Natural Language Modelling

PA154 Jazykové modelování (13)

Pavel Rychlý

pary@fi.muni.cz

May 25, 2021

Big models

- bigger is better
- many layers
- need big machines
- using advanced hardware: GPU, TPU

BERT

- Google
- pre-training on raw text
- masking tokens, is-next-sentence
- big pre-trained models available
- domain (task) adaptation

Input: The man went to the $[MASK]_1$. He bought a $[MASK]_2$ of milk .
Labels: $[MASK]_1$ = store; $[MASK]_2$ = gallon

Sentence A = The man went to the store.
Sentence B = He bought a gallon of milk.
Label = IsNextSentence

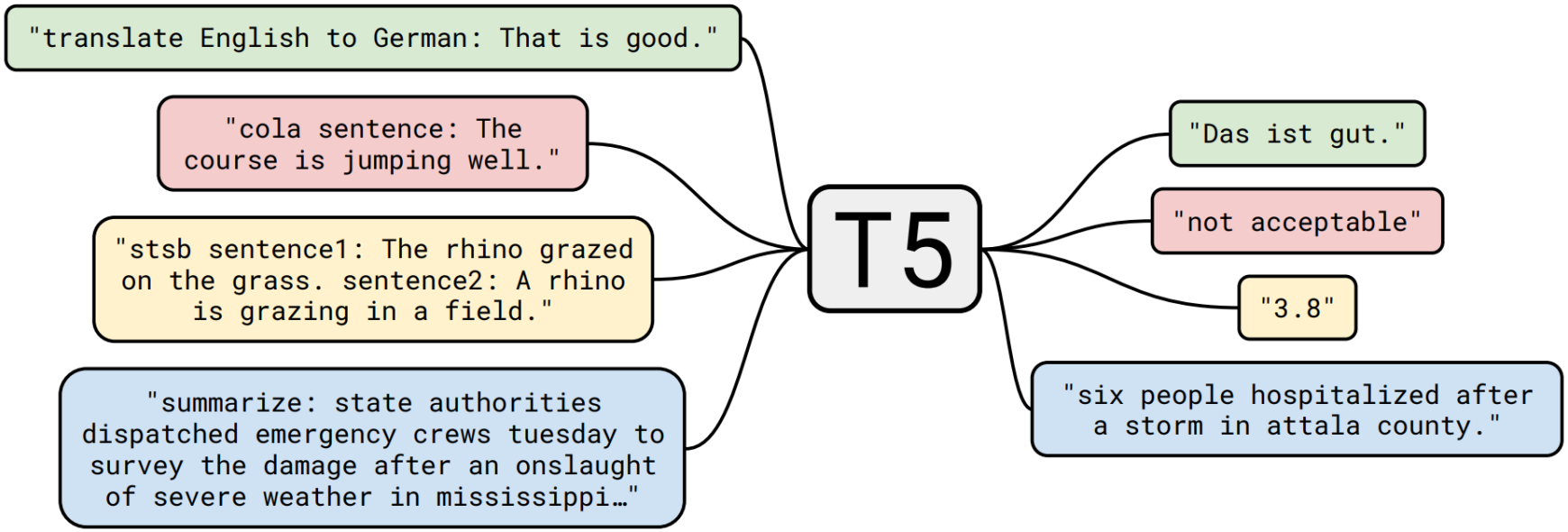
Sentence A = The man went to the store.
Sentence B = Penguins are flightless.
Label = NotNextSentence

GPT

- Open AI
- GPT-2: 1.5 billion parameters
- GPT-3: 175 billion parameters
- very good text generation
  $\rightarrow$ potentially harmful applications
- Misuse of Language Models
- bias – generate stereotyped or prejudiced content: gender, race, religion
- Sep 2020: Microsoft have "exclusive" use of GPT-3

T5: Text-To-Text Transfer Transformer

- Google AI
- transfer learning
- C4: Colossal Clean Crawled Corpus



Pretrained models

- huge training data
- long training time
- *small* model
- fine tuning on target task
- multi-language models
- universal tokenization: subword units
  - Byte-Pair Encoding (BPE)
  - WordPiece
  - SentencePiece

ALBERT

- A Lite BERT
- factorized embedding parameters
- cross-layer parameter sharing
- inter-sentence coherence loss
  Next Sentence Prediction $\rightarrow$ Sentence-Order Prediction
- much smaller: No. parameters: 108M $\rightarrow$ 12M (base)

Sentence A = `The man went to the store.`
Sentence B = `He bought a gallon of milk.`
Label = `IsNextSentence`

Sentence A = `The man went to the store.`
Sentence B = `Penguins are flightless.`
Label = `NotNextSentence`

Intrinsic evaluation

- direct evaluation of word embeddings
- semantic similarity (WordSim-353, SimLex-999, ...)
- word analogy (Google Analogy, BATS (Bigger Analogy Test Set))
- concept categorization (ESSLLI-2008)

Extrinsic evaluation

- using the model in a downstream NLP task
- Part-of-Speech Tagging, Noun Phrase Chunking, Named Entity Recognition, Shallow Syntax Parsing, Semantic Role Labeling, Sentiment Analysis, Text Classification, Paraphrase Detection, Textual Entailment Detection

Multi-task benchmarks

- GLUE (`https://gluebenchmark.com`)
  nine sentence- or sentence-pair language understanding tasks
- SuperGLUE (`https://super.gluebenchmark.com`)
  more difficult language understanding tasks
- XTREME – Cross-Lingual Transfer Evaluation of Multilingual Encoders
  (`https://sites.research.google/xtreme`)
  40 typologically diverse languages, 9 tasks

Libraries and Frameworks

- Dive into Deep Learning: online book
  `https://d2l.ai`
- Hugging Face Transformers: many ready to use models
  `https://huggingface.co/transformers`
- jiant: library, many tasks for evaluation
  `https://jiant.info`
- GluonNLP: reproduction of latest research results
  `https://nlp.gluon.ai`
- low level libraries: NumPy, **PyTorch**, TensorFlow, MXNet