Subject:

Year. Month Date. ( )

(Pre-trained Language models)

embeddings
- Sparse representation
- Dense representation
  - NN (simple) like Word2vec (1)
  - NN (complex) → Contextualized (2)

(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 layer) + backward

forward and backward have separate infos and are concatinated at the end

* BERT

multi-head self attention, bidirectional transformer

(having forward and backward info at the same time)

2 layer (12 headed) bidirectionally connected transformer

→ train scenario:

2 steps: 1. semi supervised (pre-train) 2. supervised (fine tune)

having only a few labeled data, we train and predict the unlabled ones. selecting the ones with more confidence than a threshold, we will have more labled data and can perform supervised learning

SAHEL

Subject:

Year. Month. Date. ( )

→ In NLP, text data is kind of a supervised (labled) 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 & <cls> S_1 <sep> S_2

Semi supervised part (back to backe)

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 <cls> token to output if S_2 is the next sentence of S_1.

(1402 2 9)

Supervised part (finetune)

finetunning the weightes 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 labling

① Sentence pair classification
input 2 sentences → output is cls token output
paraphrase identification, answer retrival, textual entailment

② Sentence classification
input cls token and first sentence → output of cls token

SAHEL

Subject:

Year. Month. Date. ( )

③ Question Answering → جواب سوال باید درست باشد

پرسش و پاسخ از یک متن / یک مطلب

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 fine tunning)

→ problem using BERT in mobiles (large model)

«Parsing»

هدف: پیدا کردن ارتباط ساختاری بین واژه های یک جمله

applications: spell checking, speech recognition, translation, LM

«constituency parsing»

→ phrase structure (each with a head)

پیدا کردن گروه های متشکل جمله که می‌توانند با یکدیگر جابجا شوند اما داخل گروه ها

آنها جابجایی نیست وجود ندارد و معنی عوض می‌شود

to parse a sentence we need a grammar → Context Free grammar (CFG)

CFG
- terminal (T) → words
- non-terminal (N) → phrases
- start symbol (S)
- Rules (R)

SAHEL