

一种新的智能公交出行建模方法及其实现

常 飞,方 钰,张栋良

(同济大学 计算机科学与技术系,上海 201804;

同济大学 嵌入式系统与服务计算教育部重点实验室,上海 201804)

摘 要:城市公共交通是城市交通网络的基础组成部分,为了给市民提供快捷、方便、经济地公交出行线路方案,智能交通系统中的公交出行系统的设计与实现显得非常重要。提出了一种新的智能公交出行模型,该模型采用由下而上三层结构,通过运用多标准评价的数学方法将公交多目标复杂因素包装成权值,同时给出了结合邻居结点查询技术的基础建模过程。最后经性能测试与项目测试实验证明这种方法具有较强的实用性。

关键词:智能公交;建模;多目标;邻居结点查询

中图分类号:TP338

文献标识码:A

文章编号:1673-629X(2008)12-0220-04

A New Modeling Method and Implementation of Intelligent Public Transportation

CHANG Fei, FANG Yu, ZHANG Dong-liang

(Department of Computer Science and Technology, Tongji University, Shanghai 201804, China;

Ministry of Education Key Laboratory of Embedded System and Service Computing,

Tongji University, Shanghai 201804, China)

Abstract: City public transportation is the basic part of the public transportation network. To offer a fast, convenient and economy public transportation line scheme, the design and implementation of public transportation system in ITS becomes important. Presents a new model of intelligent public transportation. This model uses three-level structure from the bottom up to the apex, this approach regard more complex multi-target factors as weight through the use of multi-criteria evaluation of mathematical methods, at the same time gives basic modeling process including the neighbor node technology. At last, performance test and experimental study show that this method has great practicability.

Key words: intelligent public transportation; modeling; multi-target; neighbor node search

0 引 言

城市公共交通是城市交通网络中的最重要组成部分,为了给市民提供快捷、方便、经济的公交出行线路方案,智能交通系统中的公交出行系统的设计与实现就显得非常重要。

关于公交出行的研究大多关注于公交出行的算法,公交出行的算法主要可以分为三类:第一类是基于矩阵相乘的方法^[1],即把公交站点可达情况以邻接矩

阵的方式保存下来,每次矩阵相乘即相当于换乘一次,这种方法是以前换乘最少为目标的方法,由于是以矩阵的方法保存信息,空间开销较大并且矩阵相乘时间复杂度较大为 $O(n^3)$ 。第二类是苏啸^[2]提出的关联的方法和傅冬绵^[3]提出的广度优先算法是通过关系的演算,即通过可达集合的扩展方法来减少计算量,虽然通过关系的演算简化了结果的计算,如在公交应用中需要考虑步行的情况,则这种方法无法满足需求。第三类是以王莉^[4]等的基于Dijkstra及其改进的方法,这种方法的优点是可以根据相应的权值算出最优的乘车路线,缺点是速度比较慢。

近年来对公交出行模型的研究也有零星研究,如杨新苗^[5]提出了以前换乘次数最少为首要目标,出行距离最短为第二目标的基于GIS的公交乘客出行路径选择模型,但是这种模型定位于换乘次数和出行距离

收稿日期:2008-06-09

基金项目:国家高技术研究发展计划(863计划)资助项目(2007AA01Z425);国家重点基础研究发展计划(973计划)前期研究专项(2007CB316502);国家自然科学基金项目(90718015,60673157)

作者简介:常 飞(1985-),男(回族),安徽芜湖人,硕士研究生,研究方向为并行计算、WebGIS-T;导师:蒋昌俊,博士生导师,教授,研究方向为网格计算、Petri网。

为主要考虑因素具有不小的局限性。

1 基于多目标决策的公交模型

1.1 模型架构

结合公交情况的特点,文中提出了一种公交模型,具体的架构如图 1 所示。

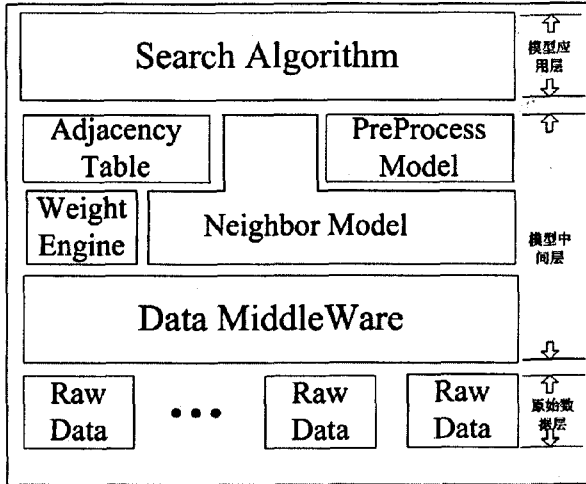


图 1 公交模型架构

该模型自下而上分为三层:原始数据层、模型中间层和模型应用层,构成了从原始数据采集到模型构建及具体公交查询应用的完整架构。

其中 Data MiddleWare 是通过对原始数据的处理来达到模型所需数据,并为架构其上的其它部分提供基础性的统一信息服务,WeightEngine 是计算权值模块,它的功能是通过 Data MiddleWare 中多目标因素的抽取,然后运用相关技术对多目标因素进行平滑,最终完成权值的计算引擎,Neighbor Model 是处理公交站点的邻居站点的模块,它所承担的任务是为 Adjacency Table 的创建和 PreProcess Model 过程中需要寻找公交站点的邻居站点的时候提供服务,Adjacency Table 是一个邻接表,它是模型中间层处理过程中的最终数据结构,Search Algorithm 则是具体算法的使用。

1.2 多目标评价体系

为了解决在公交系统中衡量公交多目标因素如乘车因素中时间、车票价格,步行中的距离和时间等因素对出行线路选择的决策问题,引入多目标评价体系,即建立相应的目标因素与代价的映射关系,由此文中建立相应的多目标体系数学表达式。

定义 1: $Cost = \sum_{i=1}^n w_i f_i$, w_i : 对应于因素的权值, f_i : 对应于代价值的平滑因子。

在这代价函数的求解过程中困难的是平滑因子的分配,当系数较少时,可以通过多变量的数据拟合合法^[6]来修正,当因素较多时,可以利用二项式系数法^[7],这

种方法近似服从正态分析比较接近实际,当然这些方法都可以在实践中根据具体情况进行调整。

1.3 模型基本概念

文中给出以下基本定义,作为模型的依据:

定义 2: 同线路可达集 $lineSet(s) = \{d \mid index(s) < index(d) \wedge Line(s) = Line(d)\}$ ($index$: 站点在线路中的下标, $Line$: 线路)

定义 3: 同站点换乘集 $stationSet(s) = \{d \mid stationName(s) = stationName(d) \wedge Line(s) \neq Line(d)\}$ ($stationName$: 站点名, $Line$: 线路)

定义 4: 步行换乘集 $footSet(s) = \{d \mid distance(s, d) < r\}$ ($distance$: 站点 s 与站点 d 的距离)

定义 5: 搜索集 $search(s) = \{d \mid d \in lineSet(s) \vee d \in stationSet(s) \vee d \in footSet(s)\}$ ($lineSet$: 同线路可达集, $stationSet$: 同站点换乘集, $footSet$: 步行换乘集)

如图 2 所示,图中有四条线路,线路自左而右,以 B 站点为例, B 站点经一号线可到达 C、D,经二号线可以到达 G、H,同站点但不同线路即线路一、线路二的 B 站点, B 站点步行一段 r 距离可以到达 I、K,由此得到 $lineSet(s) = \{G, H\}$, $stationSet(B) = \{B'\}$, $footSet(B) = \{I, K\}$, $search(s) = \{G, H, F, I, K\}$ 。

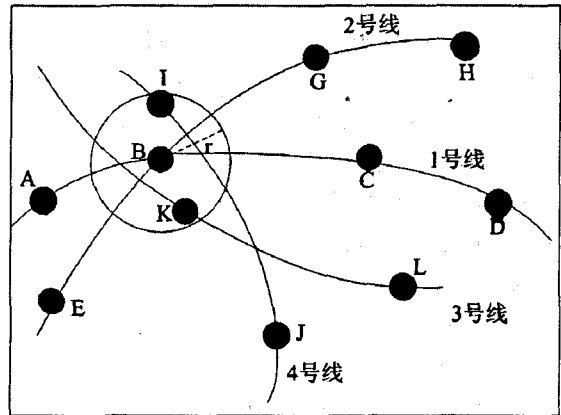


图 2 公交线路示意图

1.4 建模过程

在定义 4 中文中提到步行换乘集或者 1.1 中的预处理模块用到的 k 个最近邻居,都需要查找站点的邻居站点,找邻居站点分为二种情况:

(1) 网格文件的方法。网格文件^[8]是一种二维索引结构,而且它的插入和查询都比较简单,当城市面积不大并且公交站点分布比较均衡时可以采用这种方法,但是对于城市面积比较大,可能存在大量的索引结点是空的情况,空间利用率过低的问题。

(2) R-Tree^[9]的方法。R-Tree 最大的优点是作为利用 B 树特性的一种多维索引,可以很好地支持点

集和区域集的范围查询,并且时间复杂度和空间利用率都比较好,比较适合文中的公交站点集查询。

前面介绍了公交模型的数据结构及实例,公交建模的思想主要就是根据公交线路的信息和站台地理信息,并在内存中建立一个 R-Tree 用于站点集的插入和搜索,逐步建立搜索集并最后构建出公交模型邻接表的过程,建模算法如下:

输入: 公交线路表 $l = \bigcup_{i=1}^m (\text{lineName}, \bigcup_{j=1}^n s_j)$ (lineName : 线路名, s_j : 站点, m : 线路数, n : 线路 s_j 的站台数)、站点经纬度 $d = \bigcup_{j=1}^n (\text{stationName}, x_i, y_i)$ (stationName : 站点名, x_i : 经度, y_i : 纬度, n : 站点数)

输出: 邻接表 $G = (V, E)$, V 为结点集合, E 为边集合

算法:

(1) 初始化站点集 Set 为空集,初始化站点距离搜索树 R-tree,初始化哈希表 LineNameHash,令 $i = 1$;
(2) $\forall (S_j, S_k) \in S \wedge \text{index}(j) < \text{index}(k)$, $(x'_k, y'_k) = \text{Projector}(x_k, y_k)$,将 $\text{weight}_{j,k}, s_k$ 信息作为一个结点插入邻接表 lineSet 项;如果 $S_j \notin \text{Set}$, $\text{Set} = \text{Set} \cup \{S_j\}$,通过坐标 (x'_j, y'_j) 将 S_j 插入 R-tree,构造 $(\text{key}: S_j, \text{value}: \text{lineName})$ 插入 LineNameHash 中, S_k 同理;Projector: 将经纬度转换为大地坐标的投影函数,如麦卡托投影、高斯投影

(3) $i = i + 1$,如果 $i < m$,则转(2);

(4) 遍历 StationNameHash, 计算插入邻接表 stationSet;

(5) 遍历集合,设当前坐标为 (x_0, y_0) 并构造要查找的矩形区域 $D = ((x_0 - r, y_0 - r), (x_0 + r, y_0 + r))$,通过 R-tree 查询算法找出邻居站点信息,插入邻接表 footSet;

(6) 建模完成。

2 模型实验

2.1 测试系统

文中所提出的公交模型思想被应用于智能信息服务平台(如图 3 所示)的公交子系统中,作为测试平台,对杭州市的 256 条线路,1132 个站点进行了测试。本测试系统环境为内核英特尔® 至强® 5310 系列处理器,主频为 2.00GHz,内存容量为 4GB,主机系统为 Unix 操作系统,图 3 为系统的效果图。

效果图所示为系统点选从起点经纬度坐标为



图 3 公交系统查询方案效果图

(120.125564575195, 30.2663440704346), 到终点坐标为 (120.176109313965, 30.2688827514648) 的乘车方案(见表 1)。系统从起点搜索到求是路、浙大附中和黄龙洞三个站点,终点搜索到新华路口、宝善路口和潮鸣寺巷等站点,然后进行搜索,第一种方案虽然换乘了一次,但是路途比较短,站点较少,而第二种方案不需要换乘,但是路途较第一种方案较远,可见系统找出的两条路径效果较好。

表 1 公交系统查询方案

名称	方案	标识	备注
方案 1	步行约 200 米到黄龙洞乘 K850 下行到武林门西,然后换乘 11/K11 路下行到宝善桥建国路口,然后步行约 50 米到终点	图 3 偏直且经过松木场线路	换乘了一次,路途比较短,站点较少
方案 2	步行约 100 米到浙大附中,然后乘 K101 下行到新华路口,然后步行约 150 米到终点	图 3 弯形且经过黄龙体育中心线	不需要换乘,但路途较第一种方案远

2.2 性能测试

文中提出的公交出行模型刻画了三种基本公交出行子模型,文中选取改进的换乘最小算法^[3]和改进 Dijkstra 算法^[10]作为测试算法,通过对起始点到终点站的随机选取作为查询值,通过多次查询来测试系统的模型的平均响应时间从而测试模型的可用性,测试结果如图 4 所示。

图 4 给出了运用本模型 2.1 所述两种算法作为测试算法,分别以 1000, 2000, 3000, 4000, 5000 次请求来计算出平均响应时间,可以看出,两种算法的平均响应时间都要小于 1 秒,而且随着请求次数的增加,刚开始请求次数较小时,个体响应时间可能差异比较大,随着响应次数的增加,平均响应时间呈现缓慢的减少,最后基本维持不变,可见只要运用有效的查询算法,本模型完全可以适应大规模路网情况下的公交查询。

3 结束语

文中提出了一种新的智能公交出行模型,运用数学的方法解决了公交多目标决策中的问题,结合邻居结点查询技术,具有一定的智能性,最后运用两种搜索

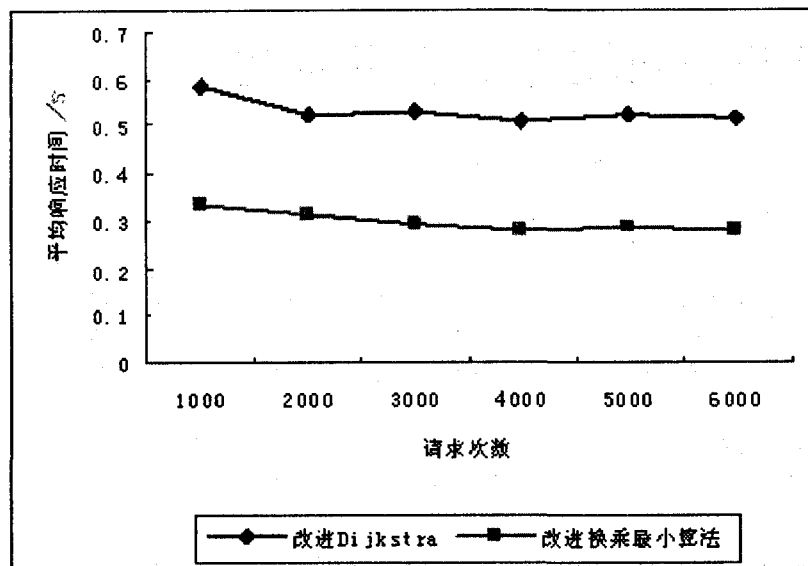


图 4 采用改进 Dijkstra 和改进的换乘
最小算法平均响应时间图

算法对本模型进行验证,并成功运用在智能交通信息平台测试版中,取得了较好的效果,文中通过公交乘车的宏观因素来分析问题,对以后相应的公交乘车系统的开发具有一定的实用意义。但该模型同样也存在需要改进的地方,如对实时公交路况等的支持上,这也将成为今后工作的重点。

参考文献:

[1] 张林峰,范炳全. 公交网络换乘矩阵的分析与算法[J].

系统工程,2003,21(6):92-96.

[2] 苏 啸,曾子维. 基于关联的城市公交换乘查询算法[J]. 计算机工程与设计, 2006,27(3):519-521.

[3] 傅冬绵. 交通系统中最少换乘算法及其实现[J]. 华侨大学学报:自然科学版, 2002,22(4):348-350.

[4] 王 莉,李文权. 公共交通系统最佳路径算法[J]. 东南大学学报:自然科学版, 2004,34(2):264-267.

[5] 杨新苗,王 炜,马文腾. 基于 GIS 的公交乘客出行路径选择模型[J]. 东南大学学报:自然科学版,2000,30(6):87-91.

[6] 徐萃薇,孙绳武. 计算方法引论[M]. 北京:高等教育出版社,2002:67-68.

[7] 张维明,邓 苏. 信息系统建模技术与应用[M]. 北京:电子工业出版社,1997: 234-235.

[8] Nievergelt J, Hinterberger H, Sevcik K C. The Grid File: An Adaptable, Symmetric Multikey File Structure [J]. Acm Transactions On Database Systems, 1984, 9(1):38-71.

[9] Guttman A. R-trees: a dynamic index structure for spatial searching[C]//proceedings of the 1984 ACM SIGMOD international conference on Management of data. New York: ACM Press, 1984:47-57.

[10] 叶 青,陈国中. 基于预处理剪枝的最短路径算法[J]. 计算机工程,2007,43(9):205-207.

(上接第 216 页)

[4] 郭 健,陈孝威. 基于颜色分布的图像检索技术[J]. 计算机工程与应用,2006(6):44-47.

[5] 王永良,陈新度,刘 强,等. 基于内容的墙纸 Web 检索系统的研究与实现[J]. 计算机应用研究,2006,23(6):167-169.

[6] 刘 笛,朱学峰,苏彩虹. 一种新型的模糊 C 均值聚类初始化方法[J]. 计算机仿真,2004,21(11):148-151.

[7] Kim Tae-Wan, Li Ki-Joune. A distance based packing method for high dimensional data[C]//Proceedings of the Fourteenth Australasian database conference on Database technologies. Adelaide, Australia: [s. n.], 2003:135-144.

[8] 张培珍,付 平,肖 军,等. 基于快速聚类索引的图像检索系统[J]. 吉林大学学报:信息科学版,2004,22(6):638-642.

(上接第 219 页)

schemes for IP traceback[C]//In: Proceedings of the 2001 IEEE INFOCOM Conference. Anchorage, Alaska: [s. n.], 2001.

[2] Carter R, Crovella M. Dynamic Server Selection Using Dynamic Path Characterization in Wide-Area Networks[C]//In: Proceedings of the 1997 IEEE INFOCOM Conference. Kobe, Japan: [s. n.], 1997.

[3] Cooperative Association for Internet Data Analysis, Skitter analysis[EB/OL]. 2000. <http://www.caida.org/tools/measurement/skitter/>.

[4] Savage S, Wetherall D, Karlin A, et al. Practical Network Support for IP Traceback[C]//Proceedings of the 2000 ACM SIGCOMM Conference. Stockholm, Sweden: [s. n.], 2000: 295-306.

[5] Savage S, Wetherall D, Karlin A, et al. Network support for IP traceback[J]. IEEE/ACM Transactions on Networking, 2001,20(2):226-237.

[6] Deering S, Hinden R. Internet Protocol Version 6 (IPv6) specification[S]. RFC 2460. 1998.