

MAD Style: Multivalent Authorship Detection (MAD) Topic Models

David Dohan, Charles Marsh, Shubhro Saha, Max Simchowitz

Princeton University, Department of Computer Science

Goals

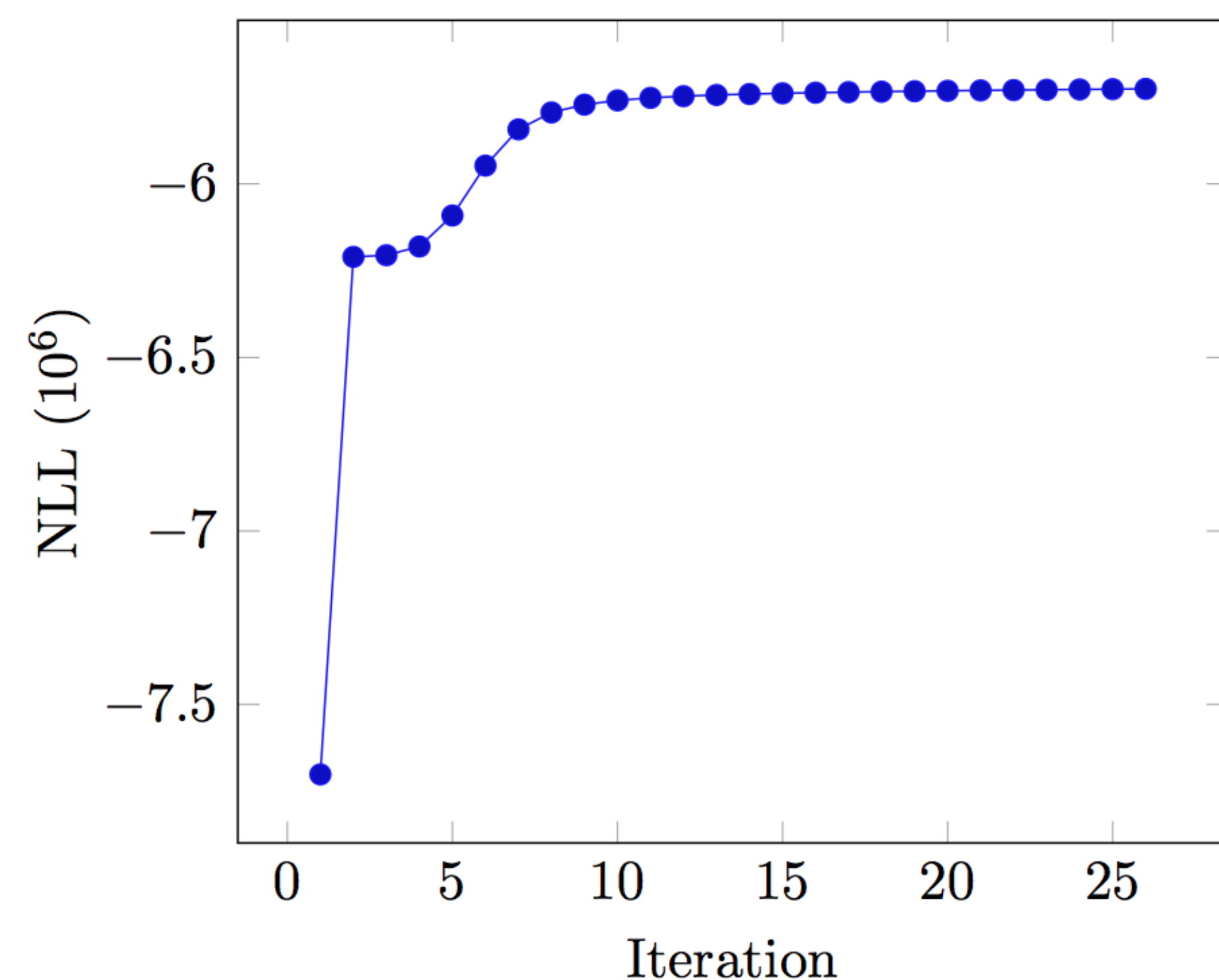
- Classify author writing style across a wide range of lexical input.
- Determine which features are most characteristic of authors' writing styles.
- Extract compact representations of per-author stylistic tendency.

Introduction

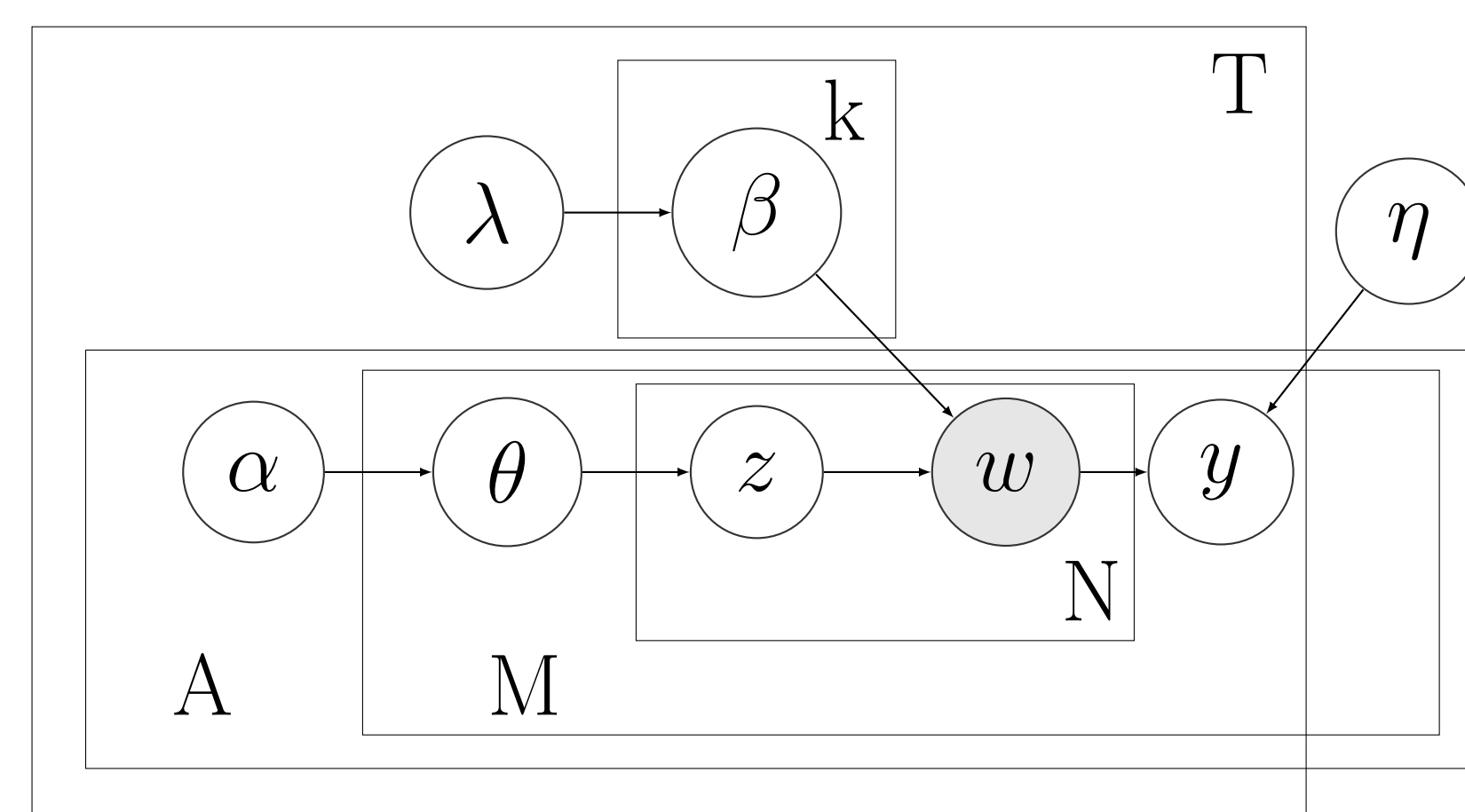
In the *authorship detection* problem, one is given:

- A set of documents labeled (by author) on which to train.
- A set of anonymized documents to classify.

Methods for authorship detection traditionally depend on careful feature extraction and rather black-box methods. Hence, they rely on extensive domain specific knowledge, and can be difficult to decipher. We present the *MAD Topic Model*, which uses syntactic and stylometric n -gram features (e.g., part-of-speech tags, meter). MAD fits separate topic models to each of these n -gram vocabularies and combines the models through a multi-class logistic regression classifier. After fitting the topic model parameters, new documents can be classified using the multi-class component. As a by-product, MAD also breaks stylistic features into vocabularies over topics, creating a compact representation of stylistic tendency.



Model



Graphical Model for the MAD Topic Model

The MAD Topic Model combines the SLDA algorithm presented in [2] with an Author Topic Model, and extends both to account for multiple word types. The model is based on variational inference, following the coordinate ascent updates in [2]. Stochastic variational inference was also tested, but proved impractical for these rather small data sets.

Summary

The Multivalent Authorship Detection (MAD) Topic Model extends Latent Dirichlet Allocation to identify authorship in documents with many separate types (“multivalent”) of count features. MAD is “doubly supervised”: it includes a multi-class logistic regression and also fits per-author Dirichlet distributions for each feature type. We test the MAD Topic Model on several real world corpora using a variety of n -gram features, including part-of-speech, syllable stress, and sequences of word lengths.

Data

We focused on three corpora: Project Gutenberg, Nassau Weekly, and Quora.

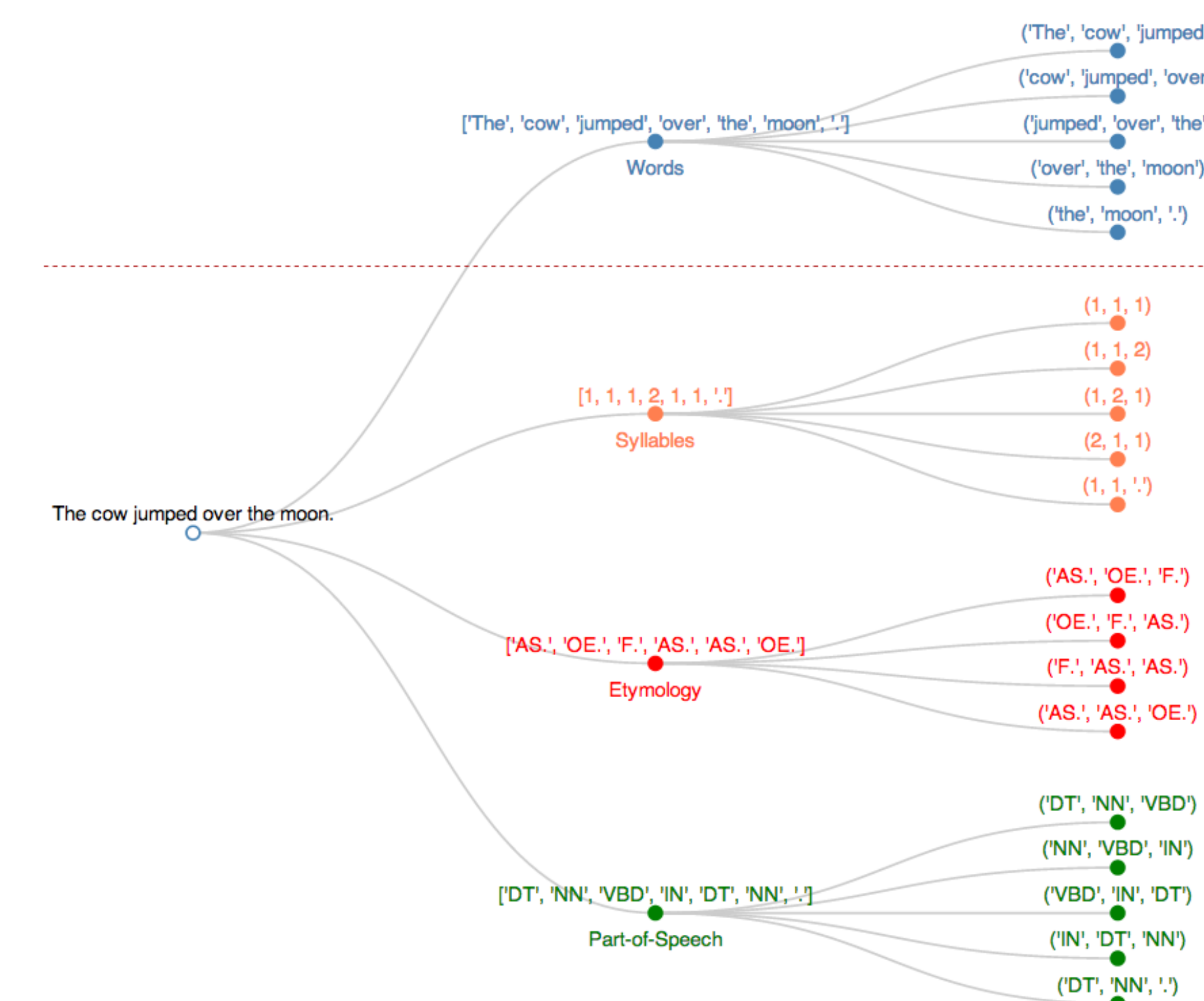
Datasets collected for training and testing

Source	Authors	Documents
Project Gutenberg	5	250
Nassau Weekly	200	550
Quora	100	1600

Project Gutenberg contains excerpts from fictional books. Nassau Weekly features narrative & editorial articles from a campus publication. Quora captures responses a question-answer site. Diversity in topic and length challenges our model to detect consistent features in a variety of contexts.

Features

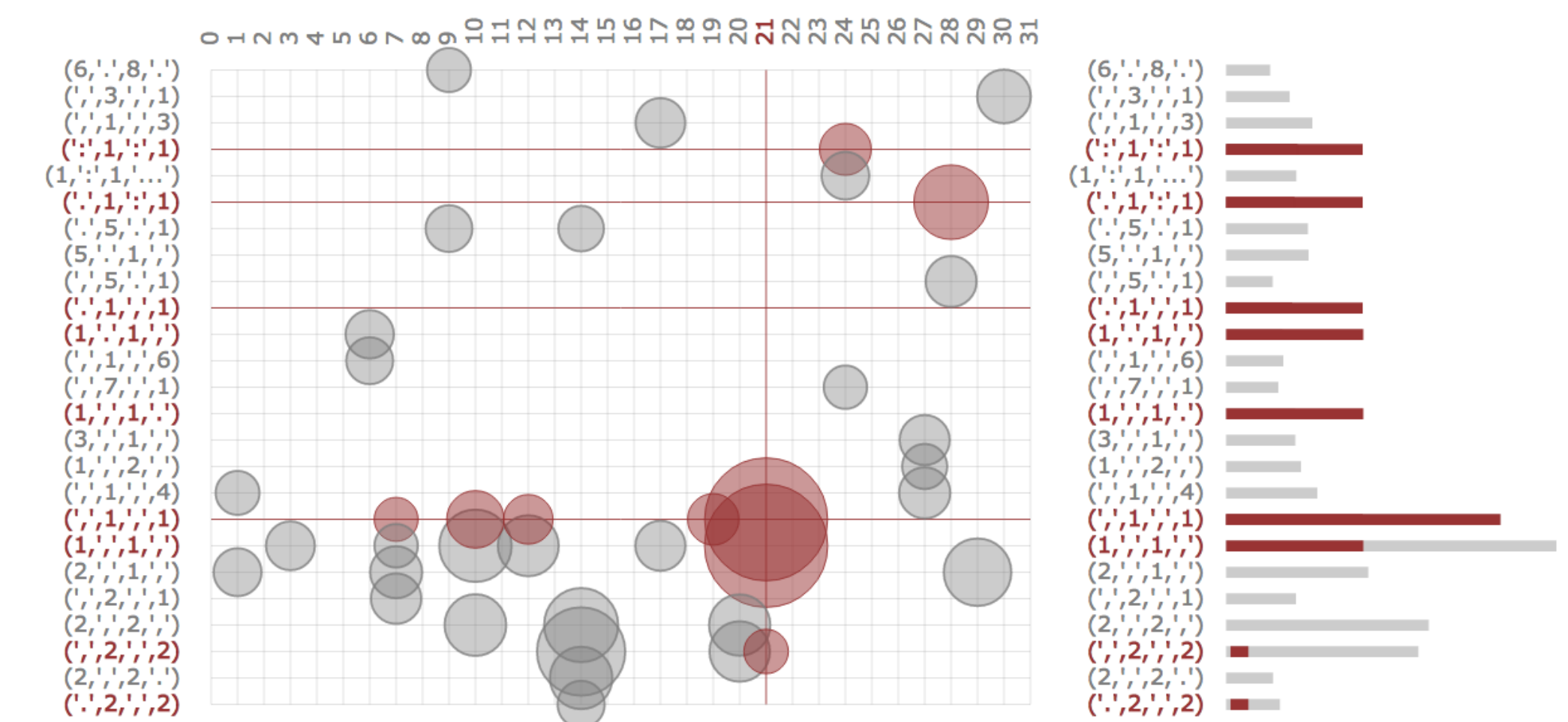
MAD works on n -grams of stylistic features, such as part-of-speech tags and syllable counts.



Feature extraction for the MAD Topic Model. Word and syllable counts (between punctuation) were also included.

Visualization

The MAD Topic Model generates topics over n -grams of stylistic features, which in turn highlights the underlying stylistic structure of the documents.



A topic model for word count (between punctuation) n -grams. Topic 21 represents short, staccato sentences. Graphic generated with the Termite tool [1].

Conclusion

- The MAD Topic Model underperformed compared to alternative classification schemes.
- However, MAD remains useful as a means of generating topic models over stylistic features and discovering the hidden structure of natural language.
- MAD's performance suggests that the topic model analogy may not be appropriate for stylistic features, at least in the classification setting.

Acknowledgements

Thanks to Chong Wang, whose sLDA codebase served as a starting point for the MAD Topic Model, as well as to David Blei for taking the time to consult with us on potential methods.

References

- J. Chuang, C. D. Manning, and J. Heer. Termite: Visualization Techniques for Assessing Textual Topic Models. In *Advanced Visual Interfaces*, 2012.
- C. Wang, D. Blei, and F.-F. Li. Simultaneous Image Classification and Annotation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1903–1910. IEEE, 2009.

Results

Preliminary results show that MAD fares worse than alternative classification schemes. However, MAD's topic models can be used to extract compact representations of stylistic tendency.

