

基于 LSTM 的 CSI 手势识别方法

刘佳慧 王昱洁 雷 艺

(合肥工业大学计算机与信息学院 合肥 230001)

摘要 基于 WiFi 的信道状态信息(Channel State Information, CSI)的手势识别在人机交互中具有广泛的应用前景。目前,大多数的 CSI 手势识别方法需人工提取特征,特征提取的过程繁琐,且只能识别特定方向的手势,限制了人的活动范围。针对上述问题,提出了利用长短时记忆神经网络(Long Short-Term Memory, LSTM)训练的方法,设计了一个基于 LSTM 的 CSI 手势识别系统。该系统将采集到的 CSI 数据首先进行异常点去除、最优子载波选择和离散小波变化去噪等预处理操作;然后通过 LSTM 网络训练分类,无需人工提取手势特征;最终实现推、拉、左挥、右挥 4 种手势在 4 个不同方向的识别,平均准确率达到 82.75%。文中分别讨论了发送到接收端的距离与数据集大小对手势识别准确率的影响,并对比 WiG 和 WiFinger 方法识别 4 个方向手势的识别准确率,结果表明文中所提方法具有更高的识别效果。

关键词 信道状态信息,长短时记忆神经网络,离散小波变换,手势识别, WiFi

中图分类号 TP183 **文献标识码** A

CSI Gesture Recognition Method Based on LSTM

LIU Jia-hui WANG Yu-jie LEI Yi

(School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230001, China)

Abstract Gesture recognition based on WiFi channel state information (CSI) has broad application prospects in human-computer interaction. At present, most methods require manual extraction of features, and the feature extraction process is cumbersome. It can only recognize gestures in a specific direction, which limits the range of people's activities. To solve the above problems, this paper proposed a method based on Long Short-Term Memory (LSTM) training to design a CSI gesture recognition system based on LSTM. The system preprocesses the collected CSI data through such as abnormal point removal, optimal subcarrier selection and discrete wavelet variation denoising. The LSTM network trains the classification without manual extraction of gesture features. Finally, the recognition of four gestures is achieved, which are pushing, pulling, left swing and right swing in four different directions, and an average recognition accuracy of 82.75% is reached. This paper discussed the influence of the distance between sender and receiver and the size of the data set on the accuracy of gesture recognition, and compared the gestures in four directions by WiG and WiFinger. The results show that the proposed method has higher recognition effect.

Keywords Channel state information, Long short-term memory, Discrete wavelet transform, Gesture recognition, WiFi

1 引言

随着科技的发展和水平的提高,以人为中心的人机交互模式越来越多地进入人们的生活,手势识别是人机交互模式的重要方式之一。当前手势识别的方法主要有基于计算机视觉、基于专用传感器和基于无线射频的手势识别。

基于计算机视觉的手势识别主要是利用摄像头采集手势动作的图像或视频,通过图像处理技术和计算机视觉的方法识别手势。例如, NÚñez 等用 Kinect 摄像机采集骨骼信息进行识别^[1-2]。Leap 公司^[3]研发的配备双摄像头的体感控制器采用 stereo-vision 原理生成 3D 数据,从中分析数据、提取手势动作信息来实现手势的识别。PointGrab 公司^[4]推出了基于 2D 摄像头的手势识别技术,其可以识别手指动作、相互位

置和手指指向。但是,该方法计算量巨大,且容易受到光照条件以及障碍物的影响。

基于专用传感器技术进行手势识别主要是通过将专用传感器(如加速度计)安装到人体上,采集相关的动作信息,从而实现手势识别。例如,哈工大的 Wu 和 Gao 等研发了可识别汉语手指字母的数据手套 CyberGlove,加州大学研发了一款利用超声波(Ultrasonic Wave)的微型芯片可以实现 3D 手势识别的可穿戴设备,百度公司推出了“Baidu Eye”^[5], InBody 公司开发了拥有手势识别功能的 InBody Band 智能手环等。专用传感器可以实现细粒度的行为感知,但是造价昂贵,安装和携带不便,难以广泛应用。

基于无线射频的手势识别主要是从特定接收设备上获取包含人体行为信息的信号,经过滤波等预处理后,提取特

本文受国家自然科学基金项目(61801162)资助。

刘佳慧(1995—),女,硕士生,主要研究方向为智能信息处理、手势识别和 WLAN 室内定位, E-mail: 1572356908@qq.com(通信作者);王昱洁(1980—),女,博士,讲师,主要研究方向为智能信息处理、WLAN 室内定位和音频信号处理, E-mail: 63249012@qq.com。

征^[6],利用其训练分类实现手势识别。而基于 WiFi 的手势识别因其具有不受光照条件限制、被动感知(无需携带传感器)、非视距、扩展性强、成本低等一系列优点,引起了人们的广泛关注。2013 年华盛顿大学的 Pu 等提出的 WiSee^[7]在 USRP-N210 实验平台上利用无线射频的多普勒效应实现了 9 个手势的识别,如推、拉、踢、闪等识别,该方法需要特殊的无线电外设。2014 年,哥延根大学的 Sigg 等^[8]从简单方便的思路出发,提取了接收信号强度指示器(Received Signal Strength Indicator, RSSI)的强度来进行简单的动作识别。2015 年 Abdelnasser 等提出的 WiGest^[9]通过分析 RSS(Received Signal Strength)信号变化的上升沿、下降沿等特征进行手势识别,单接入点的准确率为 87.5%,三接入点准确率为 96%。相比于 RSS 和 RSSI,用 CSI tool 工具从商用网卡里提取的信道状态信息是细粒度物理信息,对环境更加敏感,成为手势识别的新潮流。2015 年,WiG 系统^[10]针对推、拉、左挥、右挥 4 种手势,通过提取 CSI 信号的均值、标准偏差、中位数绝对偏差、最大值 4 个统计特征,采用 SVM 分类器,视距(LOS)下的识别率达 92%,非视距(NLOS)下的识别率达 88%。WiGest 和 WiG 的方法对不同手势提取不同的特征,提取特征的过程繁琐。2016 年, Li 等提出的 WiFinger^[11]使用商用 WiFi,经滤波、加权平均等预处理,利用 KNN 算法实现对 9 种美国符号语言(ASL)的识别,准确率达到 90.4%。2017 年, Qian 等提出的 Widance^[12]将人体运动引起的多普勒特征与运动方向结合起来,实现了单一的脚步动作在 8 个方向的识别,识别率为 92%。但该方法要求实验人员以面对信号接收端的方向,且只能识别脚部的单一动作。

以上方法中,手势特征的多样性导致的人工提取特征复杂,只能识别单方向的手势,限制了做手势的人的活动方向等,是目前 CSI 手势识别面临的问题。针对上述问题,本文提出了一种用 LSTM 训练不同手势的 CSI 数据从而实现手势识别的方法,设计了一个基于 LSTM 的 CSI 手势识别系统。该系统的创新点在于:

(1)提出了用 LSTM 网络识别手势。将采集到的 CSI 数据进行异常点去除,选择最优子载波和离散小波变化(DWT)进行去噪处理,预处理后的数据通过 LSTM 进行训练和识别,无需人工提取手势特征。

(2)采集了大量 CSI 数据作为数据集,选取其中一部分作为训练集,其余的作为测试集。最终实现了对 4 个方向的 4 种手势的识别,平均识别准确率达 82.75%,其中“推”手势的识别率最高达 87.5%。

2 WiFi 感知的基本原理

2.1 信道状态信息

正交频分复用(OFDM)技术将无线信道分成若干个彼此正交的子信道,将子数据流调制到正交子信道中传输,每个子数据流称为子载波。信道状态信息是一个在子载波尺度上的物理层信息,是指一条通信链路的信道属性^[13],反映了信号在每条传输路径上的衰弱因子^[14],比如信号散射、环境衰弱、功率随距离衰减等。利用无线网卡可从无线信号中提出 CSI 数据,每组 CSI 数据包含幅度和相位信息,具体如式(1)所示:

$$H(k) = \|H(k)\| e^{-j\angle H(k)} \quad (1)$$

其中, $H(k)$ 为第 k 个子载波的 CSI 数据, $\|H(k)\|$ 和 $\angle H(k)$ 分别为第 k 个子载波的幅值和相位。本文利用 CSI tool 工具从无线网卡中采集到的数据为包含 30 个连续子载波的 CSI 数据。

2.2 信号的动态传播模型

WiFi 信号在传播过程中遇到人体时,会发生反射、折射、衍射以及散射等现象,对信号的正常传播产生扰动,通过分析接收信号和检测信号变化的特征,感知信号传播过程中所遇到人体的状态,以到达行为识别的目的。

WiFi 信号的动态传播模型^[15]如图 1 所示,人体运动造成信号传播路径不断变化,影响接收信号。信号传播在频域中的模型简化为:

$$Y(f, t) = H(f, t) \times X(f, t) \quad (2)$$

其中, $Y(f, t)$ 为接收信号, $X(f, t)$ 为发送信号, $H(f, t)$ 为信道频率状态的响应(Channel Frequency Response, CFR),也是信道状态信息 CSI。

当信号通过 N 条路径达到接收端时,信道频率状态的响应如式(3)所示:

$$H(f, t) = e^{-j2\pi\Delta f t} \sum_{k=1}^N a_k(f, t) e^{-j2\pi f r_k(t)} \quad (3)$$

其中, $a_k(f, t)$ 代表第 k 条路径的衰减和初始相位, $e^{-j2\pi f r_k(t)}$ 代表第 k 条路径的相位偏, $r_k(f, t)$ 为时延, $e^{-j2\pi\Delta f t}$ 是收发端的相位差。

总的 CSI 包括静态路径和动态路径的频率状态响应,公式如下:

$$H(f, t) = e^{-j2\pi\Delta f t} (H_s(f, t) + H_d(f, t)) \quad (4)$$

$$H_d(f, t) = \sum_{k \in P_d} a_k(f, t) e^{-j2\pi d_k(t)/\lambda} \quad (5)$$

其中, $H_s(f, t)$ 表示静态频率状态响应, $H_d(f, t)$ 表示动态路径的频率状态响应, P_d 是所有动态路径。当人做手势时,信号的动态路径变化,总的 CSI 也发生变化。

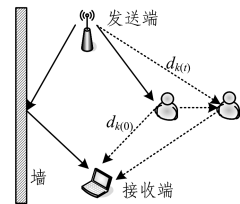


图1 WiFi 信号的动态传播模型

3 系统设计

该手势识别系统主要包括训练阶段和测试阶段,系统流程如图 2 所示。实验采集到的不同手势的 CSI 信息的幅度信息作为本系统所需的 CSI 数据,部分数据作为训练集,其余作为测试集。由于 WiFi 信号易受多径传播、信号衰减等因素干扰,直接获取到的 CSI 数据含有很多干扰噪声,因此需对 CSI 数据进行预处理,去除噪声数据。在训练阶段,训练集数据用局部异常因子算法去除数据中的异常点,然后根据方差大小分别选出 3 个最优的子载波,通过离散小波变换对子载波去除干扰噪声。不同手势预处理后的数据分别做相应的标签,CSI 数据和标签输入 LSTM 网络中训练。测试阶段中测试集数据经异常点去除、子载波选择、DWT 去噪等预处理后,输入已经训练好的网络中进行测试识别,并统计识别准确率。

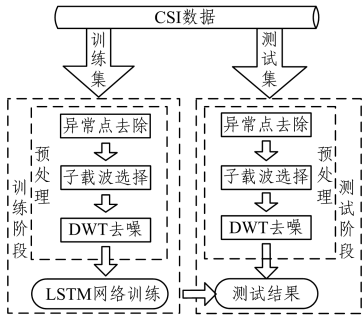


图 2 系统流程图

3.1 CSI 数据采集

实验发送端采用 TP-LINK 双频普通商用路由器发送 WiFi 信号,工作频段可为 2.4 GHz 和 5 GHz。接收端采用 Intel5300 无线网卡接收 CSI 信号数据,在 Ubuntu10.04 操作系统下运行。接收端配备 3 根 6 dB 增益的接收天线,组成了 1×3 的 MIMO (多输入多输出)阵列。考虑到用 2.4 GHz 频段的设备较多,噪声干扰太大,实验选用 5 GHz 工作频段。通过终端命令无线路由以一定的速率连续不断地发送数据,实验人员在发送和接收端中间做手势动作,接收端采样得到不同手势的 CSI 数据。实验环境如图 3 所示。

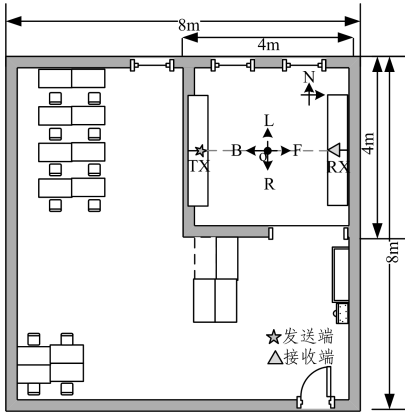


图 3 实验环境

发送端、接收端和实验人员的位置分别为 TX, RX, O 点。根据菲涅尔区模型^[16-17],距发送端 (TX) 和接收端 (RX) 的连线越近,相同手势动作引起 CSI 信号的变化越明显,接收到的 CSI 数据越能体现动作特征。考虑到菲涅尔区的对称性,实验人员位于 TX 和 RX 连线的中点 O 处,面向 RX 的方向为前 (F),其他方向依次为后 (B)、左 (L)、右 (R)。

对每个方向的动作采集 4 s 数据,开始采集时实验人员静止 2 s,以常速做动作,然后静止至 4 s 停止接收数据。采集到的每个手势的 CSI 数据为 $1 \times 3 \times 30 \times N$ 的矩阵,其中 1 代表发送端 1 根天线,3 代表接收端 3 根天线,30 代表采集到的每个 CSI 数据包包括 30 个正交频分复用子载波的信息, N 代表一段时间内采集到的 CSI 数据包的个数。每个数据包的 CSI 数据可以表示为:

$$CSI = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,30} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,30} \\ H_{3,1} & H_{3,2} & \cdots & H_{3,30} \end{bmatrix} \quad (6)$$

其中, H_{ij} 是每个子载波的 CSI 值;在 H_{ij} 中 i 表示第 i 根天线, j 表示第 j 个子载波。采集到的 CSI 值包含幅度和相位信息,本文提取幅度信息作为各手势的 CSI 信息。

3.2 预处理

由于 WiFi 信号易受多径传播、信号衰减等因素干扰,直接获取到的手势数据包含有很多干扰噪声,不能直接使用。因此,要对采集到的 CSI 数据进行预处理,去除噪声干扰。

3.2.1 异常点去除

没有手势动作(即静止的情况)的环境下,无线信道状态信息 CSI 数据很平稳,幅度波形没有太大的起伏。静止的动作不会对无线信号造成影响,即在时域里面,静止的动作对 CSI 信号的影响较小。做手势动作时,CSI 数值发生变化,幅度变化较大的部分为手势动作的部分。由传输功率变化和传输速率自适应等引起的设备内部本身状态的变化,导致采集到的 CSI 数据中存在一些较大的突发噪声,这些异常值并非手势动作引起的,会对手势识别造成干扰。局部异常因子算法 (Local Outlier Factor, LOF)^[18] 通过比较数据点与相邻点的相对密度来衡量它们的孤立程度,并对每个点赋予一个局部异常系数 LOF 。

$$LOF(P) = \frac{\sum_{O \in K(P)} \frac{Lrd_k(O)}{Lrd_k(P)}}{k} \quad (7)$$

其中, $Lrd_k(p)$ 为 p 点的 k -局部可达密度; p 点的局部异常系数 $LOF(P)$ 等于 p 点邻域的平均可达密度与 p 点自身可达密度之比,用来衡量 p 点相对于它 k -距离近邻点而言成为异常点的概率。如果 LOF 系数比较大,则反映该点局部范围包含的点比较稀疏,成为异常点的可能性就大,反之成为异常点的可能性就小。这里取 $LOF > 1$ 时的点作为异常点,将其 CSI 数据中去除。

3.2.2 子载波选择

图 4 给出手势“推”在“前”向的 CSI 数据中一根接收天线的幅度信息。横坐标表示该手势在一段时间内采集了 400 个数据包,纵轴表示 CSI 的幅度,不同的线条分别表示该接收天线 30 个子载波在采集时间内 CSI 信息幅度的变化。根据 WiFi 的动态传播模型可知,由环境引起的静态响应不反映手势动作信息,而由人体态变化引起的动态响应可以反映手势动作信息。图中, A 区域为静止时的 CSI 幅度信息,该区域幅度变化较小; B 区域幅度变化相对较大的部分为该手势动作的部分。从图中看出,同一手势不同子载波的信道状态信息存在差异。由于具有较高方差的子载波具有较大的动态响应,所以根据 3 根接收天线中 30 个子载波方差的不同,动态选取每根天线中方差最大的子载波。

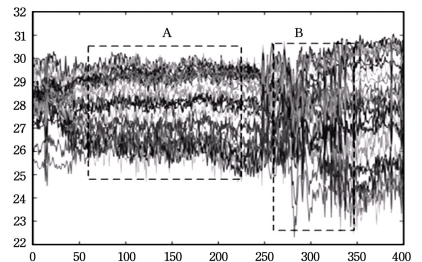


图 4 30 个子载波的幅度信息

3.2.3 DWT 去噪

本文采用离散小波变换 (DWT) 对原始的 CSI 数据去噪, DWT 可以把一个信号转化为含有时频两域信息的信号,可以根据需要把信号的噪声尽可能过滤掉,被称为信号处理中的数学显微镜。DWT 将信号分解为近似系数和细节系数,近

似系数描述了保持 CSI 模式的大规模特征的信号的形状、趋势,细节系数捕获表示高频噪声和 CSI 模式的精细细节的低比例分量。由于我们想去除高频噪声分量,同时保持 CSI 模式的足够细节以区分相似的手势,因此将动态阈值应用于细节系数以去除噪声分量。根据需要,在预处理中采用动态阈值,5层 Symlet 小波分解后重构。

本文截取了每个手势的 CSI 数据 400 个包中包含手动动作部分的 200 个包,通过 DWT 对该数据去噪。图 5 是“推”手势“前”向的 CSI 数据经子载波选择得到一个子载波信号,在 DWT 消噪前后的对比图。从原信号和 DWT 去噪后信号的对比中可以看出,DWT 去噪可以保护有用信号尖峰和突变信号,抑制高频噪声的干扰。

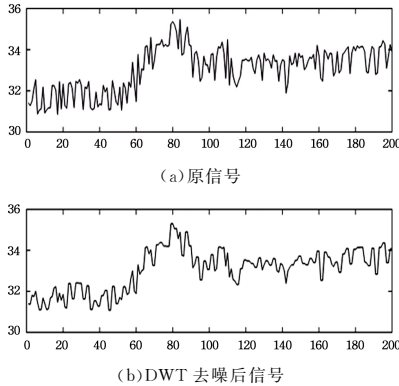


图 5 DWT 去噪前后对比图

3.3 LSTM 网络

长短时记忆神经网络^[19]是一种节点定向连接成环的时间递归神经网络,适合于处理和预测时间序列中间隔和延迟相对较长的重要事件。图 6 所示为 LSTM 模型结构,其中 X_t 为 t 时刻的输入值, h_t 为 t 时刻的输出值。

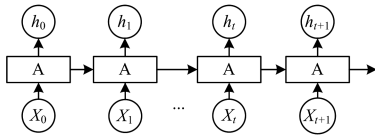


图 6 LSTM 模型结构

CSI 数据经预处理后,由 $1 \times 3 \times 30 \times N$ 的四维降维成 3×200 的二维矩阵,可以表示为:

$$X = \begin{bmatrix} H_{1,m} \\ H_{2,n} \\ H_{3,p} \end{bmatrix} \quad (8)$$

其中, $H_{1,m}$, $H_{2,n}$, $H_{3,p}$ 是 1×200 的向量,分别表示第一、二、三根天线中的 m, n, p 个子载波的 CSI 值。每个手势的 CSI 数据分别对应各手势标签 Y 。将 X 和 Y 分别作为该手势的特征向量和对应的标签输入到 LSTM 网络中。

LSTM 的特点在于它在算法中加入了一个判断信息有用与否的“处理器”,这个处理器作用的结构被称为 cell。一个 cell 中被放置了 3 扇门,分别叫作输入门、遗忘门和输出门。当 CSI 数据输入网络中,只有符合算法认证的信息才会留下,不符的信息则通过遗忘门被遗忘。如图 7 所示的 LSTM 具体结构,在 t 时刻 LSTM 的输入有 3 个:当前时刻网络的输入值 X_t ,上一时刻 LSTM 的输出值 $h_{(t-1)}$,以及上一时刻的单元状态 $c_{(t-1)}$ 。 t 时刻的输出有两个:当前时刻输出值 h_t 和当前时刻的单元状态 c_t 。

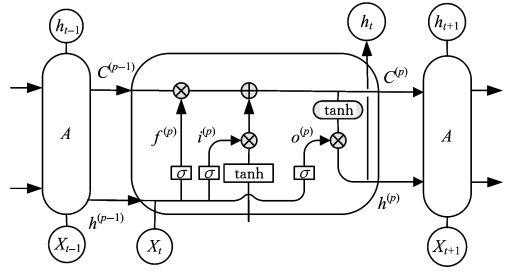


图 7 LSTM 的具体结构图

LSTM 网络中的 3 个门用来控制长期状态 C 。遗忘门决定上一时刻的单元状态 c_{t-1} 有多少保留到当前时刻 c_t 。公式如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

其中, W_f 是遗忘门的权重矩阵, $[h_{t-1}, x_t]$ 表示把两个向量连接成一个更长的向量, b_f 是遗忘门的偏置项, σ 是本文选用的激活函数 sigmoid 函数。

输入门决定当前时刻网络的输入 X_t 有多少保存到单元状态 C_t 。 W_i 是遗忘门的权重矩阵, b_i 是输入门的偏置项。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

根据上一次的输出和本次的输入来计算当前输入的单元状态,公式如下:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (11)$$

当前时刻的单元状态 c_t 的计算过程为: \circ 表示按元素乘,由上一次的单元状态 c_{t-1} 按元素乘以遗忘门 f_t ;再用当前输入的单元状态 \tilde{c}_t 按元素乘以输入门 i_t ,最后将两个积加和。公式如下:

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (12)$$

这样,就可以把当前的记忆 \tilde{c}_t 和长期的记忆 c_{t-1} 组合在一起,从而形成了新的单元状态 c_t 。

输出门控制单元状态 c_t 有多少输出到 LSTM 的当前输出值 h_t ,控制了长期记忆对当前输出的影响。公式如下:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (13)$$

$$h_t = o_t \circ \tanh(c_t) \quad (14)$$

本文把数据集分成两个非重叠集:训练集和测试集。训练集用于 LSTM 网络中的训练。LSTM 网络设置隐含层个数为 64,输入为预处理后的训练集 CSI 数据,输出为分类结果的种类。识别手势为 4 种,所以输出维度为 4,通过调整 LSTM 网络的各个参数调节输出状态。

测试阶段用 CSI 数据集中的测试集,数据对应不同的手势做相应的标签。测试集的 CSI 数据经预处理操作后输入 LSTM 网络进行测试。测试结果与对应的标签进行对比,统计不同手势对应的识别准确率。

4 实验结果与分析

4.1 实验结果

实验人员在图 3 所示实验环境中右上角 $4\text{m} \times 4\text{m}$ 的房间中采集手势的 CSI 数据,采样频率为 100 Hz。TX 和 RX 距离为 3 m 且在同一水平高度,实验人员位于 TX 和 RX 连线的中点 O 处。实验人员分别做推、拉、左挥、右挥 4 个动作,如图 8 所示。

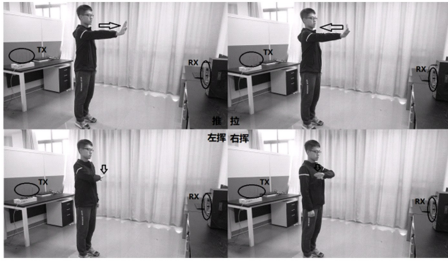


图 8 四手势

面向 RX 的方向为前(F),其他方向依次为右(R)、后(B)、左(L)。每个动作采集 4 个不同方向的数据,图 9 所示为“推”的前、后、左、右 4 个方向。

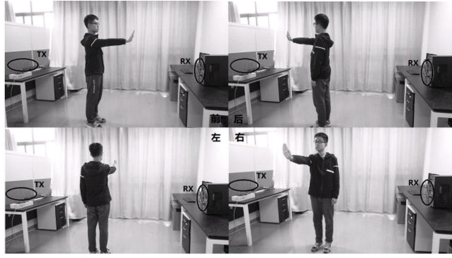


图 9 手势的四方向

每个动作采集 4s CSI 数据,开始采集时实验人员静止 2s 后匀速做完动作,静止至 4s 停止接收数据。4 个手势分别针对 4 个不同方向,每个方向采集 100 组数据,每个手势采集 1600 组有效数据,最终完成 4 个手势共计 6400 组数据的采集。选取所采集数据集中 80% 的数据作为训练集,即随机选取每个手势在每个方向的 80 组数据,剩余的 20% 数据作为测试集。训练集数据经异常点去除、子载波选择和 DWT 去噪等预处理后输入 LSTM 网络进行训练。测试集数据经同样的预处理操作,输入已训练好的网络测试,得到识别准确率。

图 10 的混淆矩阵为每个手势被识别成 4 个手势的概率。从图中可以看出推、拉、左挥、右挥的识别准确率分别为 87.5%,86.2%,79.8%,77.5%,其中“推”手势识别率最高,四手势的平均准确率为 82.75%。“推”和“拉”手势动作由于穿过近似数量的菲涅尔区,因此识别成其他手势的误判率中互相识别率较高。同理,“左挥”和“右挥”在误判率中互相识别率比被识别成其他手势的概率高。

	推	拉	左挥	右挥
推	87.5%	12.5%	2.7%	2.4%
拉	11.3%	86.2%	19%	3.2%
左挥	0.7%	0.6%	79.8%	16.9%
右挥	0.5%	0.7%	15.6%	77.5%

图 10 实验结果的混淆矩阵

4.2 对比分析

4.2.1 不同采样速率的影响

为选择合适的采样频率,将采样频率设置为 20 Hz,100 Hz,500 Hz,1000 Hz。实验发现,当采样频率低于 50 Hz 时采集到的 CSI 数据不能完整表示手势动作,采样频率高于 500 Hz 时数据丢包较严重。考虑到实际应用,采样频率过大,包含的噪声数据也会因此增加,同时数据冗余量过大,系统响应缓慢,本文实验将采样率选为 100 Hz,手势数据可清楚描述手势动作,且包含的噪声数据相对较少。

4.2.2 不同距离的影响

当发送端的无线路由和接收端天线间距离变化时,菲涅

尔区的半径发生变化,相同手势穿过菲涅尔区产生的 CSI 信号的变化也不同。为研究不同距离对 CSI 手势识别的影响,以 100 Hz 的采样速率采集数据,当发送端的无线路由和接收端天线间的距离分别为 2m,3m 和 4m 时,分析每个手势识别率。不同距离下的 4 种手势识别率如图 11 所示。

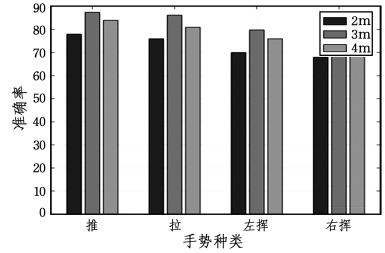


图 11 发送和接收端距离的影响

当发送端和接收端距离 3m 时,手势的识别率最高。由菲涅尔区模型可知,发送和接收端距离越近,第一菲涅尔区的半径越小,相同的手势动作经过的菲涅尔区数量越多,引起的 CSI 信号变化较大,训练识别率较高。相反,发送端和接收端距离越远,相同的手势动作引起的 CSI 信号变化越小,识别率越低。因此,发送端和接收端距离 3m 时的识别率比距离 4m 时的识别率高。但发送端和接收端距离太近时,相对于手势动作引起的 CSI 信号变化,身体状态的改变也造成较大的干扰。因此,发送端和接收端距离 2m 时比距离 3m 时的识别率低。

4.2.3 训练样本大小的影响

训练样本的数据不仅包含手势动作信息,还包含环境的丰富度,因此训练样本大小的不同会对手势识别率产生影响。选取 3 个不同大小的训练集,训练集中样本的大小分别为 320,640,1280,即分别选取每个手势的 80,160,320 个样本且这些样本平均取自 4 个方向的数据,训练样本并观察测试结果。

由图 12 可以看出,识别率随着训练样本数的增加而呈现增长趋势,这是由于训练样本越多,场景越丰富,LSTM 训练越精确。当每个手势的训练样本为 320 个时,识别率最高。但是,训练样本数目越多,训练所用时间也会越长。

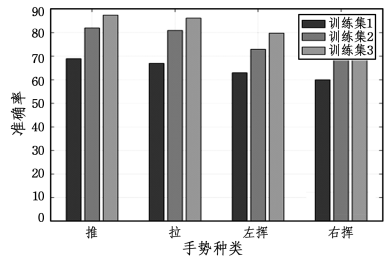


图 12 训练样本大小的影响

4.2.4 方法对比

我们采用文献[10]中 WiG 系统的方法 1、文献[11]中 WiFinger 的方法 2 与本文方法对训练集进行训练分类,并统计各手势的识别准确率。3 种方法的识别准确率如图 13 所示,其中方法 3 为本文采用的方法。可以看出,对于不同方向的 4 种手势,本文 LSTM 网络的方法识别率比方法 1 和方法 2 更高。方法 1 在单一方向的 4 个手势上通过 SVM 分类,方法 2 对特定的美国符号手势通过 KNN 分类,均可达到 90%

以上的识别率;但它们对不同方向的同一手势识别的准确率较低。这是因为这些人工提取特征的方法,只提取了单方向的手势特征并分类;SVM和KNN分类器对人工提取的不同方向手势特征的分类效果有限。而LSTM网络能够自动提取特征,且网络中的输入门、遗忘门和输出门能保留手势信息中的有效特征,通过循环训练最终达到更好的识别效果。

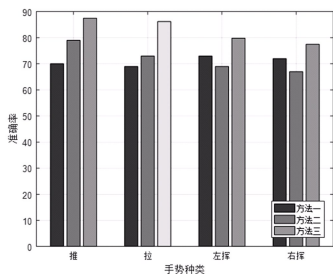


图 13 3种方法的对比

结束语 本文提出了一个基于LSTM的CSI手势识别系统,该系统利用长短期记忆神经网络训练的方法,实现了不同方向的手势识别。首先,通过Intel 5300无线网卡连接商用的路由器来采集大量手势CSI数据,得到本系统所需的数据库,将其分成训练集和测试集两部分。训练集的CSI数据经异常点去除、子载波选择和DWT去噪等预处理后,输入到LSTM网络调节各参数以进行训练。将测试集数据进行相同的预处理后输入训练好的网络中进行测试识别,最终实现推、拉、左挥、右挥4种手势在4个不同方向的识别。另外,本文还讨论了不同采样速率、不同距离和数据集大小对识别准确率的影响。对比了WiG和WiFinger中方法对多方向手势的识别,证明基于LSTM的CSI手势识别方法具有较高的识别率。不同方向的4种手势平均识别准确率为82.75%,其中“推”的手势识别率最高达87.5%。

鉴于CSI信息的细粒度,以及基于CSI的手势识别对不同环境和不同实验人员动作的CSI信息差异较敏感,为提高识别率,要求数据采集时控制环境变量,且手势动作标准化。如何减小环境变化对识别率的影响,提高手势识别的鲁棒性,是未来基于CSI信息的手势识别可研究的方向。

参考文献

- [1] NÚÑEZ J C, CABIDO R, PANTRIGO J J, et al. Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition[J]. *Pattern Recognition*, 2018, 76: 80-94.
- [2] ZHANG J, LI W, OGUNBONA P O, et al. RGB-D-based action recognition datasets: a survey[J]. *Pattern Recognition*, 2016, 60: 86-105.
- [3] Leap Motion. Leap Motion Controller[EB/OL]. [2014-12-28]. <https://www.leapmotion.com>.
- [4] Point Grab. Virtual Touch [EB/OL]. [2014-12-29]. <http://www.pointgrab.com>
- [5] CHOI J J U, AE CHUN S, CHO J W. Smart Secure-Gov: mobile government security framework [C]// Proceedings of the 15th Annual International Conference on Digital Government Research. New York: ACM, 2014: 91-99.
- [6] 鲁勇, 吕绍和, 王晓东, 等. 基于WiFi信号的人体行为感知技术研究综述[J]. *计算机学报*, 2017, 53(9): 170-183.
- [7] PU Q, GUPTA S, GOLLAKOTA S, et al. Whole-home gesture recognition using wireless signals[C]// Proceedings of the 19th Annual International Conference on Mobile Computing and Networking. New York, USA, 2013: 27-3.
- [8] SIGG S, BLANKE U, TROSTER G. The telepathic phone: Frictionless activity recognition from wifi-rssi [C]// Proceedings of the IEEE international conference on Pervasive Computing and Communications. Budapest: IEEE, 2014: 148-155.
- [9] ABDELNASSER H, YOUSSEF M, HARRAS K A. Wigest: A ubiquitous wifi-based gesture recognition system [C]// Proceedings of the 2015 IEEE Conference on Computer Communications. Hong Kong, China, 2015: 1472-1480.
- [10] HE W, WU K, ZOU Y, et al. Wig: Wifi-based gesture recognition system[C]// Proceedings of the 24th International Conference on Computer Communication and Networks. Las Vegas, USA, 2015: 1-7.
- [11] LI H, YANG W, WANG J, et al. Wifinger: Talk to your smart devices with finger-grained gesture [C]// Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, USA, 2016: 250-261.
- [12] QIAN K, WU C S, ZHOU Z M, et al. Inferring Motion Direction using Commodity Wi-Fi for Interactive Exergames [C]// ACM SIGCHI Conference on Human Factors in Computing Systems (CHI). ACM, 2017: 6-11.
- [13] HALPERIN D, HU W, SHETH A, et al. Tool release: gathering 802.11n traces with channel state information [J]. *ACM SIGCOMM Computer Communication Review*, 2011, 41(1): 53-53.
- [14] 杨铮, 刘云浩. Wi-Fi 雷达: 从 RSSI 到 CSI [J]. *中国计算机学会通讯*, 2014, 10(11): 55-60
- [15] WANG W, LIU A. X, SHAHZAD M, et al. Understanding and modeling of wifi signal based human activity recognition [C]// Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. New York, USA, 2015: 65-76.
- [16] WANG H, ZHANG D, MA J, et al. Human respiration detection with commodity wifi devices: Do user location and body orientation matter [C]// Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, USA, 2016: 25-36.
- [17] WU D, ZHANG D, XU C, et al. Widir: Walking direction estimation using wireless signals [C]// Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, USA, 2016: 351-362.
- [18] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers [C]// Proceedings of the ACM SIGMOD 2000 Int. Conf. on Management of Data. ACM, 2000: 93-104.
- [19] CHOI J S, LEE W H, LEE J H, et al. Deep Learning Based NLOS Identification with Commodity WLAN Devices [J]. *IEEE Transactions on Vehicular Technology*, 2018, 67: 3295-3303.