

# Utilizing Spatial Constraints in Topic Modeling

## CMSC 828G Project Proposal

Roman Stanchak  
roman@cs.umd.edu

Target publication venue: *NIPS? Suggestions?*

## 1 Introduction

Traditional topic modeling methods generally assume a 'bag of features' model in which there is no notion of order, location or distance associated with each *feature* in the model. In the real world, order plays a large role in defining the context in which features can appear. For instance, the meaning of a textual document is lost if the words are randomly permuted. Similarly, an image of a human face ceases to be recognizable if the relative positions of the eyes, nose and mouth are scrambled. Although topic models have seen much success in their application to various domains, these small examples suggest that there is something to be gained by considering the spatial relationships between features.

## 2 Related Work

Spatial relationships have been considered in the context of topic modeling. Wallach[8] describes a bi-gram model for text corpora, Griffiths and Steyvers [3] describe LDA Collocation which also considers bi-grams, and Wang extends their results for N-grams [10]. These works consider only the sequential order of words, and do not explicitly extend to multi-dimensional spaces. Topic modeling has been considered in the Computer Vision literature for the problem of general object recognition in images. A number of researchers have considered how to integrate spatial relationships with topic models in this context. Some methods are ad-hoc [5],[6], while others incorporate spatial location directly into the underlying generative topic model [9],[7]. The goal of these works is object recognition from images, so evaluation is in terms of recognition rate. In each work, slightly different features and codebook generation methods are used, and typically only one or two im-

age databases are evaluated. As a result, it is difficult to assess the relative effectiveness of the addition of spatial information to the topic modeling.

LabelMe is a web-application that allows casual users to upload photos and then annotate them with tagged polygons [?]. At the time of writing, the full database consists of 162993 images, 43182 of which have at least one annotation. A subset of the database is available for benchmarking purposes. The training set contains 2920 images, 32164 total annotations and 127 unique tags. The test set contains 1033 images and 32853 total annotations.

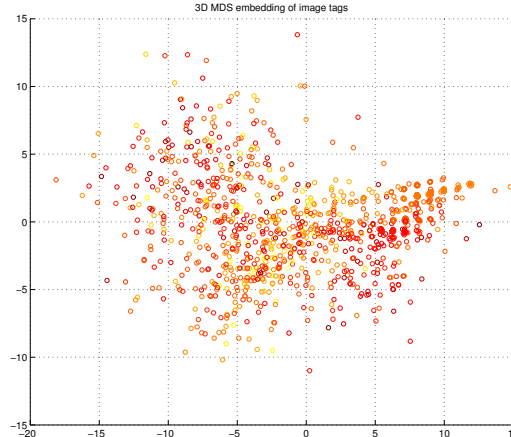


Figure 1: MDS embedding of images according to similarity in occurring tags. The metric used to create the embedding was the L1 distance between the binary tag feature vectors.

## 3 Methods

### 3.1 Algorithms

- Latent Dirichlet Allocation
- Author Topic Model

### 3.2 Evaluation

evaluated in terms of (a) document modeling and (b) document classification. Following Hofmann [4], and Blei [2], document modeling will be evaluated in terms of the *perplexity* measure on a held out test set. Similarly, document classification will be evaluated using standard precision-recall on a held out test set.

## References

- [1] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.
- [2] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation, 2003.
- [3] TL Griffiths, M Steyvers, DM Blei, and JB Tenenbaum. Integrating topics and syntax. In *Advances in Neural Information Processing Systems 17*, 2005.
- [4] Thomas Hofmann. Probabilistic latent semantic analysis. In *Proc. of Uncertainty in Artificial Intelligence, UAI'99*, Stockholm, 1999.
- [5] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *cvpr*, 02:2169–2178, 2006.
- [6] Bryan C. Russell, William T. Freeman, Alexei A. Efros, Josef Sivic, and Andrew Zisserman. Using multiple segmentations to discover objects and their extent in image collections. *cvpr*, 02:1605–1614, 2006.
- [7] Erik B. Sudderth, Antonio Torralba, William T. Freeman, and Alan S. Willsky. Learning hierarchical models of scenes, objects, and parts. *iccv*, 2:1331–1338, 2005.
- [8] Hanna M. Wallach. Topic modeling: beyond bag-of-words. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, pages 977–984, New York, NY, USA, 2006. ACM.
- [9] Xiaogang Wang and Eric Grimson. Spatial latent dirichlet allocation. In *Proceedings of Neural Information Processing Systems Conference*, 2007.
- [10] Xuerui Wang and Andrew McCallum. A note on topical n-grams. Technical Report UM-CS-2005-071, University of Massachusetts Amherst, Amherst, MA, 2005.