Chapter 8 (10)
Innleiðing í tilgjørdum viti

Notes

Contents

1 Overview of NLP

Natural Language Processing (NLP) is a field at the intersection of linguistics, computer science, and artificial intelligence. It enables machines to understand, interpret, and generate human language.

1.1 Key Applications

- Automatic Summarisation
- Information Extraction
- Machine Translation
- Question Answering
- Text Classification

2 Challenges in NLP

- **Syntax:** Structure of a sentence.
- **Semantics:** Meaning of words and sentences.

3 Approaches in NLP

3.1 Rule-Based

- Hand-coded rules based on expert knowledge.
- Easy to interpret and debug.
- Does not require large volumes of data.

3.2 Data-Driven

- Infers rules from data.
- Accuracy improves with more data.
- May be biased or hard to interpret.
- More flexible and scalable.

4 Probabilistic Models

4.1 Markov Chains

Used in both rule-based and data-driven systems depending on how transition probabilities are derived (e.g., n-gram frequency).

4.2 Bayes' Rule

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

4.3 Naive Bayes Classifier

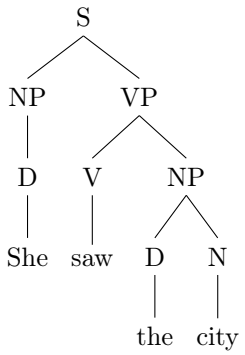
Assumes independence among features to compute the probability of a class given the features.

5 Formal Grammars

5.1 Context-Free Grammar (CFG)

Abstracts meaning from text to represent structure. It uses production rules to define possible sentence structures.

5.2 Syntax Tree Example



6 Word Representation

6.1 Tokenisation

- **Character-level:** Good for spell-checking, compact
- **Word-level:** Intuitive, but memory-intensive
- **Subword-level:** Balanced approach, most commonly used

6.2 One-Hot Encoding

Represents words as binary vectors.

6.3 Word2Vec

Learns vector representations (embeddings) of words based on their context.

6.3.1 Continuous Bag of Words (CBOW)

Predicts target word from context.

CBOW Architecture

- **Input:** One-hot encoded vectors of context words
- **Hidden Layer:** Trains word embeddings
- **Output:** Softmax over vocabulary

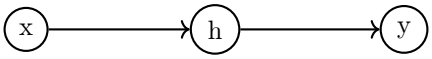
7 Neural Networks in NLP

7.1 Perceptron

A basic unit in neural networks:

$$y = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

7.2 Feedforward Neural Network



7.3 Encoder-Decoder Architecture

Used for sequence-to-sequence tasks like translation.

7.4 Recurrent Neural Networks (RNNs)

Processes sequences step by step and maintains a hidden state.

- Captures sequential dependencies.
- Suffers from vanishing gradients and depth issues.

7.5 Attention Mechanism

Weighs hidden states by importance.

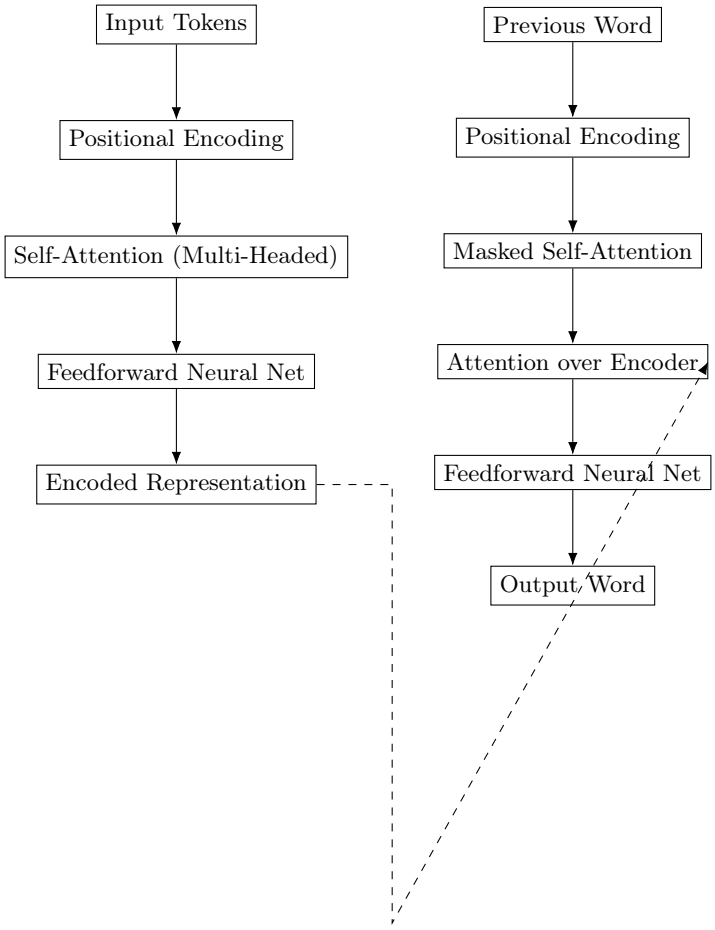
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

8 Transformers

8.1 Key Features

- Self-attention
- Parallel processing
- Positional encoding

8.2 Architecture Diagram



8.3 Advantages

- Enables parallel training.
- Captures long-term dependencies.

8.4 Limitations

- Memory intensive for long sequences.

9 Exam Preparation Summary

- Activation Functions: Step, Sigmoid
- Perceptron = Linear function + Step
- NLP Language Models: N-grams, Neural, Transformers