

Derin Öğrenme

Sequences and RNNs

Emir Öztürk

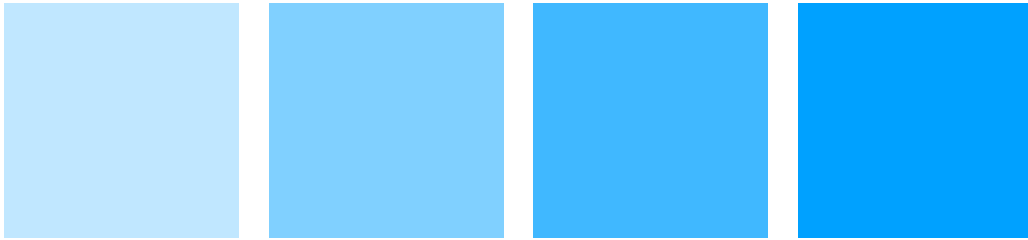
Sequence

Has some consequences

- Sequence
 - Ardışık tekrar eden elemanlar
 - Sıralama önemli
 - Cümlelerdeki kelime sırası
 - Videodaki frame sırası
 - Ses verisi
 - Karar algoritmalarındaki karar sırası
 - Farklı uzunlukta olabilir.

Sequence'in tahmindeki etkisi

Really? A slide with four squares?



Sequence için CNN-DNN

Not CuDNN

- Dense network ve CNN sabit uzunlukta
- Sıralamayı hesaba katmaz.
- Eğitimde önemli olan detaylar kaybolur.

Sınıflandırmadan farklı

Do not classify people as sequences

- Sınıflandırmada girdi verisine karşılık gelen bir sınıf verilir
- Sequence'lerde ise yalnızca girdi verisi bulunur
- Çıktı, istenilen sonucun olasılığıdır
- Maksimizasyon problemi

Doğal dilin modellenmesi

I think someone got excited

- Her kelimenin bağımsız olasılığının hesaplanması
 - The man sees the bird
 - %40
 - %20
- The _____
- The the _____
- The the the _____

Doğal dilin modellenmesi

Nothing is easy

- Bir kelimenin kendinden önce gelen n kelimeden sonra gelme olasılığı
- Daha başarılı sonuç elde edilir
- Her kelimenin bir önceki kelimeden sonra gelme olasılığı hesaplanmalı
- Kelime sayısı ile tablonun lineer artmaması
- Önce gelen kelime sayısı 1'den büyük olduğunda tablonun daha da büyümesi

Doğal dilin modellenmesi

As the examples

- The man sees the man and the man sees the boy (n=2)
 - The man ➡ sees, and
 - man sees ➡ the
 - sees the ➡ man, boy
 - and the ➡ man
 - man and ➡ the

Doğal dilin modellenmesi

And it always gets harder

- The man sees the man and the man sees the boy (n=3)
 - The man sees ➡ the
 - man sees the ➡ man, boy (%50)
 - sees the man ➡ and
 - the man and ➡ the
 - man and the ➡ man
 - and the man ➡ sees

Doğal dilin modellenmesi

What about simplification?

- Dahil edilen geçmiş kelime sayısının sınırlandırılması?
 - Yer ihtiyacı çok küçülmüyor
 - Context ile ilgili bilgi kaybı olabilir
 - Bazı kelimeler çok uzakta olsa da son kelime ile ilgili olabilir

Doğal dilin modellenmesi

Of course not

- Bunun yerine
 - Olasılıkları öğrenme işlemi
 - Context ve sırayı koruyacak şekilde
- Bunu sağlayabilmek adına
 - İçeriği vektörleştirmek

Vektörleştirme

Sublimating, Evaporating, Vectorizing

- Embedding
 - The man sees the man and the man sees the boy
 - Sözlük çıkarımı
 - İndis belirleme

The	0
man	1
sees	2
and	3
boy	4

Vektörleştirme

That is a one hot encoding

- One hot encoding
 - Toplam kelime sayısı uzunluğunda
- Kelimenin bulunduğu indis 1
- Diğerleri 0
- Man \Rightarrow [0 1 0 0 0]

The	0	10000
man	1	01000
sees	2	00100
and	3	00010
boy	4	00001

Bag Of Words

We need a really big bag

- Tüm kelimelerin geçme sayısını bir vektör olarak oluşturmak
- Sıra tutulmuyor
- Aynı kelimelere sahip iki cümle farklı anlamlara gelebilir
- I'm good and you're bad ["I":1,"good":1 ...]
- I'm bad and you're good ["I":1,"good":1 ...]

The	0
man	1
sees	2
and	3
boy	4

Doğal dilin modellenmesi

It sounded like big bang right?

- Alınan vektörün elemanlarının sırası korunmalı
- Farklı uzunlukta vektörler olabilmeli
- Zamanla öğrenebilmeli
- İkili - üçlü gruplar hesaba katılmalı
- Uzun sequence'lerde içerik korunabilmeli

Doğal dilin modellenmesi

Too much to do

- N adet eleman alındığında
 - Alınan vektörün elemanlarının sırası korunmalı
 - Farklı uzunlukta vektörler olabilmeli
 - Zamanla öğrenebilmeli
 - İkili - üçlü gruplar hesaba katılmalı
 - Uzun sequence'lerde içerik korunabilmeli

Doğal dilin modellenmesi

Goals that will never be reached

- Tüm kelimelerin olasılıkları toplandığında
 - ~~Alınan vektörün elemanlarının sırası korunmalı~~
 - Farklı uzunlukta vektörler olabilmeli
 - Zamanla öğrenebilmeli
 - ~~İkili – üçlü gruplar hesaba katılmalı~~
 - Uzun sequence'lerde içerik korunabilmeli

Doğal dilin modellenmesi

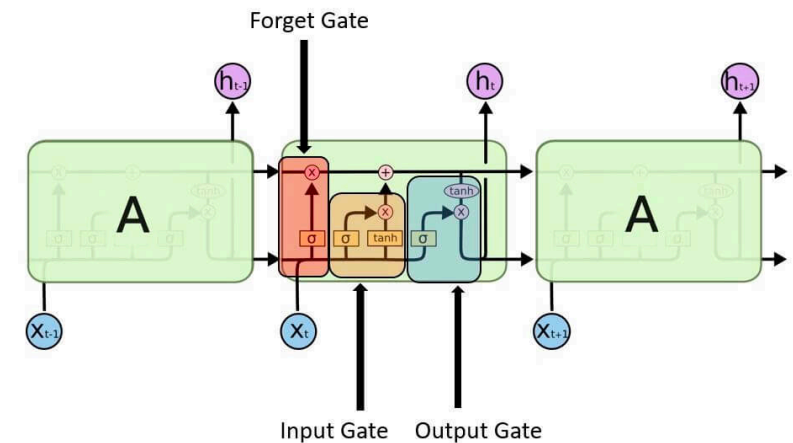
And it is getting worse

- Bag of words kullanıldığında
 - ~~Alınan vektörün elemanlarının sırası korunmalı~~
 - Farklı uzunlukta vektörler olabilmeli
 - Zamanla öğrenebilmeli
 - ~~İkili – üçlü gruplar hesaba katılmalı~~
 - Uzun sequence'lerde içerik korunabilmeli

RNN - Recurrent Neural Networks

R nearest neighbours?

- Sistemin ağırlıkları bir önceki adımın ağırlıklarını da içerir
 - Bu sayede context korunur
- Elde edilen çıktı tüm kelimelerin olasılıklarını içeren bir vektör
 - Tüm olasılıkların toplamı 1.0



RNN - Recurrent Neural Networks

Getting started

- Veri hazırlamak için her adımda
 - Bir önceki adımın ilk kelimesi eksiltilir
 - Son kelime eklenir
 - Ağırlıklar güncellenir

RNN - Recurrent Neural Networks

I told you it will get worse

- The man sees the man and the man sees the boy (N=5)
 - “The man sees the man”, and
 - “man sees the man and”, the
 - “sees the man and the”, man
 - “the man and the man”, sees
 - “man and the man sees”, the
 - “and the man sees the”, boy

RNN - Recurrent Neural Networks

Entropy... again

- Kayıp hesabı için
 - Girdi ve sıradaki çıktı kelimesi biliniyor
 - Çıktı kelime olasılıkları biliniyor.
- Birebir eşleme yok
 - Loss tümü için hesaplanmalı
- Cross Entropy

$$H(P^* | P) = - \sum_i \underbrace{P^*(i)}_{\text{TRUE CLASS DISTIRBUTION}} \log \underbrace{P(i)}_{\text{PREDICTED CLASS DISTIRBUTION}}$$

Exploding Gradients

TNT

- Ağırlıkların güncellenmesi = Farkın eklenmesi
- Gradyan
- Ağırlığı yüksek düğümler özyinelemeli olarak geriye dönük çarpılır
- Belirli bir durumdan sonra değişim durabilir
- Değişim durduğunda öğrenme duracaktır

Vanishing Gradients

Casper

- Exploding Gradients ile aynı durum
- Düşük olasılıklı düğümlerin özyinelemeli olarak azalması
- Gradyan'ın 0'lanması
- Öğrenmenin durması
- Vanish veya explode olasılığını azaltmak için step sayısı azaltılabilir

RNN - Recurrent Neural Networks

You are aware that the conditions are still not met

- RNN
 - Alınan vektörün elemanlarının sırası korunmalı
 - Farklı uzunlukta vektörler olabilmeli
 - Zamanla öğrenebilmeli
 - İkili – üçlü gruplar hesaba katılmalı
 - Uzun sequence'lerde context korunabilmeli

Uzun vade bağımlılıkları

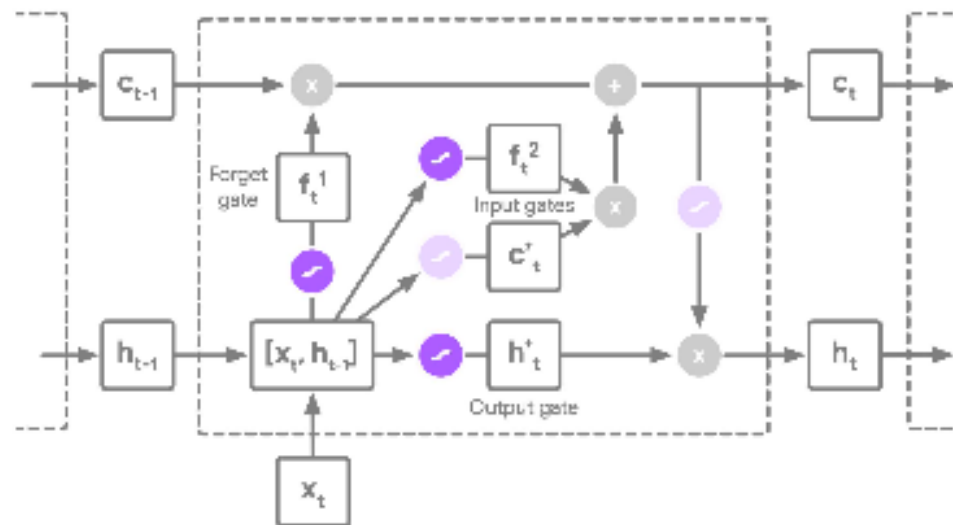
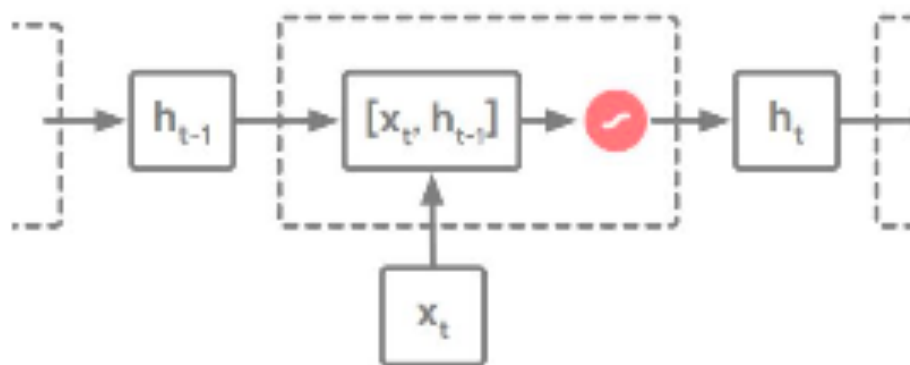
Is loyalty a district name?

- Uzun vadede belirli kelimeler cümleyi etkileyebilir
- Bu verinin kaybedilmemesi gerekmektedir.
- Rnn'lerde context uzun tutulabilir
 - Gradyan problemi

LSTM - Long-short Term Memory

Low resolution images are the key

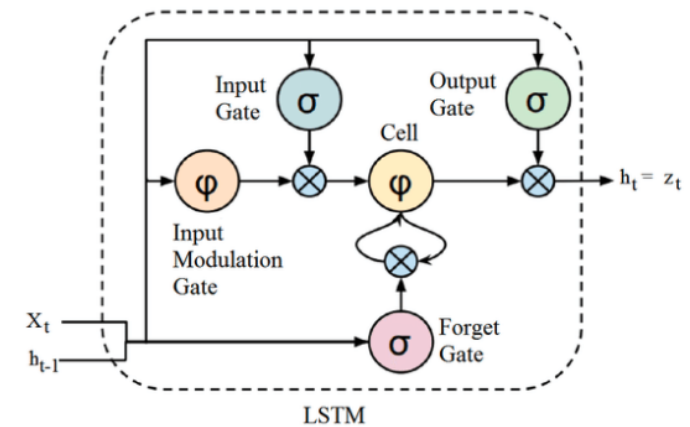
- RNN vs LSTM



LSTM - Long-short Term Memory

And now it is not readable

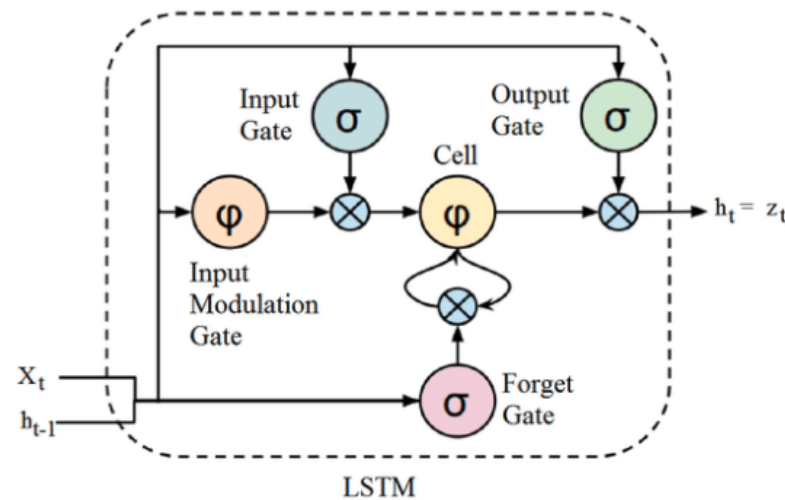
- RNN'den farklı olarak uzun dönem özelliklerini saklamayı amaçlar
 - Kısa vade özellikleri yakalayacak bir yapı
 - Uzun vade özelliklerini yakalayacak bir yapı
 - Bu yapının gradyan problemlerini engelleyecek bir yapı



Forget gate

It just switched position

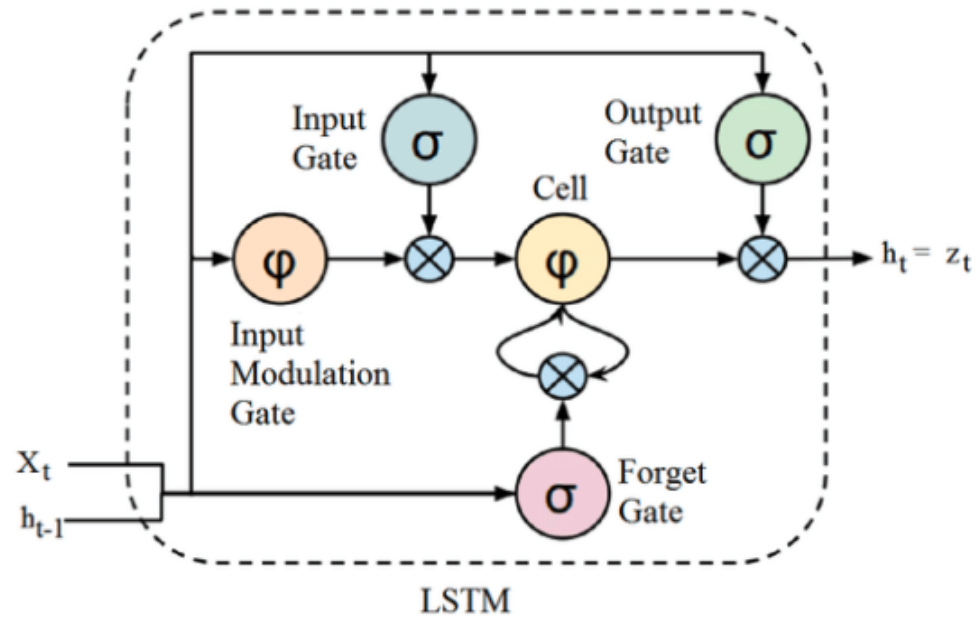
- Belirli durumlarda bazı özelliklerin unutulması gerekmektedir.
- 0 olan (vanished) gradyanların temizlenmesi
- t anındaki girdi ve $t-1$ anındaki hidden state'i alır ve sigmoid'e verir



Input gate

And it didn't do anything more

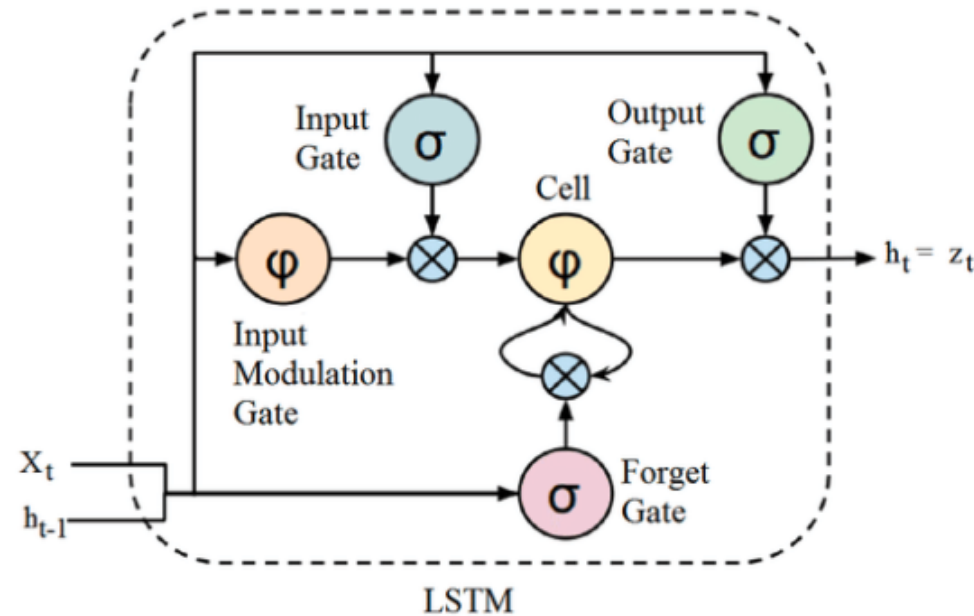
- Girdi verisine yeni bilginin eklendiği kısım
- Eski veri de uzun vade için bulunmaktadır



Output gate

Last one I swear

- RNN'de olduğu gibi çıktının gradyanının öğrenildiği kısım
- Önceki kapıdan unutulmuş, uzun vadeli ve yeni veri gelir ve yeni değerler hesaplanır



LSTM

Getting the end of it

- LSTM
 - Alınan vektörün elemanlarının sırası korunmalı
 - Farklı uzunlukta vektörler olabilmeli
 - Zamanla öğrenebilmeli
 - İkili – üçlü gruplar hesaba katılmalı
 - Uzun sequence'lerde context korunabilmeli

Sequence to sequence models

Face to face

- Girdi ve çıktı verileri farklı sequence'ler olarak verilebilir.
- Her dildeki veriyi birer sequence olarak düşünebiliriz
- Translation için kullanılabilir

Sequence to sequence models

Everything has some use cases

- Sequence sayısı eşit olmak zorunda değil
- Bir girdiye birden fazla çıktı
 - Text generation
- Birden fazla girdiye bir çıktı
 - Sentiment classification
- Birden fazla girdiye birden fazla çıktı
 - Translation, Generation, Summarization

Sequence to sequence models

Half Turkish half English

- Resim verisi sequence kabul edilebilir
- Language-Language sequence'ler yerine language-image sequence'ler
 - Translation benzeri
- Ses verisi sequence kabul edilebilir
 - Language-Ses
 - Ses-Language
- Transformer'lar

Generation

Z generation is the w... wait that's not it

- RNN veya LSTM'ler ile generation
- Bir önceki adımın çıktısı bir sonraki adıma verilir
 - Text generation
 - Image generation

PixelRNN

It became less boring with the help of pictures

- Resimler de sequence'ler olarak düşünülebilir ve RNN-LSTM'ler ile kullanılabilir
- PixelRNN

