



基于用户评分的视频 推荐系统设计与实现



2024全国高校程序设计教育大会

报告人：邹恩岑

Email: eczou@mail.ustc.edu.cn

蘇州科技大學
SUZHOU UNIVERSITY OF SCIENCE AND TECHNOLOGY

案例项目开源地址：https://github.com/zouencen/recommender_system_4teach.git

教学改革项目支持：苏州科技大学品牌本科建设项目2022DZPP-06



目录 CONTENTS

- 01 视频推荐系统介绍
- 02 实训内容与任务
- 03 相关知识及背景
- 04 教学目的
- 05 实训教学与指导
- 06 实训原理及方案
- 07 实验报告要求
- 08 考核要求与方法
- 09 案例特色或创新
- 10 案例应用

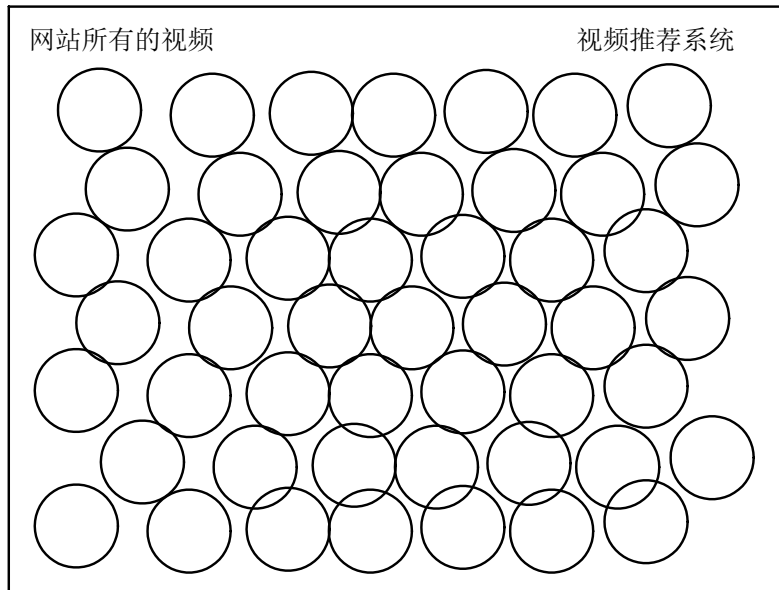
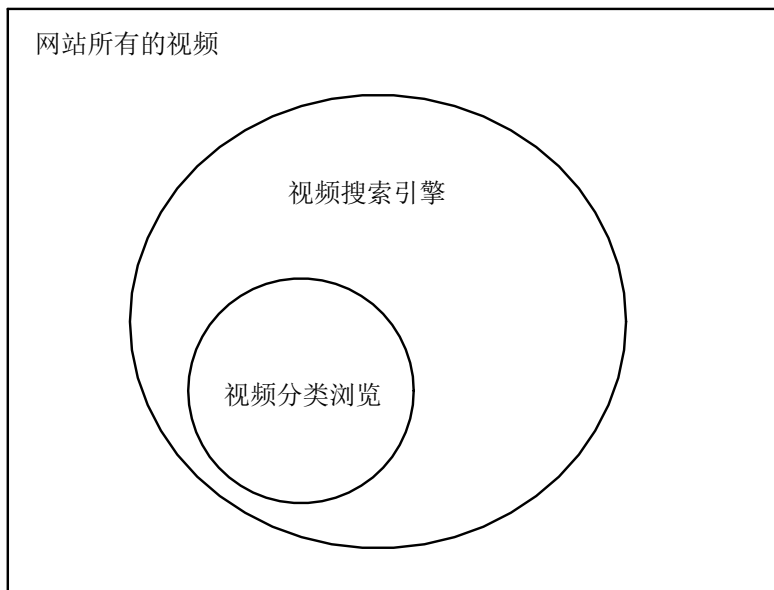




视频推荐系统介绍



電子与信息工程学院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING



数据爆炸问题凸显

由于互联网上的数据日益增多，分类浏览和搜索引擎已经越来越无法全面覆盖到互联网上的内容了。于是，互联网上出现了很多“暗数据”，始终无法被用户访问到。用户很难从海量数据中筛选出有价值的数据，造成了巨大的浪费。

个性化需求日益增强

上世纪90年代，人们提出了推荐系统的概念。推荐系统根据用户的兴趣的不同，有区别的给每位用户推荐其可能感兴趣的内容。用户在面对大量信息时，更加追求个性化的服务和产品，以满足自身独特的需求和兴趣。



实训内容与任务



電子与信息工程學院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING

实训案例任务

通过Java程序设计语言实现一个视频推荐系统，系统需结合Java、算法分析和软件工程等课程内容。根据学生年级、水平的不同，定制难度和内容。

实训过程及要求

自学预习、需求分析、实训设备与环境、算法描述、总体设计、详细设计、实验报告撰写、总结交流。

团队完成子系统

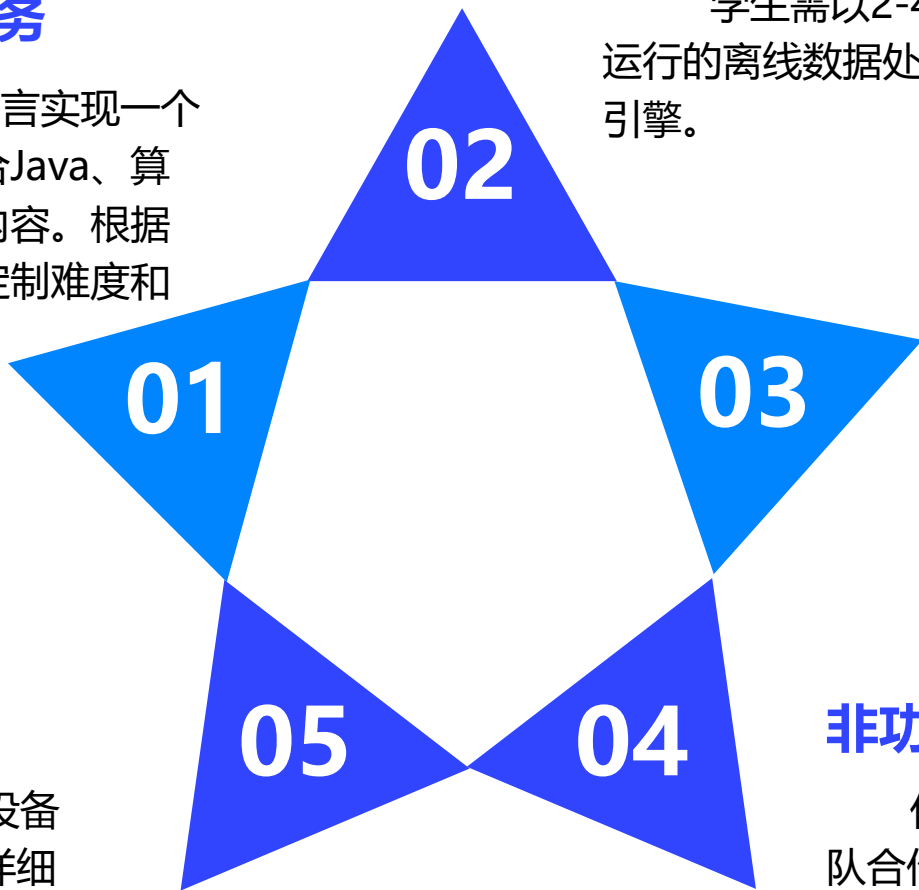
学生需以2-4人团队形式完成可运行的离线数据处理系统和在线推荐引擎。

提交结果

演示PPT、程序源码和资源包、实验报告。

非功能性要求

代码质量、性能、可扩展性、团队合作。





教学目的



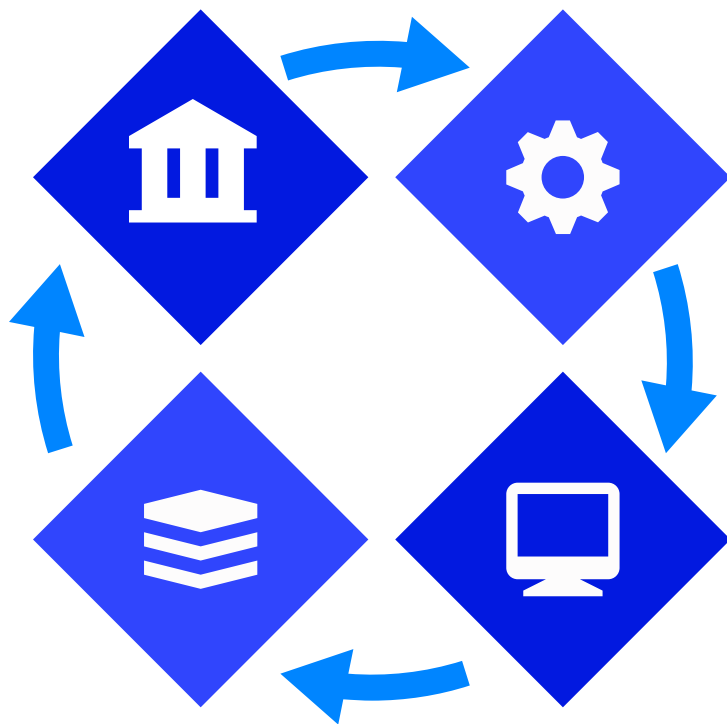
電子与信息工程學院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING

目标1

能够运用计算机**算法分析与设计基本方法**、设计范式，建立算法模型，设计算法并使用代码实现。

思政目标

学习**国家大数据战略思想**与开放共享的理念，引导学生提升程序设计、数据结构和算法能力，强调实践，培养学生正确的人生观、价值观和世界观。



目标2

能够综合运用计算机算法分析与设计的基本理论、方法，设计合理的数据结构、算法逻辑等要素，能体现出**创新意识**。

目标3

能够选择合适的建模、开发以及测试工具，对所设计的算法模型和思路进行有效的**分析和表达**，输出规范的**文档**。



相关知识及背景



電子与信息工程學院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING

1

案例核心知识

Java程序设计基础、Eclipse IDE使用、文件I/O操作、数据结构（如ArrayList和Map）和推荐系统的基本原理。

2

实践操作

学生将通过实践，学习如何处理和分析大规模数据集，以及如何设计算法来预测和推荐用户可能感兴趣的视频内容。

3

版本控制系统

学生将使用版本控制系统（如Git）进行代码管理和团队开发的技能，这对于将来的职业是重要的。



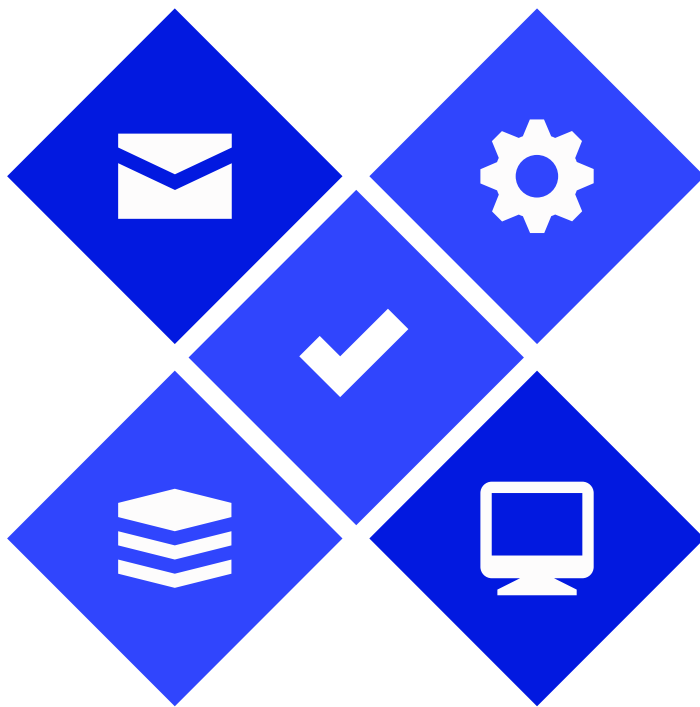


(1)知识讲解

Java程序设计基础知识，变量、数据类型、控制结构、类和对象、文件I/O、ArrayList、Map容器等。

(3)方法引导

如何将已学过的编程知识组合使用用于完成推荐系统的编写，如何用已学习的软件知识来理解推荐系统的结构。



(2)背景解释

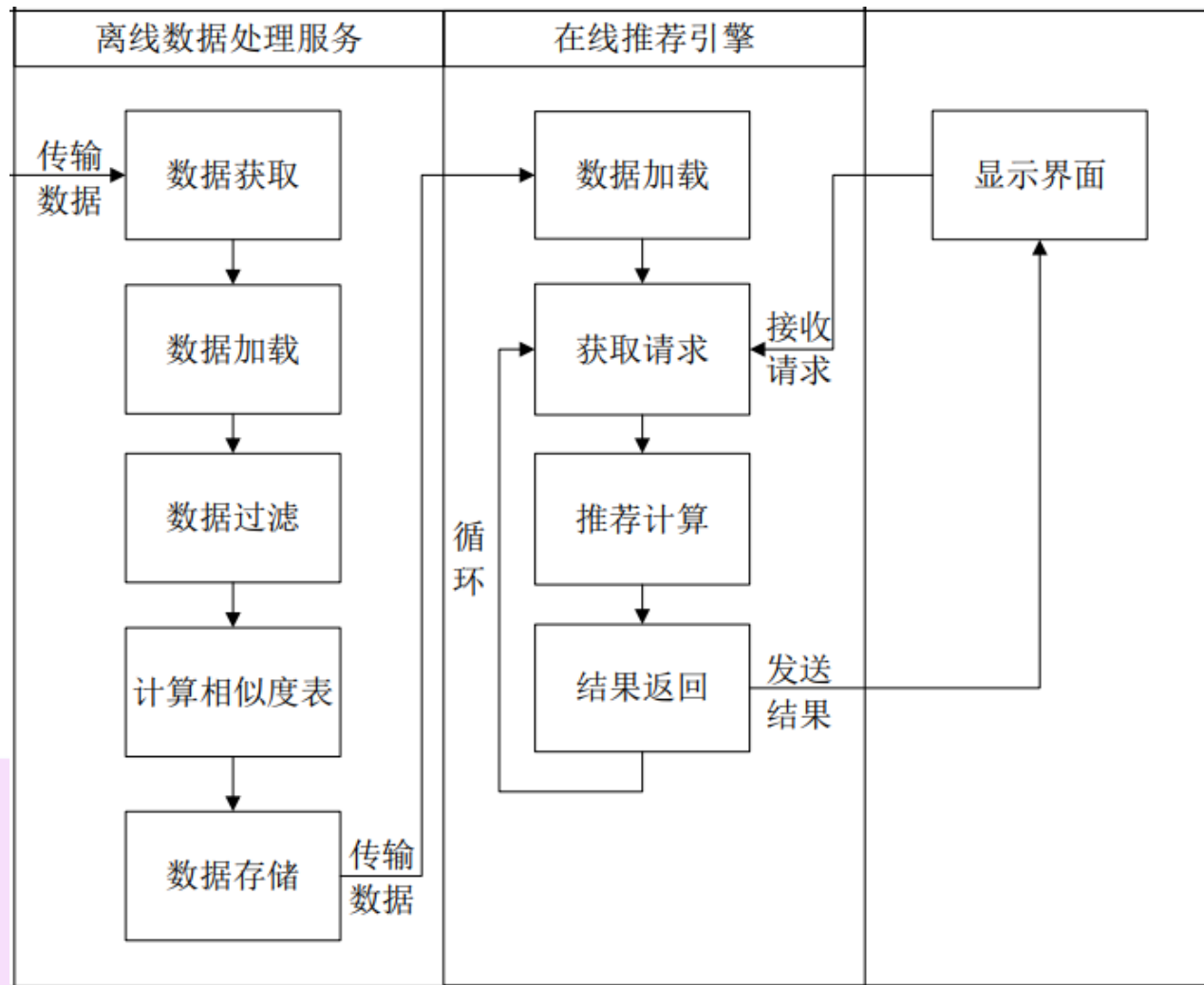
介绍推荐系统背景知识、工作机制，讨论离线数据处理和在线推荐引擎的区别，以及如何协同工作完成任务。

(4)实训中的指导或引导

提供代码示例和练习，讲解关键算法的构建步骤，引导学生实现各模块细节，并调试解决实际问题。



基本原理



推荐系统工作机制

推荐系统的核心在于理解其工作机制，即根据用户历史行为和偏好预测并推荐感兴趣的项目。

视频推荐系统组成

视频推荐系统包括离线数据处理和在线推荐两个主要部分，前者负责从数据集中提取和处理用户的评分数据，后者则使用这些数据来生成个性化推荐。

系统工作流程

系统工作流程如图所示，用户请求触发推荐引擎，引擎根据用户历史数据生成个性化推荐列表并返回给用户。



一、离线数据处理系统实现

CHAPTER





1. 数据输入



数据来源

Movie Lens网站的数据记录了用户对影片的评分，推荐系统的离线模块需要从数据仓库中获取这些数据，并对其进行处理。

数据内容

数据包含了用户编号、电影编号、用户评分和评分时间，这些数据已经经过整理和规范，方便推荐系统进行数据处理。

数据表示

数据记录的每一行都代表了一次用户的评分，例如数据1::185::5::838983525就代表了用户1对电影185《Net》的评分是5分。

数据存储

在下载完成之后，数据可以存储到文件系统中的相应位置，以供推荐系统进行处理，为用户提供更个性化的电影推荐。





2.寻找 K 邻近邻居

(1)1个电影和K个电影

推荐系统推荐电影给用户的任务是寻找与用户平时喜欢的影片相似的电影，是用用户已经看过的1个电影寻找到有相似可能的K个电影

(2)如何寻找

通过1个电影在item-to-user map里找到对应的N个用户，N个用户再再user-to-item map里找到对应K个电影



MovieID

item-to-user map

UserID1 UserID2 UserID3 ...

user-to-item map

user-to-item map

user-to-item map

MovieID11 MovieID12 MovieID13 ...

MovieID21 MovieID22 MovieID23 ...

MovieID31 MovieID32 MovieID33 ...





3. 相似度计算



01

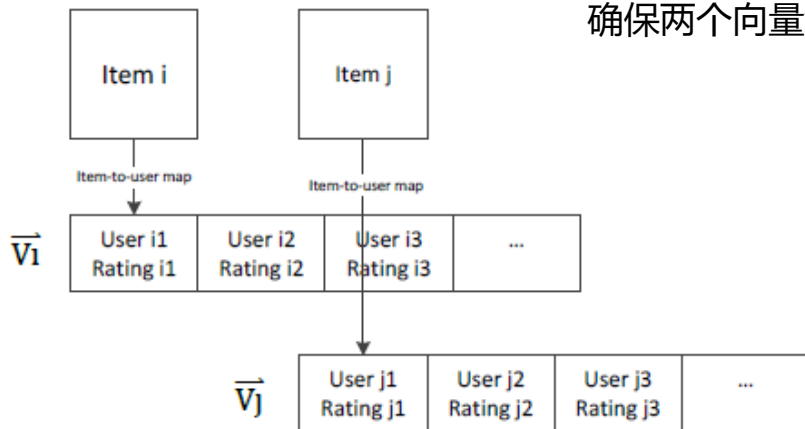
相似度计算介绍

在推荐系统中，电影的评分被视为向量，通过计算向量间的夹角来得出两个电影之间的相似度。

02

用户向量表示

将用户对项目的评分列出两个向量，分别用 V_i 和 V_j 表示，用于计算相似度。



03

维度补0

当用户对某些项目有评分而对其他项目无评分时，需将无评分的维度补0，以确保两个向量维度相同。

04

相似计算

$$\frac{\vec{V}_i \cdot \vec{V}_j}{|\vec{V}_i| \cdot |\vec{V}_j|}$$
$$\frac{\sum_{u \in U_{ij}} (R_{i,u} \cdot R_{j,u})}{\sqrt{\sum_{u \in U_{ij}} R_{i,u}^2} \sqrt{\sum_{u \in U_{ij}} R_{j,u}^2}}$$

05

sim table存储

计算得到的相似度结果存成txt文档，格式为

```
MovieID:MovieID1,simValue1
MovieID2,simValue2
MovieID3,simValue3...
```

一般先使用小量数据测试，进阶内容可要求计算全量数据



二、在线推荐引擎实现

Inputed Movie:

1 - Toy Story (1995)::Adventure|Animation|Children|Comedy|Fantasy

Recommender movies:

260 - Star Wars: Episode IV - A New Hope (a.k.a. Star Wars) (1977)::Action|Adv

780 - Independence Day (a.k.a. ID4) (1996)::Action|Adventure|Sci-Fi|War

1210 - Star Wars: Episode VI - Return of the Jedi (1983)::Action|Adventure|Sci

648 - Mission: Impossible (1996)::Action|Adventure|Mystery|Thriller

1073 - Willy Wonka & the Chocolate Factory (1971)::Children|Comedy|Fantasy

1270 - Back to the Future (1985)::Adventure|Comedy|Sci-Fi

588 - Aladdin (1992)::Adventure|Animation|Children|Comedy|Musical

32 - 12 Monkeys (Twelve Monkeys) (1995)::Sci-Fi|Thriller

356 - Forrest Gump (1994)::Comedy|Drama|Romance|War

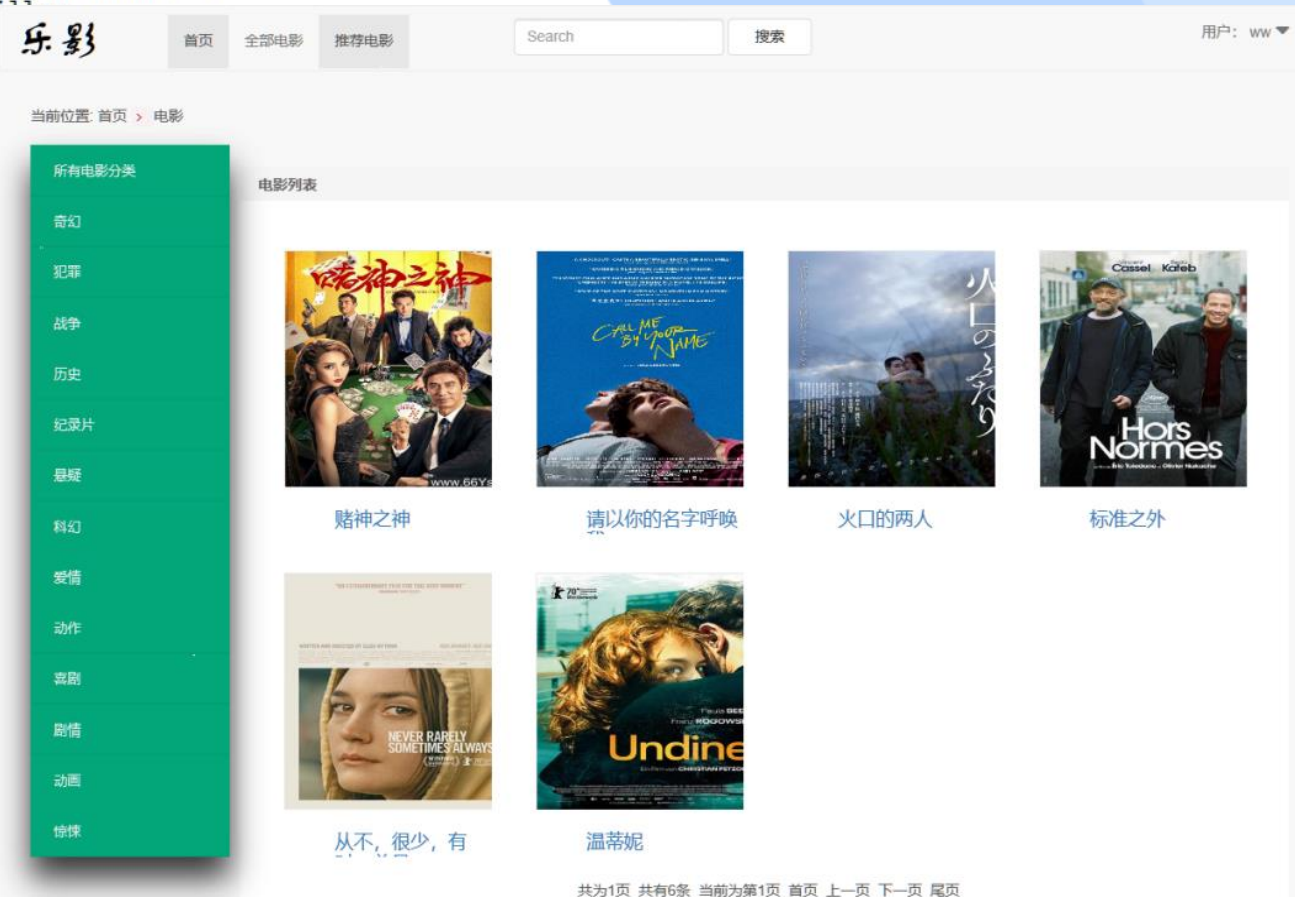
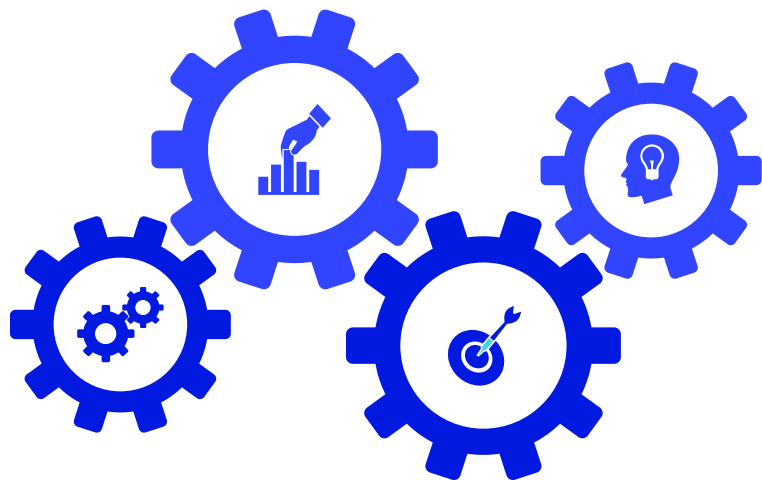
1196 - Star Wars: Episode V - The Empire Strikes Back (1980)::Action|Sci-Fi

736 - Twister (1996)::Action|Adventure|Romance|Thriller

1265 - Groundhog Day (1993)::Comedy|Fantasy|Romance

3114 - Toy Story 2 (1999)::Adventure|Animation|Children|Comedy

1198 - Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark)





二、在线推荐引擎实现



在线推荐引擎介绍

在线推荐引擎实时接收用户查询请求，经推荐引擎计算后实时返回结果，依托计算机的高速运算能力，实现高实时性的推荐，提升用户体验。



(2)相似度矩阵应用

推荐引擎使用相似度矩阵查找与给定电影记录相似的邻居电影集合，作为推荐候选电影，提高推荐准确性和用户体验。



(4)实时计算与反馈

依托计算机的高速运算能力，在线推荐引擎能够实现快速的实时计算，用户无需等待即可获得为其推荐的电影，提高用户体验。**进阶内容可要求提高每秒请求数。**



(1)相似度表加载

将相似度表数据加载到系统内存中，并使用Map数据结构进行存储，以便快速查找和计算相似度值。



(3)推荐结果生成与返回

根据用户输入的MovieID，在相似度表中查找相似的MovieID，并返回相似度最高的15个结果，以生成推荐结果。**进阶内容可要求加入混合推荐提高用户体验。**



(5)电影名称匹配与显示

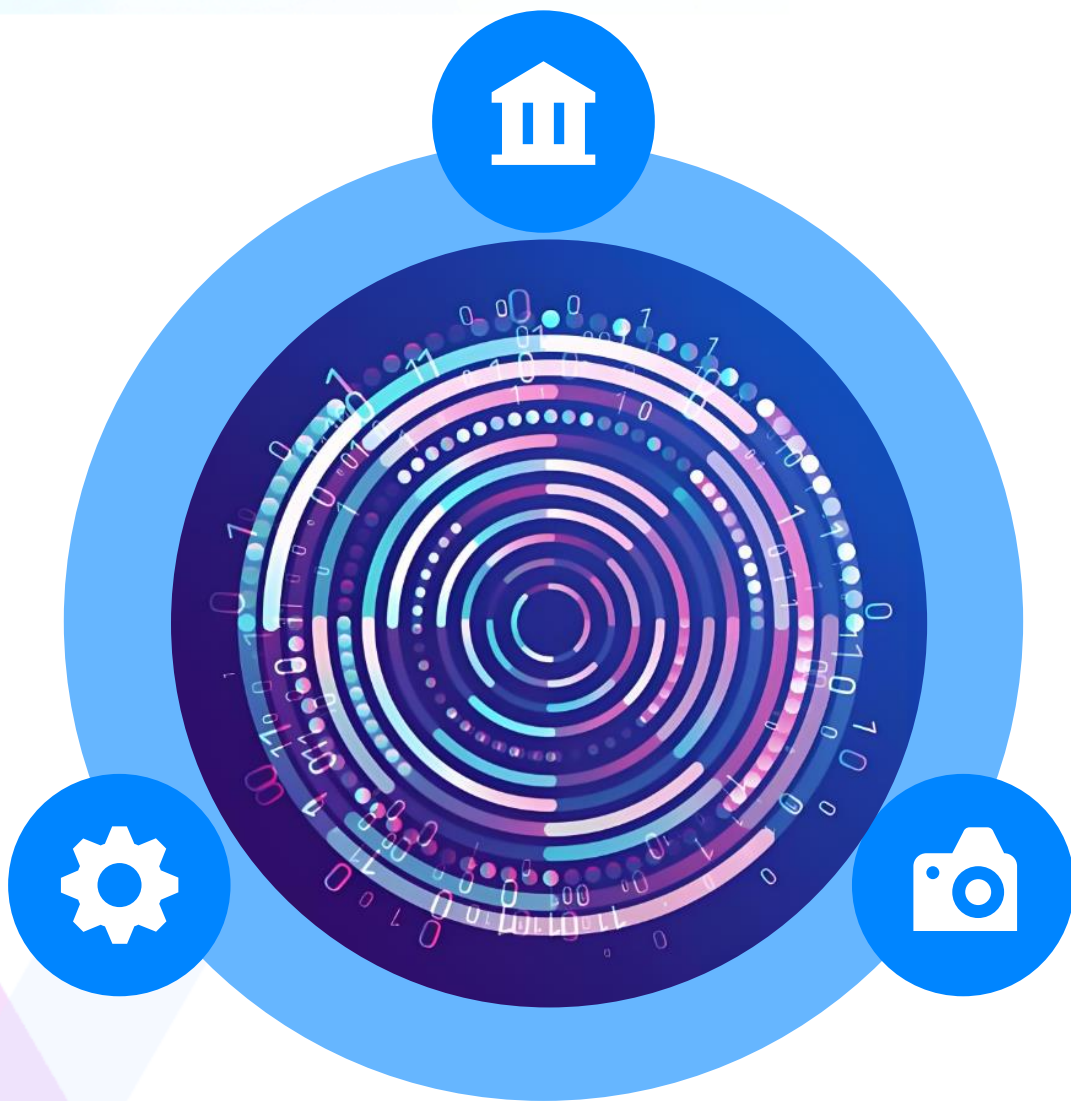
为这15个MovieID匹配电影名称的信息并显示，完成截图。用户输入一个MovieID号码，系统显示出该号码对应的电影名称和信息。



采用的程序设计思想、函数和编程技巧



電子与信息工程学院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING



模块化

将系统分为不同模块，每个模块负责不同功能，如数据读取、映射构建、推荐生成等。

面向对象

使用Java的面向对象特性来设计数据和操作，如创建Movie和User类来表示电影和用户。

文件I/O

使用FileReader和BufferedReader读取文件，使用FileWriter写入文件。

数据结构

使用HashMap存储映射，使用ArrayList存储用户和电影的列表。

异常处理

使用try-catch块处理可能的I/O异常。



实验报告

实训设备与环境

描述实验的硬件和软件环境，包括计算机配置、操作系统、Java版本以及Eclipse IDE等其他资源。

设计思路

阐述实现视频推荐系统时的整体设计思路，解释推荐系统原理在程序设计任务中的应用，包括数据结构选择和算法实现。

程序清单

提供完整的源代码清单，并使用适当的注释来解释代码的功能，确保代码的可读性和可维护性。

运行结果

展示系统运行的截图或视频，证明推荐系统正常工作，包括测试案例和输出结果，确保系统的正确性。



電子与信息工程学院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING



程序使用说明

详细说明如何运行和使用推荐系统，包括启动程序、输入数据和解读输出结果的步骤，确保他人能顺利使用。

出现的问题和解决方法

记录编写代码和运行程序时遇到的问题，以及解决问题的方法和步骤，以便于后续查阅和修改。

心得体会

反思实训过程中的挑战和学习到的经验，讨论与团队合作的重要性，以及合作如何帮助完成任务。



考核要求与方法



電子与信息工程学院
SCHOOL OF ELECTRONIC & INFORMATION ENGINEERING

01

考核目的：评估学生在实训过程中的学习成效、程序设计技能、问题解决能力以及团队合作能力。

02

考核内容：实训过程中的积极参与度和贡献；视频推荐系统的功能完整性和性能；代码的规范性、可读性和注释的充分性；实验报告的详尽程度、解决问题的难度和准确性。

03

考核时间节点：实训过程中的定期检查；实训结束时的最终评估，需要每位同学讲述自己的工作；实验报告的提交。

04

评分标准：实训参与度和团队贡献（30%）.系统功能和性能（30%）；代码质量（10%）；实验报告（30%）。

05

考核方法：自我评估；教师评审。



案例项目特色

案例项目的特色可体现于项目背景的工程性，知识应用的综合性，实现方法的多样性。

提高多样性

学生们通过参与案例项目，可以提高解决实际工程问题中多样性、创造性的能力。

多种实现方法

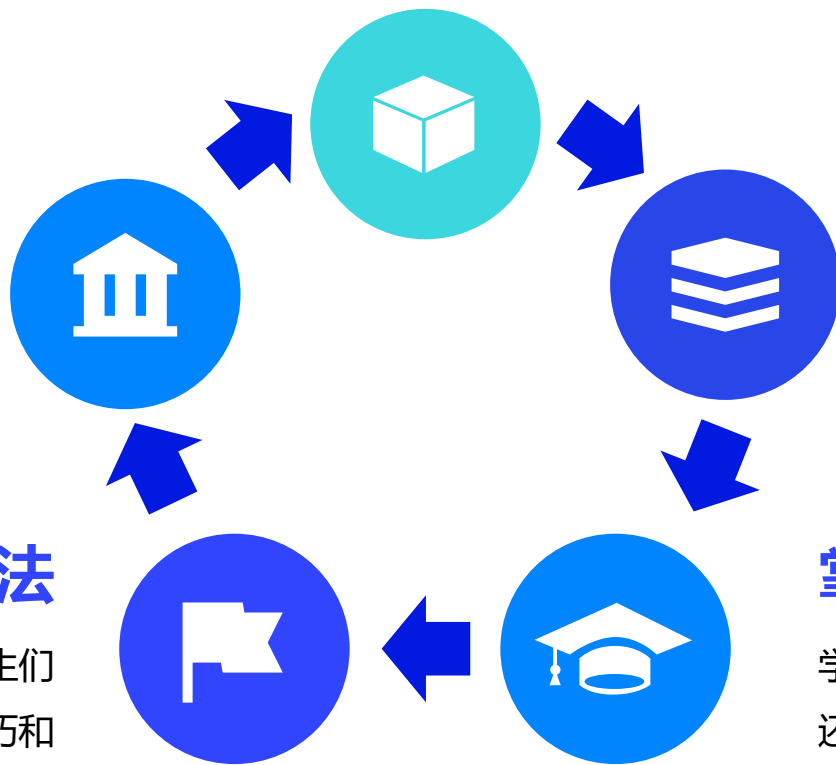
案例通过提供多种实现方法，鼓励学生们发挥创造力，探索不同的程序设计技巧和算法。

强调实用

案例特色在于其强调了实际工程问题解决的实用性，将理论知识与实际应用紧密结合。

掌握应用

学生们不仅需要掌握程序设计基础知识，还要学会如何将这些知识应用于构建一个复杂的视频推荐系统。



锻炼实践能力

低年级同学：

- 要求实现出功能
- 要求能够读懂完整的代码
- 能运行小型数据集

高年级同学：

- 实现较高每秒请求数(QPS,200以上),
- 使用多线程技术, 高速缓存技术(redis, memcached)
- 运行完整的、大型的数据集,
- JavaEE的WEB界面, 使用数据库,
- 算法精度提升(混合推荐, 准确率, 召回率)
- 完整的软件工程开发流程

课程应用

程序设计相关课程

- 分模块布置作业, 数据I/O模块、数据预处理模块、相似计算模块、在线推荐模块等。

算法分析设计相关课程

- 相似推荐核心算法实现, 多种推荐算法实现, 多种推荐算法准确率分析对比。

软件工程课程及实践项目

- 按照软件流程实现整个推荐系统, 包括需求分析、概要设计、详细设计、系统实现和软件维护等步骤以及相关的工程文档的产出。

提升能力

项目促进了学生在之后的考研深造和工作中创新能力和工程能力的提升, 为学生未来的发展奠定了坚实的基础。

案例扩展

- 基于相似计算的电子作业查重系统——应用于实际作业批改
- 基于相似计算的非标准地址标准化匹配系统——应用于公安系统
- 基于Spark大数据平台的视频推荐系统
- 基于CNN卷积神经网络的视频推荐系统——未来课程案例应用探索



THANKS

感谢观看

案例项目开源地址：

https://github.com/zouencen/recommender_system_4teach.git

教学改革项目支持：苏州科技大学品牌本科建设项目2022DZPP-06

蘇州科技大學

SUZHOU UNIVERSITY OF SCIENCE AND TECHNOLOGY