

Natural Language Processing

Introduction to language modelling

Marco Kuhlmann

Department of Computer and Information Science

Language modelling

- **Language modelling** is the task of predicting which word comes next in a sequence of words.
- More formally, given a sequence of words w_1, \dots, w_t , we want to know the probability of the next word, w_{t+1} :

$$P(w_{t+1} | w_1, \dots, w_t)$$

- We are assuming that w_{t+1} comes from a finite vocabulary V .

language models = classifiers



language modelling

- language modelling
- language modelling **nlp**
- language modelling **using lstm networks**
- language modelling **makes sense**
- language modelling **in python**
- language modelling **with rnn**
- language modelling **pytorch**
- language modelling **approach**
- language modelling **toolkit**
- language modelling **dataset**

Sweden

Advertising Business About

Privacy Terms Settings

Google Search

I'm Feeling Lucky

An alternative view on language models

- Rather than as predictive models, language models can also be viewed as models that assign a probability to a piece of text.

How likely is it that this piece of text is written in Swedish? French?

- These two views are equivalent, as the probability of a sequence can be expressed as a product of conditional probabilities: *

$$P(w_1 \cdots w_N) = \prod_{t=1}^N P(w_t | w_1, \dots, w_{t-1})$$

* There are some subtleties here that we gloss over.

Language models for translation

Rank candidate translations for the following Chinese sentence:

他 向 记者 介绍了 主要 内容

he to reporters introduced main content

- he introduced reporters to the main contents
- he briefed to reporters the main contents
- he briefed reporters on the main contents

Example from Jurafsky and Martin (2017)



ChatGPT 4 (2023-12-31)

Evaluating language models

Eisenstein § 6.4

- **Intrinsic evaluation**

How does the method or model score with respect to a given evaluation measure?

examples from classification: precision and recall

- **Extrinsic evaluation**

How much does the method or model help the application in which it is embedded?

predictive input, machine translation, speech recognition

Perplexity

Eisenstein § 6.4

- Intrinsic evaluation of language models is based on the likelihood that a model assigns to held-out data.
- Formally, we compute the cross-entropy between two probability distributions: a language model and the empirical distribution.
- This cross-entropy is usually presented as **perplexity**:

$$e^{-\frac{1}{N} \log P(w_1 \cdots w_N)}$$

Perplexity

