Daniel de Souza Severo

For more information: https://dsevero.com

EDUCATION

University of Toronto 2013 - 2014
Electrical & Computer Engineering
1 year Exchange Program

Federal University of Santa Catarina, Brazil 2010 - 2015
Bachelor of Science in Electronics Engineering
First Class Honours, 99th percentile.

AWARDS

Student Merit Award 2015
Graduated with the highest GPA ever obtained (at the time) for my major.

Student Merit Medal 2015
Elected "Best Student" by the faculty of Electrical & Electronics Engineering at the Federal University of Santa Catarina.

Science Without Borders Scholarship 2013
Awarded a full scholarship that covered tuition, transportation, necessary materials and living costs to study 2 academic semesters at the University of Toronto.

RESEARCH EXPERIENCE

Syrian-Lebanese Hospital 2018 - Current

Ward2ICU: A Vital Signs Dataset of Inpatients from the General Ward
We present a proxy dataset of vital signs with class labels indicating patient transitions from the ward to intensive care units called Ward2ICU. Patient privacy is protected using a Wasserstein Generative Adversarial Network to implicitly learn an approximation of the data distribution, allowing us to sample synthetic data. The quality of data generation is assessed directly on the binary classification task by comparing specificity and sensitivity of an LSTM classifier on proxy and original datasets. We initialize a discussion of unintentionally disclosing commercial sensitive information and propose a solution for a special case through class label balancing. ***arXiv:1910.00752***

Predicting Patient Health Risk through Financial Data.
Adoption of Electronic Health Records in Brazilian hospitals is slow due to administrative inertia. Most hospitals collect and organize only financial claims of patient spendings, making it difficult to apply ML techniques to predict patients with high health risk. We show that it is possible to use financial risk as a proxy for health risk and apply it to a Learning to Rank task for medical prevention. ***Writing in progress.***

Reducing Readmission Rates with Time Series Data of Patient Vital Signs.
Hospital Readmission Rates are a key indicator of efficiency as it depends on a hospital's ability to prioritize admitted patients and administer the allocation of intense-care units. Since patient prioritization must be transparent, explainability in ML applications are an important factor. Hence, traditional ML is still preferred over Neural Networks despite the latter having superior performance. We propose a middle-ground approach by using Deep Feature Synthesis to learn features from patient vital signs as inputs to highly explainable models such as Random Forests. ***Writing in progress.***

Linx Impulse 2016 - 2017

Hypothesis Testing for Competitive A/B Experiments with High Data Volume.
In the ecommerce sector, high paying customers usually demand that an A/B experiment take place. This pins competitors against each other in an online testing environment devised to measure the performance of each recommender system at once. Due to the dynamic nature of evaluation metric (e.g. click-through-rate, revenue-per-user) chosen for each test, my research focused on developing a generic statistical hypothesis test that could be used to prove the effect our product had on the website, while handling unstructured data ingestion values averaging 1TB per day. Our results showed that a specific sub-sampling strategy together with Bootstrapping significantly reduced the computational complexity sacrificing very little power. ***Publishing not authorized by employer.***

University of Toronto 2014 - 2015

A Report on the Ziggurat Method.
Pseudo-random number generators (PRNG's) are crucial in the context of simulating noise in communication channels. We present a report on an efficient method for generating pseudo-random samples from any decreasing probability distribution called the Ziggurat Method. Specifically, we will show the latest and most efficient version presented by McFarland. In the latter paper, the method shows a speedup of over 3 times compared to traditional algorithms such as Marsaglia's Polar Method. We present a speed comparison in C implemented on an Intel i7-4790 clocked at 3.60 GHz. A proof that the samples from this method are truly Gaussian is also provided. ***Source code and writing available at*** `https://github.com/dsevero/A-Report-on-the-Ziggurat-Method`

Optimal Decision Methods and Feedback in Physical Layer Network Coding.
Interference is becoming a fundamental limitation for modern wireless networks. To mitigate interference, a number of advanced signal-processing techniques have recently been proposed. One such example is Physical Layer Network Coding (PNC), which allows each relay node in the network to decode a function of transmitted messages from interfering signals. With relays decoding and forwarding these functions to a central destination, the destination can potentially recover all the transmitted messages in the network. In this way, interference can be cleverly harnessed and the network throughput can be greatly improved. Despite a large body of work on PNC, the following two questions remain unanswered: (1) How can a relay node decode a function in the most reliable way? (2) How can a relay node make use of feedback to further improve the decoding performance? Our project is motivated by these two questions. In particular, we aim to discover how to decode said function in order to minimize the probability of error detection. ***Presented at the SP Coding and Information School 2015.***

TEACHING EXPERIENCE

Federal University of Santa Catarina
Teaching Assistant
Assisted professors by ministering tutorials, preparing lecture materials and helped students individually at regular office hours.

- Communications Theory 2015
- Non-linear Electronic Circuits 2013
- Single-Variable Calculus 2010

CERTI Foundation 2010 - 2013
Intern Programming Instructor
Responsible for the technical training of new and current interns. Created a training course in LabVIEW programming that is still in use as of 2019.

PROFESSIONAL SERVICE

NeurIPS 2019: Conference on Neural Information Processing Systems
Reviewer for the Machine Learning for Health (ML4H) workshop.

OPEN SOURCE CONTRIBUTIONS

Dask: Scalable analytics in Python
`https://github.com/dask/dask/pulls?q=author:dsevero`

Dask-ML: Scalable Machine Learn with Dask
`https://github.com/dask/dask-ml/pulls?q=author:dsevero`

Generative Models: Self-teaching of Generative Models
`https://github.com/dsevero/generative-models`

PROFESSIONAL EXPERIENCE

Independent Contractor 2018 - Current
Machine Learning Engineer & Researcher
Developed a Fast Healthcare Interoperability Resources DataLake for running high volume machine learning models; Feature engineering and mathematical modeling for clustering algorithms used to segment patients into similar health groups; Ranked patients by future spendings using financial data achieving a precision at n=1,000 of 50% from a 15,000 total; Predicted patient LoS (Length of Stay) with regression techniques and hospital sensor data; Modified CoSimRank to create a similarity measure between developers and companies using Stack OverFlow data using Neo4j and Python.

Linx Impulse 2016 - 2018
Head of Data Science
Developed recommendation algorithms for E-commerce customers; Provided ad-hoc big data analyses to find insights from our data; Designed and monitored competitive A/B experiments devised to validate our systems performance in the face of competition; Internal A/B testing tool using the SciPy and Jupyter stack; Bandit algorithms for online optimization

Wavetech Technology Solutions 2015
Embedded Systems Engineering Intern.
Worked on microcontroller programming in C/C++ for cochlear implants.

CERTI Foundation 2010 - 2013 (Intern.)
2015 - 2016 (R. Eng)
Implemented signal processing routines (filter design and realization) in C; Programmed back-end and front-end Python software for Raspberry Pi; Embedded eLua on a platform previously developed by CERTI.

WEG Industries Summers 2011 and 2012
Electrical Engineering Intern.
Software upgrade, in LabVIEW, of an automatic calibrator of multimeters in order to account for different input frequencies; Conception and implementation of a hardware and software (LabVIEW) system that acquires, processes and stores data of specific parameters of electric motors.

REFERENCES

Prof. Frank R. Kschischang University of Toronto
Distinguished Professor of Digital Communication
Department of Electrical & Computer Engineering
`https://www.comm.utoronto.ca/frank/`

Prof. Estevam Rafael Hruschka Junior Carnegie Mellon University
Visiting Professor
Machine Learning Department
`http://www.cs.cmu.edu/~estevam/`

Prof. Chen Feng The University of British Columbia
Assistant Professor
School of Engineering
`https://people.ok.ubc.ca/cfeng01/index.html`

Prof. Danilo Silva Federal University of Santa Catarina
Associate Professor
Department of Electrical and Electronic Engineering
http://danilosilva.sites.ufsc.br/index.html