

数据库系统中公交网络换乘线路的优化选择模型

An Optimized Choosing Model of Bus - Line Changing in PTN in Database System

王世祥 饶维亚 (长春大学理学院 吉林长春 130022)

摘要: 本文给出了超大城市公交网络换乘线路的优化选择模型。以北京市为实例,运用动态 SQL 查询技术,在数据库 SQL Server 中,实现了公交线路换乘查询系统。求出了北京市公交网络的换乘次数上确界,可以查询出任意两个站点之间的换乘线路,给出了上下行线路的处理模型,优化了查询能力,具有实用性。

关键词: 公交网络 换乘线路查询 动态 SQL SQL Server 换乘次数上确界

1 引言

文 [1] 给出了大中城市公交线路查询的数据结构及其算法的实现,文 [2] ~ [6] 也研究了公交乘车线路选择的问题。

由于人口 1000 万以上的超大城市,线路和站点众多,已超出大型数据库数据表 1024 列的限制,因此文 [1] 中采用的线路与站点的对称式数据表的方式已不适合于这种情况,需要另辟途径。

首都北京人口已有 1700 多万,市内公交有 412 条线路、2872 个站点^[9]。我们在大型数据库 SQL Server 中所实现的北京公交网络换乘线路查询的系统,输入公交网络的任意起始站和终到站后,可以很快地给出优选的换乘路线。

首先构造出线路与线路之间的关系表,简称为线线表 *lltable*, 查询换乘线路时,先由表 *lltable* 确定最小换乘次数。如果最小换乘次数为 3,则直接查询换乘 3 次的线路,不必先测试是否有换乘 1 次和 2 次的线路。当查询换乘 3 次的线路时,先由线线表,确定相互之间线距为 1 的四个换乘线集(含起始站和终到站的换乘线集) *hcsta*, *hcl*, *hc2*, *hcend*, 然后它们两两线路集确定换乘站点(同时在两条线路上的站点),得到 3 段线路,再进行 3 段线路的组合连接。在连接前,按站点在两条线路上的站序之和的大小删除一些线路,即进行一次优选。否则最后的换乘线路可能达到上千条甚至几万条。

系统在线路优化选取上主要引入站距(一条线路中所含站点总数)的概念,所有可能的线路按总的站距由小到大排序,进而,在两边线路相同、只是换乘站点不同的线路中,选择站距最小的线路,这是第二次优选,优化效果明显。

本文改进了文^[1]中求公交线路网络的换乘次数上确界的算法。采用了换线行表 *hxh* 代替^[1]中的线站表 *ltable*。求出了北京市公交网络换乘次数的上确界为 4,即任选一站点最多经 4 次换乘就可以到达整个网络的其它任一站点。系统主程序结构清晰、简洁、运行速度快,可以在数据库系统的限制内推广到换乘次数更高的情况(北京市公交网络换乘次数的上确界为 4,线线表 *lltable* 中最大值为 4,而且有换乘 4 次的例)。线路优选的参数可以改为运行里程或时间,具有实用性和可推广性。

对于北京公交网络的新数据^[10],我们给出了处理上行线和下行线的模型,依此对系统作了较大修改,运行正常,所选线路带有乘车费用。

2 数据表结构和主程序的算法

2.1 主要数据表及基本概念

系统的主要数据表有三个:站线表 *stable*、线线表 *lltable* 和换线行表 *hxh*。

表 *stable* 的行是站点集(即主要字段是站点 ID 和名称),列名称是线路 ID。表 *lltable* 的行和列都是线路

集,其列值是交叉的两个线路的距离(简称线距,下面给出定义)。这两个表的简图如下

stable	L1	L2	...	L412	ltable	L1	L2	...	L412
S1	3	2	...	0	L1	0	2	...	1
S2	0	1	...	5	L2	2	0	...	3
...
S2872	0	16	...	0	L412

在表 **stable** 中,第一行 S1 的各列值表示 S1 是线路 L1 的第 3 个站点、L2 的第 2 个站点、...非 L412 的站点,等等。如果线路 Li 有 16 个站点,则列 Li 上依次对应的各站点的值为 1、2、...、16,各个值叫该站点在该线路上的站序。同一线路上两站点在表 **stable** 中的相应列值之差的绝对值,称为两站点对应于该线路的站距。

对于站点 Si 而言,经过 Si 的线路叫 Si 的换线行(换乘线的别名,而换乘线一词代表这个线路及其上的站点)。每个站点的换线行集是确定的,长时间不变的。换线行表 **hxh** 的行是站点和它的换线行集。

表 **ltable** 中第一行 L1 的各列值表示 L1 与自身线距为 0,与 L2 线距为 2, ..., 与 L412 线距为 1,等等。线距的含义是:给定一条线路,如 Li,其上所有站点的换线行,其集合记为 A,都是与 Li 线距为 1 的;A 中的这些换乘线上所有站点的换线行(除去前面出现过的线路),其集合记为 B,称为与 Li 的线距为 2;集 B 中线路的所有站点的换线行(除去前面出现过的线路),称为与 Li 的线距为 3;依此类推。线距是两条线路之间关系的一种描述,它给出了两条线路的站点之间换乘次数的一个上界,因为若两条线路的线距为 3,意味着一条线路上的站点到另一条线路上的站点最多要换乘 3 次,有些站点(在两条换乘线上的站点)可能换乘两次就行。两条线路的线距是唯一确定的且是对称的。

起始站设为 **start**,终到站设为 **end**。

2.2 换乘线路查询的主程序算法

换乘线路查询的主程序算法:

(1) 根据所给起始站和终到站,从站点表 **bjstation** 中查询相应的站点 ID。

(2) 分别根据站点 ID,调用子程序 **prohxh**,求出起始站和终到站的换线行集 **hcsta** 和 **hccnd**。

(3) 判断 **hcsta** 和 **hccnd** 中是否有相同线路。

若有,即起始站和终到站有同一线路,调用子程序 **projoin** 将站点名称、线路名称和站距等信息放入计算表 **hcjs**,调用 **projs** 计算站距等,再调用 **projg** 显示整个线路,结束主程序;

若没有,继续。

(4) 调用 **promin**,确定起始站到终到站的最小换乘次数 **@hcmin**。

(5) 根据值 **@hcmin** 相应进行查询和处理,比如 **@hcmin = 3** 时,有

IF **@hcmin = 3** — 查询换乘 3 次的线路

BEGIN

exec **prohcl 2** — 确定相互线距为 1 的四个换线行集 **hcsta**, **hc1**, **hc2**, **hccnd**

exec **prohcl 'hcs1', 'hcjoin1'** — 按 3 段产生分段线路 **hcjoin1 ~ hcjoin3**

exec **prohcl 'hcs2', 'hcjoin2'** — 即在线距为 1 的集中加入站点并进行筛选

exec **prohcl 'hcs3', 'hcjoin3'**

exec **projoin 3** — 连接三段线路集

exec **projs 3** — 计算参数,筛选线路

exec **projg 3** — 显示结果

END

(6) 主程序结束。

由此可以归纳地得出换乘 N 次的规律。

下面是关于主程序算法的一个说明(以换乘 3 次为例)。

设已查出从 **start** 到 **end** 的最小换乘次数为 **@hcmin = 3**,且已求出它们的相互线距为 1 的四个换线行集 **hcsta**, **hc1**, **hc2**, **hccnd**。将线路分为 3 段,即 **hcsta** 与 **hc1**, **hc1** 与 **hc2**, **hc2** 与 **hccnd**。

对于第一段的两个线路集 **hcsta** 与 **hc1**,因为 **hc1** 中的元至少与 **hcsta** 中的一个元满足线距为 1,因此下面还要生成准确的线距为 1 的组合序对线路集 **hcs1(LID, LID1)**,除去线距不为 1 的组合序对。例如下面线距表:

hcsta\hc1	Li	Lj	Lk
L6	1	2	3
L85	2	1	2
L172	1	1	1

通过 prohcl 中程序选出序对 $(L_6, L_i), (L_{85}, L_j), \dots, (L_{172}, L_k)$ 共 5 个元 (5 个交叉位置为 1 的) 送入表 hcs1。另两段同样处理, 得到表 hcs2 和表 hcs3。

2.3 各子程序的功能说明

(a) prohxh —— 功能: 求出起始站和终到站的换线行集。

(b) promin —— 功能: 求出起始站的换线行集 hcsta 到终到站的换线行集 hccnd 的线距, 即最小换乘次数 @ hccmin。

(c) prohcl —— 功能: 求出 hcsta 到 hccnd 之间相互线距为 1 的各线路集。

(d) prohclzh —— 功能: 由表 @ tab1, @ tab2 求出准确的线距为 1 的组合序对线路集 @ tab3。它是 prohcl 的子程序。

(e) prohcs —— 功能: 根据表 @ tab1 的值, 查询出站点组成线路集 @ tab2。

(f) projoin —— 功能: 连接几段线路组合而成一个完整的换乘线路集。

(g) projs —— 功能: 计算线路的站距、级别等参数的值并挑选 bj 值相同的线路。

(h) projg —— 功能: 显示换乘线路集结果。

3 实现换乘线路查询的一些问题

3.1 准备工作

第一项工作是检查各站点的名称是否统一、标准。这是查询成功的关键, 否则就会查不到结果甚至是错误的线路。地名不标准主要有三种情况, 第一种是缩写问题。例如, 人大与人民大学, 北航与航空航天大学, 采用哪一种名称应一致。因为虽然大家都明白这些名称, 但电脑却会作为不同的地名处理; 第二种是音似字不同。如滨角园 (414 路) 与滨角园, 安和桥 (394 路) 与安河桥, 京源路口与京原路口等等; 最易出问题的是第三种的站名相同。如 976 路的乡政府应为十八里店乡政府, 350 路的乡政府应为楼梓庄乡政府; 952 支的北工大应改为北京工业大学, 965 路的北工大应改为北方工业大学, 等等。这些问题不是地图册的问

题, 而是电子信息化带来的问题。

第二项工作是建立主要数据表。首先是在 excel 文件的行中输入线路名和相应的站点名, 再利用数据库 VF 中的几个小程序和 VF 与 SQL Server 的数据转换功能, 自动建立主要数据表 stable、hxh 和 litable。维护的自动性是提高系统效率的重要因素。城市发展得很快, 公交线路的增加、改变十分频繁, 需要经常对系统进行维护。

存储过程中, 一要注意数据表的初始化问题, 二要运用动态 SQL 查询技术^{[7][8]}。

2.2 查询结果示例和分析

下面给出两个具体的查询结果:

表 2 从天安门到中关村的 76 条换乘 1 次的乘车路线

级别	站距	起始站	换乘线	换乘站 1	换乘线	终到站
4	9	天安门	特 1 路	木樨地	特 6 路	中关村
6	10	天安门	特 1 路	大北窑	801 路	中关村
6	10	天安门	802 路	木樨地	特 6 路	中关村
8	10	天安门	802 路	大北窑	801 路	中关村
7	11	天安门	4 路	大北窑	801 路	中关村
7	11	天安门	1 路	大北窑	801 路	中关村
8	11	天安门	728 路	公主坟	811 路	中关村
5	12	天安门	1 路	木樨地	特 6 路	中关村
5	12	天安门	4 路	木樨地	特 6 路	中关村
9	12	天安门	728 路	公主坟	944 支	中关村

省略其余 66 条, 这 76 条线路的站距最小是 9 最大是 41, 而且这 76 条线路是从所有 125 条线路中选出的。本次查询用时 2 秒 (清华同方微机, 奔腾 3 667MHz, 512MB 内存)。

如果将上面的起点站和终点站交换, 得到的线路集合是相同的。

表 3 从雍王府到松山道口的 2 条换乘 4 次的乘车路线

zi	起点	换线	换站 1	换线	换站 2	换线	换站	换线	换站 4	换线	终点
19	雍王府	311 路	金顶街	326 路	火车站	971 路	东关	919 路	延庆汽车站	920 路	松山道口
19	雍王府	311 路	金顶街	396 路	火车站	971 路	东关	919 路	延庆汽车站	920 路	松山道口

在同一微机上查询用时 8 秒(保留最小站距相同的线路)。

在测试的线路数据中选出有代表性的 14 组,其中换乘 1 次的有 2 组、换乘 2 次的有 6 组、3 次的 3 组、4 次的 3 组。主要数据如下:

换乘次数	1	2	3	4
平均用时(秒)	1.5	2.7	7	5.1
平均最小站距	10	22	23	28

查询总平均用时 4 秒,考虑到查询换乘 1 次、2 次的分别在 2 秒和 3 秒内,而且所用的是奔腾 III 的机器,可以说速度是快的,因为平均每秒要处理 257 条线路。例如,用时最长 10 秒的线路:四惠至亮甲店,换乘次数 2,所有线路有 932 条(站点组合所致),选出 204 条显示,显示线路的最小站距 14,最大站距 53。用时多于 6 秒的多数是处理的线路达 300 条以上。实际使用时可以限定显示的线路的数量。

3.3 关于换乘次数的讨论

首先修改了求公交网络换乘次数上确界的算法。对于文[1]中,给出站点集去求换乘线集时,修改为使用构造的换线行表 hxh,并将各站点的换线行的行数据转为一列数据。

下面的数据是根据线线表 litable 得到的:

表 4 北京市公交网络的换乘情况统计表

项目	换乘 0 次	换乘 1 次	换乘 2 次	换乘 3 次	换乘 4 次	合计
换乘次数(对)	206	18022	62735	3896	13	84872
平均换乘	1	87.5	304.5	18.9	0.1	412
平均换乘%	0.24	21.23	73.92	4.59	0.02	100

从中可以看出换乘 4 次的有 13 对线路(两条线路中至少有一对站点需换乘 4 次到达),换乘 2 次的最多,占总数的 74%,换乘 1 次和 2 次的合计占 95% 以上,其余合计约占 5%。

经检查,换乘 4 次的情况都是在 920 路(下营至延庆汽车站)出现的。920 路与其余 411 条线路换乘 1 次

的有 1 条,2 次的有 61 条,换乘 3 次的有 336 条,4 次的有 13 条。它的换乘 3 次的情况也是最多的,占有换乘 3 次的总数的 8.6%。

站点中换线行最多的是公主坟站,有 42 条线路经过它。有 14 条线路(如特 8 内环和特 8 外环)与其它线路最多换乘 2 次,即在这些线路上公共交通最方便。为简洁起见,有 7 条环路线作为非环路处理。

4 北京公交网络的新数据的一些问题

根据[10]中的北京公交网络的新数据,重新修改了系统。主要增加了票价费用参数,最后的线路给出了往返的参考乘车票价,特别研究了上行线和下行线的处理,给出了比较合理的解决方案。

上行线和下行线共有 818 条,占线路总数(929 条)的 88%。如果某站点在上行(下行)线出现而不在下行(上行线)线中,称这个站点为非对称站点。这样的站点共有 886 个,占站点总数(3957 个)的 22.4%。

非对称站点使得查询系统更加复杂。在实际情况中,如果非对称站点是人们出行的起始站或终到站,一般地,人们会知道应该在哪一个位置上下车。因此,本系统的起始站和终到站自然包含所有站点,但非对称站点不作为中间换乘站。这样,既能够使整个公交网络连通(任意两站点都有线路连通),又化简了查询数据量(只处理 522 条线路,含两条地铁线)。在此方案下,北京公交网络新数据的换乘次数上确界仍为 4。

关于站距的计算,构造了上行线表 shangdb 和下行线表 xiadb,分别计算站距,以使数据接近实际。

对[10]进行了 11700 多组数据的测试,查询全部成功。其中,换乘 1 次的有 2080 组,查询平均用时 0 秒;换乘 2 次的有 7484 组,查询平均用时 2 秒;换乘 3 次的有 2198 组,查询平均用时 7 秒(最长用时 1 分半);换乘 4 次的较少,测试数据中未遇到(下面的例是特别寻找的)。每次平均给出线路 3~8 条。

下面是公交网络新数据下站点 S3170 到站点

S3608 换乘 4 次的例。

zj	pi	s0	l1	s1	l2	s2	l3	s3	l4	s4	l5	s5
24	10	S3170	L111	S1955	L280	S0584	L189	S1920	L457	S3046	L208	S3608
33	10	S3170	L111	S3543	L286	S0353	L058	S3457	L100	S2313	L208	S3608
33	10	S3170	L111	S3543	L049	S0353	L058	S3457	L100	S2313	L208	S3608
33	10	S3170	L111	S3543	L507	S0241	L058	S3457	L100	S2313	L208	S3608
33	10	S3170	L111	S3543	L513	S3803	L058	S3457	L100	S2313	L208	S3608
35	14	S3170	L111	S1955	L059	S1956	T2	S0497	L283	S2781	L208	S3608
41	10	S3170	L111	S2321	L387	S3186	L189	S0771	L450	S3046	L208	S3608

其中 zj 表示平均单程站点数, pi 为往返车费和, 其余为换乘线和换乘站。查询运行时间 20 秒。

参考文献

- 1 王世祥, 饶维亚. 大中城市公交线路查询的数据结构及其算法的实现[J]. 计算机系统应用, 2007, 16(9).
- 2 赵巧霞, 马志强, 张发. 以最小换乘次数和站数为目标的公交出行算法[J]. 计算机应用, 2002, 24(12): 136 - 146
- 3 马良河, 刘信斌, 廖大庆. 城市公交线路网络图的最短路与乘车路线问题[J]. 数学的实践与认识, 2004, 34(6): 38 - 44.
- 4 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.
- 5 翁敏, 毋河海, 杜清运, 蔡忠亮. 基于公交网络模型的最优出行路径选择的研究. 武汉大学学报, 2004, 29(6).
- 6 向万里, 刘洪升. 城市公交网络出行路径选择的计算机算法研究[J]. 兰州交通大学学报, 2006, (1).
- 7 Nielsen P. Microsoft SQL Server 2000 宝典[M]. 刘瑞, 陈微, 闫继忠, 刘文, 等译. 北京: 中国铁道出版社, 2004.
- 8 邹建. 中文版 SQL Server 2000 开发与管理应用实例[M]. 北京: 人民邮电出版社, 2005.
- 9 中国地图出版社地图开发室编制. 北京市实用地图册(第二版)[M]. 中国地图出版社, 2005.
- 10 2007 高教社杯全国大学生数学建模 B 题数据文件 B2007data.rar