

Yann Dubois

yd292@cam.ac.uk
+41 79 655 89 40
Blog: yanndubs.github.io

Ch. Aux Channey 20
1077 Servion, Switzerland
GitHub: github.com/YannDubs

Education

- 2018 – 2019 **The University of Cambridge**, Trinity College, UK
MPhil in Machine Learning and Machine Intelligence
Honours: Department of Engineering Scholarship
Coursework includes: CV, NLP, Probabilistic ML, Advanced ML, Speech Synthesis, RL, DL.
- 2016 – 2017 **The University of British Columbia (UBC)**, Canada
BSc in Machine Learning and Computer Science. Full year exchange program from EPFL.
Average Grade: 93% (A+).
- 2014 – 2017 **École Polytechnique Fédérale de Lausanne (EPFL)**, Switzerland
BSc in Bioengineering and Technologies.
Average Grades: 5.75 out of 6.
Honours: Best bachelor grades of the faculty.

Research Experience

- May 2018 – **Research Assistant, ILLC - The University of Amsterdam**, Netherlands
Oct. 2018 Supervisor: Dr. Elia Bruni.
Topic: Deep Natural Language Processing - Specialising Attention Mechanisms.
 - Improved the robustness of seq2seq models by reducing cross-parameter dependencies.
 - Enabled extrapolation to longer sequences by splitting content and position attention.
- May 2017 – **Research assistant, UBC and Sensible Building Science**, Canada
Oct. 2017 Topic: Machine Learning - Real-world multi-step forecasting of multi-dimensional time series.
 - Developed an improved ARIMA, more appropriate for multi-step forecasting.
 - Developed a gradient boosted regression augmented with clustering and ARIMA.
 - Implemented for a company predicting room occupancy to improve energy efficiency.
- Feb. 2016 – **Research assistant, Swiss Institute of Bioinformatics**, Switzerland
Sep. 2016 Topic: Computational genomics - optimising whole exome sequencing pipelines.
 - Developed a sample-efficient ranking algorithm for genetic variants.
 - The algorithm was used to analyse 183 infants suffering from severe sepsis.
- Aug. 2015 – **Summer student researcher, EPFL**, Switzerland
Sep. 2015 Topic: Neuroscience - link between stress induced glucocorticoid and behaviour.

Relevant Professional Experience

- Oct. 2017 – **Data Scientist Trainee, Grab**, Singapore
Apr. 2018 Company: Asia Pacific's largest start-up (~40 million users), mostly known for ride-hailing.
 - In charge of classifying 40'000 daily reviews in Thai, Vietnamese, Bahasa, English...

- I have put into production an algorithm which achieved a 20% F1-score improvement.
- Built a general NLP pipeline for the company: the core of the work was to train a sentence embedder in a semi-supervised, multi-task and language independent setting.

Sep 2015 – **Teaching assistant for Analysis I and II, EPFL, Switzerland.**

Jul. 2016 Real and multivariable analysis. 20-30 undergraduates.

Sep 2015 – **Teaching assistant for Biology, EPFL, Switzerland**

Jan. 2016 70-80 undergraduates.

Conference and Seminar Attendance

Aug. 2017 **Deep Reinforcement Learning Bootcamp, OpenAI and BAIR, USA**

Learning state-of-the-art Deep RL with P. Abbeel, S. Levine, A. Karpathy, J. Schulman, V. Minh.

Relevant Writing

- (*in preparation*) **Y. Dubois**, D. Hupkes, E. Bruni. Specializing Attention to Improve Robustness and Extrapolation.
- (*in preparation*) **Y. Dubois**, O. Naret, A. Borghesi, J. Fellay. Prioritization of variants from next-generation sequencing experiments with ScoVar.
- Machine Learning blog and glossary: yanndubs.github.io . Since Oct. 2017.
- K. Askelson*, S. Blain*, **Y. Dubois***, K. Jepson*, M. Kinney*, G. McCarthy*. Express Yourself: Exploring miRNA expression in HL-60 cells following ATRA-induced differentiation. Poster presented at: UBC Statistics - Methods for High Dimensional Biology. 2017 Apr 5; Vancouver, Canada.

Most Relevant Projects

June. 2018 - (*Member*) **I-machine-think**

Now Modular Pytorch implementation for seq2seq research: <https://github.com/i-machine-think>.

Jan. 2018 - (*Still developing*) **Rolling Average Word Embeddings (RAW-Embeddings)**

Feb. 2018 Fast and interpretable word embeddings: github.com/YannDubs/RAW-Embeddings.

Dec. 2017 - **Global NIPS 2017 Paper Implementation Challenge: Hash-Embeddings**

Jan. 2018 Won the challenge by successfully replicating and improving: “Hash Embeddings for Efficient Word Representations”. Code: github.com/YannDubs/Hash-Embeddings.

Skills

Descending **Programming:** Python (ML/DL/RL libraries), C++, Matlab, R, Bash , SQL.

Order **Languages:** French (native), Polish (native), English (C2), German (B2), Italian (B1).

Miscellaneous

2015– 2016 **Association LBC, Switzerland**

Co-founder and President. Association which unites students from various universities.

Aug 2016 **BCV Silicon Valley Start-up Camp – Young Entrepreneurs, USA**

Oct. 2015 **Runner, 4th junior (3h20), Lausanne’s Marathon, Switzerland**