MTP Progress

Nikhilkumar Jadhav

Evaluation of Topic Models

- Topic modeling is useful for analyzing large unstructured text collections

- Due to unsupervised nature, it is difficult to different models

- Extrinsic tasks like document classification or IR can be used for evaluation

- Need of a evaluation method which is
  - Accurate
  - Computationally efficient
  - Independent of specific application

9/3/2013

Evaluation of Topic Models

- Estimating probability of held-out documents is one good method

- Probability of a set of held-out documents, $\mathcal{W}$ given a set of training documents, $\mathcal{W}'$ can be written as,

- $p(\mathcal{W}|\mathcal{W}') = \int p(\mathcal{W}|parameters)\, p(parameters|\mathcal{W}')$

References

- *Wallach, Hanna M., et al. "Evaluation methods for topic models." Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009.*

9/3/2013