INT344:NATURAL LANGUAGE PROCESSING

L:2 T:0 P:1 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: study the fundamentals of Natural Language Processing, linguistic components, and basic text preprocessing tasks.

CO2 :: elaborate vector space models and transformation techniques for semantic similarity, word relationships, and document search.

CO3 :: determine the probabilistic models for tasks such as spell correction, POS tagging, and word prediction.

CO4 :: investigate the classification-based models (Naïve Bayes, Logistic Regression) for sentiment analysis.

CO5 :: evaluate sequence models (RNN, LSTM, Attention) and perform NLP tasks such as NER, POS tagging, and summarization.

CO6 :: develop and evaluate end-to-end NLP systems using transformer-based architectures for applications like chatbots and QA systems.

Unit I

Introduction to NLP and Text Processing : NLP introduction, origin of NLP, Language and Knowledge, The challenges of NLP, Language and Grammar, NLP applications, Linguistic essentials, Morphology, Syntax, Semantics, Basic Text Processing, Tokenization, Lemmatization, Stemming, Stop words, Capturing word dependency using TF-IDF

Unit II

Vector Space Models : Vector Space Models, Capture Semantic Meaning, Continues Bag-of-Words, Relationships between, Relationships between Words, Capture Dependencies between Words, Visualize the Relationships in Two Dimensions Using PCA, Machine Translation and Document Search, Transform Word Vectors

Unit III

Natural Language Processing with Probabilistic Models : Autocorrect: Minimum Edit Distance, Spellchecker to Correct Misspelled Words, Part of Speech Tagging and Hidden Markov Models: About Markov Chains and Hidden Markov Models, Part-Of-Speech Tags using a Text Corpus, Autocomplete and Language Models: N-gram Language Models work by Calculating Sequence Probabilities, Autocomplete Language Model using a Text Corpus, Word Embeddings with Neural Networks: Word Embeddings, Semantic Meaning of Words

Unit IV

Natural Language Processing with Classification Models : Sentiment Analysis with Logistic Regression, Extract Features from Text into Numerical Vectors, Binary Classifier using a Logistic Regression, Sentiment Analysis with Naïve Bayes, Bayes' rule for Conditional Probabilities, Naive Bayes Classifier

Unit V

Natural Language Processing with Sequence and Attention Models : Introduction to Sequence Models, Recurrent Neural Networks and their limitations, Long Short-Term Memory (LSTM), Applications, POS Tagging, Named Entity Recognition, Neural Machine Translation, Shortcomings of a Traditional seq2seq Model, Introduction to Attention Mechanism

Unit VI

Building Models/ Case Studies : Question Answering: Transfer Learning with State-Of-The-Art Models, T5 and Bert, Model for Answering Questions, Chatbot: Examine Unique Challenges, Transformer Models Face and their Solutions, Chatbot using a Reformer Model

List of Practicals / Experiments:

List of Practical/ Experiments

- Apply and compare stemming and lemmatization techniques to normalize text.
- Extract TF-IDF features to capture word importance across multiple documents.
- Project high-dimensional word vectors into 2D space for visual interpretation using dimensionality reduction.

- Build a simple document search system using TF-IDF vectors and cosine similarity.
- Implement a basic autocorrect system that suggests correct spellings based on minimum edit distance.
- Apply HMM for tagging words in a sentence with their respective POS tags.
- Develop unigram, bigram, and trigram models to predict the next word in a sequence (autocomplete).
- Use pretrained Word2Vec or GloVe embeddings to find semantically similar words and perform vector operations.
- Build a binary sentiment classifier (positive/negative) using Logistic Regression.
- Train and evaluate a Naïve Bayes model for text sentiment analysis.
- Identify distinguishing words in positive and negative reviews using frequency analysis.
- Perform a side-by-side comparison of both classifiers on the same dataset.
- Understand the architecture of a simple RNN and implement it for character-level text generation or next-word prediction.
- Use an LSTM model to tag each word in a sentence with its POS tag.
- Use a pre-trained BERT model to build a context-based question answering system.

Text Books:

1. NATURAL LANGUAGE PROCESSING by ELA KUMAR, DREAMTECH PRESS

References:

1. SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS AND SPEECH RECOGNITION by DANIEL JURAFSKY, JAMES H. MARTIN, PEARSON