

神经网络紧凑集成模式

汪庆华^{1,2}, 张优云¹, 朱永生¹

(1. 西安交通大学润滑理论及轴承研究所, 710049, 西安; 2. 西安工业大学机电学院, 710032, 西安)

摘要: 在分析现有神经网络集成构造过程的基础上,提出了一种神经网络紧凑集成模式,集成中成员网络的训练和网络组合权重的优化在同一个学习过程中进行,各参数的调整以提高集成泛化性能为目的.与现有神经网络集成模式相比,集成构造过程更加紧凑,它将个体网络生成阶段与结论合成阶段合二为一,并且网络之间的信息交互建立在实时动态的集成结构基础上,保证了成员网络训练与结论合成之间信息传递的始终一致性.为验证该模式的有效性及其优越性,采用4种典型的分类数据集对神经网络紧凑集成模式与CNNE、Bagging、Boosting等现有的集成模式在泛化性能上进行了比较,结果表明神经网络集成模式在测试数据集上的错误率降低了8%~16%.

关键词: 神经网络集成;紧凑集成模式;组合权重;泛化性能

中图分类号: TP18 **文献标识码:** A **文章编号:** 0253-987X(2007)03-0295-04

Compact Ensemble Model of Neural Networks

Wang Qinghua^{1,2}, Zhang Youyun¹, Zhu Yongsheng¹

(1. Theory of Lubrication and Bearing Institute, Xi'an Jiaotong University, Xi'an 710049, China;

2. School of Mechatronic Engineering, Xi'an Technological University, Xi'an 710032, China)

Abstract: Analyzing the construction process of the existing neural network ensembles, a novel compact ensemble model of neural networks is proposed. Training members in an ensemble and optimizing the combination weights for members are carried out simultaneously in the same learning process, and all parameters are adjusted to improve the generalization ability of the ensemble. In comparison with other existing ensemble model, the ensemble construction process is more compact by integrating the two stages into one, and the communication among neural networks is based on the real-time and dynamic structure of an ensemble so that the information conveyed between integration and training always keeps coincident. To validate the validity and advantage of this ensemble model, four well known classification problems are considered to compare the generalization error of compact ensemble with the generalization error of CNNE, Bagging, Boosting and other existing neural network ensemble model. The experimental results show that the error rate on testing sets can be decreased with compact ensemble model by 8% to 16%.

Keywords: neural network ensemble; compact ensemble model; combination weight; generalization ability

1990年,Hansen和Salamon^[1]开创性地提出了神经网络集成方法(neural network ensemble),他们证明可以简单地通过训练多个神经网络并将其结果进行合成,就能显著地提高神经网络的泛化能

力.构造神经网络集成的典型方法分为两步:首先训练一批神经网络,然后以某种方式组合成员网络的输出结果.因此,就产生了两大研究方向:即如何生成成员网络和如何合成成员网络的输出.1995年,

Krogh 和 Vedelsby 给出了神经网络集成泛化误差的计算公式^[2],揭示了神经网络集成中成员网络的精度、成员网络之间的相关程度与神经网络集成泛化误差的关系.目前,集成方法的研究主要集中在通过产生差异网络来增加集成的泛化性能,相关研究见文献[3-6].

现有的神经网络集成方法普遍存在这样一个现象:在生成差异性成员网络时,通常假设每个网络的组合权重一样,即成员网络之间的信息交互建立在固定的集成结构上,而在合成网络输出过程中却对成员网络的组合权重进行了修改,这样就造成训练过程与集成决策过程不一致的现象,势必会影响其泛化性能和稳定性.

本文提出一种新的神经网络集成模式——紧凑集成模式,该模式将传统集成模式的两步合二为一,训练成员网络和集成结构的确定在同一个过程中完成,即在同一个学习过程中同时优化成员网络结构参数和成员网络的组合权重.

1 神经网络紧凑集成模式

假设数据集 $D = \{(x(1), d(1)), \dots, (x(N), d(N))\}$ 由 N 个样本组成,神经网络集成 S 由 M 个成员网络 $s_1, \dots, s_M$ 组成,其组合权重表示为 $V_m (m = 1, \dots, M)$, $V_m \geq 0$ 且 $\sum_{m=1}^M V_m = 1$. 对于第 n 个样本输入,设第 m 个成员网络的输出为 $f_m(n)$,则神经网络集成 S 在该样本下的输出为

$$f(n) = \sum_{m=1}^M V_m f_m(n) \quad (1)$$

成员网络 s_m 在数据集上的泛化误差为

$$E_m = \frac{1}{N} \sum_{n=1}^N \frac{1}{2} (d(n) - f_m(n))^2 \quad (2)$$

神经网络集成 S 在数据集上的泛化误差为

$$E = \frac{1}{N} \sum_{n=1}^N \frac{1}{2} (d(n) - f(n))^2 \quad (3)$$

神经网络紧凑集成模式中,成员网络采用3层前向网络,其中隐含层与输出层取 log-sigmoid 激励函数,输入与输出层的单元数目根据输入数据集的属性个数来确定.与一般集成方法不同,在该模式集成中,成员网络的结构参数(神经元的连接权 $w_{i,j}$ 与阈值 θ_i)的优化和组合权重 V_m 的优化都是以降低神经网络集成输出误差为目的,即该算法对这些参数的调整都是以使集成输出的均方误差最小化.集成权重 V_m 的调整算法为

$$V_m(k+1) = V_m(k) - \eta \frac{\partial E}{\partial V_m} \quad (4)$$

成员网络结构参数的调整算法为

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \eta \frac{\partial E}{\partial w_{i,j}^m} \quad (5)$$

$$\theta_i^m(k+1) = \theta_i^m(k) - \eta \frac{\partial E}{\partial \theta_i^m} \quad (6)$$

式中: η 为学习步长; $m=1, \dots, M$. 因此,该紧凑集成模式是在同一个学习过程中同时训练所有成员网络并优化组合权重,在获得一批成员网络的同时也获得集成结构,其具体程序流程见图1.

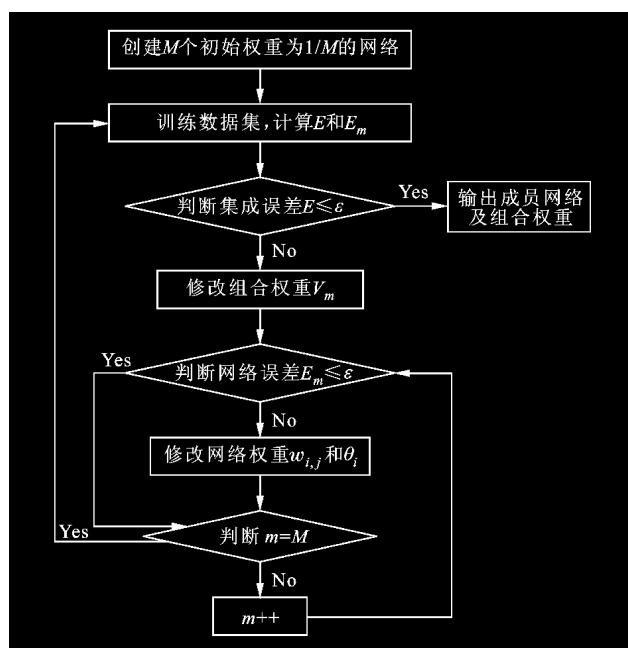


图1 神经网络紧凑集成模式流程图

2 实验与结果分析

2.1 实验设计

选择 Heart Disease(简称 Heart)、Pima Indians Diabetes(简称 Diabetes)、Australian Credit Card(简称 Australian)和 Breast Cancer(简称 Cancer)4 个数据集进行实验,它们均为两类分类问题,该数据集可从 UCI 数据集库获得^[7],数据集特征见表1.

表1 用于实验的数据集特征

数据集	样例数	输入属性个数	测试类别
Heart	270	14	分类
Diabetes	768	9	分类
Australian	690	15	分类
Cancer	683	11	分类

实验时,随机从原始样本集里抽取 3/5 的样本点作为训练集,2/5 为测试集,数据集样本点批量读入,运行次数为 10 次. 算法中部分参数设置如下: 最大循环次数设置为 2 000,对于 Heart 数据集学习步长为 0.3,Cancer 数据集学习步长为 1.5,其余两个为 0.7,成员网络个数为 3,成员网络为 3 层前向结构,隐含层与输出层取 log-sigmoid 激励函数,输入与输出层的单元数目根据输入数据集的属性个数来确定. 单个网络泛化误差和集成泛化误差为

0.005.

2.2 实验结果与比较

采用神经网络紧凑集成方法对 4 个数据集进行训练和测试的结果(错误识别率)分别见表 2 和表 3. 其中,均值是指训练集/测试集上 10 次运行结果的集成输出平均错误识别率,最小值是训练集/测试集上 10 次运行结果所得最小集成输出错误识别率,最大值是训练集/测试集上 10 次运行结果所得最大集成输出错误识别率.

表 2 训练集结果

数据集	均值	最小值				最大值			
		网络 1	网络 2	网络 3	集成	网络 1	网络 2	网络 3	集成
Heart	0.084 0	0.166 7	0.506 2	0.493 8	0.061 7	0.506 2	0.487 7	0.438 3	0.166 7
Diabetes	0.262 9	0.442 5	0.639 9	0.639 9	0.253 8	0.639 9	0.635 6	0.635 6	0.275 5
Australian	0.084 8	0.082 1	0.410 6	0.410 6	0.070 0	0.574 9	0.410 6	0.589 4	0.118 4
Cancer	0.011 5	0.322 0	0.012 2	0.136 6	0.009 8	0.322 0	0.322 0	0.031 7	0.012 2

表 3 测试集结果

数据集	均值	最小值	最大值
Heart	0.113 0	0.083 3	0.138 9
Diabetes	0.204 2	0.192 2	0.234 5
Australian	0.099 6	0.072 5	0.144 9
Cancer	0.011 4	0.007 3	0.014 7

从表 2 可以看出,神经网络紧凑集成模式主要强调的是集成泛化性能的提高,而不是成员网络的精确度,如在最小值一栏,对于 Heart 数据集,集成输出的错误识别率为 0.061 7,而 3 个成员网络的错误识别率却为 0.166 7、0.506 2、0.493 8,特别是后两个网络几乎没有分类能力,但它们的集成效果却很好. 另外 3 个数据集也是一样.

表 4~7 是将紧凑集成模式与其他集成模式在 4 个分类数据集上进行的比较. 需说明的是,在上述集成方法中,CNNE 和 EPNet 是 30 次运行结果的平均,CELS 是 25 次运行结果的平均,本文方法是 10 次运行的平均,其余方法是 5 次 10 重交叉验证的平均.

将文献[7]中的网络集成方法与本文方法进行比较的结果见表 4~7. 从这些表中,可以看出本文的神经网络紧凑集成模式在 Heart、Diabetes、Australian和 Cancer 4 个数据集上,其泛化性能较

目前的集成方法都好,其错误识别率在 Heart 数据集上较 CNNE 方法低近 16%,在 Diabetes 数据集上较 EENCL 方法低 8%,在 Australian 数据集上较 EPNet方法低 14%,在 Cancer 数据集上较 CNNE

表 4 Heart 数据集上各集成方法的错误识别率

算法	错误识别率
本文方法	0.113
CNNE	0.134
EPNet	0.167
Bagging	0.170
Arc-boosting	0.207
Ada-boosting	0.211

表 5 Diabetes 数据集上各集成方法的错误识别率

算法	错误识别率
本文方法	0.204
EENCL	0.221
EPNet	0.224
Bagging	0.228
Arc-boosting	0.244
Ada-boosting	0.233

表6 Australian数据集上各集成方法的错误识别率

算法	错误识别率
本文方法	0.099
EENCL	0.132
CELS	0.120
EPNet	0.115
Bagging	0.138
Arc-boosting	0.158
Ada-boosting	0.157

表7 Cancer数据集上各集成方法的错误识别率

算法	错误识别率
本文方法	0.011 4
CNNE	0.013 0
EPNet	0.014 0
Bagging	0.034 0
Arc-boosting	0.038 0
Ada-boosting	0.040 0

方法低12%。

神经网络紧凑集成模式在上述4个数据集上的表现比其他集成方法优越,分析其原因,大致有两个因素:第一,紧凑集成模式更加强调集成精度,集成中成员网络的参数优化与集成权重的优化都是以降低集成泛化误差为目的,EENCL、CELS、Bagging、Arc-boosting、Ada-boosting这些集成方法主要强调成员网络之间的差异性,EPNet虽然强调的不是成员网络的差异性,但它强调成员网络的精确性;第二,神经网络紧凑集成模式的信息交互建立在动态集成结构基础上,其组合权重优化过程和成员网络生成过程是在同一个学习过程中同时进行的,强调了在训练过程和集成决策过程中组合权重始终保持一致。其他集成方法在成员网络的训练时,由于网络的性能处于不断变化中,组合权重确定比较困难,因此都设定所有网络权重均相等,待成员网络训练过程结束,再确定集成权重。

3 结 论

本文提出一种神经网络紧凑集成模式,它将典型集成模式构造的两步合二为一,在同一个学习过程中同时优化成员网络结构参数和成员网络的组合权重,而且其优化过程是以降低集成泛化误差为目的。该紧凑集成模式在训练过程完成时即获得一批成员网络,并确定集成结构,改变了传统集成方法中成员网络生成过程与结论合成过程中组合权重不一致的现象。该模式算法结构紧凑,一致性好,增加了网络集成在输入空间样本输出的稳定性,集成效果优于传统集成模式。

参考文献:

- [1] Hansen L K, Salamon P. Neural network ensembles [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 1990, 12(10): 993-1001.
- [2] Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning [C] // Touretzky D, Leen T. Advance in Neural Information Processing Systems. Cambridge, USA: MIT Press, 1995: 231-238.
- [3] 周志华, 陈世福. 神经网络集成 [J]. 计算机学报, 2002, 25(1): 1-8.
Zhou Zhihua, Chen Shifu. Neural network ensemble [J]. Chinese J Computers, 2002, 25(1): 1-8.
- [4] Brown G, Wyatt J. Negative correlation learning and the ambiguity family of ensemble methods [C] // Windcatt T, Roli F. Proceedings of 4th International Workshop on Multiple Classifier Systems. Berlin: Springer-Verlag, 2003: 266-275.
- [5] Jang M, Cho S. Observational learning algorithm for an ensemble of neural networks [J]. Pattern Analysis and Applications, 2002, 5(2): 154-167.
- [6] 王正群, 陈世福, 陈兆乾. 优化分类型神经网络线性集成 [J]. 软件学报, 2005, 16(11): 1902-1908.
Wang Zhengqun, Chen Shifu, Chen Zhaoqian. An optimized neural network linear ensemble for classification [J]. Journal of Software, 2005, 16(11): 1902-1908.
- [7] Islam M M, Yao Xin. A constructive algorithm for training cooperative neural network ensembles [J]. IEEE Transactions on Neural Networks, 2000, 14(4): 820-834.

(编辑 荆树蓉)