

基于双向长短期记忆网络及注意力机制的 室内行人模式识别

梁玉杰^a, 崔博^{a,b}

(华北理工大学 a.人工智能学院; b.河北省工业智能感知重点实验室, 河北 唐山 063210)

摘要: 在室内空间进行准确的行人模式识别/场景感知,特别是与位置关联的识别/感知,对于行人的聚集或追踪具有重要意义.针对传统机器学习方法特征提取困难、分类精度低,非正常性行为造成较大识别误差等问题,提出一种基于注意力机制和双向长短期记忆(bidirectional long short-term memory, Bi-LSTM)网络的室内实时行人模式识别的模型.建立 Bi-LSTM 网络提取滑动窗口内行人模式时序特征,评估模型网络结构的性能与时效性,优化所提网络的 Bi-LSTM 层数和隐藏层节点数,并确定最优的网络结构;为了削减噪声数据对模型的影响,提高网络筛选信息特征的能力,引入注意力机制对所提取的时序特征进行权重参数优化.实验结果表明,相比传统机器学习算法,优化参数后的 Bi-LSTM 网络,行人模式识别准确度平均提高 6.37%,进一步引入注意力机制后,识别准确度平均提高 9.21%,最终准确度可达 99.32%.所提模型可以有效对行人模式/场景感知进行分类,为室内精准定位追踪提供方法支持.

关键词: 行人模式识别;滑动窗口;时序特征;Bi-LSTM;注意力机制

中图分类号: TP391

文献标志码: A

文章编号: 1000-2367(2024)03-0088-10

随着社会的发展,对精准定位、准确智能感知的要求越来越高.在室外,可以通过卫星定位技术以及通讯信号进行定位,获取行程轨迹;在室内,行人模式识别/场景感知是实现行人聚类与追踪的重要环节和前提,实现准确、可靠的室内行人时空聚类问题亟须解决.近年来很多学者与科研团队投入行人模式识别算法的研究中,极大地推进室内行人追踪的研究与发展.

实现行人模式识别的方法主要有以下两种途径:1)视觉识别:基于图片或视频的行人模式识别^[1-5],具有精度高的优点,但其对环境要求较高,对摄像头的位置比较敏感,无法进行个体的长时间行人追踪,且存在人们隐私保护问题的争议.2)基于传感器的行人模式识别^[6-14],随着惯性传感器在人们生活中的应用与发展,在室内环境中,不难通过惯性传感器检测行人的行为模式及定位信息.本文基于惯性传感器以及气压计相关时序数据对室内行人模式进行识别,达到对室内行人聚类与追踪的目的.

运用惯性传感器进行行为模式识别的研究众多.刘宇等^[6]提出基于加速度时域特征的人体行为模式识别算法,运用前馈型 BP 神经网络分类器对行走、上楼、下楼、跑、跳 5 种行为模式进行分类,综合识别精度达 85%以上.衡霞等^[7]利用手机内置三轴加速度采集人体日常行为,对数据进行提取多种统计特征,利用支持向量机(support vector machine, SVM)分类器进行分类识别.段小虎等^[8]提取三轴加速度数据的方差、均值和

收稿日期: 2023-03-01; **修回日期:** 2023-03-22.

基金项目: 2021 年度教育部产学研合作协同育人项目(202101138019); 2021 年度教育部高等学校电子信息类专业教学指导委员会项目(2021-JG-04).

作者简介(通信作者): 崔博(1979-),男,河北唐山人,华北理工大学副教授,研究方向为信号与信息智能处理、仿真方法与算法及应用, E-mail: mikecui@ncst.edu.cn.

引用本文: 梁玉杰, 崔博. 基于双向长短期记忆网络及注意力机制的室内行人模式识别[J]. 河南师范大学学报(自然科学版), 2024, 52(3): 88-97. (Liang Yujie, Cui Bo. Indoor pedestrian pattern recognition based on bidirectional long short-term memory network and attention mechanism[J]. Journal of Henan Normal University(Natural Science Edition), 2024, 52(3): 88-97. DOI: 10.16366/j.cnki.1000-2367.2023.03.01.0002.)

协方差作为特征,采用 SVM、K 最近邻和随机森林算法等传统的机器学习方法进行模式识别.张烈平等^[9]运用 BP 神经网络分析行为数据,建立人体活动行为分类模型,得到了较高的准确率.刘旭等^[10]利用 SVM 和自适应力增强(Adaboost)分类算法对站立、侧躺、仰卧、走路和跑步等姿态进行识别,但在人体运动其他模式上缺乏研究.王玉坤等^[11]采用 SVM 多类分类器对走、跑、上下楼等人体姿态进行识别,整体识别率达到 87.00% 以上,在准确率以及识别的实时性上有待提高.以上研究大部分运用机器学习面临特征提取困难、分类精度低、实时性较差等突出问题.

近年来,利用深度学习解决模式识别与室内定位的实践层出不穷.深度学习方法用样本数据训练模型,具有无需提取样本特征的优势,主动提取数据内在规律与层次结构.杨韞韬等^[8]提出一种解决监控视频中行人相互遮挡问题的人体姿态补全法,运用了循环神经网络(recurrent neural network, RNN)和注意力机制对人体姿态序列进行双向补全,估测监控视频中行人被遮挡时的状况.闫昊雷等^[4]将深度学习运用于行人重识别领域,通过双向长短期记忆网络寻求图像的上下文信息从而进行图像遮挡部分的补全.欧群雍等^[5]结合卷积神经网络(convolutional neural network, CNN)和 Bi-LSTM 的模型减少图像中长期遮挡导致的行人模式误识别的问题,提升了多行人目标的检测跟踪性能.高丽丽等^[14]提出一种基于智能手机的长短期记忆(long short-term memory, LSTM)室内定位方法,运用卡尔曼滤波算法剔除数据中的噪声,建立 LSTM 深度学习网络抑制陀螺仪数据中的漂移现象,有效提高了室内定位的精度,但其未考虑室内行人存在多种行为模式的问题.文献[3-5]中均将深度学习的算法运用在行人模式识别当中.文献[13]中尽管对深度学习网络在室内定位追踪中进行了充分研究,但仍然面临持续追踪导致功耗高的问题.

综上所述,传统机器学习模型识别精度较低,无法满足现在需求;室内连续定位追踪可能导致数据堆积造成定位不准确,且存在功耗较大以及私密性差等问题.为实现一种基于模式识别/场景感知的室内空间行人聚集与追踪方法,本文重点考虑能够反映室内行人位置特征的 4 种典型行人模式/场景,基于惯性传感器和气压计进行数据采集,提出一种注意力机制与 Bi-LSTM 相结合的网络结构模型.该模型利用 Bi-LSTM 双向结构,同时引入注意力机制使模型加强提取与行人模式相关性大的特征,实现低功耗、高精度、实时识别室内行人模式.

1 相关工作

1.1 数据预处理

数据预处理主要包括数据标准化处理和数据集构造.本文采用标准差(Z-score)标准化法^[15],基于原始数据的均值和标准差将数据标准化,转换函数如式(1)所示:

$$x' = \frac{x - \mu}{\sigma}, \quad (1)$$

式中 μ 表示样本数据的均值, σ 表示样本数据标准差, x 表示原始数据, x' 表示标准化后的数据.

单一的传感器观测点不足以挖掘动态特征,过小的窗口很难提取充足的特征,过大的窗口会导致训练参数快速增长,从而增加网络的训练时长.本文对采集到的时序数据进行滑动窗口切片处理.实施过程中,每一个样本窗口包含 300 个采样点,即窗口的大小设置为 3 s,按照 50% 的采样重叠率进行样本的连续获取.

1.2 传统机器学习方法

1.2.1 算法介绍

1) 决策树分类器主要用于分类问题的研究,模型呈现树形结构,是一种以样本实例为基础的归纳学习分类算法^[16-17].根据损失函数最小化原则建立决策树模型,即决策树的生成^[18].利用已经建立的决策树以同样的方式进行属性判断,从而预测所属类别.本文采用基尼指数(Gini)作为节点分裂评价准则进行分类,其计算如式(2)所示:

$$\text{Gini}(P) = \sum_{m=1}^n P_n (1 - P_n), \quad (2)$$

式中 n 表示所有样本的类别数目, P_n 表示第 n 类样本被选中的概率.决策树分类器具备易于解释和理解的优点,但同时存在处理连续数据较难预测的不足.

2)随机森林^[19]分类器是以决策树为基本模型,综合了装袋(Bagging)规则与随机特征子空间的算法^[20].该分类器弥补了决策树中可能存在过拟合情况的缺点,分类精度更高,具有很好的抗噪能力.本文用随机森林算法建立多棵 CART 算法决策树,最终通过投票决定行人模式的类别.

3)SVM 算法是通过最大化类间间隔构造最优超平面的方法来进行分类,是一种监督机器学习算法,已经被广泛应用于模式识别、文本分类等领域^[21].该算法可以解决高维及非线性问题,提高了分类器的泛化能力和模型的鲁棒性,但其存在对于大规模的训练样本难以实施的缺陷^[22].

1.2.2 特征提取

本文选取合加速度的幅值、合加速度标准差、气压计的标准差具体计算如下.

1)加速度 a_t 为三轴加速度各方向加速度的矢量模,即合加速度的大小,如式(3)所示:

$$a_t = \sqrt{a_{tx}^2 + a_{ty}^2 + a_{tz}^2}, \quad (3)$$

式中 a_{tx}, a_{ty}, a_{tz} 分别表示加速度在 x, y, z 轴 t 时刻的加速度, a_t 表示 t 时刻的合加速度.

2)标准差 σ 为数据样本方差的算术平方根,该特征可反映数据样本的离散程度,本文对合加速度以及气压数据进行标准差特征的提取,如式(4-5)所示:

$$\sigma_1 = \sqrt{\frac{\sum_{t=1}^n (a_t - \mu)^2}{n-1}}, \quad (4)$$

$$\sigma_2 = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}}, \quad (5)$$

式(4)、(5)中: a_t 为合加速度数据, $t = 1, 2, 3, \dots, n$, n 为数据样本个数, σ_1 为 n 个样本合加速度的标准差. x_i 为大气压数据, $i = 1, 2, 3, \dots, n$, n 为数据样本个数, σ_2 为 n 个样本大气压的标准差.

上述特征可以有效区分行人在室内走路、爬楼、电梯、扶梯 4 种行人模式/场景.其特征分布如图 1~3 所示.

图 1 表明,爬楼模式的合加速度标准差在 1.2~1.4 之间、行走模式的合加速度标准差在 1.1~1.2 之间,运动较为剧烈,且剧烈程度不同,较易区分;但是直梯、扶梯这两种模式的合加速度标准差均在 0.8~0.9 之间,虽有区分但仍有交叉,基于合加速度标准差特征进行区分容易混淆.

图 2 表明,爬楼、行走模式与其他模式的交叉较少,易于区分,扶梯与直梯这两种模式交叉较多,基于合加速度幅值特征很难区分.为了进一步区分直梯与扶梯这两种行人模式,本文进一步引入气压标准差特征.

图 3 表明,直梯模式在一个滑动窗口时间内气压变化大,数据较为离散,标准差在 0.10~0.12 之间,行走模式在一个滑动窗口时间内的气压变换非常小,样本数据非常集中,标准差较小,易于区分;爬楼与扶梯在一个滑动窗口时间内气压变化的大小相似,标准差均位于 0.03~0.06 之间;直梯与扶梯模式的气压标准差相差较大,易于区分.

2 模型构建

2.1 系统整体设计

本文设计的室内时空聚类总体系统结构如图 4 所示.基于智能手机或手环采集惯性传感器以及大气压的时序数据,构建一种结合注意力机制和 Bi-LSTM 网络的行人模式识别/场景感知模型,对室内行人模式

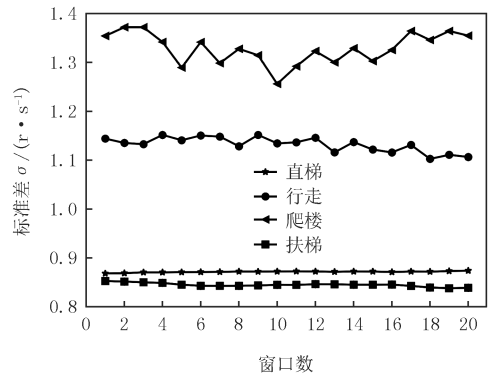


图1 合加速度标准差特征

Fig.1 Combined acceleration standard deviation features

进行实时精准识别,在行人模式识别模型的基础上,通过大气压数据进行楼层识别,同时提取数据的时间标签,将相同时间的行人模式以及所处楼层进行聚类分析,最终得到行人在室内的时空位置,以实现室内高精度、低功耗的行人聚集分析,达到精准定位的目标.

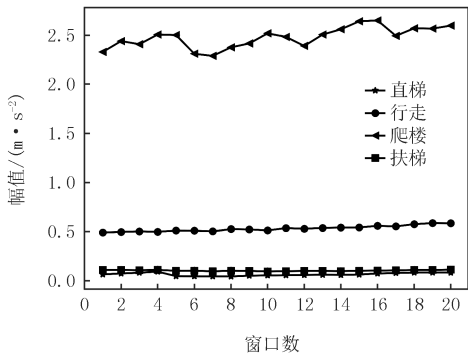


图2 合加速度幅值特征

Fig.2 Combined acceleration amplitude characteristics

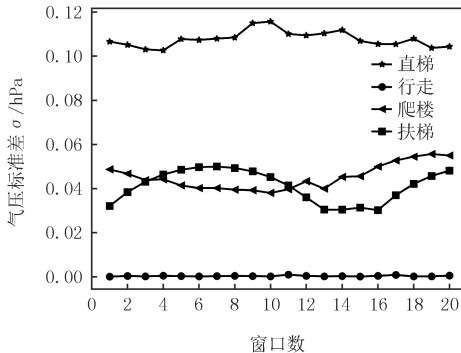


图3 气压标准差特征

Fig.3 Characteristic of atmospheric standard deviation

2.2 结合注意力机制和 Bi-LSTM 的网络架构

LSTM 网络是一种特殊的循环神经网络,通过引入遗忘门、输入门和输出门的机制,有效改善了 RNN 梯度爆炸和梯度消失的问题^[23],特别适用于序列数据处理^[23].Bi-LSTM 通过使用两个独立的隐藏层,一个 LSTM 正向处理序列,一个 LSTM 反向处理序列,这种结构充分利用序列中前后两个方向的信息,可以有效提取数据前后的时序特征^[24].

对于室内行人模式识别/场景感知,本文研究提出结合注意力机制和 Bi-LSTM 网络结构的模型,模型结构如图 5 所示,模型结构由 Bi-LSTM 层、注意力机制层和全连接层组成.该模型基于固定大小的滑动窗口,获取 4 种典型室内行人模式的时序数据进行分类识别,其主要步骤为:第一步,将输入的行人模式数据序列经过 Bi-LSTM 层,结合整个窗口数据进行正向与反向的特征提取,并输出特征值;第二步,将 Bi-LSTM 层的输出作为注意力机制层的输入,由于 Bi-LSTM 层各时刻输出的隐藏状态对结果的影响度不同,运用注意力机制让模型关注影响度较大的隐藏状态,实时分配行人模式数据中提取的权重,增加有效特征的权重,提高准确度;第三步,将注意力输出向量输入全连接层,经过两层全连接层之后降低维度至分类维度,最终输出分类结果.

1)输入层:室内行人模式的时序数据经过滑动窗口提取之后,每一个窗口有 n 个采样点,每一个采样数据中包含 m 个数据,从而得到样本输入矩阵 $\mathbf{R}^{n \times m}$, n 为每个样本输入的序列长度, m 为输入数据维度.

2)Bi-LSTM 层:室内行人模式需要通过整个滑动窗口的时序数据的信息才能更加精准地判断,因此本文运用 Bi-LSTM 网络提取数据特征,然后前馈到同一个输出层.该层生成了一系列的隐藏状态 $h_i \in \mathbf{R}^{2n \times m}$, $2n$ 表示数据序列中采样点个数的 2 倍, m 表示隐藏状态的向量维度.

3)参数舍弃(Dropout)^[25]:在 Bi-LSTM 层中加入 Dropout,大规模的神经网络存在费时和容易过拟合的缺点,加入 Dropout 之后每一次训练都会随机剔除神经网络中的部分神经元,解决模型过拟合的问题,提高模型的泛化能力.

4)注意力机制层:深度学习的注意力机制^[26-28]主要体现为在信息处理的过程中,对于不同的内容分配

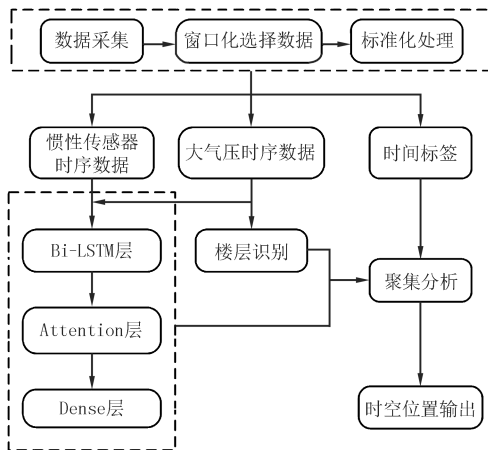


图4 结合注意力机制和Bi-LSTM的行人模式识别室内疫情追踪流程图
Fig.4 Flowchart of indoor epidemic tracking for pedestrian pattern recognition combined with attention mechanism and Bi-LSTM

不同的注意力权重,在一定程度上提升神经网络处理长序列问题的能力。

先使用注意力机制层再使用 Bi-LSTM 层可以提高模型的性能和效率,尤其是在长序列数据集。先使用 Bi-LSTM 层再使用注意力机制层也可以提高深度学习模型的性能,同时也可以提高模型在复杂序列数据上的鲁棒性,本文考虑到在室内行人模式识别中选取滑动窗口较小,滑动窗口内代表行人模式的惯性传感器数据较为复杂,将该数据先通过 Bi-LSTM 层提取当前时刻之前之后的特征信息,再将这些信息作为注意力机制层的输入,更好地捕捉特征信息中的关键信息,加强与行人模式相关性大

的输入特征,进一步提升神经网络对特征的挖掘能力,从而提高模型的学习性能以及鲁棒性。

因此本文在 Bi-LSTM 层后引入注意力层,将得到 Bi-LSTM 层的输出 $H = \{h_1, h_2, h_3, \dots, h_n\}$ 作为注意力层的输入,其中 n 是输入样本时序数据的长度,然后通过打分函数 $s(h, q)$ 计算查询向量 q 和每个输入 h_i 之间的相关性,得出分数,最后使用 softmax 函数对分数进行归一化,归一化后的结果便是注意力权重 $a = \{a_1, a_2, a_3, \dots, a_n\}$,以 a_i 为例,具体公式如下:

$$a_i = \text{softmax}(s(h, q)) = \frac{\exp(s(h_i, q))}{\sum_{j=1}^n \exp(s(h_j, q))} \quad (6)$$

将当前时刻得到的注意力权重与对应特征相乘得到加权后注意力输出向量 s_i 。

5)全连接层:全连接层连接注意力机制输出的,通过两层全连接层,综合之前神经网络的所有信息,降低维度至 4 维,便于输出分类,得到输出特征向量。

6)输出层:将特征向量 s_i 输入 softmax 分类器,输出对室内行人模式分类的预测概率,最终取最大的预测概率对应的类别作为分类结果。

2.3 楼层识别

实现高精度的室内行人聚类或追踪不仅要求对行人进行水平位置的模式识别,同样需要对行人进行准确的三维空间的定位,因此能够实时提供行人当前所在的楼层位置是必要的。基于差分气压的楼层切换识别技术成本较低,精度较高,本文使用以差分气压测高、通过室内行人模式修正的楼层识别方法,在较低成本的前提下保证提升楼层识别率^[29-30]。

根据气象学中两个不同海拔高度的大气压强差可表述为:

$$h - h_0 = \frac{R_d T}{g} \ln \frac{P_0}{P}, \quad (7)$$

式(7)中 h 为测量气压值对应的高度, m; h_0 为基准气压值对应的高度, m; g 为重力加速度 ($g = 9.8067 \text{ m/s}^2$); R_d 为干燥空气气体常数 ($R_d = 287.0529 \text{ m}^2/(\text{s}^2\text{K})$); T 为空气的热力学温度, K; P 为测量气压值; P_0 为基准气压值。

记 $\Delta h = h - h_0$, 则 Δh 表示测量气压的盖度与水平面的相对高度, 将热力学温度 T 转换为摄氏度 t , 带入相关参数的数值得到以下表达式:

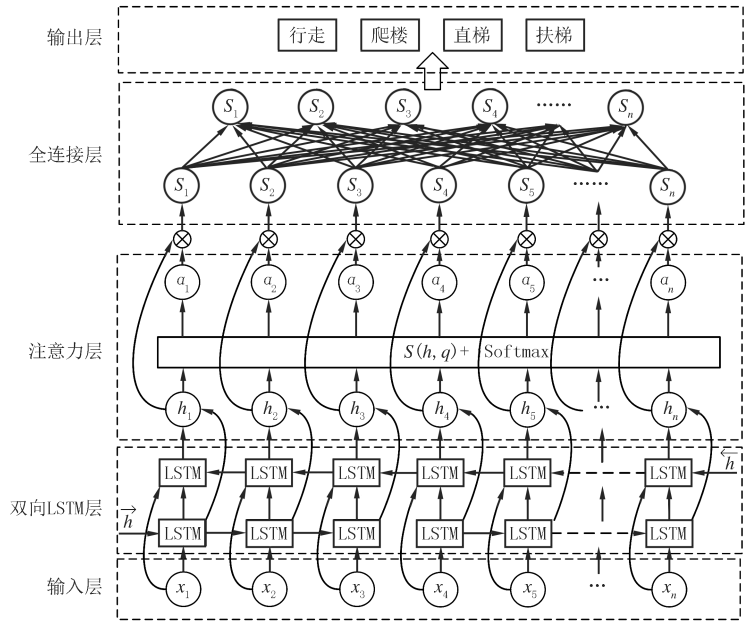


图5 结合注意力机制和Bi-LSTM网络结构图

Fig.5 Combined attention mechanism and Bi-LSTM network structure diagram

$$\Delta h = 67.399\ 5(273.15 + t)\lg \frac{P_0}{P}. \quad (8)$$

根据式(8)即可求得测量位置的高度信息、每层楼的高度及初始楼层信息确定行人的楼层号.当检测到行人处于行走模式时,通过求得的高度信息,确定行人所在楼层;当检测到行人处于爬楼、扶梯、直梯模式时,通过对窗口内气压数据的改变结合求得的层高数据,确定行人所在的具体位置,如式(9):

$$F = F_0 + \text{round}\left(\frac{1}{h_0}67.399\ 5(273.15 + t)\lg \frac{P_0}{P}\right), \quad (9)$$

式中, F 为行人所在的楼层号, F_0 为行人的初始楼层号, h_0 为建筑每层楼的高度,round 函数实现对求得高度的取整运算.

3 实验验证

3.1 实验设置

实验平台操作系统为 64 位 Windows 10 系统,CPU 为 AMD Ryzen 9 5950X 16-Core Processor,GPU 为 NVIDIA GeForce RTX 3090 24 G 独立显卡,使用 pytorch1.12.1 构建深度学习模型.

本文数据基于智能手机传感器采集,实验选择 8 名志愿者在自然环境中采集数据,其中男生 5 人,女生 3 人.在采集过程中,数据采样频率是 100 Hz,每名志愿者每种状态的测量时间约为 5 min,共采集了室内典型的 4 种行人模式步行、爬楼、扶梯、直梯数据,采集的数据均包含 4 种行人模式,且在采集时不受任何约束,不规定出脚顺序,完全按照自己的行为习惯执行相应的动作.

为了验证本文提出的结合注意力机制和 Bi-LSTM 网络的室内行人模式识别模型的实际效果,在实验前,对数据集进行滑动窗口化采样、标准化等预处理,最终得到 45 391 个数据样本,为了评估深度学习模型的稳定性与鲁棒性,将数据样本以 8 : 1 : 1 的比例随机分为训练集、验证集和测试集对模型进行训练和测试.

3.2 模型试验结果及分析

3.2.1 分类实验评估

本文采用决策树、随机森林和 SVM 3 种较为典型的机器学习算法利用准确度指标,使用 Scikit-learn 作为算法构建工具.通过对传统机器学习及深度学习的方法与本文提出的结合注意力机制和 Bi-LSTM 网络的模型形成对比实验,进行讨论.结果如表 1 所示.

表 1 分类实验评估

Tab. 1 Comparative Experiments on Traditional Machine Learning

名称	精准度/%	召回率/%	准确度%	名称	精准度/%	召回率/%	准确度%
决策树	91.70	87.50	88.34	LSTM	93.43	94.64	90.75
随机森林	90.24	90.52	89.83	本文模型	98.75	97.79	99.32
SVM	90.70	92.86	90.17				

分析对比实验结果可知,结合注意力机制和 Bi-LSTM 网络结构的模型在行人模式识别的准确度为 99.32%,且精准度平均提高了 7.23%,召回率平均提高了 6.41%,均优于其他机器学习算法.

3.2.2 超参数选择及性能评价

在深度学习的神经网络中,超参数是通过人为提前设定好网络的相关参数用于训练模型.为了更加客观全面地证明本文所提方法的准确性与高效性,运用控制变量的方法通过调整 Bi-LSTM 网络的层数、隐藏层节点数寻求最优参数组合.

(1)Bi-LSTM 网络层数.Bi-LSTM 的层数对模型的准确度、实时性与计算复杂程度有很大的影响.如果层数设置得过小,模型训练速度变快,但是精准度会有所下降,收敛速度也会变得很慢;而如果层数设置得过大,由于 Bi-LSTM 为时序模型,层数的增加会造成训练时间的增加以及占用内存指数级增长.因此设置实验的不变量:学习率为 0.000 1,隐藏层节点数为 128,全连接层层数为 2 层,全连接层神经元分别为 128 和 32

(未加激活函数),训练次数(epoch)为 150,对 Bi-LSTM 网络层数改变进行对比实验讨论,具体如表 2。

通过上述对比实验分析可知:Bi-LSTM 网络层数由 1 层增加到 5 层模型准确度上升了 11.01%,由第 5 层增加到第 7 层模型准确度下降了 1.56%。可以看出,改变 Bi-LSTM 网络层数模型的准确度会随着层数的增加而增加,从第 2 层开始模型随着层数增加准确度增长缓慢,增加到第 5 层时开始呈现下降趋势,但不明显;直到层数为第 8 层时,准确度下降了 5.30%,下降幅度较大。网络层数的增加导致神经元个数增多,极大地增加了模型的训练时间和占用内存。

表 2 不同 Bi-LSTM 网络层数模型对比

Tab. 2 Comparison of models with different layers of Bi-LSTM networks

网络层数	1	2	3	4	5	6	7	8
准确度/%	83.80	94.73	94.84	95.03	94.81	94.25	93.47	88.17
模型大小/kB	277	793	1 311	1 827	2 343	2 860	3 376	3 892
训练时间/s	17.12	18.46	19.88	21.074	22.67	24.21	26.21	28.31

(2)隐藏层节点数。一般来讲,隐藏层节点数是导致模型过拟合的直接原因,节点数过小,模型不具有泛化能力^[31];节点过多,模型容易陷入局部最优,甚至导致过拟合的情况^[32]。因此设置实验的不变量:学习率为 0.000 1,Bi-LSTM 网络层数为 2 层,全连接层层数为 2 层,全连接层神经元个数分别为 128 和 32(未加激活函数),epoch 为 150,对隐藏节点数改变进行对比试验讨论,具体如表 3。

表 3 不同隐藏层节点数模型对比

Tab. 3 Comparison of node number models of different hiding layers

隐藏层节点数	32	64	128	256
准确度/%	86.34	92.47	94.82	95.03
模型大小/kB	56	206	794	3 122
训练时间/s	15.12	16.46	18.88	25.74

通过上述对比实验分析可知,隐藏层节点数由 32 增加到 256,模型准确度上升了 8.69%,模型大小增加了 3 066 kB,训练一个 epoch 耗费的时间增加 10.62 s,可以看出,改变隐藏层节点数模型的准确度随着隐藏层节点数的增加而增加,但增加速率呈下降趋势;模型所占储存空间的大小、训练一个 epoch 耗费的时间与节点个数呈正相关。

移动端模型需满足模型尺寸小、模型计算复杂程度低、耗电量低等条件。为了满足室内行人模式识别/场景感知的实时性、高效性和准确性,模型最终确定相关超参数如表 4。

表 4 最优模型超参数设置

Tab. 4 Optimal model hyperparameter settings

超参数	训练次数	学习率	Bi-LSTM 层数	隐藏层节点数	Dropout
数值	150	0.000 1	2	128	0.5

3.2.3 模型结构改进及必要性讨论

(1)全连接层改进。全连接层层数以及是否增加激活函数。本文中将注意力机制层输出结果作为全连接层的输入,全连接层具有局部特征整合,实现最终预测分类的作用,会对数据特征的选取以及模型的准确度造成影响。因此设置实验不变量:学习率为 0.000 1,Bi-LSTM 网络层数为 2 层,隐藏层节点数为 128,epoch 为 150,设置 A₁、B₁、C₁、D₁ 组别,对全连接层的层数和是否增加激活函数进行对比试验讨论,具体如表 5。

对比实验结果显示,神经元个数越增加,模型越复杂,模型大小与训练时间均增加;全连接层数增加,模型非线性表达能力提高,提高了模型的学习能力。虽然 D₁ 组模型在模型大小以及训练时间上不是最优选择,但其行人模式识别的准确度为 96.48%,较其他组别模型有较大提升。

(2)消融实验。为了进一步验证结合注意力机制和 Bi-LSTM 模型在室内行人模式识别上的优势,探究模型架构中各部分所起的作用,在上文确定的超参数的基础上,加入多组消融实验分析模型各个模块对整体模型性能的影响程度。消融实验步骤如下:1)A₂ 组设置网络结构为单向,不引入注意力机制,模型其余超参数

不变;2)B₂组设置网络结构为双向,不引入注意力机制,模型其余超参数不变;3)C₂组设置网络结构为单向,引入注意力机制,模型其余超参数不变;4)D₂组为本文网络结构,将注意力机制与Bi-LSTM相结合.消融实验分组及实验结果如表6.

表 5 不同全连接层网络结构对比

Tab. 5 Comparison of network structures with different fully connected layers

组别	层数	神经元个数	激活函数	准确度/%	模型大小/kB	训练时间/s
A ₁	1	128	无	90.34	437	17.04
B ₁	1	128	ReLU	91.47	544	17.63
C ₁	2	128,32	无	94.79	794	18.57
D ₁	2	128,32	ReLU	96.48	907	19.05

表 6 消融实验

Tab. 6 Ablation experiment

组别	LSTM 网络结构为双向	注意力机制	准确度/%	模型大小/kB	训练时间/s
A ₂			90.75	586	17.52
B ₂	✓		96.48	907	19.05
C ₂		✓	95.58	854	18.58
D ₂	✓	✓	99.32	1 097	19.97

由表6可知,B₂组相较于A₂组将LSTM网络结构变为双向,准确度上升了5.73%,模型大小增加了321 kB,每个epoch训练时间增加了1.53 s;C₂组相较于A₂组引入注意力机制,准确度上升了4.73%,模型大小增加了286 kB,每个epoch训练时间增加了1.06 s;本文所提出的完整模型D₂组相较于A₂组,虽然模型大小与训练时间不是最优,但每个epoch训练时间仅增加了2.45 s,在实际应用中细分到每个滑动窗口增加的预测时间可忽略不计,其准确度上升了8.57%,得到大幅增加.通过将D₂组与B₂、C₂组对比可得,使用Bi-LSTM网络结构和引入注意力机制均可增加行人模式识别准确度,这在一定程度上反映了在室内行人模式识别中Bi-LSTM通过双向处理时序数据的特性,更能充分挖掘数据深层次的特征,提高数据的利用率;引入注意力机制后将样本数据的关键信息突出,使得网络能够更有效地学习数据的相互关系,提升了模型的准确度与鲁棒性.综上所述,本文充分发挥了注意力机制与Bi-LSTM网络结构的优势,使模型在室内行人模式识别/场景感知中有效运用.

(3)模型性能验证.在测试集上对所提模型进行了5次评估,表7给出了分类的混淆矩阵.结果表明,在行走、爬楼、直梯和扶梯分类中,本文算法对各种行人模式识别的准确度平均为99.32%.

表 7 混淆矩阵

Tab. 7 Confusion matrix

模式	行走	爬楼	直梯	扶梯	准确度/%	模式	行走	爬楼	直梯	扶梯	准确度/%
行走	1 071	18	0	0	99.17	直梯	0	0	1 091	2	99.81
爬楼	3	1 207	0	5	99.34	扶梯	0	4	8	1 131	98.95

4 结 论

本文以实现室内行人聚类与追踪中的行为模式识别/场景感知为出发点,提出了一种结合注意力机制和Bi-LSTM网络的行人模式识别模型,对室内典型的行人模式进行准确识别.通过进行多组对比实验与消融实验,对模型参数进行了最优化的组合,实验结果表明,与传统机器学习的方法相比,所提模型准确度提高了7.49%~10.98%,与不同网络结构的深度学习模型相比准确度平均提高7.36%.并且在构建优化模型结构时综合考虑了模型大小与训练时间,本文提出的模式识别方法能够实时、高效、准确地识别出室内行人与位置强关联的典型行为模式,能够满足室内行人模式识别的需求,为室内精准定位追踪奠定基础.在未来的工作

中,计划将引入更复杂的手机携带模式,使其更加多元化,提高模型的泛化能力与室内行人模式识别/场景感知的性能,提供一种应用更广泛的模型.

参 考 文 献

- [1] 王竣,王修晖.特征融合的多视角步态识别研究[J].中国计量大学学报,2017,28(2):234-240.
WANG J,WANG X H.Research on multi-perspective gait recognition using feature fusion[J].Journal of China University of Metrology, 2017,28(2):234-240.
- [2] 杨凯文,李双群,胡星.显著性时空特征融合的多视角步态识别算法[J].现代计算机,2022,28(17):9-15.
YANG K W,LI S Q,HU X.Fusion of salient spatio-temporal features based multi-view gait recognition[J].Modern Computer,2022, 28(17):9-15.
- [3] 杨韞韬,聂勇伟,张青,等.基于RNN和注意力机制的双向人体姿态补全方法[J].计算机辅助设计与图形学学报,2022,34(11):1772-1783.
YANG Y T,NIE Y W,ZHANG Q,et al.Bi-directional human pose completion based on RNN and attention mechanism[J].Journal of Computer-Aided Design & Computer Graphics,2022,34(11):1772-1783.
- [4] 闫昊雷,李小春,张仁飞,等.融合多尺度注意力和双向LSTM的行人重识别[J].空军工程大学学报,2022,23(5):71-76.
YAN H L,LI X C,ZHANG R F,et al.A pedestrian re-ID with multi-scale attention and bidirectional LSTM[J].Journal of Air Force Engineering University,2022,23(5):71-76.
- [5] 欧群雍,谭同德,袁红斌.结合CNN和Bi-LSTM的多行人目标检测跟踪方法[J].无线电工程,2022,52(9):1633-1641.
OU Q Y,TAN T D,YUAN H B.Research of detection and tracking method for multiple pedestrians based on CNN and Bi-LSTM[J]. Radio Engineering,2022,52(9):1633-1641.
- [6] 刘宇,江宏毅,王仕亮,等.基于加速度时域特征的实时人体行为模式识别[J].上海交通大学学报,2015,49(2):169-172.
LIU Y,JIANG H Y,WANG S L,et al.Real-time human activity pattern recognition based on time domain features of acceleration[J]. Journal of Shanghai Jiao Tong University,2015,49(2):169-172.
- [7] 衡霞,王忠民.基于手机加速度传感器的人体行为识别[J].西安邮电大学学报,2014,19(6):76-79.
HENG X,WANG Z M.Human activity recognition based on accelerometer data from a mobile phone[J].Journal of Xi'an University of Posts and Telecommunications,2014,19(6):76-79.
- [8] 段小虎,蒋刚,留沧海.基于手机加速度传感器的人体步态识别研究[J].传感器与微系统,2021,40(1):30-33.
DUAN X H,JIANG G,LIU C H.Study on human gait recognition based on cell phone acceleration sensor[J].Transducer and Microsystem Technologies,2021,40(1):30-33.
- [9] 张烈平,匡贞伍,李昆键,等.基于加速度传感器和神经网络的人体活动行为识别[J].现代电子技术,2019,42(16):71-74.
ZHANG L P,KUANG Z W,LI K J,et al.Human activity behavior recognition based on acceleration sensor and neural network[J].Modern Electronics Technique,2019,42(16):71-74.
- [10] 刘旭,徐正蕻,朱金鑫,等.辅助室内定位的关键人体姿态识别[J].科学技术与工程,2017,17(12):211-217.
LIU X,XU Z Y,ZHU J X,et al.Key human gesture recognition for assisting indoor positioning[J].Science Technology and Engineering, 2017,17(12):211-217.
- [11] 王玉坤,高炜欣,王征,等.基于加速度传感器的人体姿态实时识别[J].计算机工程与设计,2016,37(11):3092-3096.
WANG Y K,GAO W X,WANG Z,et al.Real-time human activity pattern recognition based on acceleration[J].Computer Engineering and Design,2016,37(11):3092-3096.
- [12] 邓平,吴明辉.基于机器学习的人体运动姿态识别方法[J].中国惯性技术学报,2022,30(1):37-43.
DENG P,WU M H.Human motion attitude recognition method based on machine learning[J].Journal of Chinese Inertial Technology, 2022,30(1):37-43.
- [13] ZHANG Y S,YANG A R,XIONG C,et al.Feature selection using data envelopment analysis[J].Knowledge-Based Systems,2014,64:70-80.
- [14] 高丽丽,赵焯焯.基于智能手机的LSTM室内定位算法研究[J].计算机仿真,2022,39(9):525-531.
GAO L L,ZHAO Y Z.Research on indoor location algorithm of LSTM based on smart phone[J].Computer Simulation,2022,39(9):525-531.
- [15] HAN J W,KAMBER M.数据挖掘:概念与技术[M].范明,孟小峰译.北京:机械工业出版社,2012.
- [16] BHARATI M,RAMAGERI B.Data mining techniques and applications[J].Indian Journal of Computer Science and Engineering,2010, 1(4):25-47.
- [17] JIANG L X,LI C Q.Scaling up the accuracy of decision-tree classifiers:a naive-bayes combination[J].Journal of Computers,2011,6(7): 1325-1331.
- [18] QUINLAN J R.Induction of decision trees[J].Machine Learning,1986,1(1):81-106.
- [19] BREIMAN L.Bagging predictors[J].Machine Learning,1996,24(2):123-140.
- [20] BREIMAN L.Random forests[J].Machine Learning,2001,45(1):5-32.

- [21] GAO L Y, YE M Q, LU X J, et al. Hybrid method based on information gain and support vector machine for gene selection in cancer classification[J]. *Genomics, Proteomics & Bioinformatics*, 2017, 15(6): 389-395.
- [22] NGUYEN T H, PHAM T P, NGO C Q, et al. A SVM algorithm for investigation of tri-accelerometer based falling data[J]. *Am J Signal Process*, 2016, 6: 56-65.
- [23] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. *IEEE Transactions on Signal Processing*, 1997, 45(11): 2673-2681.
- [24] 金宸, 李维华, 姬晨, 等. 基于双向 LSTM 神经网络模型的中文分词[J]. *中文信息学报*, 2018, 32(2): 29-37.
JIN C, LI W H, JI C, et al. Bi-directional long short-term memory neural networks for Chinese word segmentation[J]. *Journal of Chinese Information Processing*, 2018, 32(2): 29-37.
- [25] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [26] ZHANG H Y, HUANG H M, HAN H. Attention-based convolution skip bidirectional long short-term memory network for speech emotion recognition[J]. *IEEE Access*, 2020, 9: 5332-5342.
- [27] 陈海涵, 吴国栋, 李景霞, 等. 基于注意力机制的深度学习推荐研究进展[J]. *计算机工程与科学*, 2021, 43(2): 370-380.
CHEN H H, WU G D, LI J X, et al. Research advances on deep learning recommendation based on attention mechanism[J]. *Computer Engineering & Science*, 2021, 43(2): 370-380.
- [28] 任欢, 王旭光. 注意力机制综述[J]. *计算机应用*, 2021, 41(S1): 1-6.
REN H, WANG X G. Review of attention mechanism[J]. *Journal of Computer Applications*, 2021, 41(S1): 1-6.
- [29] 滕帆. 基于气压测高的楼层判别方法研究[D]. 徐州: 中国矿业大学, 2019.
- [30] 王玮. 基于差分气压测高的楼层切换和卫星定位关键技术研究[D]. 北京: 北京邮电大学, 2018.
- [31] 鲁南, 欧阳权, 黄俊卉, 等. 基于注意力机制和多任务 LSTM 的锂电池容量预测方法[J]. *电气工程学报*, 2022, 17(4): 41-50.
LU N, OUYANG Q, HUANG L H, et al. Capacity prediction of lithium-ion batteries based on multi-task LSTM with attention mechanism[J]. *Journal of Electrical Engineering*, 2022, 17(4): 41-50.
- [32] 马晓伟. 基于双向 LSTM 模型的英语语法错误检测[J]. *信息技术*, 2022, 46(9): 56-60.
MA X W. English grammar error detection based on bidirectional LSTM model[J]. *Information Technology*, 2022, 46(9): 56-60.

Indoor pedestrian pattern recognition based on bidirectional long short-term memory network and attention mechanism

Liang Yujie^a, Cui Bo^{a,b}

(a. Artificial Intelligence; b. Hebei Key Laboratory of Industrial Intelligent Perception North China University of Science and Technology, Tangshan 063210, China)

Abstract: In indoor space, accurate pedestrian pattern recognition/scene perception, especially the recognition/perception associated with location, is of great significance for pedestrian gathering or tracking. Aiming at the problems of traditional machine learning methods such as difficulty in feature extraction, low classification accuracy, and large recognition errors caused by abnormal behaviors, this paper proposes an indoor real-time pedestrian pattern recognition model based on attention mechanism and bidirectional long short-term memory (Bi-LSTM) network. The Bi-LSTM network was established to extract the temporal characteristics of the pedestrian mode in the sliding window, evaluate the performance and timeliness of the model network structure, optimize the number of Bi-LSTM layers and the number of hidden layer nodes, and determine the optimal network structure. In order to reduce the influence of noise data on the model and improve the ability of the network over screen information features, the attention mechanism is introduced to optimize the weight parameters of the extracted temporal features. The experimental results show that compared with the traditional machine learning algorithm, the accuracy of pedestrian pattern recognition in the optimized Bi-LSTM network is improved by 6.37% on average. After further introducing the attention mechanism, the accuracy of pedestrian pattern recognition is improved by 9.21% on average, and the final accuracy can reach 99.32%. The proposed model can effectively classify the pedestrian mode/scene perception, and provide method support for accurate indoor positioning and tracking.

Keywords: pedestrian pattern recognition; sliding window; temporal feature; Bi-LSTM; mechanism of attention