Statistical Model Aggregation via Parameter Matching

MIT-IBM Watson AI Lab

Mikhail Yurokchin, Mayank Agarwal, Soumya Ghosh, Kristjan Greenewald, Nghia Hoang

MIT-IBM Watson AI Lab, IBM Research, Cambridge, MA

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 is 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 non-parametrics allows adaptive learning of global model size.

We make minimal assumptions; widely applicable.

Bayesian nonparametric Meta Model

Draw a measure from a Beta process

$$G = \sum_i p_i \delta_{\theta_i}; \theta_i \sim H$$

For each model *j* draw from a Bernoulli process

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

$$Q_j = \sum_i b_{ji} \delta_{\theta_i}$$

Generate parameters of model *j* as noisy variants of the set of global parameters selected by the Bernoulli process

$$\tilde{\theta}_{jl} \mid \theta_{c(j,l)} \sim F(\cdot \mid \theta_{c(j,l)})$$

Inference

$$\hat{\theta}_i \mid \hat{b}_{\cdot}, \hat{\hat{\theta}}_{\cdot}$$

MAP estimate of global parameters

$$\hat{b}_{j\cdot} \mid \hat{\hat{\theta}}_{j\cdot}, \hat{b}_{\not{j}\cdot}$$

Optimize binary assignments by iteratively applying the Hungarian algorithm

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

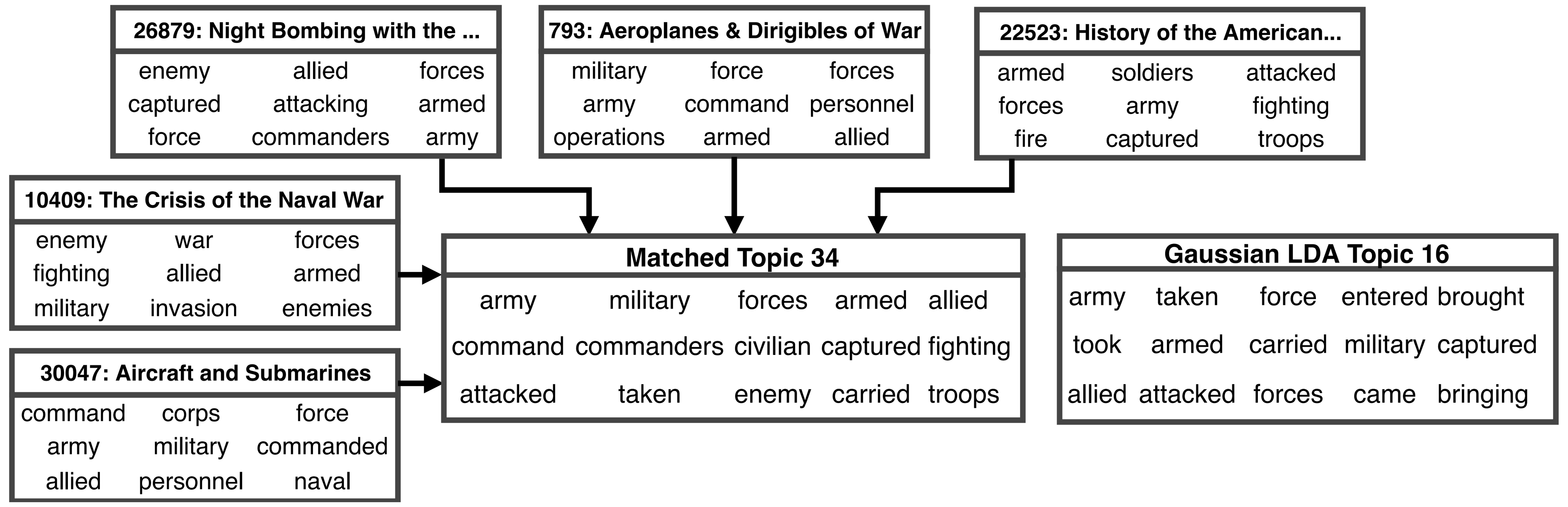
Local inference using your favorite algorithm

Results

Topic Modeling

Gaussian topics learned from Gutenberg books

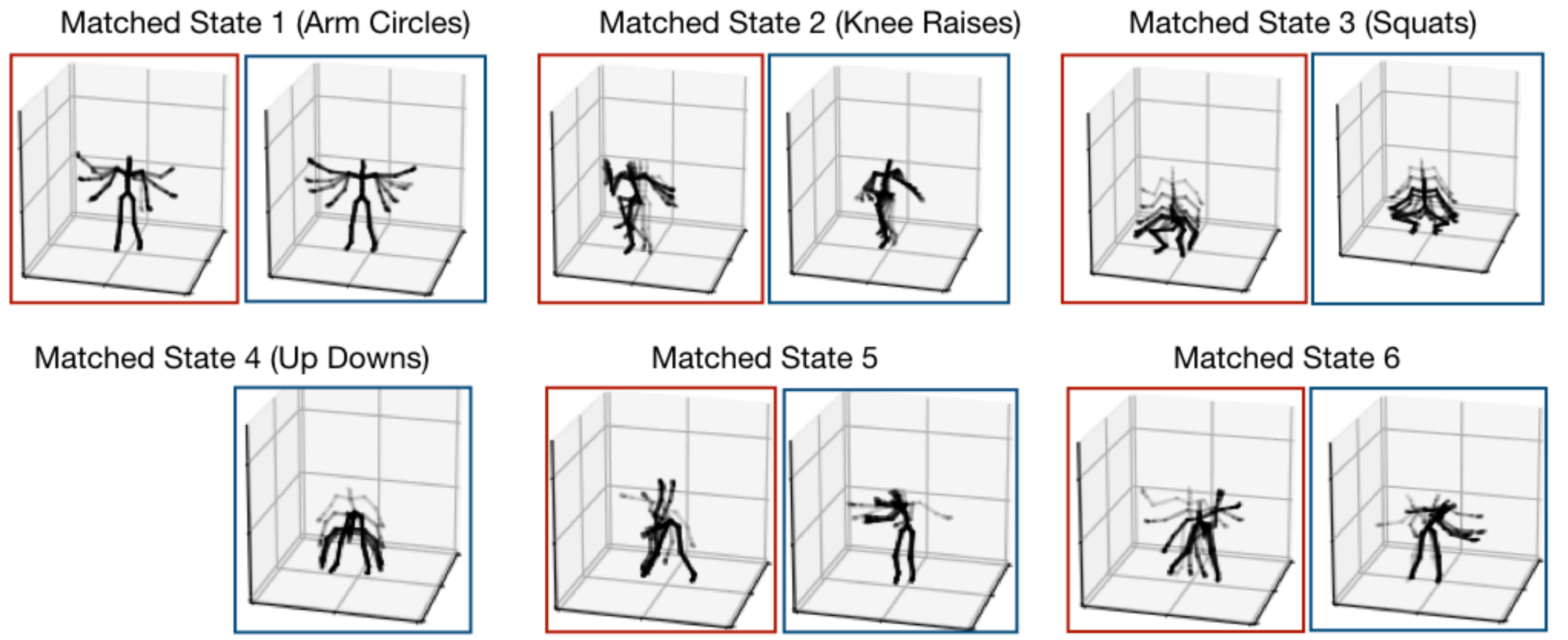
1400 times faster than Gibbs sampling, higher coherence



Motion Capture Activity Discovery

HDP-HMM for modeling MoCAP exercise sequences

Twice as fast with comparable performance to memorized VI



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