XIANG REN

TEACHING STATEMENT

Part of the reason why I worked on the tools of StructNet for my research was that I deeply enjoy helping people explore the factual knowledge in massive literature, making it easy to understand how things are connect, summarizing things into coherent stories, gaining insights from the known facts, and applying the insights to solve new problems. I find the experience of *sharing knowledge with the eager minds* and watching such activity *changing their career and life in a positive way*, to be highly rewarding and impactful. Moreover, I consider teaching as an indispensable component of my academic career and I firmly believe that the *best way to learn is to teach*.

As a TA and also a mentor, I always aim to not only provide students with the necessary tools to complete the assignments at hand, but also inculcate them with *general problem-solving skills* that apply well beyond the tasks. For instance, I often demonstrated to my students on how to break difficult and complex problems into smaller, more tractable subproblems, shared my way of understanding the problem, and explained the rationale behind the way I tackle the problem. I view teaching is successful if students *learn about the way of critical thinking* that is taught in class, or *find their own ways of learning and discovering new knowledge* by the end of my mentorship.

LECTURING EXPERIENCE

I put my teaching philosophies to work primarily in classroom. I have been teaching university students since I joined the TA team of Zhejiang University's Data Structures course as a junior. Since then, I have taught areas ranging from introductory computer science to data mining. In addition, I have given three tutorials in top conferences, and lectured about trending machine learning topics in research seminars for the past four years.

My core teaching experience is in data mining. I have been fortunate to help teach the introductory data mining course (CS412: *An Introduction to Data Warehousing and Data Mining*) at UIUC with Professor Jiawei Han. As a lecturer, I am convinced that *students have diverse background and should be treated with tailored requirements*. In my class, while students with data science background were asked to fully understand how a data mining algorithm works and why a specific model is designed in such a way, I asked students with limited math and statistics background (e.g., students from business school) to grasp the high-level intuitions of the models and know when the algorithms are applicable. I love the challenge of *adapting my presentations to audiences with diverse background*. For example, in my lecture on cluster analysis, I provided examples in both computer science and business intelligence to illustrate the k-means algorithm, in order to clearly explain the concept to different audiences. *Being interactive* is a critical part of my lecturing. I frequently raised questions related to the ongoing lecture content to engage with the students, and collected feedback from the class to check their pace. Furthermore, I actively *explored the emerging online tools to facilitate teaching and learning*. For example, I encourage students to use Piazza (`piazza.com`) to raise questions and form discussion, use stackoverflow (`stackoverflow.com`) to seek answers for small technical issues they faced, and use Coursera (`coursera.org`) to find more course-relevant resources for further reading.

Giving tutorials in academic conferences provides me a different yet complementary experience in lecturing. My tutorials were about the recent advances on automated entity recognition and typing in massive text data. As my audiences came from different areas and disciplines, I learned to *make clear connections between the techniques introduced in the tutorial and the problems and downstream applications in different domains*. For instance, I gave concrete examples in the motivation section of my tutorials to illustrate that the entities extracted from biomedical literature can be served in literature search systems, and thus facilitate scientific research. This helped attract more audiences and motivate the tutorial topic. In addition, I found audiences from different areas have quite diverse technical background. Most of them might get lost if I emphasize too much on the technical details (e.g., problem formulations and formula derivations). Therefore, I tried to distill the high-level ideas and insights from the technical approaches, and use intuitive examples to walk audiences through the ideas. I believe that the high-level insights will become the take-home message to inspire the audiences in the future exploration of their own research problems.

MENTORSHIP

Collaborating with others has been one of the most enjoyable parts of my PhD, so I have actively sought out opportunities to advise students. I have had the pleasure of mentoring eleven students in research: I advised four undergraduate students in their senior theses, guided two master student in turning their class projects into published papers, managed three master students and a PhD student who worked in our group as research assistants. These

collaborations resulted in seven papers at major data mining and NLP conferences (e.g., KDD, WSDM, EMNLP). In particular, one under-represented undergraduate student I advised, Urvashi Khandelwal, won CRA Outstanding Undergraduate Research Award and got admitted by the PhD program at Stanford. Wenqi He, an under-represented master student I advised, was named Siebel Scholar and published three research papers in her first year.

In all cases of my mentorship, I *participated during the entire process*, from interviewing and selecting candidates, to choosing projects for them, to guiding and supporting their work. Each student is different, and it has been an interesting challenge to *support each one as it best suited them*. Some are extremely efficient in learning and implementing, but need help with ideas; I try showing them a clear path, while also offering opportunities for developing their own thoughts. Some have great ideas but are inexperienced and inefficient; I try teaching them detailed research approaches, e.g., coming up with a concrete example to illustrate a specific solution idea. Some have strong theoretical background, others are on the practical side; I try bringing their respective strengths to full fruition, while helping them overcome their weaknesses. Regardless of student type, I *consider advising a mutual learning and enhancing process*: just as I have helped my students grow, they have helped me by teaching me many things and providing feedbacks on my proposed ideas.

BUILDING COMMUNITY

I believe that *vibrant community play a critical role in editing and improving research ideas*. Therefore I took on responsibilities organizing the research team on *Multi-Genre Knowledge Networks Construction* at UIUC, which gathers researchers from different disciplines (computer science, biology, statistics), and put them in direct contact with external collaborators from other universities and institutes (USC, PSU, RPI, US Army Research Lab). In collaboration with Jiawei Han, I helped lead this project end to end, drafting the proposals, recruiting students, tracking and reporting progress, organizing teleconferences, and making paper submissions. This experience was challenging and rewarding: the team published over ten research papers in top data mining conferences in the past two years, and the funding of the project was extended by the US Army Research Lab last year. My interest on text mining further motivated me to expand the community of text mining scholars at UIUC. I began with a weekly reading group that drew in participants from computer science, biology, electrical and computer engineering, statistics. I actively contributed to this group by organizing the group's weekly meetings, setting the agenda, scheduling and hosting guest speakers, deciding on readings, etc. I look forward to fostering such a community as a faculty member.

POTENTIAL COURSES TO TEACH

As my research spans multiple areas of computer science, I can teach a broad spectrum. I would be happy to teach any introductory computer science course (e.g., *Introduction to Computer Science*, *Data Structures*, *Algorithms*), as well as advanced undergraduate and graduate courses on ***information management*** (e.g., *Data Mining*, *Database Systems*, *Information Retrieval*), ***machine learning***, ***natural language processing*** (NLP), and ***artificial intelligence***. I also look forward to designing new interdisciplinary courses and seminars, as sketched below:

Introduction to Data Science. This project-based course is intended to introduce students to modern programs and technologies that are useful for organizing, manipulating, analyzing, and visualizing data. It is designed to familiarize students with all steps of data analysis: data collection and cleaning, exploratory analysis, hypothesis testing, choice of data mining and machine learning algorithms, system evaluation, data and result visualization, etc. Many of these topics are offered as separate courses, but I believe it is important to show students how they all fit together by uniting them around a practical domain.

Data-Driven Text Analytics. This undergraduate-level course will cover the basic concepts, principles and major techniques for mining and analyzing text data to discover interesting patterns, extract useful knowledge, and support decision making, with an emphasis on data-driven machine learning approaches that can be generally applied to arbitrary text data in any natural language with no or minimum human effort. It will discuss the potential applications of text mining techniques in multiple disciplines, such as biology, health science, and public policy.

Advanced Topics in Text Mining. This graduate-level course (or seminar) covers the major research topics and recent advances in text data mining, with an emphasis on the use of machine learning techniques in solving NLP and text mining problems and applications in other disciplines (e.g., health science, life science, social science). Topics include: deep learning for NLP, information extraction, biomedical text analysis, social media analysis. Readings from the literature are paired with student presentations and projects to recreate or extend well-known work.

I see teaching and advising as integral parts of my academic career in scientific research, and I believe that the best way to learn is to teach.