Lect 33: Attention Model & Transformers

review of previous lecture RNN & its types

The Seq2Seq Model

model with RNNs / Seq2Seq

model in NLP used to convert seq of type A to seq of Type B

earlier classification seq → label

auto-encoder (Sequence – Sequence)

earlier with fully connected layer

now we can use different network of our choice

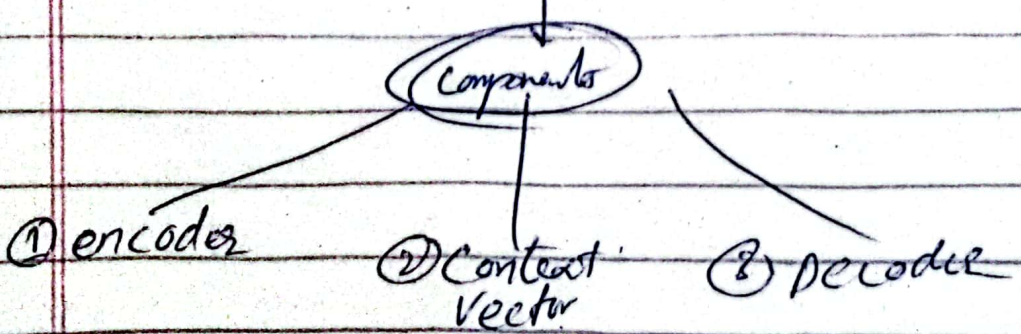
auto-encoder

Encoder/decoder model (will be using RNN)

example

(i) translation of english sentence to Spanish

Encoder-decoder Network / Seq2Seq Model Architecture



encoders are good at learning representation of any information.
it could be audio/video/image/text

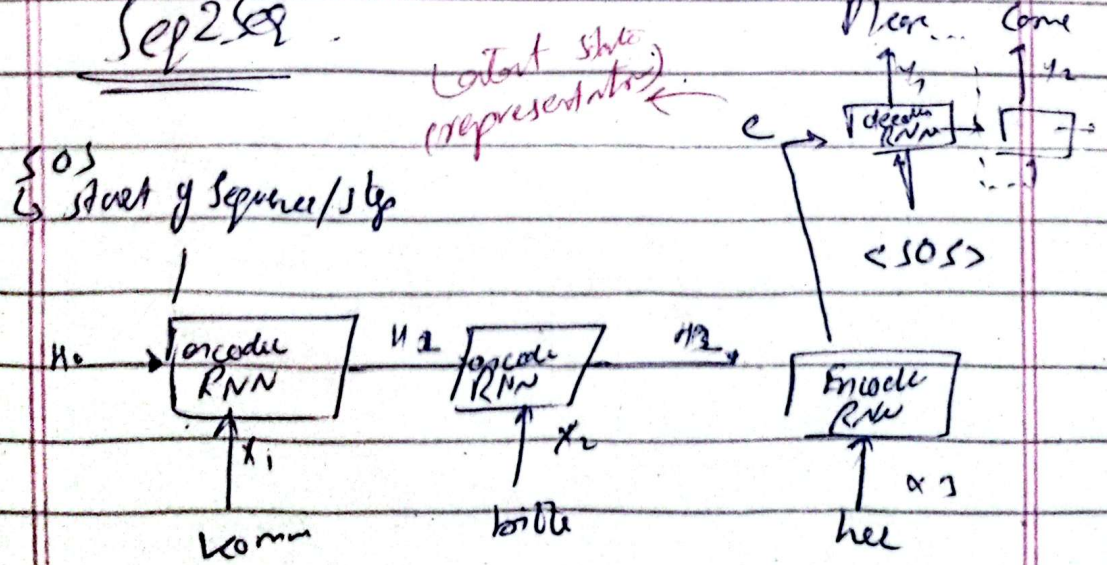
encoder → • Takes input as sequence and output as fixed representation
• read variable-length input

Decoder → Take context vector (hidden representation / latent representation) and give sequence as output.

where they used
(i) Language related tasks → Summarization
→ Translation
→ Question Answering

Context Vector
• (fixed dimension representation of input sequence)
• hidden state which is a vector that represents the learned features
• output of encoder

Seq2Seq



- a (a, b as input (compound sequence) of different) - (different time steps)

Information (input) + summary ⇒ output

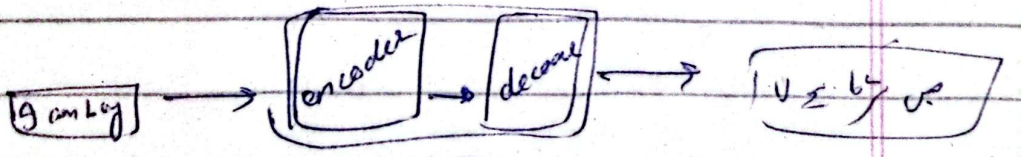
Cell can be either vanilla, LSTM, GT can be any.

input ⇒ x_1, x_2, x_3 } variable set
output ⇒ y_1, y_2, y_3 } variable set

applications & popularity

① Sentiment Analysis
② Question Answering
③ Machine Translation

⇒ Translation



Neural Machine Translation

Seq-2-Seq Model

→ it learned this, due to training.
Train together
& then they can work independently.

Problems with RNN

① They process input data sequentially, one after the other.
So not make use of modern GPU, which were designed for parallel computation. So slow training

② Long-term dependency problem
- quite ineffective when elements are distant from one another.
- due to fact, information is passed at each step
- Longer the chain is, the more probable the information lost along chain.

③ bottleneck problem
- due to fixed-size context vector
- it may fail to capture all relevant details, leading to information loss.
- So can not adequately represent long-range dependencies within input seq.

Problems with Traditional learning model

To struggle with remembering long stories or instructions

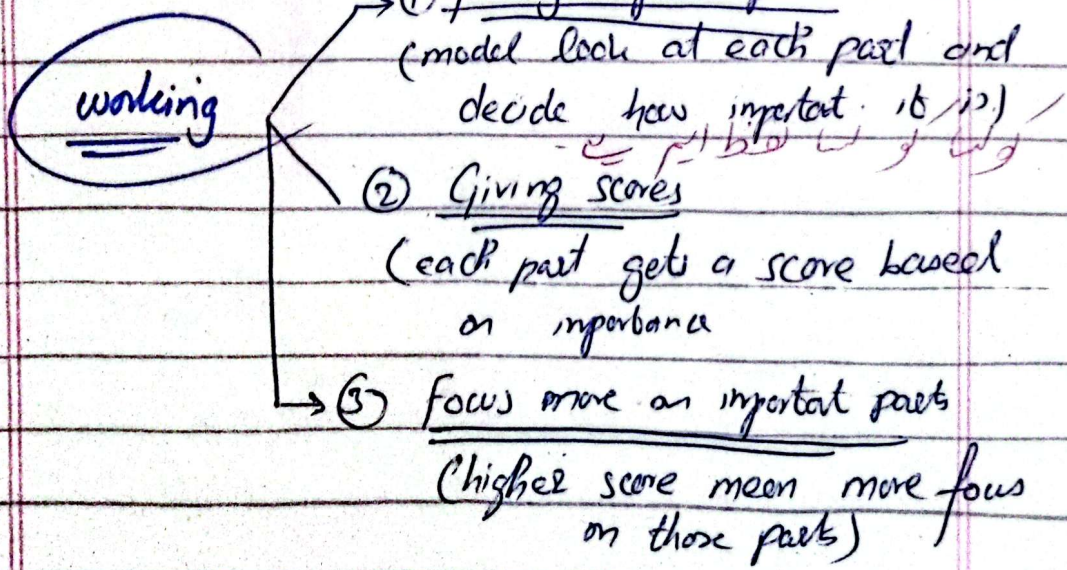
example

reading a long book & forgetting key details

Concept of Attention

Like highlighter. it highlights important information and then generate output considering that highlighter (Attention)

Such similar seq-seq model are called attention models.



Attention in actn

Translating a sentence with attention. Computer focus on key words for better translation

Example

" The Proffesor assigned homework on differential equation."

incorrect translation without attention:

ustad nay farq kay equation par homework diya

correct translate with attention

Professor ne differential equation par homework diya.

Attention — focus on important parts

helps model make better decision

can be applied b/w different input or different part of same input

Solve the problems of RNNs

(Attention)

- all neural network use weights.
- weights are random at start.

- when we train the model weights are trained.
- So they update values

2 > VT

- attention weights are random in start
- during training time the weights of attention.
