



UNIVERSITÄT
ZU KÖLN

Large Language Models, Wissenschaftliche Poster

Analyse sozialer Medien mit NLP-Methoden

Nils Reiter


`nils.reiter@uni-koeln.de`

25. Januar 2024

Section 1

Large Language Models

Epochen und Veränderungen

- ▶ Drei Generationen von Systemen
 - ▶ 1950er – 2010: Regelbasierte Systeme
 - ▶ 2005 – 2015: Hybride und statistische Systeme
 - ▶ “klassisches machine learning”
 - ▶ Seit 2015: Neuronale Systeme, große Sprachmodelle
 - ▶ Pre-training / fine-tuning
 - ▶ Zero-shot learning
 - ▶ In-context learning
- ▶  Laufende Entwicklung

BERT & co.

- ▶ BERT/GPT have outperformed the state of the art in many tasks
- ▶ Breakthrough in natural language processing
- ▶ But not over night

BERT & co.

- ▶ BERT/GPT have outperformed the state of the art in many tasks
- ▶ Breakthrough in natural language processing
- ▶ But not over night
- ▶ General idea
 - ▶ Pre-training and fine-tuning on different tasks
 - ▶ Process whole input at once (max. 512 tokens, = bidirectional)
 - ▶ Encoder-Attention-Decoder architecture (= transformer)

Jacob Devlin/Ming-Wei Chang/Kenton Lee/Kristina Toutanova (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, S. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: <https://aclanthology.org/N19-1423>

Pre-Training and Fine-Tuning

- ▶ BERT models are trained on huge data sets
- ▶ Training one from scratch requires significant resources (time/money)
- ▶ (Some) pre-trained models are shared freely
- ▶ Recipe: Take a pre-trained model and fine-tune it on your task
 - ▶ Pre-trained model contains an abstract language representation

Pre-Training and Fine-Tuning

- ▶ BERT models are trained on huge data sets
- ▶ Training one from scratch requires significant resources (time/money)
- ▶ (Some) pre-trained models are shared freely
- ▶ Recipe: Take a pre-trained model and fine-tune it on your task
 - ▶ Pre-trained model contains an abstract language representation
- ▶ Fine-tuning
 - ▶ Any language-related task!

BERT Training Tasks

Masked Language Modeling (MLM)

- ▶ Sentence-wise
- ▶ 15% of the tokens are “masked” by a special token
- ▶ Model predicts these, having access to all other tokens

BERT Training Tasks

Masked Language Modeling (MLM)

- ▶ Sentence-wise
- ▶ 15% of the tokens are “masked” by a special token
- ▶ Model predicts these, having access to all other tokens

Next sentence prediction (NSP)

- ▶ Two (masked) sentences are concatenated
- ▶ Model has to predict whether second sentence follows on the first or not

Discussion

- ▶ Pre-training collects knowledge about language and world from large data set
- ▶ Fine-tuning adapts the model to a specific task

Discussion

- ▶ Pre-training collects knowledge about language and world from large data set
- ▶ Fine-tuning adapts the model to a specific task
- ▶ 'Knowledge': Neural networks, matrices, weights
 - ▶ Opaque for us
- ▶ But it's the first time that we can store generic linguistic knowledge that is 'specializable' to specific tasks

GPT, ChatGPT, OpenAI

- ▶ ChatGPT launch: November 30, 2022 (ca. 14 months)
- ▶ Since then: Extreme hype mode
 - ▶ 30 papers on “language model” uploaded to arxiv.org – yesterday

GPT, ChatGPT, OpenAI

- ▶ ChatGPT launch: November 30, 2022 (ca. 14 months)
- ▶ Since then: Extreme hype mode
 - ▶ 30 papers on “language model” uploaded to arxiv.org – yesterday
- ▶ Noteworthy for everyone
 - ▶ OpenAI is not open, many claims cannot be proven
 - ▶ Given the principled shortcomings of the models, might be difficult to find a sustainable business model
 - ▶ Language models generate probable word sequences – anthropomorphisation is not appropriate
 - ▶ E.g., the models do not “understand” anything

Usage Scenarios

- ▶ Extract contextualized embeddings
- ▶ Fill masked tokens
- ▶ Fine-Tuning for specific classification task
- ▶ Zero-Shot classification
- ▶ Few-Shot classification / In-Context learning

Section 2

Wissenschaftliche Poster

Einführung

- ▶ Klassische wiss. Kommunikation: Texte (Artikel/Bücher), Vorträge (Präsentationen)
- ▶ Relativ neue Form: Poster

Einführung

- ▶ Klassische wiss. Kommunikation: Texte (Artikel/Bücher), Vorträge (Präsentationen)
- ▶ Relativ neue Form: Poster
- ▶ Wie funktionieren Poster-Sessions?
 - ▶ Festes Zeitfenster
 - ▶ Die Posterpräsentatoren stehen vor ihrem Poster
 - ▶ Das Publikum geht umher, liest die Poster und spricht mit den Präsentatoren
 - ▶ Kleine Gruppen bilden sich vor dem Poster

Einführung

- ▶ Klassische wiss. Kommunikation: Texte (Artikel/Bücher), Vorträge (Präsentationen)
- ▶ Relativ neue Form: Poster
- ▶ Wie funktionieren Poster-Sessions?
 - ▶ Festes Zeitfenster
 - ▶ Die Posterpräsentatoren stehen vor ihrem Poster
 - ▶ Das Publikum geht umher, liest die Poster und spricht mit den Präsentatoren
 - ▶ Kleine Gruppen bilden sich vor dem Poster

Vorteil: Diskussion in kleinen Gruppen

- ▶ Offenere Diskussion
- ▶ Weniger Lampenfieber
- ▶ Vertiefung der Diskussion

Wie präsentiert man ein Poster?

- ▶ Zwei verschiedene Arten von Besuchern
- ▶ Menschen mit spezifischen Fragen
 - ▶ Sie springen auf ihre spezifischen Fragen an
 - ▶ Haben manchmal Missverständnisse über Dinge, die nicht zu ihrem Gebiet gehören
- ▶ Menschen ohne spezifische Fragen (das ist die Mehrheit)
 - ▶ Sind dankbar, wenn man ihnen die Führung gibt.
 - ▶ D.h., sie durch das Plakat führen

Wie bereitet man ein Poster vor

- ▶ Denken Sie an “die Tour” – die Geschichte, die Sie erzählen wollen
- ▶ Teilen Sie die Geschichte in verschiedene Abschnitte auf
 - ▶ Ähnlich wie bei Vorträgen
- ▶ Jeder Abschnitt erhält einen Bereich auf dem Plakat
- ▶ Bereiche werden ausgefüllt durch
 - ▶ Aufzählungspunkte
 - ▶ Schaubilder/Diagramme
 - ▶ Bilder/Fotos/Screenshots
 - ▶ Tabellen
- ▶ Vergessen Sie nicht den üblichen wissenschaftlichen Kram: Referenzen

demo

Poster für die DHCon 2024

- ▶ Fragestellung: Was haben Sie gelernt? / Was nehmen Sie mit?
- ▶ Fokus auf die Inhalte, “hübsch machen” kann man das später noch
- ▶ Gerne Material aus den Sitzungen übernehmen

References I



Devlin, Jacob/Ming-Wei Chang/Kenton Lee/Kristina Toutanova (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, S. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: <https://aclanthology.org/N19-1423>.