Discovering visual phenomena at multiple scales using the hierarchical Latent Dirichlet Allocation

Genevieve Flaspohler

Abstract

Marine robots operate in the most communication constrained environment on earth. These challenging conditions require marine robots to display high levels of autonomy. Semantic understanding is an essential precursor to meaningful autonomy. Throughout the scientific community, hierarchical trees appear as a powerful and concise manner of representing complex semantic knowledge. In this work, I apply the nonparametric hierarchical Latent Dirichlet Allocation (hLDA) model to learn a flexible hierarchy over image data collected by the Jaguar AUV during a mission off the Hannibal Sea Mount, Panama to discover hierarchical semantic associations among the observed visual data.

1 The Hierarchical Latent Dirichlet Allocation model

The Hierarchical Latent Dirichlet Allocation (hLDA) model was introduced by Blei et al. [2010]. Compared to a flat LDA topic model, hLDA is able to model a richer class of hierarchical relationships among topics, while maintaining the nonparametric properties of HDP topic models. The foundation of this model is the nested Chinese Restaurant Process (nCRP). The nCRP is nonparametric prior over infinitely deep, infinitely branching trees. Each node k in this infinite tree represents a topic β_k, and each document is defined by a path down the tree $\mathbf{c}_d$. Each path starts with the most general topic at the tree root. Therefore, each document shares this most general topic. The topics in the leaf nodes will be shared by a much more restricted subset of documents, and represent more specific semantic constructs. The full generative model for hLDA presented in Blei et al. [2010] is reproduced in the Appendix.

1.1 Gibbs Sampling Inference

Blei et al. [2010] introduces a collapsed Gibbs sampler to do inference in the hLDA model. The topic parameters β_k and the per-document topic proportions θ_d, are integrated out to speed up the chain's convergence. For each iteration of the sampler, the per-document paths $\mathbf{c}_d$ and the per-word level allocations to topics in those paths $z_{d,n}$ are sampled from the equations reproduced in the Appendix.

1.2 Code implementation

Blei et al. [2010] provides a C implementation of the hLDA model. The code is flexible enough to work directly on appropriately discretized image data from a marine robotic mission. The documentation for this code is fairly poor, so I have spent a good amount of time understanding how to run the code and interpret and visualize the output. I then tested Blei's implementation on downloaded articles from the Associated Press and wrote my own generative model for hLDA in order to test the code on simulated data. The results of these experiments are discussed in Section 2.

2 Data experiments

2.1 Associated Press abstracts

I applied the hLDA code to discover hierarchical structure in Associated Press news articles. In Blei et al. [2010], results are presented for the full dataset of 2246 documents. However, after running the code for three days, it had only completed 1,000 of the ten thousand suggested epochs. Instead, I tested the algorithm on a 500 document subset of these articles. The highest-likelihood recovered hierarchical topic structure is visualized in Figure 1 (B).

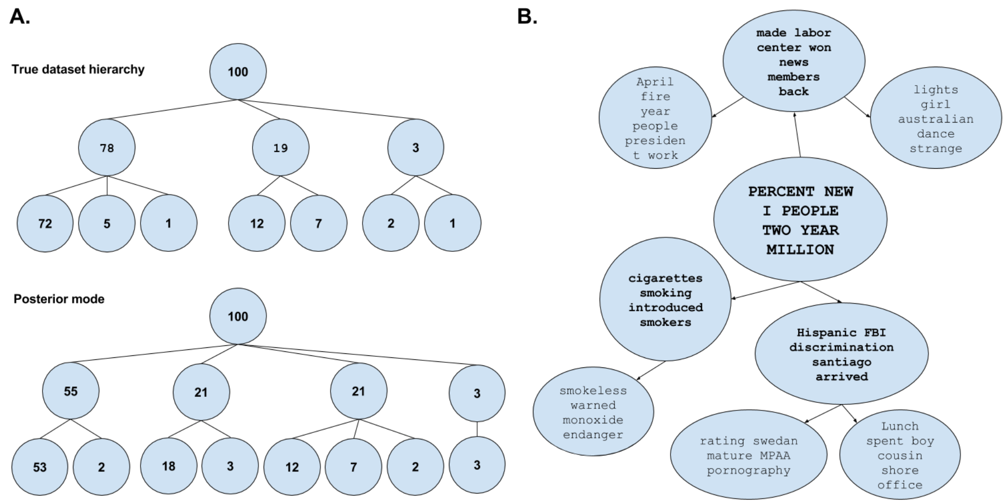

Figure 1: The results of running hLDA posterior inference on 100 simulated documents and 500 Associated Press news articles. A) The performance of the algorithm on 100 simulated documents, 250 words each. Each node in the tree represents a topic, and the number of documents containing that topic is shown directly on the node. The true topic hierarchy (top) was always smaller than the trees recovered by posterior inference (bottom). B) The three-level topic hierarchy recovered from 500 Associated Press news articles.

2.2 Simulated generative model

Blei et al. [2010] also presents the results of using the described inference procedure to recover model parameters on data simulated from the generative model. With hyperparameter settings $\eta = 0.005$ and $\gamma = 1$, they generate 100 documents of 250 words each. In the simulations, the stick breaking procedure, which determines the depth of the tree, was truncated after three levels; arbitrary branching factor was allowed. To reproduce the results, I implemented the generative model described[1] and ran it with the same hyperparameter settings. After 10,000 rounds of Gibbs sampling, the tree that produced the highest log-likelihood (the "mode") was chosen as the true tree, as described in the paper. However, unlike the results presented in the paper, the provided hLDA code does not seem to recover the true tree structure well, even after five cross-validation runs. An example of the true simulated tree structure and the structure discovered using inference is shown in Figure 1 (A).

3 Remaining action items

Despite the algorithm's failure on simulated data, it appears to recover a reasonable topic hierarchy from the Associated Press articles. It is possible that there is a mistake or misinterpretation in my simulated model, or that the authors in Blei et al. [2010] used some other methods not described in their paper to infer the simulated models with such high accuracy. I will continue to adjust the simulation, but meanwhile I will move ahead to applying the model on image data.

In Sivic et al., the hLDA model is applied to images to recover a semantic hierarchy among simple classes of objects (e.g. cars from several vantage points, stoplights, computer screens). I will follow the word-extraction process introduced in this paper (SIFT features at multiple scales) to convert my image data into discrete words, which can then be fed directly into the hLDA model. One potential issue raised by Sivic et al. is that initialization has a large impact on the optimality of the tree structure recovered. They use a heuristic that assigns a good initial guess to the level of abstraction of visual words using properties of the image. Given time, I will also attempt to also explore the effect of initialization on the recovered model.

[1]https://github.com/geflaspohler/hlda/tree/master/generative

References

David M Blei, Thomas L Griffiths, Michael I Jordan, D M Blei, and M I Jordan. The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies. *J. ACM Journal of the ACM*, 57(7), 2010. doi: 10.1145/1667053.1667056. URL `https://cocosci.berkeley.edu/tom/papers/ncrp.pdf`.

Josef Sivic, Bryan C Russell, Andrew Zisserman, William T Freeman, and Alexei A Efros. Unsupervised Discovery of Visual Object Class Hierarchies. URL `http://people.csail.mit.edu/billf/publications/Unsupervised{_}Discovery.pdf`.

Appendix

Generative model

The Hierarchical Latent Dirichlet Allocation (hLDA) model defines the following generative process for documents

(1) For each table $k \in \mathcal{T}$ in the infinite tree,
   (a) Draw a topic $\beta_k \sim$ Dirichlet(η)
(2) For each document, $d \in \{1, 2, \ldots, D\}$
   (a) Draw a path through the tree, $\mathbf{c}_d \sim$ nCRP(γ).
   (b) Draw a distribution over levels in the tree, $\theta_d|\{m, \pi\} \sim$ GEM(m, π)
   (c) For each word,
      i. Choose level $Z_{d,n}|\theta_d \sim$ Discrete(θ_d)
      ii. Choose word $W_{d,n}|\{z_{d,n}, \mathbf{c}_d, \beta\} \sim$ Discrete$(\beta_{\mathbf{c}_d}[z_{d,n}])$, which is parameterized by the topic in position $z_{d,n}$ on the path $\mathbf{c}_d$.

Gibbs Sampling Updates

The Gibbs sampler described in Blei et al. [2010] iteratively samples the per-document path assignment variables and the per-word level assignment variables.

(1) For each document $d \in \{1, \ldots, D\}$
   (a) Randomly draw $\mathbf{c}_d^{(t+1)}$ from Eq. 2
   (b) Randomly draw $z_{n,d}^{(t+1)}$ from Eq. 1 for each word, $n \in \{1, \ldots, N_d\}$

Sampling Level Allocations

The posterior level assignment for word n in document d can be written as:

$$
\begin{aligned}
P(z_{d,n}|\mathbf{z}_{-(d,n)}, \mathbf{c}, \mathbf{w}, m, \pi, \eta) &\propto P(z_{d,n}|\mathbf{z}_{d,-n}, m, \pi) P(w_{d,n}|\mathbf{z}, \mathbf{c}, \mathbf{w}_{-(d,n)}, \eta) \\
P(z_{d,n}|\mathbf{z}_{d,-n}, m, \pi) &= \frac{m\pi + \#[\mathbf{z}_{d,-n} = k]}{\pi + \#[\mathbf{z}_{d,-n} \geq k]} \prod_{j=1}^{k-1} \frac{(1-m)\pi + \#[\mathbf{z}_{d,-n} > j]}{\pi + \#[\mathbf{z}_{d,-n} \geq j]} \qquad (1) \\
P(w_{d,n}|\mathbf{z}, \mathbf{c}, \mathbf{w}_{-(d,n)}, \eta) &\propto \#[\mathbf{z}_{-(d,n)} = z_{d,n}, \mathbf{c}_{z_{d,n}} = c_{d,z_{d,n}}, \mathbf{w}_{-(d,n)} = w_{d,n}] + \eta
\end{aligned}
$$

Sampling Paths

The posterior level assignment for word n in document d can be written as:

$$
\begin{aligned}
P(\mathbf{c}_d|\mathbf{w}, \mathbf{c}_{-d}, \mathbf{z}, \eta, \gamma) &\propto P(\mathbf{c}_d|\mathbf{c}_{-d}, \gamma) P(\mathbf{w}_d|\mathbf{c}, \mathbf{w}_{-d}, \mathbf{z}, \eta) \\
P(\mathbf{c}_d|\mathbf{c}_{-d}, \gamma) &\text{is given by the nCRP prior} \qquad (2) \\
P(\mathbf{w}_d|\mathbf{c}, \mathbf{w}_{-d}, \mathbf{z}, \eta) &\propto \frac{\prod_w \Gamma(\#[\mathbf{z} = l, \mathbf{c}_l = c_{d,l}, \mathbf{w} = w] + \eta}{\Gamma(\sum_w \#[\mathbf{z} = l, \mathbf{c}_l = c_{d,l}, \mathbf{w} = w] + V\eta)}
\end{aligned}
$$