Wei Xu

Teaching Statement

I am pursuing an academic position in large measure because I find it deeply fulfilling to teach and mentor students. I believe it is one of the best ways to make a lasting impact in the world, as students launch careers in academia and industry, or become entrepreneurs.

Teaching Experience I designed and taught a completely new course *Social Media and Text Analytics* (`http://socialmedia-class.org`) at University of Pennsylvania in 2015. It was a graduate-level summer course attended by 20 students and it received very positive feedback from students in the anonymous post-class survey (attached at the end). The course covered various machine learning algorithms for text processing and social media analysis. I was excited to see that students from both CS and non-CS backgrounds were very interested in my research areas and the Twitter API tutorial I wrote as part of the course assignments is being used by other teachers and researchers worldwide. I was selected to offer this course again at the North American Summer School on Logic, Language, and Information (NASSLLI) in 2016 after a competitive peer review process by the program committee. The acceptance rate of course proposals was 38% and NASSLLI instructors are mostly tenured or tenure-track professors at research universities. It will be straightforward for me to extend this course or alter it to the undergraduate level.

During my PhD and postdoc, I also have had the privilege to gain teaching experience at different universities. I have enjoyed giving guest lectures on *Natural Language Processing* at Ohio State University, Rutgers University and University of Pennsylvania. I worked as a teaching assistant at the New York University for two CS courses: *Web Search* and *Programming Languages*.

Teaching Philosophy My main goal for teaching is to give students the technical knowledge while inspiring them to enjoy learning and have critical thinking. Lectures are very crucial and can be exciting for students if they can actively participate. Encouraging students to ask questions and having short discussions helps them understand and think more deeply about a topic. After giving many invited talks and conference presentations, I also have learned to read audiences' facial and body language and adjust pace accordingly. Depending on the students' prior experience, course projects can vary from well-defined to open-ended research questions. Interesting and challenging problem sets help to internalize the knowledge. I am interested in having open online video-recorded lectures and in-class interactive exercises to provide students more hands-on instructions and an opportunity to review lectures at anytime.

Teaching Interest I am eager to teach courses in *Machine Learning*, *Natural Language Processing*, *Data Science* at all levels. These fields have many broadly applicable technologies and interesting real-world applications I want to cover. I would also enjoy teaching artificial intelligence, algorithms, data structures, crowdsourcing and other computer science introductory courses.

Mentoring Experience I have had the pleasure to mentor several students on research projects and help them achieve their first publications. One masters student I advised at University of Pennsylvania wrote a first-authored paper with me that was published at NAACL. He is now a PhD student at UIUC. I also advised two undergraduates at Penn. Their project on automatic poetry generation was published at the AAAI's HCOMP conference, received press coverage from New Scientist, and won 2nd place among all undergraduate and graduate student projects in the CIS department. I also mentored a junior PhD student at New York University to and helped her publish her first paper at the ACL conference.

Mentoring Philosophy Mentoring students requires effort both to keep up to date on current technical advances, and to build a trusting relationship. It is very interesting and rewarding to observe students as they grow. It is important to give students consistently good advice and timely feedback. It is also very important to give students the right amount of guidance and freedom and this amount will vary depending on students and their experiences. I look forward to building a collaborative and energetic research team with students at all levels. I also hope to collaborate with other research groups via joint reading groups and seminars with internal and external speakers. As a woman in computer science, I want to mentor and inspire more female students to be successful in the field.

(Attached)
Syllabus and teaching evaluation of the course I taught at University of Pennsylvania

Course Title: Social Media and Text Analytics

Course Website: **http://socialmedia-class.org/**

Course Syllabus:

Social media provides a massive amount of valuable information and shows us how language is actually used by lots of people. This course will cover: 1) machine learning algorithms and off-the-shelf tools for processing text data; 2) prominent research findings on language use in social media.

Lecture 1: Big Data Social Science
- Relationship between social media, big data, data science, natural language processing
- Social science research on finance, politics, health, human behavior, etc.

Lecture 2: Twitter and Twitter API Tutorial
- Brief history of Twitter
- Key features of Twitter and Twitter API

Lecture 3: Natural Language Processing for Social Media (part 1)
- Main research areas and publication venues of NLP
- Domain/Genre Difference
- Language identification
- Classification method: Naïve Bayes + feature selection (Information Gain)

Lecture 4: Natural Language Processing for Social Media (part 2)
- Challenges in processing social media text
- Tokenization and Text Normalization

Lecture 5: Natural Language Processing for Social Media (part 3)
- Part-of-speech tagging, Chunking, Named entity recognition
- Sequential tagging methods (HMM and CRF) and their relations to Naïve Bayes
- Unsupervised learning method: Brown Clustering

Lecture 6: Case Study #1 - Automatic Summarization for Twitter and the PageRank Algorithm
- SumBasic algorithm
- PageRank algorithm
- Graph visualization
- Event-based summarization system

Lecture 7: Case Study #2 - Learning Large-Scale Paraphrases from Twitter
- Overview of paraphrase research
- Crowdsourcing for data collection
- Multi-instance learning and latent variable model

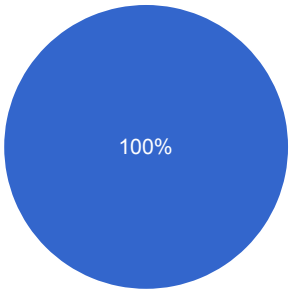
Lecture 8: Deep Learning for Natural Language Processing
- Intro to deep learning
- Language model
- Word vector representations
- Skip-gram and CBOW model

7 responses

View all responses

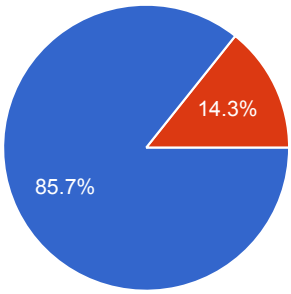
Summary

Did you enjoy the class?



Yes	**7**	100%
No	**0**	0%

Would you recommend the class to your friends?



Yes	**6**	85.7%
No	**1**	14.3%

What is your affiliation? Which department or institution?

Positive Psychology Center at Penn/ MCIT student SEAS at Penn

CHOP

UPenn

WWBP

Department of Positive Psychology

Department of Linguistics at Penn

What is your occupation?

data analyst

Web Developer

Student

Research Programmer

PhD student

research programmer

What is your major/background?

Computer Science

Economics/Comp Sci/Bioethics

Electrical Engineering

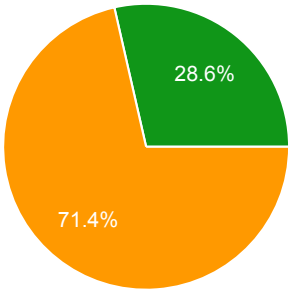
Comp sci

Physics, Computer Science

Linguistics

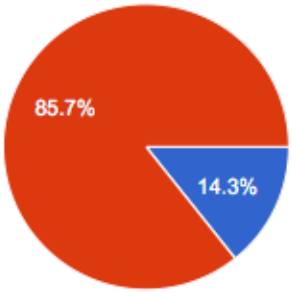
CS

What is the highest degree you have completed/obtained?



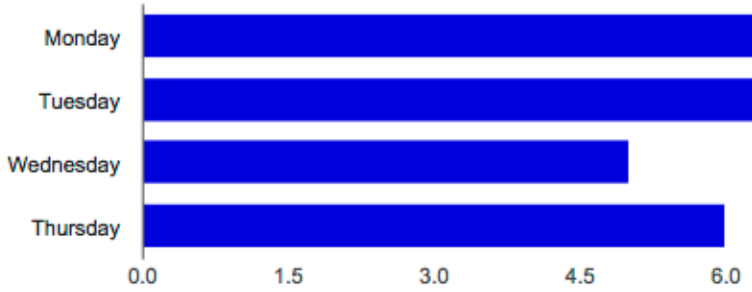
Doctorate degree (for example: PhD, EdD)	**0**	0%
Professional degree (for example: MD, DDS, DVM, LLB, JD)	**0**	0%
Master's degree (for example: MA, MS, MEng, MEd, MSW, MBA)	**5**	71.4%
Bachelor's degree (for example: BA, AB, BS)	**2**	28.6%
Associate degree (for example: AA, AS)	**0**	0%
1 or more years of college, no degree	**0**	0%
Other	**0**	0%

What did you think about the level of difficulty of the course?



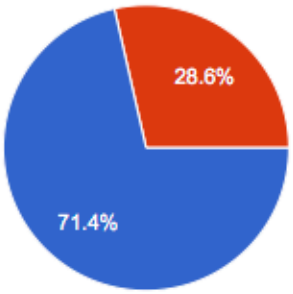
Easy	1	14.3%
Just about right	6	85.7%
Difficult	0	0%

Which dates did you attend the class?



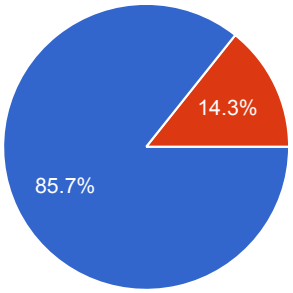
Monday	7	100%
Tuesday	7	100%
Wednesday	5	71.4%
Thursday	6	85.7%

How do you rate your course?



Above average	5	71.4%
Average	2	28.6%
Below average	0	0%
Other	0	0%

How do you rate your instructor?



Above average	**6**	85.7%
Average	**1**	14.3%
Below average	**0**	0%
Other	**0**	0%

What do you like the most about the instructor and/or the course?

I thought it was a nice mixture of NLP basics with some more detailed explanations of clusters algorithms.

Very cool slides, nice graphics, very good at inspiring curiosity!

The instructor is well-prepared and very knowledgeable about what she's doing. The content of the class is also interesting.

The instructor is clearly very knowledgable and friendly. The course gives a good broad level overview.

I really liked the instructor's enthusiasm for her subject. It was an easy course, but appropriate - I thought - for a mini-course. I liked that it was a good survey of tools and methods, and gave plenty of information to find out more for those of us who were interested.

Really gave a good overview into what's going on with Twitter text analysis

Having worked a lot in NLP, I had to learn a lot on the fly, but I had never taken a class in it, so this class was really useful to fill in the gaps of the things I had seen but not totally understood yet. I also liked the aspect of the course that explained the "state of NLP", like how the conferences work and where the NLP research landscape stands, etc.

Any feedback on how to be a better teacher, or on how to improve the class?

More concrete examples would be helpful.

It would have been really cool to do live examples of most of things talked about. For instance, when talking about POS tagging, you could ask the class for a sentence, and then input it into a POS tagger jar program, and show them that. Same for like parsing or paraphrase etc. I think that would really add a little extra to the course

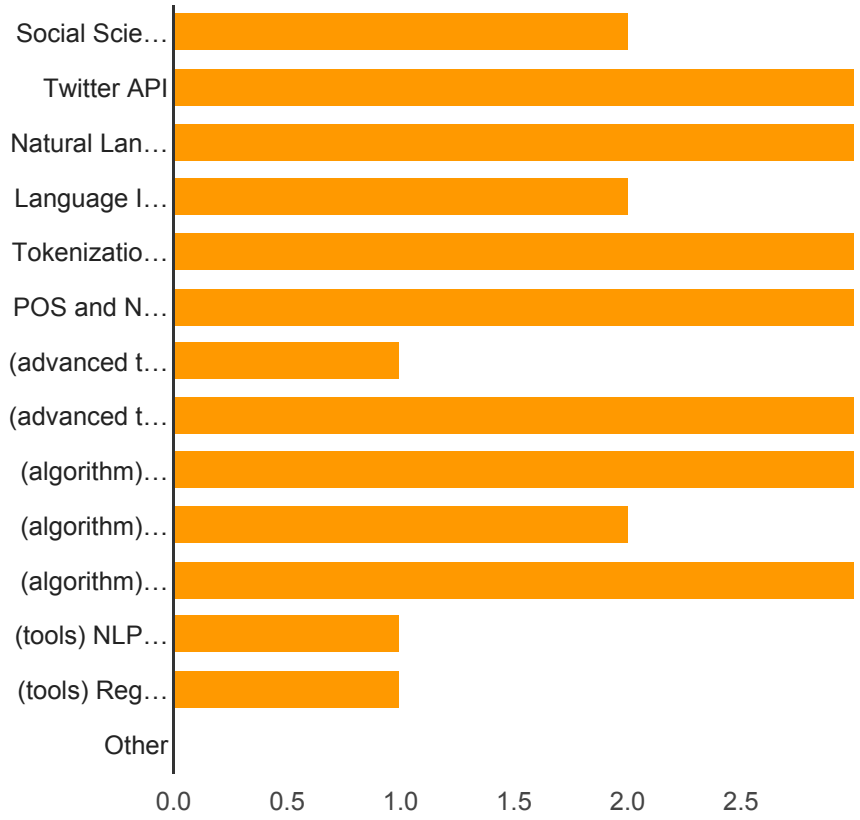
I would love to have had some hands-on time, individually or in groups, to actually do some coding and processing of Twitter data. For my purposes, practical application is better than theoretical overviews.

For this particular course, I think you could remove the information about NLP conferences and paper publishing guidelines. I'm not sure that it was of interest to those from other fields.

This is already quite good. It might be nicer to have this as a sequential class after a class for python or some background stuff. The course is a little too short but it's not about the instructor herself.

Hard to understand accent sometimes. Would have enjoyed a mix of lecture and hands on during class time.

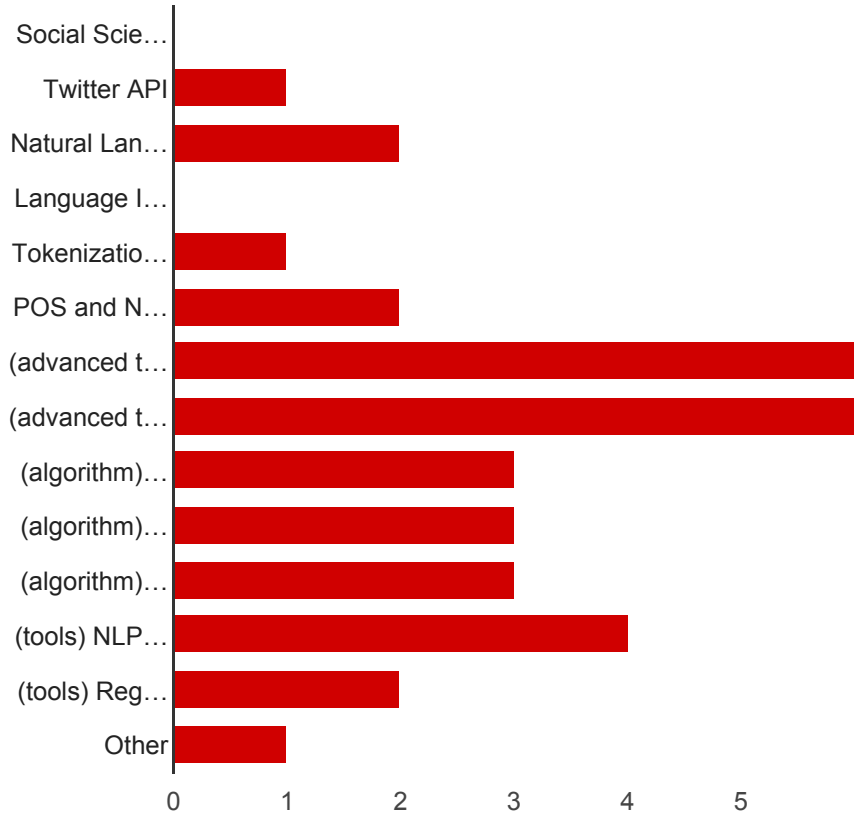
Which were your favorite topics?



Social Science Research	**2**	28.6%
Twitter API	**3**	42.9%
Natural Language Processing 101	**3**	42.9%
Language Identification	**2**	28.6%
Tokenization/Stemming/Normalization	**3**	42.9%
POS and Named Entity Tagging	**3**	42.9%
(advanced topics) Summarization	**1**	14.3%
(advanced topics) Paraphrase	**3**	42.9%

(algorithm) classification method - Naive Bayes	**3**	42.9%
(algorithm) clustering method - Brown Cluster	**2**	28.6%
(algorithm) ranking method - PageRank	**3**	42.9%
(tools) NLP pipeline tools	**1**	14.3%
(tools) Regular Expression	**1**	14.3%
Other	**0**	0%

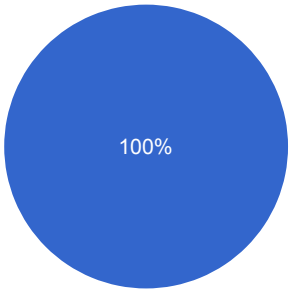
Which topics you would like the instructor teach more in details, if it is a full-length course?



Social Science Research	**0**	0%
Twitter API	**1**	14.3%
Natural Language Processing 101	**2**	28.6%
Language Identification	**0**	0%
Tokenization/Stemming/Normalization	**1**	14.3%
POS and Named Entity Tagging	**2**	28.6%
(advanced topics) Summarization	**6**	85.7%
(advanced topics) Paraphrase	**6**	85.7%
(algorithm) classification method - Naive Bayes	**3**	42.9%

(algorithm) clustering method - Brown Cluster	**3**	42.9%
(algorithm) ranking method - PageRank	**3**	42.9%
(tools) NLP pipeline tools	**4**	57.1%
(tools) Regular Expression	**2**	28.6%
Other	**1**	14.3%

Would you like the instructor to offer this course in full-length in the future?



Yes	**7**	100%
No	**0**	0%

Any suggestions on what else topics to include into a full-length course?

More step-by-step "labs" that include details on downloading and installing tools, tweaking code, and practicing the skills we're learning about. Very, very clear and easy to follow starter homeworks.

Twitter topic modeling, more details on how POS tagging and NE recognition are accomplished

Some more exercises on python maybe.

Just maybe more types of algorithms for the different NLP tasks, and which ones are used nowadays. Also, neural networks ?

Did your instructor have any unconscious ticks that were distracting?

No!

No.

No

I didn't notice anything haha

Haha! No, she was fine.

Any suggestions on how IRCS or the instructor can advertise the course better?

This is already good.

Not that I can think of