# 《基于标签的个性化旅游推荐算法》

课程名称： 大数据与云计算概论

专 业： 网络工程 班 级： 17-1

姓 名： 贾西腾

学 号： 201701060711

**山 东 科 技 大 学**

**2020 年 10 月 28 日**

### 一.引言

世界旅游组织的行业预测，未来20年世界旅游行业将持续保持较高的增长趋势。国际旅游人数预计平均增长率为每年3.3％。与此同时，大量的旅游用户也面临着一些旅游规划的问题。面对大量的景点选择，用户需要考虑诸多因素，如爱好、时间、价格、位置、交通等，这将耗费用户相当多的时间和精力去选择满意的景点。近年来个性化推荐系统在各领域得到广泛应用。在线旅游网站通过有效利用推荐系统提高了用户服务质量，进而提高自身的竞争力。在线旅游推荐系统基于用户的历史数据分析用户的兴趣特点，可以从海量的旅游景点中为用户提供个性化的景点推荐服务，从而提高服务质量，提升在线旅游网站的竞争力。

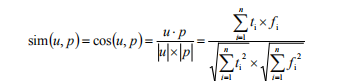
数据大爆炸时代的到来，面对信息过载问题，如何让人们在海量的数据面前能够获取到自己想要的信息，于是设计出了推荐系统。个性化推荐算法作为系统的核心，目前主要有以下5个分类：基于内容的推荐、基于关联规则的推荐、基于知识的推荐、协同过滤推荐和混合推荐。

### 二.个性化推荐算法

除协同过滤算法外，个性化推荐算法目前主要有以下4个分类：基于内容的推荐、基于关联规则的推荐、基于知识的推荐、混合推荐。

2.1基于内容的推荐算法

基于内容的推荐算法是一个简单但又重要的推荐思想。首先，该算法需要项目的各个属性特征，构建项目属性向量。接着分析用户历史行为记录，构建用户兴趣偏好向量，通过计算用户兴趣偏好向量与未评价的各个项目自身的属性向量的相似度大小比较，对目标用户产生项目预测评分或top-N推荐。假设用户兴趣偏好向量为u，未评价某项目特征向量为p，u=(t1,t2,…,ti,…,tn),i=1,2,…,n，p=(f1,f2,…,fi,…,fn),i=1,2,…,n，可以利用相似度计算公式如余弦相似度得出用户兴趣偏好向量与项目属性向量的相似度，相应的计算公式如下图所示:



计算的相似度值越大，说明用户对该未评分项目的特征偏好程序越高。推荐的结果与其他用户无关，只与项目本身的属性特征信息和该目标用户对若干项目产生历史行为数据

有关。该算法优点是推荐结果直观，缺点是存在数据的稀疏性、复杂的项目难以提取出项目属性特征向量和项目属性之间的重要程度权重值分配等问题。因此，研究了特

征权重的选取方法对推荐效果的影响；文献[5]将项目语义应用于个性化推荐;通过分析项目属性关系将项目粒度化，提出了一种基于内容的加权粒度序列推荐算法；提出基于内容和兴趣漂移模型应用于电影推荐算法中。

2.2基于关联的推荐算法

关联规则问题是Agrawal等人于1993年提出来的。为寻找数据库数据项之间的关系，提出频繁项集概念，再根据频繁项集定义关联规则。关联规则推荐过程：第一，用户

或专家指定最小支持度阈值和最小置信度阈值；第二，从数据库中找到不低于最小支持度的频繁项集；第三，利用第二步中得到的高频项集来产生满足最小置信度的强规则；第四，

根据强规则实施个性化推荐。当数据集比较大时，比较影响算法效率。利用关联规则挖掘的特性，挖掘用户属性与项目之间的关联，提出了基于关联规则挖掘的分类随机

游走算法；将关联规则挖掘应用与商品推荐中，通过用户历史记录分析挖掘，得到不错的推荐效果。

2.3基于知识的推荐算法

基于知识推荐，需要建立知识网络图谱，分析用户已有的知识和需求的知识之间的关联，为用户推荐新知识。考虑将知识系统以知识网络进行表达，引入最近邻优先的候选知识选择策略，提出一种基于建构主义学习理论的个性化知识推荐方法-建构推荐模型，通过知识网络的知识关联结构挖掘用户知识需求，并推荐出最具建构学习价值的新知识。针对不同的应用场景，如何构建一个最为合理的知识网络是一个关键性的问题。

2.3混合推荐算法

混合推荐是将若干种个性化推荐算法通过某种方式进行结合。针对不同的应用场景，考虑各单一的个性化推荐算法的优缺点，增强推荐效果。常用的有以下5种方式进行结合：加权、变换、混合、特征组合和层叠等。利用数据挖掘知识和协同过滤算法，提出一种结合用户聚类和评分偏好的混合推荐算法；提出一种基于用户特征聚类和SlopeOne填充的协同过滤混合推荐算法，试图解决数据稀疏性、可扩展性问题；提出搭建Hadoop分布式平台，并结合Mahout工具进行组合推荐算法的设计与实现，应用于电影推荐；提出协同过滤推荐算法在融合大数据技术、社会网络分析技术及关键用户分析技术方面的研究。可见，结合机器学习、数据挖掘等知识的混合推荐算法，是

今后研究与应用的主要方向。

### 三.模型

3.1景点和用户数据的挖掘与处理

3.1.1景点分类因素和标签

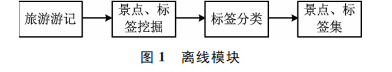
旅游计划中，用户对旅游景点的选取，需要考虑很多因素。如景点的所属区域、当季是否为最佳旅游时间等。为了方便起见，我们称这两个因素分别为区域和时间。在目前的旅游网站上，存在着海量的旅游游记。通过对游记的分析，我们发现用户在计划旅游选取景点时，还会考虑旅游的主题和景点的类型因素。关于主题，在计划旅游时，不同的用户会偏好不同的主题，如一些用户会偏好符合“徒步”主题的景点，而另一些用户可能会偏好符合“亲子”主题的景点。关于类型，是景点本身的特性。有的景点的类型是“古镇”，如凤凰古镇、束河古镇；而香港和上海则可以归到“城市”类型中。关于主题中的“亲子”、“情侣”这样的词汇，我们称之为景点主题标签。同样地，“城市”、“古镇”则称为景点类型标签。通过以上分析，我们将景点划分为区域、时间、主题、类型４种因素。此外，从海量的旅游游记中挖掘信息，我们提取出关于４种因素的标签。

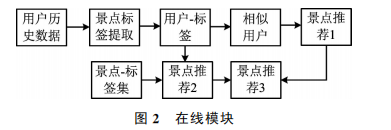
3.1.2用户历史数据和标签

得到用户历史数据之后，将会得到非常稀疏的用户－景点矩阵，这样稀疏的数据不适合用在协同过滤算法中。为了改善这种状况，我们提出设置用户标签。在旅游推荐中，用户标签和景点标签相关。每个用户去过多个景点，这一系列景点有自己的特征标签，同样的标签可能会出现多次。我们用出现频率最高的多个景点特征标签来描述该用户的兴趣标签。

3.2模型

我们的个性化旅游推荐系统模型分为两部分。一部分是离线模块，用于景点－标签集的构建。另一部分是在线模块，用于为用户做旅游景点的推荐。如图１、图２所示。





离线模块的目标是构建景点－标签数据库。首先从海量的游记中挖掘出景点和描述景点的特征标签，将景点和对应的标签联系起来；然后对标签进行分类；最后再将景点和对应的４种因素下的标签联系起来，构建景点－标签数据库。

在线模块为用户作出个性化的景点推荐。首先根据用户历史旅游景点数据，结合离线模块构建的景点－标签数据，将用户旅游景点的相应标签提取出来，用这些景点标签描述用户的兴趣特征，构建用户－标签矩阵；然后我们用两种方法推荐景点：一种方法是基于标签的协同过滤算法，这需要计算相似用户，得到景点推荐１；另一种方法是结合景点－标签数据的基于标签内容的推荐算法，得到景点推荐２；最后，我们用这两种算法加权后的混合推荐算法，进行更进一步的推荐，得到景点推荐３。

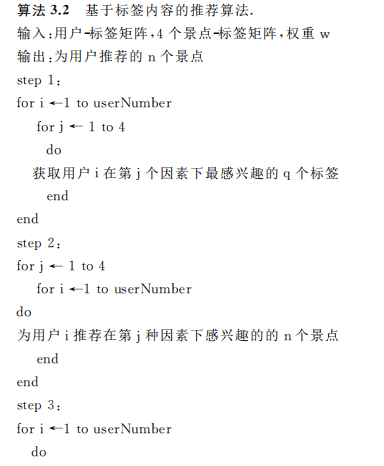
### 四.算法

4.1基于标签内容的推荐算法

该算法的思路是，对于目标用户而言，当景点在内容特征上与该用户以往感兴趣的景点相似时，用户将来很可能对这些景点同样感兴趣。不同于协同过滤算法，基于内容的推荐算法不需要通过相似用户来推荐，即向目标用户推荐时不需要其他用户的数据，没有冷启动问题和数据稀疏问题，因此我们尝试用该方法向目标用户推荐。

我们通过目标用户去过的景点的特征来学习目标用户的兴趣特征；通过用户的兴趣特征与候选景点的特征数据比较，为用户推荐相关性最大即用户最可能感兴趣的景点。

如下，描述了基于标签内容的算法。



4.2加权的混合推荐算法

基于标签的协同过滤算法和基于标签内容的推荐算法各有优劣。首先，向目标用户推荐景点时，基于标签的协同过滤算法考虑了相似用户，推荐依赖于其他用户；而基于标签内容的推荐算法具有用户独立性，以目标用户个人的历史数据构建用户特征信息。其次，基于标签的协同过滤算法无法将一个新的景点推荐给用户，因为没有相似用户去过该景点；而基于标签内容的推荐算法却可以较好地为用户推荐新的景点。此外，基于标签的协同过滤算法由于考虑了相似用户的兴趣，所以在一定程度上可以向目标用户推荐新颖的景点；而基于标签内容的推荐算法仅基于目标用户个人的信息作出推荐，具有过度个性化的缺点，在景点推荐时较难推荐新颖的景点。基于以上考虑，为达到更好的推荐效果，我们采用线性加权的方式

结合这两种算法做出景点推荐，公式表示如下：



Ｕ表示用户集；recom3U表示在加权混合推荐算法下对用户的景点推荐结果；recom1Ｕ表示在基于标签的协同过滤算法下对用户的景点推荐结果；recom2Ｕ表示在基于标签内容的推荐算法下对用户的景点推荐结果；ｗ１，ｗ２分别表示两种推荐结果线性加权的权重。

### 5.结论

本文提取景点标签并用景点标签描述用户兴趣特征，利用基于标签的协同过滤算法、基于标签内容的推荐算法、基于标签的混合推荐算法向用户作个性化景点推荐。景点标签和４种因素相关，分别是区域、时间、主题、类型，丰富了现有的通过主题模型提取的标签类别。基于４种标签的算法计算用户相似度进而作出推荐，改善了推荐效果。同时本文将基于标签的协同过滤算法与基于标签内容的推荐算法相结合，这种混合算法结合了两种算法的优势，进一步提高了推荐结果的准确度。

### 六.参考文献

[1]刘树栋，孟祥武．基于位置的社会化网络推荐系统［Ｊ］．

计算机学报，２０１５，３８（２）：３２２－３３６．

[2]聂恩伦，陈黎，王亚强，等．基于Ｋ近邻的新话题热度预测算法［Ｊ］．计算机科学，２０１２，３９（６Ａ）：２５７－２６０．

[3]项亮．推荐系统实践［Ｍ］．北京：人民邮电出版社，２０１２．

[4]冷亚军，陆青，梁昌勇．协同过滤推荐技术综述［Ｊ］．模式识别与人工智能，２０１４，２７（８）：７２０－７３４．