

基于上下文感知和移动位置预测的切换机制

王玉祥 王进 谢胜东

(南京信息工程大学江苏省网络监控中心 南京 210044) (南京信息工程大学计算机与软件学院 南京 210044)

摘要 切换机制是无线通信系统中为用户提供无缝移动应用服务的一项关键技术,切换成功与否直接影响着用户业务体验和使用。提出了基于上下文感知和移动位置预测的切换机制,即采用多种类型上下文信息并利用修正证据理论预测移动用户的准确位置。提出的方法能够避免“乒乓切换”,缩短切换延迟,提前预留资源,为快速、准确和可靠的切换判决提供依据,显著地改善了网络服务质量。理论分析和仿真结果表明,该方法位置预测更接近用户真实的运动轨迹,位置预测的均方根误差较小,大大提高了切换成功率。

关键词 切换机制,上下文,证据理论,地理信息系统 GIS(Geographic information system), Hausdorff 距离
中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.09.023

Handoff Mechanism Based on Context-aware and Mobile Location Prediction

WANG Yu-xiang WANG Jin XIE Sheng-dong

(Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing 210044, China)
(School of Computer and Software, Nanjing University of Information & Science, Nanjing 210044, China)

Abstract The handoff mechanism is a key technology which can provide user with seamless mobile application services in wireless communication system. It also has very important impact on user oriented service experience and application. In this paper, a context-aware and mobile location prediction based handoff mechanism was proposed, which uses the modified Dempster-Shafer theory together with a variety of context information to predict the location of mobile users. The proposed method can not only efficiently avoid "ping-pong" handoff, but also shorten handoff delay, and provide basis for fast, accurate and reliable decisions, and reserve resources in advance. Thus, it can significantly improve the QoS (Quality of Service) of network. Theoretical analysis and simulation results show that location prediction is closer to user's real trajectory. The RMSE of position is smaller, which means the success rate of handoff is largely improved.

Keywords Handoff mechanism, Context, Dempster-Shafer (D-S) theory, GIS (Geographic information system), Hausdorff distance

1 引言

随着无线通信技术的快速发展,功能强大的便携式终端以及多媒体终端被广泛地使用,为普适计算实现任何人在任何时间、任何地点以任何方式均可以进行通信的目标奠定了坚实基础。但是用户移动的不确定性,以及用户所处物理环境复杂多变性,使得无线网络服务质量 QoS(Quality of Service)保障面临极大挑战。切换机制成为移动通信系统的必要条件,是为用户提供无缝移动服务应用的一项关键技术,在移动通信发展中切换技术起着举足轻重的作用。切换机制是为保证移动用户通信服务质量或保持网络负载均衡,而将用户当前的通信信道切换到其他基站的过程。切换是移动通信系统必须具备的一项基本功能,也是移动通信系统的极为复杂和关键的技术。切换机制能够保障通信服务连续性,为用户提供持续稳定的服务,切换不但对单用户的 QoS 产生影

响,同时也会影响全网的性能,如系统响应速度、容量等。因此一个网络的切换性能,不但对切换本身的性能,而且对业务的影响也有着重要的意义,是影响全网通信质量的重要性能指标。

目前对切换机制出现诸多研究,比如李鹏^[1]等通过预测终端信号强度判断终端在不同基站之间的切换,把信号强度作为切换判决的唯一依据,在高建筑物阴影下容易导致“乒乓切换”;罗涛^[2]等综合考虑终端当前位置、运动方向和速度矢量等因素进行动态速率切换,取得了一定效率,可是当终端运动方向发生改变时,可能需要重切换;Zigang Yang^[3]等运用序列蒙特卡罗滤波(Sequential Monte Carlo Filter)预测终端位置并辅助终端进行切换,设计了状态转移和粒子权重调整方程,在每次位置预测的时间间隔很短时,此方法有较好的切换准确率,然而如果采样间隔较长或终端运动轨迹不确定时,可能会出现位置预测失败;N. Marmasse^[4]等建立马尔可夫模

到稿日期:2013-11-02 返修日期:2014-02-07 本文受南京信息工程大学科研基金(20100393),江苏高校优势学科建设工程项目,国家自然科学基金(60802034,41071253)资助。

王玉祥 男,博士,讲师,主要研究方向为网络与通信, E-mail: wangyx@nuist.edu.cn; 王进 男,教授,主要研究方向为物联网技术与应用; 谢胜东 男,副教授,主要研究方向为无线传感网络。

馆等等,可以形式化表示为 $B_{GIS} = \{B_1, B_2, \dots, B_n\}$, 其中 $B_i (i = 1, 2, \dots, n)$ 表示用户所处某个位置即用户所在的标志建筑物; 然后选择用户到达预测位置所经过的街道(Street)比如北京西路、中山北路等等, 这些街道可以形式化表示为 $S_{GIS} = \{S_1, S_2, \dots, S_m\}$, 其中 $S_i (i = 1, 2, \dots, m)$ 表示用户到达预测位置所经过的某一街道; 最后每条街道 $S_i (i = 1, 2, \dots, m)$ 可细分为相互连接的 k 条路径小段 $S_i = \{\alpha_1^i, \alpha_2^i, \dots, \alpha_k^i\}$, 其中 $\alpha_j^i (j = 1, 2, \dots, k)$ 细分的原则是:

- 1) 与 S_i 邻接的标志建筑物单独划分为一段 α_j^i 。
- 2) 各 S_i 之间的交叉单独划分为一段 α_j^i 。
- 3) 基站边界单独划分为一段 α_j^i , 即一段 α_j^i 不能出现不同的两个 $cell_id$ 。

南京鼓楼一段区域 GIS 信息化表示如图 2(b) 所示。

文中移动位置预测主要是指根据用户上下文信息(比如偏好、日程表等)预测用户所有可能将要到达的位置, 即到达的标志建筑物的路径序列。因此, 首先定义标志建筑物为自变量函数 $F(B_i)$, 其值域为标志建筑物属性(attribute)集合, 例如若标志建筑物 B_i 为购物商场, 则 $F(B_i) = \{a_1^i, a_2^i, \dots, a_l^i\}$, 其中 $a_1^i =$ 商场, $a_2^i =$ 法国, $a_3^i =$ 价格低廉; 若 B_i 为图书馆, 则 $F(B_i) = \{a_1^i, a_2^i, \dots, a_l^i\}$, 其中 $a_1^i =$ 图书馆, $a_2^i =$ 中国, $a_3^i =$ 免费等等。更进一步对于高层建筑物 B_i 可以细分为楼层 B_h , 其中 h 表示楼层, 每个楼层具有不同属性 $F(B_h) = \{a_1^h, a_2^h, \dots, a_l^h\}$ 。因此, 引入标志建筑物楼层参数, 空间地理信息变成了三维空间。

为根据路径信息预测用户路径序列, 文中引入以路径小段 α_i, α_j 为行列的矩阵 R , 如图 3 所示。矩阵 R 表示形式化 GIS 信息, 矩阵 R 中元素 r_{ij} 可以表示为:

$$r_{ij} = \begin{cases} \text{相邻 } \alpha_i \text{ 到 } \alpha_j \text{ 的方向值,} & i \neq j \\ \alpha_i \text{ 相邻建筑物 } B_i \text{ 的属性信息,} & i = j \end{cases}$$

其中, 当 $i \neq j$ 时表示相邻路径小段 α_i 到 α_j 的方向值, 其取值范围可以是 0° 到 360° , 为简单起见使用东(E)南(S)西(W)北(N)方向值表示, 由于本文引入标志建筑物楼层参数, 因此添加上(U)下(D)两个方向值, 构成了 GIS 三维空间信息模型, 如图 3 所示; 当 $i = j$ 时表示与道路小段 α_i 邻接的标志建筑物 B_i 的属性信息 $F(B_i) = \{a_1^i, a_2^i, \dots, a_l^i\}$ 。对于多层建筑物 B_h , 其属性信息为 $F(B_h) = \{a_1^h, a_2^h, \dots, a_l^h\}$, 此时矩阵 R 为嵌套矩阵 R^* 。另外, 由于 r_{ij} 表示相邻路径小段的邻接、方向以及标志建筑物属性等信息, 对于不相邻的路径小段矩阵元素为空, 因此, R^* 为稀疏矩阵, 如图 4 所示。

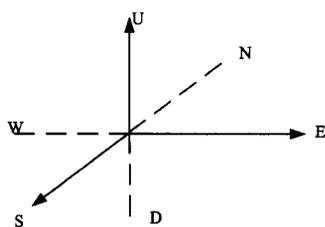


图 3 GIS 三维空间信息模型图

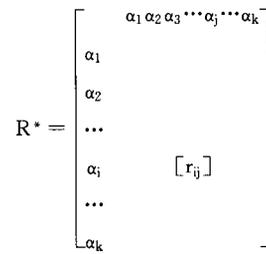


图 4 GIS 信息矩阵

3.2 推理引擎

本节在上述各类上下文信息形式化表示基础上, 利用修正证据理论 Dempster-Shafer (D-S) 对移动用户位置进行预测。D-S 理论是 Dempster 在 1968 年提出的, 他的学生 Shafer^[14] 在 1971 年进行了发展。D-S 理论将所研究的问题划分到相互排斥且穷尽的辨识框架 Θ 中, 对于任何一个属于辨识框架 Θ 的子集 A , 如果满足 $m(\Phi) = 0$ 且 $\sum_{A \subseteq \Theta} m(A) = 1$, 则称 m 为 Θ 上的基本概率分配 BPA (Basic Probability Assignment) 函数, 又称 mass 函数。为了表示对命题的信任程度, D-S 理论引入信度函数 Bel 和似然函数 Pls 的概念, 它与 BPA 的关系满足 $Bel(A) = \sum_{B \subseteq A} m(B)$, $Pls(A) = \sum_{A \cap B \neq \Phi} m(B)$, 其中 B 为辨识框架 Θ 的子集, Bel 和 Pls 分别为假设 A 的信度函数和似然函数, 区间 $[Bel, Pls]$ 表示对该假设的确信程度。

为了合成不同来源的证据, D-S 理论提出了 Dempster 合成规则。设 Bel_1 和 Bel_2 是同一辨识框架 Θ 上的两个信度函数, m_1, m_2 分别是相对应的基本可信度分配, A_i 和 B_j 为焦点, 则两个信度函数的合成法则为:

$$m(A) = m_1 \oplus m_2 = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \Phi} m_1(A_i) m_2(B_j)} \quad (1)$$

$m(A)$ 反映了 m_1 和 m_2 对应的两个证据对命题 A 的联合支持程度, $k = \frac{\sum_{A_i \cap B_j = \Phi} m_1(A_i) m_2(B_j)}{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}$ 是冲突因子, 反映了证据之间的冲突程度。当 $k = 0$ 时, 称为完全不冲突; 当 $0 < k < 1$ 时, 称为非完全冲突; $k = 1$ 时, 称为完全冲突。

3.2.1 证据生成

证据生成就是利用证据理论并结合上述采集的上下文信息分配假设和生成一系列证据来预测移动用户未来可能位置。

首先构造辨识框架 Θ 表示用户未来所有可能位置, 即用户所处所有可能标志建筑物, $\Theta = \{B_h : B_h \in B_{GIS}\}$, B_{GIS} 表示 GIS 中用户目标位置。 Θ 的子集 2^Θ 表示用户可能位置的不同假设, 每个假设可以利用 mass 函数分配基本信度值, 利用不同上下文信息生成证据来预测用户未来目标位置。文中给出基于两类上下文信息的一系列不同证据来支持相应的假设分配。

1) 基于用户偏好(Preference)上下文证据

依据用户偏好、当前位置和日程表信息与可能位置 $F(B_h) = \{a_1^h, a_2^h, \dots, a_l^h\}$ 属性的匹配度生成一系列证据。假设任一用户的偏好 P 有 n 个属性, 对于 n 个假设 H 生成的证据为 E_p 。对于任何一假设 $H_i (i = 1, 2, \dots, n)$ 表示一组用户可能位置 $B_h \in \Theta$, B_h 的属性 $F(B_h) = \{a_1^h, a_2^h, \dots, a_l^h\}$ 与用

户偏好 P 的属性具有一定的匹配度。由此, $H_i = \{B_{jh}; B_{jh} \in \Theta | a_{jh}^i \in F(B_{jh})\}$, 令

$$m(H_i) = \frac{1}{n} \quad (2)$$

其中, $m(H_i)$ 表示假设 H_i 的基本概率分配函数。

2) 基于用户日程表上下文证据

用户日程表上下文信息是用户移动位置预测的另一个有力证据。假设 B_{yh} 表示日程表中设定的用户将来到达位置, t_y 是用户未到达设定位置 B_{yh} 所剩余时间, 由此, 可以生成 n 个假设证据 E_y , 每一个假设 H_i 代表一组可能的位置 B_{jh} , 则有 $H_i = \{B_{jh}; B_{jh} \in \Theta | (t(B_{zh}, B_{jh}) + t(B_{jh}, B_{yh})) \leq i \times \frac{t_y}{n}\}$, 其中, B_{zh} 为用户当前位置, $t(B_{zh}, B_{jh})$ 为用户从当前位置 B_{zh} 移动到目标位置 B_{jh} , 令

$$m(H_i) = \frac{1}{n} \quad (3)$$

3.2.2 证据合成

基本信度分配 m 是依据现有证据对求解问题的主观断定, 获取 m 值后, 可以对不同来源证据进行合成。经典证据组合式(1)是所有参与合成的证据都具有相同的重要性, 但在普适计算环境下, 不同证据对要解决问题的重要性与可靠性是不同的, 因此在证据合成时需要证据的基本可信度进行加权修正^[15], 以反映证据不同的重要性与可靠性, 修正后的信度公式见式(4)。

$$m^\beta(A_i) = \begin{cases} (1-\beta)m(A_i), & C \neq \Theta \\ \beta + (1-\beta)m(\Theta), & \text{其他} \end{cases} \quad (4)$$

其中, 证据重要性因子 β 的值域范围为 $[0, 1]$, 当 $\beta=0$ 时表示该证据对证据合成没有什么影响, 当 $\beta=1$ 时表示该证据对证据合成有重要影响。

因此, 本文提出利用修正证据理论即加权证据组合规则式(5)融合不同来源证据, 公式如下:

$$m(A) = m_1 \oplus m_2(A) = \frac{\sum_{A_i \cap B_j = A} w_1 m_1^{\beta}(A_i) w_2 m_2^{\beta}(B_j)}{1 - \sum_{A_i \cap B_j = \phi} w_1 m_1^{\beta}(A_i) w_2 m_2^{\beta}(B_j)} \quad (5)$$

其中, $w_i = \frac{\sum_{j=1, j \neq i}^n (1-d(m_i, m_j))}{\sum_{i=1}^n \sum_{j=1, j \neq i}^n (1-d(m_i, m_j))}$, $d(m_i, m_j)$ 为 m_i, m_j 之

间的距离。

对于式(2)和式(3)所得到的不同来源的证据, 利用修正证据合成式(5)进行合成。文中只对两类上下文信息进行证据生成, 当然还有其他类型用户上下文信息如用户社交关系以及其他环境上下文信息如天气、灯光等等。由于证据合成规则符合交换律和结合率, 即证据合成与证据顺序无关, 因此今后获得证据很方便加入到合成规则中。

3.2.3 决策形成

通过不同证据源证据合成后, 形成了用户所有可能活动位置的候选集合, 然后根据信度值大小, 选择最大信度值的用户活动位置作为预测的用户将来的目标位置。

3.3 路径预测

预测出用户未来可能活动位置后, 就可以根据用户当前

位置推理用户将来路径, 形成用户有序路径序列。路径序列每一元素可以用一个二元组表示, 形如 (α_i, D_i) , 其中 D_i 表示方向值。比如在图 2(b)中, 假设当前位置为 α_7 , 根据预测目标位置为 α_{10} , 则路径预测序列为 $(\alpha_7, ES), (\alpha_6, ES), (\alpha_5, S), (\alpha_4, S), (\alpha_3, E), (\alpha_9, E), (\alpha_{10}, B_3)$, B_3 为预测目标位置。

如果从 α_7 到 α_{10} 有多条路径可供选择, 则可以根据其他有关上下文信息进行路径预测, 比如路径距离及交通阻塞情况等。

3.4 基于上下文感知和移动位置预测的切换

在得到用户预测路径序列后, 依据地理信息 GIS 中移动基站 BS 信息, 将预测用户路径二元组扩展成为三元组, 加上一维即路径所处基站 $cell_id$, 因此路径元素可表示为 $(\alpha_i, D_i, cell_id)$, 如图 5 所示。

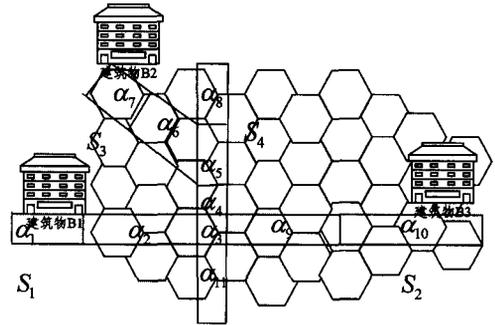


图 5 基于上下文感知和移动位置预测切换

依据图 5, 结合终端信号强度和质量以及用户运动方向 (比如用户是远离还是逐渐靠近所处基站) 对切换判决提供综合因素依据, 从而避免了由于高楼遮挡信号强度变化引起的“乒乓切换”, 提高了切换成功率。此外, 依据基于上下文感知和移动位置预测的切换机制还可以提前预留资源, 保障用户网络服务质量 QoS 。

4 仿真实验

4.1 实验环境

实验模拟环境如图 2(a)所示, 利用实验室“上下文信息处理系统”采集、表示和推理各类上下文信息, 形成智能决策。上下文信息处理系统分为 3 层: 模拟传感网络、上下文处理平台和上下文应用服务器, 如图 6 所示。

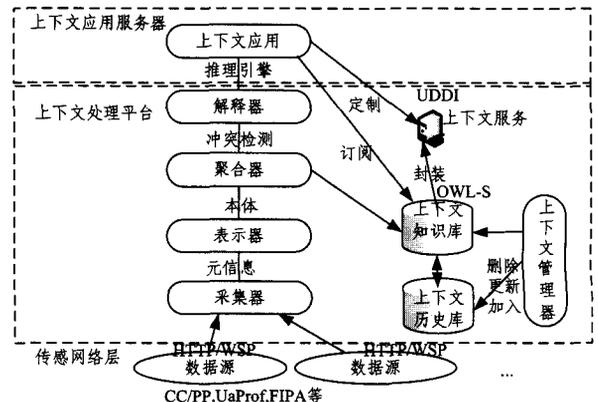


图 6 上下文信息处理系统

模拟传感网络是上下文处理系统的上下文信息来源, 上

下文处理平台是上下文处理的核心,推理引擎是其主要模块。它独立于特定上下文感知应用。上下文应用服务器主要为用户提供个性化应用服务。在模拟传感网络层通过模拟界面获取各种类型上下文,例如地理信息 GIS(标志建筑物、路径小段)、基站信息(信号强度、cell_id 等)和用户上下文(用户偏好、日程表等);上下文处理平台对采集的上下文信息利用证据理论进行聚合、推理,形成智能决策信息供上层上下文应用服务器调用。

4.2 实验方法

对基于上下文感知和修正证据理论的用户移动位置预测进行 MATLAB 仿真,与用户真实运动轨迹误差进行比对,用户的真实运动轨迹可以通过全球定位系统 GPS(Global Positioning System)设备获取。

本文对经典 Hausdorff(HD)距离算法^[16]进行改进,构造了一种新的三维距离函数。在此基础上,提出了一种三维 Hausdorff 距离算法。经典 Hausdorff 距离是描述两组点集之间相似程度的一种量度,它是集合与集合之间距离的一种定义形式。

假设真实的运动轨迹 HD_λ 和预测的运动轨迹 HD_ω 之上的采样点分别表示为 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 和 $\{\omega_1, \omega_2, \dots, \omega_m\}$, 在时刻 $t(o)$ 经过的任意一点 $\forall o \in HD_\lambda \cup HD_\omega$ 可以表示为 $o = (x, y, z, t(o))$ 。利用改进的三维 Hausdorff 距离测量公式可以计算某点 ω_j 与真实运动轨迹 HD_λ 的误差 $\delta(\omega_j, HD_\lambda)$, 见式(6)。

$$\delta(\omega_j, HD_\lambda) = \|\lambda_i - \omega_j\|_{\min_{\lambda_i} |t(\lambda_i) - t(\omega_j)|} \quad (6)$$

其中, $\|\cdot\|$ 为欧几里德范数,对于任意 $\lambda_i \in HD_\lambda$ 和 $\omega_j \in HD_\omega$, 欧几里德范数可以通过式(7)获得。

$$\|\lambda_i - \omega_j\| = \sqrt{(\lambda_i^x - \omega_j^x)^2 + (\lambda_i^y - \omega_j^y)^2 + (\lambda_i^z - \omega_j^z)^2} \quad (7)$$

因此,预测运动轨迹整条曲线 HD_ω 与真实的运动轨迹 HD_λ 的误差为:

$$\delta_m(HD_\lambda, HD_\omega) = \sum_{\omega_j \in HD_\omega} \delta(\omega_j, HD_\lambda) / m \quad (8)$$

4.3 实验结果和分析

实验 1 真实运动轨迹和各算法预测值比较

为简单起见,令式(7)中第三维数据为 0,则得到二维平面的实验数据。终端真实运动轨迹 True Location(True Loc)与基于上下文感知和证据理论用户移动位置预测 Location Prediction based on Context-aware and D-S Theory (LPCDST)以及序列蒙特卡洛滤波 Sequential Monte Carlo Filter(SMC)方法^[4]预测终端位置的对比如图 7 所示。

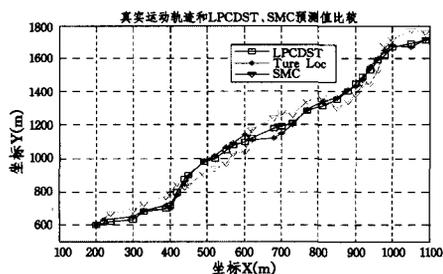


图 7 真实运动轨迹和 LPCDST、SMC 预测值结果比较

从图中可以得知,基于上下文和证据理论用户移动位置

预测的准确性比较高,采用各类上下文信息并利用修正的证据理论进行用户移动位置预测相对于其它位置预测方法有较高的准确性。

实验 2 用户运动位置 HD 误差值比较

利用式(5)可得 LPCDST 和 SMC 方法终端 Hausdorff (HD)距离位置误差值,如图 8 所示。

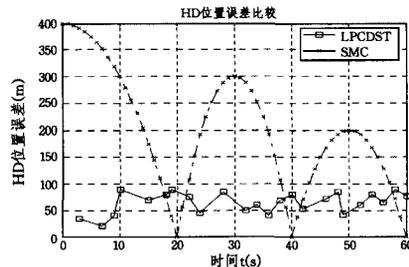


图 8 HD 位置误差比较

由图可知,基于上下文和证据理论用户移动位置预测 LPCDST 用户 HD 位置误差和比其它方法 SMC 位置预测的要小,从而有效提高了切换成功率。

结束语 切换机制是无线通信系统中保持用户移动性的必要条件,切换成功与否直接影响着用户的业务体验和应用。本文提出一种基于上下文感知和移动位置预测的切换机制,这种方法不需要用户大量历史活动信息作为先验知识,仅仅依据用户上下文(用户偏好、日程表等)以及用户所处物理环境上下文(地理信息 GIS、基站等)来准确预测用户将来可能活动位置和路径序列,并综合考虑终端信号强度和质量以及用户运动方向进行综合切换判决。仿真实验证明,该方法对用户移动位置预测有较高的准确性,从而能够为切换判决提供可靠决策依据。

文中方法仅仅涉及比较重要的几种类型上下文,上下文信息种类还需要进一步完善。另外,本体是上下文建模的重要工具和方法,在今后研究中可以利用本体对各种上下文建模并基于证据理论进行不确定性推理,从而减小计算复杂度,为用户实时性切换提供可靠保证。

参考文献

- [1] 李鹏. 一种基于业务信道信号质量的优化接力切换策略[J]. 西华大学学报, 2013, 32(4): 58-61
- [2] Luo Tao, Lv Zi-ru, Yao Zhe, et al. A Position-Based Handoff Scheme in WAVE Systems[J]. China Communications, 2011; 8(1): 95-101
- [3] Yang Zi-gang, Wang Xiao-dong. Joint Mobility Tracking and Handoff in Cellular Networks via Sequential Monte Carlo Filtering[J]. IEEE Transactions on Signal Processing, 2003, 51(1): 269-281
- [4] Marmasse N, Schmandt C. A User-Centered Location Model [J]. Personal and Ubiquitous Computing, 2002, 6(5): 318-321
- [5] Shen X, Mark J W, Ye J. User Mobility Profile Prediction: An Adaptive Fuzzy Inference Approach [J]. Wireless Networks, 2000, 6(6): 363- 374
- [6] Kumar V, Venkataram P. A Prediction Based Location Management Using Multi-Layer Neural Networks[J]. J. Indian Inst. of

[7] Tseng Chien-chao, Lin Chia-liang, Shih B-Y, et al. Topology-aware handoff scheme for surveillance patrol robot[J]. *Non-linear Dynamics*,2013,73(3):2073 - 2081

[8] Abo-Zahhad M,Ahmed S M,Mourad M. Future location prediction of mobile subscriber over mobile network using Intra Cell Movement pattern algorithm[C]//International Conference on Communications, Signal Processing, and their Applications (ICCSPA). 2013;1-6

[9] Jin Zhao-yan, Shi Dian-xi, Wu Quan-yuan, et al. Random Walk Based Location Prediction in Wireless Sensor Networks. [J]*International Journal of Distributed Sensor Networks*,2013;1-9

[10] Samaan N, Karmouch A. A Mobility Prediction Architecture Based on Contextual Knowledge and Spatial Conceptual Maps [J]. *IEEE Transactions on Mobile Computing*,2005,4(6):537-551

[11] 孙巍巍,苏寒松,滕友伟,等. 3GPP LTE 系统中结合位置预测的

[12] Yuan Yue-ming, Guan Wei, Qiu Wei. Map matching of mobile probes based on handoverlocation technology[C]//2010 International Conference on Networking Sensing and Control, IC-NSC,2010;587-592

[13] Wang Yuan-lin, Yu Zhao-yuan, Luo Wen, et al. A 3D GIS spatial data model based on conformal geometric algebra[J]. *Science China Earth Sciences*,2011,54(1):101-112

[14] Shafer G. A Mathematical Theory of Evidence [M]. Princeton, NJ:Princeton Univ. Press,1976;124-130

[15] Miao Y Z, Zhang H X, Zhang J W, et al. Improvement of the combination rules of the D-S evidence theory based on dealing with the evidence conflict[C]//2008 IEEE International Conference on Information and Automation. Shanghai,2008;331-336

[16] Peterfreund N. Robust tracking of position and velocity with Kalman snakes [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*,2000,22(6):564-569

(上接第 90 页)

来有更先进的缺陷定位技术提出时,可以使用 GenProg-FL 工具从软件错误自动修复的角度对其进行评估,验证其定位的有效性。

通过使用 GenProg-FL 工具对 Jaccard、Ample 等缺陷定位技术的评估发现,在 GenProg-FL 中使用更先进的缺陷定位技术,能明显提高软件错误自动修复的效率,所产生的候选补丁数量得到很明显的减少,而验证补丁有效性的过程一致,所以修复所占用的时间便得到很明显的减少。另外,通过对 14 种缺陷定位技术的评估,我们认为在未来有更高效率的缺陷定位技术提出之前,应该优先考虑使用 Jaccard 指导 GenProg-FL 来对软件故障进行修复^[5],Jaccard 可以有效地提高软件自动修复的效率。

GenProg-FL 工具将基于程序谱的缺陷定位技术与软件自动修复技术结合起来,对两者的发展都具有十分重要的意义。

结束语 本文系统地对 GenProg-FL 工具进行了介绍。首先,在对 GenProg 的介绍中,重点分析研究了 GenProg 的缺陷定位策略,发现其存在的不足。缺陷定位技术虽然已经得到很大的发展,传统的评估方法也对现有的缺陷定位技术进行了评估,但并没有促使其很好地运用于工业实践中。在此基础上,本文提出了一种创新性的评估方法,即从软件自动修复的角度出发,使用 GenProg-FL 工具接受某种缺陷定位技术的指导,去修复现实存在缺陷的故障程序。同时,以产生的候选补丁数量(NCP)为评估标准,对所使用的缺陷定位技术的定位有效性进行评估。本文对此评估方法的可行性进行了系统的阐述。同时,指出了更有效的缺陷定位技术对软件自动修复的意义。在更有效的缺陷定位技术的指导下,GenProg-FL 在软件错误自动修复上的效率得到明显的提升,这对软件错误自动修复技术的发展同样具有十分重要的意义。

参 考 文 献

[1] IEEE Standard Classification for Software Anomalies[S]. IEEE

Std 1044-2009,2010,(1044-1993);C1-15

[2] Le Goues C,Dewey-Vogt M,Forrest S,et al. A systematic study of automated program repair[C]//fixing 55 out of 105 bugs for \$ 8 each; International Conference on Software Engineering (ICSE),2012. 2012;3-13

[3] Le Goues C, Nguyen T, Forrest S, et al. Genprog: A generic method for automatic software repair[J]. *IEEE Transactions on Software Engineering(TSE)*,2012,38(1):54-72

[4] Abreu P Z, van Gemund A J C. Spectrum-based Multiple Fault Localization[C]//Proceedings of the ASE'09. 2009

[5] Qi Yu-hua, Mao Xiao-guang, Lei Yan, et al. Using Automated Program Repair for Evaluating the Effectiveness of Fault Localization Techniques[C]//Proceedings of the 2013 International Symposium on Software Testing and Analysis,2013. 2013;191-201

[6] Wilcoxon F. Individual comparisons by ranking methods[J]. *Biometrics Bulletin*,1945,1(6):80-83

[7] Vargha A, Delaney H D. A critique and improvement of the CL common language effect size statistics of mcgraw and wong[J]. *Journal of Educational and Behavioral Statistics*,2000,25(2):101-132

[8] Abreu R,Zoetewij P,Golsteijn R, et al. A practical evaluation of spectrum-based fault localization[J]. *Journal of Systems and Software(JSS)*,2009,82(11):1780-1792

[9] Xie X Y, Chen T Y, Kuo F-C, et al. A Theoretical Analysis of the Risk Evaluation Formulas for Spectrum-Based Fault Localization[J]. *ACM Transactions on Software Engineering and Methodology (TOSEM)*,2013,22(4)

[10] Yu Y, Jones J A, Harrold M J. An empirical study of the effects of test-suite reduction on fault localization[C]//International Conference on Software Engineering (ICSE),2008. 2008;201-210

[11] Ali S, Andrews J H, Dhandapani T, et al. Evaluating the accuracy of fault localization techniques[C]//International Conference on Automated Software Engineering (ASE),2009. 2009;76-87