# 1 引言

## 1.1编写目的

近年来，旅游成为了越来越多人娱乐休闲的的首要选择，传统的基于旅行社的旅游方式虽然会较为减少旅者的时间与精力，但往往信息过于单一，不够“个性化”，无法满足每一个旅者的需求。因此，自由行也变得越来越受欢迎，但随着旅游信息的爆发式増长，如何方便快捷的从这些信息中获取有用信息变得越来越困难。人们往往不得不花费大量时间与精力来规划旅行景点。

随着机器学习与大数据的发展，个性化的推荐系统也受到了越来越多人的重视，本文探索和研究空间聚类算法，协同过滤算法在推荐领域的应用，设计一个基于协同过滤算法的个性化旅游景点推荐系统。在本系统中，首先通过对flicker数据集的分析，通过空间聚类算法实现对景点信息的挖掘，分析和研究用户与景点之间的数据联系，通过协同过滤算法实现用户个性化的景点推荐。最后，完成个性化旅游景点推荐系统的设计。

## 1.2项目背景

据国家旅游局发布数据显示，2015年，我国国内旅游总人次突破40亿，其中出境旅游人次突破1.2亿，旅游业收入超过过４万亿元人民币。统计表明，2015年在国内旅游的40亿人次中，自由行人群达到32亿人次。根据国家旅游局发布的《2016年上半年旅游统计数据报告》显示，在2016年上半年，我国旅游市场规模稳步扩大，继续领跑宏观经济，其中，国内旅游22.36％亿人次，比上年同期增长10.47％；入出境旅游1.27亿人次，增长4.1％；上半年实现旅游总收入2.25万亿元，增长12.4％，这表明我国的旅游业仍然处于一个稳步发展的阶段，具备可研究的价值。

旅游业高速发展的信息化时代，提供了丰富的旅游信息，但大量数据信息产生的同时也附带了“信息过载”的问题，难以满足人们不断增长的个性化需求。个性化推荐算法对解决“信息过载”的问题具有很好的作用，然而通过调查国内现有的旅游推荐系统，发现大部分旅游信息系统仅根据用户输入的信息进行内容推荐，甚至只具备为用户提供基本的旅游信息查询等功能，缺少鲜明的个性化旅游推荐服务。

国内的大部分旅游信息系统，如去哪儿、携程网和妈蜂窝等网站在为用户提供旅游信息的查询，机票酒店的预定等功能上己经很完备，但是普遍缺乏针对旅游用户的个性化推荐服务，目前提供的功能很难保证用户的个性化需求。

随着GPS和移动互联网技术的发展，产生了越来越多的基于位置的服务，这些服务产生了大量的包含地理位置信息的媒体数据，如人们在Flickr、Picaso上共享的包含地理标签的照片。给照片附上地理标签，既为管理数据提供了方便，也给用户带来了更好的使用体验，从而为这些服务注入了新的活力。因此，各大网站上带地理标签的照片越来越多，数据量都以亿计。这些照片所携带的地理位置信息，都是用户拍照时所处的位置，且当用户将这些照片在互联网上共享时，都经过用户自身的筛选。因此，我们可以认为这些照片中的地理位里均隐含着分享者生活、旅行中的兴趣点。

## 1.3数据集的处理规则

对于初始数据集的挖掘，本文采取Flickr数据集作为基本数据集，Flickr数据集中每一项都包含用户信息以及geo照片信息。

照片和景点的关系是：照片描述的是一个坐标点（拍摄照片时的经度和纬度），而景点是一个面，所以在某个范围内的照片可以认为属于一个景点，可以通过空间聚类算法等方式对Flicker数据集聚类，得出旅游景点项目。

若一个用户在Flicker中上传一张照片，可以认为用户会喜欢此照片所在的景点，且对此景点的喜欢程度与其上传此景点的照片量是正相关的。

一个景点受欢迎的程度与此景点照片数量是正相关的。

一个用户的旅游经验度与其分享的照片数量是正相关的。

# 2系统分析

## 2.1可行性分析

#### 2.1.1经济可行性

本系统作为一个工程实践设计，无需开发经费，目前对于我们来说在经济上是可以接受的，所以经济上是可行的。

#### 2.1.2技术可行性

主要目的是为用户方便快捷个性化景点推荐，当前的热门景点新闻，产品的展示，提高景点介绍编人员的服务质量和服务效率，当用户浏览资讯时，终端向服务器发出请求，服务器旅游信息库的实时数据向终端发送数据，显示在终端的屏幕上。当用户参团时，终端向服务器提交用户信息后，将参团存入信息库，并及时显示该参团信息，系统管理员发布旅游信息时，将信息存入信息数据库并在页面显示该操作添加的旅游信息，当管理员修改旅游信息时，旅游信息数据库数据也做相应改变，页面及时显示更改。辑

## 2.2系统性能需求分析

#### 2.2.1正确性需求

管理员应能够进行有关的旅游信息准确地添加到数据库中。

系统用户登录后，系 统应能正确地读取用户个人信息以及用户的权限信息。

系统的操作结果与预期的结果应该是一致。

#### 2.2.2安全性需求

系统登录需要输入用户名、密码，并且需要防止SQL注入问题，用户登录后根据用户角色的不同可以访问的权限应该不同。为了防止用户未授权访问了不该访问的功能，系统需要在客户端和控制端设置权限拦截。

#### 2.2.3界面需求

系统对界面的要求窗口布局清晰，颜色搭配合理，主题以绿色为主，色调柔和，各窗体主题风格一致同时采用windows 的通用图形界面，对用户友好，必须对鼠标和键盘提供支持，界面的设计应遵循如下规则：

（1）界面具有一致性，接你按规范应遵循MS Windows软件界面概念规范。

（2）提供简单的错误处理。

（3）提供信息反馈，用多种信息提示用户当前的系统运行状态。

（4）操作可逆，其动作可以是单个操作，或者是一个相对对立的操作序列。

（5）显示启动画面，画面要简洁明了，不能太过花哨。

#### 2.2.3稳定性需求

系统部署后，在硬件条件和支持软件条件没有变化的情况下，能够一直保持运行状态，直到系统被升级或代替。

#### 2.2.4扩展性需求

系统应该支持功能扩展与支持环境的扩展。功能扩展就是在现有的功能模块的基础上可以添加信息的功能模块。

#### 2.2.4 故障处理能力需求

系统可能遇到的软件故障是数据库与应用程序服务器。为了满足信息处理的需求，可以采取数据恢复数据来解决。

## 2.3系统推荐流程分析

本系统为用户提供个性化的旅游项目推荐服务，帮助用户确立自己感兴趣的旅游项目，减少用户自行制计划的时间和精力。用户通过对旅游项目评分，是用户对系统的反馈，可以帮助系统进一步提高推荐结果的精度。



# 3 系统模块划分及模块用例图

系统主要模块划分如下。



## 3.1 用户注册

用户注册模块负责用户注册，即向系统添加新用户。



## 3.2 用户登录

用户登录模块用户用户登录。



## 3.3 旅游项目推送

旅游项目推送模块用于用户在浏览网站中的旅游项目时，向用户推送浏览的旅游项目的相关信息，可以包括项目的照片，相似项目等。



## 3.4 推荐反馈

旅游项目评分模块用于用户对以游玩的项目进行评分，这是用户对系统的反馈，可以帮助系统进一步提高推荐结果的精度。



## 3.5 用户管理

用户管理模块用于用户管理自己的信息。如登录密码等。



## 3.6 旅游项目管理

旅游项目管理模块用于管理旅游项目，即从flicker数据集中聚类出的旅游景点信息。系统管理员可以重新聚类项目，删除某个项目和修改某个项目的信息。



## 3.7 个性化推荐

个性化推荐用于生成个性化推荐结果。



# 4.开发环境

## 1.算法建模

使用python语言及其第三方库，如panda,numpy等。

## 2.web开发

使用HTML、css、javascript构建web前端页面

使用python构建后台服务器，处理业务逻辑

使用mysql数据库