Capturing Semantically Meaningful Word Dependencies with an Admixture of Poisson MRFs
David I. Inouye, Pradeep Ravikumar, Inderjit S. Dhillon

Admixture of Poisson MRFs (APM) [Inouye et al. 2014]

Open Problems in APM Model
1. High computational complexity of APM
   - No parallelism since optimizing jointly over all parameters
   - Slow convergence of proximal gradient descent
   - APM has $O(k^2)$ parameters versus $O(k)$ for LDA
2. Edge parameters of APM not directly evaluated
   - Previous metrics calculated word pair statistics for top words
   - However, APM model dependencies between words
   - How can we semantically evaluate these dependencies?

Proposed Solutions
1. Parallel alternating Newton-like algorithm
   - Split into convex problems
   - Demonstrate scaling at $p = 10,000$ and $n = 100,000$
   - Empirically, $O(k^2)$ complexity to estimate $O(k + kp^2)$ parameters
   - http://bigdata.ics.uci.edu/software/apm/
2. Evocation metric that directly evaluates word pairs
   - Develop novel metric based on notion of evocation
   - (which words "bring to mind" other words)

Evocation Metric
- Word pairs for Evoc-2 ($m = 50$) ordered by human score
- Parallel speedup on BNC Corpus
- Qualitative Analysis of Evocation
- Best APM Model ($k = 5$)

Parallel Alternating Newton-like Algorithm (Code available*)
1. Split the algorithm into alternating steps
   - Posterior is convex in $W$ or $\Phi^k$ but not both
   - Similar to EM for mixture models or ALS for NMF
   - $\arg\min_{w} \sum_{j=1}^{n} \sum_{o=1}^{k} \log \Phi_{o}^{j} w_{o}^{j} + \lambda \sum_{i} \sum_{o=1}^{k} \psi_i \Phi_{o}^{j} w_{o}^{j}$
   - $\Phi = \arg\min_{\Phi} \sum_{i=1}^{n} \sum_{o=1}^{k} \log \Phi_{o}^{j} w_{o}^{j} + \sum_{i=1}^{n} \Phi_{o}^{j} w_{o}^{j}$

Timing Results on Wikipedia
- Timing results for different sizes of a Wikipedia dataset show that the algorithm scales approximately as $O(np^2)$. (fixed $k = 5$, $\lambda = 0.5$)
- Parallel speedup is approximately linear when using a simple parallel loop in MATLAB. Subproblems are all independent so speedup could be $O(m/n p)$ on distributed system. Experimental on BNC corpus $(p = 1646$ and $n = 4049)$ fixing $k = 5$, $\lambda = 8$ and running for 30 alternating iterations.

Evocation [Boyd-Graber et al. 2006]
- Evocation denotes the idea of which words "evoke" or "bring to mind" other words
- Distinctive from word similarity or synonymy
- Types of evocation: Rose - Flower (example), Brave - Noble (kind), Yell - Talk (manner), Eggs - Bacon (co-occurrence), Snore - Sleep (setting), Wet - Desert (antonymy), Work - Lazy (exclusivity), Banana - Kiwi (likeness)
- Red highlights pairs that seem semantically uninteresting
- Blue highlights pairs that seem semantically interesting