Latent Dirichlet Analysis for Document Topic Modelling

Xiang Li
xngli@umich.edu

Zheng Luo
zlou@umich.edu

Yan Chen
yanchenm@umich.edu

Sajan Patel
sajanptl@umich.edu

Abstract

Abstract goes here. Talk about how LDA can be used for topic modeling. Our method and implementation is built on the Blei et. al. JMLR paper. We recreated the experiment for document modeling from the paper. We also applied LDA to the Yelp Dataset. We present our implementation of the method, experiment, and our analysis of the results in this paper.

1 Introduction

Here, describe the problem statement: Given a text document, model the topics of the document. (expand on that more).

1.1 Related Works

Based on Blei, briefly talk about unigram, mixture of unigram, and plsi.

2 Method

Describe the graphical model of the document and present the algorithm described in the paper. This should be like a summary of the algorithm's ideas. We can save the details of variational inference and gibbs sampling for our implementation section since we only have 8 pages worth of space. A picture here of the graphical model would also be good.

2.1 Inference Method

Show derivation and algorithm

2.2 Gibbs Sampling Method

Show derivation and algorithm

3 Our Implementation

We implemented LDA using both Gibbs sampling and inference.

4 Experimental Evaluation

4.1 Setup

Describe recreation of Document Modeling and Yelp experiment from Blei et. al. here.

4.2 Metrics

Describe perplexity here.

4.3 Results

Show graphs of the perplexity and histogram of topics.

Discuss results in table if also needed.

Also compare gibbs vs inference in each.

5 Conclusion

Discuss the subsequent conclusions we gained from this reimplementation of LDA. Summarize advantages and disadvantages as well.

Acknowledgements

Describe efforts and work division here.

References

Use Bibtex here. We will only have a few references.