

Derin Öğrenme

DL for NLP ve Transformer'lar

Emir Öztürk

Deep Learning for NLP

NLP... NLP Everywhere

- Doğal dil işleme
- Doğal dil ile ilgili tüm konuların kapsamı
 - Ses tanıma
 - Ses üretimi
 - Makine çevirisi
 - Text sınıflandırma
 - Text özetleme

Deep Learning for NLP

Too many parameters to handle

- Sembol seçimi
- Ağırlıklı olarak kelimeler
- Kelimenin birden fazla anlamı
- Cümle içerisinde geçtiği yerin önemi

Anlam eldesi

The picture is not relevant

- Yalnızca kelime değil kelime içerisindeki harfler için de uygulanması mümkün

Jack and Jill went up the hill.

The pole vault was the last went.

Anlam eldesi

Hundreds of them

- Context'in çözümleme için kullanılması
- Kelimenin diğer kelimelere göre belirlenmesi
 - **Yüz** kişi **yüz** maskesi ile denizdeki adama **yüz** diye bağırdı
- Konum veya diğer kelimelerin yetmediği durumlar
 - The man who ate the pepper sneezed
 - The cat who bit the dog barked

Anlam eldesi

Funniest one is the man's hand right?

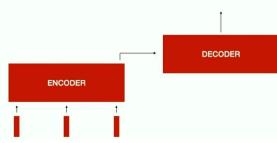
- Görüntüde bu iş daha da zorlaşmaktadır
- Detayın verilmediği durumlar
- Rengin anlaşılması olmadığı durumlar
- İnsanların bile ayırt edemediği durumlar



Encoder - decoder mimarisi

TV Decoder was old technology

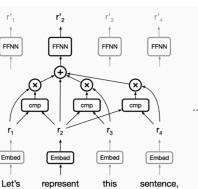
- Decoder encoder'in son hidden state'ini kullanır
- Daha önceki hidden state'lerin etkisi de olabilir



Transformer'lar

Optimus prime

- Attention is all you need
- Attention kısmına encoder'in tüm hidden state'leri verilir
- Decoder'in hidden state'leri de verilir
- Self attention
 - Bir kelime için geçmiş ve gelecek bilgisi
- Kelimenin diğer kelimelerin katkısı ile representasyonu
- Pozisyonların saklanabilmesi



Transformer'lar

One punch man

- One hot encoding başarılı değil
 - Her kelimenin birbirine olan uzaklığı eşit
 - Anlam yakınlığı öğrenimi mümkün değil

Transformer'lar

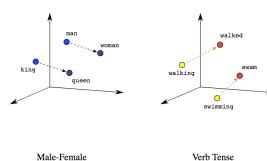
Embedded systems are absolutely different

- Embedding
 - Metin kelimelere bölünür
 - Kelimelere birer id verilir.
 - Her kelime için birer embedding vektör oluşturulur.
 - Sözlükteki kelime sayısı D, V büyüklüğünde bir vektör => VxD Matris

Word Embeddings

The same old graphic everywhere

- Kelimelerinin ifadesinin bir biçimini
- Embedding örnekleri
- Embedding'de kelimeler arası ilişkinin tanımı



Transformer'lar

And there is a more complicated scheme

- Pozisyonların saklanması
 - Embedding'e pozisyonların da eklenmesi
- Buradan elde edilen çıktı "Positional embedding"

Sequence of tokens, n	Index of token, i	Positional Encoding Matrix with $A=2\pi$, $\omega=100$			
		i=0	i=1	i=2	i=3
I	0	$P_{pos}(0) = 0$	$P_{cos}(0) = 1$	$P_{sin}(0) = 0$	$P_{cos}(0) = 1$
am	1	$P_{pos}(1) = 0.84$	$P_{cos}(1) = 0.54$	$P_{sin}(1) = 0.10$	$P_{cos}(1) = 0.96$
a	2	$P_{pos}(2) = 0.91$	$P_{cos}(2) = -0.42$	$P_{sin}(2) = 0.20$	$P_{cos}(2) = 0.88$
Robot	3	$P_{pos}(3) = 0.14$	$P_{cos}(3) = -0.99$	$P_{sin}(3) = 0.30$	$P_{cos}(3) = 0.96$

Positional Encoding Matrix for the sequence 'I am a robot'

Transformer'lar

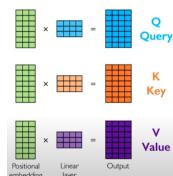
Everyone needs it... Not the transformer but attention

- Attention
- Self-Attention
- Multi-head Attention

Attention

Because 1 isn't enough

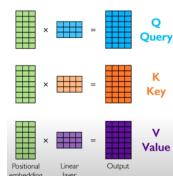
- 3 adet matris
- Her matrisin ağırlığı birbirlerinden farklı
- $V \times D$ büyüğünde bir matrisle çarpıma uygun boyutta
- Çıktıda elde edilen 3 adet vektör
 - Q
 - K
 - V



Multihead self attention

Do you know anything about Cerberus

- Farklı başlangıç matrisleri ile aynı işlemlerin n defa gerçekleştirilmesi
- Birbirinden bağımsız n adet self attention katmanı
- Katmanların lineer bir katman ile tek katmana düşürülmesi



Positional encoding

Take your positions

- Kelimenin nerede geçtiği önemli
- LSTM ve RNN'de pozisyonlar sekansların içerisinde verilmekte
 - Kelimenin n adet sonra geçmesi
 - Aralık açıldıkça aralık azalır
- Kelimelerin geçtiği indislerin kodlanması
 - Kelimenin n. yerde geçmesi
 - Uzaktaki kelimelerin de önemi kaybolmaz

Unsupervised Learning - BERT

It won't be complete without tea

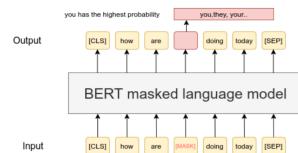
- Transformer
- Eldeki veriden eğitim
- RNN ve LSTM'deki gibi input'u kaydırma yoluna gidilebilir.
- Bunun dışında iki yönlü öğrenim için MASK



Masked Language Modelling

The Mask was a good movie

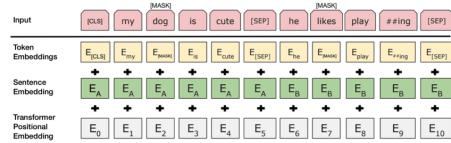
- Verilen cümlenin belirli bir orandaki kelimesini MASK ile değiştirmek
- Kelimelerin diğer kelimeler ile olan ilişkisini sistemin öğrenmesini sağlamak
- Öğrenme çift yönlü olduğu için aradaki kelimenin tahmini de mümkün



Masked Language Modelling

There was carriage return and new line

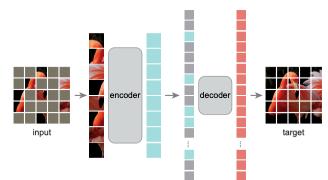
- Cümleler arası ilişkinin de kurulması
- CLS
- SEP



Masked Modelling

Yes, they tried it again

- Görüntüler için de kullanılabilir
- Görüntü içerisinde belirli kısımlar mask'lanarak öğrenme sağlanabilir



Transformer örnekleri

And by 2025 there is more

- GPT (GPT-2, GPT-3, GPT-4, ...)
- Claude
- BERT
- XLNet
- RoBERTa

Transformer kullanım alanları

Use it with caution

- NLP
 - Chatbot
- Görüntü işleme
 - ViT (Vision transformer)
- Bioinformatics
 - AlphaFold
- Ses işleme
 - Speech to text

Finetuning

To get the sound right

- RNN ve LSTM'de sıfırdan eğitim
- Önceden eğitilen bir modelin üstüne yeni veri ile eğitim işlemi gerçekleştirmek
- Zaman kazancı
- Başarının daha yüksek olması
- Daha küçük verisetinde transfer learning

Finetuning

Getting bigger and bigger

- Tüm model finetune edilmesi maliyetli
- Özellikle büyük dil modellerinde bellek ihtiyacı
- Parameter efficient fine tuning (PEFT)
- Low Rank Adaptation (LORA)

Transformer'ların kısıtları

We all have some constraints

- Yüksek işlem gücü ve bellek ihtiyacı
- Model içeriğinin yorumlanması
 - Explainable ai
 - Büyük veri ihtiyacı
- Mevcut modellerin sahip olduğu bias'in miras alınması
- Hiperparametre optimizasyonu problemleri

Transformer'lardaki yenilikler

Old buildings have some value

- LLM ve reasoning
- Sparse - Linear transformers
- Flash Attention
- Quantization
- Multimodal - Hibrit mimariler
- Ethical AI