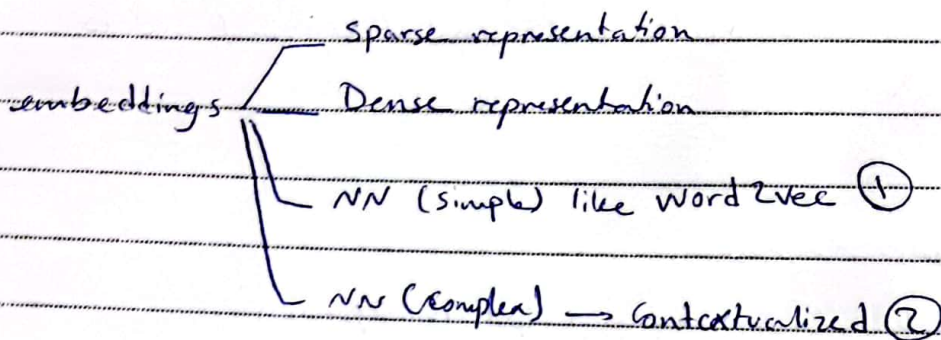


Subject:

Year, Month, Date, ()

1. Pre-trained Language models



① not containing ^{each} context information (a unique word only has one vector)

in ②, each word in each context has a vector

* ELMO (from RNN family)

giving words to LSTM blocks (2 layers) → backward

forward and backward have separate info and are concatenated at the end

* BERT

multi-head self attention, bidirectional transformer

(having forward and backward info at the same time)

2 layer (2 headed) bidirectionally connected ~~trans~~ transformer

→ train scenarios

2 steps: 1. Semi-supervised (Pre-train) 2. Supervised (fine-tune)

having only a few labeled data, we train and predict the unlabeled ones. selecting the ones with more confidence than a threshold, we will have

SAHEL ~~more~~ labeled data and can perform ~~several~~ supervised learning

Subject:

Year, Month, Date, ()

→ In NLP, text data is kind of supervised (labeled) since we always have the next word → also called semi-supervised

(BERT_{Base} → 12 layers encoder BERT_{Large} → 24 layers encoder)

we give 2 sentences to the model S_1, S_2 & $\langle \text{cls} \rangle S_1 \langle \text{sep} \rangle S_2$
(back to back)

Semi supervised part

① masked language model: removing some percentage of words to be predicted

(15% of input words usually)

② next sentence prediction: given S_1 and S_2 we want the $\langle \text{cls} \rangle$ token to output if S_2 is the next sentence of S_1 .

(1402 2 4)

Supervised part (fine tune)

fine tuning the weights learned in the previous part for a specific task

- tasks
- ① sentence pair classification → relation between 2 texts
 - ② single sentence classification → text classification
 - ③ Question Answering
 - ④ single sentence sequence labelling

① sentence pair classification

input 2 sentences → output in cls token output

paraphrase identification, answer retrieval, textual entailment

(is a paraphrase of) (is a question about)

② sentence classification

input cls token and first sentence → output of cls token

SAHEL

Subject:

Year. Month. Date. ()

③ Question Answering

پرسش و پاسخ از متن دست مطلب

→ جواب سوال با توجه به متن باشد

output: جواب سوال (span of answer from/to)

④ Sequence Labeling

* CLS token can output the representation for the whole text

* BERT can be used only for representation (no finetuning)

→ problem using BERT in mobiles (large model)

«Parsing»

هدف: بیان ارتباط ساختاری بین واژه‌های یک جمله

applications: spell checking, speech recognition, translation, LM

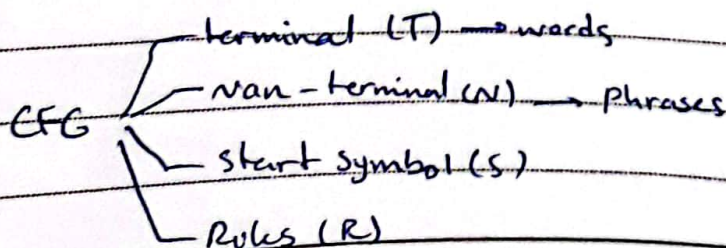
«constituency parsing»

→ phrase structure (each with a head)

ساختار گره‌های جمله را مشخص می‌کند و می‌تواند به عنوان یک ابزار برای تشخیص ساختار نحوی استفاده شود.

to parse a sentence we need a grammar

Context free grammar (CFG)



SAHEL