

Victor Prokhorov

Selwyn College, Grange Road
Cambridge, UK
07818927970
victorprokhorov91@gmail.com

- PROFILE** I am a third year PhD student in Natural Language Processing at the University of Cambridge.
- EDUCATION** **(PhD) Computation, Cognition and Language (2017-2021):**
University of Cambridge (Language Technology Lab)
- (MPhil) Advanced Computer Science (2016-2017):**
University of Cambridge (The Computer Laboratory) - Grade: Distinction
- (BEng) Computer Science and Electronics (2013-2016):**
University of Bristol - Grade: First Class (Hons)
- Other qualifications:** Google NLP Summit (2019, Zurich, Switzerland), FoPSS logic and learning school (2018, University of Oxford)
- SKILLS** *Programming Languages:* Python (Experience: 3 years), Tensorflow + Keras (Experience: 2 years), LaTeX (Experience: 3 years), C (Experience: university projects), Java (Experience: university projects), Haskell (Experience: university projects), C# (Experience: 6 months)
Languages: English, Russian (Native Speaker)
- RESEARCH FOCUS**
- **Natural Language Processing Aspect:** I am interested in advancing the research on distributed sentence representation and its interpretability.
 - **Machine Learning Aspect:** Generative models with the focus on Variational Autoencoders. As well as combination of probabilistic graphical models and deep learning in general.
- PUBLICATIONS**
- **Victor Prokhorov**, Ehsan Shareghi, Yingzhen Li, Mohammad Taher Pilehvar, Nigel Collier (2019): ***On the Importance of the Kullback-Leibler Divergence Term in Variational Autoencoders for Text Generation***, Accepted in 3rd Workshop on Neural Generation and Translation (WNGT), 2019, Hong Kong, China.
 - **Victor Prokhorov**, Mohammad Taher Pilehvar, Nigel Collier (2019): ***Generating Knowledge Graph Paths from Textual Definitions using Sequence-to-Sequence Models***, Accepted in North American Chapter of the Association for Computational Linguistics (NAACL), 2019, Minneapolis, USA, pages: "1968–1976"
 - **Victor Prokhorov**, Mohammad Taher Pilehvar, Dimitri Kartsaklis, Pietro Lio, Nigel Collier (2019): ***Unseen Word Representation by Aligning Heterogeneous Lexical Semantic Spaces***, Accepted in Association for the Advancement for Artificial Intelligence (AAAI) 2019, Honolulu, Hawaii, USA, pages: "6900–6907"

- Mohammad Taher Pilehvar, Dimitri Kartsaklis, **Victor Prokhorov**, Nigel Collier (2018): ***Card-660: Cambridge Rare Word Dataset - a Reliable Benchmark for Infrequent Word Representation Models***, Accepted in Conference on Empirical Methods in Natural Language Processing (EMNLP) 2018, Brussels, Belgium, pages: "1391–1401"
- Rui Fan, **Victor Prokhorov**, Naim Dahnoun (2016): ***Faster-Than-Real-Time linear lane detection implementation using SoC DSP TMS320C6678***, Accepted in IEEE International Conference on Imaging Systems and Techniques (IST), 2016, Chania, Greece, pages: "306-311"

TALKS and POSTERS

- ***Generating Knowledge Graph Paths from Textual Definitions using Sequence-to-Sequence Models***: NAACL 2019 - North American Chapter of the Association for Computational Linguistics (Minneapolis, USA) [poster]
- ***Unseen Word Representation by Aligning Heterogeneous Lexical Semantic Spaces***: AAAI 2019 - Association for the Advancement for Artificial Intelligence (Honolulu, Hawaii, USA) [poster]

GRANTS

- **Student Travel Grant:**
Awarded by: 3rd Workshop on Neural Generation and Translation

PROJECTS

- **Unseen Word Representations using Vector Space Alignment** (MPhil Dissertation): In this work I put forward a technique that exploits the knowledge encoded in lexical resources to enhance the coverage of a pre-trained word embeddings vocabulary for rare and unseen words. The approach adapts graph embedding techniques as well as cross-lingual vector space transformation algorithms in order to merge lexical knowledge encoded in ontologies with that derived from corpus statistics. (*Programming language: Python*)
- **Faster-Than-Real-Time linear lane detection implementation using SoC DSP TMS320C6678** (BEng Dissertation): This work mainly focuses on an optimised implementation of the linear lane detection system based on multiple image pre-processing methods and an efficient Hough transform. The algorithm was implemented on the TMS320C6678 System on Chip Digital Signal Processor. (*Programming language: C, OpenMP*)
- **2D Platformer on Unity Game Engine** (Summer Project): My role was to design an architecture of the game. (*Programming language: C#*)