Characterizing Unknown Events in MEG Data with Group Factor Analysis

Aalto University
UNIVERSITY OF HELSINKI
HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

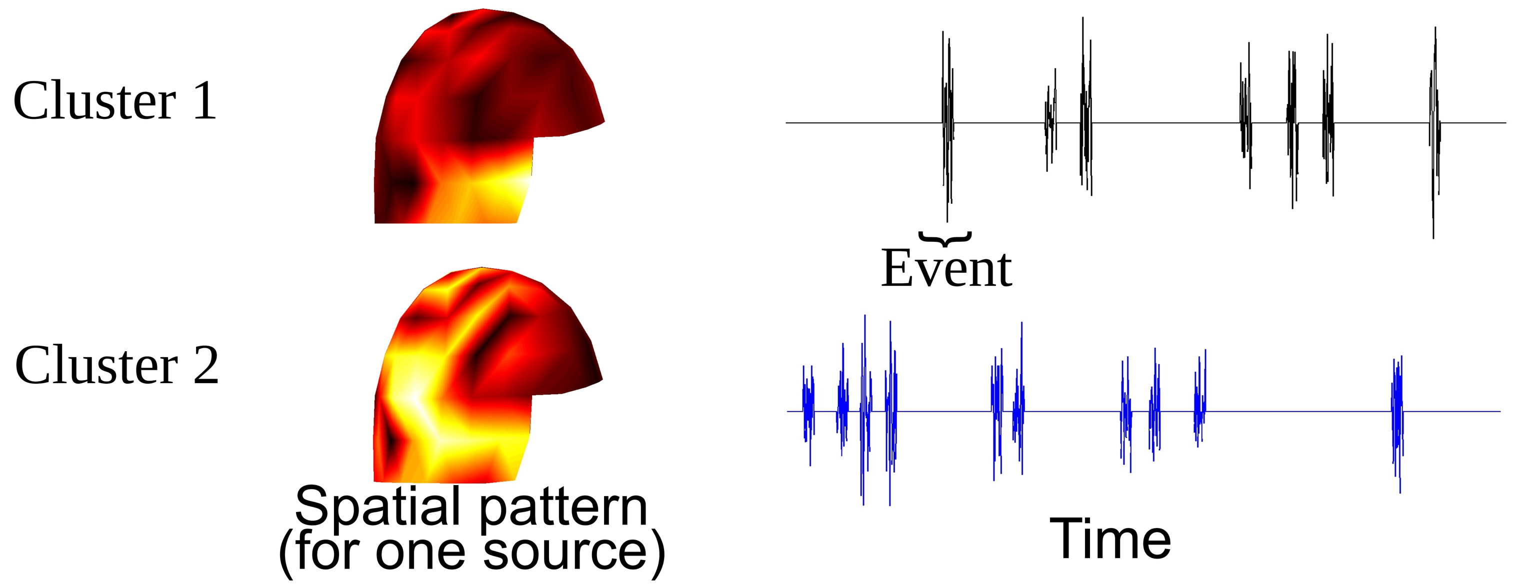
Sami Remes[1], Arto Klami[2], Samuel Kaski[1,2]

1. Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University

2. Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki

1. Motivation

- Modelling data from multiple data sources
  - Analysis of two-person MEG data that were recorded simultaneously from a pair of subjects, for example
- Interested in finding dependencies between the data sources
  - Typical solution by models such as Canonical Correlation Analysis (CCA) or Group Factor Analysis (GFA)
  - These assume that there exists global correlations over the length of a whole experiment, which is not a reasonable assumption in some cases, such as when considering the MEGs of two subjects without very specific stimulus
- Propose a joint model for multiple data sources, assuming that correlations only exists during temporally localized time windows or events, with similar events clustered into "event categories"



2. Group Factor Analysis

- GFA extends (Bayesian) CCA to model correlations between multiple data sources, or can be seen alternatively as extending factor analysis from single variables to groups of variables
- Based on group-sparse ARD prior that selects whether a factor is relevant for explaining a specific data source, a group of variables
- Likelihood for i'th sample of m'th data source written as

$$\mathbf{y}_i^{(m)} | \mathbf{W}, \mathbf{z}_i, \tau_m \sim \mathcal{N}\left(\mathbf{W}^{(m)}\mathbf{z}_i, \tau_m^{-1}\mathbf{I}\right)$$

- The latent variable is shared for all data sources to allow modelling the correlations between them
- Group-wise ARD prior is hierarchically constructed as follows

$$\mathbf{w}_k^{(m)} | \ \alpha_k^{(m)} \sim \mathcal{N}\left(\mathbf{0}, \left(\alpha_k^{(m)}\right)^{-1}\mathbf{I}\right) \qquad \alpha_k^{(m)} \sim \mathcal{G}(10^{-14}, 10^{-14})$$

- The M×K matrix consisting of the hyperparameters then acts as a kind of loading matrix that indicates which factors "load" on which data sources; large value of the parameter implies that the factor is irrelevant in modelling the source

3. Proposed Model

Model

- Assumes that most variation explained by source specific independent "noise" models, each source with their own sets of latent variables and loadings
- During each event, the samples are modelled with additional set of latent variables that are shared between sources to model the correlation, with group-sparse ARD on the loadings
- Similar events are characterized by assigning them to exactly one event category; each category has a separate loading matrix
- Model written with the help of an indicator variable γ_i^e as

$$\mathbf{y}_i^{(m)} \sim \mathcal{N}\left(\mathbf{W}^{(m)}\mathbf{z}_i^{(m)} + \sum_{e=1}^{E} \gamma_i^e \hat{\mathbf{W}}_e^{(m)} \hat{\mathbf{z}}_i^e, \tau_m^{-1}\mathbf{I}\right)$$

Inference

- Variational Bayes is used for most parameters (latent variables, loading matrices, noise precisions)
- Indicator variables are instead handled via a type-II maximum likelihood optimization step intertwined with the VB updates to maximize the variational lower bound, in order to provide binary decisions for whether an event belongs to a certain cluster/category

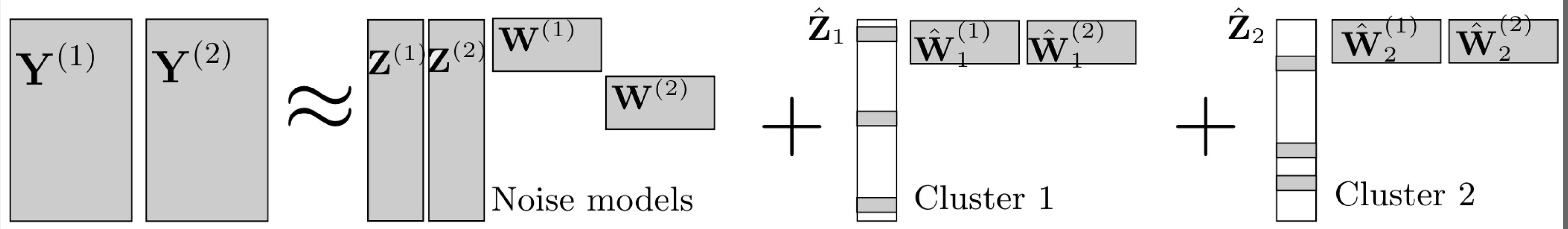


Figure. Factorization of the data matrices. Gray color indicates non-zero elements. For samples that do not belong to any event the two data sources are independent and modeled with the noise models. The samples within events choose one of the event categories and use the factors specific to that category for modeling correlation during the event.

4. Demonstration

Data description

- Collected in an experiment where pairs of participants were instructed to play a word game, where they took turns in uttering isolated words which were supposed to make up a sensible story
- Each of these isolated words was considered an event, which was defined as starting at the approximate beginning of a word and lasting about half a second
- Data were preprocessed first with signal space separation (SSS) to remove magnetic fields not emanating from brain activity; then downsampled to ~67 Hz and high-pass filtered at 3 Hz
- All 204 gradiometer channels were then used then used as features in our extended GFA model

Results

- We evaluated the proposed model by assessing whether two event categories as captured by the model correspond to the known dichotomy of two different speakers
- Results statistically significant for all pairs, and our method greatly outperformed simple comparison methods using PCA/GFA for dimensionality reduction and k-means for clustering the individual samples within an event, which was then clustered according to a majority rule

Table: Clustering accuracy of words with respect to the speaker. For each pair the best method is written in boldface.

Pair	Number of events	Proposed model	GFA + k-means	PCA + k-means
A	102	**0.90**	0.55	0.57
B	104	**0.92**	0.51	0.51
C	172	**0.62**	0.55	0.59
D	116	**0.76**	0.61	0.59
E	88	**0.96**	0.50	0.60
F	170	**0.75**	0.57	0.51
G	160	**0.81**	0.58	0.52

5. Conclusions

- We presented a novel CCA-based model that looks for temporally localized correlations during pre-defined events
- We demonstrated it on a two-person MEG data set, and showed that the model was able to successfully separate the two event categories in an unsupervised manner, and that MEG contains informative features that can be found by jointly modelling the two data sources

References

1. Seppo Virtanen, Arto Klami, Suleiman A Khan, and Samuel Kaski. **Bayesian group factor analysis**. AISTATS 2012.
2. Pamela Baess, Andrey Zhdanov, Anne Mandel, Lauri Parkkonen, Lotta Hirvenkari, Jyrki P Mäkelä, Veikko Jousmäki, and Riitta Hari. **MEG dual scanning: a procedure to study real time auditory interaction between two persons**. Front. Hum. Neurosci., 6, 2012.

Acknowledgements

This work was financially supported by the Academy of Finland (Finnish Centre of Excellence in Computational Inference Research COIN, grant no 251170; grant no 266969; BrIAn project of LASTU research program, no 135198). We thank Prof. A. Hyvärinen and Prof. R. Hari, the leaders of the BrIAn project, for fruitful collaboration. The dual-MEG data set was collected in Brain2Brain project funded by the European Research Council (Advanced Grant #232946 to Riitta Hari, Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University). We thank P. Baess, R. Hari, T. Himberg, L. Hirvenkari, V. Jousmäki, A. Mandel, J. Mäkelä, J. Nurminen, L. Parkkonen, and A. Zhdanov for the possibility to use the anonymized dual-MEG data, A. Mandel for preprocessing of the data set, and L. Hirvenkari for the stimulus timing. Computational resources were provided by Aalto Science-IT project.