

基于遗传算法的超市最短导购路径推荐

韩建妙, 刘业政

HAN Jianmiao, LIU Yezheng

1.合肥工业大学 管理学院,合肥 230009

2.合肥工业大学 过程优化与智能决策教育部重点实验室,合肥 230009

1.School of Management, Hefei University of Technology, Hefei 230009, China

2.Key Laboratory of Process Optimization and Intelligent Decision-making, Hefei University of Technology, Hefei 230009, China

HAN Jianmiao, LIU Yezheng. Genetic algorithm-based shortest shopping guide route recommendation in supermarket. Computer Engineering and Applications, 2016, 52(4): 238-242.

Abstract: The large amount of products and complexity of spatial layout in supermarkets make customers easy to get lost in the room, spending much time searching goods. This paper proposes a genetic algorithm-based shortest shopping guide route recommendation method in supermarkets. The spatial layout structure of supermarket is discretized into a graph, in which nodes represent product zones, and undirected edges represent the walk route between two adjacent zones. The goods on the shopping list which one customer wants to buy are mapped into specific zones based on the shelf-locations they belonged to. The shortest path combining the entrance, the goods zones and the check-out counter is optimized by genetic algorithm. The simulation results in Matlab environment show that the method is simple and effective, and can quickly recommend a shortest path to customers for their shopping decisions.

Key words: supermarket shopping guide; environment modeling; shortest route; genetic algorithm; personalized recommendation

摘 要:大型超市里商品数目的繁多、空间布局的复杂往往容易让消费者迷失在室内,花费大量时间进行商品搜寻。提出基于遗传算法的超市导购路径推荐方法,通过对超市的空间布局结构离散化建模,生成分别用节点和无向边表示商品区域和区域之间可行走路线的平面图;结合消费者的采购清单,根据商品所在的货架位置将商品与具体的区域做出匹配,用遗传算法优化生成一条联结超市入口、要采购的商品区域以及结账柜台的最短路线。Matlab的仿真结果显示,该方法简单、高效,能够快速为消费者推荐出最短路线,供其购物参考。

关键词:超市导购;环境建模;最短路径;遗传算法;个性化推荐

文献标志码:A **中图分类号:**TP391 **doi:**10.3778/j.issn.1002-8331.1402-0358

1 引言

大型超市如家乐福、沃尔玛等已成为都市居民日常生活中不可缺少的一部分。信息技术特别是移动商务的应用以及智能手机的普及为消费者在移动设备上下单带来了便捷;但是消费者进入超市后,却因为对超市的环境、布局的不熟悉(因为超市中商品摆放并不是固定不变的),需要花费很多时间在商品寻找上。据统计,消费者通常花费20%的时间在商品的挑选上,剩余的

80%时间用在对商品的搜寻或者其他方面^[1]。另一方面,从消费者的偏好来看,顾客希望在较短的时间内完成行走过程^[2]。因此,要实现消费者在超市中进行在线下单、现场购买(一种O2O模式),最短路线推荐就变得很有必要。

文献[3]研究发现,大部分消费者在购物时都寻求以一种有效的行走方式来进行有序地商品采购。文献[4]在此基础上指出消费者在选购商品时也面临着与旅

基金项目:国家自然科学基金(No.71371062);国家重点基础研究发展计划(973计划)(No.2013CB329603)。

作者简介:韩建妙(1989—),男,硕士研究生,研究领域为超市个性化推荐,E-mail: hanjianmiao163@163.com;刘业政(1965—),男,博士,教授,博士生导师,研究领域为电子商务和数据挖掘。

收稿日期:2014-02-27 **修回日期:**2014-05-08 **文章编号:**1002-8331(2016)04-0238-05

CNKI网络优先出版:2014-06-18, <http://www.cnki.net/kcms/doi/10.3778/j.issn.1002-8331.1402-0358.html>

行商走访城市所类似的问题。文献[5]研究多个商品区域之间的最短距离最大化,通过让消费者在超市内停留更多的时间来促进商品销售。上述研究虽然考虑到需要在超市里作路线规划,但是均没有面向消费者作最短路线的推荐。目前有关路线推荐的研究主要集中在室外环境,如基于实时路况信息共享的自驾游路线推荐^[6]、基于RFID的主题公园游玩路线规划^[7]等。不同于室外环境,在超市等室内作路线推荐需要考虑到物理约束的限制,这类约束可以用货架等物理障碍进行特征化^[8]。存在障碍约束的最短路径规划问题一直是人们的研究热点,目前针对超市等室内的路径规划问题主要集中在对移动机器人的行走路线研究上^[9-12]。文献[13-14]提出根据消费者的购物清单实现导购路线的推荐,但是没有考虑到超市空间建模问题,且所采用的算法以启发式方法为主,计算复杂度高。文献[15]旨在研究消费者的购物行为,根据消费者的历史购物记录以及当前已经购买的商品,结合关联规则和序列事件预测技术推荐出其可能感兴趣的下一个商品,引导其进行购物。

在实体商店作推荐主要可以分为商品自身的推荐以及有关商品的位置布局、导购路线设计等的服务推荐两种形式。不同于文献[15]的商品推荐,本文主要研究根据消费者的购物清单为其做出最短导购路线的推荐。首先采用图论知识对超市空间环境进行离散化建模,生成用节点和边表示的平面图。根据消费者的购物清单将商品与具体的商品区域节点进行映射,采用遗传算法优化生成一条联结超市入口,购买清单上的商品和结账柜台的最短导购路线,并推荐给用户参考。文章最后通过仿真实验,介绍了该方法的具体应用。

2 问题描述与模型建立

2.1 问题描述

本文旨在为消费者自主推荐一条最短的超市室内导购路线。假设消费者想要购买的商品已知,本文根据其在采购清单上的商品信息,识别出商品所在的货架区域,为消费者推荐出一条联结超市入口、购物清单上的商品所在的货架区域、结账柜台的最短导购路线。

2.2 超市空间环境建模

超市空间环境建模的目的是通过对超市空间环境的特征进行考虑,转化成能够进行路径规划的计算机模型。考虑到在超市等实际卖场中,由于商品货架等物理障碍的存在,消费者的移动是高度受限的,即顾客只能沿着特定的过道行进,而不能随心所欲地在整个空间里行走。在这种高度受约束的空间环境中,采用离散化的建模方式显得更为合适。

如图1所示的超市平面布局图,阴影部分为商品货架,其他空白部分为可行走部分(过道),超市入口和结账柜台(出口)分别被设置在图的左下角和右下角。为了进行路线推荐,结合商品的属性,将空白部分离散化

成若干个不重叠的区域。这些区域和商品货架共同构成超市的整个空间物理结构。具体的,在分析超市平面布局的基础上,将过道部分划分为 n 个相互不重叠的区域并对其编号。将卖场入口和结账柜台抽象成两个区域,分别用1和 n 表示。将划分后的区域用节点表示,两个相邻区域之间用一条无向边进行连接,无向边表示两个区域之间的可行走路线。

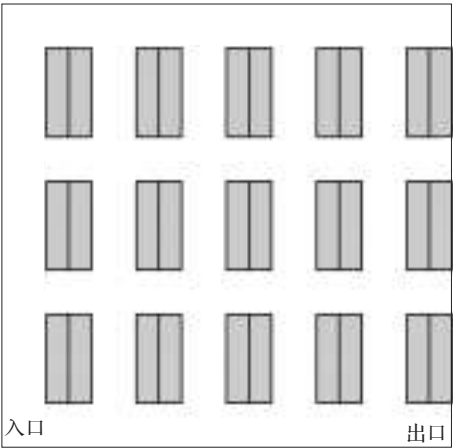


图1 超市平面布局图

简便起见,本文对两个相邻区域之间的距离作标准化处理。假设两个相邻区域之间的距离为一个单位,用户在它们之间可以一步到达,而在不相邻的区域之间则不能,需要经过其他的区域。离散化建模后的超市平面拓扑结构如图2所示。

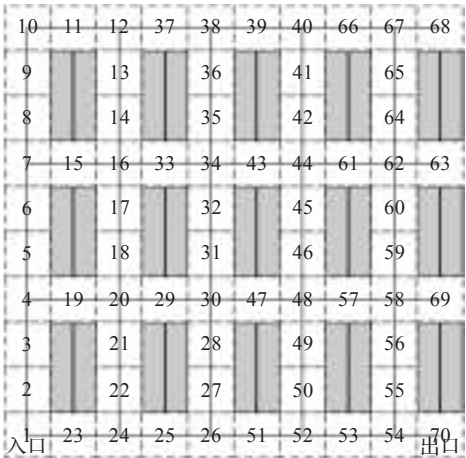


图2 离散化后的超市平面图

通过将超市空间结构映射成用节点和边表示的无向图,隐式地将货架等物理约束考虑在内。消费者在超市内的行走转化为在有限节点集合上的移动。超市内任意两个区域之间的最短行进距离被近似成是连接这两个区域节点之间的最短距离。此外,考虑到超市内同类商品通常摆放在相同货架上,根据商品的类别属性将商品与所划分的区域做出匹配,在区域水平上进行路线推荐,可以保证分析的粒度以及便于用户对同类商品进行比较查看。

2.3 最短导购路线数学模型

已知消费者的购物清单上有 m 种商品, 将商品与具体的节点进行匹配, 得到用户需要访问的节点集合 $M = \{l_i | i = 1, 2, \dots, m\}$ 。令 $G = (V, E)$ 表示离散化建模后的超市平面图, $V = \{v_i | i = 1, 2, \dots, n\}$ 为节点集合, 其中 v_1 和 v_n 分别表示超市入口节点和结账柜台(出口)节点, n 为节点总数; E 为任意相邻节点之间的无向边集合。令 $L = M \cup \{v_1, v_n\}$, 表示消费者要从超市入口 v_1 进入, 走过所有需要访问的商品区域节点集合 M , 最后从结账柜台(出口)节点 v_n 离开。令 $H = \{v_1, l_{s1}, l_{s2}, \dots, l_{si}, \dots, l_{sm}, v_n | l_{si} \in M\} = \{I_1, I_2, \dots, I_{m+2}\}$ 为 L 的一个有序节点排列, 即 H 为一条候选路径方案。 H 中任意两节点之间的最短距离 d 和最短路径 s 已知, 建立如下数学模型:

$$\min Z = \sum_{i=1}^{m+1} d(I_i, I_{i+1}), I_i \in H \quad (1)$$

式(1)的优化目标是求一条以入口节点 v_1 为路线起点, 以结账柜台(出口)节点 v_n 为路线终点, 联结消费者所有要买的商品所在的货架节点的最短路线。该问题在本质上等同于 TSP 问题, 本文采用遗传算法对上述模型进行求解。

3 基于遗传算法的消费者采购路径规划

遗传算法(Genetic Algorithm, GA)^[16]是一种模拟生物在自然界遗传和进化的全局搜索优化技术。通过选择、遗传、交叉、变异等操作, 遗传算法能够实现对路径规划等复杂问题的求解。

(1) 编码表示

本文采用整数编码的方式。如图2所示, 首先根据用户购物清单上的商品属性, 将商品与具体的区域进行匹配, 得到 m 个区域, 每个区域用对应的节点号进行编码。如分析某位用户的购物清单后, 得到节点集合 $\{56, 3, 17, 8, 12, 66, 46, 35\}$; 加上超市入口节点1和结账柜台(出口)节点70, 并分别置于编码串首端和末端, 最终用 $\{1, 56, 3, 17, 8, 12, 66, 46, 35, 70\}$ 表示一个单个体。该个体即是一种行走方案, 表示消费者从区域1进入, 分别走过区域56、3、17、8、12、66、46、35进行商品采购, 最后从区域70离开。

(2) 种群生成

编码表示之后需要初始化一定数目的个体作为种群, 并将该种群作为起始解。种群规模的大小一般视所要解决问题的复杂度而定。

(3) 适应度函数

适应度函数是用来衡量一个候选路径方案的优劣。本文适应度函数为:

$$fit = \left[1 - \frac{len - minlen}{maxlen - minlen + 0.0001} \right]^m \quad (2)$$

式中 fit 表示个体的适应度值, 个体的距离越短, 则适应度越接近1, 个体的适应度越高。 len 表示个体的路径总长度, $minlen$ 表示种群中所有个体所产生的路径长度的最小值, $maxlen$ 表示种群中所有个体产生的路径长度的最大值, m 表示适应度值归一化淘汰加速指数。个体的路径长度计算方法为: $len = \sum_{i \neq j} d_{ij}$, d_{ij} 表示节点 i 和节点 j 之间的距离, $1 \leq i, j \leq n$, n 为节点数。

为计算节点 i 和节点 j 之间的最短距离 d_{ij} , 本文采用 Floyd 算法进行求解。具体的, 首先根据图2构造如图3所示的邻接矩阵 D 。矩阵 D 中 v_i 即为图2中的第 i 个节点, $1 \leq i \leq n$, n 为节点数。定义若两个节点之间相邻, 则它们的距离等于1, 不相邻则为无穷大, 节点自身的距离为0。

$$\begin{matrix} & v_1 & v_2 & \dots & v_{n-1} & v_n \\ \begin{matrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{matrix} & \begin{pmatrix} 0 & 1 & \dots & \infty & 1 \\ 1 & 0 & \dots & \infty & \infty \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \infty & \infty & \dots & 0 & \infty \\ 1 & \infty & \dots & \infty & 0 \end{pmatrix} \end{matrix}$$

图3 邻接矩阵 D

基于 Floyd 算法求解任意两点之间的最短距离的步骤为^[17]:

① 首先初始化 $D^{(0)} = D$, $S^{(0)} = (s_{ij}^{(0)})_{n \times n}$, S 为最短路径矩阵。初始化 $S^{(0)}$: $s_{ij}^{(0)} = -1$, $1 \leq i, j \leq n$, n 为节点数;

② 分别计算矩阵 $D^{(1)}, D^{(2)}, \dots, D^{(n)}, S^{(1)}, S^{(2)}, \dots, S^{(n)}$ 。其中, 第 k 个矩阵 $D^{(k)}$ 是在矩阵 $D^{(k-1)}$ 的基础上加入第 k 个节点之后, 对 $D^{(k-1)}$ 中的节点之间的距离进行更新后所形成的新矩阵; 元素 $d_{ij}^{(k)}$ 的含义是允许中间节点的序号最大为 k 时, 从 v_i 到 v_j 的最短路径长度; 与之对应的第 k 个矩阵 $S^{(k)}$ 中的元素 $s_{ij}^{(k)}$ 表示从 v_i 到 v_j 的最短路线的第一条弧的结束节点序号。已知 $D^{(k-1)} = (d_{ij}^{(k-1)})_{n \times n}$, $S^{(k-1)} = (s_{ij}^{(k-1)})_{n \times n}$, 第 k 个最短距离矩阵 $D^{(k)} = (d_{ij}^{(k)})_{n \times n}$, 最短路径矩阵 $S^{(k)} = (s_{ij}^{(k)})_{n \times n}$, $k = 1, 2, \dots, n$, 分别定义为:

$$d_{ij}^{(k)} = \min\{d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}\} \quad (3)$$

$$s_{ij}^{(k)} = \begin{cases} s_{ij}^{(k-1)}, & d_{ij}^{(k-1)} \leq d_{ik}^{(k-1)} + d_{kj}^{(k-1)} \\ s_{ik}^{(k-1)}, & d_{ij}^{(k-1)} > d_{ik}^{(k-1)} + d_{kj}^{(k-1)} \end{cases} \quad (4)$$

③ 当 $k = n$ 时终止。此时 $D^{(n)} = (d_{ij}^{(n)})_{n \times n}$, 其中元素 $d_{ij}^{(n)}$ 的值等于 v_i 到 v_j 的最短路线长; $S^{(n)} = (s_{ij}^{(n)})_{n \times n}$, 其中元素 $s_{ij}^{(n)}$ 是 v_i 到 v_j 的最短路线的第一条弧的终点。

(4) 种群的选择操作

选择操作是从旧种群中以一定的概率选择出个体到新种群中。本文采用轮盘赌选择法^[18]进行个体选择操作。轮盘赌法的选择过程与个体的适应度值、选择概率和累积概率有关,在选择过程中需要进行多轮选择,每一轮产生 $[0, 1]$ 之间的一个随机数,并作为选择概率来确定被选的个体,若个体的当前累计概率大于或等于选择概率,则将该个体选择作为新种群的个体。

(5)交叉操作

交叉操作通过对种群内的个体进行两两分组,以交叉概率 P_c 对每组个体进行基因片段交叉重组。以 $\{1, 56, 3, 17, 8, 12, 66, 46, 35, 70\}$ 为例,采用部分交叉策略。通过产生区间 $(1, 10)$ 之间的两个整数 r_1 和 r_2 , 确定两个位置,对两个位置之间的数据进行交叉操作,如 $r_1=5, r_2=7$:

1 56 3 17 | 8 12 66 | 46 35 70
1 3 8 66 | 35 56 17 | 12 46 70

交叉为:

1 * 3 * | 35 56 17 | 46 * 70
1 3 * * | 8 12 66 | * 46 70

交叉后一个染色体中会出现重复的编号,根据中间段的对应关系将重复的数据进行消除,得到:

1 12 3 66 | 35 56 17 | 46 8 70
1 3 35 17 | 8 12 66 | 56 46 70

(6)变异操作

变异操作以变异概率 P_m 对个体染色体随机选择两个点,交换两个点的基因。如一条染色体为 $\{1, 12, 3, 66, 35, 56, 17, 46, 8, 70\}$, 随机产生 $(1, 10)$ 之间的两个数 r_1 和 r_2 , 确定两个位置,对两个位置上的数据进行交换,如 $r_1=4, r_2=7$:

1 12 3 66 35 56 17 46 8 70

交换为:

1 12 3 17 35 56 66 46 8 70

(7)进化逆转操作

为防止遗传算法陷入局部最优,提高搜索能力,在选择、交叉、变异之后对种群内的个体进行多次逆转操作。逆转的操作是在保持个体首尾编码不变的情况下,随机产生 $(1, 10)$ 之间的两个数 r_1 和 r_2 , 确定两个位置,对两个位置之间的数据进行反转,如 $r_1=2, r_2=6$:

1 | 12 3 17 35 56 | 66 46 8 70

逆转后:

1 | 56 35 17 3 12 | 66 46 8 70

逆转后的个体的适应度值进行重新计算,只有在适应度值有所提高的情况下才保留该个体,否则逆转无效。

通过对种群进行多次选择、交叉、变异、逆转操作,在全局范围内寻找最优的路径方案。整个遗传算法的流程图如图4所示。

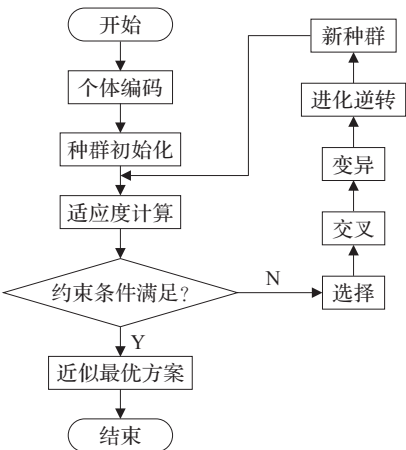


图4 遗传算法流程图

4 算例分析

为验证算法的稳定性,在 Matlab 中建立如图2所示的超市平面图。进行9组仿真实验,商品数分别为5、7、9、11、13、15、17、19和20个,每组实验中商品随机生成,分别对应图2中的其中一个区域节点。生成一组商品后,加上超市入口节点和出口节点,用遗传算法进行优化。取遗传算法的种群规模 $NIND=150$, 适应值归一化淘汰系数 $m=2$, 交叉概率 $P_c=0.9$, 变异概率 $P_m=0.15$, 优化迭代次数 $MAXGEN=100$ 。

每组实验在运行过程中记录算法在达到收敛时所需的迭代次数,每组实验重复进行20次,记录迭代数最大值、最小值及平均值,结果绘制如图5所示。

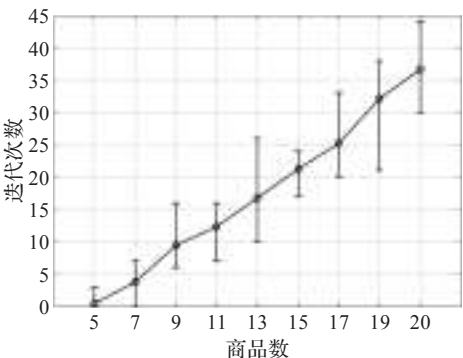


图5 迭代次数随商品个数变化图

图5中带圈的点表示每组实验的迭代数平均值,上下限分别表示迭代次数的最大值和最小值。虽然商品数的增多使得算法在每次收敛时所需的迭代次数也随之增长,但单从每组实验中看出,尽管在同一组中商品是随机生成的,算法的迭代次数却保持在一个稳定的区间内,且在商品数最多时(20个),算法仍在50次以内即可收敛,这说明本文算法是稳定的。

以顾客要买商品 $\{17, 12, 56, 46, 66, 3, 8, 35\}$ 为例,经优化生成如图6所示的最短路线,算法在迭代到接近10次时即达到收敛,总优化时间为5.2984 s。生成的最短路径长度为37,表示用户从入口节点出发,经过37步

就可以购买完全部商品,从出口节点离开。图中带圈的节点表示用户要到达的商品区域,带三角的节点分别表示入口节点和出口节点,带箭头的路线表示给用户的推荐路线过程,最终的整个行走路线为1→2→3→4→5→6→7→8→9→10→11→12→13→14→16→17→16→33→34→35→36→38→39→40→66→40→41→42→44→45→46→48→57→58→56→55→54→70。

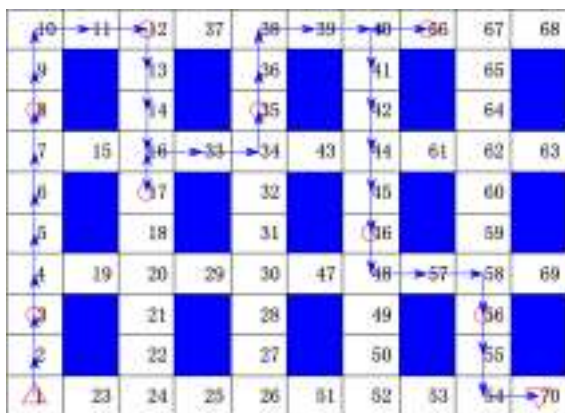


图6 最短路线生成图

5 结束语

本文提出基于遗传算法的消费者导购最短路线推荐,旨在根据消费者的采购清单,自主生成一条联结超市入口、全部欲购商品及结账柜台的最短路线,供消费者行进参考。通过将超市的空间结构离散化建模,生成用节点表示商品区域,用无向边表示两个相邻区域之间的可行走路线的平面图。将消费者要买的商品映射到具体的商品区域上,用遗传算法优化生成一条联结超市入口、所有要买的商品和结账柜台的最短路线。通过在Matlab上的仿真显示,本文的方法能够在短时间内生成一条最短路线,算法简单高效。

下一步的研究方向将根据消费者的历史购买记录、商品自身特点等信息,用数据挖掘等技术分析消费者潜在的购物需求,结合最优路径推荐,将顾客可能喜欢的商品也加入到购物路线的设计中,给用户带来购物惊喜,实现交叉销售和向上销售。

参考文献:

- [1] 索伦森 H. 卖一透视顾客心[M]. 孙静, 李铁成, 译. 北京: 中国人民大学出版社, 2011: 32-33.
- [2] Jia Hongwei, Jiang Tao, Zhao Fang. Research on route recommendation for indoor multi-travel destinations[C]//2012 IEEE World Automation Congress. Mexico: [s.n.], 2012: 1-4.
- [3] Larson J S, Bradlow E T, Fader P S. An exploratory look at supermarket shopping paths[J]. International Journal of Research in Marketing, 2005, 22(4): 395-414.
- [4] Hui S K, Fader P S, Bradlow E T. The traveling salesman goes shopping: the systematic deviations of grocery paths from TSP optimality[J]. Marketing Science, 2009, 28(3): 566-572.
- [5] Gertin T. Maximizing the cost of shortest paths between facilities through optimal product category locations[D]. Virginia: George Mason University, 2012.
- [6] Liu Long, Xu Jin, Liao S S, et al. A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication[J]. Expert Systems with Applications, 2014, 41(7): 3409-3417.
- [7] Tsai C Y, Chung S H. A personalized route recommendation service for theme parks using RFID information and tourist behavior[J]. Decision Support Systems, 2012, 52(2): 514-527.
- [8] Hui S K, Fader P S, Bradlow E T. Path data in marketing: An integrative framework and prospectus for model building[J]. Marketing Science, 2009, 28(2): 320-335.
- [9] 张捍东, 郑睿, 岑豫皖. 移动机器人路径规划技术的现状与展望[J]. 系统仿真学报, 2005, 17(2): 439-443.
- [10] AL-Taharwa I, Sheta A, Al-Weshah M. A mobile robot path planning using genetic algorithm in static environment[J]. Journal of Computer Science, 2008, 4(4): 341-344.
- [11] 石为人, 王楷. 基于 Floyd 算法的移动机器人最短路径规划研究[J]. 仪器仪表学报, 2009, 30(10): 2088-2092.
- [12] 胡喜玲, 李洪波, 胡俊. 基于自适应混沌遗传算法的路径规划[J]. 计算机工程与应用, 2013, 49(9): 68-73.
- [13] Pradhan S, Krishna P R, Rout S S, et al. Wish-list based shopping path discovery and profitable path recommendations[C]//Proceedings of 3rd IEEE International Conference on Services in Emerging Markets. India: [s.n.], 2012: 101-106.
- [14] Pardines I, Lopez V. Shop&Go: TSP heuristics for an optimal shopping with smartphones[J]. Science China Information Sciences, 2013, 56(11): 1-12.
- [15] Rudin C, Letham B, Salieb-Aouissi A, et al. Sequential event prediction with association rules[C]//Proceedings of 24th Annual Conference on Learning Theory, Budapest: [s.n.], 2011: 1-13.
- [16] Potvin J Y. Genetic algorithms for the traveling salesman problem[J]. Annals of Operations Research, 1996, 63(3): 339-370.
- [17] 李明哲, 金俊, 石端银. 图论及其算法[M]. 北京: 机械工业出版社, 2010: 72-73.
- [18] 王小平, 曹立明. 遗传算法—理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002: 10-12.