# 智能推荐引擎

## 研究目的

充分挖掘用户的历史记录和浏览记录，将这些记录与用户信息相结合，推荐出用户可能感兴趣的事物。这样既给用户节约时间搜索信息，又让用户体验到智慧推荐方便快捷的优点。

## 研究现状

目前比较常见的推荐技术主要分为：协同过滤推荐和基于内容过滤的推荐。 协同过滤技术是基于社会学研究的算法，根据协同过滤可以有效跳过信息所产生的干扰，直接对用户做出相似比较，因此可以更好更准确地获取推荐信息。近几年，协同过滤技术被利用于多种领域，采用这一技术所研发的智能推荐引擎在很多大型网站上都已出现。这一技术主要包括基于项目和基于用户的协同过滤，基于项目的协同过滤在评分矩阵的基础上分析比较项目之间的相似性，结合用户信息做出合理推荐结果。。然而，协同过滤技术或多或少要遇到稀疏矩阵问题，这个问题成为了协同过滤技术的一个难题之一。

基于内容过滤技术则不同，主要是将项目信息作为基础对象，在项目之间进行信息的对比比较得出相似性，在用户浏览操作记录的基础上创建用户对比信息，同时结合项目之间与用户之间的对比信息以完成推荐结果的输出。信息作为项目的基本组成机构，这就需要对信息进行研究，以达到推荐结果的高质量和高精确度的目的。基于内容过滤推荐技术的运用过程中，或多或少都要面对多种困难问题。比如难以得到高精准率的文本信息和文本信息难以匹配等。另外，视频、音频、图片等信息不能生成统一准确的文本信息，也就造成了这一推荐技术不能把视频、音频、图片等信息进行推荐。这是基于内容过滤推荐技术的主要缺陷

。

## 研究内容

### 混合推荐算法

混合推荐算法的基本思想是把基于内容过滤和协同过滤两种算法取长补短进行组合，第一步把用户特征、浏览记录、历史操作等作为智能引擎推荐的基础，根据KNN算法获得项目特征，计算项目之间的相关度，然后根据基于内容过滤的推荐技术对协同过滤中产生的系数矩阵进行预测性填充，在进行协同过滤算法之前，现将稀疏矩阵填充为密集矩阵，填充的数据具有一定可信度，然后对密集矩阵进行协同过滤，以此来解决矩阵的稀疏性问题。

为了获得用户对推荐结果的回馈评分，有两种主要方法。首先，让用户对推荐结果打分；其次，根据使用者的操作行为来判断喜爱程度，例如，使用者打开指定页面，页面停留时间长短，复制拷贝文件，可以把这三项与对应权重的乘积之和作为用户对该类型的喜爱程度。

基于内容过滤算法提取项目的基本特征，再对比项目间的特征相似度，从而有效地得出推荐结果。系统采用的ＫＮＮ算法目的在于匹配项目中Ｋ个相似度最高的项目，作为预测推荐输出到用具接口。

1. 提取数据源关键信息，数据源作为各个关键项目的载体，包括了项目的重要属性，通过提取部分关键信息来确定项目的特征向量，作为项目相似度比较的主要依据。用户专业和爱好这两个属性作为KNN算法的主要属性。
2. 设计相关度计算方法，构建相似度模型。
3. 生成相似度矩阵。
4. 通过相似度矩阵得到推荐结果。通过吧KNN算法思路和项目相似度计算综合考虑，将实现结果的输出。只要报与目标内容相似度最高的N个项目输出到用户接口即可。

### 混合推荐算法设计

基于协同过滤算法经常会难以解决稀疏性问题和冷启动问题，于是与 KNN 算法的内容过滤结合起来，能够有效解决这两个难题，并且根据相似度计算得出推荐结果。接下来还需要考虑到混合推荐算法在实际应用中的实时性。