

Research on WIFI Probe Data Preprocessing for Indoor Location

Zhenya Zhang¹, Huaqian Cao¹, Qinqin Nie¹, Mengjie Dong¹, Hongmei Cheng², Ping Wang¹

¹Anhui Province Key Laboratory of Intelligent Building & Building Energy Saving, Anhui Jianzhu University, Hefei Anhui

²School of Economics & Management, Anhui Jianzhu University, Hefei Anhui
Email: zyzhang@ahjzu.edu.cn

Received: Feb. 9th, 2018; accepted: Feb. 21st, 2018; published: Feb. 28th, 2018

Abstract

To verify the feasibility of identifying whether a WIFI terminal is within a designated indoor area based on RSSI value collected by multi WIFI detector, data preprocessing task for the construction of detection data set is divided into three phase such as WIFI probe data parsing, time frame Numbering for probe data and the construction of indoor occupant location oriented data set in this paper. Flow charts for those three phases are given. In this paper, a BP neural network based discriminator for the identification that whether a WIFI terminal is within a designated indoor area is implemented with multi RSSI vector as input. And experimental results show that the precision of the discriminator for indoor area location is high. It is feasibility of identifying whether a WIFI terminal is within a designated indoor area based on multi RSSI detected.

Keywords

WIFI Detector, Received Signal Strength Indication, Discriminator, BP Neural Network

面向室内定位的WIFI探针数据预处理研究

张振亚¹, 操华茜¹, 聂芹芹¹, 董梦杰¹, 程红梅², 王 萍¹

¹安徽建筑大学智能建筑与建筑节能安徽省重点实验室, 安徽 合肥

²安徽建筑大学经济与管理学院, 安徽 合肥
Email: zyzhang@ahjzu.edu.cn

收稿日期: 2018年2月9日; 录用日期: 2018年2月21日; 发布日期: 2018年2月28日

摘 要

为验证依据多探针同时感知到的同一WIFI终端的RSSI值辨识WIFI终端是否在指定区域内的可行性, 本文

文章引用: 张振亚, 操华茜, 聂芹芹, 董梦杰, 程红梅, 王萍. 面向室内定位的WIFI探针数据预处理研究[J]. 人工智能与机器人研究, 2018, 7(1): 34-42. DOI: 10.12677/airr.2018.71004

围绕多探针数据集的构造将WIFI终端数据预处理流程划分为探针探测数据集解析、探针探测数据时间帧编号、面向室内人员定位的探针数据构造等三个阶段,并设计了相关预处理任务的流程并进行了实现。实验结果表明,以预处理后的数据为输入,基于BP神经网络的判别器可以以很高的准确率判别WIFI终端是否在指定的区域内,依据多探针同时感知到的同一WIFI终端的RSSI值辨识WIFI终端是否在指定区域内是可行的。

关键词

WIFI探针, 接收信号的强度指示, 判别器, BP神经网络

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

通常,建筑物内人员数量、人员流量和人员位置分布都是建筑节能实现可利用的重要基础数据[1] [2] [3]。准确获知建筑物内人员的相关准确信息(包含数量、流量、位置、分布等)对建筑的智能运行以及建筑节能管理具有重要意义[4]。近目前,室内人员计数研究及其在建筑节能领域的应用已成为学术界和工业界的研究热点[5] [6]。其中,无需额外的硬件发射设备、基于WIFI技术的人员定位以其廉价、无需待估计人员主动参与的被动式室内人员自动计数方法由为研究者和工业界所关注[7] [8]。

WIFI通信过程中,只要WIFI模块开启,无论是否已经成功接入到某个AP,终端都会间歇性地、明文形式的突发地广播WIFI探测请求(probe request)帧。该帧中含有终端的MAC地址、接收信号强度(RSSI)等信息。WIFI探针即利用WIFI探测请求帧的明文特性,监测并解析出其中的有用信息。为实现建筑物内指定区域内人员数量比较准确的计量,我们设计了一个基于多个WIFI探针的室内人员技术系统。本文讨论了使用多个WIFI探针的探测数据计量建筑物内指定区域内人员数量时的数据预处理过程,并使用了一个基于BP神经网络的判别器对依据多探针同时感知到的同一WIFI终端的RSSI值辨识WIFI终端是否在指定的区域内的可行性进行了实验验证。

2. 探针探测数据集解析

系统中,每个探针周期性的将探测到的WIFI终端信息发送到数据库服务器,数据库服务器中,探针探测到的一个WIFI终端的信息被保存到表格TZD中。表格TZD定义了ID、tID、tData、tTime四个字段,其中:ID字段为增量为1的标识字段,用于标识每一条被探测到的WIFI终端的信息;tID为探测到该条WIFI终端的信息的探针的名字;tData为探针探测到的WIFI终端相关信息;tTime为该条探测信息被保存到数据库的时间。依据探针数据协议中规定的格式,需要将存储在tData字段中的内容解析为WIFI终端MAC地址、AP点MAC地址、信号大类类别、信号小类类别、传输信道、信号强度等信息。最终,表格TZD的每一行被保存为表格Analyze_LocTZdata的一行。数据集解析的流程如图1示意。

表格Analyze_LocTZdata的字段包括timeFrameCounter、ID、tID、tData、tTime、WIFI终端MAC地址、AP点MAC地址、信号大类类别、信号小类类别、传输信道、信号强度、解析时间、终端位置locClass、终端水平位置locX、终端水平位置locY。其中,ID、tID、tData、tTime字段的取值为表格TZD对应字段的取值;WIFI终端MAC地址、AP点MAC地址、信号大类类别、信号小类类别、传输信道、信号强

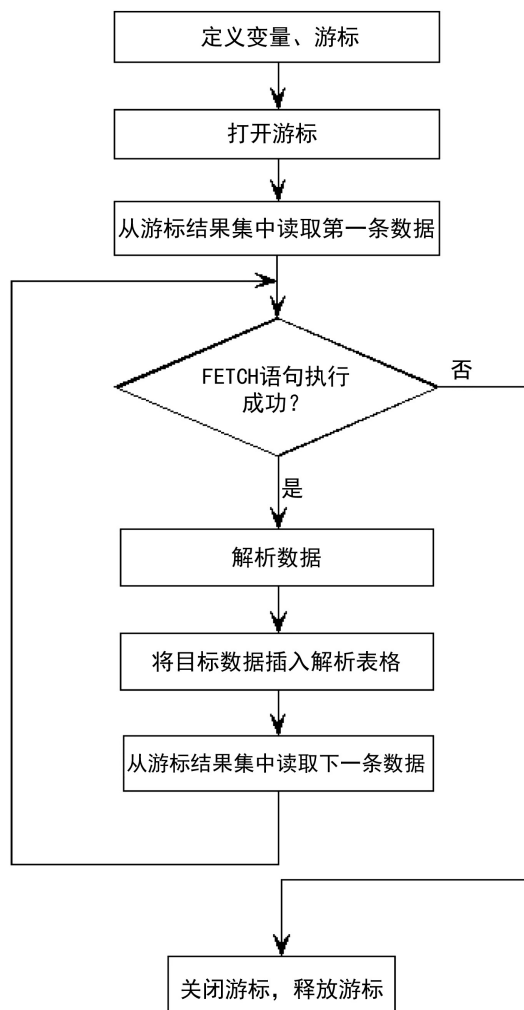


Figure 1. Flow chart for data parser

图 1. 数据解析流程

度等字段的取值从表格 TZD 中 tData 字段取值中解析获得；解析时间字段的取值为解析 tData 字段完成后的时间；终端位置 locClass 取值为 WIFI 终端所在区域的标识，取值为 0 时表示在区域外，1 表示为在区域内；终端水平位置 locX、终端水平位置 locY 字段的取值为 WIFI 终端所在位置的横坐标和纵坐标；timeFrameCounter 为数据采集时间 tTime 的时间帧编号。

探针探测数据存储表格 TZD 中的数据被解析到 Analyze_LocTZdata 表格时，timeFrameCounter、终端位置 locClass、终端水平位置 locX、终端水平位置 locY 取值均为空。其中，终端位置 locClass、终端水平位置 locX、终端垂直位置 locY 的取值需通过人工标注，而 timeFrameCounter 由专门的编码过程赋值。

3. 面向室内人员定位的探针数据构造

考虑到 WIFI 终端发送 probe 帧的随机性和突发性，本文相关的研究在使用 WIFI 探针计量建筑物内指定区域内人员数量时将一段时间内被 WIFI 探针感知到的全部 WIFI 终端信息作为一个时间帧内被探针系统感知到的 WIFI 终端信息。全部时间帧自 1 开始顺序编号。因此，对表格 Analyze_LocTZdata，需要根据每行 tTime 的取值确定 timeFrameCounter 的值。图 2 给出了对 Analyze_LocTZdata 的每一行的 timeFrameCounter 字段的赋值流程。图 2 示意的流程中，起始时间 timeStart、时间帧长度 timeInterval 是

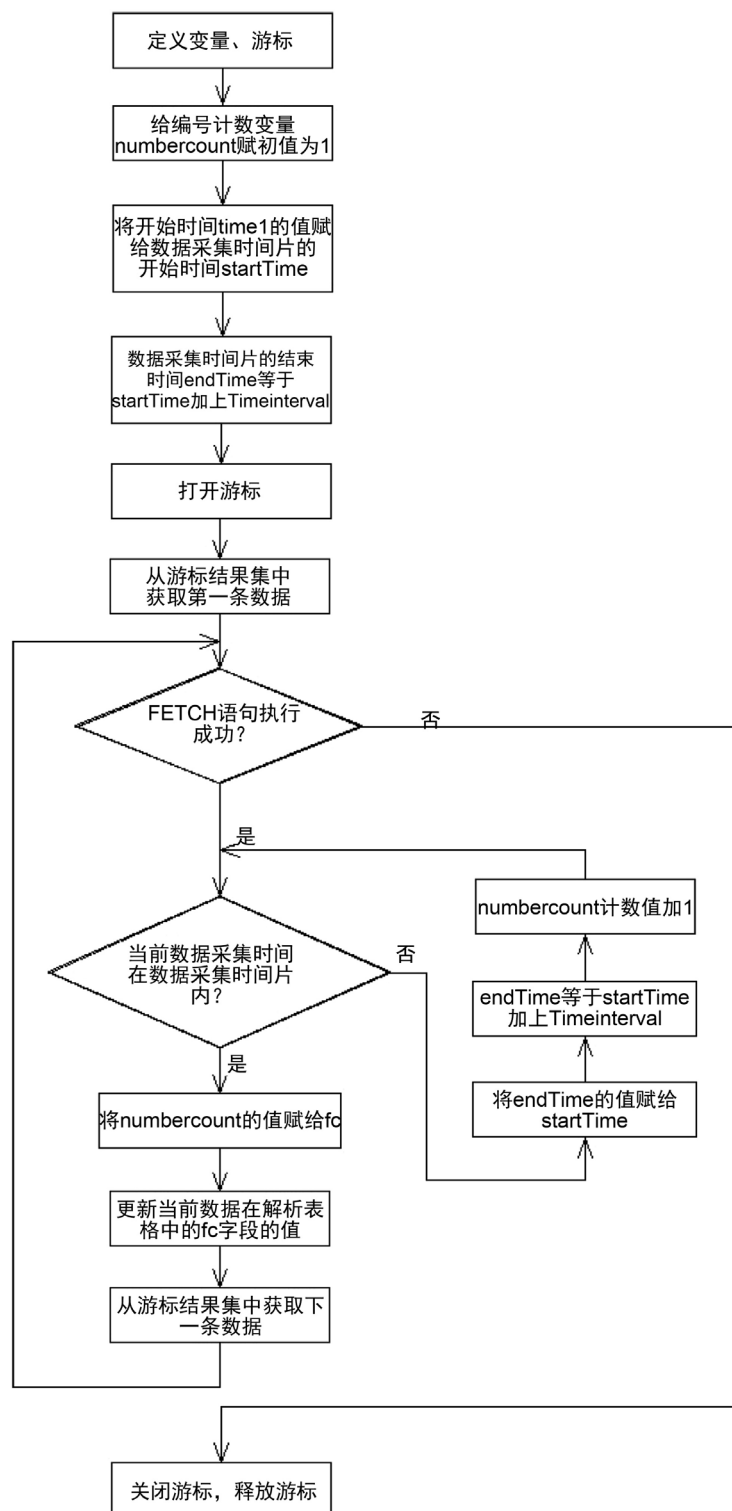


Figure 2. Flow chart for the numbering of time frame for probe data
图 2. 探针数据的时间帧计数流程

必须的输入；timeFrameCounter 的值逐 1 增加，时间格式为 YYYY-MM-DD hh:mm:ss，精确到毫秒。

为准确的判定 WIFI 探针感知到的 WIFI 终端是否在指定区域内，研究在指定区域内部署了 4 个探针，

只有当4个探针在同一时间帧内都感知到某一WIFI终端时才对该终端是否在指定区域内进行判定。所以需要从解析后的WIFI探针数据集中提取具timeFrameCounter取值相同、被4个探针分别感知的WIFI终端的数据。虽然分析的目的以及所使用的分析模型不进相同,但目标数据的基本格式可以表述为四元组 $rssiv = (RSSI_1, RSSI_2, RSSI_3, RSSI_4)$,其中 $RSSI_i$ 表示第 i 个探针感知到WIFI终端的probe帧时测得的接收信号的强度指示(Received Signal Strength Indication)值。为适合判断是否在指定区域内的室内定位模型构建,研究使用的数据是四元组 $rssiv$ 形式基本格式的扩展, $rssiv' = (RSSI1, RSSI2, RSSI3, RSSI4, locStatus)$,其中, $locClass$ 为WIFI终端所在区域的标识,取值为0时表示在区域外,1表示为在区域内。从解析后的数据表格Analyze_LocTZdata中构造目标数据集的流程由图3给出。

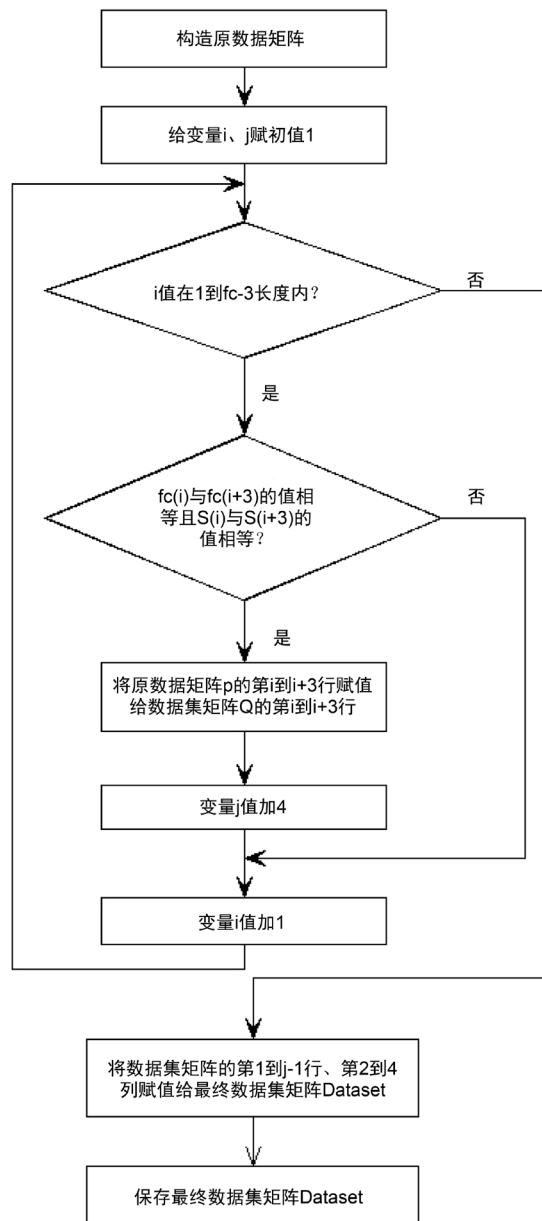


Figure 3. Flow chart for the construction of indoor occupant location oriented data set

图3. 探针定位模型数据集构造流程

图 3 中：所谓构造原始矩阵所需的列是指从 Analyze_LocTZdata 中构造具有 timeFrameCounter、WIFI 终端 MAC 地址、信号强度、终端位置 locClass 等列的子视图；而最终数据集矩阵 Dataset 是指为便于不同研究平台下数据共享而以 excel 文件格式保存好的目标数据。

4. 实验结果与分析

为快速验证基于上述探针探测数据集解析、面向室内人员定位的探针数据构造等步骤以及流程所构造的数据集的有效性，研究在 Matlab 环境中使用 BP 神经网络快速实现了一个辨识 WIFI 终端是否在指定区域判别器：1) 该判别器以四个 WIFI 探针在同一时间帧内嗅探到的同一 WIFI 终端的 RSSI 值为输入，每个 WIFI 探针对应着一个输入神经元，即该判别器使用的 BP 神经网络有四个输入神经元；2) 该判别器使用的 BP 神经网络的输出层包含一个输出神经元，该神经元输出值为 1 时表明被感知的 WIFI 终端在指定的区域内，而输出值为 0 时表明被感知的 WIFI 终端在指定的区域外；3) 该判别器使用的 BP 神经网络有 1 个隐层，包含 5 个隐含神经元；4) 该判别器使用的 BP 神经网络隐层神经元的激活函数为 $f(x) = \tan sig(x) = \frac{2}{1+e^{-2x}} - 1$ ，输出层神经元的激活函数为 $f(x) = \text{purelin}(x) = x$ 。试验中，对训练完成的判别器，若其使用的 BP 神经网络的输出值 x 时，若 $x \geq \varepsilon$ ，则认为判别器的输出为 1，否则认为判别器的输出为 0，其中 ε 为实现指定的阈值。

实验使用的数据为我们部署在智能建筑与建筑节能安徽省重点实验室物联网应用研究室的 WIFI 探针系统采集到的 WIFI 终端数据；关注的 WIFI 终端共 16 部，7 部放在室内的不同位置，9 部放在室外走廊的不同位置。

实验中，全部数据分成样本数据和测试数据两部分。实验中，对每一个采样率，判别器被重复 1000 次构造并计算判别的准确率。图 4 显示了当采样率从 0.005 到 0.995 变化时，使用基于 BP 神经网络的 WIFI 终端是否在指定区域判别器判别的准确率随着采样率的变化而变化的情况。图 4 中，横坐标表示采样率，纵坐标表示识别的准确率。图 4 中，蓝色实线表示对样本数据进行判别时判别准确率的平均值随

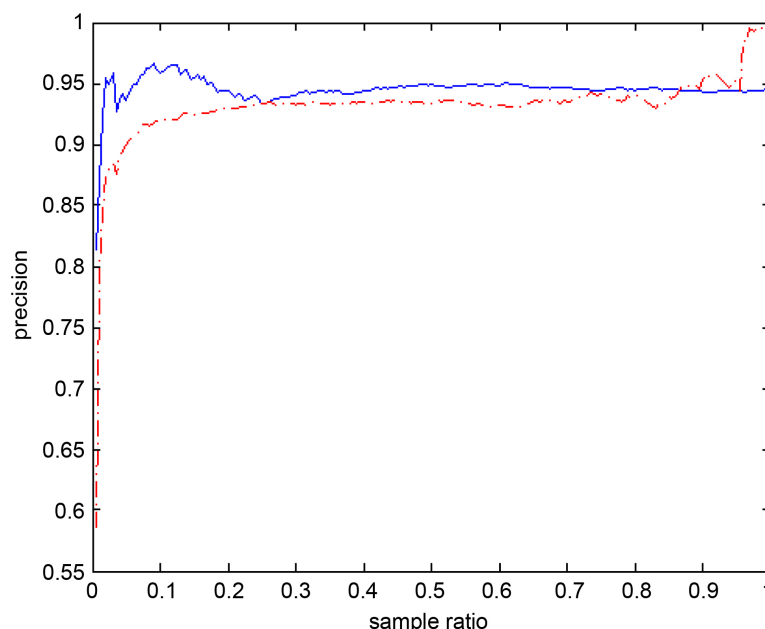


Figure 4. Effecting of sampling ratio on identification precision

图 4. 采样率对识别准确率的影响

着采样率的变化情况,红色虚线表示对测试数据进行判别时判别准确率的平均值随着采样率的变化情况。显然,在采样率不需要很大时,基于样本数据,判别器可以以较高的准确率(90%)辨识 WIFI 终端是否在指定的区域中。类似的结论对测试数据依然成立。这表明,依据多探针同时感知到的同一 WIFI 终端的 RSSI 值判别 WIFI 终端的位置是可行的。

表 1 出来了采样率为 0.1、0.2、...、0.9 时,使用判别器判别 WIFI 终端是否在指定区域内的误判情况。表 1 中,错误率 1 表示被样本数据的判别错误情况,错误率 2 表示被测试数据的判别错误情况。显然,对不同的采样率,使用基于 BP 神经网络的判别器辨识 WIFI 终端是否在指定区域内不仅保持较低的错误率,且错误率的变化较稳定,这进一步表明,依据多探针同时感知到的同一 WIFI 终端的 RSSI 值判别 WIFI 终端的位置是可行的。

从表 1 中可以发现,虽然在采用率比较小的时候无论是错误率 1 还是错误率 2 的平均值都保持较低的水平,但是由于方差都比较大(与错误率在同一数量级)。从图 5 和图 6 可以更直观地发现同样的现象:

Table 1. Sample rate VS identification error rate

表 1. 采样率与识别错误率

采样率	错误率 1		错误率 2	
	平均值	方差	平均值	方差
0.1	0.0409	0.0243	0.0785	0.0162
0.2	0.0565	0.0179	0.0715	0.0135
0.3	0.0599	0.0173	0.0672	0.0154
0.4	0.0562	0.0086	0.0651	0.0065
0.5	0.0510	0.0067	0.0658	0.0053
0.6	0.0515	0.0067	0.0677	0.0059
0.7	0.0532	0.0054	0.0645	0.0058
0.8	0.0555	0.0048	0.0594	0.0071
0.9	0.0566	0.0049	0.0469	0.0096

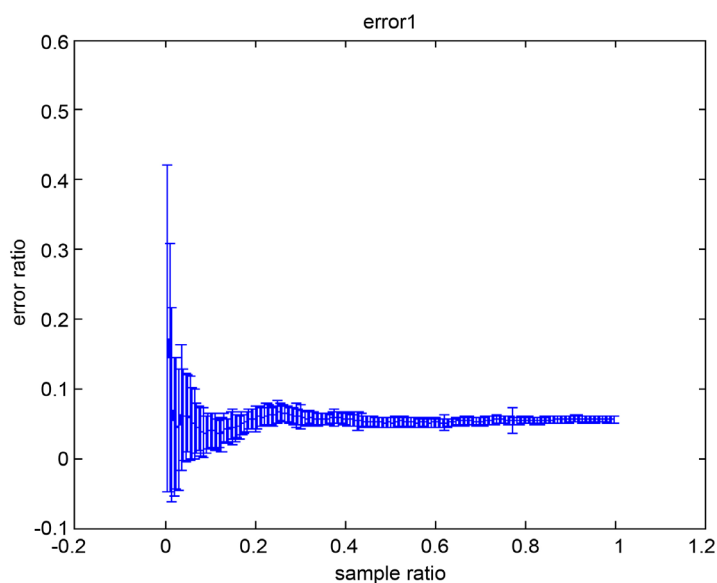


Figure 5. Effecting of sampling rate on error rate (error1, with standard deviation)

图 5. 采样率对样本数据识别错误率(error1)的影响(带标准差)

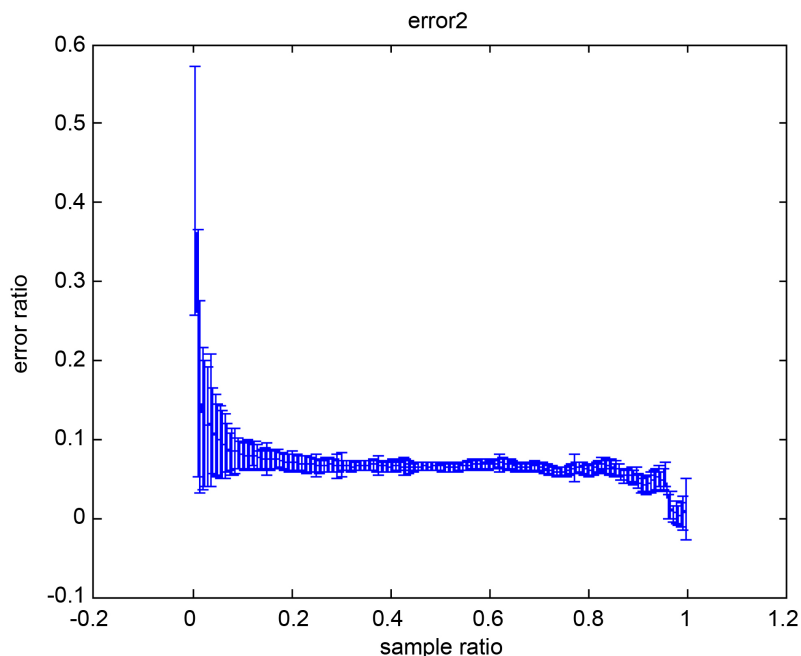


Figure 6. Effecting of sampling rate on error rate (error2, with standard deviation)
图 6. 采样率对样本数据识别错误率(error2)的影响(带标准差)

图 5 和图 6 分别以及给出的是采样率对样本数据识别错误率(error1)的影响采样率对测试数据识别错误率(error2)，显然，无论从图 5 还是图 6 都可以发现，当采样率较小时，判别器识别的误差率波动较大。因此故实际应用中需要选取合适的采样率，在判别器的训练时间、错误率以及错误率的稳定性三个性能指标之间进行平衡。

5. 结论与研究展望

围绕 WIFI 探针系统采集到的 WIFI 终端数据的预处理，本文将 WIFI 终端数据预处理流程划分为探针探测数据集解析、探针探测数据时间帧编号、面向室内人员定位的探针数据构造等三个阶段。并在对每个阶段的任务目标厘清的基础上设计了相关预处理任务的流程并进行了实现。实验结果表明，所获得的预处理后的 WIFI 终端数据被当作基于 BP 神经网络的判别器的输入时，该判别器可以以很高的准确率判别 WIFI 终端是否在指定的区域内。基于 BP 神经网络的判别器的性能表明依据多探针同时感知到的同一 WIFI 终端的 RSSI 值辨识 WIFI 终端是否在指定的区域内是可行的。

为更好的提升基于 BP 神经网络的判别器在辨识 WIFI 终端是否在指定的区域内时的准备率，基于 BP 神经网络的判别器的优化，包括 BP 神经网络的隐层数、隐层神经元的数量、判别器阈值都是值得优化的参数。进一步，由于 BP 神经网络中要求所有非输入神经元的激活函数一阶可导，而是否在指定区域内的 0/1 辨识结果是离散的，因此，结合具体的应用需求，面向离散判别的基于神经网络、辨识 WIFI 终端是否在指定区域内的判别器的前向神经网络架构以及训练方法是值得探索的楼宇智能化实现相关的研究。

基金项目

本文受国家重点研发计划项目“新型建筑智能化系统平台技术”(2017YFC0704100)、国家自然科学基金(11471304)部分资助。

参考文献 (References)

- [1] Kim, Y.S. and Srebric, J. (2017) Impact of Occupancy Rates on the Building Electricity Consumption in Commercial Buildings. *Energy and Buildings*, **38**, 591-600. <https://doi.org/10.1016/j.enbuild.2016.12.056>
- [2] Rahman, K.A., Hariri, A., Leman, A.M. and Yusof, M.Z. (2017) Energy Consumption in Residential Building: The Effect of Appliances and Human Behavior. *AIP Conference Proceeding*.
- [3] Oldewurtel, F., Sturznegger, D. and Morari, M. (2013) Importance of Occupancy Information for Building Climate Control. *Applied Energy*, **101**, 521-532. <https://doi.org/10.1016/j.apenergy.2012.06.014>
- [4] Shen, W.M., Newsham, G. and Gunay, B. (2017) Leveraging Existing Occupancy-Related Data for Optimal Control of Commercial Office Buildings: A Review. *Advanced Engineering Informatics*. <https://doi.org/10.1016/j.aei.2016.12.008>
- [5] Bahl, P. and Padmanabhan, V.N. (2000) RADAR: An In-Building RF-Based User Location and Tracking System. *Proceeding of IEEE Infocom*, 26-30 March 2000, 775-784. <https://doi.org/10.1109/INFCOM.2000.832252>
- [6] Jia, M., Srinivasan, R.S. and Raheem, A.A. (2017) From Occupancy to Occupant Behavior: An Analytical Survey of Data Acquisition Technologies, Modeling Methodologies and Simulation Coupling Mechanisms for Building Energy Efficiency. *Renewable and Sustainable Energy Reviews*, **68**, 525-540. <https://doi.org/10.1016/j.rser.2016.10.011>
- [7] Candanedo, L.M., Feldheim, V. and Deramaix, D. (2017) A Methodology Based on Hidden Markov Models for Occupancy Detection and a Case Study in a Low Energy Residential Building. *Energy and Buildings*, **148**, 327-341. <https://doi.org/10.1016/j.enbuild.2017.05.031>
- [8] Gu, Y., Lo, A. and Niemegeers, I. (2009) A Survey of Indoor Positioning Systems for Wireless Personal Networks. *IEEE Communications Surveys and Tutorials*, **11**, 13-32. <https://doi.org/10.1109/SURV.2009.090103>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2326-3415, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: airr@hanspub.org