doi:10.3772/j.issn.1002-0470.2016.04.007

基于特征融合进行活动识别的 DCNN 方法①

王金甲② 杨中玉

(燕山大学信息科学与工程学院 秦皇岛 066004)

摘 要 研究了输入是可穿戴传感器获得的多通道时间序列信号,输出是预定义的活动的活动识别模型,指出活动中的有效特征的提取目前多依赖于手工和浅层特征学习结构,不仅复杂而且会导致识别准确率下降;基于深度学习的卷积神经网络(CNN)不是对时间序列信号进行手工特征提取,而是自动学习最优特征;目前使用卷积神经网络处理有限标签数据仍存在过拟合问题。因此提出了一种基于融合特征的系统性的特征学习方法用于活动识别,用 ImageNet16 对原始数据集进行预训练,将得到的数据与原始数据进行融合,并将融合数据和对应的标签送入有监督的深度卷积神经网络(DCNN)中,训练新的系统。在该系统中,特征学习和分类是相互加强的,它不仅能处理端到端的有限数据问题,也能使学习到的特征有更强的辨别力。与其他方法相比,该方法整体精度从87.0%提高到87.4%。

关键词 融合特征, 多通道时间序列, 深度卷积神经网络(DCNN), 活动识别

0 引 言

活动识别在各个领域已经有了广泛应用,如机器人学习、健康监控、智能医院、随境游戏、智能家居等[1]。活动识别主要分为基于视觉的方法、基于无线电的方法和基于传感器的方法三种类型。基于视觉的方法利用图像和视频处理技术对相机获得的数据进行处理,进而进行活动识别。基于无线电的方法使用信号的衰减和传播特性检测活动系统的覆盖范围。基于传感器的方法,如加速度计,在活动时对时间序列采样。相比于其他方式,基于传感器的方法有三点优势:(1)不必在有限的覆盖区域内活动。(2)可以使用可穿戴传感器或者智能手机,这两种方式廉价并且可广泛应用;(3)与无线电方法不同,不用担心因发送信号对人体健康产生影响。这些优点使得基于传感器的活动识别算法发展迅速,影响广泛。

采用可穿戴传感器的活动识别依赖于传感器的组合,如加速度计、重力传感器、磁力传感器。在国外,可穿戴设备的活动识别研究已有初步成果。Roggen 和 Ordonez 使用滑动窗对原始数据进行处理后,分别用模板匹配方法[2]和隐马尔科夫模型[3]进行分类。Cao 使用简单的预分类策略[4],即通过过采样方法校正类的不均衡,然后利用数据间的顺序性对预测的标签序列进行平滑处理来提高其性能,他将所提的方法与支持向量机(support vector machine,SVM)和 k 近邻分类器(k-nearest neighbor,KNN)分类进行对比,表明了其优越性。此外,Bulling 使用了均值和协方差(means and variance,MV)[5],Platz 使用了深度置信网络(deep belief network,DBN)[6], Yang 使用了深度卷积神经网络(deep convolutional neural network, DCNN)[7]。在国内,一般用传统方法进行可穿戴设备的活动识别。如吴渊使用绝对值和简单移动平均线处理的方法[8],刘斌选择四种典型的统计学习方法(分别是k-近邻算法、支持向量机、朴素贝叶斯网络以及基于

① 国家自然科学基金(61273019,61473339),河北省自然科学基金(F2013203368),河北省青年拔尖人才支持项目([2013]17),河北省博士后专项资助(B2014010005)和中国博士后科学基金(2014M561202)资助项目。

② 男,1978 年生,博士,教授;研究方向:信号处理,模式识别及其应用;联系人,E-mail: wjj@ysu.edu.cn

(收稿日期:2016-01-07)

朴素贝叶斯网络的 AdaBoost 算法)分别创建活动识别模型,最后通过模型决策得到最优的活动识别模型[9]。在多通道时间序列信号中手工提取特征通常会忽略不同信号之间的相关性,而深度卷积神经网络(DCNN)方法可以弥补这个不足,但是使用 DCNN 方法处理有限标签数据会出现过拟合问题。受文献[10]的启发,本文提出了一种基于融合特征的系统性的特征学习方法用于活动识别,该方法采用滑动窗策略将信号转换成新的活动图像,用 ImageNet16 对原始数据集进行预训练,将得到的数据与原始数据进行融合,并将融合数据和对应的标签送入有监督的深度卷积神经网络中,训练新的系统。该系统可以自动学习最优的特征,使学习到的特征有更强的辨别力。与独立时间序列信号或者统计学特征相比,融合特征可以取得更好的分类结果。

1 预训练模型

1.1 ImageNet 数据集

ImageNet 数据集是与视觉相关的分类任务,它包含约 1500 万张带标记的高分辨率图像,近 22000 类。数据集中的图像是通过搜索引擎检索到的,是常见的多媒体数据。每年都举行 ImageNet 分类比赛,即"ImageNet 大规模视觉识别竞赛",与会者选择这个数据集的子集训练分类算法。2014 年,Simonyan 和 Zisserman 训练得到 imageNet16 模型[11]。

1.2 无监督的预训练

文献[12]表明在图像多分类任务中,封装表示通常比标准的分类方法好,这和深层网络表示相通,在有限标记数据情况下可以进行迁移学习。对数据进行预训练时,取 ImageNet16 模型的第 36 层作为输出,将得到的数据与原始数据进行融合,再将该数据和对应的标签送入有监督的 DCNN 分类器,训练新的系统。这样做的原因如下:第一,深层可以得到丰富的信息。在预训练模型中,浅层对实验结果影响较小,而深层对分类准确率的影响较大。因此,利用 ImageNet16 模型中的深层网络得到预训练数据而丢弃浅层的特征。第二,融合特征可得到更高的准确率,如由文献[13]提出从卷积神经网络的不同层提取组合信息,网络中的不同尺度信息可以共存,称为多尺度特征提取。深度学习中这种方法很常见,它可以跳过层之间的连接,将严格的时序网络转换成一个有向无环图并对分类结果产生积极的影响。文献[14]中有另外一种策略,用卷积神经网络(CNN)分别对 RGB 图像和深度图像进行深度特征提取,转换成单一向量后,将它送入最终的分类器。

2 深度卷积神经网络结构

本文研究方法流程图如图 1 所示。基于融合特征的系统性的特征学习方法用 ImageNet16 对原始数据集进行预训练,将得到的数据与原始数据进行融合,再将该数据和对应的标签送入有监督的深度卷积神经网络训练新的系统。最后用该模型对测试样本进行分类。

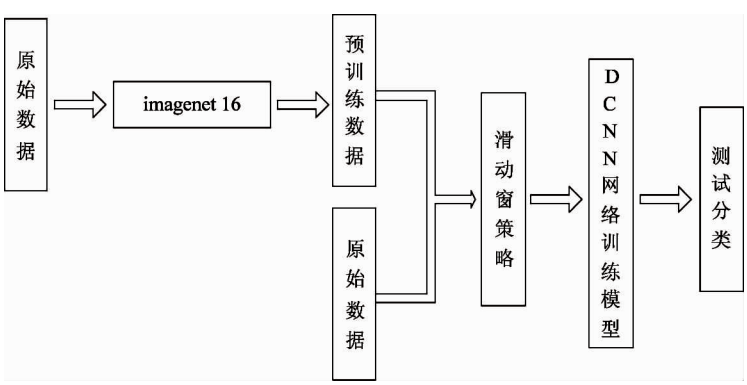


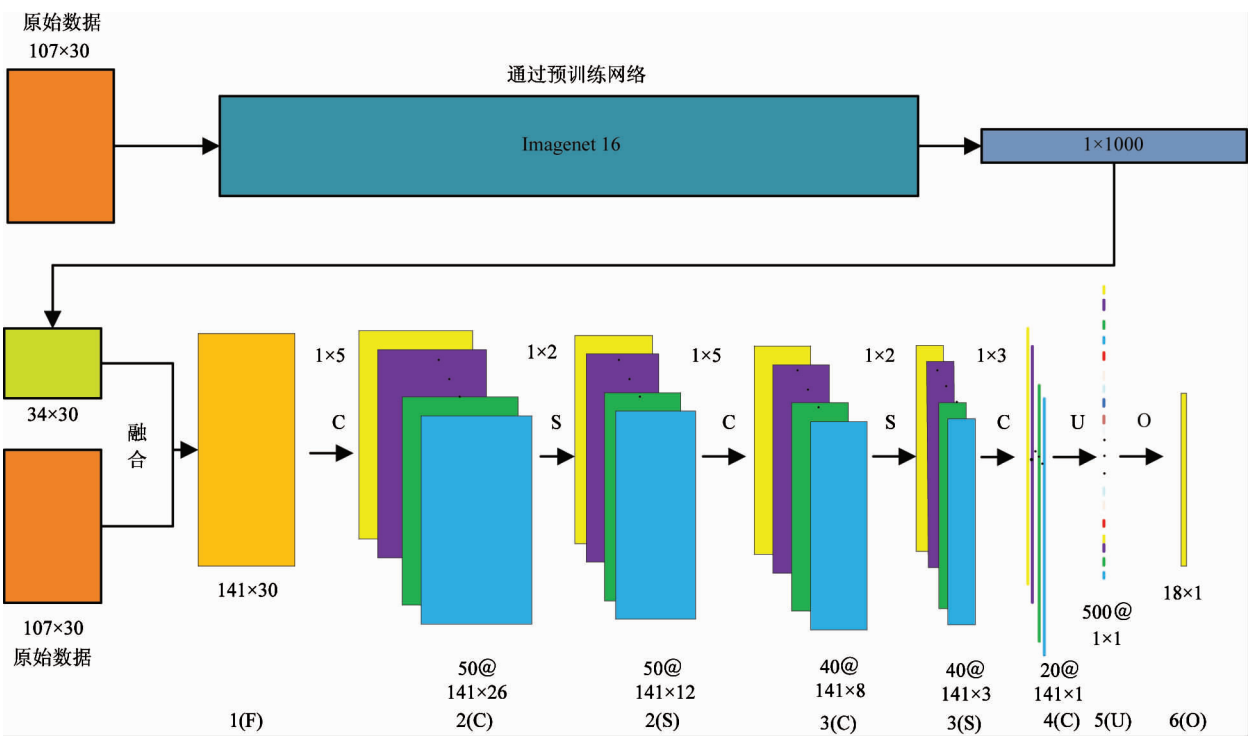
图 1 本文方法整体流程图

实验采用滑动窗策略，将时间序列信号分割成一系列短信号的集合。DCNN 结构中的输入样本是一个二维矩阵，每个输入样本包含 r 个原始样本，滑动窗口的步长为 3，每个样本有 D 个属性，在这里 r 为采样率（传感器信号的采样速率为 30Hz，因而试验中使用 $r=30$）。选择较小的步长，可以增加实例的数量，但是可能产生较高的计算成本。对于训练数据，样本的真实标签是由 r 个原始记录数据中出现次数最多的标签决定。DCNN 结构中，第 i 层的第 j 个特征图中传感器 d 第 x 行的值记作 $v_{ij}^{x,d}$。在卷积层中，上一层的特征图与卷积核进行卷积，加上偏差，然后通过激活函数，得到输出的特征图，如下式所示：

$$v_{ij}^{x,d}=\tanh\left(b_{ij}+\sum_{m}\sum_{p=0}^{P_i-1}w_{ijm}^{p}v_{(i-1)m}^{x+p,d}\right)\ \forall d=1,\cdots,D \tag{1}$$

式中，$\tanh(\cdot)$ 为双曲正切激活函数，b_{ij} 为当前特征图的偏差，m 为与当前特征图相连的第 $i-1$ 层特征图的集合，P_i 为卷积核的长度，w_{ijm}^{p} 为在位置 p 处卷积核的值，D 为 141。

我们所提出的融合的 DCNN 结构共有 6 层，如图 2 所示。这里，将第 1 层称为融合层，在深度卷积神经网络中，第 2 层的卷积层使用 1×5 大小的 50 个核来过滤 141×30 的图像，池化的大小为 1×2。第 3 层的卷积层使用第 2 层的输出作为输入，使用 40 个大小为 1×5 的核进行过滤，池化的大小为 1×2。第 4 层的卷积使用第 3 层的输出作为输入，使用 20 个大小为 1×3 的核进行过滤。第 5 层使用第 4 层的输出作为输入，使用 500 个大小为 141×1 的核进行过滤，称为合并层。目的是将上一层输出的所有特征图连接成一维向量。如图 3 所示，我们将所有特征图在这一层合并，第 5 层中每个点都是由所有特征图中的部分数据得到，从而实现了参数级联，而不是简单地将它们连接。本质上，本层进行的是卷积操作，只是得到的结果是由 500 个 1×1 的特征图级联而成的数据，第 j 个特征图的值通过 $v_{ij}=\tanh\left(b_{ij}+\sum_{m}\sum_{d=1}^{D}w_{ijm}^{d}v_{(i-1)m}^{d}\right)$ 计算得到。第 6 层为全连接层，它将第 5 层的输出送入 softmax 分类器，输出为 $v_{ij}=\dfrac{\exp(v_{(i-1)j})}{\sum_{j=1}^{C}\exp(v_{(i-1)j})}$，$C$ 为输出类的数目，它提供了分类结果的后验概率。



“@”符号前的数字表示该层特征图的数量，“@”后的数字表示该层一个特征图的维数；

图中的“F”、“C”、“S”、“U”、“O”分别表示融合操作、卷积操作、池化操作、合并操作和输出操作。

图 2　融合特征的深度卷积神经网络结构

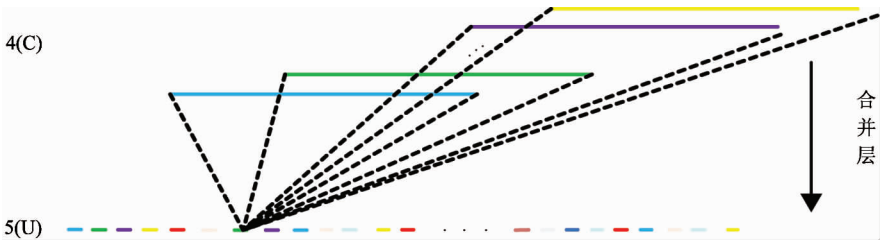


图 3　合并层示例

为了防止过拟合,得到更准确的实验结果,每次都将卷积操作的结果送入整流线性单元,即上一层的输出通过函数 $relu(v) = \max(v,0)$,每次进行池化后都对数据进行归一化。第 4 层进行卷积操作后,也对数据进行归一化。论文中的池化层采用最大池化,在一个局部时空邻域的范围内寻找最大特征图(通常涉及池化操作)。为了图文简洁,图 2 和图 3 中没有画出整流线性单元和归一化层。

3　实验结果

3.1　机会数据集

本文在机会数据集(opportunity activity recognition dataset)上进行实验,采集环境如图 4 所示。该数据集的活动涉及全身动作,与早餐情景有关,共 18 类,即空类、开门 1、开门 2、关门 1、关门 2、开冰箱、关冰箱、开洗碗机、关洗碗机、开抽屉 1、关抽屉 1、开抽屉 2、关抽屉 2、开抽屉 3、关抽屉 3、擦桌子、喝水、切换开关。训练集样本个数为 136869,测试集样本个数为 32466。空类指非相关活动或非活动。取第一名受试者数据,前两组活动和演练数据作为训练数据,第三组作为测试数据。

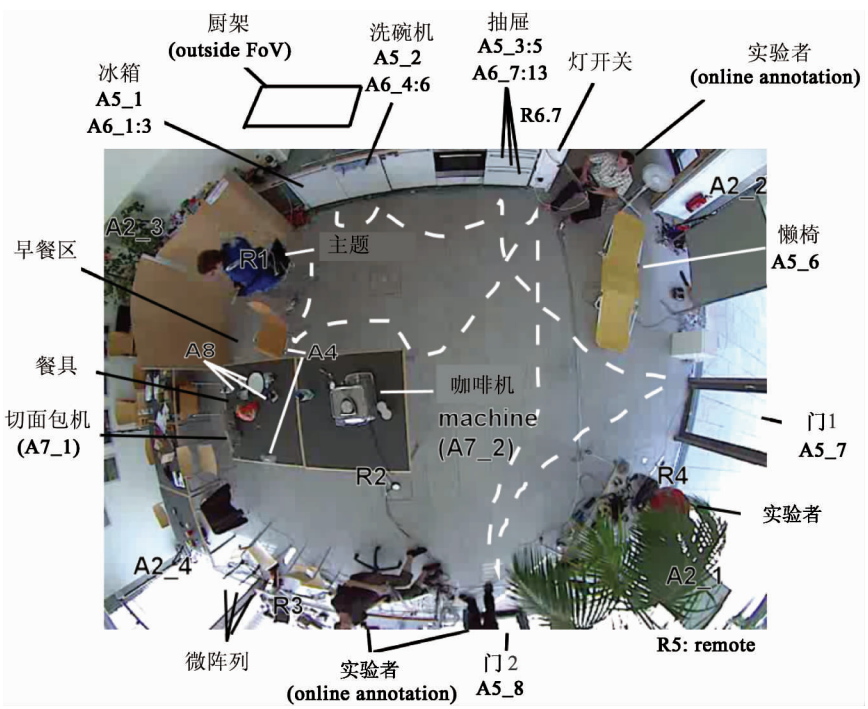


图 4　机会数据集采集环境俯视图

3.2　实验设置

在机会数据集上进行实验时,CNN 的一般操作中参数依照文献[15]进行选择,如何选择最优的参数是一个开放性的问题。我们将所提方法与传统支持向量机、k 近邻(KNN)分类、深度置信网络、深度卷积神经网络、均值和协方差比较。支持向量机

(SVM)和卷积神经网络(KNN)在该数据集上得到的准确率较好,深度置信网络(deep belief network,DBN)和 CNN 是新的可用于活动识别的深度学习算法。

SVM:基于径向基函数(RBF)核的支持向量机作为分类器,支持向量机的输入是原始时间序列。交叉验证程序用于调整支持向量机的参数。

KNN:对时间序列分类问题进行了全面的评价。在基于欧氏距离的简单 KNN 算法中,1NN 分类效果最好,因此我们将 1NN 作为分类器。同 SVM 一样,1NN 算法的输入是原始时间序列。

MV:均值和方差。与所提出的 DCNN 方法相似,首先采用滑动窗策略生成新的样本。然后,提取每个样本的平均值和方差,构成分类器的输入。采用的分类器是 $k=1$ 的 KNN。

DBN:类似于方法 DCNN 和 MV,首先采用滑动窗策略生成新的样本。然后,提取每一个样本的平均值作为 DBN 的输入。在该方法中使用的分类器是 $k=1$ 的 KNN 或多层感知器神经网络。

3.3　实验结果

本文所提方法和其他方法在机会数据集上进行实验,得到的均值 F-测度(AF)、归一化 F-测度(NF)和准确率(AC)如表 1 所示。基于融合特征的深度卷积神经网络结构得到的实验结果与 SVM、KNN、MV、DBN 相比,得到的实验结果更好。与 CNN 方法相比,结果相当,得到的准确率略有提高。实验证明将深度卷积神经网络结构用于人类活动识别行之有效,基于融合的方法可以让网络自动提取更有识别力的特征,提高识别率。

表 1　不同方法的实验结果

方法	AF(%)	NF(%)	AC(%)
SVM	45.6	83.4	83.8
KNN	42.7	80.3	79.3
MV	54.2	83.9	83.7
DBN	14.3	75.0	80.0
CNN	55.5	86.4	87.0
本文方法	54.5	86.4	87.4

本文所提方法在机会数据集上实验,得到的实验结果的混淆矩阵如图 5 所示。从结果可以看出,空闲、开门 2、关门 2、切换开关的识别率较高。实验者做无关动作或空闲会对实验结果产生很大影响,还不能准确区分空类和特定的类,这会导致一些特定的类可能错分为空类。此问题至今还没有特别行之有效的方法,有待进一步研究。

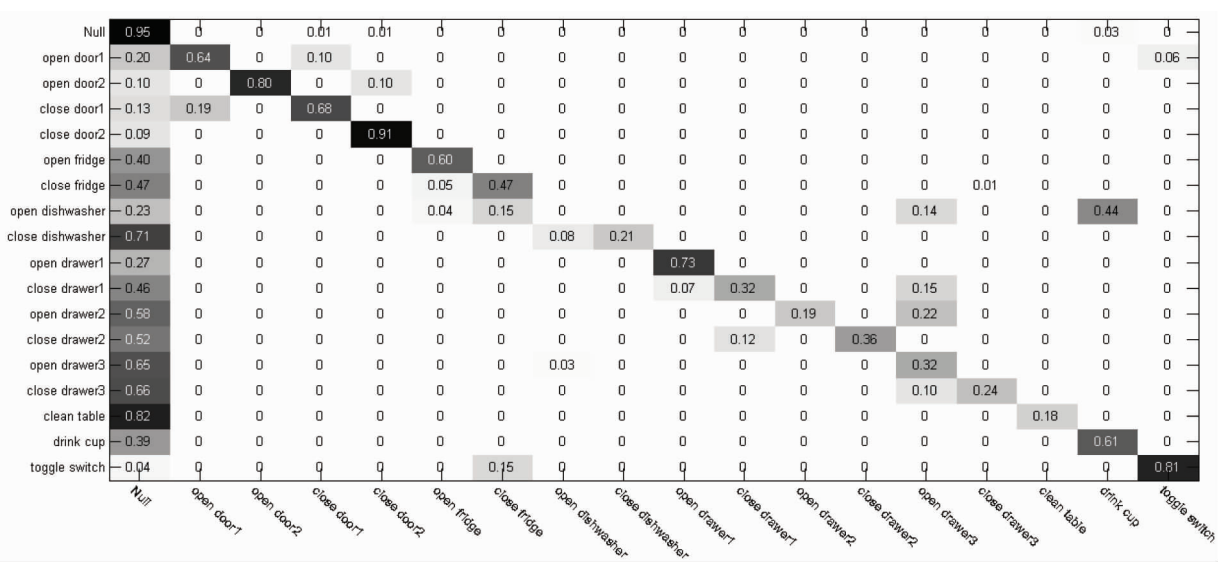


图 5　机会数据集的混淆矩阵

实验是在一台普通的装有 Matlab 2014 的电脑上运行,电脑的 CPU 是 3.1GHZ,内存是 4GB。用 imageNet16 模型预训练单个样本所用的时间为 7.0s,训练模型所用的时间约为 2.0h,测试所有样

本所用的时间为 5.7s。文中提到的其他算法所用时间在其文献中没有给出,所以没有进行实时性对比。

4 结 论

本文提出了一种基于融合特征的系统性的特征学习方法,它可以自动进行特征提取,进行活动识别,将预训练得到的数据与原始数据融合,并建立新的 DCNN 深度架构来研究多通道时间序列数据。这种架构主要采用卷积和池化操作捕捉传感器信号在不同时间尺度的显著模式。系统将所有识别的显著模式在多个通道进行统一,最终映射到不同的活动。所提出的方法的主要优点是:(1)用非手工方式提取特征,自动选择最优的特征;(2)提取的特征更具识别力;(3)对过拟合问题有所改善;(4)特征提取和分类都统一在一个模型,其性能是相互增强的。实验表明,基于融合特征的深度卷积神经网络方法优于其他方法,该方法在人类活动识别问题中可以有效地进行特征学习和分类。

参考文献

[1] Alsheikh M A, Selim A, Niyato D, et al. Deep activity recognition models with triaxial accelerometers. *Computer Science*, 2015, arxiv: 1511.04664

[2] Roggen D, Cuspinera L P, Pombo G, et al. Limited-memory warping LCSS for real-time low-power pattern recognition in wireless nodes. In: Proceedings of the 12th European Conference Wireless Sensor Networks (EWSN), Porto, Portugal, 2015. 151-167

[3] Ordonez F J, Englebienne G, De Toledo P, et al. In-home activity recognition: Bayesian inference for hidden Markov models. *IEEE Pervasive Computing*, 2014, 13(3):67-75

[4] Cao H, Nguyen M N, Phua C, et al. An integrated framework for human activity classification. *Ubicomp*, 2012:331-340

[5] Bulling A, Blanke U, Schiele B. A tutorial on human activity recognition using body-worn inertial sensors. *Acm Computing Surveys*, 2014, 46(3):57-76

[6] Plötz T, Hammerla N Y, Olivier P. Feature learning for activity recognition in ubiquitous computing. In: Proceedings of the International Joint Conference on Artificial Intelligence, Barcelona, Spain, 2011. 1729-1734

[7] Yang J B, Nguyen M N, San P P, et al. Deep convolutional neural networks on multichannel time series for human activity recognition. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina, 2015. 3995-4001

[8] 吴渊, 史殿习, 杨若松等. 手机位置和朝向无关的活动识别技术研究. 计算机技术与发展, 2016(4):

[9] 刘斌, 刘宏建, 金笑天等. 基于智能手机传感器的人体活动识别. 计算机工程与应用, 2016, 52(4):188-193

[10] Marmanis D, Datcu M, Esch T, et al. Deep learning earth observation classification using imageNet pretrained networks. *IEEE Geoscience & Remote Sensing Letters*, 2016, 13(1):105-109

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arxiv preprint*, arxiv: 1409.1556, 2014

[12] Donahue J, Jia Y, Vinyals O, et al. DeCAF: A deep convolutional activation feature for generic visual recognition. *Computer Science*, 2013, 50(1): 815-830

[13] Sermanet P, Lecun Y. Traffic sign recognition with multi-scale. *Convolutional Networks*, 2011, 42(4):2809-2813

[14] Socher R, Huval B, Bhat B, et al. Convolutional-recursive deep learning for 3D object classification. *Advances in Neural Information Processing System*, 2012: 665-673

[15] Lecun Y, Bottou L, Orr G B, et al. Efficient BackProp. Neural Networks: Tricks of the Trade. Berlin Heidelberg: Springer, 1998. 9-50

A DCNN method for human activity recognition based on feature fusion

Wang Jinjia, Yang Zhongyu

(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract

An activity recognition model, with its input being the multi-channel time series signals obtained by wearable sensors and output being a predefined activity, was studied. It was pointed that extracting effective features from activity is a key in activity recognition. Most of the existing work relies on manual extraction of the features and the shallow learning structure, which makes the work complex and the recognition unaccurate. However, the convolutional neural network (CNN) based on deep learning does not manually extract the time series signals, but automatically learns the best feature. At present, using convolutional neural network to process limited labeled data still has the overfitting problem. Therefore, a systematic feature learning method based on fusion characteristics was presented for activity recognition. The method uses the ImageNet16 to pre-train the original data set to fuse the obtained data with the original data, and puts the fused data and the corresponding tag into a supervised depth convolutional neural network (DCNN) to train the new system. In this system, the characteristics of learning and classification are mutually reinforcing, which can not only deal with the problem of limited data from end to end, but also make the learning more discriminative. Compared with other methods, the overall accuracy of the proposed method is increased from 87% to 87.4%.

Key words: fusion feature, multichannel time sequence, deep convolutional neural network (DCNN), activity recognition