Information measures and likelihood ratios

Dr. David Lucy & Dr. Phil Rose

contact: d.lucy@lancaster.ac.uk

Lancaster University & ANU

Cepstral coefficients

Cepstral coefficients:

- Derived from the spectral properties of speech.
- Are the inverse discrete Fourier transform of the log of the discrete Fourier transform of a signal invariant in time (windowing).
- Considered as features amongst the DSP community.

Can be thought of, as far as comparison problems are concerned, as equivalent to any other feature set such as trace elements, or other physico-chemical properties.

Cepstral coefficients

Cepstral coefficients:

- In this case from the windowed Japanese /sh/ phoneme.
- Observations from 100 speakers.
- Extracted 12 cepstral coefficients from each individual.

40 replicates were made on two separate occasions for each individual. The between sessions variability has not been taken into account for this work.

Cepstral coefficients

The cepstral coefficients were for between speakers speech comparison:

- Problem is what to do with them.
  - For full multivariate approach some decomposable model based on the between items covariance matrix is appropriate.
  - A reasonable number of variables here - so suitable.
- Experience suggests that use of more than five or so variables has no practical benefit.

Cepstral coefficients

The number of variables in this case presents us with an ideal opportunity to look at the relationship between benefit, and amount of information used.

Two pronged approach:

1. Univariate: looking at how variables with difference amounts of information affect the evidence for identity.
2. Dimension reduction: can some linear function of our observations encapsulate a large amount of the evidence for identity?

LANCASTER
UNIVERSITY

F-ratios

An obvious choice favours those variables with the higher between item variance component, to the within item variance component.

Cepstral coefficient	F	Cepstral coefficient	F
cc.1	4.22	cc.7	0.36
cc.2	0.23	cc.8	0.17
cc.3	0.07	cc.9	0.52
cc.4	1.41	cc.10	2.92
cc.5	1.85	cc.11	3.78
cc.6	0.01	cc.12	3.18

Here F is $1/40$ that commonly calculated in variance component procedures as they tend to not be divided by the number of replicated observations.

Dimension reduction

PCA often employed - unfortunately unsuited when the focus of interest is some form of group membership (Krzanowski; 1988: p.291).

Krzanowski (p.297) recommends the *canonical variates* as the projection of choice: canonical variates are those transformations which maximise the F ratios.

Suppose $j = \{1, \ldots, p\}$ observations, replicated k times, of $i = \{1, \ldots, g\}$ items such that x_{ijk} is the k^{th}observation on variable j from item i.

Canonical variates

The j^{th} canonical variate y is:

$$y_j = \mathbf{A}'x$$

where $\mathbf{A}$ is the j^{th} eigenvector of $\mathbf{U}^{-1}\mathbf{C}$

where $\mathbf{U}$ is the within items covariance matrix, and $\mathbf{C}$ the between items covariance matrix.

Canonical variates

- The eigenvalues of $\mathbf{U}^{-1}\mathbf{C}$ are the F ratios for the within and between variances within the canonical variate space.
- The new variates y are mutually uncorrelated, but, the variate space itself has covariance $\mathbf{U}$.

The second property would require a slight reworking of existing multivariate methods were a complete set of j variates, or a subset of the variates used.

Not too sure how to go about that at the moment.

Eigenvalues of $\mathrm{U}^{-1}\mathrm{C}$

cv	F	cumulative
cv.01	1.81	0.25
cv.02	1.49	0.45
cv.03	0.84	0.57
cv.04	0.74	0.67
cv.05	0.56	0.75
cv.06	0.45	0.81
cv.07	0.40	0.86
cv.08	0.28	0.90
cv.09	0.21	0.93
cv.10	0.20	0.96
cv.11	0.15	0.98
cv.12	0.13	1.00

LANCASTER UNIVERSITY

Eigenvectors of $\mathrm{U}^{-1}\mathrm{C}$

	cv.1	cv.2	cv.3	cv.4	cv.5
cc.1	-0.27	-0.17	0.08	0.11	0.12
cc.2	0.04	-0.53	-0.13	0.05	0.03
cc.3	0.10	-0.21	0.31	-0.25	-0.14
cc.4	0.24	-0.27	0.21	-0.24	-0.15
cc.5	0.36	0.08	0.09	0.37	-0.10
cc.6	0.15	0.03	-0.04	-0.13	-0.43
cc.7	0.27	0.12	-0.05	-0.38	-0.25
cc.8	0.07	-0.02	-0.16	-0.28	0.00
cc.9	-0.14	-0.10	0.46	0.56	-0.42
cc.10	0.45	-0.48	0.43	0.34	0.00
cc.11	0.22	-0.28	0.51	0.20	0.34
cc.12	0.59	-0.48	0.38	-0.11	0.63

Likelihood ratios

Univariate likelihood ratios calculated (Aitken & Lucy; 2004) for pairwise comparisons for each element and variate:

- 100 items yield 100 same item comparisons,
- 4950 different item comparisons.

The recorded performance indicators were:

- False positives,
- false negatives,
- F-ratio,
- maximum empirical cross entropy (Ramos).

Performance - cepstrums

cepstrum	false +ves (%)	false -ves (%)	F-ratio	max ece
cc.1	34.08	12	4.22	0.64
cc.11	45.35	15	3.78	0.82
cc.12	43.54	17	3.18	0.80
cc.10	43.90	18	2.92	0.86
cc.5	38.51	8	1.85	0.69
cc.4	43.66	18	1.42	0.80
cc.9	41.01	17	0.53	0.77
cc.7	42.22	15	0.36	0.81
cc.2	33.76	10	0.21	0.65
cc.8	44.26	24	0.17	0.93
cc.3	39.78	19	0.08	0.83
cc.6	43.74	16	0.01	0.79

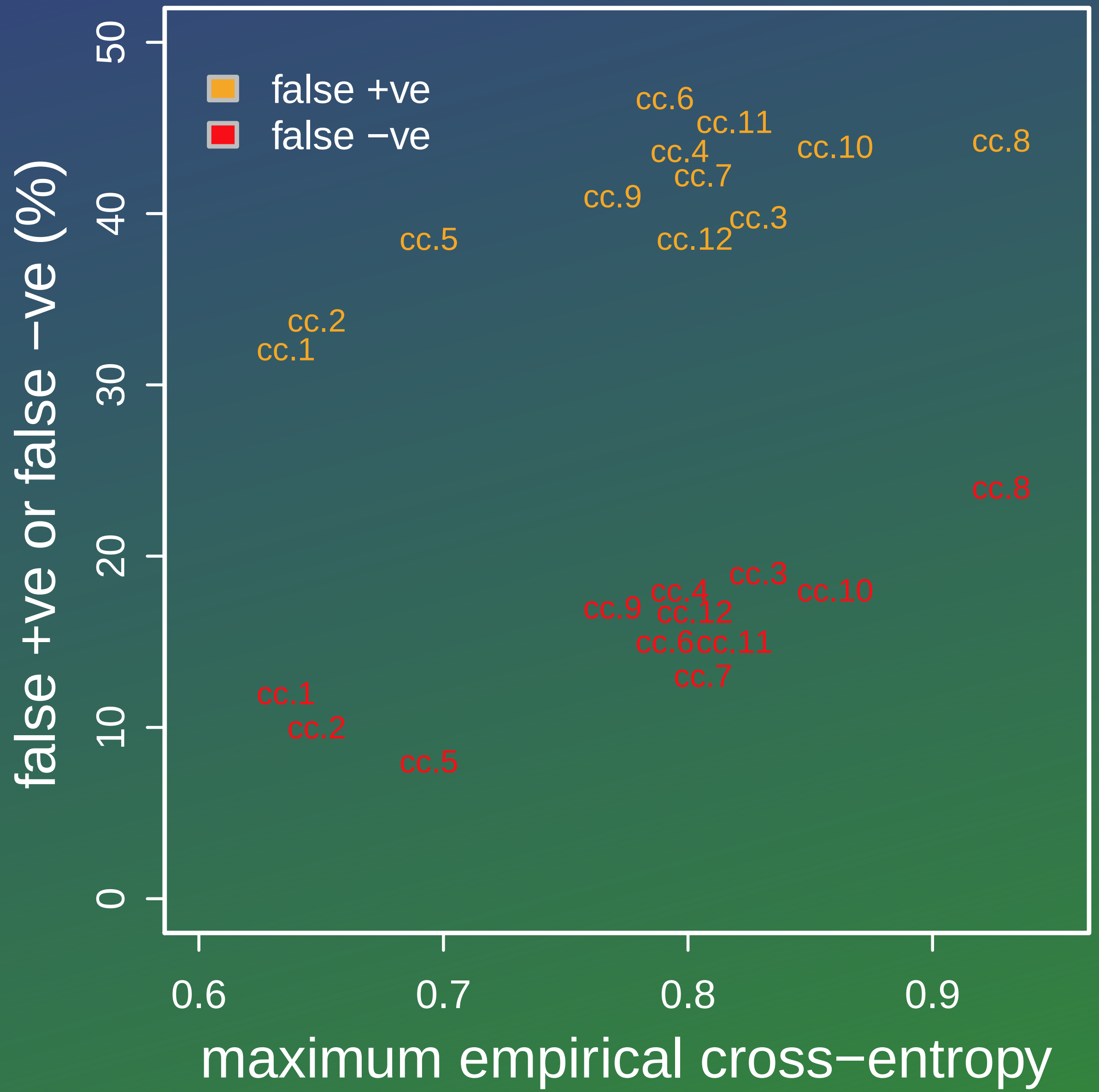
Ordered by F-ratio, then maximum ECE.

Performance - variates

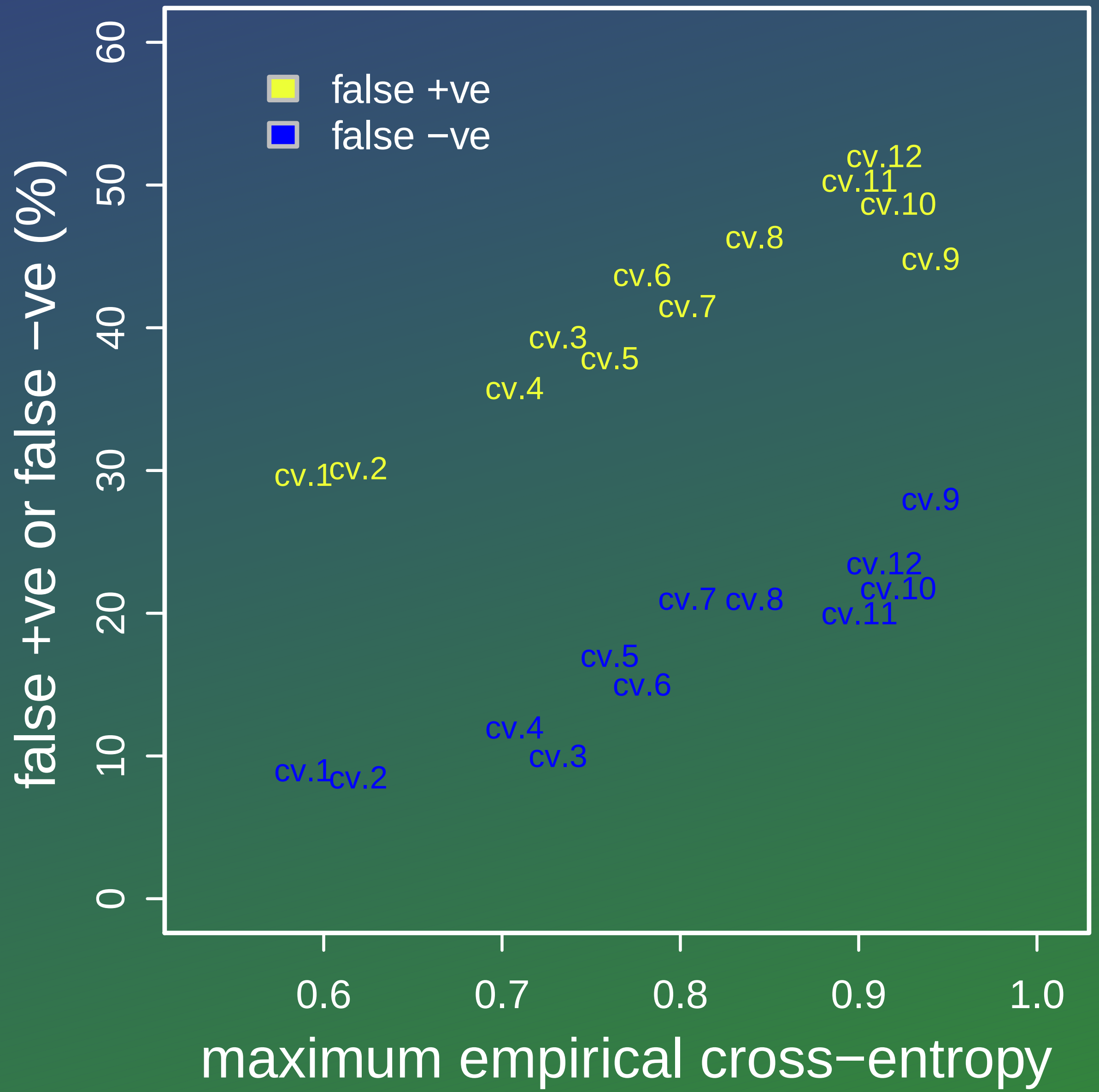
variate	false +ves (%)	false -ves (%)	F-ratio	max ece
cv.1	29.70	9	1.81	0.59
cv.2	30.16	9	1.49	0.62
cv.3	39.33	12	0.84	0.73
cv.4	35.78	12	0.74	0.71
cv.5	38.85	17	0.56	0.76
cv.6	43.70	15	0.45	0.78
cv.7	42.51	21	0.40	0.80
cv.8	46.34	21	0.28	0.84
cv.9	44.83	28	0.21	0.94
cv.10	49.19	21	0.20	0.92
cv.11	50.30	20	0.15	0.90
cv.12	51.03	21	0.13	0.91

Ordered by F-ratio, then maximum ECE.

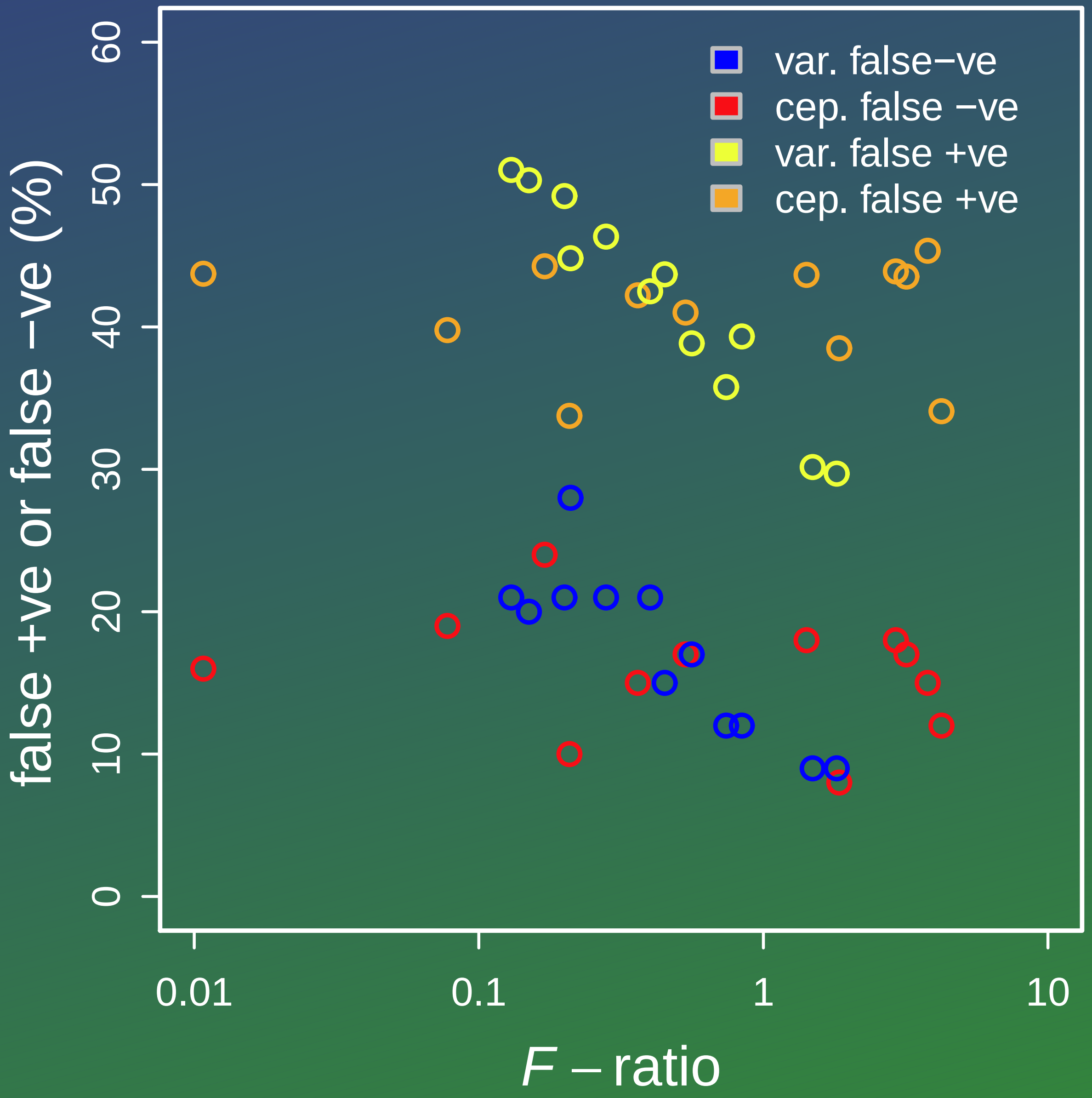
ECE - elements



ECE - variates



F-ratios



var. false−ve
cep. false −ve
var. false +ve
cep. false +ve
false +ve or false −ve (%)
0
10
20
30
40
50
60
0.01
0.1
1
10
F – ratio

Conclusions

- For reasonable spaces canonical variates may well be a useful alternative to other multivariate approaches:
  - easy to calculate,
  - have given better outcomes than the cepstral coefficients,
  - are “lossy”,
  - dimensions in their space are non-orthogonal, but known.
- Measures of “performance” can be approximately known in advance of calculation.
  - Knowledge of the F-ratio gives some knowledge of performance measure.

LANCASTER UNIVERSITY

Conclusions

The “lossy” nature of canonical variates may be an impediment to their use:

- May be that under rules of evidence **all** evidence must be used.
- Hence initially Aitken *et al.* used all observations.
- If losses of information are acceptable then multivariate comparison problems might become more manageable.

In a sense, because not all elements are observed, then even the most exhaustive of elemental analyses incur some loss of information.

Acknowledgements

The authors should like to thank Dr. Takashi Osaniai (ANU), and Dr. Chris Sherlock (Lancaster University).