

基于异步优势动作评价的 RFID 室内定位算法



李 丽 郑嘉利 王 哲 袁 源 石 静

广西大学计算机与电子信息学院 南宁 530004

广西多媒体通信与网络技术重点实验室 南宁 530004

(1114235262@qq.com)

摘 要 针对现有的 RFID 室内定位算法的精度容易受到环境因素影响的问题,提出了一种基于异步优势动作评价(Asynchronous Advantage Actor-critic, A3C)的 RFID 室内定位算法。该算法的主要步骤为:1)将 RFID 的信号强度 RSSI 值作为输入值,多个线程子动作网络并行交互采样学习,利用子评价网络评价动作值的优劣,使模型不断优化,找到最优信号强度 RSSI 值,并训练定位模型;子线程网络定期将网络参数异步更新到全局网络上,全局网络最后输出参考标签的具体位置,同时训练得到异步优势动作评价定位模型。2)在线定位阶段,当待测目标进入待测区域时,记录待测目标的信号强度 RSSI 值,将其输入异步优势动作评价定位模型中,子线程网络从全局网络中获取最新定位信息,对待测目标进行定位,最后输出目标的具体位置。实验数据表明,基于异步优势动作评价的 RFID 室内定位算法与传统的基于向量机(Support Vector Machines, SVM)定位、基于极限学习机(Extreme Learning Machine, ELM)定位、基于多层神经网络定位(Multi-Layer Perceptron, MLP)的 RFID 室内定位算法相比,定位平均误差分别下降了 66.114%, 50.316%, 44.494%;定位稳定性分别平均提高了 59.733%, 53.083%, 43.748%。实验结果表明,基于异步优势动作评价的 RFID 室内定位算法在处理大量室内定位目标时具有较好的定位性能。

关键词: RFID; RSSI; 强化学习; 异步优势动作评价; 室内定位

中图法分类号 TP301.6

RFID Indoor Positioning Algorithm Based on Asynchronous Advantage Actor-Critic

LI Li, ZHENG Jia-li, WANG Zhe, YUAN Yuan and SHI Jing

School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China

Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China

Abstract In view of the fact that the accuracy of existing RFID indoor positioning algorithm is easily affected by environment factors and the robustness is not strong, this paper proposed an RFID indoor positioning algorithm based on asynchronous advantage actor-critic (A3C). The main steps of the algorithm are as follows. Firstly, the RSSI value of RFID signal strength is used as the input value. The multi-thread sub-action network parallel interactive sampling learning, and the sub-evaluation network evaluates the advantage and disadvantage of the action value, so that the model is continuously optimized to find the best signal strength RSSI and trains the positioning model. The sub-thread network updates the network parameters to the global network on a regular basis, and the global network finally outputs the specific location of the reference tag, at the same time the asynchronous advantage actor-critic positioning model is trained. Secondly, in the online positioning stage, when the target to be tested enters the area to be tested, the signal strength RSSI value of the object to be tested is recorded and input into the asynchronous advantage actor-critic positioning model. The sub-thread network obtains the latest positioning information from the global network, locates the side target, and finally outputs the specific position of the target. RFID indoor positioning algorithm based on asynchronous advantage actor-critic was compared with the traditional RFID indoor positioning algorithm based on Support Vector Machines (SVM) positioning, Extreme Learning Machine (ELM) positioning, and Multi-Layer Perceptron positioning (MLP). Experiment results show that the mean positioning error of the proposed algorithm is respectively decreased by 66.114%, 50.316% and 44.494%; the average positioning stability is respectively increased by 59.733%, 53.083% and 43.748%. The experiment results show that the proposed RFID indoor positioning algorithm based on asynchronous advantage actor-critic has better positioning performance when dealing with a large number of indoor positioning targets.

Keywords RFID, RSSI, Reinforcement learning, Asynchronous advantage actor-critic, Indoor positioning

收稿日期:2019-01-10 返修日期:2019-05-10 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61761004)

This work was supported by the National Natural Science Foundation of China (61761004).

通信作者:郑嘉利(zjl@gxu.edu.cn)

1 引言

随着室内定位技术的不断发展,基于射频识别(Radio Frequency Identification, RFID)的室内定位技术以其非视距、非接触、快速识别等优点成为当前领域的热点研究对象^[1]。目前,RFID室内定位技术主要有4种测距思路,分别是基于信号到达时间(Time of Arrival, TOA)测距、基于信号到达时间差(Time Difference of Arrival, TDOA)测距、基于信号到达角度(Angle of Arrival, AOA)测距、基于信号到达强度(Received Signal Strength Indication, RSSI)测距^[2]。其中 TOA, TDOA, AOA 等测距法的测距误差较小,但有效距离短^[3]; RSSI 测距法的有效距离长,便于测量。因此,本文选用基于信号到达强度的测距。近年来,人们将机器学习引入 RFID 室内定位,利用机器学习的思想来提高 RFID 室内定位的精度^[4]。例如,文献[5-6]利用高斯滤波器对原始 RSSI 进行预处理,利用粒子滤波算法优化神经网络的权值和阈值,建立 RSSI 值与目标位置的关系。该方法具有比其他人工神经网络更好的性能,但是所用神经网络为浅层神经网络,不适合大量数据的定位。文献[7]将 LANDMARC 方法与前馈神经网络相结合,输入稳定的信号强度 RSSI 值,利用 LANDMARC 定位方法进行位置估计,利用神经网络调整优化定位模型,其训练得到的室内定位模型可用于难识别区域的关键目标定位。文献[8-9]利用萤火虫群优化算法在不同初始条件下快速获得最优正则化权值,同时结合极限学习机的半监督特点来减少标记参考目标的数量,从而降低了定位成本,提高了定位精度。当同时定位多个目标时,文献[10]提出将多维度尺度变换与指纹定位算法相结合来计算目标与距离的秩,然后得到目标的相对坐标,最后定位时利用坐标系匹配算法得到目标的绝对坐标。该算法的定位误差低于传统算法的误差。文献[11]根据传输距离将目标进行分组,利用同组目标在传播距离上具有的相似性,使用聚类算法对待测目标进行测距,在定位成本和灵活性方面优于其他算法。

本文提出了一种新型的室内定位算法,即基于异步优势动作评价的 RFID 室内定位算法,其主要包括室内定位模型训练和在线定位两个阶段。首先,建立全局网络和多个线程子网络,且多个线程子网络并行学习,保证每个线程环境设定不同,选出最优 RSSI 值;然后,训练室内定位模型,每隔一段时间将线程子网络更新参数异步更新到全局网络中;最后,得到 RFID 室内定位模型,并输出参考标签的坐标。在线定位时,输入待测目标的 RSSI 值,保证定位模型线程子网络的参数与全局网络参数相同,并定位输出待测目标的具体坐标值,从而有效解决了大量定位目标难定位的问题。

2 基于异步优势动作评价的 RFID 室内定位算法

实际的 RFID 室内定位中存在着多径效应、障碍物等因素,且室内定位环境不同, RSSI 信号的强度值也不同,若使用固定的损耗系数和噪声干扰,会造成很大的测距误差。现有的算法大都针对少量的 RFID 定位目标;对于大量定位目标,浅层神经网络不能很好地定位出目标位置,并且其精度容易受到环境因素的影响。强化学习^[12]作为机器学习的分支,指智能体在完成某项任务时,首先通过动作与周围环境进行交

互,在动作和环境的作用下,智能体会产生新的状态,同时环境会给出一个立即回报,经过数次迭代后,智能体最终学到完成相应任务的最优动作。将强化学习中的动作-评价方法与神经网络相结合,建立异步优势动作评价算法,并将其用于 RFID 室内定位。在 RFID 室内定位中,多个 RFID 目标定位的问题可被看作一个连续的强化学习动态决策过程,在遇到大量定位目标时,可以根据定位环境不断更新定位模型,实现动态定位,提高定位精度和效率。

2.1 RFID 室内定位模型的训练

首先,对定位区域的参考标签进行数据采集,记录参考标签的原始信号强度 RSSI 值,根据初始动作 a_t 获得初始状态 s_t 并预处理得到相应的特征输入。特征输入可用动作决策函数 $f(\text{RSSI}|s_t)$ 来表示。建立全局动作网络 $\psi(a_t|s_t; \theta_\psi)$:

$$\psi(a_t|s_t; \theta_\psi) = \frac{1}{\tau(s_t, \theta_\psi) \sqrt{2\pi}} \exp\left(-\frac{(a_t - \omega(s_t, \theta_\psi))^2}{2\tau(s_t, \theta_\psi)^2}\right) \quad (1)$$

由于定位动作环境是连续的,因此将全局动作网络定义为高斯函数。 $\omega(s_t, \theta_\psi)$ 和 $\tau(s_t, \theta_\psi)$ 为全局动作网络的输出参数,近似为参数化的线性函数:

$$\omega(s_t, \theta_\psi) = \theta_\psi^T f(\text{RSSI}|s_t) \quad (2)$$

$$\tau(s_t, \theta_\psi) = \exp(\theta_\psi^T f(\text{RSSI}|s_t)) \quad (3)$$

其中, θ_ψ 为全局动作网络的动作决策参数。建立全局评价网络 $V(s_t; \theta_v)$, 即对全局网络预期的折扣回报求期望。期望服从贝尔曼方程:

$$V(s_t; \theta_v) = E[R_t + \gamma V(s_{t+1}; \theta_v)] \quad (4)$$

其中, γ 为折扣因子, θ_v 为全局评价网络的估计参数。

建立多个线程子动作网络 $\psi'(a_t|s_t; \theta_{\psi'})$:

$$\psi'(a_t|s_t; \theta_{\psi'}) = \frac{1}{\tau(s_t, \theta_{\psi'}) \sqrt{2\pi}} \exp\left(-\frac{(a_t - \omega(s_t, \theta_{\psi'}))^2}{2\tau(s_t, \theta_{\psi'})^2}\right) \quad (5)$$

其中, $\omega(s_t, \theta_{\psi'})$ 和 $\tau(s_t, \theta_{\psi'})$ 为线程子动作网络的输出参数:

$$\omega(s_t, \theta_{\psi'}) = \theta_{\psi'}^T f(\text{RSSI}|s_t) \quad (6)$$

$$\tau(s_t, \theta_{\psi'}) = \exp(\theta_{\psi'}^T f(\text{RSSI}|s_t)) \quad (7)$$

其中, $\theta_{\psi'}$ 为线程子动作网络的动作决策参数。

建立多个线程子评价网络 $V(s_t; \theta_{v'})$, 即对子线程网络预期的折扣回报求期望。期望服从贝尔曼方程:

$$V(s_t; \theta_{v'}) = E[r_t + \gamma V(s_{t+1}; \theta_{v'})] \quad (8)$$

其中, $\theta_{v'}$ 为线程子评价网络的估计参数。

然后,多个线程与各自的训练环境交互学习,得到相应的信号强度 RSSI 值序列样本,使用序列样本完成定位模型的训练。定位模型根据 RSSI 值和定位坐标建立非线性关系,训练出参考标签坐标,得到线程总回报值:

$$R = \sum_{i=1}^n (\gamma^{i-1} r_i + V(s_i; \theta_{v'})) \quad (9)$$

其中, R 表示当前线程的总回报值, r_t 表示当前子动作的回报值。再计算每个线程的优势函数 $A(s_t; \theta_{v'})$, 用优势函数来提升评价网络。评价网络用于评价定位模型的优劣,可表示为:

$$A(s_t; \theta_{v'}) = r_t + \gamma r_{t+1} + \dots + \gamma^{n-t} r_n + \gamma^{n-t} V(s_n; \theta_{v'}) - V(s_t; \theta_{v'}) \quad (10)$$

最后,定位模型再累加线程子动作网络及线程子评价网络的梯度:

$$1) \text{ 累加线程子动作网络梯度, 总和为全局动作网络梯度:} \\ d\theta_\psi = \frac{1}{n} \sum_{i=1}^n (\nabla_{\theta_\psi} \log \psi(a_i|s_i; \theta_\psi) A(s_i; \theta_{v'}) + F) \quad (11)$$

其中, $F = \beta \nabla_{\theta_{\psi}} H(\psi'(a_t | s_t; \theta_{\psi}'))$, 为交叉熵, 其使动作拥有更好的探索性; β 为交叉熵的权重。

2) 累加线程子评价网络梯度, 总和为全局评价网络梯度:

$$d\theta_v = \frac{1}{n} \sum_{t=1}^n (\nabla_{\theta_v} (R - V(s_t, \theta_v'))^2) \quad (12)$$

当线程子动作网络及线程子评价网络计算并累加完梯度时, 再反向传播, 对全局网络参数进行异步更新, 进行多次迭代, 最终异步优势动作评价定位模型训练完成, 输出参考标签的具体坐标值, 得到异步优势动作评价定位模型。

2.2 RFID 室内定位模型的在线定位

在线定位阶段需要过滤 RSSI 测量数值中不稳定、非正常的数, 选出最优 RSSI 值, 才能得到高精度的待测目标位置信息。因此, 采用 m 次测量法对在线定位阶段数据进行相应的预处理。首先, 获取有效的 RSSI 值原始信息; 然后, 对连续 m 次测得的待测目标 RSSI 值进行累加, 得到最终输入值; 其次, 根据输入值不断调整网络参数, 动态更新网络以适应环境变化; 最后, 确定待测目标的位置。若采用 m 次测量, m 取值越大, 得到的有效数据越多, 意味着设备测量时间就越长, 从而响应用户体验。因此, m 值的计算与读写器在该位置所能接收到的待测目标 RSSI 值的概率、参数 e 和 T 直接相关。RSSI 值的接收概率由环境决定, e 反映测试的可靠性, T 反映用户对信号变化的容忍程度, m 值的计算如下:

$$1 - (1 - P(E(x) - e < x < E(x) + e))^m = T \quad (13)$$

$$m = \log_{P(E(x) - e < x < E(x) + e)} (1 - T) \quad (14)$$

其中, x 表示所测的 RSSI 值。

异步优势动作评价的 RFID 室内定位算法的流程如图 1 所示。

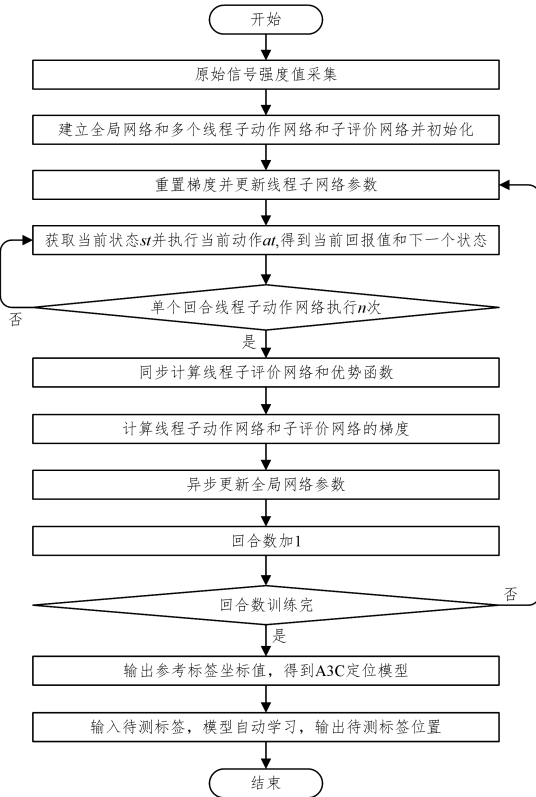


图 1 基于异步优势动作评价的 RFID 室内定位算法的流程
Fig. 1 Flow of RFID indoor localization algorithm based on asynchronous advantage actor-critic

2.3 算法参数的设置

在本文算法中, 参考目标用于训练 RFID 室内定位模型, 待测目标用于测试 RFID 室内定位模型的定位能力。用 Tensorflow^[13-14] 架构实现整个定位模型, 所使用的神经网络为前馈神经网络^[15]。在前馈神经网络中, 三层结构足以产生任意复杂的决策区域, 因此本文中前馈神经网络设置为三层神经网络, 分别为 RSSI 输入层、隐藏层、输出层, 其中输入层大小设置为 8×8 。在隐藏层中选择合适数量的神经元是整个定位网络的关键, 隐藏层神经元设置为 $64 \times 64 \times 64$, 训练隐藏层收敛于输出层; 输出层的神经元为 2, 对应于目标坐标 (x, y) 。非线性激活函数为非线性整流函数 (Rectified Linear Unit, ReLU)^[16], 异步优势动作评价定位网络的参数设置如表 1 所列。

表 1 异步优势动作评价定位网络的参数

Table 1 Parameters of asynchronous advantage actor-critic

localization network	
超参数	取值
贪婪度 ϵ	0.1
线程数 K	32
优化器	共享 RMSprop
激励函数	ReLU
回合数 M	1000000
折扣因子	0.9
学习率	1×10^{-5}
动量 α	0.2

为了解决异步优势动作评价定位模型选择最优 RSSI 值的问题, 对定位模型的动作、状态、奖励分别进行如下设置。

动作设置: 动作的决策函数 a_t 为高斯函数, 按合理的距离移动, 若 $RSSI_{\text{present}} > RSSI_{\text{best}}$, 则 $RSSI_{\text{best}} = RSSI_{\text{present}}$, 将当前值复制给最优值, 且当前动作选择最优 RSSI 值。动作决策函数表示为:

$$f(RSSI | s_t) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{RSSI - B}{2\sigma^2}\right) \quad (15)$$

其中, $RSSI$ 表示在状态 s_t 下获取的所有 RSSI 值, 表示为:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (RSSI_i - B)^2} \quad (16)$$

其中, N 表示所有的 RSSI 值, $RSSI_i$ 表示第 i 个 RSSI 值, B 表示平均值:

$$B = \frac{1}{N} \sum_{i=1}^N RSSI_i \quad (17)$$

状态设置: 状态是一个代理可以观察到的一组状态, 本文中的状态是临近坐标点的 RSSI 值。状态变化时, 动作不随机, 需要在参考目标的数据范围内从上一个目标移动到下一个目标。

奖励值设置: 将误差值取反后作为奖励值, 误差值越小, 奖励值越大。累积奖励值, 不断优化定位模型。累积奖励值越大, 定位模型的定位精度越高。

梯度更新设置: 线程子动作网络和线程子评价网络采用的梯度方法为随机梯度法^[17]。在梯度下降时, 对于梯度震荡较大的项, 减慢其下降速度; 对于震动幅度小的梯度, 加快其下降速度。每一次训练都快速遍历全部信号的强度值, 以达到全局最优训练。优化器为共享 RMSprop^[18], 梯度更新过程如下:

$$g = \alpha g + (1 - \alpha) \Delta \theta_{\psi}^2 \quad (18)$$

$$\theta_{\psi} = \theta_{\psi} - \eta \frac{\Delta \theta_{\psi}}{\sqrt{g + \delta}} \quad (19)$$

其中, g 表示平方梯度; α 表示动量; η 表示学习率; δ 表示一个很小的常数, 一般取值为 1×10^{-6} , 以防止分母为零。 g 在线程之间共享, 使线程间共享数据信息。 每次计算梯度时, 线程子动作网络直接计算全局动作网络梯度和全局评价网络梯度, 然后用优化器对全局网络参数进行异步更新。

3 实验结果及分析

实验中使用的 RFID 读写器类型为 Impinj Speedway Revolution R420 UHF RFID, 主要工作频率为 860 ~ 960 MHz, 发射功率为 10 ~ 31 dB, 灵敏度为 -84 dB, 天线端口为 4 端口, 其中端口可拓展到 32 个, 可处理多种频段或复杂环境下的定位目标数据并返回定位目标信号的强度值。 定位环境采用 10m × 10m 的室内矩形区域, 设置读写器天线 M 为 4, 标签数 N 为 732。 在室内环境建立直角坐标系, 以一个顶点为坐标原点 $O(0,0)$, 与原点相交的两条边为 x 轴和 y 轴。 4 个读写器放置在区域的 4 个顶点上, 坐标为 $(0,0)$, $(0,10)$, $(10,0)$, $(10,10)$, 每个读写器天线的摆放角度为 45° , 以获得更大的信号强度 RSSI 值。 为了分析 A3C 定位算法的性能, 本文采用距离误差来表示定位误差:

$$d_{error_j} = \sqrt{(x_j - x_j^i)^2 + (y_j - y_j^i)^2} \quad (20)$$

其中, (x_j, y_j) 是第 j 个目标的估计位置, (x_j^i, y_j^i) 为第 j 个目标的实际位置。 由于参考目标数量较大, 因此计算其平均误差:

$$average = \frac{1}{N} \sum_{i=1}^N d_{error_j} \quad (21)$$

其中, $average$ 表示定位的平均误差, N 为总的定位目标数。

3.1 环境因素影响

在训练离线模型时, 为了测试外界环境对定位精度的影响, 以标签密度为测量标准, 将标签密度分别设为 1.0 m, 0.9 m, 0.8 m, 0.7 m, 0.6 m, 0.5 m, 0.4 m, 0.3 m 进行一次实验, 不同标签密度内的标签个数如表 2 所列, 一次实验测得 3168 个数据, 共测量 4 次, 总数据量为 12672。

表 2 不同密度下的参考标签数量

Table 2 Number of reference labels under different densities

密度/m	数量/个
1.0	96
0.9	144
0.8	206
0.7	306
0.6	421
0.5	598
0.4	665
0.3	732

如图 2 所示, 当参考标签的密度从 1.0 m 降低到 0.8 m 时, 平均定位误差较大; 但从 0.8 m 降低到 0.5 m 时, 平均定位误差显著减小; 当参考标签的密度从 0.5 m 降低到 0.3 m 时, 平均定位误差逐渐减小。 因此, 参考标签的密度越大且数量越多, 定位精度越高。 实验结果表明, 对于大量标签定位, A3C 具有较好的定位效果。

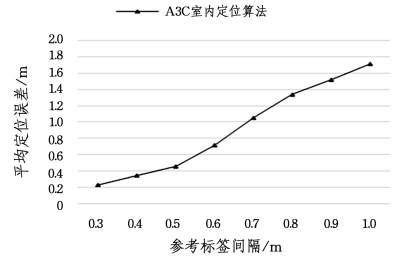


图 2 参考标签密度的影响

Fig. 2 Influence of density of reference tags

训练离线模型时, 为了测试训练集对 A3C 室内定位算法的影响, 将 12672 个数据分成不同大小的训练集。 如图 3 所示, 当训练规模小于 3168 个数据时, 平均定位误差减小得较为缓慢; 但当数据量从 3168 增加到 7920 时, 训练得到的平均定位误差迅速减小; 当数据量从 7920 增加到 11088 时, 平均定位误差减小的速度减慢; 当训练数据多于 11089 个时, 平均定位误差随着数据的增加基本保持不变, A3C 模型开始收敛。

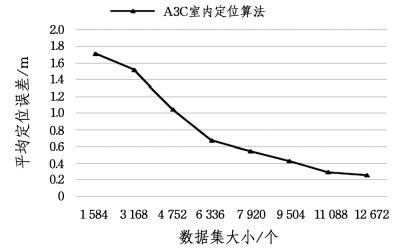


图 3 不同训练集的大小对定位模型的影响

Fig. 3 Impact of training data size on location accuracy

在线定位时, 为了测试测量次数 m 对定位精度的影响, 对 732 个定位目标进行定位, 间隔为 0.3 m, 将 m 值从 1 增加到 9, 并对不同的 e 和 T 进行对比分析。 e 反映测试的可靠性, T 反映用户对信号变化的容忍程度。 实验结果如图 4 所示, 可以看出, 随着 m 值及 e 和 T 的不断增大, 定位精度不断提高。

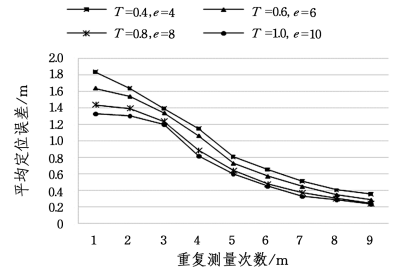


图 4 不同参数下重复测量次数对定位精度的影响

Fig. 4 Impact of number of measurements on location accuracy with different parameters

3.2 其他算法的比较

为了测试 A3C 室内定位模型的性能, 分别从模型的定位精度、稳定性、时间复杂度和定位效率方面将其与基于 SVM 定位^[19]、基于 ELM 定位^[20]、基于 MLP 定位^[21]的算法进行比较。

3.2.1 定位精度

定位精度表征测试模型的定位能力, 定位误差越小, 表示

定位能力越好。在线定位阶段,模型的定位误差分布如图 5 和表 3 所示。图 5 中,“×”表示实际标签坐标;“+”表示计算标签坐标;横轴表示标签坐标 x ;纵轴表示标签坐标 y ;圆圈表示实际坐标到计算坐标的误差,圆越大,表示定位误差越大。

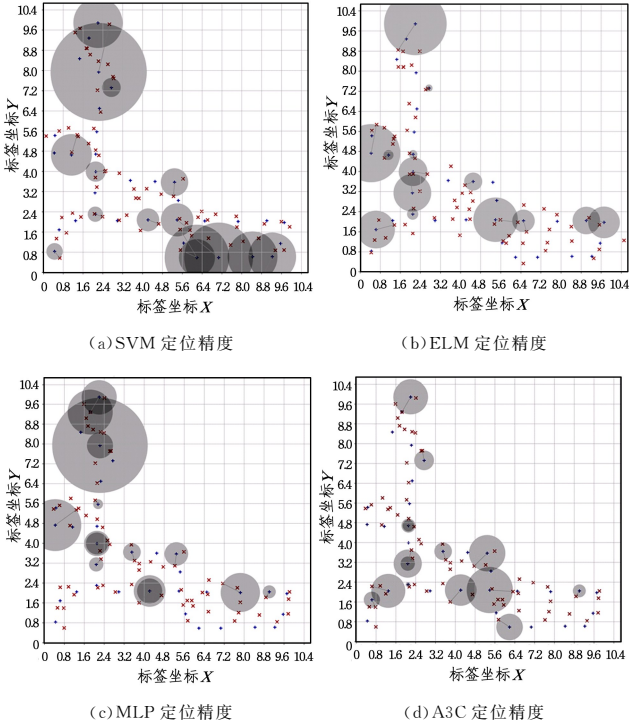


图 5 4 种算法模型定位精度的对比

Fig. 5 Localization accuracy comparison of four algorithm models

表 3 4 种算法在在线定位阶段的定位结果

Table 3 Localization result of four algorithm models in online locating phase

(单位:m)				
算法	最大值	最小值	平均值	标准差
A3C	1.7162	0.0030	0.2984	0.3713
MLP	2.0042	0.0194	0.5376	0.6531
ELM	2.3251	0.0046	0.6006	0.7914
SVM	2.5722	0.0043	0.8806	0.9221

A3C 算法在在线定位阶段的定位能力最好,MLP 的在线定位能力其次,ELM 和 SVM 的在线定位能力较差。A3C 定位模型与 SVM,ELM,MLP 定位模型相比,在线定位的平均误差分别下降了 66.114%,50.316%,44.494%。这一结果表明,浅层神经网络的建模能力有限,随着环境中定位目标的增大,其在处理大量 RSSI 信号强度信号时效果并不理想;而本文基于 A3C 的定位模型可以有效解决数据量大的问题。由表 3 可知,SVM,ELM,MLP 和 A3C 模型在在线定位时的最大定位误差分别为 2.5722m,2.3251m,2.0042m,1.7162m。出现最大定位误差是由于目标在边缘或角落时,信号强度值减弱,导致定位误差增大。

3.2.2 稳定性

稳定性表征数据的波动程度,可用标准差来测量,标准差越小,表示定位数据的波动越小,稳定性越高,如表 3 所列。A3C 的标准差最小,相比 SVM,ELM,MLP,标准差分别降低了 0.5508m,0.5427m,0.2818m,相当于稳定性提高了

59.733%,53.083%,43.748%。这也表明,浅层神经网络如 ELM 及 SVM 在处理大量定位信号时效果并不理想,而本文基于异步优势动作评价的室内定位方法可以随着环境的变化不断调整网络;同时,并行处理数据能够有效克服数据量大的问题,提高了定位的稳定性。

3.2.3 时间复杂度

表 4 比较了 4 种算法的模型训练时间和在线定位时间。ELM 的离线训练时间最短,A3C 的离线训练时间最长。A3C 训练无监督的特点导致模型训练时间增加,其总训练时间为 204.382 s,比 MLP 增加了 33.780%,比 ELM 增加了 59.258%,比 SVM 增加了 26.501%。但 A3C 的在线定位时间为 0.1340 s,比 MLP 减少了 76.872%,比 ELM 减少了 84.608%,比 SVM 减少了 82.832%。实验结果表明,A3C 可以解决在线阶段定位繁琐的问题。

表 4 4 种定位算法的计算时间

Table 4 Computing time of four localization algorithm

(单位:s)		
算法	离线训练时间	平均在线定位时间
A3C	204.382	0.1340
MLP	135.340	0.5794
ELM	83.269	0.8706
SVM	150.218	0.7805

3.2.4 定位效率

为了比较定位效率,设置定位目标个数为 770,对比 4 种算法在定位时间不断增加时能准确定位的目标个数。由图 6 可知,与 MLP,ELM,SVM 算法相比,A3C 定位模型在单位时间内能够定位的目标数较多,并且随着定位时间的不断增加,其可以逐渐定位出所有目标,在线定位效率较高;而 MLP,ELM 和 SVM 算法在定位目标数达到一定值时,随着时间的增加,定位出的目标个数基本不变,定位效率较差。ELM 和 SVM 在面对大量数据时泛化能力较差,因此定位效率较低。

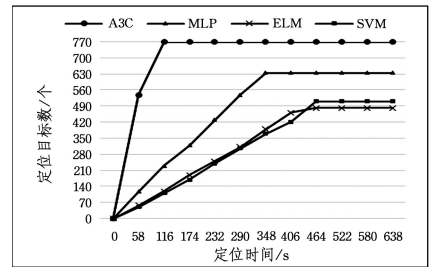


图 6 4 种算法的定位效率

Fig. 6 Localization efficiency of four algorithm

结束语 针对 RFID 室内定位,传统定位算法在定位目标数量较大时,定位精确度较差,因此本文引入异步优势动作评价的方法进行室内定位学习,包括模型训练阶段和在线定位阶段。在模型训练阶段,输入参考标签的信号强度值,进行多线程并行学习,训练定位模型,从而加快模型训练速度。在在线定位阶段,直接输入待测目标 RSSI 值,快速定位待测目标的位置。实验结果表明,与传统室内定位算法相比,本文算法在预测能力、定位精度和稳定性等方面都更优。

未来我们将着眼于 RFID 室内定位模型的训练速度和模

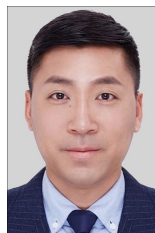
型的稳定性问题,引入深度确定性网络来训练 RFID 定位模型,通过确定性动作网络选择最优 RSSI 值,评价网络判断动作的优劣,同时引入迁移学习,将强化学习和监督学习相结合,用于 RFID 室内定位,以进一步提高定位的准确度。

参 考 文 献

- [1] SHI J Y, QIN X L, WANG L. Gradient and Constant-game Based RFID Indoor Localization Algorithm [J]. Computer-Science, 2015, 42(11): 138-143.
- [2] ZHENG J, YANG Y, HE X, et al. Multiple-port reader antenna with three modes for UHF RFID applications [J]. Electronics Letters, 2018, 54(5): 264-266.
- [3] LIU K, ZHANG W, ZHANG W D, et al. A Wireless Positioning Method Based on Deep Neural Network [J]. Computer Engineering, 2016, 42(7): 82-85.
- [4] YANG Y N, XIA B, YUAN W, et al. Research on Ranging Algorithm Based on Convolution Neural Network [J]. Journal of Chongqing University of Technology (Natural Science), 2018 (3): 172-177.
- [5] WANG C, WU F, SHI Z, et al. Indoor positioning technique by combining RFID and particle swarm optimization-based back propagation neural network [J]. Optik - International Journal for Light and Electron Optics, 2016, 127(17): 6839-6849.
- [6] WANG C, SHI Z, WU F, et al. An RFID indoor positioning system by using Particle Swarm Optimization-based Artificial Neural Network [C] // 2016 International Conference on Audio, Language and Image Processing (ICALIP). IEEE Computer Society, 2017: 738-742.
- [7] KUNG H Y, CHAISIT S, PHUONG N T M. Optimization of an RFID location identification scheme based on the neural network [J]. International Journal of Communication Systems, 2015, 28(4): 625-644.
- [8] JIANG X, LIU J, CHEN Y, et al. Feature Adaptive Online Sequential Extreme Learning Machine for lifelong indoor localization [J]. Neural Computing & Applications, 2016, 27(1): 215-225.
- [9] LIU F, ZHONG D. GSOS-ELM: An RFID-Based Indoor Localization System Using GSO Method and Semi-Supervised Online Sequential ELM [J]. Sensors, 2018, 18(7): 1995.
- [10] GAO Z, MA Y, LIU K, et al. An Indoor Multi-tag Cooperative Localization Algorithm Based on NMDS for RFID [J]. IEEE Sensors Journal, 2017, 17(7): 2120-2128.
- [11] ZHAO Y, LIU K, MA Y, et al. Similarity Analysis-Based Indoor Localization Algorithm With Backscatter Information of Passive UHF RFID Tags [J]. IEEE Sensors Journal, 2016, 17(99): 1-1.
- [12] SUTTON R, BARTO A. Reinforcement Learning: An Introduction (second edition) [M]. The MIT Press, 2018.
- [13] MURRAY D G, MURRAY D G. A computational model for TensorFlow: an introduction [C] // Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Language. New York: ACM, 2017: 1-7.
- [14] ABADI M. TensorFlow: learning functions at scale [J]. Acm Sigplan Notices, 2016, 51(9): 1-1.
- [15] SCHMIDHUBER J. Deep learning in neural networks: An overview [J]. Neural Network, 2015, 61(5): 85-117.
- [16] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. arXiv:1409.1556, 2014.
- [17] SONG R, LEWIS F, WEI Q, et al. Multiple actor-critic structures for continuous-time optimal control using input-output data [J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(4): 851-865.
- [18] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous Methods for Deep Reinforcement Learning [J]. arXiv:1602.01783v2, 2016.
- [19] BURTON A, PARIKH T, MASCARENHAS S, et al. Driver identification and authentication with active behavior modeling [C] // 12th International Conference on Network and Service Management (CNSM). IEEE Computer Society, 2017: 388-393.
- [20] ALARIFI A, ALSALMAN A M, ALSALEH M, et al. Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances [J]. IEEE Sensors, 2016, 16(5): 1-36.
- [21] ZHAI X, ALI A A S, AMIRA A, et al. MLP Neural Network Based Gas Classification System on Zynq SoC [J]. IEEE Access, 2017, 4(99): 8138-8146.



LI Li, born in 1994, postgraduate. Her main research interests include information processing and communication networks, reinforcement learning and internet of things.



ZHENG Jia-li, born in 1979, professor. His main research interests include internet of things, RFID and artificial intelligence.