Topic Modeling using Latent Dirichlet Allocation

Hannah Chen hechen@eng.ucsd.edu
Sonali Rahagude srahagud@ucsd.edu
Sneha Venkatesh Yelimeli svyelime@eng.ucsd.edu
Chun Fan c9fan@ucsd.edu
Yashodhan Karandikar ykarandi@eng.ucsd.edu

Abstract

1 Introduction

2 Background

3 Design

4 Results

5 Conclusions

A Appendix

References

[1] Charles Elkan. Clustering documents with an exponential-family approximation of the dirichlet compound multinomial distribution. In *Proceedings of the 23rd international conference on Machine learning*, pages 289–296. ACM, 2006.

[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[3] Lindsay I Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

[4] Gregor Heinrich, Jörg Kindermann, Codrina Lauth, Gerhard Paaß, and Javier Sanchez-Monzon. Investigating word correlation at different scopes—a latent concept approach. In *Workshop Lexical Ontology Learning at Int. Conf. Mach. Learning*, 2005.

[5] Marina Meilă. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[6] Gregor Heinrich. Parameter estimation for text analysis. Technical report, 2004.