Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Variational Inference

Machine Learning: Jordan Boyd-Graber
University of Colorado Boulder
LECTURE 21

Roadmap

- Big-picture questions
- VI for LDA
- More content questions
- Walkthrough of VI for LDA (HW)

Example

- Three topics, same documents as last time

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix} \tag{1}$$

- Assume uniform γ: $(2.0, 2.0, 2.0)$
- Compute update for ϕ

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right) \tag{2}$$

- For a the first word (dog) in the document: **dog cat cat pig**

Example

- Three topics, same documents as last time

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix} \tag{1}$$

- Assume uniform γ: $(2.0, 2.0, 2.0)$
- Compute update for ϕ

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right) \tag{2}$$

- For a the first word (dog) in the document: **dog** **cat cat pig**

Update ϕ for dog

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$

Update ϕ for dog

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for dog

$$\beta = \begin{array}{c} \\ \\ \end{array} \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for dog

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(2) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for dog

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi(2.000) - \Psi(2.000 + 2.000 + 2.000)\right) = 0.051$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi(2.000) - \Psi(2.000 + 2.000 + 2.000)\right) = 0.051$
- $\phi(2) \propto 0.185 \times \exp\left(\Psi(2.000) - \Psi(2.000 + 2.000 + 2.000)\right) = 0.051$
- After normalization: $\{0.333, 0.333, 0.333\}$

Update ϕ for pig

$$\beta = \begin{array}{c} \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix} \end{array}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$

Update ϕ for pig

$$\beta = \begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ \end{array}$$

$$\beta = \begin{bmatrix} .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for pig

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$\phi_{ni} \propto$
$\beta_{iv} \exp \left(\Psi \left(\gamma_i \right) - \Psi \left(\sum_j \gamma_j \right) \right)$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp \left(\Psi \left(2.000 \right) - \Psi \left(2.000 + 2.000 + 2.000 \right) \right) = 0.051$
- $\phi(1) \propto 0.185 \times \exp \left(\Psi \left(2.000 \right) - \Psi \left(2.000 + 2.000 + 2.000 \right) \right) = 0.051$

Update ϕ for pig

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(2) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for pig

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(2) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- After normalization: $\{0.333, 0.333, 0.333\}$

Update ϕ for cat

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$\phi_{ni} \propto$

$\beta_{iv} \exp \left(\Psi(\gamma_i) - \Psi\left(\sum_j \gamma_j\right)\right)$

- $\gamma = (2.000, 2.000, 2.000)$

Update ϕ for cat

$$\beta = \left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array} \right]$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.072$

Update ϕ for cat

$$\beta = \begin{array}{c} \\ \\ \end{array}\left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array}\right]$$

$\phi_{ni} \propto$

$\beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.072$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for cat

$$\beta = \begin{bmatrix} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{bmatrix}$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.072$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(2) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$

Update ϕ for cat

$$\beta = \begin{array}{c} \\ \\ \end{array}\left[\begin{array}{ccccc} \text{cat} & \text{dog} & \text{hamburger} & \text{iron} & \text{pig} \\ .26 & .185 & .185 & .185 & .185 \\ .185 & .185 & .26 & .185 & .185 \\ .185 & .185 & .185 & .26 & .185 \end{array}\right]$$

$$\phi_{ni} \propto \beta_{iv} \exp\left(\Psi\left(\gamma_i\right) - \Psi\left(\sum_j \gamma_j\right)\right)$$

- $\gamma = (2.000, 2.000, 2.000)$
- $\phi(0) \propto 0.260 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.072$
- $\phi(1) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- $\phi(2) \propto 0.185 \times \exp\left(\Psi\left(2.000\right) - \Psi\left(2.000 + 2.000 + 2.000\right)\right) = 0.051$
- After normalization: $\{0.413,\ 0.294,\ 0.294\}$

Update γ

- Document: dog cat cat pig
- Update equation

$$\gamma_i = \alpha_i + \sum_n \phi_{ni} \tag{3}$$

- Assume $\alpha = (.1, .1, .1)$

Update γ

- Document: dog cat cat pig
- Update equation

$$\gamma_i = \alpha_i + \sum_n \phi_{ni} \tag{3}$$

- Assume $\alpha = (.1, .1, .1)$

	ϕ_0	ϕ_1	ϕ_2
dog	.333	.333	.333
cat	.413	.294	.294
pig	.333	.333	.333
α	0.1	0.1	0.1
sum	1.592	1.354	1.354

- Note: **do not normalize**!

Update γ

- Document: dog cat cat pig
- Update equation

$$\gamma_i = \alpha_i + \sum_n \phi_{ni} \quad (3)$$

- Assume $\alpha = (.1, .1, .1)$

	ϕ_0	ϕ_1	ϕ_2	
dog	.333	.333	.333	
cat	.413	.294	.294	x2
pig	.333	.333	.333	
α	0.1	0.1	0.1	
sum	1.592	1.354	1.354	

- Note: **do not normalize**!

Update β

- Count up all of the ϕ across all documents
- For each topic, divide by total
- Corresponds to maximum likelihood of expected counts

Update β

- Count up all of the ϕ across all documents
- For each topic, divide by total
- Corresponds to maximum likelihood of expected counts
- Unlike Gibbs sampling, no Dirichlet prior

Plan

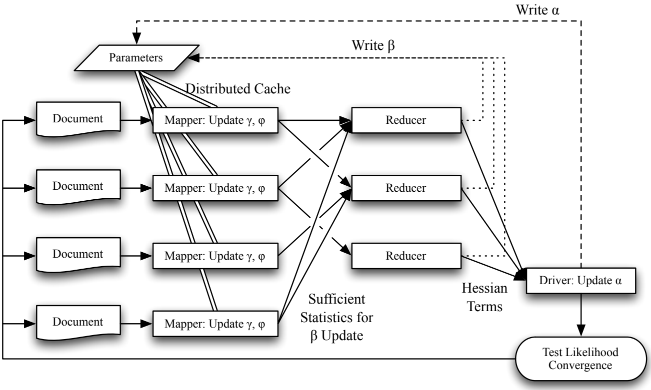
What research is going on in variational inference?

Automatic Inference

```
public void GaussianModel(double[] data)
{
  double mean = Factor.Random(new Gaussian(0, 100));
  double precision = Factor.Random(new Gamma(0, 1));
  for (int i = 0; i < data.Length; i++) {
    data[i] = Factor.Gaussian(mean, precision);
  }
  InferNet.Infer(mean);
  InferNet.Infer(precision);
}
```

Parallel LDA

Zhai et al, 2012



Online LDA

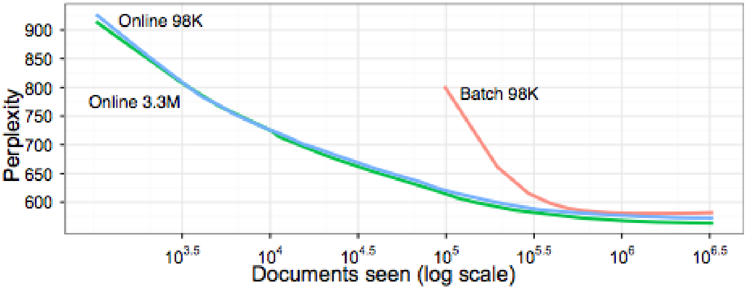
Hoffman and Blei, 2010

Algorithm 2 Online variational Bayes for LDA

Define $\rho_t \triangleq (\tau_0 + t)^{-\kappa}$
Initialize λ randomly.
for $t = 0$ to ∞ **do**
  E step:
  Initialize $\gamma_{tk} = 1$. (The constant 1 is arbitrary.)
  repeat
    Set $\phi_{twk} \propto \exp\{\mathbb{E}_q[\log \theta_{tk}] + \mathbb{E}_q[\log \beta_{kw}]\}$
    Set $\gamma_{tk} = \alpha + \sum_w \phi_{twk} n_{tw}$
  until $\frac{1}{K} \sum_k |\text{change in} \gamma_{tk}| < 0.00001$
  M step:
  Compute $\tilde{\lambda}_{kw} = \eta + D n_{tw} \phi_{twk}$
  Set $\lambda = (1 - \rho_t)\lambda + \rho_t \tilde{\lambda}$.

Online LDA

Hoffman and Blei, 2010



Online LDA

Hoffman and Blei, 2010

Initialize $\lambda^{(0)}$ randomly.
Set the step-size schedule ρ_t appropriately.
repeat
    Sample a data point x_i uniformly from the data set.
    Compute its local variational parameter,

$$\phi = \mathbb{E}_{\lambda^{(t-1)}}[\eta_g(x_i^{(N)}, z_i^{(N)})].$$

    Compute intermediate global parameters as though x_i is replicated N times,

$$\hat{\lambda} = \mathbb{E}_{\phi}[\eta_g(x_i^{(N)}, z_i^{(N)})].$$

    Update the current estimate of the global variational parameters,

$$\lambda^{(t)} = (1-\rho_t)\lambda^{(t-1)} + \rho_t \hat{\lambda}.$$

until forever

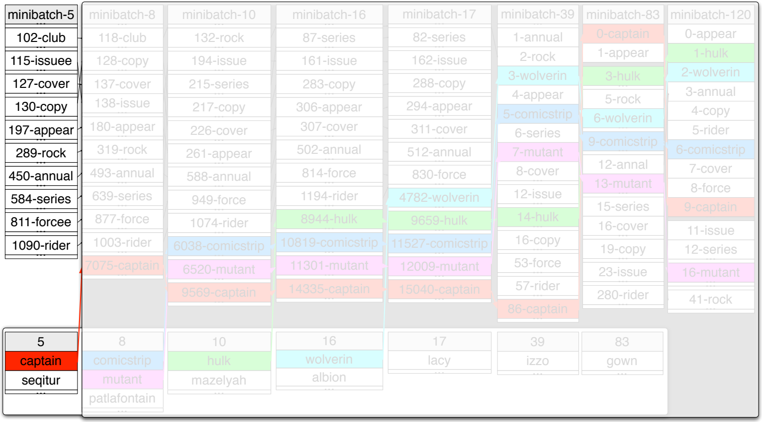
Best of Both Worlds

Algorithm 1 Algorithm for hybrid stochastic variational-Gibbs inference.

for $t \in 1, ..., \infty$ **do**
  $\rho_t \leftarrow \left(\frac{1}{t_0+t}\right)^{\kappa}$
  sample minibatch $\mathcal{B}$
  for $d \in \mathcal{B}$ **do**
    initialize $\boldsymbol{z}_d^0$
    discard B burn-in sweeps
    for sample $s \in 1, ..., S$ **do**
      for token $i \in 1, ..., N_d$ **do**
        sample $z_{di}^s \propto (\alpha + N_{dk}) e^{\mathbb{E}_q[\log \beta_{kw}]}$
      end for
    end for
  end for
  $\lambda_{kw}^t \leftarrow (1 - \rho_t)\lambda_{kw}^{t-1} + \rho_t \left(\eta + \frac{D}{|\mathcal{B}|}\hat{N}_{kw}\right)$
end for

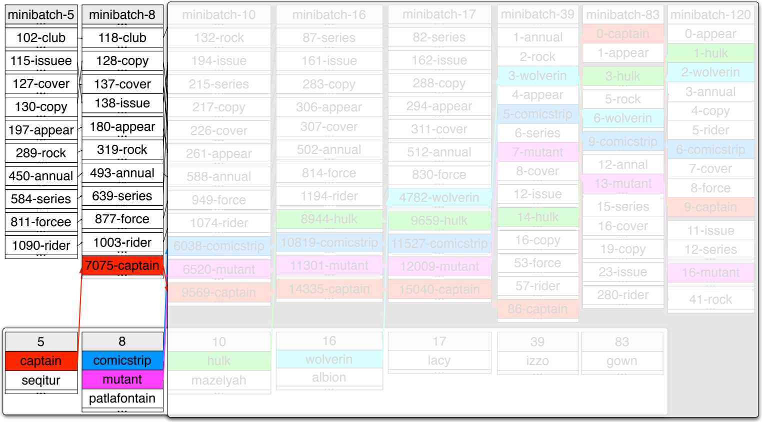
Matching Models and Inference

Zhai and Boyd-Graber, 2013



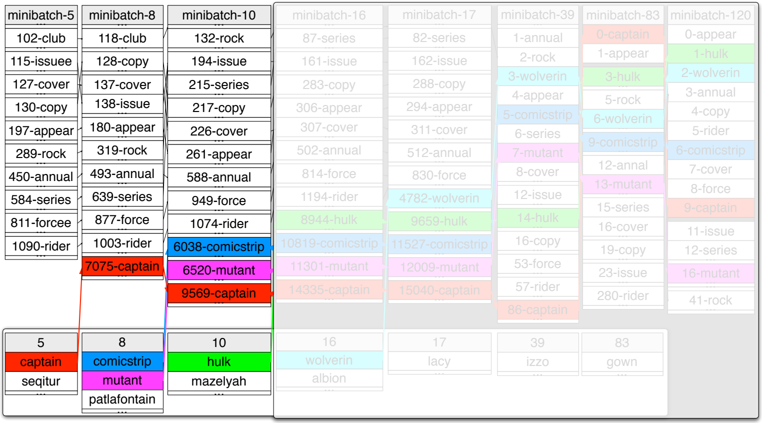
Matching Models and Inference

Zhai and Boyd-Graber, 2013



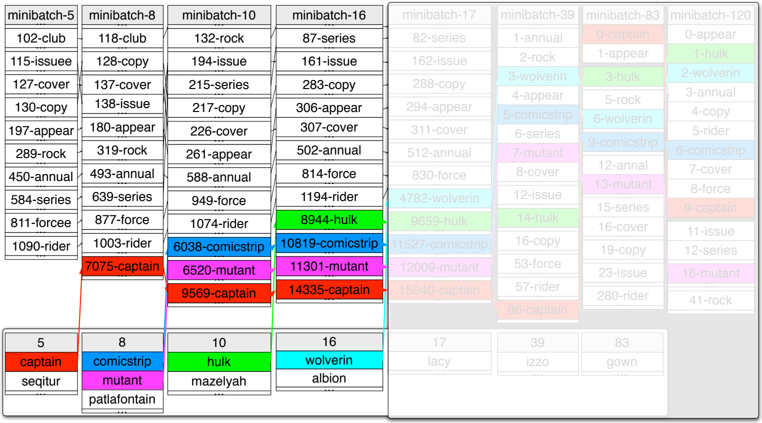
Matching Models and Inference

Zhai and Boyd-Graber, 2013



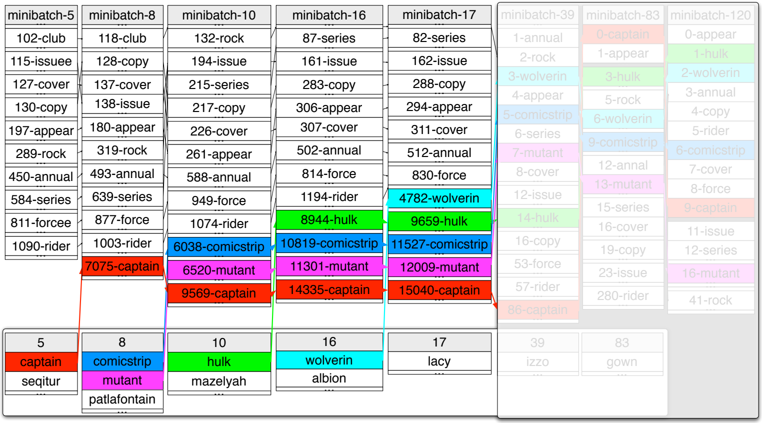
Matching Models and Inference

Zhai and Boyd-Graber, 2013



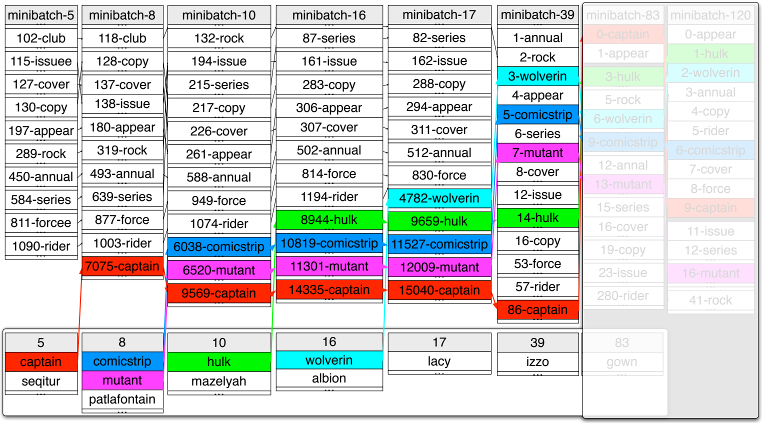
Matching Models and Inference

Zhai and Boyd-Graber, 2013



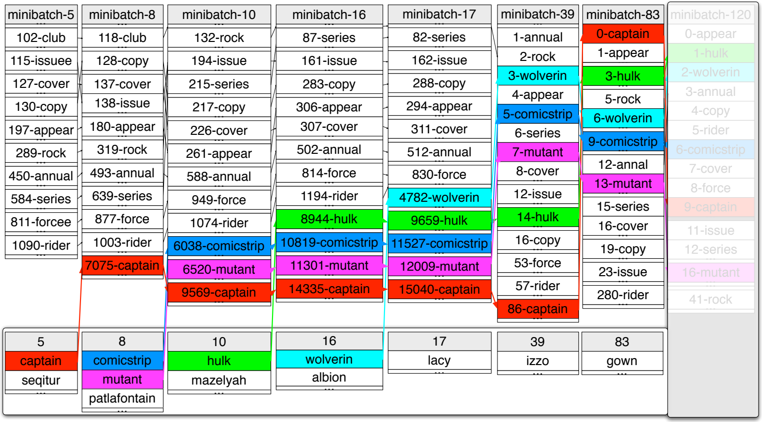
Matching Models and Inference

Zhai and Boyd-Graber, 2013



Matching Models and Inference

Zhai and Boyd-Graber, 2013



Matching Models and Inference

Zhai and Boyd-Graber, 2013