

# 混合深度学习模型 C-RF 及其在手写数字识别中的应用

侯艳路 丁世飞 孙统凤

(中国矿业大学计算机科学与技术学院, 徐州, 221116)

**摘要:** 卷积神经网络(Convolutional neural network, CNN)是一种常见的深度学习模型,受人类视觉认知机制启发而来,能够从原始图像得到有效的特征表达。CNN模型在图像识别领域不断取得突破,但是在训练过程中需要花费大量时间。随机森林(Random forest, RF)在分类和回归上具有很高的精度,训练速度快并且不容易出现过拟合的问题,现有的基于RF的分类器都依赖手工选取的特征。针对以上问题,本文提出了基于CNN的C-RF模型,把CNN提取到的特征输入RF中进行分类。由于随机权值网络同样可以得到有效的结果,所以不用梯度算法调整网络参数,以免消耗大量时间。最后在MNIST数据集和Rotated MNIST数据集上进行了实验,结果表明C-RF模型的分类精度比RF有了较大的提高,同时泛化能力也有所提升。

**关键词:** 卷积神经网络;随机森林;手写数字识别

**中图分类号:** TP391.4      **文献标志码:** A

## Hybrid Deep Learning Model C-RF and Its Application in Handwritten Numeral Recognition

Hou Yanlu, Ding Shifei, Sun Tongfeng

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, 221116, China)

**Abstract:** Convolutional neural network (CNN) is a kind of common architecture of deep learning, which is inspired by the biological visual cognition mechanism. CNN can obtain the effective feature expression from the original image. In recent years, CNN has made breakthroughs in the field of image recognition, but it takes a lot of time in the training process. As a new machine learning algorithm proposed by Leo Breiman in 2001, random forest (RF) has high accuracy in classification and regression, fast training speed and is not prone to over-fitting. The existing RF based classifiers rely on hand-selected features. Aiming at the above problems, a new C-RF model based on CNN is proposed in this paper, which puts the features extracted by CNN into RF to complete the classification. Since the network using random weights can also obtain effective results, gradient algorithm is not used to adjust the network parameters for avoiding a lot of time consumption. Experimental results on the MNIST and the Rotated MNIST datasets show that the classification accuracy of C-RF model is better than that of RF, and the generalization ability is also improved at the same time.

**Key words:** convolutional neural network; random forest; handwritten numeral recognition

## 引言

手写数字识别是模式识别的一种,包含在字符识别技术中,处理一些数据信息的关键技术就是手写数字识别,比如财务报表、邮政编码和各种票据等。1998年 Lecun 等提出了手写体数字识别模型 LeNet-5<sup>[1]</sup>,曾被广泛应用于识别美国银行支票手写体数字;文献[2]中用 K 近邻分类算法(K-nearest-neighbors, KNN)在 MNIST 数据集上实现了 2.83% 的分类错误率;支持向量机(Support vector machine, SVM)和其改进算法在分类任务中得到广泛应用<sup>[3-4]</sup>,2012年 Niu 等提出了应用在数字识别上的混合模型 CNN-SVM<sup>[5]</sup>,用 CNN 进行特征提取,SVM 作为分类器,结合双方各自的优点,在图像分类任务中取得了很好的实验效果;Luo<sup>[6]</sup>等在 2014 年提出了一种混合方法 ELM-SRC(Sparse representation based classification),把 SRC 处理噪声图像的优势和 ELM 训练速度快结合起来,在 USPS 手写数据集上进行了实验,既提高了分类精度,又保证了时间效率。

卷积神经网络(Convolutional neural network, CNN)是深度学习算法的一种,广泛应用于很多领域,如目标识别、场景分类和人脸识别等。CNN 通过一层一层的学习,每层会自动从输入图像中提取到不同的特征,从低级到高级,在面对各种图像识别任务时都有很好的效果,被认为是通用图像识别系统的代表之一<sup>[7]</sup>。通常,卷积层中的神经元通过局部感受野与上层相连接,通过卷积得到该局部的特征,在池化层用池化的方法进行二次特征提取,卷积层和池化层交替出现的结构使其能够容忍一定畸变的输入样本<sup>[8]</sup>。但是 CNN 在训练的过程中需要采用 BP 算法来调整参数,这个过程会消耗大量时间。随机森林(Random forest, RF)是 Leo Breiman 在 2001 年提出的一种新的机器学习算法<sup>[9]</sup>,在分类和回归上具有很高的精度,训练速度快并且不容易出现过拟合问题,在抗噪方面也表现良好,现有的基于 RF 的分类器都依赖手工选取的特征,然而手工选取特征非常费时费力而且需要专业知识背景,是否能够取得好的效果还依赖于一定的经验和运气。文献[10]提出网络结构在随机无预训练的权值情况下也可以取得很好的效果;文献[11]使用随机初始化的未经任何训练的网络还原生成了逼真的原始图像;Maji<sup>[12]</sup>等在 2015 年提出了深度神经网络和 RF 模型用来检测眼底图像中的视网膜血管,并在 DRIVE 数据集上达到了 93.27% 的准确率。基于以上问题,文中提出了一个混合模型 C-RF。在混合模型中,用随机权值的 CNN 提取特征,然后交给 RF 完成分类,这样使得模型在提取特征的过程中大大减少了时间,既克服了 CNN 训练时间过长的问题,又解决了 RF 人工选取特征的缺陷。

## 1 基础理论

### 1.1 卷积神经网络

CNN 是深度学习广泛应用的一个模型,尤其在图片领域非常高效<sup>[13]</sup>,它是一个可以完成特征自动提取和分类的多层神经网络。CNN 的结构如图 1 所示,由输入层、交替的卷积层和池化层、输出层组成。CNN 执行的是有监督的指导训练,过程大致为以下 4 步:(1)从样本集 $(X, Y)$ 中取出  $X$ ,输入到网络中;(2)前向传输,经过每层的变换,得到最终的实际输出;(3)计算误差值,用 BP 算法求解得到误差对网络权值的导数;(4)进行权值更新。

#### 1.1.1 全连接层

##### (1) 计算输出

对于全连接的第  $l$  层的输出函数为

$$x^l = f(u^l), u^l = \mathbf{W}^l x^{l-1} + b^l \quad (1)$$

式中:  $f(\cdot)$  代表激活函数,这里采用 sigmoid 函数。

第  $n$  个训练样例的误差损失为

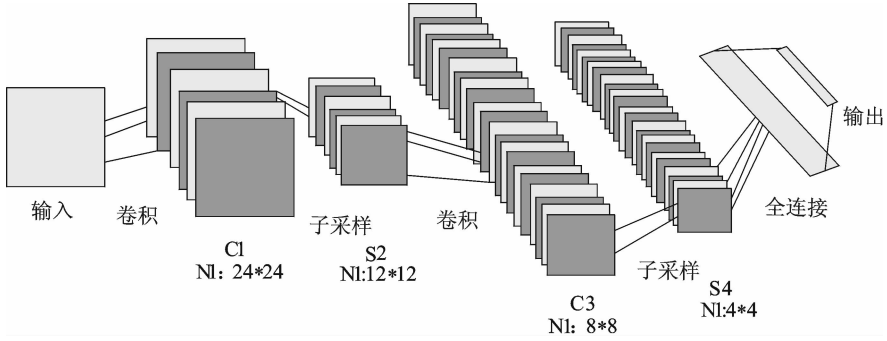


图 1 CNN 结构

Fig. 1 Structure of CNN

$$E^n = \frac{1}{2} \sum_{k=1}^c (t_k^n - y_k^n)^2 = \frac{1}{2} \|t^n - y^n\|_2^2 \quad (2)$$

(2) 权值更新

为了方便,定义灵敏度为

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial u} \frac{\partial u}{\partial b} = \delta \quad (3)$$

反向传播中第  $i$  层的灵敏度为

$$\delta^i = (\mathbf{W}^{i+1})^T \delta^{i+1} \circ f'(u^i) \quad (4)$$

对输出层的基值  $b$ ,第  $l$  层损失函数对基值  $b$  的偏导数为

$$\frac{\partial E}{\partial b^l} = \frac{\partial E}{\partial u^l} \frac{\partial u^l}{\partial b^l} = \frac{\partial E}{\partial f(u^l)} \frac{\partial f(u^l)}{\partial u} = - \sum (t - f(u^l)) f'(u^l) \quad (5)$$

对于权值  $\mathbf{W}$ ,求偏导数

$$\frac{\partial E}{\partial \mathbf{W}^l} = \frac{\partial E}{\partial u^l} \frac{\partial u^l}{\partial \mathbf{W}^l} = \frac{\partial E}{\partial u^l} x^{l-1} = x^{l-1} (\delta^l)^T \quad (6)$$

式中  $x^{l-1}$  为第  $l$  层的输入也就是上层的输出,所以最终权值的改变量为

$$\Delta \mathbf{W}^l = -\eta \frac{\partial E}{\partial \mathbf{W}^l} = -\eta x^{l-1} (\delta^l)^T \quad (7)$$

1.1.2 卷积层

在卷积层,相应的输出特征图为

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (8)$$

式中  $M_j$  表示输入特征图的集合。为了能够得到  $l$  层的灵敏度,需要对池化层的灵敏度 map 进行上采样操作,使得池化层的灵敏度 map 与卷积层的灵敏度 map 大小一致。在池化层 map 的权值都取一个  $\beta$ ,则第  $l$  层的节点  $j$  的灵敏度为

$$\delta_j^l = \beta^{l+1} (f'(u_j^l) \circ up(\delta_j^{l+1})) \quad (9)$$

式中  $up(x)$  表示对  $x$  进行上采样,具体操作要根据前面池化的方法。现在对于一个给定的 map,就可以计算得到基值  $b$  梯度为

$$\frac{\partial E}{\partial b_j} = \sum_{u,v} (\delta_j^l)_{uv} \quad (10)$$

最后,对卷积核的权值  $\mathbf{W}$  的梯度用 Matlab 的卷积函数来计算,即

$$\frac{\partial E}{\partial K_{ij}^l} = \text{rot180}(\text{conv2}(x_i^{l-1}, \text{rot180}(\delta_j^l), 'valid')) \quad (11)$$

### 1.1.3 池化层

对于池化层,不会改变输入特征图的数量,只是将输入特征图变小,有

$$x_j^i = f(\beta_j \text{down}(x_j^{i-1}) + b_j^i) \quad (12)$$

式中  $\text{down}(\cdot)$  表示一个池化函数。

当计算式(9)中卷积核的灵敏度时,这里再次借助卷积函数来得到卷积核的灵敏度,即

$$\delta_j^i = f'(u_j^i) \circ \text{conv2}(\delta_j^{i+1}, \text{rot180}(K_j^{i+1}), 'full') \quad (13)$$

这时基值  $b$  的梯度计算与卷积层一样,同式(10)。权值  $\beta$  的灵敏度为

$$\frac{\partial E}{\partial b_j} = \sum_{u,v} (\delta_j^i \circ d_j^i)_{uv} \quad (14)$$

式中  $d_j^i = \text{down}(x_j^{i-1})$ 。

## 1.2 随机森林

决策树<sup>[14]</sup>是一类常见的机器学习方法。它是一种树结构(包括二叉树和多叉树),一般由根节点、叶节点和内部节点构成。根节点包含样本全集,叶节点对应于决策结果,内部节点表示一个特征或者属性。从根节点开始根据相应的特征属性选择输出分支,直到叶节点,并将叶节点存放的类别作为分类结果,这样就可以产生一棵处理未知示例能力强的决策树。建立决策树的关键是分裂属性,即在当前状态下选择哪个属性作为分类依据,主要分两步:第一步,由样本集生成决策树;第二步,决策树的剪枝,在决策树学习中,尽可能的正确分类会造成决策树分支过多,以至于把样本自身的一些特点当作所有数据的共性而导致过拟合,所以通过主动去掉一些分支来降低过拟合的风险。

决策树算法的主要代表有 ID3<sup>[15]</sup>, C4.5<sup>[15]</sup> 和 CART<sup>[16]</sup> 等。决策树的优点很多,比如:结构简单直接,过程非常直观,易于理解和解释;决策树算法可以取得良好的分类效果;由于计算效率高,因此对于一些大型数据是可行的并且效果良好;可以很好地抑制样本中存在的噪音,也可以解决数据属性的缺失,因此可以一定程度上缓解由此带来的分类性能降低问题。主要的缺点有:对连续性的字段预测困难,效果不好;当分类类别太多时,精度下降得比较快;分类规则繁杂,一定程度上会由于过于适应噪声而导致过拟合等。

随机森林算法的实质是一个包含  $K$  个决策树的分类器,用随机方式建立一个森林,森林由决策树组成,每棵树之间没有关联。它采用 bootstrap 重抽样方法从原始样本中抽取  $n$  个样本<sup>[17]</sup>;对每个样本从所有属性中选择  $k$  个属性,选择最佳属性作为节点建立 CART 决策树;重复  $m$  次建立  $m$  棵 CART 决策树,这  $m$  棵 CART 形成随机森林,通过投票得到最终投票结果。

随机森林具有较高的分类精度,解决了决策树过拟合的缺陷,对噪声和异常点的抗干扰能力比较强;并且运行速度比较快,可以根据需要生成任意数目的树。基上以上众多优点,随机森林被广泛应用于医学研究<sup>[18]</sup>、文本分类<sup>[19]</sup> 和生物信息<sup>[20]</sup> 等领域。

## 2 混合模型 C-RF

### 2.1 C-RF 结构

混合模型 C-RF 的主要思想是用 RF 代替 CNN 原有的输出层作分类,其结构图如图 2 所示。首先用随机权值的卷积层和池化层完成对图像的特征提取,然后将提取到的图像特征输入到 RF 分类器中得到分类结果。卷积层的滤波器个数会极大地影响模型的精度和泛化能力,根据经验,C-RF 模型的 N1 和 N2 的取值分别为 10 和 20;同时 RF 中 Ntree(RF 中树的数量)值也会影响分类精度,在实验中作了不同数目 Ntree 下的分类精度来进行比较。

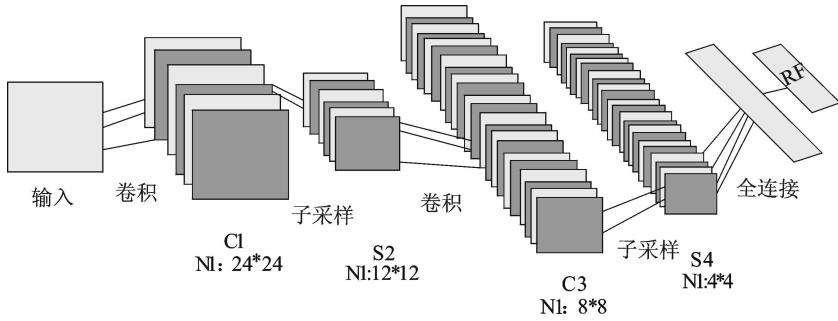


图 2 C-RF 结构

Fig. 2 Structure of C-RF

## 2.2 混合模型优点

特征提取是图像识别和分类的关键部分<sup>[21]</sup>,能否提取到反应图像属性的特征对分类精度至关重要。CNN 可以通过训练数据进行一层一层地学习,自动从输入图像中完成特征提取,避免了人工经验式的特征提取。在 CNN 中,经过简单规则化的图像直接作为网络的输入,再依次传输到网络的每一层,每层通过多个不同滤波器去计算得到不同尺度的特征。因为图像的局部感受野可以使神经元能够抽取得到一些基本的视觉特征,例如边缘或者角点等,并且每一层的神经元共享权值,每个卷积层后都跟着一个二次提取特征的池化层,这种特有的方法能够获得平移、比例缩放、倾斜和旋转不变的数据的显著特征。RF 是一种组合分类器,解决了决策树过拟合的缺陷,对噪声和异常点的抗干扰能力比较强,并且运行速度比较快,对大量数据也能保持高效性。

本文提出的混合模型 C-RF 发挥了 CNN 在特征提取方面的优势和 RF 速度快和不容易过拟合等优点,利用随机权值的 CNN 自动提取图像的特征,将提取到的特征作为 RF 分类器的输入,既避免了 CNN 在训练过程中消耗大量时间,同时又提高了 RF 的分类精度。

## 3 实验结果

为了评估混合模型 C-RF 的分类性能,作者在 MNIST 和 Rotated MNIST 数据集上进行相关的实验。实验环境:64 位 Windows 10 操作系统,4 GB RAM,主板型号为联想 To be filled by O. E. M,英特尔 Intel(R) Core(TM) i5-2400 3.10 GHz 处理器,MATLAB2010b。

### 3.1 MNIST 数据集

MNIST 数据集包含了 70 000 张手写数字图像,其中训练样本有 60 000 个,测试样本有 10 000 个,每张图像的大小为  $28 \times 28$  像素值。数据集中部分手写数字图像如图 3 所示。

假设 Ntree 为 RF 中树的数量,当 Ntree 取值过小时,分类精度就不会达到理想效果。由于 RF 不容易出现过拟合问题,因此可以使 Ntree 取值尽量大,以保证分类的精度,但与此同时会花费大量时间来构建 RF,所以 Ntree 的取值对 RF 的性能、复杂度有着重要的意义。为了避免 CNN 训练时间过长,采用了随机权值的方法,在 CNN 提取特征之后,将提取到的特征输入到 RF 中作分类,为了对比作了不同 Ntree 数目下的实验,表 1 列出了不同 Ntree 取值下的测试错误率。

从表 1 中可以看到,C-RF 的分类精度在每个 Ntree 下都要优于 RF。训练收敛的 CNN 虽然可以取得更高的精度,但是在训练迭代的过程中需要大量时间。由于随机权值的有效性,采用了随机权值的 CNN,不需要梯度算法调整参数,大大缩减了特征提取时间,同时实验结果也证明了 C-RF 模型优于 RF。

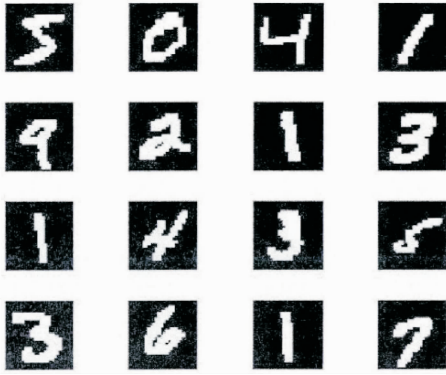


图 3 MNIST 数据集中部分图像

Fig. 3 Some images of MNIST dataset

表 1 MNIST 数据集在 RF 和 C-RF 上的实验结果

Tab. 1 Results of RF and C-RF on MNIST dataset

Ntree	RF	C-RF
100	3.03	2.17
200	3.04	2.13
300	2.93	2.04
400	2.93	1.99
500	2.91	2.00
600	2.87	2.01
700	2.87	2.06
800	2.89	2.10

### 3.2 Rotated MNIST 数据集

为了进一步说明模型的有效性,作者选取了变异的 MNIST 数据集中的—个 Rotated MNIST 数据集作了对比试验, Rotated MNIST 数据集将 MNIST 数据集中的数字图像在  $0 \sim 2\pi$  之间均匀地旋转,如图 4 所示。



图 4 Rotated MNIST 数据集部分图像

Fig. 4 Some images of Rotated MNIST dataset

从 Rotated MNIST 数据集中随机选取了 50 000 个数据为训练数据, 10 000 个为测试数据, 与 MNIST 数据集—样, 分别做了在不同 Ntree 数目下 RF 和 C-RF 的实验。表 2 是 Rotated MNIST 数据集下 RF 和 C-RF 的测试错误率。从表 2 中可以看到模型 C-RF 在不同 Ntree 取值下效果都要好于 RF, 再次验证了模型的有效性, 同时泛化能力也强于 RF。

表 2 Rotated MNIST 数据集在 RF 和 C-RF 上的实验结果

Tab. 2 Results of RF and C-RF on rotated MNIST dataset

Ntree	RF	C-RF
100	11.15	10.44
200	10.44	9.87
300	10.28	9.65
400	10.14	9.45
500	10.14	9.21
600	10.07	9.21
700	10.06	9.19
800	10.08	9.19

## 4 结束语

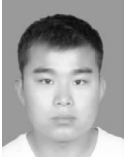
本文提出了一种基于 CNN 的混合模型 C-RF,该模型利用随机权值的 CNN 自动提取输入图像特征,然后用 RF 替换原始 CNN 的输出层,最后将 CNN 提取到的特征输入 RF 中进行分类。本文在 MNIST 和 Rotated MNIST 数据集上进行了相关实验,实验结果表明,和 RF 相比,模型 C-RF 明显提高了分类精度。关于混合模型 C-RF 的研究,还有很多问题需要去探讨,比如模型结构的优化、应用推广等,具体如下:(1)对于 CNN 中每一层卷积核个数以及卷积核大小的设定没有完整的理论来指导,或许可以进一步优化它们的取值;(2)用其他深度学习算法如自动编码器(Auto encoder, AE)、受限玻尔兹曼机(Restrict Boltzmann machine, RBM)等来提取特征是否会取得更好的效果需要进一步探讨;(3)本文只在手写数字数据集中做了实验,是否适用于一些更大更复杂的数据集值得探讨。

### 参考文献:

- [1] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [2] Amit Y, Geman D, Wilder K. Joint induction of shape features and tree classifiers[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997, 19(11): 1300-1305.
- [3] 邹永祥, 吴宗亮. 一种广义不可分的支持向量机算法[J]. 数据采集与处理, 2015, 30(2): 434-440.  
Zou Yongxiang, Wu Zongliang. Generalized C-support vector machine algorithm[J]. Journal of Data Acquisition and Processing, 2015, 30(2): 434-440.
- [4] 奚吉, 赵力, 左加阔. 基于改进多核学习的语音情感识别算法[J]. 数据采集与处理, 2014, 29(5): 730-734.  
Xi Ji, Zhao Li, Zuo Jiakuo. Speech emotion recognition based on modified multiple kernel learning algorithm[J]. Journal of Data Acquisition and Processing, 2014, 29(5): 730-734.
- [5] Niu X X, Suen C Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits[J]. Pattern Recognition, 2012, 45(4): 1318-1325.
- [6] Luo M, Zhang K. A hybrid approach combining extreme learning machine and sparse representation for image classification [J]. Engineering Applications of Artificial Intelligence, 2014, 27: 228-235.
- [7] Rousseeuw P J, Leroy A M. Robust regression and outlier detection[J]. Journal of the American Statistical Association, 1987, 31(2): 260-261.
- [8] Neubauer C. Shape position and size invariant visual pattern recognition based on principles of neocognitron and perception in artificial neural networks [M]. Amsterdam, the Netherlands: North-Holland, 1992: 833-837.
- [9] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [10] Saxe A M, Pang W, Koh Z, et al. On random weights and unsupervised feature learning[C]// Proc of the 28th International Conference on Machine Learning. New York: ACM Press, 2010: 1089-1096.
- [11] He K, Wang Y, Hopcroft J. A powerful generative model using random weights for the deep image representation[C]// 30th Annual Conference on Neural Information Processing Systems. Barcelona, Spain: Neural Information Processing Systems Foundation, 2016: 631-639.
- [12] Maji D, Santara A, Ghosh S, et al. Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images[C]// 37th Annual International Conference of the IEEE. Milan, Italy: Engineering in Medicine and Biology Society, 2015: 3029-3032.
- [13] Lecun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision[C]// IEEE International Symposium on Circuits and Systems. [S. l.]: IEEE, 2010: 253-256.
- [14] Friedl M A, Brodley C E. Decision tree classification of land cover from remotely sensed data[J]. Remote Sensing of Environment, 1997, 61(3): 399-409.
- [15] Quinlan J R. Induction of decision trees[J]. Machine Learning, 1986, 1(1): 81-106.
- [16] Breiman L, Friedman J H, Olshen R, et al. Classification and regression trees[J]. Biometrics, 2015, 40(3): 17-23.

- [17] Efron B. Bootstrap methods: Another look at the Jackknife[J]. *The Annals of Statistics*, 1979, 7(1):1-26.
- [18] Ying W, Li X, Xie Y, et al. Preventing customer churn by using random forests modeling[C]// *IEEE International Conference on Information Reuse and Integration*. Las Vegas, USA: IEEE Computer Society, 2008:429-434.
- [19] 张华伟, 王明文, 甘丽新. 基于随机森林的文本分类模型研究[J]. *山东大学学报理学版*, 2006, 41(3):5-9.  
Zhang Huawei, Wang Mingwen, Gan Lixin. Automatic text classification model based on random forest[J]. *Journal of Shandong University*, 2006, 41(3):5-9.
- [20] Pang H, Datta D, Zhao H. Pathway analysis using random forests with bivariate node-split for survival outcomes[J]. *Bioinformatics*, 2009, 26(26):250-258.
- [21] Ding S, Xu X, Nie R. Extreme learning machine and its applications[J]. *Neural Computing and Applications*, 2014, 25(3): 549-556.

## 作者简介:



侯艳路(1993-),男,硕士研究生,研究方向:深度学习, E-mail:houyanlu@126.com。



丁世飞(1963-),男,教授,研究方向:智能信息处理、人工智能与模式识别、机器学习与数据挖掘、粗糙集与软计算、粒度计算等, E-mail:dingsf@cumt.edu.cn。



孙统凤(1977-),男,副教授,研究方向:智能信息处理、人工智能与模式识别、机器学习与数据挖掘等, E-mail:suntf@cumt.edu.cn。

(编辑:刘彦东)



