

大语言模型问答系统的构建 及其在计算广告期刊内容研究中的应用 ——基于《国际品牌观察》语料

答辩人：刘瑞初

指导老师：吴殿义 王小宁

目录

01 选题的背景意义

02 实验框架与内容

03 实验评估与结果

04 论文总结与致谢

01

选题的背景意义

选题的背景意义

选题背景概括

- 1.人工智能技术发展迅速，实现可能性；
大模型在各行各业体现使用价值。
- 2.计算广告实践发展的离不开理论研究支持；
计算广告理论研究领域广阔但在期刊研究方面存在一定空白。

选题意义概括

- 1.技术层面实现了完整的大模型部署方案；
建立了可供参考的文本质量评估体系
- 2.应用层面实现了对文献内容的增强检索、知识库问答和辅助研究功能；

并用于辅助期刊内容研究。

02

实验框架与内容

研究关键词

技术实现

Ollama、DeepSeek-R1和
RAGFlow

——大语言模型部署、问答系
统模型基础和本地知识库与前
端页面

向量知识库、大模型问答、辅
助研究

——增强检索、智能回答、图
表生成

性能评估

Transformers

——问答系统生成文本质量评
估

BERTScore、BARTScore

——文本质量评估方法

Precision (P)、Recall (R)

、F1 Score (F1)

——文本质量评估指标

实际应用

《国际品牌观察》主要研究领
域是什么？

——期刊全景分析

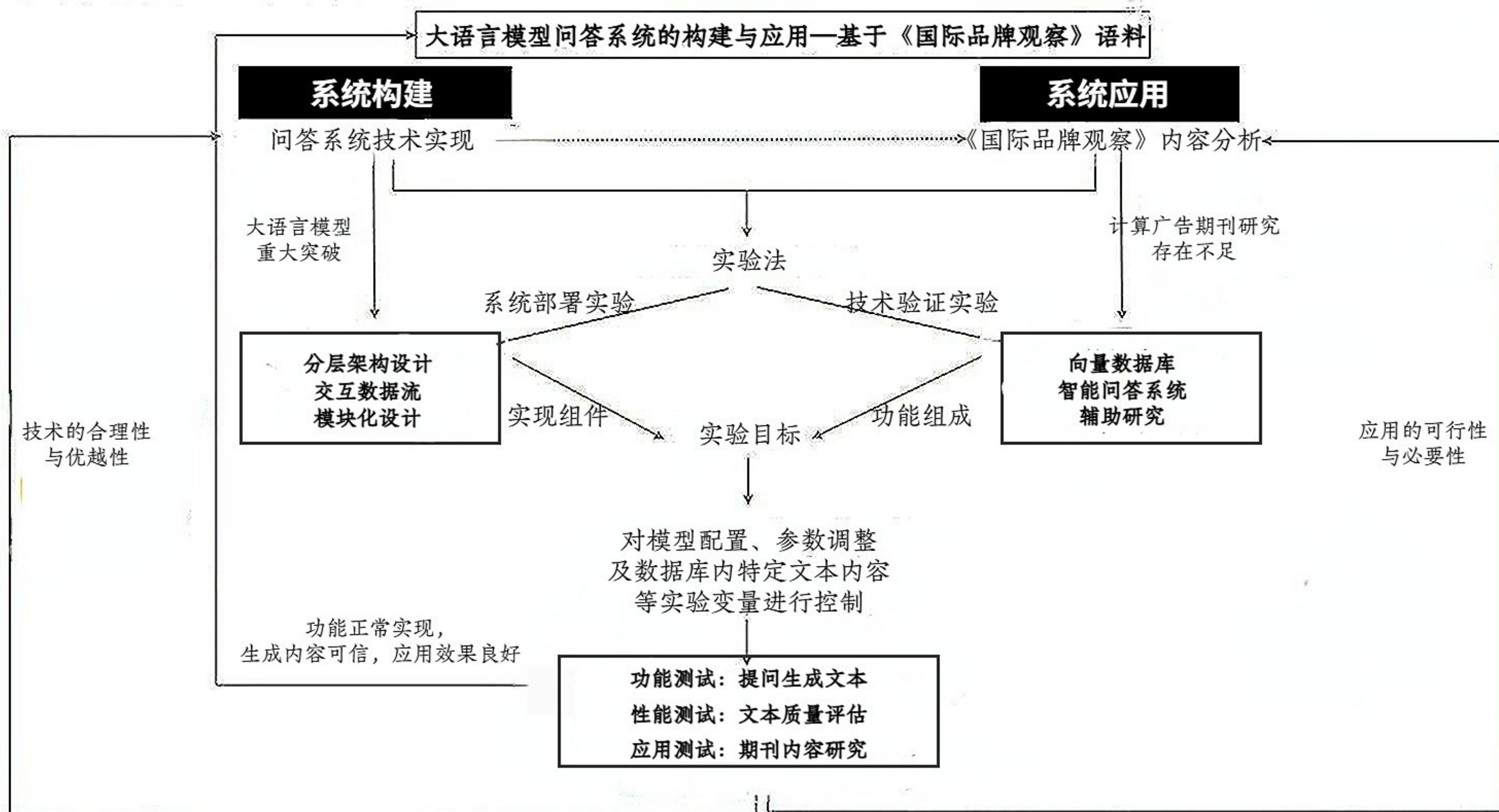
《国际品牌观察》中关于计算
广告概念不同学者的观点及演
变过程？

——期刊局部分析

文本28中作者提出的计算广告
概念及贡献？

——期刊单篇分析

实验框架



实验流程-部署实验

1.Ollama的安装部署



Get up and running with large language models.

Run Llama 3.3, DeepSeek-R1, Phi-4, Mistral, Gemma 2, and other models, locally.

Download ↓

2.Ollama部署DeepSeek模型

```
C:\WINDOWS\s × + v
Microsoft Windows [版本 10.0.26100.3194]
(c) Microsoft Corporation. 保留所有权利。

C:\Users\lrc13>ollama run deepseek-r1:8b
>>> 你好
<think>

</think>

你好！很高兴见到你，有什么我可以帮忙的吗？

>>> Send a message (/? for help)
```

3.Docker安装RAGFlow

```
C:\Windows\Sy: × + v
Microsoft Windows [版本 10.0.26100.3194]
(c) Microsoft Corporation. 保留所有权利。

D:\ragflow\ragflow-main\docker>docker compose -f docker-compose.yml up -d
time="2025-03-09T22:32:39+08:00" level=warning msg="The \"HE_ENDPOINT\" variable is not set. Defaulting to a blank string."
time="2025-03-09T22:32:39+08:00" level=warning msg="The \"MACOS\" variable is not set. Defaulting to a blank string."
[+] Running 5/5
✔ Container ragflow-mysql Healthy
✔ Container ragflow-server Running
✔ Container ragflow-minio Running
✔ Container ragflow-redis Running
✔ Container ragflow-es-01 Running

D:\ragflow\ragflow-main\docker>
```

4.RAGFlow注册登录

登录

很高兴再次见到您！

* 邮箱

lrc130322@163.com

* 密码

请输入密码

请输入密码

☐ 记住我

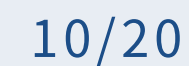
没有帐户? [注册](#)

登录



实验内容-模块实现一





实验内容-模块实现三

辅助研究功能实现图



03

实验评估与结果

实验评估-验证实验一

功能测试实现图

功能测试系统

功能测试

性能测试

应用测试

测试方法: 错误推测法 (黑盒测试)

- 基于测试人员经验预测系统可能存在的问题
- 设计测试用例通过实际操作验证系统功能

基础与专业问答测试

测试用例:

基础事实型问答

概念解释型问答

测试结果:

提供准确的事实信息与概念解释

分析类问答测试

测试用例:

比较分析型问答

热点探讨型问答

测试结果:

提供多角度比较分析与研究热点识别

系统交互测试

测试用例:

多轮对话连贯性

测试结果:

保持上下文连贯，理解指代关系

实验评估-验证实验二

应用测试实现图



实验评估-验证实验三

性能测试实现图

性能测试系统

功能测试

性能测试

应用测试

测试方法: 单元测试法 (白盒测试)

- 基于Transformers库中的AutoModelForSeq2SeqLM组件
- 使用预训练模型评估系统生成文本的质量

BERTScore评估

评估指标:

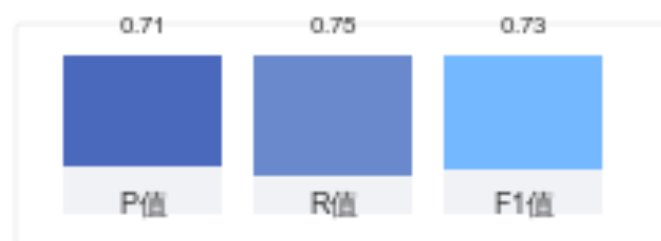
精确度(P): 生成文本与参考文本的匹配度

召回率(R): 参考文本内容被生成文本覆盖度

F1值: P和R的调和平均

该方法基于词嵌入空间中的语义相似性评估

测试结果:



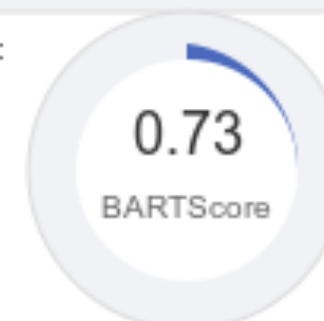
BARTScore评估

评估原理:

- 基于条件生成概率的评估方法
- 考虑文本流畅性、连贯性和信息完整性
- 评估生成内容的实用性和可读性

该方法通过BART模型计算文本编码后的相似度

测试结果:



实验评估-质量实验

1.确定评估的摘要文本和生成文本

文本2 (Text 2)	包括广告营销在内的越来越多的社会科学领域对信息、数据、算法越来越依赖，甚至已经从根本上将之视为学科持续发展和不断创新的最核心要素。事实上，计算广告的新范式也是如此。	文章阐述了广告业正经历从传统"魔法艺术"向"计算广告学范式"的根本转变。这一新范式不是否定经典广告理论，而是将其与数据、算法和人工智能有机融合，使广告从艺术形式转变为科学。计算广告已超越程序化投放，扩展至各媒体形式和广告环节，推动品牌广告与效果广告从对立走向"品效合一"。未来广告将从"广告即传播"向"广告即服务"转变，其核心目标是通过算法实现产品供给与用户需求的精准匹配，个性化的用户需求也将由个性化的算法服务来满足。
文本9 (Text 9)	在大数据时代，计算是所有的广告活动所必须具备的典型特征和关键能力，任何广告活动和任何广告活动的广告环节，都应该被数据要素、计算能力和算法工具改造、重构甚至颠覆。	文章指出计算广告就是广告本身，在大数据时代，所有广告活动都需要计算能力，不仅仅是互联网广告。计算不应被视为某种特定广告形式或环节，而是当前广告业必须具备的核心特征。广告本质上就包含着计算观念，只是在大数据出现后，计算从传统形式升级为大数据计算。文章认为广告学正在从经典范式向计算广告学范式转换，这不是对旧范式的否定，而是一种演进。应对这种转变既不能过度乐观认为经典广告已死，也不能轻蔑地认为旧方法仍能解决新问题。通过数据思维和计算能力重构广告实践和教研，广告将从魔法艺术转变为科学。

2.加载 BERT、BART 模型及其分

```
class TextScorer:
    def __init__(self):
        # BERTScore
        self.bertscore = score
        # BARTScore
        self.bart_model_name = "fnlp/bart-base-chinese"
        self.bart_tokenizer = (AutoTokenizer.from_pretrained
                                (self.bart_model_name))
        self.bart_model = (AutoModelForSeq2SeqLM.from_pretrained
                            (self.bart_model_name))
        self.bart_model.eval()
        if torch.cuda.is_available():
            self.bart_model.to('cuda')
```

3.定义bartscore、bertscore函数

```
def calculate_bertscore(self, cand, refs, lang="zh"):
    P, R, F1 = self.bertscore(cand, refs, lang=lang, verbose=True)
    return P.mean().item(), R.mean().item(), F1.mean().item()

def calculate_bartscore(self, generated_text, reference_text):
    inputs = self.bart_tokenizer([generated_text, reference_text],
                                  return_tensors='pt', padding=True, truncation=True)
    if torch.cuda.is_available():
        inputs = {k: v.to('cuda') for k, v in inputs.items()}
    if "token_type_ids" in inputs:
        del inputs["token_type_ids"]
    with torch.no_grad():
        outputs = self.bart_model(**inputs, output_hidden_states=True)
        last_hidden_state = outputs.encoder_last_hidden_state
        generated_embedding = last_hidden_state[0].mean(dim=0)
        reference_embedding = last_hidden_state[1].mean(dim=0)
    score = (torch.nn.functional.cosine_similarity
              (generated_embedding.unsqueeze(0), reference_embedding.unsqueeze(0)).item())
    return score
```

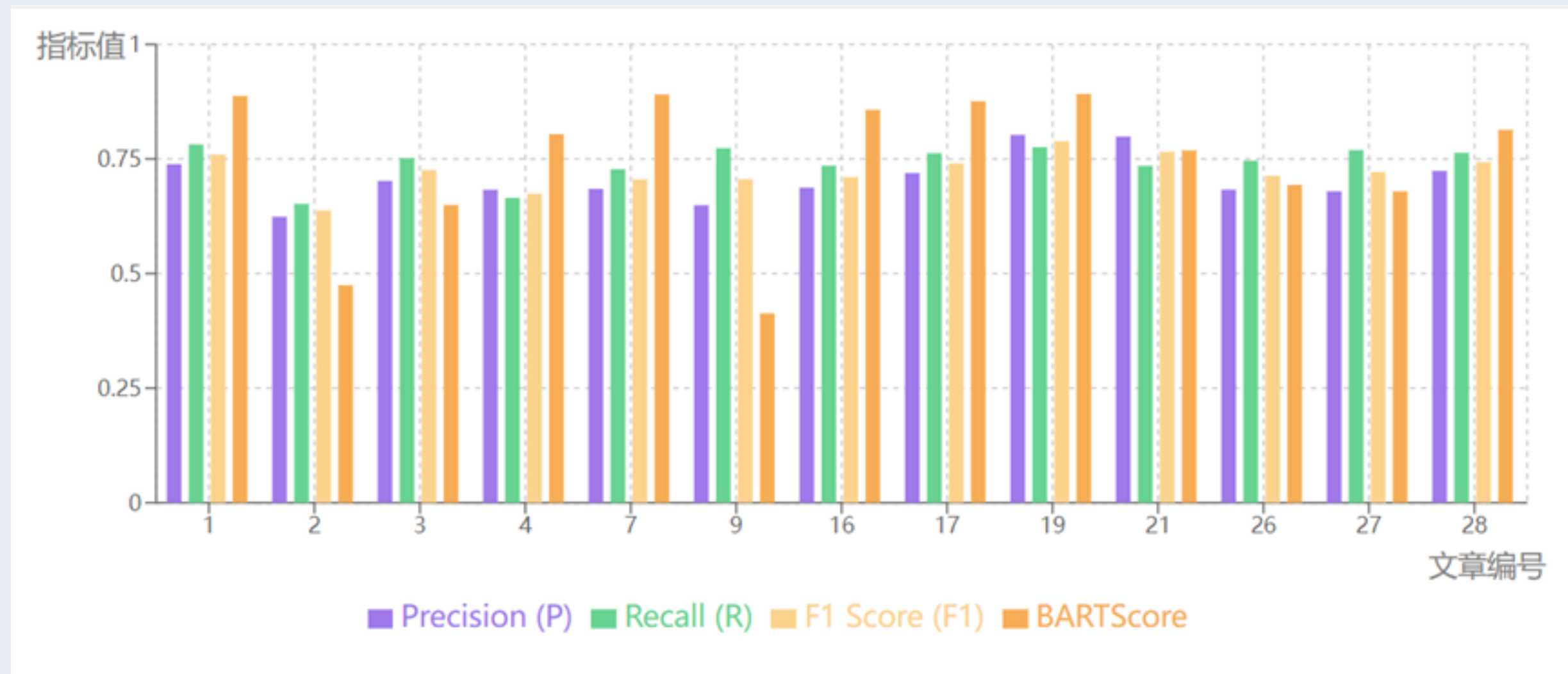
4.计算生成文本P值、R值和F1值

```
def main(generated_text, reference_text):
    scorer = TextScorer()

    # 计算 BERTScore
    P, R, F1 = (scorer.calculate_bertscore
                 (cand: [generated_text], refs: [reference_text]))
    print(f"BERTScore - Precision (P): {P:.4f}")
    print(f"BERTScore - Recall (R): {R:.4f}")
    print(f"BERTScore - F1 Score (F1): {F1:.4f}")

    # 计算 BARTScore
    bart_score = (scorer.calculate_bartscore
                  (generated_text, reference_text))
    print(f"\nBARTScore: {bart_score:.4f}")
```

实验结果-质量实验



$$P_{\text{BERT}} = \frac{1}{|\text{gen}|} \sum_{x_i \in \text{gen}} \max_{y_j \in \text{ref}} x_i^T y_j$$

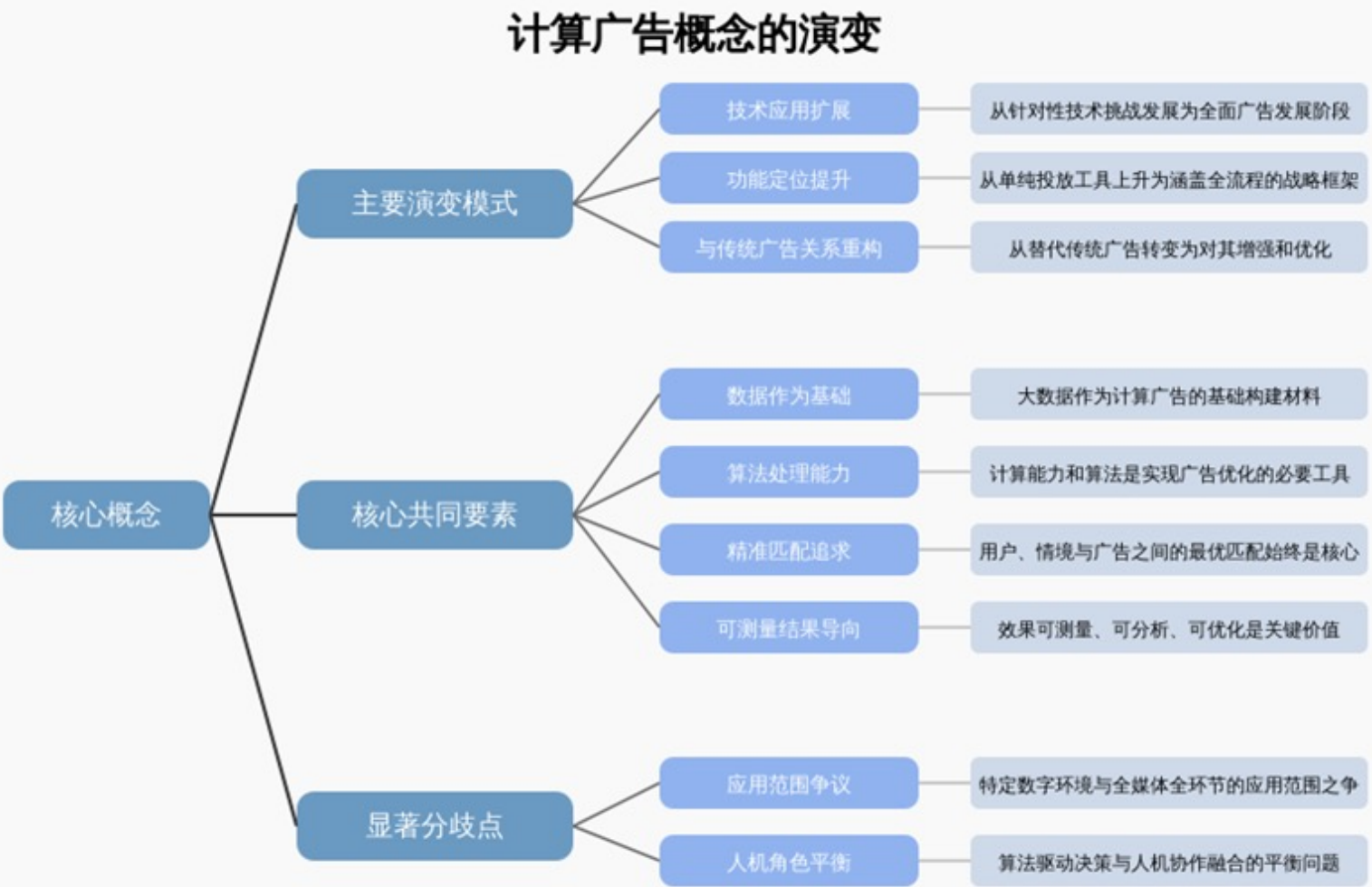
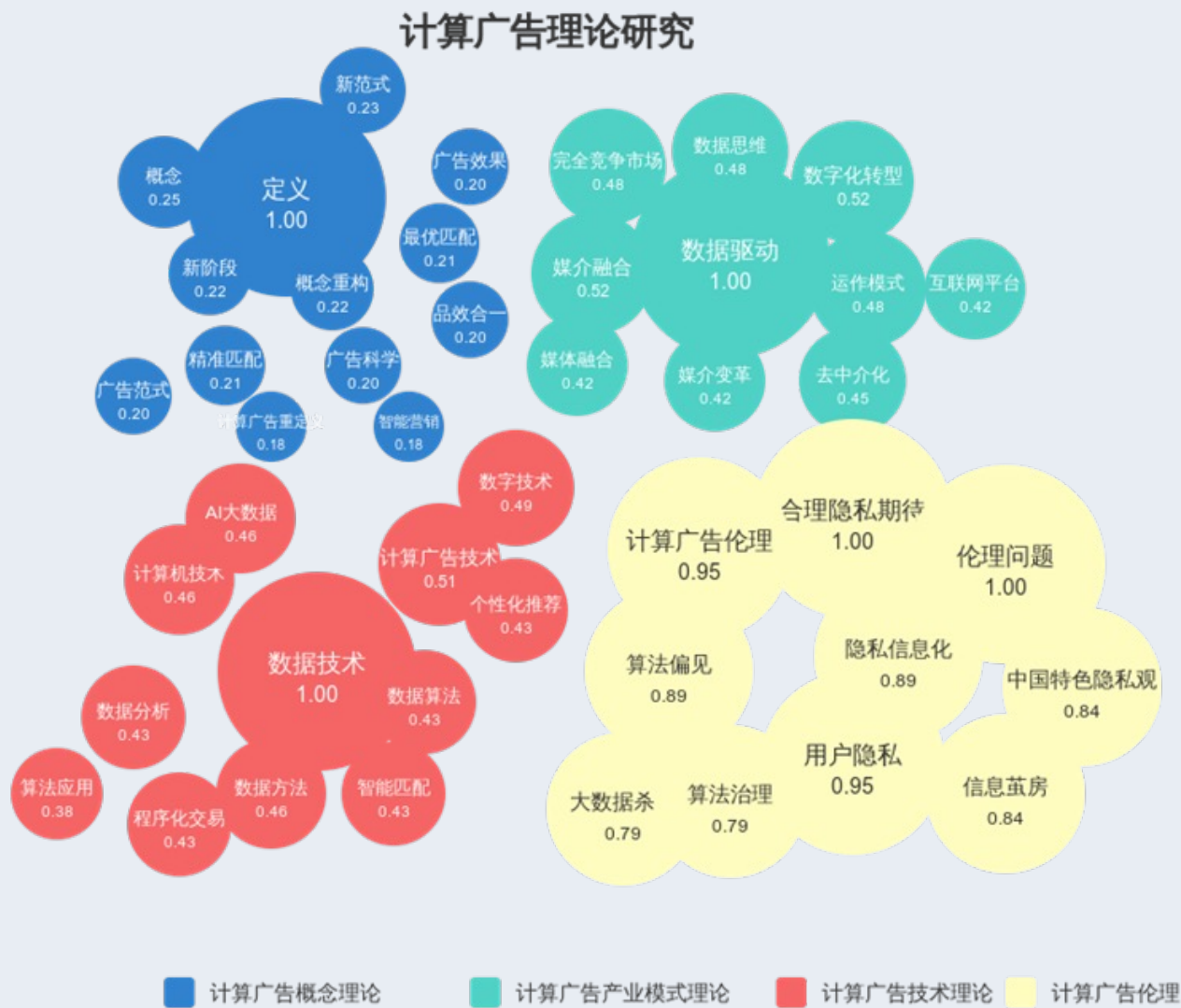
$$R_{\text{BERT}} = \frac{1}{|\text{ref}|} \sum_{y_j \in \text{ref}} \max_{x_i \in \text{gen}} x_i^T y_j$$

$$F1_{\text{BERT}} = \frac{2 \cdot (P_{\text{BERT}} \cdot R_{\text{BERT}})}{P_{\text{BERT}} + R_{\text{BERT}}}$$

$$\text{BARTScore} = \sum_{t=1}^n \log P(y_t | y_{<t}, x, \theta_{\text{BART}})$$

系统测试显示，生成的文本摘要在BERTScore的精确度、召回率和F1值三项指标上平均达到0.7以上，BARTScore平均达到0.73，证实了系统具备良好的语义理解能力和文本生成质量。

实验结果-辅助实验



通过对《国际品牌观察》期刊多层次、立体的文本分析，厘清了该期刊在计算广告领域不同层次的研究内容与研究现状，可以作为辅助研究工具进一步应用于更具体更实际的场景。

04

论文总结与致谢

论文总结

- 1.技术实现角度上，本研究成功构建了一套基于本地部署的大语言模型问答系统，实现了用户管理、知识库管理、大模型问答和辅助研究功能；
- 2.实验结果表明，本问答系统能够有效识别和提取计算广告期刊文献中的关键信息和研究内容，且生成内容通过评估检验，质量良好；
- 3.通过案例分析，本研究发现计算广告期刊研究内容主要集中在理论概念界定、技术应用和学科建设三个方面，且呈现出从技术导向全链路多领域转变的明显趋势。



致 谢

感谢指导老师们的耐心指导和无私奉献！ ☒☒☒

感谢评委老师们百忙之中的参与和耐心聆听！ ☒☒☒

感谢各位！ ！ ！ ☒☒☒