

基于时空序列搜索的最优个性化旅游规划

周春姐 曲海平 刘 莉

(鲁东大学信息与电气工程学院 烟台 264000)

摘 要 随着人们对高效、个性化服务的日益关注,其对旅游规划的要求也越来越高。然而,目前制定一个完美的、个性化的旅游规划很困难。一个合理的旅游规划应该包含以下特征:1)基于游客兴趣爱好的个性化推荐;2)覆盖尽可能多的兴趣景点;3)走最少的路程,花费最少的时间。由于每次旅游包含多个景点,因此该问题变成了每个景点都附有权值的时空序列优化问题。为了实现最优个性化旅游规划,首先,针对每个城市为每位游客提供一系列带有权值的景点;然后,综合考虑距离、权值、旅游时间和景点特色等因素,为游客推荐一个最优旅游序列。为此,提出了4种有效算法,使游客在旅游时间范围限制内走最少的路程来游览最想游览的景点。在社交网络的真实数据集上,通过实验验证了所提算法的有效性。

关键词 时空,序列,社交网络,个性化,旅游规划

中图法分类号 TP391

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2017.09.052

Optimal Personalized Trip Planning Based on Spatio-Temporal Sequence Searching

ZHOU Chun-jie QU Hai-ping LIU Li

(School of Information and Electrical Engineering, Ludong University, Yantai 264000, China)

Abstract Nowadays, due to the increasing user requirements of efficient and personalized services, a perfect trip planning is urgently needed. However, it is hard for people to make a personalized traveling plan at present. For a satisfactory trip planning, the following features are desired: i) personalized recommender based on the interest and habits of travelers; ii) maximal coverage of sites of interest; iii) minimal effort such as transporting time on the route. As a trip contains a sequence of stops at multiple scenes, the problem of trip planning becomes an optimized spatio-temporal sequence where each stop is weighted. For each city, this paper generated a set of weighted scenes for each user. Then it can retrieve the optimal sequence of scenes in terms of distance, weight, visiting time and scene features. We developed alternative algorithms for searching optimal sequences, with consideration of the weight of each scene, the preference of users, and the travel time constraint. The experiments demonstrate the efficiency of the proposed algorithms based on real datasets from social networks.

Keywords Spatio-temporal, Sequence, Social networks, Personalized, Trip planning

1 引言

随着当代社会的快速发展,人们越来越关注高效、个性化的服务。在旅游行业中,一个完美的旅游规划能够帮助人们节省大量的时间和精力,并且使其游览尽可能多的、自己感兴趣的景点。然而,目前要制定这种完美的、个性化的旅游规划十分困难。大多数游客都是通过口耳相传、旅游公司、旅游博客、旅游宣传册等简单地遵循其他人的旅游路线,没有考虑自己的兴趣爱好,也没有考虑该路线中景点的最优游览顺序。考虑这样的一个场景:某人计划在假期出游,但是没有明确的目的。为了制定一个满意的旅游规划,他在网络上浏览了各类旅游路线,或者从旅游公司寻求建议。然后根据别人的

建议选择一条最流行的旅游线路,但是他很少考虑自己的兴趣爱好。这种盲目的旅游规划可能导致以下后果:1)在不感兴趣的景点上浪费时间;2)错过了真正感兴趣的景点;3)在景点与景点之间的路上浪费了大量的时间;4)增加了很多不必要的开销。因此,一个合理的旅游规划应该包含以下特征:1)基于游客兴趣爱好的个性化推荐;2)覆盖尽可能多的兴趣景点;3)走最少的路程,花费最少的时间。

为了实现这些目标,针对每个城市为每位游客提供一系列带有权值的景点。进一步,综合考虑距离、权值、旅游时间和景点特色等因素,为游客推荐一个最优旅游序列,使其在旅游时间范围限制内,走最少的路程来游览最想游览的景点。因此,游客可以在开始旅游之前,根据自身的兴趣爱好制定一

到稿日期:2016-08-20 返修日期:2016-12-31 本文受国家自然科学基金项目(61202111,61472141,61273152,61303017),山东省高等学校科技计划项目(J12LN05),山东省自然科学基金联合专项项目(ZR2013FL009),烟台市科技发展计划项目(2013ZH092,2014JH042),鲁东大学博士基金项目(LY2012023)资助。

周春姐(1981-),女,博士,副教授,主要研究方向为物联网、数据挖掘,E-mail:lucyzcj@gmail.com;曲海平(1981-),男,博士,讲师,主要研究方向为物联网、大数据;刘莉(1982-),女,博士,主要研究方向为大数据、云计算。

个高效的、个性化的旅游规划。

为了说明该研究的必要性,图1给出了一个旅游规划的具体应用场景。当某游客到达一个城市时, OrientSTS 自动根据景点特色和该游客的兴趣爱好为各个景点赋予权值,然后选取6个具有较大权值的兴趣景点($o_1, o_2, o_3, o_4, o_5, o_6$),其权值分别是($w_1, w_2, w_3, w_4, w_5, w_6$)。针对这6个被选定的景点,结合知识库信息给出一个最优时空旅游序列(图1中的深灰色序列)。同时,根据多数游客的旅游经验产生另外一个序列(图1中的浅灰色序列)。其中,旅游耗时既包括每个景点的游览耗时,也包括景点之间转换的路程耗时(如图1中时间轴上的矩形框所示^[1])。由图可见,对同一组景点游览顺序的不同导致总的旅游耗时存在很大差异:实线序列的耗时少于6个小时,而虚线序列的耗时多于9个小时。

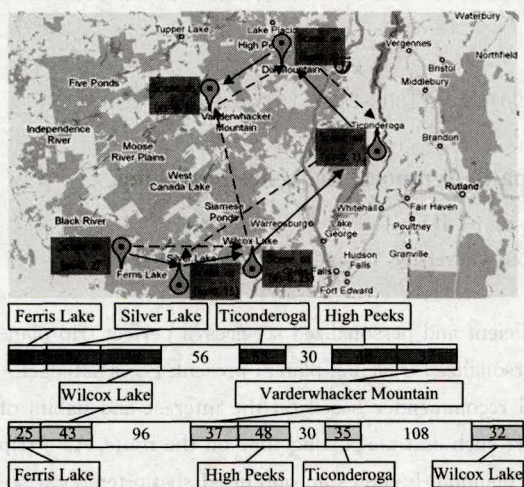


图1 两条不同序列的耗时对比

用图来解决时空旅游规划问题,其中图的顶点表示旅游景点,边表示景点之间的路径。据我们所知,目前没有任何相关工作同时考虑了成反比关系的顶点权值和边距离两个因素。本文结合游客兴趣爱好和景点特色,针对不同应用场景,提出了有效的近似算法和启发式算法来实现时空旅游规划。本文的主要贡献点总结如下:

(1)提出了3种局部近似优化算法,分别是对距离、权值和价值的近似优化。这些算法分别在以下情形中是高效的:1)当游客更关注距离时;2)当游客更关注景点特色时;3)当游客既关注距离,又关注权值时。

(2)提出了一种新颖的启发式算法,当启发式函数永远小于实际到达目标景点的最小价值时,该算法可以找到一个最优旅游序列。

(3)在社交网络的真实数据集上进行了性能测试,并证明了算法的有效性。

本文第2节介绍了相关研究工作;第3节给出了问题的定义;第4节提出了3种近似算法;第5节提出了一种启发式算法;第6节利用真实数据集并通过实验验证了所提算法的有效性;最后总结全文并对未来工作进行展望。

2 相关工作

在旅游规划研究中,文献[2]提出了一个旅游代理,为特定游客进行旅游规划推荐;他们认为要推荐一个满意的旅游

规划,与游客进行沟通和协商是非常必要的。一个好的旅游代理不仅要满足游客的个人偏好,更要考虑时间、空间、花费等方面的制约。文献[3]提出了一个基于多代理的系统进行旅游规划,他们将旅游规划问题转化成谓词逻辑,利用树形检索找出可行的旅游路线。但是该研究缺少路径选择,也没有相关算法的评估验证。文献[4]提出了基于规则的旅游规划,并且设计了基于位置、推荐服务的快捷传输方法,基于游客的当前位置、时间、个人偏好和需求提供不同的推荐。然而,这项工作需要旅游专家事先定义旅游规则,这对多数游客而言是不适用的。文献[5]提出了一种个性化路线规划网络阻抗的方法,在语义描述的各路段上仿真了各种交通规则,用户可以根据交通量、安全性、感兴趣的场所等因素选择自己偏好的路段,然后,根据用户的选择计算路网上各个路段的不同权值。文献[6]提出了一个智能推荐系统,该系统可以根据很多用户的GPS轨迹快速找到所需路线。在这个路线推荐系统中,用户需要提供起始地、目的地和旅游时间。基于有地理标志的照片信息,文献[7]能够自动地发现带有语义注释的景点序列。然而,其没有考虑行程约束和游客的偏好。文献[8]致力于解决自动旅游路线规划的问题,试图用动态规划来提高路线选择的效率。动态规划方法基于深度优先搜索遍历所有路径,在遍历过程中,如果某个节点的游览时间大于总的旅游时间,那么包含该节点的子树就会被过滤掉,以提高计算效率。文献[9]提出了路线中包含交通事故的个性化旅游规划方法,用来解决两个NP-hard优化问题(RMT和BMT)到任意精度。文献[10]提出了一个交互式旅游包配置系统,用户可以通过简单的交互来配置自己的旅游包。文献[11]提出了一个实时的旅游规划应用,该应用不仅能够规划新的旅游线路,而且能够重新规划当前旅游线路。文献[12]提出了3种优化算法,从当前景点出发,以最小的延迟游览所有的景点。文献[13]提出了旅游规划查询的3种算法,然而其中的安全区域生成法很耗时。文献[14]提出了一个交互式的旅游包整合系统,该系统将各景点虚拟成候选旅游包,通过简单交互为用户提供一个整合的旅游包。然而,上述研究的效率都与景点的数目相关。

基于台南地区的多家代理,文献[15]提出了一种语义推荐方法,即根据游客的需求,代理公司给出个性化的旅游路线推荐。但是,代理公司提供的景点数目和酒店数目都是固定且一成不变的。文献[16]采用启发式算法来快速解决城市旅游规划问题。文献[17]致力于从多个用户的GPS记录中挖掘相似的旅游序列。而文献[18]从时空元数据中抽取最大的周期模式。Girardin等^[19]基于大量带有地理标志的Flickr照片,分析了印度Florence省的旅游客流,其结果为更好地了解人们的旅游方式做出了贡献。Chen等^[20]研究了基于位置查找轨迹的问题,其目标是从数据库中查找K条最相关的轨迹,使其能够连接到指定的位置。文献[21]介绍了3种算法来实时地查找从当前位置出发,游览所有景点的最优路线。文献[22]根据奖励度量计算最优行程。文献[23]提出了一种个性化旅游推荐的框架结构,但是没有考虑旅游规划。文献[24]提出了一个实时的旅游规划应用,该应用能够基于非结构化的自然语言输入规划一次旅游。文献[25]提出了一种空间近似算法,使得在给定距离范围内或者在旅游时间限制范

围内,用户的满意度最高。文献[26]通过对旅行者的选择进行学习,提出了一种用来提高旅游推荐的相关性的方法,其更类似于一个信息抽取系统。然而,上述研究工作都没有考虑景点的旅游特色和用户的兴趣爱好,而这正是本研究的重点。

3 预备知识

表1列出了本文所使用的主要符号及其描述。用大写字母表示集合(例如, C 表示所有景点的集合),小写字母表示变量(例如 o_i)。

表1 符号表

符号	定义和描述
V	景点序列 $\{o_1, \dots, o_k\}$ 的价值
C	所有景点的集合 $\{C_1, \dots, C_m\}$
R	所选择景点的集合 $\{o_1, \dots, o_k\}$
Q	一个旅游序列
N	路网
$W^{x,y}$	对象 x 和 y 组成的邻接矩阵
T_{total}	总的旅游时间
$T(o_i, o_j)$	从景点 o_i 到景点 o_j 所花费的时间
$T(o_i)$	景点 o_i 的游览耗时
D_{dis_N}	路网上各个景点之间的距离
w_i	所选择的景点 o_i 的权值
o_i	第 i 个被选择的景点
α	平衡因子
m	所有景点的数目
k	所选择的景点的数目

用图来表示时空旅游规划问题,其中旅游景点是图节点(标注为景点名字),景点之间的路径为图的边。给定一个图 $G(O, \xi)$,该图有 n 个节点 $O = \{o_1, \dots, o_n\}$, s 条边 $\xi = \{e_1, \dots, e_s\}$,并且每个节点都有一个权值,用来表示游客对该景点的喜好程度。景点序列 $\{o_i, \dots, o_j\}$ 的价值表示为 $V(o_i, \dots, o_j) \geq 0$,代表该景点序列的打分,其定义如式(1)所示。序列的价值与所选择景点的权值总和成正比,而与总距离成反比。在此,考虑景点之间的路网距离,这与实际应用的场景更相符。假设 v' 是游客的平均速度,因此距离 $D_{dis_N}(o_i, o_j)$ 可以表示为 $v' * T(o_i, o_j)$ 。 α 被定义为权值和距离之间的平衡因子,它根据不同的应用场景发生变化,随机调整。

$$V(o_i, o_j) = \alpha * (w_i + w_j) + (1 - \alpha) * \frac{1}{D_{dis_N}(o_i, o_j)}$$

$$= \alpha * (w_{i,j}) + (1 - \alpha) * \frac{1}{v' * T(o_i, o_j)} \quad (1)$$

给定 m 个景点集合 $C = \{C_1, \dots, C_m\}$ (其中 $m \leq n$),一个映射函数 $\pi: C_j \rightarrow o_i$,将图中的每个景点 $C_j \in C$ 映射为一个节点 $o_i \in N$ 。因此,景点可以看作是特殊的节点,用节点符号标注,本文中景点和节点共用同一符号。

基于上述定义,时空旅游规划问题可以定义如下。

定义1 给定一个集合 $R \subseteq C$ ($R = \{o_1, o_2, \dots, o_k\}$),一个开始节点 S 和一个结束节点 E 。目标是找出一个从 S 到 E 的旅游景点序列 $Q = \{S, o_1, \dots, o_k, E\}$,使得该序列的价值最大,并且总的游览时间不超过 T_{total} 。即,根据式(1),对于某个特定用户(即平衡因子固定),该序列的权值最大(即游览 R 中尽可能多的兴趣景点),被选择景点之间的路程最少,并且总的游览时间在 T_{total} 范围限制内。

时间约束条件如式(2)所示,其不仅包括每个景点的游览耗时,而且包括各个景点之间的转换时间。景点的游览耗时

可以通过文献[27]得到,这一部分不是研究重点,主要关注景点之间转换的时间代价,即每对景点之间的路程长度。

$$\sum_{i=1}^k T(o_i) + \sum_{i=1, j=i+1}^k T(o_i + o_j) \leq T_{total} \quad (2)$$

从图 G 的节点 u 出发,遍历其相邻节点 o ,然后再从 o 出发遍历其相邻节点,以此类推,直到终节点 v 。这时可以得到图 G 中从 u 到 v 的一个序列 Q ,即图中的一系列节点集合,并且序列中连续的节点在图中是相邻的。序列 Q 可以表示为:

$$Q: u = o_1, o_2, \dots, o_k = v \quad (3)$$

其中, $k \geq 1$, o_i 与 o_{i+1} ($1 \leq i \leq k$)是相邻的。每个节点 o_i 和每条边 $o_i o_{i+1}$ ($1 \leq i \leq k-1$)都属于 Q 。节点的数目代表序列的长度。因此序列 Q 的长度为 k ,如式(3)所示。

4 时空旅游规划的局部优化算法

本节提出3种局部近似优化算法,分别是对距离、权值和价值的近似优化。这些算法分别在以下情形中是高效的:

1)当游客更关注距离时;2)当游客更关注景点特色时;3)当游客既关注距离又关注权值时。因为这些算法都只考虑了时间旅游规划问题中的某一方面,所以称之为局部优化算法。

4.1 对距离的近似优化

解决时空旅游规划问题的一个最直接的算法就是通过循环计算当前景点的 $\lfloor m/2 \rfloor$ 个最近邻景点,计算并比较它们的价值,找出还未被游览的景点中价值最大的景点作为下一个游览目标。同时,更新总的游览时间,将其加上该景点的游览时间以及该景点到下一景点之间的耗时。如果更新后的总时间小于 T_{total} ,则将该景点加到旅游序列中;否则,恢复总游览时间的值,并忽略该景点。通常,给定一个偏序列 Q_k , $k < m$,通过插入景点 o_{k+1} 形成序列 Q_{k+1} 。 o_{k+1} 是 o_k 的 $\lfloor m/2 \rfloor$ 个近邻景点中还没有被游览过、价值最大且满足总时间约束的景点。最后,将 o_k 与 E 连接得到最终旅游序列。称该算法为d-LOA,具体见算法1。当游客对距离更加关注时(比如游客骑自行车旅游),该方法很有效。然而,该方法选取的景点可能并不是游客最喜爱的景点。

算法1 d-LOA 算法

输入:开始景点 $o=S$;结束景点 E ;景点ID集合 $I=\{1, \dots, m\}$;初始的时空旅游序列 $Q_a=\{S\}$;总时间约束 T_{total}

输出:对距离近似优化的时空旅游序列 Q_a

1. loc=S;
2. t=T_{total};
3. WHILE(I非空并且 t>0) DO
4. 定义一个数组DS来存储从loc到其他景点的距离;
5. FOR each n∈I do
6. DS[loc, n]=D_{dis_N}(loc, n);
7. END FOR
8. HI=DS最小的 $\lceil m/2 \rceil$ 个景点的ID集合;
9. 定义一个数组V来存储从loc到其他景点的价值;
10. FOR each n∈HI do
11. IF T(n) + T(loc, n) ≤ t THEN
12. V[loc, n]= $\alpha * (w_{loc} + w_n) + (1 - \alpha) * \frac{1}{DS[loc, n]}$
13. END IF
14. END FOR
15. o=V中价值最大的景点;

16. loc=o;
17. 从 I 中去掉 o;
18. 将 o 放入 Q_a 中;
19. $t=t-(T(o)+T(loc,o))$;
20. END WHILE
21. $Q_a=Q_a+\{E\}$;

4.2 对权值的近似优化

解决时空旅游规划问题的另一个算法是通过循环执行以下操作得到一个旅游景点序列。首先,选取与当前景点相连的、还未游览过的 $\lceil m/2 \rceil$ 个权值最大的景点;其次,计算并比较这些景点的价值,选择价值最大的景点;然后,更新总的游览时间,将其加上该景点的游览耗时以及该景点和上一景点之间的路程耗时。如果总的游览时间小于 T_{total} ,则将景点加入到序列中;否则,忽略该景点,还原总游览时间的值。称该算法为 w-LOA,其算法描述与算法 1 类似,此处不再赘述。当游客更加关注景点特色时,该算法是有效的。然而,该方法可能会导致很长的旅游距离,在景点之间耗时较多。

4.3 对距离和权值的近似优化

解决时空旅游规划问题的一种混合局部优化算法就是通过循环执行下列操作得到一个旅游景点序列。首先,计算当前景点和其他任何未游览过的景点之间的价值;其次,选取价值最大的景点;然后,更新总的游览时间,将其加上该景点的游览耗时以及该景点和上一景点之间的路程耗时。如果总的游览时间小于 T_{total} ,则将景点加入到序列中;否则,忽略该景点,还原总游览时间的值。通常,给定一个偏序列 Q_k ,其中 $k < m$,通过插入景点 o_{k+1} 形成序列 Q_{k+1} 。 o_{k+1} 是 R 中还未被游览过、满足总时间约束且价值最大的景点。最后,将 o_k 与 E 连接得到最终的旅游序列。称该算法为 v-LOA,具体见算法 2。这种方法同时考虑了距离和权值,因此,能够得到一个较好的时空旅游序列,其具有最小距离和最大权值。

算法 2 v-LOA 算法

输入:开始景点 $o=S$;结束景点 E ;景点 ID 集合 $I=\{1,\dots,m\}$;初始的时空旅游序列 $Q_a=\{S\}$;总时间约束 T_{total}

输出:对距离和权值近似优化的时空旅游序列 Q_a

1. loc=S;
2. $t=T_{total}$;
3. WHILE (I 非空并且 $t>0$) DO
4. 定义一个数组 V 来存储从 loc 到其他景点的价值;
5. FOR each $n \in I$ do
6. IF $T(n) + T(loc,n) \leq t$ THEN
7. $V(loc,n) = \alpha * (w_{loc} + w_n) + (1-\alpha) * \frac{1}{D_{dis_N}(loc,n)}$
8. END IF
9. END FOR
10. $o=V$ 中价值最大的景点;
11. loc=o;
12. 从 I 中去掉 o;
13. 将 o 放入 Q_a ;
14. $t=t-(T(o)+T(loc,o))$;
15. END WHILE
16. $Q_a=Q_a+\{E\}$;

5 时空旅游规划的全局优化算法

除了前面介绍的近似算法外,本文还提出了一种新颖的

启发式算法,即全局优化算法(Global Optimization Algorithm,GOA)。与之前的 3 种算法相比,GOA 算法能得到更好的规划结果。当启发式函数永远小于实际到达目标景点的最小价值时,GOA 算法可以找到一个最优旅游序列。本文选择欧几空间中旅游序列的价值作为启发函数,如定义 2 所述。在这种场景下它永远小于或等于路网上实际序列的价值,从而可以保证得到路网上价值最优的序列。

定义 2 对于节点 u 和 $v(u,v \in N)$, $D_{dis_N}(u,v)$ 表示路网距离, $D_{dis_E}(u,v)$ 表示欧几距离。相应地, $V_N(u,v)$ 表示路网上从节点 u 到 v 的序列的价值,而 $V_E(u,v)$ 表示其在欧几空间的价值。本研究更加关注个人的兴趣爱好,因此设定平衡因子 $\alpha > 0.5$ 。由式(1)可以得到 $V_E(u,v) \leq V_N(u,v)$ 。

5.1 查找最大价值的旅游序列

利用启发式算法 GOA,可以找到从当前景点到 R 中任何景点的最大价值。其详细描述见定理 1。

定理 1 o 是从 u 到 v 的旅游序列中的一个中间景点,最大价值的旅游序列 $V_{NE}(u,o,v)$ 由经过 o 的价值最大的一系列景点组成。因此,式(9)成立。

$$V_{NE}(u,o,v) = V_N(u,o) + V_E(o,v) \quad (9)$$

证明: $V_N(u,o)$ 表示从起始景点 u 到景点 o 的价值,而 $V_E(o,v)$ 表示从景点 o 到目标景点 v 的价值的启发式函数。由启发式算法的原始概念可知式(9)成立。其中, $V_N(u,o)$ 可以由推理 1 得到,而 $V_E(o,v)$ 可以由推理 2 得到。证毕。

推理 1 假设起始景点 $u:=o_0$, 目标景点 $v:=o_k$, 一个路网上的旅游序列 $(o_0, o_1, \dots, o_{k-1}, o_k)$ 由 $k+1$ 个兴趣景点组成。旅游序列的价值 $V_N(o_0, o_k)$ 表示从景点 o_0 出发,经过景点 $o_1, \dots, o_{k-1}$, 到景点 o_k 的序列值:

$$\sum_{i=1}^k V_N(o_{i-1}, o_i) = V_N(o_0, o_k) \quad (10)$$

证明:价值函数 $V_N(o_0, o_k)$ 表示在路网上从景点 o_0 到 o_k 的旅游序列的价值总和,即,该价值是一个渐增的序列价值,从起始景点 o_0 到目标景点 o_k 的过程要经过尽可能多的景点 $o_1, \dots, o_{k-1}$ 。因此,总价值 $V_N(o_0, o_k)$ 可以被划分为 $V_N(o_0, o_1) + V_N(o_1, o_2) + \dots + V_N(o_{k-1}, o_k)$, 即 $\sum_{i=1}^k V_N(o_{i-1}, o_i)$ 。证毕。

推理 2 假设节点 o 和节点 v 分别是当前景点和目标景点, $h(o)$ 是启发式估计量,那么对于启发式价值函数 $V_E(o, v)$, 式(11)成立。

$$h(o) \leq V_E(o, v) \leq V_N(o, v) \quad (11)$$

证明:如果目标景点是可达的,那么启发式预测函数就能找到到达目标景点的、最优的旅游序列。根据定义 2,将欧几空间的价值作为启发式函数,它是从景点 o 到 v 的价值的下限。因此,当 $h(o)$ 作为启发函数且近似等于欧几空间的价值时, $h(o) \leq V_E(o, v)$ 成立。另外,尽管欧几距离小于或等于路网距离,但是由于平衡因子大于 0.5,因此从景点 o 到 v 的旅游序列价值不等式 $V_E(o, v) \leq V_N(o, v)$ 成立。具体分析如下:

$$V_N(o, v) = \alpha * (w_o + w_{o_1} + \dots + w_{o_j} + w_v) + (1-\alpha) * \frac{1}{D_{dis_N}(o, o_1) + \sum_{k=i}^{j-1} D_{dis_N}(o_k, o_{k+1}) + D_{dis_N}(o_j, v)}$$

$$V_E(o, v) = \alpha * (w_o + w_v) + (1-\alpha) * \frac{1}{D_{dis_E}(o, v)}$$

因此

$$\Delta V = V_N(o, v) - V_E(o, v) \\ = \alpha * (w_{o_1} + \dots + w_{o_j}) + (1 - \alpha) * \frac{1}{\Delta D_{dis}(o, v)} > 0$$

证毕。

5.2 有效查找最优景点

采用文献[28]中介绍的分支界定技术来查找最优的旅游序列(即距离最小、景点权值最大的旅游序列)。根据定理2,该技术能够过滤掉与当前景点相连的、所有不必要的景点,即能找到一个最优时空旅游序列,使该序列的价值最大。该方法能够选择从景点 u 出发到达邻近景点 $o_1, o_2, \dots, o_i, \dots, o_k$ 的且具有最小距离 $D_{disNE}(u, o_i, v)$ 的景点 o_i 。同样,该方法也能够从可选择的景点 $o_1, o_2, \dots, o_j, \dots, o_k$ 中选取具有最大权值 $w(u, o_j, v)$ 的景点 o_j 。因此,对最优旅游序列的定义如下。

定义3 假设 u 和 v 分别是起始景点和目标景点,最优旅游序列是从 u 到 v 的有序景点的集合,它具有最大的价值 $V_{NE}(u, o_i, v)$ (即具有最小的距离 $D_{disNE}(u, o_i, v)$ 和最大的权值 $w(u, o_i, v)$),其中 o_i 是被选中的且与 u 相邻的景点。

定理2描述了如何利用分支界定技术查找最优旅游序列,即如何由 $V_{NE}(u, o_i, v)$ 得到 $V_{NE}^*(u, o_i, v)$ 。

定理2 给定路网上的两个景点 u 和 $v(u, v \in N)$,如果存在一个选定的景点集合 $o_1, o_2, \dots, o_i, \dots, o_k$ 同时连接景点 u 和 v ,则由定义3得出式(12)成立。

$$V_{NE}^*(u, o_i, v) = \max_{1 \leq i \leq k} (V_{NE}(u, o_i, v)) \quad (12)$$

证明:假设 o_i 和 o_j 是与已选定景点 u 相连接的两个邻近

景点。如果 $V_{NE}(u, o_i, v)$ 大于其他任何两个相邻景点的价值,即 $V_{NE}(u, o_i, v) > V_{NE}(u, o_k, v) (\forall k \in (i, j])$,则 $V_{NE}(u, o_i, v)$ 是同时连接 u 和 v 的经过 o_i 的所有邻近景点中的最优旅游序列。证毕。

根据定理2,通过预先计算从邻近景点到目标景点的价值,分支界定技术能够过滤掉很多景点,从而避免不必要的景点扫描。

5.3 有效查找最优旅游序列

时空旅游规划查询的结果是一个有序景点旅游序列以及它们之间的路径。图2给出了在 k 个景点中有效查找时空旅游序列的例子。首先,利用式(11)选择一条从开始景点 S 出发到其他任何一个被选中景点的路径。这一过程包含两步,先利用式(1)选择到开始景点的价值最大的第一个景点,再找出它们之间的路径。为了避免重复查找已经确定的路径,需要一个队列来存储通过 $V_N(S, o_i)$ 计算出来的已经扫描的景点的最优价值,以及从 S 到 o_i 的路径。图2中,分别查找从开始景点 S 到景点 o_1, o_2, o_3 的时空旅游序列,并且选取队列中价值最大的景点 o_2 。然后计算 $V_{NE}(S, o_2, o_1), V_{NE}(S, o_2, o_3)$ 和 $V_{NE}(S, o_2, o_4)$ 。为了找到最优的景点,利用式(14)计算 $V_{NE}^*(S, o_2, o_4)$,并将路径存储在队列中。依此类推,可以产生一个最大价值的旅游序列 $S \rightarrow o_2 \rightarrow o_4$ 。接下来,利用式(11)查找从当前景点到其他未被游览过的景点的最大价值,从而刷新序列。如果此时序列的总时间小于或等于时间约束 T_{total} ,则以同样的方法循环刷新得到其他剩余景点的旅游序列。

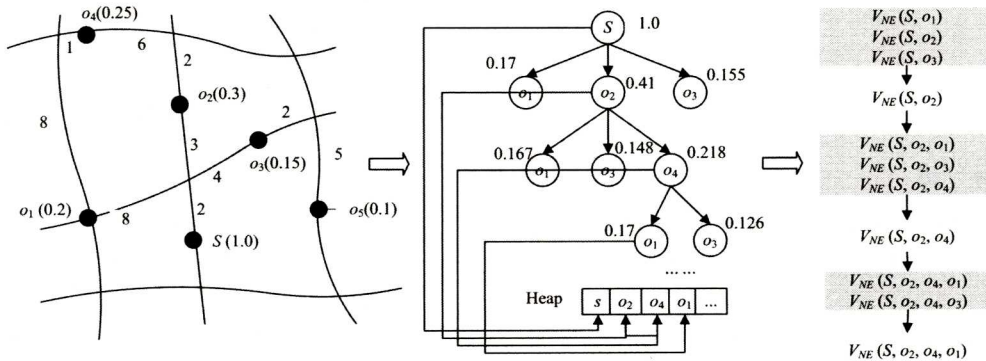


图2 寻找最优景点序列

算法3描述了GOA算法的实现过程。首先将开始景点加入到OPEN列表中,并将该景点作为当前景点。第3—6行表示如果目标景点已经被加到CLOSE列表中,则该时空旅游序列已经被找到,并停止循环;否则,重复执行第7—28行。从OPEN列表中选取价值最大的景点。通过加入该景点的游览耗时以及该景点与上一景点之间的路程耗时,更新总的游览时间。如果总游览时间小于 T_{total} ,则把该景点移到CLOSE列表中;否则,忽略该景点,并且还原总游览时间。对于与当前景点邻近的其他任何一个景点,如果是不可达的,或者它已经在CLOSE列表中,则忽略它;否则,执行如下操作:如果它没在OPEN列表中,就将其加到OPEN列表中,并将当前景点看作该景点的祖先,计算该景点的价值;如果它已经在OPEN列表中,则通过当前序列的价值判断这条到达该景

点的路径是否是最好的(价值越大表示该路径越好),如果该路径已经是最好的,则将该景点的祖先看作当前景点,并且重新计算该景点的价值。

算法3 GOA算法

输入:开始景点 $o=S$;结束景点 E ;景点ID集合 $I=\{1, \dots, m\}$;总时间约束 T_{total}

输出:全局最优的时空旅游序列 Q_a

1. 数组OPEN用来存储所有被选取但还未游览过的景点
OPEN=[S];
2. 数组CLOSE用来存储所有游览过的景点
CLOSE=[];
3. WHILE (OPEN非空并且 $T_{total} > 0$) DO
4. 抽取OPEN中的第一个景点 o ;
5. 将 o 放入CLOSE中;

```

6. IF o=E THEN
7.   break;
8. END IF
9. M 等于 I 中景点 o 的孩子景点数目;
10. FOR each s∈M do
11. IF T(o) + T(o,s)≤Ttotal THEN
12.   计算评估值;

$$EV(o,s) = \alpha * (w_o + w_s) + (1-\alpha) * \frac{1}{D_{dis_N}(o,s)}$$

13. IF(s 没在 OPEN 和 CLOSE 中) THEN
14.   从 I 中去掉 s;
15.   将 s 放入 OPEN 中;
16. ELSE IF(s 在 OPEN 中) THEN
17. IF (EV(o,s) > EV(OPEN)) THEN
18.   更新 OPEN 的值;
19. END IF
20. ELSE
21. IF (EV(o,s) < EV(CLOSE)) THEN
22.   更新 CLOSE 的值;
23.   从 CLOSE 中去掉 s;
24.   将 s 放入 OPEN;
25. END IF
26. END IF
27. END IF
28. END FOR
29. 将 o 放入 CLOSE;
30. 根据 EV 对 OPEN 中的景点进行降序排列;
31. Ttotal = Ttotal - (T(o) + T(o,s));
32. END WHILE
33. Qi = CLOSE;
34. 释放 CLOSE;
35. 返回 Qi;

```

6 实验结果与分析

本节测试了为解决时空旅游规划而提出的各种算法的性能,实验在社交网络真实数据集上实现。

6.1 数据环境

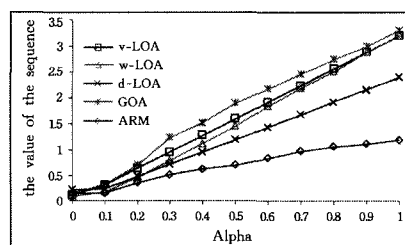
通过 Twitter, Flickr 等社交网络, 获取了北京 286 个景点、658 条边的真实数据集。搜集到的照片数目共计 4000000 张, 与这些照片相关的标签关键词大约 9100000 个。数据的维度包括照片的拍摄时间、地点、特色标签等信息。在实际应用中可以根据不同应用场景的数据定制合适的存储模式。在本系统中, 使用 XML 文件存储样例数据。在这个数据集中, 根据景点的旅游特色和游客的兴趣爱好选取了一些有趣的景点, 产生了多个兴趣景点数目、平衡因子、限定时间均不同的数据集。狄利克雷参数 β 设定为 0.01, 兴趣景点的总数目满足条件 $m \in [1, 20]$, 平衡因子满足条件 $\alpha \in [0, 1]$, 总的游览限定时间满足 $T_{total} \in [1h, 8h]$, 其中 h 表示采用的时间粒度是小时。简单起见, 只考虑一天的游览。旅游序列的价值和距离都在路网上计算。

6.2 性能评测

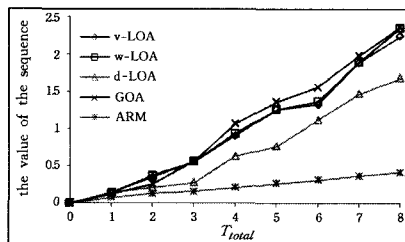
下面分别测试所提算法的性能。因为已有的相关研究工作都没有同时考虑景点特色、用户兴趣爱好和旅游时间限制

这 3 个因素, 所以他们没有可比性。为了证明所提算法的优越性, 本节将其与平均随机算法 (ARM) 的性能进行了比较。ARM 算法的思想是每次都随机选取景点序列, 并取 300 次的平均值。所有算法的性能值都是由多个用户提出的多个查询得出的平均值。同时还设计了一个用户投票界面, 用来搜集用户的满意度。统计了 2000 多个用户半年的数据量, 样本中包含了 1000 多个学生, 另外还有教师、IT 人员和其他行业人员。GUI 为用户提供一些交互问题, 如: 对所推荐景点集合的满意度, 对响应时间的满意度, 以及分别对 d-LOA, w-LOA, v-LOA 和 GOA 算法所推荐的时空旅游序列的满意度等。一律采用匿名方式进行统计, 并且多数统计结果集中在节假日和周末, 因此统计结果能够真实地体现用户的使用心得。

首先测试了 α 和 T_{total} 对时空旅游规划价值的影响。图 3(a) 给出了 $T_{total}=8h$ 时平衡因子 α 对 STS 价值的影响趋势。图 3(b) 给出了 $\alpha=0.7$ 时 T_{total} 对 STS 价值的影响趋势。在这两种情况下, GOA 算法的性能都优于 v-LOA, d-LOA, w-LOA 和 ARM, 并且 ARM 算法的价值最小。随着 α 和 T_{total} 的增加, 所有算法的性能都增加。算法 d-LOA 受景点之间相对位置的影响很大, 因为它只是贪婪地查找剩余景点中距离最近的 $|m/2|$ 个景点, 而没有考虑这些景点与目标景点 E 的方向是否相同, 所以随着 α 和 T_{total} 的增加, 算法 d-LOA 偏离正确方向的可能性也增加。图 3(b) 中算法 w-LOA 和 v-LOA 的趋势基本一致, 原因是当 $\alpha=0.7$ 时距离对整个序列的价值的的影响很小, 因此这两种算法的价值类似。



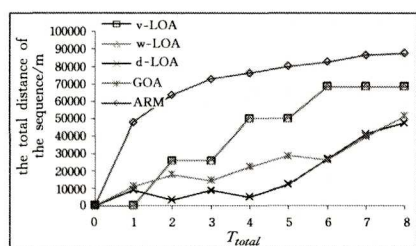
(a) 不同的 α 时价值的变化趋势



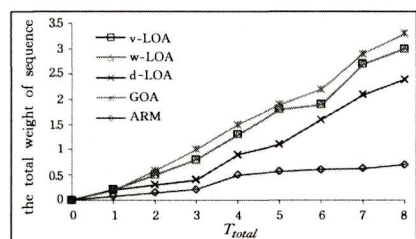
(b) 不同的 T_{total} 时价值的变化趋势

图 3 价值的变化趋势

图 4(a) 给出了 $\alpha=0.7$ 时 T_{total} 对路网上 STS 总距离的影响趋势。图 4(b) 给出了 $\alpha=0.7$ 时 T_{total} 对 STS 总权值的影响趋势。随着 T_{total} 的增加, 所有算法的距离和权值都增加。算法 d-LOA 的总距离小于其他 4 种算法的总距离, 因为它总是贪婪地查找剩余景点中距离最近的 $|m/2|$ 个景点。图 4(b) 中算法 GOA 的权值优于算法 d-LOA, w-LOA 和 v-LOA。ARM 算法的距离最大, 但权值最小。因此本文提出的 4 种算法都优于 ARM 算法。



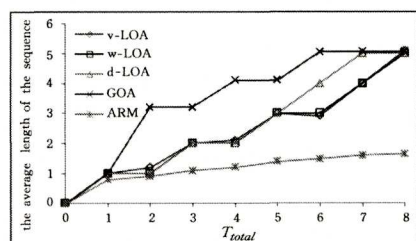
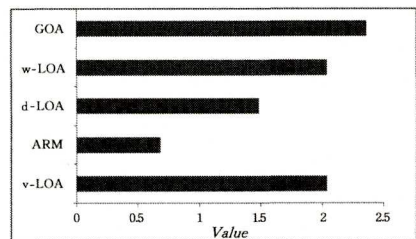
(a) 距离的变化趋势



(b) 权重的变化趋势

图4 T_{total} 对性能的影响

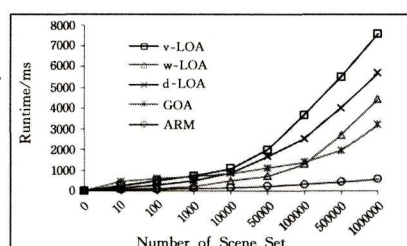
图5(a)给出了 $\alpha=0.7$ 时 T_{total} 对STS序列平均长度的影响趋势。通常,算法GOA比其他4种算法包含更多的景点个数,这是因为算法GOA可以得到路网上全局最优的旅游序列。而ARM算法中包含的景点个数最少。由图5(b)可知,GOA算法的价值是最大的,w-LOA和v-LOA的价值几乎相同,其变化情况取决于平衡因子 α 的选择。ARM算法的价值最小。

(a) 不同 T_{total} 时序列长度的变化趋势

(b) 不同算法的价值变化趋势

图5 不同算法的变化趋势

图6(a)给出了运行时间随景点数目的变化趋势。当景点数目大于50000时,算法v-LOA,d-LOA和w-LOA的运行时间快速增长。图6(b)给出了当景点数目等于10000时运行时间随不同 T_{total} 的变化趋势。在这两种情况下,随着景点数目和 T_{total} 的增加,所有算法的运行时间都会增加。其中算法v-LOA的运行时间最长,而算法ARM的运行时间最短。不同的是,图6(b)中算法运行时间的增加速度比图6(a)中的慢。



(a) 不同数目的景点集合

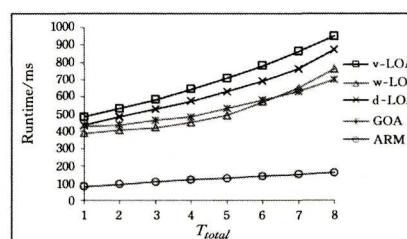
(b) 不同的 T_{total}

图6 运行时间的变化趋势

由上述测试可以得出如下结论:1)算法d-LOA的距离小于其他4种算法的距离,其适合于对距离更加关注的游客;2)算法w-LOA的权值大于其他4种算法的权值,其适合于对景点特色更加关注的游客;3)算法GOA的价值最大,当启发函数对到达目标的实际最小价值的估计不太高的情况下,该算法能够得到一个最优旅游序列。

图7给出了游客对所推荐景点集合的满意度。由图7可见,超过95%的游客对推荐结果是满意的,仅有5%的游客对结果不满意。这是因为在检测游客兴趣爱好时,不同特征的权值难以准确预测。响应时间的满意度大约为97%,如图8所示。图9给出了游客对所推荐的时空旅游序列的满意度。本文对时空旅游规划的4种不同算法都进行了测试。几乎每位游客对GOA的结果都很满意,极少数游客不满意的原因可能是他们对距离或者权值更为关注。由图9可见,算法d-LOA,w-LOA和v-LOA的满意度都低于GOA,也就是说,更多游客关注的是旅游序列的总价值。

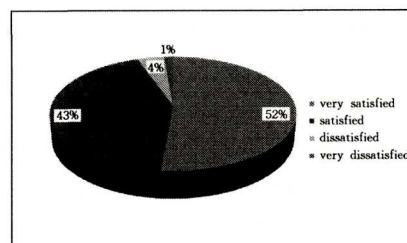


图7 游客对景点集合的满意度

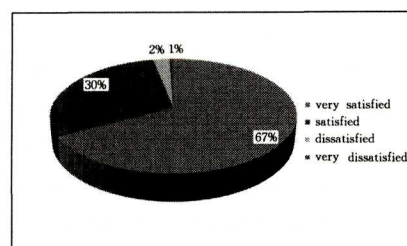


图8 游客对响应时间的满意度

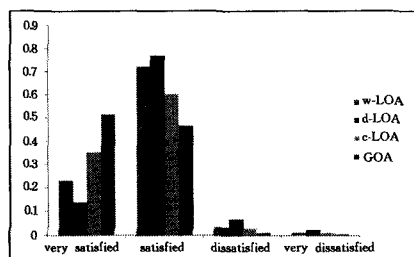


图9 游客对不同算法提供的时空旅游序列的满意度

结束语 本研究提出了时空旅游规划算法,该算法能够为游客提供最优的、个性化的旅游规划。在景点的选择和最优景点序列分析方面考虑了用户自身的兴趣爱好,具有一些新颖的特征,能够完成现有系统无法完成的功能,为用户提供高质量的服务。本文所提出的算法能够使用户在旅游时间范围限制内通过最少的路程来游览最想游览的景点。本研究提出的最优时空序列查询的框架结构具有通用性,所提出的技术算法也可以应用于其他领域。

参考文献

- [1] HALPIN H, ROBU V, SHEPHERD H. The Complex Dynamics of Collaborative Tagging[C]// The 16th International Conference on World Wide Web (WWW). 2007:211-220.
- [2] SOO V W, LIANG S H. Recommending a Trip Plan by Negotiation with a Software Travel Agent[J]. Cooperative Information Agents, 2001, 2182:32-37.
- [3] CASTILLO L, ARMENGOL E, ONAINDIA E, et al. Samap: An User-oriented Adaptive System for Planning Tourist Visits[J]. Expert Systems with Applications, 2008, 34(2):1318-1332.
- [4] YU C C, CHANG H P. Personalized Location-Based Recommendation Services for Tour Planning in Mobile Tourism Applications[J]. E-Commerce and Web Technologies, 2009, 5692:38-49.
- [5] NIARAKI A S, KIM K. Ontology based Personalized Route Planning System using a Multi-criteria Decision Making Approach[J]. Expert Systems with Applications, 2009, 36(2):2250-2259.
- [6] YOON H, ZHENG Y, XIE X, et al. Smart Itinerary Recommendation Based on User-Generated GPS Trajectories[C]// Proceedings of the International Conference on Ubiquitous Intelligence and Computing. 2010:19-34.
- [7] KISILEVICH S, KEIM D, ROKACH L. A Novel Approach to Mining Travel Sequences Using Collections of Geotagged Photos[C]// Proceedings of the 13th AGILE International Conference on Geographic Information Science. 2010:163-182.
- [8] LU X, WANG C, YANG J M, et al. Photo2Trip: Generating Travel Routes from Geo-Tagged Photos for Trip Planning[C]// Proceedings of the ACM International Conference on Multimedia. 2010:143-152.
- [9] YU J, ASLAM J A, KARAMAN S, et al. Optimal Tourist Problem and Anytime Planning of Trip Itineraries[OL]. <http://xxx.tall.ac.il/pdf/1409.8536v1>.
- [10] XIE M, LAKSHMANAN L V S, WOOD P T. IPS: An Interactive Package Configuration System for Trip Planning[J]. PV-LDB, 2013, 6(12):1362-1365.
- [11] JOSEPH R. Real-time Trip Planning with the Crowd[M]. CHI Extended Abstracts, 2013:2743-2748.
- [12] GEORGIOS A L, AMIR A S, POLVCHRONIS K. A New Location-Aware Calendar-Based Application for Dynamic Minimum Path Trip Planning[J]. Wireless Personal Communications (WPC), 2014, 78(1):29-44.
- [13] OHSAWA Y, HTOO H, WIN T N. Continuous Trip Route Planning Queries[M]// Book: Advances in Databases and Information Systems. Volume 9809 of the series Lecture Notes in Computer Science. Springer International Publishing. 2016:198-211.
- [14] XIE M, LAKSHMANAN L V S, WOOD P T. IPS: An Interactive Package Configuration System for Trip Planning[J]. PV-LDB, 2013, 6(12):1362-1365.
- [15] LEE C S, CHANG Y C, WANG M H. Ontological Recommendation Multi-Agent for Tainan City Travel[J]. Expert Systems with Applications, 2009, 36(3):6740-6753.
- [16] SOUFFRIAU W, VANSTEENWEGEN P. Tourist Trip Planning Functionalities: State-of-the-Art and Future[C]// Proceedings of the 10th International Conference on Web Engineering (ICWE). 2010:474-485.
- [17] FOSCA G, MIRCO N, DINO P, et al. Trajectory Pattern Mining[C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2007:330-339.
- [18] MAMOULIS N, et al. Indexing and Querying Historical Spatio-temporal Data[C]// Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2004:236-245.
- [19] GIRARDIN F, DAL F, BLAT J, et al. Understanding of Tourist Dynamics from Explicitly Disclosed Location Information[C]// Proceedings of the 4th International Symposium on LBS and Telecartography. 2007.
- [20] CHEN Z, SHEN H T, ZHOU X, et al. Searching Trajectories by Locations-An Efficiency Study[C]// Proceedings of the 36th SIGMOD International Conference on Management of Data (SIGMOD). 2010:255-266.
- [21] GEORGIOS A L, AMIR A S, POLVCHROIS K. A New Location-Aware Calendar-Based Application for Dynamic Minimum Path Trip Planning[J]. Wireless Personal Communications (WPC), 2014, 78(1):29-44.
- [22] YOON H, ZHENG Y, XIE X, et al. Social Itinerary Recommendation from User-generated Digital Trails[J]. Personal and Ubiquitous Computing, 2012, 16(5):469-484.
- [23] IGO R B, JOSE A F, FRANCO M N. Where Shall We Go To-

- day?: Planning Touristic Tours with Tripbuilder[C]// CIKM, 2013: 757-762.
- [24] JOSEPH R. Real-time Trip Planning with the Crowd[J]. CHI Extended Abstracts, 2013: 2743-2748.
- [25] ZHU C, XU J, LIU C, et al. Efficient Trip Planning for Maximizing User Satisfaction [C] // 20th International Conference (DASFAA 2015). Hanoi, Vietnam, Part 1, 2015: 260-276.
- [26] BORIS C. Improved Trip Planning by Learning from Travelers' Choices[C] // ICML Workshop Mining Urban Mobiltiy. At Lille, France, July, 2015: 17-26.
- [27] POPESCU A, GREFFENSTETTE G. Deducing Trip Related Information from Flickr[J]. International Conference on World Wide Web, 2009, 25(1): 1183-1184.
- [28] TAO Y, PAPADIAS D, SHEN Q. Continuous Nearest Neighbor Search[C] // Proceedings of 28th International Conference on Very Large Data Bases (VLDB). 2002: 287-298.
- (上接第 276 页)
- 曹文君, 薛善良. 基于物联网的城市路况监测系统[J]. 计算机与数字工程, 2016, 44(8): 325-424.
- [5] LIU Y, LIN Y, JIANG S, et al. Highway management information system based on Internet of Things[J]. Microelectronics & Computer, 2015, 32(1): 165-168. (in Chinese)
- 刘永, 林鹰, 蒋山, 等. 基于物联网的高速公路运行管理系统[J]. 微电子学与计算机, 2015, 32(1): 165-168.
- [6] CHAKRABORTY P S, TIWARI A, SINHA R. Adaptive and Optimized Emergency Vehicle Dispatching Algorithm for Intelligent Traffic Management System[J]. Procedia Computer Science, 2015, 57(9): 1384-1393.
- [7] ZHU X W, MENG X F. Optimizing Departure Interval for Bus Dispatching System Based on Comprehensive Objective Model [C]// Cota International Conference of Transportation Professionals. 2015: 1259-1268.
- [8] TANG J, SHI W Z, MENG L K. Time-Dependent Dynamic Vehicle Routing Based on Genetic Algorithm[J]. Geomatics and Information Science of Wuhan University, 2008, 33(8): 875-879. (in Chinese)
- 唐健, 史文中, 孟令奎. 基于遗传的算法的时相关动态车辆路径规划模型[J]. 武汉大学学报(信息科学版), 2008, 33(8): 875-879.
- [9] WANG J Q, ZHANG X. Modeling and Solving of Dynamic Vehicle Routing Problem in Complex Environment[J]. Journal of Wuhan University(Natural Science Edition), 2010, 56(4): 462-466. (in Chinese)
- 王江晴, 张潇. 复杂环境下动态车辆路径问题的建模与求解[J]. 武汉大学学报(理学版), 2010, 56(4): 462-466.
- [10] LIU F H, ZHAO J M, ZHU X Z. Research of Optimizing Physical Distribution Routing Based on Improved Genetic Algorithm [J]. Journal of Computer Technology and Development, 2016(3): 503-512. (in Chinese)
- 刘芳华, 赵建民, 朱信忠. 基于改进遗传算法的物流配送路径优化研究[J]. 计算机技术与发展, 2016(3): 503-512.
- [11] ZHANG T, LAI P Z, HE Q F, et al. Optimization of Dynamic Vehicle Routing of Urban Distribution Based on the Real-time Information[J]. Journal of Systems Engineering, 2015, 33(7): 58-64. (in Chinese)
- 张婷, 赖平忠, 何琴飞, 等. 基于实时信息的城市配送车辆动态路径优化[J]. 系统工程学报, 2015, 33(7): 58-64.
- [12] FENG L. Virtualization Scheduling Method for the Process of Coal Mining Equipment in Underground Coal Mine[J]. Computer Simulation, 2015, 32(4): 419-423. (in Chinese)
- 冯亮. 煤矿井下采煤设备工序虚拟化调度方法[J]. 计算机仿真, 2015, 32(4): 419-423.
- [13] ZHAO R, HU X, HE H D. E-commerce logistics distribution routing optimization considering customer satisfaction[J]. Journal of Shanghai Maritime University, 2015, 36(3): 64-70.
- [14] LI Z P, WU Z G. Study on the inventory routing problem of refined oil distribution based on working time equilibrium[J]. American Journal of Operations Research, 2016, 6(1): 17-24.
- [15] WANG S, BEI J. Real-time kinematic positioning algorithm of GPS/BDS[J]. Journal of Clinical Pathology, 2014, 63(2): 156-164.
- [16] ZHOU Y. Research on multi dimensional collaborative logistics management under the environment of Internet of things[D]. Changsha: Hunan University, 2013. (in Chinese)
- 周扬. 物联网环境下多维度协同物流管理研究[D]. 长沙: 湖南大学, 2013.
- [17] FENG L. Intelligent Logistics and Distribution System Based on Internet of Things[C]// IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC 2016). Xi'an, China, 2016.
- [18] ZHANG J Q, FANG W G. Improved Genetic Algorithm for Vehicle Routing Problem with Time Window[J]. Computer Engineering and Applications, 2010, 46(32): 228-231, 240. (in Chinese)
- 张建强, 方卫国. 有时间窗约束车辆路径问题的改进遗传算法[J]. 计算机工程与应用, 2010, 46(32): 228-231, 240.
- [19] MAHMOUDI M, ZHOU X. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows[J]. Transportation Research Part B Methodological, 2016, 89(2): 19-42.
- [20] YANG Z W, VAN OSTA J P, VAN VEEN B. Dynamic vehicle routing with time windows in theory and practice[J]. Natural Computing, 2017, 16(1): 119-134.