The Detection of the Deception Based on Stacked Denoising Autoencoder*

LEI Peizhi[1], *FU Hongliang*[1*], *TAO Huawei*[1], *JIANG Pengxu*[1], *ZHAO Li*[2], *YE Chao*[2]
(1.*College of Information Science and Engineering, Henan University of Technology, Zhengzhou* 450001, *China*;
2.*School of Information Science and Engineering, Southeast University, Nanjing* 210096, *China*)

Abstract: In order to improve the accuracy of detection of deception, a speech detection algorithm based on stacked denoising autoencoder is proposed(SDA-SVM). The algorithm first uses Opensimile to extract 384 dimensions voice features. Then, two layers of denoising autocoder networks are constructed to transform the voice features. Finally, the SVM classifier is used to classify the truthful and deceptive speech. The source of the speech used is the CSC lie corpus, experimental results show that compared with the traditional SVM classification, the proposed algorithm increases the accuracy by at least 1.85%.

Key words: polygraph; voice features; stacked denoising autoencoder; SVM
EEACC: 4130; 6220M **doi**: 10.3969/j.issn.1005-9490.2019.03.047

基于栈式去噪自编码器的语音测谎算法*

雷沛之[1], 傅洪亮[1*], 陶华伟[1], 姜芃旭[1], 赵 力[2], 叶 超[2]
(1. 河南工业大学信息科学与工程学院, 郑州 450001; 2.东南大学信息科学与工程学院, 南京 210096)

摘 要: 为了进一步提高谎言语音检测的准确率, 提出了一种基于栈式去噪自编码器的语音测谎算法(SDA-SVM)。该算法首先采用 OpenSMILE 提取了 384 维语音特征; 然后构建了两层去噪自编码网络对语音特征进行变换加工; 最后, 采用 SVM 分类器对语音是否为谎言进行分类识别。所用语音来源为 CSC 测谎语料库, 实验结果显示: 相比传统的 SVM 分类, 所提算法的检测准确率至少提升 1.85%。

关键词: 测谎; 语音特征; 栈式去噪自编码器; SVM
中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1005-9490(2019)03-0793-04

说谎, 字典将其定义为“说话人在知道事实的前提下, 通过刻意隐瞒并提供与事实不符的语言信息的行为”。测谎, 即检测某些语言是否为谎言。谎言的危害是巨大的, 例如盛行的电信诈骗, 传销诈骗, 通常给受害者及其家属带来毁灭性的后果, 因此检测谎言一直以来都是研究者们研究的热门领域[1]。早期的人工测谎培训, 存在着培训成本高, 识别率低的缺点, 于是研究人员开始尝试用人机进行谎言检测。开展语音测谎研究, 对于推动计算机, 语音识别等相关学科的发展具有重大的现实意义, 对提高司法鉴定水平、促进商业贸易公平等也有较强的辅助作用。

随着测谎研究的深入, 语言中包含的特征成为了测谎的依据, 例如梅尔频率倒谱系数[2], 并使用一些传统的分类器进行识别, 但由于特征提取算法难以学习到语言的深层特征以及分类器有限的泛化能力, 所以经常遇到维数灾难, 过拟合等问题。针对上述问题, 深度学习的出现提供了新的解决思路[3], 降噪自编码器则是深度学习中的一个模块。堆栈降噪自编码器 SDA(Stacked Denoising Autoencoder)[4], 使用无标签的数据进行无监督的逐层贪婪预训练来初始化网络参数, 并且用有标签的数据和反向传播算法进行微调, 进一步优化模型[5]。本文使用具有两层网络的堆栈降噪自编码器, 并后接 SVM 分类机进行检测。实验结果表明该网络可以提高谎言检测的准确率。

1 栈式去噪自编码器

1.1 去噪自编码器

自编码器 AE(AutoEncoder)[6]是一种具有 3 层

项目来源: 国家自然科学基金项目(61673108); 国家自然科学基金项目(61601170); 河南省高等学校重点科研项目(19A510009)
收稿日期: 2018-05-29 **修改日期**: 2018-07-10

的神经网络,可以提取出机器学习所需要的某些数据特征,包括输入层,隐藏层和输出层。在现实应用场景中,人们需要更具有鲁棒性的特征[7],因此引入去噪自编码器 DAE(Denoising Autoencoder),其结构如图 1 所示。

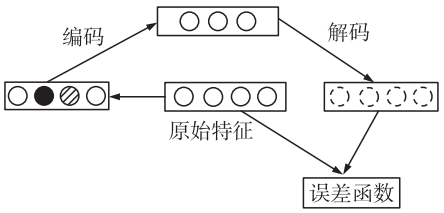

图 1 去噪自编码器结构

去噪自编码器有编码和解码两个部分。对于编码部分,可以表示为:

$$Y=S(W_1X+B_1) \tag{1}$$

式中:S 为非线性激活函数,一般为 Sigmoid 函数,W_1 为 $n_1\times n_2$ 维权重矩阵,B_1 为 n_2 维偏置向量,X 为加入噪声后的数据,Y 为隐层的输出数据。

对于解码部分:

$$Z=S(W_2Y+B_2) \tag{2}$$

式中:W_2 为 $n_2\times n_3$ 维权重矩阵,B_2 为 n_3 维偏置向量,Z 为输出层的输出数据。它的训练过程就是用随机梯度下降法来实现重构误差最小化,其误差函数可以表示为:

$$L=\frac{1}{2}|Z-X|^2 \tag{3}$$

1.2 栈式去噪编码器

栈式去噪自编码器(Stacked Denoising Autoencoder)是由多个训练好的去噪自编码器堆叠而成,可以提取更深层的特征。首先利用不带标签的数据使用贪婪逐层预训练每一个自编码器。完成预训练之后利用分类器和用反向传播算法对整个网络系统进行微调,参数更新公式为:

$$W=W-\eta\frac{\partial J(y,y\_)}{\partial W} \tag{4}$$

$$b=b-\eta\frac{\partial J(y,y\_)}{\partial b} \tag{5}$$

式中:W,b 为网络权重和偏置向量,η 为学习率,$J(y,y\_)$ 为分类结果和正确结果的误差函数,一般为相对熵函数。

2 支持向量机

SVM 是 Corinna Cortes 和 Vapnik 等[8]提出的一种有监督的分类模型,可以用来做回归,分类等分析,是将原始数据映射到高维空间中进行线性分类。若原数据为 x,映射后的数据为 $\varphi(x)$,则其超平面可表示为式(6),目标就是找到最佳超平面。

将数据映射到不同的空间需要不同的核函数,一般常用的核函数有:线性核函数式(7),多项式核函数公式(8)[9],本文使用线性核函数。

$$f(x)=W\cdot\varphi(x)+b \tag{6}$$

$$K(X,Y)=X\cdot Y \tag{7}$$

$$K(X,Y)=[(X\cdot Y)+1]^d \tag{8}$$

3 “SDA-SVM”算法模型

本文的模型共有 3 个步骤:

(1)特征提取

特征是测谎系统中的重要分类依据。在 2009 年首次举办的国际语音情感识别挑战赛 INTERSPEECH 2009 EC 的分类器子挑战中,为了更好地描述语音中蕴含的特征,举办方根据声学特征中使用最为广泛的特征和函数[10],创建了包含 384 维向量的基本特征集,其中有 16 个低层描述符和 12 个统计函数,如表 1 所示。本文将实验所用的语音经 OpenSMILE 软件依该特征集提取出 384 维特征。

表 1 2009 国际语音情感识别挑战赛中的基本特征集[11]

低层描述符(16*2)	统计函数(16)
均方根能量	标准差
基频	峰度,偏度
梅尔频率倒谱系数 1~12	线性回归:斜率,偏量,均方误差
过零率	均值
谐波噪声比	极限值:最大最小值,相对位置,范围

(2)特征重构及表达

图 2 为本文所搭建的重构特征的栈式去噪自编码器。

将谎言库中语音划分为训练集和测试集,用训练集来训练整个分类系统,$H(n)$ 为第一隐层输出,

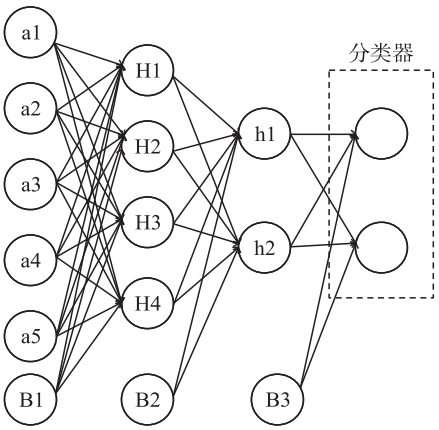

图 2 栈式去噪自编码器结构图

$h(m)$为第二隐层输出,训练的相关参数及最终系统的隐层单元数如表 2 所示。

表 2　训练及测谎系统的参数

网络参数	值
第一隐层节点数	120
第二隐层节点数	80
预训练次数	200
微调次数	100
预训练学习率	0.001
反向传播算法学习率	0.001
激活函数	sigmoid

(3)分类检测

在已经训练好的栈式去噪自编码器的后面接 SVM 分类器,将测试集数据经栈式去噪自编码器重构完特征之后进行检测,并记录其准确率。

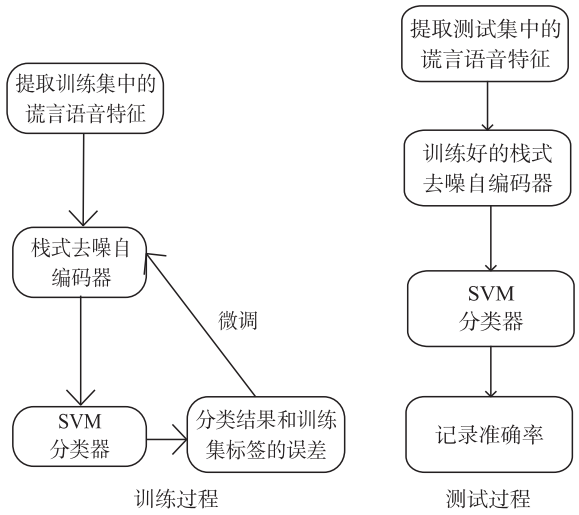

图 3　训练及测试过程框图

4　实验分析

4.1　实验设置

本文所用的谎言数据库为 CSC 数据库,参与录制的为 32 名哥伦比亚大学的学生或教师,其中男性和女性各占一半。这 32 名人员被告知即将参加“寻找符合美国顶级优秀企业家人才”的活动,形式是对音乐,人际交流,地理,公民等 6 个方面进行提问测试,由面试官将参与者的得分和面试交流来判断他们是否具有“优秀企业家”的特质。并最终得到约 7 h 的谎言语音样本[12]。

从该库中剪切出 5 000 条语音,其中男女各 15 名,经 OpenSMILE 软件,提取出了 384 维的特征集用于实验。并采用“准确率”作为分类性能指标。

4.2　实验结果与分析

实验分两次,第 1 次是将测试集经过训练好的栈式去噪自编码器之后再用 SVM 进行分类,第 2 次直接使用 SVM 进行分类,并将结果进行比对。SVM 分类器的 C 值选择 10 组,值为从 1 到 10,对于每个 C 值都进行 10 次实验并取最后的平均值。

由图 4 可以看出,将数据经栈式自编码器处理之后再进行分类与直接将数据进行分类相比,准确率最低提升了 1.85%,最高提升了 2.47%。因此,本文设计的栈式去噪自编码器可以提取出更高阶的特征,提升了 SVM 对于谎言检测的准确率。

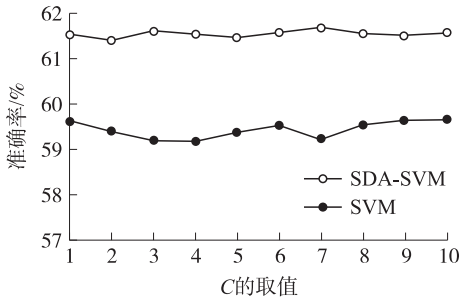

图 4　重复试验次数及其准确率

为进一步验证算法的有效性,将所提算法与其他分类器进行比较,对比分类器包括 SVM、Softmax、MLP,其中 SVM 参数依照图 4,C 值选取 7,Softmax 和 MLP 中的学习率取 0.01。对于每种分类器都进行 10 次实验并取平均值。

表 3　各个分类器的检测准确率

分类器	平均识别率
softmax	58.34%
MLP	59.41%
SVM	59.40%
SDA+SVM	61.53%

由比对结果可以看出,本文设计的测谎系统相较于传统的分类器可以提高谎言检测的准确率。

5　结束语

针对传统 SVM 分类器对谎言分类的准确率不高的问题,本文通过栈式去噪自编码器对数据进行预处理,去除了特征中的一些冗余信息,更有效的提取了语音中的更高阶特征以助于分类,并在 CSC 数据库上进行了实验,结果表明该算法可以提高 SVM 对谎言检测的正确率。下一步,将会对自编码器的个数,以及网络中隐层的单元数进行研究,重构更高阶更有助于分类的特征,进一步提升检测的正确率。

参考文献:

[1] Eitan Elaad. Cognitive and Emotional Aspects of Polygr-Aph Diagnostic Procedures:A Comment on Palmatier and Rovner(2015)[J]. International Journal of Psychophysiology,2015,95(1):14-15.

[2] 余华,章勤杰,赵力. 语音情感识别算法中新型参数研究[J].

电子器件,2017,40(5):1234-1237.
[3] Anthes G. Deep Learning Comes of Age[J]. Communication of the ACM,2013,56(6):13-15.
[4] Moumita Saha,Pabitra Mitra,Ravi S Nanjundiah. Predictor Discovery for Early-Late Indian Summer MonsoonUsing Stacked Autoencoder[J]. Procedia Computer Science,2016(80):565-576.
[5] 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.
[6] Kai Sun,Jiangshe Zhang,Chunxia Zhang,et al. Generalized Extreme Learning Machine Autoencoder and a New Deep Neural Network [J]. Neurocomputing,2016,230:374-381.
[7] Hiroshi Ohno. Linear Guided Autoencoder:Representation Learning with Linearity[J]. Applied Soft Computing,2017,55(6):566-578.
[8] 陶坚,喻擎苍. 支持向量机在入侵检测系统中的应用[J]. 电子器件,2007(6):2226-2228.
[9] Hari Seetha,Saravanan R,Narasimha Murty M. Pattern Synthesis Using Multiple Kernel Learning for Efficient SVM Classification [J]. Cybernetics and Information Technologies,2012,12(4):77-94.
[10] Schuller B,Batliner A,Seppi D,et al. The Relevance of Feature Type for the Automatic Classification of Emotional User States:Low Level Descriptors and Functionals[C]//Interspeech. 2007:2253-2256.
[11] Schuller B,Seppi S,Batliner A. The Interspeech 2009 Emotion Challenge[C]//Proceedings Interspeech 2009,10th Annual Conference of the International Speech Communication Association,2009:312-315.
[12] Enos F,Benus S,Cautin R L,et al. Personality Factors in Human Deception Detection:Comparing Human to Machine Performance [C]//Interspeech. Pittsburgh:ISCAINST Speech Communication Assoc,2006:813-816.

雷沛之(1994-),男,汉族,河南郑州人,河南工业大学在读研究生,主要研究方向为信号与信息处理,pzmz_9394@163.com;

傅洪亮(1965-),男,汉族,河南郑州人,博士,河南工业大学教授,主要研究方向为宽带无线通信及现代信号处理,jackfu_zz@163.com。