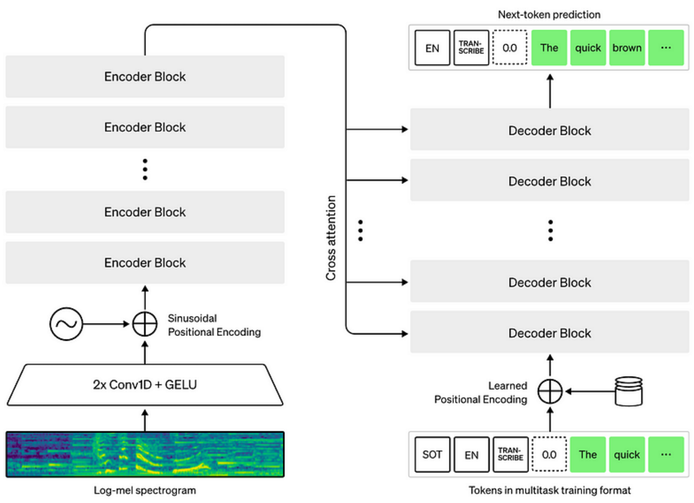
Fino podešavanje Whisper modela za izradu AI asistenta

Glasovni AI asistent na srpskom jeziku

Tina Mihajlović, Softversko inženjerstvo i infromacione tehnologije

Uvod

Whisper je **ASR** (*Automatic Speech Recognition*) model razvijen od strane OpenAI-a. Baziran je na klasičnoj enkoder-dekoder transformer arhitekturi.
Treniran je nas preko 680,000 sati audio snimaka na više od 96 jezika, između ostalog i na srpskom.



Arhitektura Whisper modela

Cilj projekta jeste poboljšanje vrednosti metrike (WER - *word error rate*) **"small" Whisper** modela za srpski jezik, te primena tako unapređenog modela u integraciji sa ChatGPT Turbo3.5 API-jem i Google-ovog TTS (*text to speech*) modela za implementaciju glasovnog pametnog asistenta.

Metodologija

Projekat se razvijao u dve etape:

- Fino podešavanje i treniranje Whisper modela
- Primena istreniranog modela za izradu pametnog asistenta

Fino podešavanje i treniranje

Za treniranje je korišćeno *Google Colab* okruženje sa GPU akceleracijom. Učitavanje pre-treniranog Whisper small modela kao i čitavog skupa podataka ze trening rađeno je preko *Hugging Face* platforme.

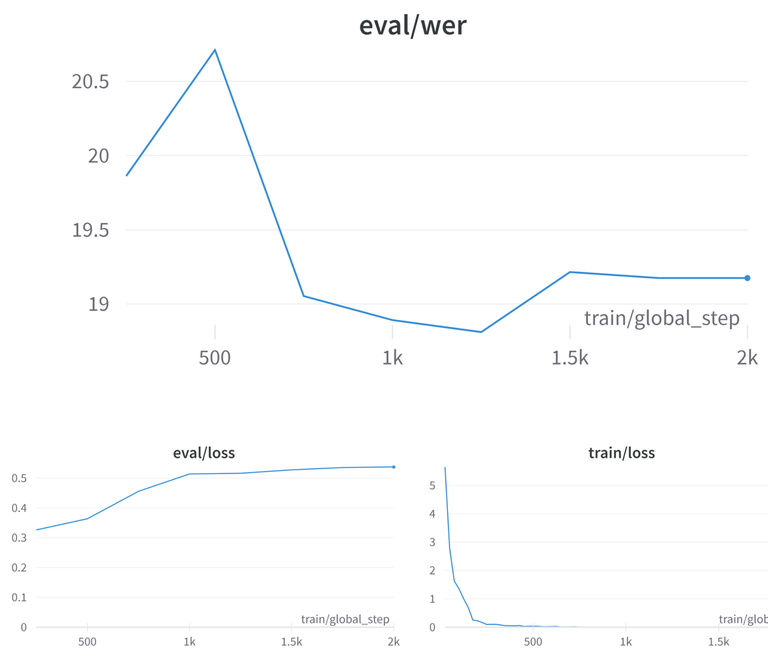
Korišćen je **common_voice_11_0** skup podataka, deo na srpskom jeziku, koji sadrži labelirane podatke namenjene treningu ASR modela.
Tokom treniranja je uzet podskup podataka koji objedinjuje test i validation split (ukupno oko 1.3h audio materijala), dok je za evaluaciju korišćen test split seta podataka.

Izrada AI glasovnog asistenta

Nakon što je fino podešavanje i trening modela urodio zadovoljavajućim rezultatima, model je iskorišćen za transkripciju izgovorenog pitanja u tekst, taj tekst se šalje **ChatGPT** API-u koji vraća odgovor na pitanje, nakon čega **Google-ov TTS** model odgovor izgovara nazad korisniku.

Rezultati

Tokom finog podešavanja parametara model je pokretan više puta sa različitim konfiguracijama kako bi se empirijski došlo do optimalnih vrednosti parametara.
Poslednje treniranje modela je izvršeno u 2000 koraka, sa evaluacijom na svakih 250, tokom čega su dobijeni sledeći rezultati:



Funkcija gubitka (*loss* funkcija) treniranja i WER opada gotovo kontantno iz iteracije u iteraciju, što pokazuje da model dobro uči na podacima.

Međutim, evaluciona loss funkcija raste sa svakom iteracijom, što dovodi u pitanje poboljšanje mogućnosti generalizacije modela - ovo je ponašanje koji se primećuje i u sličnim istraživanjima i radu na finom podešavanju Whisper small modela. Povećanje loss funkcije je minimizovano dodavanjem dropout parametra, augmentacijom podataka i povećanjem skupa podataka.

Dalje poboljšanje bi se potencijalno moglo ostvariti značajnijim povećanjem trening skupa i dodatnom randomizovanom augmentacijom dela tog skupa - što se ostavlja za dalje istraživanje.

Detaljniji pregled vrednosti parametara po iteracijama dat je u tabeli:

Step	Training Loss	Validation Loss	Wer
250	0.095400	0.326219	19.862460
500	0.033200	0.363609	20.711974
750	0.005000	0.455462	19.053398
1000	0.001400	0.513732	18.891586
1250	0.000500	0.515995	18.810680
1500	0.000300	0.527528	19.215210
1750	0.000300	0.535277	19.174757
2000	0.000300	0.537420	19.174757

Modeli na kraju svake od evalucionih tačaka su sačuvani i dostupni za korišćenje (**checkpoint**-i).
Za AI asistenta je iskorišćen model na koraku 1250, budući da je za glavnu metriku performanse modela uzet WER.

WER	eval/loss
18.81	**0.51**

WER i vrednost funkcije gubitka tokom evaluacije za korak 1250

Gledajući kao glavnu metriku vrednost validacione loss funkcije, za najbolji model bio bi izabran model na koraku 250 - potpuno validan izbor ostavljem korisnicima.

Zaključak

Whisper small checkpoint model u svom originalnom stanju pokazuje korektne performanse transkripcije na srpski jezik. Međutim, sa relativno malim setom trening podataka, ograničenom procesnom moći i vremenom treniranja, postignuti su zadovoljavajući rezultati u pogledu smanjenja WER-a početnog modela. Istraživanje daje metodologiju i konfiguraciju parametra za uspešno fino podešavanje modela, kao i motivaciju za dalji rad i unapređenja.

AI asistent nastao kao primena istreniranog modela, pokazuje izuzetnu tačnost transkripcije, dok su minorne greške u istoj gotovo neprimetne za ChatGPT API koji uspešno pogađa suštinu pitanja i pored tih grešaka.

Fakultet tehničkih nauka
Novi Sad, Srbija
jul 2023.