

## 无监督特征学习的人体活动识别\*

史殿习, 李勇谋, 丁博

(国防科技大学 计算机学院, 湖南 长沙 410073)

**摘要:** 针对人的局限性可能会导致在提取特征中丢失重要信息, 从而影响最终的识别效果问题, 提出无监督特征学习技术的惯性传感器特征提取方法。其核心思想是使用无监督特征学习方法学习多个特征映射, 再将所有特征映射拼接起来形成最终的特征计算方法。其优点是会造成重要信息的损失, 而且可以显著减少所使用的无监督特征学习模型的规模。为了验证所提出的特征提取方法在活动识别中的有效性, 运用一个公开的活动识别数据集, 使用三种常用无监督模型进行特征提取, 并使用支持向量机进行活动识别。实验结果表明, 特征提取方法取得了良好的效果, 与其他方法相比具有一定的优势。

**关键词:** 人体活动识别; 无监督特征学习; 智能手机; 传感器

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1001-2486(2015)05-128-07

## Unsupervised feature learning for human activity recognition

SHI Dianxi, LI Yongmou, DING Bo

(College of Computer, National University of Defense Technology, Changsha 410073, China)

**Abstract:** To solve the problems that human limitations may cause the loss of important information, thus affecting the classification results, a feature extraction method based on unsupervised feature learning techniques was proposed. Unsupervised feature learning method to learn multiple feature maps was used and concatenated together. This method can avoid the loss of important information, and also can significantly reduce the scale of unsupervised feature learning model used. To evaluate the proposed method, experiments on a public human activity recognition dataset were performed, using three commonly used unsupervised feature learning models, and finally using support vector machines to classify activities. The results show that the proposed feature extraction method achieves good results, and has certain advantages compared with other methods.

**Key words:** human activity recognition; unsupervised feature learning; smartphone; sensors

目前广泛使用的智能手机如 Apple iPhone 以及 Android 的各种型号的智能机, 大都内置了多种传感器, 例如加速度传感器、陀螺仪传感器、距离传感器、光传感器以及 GPS 等。内嵌了各种传感器的智能手机的广泛普及和应用, 催生了崭新的基于智能手机传感器的应用和研究, 例如人体活动识别、跌倒检测、身份识别等<sup>[1]</sup>。其中, 人体活动识别是当前的热点研究问题。活动识别可以作为健康评估和医疗保健的干预工具, 通过长时间识别和记录用户的活动, 可以计算出用户的活动水平和能量消耗, 为活动太多或太少的用户提供反馈信息, 帮助其建立健康的生活方式。

文献[1]中对现有的活动识别研究做了比较全面的论述, 按识别的活动的类别可以分为: 运动类、交通模式类、锻炼类、日常类和手机使用方式类活动等。目前, 运动类活动识别是当前的研究

热点问题, 也是史殿习等关注的重点, 通常包括走路、上楼梯、下楼梯、跑步、坐、站、躺等, 通常涉及加速度计和陀螺仪两种惯性传感器的一种或者两种。

活动识别是一个典型的时间序列分类问题, 其问题可以定义为: 确定传感器数据流的某一连续部分是由哪种人体活动产生的。活动识别的处理过程一般分为以下三个过程: 数据预处理、特征提取和分类<sup>[1]</sup>。

数据预处理通常包括噪声消除和数据流分割。采集数据时, 由于外界环境的影响如路面不平以及智能手机位置的变动, 会使数据包含大量的噪声, 噪声消除通常采用的是低通滤波器技术。由于从连续的传感器数据流提取有用信息比较困难, 并且大部分的分类算法无法处理连续的数据流, 因此一般使用具有 50% 重叠、长度固定的滑

\* 收稿日期: 2015-05-03

基金项目: 国家自然科学基金资助项目(61202117, 91118008)

作者简介: 史殿习(1966—), 男, 山东龙口人, 教授, 博士, E-mail: dxshi@nudt.edu.cn

动窗口将数据分割为等长的段,如图1所示。

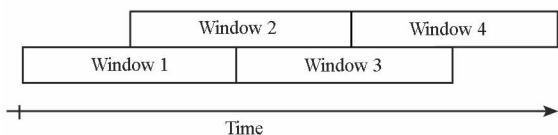


图1 50%重叠的滑动窗口

Fig. 1 Sliding windows with 50% overlap

经过预处理后,得到固定长度的数据,直接用于分类时效果较差,一般需要先进行特征提取。所谓特征,通常是指原始数据的抽象表示,好的特征是分类成功的关键。特征提取的本质是确定从原始数据到特征的变换方法,这种变换方法大部分都是由研究人员,根据一些特定领域的专业知识人工设计的。常用的特征分为时域特征和频域特征:时域特征是直接从时域数据中计算出的一些统计值,例如均值、标准差、能量、熵和相关系数等;频域特征通常是使用快速傅里叶变换得到原始数据的频域表示,然后基于频域表示提取特征。常用的这两类特征依赖领域知识,在应用于不同类别的活动时,一是可能会存在不适应特定数据的问题;二是它们通常是由人工设计的,可能由于设计的失误,导致在提取特征后,丢失重要信息,从而影响分类效果。为了减少这两个问题的影响,可以使用无监督特征学习技术<sup>[2]</sup>,该技术能够直接从传感器数据中学习特征的变换方法。在使用无监督特征学习技术提取特征时,不需要依赖领域知识,而且由于是从数据中学习特征,因此也不存在数据不适应问题。据悉,目前使用该技术对惯性传感器数据进行特征提取,进而进行活动识别的研究仅有文献[3-5]。文献[3]最早使用无监督特征学习技术如深度信念网络(Deep Belief Networks, DBN)和主成分分析(Principal Components Analysis, PCA)技术对加速度传感器数据进行处理,学习特征变换方法,进而提取相应的特征对厨房活动、日常活动以及汽车生产流水线中工人的活动等进行识别;文献[4]在文献[3]的工作基础上,使用平移不变性稀疏编码技术提取特征;文献[5]亦使用稀疏编码技术进行特征提取,进而对运动类活动和交通模式类活动如静止、走路、乘车、骑自行车等进行识别。

活动识别的最后一步是分类。首先,选择合理的分类算法对分类器进行训练;然后,使用通过训练得到的分类器进行分类。目前,活动识别中所用的分类算法大多数是有监督分类算法<sup>[1]</sup>,比如支持向量机(Support Vector Machine, SVM)、人工神经网络(Artificial Neural Network, ANN)、决

策树 C4.5、朴素贝叶斯和 K 近邻等算法。相对而言,目前使用半监督的分类方法的研究相对较少,文献[6]研究了自我学习、协同学习和主动学习三种经典的半监督的分类方法在活动识别领域的应用;文献[7]使用无监督聚类算法来提高已有分类器的性能,亦属于半监督的分类方法。对于半监督的分类方法来说,由于其可以利用无标注的数据,在实际应用中效果较好,其识别准确率可以达到80%以上<sup>[1]</sup>。虽然目前有很多分类算法,选择合适的分类算法仍然是比较烦琐和具有挑战性的工作。

在活动识别的处理过程中,虽然活动类型是由分类器来识别和分类的,但是活动识别的前两个阶段也都具有十分重要的作用,预处理和特征表示都是获取高性能的分类结果的关键因素。针对活动识别领域使用的常用特征的一些缺陷,史殿习等提出一种基于无监督特征学习技术的惯性传感器特征提取方法,在此基础上,使用支持向量机 SVM 分类方法,在一个活动识别公开数据上进行了实验验证。

## 1 惯性传感器数据特征提取方法

### 1.1 基本思想

所涉及的惯性传感器包括三轴加速度传感器和陀螺仪传感器。这两种传感器都具有三个方向,与生活的三维空间相对应。但是手机的三维坐标系是基于手机屏幕的,如图2<sup>[15]</sup>所示。

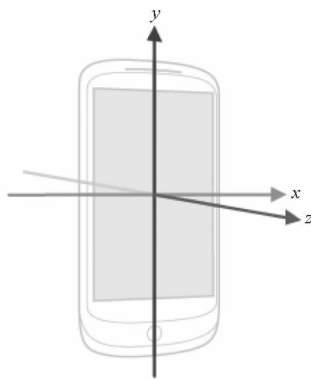


图2 智能手机坐标系

Fig. 2 Coordinate system of smartphones

当智能手机在人身上的位置和方向相对固定时,其采集的三维数据比较稳定,能够正确地反映身体的运动;当方向变化较大时,其合成量仍然可以可靠地测量身体运动的程度。基于上述考虑,增加一个维度的信息,特征提取时将使用四个维度的信息,即 $\{X, Y, Z, R\}$ ,将每个维度称为一个

channel, 即四个 channel。在此基础上, 提出一种基于无监督特征学习的惯性传感器特征提取方法, 并将其集成到经典的分类问题处理流程中, 如图 3 所示。首先, 将传感器数据流使用滑动窗口技术分割为等长的数据段, 分割后得到等长数据段有三个 channel, 即  $\{X, Y, Z\}$ ; 然后, 计算合成量  $R = \sqrt{X^2 + Y^2 + Z^2}$ , 将这四个 channel 的数据作为无监督特征学习模型的输入。该特征提取方法主要包含以下两个步骤:

1) 对每个 channel 的数据使用无监督特征学习技术, 学习特征变换方法, 如图 3 所示,  $\{X, Y, Z, R\}$  四个 channel 分别得到四个特征映射方法:  $\text{map}_X, \text{map}_Y, \text{map}_Z, \text{map}_R$ 。

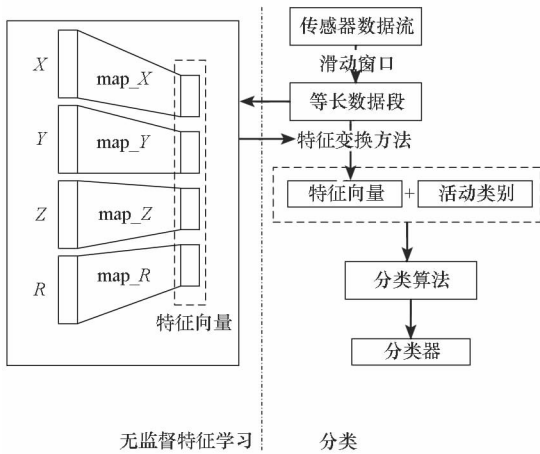


图 3 基于无监督特征学习的活动识别

Fig. 3 Activity recognition based on unsupervised feature learning

2) 把  $\{X, Y, Z, R\}$  四个 channel 的数据分别使用对应的特征映射方法  $\text{map}_X, \text{map}_Y, \text{map}_Z, \text{map}_R$  得到的特征拼接起来, 作为最终的特征向量, 用于分类。

将这种首先分 channel 使用无监督特征学习技术进行学习特征映射, 然后将其拼接到一起的特征提取方法, 称作基于无监督特征学习的 channel-wise 特征提取方法<sup>[8]</sup>。与之相对应的是不分 channel 的无监督特征学习<sup>[3-5]</sup>, 即将三个 channel 或者四个 channel 的原始数据拼接起来, 作为一个无监督特征学习模型的输入, 学习特征表示。这种基于无监督特征学习的 channel-wise 特征提取方法的优点主要包括: ①使用特征学习技术能够克服对领域知识的依赖, 也不会造成重要信息的损失; ②channel-wise 方式使用无监督特征学习技术, 与不分 channel 的方式相比, 可以显著减少所使用的无监督特征学习模型的规模, 例

如在基于神经网络的无监督特征学习方法中, 可以将权值数量降为 1/4, 从而减少训练时间和所需要的数据规模; ③channel-wise 方式更为灵活, 可以在不同的 channel 使用不同的超参数, 甚至不同的模型; ④使用 channel-wise 方式, 使实现模型的并行训练更为容易。虽然, 这种 channel-wise 方式不能提取不同 channel 间的相关信息, 但是由于无监督特征学习技术能够控制好特征提取带来的信息损失, 所以不同 channel 的相关信息仍然可以被分类算法所利用。

### 1.2 特征学习模型

在所提出的特征提取方法中, 需要用到无监督特征学习技术从每个 channel 学习特征映射方法。在经过对常见的无监督特征学习技术分析之后, 主要研究将两种基于自编码器 (Auto-Encoder, AE) 的模型和主成分分析 (Principal Components Analysis, PCA) 应用于传感器数据特征提取。其原因在于这两种模型具有一定的代表性, AE 模型是非线性模型, 而 PCA 是线性模型。

AE 通常定义为一个单隐藏层神经网络, 其输入单元的数量与输出单元的数量相同, 在训练时将输入值作为目标值进行训练。本质上 AE 是尝试去学习一个恒等函数, 换句话说, 它尝试逼近一个恒等函数, 从而使得输出  $\hat{x}$  接近于输入  $x$ 。

若令  $N$  和  $K$  分别表示输入单元的数量和隐藏单元的数量, 那么一个 AE 的参数可以形式化地表达为:  $W^{(1)} \in \mathbb{R}^{(K \times N)}$  和  $W^{(2)} \in \mathbb{R}^{(N \times K)}$  ( $W^{(1)}, W^{(2)}$  分别表示从输入层到隐藏层的权值和从隐藏层到输出层的权值),  $b^{(1)} \in \mathbb{R}^K$  和  $b^{(2)} \in \mathbb{R}^N$  ( $b^{(1)}, b^{(2)}$  分别表示隐藏层偏置和输出层偏置)。一个 AE 可以分解为两部分, 编码部分和解码部分。编码部分将输入  $x \in \mathbb{R}^N$  变换为隐藏层表示  $h(x) \in \mathbb{R}^K$ , 解码部分是使用隐藏层表示重建原始输入。隐藏层表示  $h(x)$  和原始输入的重建  $\hat{x}$  的计算方法为:

$$h(x) = \sigma(W^{(1)}x + b^{(1)}) \quad (1)$$

$$\hat{x} = \sigma(W^{(2)}h(x) + b^{(2)}) \quad (2)$$

其中,  $\sigma(\cdot)$  是激活函数, 通常使用 sigmoid 函数和 tanh 函数。AE 的目标是学习隐藏层表示  $h(x)$ , 也可以将  $h(x)$  看作特征映射。将用于训练 AE 的无标签的数据集记作  $X = \{x^{(1)}, \dots, x^{(m)}\}$ , 其中  $x^{(i)} \in \mathbb{R}^N$ , 采用均方误差作为衡量输入  $x$  和输出  $\hat{x}$  的相似程度的标准, 所以 AE 的代价函数可以表示为:

$$J(\theta) = \frac{1}{m} \sum_{x \in X} \|\hat{x} - x\|^2 \quad (3)$$

其中,  $\theta$  表示 AE 的参数集合  $\{W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}\}$ 。

不作限制地使用 AE 通常并不能取得好的效果,需要对 AE 增加某些限制。为了学习获取有用的表示,一种常用的限制是对隐藏层单元施加稀疏性约束,一般认为超完备的稀疏表示能够取得的性能高于规模更小的表示<sup>[9]</sup>。稀疏性约束是指限制每个隐藏层单元的平均激活值,使其接近于一个固定的常数  $\rho \in (0, 1)$ ,  $\rho$  通常是一个接近于 0 的较小的值(比如  $\rho = 0.05$ )。其中,隐藏神经元  $j$  在训练集上的平均活跃度用  $\hat{\rho}_j$  表示。这种具有稀疏性约束的 AE 通常被称为稀疏自编码器(Sparse Auto-Encoder, SAE)。稀疏性约束是通过在 AE 的代价函数加入用于描述神经元的平均激活值  $\hat{\rho}_j$  和  $\rho$  的相差程度的 KL(Kullback-Leibler)散度(又称相对熵)实现的,其中 KL 散度的定义为:

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log_2 \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log_2 \frac{1 - \rho}{1 - \hat{\rho}_j}。$$

因此,SAE 的代价函数为:

$$J(\theta) = \frac{1}{m} \sum_{x \in X} \|\hat{x} - x\|^2 + \beta \sum_{j=1}^K KL(\rho \parallel \hat{\rho}_j) \quad (4)$$

式(4)中,第一项为均方误差,第二项为稀疏性惩罚,  $\beta$  为调节两项的比例的参数。由于 KL 散度是平滑且可微的,所以其代价函数也是可微的,可以使用基于梯度的优化方法进行训练。当最小化代价函数时,第一项是用于减小重构时的信息损失,第二项使隐藏层单元的平均激活值保持在预设的水平。值得注意的是,由于 KL 散度只在  $(0, 1)$  上有定义,在使用 tanh 等值域为  $(-1, 1)$  的函数作为激活函数时,需要将隐藏单元的平均激活值调整到  $(0, 1)$  上。

AE 模型的另外一个扩展是去噪自编码器(De-noising Auto-Encoder, DAE), DAE 最初是在文献[10]中作为神经网络的基本构造块引入的。为了使 AE 能够学习到鲁棒性更好的特征,减轻噪声对特征的影响, DAE 试图从加入了噪声破坏的原始输入重建出无噪声的原始输入。在 DAE 中常用的噪声有以下两种<sup>[11]</sup>:

1) 高斯噪声:用  $GN(x)$  表示加入高斯噪声后的数据,则  $GN(x) = x + N(0, \sigma^2 I)$ , 其中  $\sigma^2 I$  为所加入的噪声的协方差矩阵。

2) Mask 噪声:对于每个输入  $x$ , 从所有元素中随机选择占比例  $p$  的元素,将其设置为激活函

数的值域的最小值,若使用 sigmoid 函数则置 0,若使用 tanh 函数则置 -1。

最后,在使用基于 AE 的模型进行特征学习时,为了克服单隐藏层网络表示能力的一些缺陷,通常的做法是将多个 AE 堆叠起来,即将一个 AE 的隐藏层输出作为另一个 AE 的输入。在训练时,先用原始数据训练最外层的 AE,得到其隐藏层输出后将其作为输入数据训练下一层 AE。

PCA 是一个被广泛使用的具有良好数学基础的维度缩减方法。它通过使用正交变换,将可能存在相关性的一定维度的变量转换成维度更小的无相关的变量,同时在维度缩减的过程中保持信息损失最小化。在 PCA 中,数据集的协方差矩阵的最大的几个特征值所对应的特征向量所张成的低维子空间,能够保持原始数据最大的方差。从原始数据到方差最大化的子空间的投影可以作为特征表示,也可以用于数据的可视化。

PCA 可以看作是一种无监督特征学习方法,主要是由于这种方法可以自动地从数据中学习得到紧凑且有意义的原始数据的表示,并且不需要依赖于领域知识。本质上,PCA 学习到一个线性变换  $f(x) = W^T x$ , 把数据映射为特征,这与使用线性激活函数的 AE 具有相似的效果。

### 1.3 超参数选择方法

机器学习模型中的参数可以分为两类:参数和超参数。寻找最优参数问题已经得到很好的解决,目前有很多优化算法可以使用。而对于超参数的选择,尚未得到很好解决,更多地依赖于经验和穷举搜索。然而,合理的超参数设置也对机器学习方法的性能具有至关重要的影响<sup>[12]</sup>。对于超参数的选择,采用基于经验与穷举搜索相结合的方法,其中,对于数值型超参数,选取合理的取值范围和搜索步长,进行网格搜索。

## 2 活动识别实验验证和结果分析

为了验证提出的特征提取方法在活动识别中的有效性,在公开数据集上设计实验,共提取了八种特征,最后使用 SVM 对常见的六种活动进行识别并分析实验结果。

### 2.1 活动识别数据集

实验的数据来源于一个常用的活动识别的数据集:UCI 机器学习资源库<sup>[13]</sup>中的“Human Activity Recognition Using Smartphones Data Set”<sup>[14]</sup>,简记为 UCI-HAR 数据集。该数据集所收集的数据来自一组 30 个年龄分布在 19 ~ 48 岁

的志愿者,每个人要求执行六种活动: WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING 和 LAYING,在执行这些活动的时候将一部智能手机固定在腰部。在活动的过程中,使用智能手机内置的加速度计和陀螺仪两种惯性传感器,以 50Hz 固定频率记录三轴线性加速度和三轴角速度。在实验进行过程中,使用视频记录用于人工的标记数据。所获得的数据集被随机分成两组,其中 70% 的志愿者被选择用于生成训练数据,另外 30% 用于生成测试数据。

该数据集中的传感器信号已经经过噪声消除,并且使用长度 2.56s 具有 50% 的滑动窗口进行分割,分割后的数据每个窗口 128 个采样点。滑动窗口分割后,训练集有 7352 个样例,测试集有 2947 个样例。每个样例的属性包括三轴加速度、三轴角速度、活动标签以及生成数据的人的编号。为了便于在 AE 中使用 sigmoid 函数作为编码函数和解码函数,将所有传感器数据都调整到  $[0.1, 0.9]$ 。

## 2.2 特征提取实验

为了验证特征提取方法的有效性,对比不同特征的效果,不仅使用两种基于特征学习的方法提取特征,还提取了常用的统计特征和频域特征。本节将简单描述以 channel-wise 和不区分 channel 两种方式使用特征学习技术进行特征提取的过程,以及统计特征和频域特征的定义和计算方法。

在进行特征学习时,使用 UCI-HAR 数据集中训练集与测试集的所有数据,共 10 299 个训练样例对特征学习模型进行训练。

在基于 AE 的方法中级联两层 AE,SAE 和 DAE 第一层网络结构参照 PCA 的合适压缩维度进行选择,SAE 第二层仍然采取同样方法,对于 DAE 第二层网络结构则参照 SAE 第二层。

### 2.2.1 基于 SAE 的 channel-wise 方式特征学习

通过使用 PCA 模型进行分析,发现当将 128 维的各个 channel 的数据压缩为 64 维时,所有 channel 均可以保持超过 99% 的方差比例。因此将隐藏层单元数目设置为 64 的 1.5 倍,所以第一层 SAE 网络结构为  $128 \times 96 \times 128$ 。根据 1.3 节提出的方法,稀疏性稀疏  $\beta = 128/96 \approx 1.33$ 。再在各个 channel 的数据中搜索最优的  $\rho$ ,搜索时对 SAE 使用 Pylearn2 中的带线搜索的批处理梯度下降法训练 200 个周期,模型的评价标准为最后 5 个周期 MSE 平均值。搜索结果(见表 1)表明,对于加速度传感器数据  $\{X, Y, Z, R\}$  最优  $\rho$  分别为  $\{0.25, 0.25, 0.30, 0.15\}$ ,陀螺仪传感器数据

$\{X, Y, Z, R\}$  最优  $\rho$  值分别为  $\{0.25, 0.10, 0.10, 0.40\}$ 。

表 1 所有 channel 第一层 SAE  $\rho$  搜索结果

Tab. 1 Search results of SAE  $\rho$  in first layer for all channels

channel	$\rho$							
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
Acc-R	0.051	0.037	0.034	0.037	0.040	0.044	0.046	0.049
Acc-X	0.119	0.083	0.074	0.071	0.069	0.081	0.085	0.094
Acc-Y	0.165	0.146	0.134	0.125	0.118	0.122	0.121	0.121
Acc-Z	0.113	0.089	0.083	0.079	0.076	0.071	0.080	0.076
Gyro-R	0.048	0.033	0.029	0.027	0.026	0.025	0.022	0.021
Gyro-X	0.057	0.054	0.052	0.051	0.048	0.049	0.053	0.055
Gyro-Y	0.062	0.053	0.058	0.061	0.062	0.055	0.060	0.057
Gyro-Z	0.051	0.044	0.046	0.044	0.049	0.053	0.058	0.059

在确定第一层 SAE 的各个超参数后,使用 Pylearn2 中的带线搜索的批处理梯度下降法,对其训练 200 个周期。训练后的第一层 SAE 的隐藏层输出作为下一层 SAE 的输入。

第二层 SAE 使用与第一层 SAE 相似的方法确定超参数。通过对第一层 SAE 的隐藏层输出使用 PCA 分析,发现 16 维为合适的压缩维度,因此第二层 SAE 的网络结构为  $64 \times 24 \times 64$ 。使用同样的方法搜索最优  $\rho$  取值,根据搜索结果,加速度传感器数据  $\{X, Y, Z, R\}$  最优  $\rho$  分别为  $\{0.30, 0.30, 0.40, 0.40\}$ ,陀螺仪传感器数据  $\{X, Y, Z, R\}$  最优  $\rho$  值分别为  $\{0.40, 0.05, 0.30, 0.20\}$ 。

在确定各个 channel 第二层 SAE 的最优稀疏性参数  $\rho$  的取值后,使用与第一层类似的训练方法和设置对其进行训练。将每个 channel 第二层 SAE 的隐藏层输出作为特征,每个 channel 得到 24 个特征,这种特征提取方法共提取了 192 个特征。

### 2.2.2 基于 DAE 的 channel-wise 方式特征学习

根据使用 SAE 时对数据的分析,将第一层 DAE 隐藏层单元数设为 64,第二层 DAE 隐藏层单元数设为 16。采用网格搜索方式来确定 DAE 的噪声类型和噪声水平,模型的选择标准基于训练集的 10-折交叉验证结果,分类器使用 one-vs-one 多类扩展的、带有径向基核函数(Radial Basis Function, RBF)核函数的 SVM, SVM 的两个超参数的搜索空间分别为  $C \in \{2^{-5}, 2^{-4}, \dots, 2^5\}$  和  $\gamma \in \{2^{-4}, 2^{-3}, \dots, 2^4\}$ 。搜索结果(见表 2)表明,第一层 DAE 使用  $\sigma^2 = 0.0025$  的 GN 噪声、DAE2 使用  $\sigma^2 = 0.0100$  的 GN 噪声所取得的效果最好。

这种方式共提取了 128 个特征。

表 2 channel-wise DAE 超参数搜索结果

Tab. 2 Search results of channel-wise DAE's hyper-parameters

第一层 DAE	第二层 DAE	CV 准确率/%
GN; $\sigma^2 = 0.0025$	GN; $\sigma^2 = 0.0025$	94.8
	GN; $\sigma^2 = 0.0100$	94.9
	MN; $p = 0.10$	94.5
	MN; $p = 0.20$	94.3
GN; $\sigma^2 = 0.0100$	GN; $\sigma^2 = 0.0025$	95.0
	GN; $\sigma^2 = 0.0100$	94.9
	MN; $p = 0.10$	94.5
	MN; $p = 0.20$	94.4
MN; $p = 0.10$	GN; $\sigma^2 = 0.0025$	94.5
	GN; $\sigma^2 = 0.0100$	94.6
	MN; $p = 0.10$	94.1
MN; $p = 0.20$	MN; $p = 0.20$	94.5
	GN; $\sigma^2 = 0.0025$	94.3
	GN; $\sigma^2 = 0.0100$	94.4
MN; $p = 0.20$	MN; $p = 0.10$	94.2
	MN; $p = 0.20$	94.1

### 2.2.3 基于 PCA 的 channel-wise 方式特征学习

为了使提取的特征具有可比性,使用 PCA 将每个 channel 128 维数据压缩为 16 维。最终也提取了 128 个特征。

### 2.2.4 不区分 channel 方式特征学习

为了对比 channel-wise 方式与不区分 channel 方式的特征学习方式,对应提取了三种不区分 channel 方式的特征。为了使结果具有可比性,对于不区分 channel 方式的学习,对每种传感器数据使用与 channel-wise 方式相对应的结构和参数。

在使用 SAE 模型进行不区分 channel 方式特征学习时,即第一层 SAE 为  $512 \times 384 \times 512$ ,第二层 SAE 为  $384 \times 96 \times 384$ 。使用与 channel-wise 方式相似的稀疏性参数搜索过程。由于网络规模较大导致训练时间较长,在不分 channel-wise 方式特征学习搜索中仅对模型训练 100 个周期。第一层搜索结果为:加速度数据 0.05,陀螺仪数据 0.10。在第一层 SAE 确定最优  $\rho$  后,仍然对其训练 200 个周期,然后将其隐藏层输出作为第二层 SAE 的输入对其进行参数搜索,搜索结果为:加

速度数据 0.10,陀螺仪数据 0.05。最后对第二层 SAE 使用最优参数训练 200 个周期,每种传感器数据得到 96 个特征,共计 192 个特征。

在使用 DAE 模型进行不区分 channel 方式特征学习时,第一层 DAE 网络结构为  $512 \times 256 \times 512$ ,第二层 DAE 网络结构为  $256 \times 64 \times 256$ 。对于不区分 channel 方式,DAE1 使用  $\sigma^2 = 0.0025$  的 GN 噪声、DAE2 使用  $p = 0.020$  的 MN 噪声所取得效果最好。这种方式共提取了 128 个特征。

在使用 PCA 模型进行不区分 channel 方式特征学习时,将所有加速度数据使用 PCA 压缩为 64 维度,陀螺仪数据也压缩为 64 维度。最终,同样提取了 128 维数据。

### 2.2.5 时域统计特征和频域特征

时域统计特征是基于智能手机的活动识别研究中使用最多的特征,其主要优点是计算简单,适用于在资源有限的智能手机上实现。选取最为常用的几个统计特征,包括从每个 channel 提取的均值、标准差、能量和熵,以及每个传感器提取的 6 个相关系数,共计 44 个特征。

频域特征也是基于智能手机的活动识别研究的常用特征,频域特征效果一般比统计特征要好,但是频域特征计算复杂度更高。频域特征一般是先对信号进行离散傅里叶变换得到傅里叶系数,然后基于傅里叶系数提取特征。对每个 channel 的数据进行离散傅里叶变换后,取其最大的 8 个系数作为特征,每个系数记录其频率和振幅,每个 channel 提取 16 个特征,所有 channel 共 128 个频域特征。

## 2.3 实验结果与分析

在完成特征提取后,使用支持向量机 SVM 识别上文所提的 6 种活动。所使用的 SVM 的多分类扩展使用 one-vs-one 方式,核函数使用常用的高斯核函数。实验过程中,首先在训练集上使用 10-折交叉验证,在  $C \in \{2^{-5}, 2^{-4}, \dots, 2^5\}$  和  $\gamma \in \{2^{-4}, 2^{-3}, \dots, 2^4\}$  网格中搜索最优的  $C$  和  $\gamma$  值,然后使用最优的  $C$  和  $\gamma$  在训练集上训练出分类器,最后报告交叉验证的准确率和在测试集上的准确率,见表 3。从表 3 中的结果可以看出:

1) 对于两种特征学习方式,channel-wise 方式要优于不区分 channel 方式,所以在处理惯性传感器数据时,channel-wise 方式使用特征学习技术更为合适。

2) 使用 channel-wise 方式的三种特征学习技术都明显优于常用的统计特征和频域特征,这说明该特征提取方法具有一定的优势。

3) 在使用 channel-wise 方式的三种特征学习技术中,线性方法 PCA 取得最好的结果,非线性的 SAE 的准确率略低于 PCA,这是由于用于特征学习的数据较少。在两种基于 AE 的技术中,SAE 要比 DAE 更好,说明 SAE 更适合处理传感器数据。

表 3 不同特征使用 SVM 识别准确率

Tab.3 Recognition accuracy of SVM for different features

特征	交叉验证准确/%	测试准确率/%
SAE channel-wise	95.14	91.08
SAE Mono	95.05	88.60
DAE channel-wise	95.01	88.87
DAE Mono	92.87	79.10
PCA channel-wise	95.29	91.82
PCA Mono	94.98	90.53
时域统计特征	97.55	88.56
频域特征	92.86	85.14

### 3 结论

本文提出了一种基于无监督特征学习的惯性传感器数据特征提取方法。这种方法具有以下特点:由于无监督特征学习一般使用信息损失最小化方法学习特征映射,所以可以避免重要信息损失;能够自动地从数据学习特征表示,对于领域专业知识的依赖更少;细粒度的特征提取方式速度快、规模小,对于其他问题也有一定的借鉴意义。

基于上述方法提取的特征在公开数据集上进行活动识别实验。实验结果表明,channel-wise 方式特征学习在活动识别的准确率上优于不分 channel 方式,并且也优于时域统计特征和频域特征。在所研究的三种特征学习模型中,SAE 和 PCA 大致相当并优于 DAE。由于实验所使用的数据较少,而无监督特征学习模型通常需要大量数据,在下一步工作中将会收集更多数据用于学习特征表示。

### 参考文献 (References)

[1] Incel O D, Kose M, Ersoy C. A review and taxonomy of activity recognition on mobile phones[J]. BioNanoScience,

2013, 3(2): 145 - 171.

[2] Bengio Y, Courville A C, Vincent P. Unsupervised feature learning and deep learning: a review and new perspectives[J]. CoRR, 2012;1.

[3] Plötz T, Hammerla N Y, Olivier P. Feature learning for activity recognition in ubiquitous computing[C]//Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 2011, 2: 1729 - 1734.

[4] Vollmer C, Gross H M, Eggert J P. Learning features for activity recognition with shift-invariant sparse coding[C]// Proceedings of ICANN, 2013; 367 - 374.

[5] Bhattacharya S, Nurmi P, Hammerla N, et al. Using unlabeled data in a sparse-coding framework for human activity recognition[J]. Pervasive and Mobile Computing, 2014, 15: 242 - 262.

[6] Longstaff B, Reddy S, Estrin D. Improving activity classification for health applications on mobile devices using active and semi-supervised learning[C]//Proceedings of Pervasive Computing Technologies for Healthcare (Pervasive Health), 2010 4th International Conference on No Permissions, IEEE, 2010; 1 - 7.

[7] Zhao Z T, Chen Y Q, Liu J F, et al. Cross-people mobile-phone based activity recognition[C]//Proceedings of International Joint Conference on Artificial Intelligence, 2011, 22(3): 2545 - 2550.

[8] Li Y M, Shi D X, Ding B, et al. Unsupervised feature learning for human activity recognition using smartphone sensors[C]//Proceedings of Second International Conference, MIKE, 2014; 99 - 107.

[9] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2009; 1794 - 1801.

[10] Vincent P, Larochelle H, Bengio Y S, et al. Extracting and composing robust features with denoising autoencoders[C]// Proceedings of the 25th International Conference on Machine Learning, ACM, 2008; 1096 - 1103.

[11] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. The Journal of Machine Learning Research, 2010, 11; 3371 - 3408.

[12] Bergstra J, Bengio Y. Random search for hyper-parameter optimization[J]. The Journal of Machine Learning Research, 2012, 13(1): 281 - 305.

[13] Frank A, Asuncion A. UCI machine learning repository[M/OL]. <http://archive.ics.uci.edu/ml>.

[14] Anguita D, Ghio A, Oneto L, et al. A public domain dataset for human activity recognition using smartphones[C]. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN, 2013.

[15] Android. Sensor Event Class[M/OL]. [http:// developer.android.com/reference/android/hardware/SensorEvent.html](http://developer.android.com/reference/android/hardware/SensorEvent.html).