Give me a title!

Virginia Partridge
University of Massachusetts Amherst
vcpartridge@umass.edu

Abstract

TODO

1 Motivation

Latent Dirichlet Analysis (LDA) is a widely adopted approach for unsupervised topic modeling and has been used across disciplines for exploring themes and trends in collections of documents. LDA has been applied to explore the ever-growing collections of text from online platforms and social media and analyze language changes in academic fields over time (Koltsova and Koltsov, 2013; McFarland et al., 2013; Vogel and Jurafsky, 2012; Mitrofanova, 2015).

2 Related Work

3 Methods

3.1 Framework for discussing morphological complexity

lemmas vs lexemes vs types, paradigms tokens vs surface forms slots Russian as a flective langauage inflectional vs derivational morphology

3.2 Latent Dirichlet Analysis

symmetric vs asymmetric prior mallet gibbs sampling implementation

3.3 Evaluation metrics

4 Corpus

4.1 Conflation methods and vocabulary reduction

References

Olessia Koltsova and Sergei Koltsov. 2013. Mapping the public agenda with topic modeling: The case of the russian livejournal. *Policy & Internet*, 5.

Daniel A McFarland, Daniel Ramage, Jason Chuang, Jeffrey Heer, Christopher D Manning, and Daniel Jurafsky. 2013. Differentiating language usage through topic models. *Poetics*, 41(6):607–625.

Olga Mitrofanova. 2015. Probabilistic topic modeling of the russian text corpus on musicology. In *International Workshop on Language, Music, and Computing*, pages 69–76. Springer.

Adam Vogel and Dan Jurafsky. 2012. He said, she said: Gender in the ACL Anthology. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*, pages 33–41, Jeju Island, Korea. Association for Computational Linguistics.