Artificial Neural Networks for Biometric Applications

PRESENTED BY:

LI GUOJIE

SEPT. 2003

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

ANN for Biometric Applications

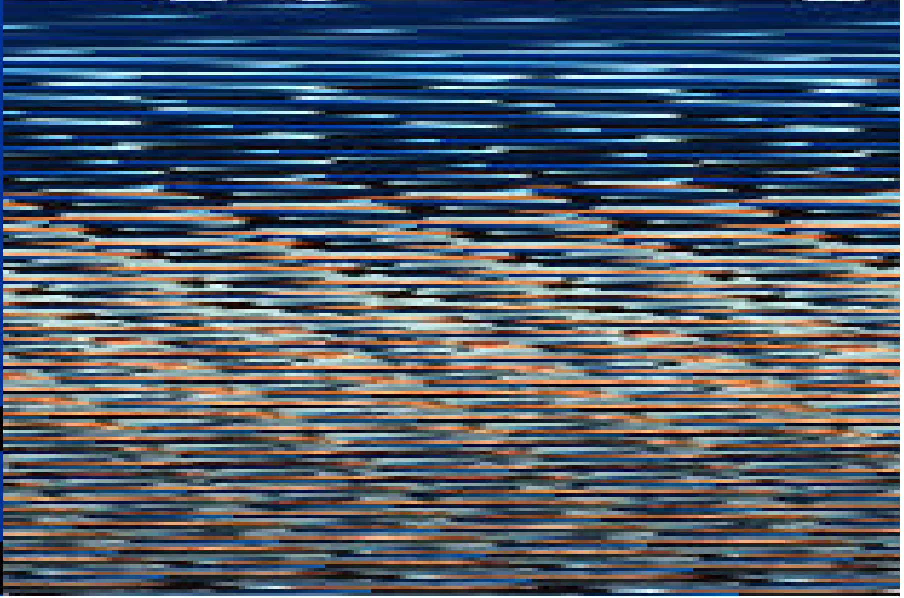
- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

How to Tell the Real Saddam from His Doubles?

1. Saddam has at least three doubles – Reuters, Sept. 27, 2002.
2. Hard to know if Saddam tape is genuine – CNN, July 30, 2003.

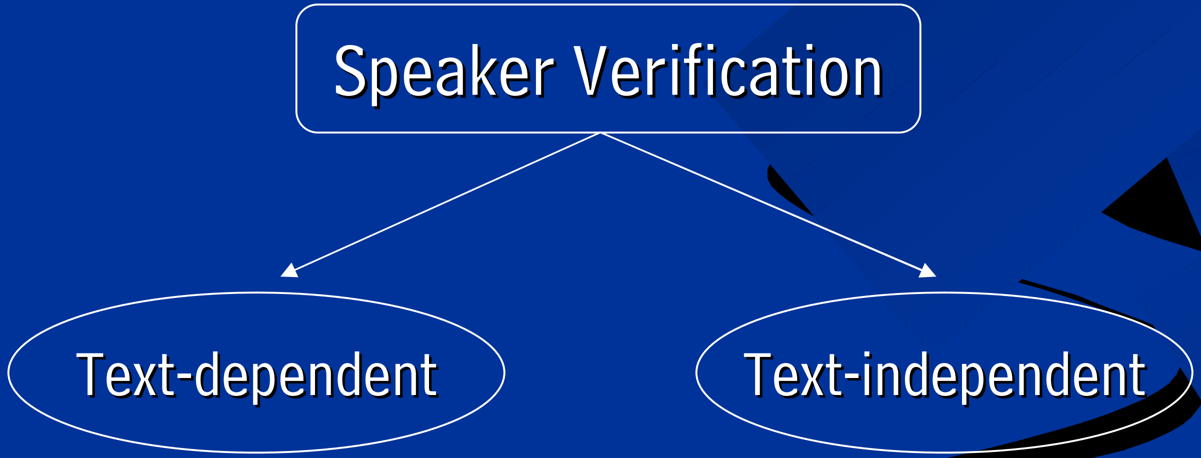
- How to tell whether a 'Saddam' is the real one or not?
- Possible approaches

  - DNA Test
  - Face Recognition
  - Speaker Verification



Introduction to Speaker Verification

- Speaker Verification is aimed to decide, given a sample of speech, whether a specified candidate speaker said it.
- Speaker Verification can be categorized into:



Speaker Verification

Text-dependent

Text-independent

The speaker is supposed to say predefined words and he must cooperate

No restriction. The speaker can say anything.

Applications of Speaker Verification

- Access control, eg. computer access control
- Telephone voice authentication
- Automatic speaker labeling of recorded meetings for audio indexing
- Many other applications …

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

Review of Speaker Verification

- Statistical method
  - Using long-term statistics calculated on a very long acoustic sequence
  - Problem: Sensitive to the variability of the transfer function of the channel
  - C. Montacie improved it using a multi-dim autoregressive model in 1992
- Vector quantization
  - Proposed by F. K. Soong
  - Vector quantization of spectral or cepstral vectors
- Hidden Markov Models (HMM)
  - 'Linear predictive HMM and the speech signal' – A. B. Poritz, 1982.
- Artificial neural networks
  - Multilayer perceptrons
  - RBF and EBF - M. Man-Wai & K. Sun Yuan
  - EBF performs better than vector quantization method and large RBF networks

Review of Speaker Verification

- Can other kinds of RBF networks outperform EBF networks?
- MRAN, a sequential learning RBF, performs better than conventional RBF networks and some of EBF networks

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function (RBF) Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

RBF Networks: Fundamentals

- Exact interpolation problem:

$$\{X_i\} \xrightarrow[y_i = f(X_i)]{f(.)} y_i$$

- The RBF approach to exact interpolation problem:
  - Form a set of basis functions, one for each data set.
  - RBF function is of the form: $\mathbf{F}\,(\,\|\,\mathbf{x} - \boldsymbol{\mu}\,\|\,)$
  - f(x) can be taken to be a linear combination of the outputs of all RBFs:

$$f(x) = \sum_k \underset{\text{Weight factor}}{\underset{\uparrow}{w_k}} \Phi\left(\left\| x - \underset{\text{Center of the function}}{\underset{\uparrow}{m_k}} \right\|\right)$$

RBF Networks: Fundamentals

- RBF Function:
  - Non-linear
  - Choices: thin plate spline, multi-quadratic, Gaussian kernel
  - Gaussian function is the most common one:

$$\Phi(x) = \exp\left(-\frac{x^2}{2s^2}\right)$$

Factor controlling radius of influence of the function

- Generalization to multi-variable output can be done by mapping input vectors X_i onto corresponding output vectors Y_i

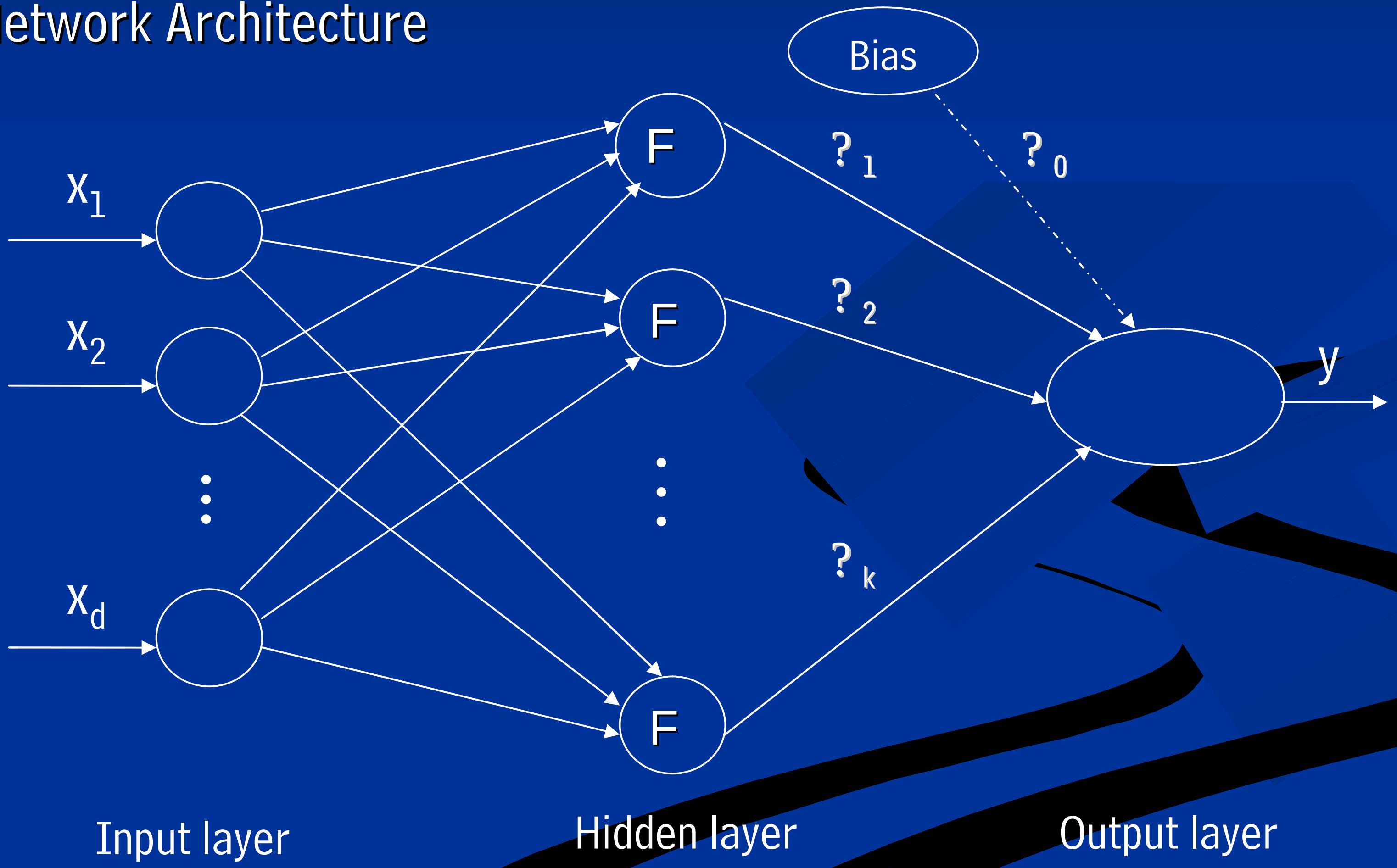
Radial Basis Function Networks

- RBF Network model is built on the base of the exact interpolation
- Improvements:
  - Restriction for strict function interpolation removed – Broomhead & Lowe
  - Centers of basis functions are not constrained to the values of input data set
  - Bias parameters can be included
- RBF Network model after improvements made:

$$\underbrace{y}_{\text{Output}} = \sum_{k}^{K} \underbrace{w_k}_{\text{Weight factor}} \underbrace{\Phi_k}_{\text{Radial basis function}}(x) + \underbrace{w_0}_{\text{Bias}}$$

Architecture of RBF Networks

- Network Architecture



Radial Basis Functions

- RBF Network model

$$y = \sum_{k}^{K} w_k \Phi_k(x) + w_0$$

Number of hidden neurons

Bias

Connection weight for each hidden unit

- Gaussian function as the radial basis function

$$\Phi(x) = \exp\left(- \frac{\left\| x - m_k \right\|^2}{2s^2} \right)$$

Center position of the hidden neuron

Width value for the Gaussian function

Radial Basis Function Networks

- Problem: Difficult to choose the right number of hidden neurons
- S. Lee & R. Kil developed a Hierarchically Self-Organizing Learning algorithm to determine the number of hidden units

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function (RBF) Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

Resource Allocation Networks

- Batch learning vs. sequential/incremental learning
- Platt proposed a sequential learning method for Resource Allocation Networks (RAN)
  - Hidden neurons are added based on the novelty of the training pattern inputs. If not novel, LMS is used to update network params.
- Kadirkamanathan & Niranjan improved RAN by replacing the LMS with the Extended Kalman Filter (EKF) – RANEKF
- Drawback of RAN – noise data points may trap the network to create redundant hidden neurons

Minimal Resource Allocation Networks

- L. Yingwei, N. Sundararajan and P. Saratchandran improved RAN by adding a pruning strategy. The resulting network is known as Minimal Resource Allocation Networks (MRAN).
- Growth strategy of MRAN: same as that of RAN, comprised of allocation of new hidden units and adjustment of network params.
- Criteria to add a new hidden unit:

$$\|x_n - u_{nr}\| > e_n \qquad\qquad e_n = |y_n - f(x_n)| > e_{\min}$$

$$e_{rmsn} = \sqrt{\frac{\sum_{i=n-(M-1)}^{n} e_i^2}{M}} > e_{\min 1}$$

Minimal Resource Allocation Networks

- If one or more criteria can not be satisfied, no hidden neurons will be added and network parameters will be updated using an Extended Kalman Filter.
- Pruning strategy of MRAN: a hidden unit should be pruned if its output contribution is less than a threshold over a certain number of consecutive iterations.

$$y_n = \sum_{k=1}^{K} w_k \Phi_k(x_n) + w_0$$

Total network output

$$o_k = w_k \Phi_k$$

A hidden unit’s contribution to the total output

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function (RBF) Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Works

Application of MRAN to Speaker Verification – The Speech Database

- TIMIT – a phonetically balanced continuous speech corpus, containing speech from 630 speakers representing 8 major dialect divisions of American English.
- 258 speakers from the first four regions of the corpus are involved in the experiments:

SET	REGION	MALE	FEMALE	TOTAL
Speaker set	2	46	30	76
Anti-speaker set	1	20	18	38
Pseudo-imposter set	4	38	30	68
Imposter set	3	53	23	76

Application of MRAN to Speaker Verification – The Speech Database

- Speaker set: contains all the target speakers;
- Data of anti-speakers in the anti-speaker set is used to train the neural network for each target speakers during speaker enrollment;
- Data of pseudo-imposters in pseudo-imposter set is used to determine the decision threshold for each network;
- Data of imposters in imposter set is used in verification phase to calculate the false acceptance rate.

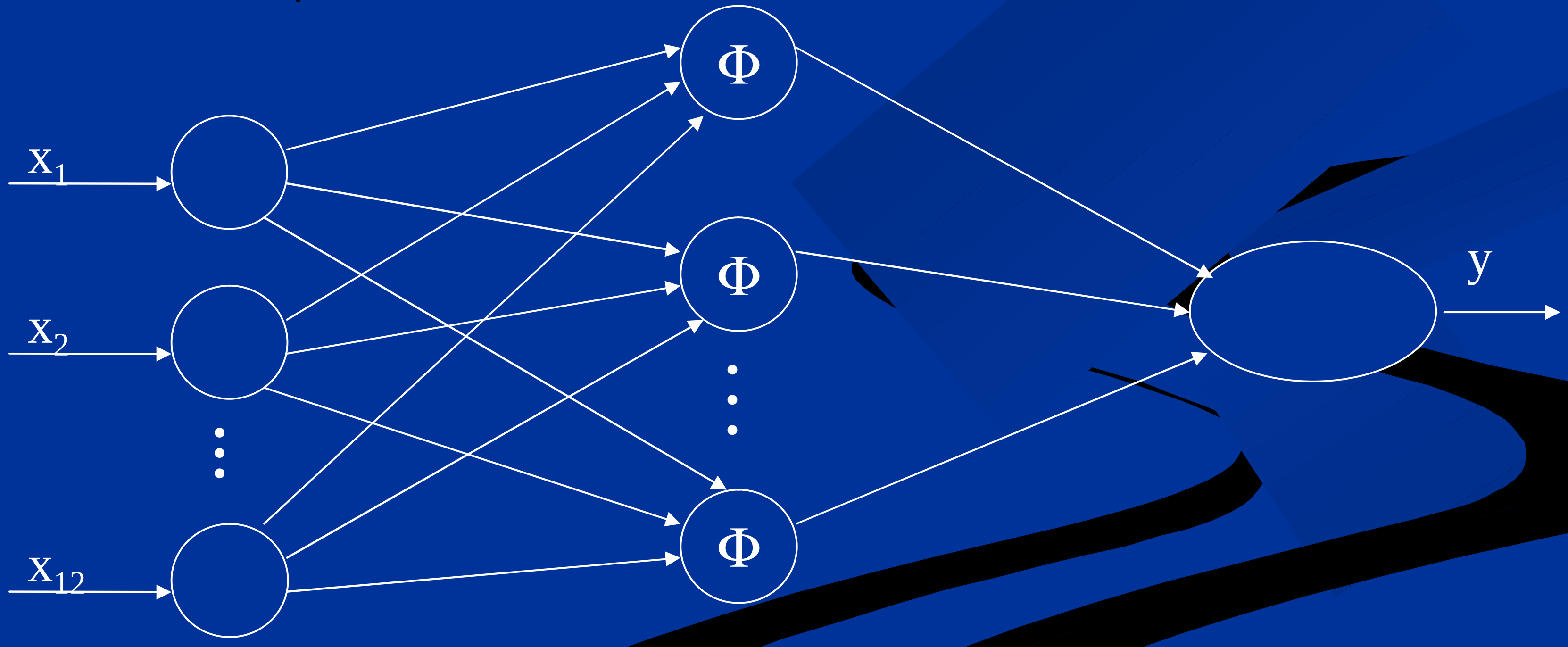
- Each speaker reads three types of sentences: 2 Dialect (SA), 5 Compact (SX) and 3 Diverse (SI) sentences.

Application of MRAN to Speaker Verification – Feature Extraction

- Cepstral coefficients are the best features for speaker verification
- 12th order LP-derived cepstral coefficients are used as feature vectors in the experiments
- Speech signals are pre-emphasized by a filter with transfer function $H(z) = 1 - 0.95z^{-1}$
- Silent regions are removed

Application of MRAN to Speaker Verification – Enrollment

- Each speaker in the speaker set is assigned a radial basis network which models the characteristics of his or her voice
- Network setup:



Input layer Hidden layer Output layer

Application of MRAN to Speaker Verification – Steps of Enrollment

1) Using the target speaker’s data and all anti-speakers’ data to train a MRAN network

   Ratio of training vectors between the target speaker class and anti-speaker set class is around 1:38 ---- Highly unbalanced!

Anti-speaker 1’s data	The speaker’s data	Anti-speaker 2’s data	The speaker’s data	…	Anti-speaker 38’s data	The speaker’s data

   Supervised training: target speaker’s output is 1; -1 otherwise.

2) Determining the decision (acceptance or rejection) threshold using pseudo-imposter set

Application of MRAN to Speaker Verification – Verification Phase

- Each network trained will be verified against the target speaker and a set of imposters
- False Rejection Rate (FRR) – probability of not detecting the target speaker when present

$$E_{FRR} = N_{MISS} / N_{TARGET}$$

- False Acceptance Rate (FAR) – probability of falsely detecting the target speaker when not present

$$E_{FAR} = N_{FA} / N_{IMPOSTER}$$

Application of MRAN to Speaker Verification – Verification Phase

- Equal Error Rate (ERR) – when the decision threshold is properly adjusted, FRR and FAR can be equal and the value of ERR equals to that of FAR or FRR
- Verification Process: an input feature vector sequence $T=[X_1, X_2, \cdots, X_t]$ extracted from an utterance spoken by an unknown speaker is fed into the network. Average output:

$$\bar{y} = \frac{1}{t}\sum_{i=1}^{t} f\left(x_i\right)$$

- Verification decisions:

Decision threshold

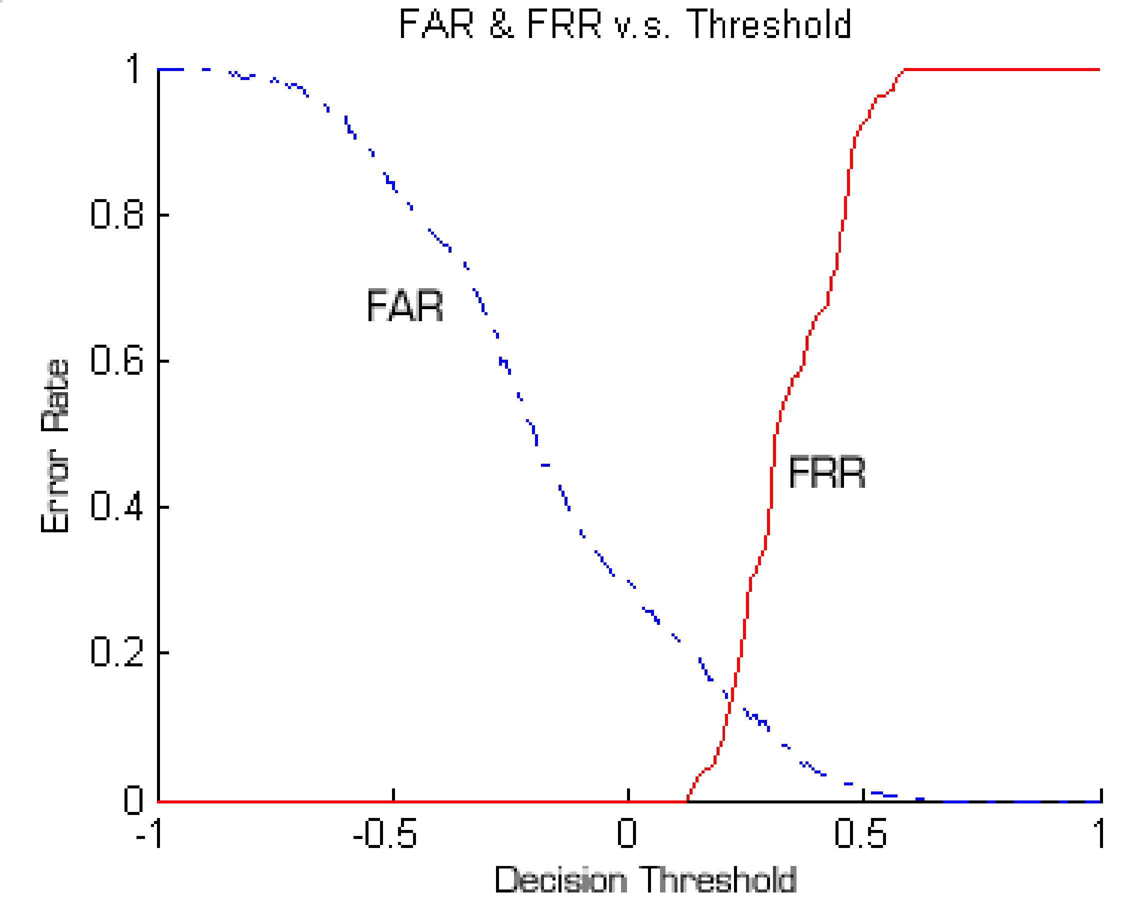
$$\begin{cases} \bar{y} > x \Rightarrow accept \\ \bar{y} \leq x \Rightarrow reject \end{cases}$$

Application of MRAN to Speaker Verification – Decision Threshold

- Decision threshold is determined in the enrollment phase
- Decision threshold is determined using feature vectors derived from utterance of pseudo-imposters
- Decision threshold is set to the threshold at which False Acceptance Rate reaches a predefined value (2 percent in our experiments) – feed data of pseudo-imposters into the network and adjust the threshold until the FAR reaches 2 percent

Application of MRAN to Speaker Verification – Decision Threshold

- Many ways to calculate the decision threshold



ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function (RBF) Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

Application of MRAN to Speaker Verification – Experimental Results

- 76 experiments for all the speakers in the speaker set. Experimental results are then compared with those by M. Ran-Wai and K. Sun-Yuan
- Summary of experimental results over total 76 networks corresponding to 76 speakers:

MRAN Experimental Results

Rates	Equal Error Rate	False Acceptance Rate	False Rejection Rate	No. of hidden neurons
Avg. Value	**0.948%**	**3.183%**	**5.306%**	**57.46**
STDEV	**1.399%**	**2.471%**	**15.636%**	**24.85**

Application of MRAN to Speaker Verification – Experimental Results

- Geometric Mean Error (GME) – another statistic offering the desired ability of comparing systems

$$E_{GME} = \sqrt{E_{FAR} * E_{FRR}}$$

- GME is not perfect – it tends to decrease as the FRR is increased

Application of MRAN to Speaker Verification – Results Comparison

- Results compared with those by M. Ran-Wai and K. Sun-Yuan

NETWORKS	NUMBER OF HIDDEN UNITS	COMPLEXITY	EER	FAR	FRR	GME
RBF	61	917	7.46%	20.72%	53.93%	33.43%
MRAN	57	804	0.95%	3.18%	5.30%	4.10%
EC	10	922	0.44%	0.46%	14.99%	2.62%
EED	35	912	0.47%	1.78%	15.07%	5.18%
EEF	10	922	0.37%	3.52%	6.74%	4.87%

Application of MRAN to Speaker Verification – Results Comparison

- MRAN outperforms the conventional RBF and some of the EBF networks in terms of geometric mean error (GME) with the minimum complexity, i.e., number of parameters of the network

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function (RBF) Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

Conclusions

- With the minimum complexity, MRAN outperforms conventional RBF networks and
- its performance is comparable or even better than EBF networks

Future Works

- Extend MRAN to elliptical basis function
- Compare the extended EBF MRAN with RBF MRAN and other kinds of neural networks in terms of error measures and complexity

ANN for Biometric Applications

- Problem Statement: Speaker Verification
- Review of Speaker Verification
- Review of Radial Basis Function Networks
- Minimal Resource Allocation Networks (MRAN)
- Application of MRAN to Speaker Verification
- Experimental Results
- Conclusions and Future Work

Thank You

Q & A

Supplemental Slides

PRESENTED BY:

LI GUOJIE

SEPT. 2003

Elliptical Basis Function Networks

- RBF units are hyperspherical. High recognition accuracy can be achieved when the components of the training vectors are independent. If this is not the case, more basis functions are required.
- It would be beneficial if full covariance matrices could be incorporated into the RBF structure so that complex distribution could be represented without the need for using a large number of basis functions → EBF.

Elliptical Basis Function Networks

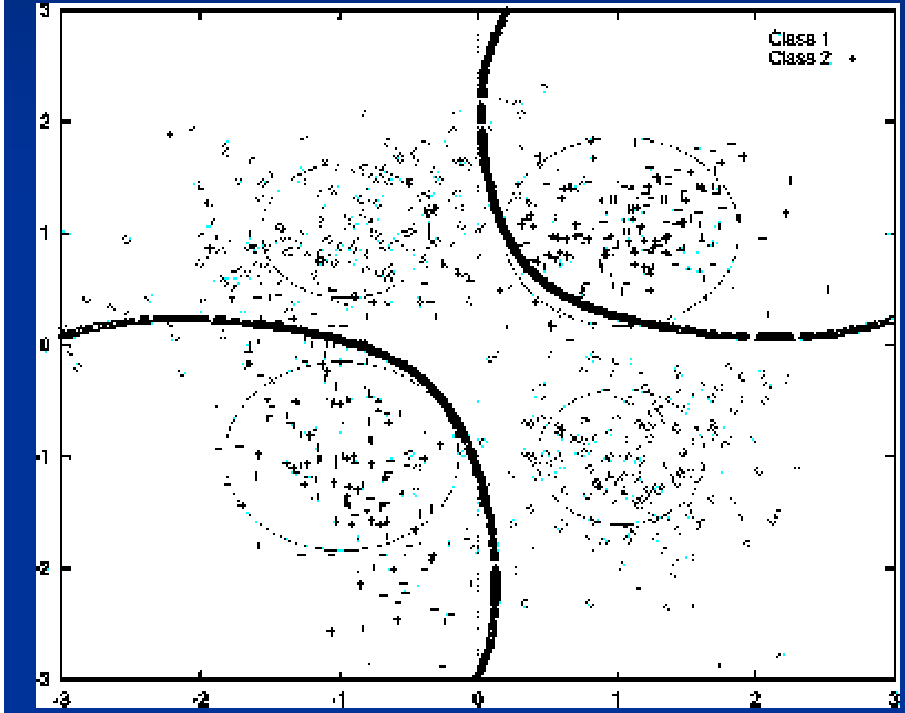
- EBF networks can be considered as an extension of the RBF networks
- The kth output of an EBF network with I inputs and K function centers has the form: (N inputs, M outputs, K hidden units)

$$y_m\left(\vec{x}_n\right) = w_{m0} + \sum_{k=1}^{K} w_{mk}\Phi_k\left(\vec{x}_n\right)$$

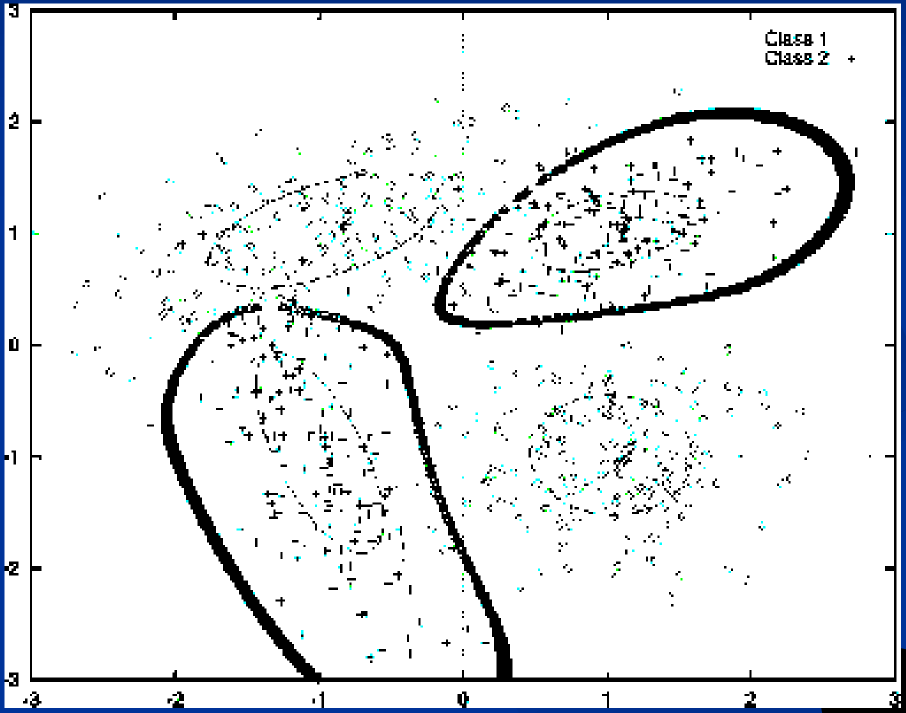
where $$\Phi_k\left(\vec{x}_n\right) = \exp\left\{-\frac{1}{2g_k}\left(\vec{x}_n - m_k\right)^T \Sigma_k^{-1}\left(\vec{x}_n - m_k\right)\right\}$$

and heuristically, $$g_k = \frac{3}{5}\sum_{j=1}^{5}\left\|\vec{m}_j - \vec{m}_k\right\|$$

EBF vs. RBF



RBF with 4 hidden units



EBF with 4 hidden units