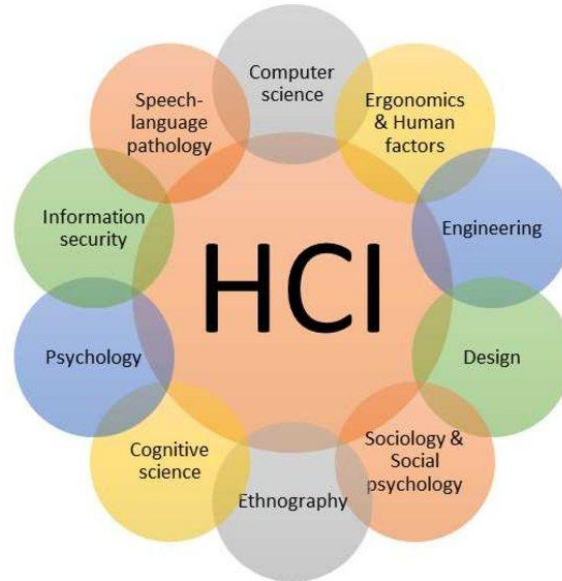


Doğal Dil İşleme Atölyesi

- Bilgisayarlar insan dilini nasıl anlar?
- Doğal dil işleme nedir? Kullanım alanları nelerdir?
- Pretrained language model, transfer learning, attention mechanism nedir?
- Pytorch Lightning, PyTorch ve HuggingFace kütüphanelerini kullanarak pretrained language modeller nasıl fine-tune edilir?

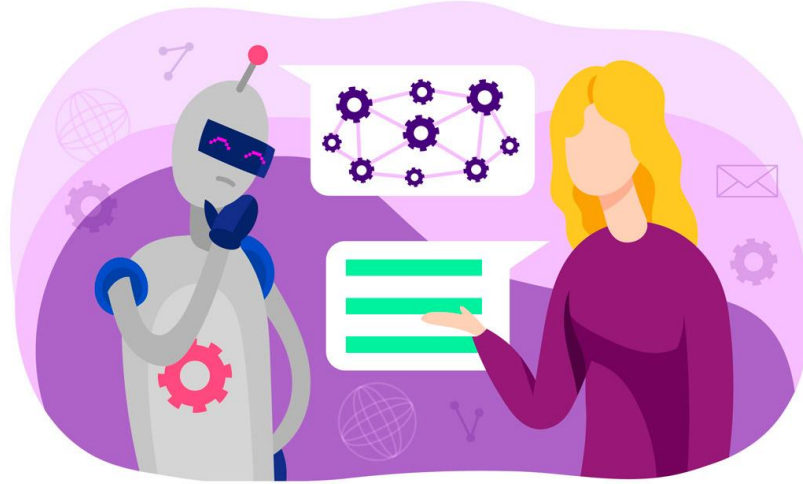
Doğal Dil İşleme

Bilgisayar - İnsan etkileşimini doğal dil yoluyla kuran Bilgisayar Bilimi alt-alanı



Doğal Dili nasıl işleriz?

Bilgisayarlar kelimelerden anlamazlar.



Tokenize etmek yeter mi?

Dil bağlama göre oldukça değişebilen kompleks bir yapıdır. İnsan, dili işlerken gördüğünü, deneyimlerini, bağlamı ve birçok şeyi hesaba katar ve çoğu zaman konuşulan dil belli kurallara uymaz.

- Paket kutuya küçük olduğu için sığmadı.
- Paket kutuya büyük olduğu için sığmadı.

Doğal Dil İşleme != Makine Öğrenmesi

“we mathematically define classes of linguistic representations and formal grammars (which are usually probabilistic models nowadays) that seem adequate to capture the range of phenomena in human languages. We study their mathematical properties, and devise efficient algorithms for learning, production, and comprehension.”

How is computational linguistics different from natural language processing?

This question previously had details. They are now in a comment.



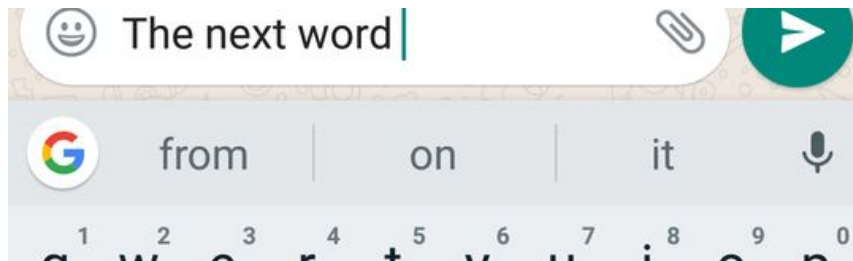
Jason Eisner, computer science professor at Johns Hopkins

Updated Aug 20, 2016 · Upvoted by Rishi Bommasani, Ph.D. Computer Science & Natural Language Processing, Stanford University (2025) and Moukthika Yerramilli, MA Computational Linguistics & Linguistics, English and Foreign Languages University (2018)

Dil Modelleri

Olasılıksal Dil Modelleri

- Kelimelerin ya da kelime öbeklerinin dil içerisinde dağılımlarını öğrenerek gösterim oluşturma



Trigram Probability

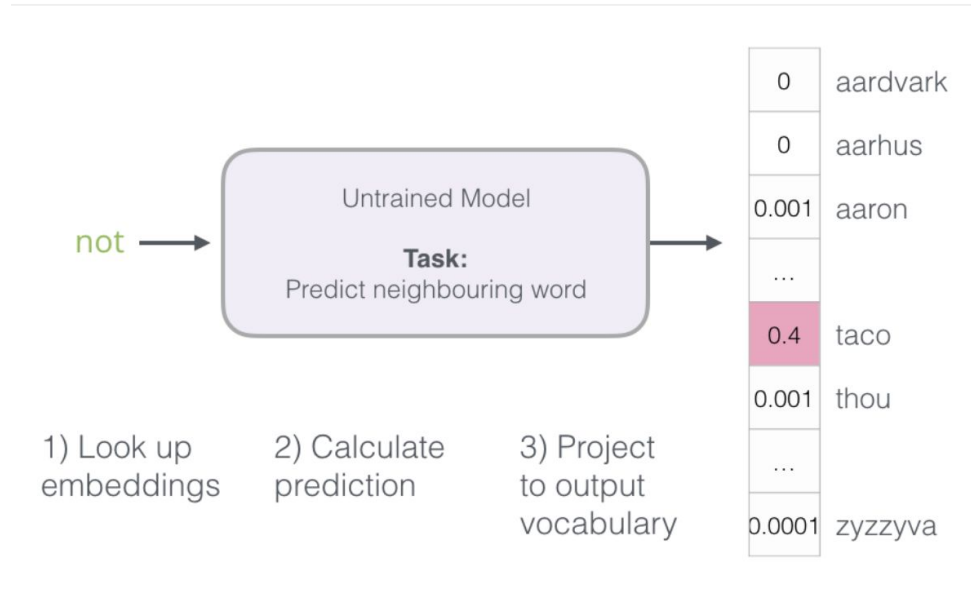
Corpus: I am happy because I am learning

$$P(happy|I\ am) = \frac{C(I\ am\ happy)}{C(I\ am)} = \frac{1}{2}$$

Probability of a trigram: $P(w_3|w_1^2) = \frac{C(w_1^2\ w_3)}{C(w_1^2)}$

Deep NLP, ilk yaklaşımlar

Derin Öğrenme ile Dil Modelleme



Deep NLP, ilk yaklaşımlar

Word2Vec

- 300 boyutlu
- Toplam token sayısı 3M
- 22 saat eğitim süresi
- Google News Dataset, 3B kelime

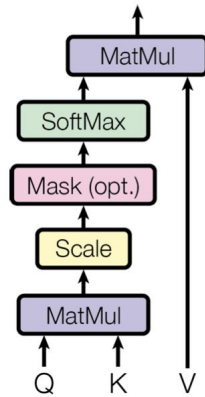
Oyun Zamani

<https://github.com/alierenak/iSemanticSpace>

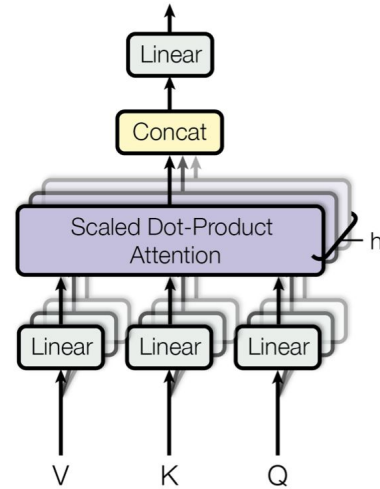
Transformers Yapıları

Attention Mekanizması

Scaled Dot-Product Attention



Multi-Head Attention



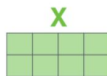
Transformers Yapıları

Attention Mekanizması

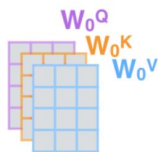
1) This is our input sentence*

Thinking
Machines

2) We embed each word*



3) Split into 8 heads.
We multiply X or R with weight matrices



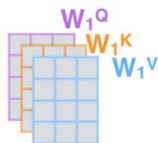
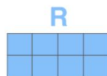
4) Calculate attention using the resulting $Q/K/V$ matrices



5) Concatenate the resulting Z matrices, then multiply with weight matrix W^O to produce the output of the layer



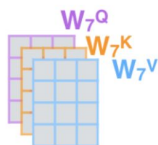
* In all encoders other than #0, we don't need embedding. We start directly with the output of the encoder right below this one



...

...

...



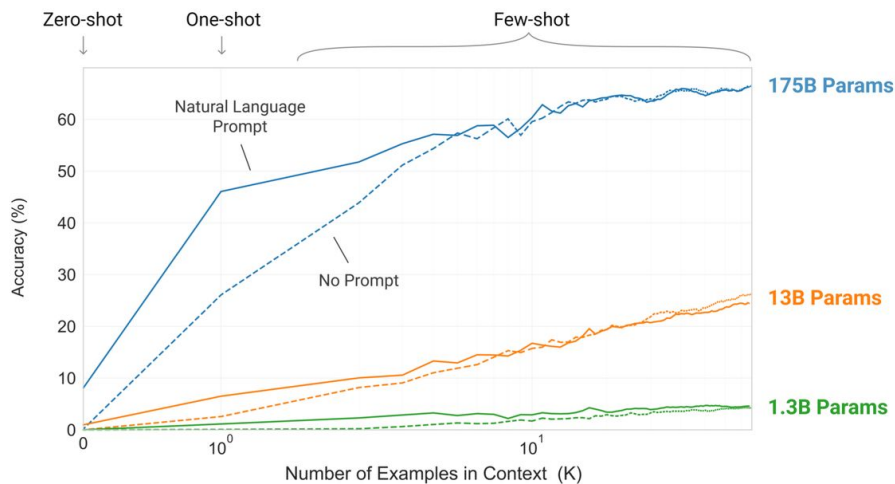
Transformers Modeller

Eğitim teknikleri

- Masked Language Modelling (BERT, RoBERTa, T5)
- Casual Language Modelling (GPT, Decoder-only architectures)
- Permutation Language Modelling (XLNet)

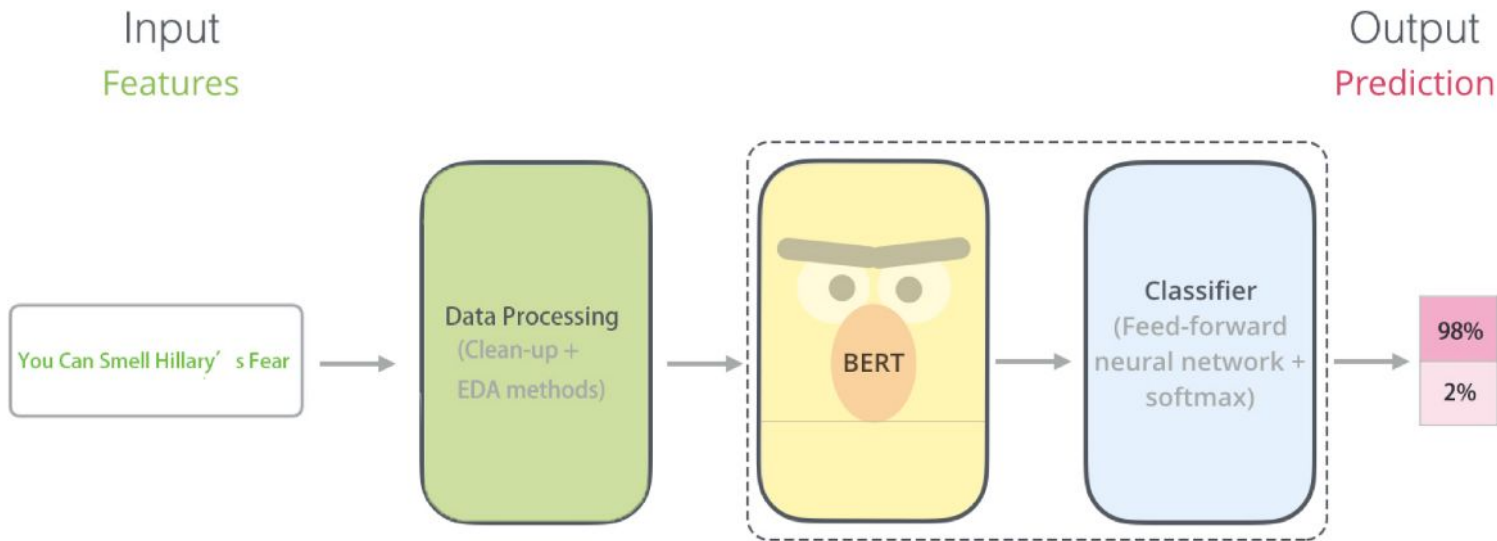
Transformers Yapıları

Modeller gittikçe büyüyor

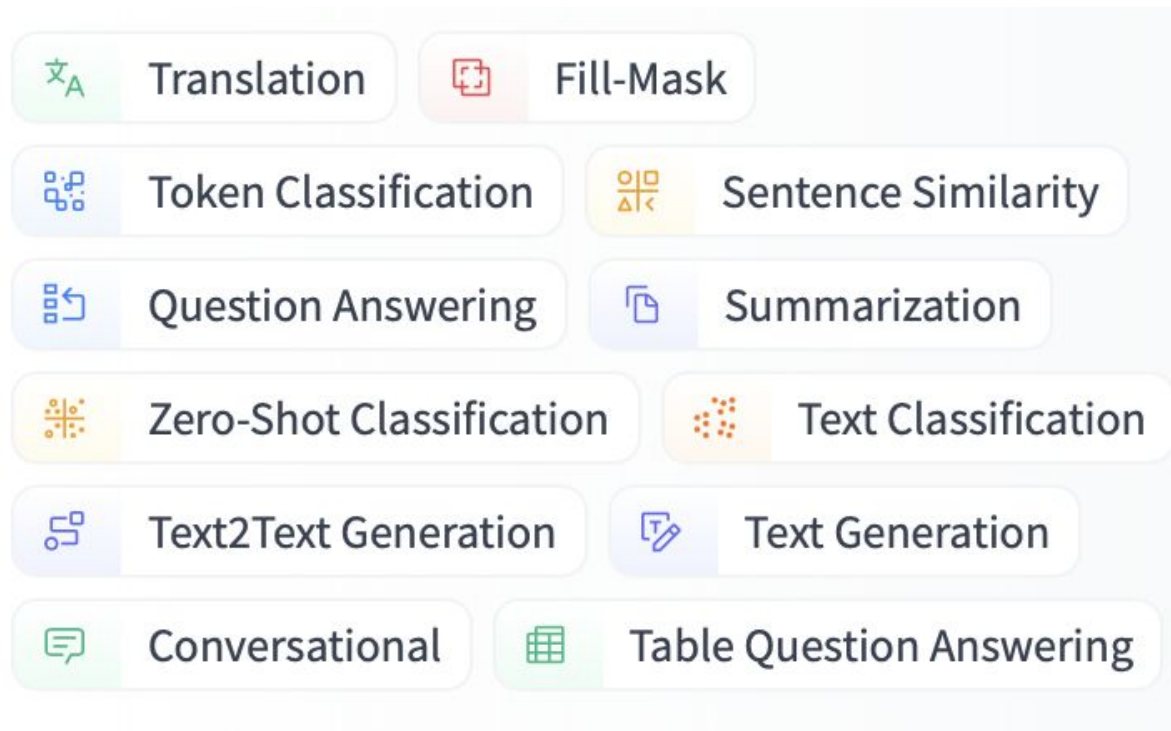


Transformers Pipeline

Transformers yapılarında text verisinin izlediği yol



Transformers Kullanım Alanları



Kodlama Zamanı

https://colab.research.google.com/drive/1U3w37nTK9i_UFNq4rwVaX3mn79veJNa-?usp=sharing