Summer of Science 2024: Natural Language Processing

Mentee: Abhi Jain
Mentor: Nilesh Choudhary

Plan of Action

Timeline

- **Week 1:** Introduction to NLP and Text Preprocessing
  - Overview of NLP
  - Tokenization
  - Stop-word removal
  - Stemming and Lemmatization
  - Punctuation removal
- **Week 2:** Text Representation
  - Bag of Words
  - Count Vectorization
  - TF-IDF
  - Word Embeddings (Word2Vec, Doc2Vec)
  - n-gram models (uni-gram, bi-gram, tri-gram)
- **Week 3:** Deep Learning for NLP
  - Neural Networks
  - Recurrent Neural Networks (RNNs)
  - Long Short-Term Memory Networks (LSTMs)
- **Week 4:** Attention Mechanism, Seq2seq, and Encoder-decoder models
  - Understanding Attention Mechanism
  - Sequence-to-Sequence (Seq2seq) models
  - Encoder-decoder architecture
- **Week 5:** Transformers and BERT
  - Introduction to Transformers
  - BERT (Bidirectional Encoder Representations from Transformers)
  - Transfer Learning with BERT
- **Midterm Report Submission:** Late June
- **Week 6:** Large Language Models (LLMs)
  - Understanding LLMs
  - Applications and case studies
- **Week 7:** Applications of NLP
  - Overview of NLP applications

- Project analysis and workflow explanation

- **Week 8:** Advanced Topics and Additional Materials (if time permits)
  - Exploring additional advanced topics in NLP
  - Reviewing supplementary materials and lectures

- **Week 9:** Final Review and Preparation for Endterm Submission
  - Comprehensive review of all topics
  - Preparation for final report submission

Final Report Submission: Late July

References

- *YouTube Series:*
  - Deep Learning for NLP
    `https://www.youtube.com/watch?v=ERibwqs9p38&list=PL3FW7Lu3i5Jsnh1rnUwq_TcylNr7EkRe6&index=2`)
  - CS224N: NLP with Deep Learning
    `https://www.youtube.com/watch?v=PLryWeHPcBs&list=PLoROMvodv4rMFqRtEuo6SGjY4XbRIVRd4&index=5`)

- *Paper:*
  - Attention is All You Need
    `https://papers.neurips.cc/paper/7181-attention-is-all-you-need.pdf`)

- *Online Course:*
  - Hugging Face NLP Course
    (`https://huggingface.co/learn/nlp-course/en/chapter1/4?fw=pt`)

- *YouTube Lectures:*
  - What are Transformer Models and how do they work?
    `https://youtu.be/qaWMOYf4ri8?si=sHIzhXrLJj4uqCOj`
  - Long Short-Term Memory (LSTM), Clearly Explained
    `https://youtu.be/YCzL96nL7j0?si=U7rbfVdoxyxddwPD`
  - Sequence-to-Sequence (seq2seq) Encoder-Decoder Neural Networks
    `https://youtu.be/L8HKweZIOmg?si=_fFYPV1sS0bkht4B`
  - Transformer Neural Networks, ChatGPT's foundation
    `https://youtu.be/zxQyTK8quyY?si=1e0M2grZ9nGDhygi`
  - Encoder Decoder — Sequence-to-Sequence Architecture
    `https://youtu.be/KiL74WsgxoA?si=YwSCC7beAvIelwGR`
  - Attention Mechanism in 1 video — Seq2Seq Networks
    `https://youtu.be/rj5V6q6-XUM?si=RrSj_JfaQAqSfTT9`