目 录

[第1章 基于时序分析的协同过滤算法 1](#_Toc498020500)

[1.1 问题引入 1](#_Toc498020501)

[1.2 SeqBasedMF推荐算法描述 1](#_Toc498020502)

[1.3 核密度估计的引入(未完成) 1](#_Toc498020503)

[1.4 本章小结 1](#_Toc498020508)

# 基于时序分析的协同过滤算法

**（说明）**

## 问题定义

**（说明）**

### 问题引入

伴随着互联网的发展，协同过滤算法是至今为止发展最为成熟、应用最为广泛的推荐算法。该算法的诞生标志着推荐系统的诞生【9】，也是本文的理论依据【16】。本文的推荐算法主要是基于用户与商品评分矩阵进行相应的计算处理，目的是对推荐系统中用户的评分行为进行预测。本文所指“用户”，即大用户，为可接入较高的电压等级，具备一定购电规模电力用户【17】。“商品”即电力能源，发生交易的前提是大用户直购电工作的开展。在电力体制中，大用户直购电是电厂和终端购电大用户之间通过直接交易的形式协定电量和购电价格，然后委托电网企业将协议电量由发电企业输配终端购电大用户，并另支付电网企业所承担的输配服务【18】。随着电力改革的推行，大用户可以直接与发电企业达成交易，电网企业不再是众多大用户的唯一售电方。发电企业面对的销售对象逐渐增加以及这样所产生的竞争效应，势必会大大提高电力生产技术和价格定位的灵活性。减少对电力市场的约束，让市场自动调节，使生产和消费双方获得双赢。

### 电力交易推荐系统拟解决问题

大用户通过自身需求与一个或者多个发电企业进行直接的自主选择交易，这种直购电模式在试点地区运营过一段时间后带来了一定的改革红利。电力交易推荐系统的最终目的是针对电力改革后对电力市场中要出现的新大用户推荐令其满意的发电企业。为电力市场中新出现的大用户推荐其满意的发电企业作为其要进行交易的候选对象。电力交易的推荐系统主要解决以下几个问题。

1. 为用户生成满意度较高的推荐列表

为用户生成推荐列表首先要获取用户偏好，要了解用户究竟想要什么。最好的办法是用户注册系统时就主动描述其偏好告知系统，但是这样存在三个不可行之处：首先，用户的偏好无法用当前的自然语言处理技术完全理解；其次，用户的偏好是不断产生变化的，总会产生新的变化，然而用户却不会经常在系统中更新他们的偏好；另外，有些偏好无法用语言来描述，以至于某些用户无法明确自己真实的偏好【16,19】。基于以上问题，用户的历史行为和偏好是推测用户未来行为和偏好的宝贵资料，利用这些数据可以巧妙的避开上述不可行之处，因此，大量的用户数据是推荐系统的重要元素，要从现有的数据中挖掘出用户不断变化的、难以表达和理解的偏好。如何准确把握用户的偏好是影响准确率（推荐系统的重要指标）的关键问题。

1. 评分矩阵数据稀疏性问题

数据稀疏指的是评分矩阵中只有少数单位（在大量的记录中）被有效地用来表示典型的数据向量【20】，而实际上有大量单位的值为零，只有少数单位的值是非零的。数据稀疏问题作为推荐系统的经典挑战已经成为众多研究者急于攻克的问题之一。在协同过滤算法中，通过分析用户交易行为以及用户交易后对所购商品的评分来预测用户对新商品的偏好评分进而给用户推荐商品，如果用户评分积极性较高则用户-商品评分矩阵会有较多有价值数据；然而，即使用户积极性再高，也无法针对所有商品给出有价值的评分，而评分矩阵中每个单位是某个用户对某一个商品的评分，因而评分矩阵会有很多空位；加之多数用户并不会有充分的积极性来对商品评分，数据的稀疏性可想而知。在电力交易系统中，由于电力改革前的交易模式单一，大用户与发电企业之间的交易一直被电网公司驱动，大用户的购电需求以及满意度评价往往只针对电网公司并非发电企业，历史遗留的原因造成数据稀疏性的问题更加严重。电力改革后，大用户可以对发电企业自主进行点对点交易，其选择要综合多方面因素，若大用户对发电企业给出的电能越满意则相应的交易次数就会增加，反之，交易次数会减少，因此将大用户与发电企业的交易次数设定为初始评分矩阵中的元素是完全可行的。 其实，数据稀疏的问题在本质上是无法完全克服的，为了解决这个问题，很多文献中已经提出并使用了许多稀疏措施【21】。例如：扩散算法【22】、迭代寻优算法【23】、转移相似性算法【24】。本文结合概率矩阵分解算法可以一定程度上解决数据稀疏性问题。

1. 新大用户进入系统的冷启动问题

正如前面提及的,用户的历史行为和偏好是推测用户未来行为和偏好的宝贵资料,因此大量的用户历史行为数据就成为推荐系统的极其重要的组成要素以及先决条件【9】。这些问题在多数互联网公司里面或许算不上问题，因为经过长时间的运营这些互联网公司的互联网应用积累了大量的用户行为数据。但是对于电力推荐系统来说，大用户历史行为数据的缺乏是关键性问题，“巧妇难为无米之炊”，这类由于用户历史行为数据缺失而导致推荐不准确的问题，称之为“冷启动问题”（cold start）。一般来说，冷启动问题主要分为三类：用户冷启动、物品冷启动、系统冷启动【9】。用户冷启动就是前面提到的缺乏用户历史行为数据而造成无法借此对其个习惯化推荐；物品冷启动意味着将新的商品推荐给可能对它感兴趣的用户；系统冷启动代表在一个全新开发的网站上，没有用户以及用户交易行为，也要能给新用户带来个性化推荐的服务体验。针对这三类问题，有不同的解决方案，大体上有几类：提供非个性化推荐服务、利用新用户注册时提供的基本信息做粗糙的个性化和利用社交网络导入社交网站上好友信息等【9】。

在电力交易推荐系统中，由于新大用户没有明确的历史交易行为作为依据，无法提供准确的推荐给这类用户。因此，本文给出这样一种解决方案：当新用户注册时，即需要添加一些需求偏好，利用这些偏好数据作为它的历史交易行为数据，而匿名登录用户（即未注册用户）仅有浏览一定范围内的信息的权限，仅提供交易量较多的发电企业作为推荐结果。

### 问题描述

本文的推荐算法主要是基于用户与商品评分矩阵进行相应的计算处理，目的是对推荐系统中用户的评分行为进行预测。形式上，设系统中存在个用户，其构成的集合为.存在个商品，其构成的集合为，用户-商品评分矩阵为，其中，代表用户对商品的评分。协同过滤算法利用矩阵分解模型学习用户或商品的特征向量，然后基于此特征向量预测未知评分。

## 基于时序分析的用户社交关系选择

## 概率矩阵分解

## SeqBasedMF推荐算法

复杂度分析



## 本章小结

# 