

# 一种预测群体用户访问行为的算法

刘清华,黄明和,唐权华,王 渊

(江西师范大学 软件学院,江西 南昌 330022)

**摘 要:**在电子商务发展中,商家需要理解用户访问网站的行为,为用户提供个性化服务,从而吸引用户购买商品。挖掘用户访问网站的行为是商家一个急需解决的问题,通过对 Web 日志进行挖掘是解决该问题的重要研究方法。提出了网页兴趣信息素的新概念,它是由页面相对浏览时间和点击率构建而成,利用兴趣信息素设计了基于蚁群算法的群体用户访问路径挖掘算法,根据挖掘结果预测用户访问行为。实验结果表明,兴趣信息素可以有效地预测用户的兴趣变化,能准确地反映用户访问模式,提高了预测群体用户访问行为的准确率。

**关键词:**访问行为;兴趣信息素;蚁群算法;Web 日志挖掘

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2014)02-0059-04

doi:10.3969/j.issn.1673-629X.2014.02.014

## An Algorithm of Predicting Users Access Behaviors

LIU Qing-hua, HUANG Ming-he, TANG Quan-hua, WANG Yuan

(School of Software, Jiangxi Normal University, Nanchang 330022, China)

**Abstract:** During the developing of e-commerce, enterprises should understand the behaviors of users browsing the Website, offer personal services to those that be leaded to buy their commodities. Mining user access behaviors is one problem that is badly in need of being solved by businessmen, while Web log mining suggests an important research method. In this paper, put forward a new concept of Web page interest pheromone. Both text relative browsing time and click rate define Web page interest pheromone. Group user access path mining algorithm is based on ant colony algorithm, interest pheromone being involved in designing. And according to mining results, user access behaviors are predicted. Experimental results show interest pheromone can efficiently forecast how user interests change, accurately reflect user access mode and improve the predicting accuracy of group user access behaviors.

**Key words:** access behaviors; interest pheromone; ant colony algorithm; Web log mining

## 0 引 言

近年来,电子商务网站发展速度快,竞争也异常激烈,商家使出各种营销手段吸引用户访问网站,但效果不佳。如何准确理解用户访问网站的行为是一个急需解决的问题,将根据用户访问网站的行为,为用户提供个性化服务,从而吸引用户购买商品。因此,建立预测用户访问网站模型具有重要意义。

建立预测用户访问网站模型的关键问题是需要对大量匿名用户访问网站的行为进行深层次地识别理解。网站的 Web 日志数据中隐含了匿名用户的各种个性化信息,对 Web 日志进行深层的挖掘是解决上述问题的有效途径<sup>[1-2]</sup>。因此,通过 Web 日志挖掘可以理解用户的访问行为,分析用户的偏好,预测用户下一次的访问行为,并主动为用户提供网站的导航服

务<sup>[3-4]</sup>。

一般基于 Web 日志文件分析用户行为使用 Markov 模型的随机方法、关联规则挖掘方法等<sup>[5]</sup>。这些方法具有一定的有效性,但没有考虑网站的层次结构,不能很好地动态反映用户兴趣的变化。蚁群算法是由 Dorigo 等人为解决旅行商问题而提出的“蚂蚁系统”,即基本蚁群算法<sup>[6]</sup>。它是受蚁群在觅食行为的启发而逐步发展起来的一种模拟蚂蚁群体智能行为的算法。作为群体智能算法的一个分支,在求解复杂优化问题方面具有优势。

由于用户访问网站的行为与蚁群觅食的行为非常相似,用蚁群来模拟用户,对 Web 日志进行处理、分析来挖掘用户访问模式<sup>[7-8]</sup>。

收稿日期:2013-04-19

修回日期:2013-07-25

网络出版时间:2013-11-29

基金项目:国家自然科学基金资助项目(6126037);江西师范大学青年成长基金项目(4499)

作者简介:刘清华(1984-),男,江西九江人,硕士,研究方向为 Web 数据挖掘。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20131129.1006.046.html>

## 1 蚁群算法优化模型

### 1.1 蚁群算法原理

蚁群在觅食过程中,蚂蚁之间通过释放激素传递信息,通过触觉感知外激素的浓度来选择觅食路径。外激素浓度高的路径,蚂蚁选择的概率越大,反之亦然。虽然单个蚂蚁的能力有限,但是许多蚂蚁组成一个群体却能够完成相当复杂的活动,科学研究表明,蚂蚁所到之处都会释放一种化学物质(外激素),外激素浓度越高就表明所走过的蚂蚁越多,通常这就是一条合理的觅食路线<sup>[9]</sup>。

在刚开始从蚁巢到食物源的所有路径上的蚂蚁数量基本相等,在这个时期蚂蚁的觅食路线是盲目的,它们在各条路径上所释放的外激素相差不大,后续蚂蚁选择路径的概率亦基本相等,如图 1 所示。随着时间推移,后续的蚂蚁会根据前面蚂蚁释放的外激素的浓度进行选择,很明显路径越短,外激素的浓度越高。到最后几乎所有蚂蚁都会选择最短路径寻找食物,只有极个别的蚂蚁会选择其他路径,如图 2 所示。

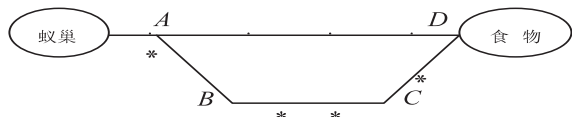


图 1 初始的蚂蚁分布图



图 2 后期的蚂蚁分布图

### 1.2 蚁群算法优化模型

蚂蚁的觅食过程实际上就是在寻找最短路径的过程,即在候选解的约束条件下寻求最优解决方案<sup>[10]</sup>。

定义 1:组合优化模型。一个组合最优问题 $(\bar{S}, \Omega, f)$ ,  $\bar{S}$  表示候选解集,  $\Omega$  表示可行候选解的约束条件的集合,  $f$  表示目标函数。

运用蚁群算法求解组合优化问题可视为:搜索  $s^* \in \bar{S}$ , 满足  $f(s^*) \leq f(s), \forall s \in \bar{S}$ 。记  $s$  中的组成元素来源于一个离散集  $V = \{v_1, v_2, \dots, v_n\}$ ,  $s = \{s_1, s_2, \dots, s_n\}$ ,  $\forall s_i \in V, i=1, 2, \dots, n$ 。蚁群算法的执行过程可以被看作蚂蚁在图  $G = (V, E, T)$  上的随机游走。离散集  $V$  为图  $G$  的节点集,  $E$  为边集,  $T$  为信息素向量, 其中  $\tau_{ij}$  对应节点  $v_i$  到  $v_j$  边  $(i, j)$  上的信息素。

设蚂蚁的数量为  $m$ , 用  $d_{ij}(i, j=1, 2, \dots, n)$  表示在图  $G$  中节点  $v_i$  和节点  $v_j$  之间的距离,  $\tau_{ij}(t)$  表示在  $t$  时刻节点  $v_i$  和节点  $v_j$  之间的路径上残留的信息素强度。用禁忌表  $\text{tabu}_k (k=1, 2, \dots, n)$  记录蚂蚁  $k$  当前所走过的节点, 节点集合随着  $\text{tabu}_k$  进化过程作动态调整。在搜索过程中, 蚂蚁根据各条路径上的信息量  $\tau_{ij}$  及路径

的启发信息来计算状态转移概率。 $P_{ij}^k(t)$  表示在  $t$  时刻蚂蚁  $k$  由节点  $v_i$  转移到节点  $v_j$  的状态转移概率, 其表达式为<sup>[2]</sup>:

$$P_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta}{\sum_{s \in \text{allowed}_k} [\tau_{is}(t)]^\alpha \cdot [\eta_{is}(t)]^\beta} & \text{If } j \in \text{allowed}_k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

其中,  $\text{allowed}_k = \{V - \text{tabu}_k\}$  表示允许蚂蚁  $k$  下一步选择的节点,  $\text{tabu}_k$  表示蚂蚁  $k$  已经走过的节点集合;  $\alpha$  为信息启发式因子;  $\beta$  表示期望启发式因子;  $\eta_{ij}(t)$  为启发函数, 其表达式为:

$$\eta_{ij}(t) = 1/d_{ij} \quad (2)$$

对于蚂蚁  $k$  而言,  $d_{ij}$  越小, 则  $\eta_{ij}(t)$  越大,  $P_{ij}^k(t)$  也越大, 显然, 该启发函数表示蚂蚁从节点  $v_i$  转移到节点  $v_j$  的期望程度。

在每只蚂蚁完成对图  $G$  中所有  $n$  个节点的遍历后, 要对残留信息进行更新处理。其表达式为:

$$\tau_{ij}(t+n) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) \quad (3)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t) \quad (4)$$

$$\Delta\tau_{ij}^k(t) = \begin{cases} \frac{Q}{L_k} & \text{若第 } k \text{ 只蚂蚁在经过路径}(v_i, v_j) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

式中,  $\rho$  为信息素挥发系数,  $1-\rho$  为信息素存留因子,  $0 \leq \rho < 1$ ;  $\Delta\tau_{ij}(t)$  表示此次循环中路径  $(v_i, v_j)$  上的信息素增量;  $\Delta\tau_{ij}^k(t)$  表示第  $k$  只蚂蚁在此次循环中留在路径  $(v_i, v_j)$  上的信息素;  $Q$  表示信息素强度, 它在一定程度上影响算法的收敛速度;  $L_k$  表示第  $k$  只蚂蚁此次循环所走路程的总长度。

## 2 用户访问行为与 TSP 相似

定义 2: 用户访问网站行为。用户访问网站行为是指用户对 Web 页面的浏览顺序。

定义 3: Web 模型。Web 模型可以抽象为有向图  $(V, E)$  来描述。其中  $V$  是页面的集合, 称为“节点集合”;  $E$  是所有超链接的集合, 称为“边集合”。有向边  $(V_i, V_j) \in E$  表明从页面  $V_i$  至页面  $V_j$  有一个超链接, 其中  $V_i, V_j \in V$ 。

用户访问 Web 站点时, 通常根据自身的兴趣爱好进行访问, 不同用户访问页面的顺序可能不同, 表现出个体随机性, 而对于特定的人群来说, 浏览网站的目的通常具有相应的针对性, 所以通过 Web 日志挖掘可以寻找出某种群体用户访问行为。

TSP (Travelling Salesman Problem) 是一个经典的

组合优化问题,属于 NP 完全问题,该问题描述如下:

假设有  $m$  个城市  $C = \{C_1, C_2, \dots, C_m\}$ , 城市之间相互连接, 城市  $C_i, C_j$  间的距离记为  $e_{ij}(i, j = 1, 2, \dots, m)$ , 求一条路径  $R = \{r_1, r_2, \dots, r_m\}$ , 它经过每个城市一次而且只经过一次, 并且它是所有满足此条件的路径的最小者。

$$E = \sum_{i=1}^{m-1} e_{c_{F_{i+1}}} \quad (6)$$

求解 TSP 要求每只蚂蚁以最短路径无重复地遍历每个城市后返回原出发地。蚁群算法成功解决了 TSP, 首先将蚂蚁随机地分布到某一城市, 然后根据转移概率公式选择信息素浓度最高的路径(信息素浓度高的路径距离相对较短), 随着时间的推移, 在较短的路径上, 蚂蚁所释放的信息素浓度会越来越高, 最终蚂蚁都聚集到最短的路径上。

用户访问某个站点页面的过程就是对该页面所链接的其他页面的一个随机选择过程, 通常用户在浏览某个页面时, 可能被当前页面上链接的其他页面所吸引。访问最多的页面当然会吸引更多的用户来浏览, 这与 TSP 的根本原理是相同的, 蚂蚁在寻找食物过程中会选择信息素浓度高的路径, 路径越短, 则释放的信息素的浓度就越多, 即用户访问网站的行为与蚂蚁觅食的行为是非常相似的。

### 3 用户访问行为建模与预测

#### 3.1 蚁群访问模型

定义 4: 蚁群访问模型。蚁群访问模型用一个有向图  $G(V, E, W)$  表示。其中,  $V$  是节点的集合, 每一个节点代表一个页面;  $E$  是边的集合, 边代表页面之间的链接;  $W$  是边上权值的集合, 代表网页的兴趣度。

用户访问某个页面时, 可能会被当前页面上的其他链接所吸引而点击查看。一个正在访问站点的用户假设是一个蚂蚁个体, 在选择页面时随机选择链接。越令人感兴趣的链接将更能吸引蚂蚁个体, 并且蚂蚁个体在选择页面上停留的时间越长, 表示蚂蚁个体所代表的用户对此页面的兴趣越浓, 在此链接上释放的信息素越多<sup>[11-12]</sup>。文中使用访问网页的相对浏览时间和点击率为参数来定义网页兴趣度。

定义 5: 页面点击率。

$$\text{ClickRatio}(i) = \frac{\text{visits}(i)}{\sum_{i \in \text{tabu}_k} \text{visits}(i)} \quad (7)$$

其中,  $\text{visits}(i)$  表示网页  $i$  的访问数;  $\text{tabu}_k$  表示蚂蚁  $k$  已访问过的网页集合。

定义 6: 页面相对浏览时间。

$$\text{Time}(i) = \frac{\text{TotalTime}(i)}{\text{Average}_{i \in \text{tabu}_k} \text{TotalTime}(i)} \quad (8)$$

其中,  $\text{TotalTime}(i)$  表示网页  $i$  的浏览时间;  $\text{tabu}_k$  表示蚂蚁  $k$  已访问过的网页集合;  $\text{Average}$  表示所访问过的页面平均浏览时间。

定义 7: 网页兴趣度。

$$\text{Interest}(i) = x \text{ClickRatio}(i) + y \text{Time}(i) \quad (9)$$

其中,  $\text{Interest}(i)$  表示页面  $i$  的兴趣度;  $x$  是浏览次数在网页兴趣度计算中所占的权重,  $x$  越大, 表示浏览次数在网页兴趣度计算中越重要;  $y$  是停留时间在信息量计算中所占的权重,  $y$  越大, 表示停留时间在网页兴趣度中越重要。  $x + y = 1 (x > 0, y > 0)$ ,  $x, y$  各取 0.5。

定义 8: 选择偏爱度。选择偏爱度  $\eta_{ij}$  表示用户访问从页面  $i$  到页面  $j$  的路径的频率。定义如下:

$$\eta_{ij} = \frac{C_{ij}}{\sum_{k=1}^n C_{ik}} \quad (10)$$

式中,  $C_{ij}$  表示访问节点  $i$  到  $j$  的路径次数, 从节点  $i$  到所有的下一节点有  $n$  种不同的选择。

定义 9: 兴趣信息素。函数  $\tau_{ij}(t)$  表示用户对从节点  $i$  到节点  $j$  的路径的访问兴趣。其兴趣信息素函数定义如下:

$$\tau_{ij}(t+1) = (1 - \rho) \tau_{ij}(t) + \Delta \tau_{ij} \quad (11)$$

$$\Delta \tau_{ij} = Q * \text{Interest}(j) \quad (12)$$

式(11)中,  $\rho$  为兴趣信息素挥发系数,  $1 - \rho$  为兴趣信息素存留因子,  $0 \leq \rho < 1$ 。式(12)中,  $Q$  为常数;  $\Delta \tau_{ij}$  为兴趣信息素增量, 它的值依赖于  $\text{Interest}(j)$ ;  $\text{Interest}(j)$  表示路径  $i \rightarrow j$  的兴趣度。

定义 10: 路径选择函数。路径选择函数  $p_{ij}(t)$  表示用户访问从节点  $i$  到节点  $j$  路径的偏好, 定义如下:

$$p_{ij}(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{k=1}^n [\tau_{ik}]^\alpha \cdot [\eta_{ik}]^\beta} \quad (13)$$

#### 3.2 预测群体用户访问行为的算法

首先, 对 Web 日志进行数据预处理, 按照数据清洗、数据降维、用户识别、会话识别、事物识别、构建蚁群访问 Web 模型等步骤进行, 数据预处理流程如图 3 所示。

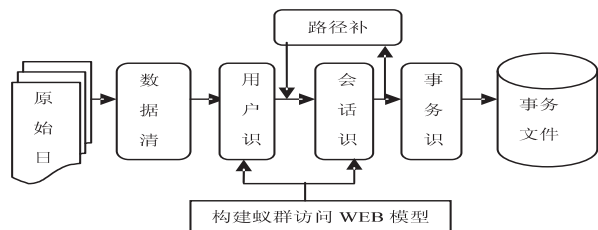


图3 Web日志数据预处理流程

其次, 利用蚁群算法对蚁群访问 Web 模型进行挖掘, 得出用户访问路径。设  $m$  表示用户的数量,  $n$  表示页面的数量,  $N_{\max}$  表示最大迭代次数。下面给出该算法的基本步骤:



Step1: 参数初始化操作, 将  $\rho, \alpha, \beta, m, n, Q, C, N_c, N_{cmax}$  等参数进行初始化;

Step2: 按照(9) 式计算各页面兴趣度, 按(12) 式设置各路径上的初始兴趣信息素;

Step3: 将  $m$  只蚂蚁随机放入到  $n$  个页面中;

Step4: 每只蚂蚁分别进行循环操作, 即 for( $k=1; k \leq m; k++$ ), 搜索路径; 随机选择每只搜索蚂蚁的初始位置, 并将该位置放入每只蚂蚁对应的  $tabu_k$  表中;

Step5: 每只蚂蚁按式(10) 计算下一页面选择偏爱度, 然后按式(13) 选择下一个页面  $j$ , 并将页面  $j$  置于禁忌表  $tabu_k$  中;

Step6: 若禁忌表  $tabu_k$  未滿, 则转至 Step5 继续;

Step7: 若禁忌表滿时, 记录第  $k$  只蚂蚁此次遍历的路径, 并置空禁忌表;

Step8: 对蚂蚁一次遍历完毕的路径上的信息素按(11), (12) 式进行全局更新;

Step9: 如果  $N_c \leq N_{cmax}$ , 则  $N_c = N_{c+1}$ , 跳转至 Step4 继续进行, 否则, 跳转到步骤 10;

Step10: 每个蚂蚁个体完成了若干次的迭代后, 得出蚂蚁访问较集中的路径; 这些集中访问的路径即为群体用户访问模式;

Step11: 输出群体用户访问模式, 算法终止。

最后, 应用群体用户访问模式, 预测用户访问行为。经过文中设计的蚁群算法处理 Web 日志后, 得出群体用户访问模式。将访问模式路径长度为 3 及以上存储到用户访问行为模式库中。当用户在访问网站时, 识别当前用户所访问的页面, 若与访问模式前 3 个页面相同, 则对用户访问行为进行预测, 向用户推荐访问模式中的第 4 个页面, 实现个性化服务。

## 4 实验结果及分析

该项目实验数据来自山货吧电子商务网站的服务器, 数据集大小为 500 MB 左右, 该 Web 日志数据是用用户一个星期访问该网站的记录<sup>[13]</sup>。按上述步骤对 Web 日志进行数据清洗、数据降维、用户识别、会话识别、事物识别等步骤处理, 将数据集分为训练集和测试集, 其中训练集包含 90 132 条记录, 测试集包含 36 238 条记录。为了提高预测用户访问行为的正确率, 在会话识别阶段将长度小于 3 的路径删除。实验所用计算机处理器为 Intel Core(TM)2 Duo, 内存 2 G, 测试程序使用 JAVA 语言编写, 在 Windows XP 操作系统上运行。在实验中, 发现选择概率  $P$  的值对实验结果有重要影响。如果选择概率  $P$  取较高的值, 则产生很少的浏览路径, 如果  $P$  取较低的值, 则产生用户不感兴趣的路径。将  $P$  设为 0.3,  $\rho$  设为 0.8,  $Q$  为 0.5。

为了验证兴趣信息素对群体用户访问模式挖掘结

果的影响, 实验对  $\alpha, \beta$  等参数进行设置, 将  $\alpha$  设为 0 表示与兴趣信息素无关, 设  $\alpha$  为 1 表示该算法使用了兴趣信息素来进行模式挖掘; 将  $\beta$  设为 1 表示该算法考虑了选择偏爱度。因为实验主要对兴趣信息素进行验证, 故分 2 组实验进行对比, 即  $\alpha=0, \beta=1$  为一组;  $\alpha=1, \beta=1$  为另外一组, 经过实验得出结果如图 4 所示。

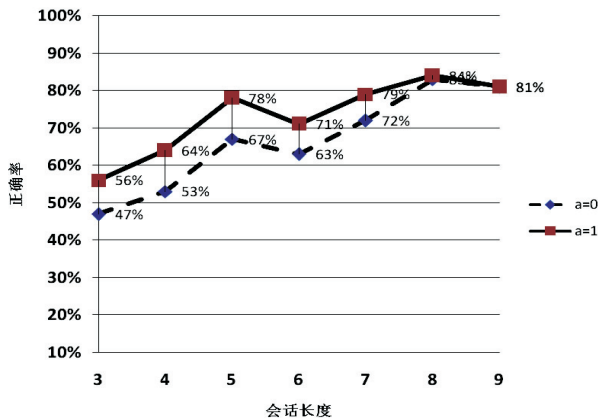


图 4 两种情况下的准确度比较

在蚁群算法中通过兴趣信息素和选择偏爱度的使用能够提高偏好路径的准确率, 可以更有效地预测用户的兴趣变化, 能够对预测群体用户访问行为的结果产生好的影响。

## 5 结束语

文中利用兴趣信息素设计了基于蚁群算法的挖掘群体用户访问网页路径算法, 实验结果表明该算法是切实可行的, 能准确地反映用户访问模式, 从而预测用户访问行为。此算法存在不足之处是只考虑了用户在每个页面的浏览次数和访问时间所反映出的用户兴趣程度, 没有考虑到用户对页面的内容的兴趣度, 因此仍然没有完全反映出用户的访问兴趣。

## 参考文献:

- [1] Mulvenna M D, Anand S S, Buchner A G. Personalization on the net using web mining: Introduction[J]. Comm of ACM, 2000, 43: 122-125.
- [2] Shanhab C, Banaei-Kashnai F. Efficient and anonymous web -usage mining for web personalization[J]. INFORMS journal on computing, 2003, 15(2): 123-147.
- [3] Saxena D K, Shukla M R. Significant interval and frequent pattern discovery in web log data[J]. International journal of computer science issues, 2010, 7(1): 394-399.
- [4] Mobasher B. Discovery and evaluation of aggregate usage profiles for web personalization[J]. Data mining and knowledge discovery, 2002, 6(1): 61-82.
- [5] Shahabi C, Kashani F B. A framework for efficient and anonymous web usage mining based on client-side tracking[C]//

记忆遗传算法计算个体适应度的时间消耗对比。

从表 2 中可以看出,相比于基本遗传算法,自适应记忆遗传算法在计算种群中个体适应度值方面消耗的时间更少。

表 2 解 pr76. tsp 时基本遗传算法和自适应记忆遗传算法计算个体适应度的时间消耗(s)

不同遗传算法	Popsize = 50	Popsize = 100	Popsize = 150
SGA	5.694 59	10.422 68	15.284 91
AMGA (GWS=0.1 * Popsize)	2.286 08	4.654 93	7.790 41
AMGA (GWS=0.2 * Popsize)	2.300 86	5.020 27	8.963 41
AMGA (GWS=0.3 * Popsize)	2.358 66	5.394 06	10.240 41
AMGA (GWS=0.4 * Popsize)	2.394 56	5.696 74	10.532 38
AMGA (GWS=0.5 * Popsize)	2.464 53	6.133 74	11.503 58
AMGA (GWS=1.0 * Popsize)	3.645 79	8.688 44	17.170 02

3.2 算法收敛性测试

以 p76 为 TSP 测试实例,对文中改进后的算法的收敛性能进行测试。测试实例进行 10 次重复测试实验,其收敛代数与最短路径长度的测试结果如表 3,不同算法求得的最优解与 TSPLIB 提供的最优解的平均相对误差如表 4 所示。

表 3 不同算法求解实例 pr76. tsp 的性能对比

不同算法	平均收敛代数	平均最短路径长度	最短路径长度	实际最短路径长度
SGA	2 067.9	189 815.138 8	179 497.362 5	108 159.438 3
Srinivas AGA	1 123.6	136 544.963 1	128 350.335 0	108 159.438 3
AMGA	1 521.7	113 108.868 0	109 679.316 4	108 159.438 3

表 4 不同算法求解的最优解与 TSPLIB 提供的最优解的平均相对误差

不同算法	bayg29. tsp/%	pr76. tsp/%	ch150. tsp/%
SGA	11.75	75.50	82.53
Srinivas AGA	4.66	26.24	21.91
AMGA	1.00	4.58	9.38

上述实验测试数据表明:文中的算法的收敛结果明显好于其他两种算法的收敛结果,最终收敛结果也接近于实际最优解,与 TSPLIB 提供的最优解的平均相对误差最大不超过 9.38%;其次,文中的算法具有较快的收敛速度。

(上接第 62 页)

Proc of WEB KDD 2001: Mining web log data across all customer touch points. [s.l.]: Springer-Verlag, 2002: 113-114.

[6] Dorigo M, Caro G D, Gamardella L M. Ant algorithm for discrete optimization[J]. Artificial life, 1999, 5(3): 137-172.

[7] 鲍 钰. 基于 Web 日志的个性化搜索引擎模型的发现[J]. 计算机应用研究, 2009, 26(5): 1806-1809.

[8] 郭 岩, 白 硕, 杨志峰, 等. 网络日志规模分析和用户兴趣挖掘[J]. 计算机学报, 2005, 28(9): 1483-1496.

[9] 马溪骏, 凌海峰, 刘业政, 等. 基于蚁群算法的群体用户兴

4 结束语

文中提出了一种自适应记忆遗传算法, 分别通过引入基因库和利用 Logistic 曲线方程对调节公式进行调整等策略对遗传算法进行改进, 经实验验证文中算法能够有效减少算法的时间复杂度且收敛速度较快。

参考文献:

[1] Holland J H. Adaptation in natural and artificial systems[M]. Ann Arbor: University of Michigan Press, 1975.

[2] 金 芬. 遗传算法在函数优化中的应用研究[D]. 苏州: 苏州大学, 2008.

[3] Sharapov R R, Lapshin A V. Convergence of genetic algorithms[J]. Pattern recognition and image analysis, 2006, 16(3): 392-397.

[4] 黄樟灿, 李 炜. 有界区域上多峰函数全局优化问题的改进演化算法[J]. 武汉大学学报(理学版), 2007, 53(1): 55-58.

[5] 陈一虎, 刘淳安. 基于实数编码的遗传算法收敛性研究[J]. 西南民族大学学报(自然科学版), 2006, 32(4): 666-669.

[6] 徐宗本, 聂赞坎, 张文修. 遗传算法的几乎必然强收敛性-鞅方法[J]. 计算机学报, 2002, 25(8): 785-793.

[7] Syswerda G, Schaffer J D. Uniform crossover in genetic algorithms[C]//Proceeding of 3rd international conference on genetic algorithms. USA: Morgan Kaufmann, 1989: 2-9.

[8] Kennedy J, Eberhart R C. Particle swarm optimization[C]//Proceeding of IEEE international conference on neural networks. USA: NJ, 1995: 1942-1948.

[9] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithm[J]. IEEE transaction on systems man and cybernetics, 1994, 24(4): 656-667.

[10] 任子武, 伞 冶. 自适应遗传算法的改进及在系统辨识中的应用研究[J]. 系统仿真学报, 2006, 18(1): 41-43.

[11] 殷祚云. Logistic 曲线拟合方法研究[J]. 数理统计与管理, 2002, 21(1): 41-46.

[12] Applexgate D, Bixby R. Implementing the Danting-Fulkerman-Johnson algorithm for large traveling salesman problems[J]. Mathematical programming, 2003, 97(1-2): 91-98.

趣导航路径发现[J]. 中国管理科学, 2006, 14(3): 56-59.

[10] 凌海峰. 基于 ACO 的 Web 使用挖掘方法研究[D]. 合肥: 合肥工业大学, 2009.

[11] 霍晓钢. 由 Web 日志确定页面访问时间的方法研究[J]. 计算机工程与设计, 2010, 31(10): 2416-2419.

[12] 王有为, 张健斌. 一种新的层次结构网站用户兴趣模式变化识别算法[J]. 系统工程理论与实践, 2008, 28(10): 89-95.

[13] 刘清华. 基于 ACO 的 Web 日志挖掘研究[D]. 南昌: 江西师范大学, 2011.

# 一种预测群体用户访问行为的算法

作者：[刘清华](#)，[黄明和](#)，[唐权华](#)，[王渊](#)，[LIU Qing-hua](#)，[HUANG Ming-he](#)，[TANG Quan-hua](#)，[WANG Yuan](#)

作者单位：[江西师范大学 软件学院, 江西 南昌, 330022](#)

刊名：[计算机技术与发展](#)

ISTIC

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2014(2)

本文链接：[http://d.g.wanfangdata.com.cn/Periodical\\_wjfz201402015.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjfz201402015.aspx)