3rd NOSE Short Course

Alpbach (Austria) 21-26 March 2004

Feature Selection Techniques

Professor Julian Gardner
School of Engineering
University of Warwick, UK

Some books on electronic noses

Acknowledgements

1988

Worked example

Dr Evor Hines
Dr Pascal Boilot

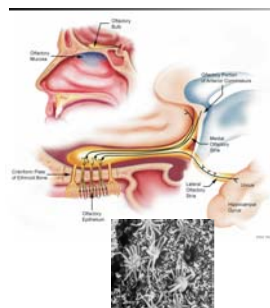
Contents

- Introduction to Feature Selection & Extraction
- Feature Selection Criteria
- Search Algorithms for Feature Selection
- Suboptimal Search Algorithms
  - Monte Carlo techniques with simulated annealing
  - Node pruning/growing algorithms in neural networks
  - Genetic algorithms
- Worked Example on Electronic Nose
- Summary

Introduction to Feature Selection

- Objective: Represent high-dimensionality data in a reduced number of data-sets
- Reason:
  - Simpler/faster subsequent analysis
  - Improved classification performance
  - Removal of redundant/irrelevant information

Dimensionality of human olfactory system



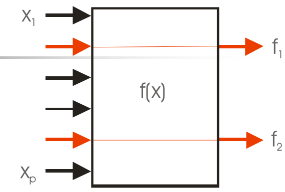
- 1-100 million olfactory receptor cells
- 300 genes that encode olfactory binding proteins
- 1,000s glomeruli nodes
- Mitral/tufted cells
- 2-3% genome coding!

From Handbook of Machine olfaction (2003), p3

Representing Data in Reduced Dimensions

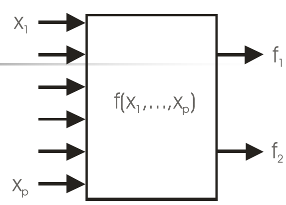
- Multivariate analysis perspective:
  - Ordination or geometrical methods
    - E.g. Principal components analysis
    - E.g. Multidimensional scaling
- Pattern recognition perspective:
  - Feature selection methods
  - Feature extraction methods
    - E.g. Linear discriminant analysis
    - E.g. Karhunen-Loeve expansion

Feature Selection



- Identify those variables *x* that do not contribute to the classification task, e.g. class separability in discrimination
- Seek *d* features out of a set of *p* measurements
- This is called:
  - Feature Selection in the Measurement or Sensor Space
  - Or just Feature Selection

Feature Extraction



- Find a transformation from the *p* measurements to a lower dimension space
- This is called:
  - Feature Selection in the Transformed Space
  - Or Feature Extraction

Transformation

- May be linear or non-linear
  - For linear see PCA/Karhunen-Loeve transform
- May be supervised or unsupervised
- For supervised case:
  - Maximise the class separability

Optimisation of Criterion Function *J*

- Feature selection
  - Find best subset X_d of size *d* over all subsets χd of the *p* possible measurements

$$J(\tilde{X}_d) = \max_{X \in \chi_d} J(X)$$

- Feature extraction
  - Find best transformation A of the variables x over the set of all allowable transforms

$$J(\tilde{A}) = \max_{A \in \mathrm{A}} J(A(x)) \qquad y = \tilde{A}(x)$$

What is criterion J?

- Some measure of distance or dissimilarity between distributions
  - E.g. Euclidean distance metric in CA

Simple Feature Selection

The Problem:

"Given a set of measurements on *p* variables, what is the best subset of size *d*?"

Thus we are not considering a transformation of the measurements, merely selecting those *d* variables that contribute most to the discrimination problem.

Simple Feature Selection

The Solution:

Evaluate the optimality criterion for all possible combinations of *d* variables selected from *p* and select the combination that maximises this criterion.

Thus we are not considering a transformation of the measurements, merely selecting those *d* variables that contribute most to the discrimination problem.

Size of Sensor arrays in e-Noses

Manufacturer	Technology	Number of sensors
Agilent Technologies	MS or GC/MS	550 masses
Airsense Analytics	MOS and MS	4 …
Alpha M.O.S.	Sensor array (MOS, CP, QMB) and MS	8, 16,….550
Applied Sensor	Field effect MOS, MOS, QMB	22 ….
Cyrano Sciences Inc.	CP	32
Electronic Sensor Technology	GC and SAW	100s … 8
HKR Sensorsysteme	QMB and MS	4 … 100s
Illumina Inc.	BeadArray fiber optic	100s
Microsensor Systems Inc.	SAW or GC	8 to 100s
Osmetech plc	CP	32
SMart Nose	MS	100s

Size of Search Space for *p* Sensor Array?

$$n_d = \frac{p!}{(p-d)!d!}$$

d Features	p variables	Number
4	8	70
4	32	35,960
10	25	3,268,760

$$N_d = \sum_{d=1}^{d=p} \frac{p!}{(p-d)!d!}$$

Osmetech array

Evaluate optimality criterion J for each set

- Optimal methods:
  - Exhaustive search methods
  - Accelerated search
  - Monte Carlo methods (e.g. simulated annealing and genetic algorithms)
- Suboptimal methods: trade off searching all space for computational efficiency

Feature Selection Criteria

- We need to find a means of measuring the ability of a feature set to accurately discriminate between two or more classes. Two ways:
- Choose the feature sets for which the classifier performs well on a separate test/validate set, e.g. percentage of correct classifications.
  - Feature set may differ with choice of classifier
- Estimate the overlap between the distributions from which the data are drawn and favour those sets with minimal overlap, i.e. maximise separability.
  - Feature set is independent of choice of classifier

Feature Selection Criteria

- Confusion matrix to calculate error rate
  - True versus apparent error rate?
- Probabilistic distance between two distributions
- E.g. average, Chernoff, Patrick-Fischer
- Estimating the multivariate PDF and then integrating it is very time consuming

Search Algorithms for Feature Selection

- Bottom-up Approach:
  - Start with the empty set (*d=0*) and build up incrementally
- Top-down Approach:
  - Start with the full set (*d=p*) and build up incrementally

Suboptimal Search Algorithms: Best Individual N

- Simplest one is to assign a discrimination power estimate to each of the features in the original set. Thus features are ordered such that:

$$J(x_1) \geq J(x_2) \ldots J(x_p)$$

- Select as our best set of N features with the best individual scores

$$\{x_i | i \leq N\}$$

Suboptimal Search Algorithms: Best Individual N

- Poor estimates when features in the original set are highly correlated – as often the case for electronic nose sensors.

Suboptimal Algorithms: Sequential Forward Selection (SFS)

- Bottom-up approach
- Start with null set
- Add a new feature that has maximum value of the optimality function J, i.e. maximum selection criterion
- When the best feature added makes the feature set worse terminate or when the maximum number of features is reached
- Disadvantage: cannot delete already added features that may be rendered redundant

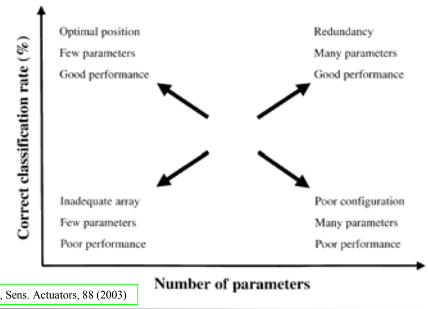
Suboptimal Algorithms: Sequential Backward Selection (SBS)

- Top-down approach
- Start with the complete set
- Delete a new feature that has minimum value of the optimality function J, i.e. minimum selection criterion
- When the worst feature eliminated makes the feature set worse terminate
- Disadvantage: computational more demanding than SFS

Suboptimal Algorithms: Plus L – take away r selection

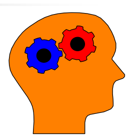
- Bottom approach with some back tracking L<r
- L features are added to the feature set using SFS and then the worst r eliminated using SBS
- Top-down approach with L>r

Sensor Selection and Optimal Feature Set



From Boilot et al., Sens. Actuators, 88 (2003)

Sensor Selection Techniques: Neural based



- Pruning/growing perceptrons
- Neurofuzzy/Genetic methods (Pardo)

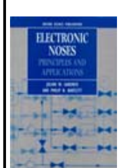
From Electronic Noses (2000), IOP, p83

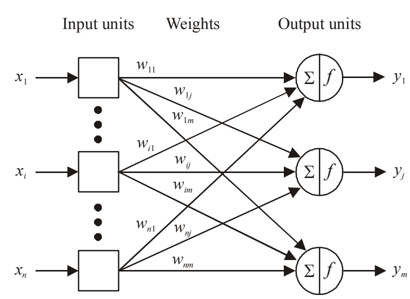
Worked Example:

- GA search engine/PNN classifier

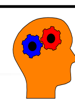
From Boilot et al., Sens. Actuators, 88 (2003)

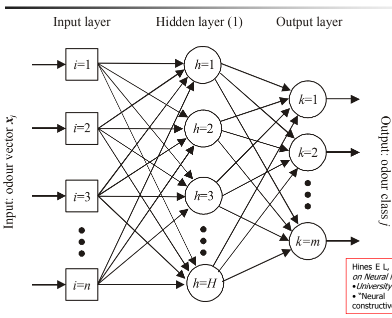
Feature selection: Pruning of Neural Nets





Feature selection: Pruning of MLP

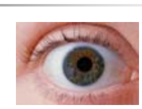


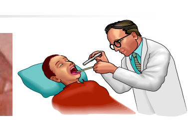


Hines E L, Gianna C and Gardner J W 1992 *Workshop on Neural Networks: Techniques and Applications,*
- *University of Liverpool, UK, 7-8 September 1992*
- "Neural network based electronic nose using constructive algorithms"

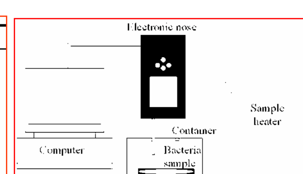
Worked Example: Bacteria detection

- 32 polymer sensor array (Commercial C320 unit)
- Two data-sets: eye bacteria and ENT bacteria

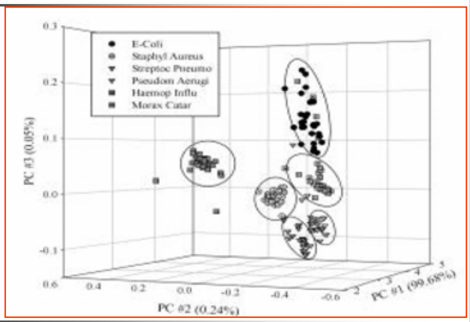




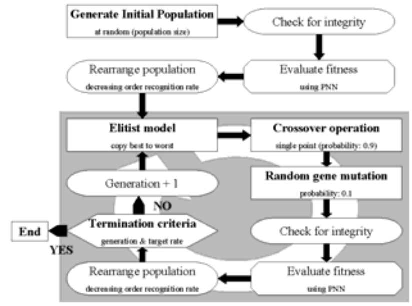
Eye bacteria experiment	no.
Staphylococcus aureus (sta)	60
Haemophilus influenzae (hai)	60
Streptococcus pneumoniae (stp)	60
Escherichia coli (eco)	60
Pseudomonas aeruginosa (psa)	60
Moraxella catarrhalis (moc)	60

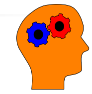


Feature Extraction: PCA Clustering



v-integer genes GA to Select Sensor Subset





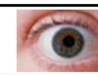
Eye Bacteria: Sensor Selection

- 6 bacteria classes
- 32 sensor array with PNN classifier
- SFS and SBS results both suggest 6 sensor subset

V-integer (No. of sensors)	Population (No. chromo.)	Random (Avg. init pop.)	GA Best % of all	GA Avg. %
12	12	83.7	90.6	90.4
10	15	82.0	90.6	90.2
8	20	78.6	90.0	89.4
6	25	75.6	90.6	89.4
4	40	69.2	89.4	87.8

From Boilot et al., Sens. Actuators, 88 (2003)

Eye Bacteria: Best Classifier



- Optimizing results for best set of 3 and 6 sensors

No. of sensors	Selected sensors	CA with 14 gps	FCM	MLP BPGDM (lr=0.1)	MLPBP LevMar	RBF sc=5	PNN sc=0.05
6	8,11,15, 23,31,32	65%	90% (16 clusters)	87.8%	96.7% (6×8×6)	70.6%	92.2%
3	8,11,23	50.5%	88.3% (13 clusters)	90.0%	93.3% (3×6×6)	65.0%	90.6%

From Boilot et al., Sens. Actuators, 88 (2003)

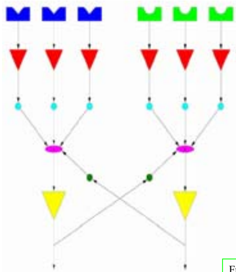
Generalised Summary

- Feature selection is the process of selecting from the original features or variables those important for classification
- Statistical approaches can be used that are optimal or suboptimal
- Some criteria J depend upon the classifier choice
- Search algorithms can be very computer intensive and genetic algorithms can be better that SFS/SBS methods
- Feature selection in a transformed space may be a better approach for low dimensionality problems

Sensor Selection and Electronic Noses

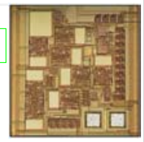
- Sensor Selection/Extraction and large arrays may help solve e-nose applications
- V-integer genes GA to select sensors and PNN classifier offer considerable benefits
  - Subset of 6 sensors identified in 5 runs and gave 90.6% *cf* 91.7% for 32 sensors
- Future Sensor selection could be adaptive, i.e. change with time as sensors drift, foul, etc?

Silicon Implementation of Olfactory Bulb



aVLSI Sensor Chip

From Gardner et al.,
IEEE Sensors Conf, 2003



aVLSI Neuromorphic Chip

From Pearce et al.,
Brain Inspired Cognitive
Systems, August, 2004

Thank You for you attention!