ABSTRACT

Title of Dissertation: Improving Knowledge Discovery:
Advanced Topic and Language Models

Kyle Seelman
Doctor of Philosophy, 2025

Dissertation Directed by: Professor Jordan Boyd-Graber
Department of Computer Science

Knowledge discovery in textual data is a cornerstone of natural language processing (NLP), driving innovations that enable machines to uncover, interpret, and interact with human language in unprecedented ways. Topic modeling is a popular method for distilling vast corpora into comprehensible themes. In parallel, the large language models (LLM) have revolutionized NLP, offering versatility and power that encapsulates world knowledge.

We improve traditional topic modeling techniques by harnessing the capabilities of LLM. We refine and extend the utility of time-tested models while simultaneously exploring the potential of LLM to perform complex language tasks traditionally reliant on tailor-made models. In doing so, this work addresses the balance between the interpretability and accessibility of topic modeling and the broad yet sometimes opaque knowledge within LLM.

The core proposition of this research is twofold: first, it introduces enhancements to topic modeling methodologies that increase their adaptability and user engagement, thereby improving the granularity and relevance of extracted topics in an age dominated by LLM. Second, it demonstrates

Chapter 6: Conclusion

This dissertation has advanced the study of topic modeling and predictive language systems through a series of contributions that place human interaction and interpretability at the center of model design. Across probabilistic, neural, and large language model–based approaches, the research demonstrates how incorporating human knowledge, lightweight interventions, and contextual priors can yield systems that are both more effective and more aligned with practical use cases.

6.1 Summary of Contributions

The first major contribution was interactive neural topic modeling (I-NTM), which introduced topic embeddings as moveable embeddings that users can manipulate directly through easy labeling. Unlike earlier interactive models for probabilistic models that involved many constraints, I-NTM allows user guidance to reshape the latent space. Experiments showed that even minimal feedback, such as a single labeled word per topic, can improve topic alignment with human expectations. Downstream retrieval experiments further demonstrated that these refinements improve both relevance and efficiency compared to static models.

The second contribution was Archivist, a hybrid topic modeling framework that integrates BERT-based priors into collapsed Gibbs sampling. By injecting contextual knowledge into probabilistic inference, Archivist achieved more coherent and stable topics across corpora of varying size. Evaluations on perplexity, coherence, topic stability, and word intrusion tasks confirmed that Archivist outperforms standard LDA and is competitive with, or superior to, neural baselines such as BERTopic—particularly in low-resource settings where finding coherent topics early is critical.

The third contribution extended this human-in-the-loop philosophy to predictive language modeling. Using LLaMA for personality prediction, the dissertation examined both few-shot prompting and parameter-efficient fine-tuning via LoRA. Results showed improvements in difficult traits, such as neuroticism facets, though the fine-tuned models tended toward normalized output distributions. This raised a key methodological question: should models be optimized to match self-reported labels, or should they instead aim to model underlying dispositions more faithfully? This tension highlights the need for psychologist-in-the-loop feedback and more principled training objectives, such as RLHF, in future work.

A unifying theme of these contributions is the integration of human knowledge and constraints into machine learning systems. At the same time, several limitations persist. Archivist depends on pretrained language models whose biases shape priors in ways not always transparent. I-NTM provides greater flexibility but raises computational challenges when applied to large-scale or streaming corpora. The personality prediction experiments underscored both the promise of LLMs and the difficulty of treating noisy, self-reported psychological labels as gold truth.

6.2 Future Directions

This work opens a number of directions for future research:

Scaling Interactive Models. Extending I-NTM to streaming, multilingual, and multimodal corpora, while incorporating active learning to minimize annotation effort.

Deeper Probabilistic–LLM Integration. Beyond priors, probabilistic inference could be co-trained or periodically updated with embeddings, closing the loop between contextual knowledge and sampling.

Psychologist-in-the-Loop Fine-Tuning. Future predictive systems should explore RLHF guided by expert annotations, targeting constructs of actual psychological interest rather than self-report artifacts.

Human-Centered Evaluation. Building on this dissertation's combination of automatic metrics, LLM-based evaluators, and user studies, future work should move toward standardized interactive evaluation protocols that reflect real-world use.

6.3 Closing Statement

Together, these contributions push topic modeling and predictive language research toward systems that are not only technically strong but also interpretable, interactive, and human-aligned. By demonstrating how contextual embeddings, user constraints, and expert-informed fine-tuning can reshape probabilistic and neural models, this work underscores that the next advances in NLP will come not just from larger models, but from centering human needs and agency within the modeling loop.