**目 录**

[第1章 前言 3](#_Toc515003610)

[1.1车牌识别介绍 3](#_Toc515003611)

[1.2车牌识别应用领域 3](#_Toc515003612)

[1.3实现方法简介 4](#_Toc515003613)

[1.4 OpenCV 4](#_Toc515003614)

[1.4 开发环境 6](#_Toc515003615)

[第2章 方案论证 7](#_Toc515003616)

[2.1 车牌定位方案论证 7](#_Toc515003617)

[2.1.1 基于车牌颜色分布的定位方法 7](#_Toc515003618)

[2.1.2 基于Radon变换的车牌定位方法 8](#_Toc515003619)

[2.1.3 基于边缘检测的车牌定位方法 8](#_Toc515003620)

[2.2 字符识别方案论证 9](#_Toc515003621)

[2.2.1 基于神经网络的字符识别 9](#_Toc515003622)

[2.2.2 基于KNN最邻近算法的字符识别 11](#_Toc515003623)

[2.3 方案比较与选择 12](#_Toc515003624)

[第3章 车牌位置检测 13](#_Toc515003625)

[3.1 图像预处理 13](#_Toc515003626)

[3.1.1 RGB空间与HSV空间的转换 13](#_Toc515003627)

[3.1.2 图像滤波 14](#_Toc515003628)

[3.1.3 图像二值化 15](#_Toc515003629)

[3.2 轮廓特征提取 16](#_Toc515003630)

[3.2.1 边缘检测 16](#_Toc515003631)

[3.2.2 轮廓筛选 17](#_Toc515003632)

[3.2.3 轮廓提取与分组 18](#_Toc515003633)

[3.3 车牌切割 20](#_Toc515003634)

[第4章 基于KNN的字符识别 21](#_Toc515003635)

[4.1 KNN简介 21](#_Toc515003636)

[4.2 字符预处理 22](#_Toc515003637)

[4.2.1 字符分割 22](#_Toc515003638)

[4.2.2 图像归一化 23](#_Toc515003639)

[4.3 样本训练 23](#_Toc515003640)

[4.3.1 贪心算法 23](#_Toc515003641)

[4.3.2 训练过程 24](#_Toc515003642)

[4.4 KNN算法的字符识别 27](#_Toc515003643)

[第5章 系统功能、指标参数 29](#_Toc515003644)

[谢辞 32](#_Toc515003645)

[参考文献 33](#_Toc515003646)

[附录 34](#_Toc515003647)

**摘要**

车牌识别是指带有车牌的图片经过特定系统的处理以字符形式输出车牌信息，是计算机视觉领域的一部分。本文研究内容的便是实现这样一个系统，主要包括车牌位置检测，字符识别两个部分。OpenCV是一个广受欢迎的开源计算机视觉库，提供了完善的图像处理函数，使用OpenCV可以大大节省开发时间加快开发效率，本系统的便是在OpenCV的基础上进行开发。系统在位置检测与字符分割部分取得了极高的准确度，字符识别部分准确率还有待提高。

**关键字：车牌识别，OpenCV，计算机视觉**

**Abstract**

Vehicle license plate Recognition is a part of computer vision field, which is a kind of picture with license plate, which is processed by special system to output license plate information in character form. The content of this paper is to realize such a system, mainly including license plate location detection, character recognition two parts. OpenCV is a popular open source computer Vision Library, providing a perfect image processing function, using OPENCV can greatly save development time and speed up development efficiency, this system is based on the OpenCV development. The system obtains the extremely high accuracy in the position detection and the character segmentation part, the character recognition part accuracy rate still needs to improve.

**Keywords: vehicle license plate recognition, OpenCV, computer vision**

# 第1章 前言

## 1.1车牌识别介绍

随着大数据与人工智能领域的发展，计算机视觉技术逐步得到应用，车牌识别作为其中一个炙手可热的应用项目，天生拥有广阔的市场与发展前景，从而吸引了大量企业从事相关研发。

车牌识别一般指车牌识别系统（vehicle license plate recognition，VLPR），即通过摄像头采集车辆信息，通过对视频图像的处理得到车牌信息，从而达到识别车辆的目的，是计算机视觉在图像处理识别方面的一种应用。

## 1.2车牌识别应用领域

作为使用最为广泛的交通工具，汽车的身影无处不在，车牌识别的应用同样无处不在，以下例举了几个常见的应用。

不停车电子收费系统（,Electronic Toll Collection，ETC），通过在高速公路出入口安装ETC系统大大加快了车辆通行速度，有效缓解高速公路收费处的拥堵问题，节省大量人力，并能通过系统记录行程信息，便于随时查证。

停车场出入管理系统，传统停车场出入由人工给卡并人工计算停车时间，不仅耗费人力，速度还很慢。应用停车出入管理系统自动识别车辆，配合缴费系统便可以实现无人操作。

超速违章处罚，通过在高速路安装超速监测系统，能够自动监测超速车辆并识别车辆信息，从而加大威慑力减少违法行为。通过遍布的摄像头能够迅速定位犯罪分子的车辆位置加快破案效率。

总之，车牌识别已经成为人类离不开的一项技术，使得车辆的监督管理变得极为便利，因此提高车牌识别的准确率，加快识别速度，降低成本等变得极为重要，虽然这一领域的实际应用已经比较成熟，但依然存在较大的进步空间，依旧存在很大的研究价值，本设计便是本着这样的态度进行研究希望能够理清车牌识别的逻辑并寻求突破。

## 1.3实现方法简介

车牌识别系统的实现主要是两个方面，车牌检测与字符识别。车牌检测即车牌图像定位，在一副完整图像中找到车牌位置并进行分割。车牌检测部分即对分割得到的车牌部分进行字符识别得到车牌号。

车牌图像定位是车牌识别中至关重要的一步，由于字符识别的输入正好是车牌图像定位的输出结果，所以车牌图像定位的准确与否直接关系到最后结果的正确与否，并且需要考虑到周围环境的影响。车牌图像定位主要用到图像变换与图像形态学处理等知识进行处理。对图像进行预处理实现灰度变换以及滤波除去噪声等操作，利用车牌部分明显的方形轮廓作为特征从而实现车牌位置的查找。

车牌识别部分使用K最邻近（K-NearestNeighbor）分类算法，为数据挖掘分类算法中的一种，通过先验知识提取车牌中各个字符特征记录与XML文件中，在识别时提取上一步车牌图像定位得到切割图像的特征进行特征对比，选取与模板特征最邻近的特征群所对应的字符为结果字符。

## 1.4 开发环境

本设计采用的开发工具为Visual Studio2012，以下简称VS。VS是微软公司推出的开发工具集，所编写的代码适用于Windows所有平台，具有极其完善的功能，包括了整个软件生命周期中所需要的大部分工具。

在Visual Studio2012环境下配置OpenCV3.0的主要步奏包括：

1. OpenCV库的下载与安装
2. Windows环境变量的添加
3. Visual Studio调用文件链接的添加

# 第2章 方案论证

## 2.1 车牌定位方案论证

### 2.1.1 基于车牌颜色分布的定位方法

由于汽车车牌颜色对比度很高且颜色相对固定，例如普通车牌大多是白底黑字或者蓝底黄字，车牌与字符，车牌与车声的颜色差异较大，即可以采用区域生长法的方法对彩色图像进行分割。

获取的汽车图像大多是RGB格式，处理起来比较麻烦，所以需要将图片从RGB空间转换到HSV空间。在HSV空间中只需要色度（Hue）和饱和度（Saturation）两个分量便可以对车牌颜色分布进行计算。基于车牌颜色分布的定位流程如图2.1所示。



图2.1 基于车牌颜色分布定位流程

此方法减少了车牌的漏检情况，在车牌倾斜情况下也有很高的检测准确度，提高了检测的准确性与稳定性。但是不适合极端天气下的检测，在车牌与背景对比度较低的情况下检测的准确性与速度都有所下降，而且由于车牌的颜色不一致，比如小轿车是蓝色背景，货车是黄色背景，除此之外还有很多其它不同的车牌，所以情况较多加大了检测的难度。

### 2.1.2 基于Radon变换的车牌定位方法

Radon变换在数学上指一种积分变换，在图像处理领域主要用于对倾斜的图象进行复原。现实生活中车牌都是矩形的，拥有固定不变的形状，边缘检测后车牌的长宽相对集中，其中长宽之间的距离相对固定，边缘都由一定比例的线段组成。但由于拍摄角度的原因，导致成像后的车牌发生变形扭曲，并非得到正面严格垂直角度的图像，与预定的特征不符，使用Radon变换算法首先将图像在各个方向进行投影得到投影后的图像信息，通过对投影后的图像做特定的分析得到原始图像的偏移角度，然后通过对图像进行旋转，拉伸，变形等操作对偏移角度进行纠正得到矩形或者近似矩形的车牌图像。基于Radon算法的车牌角度纠正定位方法流程如图2.2所示。



图2.2 基于Radon变换的车牌定位流程图

### 2.1.3 基于边缘检测的车牌定位方法

边缘指图像中色彩变化较大的分界部分，灰度变换之后的图像边缘即为灰度值急剧变化的像素区域，边缘是图像的基本特征，车牌图像更是明显，车牌的字符与车牌背景之间，车牌与车声部分都存在明显的边缘，所以边缘检测是车牌检测中最常用的方法之一，在车牌图像定位中占据重要的地位。边缘检测的实现方法是通过使用某种算子来提取图像中的物体与边缘交界部分，常见的算子有sobel算子，拉普拉斯边缘检测算子，prewitt算子，roberts算子等。

边缘检测的基本思想是基于边缘部分与背景具有较大的对比度，即差分值较大，通过检测发现当前点在临近区域为局部最大值且差值大于所规定的阈值，则判定当前点为边界点，众多点连在一起便形成了物体边缘的轮廓。车牌边缘检测便是通过这样的原理找到车牌位置，实现定位。基于边缘检测的车牌定位流程如图2.3所示。.



图2.3 基于边缘检测的车牌定位流程图

此方法具有检测速度快，精准度与效率高，且对噪声有很好的适应能力等优点，所以适用范围广，对于一张图片中有多个车牌的情况有也有很好的效果。但是对于车牌破损，边缘有污渍的车辆检测效果大大下降，对于背景复杂的情况，可能存在有类似于车牌的边缘，错把一些具有类似性质的其他标志物当做车牌导致定位错误。

## 2.2 字符识别方案论证

### 2.2.1 基于神经网络的字符识别

作为机器学习中的热门算法，神经网络模拟人类大脑的工作原理，以神经元为基本单位建立复杂的神经网络系统。人工神经网络结构如图2.4所示，由细胞体，树突，轴突构成的细胞体组成。

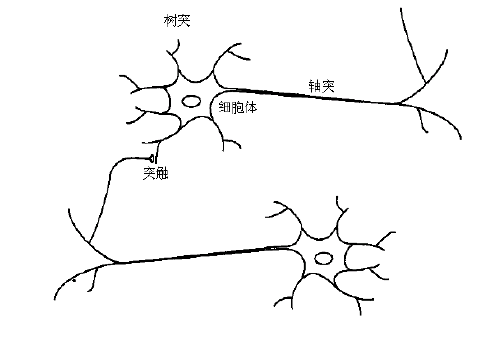


图2.4 神经网络结构

抽象后的神经元数学模型如图2.5所示，输入信号P经过w的加权后进过累加器与b相加，进过函数f的处理最终输出a。图2.5表示的为单层神经网络模型，实际运用中常常使用单层神经网络的级联形式即多层神经网络模型。

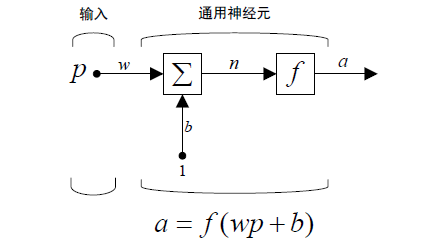


图2.5 神经元数学模型

多层神经网络分为三层，输入层，隐含层，输出层。通常来说，隐含层层数越多，神经网络的分析能力越强，训练的数据也更多，多层神经网络模型如图2.6所示。神经网络的训练即为对模型中各个参数的不断修正的过程，首先批量输入学习样本并对输入量进行归一化处理，并对最大训练次数，学习精度，隐含层节点数，初始权值，阈值，学习速率进行初始化，然后计算各层已经最后的输出情况，计算输出层误差从而作为反馈对参数进行调整直到误差减小到允许的范围。

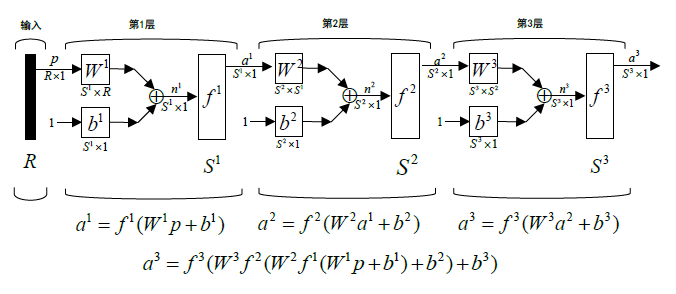


图2.6 多层神经网络模型

### 2.2.2 基于KNN最邻近算法的字符识别

KNN（K-Nearest Neighbor）属于非参数方法，在非参数估计（nonparametric estimation）中，我们只假设相似的输入具有相似的输出，这是一种合理的假设：世界是平稳的，并且无论密度，判别式还是回归函数都是缓慢地变化[2]。KNN算法则是这种思想的具体体现，通过计算判定样本与训练样本集的距离，在指定距离内涵盖最多样本所属的类别即为需要判定样本所属类别，如图2.7所示选取k值，计算以欧氏距离k为半径的圆内其他类别的个数，图中中心小红点以k为半径的圆内三角形个数最多，则判定中心小红点为三角形。



图2.7 KNN示意图

KNN算法为惰性算法，不需要和神经网络一样训练参数得到判别函数，而是直接将训练集记录下来在判别的时候依次比较，所以此算法的训练成本较低，但是在数据量很大的时候使用knn算法有很大的计算量，所以一般用在训练量不大的地方。KNN算法可以归纳为以下几步：

1. 选择邻近的数量K和距离度量方法[1]。
2. 找到待分类样本的K个最近邻居[1]。
3. 根据最邻近的类标进行多数投票[1]。

## 2.3 方案比较与选择

通过对以上主流车牌方法识别的研究学习，发现各种方法均有相应的优点与缺点，考虑到车牌普遍拥有一样的方形结构并且车牌内有比较规律的字符，所以最终决定采用基于轮廓提取的定位方法查找车牌的位置。本次车牌识别主要对象限于只有字母与数字的车牌，所以仅有26个字母与10个数字，也就是在36个类别中进行匹配，所以类别并不是很多，若采用神经网络的方法进行识别会加大训练难度，所以最终选择了KNN算法实现车牌字符识别。综上，最终选择的方案为轮廓提取+K-邻近算法。

# 第3章 OpenCV介绍

## 3.1 OpenCV简介

OpenCV(Open Source Computer Vision Library)是一个开源的计算机视觉与机器学习库。其建立的目的是为了提供一个通用的计算机视觉库与加速机器学习在通用商业产品中应用。由于其使用BSD产品协议（一个非常宽松的协议），使得在商业应用中可以自由地使用与修改OpenCV。最开始有因特尔公司出资赞助开发，如今由Willow Garage提供支持。

OpenCV有超过2500中可操作算法，其中包括了全套的经典与最新的计算机视觉与机器学习算法。这些算法可以用于检测，识别人脸，或者其他的物体，在视频中识别人类的行为，在相机功能中追踪移动的物体，提取物体的3D模型，操控立体视频中的点，拼割图像从而产生高分辨率的全景图像，在图像库中找到相同的图像，使用flash功能达到消除红眼的目的。OpenCV有着超过四万七千的开发者社区与超过一千四百万的下载量，被大量地用于公司，研发团队与政府部门。

伴随着许多高科技公司，例如谷歌，雅虎，微软，因特尔，IBM，宏达，Toyota对OpenCV技术人员的招聘，从而产生了许多类似Applied Minds，Zeitera，和VideoSurf之类的应用。使得OpenCV的社区更为活跃，OpenCV的应用从对街景图片的拼接到以色列国防安全的监测，到中国的矿井监测，又或者是Willow Garage公司的机器人项目，欧洲的泳池安全监测工作等等，OpenCV的应用遍布世界各处各个领域。

OpenCV同时具有C++,Python，Java，Matble的接口，支持Windows，Linux，Android以及Mac OS等平台。其应用主要依靠于实时视觉应用，必要的时候也会使用MMX以及SSE指令。全功能的CUDA和OPENCL接口正在开发之中，有超过500种算法以及10倍多的函数及组件用于支持这些算法。

## 3.2 OpenCV常用模块介绍

OpenCV的基本结构主要包括5个模块，分别为图像处理和视觉算法，统计分类器（MLL），图像和水平输入输出，基本机构与算法，CnAux。其结构关系如图3.1所示。



图3.1 OpenCV结构关系

首先，CV模块是图像处理时最常用的基本算法，用于对图像进行一般的数字形态学处理。MLL（Machine Learning Library）即统计分类器，随着大数据与计算机硬件速度的发展，机器学习也取得了重大的突破，OpenCV3.0新加入了很多机器学习算法用于图像识别与分类，这也是OpenCV更新最为频繁的一个模块。HighGUI则是针对视频图像的读取显示部分的，能够方便地读取与显示视频图像[5]。CXCORE包含了OpenCV基本的数据结构，动态数据库，绘图函数，数组操作相关函数，辅助功能与系统函数的宏[5]。结构图2.1中没有包括第五个模块CnAux，该模块的作用是包含即将淘汰的算法与结构。

从OpenCV出现至今已经更新了多个版本，本实验采用OpenCV3.0进行操作，相对于OpenCV2.X大体保留了OpenCV2经典的C++和Python编程接口风格，添加了Python3.X的支持，优化了Java的支持，并新接入了MATLAB的支持。对图像存储结构进行了调整，添加了一些新的图像处理算法与机器学习模型，新版本包括了目标检测算法，鱼眼镜头模型等全新算法。还增加了许多更高层次的高级封装，比如车牌检测，人脸检测等。相比之前新优化了更多指令集，值得一提的是不仅对Intel平台的处理器指令进行了优化，还增加了对ARM指令集的支持，使得在嵌入式设备进行处理时速度有了一定的提升。

### 3.2.1 CV模块

### 3.2.2 MLL统计分类器

### 3.2.3 HIGH GUI

# 第4章 车牌位置检测

## 4.1 图像预处理

图像预处理是车牌提取中的重要环节，通过输入设备输入的通常为多通道的彩色图像，数据量过大不利于图像的处理，图像的预处理包括图像格式空间转换，图像的二值化，图像的滤波。

### RGB空间与HSV空间的转换

RGB色彩模式主要面向硬件描述，是基于三原色原理的一种描述方式，通过红，黄，蓝三种颜色按不同比例混合可以合成各式各样的颜色，实际应用中输入设备采集的图像正是这种格式，但不便于人感官的辨认，比如我们看一种颜色，并不知道它由什么什么比例的三原色混合得到。HSV(hue，saturation，value)色彩空间则是面向用户 的一种颜色空间，也称作六角椎体模型，由色调H，饱和度S，前度V组成。颜色模型如图3.1所示。

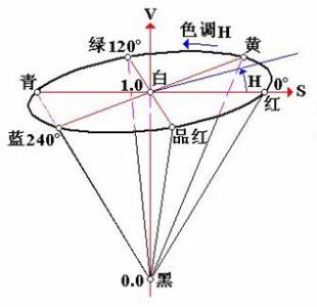


图3.1 HSV颜色模型

当图像处于RGB空间时数据量太大且不利于对图像的处理，转化为HSV空间后直接取V(Value)通道便得到对应的灰度图像，图像处理时主要对灰度图像进行处理，转换后的灰度图像分别如图3.2所示。



图3.2 灰度变换后的汽车图像

### 图像滤波

图像滤波指在尽量保留图片特征信息的条件下对目标的噪声进行滤除，是图像处理中重要的部分。由于灰度图像中存在很多细小的颗粒噪声，如果直接进行二值化处理不仅会增加计算量，还会提取到很多噪声信息不利于后面的轮廓提取，进行滤波处理后可以滤除细小的颗粒物与复杂背景中的细微信息，利于后续对图像的处理。

数学形态学是图像处理中最常用也是最基本的算法，此处理方法的基本思想为用具有一定形态或形状的结构元素去提取和度量图像中对应的形状，从而实现分析图片、识别图片与处理图片的目的。其数学基础为集合论，完备的数学基础是形态学用于图像分析处理，形态滤波器等领域的坚实后盾。数学形态学主要有四种基本操作：膨胀，腐蚀，开运算，闭运算。由这四种基本运算延伸出了很多新的运算，利于骨架抽取，击中与击不中变换，灰值形态学梯度，top-hat变换，颗粒分析，流域变换等。形态学运算的主要功能为：（1）图像的预处理（去除噪声，简化形状）；（2）图像增强（改善图像质量，丰富图像信息）；（3）物体检测（从背景中分割兴趣区域）；（4）物体量化描述（周长，面积，投影）。

膨胀操作使得物体轮廓向外扩张，若物体中存在细小空洞或者某些边缘有细小的断口，可以使用膨胀操作对断口进行填补[5]。腐蚀操作是膨胀的反操作，会去掉物体的边缘点，细小物体会直接被认定为边缘点而被消除。腐蚀操作将图像轮廓边缘缩小，从而达到除去孤立的细小噪声点，使得轮廓平滑[5]。

对图像进行先膨胀再腐蚀的操作称为闭运算，相反，对图像进行先腐蚀再膨胀的操作称为开运算。开运算用于消除孤立的小物体，在纤细点处分离物体，平滑物体轮廓并不明显改变其面积[5]。顶帽操作是原图像与开运算结果之差，黑帽操作为闭运算结果与开运算之差。噪声无处不在，通常来说，在噪声的影响下，二值化后的图像边缘往往并不平滑，存在着由噪声引起的断层或者小孔，灵活地使用开闭运算能够对此类情况进行改善。本文真是使用了此类方法进行滤波处理，通过顶帽与黑帽操作达到降低鼓励噪声，平滑轮廓曲线，填补细小空洞，链接邻域的目的。

高斯滤波是一种线性平滑滤波，用于消除高斯噪声，本实验中采取的滤波方式正是高斯滤波，其一般有两种实现方式，一种是通过傅里叶变换去除高频分量，另外一种是用离散化窗口滑窗卷积，最常见的是用滑窗实现，本文使用高斯函数对灰度图像进行处理从而达到滤除高频噪声的目的，对于零均值的高斯函数为：

(3-1)

### 4.1.3 图像二值化

进行轮廓提取前需要找到合适的阈值对图像二值化，二值化后图像中就只有黑色与白色两种颜色，便于进行轮廓提取，所以找到合适的阈值便显得及其重要，一般采取试探法尝试不同阈值观察效果，当车牌颜色与车牌内文字颜色能清晰分开时，便认定为合格。为了加大图像中黑白区域对比度使用图像处理形态学方法（灰度+闭运算-开运算）处理，二值化后的结果如图3.3所示。观察图像易知进过二值化处理后图像中的关键信息（车牌与字符）并未损失反而更加突出，接下来的处理便基于此二值图像，由于仅有两种颜色，所以处理起来的信息量大大降低。



图3.3 二值化后图像

## 4.2 轮廓特征提取

轮廓提取指在包含目标与背景的数字图像中，忽略图像背景，噪声等不需要的信息，对目标进行提取检测的技术。目前轮廓提取主要有两种方式，第一种利用传统的形态学方法使用边缘检测算子对边缘进行检测，另外一种方法是使用从人类视觉中提取出来的数学模型提取边缘信息，这儿使用传统的形态学方法进行边缘检测。

### 4.2.1 边缘检测

边缘检测属于计算机视觉与图像处理中的基本问题，其目的是检测图像中亮度变化明显的点，在二值化图像中即是提取白色与黑色交界处的点。通过边缘检测大幅度地剔除了与目标无关的信息，利用OpenCV中的cv::findContours函数可以轻易得获取图像的边缘信息，二值化图像进过边缘检测函数后输出图像的轮廓信息，每个轮廓为一个整体。提取轮廓后的结果如图3.4所示。



图3.4 提取轮廓信息后的图像

对比二值化图像与灰度图像容易发现，提取轮廓后的图像由线条构成，以黑色为背景，白色即为提取之后的轮廓。

### 4.2.2 轮廓筛选

由轮廓提取结果可知，轮廓提取后拥有大量的背景信息，所以需要对轮廓进行筛选去除大部分没必要的轮廓信息。通过计算轮廓的半径，周长等信息对轮廓进行筛选除去太大或者太小的轮廓从而达到筛选轮廓的目的。使用OpenCV中possibleChar.boundingRect.area()函数得到轮廓的面积，使用OpenCV中内置函数possibleChar.boundingRect.width得到轮廓的最大宽度，使用内置函数possibleChar.boundingRect.height获取轮廓最大高度，设定最小area为80，最小width为2，最小height为8，筛选后的轮廓如图3.5所示。

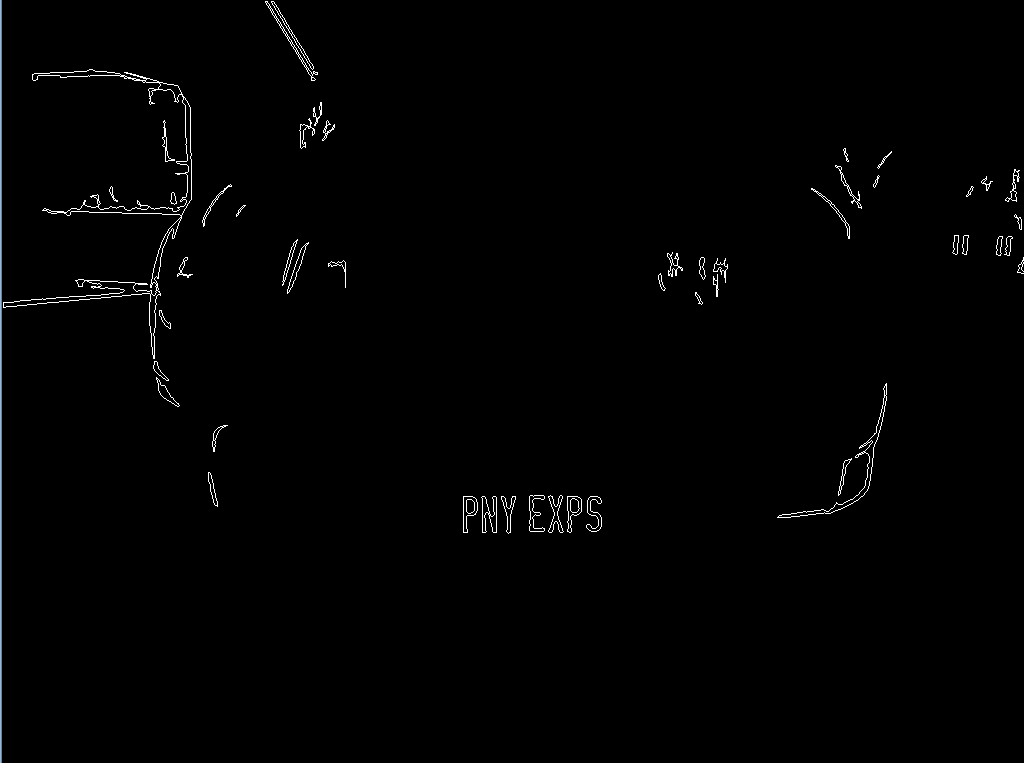


图3.5 轮廓筛选结果

进过筛选，去除了大量不相关的轮廓，并且完整保留了关键信息，大大地减少了后续计算量。

### 4.2.3 轮廓提取与分组

由图3.5可知，筛选后的轮廓依旧包含与车牌无关的部分，利用车牌字符轮廓之间的关联性可以有效提取有用轮廓与多余轮廓，车牌部分字符中心间距较窄，并且分布规律，对轮廓规律的提取流程如图3.6所示。将满足相关性的轮廓分到同一个组别，相关性的条件正是为了满足车牌的字符而量身定制，也就是说，车牌字符的轮廓都会被分到同一个类别并且成为一个整体，通过对现存轮廓的分组使得所要提取的所有信息都包含于此，后续处理便只需要针对组别进行便可。处理至此便完成了车牌检测的大部分工作，用红色的框将每个组括起来便能清楚地观察到处理后的结果，如图3.7所示。



图3.6 轮廓分组流程

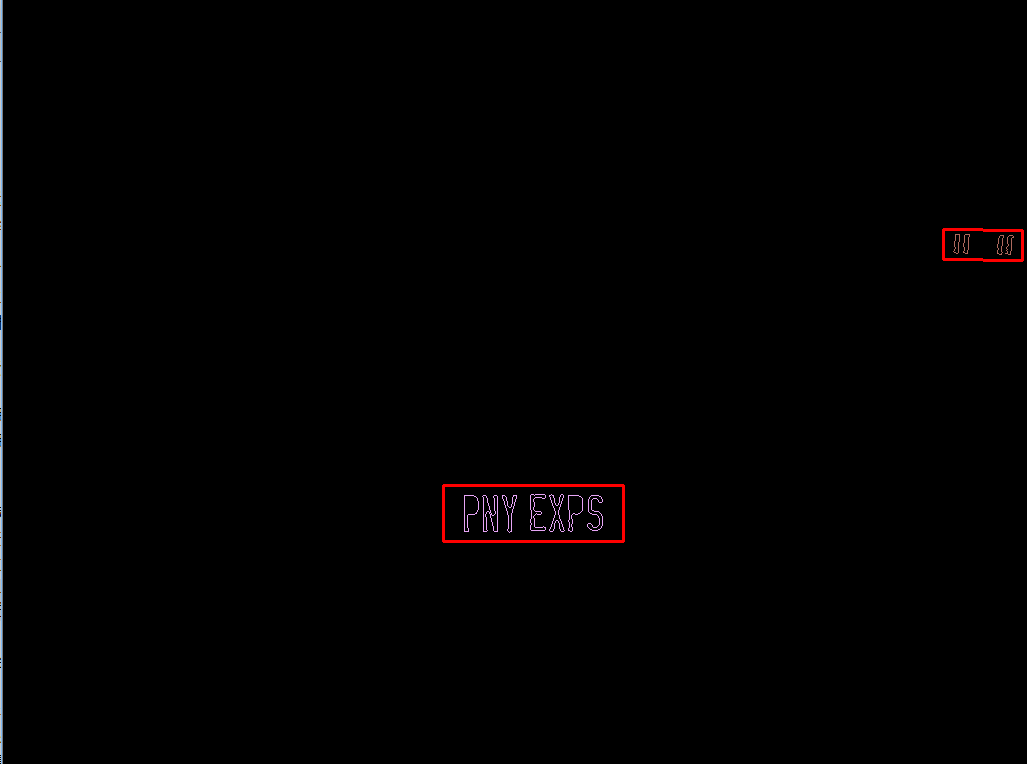


图3.7 轮廓分组筛选结果

对分组后的轮廓用红色的方框包含，计算方框的旋转角度，对不满足条件的组别直接剔除。

## 4.3 车牌切割

之前的所有处理均是在原图像上进行，目的都是为了提取图像中的车牌部分信息以便后续识别部分的处理，进过预处理与轮廓提取操作后已经成功提取出车牌部分信息，现在只需要对兴趣区域定位裁剪即可，即裁剪轮廓提取部分的红色区域。裁剪结果如图3.8所示。

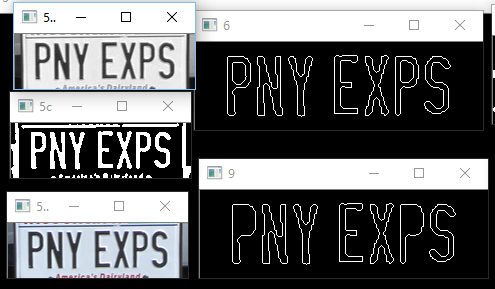


图3.8 车牌切割后结果

# 第5章 基于KNN的字符识别

## 5.1 KNN简介

所谓最邻近算法，即使给定一个训练数据集，对新的输入实例，在训练数据集中找到与该实例最邻近的K个实例，这K个实例的多数属于某个类，就把该输入实例分类到此类中。

KNN算法虽然为最简单的机器学习方法之一，但是应用得能得到极好的识别效果，本实验使用精简的最邻近算法。非参数方法的时间和空间复杂度正比于训练集的大小，目前已经提出了一些精简算法，以减少存放的实例数而不降低其性能。其基本思想是选择X的最小子集Z使得用Z代替X时，误差不增加（Dasarathy 1991）。最著名的方法是精简的最邻近（condensed nearest neighbor），其使用1-nn作为分类的非参数估计（Hart 1968）。1-nn以分段线性的方式近似判别式，并且只需要保存定义判别式的实例。类区域内部的实例不必作为它的同一类的最邻近存放，并且它的缺失不会导致（训练集上的）任何错误。这样的子集称为相容子集，并且我们希望找出最小的相容子集。

早在二十世纪Hart提出一种找到Z的贪心算法。即找到当前最佳是配置，训练集充空集开始，以空集去识别整个训练实例，每当出现一个错误时，便将错误实例添加至训练集中，直到没有错误出现为止，这样便将训练集有效地压缩并且实现误差不增加，此方法大大减小了训练样本的冗余度，尤其是在大样本的情况下大大降低了计算量提高了计算速度，训练过程如图4.1所示。

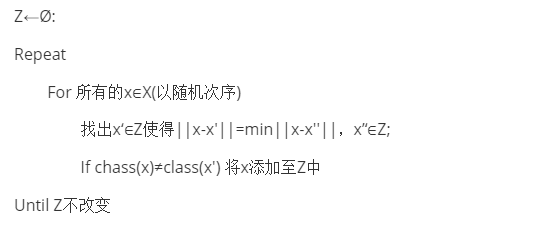


图4.1 压缩的最邻近算法

## 5.2 字符预处理

在对图片进行训练之前有必要对图像进行预处理使之满足统一的规则，从网上得到的图片各式各样，进过预处理之后的图像才能用于计算相互之间的距离。首先经过与3.1章相同的预处理得到车牌字符的二值化图像，之后进行字符分割与归一化处理。

### 5.2.1 字符分割

字符分割是车牌识别系统中的重要部分，车牌识别的时候也是对单个字符进行识别，所以字符识别的准确与否直接影响了最终的识别结果。

本实验采取计算字符轮廓大小的方法实现字符分割，直接计算字符轮廓大小然后以方形对图形进行分割。这样的方法分割的结果十分准确并且具有自动适配性，不会因为车牌的型号不同而受影响，处理之后的结果如图4.2所示。



图4.2 字符分割结果

计算出字符的轮廓之后，按照轮廓对车牌进行分割并将分割结果存入新的数组中，以便后续训练。

### 5.2.2 图像归一化

归一化是一种简化的计算方式，用以简化量纲以便统一计算，使物体系统绝对值的计算方式变为相对值的计算方式。

截取之后的字符图片大小大小各异，没法直接进行欧式距离的计算，所以需要对图像进行归一化处理，在综合考虑归一化后的速度与效率之后选择了200×300作为归一化的结果，将所有切割后的字符图像统一重新缩放至200×300的分辨率。

## 5.3 样本训练

为了减少样本个数，提高识别速度，训练过程中采取了贪心算法的思想，并将训练过后的结果数字化之后保存于xml文件中。

### 5.3.1 贪心算法

贪心算法又称贪婪算法，指在对某一问题进行求解时，只考虑适配与当前样本的最佳结果而不从整体最优解上考虑取值，所得的结果为局部最优解。精简的最邻近算法就是一种贪心算法，其目的是最小化训练误差与复杂度。可以用一个增广误差函数表示如式4-1：

E(Z|X)=E(X|Z)+λ|Z| (4-1)

其中E(X|Z)用于存放Z在X上的误差。|Z|是z的基数，第二项为对复杂度的惩罚。与其他的正则化方案相同，λ体现了误差与复杂度的折中，使对于较小的λ，误差变得更为重要，随着λ的逐渐增大，对模型的惩罚则会更大。

贪心算法的一般训练流程如图4.3所示。

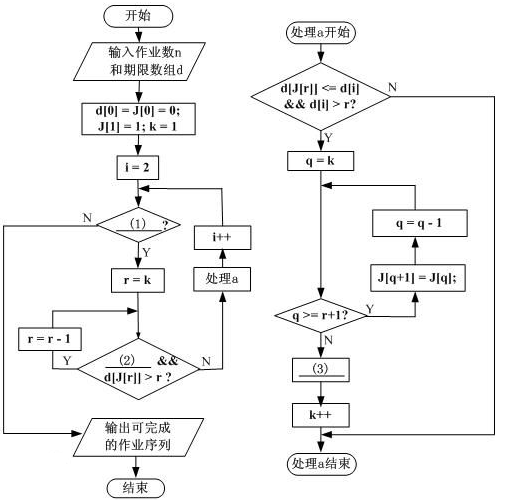


图4.3 贪心算法一般流程

### 5.3.2 训练过程

因为KNN是非参数化模型，所以其训练过程与神经网络，SVM（支持向量机），核机器等模型的训练方法不同，不用对未知参数进行训练。简单来说，作为惰性学习算法，KNN的训练过程就是对样本的存储过程，同时因为采用了贪心算法，样本的存储与筛选便成了训练过程中的首要问题。

首先从训练样本集中读出第一个样本，此时没有任何数据对样本进行识别，所以识别结果自然错误，将此样本值保存到用于识别未筛选样本的识别文件中。如此不断往复，使用识别文件中的数据对未筛选内容进行识别，若识别成功则舍弃此样本，若识别失败，将识别失败的图片数据存入识别文件中，直到达到要求的正确率为止，训练流程如图4.4所示。



图4.4 KNN训练流程

训练用的样本来自互联网，训练的时候只需要对图片就行归一化后便可以直接使用。样本如图4.5所示。

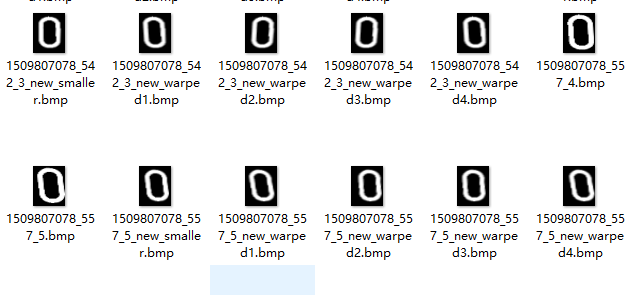


图4.5 训练样本

训练之后的数据以XML的格式保存起来，两个XML文件分别保存图片数据与数据对应的分类。本次设计以classification.xml文件保存对应的分类，images.xml与classification.xml数据一一对应，用于保存每个训练样本数据所处的分类。本次车牌识别仅仅针对只有字母与数字的车牌，所以车牌字符类别共有26个字母与10个数字一共36个类别，每个类别有5组不同的数据与之对应。判断类别的数据如图4.6所示。

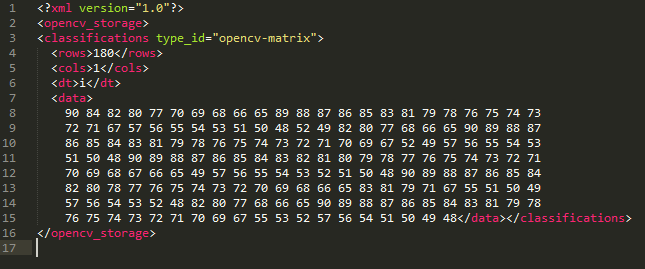


图4.6 KNN车牌字符类别数据

由图可以看出，数据均为数字，这是因为所有的类别均用相应的ASCII码表示，字符0-9对应着ASCII码的48-57，字符A-Z对应着ASCII码的65-90。这就是上图中数字所表示的意义。上面说到，每一个分类都有对应的数据，数据即是用作与待识别字符进行欧式距离计算的，距离最近的数据所对应的类别即被判定为待识别字符的类别。其部分数据如图4.7所示（数据太大只截取了开头与结尾处数据）。由图可以看出，样本图片的数据为180\*600的矩阵，180与classification.xml中的180行对应，对应着36\*5=180个类别，600则是字符图像归一化后的分辨率20\*30=600，即对应的图片数据，目前为止，KNN模型的训工作练便完成了。

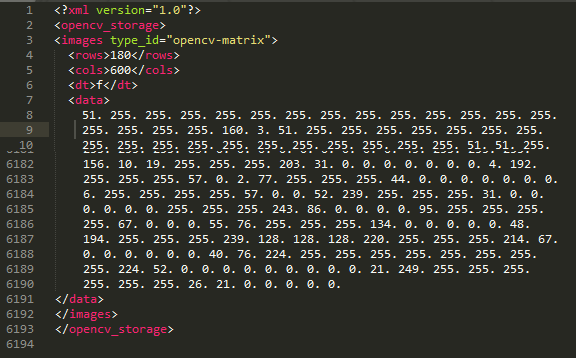


图4.7 部分字符图片数据

## 5.4 KNN算法的字符识别

关于KNN算法的字符识别过程前面有过简要的介绍，即通过计算待识别字符图片与训练样本中的已分类数据之间的距离得到相似度，根据相似度来对未分类数据进行判定，字符识别的流程如图4.8所示，识别开始时读取待识别数据，然后用待分类数据依次与已训练数据计算欧式距离，进行遍历过后比较得到最小距离，最小距离对应数据所在的分类即是我们需要得到的分类结果。OpenCV中内置的机器学习算法中直接包含了KNN算法，利用内置的KNearest->setDefaultK(1)设置K的初始值为1；使用kNearest->train()读取4.3节中训练好的样本数据；使用kNearest->findNearest()读取待识别数据并且与样本数据进行对比并返回对应类别的ASCII编码。为了将识别后的结果更清晰地显示出来，直接将字符写入原图像车牌下面进行对比。识别后的结果如图4.9所示



图4.8 识别流程图



图4.9 识别结果

# 第6章 系统功能、指标参数

## 6.1 系统功能

此系统实现了图像预处理，图像滤波，图像阈值化，图像

## 6.2 指标参数

系统的主要任务大致分为车牌位置检测与字符识别两个部分，车牌位置检测的结果直接影响字符识别的准确度。所以以下指标分为两个部分，车牌位置检测以随机的带有车辆的图片进行识别，字符识别的指标建立在车牌位置识别正确的前提下进行指标测试。

* 识别特征：车牌号码
* 输出结果：车牌号码
* 车牌倾斜角度范围：-10°-- +10°
* 车牌位置检测准确率：100%
* 字符识别准确率：98.7%

# 第7章 总结与展望

随着计算机芯片的快速发展，计算机的处理速度越来越快，许多之前受限于硬件的技术得到快速的发展。计算机视觉便是这样一个领域，在大数据的背景下机器学习得以发展，车牌识别由于巨大的市场需求成为最早落地的项目之一。本次设计的实现过程中通过互联网了解了很多相关知识，收获颇丰，虽然过程中遇到了很多问题，例如OpenCV库的不熟悉，图像处理相关知识的匮乏，C++语法及思想的不充分理解，机器学习相关的知识更是接触甚少。图书馆丰富的藏书量与互联网的便利很大程度地解决了这些问题，当静心潜力地投入到对这个领域的探索中时收获的知识与快乐都是不可言喻的。

本次设计针对目前计算机视觉领域的热门问题车牌识别给出了边缘检测+KNN最邻近算法的识别方案，旨在提出一种快捷简单的方法实现车牌识别的工作。整个项目的开发环境为Visual Studio2012，使用开源的计算机视觉库OpenCV。结合前人经验与个人的实践成功实现了车牌的识别工作。

首先，使用图像形态学领域的知识结合OpenCV便利的接口函数，通过图像空间变换，灰度变换，滤波，阈值化，轮廓提取等操作定位车牌位置，确定位置后切割车牌，则车牌位置检测部分工作基本完成；字符识别使用KNN算法，作为惰性算法，KNN的训练过程即使对训练样本的加工与分类存储，整个过程包括字符预处理，字符分割，图像归一化，计算欧式距离得到待分类字符所属类别，基本完成了预定的目标，实现的车牌字符的提取。

该系统在车牌识别中通用性强，全程使用C++进行开发，可以方便地移植到移动设备与嵌入式系统上工作。本次设计虽然完成了车牌识别的工作，但是整个过程中依然存在很多不足的地方有待提高。

1. 对于车牌角度纠正的问题没有给出合适的方案，所以对于倾斜的车辆图片不能很好地进行识别，此处还有较大的改进空间。
2. 系统对于复杂背景与复杂环境的抗干扰能力不足，当图像背景复杂（比如有类似车牌的标志，多个车牌的情况）时，识别效果并不理想。
3. 在夜晚或是光线不均匀的环境中，阈值化的效果不佳，造成车牌定位错误的情况，或是由于反光导致识别过程中分类错误，不能有效地提取到车牌字符。
4. 识别的准确率与实时性有待进一步地提高，车牌识别的实时性至关重要，正确识别车牌只是最基本的要求，当输入设备的分辨率较高，输入图像的实时性要求高的时候，系统的识别能力下降，有卡顿的现象发生，算法依然需要完善。

通过此次毕业设计我收获到了很多新的知识，也更清楚了不足的地方，需要改进的地方还有很多，路漫漫其修远兮，吾将上下而求索。

# 致谢

在熬过了漫长的两个月后，毕业论文终于接近尾声，毕业设计能够正常完工离不开老师同学们的默默支持，在这里首先要对我的指导老师卿曹进教授表示诚挚的感谢，卿老师为毕业设计的方向与实施方案给出了指导性的意见。以及CSDN论坛上的博主们，你们的博客解决了我遇到的很多难题。

同时，我还要感谢微软公司，感谢你们开发出visual studio这么好用的IDE以及Microsoft Word，没有这些我是完成不了代码的调试以及论文的编写工作；感谢Willow Garage一直以来对OpenCV的开发维护，使得我们能免费用上如此好用的计算机视觉库，没有你们的支持，此次毕业设计就要更换题目了；感谢Image-Net网站维护人员十多年来的搜集素材努力，是你们为机器学期提供了大量的训练样本，极大地减少了此次论文的工作量；感谢学校及时地装上了空调，使得同学们能在炎炎夏日睡个好觉，否则是不会有如此精力持续地接收毕业论文的摧残；感谢学校二食堂年复一年坚持不懈地做出难以下咽的饭菜，让我时刻体验到生活的艰辛与不易，给了我们不断奋斗的强大动力，让我们明白吃得苦中苦方位人上人的祖传真理；感谢少年JUMP一直以来推出的优秀漫画，让我能够在论文的重压下找到一处喘息的净土，重新发现生活的美好之处；感谢篮球，好的身体是完成毕业设计不可缺少的重要支撑，每周两次的篮球运动使得身体一直保持着不错的状态。感谢维普论文检测系统，你精准地检测使我清晰地了解到论文的重复内容，论文查重能过你功不可没。一路走来需要感谢的人和事太多太多，怀有一颗感恩的心，才能更好地向前，小草因为感谢土地使它有了栖身之所，因此献给人间万顷碧绿；鲜花由于感恩大地对它的滋养之恩，所以呈现给人间万紫千红。感恩是一束金色的阳光，它能融化冰雪，温暖万物，传播大爱。再次感谢大家的帮助与支持。

# 参考文献

1. Sebastian Raschka.Python机器学习[M].机械工业出版社，2017.3
2. Ethem Alpaydin.机器学习导论[M].机械工业出版社，2015.11
3. 王志明.数字图像处理与分析[M].清华大学出版社，2012.2
4. 邬延辉、王小权、陈叶芳.C++程序设计教程基于案例与试验驱动[M]，2010.6
5. 杨思源. 基于OPENCV的车辆牌照识别系统研究[D]. 西安电子科技大学, 2013.
6. 滕俊 王弟林 文汉云. 基于OpenCV下的Visual C数字图像处理方法[D]. 现代计算机(专业版), 2012.
7. 刘耀军 张姗梅 冯娟. 基于有限自动机的二值图像膨胀[D]. 西北师范大学学报(自然科学版), 2012.
8. 宋佳星. 基于机器视觉的插件机定位系统的研究与应用[D]. 大连理工大学, 2014.
9. 彭晓庆. 道路交通标志图像处理[D]. 西安电子科技大学, 2010.
10. 程成. 基于电磁扫描的高密度电路板故障器件检测技术[D]. 西安电子科技大学, 2011.
11. 吴亮. 运动目标提取、阴影和鬼影检测及去除算法研究[D]. 国防科学技术大学, 2007.
12. 张鑫. 智能视频监控中运动目标的检测与跟踪[D]. 杭州电子科技大学, 2012.
13. 韩志远. 基于粒子滤波的运动行人跟踪方法的研究[D]. 苏州大学, 2012.
14. 孙红霞. 地下输油管道红外图像的分析研究[D]. 天津大学, 2004.
15. 刘峰. 基于面部特征的性别鉴别方法研究[D]. 计算机光盘软件与应用, 2013.
16. 吕联盟. 基于云计算的人脸识别系统研究与设计[D]. 长安大学, 2014.
17. 陈筱. 基于区域生长和数学形态学的MRI图像处理研究[D]. 中南民族大学, 2012.
18. 闫海霞 刘岩俊. 基于顺序形态学自适应边缘检测方法[D]. 计算机应用研究, 2011.
19. 钟华. 水火弯板机系统中大尺寸钢板三维重建过程研究[D]. 广东工业大学, 2014.
20. 苏超 彭泽荣 黄志文. 基于图像处理的绝缘子缺陷检测[D]. 广东电网有限责任公司清远供电局, 2016.
21. 莫国影. CCD图像识别技术在疲劳裂纹检测中的应用基础研究[D]. 南京航空航天大学, 2008.

# 附录一：关键源代码

// Main.cpp

// add-self-round

double round(double val)

{

return (double)val;

}

#include "Main.h"

///////////////////////////////////////////////////////////////////////////////////////

int main(void) {

bool blnKNNTrainingSuccessful = loadKNNDataAndTrainKNN(); // attempt KNN training

if (blnKNNTrainingSuccessful == false) { // if KNN training was not successful

// show error message

std::cout << std::endl << std::endl << "error: error: KNN traning was not successful" << std::endl << std::endl;

return(0); // and exit program

}

cv::Mat imgOriginalScene; // input image

imgOriginalScene = cv::imread("image9.png"); // open image

if (imgOriginalScene.empty()) { // if unable to open image

std::cout << "error: image not read from file\n\n"; // show error message on command line

\_getch(); // may have to modify this line if not using Windows

return(0); // and exit program

}

std::vector<PossiblePlate> vectorOfPossiblePlates = detectPlatesInScene(imgOriginalScene); // detect plates

vectorOfPossiblePlates = detectCharsInPlates(vectorOfPossiblePlates); // detect chars in plates

cv::imshow("imgOriginalScene", imgOriginalScene); // show scene image

if (vectorOfPossiblePlates.empty()) { // if no plates were found

std::cout << std::endl << "no license plates were detected" << std::endl; // inform user no plates were found

}

else { // else

// if we get in here vector of possible plates has at leat one plate

// sort the vector of possible plates in DESCENDING order (most number of chars to least number of chars)

std::sort(vectorOfPossiblePlates.begin(), vectorOfPossiblePlates.end(), PossiblePlate::sortDescendingByNumberOfChars);

// suppose the plate with the most recognized chars (the first plate in sorted by string length descending order) is the actual plate

PossiblePlate licPlate = vectorOfPossiblePlates.front();

cv::imshow("imgPlate", licPlate.imgPlate); // show crop of plate and threshold of plate

cv::imshow("imgThresh", licPlate.imgThresh);

if (licPlate.strChars.length() == 0) { // if no chars were found in the plate

std::cout << std::endl << "no characters were detected" << std::endl << std::endl; // show message

return(0); // and exit program

}

drawRedRectangleAroundPlate(imgOriginalScene, licPlate); // draw red rectangle around plate

std::cout << std::endl << "license plate read from image = " << licPlate.strChars << std::endl; // write license plate text to std out

std::cout << std::endl << "-----------------------------------------" << std::endl;

writeLicensePlateCharsOnImage(imgOriginalScene, licPlate); // write license plate text on the image

cv::imshow("imgOriginalScene", imgOriginalScene); // re-show scene image

cv::imwrite("imgOriginalScene.png", imgOriginalScene); // write image out to file

}

cv::waitKey(0); // hold windows open until user presses a key

return(0);

}

///////////////////////////////////////////////////////////////////////////////////////////////////

void drawRedRectangleAroundPlate(cv::Mat &imgOriginalScene, PossiblePlate &licPlate) {

cv::Point2f p2fRectPoints[4];

licPlate.rrLocationOfPlateInScene.points(p2fRectPoints); // get 4 vertices of rotated rect

for (int i = 0; i < 4; i++) { // draw 4 red lines

cv::line(imgOriginalScene, p2fRectPoints[i], p2fRectPoints[(i + 1) % 4], SCALAR\_RED, 2);

}

}

///////////////////////////////////////////////////////////////////////////////////////////////////

void writeLicensePlateCharsOnImage(cv::Mat &imgOriginalScene, PossiblePlate &licPlate) {

cv::Point ptCenterOfTextArea; // this will be the center of the area the text will be written to

cv::Point ptLowerLeftTextOrigin; // this will be the bottom left of the area that the text will be written to

int intFontFace = CV\_FONT\_HERSHEY\_SIMPLEX; // choose a plain jane font

double dblFontScale = (double)licPlate.imgPlate.rows / 30.0; // base font scale on height of plate area

int intFontThickness = (int)round(dblFontScale \* 1.5); // base font thickness on font scale

int intBaseline = 0;

cv::Size textSize = cv::getTextSize(licPlate.strChars, intFontFace, dblFontScale, intFontThickness, &intBaseline); // call getTextSize

ptCenterOfTextArea.x = (int)licPlate.rrLocationOfPlateInScene.center.x; // the horizontal location of the text area is the same as the plate

if (licPlate.rrLocationOfPlateInScene.center.y < (imgOriginalScene.rows \* 0.75)) { // if the license plate is in the upper 3/4 of the image

// write the chars in below the plate

ptCenterOfTextArea.y = (int)round(licPlate.rrLocationOfPlateInScene.center.y) + (int)round((double)licPlate.imgPlate.rows \* 1.6);

}

else { // else if the license plate is in the lower 1/4 of the image

// write the chars in above the plate

ptCenterOfTextArea.y = (int)round(licPlate.rrLocationOfPlateInScene.center.y) - (int)round((double)licPlate.imgPlate.rows \* 1.6);

}

ptLowerLeftTextOrigin.x = (int)(ptCenterOfTextArea.x - (textSize.width / 2)); // calculate the lower left origin of the text area

ptLowerLeftTextOrigin.y = (int)(ptCenterOfTextArea.y + (textSize.height / 2)); // based on the text area center, width, and height

// write the text on the image

cv::putText(imgOriginalScene, licPlate.strChars, ptLowerLeftTextOrigin, intFontFace, dblFontScale, SCALAR\_YELLOW, intFontThickness);

}

// Main.h

#ifndef MY\_MAIN // used MY\_MAIN for this include guard rather than MAIN just in case some compilers or environments #define MAIN already

#define MY\_MAIN

#include<opencv2/core/core.hpp>

#include<opencv2/highgui/highgui.hpp>

#include<opencv2/imgproc/imgproc.hpp>

#include<math.h>

#include "DetectPlates.h"

#include "PossiblePlate.h"

#include "DetectChars.h"

#include<iostream>

#include<conio.h> // may have to modify this line if not using Windows

//#define SHOW\_STEPS // un-comment or comment this line to show steps or not

// global constants ///////////////////////////////////////////////////////////////////////////////

const cv::Scalar SCALAR\_BLACK = cv::Scalar(0.0, 0.0, 0.0);

const cv::Scalar SCALAR\_WHITE = cv::Scalar(255.0, 255.0, 255.0);

const cv::Scalar SCALAR\_YELLOW = cv::Scalar(0.0, 255.0, 255.0);

const cv::Scalar SCALAR\_GREEN = cv::Scalar(0.0, 255.0, 0.0);

const cv::Scalar SCALAR\_RED = cv::Scalar(0.0, 0.0, 255.0);

// function prototypes ////////////////////////////////////////////////////////////////////////////

int main(void);

void drawRedRectangleAroundPlate(cv::Mat &imgOriginalScene, PossiblePlate &licPlate);

void writeLicensePlateCharsOnImage(cv::Mat &imgOriginalScene, PossiblePlate &licPlate);

# endif // MAIN

// PossibleChar.cpp

#include "PossibleChar.h"

///////////////////////////////////////////////////////////////////////////////////////////////////

PossibleChar::PossibleChar(std::vector<cv::Point> \_contour) {

contour = \_contour;

boundingRect = cv::boundingRect(contour);

intCenterX = (boundingRect.x + boundingRect.x + boundingRect.width) / 2;

intCenterY = (boundingRect.y + boundingRect.y + boundingRect.height) / 2;

dblDiagonalSize = sqrt(pow(boundingRect.width, 2) + pow(boundingRect.height, 2));

dblAspectRatio = (float)boundingRect.width / (float)boundingRect.height;

}

// PossibleChar.h

#ifndef POSSIBLE\_CHAR\_H

#define POSSIBLE\_CHAR\_H

#include<opencv2/core/core.hpp>

#include<opencv2/highgui/highgui.hpp>

#include<opencv2/imgproc/imgproc.hpp>

///////////////////////////////////////////////////////////////////////////////////////////////////

class PossibleChar {

public:

// member variables ///////////////////////////////////////////////////////////////////////////

std::vector<cv::Point> contour;

cv::Rect boundingRect;

int intCenterX;

int intCenterY;

double dblDiagonalSize;

double dblAspectRatio;

///////////////////////////////////////////////////////////////////////////////////////////////

static bool sortCharsLeftToRight(const PossibleChar &pcLeft, const PossibleChar & pcRight) {

return(pcLeft.intCenterX < pcRight.intCenterX);

}

///////////////////////////////////////////////////////////////////////////////////////////////

bool operator == (const PossibleChar& otherPossibleChar) const {

if (this->contour == otherPossibleChar.contour) return true;

else return false;

}

///////////////////////////////////////////////////////////////////////////////////////

bool operator != (const PossibleChar& otherPossibleChar) const {

if (this->contour != otherPossibleChar.contour) return true;

else return false;

}

// function prototypes ////////////////////////////////////////////////////////////////////////

PossibleChar(std::vector<cv::Point> \_contour);

};

#endif // POSSIBLE\_CHAR\_H

// PossiblePlate.h

#ifndef POSSIBLE\_PLATE\_H

#define POSSIBLE\_PLATE\_H

#include <string>

#include<opencv2/core/core.hpp>

#include<opencv2/highgui/highgui.hpp>

#include<opencv2/imgproc/imgproc.hpp>

///////////////////////////////////////////////////////////////////////////////////////////////////

class PossiblePlate {

public:

// member variables ///////////////////////////////////////////////////////////////////////////

cv::Mat imgPlate;

cv::Mat imgGrayscale;

cv::Mat imgThresh;

cv::RotatedRect rrLocationOfPlateInScene;

std::string strChars;

///////////////////////////////////////////////////////////////////////////////////////

static bool sortDescendingByNumberOfChars(const PossiblePlate &ppLeft, const PossiblePlate &ppRight) {

return(ppLeft.strChars.length() > ppRight.strChars.length());

}

};

#endif // end #ifndef POSSIBLE\_PLATE\_H

# 附录二：外文翻译

## 原文：

**The Introduction Of Machine Learning**

**1 Introduction**

* 1. **what is machine learning**

To solve a problem on a computer, we need an algorithm. An algorithmis a sequence of instructions that should be carried out to transform the input to output. For example, one can devise an algorithm for sorting. The input is a set of numbers and the output is their ordered list. For the same task, there may be various algorithms and we may be interested in finding the most efficient one, requiring the least number of instructions or memory or both.

For some tasks, however, we do not have an algorithm—for example,to tell spam emails from legitimate emails. We know what the input is: an email document that in the simplest case is a file of characters. We know what the output should be: a yes/no output indicating whether the message is spam or not. We do not know how to transform the input to the output. What can be considered spam changes in time and from individual to individual.

What we lack in knowledge, we make up for in data. We can easily compile thousands of example messages some of which we know to be spam and what we want is to “learn” what consititutes spam from them. In other words, we would like the computer (machine) to extract automatically the algorithm for this task. There is no need to learn to sort numbers, we already have algorithms for that; but there are many applications for which we do not have an algorithm but do have example data.

With advances in computer technology, we currently have the ability to store and process large amounts of data, as well as to access it from physically distant locations over a computer network. Most data acquisition devices are digital now and record reliable data. Think, for example, of a supermarket chain that has hundreds of stores all over a country sellingthousands of goods to millions of customers. The point of sale terminals record the details of each transaction: date, customer identification code, goods bought and their amount, total money spent, and so forth. This typically amounts to gigabytes of data every day. What the supermarket chain wants is to be able to predict who are the likely customers for a product. Again, the algorithm for this is not evident; it changes in time and by geographic location. The stored data becomes useful only when it is analyzed and turned into information that we can make use of, for example, to make predictions.

We do not know exactly which people are likely to buy this ice cream flavor, or the next book of this author, or see this new movie, or visit thiscity, or click this link. If we knew, we would not need any analysis of the data; we would just go ahead and write down the code. But because we do not, we can only collect data and hope to extract the answers to these and similar questions from data.

We do believe that there is a process that explains the data we observe. Though we do not know the details of the process underlying the generation of data—for example, consumer behavior—we know that it is not completely random. People do not go to supermarkets and buy things at random. When they buy beer, they buy chips; they buy ice cream in summer and spices for Gluhwein in winter. There are certain patterns in the data.

We may not be able to identify the process completely, but we believe we can construct a good and useful approximation. That approximation may not explain everything, but may still be able to account for some part of the data. We believe that though identifying the complete process may not be possible, we can still detect certain patterns or regularities. This is the niche of machine learning. Such patterns may help us understand the process, or we can use those patterns to make predictions: Assuming that the future, at least the near future, will not be much different from the past when the sample data was collected, the future predictions can also be expected to be right.

Application of machine learning methods to large databases is called data mining. The analogy is that a large volume of earth and raw material is extracted from a mine, which when processed leads to a small amount of very precious material; similarly, in data mining, a large volume of data is processed to construct a simple model with valuable use, for example, having high predictive accuracy. Its application areas are abundant: In addition to retail, in finance banks analyze their past data to build models to use in credit applications, fraud detection, and the stock market. In manufacturing, learning models are used for optimization, control, and troubleshooting. In medicine, learning programs are used for medical diagnosis. In telecommunications, call patterns are analyzed for network optimization and maximizing the quality of service. In science, large amounts of data in physics, astronomy, and biology can only be analyzed fast enough by computers. The World Wide Web is huge; it is constantly growing, and searching for relevant information cannot be done manually.

But machine learning is not just a database problem; it is also a part of artificial intelligence. To be intelligent, a system that is in a changing environment should have the ability to learn. If the system can learn and adapt to such changes, the system designer need not foresee and provide solutions for all possible situations.

Machine learning also helps us find solutions to many problems in vision, speech recognition, and robotics. Let us take the example of recognizing faces: This is a task we do effortlessly; every day we recognize family members and friends by looking at their faces or from their photographs, despite differences in pose, lighting, hair style, and so forth. But we do it unconsciously and are unable to explain how we do it. Because we are not able to explain our expertise, we cannot write the computer program. At the same time, we know that a face image is not just a random collection of pixels; a face has structure. It is symmetric. There are the eyes, the nose, the mouth, located in certain places on the face. Each person’s face is a pattern composed of a particular combination of these. By analyzing sample face images of a person, a learning program captures the pattern specific to that person and then recognizes by checking for this pattern in a given image. This is one example of pattern recognition.

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both.

Machine learning uses the theory of statistics in building mathematical models, because the core task is making inference from a sample. The role of computer science is twofold: First, in training, we need efficient algorithms to solve the optimization problem, as well as to store and process the massive amount of data we generally have. Second, once a model is learned, its representation and algorithmic solution for inference needs to be efficient as well. In certain applications, the efficiency of the learning or inference algorithm, namely, its space and time complexity, may be as important as its predictive accuracy.

Let us now discuss some example applications in more detail to gain more insight into the types and uses of machine learning.

* 1. **Examples of Machine Learning Applications**
     1. **Learning Associations**

In the case of retail—for example, a supermarket chain—one application of machine learning is basket analysis, which is finding associations between products bought by customers: If people who buy X typically also buy Y, and if there is a customer who buys X and does not buy Y, he or she is a potential Y customer. Once we find such customers, we can target them for cross-selling.

In finding an association rule, association rule we are interested in learning a conditional probability of the form P(Y|X) where Y is the product we would like to condition on X, which is the product or the set of products which we know that the customer has already purchased.

Let us say, going over our data, we calculate that P(chips|beer) = 0.7. Then, we can define the rule:

70 percent of customers who buy beer also buy chips.

We may want to make a distinction among customers and toward this, estimate P(Y|X,D) where D is the set of customer attributes, for example, gender, age, marital status, and so on, assuming that we have access to this information. If this is a bookseller instead of a supermarket, products can be books or authors. In the case of a Web portal, items correspond to links to Web pages, and we can estimate the links a user is likely to click and use this information to download such pages in advance for faster access.

* + 1. **Classification**

A credit is an amount of money loaned by a financial institution, for example, a bank, to be paid back with interest, generally in installments. It is important for the bank to be able to predict in advance the risk associated with a loan, which is the probability that the customer will default and not pay the whole amount back. This is both to make sure that the bank will make a profit and also to not inconvenience a customer with a loan over his or her financial capacity.

In credit scoring (Hand 1998), the bank calculates the risk given the amount of credit and the information about the customer. The information about the customer includes data we have access to and is relevant in calculating his or her financial capacity—namely, income, savings, collaterals, profession, age, past financial history, and so forth. The bank has a record of past loans containing such customer data and whether the loan was paid back or not. From this data of particular applications, the aim is to infer a general rule coding the association between a customer’s attributes and his risk. That is, the machine learning system fits a model to the past data to be able to calculate the risk for a new application and then decides to accept or refuse it accordingly.

This classification is an example of a classification problem where there are two classes: low-risk and high-risk customers. The information about a customer makes up the input to the classifier whose task is to assign the input to one of the two classes.

After training with the past data, a classification rule learned may be of the form

IF income> θ1 AND savings> θ2 THEN low-risk ELSE high-risk

for suitable values of θ1 and θ2 (see figure 1.1). This is an example of a discriminant; it is a function that separates the examples of different classes.

Having a rule like this, the main application is prediction: Once we have a rule that fits the past data, if the future is similar to the past, then we can make correct predictions for novel instances. Given a new application with a certain income and savings, we can easily decide whether it is lowrisk or high-risk.

In some cases, instead of making a 0/1 (low-risk/high-risk) type decision, we may want to calculate a probability, namely, P(Y|X), where X are the customer attributes and Y is 0 or 1 respectively for low-risk

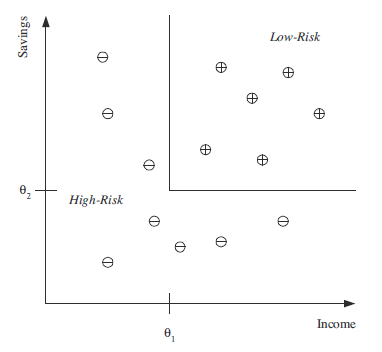


Figure 1.1 Example of a training dataset where each circle corresponds to one data instance with input values in the corresponding axes and its sign indicates the class. For simplicity, only two customer attributes, income and savings, are taken as input and the two classes are low-risk (‘+’) and high-risk (‘−’). An example discriminant that separates the two types of examples is also shown.

and high-risk. From this perspective, we can see classification as learning an association from X to Y. Then for a given X = x, if we have P(Y = 1|X = x) = 0.8, we say that the customer has an 80 percent probability of being high-risk, or equivalently a 20 percent probability of being low-risk. We then decide whether to accept or refuse the loan depending on the possible gain and loss.

There pattern are many applications of machine learning in pattern recognition . One is optical character recognition, which is recognizing character codes from their images. This is an example where there are multiple classes, as many as there are characters we would like to recognize. Especially interesting is the case when the characters are handwritten—for example, to read zip codes on envelopes or amounts on checks. People have different handwriting styles ; characters may be written small or large, slanted, with a pen or pencil, and there are many possible images corresponding to the same character. Though writing is a human invention, we do not have any system that is as accurate as a human reader. We do not have a formal description of ‘A’ that covers all ‘A’s and none of the non-‘A’s. Not having it, we take samples from writers and learn a definition of A-ness from these examples. But though we do not know what it is that makes an image an ‘A’, we are certain that all those distinct ‘A’s have something in common, which is what we want to extract from the examples. We know that a character image is not just a collection of random dots; it is a collection of strokes and has a regularity that we can capture by a learning program.

If we are reading a text, one factor we can make use of is the redundancy in human languages. A word is a sequence of characters and successive characters are not independent but are constrained by the words of the language. This has the advantage that even if we cannot recognize a character, we can still read t?e word. Such contextual dependencies may also occur in higher levels, between words and sentences, through the syntax and semantics of the language. There are machine learning algorithms to learn sequences and model such dependencies.

In the case of face recognition, the input is an image, the classes are people to be recognized, and the learning program should learn to associate the face images to identities. This problem is more difficult than optical character recognition because there are more classes, input image is larger, and a face is three-dimensional and differences in pose and lighting cause significant changes in the image. There may also be occlusion of certain inputs; for example, glasses may hide the eyes and eyebrows, and a beard may hide the chin.

In medical diagnosis, the inputs are the relevant information we have about the patient and the classes are the illnesses. The inputs contain the patient’s age, gender, past medical history, and current symptoms. Some tests may not have been applied to the patient, and thus these inputs would be missing. Tests take time, may be costly, and may inconvience the patient so we do not want to apply them unless we believe that they will give us valuable information. In the case of a medical diagnosis, a wrong decision may lead to a wrong or no treatment, and in cases of doubt it is preferable that the classifier reject and defer decision to a human expert.

In speech recognition, the input is acoustic and the classes are words that can be uttered. This time the association to be learned is from an acoustic signal to a word of some language. Different people, because of differences in age, gender, or accent, pronounce the same word differently, which makes this task rather difficult. Another difference of speech is that the input is temporal; words are uttered in time as a sequence of speech phonemes and some words are longer than others.

Acoustic information only helps up to a certain point, and as in optical character recognition, the integration of a “language model” is critical in speech recognition, and the best way to come up with a language model is again by learning it from some large corpus of example data. The applications of machine learning to natural language processing is constantly increasing. Spam filtering is one where spam generators on one side and filters on the other side keep finding more and more ingenious ways to outdo each other. Perhaps the most impressive would be machine translation. After decades of research on hand-coded translation rules, it has become apparent recently that the most promising way is to provide a very large number of example pairs of translated texts and have a program figure out automatically the rules to map one string of characters to another.

Biometrics is recognition or authentication of people using their physiological and/or behavioral characteristics that requires an integration of inputs from different modalities. Examples of physiological characteristics are images of the face, fingerprint, iris, and palm; examples of behavioral characteristics are dynamics of signature, voice, gait, and key stroke. As opposed to the usual identification procedures—photo, printed signature, or password—when there are many different (uncorrelated) inputs, forgeries (spoofing) would be more difficult and the system would be more accurate, hopefully without too much inconvenience to the users. Machine learning is used both in the separate recognizers for these different modalities and in the combination of their decisions to get an overall accept/reject decision, taking into account how reliable these different sources are.

Learning knowledge a rule from data also allows knowledge extraction. The rule is a simple model that explains the data, and looking at this model we have an explanation about the process underlying the data. For example, once we learn the discriminant separating low-risk and high-risk customers, we have the knowledge of the properties of low-risk customers. We can then use this information to target potential low-risk customers more efficiently, for example, through advertising.

Learning also performs compression in that by fitting a rule to the data, we get an explanation that is simpler than the data, requiring less memory to store and less computation to process. Once you have the rules of addition, you do not need to remember the sum of every possible pair of numbers.

Another use of machine learning outlier detection is outlier detection, which is finding the instances that do not obey the rule and are exceptions. In this case, after learning the rule, we are not interested in the rule but the exceptions not covered by the rule, which may imply anomalies requiring attention— for example, fraud.

## 翻译：

**机器学习导论**

* 1. 什么是机器学习

为了在计算机上解决问题，我们需要算法。算法是指令的序列，它把输入变换成输出。例如，我们可以为排序设计一个算法，输入是数的集合，而输出是它们的有序列表。对于相 同的任务可能存在不同的算法，而我们感兴趣的是如何找到需要的指令或内存最少，或者二 者都最少的最有效算法。

然而，对于某些任务，我们没有算法；例如，我们没有将垃圾邮件与正常邮件分开的算 法。我们知道输人是邮件文档，最简单的情况是一份字符文件。我还知道输出应该是指出消 息是否为垃圾邮件的“是”或“否”，但是我们不知道如何把这种输人变换成输出。所谓的垃 圾邮件随时间而变，因人而异。

我们缺乏的是知识，作为补偿我们有数据。我们可以很容易地编辑数以千计的实例消息，其中一些我们知道是垃圾邮件，而我们要做到的是希望从中“学习”垃圾邮件的结构。换言之，我们希望计算机(机器）自动地为这一任务提取算法。不需要学习如何将数排序，因为我们已经 有这样的算法；但是，对于许多应用而言，我们确实没有算法，而是有实例数据。

随着计算机技术的发展，我们现在已经拥有存储和处理海量数据以及通过计算机网络从 远程站点访问数据的能力。目前大多数的数据存取设备都是数字设备，记录的数据也很可rr 靠。以一家连锁超市为例，它拥有遍布全国各地的数百家分店，并且在为数百万顾客提供数 千种商品的零售服务。销售点的终端设备记录每笔交易的详细资料，包括日期、顾客识别 码、购买商品和数量、消费总额等。这是典型的每日几个g字节的数据。连锁超市希望能 够预测某种产品可能的顾客。对于这一任务，算法同样并非是显然的；它随时间而变，因地 域而异。只有分析这些数据，并且将它转换为可以利用的信息时，这些存储的数据才能变得 有用，例如做预测。

我们并不确切地知道哪些人倾向于购买这种口味的冰淇淋，或者这位作家的下一本书是什 么，也不知道谁喜欢看这部新电影、访问这座城市，或点击这一链接。我们不能确切地知道哪 些人比较倾向于购买哪种特定的商品，也不知道应该向喜欢读海明威作品的人推荐哪位作者。如果我们知道，我们就不需要任何数据分析；我们只管供货并记录下编码就可以了。但是，正因为我们不知道，所以才只能收集数据，并期望从数据中提取这些问题或相似问题的答案。

我们确信存在某种过程，可以解释我们所观测到的数据。尽管我们不清楚数据产生过程 (例如顾客行为）的细节，但是，我们知道数据产生不是完全随机的。人们并不是去超市随 机购买商品。当人们买啤酒时，也会买薯片；夏天买冰淇淋，而冬天则为Gluhweine买香料。数据中存在确定的模式。

我们也许不能够完全识别该过程，但是我们相信，我们能够构造一个好的并且有用的近 似（good and useful approximation)。尽管这样的近似还不可能解释一切，何其仍然可以解释 数据的某些部分。我们相信，尽管识别全部过程也许是不可能的，但是我们仍然能够发现某 些模式或规律。这正是机器学习的定位。这些模式可以帮助我们理解该过程，或者我们可以 使用这些模式进行预测：假定将来，至少是不远的将来，情况不会与收集样本数据时有很大 的不同，则未来的预测也将有望是正确的。

机器学习方法在大型数据库中的应用称为数据挖掘（data mining)。类似的情况如大量的 金属氧化物以及原料从矿山中开采出来，处理后产生少量非常珍贵的物质。同样地，在数据 挖掘中，需要处理大量的数据以构建简单有用的模型，例如具有高精度的预测模型。数据挖 掘的应用领域非常广泛：除零售业以外，在金融业，银行分析他们的历史数据，构建用于信 用分析、诈骗检测、股票市场等方面的应用模型；在制造业，学习模型可以用于优化、控制 以及故障检测等；在医学领域，学习程序可以用于医疗诊断等；在电信领域，通话模式的分 析可用于网络优化和提高服务质量；在科学研究领域，比如物理学、天文学以及生物学的大 量数据只有用计算机才可能得到足够快的分析。万维网（World Wide Web)是巨大的，并且 在不断地增长，因此在万维网上检索相关信息不可能依靠人工完成。

然而，机器学习不仅仅是数据库方面的问题，它也是人工智能的组成部分。为了智能 化，处于变化环境中的系统必须具备学习的能力。如果系统能够学习并且适应这些变化，那 么系统的设计者就不必预见所有的情况，并为它们提供解决方案了。

机器学习还可以帮助我们解决视觉、语音识别以及机器人方面的许多问题。以人脸识别 问题为例：我们做这件事毫不费力；即使姿势、光线、发型等不同，我们每天还是可以通过 看真实的面孔或其照片来认出我们的家人和朋友。但是我们做这件事是下意识的，而且无法 解释我们是如何做的。因为我们不能够解释我们所具备的这种技能，我们也就不可能编写相 应的计算机程序。但是我们知道，脸部图像并非只是像素点的随机组合；人脸是有结构的、 对称的。脸上有眼睛、鼻子和嘴巴，并且它们都位于脸的特定部位。每个人的脸都有各自的 眼睛、鼻子和嘴巴的特定组合模式。通过分析一个人脸部图像的多个样本，学习程序可以捕 捉到那个人特有的模式，然后在所给的图像中检测这种模式，从而进行辨认。这就是模式识别（pattern recognition )的一个例子。

机器学习使用实例数据或过去的经验训练计算机，以优化某种性能标准。我们有依赖于 某些参数的模型，而学就是执行计算机程序，利用训练数据或以往经验来优化该模型的参数 的。模型可以是预测性的（predictive),用于未来的预测，或者是描述性的（descriptive),用于 从数据中获取知识，也可以二者兼备。

机器学习在构建数学模型时利用了统计学理论，因为其核心任务就是从样本中推理。计 算机科学的角色是双重的：第一，在训练时，我们需要求解优化问题以及存储和处理通常所 面对的海量数据的高效算法。第二，一旦学习得到了一个模型，它的表示和用于推理的算法 解也必须是高效的。在特定的应用中，学习或推理算法的效率，即它的空间复杂度和时间复 杂度，可能与其预测精确度同样重要。

现在，让我们更详细地讨论一些应用领域的例子，以进一步深人了解机器学习的类型和用途。

* 1. **机器学习的应用实例**
     1. **学习关联性**

在零售业，例如超市连锁店，机器学习的一个应用是购物篮分析（baska analysis)。它 的任务是发现顾客所购商品之间的关联性：如果人们在购买商品X时也通常购买商品F,而 有一名顾客购买了商品X却没有购买商品则他（或她）即是商品^的潜在顾客。一旦我们 发现这类顾客，我们就能针对他们实行打包销售策略。

为发现关联规则（association rule)，我们对学习形如P(Y|X)的条件概率感兴趣，其中 X是我们知道的顾客已经购买的商品或商品集，Y表示在条件X下可能购买的商品。

假定考察已有的数据，计算得到P(chips|beer)=0.7,那么我们就可以定义规则：

购买啤酒（beer)的顾客中有70%的人也买了薯片（chips)。

我们也许想要区分不同的顾客。针对这个问题，我们需要估计p(X|Y,D)，其中D是顾客的一组属性，如性别、年龄、婚姻状况等，这里假定我们已经得到了这些属性信息。如果是考虑书店而不是超市的销售问题，则商品就可能是书或作者等。对于Web门户网站人 口问题，项对应着到Web网页的链接，而我们可以估计用户可能点击的链接，并利用这些 信息来预先下载这些网页，以取得更快的网页存取速度。

**1.2.2分类**

信贷是金融机构（例如银行）借出的一笔钱，需要连本带息偿还，通常是分期偿还。对银行来说，重要的是能够提前预测贷款风险。这种风险是客户不履行义务和不全额还 款的可能性。既要确保银行获利，又要确保不会因提供超出客户财力的贷款而给客户带来不便。

在信用评分（credit scoring) ( Hand 1998)中，银行要计算在给定信贷额度和客户信息情 况下的风险。客户信息包括我们已经获取的数据以及与计算客户财力相关的数据，即收人、 存款、担保、职业、年龄、以往经济记录等。银行有以往贷款的记录，包括客户数据以及贷 款是否偿还。通过这类特定的申请数据，我们可以推断出一般规则，表示客户属性及其风险 性的关联性。也就是说，机器学习系统用一个模型来拟合过去的数据，以便能够对新的申请 计算风险，从而决定接受或拒绝该项申请。

这是一个分类（classification)问题的例子，这里有两个类：低风险客户和高风险客户。 客户信息作为分类器的输入（input)，分类器的任务是将输人指派到其中的一个类。

利用以往数据进行训练后，学习得到的规则可能具有如下形式

**IF income>θ1, AND savings>θ2 THEN low-risk ELSE high-risk**

其中θ1,和θ2是合适的值（参见图1-1)。这是判别式（discriminant)的一个例子，它是将不|r〇 类的样本分开的函数。

有了这样的规则，其主要用途就是预测（prediction): —旦我们拥有拟合以往数据的规 则，如果未来与过去类似，那么我们就能够对新的实例做出正确的预测。如果给定一个新的 具有特定收人（income )和存款（savings )的申请，我们就可以很容易地判断出它是低风险（low-risk）还是高风险（high-risk）了。

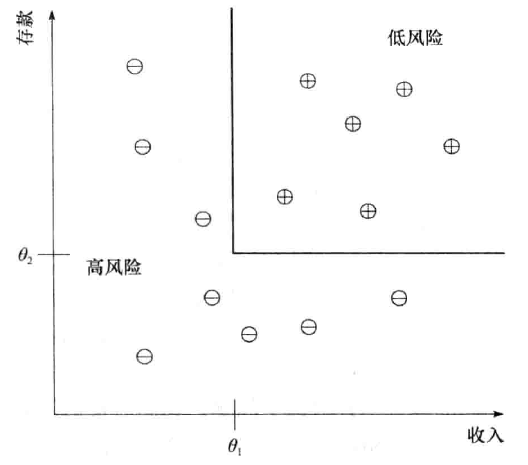


图1-1.训练数据集示例，其中每个圆圈对应一个数据实例，输人值在对应的坐标上，符号则指示 着类别。为简单起见，输人只包括客户的收人（income)和存款（savings)两种属性，两个类 分别为低风险（“ +”）和高风险（“-”）。图中还显示了分隔两类样本的判别式样例

在某些情况下，我们可能不希望做0/1(低风险/高风险）类型的判断，而是希望计算 一个概率值p(Y|X)，其中x是顾客属性，Y是0或1，分别表示低风险和高风险。从这 个角度来看，我们可以将分类看作学习从X到Y的关联性。于是，给定如果有 P(Y=1|X=x)=0.8,则我们就说该客户为高风险的可能性有80%，或者等价地说该客户为低风险的可能性有20%。然后，我们可以根据可能的收益和损失来决定接受或拒绝这笔贷款业务。

机器学习在模式识别（pattern recognition)方面有很多的应用。其中之一是光学字符识别 (optical character recognition, OCR)，即从字符图像识别字符编码。这是一个多类问题的例 子，类与我们想要识别的字符一样多。特别有趣的是手写体字符的识别问题。人们有不同的 书写风格；字体有大有小，倾斜角度不同，还有用钢笔或用铅笔之别，所以同一个字符可能 会有许多种可能的图像。尽管书写是人类的发明创造，但是我们还没有像人类读者一样准确 的系统。我们没有字符“A”的形式化描述，涵盖所有“A”而不涵盖任何非“A”。没有这种形 式化描述，我们就要从书写者那里取样，从这些实例中学习关于“A”的定义。然而，尽管我 们不知道是什么因素使得一个图像被识别为“A”，但是我们确信所有这些不同的“A”的图像 都具有某些共同的特征，这正是我们希望从实例中提取的。我们知道，图像不只是随机点的 集合，它是笔画的集合，并且是有规律的，通过学习程序我们能够補获这些规律。

阅读文本时，我们能够利用的一个因素是人类语言的冗余性。同是字符的序列 (sequence),并且相继的符号不是独立的，而是被语言的词所约束。这有好处，即便有一个 符号不能识别，我们仍可以读出词t?ee。根据语言的语法和语义，这种上下文的依赖性还可能出现在词和句子之间等较高的层次上。目前有用于学习序列和对这种依赖性建模的机器 学习算法。

对于人脸识别（face recognition),输人是人脸图像，类是需要识别的人，并且学习程序 应当学习人脸图像与身份之间的关联性。这个问题比OCR更困难，原因是人脸会有更多的 类，输人图像也更大一些，并且人脸娃三维的，不同的姿势和光线等都会导致图像的显著变 化。另外，对于特定人脸的输人也会出现问题，比如说眼镜可能会把眼睛和眉毛遮住，胡子 可能会把下巴盖住等。

对于医学诊断（medical diagnosis),输人是关于患者的信息，而类是疾病。输人包括患 者的年龄、性别、既往病史、目前症状等。当然，患者可能还没有做过某些检查，因此这些 输入将会缺失。检查需要时间，还可能要花很多钱，而且也许还会给患者带来不便。因此， 除非我们确信检查将提供有价值的信息，否则我们将不对患者进行检查。在医学诊断的情况 下，错误的诊断结果可能会导致我们采取错误的治疗或根本不进行治疗。在不能确信诊断结 果的情况下，分类器最好还是放弃判定，而等待医学专家来做决断。

在语音识别（speech recognition)，输人是语音，类是可以读出的词汇。这里要学习的是 从语音信号到某种语言的词汇的关联性。由于年龄、性别或口音方面的差异，不同的人对于 相同词汇的读音不同，这使得语音识别问题相当困难。语音识别的另一个特点是其输人信号 是时态的（temporal),词汇作为音素的序列实时读出，而且有些词汇的读音会较长一些。

语音信息的作用有限，并且与光学字符识别一样，在语音识别中，“语言模型”的集成 是至关重要的，而且提供语言模型的最好方法仍然是从实例数据的大型语料库中学习。机器 学习在自然语言处理（natural language processing)方面的应用与日俱增。垃圾邮件过滤就是一 种应用，其中垃圾邮件的制造者为一方，过滤者为另一方，一直都在寻找越来越精巧的方法 以便超越对方，也许最吸引人的是机器翻译（machine translation)。经历了数十年手工编写翻 译规则的研究之后，最近人们认识到最有希望的方法是提供大量翻译文本实例对，并且让程 序自动地提取一个字符串映射到另一个字符串的规则。

生物测定学（biometrics)使用人的生理和行为特征来识别或认证人的身份，需要集成来 自不同形态的输人。生理特征的例子是面部图像、指纹、虹膜和手掌；行为特征的例子是签 字的力度、嗓音、步态和击键。与通常的鉴别过程（照片、印刷签名或口令）相反，会有许 多不同的（不相关的）输入，伪造（欺骗）更困难，并且系统更准确，不会对用户太不方便。 机器学习既用于针对这些不同形态而构建不同的识别器，也考虑这些不同数据源的可靠性， 用于组合它们的决策，以便得到接受或拒绝的总体决策。

从数据中学习规则也为知识抽取（knowledge extraction)提供了可能性。规则是一种解释 数据的简单模型，而观察该模型我们能得到潜在数据处理的解释。例如，一旦我们学会f区 分低风险客户和高风险客户的判别式，我们就拥有了关于低风险客户特性的知识。然后，我 们就能够利用这些知识，通过比如广告等方式，更有效地争取那些潜在的低风险客户。

机器学习还可以进行压缩（compression)。用规则拟合数据，我们能得到比数据更简单的 解释，需要的存储空间更少，处理所需要的计算更少。例如，一旦你掌握了加法规则，你就 不必记忆每对可能数字的和是多少。

机器学习的另一种用途是离群点检测（outlier detection)，即发现那些不遵守规则的例外实例。在这种情况下，学习规则之后，我们感兴趣的不是规则，而是规则未能覆盖的例外，他们可能暗示出我们需要注意的异常，如诈骗。