

doi 10.3969/j.issn.1007-855x.2009.04.014

公交网络的分层数据模型与智能查询算法

舒新峰

(西安邮电学院 计算机系, 陕西 西安 710121)

摘要:为解决当前公交查询系统普遍存在的查询效率低且仅能进行同城公交换乘查询等问题,提出了基于分层模式的公共交通网络模型,可对全国范围内的交通网络进行统一管理,并设计了新型的交通查询算法,实现了任意两地间的最佳乘车方案查询.和现有系统相比,新方案具有良好的可扩充性和较低的时间复杂度,可用于各类型公交查询系统开发.

关键词: 交通网络; 公交查询; 最优路径; 数据结构

中图分类号: TP302 **文献标识码:** A **文章编号:** 1007-855X(2009)04-0063-05

Tiered Data - Model for Public Transit Network and Intelligent Traffic Inquiring Algorithm

SHU Xin-feng

(Department of Computer Science and Technology, Xi'an Institute of Posts and Telecommunications, Xi'an 710121, China)

Abstract: To solve the problems existing in the current public transport inquiring systems such as low efficiency and limitation to route search between two sites only in the same city, a tiered data model is proposed to organize all the public transit networks within China. A novel traffic inquiring algorithm is also presented with the ability of searching the optimized route between any two cities. Compared with the available solutions, the new technique has a good extendibility and lower time complexity, and can be employed to the development of various types of public transport inquiring systems.

Key words: transit network; public transport query; optimized route; data structure

0 引言

公共交通是交通网络的重要组成部分,具有方便、经济的特点而成为人们出行的首选.随着国民经济的发展和城市化进程的推进,城市间以及城市内部建立了庞大的公共交通网络,为人们出行提供了极大的便利.然而面对错综复杂的交通网络,即使在同一城市两地之间选择一个最佳乘车路线也绝非易事,更不用说跨城市出行.因此十分有必要开发一套通用、高效,且易于管理与扩充的智能交通查询系统,能搜索任意两地间的最佳乘车方案,为人们的本地或异地出行提供便利.

国内在交通查询系统的技术研发方面已经取得一些重要成果.在出行路线的选择上,调查研究表明乘客主要考虑的因素是换乘次数,其次是时间和票价^[1].在公交网络的描述上,选用邻接表已经成为普遍共识,具体数据则使用数据库或者 XML 文件形式进行存储^[2,3].在最佳乘车路线的计算上,除了使用各种形式的 Dijkstra 和 Floyd 算法^[4~6]外,文献 [2] 通过数据库的查询功能和关系运算功能实现了最优换乘的搜索;龚翱等 [7] 将网络路由扩展算法移植到公交线路查询中,利用嵌入式系统中有限的软硬件资源实现了最佳路径查询.在查询方式上,将 Web-GIS 和移动通讯等技术融合到查询系统的开发中,支持字符或图

收稿日期: 2009-05-12

作者简介: 舒新峰 (1975-), 男, 硕士, 讲师. 主要研究方向: 软件工程、算法与数据结构、形式化技术.

E-mail: shuxinfeng@gmail.com

形界面模式,并提供单机查询、基于 Web 的查询以及短信查询^[7~9]等。

目前,虽然北京、上海等城市已经有交通查询系统投入使用,实现了基本站点查询、线路查询和两地间的乘车路线查询,获得了可观的经济和社会效益。然而,现有的系统效果不太理想,存在的主要问题包括:

最短路径评估简化为仅统计经过的车站数或者公里数,没有对出行影响因素进行综合考虑。现有查询算法均是以公共交通网络为基础,从出发站点开始逐站点以扩散方式向四周搜索,算法效率低下。特别在大城市站点多、交通线路复杂的情况下,搜索时间难以容忍。查询仅限于城市内部,不能跨城市查询,对异地出行十分不便。

针对以上问题,在对国家交通网络特征分析的基础上,提出了基于分层模式的公共交通网络数据模型,实现了城市内部和城市间交通网络的统一管理,并设计了新型的查询算法支持任意两地间最佳乘车方案的快速查询。

1 交通网络数据建模

1.1 数据组织模型

我们国家的公共交通网络由城市交通网和城际交通网复合而成。对于这样一个规模极其庞大的交通网络,如果用一个图来存储,一方面由于结构过于复杂而难以管理和维护,另一方面在这样的一个图中进行任意两地间的最佳乘车路线查询基本上是一项不可完成的任务。

考虑到公交网络里面的交通线路绝大多数是运行于城市内的公交线路,而城际线路运行于城市之间且数量较少,因此在数据的存储与组织上可以采用分层模式以不同抽象粒度对交通网络进行描述。如图 1 所示,第 1 层为城市级交通网络,用一个有向图来描述,图中将每个城市抽象成一个节点,节点间通过城际交通线路连接起来;第 2 层为站点级交通网络,由多个有向图组成,每个有向图对应一个城市。

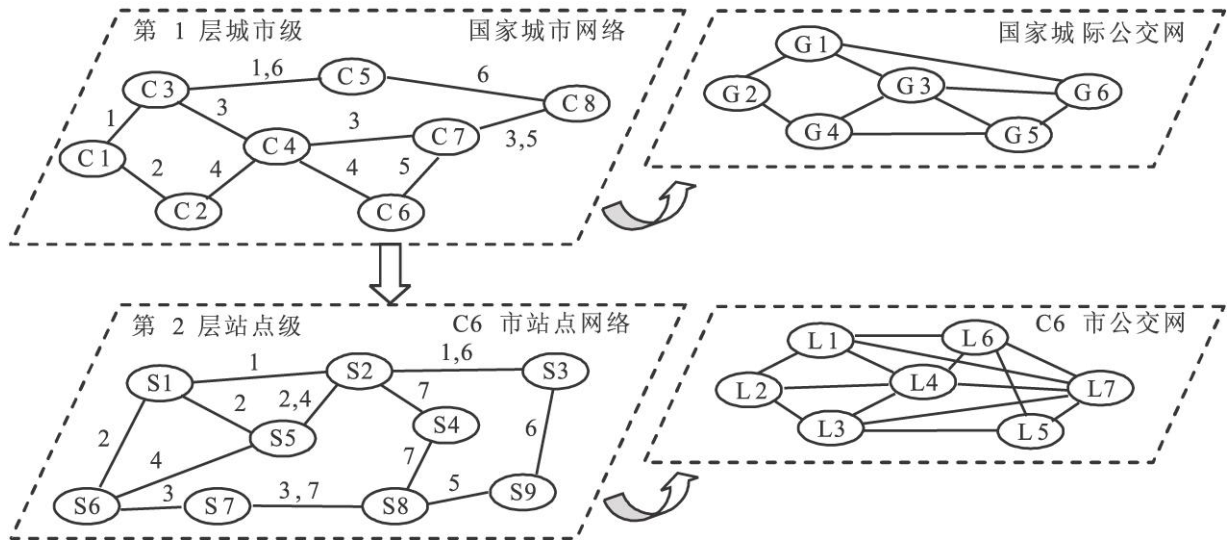


图 1 系统的分层数据模型
Fig 1 Tiered data - model of the system

虽然采用国家城市网络与城市站点网络相结合的模式已经具备了足够的信息来求出任意两地之间的最短路径,然而为了提高查询速度,避免在图中按节点进行最佳换乘方案和乘车路线的搜索,将站点网络和城市网络里面出现的公交线路提取出来,分别构成城市公交网(每城市均有 1 个)和国家城际公交网。在公交网里,将每个公交线路抽象成一个节点,线路间的换乘关系描述为节点间的邻接关系。

1.2 数据描述

交通网络的主要数据定义如表 1 所示. 城市网络及站点网络分别给出了城市和站点在地理位置上的拓扑关系, 由于公交车的出发路线和回程路线可能有所不同, 城市及站点网络均为有向图. 城市(站点)间的弧直观的体现为通过相关联一对顶点的公交线路的集合, 然而不同的公交线路在 2 个城市(站点)之间实际走的路线可能有所不同, 因此在图中允许 2 个顶点间存在多条弧. 为提高查询效率, 这里给不同的公交线路定义不同的弧. 城市网络及站点网络均为稀疏图, 采用邻接表存储.

表 1 交通网络的主要数据定义

Tab 1 Definitions of the key elements in the transit networks

交通子图	数据名称	数据格式
城市网络	城市节点	城市编号 + 城市名称 + {城际车次编号 + 停靠站点 + 到达时间 + 离开时间}
	弧	弧尾城市 + 弧头城市 + 城际车次编号 + 长度
城际公交网络	线路节点	车次编号 + 名称 + 类型 + 出发城市及站点 + 终点城市及站点 + 总公里数 + 计费模式 + 发车间隔 + 发车时间 + 到达时间 + 车速
城市站点网络	车站节点	车站编号 + 名称 + 别名 + 车站级别
	弧	弧尾站点 + 弧头站点 + 车次编号 + 长度
城市公交网络	线路节点	车次编号 + 名称 + 类型 + 起点站 + 终点站 + 总公里数 + 计费模式 + 发车间隔 + 起始发车时间 + 收车时间 + 车速 + 每站停车时间

城际交通网和城市交通网给出了公交车之间的换乘关系. 一般来讲, 2 个城市之间以及同一城市的两地之间换乘次数不应该多于 3 次, 因此公交图是一个稠密图, 使用邻接矩阵存储比较合适.

为解决同名问题, 每个城市和车站均有一个唯一的编号, 其中城市编号为 4 位数字串, 车站编号为长度为 8 位的数字串且前 4 位为所在城市的编号. 此外, 每条城市和城际公交线路均有一个唯一编号; 线路类型为公共汽车、地铁、火车、长途汽车等; 线路的计费模式包括一票制、按公里计费和分段计费(如每 6 站进位 1 元)等.

在对交通网络进行数据建模时, 可结合 Map Info 等地理信息系统软件进行. 首先将国家和城市的交通地图扫描后进行矢量化处理, 抽取出交通网络的节点和边, 并在此基础上定义交通网络的站点、弧和公交的车次及线路信息.

2 智能搜索算法设计

为了描述方便, “换乘方案”指从一个站点到达另外一个站点需要乘坐的公交车序列 $H_1: \dots: H_n$, 其中 H_i 表示城市或城际公交车编号; “乘车路线”或“乘车方案”则表示从一个站点到达另外一个站点需要乘坐的公交车及换乘站点序列 $Q_1, H_1, \dots, Q_m, H_m$, 其中 Q_i 表城市编号或者站点编号.

当出发站点和目标站点间有多个乘车方案可供选择时, 虽然多数人考虑的主要因素依次为转车次数、时间和费用, 然而由于不可预知因素导致公交车进站的随机性以及乘客乘车的随意性, 仅以这些因素为依据对计算出来的乘车方案进行排序, 提供多种方案供乘客选择. 下面, 先给出同城两地之间的最佳乘车路线查询算法, 然后将其推广到任意两地之间.

2.1 同城两地之间查询

令存储车站网络的有向图为 $G_s = (V_s, E_s)$, 存储公交网络的无向图为 $G_b = (V_b, E_b)$, 对于任意给定的乘车出发站点 S_i 和目标终点站点 S_e , 搜索算法分为 2 个阶段进行, 首先在公交网络里搜索最佳换乘方案, 然后根据得到的换乘方案在站点网络里搜索最佳乘车路线.

第 1 阶段: 搜索最佳换乘方案. 在城市公交网络里, 以最小换乘次数为约束条件计算所有可能的换乘方案并存储在集合 C 中. 换乘方案的计算是基于图广度优先遍历的改进算法, 可一次计算出从 B_s 出发, 经

过最少 i 次换乘后到达 B_e 的所有可能方案,具体过程如下:

1) 在站点图 G_s 中,分别求出经过站点 S_i 和 S_e 的公交车集合 B_s 和 B_e

2) 令 $C = B_s \cap B_e$, 如果 C 不为空,则该集合给出了所有可直接到达的公交线路,计算停止;否则转第 3) 步.

3) 令变量 $i = 0$, 集合 $A = B[0] = B_s, N = B[1] = \emptyset$, 其中 A 是已访问到的公交车集合, $B[i]$ 存储的是从 S_i 出发的所有可能的 i 次换乘方案.

4) 对于集合 $B[i]$ 中每一个换乘序列 $L = L_1: \dots: L_i$, 对于任何 $\langle L_i, L_k \rangle \in E_B$, 则

a) 如果 $L_k \in B_e$, 则令 $C = C \cup \{L: L_k\}$;

b) 否则, 如果 $L_k \notin A$, 令 $B[i+1] = B[i+1] \cup \{L: L_k\}, N = N \cup \{L_k\}$.

5) 如果 $C \neq \emptyset$, 则 C 中为 $i+1$ 次的换乘方案, 搜索停止; 否则 $A = A \cup N, i = i+1, B[i+1] = \emptyset$, 转第 4) 步继续搜索.

第 2 阶段: 搜索最佳乘车方案. 由于人们实际的出行成本包含时间与金钱两方面, 为了便于比较, 需要将票价 P 转换为时间(分钟), 其转换函数^[10]为

$$T(P) = \frac{\text{票价 } P \times \text{年法定工作天数} \times 480}{\text{人均年收入}} \quad (1)$$

此外, 乘车总时间包括车辆行进、红灯停车、车站停车和换车等候的时间, 而且这些时间容易受到天气、乘客流量、车流量以及道路状况等诸多因素影响. 为了简化计算, 这里设定所乘公交车的速度与每站停靠时间均恒定(取决于线路的类型和站点的级别), 且每次换车等候时间为换车车辆发车周期的 $1/2$

下面结合城市车站图 G_s , 依次计算 C 中每一个换乘方案对应的最佳乘车路线, 并以三元组 $\langle ; P; D \rangle$ 的形式存储在数组 MinPath 中, 这里 $\langle ; P; D \rangle$ 为乘车路线, P 为总票价, D 为乘车代价. 为了方便计算, 还需要定义乘车信息六元组 $\langle ; ; ; P_1; P_2; D \rangle$, 其中 $\langle ; ; ; P_1; P_2; D \rangle$ 为从当前站点到达目标站点的换乘方案, $\langle ; ; ; P_1; P_2; D \rangle$ 为从出发站点到最近一次换乘前所经过的乘车路线, $\langle ; ; ; P_1; P_2; D \rangle$ 为最近一次换乘到当前站点所经过的站点序列, P_1 与 P_2 分别为这两段路线的票价, D 为到从起点到达当前站点代价. 最佳乘车路线算法的核心思想和 Dijkstra 最短路径算法类似, 不同之处在于进行路径搜索时是在换乘方案的引导下进行的, 避免了大量的无效计算, 具体算法过程如下:

1) 令变量 $i = 1$ (i 为 C 中换乘方案的编号, $1 \leq i \leq |C|$)

2) 对于 C 中第 i 个换乘方案 $L_i = L_1: \dots: L_k$, 按照下面步骤计算其对应的最佳乘车路线:

a) 令集合 $A = \{ \langle ; S_i; S_i; 0; 0; R(L_1)/2 \rangle \}$. A 用于存储所有可能的乘车路线, $R(L)$ 为公交车 L 的发车周期函数.

b) 在集合 A 中选择一个 D 值最小的六元组 $\langle L_j; L_{j+1}: \dots: L_k; ; S_1, \dots, S_k; P_1; P_2; D \rangle$. 如果 $S_k = S_e$, 转 e), 否则转 c).

c) 对于从 S_k 出发的任何一个弧 $\langle S_k, S', L, d \rangle$, 如果 S' 不在 S_1, \dots, S_k 中, 则: 如果 $L = L_j$, 令 $D' = D + d/\text{Speed}(L_j) + \text{STime}(L_j, S_k)$; 如果 $L = L_{j+1}$, 令 $D' = D + d/\text{Speed}(L_j) + \text{STime}(L_j, S_k) + T(\text{Price}(L_j, S_k) - P_2)$. 给 A 中添加六元组 $\langle L_j; L_{j+1}: \dots: L_k; ; S_1, \dots, S_k; P_1; P_2; D' \rangle$. 此处 $\text{Speed}(L)$ 及 $\text{STime}(L, S_k)$ 为公交车 L 的速度及 L 在站点 S_k 的停车函数, $\text{Price}(L, S_k)$ 为公交车 L 在 S_k 段的票价函数. 如果 $L = L_{j+1}$, 令 $P_1' = P_1 + P_2, P_2' = \text{Price}(L_{j+1}, S_k, S')$, $D' = D + R(L_{j+1})/2 + d/\text{Speed}(L_{j+1}) + \text{STime}(L_{j+1}, S_k) + T(\text{Price}(P_2'))$, 给 A 中添加新的六元组 $\langle L_{j+1}: \dots: L_k; ; S_1, \dots, S_k; S_k, S'; P_1'; P_2'; D' \rangle$.

d) 将六元组 $\langle L_j; L_{j+1}: \dots: L_k; ; S_1, \dots, S_k; P_1; P_2; D \rangle$ 从 A 中删除, 转 b).

e) 此时, 六元组 $\langle L_k; ; S_1, \dots, S_k; P_1; P_2; D \rangle$ 保存的是按照换乘方案 L_i 得到最佳乘车路线, 令 $\text{MinPath}[i] = \langle ; L_k, S_k; P_1 + P_2; D \rangle$.

3) 如果 $i < |C|$, 令 $i = i+1$, 转 2); 否则转 4).

4) 对数组 MinPath 里的三元组按照代价域 D 排序, 计算完成.

2.2 任意两地之间查询

对于任意给定的出发站点 s_i 和目标站点 s_e , 最佳乘车方案搜索过程如下:

1) 获取 s_i 和 s_e 所在的城市 C_i 和 C_e 如果 C_i 和 C_e 相同, 按照同城搜索算法进行; 否则转第 2) 步.

2) 利用城市网络和城际交通网络搜索出 C_i 和 C_e 间的最佳城际乘车路线, 策略与同城两地间的搜索算法类似, 此处不再赘述. 由于城际交通线路以火车和长途汽车为主, 往往有固定发车时间, 所以换乘时间应该以到达时间和发车时间差为准. 把得到的最佳城际乘车方案保存在数组 CityminPath 中, 数组元素格式为三元组 $\langle C_1, G_1, \dots, C_{k-1}, G_{k-1}, C_k; P; D \rangle$, 其中 $C_1, G_1, \dots, C_{k-1}, G_{k-1}, C_k$ 为城际乘车路线, P 为总票价, D 为总代价.

3) 对于 CityminPath 中每一个城际乘车方案 $\langle C_1, G_1, \dots, C_{k-1}, G_{k-1}, C_k; P; D \rangle$,

a) 在城市网络依次求出每一个城际线路 $G_i (1 \leq i \leq k-1)$ 在城市 C_i 及 C_{i+1} 的停靠站点 S_i' 及 S_{i+1}' ;

b) 在城市 $C_1, C_i (2 \leq i \leq k-1), C_k$ 中按照同城搜索算法分别求出 s_i 到 S_i' , S_i' 到 S_{i+1}' 以及 S_k 到 s_e 的最佳乘车路径, 并将城际乘车线路 $C_1, G_1, \dots, C_{k-1}, G_{k-1}, C_k$ 中出现的城市按顺序替换成对应的最佳乘车路径, 同时将得到的票价和代价累加到 P 与 D 中. 由于城内乘车代价相对与城际乘车代价要小得多, 替换时每城市只选取 MinPath 中排名第一的乘车方案既可.

4) 对 CityminPath 中的所有最终乘车方案按照 D 值排序, 计算完毕.

3 算法性能分析与比较

由于现有算法不能实现城市间乘车路线查询, 此处仅将本文的同城搜索算法和广泛使用的 Dijkstra 算法加以分析比较. 假定待查询城市交通网络的站点个数为 m , 公交线路个数为 n 无论是哪种算法, 在查询最佳乘车路线时既要找出所有最佳换乘方案, 同时还需要对每一种换乘方案求出对应的最佳乘车路线.

研究将搜索过程分两步进行: 首先在公交网络中搜索最佳换乘方案, 搜索过程如图 2(a) 所示, 第 i 层为最少经过 $i-1$ 次换乘所到达的公交车集合, 由于各层的集合互不相交, 不难看出最坏情况下的时间复杂度为 $O(n \log n)$; 然后, 在站点网络里依次计算每一个换乘方案对应的最佳乘车路线, 搜索过程如图 2(b) 的粗实线部分所示, 换乘方案确保了计算沿着正确方向进行, 最坏的时间复杂度为 $O(m^2)$. 由于公交线路数量 n 远小于站点数量 m , 故总的时间复杂度最坏为 $O(m^2)$.

传统的 Dijkstra 算法在进行查询时, 直接在站点网络上从起始站点 s_i 开始以扩散方式向四周搜索并计算所有可能乘车路线, 搜索过程如图 2(b) 虚线部分所示, 显然进行了大量的无效计算, 最坏情况下的时间复杂度为 $O(n^2 * m^2)$. 虽然文献 [6] 对传统的 Dijkstra 算法做了改进, 也将查询过程分为最佳换乘方案计算和最佳乘车方案计算两步进行, 并详细给出了最佳换乘方案的求解算法, 然而该算法是基于站点网络完成的, 搜索过程和传统 Dijkstra 算法相同, 效率明显低于论文的算法.

4 结论

交通查询系统在现代人们的出行中扮演着越来越重要的角色. 针对现有查询系统存在的缺陷, 提出了基于分层模式的交通网络模型, 可对国家范围内的交通网络进行统一管理, 并基于此模型设计了新型查询算法, 实现了真正意义上的任意两地间的最佳换乘路线查询. 论文提出的解决方案, 一方面可以通过裁剪以城市为单位发布查询系统来实现同城任意两地间的最佳乘车路线查询, 相反也可以把相互独立的城市公交数据集成起来实现跨城市的最佳路径查询, 具有良好的可扩充性.

(下转第 88 页)

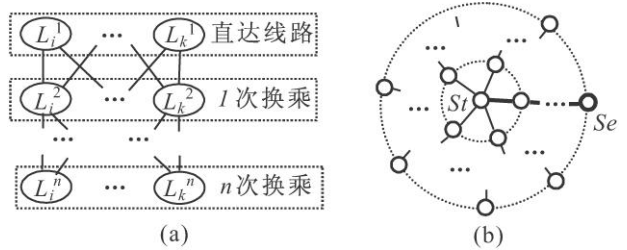


图 2 搜索方案示意图
Fig 2 Sketch of the searching strategy

内涵和企业的视觉形象.而对于“B”而言,需改进的几个环节主要是品牌联想、广告水平、企业视觉形象和产品外观设计等.

根据上面的分析,对 2 个企业提高品牌竞争力可提出以下建议:

A 卷烟品牌生产企业: 提高卷烟用原辅材料的质量; 进一步提高包装质量; 深入挖掘并传播产品的文化内涵; 改善企业的行为准则; 重新设计企业的标志,改善企业视觉形象; 注意创造品牌的联想性及独特的品牌个性; 进一步提升品牌的美誉度及熟悉度; 加大广告、促销的投入力度; 改善企业的公共关系.

B 卷烟品牌生产企业: 深入研究消费者的口味,不断改善产品吸味; 改革传统的包装设计,设计出能够更好迎合消费者心理的外观包装; 深入挖掘并传播产品的文化内涵; 进一步改善企业的行为准则及标识等,树立良好的企业视觉形象; 在创造品牌的联想性方面要再下工夫; 适当调整产品价位,满足目标消费者需求; 提高渠道管理水平; 提高广告质量.

参考文献:

- [1] (美)菲利普·科特勒. 市场营销管理(亚洲版)[M]. 北京:中国人民大学出版社,1997.
- [2] (美)汤姆·邓肯,桑德拉·莫里亚蒂. 品牌至尊——利用整合营销创造终极价值[M]. 北京:华夏出版社,2000.
- [3] (美)迈克尔·波特. 竞争优势[M]. 北京:华夏出版社,1997.
- [4] 冯云廷,李怀斌. 现代营销管理教程[M]. 吉林:东北财经大学出版社,1998.
- [5] 汪培庄. 应用模糊数学[M]. 北京:首都经济贸易大学出版社,1989.
- [6] 牛映武. 运筹学[M]. 陕西:西安交通大学出版社,1994.

(上接第 67 页)

在未来的研究中,将在本文的基础上结合 Web - GIS 等技术以增量模式开发智能交通查询系统:首先建立城际交通网络,并实现城市间的最佳乘车路线查询,该阶段的软件可以满足乘坐火车和长途汽车出行的查询需要;接着逐个建立每个城市内部的交通网络,对于每一个新加入的城市,可满足城市内部以及和其它已有城市间的出行最佳乘车路线查询需要.通过实际系统对本文提出的方案进行进一步的验证和完善.

参考文献:

- [1] 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,2000,30(6): 88 - 91.
- [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] 杨新敏,孙静怡,钱育渝. 城市交通流配流问题的遗传算法求解[J]. 昆明理工大学学报:理工版,, 2002, 27(5): 144 - 147.
- [4] 罗跃军,李霖,朱敦尧,等. 车辆导航系统中最短路径计算的数据模型[J]. 昆明理工大学学报:理工版,2004,29(3): 106 - 109.
- [5] 鲍江宏,关毅璋. 基于矩阵运算的公交查询高效算法[J]. 计算机工程与应用,2008,44(10), 198 - 200.
- [6] 陈箫枫,蔡秀云,唐德强. 最短路径算法分析及其在公交查询应用[J]. 工程图学学报,2001(3): 20 - 24.
- [7] 朱茵,陆化普,周海淞,等. 基于 C/S 与 B/S 混合体系结构的智能交通管理信息系统的设计[J]. 公路交通科技,2005,22(11): 147 - 151.
- [8] 龚翔,朱宁波,史长琼,等. 基于便携设备的城市公交线路查询算法的研究[J]. 计算机工程与设计,2008,29(23): 6124 - 6126.
- [9] 陈小宾,葛新伟,林鸿飞. 基于语义计算的公交移动问答系统[J]. 计算机工程与科学,2008,30(10): 118 - 121.
- [10] 何胜学,范炳全,严凌. 公交网络最优路径的一种改进求解算法[J]. 上海理工大学学报,2006,28(1): 63 - 67.