[DOWNLOAD A USC IDENTITY GUIDELINES COMPLIANT LOGOTYPE, THEN DELETE THIS TEXT AND DRAG/DROP THE FILE HERE]

Course ID and Title "Machine Learning for Knowledge Extraction and Reasoning"

Units: 4.0

Term—Day—Time: Spring 2019—Monday—2:00-5:20pm

IMPORTANT:

The general formula for contact hours is as follows:
Courses must meet for a minimum of one 50-minute session per unit per week over a 15-week semester. Standard fall and spring sessions (001) require a final summative experience during the University scheduled final exam day and time.

(Please refer to the *Contact Hours Reference* guide.)

Location: TBD; (website will be up at http://www-bcf.usc.edu/~xiangren/cs699_IE_fall2018.html)

Instructor: Xiang Ren

Office: SAL 308
Office Hours: by appointment
Contact Info: xiangren@usc.edu, 213-821-4067, Timeline for replying to emails/calls: within 48 hours

Teaching Assistant: Hongtao Lin

Office: TBD
Office Hours: TBD
Contact Info: lin498@usc.edu

IT Help: N/A
Hours of Service:
Contact Info: Email, phone number (office, cell), Skype, etc.

Course Description

In today's computerized and information-based society, people are inundated with vast amounts of text data, ranging from news articles, social media posts, scientific publications, to a wide range of textual information from various vertical domains (e.g., corporate reports, advertisements, legal acts, medical reports). How to turn such massive and unstructured text data into structured, actionable knowledge, and how to enable effective and user-friendly access to such knowledge is a grand challenge to the research community. This course will introduce and discuss many of the sub-problems and machine learning approaches for knowledge extraction and reasoning, including use of language features, sequence learning models, rule learning, relational learning, and deep learning techniques. We will discuss segmentation of text sequences, classification of segments into types, clustering and de-duplication of records, knowledge graph embedding, knowledge reasoning.

Learning Objectives

At a high-level, through this course students will have a concrete idea of what knowledge extraction and reasoning is about, what the state-of-the-art is, and what the open problems are. Along the way, students will explore many of the mainstays of machine learning, including maximum likelihood, maximum entropy methods, discriminative training, mixture models, and semi-supervised training methods, as well as deep learning models such as recurrent neural networks, convolutional neural networks, sequence-to-sequence learning, relational learning, and graph neural networks. The hope is that by the end of this course students will have in-depth understanding about information extraction and knowledge graph reasoning as well as the related machine learning models, and can develop practical algorithms and implement systems for solving different knowledge extraction and reasoning problems.

Prerequisite(s): CSCI 567 (Machine Learning), CSCI 544 (Applied Natural Language Processing); Familiarity with probability and algorithms.

Co-Requisite(s): CSCI 567 (Machine Learning), CSCI 544 (Applied Natural Language Processing)

Concurrent Enrollment: N/A

Recommended Preparation: sufficient mathematical background; general background on optimization; good programming skills

Course Notes

Lecture notes will be available online after each class.

Technological Proficiency and Hardware/Software Required

N/A

Required Readings and Supplementary Materials

- Ren & Han. *Mining Structures of Factual Knowledge from Text: An Effort-Light Approach*
- *Speech and Language Processing*, Daniel Jurafsky and James Martin, Prentice-Hall (second edition).
- Christopher D. Manning and Hinrich Schutze. *Foundations of Statistical Natural Language Processing.* MIT Press.
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*.
- Trevor Hastie, Robert Tibshirani and Jerome Friedman. The Elements of Statistical Learning.
- Convex Optimization (Boyd)
- Deep Learning (Goodfellow & Bengio)
- Deep Neural Methods for NLP (Goldberg)

Description and Assessment of Assignments

What kind of work is to be done and how should it be completed, i.e. how the learning outcome will be assessed. Include any assessment and grading rubrics to be used.

Grading Breakdown

Including the above detailed assignments, how will students be graded overall? Participation should be no more than 15%, unless justified for a higher amount. All must total 100%.

Assigments	Points	% of Grade
HW1	100	10%
HW2	100	10%
Project survey report	100	15%
Project mid-term report	100	15%
Project final presenatation	100	15%
Project final report	100	35%

Grading Scale

A	95-100
A-	90-94
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	59 and below

Assignment Rubrics

N/A

Assignment Submission Policy

By email, before 11:59pm of the due date.

Grading Timeline

Assignments will be graded within one week after the due date.

Additional Policies

Late homework policy: **you are given 4 late days** for the assignments and project proposal/survey (no late days for the final project report), to be used in integer amounts and distributed as you see fit. Additional late days will each result in a deduction of 10% of the grade of the corresponding assignment.

Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings and Homework	Deliverable/ Due Dates
Week 1 Jan 21	**Class Introduction and Outline.** - Self-introductions - Course overview **Tasks for Knowledge Extraction and Reasoning** - source of data; amount of labeled data; closed/open-domain		
Week 2 Jan 24 (make-up)	**Sequence tagging & entity recognition** - Different problem formulation for NER: pattern boostrapping, sequence labeling, etc. - Sequence labeling: HMM, MEC, CRF - semi-supervised models - pattern-based methods - co-training, boostrapping - distant supervision	Project team-up, prepare proposal	Project proposals due by end of Jan 27
Week 3 Jan 28	**Entity typing, linking & co-reference** - supervised learning methods - typing under noisy labels - fine-grained typing models	- HW1 (on named entity recognition) - project survey	
Week 4 Feb 4	**Parsing & structured prediction** - dependency parsing - semantic parsing - statistical models for structured prediction - neural parsing models		Project survey due
Week 5 Feb 11	**Neural sequence models** - Recurrent neural network for sequence tagging - Convolutional neural nets - Character-level & subword methods		
Week 6 Feb 18	**Relation Extraction I** - rule-based methods - pattern bootstrapping - feature-based classifiers **Paper presentation**		HW1 due
Week 7 Feb 25	**Relation Extraction II** - neural models - distantly-supervised methods **Paper presentation**	HW2 (on relation extraction)	
Week 8 Mar 4	**Learning on knowledge graphs** - knowledge graph embedding & completion - knowledge-based question answering **Paper presentation**		Project mid-term report due by end of the week

Week 9 Mar 11	**No class – spring break**		
Week 10 Mar 18	**Knowledge reasoning techniques (guest lecture: Jay Pujara)** **Paper presentation**		HW2 due
Week 11 Mar 25	**Genrative models for NLP** - seq2seq models - machine reading comprehension, text summarization, machine translation **Paper presentation**		
Week 12 Apr 1	**Learning on graph-structured data** - spectral methods - matrix factorization - multi-relation learning methods - graph neural networks **Paper presentation**		
Week 13 Apr 8	**Low-resources learning for NLP (guest lecture: Violet Peng)** **Paper presentation**		
Week 14 Apr 15	**Graph mining and applications (Guest lecture by Neil Shah)** **Paper presentation**		
Week 15 Apr 22	**Project final presentation**		Project final report due by May 5th.