

土木建筑CAE 结题报告

数字识别

张云翼

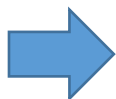
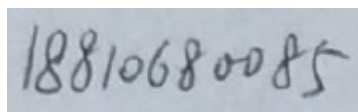
2016年5月26日



背景

描述

- 输入一个含有阿拉伯数字字符串的图象
- 输出字符串的内容



18810680085

步骤

图像预处理

字符切分

训练分类器

字符识别



方法-样本准备

训练样本

4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4

包含每个数字的各种写法100个，
共1000个样本。

测试样本

手机号码16个，
共176个数字。

13120180221
13261762515
18800191525
18810310328
18813062678

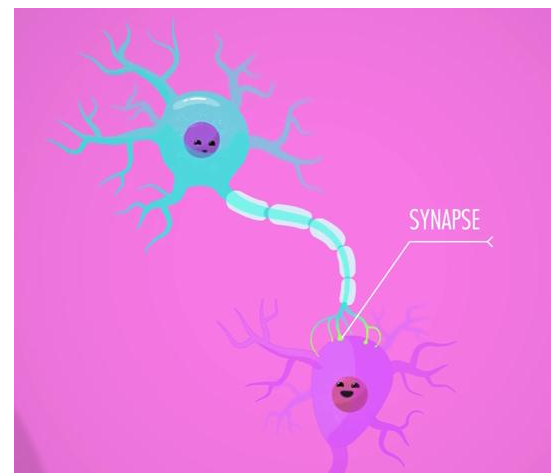
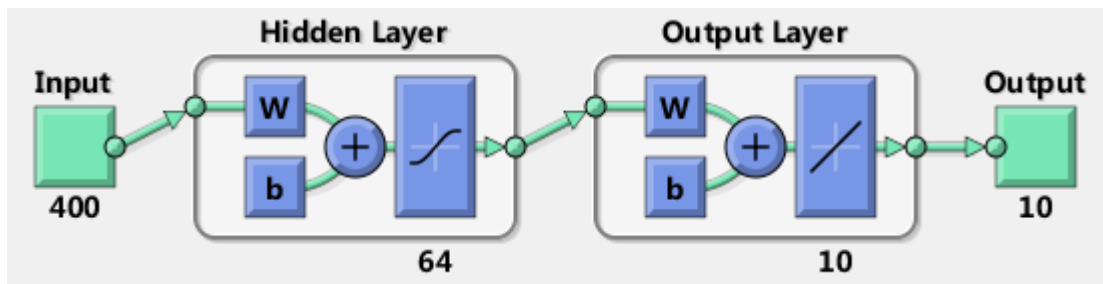
特征提取

识别每个数字的包围盒
缩放到20*20像素的大小
进行二值化



方法-分类器训练

人工神经网络



输入层包含400个特征。

隐藏层64个神经元，双曲正切变换。

输出层10个神经元，线性变换。

可以萃取隐藏的特征。

训练网络

读入特征

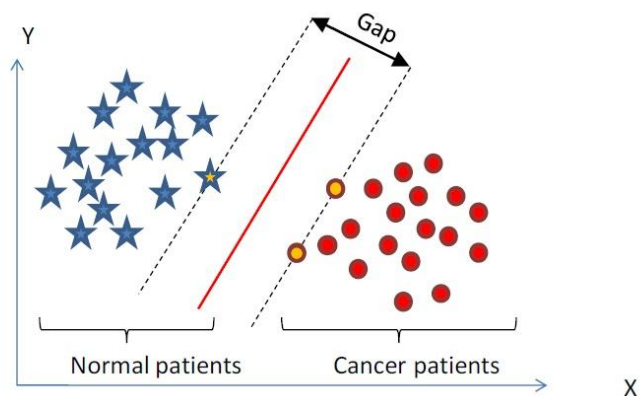
送入网络

得到分类
结果



方法-分类器训练

支持向量机



最大化切割平面到最近样本的距离。

1vn方法

训练10个分类器。

1v1方法

训练45个分类器。

训练分类器

读入特征

送入分类器

投票

得到分类结果



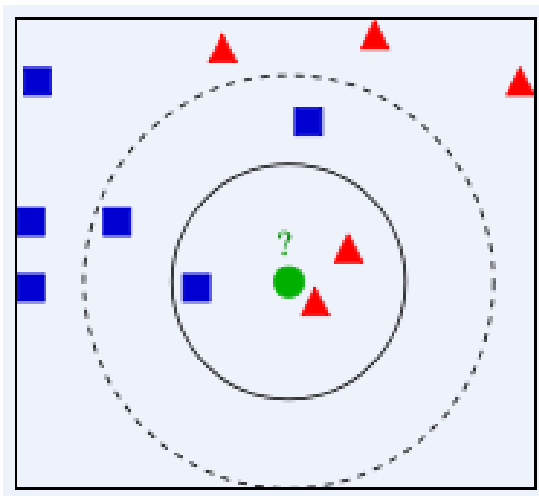
方法-分类器训练

K-最近邻

只需对样本进行预处理和特征提取。

无需训练。

设样本数为 n ，特征数为 m 。



读入特征

计算与各样本
的差异

选择前5个

得到分类
结果

空间复杂度 $O(mn)$

时间复杂度 $O(mn)$



方法-分类器训练

方法对比

	训练时间/s	分类时间/s	准确率
人工神经网络	27	1.12	93.75%
K-最近邻	0	3.27	95.45%
决策树	1.44	0.54	82.39%
SVM 1vn	3.58	2.8	94.89%
SVM 1v1	2.57	9.88	98.30%

人工神经网络训练时间最长，但分类时间短，适宜嵌入大量特征；
K-最近邻算法为被动学习算法，无需训练，但难以处理大量样本或特征。



方法-分类器训练

方法对比

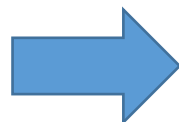
	训练时间/s	分类时间/s	准确率
人工神经网络	27	1.12	93.75%
K-最近邻	0	3.27	95.45%
决策树	1.44	0.54	82.39%
SVM 1v1	3.58	2.8	94.89%
SVM 1v1	2.57	9.88	98.30%

决策树算法准确率较低，不宜处理大量特征。

Svm算法准确率很高，但分类时间较长，这是因为需要进行多次分类并投票。



扩展-语音识别



18810680085

图像预处理

字符切分

训练分类器

字符识别



音频预处理

端点识别

训练分类器

字符识别



扩展-语音识别

语音学基础

声音三要素

响度

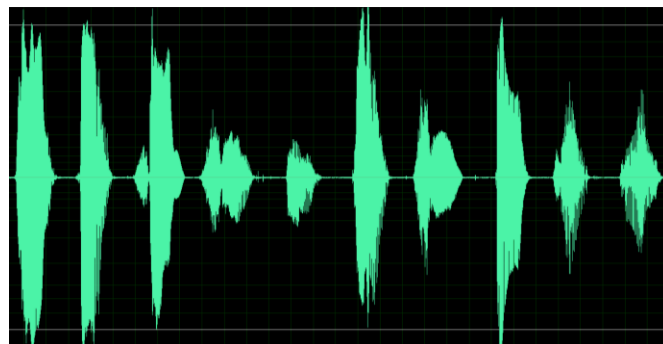
振幅

音高

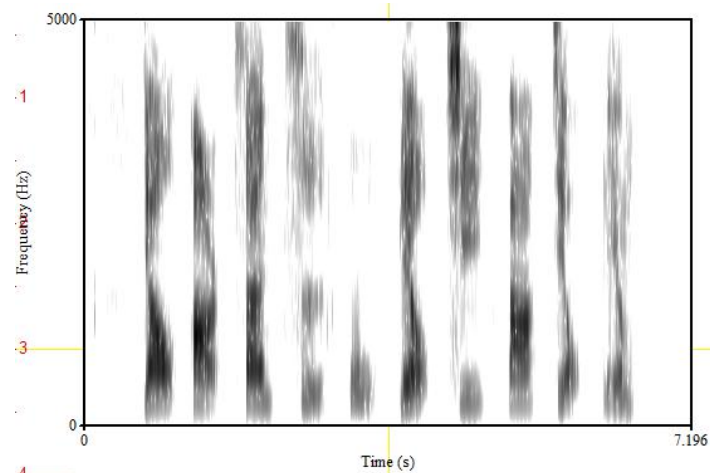
基频

音色

谐波



功率谱

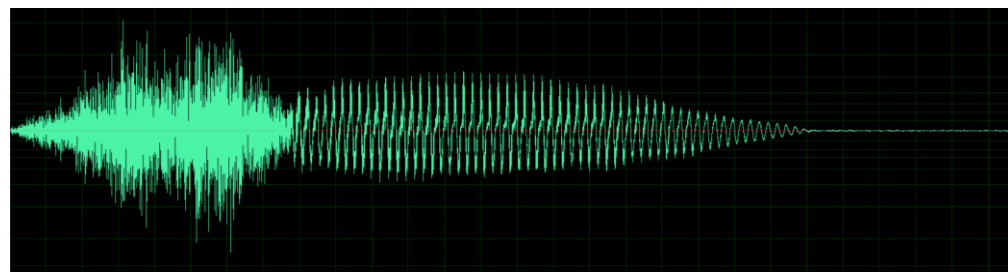
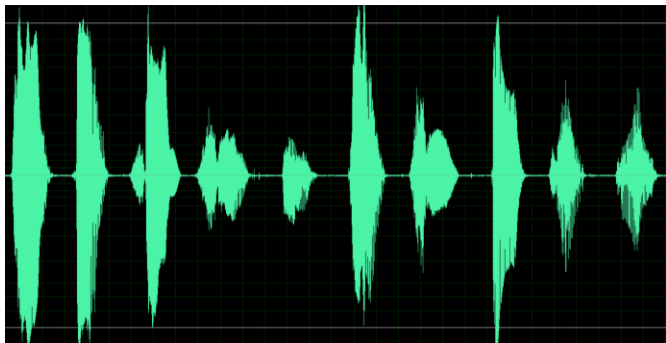


频率谱



扩展-语音识别

功率谱分析



功率谱分析可以对语音进行切分。

分帧

短时平均能量

端点判断

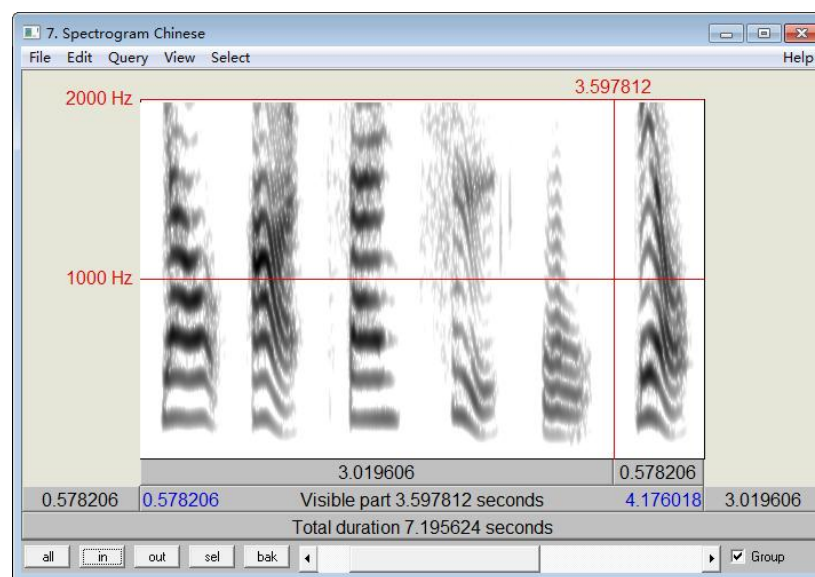
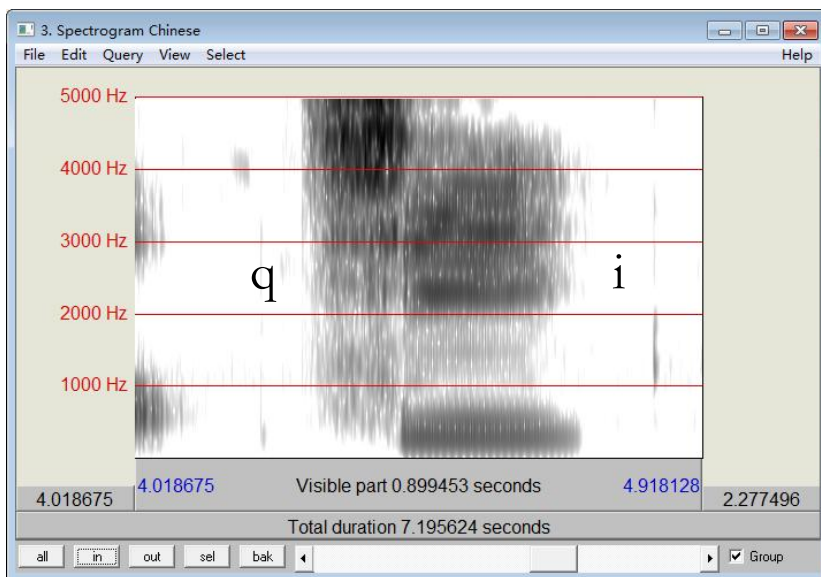
归一化

短时过零率



扩展-语音识别

频率谱分析

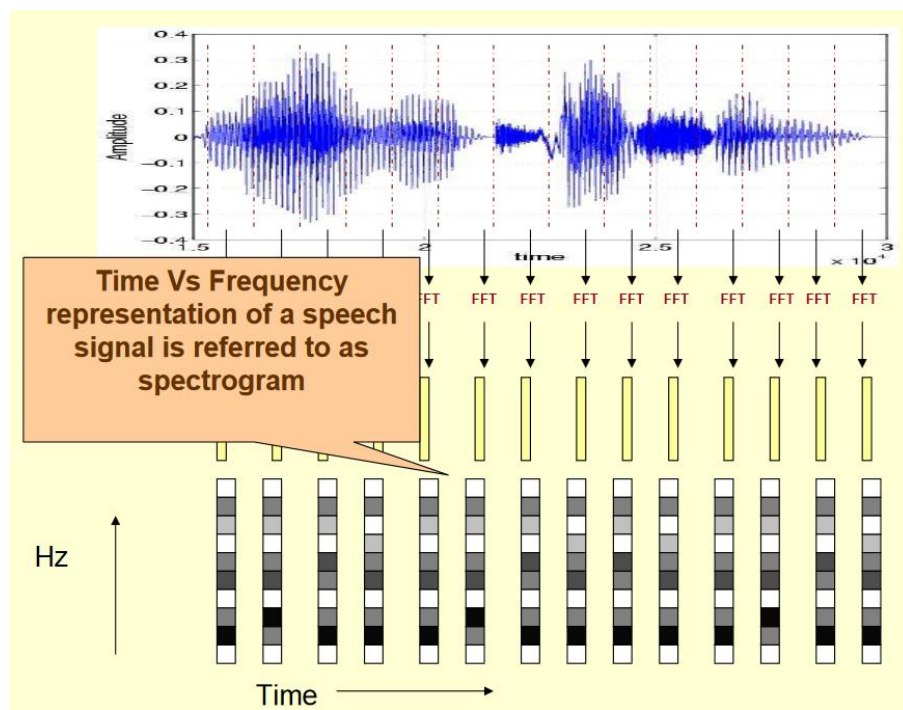
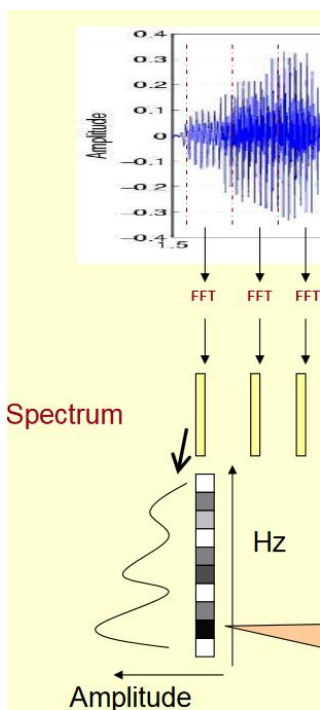


频率谱分析可以提供语音的特征。



扩展-语音识别

频率谱分析



加窗

快速傅里叶变换

得到频率谱

特征提取



扩展-语音识别

识别结果

每段语音长0.29秒，采样率44100Hz，每帧256个采样点，重叠80个采样点。每帧提取24个MFCC参数，每段语音共3672个特征。

使用人工神经网络进行识别，隐藏层选100个神经元。

识别结果不佳，正确率只有55%。

- 相对人工神经网络的VC维（自由度）来讲，样本数量远远不够。
- 采集的训练样本与测试样本存在一定的不一致。



总结

- 尝试了多种机器学习的智能化算法，并对其效率和性能进行了对比。
- 尝试了图象处理和语音处理的基本方法。
- 尝试了对图象和语音的数字识别。



谢谢

2016年5月26日