

CSCE 771: Computer Processing of Natural Language Fall 2020

Section 001: Hybrid

Asynchronous Online (via Blackboard),
Synchronous (via Microsoft Team COTEAM-BIPLAVS-CSCE-771-001-FALL-2020),
In class – only when announced

Biplav Srivastava, Ph.D. Professor, AI Institute, Computer Science & Engg., 541 Main St, Horizon 1, 4th Floor, Univ. South Carolina, Columbia, SC 29208 (518) 496-0128 biplav.s@sc.edu	Course Website: https://blackboard.sc.edu Supplementary Website: https://sites.google.com/site/biplavsrivastava/teaching/csce-771-computer-processing-of-natural-language
	Teaching Assistant: N/A
	Office Hours: TTh 11:30 am – 12:30 pm or by appointment
	Class Timings: MW 3:55 pm – 5:10 pm

Course Syllabus

Catalog Description

This will be an advanced course on computer-based processing of communication between people (languages) with a focus on text (NLP), and appreciation of multi-modal communication involving speech and images. The course will have lessons on parsing of input, representation of syntax and semantics, and analysis to derive insights. It will explain learning and reasoning based methods for extracting entities, disambiguating and linking them, and applying them for real-world problems. We will also discuss issues in building advanced systems for conversation (“chatbots”) and machine translation, and ethical concerns while testing and fielding them with people.

Course Outcomes

- L1: Appreciate diversity and similarity in natural languages – text, speech and visual; focus of course will, however, be text (NLP) and English
L2: Understand issues related to data and tools. Experiment design, Metrics for evaluation and to detect bias, Methods to build trust in processing – transparent assessment, Providing explanations for output
L3: Able to perform data processing: (a) Structured data representation from unstructured text; (b) Extract entities and relationships; (c) Extract contexts; (d) representation learning – word embedding
L4: Understand AI methods in NLP: (a) Learning methods – including language models, (b) Reasoning, (c) Representation – knowledge graphs/ ontology
L5: Understand NLP applications – (a) Document intelligence: sentiment, translation; (b) collaborative assistants

All learning outcomes in this primarily Distributed Learning course, with possibility for in-class experience attendance when safely possible, are equivalent to face-to-face (F2F) version of this course.

Prerequisites

Experience with a first course on Artificial Intelligence is desirable, example: CSCE 580.

Students without this course may still enroll knowing that necessary AI background can be found in the standard AI book, “AI – A Modern Approach”; <http://aima.cs.berkeley.edu/>. Relevant material is in chapters 3,4,5 and 6.

Textbooks and Reading Materials

1. The required textbook for this course is: Speech and Language Processing (3rd ed. draft) Dan Jurafsky and James H. Martin
Available online at: <https://web.stanford.edu/~jurafsky/slp3/>
2. The optional reference book, specially suggested for students without CSCE 580, is: Artificial Intelligence: A Modern Approach (Fourth edition, 2020)
Stuart Russell and Peter Norvig
<http://aima.cs.berkeley.edu/>
ISBN-13: 978-0134610993
3. Research Papers
Links to PDFs will be advertised in class material
4. Open Datasets - Illustration
 - a. Text of legislations - LegiScan, <https://legiscan.com/>
 - b. COVID-19 research papers - <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/> ; <https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19>
 - c. Text of patents, Google patents - <https://patents.google.com/>

All readings/materials comply with copyright and fair use policies.

Required Software

In order to access course materials and complete the course assignments you must have access to:

- A personal computer (PC) or laptop with the **Microsoft Office Suite** (Word, Excel, PowerPoint).
- The Internet (for using email, browsing the web, accessing the course website, and submitting assignments)
- Programming environment setup using open source. Python using Jupyter Notebook or Java using Eclipse will be the supported languages.
- For any proprietary tool, please consult the instructor. As a general rule, the instructor will ensure that no student may get an undue advantage due to usage of a particular technology that is unavailable to others.

Course Format

This course will be delivered **asynchronously through Blackboard** and **synchronously via Microsoft Team** (COTEAM-BIPLAVS-CSCE-771-001-FALL-2020). The primary medium of course delivery will be online with a mix of pre-recorded videos and live online lecture. No student needs to come to physical class if they do not want to.

The instructor will attempt to familiarize with students and be on campus as much is safely possible. If conditions improve and students favor, we will allow a small fraction of students to come to class (like 5-10%) on a rotating basis. Hence, in class will be only when announced and only for interested students. In class attendance will have no impact on course outcome.

- Student-to-Instructor (S2I) Interaction: Students will listen/view lectures online via videos and interact with the professor through email, MS Teams or Blackboard Collaborate Ultra, and (possibly) discussion boards. The professor will post regular announcements, provide individual feedback to students, and hold online office hours via MS Teams or Blackboard Collaborate Ultra.
- Students-to-Student (S2S) Interaction: Students will engage in discussions through email, the discussion board, and MS Teams or Blackboard Collaborate Ultra.
- Student-to-Content (S2C) Interaction: Students will engage with course content by completing assignments and participating in discussion boards and video conference meetings.

The instructor will reply to all feedback in a reasonable amount of time; the same is expected of the students. Specifically,

- **Communication:** Responses to email communication and questions will be provided within 48 hours. In subject, please prefix with “**CSEC771**.” If you have a question about any deliverable (e.g., project, assignment, presentation) with a deadline, you are advised to email me at least 5 days before it is due so that there is sufficient time between response and deadline.
- **Test Grading:** Grades for assignments will be returned within 1 week of due date.

Time Commitment and Planning

Any university graduate course requires a large amount of work outside of lecture. I assume that when you register for this course you will allocate an average of at least 8-10 hours per week, in addition to lectures, to study the textbook material, complete the homework assignments, and prepare for course project. It is your responsibility to manage your workload. If you procrastinate starting your assignments, you may find that you do not have enough time to complete the project or assignments, or that a technology problem may prevent you from completing your assignment. Note that not being able to access a computer or network will not be considered an acceptable excuse for submitting your assignment late.

Assessments

- Project: 50% for project report and code, plus 10% for elevator presentation to class
 - Analysis project OR
 - Dataset must be from given catalog
 - Use NLP methods to present new insights
 - New method (Research) project
 - Problem to be discussed with instructor
 - Example: Translate from a mixed language to standard English
- Quiz: 20%
 - At least 4 based on preceding lectures
- Exam: Research paper reading (10%) and presentation to class (10%) - Total 20%
 - Read a paper accepted at a top NLP / AI conference: ACL 2019-2020, AAAI 2019-2020, IJCAI 2019-2020, NeurIPS 2019-2020. Make a 1-slide summary based on given presentation template.
 - Present a 1-slide summary to class (10%)

Your overall final course letter grade will be determined by your grades on the following assessments.

Tests	1000 points
• Course Project – report, in-class presentation	600 points
• Quiz – best of 4 from 5	200 points
• Final Exam – Paper summary, in-class presentation	200 points
Total	1000 points

Your final grade is based on the total points you have earned over the semester. Letter grades will be assigned as follows:

A	=	[900-1000]
B+	=	[870-899]
B	=	[800-869]
C+	=	[770-799]

C	=	[700-769]
D+	=	[670-699]
D	=	[600-669]
F	=	[0-599]

If everyone performs very well, I do not have a problem with assigning everyone A's. However, poor performance (particularly failure to do project on time) will result in a low grade.

Important Note Regarding Grade Appeals

My teaching assistant or I will grade all assignments. If you have a question about a grade you have received, or you believe that you were graded incorrectly, please see me during **office hours** or set up an online appointment. If you wish to appeal an assignment grade you must do so within one (1) week of my posting the grade to Blackboard. If you want to make a case for re-grading your work based on another student's grade on the same task, I will review and then re-grade your work as well as the other student's work entirely from scratch.

Missing a Quiz: There will be 5 quizzes on announced dates. A student's lowest score from the 5 will be removed and the rest will be considered for assessment. If a person misses 1 quiz, the score of that quiz will be ignored. If a person misses more quizzes, this will impact their assessment.

Missing slot on project or paper presentation: Unless the instructor is informed 1 week in advance, missing the presentation slot will lead to a zero on presentation component.

Missing the project report: A delay of 7 days will be allowed with a penalty of 20% (200 points). No point will be awarded for any delay beyond a week for project report.

Request for Accommodations

The University of South Carolina is committed to providing access to programs and services for qualified students with disabilities. If you are a student with a disability and require accommodation to participate and complete requirements for this class, notify me immediately and contact the Student Disability Resource Center (<http://www.sa.sc.edu/sds>, 1523 Greene Street, LeConte College Room 112A, 803-777-6142, sasds@mailbox.sc.edu) for verification of eligibility and determination of specific accommodations. In addition, please provide me the required accommodation letter from the Student Disability Resource Center. *All course materials are available in alternative format upon request.*

Academic Integrity

University policies and procedures regarding academic integrity are defined in policy STAF 6.25, Academic Responsibility - The Honor Code (see <http://www.sc.edu/policies/ppm/staf625.pdf>). Prohibited behaviors include

plagiarism, cheating, falsification, and complicity. All potential Honor Code violations will be reported to the Office of Student Conduct and Academic Integrity, which has the authority to implement non-academic penalties as described in STAF 6.25. Academic penalties for Honor Code violations in this course range from a zero on the assignment to failure of the course.

In reference to this course, students are expected to do their own work when assignments require individual work. For example, students may not copy the work of others, either manually or electronically, under these conditions. Further, students who allow their work to be copied by others risk violation of the University Honor Code. If situations arise in which the application of the University Honor Code is unclear, students should seek the interpretation of the instructor.

The faculty takes violations of the University Honor Code (<http://www.sc.edu/policies/ppm/staf625.pdf>) seriously. Students are encouraged to review the Honor Code and to understand the consequences of any action that is proven to be a violation of the code.

Remember that the first tenet of the Carolinian Creed is, "I will practice personal and academic integrity."