

# “信息引航者”——基于 MGCA 的新闻推荐系统

## 摘要

信息技术不断发展，互联网上存在着大量信息，通过小红书、今日头条、bilibili 等互联网平台获取信息是主要的信息获取途径之一。随着互联网上信息的越来越多，想要在海量信息中获取我们感兴趣或有需求的信息愈加困难。

推荐系统的出现正是为了解决这种“信息过载”的问题。推荐系统已在互联网中得到了广泛的应用，并给应用它的企业带来了丰厚的利润。推荐系统给亚马逊带来了 35% 的销售收入，给 Netflix 带来了高达 75% 的消费，并且 Youtube 主页上 60% 的浏览来自推荐服务。有关推荐系统的研究具有十分深远的意义与巨大的实用价值。

# 目录

一、引言	3
1.1 研究背景及意义	3
1.1.1 信息过载问题的解决迫在眉睫	3
1.1.2 互联网与推荐系统的融合带来巨大的经济价值	3
1.1.3 响应国家发展和改革委员会“新基建”号召	4
1.2 相关研究现状	5
1.2.1 推荐系统在工业界的应用现状	5
1.2.2 推荐系统在学术界的研究现状	5
1.3 项目架构总览	5
1.3.1 项目整体框架	5
1.3.2 项目优势与创新性	5
二、模型技术路线及实现方案	6
2.1 数据集选取	6
2.1.1 MIND 数据集	6
2.1.2 glove 词嵌入	6
2.1.3 数据预处理	6
2.2 方法介绍	6
2.2.1 图神经网络	6
2.2.2 注意力机制	6
2.2.3 Transformer	6
2.2.4 Fastformer	6
2.3 模型架构	6
2.3.1 总体架构	6
2.3.2 候选新闻编码	6
2.3.3 单词粒度感知候选新闻	6
2.3.4 新闻粒度感知候选新闻	6
2.3.5 实体粒度感知候选新闻	6
2.4 模型评估	6
2.4.1 与之前模型的性能对比	6
2.4.2 消融实验	6

<b>三、软件工程与开发架构方案</b>	<b>7</b>
3.1 开发工具与框架	7
3.2 新闻推荐系统软件落地示例	7
3.2.1 软件周期模型	7
3.2.2 软件开发模型	9
3.2.3 数据库设计	9
3.2.4 UML 设计	9
3.2.5 软件功能演示	9
3.2.6 兼容性测试与稳定性测试	9
<b>四、商业模式构建与经营管理</b>	<b>11</b>
4.1 市场竞争分析	11
4.2 投资回报分析	11
4.3 盈利方式	11
4.3.1 提供整套低耦合工业界解决方案	11
4.3.2 提供整套成熟软件系统	11
4.4 财务管理	11
4.5 营销战略	11
<b>参考文献</b>	<b>12</b>

# 一、引言

## 1.1 研究背景及意义

### 1.1.1 信息过载问题的解决迫在眉睫

互联网高速发展, 信息呈爆炸式增长, 用户逐渐由信息匮乏时代迈入了信息过载时代——过量信息反而使得用户无法找到自己需要的信息。信息过载问题会产生一系列问题, 比如用户体验变差、用户数量骤减、用户粘性降低等。信息爆炸的今天, 个性化新闻推荐技术已经变成了许多新闻网站和 App 的关键技术。新闻的数据来源众多, 可能一分钟就有成千上万条新的数据产生, 在数据量激增的情况下, 需要付出更多的人力成本, 并且人工处理速度慢, 效率十分低下。一套个性化新闻推荐系统刚好可以应用于有效缓解这种信息过载问题。如图 1 是新闻推荐系统通过个性化推荐实现精准化投放信息的流程示意。

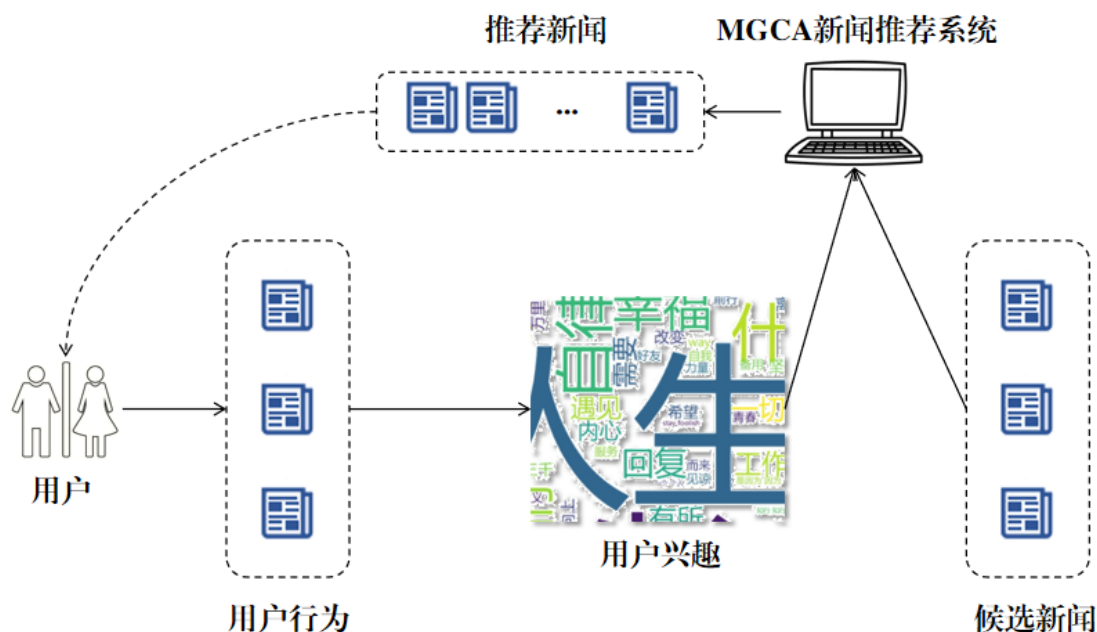


图 1 个性化新闻推荐系统实现精准化投放信息

### 1.1.2 互联网与推荐系统的融合带来巨大的经济价值

目前互联网变现的主要方式均可以与推荐系统融合起来, 创造巨大的经济价值, 如图 2 展示了推荐系统与互联网融合时创造价值的几种常见途径:

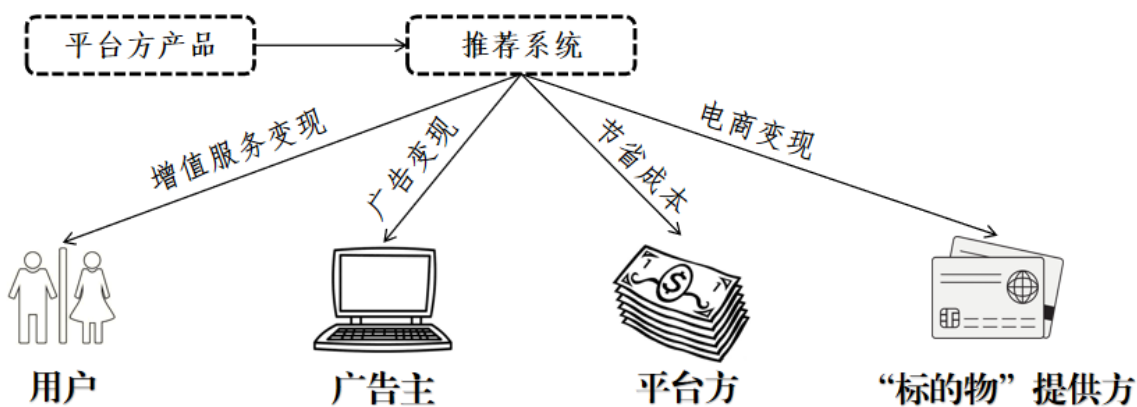


图 2 推荐系统与互联网融合体现的经济价值

★ **增值服务变现：**增值服务主要是指基本服务以外的特色服务，如新闻平台常见的会员制。推荐系统通过精准把握用户兴趣，引导用户购买会员增值服务产品。推荐系统的个性化推荐服务不仅可以更加深层次的挖掘用户需求，还可以使更多相对冷门但优质的内容得到曝光与展示。

★ **广告变现：**用户方产品在互联网上投放广告，通过提高广告曝光以及用户点击广告获取权益。通过推荐系统的应用，可以在新闻 APP 中实现广告精准投放。广告业务偏轻公司一般都会选择流量外包方式做广告变现，通过新闻推荐系统提高广告精准投放率与用户点击率即是一种非常好的方式。

★ **节省成本变现：**推荐系统对平台方的贡献不仅在于直接的商业价值或是促进用户粘性、用户数量增长等，还可以用更少的成本实现最大程度的个性化新闻推荐，提高内容的分发效率。

★ **电商变现：**无论是传统意义的实体电商产品，还是网络小说、网络课程等虚拟电商产品，都可以通过提高分发商品效率，精准化投放促进商品售卖而提高商家分成，吸引商家入驻。总的来说，在电商变现方面，新闻推荐系统可以做到促进“标的物”提供方的生态逐步健全繁荣，获取更多经济效益。

### 1.1.3 响应国家发展和改革委员会“新基建”号召

国家发改委于 2020 年 4 月首次明确了新型基础设施的范围包括以人工智能、云计算、区块链等为代表的新技术基础设施，新型基础设施建设对于打造经济新动能、重塑供应新链条、促进惠民新模式有重大意义，是当前稳经济、促增长的核心支撑。人工智能、云计算、物联网等现代信息科技依托新型基础设施建设得到了高速发展。

特别是处于疫情后时代，线下娱乐受到重创，线上娱乐成为青年群体最重要的娱乐方式之一。在当前的时代背景下，新闻推荐的研究有助于促进青年群体的幸福感，扩大新型基础设施建设的有效投资，有效对冲经济下行压力，有力支撑经济社会高质量恢复与持续发展。

## **1.2 相关研究现状**

### **1.2.1 推荐系统在工业界的应用现状**

### **1.2.2 推荐系统在学术界的研究现状**

## **1.3 项目架构总览**

### **1.3.1 项目整体框架**

### **1.3.2 项目优势与创新性**

## 二、模型技术路线及实现方案

### 2.1 数据集选取

#### 2.1.1 MIND 数据集

#### 2.1.2 glove 词嵌入

#### 2.1.3 数据预处理

### 2.2 方法介绍

#### 2.2.1 图神经网络

#### 2.2.2 注意力机制

#### 2.2.3 Transformer

#### 2.2.4 Fastformer

### 2.3 模型架构

#### 2.3.1 总体架构

#### 2.3.2 候选新闻编码

#### 2.3.3 单词粒度感知候选新闻

#### 2.3.4 新闻粒度感知候选新闻

#### 2.3.5 实体粒度感知候选新闻

### 2.4 模型评估

#### 2.4.1 与之前模型的性能对比

#### 2.4.2 消融实验

## 三、 软件工程与开发架构方案

### 3.1 开发工具与框架

系统开发操作系统环境：Windows11

数据库：MySQL

开发语言：JAVA，Python

JDK：jdk1.8

Python：3.6

开发平台：IntelliJ IDEA 2022.1.3

前端框架：Vue3，bootstrap

后端框架：springboot

### 3.2 新闻推荐系统软件落地示例

#### 3.2.1 软件周期模型

在软件周期模型阶段，我们将采用经典的软件开发周期，包括可行性分析、需求分析、设计、开发、测试和部署等阶段。在可行性分析阶段，我们将对新闻网站个性化推荐系统的可行性进行评估，以确定项目是否值得投入，并识别潜在的风险和挑战。在需求分析阶段，我们将利用敏捷开发方法论，通过迭代方式逐步完善前端与后端的功能。在设计阶段，我们将着重考虑系统架构和界面设计，确保软件具备良好的可扩展性和用户体验。在开发阶段，我们将按照设计方案进行编码和实现，并且进行代码审查以确保代码质量。在测试阶段，我们将进行单元测试、集成测试和系统测试，以发现和修复可能存在的问题。最后，在部署阶段，我们将准备好软件的发布版本，并进行部署和配置，确保软件能够正常运行并满足用户需求。

**1. 可行性分析阶段** 项目面临着一些技术挑战，包括评估网站新闻个性化推荐算法的复杂性和实现难度，以及与多个数据源对接的挑战。为了应对这些难题，我们将记录用户历史点击新闻，并将其与候选新闻输入到算法中，通过智能计算得到匹配度，以优先显示匹配度较高的新闻。市场调研表明，在当前信息过载的环境下，一套个性化新闻推荐系统更符合市场需求且具有竞争优势。在成本估算方面，我们将综合考虑项目开发、运维、推广等各方面的成本，包括硬件、软件、人力等，以确定项目值得投入的程度。

**2. 需求分析阶段** 随着互联网的高速发展，信息量呈爆炸式增长，用户逐渐由信息匮乏时代迈入信息过载时代。在这样的背景下，我们的目标是建立一个新闻网站，通过个性化推荐系统帮助用户从海量信息中快速准确地找到感兴趣的内容，提高用户体验、用户数量和用户粘性。



★ 个性化推荐：根据用户的历史浏览记录、点击行为、点赞收藏等个性化信息，推荐用户感兴趣的新闻内容。

★ 实时更新：系统需要能够实时从各个新闻源获取最新的新闻内容，并将其纳入推荐范围。

★ 多样化推荐算法：结合协同过滤、内容分析、深度学习等技术，提供多种推荐算法，确保推荐结果的准确性和多样性。

★ 用户反馈机制：提供用户反馈功能，用户可以对推荐结果进行点赞、收藏、举报等操作，以优化推荐算法。

★ 个人信息管理：提供用户个人信息管理功能，用户可以编辑个人兴趣标签、设置推送频率等。

★ 跨平台兼容性：支持多平台访问，包括 PC 端、移动端和平板电脑端，保证用户在不同设备上的使用体验一致。

**3. 设计阶段** 新闻推荐系统分为用户管理模块、新闻管理模块、数据处理模块和推荐算法模块，用户管理模块主要负责对用户信息的管理，用户管理模块是新闻推荐系统中的重要功能模块之一，主要负责对用户信息的管理，包括用户的注册、登录、用户信息管理等功能。新闻管理模块包括新闻的采集、存储以及新闻的更新管理等功能。数据处理模块在处理新闻内容以及用户行为产生的数据后，将会对数据进行数据清洗和分析，处理后的结果将会存储在数据库中。推荐算法构建推荐系统模块在已有数据的基础上构建推荐算法，基于数据进行分析，根据推荐算法得出对目标用户的推荐结果，并向用户推送。

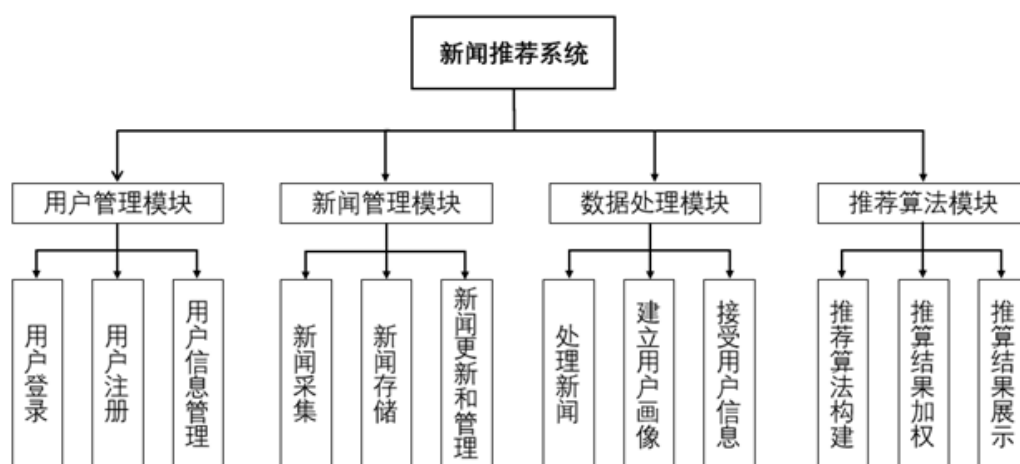


图 3 系统功能模块图

**4. 测试阶段** 在测试阶段，我们将进行系统的测试以确保其质量和稳定性，通过单元测试、集成测试和系统测试对系统进行正向测试。通过漏洞扫描对系统进行逆向测试，保

障系统的安全性，通过授权和认证测试，测试系统的用户授权和认证机制。

- ★ 单元测试：对系统中的各个单元（函数、模块）进行测试，验证其功能的正确性。

- ★ 集成测试：将系统中各个单元集成在一起，测试其之间的交互和整体功能。

- ★ 系统测试：对整个系统进行端到端的测试，模拟用户使用场景，验证系统的功能和性能。

- ★ 漏洞扫描：使用安全扫描工具对系统进行扫描，发现并修复潜在的安全漏洞。

- ★ 授权和认证测试：测试系统的用户授权和认证机制，确保用户的身份和权限受到正确的控制。

### 3.2.2 软件开发模型

### 3.2.3 数据库设计

### 3.2.4 UML 设计

### 3.2.5 软件功能演示

### 3.2.6 兼容性测试与稳定性测试

- ★ 可执行测试：分别部署到不同性能的电脑上运行测试，均成功运行，网页反馈均在 35ms 到 45ms 左右，不存在延迟和卡顿现象，该系统对电脑的配置要求较低。

- ★ 功能测试：所有功能均可正常运行使用，可视化动态大屏显示正常，首页展示效果正常，查询功能正常，链接页面跳转正常。

- ★ 兼容测试：在 Windows、Linux、Mac 等操作系统可以正常访问。谷歌浏览器、火狐浏览器、QQ 浏览器、360 浏览器等运行均正常。Windows 操作系统的 python 版本为 3.6 以上、tensorflow2.14.0 版本环境可以正常运行平台。

- ★ 安全测试：本项目采取的是 https 协议，安全性和可靠性比较高，jupyter 环境提供了三种安全级服务配置，可以按照实际的需求提升安全等级。

## 四、 商业模式构建与经营管理

### 4.1 市场竞争分析

### 4.2 投资回报分析

### 4.3 盈利方式

#### 4.3.1 提供整套低耦合工业界解决方案

#### 4.3.2 提供整套成熟软件系统

### 4.4 财务管理

### 4.5 营销战略

## 参考文献

- [1] 陈昌凤, 袁雨晴. 智能新闻业: 生成式人工智能成为基础设施 [J]. 内蒙古社会科学, 2024, 45(01): 40-48. DOI: 10.14137/j.cnki.issn1003-5281.2024.01.006.