

Metódy transformérov a ich aplikácia v oblasti textovej analýzy

Študent: Martin Bača

Vedúci: RNDr. Šimon Horvát, PhD.

Konzultant: doc. RNDr. Ľubomír Antoni, PhD.

Motivácia

- Vytvoriť chatbot-a, ktorý pri poskytnutí textu a otázky viazanej k danému textu, dokáže spracovať text a nájsť v ňom odpoveď na danú otázku

Ciele

- ▶ 1. Spracovať prehľad nových metód neurónových sietí, transformérov, popísať ich princípy a oblasti ich súčasného využitia.
- ▶ 2. Implementovať metódy neurónovej siete založenej na transforméroch na tréning pomocou vybraných dátových súborov.
- ▶ 3. Otestovať natrénovaný model vo vybranej aplikačnej oblasti.
- ▶ 4. Porovnať dosiahnuté výsledky s inými dostupnými štúdiami.

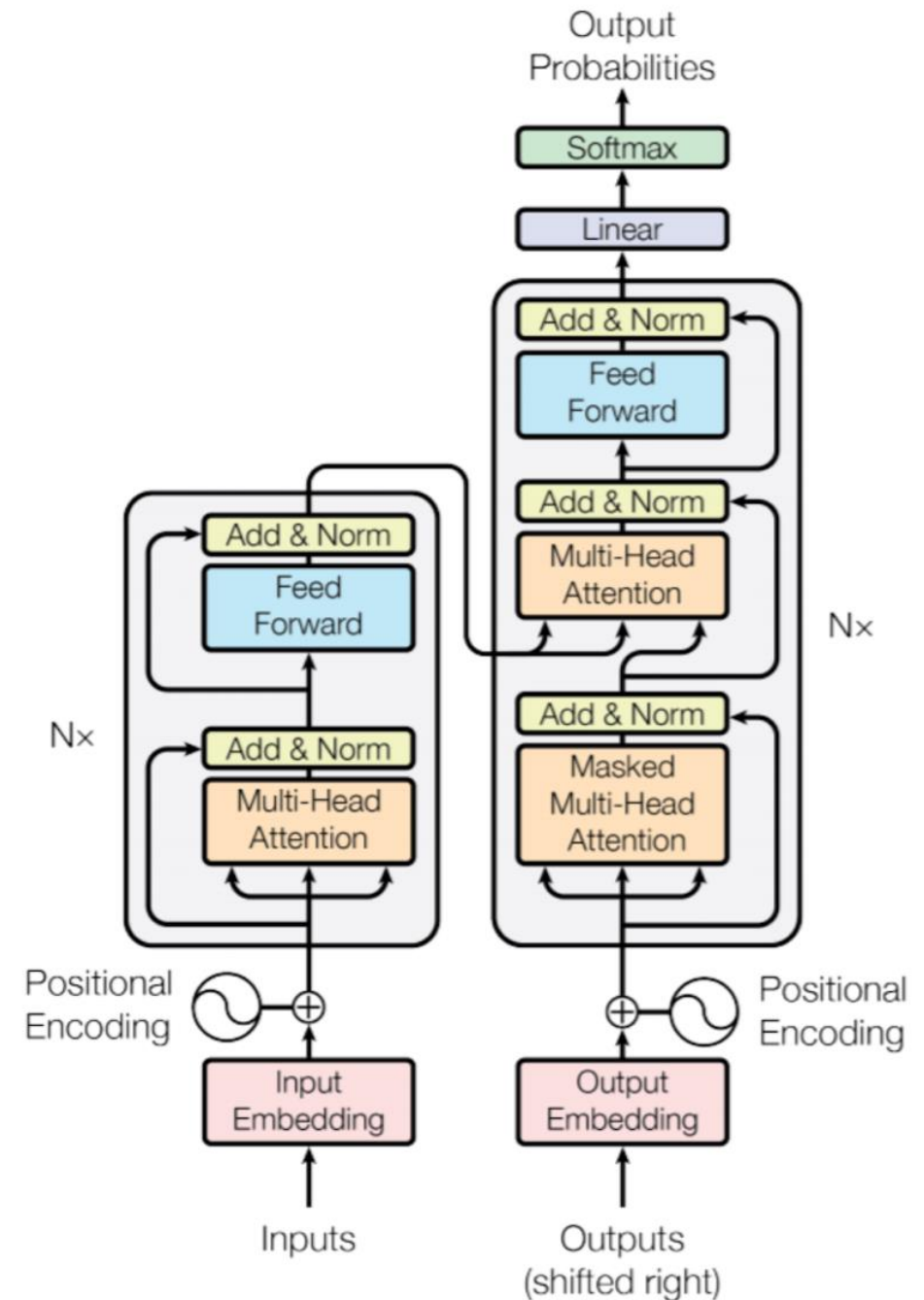
Transforméry

- ▶ Model transformérov je neurónová sieť, ktorá sa učí kontext a tým aj významprostredníctvom sledovania vzťahov v sekvenčných údajoch
- ▶ Architektúra bola prvý krát predstavená v článku “Attention Is All You Need” vedcami zo spoločnosti Google
- ▶ Dnes je to štandardne používaná technika v oblastiach ako je spracovanie prirodzeného jazyka, počítačové videnie, generovanie obsahu a učenie posilňovaním

Transforméry

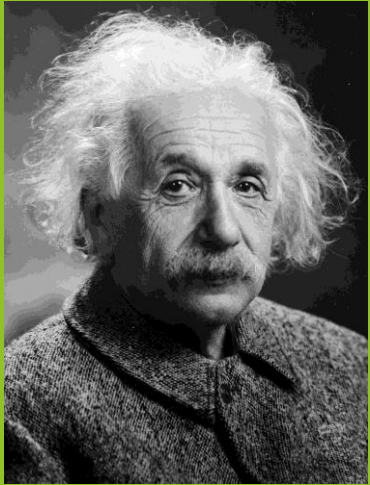
- Zložené z 2 častí:

- Encoder - transformuje vstupy na latentné reprezentácie, ktoré obsahujú mechanizmus pozornosti (dokáže zachytávať vzťahy a súvislosti). Výstupom tejto časti je vektorová reprezentácia vstupu
- Decoder - generuje výstup za použitia reprezentácií získaných z encoderu spolu s mechanizmom pozornosti na predchádzajúce výstupy. Týmto spôsobom postupne vytvára novú sekvenciu, pričom zohľadňuje aj predošlé generované texty

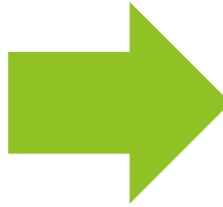


Ilustračný príklad

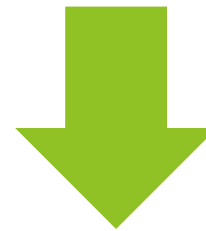
Text



Albert Einstein bol švajčiarsko-americký teoretický fyzik nemecko-židovského pôvodu. Narodil sa 14. marca 1879 v nemeckom Ulme.



Model



Otázka

Kedy sa narodil Albert Einstein?

Odpoved'

Albert Einstein sa narodil 14. marca 1879.

Vykonané kroky

- ▶ Štúdium odporúčanej literatúry
- ▶ Študovanie pozadia a využitia transformérov
- ▶ Hľadanie vhodného datasetu

Ďalšie kroky

Rozšírenie datasetu - generovanie otázok

- ▶ Dataset obsahuje len texty - otázky k nim je potrebné vygenerovať
- ▶ Prínosom práce je okrem iného aj preskúmanie generovania otázok na základe textov, čo pomôže automatizácii vytvárania datasetu
- ▶ Na generovanie otázok použijeme už existujúci jazykový model Ollama
- ▶ Ku každej otázke bude uložená aj odpoveď a jej pozícia v texte
- ▶ Údaje budú uložené vo forme JSON-u

Budovanie modelu

- ▶ Použijeme už natrénovaný model transformérov
- ▶ Tento model upravíme pre naše potreby
- ▶ Budeme sledovať ako si model poradí s odpovedaním na otázky a budeme ho usmerňovať a vylepšovať

Odporúčaná literatúra

- ▶ 1. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
- ▶ 2. Tunstall, L., Von Werra, L., & Wolf, T. (2022). Natural language processing with transformers. " O'Reilly Media, Inc.".

Ďakujem za pozornosť