



(12) 发明专利申请

(10) 申请公布号 CN 112800983 A

(43) 申请公布日 2021.05.14

(21) 申请号 202110138933.7

(22) 申请日 2021.02.01

(71) 申请人 玉林师范学院

地址 537000 广西壮族自治区玉林市教育
东路299号

(72) 发明人 杨小凤 韦艳芳 王强

(74) 专利代理机构 广州海心联合专利代理事务
所(普通合伙) 44295

代理人 王洪娟 马赞斋

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

G06N 20/00 (2019.01)

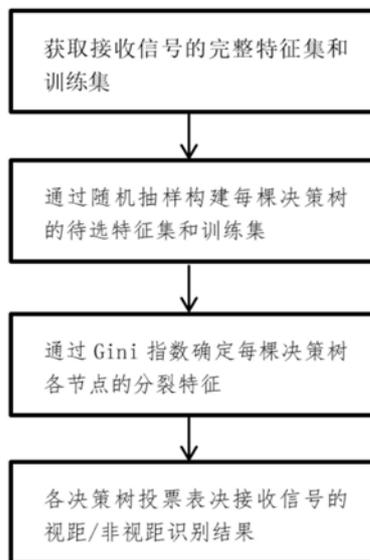
权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种基于随机森林的非视距信号识别方法

(57) 摘要

本发明公开了一种基于随机森林的非视距信号识别方法,涉及无线定位技术领域,解决现有定位方法实时性差、精度低的技术问题,所述方法为:构建由若干决策树构成的随机森林模型,依次测量目标节点到各个基站的接收信号 $r_*(t)$,将 $r_*(t)$ 的特征参数输入随机森林模型,得到目标节点与各个基站接收信号的视距/非视距识别结果,利用接收信号定位时可将识别为非视距的信号去掉。



1. 一种基于随机森林的非视距信号识别方法,其特征在于,包括步骤如下:

S1. 在包含1个目标节点和A个基站的测试区域内随机选取N个训练位置,将训练通信节点依次放置在各所述训练位置,测量该训练通信节点在每一个训练位置 $n \in N$ 到第K个基站的接收信号 $r_n(t)$, $K=1, 2, \dots, A$;

分别计算得到 $r_n(t)$ 的6个特征参数,包括:

能量参数 $e = \int |r_n(t)|^2 dt$ 、

最大幅值参数 $r_{\max} = \max_r |r_n(t)|$ 、

上升时间参数 $t_{\text{rise}} = \min_t \{t: |r_n(t)| \geq 0.6r_{\max}\} - \min_t \{t: |r_n(t)| \geq 6\sigma_n\}$ 、

平均延时参数 $\tau_m = \frac{\int t|r_n(t)|^2 dt}{\int |r_n(t)|^2 dt}$ 、

均方根延时参数 $\tau_r = \frac{\int (t-\tau_m)^2 |r_n(t)|^2 dt}{\int |r_n(t)|^2 dt}$ 、

峰度参数 $\kappa_s = \frac{E[(|r_n(t)| - \mu_r)^4]}{\sigma_r^4}$;

将所述6个特征参数组成特征集 $F = \{e, r_{\max}, t_{\text{rise}}, \tau_m, \tau_r, \kappa_s\}$,将N个位置得到的 $r_n(t)$ 组成训练输入矩阵 $\mathbf{X} = \{r_n(t)\}_{n=1}^N$,建立训练输出矩阵 $\mathbf{y} = \{y_n\}_{n=1}^N$,其中 y_n 为 $r_n(t)$ 的视距或非视距识别标记,若识别为视距,则 $y_n = 1$,若识别为非视距,则 $y_n = 0$,得到完整的训练集 $\mathbf{D} = \{(\mathbf{X}, \mathbf{y})\}_{n=1}^N$;

S2. 测量所述目标节点到该基站的接收信号 $r_*(t)$,并计算 $r_*(t)$ 的6个特征参数;

S3. 构建由若干决策树构成的随机森林模型,对于各个决策树:从完整的训练集 \mathbf{D} 中以有放回抽样的方式抽样N次,形成该决策树的训练集 \mathbf{D}' ,从特征集 F 中随机选取M个特征参数作为该决策树的待选特征, $M < 6$,计算所述待选特征的Gini指数,并按照Gini指数从小到大的顺序将这些待选特征依次作为该决策树的根节点、中间节点和叶节点的分裂特征,

$$Gini(\mathbf{D}', f) = \frac{|\mathbf{D}_1|}{|\mathbf{D}'|} Gini(\mathbf{D}_1) + \frac{|\mathbf{D}_2|}{|\mathbf{D}'|} Gini(\mathbf{D}_2),$$

$$Gini(\mathbf{D}_i) = 1 - \sum_{i=1}^2 p_i^2,$$

其中 $Gini(\mathbf{D}', f)$ 为根据待选特征 f 将训练集 \mathbf{D}' 识别分为 \mathbf{D}_1 (视距)和 \mathbf{D}_2 (非视距)两类的Gini指数, $Gini(\mathbf{D}_i)$ 为类别 \mathbf{D}_i ($i=1, 2$)的Gini指数, $|\mathbf{D}_1|$ 、 $|\mathbf{D}_2|$ 、 $|\mathbf{D}'|$ 分别为集合 \mathbf{D}_1 、 \mathbf{D}_2 、 \mathbf{D}' 中样本的个数, p_i 为类别 \mathbf{D}_i 中识别分类正确和错误的概率;

S4. 将目标节点接收信号 $r_*(t)$ 的特征参数输入所述随机森林模型,每棵决策树将其与各个节点的分裂特征值区间范围比较并输出其视距/非视距识别类别,全部决策树的判断结果中的大多数作为该基站最终的视距/非视距识别结果;

S5. 对所述测试区域内的其余基站重复步骤S1~S4,得到所述目标节点与各个基站接收信号的视距/非视距识别结果,利用接收信号定位时可将识别为非视距的信号去掉。

2. 根据权利要求1所述的一种基于随机森林的非视距信号识别方法,其特征在于,所述 $A \geq 3$ 。

一种基于随机森林的非视距信号识别方法

技术领域

[0001] 本发明涉及无线定位技术领域,更具体地说,它涉及一种基于随机森林的非视距信号识别方法。

背景技术

[0002] 无线定位(Wireless Localization)广泛应用于军事、物流、安全、医疗、搜索和营救等领域。提高定位系统在复杂多径,非视距(None-line-of-sight,NLOS)环境下的定位精度,这是当前基于波达时间(Time-of-Arrival,TOA)的无线定位的研究热点,其关键问题之一为非视距信号识别。

[0003] 非视距信号识别是指当距离测量值较多时,将非视距测量值识别出来并去掉,仅利用视距测量值进行定位。目前主要有三类方法:

[0004] 1) 基于距离测量的方法,根据非视距环境下的距离测量值的方差大于视距环境下的距离测量值的方差,将若干个距离测量值的方差和预先设定的阈值作比较,大于阈值可判定为非视距信号,小于阈值可判定为视距信号,该方法适用于静态目标定位,当目标处于动态时,距离测量值的方差增大,容易将视距信号误判为非视距信号;

[0005] 2) 基于信道统计特性的方法,通过分析信道冲激响应的幅值的累积分布函数来识别非视距信号,或通过比较信道的峰度(kurtosis)、平均时延(mean excess delay)和均方根时延(root mean square delay)的联合似然函数值和门限值来识别非视距信号,但是判决门限的定义比较模糊;

[0006] 3) 通过环境的地理几何信息识别信道,利用光线追踪算法来识别非视距信号,需要事先知道环境的布局。

[0007] 以上算法都属于统计方法,其共同缺点是:(一)一般需要事先知道样本的先验分布,并需要收集足够多的样本数据,而这些要求在实际应用中往往难以达到,并且算法实时性不高;(二)算法需要的特征联合概率分布有时候难以确定,稳定性差。

发明内容

[0008] 本发明要解决的技术问题是针对现有技术的上述不足,本发明的目的是提供一种实时性高、稳定性好的基于随机森林的非视距信号识别方法。

[0009] 本发明的技术方案是:一种基于随机森林的非视距信号识别方法,包括步骤如下:

[0010] S1. 在包含1个目标节点和A个基站的测试区域内随机选取N个训练位置,将训练通信节点依次放置在各所述训练位置,测量该训练通信节点在每一个训练位置 $n \in N$ 到第K个基站的接收信号 $r_n(t)$, $K=1,2,\dots,A$;

[0011] 分别计算得到 $r_n(t)$ 的6个特征参数,包括:

[0012] 能量参数 $e = \int |r_n(t)|^2 dt$ 、

[0013] 最大幅值参数 $r_{\max} = \max_r |r_n(t)|$ 、

[0014] 上升时间参数 $t_{\text{rise}} = \min_t \{t: |r_n(t)| \geq 0.6r_{\max}\} - \min_t \{t: |r_n(t)| \geq 6\sigma_n\}$ 、

[0015] 平均延时参数 $\tau_m = \frac{\int t|r_n(t)|^2 dt}{\int |r_n(t)|^2 dt}$ 、

[0016] 均方根延时参数 $\tau_r = \frac{\int (t-\tau_m)^2|r_n(t)|^2 dt}{\int |r_n(t)|^2 dt}$ 、

[0017] 峰度参数 $\kappa_s = \frac{E[(|r_n(t)|-\mu_r)^4]}{\sigma_r^4}$;

[0018] 将所述6个特征参数组成特征集 $F = \{e, r_{\max}, t_{\text{rise}}, \tau_m, \tau_r, \kappa_s\}$, 将N个位置得到的 $r_n(t)$ 组成训练输入矩阵 $\mathbf{X} = \{r_n(t)\}_{n=1}^N$, 建立训练输出矩阵 $\mathbf{y} = \{y_n\}_{n=1}^N$, 其中 y_n 为 $r_n(t)$ 的视距或非视距识别标记, 若识别为视距, 则 $y_n = 1$, 若识别为非视距, 则 $y_n = 0$, 得到完整的训练集 $\mathbf{D} = \{(\mathbf{X}, \mathbf{y})\}_{n=1}^N$;

[0019] S2. 测量所述目标节点到该基站的接收信号 $r_*(t)$, 并计算 $r_*(t)$ 的6个特征参数;

[0020] S3. 构建由若干决策树构成的随机森林模型, 对于各个决策树: 从完整的训练集 \mathbf{D} 中以有放回抽样的方式抽样N次, 形成该决策树的训练集 \mathbf{D}' , 从特征集 F 中随机选取M个特征参数作为该决策树的待选特征, $M < 6$, 计算所述待选特征的Gini指数, 并按照Gini指数从小到大的顺序将这些待选特征依次作为该决策树的根节点、中间节点和叶节点的分裂特征,

[0021] $Gini(\mathbf{D}', f) = \frac{|\mathbf{D}_1|}{|\mathbf{D}'|} Gini(\mathbf{D}_1) + \frac{|\mathbf{D}_2|}{|\mathbf{D}'|} Gini(\mathbf{D}_2),$

[0022] $Gini(\mathbf{D}_i) = 1 - \sum_{i=1}^2 p_i^2,$

[0023] 其中 $Gini(\mathbf{D}', f)$ 为根据待选特征 f 将训练集 \mathbf{D}' 识别分为 \mathbf{D}_1 (视距) 和 \mathbf{D}_2 (非视距) 两类的Gini指数, $Gini(\mathbf{D}_i)$ 为类别 \mathbf{D}_i ($i=1, 2$) 的Gini指数, $|\mathbf{D}_1|$ 、 $|\mathbf{D}_2|$ 、 $|\mathbf{D}'|$ 分别为集合 \mathbf{D}_1 、 \mathbf{D}_2 、 \mathbf{D}' 中样本的个数, p_i 为类别 \mathbf{D}_i 中识别分类正确和错误的概率;

[0024] S4. 将目标节点接收信号 $r_*(t)$ 的特征参数输入所述随机森林模型, 每棵决策树将其与各个节点的分裂特征值区间范围比较并输出其视距/非视距识别类别, 全部决策树的判断结果中的大多数作为该基站最终的视距/非视距识别结果;

[0025] S5. 对所述测试区域内的其余基站重复步骤S1~S4, 得到所述目标节点与各个基站接收信号的视距/非视距识别结果, 利用接收信号定位时可将识别为非视距的信号去掉。

[0026] 作为进一步地改进, 所述 $A \geq 3$ 。

[0027] 有益效果

[0028] 本发明与现有技术相比, 具有的优点为:

[0029] 本发明的方法将非视距信号识别作为一个视距/非视距两类别分类问题处理, 采用机器学习方法——随机森林进行识别分类, 随机森林是一个包含多个决策树的高精度分类器, 其输出的类别是由各决策树输出的类别的大多数决定, 与现有的统计方法相比, 本发明的方法属于机器学习方法, 不必依赖于样本所从属的总体的分布形式, 仅依赖少量数据观测值的与总体分布无关的性质进行检验和估计, 能够有效减小推断偏差、提高非视距定位精度, 实时性高、稳定性好, 非常适用于基于波达时间的无线定位系统。

附图说明

[0030] 图1为本发明的流程图。

具体实施方式

[0031] 下面结合附图中的具体实施例对本发明做进一步的说明。

[0032] 参阅图1,一种基于随机森林的非视距信号识别方法,包括步骤如下:

[0033] S1.在包含1个目标节点和A个基站的测试区域内随机选取N个训练位置,将训练通信节点依次放置在各训练位置,测量该训练通信节点在每一个训练位置 $n \in N$ 到第K个基站的接收信号 $r_n(t)$, $A \geq 3, K=1, 2, \dots, A$;

[0034] 分别计算得到 $r_n(t)$ 的6个特征参数,包括:

[0035] 能量参数 $e = \int |r_n(t)|^2 dt$ 、

[0036] 最大幅值参数 $r_{\max} = \max_r |r_n(t)|$ 、

[0037] 上升时间参数 $t_{\text{rise}} = \min_t \{t: |r_n(t)| \geq 0.6r_{\max}\} - \min_t \{t: |r_n(t)| \geq 6\sigma_n\}$ 、

[0038] 平均延时参数 $\tau_m = \frac{\int t|r_n(t)|^2 dt}{\int |r_n(t)|^2 dt}$ 、

[0039] 均方根延时参数 $\tau_r = \frac{\int (t-\tau_m)^2 |r_n(t)|^2 dt}{\int |r_n(t)|^2 dt}$ 、

[0040] 峰度参数 $\kappa_s = \frac{E[(|r_n(t)| - \mu_r)^4]}{\sigma_r^4}$;

[0041] 将6个特征参数组成特征集 $F = \{e, r_{\max}, t_{\text{rise}}, \tau_m, \tau_r, \kappa_s\}$,将N个位置得到的 $r_n(t)$ 组成训练输入矩阵 $\mathbf{X} = \{r_n(t)\}_{n=1}^N$,建立训练输出矩阵 $\mathbf{y} = \{y_n\}_{n=1}^N$,其中 y_n 为 $r_n(t)$ 的视距或非视距识别标记,若识别为视距,则 $y_n = 1$,若识别为非视距,则 $y_n = 0$,得到完整的训练集 $\mathbf{D} = \{(\mathbf{X}, \mathbf{y})\}_{n=1}^N$;

[0042] S2.测量目标节点到该基站的接收信号 $r_*(t)$,并计算 $r_*(t)$ 的6个特征参数;

[0043] S3.构建由若干决策树构成的随机森林模型,对于各个决策树:从完整的训练集D中以有放回抽样的方式抽样N次,形成该决策树的训练集 D' ,从特征集F中随机选取M个特征参数作为该决策树的待选特征, $M < 6$,计算待选特征的Gini指数,并按照Gini指数从小到大的顺序将这些待选特征依次作为该决策树的根节点、中间节点和叶节点的分裂特征,

[0044] $Gini(D', f) = \frac{|D_1|}{|D'|} Gini(D_1) + \frac{|D_2|}{|D'|} Gini(D_2)$,

[0045] $Gini(D_i) = 1 - \sum_{i=1}^2 p_i^2$,

[0046] 其中 $Gini(D', f)$ 为根据待选特征 f 将训练集 D' 识别分为 D_1 (视距)和 D_2 (非视距)两类的Gini指数, $Gini(D_i)$ 为类别 D_i ($i=1, 2$)的Gini指数, $|D_1|$ 、 $|D_2|$ 、 $|D'|$ 分别为集合 D_1 、 D_2 、 D' 中样本的个数, p_i 为类别 D_i 中识别分类正确和错误的概率;

[0047] S4.将目标节点接收信号 $r_*(t)$ 的特征参数输入随机森林模型,每棵决策树将其与各个节点的分裂特征值区间范围比较并输出其视距/非视距识别类别,全部决策树的判断结果中的大多数作为该基站最终的视距/非视距识别结果;

[0048] S5.对测试区域内的其余基站重复步骤S1~S4,得到目标节点与各个基站接收信号的视距/非视距识别结果,利用接收信号定位时可将识别为非视距的信号去掉。

[0049] 本发明的方法将非视距信号识别作为一个视距/非视距两类别分类问题处理,采用机器学习方法——随机森林进行识别分类,随机森林是一个包含多个决策树的高精度分类器,其输出的类别是由各决策树输出的类别的大多数决定,与现有的统计方法相比,本发明的方法属于机器学习方法,不必依赖于样本所从属的总体的分布形式,仅依赖少量数据观测值的与总体分布无关的性质进行检验和估计,能够有效减小推断偏差、提高非视距定位精度,实时性高、稳定性好,非常适用于基于波达时间的无线定位系统。

[0050] 以上仅是本发明的优选实施方式,应当指出对于本领域的技术人员来说,在不脱离本发明结构的前提下,还可以作出若干变形和改进,这些都不会影响本发明实施的效果和专利的实用性。

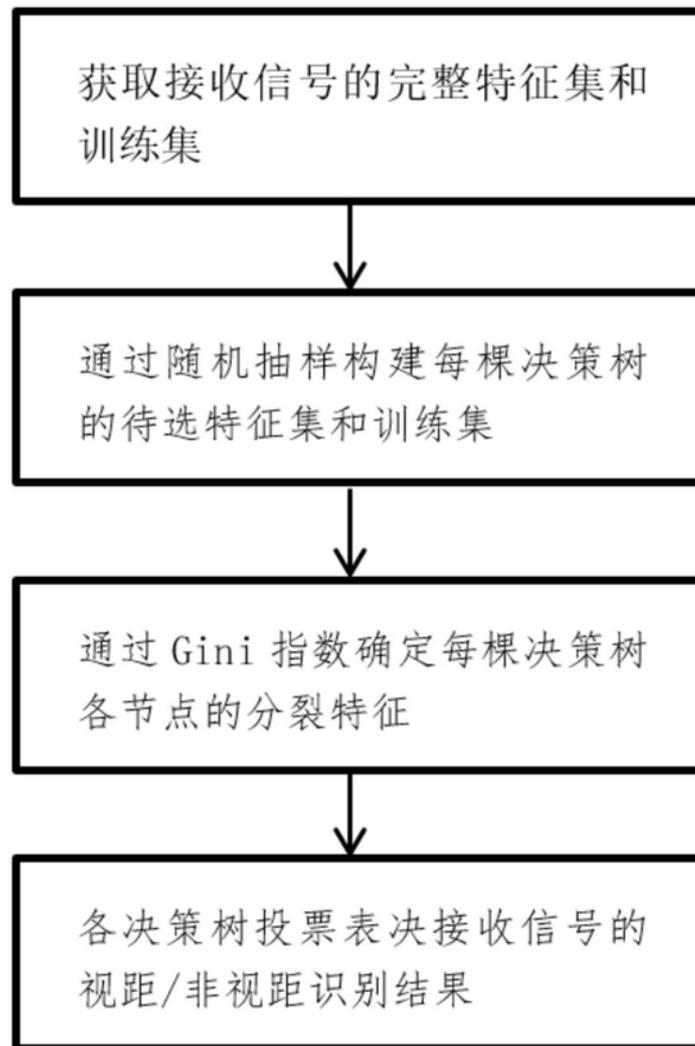


图1