## 实习项目：餐饮智能推荐服务

[实习项目：餐饮智能推荐服务 1](#_Toc3515)

[一、 项目背景介绍 2](#_Toc24393)

[1.行业背景 2](#_Toc18317)

[2. 分析目标 3](#_Toc8935)

[二、 数据说明 3](#_Toc5014)

[1.数据来源 3](#_Toc15297)

[2.数据表介绍及其数据说明 3](#_Toc3404)

[3. 数据量大小 4](#_Toc20597)

[三、 数据处理 4](#_Toc24013)

[1. 数据处理任务流程 4](#_Toc30398)

[2. 结果展示 4](#_Toc22706)

[四、 热销度分析 5](#_Toc7536)

[1. 热销度分析及任务流程 5](#_Toc31897)

[2. 结果展示 6](#_Toc20322)

[五、 智能推荐模型 7](#_Toc23781)

[1. 模型构建准备 7](#_Toc24241)

[2. 模型构建及结果 8](#_Toc26238)

[3. 模型评价 10](#_Toc17145)

## 项目背景介绍

### 1.行业背景

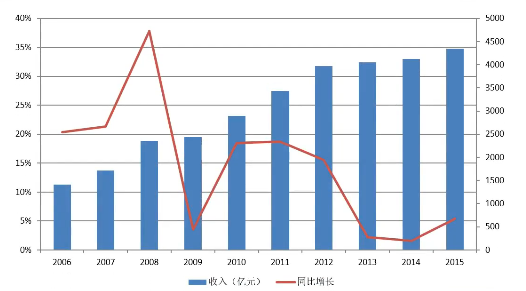
餐饮行业作为我国第三产业中的一个传统服务性行业，始终保持着旺盛的增长势头，取得了突飞猛进的发展，展现出繁荣兴旺的新局面。与此同时，我国餐饮业发展的质量和内涵也发生了重大变化。根据国家统计局数据显示，餐饮行业餐费收入从2006到2015年都处于增长的趋势，但是同比却又很大的波动。如下图1所示随着年份增加餐费收入净额实现较大幅度增长，但同比增长却在部分年份实现了负增长且增长量并不稳定。

如果仔细分析原因，餐饮企业正面临以下问题。

1. 同行竞争压力大，由于餐饮业也伴随着数量增多和网红点等创新型餐厅的流行而导致传统行业在竞争方面存在恶性激烈竞争。
2. 对消费者的需求认识不到位，客户忠诚度较低，传统餐饮业忽略了客户产品推荐等创新型手段吸引消费者，导致消费者体验感较差，且传统餐饮业没有主动展开来自消费者对于餐饮的满意度与回馈调查，部分导致消费者回头率较低，并使企业对客户把握度较低。
3. 菜品和服务缺乏个性化，菜品口感与菜品名称与其他竞争企业较为类似，产品菜品同质化现象严重，服务以日常端茶倒水等服务为主，忽略了消费者的心理，并使企业吸引力不足。

而目前解决方法可以借助互联网电商展开思路拓展，以淘宝京东等互联网电商企业为例，它们借助自身平台打造相似餐饮产品以及其他产品的类似相关物品推荐，从而使得个性化实现度较高，造成交易量与营业额的同比化增加，将其餐饮职能推荐则对传统餐饮业的转型和保住新兴餐饮业发展势头具有重大意义。

图1 餐饮业餐费收入增长图



### 分析目标

本项目对某餐饮企业的订单数据进行分析，利用协同过滤算法对就餐的客户进行个性化菜品智能推荐

## 数据说明

### 1.数据来源

项目数据某餐饮企业的系统数据库中积累了大量的与客户用餐相关的数据，包含菜品详情表，订单表，订单详情表等。数据来源时间段：2016年8月1日至2016年8月31日

### 2.数据表介绍及其数据说明

项目数据收集在订单表(meal\_order\_info)和详情表(meal\_order\_detail)中，具体数据说明如下表1所示

表1 数据说明

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 含义 | 名称 | 含义 |
| info\_id | 订单ID | lock\_time | 锁单时间 |
| emp\_id | 客户ID | cashier\_id | 收银ID |
| number\_consumers | 消费人数 | pc\_id | 终端ID |
| mode | 消费方式 | order\_number | 订单号 |
| dining\_table\_id | 桌子ID | org\_id | 门店ID |
| Dining\_table\_name | 桌子名称 | Print\_doc\_bill\_num | 打印doc账单编码 |
| expenditure | 消费金额 | Lock\_table\_info | 桌子关闭信息 |
| Dishes\_count | 总菜品数 | Order\_status | 0：未结算，1：结算,2:已锁单 |
| Accounts\_payable | 付费金额 | phone | 电话 |
| Use\_start\_time | 开始时间 | name | 名字 |
| Check\_closed | 支付结束 |  |  |

### 数据量大小

订单表(meal\_order\_info)：945条记录

详情表(meal\_order\_detail)：10037条记录

## 数据处理

### 数据处理任务流程

图2 任务流程图

统计每个订单状态的占比

选取有效的订单数据

选取主要特征

1. 统计每个订单的占比：因为在order\_status的订单状态中，部分数据会出现异常指标(当数值为2和0时代表订单状态异常)，故需要统计出order\_status为1的订单数，而具体的占比情况如表所示:

表2 订单状态占比情况

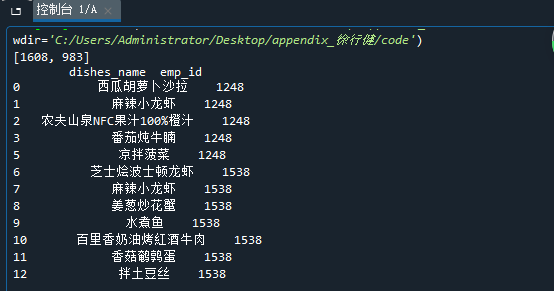
|  |  |  |  |
| --- | --- | --- | --- |
| 订单状态 | 0 | 1 | 2 |
| 占比 | 0.0095 | 0.9873 | 0.0032 |

1. 选择有效的订单数据：根据常识和数据可知，客户可能出现在订单详情表中有菜品订单记录但在订单表中不存在的情况，所以需要将该部分多出来的订单记录清除
2. 选取主要特征：关于智能推荐模型的特征，只选择菜品名称和用户ID两种特征即可完成操作，故提取该两种特征即可，之后并根据订单情况去除菜品名称为白饭这样的非特征性影响模型的菜品名称数据

### 结果展示

数据初步处理好的表格及部分值的情况如下图3所示

图3 运行表格结果表头及12个数据展示



## 热销度分析

### 热销度分析及任务流程

图4 任务流程图

读取并探索订单数据

处理菜品名称字符数据

绘制条形图展示热销图

构建热销度评分指标

(1)读取并探索订单数据:运用pandas读取文件数据，选取主要特征，观察特点。

(2)处理菜品名称字符数据：因为在pandas读取主要特征数据后，发现部分菜品名称数据存在大量分隔符，故在进行下一步前需要清洗处理掉多余字符数据

(3)构建热销度评分指标:热消毒即在一定时间类产品销量的程度，根据餐饮企业近30天（即2016年8月1日至2016年8月31日）的菜品销售统计每个菜品的热销度，其计算公司如下所示。经Min-Max标准化后计算得分，最高分为1分，最低分为0分。

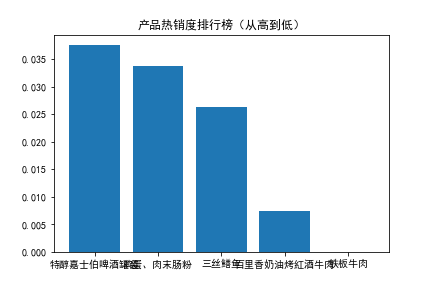


其中，热销度评分为某项菜品的热销度评分，其值范围在0到1之间：为某项菜品的销售份数；为该餐饮企业最近30天内有销售记录的菜品中的最大销售份数；为该餐饮企业最近30天内有销售记录的菜品中的最小销售份数。

1. 绘制条形图用以展示热销图:根据热销度排序，选出热销排名前5名和后5名数据，观察情况。

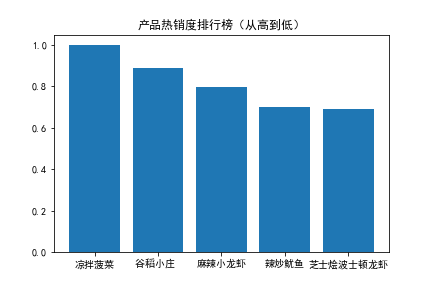
### 结果展示

图5 产品热销度后5名排行榜



如上图5所示，产品热销度排行榜后5名的是特醇嘉士伯啤酒，肉末肠粉，三丝鳝鱼，百里香奶油烤红酒牛肉和铁板牛肉，也可说明该餐饮店需要对后5名的产品进行计划调整，可以选择减少供应或更换新菜品替代等方式处理销量不足的问题

图6 产品热销度前5名排行榜



如上图6所示，产品热销度排行榜前5名的是凉拌菠菜，谷稻小庄，麻辣小龙虾，辣炒鱿鱼和芝士烩波士顿龙虾，也可说明该餐饮店可以对前5名的产品加大供应或提供菜单热销量推荐等方法加大该部分餐饮产品的销量。

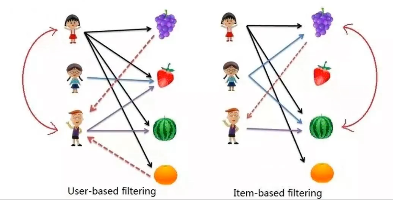
## 智能推荐模型

### 模型构建准备

其模型构建需要进行以下步骤

1. 将模型数据划分为训练集和测试集：用于检测数据模型准确率时使用，而同时为了运算方便将客户ID处理为只保留点了3种菜以上的客户ID
2. 构建数据透视表下的训练集数据客户-菜品二元矩阵，观察其特征

图6 协同过滤算法介绍图



1. 运用协同顾虑推荐算法模型原理处理数据：协同过滤推荐算法的原理是找到相似的用户或物品进行推荐，最常用的是基于用户的协同过滤算法(base collaborative filtering)和基于物品的协同过滤算法(item-based collaborative filtering)，而本项目适用于基于物品的协同过滤算法，个性化推荐主要分为两步：计算菜品于菜品之间的相似度和根据菜品相似度和客户的历史行为给客户生成推荐列表，而计算相似度存在以下方法，最终我们选择Jaccard相似度方法计算以便于简便运算过程：
2. 欧几里得距离



1. 皮尔逊相关系数



1. Jaccard相似度



之后生成的推荐列表将采用下列计算推荐算法中目标客户对所有菜品的感兴趣程度，并根据感兴趣程度给目标客户生成推荐列表。



其中P为目标客户对所有菜品的感兴趣程度，SIM为所有菜品之间的相似度，R代表了客户对物品的心趣。在原始数据中之记录了客户用餐之后的订单，说明客户的行为是用餐与否，并没有类似电子商务网站上的购买，评分和评论等行为，因此R只存在0和1

### 模型构建及结果

1. 客户-菜品二元矩阵展示

由于矩阵的数据量庞大，只截取到测试集的一部分情况，如下表所示，用该种图的好处在于可以将文本型数据转化为数值型数据以便于之后操作

表3 客户-菜品二元矩阵

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| emp\_id/dishes\_name | 42度海之蓝 | 北冰洋汽水 | 38度剑南春 | 50度古井贡酒 | 52度泸州老窖 | 53度茅台 | 一品香酥藕 |
| 984 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 986 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 987 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 988 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 989 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 996 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 997 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 998 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1001 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1004 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1006 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

1. 训练集测试集数据基本情况

关于训练集和测试及的数据基本情况如下表4所示

表4 数据基本情况

|  |  |  |
| --- | --- | --- |
|  | 测试集 | 训练集 |
| ID数量 | 374 | 93 |
| 总记录数量 | 7517 | 1988 |

1. 训练集推荐列表及排序数值

表5 训练集推荐列表实例

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 客户ID | 菜品名称 | 排序分数 | 客户ID | 菜品名称 | 排序分数 |
| 1272 | 凉拌菠菜 | 1.0692122196494607 | 1401 | 焖猪手 | 1.3894581280788176 |
| 五色糯米饭(七色) | 1.024120939042756 | 五色糯米饭(七色) | 0.9195032998954566 |
| 水煮鱼 | 0.6706906788046705 | 西瓜胡萝卜沙拉 | 0.2724 |
| 自制猪肉脯 | 0.3259668508287293 | 辣炒田螺 | 0.22972972972972974 |
| 蛋挞 | 0.272 | 意文柠檬汁 | 0.22764227642276422 |
| 蒙古烤羊腿 | 0.2569444444444444 | 清蒸蝶鱼 | 0.20212765957446807 |
| 剁椒鱼头 | 0.2440944881889764 | 番茄炖牛腩 | 0.17204301075268819 |
| 爆炒双丝 | 0.23275862068965517 | 打卤面 | 0.16363636363636364 |
| 香烤牛排 | 0.22018348623853212 | 芹黄鳝丝 | 0.23529411764705882 |
| 爆炒鸡翅 | 0.20987654320987653 | 玉竹南北杏鸭腿汤 | 0.15555555555555556 |
| 培根花菜 | 0.1875 | 重庆特色油烧兔 | 0.15555555555555556 |

则可以通过表6看出对于ID为1272的客户，餐饮企业可以选择着重推荐凉拌菠菜，五色糯米饭(七色)和水煮鱼三种菜，而对于ID为1401的客户，餐饮企业可以选择着重推荐焖猪手，五色糯米饭(七色)和西瓜胡萝卜沙拉三种菜品。

1. 测试集推荐列表及排序数值

表6 测试集推荐列表实例

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 客户ID | 菜品名称 | 排序分数 | 客户ID | 菜品名称 | 排序分数 |
| 1146 | 木须豌豆 | 1.2152777777777777 | 1522 | 凉拌菠菜 | 0.42857142857142855 |
| 自制猪肉脯 | 0.799749687108886 | 白胡椒胡萝卜羊肉汤 | 0.3888888888888889 |
| 谷稻小庄 | 0.39215686274509803 | 番茄炖牛腩 | 0.3333333333333333 |
| 辣炒鱿鱼 | 0.38461538461538464 | 山药养生粥 | 0.32 |
| 芝士烩波士顿龙虾 | 0.38461538461538464 | 冰糖红豆薏米粥 | 0.3125 |
| 杭椒鸡珍 | 0.375 | 焖猪手 | 0.29411764705882354 |
| 番茄炖牛腩 | 0.3333333333333333 | 蛋挞 | 0.28 |
| 焖猪手 | 0.3333333333333333 | 啤酒鸭 | 0.25 |
| 黄花菜炒木耳 | 0.3333333333333333 | 芹黄鳝丝 | 0.23529411764705882 |
| 重庆特色油烧兔 | 0.3333333333333333 | 自制猪肉脯 | 0.22857142857142856 |
| 炝炒大白菜 | 0.32142857142857145 | 爆炒双丝 | 0.22727272727272727 |

可以根据对于训练集的推荐情况作出关于测试集的推荐情况，可以看出对于ID为1146的客户，餐饮企业可以选择着重推荐木须豌豆，自制猪肉脯和谷稻小庄三种菜，而对于ID为1522的客户，餐饮企业可以选择着重推荐凉拌菠菜，白胡椒胡萝卜羊肉汤和番茄炖牛腩三种菜品。

### 模型评价

1. 构建测试机客户IP-菜品字典,展示如下:

{'1429': {'香酥两吃大虾': 1.0, '自制猪肉脯': 1.0, '水煮鱼': 1.0, '剁椒鱼头': 1.0, '小米南瓜粥': 1.0, '凉拌蒜蓉西兰花': 1.0, '红酒炖羊肉': 1.0, '凉拌菠菜': 1.0, '皮蛋瘦肉粥': 1.0, '冰糖红豆薏米粥': 1.0, '辣炒鱿鱼': 1.0, '大理石奶油蛋糕': 1.0, '南瓜枸杞小饼干': 1.0, '番茄炖牛腩': 1.0, '紫薯面包卷': 1.0, '蒙古烤羊腿': 1.0, '核桃葡萄干土司': 1.0, '清爽拌凉面': 1.0, '鱼香肉丝拌面': 1.0, '独家薄荷鲜虾牛肉卷': 1.0, '五色糯米饭(七色)': 1.0, '黑米恋上葡萄': 1.0}, '1522': {'麻辣小龙虾': 1.0, ' 北冰洋汽水 ': 1.0, '桂圆枸杞鸽子汤': 1.0, '小米南瓜粥': 1.0}, '1338': {'芝士烩波士顿龙虾': 1.0, '倒立蒸梭子蟹': 1.0, '冰镇花螺': 1.0, '快炒黄鳝': 1.0, '爆炒双丝': 1.0, '泡椒凤爪': 1.0, '白胡椒胡萝卜羊肉汤': 1.0, '肉丁茄子': 1.0, '长城窖酿解百纳红酒干红葡萄酒': 1.0, '清炒菊花菜': 1.0, '不加一滴油的酸奶蛋糕': 1.0, '五色糯米饭(七色)': 1.0}, '1096': {'麻辣小龙虾': 1.0, '百里香奶油烤红酒牛肉': 1.0, '清蒸海鱼': 1.0, '西瓜胡萝卜沙拉': 1.0, '纸杯蛋糕': 1.0, '53度茅台': 1.0, '党参黄芪炖牛尾': 1.0, '焖猪手': 1.0, '干锅田鸡': 1.0, '五彩藕苗': 1.0, '黑米恋上葡萄': 1.0, '凉拌菠菜': 1.0, '山药养生粥': 1.0, '豌豆薯仔猪骨汤': 1.0, '芹菜炒腰花': 1.0, '自制猪肉脯': 1.0, '爆炒猪肝': 1.0, '剁椒鱼头': 1.0, '辣炒鱿鱼': 1.0, '蒙古烤羊腿': 1.0, '姜葱炒花蟹': 1.0, '香烤牛排': 1.0, '蒜香辣花甲': 1.0, '美妙绝伦之白莲花': 1.0, '红酒炖羊肉': 1.0, '香辣腐乳炒虾': 1.0, '谷稻小庄 ': 1.0, '孜然羊排': 1.0, '南瓜芝士包': 1.0, '鱼香肉丝拌面': 1.0, '照烧鸡腿': 1.0, '拌土豆丝': 1.0, '番茄炖秋葵': 1.0, '爆炒双丝': 1.0, '番茄炖牛腩': 1.0, '香酥两吃大虾': 1.0, '小炒羊腰': 1.0, '金玉良缘': 1.0, '酸辣藕丁': 1.0, '张裕葡萄酒张裕赤霞珠干红': 1.0, ' 北冰洋汽水 ': 1.0, '哈尔滨啤酒罐装': 1.0, '蒜香包': 1.0, '牛奶卷': 1.0, '老式面包': 1.0, '南瓜枸杞小饼干': 1.0, '打卤面': 1.0, '小米南瓜粥': 1.0}, ...}

(2)准确率计算

根据正确推荐的菜品数和所有推荐的菜品数的比值，来计算菜品推荐准确率，运用下列公式计算出的准确率为如下表所示等序列，则说明模型良好



表7 准确率表

|  |  |  |  |
| --- | --- | --- | --- |
| 训练集 | | 测试集 | |
| ID:1272 | ID:1401 | ID1146 | ID1522 |
| 0.9090909090909091 | 0.9473684210526316 | 0.913333333333 | 0.875 |