Computer Science 480/697- Syllabus
Special Topics (Speech and Language Processing), Spring 2024

Instructor Information

Xiaohui Liang, PhD
Xiaohui.Liang@umb.edu
Co-teach: Youxiang Zhu
Youxiang.Zhu001@umb.edu
Office Hours: Tu/Th 1:00 PM - 2:00 PM or by appointment
Class time: Tu/Th 4:00 PM - 5:15 PM

Note: The following link will assist you in forwarding your UMB email account to your personal account: Use this link. Throughout the semester, I will communicate with you via your UMB *e*-mail account. You may have e-mail redirected from your official UMass Boston address to another *e*-mail address at your own risk. The University will not be responsible for the handling of *e*-mail by outside vendors or by departmental servers.

Course Information

Course Title: Speech and Language Processing

Credits: 3

Description: This course offers a comprehensive exploration of the vibrant landscape of natural language processing (NLP) and speech technologies, spanning from foundational principles to the latest advancements. Throughout the 14-week curriculum, students will dive deep into topics including deep learning in NLP, tokenization, machine translation, sentiment analysis, question answering systems, text summarization, large language models like GPT-4, and speech processing techniques such as automatic speech recognition, speaker identification, and text-to-speech synthesis. Enriched with practical applications, real-world challenges, and ethical considerations, this course not only equips learners with technical proficiency but also fosters critical thinking about the societal impacts of these technologies. As the course culminates, students will gain insights into emerging trends, thereby positioning them at the forefront of speech and language processing innovations.

Context: This course serves one of the electives in computer science.

Prerequisites: CS310 or permission by lecturer

Course Rubric:

Tests/Assignments/Deliverables	Number	Grade %
1. Assignments/Projects	**5**	**60%**
2. Paper Presentation	**1**	**30%**
3. Attendance/participation	**1**	**10%**

Assignment Description:

1. Sentiment Analysis of Movie Reviews

- **Guidelines**:
  - Use a dataset of movie reviews (e.g., IMDB dataset).
  - Preprocess the dataset: Tokenization, stopword removal, and stemming/lemmatization.
  - Build a model to classify the reviews into positive or negative sentiment.
  - Compare the performance of traditional ML methods (e.g., Naive Bayes, SVM) with a deep learning approach (e.g., a basic RNN or LSTM).

- **Deliverables**:
  - Code (Jupyter notebook or Python script).
  - Report detailing the preprocessing steps, choice of models, and evaluation metrics.
  - Presentation summarizing the findings.

- **Evaluation Criteria**:
  - Quality of data preprocessing.
  - Model performance metrics (accuracy, F1 score, etc.).
  - Clarity and depth of the report and presentation.

---

2. Text Summarization using Transformer Models

- **Guidelines**:
  - Choose a set of long-form articles or documents.
  - Implement both extractive and abstractive summarization techniques.
  - Explore the use of a transformer model like BERT or T5 for abstractive summarization.

- **Deliverables**:
  - Code (Jupyter notebook or Python script).
  - Report comparing the pros and cons of extractive vs. abstractive methods, the challenges faced, and the results.
  - Presentation of key findings.

- **Evaluation Criteria**:
  - Quality of summaries (coherence, relevance).
  - Implementation of both summarization techniques.

  - Depth of analysis in the report.

---

3. Question-Answering System on Custom Dataset

- **Guidelines**:
  - Choose or create a dataset, e.g., a set of science textbooks or company manuals.
  - Implement a question-answering system.
  - Use models like BERT or DistilBERT for the task.

- **Deliverables**:
  - Code (Jupyter notebook or Python script).
  - Report detailing the dataset, model architecture, and performance.
  - Presentation showcasing a demo and findings.

- **Evaluation Criteria**:
  - Accuracy and relevance of answers.
  - Quality and clarity of code.
  - Depth of analysis in the report.

Speech Projects:

4. Speech-to-Text Transcription Service

- **Guidelines**:
  - Use a dataset of audio clips (e.g., LibriSpeech).
  - Implement a system to convert speech to text.
  - Experiment with models like Wav2Vec or DeepSpeech.

- **Deliverables**:
  - Code (Jupyter notebook or Python script).
  - Report detailing the preprocessing steps, choice of models, and evaluation metrics.
  - Presentation summarizing the findings.

- **Evaluation Criteria**:
  - Accuracy of transcription.
  - Quality and clarity of code.
  - Depth of analysis in the report.

---

5. Speaker Identification System

- **Guidelines**:
  - Use a dataset with multiple speakers, like VoxCeleb.
  - Implement a system to identify the speaker based on their voice.
  - Explore feature extraction techniques and deep learning models for identification.

- **Deliverables**:
  - Code (Jupyter notebook or Python script).
  - Report discussing the methodology, challenges, and results.
  - Presentation showcasing a demo and findings.

- **Evaluation Criteria**:
  - Accuracy of speaker identification.
  - Quality and clarity of code.
  - Depth of analysis in the report.

Paper presentation:

Students will present and discuss the most recent papers about speech and language processing. Papers can be picked by students or suggested by instructors. Each presentation can be 10-20 minutes.

Course Policies:

Attendance – A roll call will be conducted at the end of each class. The attendance score is calculated as the number of the attended classes divided by the total number of the classes.

Assignments submission– For assignments, no late submissions are accepted unless you have made prior arrangements with me.

Collaboration Policy- You are allowed and encouraged to collaborate with anybody, use any external rescources, and use any AI tools including ChatGPT. However, proper citation or acknowledgment should be given if you do so.

Grading

Grading: Grade type for the course is a whole or partial letter grade. (Please see table below)

Grading Policy for Undergraduate Students		
Letter Grade	**Percentage**	
A	90-100%	
A-	87-89%	
B+	84-86%	
B	80-83%	
B-	77-79%	
C+	74-76%	
C	70-73%	
C-	67-69%	
D+	64-66%	
D	60-63%	
F	<60%	

Note: the lowest passing grade for a graduate student is a “C”. Grades lower than a “C” that are submitted by faculty will automatically be recorded as an “F”.
Please see the Graduate Catalog or website for more detailed information on the University’s grading policy.

Grading Policy for Graduate Students		
Letter Grade	**Percentage**	**Quality Points**
A	93-100%	4.00
A-	90-92%	3.70
B+	87-89%	3.30
B	83-86%	3.00
B-	80-82%	2.70
C+	77-79%	2.30
C	73-76%	2.00
F	0-72%	0.0
W	Received if withdrawal occurs before the withdrawal deadline.	N/A
AU	Audit (only permitted on space-available basis)	N/A
NA	Not Attending (student appeared on roster, but never attended class. Student is still responsible for tuition and fee charges unless withdrawal form is submitted before deadline. NA has no effect on cumulative GPA.)	N/A

Text book:
- **Speech and Language Processing (3rd ed. draft), by Dan Jurafsky and James H. Martin**

Course Schedule

Week 1: Introduction to Speech and Language Processing

Core Topic(s):
1. Overview of the course and the field's historical evolution
2. Introduction to deep learning: Why deep learning outperforms traditional methods
3. Supervised vs. unsupervised learning: Advantages, disadvantages, and applications
4. Deep dive into loss functions: Mean Squared Error, Cross-Entropy, and others
5. Evaluating models: Accuracy, Precision, Recall, F1-Score, ROC, and AUC

Learning Objectives:
1. Understand the foundational concepts of speech and language processing.
2. Recognize the impact and advantages of deep learning in NLP tasks.
3. Differentiate between supervised and unsupervised learning and their applications.
4. Grasp the importance of selecting appropriate loss functions and evaluation metrics.

Week 2: NLP Basics - Part 1

Core Topic(s):
1. Introduction to regular expressions: Patterns, metacharacters, and practical usage
2. Edit distance in text processing: Applications in spell correction and DNA sequence analysis
3. Part of speech tagging: Definition, algorithms, and challenges
4. Comprehensive guide to tokenizing: Segmenting, normalizing, stemming, and lemmatization

Learning Objectives:
1. Gain proficiency in using regular expressions for text manipulation.
2. Understand the significance of edit distance in various applications.
3. Learn the importance of part-of-speech tagging in understanding sentence structures.
4. Master the techniques for tokenizing and preprocessing text data for NLP tasks.

Week 3: NLP Basics - Part 2

Core Topic(s):
1. Word embeddings: From one-hot encodings to Word2Vec, GloVe, and FastText
2. Introduction to language models: From Markov chains to neural models

3. N-grams: Definition, applications, and challenges
4. Evaluating language models: Perplexity, overfitting, and smoothing techniques

Learning Objectives:
1. Comprehend the evolution and significance of word embeddings in capturing semantic meanings.
2. Understand the underlying principles of language models and their applications.
3. Dive into the concept of N-grams and their role in statistical language modeling.
4. Learn to evaluate and refine language models effectively.

Week 4: Machine Translation and Transformers

Core Topic(s):
1. History and evolution of machine translation: Rule-based, statistical, and neural approaches
2. Introduction to transformers in NLP: Why they revolutionized the field
3. RNNs vs. Transformers: Understanding the shift
4. Delving deep into transformer architecture: Self-attention, multi-head attention, positional encoding
5. Encoder-decoder models in translation

Learning Objectives:
1. Trace the progress of machine translation techniques over the decades.
2. Understand the groundbreaking impact of transformer architecture in NLP.
3. Recognize the limitations of RNNs and advantages of transformers.
4. Deeply understand the components and working of the transformer architecture.

Week 5: Sentiment Analysis and BERT

Core Topic(s):
1. Sentiment analysis: From rule-based to deep learning techniques
2. BERT's architecture: Attention mechanisms and token representations
3. Fine-tuning BERT for sentiment analysis: Practical walkthrough
4. Limitations and challenges of sentiment analysis: Sarcasm, context, and multi-domain sentiments
5. Beyond BERT: RoBERTa, DistilBERT, and other variations

Learning Objectives:
1. Grasp the core concepts of sentiment analysis and its importance in real-world applications.
2. Dive deep into BERT's architecture and understand its components.
3. Learn to utilize pre-trained models like BERT for specific NLP tasks.
4. Understand the challenges faced in sentiment analysis and ways to mitigate them.

Week 6: Question and Answering Systems

Core Topic(s):
1. Historical perspective: From rule-based to deep learning-driven QA systems
2. Techniques and models for building QA systems: Retrieval-based and generative models
3. Drilling down on datasets: SQuAD, Natural Questions, and others
4. Challenges in QA: Ambiguity, multi-step reasoning, and context preservation
5. Real-world applications: Search engines, virtual assistants, and customer support bots

Learning Objectives:
1. Understand the progress and significance of QA systems in information retrieval.
2. Learn about various models and techniques used in modern QA systems.
3. Familiarize oneself with popular datasets and benchmarks in QA.
4. Recognize the challenges and ongoing research in the QA domain.

Week 7: Text Summarization Techniques

Core Topic(s):
1. Text summarization: Extractive vs. abstractive approaches
2. Metrics for evaluating summaries: ROUGE, BLEU, METEOR
3. Advanced models in summarization: GPT-2, T5, BART, and their comparison
4. Challenges in summarization: Maintaining coherence, avoiding bias, and handling long documents
5. Real-world use-cases: News, research papers, and content aggregation

Learning Objectives:
1. Grasp the techniques and importance of text summarization.
2. Understand the metrics and their significance in evaluating summarization models.
3. Dive deep into the workings of advanced models like GPT-2, T5, and BART.
4. Recognize challenges in summarization and potential solutions.

Week 8: Large Language Models and In-Context Learning

Core Topic(s):
1. The rise of large language models: GPT-3, ChatGPT, GPT-4
2. In-context learning: Definition, advantages, and challenges
3. Multi-modal models: CLIP, GPT-4, LLaVA, integrating vision and language
4. Ethical implications: Bias in models, misinformation, and responsible usage
5. Future prospects: Few-shot learning, efficiency, and domain-specific LLMs

Learning Objectives:
1. Understand the capabilities and significance of large language models.
2. Dive into the concept of in-context learning and its advantages.
3. Explore models integrating multiple modalities like vision and language.

4. Reflect on the ethical concerns and challenges posed by LLMs.

Week 9: Speech Basics

Core Topic(s):
1. Introduction to speech processing: Evolution and importance
2. Speech signals: Fundamentals, spectrograms, and feature extraction
3. Speech vs. text: Unique challenges in processing audio data
4. Overview of SUPERB: Benchmarks in speech processing
5. Aspects of speech: Content, speaker identification, semantics, emotion detection

Learning Objectives:
1. Gain foundational knowledge in the field of speech processing.
2. Understand the nature of speech signals and techniques for their analysis.
3. Recognize the unique challenges posed by audio data compared to text.
4. Familiarize oneself with benchmarks and their role in the speech domain.

Week 10: Automatic Speech Recognition (ASR)

Core Topic(s):
1. Evolution of ASR: From template matching to deep learning
2. Core components of modern ASR: Acoustic model, language model, and pronunciation model
3. Models in ASR: DeepSpeech, Wav2Vec, RNN-Transducer
4. Challenges in ASR: Handling accents, noise, and real-world scenarios
5. Applications: Transcription services, voice assistants, and accessibility tools

Learning Objectives:
1. Understand the foundational concepts and historical progression of ASR.
2. Dive deep into the workings of modern ASR systems and their components.
3. Explore state-of-the-art models and their applications in ASR.
4. Recognize the challenges in real-world ASR applications and potential solutions.

Week 11: Speaker Recognition and Identification

Core Topic(s):
1. Introduction to speaker recognition: Verification vs. identification
2. Techniques in speaker recognition: Feature extraction, modeling, and classification
3. Datasets: VoxCeleb1, VoxCeleb2, and their significance
4. Challenges: Handling variability, imposters, and real-world scenarios
5. Advanced topics: Speaker diarization, emotion detection, and voice biometrics

Learning Objectives:
1. Grasp the core concepts behind speaker recognition and its real-world implications.
2. Dive into the techniques used for recognizing and distinguishing between speakers.
3. Familiarize oneself with benchmark datasets in speaker recognition.
4. Understand the challenges and frontiers in speaker recognition research.

Week 12: Voice Conversion Techniques

Core Topic(s):
1. Introduction to voice conversion
2. Techniques and applications of voice conversion
3. Voice transformation techniques: Spectral, Prosodic, and Waveform modifications
4. Neural-based voice conversion: CycleGAN, StarGAN
5. Challenges in maintaining speaker identity during conversion
6. Real-world applications: dubbing, voice assistants customization, and voice identity protection.

Learning Objectives:
1. Understand the foundational concepts behind voice conversion and its significance.
2. Explore both traditional and neural-based voice conversion techniques.
3. Delve into the intricacies of preserving speaker uniqueness during voice transformation.
4. Recognize the potential real-world applications and ethical implications of voice conversion.

Week 13: Text-to-Speech Synthesis

Core Topic(s):
1. Basics of text-to-speech (TTS) synthesis
2. Evolution of TTS: Concatenative, Parametric, and Neural TTS
3. Deep learning models in TTS: Tacotron, FastSpeech, WaveNet
4. Expressive TTS: Adding emotions and styles to synthesized speech
5. Applications: Assistive technologies, audiobooks, and voice assistants
6. Challenges: Natural prosody generation, multilingual TTS, and low-resource languages

Learning Objectives:
1. Acquire a comprehensive understanding of TTS synthesis and its historical progression.
2. Explore the leap from traditional to deep learning-based TTS techniques.
3. Understand the role of expressive TTS in generating emotion-rich synthesized speech.
4. Identify real-world applications where TTS has a significant impact.
5. Recognize the challenges and research frontiers in producing natural and diverse sounding speech from text.

Week 14: Course Recap and Future Directions

Core Topic(s):
1. Comprehensive recap of key concepts from the course: From foundational NLP techniques to advanced speech processing.
2. Reflective discussions: Real-world applications and impacts of speech and language processing.
3. Looking ahead: Emerging trends in speech and language processing.
   - Zero-shot and few-shot learning
   - The intersection of neuroscience and NLP
   - Federated learning for NLP
   - Continual and lifelong learning in language models
   - The rise of unsupervised and self-supervised methodologies
4. Ethical considerations and responsible AI:
   - Bias and fairness in NLP models
   - Privacy concerns in speech processing
   - The role of NLP in misinformation and content moderation
   - Sustainable AI: Energy consumption of large models

Learning Objectives:
1. Consolidate and reinforce knowledge acquired throughout the course.
2. Engage in critical discussions on the applications, challenges, and societal impacts of speech and language processing technologies.
3. Familiarize oneself with the cutting-edge trends and research directions in the field.
4. Reflect on ethical considerations and the importance of responsible AI in the development and deployment of NLP systems.

Methods of Instruction

Methods:

- Lecture slides via projector
- Handwriting on blackboard
- Live demo on Python programs
- In-class discussion
- Online discussion in Piazza

Accommodations

The University of Massachusetts Boston is committed to providing reasonable academic accommodations for all students with disabilities. This syllabus is available in alternate format upon request. Students with disabilities who need accommodations in this course must contact the instructor to discuss needed accommodations. Accommodations will be provided after the student has met with the instructor to request accommodations. Students must be registered with the Ross Center for Disability Services, CC UL

211 (617.287.7430) before requesting accommodations from the instructor. http://www.umb.edu/academics/vpass/disability/. After registration with the Ross Center, a student should present and discuss the accommodations with the professor. Although a student can request accommodations at any time, we recommend that students inform the professor of the need for accommodations by the end of the Drop/Add period to ensure that accommodations are available for the entirety of the course.

Academic Integrity and The Code of Student Conduct

It is the expressed policy of the University that every aspect of academic life not only formal coursework situations, but all relationships and interactions connected to the educational process shall be conducted in an absolutely and uncompromisingly honest manner. The University presupposes that any submission of work for academic credit indicates that the work is the student's own and is in compliance with University policies. In cases where academic dishonesty is discovered after completion of a course or degree program, sanctions may be imposed retroactively, up to and including revocation of the degree. Any student who reasonably believes another student has committed an act of academic dishonesty should inform the course instructor of the alleged violation. These policies are spelled out in the Code of Student Conduct. Students are required to adhere to the Code of Student Conduct, including requirements for academic honesty, as delineated in the University of Massachusetts Boston Graduate Catalogue and on their Website and in relevant program student handbook(s) or websites. UMB Code of Student Conduct
You are encouraged to visit and review the UMass website on Plagiarism*:* Plagiarism Prevention & Awareness: Home

Other Pertinent and Important Information

You are advised to retain a copy of this syllabus in your personal files for use when applying for future degrees, certification, licensure, or transfer of credit.