FEATURE-TASK BICLUSTERING IN MULTITASK REGRESSION

- WORK IN PROGRESS -

Madalina Fiterau
mfiterau@cs.cmu.edu

Leila Wehbe
lwehbe@cs.cmu.edu

Carnegie Mellon University
MACHINE LEARNING DEPARTMENT

WiML 2012

MOTIVATION

We are interested in multi-task regression problems in which:

- a group structure can be identified among tasks such that features within a group can be learned using approximately the same set of input features
- tasks in different groups depend on disjoint/nearly-disjoint highly sparse sets of features

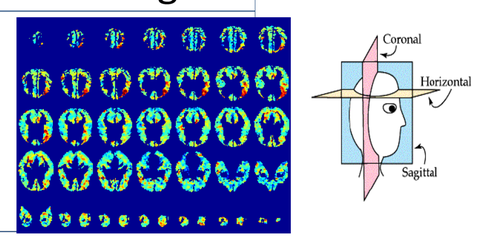
Unlike other related work, we do not assume :

- Information on the group structure among the tasks
- domain knowledge the features that are relevant to each group

MEANING REPRESENTATION IN THE BRAIN

9 subjects saw word-picture pairs from 12 semantic categories:

	Semantic Features	fMRI Image
chisel	Is it edible? 1 Is it alive? 1 Can you hold it? 5 Is it bigger than a car? 1	



Does the brain represent semantic categories in disjoint patterns?

ALGORITHM

ORTHOGONAL BICLUSTERING METHOD

Idea: start by finding overlapping bi-clusters, then gradually increase the orthogonality restrictions.

Soft Clustering
- Initialize set of bi-clusters B
- While there are unassigned tasks
  - $\beta_k = argmin_\beta |Y - XB|_2^2 + \lambda_1 \beta^T B + \lambda_2 B^T \beta$
  - Retrieve cluster from β_k and add to B
- Obtain optimal values for learned structure

Separate Clusters
- While items in B are not orthogonal
  - Increase λ_1 and λ_2
  - Perform block coordinate descent over bi-clusters until convergence

BI-CLUSTER RETRIEVAL HEURISTICS

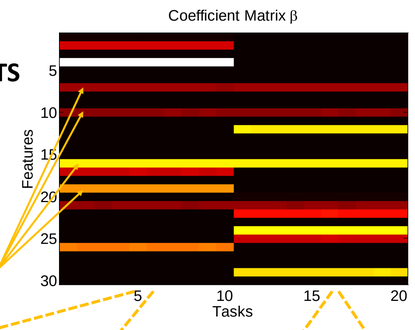
1. Common sparsity patterns for features and tasks
   - for sparse structure recovery
2. Group features used frequently for same tasks
   - task-focused grouping
3. Group tasks by how well they are learned
   - targets improved accuracy

EXAMPLE

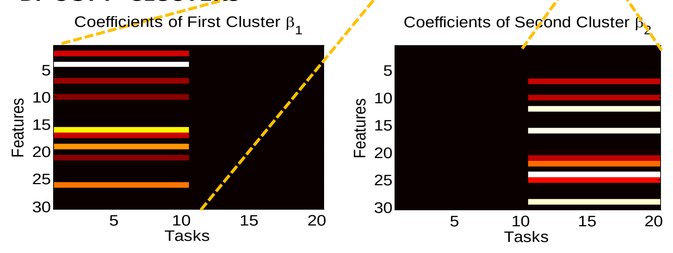
A. TRUE COEFFICIENTS

- Two task groups
- Each group depends on its own set of features plus a set of common ones

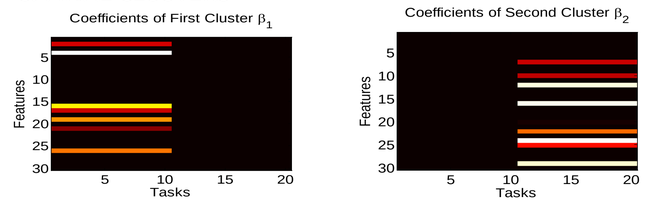
Informative Features Common Across Tasks



B. 'SOFT' CLUSTERS

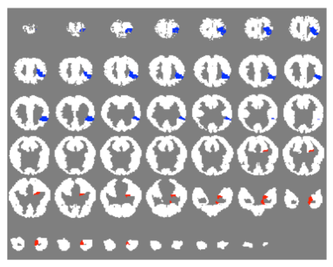


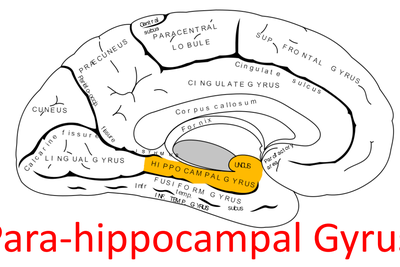
C. FINAL CLUSTERS



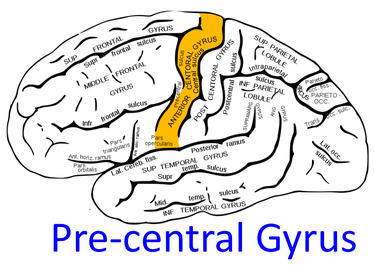
EXPERIMENTS

FMRI EXPERIMENT





Para-hippocampal Gyrus
recognition of places



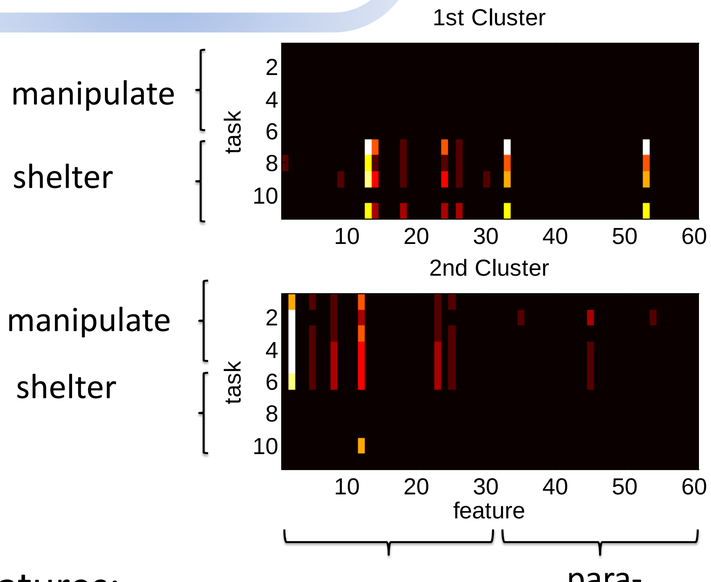
Pre-central Gyrus
motor cortex

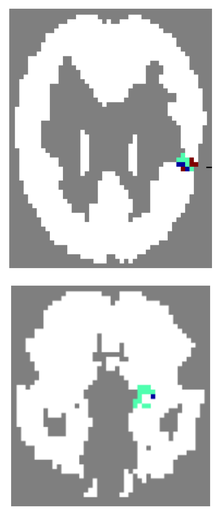
We simplify the problem to 2 categories:

- **Buildings**: 'chisel' 'hammer' 'pliers' 'saw' 'screwdriver'
- **Tools:** 'apartment' 'barn' 'church' 'house' 'igloo'

The method recovered a clustering of the manipulate and shelter features:

- Both features are represented in the pre-central voxels
- The clustering does not correspond to the predicted anatomical separation





CONCLUSION

Contributions

- Introduced a new method for multitask regression to recover the grouping between tasks and features
- The method correctly retrieves the underlying bicluster structure in artificial data
- Obtained preliminary results on real data, which verified the correlation of nearby voxels, but the hypothesis is not yet fully verified

Research Directions

- Determine if results hold across multiple subjects
- Test to see if they generalize to other semantic categories – such as food – and different brain regions – the gustatory cortex for instance
- Explore theoretical properties of the method; under what conditions is the coefficient matrix recovered and what bounds on recovery error and risk