

公交线路查询算法^①

王海帅, 冀振燕, 王 森

(北京交通大学 软件学院, 北京 100044)

摘 要: 公共交通不仅是衡量城市现代化程度的重要标志也是解决交通拥堵问题的途径. 而公交线路查询系统的关键技术是公交线路查询算法, 它对提高公交资源的利用率有着重要的意义. 总结了国内外城市公交最优路径算法并在此基础上分析了高效运行城市公交系统的条件和影响因素. 介绍了最短路径问题及 Dijkstra 算法及其在查询系统应用中的弊端. 然后提出了基于换乘最小的广度优先算法的数学模型, 给出了算法的实现, 并以银川市公共交通公司的公交部分数据为基础, 完成了公交信息查询系统的设计与开发.

关键词: 最优路径; 最小换乘; 最短路径; 查询算法; 公交信息查询

Bus Transport Transfer Algorithm

WANG Hai-Shuai, JI Zhen-Yan, WANG Sen

(School of Software, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Public transportation is not only a significant symbol of the urban modernization, but also an optimal approach of solving urban crowded traffic problem. The bus line search algorithm is the key technical query system. This paper summarizes the algorithm of public transportation optimum route choice of foreign and urban cities, and analyses the realization conditions and influence factors for effective running. The paper introduces the shortest path problem and Dijkstra algorithm and analyzes the drawbacks of its application in bus line query. Then the optimal travel route choice algorithm of mathematical model based on the smallest transfer is proposed, and the implementation of the algorithm is given. Based on the data of the public transportation company Yinchuan, a bus information query system design and development is completed.

Key words: the optimum path; minimal transfer; shortest path; bus information query

1 引言

随着社会经济的迅速发展, 公共交通在现代化人们生活中起着越来越重要的作用. 城市公共交通的基本任务是通过为乘客提供安全、方便、迅速、准点、舒适的乘车条件, 最大限度地发挥现有公共交通设施的营运能力, 节约全社会的总乘车出行时间^[1]. 但是由于城市公交线路的不断增多以及新公交线路的延伸和开辟, 选择合适的乘车线路存在一定的困难, 因此, 建立符合公交线路查询特点的公交数据以及公交查询系统, 是城市公交信息化和数字化的迫切需要.

公交查询系统中的核心是查询算法的设计与实

现. 经过研究现有公交最优路径算法, 发现很多没有考虑影响乘客出行选择的实际因素, 缺乏实用性. 如一些查询算法在研究最短出行距离或最短时间时, 要经过多次换乘才能到达目的地, 但这对乘客的乘车时间和费用等方面有很多不确定的因素, 多次换乘没有实际意义. 所以通过对公交网络模型的最优路径算法的研究, 本文提出一种基于换乘次数最少为第一目标, 最短距离或最少费用为第二目标的综合评价指数的实用性强的最优出行路径选择算法, 最大程度满足乘客的乘车要求, 提高城市公共交通的利用率.

本文研究的具体目标为:

^① 收稿时间:2012-07-04;收到修改稿时间:2012-08-27

1) 对经典的最短路径算法进行分析并指出了在现实的公交查询系统应用中的弊端. 在对公交线路最优出行路径选择的研究和实践基础上分析城市公共交通线路选择的影响因素和评价指标及最优出行路径选择的实现条件;

2) 根据公民出行的实际需求, 对最优评价指标进行分析, 设计, 并给出计算方法;

3) 综合考虑上述三个评价指标, 为线路查询算法建立数学模型;

4) 针对上述建立的模型, 实现了基于换乘最小的实用性强的最优出行线路查询算法的设计和实现.

2 最短路径问题及算法调研

良河^[2]等根据公共网络图提出多目的地乘车线路问题, 并通过数学模型设计出了最佳乘车路线. 针对目前城市公交查询系统存在的问题及乘客的需求, 梁虹^[3]等人通过对经典 A* 算法的改进而提出了一种新的针对公交网络的最短路径算法, 在此算法的基础上通过对昆明市公交系统的数据建模实现了相应的公交查询系统, 该公交系统有较好查询速度. 针对道路网络不同于其它网络, Liu C L^[4]采用矩阵记录约束并结合 Q 矩阵, 在约束条件下使用 A* 算法来对以往设计进行改进, 对求得的最短路径有较理想的效果. Peng Z R^[5]等人通过对整体服务水平以及同一条线路上的时间表的综合考虑来确定最短路径和换乘点, 研究并提出了一种适合道路网络特殊性的最短路径算法.

由于城市公交网络的快速发展, 城市公交查询算法的研究越来越受到国内外学者的重视. 冯林^[6]等人采用快速搜索策略而提出了基于层次空间推理的最优乘车算法, 并获得了较满意的效果. Salzborn^[7]等人通过对若干直路与一条主干道路的调查提出公交时刻表可以与主干道路与直流换乘协调制定. 针对多种公交方式换乘过程中计算换乘方式数目及换乘中的非线性费用两个问题, Lo H K^[8]通过建立模型解决了这种问题.

3 算法数学模型的建立

3.1 假设和符号说明

假设:

1) 任意公交车在相邻站点间的行驶时间为 2min, 乘客在一个换乘站点的平均等待时间为 10min;

2) 选择的评价指标为换乘次数、时间和费用, 换乘

次数为首要考虑要素, 在换乘次数一样的条件下, 综合考虑后两个评价指标来选择合适的线路;

3) 根据实际情况, 如果公众使用过程中不能找到换乘车次少于两次的最优公交线路, 建议改为其他交通方式出行;

符号说明:

\bar{t} 为换乘平均耗时, n_i 为第 i 条线路上换乘总次数 ($n_i \leq 2$), l_i 为公交站点计数, s_i 为第 i 条路线上包括站点 A 与 B 在内的总公交站点数, t_i 为第 i 条线路相邻公交站点间平均行驶时间. M_i 为从站点 A 到 B 的所有路线中第 i 条总费用值. T_i 为从站点 A 到 B 的所有路线中第 i 条的总乘车时间值. $n_i - (m_i - 1)$ 为 c_k 为从站点 A 到站点 B 第 i 条线路上分段计价线路的换乘次数, 第 k 条单一票价制线路的票价取值为 1 元, $r(a_{ij})$ 为第 j 条分段计价线路的票价,

3.2 数学模型描述

出行者根据实际情况选择起点 A 与目的地 B. 默认情况下, 系统自动选择性价比最高的线路方案, 但出行者可以更改时效性、经济性的要求等级, 从而确定一个 α 值.

第一步, 系统首先要确定有无一条公交线路从站点 A 直达站点 B. 如果有一条, 那么这条线路即是解. 如果查出有多条线路, 那么比较这几条线路的综合评价指数 R_0 , 综合评价指数最小的线路是最优解. 如果没有直达线路, 执行下一步.

第二步, 系统要试着找出一个过渡站点 C, 即从站点 A 到站点 C, 再到站点 B. 查找经过 A、B 这两个站点的所有公交线路所经过的共同站点 C, 如果存在, 将站点 A 到站点 C 的所有直达公交线路存入集合 X, 再将站点 C 到站点 B 的所有直达公交线路存入集合 Y. 集合 X 与集合 Y 的各种组合即是站点 A 经过站点 C 到达站点 B 的线路, 同第一步的操作一样, 比较这些线路的综合评价指数, 综合评价指数最小的线路 R_1 即是最优解. 如果不存在这样的过渡站点, 执行第三步.

第三步, 系统试着找出两个过渡站点 D 和 E, 即从站点 A 到站点 D, 再到站点 E, 最后到达站点 B. 首先, 系统查找数据库, 计算经过站点 A 的所有公交线路的所有站点并将结果存入集合 O. 继续查找数据库, 计算经过站点 B 的所有公交线路的所有站点并将结果存入集合 P. 然后, 同理计算经过集合 O 每个站点的所

有公交线路并存入集合 H, 计算经过集合 P 每个站点的所有公交线路并存入集合 I. 比较集合 H 和 I, 若这两个集合有相同公交线路 S, 则公交线路 S 分别对应于集合 O、集合 P 中的 D、E 为换乘站点, 即乘客从站点 A 经过站点 D 和 E, 到达站点 B. 该方案会产生多种组合, 与上几步同理, 通过比较综合评价指数得到最优公交线路 R2.

假设只考虑不超过两次的换乘情况, 如果经过以上三步处理, 还是无法到达目的地, 则说明换乘次数过多. 系统会建议出行者改乘其他交通工具, 例如出租车等.

通过以上分析, 选择 A 到 B 所有路线中第 i 条路线的无量纲化综合评价指数^[9].

$$\theta_i = f(T_i, M_i) = F(T_i', M_i') = \alpha T_i' + (1-\alpha)M_i', 0 \leq \alpha \leq 1 \quad (1)$$

经无量纲化处理后 T_i 的值为

$$T_i' = \frac{s_i}{\sum_{l_i=2}^{s_i} t_{li} + n_i \bar{t}} \quad (2)$$

经无量纲化处理后 M_i 的值 M_i' 为:

$$M_i' = \frac{m_i}{\sum_{k=1}^{m_i} c_k} + \frac{n_i - (m_i - 1)}{\sum_{j=1}^{n_i} r(a_{ij})} \quad (3)$$

站点数 a_{ij} 的分段函数定义为:

$$r(a_{ij}) = \begin{cases} 1, & 0 \leq a_{ij} < 9 \\ 2, & 9 \leq a_{ij} < 18 \\ 4, & a_{ij} \geq 18 \end{cases} \quad (4)$$

由上述分析得到任意两公交站点所有路线中综合评价指数的最小值(5):

$$\min\{\theta_i\} = \min\left\{ \alpha \left[\frac{s_i}{\sum_{l_i=2}^{s_i} t_{li} + n_i \bar{t}} \right] + (1-\alpha) \left[\frac{m_i}{\sum_{k=1}^{m_i} c_k} + \frac{n_i - (m_i - 1)}{\sum_{j=1}^{n_i} r(a_{ij})} \right] \right\}, 0 \leq \alpha \leq 1 \quad (5)$$

4 最优出行线路选择算法设计

根据上述模型, 具体算法设计流程图如图 1.

5 最优出行线路选择算法实现和应用研究

5.1 实验数据

本文所使用的实验数据由宁夏银川市公交公司提供, 提供的部分银川市公交线路和站点信息和自己在现实的基础上虚拟的线路. 其中公交线路更新时间截止

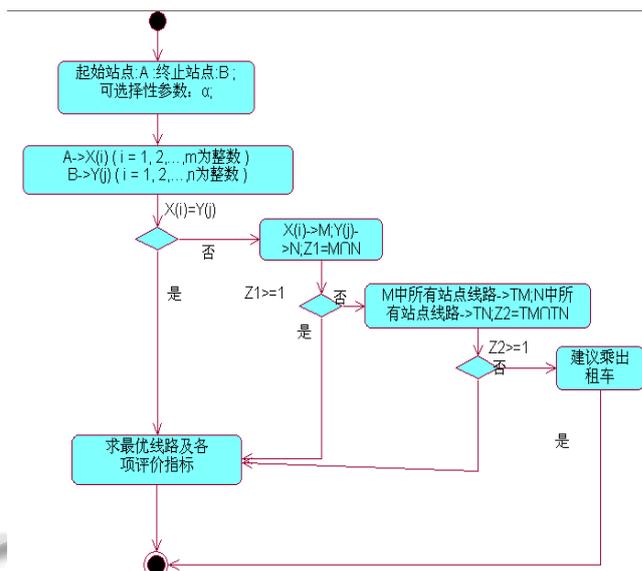


图 1 算法流程图

2011-07-30.

5.2 算法的应用研究

为了充分验证算法的性能, 采用 VB 2005 编程实现. 实验抽取互不相同的公, 分别作为出发地和目的地, 进行线路查询. 0 次换乘, 设计三种不同的偏好, 实验结果如下:



图 2 0 次换乘时效性选择

对以上的不同选择进行对比(如表 1)可以看出系统能够很好的根据乘客的需求给出最优的出行线路.

从实验所得的结果以及根据换乘次数、耗时和车费等指标, 效果良好, 切实可行. 此查询算法能够满足乘客不同的要求, 得到最优的乘车线路.



图 3 0 次换乘经济性选择



图 4 0 次换乘综合性选择

表 1 0 次换乘数据分析

起始始->终点站	换乘次数	性能要求	综合评价指数	费用/元	耗时/min	最优线路/路
五医院->市医院	0	时效性	0.4	2	20	102
五医院->市医院	0	经济性	0.25	1	24	23
五医院->市医院	0	综合性	0.365	1	24	23

1 次换乘, 设计三种不同的偏好, 实验结果如下:

表 2 1 次换乘数据分析

起始始->终点站	换乘次数	性能要求	综合评价指数	费用/元	耗时/min	最优线路/路
附属医院->双城门	1	时效性	0.64	3	32	附属医院 302->南门->双城门 3
附属医院->双城门	1	经济性	0.5	2	34	附属医院 12->西京女子医院->双城门 3
附属医院->双城门	1	综合性	0.59	2	34	附属医院 12->西京女子医院->双城门 3

2 次换乘, 设计三种不同的偏好, 实验结果如下:

表 3 2 次换乘数据分析

起始始->终点站	换乘次数	性能要求	综合评价指数	费用/元	耗时/min	最优线路/路
西门南->西京女子医院	2	时效性	1.12	4	56	西门南 102->电机厂->302->南门->3 西京女子医院
西门南->西京女子医院	2	经济性	0.75	3	78	西门南 102->市医院->23->林业学校->12 西京女子医院
西门南->西京女子医院	2	综合性	1.06	4	56	西门南 102->电机厂->302->南门->3 西京女子医院

6 总结

本文以银川城市公交有限公司提供的部分交通数据为基础, 编程实现了公交信息查询系统, 基本解决了乘客在公交出行中选择线路困难的问题. 不仅可以

给市民和游客的乘车提供向导, 同时实现了城市公交的数字化和信息化.

(下转第 137 页)

表 3 决策树正确识别率统计

样本类别	样本数	错误识别数	正确率 (%)	平均识别率
非畅销	440	36	91.8%	84.5%
畅销	410	96	76.6%	

表 4 规则正确率统计

规则	分类结果	样本数	错误识别数	正确率
Rule 1	非畅销	32	3	90.6%
Rule 2	畅销	141	17	87.9%
Rule 3	畅销	46	9	80.4%
Rule 4	非畅销	370	76	79.5%
Rule 5	畅销	163	10	93.9%
Rule 6	非畅销	98	17	82.7%

3 结语

本文运用 C4.5 算法,对团购网站与销量相关的各项指标进行了研究,初步得到了影响畅销与否的一些因素.预测结果表明:在该网站历史数据中,在第二季度服饰箱包鞋类目下,历史收藏数是预测团购销量水平的较好指标,同时,销量又受到价格的影响较大,卖家等级、是否选择更快捷的物流服务以及商品历史好评率也会在一定程度上反映畅销与否.使用同样的方法,可以针对不同商品类目得到相应的规则,下一步可以加入时间的因素,探求不同季节下分类规则的变化.这需要决策支持者根据自身的业务需求进行模型的调整.决策树算法的应用使团购网站选择商品时

多了决策的依据,同时生成的规则也较容易理解和应用.希望网站的决策层、业务人员在开发新商家的过程中,不仅仅运用以往的经验,更能使用一些数据分析来辅助自己进行选品,甚至可以进行反向招商,主动寻找适合在团购平台进行营销的商家.

参考文献

- 1 游超,姚振晔.网络团购模式分析与发展趋势预测.商品与质量,2011,(4):63.
- 2 Taherkhani A. Using Decision Tree Classifiers in Source Code Analysis to Recognize Algorithms: An Experiment with Sorting Algorithms.2011,54(11).
- 3 Quinlan JR. C4.5:Programs for machine learning. Morgan Kaufman. 1993: 81-106.
- 4 李强.创建决策树算法的比较研究——ID3, C4. 5, C5.0,算法的比较.甘肃科学学报,2006,18(4):84-87.
- 5 薛薇,陈欢歌.基于 Clementine 的数据挖掘.北京:中国人民大学出版社,2012.213-216.
- 6 李楠,段隆振,陈萌.决策树 C4.5 算法在数据挖掘中的分析及其应用.计算机与现代化,2008,(12):160-163.
- 7 魏晓云.决策树分类方法研究.计算机系统应用,2007,16(9):42-45.
- 8 段富,曾祥东,牛保宁.决策树方法在煤炭物流客户分析中的应用.计算机工程与应用,2010,46(10):245-248.
- 9 王晓国,黄韶坤,朱炜,李启炎.应用 C4.5 算法构造客户分类决策树的方法.计算机工程,2003,29(14):89-91.
- 10 桂现才,彭宏,王小华.C4.5 算法在保险客户流失分析中的应用.计算机工程与应用,2005,(17):197-199.

(上接第 91 页)

参考文献

- 1 匡星.城市常规公共交通服务水平研究[硕士学位论文].长春:吉林大学交通运输学院,2004.
- 2 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.数学的实践与认识,2004,34(6):38-44.
- 3 梁虹,袁小群,刘蕊.一种新的公交数据模型与公交查询系统实现.计算机工程与应用,2007,43(3):234-238.
- 4 Liu CL. Best-path planning for public transportation systems. Proc. of the 5th International IEEE Conference on Intelligent Transportation Systems. Singapore, 2002: 834-839.
- 5 Peng ZR, Huang RH. Design and development of interactive trip planning for web-based transit information systems. Transportation Research part C, 2000,(8):409-425.
- 6 冯林,孙宇哲.基于层次空间推理的公交最优乘车方案.计算机工程,2005,31(21):55-56.
- 7 Salzbom FJM. Scheduling bus systems with interchanges. Transportation Science, 1980,(14):211-231.
- 8 Lo HK, Yip CW, Wan KH. Modeling transfer and non-Linear fare structure in multi-modal network. Transportation Research Part B,2003,(37):149-170.
- 9 王惠文.偏最小二乘回归方法及其应用.北京:国防工业出版社,1999.