

# 基于数据挖掘的南阳市区智能旅游路线规划算法研究

詹可迪 段肖寒 廉幸艺 王 萌  
南阳理工学院 河南 南阳 473003

**【摘要】**随着收入和生活水平的提高,自由行成为人们出行的首选。本文针对自由旅行游客设计个性化的旅游路线,研究了一种基于数据挖掘的最优旅游路线规划算法。以南阳市各景区为例,利用数据挖掘技术充分挖掘旅游路线规划中的有效信息,利用分支定界模型探索旅游过程中的最佳路线,满足游客个性化需求。

**【关键词】**数据挖掘; 分枝定界法; 关联规则; 旅游路线规划

## 引言:

游客在游览某个城市之前,往往会有自己的主观意识,查询旅游信息,确定旅游路线。当下,旅游行业相关网站虽然可以为游客提供景点等相关信息,但规划路线可能会遇到很多现实问题,使得规划路线不尽如人意。

本文利用数据挖掘中的“关联法”和“分支定界法”的数学模型研究旅游线路规划,以解决游客在有限的时间内游览尽可能多的景点的问题。在为游客出行带来便利的同时,也促进了南阳的经济发展。

## 一、南阳市区景点的确立

首先,利用高德地图、百度导航等网站确定景区大概位置、实地考察、确定各景区的具体位置和景区特色。

最后确定研究的南阳市的热门景点有:南阳汉画馆、武侯祠、医圣祠、张衡墓、南阳府衙、白河国家湿地公园、独山、南阳世界月季大观园、南阳月季博览园、王府山。

根据景点的不同风格特点,将景点划分为人文景观和自然景观两种类型,具体见表1。

表1 南阳市区景点类型

类型	景点名称
人文景观	南阳汉画馆、武侯祠、医圣祠、张衡墓、南阳府衙
自然景观	白河国家湿地公园、独山、南阳世界月季大观园、南阳月季博览园、王府山

## 二、数据挖掘

数据挖掘是从大量数据中自动搜索具有特殊关系的信息的过程。利用数据挖掘技术可以提出更合理、更受欢迎的旅游路线。本文使用的数据挖掘算法是一种关联规则方法。

关联规则反映了一个景点与其他景点之间的相互依存关系。如果两个或多个景点之间存在一定的关系,那么一个景点就可以被其他景点预测。该算法可以利用大数据挖掘相关联景点,并帮助游客规划景点旅行路线参考意见。

### 1. 项集的最小支持度与频繁项集

为了满足一定的要求,用户需要制定规则必须满足的支持度和置信度阈值。这个阈值称为项目集的最小支持度,记为  $\text{sup}_{\min}$ 。频繁项集是指支持度大于或等于最小支持度的项集,即频繁景点的集合;小于最小支持度的项集为非频繁项集,即非频繁景点的集合。

### 2. Apriori 算法生成频繁项集, 频繁项集挖掘关联规则

如果存在一个频繁项集 { 汉画馆, 武侯祠 }, 可能有一条关联规则就为“汉画馆  $\Rightarrow$  武侯祠”, 即若某位游客选择汉画馆, 则极大可能该游客下一个景点会选择武侯祠。但反过来则不一定。这里频繁项集使用支持度量化, 关联规则使用置信度量化。

对于一个频繁项目集 M, 可以生成规则  $R: M \Rightarrow N$ , 置信度定义为:

$$\text{confidence}(M \Rightarrow N) = \frac{\text{support}(M \cup N)}{\text{support}(M)}$$

其中  $\text{support}(M)$  表示项集的支持度。

### 3. 实际意义

根据游客在各个软件和网站上提供的旅游景点搜索记录, 确定游客想去的旅游景点。利用关联规则对系统中的数据进行挖掘, 可以更好地收集游客喜欢的景点信息, 生成对旅游路线规划具有指导意义的关联景点。获得关联景点后, 根据获得的关联景点进行最优路线规划。

## 三、南阳市最优旅游路线规划

### 1. 问题重述

如何轻松、不重复地观光南阳人最喜欢的旅游景点问题。

### 2. 符号说明

(1)  $V_i, V_j$ : 无向图的顶点即南阳市区的各旅游景点位置;

(2)  $D$ : 由各景点间的距离值所构成的矩阵;

(3)  $S$ : 起点与其他景点的距离为零的元素;

(4)  $D_i$ : 各景点间的距离构成的矩阵中每一行减去该行最小元素及每一列减去该列最小元素后构成的矩阵;

(5)  $(V_i, V_j)$ : 无权图的边, 即权, 表示两景点间距离;

(6)  $d(V_i, V_j)$ : 为任意两顶点与顶点在图中最短路径长度  $d(V_i, V_j) = d_{ij}$ 。

### 3. 模型假设

- (1) 到景区的距离是可测的。
- (2) 游客在景区花费的时间和费用相等。
- (3) 南阳汉画馆是游客的旅游景点，游客最终会回到汉画馆。
- (4) 游客经过道路畅通无阻，各景区距离与时间和

油耗成正比。  
4. 模型建立与求解

(1) 根据旅游中游客的不同出行方式(自驾游等)，得到南阳部分旅游景点的路程情况见表2(每两个旅游景点之间有一个可行的距离，单位：公里)。

表2 每两景点间的路程

	南阳 东站	南阳汉 画馆	武侯祠	张衡墓	医圣祠	府衙	白河国家 湿地公园	独山	南阳世界月 季大观园	南阳月季 博览园
南阳东站	0	20	22	33	16	18	14	21	7.4	20
南阳汉画馆	20	0	1.3	26	6	4.8	9.2	13	18	17
武侯祠	22	1.3	0	27	6.3	4.5	9.7	13	17	17
张衡墓	33	26	27	0	21	26	21	15	25	13
医圣祠	16	6	6.3	21	0	3.6	3.8	7.8	11	11
府衙	18	4.8	4.5	26	3.6	0	6.1	9.8	13	13
白河国家湿 地公园	14	9.2	9.7	21	3.8	6.1	0	8.2	8.1	8.2
独山	21	13	13	7.8	7.8	9.8	8.2	0	14	6.5
南阳世界月 季大观园	7.4	18	17	11	11	13	8.1	8.1	0	15
南阳月季博 览园	20	17	17	11	11	13	8.2	8.2	15	0

(2) “分枝定界法”模型

将每个景点到其他景点的距离看做  $n$  阶矩阵  $D$ ，即矩阵  $D$  为对称方阵。首先，令  $D$  中的每行减去各行最小非零元素，得到矩阵  $D_1$ 。再令矩阵  $D_1$  的每列减去各列的最小非零元素，得到新矩阵  $D_2$ ，使得新矩阵  $D_2$  的每行每列中都至少存在一个零元素。然后，将矩阵  $D_2$  中起点与其他景点的距离为零的元素所在行与列去除得到矩阵  $D_3$ 。此时起点与元素  $S$  组成一条路。再次，将矩阵  $D$  变化到矩阵  $D_3$  的这一步骤对进行重复操作，得出新的景点添到最近一次循环操作确定路线的后面，使其成为一条新路，直到最后矩阵形式为  $\begin{bmatrix} 0 & \infty \\ \infty & 0 \end{bmatrix}$ 。最后要求求得路线中所有的景点在一条路线只能出现一次，且包含所有选中的景点。满足所需的两个要求之后即可停止，否则一直重复上面的步骤，直到满足条件为止。

(3) “分枝定界法”寻找最佳旅游路线算法求解

步骤1: 使用 Floyd(弗劳德)算法得到任意两个景点的距离，构造无向图。

步骤2: 在所有无向图中随机搜索初始哈密顿圈或几个具有指定起点的哈密顿圈。

步骤3: 检索是否有新节点可用，从而得到近似的最佳哈密顿圈。

步骤4: 比较得到的哈密顿圈的权重，最小的权重就是最佳哈密顿圈的近似解。

(4) 模型评价

规划最佳旅行路线的方法有很多。在“最近邻插值法”和“分枝定界法”两种算法中，“最近邻插值法”和“分枝定界法”在给定权重下都转化为无向，找出图中总权重最小的哈密顿环。但是，由于模型和算法不同，“最近邻插值”的方法更直观易懂，但效果不佳，会造成严重的图像失真；“分枝定界法”在景区地点较多的情况下，可以使用计算机编程来解决问题。

因此，在实践中使用“分枝定界法”来寻找近似的最佳旅游路线更有优势。

四、结束语

针对自由行游览南阳景区的游客旅游路线选择问题，本文采用数据挖掘和分枝定界相结合的方法，基于游客关注点，对南阳景区内部相关性进行建模，确定最优路径规划算法，解决游客在有限的时间内游览尽可能多的景点的问题。

参考文献:

[1] 周溪召主编. 运筹学及应用——北京：化学工业出版社，2009. 1.  
 [2] 王文平等编著. 运筹学——北京：科学出版社，2007.  
 [3] 姜启源，谢金星，叶俊. 数学模型(第三版). [M] 北京高等教育出版社，2003. 8.  
 [4] 杜治涵. 数据挖掘研究[J]. 信息与电脑(理论版)，2021, 33(01):169-171.  
 [5] 颜雪松，蔡之华，蒋良孝等. 关联规则挖掘综述[J]. 计算机应用研究，2002(11):3-6.  
 [6] 包剑. 关联规则挖掘研究[J]. 计算机系统应用，2005(11):56-58.  
 [7] 陆丽娜，陈亚萍，魏恒义，等. 挖掘关联规则中 Apriori 算法的研究[J]. 小型微型计算机系统，2000, 021(009):940-943.  
 [8] 栗晓聪，滕少华. 频繁项集挖掘的 Apriori 改进算法研究[J]. 江西师范大学学报(自然版)，2011(05):498-502.  
 [9] 李少年，孟志青，田媛. 一个基于频繁项集的时态数据挖掘算法[J]. 湘潭大学自科学报，2003, 025(003):21-25.

作者简介:姓名:詹可迪(2000.1—)性别:女,民族:汉族,籍贯:河南省洛阳市,高中学历,南阳理工学院数理学院学生,路线规划。