

# 上海杉达学院本科生毕业论文（设计）开题报告

生成时间：2025 年 11 月 8 日

学生姓名	宋子杰	学号	B24015220	班级	B240152
院系	信息科学与技术学院		专业	软件工程	
课题题目	基于主题模型的诗歌推荐系统的设计与实现				
课题性质	设计		课题来源	其他	
指导教师	尹小花		职称		
课题目的与意义	<p>在数字化洪流席卷全球、社会生活节奏日益加速的时代背景下，现代人普遍面临着巨大的精神压力与内心的迷茫。诗歌作为人类情感与智慧的结晶，自古以来便是我们寻求精神慰藉、发现自我、甚至重塑人生理想的宝贵路径。它以凝练的文字承载着深刻的意境与人文关怀，是连接个体心绪与历史文化的桥梁。然而，海量诗歌资源的数字化普及，反而加剧了读者的选择困境，使得那些契合读者当下心绪、蕴含独特艺术价值的佳作，难以被高效地发现和欣赏。因此，构建一个能够深度理解用户审美偏好、精准捕捉情感需求的个性化诗歌推荐系统，具有重要的应用价值和人文意义。然而，现有的传统诗歌推荐系统多采用基于协同过滤（Collaborative Filtering, CF）的推荐算法，该算法主要依赖用户的历史评分数据（如1-5星打分）来计算用户间的相似度。然而，在实际应用中，用户往往存在“只读不评”或“评分稀疏”的现象，导致评分矩阵极度稀疏，系统难以找到相似用户（即“数据稀疏性”问题）。此外，单一的数字评分（Scalar Rating）无法反映用户复杂的审美偏好——例如，两个用户都给《静夜思》打了5分，用户A是因为喜欢“思乡”的情感，而用户B是因为喜欢“月亮”的意象，传统算法无法区分这种深层动机的差异。</p> <p>本课题旨在解决上述“评分稀疏”与“语义理解缺失”的难题，构建一个基于Flask与Vue的前后端分离诗歌推荐系统。核心思路是引入LDA(Latent Dirichlet Allocation)主题模型作为“语义翻译官”，挖掘用户评论文本中隐含的主题信息（如情感基调、意象风格）。具体而言，本研究将利用LDA模型分析用户的文本评论，提取出“边塞”、“田园”、“悲秋”等潜在主题概率分布，以此量化用户的深层偏好。随后，通过改进传统的余弦相似度计算公式，将单一的评分相似度扩展为“评分+主题”的综合相似度计算。这就好比在寻找书友时，不仅看“是否打过分”，更看“是否聊得来”，从而在数据稀疏的情况下也能精准定位兴趣相投的“邻居”用户。</p>				

研究方案和 主要内容	<p>本课题拟构建一个 B/S 架构的诗歌个性化推荐系统，前端采用 Vue.js 框架构建交互界面，后端采用 Python Flask 框架处理业务逻辑与算法模型。</p> <p><b>系统应用逻辑与用户流程设计：</b></p> <p>用户交互层（Vue 前端）：用户注册登录后，可浏览、搜索不同朝代与流派的诗歌。用户在阅读过程中，可对诗歌进行评分（1-5 星）或发表文本评论（如“这首诗的离愁别绪很感人”）。</p> <p>数据传输层（API 交互）：前端通过 Axios 向 Flask 后端发送用户的行为数据（评分、评论），后端将数据存入数据库，并实时响应前端的推荐列表请求。</p> <p>核心推荐层（Flask 后端算法）：系统定期触发 LDA 主题模型训练任务。当用户产生新的评论数据时，系统提取其中的主题特征（如“边塞”、“田园”、“悲秋”），结合协同过滤算法更新用户的推荐列表。</p> <p><b>推荐机制运作逻辑（LDA-CF 算法实现）：</b></p> <p>数据预处理：利用 Jieba 分词对诗歌内容及用户评论进行清洗，去除停用词，构建语料库。</p> <p>LDA 主题建模：将每首诗及其评论集视为“文档”，利用 LDA 模型挖掘隐含的 <math>K</math> 个主题（如意象、情感基调）。生成“文档-主题”概率分布矩阵，从而量化用户对特定主题（如“爱国”、“思乡”）的偏好程度。</p> <p>用户相似度计算：改进传统的余弦相似度计算公式。不仅计算用户对诗歌评分的相似度，还结合用户在 LDA 主题层面上的偏好相似度。如果两个用户都频繁评论“边塞”主题的诗歌且情感倾向一致，则判定为高相似度邻居。</p> <p>预测与推荐：根据 Top-N 相似用户的评分记录，预测目标用户对未读诗歌的评分，将预测分最高的 Top-K 首诗歌推送至前端“猜你喜欢”模块。</p>
初拟 论文提纲	<p>摘要</p> <p>ABSTRACT</p> <p>第一章 绪论</p> <p>1.1 项目的背景意义</p> <p>1.2 国内外研究现状</p> <p>1.3 本文的主要工作和章节安排</p> <p>第二章 主要技术</p> <p>2.1 推荐算法理论</p> <p>2.2 协同过滤（CF）算法原理及局限性</p> <p>2.3 LDA 主题模型技术</p> <p>2.4 本章小结</p> <p>第三章 需求分析</p> <p>3.1 功能需求分析</p> <p>3.2 非功能需求分析</p> <p>3.3 本章小结</p> <p>第四章 系统设计</p> <p>4.1 系统总体架构设计</p> <p>4.2 算法总体设计思路与流程</p> <p>4.3 诗歌评论数据的预处理与 LDA 主题提取</p> <p>4.4 融合主题偏好的改进型用户相似度计算</p> <p>4.5 基于 Top-N 邻域的评分预测与推荐生成</p> <p>4.6 数据库设计（用户、诗歌、评论、评分表结构）</p>

	<p>4.7 本章小结</p> <p>第五章 系统实现</p> <p>5.1 Web 后端 Flask 实现</p> <p>5.2 Web 前端 Vue 实现</p> <p>5.3 推荐功能核心代码实现</p> <p>5.4 本章小结</p> <p>第六章 系统测试</p> <p>6.1 测试环境（方法）</p> <p>6.2 测试用例</p> <p>6.3 测试结果</p> <p>6.4 本章小结</p> <p>第七章 总结与展望</p> <p>7.1 总结</p> <p>7.2 不足与展望</p> <p>致谢</p> <p>参考文献</p> <p>附录 1</p> <p>附录 2</p>
<b>主要 参考资料</b>	<p>[1] 张宇, 吴静. 基于 LDA 主题模型的协同过滤推荐算法 [J]. 智能计算机与应用, 2024, 14(02) : 190–194.</p> <p>[2] 祁紫冉. 混合协同过滤算法及其在图书推荐中的应用研究 [D]. 河北经贸大学, 2024. DOI:10.27106/d.cnki.ghbju.2024.000482.</p> <p>[3] 姜威, 于琦, 于彤, 等. 基于 LDA 模型的中医药新闻主题识别及其热点演化分析 [J]. 中国中医药图书情报杂志, 2025, 49(06) : 112–117.</p> <p>[4] 王睿. 基于协同过滤算法的图书推荐系统研究与设计 [D]. 沈阳理工大学, 2023. DOI:10.27323/d.cnki.gsgyc.2023.000168.</p> <p>[5] 杜梦晗, 崔仙姬, 姜雨蒙, 等. 基于 LDA 主题模型的 MOOC 推荐方法研究 [J]. 计算机与数字工程, 2024, 52(12) : 3616–3622+3760.</p> <p>[6] 杜宇灏, 林晓霞, 李环宇, 等. LDA 算法在微博热搜可视化中的应用 [J]. 福建电脑, 2023, 39(12) : 15–19. DOI:10.16707/j.cnki.fjpc.2023.12.003. [7] Uma Devi G. Sentiment Analysis with Python: A Hands-on Approach [Z]. Magestic TECnology Solutions (P) Ltd, 2023(2023–02–12).</p> <p>[8] Unpingco J. Python Programming for Data Analysis [M]. Springer International Publishing, 2021.</p> <p>[9] Cansdale A. How exploring Python can level up your data Visualization [J]. The Biochemist, 2021, 43(5) : 4–7.</p> <p>[10] 罗茶连. 基于意象词特征的诗歌情感分类方法研究 [D]. 厦门大学, 2021. DOI:10.27424/d.cnki.gxmdu.2021.002496.</p> <p>[11] 于海涛, 刘竞泽, 唐尧杰. 基于情感词典的民宿评论情感分析——以桂林阳朔为例 [J]. 重庆交通大学学报(社会科学版), 2025, 25(02) : 80–89.</p> <p>[12] 胡海莹, 陈学文. 基于作者的协同过滤图书推荐算法 [J]. 现代信息科技, 2024, 8(18) : 108–111+115. DOI:10.19850/j.cnki.2096–4706.2024.18.02</p>

	<p>2.</p> <p>[13] 黄勃, 严非凡, 张昊, 等. 推荐系统研究进展与应用[J]. 武汉大学学报(理学版), 2021, 67(06) : 503–516. DOI:10.14188/j.1671-8836.2021.1001.</p> <p>[14] 刘炜, 叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报, 2017, 43(05) : 32–41. DOI:10.13530/j.cnki.jlis.170020.</p> <p>[15] 惠调艳, 王智, 何振华, 等. 基于词典-TextCNN-Word2Vec 组合模型的在线评价细粒度情感分析[J]. 情报理论与实践, 2025, 48(02) : 168–177. DOI:10.16353/j.cnki.1000-7490.2025.02.018.</p>
进度安排	<p>2025 年 11 月 29 日-2026 年 12 月 5 日 进行 LDA 主题分析</p> <p>2025 年 12 月 6 日-2026 年 12 月 15 日 进行推荐算法设计</p> <p>2026 年 12 月 16 日-1 月 4 日 完成系统设计</p> <p>2026 年 1 月 5 日-3 月 1 日 完成论文初稿</p> <p>2026 年 3 月 2 日-4 月 1 日 进行论文修改和定稿。</p> <p>2026 年 4 月 2 日-4 月 30 日 准备论文答辩</p>
指导教师 审核意见	<p style="text-align: right;">指导教师 : _____</p> <p style="text-align: right;">年   月   日</p>
专业负责人 审核意见	<p style="text-align: right;">专业负责人: _____</p> <p style="text-align: right;">年   月   日</p>

学生: \_\_\_\_\_

年   月   日

# 毕业设计（论文）开题报告要求

## 题目基本信息

应具体明确，不要太长（10-25字），如：“基于 JEE 的人事管理网站开发与实现”。

## 内容详情

### 一、课题目的与意义（一般 200-300 字）

针对国内外计算机学科或计算机行业发展过程中需要解决的有关技术问题或者在社会经济发展过程需要解决的应用问题，提出本毕业设计题目。简要介绍国内外在与毕业设计课题相关方面的有关技术，列举一些技术研究或国内外已经开展的工作，但可能还存在哪些问题或我们还可以做哪些方面的工作，选取其中的一个问题/方面进行解决，以期望达到怎样的目的，获得怎样的成效，具有什么样的理论意义或应用价值。

### 二、研究方案和主要内容（一般 150-250）

#### 1. 研究内容

介绍本课题涉及的内容与范围，在简要介绍涉及的功能模块或主要的研究要点，有什么预期研究成果（100-200字）。研究内容应具体、清晰，越规范、准确，越好。

#### 2. 研究目标

研究目标：较详细地介绍完成本课题的若干功能模块。

功能与系统环境：课题功能模块应达到的具体要求和技术指标；系统的运行环境或平台。

### 三、初拟论文提纲(这里可按二级目录列出，正式论文目录可用三级目录)

例如：

摘要

ABSTRACT

第一章 绪论

1.1 深度学习的研究现状

1. 2 本文的研究意义

1. 3 本文的主要工作和章节安排

第二章 深度学习的相关理论和有关实现技术

2. 1 深度学习的相关概念

2. 2 卷积神经网络算法的思想和有关技术

2. 3 深度学习技术对手写数字识别的有关算法及测试

第三章 手写字符识别系统总体设计

3. 1 基于 XXXX 的手写字符识别系统框架设计概述

3. 2 XXXX 平台的使用

3. 3 神经网络的设计

3. 4 前向传播的设计

3. 5 反向传播的设计

3. 6 梯度更新的设计

3. 7 激活函数的设计

3. 8 模型参数初始化的设计

第四章 手写字符识别系统在 XXXX 平台的实现

4. 1 基于 XXXX 的运算实现

4. 2 神经网络中数据结构的实现

4. 3 前向传播的实现

4. 4 反向传播的实现

4. 5 梯度更新的实现

4. 6 模型参数初始化的实现

4. 7 手写字符识别系统的实现

第五章 相关测试

5. 1 测试环境

5. 2 测试方案

5. 3 测试结果

第六章 总结与展望

6.1 总结

6.2 不足与展望

致谢

参考文献

附录 1

附录 2

在正式开始写毕业设计论文时，请注意参考文献的写法

### 参考文献

按学校要求，篇目至少要 15 篇，其中至少 2 篇为英语文献，以近 5 年为主。

### 参考文献书写要求：

出处明确，适用性强，格式要规范。对书或文章按如下要求书写：

(1) 期刊：[序号]作者姓名. 文题[类型标识]. 刊名, 年, 卷(期): 起讫页. (类型标识：专著 M, 论文集 C, 报纸文章 N, 期刊文章 J, 学位论文 D, 研究报告 R, 标准 S, 专刊 P, 其它 Z)

例：

[1]孙长东. ERP管理思想及其实施风险分析[J]. 管理世界, 2002, No.8: 143-144

[2]Geske R. The valuation of compound options[J]. Journal of Financial Economics, 1977, No. 7: 63-81.

(2) 图书：[序号]作者姓名. 书名[类型标识]. 版次(第 1 版略). 出版地：出版者, 年. 起讫页.

例：

[1] 薛华成. 管理信息系统[M]. 第3版. 北京：清华大学出版社, 2000. 124-156.

[2] Bielecki T R, Rutkowski M. Credit Risk: Modeling , Valuation and Hedging[M]. Berlin: Springer, 2002. 43—65.

(3) 网上资料类：作者、文章名、网址、日期（阅读日期）

## **进度安排**

**毕业设计（论文）进度安排：**与系论文工作总体安排一致，可以指导教师微调。

## **论文查重**

为了保证本科生毕业论文质量，学校规定每年对所有本科毕业论文在知网上进行查重。为了配合学校对本科毕业论文的查重工作，学院要求每个学生必须对毕业论文在知网进行查重。

学生自行查重可在任意平台进行，但查重结果必须符合学校有关规定（论文文字重合率 $<30\%$ ）。结果未必与知网一致。在毕业论文答辩之前完成。每人最多3次查重机会。