

Shifting Tone of a Body of Text

Cole Hollant

**Adviser: Stefan Méndez-Diez, Robert McGrail,
Kerri-Ann Norton**

Computer Science and Mathematics

November 2019

1. Abstract

- Natural language processing
 - Tone recognition: label input text
 - Tone alteration: align text with a different category
- Plutchik's basic emotions
 - Joy, trust, fear, surprise, sadness, disgust, anger, anticipation

2. High Level Plans

- Build out datasets / lexicons
 - Sourced vs created
 - * NRC Emotion / Affect lexicons
 - Unigram limitations
 - * Twitter scraper
 - * Scraped thesaurus
- Make different models
 - Arbitrary replacement (control)
 - Latent Dirichlet Allocation
 - Neural nets?
 - * Potential n-gram benefits
- User testing

3. Latent Dirichlet Allocation

- Generative statistical topic model
 - document \rightarrow set of topics
- Bayesian inference for topic mixture
- Soft clustering
 - fuzzy membership
 - topic reduction

4. Current Progress

- Thesaurus
- Lexicons
- Website
 - Frontend
 - API
 - Server setup, etc

Next Steps

- More server configuration
- Build LDA model
- Alter data visualization

5. Thank you!

Here are links if anyone is interested

- Frontend: sproj.colehollant.com
- Backend: sproj.api.colehollant.com
- Version control: [my github](#)

References

- [1] Big huge thesaurus api. <https://words.bighugelabs.com/>. (Accessed on 11/12/2019).
- [2] Ng David M., Andrew Y., and Michael I. Latent dirichlet allocation, Jan 1970.
- [3] Thomas L Griffiths and Mark Steyvers. Finding scientific topics, Apr 2004.
- [4] Saif M. Mohammad. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia, 2018.
- [5] Saif M. Mohammad. Word affect intensities. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan, 2018.
- [6] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. 29(3):436–465, 2013.
- [7] Jonathan K. Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.