April 5, 2017

Dear Bombora members,

I'm writing to your recent advertisement for a Data Science Intern position that I heard from one faculty of my school. I believe I'm a strong candidate for this position.

YOUR NEEDS	MY QUALIFICATION
solid fundamentals across the mathematics that allows us to interpret the world, empirically.	I have a good mathematic foundation, including statistics and linear algebra. Besides, I'm also good at real analysis (a more advanced topic in math).
solid knowledge of natural language processing in both classical theories and deep learning approaches.	I have implemented my deep learning models for image classification and natural language processing (which shows very promising results). Besides, I have a good understanding on PAC learning, online learning and kernel methods.
In Python, Matlab, or Scala or perhaps a few other languages, (we like, R and Julia).	Fluence in Python and Scala, experience with Julia and R.
Graphical Theory + Inference	My first internship is about probabilistic graphical model and MAP inference.
Algorithms / Data Structures	Current TA for undergraduate Algorithm course, especially good at repurposing algorithm.
Functional Programming	Experience with Scala and Scheme

Thank you for your considerations. If you would like to discuss further about my qualifications, please do not hesitate to contact me via email at happyhackingyy@gmail.com or via telephone at 917-399-5754. I look forward to talking with you.

Sincerely,

Yunjian Yang

Yunjian Yang

10420 Queens Blvd, Forest Hills, NY, 11375
917-399-5754 happyhackingyy@gmail.com

Motived data science intern with statistical machine learning experience (PAC Learning, Graphical Model), deep learning experience (in visual recognition and language modeling) and solid mathematical analysis skills (real analysis). Especially good at improving and implement machine learning algorithm.

Experience

Research Assistant Intern
China Internet Network Information Center, Beijing, China (08/2015 – 02/2016)

- Improved the original Internet of Things(IoT) clustering algorithm by more than 10 percent.
- Responsible for IoT identifiers structured learning (structured prediction) algorithm by introducing the graphical model and implementing it.
- Developed 10+ components in Java to help identifier recognition and text segmentation.

Education

Courant Institute of Mathematical Science, New York University **(09/2016 – 05/2018)**
Master of Science in Computer Science **GPA: 3.90/4.00**
Course Highlights:

- Fundamental: Operating Systems, Fundamental Algorithm, Programming Language
- Machine Learning: Advanced Machine Learning, Deep Learning, Realtime Big Data Analytic

University of International Business and Economics, Beijing, China **(09/2012 – 07/2016)**
Bachelor in Management Information Systems and Finance (Dual Major) **GPA: 3.64 / 4.00**

Project

Language Modeling (course project)

- Language modeling with LSTM and GRU, achieving the state of the art result.
- Improve the original model by implementing better optimizer and initialization methods.

Semi-Supervised Learning on MNIST Dataset (course project)

- Achieved an accuracy of 99.2 percent with only 3000 labeled data, which is remarkable.
- Implemented ladder networks that improved the original semi-supervised method by 2%.

Fraud Detection in Social Networks (course project, 1st in class)

- Tech lead on developing classification algorithm. Our embarrassing simple algorithm obtained decent results on fraud detection.
- Responsible for the parallel computation part, where I leverage the flexibility of spark for computation intensive process and use MapReduce for data ETL.

Optimization on Dominating Set Problems (Project by Natural Science Foundation of China)

- Designed hybrid simulated annealing and stochastic simulation algorithm, improving the result by 5 percent than the original algorithm.

Skill

Framework: Torch (proficient), Hadoop (familiar), Spark (familiar), Cuda (familiar)

- Being able to implement different neural network structures (ConvNet, LSTM).
- Being able to write MapReduce and Spark program for data processing.
- Being able to write general purpose GPU program within Cuda framework.

Relevant Language: C++ (proficient), Python (proficient), Scala (familiar)