



Elektronski fakultet u Nišu

Augmentacija audio podataka u mašinskom učenju

Student: Danilo Milošević

Profesor: Aleksandar Stanimirović

Sadržaj

1.	Uvod	3
2.	Osnove audio signala	4
2.1	Jedan	4
2.2	Dva	4
2.3	Tri	4

1. Uvod

U toku poslednje decenije smo svedoci sve boljih i naprednijih modela veštačke inteligencije, kako kod obrade slika, teksta, videa tako i kod audio podataka. Kao i kod drugih oblasti mašinskog učenja, performanse modela zavise dosta od kvaliteta i količine podataka nad kojima se trenira model. Prikupljanje velike količine kvalitetnih audio podataka (pogotovo anotiranih) je dosta skupo, vremenski zahtevno a u nekim slučajevima i nemoguće. Zbog ovoga primenjujemo augmentaciju podataka.

Augmentacija audio podataka, i uopšte bilo koja vrsta augmentacije podataka, predstavlja skup tehnika kojima se od postojećeg skupa podataka generišu dodatni podaci primenom različitih transformacija. Za razliku od augmentacije slika, koja je danas standardna praksa, augmentacija audio podataka ima dodatne izazove zbog vremenske prirode audio signala, kao i kompleksnosti percepције zvuka kod ljudi.

Osnovni cilj augmentacije je da se poveća raznovrsnost i veličina trening skupa bez sakupljanja novih podataka, čime se model čini otpornijim na varijacije koje se mogu desiti tokom primene modela (šum npr.). Jedan primer je prepoznavanje ljudskog govora. Ukoliko je model treniran na ljudskim glasovima snimljenim u studiju, postoji mogućnost da će performanse modela biti loše u realnim uslovima - na ulici, u kafiću, kod kuće itd.

U ovom radu ćemo prvenstveno obraditi šta je zvuk, kako se može opisati i predstaviti, kao i kakve različite tehnike augmentacije podataka. Razmotrićemo kako jednostavnije tako i kompleksnije transformacije, u različitim domenima, kada i kako treba primeniti date tehnike kao i njihove prednosti i nedostatke. Na kraju ćemo spomenuti kako možemo meriti efikasnost tehnika augmentacije koje će kasnije biti obrađene u praktičnom delu projekta.

2. Osnove audio signala

2.1 Jedan

2.2 Dva

2.3 Tri