

DS7337: Natural Language Processing

Overview

Natural language processing from an application-builder's perspective, using Python with NLTK, SpaCy, Tensorflow and other packages. Emphasis on readying the student to become a working professional in real-world scenarios.

Course Description

This class introduces natural language processing (NLP) as applied to text mining and other tasks with unstructured big data. Students will receive a broad survey of the major tasks in natural language processing. Topics include document clustering and classification, automated tagging, topic modeling, and feature extraction to support a variety of applications. The focus is on best practices for choosing the right tool and method for an application, illustrated with real-world case studies. The class offers experience building solutions from real-world data sets, utilizing WordNet and the data interfaces of leading websites.

Learn: Feature extraction from unstructured text, text classification, document clustering, sentiment analysis, corpus analytics, syntax parsing, topic modeling

Core Resources:

Python 3.6.X or higher: available at
<https://www.python.org/download/releases/>

NLTK (Natural Language Toolkit) available at
<http://www.nltk.org/install.html>

Bird, Steven, Ewan Klein, and Edward Loper, *Natural Language Processing with Python* (NLTK, 2009, 2014). Available free online at <http://www.nltk.org/book/>

Sarkar, Dipanjan, *Text Analytics with Python* (New York: Apress, 2016). Available on Amazon or at <https://www.apress.com/us/book/9781484223871>

Tensorflow 2.0 <https://www.tensorflow.org/>

Spacy 2.0 <https://spacy.io/>

Schedule: (reminder the professor of record may change this as needed at their discretion)

Unit	Class Topic	Async Topic	Due
1	Neural Networks	Introduction: What is NLP? Two sides of NLP Applications of NLP	Quiz 1
2	Vectors/Glove	Levels of Analysis in NLP Lexical analysis Syntactic analysis Semantic analysis Discourse analysis	HW1, Quiz 2
3	Vectors/Word2Vec	Shallow vs. Deep Statistical vs. Symbolic Feature engineering vs. feature learning	Quiz 3

		<p>Top-down vs. bottomup</p> <p>Transparent vs. opaque (AI vs. XAI)</p>	
4	Recurrent Neural Network	<p>Working in NLP</p> <p>NLP and data science</p> <p>Job roles that utilize NLP</p> <p>Sectors that utilize NLP</p> <p>Organizations that relate to NLP</p>	HW2
5	LSTM/GRU	<p>Low-Level Analysis</p> <p>Text preprocessing</p> <p>Text normalization</p> <p>Low-level document feature extraction</p>	Quiz 4
6	Sentence Classification	<p>Lexical Knowledge Bases</p> <p>Lexical knowledge bases</p> <p>Resources for creating or extending lexical knowledge bases</p>	HW3, Quiz 5

		Applications of lexical knowledge bases	
7	Token Classification	Syntactic Analysis: POS-Tagging POS-tagging Using POS tags	HW4
8	Review	Syntactic Analysis: Parsing Shallow parsing Using chunks Full grammar parsing Uses for full parse trees	, Quiz 6
9			Midterm
10	Seq to Seq	Semantic Analysis: Semantic Relatedness Word similarity Document similarity Applications of semantic similarity	HW5, Quiz 7
11	Attention	Semantic Analysis: Document Clustering	Quiz 8

		<p>Methods of Clustering</p> <p>Working with Clusters</p>	
12	Attention is All you need	<p>Semantic Analysis: Text Classification</p> <p>Document Classification</p> <p>SVMs for text classification</p> <p>Descriptor-based text classification</p>	HW6
13	BERT	<p>Semantic Analysis: Topic Modeling</p> <p>Topic Model Paradigms</p> <p>Organic Topic Modeling</p> <p>Canonical Topic Modeling</p> <p>Entity-Centric Topic Modeling</p>	HW7, Quiz 9
14	Hub Models	<p>Semantic Analysis: Sentiment and Rhetoric</p> <p>General Sentiment Analysis</p>	HW8, Quiz 10

		Advanced Sentiment Analysis	
15	Final Presentation		,Final

Assessments

All assessments will be graded on a standard scale (not a curve).

Quizzes

Quizzes are multiple choice and will be taken online. They may relate to material from the video and/or reading materials. Quizzes will be given most weeks that we are not reviewing for a midterm or final. Each quiz has the same weight for grading purposes as every other quiz regardless of the number of questions in each quiz.

LATE QUIZES WILL BE ASSESSED AS 0

Homework Assignments

Other assignments may involve manually retrieving and analyzing textual data or writing some Python code to process text. You may be required to provide inputs, outputs, source code, and a written summary of your findings. Some of the assignments will build upon your work in previous assignments, so that sequentially, you will be building up a semester project.

All assignments have the same weight for determining the Homework Assignments portion of the final grade.

Late assignments will be assessed at 10% reduced grade PER DAY after deadline

Midterm and Final Exams

These will comprise open-ended questions, requiring short answer or short essay responses. Most of the questions will involve application or problem-solving scenarios, requiring you to explain one or more problems, solutions, methods, apparatuses, and alternatives. Instructor will provide exam details directly to students.

Assignment Type	Weight on Final Grade
Quizzes (multiple choice, online)	20%
Homework Assignments (working with data and/or coding)	30%
Midterm (mainly essay questions)	25%
Final (Project/Coding)	25%

1. Students needing academic accommodations for a disability must first register with Disability Accommodations & Success Strategies (DASS). Students can call 214-768-1470 or visit <http://www.smu.edu/Provost/SASP/DASS> to begin the process. Once approved and registered, students will submit a DASS Accommodation Letter to faculty through the electronic portal *DASS Link* and then communicate directly with each instructor to make appropriate arrangements. Please note that accommodations are not retroactive and require advance notice to implement

2. The DASS office (Disability Accommodations & Success Strategies) works with any SMU student with a disabling condition who might require a reasonable accommodation while at SMU. Most accommodations are test-taking related (extended time), but others may operate behind the scenes. It is important to be aware of any disability-related needs for your enrolled students. Please include the recommended syllabus statement shared by the Provost (see below), which invites students to contact you about their accommodation needs. If they've been approved for any accommodations by DASS, the student will share an electronic letter through the DASS Link portal. Instructors view the letter, sign it once the instructor and student have discussed the plan for implementing the accommodation, and refer to the responsible DASS staff member on the letter for any questions. At times, accommodations in an online program may be technical in nature. Students who are blind or have low-vision may require alternate formats of text or images. Students who are deaf/Hard of Hearing may require a remote captioner who types everything spoken during class. In this case, it is important to remember that cancelling class will impact this contracted service provider. Due to SMU's contract with the service provider, please provide at least 48hrs notice to your students if you need to cancel class, in order for the service to be cancelled without charge. This expectation applies to in-person classes and events, as well, when an interpreter or captioner has been hired to provide access to a

student with a disability. For any questions, don't hesitate to contact the DASS office. We're here as a resource to you.

<https://www.smu.edu/Provost/SASP/DASS/FAQ/ForFaculty#statement>