A class-modular feedforward neural network for handwriting recognition

Authors: Il-Seok Oh, Ching Y. Sueh
Pattern Recognition 35 (2002)

Presented by: Nishanth Aravamudan

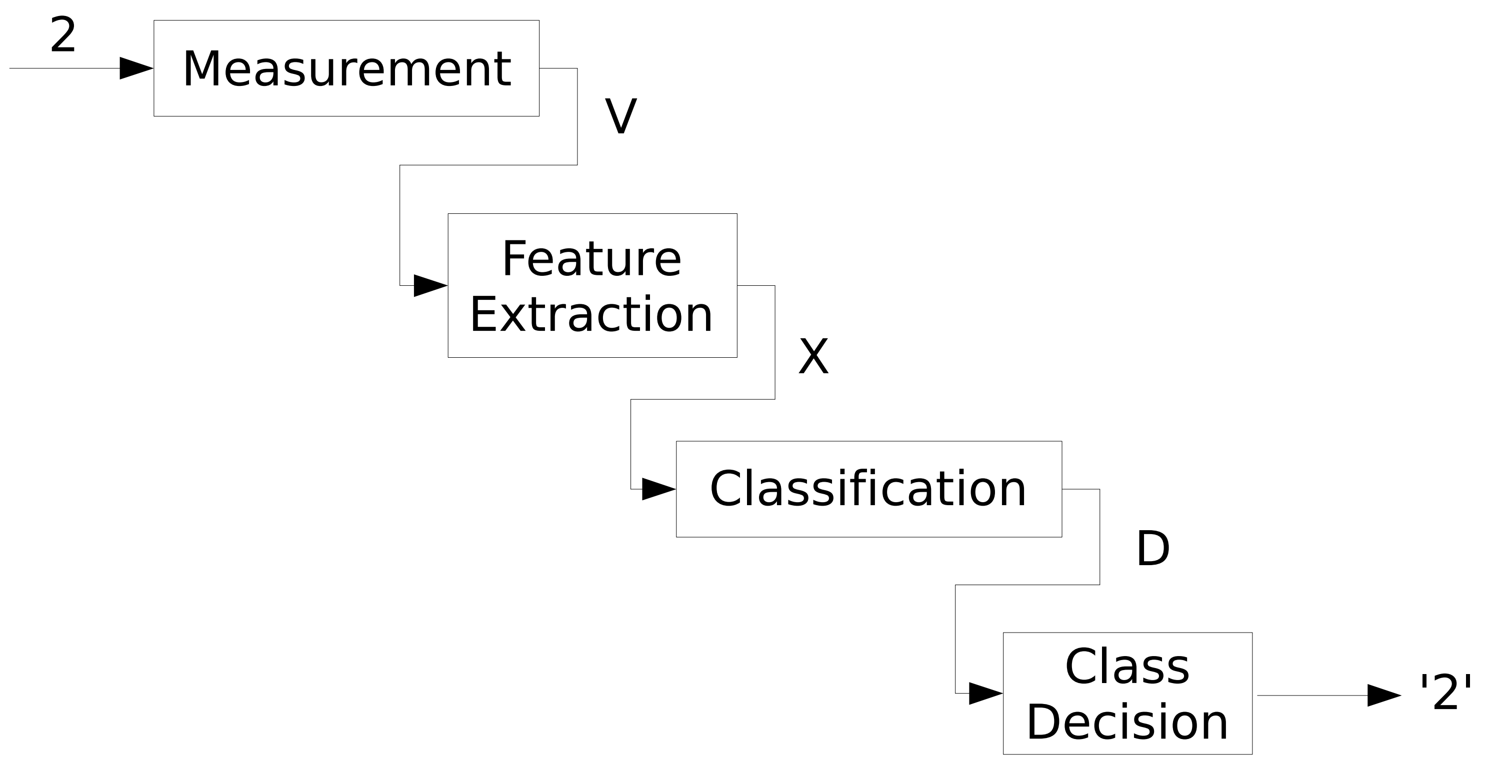
Results

- Class-modular approach superior to nonmodular approach
  - Convergence
  - Recognition
  - Large-set classification

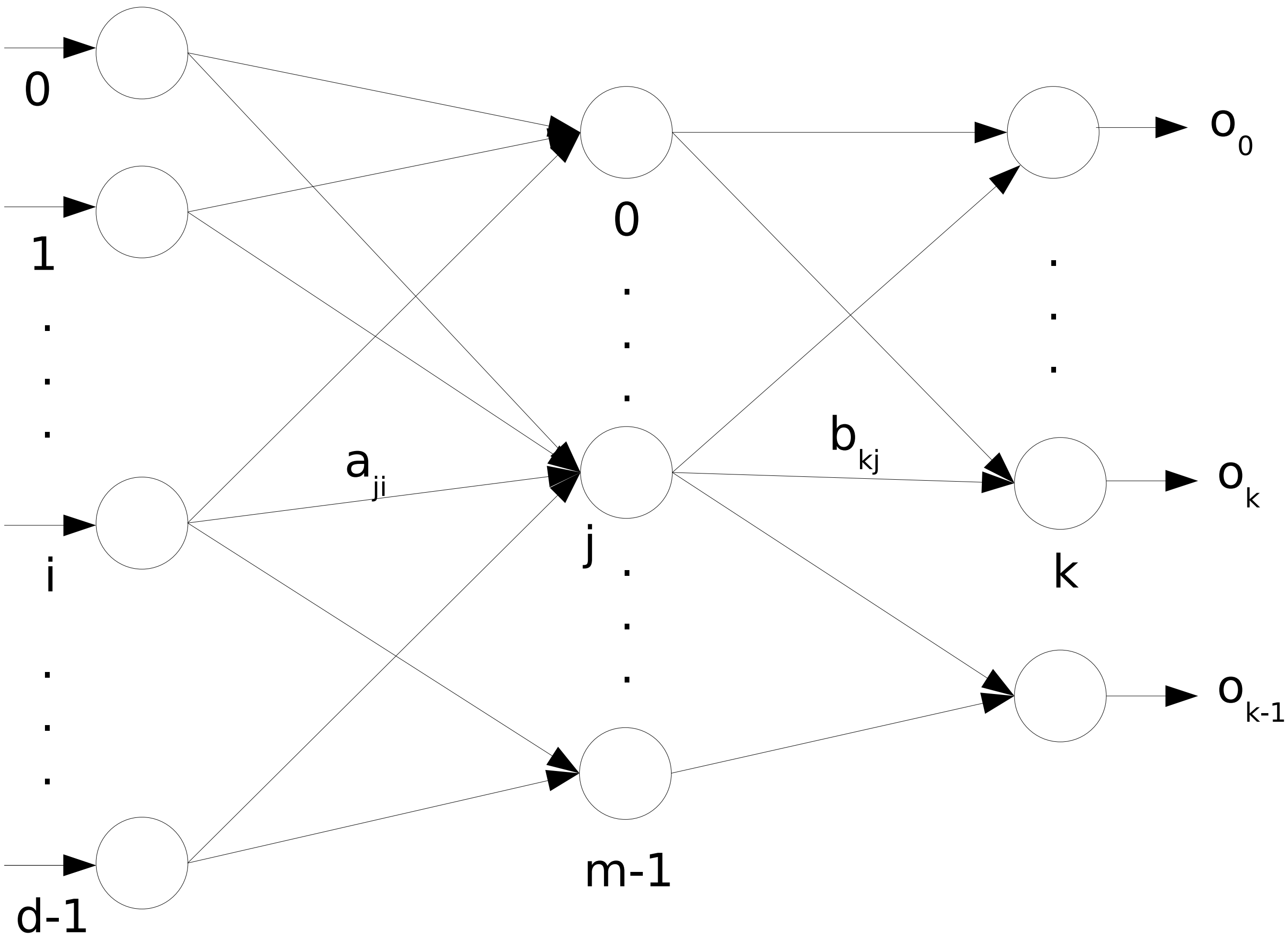
Datasets

- Handwritten numerals
  - 10 classes; 4,000/2,000 training/test samples
- English capital letters
  - 26 classes; 26,000/8,920 training/test 8,920 samples
- Touching numeral pairs (synthetic)
  - 100 classes; 50,000/1,374 training/test samples
- Korean characters in postal addresses
  - 352 classes; 26,460/10,560 training/test samples

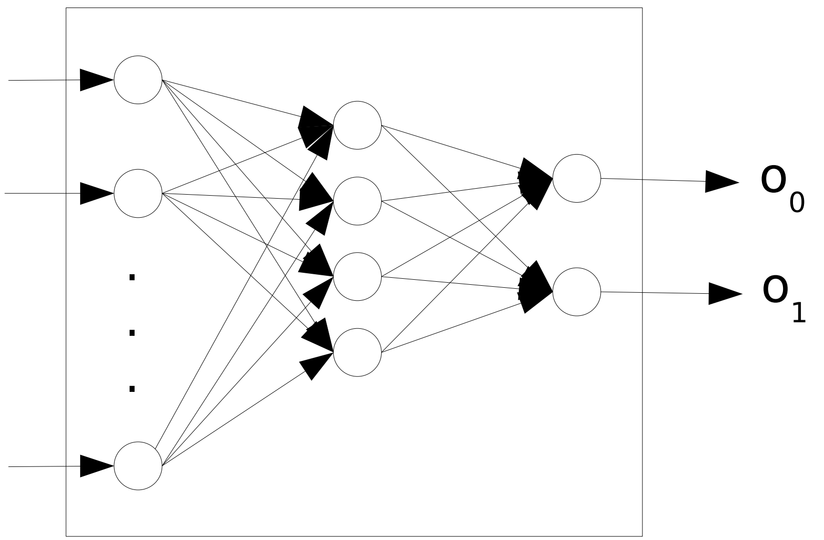
Typical nonmodular network

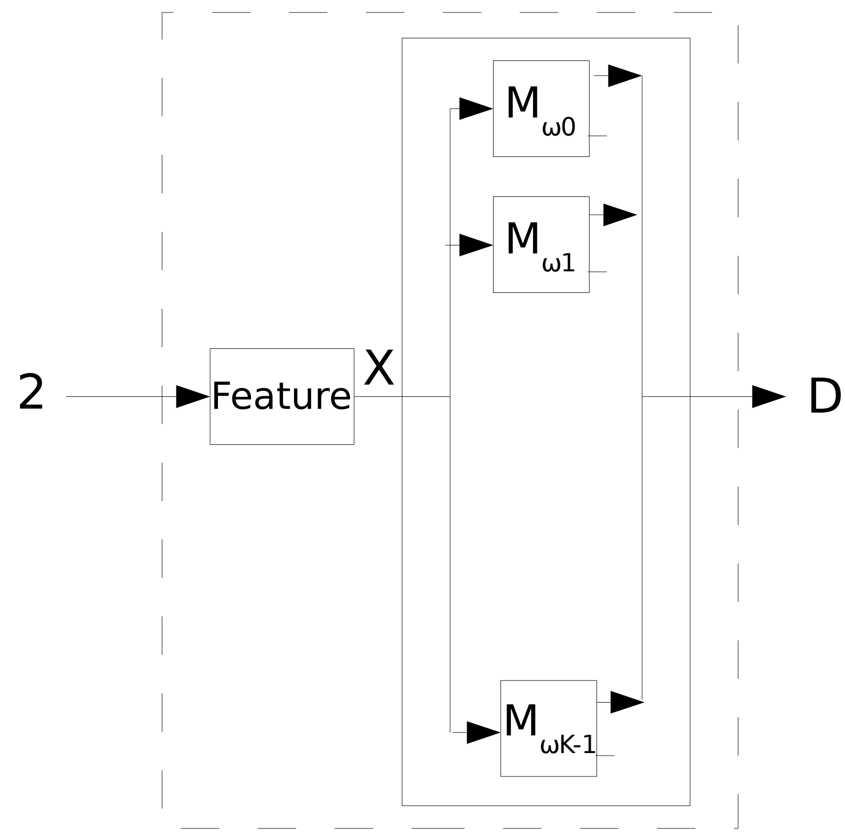


Nonmodular Network Architecture



Class-modular Network Architecture





Modular Architecture

- K 2-classifiers (M_{ω_i})
  - Each discriminates class ω_i (Ω_0) from K-1 other classes (Ω_1)
  - 4 hidden nodes
  - Apply input X to all K discriminators
    - o_0 integrated to form decision vector

Comparing Model Complexity

		Conventional	Modular
Number of nodes in layers	Input	d	d
	Hidden	m	m
	Output	K	2
Number of connections between layers	Input-Hidden	dm	dm
	Hidden-Output	mK	2m
	Total	dm + mK	(dm+2m)K
	Total under assumption	Assuming m = 2K	Assuming m = 4 4(d+2)K

More Model Complexity

	Number of total connections		Training Set Size / Classifier Size	
	Conventional	Modular	Conventional	Modular
Numerals	5,320	10,320	0.75	3.88
English capital letters	14,664	26,832	1.77	25.19
Touching numeral pairs	84,000	128,800	0.6	38.82
Hangul	425,216	357,632	0.06	24.25

Modular Training Algorithm

- M_{ω_i} trained independently
  - Partition and relabel training set appropriately for each classifier
  - Sigmoid activation function
  - Backpropagation
- Variables
  - $\eta = 0.2$
  - initial weights between 0.0 and 0.2

Some Details

- Training set not balanced between 2-class positives and negatives
- Terminate training if
  1. MSE $< \varepsilon_1$; or
  2. Most recent n epochs change the MSE less than ε_2 on the average; or
  3. Number of epochs exceed T
- Winner-takes-all determines the eventual classification given D

Convergence

Character Set	Convergence (epochs)		Final MSE	
	Nonmodular	Modular	Nonmodular	Modular
Numerals	212	156	0.012	0.012
English Capital Letters	181	173	0.116	0.063
Touching Numeral Pairs	300	268	0.248	0.098
Hangul	300	300	0.432	0.076

- MSE versus epochs curve “smoother” near termination for modular network

Performance (Recognition)

Character Set	Training Set (%)		Test Set (%)	
	Nonmodular	Modular	Nonmodular	Modular
Numerals	98.12	99.20	94.15	94.15
English Capital Letters	85.85	96.89	81.03	91.11
Touching Numeral Pairs	67.03	95.31	57.06	75.18
Hangul	26.66	93.53	22.46	68.75

- For capitals, modular test accuracy is better than nonmodular training accuracy
- Gap in recognition rate increases with number of classes
- Recognition rate versus epochs curve “smoother” for modular network

Results

- Class-modular approach superior to nonmodular
  - Convergence
    - Faster
  - Recognition
    - Better
  - Large-set classification
    - Scales

Conclusions

- Why is modular better?
  - Ratio of classifier size to training set size
  - Smaller subclassifiers with fewer parameters to estimate but full training set
  - Each sample is seen K times each epoch
- Future considerations
  - Class-dependent feature sets
  - Class-dependent classifier