Krishnapriya Vishnubhotla

Room 276, D.L. Pratt Building, 6 King's College Road, Toronto, Ontario – M5S 3H5

vkpriya@cs.toronto.edu • https://priya22.github.io/

I am a third-year PhD student in the Computational Linguistics group at the University of Toronto, supervised by **Prof. Graeme Hirst**, and co-supervised by **Prof. Frank Rudzicz**. My main research area is **Natural Language Processing**. My expertise is in developing datasets and models for the computational analysis of literary texts, quantifying variations in style and meaning, and generative models of text. I am particularly interested in the stylistic patterns of similarity in the writing styles of people and communities. My projects computationally model this variation in (a) literary texts, and (b) social media data, using predictive and generative machine learning models.

Education

- **University of Toronto** — **Toronto**
  PhD in Computer Science, Computational Linguistics Group, GPA: 3.97/4.0 — *2019–Present*
- **University of Toronto** — **Toronto**
  Master of Science in Computer Science, Thesis option, GPA: 4.0/4.0 — *2017–2019*
- **National Institute of Technology Karnataka-Surathkal** — **Mangalore, India**
  B.Tech Computer Science and Engineering , CGPA 8.93/10 — *2013–2017*
- **K.V.I.I.Sc** — **Bangalore, India**
  Primary, Secondary and High School (Grades 1 to 12), 97% AISSCE — *2001–2013*

Work Experience

- **University of Toronto** — **Toronto**
  Teaching Assistant — *September 2017–Present*
  My duties include preparing tutorials, quizzes and taking lab hours.
  - CSC108: Introduction To Computer Programming.
  - CSC148: Introduction to Computer Science
  - CSC309: Programming on the Web.
  - CSC401/2511: Natural Language Computing.
- **Georgian Partners** — **Toronto**
  Research Intern — *May 2020 – Dec 2020*
  As a research intern at Georgian, I worked on developing unsupervised clustering models of text embeddings for internal company applications.
- **Samsung AI Research Center** — **Toronto**
  Research Intern — *May 2019 – September 2019*
  As a research intern at Samsung AI, I worked on multi-modal representation learning and semi-supervised methods of text and video alignment. I was a part of the winning submission for the Samsung Retail Robot Challenge, for which we built an interactive clip retrieval system for customer support videos.

Research Projects

- **Quotation Attribution and Character Voice in Literary Texts (Sept 2019-Current)**
  In this thesis project, I am exploring models that can capture the stylistic variation across characters and authors in literary novels. This partly involved designing neural models that can accurately attribute quotations within texts to their speakers. Current outcomes: a web-platform for annotating quotations and related information; a dataset for quotation attribution; a BERT–based ensemble speaker prediction model.
- **Measuring emotion granularity in Twitter data (Jan 2022-Current)**
  Developing computational metrics to measure the level of specificity of emotion expression in social media data,

particularly tweets. These metrics will be used to evaluate correlations between emotion granularity and the mental, physical and emotional health of populations.

- **Characterizing Emotion Dynamics (Oct 2021-Current)**

  Modeling variation in the emotional states of speakers across time. This work builds on previously-proposed metrics to measure the emotional state of a speaker using natural language utterances. We focus on adapting these metrics to (a) the literary domain, and (b) a dataset of temporally-distributed tweets.

- **A Dataset of Semantic Textual Relatedness (March 2020-Jan 2021)**

  Created a dataset of sentence pairs annotated for semantic relatedness using Best–Worst Scaling. Explored the contribution of various linguistic features to semantic relatedness, and evaluated state-of-the-art sentence representation models on the dataset.

- **Disentangling Content and Style in Texts (Sept 2019-Jan 2020)**

  Evaluated autoencoder variants that learn disentangled representations of content and style on a highly-structured Natural Language Generation dataset. Our findings (published) highlight the limitations of current learning methods.

- **Masters Thesis Project (2018-2019):** *'A Stylometric Investigation of Character Voices in Literary Fiction'*

  Investigated the stylistic characteristics of character dialogue in a corpus of modern-era plays. We demonstrate the effectiveness of Bayesian lexical models combined with word embeddings on this task, and uncover some interesting commonalities in character voices across authors and texts. We further use our stylistic features to build a semi-supervised quote attribution algorithm that achieves performance comparable to the state-of-the-art, while requiring lesser manual annotation.

- **Learning Dicrete Latent Structure: Course Project (Winter 2018):** *'GANs for Text Generation using word2vec'*

  Developed a text generation model using Generative Adversarial Networks (GANs), with word2vec embeddings as the input and output of the system. This bypasses the problem of differentiating through a discrete space, i.e, words. A conditional variant of the model was also implemented.

Publications

Working Papers

- Abdalla, M., Vishnubhotla, K. and Mohammad, S.M., 2021. What Makes Sentences Semantically Related: A Textual Relatedness Dataset and Empirical Study. *arXiv preprint arXiv:2110.04845.*

Refereed Publications

To Appear:

- Vishnubhotla, K. and Mohammad, S.M., 2022. Tweet Emotion Dynamics: Emotion Word Usage in Tweets from US and Canada. *To appear in the Proceedings of the 13th Language Resources and Evaluation Conference 2022*
- Vishnubhotla, K., Hammond, A. and Hirst, G., 2022. The Project Dialogism Novel Corpus: A Dataset for Quotation Attribution in Literary Texts. *To appear in the Proceedings of the 13th Language Resources and Evaluation Conference 2022*
- Hammond, A., Vishnubhotla, K., and Hirst, G., 2022. Voices Speaking To and About One Another: Introducing the Project Dialogism Novel Corpus. *Proceedings of the Digital Humanities Conference 2022*

Published:

- Vishnubhotla, K., Hirst, G. and Rudzicz, F., 2021, August. An Evaluation of Disentangled Representation Learning for Texts. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 1939-1951).
- Hammond, A., Hirst, G. and Vishnubhotla, K., 2020. The Words Themselves: A Content-based Approach to Quote Attribution. *Digital Humanities 2020*
- Vishnubhotla, K., Hammond, A. and Hirst, G., 2019, June. Are Fictional Voices Distinguishable? Classifying Character Voices in Modern Drama. In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature 2019* (pp. 29-34).
- Budhkar, A., Vishnubhotla, K., Hossain, S. and Rudzicz, F., 2019, August. Generative Adversarial Networks for Text Using Word2vec Intermediaries. In *Proceedings of the 4th Workshop on Representation Learning for NLP 2019* (pp. 15-26).

- Jeblee, S., Budhkar, A., Milic, S., Pinto, J., Pou-Prom, C., Vishnubhotla, K., Hirst, G. and Rudzicz, F., 2018. TorontoCL at the CLEF 2018 eHealth Challenge Task. In *CLEF 2018 Evaluation Labs and Workshop: Online Working Notes, CEUR-WS, September 2018*.

Academic Service

- I have served as a reviewer for:
  - ACL Rolling Review: 2022, 2021
  - *ACL Conferences (ACL, NAACL): 2020, 2019, 2018
  - EMNLP: 2020, 2019
- Volunteer mentor for the Graduate Application Assistance Program, 2021.
- Served as a Triager for the DCS Admissions Program in 2020.
- Maintained the official webpage of the Computational Linguistics group from 2018-2020.

Relevant Courses

- (Advanced) Computational Linguistics
- Natural Language Computing
- Introduction to Machine Learning
- Learning Discrete Latent Structure
- Algorithms for Private Data Analysis
- Topics in Computational Social Science
- Computational Models of Semantic Change

Technical skills

- **Machine Learning Frameworks:** PyTorch, TensorFlow, FastAI, HuggingFace
- **Programming Languages:** Proficient in: C, C++, Python, Javascript, Ruby
  Basic ability with: Java.