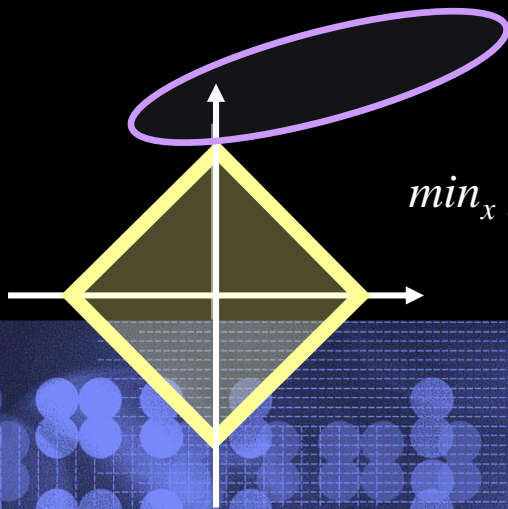


Using Machine Learning to Learn about Brain



$$\min_x \|y - Ax\|^2 + \lambda \|x\|_1$$

Irina Rish (and many collaborators)

Computational Biology Center
IBM T.J. Watson Research Center, NY

Collaborations and Recent Books

- **Computational Biology Center @ IBM Watson:**

- Guillermo Cecchi, James Kozloski, Jeremy Rice

- **IBM Watson**, other departments

- Steve Heisig, Ravi Rao, Sasha Aravkin, Melissa Carroll

- **Neurospin** (France):

- JB Poline, Bertrand Thirion et al

- **Mt Sinai** (New York):

- Rita Goldstein

- **Northwestern U.** (Chicago)

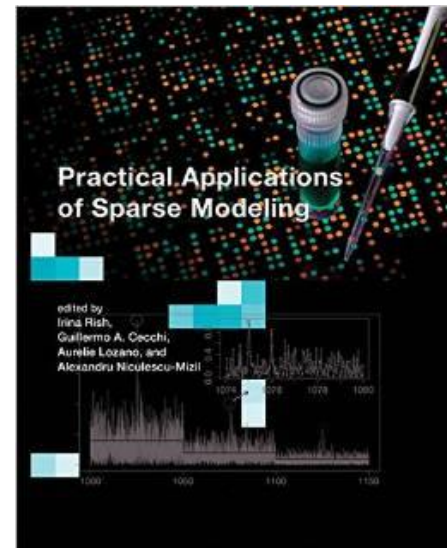
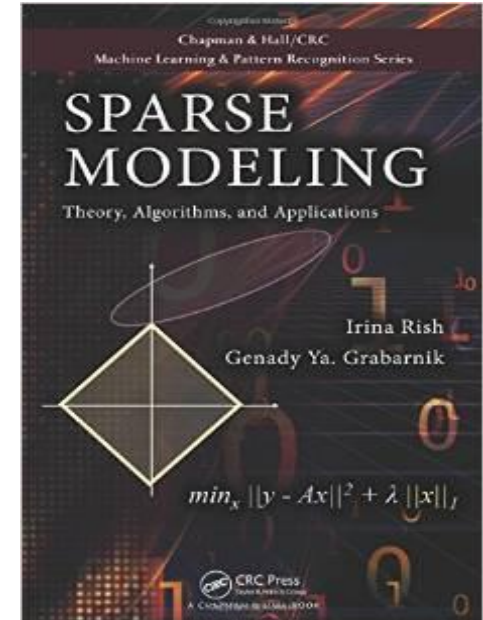
- A.V. Apkarian

- **SUNY Stony Brook:**

- Jean Honorio (now at MIT), Dimitris Samaras

- **Lehigh University:**

- Katya Scheinberg



Functional Magnetic Resonance Imaging (fMRI)

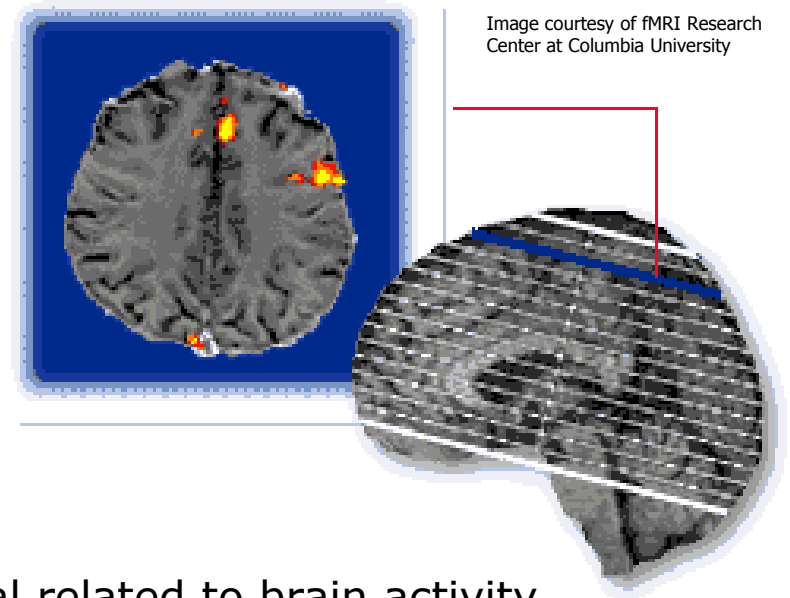


Image courtesy of fMRI Research Center at Columbia University

- Blood-oxygen-level-dependent (**BOLD**) signal related to brain activity while subject performs some task in scanner

- 4D '**brain movie**': a sequence of 3D brain volumes

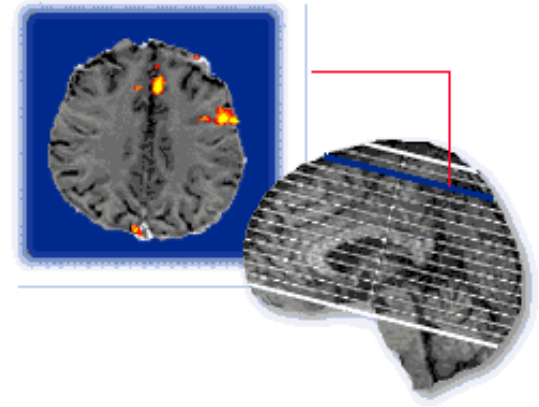
3D voxels $\sim 3 \times 3 \times 3$ mm, time repetitions (TR) $\sim 1-2$ s

- Challenge: **high-dimensional, small-sample data**

10,000 to 100,000 variables (voxels), but only **100s of TRs** (samples), and **less than 100 subjects**

What are we looking for in fMRI data?

- Mainstream fMRI analysis **objective**:
discovering brain areas **relevant** to a mental state or a task



- **But how to measure ‘relevance’?**

- Ultimately: **mutual information** (but computationally intractable to evaluate on all voxel subsets)
- Simplest approximation: **univariate (voxel-wise) correlations with the task** (GLM approach)
But informative multivoxel patterns are often missed (Haxby et al, many other studies, this work)
- This work: **predictive accuracy of multivariate sparse models** as a better proxy for relevance

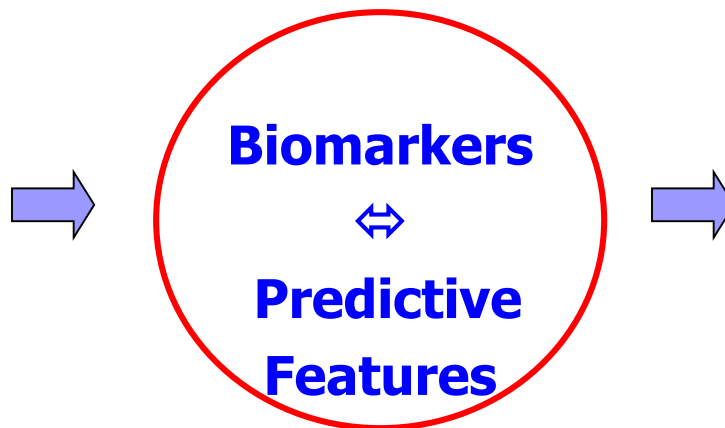
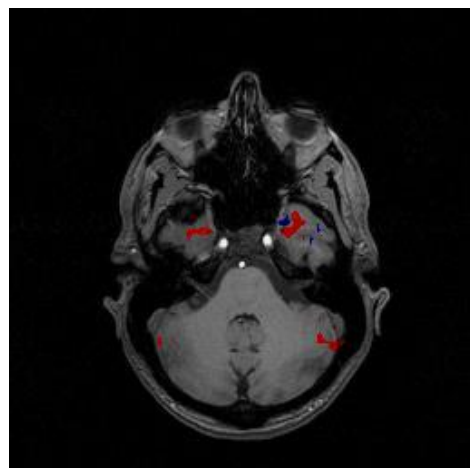
- **Questions:**

How is task-related information distributed in the brain?

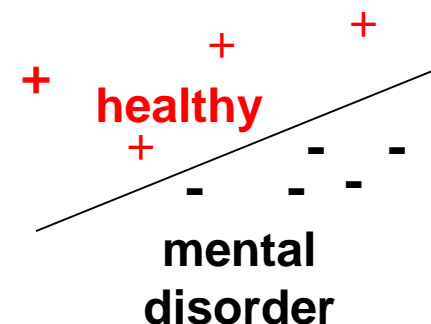
Is there a sharp boundary between relevant vs. irrelevant brain areas?

Or is the information distribution through the brain almost ‘holographic’?

Our Goal: Interpretable Multivariate Predictive Models



Predictive Model



Feature Selection

- [Carroll et al, Neuroimage 2009]
- [Rish et al, Brain Informatics 2010]
- [Rish et al, SPIE Med.Imaging 2012]

Sparse regression (LASSO, Elastic Net)

$$\hat{\beta} = \arg \min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1$$

Feature Construction:

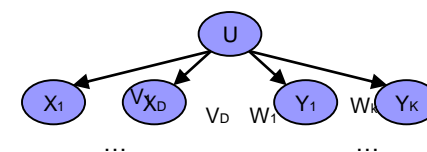
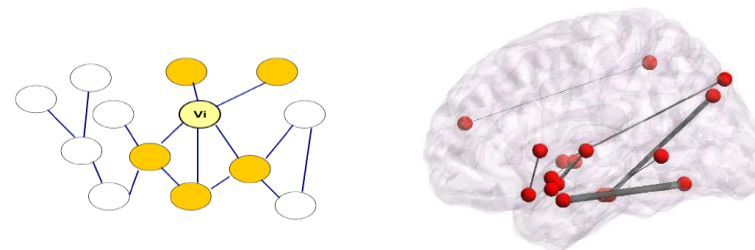
- feature engineering (network properties etc.)

- [Rish et al, PLoS One 2013, Cecchi et al, NIPS 2009]
- [Rish et al, SPIE Med.Imaging 2012]

- automated feature extraction:

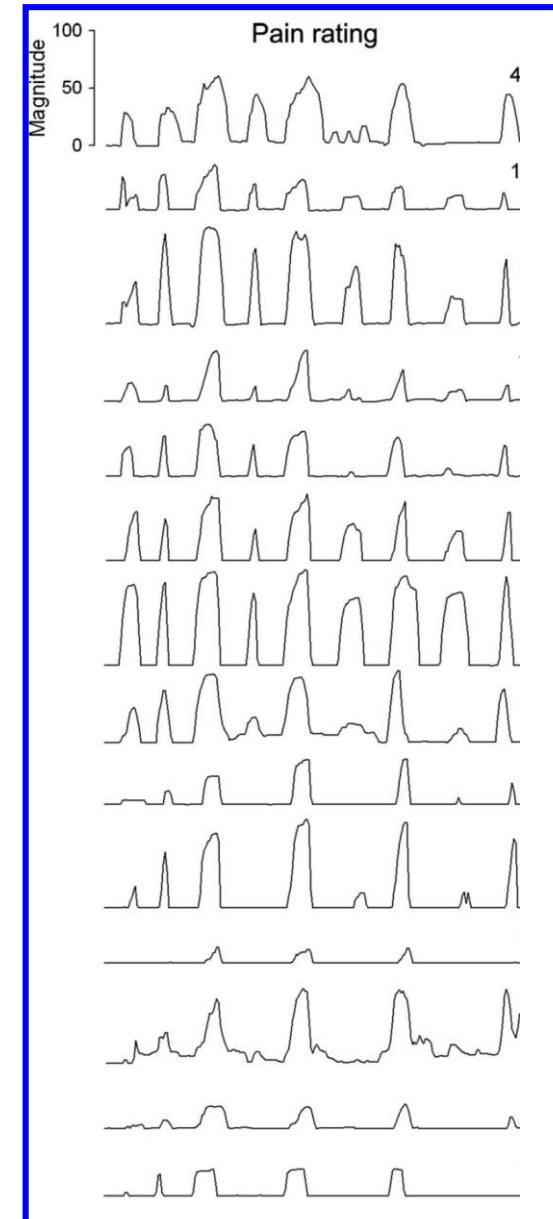
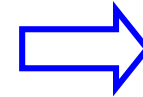
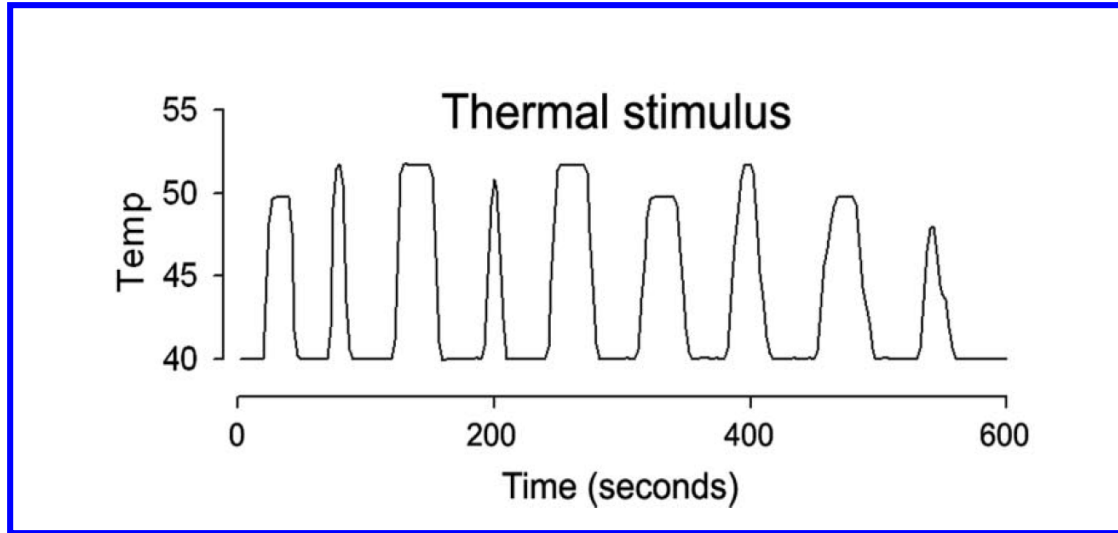
dictionary learning, deep learning, and so on

- [Rish et al, ICML 2008], ongoing work



Example: Pain Perception Studies

Data from
[Baliki, Geha, Apkarian 2008]



14 healthy subjects presented with painful thermal stimuli while in fMRI scanner, and asked to rate their pain level (using a finger-span device).

Where are pain-related brain areas?

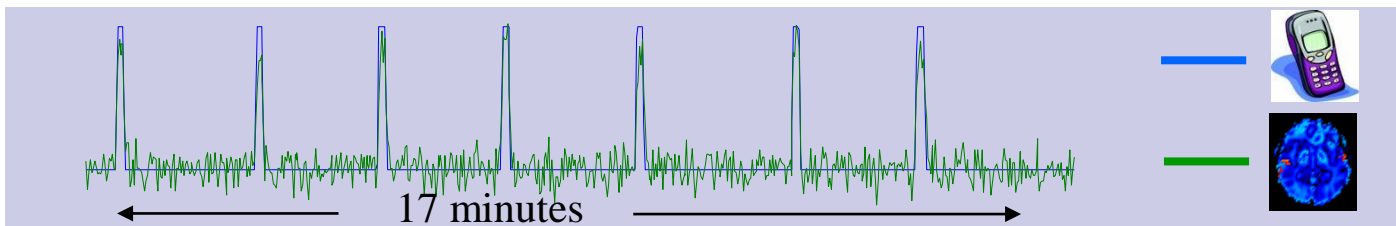
Can we predict pain perception and/or stimulus from fMRI?

Example: Videogame Playing (PBAIC07)

- subjects play a videogame in a scanner
- 24 continuous response variables
(convolved with HRF function)
 - Listening to Instructions
 - Annoyance
 - Anxiety
 - Looking at faces, etc.



Instructions variable:



Can we find brain areas involved in this task?

Can we predict response variables from fMRI?

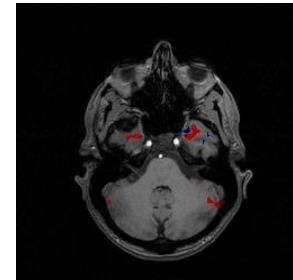
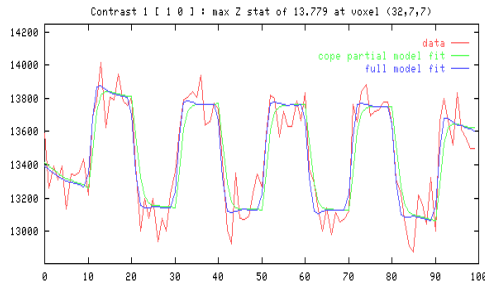
Reversing GLMs: Predicting Mental State from Voxels

$$y = X\beta + \text{noise}$$

Measurements:
mental states, behavior,
tasks or stimuli

fMRI data ("encoding")
rows – samples (~500)
Columns – voxels (~30,000)

Unknown
parameters
(“signal”)



Find a small number of the **most relevant** voxels (brain areas)
Relevant ⇔ **(jointly) predictive** about task (vs. individually correlated)

Feature Selection via Sparse Regression: LASSO and Elastic Net

ISSUE: high-dimensional, small-sample problem

- solutions are *overfit* to data: poor generalization
- difficult to *interpret* (determine relevant voxels)

APPROACH:

- **LASSO:** adds ℓ_1 -norm regularization

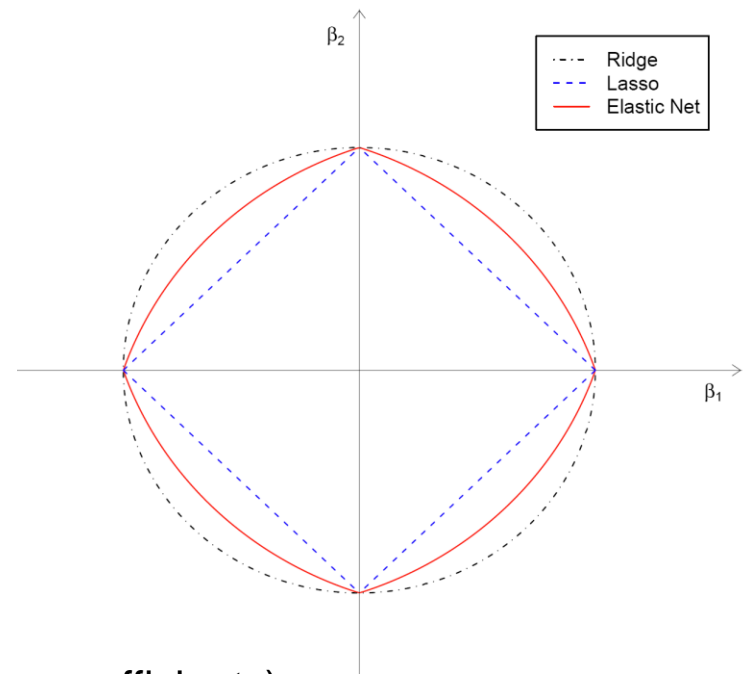
$$\min_{\beta} ||y - X\beta||_2^2 + \lambda ||\beta||_1$$

- selects relevant voxels (sparse solution \Leftrightarrow many zero coefficients)
- improving LASSO: **Elastic Net** - sparsity + grouping of correlated variables

$$\hat{\beta} = \arg \min_{\beta} ||y - X\beta||_2^2 + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||_2^2$$

• l_1 keeps singularities at vertices \Rightarrow sparsity

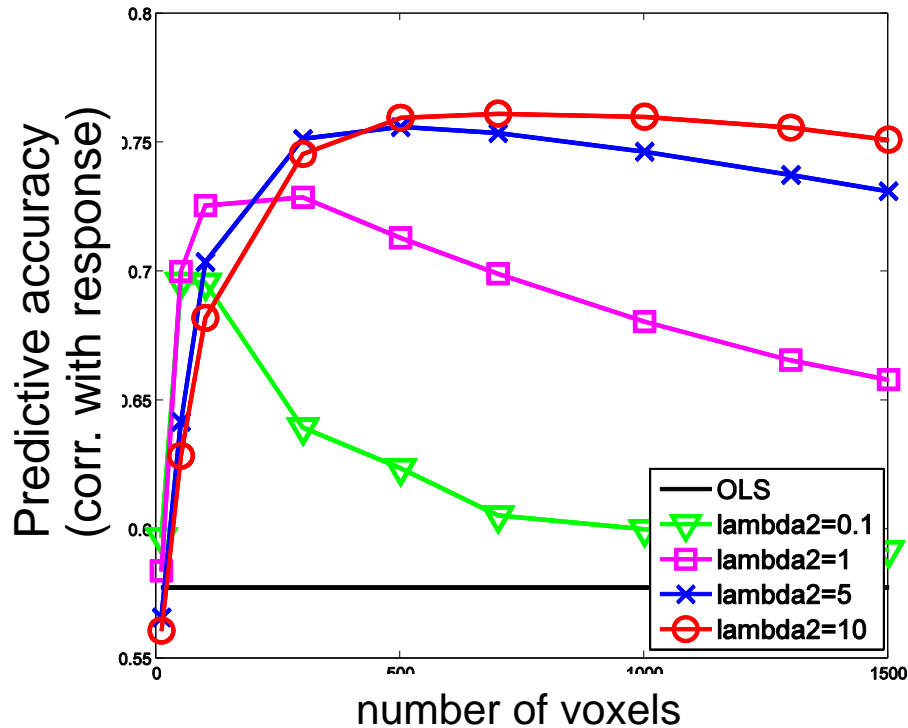
• l_2 enforces strictly convex edges \Rightarrow grouping effect



Sparse Models Can Accurately Predict Mental States

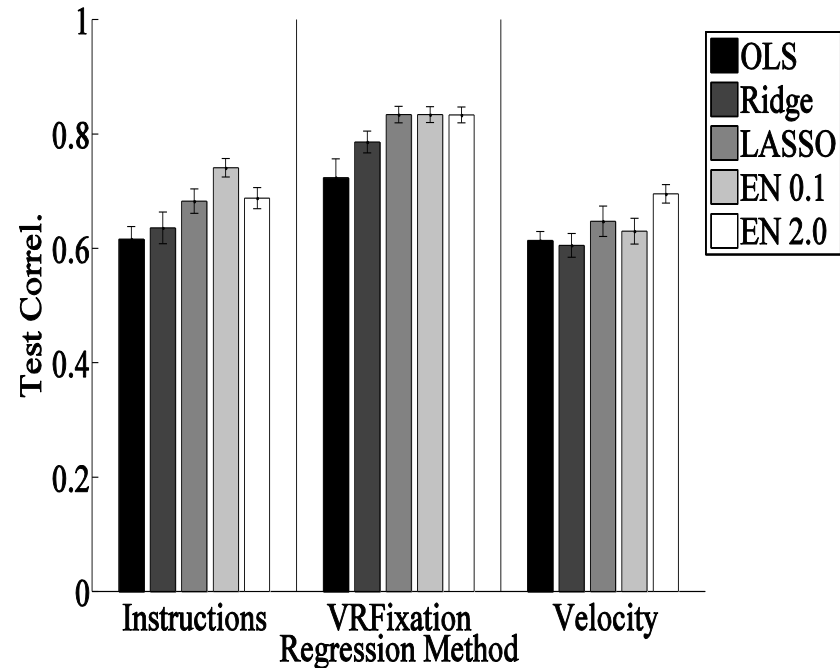
[SPIE Med.Imaging 2012], [Rish et al, BI 2010], [Carroll et al, Neuroimage 2009]

Pain Rating Task



[Rish et al. 2010]

PBAIC-07 Tasks



[Carroll et al. 2009]

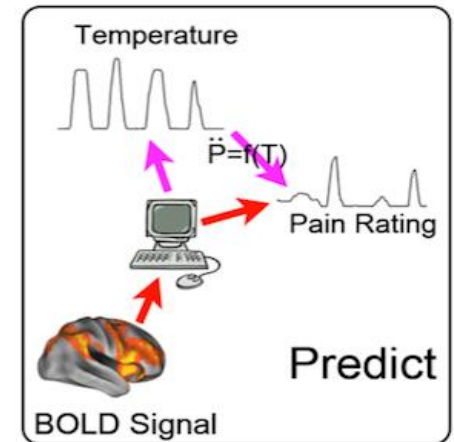
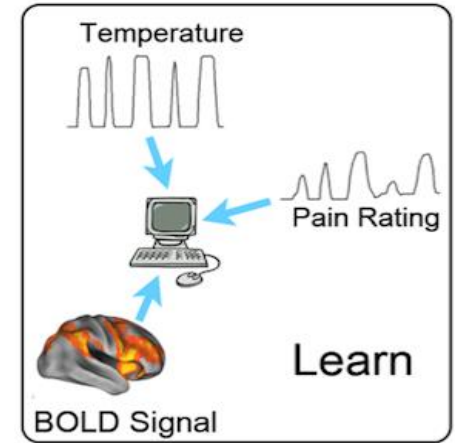
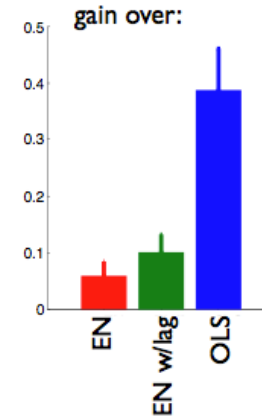
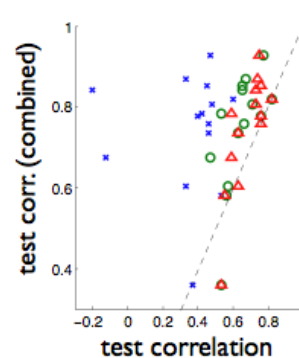
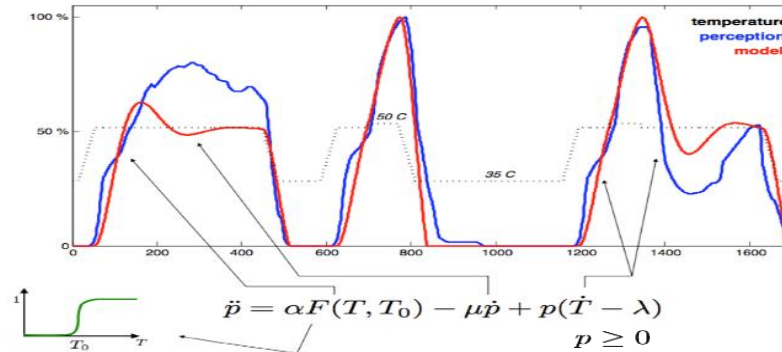
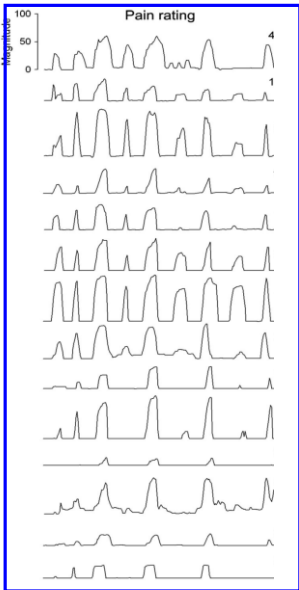
Elastic Net: close to 0.8 prediction accuracy for pain ratings [Rish et al, BI-2010], as well as for several tasks in PBAIC-07 dataset [Carroll et al, Neuroimage 2009]

Even Better Results: Combine Data-Driven and Analytical Models

- Dynamical model (1st order, only 3 parameters) captures inter-subject variability in pain response given stimulus
- Stimulus not available? Predict from fMRI, then apply the model!

[PLoS Comp Bio 2012]

Varying Pain Perception



Incorporating nonlinear dynamical model into sparse learning (via hidden stimulus variable) improves over 'direct' sparse regression – due to very high accuracy of analytical model !

But How is Predictive Information Distributed in the Brain?

- Are there **multiple** sparse models that predict well?

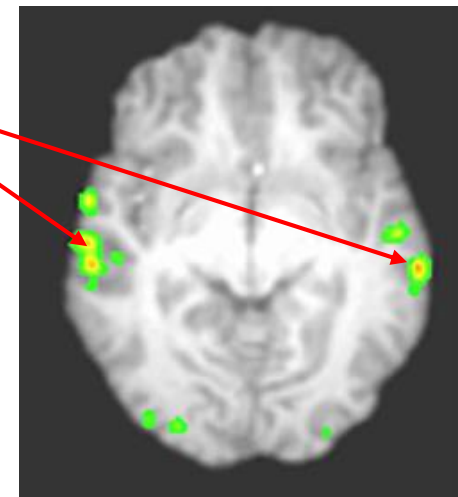
Best sparse solution for PBAIC 'Instructions' task



Voxels in a sparse solution **predict well** \Rightarrow **task-relevant**



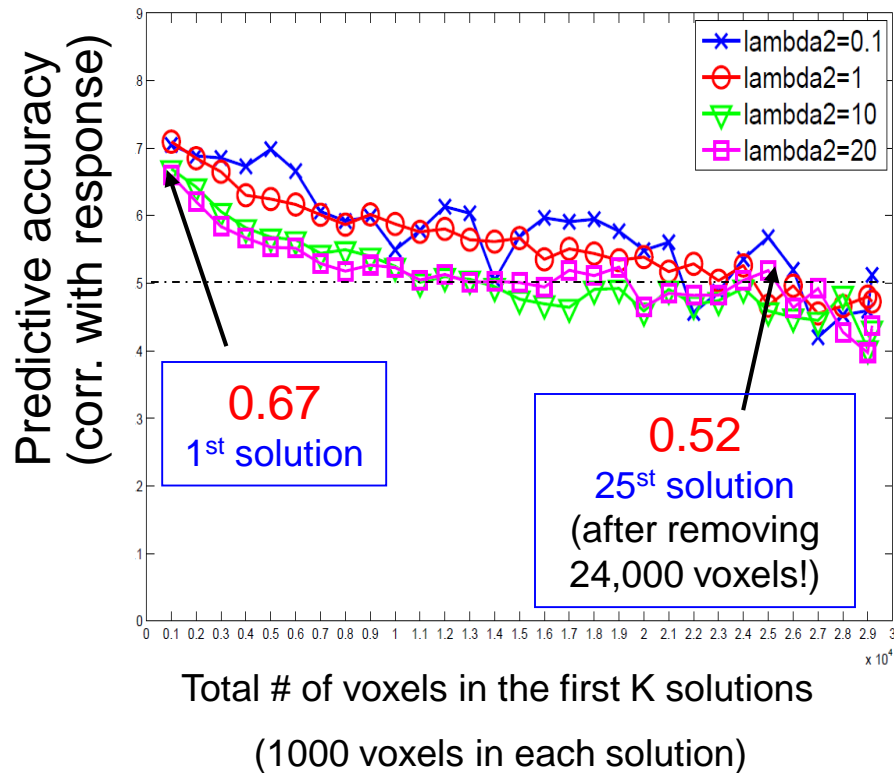
But does this mean **ONLY** these voxels are relevant?
Not necessarily – multiple good solutions may exist!



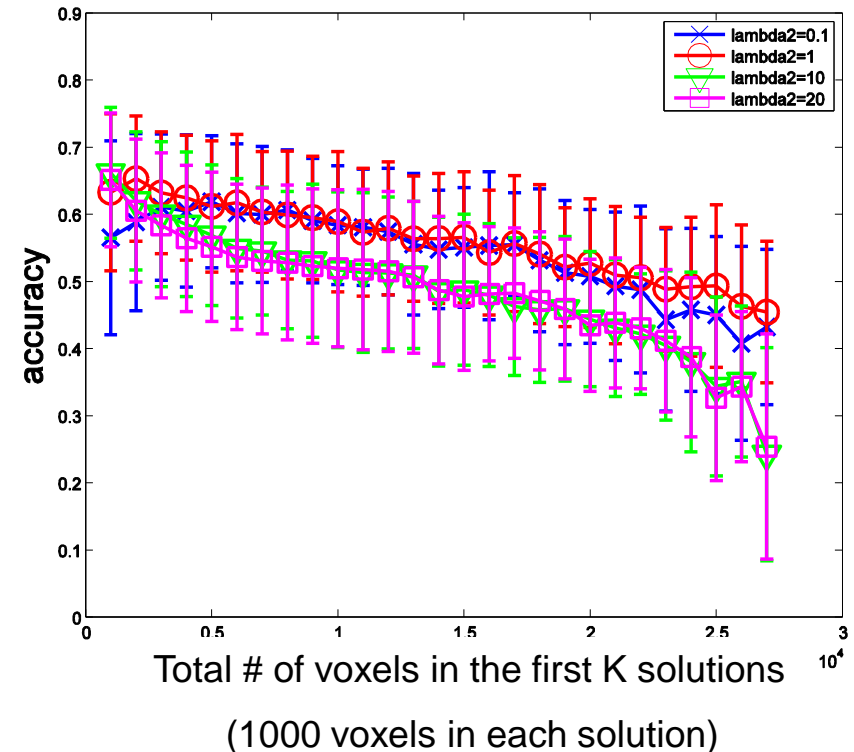
- **Whole-brain exploration of relevant areas:** find sparse solution, remove its voxels, find another one; repeat until no voxels are left:
 1. Run Elastic net to find one solution of size $k = \min(1000; \# \text{ of remaining voxels})$
 2. Remove the solution voxels from the data
 3. If no more voxels left, stop, otherwise Go to step 1
- How does the predictive power degrade as more voxels are removed?
- Is there a sharp transition between relevant and irrelevant voxels?

Full-Brain (Holographic) Information Distribution for Pain

Pain prediction: subject 6

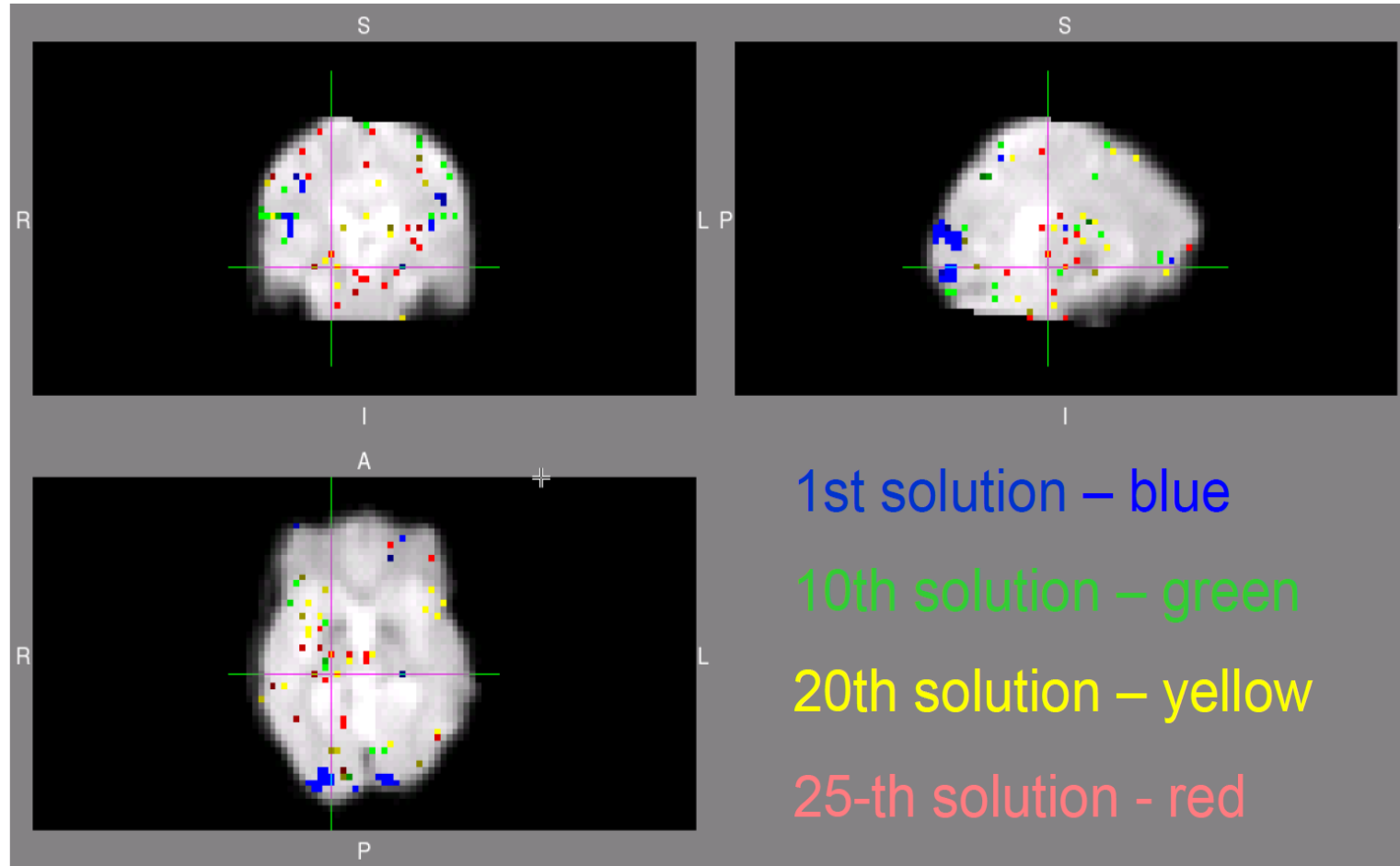


Pain prediction: avg. over 14 subjects



- Surprisingly slow degradation of predictive accuracy!
- No sharp transition between relevant and irrelevant voxels

Pain-Predictive Solutions Are Spatially Distributed Throughout the Brain

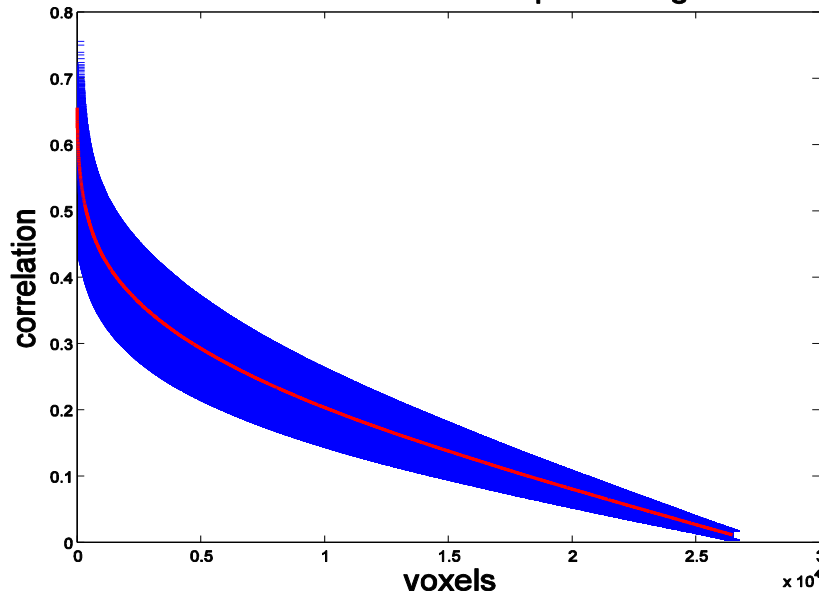


Highly predictive solutions are spread throughout the brain:

later solutions DO NOT use voxels from same predictive areas as former ones

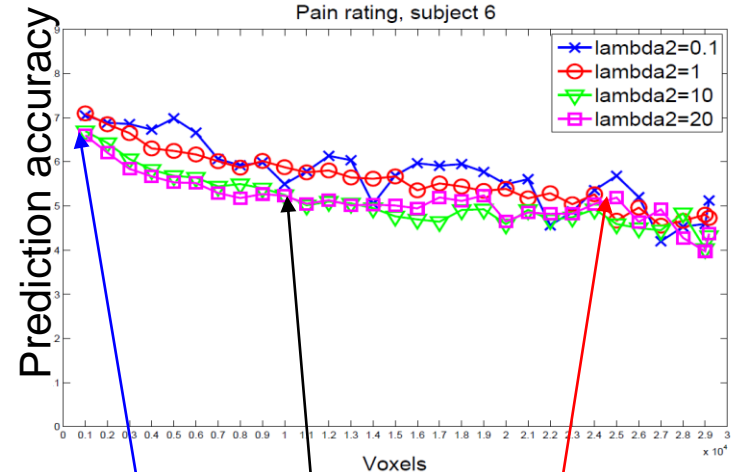
Univariate Correlations DO NOT Properly Capture Voxel Relevance!

Voxel correlations with pain ratings

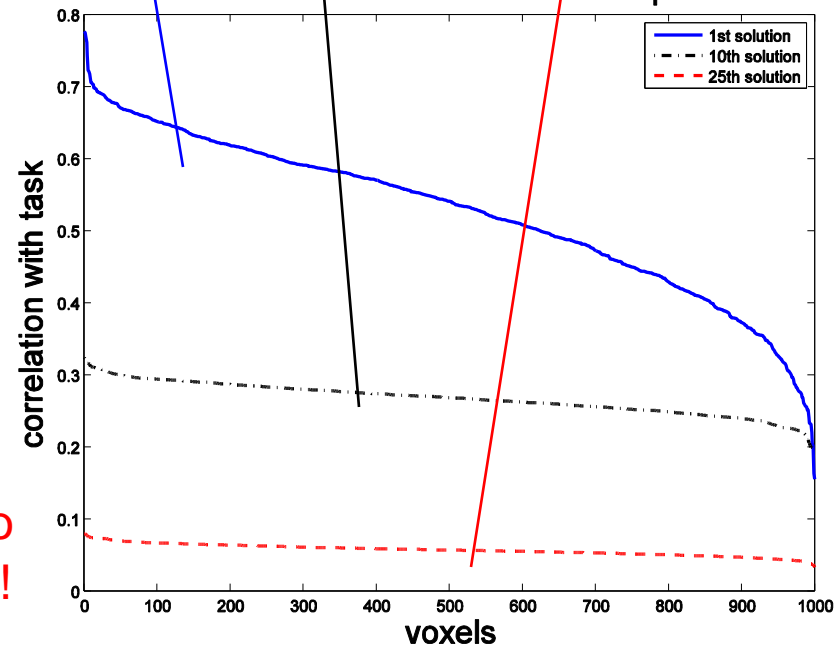


- **Exponential** decay in relevance measured by **univariate correlation** with the task vs. **linear** decay for **prediction accuracy**
- Still **highly predictive solution #25** (0.52 accuracy vs. 0.67 of the 1st solution) has **no voxels with individual correlation above 0.1!**

Pain rating, subject 6

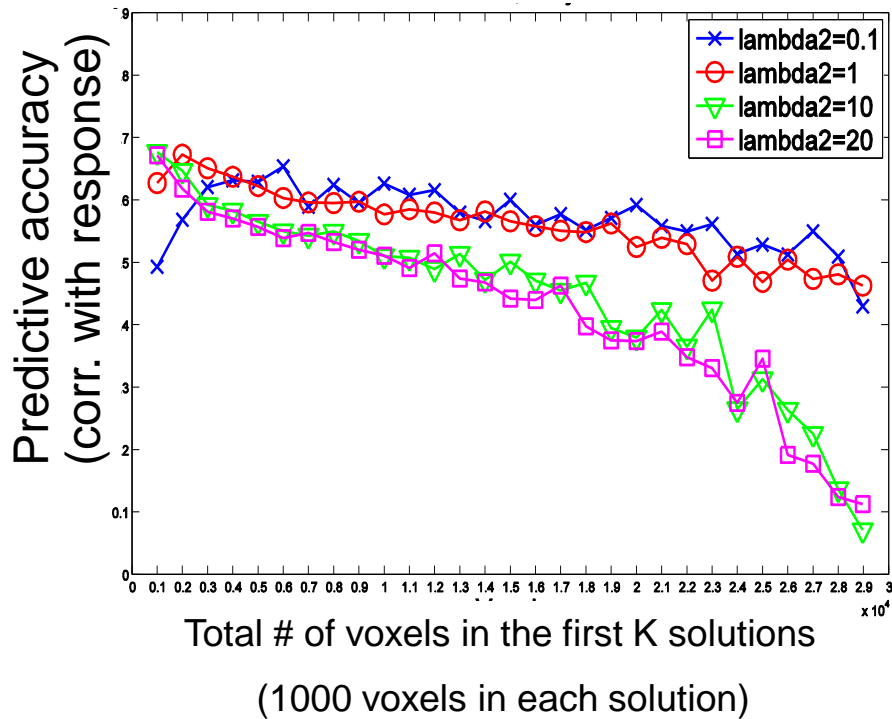


Correlation of EN-solutions with pain

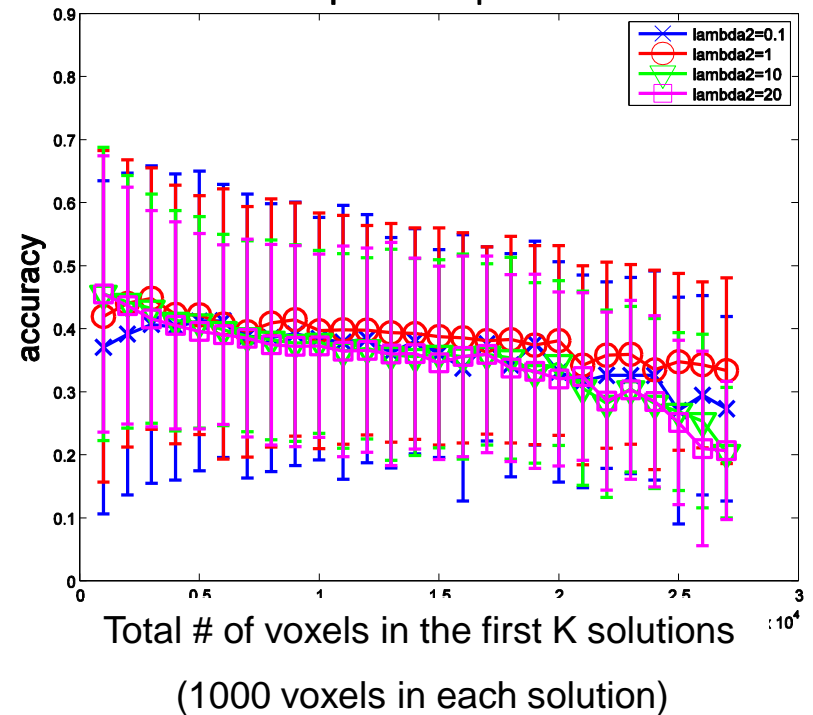


Visual Task: Similar Results to Pain

Visual task: subject 6

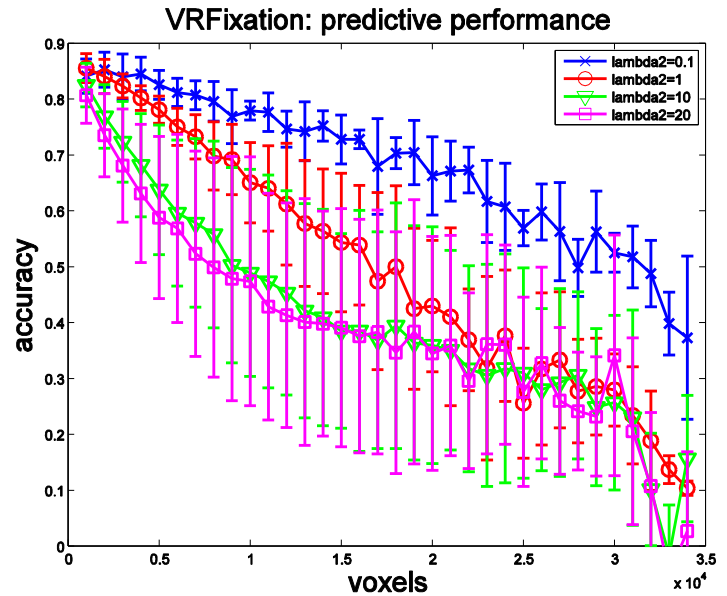
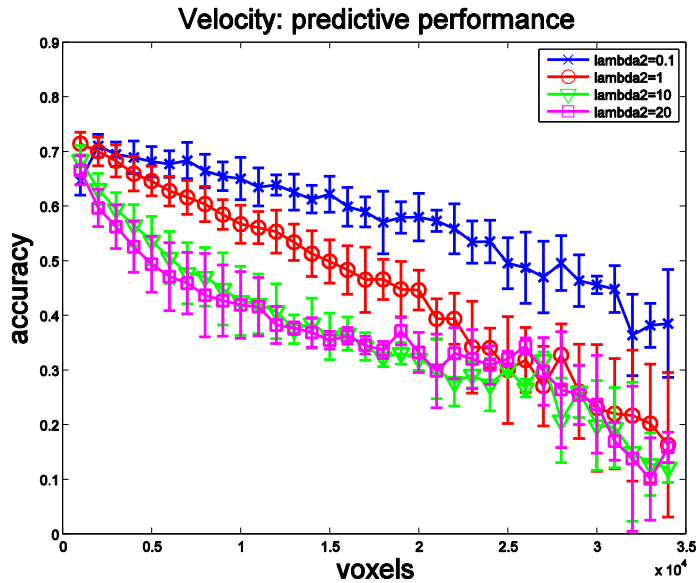


Visual task: avg. over 14 subjects

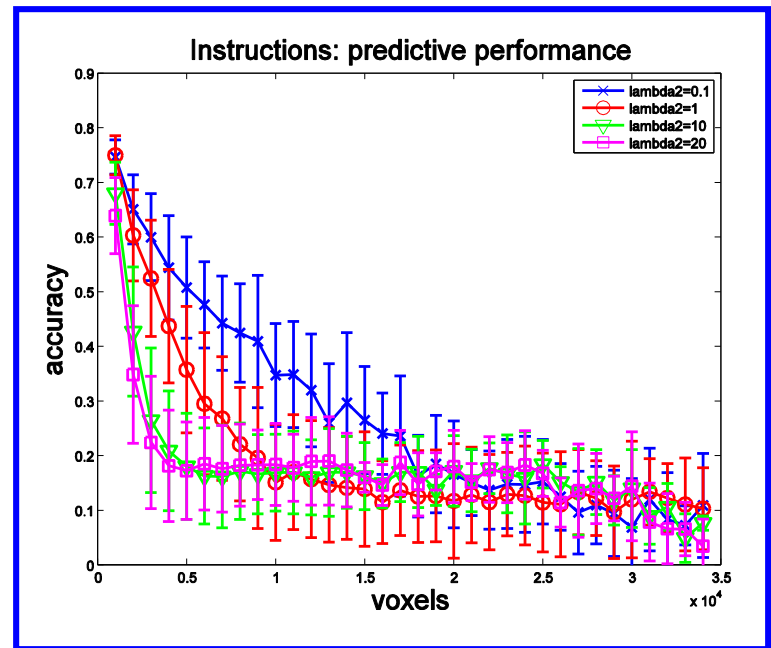


- Similar to pain: linear decay of predictive accuracy (\Leftrightarrow relevance)
- Stronger grouping (larger λ_2) \Leftrightarrow faster separation of more relevant voxels from less relevant ones

PBAIC Tasks: Results Depend on the Task!

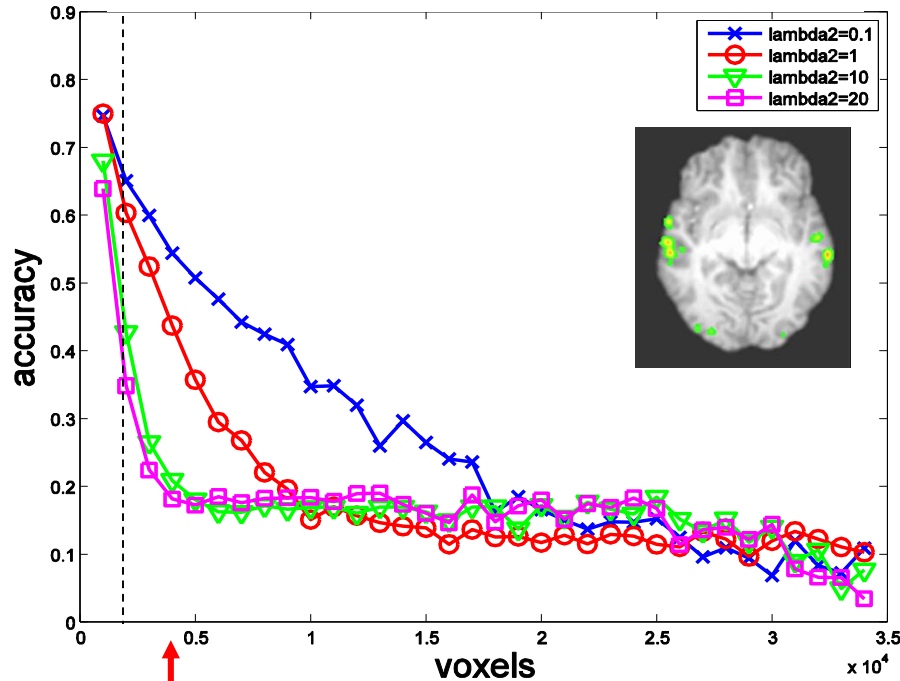


- Solution relevance (accuracy) degrades faster than for pain, though for the 2 tasks in the top panel decay is still close to linear
- 3rd task (instructions): fast (exponential) decay; relevant voxels are a small fraction of the brain (NO holographic effect)
- Decay is slower for low grouping (small λ_2), but this is likely an artifact of the method; larger $\lambda_2 \Leftrightarrow$ are more reliable results



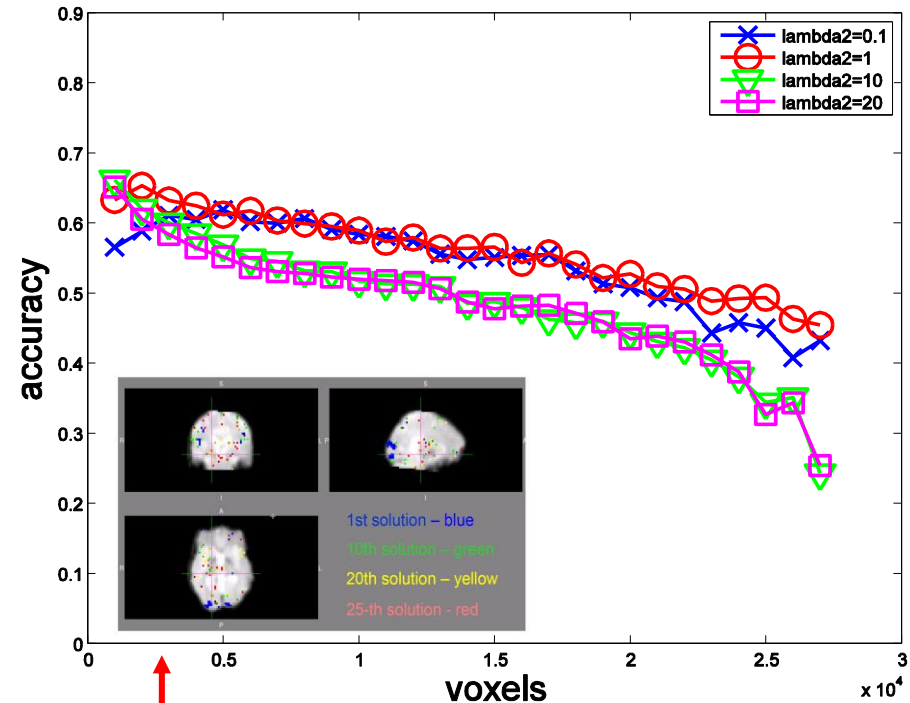
Hypothesis: Simpler Tasks More 'Localized' than Complex Ones?

Relatively simple auditory task:
listening to short instructions



Sharp transition from highly relevant first two solutions (2000 voxels), to practically irrelevant remaining voxels (0.2 and lower accuracy)

More complex task:
perceiving and rating pain



No such sharp transition, slow linear decay from best (on average) 0.65 accuracy (1st solution) to 0.5 (10th sol.) and 0.4 accuracy (24th solution, 23,000 voxels removed)

Sparse Regression and fMRI: Summary

■ Main questions:

- How is task-related information distributed in the brain?
- Is there a sharp boundary between relevant vs. irrelevant brain areas?

■ Approach:

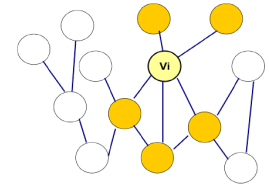
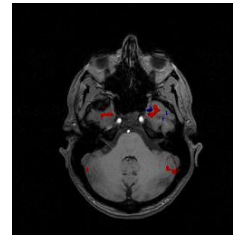
- exploring solution space of multivariate sparse regression, where sparse solutions \Leftrightarrow task-relevant voxel subsets/areas

■ Results:

- contrary to traditional univariate correlation (or GLM) approach, multivariate sparse regression reveals full-brain ('holographic') spread of task-relevant information
- tasks such as pain rating and visual rating seem to involve most of the brain rather than just specific areas (involvement measured by predictive accuracy)
- however, not all tasks are holographic (e.g., Instructions in PBAIC dataset is not)

■ Hypothesis (requires further empirical investigation):

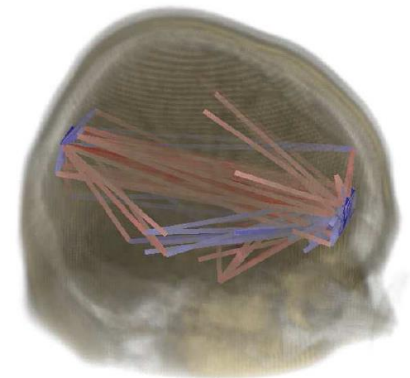
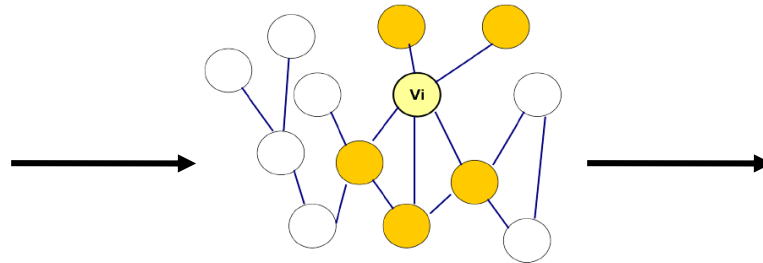
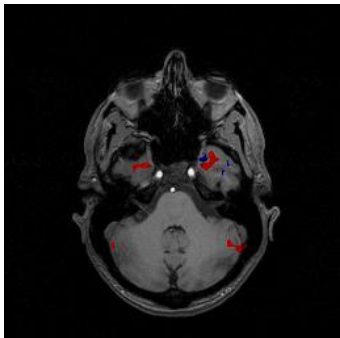
- widespread activation (measured by multivariate predictive information) is more characteristic of complex tasks such as pain perception; simpler tasks have more clear separation between relevant and irrelevant brain areas.



Part 2: Learning (Sparse) Brain Networks

Overview

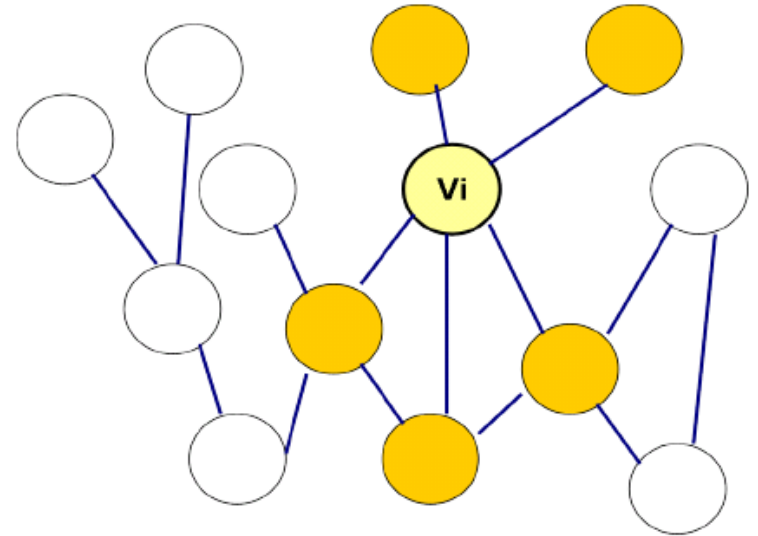
- **Our goal:** learning full-brain **interpretable** probabilistic network models
- **Problem:** full-brain networks, even edge-sparse, are hard to interpret; can we identify most relevant nodes/voxels?
- **Proposed approach:** variable (node) selection, besides the usual edge selection, using group-Lasso type of penalty
- **Application:** study of cocaine addicts vs. controls (Goldstein et al., 2007) performing a visual attention task with a monetary reward
- **Results:** significantly more interpretable and statistically more accurate networks that discover most important clusters of interacting voxels



Markov Networks (Markov Random Fields)

$$\mathbf{X} = \{X_1, \dots, X_p\}, \quad G = (V, E)$$

$$P(\mathbf{X}) = \frac{1}{Z} \prod_{C \in \text{Cliques}} \Phi_C(\mathbf{X}_C)$$

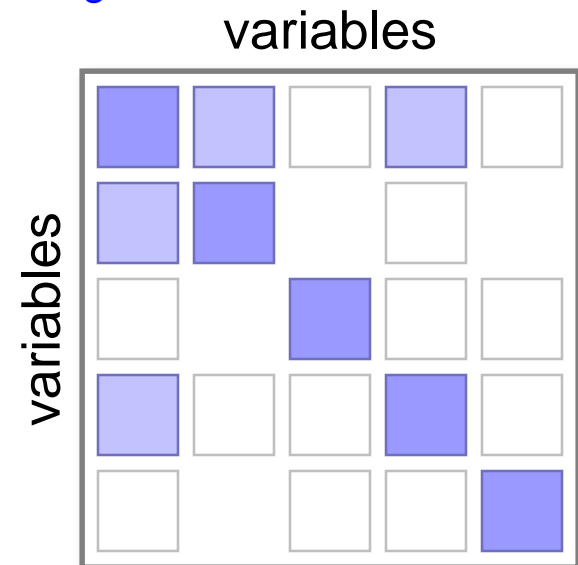
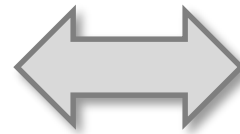
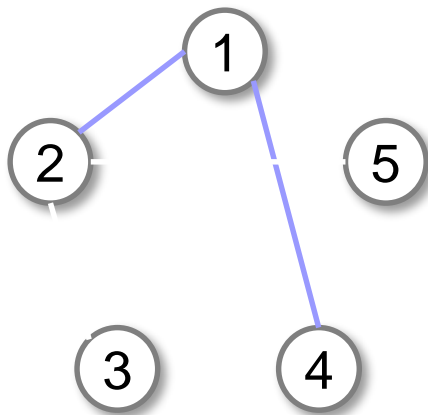


Lack of edge $(i, j) \rightarrow$
conditional independence $X_i \perp X_j | \text{rest}$

Gaussian Markov Random Fields (GMRFs)

Markov random field of jointly Gaussian variables

- $P(\mathbf{x}) = (2\pi)^{-\frac{P}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$
- Σ - covariance matrix, Σ^{-1} - precision (concentration) matrix
- Zeros in Σ : marginal independence
- Zeros in $\Sigma^{-1} \Leftrightarrow$ conditional independence \Leftrightarrow lack of edge



Sparse $\Sigma^{-1} \Leftrightarrow$ sparse Markov network

Precision matrix
(Inverse covariance)

Maximum Likelihood Estimation of the Inverse Covariance Matrix

Assume the data \mathbf{X} are centered to have zero mean. Then:

$$\hat{\Sigma}^{-1} = \arg \max_{C \succ 0} \log p(C|\mathbf{X}) = \arg \max_{C \succ 0} \log p(\mathbf{X}, C) =$$

$$= \arg \max_{C \succ 0} \log \det(C) - \text{tr}(SC)$$

where $S = \frac{1}{N} \sum_{i=1}^N x_i^T x_i$ is the empirical covariance matrix (MLE of Σ)

Why not just use $\hat{\Sigma}^{-1} = S^{-1}$?

- in small-sample case ($n < p$), S may not be even invertible
- even if it is, S^{-1} almost never contains exact zeros
- **needed: an explicit sparsity-enforcing constraint (regularization)**

l_1 -Regularized Maximum Likelihood Problem

Primal:

$$\hat{\Sigma}^{-1} = \arg \max_{C \succ 0} \log \det(C) - \text{tr}(SC) - \lambda \|C\|_1 \quad (1)$$

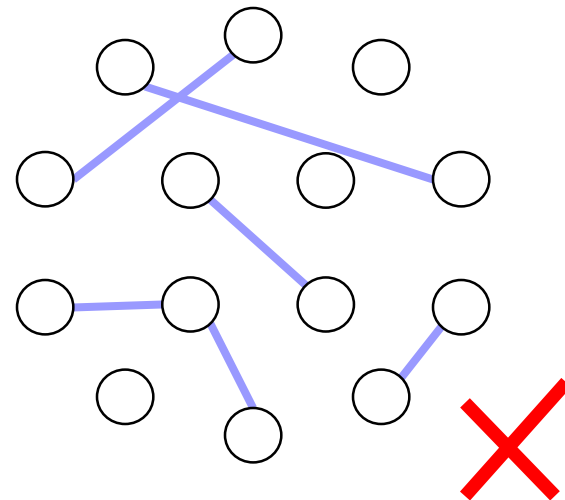
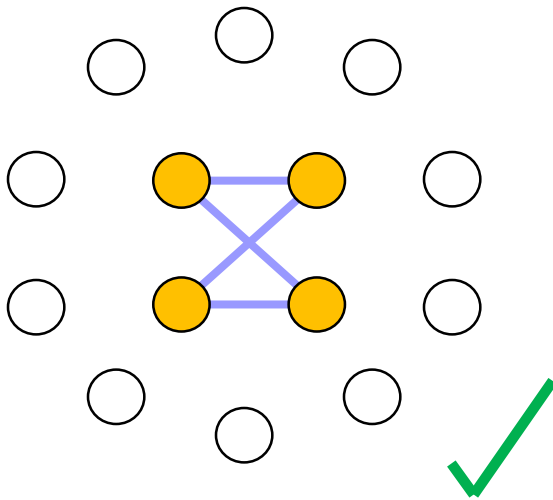
Various algorithms:

- Approximation: LASSO for each node (Meinshausen&Buhlman, 2006)
- Block-coordinate descent: COVSEL (Banerjee et al, 2006),
glasso (Friedman et al, 2007)
- Projected gradient (Duchi et al, 2008)
- Greedy ascent (Scheinberg and Rish, 2010)
- Alternating Linearization Method (Scheinberg et al, 2011)
- several more recent efficient techniques available

Various sparse structure, besides basic edge-sparsity:

- diagonal structure (Levina et al., 2008)
- block structure for known block-variable assignments (Duchi et al., 2008)
- unknown block-variable assignments (Marlin & Murphy, 2009; Marlin et al., 2009)
- spatial coherence (Honorio et al, 2009)
- common structure among multiple tasks (Honorio et al, 2010)

- Datasets with thousands of variables:
fMRI, gene expression, stock prices, world weather
- Hypothesis: often, only a relatively few variables are interacting with each other, forming network clusters; the rest are not relevant



- Goal: select these *important* nodes, and find their interaction pattern

Variable-Selection Regularizer: Block-Sparsity over Node's Neighbors

- An additional (last) term to encourage variable

A proxy for the number of connected nodes \rightarrow NP-hard learning

$$\max_{C \succ \mathbf{0}} \underbrace{\log \det C - \text{tr}(SC)}_{\text{log-likelihood of the dataset}} - \underbrace{\lambda \|C\|_1}_{\text{sparseness prior}} - \underbrace{\tau \|C\|_{1,p}}_{\text{our variable-selection prior}}$$

- Variable-selection prior: block l1/lp norm, for $p \in \{2, \infty\}$

$$\|C\|_{1,p} = \sum \|c_{n,1}, \dots, c_{n,n-1}, c_{n,n+1}, \dots, c_{n,N}\|_p$$

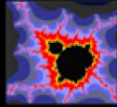


- We use Block-Coordinate Descent (BCD) on the primal (not dual!): a sequence of quadratic subproblems with closed form solutions, see (Honorio et al, AISTATS 2012)

Cocaine Addiction fMRI Data

- fMRI dataset previously collected by (Goldstein et al, 2007)
 - 15 cocaine addicted subjects and 11 control subjects
 - 87 scans/TRs (3.5 s), 53x63x46 voxels
- Subsampling to reduce dimensionality: 4x4x4 voxel cubes \Leftrightarrow 869 nodes
- Task: visual attention, with monetary reward

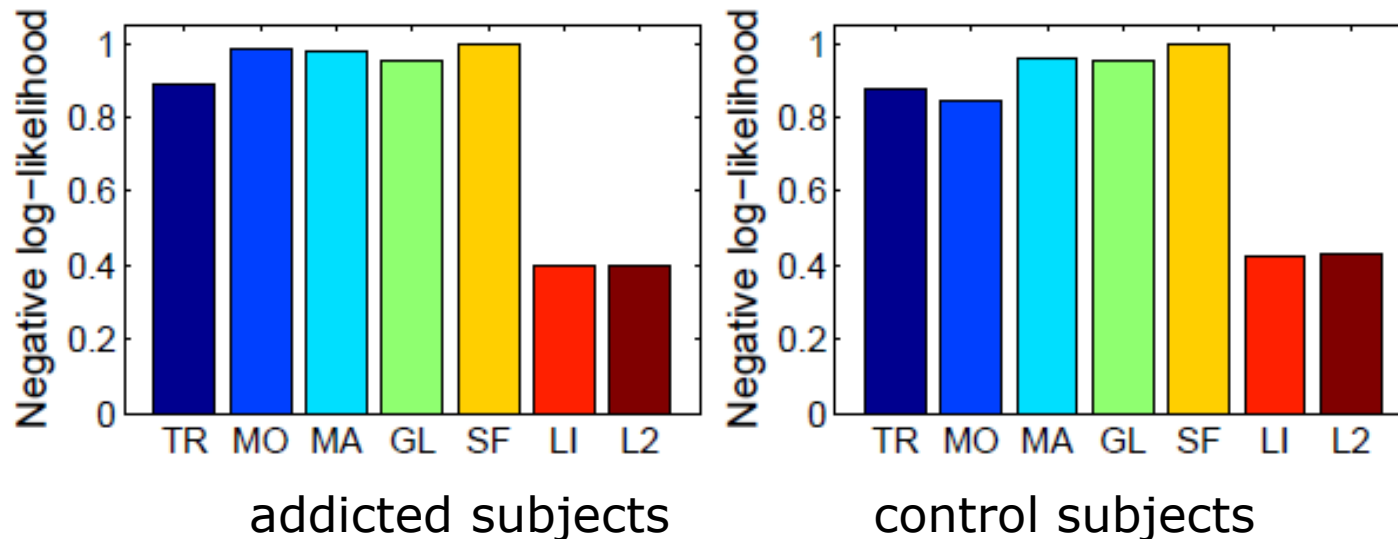
Instructions were to press a response button (using the thumb of the dominant hand) with speed and accuracy upon seeing the target (red square) after a “Go” but not after a “No-go” instruction stimulus.

B. Each 3.5-sec trial (81 go and 81 no-go trials in each block):

Fixation 1000 ms	Instruction 500 ms	Fixation 1000 ms	Trigger/Response 500 ms	Feedback 500 ms
+	 or  “Go” “No-go”	+		\$0.00 \$0.01 or \$0.45

Results: Better Model Fit (Better Likelihood)

- 1/3 of the data was used for training, 1/3 for validation and 1/3 for testing
- Results comparing negative log-likelihood (lower \Leftrightarrow better) of networks learned using various methods:



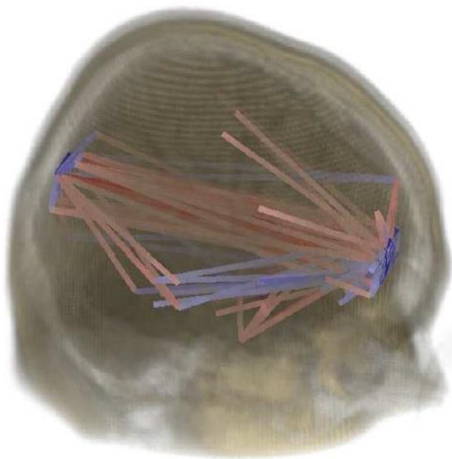
Our methods (LI,L2) outperform competitors, e.g. Meinshausen-Buhlmann (MO,MA), graphical lasso (GL), scale-free networks (SF) and Tikhonov regularization (TR).

Variable-selection assumption seem to fit the data better than standard sparse GMRFs.

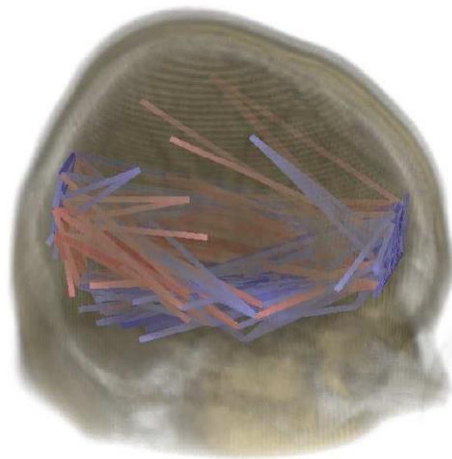
(Much) Better Interpretability

our $\ell_{1,2}$ method

cocaine subjects



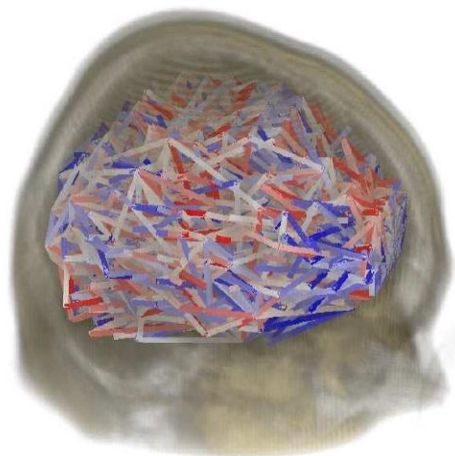
control subjects



Blue - positive interactions
red - negative interactions

Our structures involve fewer connected variables (\sim **50 *connected nodes***) and have higher log-likelihood than graphical lasso).

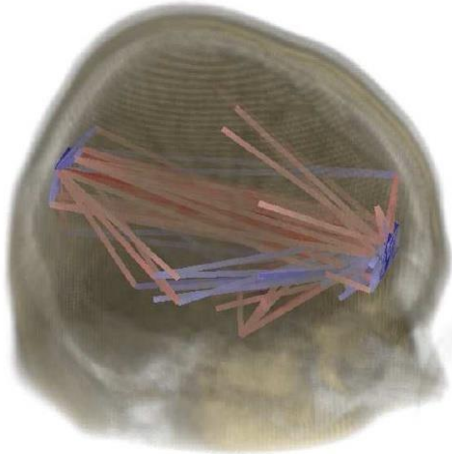
graphical lasso



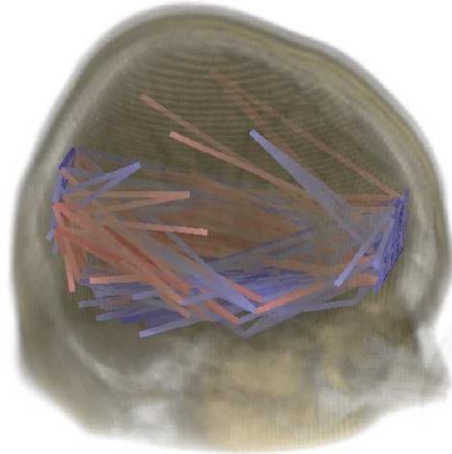
When performing classification of cocaine vs. control by using GMRFs, all methods obtain **84.6%** leave-one-subject-out accuracy

Discussion

cocaine subjects



control subjects



Blue - positive interactions
red - negative interactions

In cocaine addicts, as compared to controls, we observe

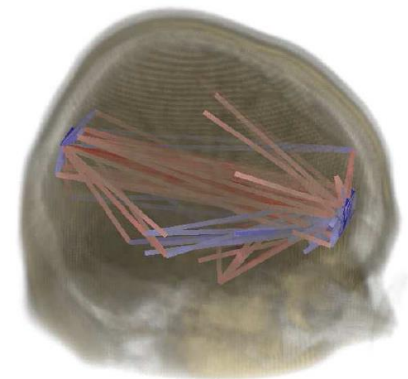
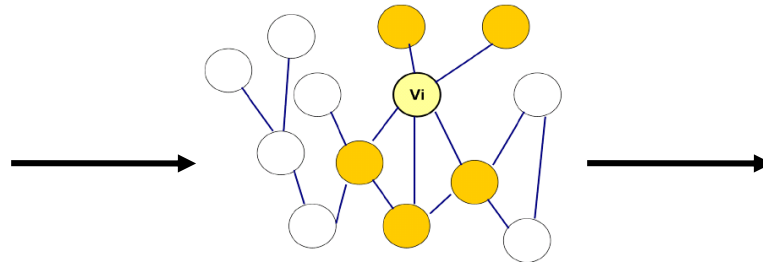
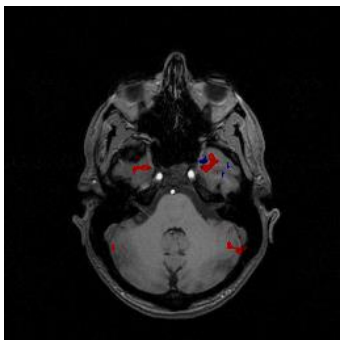
- increased interactions between the visual cortex (left) and the prefrontal cortex (right)
- decreased density of interactions between the visual cortex with other brain areas

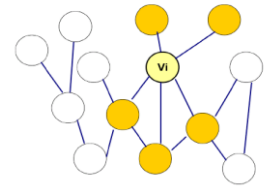
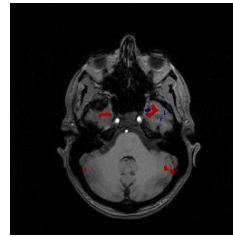
Note that the trigger for reward was a visual stimulus and that abnormalities in the visual cortex was reported in (Lee et al, 2003) when comparing cocaine abusers to control subjects

Also, prefrontal cortex is involved in decision making and reward processing, and abnormal monetary processing in the prefrontal cortex was reported in (Goldstein et al, 2009) when comparing cocaine addicted individuals to controls.

Conclusions

- We introduced **variable-selection** into sparse Gaussian MRF learning
- Our models **fit data better** than competitors without variable selection
- Most importantly, our method produces much more interpretable networks
- **Application:** study of cocaine addicts vs. controls (Goldstein et al., 2007) performing a visual attention task with a monetary reward
- **Results:** significantly more interpretable and statistically more accurate networks that discover most important clusters of interacting voxels





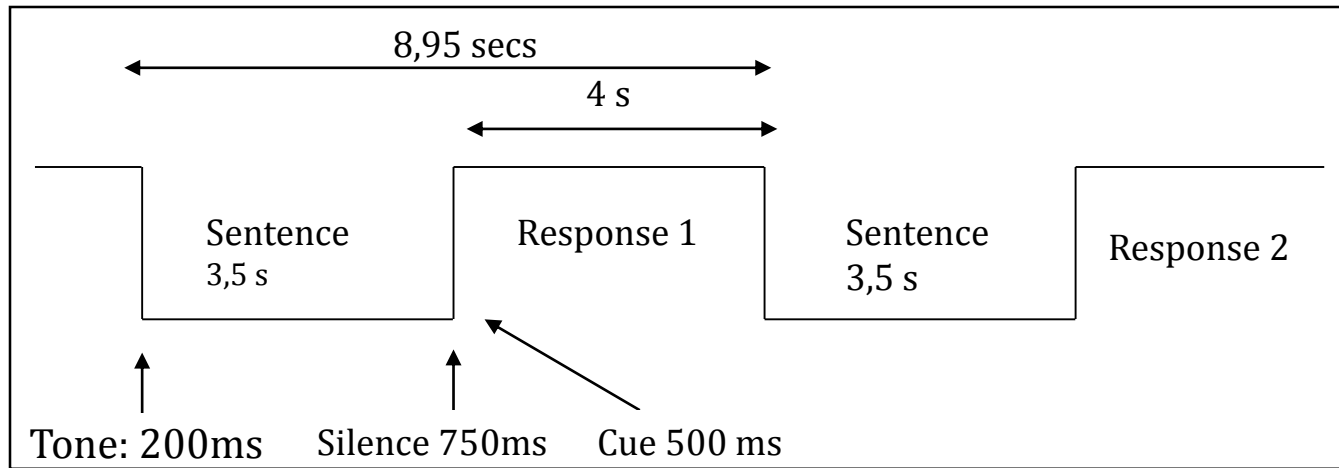
Part 3: Importance of Good Feature Construction

Schizophrenia Study

- Not a localized dysfunction, spatially or mechanistically (e.g., unlike depression, epilepsy, stroke, Parkinson's)
- Hypothesized to be a disconnection syndrome [Wernicke 1906; Bleuler, 1911; Friston & Frith, 1995]

Our objective: discover schizophrenia 'biomarkers', i.e. brain activity patterns associated with this disorder

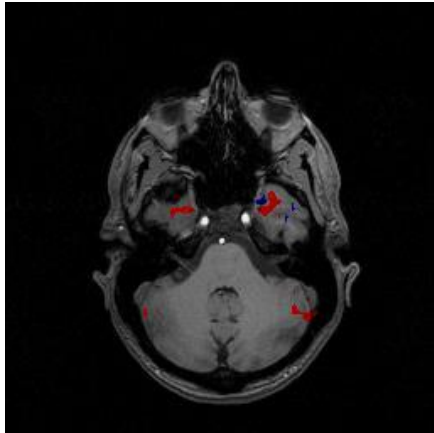
Experiment: Simple Auditory Task in fMRI Scanner*



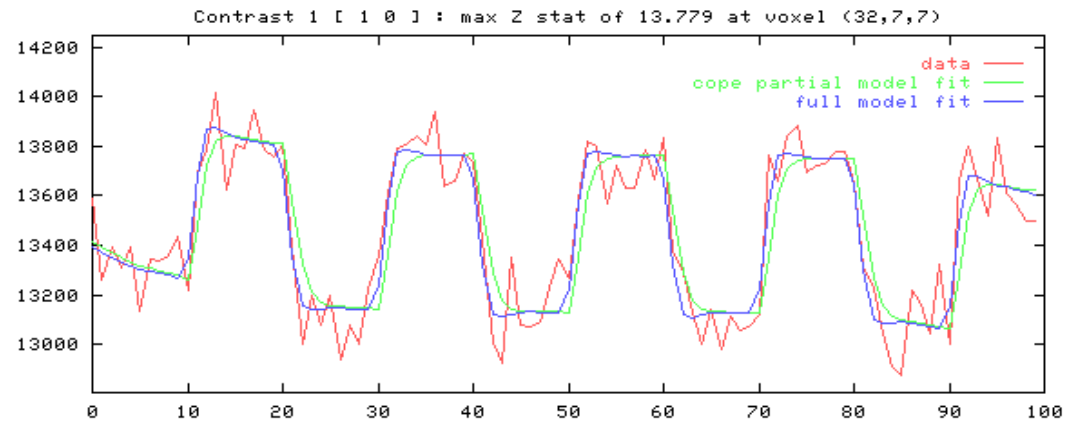
96 trials, with 32 sentences in French (native), 32 sentences in foreign languages, and 32 silence interval controls. Two runs.

- Patient Group (11 subjects)
 - Prone to auditory hallucinations
 - Native French speakers, right-handed, 3+ yrs. illness
- Normal Group (11 subjects)

Standard Approach: Univariate, Task-Related Activations



fMRI activation image and time-course
courtesy of Steve Smith, FMRI



- For each voxel, compute a score (e.g., correlation, or GLM coefficient) reflecting how well its activity matches the stimulus sequence
- Threshold the scores to select only statistically significant ones

However, no statistically significant differences were found across groups; also, classification based on activation features was close to chance level*

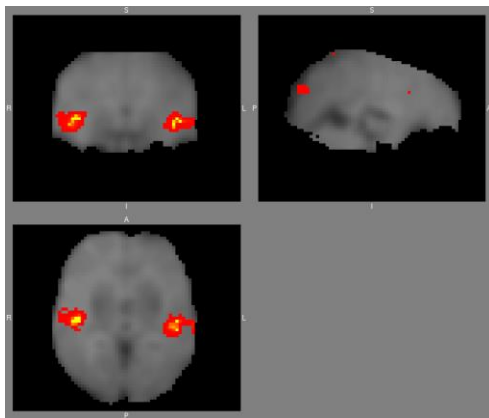
*G. Cecchi, et al., *Neural Information Processing Systems (NIPS-2009)*

Network Features Greatly Outperform Task-Activation Features

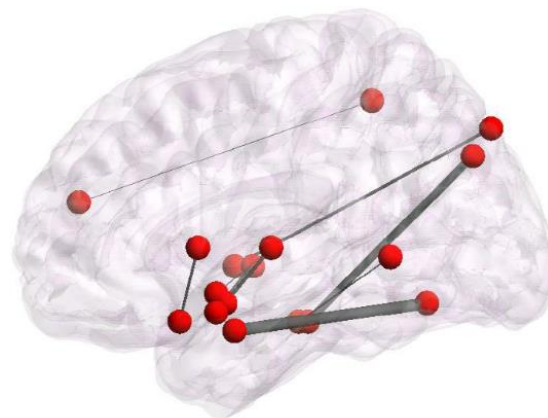
[NIPS 2009] [PLoS ONE 2013]

Functional networks: (thresholded) voxel-level correlation matrices

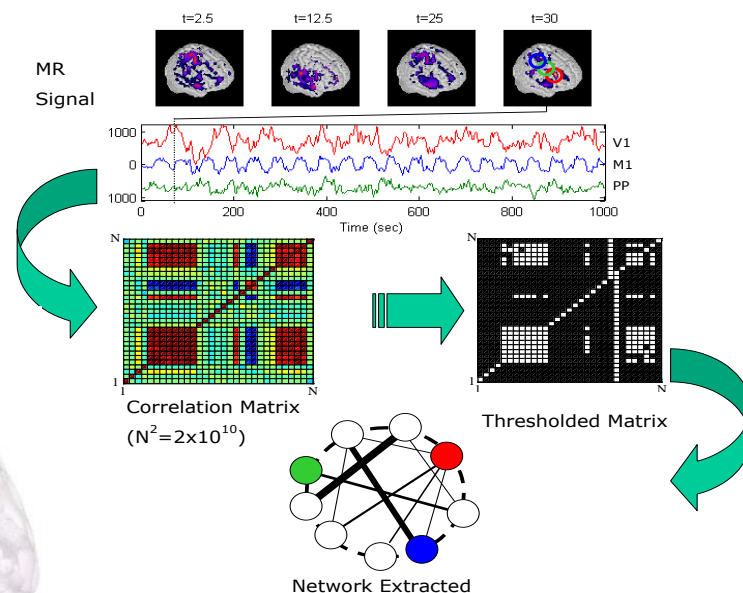
Explore functional network features vs local activations:



Voxel degrees +
GMRF = 86% accuracy



Top cross-voxel correlations
+ SVM = 93% accuracy



No matter which classifier we used, network features outperformed local activations, thus serving as much better biomarkers.

Best results: specific combination of a degree feature + classifier

Schizophrenia and Networks: Summary

- Functional networks contain large amount of schizophrenia-related information that may not be present in task-related activations
- Network properties, as opposed to activations, allow for impressively high prediction accuracy (up to 93%) given a simple auditory task
- Simplest features (a dozen of top-ranked pairwise correlations) are most predictive among all network features we tried so far



Beyond fMRI:

'Mind-Reading' from Cheaper Sensors?



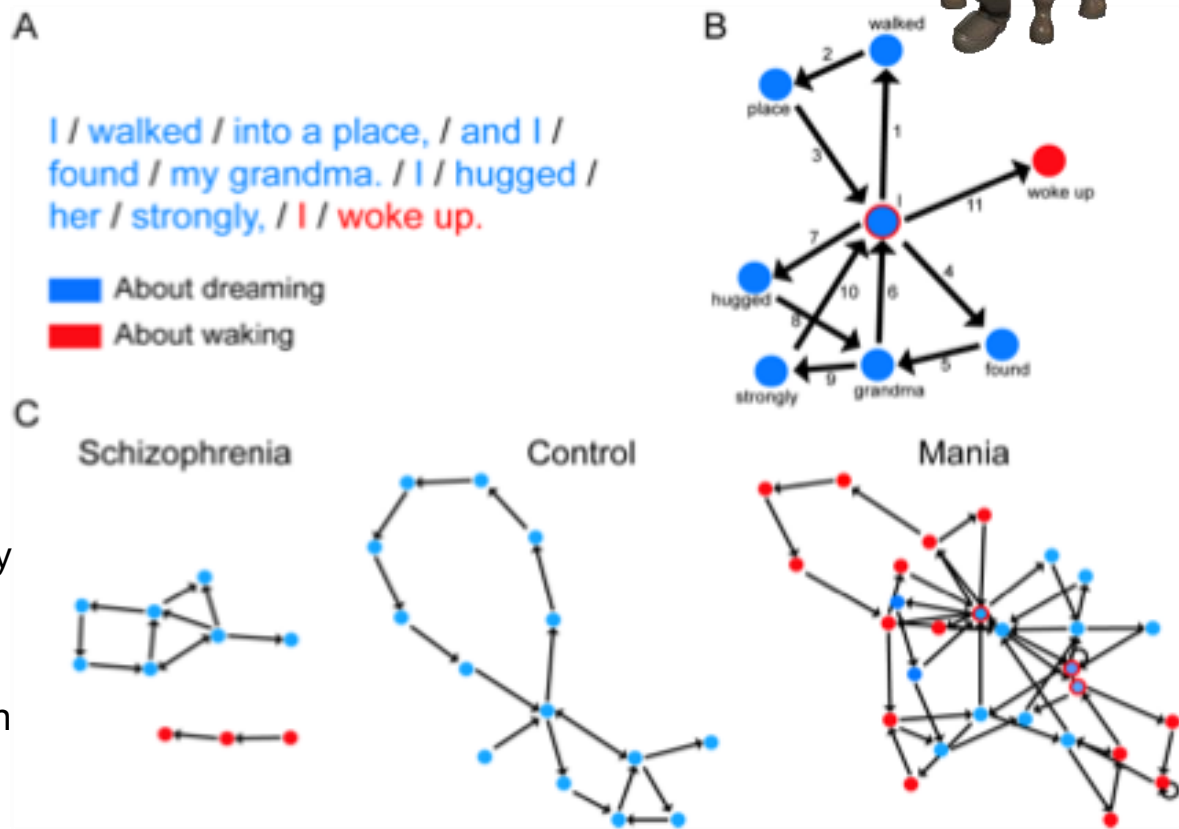


Text Analytics for “Computational Psychiatry”

“Language is a window into the brain” - M. Covington



- **93% accuracy** discriminating schizophrenics from manics based on syntactic speech graphs [PLoS One, 2012]
- nearly **100% accuracy** predicting 1st psychotic episode ONE YEAR in advance (!!) via coherence and a few other features (ongoing work)
- **88% accuracy** discriminating ecstasy and meth users from controls, using semantic features such as proximity to ‘empathy’ concept, etc., and graph features [Neuropsychopharmacology, 2014]





Current Work: Speech Coherence



Thoughts On Being Yourself

JennaMarbles 2,687,561
Subscribe 13,530,909

https://www.youtube.com/watch?v=6xx_pwu7n-Y



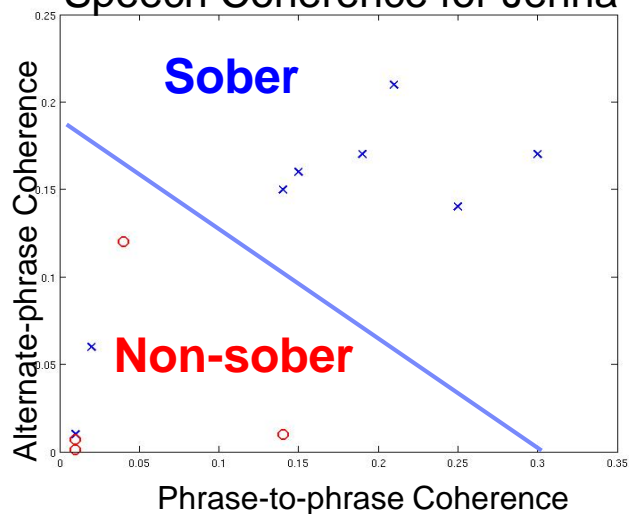
Drunk Makeup Tutorial

JennaMarbles 18,752,753
Subscribe 13,534,842

<https://www.youtube.com/watch?v=MXzwAXzUwwE>

Sober vs. Non-sober

Speech Coherence for Jenna



Text coherence:

Currently measured as the angle between vector representations of consecutive sentences (word vectors computed by LSA)

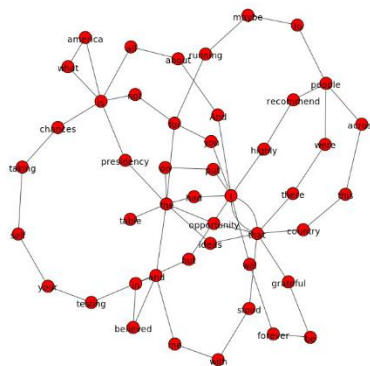


Current Work: Speech Coherence



<https://www.youtube.com/watch?v=e2h-DgYcCtw>

<https://www.youtube.com/watch?v=21z30aNO3cA>



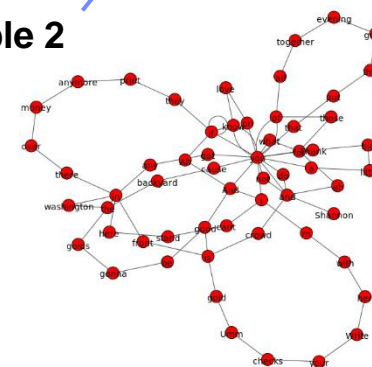
FEATURE

Lemmatized-Nodes	352
Lemmatized-Edges	985
Lemmatized-Loop Len 1	8
Lemmatized-Loop Len 2	44
Lemmatized-Loop Len 3	284
Lemmatized-Loop Len 4	1712

Example 1

Example 2

Lemmatized-Nodes	350
Lemmatized-Edges	1073
Lemmatized-Loop Len 1	4
Lemmatized-Loop Len 2	64
Lemmatized-Loop Len 3	442
Lemmatized-Loop Len 4	2956



Coherence:

Phrase to phrase-median	0.2201
Alternate phrases-median	0.2895

Phrase to phrase-median	0.0049
Alternate phrases-median	0.0045



Goal: Mental State Detection to Improve Mental Function

theguardian

News Sport Comment Culture Business Money Life & style

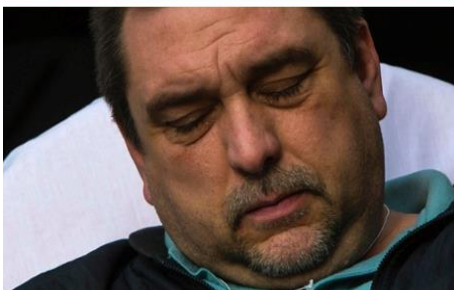
News World news New York

Train driver in deadly New York crash had 'severe' sleep disorder, NTSB says

- Report found William Rockefeller had 'obstructive sleep apnea'
- Commuter train derailment in December 2013 left four dead

Associated Press in White Plains, New York
theguardian.com, Monday 7 April 2014 20:39 BST

Jump to comments (29)



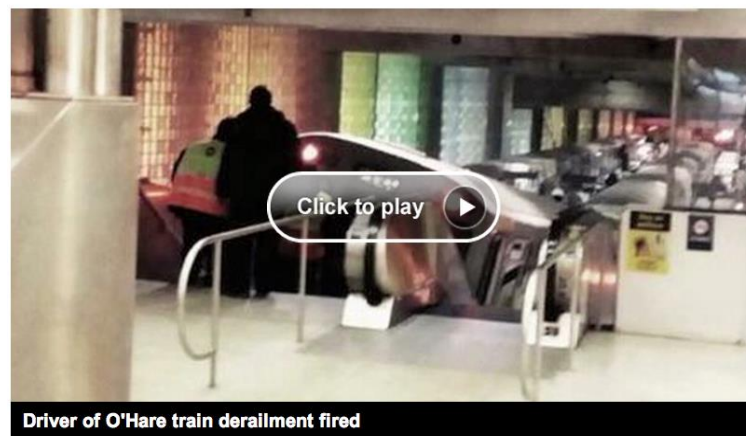
Metro-North engineer William Rockefeller Jr is loaded into an ambulance after a derailment in the Bronx borough of New York 1 December 2013. Photograph: Eric Thayer/Reuters

The driver of a New York commuter train that derailed at high speed last year, killing four people, had a serious sleep disorder that interrupted his rest dozens of times each night, federal investigators disclosed on Monday.



Operator of train that jumped tracks at Chicago's O'Hare Airport fired

By Suzanne Presto and Greg Botelho, CNN
updated 9:49 AM EDT, Sat April 5, 2014



Driver of O'Hare train derailment fired

STORY HIGHLIGHTS

• **NEW:** Official: No reason, based on schedule, that fatigue should have been a factor

(CNN) -- The driver of a train that jumped the tracks last month at Chicago's O'Hare International Airport -- after having reportedly "dozed off" -- has been fired, a transit authority spokeswoman said Friday.

Can we avoid such tragic accidents by monitoring driver's mental state and performing preemptive actions in real-time?

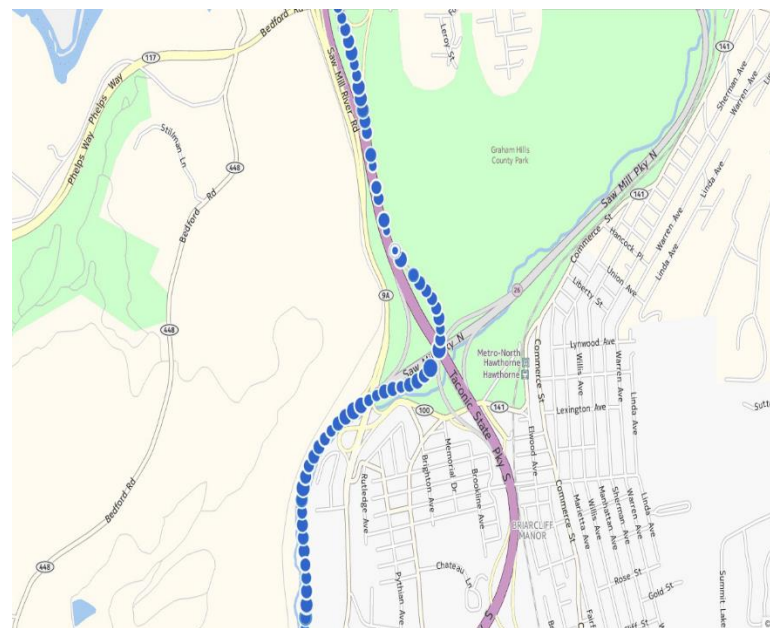
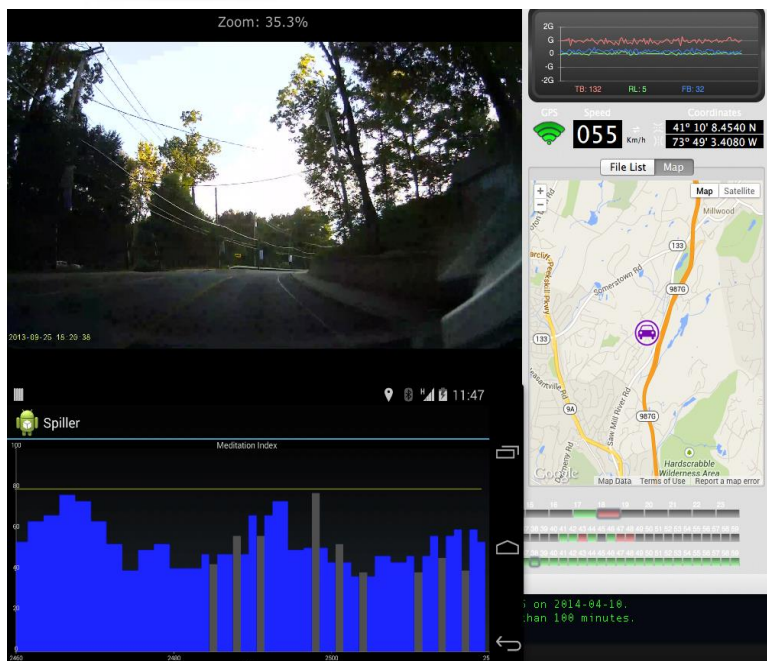


[Heisig et al, 2014]

Ongoing Work: Driver Cognitive Load from EEG



EEG: Raw waveform is FFTed to power in frequency bands (e.g., from NeuroSky or Muse device)



Example: Geographic EEG Plot of relaxation index. Merging onto a highway requires extra concentration. Sensitive software would not interrupt the driver prior to and during transit of this area.



Preliminary Results with Sparse Regression

- We used **sparse regression (LASSO/Elastic Net)** to predict Attention and Relaxation from 'raw' frequency band EEG data
- Accuracy measure: correlation between true and predicted signals

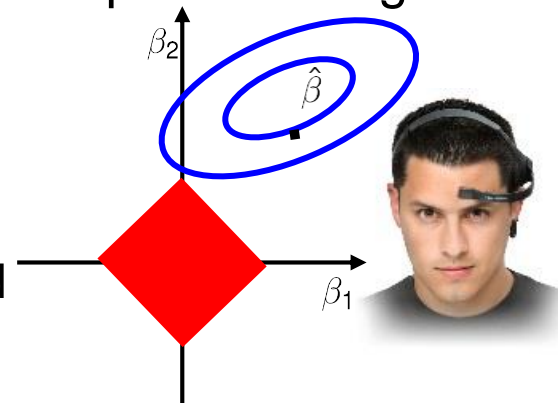
- Encouraging results:

Attention: 0.91 correlation

best model uses smoothed normalized Theta band

Relaxation: 0.87 correlation

best model uses smoothed normalized hiAlpha, Theta, MidGamma

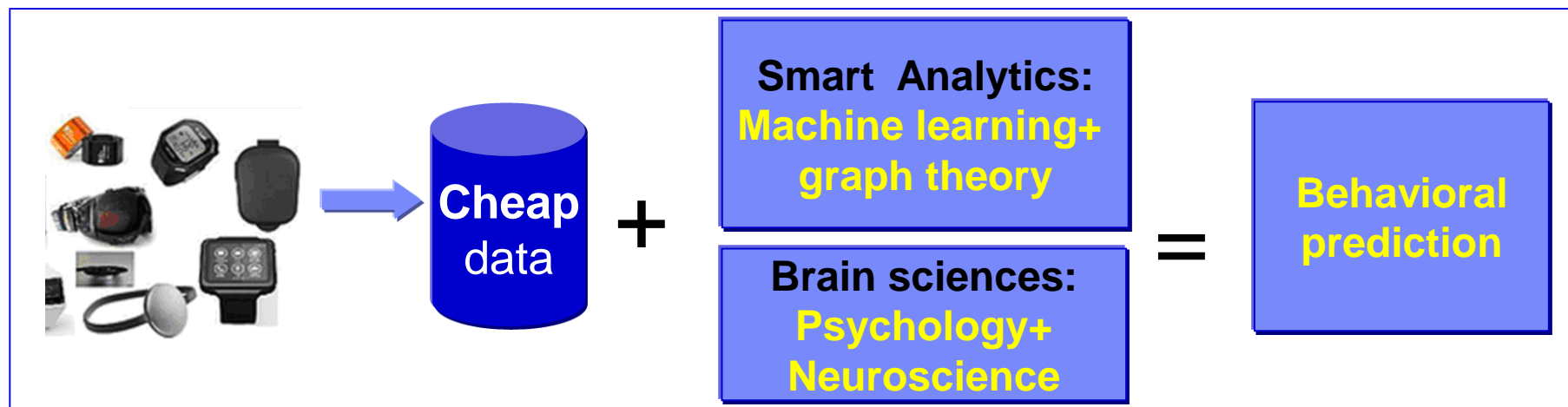
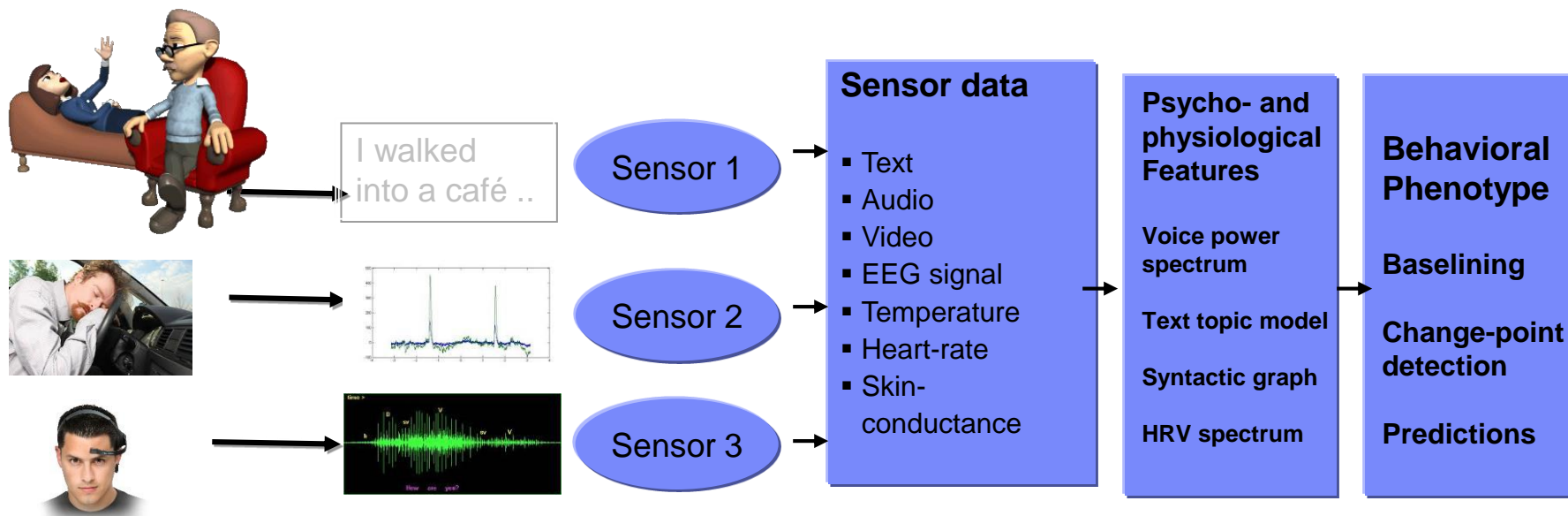


- Note that predictive model significantly outperforms the best single-variable correlations:

- Attention: highest correlation was (negative) 0.59 with sHiAlphaNorm
- Relaxation: highest correlation was (positive) 0.71 with sHiAlphaNorm



Towards “Augmented Human”: Real-Time Mind-Reading from Cheap Sensors





References

[Heisig et al, 2014] S. Heisig, G. Cecchi, R. Rao and I. Rish. Augmented Human: Human OS for Improved Mental Function. AAAI 2014 Workshop on Cognitive Computing and Augmented Human Intelligence.

[Neuropsychopharmacology, 2014] A Window into the Intoxicated Mind? Speech as an Index of Psychoactive Drug Effects. Bedi G, Cecchi G A, Fernandez Slezak D, Carrillo F, Sigman M, de Wit H. *Neuropsychopharmacology*, 2014

[PLoS ONE, 2013] Schizophrenia as a Network Disease: Disruption of Emergent Brain Function in Patients with Auditory Hallucinations, I Rish, G Cecchi, B Thyreau, B Thirion, M Plaze, M-L Paillere-Martinot, C Martelli, J-L Martinot, J-B Poline. *PLoS ONE* 8(1), e50625, Public Library of Science, 2013.

[PLoS One, 2012] Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. N.B. Mota, N.A.P. Vasconcelos, N. Lemos, A.C. Pieretti, O. Kinouchi, G.A. Cecchi, M. Copelli, S. Ribeiro. *PLoS One*, 2012

[SPIE Med.Imaging 2012] Sparse regression analysis of task-relevant information distribution in the brain. Irina Rish, Guillermo A Cecchi, Kyle Heuton, Marwan N Baliki, A Vania Apkarian, SPIE Medical Imaging, 2012.

[AISTATS 2012] J. Honorio, D. Samaras, I. Rish, G.A. Cecchi. Variable Selection for Gaussian Graphical Models. AISTATS, 2012.

[PLoS Comp Bio 2012] Predictive Dynamics of Human Pain Perception, GA Cecchi, L Huang, J Ali Hashmi, M Baliki, MV Centeno, I Rish, AV Apkarian, *PLoS Comp Bio* 8(10), e1002719, Public Library of Science, 2012.

[Brain Informatics 2010] I. Rish, G. Cecchi, M.N. Baliki and A.V. Apkarian. Sparse Regression Models of Pain Perception, in Proc. of Brain Informatics (BI-2010), Toronto, Canada, August 2010.

[NeuroImage, 2009] Prediction and interpretation of distributed neural activity with sparse models. Melissa K Carroll, Guillermo A Cecchi, Irina Rish, Rahul Garg, A Ravishankar Rao *NeuroImage* 44(1), 112--122, Elsevier, 2009.

[NIPS, 2009] Discriminative network models of schizophrenia, GA Cecchi, I Rish, B Thyreau, B Thirion, M Plaze, M-L Paillere-Martinot, C Martelli, J-L Martinot, J-B Poline. *Advances in Neural Information Processing Systems (NIPS 2009)*, pp. 252--260, 2009.



Links

Publication page:

http://researcher.watson.ibm.com/researcher/view_person_pubs.php?person=us-rish&t=1

Books:

Practical Applications of Sparse Modeling, I Rish, GA. Cecchi, A Lozano, A Niculescu-Mizil (editors), MIT Press, 2014.

http://www.amazon.com/Practical-Applications-Modeling-Information-Processing/dp/0262027720/ref=sr_1_2?ie=UTF8&qid=1427846244&sr=8-2&keywords=sparse+modeling

I Rish and G Grabarnik. Chapman and Hall/CRC Machine Learning and Pattern Recognition, 2014.

http://www.amazon.com/Sparse-Modeling-Algorithms-Applications-Recognition/dp/1439828695/ref=sr_1_1?ie=UTF8&qid=1427846244&sr=8-1&keywords=sparse+modeling