

MTP Progress

17/09/2013

Evaluation of topic models

- Leave out some documents called held-out documents
- Find the probability of the held-out documents using parameters learned from the training documents

Evaluation of topic models

- **Read** *Wallach, Hanna M., et al. "Evaluation methods for topic models." Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009.*
- **Downloaded** PubMed central abstracts
- **Implemented** Left-to-right evaluation of held-out documents

Left-to-right evaluation

- Algorithm
 - $L = 0$
 - For each word w_n of the document
 - $P_n = 0$
 - For each topic t from 1 to K
 - $P_n = P_n + p(w_n, z_n=t \mid \text{topic assignments of all words before } w_n)$
 - $L = L + \log(P_n)$
 - L is the log probability of the document

Evaluation using PubMed central abstracts

- 2065 abstracts
- Held out documents : 207
- Training documents : 1858
- Total log probability of held-out documents :
-81429.23699786628

Future plan

- Reading papers on Joint topic sentiment modeling