

doi:10.6041/j.issn.1000-1298.2018.03.022

基于深度信念网络的猪咳嗽声识别

黎 焯^{1,2} 赵 建^{1,2} 高 云^{1,2} 雷明刚^{2,3} 刘望宏^{2,3} 龚永杰^{1,2}

(1. 华中农业大学工学院, 武汉 430070; 2. 生猪健康养殖协同创新中心, 武汉 430070;

3. 华中农业大学动物科技学院动物医学院, 武汉 430070)

摘要: 为了在生猪养殖产生呼吸道疾病的初期,通过监测猪咳嗽声进行疾病预警,提出了基于深度信念网络(DBN)对猪咳嗽声进行识别的方法。以长白猪咳嗽、打喷嚏、吃食、尖叫、哼哼、甩耳朵等声音为研究对象,利用基于多窗谱的心理声学语音增强算法和单参数双门限端点检测对猪声音进行预处理,实现猪声音信号的去噪和有效信号检测。基于时间规整算法提取300维短时能量和720维梅尔频率倒谱系数(MFCC)组合成1020维特征参数,将该组合特征参数作为DBN学习和识别数据集,选定3隐层神经元个数分别为42、17和7,构建网络结构为1020-42-17-7-2的5层深度信念网络猪咳嗽声识别模型。通过5折交叉实验验证,基于DBN的猪咳嗽声识别率和总识别率均达到90%以上,误识别率不超过8.07%,最优组猪咳嗽声识别率达到94.12%,误识别率为7.45%,总识别率达到93.21%。进一步基于主成分分析法(PCA)提取1020维特征参数98.01%主成分得到479维特征参数,通过5折交叉实验验证,猪咳嗽声识别率和总识别率相对降维前均有所提高,误识别率有所降低,最优组猪咳嗽声识别率达到95.80%,误识别率为6.83%,总识别率达到94.29%,实验结果表明所建模型是有效可行的。

关键词: 生猪; 咳嗽; 深度信念网络; 特征参数; 识别

中图分类号: TN912.34 **文献标识码:** A **文章编号:** 1000-1298(2018)03-0179-08

Recognition of Pig Cough Sound Based on Deep Belief Nets

LI Xuan^{1,2} ZHAO Jian^{1,2} GAO Yun^{1,2} LEI Minggang^{2,3} LIU Wanghong^{2,3} GONG Yongjie^{1,2}

(1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. The Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China

3. College of Animal Science and Technology, College of Animal Medicine, Huazhong Agricultural University, Wuhan 430070, China)

Abstract: In the early stage, pig cough sound could be detected for early disease warning, and a method based on deep belief nets (DBN) was proposed to construct a pig cough sound recognition model. Pig sounds of Landrace pigs, including cough, sneeze, eating, scream, hum and shaking ears sounds were automatically recorded. The samples were preprocessed by speech enhancement algorithm based on a psychoacoustical model and speech endpoint detection algorithm based on short-time energy to reduce the noise and get useful parts of samples. Based on the dynamic time warping (DTW) algorithm, the short-time energy characteristics were scaled to a 300-dimensional short-time energy feature vector, while the 24-dimensional MFCC characteristics were scaled to a 720-dimensional MFCC feature vector. And then the 300-dimensional short-time energy feature vector and the 720-dimensional MFCC feature vector were combined to construct a 1020-dimensional vector as the input of the deep belief nets. The number of neuron of the three hidden layers were set to be 42, 17 and 7, respectively, so the pig sound recognition model based on DBN was finally designed to be 1020-42-17-7-2. The 5-fold cross validation experiment showed that recognition rate, error recognition rate and total recognition rate of the best experimental group were 94.12%, 7.45% and 93.21%, respectively. Furthermore, the first 479 principal components of 1020 dimension feature parameters were obtained by PCA dimensionality reduction. The recognition rate, error recognition rate and total recognition rate obtained better performance, and the best experimental group reached 95.80%, 6.83% and 94.29%, respectively. The result demonstrated that the DBN model was effective for the pig cough recognition.

Key words: pig; cough; deep belief nets; feature parameters; recognition

收稿日期: 2017-07-21 修回日期: 2017-09-19

基金项目: 华中农业大学大北农青年学者提升专项项目(2017DBN005)、现代农业产业技术体系项目(CARS-36)、国家重点研发计划项目(2016YFD0500506)和中央高校基本科研业务费专项资金项目(2015PY079)

作者简介: 黎焯(1980—),男,副教授,博士,主要从事生猪信息智能感知与行为识别研究,E-mail: lx@mail.hzau.edu.cn

0 引言

近年来,国家着力转变生猪产业经济发展方式,引导生猪产业向规模化、集约化、标准化饲养方向发展。生猪产业规模化发展的同时,呼吸道疾病已成为各大养猪场最常见、危害最严重的疾病之一。而咳嗽是猪呼吸道疾病的主要症状,尤其是早期症状,因此可以通过监测咳嗽声进行猪早期呼吸道疾病预警^[1-3]。目前所采用的方法一般为简单易行的人工检测,但是人工检测不仅人力成本高,并且识别率难以得到保证。随着现代信息、数字信号处理、传感器等技术的快速发展,将计算机技术与声音特征分析相结合,利用语音识别技术分析猪咳嗽声与猪非咳嗽声,对猪咳嗽声进行识别,有利于生猪呼吸道疾病的预警,促进生猪健康养殖的发展^[4]。

国外开展猪声音识别的研究相对较早。MITCHELL等^[2]发现了健康猪和病猪咳嗽声短时能量动态变化的差异,测得健康猪咳嗽声持续时间一般为0.21 s,而病猪咳嗽声持续时间一般为0.3 s。SARA等^[3]通过实验发现,病猪咳嗽音频的标准化压力均方差及峰值频率均值均低于健康猪,而病猪咳嗽声持续时间和频率均高于健康猪。EXADAKTYLOS等^[5]采用改进的模糊c均值聚类算法识别猪咳嗽,总识别率达到85%,其中病猪咳嗽声识别率达到82%。GUARINO等^[6]采用动态时间规整(Dynamic time warping, DTW)算法识别猪咳嗽,识别率达到85.5%,非咳嗽声识别率达到86.6%。HIRTUM等^[7]考虑低频环境噪声对咳嗽声的影响,建立环境噪声模型,在此基础上构建基于模糊c均值聚类算法的咳嗽声识别模型,非实时单个咳嗽声识别率能够达到92%,但错误率达到21%。

目前,国内针对猪声音识别的研究尚处于起步阶段。马辉栋等^[8]设计了基于短时能量和短时过零率的猪咳嗽声双门限端点检测算法。刘振宇等^[9]采用隐马尔科夫模型(Hidden markov model, HMM)对猪咳嗽声进行识别,识别率达到80%。徐亚妮等^[10]利用模糊c均值聚类算法进行猪咳嗽声与尖叫声识别,识别率分别达到83.4%和83.1%。

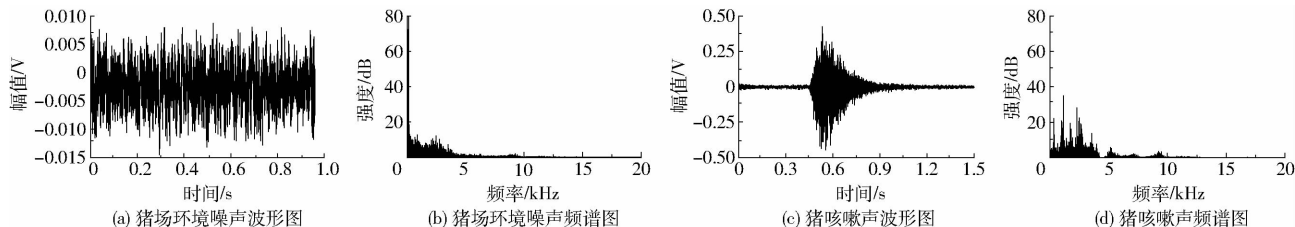


图1 猪场环境噪声分析

Fig. 1 Analysis of environmental noise in pig farms

深度信念网络(Deep belief nets, DBN)由HINTON等^[11]于2006年提出,DBN通过对人脑组织结构和功能的模拟,有着与人类类似的记忆能力、概括推理能力以及强大的分类、预测能力^[12]。最近几年来,深度学习在语音识别领域得到了很好的应用,体现了其强大的声学建模能力^[13-15]。本文把深度信念网络引入猪声音识别领域,以长白猪咳嗽、打喷嚏、吃食、尖叫、哼哼、甩耳朵等声音为研究对象,构建猪咳嗽声识别模型,为生猪健康养殖过程中猪咳嗽声的识别提供一种新方法。

1 猪声音采集与预处理

1.1 猪声音采集

猪声音采集在校属精品猪场进行。采集工具为美博-M66录音笔,采样频率为48 000 Hz,采样精度为16位,可连续录音24 h。声音采集在3—4月猪病多发期进行,共采集10头75 kg左右长白猪的声音,经兽医诊断10头猪中5头猪感染呼吸道疾病,咳嗽明显。采用专家分类法对录音笔采集的猪声音进行分类标记,选取咳嗽、打喷嚏、吃食、尖叫、哼哼和甩耳朵声作为研究对象。得到猪声音样本1 400个,其中咳嗽样本594个,打喷嚏样本241个,吃食样本152个,尖叫样本130个,哼哼样本125个,甩耳朵样本158个,保存为wav格式。

1.2 猪声音预处理

猪场环境下采集到的猪声音样本包含很多噪声和无效声音,为提高猪咳嗽声与非咳嗽声识别率,在特征参数提取之前需要进行去噪和端点检测。

1.2.1 猪声音样本噪声分析

从图1b和图1d可以看出,猪场环境噪声频段主要集中在5 kHz以下,并且与猪咳嗽声频段(0.3 ~ 8 kHz)^[8]有重叠,传统数字滤波器(低通、高通或带通)难以对猪声音样本进行有效去噪,本文采用基于多窗谱的心理声学语音增强算法对猪声音样本实现去噪。

1.2.2 基于多窗谱的心理声学语音增强

多窗谱即对待估计猪声音样本序列加多个相互正交的窗,分别计算频谱,然后求平均的非参数谱估

计方法。人耳掩蔽阈值的心理声学约束可以对带噪猪声音信号中噪声失真进行限制^[16-17]。

通过原始猪声音样本的多窗谱估计原始猪声音样本中噪声与含噪信号之比(即先验信噪比倒数),用基于噪声与含噪信号之比的幅度谱减法得到预估计猪声音信号,通过预估计猪声音信号计

算隐蔽阈值,用心理声学算法得到去噪后的猪声音信号。

图 2 所示为基于多窗谱的心理声学语音增强算法处理前后猪声音样本波形图,对比语音增强前后波形图可知,语音增强后猪声音样本噪声明显减少,声音信号波形几乎没有发生失真。

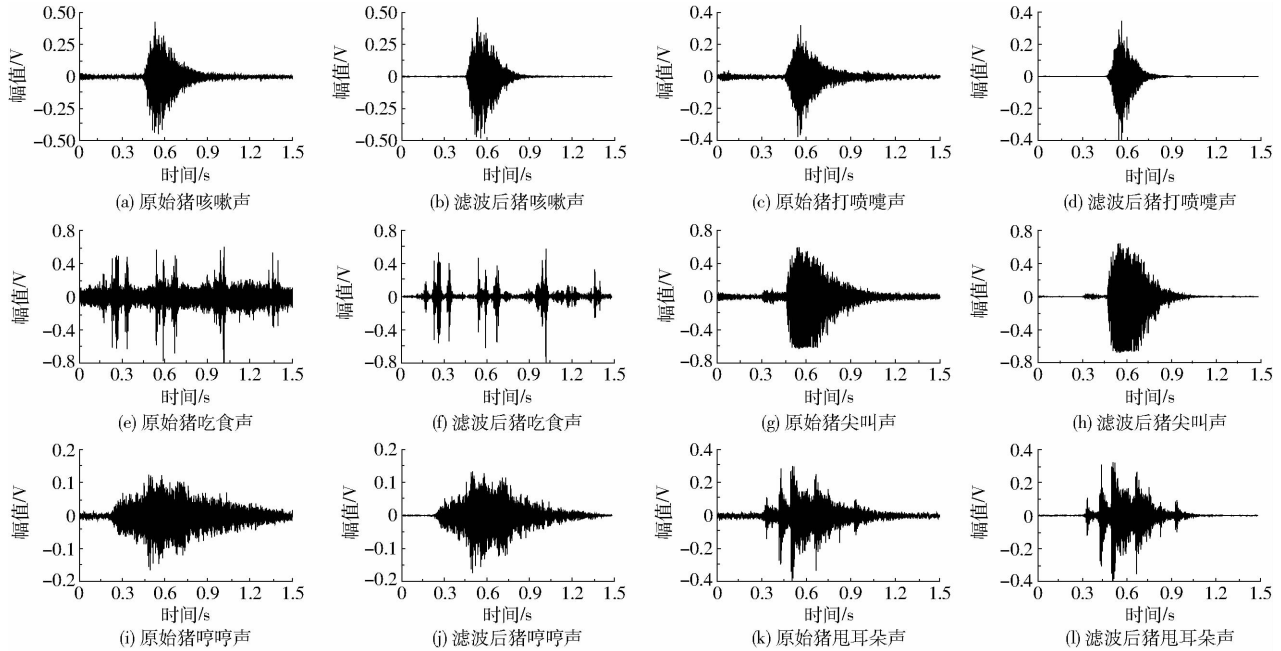


图 2 猪声音样本滤波前后对比

Fig. 2 Comparison of original pig sounds and sounds after denoising

1.2.3 单参数双门限端点检测

语音信号端点检测是指从包含语音的一段信号中找出语音的起止点,把起止点之间的信号定义为有效信号。猪声音样本经过语音增强后,噪声显著减少,本文选择基于短时能量的方法进行猪声音样本端点检测^[18]。对猪声音样本 $y(n)$,分帧后第 v 帧表示为 $y_v(n)$,此帧猪声音信号的短时能量 E_v 计算公式为

$$E_v = \sum_{n=1}^L y_v^2(n) \quad (1)$$

式中 L ——帧长,根据声音信号的短时平稳特性取为 200 个采样点

n ——猪声音样本采样点序号

基于短时能量的单参数双门限端点检测中,单参数是幅值归一化后的短时能量 e_v ,计算公式为

$$e_v = \frac{E_v}{V} \quad (2)$$

式中 V ——猪声音样本总帧数

设定 2 个阈值 T_1 和 T_2 ,其公式为

$$T_1 = 1.5 \max(e_1, e_2, \dots, e_{F_{INS}}) \quad (3)$$

$$T_2 = 1.1 \max(e_1, e_2, \dots, e_{F_{INS}}) \quad (4)$$

式中 F_{INS} ——猪声音样本前导无话段帧长

$e_{F_{INS}}$ ——第 F_{INS} 帧幅值归一化后的能量

由式(3)和式(4)计算得到 T_1 、 T_2 分别约为 0.02、0.008。当 e_v 高于 T_1 时判定为语音帧,低于或高于 T_2 时确定为猪声音样本起止点。图 3 所示为单参数双门限端点检测对应的猪声音样本检测结果。从图 3 可知,基于短时能量的端点检测可以较好地检测出有效信号。

2 基于时间规整算法的猪声音特征参数提取

2.1 时间规整算法

由图 3 可知,经过端点检测后的猪声音信号长度不同,而一个结构确定的神经网络输入层神经元个数也是确定的,因此需要运用时间规整算法^[19]将预处理后的猪声音样本规整为同一长度再进行特征参数提取。

时间规整网络结构如图 4 所示,对于一个有 V 帧的猪声音样本,规整网络输入层有 V 个节点,第 v 个节点对应猪声音样本第 v 帧的特征矢量 $A_{0,v}$ ($v = 1, 2, \dots, V$)。规整至第 l 层后将距离最近的 2 个特征矢量以加权平均合并,其余特征矢量不变,第 l 层就具有 $V - l$ 个节点以及与之联系的 $V - l$ 个矢量

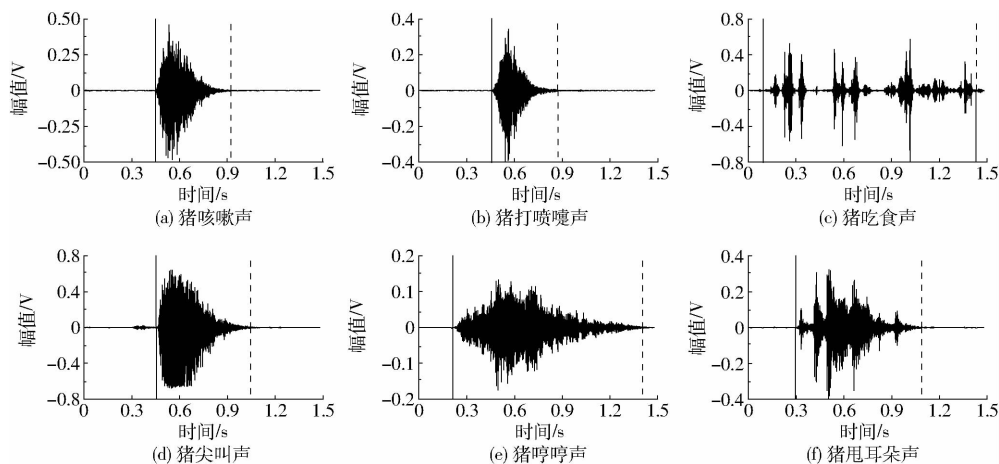


图3 猪声音样本端点检测结果

Fig. 3 End point detection of pig sound

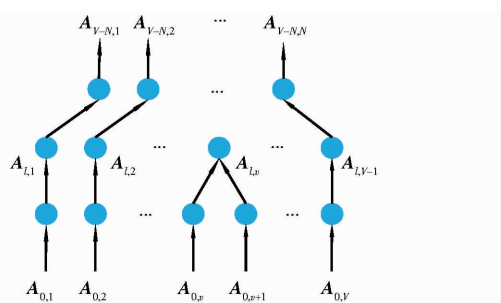


图4 时间规整网络结构图

Fig. 4 Structure diagram of time warping net

$A_{l,v} (v=1, 2, \dots, V-l)$ 。以此类推, 经过 $V-N$ 步合并, 最终网络输出层具有 N 个节点以及与之联系的 N 个特征矢量 $A_{v-N,v} (v=1, 2, \dots, N)$ 。

2.2 短时能量特征参数提取

由图2可知, 不同种类猪声音的波形不同, 为了反映这一区别, 利用式(1)计算出猪声音信号的短时能量, 6种猪声音信号的短时能量变化情况如图5所示。不同种类猪声音持续时间不同, 同种猪声音持续时间也不完全相等。本文研究的6种猪声音中咳嗽、打喷嚏持续时间一般在0.4 s左右, 而吃食、尖叫、哼哼和甩耳朵持续时间在0.6~1.4 s不等。为了进行统一衡量, 可采用时间规整算法将所有猪声音样本规整到300帧组成的一个300维的短时能量特征参数数组。

V 帧猪声音信号对应采样点数 N_v 的计算公式为

$$N_v = (V-1)L_{inc} + L \quad (5)$$

式中 L_{inc} ——帧移, 取为帧长 L 的40%

N_v 采样点对应时长 Δt 的计算公式为

$$\Delta t = N_v / F_s \quad (6)$$

式中 F_s ——采样频率

由式(5)可以计算出300帧对应采样点数 N_v 为24120, 进一步由式(6)得到对应时间长度 Δt 为

0.5025 s。时间规整为300帧(即约0.5 s)后6种猪声音样本短时能量分布如图5所示。

对比时间规整前后猪声音短时能量变化图可知, 猪咳嗽声和打喷嚏声短时能量波形图在0.4~0.5 s部分相对其他种类猪声音能量低, 造成特征数组后面部分值相对较小。基于时间规整的短时能量特征参数可以同时反映不同种类猪声音信号短时能量动态变化的差异和不同种类猪声音信号持续时间的差异性。

2.3 梅尔频率倒谱系数特征参数提取

梅尔频率倒谱系数(Mel frequency cepstral coefficients, MFCC)的分析是基于人耳的听觉机理, 依据人的听觉实验结果来分析声音的频谱特性^[20-21]。将声音在频域上划分成若干频率群, 选择梅尔频率滤波器组为24组。图6a所示为猪咳嗽声样本MFCC三维图, 对于一个300帧的猪咳嗽样本, 24维的MFCC数据量是比较大的, 可以采用时间规整算法将300帧的MFCC规整到30帧组成720维的MFCC特征向量。图6b所示为时间规整后猪咳嗽声MFCC三维图。

由猪咳嗽样本MFCC时间规整前后三维图可知, 图6b相对图6a帧数从300帧减少到30帧, 数据量大大减少。同时, 图6b也保留了图6a在时序上的动态变化特性。

3 猪咳嗽声识别

3.1 深度信念网络猪咳嗽声识别模型

深度信念网络是由多层受限玻尔兹曼机(Restricted boltzmann machine, RBM)堆叠而成的网络模型^[11]。RBM是一个由两层神经元组成的层间全连接、层内无连接的网络结构, 显层为输入层, 隐层为特征提取层。其结构如图7所示。

RBM中神经元通常只取0或1两种状态, 状态

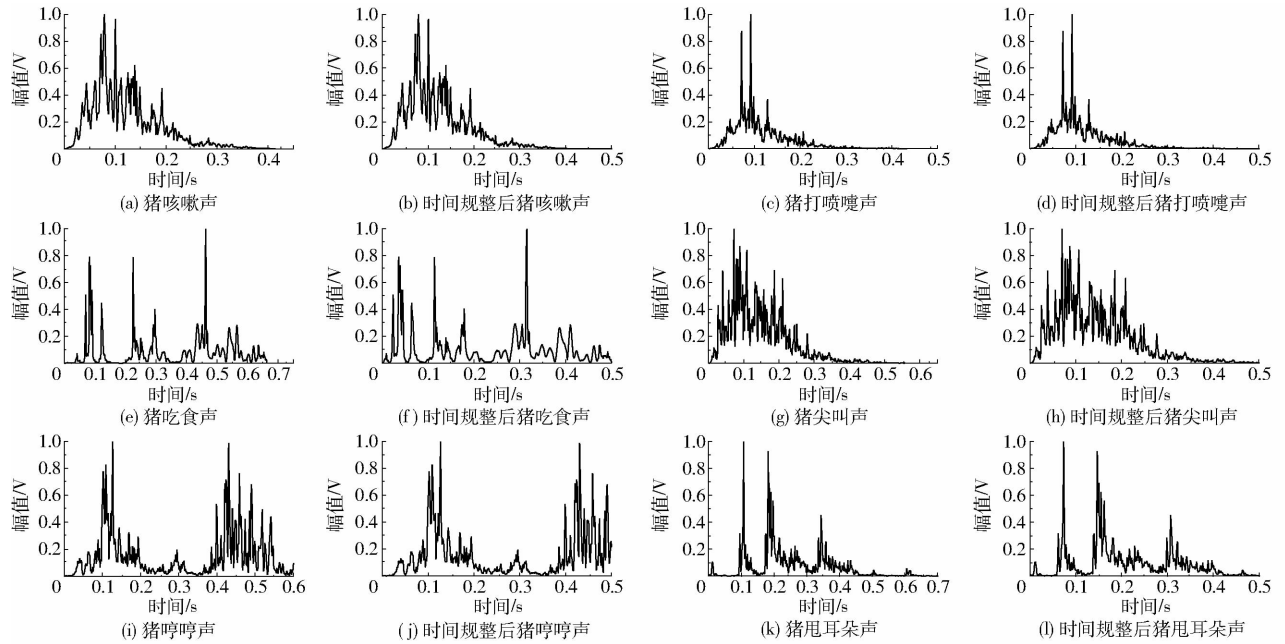


图 5 猪声音样本短时能量时间规整前后对比

Fig. 5 Comparison of short-time energy of pig sound and sound after time warping

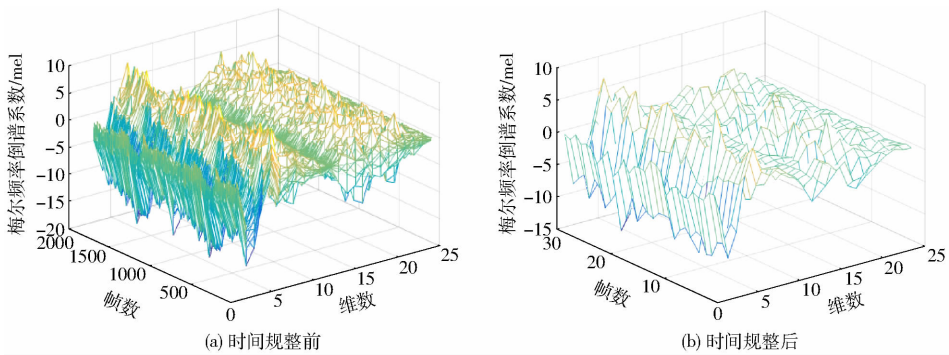


图 6 猪咳嗽声的 MFCC 图

Fig. 6 MFCC of pig cough

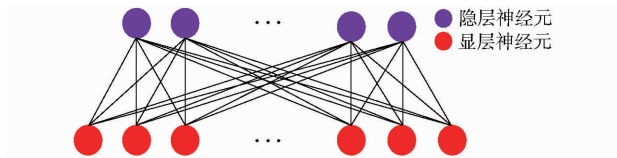


图 7 受限玻尔兹曼机网络结构

Fig. 7 Structure of RBM

0 表示神经元处于抑制状态, 状态 1 表示神经元处于激活状态。令向量 $s \in (0, 1)_d$ 表示 d 个神经元的状态, w_{ij} 表示神经元 i 与 j 之间的连接权值, θ_i 表示神经元 i 的阈值, 则状态向量 s 所对应的玻尔兹曼机能量计算公式为

$$E_s = - \sum_{i=1}^{d-1} \sum_{j=i+1}^d w_{ij} s_i s_j - \sum_{i=1}^d \theta_i s_i \quad (7)$$

若网络中的神经元不依赖于输入值的顺序进行更新, 则网络最终会达到玻尔兹曼分布。此时状态向量 s 出现的概率仅由其能量与所有可能状态向量的能量确定, 状态向量 s 出现的概率计算公式为

$$P_s = \frac{e^{-E_s}}{\sum_s e^{-E_s}} \quad (8)$$

本文采用 3 层 RBM 堆叠形成的 DBN 作为猪咳嗽声识别模型, 网络结构如图 8 所示。

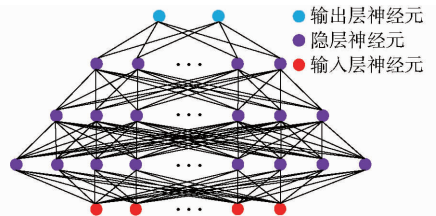


图 8 深度信念网络猪咳嗽声识别模型

Fig. 8 Pig cough sound recognition model based on DBN

将经过时间规整算法提取的 300 维短时能量和 720 维 MFCC 组合得到 1 020 维组合特征参数作为 DBN 的输入, DBN 输入层神经元个数选为 1 020 个。最后一层为输出层, 根据识别对象为猪咳嗽声与猪非咳嗽声, 将输出层神经元个数选为 2 个。隐层神经元个数选取不合理会导致 DBN 识别能力和网络

容错性降低,设每个隐层神经元个数分别为 l_1 、 l_2 和 l_3 ,则 DBN 模型可表示为 $1020-l_1-l_2-l_3-2$ 。隐层神经元个数经验公式为

$$l_h = \sqrt{l_{h-} + l_{h+}} + \delta \quad (9)$$

式中 l_h ——隐层神经元个数

l_{h-} ——前一层神经元个数

l_{h+} ——后一层神经元个数

δ ——平衡参数,取 0~10 之间的常数

根据式(9)得到求解各隐层神经元个数的方程组为

$$\begin{cases} l_1 = \sqrt{1020 + l_2} + \delta \\ l_2 = \sqrt{l_1 + l_3} + \delta \\ l_3 = \sqrt{l_2 + 2} + \delta \end{cases} \quad (10)$$

通过选取合适的 δ 值,解方程组(10)得 $l_1 = 37$, $l_2 = 12$, $l_3 = 7$ 。

为了得到最优的各隐层神经元个数,按照梯度 5 依次将 l_1 取值为 32、37、42 和 47,将 l_2 取值为 7、12、17 和 22,将 l_3 取值为 2、7、12 和 17。经过多次训练对比重构误差和训练收敛时长,最终将 l_1 、 l_2 和 l_3 分别取值为 42、17 和 7,故 DBN 网络模型结构设定为 $1020-42-17-7-2$ 。

3.2 深度信念网络猪咳嗽声识别模型训练

本模型采用的非监督贪婪算法逐层预训练 RBM 和 BP 算法^[22-23]以整体微调的方式来训练 DBN。预训练的过程即逐层训练 RBM 的过程。将每一个猪声音样本特征参数作为一个状态向量 s ,RBM 训练的目的是通过式(7)最小化玻尔兹曼机的能量,同时通过式(8)最大化状态向量 s 出现的概率,进而得到对应的 RBM 权值 w_{ij} 和阈值 θ_i 参数。这个过程用对比散度(Contrastive divergence, CD)算法来实现^[24-25]。

对于一个含有 l_d 个显层神经元、 l_q 个隐层神经元的 RBM,令 v 和 h 分别表示显层与隐层神经元的状态向量,由 RBM 的特殊网络结构可得,由隐层计算显层和显层计算隐层的条件概率分别为

$$P(v/h) = \prod_{i=1}^{l_d} P(v_i/h) \quad (11)$$

$$P(h/v) = \prod_{j=1}^{l_q} P(h_j/v) \quad (12)$$

由 CD 算法原理,对于每一个猪声音样本 v ,首先根据式(12)计算出隐层神经元的状态分布,然后由此概率分布通过吉布斯采样得到 h ;接着根据式(11)和式(12)得到 RBM 权值更新公式为

$$w_{i+1} = w_i + \eta(v\mathbf{h}^T - v'\mathbf{h}'^T) \quad (13)$$

式中 η ——学习率,本文设置为 0.1

v' —— v 经过吉布斯采样的结果

h' —— h 经过吉布斯采样的结果

w_i ——第 i 次训练得到的权值

w_{i+1} ——第 $i+1$ 次训练得到的权值

为了提高 DBN 训练效率,本文采用数据包(每包 50 个样本)分包进行学习训练。为防止 DBN 训练时陷入过拟合,在权值更新过程中引入权重衰减(Weight decay)进行修正,通常情况下权重衰减取值范围为 0.000 01~0.01^[26]。将式(13)修改为

$$w_{i+1} = w_i + \eta(v\mathbf{h}^T - v'\mathbf{h}'^T + rw_i) \quad (14)$$

式中 r ——权重衰减,本文设置为 0.000 2

3.3 猪咳嗽声识别结果分析

采用 3 个指标来衡量实验结果:猪咳嗽声识别率(正确识别的猪咳嗽样本占测试集中猪咳嗽样本总数百分比)、猪咳嗽声误识别率(被误识别为猪咳嗽声的猪非咳嗽样本占测试集中猪非咳嗽样本总数百分比)和总识别率(正确识别的猪咳嗽样本和猪非咳嗽样本占测试集样本总数百分比)。采用 5 折交叉验证方法进行交叉验证,即将猪咳嗽声与 5 种非咳嗽样本平均分成 5 等分,并按照训练集与测试集样本 4:1 的比例分成 5 组,进行交叉验证,实验结果如表 1 所示。

表 1 5 折交叉验证结果

Tab. 1 Comparison of recognition rate of cross

组别	validation method		%
	猪咳嗽声识别率	猪咳嗽声误识别率	
1	94.12	7.45	93.21
2	90.76	8.07	91.43
3	91.60	8.07	91.79
4	92.44	7.45	92.86
5	92.44	8.07	92.14

通过表 1 可知,各组猪咳嗽声识别率和总识别率均高于 90.00%,猪咳嗽声误识别率不超过 8.07%,说明了本文采用的基于 DBN 的猪咳嗽声识别模型是有效的。表 1 中第 1 组猪咳嗽声识别率达到 94.12%,误识别率达到 7.45%,总识别率达到 93.21%,为最佳实验组。

3.4 猪声音信号特征参数降维

经过时间规整算法提取的 300 维短时能量和 720 维 MFCC 组合的 1 020 维猪声音信号特征参数维数高,为了提炼出有效的特征参数进而提高猪咳嗽声识别效率,本文采用主成分分析(Principal component analysis, PCA)^[27]舍去原 1 020 维特征参数中表征猪声音信号能力弱的特征分量。

其主要实现过程为:

设 m 为猪声音样本个数, n 为每个样本特征参

数维度,原始数据矩阵 $X_{m \times n}$ 可表示为

$$X_{m \times n} = [x_{ab}]_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (15)$$

将所有样本进行中心化得到初始化后的数据

$$x_{ab} = x_{ab} - \frac{\sum_{b=1}^n x_{ab}}{n} \quad (a = 1, 2, \dots, m) \quad (16)$$

初始化后的数据矩阵 $X'_{m \times n}$ 的协方差矩阵为

$$\Sigma = \frac{\sum_{a=1}^m X'_{m \times n} X'_{m \times n}^T}{m - 1} \quad (17)$$

Σ 的特征值和特征向量为

$$\begin{cases} \Sigma W = \lambda W \\ W = [w_1 \quad w_2 \quad \cdots \quad w_n] \\ \lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_b, \dots, \lambda_n) \end{cases} \quad (18)$$

式中 λ_b —— Σ 的特征值

W —— 特征值 λ_b 对应的特征向量

将特征值 λ_b 从大到小排列,取前 p 个特征值对应的主成分, p 值的选取通过特征值贡献率确定,其计算公式为

$$L(p) = \frac{\sum_{b=1}^p \lambda_b}{\sum_{b=1}^n \lambda_b} \quad (19)$$

1 020 维短时能量和 MFCC 特征参数与贡献率的关系如图 9 所示。

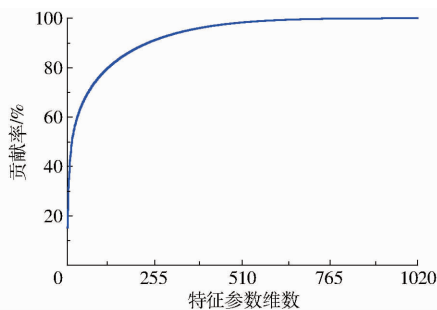


图 9 特征参数维数的贡献率

Fig. 9 Contribution rate of dimension of feature parameters

图 9 中,当特征值的贡献率大于 90% 时,可认为这 p 个主成分的特征参数就能反映原猪声音信号的信息。经过计算,当 p 取 479 时,贡献率达到 98.01%。于是,将 1 070 维特征参数降到 479 维。通过式(9)再次计算 DBN 网络结构为 479 - 27 -

14 - 6 - 2,5 折交叉验证实验结果如表 2 所示。

表 2 特征参数降维后识别结果

Tab. 2 Recognition rate after feature parameters

dimension reduction			%
组别	猪咳嗽声识别率	猪咳嗽声误识别率	总识别率
1	95.80	6.83	94.29
2	93.28	6.21	93.57
3	93.28	6.83	93.21
4	94.12	7.45	93.21
5	92.44	5.60	93.57

对比表 1 和表 2 识别结果可知降维后的各实验组猪咳嗽声识别率、误识别率和总识别率较降维前均有相应改善。降维后最佳组为第 1 组,相对降维前最佳组猪咳嗽声识别率提高了 1.68 个百分点,误识别率降低了 0.62 个百分点,总识别率提高了 1.08 个百分点,因此经过主成分分析法优化得到的 479 维猪声音特征参数可以代替原 1 020 维猪声音特征参数。

4 结论

(1) 利用时间规整算法提取猪咳嗽、打喷嚏、进食、尖叫、哼哼和甩耳朵 6 种声音短时能量 300 维, MFCC 720 维, 以此 1 020 维组合特征参数作为猪咳嗽声与猪非咳嗽声的衡量指标。进一步采用主成分分析法提取原 6 种声音的 1 020 维特征参数对应的贡献率为 98.01% 时的 479 维特征参数。实验表明,短时能量和 MFCC 的结合可以很好地衡量猪咳嗽声与猪非咳嗽声的差异。

(2) 将深度信念网络引入猪声音识别领域,构建一个 5 层的深度信念网络猪咳嗽声识别模型,采用非监督贪婪算法逐层预训练 RBM 和 BP 算法以整体微调的方式训练 DBN。设定 50 个样本的小批量数据包训练模式,学习率为 0.1,引入权重衰减 0.000 2,能够获取收敛较优的猪咳嗽声识别模型。

(3) 通过 5 折交叉实验验证,采用 1 020 维特征参数,基于 DBN 的猪咳嗽声识别模型最佳组猪咳嗽声识别率达到 94.12%,误识别率为 7.45%,总识别率达到 93.21%,经过 PCA 降维至 479 维后,猪咳嗽声识别率和总识别率相对降维前均有所提高,误识别率有所下降,最佳组猪咳嗽声识别率为 95.80%,误识别率为 6.83%,总识别率为 94.29%。

参 考 文 献

1 SILVA M, FERRARI S, COSTA A, et al. Cough localization for the detection of respiratory diseases in pig houses[J]. Computers and Electronics in Agriculture, 2008, 64(2):286 - 292.

- 2 MITCHELL S, VASILEIOS E, SARA F, et al. The influence of respiratory disease on the energy envelope dynamics of pig cough sounds[J]. *Computers and Electronics in Agriculture*, 2009, 69(1):80–85.
- 3 SARA F, MITCHELL S, MARCELLA G, et al. Cough sound analysis to identify respiratory infection in pigs[J]. *Computers and Electronics in Agriculture*, 2008, 64(2):318–325.
- 4 何东健, 刘冬, 赵凯旋. 精准畜牧业中动物信息智能感知与行为检测研究进展[J/OL]. *农业机械学报*, 2016, 47(5):231–244. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160532&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.05.032.
- HE Dongjian, LIU Dong, ZHAO Kaixuan. Review of perceiving animal information and behavior in precision livestock farming[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016,47(5):231–244. (in Chinese)
- 5 EXADAKTYLOS V, SILVA M, AERTS J M, et al. Real-time recognition of sick pig cough sounds[J]. *Computers and Electronics in Agriculture*, 2008, 63(2):207–214.
- 6 GUARINO M, JANS P, COSTA A, et al. Field test of algorithm for automatic cough detection in pig house[J]. *Computers and Electronics in Agriculture*, 2008, 62(1):22–28.
- 7 HIRTUM A V, BERCKMANS D. Fuzzy approach for improved recognition of citric acid induced piglet coughing from continuous registration[J]. *Journal of Sound and Vibration*, 2003, 266(3):677–686.
- 8 马辉栋, 刘振宇. 语音端点检测算法在猪咳嗽检测中的应用研究[J]. *山西农业大学学报:自然科学版*, 2016,36(6):445–449.
- MA Huidong, LIU Zhenyu. Application of end point detection in pig cough signal detection[J]. *Journal of Shanxi Agricultural University:Natural Science Edition*, 2016,36(6):445–449. (in Chinese)
- 9 刘振宇, 赫晓燕, 桑静, 等. 基于隐马尔可夫模型的猪咳嗽声音识别的研究[C]//中国畜牧兽医学会信息技术分会第十届学术研讨会论文集, 2015:99–104.
- 10 徐亚妮, 沈明霞, 闫丽, 等. 待产梅山母猪咳嗽声识别算法的研究[J]. *南京农业大学学报*, 2016,39(4):681–687.
- XU Yani, SHEN Mingxia, YAN Li, et al. Research of predelivery Meishan sow cough recognition algorithm[J]. *Journal of Nanjing Agricultural University*, 2016,39(4):681–687. (in Chinese)
- 11 HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006,18(7):1527–1554.
- 12 HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786):504–507.
- 13 LECUN Y, BENGIO Y, HINTON G E. Deep learning[J]. *Nature*, 2015,512:436–444.
- 14 MOHAMED A R, DAHL G E, HINTON G E. Acoustic modeling using deep belief networks[J]. *IEEE Transactions on Audio Speech and Language Processing*, 2011, 20(1):14–22.
- 15 LI Xiangang, YANG Yuning, PANG Zaihu, et al. A comparative study on selecting acoustic modeling units in deep neural networks based large vocabulary Chinese speech recognition[J]. *Neurocomputing*, 2015,170:251–256.
- 16 HU Y, LOIZOU P C. Incorporating a psychoacoustical model in frequency domain speech enhancement[J]. *Signal Processing Letters IEEE*, 2004, 11(2):270–273.
- 17 吴红卫, 吴镇扬, 赵力. 基于多窗谱的心理声学语音增强[J]. *声学学报*, 2007, 32(3):275–281.
- WU Hongwei, WU Zhenyang, ZHAO Li. Psychoacoustical enhancement of speech based on multitaper spectrum[J]. *Acta Acustica*, 2007,32(3):275–281. (in Chinese)
- 18 戴元红, 陈鸿彬, 乔德江, 等. 基于短时能量比的语音端点检测算法的研究[J]. *通信技术*, 2009,42(2):181–183.
- DAI Yuanhong, CHEN Hongchang, QIAO Dejiang, et al. Speech endpoint detection algorithm analysis based on short-term energy ratio[J]. *Communications Technology*, 2009,42(2):181–183. (in Chinese)
- 19 王山海, 景新幸, 杨海燕. 基于深度学习神经网络的孤立词语音识别的研究[J]. *计算机应用研究*, 2015,32(8):2289–2292.
- WANG Shanhai, JING Xinxing, YANG Haiyan. Study of isolated speech recognition based on deep learning neural networks[J]. *Application Research of Computers*, 2015,32(8):2289–2292. (in Chinese)
- 20 林玮, 杨莉莉, 徐柏龄. 基于修正 MFCC 参数汉语耳语音的话者识别[J]. *南京大学学报:自然科学版*, 2006,42(1):54–62.
- LIN Wei, YANG Lili, XU Boling. Speaker identification in Chinese whispered speech based on modified-MFCC[J]. *Journal of Nanjing University: Natural Sciences*, 2006,42(1):54–62. (in Chinese)
- 21 李志忠, 腾光辉. 基于改进 MFCC 的家禽发声特征提取方法[J]. *农业工程学报*, 2008,24(11):202–205.
- LI Zhizhong, TENG Guanghui. Feature extraction for poultry vocalization recognition based on improved MFCC[J]. *Transactions of the CSAE*, 2008,24(11):202–205. (in Chinese)
- 22 YU C C, LIU B D. A backpropagation algorithm with adaptive learning rate and momentum coefficient[C]//International Joint Conference on Neural Networks. *IEEE*, 2002:1218–1223.
- 23 HAMEED A A, KARLIK B, SALMAN M S. Back-propagation algorithm with variable adaptive momentum[J]. *Knowledge-Based Systems*, 2016,114:79–87.
- 24 ERHAN D, BENGIO Y, COURVILLE A, et al. Why does unsupervised pre-training help deep learning? [J]. *Journal of Machine Learning Research*, 2010, 11(3):625–660.
- 25 HINTON G E. Learning multiple layers of representation[J]. *Trends in Cognitive Sciences*, 2007, 11(10):428–434.
- 26 HINTON G E. A practical guide to training restricted boltzmann machines[J]. *Momentum*, 2012, 9(1):599–619.
- 27 李慧, 祁力钧, 张建华, 等. 基于 PCA-SVM 的棉花出苗期杂草类型识别[J/OL]. *农业机械学报*, 2012,43(9):184–189. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20120934&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2012.09.034.
- LI Hui, QI Lijun, ZHANG Jianhua, et al. Recognition of weed during cotton emergence based on principal component analysis and support vector machine[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2012,43(9):184–189. (in Chinese)