Teaching Statement

Jia Yu
jiayu2@asu.edu

One of my primary reasons for seeking a faculty position is to have the privilege to teach students. I believe that education is the foundation on which our society achieves progress. Although we, as computer scientists, write programs to teach computers how to execute a sequence of repeated tasks, human learning is a subtle and complex process. Teaching is not to repeat what is written in the textbooks, but to communicate the key ideas based on critical thinking and hands-on experience.

Teaching experience. My teaching experience started back in 2014 when I was a teaching assistant for the undergraduate course *Introduction to Programming Languages* (CSE 240). In the past six years, I served as a teaching assistant for various courses including the undergraduate-level *Database Management* (CSE 412) and the graduate-level *Information Assurance* (CSE 543) and *Distributed Database Systems* (CSE 512). In Fall 2019, I was a course instructor of *Data Processing at Scale* (CSE 511) which is a sister course to CSE 512. These collegiate teaching positions provided me with valuable experience in developing substantial teaching materials. I also deeply participated in the design of projects and assignments for the ASU database courses [1]. Recently, I helped ASU move CSE 511/512 to Coursera [2] with fully automated lectures, quizzes, assignments and projects, such that we are able to offer the ASU Online Degree in Master of Computer Science to the entire world. More than 10,000 students enrolled in the free sections of this course. Over time, CSE 511/512 (co-designed by my advisor Prof. Mohamed Sarwat and me) has become one of the most popular graduate computer science courses at ASU.

Teaching interest. My teaching interests lie in the area of big data systems including database management systems, distributed and parallel databases. I wish to teach the students how to manage big data using existing databases or their home-made systems. Since the volume of available data increases at a staggering rate, I strongly believe that big data management comprising storage, querying and visualization has become an essential skill that every qualified computer engineer should have. I will systematically develop lectures and hands-on projects for both undergraduate and graduate courses. My direct experience in mentoring students from diverse technical and cultural backgrounds, instructing real courses on various topics, and interacting with many great educators, would inter alia further me in achieving such goals.

Interactive teaching. My pedagogical style is student-centered and hence I lean toward establishing an active learning classroom where students contribute in discussions and group activities. Based on my experiences in teaching and mentorship, I have found that students learn the best when their minds are actively engaged. Despite the fact that computer science lectures always contain a substantial amount of mathematics, statistics and programming, I believe that teaching is more effective when organizing two-way discussions instead of one-way lecturing. For example, while teaching a graduate database course *Data Processing at Scale* (CSE 511) at ASU, I frequently engaged students with open discussions on relevant hot topics such as Relational DBMS versus NoSQL and Hadoop MapReduce versus Spark. In the future, I plan to give some in-class demonstration opportunities to students who perform outstandingly well in assignments and projects. During the live demonstration, students will be able to showcase their systems, interact with fellow classmates and collect helpful feedback.

Hands-on practice. Another focus in my courses would be to prepare students to apply their knowledge in practice. To this end, I strive to show students how to connect theoretical concepts to applications from different domains. Hands-on projects are an important part of my course. The projects will enable students to practice the course material through programming exercises. For instance, I developed several interesting projects for graduate database courses CSE 511/512 both at ASU and for Coursera. One of the projects is to implement distributed query algorithms in Apache Spark SQL to facilitate large-scale geospatial hot spot analysis (see the footnote). Over the span of this project, students not only get in touch with the cutting-edge big data system, Apache Spark, but also learn how to apply their new skill to solve a realistic spatial analysis problem in Uber and Lyft which is about identifying statistically significant spatial hot spots such that taxi companies can deploy taxis accordingly. Many past students highlight our projects on their LinkedIn profiles to attract recruiters. In the future, I plan to continue incorporating state-of-the-art programming paradigms in my projects for real-life applications of streaming analytics, big data visualization and big graph management.

Collaborative learning. I also put a great emphasis on collaborative learning. Well-arranged groups form inclusive learning environments and facilitate internal collaboration among group members. Students from various disciplines with different backgrounds can work together and participate in open discussions. As a result, they can educate each other and learn how to tackle problems from different perspectives. While teaching CSE 511 at ASU, to make sure that students actively contribute to their group work, I added the peer evaluation phase and asked students to show their GitHub contribution graph at the end of the course. During that semester, each group was also required to turn in a status report every few weeks so that I could intervene to solve problems under unexpected circumstances and facilitate the smooth completion of all the student group projects.

[1] One of the projects in CSE 511/512 graduate database course: hot spot analysis
[2] ASU Online Degree in Master of Computer Science: main page on Coursera, syllabus, enrollment statistics