手写数字识别实验报告

李成扬*，胡永辉老师†

华北电力大学 控制与计算机工程学院

【摘　要】 近些年，各种模式识别方法的发展促进手写数字识别取得了长足的进步。作为计算机视觉的一项典型应用，手写数字识别成果可广泛应用于邮政编码识别、统计报表识别、考试成绩判定等领域。本实验基于随机森林、K 近邻、卷积神经网络 3 种模式识别算法对来自 MNIST 数据集的手写数字图片进行识别，实验结果显示 3 种算法的准确率都在 90% 以上, 其中卷积神经网络算法的识别准确率高达 99.17%，充分说明了模式识别方法在手写数字识别领域的有效性。

【关键词】 模式识别，随机森林，K 近邻，卷积神经网络，手写数字识别

1　引 言

手写数字识别是通过计算机技术判定人类笔迹对应的数字，有助于提高社会的数字化水平，准确高效地获得手写数字中的有效信息。MNIST 数据集是手写数字识别领域最流行的公开数据集，包含 60000 个训练样本和 10000 个测试样本。手写数字图像和标签如图1所示，其中手写数字图像是 28*28 像素的灰度图。本实验在 MNIST 数据集上展开，对比分析随机森林、K 近邻和卷积神经网络 3 种模式识别算法的识别效果。

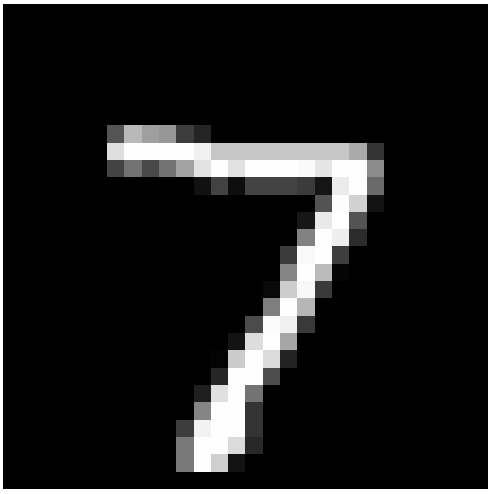

图 1 MNIST 样本示例

姓名: **李成扬**
学号: 120222227119
E-mail：chengyangli@ncepu.edu.cn
代码仓库：https://github.com/yangyangyue/pattern.git

2　随机森林

随机森林是一种用于分类和其他方法的组合回归学习方法，其采用 Bagging 算法有放回地随机采样多个数据子空间以训练若干颗决策树，并在决策树的训练过程中引入了随机属性选择机制，数据子空间和分裂属性的随机性能够避免模型过拟合，提升模型对未知数据的泛化能力。对于手写数字图像来说，其每个像素点都可以看作一个特征，所以每张手写数字图像包含 784 个特征。使用随机森林进行手写数字识别的原理就是根据若干个像素点的特征值判定该图像对应的数字。本实验中随机森林包含 10 颗决策树，如图2所示，每颗决策树由一个根节点、若干个内部节点和若干个叶结点构成，内部节点表示对一个特征的判定，叶结点表示一个类别。

在训练阶段，本实验针对每颗决策树从训练数据集中随机抽取 30% 的样本作为训练数据。在决策树分裂时，首先从所有候选特征中随机选择大小为特征数量的算术平方根的特征子集合，然后从该子集合中选择最佳分裂特征，理想的分裂结果是各个子分区纯度尽量高。本实验采用基尼指数作为度量数据集不纯度的指标，数据集 D 分裂前的基尼指数定义如式(1)所示。

$$\mathrm{Gini}(D) = 1 - \sum_{i=1}^{m} p_i^2 \tag{1}$$

式中，m 是类别数量；p_i 是随机选择一个样本属于

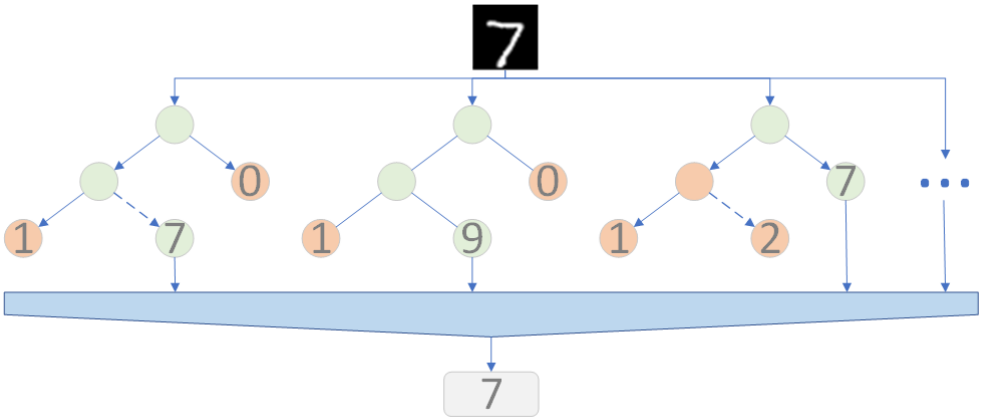

图 2 随机森林分类流程

第 i 类的概率。数据集 D 分裂为两个数据子集 D_1 和 D_2 后的基尼指数定义如式(2)所示。

$$\text{Gini}_A(D) = \frac{|D_1|}{|D|}\text{Gini}(D_1) - \frac{|D_2|}{|D|}\text{Gini}(D_2) \quad (2)$$

式中，$|D|$ 表示数据集 D 的样本数量。

在测试阶段，测试样本经过内部节点的判定最终达到叶结点，从根节点到叶子节点的特定判别路径对应于决策树的分类过程。若图2中 3 颗决策树的判别路径如绿色节点所示，则它们的分类结果依次是 7、9 和 7，所有分类结果中占比最大的结果被认定为样本最终类别，所以图2中样本被正确识别为 7。

3 K 近邻

K 近邻算法的核心思想是如果一个样本在特征空间中的 K 个最相近的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。如图3所示，应用 K 近邻算法来识别手写数字的基础是相同数字的图像是相似的，而不同数字的图像差异较大。本实验采用欧式距离衡量样本之间的差异，其定义如式(3)所示。

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{m} |x_i - y_i|^2} \quad (3)$$

式中，m 是特征数量；x_i 和 y_i 分别是样本 x 和 y 在第 i 个特征的取值。

K 近邻算法无需事先训练，而是在测试时直接基于训练数据集计算出 K 个近邻。如图3所示，本实验中首先计算每个训练样本到测试样本的距离，然后取和待分类样品距离最近的 10 个训练样例，其中占比最多的类别被认为是测试样本类别。在图3所示场景下，测试样本被正确识别为 7。

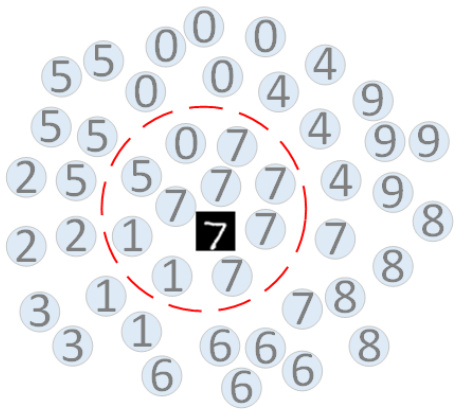

图 3 KNN 分类原理

4 卷积神经网络

相对于传统的全连接神经网络，卷积神经网络更加关注数据的局部特征，能够更好地提取手写数字图像中的局部特征。由于图像中临近的像素间常常具有强烈的空间局部性和相关性，因此使用卷积操作来提取图像中的特征比使用全连接网络更加有效。此外，卷积层在训练时的参数量相较于全连接层更少，这有助于减少训练模型的资源开销。

本实验采用的卷积神经网络结构如图4所示，网络输入的手写数字图像形状为 (1×28×28)，2 个卷积层及嵌入其中的批归一化层、激活层和最大池化层将其形状映射为 (32×14×14)，其中第一个

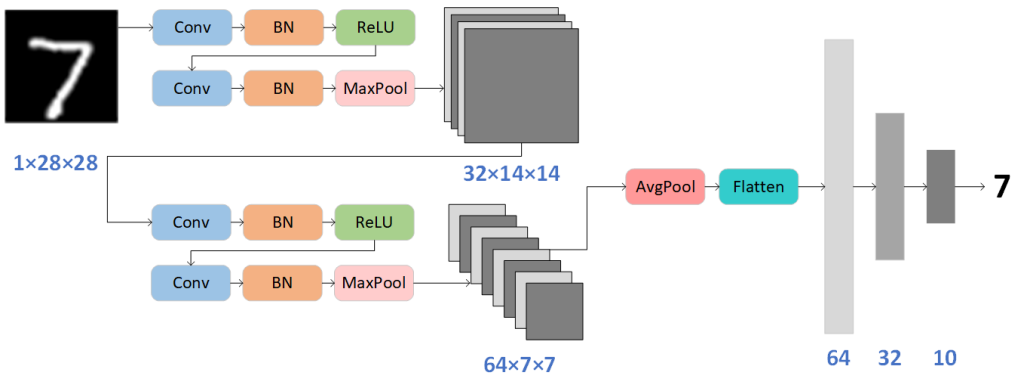

图 4 卷积神经网络模型结构

卷积层将图像通道数提升到 32，最大池化层将图像长宽缩减为原来的 $\frac{1}{2}$。随后，一个相似的结构将图像映射为 (64×7×7) 的特征图。接下来，通过平均池化层求取特征图每个通道上的特征平均值，通过 Flatten 操作得到长度 64 的特征向量。最后，连续两个全连接层将结果映射成长度为 10 的特征向量，对应于 0 到 9 共 10 个类别标签。

在训练阶段，网络的初始学习率为 $1e-4$，每 30 轮学习率降为原来的 0.1 倍，共训练 100 轮。网络中的权重参数在误差反向传播过程中更新，使用交叉熵作为损失函数计算输出和标签的误差，交叉熵损失函数定义如式(4)所示。

$$CCE = -\sum_{i=1}^{M} p(x_i) \cdot \log q(x_i) \tag{4}$$

式中，M 为类别数量，$p(x_i)$ 表示样本实际属于第 i 类的概率，$q(x_i)$ 表示模型预测样本属于第 i 类的概率。

5 实验结果及分析

本实验使用准确率、精确率、召回率和 F1 衡量分类算法的效果，准确率是分类正确的样本占总样本的比例，精确率是所有识别为正的样本中识别正确的比例，召回率是正样本中识别正确的比例，F1-Score 是综合了精确率与召回率的指标，准确率、精确率、召回率和 F1 的定义如式(5)到(8)所示。

$$Acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{5}$$

$$Pre = \frac{tp}{tp + fp} \tag{6}$$

$$Rec = \frac{tp}{tp + fn} \tag{7}$$

$$F1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \tag{8}$$

式中，tp 表示正确的正样本数量；tn 表示识别正确的负样本数量；fp 表示识别错误的负样本数量；fn 表示识别错误的正样本数量。

表 1 算法准确率指标

算法	随机森林	K 近邻	卷积神经网路
准确率	0.9381	0.9600	0.9917

注：K 近邻算法仅测试 MNIST 测试集前 100 个样本

实验结果如表1所示，从中可以看到所有算法的识别准确率都超过 90%，其中随机森林准确率最低，K 近邻算法次之，卷积神经网络准确率最高，这可以被归因为以下几点：1) 随机森林算法能够处理高维数据，且每颗决策树训练数据和分裂属性的随机性能够提升模型的泛化能力，所以能够其手写数字识别准确率能达到 90% 以上。但是，由于训练决策树时每次仅考虑一个特征，所以决策树会忽略属性间的相关性。此外，手写数字图像每个特征有 256 个可能的取值，特征级别划分过多时也会对随机森林的性能产生较大影响。2) K 近邻算法适合于类域交叉范围较大的样本集，由于相同手写数字的图像相似度较高，所以 K 近邻算法取得了 96% 的识别准确率。然而，MNIST 数据集中不同手写数字的图像也可能会有很大的相似度，这使得 K 近邻算法对相似手写数字的识别准确率降低。此外，K 近邻算法是一种懒惰算法，在测试样本时需要较大开销，这使得其实际场景下的实用性降低。3) 卷积神经网络是一种端到端的分类模型，不需要人工事先设计分类特征，加

之卷积操作能够很好地关注到图像地局部相关性，因此卷积神经网络能够提取到更加具有泛化性的特征，使得其识别准确率高达 99.17%。最后，部分手写数字即便是人类都难以区分，这部分噪声的影响使得测试准确率难以达到 100%。

6 结 论

本实验实现了随机森林、K 近邻、卷积神经网络 3 种模式识别算法，并且在 MNIST 数据集上应用它们处理手写数字识别任务。实验结果表明 3 种算法均能够取得较高的识别准确率，其中卷积神经网络算法效果最佳，这表明深度学习方法在手写数字识别等复杂领域有很大的潜力。然而，卷积操作更加关注图像中的局部特征，这可能会导致对全局特征的忽视。因此，下一步实验将结合卷积模块与自注意力模块搭建识别模型，以期能够融合局部特征和全局特征取得更好的实验结果。