

LDA in the Wild: How Practitioners Develop Topic Models

Theo Bayard de Volo, Alfredo Gomez, Tatsuki Kuze, and Alexandra Schofield

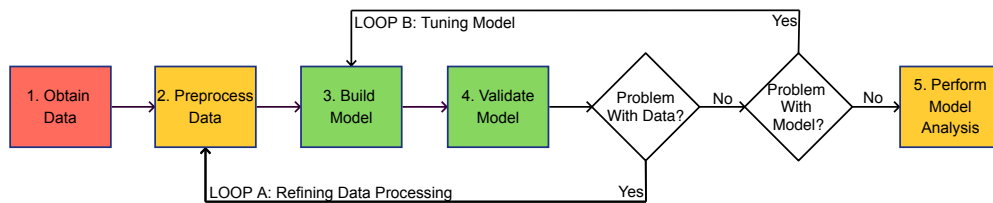


Figure 1: A unified workflow matching the descriptions of all interviewed participants.

Probabilistic topic modeling has exciting applications in humanities and social sciences, but researchers in these fields often encounter a steep learning curve to understand and adapt topic modeling tools for their work. While existing tools provide insightful visualizations [1, 2, 3, 4, 5], they often rely on an abstract idea of how topic models research progresses that may diverge from researcher experience. To better understand how to design probabilistic topic modeling tools for non-experts, we interviewed 15 individuals who use topic models for research, specifically emphasizing people whose home field is outside of computer science. From these interviews, we have developed novel design insights, which we are currently implementing in a new version of the jsLDA topic model tool [5].

Methods To develop cognitive task analysis (CTA) [6] for topic modeling projects, interviews with our participants were split into two parts. In the first part, participants broke down their task first into phases, then into *cognitive work requirements*, or tasks and questions that needed to be resolved to finish a phase [7]. In the second part, scholars addressed broader questions about their workflows such as how they found data, what tools they preferred to use, and what steps they would take when their intuitions and a model disagreed.

Preliminary Results We list a few common themes from our interviews. These will contribute to ongoing work both towards a full cognitive work analysis model for LDA research and

towards implementation of specific design recommendations in jsLDA [5].

- **LDA research follows a standard workflow.** The structure diagrammed in Figure 1 summarizes the common parts of participants’ workflows. All participants emphasized that the central steps are iterative, with trained models feeding into decisions about both data cleaning and model parameters. This contrasts with existing tools that often assume a linear progression through these steps.
- **Some phases vary more between projects than others.** In Figure 1, we use green, yellow, and red coloring to indicate variance from least to greatest respectively. This has important implications for LDA tools, as tools supporting phases with less variance can more easily support a wider range of research.
- **Practitioners value a common set of qualities in their tools.** We summarize three properties of good LDA tools: they are straightforward, continuous, and flexible. *Straightforward* tools clearly indicate how a user should expect each interaction to affect the text, model, and/or visualizations. *Continuous* tools allow for easy transition between subtasks and between other tools. Finally, *flexible* tools allow for higher levels of customization and control.

Acknowledgments This work was funded by Harvey Mudd College via a Sprague grant.

References

- [1] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher, “Serendip: Topic model-driven visual exploration of text corpora,” in *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 173–182, Oct. 2014.
- [2] C. Sievert, “cpsievert/LDAvis,” May 2020. original-date: 2014-03-05T06:17:16Z.
- [3] A. Ganesan, K. Branley, S. Pan, and J. Chen, “LDAExplore: Visualizing Topic Models Generated Using Latent Dirichlet Allocation,” in *Proceedings of the TextVis Workshop - Intelligent User Interfaces (IUI)*, p. 7, 2015.
- [4] Steffen Pielström, Severin Simmler, Thorsten Vitt, and Fotis Jannidis, “A graphical user interface for LDA topic modeling – DH2018.” Library Catalog: dh2018.adho.org.
- [5] D. Mimno, “mimno/jsLDA,” May 2020. original-date: 2013-04-23T14:37:17Z.
- [6] J. M. Schraagen, S. F. Chipman, and V. L. Shalin, *Cognitive task analysis*. Psychology Press, 2000.
- [7] W. C. Elm, S. S. Potter, J. W. Gualtieri, E. M. Roth, and J. R. Easter, “Applied Cognitive Work Analysis: A Pragmatic Methodology for Designing Revolutionary Cognitive Affordances,” *Handbook of cognitive task design*, pp. 357–382, 2003.