神经网络
Neural Networks
第八章

神经网络集成

史忠植

中国科学院计算技术研究所
http://www.intsci.ac.cn/

内容提要

- 8.1 概述
- 8.2 神经网络集成的基本原理
- 8.3 神经网络集成的方法
- 8.4 结论生成方法
- 8.5 个体生成方法
- 8.6 基于Bagging的聚类
- 8.7 神经网络集成系统的规则获取
- 8.8 神经专家系统

集成学习

1990年汉森(L.K. Hansen)和萨拉蒙(P. Salamon)提出了神经网络集成(neural network ensemble)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拟合，显著地提高神经网络系统的泛化能力。神经网络集成可以定义为用有限个神经网络对同一个问题进行学习，集成在某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。

集成学习

Boosting是一大类算法的总称，最早由沙皮尔(R.E. Schapire)提出。1995年弗洛德(Y. Freund)对Schapire的算法进行了改进，提高了算法的效率。但沙皮尔(R.E. Schapire)和弗洛德(Y. Freund)的算法在解决实际问题时有一个重大缺陷，即要求事先知道弱学习算法学习正确率的下界，这在实际问题中很难做到。1997年，沙皮尔(R.E. Schapire)和弗洛德(Y. Freund)提出了著名的Adaboost(Adaptive Boost)算法，该算法的效率与Freund算法很接近，却可以非常容易地应用到实际问题中，因此，该算法已成为目前最流行的Boosting算法。

集成学习

1996年，Breiman从文献可重复取样技术(Bootstrap Sampling)入手，提出了著名的Bagging方法。在该方法中，各学习器的训练集由从原始训练集中随机选取若干示例组成，训练集的规模与原始训练集相当，训练例允许重复选取。这样，原始训练集中某些示例可能在新的训练集中出现多次，而另外一些示例则可能一次也不出现。在预测新的示例时，所有学习器的结果通过投票的方式来决定新示例的最后预测结果。

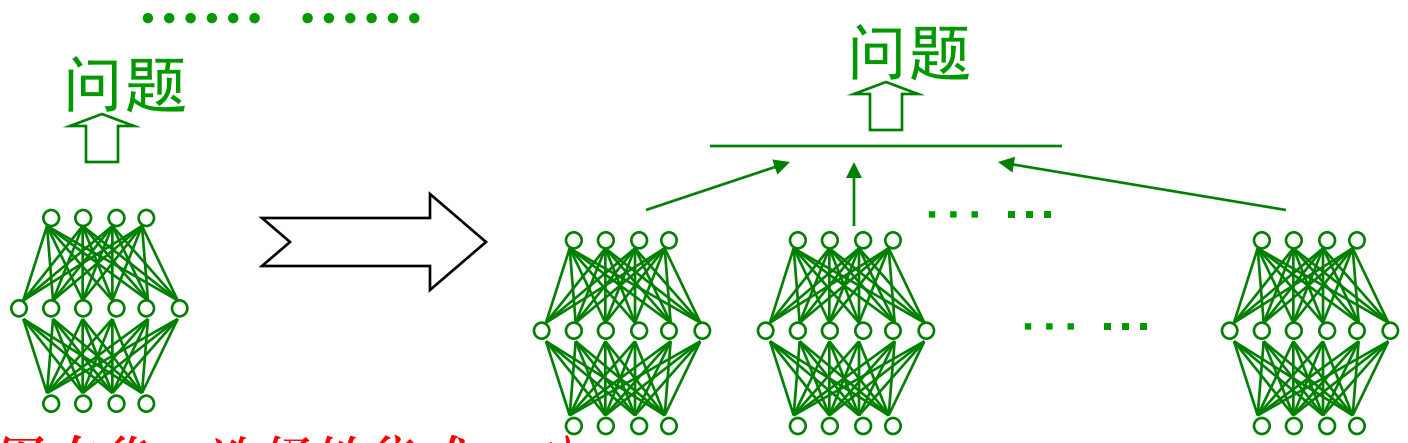
集成学习

集成学习（Ensemble Learning）是一种机器学习范式，它使用多个（通常是同质的）学习器来解决同一个问题

集成学习中使用的多个学习器称为个体学习器

当个体学习器均为决策树时，称为“决策树集成”

当个体学习器均为神经网络时，称为“神经网络集成”



(引自周志华：选择性集成ppt)

集成学习的重要性

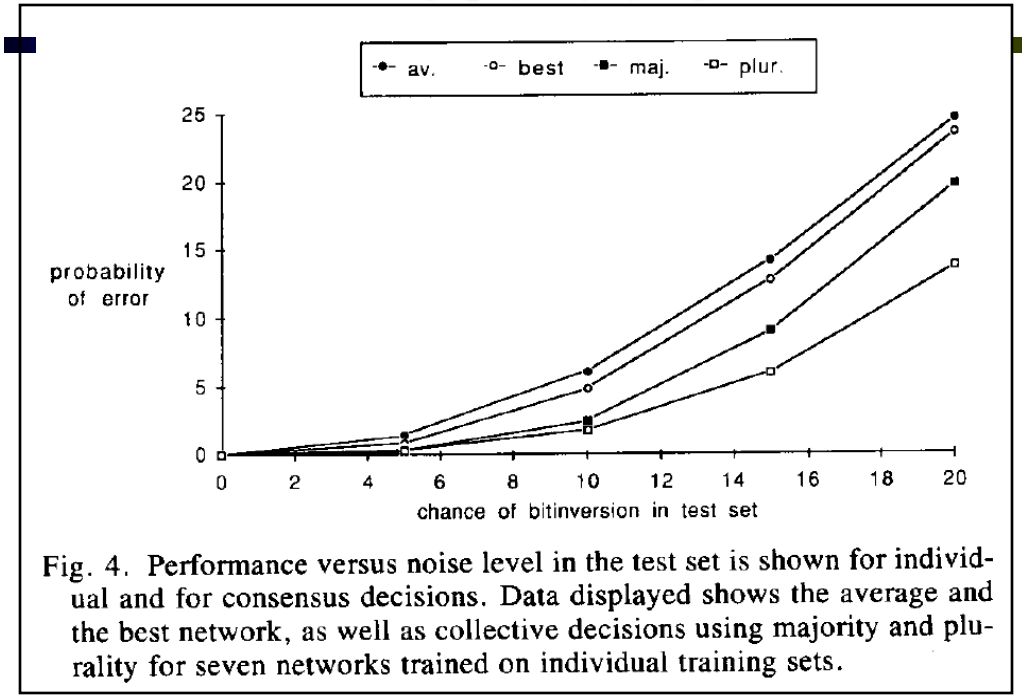


Fig. 4. Performance versus noise level in the test set is shown for individual and for consensus decisions. Data displayed shows the average and the best network, as well as collective decisions using majority and plurality for seven networks trained on individual training sets.

[L.K. Hansen & P. Salamon, TPAMI90]

问题：对20维超立方体空间中的区域分类

左图中纵轴为错误率

从上到下的四条线分别表示：

平均神经网络错误率

最好神经网络错误率

两种神经网络集成的错误率

由于集成学习技术可以有效地提高学习系统的泛化能力，因此它成为国际机器学习界的研究热点，并被国际著名学者 Dietterich 称为当前机器学习四大研究方向之首 [T.G. Dietterich, AIMag97]

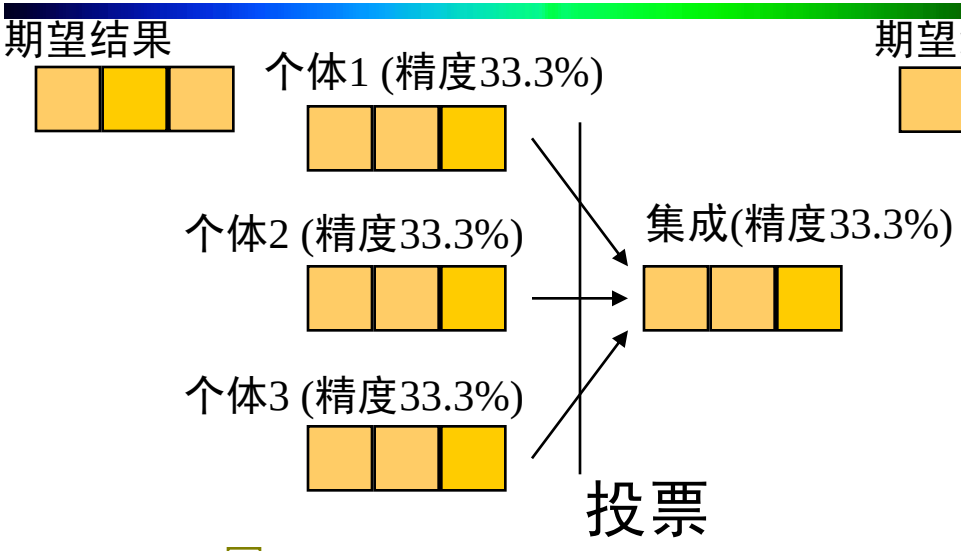
令人惊奇的是，集成的错误率比最好的个体还低

集成学习的用处

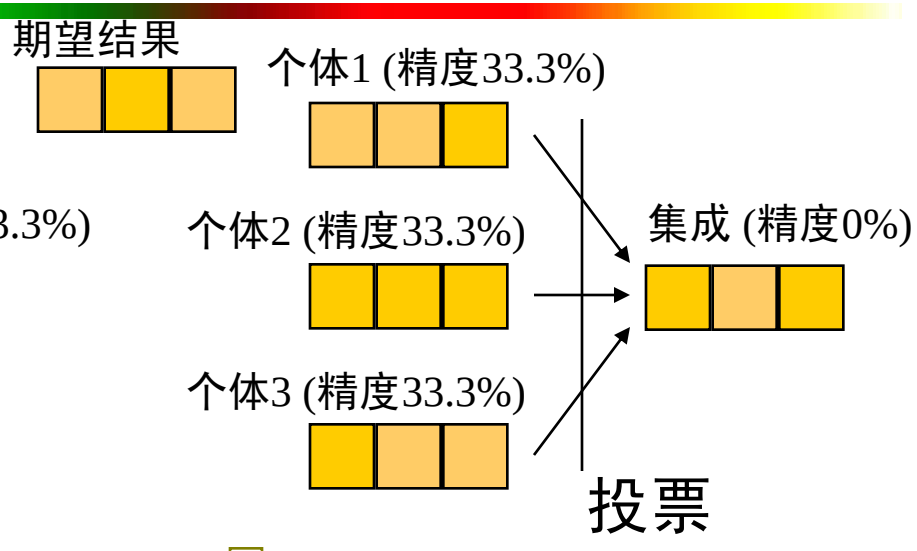
集成学习技术已经在行星探测、地震波分析、Web信息过滤、生物特征识别、计算机辅助医疗诊断等众多领域得到了广泛的应用

只要能用到机器学习的地方，就能用到集成学习

如何构建好的集成



个体必须有差异



个体精度不能太低

$E = \bar{E} - \bar{A}$ [A. Krogh & J. Vedelsby, NIPS94]

个体学习器越精确、差异越大，集成越好

(引自周志华：选择性集成ppt)

个体越多越好吗？

既然多个个体的集成比单个个体更好，那么是不是个体越多越好?

更多的个体意味着：

- 在预测时需要更大的计算开销，因为要计算更多的个体预测
- 更大的存储开销，因为有更多的个体需要保存

个体的增加将使得个体间的差异越来越难以获得

选择性集成

Many Could be Better Than All：在有一组个体学习器可用时，从中选择一部分进行集成，可能比用所有个体学习器进行集成更好 [Z.-H. Zhou *et al.*, AIJ02]

从一组个体学习器中排除出去的个体（k）应满足：

分类 $$(2N-1)\sum_{i=1}^{N}\sum_{j=1}^{N}C_{ij} \le 2N^2\sum_{\substack{i=1\\ i\neq k}}^{N}C_{ik} + N^2E_k$$

回归 $$\sum_{\substack{j=1\\ j\in\{j\,|\,|Sum_j|\le 1\}}}^{m} Sgn\left(\left(Sum_j + f_{kj}\right)d_j\right) \le 0$$

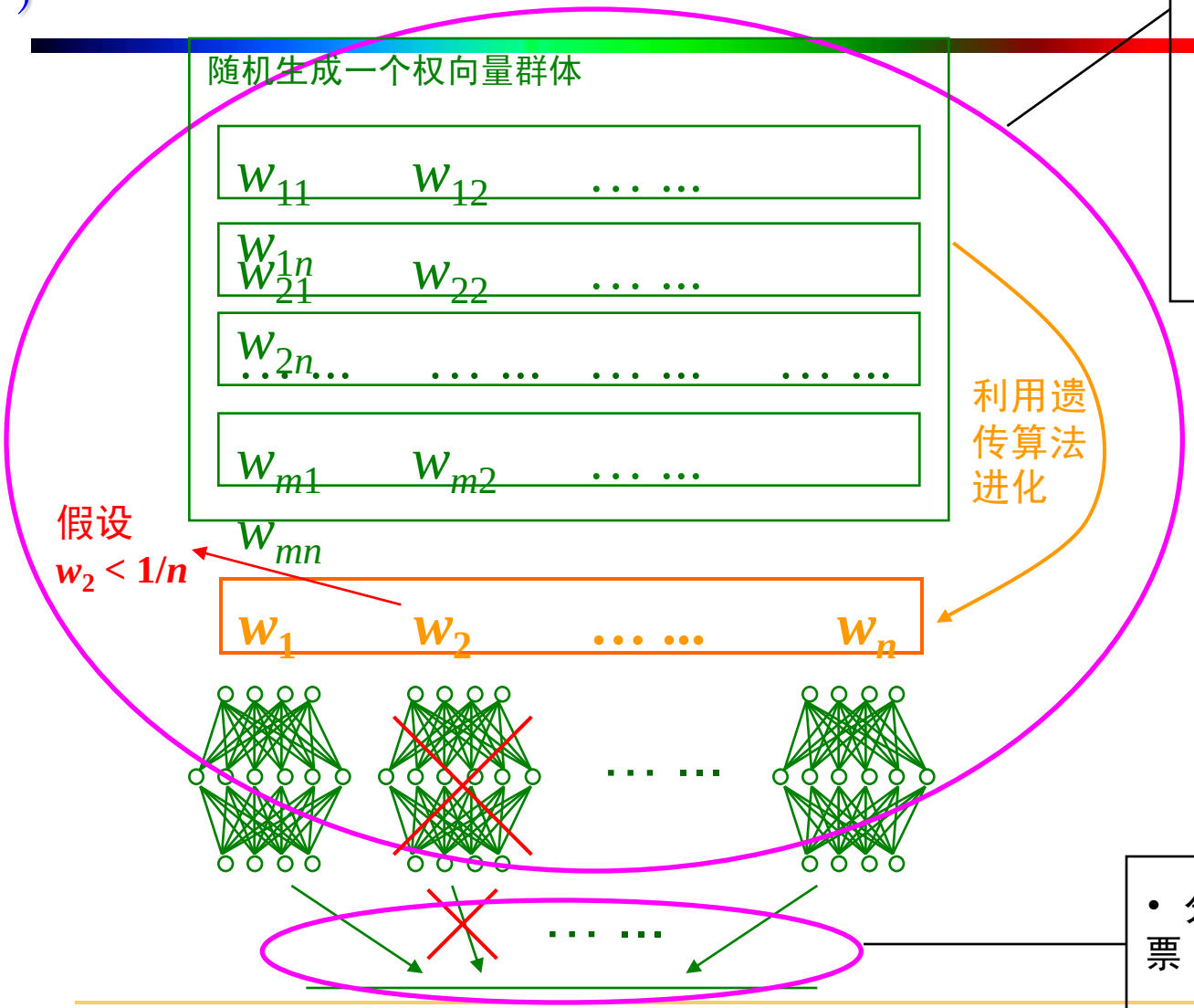
遗憾的是，上述公式在解决实际问题时难以直接使用

(引自周志华：选择性集成ppt)

GASEN（基于遗传算法的选择性集成学习算法）

遗传算法选择

随机生成若干权向量，权向量的每个分量对应了一个个体学习器，这些权向量被遗传算法进化，得到一个最优权向量，它表示了各个体学习器在构成集成时的“重要性”，据此进行个体的选择



为了证明选择性集成学习的可操作性，我们提出了GASEN算法

- 分类：有排除的投票
- 回归：有排除的平均

[Z.-H. Zhou *et al.*, AIJ02]

实验结果

与著名的集成学习算法Bagging和Boosting相比，GASEN 获得了更高的(或相当的) 精度，而且使用的个体学习器少得多 (回归：19% (3.71/20)；分类：36% (7.10/20.0))

[Z.-H. Zhou *et al.*, AIJ02]

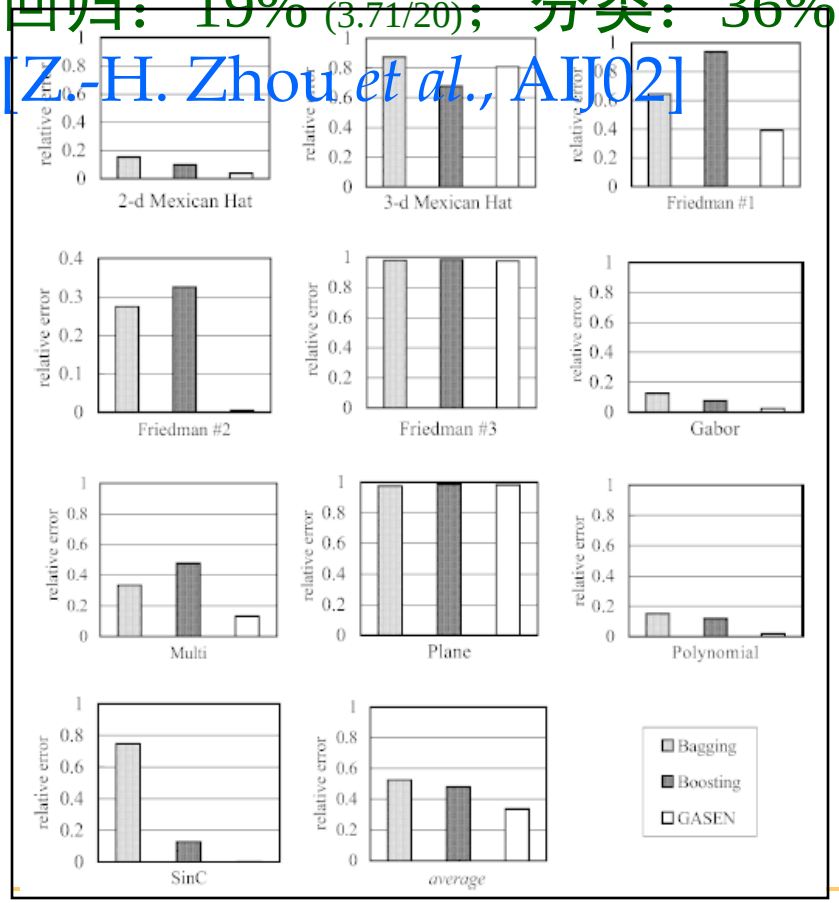


Fig. 2. Comparison of the relative error of Bagging, Boosting, and GASEN on regression tasks.

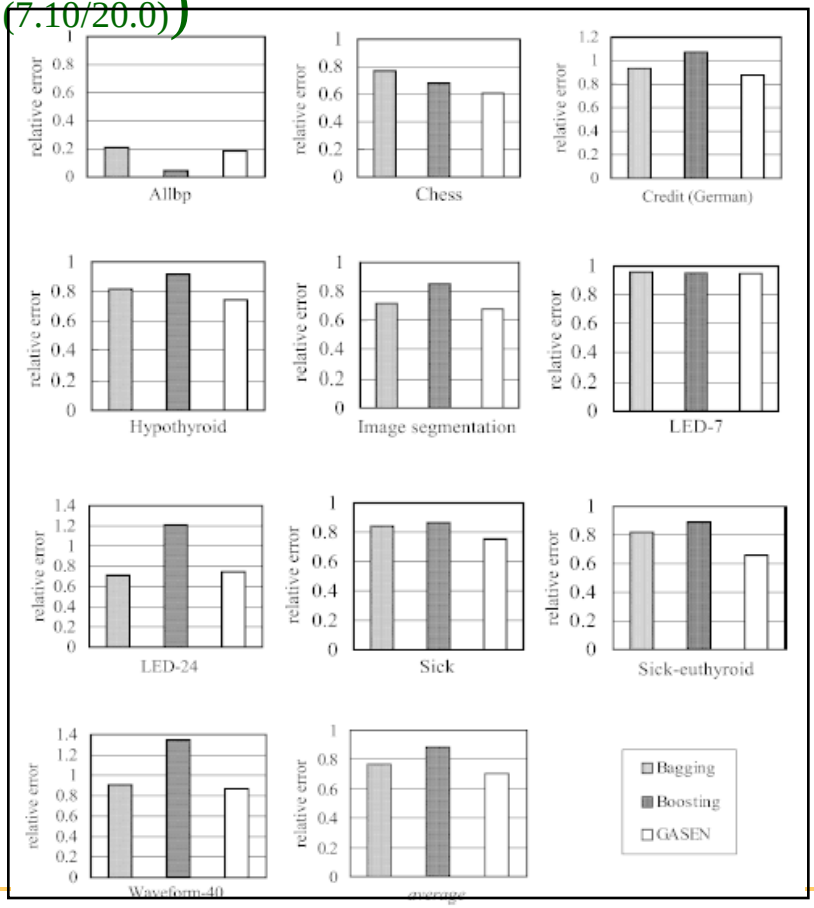


Fig. 3. Comparison of the relative error of Bagging, Boosting, and GASEN on classification tasks.

Bias-Variance分解

给定学习目标和训练集规模，

bias 度量了学习算法的平均估计结果与目标的接近程度

variance 度量了在同样规模的不同训练集上，学习算法的估计结果的扰动程度

我们采用的分解机制为 [R. Kohavi & W.H. Wolpert, ICML96]

$$\text{bias}_x^2 = \frac{1}{2}\sum_{y\in Y}\left[P\left(Y_F = y|x\right) - P\left(Y_H = y|x\right)\right]^2$$

$$\text{variance}_x = \frac{1}{2}\left(1 - \sum_{y\in Y} P\left(Y_H = y|x\right)^2\right)$$

以往研究表明，Bagging主要减小variance，而Boosting主要减小bias

[E. Bauer & R. Kohavi, MLJ99; L. Breiman, TechRep96]

分解结果

回归　　分类

bias

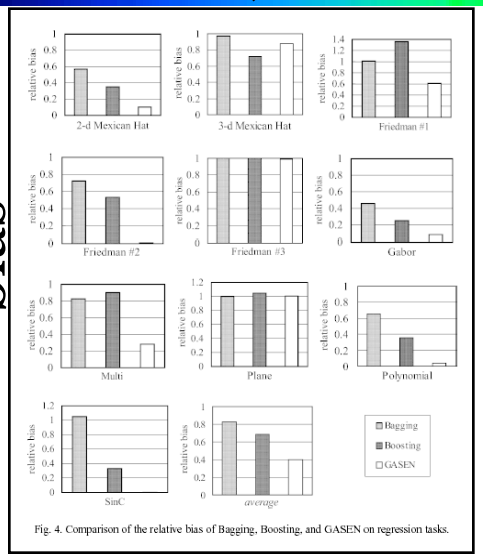


Fig. 4. Comparison of the relative bias of Bagging, Boosting, and GASEN on regression tasks.

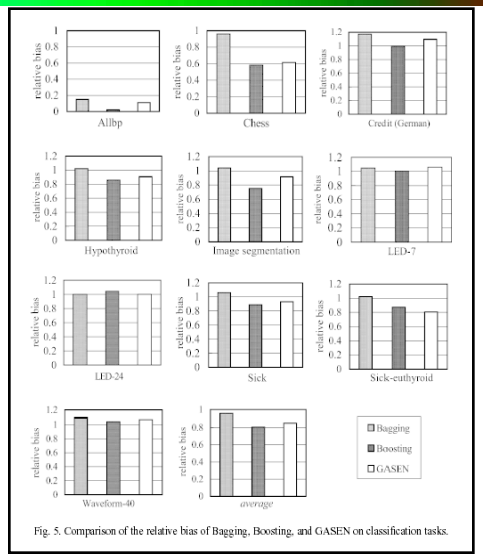


Fig. 5. Comparison of the relative bias of Bagging, Boosting, and GASEN on classification tasks.

variance

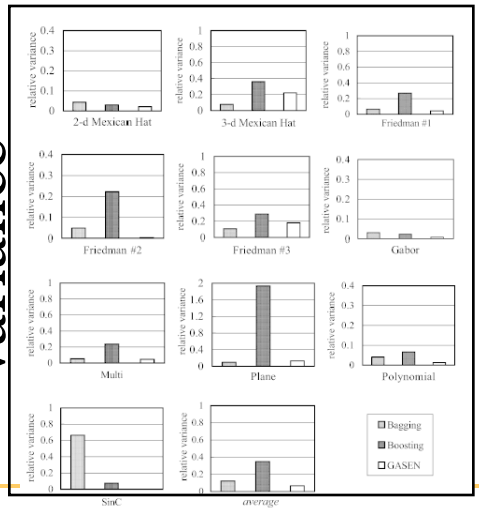


Fig. 6. Comparison of the relative variance of Bagging, Boosting, and GASEN on regression tasks.

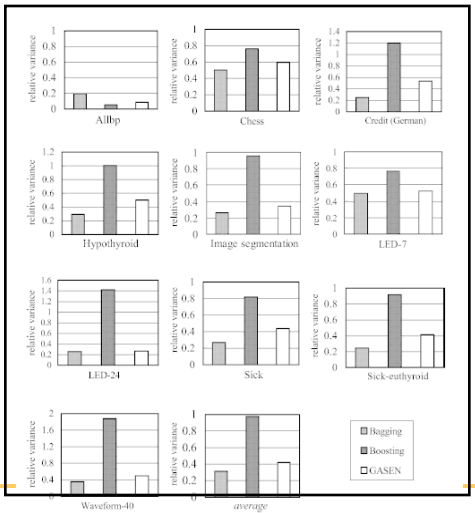


Fig. 7. Comparison of the relative variance of Bagging, Boosting, and GASEN on classification tasks.

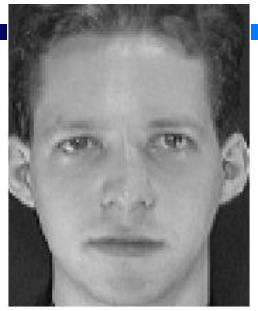
在回归任务中，GASEN在减小bias和variance方面都优于Bagging和Boosting

在分类任务中，GASEN在减小bias方面优于Bagging，在减小variance方面优于Boosting

…… GASEN的成功在于其既可以有效地减小bias，又可以有效地减小variance

[Z.-H. Zhou *et al.*, AIJ02]

生物特征识别：选择性集成的一个应用

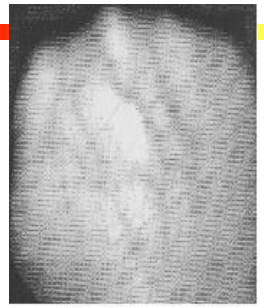


人脸

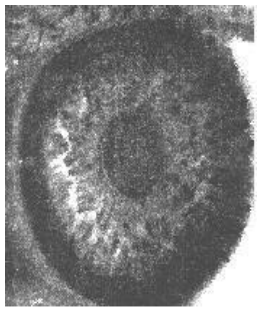
脸部热量图

指纹

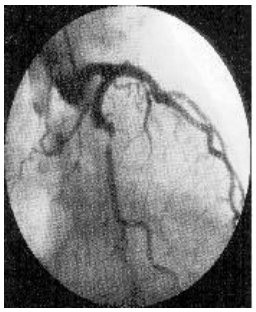
手形



手部血管分布

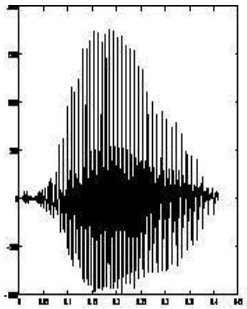


虹膜



视网膜

签名



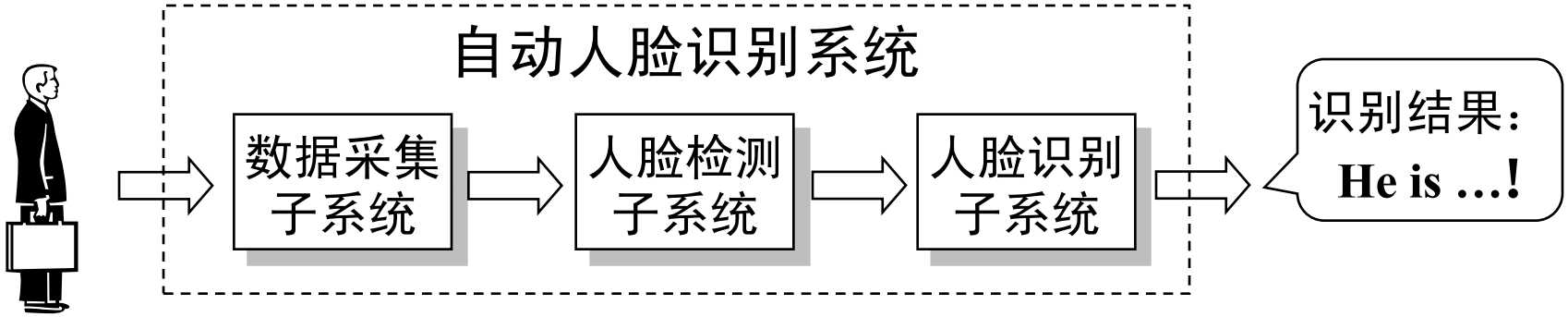
语音

Bill Gates: 以人类生物特征进行身份验证的生物识别技术，在今后数年内将成为IT产业最为重要的技术革命

人脸识别

人脸识别因识别方式友好、可隐蔽而备受学术界和工业界关注（但人脸识别不是万能的）

自动人脸识别系统



- 所谓自动人脸识别系统，是指不需要人为干预，能够自动获取人脸图像并且辨别出其身份的系统
- 一个自动人脸识别系统至少要包含三个部分，即数据采集子系统、人脸检测子系统和人脸识别子系统

“人脸识别”有时是指整个自动人脸识别系统所做的工作，有时是指人脸识别子系统所做的工作

(引自周志华：选择性集成ppt)

本征脸（eigenface）方法

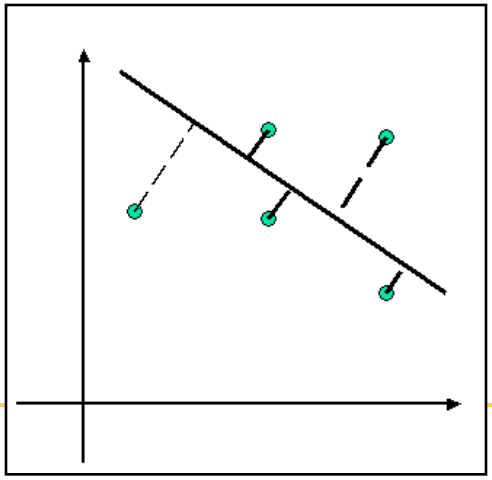
是人脸识别的基准技术，并已成为事实上的工业标准

该方法基于主成分分析（PCA）

PCA是将分散在一组变量上的信息集中到某几个综合指标（主成分）上的数学方法，实际上起着数据降维的作用，并保证降维过程最大化保留原数据的差异

这对最大化类间差异（即不同人之间的差异）并最小化类内差异（即同一人的不同图像间的差异）很有效

用PCA将2维数据降到1维的例子，绿色点表示二维数据，PCA的目标就是找到这样一条直线，使得所有点在这条直线上的投影点之间的平均距离最大。也就是最大化地保留了原数据的差异性



本征脸方法

- 高N_1，宽N_2的图像P 可以转化为$N_1 \times N_2$维的向量x
- 线性变换: $y = \mathbf{W}^{\mathrm{T}}(x - \mu)$，其中$y$的维数$m$ 远远小于x的维数n
- 寻找$\mathbf{W}$ 使得y 最大程度地保持x原有的差异（variance）
- $\mathbf{W}$ 的求法：

  1) 样本集的总体散布矩阵：$C = E\{(x-\mu)(x-\mu)^{\mathrm{T}}\}$

  2) 求出C 的本征向量和对应的本征值;

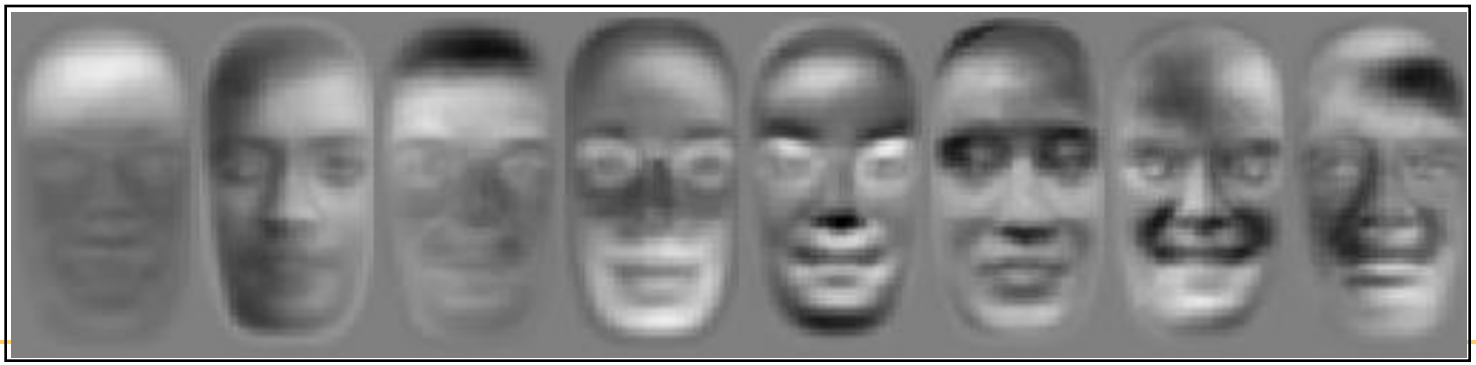
  3) 将本征值排序为 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$，它们对应的本征向量分别为 $w_1, w_2, \ldots, w_n$

  4) 取最前面的m个本征向量 $w_1, w_2, \ldots, w_m$ 组成变换矩阵$\mathbf{W}$

本征脸方法

- 直接计算C的本征值和本征向量是困难的，可以通过对矩阵 $\mathbf{X}=[(x_1-\mu),(x_2-\mu),\ldots,(x_D-\mu)]$ 做奇异值分解间接求出
- m值的选择：$\dfrac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{n}\lambda_i}\geq T$

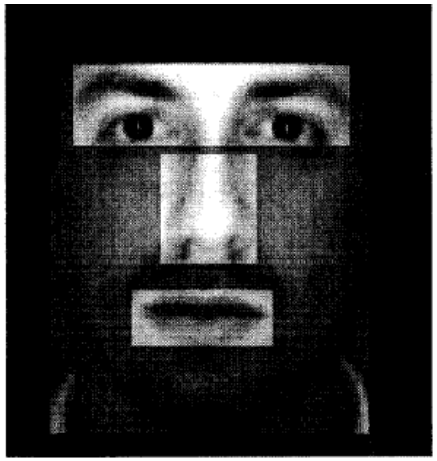
如果将本征向量恢复成图像，这些图像很像人脸，因此称为“本征脸”



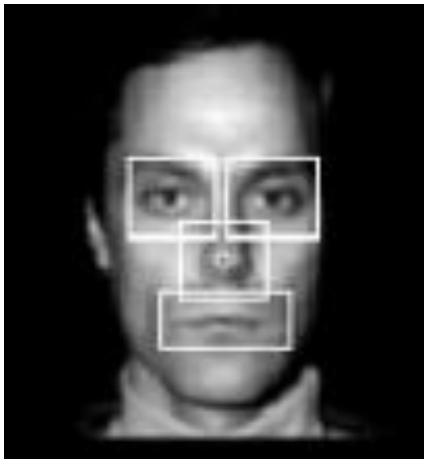
[M. Turk & A. Pentland, JCN91]

本征特征（eigenfeature）方法

利用PCA分析眼、鼻、嘴等局部特征，即本征特征方法



[R. Brunelli & T. Poggio, TPAMI93]



[A. Pentland *et al.*, CVPR94]

这实际上相当于：为若干重要的特征建立本征空间，然后将多个本征空间集成起来

本征脸 vs. 本征特征

本征脸利用全局特征，本征特征利用局部特征，二者各有优势

待识别图像

本征脸识别结果

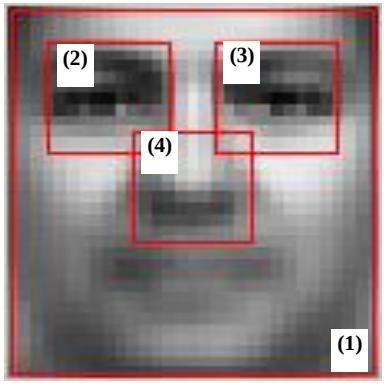
本征特征识别结果

[A. Pentland *et al.*, CVPR94]

本征脸 vs. 本征特征

将二者结合，可以得到更好的识别效果

同样，这实际上相当于：为若干重要的特征建立本征空间，然后将多个本征空间集成起来



由于嘴部受表情影响很严重，因此未考虑嘴部特征

难题——能否自动确定：

该用哪些特征？（眼睛？鼻子？嘴？……）

特征的确切位置在哪儿？（从哪儿到哪儿算眼睛？……）

SEME（选择性多本征空间集成算法）

将人脸图像中所有的矩形区域都看做一个可能的特征，这样，在每一个矩形区域都建立一个本征空间，最后将重要的本征空间集成起来

图像中包含的矩形区域的数量是非常巨大的（例如一幅34×31的图像包含的矩形区域就多达295,120个），不可能使用所有的本征空间组成集成，但可以运用选择性集成思想，从中选择出部分本征空间组成集成

考虑 $E = \bar{E} - \bar{A}$，选择的原则应该是：

- 本征空间本身的误差较小
- 本征空间之间的差异较大（即互补性较大）

(引自周志华：选择性集成ppt)

SEME（选择性多本征空间集成算法）

- 给定k个人脸图像样本（每人两张图像，一张为gallery图像，另一张为probe图像），算法将从所有N个矩形（R_1，R_2，…，R_N.）中选择出m个
- For $i = 1, 2, \ldots, N$:
  - 1 以gallery图像为训练集，在R_i上训练出一个本征空间
  - 2 利用该本征空间识别所有的probe图像，记下识别率r_i
- 将R_i按照相应的r_i从大到小排序
- S = {R_1}，A = {R_2, …, R_n}，这里 $m << n << N$
- For $t = 1, 2, \ldots, m$-1:
  - 1 对A中的每一个矩形R_i，计算R_i 能够识别正确，而S中至少有一个矩形识别错误的probe图像的数目，记为c_i
  - 2 找到具有最大纠正误识数目c_t的矩形R_t
  - 3 将R_t从A中删除并添加到S中
- 集成与S中m个矩形相对应的本征空间用于人脸识别

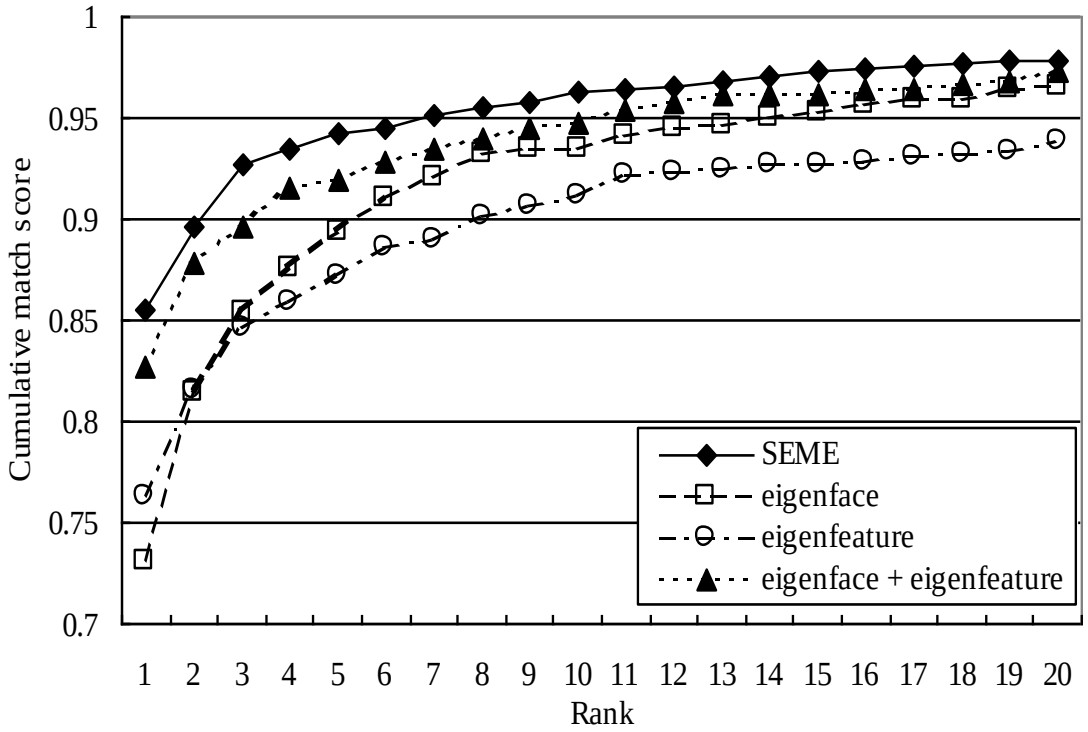
选择误差小的矩形特征

选择差异大的矩形特征

[X. Geng & Z.-H. Zhou, unpub04]

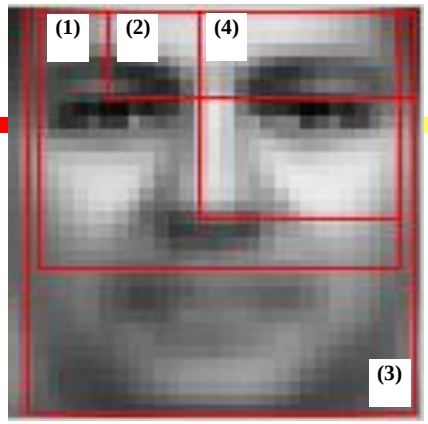
实验结果

FERET人脸数据库上的结果

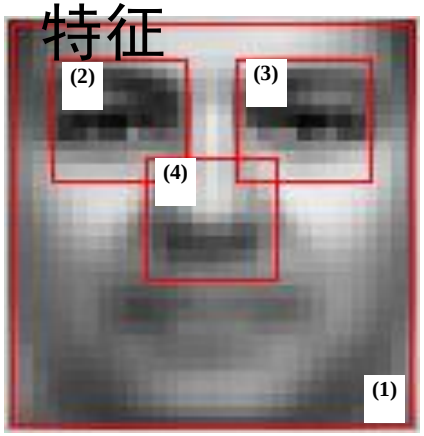


待识别图像出现在算法返回的前Rank个图像中

[X. Geng & Z.-H. Zhou, unpub04]



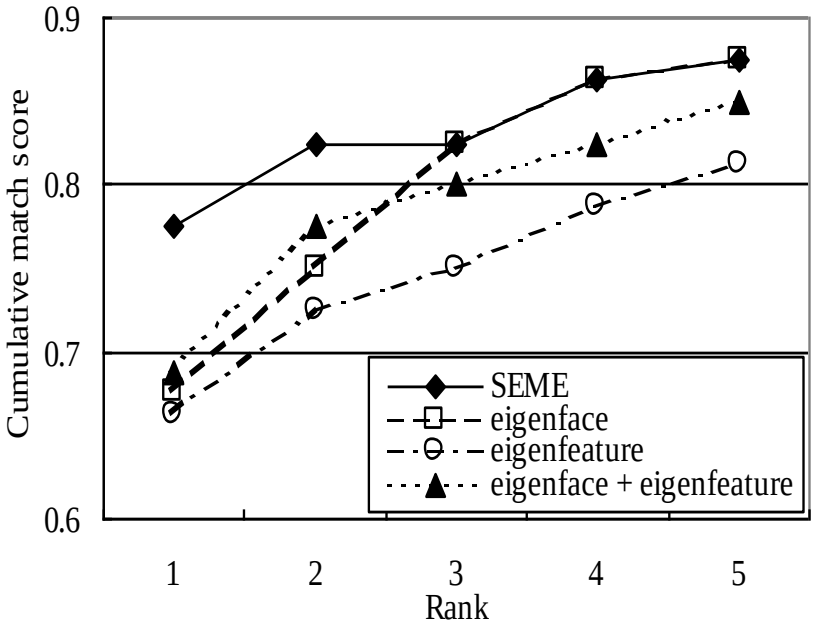
SEME选择的特征

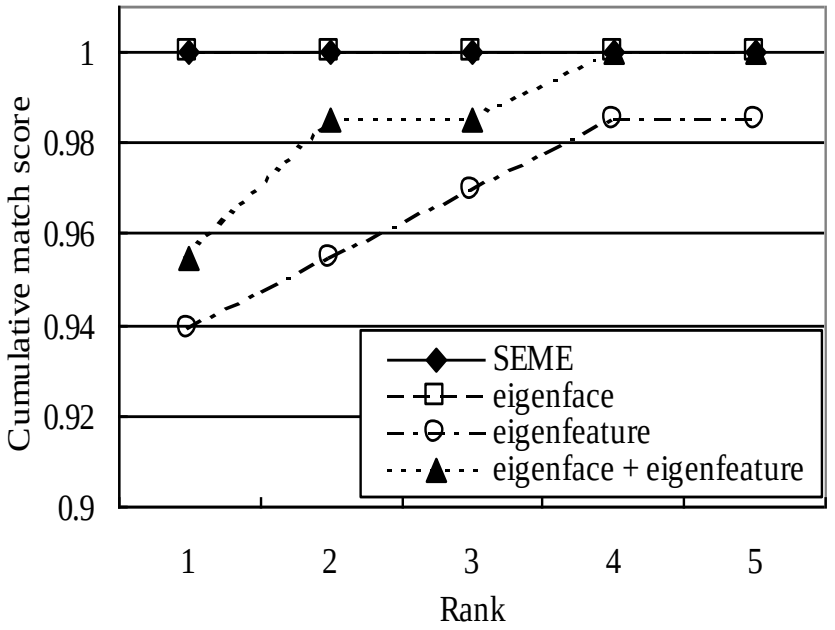


本征脸+本征特征所用的特征

SEME的可扩展性

SEME的训练（计算）开销很大，但只需训练一次



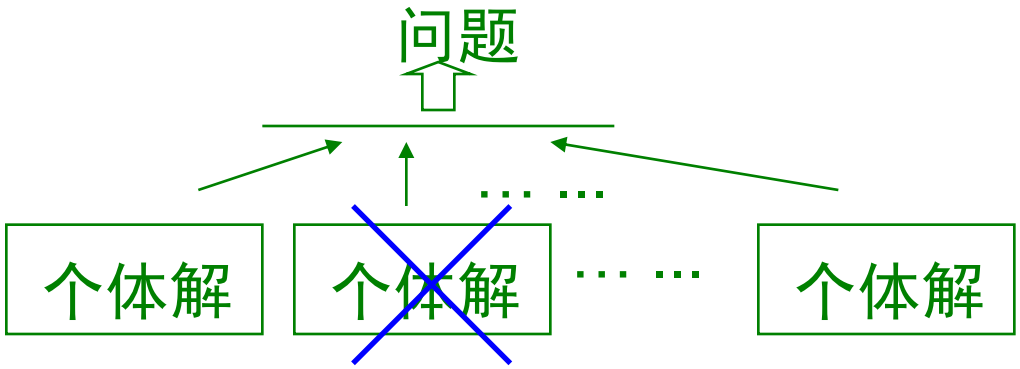


将FERET人脸数据库上选择出的本征空间集成直接用于ORL（左）和BioID（右）这两个人脸数据库的结果

[X. Geng & Z.-H. Zhou, unpub04]

选择性集成

选择性集成的思想：利用多个个体，并通过对个体进行选择，可以获得更好的结果



选择性集成的思想可以用到更多的领域中去

选择的基本原则：个体的效用高、差异大

(引自周志华：选择性集成ppt)

基于Bagging的聚类

聚类分析时若需要处理的对象数非常庞大，对所有对象进行聚类运行开销很大。可以考虑对整个空间中分布在某个局部的对象进行聚类从而得到结果。但是，由于聚类对象的空间分布可能很不均匀，所得到的聚类结果可能只代表某个局部信息而不是全局的聚类结果。为了解决这个问题，可以采用类似Bagging算法中产生个体训练集的方式产生用于聚类的训练集，即通过可重复取样技术从原向量集χ中产生若干训练集$\{S_i\}$, $i=1$，…T, 对每个训练集S_i用k均值聚类器进行聚类。

神经网络集成系统的规则获取

利用神经网络集成系统可以获取规则。以从神经网络集成系统获取规则的算法REFNE(Rule Extraction From Neural network Ensemble)为例，说明学习的过程和算法。算法中使用if-then形式的规则来解释集成学习到的知识，采用神经网络作为集成的基学习器。

神经专家系统

神经专家系统的知识表示方式与传统人工智能完全不同。传统的知识表示，不管是产生式系统，还是语义网络，都可以看作是知识的一种显式表示，而神经专家系统中的知识表示可看作是一种隐式表示。在这里知识并不像在产生式系统中那样独立表示每一规则，而是将某一问题的若干知识在同一网络中表示。

神经专家系统

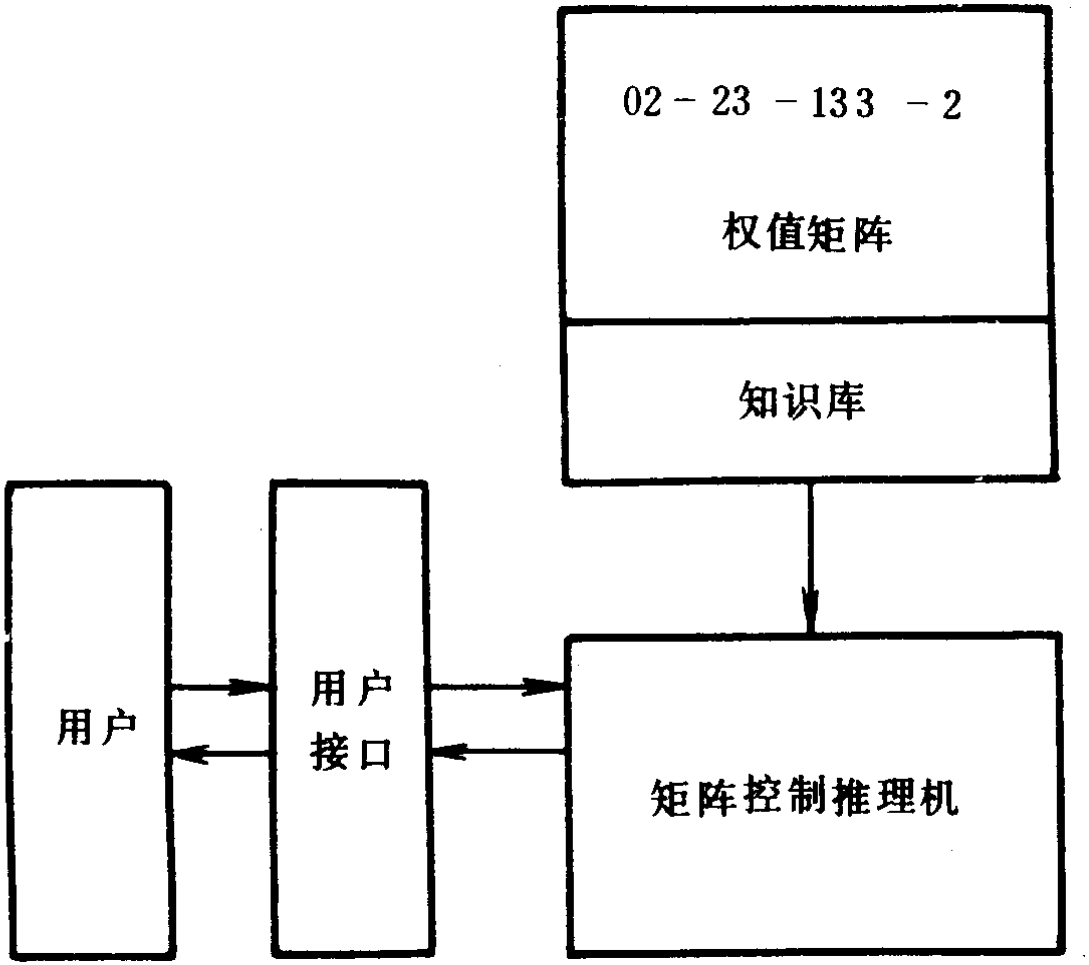
神经专家系统的推理机制与现有的专家系统所用的基于逻辑的演绎方法不同，它的推理机制基本上是数值计算过程，主要由以下三部分组成：

(1) 输入逻辑概念到输入模式的变换，并根据论域的特点，确定变换规则，再根据相
应规则，将目前的状态变换成神经网络的输入模式；

(2) 网络内的前向计算：根据神经元特征，其输入为x_i = ，w_{ij}为连接权值系数，y_j为前层神经元的输出。本层神经元的输出$y_i = f_i(x_i + \theta_i)$。其中的$\theta_i$为神经元的阈值，$f_i$为单调递增非线性函数。通过上述计算即可产生神经网络的输出模式；

(3) 输出模式解释：随着论域的不同，输出模式的解释规则亦各异。解释的主要目的是将输出数值向量转换成高层逻辑概念。

神经专家系统的结构



Thank You

Intelligence Science

http://www.intsci.ac.cn/