How data is processed input/output, base of language modelling 10 min

Transformer XL 10 min (5 slides)

Problem: attention can only deal with fixed-length context, split the corpus into chunks, the problem is if you do this you can lose the context across those fragment (context fragmentation)

Recurrence in deep self-attention networks. Transformer XL is a reccurent attention model

Self Attention presentation

Mask you can prevent a network from extracting forbidden information

No information flow across the different chunks

Largest possible dependency is bound by sequence length

Concatenate the previous segment to the next segment

Concatenating every previous hidden state with the current one

~~With the vanilla transformer :~~

~~Consumes one segment at a time and one prediction at a time~~

~~Extremely expensive~~

Explain formula and each term of it

Explain HEAD utility

Transformer 15 min (8 slides)

EXPLAIN CLEARLY POSITIONAL EMBEDDING

EXPLAIN VERY VERY VERY CLEARLY THE ATTENTION SYSTEM

Adaptive softmax 10 min (5 slides)

Introduction to Language Model

Character level :

Data :

* The data (Markets360)
* How is it tokenized
* Statistics on data

Model :

* LSTM
* Memory performance and Accuracy in a word
* Results
* Limit ?

Word level

* Data (Wikitext-103 and Markets 360)
* Markets 360 tokenization

Transformer :

* Why using a transformer
* Positional Embedding
  + Why
  + How does it work ?
* Self Attention :
  + Principle, why?
  + Detail of the implementation
* Multi Head self attention
  + Why ?
  + Detail of the implementation
* Results for Self Attention
* Feed forward Layer
* Decoder and final softmax layer
* Problem ?

Transformer XL

* Differences with transformer
* Memory added
* Absolute Positional Embedding problem
  + How is it solved
* Formula and explaination of each term
* Graphics about heads

Adaptive Softmax

* Softmax problem
* Adaptive softmax
* How does it work ?

Our model

Result of the model

What in the future ?

* Increasing size of the dataset ?
* Model GPT2 (BPE)