目 录

[第1章 基于用户偏好估计的协同过滤算法 1](#_Toc499989762)

[1.1 相关工作 1](#_Toc499989763)

[1.1.1 用户兴趣模型的建立 1](#_Toc499989764)

[1.1.2 相似性度量 1](#_Toc499989765)

[1.1.3 评分预测规则 2](#_Toc499989766)

[1.1.4 矩阵填充技术 3](#_Toc499989767)

[1.2 算法设计 3](#_Toc499989768)

[1.2.1 分类相似度 3](#_Toc499989770)

[1.2.2 用户兴趣估计 3](#_Toc499989771)

[1.2.3 用户相似性计算 3](#_Toc499989772)

[1.2.4 算法描述 3](#_Toc499989773)

[1.3 实验结果及分析 3](#_Toc499989774)

[1.3.1 数据集描述 3](#_Toc499989776)

[1.3.2 评估指标 3](#_Toc499989777)

[1.3.3 实验结果 3](#_Toc499989778)

[1.4 本章小结 4](#_Toc499989783)

# 基于用户偏好估计的协同过滤算法

**(说明)**

本章针对电力领域的数据稀疏性问题，提出了基于用户偏好估计的协同过滤算法。

## 相关工作(未完成)

### 用户兴趣模型的建立(完成ing)

设有个用户，构成集合；个商品，构成商品集合；用户对商品的评分为，则用户-商品评分矩阵为



正如前文所述，由于数据稀疏性原因，该矩阵不是所有评分数据都存在，而要解决数据稀疏性问题是利用已经存在的而尽可能地预测和填补缺失的评分数据。

### 相似性度量(完成ing)

余弦相似度和皮尔逊相关系数。

1. 余弦相似度

设为用户评分过的商品集合，即，为指示函数，当用户对商品给出打分则其值为1，否则为0；余弦相似度是将用户的评分作为维商品空间上的向量，利用两个用户评分向量夹角的余弦值来衡量两个用户的相似性，如式所示。



由于使用余弦相似度来衡量两个用户的相似性没有考虑用户评分尺度的差异，因此，通过将所有评分与用户的商品的平均评分作差来消除这种差异，如式所示。



1. 皮尔逊相关系数

皮尔逊相关系数是余弦相似度在维度值缺失情况下的一种改进,由皮尔逊相关系数计算用户相似性的方法如式



其中，为用户给出评分的平均值。

正如前面两种相关性的描述，这些相关性度量方法都只考虑那些有共同的商品，对于那些缺失的数据中，缺少评分的商品会隐藏着用户潜在的偏好，只是没有做出评分而已。因此，若数据的性属性问题足够大则基于此计算出的相似性会有很大偏差。

### 评分预测规则(完成ing)

利用用户间相似度能在数据集中搜索到目标用户的邻居用户集合，然后对该邻居集合中所用用户对目标商品的评分的加权平均值，权重为用户间相似度。那么目标用户对目标商品的评分的预测值就可以定义为该加权平均值。又由于用户间的评价尺度差异，所以在评分的预测值方法中引入偏置量，，如式



其中，为用户对商品的评分预测值，为用户的平均评分，为用户的邻居集合。

### 矩阵填充技术(未完成)

## 算法设计(未完成)



### 分类相似度(未完成)

### 用户兴趣估计(未完成)

### 用户相似性计算(未完成)

### 算法描述(未完成)

## 实验结果及分析(未完成)



### 数据集描述(未完成)

### 评估指标(未完成)

### 实验结果(未完成)



## 本章小结(未完成)

# 