Elektronski fakultet Niš



**Prevođenje teksta u govor, i prevodjenje govora u tekst korišćenjem Google API-ja i python-a.**

Tim: pow(B,2)J

David Bosnić 16489

Nikola Begović 16481

Jovana Jović 16639

Sadržaj

[Odeljak 1 - Kratak opis problema 3](#_Toc61721627)

[Odeljak 2 - Pregled i kratak opis tehnika i algoritama 3](#_Toc61721628)

[*Hidden Markov model (HMM)* 3](#_Toc61721629)

[*Neuronske mreže* 3](#_Toc61721630)

[*Duboke neuronske mreže* 4](#_Toc61721631)

[*NLP algoritmi* 4](#_Toc61721632)

[Odeljak 3 - Formulacija problema 4](#_Toc61721633)

[Odeljak 4 - Kratak opis rešenja 5](#_Toc61721634)

[Reference 5](#_Toc61721635)

Odeljak 1 - Kratak opis problema

Rešavanje problema prevođenja govora u tekst i teksta u govor zasniva se na tranformaciji ulaznih zvučnih signala registrovanih na ulaznom uređaju u tekstualni zapis i formiranju izlaznog zvučnog signala (govora) na osnovu ulaznog teksta, respektivno.

Sinteza govora (TTS) i prepoznavanje govora(STT) uključuju mnoge discipline kao sto su akustika, lingvistika, digitalna obrada signala, računarske nauke itd. To su vrhunske tehnologije u oblasti obrade informacija, naročito za trenutne inteligentne sisteme za interakciju govorom.

Odeljak 2 - Pregled i kratak opis tehnika i algoritama

I akusticno modeliranje i modeliranje jezika su važni delovi modernih statičkih algoritama za prepoznavanje govora. Hidden Markov model (HMM) je široko korišćen u mnogim sistemima. U današnje vreme najkorišćeniji su algoritmi i modeli zasnovani na neuronskim mrežama i principima mašinskog učenja.

Hidden Markov model (HMM)

Moderni savremeni sistemi za prepoznavanje govora opšte namene se zasnivaju na Hidden Markov-om modelu. To su statički modeli koji daju sekvencu simbola ili veličina. HMM se koristi za prepoznavanje govora jer se govorni signal može posmatrati kao komandni stacionarni signal ili kratkotrajni stacionarni signal. U kratkom vremenskom periodu (npr 10ms), govor se može smatrati stacionarnim procesom.

Neuronske mreže

Neuronske mreže se upotrebljavaju kao karakteristike transfomacija ili redukcije dimenzionalnosti, korak pre HM modela. Međutim, u novije vreme *“Long short-term memory”* - (LSTM), povezane rekurentne neuronske mreže *“Recurrent neural network” -* (RNNs) i neuronske mreže sa vremenskim odlaganjem *“Time Delay Neural Networks”* - (TDNN's) su pokazale poboljšanje performansi u ovoj oblasti.

Duboke neuronske mreže

Duboka neuronska mreža (DNN -*“Deep neural network”*) je veštačka neuronska mreža sa više skrivenih slojeva jedinica između ulaznog i izlaznog sloja. Slično plitkim neuronskim mrežama, DNN arhitekture generišu kompozitne modele gde viši slojevi omogućavaju sastavljanje karakteristika iz nižih slojeva dajući ogroman kapacitet za učenje, a i potencijal modeliranja složenih obrazaca govornih podataka.

NLP algoritmi

NLP algoritmi su obično zasnovani na algoritmima mašinskog učenja. Umesto ručnog kodiranja velikih skupova pravila, NLP se može osloniti na mašinsko učenje da bi ih automatski naučilo analizirajući niz primera (tj. poput knjige, zbirke rečenica) i donoseći statički zaključak. Generalno, što više podataka bude analizirano, model će biti tačniji.

Odeljak 3 - Formulacija problema

Pri prevođenju teksta u govor cilj je da dobijeni zvučni signal bude što precizniji, bez naglih promena frekvenije izlaznog signala. Pri prevođenju govora u tekst potrebno je da ulazni signal bude što čistiji, odnosno bez primesa koje bi uzrokovale distorziju samog signala i time umanjile kvalitet prevođenja istog.

Kada se koriste HMM i DNN metode za analizu teksta, sistem se obično koristi za izvršavanje fonemske segmentacije i anotacije. On obično sadrži sledećih pet nivoa:

* Fonetski nivo
* Nivo sloga
* Nivo reči
* Nivo fraze
* Nivo rečenice

Analiza kroz ovih pet nivoa omogućava izlazni zvučni signal bude gladak i što približniji prirodnom ljudskom govoru, upravo zbog sinteze govora kroz duboku analizu svih segmenata govora.

Google koristi kombinaciju tehnika dubokog učenja i obrade prirodnog jezika (NLP) da bi raščlanio naš upit, preuzeo odgovor i predstavio ga u obliku zvuka i teksta. Takođe Google koristi HMM kroz analizu što kraćeg intervala zvučnog signala, radi postizanje što veće konstantnosti i konzistentnosti samog određivanja glasa.

Odeljak 4 - Kratak opis rešenja

Za ovaj projekat, što se tiče prepoznavanja zvuka, korišćene su sledeće biblioteke: PyAudio i SpeechRecognition. Speech Recognition biblioteka nam omogućava prepoznavanje govora kroz interakciju sa različitim engine-ima i API-ima. U našem slučaju to su: Google Speech Recognition i Google Cloud Speech API.

Kada radimo prepoznavanje govora kroz mikrofon, potrebno je da prvi snimimo govor, a zatim tako snimljen govor prosledimo Google speech to text recognition engine-u, koji će da obavi prepoznavanje i vrati nam prepoznati tekst. Prolikom slanja zvučnog zapisa, specificiramo jezik koji želimo da bude prepoznat i preveden (u našem slučaju imamo opciju za engleski („en-US“) i srpski („sr“).

Što se tiče prevođenja teksta u govor, korišćene su sledeće biblioteke : gTTS (Google Text-To-Speech) i playsound. gTTS je pzthone biblioteka koja lako konvertuje tekst u zvuk, dok playsound koristimo za puštanje audio fajlova.

Kod prevođenja teksta u govor, potrebno je da prihvatimo uneti tekst kroz formu i da ga prosledimo funkciji gTTS uz naziv jezika na kojiželimo da dobijemo govor. Nakon toga dobijeni audio fajl se reprodukuje korisniku pozivom funkcije playsound i čuva u memoriji.

Reference

https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a

https://heartbeat.fritz.ai/the-3-deep-learning-frameworks-for-end-to-end-speech-recognition-that-power-your-devices-37b891ddc380

https://towardsdatascience.com/easy-speech-to-text-with-python-3df0d973b426

https://www.javatpoint.com/how-to-convert-text-to-speech-in-python

https://towardsdatascience.com/easy-text-to-speech-with-python-bfb34250036e

https://ai.googleblog.com/2015/08/the-neural-networks-behind-google-voice.html?m=1

https://medium.com/sciforce/text-to-speech-synthesis-an-overview-641c18fcd35f

https://www.quora.com/What-speech-recognition-algorithms-are-used-by-Google

https://www.mdpi.com/2076-3417/9/19/4050/htm

https://en.m.wikipedia.org/wiki/Speech\_recognition