NLP

1. intro of nlp
2. Tokenization of words / paragraph
3. One-hot encoding of words and characters
4. Using word embeddings
5. Example the IMDB data for use with an Embedding layer
6. DOWNLOADING THE IMDB DATA AS RAW TEXT and download pre trained word Embedding
7. Training the same model without pre trained word embeddings

1.intro

1.The deep-learning sequence-processing models introduced in the following sections can use text to produce a basic form of natural-language understanding, sufficient for applications including document classification, sentiment analysis, author identification, and even question-answering (QA) (in a constrained context)

* Sequence-to-sequence learning, such as decoding an English sentence into French Sentiment analysis, such as classifying the sentiment of tweets or movie reviews as positive or negative
* Time Series forecasting, such as predicting the future weather at a certain location, given recent weather data

2.Deep learning for natural-language processing is pattern recognition applied to words, sentences, and paragraphs.

* Segment text into words, and transform each word into a vector.
* Segment text into characters, and transform each character into a vector

3.the different units into which you can break down text (words, characters, or n-grams) are called tokens, and breaking text into such tokens is called tokenization.

4.There are multiple ways to associate a vector with a token. In this section, I’ll present two major ones: one-hot encoding of tokens, and token embedding (typically used exclusively for words, and called word embedding).

