

一种基于模糊决策树的运动预测机制^{*})

王兴伟¹ 原常青¹ 黄敏²

(东北大学计算中心 沈阳 110004)¹ (东北大学信息科学与工程学院 沈阳 110004)²

摘要 通过考虑环境约束及用户运动的随机性,设计了一种基于模糊决策树的运动预测机制。该机制在示例集的基础上,利用属性模糊化和模糊分类熵建立初始模糊决策树,由此生成模糊决策规则进行预测,根据变化情况适时对模糊决策树进行必要的维护。仿真研究表明,该机制预测准确率较高,预测开销较小,是可行和有效的。

关键词 运动预测,模糊决策树,模糊分类熵,移动 Internet

A Fuzzy-Decision-Tree-Based Motion Prediction Mechanism

WANG Xing-Wei¹ YUAN Chang-Qing¹ HUANG Min²

(Computing Center, Northeastern University, Shenyang 110004)¹

(School of Information Science and Engineering, Northeastern University, Shenyang 110004)²

Abstract Considering environment factor and user movement randomness, a fuzzy decision tree based motion prediction mechanism in mobile Internet is proposed. Based on the instance set, an initial fuzzy decision tree is built up with fuzzy attribute and classification entropy. The tree produces fuzzy decision rules for instant motion predictions and evolves with time. Simulation results have shown that the proposed mechanism is both feasible and effective with relatively high accuracy and low overhead of motion prediction.

Keywords Motion prediction, Fuzzy decision tree, Fuzzy classification entropy, Mobile Internet

1 引言

在移动 Internet 中,当用户从一个小区运动到另一个小区时,由于无线信道不可靠且资源有限,提前预测目标小区并进行资源预留是保证切换期间通信正常的有效手段^[1]。预测准确与否关系到对用户服务质量(QoS, Quality of Service)的保证^[2]。

通常,小区内地形、地貌、道路和建筑物等很少改变,据此可以对用户运动进行预测。但是,由于用户运动的随机性,因此很难准确预测。

已有的运动预测机制大体可分为三类。第一类通过比较用户监测到的周围小区信号或者周围小区监测到的用户信号进行预测。文[3]选取信号强度最大的小区作为目标小区,但由于障碍物以及多径衰落等原因导致信号强度值不准,预测效果较差;文[4]对此作了改进,定义了比信号强度更贴合实际的信道质量,选取该值较好的小区作为目标小区,在一定程度上克服了多径衰落等的影响,但预测过程需要周围小区配合;文[5]考虑用户运动随机性的影响,比较在本小区和相邻小区追踪获得的信号强度值,引入多目标预测,确定一个或多个目标小区,预测准确性有所提高,但开销增大。第二类根据运动概率模型进行预测。文[6]根据当前及周围小区监测到的用户信号,确定其运动速度,计算用户到达周围小区的概率,从而确定目标小区,但未考虑速度对转向的影响;文[7]中的运动概率模型虽然考虑了速度与转向可能性之间的关系且预测一扇形区域内的一个或多个小区作为目标小区,但也存在开销较大的问题。第三类根据地理信息进行预测。文[8]使用 GPS (Global Positioning System)^[8] 系统周期性收集用户地理位置信息,据此确定目标小区,但未充分考虑用户运动随机性的影响。

本文设计了一种基于模糊决策树的运动预测机制,力图在尽量减小预测开销的同时,增大预测准确性。仿真研究表明,该机制是可行和有效的。

2 模糊决策树

决策树是归纳学习领域的一个重要分支。通常从示例集出发,依据启发式信息,产生一棵决策树,并使用由决策树转化而来的规则预测待定事件。ID3 算法^[9~11]是国际上非常有影响的决策树算法,它使用分类熵作为启发式变量,选取使得类的混乱程度最小(即具有最小分类熵)的属性作为扩展属性,导出一棵清晰决策树。然而,在现实世界中有很多事物难以准确描述,导致模糊决策树^[9~11]的出现。

设论域 $X = \{e_1, e_2, \dots, e_n\}$, 表示总结预测规则的示例集, n 值根据实际情况确定。示例集中的任一元素 e_k ($k=1, 2, \dots, n$) 都有如下属性: $A^{(1)}, A^{(2)}, \dots, A^{(s)}$ 。 $A^{(s)}$ 是类属性,即需要预测的内容。除 $A^{(s)}$ 外,其它属性 $A^{(i)}$ ($i=1, 2, \dots, s-1$) 都有与之对应的属性描述集合 $T(A^{(i)}) = \{T_1^{(i)}, T_2^{(i)}, \dots, T_{m_i}^{(i)}\}$ ($j=1, 2, \dots, m_i$) 对应定义在论域 X 上的一个模糊集合。

用 $\mu_{T_j^{(i)}}(e_k)$ 表示 e_k 对 $T_j^{(i)}$ 的隶属度函数,则 $T_j(A^{(i)}) = \frac{\mu_{T_j^{(i)}}(e_1)}{e_1} + \frac{\mu_{T_j^{(i)}}(e_2)}{e_2} + \dots + \frac{\mu_{T_j^{(i)}}(e_n)}{e_n}$ ^[11]。

3 基于模糊决策树的运动预测机制

图 1 是通常在移动 Internet 中的小区布局示意图。假设在 0 小区中有用户即将切换,则目标小区必然是 1~6 小区之一。所有用户从 0 小区切换到其他小区的历史记录组成用户在小区 0 切换时的决策树示例集。本文设计的运动预测机制基于示例集建立一棵初始模糊决策树,用于预测切换的目标

^{*} 国家自然科学基金项目(60473089, 60003006, 70101006); 辽宁省自然科学基金项目(20032018, 20032019); 教育部现代远程教育工程项目。王兴伟 博士, 教授, 博士生导师, 主要研究领域为下一代互联网、移动无线 Internet 及 IP/DWDM 光 Internet。

小区,并在线适时维护该树。

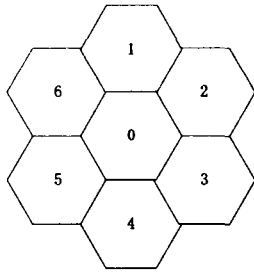


图1 小区布局

3.1 初始模糊决策树建立

3.1.1 属性设置 设置6类11个属性,即 $s=11$ 。

① 切换前位置 $A^{(1)}$ 用一个二元组 $\langle r, a \rangle$ 表示。如图2,切换通常发生在小区边界,根据 r 和 a 可确定切换发生位置。

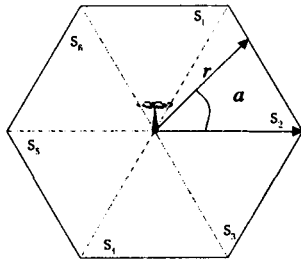


图2 切换位置

② 切换时间 $A^{(2)}$ 记录切换发生的时刻。根据具体情况设置计时单位。

③ 运动速率 $A^{(3)}$ 用户发生切换时即时速率的径向分量 $|v_r|$ 。

④ 用户 QoS 等级 $A^{(4)}$ 参照 DiffServ^[12],将用户 QoS 需求分成多个等级。

⑤ 用户接收到的周围6个小区的信号强度 $A^{(5)}, A^{(6)}, A^{(7)}, A^{(8)}, A^{(9)}, A^{(10)}$ 。

⑥ 切换后位置 $A^{(11)}$ 该属性是类属性。对于图1,如果用户切换前位于0小区,则 $A^{(11)}$ 的取值范围为 $\langle 1$ 小区,2 小区,3 小区,4 小区,5 小区,6 小区 \rangle 。

3.1.2 属性模糊化 在上述11个属性中, $A^{(11)}$ 是类属性,是模糊决策树的分类标准和预测内容,毋需模糊化。对其它10个属性,需按其属性描述对示例集进行模糊分类。

$A^{(1)}$ 设置 $m_1=6$ 个属性描述,即六边形的每条边各对应一个属性描述一扇区(如图2), $T(A^{(1)}) = \{T_1^{(1)}, T_2^{(1)}, T_3^{(1)}, T_4^{(1)}, T_5^{(1)}, T_6^{(1)}\} = \{S_1, S_2, S_3, S_4, S_5, S_6\}$, 对应的隶属度函数 $\mu_{11}(\alpha), \mu_{12}(\alpha), \mu_{13}(\alpha), \mu_{14}(\alpha), \mu_{15}(\alpha), \mu_{16}(\alpha)$ 如图3所示。

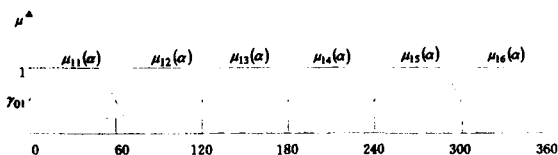


图3 切换前位置模糊化

$A^{(2)}$ 可根据实际情况设置 m_2 个属性描述, $T(A^{(2)}) = \{T_1^{(2)}, T_2^{(2)}, \dots, T_{m_2}^{(2)}\} = \{t_1, t_2, \dots, t_{m_2}\}$, 如 $\{8:00:00, 9:00:00, \dots, 22:00:00\}$ 。设 t 为时间变量,对应的隶属度函数 $\mu_{21}(t), \mu_{22}(t), \dots, \mu_{2m_2}(t)$ 分别如图4所示。

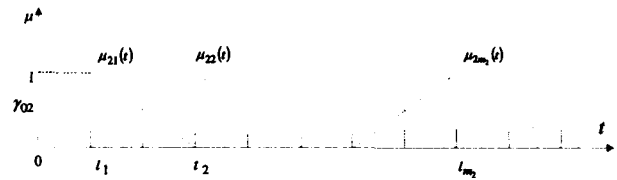


图4 切换时间模糊化

$A^{(3)}$ 可根据实际情况设置 m_3 个属性描述, $T(A^{(3)}) = \{T_1^{(3)}, T_2^{(3)}, \dots, T_{m_3}^{(3)}\} = \{V_1, V_2, \dots, V_{m_3}\}$, $V_1 < V_2 < \dots < V_{m_3}$, 对应的隶属度函数 $\mu_{31}(|v_r|), \mu_{32}(|v_r|), \dots, \mu_{3m_3}(|v_r|)$ 分别如图5所示。

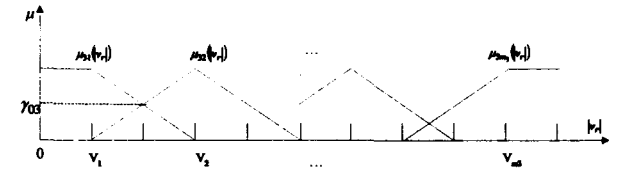


图5 运动速率模糊化

$A^{(4)}$ 分为 $1, 2, \dots, Q$ 个等级, QoS 保证程度依次提高,设置 $m_4=Q$ 个属性描述,即 $T(A^{(4)}) = \{T_1^{(4)}, T_2^{(4)}, \dots, T_{m_4}^{(4)}\} = \{1, 2, \dots, Q\}$, 对应的隶属度函数 $\mu_{41}, \mu_{42}, \dots, \mu_{4m_4}$ 分别如图6所示。

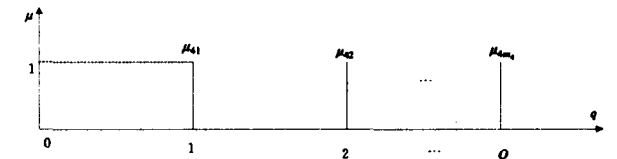


图6 QoS等级模糊化

$A^{(5)}, A^{(6)}, A^{(7)}, A^{(8)}, A^{(9)}, A^{(10)}$ 可根据实际情况分别设置 $m_5, m_6, m_7, m_8, m_9, m_{10}$ 个属性描述, $T(A^{(l)}) = \{T_1^{(l)}, T_2^{(l)}, \dots, T_{m_l}^{(l)}\} = \{I_{l1}, I_{l2}, \dots, I_{lm_l}\} (l=5, 6, 7, 8, 9, 10), I_{ln} < I_{l2} < \dots < I_{lm_l}$ 。设 is_l 为信号强度变量,对应的隶属度函数 $\mu_{51}(is_1), \mu_{52}(is_1), \dots, \mu_{5m_5}(is_1)$ 分别如图7所示。

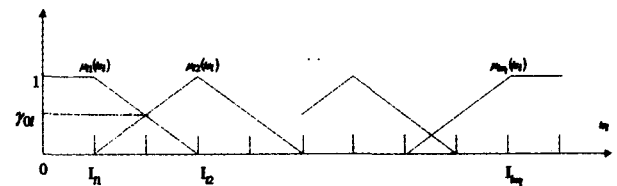


图7 接收信号强度模糊化

3.1.3 模糊分类熵 根据文[13],作出如下定义:

定义1 第 i 个属性 $A^{(i)}$ 的第 j 个属性描述 $T_j^{(i)}$ 的模糊集合 $T_j(A^{(i)})$ 的势定义为: $M_j(A^{(i)}) = \sum_{k=1}^n \mu_{ij}(e_k)$ 。

定义2 属性描述 $T_j^{(i)}$ 的关于类属性的第 k 个属性值相关频度定义为: $p_k^{(i)} = \frac{M(T_j(A^{(i)}) \cap T_k(A^{(i)}))}{M(T_j(A^{(i)}))}$ 。

定义3 属性描述 $T_j^{(i)}$ 的分类熵定义为: $entropy_j^{(i)} = \sum_{k=1}^{m_j} p_k^{(i)} \log_2 \frac{1}{p_k^{(i)}}$ 。

$$-p_{jk}^{(i)} \log_2 p_{jk}^{(i)}。$$

定义4 第*i*个属性 $A^{(i)}$ 的平均分类熵定义为:

$$E^{(i)} = \sum_{j=1}^{m_i} \frac{M(T_j(A^{(i)}))}{\sum_{j=1}^{m_i} M(T_j(A^{(i)}))} \text{entropy} p_{j}^{(i)}。$$

3.1.4 初始模糊决策树建立方法

步骤1:将示例集中的所有元素模糊化;

步骤2:从所有属性中,选取 $E^{(i)}$ 最小的属性作为模糊决策树根节点的属性,并设根节点为当前节点;

步骤3:从当前节点属性的每一属性描述对应的模糊集合中,选取隶属度函数值大于门限值 $\gamma(0 < \gamma \leq \gamma_{0h}, h=1, 2, 3, 5, 6, 7, 8, 9, 10)$,参见图3,4,5和7)的元素组成模糊子类集,每一个模糊子类集产生当前节点的一个孩子节点。

步骤4:对当前节点的每一个孩子节点,若其模糊子类集中的所有元素都具有相同类属性值或在从根节点到当前节点的分枝上的所有节点的属性都已使用过,则设置其为叶子节点;若所有孩子节点都设置成叶子节点,则模糊决策树建立完毕,否则选取未使用过的且 $E^{(i)}$ 最小的属性作为该孩子节点的属性,并设该孩子节点为当前节点,转步骤3。

图8是使用上述方法生成的一棵模糊决策树。

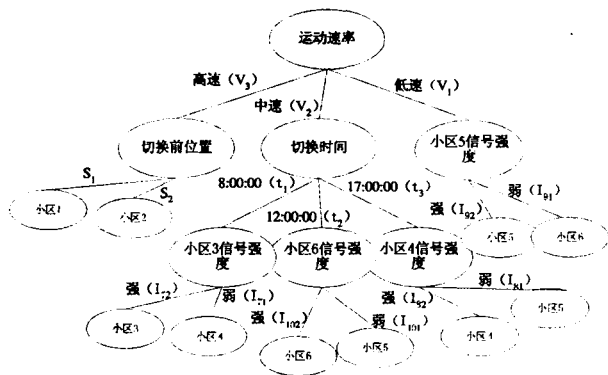


图8 模糊决策树举例

3.2 基于模糊决策树的运动预测机制

3.2.1 模糊规则

定义5^[11] 设A,B是定义在论域X上的模糊集合。称“IF A THEN B”为一条模糊规则,简记为 $A \rightarrow B$,A称为条件模糊集,B称为结论模糊集。

按照定义5,可以将模糊决策树转化为一系列的模糊决策规则,每个叶子节点对应一条规则。例如,对图8而言,可以产生如下10条规则:

规则1:IF $A^{(1)}$ 隶属于“ S_1 ”(即 $\mu_{s1}(\alpha) > \gamma$) AND $A^{(3)}$ 隶属于“高速(V_3)”(即 $\mu_{s3}(v) > \gamma$) THEN $A^{(11)}$ 隶属于“小区1”;

规则2:IF $A^{(1)}$ 隶属于“ S_2 ”(即 $\mu_{s2}(\alpha) > \gamma$) AND $A^{(3)}$ 隶属于“高速(V_3)”(即 $\mu_{s3}(v) > \gamma$) THEN $A^{(11)}$ 隶属于“小区2”;

规则3:IF $A^{(2)}$ 隶属于“8:00:00(t_1)”(即 $\mu_{t1}(t) > \gamma$) AND $A^{(3)}$ 隶属于“中速(V_2)”(即 $\mu_{s2}(v) > \gamma$) AND $A^{(7)}$ 隶属于“强(I_{72})”(即 $\mu_{r2}(r_{s3}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区3”;

规则4:IF $A^{(2)}$ 隶属于“8:00:00(t_1)”(即 $\mu_{t1}(t) > \gamma$) AND $A^{(3)}$ 隶属于“中速(V_2)”(即 $\mu_{s2}(v) > \gamma$) AND $A^{(7)}$ 隶属于“弱(I_{71})”(即 $\mu_{r1}(r_{s3}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区4”;

规则5:IF $A^{(2)}$ 隶属于“12:00:00(t_2)”(即 $\mu_{t2}(t) > \gamma$) AND $A^{(3)}$ 隶属于“中速(V_2)”(即 $\mu_{s2}(v) > \gamma$) AND $A^{(10)}$ 隶属于“强(I_{102})”(即 $\mu_{102}(r_{s6}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区6”;

规则6:IF $A^{(2)}$ 隶属于“12:00:00(t_2)”(即 $\mu_{t2}(t) > \gamma$) AND $A^{(3)}$ 隶属于“中速(V_2)”(即 $\mu_{s2}(v) > \gamma$) AND $A^{(10)}$ 隶属于“弱(I_{101})”(即 $\mu_{101}(r_{s6}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区5”;

规则7:IF $A^{(2)}$ 隶属于“17:00:00(t_3)”(即 $\mu_{t3}(t) > \gamma$) AND $A^{(3)}$ 隶属于“中速(V_2)”(即 $\mu_{s2}(v) > \gamma$) AND $A^{(8)}$ 隶属于“强(I_{82})”(即 $\mu_{82}(r_{s4}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区4”;

规则8:IF $A^{(2)}$ 隶属于“17:00:00(t_3)”(即 $\mu_{t3}(t) > \gamma$) AND $A^{(3)}$ 隶属于“中速(V_2)”(即 $\mu_{s2}(v) > \gamma$) AND $A^{(8)}$ 隶属于“弱(I_{81})”(即 $\mu_{81}(r_{s4}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区5”;

规则9:IF $A^{(3)}$ 隶属于“低速(V_1)”(即 $\mu_{s1}(v) > \gamma$) AND $A^{(9)}$ 隶属于“强(I_{92})”(即 $\mu_{92}(r_{s5}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区5”;

规则10:IF $A^{(3)}$ 隶属于“低速(V_1)”(即 $\mu_{s1}(v) > \gamma$) AND $A^{(9)}$ 隶属于“弱(I_{91})”(即 $\mu_{91}(r_{s5}) > \gamma$) THEN $A^{(11)}$ 隶属于“小区6”。

3.2.2 运动预测过程

步骤1:记录用户的切换前位置、切换时间、运动速率、QoS等级及其接收到的周围6个小区的信号强度并进行模糊化;

步骤2:依照模糊决策树寻找条件模糊集与用户运动情况匹配的规则。若没有匹配的规则,则预测失败,结束;否则,转步骤3;

步骤3:若有一条或多条规则预测用户目标小区隶属于同一小区,则将其作为目标小区;若有多条规则预测用户目标小区隶属于不同小区,则将其都作为目标小区(多目标预测),结束。

3.3 模糊决策树的维护

小区内道路或建筑物的改变以及基站的移动(如移动通信车)会引起小区覆盖范围变化,从而导致依据现有模糊决策树进行预测的准确性下降。因此,需要根据变化情况适时进行必要的维护。

定义6 对一条模糊决策规则,其预测正确的次数与总预测次数之比称为其预测准确率。由于模糊决策树的每个叶子节点都对应一条规则,因此也可称之为叶子节点预测准确率,用 p_f 表示。

定义7 模糊决策树中所有叶子节点的平均预测准确率称为树的预测准确率,用 p_t 表示。

定义8 模糊决策树中所有叶子节点预测准确率的方差称为树的预测方差,用 σ_t 表示。

一般地,对模糊决策树的维护操作主要有三种:重构,即基于示例集重新建立模糊决策树,并产生模糊决策规则;剪枝,即将叶子节点(模糊决策规则)以及仅与之相关的分枝或节点删除;分叉,即将叶子节点修改为孩子节点,利用初始模糊决策树建立方法计算一棵嫁接于该处的模糊决策子树。

设 p_f 的门限为 P_f , p_t 的门限为 P_t 且 σ_t 的门限为 Δ_t ,则对模糊决策树的维护主要是进行如下判断和操作:

① 若 $p_t < P_t$,则重构;

② 若 $p_t \geq P_t$ 且 $\sigma_t > \Delta_t$,则遍历每个叶子节点。若某叶子节点预测次数为零,则剪枝;若预测次数不为零且 $p_f < P_f$,则分叉;否则,不进行任何操作。

4 仿真实现与性能评价

在SAS 8.2平台^[14]上进行了仿真实现与性能评价。图9

为某园区的部分平面图,0~6 是其中的 7 个小区。为了评价本文设计的运动预测机制,对从 0 小区切换到周围 1~6 小区的情况进行实际观测,共获得 6000 条观测记录。根据记录时间先后,将其中的前 70%记录构成示例集,用来生成初始模糊决策树,后 30%构成检验集,用来检验预测效果。

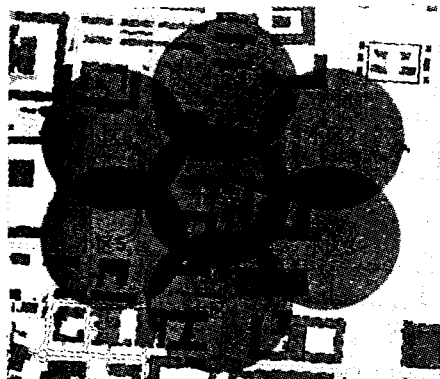


图 9 某园区小区示意图

如表 1 所示,随着 γ (参见 3.1.4)值减小,模糊决策树深度增大,叶子节点(即规则)增多,预测准确率提高,但多目标预测比率增大,开销随之上升。在同一条件下,文[7]中预测方法的多目标预测比率约为 26.3%,而本文提出的机制最多不超过 5.2%,因此预测开销相对较小。

表 1 不同 γ 值下的预测准确率

γ	规则数(条)	模糊决策树深度	多目标预测比率(%)	预测准确率(%)
0.5	1	1	1.2	11.6
0.4	2	2	1.2	30.0
0.3	4	3	1.4	62.2
0.25	5	4	1.7	86.8
0.2	7	5	1.9	95.3
0.15	10	6	2.5	98.8
0.1	14	7	2.9	99.5
0.05	20	8	3.5	99.5
0.02	38	9	5.2	99.5

取 $\gamma=0.1$ 并将本文提出的运动预测方法与文[7]中的预测方法在不同速度下作比较,结果如图 10 所示。用户速度较低时,运动随机性较高,本文设计的方法的预测准确率较高;而用户速度较高时,运动随机性较低,由于文[7]采用了针对速度影响的运动概率模型,预测准确率较高。

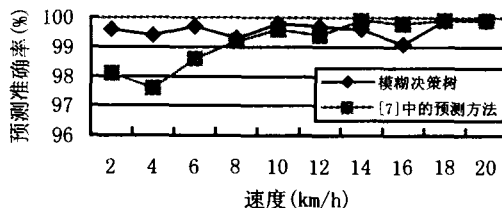


图 10 不同速度下的预测准确率比较

结束语 本文提出一种移动 Internet 中的基于模糊决策树的运动预测机制。仿真研究表明,该机制预测准确率较高,开销较小,是可行和有效的。

参考文献

- 1 Talukdar A K, Badrinath B R, Acharya A. MRSVP: A resource reservation protocol for an integrated services network with mobile hosts[J]. Wireless Networks, 2001, 7(1): 5~19
- 2 王兴伟, 张应辉, 刘积仁, 李华天. 分布式多媒体系统服务质量管理实现支持机制[J]. 软件学报, 1998, 9(2): 87~90
- 3 Hou J, Fang Y. Mobility-based call admission control schemes for wireless mobile networks. Wireless Communications and Mobile Computing[J]. 2001, 5(1): 269~282
- 4 Wang X W, Yuan C Q, Song B, Huang M. A QoS-based handoff supporting mechanism[A]. The Tenth Joint International Computer Conference[C]. 2004, 11: 282~285
- 5 Das S K, Jayaram R, Kakani N K, Sanhoy K. A call admission and control scheme for quality-of-service(QoS) provisioning in next generation wireless networks[J]. Wireless Networks, 2000(6): 17~30
- 6 黄秋波, 李昊, 高传善. 下一代移动网络接纳控制的改进方案[J]. 小型微型计算机系统, 2003, 24(8): 1437~1440
- 7 Wang X W, Yuan C Q, Song B, Huang M. A fan-shaped flexible resource reservation mechanism in mobile wireless Internet[A]. The fifth international conference on parallel and distributed computing, applications and technologies[C], 2004, 12 singapore, LNCS3320: 660~664
- 8 Lee D S, Hsueh Y H. Bandwidth-reservation scheme based on road information for next-generation cellular networks [J]. IEEE Transactions on Vehicular Technology, 2004, 53(1): 243~252
- 9 Dong M, Kothari R. Look-ahead based fuzzy decision tree induction[J]. IEEE Trans. on Fuzzy Systems, 2001, 9(3): 461~468
- 10 Olaru C, Wehenkel L. A complete fuzzy decision tree technique [J]. Fuzzy Sets and Systems, 2003, 138: 221~254
- 11 王熙照, 孙娟, 杨宏伟, 赵明华. 模糊决策树算法与清晰决策树算法的比较研究[J]. 计算机工程与应用, 2003, 23(21): 72~76
- 12 Blake S, Black D, Carlson M, Davies E. An architecture for differentiated services[Z]. RFC2475, 1998
- 13 赵明华, 孙娟, 杨宏伟, 王熙照. 模糊决策树中参数对模糊熵的敏感性分析[J]. 计算机工程, 2003, 29(11): 90~93
- 14 高惠璇. BASE SAS 软件使用手册[M]. 北京: 中国统计出版社, 1997