



## 一种卷积神经网络集成的多样性度量方法

汤礼颖, 贺利乐, 何林, 屈东东

引用本文:

汤礼颖, 贺利乐, 何林, 等. 一种卷积神经网络集成的多样性度量方法[J]. *智能系统学报*, 2021, 16(6): 1030–1038.

TANG Liying, HE Lile, HE Lin, et al. Diversity measuring method of a convolutional neural network ensemble[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1030–1038.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202011023>

## 您可能感兴趣的其他文章

### 基于增强AlexNet的音乐流派识别研究

Music genre recognition research based on enhanced AlexNet

智能系统学报. 2020, 15(4): 750–757 <https://dx.doi.org/10.11992/tis.201909032>

### 层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification

智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

### 一种高效的稀疏卷积神经网络加速器的设计与实现

Design and implementation of an efficient accelerator for sparse convolutional neural network

智能系统学报. 2020, 15(2): 323–333 <https://dx.doi.org/10.11992/tis.201902007>

### 基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

### 多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene

智能系统学报. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

### 一种多样性和精度加权的数据流集成分类算法

An ensemble classification algorithm based on diversity and accuracy weighting for data streams

智能系统学报. 2019, 14(1): 179–185 <https://dx.doi.org/10.11992/tis.201806021>

 微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202011023

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210923.1720.006.html>

# 一种卷积神经网络集成的多样性度量方法

汤礼颖<sup>1</sup>, 贺利乐<sup>1</sup>, 何林<sup>2</sup>, 屈东东<sup>1</sup>

(1. 西安建筑科技大学机电工程学院, 陕西 西安 710055; 2. 西安建筑科技大学理学院, 陕西 西安 710055)

**摘要:** 分类器模型之间的多样性是分类器集成的一个重要性能指标。目前大多数多样性度量方法都是基于基分类器模型的 0/1 输出结果(即 Oracle 输出)进行计算, 针对卷积神经网络的概率向量输出结果, 仍需要将其转化为 Oracle 输出方式进行度量, 这种方式未能充分利用卷积神经网络输出的概率向量所包含的丰富信息。针对此问题, 利用多分类卷积神经网络模型的输出特性, 提出了一种基于卷积神经网络的概率向量输出方式的集成多样性度量方法, 建立多个不同结构的卷积神经网络基模型并在 CIFAR-10 和 CIFAR-100 数据集上进行实验。实验结果表明, 与双错度量、不一致性度量和  $Q$  统计多样性度量方法相比, 所提出的方法能够更好地体现模型之间的多样性, 为模型选择集成提供更好的指导。

**关键词:** 卷积神经网络; 集成学习; 多样性度量; 机器学习; 分类器集成; 概率向量输出; Oracle 输出; 基模型  
**中图分类号:** TP181; TP391   **文献标志码:** A   **文章编号:** 1673-4785(2021)06-1030-09

中文引用格式: 汤礼颖, 贺利乐, 何林, 等. 一种卷积神经网络集成的多样性度量方法[J]. 智能系统学报, 2021, 16(6): 1030-1038.

英文引用格式: TANG Liying, HE Lile, HE Lin, et al. Diversity measuring method of a convolutional neural network ensemble[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1030-1038.

## Diversity measuring method of a convolutional neural network ensemble

TANG Liying<sup>1</sup>, HE Lile<sup>1</sup>, HE Lin<sup>2</sup>, QU Dongdong<sup>1</sup>

(1. School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China; 2. School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China)

**Abstract:** Diversity among classifier models has been recognized as a significant performance index of a classifier ensemble. Currently, most diversity measuring methods are defined based on the 0/1 outputs (namely Oracle outputs) of the base model. The probability vector outputs of a convolutional neural network (CNN) still need to be converted into Oracle outputs for measurement, which fails to fully use the rich information contained in the CNN probability vector outputs. To solve this problem, a new diversity measuring method for probabilistic vector outputs based on CNNs is proposed. Several base models of CNN models with various structures are established and tested on the CIFAR-10 and CIFAR-100 datasets. Compared with double-fault measure, disagreement measure, and Q-Statistic, the proposed method can better reflect the differences between the models and provide better guidance for a selective ensemble of CNN models.

**Keywords:** CNN; ensemble learning; diversity measures; machine learning; multiple classifier ensembles; probability vector outputs; Oracle outputs; basic model

集成学习是一种重要的机器学习方法, 主要思路是多个分类器基于某种组合方式组合在一起, 来取得比单个分类器更好的性能, 从而可显著提高学习系统的预测精度和泛化能力<sup>[1]</sup>。

基分类器之间的多样性是影响集成系统泛化

性能的重要因素<sup>[2]</sup>, 产生泛化能力强、差异性大的基学习器是集成算法的关键。但如何有效的度量并利用这些多样性还是一个悬而未决的问题。目前大多数多样性度量方法都是基于基分类器的 0/1 输出方式 (Oracle 输出), 主要分为两大类: 成对度量和非成对度量。成对度量方法主要有  $Q$  统计方法<sup>[3]</sup>、不一致度量方法<sup>[4]</sup>、双错度量方法<sup>[5]</sup> 和相关系数法<sup>[6]</sup> 等; 非成对度量方法主要有 KW 差

收稿日期: 2020-11-20. 网络出版日期: 2021-09-24.

基金项目: 国家自然科学基金项目 (61903291).

通信作者: 贺利乐. E-mail: [hllnh2013@163.com](mailto:hllnh2013@163.com).

异度量方法<sup>[7]</sup>、 $\kappa$ 评判间一致性度量方法<sup>[8]</sup>、熵度量方法<sup>[9-10]</sup>、难度度量<sup>[11]</sup>、广义差异性度量<sup>[12]</sup>等。

近年来国内学者们也提出了一些新的研究多样性度量方法。例如: 邢红杰等<sup>[13]</sup>提出了一种关系熵多样性度量方法, 结合相关熵和距离方差描述支持向量数据来实现选择性集成; 李莉<sup>[14]</sup>提出了一种距离信息熵的多样性描述方法实现基分类器集成, 该方法在 UCI 数据集上具有与成对差异性度量方法一致的效果; 赵军阳等<sup>[15]</sup>针对多分类器系统差异性评价中无法处理模糊数据的问题, 提出了一种采用互补性信息熵的分类器差异性度量方法, 该方法省略了对分类器输出结果的正确性判别过程, 直接处理分类器的输出结果, 但是未考虑集成规模与精度之间的平衡问题, 对系统的泛化性能有一定的影响; 周刚等<sup>[16]</sup>提出了一种联合信息增益的过程多样性度量方法, 该方法在西瓜数据集上拥有与传统方法相近的准确性。上述几种方法都是基于信息论角度来定义多样性的, 基模型的训练和集成都采用机器学习方法得到, 在应用卷积神经网络集成的多样性研究方面较少。

随着深度学习的快速发展, 卷积神经网络 (convolutional neural network, CNN)<sup>[17]</sup> 因其独特的结构已成为当前图像识别领域的研究热点。卷积神经网络分类器的输出方式与 Oracle 输出方式不同, 采用的是  $c$  维向量输出方式。在处理多分类 ( $c > 2$ ) 问题中, 分类器输出单元都需用 Softmax 激活函数进行数值处理, 将多分类的输出数值向量形式转化为类标签的相对概率向量方式, 即  $c$  维向量形式  $\mathbf{D}(z) = [d_0(z) \ d_1(z) \ \cdots \ d_{c-1}(z)]$ ,  $d_{c-1}(z)$  为分类器  $\mathbf{D}$  给出的样例属于第  $c-1$  类的置信度,  $i = 1, 2, \dots, c-1$ ,  $c$  为不同类标签的数目。针对卷积神经网络的概率向量输出结果, 使用 0/1 输出准则来统计分类器对样本分类结果的正确/错误个数来计算多样性, 是一种硬性的评判标准<sup>[18]</sup>。如两个 5 分类器的输出概率向量是  $\mathbf{D}_i(z) = (0.17, 0.03, 0.6, 0.05, 0.15)$  和  $\mathbf{D}_j(z) = (0.1, 0.2, 0.5, 0.17, 0.13)$ , 将它们转化为 Oracle 输出方式, 两个分类器的结果都为第三类, 二者输出结果一致, 采用基于模型的 Oracle 输出结果多样性度量方法将会判定二者没有任何不同。但事实上, 两个分类器输出结果存在明显的不同, Oracle 输出方式未能充分利用卷积神经网络输出的概率向量所包含的丰富信息。

针对此问题, 本文充分利用多分类卷积神经网络的输出特性信息, 提出了一种基于卷积神经网络集成的多样性度量方法。采用概率向量输出方式来表示模型分类结果, 引入两个向量的差异

性, 针对每一个训练样本计算模型输出结果之间的差异性来评估多样性。

### 1 3 种传统多样性方法

假设存在分类器  $h_i$  和  $h_j (i, j = 1, 2, \dots, T, i \neq j)$ , 1 表示分类器正确, 0 表示分类器分类错误, 两个分类器的联合输出结果用表 1 中 a、b、c、d 表示。a 表示在训练过程中分类器  $h_i$  和  $h_j$  都对样本分类正确的样本个数, b 表示在训练过程中分类器  $h_i$  对样本分类错误而分类器  $h_j$  分类正确的样例个数, c 表示分类器  $h_i$  对样本分类正确而分类器  $h_j$  错分的样例个数, d 表示分类器  $h_i$  和  $h_j$  都对样本错分的样例个数。由此, 总的样本个数可以表示为  $m = a + b + c + d$ , 两个分类器之间的关系如表 1 所示。

表 1 两个分类器之间的关系  
Table 1 Relationship between a pair of classifiers

联合输出结果	$h_j = 1$	$h_j = 0$
$h_i = 1$	a	c
$h_i = 0$	b	d

#### 1) $Q$ 统计方法 ( $Q$ )

$$Q_{ij} = \frac{ad - bc}{ad + bc} \quad (1)$$

$Q$  的取值为  $[-1, 1]$ , 当两个分类器总是对其正确分类或者错误分类, 说明其行为是一致的, 则有  $b = c = 0$ , 即  $Q_{ij} = 1$ , 此时它们的多样性最低; 反之, 如果两个分类器的分类刚好一个正确一个错误, 即  $Q_{ij} = -1$ , 这种情况下多样性最好。

#### 2) 不一致性度量 ( $D_{is}$ )

不一致性度量表示两个分类器之中一个判断正确一个判断错误的测试个数与总测试个数的比值。

$$D_{is} = \frac{b + c}{m} \quad (2)$$

#### 3) 双错度量方法 ( $D_F$ )

$$D_F = d/m \quad (3)$$

$D_F$  取值范围为  $[0, 1]$ , 值越大, 说明两个分类器都对其分类出错, 多样性越差。

## 2 基于概率向量输出方式的多样性度量方法

首先定义两个多类分类的分类器 A 和 B,  $z_j (j = 1, 2, \dots, m)$  为  $m$  个样本中的第  $j$  个样本, 则对样本  $z_j$  的模型输出概率向量为

$$\begin{aligned} \mathbf{D}_A(z) &= [d_{A_0} \ d_{A_1} \ d_{A_2} \ \cdots \ d_{A_{c-1}}] \\ \mathbf{D}_B(z) &= [d_{B_0} \ d_{B_1} \ d_{B_2} \ \cdots \ d_{B_{c-1}}] \end{aligned}$$

其中, 向量中的各元素是模型预测概率值, 索引  $(A_0, A_1, A_2, \dots, A_{c-1})$  和  $(B_0, B_1, B_2, \dots, B_{c-1})$  分别为向量  $D_A$  和  $D_B$  各概率对应的类标签值, 有  $\sum_{k=0}^{c-1} d_{A_k} = 1, \sum_{k=0}^{c-1} d_{B_k} = 1$ 。

模型经过训练之后, 输出层经过 Softmax 激活函数得到模型概率向量结果, 定义卷积神经网络的多样性度量方法, 基本思路如下:

1) 将 2 个分类器概率向量中的元素从大到小进行排列, 并返回其对应的 index(类标签) 值, 得到分类器最大置信度概率由大到小排列对应的类标签向量  $L_A[i]$  和  $L_B[i], i = 0, 1, 2, \dots, c-1, c$  为不同类标签的数目。

$$L_A[i] = \text{np.argsort}(D_A(z))[:, -1] \quad (4)$$

$$L_B[i] = \text{np.argsort}(D_B(z))[:, -1]$$

2) 对于已标记单一样本  $z_j$ , 两个分类器的多样性  $D_{AB}(z_j)$  定义如式 (5)、(6):

$$f_{AB}(i) = \begin{cases} 1, L_A[i] \neq L_B[i] \\ 0, L_A[i] = L_B[i] \end{cases}, \quad i = 0, 1, 2, \dots, c-1 \quad (5)$$

$$D_{AB}(z_j) = \frac{1}{c} \sum_{i=0}^{c-1} (k_i f_{AB}(i))^2 \quad (6)$$

式中:  $k_i$  为不同排序位置差异的权重 (越靠前的输出结果不一样说明差异性越大), 本文通过多次实验对比, 最终取值为  $\{32, 8, 4, 2, 1, 0, \dots, 0\}$ , 采用指数递减值。

式 (4) 返回两个分类器元素之间的差异性结果序列。分类器之间的多样性只是反映两个模型对于已标记样本差异性, 与模型类标签值无关, 与输出结果的位置无关。因此, 为了保证度量的准确性和有效性, 定义式 (5) 在分类器相同的索引值情况下, 当两个分类器输出类标签是一致时, 返回为 0; 当输出类标签不同时返回为 1。将式 (5) 代入式 (6), 得到两个分类器对于单一样本多样性的定义。  $D_{AB}(z_j)$  值越大, 两个概率向量差异性越大, 分类器之间的多样性也就越大。

假设对于样本  $z_j$  两个 5 分类的输出概率向量是  $D_1(z_j) = (0.03, 0.07, 0.6, 0.05, 0.25)$  和  $D_2(z_j) = (0.1, 0.17, 0.4, 0.2, 0.13)$ , 得到分类器最大置信度概率由大到小排列对应的类标签向量  $L_1[z_j] = (2, 4, 3, 1, 0)$  和  $L_2[z_j] = (2, 3, 1, 4, 0)$ , 得到向量  $f_{AB}(i) = (0, 1, 1, 1, 0)$ 。对于单一样本  $z$  得到

$$D_{AB}(z_j) = \frac{1}{5} \sum_{i=0}^4 (k_i f_{AB}(i))^2 =$$

$$\frac{1}{5} ((8 \times 0)^2 + (4 \times 1)^2 + (2 \times 1)^2 + (1 \times 1)^2 + (0 \times 0)^2) = 4.2$$

而如果采用基于 Oracle 输出的多样性度量方式, 对于该样本两分类器输出结果相同, 得到的多样性为 0。

3) 对于已标记样本集  $Z = \{z_1, z_2, \dots, z_j, \dots, z_m\}$ , 两个基模型之间的多样性为所有样本的集合多样性后的平均值, 如式 (7) 所示:

$$D_{AB}(Z) = \frac{1}{m} \sum_{j=0}^{m-1} D_{AB}(z_j) \quad (7)$$

4) 对于多分类器系统  $\{D_1, D_2, \dots, D_T\}$ , 其多样性为计算每对基模型之间的多样性的平均值:

$$D_{PV}(Z) = \frac{2}{T(T-1)} \sum_{i=1}^{T-1} \sum_{j=i+1}^T D_{ij}(Z) \quad (8)$$

### 3 实验验证与分析

为了验证新的多样性度量方法的有效性和先进性, 采用 4 种方法计算基模型的多样性:  $Q$  统计、不一致性度量、双错度和本文提出的基于概率向量方法  $D_{PV}$ 。在 CIFAR-10 与 CIFAR-100 数据集上分别进行了实证对比分析。

实验分析主要分为 3 部分内容: 1) 训练 19 个基学习器模型, 并得到单个基模型以及预测结果; 2) 基于 CIFAR-10、CIFAR-100 数据集采用 4 种方法计算各基模型之间的多样性; 3) 在数据集上进行基模型的集成预测, 采用简单平均集成策略集成基模型, 并对分析单个模型与集成模型之间的预测结果。

#### 3.1 模型结构

卷积神经网络是一种具有深度结构的前馈神经网络, 对于大型图像处理具有十分出色的表现<sup>[19]</sup>。CNN 的基本结构由输入层、卷积层、池化层、全连接层和输出层构成。本文构造了 19 种不同结构的 CNN 网络模型, 网络结构都包括输入层、卷积层、池化层、Dropout 层、Flatten 层和全连接层, 各模型结构不同之处在于卷积层、池化层和 Dropout 层的不同。表 2、3 为其中 9 个模型的具体结构。

表 2 4 个候选基模型结构

Table 2 Structures of four-candidate basic models

结构	Model_2	Model_3	Model_5	Model_6
输入层	32×32×3	32×32×3	32×32×3	32×32×3
卷积层1	Conv3-64 3×3	Conv3-512 3×3	Conv3-64 3×3	Conv3-512 3×3

续表 2

结构	Model_2	Model_3	Model_5	Model_6
卷积层2	Conv3-128 3×3	Conv3-256 3×3	Conv3-128 3×3	Conv3-256 3×3
卷积层3	—	—	—	—
池化层1	Maxpool	Maxpool	Maxpool	Maxpool
卷积层4	Conv3-256 3×3	Conv3-128 3×3	Conv3-256 3×3	Conv3-128 3×3
卷积层5	Conv3-512 3×3	Conv3-64 3×3	Conv3-512 3×3	Conv3-64 3×3
卷积层6	Conv3-64 3×3	—	Conv3-512 3×3	—
池化层2	Maxpool	Maxpool	Maxpool	Maxpool
过拟合层1	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)
卷积层7	Conv3-128 3×3	Conv3-512 3×3	Conv3-256 3×3	Conv3-64 3×3
卷积层8	Conv3-256 3×3	Conv3-256 3×3	Conv3-128 3×3	Conv3-128 3×3
卷积层9	Conv3-512 3×3	—	Conv3-64 3×3	Conv3-256 3×3
池化层3	—	Maxpool	—	—
卷积层10	—	Conv3-128 3×3	—	Conv3-512 3×3
卷积层11	—	Conv3-64 3×3	—	—
卷积层12	—	—	—	—
池化层4	Global Maxpool	Global Maxpool	Global Maxpool	Global Maxpool
拉平	fatten	fatten	fatten	fatten
过拟合层2	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)
全连接层	Dense(64)	Dense(65)	Dense(66)	Dense(67)
输出层	Softmax	Softmax	Softmax	Softmax

表 3 5 个候选基模型结构

Table 3 Structures of five-candidate basic models

结构	Model_8	Model_10	Model_12	Model_13	Model_14
输入层	32×32 ×3	32×32 ×3	32×32 ×3	32×32 ×3	32×32 ×3
卷积层1	Conv3-256 3×3	Conv3-512 3×3	Conv3-64 3×3	Conv3-64 1×1	Conv3-64 3×3
卷积层2	Conv3-128 3×3	Conv3-64 3×3	Conv3-64 1×1	Conv3-64 3×3	Conv3-64 1×3
卷积层3	—	—	—	Conv3-64 1×1	Conv3-64 3×1
池化层1	Maxpool	Maxpool	Maxpool	Maxpool	Maxpool
卷积层4	Conv3-256 3×3	Conv3-128 3×3	Conv3-256 3×3	Conv3-256 1×1	Conv3-128 3×3
卷积层5	Conv3-128 3×3	Conv3-64 3×3	Conv3-256 1×1	Conv3-256 3×3	Conv3-128 1×3
卷积层6	Conv3-256 3×3	—	—	Conv3-256 1×1	Conv3-128 3×1
池化层2	Maxpool	Maxpool	Maxpool	Maxpool	Maxpool
过拟合层1	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)
卷积层7	Conv3-128 3×3	Conv3-512 3×3	Conv3-512 3×3	Conv3-512 1×1	Conv3-256 3×3
卷积层8	Conv3-256 3×3	Conv3-64 3×3	Conv3-512 1×1	Conv3-512 3×3	Conv3-256 1×3
卷积层9	Conv3-64 3×3	—	—	Conv3-512 1×1	Conv3-256 3×1
池化层3	—	Maxpool	Maxpool	Maxpool	Maxpool
卷积层10	—	Conv3-128 3×3	Conv3-128 3×3	Conv3-128 1×1	Conv3-512 3×3
卷积层11	—	Conv3-64 3×3	Conv3-128 1×1	Conv3-128 3×3	Conv3-512 1×3
卷积层12	—	—	—	Conv3-128 1×1	Conv3-512 3×1
池化层4	Global Maxpool	Global Maxpool	Global Maxpool	Global Maxpool	GlobalMaxpool

续表 3

结构	Model_8	Model_10	Model_12	Model_13	Model_14
拉平	fatten	fatten	fatten	fatten	fatten
过拟合层2	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)	Dropout(0.25)
全连接层	Dense(68)	Dense(69)	Dense(70)	Dense(71)	Dense(72)
输出层	Softmax	Softmax	Softmax	Softmax	Softmax

### 3.2 实验环境及评价指标

本实验采用的是 Windows10 系统的 GTX1 660Ti 6 GB 独立显卡实验平台。在 Window10 系统下,所有网络均基于深度学习框架 Tensorflow2.0 实现。

模型的性能评价指标包括训练集、验证集、测试集的准确性 (Accuracy, 简称  $A_{cc}$ ), 是衡量模型正确分类的标准。其中测试集的准确率直接反映了模型的预测能力, 见式 (9) 所示:

$$A_{cc} = \frac{1}{m} \sum_{j=1}^m I(y_j = f(x_j)) \quad (9)$$

式中:  $m$  是已标记样本集  $Z = \{z_1, z_2, \dots, z_m\}$  中样本个数;  $y_j$  是类标签;  $f(x_j)$  为模型预测结果;  $I$  是条件判断函数。

### 3.3 实验数据集

CIFAR-10 和 CIFAR-100 数据集是用于普适物体识别的图像分类数据集, 由 Alex 等<sup>[20-21]</sup> 收集。CIFAR-10 和 CIFAR-100 数据集都包含 60 000 张 32 像素×32 像素的彩色图片, 分别为训练集 50 000 张和测试集 10 000 张, 训练集用于训练基

学习器, 并计算模型间多样性值; 测试集则用于对各种基模型和集成模型的预测效果进行验证和对比。CIFAR-10 数据集分为 10 个类, 每个类有 6 000 张图片; CIFAR-100 数据集有 20 个超类, 每个超类分为 5 个小类, 总计 100 分类, 每个分类包含 600 张图片。

### 3.4 实验结果分析

#### 3.4.1 基模型预测结果

基学习器模型分别在 CIFAR-10 和 CIFAR-100 数据集上进行训练, 使用 SGD 优化器<sup>[22]</sup> 更新权重, 对训练数据集首先进行图像预处理, 包括图像归一化和数据集打乱, 批量规范化之后训练单批次大小为 128。经过反复实验, 卷积神经网络模型大概在 200 次迭代之后已经完全收敛, 识别率达到了最高, 故 Epochs 值设置为 200; 在卷积操作中使用 ReLU<sup>[23]</sup> 激活函数, 添加  $L_2$  正则化处理, 参数设置  $\lambda$  为 0.001; 利用 Dropout 防止过拟合, 设置为 0.25; 采用分段学习率训练方式, 初始学习率设为 0.01, 动量系数均为 0.9; 权重初始化采用 He 正态分布初始化器, 偏差初始化为零。单个基模型经过训练后的预测结果如表 4 所示。

表 4 19 个候选基模型预测结果  $A_{cc}$

Table 4 Prediction results  $A_{cc}$  of 19 candidate basic models

%

基模型	CIFAR-10	CIFAR-100	基模型	CIFAR-10	CIFAR-100
model_1	90.81	67.41	model_11	88.85	62.55
model_2	91.23	65.43	model_12	88.33	61.30
model_3	89.44	63.12	model_13	88.97	62.04
model_4	90.25	65.07	model_14	88.73	62.98
model_5	90.82	66.46	model_15	51.47	63.29
model_6	87.99	62.12	model_16	80.92	53.89
model_7	90.72	65.13	model_17	79.17	50.65
model_8	91.00	66.50	model_18	81.00	49.80
model_9	88.87	61.75	model_19	60.76	36.75
model_10	89.11	62.86			

由表 4 可知, 基模型 15~19 的预测准确性较低, 考虑到模型集成之后的效果, 初步筛选出基模型 1~14 作为初筛选基模型。

#### 3.4.2 基于 CIFAR-10 数据集的多样性分析

表 5 是 3 种传统多样性方法和本文方法分别

在 CIFAR-10 数据集上得到的多样性统计值最大的两分类器集成模型, 表 6 是表 5 得到的两分类器集成模型在测试集上的预测结果。同理, 表 7 是得到的多样性统计值最大的三分类器集成模型, 表 8 是其预测结果; 表 9 是得到的多样性统计

值最大的四分类器集成模型结果, 表 10 是其预测结果。其中, 平均集成精度是指通过多样性方法筛选得到的三组集成模型组合采用简单平均集成策略得到的在测试集上的集成精度, 并求取平均值的结果; 基模型平均精度是指表 6 中筛选出的单个基模型在测试集上的分类精度的平均值。如表 6 中 DF 方法的平均集成精度就等于表 5 中对应的 3 个两分类器集成模型 2-14、2-5 和 5-14 分别在测试集上的集成精度的平均值, 而相应的基模型平均精度是指单一基模型 2、5 和 14 测试精度的均值结果。

表 5 CIFAR-10 上多样性最大的两分类器集成模型  
Table 5 Results of an optimal two-classifiers model on CIFAR-10

降序排列	$D_F$	$D_{is}$	$Q$	$D_{PV}$
1	2-14	6-12	6-13	2-6
2	2-5	6-13	6-8	2-3
3	5-14	12-14	5-6	2-5

表 6 两分类器集成模型在 CIFAR-10 上预测结果  
Table 6 Prediction results of two-classifier models on CIFAR-10 %

类别	平均集成精度	基模型平均精度	提高百分比
$D_F$	91.987	90.260	1.727
$D_{is}$	90.093	88.510	1.583
$Q$	90.920	89.700	1.220
$D_{PV}$	92.067	89.870	2.197

表 7 CIFAR-10 上多样性最大的三分类器集成模型  
Table 7 Results of optimal three-classifier models on CIFAR-10

降序排列	$D_F$	$D_{is}$	$Q$	$D_{PV}$
1	2-5-14	6-12-13	12-13-14	2-3-6
2	2-5-6	6-12-14	6-12-13	2-5-6
3	2-3-5	6-13-14	5-6-13	2-6-8

表 8 三分类器集成模型在 CIFAR-10 上预测结果  
Table 8 Prediction results of three-classifier models on CIFAR-10 %

类别	平均集成精度	基模型平均精度	提高百分比
$D_F$	92.490	89.964	2.526
$D_{is}$	91.003	88.505	2.498
$Q$	91.010	88.968	2.042
$D_{PV}$	92.293	89.718	2.575

由表 5、6 可知, 本文方法筛选出的二分类器集成模型是 2-6、2-3 和 2-5, 平均集成精度相比于基模型平均精度提升效果最优, 提升了 2.197%;

由表 7 和表 8 可知, 本文方法筛选出的三分类器集成模型为 2-3-6、2-5-6 和 2-6-8, 集成精度相比于基模型平均精度提升效果最优, 提升了 2.575%; 由表 9、10 可知, 本文方法筛选出的四分类器模型是 2-3-5-6、2-3-6-8 和 2-3-6-10, 集成精度相比于基模型精度提高了 2.580%, 比  $Q$  统计方法效果好, 比不一致性方法相差 0.397%, 比双错度量方法相差 0.074%。同时, 本文方法筛选出的二分类器的平均集成精度是最高的, 达到了 92.067%, 性能相比于其他 3 种方法是最优的, 三分类器和四分类器的平均集成精度与双错度量得到的结果相近, 且高于不一致性方法和  $Q$  统计方法, 说明本文提出的基于概率向量的多样性度量方法是有效的。

表 9 CIFAR-10 上多样性最大的四分类器集成模型  
Table 9 Results of optimal four-classifier models on CIFAR-10

降序排列	$D_F$	$D_{is}$	$Q$	$D_{PV}$
1	2-5-8-14	6-12-13-14	6-12-13-14	2-3-5-6
2	2-5-6-8	6-9-12-13	5-12-13-14	2-3-6-8
3	2-5-13-14	11-12-13-14	2-12-13-14	2-3-6-10

表 10 四分类器集成模型在 CIFAR-10 的预测结果  
Table 10 Prediction results of four-classifier models on CIFAR-10 %

类别	平均集成精度	基模型平均精度	提高百分比
$D_F$	92.767	89.790	2.977
$D_{is}$	91.277	88.623	2.654
$Q$	91.810	89.495	2.315
$D_{PV}$	92.511	89.931	2.580

### 3.4.3 基于 CIFAR-100 数据集的多样性分析

为进一步验证本文所提方法的有效性, 在 CIFAR-100 数据集上进行验证实验。表 11 是 3 种传统多样性方法和本文方法分别在 CIFAR-100 数据集上得到的多样性统计值最大的两分类器集成模型, 表 12 是表 11 得到的两分类器集成模型的预测结果; 表 13 是得到的多样性统计值最大的三分类器集成模型, 表 14 是其预测结果; 表 15 是得到的多样性统计值最大的四分类器集成模型, 表 16 是其预测结果。

由表 11、12 可知, 本文方法在 CIFAR-100 数据集上筛选出的二分类器集成模型是 5-14、12-14 和 3-14, 平均集成精度相比于基模型平均精度提升效果最优, 提升了 5.812%; 由表 13、14 可知, 本文方法筛选出的三分类器集成模型为 5-12-14、5-8-14 和 3-8-14, 平均集成精度相比于基模型平

均精度提升效果最优,提升了 7.421%;由表 15、16 可知,本文方法筛选出的四分类器模型是 3-5-12-14、3-5-8-14 和 5-8-12-14,平均集成精度相比于基模型平均精度提高了 7.846%,集成精度相比于基模型精度提升效果最优。同时,在 CIFAR-100 数据集上,本文方法相比于其他 3 种方法,筛选出的集成模型的平均集成精度无论是二分类器、三分类器还是四分类器都是最高的,集成性能优于双错度量方法、不一致性方法和  $Q$  统计方法,再次验证了本文方法的有效性。

表 11 CIFAR-100 上多样性最大的两分类器集成模型  
Table 11 Results of optimal two-classifier models on CIFAR-100

降序排列	$D_F$	$D_{is}$	$Q$	$D_{PV}$
1	5-6	6-12	12-14	5-14
2	5-14	12-14	6-8	12-14
3	5-13	11-12	13-14	3-14

表 12 两分类器集成模型在 CIFAR-100 的预测结果  
Table 12 Prediction results of two-classifier models on CIFAR-100 %

类别	平均集成精度	基模型平均精度	提高百分比
$D_F$	67.740	63.400	4.340
$D_{is}$	65.880	62.240	3.640
$Q$	67.273	62.988	4.285
$D_{PV}$	69.278	63.466	5.812

表 13 CIFAR-100 上多样性最大的三分类器集成模型  
Table 13 Results of optimal three-classifier models on CIFAR-100

降序排列	$D_F$	$D_{is}$	$Q$	$D_{PV}$
1	5-13-14	6-12-13	12-13-14	5-12-14
2	5-12-14	12-13-14	9-13-14	5-8-14
3	5-6-13	11-12-13	6-12-13	3-8-14

表 14 三分类器集成模型在 CIFAR-100 的预测结果  
Table 14 Prediction results of three-classifier models on CIFAR-100 %

类别	平均集成精度	基模型平均精度	提高百分比
$D_F$	70.230	62.980	7.250
$D_{is}$	68.427	62.198	6.229
$Q$	68.413	62.038	6.375
$D_{PV}$	71.493	64.072	7.421

表 15 CIFAR-100 上多样性大的四分类器集成模型  
Table 15 Results of optimal four-classifier models on CIFAR-100

降序排列	$D_F$	$D_{is}$	$Q$	$D_{PV}$
1	2-5-13-14	9-12-13-14	9-11-13-14	3-5-12-14
2	5-8-13-14	6-9-12-13	9-12-13-14	3-5-8-14
3	5-12-13-14	11-12-13-14	6-12-13-14	5-8-12-14

表 16 四分类器集成模型在 CIFAR-100 的预测结果  
Table 16 Prediction results of four-classifier models on CIFAR-100 %

类别	平均集成精度	基模型平均精度	提高百分比
$D_F$	71.697	64.120	7.577
$D_{is}$	69.590	62.123	7.467
$Q$	69.503	62.123	7.380
$D_{PV}$	71.918	64.072	7.846

在基模型准确性比较接近的情况下,在 CIFAR-10 数据集上  $D_{PV}$  方法和双错度量方法筛选出的集成模型的集成性能较接近,都优于不一致性和  $Q$  统计方法;在 CIFAR-100 数据集上,  $D_{PV}$  方法筛选出的集成模型无论是平均集成性能还是相对于基模型提升的效果都是最佳的。由此说明,在基模型相对较弱,模型之间有更大的互补空间可以利用时,  $D_{PV}$  方法更能够体现模型之间的互补性,可选择更好的模型组合进行集成。通过在 CIFAR-10 和 CIFAR-100 数据集上验证,本文提出的  $D_{PV}$  方法可以充分利用卷积神经网络的概率向量的输出特性,深度挖掘模型内部互补信息,相较传统多样性度量方法更能够体现出模型之间的互补性和多样性。

#### 4 结束语

针对目前模型集成中多样性度量方法在基模型为卷积神经网络时未充分利用基模型输出的概率向量信息的问题,本文提出一种基于卷积神经网络概率向量的多样性度量方法。通过在 CIFAR-10 和 CIFAR-100 数据集上训练多个不同结构的卷积神经网络进行基于多样性方法的选择对比实验,结果表明,在 CIFAR-10 数据集上本文方法和双错度量方法相近,且优于不一致性和  $Q$  统计方法;在 CIFAR-100 数据集上,相比于双错度量、 $Q$  统计和不一致性方法,本文提出的方法是最优的。本文提出的方法能够充分利用卷积神经网络的概率向量输出特性,更好地体现模型之间的多样性,特别是在基模型整体较弱时,能够选

出集成效果相对更好的模型组合,且集成模型的性能对比基模型平均性能的提升效果也更优。此方法为基于概率向量输出的卷积神经网络模型集成选择提供了一种新的多样性度量思路。

## 参考文献:

- [1] OPITZ D, MACLIN R. Popular ensemble methods: an empirical study[J]. *Journal of artificial intelligence research*, 1999, 11: 169–198.
- [2] ZHOU Zhuhui. Ensemble methods: foundations and algorithms[M]. New York: CRC Press, 2012: 236.
- [3] YULE G U. On the association of attributes in statistics: with illustrations from the material of the childhood society, &c[J]. *Philosophical transactions of the royal society of London. Series A*, 1900, 1900, 194: 257–319.
- [4] SKALAK D B. The sources of increased accuracy for two proposed boosting algorithms[C]//*Proceedings of American Association for Artificial Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop*. Portland, USA, 1996: 1133.
- [5] GIACINTO G, ROLI F. Design of effective neural network ensembles for image classification purposes[J]. *Image and vision computing*, 2001, 19(9/10): 699–707.
- [6] KUNCHEVA L I, WHITAKER C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy[J]. *Machine learning*, 2003, 51(2): 181–207.
- [7] KOHAVI R, WOLPERT D H. Bias plus variance decomposition for zero-one loss functions[C]//*Proceedings of the 13th International Conference on Machine Learning*. San Francisco, USA, 1996: 275–283.
- [8] CONOVER W J. Statistical methods for rates and proportions[J]. *Technometrics*, 1974, 16(2): 326–327.
- [9] SHIPP C A, KUNCHEVA L I. Relationships between combination methods and measures of diversity in combining classifiers[J]. *Information fusion*, 2002, 3(2): 135–148.
- [10] HANSEN L K, SALAMON P. Neural network ensembles[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2002, 24(10): 993–1001.
- [11] CUNNINGHAM P, CARNEY J. Diversity versus quality in classification ensembles based on feature selection[C]//*Proceedings of the 11th European Conference on Machine Learning*. Catalonia, Spain, 2000: 109–116.
- [12] PARTRIDGE D, KRZANOWSKI W. Software diversity: practical statistics for its measurement and exploitation[J]. *Information and software technology*, 1997, 39(10): 707–717.
- [13] 邢红杰, 魏勇乐. 基于相关熵和距离方差的支持向量数据描述选择性集成[J]. *计算机科学*, 2016, 43(5): 252–256, 264.  
XING Hongjie, WEI Yongle. Selective ensemble of SVDDs based on correntropy and distance variance[J]. *Computer science*, 2016, 43(5): 252–256, 264.
- [14] 李莉. 基于差异性度量的分类器集成优化方法研究与应用[D]. 大连: 大连海事大学, 2017.  
LI Li. Optimization method research and application of multiple classifiers ensemble based on diversity measure[D]. Dalian: Dalian Maritime University, 2017.
- [15] 赵军阳, 韩崇昭, 韩德强, 等. 采用互补信息熵的分类器集成差异性度量方法[J]. *西安交通大学学报*, 2016, 50(2): 13–19.  
ZHAO Junyang, Han Chongzhao, Han Deqiang, et al. A novel measure method for diversity of classifier integrations using complement informationentropy[J]. *Journal of Xi'an Jiaotong University*, 2016, 50(2): 13–19.
- [16] 周钢, 郭福亮. 基于信息熵的集成学习过程多样性度量研究[J]. *计算机工程与科学*, 2019, 41(9): 1700–1707.  
ZHOU Gang, GUO Fuliang. Process diversity measurement of ensemble learning based on information entropy[J]. *Computer engineering and science*, 2019, 41(9): 1700–1707.
- [17] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. *计算机学报*, 2017, 40(6): 1229–1251.  
ZHOU Feiyan, JIN Linpeng, DONG Jun. DONG Jun. Review of convolutional neural network[J]. *Chinese journal of computers*, 2017, 40(6): 1229–1251.
- [18] FAN Tiegang, ZHU Ying, CHEN Junmin. A new measure of classifier diversity in multiple classifier system [C]//*Proceedings of 2008 International Conference on Machine Learning and Cybernetics*. Kunming, China, 2008.
- [19] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. *自动化学报*, 2016, 42(9): 1300–1312.  
CHANG Liang, DENG Xiaoming, ZHOU Mingquan, et al. Convolutional neural networks in image understanding[J]. *Acta Automatica Sinica*, 2016, 42(9): 1300–1312.
- [20] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[J]. *Handbook of systemic autoimmune diseases*, 2009, 1(4): 7.
- [21] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, USA, 2012.
- [22] SINHA N K, GRISCIK M P. A stochastic approxima-

tion method[J]. IEEE transactions on systems, man, and cybernetics, 2007, SMC-1(4): 338-344.

[23] GLOOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[J]. Journal of machine learning research, 2011, 15: 315-323.

作者简介:



汤礼颖, 硕士研究生, 主要研究方向为图像识别与目标检测。



贺利乐, 教授, 博士生导师, 主要研究方向为机器人智能化技术、机器学习。2015 年获陕西省高等学校科学技术奖二等奖, 2016 年获陕西省科学技术奖三等奖。获发明专利授权 5 件, 出版专著 1 部, 教材 4 部, 发表学术论文 86 篇。



何林, 讲师, 主要研究方向为深度学习。

### AI 好书推荐——《智能算法导论》

智能算法是在进化计算、模糊逻辑、神经网络三个分支发展相对成熟的基础上, 通过相互之间的有机融合而形成的新的科学方法, 也是智能理论和技术发展的崭新阶段。本书对智能算法的前沿领域进行了详细的论述, 主要内容包括: 遗传算法、免疫算法、Memetic 算法、粒子群算法、蚁群算法、狼群算法、蜂群算法、细菌觅食算法、分布估计算法、差分进化算法、模拟退火、贪心算法、雨滴算法、禁忌搜索算法、量子算法、最小二乘法、A\*算法、神经网络算法、深度学习算法、混合智能算法。

作者简介:

尚荣华, 女, 西安电子科技大学华山学者特聘教授, 博导, 主要研究方向为智能感知与自然计算、类脑计算与大数据的学习、优化与应用。

焦李成, 男, 西安电子科技大学华山学者领军教授, 博导, 入选欧洲科学院外籍院士, 俄罗斯自然科学院外籍院士, 主要研究方向为智能感知与计算、图像理解与目标识别、深度学习与类脑计算。

刘芳, 女, 西安电子科技大学华山学者特聘教授, 博导, 享受国务院特殊津贴的专家, 国家自然科学基金获得者, 建国七十周年奖章获得者, 主要研究方向为大数据感知与模式识别、机器学习与智能图像处理。

李阳阳, 女, 教授, 华山学者特聘教授, 博导, 2012 年入选教育部“新世纪优秀人才支持计划”, 2014 年入选“陕西省科技新星”, 主要研究方向为计算智能、量子计算和图像处理。