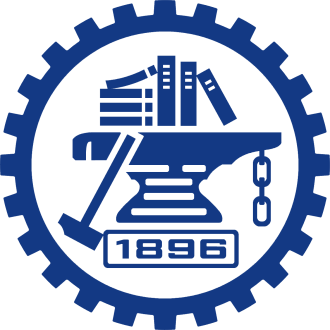
申请上海交通大学硕士学位[论文](#_OFDM技术简介_1)

基于强化学习的机票动态定价策略研究

****

**学校：** 上海交通大学

**院系：** 电子信息与电气工程学院

**班级：** B1703391

**学号：** 11703910026

**硕士生：** 刘泽霖

**专业：** 计算机科学与技术

**导师：** 曹健

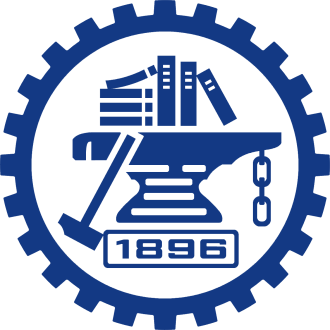
**上海交通大学电子信息与电气工程学院**

**2020年1月**

**A Dissertation Submitted to Shanghai Jiao Tong University for the**

**Degree of Master**

Airline Tickets Dynamic Pricing Based on Reinforcement Learning

****

**Author:** Zelin Liu

**Specialty:** Computer Science

**Advisor:** Jian Cao

School of Electronic Information and Electrical Engineering

Shanghai Jiao Tong University

Shanghai, P.R.China

January, 2020

**上海交通大学**

**学位论文原创性声明**

本人郑重声明：所呈交的学位论文《机票价格预测系统的设计与实现》，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

**上海交通大学**

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

**保密**□，在年解密后适用本授权书。

本学位论文属于

**不保密**□**√**。

（请在以上方框内打“**√**”）

学位论文作者签名： 指导教师签名：

日期： 年 月 日 日期： 年 月 日

**基于强化学习的机票动态定价策略研究**

摘要

作为在线旅行社（OTA），它必须确定机票的合理价格，该价格等于佣金和采购成本的总和。当前，大多数OTA在一段时间内对所有航空公司采用固定的佣金费率。但是，旅客可能会根据不同的航空公司和机票供应情况来调整对价格的期望。因此，有可能并且有必要动态地调整每个航空公司的机票价格，以增加其收入。不幸的是，机票的需求和可用性会随着非常复杂的模式而变化，因此，即使不是不可能，也很难采用数学模型来描述它们并得出解析解。因此，我们采用强化学习方法来学习动态定价规则。但是，不可能通过调整价格来观察乘客的反应，这对于强化学习是必不可少的。因此，我们开发了一种乘客模拟器，可以根据机票价格生成乘客的响应。为了使旅客模拟器更加逼真，它会根据历史数据和最新数据不断调整其固有模型。在真实数据集上的实验结果表明，我们的方法可以有效地学习动态定价规则。

**关键词**：城际出行路径搜索 出行偏好 个性化路径推荐 路径缓存

Air Ticket Dynamic Pricing Rules based on Reinforcement Learning

ABSTRACT

As an online travel agency (OTA), it has to decide a reasonable price for flight tickets which equals to the sum of commission fees and procurement costs. Currently, most OTAs adopt a fix rate of commission fees for all airlines in a period of time. However, passengers may adjust their expectations to the prices in terms of different airlines and availability of flight tickets. Therefore, it is possible and necessary to adjust the flight ticket prices for each airlines dynamically in order to increase their revenues. Unfortunately, the demands and the availability of flight tickets change following very complex patterns so that it is very hard, if not impossible, to adopt mathematical models to describe them and to derive analytical solutions. Therefore, we apply reinforcement learning approach to learn dynamic pricing rules. However, it is not possible to adjust the prices to observe passengers' responses, which is a necessary for reinforcement learning. Therefore, we develop a passenger simulator which can generate passengers' responses according to flight tickets' prices. In order to make passenger simulator more realistic, it adjusts its inherent models based on historical data and up-to-date data continuously. The experimental results on a real-world data set show that our approach can learn dynamic pricing rules efficiently.

Keywords:

目录

[摘要 I](#_Toc469747835)

[ABSTRACT II](#_Toc469747836)

[第一章 绪论 1](#_Toc469747837)

[1.1 研究背景及研究意义 1](#_Toc469747838)

[1.2 国内外研究现状 3](#_Toc469747839)

[1.2.1 路径搜索的研究现状 3](#_Toc469747840)

[1.2.2 路径推荐的研究现状 4](#_Toc469747841)

[1.2.3 既有内容总结 5](#_Toc469747842)

[1.3 本文研究内容和结构安排 5](#_Toc469747843)

[1.3.1 研究内容 5](#_Toc469747844)

[1.3.2 文章结构安排 6](#_Toc469747845)

[第二章 相关技术 9](#_Toc469747846)

[2.1 路径规划 9](#_Toc469747847)

[2.2 推荐系统 11](#_Toc469747848)

[2.2.1 推荐系统概况 11](#_Toc469747849)

[2.2.2 常用相似度的计算方法 13](#_Toc469747850)

[2.2.3 常用评价指标 15](#_Toc469747851)

[2.3 缓存策略 16](#_Toc469747852)

[2.4 本章小结 17](#_Toc469747853)

[第三章 个性化路径推荐模型 19](#_Toc469747854)

[3.1 引言 19](#_Toc469747855)

[3.2 个性化路径推荐模型 20](#_Toc469747856)

[3.3 本章小结 23](#_Toc469747857)

[第四章 带约束的top-K路径搜索算法 25](#_Toc469747858)

[4.1 引言 25](#_Toc469747859)

[4.2 路径搜索算法KSPG 25](#_Toc469747860)

[4.2.1 路径搜索问题定义 25](#_Toc469747861)

[4.2.2 KSPG算法 26](#_Toc469747862)

[4.3 可到达距离计算方法 32](#_Toc469747863)

[4.4 实验结果分析 32](#_Toc469747864)

[4.4.1 实验准备 33](#_Toc469747865)

[4.4.2 实验结果与分析 35](#_Toc469747866)

[4.5 本章小结 36](#_Toc469747867)

[第五章 路径推荐算法 37](#_Toc469747868)

[5.1 引言 37](#_Toc469747869)

[5.2 特征预处理 37](#_Toc469747870)

[5.3 基于信息熵的路径推荐算法 38](#_Toc469747871)

[5.3.1 路径相似度计算 39](#_Toc469747872)

[5.3.2 基于信息熵的路径推荐算法 41](#_Toc469747873)

[5.4 实验结果分析 44](#_Toc469747874)

[5.4.1 实验数据和相关设置 44](#_Toc469747875)

[5.4.2 推荐评价指标 49](#_Toc469747876)

[5.4.3 结果分析 49](#_Toc469747877)

[5.5 本章小结 54](#_Toc469747878)

[第六章 路径搜索的缓存策略 55](#_Toc469747879)

[6.1 行程搜索数据分析 55](#_Toc469747880)

[6.2 行程缓存模块 56](#_Toc469747881)

[6.3 路径缓存模块 57](#_Toc469747882)

[6.3.1 路径缓存策略 58](#_Toc469747883)

[6.3.2 路径缓存结果分析 59](#_Toc469747884)

[6.4 路径缓存更新策略 60](#_Toc469747885)

[6.4.1 路径缓存更新策略 60](#_Toc469747886)

[6.4.2 缓存路径更新策略结果分析 62](#_Toc469747887)

[6.5 本章小结 63](#_Toc469747888)

[第七章 全文总结与展望 65](#_Toc469747889)

[7.1 本文工作总结 65](#_Toc469747890)

[7.2 未来工作展望 65](#_Toc469747891)

[参考文献 67](#_Toc469747892)

[致谢 71](#_Toc469747893)

[附录一 符号与标记 72](#_Toc469747894)

[附录二 英文缩略语表 73](#_Toc469747895)

[攻读硕士学位期间已发表的论文 74](#_Toc469747896)

1. 绪论
   1. 研究背景及研究意义

近年来，在线旅行社（OTA）成为许多想要购买机票，预订酒店或购买度假机票的乘客的首选，因为OTA可以提供更多选择和服务。以机票为例，在旅客输入查询信息后，将显示所有满足其要求的可用机票。通常，这些机票的信息包括航班号，起飞时间，航空公司，飞机类型，机票等级和价格。具体来说，价格是佣金和采购费的总和。由于采购费用是由航空公司决定的，因此佣金是由OTA自己决定的。目前，佣金是所有航空公司机票的固定费率，可以在一段时间后进行调整。

显然，不同航空公司对机票价格的期望不同。即使是同一家航空公司的机票，起飞前的剩余时间和机票供应情况也会影响期望值。同时，如果无法在起飞时间之前售出机票，则意味着OTA错过了从这张机票中赚钱的机会。因此，有可能并且有必要动态地调整每个航空公司的机票价格，以增加其收入。

动态定价是一种业务策略，可以及时调整产品价格，以便在正确的时间向正确的客户分配正确的服务[1]。先前的许多研究都通过使用各种数学公式对客户的需求曲线进行建模来关注动态定价问题[3] [4]。然而，机票是易腐产品，具有有限的时间范围和较低的边际成本[2]。尤其是易腐产品的需求曲线是非平稳的，由于易腐产品的需求受众多因素的影响，因此很难用数学函数来描述需求。本文提出了一种利用动态定价规则学习方法强化学习，这是一种无模型的方法。 DeepMind提出的Deep Q网络算法将神经网络与强化学习相结合，取得了巨大的成功[5]。本文提出了一种深度强化学习方法来学习动态定价规则。此外，注意到时间相关性时，基础模型变成了部分观测的马尔可夫决策问题。本文提出了一种DRQN模型，该模型在传统的深度Q网络算法中使用循环网络而不是全连接网络。

不可能调整价格以观察乘客在真实环境中的反应，这对于强化学习是必不可少的。我们实现了一个乘客模拟器平台，该平台可以根据模拟乘客的喜好预测其是否会购买机票。该平台可以在培训DRQN时提供反馈。与现有的旅客模拟器平台（例如PODS [5]）进行比较，我们的旅客模拟器可以根据历史数据和最新数据来调整其固有模型。

本文其余部分的结构如下。第2节介绍相关工作。在第3节中，将介绍整个框架。在第4节中，提出了基于旅客偏好模型，旅客选择模型和差距模型的旅客模拟器。第5节介绍了一种称为DRQN的强化学习方法，用于学习动态定价规则。第6节介绍了评估乘客模拟器的准确性以及DRQN和其他同行学习的动态定价规则的性能的实验结果，最后在第7节得出了结论

1. 相关技术
   1. 路径规划

动态定价问题，也称为收益管理，收入管理问题，着重于寻找一种调整产品价格的策略，以使总收入最大化。在不同类型的乘客需求模型假设下已经讨论了这个问题。 Gallego和Van Ryzin使用强度控制和形成的强度作为库存水平和视线长度的函数提出了这个问题[6]，而Bitran提出了使用非均匀泊松分布来描述需求模型的方法[7]。其他研究人员则重点放在旅客到达模型上，而不是弄清楚旅客需求模型[8]。在需求和环境模型的假设下，提出了多种方法来解决收益最大化问题。 Anjos提出了一种分析模型来描述单向定价下的航班预订模式，并得出了英国航空公司最优价格结构的表达式[2] [4]。显然，提供一个分析模型来描述不断变化的需求是相当困难的。

旅客模拟器是一种可以模拟环境中旅客选择行为的工具。选择行为得到了广泛的研究。例如，[10]提出了一种服务质量差距模型作为选择模型的效用函数。 Cao J.等人提出了描述乘客对航班偏好的模型[11]。此外，还开发了一套乘客模拟器。其中，旅客起源地模拟器（PODS）是最著名的一种，它可以模拟航空业中各种可能的旅行选择中个人旅行者的选择[9]。具体来说，在PODS中应用Logit选择模型来处理乘客的偏好。诸如重新计划之类的其他因素被替换为无效功能，并添加到给定航班的总成本中。 PODS需要输入订单号。在我们的模拟器中，请求数量将基于历史数据进行预测。此外，我们的模拟器可以根据最新数据调整其固有模型。

强化学习（RL）已成为解决复杂的顺序决策问题的强大工具。 Q学习[12]是一种流行的RL方法，它训练一个查询表以找出代理在每个状态-动作对中获得的预期奖励。 Q学习已用于机票动态定价问题[13] [14]和其他定价问题，例如云服务[10]。然而，使用查找表扩大了训练空间，结果，Q学习方法收敛得很慢。而且，动作空间和状态空间的可能大小受到限制。为了解决这些问题，Deep Q Network（DQN）应运而生。 Volodymyr Mnih等。 Al在Q学习中介绍了一种使用神经网络代替查找表的算法[15]。提出了经验重播和固定q目标方法来解决过度拟合问题。此外，在许多问题中，不仅根据当前状态而且还根据过去的状态来采取行动，因此马尔可夫决策问题变成了部分观察到的马尔可夫决策问题。 [16]提出了使用递归网络的方法来解决这些问题。我们借鉴[16]的想法来实现DRQN模型。

1. 机票购买仿真模型

本章将重点讨论如何对用户的购买行为进行建模，进而构建出用户购买的仿真模型，也即如何构建上文所提到的强化学习模型中的环境。本文通过对市场上可购买的机票以及历史成交订单进行分析，提出了一种结合用户特征和选择模型的用户机票购买仿真模拟模型。该模型可以在给定机票信息以及用户量的条件下，生成拥有独特特征的模拟用户，然后通过模拟用户购买行为，来得出最终的购买量，完成仿真模拟这一过程。所以本章的内容将会从用户特征分析、用户购买决策、带特征的模拟用户生成这几部分进行展开。

* 1. 用户特征分析
     1. 用户特征内容

用户是机票购买行为的重要角色之一，也是本章研究的机票购买仿真模拟模型中的主要角色。在仿真模拟中，尤其是在其中的用户购买决策中，对于用户的特征建模是至关重要的，所以本小节从用户特征出发，分析并讨论哪些用户的特征是我们所关心的。

用户的历史数据由两部分组成，一部分是用户的历史购买记录，另一部分是用户的历史查询信息。历史购买记录反映了用户钟意什么样的机票，而查询记录体现了用户对于机票的选择过程。由此可知，用户的特征必然有一部分可以用机票的特征来表达。



图3-1上海-成都机票搜索结果

Figure 3-1 Flight ticket searching results from Shanghai to Beijing

机票作为一种可以在线出售的商品，每一张机票都有各种各样的属性特征，而每个特征都会影响到最后用户是否会进行购买。例如，如图3-1所示，在国内某大型在线旅行社（后文简称 OTA）的官网上搜索从上海到成都，出发日期为2020年3月1日的单程机票的搜索结果，搜索日期为2019年12月9日。在搜索结果中我们可以看到各种不同的机票以及它们相应的特征，其中包括起飞日期、起飞时间、到达日期、到达时间，起飞和到达的城市与机场，机票价格，航空公司，舱位，退改签政策等等。而在本章的研究中，我们主要的数据源来自该大型OTA的脱敏数据记录。其中就包括了顾客购买的机票信息，基于该行业的专家经验以及我们每个人自己日常生活中购买的经验，我们主要从机票繁杂的信息中提取了一些相对重要的特征信息：

表3 - 1 机票特征以及其含义

Table 4 - 4 Features of Air Tickets

|  |  |  |
| --- | --- | --- |
| **特征** | **意义** | **样例** |
| 购买时间 | 用户进行机票查询的时间 | 2019.12.01 |
| 出发时间 | 用户搭乘航班的出发时间 | 2020.03.01 |
| 到达时间 | 用户搭乘航班的到达时间 | 2020.03.02 |
| 出发城市 | 用户搭乘航班出发的城市 | 上海，厦门 |
| 出发机场 | 用户搭乘航班出发的机场 | SHA：上海虹桥机场 |
| 到达城市 | 用户搭乘航班到达的城市 | 北京，深圳 |
| 到达机场 | 用户搭乘航班出发的机场， | PEK：北京首都国际机场 |
| 航空公司 | 所搭乘航班所属的航空公司 | 东方航空 |
| 舱位 | 表示座位的等级 | 头等，公务，经济 |
| 退改签 | 退票改签政策 |  |
| 托运重量 | 最大允许的行李托运重量 | 20Kg |
| 销售途径 | 机票通过何途径、平台进行销售 | 散团，代理商 |

表3-1展示了行业领域内，同时也是每个用户在日常生活中购买机票最关心的机票特征。大多数特征都是在日常生活中人们比较熟知的概念，例如城市、机场和舱位，除此之外也有一些需要额外说明的特征：

1）机票的购买时间和到达时间：在用户发生一次购票动作时，会有两个时间牵涉其中。机票的购买时间表示当用户完成了机票搜索行为以及购买行为后，切实在OTA平台上预订机票的这个行为所发生的时间，而航班时间则表示用户预订的航班真实起飞的时间，往往会晚于机票的搜索时间。例如：用户A想要在假期飞往成都度假，A在2019年12月01日在OTA平台上搜索了机票，并获得了数量丰富的可供购买的各种机票。A最终购买了一张2020年03月1日出发的机票。那么在这个行为中，购买时间为2019年12月01日而出发时间则为2020年03月01日。

2）销售途径：销售途径特征在绝大多数情况下，并不是用户购买机票时会考虑到的机票特征。事实上，该特征一般被用在机票销售平台的分析中。销售途径体现了一张机票通过何种途径销售给用户。而根据用户的属性和预订机票的方式，他们可以分为五类[18]。这五种类型分别是：

·直客：出于自身旅行、学习、探亲访友等原因进行购票行为的个人，也常被成为散客。直客占据市场大多数份额，特点是对价格敏感程度高、对起落时间要求低，购买频次低，购买数量低。

·散团：由旅行社等经营团体牵头而组成的小型团体，相比直客人数多更多，价格敏感，对时间要求低，购买频次与数量适中。

·商旅：因公务原因需要购买机票的个人或团体，特点是对价格敏感度低，对时间要求高，舱位等级也比其他类型的要高。

·代理：由各大航空公司授权的代理商，可以认为是航空公司对用户的接口，往往会有打折与优惠，同散客一起占据了机票市场的大多数份额。

·分销商：航空公司与用户的中间环节，分销商从航空公司购买机票并出售给个人用户，随着OTA的发展，分销商占据的市场份额已经变得十分低。

由此可知，虽然对于用户而言，销售途径并不重要，然而对于机票用户模拟来说，这一特征却有一定的重要性。不同的销售途径中，机票的特征会有所差异。相应地，每个途径中用户的特征也会有所差异。这也是我们要将该特征作为用户特征建模的一部分的原因。

* + 1. 用户特征模型

上一小节中阐述了机票的各种各样的特征，而基于这些特征我们也可以对用户的特征集进行建模，进而才能够模拟出给定数量的用户群体中每一个用户的特征刻画。事实上，机票中任何一条信息都可以被作为一个特征来进行建模，而在论文 基于配对选择模型的机票个性化推荐 中，提出了在机票领域中，最常用也最能刻画机票的三个特征为：航空公司，舱位，以及价格。在本文的研究中，我们主要针对上述三个特征，再结合该问题中的销售途径特征，对用户的特征进行建模。

1） 航空公司偏好特征：

该特征代表了用户对机票的航空公司属性的喜好程度。对于航司的偏好特征可以用一个单位向量表示，单位向量的每个分量代表着一个航空公司，而该分量的值则代表了该用户对该航司的偏爱程度

其中， 为航司特征偏好，n 为向量的维数，代表了航空公司的总数。例如，市面上一共有四家航空公司：MU, AA, DL, UA。 用户U的航司特征偏好 则表示了用户U最偏好AA航司，最不偏好UA航司，而对MU和DL航司的偏好程度是一致的。

2） 舱位等级偏好特征：

该特征体现了用户对于不同舱位等级的偏好程度。追求享受不受预算限制的用户会更倾向于更豪华的舱位等级而对飞行体验不那么重视同时也受到预算限制的用户将会更倾向于经济舱。在本文讨论的问题中，舱位等级分为三类：头等舱，公务舱以及经济舱。舱位等级的偏好特征向量定义与1）中相似：

假设用户U的航司特征偏好 ，这表明了用户U最偏好经济舱，其次是公务舱，对头等舱则没有任何兴趣。

3） 价格特征：

  价格特征自然是一个非常重要的特征。通常情况下，我们会用价格的绝对大小来作为这个特征的值。然而事实上，价格的绝对高低，受到很多其他因素影响，其中甚至还有与用户本身无关的外部因素。航空里程就是其中之一，用上海飞往北京的机票价格和上海飞往纽约的价格进行比较，来得出用户对于价格的偏好，是十分不合理的。所以我们提出了相对价格来进行特征提取，该特征也被称为价格敏感度：

此功能代表乘客的价格敏感性。由于不同航班的价格范围各不相同，这意味着绝对价格不能表示乘客的价格偏好。在这里，我们引入了一种称为价格敏感度的度量，以衡量人们偏爱便宜机票的程度。对于给定的出发和到达城市对，价格敏感性等于：

其中 和分别代表该城市对中的所有机票中（这些机票飞行里程一致，因此可以消除里程对价格的影响），最高的价格和最低的价格，而p代表的是用户购买的机票价格。这样价格特征就从映射到了，并且敏感度越接近小，代表用户对于价格越不敏感，越不倾向于为价格因素放弃其他条件。

4） 用户类别：

该特征刻画的是上一小节中机票的销售途径，基于上述分析结果，我们将销售途径当做用户的类别。该特征的形式：

其中是一个维度为5的向量，并且以one-hot形式表达，也即只有一个维度上的值为1，其余为0。每个维度相应代表了该用户的类型是直客、散团、商旅、代理和分销商。

结合上述讨论，我们可以定义用户u的特征偏好：

* 1. 基于用户特征与选择模型的机票购买仿真模型

上一节我们重点讨论了如何对用户的特征进行建模，这一节将会基于上述结果，主要讨论我们的仿真模型如何随机生成用户、模拟用户购买，并最终得出用户的购买情况。

* + 1. 机票购买仿真模型系统设计

如图3-2所示，机票购买仿真模型由三部分组成：

1）用户生成模型：主要负责在给定用户规模的情况下，通过对历史数据分析得出用户特征的分布情况，再结合该分布随机生成符合分布的拥有自身特征属性的模拟用户群体

2）用户购买模型：对于1）生成的模拟用户，使用合适的算法，模拟用户的购买行为，也即从若干张备选机票中选择一张进行购买或者选择不够买。

3）参数训练模型：机票仿真模型不论是在训练中还是在实际应用中，都会对一段时间内的数据进行训练和预测，随着时间的不断推进，会不断有新的用户数据、购买数据产生。参数训练模型主要负责在这个过程中，不断自适应地计算出最新的参数，进而更新机票购买仿真模型中的内容。

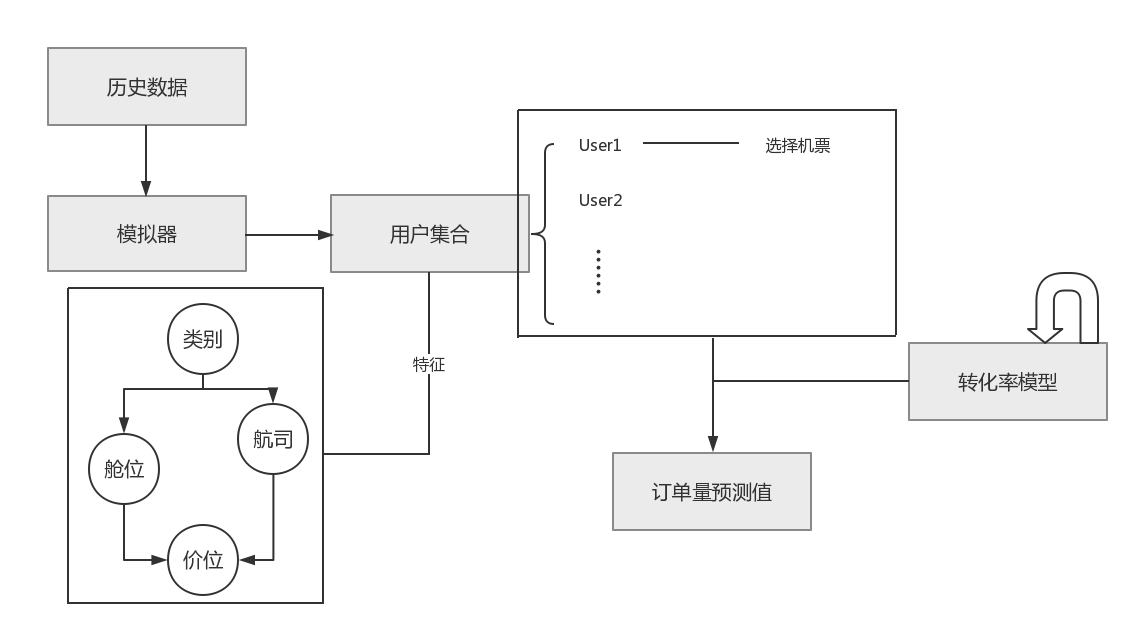


图3-2 机票购买仿真模拟系统设计

Figure 3-2 System Design of Air Ticket Purchasing Simulator

* + 1. 用户生成模型

基于上文对于机票特征和用户特征的分析，我们将对用户的机票购买历史记录，结合对于各个特征分布的分析，学习得到用户的特征分布模型，进而生成带有特征的模拟用户。如图3-2所示，图中左边部分对应着用户生成模型。在生成模拟用户的过程中，模型将会根据概率统计结果，先确定用户的类别，然后再根据类别对应的特征分布概率，生成用户完整的特征。

用户的购买记录和机票是一一对应的，在讨论用户的历史购买记录时，我们假设用户的全集记为**，**任一用户所购买的机票集合记为，那么对于任何一张机票，它的特征模型记为：

其中 皆采用one-hot的形式结合机票的具体信息进行表达。例如，假设一个用户购买的机票是直客类型的票，那么相应的。同样地，假设市面上一共有四家航空公司：MU, AA, DL, UA。而机票是DL航空公司的，那么。

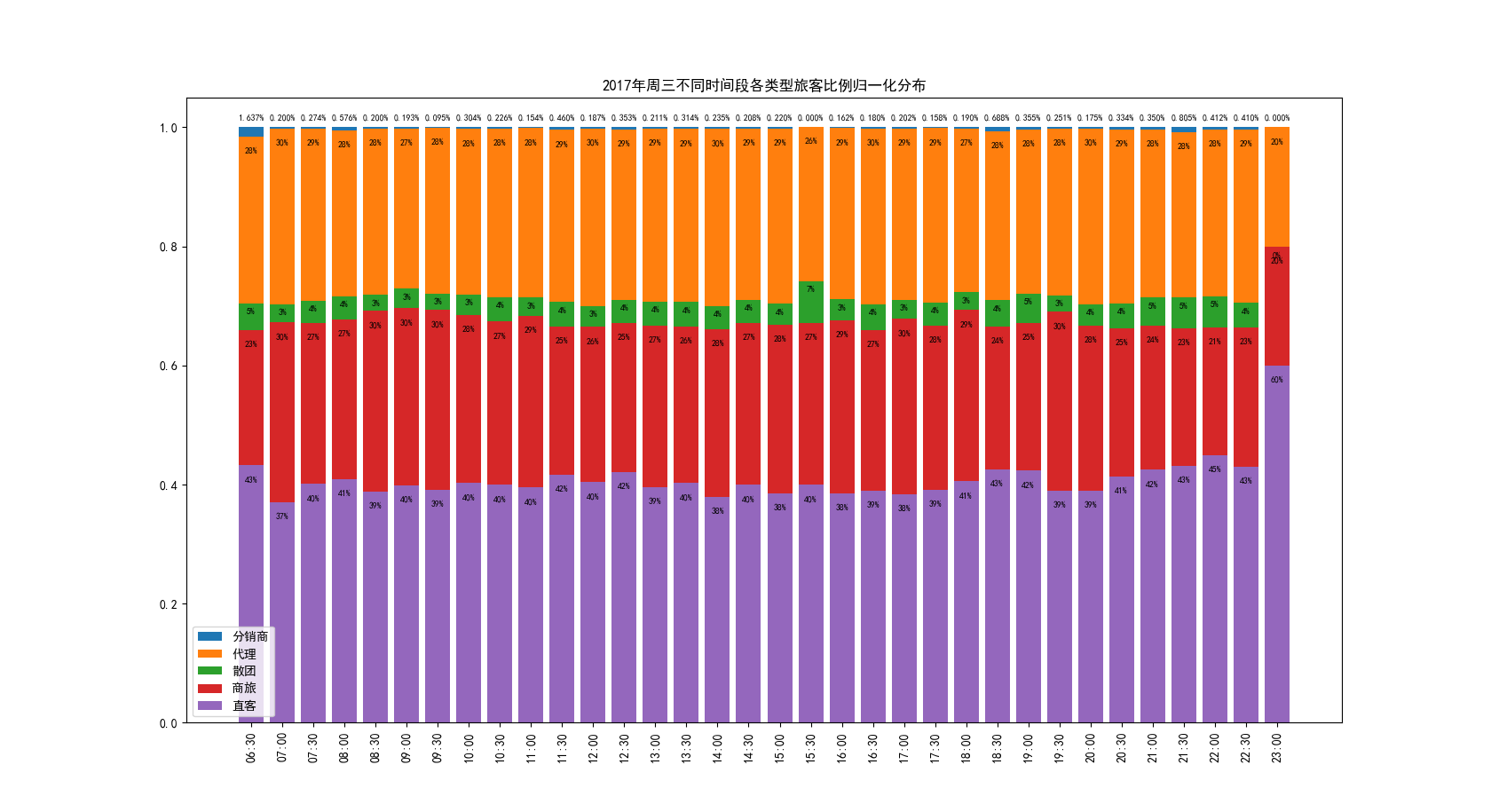


图3-3 2017年某日不同时段机票类型分布

Figure 3-3 Distribution of ticket type in different time in 2017

根据图3-3我们可以发现，在同一天中的不同时段内，五种类型的机票购买数量的分布近乎是一致的。事实上，我们发现在更长的时间范围内，这个现象依然成立，该分布仅仅受到星期几的影响。同一周内，周末的分布和周中不一致，而在不同周内，周中和周中的分布近乎一致，周末与周末的分布也符合这个结论。故我们可以通过机票购买历史记录得出用户类别的概率分布：

其中Norm代表将向量归一化。该结果实际上等于用户类别分别取五个值所对应的概率，例如 那么用户类别为直客和散团的概率皆为10%，为商旅的概率为30%，为代理的概率为50%，而类别为分销商的概率为0.

在确定了用户类别之后，可以进一步的确定用户的偏好航司特征的概率分布。由图3-4可知，对于同一类别的用户，在一周内不同天各航司的购买数量分布。可以看到随着天数的变化，分布的变化较大。所以在生成航司的概率分布时，我们需要对不同的日期进行处理。

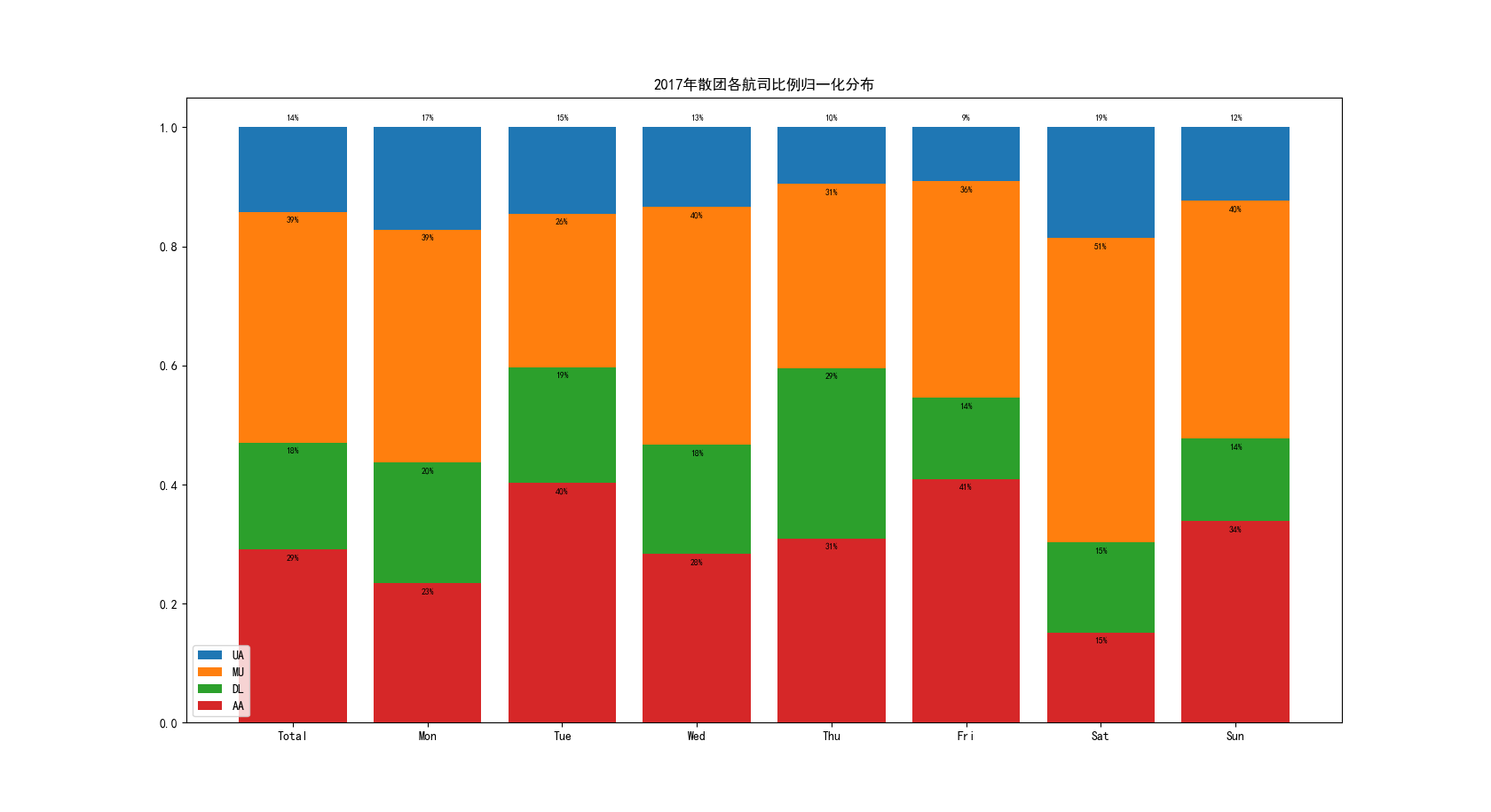


图3-4 2017年不同星期日下机票航空公司分布

Figure 3-4 Distribution of ticket airline in different weekdays in 2017

通过机票购买历史记录，我们得出用户偏好航司的概率分布：

其中为狄利克雷分布，也即Beta分布的多维扩展。

与之类似的，我们也可以刻画用户偏好舱位特征的概率分布，如图3-5所示，我们可以发现，在同一个用户类别下机票的舱位分布基本不受到星期几的影响，并且基本保持一致，故对于偏好舱位特征的概率分布，定义如下：

对于价格特征，我们采取上文所提到的价格敏感的处理方法。如图3-5所示，价格敏感度的处理方式，可以有效地将不同城市对之间不一致的价格范围转化成统一的一致的0到1之间的特征区间。

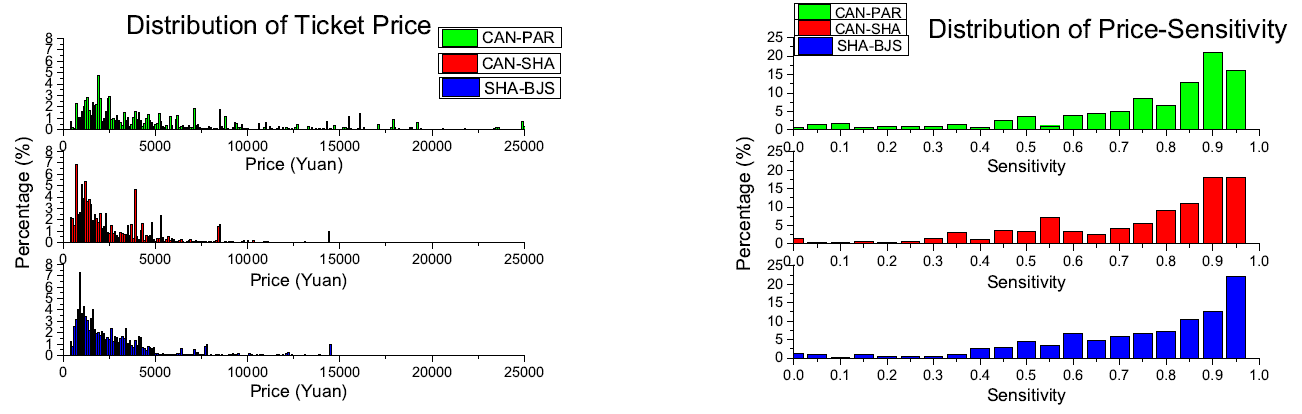


图3-5 价格分布于价格敏感度分布

Figure 3-5 Distribution of price and price sensitivity

* + 1. 用户购买模型

这一小节将在上面的结论下，进一步讨论用户的购买模型。我们想要解决的问题是，如何在给定了用户的特征以及用户可供选择的机票（包含相应特征）之后，模拟用户的决策和购买的行为，并且希望结果尽可能准确。

事实上，用户购买模型可以被看成一个选择问题或者称为选择模型。不仅仅在计算机领域，在经济学中这个问题也被广泛讨论。人们倾向于用效用函数（Utility Function）来对该问题进行建模。在MIT的项目PODS（Passenger Orient Destination Simulator）中，机票的各个特征在用户的眼里被理解为一个有一个的效用（Utility），例如对于一个希望快速到达目的地的用户来说，飞行时间上的节约可以被视为效用，少飞行一小时对他带来的效用可以被同等的看为100元的收益。当他面对两张机票，一张飞行10小时价值1000元，而另一张飞行9小时价值1050元，评估两张机票的效用之后，该用户最终会选择购买第二张机票。相应的，舱位偏好、航司偏好、飞机是否红眼航班等等特征，都可以被视为效用并且加以计算。

而在我们的问题当中，机票和用户都被特征 所表示，我们希望设计出一个效用函数来体现出一张机票被一个用户所喜爱的程度。对于航司和舱位偏好而言，一个很好的方法是通过向量的点乘来体现，因为点乘实际上表示了两个向量之间的夹角，体现了向量相似的程度。而对于价格敏感度这个特征，我们需要的是越靠近效用越高。用特征的差值来刻画是一个解决方法，但是一个更好的方法是使用Sigmoid函数的变体来刻画，因为这可以体现出不同的变化率。因此，对于一个用户而言购买一张机票的效用函数 被定义为：

用户效用函数的值反映了用户对给定机票的满意程度，对于若干张机票，用户将会购买对他而言效用函数取值最高的一张机票，这十分符合逻辑。

考虑到现实生活中真实发生的情况：我们在做选择的时候并不会永远维持理性，偶尔会发生一些非理性的行为，所以我们可以在效用函数中增加一个噪音以体现不确定性。可以理解为噪音，是一个满足以0为均值，1为方差的正态分布的随机变量。这样我们就得到了修正过的效用函数：

进一步，不进行购买也是一种购买行为。从效用函数的角度看出发，如果一个用户能够选择的所有机票对他带来的效用都很低，他理应不进行购买。我们定义了最低购买效用阈值，任何效用函数低于该值的机票都不会被购买。

这样用户的购买模型就被一组参数 所确定，通过历史数据中的机票信息、用户搜索量以及用户购买量作为模型的输入和输出，我们可以对参数进行训练找到最符合真实情况的模型参数。

* + 1. 机票购买率模型

在现实生活场景中，OTA平台只能收集到用户对于机票的搜索量，也即有多少用户进行了机票的搜索。而上述讨章节中讨论的是对于有意愿进行购买的用户，如何模拟他们的特征以及购买的决策行为。然而在行业领域中有一个长期存在的现象，那就是搜索量和真正打算购买的用户的量并不相等。领域内专家用购买率来表达有购买意愿的用户和进行过搜索的用户的数量的比值。出于很多原因，例如很多用户只是抱着随便看看的态度进行搜索，部分搜索量产生的原因仅仅是需要和其他产品进行对比，甚至还有些搜索量是来自网络爬虫对于机票信息进行爬取，导致真正有意愿购买的用户数量变得不可知。这里将对机票购买率进行讨论，首先我们要明确几个定义

1）机票的搜索日期：在一次机票搜索行为中，行为发生的真实日期就是机票的搜索日期。这里我们记为。

2）搜索出发的日期差：在一次机票搜索行为中，用户想要购买的机票出发的日期，和这个搜索行为发生时的真实日期之间的天数差，称为日期差，记为。一般来说，越大，代表顾客正在搜索越远期的机票，相应的购买率就会下降。相反，越小，代表着搜索行为越紧急，购买率也会上升。

3）机票的出发日期：在机票搜索行为中，用户想要购买的机票的出发日期。显然，出发日期 = 。

对于用户而言，虽然可以搜索非常非常未来的机票，但是这显然没有意义。所以我们认为当日期差大过一个阈值之后，对应的购买率都将变为0。这个阈值记为。在一个搜索行为中，对于搜索日期和日期差，我们对购买率线性建模：

其中代表搜索日期为，出发日期为天之后的购买率。是购买系数，可以根据实际订单数据进行推导。

购买系数体现了在所有的搜索行为中有多少比例的搜索行为是发自有购买意愿的个体。购买系数是一个随着时间不断变化的量，我们可以通过自回归的方法来预测未来的购买系数:

其中为常数，表示第t天的购买系数。

* + 1. 机票购买量预测

结合本章之前几小结所讨论的内容，我们的机票购买仿真模型的几个子系统都已经做了非常详细的阐述。这一节将会回到模型的最顶层，来阐述模型是如何运转的。

首先，机票仿真购买模型的输入有：

1）用户的历史购买订单信息

2）可供用户购买的机票信息

3）需要模拟的用户数量

而我们的仿真模型要根据这些输入，随机生成出满足条件的模拟用户，最终预测在这些条件下，会产生多少购买量。

对于一个用户，可供选择购买的机票集合为，那么最终购买的机票数量为

如果可够买机票中效用函数的最大值超过了阈值，则会产生一个购买，反之则为0个。

对于整个仿真模型而言，关心的是最终的购买量，发生在搜索日期的购买量为：

其中为购买率日期差阈值，为购买系数，为在进行搜索日期为，日期差为的搜索的用户集合。在每天仿真模拟预测的过程中，购买系数采用的是预测值，通过式（3-11）进行预测。而当一天结束，收集到了真实的购买量之后，我们可以更新真实的购买系数：

然后再根据式（3-11）预测出下一天的购买系数，准备好下一天的机票购买仿真。

* 1. 基于生成对抗的用户决策模型

在上一小节中我们提出了一种基于效用函数的用户购买决策模型。该模型的优点是能够结合用户和机票的特征进行建模，结合能够真实反映用户心理活动的效用函数来对购买行为进行刻画。模型并不复杂，训练过程也相对容易，只需通过拟合找到一组参数即可。

然而这个模型也有一定的缺点。首先，该模型由于对价格敏感度的建模是基于经验给出的公式，所以在价格特征上的准确程度有待商榷。同时，该模型对于价格调整带来的变化，并没有进行充分的讨论。最后，该模型的训练过程仅仅是对于效用函数中的参数进行训练，对于用户和机票的特征值没有做任何的修正。

事实上，在这个问题中，我们有用户每一次进行搜索的记录。详细地说，对于每一个用户的每一次搜索，他能看到多少不同的机票，并且最终从这些机票中选择了哪一张进行购买，或者放弃购买。我们希望通过这些数据来训练得出用户购买的模型和用户对机票特征如何进行分析的模型。事实上，这两个模型是可以一起被训练的。在这一小结中我们将会做详细的阐述。

* + 1. 用户选择模型

用户，在现实生活中对应着一个个真实的人。人在做选择的时候并不是盲目的，而是趋利避害的。与此同时，在一段时间范围内，人做选择时所考虑的因素也会不一样，人不是一成不变的。基于这两个事实，我们在对用户行为进行建模的过程中，将会考虑到下述两个重要机制：

1）用户是会主动追求更高的回报的：正如上文所讨论的，对于一个用户而言，当他被给与了k张机票进行购买选择的时候，他一定会选择使得他效用最高的一张机票。这里我们将效用值用另一个概念记录，奖励（reward）r。奖励r衡量了一个用户对于一张机票的满意程度。当然，用户也可以选择一张机票也不购买，相应的，他就能收获到不做购买所带来的奖励。

2）用户的偏好随着时间而不断变化：考虑这样一个真实存在的例子，一个用户可能一直都搭乘航空公司A的航班，可能对航司B完全不感兴趣。然而在一次偶然的机会下，该用户搭乘了航司B的航班并且体验非常好，从此对航司B产生了很大的喜好。同样的，用户也会因为经常做航司B，渐渐发现了航司B的一些问题，于是之后对航司B的航班兴趣降低。因此，我们需要对用户的状态进行建模，用户每完成一次搜索行为，都会通过他的历史行为来更新他的用户状态。

为了同时考虑到上述的这些因素，我们提出了一个用户选择模型的生成方法。在这个模型中，我们即会考虑到用户选择的机票，也会考虑到用户的历史状态。该模型可以类比为上一小节中的用户效用函数。对于一个用户，我们会考虑一段时间内他所有的搜索和购买行为，记每一个阶段为，在t时刻用户的选择记为，如果用户的行为是购买了第i张机票，那么可以记为。特别的，如果用户没用进行购买，可以额外定义一个行为a，用来代表没有购买这个行为，也即不购买别认为是一种特殊的购买行为。用户的状态记为，相应的用户在这样的条件下选择了机票所能获得的奖励记为。同时，我们将在时刻t用户可供选择的所有机票的集合记为，这些机票的特征相应的记为，其中k为可供购买的机票数量。而用户选择模型实际上就是如何从中选择，进而用户的选择模型实际上就是在这k个行为上的一个概率分布，我们可以将其记为，为概率单纯形。考虑到用户的购买行为总是以最大化自己的奖励为目标，那么用户的选择模型应该满足：

其中代表了在策略下，获得的奖励的期望。由于是一个概率单纯形，或者说实际上表示了选择各个行为的概率，所以这里最大化的其实就是一个期望奖励。为正则项，正则项的加入是为了避免过于稀疏。如果不加上正则项，那么用户的选择模型将会退化成一个在行为 上为1，其他全为0的one-hot向量。映射到实际生活中，就是用户永远没有可能去选择现在这个状态下看起来奖励并不是最高的行为，然而该行为可能会给用户带来更好的未来的奖励。所以这里的正则项，实际上也充当了探索未知的功能。而则控制了正则的强度，越大，正则惩罚就越小，就会导致结果越稀疏。关于正则项，我们进行更深入的讨论：

1）L1正则化：L1正则化是十分常见的一种正则方法，不仅在我们的问题中，在其他机器学习问题中也经常被使用到。假设，那么对于我们的用户模型，L1正则化可以被表示为：

由此可见在我们这个特殊的问题中，L1正则化并不能起到实质性的作用。

2）L2正则化：L2正则化也是一种十分常见的正则方法，L2正则化和L1正则化不同，相比于L1正则化有着更好的特征筛选的功能，L2正则化的结果比起L1不会那么稀疏，而是会更加分散一些。L2正则化可以被表示为：

3）信息熵正则化：另一种可以用于正则惩罚的函数是香农信息熵（Shannon Entropy），香农信息熵一般用来衡量一个系统的混乱程度，定义如下：

其中U为一个系统，为系统成为第i种状态发生的概率。当然我们可以看到，一个系统的不确定性越大（也即都相等），H(U)就会越大，这可以由琴声不等式（Jensen Inequality）来证明，而当系统的不确定性越小（例如，其余全为0），那么系统的信息熵H(U)就会越小。显然在我们的问题中，我们加入正则项的目的是为了让用户的选择模型的确定性降低，所以我们的正则项将会取用负的香农信息熵：

**引理3.1.** 当使用负香农信息熵作为正则项时，用户的选择模型拥有一个闭式解

**证明：** 证明过程并不复杂，该问题本质上是求解一个带约束条件的多元函数的极值。我们只需要使用拉格朗日乘子法就可以进行求解。

设，并且，拉格朗日函数为：

对和求导，可以得到：

通过（3-22）式可以解得：

代入（3-23）式可解得：

故**引理3.1.** 得证。

事实上，并不是所有的正则惩罚项，我们都可以求解出闭式解。而闭式解将会使得问题的讨论更加简化。所以在我们的实验中，将会主要用负信息熵作为正则惩罚项。我们依然可以按照上一小节中所作的，在用户模型中加入一个随机变量，，来刻画用户的一些随机行为。而从式（3-20）中我们可以进行一些进一步的分析。首先我们可以看到用户选择第i张机票的概率 的大小和用户购买这张机票能获得的奖励是正相关的，这一点符合我们之前的假设。对于正则系数，我们来分析他对于用户选择模型的影响。考虑两张机票对应的两个行为，用户选择i和j的概率比例为。可以看到，正则系数越小，用户选择i和j的概率比例就会越小，用户就会更大的可能性去尝试和探索。在我们的实验中，我们将正则系数的值设为1.

基于上树的讨论与分析，我们得到了一个更为合理的用户选择模型，该模型能够表现出用户在真实生活中的特性。而该模型相比上一节中的模型更为灵活，我们不仅可以调节正则系数来改变用户对于尝试新机票的倾向程度，还可以通过调节正则项来刻画不同的用户选择逻辑。用户模型的训练也变得更加合理，通过最大化期望奖励来确定模型。只需要得到用户的奖励模型和用户的状态模型，就可以训练出用户的选择模型。

* + 1. 用户奖励模型与状态模型

前文已经描述了如何对用户的选择模型进行建模，而选择模型的训练，需要用户的状态模型和奖励模型作为基础。所以这一节将会主要对奖励模型和状态模型进行建模。事实上，奖励模型和状态模型是息息相关的，至少我们期望能够将它们建模成相关的模型。一个合理的解决方案是通过用户的历史购买记录中机票的额特征来刻画用户当前的状态，结合当前将要购买的机票的特征来生成用户购买某张机票的奖励函数。

首先我们对用户的状态

数据。行为模型试图模仿动作

由实际用户提供的序列，该行为可以最大程度地发挥作用

她的奖励函数类似于生成对抗

网络，（i）充当生成器

用户根据其历史记录进行的下一步操作，以及（ii）

r充当区分器，试图区分

用户的实际行为（由行为产生的行为）

模型。因此，受GAN框架的启发，我们估计

和r同时通过mini-max公式表示。

更准确地说，给定T个观察到的动作的轨迹

fa1

真正; a2

真正; ：：：;在

用户及其对应的信息

点击项功能ff1

; f2

; ：：：; T

g，我们学习r和

通过解决以下最小-最大优化联合

生成式对抗训练：

分

</ s> </ s> </ s>

最高

􀀀

Ë

PT

t = 1r（st

真正;在）

</ s> </ s> </ s>

􀀀R（）=

</ s> </ s> </ s>

􀀀

PT

t = 1r（st

真正;在

真正）; （5）

我们在哪里使用st

强调指出这一点是正确的

数据。通过以上优化，可以看到

奖励r将从实际用户中提取一些统计信息

操作并为用户操作建模，并尝试放大其

差异（或使它们的负差距更大）。相反，

用户模型将尝试缩小差异，

因此更类似于真实的用户行为。或者，

最小最大优化也可以解释为

作为对手和学习者之间的游戏，

对手试图通过调整来最小化学习者的报酬

r，而学习者试图通过

进行调整以抵消对抗动作。这给

用户行为培训过程中的大幅度培训

风味，我们甚至想为

最坏的情况。

对于一般的正则化函数R（），最优解

在等式中（5）没有封闭形式，通常

需要通过交替更新和r来解决。

8 <

：

+

1r E

PT

t = 1

一世

􀀀

1r R（）=;

􀀀

2E

PT

t = 1 r r

一世

+

2

PT

t = 1 r r：

（6）

由于非凸性，该过程可能不稳定

问题。为了稳定培训过程，我们将

利用特殊的正则化来初始化培训

处理。更具体地说，对于熵正则化，我们

可以获得内在最大化的封闭形式解

用于用户行为模型，这使得学习奖励

功能简单（请参阅下面的引理2和附录A

证明）。一旦奖励了熵的奖励函数

正则化，可用于初始化学习

其他可能归纳为正则化函数的情况

* 1. 实验结果分析
  2. 本章小结

本章主要完成了对机票购买仿真模型的建模与实现。本章从机票和用户的特征出发，结合实际领域内专业知识和科学的计算机算法和概念，对每张机票和每个用户进行了特征模型的刻画。基于此，本章应用概率分布模型提出了模拟用户的生成算法。本章通过两种方法对用户的购买行为进行分析研究，一种是较为朴素的基于定式的效用函数模型，另一种是基于对抗生成的用户模型和决策模型都可以博弈学习的选择模型。最后本章提出了仿真模型的架构设计，包括如何计算出预测的购票数量、如何对购买率进行建模、如何预测购买率以及模型中系数的更新算法。在本章最后，通过相关的实验，进一步对购买模型的准确性以及机票购买仿真模型的准群度进行测试，评估以及分析。

本章讨论的内容为强化学习算法的研究打下了坚实的、必不可缺的基础。

1. 带约束的top-K路径搜索算法
   1. 引言

， 路径推荐算法

* 1. 引言

1. 全文总结与展望
   1. 本文工作总结

参考文献

[]Ctrip. [cited 2016 Dec]; Available from: <https://www.ctrip.com/>.

[2]Qunar. [cited 2016 Dec]; Available from: <https://www.qunar.com/>.

[3]Triplehop’s TripMatcher.[cited 2016 Dec]; Available from: <https://www.ski-europe.com>.

[4]Resnick, P. and H.R. Varian, Recommender systems. Communications of the ACM, 1997. 40(3): p. 56-58.

[5]Ricci, F., L. Rokach, and B. Shapira, Introduction to recommender systems handbook.

SIGIR Forum. Vol. 33. No. 1. ACM, 1999.

致谢

值此论文完成之际，谨向在两年半研究生学习阶段所有教导、关心、帮助过我的人表示最诚挚的感谢。

首先，感谢我的导师曹健教授。从研一到现在，无论科研还是生活，曹老师给予我很多指导和鼓励，在毕业论文的整个研究过程中，曹老师一直悉心教导、指点，当研究和实验陷入低谷时，总是给予帮助，每周思辨性的讨论为我提供开拓思维，学会如何从多方面考虑问题和解决问题，最后成功的化解瓶颈，让我受益匪浅。曹老师对科研的认真谨慎负责任的态度令我钦佩，生活中平易近人且不失风度，这些也是我今后在工作、生活和学习中应有的态度。

同时，也非常感谢已经毕业的陈誉文和杨芳洲学长。项目研究中，在他们的指导下逐渐理解科研方法，每周的交流使得科研目标更加明确，同时也要感谢赵亚峰同学，在同一项目组中共同成长、互帮互助，在科研实验中也给了我很大帮助，使得论文中实验才可以进展如此顺利。还要感谢姚艳博士、冯珊珊博士、张棪等同学，让我再SJTU-CIT的大家庭中度过最美好的研究生时光，是我莫大的荣幸。

最后，由衷的感谢一直陪伴的父母、家人和朋友，谢谢你们的支持和关心，在困难的时候给予我坚强的后盾，这也是我以后在工作中前进的最大动力。

附录一 符号与标记

|  |  |
| --- | --- |
|  | 极小值 |
|  | 极大值 |
|  | 平均值 |
|  | 连加运算符 |
|  | 积分运算符 |
|  | 集合A中元素个数 |
| # | 数量 |
|  | 以k为底，m的对数 |

附录二 英文缩略语表

|  |  |  |
| --- | --- | --- |
| OTA | Online Travel Agent | 在线旅游代理 |
| CF | Collaborative Filtering | 协同过滤 |
| KMCSP | K Multiple Constraints Shortest Paths | top-带约束最短路径问题 |
| KSPG | K Shortest Path Generator | 带约束的启发式路径搜索算法 |
| IERA | Information Entropy Recommendation Algorithm | 基于信息熵的路径推荐算法 |
| KNN | K-Nearest Neighbors | K最近邻模型 |
| FIFO | First In First Out | 最先进入最先换出 |
| LRU | Least Recently Used | 最近最少使用页面置换 |
| LFU | Least Frequently Used | 最近最不常用页面置换 |
| FBR | Frequency-Based Replacement | 基于访问频率页面置换 |
| MAE | Mean Absolute Error | 平均绝对误差 |
| MSE | Mean Square Error | 均方误差 |
| MAP | Mean Average Precision | 平均准确率均值 |
| BJS | Beijing | 北京城市代码 |
| SHA | Shanghai | 上海城市代码 |
| CAN | Guangzhou | 广州城市代码 |
| XMN | Xiamen | 厦门城市代码 |
| TSN | Tianjin | 天津城市代码 |
| SYX | SanYa | 三亚城市代码 |
| CKG | ChongQing | 重庆城市代码 |

攻读硕士学位期间已发表的论文

**已录用的论文：**

1. 华逸群 and 曹健. "基于模糊时间序列的机票价格预测." 小型微型计算机系统 (2015). （第一作者）