

Generisanje game muzike na osnovu slike

Uvod

- Cilj programa je da na osnovu slike igrice generiše muziku. Program je podeljen na 2 dela, gde jedan služi za generisanje početne muzike na osnovu slike, a drugi na osnovu početne muzike generiše ostatak muzike.
- Muzika se generise u formi MIDI file-ova

Skup podataka

- Za skup podataka smo "scrap-ovali" sa sledeca dva sajta
- <https://www.vgmusic.com/> - za MIDI file-ove
- <https://www.giantbomb.com/> - za slike
- ukupno 4.1GB podataka
- MIDI file-ove smo pokupili tako sto smo svaku pesmu skinuli pomocu seleniuma
- dok smo slike pokupili tako sto bi vrsili pretragu na "giantbomb" sajtu sa imenom igrice, nakon cega bi skinuli slike iz rezultata – naravno koristeći Selenium

Metodologija

MIDI transformacije

- [Transformacija](#) MIDI tip 1 file-ova u tip 0
- [Transformacija](#) MIDI u tekst
- Normalizacija podataka – tempo, otkucaj, jacina zvuka
- Izbacivanje bespotrebnih metapodataka
- Uproscavanje I svodjenje MIDI-ja na 2 instrukcije
- Kompresija I svodjenje MIDI-ja na 2 instrukcije – vremensku I notnu.

Slika u tekst

- Korišten je model koji uči da generise tekst (note) na osnovu slike. Korišten je dropout 0.5, 2 dense sloja i LSTM sloj za rekurentnu mrežu. Aktivaciona funkcija je bila relu, dok za poslednji sloj je korišten softmax.
- Mreža bi od slike generisala kompresovanu verziju MIDI file-a koja bi sluzila kao input za GPT-2 mrežu

GPT-2

- Za generisanje pesme iz pocetnih nota pesme (generisane od strane prethodne mreže), koristili smo vec implementiranu [GPT-2 mrežu](#) sa pretreniranim modelom od 124MB sa jos dodatnim fine-tuningom nad nasim skupom podataka
- Parametri koje smo koristili su - Temperature=0.7, top_k=0, n_samples= 1, batch_size=1.
- Nakon generisanja kompresovane verzije, file bi se dekompresovao, zatim pretvorio iz teksta u MIDI.

Rezultati I zakljucak

- Sistem uspesno generise muziku koja ima ritmickog smisla I nije bolna za slusanje
- Sistem znacajno bolje generise kracu muziku – do 30 sekundi – nego duzu
- Kod odredjenih kombinacija nota, GPT-2 upadne u "loop" gde ih konstantno ponavlja inkrementirajuci jacinu zvuka svaki put.
- GPT-2 nije najbolji izbor, bilo bi bolje da smo koristili regresioni model koji je predvidjen da barata muzikom I ima "bolji osecaj za vreme"
- Sistem bi se bolje ponasao kada bi imali dovoljno resursa da iskoristimo sve nase podatke za treniranje

