

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 720

**SUSTAV ZA PREPOZNAVANJE GOVORNIH
NAREDBI U STVARNOM VREMENU NA
RUBNIM UREĐAJIMA**

Luka Balić

Zagreb, 14. veljače 2025.

Zagreb, 30. rujna 2024.

DIPLOMSKI ZADATAK br. 720

Pristupnik: **Luka Balić (0036513380)**

Studij: Računarstvo

Profil: Računalno inženjerstvo

Mentor: prof. dr. sc. Hrvoje Džapo

Zadatak: **Sustav za prepoznavanje govornih naredbi u stvarnom vremenu na rubnim uređajima**

Opis zadatka:

Proučiti metode i algoritme za analizu i klasifikaciju zvučnih sekvenci u stvarnom vremenu, s naglaskom na pristupe temeljene na dubokom učenju i primjenu u prepoznavanju govornih naredbi. Proučiti programske okvire za razvoj modela prikladnih za prepoznavanje govornih naredbi i implementaciju u ugradbenim računalnim sustavima s ograničenim resursima. Proučiti mogućnosti i ograničenja u primjeni procesora iz porodice ESP32 za implementaciju rješenja za prepoznavanje glasovnih naredbi u stvarnom vremenu te odabrati prikladnu procesorsku porodicu. Implementirati pokazni sustav koji će omogućiti prepoznavanje glasovnih naredbi u stvarnom vremenu. Razviti programsku potporu za mikrokontroler koja će omogućiti snimanje zvuka preko mikrofona, izvođenje modela za prepoznavanje glasovnih naredbi te poduzimanje jednostavnih akcija na svaku prepoznatu naredbu. Detaljno dokumentirati postupak treniranja i prilagodbe modela za izvođenje na ugradbenom računalnom sustavu. Eksperimentalno provesti ispitivanje značajki sustava i uspješnosti prepoznavanja glasovnih naredbi pod različitim uvjetima rada.

Rok za predaju rada: 14. veljače 2025.

Hvala svima...

Sadržaj

1. primjer

Neki od radova koje ćemo citirati su [?, ?, ?, ?, ?, ?]. Trebaju nam samo radi testiranja kako izgleda referenciranje rada s konferencije, rada iz časopisa, knjige i Internetske stranice.



Slika 1.1. Moja prva slika

Referenciramo se na sliku ?? u sredini rečenice, zatim prije zareza ??, te zatim na kraju rečenice ?? Upravo smo testirali radi li naredba \ref ispravno u slučaju kada nakon nje slijedi točka.

Sada slijedi jedna jednačba:

$$\int_{-\infty}^{+\infty} f(t) dt = F(\omega) \tag{1.1}$$

Jednačba (??) je moja prva jednačba koja defnira par $f(t) \circ \bullet F(\omega)$ ili $F(\omega) \bullet \circ f(t)$.

2. Uvod

2.1. Tiny ML

Modeli strojnog učenja revolucionarizirali su tehnologiju koju koristimo u svakodnevnom životu tako što su omogućili računalima učenje iz podataka kojima raspolažu u svrhu donošenja odluka u situacijama za koje nisu eksplicitno programirana. Tradicionalno su takvi modeli bili namijenjeni za računala visokih performanci i neograničavajućih resursa, međutim sve bržim razvojem IoT (eng. Internet of Things) područja te zahvaljujući pristupačnoj cijeni mikrokontrolerskih sustava, počela je prilagodba modela strojnog učenja za takve sustave. Na prvi pogled dva nespojiva svijeta su se susrela te pronašla svoju primjenu u raznovrsnim sustavima. Tiny ML (eng. Tiny Machine Learning) je naziv koji se odnosi na implementaciju modela strojnog učenja na uređaje ograničenih resursa kao što su mobilni telefoni i mikrokontroleri. Glavne karakteristike modela namijenjenih za takve sustave su relativno malen memorijski otisak te mogućnost odziva u stvarnom vremenu. Sustav implementiran kroz ovaj rad ima zadatak prepoznati unaprijed zadane glasovne naredbe te pokrenuti izvršavanje određenog posla vezanog uz specifičnu naredbu.

2.2. Opis problema

Okruženi smo digitalnim glasovnim asistentima kao što su Googleov Assistant, Appleova Siri te Amazonova Alexa. Ovakvi sustavi mogu u vrlo kratkom roku pružiti zatražene informacije te bez ikakvog problema komunicirati s osobom koja ih koristi. Za prepoznavanje i obradu ljudskog govora te dohvaćanje bitnih informacija zaduženi su modeli kojima je potrebna velika procesorska moć i dovoljno prostora za pohranu te se zbog toga taj dio posla odrađuje na serverskim računalima. Takav sustav podrazumijeva

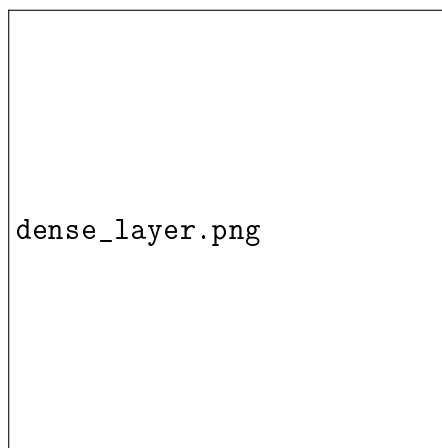
konstantnu internetsku vezu te stabilni i dugotrajni izvor električne energije. Kada bi mobilni uređaji slali konstantan tok audio podataka na server, brzo bi ispraznili bateriju te nepotrebno koristili mobilne podatke za pristup internetu. Zbog toga su takvi sustavi osmišljeni da čekaju naredbu za početak komunikacije, a tek onda uspostave vezu sa serverom. Međutim, i dalje nam ostaje problem konstantne akvizicije ulaznih audio podataka te prepoznavanje naredbe kao što je "Hey Google" ili nešto slično. U ovoj situaciji savršenu primjenu pronašli su procesori izrazito male potrošnje na kojima je moguće implementirati optimizirane modele strojnog učenja. Takav procesor bi konstantno akvizirao podatke s audio ulaza te lokalno, uz pomoć treniranog modela, čekao ključnu riječ te nakon prepoznavanja dao znak cijelom sustavu da se može "probuditi" iz stanja niske potrošnje te odraditi svoj posao. Ovakvim pristup, postignuta je efikasnost, niska potrošnja, brz odaziv, smanjena je potrošnja internetskih podataka te možda i najvažnija stvar - privatnost. Naime, nema potrebe za konstantnim slanjem glasovnih podataka na server što omogućuje da na server dospiju samo glasovni isječci u trenucima u kojima želimo.

3. Struktura sustava

4. Neuronska mreža

Okosnica cjelokupnog sustava za prepoznavanje izgovorenih naredbi je neuronska mreža. Neuronska mreža ili preciznije umjetna neuronska mreža je računalni model inspiriran biološkom strukturom neurona u mozgu čovjeka. Sastoji se od čvorova (neurona) i jednosmjernih veza između njih (sinapsa) koji na taj način tvore usmjereni graf. Čvorovi su grupirani u slojeve, a svaki od njih je povezan s čvorovima iz susjednog sloja na određeni način. Način na koji su određeni slojevi međusobno povezani određuje vrstu sloja.

Jednostavan primjer strukture neuronske mreže je potpuno povezani sloj (engl. fully connected layer ili dense layer) koji se često koristi kao osnovni građevni blok u umjetnim neuronskim mrežama. U potpuno povezanom sloju, svaki čvor jednog sloja povezan je sa svakim čvorom susjednog sloja. Ovakva struktura omogućava mreži fleksibilno učenje složenih odnosa između ulaznih i izlaznih podataka.



Slika 4.1. Potpuno povezani sloj

$$\int_{-\infty}^{+\infty} f(t) dt = F(\omega) \quad (4.1)$$

5. Zaključak

zajkkjbdv

Sažetak

Sustav za prepoznavanje govornih naredbi u stvarnom vremenu na rubnim uređajima

Luka Balić

Unesite sažetak na hrvatskom.

Ključne riječi: prva ključna riječ; druga ključna riječ; treća ključna riječ

Abstract

A system for recognizing voice commands in real-time on edge devices

Luka Balić

Enter the abstract in English.

Keywords: the first keyword; the second keyword; the third keyword

Privitak A: The Code