Statistical Model Aggregation via Parameter Matching

Mikhail Yurochkin, Mayank Agarwal, Soumya Ghosh, Kristjan Greenewald, Nghia Hoang
MIT-IBM Watson AI Lab, IBM Research, Cambridge, MA

MIT-IBM Watson AI Lab

IBM Research

Goal

Learn statistical models from heterogeneous data produced by related, but unique, generative processes.

Data may be large, siloed, and private making pooling impractical.

Distributed hierarchical Bayesian modeling too communication inefficient to be applicable.

Contributions

We develop meta models for efficiently combining models learned from local data.

Local parameters can exhibit arbitrary permutations. We learn permutation invariant global parameters.

Bayesian nonparametrics allows adaptive learning of global model size.

We make minimal assumptions; widely applicable.

Bayesian nonparametric Meta Model

Draw a measure G from a Beta process with base measure H

$$G = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}; \theta_i \sim H$$

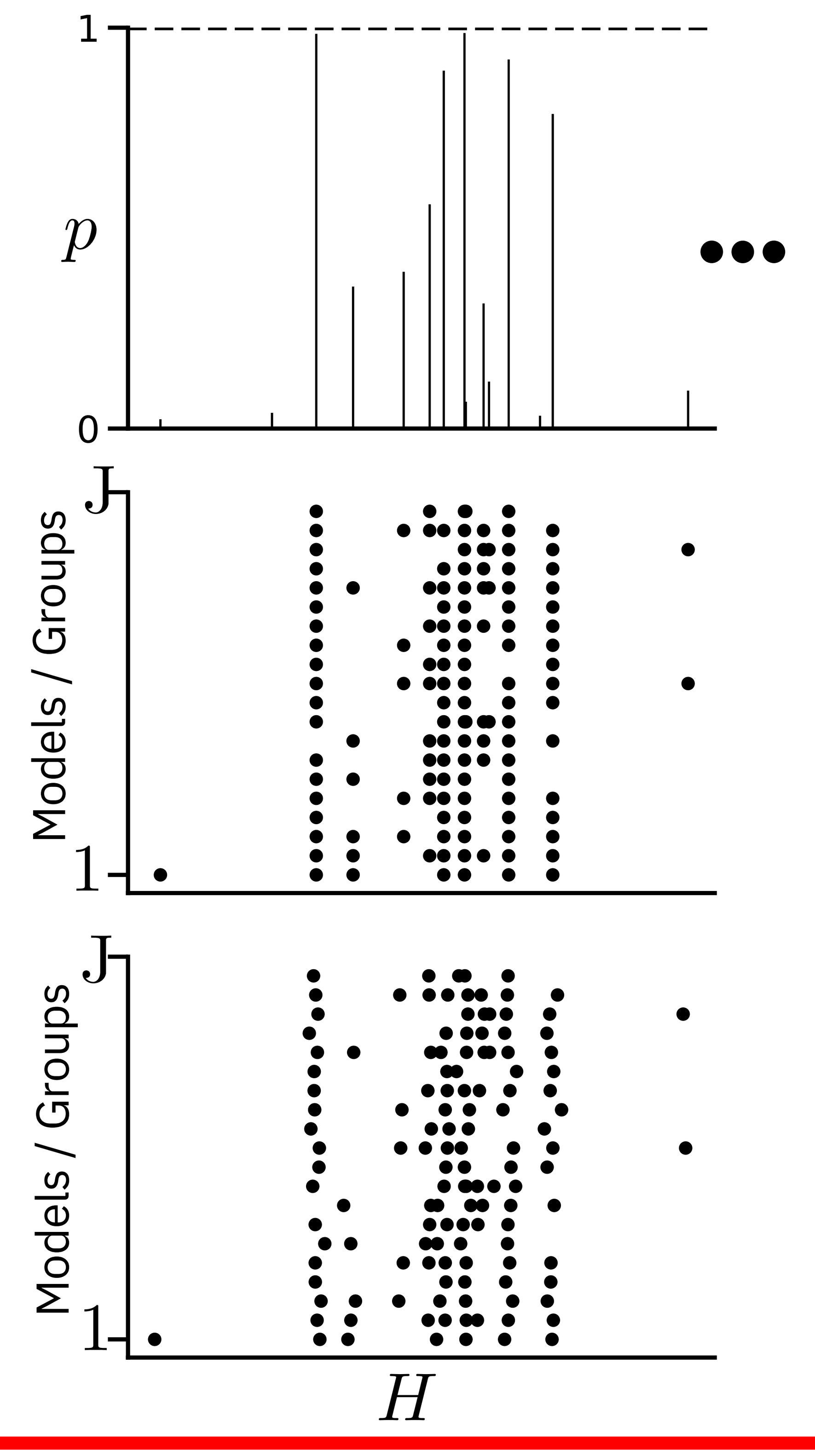
For each model j draw from a Bernoulli process (BeP) centered at G, $Q_j \mid G$

$$b_{ji} \mid p_i \sim \text{Bern}(p_i)$$

$$Q_j = \sum_{i=1}^{\infty} b_{ji} \delta_{\theta_i}$$

Generate parameters of model j centered around global parameters selected by Q_j indexed by l

$$\tilde{\theta}_{jl} \mid \theta_{jl} \sim F(\cdot \mid \theta_{jl})$$



Inference

$$\theta_i \mid b_{\cdot \cdot}, \tilde{\theta}_{\cdot \cdot}$$

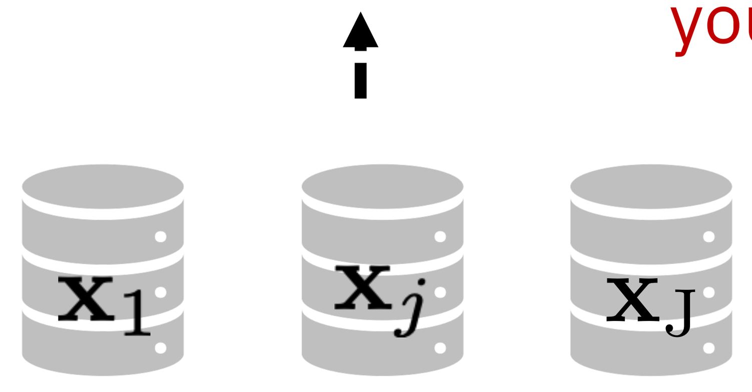
MAP estimate of global parameters

$$b_{j \cdot} \mid \tilde{\theta}_{j \cdot}, b_{\not{j} \cdot}$$

Estimating binary assignments can be cast as a linear sum assignment problem. We use the Hungarian algorithm to solve it

$$\tilde{\theta}_{jl} \mid \mathbf{x}_j$$

Local inference using your favorite algorithm

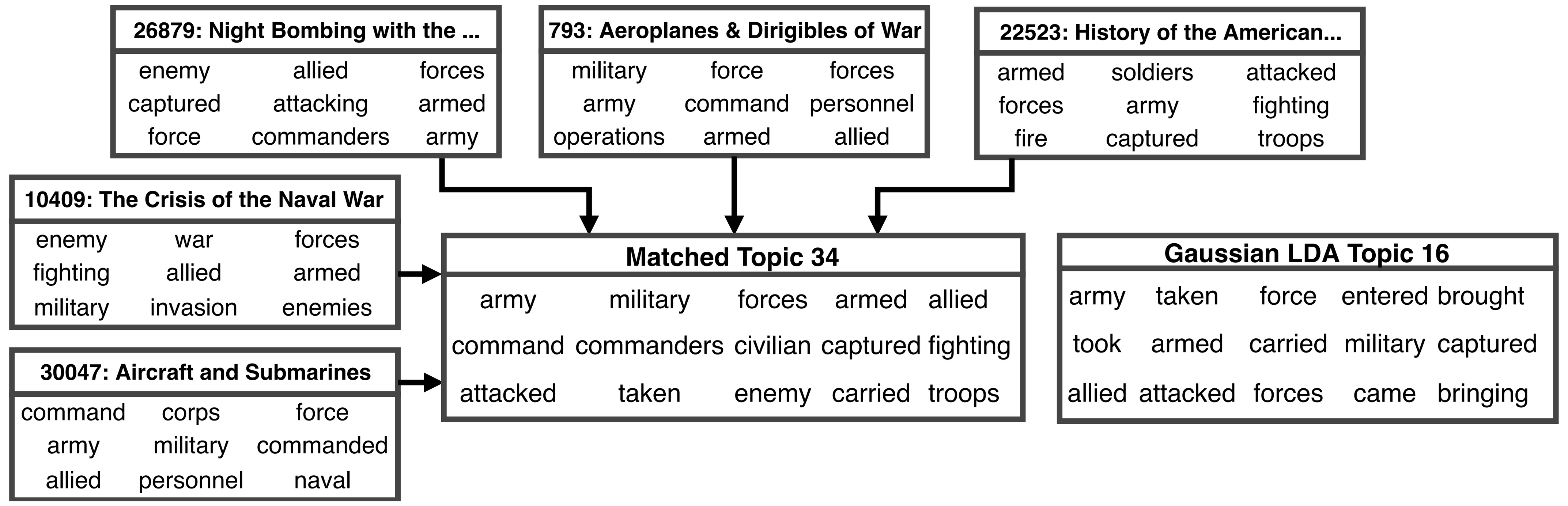


Results

Code: https://github.com/IBM/SPAHM

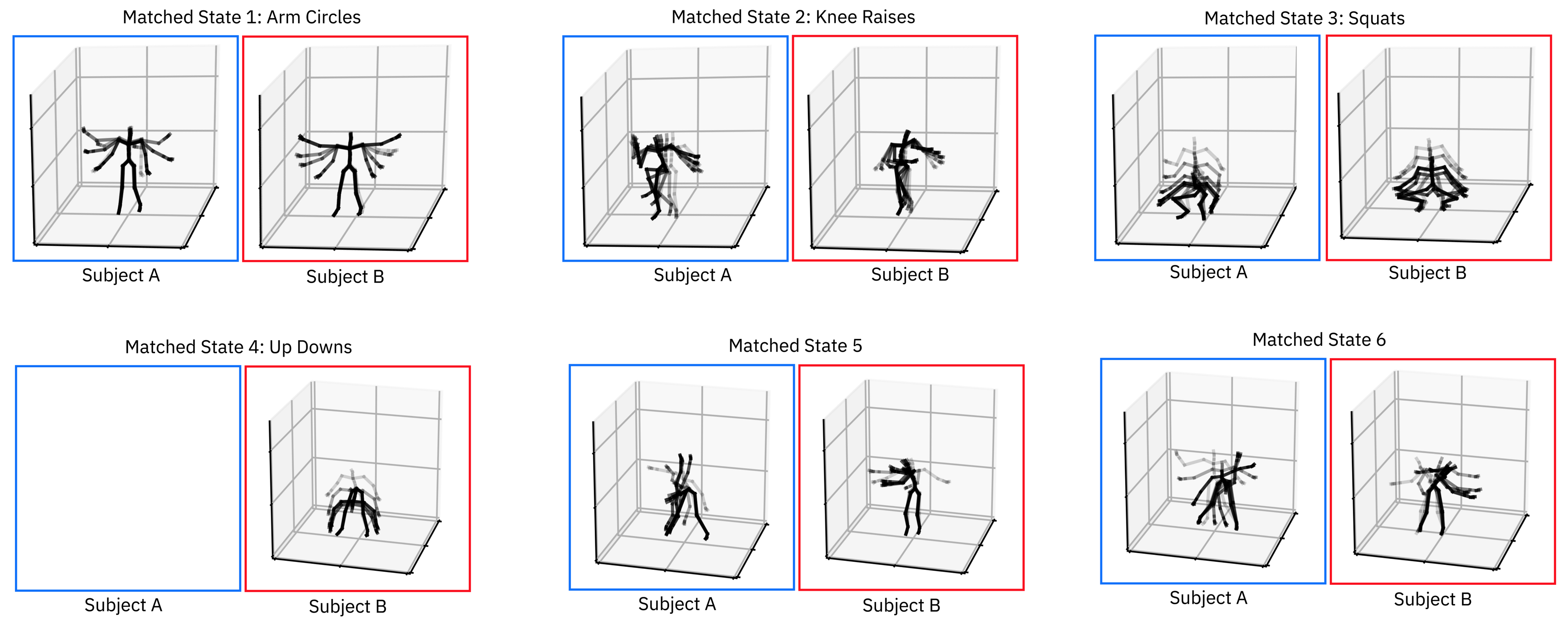
Topic Modeling

Gaussian topics learned from Gutenberg books
1400 times faster than Gibbs sampling, higher coherence



Motion Capture Activity Discovery

HDP-HMM for modeling MoCAP exercise sequences
Twice as fast with comparable performance to memoized VI



Related:
a) Scalable inference of topic evolution via models for latent geometric structures. NeurIPS’19. *Wed 10:45 AM -- 12:45 PM @ East Exhibition Hall B + C #190*
b) Bayesian nonparametric federated learning of neural networks. ICML’ 19.