

优化分类型神经网络线性集成*

王正群^{1,2+}, 陈世福¹, 陈兆乾¹

¹(计算机软件新技术国家重点实验室(南京大学),江苏 南京 210093)

²(扬州大学 信息工程学院,江苏 扬州 225009)

An Optimized Neural Network Linear Ensemble for Classification

WANG Zheng-Qun^{1,2+}, CHEN Shi-Fu¹, CHEN Zhao-Qian¹

¹(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210093, China)

²(School of Information Engineering, Yangzhou University, Yangzhou 225009, China)

+ Corresponding author: Phn: +86-25-83593163, Fax: +86-25-83593163, E-mail: yzwzq@yzcn.net, <http://www.nju.edu.cn>

Received 2004-03-18; Accepted 2005-02-04

Wang ZQ, Chen SF, Chen ZQ. An optimized neural network linear ensemble for classification. *Journal of Software*, 2005,16(11):1902–1908. DOI: 10.1360/jos161902

Abstract: Neural network ensembles have been used increasingly in recent years to improve classifier's generalization ability. There are several methods of designing neural network ensembles, such as weighted average linear ensemble. The output of the weighted average linear ensemble is a weighted average of the output of each component neural network, with the weights determined by a function. Based on the characteristic of the classification issue, a function is defined, which is the ratio of the pattern separability within class to the pattern scatter between classes. The minimum of the function corresponds to the optimal weights, so an optimal linear ensemble is obtained. The optimal weights are searched by a genetic algorithm. The rationale behind the function is also analyzed, showing that it accords with the Bayesian decision rule. Finally, to estimate the performance of this linear ensemble, two handwritten Chinese character feature sets and four data sets from UCI machine learning depository are used. Empirical study shows that the optimal linear ensemble method can produce ensemble of the neural network with a stronger generalization ability.

Key words: neural network; neural network ensemble; classifier; genetic algorithm; optimization

摘要: 构造多神经网络集成系统,系统的输出由个体神经网络的输出线性加权产生.提出了一种度量个体神经网络在不同的权重下集成性能的判别函数,函数表示了由个体神经网络输出刻画的模式类内会聚性和类间散布性.应用遗传算法解决了求解最优个体网络集成权重问题.分析了该判别函数的合理性及其与 Bayes 决策规则的关系.

* Supported by the National Natural Science Foundation of China under Grant No.602730333 (国家自然科学基金); the Natural Science Foundation of Jiangsu Province of China under Grant No.BK2004079 (江苏省自然科学基金); the Natural Science Foundation of Yangzhou University of China under Grant No.KK0413160 (扬州大学自然科学基金)

作者简介: 王正群(1965 -),男,江苏如东人,博士,副教授,主要研究领域为机器学习,神经网络,模式识别;陈世福(1938 -),男,教授,博士生导师,主要研究领域为机器学习,知识工程,分布式人工智能,图像处理;陈兆乾(1940 -),女,教授,博士生导师,主要研究领域为机器学习,知识工程,神经网络.

用两个手写体汉字特征数据集和4个UCI数据库中的数据集比较了所提出的神经网络集成方法和其他几种神经网络集成方法的性能。

关键词: 神经网络;神经网络集成;分类器;遗传算法;优化

中图法分类号: TP183 文献标识码: A

自20世纪80年代以来,神经网络吸引了大批研究人员的注意.然而,由于缺乏严密理论体系的指导,神经网络技术的应用效果完全取决于使用者的经验,即在解决真实世界问题时,由于缺乏问题的先验知识,往往需要经过大量费力耗时的实验摸索才能确定合适的神经网络模型、算法以及参数设置,其使用效果完全取决于使用者的经验.

1990年,Hansen和Salamon^[1]开创性地提出了神经网络集成方法.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,显著地提高神经网络系统的泛化能力.由于该方法易于使用且效果明显,即使是缺乏神经计算经验的普通工程技术人员也可以从中受益,因此它被视为一种非常有效的工程化神经计算方法.

1 相关工作

1996年,Sollich和Krogh^[2]为神经网络集成下了一个定义,“神经网络集成是用有限个神经网络对同一个问题进行学习,集成在输入示例下的输出由构成集成的各神经网络在该示例下的输出共同决定”.由于认识到神经网络集成所蕴涵的巨大潜力和应用前景,很多研究者都进行了这方面的研究.

对神经网络集成的研究主要集中在两方面,即如何生成集成中的个体神经网络以及如何将多个个体神经网络的结论进行结合.神经网络集成的应用也主要集中在两方面,即回归估计问题和分类识别问题.

理论和实验研究表明^[1,3],神经网络的泛化性能极大地信赖集成中个体神经网络的精度和差异性.为了保持个体网络之间的差异性,在生成个体神经网络时,常采取以下方法^[4]:(1)不同的个体神经网络采用不同的初始权重;(2)不同的个体神经网络采用不同的拓扑结构;(3)不同的个体神经网络采用不同的训练数据;(4)不同的个体神经网络采用不同的训练算法.以上这些方法对个体神经网络的训练都是单独进行的,训练过程中它们相互之间没有信息交互.Liu等人^[5,6]认为,对于神经网络集成,个体神经网络之间的不相关性,甚至负相关性比它们之间的差异性更重要.对于回归估计问题,Liu等人^[5]提出了个体神经网络负相关集成学习方法,使个体神经网络的输出与其余个体神经网络集成的输出负相关,从而降低了个体神经网络集成的泛化误差;Jang等人^[6]提出了神经网络集成观察学习方法,个体神经网络在训练过程中,观察其他个体神经网络的输出,决定自身均方误差的调整幅度.这两种神经网络集成方法,所有参与集成的个体神经网络同时训练,训练过程中个体神经网络之间交换信息,以集成泛化误差的降低作为最终目标.

对于回归估计问题,神经网络集成常采用线性集成方法,即集成的输出由个体神经网络的输出通过简单平均^[5,6]或加权平均产生;对于分类识别问题,多个神经网络结论结合采用投票方法^[7](绝对多数投票法和相对多数投票法),也采用线性集成方法.在线性集成方法中,加权平均集成又分为两类,一类个体神经网络的集成权重由训练样本确定^[8],另一类由它们的输出结果确定^[9],因此这一类神经网络集成又称为动态集成.

为了降低参与集成的神经网络数目,提高神经网络集成的泛化性能,Zhou等人^[7,10]提出了选择性神经网络集成方法,即从训练好的一批个体神经网络中,选择部分神经网络进行集成.Zhou等人^[7]的神经网络集成方法,运用遗传算法确定个体神经网络的权重,权重反应个体神经网络在集成中的重要性,根据个体神经网络的权重选择部分神经网络参与集成;Giacinto等人^[10]的聚类神经网络集成方法,根据个体神经网络的输出对个体神经网络进行聚类,从每类中各选出一个个体神经网络参与集成;Bakker等人^[11]的聚类神经网络集成方法,也根据个体神经网络的输出对个体神经网络进行聚类,最后根据聚类的结果重新构造一个个体神经网络参与集成.这几种方法都取得了较好的集成效果.

对神经网络集成进行定量和定性分析,是估计神经网络集成性能的两个方面.对于回归估计问题,均方误差、方差是定量分析网络集成的重要指标.Carney等人^[12]依据个体神经网络的性能给出了神经网络集成的信

度区间和预测区间,对神经网络集成性能进行了定性分析.对于分类识别问题,Cunningham 等人^[13]用不同数量的训练数据和验证数据,定量分析了神经网络和神经网络集成的稳定性问题;Tumer^[14]以最小错误率 Bayes 决策规则为依据,分析了个体神经网络的判决区域与神经网络集成判决区域的关系,在特定条件下,随着网络集成中个体神经网络数量的增加,网络集成的判决区域趋向于最小错误率 Bayes 决策区域.

2 优化分类型神经网络集成

对于分类识别问题,神经网络被认为是从特征空间到类标签空间的映射装置.本文中假设个体神经网络为二层前馈型神经网络,用 Delta 学习规则训练成分类器.输入层神经元的个数与模式特征向量的维数相同.输出层神经元的个数与模式类别数相同.在对神经网络训练时,如果训练样本来自第 i 类,则设定输出层第 i 个神经元的输出为 1,其他神经元的输出为 0.输入某个模式的特征向量,神经网络的输出是对模式所属类别的后验概率估计^[15],因此采用最小错误率 Bayes 决策规则对模式进行分类决策.

2.1 问题的提出

分类型神经网络集成大多采用多数投票法和线性集成法.多数投票法把个体神经网络对待识模式所属类别的后验概率估计转化为对待识模式所属类别的确定判断,依据多数个体神经网络对模式所属类别的判断确定模式所属类别.在线性神经网络集成方法中,简单平均法通过个体神经网络对待识模式所属类别后验概率估计的平均产生集成对待识模式所属类别后验概率估计,再采用最小错误率 Bayes 决策规则对模式进行分类决策.

由于个体神经网络对模式所属类别后验概率的估计存在不确定性,因而对模式所属类别的判断存在不确定性,多数投票法根据其他个体神经网络的判断校正这种不确定性,但丢掉了个体神经网络对待识模式所属类别的后验概率估计信息.简单平均法利用了个体神经网络对待识模式所属类别的后验概率估计信息,但没有对模式所属类别的后验概率估计的确定性进行衡量.

在线性神经网络集成方法中,加权平均法不仅利用了个体神经网络对待识模式所属类别的后验概率估计信息,还利用权重对模式所属类别的后验概率估计的确定性进行了衡量.

对于输入模式,设第 l ($l=1,2,\dots,N$) 个个体神经网络的输出为 $F_l=(f_{l1},f_{l2},\dots,f_{lc})$,其中, f_{li} ($i=1,2,\dots,C$) 表示第 i 个输出神经元的输出值.加权平均线性集成的输出表示为 $\bar{F}=(\bar{f}_1,\bar{f}_2,\dots,\bar{f}_c)$,其中 $\bar{f}_i=\sum_{l=1}^N w_l f_{li}$.设集成系统的理想输出为 $D=(d_1,d_2,\dots,d_c)$,加权平均神经网络集成通过极小化均方差函数 $E(\|\bar{F}-D\|)$ 来确定个体网络集成的权重^[14].对于训练模式,由于已知类别信息,如果它来自第 i 类,设定 $d_i=1, d_j=0(j \neq i)$,从而由求极值算法得到权重^[8].

以第 1 类模式为例,极小化均方差函数使集成系统对第 1 类模式判决的输出与点 $(1,0,\dots,0)$ 的距离最小. $(1,0,\dots,0)$ 是第 1 类模式的代表.上述确定权重的方法只考虑了同类模式之间的会聚性,没有考虑不同类模式之间的散布性.下面重新给出一个度量个体神经网络集成权重性能的判别函数.

设已知第 i 类模式的代表为 D ,输入第 i 类模式,集成系统的输出为 \bar{F} ;第 j 类模式的代表为 D_j .定义判别函数:

$$J(W) = \frac{S_w}{S_B} \quad (1)$$

满足:

$$0 \leq w_l \leq 1 \quad (2)$$

$$\sum_{l=1}^N w_l = 1 \quad (3)$$

其中, $S_w = E(\|\bar{F} - D\|)$, $S_B = E\left(\frac{1}{C-1} \sum_{\substack{j=1 \\ j \neq i}}^C \|\bar{F} - D_j\|\right)$. 式(2)、式(3)保证集成的输出值范围与个体神经网络输出值范围相同.

2.2 理论分析

把神经网络集成的输出看作 C 维空间的一个向量.上一节给出的判别函数是 S_w 与 S_B 之比. S_w 是模式到它所属模式类代表距离的期望, S_B 是模式到其他模式类代表平均距离的期望. S_w 反映了同类模式之间的会聚性, S_B 反映了不同类模式之间的散布性.因此,判别函数的最小值对应着线性集成的最优权重.由最优权重确定的神经网络集成称为最优神经网络集成(optimal neural networks ensemble,简称 ONNE).

判别函数 $J(W)$ 根据模式到模式类代表的距离度量个体网络权重对网络集成的影响,最优权重对应着判别函数的最小值,因而对于待识模式,加权平均神经网络线性集成系统合理的决策规则为:

规则 1. 如果 $\|\bar{F} - D_i\| = \min_{j=1,2,\dots,C} \{\|\bar{F} - D_j\|\}$, 则模式属于第 i 类.

个体神经网络的输出是对后验概率的估计^[15],在限制条件(式(2),(3))下,神经网络加权平均线性集成也是对后验概率的估计,按照 Bayes 决策理论,集成系统合理的决策规则为:

规则 2. 如果 $\bar{f}_i > \bar{f}_j$ ($j \neq i, j = 1, 2, \dots, C$), 则模式属于第 i 类.

如果集成系统根据规则 1 对模式分类识别,则集成系统是最小距离分类器;如果集成系统根据规则 2 对模式分类识别,则集成系统是最小错误率 Bayes 分类器.下面的定理说明了这两个分类器的等价性.

定理. 规则 1 与规则 2 等价.

证明:首先证明规则 1 蕴含规则 2.

显然, $\|\bar{F} - D_i\| = \min_{j=1,2,\dots,C} \{\|\bar{F} - D_j\|\} \Leftrightarrow \|\bar{F} - D_i\| < \|\bar{F} - D_j\|$ 对一切 $j \neq i$ 成立.

向量 D_i 的第 i 个分量为 1,其余分量为 0;向量 D_j 的第 j 个分量为 1,其余分量为 0,因而 $\|\bar{F} - D_i\| < \|\bar{F} - D_j\| \Leftrightarrow (\bar{f}_i - 1)^2 + (\bar{f}_j)^2 < (\bar{f}_i)^2 + (\bar{f}_j - 1)^2$.

而 $(\bar{f}_i - 1)^2 + (\bar{f}_j)^2 < (\bar{f}_i)^2 + (\bar{f}_j - 1)^2 \Rightarrow \bar{f}_i > \bar{f}_j$.

所以,规则 1 蕴含规则 2.

类似可证,规则 2 蕴含规则 1.

根据以上分析,由第 2.2 节的判别函数确定集成中个体神经网络的最优权重,再由规则 1 或规则 2 对模式进行分类识别,集成系统性能最好.

2.3 最优权重求解

最优神经网络集成的个体网络权重与判别函数的最小值对应.一般求解函数极值方法求得的函数极值是局部极值.而遗传算法对优化问题具有全局寻优能力,被广泛用于解决优化问题^[7].我们用遗传算法求解最优神经网络集成的个体网络权重.

遗传算法的实现采用了 Houck 等人^[16]开发的 GATO 工具箱.遗传算法的选择算子、交叉算子、变异算子、交叉概率、变异概率、停止准则等系统参数均采用 GATO 工具箱的缺省设置.

个体网络权重采用实数编码方式.假设某遗传个体与 $W = (w_1, w_2, \dots, w_N)$ 相对应,验证集 V 中 m_i ($i = 1, 2, \dots, C$) 个样本来自第 i 类,计算:

$$\hat{f}(W) = C_0 - \frac{\hat{S}_w}{\hat{S}_B} \tag{4}$$

其中, $\hat{S}_w = \frac{1}{C} \sum_{i=1}^C \left(\frac{1}{m_i} \sum_{k=1}^{m_i} \|F_k - D_i\| \right)$, $\hat{S}_B = \frac{1}{C} \sum_{i=1}^C \left(\frac{1}{m_i} \sum_{k=1}^{m_i} \left(\frac{1}{C-1} \sum_{j \neq i}^C \|F_k - D_j\| \right) \right)$, F_k 是输入第 i 类中第 k 个样本后网络

集成的输出. C_0 是设定的常数, 保证 $\hat{f}(W)$ 为正值. 只要每个个体网络对验证集的识别正确率超过 50%, 就有 $\hat{S}_w < \hat{S}_B$, 因而设定 $C_0 = 1$.

以 $\hat{f}(W)$ 作为遗传个体的适应度. 由于存在限制条件 $\sum_{l=1}^N w_l = 1$, 从遗传个体中分离出个体网络权值 w_l 之后, 需进行标准化处理:

$$w_l \leftarrow \frac{w_l}{\sum_{l=1}^N w_l} \quad (5)$$

3 实验比较

实验中采用了 5 倍交叉验证技术, 实验结果是 5 次实验结果的平均值. 把样本集分成 5 等分, 每次用其中 1 等分作为测试集来测试网络集成的性能, 其余的 4 等分(集成原始训练集)用于设计网络集成. 网络集成规模设定为 20.

Bagging 算法^[17]是生成神经网络集成中个体神经网络的一种常用技术, 它的基础是可重复取样(bootstrap sampling). 个体神经网络的训练集从原始训练集中随机选取若干示例组成. 基本集成方法^[9](basic ensemble method, 简称 BEM)中集成的输出由个体神经网络输出通过简单平均生成, 而多数投票法(majority voting)^[7]神经网络集成把个体神经网络的输出转换成对待识模式的确定判决, 由多数个体神经的判决产生集成判决. Jimenez^[9]提出了一种神经网络动态权重线性集成方法(dynamically averaged network, 简称 DAN), 个体神经网络权重由个体神经网络的输出确定. 实验中比较了神经网络集成基本集成方法、动态权重线性集成方法、多数投票法和本文提出的神经网络集成方法 ONNE.

为了使个体网络性能具有一定的差异性, 采用可重复取样技术^[17]从集成原始训练集中生成个体神经网络的训练集, 训练集的规模是集成原始训练集的一半. 用于确定集成最优权重的验证集也由集成原始训练集采用可重复取样技术产生.

3.1 实验数据

实验中使用了两部分数据集. 第 1 部分数据集取自南京理工大学手写体汉字样本库 NUST603HW. 该样本库包含银行支票上金额大写常用的 19 个汉字的样本, 每类取 400 个样本, 分别抽取它们的交叉数特征(cross number)^[18]和周边特征(periphery feature)^[19]用于实验. 第 2 部分数据集取自 UCI 机器学习数据库. 该数据库被广泛用于测试分类器性能. 在 Heartdisease 数据集中, 绝大多数数据缺少后 3 个特征, 在实验中去掉了这些特征. 缺少其他特征的数据不包括在本文实验中. 对于 Letter 数据集, 我们选取了其中的 5 000 个样本用于本文的实验.

这两部分数据集中与本文实验有关的属性参数见表 1.

Table 1 The test data sets

表 1 测试数据集

Data set	Classes	Features	Size
NUST603HW	Cross number	128	7 600
	Periphery feature	128	7 600
UCI repository of machine learning	Heartdisease (Hungarian)	2	261
	Letter	26	5 000
	Waveform	3	5 000
	Glass	6	214

3.2 实验结果

在神经网络集成中,对于分类识别问题,由于难以用表达式表示个体神经网络输出结论之间的关系^[7],个体神经网络输出结论大多采用多数投票法进行结合^[1,7].本文把神经网络看作是数据的映射装置,根据个体神经网络的输出结果,用一个表达式(式(1))表示个体神经网络输出结论之间的关系,进而提出了一种最优神经网络线性集成方法 ONNE.

表 2 给出了不同的神经网络集成方法在几个数据集上的分类识别结果(以识别正确率表示).本文的神经网络集成方法平均正确率比神经网络集成方法 BEM 平均高 2.31%,比 DAN 平均高 1.56%,比多数投票法平均高 1.79%.这表明本文的神经网络集成方法 ONNE 是有效的.由表 2 还发现,动态权重线性集成方法 DNA 泛化性能不稳定,一般能取得比多数投票法更好的集成效果,但有时比基本集成方法集成效果还差,而本文的神经网络集成方法 ONNE 总能取得较好的集成效果.

Table 2 Experiment on comparing Bagging and majority voting with ONNE

表 2 Bagging 方法、多数投票法和 ONNE 方法的比较实验

Data set	NUST603HW		UCI repository of machine learning			
	Cross number	Periphery feature	Heart-Disease	Letter	Wave-Form	Glass
BEM	85.34	87.88	76.90	79.25	86.15	87.88
DAN	87.21	87.61	77.34	81.06	86.78	87.87
Majority voting	86.66	87.67	77.55	80.97	85.33	88.32
ONNE	88.91	88.49	79.76	82.45	87.12	90.53

4 结束语

目前,神经网络集成技术的研究主要集中在两方面,即个体神经网络生成方法和个体神经网络集成结论结合方法.本文对个体神经网络集成结论结合方法进行了研究,根据分类型神经网络的特点,提出了一种优化神经网络线性集成方法 ONNE,既考虑了同类模式之间的会聚性,又考虑了不同模式类之间的散布性.神经网络集成的输出是对模式所属类别的后验估计,因而,它的决策基于最小错误率 Bayes 决策规则.

在进一步的神经网络集成技术研究中,我们将对如下问题进行研究:

- (1) Zhou 等人^[7]基于多数投票法提出一种选择部分神经网络参与集成的方法,取得了很好的分类识别效果.本文的神经网络集成原理不同于多数投票法.基于本文的神经网络集成原理,提出一种新的选择部分神经网络参与集成方法,这是我们下一步要做的工作.
- (2) 对于回归性神经网络集成,Liu 等人^[5]提出了负相关个体神经网络训练方法.个体神经网络训练以降低集成系统的泛化误差为目的.对于分类型神经网络,集成系统的正确识别率与个体网络性能联系的解析表达式难于给出,但本文给出的表达式(3),刻画了个体网络的输出性能与其他网络输出性能之间的关系,基于表达式(3),研究个体网络的生成方法,也是我们下一步的工作.

致谢 感谢南京理工大学 603 教研室提供了手写体汉字符本库 NUST603HW;感谢 Andras Janosi 收集了 Heart-disease 数据集,David J. Slate 提供了 Letter 数据集,David Aha 提供了 Waveform 数据集,Vina Spiehler 提供了 glass 数据集.

References:

- [1] Hansen LK, Salamon P. Neural network ensembles. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1990,12(10): 993-1001.
- [2] Sollich P, Krogh A. Learning with ensembles: How over-fitting can be useful. In: Touretzky D, Mozer M, Hasselmo M, eds. Advances in Neural Information Processing Systems 8. Cambridge: MIT Press, 1996. 190-196.
- [3] Hashem S. Optimal linear combinations of neural networks. Neural Networks, 1997,10(4):599-614.
- [4] Sharkey AJC. On combining artificial neural nets. Connection Science, 1996,8(3-4):299-314.
- [5] Liu Y, Yao X. Ensemble learning via negative correlation. Neural Networks, 1999,12(10):1399-1404.

- [6] Jang M, Cho S. Observational learning algorithm for an ensemble of neural networks. *Pattern Analysis and Applications*, 2002,5(2): 154–167.
- [7] Zhou ZH, Wu JX, Tang W. Ensembling neural networks: many could be better than all. *Artificial Intelligence*, 2002,17(1-2): 239–263.
- [8] Benediktsson JA, Sveinsson JR, Ersoy OK. Optimized combination of neural networks. In: *Proc. of the IEEE Int'l Symp. on Circuits and Systems*. Atlanta, 1996. 535–538.
- [9] Jimenez D. Dynamically weighted ensemble neural network for classification. In: *Proc. of the IEEE Int'l Joint Conf. on Neural Networks*. Anchorage, 1998. 753–756.
- [10] Giacinto G, Roli F. Design of effective neural network ensembles for image classification purposes. *Image and Vision Computing*, 2001,19(9-10):699–707.
- [11] Bakker B, Heskes T. Clustering ensembles of neural network models. *Neural Networks*, 2003,16:261–269.
- [12] Carney J G, Cunningham P. Confidence and prediction intervals for neural network ensembles. In: *Proc. of the Int'l Joint Conf. on Neural Networks*. Washington, 1999. <http://www.cs.tcd.ie/Padraig.Cunningham/online-pubs.html>
- [13] Cunningham P, Carney J, Jacob S. Stability problems with artificial neural networks and the ensembles solution. *Artificial Intelligence in Medicine*, 2000,20(3):217–225.
- [14] Tumer K, Ghosh J. Analysis of decision boundaries in linearly combined neural classifiers. *Pattern Recognition*, 1996,29(2): 341–348.
- [15] Cho SB, Kim JH. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Trans. on SMC*, 1995,25(2): 380–384.
- [16] Houck CR, Joines JA, Kay MG. A genetic algorithm for function optimization: A matlab implement. Technical Report, NCSU-IE-TR-95-09, Raleigh: North Carolina State University, 1995.
- [17] Breiman L. Bagging predictors. *Machine Learning*, 1996,24(2):123–140.
- [18] Tseng YH, Kuo CC, Lee HJ. Speeding up Chinese character recognition in an automatic reading system. *Pattern Recognition*, 1998, 31(11):1601–1612.
- [19] Tang YT, Tu LT, Lee SW, Lin WW, Shyu LS. Offline recognition of Chinese handwriting by multifeature and multilevel classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998,20(5):556–561.