北京工业大学

2017 - 2018 学年 第1学期

信息学部 计算机学院

|  |  |  |  |
| --- | --- | --- | --- |
| 课程名称： | 数据结构课程设计 | | |
| 报告性质： | 实验报告 | | |
| 学号： | 15143103 | 姓名： | 刘方瑞 |
| 任课教师： | 王众 | 课程性质： | 学科基础必修课 |
| 学分： | 2 | 学时： | 60 |
| 班级： | 150743 | 成绩： |  |
| 教师评语： |  | | |

|  |  |  |
| --- | --- | --- |
| 需求分析 | 根据题中需求，提供功能划分说明 |  |
| 设计 | 逻辑结构、存储结构设计、算法描述 |  |
| 使用说明 | 界面是否友好 |  |
| 总结 | 是否感悟有收获 |  |
| 摘要 | 考察文字抽象能力 |  |
| 格式 | 是否有目录、页号 |  |

2017年 月 日

目录

[一、 需求分析 - 3 -](#_Toc498030509)

[1.1 概述 - 3 -](#_Toc498030510)

[1.2 应用场景 - 3 -](#_Toc498030511)

[1.3 算法模型 - 3 -](#_Toc498030512)

[1.4 架构思路 - 4 -](#_Toc498030513)

[二、 数据结构设计 - 5 -](#_Toc498030514)

[2.1 数据结构模型 - 5 -](#_Toc498030515)

[2.1.1 逻辑模型 - 5 -](#_Toc498030516)

[2.1.2 存储模型 - 6 -](#_Toc498030517)

[2.2 算法思路 - 6 -](#_Toc498030518)

[2.2.1 概述 - 6 -](#_Toc498030519)

[2.2.2 遍历路径树 - 7 -](#_Toc498030520)

[2.2.3 最短优先 - 8 -](#_Toc498030521)

[2.2.4 重复排除 - 9 -](#_Toc498030522)

[2.3 算法特点 - 9 -](#_Toc498030523)

[三、 详细设计 - 9 -](#_Toc498030524)

[3.1 架构设计 - 9 -](#_Toc498030525)

[3.1.1 Web前端 - 9 -](#_Toc498030526)

[3.1.2 Flask与后台 - 9 -](#_Toc498030527)

[3.1.3 算法模块 - 12 -](#_Toc498030528)

[3.2 算法实现 - 12 -](#_Toc498030529)

[3.2.1 Floyd矩阵与距离更新 - 12 -](#_Toc498030530)

[3.2.2 路径树的生成与选路策略 - 12 -](#_Toc498030531)

[3.2.3 路径输出 - 12 -](#_Toc498030532)

[3.3 算法分析 - 12 -](#_Toc498030533)

[3.3.1 时间复杂度 - 12 -](#_Toc498030534)

[3.3.2 空间复杂度 - 12 -](#_Toc498030535)

[3.3.3 可行性分析 - 12 -](#_Toc498030536)

[四、 测试 - 12 -](#_Toc498030537)

[4.1 模块测试 - 12 -](#_Toc498030538)

[4.1.1 前端测试 - 12 -](#_Toc498030539)

[4.1.2 后端测试 - 12 -](#_Toc498030540)

[4.1.3 算法模块测试 - 12 -](#_Toc498030541)

[4.2 整体测试 - 13 -](#_Toc498030542)

[4.3 测试总结 - 13 -](#_Toc498030543)

[五、 总结与提高 - 13 -](#_Toc498030544)

[5.1 总结与提高 - 13 -](#_Toc498030545)

[5.2 写给自己的话 - 13 -](#_Toc498030546)

[六、 附录 - 14 -](#_Toc498030547)

[6.1 算法API手册 - 14 -](#_Toc498030548)

[6.2 服务器API手册 - 14 -](#_Toc498030549)

[6.3 web交互手册 - 14 -](#_Toc498030550)

1. 需求分析
   1. 概述

本次数据结构课程设计的目标是完成一个校园导游系统。假定学校有十个场所，而任意两个场所之间的距离各不相同。系统可以在用户选定的场所中规划出最优的游览路线，并且对于其中的一些路径点能够考虑用户对时间的要求，最终回到起点。而这个规划的路线应该尽可能得短，同时还要满足用户对于时间上的要求。系统在给出路径规划的同时，应该给出路径的时间消耗信息和路径经过的具体点。

对于这个问题我们可以使用数据结构中与图相关的算法来求解得到。由于涉及到不同类型的策略，所以我们不能完全“照搬”图中求解单源最短路径的方法。在这次我们解决实际问题的过程当中混合了不同的求解方法，而这些子算法中都是运用数据结构的具体知识实现的。

* 1. 应用场景

首先我们的目标用户是参观大学的普通人群，因此我们在准备验证用的终端程序的同时，还要向用户提供界面友好的接口，来提供可视化的结果，给用户更直观的反馈。

由于现在移动设备的增加，如果继续使用客户端/服务器架构来提供服务，会增加用户的使用成本。对于地图线路规划的这种轻量级服务，下载客户端会成为用户使用服务的累赘。为此我们选择了浏览器/服务器的架构来提供服务。

浏览器/服务器架构可以支持多种操作系统，可以在几近所有的设备上运行。由于互联网服务在快速发展，通讯成本在迅速下降。这样使得将服务部署到互联网与部署到本地存储几乎无异。同时，浏览器/服务器结构可以很好地保护知识产权。所有的算法核心代码全部都在远程的服务器中，用户无需存储任何核心算法代码。因此也就没有任何被侵权的可能性了。

* 1. 算法模型

我们在课程设计中使用的核心算法是Floyd算法。

对比Dijkstra算法，Floyd的时间复杂度也许并没有很大优势。然而由于我们需要随机从图中取出最短路径，而Floyd恰好在一次更新邻接关系和距离以后能够得到所有点之间两两的关系。因此比每次都要单独求解的Dijkstra算法更加方便。在地图不变化的前提条件下，甚至可以直接存储更新以后的Floyd邻接矩阵和距离矩阵，大大减少了计算量，将O(n3)的时间复杂度直接下降至O(n)[[1]](#footnote-1)。

我们在算法中使用矩阵存储图。因为Floyd算法与矩阵更加友好，而且矩阵更加便于存取。矩阵中分两种顶点，一种是可见顶点，这种顶点是可以被用户选中作为路径点的；而另一种是不可见顶点，这种顶点比较特殊：不可见顶点对于用户是不可见的，而对于整个规划系统是可见的。这些顶点一般都是多个顶点的中转点，类似十字路口等的顶点。我们在将图抽象成邻接矩阵的时候，会将邻接矩阵一并计算在内，这样可以方便在添加顶点后的矩阵设计，也可以更规整地设计整个抽象地图与实际地图的映射关系。

由于在需求中规定需要考虑部分顶点的时间限制要求，那么我们就需要考虑在规定的时间中我们是否能够完成我们规划的路径。考虑到给用户提供最大的自由度，我们选择遍历所有可能的路径点来选择最优的路径解。我们如此抽象问题：假设用户给定的具有时间限制的顶点都是按照时间顺序排列好的。而这些顶点中带有的时间限制仅仅是单纯的从上一个出发点到这个顶点过程中能够消耗的最大时间。我们就针对这个顶点构建一棵树。如图 1所示。



图 1 遍历路径树

我们将这种图的结构称为遍历路径树[[2]](#footnote-2)。树的起点是上一个出发顶点，所有叶子结点是所有可能的到达顶点，也就是所有的叶子顶点的下一个路径点都是带有时间要求的这个节点本身。不考虑剪枝情况且将目的顶点考虑在内的遍历路径树，如图 2所示。



图 2 完全遍历路径树

其中顶点0是上一个出发点，顶点5是具有时间要求的顶点。中间的顶点都是可能的路径点。当然我们在构建这样一棵伪树的时候，是必须根据路径长度进行剪枝的。

选择遍历的方法是因为我们可以从中获益：用户可以选择自己不同时间关键点[[3]](#footnote-3)之间需要途径的节点个数。从遍历路径树的角度而言，就是可以通过限制树高就可以完成这个功能。同时我们在对完全遍历路径树[[4]](#footnote-4)剪枝的过程中就完成了对时间限制因素的考量。同时我们可以对路径树中的每层节点进行升序排序，并且按照最左优先的原则选择路径，就可以得到我们期望的最短（且最深）的路径。

这一方法最大的特点就是灵活，并且可以根据用户的不同需求求解不同策略下的最优路径解。

* 1. 架构思路

本次课程设计采用了浏览器/服务器的架构设计。因此我们需要将整个应用划分为前端服务页面、后端服务应用和算法服务三个部分。

前端服务页面主要由动态网页框架与网页脚本两部分组成。动态网页框架主要定义服务页面的布局与外观，用于给用户一个良好的可视化界面。网页框架主要将内容分为三层，分别用于显示内容，更新内容和监听用户行为来作出反应。网页脚本主要负责对已经收到的用户行为做出相应，检查和组装数据，并把合法数据发送给服务器。

后端服务应用由软路由和中间件组成。我们通过软路由将网页分为多个独立可服务的页面。用过对于每个可路由的地址我们都编写相应的服务处理程序来响应所有合法的请求。中间件的主要功能是将请求的信息转化为能够被算法服务解析的数据结构，例如数组等。

算法服务时整个应用的核心。我们主要将算法应用根据功能划分为了三个部分。

首先是单源最短路径求解模块。这个模块主要负责使用Floyd算法更新并存储链接状态矩阵组[[5]](#footnote-5)和计算单源最短路径。这个模块是我们整个算法服务的核心。我们之后的模块都是以此为基础的。

在这之上的是多路求解模块，该模块负责接收两组数据，分别是包含普通路径点的无序集合[[6]](#footnote-6)N和包含时间关键点有序序列K。我们的算法主要是对于给定的N与K，规划一条最优路径规划。模块最终的输出是一组由二元组构成的路径序列。其中每一个二元组都是对一条边的描述。

第三个模块就是路径树处理模块。该模块负责对遍历路径树进行构建、剪枝、以及选路的操作。我们的路径树是由链式存储的，因此对于每一次多路求解的迭代，都需要动态地构建并对路径树剪枝。该模块输入是路径树的起点、终点以及一个中继顶点集合，而这些中继顶点都是由普通路径点构成的。

以上就是本次课程设计的架构设计思路，具体每个模块的设计细节会在第三章中具体介绍。

1. 数据结构设计
   1. 数据结构模型
      1. 逻辑模型

在此次数据结构课程设计中，我们一共运用了以下数据结构：队列、n叉树和图。

首先队列是我们在程序设计当中最常用的数据结构。由于我们的算法中涉及到迭代的算法，队列是我们判定迭代结束的重要条件。我们利用队列先进先出的特性，对于固定顺序的时间关键点分别求解最短路径。

N叉树用于我们构建遍历路径树的过程中。我们使用一棵N叉树来存储遍历路径树。对于一个长度为k的普通路径点集合，路径树第一层的树枝顶点数为k个，因为遍历路径树会穷举所有的可能性。对于路径树的第二层树枝节点，我们会安排k-1个顶点。以此类推，当我们完成遍历路径树的穷举时，一共会生成k！条路径。当然这是没有剪枝的情况，通过剪枝，我们会得到更加精炼的结果。

由于我们的数据规模较小，因此在面对k!条路径并不是很困难，因此我们就选择了遍历路径树来解决这个问题。因为路径树可以给我们更灵活的选择，并且在这种数据规模下并不会影响算法的效率。但是如果数据量扩大十倍，也许我们就需要更加高效率的算法来实现选路的问题了。

接下来是图结构。由于本次数据结构课程设计的主要操作对象是地图，因此图作为基本的载体，是我们主要操作的对象。对于一张图G，拥有边集E与点集V，记作G<V,E>。

这就是我们使用的数据结构逻辑模型的简单介绍。

* + 1. 存储模型

我们使用数组来存储队列。在使用数组存储队列的时候需要注意一点，就是数组是一个固定长度的结构，如果我们向数组里面插入的元素超出了缓冲区的长度，我们就需要更新数组的容器，也就是创建一个更大的数组，并将数组的内容拷贝过去，更新缓冲区长度的值，并释放掉原来的数组。这样我们就完成了一个变长的数组队列。

我们使用链式存储树。我们使用了链式方法存储树，也就是说树的子节点都是用指针存储的。如果想要访问树的子节点，仅需要访问指针指向的结构体就可以完成。链式存储有几点好处，我们可以更加动态地创建树的结构，并且更加形象地完成剪枝的动作。

我们使用矩阵存储图。由于我们使用了Floyd算法作为我们的单源最短路径求解方法，所以最好的选择就是使用矩阵存储图。矩阵存储图访问速度更快一些，并且会摆脱指针的痛苦。我们知道图有三种存储方法，邻接矩阵是一种，剩下的两种[[7]](#footnote-7)都需要维护指针对应的表结构。这样可以让我们随机读取边的信息，可以有效提高我们算法执行的性能。

* 1. 算法思路
     1. 概述

在课程设计中我们主要是围绕这如何将时间关键点纳入到路径规划算法当中。我们会阐述之前的一些设计思路，并详细解释为什么选择现有的算法。

首先要考虑时间关键点的问题，就要考虑时间关键点为什么出现。经过仔细阅读课程设计说明，并向老师请教核实以后，我认为，我们要规划的路径要在每对时间关键点之间插入任意数量的路径点。然而选择多少点插入，并且选择那些点插入都是需要算法给出的。对于这样的多自由度的问题，我们就需要制定策略来规划具体路径了。由于对于这样的问题求解全局最优解也许会比较繁琐，考虑到运筹学中的部分知识，我们从局部最优解考虑此类问题或许是一个比较好的办法。

于是我们想到了直接向规划好的路径进行插入的方法。然而这种方法很难预期下一步得到的新路径的消耗是否能够满足时间关键点的要求。因此其实也相当于在遍历整个路径的可能性。而且这样的方法很难有条理地理解路径的走向。我们需要一个在生成路径的同时完成插入路径点的算法。因此我们引入了遍历路径树来解决这个问题。

我们将问题抽象成两个层次的子问题：宏观的关键点层次[[8]](#footnote-8)的路径求解问题和低层次的图中顶点的路径求解问题。由于我们已经将图中任意两个顶点之间的链接状态进行了求解，因此我们可以对任意的两点快速得到最短路径，甚至这些路径可能由多个顶点构成。于是我们可以先完成在关键点层次的路径求解，通过求解出宏观路径而得到低层次的也就是具体在图中的路径，如图 3所示。



图 3 左图是关键点层次路径规划，右图是顶点层次路径规划

然而我们在抽象问题的过程中发现了会出现在规划完关键点层次的宏观路径以后，再重新求得实际路径的时候，会出现关键点在得到的实际路径中重复出现中出现。这样就浪费了我们的路程消耗，不过通过后来的技术实现我们解决了这个问题。详细设计请参照2.2.4。

* + 1. 遍历路径树

正如我们在1.4所讨论的，遍历路径树是我们选择时间关键点之间路径点策略的工具。遍历路径树可以灵活地把我们的策略与具体的数据结构联结起来。这赋予了算法其他方法不可比拟的灵活性，在部署的时候可以根据需求灵活地更改策略，也可以让用户自定义策略，完成不同目标的路径规划任务。我们认为遍历路径树的主要操作涉及构建、剪枝、以及排序。

首先我们来抽象地描述一下我们的规划算法：

定义 1：假设由起始时间关键点S，目的时间关键点D，以及由普通路径点组成的集合G={N1, N2, N3,… …, Nk} 。在时间关键点S与时间关键点D之间，有时间要求L。我们的是找出在满足时间要求L所有可能的P⊂G，也就是所有满足时间要求的G的子集。之后从子集中选出符合目前策略的集合Pr 。最终我们使用路径函数H(S, D, Pr)将整个集合转化为路径R。这样得到的路径R就是局部最优解。对于处于时间关键点中所有的有序对<S, D>求解路径R，就得到了该问题的一个全局解。

其中对于路径树的生成方法，已经给出了一个比较清晰的描述：遍历路径树的目标就是穷举在所有可能的P⊂G，也就是G的子集[[9]](#footnote-9)。如果我们能够构建这样一个结构，就相当于把所有的解都进行了遍历，就是定义了我们求解的解空间的范围。

细心的你可能会发现，我们这里构建所谓的“遍历路径树”并不是一棵树，而是个严格意义上的图。在这里我们需要重新说明一下，虽然我们将遍历路径树称为树，然而它其实不是真正的树。我们在构建的是一个非平面图，但是它并不平凡，是因为如果我们抛去目的时间关键点D不看，那么整个我们构建的图的结构就是一棵树。如图 4所示。



图 4遍历路径树的立体表示

这样的一个结构为什么能够遍历所有的G的子集呢？思考这样一个问题，如果我们将元素的所有有序排列全部找到了，那么对于每一个单独的元素来说，它之前（包含它本身）构成的集合都能构成一个唯一的G的子集。这里可能描述的有些抽象，还可以参考一下图 1帮助理解。

而且我们之所以把这样的结构称作遍历路径树，是因为我们实际上是使用树的结构来实现对这样一个非平面图的存储。由于涉及到具体实现过程，所以在此不再赘述，请您参见3.2.2关于路径树生成的详细设计。

之后我们来详细探讨有关遍历路径树的剪枝方法。我们通过上述生成算法得到的仅仅是一个没有考虑到时间限制的一个简单遍历。在我们的整个解空间里还充斥着不符合条件限制的解。那么我们就需要将解空间进一步缩小，以得到最精简的可行解空间。

换言之，遍历路径树的剪枝也就是在解集P中选取所有满足时间要求的子集PL。如果按照完全的集合操作，那么我们的时间成本会直线上升。这里又体现了使用路径树的一个优点。我们仅仅需要剪短一条边，就可以去掉后续所有的不可行解，如图 5所示



图 5 遍历路径树的剪枝

我们将从起点到该普通路径点的时间消耗以数值的形式存储，并将其与时间限制做比较。如果时间消耗超出了要求，就将该普通路径点从树中剪下，与此同时，其子节点也将一并被舍弃。例如图 5中红色标出的连接，是在该限制条件下的剪枝结果。

我们在完成了对路径树的剪枝以后，就需要对整个路径树进行排序。这样是方便我们根据策略选择具体的路径。如果我们将所有节点的孩子节点进行升序排序，那么我们就可以使用全局最小的策略选择路径，并根据用户给定的深度限制，得出该策略下的局部最优解。

* + 1. 最短优先

我们在完成定义 1所描述的算法的过程中，也许在时间关键点有序对<S, D>用尽的同时，普通路径点集G仍然非空。这就意味着还有普通路径点没有被分配。这个时候不能将这些路径点弃之不管，所以需要有一个候补子算法来弥补这个缺陷。



图 6 最短优先

* + 1. 重复排除
  1. 算法特点

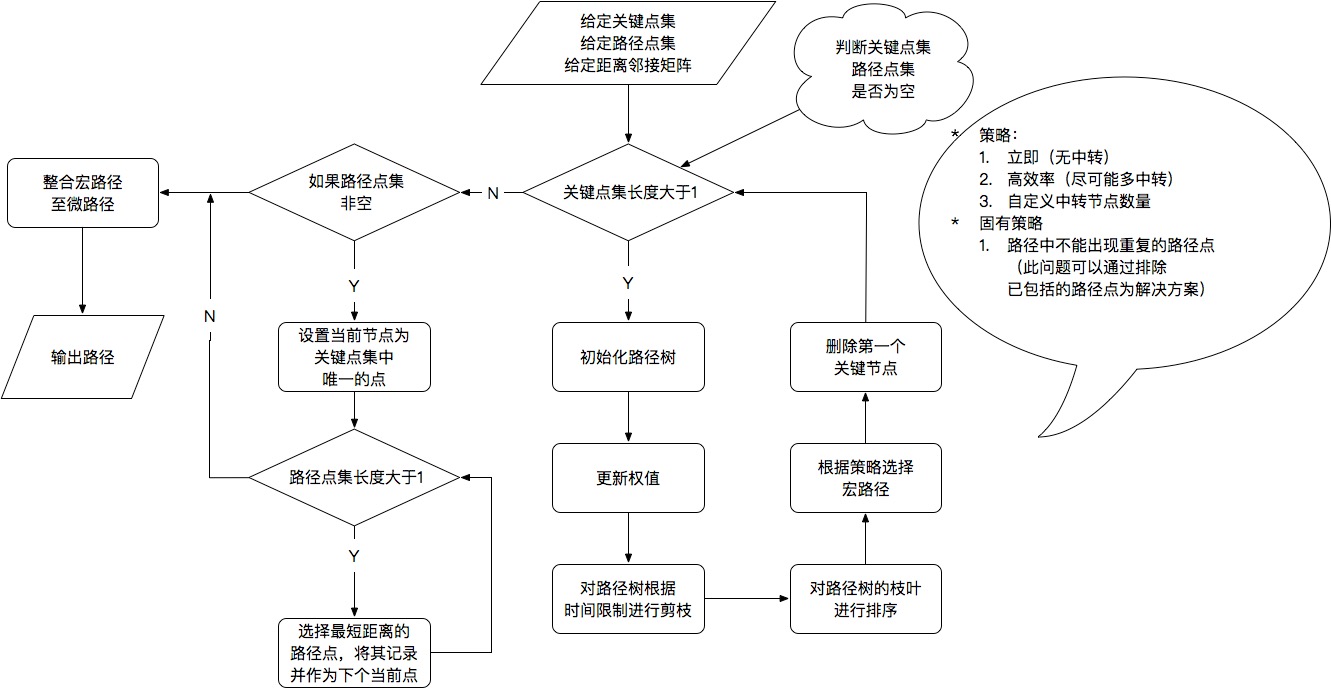
1. 详细设计
   1. 架构设计
      1. Web前端



* + 1. Flask与后台







* + 1. 算法模块
  1. 算法实现
     1. Floyd矩阵与距离更新
     2. 路径树的生成与选路策略
     3. 路径输出
  2. 算法分析
     1. 时间复杂度
     2. 空间复杂度
     3. 可行性分析

1. 测试
   1. 模块测试
      1. 前端测试
      2. 后端测试
      3. 算法模块测试





* 1. 整体测试
  2. 测试总结

1. 总结与提高
   1. 总结与提高
   2. 写给自己的话
2. 附录
   1. 算法API手册
   2. 服务器API手册
   3. web交互手册

1. 由于Floyd求解路径时涉及到回朔(back trace)路径，因此查询复杂度与回溯复杂度为O(n+1)=O(n)。 [↑](#footnote-ref-1)
2. 遍历路径树：通过遍历所有可能的顶点形成的图结构，后文会简称为路径树。实际上并不是逻辑上的树结构，我们会在2.2.2以及3.2.2中详细探讨这个问题。 [↑](#footnote-ref-2)
3. 时间关键点：具有时间限制要求的顶点 [↑](#footnote-ref-3)
4. 完全遍历路径树：没有经过时间限制剪枝的遍历路径树 [↑](#footnote-ref-4)
5. 链接状态矩阵组由邻接矩阵和距离矩阵组成 [↑](#footnote-ref-5)
6. 普通路径点是不带有时间限制的顶点 [↑](#footnote-ref-6)
7. 这两种表分别为邻接表以及十字链表。邻接表是将与此顶点相邻的节点标识符以链表形式存储起来。十字链表可以理解为将邻接矩阵以链式存储的结果。两种方法都需要维护指针，尤其是后者。 [↑](#footnote-ref-7)
8. 关键点指的是时间关键点与普通路径点组成的问题所涉及的抽象层次。 [↑](#footnote-ref-8)
9. 这里之所以与上文描述不符，是因为生成算法未包含剪枝算法。为了区分开来，我们在这里分开定义。 [↑](#footnote-ref-9)