

# 基于条件对数似然的 BP 神经网络多类分类器<sup>①</sup>

任 方, 马尚才

(山西财经大学 信息管理学院, 太原 030006)

**摘 要:** BP 神经网络分类器存在收敛速度慢的缺陷, 为了提高分类器性能, 针对这一缺陷对 BP 算法进行改进. 提出将条件对数似然(CLL)准则融入到监督性 BP 神经网络多类型分类过程中, 利用 CLL 的可分解性优势, 计算测试样本的条件概率, 在误差反向传播时利用条件概率对权值进行相应的加权降权操作, 简化误差反馈过程中的计算量. 在实验中对改进算法的收敛速度和准确率进行了测试, 说明了该算法的有效性及其实用性.

**关键词:** BP 神经网络; 条件对数似然; 多类分类器; 收敛速度; 监督性神经网络

## Multi-class BP Neural Network Classifier Based on the Conditional Log-Likelihood

REN Fang, MA Shang-Cai

(Department of Information Management, Shanxi University of Finance & Economics, Taiyuan 030006, China)

**Abstract:** BP neural network classifier has a slowly convergence rate, in order to improve the performance of the classifier, there is an improvement in BP algorithm for the problem. The Conditional Log-Likelihood (CLL) is applied into the supervisory neural network classification for the multi-class selection. By using the decomposability of CLL, calculate the conditional probability of the test samples. In the error back-propagation process, increasing or reducing the corresponding weights by using the conditional probabilities, which can simplify the computation in the process of error feedback. In the paper, we test the convergence speed and accuracy for the improved algorithm in the experiment. It illustrates the effectiveness and the practicality of the algorithm.

**Keywords:** BP neural network; conditional log-likelihood; multi-class classifier; convergence speed; supervisory neural network

BP 学习算法在人工神经网络中应用广泛, 目前在计算机科学、模式识别、机器人、预测等方面的应用中均取得了显著成效. 实验已经证明, 利用 BP 神经网络算法构造的分类器可以展现出很好的分类性能. BP 算法具有很好的非线性映射能力、泛化能力及容错能力, 但 BP 神经网络是基于误差反向传播机制, 存在收敛速度较慢、易陷入局部收敛等缺陷.

目前, 国内外学者已经针对这些缺陷做出一些研究. 常见的改进方法有优化初始权值、增加动量项和自适应学习. 董国君<sup>[1]</sup>等提出了一种基于随机退火机制的竞争层神经网络学习算法, 通过采用退火技术避免网络收敛到能量函数的局部极小点, 从而得到全局最优值; 胡洁<sup>[2]</sup>等提出一种快速且全局收敛的神经网络学习算法, 提高了 BP 算法的效率. Y.H.Zweiri<sup>[3]</sup>等人在动量项函数中再添加一项与误差成比例的项, 构成三项 BP 算法. Wei Gao<sup>[4]</sup>把 BP 神经网络与进化计算相结合, 设计出更高效的 BP 网络. 针对收敛速度慢这一缺陷, 本文提出一种基于 CLL(Conditional Log-Likelihood)准则的 BP 分类算法, 这一准则因其可分解性等优势, 在分类器中得到广泛应用, 将其应用于 BP 分类器, 可以有效提高 BP 分类器的性能.

1 BP网络分类器的结构和机理

BP 神经网络, 也即误差反向传播网络, 它是一种按误差逆传播算法训练的多层前馈网络. 一般由输入层、输出层与隐层三部分组成. BP 网络分类器采用三层结构, 其拓扑结构如图 1 所示. 网络的输入层节点

① 收稿时间:2013-10-19;收到修改稿时间:2013-11-08

数等于样本的属性个数, 输出层节点数等于类别数. BP 网络分类器在误差反馈机制下, 反馈信号会引起网络输出的不断变化, 当变化最后消失时, 网络达到平衡状态, 也即分类过程达到收敛. 在测试过程中, 当样品与曾学习过的样本十分相似时, 分类器就会做出正确的响应.

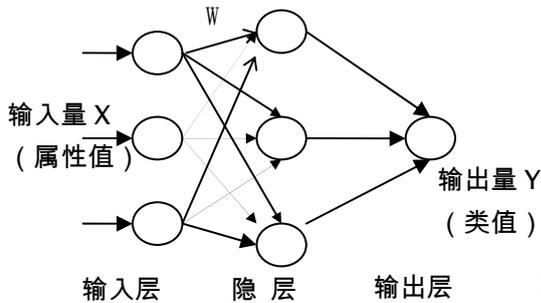


图 1 BP 神经网络拓扑结构

传统 BP 神经网络分类基本原理过程: (1)初始化: 对所有连接权  $W$  和阈值  $\theta$  赋予随机任意小值;

(2)从  $N$  组输入输出样本中取一组样本:  $X=[x_1, x_2, \dots, x_{m-1}]$  输入到网络中;

(3)正向传播: 计算各层节点的输出,  $Y_i^m=f(net), net=XW$ ;

(4)计算实际输出与相应的理想输出的误差:  $e_i = d_i - y_i^m (i = 1, \dots, p_m)$ ;

(5)反向传播: 按极小化误差的方式调整权矩阵;

(6)取出另一组样本重复(2)-(5)过程, 直到  $N$  组输入输出样本的误差达到要求时为止.

由上述过程可以看出, 标准的 BP 算法的权值及学习率的确定是随机的, 导致产生局部极小而得不到全局最优; 训练次数多学习效率低下, 产生收敛速度慢的问题; 在训练学习时存在覆盖趋势, 即学习新样本遗忘旧样本; BP 神经网络隐节点的选取也缺乏理论支持.

## 2 基于CLL的BP网络分类器

在传统 BP 网络分类器工作过程中, 假设有  $C_t$  个输出节点, 分类器一次分类循环过程后, 所产生的分类结果为 1 的只有一个, 而其余  $C_t-1$  个节点都为 0, 接下来的反向误差传播过程即会对所有  $C_t$  个节点进行, 直至此过程收敛. 在此过程中, 对结果为 0 的分类器也同样进行反向误差传播, 从而导致不易收敛、收敛速度慢. 针对这个问题, 本文提出将条件对数似然的

思想运用到分类中, 其基本原理是: 在每一次得到输出  $C_t^m$  后, 利用条件对数似然的概率思想, 计算每一个输出的概率, 从而获得  $C_t$  个不同的概率, 显然最终分类结果为低概率输出的可能性小, 而这种低概率输出的存在却大大增加了神经网络分类的工作量, 因此可以同样采用梯度下降的思想对其中的  $p$  个最小概率输出进行大力度降低权重, 同时也可以对概率极大的分类器增加权重, 在保证准确率甚至提高准确率的基础上, 使整个分类过程得到快速收敛, 在一定程度上加速 BP 算法的收敛速度.

首先对一次分类过程进行简要分析: 假设有  $M$  个输入节点, 即样本拥有  $M$  个属性,  $C_t$  个输出节点( $t>3$ ), 三个隐层节点. 传统采用 sigmoid 函数作为激励函数<sup>[5]</sup>, 若采用批型学习方法, 则网络误差函数为公式(1):

$$E(W) = \frac{1}{2 * M * C_t} \sum_{m=1}^M \sum_{t=1}^{C_t} (d^m - y^m)^2 \quad (1)$$

根据梯度下降算法, 我们可以得到求解(1)式最优权值的迭代算法:

$$W(k+1) = W(k) - \mu * \Delta E(k) \quad (2)$$

其中  $\mu > 0$  为学习率.  $o^m$  代表网络隐节点输出向量:

$$\frac{\partial E(W)}{\partial w_r} = \frac{1}{M * C_t} \sum_{m=1}^M (y^m - d^m) y^m (1 - y^m) o^m \quad (3)$$

$$\frac{\partial E(W)}{\partial w_i^1} = \frac{1}{M * C_t} \sum_{m=1}^M \sum_{t=1}^{C_t} (y^m - d^m) y^m (1 - y^m) w_i o^m (1 - o^m) x^m \quad (4)$$

最终, 从所有  $C_t$  个分类中选择出最大的网络输出作为分类结果. 实际上在网络的拓扑结构与学习数据确定后, 网络期望输出就已设定好了, 其实际输出则是由激励函数计算而来, 也即本文是基于监督性神经网络的学习. 因此, 由给定的数据集样本可以得知每个样本的先验概率及条件概率, 这就给一次分类结果后进行误差反向传播前的有效调整权重过程提供了契机, 即将  $C_t$  个分类结果根据条件概率进行相应的权重更改, 最终达到收敛过程加快的目标.

假设 BP 神经网络表示为  $G$ , 数据样本集为  $D$ , 样本  $X$  的网络期望输出为  $C_t$ , 由样本集可知其网络期望输出的概率为  $P(C_t)$ , 则对应的条件概率定义为  $P(C_t|C)$ . 对于给定的数据样本  $X=[x_1, x_2, \dots, x_{m-1}]$  和误差目标函数, 要做的工作是找到最大的输出  $Y^m$ . 在一次分类过程后, 假设其计算输出也即实际输出为  $C_1$ , 而网络期望输出为  $C_t (t \neq 1)$ , 则存在条件概率  $P(C_t|C_1)$ . 数据集  $D$  中

不同实例取值的数目定义为  $N_{ijk}$ , 则利用对数似然准则(LL)进行计算, 见公式(5):

$$LL(G|D) = \sum_{i=1}^N \log P(y_i^1, \dots, y_i^n, C_i) = \sum_{c=1}^s N_c \log \left( \frac{N_c}{N} \right) + \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \left( \frac{N_{ijk}}{N_{ijk}} \right) \quad (5)$$

但 LL 准则存在不可分解的缺点, 将其改进为 CLL<sup>[6]</sup>形式如下公式(6), 其中 B 表示重新构造形成的神经网络:

$$CLL(B|D) = \sum_{i=1}^N \log P(C_i | y_i^1, \dots, y_i^n) \quad (6)$$

对于一个已知样本  $X_t$ , 根据不同分类器计算出的输出为  $y_t^1, \dots, y_t^n$ , 则对应的概率可由公式(7)求得, 可将公式(7)简化表示为公式(8):

$$P(c_i | y_t^1, \dots, y_t^n) = \frac{P(y_t^1, \dots, y_t^n, c_i)}{P(y_t^1, \dots, y_t^n, c_1) + \dots + P(y_t^1, \dots, y_t^n, c_n)} \quad (7)$$

$$P(c_i | y_t^1, \dots, y_t^n) = \frac{U_i}{U_1 + \dots + U_i + \dots + U_n} \quad (8)$$

在此, 利用对数似然的思想, 将概率转化为对数形式, 表示为:

$$f(U_1, \dots, U_i) = \log \left( \frac{U_i}{U_1 + \dots + U_i + \dots + U_n} \right) \quad (9)$$

通过 f 函数的计算, 可得出每个分类器的分类结果概率, 根据梯度下降思想对概率较低的分类按照规则进行降权, 对概率高的分类增加权重, 在不断的优化中加速网络中的反向误差传播过程. 这个计算概率的过程采用了条件对数似然思想, 计算简便, 时间复杂度为  $O(n)$ , 相较于传统反向误差传播的时间复杂度要低很多, 因此具有研究意义.

### 3 实验

#### 3.1 实验环境

本实验在机器学习平台 weka 环境下进行测试和分析, 所采用的数据来源于 UCI 数据仓库中的 20 个数据集, 网址为 <http://kdd.ics.uci.edu/>. 表 1 简单描述了这 20 个数据集的名称、实例个数、属性个数及类个数, 数据集的分类数均大于 3, 其中 letter 数据集最大, 分类数也最多. 针对这 20 个数据集, 分别采用传统的 BP 神经网络分类器和改进的 BP 网络分类器进行实验测试, 对实验结果中的收敛速度及分类准确率进行比较分析.

改进的 BP 算法的伪码简单如下:

1)初始化每个属性所对应权值  $W(k)$ 及阈值  $\theta$ 、 $\mu$ .

2)定义误差函数初值  $E$ :  $E=0+1$ .

3)在阈值范围内进行误差反向传播.

3.1  $E=0$ ;

3.1.1 对 D 中的每一个样本  $(X_n, Y_n)$ : 根据公式(6), 计算出测试样本的条件概率  $P_n$ . 将  $X_n$  的属性值与相应的权值进行乘积运算, 得到输出  $O_n$ ; 根据网络误差函数, 即公式(1)计算出对应的  $E_n$ ;

3.1.2 计算输出值:  $E=E+E_n$ ;

3.1.3 根据 CLL 准则下计算出的样本对应分类概率, 对取值  $P_{max}$  的概率大的分类输出  $C_{max}$  对应的  $W$  加权, 同时对取值  $P_{min}$  的概率小的分类输出  $C_{min}$  对应的  $W$  进行降权. 加权降权操作主要结合梯度下降的思想, 对公式(2)进行变形, 则加权降权操作分别为:

$$W(k+1) = W(k) - \mu * \Delta E(k) * (P_{min} / P_{max})$$

$$W(k+1) = W(k) - \mu * \Delta E(k) * (P_{max} / P_{min})$$

3.2 进入下一轮误差反向传播过程:  $E=E/2.0$ .

表 1 实验数据集描述

序号	数据集	实例	属性	类值
1	anneal	898	38	6
2	artificial	5109	7	10
3	audiology	226	69	24
4	autos	205	25	7
5	bridges2	108	11	6
6	flag	194	27	6
7	glass	214	9	7
8	heart-c	303	13	5
9	heart-h	294	13	5
10	hypothyroid	3772	29	4
11	letter	20000	16	26
12	lymph	148	18	4
13	machine	209	7	8
14	page-blocks	5473	10	5
15	primary-tumor	339	17	22
16	segment	2310	19	7
17	solar-flare	333	10	8
18	soybean	687	35	19
19	vowel	990	13	11
20	zoo	104	17	7

#### 3.2 实验结果及分析

实验收集了算法改进前后的收敛速度和准确率数据, 见表 2 所示.

表 2 算法改进前后分类器实验比较结果

序号	数据集	传统 BP 算法		改进 BP 算法	
		收敛速度	准确率	收敛速度	准确率
1	anneal	1.53s	0.7825	1.46s	0.7901
2	artificial	2.03s	0.4529	1.92s	0.5964
3	audiology	0.98s	0.6376	0.95s	0.6564

4	autos	0.83s	0.5136	0.81s	0.5763
5	bridges2	0.76s	0.7187	0.71s	0.7165
6	flag	0.85s	0.5289	0.76s	0.5325
7	glass	0.92s	0.5392	0.89s	0.5568
8	heart-c	1.05s	0.7344	0.96s	0.7368
9	heart-h	1.03s	0.6482	0.94s	0.6791
10	hypothyroid	1.86s	0.8134	1.80s	0.8006
11	letter	5.62s	0.8961	4.92s	0.9456
12	lymph	0.75s	0.8501	0.55s	0.8513
13	machine	1.05s	0.8688	0.92s	0.8733
14	page-blocks	2.29s	0.9363	2.04s	0.9376
15	primary-tumor	1.34s	0.7766	1.20s	0.8223
16	segment	1.79s	0.8347	1.76s	0.8425
17	solar-flare	1.37s	0.8762	1.29s	0.8812
18	soybean	1.46s	0.9084	1.42s	0.9425
19	vowel	1.73s	0.7944	1.68s	0.8512
20	zoo	0.52s	0.7364	0.48s	0.7276

从实验结果发现,改进的BP网络分类器的收敛速度明显提高,对于分类数少的数据集,改进分类器的收敛速度有所提高,但不是很明显,对于分类数越多的数据集,改进分类器的分类收敛速度提高则越明显,如图2所示.实验在提高了收敛速度的基础上发现,改进分类器的准确率也有所提高,对于分类数少的数据集,改进分类器的准确率有提高,但也有少数数据集的分类准确率略微降低,但对于分类数多的数据集,则准确率提高明显,如图3所示.综上所述,相较于传统BP分类器,改进的BP网络分类器的分类收敛速度加快,分类准确率有所提升,而该分类器在多类型分类的应用中可以获得更加显著的良好性能.

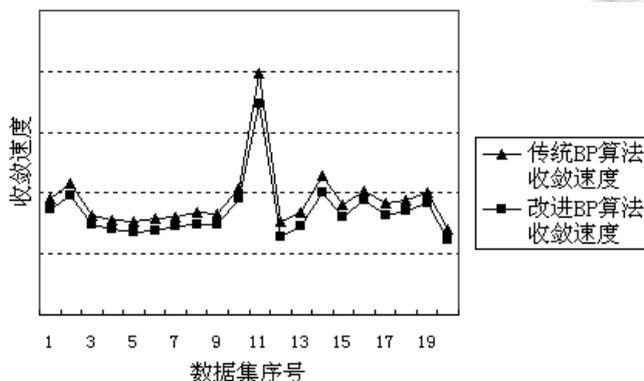


图2 BP算法改进前后收敛速度比较

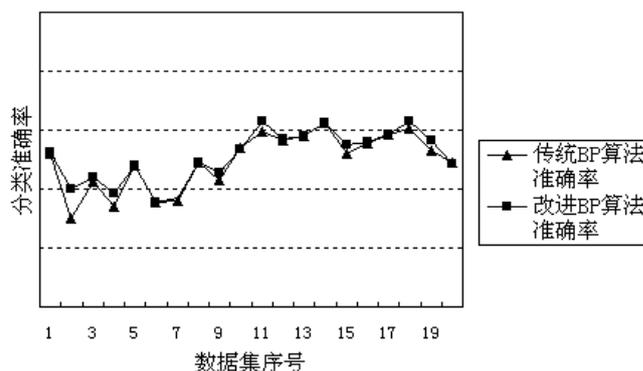


图3 BP算法改进前后分类准确率比较

#### 4 结语

BP网络分类器是基于BP算法的,而BP算法有其自身的缺点,本文主要就收敛速度慢的问题进行改进,利用对数似然能够在数量级上降低计算量的优势,将其应用到每次分类结束后的实际输出的删减工作中,从而在一次循环后,可以实现降低神经网络复杂度、减少计算量的效果,并增强了网络的更精确分类能力.

#### 参考文献

- 董国君,哈力木拉提.基于随机退火的神经网络算法及其应用.计算机工程与应用,2010,19.
- 胡洁,曾祥金.一种快速且全局收敛的BP神经网络学习算法.系统科学与数学.2010,30(5):604-610.
- Zweiri YH, Whidborne JF, Seneviratne LD. A three-term back-propagation algorithm. Neurocomputing, 2003, 50: 305-318.
- Gao W. Study on new evolutionary neural network. Proc. of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2003,11: 2-5.
- 毕小梅,陈建斌.修正激励函数对标准BP算法的改进.算法分析,2012,2(7):100-102.
- Carvalho AM, Roos T, Oliveira AL. Discriminative learning of bayesian networks via factorized conditional log-likelihood. Journal of Machine Learning Research, 2011, 12: 2181-221.