值（多标签场景）。
- 消融实验（验证行业特征、注意力机制的贡献）。



03

主实验结果

仅 10% 精确标注数据时，宏平均 $F1=0.847$ ，加权平均 $F1=0.888$ ，显著优于传统弱监督方法。



04

消融实验

行业特征引入使召回率提升 19.8%，双通道注意力机制贡献 10.2% 的精度增益。

论文结论

01

方法有效性

多模态弱监督信号融合机制，成功将商品标题、行业分类等弱标注数据与少量精确标注结合，通过跨模态特征对齐与噪声鲁棒训练策略，实现了对电商图像关键信息的高效识别。

02

行业应用

拓展了弱监督学习在多模态场景下的应用边界，构建的轻量化模型满足电商平台实时标注需求，在搜索等多种电商场景中具有直接落地价值，为行业降低运营成本提供了可复制的技术方案。

03

模型创新

设计的双通道注意力网络与渐进式训练策略，有效提升了模型对跨模态语义关联的捕捉能力，能动态聚焦图像关键区域与文本关键词的对应关系，证明了语义引导下视觉特征提取的有效性。

不足与展望



局限性

- 行业数据利用较浅（仅 One-Hot 编码），未挖掘深层语义关联。
- 固定融合策略难以适应季节性营销场景变化。



未来方向

- 引入知识图谱，增强品牌 LOGO 变形、赠品组合的语义理解。
- 设计动态权重机制，基于实时数据自动调整模态融合策略。
- 拓展跨模态生成任务（如文案生成商品主图），构建全链路智能系统。



THANK YOU FOR WATCHING

恳请各位老师批评指正

答辩学生：张世雨

指导老师：王薇 王小宁

中国传媒大学广告学院
2021级广告学（计算广告双学士学位复合型人才培养项目）