

Teaching Statement

William Yang Wang, CMU

Many areas in *Artificial Intelligence*, in particular *Machine Learning* and *Data Science*, are taking increasingly vital roles in Computer Science education in the era of big data. For many schools, with hundreds of enrolled students and bounded resources, one of the biggest challenges is to teach these classes **at scale**, without trading off the quality. For me, I was very fortunate to be involved in teaching graduate *Machine Learning* and *Machine Learning with Large Datasets* at CMU, where I have attempted to tackle some of the problems in learning at scale: I believe that a core aspect for empowering students is to encourage **meaningful technology use** for both students and teachers in the classroom.

In this statement, I will describe some of my past teaching experiences, including:

- **Lecturing at a Large Summer School.** I have taught an Information Extraction course right before the largest Natural Language Processing conference, with 250 students, researchers, and professors attending.
- **Teaching Machine Learning at Scale.** I co-designed and developed two popular machine learning classes at Carnegie Mellon, where I led a team of teaching assistants to work with more than 200+ enrolled students.
- **Mentoring Students.** Over the years, I have also designed Summer project research plans, and mentored undergraduate and graduate students.

Additionally, I have given invited presentations for smaller audiences at leading computer science research institutions such as Microsoft Research (Redmond and Asia labs), Yahoo! Labs (Sunnyvale and NYC), Carnegie Mellon, Columbia University, University of Southern California, Academia Sinica (Taiwan), New York University, Peking University etc. I was also a recipient of the notable presentation prize at the student research colloquium of the Language Technologies Institute of Carnegie Mellon's School of Computer Science.

1 Lecturing at the Pre-ACL Summer School at Peking University

As a part of the welcome event of the first Annual Meeting of the Association for Computational Linguistics (ACL) being held in mainland China, the local organizers of ACL-IJCNLP 2015 decide to invite five world-renowned experts in NLP, and hold a summer school at Peking University. Organized by the Chinese Information Processing Society, this event is the 10th *Summer School on Human Language Technology*¹. Since a total of **250 students and professors** all over China enrolled in this year's summer school, Peking University has decided to hold the event at the largest available lecture hall on campus.

As the only invited graduate student instructor (other four instructors include tenured professors David Chiang, Nianwen Xue, Yang Liu, and Vincent Ng), I taught a 3.5-hour lecture on Information Extraction (IE), covering various aspects of the theories and practices of modern information extraction techniques. This lecture includes introductory materials in information extraction, where I use intuitive examples to illustrate different tasks in IE, and introduce simple classification and sequential models for named entity recognition. As a part of the course, I also cover recent advances in IE, including distant supervision and latent factor models. To transform theories into practices, I also provide hands-on command-line demonstrations on conditional random fields, Open IE, and ProPPR throughout the lecture. The lecture was well-received: I was surrounded by a few dozens of students for further discussions during the break and after class, and some professors further suggested that next time the organizers should give me more time.

¹<http://www.cipsc.org.cn/summerschool/>

2 Teaching Machine Learning at Scale

In the Fall semester of 2013, I was nominated as the head teaching assistant (TA) for a graduate-level Machine Learning class at Carnegie Mellon University. The class had a record number of enrollment: a total of 150 students from various departments across the campus have registered the class. As a head TA, I am responsible to hold weekly 1.5 hour recitation session, where I cover difficult topics and reinforce core concepts for the class. My recitation materials include various unsupervised, semi-supervised, and supervised algorithms, learning theories, and inference methods.

Given the limited size of the teaching team, a major challenge is to design a set of bi-weekly homework assignments that not only cover the broad areas of theories and practices in machine learning, but also can be easily manageable by the teaching team. In particular, the biggest issue is to design and grade the bi-weekly programming assignments for the class. To do this, I used Autolab², an online judge system that automatically handles and grades of programming assignments. As an early user of Autolab, I designed a Kaggle-like grading mechanism for the machine learning class: the handout includes a training dataset and a validation set, and the students need to implement the learning algorithms. To encourage healthy peer competitions, I designed grading standards based on the accuracy on the holdout test set, the training and testing runtime, as well as the memory footprint. Since the scoreboard was visible to all, students found that unlike traditional machine learning classes where it might be hard to connect practices to theories, the homework assignments in my class were both fun and educational.

Due to the size of the class, it is difficult to interact with students via emails. To solve this issue, I used Piazza, where we answer questions in a discussion forum setup. Another innovation is that we encourage students to answer each others' questions by awarding extra credits to students who have both answered a large quantity of questions and received a lot of positive feedback. By the end of the semester, we have a few stellar students who have answered hundreds of peer questions while strengthening their own understandings of the concepts. Another bonus of this innovative strategy is that: I have found new teaching assistants to replace me for next year's class at no extra cost.

In the Spring semester of 2014, I became the head TA for another popular class—*Machine Learning with Large Datasets*, where we teach scalable machine learning algorithms and practical implementations. Given my prior experience with Autolab, I have further augmented Autolab's virtual grading machines with the capability of running distributed machine learning algorithms using Hadoop. To prepare professional Masters students for their industrial jobs, I have also created Pig and Hadoop homework assignments that run on Amazon Web Services (AWS) with large clusters.

3 Mentoring Students

At the very early stage of my graduate school, I have acted as a mentor for a minority undergraduate student who does not have a background in programming, machine learning, or speech processing. Under my mentorship, she was able to perform data analysis based on WEKA, and understood the basics of machine learning. Many years later, when I was in my final years as a PhD student, I have also mentored an undergraduate and two graduate students for their summer research projects in scalable machine learning for probabilistic first-order logics. The undergraduate student later went on to a graduate program at a prestigious school, and one graduate student got into a top PhD program in machine learning in the south central.

4 Conclusion

All these varied experiences have given me excitements in advancing modern Computer Science education. I am extremely excited to teach introductory undergraduate classes in *Data Science* and *Artificial Intelligence*, and I also look forward to teaching graduate classes in *Machine Learning*, *Natural Language Processing*, *Computational Social Science*, and *Information Extraction*. As a computer science student for the past ten years, I am now looking forward to expanding my knowledge and taking a new role: a computer science professor who aspires to inspire students to develop computational thinking.

²<https://autolab.cs.cmu.edu/>