



李英女

铁路客运信息查询算法

李英女 郑国雄

摘要 铁路客运信息查询算法是建立在数据库基础上的。为用户提供全国范围内任意两个车站之间合理的乘车方案,该算法需要建立的数据库包括:车次表、车站表、时刻表和相关局表,通过这些基本数据库,使用中转算法可以产生中转站表,从而构成铁路客运信息查询算法的基础。当用户任意给定两个车站时,查询算法能在短时间内给出合理的乘车路线、乘车车次、乘车时间以及中转位置等信息,用户可以利用这些信息指导自己的铁路旅行。

关键词 客运 数据库 查询 算法

Inquiry Arithmetic for Railway Passenger Transport Information

Li Yingnu Zheng Guoxiong

(Beijing Institute of Tracking and telecommunication Technology, 100094)

Abstract: The inquiry arithmetic for railway passenger transport information is based on the database, and its function is to serve the user an available scheme for them to transfer between each two railway stations all over the country. The databases for this method include: table of train numbers, table of railway stations, table of schedule and table of correlative railway bureau. Using these original tables, we can get transfer tables with Transfer Arithmetic, which are the foundation of the inquiry arithmetic for railway passenger transport information. When given any two railway stations, the inquiry arithmetic can offer a reasonable route, train numbers, starting times and transfer stations ect. . With these information, user can guide one's railway travel life effectively.

Keywords: railway passenger transport, database, inquiry, arithmetic

1 引言

铁路运输具有运能大、速度快、能耗小、安全可靠、

对环境污染小、运费低,选择铁路旅行是目前我国绝大多数人和许多国外来华旅游团主要的中长途旅行方式。

李英女 北京跟踪与通信技术研究所 工程师 100094 北京市
郑国雄 北京航天指挥控制中心 在职博士研究生 100094 北京市

我国正在运营的铁路客运系统中,有8种列车,分别是:准高速列车、快速列车、旅游列车、特别快车、普

则挖掘方面我们还有许多问题需要进一步研究,如怎样提高算法的效率,怎样对挖掘出的规则进行进一步的剪裁和总结等等。

5 参考文献

- 1 R. Agrawal, T. Imielinski, and A. Swami, Mining association rules between sets of items in large databases. In Proc. 1993 ACM—SIGMOD Int. Conf. Management of Data,

pp. 207~216, Washington, D. C., May 1993.

- 2 N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal. Efficient Mining Of Association Rules Using Closed Itemset Lattices. Information Systems Vol. 24, No. 1, pp. 25~46, 1999
- 3 王珊. 数据仓库技术与联机分析处理. 北京: 科学出版社, 1998

(责任编辑:赵存义 收稿日期:2000-01-12)

通快车、普通旅客快车、管内旅客列车和广深旅客快车;有列车车次约 2242 趟,其中准高速列车 58 趟、快速列车 112 趟、旅游列车 126 趟、特别快车 288 趟、普通快车 379 趟、普通旅客快车 411 趟、管内旅客列车 834 趟和广深旅客快车 34 趟,这些统计还不包括为临时需要而增加的车种和车次。列车所停靠的车站约 4628 个,其中有约 230 个车站处在两条铁路线的交叉点上;有约 274 个车站有与自己相关的同一城市站;有约 138 个车站有始发列车,有约 1316 个车站有快车列车车次(后面定义)停靠,其中约 719 个车站所停靠的快车列车车次大于 5 个。所有车站在遍布全国各地 14 个铁路局的 55 个铁路分局中。如果按列车的停靠时刻点来统计,有约 34476 个停靠时刻点。从这些基本的数据可以看到,我国的铁路客运系统非常复杂。

如此复杂的客运系统,为了选择某两站之间的客运列车运行线,查找到的某两个站之间的合理乘车方案非常不容易。

铁路客运信息查询算法就是在这种需求下提出的,为旅客提供全国范围内任意两个车站之间合理的乘车方案。算法中建立的数据库包括:车次表、车站表、时刻表和相关局表,通过这些基本数据库,使用中转算法可以产生中转站表,从而构成铁路客运信息查询算法的基础。当用户任意给定两个车站时,使用查询算法能在短时间内给出合理的乘车路线、乘车车次、乘车时间以及中转位置等信息,用户可以利用这些信息有效指导自己的铁路旅行生活。

2 数据库的表结构

下面定义一些术语以备数据表描述时使用:

枢纽站:从一条铁路线转到另一条铁路线客运列车必须经过的车站叫枢纽站;枢纽站往往在两条或多条铁路线的交叉点上。枢纽站常常会成为中转站。

相关站:在同一城市的客运列车停靠车站之间叫相关站;相关站在查询时非常有用,当用户给定一个车站时,程序能自动检查其相关站,在更大的范围内选择更合理的结果。

快车、慢车:客运列车有准高速列车、快速列车、旅游列车、特别快车、普通快车、普通旅客快车、管内旅客列车和广深旅客快车 8 种,定义普通旅客快车和管内旅客列车为慢车,其它的车种为快车;列车有快慢之分,这样在选择中转时可以尽量使用快车,从而有更优的结果。

大站、小站:按客运车站所经由的快车数目来确定车站的大小属性,称所由的快车趟数大于 5 的站为大站,其它的即为小站。

设计铁路客运信息查询算法首先需要对客运相关的数据进行描述,客运相关的数据包括描述个体的数据和描述关联的数据两类,下面分别对使用数据库表的形式给予描述。

铁路局(分局)、客运车站、客运列车和客运列车时刻表是铁路客运信息查询算法的个体数据,在查询系统中将它们分别设计为分局表、车站表、车次表和时刻表;分局与分局之间、同一城市的客运车站之间有特殊的关联,在查询系统中将它们分别设计为分局相关表和车站相关表。它们的表结构是:

(1)分局表是由:分局编号,分局名称两个字段组成。

分局表用于存放铁路分局信息,每个记录描述一个铁路分局。分局编号对全国所有的铁路分局进行顺序编码。

(2)车站表是由:车站编号,车站名称,分局编号,车站级别,枢纽属性,相关属性,起点属性等 7 个字段组成。

车站表用于存放客运车站信息,每个记录描述一个客运车站。车站编号对全国所有的客运车站进行顺序编号,这个编号是为了解决车站的重名问题,使得算法查询运算时有唯一性,能得到正确的查询结果,车站编号在查询过程中使用。车站名称的内容是字符串形式的站名。车站名称在查询结果输出时使用。分局编号描述该车站归属的铁路分局。车站级别是指经由该车站的所有快车车次的个数,车站级别的值需要通过查询时刻表,计算每个车站经过非 M、L 列车车次的个数得到。车站级别的值越大表示该站所处的地位越高。枢纽属性描述该客运车站是否为枢纽站,枢纽属性为 1 的站称为枢纽站,枢纽站常常成为中转站。相关属性指定该客运车站在同一城市内等价车站在车站相关表中的记录位置。起点属性描述该客运车站所有起点车数目。

(3)车次表是由:车次编号,车次名称,车种标等 3 个字段组成。

车次表用于存放客运列车信息,描述每个记录描述一趟列车车次。车次编号对全国所有的客运列车进行顺序编号,这个编号与车次名称的列车编号是不一样的,车次名称的列车编号在全国范围内不唯一,设计

车次编号后,对每个车次使用一个唯一的编号,使得在系统查询运算时有唯一性,能得到正确的查询结果。对于往返客运列车使用两个车次编号描述,对于变号客运列车使用一个车次编号描述。车次名称是指铁路系统为某个客运列车所指定的代号,如45次、123/135次等。车种标识有Z—准高速列车、K—快速列车、Y—旅游列车、T—特别快车、P—普通快车、L—普通旅客快车、M—管内旅客列车和S—广深旅客快车八种,这些内容可以用于简单确定列车车种,车种不同其价格和速度等均不一样,有的甚至差别很大。

(4)时刻表是由:车次编号,车站序号,车站编号,到达时间,到达时间1,离开时间,离开时间1等7个字段组成。

时刻表用于描述客运列车停靠点信息,每个记录描述了一个客运列车停靠时刻点。车站序号描述列车车站在列车车次停靠的顺序位置,如某车次的起点站的车站序号为1,第一个停靠车站的车站序号为2,依此类推。到达时间和到达时间1描述列车车次到达车站的当天时间(HH:MM格式)和天数,当列车时刻跨一天到达时间1为1,依此类推。离开时间和离开时间1描述列车车次离开车站的当天时间(HH:MM格式)和天数,当列车时刻跨一天离开时间1为1,依此类推。

(5)分局相关表是由:分局编号1,分局编号2,相关个数,相关分局编号1,相关分局编号2,...相关分局编号20等多个字段组成。

分局相关表用于存放从一个分局乘车到另一个铁路分局可能经过的铁路分局列表,每个记录描述一个从分局编号1到分局编号2的相关铁路分局信息。设定最多相关分局数为20个,事实上可以根据实际的相关分局数做调整,当两个分局之间的相关分局超过20个时,可以人为减少其数量。

(6)车站相关表是由:车站编号,相关个数,车站编号1,...车站编号20等多个字段组成。

车站相关表用于存放同一城市内等价车站的信息,每个记录描述一个车站。设定每个车站的最多相关站数为20个(事实上不可能超过20个)。

3 查询算法

任意两个客运车站之间的乘车方案分有直达车和无直达车两种情况,当两个车站有直达车时,可以通过查询时刻表的方法,方便快速地得到乘车方案。而当两个车站无直达车时,需要依托第三个站或第 $n(n > 3)$

个站中转,将无直达车的解变为起点站到中转站、中转站到中转站(可能多个)、中转站到终点站的多个有直达车乘车方案的集合。因此如何在短时间内找到无直达车情况下合理的中转站是查询算法的重点,因为中转站确定了,问题的解也就得到了。

在遇到无直达车乘车方案的查询问题时,一般地,可以从起点站开始,将经过该站的所有客运列车所停靠的客运车站定为中转站,称第一层中转站,查该中转站是否有到达终点站的直达车,如果有,得到中转一次的乘车方案,如果没有,以某个第一层中转站为起点站,进行递归运算,直到查到第 n 层中转站到终点站有直达车,得到中转 n 次的乘车方案,将无直达车乘车方案的解分离为多段有直达车乘车方案的解。

这个最直观的查询算法是可以实现的,但它所搜索的中转站是遍历的,运算量大,没有实用价值。

为了加快搜索速度,这里提出一个使用中转站表过渡的方法来解决,即事先依据时刻表的信息,产生一个中转站表,形成每个客运车站可能的中转站列表,把大量查询时必须计算的工作放在系统准备时进行,这样在实际查询时的运算可以仅通过比中对转站列表来得到合理的中转站,减少查询时的计算量;另外,为了使得查询更有方向 and 选择性,提出一个使用分局相关表的方法来解决,即在选择中转站时必须是有条件的,分局相关表定义了一个从铁路分局到另一个铁路分局所必须经过的铁路分局,这个信息可以从铁路线布局中得到,有了这个信息,所选择的中转站必须落在相关分局内,而其他的中转站可以筛选去掉,这样不仅确保了其方向性,而且缩小选择范围,减少计算量。

中转站表是由:车站编号,中转站数,中转车站编号1,中转车站编号2,...中转车站编号120等多个字段组成。

中转站表用于存放从某个客运车站出发,可能成为其中转车站的客运车站信息,每个记录描述一个客运车站的中转站列表。中转站表是为建立查询算法而根据基本数据产生的中间结果数据。

查询算法包括中转站产生算法和查询乘车方案算法。

中转站产生算法:

(1)逐个提取车站表中的车站编号,查时刻表经过该车站编号的所有快车车次编号,得到各个车次编号;

(2)逐个提取得到的各个车次编号,查时刻表该车次编号所经过的所有车站编号,得到各个车站编号;

(3)在得到的各个车站编号中选择起点属性大于0或枢纽属性等于1的客运车站为该站的中转站,增加中转站数和中转车站编号 n ;

(4)对于某些小站,经过上述运算后仍没有一个中转站,逐个提取中转站表的中转站数为零的车站编号,查时刻表经过该车站编号的所有慢车车次编号,得到各个车次编号;

(5)逐个提取得到的各个车次编号,查时刻表该车次编号所经过的所有车站编号,得到各个车站编号;

(6)在得到的各个车站编号中选择与指定小站相临的两端大站为该站的中转站,修改中转站表的中转站数为2;

查询乘车方案算法:

查询乘车方案算法分主处理程序、直达方案查询和中转方案查询3部分。

主处理程序:

用户可以输入任意两个客运车站做为起点站和终点站,程序根据用户输入的两个站进行检索,给用户可以提供可以选择的乘车方案,乘车方案数目最多4个,最大中转数为3次。

(1)用户输入两个客运车站的站名,进行输入确认处理,判断其合理性,查车站表将其转换为两个车站编号;

(2)使用两个车站编号,进行直达乘车方案处理,判断是否有直达乘车方案;

(3)如果有直达乘车方案,进行第七步骤;如果无直达乘车方案继续;

(4)进行中转乘车方案处理,判断是否有中转乘车方案;

(5)如果有中转乘车方案,存储中转乘车方案结果,增加乘车方案数目。如果没有中转乘车方案继续;

(6)判断乘车方案数目够了吗?如果方案数目够了,进行第七步骤;否则中转次数加1;判断中转次数大于最大中转数了吗?如果是小于等于,将符合条件的中转站集设为起点站,继续第四步骤;如果大于继续;

(7)显示存储的乘车方案结果,如果乘车方案数目为零,显示无解;

直达方案查询:

直达方案查询针对某两个客运车站的车站编号,并不一定是用户给定的起点站和终点站,还可能是某个中转站和终点站,因此使用站1和站2表示:

(1)查时刻表中车站编号为站1的车次编号,记为

车次编号组1,表示站1的所有经过车次;

(2)查时刻表中车站编号为站2的车次编号,记为车次编号组2,表示站2的所有经过车次;

(3)在车次编号组1和车次编号组2中查相同的车次编号,而且,要求相应的站1在该车次的车站序号小于站2在该车次的车站序号,记为车次编号组3;

(4)判断车次编号组3是否有信息,如果有信息,表示有直达乘车方案,记录方案数目,生成输出结果,设置是否有结果标志。

中转方案查询:

中转方案查询与直达方案查询类似,针对的某两个客运车站的车站编号是任意的,在第一次调用时是用户输入的起点站车站编号和终点站车站编号。

(1)在分局有关表中查站1和站2的相关铁路分局记录,记为分局值;

(2)在中转站表中查站1的中转站记录,记为中转站值1;

(3)在中转站表中查站2的中转站记录,记为中转站值2;

(4)比对中转站值1和中转站值2,提取车站编号相应的分局编号在分局值内,而且车站编号一致的值,记为中转站集;

(5)判断是否有一致的中转车站编号,即中转站集的内容,如果有,该车站编号为站1和站2之间的中转站,获取结果和设置结果标识。

4 结束语

应用 Visual Basic 语言和 Access 数据库,我们已经在微机上实现了给定的算法,把产生中转表的算法称为维护算法,把查询乘车方案算法称为查询算法,在主板为 P II 266,内存为 32MB 的计算机上,查询任意两个客运车站之间乘车方案的最大延时为 3 秒。这么小的时间延时值,对应用推广很有价值。该算法可以应用于铁路客运信息查询系统。

5 参考文献

- 1 铁道部运输局. 全国 98' 铁路旅客列车时刻表. 中国铁道出版社, 1998
- 2 Microsoft corp. MS Visual Basic. Microsoft corp., 1997
- 3 Microsoft corp. MS Access 97. Microsoft corp., 1997

(责任编辑:赵存义 收稿日期:2000-02-08)