Victor Prokhorov

https://victorprokhorov.github.io | victorprokhorov91@gmail.com | (+44)07818927970

Skills

Programming Python, PyTorch (see: pps-vae), TensorFlow (see: hsvae, kl-text-vae, text2path)

Research Communication communicated my research via publications at: ICML, AAAI, NAACL, EMNLP; communicated research for public audiences of peers: poster presentation at Microsoft Research Cambridge, UK, 2020 and poster presentation at Google NLP Summit, Zurich, Switzerland, 2019.

Work Experience

Research Scientist (NLP), Huawei, London Sep 2024 - Present

Postdoctoral Researcher, University of Edinburgh, ExLab Apr 2022 - Apr 2024

– conducted a research on deep generative models with interpretable latent representations see project [**tl;dr:** *the objective is to build an interpretable latent variable model for visual and language modalities by grounding the latent representation in the images*]

– was responsible for leading of the project (e.g., defining current priorities, managing work with an external collaborator), a mathematical formulation of a variational autoencoder model, its implementation [code] and communicating the findings via a publication [paper]

Research Intern, Montreal Institute for Learning Algorithms (MILA) Sep 2020 - Dec 2020

– conducted a research on deep generative models with symbolic latent representations for text [**tl;dr:** *instead of pretraining a model to learn distributed representation of text which has been proven useful in transfer learning setup, in the project we aimed to induce reusable, graph-based symbolic representation of text*]

– was responsible for a mathematical formulation of a variational autoencoder model and its implementation

Selected Publications

1. **Victor Prokhorov** and Ivan Titov and Siddharth N, Autoencoding Conditional Neural Processes for Representation Learning in **ICML 2024** [code] [**tl;dr:** *develops the Partial Pixel Space Variational Autoencoder that casts Conditional Neural Process (CNP) context as latent variables learnt simultaneously with the CNP*] [**keywords:** generative models; variational autoencoders; conditional neural processes; representation learning; interpretability; vision]

2. Mattia Opper and **Victor Prokhorov** and Siddharth N, StrAE: Autoencoding for Pre-Trained Embeddings using Explicit Structure in **EMNLP 2023** (main) [**tl;dr:** *investigates how explicit structure affects embedding quality*] [**keywords:** autoencoders; representation learning; interpretability; structure induction; NLP]

3. **Victor Prokhorov** and Yingzhen Li and Ehsan Shareghi and Nigel Collier, Learning Sparse Sentence Encoding without Supervision: An Exploration of Sparsity in Variational Autoencoders in **RepL4NLP 2021** (workshop) [code] [**tl;dr:** *presents a novel VAE model with sparse latent representations for text representation and generation*] [**keywords:** generative models; variational autoencoders; sparsity; representation learning; interpretability; NLP]

4. Lan Zhang and **Victor Prokhorov** and Ehsan Shareghi, Unsupervised Representation Disentanglement of Text: An Evaluation on Synthetic Datasets in **RepL4NLP 2021** (workshop) [code] [**tl;dr:** *uses the existing disentanglement VAE models to study the challenges of disentanglement for text*] [**keywords:** generative models; variational autoencoder; disentanglement; representation learning; NLP]

5. **Victor Prokhorov** and Ehsan Shareghi and Yingzhen Li and Mohammad Taher Pilehvar and Nigel Collier, On the Importance of the Kullback-Leibler Divergence Term in Variational Autoencoders for Text Generation in **WNGT 2019** (workshop) [code] [**tl;dr:** *uses an information-theoretic framework to study an effect of the rate and distortion on text generation*] [**keywords:** generative models; variational autoencoder; representation learning; model analysis; information theory; NLP]

6. **Victor Prokhorov**, Mohammad Taher Pilehvar, Dimitri Kartsaklis, Pietro Lio, Nigel Collier, Unseen Word Representation by Aligning Heterogeneous Lexical Semantic Space in **AAAI 2019** [**tl;dr:** *uses knowledge graphs to learn representations for rare and out-of-vocabulary words*][**keywords:** representation learning; knowledge graphs; NLP]

7. **Victor Prokhorov**, Mohammad Taher Pilehvar, Nigel Collier, Generating knowledge graph paths from textual definitions using sequence-to-sequence models in **NAACL 2019** [code] [**tl;dr:** *sequence-to-sequence model that grounds sentence meaning in a knowledge graph*] [**keywords:** representation learning; interpretability; knowledge graphs; NLP]

Education

2017 - 2022 PhD in Computation, Cognition and Language at **University of Cambridge**
- *Ph.D. Thesis:* Injecting Inductive Biases into Distributed Representations of Text, under supervision of Nigel Collier and Ehsan Shareghi

2016 - 2017 MPhil in Advanced Computer Science at **University of Cambridge** (Distinction)

2013 - 2016 BEng in Computer Science and Electronics at **University of Bristol** (First Class)

Honors & Awards

- Mitacs Globalink Research (Mitacs, **Awarded**; the award required in person presence in Canada; due to COVID outbreak did not accept)
- Student Travel Grant (Workshop: Neural Generation and Translation)

Activities

- Co-organiser for Edinburgh NLP Meetings, 2022-2023
- Co-organiser for Dagstuhl ELLIS NLP Workshop, 2022
- Reviewer for ICLR, 2022 and ACL ARR, 2021