

Autumn 2024, Part 1

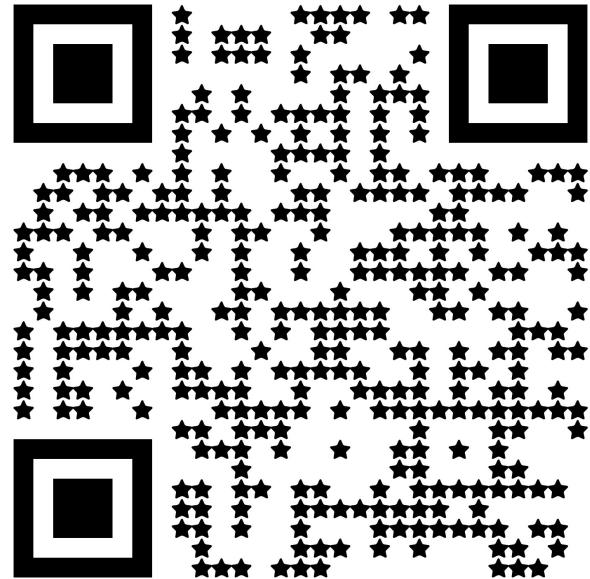
Introduction to NLP

Eliza Vialykh, [@elf_lesnoy](https://twitter.com/elf_lesnoy)



Course program

1. Introduction to NLP
2. Text preprocessing,
Word embeddings
3. Classification and Naive Bayes
4. Language modeling, RNN, LSTM
5. Attention, Transformers
6. Sesame Street
7. LLM Fine-Tuning
8. Inference Optimization for LLMs
9. PEFT
10. Retrieval-Augmented Generation (RAG)



Prerequisites

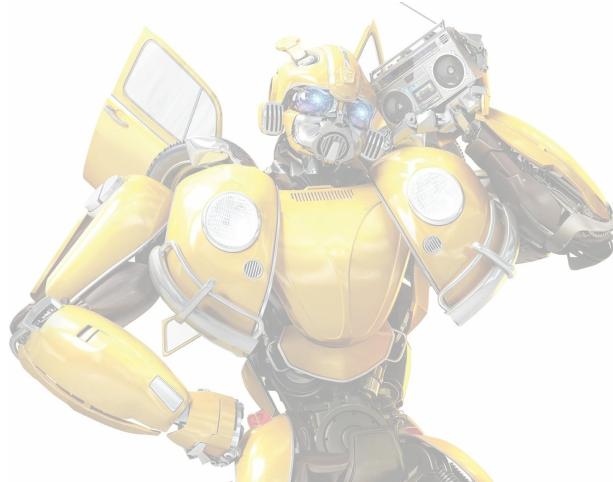
1. Fundamentals of Python Programming
2. Introduction to Machine Learning
3. Mathematics and Statistics for Machine Learning
4. Introduction to Deep Learning

Mark

Tests	HW	Final Project	Exam
10%	30%	20%	40%

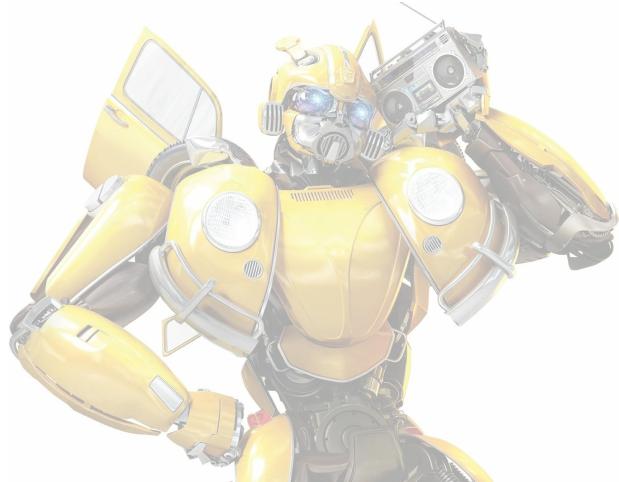
Content

1. Introduction to NLP
2. History
3. Tasks
4. Metrics
5. Pipeline
6. Stack



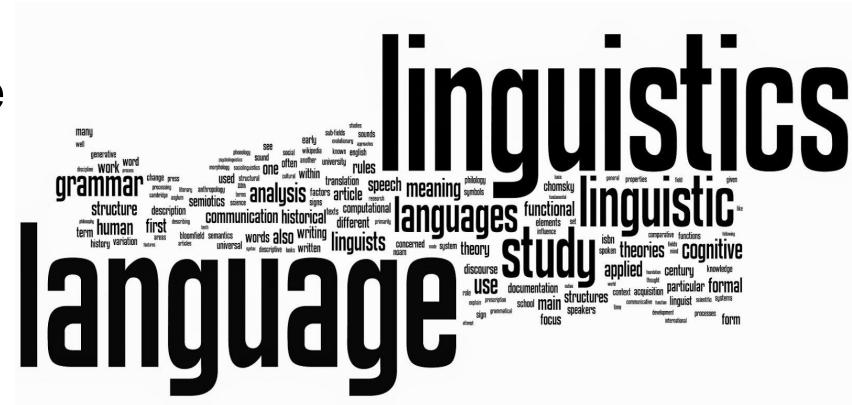
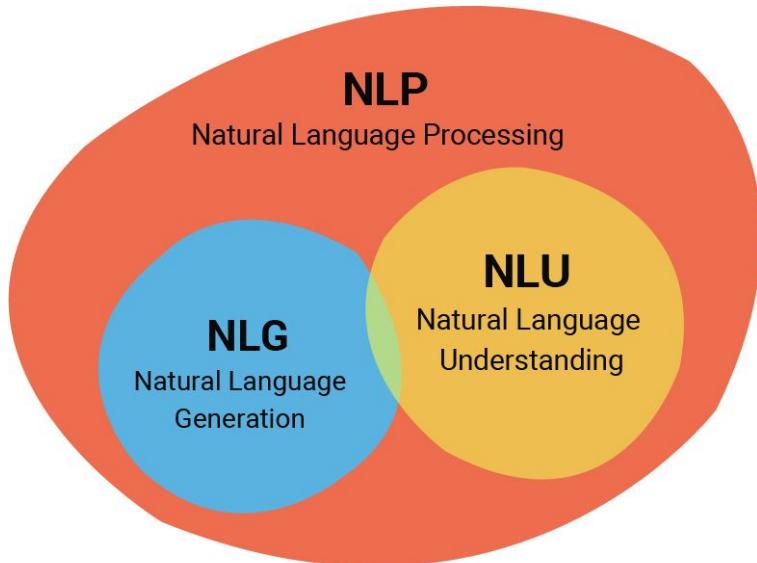


Introduction to NLP



What is NLP?

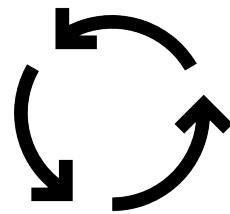
NLP is a field of linguistics and machine learning focused on understanding everything related to human language.



A

NLP approaches

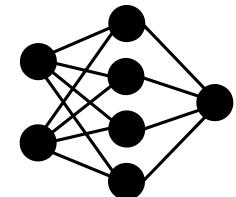
Rules



ML

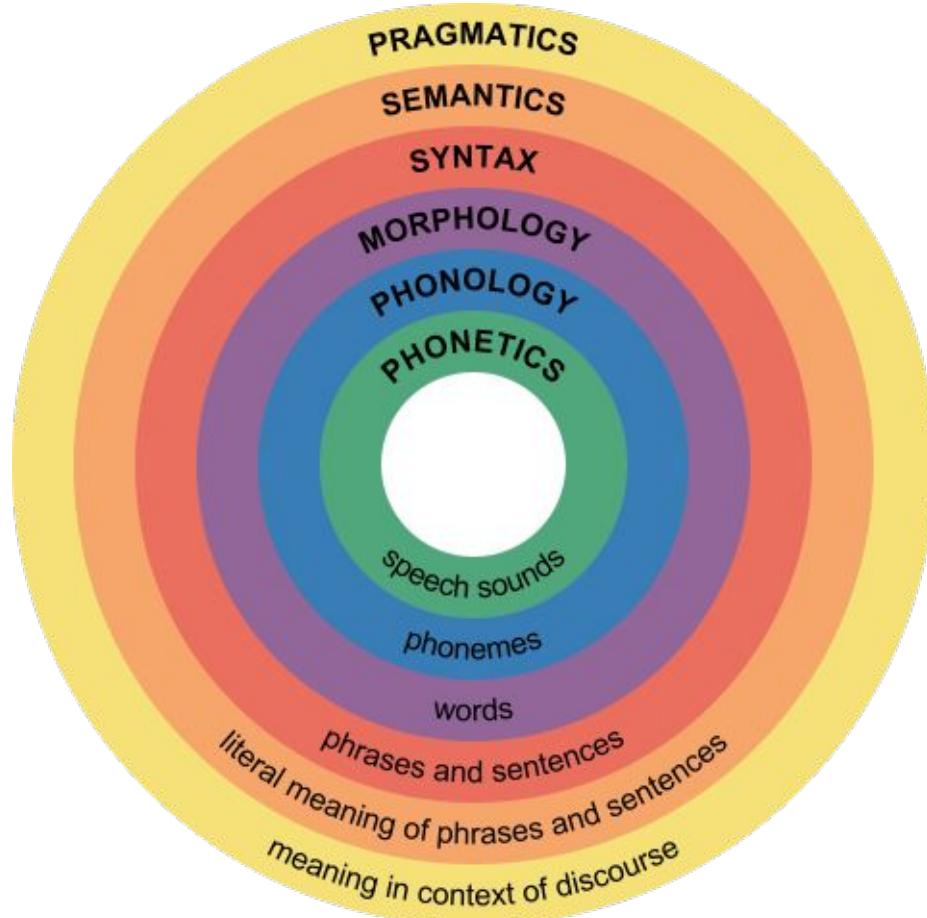


DL



Why is it so difficult?

Linguistics is the scientific study of language. Modern-day linguistics is considered a science because it entails a comprehensive, systematic, objective, and precise analysis of all aspects of language – i.e., **the cognitive, the social, the cultural, the psychological, the environmental, the biological, the literary, the grammatical, the paleographical, and the structural.**



Why is it so difficult?

Теорема 1.1.1. Класс языков, порождаемых грамматиками, совпадает с классом рекурсивно перечислимых языков (стр. 49; Davies 1958 и Гладкий 1966).

Теорема 1.1.2. Класс языков, порождаемых неукорачивающими грамматиками, совпадает с классом НС-языков; более того, для всякой неукорачивающей грамматики можно эффективно построить эквивалентную ей НС-грамматику (стр. 54; Хомский 1966, стр. 165, и Гладкий 1966, стр. 72).

Why is it so difficult?

«Я иду домой» → «Я **домой** иду» → «Домой иду **я**»

Средства
художественной
выразительности

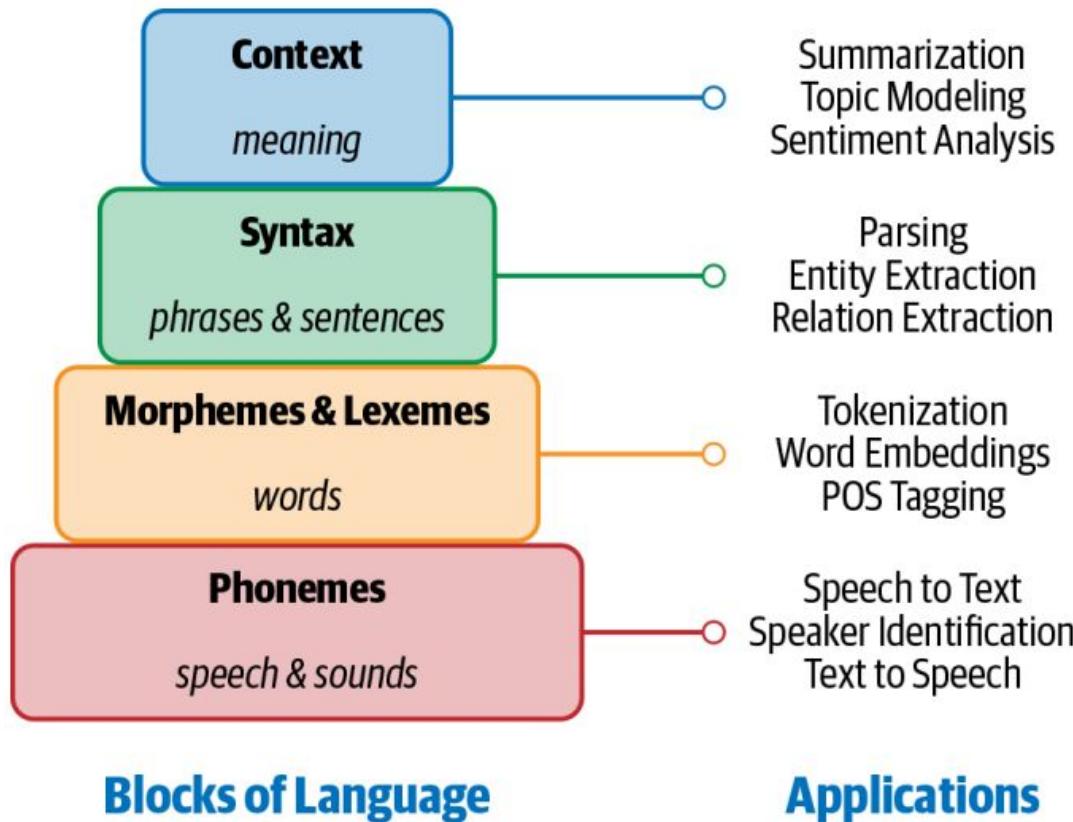


Why is it so difficult?

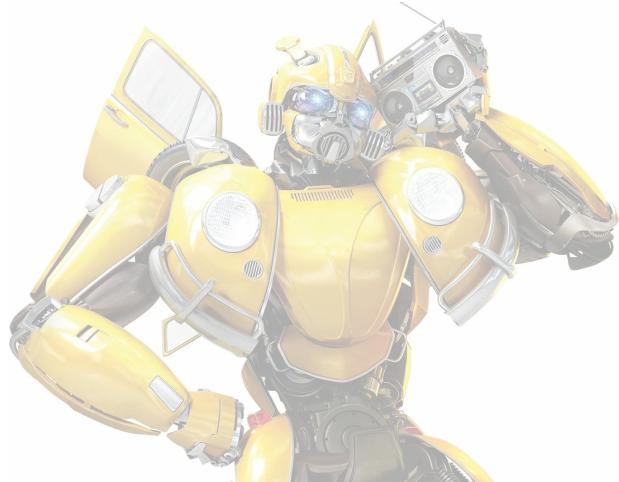
«Руки не доходят посмотреть»



Why is it so difficult?



History



AN NLP TIMELINE AND THE TRANSFORMER FAMILY

BAG OF WORDS (BOW)

Count the occurrences of each word in the documents and use them as features.

1954

TF-IDF

The BOW scores are modified so that rare words have high scores and common words have low scores.

1972

WORD2VEC

Each word is mapped to a high-dimensional vector called word embedding, which captures its semantic. Word embeddings are learned by a neural network looking for word correlations on a large corpus.

2013

RNN

RNNs compute document embeddings leveraging word context in sentences, which was not possible with word embeddings alone.

LSTM

Capture long term dependencies.

1997

Bidirectional RNN

Capture left-to-right and right-to-left dependencies.

1997

Encoder-decoder RNN

An RNN creates a document embedding (i.e. the encoder) and another RNN decodes it into text (i.e. the decoder).

2014

1986

TRANSFORMER

An encoder-decoder model that leverages attention mechanisms to compute better embeddings and to better align output to input.

2017

BERT

Bidirectional Transformer pretrained using a combination of Masked Language Modeling and Next Sentence Prediction objectives. It uses global attention.

2018

GPT

The first autoregressive model based on the Transformer architecture.

GPT-2

A bigger and optimized version of GPT, pre-trained on WebText.

2019

GPT-3

A bigger and optimized version of GPT-2, pre-trained on Common Crawl.

2020

2018

CTRL

Similar to GPT but with control codes for conditional text generation.

2019

TRANSFORMER-XL

It's an autoregressive Transformer which can reuse previously computed hidden-states to attend to longer context.

2019

ALBERT

A lighter version of BERT, where (1) Next Sentence Prediction is replaced by Sentence Order Prediction, and (2) parameter-reduction techniques are used for lower memory consumption and faster training.

2019

ROBERTA

Better version of BERT, where (1) the Masked Language Modeling objective is dynamic, (2) the Next Sentence Prediction objective is dropped, (3) the BPE tokenizer is employed, and (4) better hyperparameters are used.

2019

XLM

Transformer pre-trained on a corpus of several languages using objectives like Causal Language Modeling, Masked Language Modeling, and Translation Language Modeling.

2019

XLNET

Transformer-XL with a generalized autoregressive pre-training method that enables learning bidirectional dependences.

2019

PEGASUS

A bidirectional encoder and a left-to-right decoder pre-trained with Masked Language Modeling and Gap Sentence Generation objectives.

2019

DISTILBERT

Same as BERT but smaller and faster, while preserving over 95% of BERT's performances. Trained by distillation of the pre-trained BERT model.

2019

XLM-ROBERTA

RoBERTa trained on a multilingual corpus with the Masked Language Modeling objective.

2019

BART

A bidirectional encoder and a left-to-right decoder trained by corrupting text with an arbitrary noising function and learning a model to reconstruct the original text.

2019

CONVBERT

Better version of BERT, where self-attention blocks are replaced with new ones that leverage convolutions to better model global and local context.

2019

FUNNEL TRANSFORMER

A type of Transformer that gradually compresses the sequence of hidden states to a shorter one and hence reduces the computation cost.

2020



NLPLANET

The community of
NLP enthusiasts!



<https://www.linkedin.com/company/nlplanet>



<https://medium.com/nlplanet>



[@nlplanet_](https://twitter.com/nlplanet_)

REFORMER

A more efficient Transformer thanks to local-sensitive hashing attention, axial position encoding and other optimizations.

2020

T5

A bidirectional encoder and a left-to-right decoder pre-trained on a mix of unsupervised and supervised tasks.

2020

LONGFORMER

A Transformer model replacing the attention matrices with sparse matrices for higher training efficiency.

2020

PROPHETNET

A Transformer model trained with the Future N-gram Prediction objective and with a novel self-attention mechanism.

2020

ELECTRA

Same as BERT but lighter and better. The model is trained with the Replaced Token Detection objective.

2020

SWITCH TRANSFORMER

A sparsely-activated expert Transformer model that aims to simplify and improve over Mixture of Experts.

2021

~1950s-1970s

Rule-based Systems

~1980s-1990s

Statistical Approach

~2000s-2010s

Machine Learning



ELIZA is an early NLP computer program developed from 1964 to 1967 at MIT by Joseph Weizenbaum.

```
Welcome to
      EEEEEE  LL      IIII    ZZZZZZ  AAAAAA
      EE      LL      II      ZZ      AA      AA
      EEEEEE  LL      II      ZZZ     AAAAAAAA
      EE      LL      II      ZZ      AA      AA
      EEEEEE  LLLLLL  IIII    ZZZZZZ  AA      AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```



Eliza

Regex

User: I am unhappy.

ELIZA: DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

User: I need some help, that much seems certain.

ELIZA: WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

User: Perhaps I could learn to get along with my mother.

ELIZA: TELL ME MORE ABOUT YOUR FAMILY

User: My mother takes care of me.

ELIZA: WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

User: My father.

ELIZA: YOUR FATHER

User: You are like my father in some ways.

Weizenbaum (1966)

“I need X” —> “What would it mean to you if you got X?”.
https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Eliza

Regex

User: I am unhappy.

ELIZA: DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

User: I need some help, that much seems certain.

ELIZA: WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

User: Perhaps I could learn to get along with my mother.

ELIZA: TELL ME MORE ABOUT YOUR FAMILY

User: My mother takes care of me.

ELIZA: WHO ELSE IN YOU FAMILY TAKES CARE OF YOU

User: My father.

```
s/.* YOU ARE (depressed|sad) .*/I AM SORRY TO HEAR YOU ARE \1/  
s/.* YOU ARE (depressed|sad) .*/WHY DO YOU THINK YOU ARE \1/  
s/.* all .*/IN WHAT WAY/  
s/.* always .*/CAN YOU THINK OF A SPECIFIC EXAMPLE/
```

SHRDLU

SHRDLU is an early natural-language understanding computer program that was developed by Terry Winograd at MIT in **1968–1970**.

Person: PICK UP A BIG RED BLOCK.

Computer: OK. (does it)

Person: GRASP THE PYRAMID.

Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.

Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.

Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Computer: OK. (does it)

Person: WHAT DOES THE BOX CONTAIN?

Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

Person: WHAT IS THE PYRAMID SUPPORTED BY?

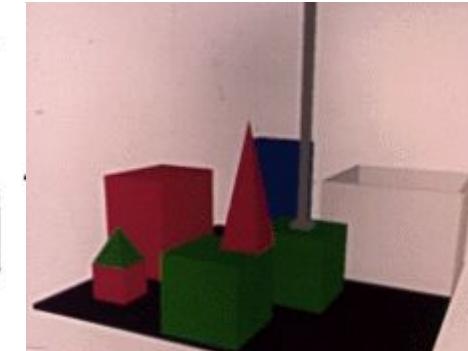
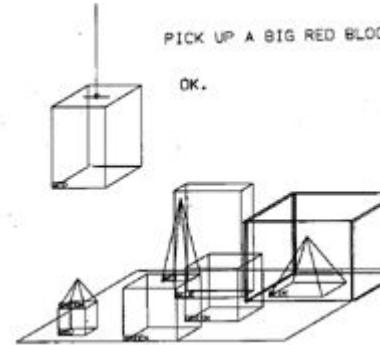
Computer: THE BOX.

Person: HOW MANY BLOCKS ARE NOT IN THE BOX?

Computer: FOUR OF THEM.

Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?

Computer: YES, THE RED CUBE.





2017

~2017s-...



arXiv > cs > arXiv:1706.03762

Computer Science > Computation and Language

[Submitted on 12 Jun 2017 (v1), last revised 2 Aug 2023 (this version, v7)]

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolution. On two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves state-of-the-art performance on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model achieves a BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer translates text successfully to English constituency parsing both with large and limited training data.

Comments: 15 pages, 5 figures

Subjects: Computation and Language (cs.CL); Machine Learning (cs.LG)

Cite as: arXiv:1706.03762 [cs.CL]

(or arXiv:1706.03762v7 [cs.CL] for this version)

<https://doi.org/10.48550/arXiv.1706.03762> 

2018

2019

2020

2021

GPT

GPT-2

T5

GPT-3

XLM

ALBERT

M2M100

BERT

RoBERTa

DeBERTa

XLNet

BART

LUKE

DistilBERT

Longformer

June 2018: GPT, the first pretrained Transformer model, used for fine-tuning on various NLP tasks and obtained state-of-the-art results

Hello GPT-4o

Придумай стихотворение об искусственном интеллекте и еноте

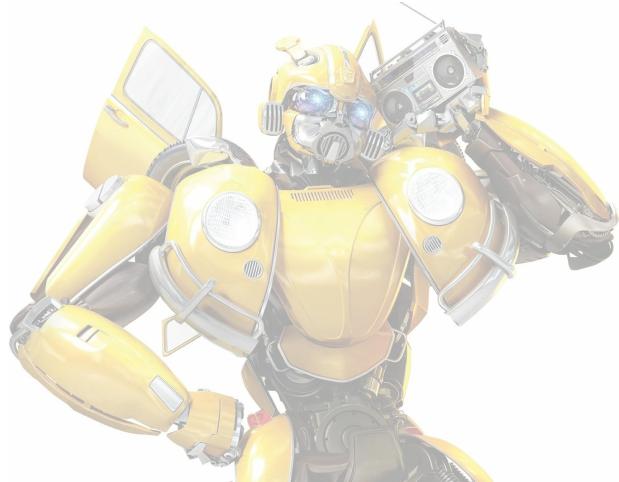


В лесу под звёздным небом ночи
Жил енот, он был очень точен:
Во всём, что делал, был педант,
Прямо как интеллект искусственный, гигант.

Енот был мудрым, любил читать,
И книжки с техникой листать.
Однажды в лес ему привезли
Компьютер новый, полон магии.

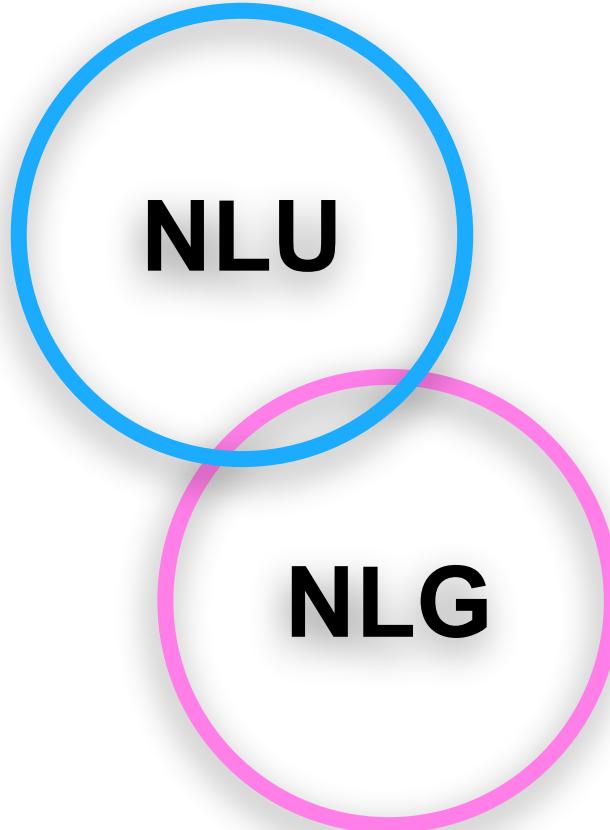


Tasks



Tasks

1. Text Classification
2. Text Similarity
3. Machine-learned Ranking
4. Information Retrieval
5. NER
6. Chatbots
7. Text Summarization
8. Machine Translation



Text Classification

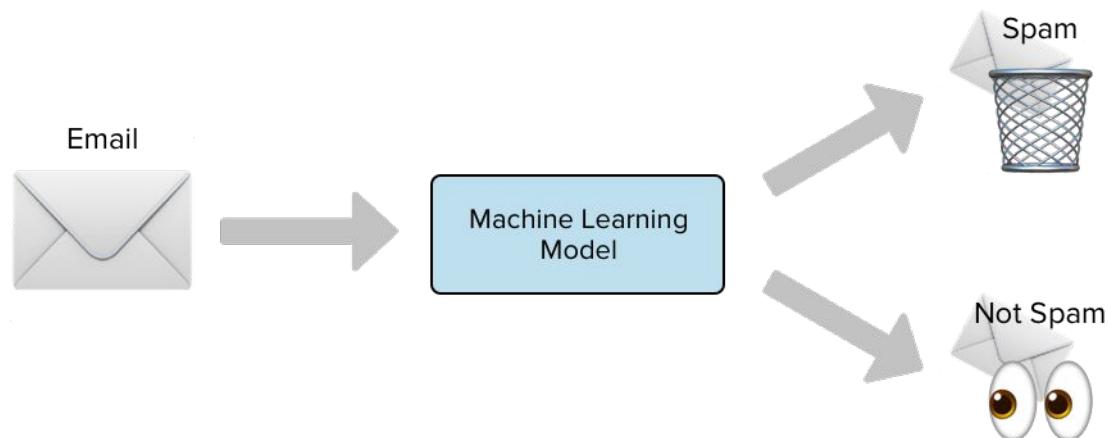
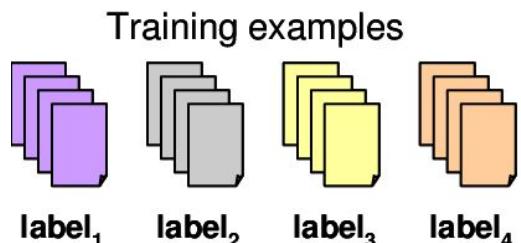
1. Sentiment analysis
2. Author identification
3. Organizing documents by topic
4. Triaging inbound emails



Positive

Negative

Neutral



Text Similarity

Vacancy

Satellite Nepal Pvt. Ltd. an Authorized Distributor of Ncell Pvt. Ltd. is seeking to fulfil vacancy for the following position;

Position: MARKETING REPRESENTATIVE

Location:

- Mechi-1, Koshi-1, Sagarmatha-1, Janakpur-1, Narayani-1, Bagmati-1, Lumbini-1
- Gandaki/Dhaulagiri-1 (Candidate must be citizen of Gandaki zone)
- Karnali/Rapti/Bheri-1 (Candidate must be citizen of Bheri zone)
- Seti/Mahakali-1 (Candidate must be citizen of Mahakali zone)

Required Qualification and skills.

1. Minimum +2 passed (Bachelor in Marketing preferred) having 2- 4 years



Marjorie D. McGahey

Sale Executive

PROFILE

- ✉ April 4, 1987
- ✉ Mail
- 📞 718-564-6972
- ✉ marjorie@jourapide.com
- 📍 526 Longview Avenue, Brooklyn, NY
- 🌐 http://marjorie-info.me

OBJECTIVE

Take advantages of sales skills & experience and

EDUCATION

Sep 2010 - June 2015
FOREIGN TRADE UNIVERSITY
Major: Economics and International Business
GPA: 7.34/10

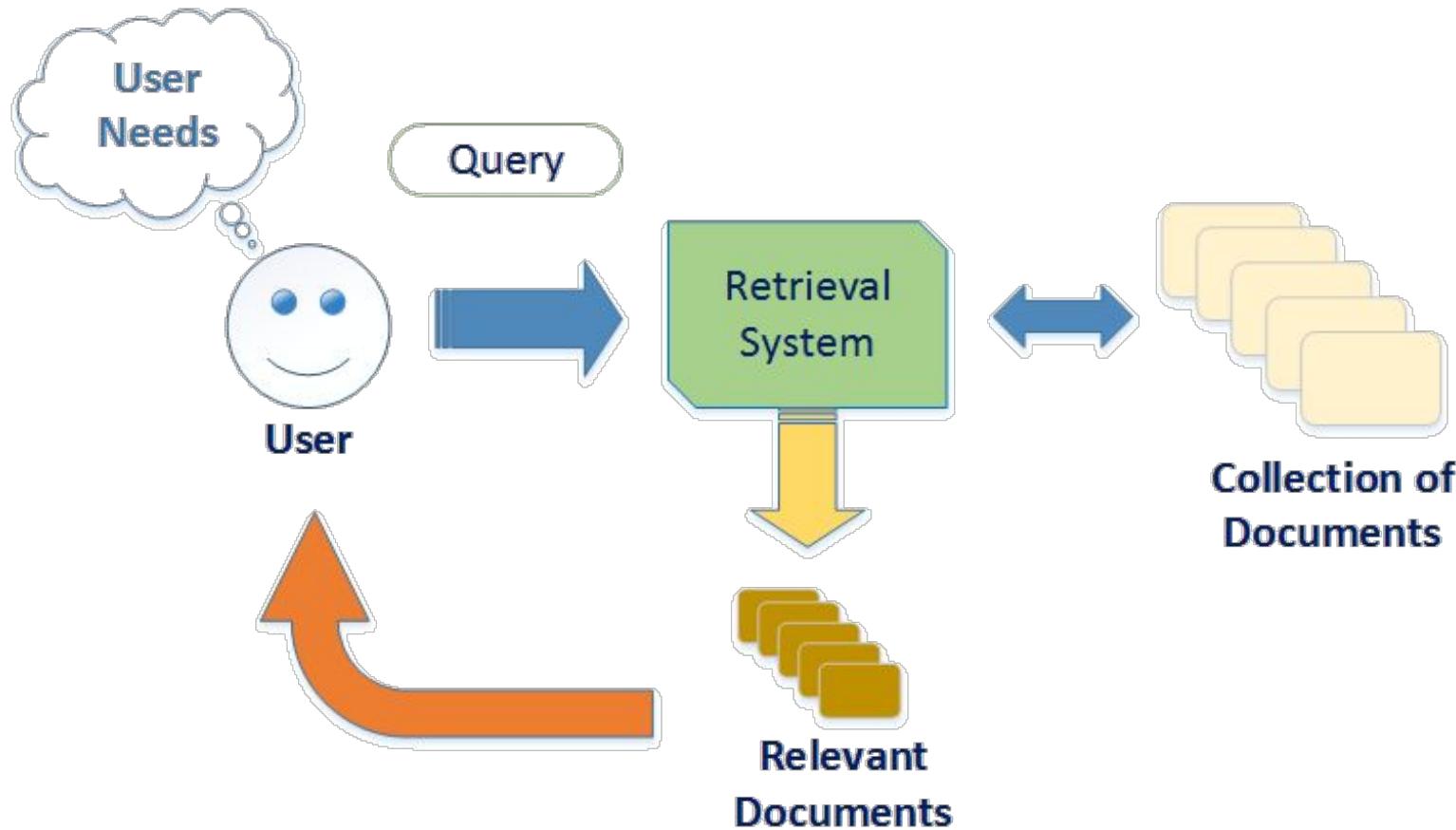
Sep 2012 - June 2014
UNIVERSITY OF ECONOMICS
Major: Chief accountant
GPA: 7.5/10

WORK EXPERIENCE

May 2011 - Now
UHE TRADING COMPANY
Sales Executive
- Manage a retail shop in NeyOm province
- Holding on current customers and develop new distributors
- Handling the problem related to distributors: Display, discount, event plan, network development.

Jul 2011 - Dec 2013
IPSOS BUSINESS CONSULTING VIETNAM CO.
Marketing Executive
- Interviewing users according to content of different

Information Retrieval



ML Ranking

Google

how to launch the rocket



NASA Space Place – NASA Science for Kids (.gov)

<https://spaceplace.nasa.gov> > ... · Перевести эту страницу



How Do We Launch Things Into Space?

We **launch** satellites and spacecraft into space by putting them on **rockets** carrying tons of propellants. The propellants give the **rocket** enough energy to boost ...



NASA (.gov)

<https://www.nasa.gov/uploads/2012/03/rockets.pdf>



How Rockets Work

Naturally, **launching rockets** into space is more complicated than Newton's laws of motion imply. Designing rockets that can actually **lift off Earth** and reach ...

NER

Cristiano Ronaldo PERSON dos Santos Aveiro PERSON GOIH ComM (Portuguese pronunciation: [kriˈʃtjɐnu ʁoˈsɐ̃tu ɐvɐiɾu]; born 5 February 1985 DATE) is a Portuguese professional footballer who plays as a forward for Premier League club Manchester United and captains the Portugal national team. He has won 32 trophies in his career, including seven league titles, five UEFA Champions Leagues, one UEFA European Championship EVENT, and one UEFA Nations League. Ronaldo PERSON holds the records for most appearances (183), most goals (140), and assists (42) in the Champions League, most goals in the European Championship EVENT (14), most international goals by a male player (115), and most international appearances by a European male (186).

Знакомьтесь —
наш виртуальный помощник!

Chatbots

A

Альфа-Банк ✓

12.01.2023

Как в Альфа-Банке создали чат-бот, который
и отвечает, словно человек

ТИНЬКОФФ

Олег, сколько
стоит доллар?



СБЕР
БИЗНЕС БОТ



Text Summarization

<https://www.chatpdf.com/c/Cn7Yp0N2QBWj6LBwiJq7u>

Chat



word2vec Parameter Learning Explained

Xin Rong
ronxin@umich.edu

Abstract

The word2vec model and application by Mikolov et al. have attracted a great amount of attention in recent two years. The vector representations of words learned by word2vec models have been shown to carry semantic meanings and are useful in various NLP tasks. As an increasing number of researchers would like to experiment with word2vec or similar techniques, I notice that there lacks a material that comprehensively explains the parameter learning process of word embedding models in details, thus preventing researchers that are non-experts in neural networks from understanding the working mechanism of such models.

This note provides detailed derivations and explanations of the parameter update equations of the word2vec models, including the original continuous bag-of-word (CBOW) and skip-gram (SG) models, as well as advanced optimization techniques, including hierarchical softmax and negative sampling. Intuitive interpretations of the gradient equations are also provided alongside mathematical derivations.

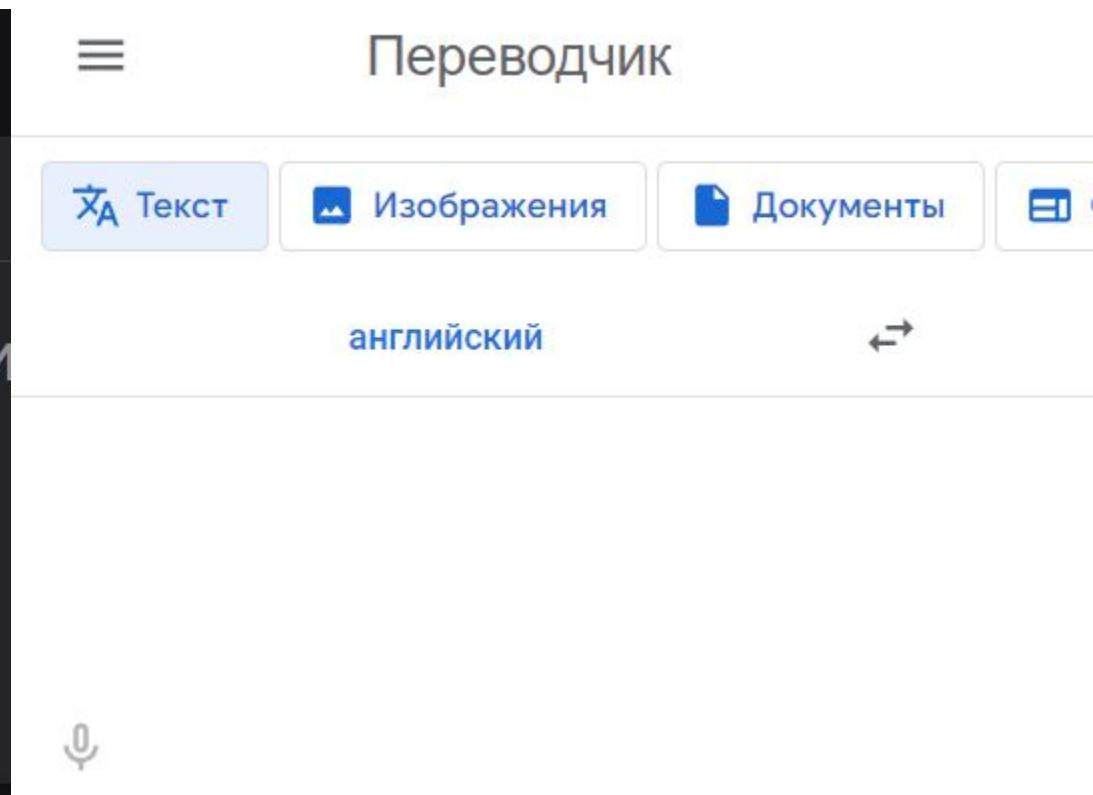
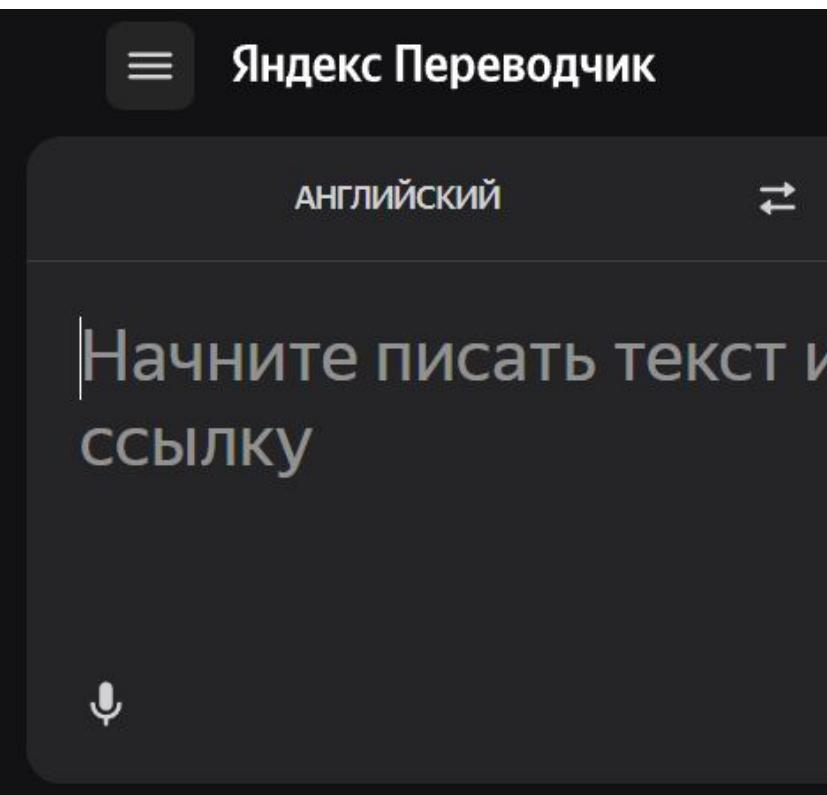
In the appendix, a review on the basics of neuron networks and backpropagation is provided. I also created an interactive demo, wevi, to facilitate the intuitive understanding of the model.^[1]

Hello there! 🎉 Welcome to the fascinating world of word2vec! This PDF file dives into the intricacies of the word2vec model, a groundbreaking approach to word embeddings that captures the semantic meanings of words. Authored by Xin Rong, it provides a comprehensive explanation of the parameter learning process, covering both the continuous bag-of-words (CBOW) and skip-gram models, along with advanced optimization techniques like hierarchical softmax and negative sampling. Whether you're a seasoned researcher or just starting out, this document is designed to enhance your understanding of neural networks and their applications in natural language processing. Enjoy your exploration!

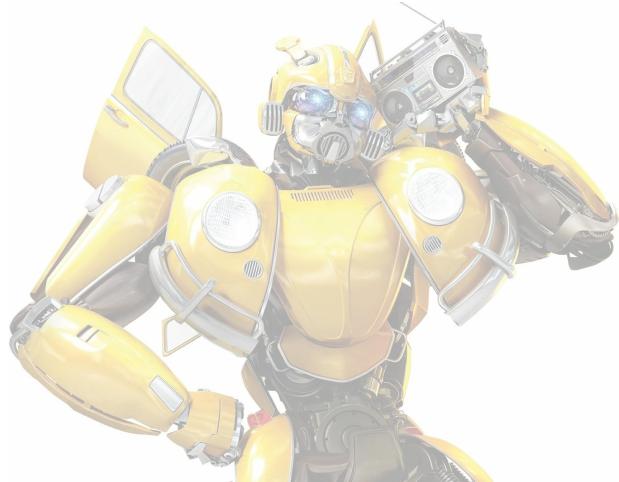
Here are three example questions you can ask about the file:

- What are the main differences between the continuous bag-of-words (CBOW) and skip-gram models in word2vec?
- Can you explain the concept of negative sampling and how it improves the efficiency of word2vec?
- What foundational knowledge about neural networks is covered in the appendix of the PDF?

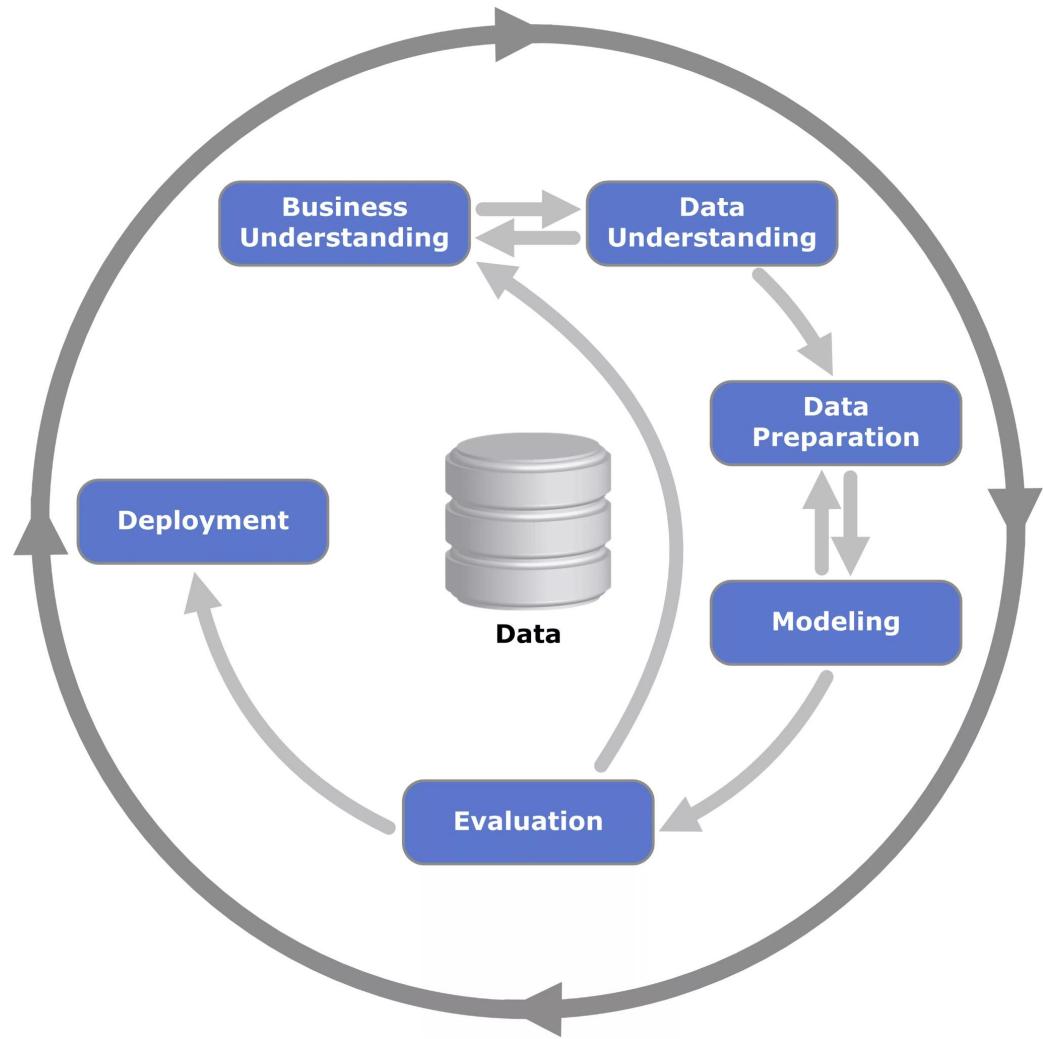
Machine Translation



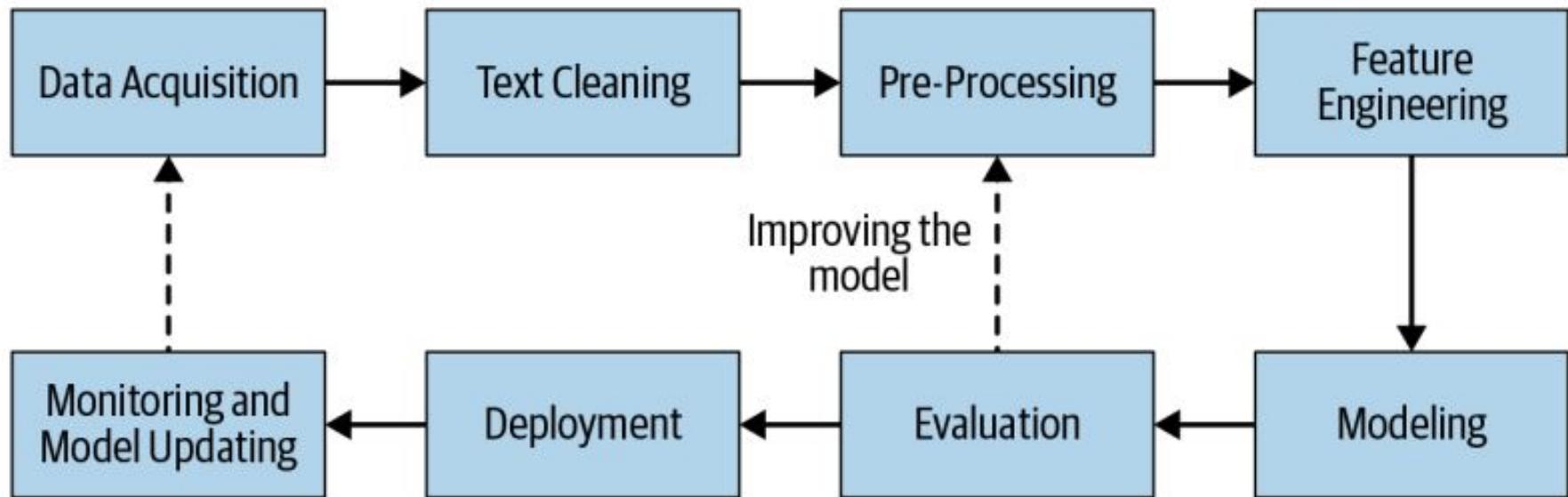
Pipeline



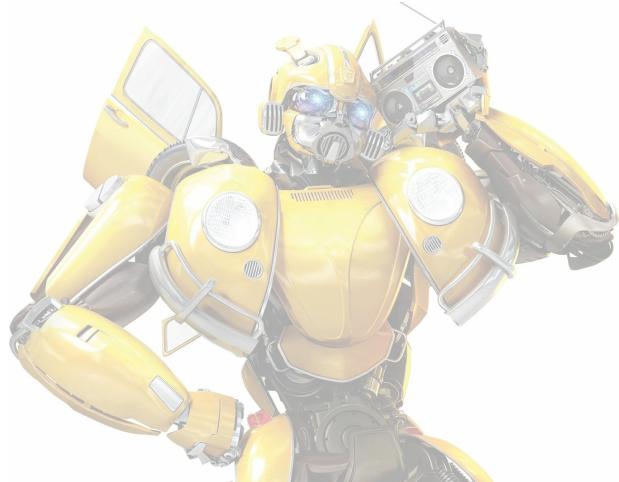
CRISP-DM



Pipeline



Metrics



Metrics



$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}}$$

or

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}}$$

or

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}$$

BLEU (Bilingual Evaluation Understudy)

Measures the similarity between machine-generated text and human-generated reference translations, based on n-gram overlap.

Pros: Widely used for evaluating machine translation models.

Cons: Doesn't capture fluency or grammatical correctness, sensitive to the choice of reference translations.

$$BLEU = BP \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right)$$

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases} \quad p_n = \frac{\sum_{C \in \{\text{Candidates}\}} \sum_{\text{n-gram} \in C} \text{Count}_{\text{clip}}(\text{n-gram})}{\sum_{C' \in \{\text{Candidates}\}} \sum_{\text{n-gram}' \in C'} \text{Count}(\text{n-gram}')}$$

This is illustrated in the following example from Papineni et al. (2002):

Example of poor machine translation output with high precision

Candidate	the	the	the	the	the	the	the
Reference 1	the	cat	is	on	the	mat	
Reference 2	there	is	a	cat	on	the	mat

$$P = \frac{m}{w_t} = \frac{7}{7} = 1$$

$$P = \frac{2}{7}$$

Comparing metrics for candidate "the the cat"

Model	Set of grams	Score
Unigram	"the", "the", "cat"	$\frac{1 + 1 + 1}{3} = 1$
Grouped Unigram	"the"**2, "cat"**1	$\frac{1 + 1}{2 + 1} = \frac{2}{3}$
Bigram	"the the", "the cat"	$\frac{0 + 1}{2} = \frac{1}{2}$

**Spaces**evaluate-metric/**bleu**

like

43



Running

Metric: bleu

data

predictions	references
"Привет! Давай поужинаем?"	"Привет! Давай поужинаем в ресторане?"

bleu

```
{'bleu': 0.4739878501170792, 'precisions': [1.0, 0.75, 0.6666666666666666, 0.5],  
'brevity_penalty': 0.6703200460356393, 'length_ratio': 0.7142857142857143,  
'translation_length': 5, 'reference_length': 7}
```

<https://huggingface.co/spaces/evaluate-metric/bleu>

ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

Measures the overlap between machine-generated text and human-generated summaries, based on n-gram recall and precision.

Pros: Useful for evaluating text summarization models.

Cons: Sensitive to the choice of reference summaries, may not capture semantic similarity.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

I really loved reading the Hunger Games

Machine generated summary

$$\text{ROUGE-1}_{\text{recall}} = \frac{\text{Num word matches}}{\text{Num words in reference}} = \frac{6}{6}$$

I loved reading the Hunger Games

Human reference summary

$$\text{ROUGE-1}_{\text{precision}} = \frac{\text{Num word matches}}{\text{Num words in summary}} = \frac{6}{7}$$

$$\text{ROUGE-1 F1-score} = 2 \left(\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \right)$$

ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

I really

really loved

loved reading

reading the

the Hunger

Hunger Games

Generated summary
bigrams

I loved

loved reading

reading the

the Hunger

Hunger Games

Reference summary
bigrams

$$\text{ROUGE-2 recall} = \frac{\text{Num bigram matches}}{\text{Num bigrams in reference}} = \frac{4}{5}$$

$$\text{ROUGE-2 precision} = \frac{\text{Num bigram matches}}{\text{Num bigram in summary}} = \frac{4}{6}$$

Perplexity

Measures how well a language model predicts the next word in a sequence.

Pros: Simple and easy to interpret, can compare models of different sizes.

Cons: Doesn't directly measure the quality of the generated text, sensitive to rare words and n-grams.

$$PP(W) = \sqrt[N]{\frac{1}{P(w_1, w_2, \dots, w_N)}}$$

Test Set

“Yesterday I went to the cinema”
“Hello, how are you?”
“The dog was wagging its tail”

High probability
Low perplexity

Fake/incorrect sentences

“Can you does it?”
“For wall a driving”
“She said me this”

Low probability
High perplexity

Perplexity

Hugging Face is a startup based in New York City and Paris
p(word)

Word Error Rate (WER)

Ratio of the number of errors in a speech recognition output to the total number of words in the reference transcript.

Pros: Widely used for evaluating speech recognition models.

Cons: Doesn't capture semantic errors, may be sensitive to pronunciation variations.

$$\text{WER} = (S + D + I) / (S + D + C)$$

Where:

S is the number of substitutions (i.e. 'Dolly' vs the actual text 'DALL·E')

D is the number of deletions (i.e. 'I speech-to-text' vs the actual text 'I like speech-to-text')

I is the number of insertions (i.e. 'I really like speech-to-text' vs the actual text 'I like speech-to-text')

C is the number of correctly predicted words.

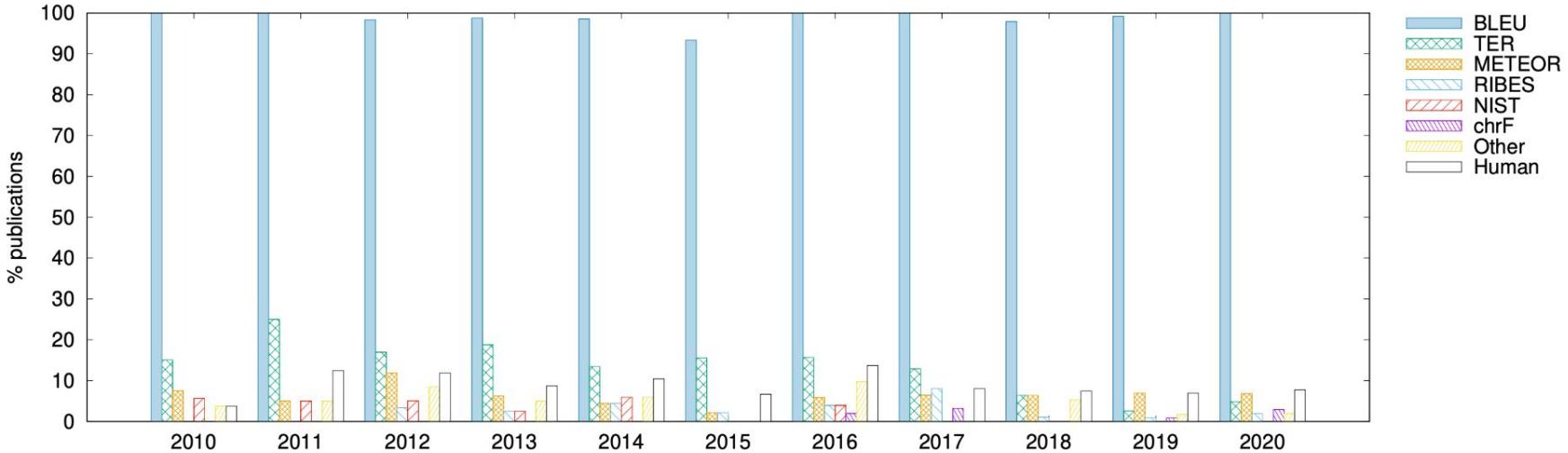
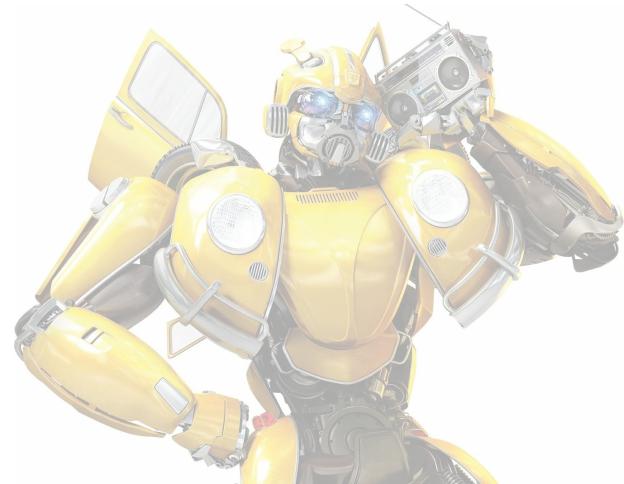


Figure 1: Percentage of papers using each evaluation metric per year. Metrics displayed are used in more than five papers. “Other” denotes all other automatic metrics. “Human” denotes that a human evaluation has been conducted.

Stack

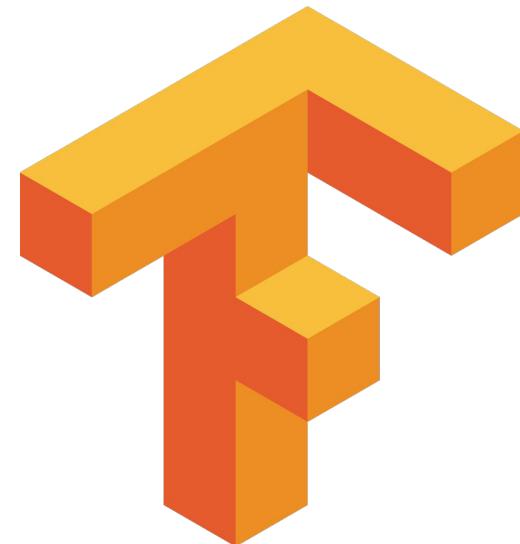


Stack



spaCy

gensim



Regex

A regular expression is an algebraic notation for characterizing a set of strings. Regular expressions are particularly useful for searching in texts, when we have a pattern to search corpus for and a corpus of texts to search through.

•[RegEx]*

regular expressions 101

REGULAR EXPRESSION

1 match (26 steps, 1.3ms)

: / https?:\/\/(?:[-\w]+\.)?(?:[-\w]+)\.\w+(?:\.\w+)?\/?.*

/ i

□

TEST STRING

https://mattweb.ru/moj-blog/raznoe/item/142-30-primerov-regulyarnykh-vyrazhenij

Hugging Face Hub

 **Hugging Face**

[Models](#) [Datasets](#) [Spaces](#) [Posts](#)

[Tasks](#) [Libraries](#) [Datasets](#) [Languages](#) [Licenses](#)
[Other](#)

Multimodal

 [Image-Text-to-Text](#)

 [Visual Question Answering](#)

 [Document Question Answering](#)

Computer Vision

 [Depth Estimation](#)

 [Image Classification](#)

 [Object Detection](#)

 [Image Segmentation](#)

Models 789,568

 [meta-llama/Meta-Llama-3.1-8B-Instruct](#)
Text Generation • Updated 2 days ago • ↓ 204k • ❤ 858

 [apple/DCLM-7B](#)
Updated about 13 hours ago • ↓ 2.93k • ❤ 652

 [mistralai/Mistral-Large-Instruct-2407](#)
Text Generation • Updated 1 day ago • ↓ 3.94k • ❤ 467

 [mistralai/Mistral-Nemo-Instruct-2407](#)
Text Generation • Updated 2 days ago • ↓ 69.2k • ❤ 788

Transformers



Transformers

build failing license Apache-2.0 website online release v4.42.4
Contributor Covenant v2.0 adopted DOI 10.5281/zenodo.7391177

English | 简体中文 | 繁體中文 | 한국어 | Español | 日本語 | हिन्दी |
Русский | Português | தென்றி | Français | Deutsch | Tiếng Việt |

State-of-the-art Machine Learning for JAX, PyTorch and TensorFlow

Part of the Hugging Face course!

```
# Sentiment analysis pipeline
analyzer = pipeline("sentiment-analysis")
```

feature-extraction (get the vector representation of words)

fill-mask

ner (named entity recognition)

question-answering

sentiment-analysis

summarization

text-generation

translation

zero-shot-classification



sentence-transformers



Sentence Transformers

University

🔗 <https://www.SBERT.net> 🌐 nreimers

SentenceTransformers Documentation

Sentence Transformers (a.k.a. SBERT) is the go-to Python module for accessing, using, and be used to compute embeddings using Sentence Transformer models (quickstart) or to ca unlocks a wide range of applications, including [semantic search](#), [semantic textual similarit](#)

A wide selection of over [5,000 pre-trained Sentence Transformers models](#) are available fo the-art models from the [Massive Text Embeddings Benchmark \(MTEB\) leaderboard](#). Addi Transformers, enabling you to create custom models for your specific use cases.

⚡ Inference API ⓘ

💡 Sentence Similarity

Examples ⏺

Source Sentence

That is a happy person

Sentences to compare to

That is a happy dog

That is a very happy person

Today is a sunny day

I love NLP

I love nlp so I am very happy

Compute

Computation time: 0.021 s

That is a happy dog

0.695

That is a very happy person

0.943

Today is a sunny day

0.257

I love NLP

0.210

I love nlp so I am very happy

0.391

JSON Output

Maximize

TensorFlow Hub

TensorFlow Hub — это хранилище обученных моделей машинного обучения.

TensorFlow Hub — это хранилище обученных моделей машинного обучения, готовых к тонкой настройке и развертыванию в любом месте. Повторно используйте обученные модели, такие как BERT и Faster R-CNN, всего несколькими строками кода.



См. руководство

Узнайте, как использовать TensorFlow Hub и как это работает.



См. учебные пособия

В учебных пособиях показ...

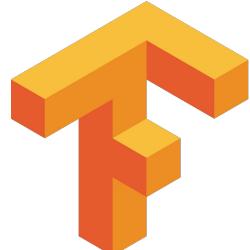
Trending Models

See All



Посмотреть модели

Найдите обученные модели



Llama 3.1

Meta

14 Variations · 11 Notebooks

The Meta Llama 3.1 collection of m...

▲ 146



Gemma 2

Google

12 Variations · 4 Notebooks

Gemma is a family of lightweight, st...

▲ 246



Gemma 2

Keras

4 Variations · 6 Notebooks

Keras implementation of the Gemm...

▲ 66



gemma-2

Eisuke Mizutani

2 Variations · 6 Notebooks

▲ 36



```
!pip install --upgrade tensorflow_hub  
  
import tensorflow_hub as hub  
  
model = hub.KerasLayer("https://tfhub.dev/google/nnlm-en-dim128/2")  
embeddings = model([ "The rain in Spain.", "falls",  
                    "mainly", "In the plain!" ])  
  
print(embeddings.shape) #(4, 128)
```

Practical work_1

Implement an ELIZA-like program 

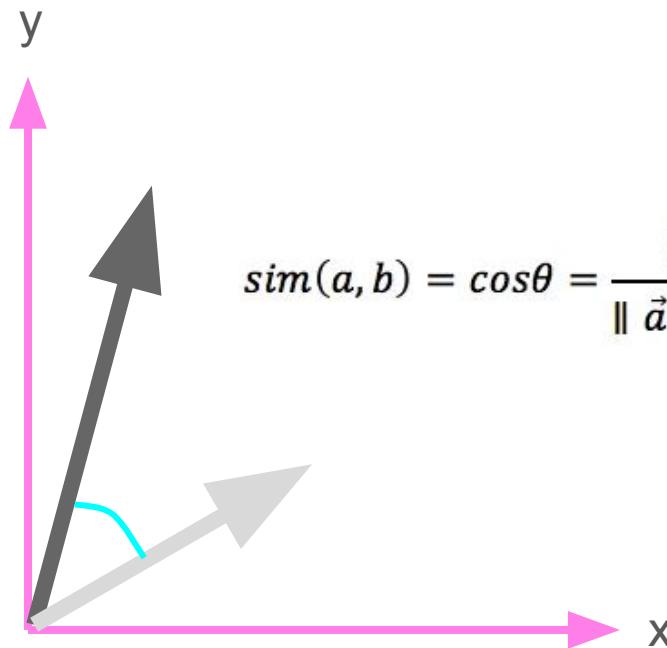
• [RegEx]*

How to work with words?

“Hello, I want to go to Italy”

“Hello, I want to go to Italy”

$$\begin{pmatrix} 0.286 \\ 0.884 \\ 0.428 \\ 0.922 \\ -0.321 \\ 0.286 \\ 0.684 \\ 0.726 \\ 0.912 \\ -0.921 \\ 0.726 \\ 0.912 \\ -0.920 \end{pmatrix}, \begin{pmatrix} -0.322 \\ 0.222 \\ -0.122 \\ 0.002 \\ 0.221 \\ 0.286 \\ 0.111 \\ 0.434 \\ 0.112 \\ -0.111 \\ -0.098 \\ 0.323 \\ -0.220 \end{pmatrix}, \dots, \dots, \dots, \dots, \dots, \dots$$



Perplexity

$$P(W) = P(w_1, w_2, \dots, w_N) = P(w_1)P(w_2) \dots P(w_N) = \prod_{i=1}^N P(w_i)$$

$$\ln(P(W)) = \ln \left(\prod_{i=1}^N P(w_i) \right) = \sum_{i=1}^N \ln P(w_i)$$

$$\frac{\ln(P(W))}{N} = \frac{\sum_{i=1}^N \ln P(w_i)}{N}$$



$$\begin{aligned} e^{\frac{\ln(P(W))}{N}} &= e^{\frac{\sum_{i=1}^N \ln P(w_i)}{N}} \\ (e^{\ln(P(W))})^{\frac{1}{N}} &= (e^{\sum_{i=1}^N \ln P(w_i)})^{\frac{1}{N}} \\ P(W)^{\frac{1}{N}} &= \left(\prod_{i=1}^N P(w_i) \right)^{\frac{1}{N}} \end{aligned}$$

