Hyvärinen A., Hoyer P. O., Inki M.:

The Independence Assumption: Analyzing The Independence Of The Components By Topography

Structure of Presentation

- Motivation of the topographic ICA (TICA)
- Background
- TICA model
- Experiments
- Summary

Motivation of the topographic ICA

- 'Weakness': in classical ICA,

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

  the estimated components s_i have no particular order and are sometimes diffucult to interpret

- Observation: in practice, estimated components often violate the independence assumptions

- Remedy: estimation of the residual dependence for ordering the components

Motivation of TICA

How to estimate this 'residual' dependence?:

- the use of higher-order cross-cumulants (mutual information)
- Note: because the s_i do not form a well-defined set, simple dependency estimation from the estimates of the s_i can be problematic

How to utilize these estimates?:

- estimates define *topographic order* between the components

Background: Independent Subspace Analysis

- in the ISA it is assumed that the s_i can be divided into m-tuples, such that the s_i inside a given m-tuple can be dependent on each other, but dependencies between different m-tuples are not allowed
- in addition, the probability distributions for the m-tuples of s_i are spherically symmetric, that is, they depend only on the norm
- ISA finds invariant-feature subspaces from the data
- TICA can be seen as a generalization of the ISA

Background: ISA likelihood

- log-likelihood of the ICA model, given T observations $\mathbf{x}$, is

$$\log\ L(\mathbf{w}_i, i = 1, .., n) = \sum_{t=1}^{T} \sum_{i=1}^{n}\ \log\ p_i(\mathbf{w}_i^T \mathbf{x}(t)) + T\ \log|det\mathbf{W}|,$$

  where p_i are the density functions of the s_i

- the likelihood of the independent subspace analysis is

$$\log L(\mathbf{w}_i, i = 1, .., n) = \sum_{t=1}^{T} \sum_{q=1}^{Q} G(\sum_{i \in S_q} (\mathbf{w}_i^T \mathbf{x}(t))^2) + T\ \log|det\mathbf{W}|,$$

  where $G(\sum_{i \in S_q} s_i^2)$ gives the logarithm of the probability density inside the q-th m-tuple of s_i

Background: ISA properties

- because of the assumption that the distribution in the subspace is sparse, G is chosen to reflect that property

$$G(\sum_{i \in S_j} s_i^2) = \log p_q(s_i, i \in S_q) = -\alpha[\sum_{i \in S_j} s_i^2]^{\frac{1}{2}} + \beta$$

- this implies that the two components s_i and s_j that belong to the same subspace tend to be non-zero simultaneously = they have *correlated energies*

Topographic ICA model

Key idea: components close to each other in the topography are not assumed to be independent, whereas the components not close to each other are independent

- allowed dependence is *correlation of energies*, which means

$$cov(s_i^2, s_j^2) = E\{s_i^2 s_j^2\} - E\{s_i^2\}E\{s_j^2\} > 0$$

- topography can be defined by lattice or grid

TICA model 1: modified ISA

TICA is a generalized ISA in the sense that it models a dependence between all components neighbouring in the chosen topography, not just inside the m-tuples.
TICA model 1 defined by log-likelihood is

$$\log L(\mathbf{w}_i, i = 1, .., n) = \sum_{t=1}^{T} \sum_{j=1}^{n} G\Big(\sum_{i=1}^{n} h(i,j)(\mathbf{w}_i^T \mathbf{x}(t))^2\Big) + T \log|det\mathbf{W}|,$$

where $h(i,j)$ is a neighbourhood function expressing the strength of the connection between i:th and j:th components.

Learning rule: $\Delta \mathbf{w}_i \propto E\{\mathbf{x}(\mathbf{w}_j^T \mathbf{x}) r_i\}$,
where $r_i = \sum_{k=1}^{n} h(i,j) g(\sum_{j=1}^{n} h(k,j)(\mathbf{w}_j^T \mathbf{x})^2)$ and g is a derivative of G.

TICA model 2: The generative model

The generative TICA model defines the distribution of the components s_i as follows:

- the variances σ_i^2 of s_i are assumed to be random variables
- s_i are generated independent from each other, given their variances
- dependence among the s_i is implied by the dependence of their variances

TICA model 2: The generative model

- the variances are generated by

$$\sigma_i = \phi(\sum_{k=1}^{n} h(i,k)u_k),$$

  where u_i are 'higher-order' independent components and ϕ some scalar nonlinearity

- the actual independent components are generated as follows

$$s_i = \phi(\sum_{k=1}^{n} h(i,k)u_k)z_i = \sigma_i z_i,$$

  where z_i is a random variable with the same distribution as s_i given that σ_i^2 is fixed to unity

Basic Properties of the TICA model

- all the components s_i are uncorrelated
- components far from each other in the topography are independent
- components near to each other tend to be non-zero at the same time (correlation of energies)
- if every $p(s_i|\sigma_i^2)$ is assumed to be Gaussian, unconditional distributions of s_i are mixtures of Gaussians and have positive kurtosis (i.e are super-Gaussian), which is the same property as the independent components of the real data have
- classic ICA is a special case of the TICA (when $h(i,j) = \delta(i,j)$)

TICA vs. Other topographic mappings (OTM)

1) every vector represents a direction in TICA
   (OTM: usually a single point)

2) components in TICA are estimated based on their independence
   (OTM: usually similar to VQ/clustering)

3) in TICA similarity defining the topography is defined by higher order correlations
   (OTM: usually by Euclidian similarity)

Experiments: Feature extraction of image data

- basis vectors obtained were similar to those obtained by ordinary ICA; in addition they had a clear topographic organization
- correlations decreased as the distance increased
- neighbouring vectors tend to have same orientation and frequency, whereas their phases were very different $\rightarrow$ neighbourhood of basis vectors functions as a complex cell

Other experiments

- speech data: features organized according to frequency and phase
- MEG data: found signals similar to previous MEG results; in topography: clustering according to physical source of the signal

Summary

- topographic ICA is a generative model combining ICA with topographic mapping
- distance in the topographic space is represented by the higher order correlations
- estimation of the model by gradient learning rule
- results from the experiments imply that the neighbourhoods have properties similar to complex cells
- topographic ordering helps in the interpretation of the results with the high-dimensional data