

Daniel C. FERREIRA

daniel.ferreira.1@gmail.com

github.com/dcferreira

HIGHLIGHTS

- Strong mathematical background and experience in Machine Learning
- Experience in network security and natural language processing
- Curious and fast learner

SHORT BIO

- 2019– Data Scientist at [Cyan Security](#)
– Currently using state-of-the-art Machine Learning methods to classify websites in multiple languages.
- 2016–2019 Project Assistant in the Communications Networks group, TU Wien, integrated in the [BigDAMA project](#).
– Research topics mostly related to how to represent network traffic for detecting attacks at the network level (i.e., which features to use). Experimented with both classical features (what have people been using and why?), and feature learning approaches using Deep Learning techniques, and in particular representing traffic in 2-dimensional spaces.
- 2016 Junior Researcher at Priberam Labs, integrated in the [SUMMA project](#).
– Tackled the problem of *named entity recognition* in which, given a text with some entity mentions (e.g., David and Victoria), the goal is to find the corresponding entities in Wikipedia (e.g., David Beckham and Victoria Beckham). We used both classical methods (such as SVM) and modern Deep Learning approaches.
- 2015 MSc in [Applied Mathematics](#) at [Instituto Superior Técnico](#)
MAJOR: Computation
THESIS in cross-lingual classification (grade: 19/20)
– We wanted to classify news documents in German, given that our training data only has English documents. We proposed two novel approaches to find word representations in a bilingual space, one using CCA and another using a new formulation. Details can be found [here](#).
Advisors: [André F. T. MARTINS](#), [Mariana S. C. ALMEIDA](#), [M. Rosário OLIVEIRA](#)
- 2013 BSc in [Applied Mathematics and Computation](#) at [Instituto Superior Técnico](#)
- 2010 Completed high-school
- 2009 Certificate of Proficiency in English (CPE) by [University of Cambridge](#)
- 1992 Born in Lisbon, Portugal

RELEVANT PUBLICATIONS

- 2019 Extreme Dimensionality Reduction for Network Attack Visualization with Autoencoders (*accepted at IJCNN 2019*)
- 2017 [A meta-analysis approach for feature selection in network traffic research](#), ACM SIGCOMM 2017 Reproducibility Workshop
- 2016 [Jointly Learning to Embed and Predict with Multiple Languages](#), ACL 2016

RELEVANT COURSES

- Machine Learning [Lisbon Machine Learning School](#) (participant in 2014, and monitor in years 2015, 2016, 2017, 2018); [Deep Learning for Visual Computing](#) (2019); [DeepLearn Bilbao](#) (2017); [Statistical Learning](#) (2015); [Statistical Methods in Data Mining](#) (2014)
- Security [Network Security – Advanced Topics](#) (TU Wien, 2018); [PhD School on Traffic Monitoring and Analysis](#) (2017); [Advanced Internet Security](#) (TU Wien, 2017); [Software Security](#) (2015)

SOFTWARE EXPERIENCE

- Languages Python, R, Mathematica, Go, C, Java, Latex
- Libraries NumPy, Pandas, TensorFlow, PyTorch, Keras, Theano, scikit-learn, Scapy
- Other Linux, Docker, Git, Wireshark, Inkscape, Photoshop

OTHER

- Developed the [NTARC database](#) of network traffic research, for meta-analysis purposes.
- Participated in Capture the Flag (CTF) competitions with [WE_OWN_YOU](#)
- Developed task-oriented [multilingual word embeddings](#) – useful for text tasks involving multiple languages
- Developed [mdcgenpy](#) – a generator of random clustered data, for evaluation of clustering algorithms
- Developed [Traffic Flow Mapper](#) – a prototype to visualize network traffic flows in real time in a 2d plot