Mucong Ding

PERSONAL INFORMATION

Name: DING, Mucong
Date of Birth: July 17, 1996
Citizenship: China
Email: mcding@connect.ust.hk
johnding1996@hotmail.com
Personal Homepage:
http://mcding.student.ust.hk/

Mailing Address:
424T, UG Hall V,
HKUST, Clear Water Bay,
Kowloon, Hong Kong.
Google Scholar
Github (Personal)
Github (Organization)

EDUCATION

Hong Kong University of Science and Technology (HKUST), Hong Kong, China
B.Sc (double major) in **Computer Science** *and* **Mathematics** **September, 2015 - present**

Expected to graduate in June, 2019

Courses and Grades:

- 9 Computer science undergraduate-level courses
- **5 Computer science graduate-level courses**
  - Deep Learning for Computer Vision, Advanced Machine Learning, Knowledge Discovery in Databases, Advanced Computer Graphics, Parallel Programming.
- **10 Mathematics courses**
  - Honors Real Analysis I and II, Honors Linear and Abstract Algebra I and II, Probability Theory, Complex Analysis, Partial Differential Equations.
- 8 Physics undergraduate-level courses
- **4 Physics graduate-level courses**
  - Solid State Physics I, Advanced Quantum Mechanics, Statistical Mechanics I, General Relativity.

GPA: 4.137/4.3, Major GPA: 4.196/4.3, 42 courses for credits.

Academic Honors:

- **University Full Scholarship**
- On the Dean's List every semester (Term GPA $>$ 3.7/4.3)

Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA
Special Student at EECS department **February, 2018 - June 2018**

Courses and Grades: 3 computer science (Course 6) graduate-level courses:

- **Inference and Information**
  - Math course on decision theory and information theory.
- **Natural Language Processing (NLP) I**
  - Classical NLP methods with an emphasis on linguistic and cognitive understandings.
- **Science of Deep Learning**
  - Theoretical topics on optimization/generalization, generative models, adversarial examples, and reinforcement learning.
  - **Completed a research project on adversarial generation and perturbation elimination using generative adversarial networks (GANs)**, led to an unpublished thesis.
  - See research experiences #3 below for details.

GPA: 5.0/5.0, 3 courses for credits.

An Independent Research Project:

- Improved transfer learning and representation learning in massive open online courses (MOOCs).
- See research experiences #1 below for details.

Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA
Research Intern at **CSAIL** **June, 2017 - August 2017**

Summer Research in Any Scale Learning for All (ALFA) Group:

- Worked on MOOC-Learner-Project: a data analytics and visualization framework for e-learning.
- See research experiences #4 below for details.

The High School Affiliated to Renmin University of China (Senior High), Beijing, China
High School Student **September, 2012 - June, 2015**

Won the 6th place in the 16th Asian Physics Olympiad (APhO), 2015.
Selected as one of the 8 members of the China team for the 16th APhO, from a pool of more than 10,000 applicants.
Selected as one of the 5 members of the China team for the 27th International Young Physicists' Tournament (IYPT), 2014.

Accepted Papers

1. Mucong Ding, Yanbang Wang, Erik Hemberg, and Una-May O'Reilly. 2019. **Transfer Learning using Representation Learning in Massive Open Online Courses**. In Proceedings of International Learning Analytics and Knowledge Conference (LAK'19). ACM, New York, USA, 10 pages.

2. Mucong Ding, Kai Yang, Dit-Yan Yeung, and Ting-Chuen Pong. 2019. **Effective Feature Learning with Unsupervised Learning for Improving the Predictive Models in Massive Open Online Courses**. In Proceedings of International Learning Analytics and Knowledge Conference (LAK'19). ACM, New York, USA, 10 pages.

3. Mu Cong Ding and Kwok Yip Szeto. 2017. **Selection of Random Walkers that Optimizes the Global Mean First Passage Time for Search in Complex Networks**. In Proceedings of the International Conference on Computational Science (ICCS'17). Procedia Computer Science, Zürich, Switzerland, 5 pages.

Honors and Awards

2017/04 **The First Runner-Up** of *Mr Armin and Mrs Lillian Kitchell Undergraduate Research Award 2017*, HKUST
2016/05 **The Champion** of *The 4th HKUST UG Math Competition* (Junior Group), HKUST
2015/05 **Gold Medalist** of *Asian Physics Olympiad (APhO)*, Hangzhou, China
- Ranked 6th in Asia
2014/11 **Gold Medalist** of *Chinese Physics Olympiad (CPhO)*, Hangzhou, China
2014/03 **The 4th Place** of *Chinese Young Physicists' Tournament (CYPT)*, Tianjin, China
2014/02 **The First Prize** of *Physics Olympiad of Pan Pearl River Delta Elites and Prestige Schools of China*, Hong Kong, China

Scholarships

2015 - 2019	**University Full Scholarship**	**190,000 HKD/Yr**
2017/08	HKSAR Government Scholarship Fund - Reaching Out Award	10,000 HKD
2017/04	Mr. Armin and Mrs. Lillian Kitchell Research Award	10,000 HKD
2016/08	UROP Summer Research Stipend	4,000 HKD
2016/05	The 4th HKUST Undergraduate Math Competition Award	2,100 HKD

Research Experiences

1. Transfer Learning and Representation Learning in Massive Open Online Courses
Supervisor: **Dr. Una-May O'Reilly** (MIT) (Website) and **Dr. Erik Hemberg** (MIT)
Spring 2018 - Present

Transfer Learning using Learned Representations

- **Proposed an automated transductive transfer learning approach to transfer knowledge from a finished course to an ongoing one**, which only relies upon the problem-agnostic temporal organization of the clickstream data. **Enabled using data from more than one course to model student behavior.**

- **Proposed two novel transfer algorithms using the representations learned by auto-encoders, which can reduce the AUC (area under ROC curves) scores of student dropout prediction by 8% on average and consistently outperforms the other three baseline methods.**
- **Led to a first-author paper accepted by the 9th International Learning Analytics and Knowledge Conference (LAK'19), Tempe, Arizona, USA.**

Developing a Framework for Training and Evaluating Models on MOOC data

- Designed a pre-processing pipeline which was **15 times faster and more stable** than the original MOOC-Learner-Curated software.
- **Developed a distributed pre-processing, training, and visualization toolkit for edX courses named MOOC-Learner-Data-Science-Analytics** (MLDSA) **using PyTorch, in which many useful models and transfer algorithms were implemented.**
- **Led to a first-author paper in preparation for the 12th Educational Data Mining Conference (EDM'19), Montréal, Canada.**

2. Representation Learning for Student Performance Prediction in Massive Open Online Courses

Supervisor: **Prof. Dit-Yan Yeung** (HKUST) (Website)

Fall 2017 - Spring 2018

Representation Learning for Performance Prediction

- **Replaced the conventional feature design and extraction step with an automated pre-processing algorithm and deep learning.**
- **Solved the feature learning problem using unsupervised learning to learn a compact representation from the raw features.**
- Proposed a **modified long short-term memory (LSTM) auto-encoder (AE)** which learns a fixed-length embedding for each input sequence.
- Learned features can **reduce the prediction mean-squared error (MSE) by up to 17%** compared with completely supervised models in the task of predicting students' performance.
- **Led to a first-author paper accepted by the 9th International Learning Analytics and Knowledge Conference (LAK'19), Tempe, Arizona, USA.**

3. Adversarial Generation and Perturbation Elimination with GANs

Supervisor: **Prof. Aleksander Mądry** (MIT) (Website) and **Prof. Constantinos Daskalakis** (MIT) (Website)

Spring 2018

Adversarial Attacks and Defenses using GANs

- Inspired by AdvGAN (IJCAI'18) (attack), APE-GAN and Defense-GAN (ICLR'18) (defense), explored the possibility of improving both the attack's and defense's performance by combining their strengths.
- **Proposed a combination of AdvGAN and APE-GAN which could be trained as one GAN, and investigated the problem of adversarial generation and perturbation elimination in terms of its min-max formulation.**
- Evaluated the combined model on MNIST, found that **the adversarial examples generated by the combined GAN can fool the classifiers protected by a pre-trained APE-GAN.**
- Led to a 13-page unpublished thesis at MIT.

4. MOOC-Learner-Project: A Data Analytics and Visualization Framework for E-Learning

Supervisor: **Dr. Una-May O'Reilly** (MIT) (Website) and **Dr. Erik Hemberg** (MIT)

Summer 2017

Observing and Understanding Video Watching Behavior by Visualization

- Designed 10 classes of features to characterize users' video watching behavior.
- **Answered many data analytics questions on dropout classification, longitudinal behavioral of students, and correlations between watching behavior and video contents.**

- Led to a 17-page unpublished thesis at MIT.

Building a Data Analytics and Visualization Framework for Open edX

- **Developed MOOC-Learner-Visualized** (MLV) as a platform which plots interactive and static figures for learning analytics based on the proposed features. It is a **powerful tool for educational scientists to initiate ideas and examine theories**.
- **Rebuilt MOOC-Learner-Curated** (MLC) which translates and curates activities captured from a MOOC learner into a relational database, and **MOOC-Learner-Quantified** (MLQ) which quantifies the MOOC learner behavior.
- **Accomplished a year-long proposal** founded by the HKUST-MIT Research Alliance Consortium. Presented at its annual meeting with a poster.

5. Classical Random Walk on Complex Networks

Supervisor: **Prof. Kwok Yip Szeto** (HKUST) (Website)

Fall 2015 - Fall 2017

First-Passage Process of Random Walker on Complex Networks

- **Obtained a new formula of the first-passage probability function via inverse Laplace transform.**
- **Developed a mean-field approximation method to calculate the mean first-passage time and the characteristic relaxation time with high precision.**
- Provided a intuitive viewpoint of complex networks by mapping the structural information to poles on complex plane via an ensemble of random walks.
- **Led to a first-author paper in preparation for Physical Review E.**

Optimization of First-Passage Time for Search with Multiple Random Walkers

- **Developed a novel genetic optimization algorithm to select the initial positions of random walkers.**
- Evaluated the genetic algorithm on artificial random networks and real-world scale-free networks.
- **Led to a first-author paper accepted by the International Conference on Computational Science Conference (ICCS'17), Zürich, Switzerland.**

Papers in Preparation

1. Mucong Ding, Erik Hemberg, and Una-May O'Reilly. 2018. **MOOC Learner Data Science Analytics**. Preparing to submit to the 12th Educational Data Mining (EDM'19), Montréal, Canada. 10 pages.

2. Mu Cong Ding and Kwok Yip Szeto. 2018. **First-passage time distribution for random walks on complex networks using inverse Laplace transform and mean-field approximation**. Preparing to submit to Physical Review E. 8 pages.

Theses

1. Mucong Ding, Sirui Lu, and Zhiwei Ding. 2018. **Adversarial Generation and Perturbation Elimination with GANs**. 13 pages.

2. Mucong Ding and Erik Hemberg. 2017. **Observing and Understanding the Video Watching Behavior in Online Lectures**. 17 pages.

Posters

1. Mucong Ding, Erik Hemberg, and Una-May O'Reilly. 2018. **MOOC-Learner-Project Overview**. 2 pages.

Software

1. **Core Developer.** **MOOC-Learner-Project (MLP)**: taps the potential of Massive Open Online Course student behavioral data by providing data science technology that makes the data accessible for teaching and learning research, thus enables insights into how students learn and how instructors can effectively teach.

2. **Lead Developer and Maintainer.** **MOOC-Learner Data Science Analytics (MLDSA)**: an end-to-end solution (data-processing + modeling + predicting) for MOOC data analytics which provides an easy-to-use and research-friendly interface for interactive data science, which also sup-

ports online and at-scale working.

3. **Lead Developer and Maintainer.** **MOOC-Learner-Visualized (MLV)**: plots static and interactive figures for learning analytics. A powerful tool for educational scientists to initiate ideas and examine theories.

4. **Core Developer and Maintainer.** **MOOC-Learner-Quantified (MLQ)**: quantifies the MOOC learner behavior as longitudinal features.

5. **Maintainer.** **MOOC-Learner-Curated (MLC)**: translates and curates activities captured from a MOOC learner into a relational database.

6. **Lead Developer.** **MOOC-Learner-Modeled (MLM)**: serves as an interface to train and test all kinds of classifier models on all possible set of user longitudinal features, and transfer models among weeks and courses.

7. **Lead Developer and Maintainer.** **MOOC-Learner-Docker (MLD)**: connects all MOOC-Learner-Project pipelines (MLC, MLQ, MLV, MLM) and embeds them into Docker containers with a unified configuration.

8. Personal Project. Rendering Thin Film Interference on Soap Bubbles: renders real looking soap bubbles using *WebGL*. Implements an approximated thin-film interference formula, and simulates the film-thickness distribution on the surface by considering the drifting and sloshing effects.

9. Personal Project. In Browser Demo of Variational Auto-Encoders: demonstrates the training and inference processes of variational auto-encoders on *MNIST* digits using *Keras.js*.

Activities

- **Speaker**, Undergraduate Research Sharing Session, HKUST, 2017
  - Shared my life and experiences at MIT to other students.
  - Helped them prepare for summer research and exchange.
- **Volunteer**, 17th Asian Physics Olympiad (APhO), HKUST, 2017
  - Shared my experiences on the official Facebook account.
  - Discussed with participants and helped them relieve stress.
- **Teaching Assistant**, AP Physics Classes, High School, 2015 - 2016
  - Gave tutorials and full lectures to schoolmates after I graduated from the high school.
- **Founder & Organizer**, Physics Discussion Group, High School, 2013 - 2015
  - Aimed to discuss advanced materials which are not covered in the curriculum.
  - Registered association at the students' union.

Community Engagements

- **Volunteer**, HKUST Connect (Serve Learn Act), Hong Kong, 2018
  - **Feeding Hong Kong (FHK) warehouse session volunteer.** Provided support to assist in inspecting, sorting and packing donated food for charity deliveries to the needy.

Other Interests

- Academic Interests apart from Computer Science:
  - In Math: Stochastic Processes, Graph theory, Differential Equations, Differential Geometry.
  - In Physics: Statistical Mechanics, Condensed Matter Physics, Quantum Computation.
- Sports and Activities:
  - Badminton, Mountaineering, Swimming, Traveling in the Countryside.

Skill Sets

- **Proficient in Python and C/C++**:
  - Developed 6 Python repositories on machine learning, data mining, and visualization.
  - Completed 7 C++ course projects on computer graphics and parallel programming.
  - Have written around **35K lines of Python**, and **15K lines of C/C++**.
- Proficient with Linux, Git, and LaTeX.
- **Experienced software developer. Know how to collaborate in a team.**
- Intermediate proficiency in JavaScript, Java, MATLAB and LISP.