SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1176

NENADZIRANO UČENJE ZNAČAJKI GOVORA KORIŠTENJEM NEURONSKIH MREŽA BAZIRANIH NA AUTOENKODERSKIM ARHITEKTURAMA

Luka Murn

Zagreb, lipanj 2018.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA
ODBOR ZA DIPLOMSKI RAD PROFILA

Zagreb, 9. ožujka 2018.

DIPLOMSKI ZADATAK br. 1176

Pristupnik: **Luka Murn (0036474720)**
Studij: Informacijska i komunikacijska tehnologija
Profil: Obradba informacija

Zadatak: **Nenadzirano učenje značajki govora korištenjem neuronskih mreža baziranih na autoenkoderskim arhitekturama**

Opis zadatka:

Nenadzirano učenje značajki i niždimenzionalnih reprezentacija podataka nalaze primjene u strojnom učenju, kompresiji s gubitkom i sl. Postojeće metode poput analize osnovnih komponenata (PCA) se često baziraju na linearnim transformacijama podataka, dok su modernije metode bazirane na neuronskim mrežama bolje opremljene za prepoznavanje nelinearnih odnosa u podacima. U okviru diplomskog rada potrebno je implementirati sustav za nenadzirano učenje značajki govora korištenjem neke od autoenkoderskih arhitektura neuronskih mreža. Dodatno, potrebno je evaluirati implementaciju nad nekim od klasičnih glasovnih klasifikacijskih problema (klasifikacija emocija, verifikacija govornika, i sl.) ili treniranjem nad specifičnim glasovnim podacima ili korištenjem prijenosa učenja (eng. transfer learning). Niždimenzionalne reprezentacije podataka potrebno je usporediti ovisno o strukturi ulaznih glasovnih podataka (čisti glasovni podaci, glasovni podaci složeni u 2D oblik, spektrogrami, mel-filterirani spektrogrami, itd.). Optimalnu arhitekturu mreže (dubinu/širinu komponenti) potrebno je evaluirati i diskutirati u ovisnosti o dostupnoj količini podataka za učenje/testiranje.

Zadatak uručen pristupniku: 16. ožujka 2018.
Rok za predaju rada: 29. lipnja 2018.

Mentor:
Prof. dr. sc. Davor Petrinović

Djelovođa:
Izv. prof. dr. sc. Marko Subašić

Predsjednik odbora za diplomski rad profila:
Prof. dr. sc. Sven Lončarić

Zahvaljujem se mentoru prof. dr. sc. Davoru Petrinoviću i asistentu Igoru Mijiću, mag. ing. na strpljenju, razumijevanju i ukazanoj pomoći.

SADRŽAJ

1. Uvod

Govorni signal je najbrži i najprirodniji način komunikacije između ljudi. Ova činjenica potaknula je istraživače da razmišljaju o govoru kao o brzoj i učinkovitoj metodi interakcije između čovjeka i stroja. Međutim, to zahtijeva da stroj mora biti dovoljno inteligentan za prepoznavanje ljudskog glasa. Od kraja pedesetih godina provedena su mnoga istraživanja u području prepoznavanja govora. Unatoč velikom napretku u spomenutom području, još uvijek nije moguća prirodna interakcija između čovjeka i stroja jer stroj ne razumije emocionalno stanje govornika.

Ovo je dovelo do novog područja istraživanja, prepoznavanja emocija u govoru, koje je definirano kao ekstrakcija emocionalnog stanja govornika iz njegovog ili njezinog govora. Vjeruje se da se prepoznavanje emocija u govoru može koristiti za izdvajanje korisne semantike iz govora i time poboljšati učinak sustava za prepoznavanja govora [74]. Prepoznavanje emocija u govoru osobito je korisno za primjene koje zahtijevaju prirodnu interakciju čovjeka i stroja, kao što su tutorske aplikacije na računalima čiji odgovor korisniku ovisi o njegovoj emociji [86]. Također je korisno za upravljačku ploču u automobilu gdje informacije o emocionalnom stanju vozača mogu biti dane sustavu kako bi povećali njegovu sigurnost [86]. Može se koristiti i kao dijagnostički alat za terapeute [42]. U aplikacijama za e-učenje, prepoznavanje emocija u govoru se može iskoristiti za poboljšanje iskustva učenika prilagođavanjem dostave materijala za učenje na temelju njihovih emocionalnih stanja [101]. Stoga ne iznenađuje da je prepoznavanje emocija u govoru tijekom posljednjih desetljeća jedna od glavnih tema istraživanja u obradi govora, interakciji čovjeka i računala i kompjuterski posredovane ljudske komunikacije [77, 89, 19, 3].

Općenito, prepoznavanje emocija u govoru usredotočuje se na korištenje metoda strojnog učenja kako bi se automatski predvidjela "ispravna" emocionalna stanja iz govora. Cilj algoritama strojnog učenja je otkrivanje statističke strukture podataka. Konkretno, algoritmi učenja reprezentacije pokušavaju pretvoriti neobrađene podatke u obrazac iz kojeg je lakše obavljati nadzirane zadatke učenja, poput klasifikacije. To je osobito važno kada je klasifikator koji prima ovu reprezentaciju kao unos linearan i

kada je broj dostupnih labeliranih primjera mali.

Učenje s prijenosom značajki (engl. *feature transfer learning*) je predloženo kao mogući put u razvoju metoda za prijenos korisnih informacija iz jedne ili više izvornih neuronskih mreža u povezanu ciljnu mrežu [106]. Empirijski i teoretski se pokazuje da navedeno učenje može uvelike poboljšati uspješnost učenja, osobito kada je samo mali broj podataka dostupan u ciljnoj domeni. Zbog toga je jasno kako prepoznavanje emocija u govoru može imati koristi od upotrebe učenja s prijenosom značajki.

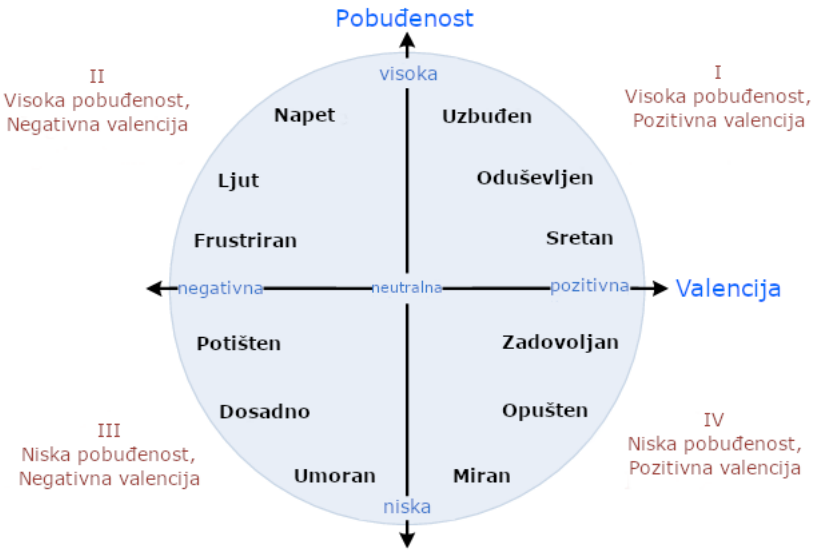
U nedavnim je radovima predloženo učenje s prijenosom značajki temeljeno na metodi rijetkih autoenkodera - vrsti neuronske mreže s ograničenjima rijetkosti na skrivenim slojevima - za otkrivanje znanja u akustičnim značajkama kako bi se poboljšale performanse prepoznavanja govornih emocija pri primjeni tog znanja na izvorne podatke [27, 26]. Pristup učenja značajki pomoću rijetkog autoenkodera sastoji se od dvije faze: prvo se nauči reprezentacija pomoću jednoslojnog autoenkodera koji se trenirao na primjerima iz podskupa za treniranje. Zatim se ova reprezentacija primjenjuje na testni podskup te se potom koristi za klasifikacijski zadatak kako bi se izgradili standardni emocionalni modeli.

Upravo zato će se u ovom radu iskoristiti učenje s prijenosom značajki uz rijetke autoenkodere kako bi se zatim, uz korištenje klasifikatora, obilježile izvučene značajke iz govornog signala prema iskazanim emocijama. U posljednje vrijeme pojavio se trend u zajednici strojnog učenja prema izvođenju reprezentacije ulaznog signala izravno iz sirovih, neobrađenih podataka [107]. Motivacija iza ove ideje je da, u konačnici, mreža automatski uči posrednu reprezentaciju neprekinutog ulaznog signala koji bolje odgovara ciljnom zadatku i time dovodi do poboljšane izvedbe.

2. Prepoznavanje emocija u govoru

2.1. Osnovna podjela emocija

Emocija nema zajednički dogovorenu teorijsku definiciju [55]. Međutim, ljudi prepoznaju emocije kad ih osjete. Zbog toga su istraživači mogli proučavati i definirati različite vrste emocija. Razni su autori predložili da postoji od dviju pa do dvadeset osnovnih ili prototipnih emocija [79, 65]. Najčešće četiri koje se pojavljuju na ovim popisima su: strah, bijes, tuga i radost. Neki su autori manje zabrinuti s osam prototipnih emocija i prvenstveno se fokusiraju na dimenzije emocija, poput negativnih ili pozitivnih emocija. Naširoko je prihvaćeno da se emocija može karakterizirati u dvije dimenzije: "pobuđenost" (mirno/uzbuđeno) i "valencija" (negativna/pozitivna) [41], kao što je vidljivo na Slici 2.1.



Slika 2.1: Dvodimenzionalni prostor valencija-pobuđenost [112]

Pobuđenost se odnosi na količinu energije potrebnu za izražavanje određene emocije. Prema nekim fiziološkim studijama mehanizma proizvodnje emocija [36], otkriveno je da se simpatički živčani sustav pobuđuje emocijama radosti, ljutnje i straha.

To uzrokuje povećanu brzinu otkucaja srca, viši krvni tlak, suhoću usta i povremeno podrhtavanje mišića. Dobiveni govor je zato glasan, brz i izražen jakom visokofrekventnom energijom, višom prosječnom visinom glasa i širim rasponom visine glasa. S druge strane, s pobuđivanjem parasimpatičkog živčanog sustava, recimo s tugom, otkucaji srca i krvni tlak se smanjuju, a salivacija se povećava, stvarajući govor koji je spor, slab, i s malom visokofrekventnom energijom. Dakle, akustične značajke kao što su visina, dužina, kvaliteta glasa i artikulacija govornog signala vrlo su korelirane s ishodišnim emocijama [18]. Međutim, emocije se ne mogu razlikovati samo pomoću aktivacija. Na primjer, emocije i bijesa i sreće odgovaraju visokim aktivacijama, ali one prenose različite izražaje. Ovu razliku karakterizira dimenzija valencije. Na žalost, među istraživačima nema dogovora o tome kako, ili čak ako, se akustične značajke povezuju s tom dimenzijom [67]. Stoga, iako se klasifikacija između emocija s visokom aktivacijom (koje se nazivaju i visoka pobuđenost) i emocija s niskom aktivacijom može postići uz visoku točnost, razvrstavanje između različitih emocija i dalje je izazovan problem. No ima smisla pojednostaviti moguće kategorije emocija za računala, kako bi mogli početi s prepoznavanjem jednostavnih, najočitijih emocija.

Za afektivno računarstvo, problemi prepoznavanja i modeliranja pojednostavljeni su pretpostavkom malog broja diskretnih emocija ili malog broja dimenzija. Ako su poželjne eksplicitne dimenzije, mogao bi se izračunati svojstveni prostor promatranih značajki i tražiti prepoznatljiva "svojstvena raspoloženja". Ona mogu odgovarati čistim emocijama ili mješavini emocija. Dobiveni svojstveni prostori bili bi zanimljivi za uspoređivanje s prostorima koji se nalaze u istraživanjima faktorskih emocija; takvi prostori obično uključuju osi valencije (pozitivna, negativna) i pobuđenosti (mirno, uzbuđeno). Prostori se mogu procijeniti u različitim uvjetima, kako bi se bolje karakterizirale značajke izražavanja emocija i njihove ovisnosti o vanjskim (npr., ekološkim) i kognitivnim (npr., osobnim) značajkama. Mogu se obilježiti putanje u tim prostorima, koje bi obuhvatile dinamičke aspekte emocija. Uzimajući da se jedan od tih modela dimenzija prostora, treniran na pojedinačnim izlazima, a zatim i značajkama nepoznatog izlaza, može sakupljati u vremenu i koristiti s alatima kao što je *maximum a posteriori* odlučivanje da se prepozna nova nerazvrstana emocija. Budući da se prepoznavanje emocionalnog stanja može postaviti kao nadzirani klasifikacijski problem, tj. onaj gdje su klase navedene *a priori*, raspoloživi su različiti modeli prepoznavanja uzoraka te učenja značajki [34, 104].

2.2. Modeli za prepoznavanje emocija u govoru

Sustavi koji prepoznaju parajezične signale temeljene na govoru, kao što su sustavi klasifikacije emocija, općenito djeluju u dvije široke faze. Prednji sustav koji ekstrahira značajke karakteristične za parajezične informacije od interesa i pozadinski sustav koji donosi klasifikacijske odluke temeljene na tim značajkama. Značajke su gotovo uvijek vektorske reprezentacije govornih signala te se klasifikacijske odluke temelje na razlikama u statističkim svojstvima distribucije značajki vektora. Posljedično, performanse sustava klasifikacije emocija u govoru ovise o dva čimbenika, o stupnju do kojega se razlikuju statistička svojstva distribuiranih vektora značajki iz govora koji odgovaraju različitim emocijama i točnost s kojom se te razlike mogu modelirati u pozadinskom sustavu. Prvi faktor određuje gornju granicu točnosti klasifikacije bilo kojeg sustava klasifikacije emocija s obzirom na skup značajki, dok drugi faktor vodi do razlika u točnosti klasifikacije različitih sustava.

Idealno, statistička svojstva distribucije vektorskih značajki značajno će se razlikovati između različitih emocija (emocionalna varijabilnost) i ne razlikuju se zbog bilo kojeg drugog razloga. Međutim, u stvarnosti oni također značajno variraju zbog razlika između različitih govornika (varijabilnost govornika), zbog razlika u jezičnom sadržaju (fonetska varijabilnost), kao i razlika u ostalim parajezičnim detaljima [99]. Ti dodatni izvori varijabilnosti zauzvrat utječu na "klasifikacijske zakone" koji su izvedeni u pozadinskom sustavu i degradiraju učinak klasifikacije [6, 17, 89]. Iako su fonetska i varijabilnost govornika vjerojatno najznačajnije utječu na sustav klasifikacije emocija, smatra se da je varijabilnost govornika veći problem u mnogim uobičajenim značajkama [98].

Kulturni aspekti su među najznačajnijim varijancama koje se mogu pojaviti kada se istovremeno koriste različiti skupovi za dizajn sustava prepoznavanja emocija [95]. Stoga je važno sustavno ispitati moguće razlike i razviti strategije za rješavanje kulturne raznovrsnosti u emocionalnom izražavanju. Da bi se bolje nosili s razlikama u skupovima, predlaže se prilagodba skupova značajki [38], poduzorkovanje primjera skupa, a ne uzimanje svih podataka [90], dodavanje neoznačenih podataka za samotreniranje sustava [113], sintetiziranje dodatnih podataka [92], ili primjena metoda učenja s prijenosom znanja kako bi podaci bili "sličniji" [29].

2.3. Smanjivanje dimenzionalnosti i ekstrakcija značajki

Kako podaci sve više postaju visokodimenzionalni, poput genomske informacije, slika, videozapisa i teksta, smanjivanje dimenzionalnosti za generiranje reprezentacije na visokoj razini ne smatra se samo važnim, već i neophodnim korakom pretprocesiranja podataka. Iako modeli strojnog učenja teoretski trebaju raditi na bilo kojem broju značajki, visokodimenzionalni skupovi podataka uvijek dovode do niza problema, uključujući prenaučenost, visoku računalnu složenost i pretjerano kompliciranе modele, što dovodi do dobro poznatog problema - prokletstva dimenzionalnosti [50]. Drugi razlog za smanjenje dimenzionalnosti jest što reprezentacije na visokoj razini mogu pomoći ljudima da bolje razumiju intrinzičnu strukturu podataka.

Različite metode smanjenja dimenzionalnosti [110, 25, 1] mogu se grubo podijeliti u dvije kategorije: odabir značajki i ekstrakcija značajki. Glavna razlika između njih leži u iskorištavanju dijela ili svih ulaznih značajki. Na primjer, odabir značajki pronalazi podskup svih značajki, a broj odabranih značajki manji je od izvornika, dok ekstrakcija značajki generira novi skup značajki koje su nastale kao kombinacija izvornih. Odabir najboljeg podskupa ulaznih varijabli je nedeterminističko-polinomijalni kombinatorni problem. Štoviše, metode odabira značajki ocjenjuju svaku varijablu neovisno, no varijable koje zasebno ne daju korisne informacije mogu to učiniti kada se zajedno koriste. Zbog toga su se pojavile druge mogućnosti, prvenstveno ekstrakcija [68]. Predložene su početne metode ekstrakcije značajki [53] zasnovane na projekciji, odnosno mapiranju ulaznih značajki iz izvornog visokodimenzionalnog prostora u novi niskodimenzionalni prostor, pritom minimizirajući gubitak informacija.

Pokazano je da autoenkoderske mreže s jednim skrivenim slojem izvlače osnovne komponente podataka [5]. Takve su mreže korištene za ekstrakciju značajki i razvijanje kompaktnog kodiranja podataka. Analiza osnovnih komponenata (engl. *principal component analysis*) projicira podatke u linearni potprostor s minimalnim gubitkom podataka, množenjem podataka pomoću svojstvenih vektora kovarijacijske matrice uzorka. Proučavanjem veličine odgovarajućih svojstvenih vrijednosti može se procijeniti minimalna dimenzionalnost prostora u koji se podaci mogu projicirati i procijeniti gubitak. Međutim, ako podaci leže na nelinearnom potprostoru prostora značajki, tada će osnovne komponente precijeniti dimenzionalnost. Dodavanje skrivenih slojeva i nelinearnih elemenata između ulaza i reprezentacijskog sloja, te između sloja reprezentacije i izlaza gradi mrežu koja je sposobna za učenje nelinearnih reprezentacija [56, 75, 108]. Takve mreže mogu izvesti nelinearnu projekciju analognu analizi osnovnih komponenata i izvući "osnovne značajke".

2.4. Učenje s prijenosom znanja

Uobičajeni algoritmi strojnog učenja tradicionalno se bave izoliranim zadacima. Učenje s prijenosom znanja (engl. *transfer learning*) pokušava to promijeniti pomoću razvijanja metoda prijenosa znanja naučenih u jednom ili više izvornih zadataka i upotrebe za poboljšanje učenja u povezanom ciljnom zadatku [106]. Metode koje omogućuju takvo učenje predstavljaju napredak prema tome da strojno učenje bude jednako učinkovito kao ljudsko učenje.

U induktivnom zadatku učenja cilj je izvesti prediktivni model iz skupa primjera za treniranje [72]. Često je cilj klasifikacija, tj. dodjeljivanje oznaka klase primjerima. Prediktivni model naučen induktivnim algoritmom učenja trebao bi dati točna predviđanja ne samo na primjerima za treniranje, već i na budućim primjerima koji dolaze iz iste distribucije. Da bi se proizveo model s ovom generalizacijskom sposobnošću, algoritam učenja mora imati induktivnu pristranost - skup pretpostavki o pravoj raspodjeli podataka za treniranje. Odabir ili prilagodba induktivne pristranosti ciljnih zadataka temelji se na znanju prenesenom iz izvornih zadataka. Način na koji se to radi ovisi o tome koji induktivni algoritam učenja se koristi za učenje izvornih i ciljnih zadataka. Neke metode prijenosa sužavaju prostor hipoteza, ograničavajući moguće modele ili uklanjajući korake pretraživanja. Druge metode proširuju prostor te se pretraživanjem omogućuje otkrivanje složenijih modela, kao i dodavanje novih koraka pretraživanja.

Među različitim načinima učenja s prijenosom znanja, duboke neuronske mreže koje imaju mnogo skrivenih slojeva i koje se treniraju primjenom novih metoda pokazale su da odgovaraju za spomenuto učenje [9]. Prethodni radovi [10, 12] su pokazali da je duboka arhitektura neophodna za kompaktno predstavljanje mnogih funkcija i takve arhitekture vode do korisnih reprezentacija koje idealno razdvajaju faktore varijacija prisutnih u ulaznim podacima. Jedna od uspješnih primjena dubokih neuronskih mreža proizlazi iz iskorištavanja informacija u susjednim okvirima u području prepoznavanja govora i korištenja zavisnih stanja ovisnih o kontekstu u akustičkom modeliranju, koja nadmašuje najsuvremenije metode prepoznavanja govora, ponekad i s velikom razlikom [32, 45].

Učenje značajki, odnosno učenje nekih transformacija podataka koji olakšavaju izdvajanje korisnih informacija pri gradnji klasifikatora ili drugih prediktora, razmatrano je iz mnogih perspektiva unutar područja strojnog učenja [28, 13, 27, 111]. Ključna ideja učenja značajki jest korištenje dubokih arhitektura, što rezultira apstraktnom reprezentacijom. Općenito, apstraktni koncepti nepromjenjivi su za većinu lokalnih pro-

mjena ulaza.

Nakon koncepta učenja značajki, predloženo je učenje s prijenosom značajki kako bi se riješio problem ponovnog korištenja prethodno naučenih znanja iz "drugih" podataka ili značajki [76]. Ovo prilično bitno obilježje sugerira da bi učenje s prijenosom značajki bilo dobro prilagođeno scenarijima gdje je distribucija podataka u testnoj domeni različita od one u području treninga, ali gdje zadatak ostaje isti [28, 27]. Na primjer, predloženo je učenje s prijenosom značajki temeljeno na metodi rijetkih autoenkodera za ekstrakciju znanja iz akustičkih značajki s malim brojem označenih ciljnih podataka kako bi se poboljšala učinkovitost prepoznavanja govornih emocija kod izvornih podataka [27].

3. Nenadzirane metode učenja u neuronskim mrežama

Algoritmi nenadziranog učenja za cilj imaju otkriti strukturu skrivenu u podacima i naučiti reprezentacije koje su prikladnije kao ulaz nadziranom modelu od sirovog ulaza. Mnoge nenadzirane metode temelje se na rekonstrukciji ulaza iz reprezentacije, pritom ograničavajući reprezentaciju da ima određena poželjna svojstva (npr. niska dimenzija, rijetkost itd.). Drugi se temelje na aproksimaciji gustoće stohastičkom rekonstrukcijom ulaza iz reprezentacije.

Jedna od glavnih svrha nenadziranog učenja je proizvodnja dobre reprezentacije podataka koji se mogu koristiti za detekciju, prepoznavanje, predikciju ili vizualizaciju. Dobre reprezentacije uklanjaju nevažne varijabilnosti ulaznih podataka, uz očuvanje informacija korisnih za krajnji zadatak. Jedan od uzroka interesa za nenadzirano učenje je mogućnost stvaranja dubokih hijerarhija značajki slaganjem jednog nenadziranog modula na drugi [46, 12, 70, 80]. Nenadziranom modulu na jednom sloju u hijerarhiji se šalju reprezentacijski vektori proizvedeni u sloju ispod. Viša razina reprezentacije obuhvaća visoke razine ovisnosti između ulaznih varijabli, čime se poboljšava sposobnost sustava da obuhvati temeljne pravilnosti podataka. Izlaz zadnjeg sloja u hijerarhiji može se unijeti u konvencionalni nadzirani klasifikator. Nenadzirane inicijalizacije nastoje izbjeći lokalne minimume i povećati stabilnost učinka mreže [37].

Prirodni način dizajniranja sustava za nenadzirano učenje koji se može slagati jedan na drugi je paradigma enkoder-dekoder [81]. Enkoder pretvara ulaz u reprezentaciju (također poznat kao kod ili vektor značajki), a dekoder rekonstruira ulaz (možda stohastički) iz reprezentacije. Analiza osnovnih komponenata, autoenkoderske neuronske mreže te ograničeni Boltzmannovi strojevi samo su primjeri ove vrste arhitekture. Arhitektura enkodera/dekodera je atraktivna iz dva razloga: 1. nakon treninga, računanje koda je vrlo brz proces koji se sastoji samo u provođenju ulaza kroz enkoder; 2. rekonstrukcija unosa s dekoderom osigurava način za provjeru je li kod zabilježio relevantne podatke u podacima.

3.1. Autoenkoderske mreže

Autoenkoderi su neuronske mreže sa simetričnom strukturom, gdje srednji sloj predstavlja enkodirane ulazne podatke. Autoenkoderi se treniraju za rekonstrukciju njihovog ulaza na izlaz, uz zadovoljavanje određenih ograničenja koja sprječavaju jednostavno kopiranje podataka kroz mrežu. U svom najjednostavnijem obliku, autoenkoder se sastoji od dva dijela, enkodera i dekodera. Uveden je u kasnim osamdesetim godinama [35, 5] kao metoda za smanjivanje dimenzionalnosti, pri čemu izlaz enkodera predstavlja smanjenu reprezentaciju i gdje je dekoder podešen za rekonstrukciju inicijalnog unosa iz enkoderske reprezentacije kroz minimiziranje funkcije gubitka. Kada su aktivacijske funkcije enkodiranja linearne te je broj skrivenih jedinica manji od dimenzije ulaza (time formirajući usko grlo), pokazalo se da su naučeni parametri enkodera potprostor osnovnih komponenti ulaznog prostora [5]. Međutim, uz upotrebu nelinearnih aktivacijskih funkcija, može se očekivati da će autoenkoder naučiti korisnije detektore značajki od onih koje se mogu dobiti jednostavnom analizom osnovnih komponenti [52].

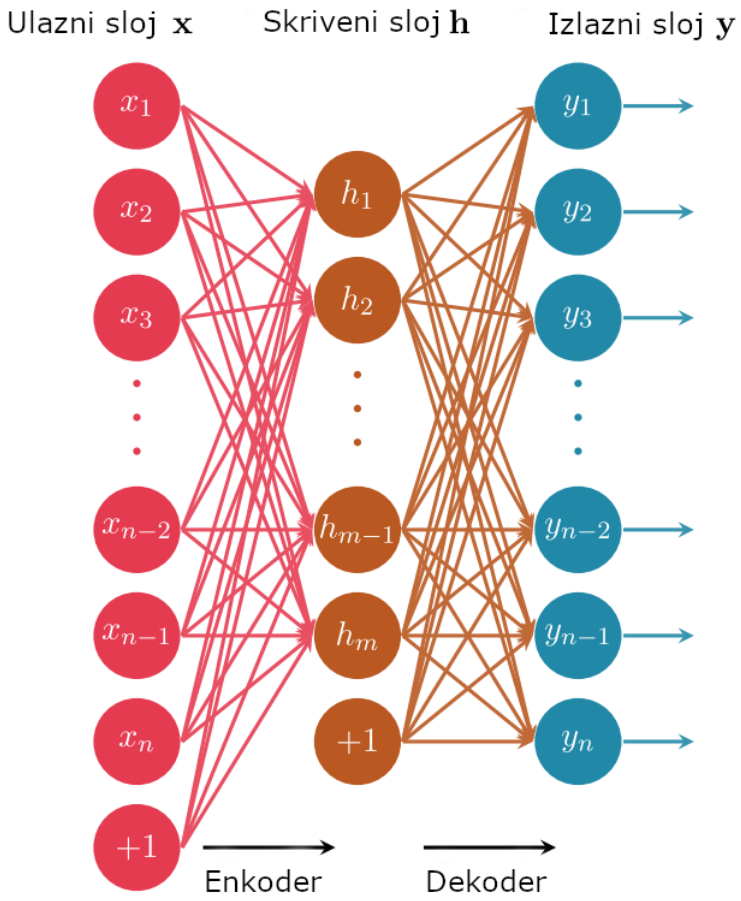
U novije vrijeme, autoenkoderi su ponovno zauzeli glavnu poziciju u pristupu "dubokom arhitekturom" [47, 46, 10, 37], gdje se slojevi autoenkodera slažu i nenadzirano treniraju od dna prema vrhu. Zbog toga autoenkoderi i podvrste kao što su rijetki autoenkoderi [27], autoenkoderi s uklanjanjem šuma [109] i varijacijski autoenkoderi [44], pokazuju obećavajuću sposobnost izdvajanja važnih značajki, posebno u obradi slike i obradi prirodnog jezika. Faza treniranja od dna prema vrhu je indiferentna s obzirom na konačni cilj i stoga se očigledno može koristiti u pristupima učenja s prijenosom znanja [4].

3.1.1. Osnovni autoenkoder s jednim skrivenim slojem

Tradicionalni model autoenkodera [12], poput osnovne neuronske mreže, sastoji se od ulaznog sloja, skrivenog sloja i izlaznog sloja, kao što je prikazano na Slici 3.1. Formalno, kao odgovor na ulazni primjer $\mathbf{x} \in \mathbb{R}^n$, skrivena reprezentacija $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^m$, ili kod je

$$\mathbf{z}^{(1)} = \mathbf{W}^{(1)} \cdot \mathbf{x} + \mathbf{b}^{(1)} \tag{3.1}$$

$$\mathbf{h} = f(\mathbf{z}^{(1)}) \tag{3.2}$$



Slika 3.1: Arhitektura autoenkodera. Autoenkoder se sastoji od ulaznog sloja, skrivenog sloja i izlaznog sloja [26]

gdje je $f(\cdot)$ određen kao aktivacijska funkcija (obično se koristi logistička sigmoidalna funkcija ili nelinearni tangens hiperbolni primijenjen na svaku komponentu zasebno), $\mathbf{W}^{(1)} \in \mathbb{R}^{m \times n}$ je težinska matrica, i $\mathbf{b}^{(1)} \in \mathbb{R}^m$ je vektor pomaka. Ovaj proces koji nelinearno pretvara ulaz u novu reprezentaciju poznat je kao enkoder. Nedovoljno kompletan autoenkoder odgovara autoenkoderu u kojem je broj ulaznih jedinica veći od skrivenih jedinica, tj. $n > m$. Enkoder obično daje prikaz robusniji od originalnog ulaza, koji se može primijeniti na sljedeći proces. Dekoder mapira skrivenu reprezentaciju $\mathbf{h}$ natrag na rekonstrukciju $\mathbf{y} \in \mathbb{R}^n$:

$$\mathbf{z}^{(2)} = \mathbf{W}^{(2)} \cdot \mathbf{h} + \mathbf{b}^{(2)} \tag{3.3}$$

$$\mathbf{y} = f(\mathbf{z}^{(2)}) \tag{3.4}$$

gdje je $\mathbf{W}^{(2)} \in \mathbb{R}^{n \times m}$ težinska matrica, i $\mathbf{b}^{(2)} \in \mathbb{R}^n$ je vektor pomaka. Ako su dvije težinske matrice ograničene da budu u obliku $\mathbf{W}^{(2)} = (\mathbf{W}^{(1)})^T$, to se naziva vezanim

težinama i time se smanjuje broj prilagodljivih parametara. S obzirom na skup ulaznih primjera $\mathbf{X}$, treniranje autoenkodera sastoji se od pronalaženja skupova parametara $\theta = \{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(1)}, \mathbf{b}^{(2)}\}$, što odgovara minimizaciji sljedeće ciljne funkcije:

$$\mathcal{J}^{AE}(\theta) = \sum_{\mathbf{x} \in \mathbf{X}} L(\mathbf{x}, \mathbf{y}) \tag{3.5}$$

gdje je L pogreška rekonstrukcije. Tipični primjeri funkcija uključuju kvadratnu pogrešku $L(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||^2$ upotrijebljenu u slučaju linearne rekonstrukcije i gubitak unakrsne entropije kad je aktivacijska funkcija sigmoidalna (i ulazi su u intervalu $[0, 1]$): $L(\mathbf{x}, \mathbf{y}) = -\sum_{i=1}^{N} x_i \log(y_i) + (1 - x_i) \log(1 - y_i)$.

Minimizacija se obično ostvaruje pomoću učenja s unatražnim rasprostiranjem (engl., *backpropagation*) sa stohastičkim gradijentnim spustom ili naprednijim metodama optimizacije kao što su algoritam Broyden-Fletcher-Goldfarb-Shanno (L-BFGS), konjugirana gradijentna metoda ili ELM metode.

3.1.2. Aktivacijske funkcije

Jedinica koja se nalazi u bilo kojem od skrivenih slojeva neuronske mreže prima nekoliko ulaza iz prethodnog sloja. Jedinica izračunava težinsku sumu ovih ulaza i na kraju izvodi određenu operaciju, tzv. aktivacijsku funkciju, kako bi se dobio izlaz. Nelinearnost ponašanja većine neuronskih mreže zasnovana je na odabiru aktivacijske funkcije koja će se koristiti [20].

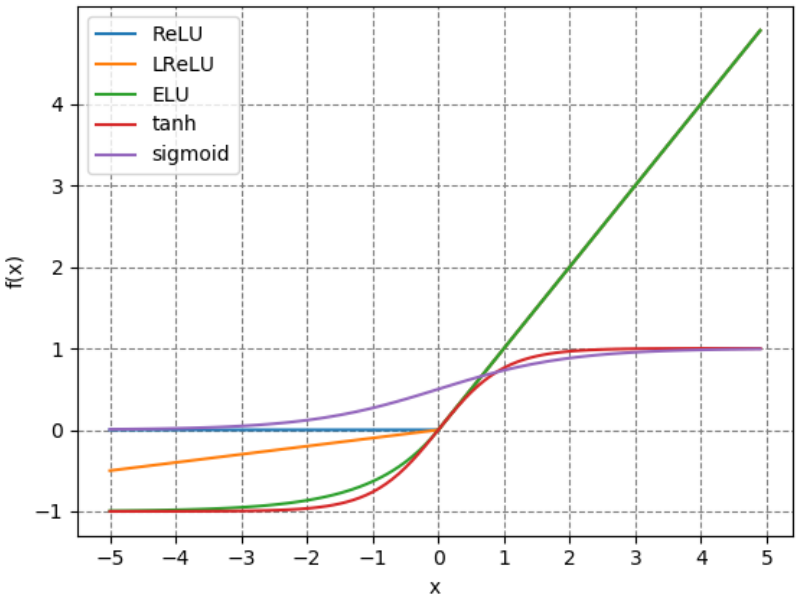
Trenutno najpopularnija aktivacijska funkcija za neuronske mreže je ispravljena linearna funkcija (ReLU), koja je isprva bila predložena za ograničene Boltzmannove strojeve, a zatim se uspješno počela koristiti za neuronske mreže [43]. ReLU aktivacijska funkcija jest $f(x) = \max(0, x)$. Osim što proizvode rijetke kodove, glavna prednost ReLU-ova je što olakšavaju problem nestajućih gradijenata [48]. Međutim, ReLU-ovi su nenegativni i stoga imaju srednju vrijednost aktivacije veću od nule.

Jedinice koje imaju ne-nul srednju vrijednost aktivacije služe kao pristranost za sljedeći sloj. Ako se takve jedinice ne poništavaju, učenje uzrokuje pomak pristranosti za jedinice u sljedećem sloju. Što su jedinice više povezane, to je veći njihov pomak pristranosti. Za neuronsku mrežu je poznato da centriranje aktivacije ubrzava učenje [24, 84]. "Normalizacija serije" (engl. *batch normalization*) također centrira aktivacije s ciljem da se suprotstavi unutarnjem kovarijacijskom pomaku (engl. *covariate shift*)[51]. Alternativa centriranju jest guranje srednje vrijednosti aktivacije prema nuli odgovarajućom aktivacijskom funkcijom. Stoga je tangens hiperbolni preferiran nad ostalim logističkim funkcijama [24]. Nedavno su se *leaky ReLUs* (LReLU-ovi)

koji zamjenjuju negativni dio ReLU-a linearnom funkcijom pokazali superiornijima od običnih ReLU-ova [69]. Formula LReLU-a jest $f(x) = \max(\alpha x, x)$ $(0 < \alpha < 1)$.

Za razliku od ReLU-ova, aktivacijske funkcije poput LReLU-ova ne osiguravaju stanje deaktivacije neosjetljivo na šum. Aktivacijska funkcija koja ima negativne vrijednosti kako bi omogućila srednje vrijednosti aktivacija blizu nule, ali koja zasićuje ka negativnoj vrijednosti s manjim argumentima jest ELU [22]. Eksponencijalna linearna funkcija (ELU) s $0 < \alpha$ je

$$f(x) = \begin{cases} x & \text{ako } x > 0 \\ \alpha(\exp(x) - 1) & \text{ako } x \leq 0 \end{cases}. \tag{3.6}$$



Slika 3.2: Ispravljena linearna funkcija (ReLU), *leaky ReLU* (LReLU, α = 0.1), eksponencijalna linearna funkcija (ELU, α = 1.0), tangens hiperbolni i sigmoida

ELU hiperparametar α kontrolira vrijednost na koju ELU zasićuje za negativne ulaze (vidi Sliku 3.2). ELU-ovi smanjuju efekt nestajućeg gradijenta kao i ispravljene linearne funkcije i LReLU-ovi. Problem nestajućeg gradijenta ublažava se jer je pozitivan dio tih funkcija jednak identitetu ulaza, stoga je njihova derivacija jedan i ne sužava se. Suprotno tome, tangens hiperbolni i sigmoidalne aktivacijske funkcije gotove se svugdje sužavaju i nestaju. Za razliku od ReLU-ova, ELU-ovi imaju negativne vrijednosti koje guraju srednju vrijednost aktivacija bliže nuli. One omogućuju brže učenje jer približavaju gradijent prirodnom gradijentu. ELU-ovi zasićuju prema ne-

gativnoj vrijednosti kada se argument smanjuje. Zasićenje znači malu derivaciju koja smanjuje varijaciju i informaciju koja se prenosi na sljedeći sloj. Stoga je reprezentacija i robusna i niske kompleksnosti [49]. Pozitivne značajke ovih interpretacija jest da se aktivacija može jasno razlikovati od deaktivacije i da samo aktivne jedinice nose važne informacije.

U eksperimentima ELU-ovi vode ne samo ka bržem učenju, nego i značajno boljim performansama generalizacije od ReLU-ova i LReLU-ova na mrežama s više od 5 slojeva [22].

3.1.3. Optimizatori

Gradijentni spust često se koristi za opetovanu prilagodbu težina sitnim koracima prema smjeru negativnog gradijenta

$$\mathbf{W}^{\tau+1} = \mathbf{W}^{\tau} - \eta \nabla \mathcal{J}(\mathbf{W}^{\tau}) \tag{3.7}$$

gdje τ označava korak iteracije, a varijabla $\eta > 0$ je stopa učenja. U svakom koraku, serijske metode koriste dio podskupa za treniranje za izračunavanje gradijenta $\mathcal{J}(\mathbf{W}^{\tau})$, a zatim ažuriraju težine. Stohastički gradijentni spust, *on-line* inačica gradijentnog spusta, prevladavajuća je metoda optimizacije za treniranje neuronskih mreža na velikim skupovima podataka [59]. U ovoj metodi ažuriranje težina temelji se na gradijentnoj vrijednosti cilja samo za jedan primjer. Ovo ažuriranje se ponavlja za niz malih skupova primjera (engl. *batch*) odabranih iz podskupa za treniranje. Stohastički gradijentni spust koristi učenje s unatražnim rasprostiranjem kao metodu računanja gradijenata [83, 58]. Učenje s unatražnim rasprostiranjem izračunava gradijente u dva prolaza, prema naprijed i prema natrag. Za višeslojnu mrežu, unaprijedni prolaz izračunava aktivaciju svih slojeva uzastopnim korištenjem jednadžbi (3.3) i (3.4). Prema pravilu lanca, unazadni prolaz izračunava ciljni gradijent s obzirom na parametre mreže. Kad je unaprijedni prolaz završen, unazadni prolaz proširuje gradijente kroz sve slojeve, počevši od gornjeg sloja (tj. izlaznog sloja) i kreće se unatrag dok ne dosegne dno (tj. ulazni sloj).

Ovaj jednostavan postupak često je iznenađujuće brz, što rezultira dobrim skupom težina i lako se skalira s brojem primjera treniranja u usporedbi sa složenijim metodama optimizacije [15]. Najočitiji nedostatak neuronskih mreža je da su vrlo sklone zaglaviti u lokalnim minimumima [14]. Pojavljuju mnogi ozbiljni problemi, kao što su prenaučenost modela i nestajući gradijenti [11], tijekom treninga s postupkom gradijentnog spusta. Stoga valja obratiti posebnu pozornost kako bi se osigurao brz postupak

konvergencije, ali i dobar skup parametara.

Adam je metoda za učinkovitu stohastičku optimizaciju koja zahtijeva samo gradijente prvog reda uz malu računalnu složenost [54]. Metoda izračunava pojedine prilagodljive razine učenja za različite parametre iz procjena prvog i drugog momenta gradijenata; ime Adam proizlazi iz procjene prilagodljivog momenta. Metoda je dizajnirana kao kombinacija prednosti dviju metoda: AdaGrad [33], koja dobro funkcionira s malim gradijentima, te RMSProp [105], koja dobro funkcionira na *on-line* i ne-statičkim postavkama. Neke od prednosti Adama jesu da su dimenzije ažuriranja parametara invarijantne za skaliranje gradijenta, njegovi koraci su ograničeni veličinom koraka hiperparametra, ne zahtijeva stacionarni cilj, radi s rijetkim gradijentima i prirodno izvodi korekciju koraka.

Metoda je jednostavna za implementaciju i zahtijeva mali računalni trošak. Eksperimenti potvrđuju analizu na brzini konvergencije u konveksnim problemima [54]. Sve u svemu, Adam je robustan i dobro prilagođen širokom rasponu ne-konveksnih optimizacijskih problema u području strojnog učenja.

3.1.4. Rijetki autoenkoder

Govor se proizvodi modulacijom relativno malog broja parametara dinamičkog sustava [30, 31], a to podrazumijeva da je njegova istinska temeljna struktura puno manjih dimenzija nego što je odmah vidljivo iz prozora koji sadrži stotine koeficijenata [45]. Dakle, vjeruje se da govorna emocionalna obilježja također imaju takvu temeljnu strukturu, ako postoji metoda koja može učinkovito iskoristiti informacije ugrađene u veliki skup podataka. Da bi se omogućilo učenje s prijenosom značajki, može se upotrijebiti osnovna struktura značajki naučenih iz ciljnih podataka kako bi rekonstruirali ostale izvorne podatke. Rijetki autoenkoder koristi se za iskorištavanje temeljne strukture značajki iz ciljnih podataka, predstavljenih skupom težinskih matrica i vektorom pomaka [26].

Rijetkost ima mnogo značajnih prednosti. Rijetka reprezentacija može olakšati otpetljavanje informacija u algoritmima dubokog učenja. Gusta reprezentacija je vrlo osjetljiva na svaku promjenu podataka. Nasuprot tome, rijetka reprezentacija je robusnija jer male promjene ulaza stvaraju gotovo zanemarive učinke na skup značajki različitih od nule. Druga prednost je da je rijetka reprezentacija sklonija dekodiranju linearnim modelom uz vrlo niske troškove računanja. Ipak, vrijedno je napomenuti da preveliko uvođenje rijetkosti u model može nepovoljno utjecati na učinak generalizacije jer ograničava kapacitet modela.

Uobičajeni način uvođenja rijetkosti u autoenkodere je dodavanje kazne u funkciju gubitka [64], tj. dodavanje regularizacije koja penalizira odstupanje očekivane aktivacije skrivenih slojeva od (niske) fiksne razine ρ. Optimizacijski problem je sljedeći:

$$\mathcal{J}^{SAE}(\theta) = \sum_{\mathbf{x}\in\mathbf{X}} ||\mathbf{x} - \mathbf{y}||^2 + \beta \sum_{j=1}^{m} \mathrm{SP}(\rho||\hat{\rho}_j) \tag{3.8}$$

gdje je $\mathrm{SP}(\rho||\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j}$ pojam regularizacijskog člana rijetkosti, $\hat{\rho}_j = \frac{1}{N} \sum_{i=1}^{N} h_j(\mathbf{x}_i)$ je prosječna aktivacija skrivenog sloja j (prosjek na podskupu za treniranje), ρ je razina rijetkosti, a β kontrolira težinu rijetke regularizacije. Ako je broj skrivenih jedinica m manji od broja ulaznih jedinica n, onda je mreža prisiljena naučiti komprimirani i rijetki prikaz ulaza.

Kazna $\mathrm{SP}(\rho||\hat{\rho}_j)$ je Kullback-Leiblerova divergencija između Bernoullijeve slučajne varijable sa srednjom vrijednošću ρ i Bernoullijeve slučajne varijable sa srednjom vrijednošću $\hat{\rho}_j$. KL divergencija je mjera razlike između dvije različite distribucije [57]. KL divergencija zadovoljava $\mathrm{KL}(p||q) \geq 0$ i ima ključno svojstvo da $\mathrm{KL}(p||q) = 0$ ako i samo ako $q = p$, a inače se monotono uzdiže kako q odstupa od p. Analogno KL divergenciji, funkcija $\mathrm{SP}(\rho||\hat{\rho}_j)$ doseže svoj minimum 0 za $\hat{\rho}_j = \rho$, i drastično raste kako se $\hat{\rho}_j$ približava 0 ili 1 [26]. Prema tome, očekuje se da je $\hat{\rho}_j$ približno jednak ρ u fazi učenja.

Za sve trening podatke, rijetki autoenkoder prvo se primjenjuje na primjerima $x_i^t \in \mathbb{R}^n$ kako bi naučili skup parametara W_1, W_2, b_1 i b_2. Za prijenos svakog od primjera $x_i^s \in \mathbb{R}^n$ iz izvorne do ciljne domene, značajke $\tilde{x}_i^s \in \mathbb{R}^n$ izračunavaju se na temelju naučenog skupa parametara rješavanjem jednadžbe:

$$\tilde{x}_i^s = \mathrm{SA}_{Latent}(x_i^s) \tag{3.9}$$

gdje je $\mathrm{SA}_{Latent}(x_i^s) = f(W_1 x + b_1)$ latentni prostor jednoslojnog autoenkodera. Jednadžba (3.9) prisiljava ulaz x_i^s da se rekonstruira kroz izračunavanje rijetke nelinearne kombinacije parametara naučenih na ciljnim podacima. Postupak rekonstrukcije smanjuje razliku između izvornih podataka i ciljnih podataka te dovodi do prijenosa značajki iz izvorne u ciljnu domenu.

Kao što se može vidjeti iz algoritma 1, na svakom koraku iteracije, uzorci specifični za labelu u ciljnom skupu koriste se za treniranje jednoslojnog autoenkodera označenog sa $\mathrm{SA}^l(W, b)$ koji obuhvaća opću strukturu mapiranja za ulazne uzorke. Za ulazni skup, uzorci s odgovarajućom klasom rekonstruirani su koristeći $\mathrm{SA}^l_{Latent}(T_s^{C_l})$, kako je opisano u jednadžbi (3.9), prema strukturi mapiranja koju je naučio trenirani autoenkoder $\mathrm{SA}^l(W, b)$. Zatim, kao i većina sustava za prepoznavanje govornih emocija, ove

rekonstruirane značajke koriste se kao ulaz standardnim nadziranim klasifikacijskim algoritmima $\mathcal{H}$ - strojevima potpornih vektora. Konačno, testni podskup koristi se za procjenu klasifikatora.

Algoritam 1: Učenje s prijenosom značajki rijetkim autoenkoderom

Ulaz: Dva označena skupova podataka T_t i T_s, i odgovarajuće oznake klase C_1, ..., C_L.

Izlaz: Naučeni klasifikator $\mathcal{H}$ za ciljni zadatak

Inicijaliziraj latentni skup $\tilde{T}_s = \emptyset$.

za *l=1 to L* **čini**

- Inicijaliziraj jednoslojni autoenkoder $\mathrm{SA}^l(W, b)$.
- Odaberi primjere $T_t^{C_l}$ specifične za labelu iz T_t.
- Treniraj $\mathrm{SA}^l(W, b)$ koristeći $T_t^{C_l}$.
- Odaberi primjere $T_s^{C_l}$ specifične za labelu iz T_s.
- Rekonstruiraj podatke i uzmi značajke iz skrivenog prostora $\tilde{T}_s^{C_l} = \mathrm{SA}^l_{Latent}(T_s^{C_l})$.
- Ažuriraj latentni skup podataka $\tilde{T}_s = \tilde{T}_s^{\cup}\tilde{T}_s^{C_l}$

Nauči klasifikator $\mathcal{H}$ primjenom nadziranog algoritma učenja (npr., strojevi potpornih vektora) na latentnim podacima $\tilde{T}_s$.

vrati *Naučeni klasifikator* $\mathcal{H}$

3.2. Klasifikatori

Niz čimbenika diktira izbor odgovarajuće klasifikacijske paradigme. Od ključne važnosti za ovo područje su: 1. tolerancija visoke dimenzionalnosti; i 2. sposobnost iskorištavanja malih skupova podataka [3]. Manje ključna, iako još uvijek važna, su i druga razmatranja poput sposobnosti rješavanja nelinearnih problema, prilagodbe učinkovitosti i računalnih troškova. Problem skupova značajki visokih dimenzija obično se bolje rješava izborom i uklanjanjem značajki prije nego što se izvodi stvarna klasifikacija. Popularni klasifikatori za prepoznavanje emocija, kao što su linearni diskriminativni klasifikatori (LDC) i klasifikatori k-najbližih susjeda (kNN), pokazali su se vrlo uspješnim i na stvarnom emocionalnom govoru [7, 60, 100]. Međutim, oni pate od sve većeg broja značajki koji vode do područja prostora značajki gdje su podaci vrlo rijetki ("prokletstvo dimenzionalnosti" [8]). Klasifikatori poput kNN-a koji dijele prostor značajki u ćelije, pogođeni su prokletstvom dimenzionalnosti i osjetljivi su na stršeće vrijednosti (engl. *outlier*). Prirodno proširenje LDC-a su strojevi potpornih vektora (engl. *support vector machines*, SVM). Ako ulazni podaci nisu prethodno

(implicitno) linearno transformirani, što može povećati ili smanjiti broj značajki, i ako linearni klasifikator poštuje kriterij prilagodbe maksimalne margine, dobivamo SVM. Iako SVM nije nužno najbolji klasifikator za svaku konstelaciju [71], oni daju dobra svojstva generalizacije [62, 21, 73], a danas se smatraju vrhunskim klasifikatorima.

SVM klasifikatori uglavnom se temelje na korištenju jezgrenih funkcija kako bi nelinearno mapirali izvorne značajke na visoki dimenzionalni prostor gdje se podaci mogu uspješno klasificirati pomoću linearnog klasifikatora. SVM klasifikatori se često koriste u mnogim primjenama za prepoznavanje uzoraka i nadmašuju druge poznate klasifikatore [61]. Oni imaju neke prednosti u odnosu na skrivene Markovljeve modele, uključujući globalnu optimalnost algoritma treniranja [16], te postojanje izvrsnih granica generalizacije ovisnih o podacima [23]. SVM klasifikatori se također opsežno koriste za problem prepoznavanja govornih emocija [86, 63, 78].

Formalno, uz primjere za treniranje i odgovarajuće binarne oznake $(\mathbf{x}_i, y_i)$, i = 1,...,N, gdje je $\mathbf{x}_i \in \mathbb{R}^n$ i $y_i \in \{-1, 1\}$, SVM rješava sljedeći problem minimizacije

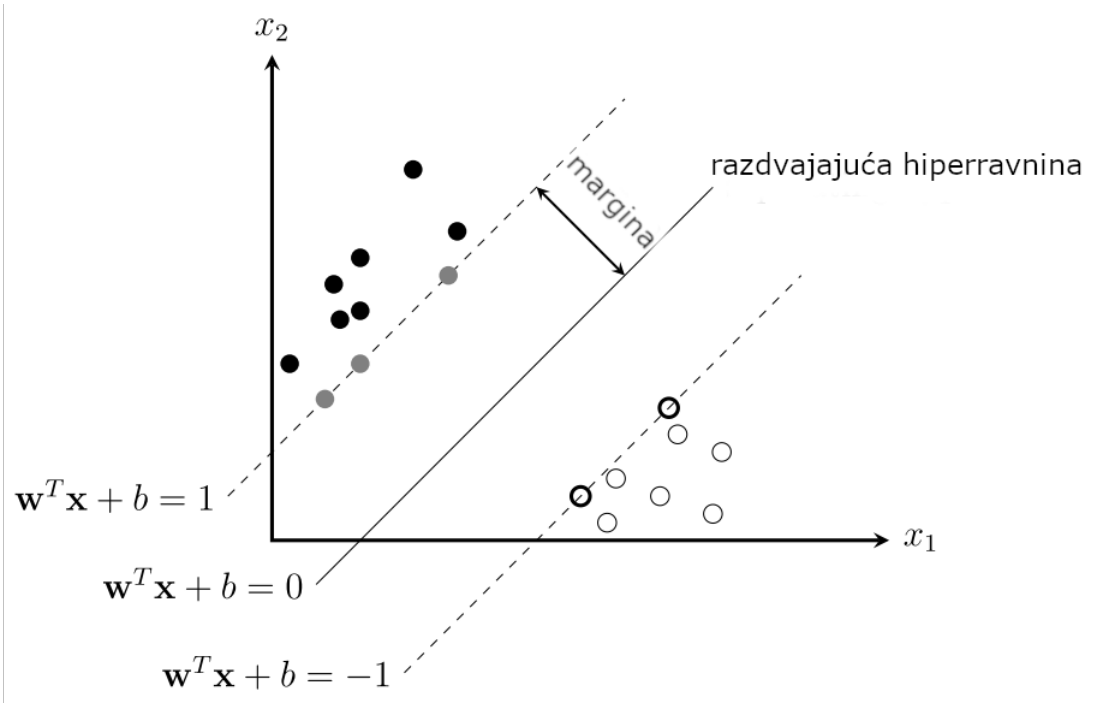
$$\underset{x}{\operatorname{argmin}} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i, \tag{3.10}$$

podložno $y_i(\mathbf{w}^T\phi(\mathbf{x}_i + b)) \geq 1 - \xi_i, \quad \xi_i \geq 0$. Ovdje regularizacijski parametar C kontrolira varijable ξ_i kako bi kažnjavao podatkovne točke koje krše zahtjeve margine, a $\phi(\cdot)$ označava funkciju mapiranja prostora značajki. U prostoru značajki definiranim s $\phi(\cdot)$, SVM-ovi traže hiperravninu za linearno razdvajanje maksimiziranjem margine. Izgrađena margina pivotira oko podskupa točaka podataka za treniranje, koji se nazivaju potpornim vektorima jer podržavaju hiperravnine na obje strane margine.

Slika 3.3 ilustrira hiperravninu, margine i potporne vektore za SVM. Važno je naglasiti važnost primjene funkcije mapiranja prostora značajki $\phi(\cdot)$ u formiranju SVM-a. Očigledna, ali presudna primjedba je da nelinearna klasifikacijska funkcija igra ključnu ulogu u optimalnom klasificiranju nelinearno razdvojivih podataka. Pri primjeni SVM-a za nelinearno razdvojive podatke, ulazni podaci linearno su mapirani na mnogo veći (ili čak beskonačno) dimenzionalni prostor u kojem su podaci linearno razdvojivi. Takva strategija, nazvana jezgrenim trikom, omogućuje strojevima potpornih vektora učinkovito klasificiranje podataka u vrlo visokodimenzionalnim prostorima. S obzirom na nelinearnu funkciju $\phi(\cdot)$, jezgrena funkcija na dva primjera $\mathbf{x}_i$ i $\mathbf{x}_j$ definirana je kao

$$k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \tag{3.11}$$

Najjednostavniji primjer jezgre je linearna jezgra s funkcijom identiteta, u tom



Slika 3.3: Razdvajajuća hiperravnina i margine za SVM. Uzorci na margini poznati su kao potporni vektori, a označeni su sivim točkama i podebljanim krugovima. Razdvajajuća hiperravnina se postiže maksimiziranjem margine [26]

slučaju $\phi(\mathbf{x}) = \mathbf{x}$ i $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$. Druga često korištena jezgra je Gaussova jezgra koju definira

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(\frac{-||\mathbf{x}_i - \mathbf{x}_j||}{2\sigma^2}) \tag{3.12}$$

gdje je σ širina.

3.2.1. Ocjena klasifikacije

Metode procjene klasifikacije pružaju način ocjenjivanja kvalitete različitih klasifikacijskih sustava. Kriteriji za procjenu obično se dobivaju uspoređivanjem diskretnih predviđenih klasa s istinitim ciljnim klasama [85]. Klasifikacijski zadatak matematički se može vidjeti kao mapiranje f od vektora $\mathbf{x}$ do skalara $y \in \{1, , ..., c\}$

$$f : \mathbf{X} \to \{1, ..., c\} \quad \mathbf{x} \mapsto y \tag{3.13}$$

Nakon treniranja, sustav klasifikacije se procjenjuje na drugom skupu primjera, koji je poznat kao testni podskup. S obzirom na testni podskup $\mathbf{X}^{te}$, svaki testni primjer označen je jednim ciljnom klasom $t \in \{1, , ..., c\}$, tako da testni podskup zadovoljava

sljedeće uvjete

$$\mathbf{X}^{te} = \bigcup_{i=1}^{c} \mathbf{X}_t^{te} = \bigcup_{i=1}^{c} \{x_{t,n} | n = 1, ..., N_t\} \tag{3.14}$$

gdje je N_t broj primjera u testnom podskupu koji pripadaju klasi t, što dovodi do veličine ispitnog skupa $|X^{te}| = \sum_{t=1}^{c} N_i$. Jedna uobičajena mjera se naziva osjetljivost (engl. *recall*), koja ocjenjuje učinak specifičan za labelu. Osjetljivost ovisi samo o primjerima klase t

$$\text{Recall}_t = \frac{|\{\mathbf{x} \in \mathbf{X}_t^{te} | y = t\}|}{N_t} \tag{3.15}$$

Obično je poželjno uzeti u obzir raspodjelu svih klasa pri procjeni opće učinkovitosti sustava klasifikacije. Neka $p_t = N_t/|X^{te}|$ označava prijašnju vjerojatnost labele t u testnom podskupu, Neusrednjena osjetljivost (engl. *weighted average recall*, WAR) je zadana kao

$$\text{WAR} = \sum_{t=1}^{c} p_t \text{Recall}_t \tag{3.16}$$

Nadalje, ako je raspodjela primjera među klasama izrazito neuravnotežena, moguće je radije zamijeniti prijašnje p_t za sve klase s konstantnom težinom $1/c$. To je poznato pod nazivom neusrednjena osjetljivost (engl., *unweighted average recall*, UAR)

$$\text{UAR} = \frac{\sum_{t=1}^{c} \text{Recall}_t}{c} \tag{3.17}$$

Važno je napomenuti da se UAR često koristi kao službena preporučena mjera za paralingvističke zadatke [88, 91, 94]. Zbog toga je UAR prihvaćen kao primarni mjerni podatak za procjenu uspješnosti prepoznavanja emocije u govoru.

4. Eksperimenti

4.1. Predobradba podataka

4.1.1. Normalizacija

Jedan od specifičnih izazova u otkrivanju emocionalnog stanja korisnika je sposobnost rješavanja varijabilnosti između govornika promatranih u ekspresivnom govoru [17]. Proizvodnja govora kod ljudi rezultat je kontroliranog anatomskog kretanja pluća, dušnika, grkljana, ždrijela, usne šupljine i nosne šupljine. Kao rezultat toga, svojstva govora su inherentno ovisna o govorniku. Iako postoje obrasci koji se očuvaju neovisno o govorniku (npr. povećanje osnovne frekvencije F_0 u ljutitim rečenicama), izraz emocija predstavlja specifične razlike. Prethodni radovi pokazali su da klasifikatori ovisni o govornicima daju veću učinkovitost od onih koji su neovisni o govornicima [87]. Stoga ne iznenađuje da normalizacija govora može poslužiti kao ključni korak u izgradnji snažnog sustava prepoznavanja emocija.

Izraz predobrada odnosi se na sve operacije, koje se moraju izvesti na vremenskom uzorku govornog signala prije ekstrakcije značajki. Prije faze ekstrakcije značajki pretprocesiraju se podaci pomoću različitih metoda. Neka $\mathcal{D} = \{x^{(j)}\}_{j=1,\ldots,n}$ bude skup za treniranje gdje je $x_{(j)} \in \mathbb{R}^d$. Za svaku značajku izračunava se njezina srednja vrijednost $\mu_k = \frac{1}{n}\sum_{j=1}^{n} x_k^{(j)}$ i varijanca σ_k. Zatim, svaka transformirana značajka $\tilde{x}_l^{(j)} = (x_k^{(j)} - \mu_k)/\sigma_k$ ima srednju vrijednost nula i varijancu jedan.

4.1.2. Rano zaustavljanje

Nenadzirano predtreniranje ima učinak koji, suprotno klasičnim regularizatorima, ne nestaje s povećanjem broja podataka [37]. Predtreniranje postavlja parametre u blizini rješenja te je niske prediktivne pogreške generalizacije.

Uz mali skup za treniranje, obično se ne pridaje velika važnost minimizaciji pogreške u treniranju, jer je prenaučenost glavni problem; pogreška treniranja nije dobar

način razlikovanja izvedbe generalizacije dvaju modela. Nenadzirano predtreniranje pomaže pronaći prividne lokalne minimume koji imaju nižu grešku generalizacije. S velikim podskupom za treniranje, empirijske i stvarne distribucije konvergiraju. U takvom scenariju, pronalaženje boljeg lokalnog minimuma bit će važno, a jače (bolje) strategije optimizacije trebaju imati značajan utjecaj na generalizaciju kada je skup za treniranje vrlo velik. Prethodni rezultati [46] pokazuju da se ne samo testna pogreška, već i pogreška pri treniranju znatno smanjuje uz nenadzirano predtreniranje pri treniranju dubokih autoenkodera (s uskim grlom) što je snažan pokazatelj učinka optimizacije.

Predtreniranje kao inicijalizacija se može smatrati ograničavanjem postupka optimizacije na relativno mali volumen parametarskog prostora koji odgovara lokalnom području nadzirane funkcije gubitka. Može se vidjeti da rano zaustavljanje ima sličan učinak, ograničavajući postupak optimizacije na područje parametarskog prostora koji je blizu početne konfiguracije parametara. Uz τ broj iteracija treninga i η stopa učenja (engl. *learning rate*) koji se koriste u postupku ažuriranja, $\tau\eta$ se može smatrati recipročnim parametrom normalizacije. U slučaju optimizacije jednostavnog linearnog modela korištenjem funkcije kvadratne pogreške i jednostavnog gradijentnog spusta, rano zaustavljanje imat će sličan učinak kao tradicionalna regularizacija. Dakle, i u predtreniranju i u ranom zaustavljanju, parametri nadzirane funkcije gubitka su ograničeni da budu blizu početnih vrijednosti. Nakon svake iteracije treninga može se izračunati indikator pogreške generalizacije (bilo od primjene nenadziranog kriterija učenja na validacijski skup ili čak i treniranjem linearnog klasifikatora na skup pseudo-validacije). Formalniji opis ranog zaustavljanja kao regularizatora se može naći u [102, 2].

4.1.3. Stopa učenja

Najvažniji hiperparametar za većinu algoritama je stopa učenja [9]. Preniska stopa učenja znači sporu konvergenciju ili konvergenciju ka lošim svojstvima s obzirom na konačno vrijeme računanja. Previsoka stopa učenja daje slabe rezultate jer se kriterij treniranja može povećati ili oscilirati. Optimalna stopa učenja za jedan skup podataka i arhitekturu može biti prevelika ili premalena pri mijenjanju jednog ili drugoga, pa je potrebno optimizirati učenje. Da bi se učinkovito tražila optimalna stopa učenja, potrebno je započeti s visokom stopom učenja i smanjivati ju sve dok treniranje ne počne divergirati. Najveća stopa učenja koja ne daje divergentno treniranje obično je vrlo dobar izbor za razinu učenja.

Iz tog razloga se opadanje polinoma (eng. *polynomial decay*) primjenjuje na stopi učenja. Ova funkcija monotono smanjuje stopu učenja u određenim koracima do zadane krajnje vrijednosti.

4.2. Baza podataka

Vremenski kontinuirana predviđanja prirodnih emocija (pobuđenost i valencija) na govoru i vizualnim podacima se istražuju korištenjem *REmote COLlaborative and Affective* (RECOLA) baze podataka [82]. U korpus su uključena četiri modaliteta: audio, video, elektrokardiogram (EKG) i elektrodermalna aktivnost (EDA). Ukupno je zabilježeno 9.5 sati multimodalnih snimki 46 sudionika s francuskog govornog područja, obavljajući zadatak u parovima tijekom video konferencije, te su one podijeljene u datoteke u trajanju od 5 minuta. Među sudionicima, 17 je francuskih parova, tri su njemačka i tri talijanska. Skup podataka je podijeljen u tri podskupa - treniranje (16 ispitanika), validacija (15 ispitanika) i test (15 ispitanika) - balansiranjem spola i dobi govornika. Konačno, 6 ocjenjivača (tri muška, tri ženska) labelirali su sve snimke. Svakih 40 milisekundi snimke označene su vrijednosti valencije i pobuđenosti. Te vrijednosti smještaju se unutar 4 kvadranta, odnosno 4 klase.

U okviru ovog rada, koristit će se isključivo audio snimke za treniranje. Svaka snimka je duga 5 minuta, a raspoloživo je 9 audio datoteka, što znači da je skup podataka trajanja 45 minuta. Kako bi se uzele u obzir varijacije u različitim razinama glasnoće između govornika, vremenske sekvence se pretprocesiraju da imaju srednju vrijednost nula i varijancu jedan, a nakon toga se valni oblik segmentira na sekvence duge 100 ms. Pri frekvenciji uzorkovanja od 22.05 kHz, to odgovara 2205-dimenzionalnom ulaznom vektoru i skupu od 27000 primjera za treniranje.

Nadalje, ako je veličina podatkovnog skupa za treniranje mala, treniranje može uzrokovati prenaučenost, što može rezultirati lošom generalizacijom, tj. mreža ima loša svojstva kada je izložena uzorku koji nikada nije vidjela. Međutim, trening s velikom količinom podataka je dugotrajan, i mreža se može sporo ažurirati nakon što se trenira na određen način. Zato je obično učinkovitije izračunati derivaciju na malom, slučajno odabranom skupu primjera koji se sastoji od podataka za treniranje, umjesto cijelog skupa, prije ažuriranja težina u odnosu na gradijent. Zbog toga se ulazni podskup od 27000 jedinica, dijeli u manje skupove od 100 primjera, koji se potom provode kroz neuronsku mrežu rijetkog autoenkodera.

4.3. Rezultati

Pripremljeni ulazni podaci provedeni su kroz rijetku autoenkodersku mrežu uz različit broj trening epoha te različite aktivacijske funkcije između ulaznog i reprezentacijskog sloja. Broj značajki u latentnom potprostoru postavljen je na 30, a koristi se Adam optimizator uz opadajuću vrijednost stope učenja.

Ranija istraživanja strategija ocjenjivanja korištenih u automatskom prepoznavanju emocija počela su prepoznavanjem emocija ovisnih o govorniku, baš kao i u prepoznavanju govora. Stoga se i danas najveći dio istraživanja oslanjanja na testiranja unakrsnom provjerom ovisna o govorniku. Međutim, samo *Leave-One-Subject-Out* (LOSO) [97, 103] ili *Leave-One-Subject-Group-Out* (LOSGO) [88] (unakrsna) provjera osigurava istinsku neovisnost govornika. Zbog spomenutog razloga, za klasifikaciju emocija iz latentnih značajki koristi se osnovni SVM klasifikator uz LOSO kros-validaciju i neusrednjena osjetljivost (UAR) kao ocjena klasifikacije.

Tablica 4.1: Ocjena klasifikacije emocija u govoru autoenkoderskim arhitekturama

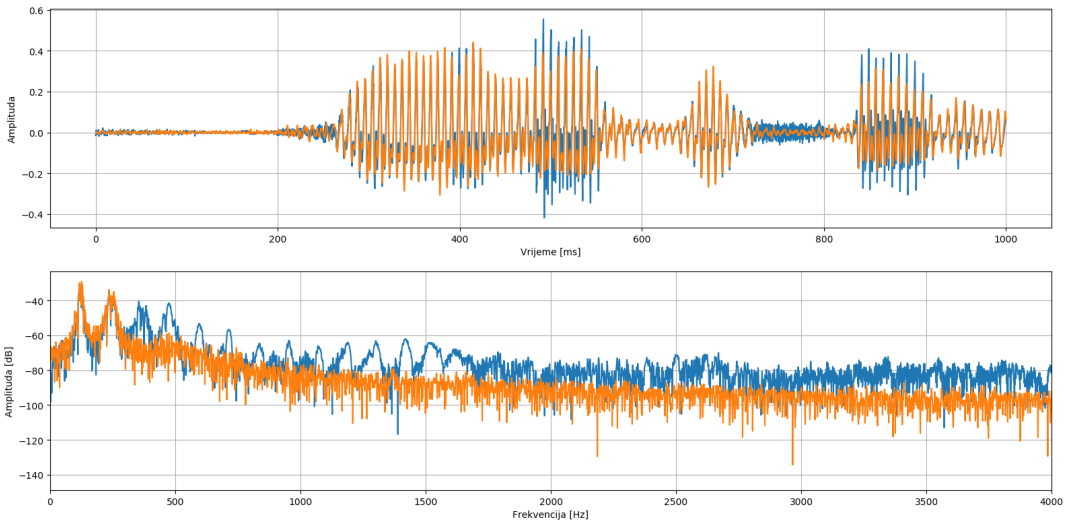
Autoenkoderska mreža		UAR(%)		
Aktivacijska funkcija	Broj epoha	najgori	prosječni	najbolji
ReLU	500	14.83	51.85	94.74
LReLU	500	14.88	51.81	93.32
ELU	200	14.71	51.68	92.75
	500	15.34	51.99	93.03
	1000	14.94	51.98	95.53
tanh	500	14.93	52.02	95.63
sigmoida	500	14.89	51.99	93.79

Iz tablice 4.1 se može iščitati kako su rezultati, odnosno neusrednjena osjetljivost klasifikacije emocija, vrlo neosjetljivi na promjenu aktivacijske funkcije. Budući da ReLU funkcija uvijek daje nulu za negativne ulaze i time slabi proces rekonstrukcije ulaznih značajki, očekivano je da nakon 500 epoha postiže najgore rezultate među odabranim funkcijama. LReLU ipak ostvaruje slabiju klasifikaciju, iako je zaista riječ o zanemarivim razlikama. Sigmoidalne funkcije se najčešće koriste u autoenkoderskim arhitekturama i vidljivo je zašto, postižu najbolje rezultate od ostalih odabranih aktivacijskih funkcija. Tangens hiperbolni (*tanh*) također je sigmoidalna funkcija, simetrična

i uz veće strmine od standardne logističke funkcije pa je poželjan izbor u autoenkoderskim arhitekturama jer stvara veće gradijente od obične sigmoide.

Više epoha za treniranje ne znači nužno i veću točnost klasifikacije, a takav slučaj je vidljiv i kod treniranja autoenkoderske mreže s ELU aktivacijskom funkcijom. Razlike među klasifikacijama s 500 i 1000 epoha nema, odnosno klasifikacija je čak i neznatno bolja uz manji broj epoha. Moguće je kako je model upao u lokalni minimum, ali i da se model približava prenaučenosti. Iz toga se može zaključiti kako bi optimalan broj trening epoha za RECOLA bazu podataka i prepoznavanje emocija u govoru zaista mogao biti manji od brojke 1000. Detaljni rezultati za autoenkoderske mreže podijeljeni po pojedinačnom paru govornika nalaze se u dodatku A.

Razlog polovičnoj uspješnosti klasifikacije emocija leži u veoma slaboj rekonstrukciji spektra originalnog govornog signala. Na slici 4.1 je prikazano kako rijetki autoenkoder s *tanh* aktivacijskom funkcijom precizno rekonstruira ulazni signal u vremenskoj domeni, no stvarni gubitak informacije se događa u frekvencijskoj domeni, gdje rekonstruirani signal, odnos izlaz iz autoenkodera, uspijeva tek djelomično opisati formantne frekvencije u govoru. Zbog toga izlazni govorni signal zvuči prigušeno, no prati promjene u dinamici razgovora, dok se čuju i pucketanja kao posljedica izravnog nadovezivanja uzoraka od 100 milisekundi gdje dolazi do mogućih gubitaka uzoraka i naglog faznog pomaka.



Slika 4.1: Vremenska i frekvencijska karakteristika jedne sekunde rekonstruiranog govornog signala iz autoenkoderske mreže s *tanh* aktivacijskom funkcijom

Kako bi ocijenili koliko je uspješnost klasifikacije rijetkim autoenkoderom zapravo korisna i upotrebljiva, rezultate je potrebno usporediti s jednom od modernih metoda za ekstrakciju značajki iz govora.

4.3.1. Usporedba s GeMAPS značajkama iz openSMILE alata

openSMILE verzija 2.0 je alat otvorenog koda [39], koji se nametnuo kao standard za ekstrakciju značajki u prepoznavanju govornih emocija. SMILE je kratica za govorno i multimedijsko tumačenje ekstrakcijom u visokoj dimenziji (engl. *Speech and Multimedia Interpretation by Large-space Extraction*). openSMILE je jednostavna aplikacija čije se modularne komponente za ekstrakciju značajki mogu slobodno konfigurirati i povezati putem konfiguracijske tekstualne datoteke. Služi kao službeni temelj za niz INTERSPEECH izazova u području računalne paralingvistike [96, 93] gdje je pokazano da daje vrhunske rezultate te se može primijeniti na vrlo širokom rasponu zadataka s jedinstvenom shemom ekstrakcije značajki.

Ženevski minimalistički akustički skup parametara (engl. *Geneva Minimalistic Acoustic Parameter Set*, GeMAPS) osnovni je akustični skup parametara za razna područja automatske analize glasa [40], kao što je paralingvistička ili klinička analiza govora. Ti su parametri odabrani na temelju: a) njihovog potencijala za indeksiranje afektivnih fizioloških promjena u proizvodnji glasa, b) dokazane vrijednosti u bivšim istraživanjima, i c) njihovom teoretskom značaju. GeMAPS implementacija je javno dostupna s openSMILE alatima. Komparativna usporedba predloženog seta značajki i velikih osnovnih skupova značajki INTERSPEECH izazova pokazuju visoku učinkovitost spomenutog skupa u odnosu na njegovu veličinu.

U tablici 4.2 je prikazana ocjena klasifikacije neusrednjenom osjetljivošću za 23 značajke iz skupa GeMAPS. Prosječni postotak uspješne klasifikacije veći je za 0.6% od najbolje klasifikacije autoenkoderskom arhitekturu, dok je za jednog govornika postotak točnosti klasifikacije čak 100%. Kako je u prethodnom slučaju rijetkog autoenkodera klasifikacija izvršena na skupu od 30 značajki, jasno je da bi se postigli još bolji rezultati na povećanom broju GeMAPS značajki. Detaljni rezultati za GeMAPS značajke podijeljeni po pojedinačnom paru govornika nalaze se u dodatku B.

Tablica 4.2: Ocjena klasifikacije emocija u govoru provedene na GeMAPS značajkama

	UAR(%)		
openSMILE skup značajki	najgori	prosječni	najbolji
GeMAPS	15.99	52.62	100

Zbog minimalno lošijih rezultata ekstrahiranih značajki iz autoenkoderske arhitekture naspram široko prihvaćenih značajki iz openSMILE biblioteke, potrebno je usporediti kolika je koreliranost tih značajki. Maksimalni koeficijent korelacije sukladnosti (engl. *concordance correlation coefficient*) [66] između pojedinačnih značajki iznosi 0.000035, a minimalni iznosi -0.000038, što sugerira da rijetki autoenkoder uči potpuno različite značajke od poznatih GeMAPS značajki.

Budući da se u ovom radu koristio rijetki autoenkoder sa samo jednim skrivenim slojem, ovakvi rezultati su obećavajući. Kombinacijom više različitih vrsta autoenkodera, poput varijacijskog [44], ili autoenkodera s uklanjanjem šuma [109], te dodavanjem više skrivenih slojeva uz različite aktivacijske funkcije, jasno je da su poboljšanja trenutnih rezultata moguća i da postoji budućnost u prepoznavanju emocija u govoru iz čistog valnog oblika govornog signala pomoću autoenkodera.

5. Zaključak

Predstavljena je metoda učenja s prijenosom značajki bazirana na rijetkom autoenkoderu. Osnovna ideja je korištenje jednoslojnog autoenkodera za rekonstrukciju izvornih podataka, a zatim primjena takve strukture za pronalaženje zajedničke strukture u podacima kako bi se dovršio korisni prijenos znanja iz izvornog u ciljni zadatak. Latentne značajke koriste se za izgradnju klasifikatora za prepoznavanje govornih emocija.

Eksperimentalni rezultati pokazuju da predloženi algoritam učinkovito prenosi znanje i postiže točnost klasifikacije približnu trenutno najsuvremenijim metodama ekstrakcije značajki za prepoznavanje emocija u govoru. Kao mogući napredak, mrežu je preporučljivo prenaučiti na puno većem skupu ulaznih podataka kako bi se ciljni podaci mnogo učinkovitije mapirali u značajke u latentnom prostoru. Baza podataka RECOLA je vremenskim trajanjem prekratka i morali su se uzeti okviri od 100 milisekundi kako bi se stvorilo dovoljno podataka za treniranje. A okviri od 100 milisekundi su premaleni za kvalitetno obuhvaćanje svih specifičnosti u glasu pojedinca.

Moguće poboljšanje uključuje i promjenu iz jednoslojne arhitekture u duboku arhitekturu kako bi se pronašle još korisnije informacije u emocionalnim značajkama, kao i korištenje različitih aktivacijskih funkcija u različitim skrivenim slojevima. Novonastali autoenkoderi bi time spojili karakteristike i prednosti tih funkcija. Također je moguće iskoristiti svojstva različitih vrsta autoenkodera kako bi se stvorila uspješnija arhitektura za ekstrakciju govornih značajki.

LITERATURA

[1] N. Akkarapatty, A. Muralidharan, N. Raj, i P Vinod. *Dimensionality Reduction Techniques for Text Mining*, stranice 49–72. IGI Global, 01 2017.

[2] S. Amari, N. Murata, K. R. Muller, M. Finke, i H. H. Yang. Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks*, 8(5):985–996, 1997.

[3] M. El Ayadi, M. S. Kamel, i F. Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3):572 – 587, 2011.

[4] P. Baldi. Autoencoders, unsupervised learning, and deep architectures. U I. Guyon, G. Dror, V. Lemaire, G. Taylor, i D. Silver, urednici, *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, svezak 27 od *Proceedings of Machine Learning Research*, stranice 37–49, Bellevue, Washington, USA, 2012. PMLR.

[5] P. Baldi i K. Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2(1):53 – 58, 1989.

[6] A. Batliner i R. Huber. *Speaker Characteristics and Emotion Classification*, stranice 138–151. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[7] A. Batliner, K. Fischer, R. Huber, i J. Spilker. Desperately seeking emotions or: Actors, wizards, and human beings. 2000.

[8] Richard E. Bellman. *Adaptive Control Processes: A Guided Tour*. MIT Press, 1961.

[9] Y. Bengio. Deep learning of representations for unsupervised and transfer learning. U I. Guyon, G. Dror, V. Lemaire, G. Taylor, i D. Silver, urednici, *Proce-*

edings of ICML Workshop on Unsupervised and Transfer Learning, svezak 27 od *Proceedings of Machine Learning Research*, stranice 17–36, Bellevue, Washington, USA, 2012. PMLR.

[10] Y. Bengio i Y. Lecun. *Scaling learning algorithms towards AI*. MIT Press, 2007.

[11] Y. Bengio, P. Simard, i P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.

[12] Y. Bengio, P. Lamblin, D. Popovici, i H. Larochelle. Greedy layer-wise training of deep networks. U B. Schölkopf, J. C. Platt, i T. Hoffman, urednici, *Advances in Neural Information Processing Systems 19*, stranice 153–160. MIT Press, 2007.

[13] Y. Bengio, A. Courville, i P. Vincent. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1798–1828, Kolovoz 2013.

[14] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.

[15] L. Bottou i O. Bousquet. The tradeoffs of large scale learning. U J. C. Platt, D. Koller, Y. Singer, i S. T. Roweis, urednici, *Advances in Neural Information Processing Systems 20*, stranice 161–168. Curran Associates, Inc., 2008.

[16] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[17] M. Busso, C. Bulut i S.S. Narayanan. Toward effective automatic recognition systems of emotion in speech. U J. Gratch i S. Marsella, urednici, *Social Emotions in Nature and Artifact*. Oxford University Press, 2013.

[18] J. Cahn. The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, 8:1–19, 1990.

[19] R. A. Calvo i S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37, 2010.

[20] David Charte, Francisco Charte, Salvador García, María J. del Jesus, i Francisco Herrera. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Information Fusion*, 44:78 – 96, 2018.

[21] Ze-Jing Chuang i Chung-Hsien Wu. Emotion recognition using acoustic features and textual content. U *2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763)*, svezak 1, stranice 53–56, 2004.

[22] Djork-Arné Clevert, Thomas Unterthiner, i Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *CoRR*, abs/1511.07289, 2015.

[23] N. Cristianini i J. Shawe-Taylor. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA, 2000.

[24] Yann Le Cun, Ido Kanter, i Sara A. Solla. Eigenvalues of covariance matrices: Application to neural-network learning. *Phys. Rev. Lett.*, 66:2396–2399, 1991.

[25] J. P. Cunningham i Z. Ghahramani. Linear dimensionality reduction: Survey, insights, and generalizations. *Journal of Machine Learning Research*, 16:2859–2900, 2015.

[26] J. Deng. *Feature Transfer Learning for Speech Emotion Recognition*. Doktorska disertacija, Technische Universität München, 09 2015.

[27] J. Deng, Z. Zhang, E. Marchi, i B. Schuller. Sparse autoencoder-based feature transfer learning for speech emotion recognition. U *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, stranice 511–516, 2013.

[28] J. Deng, R. Xia, Z. Zhang, Y. Liu, i B. Schuller. Introducing shared-hidden-layer autoencoders for transfer learning and their application in acoustic emotion recognition. U *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, stranice 4818–4822, 2014.

[29] J. Deng, Z. Zhang, i B. Schuller. Linked source and target domain subspace feature transfer learning – exemplified by speech emotion recognition. U *2014 22nd International Conference on Pattern Recognition*, stranice 761–766, 2014.

[30] L. Deng. *Computational Models for Speech Production*, poglavlje K. Ponting (ed.), stranice 199–213. Springer Verlag, 1999.

[31] L. Deng. Switching dynamic system models for speech articulation and acoustics. U M. Johnson, S. P. Khudanpur, M. Ostendorf, i R. Rosenfeld, urednici, *Mathematical Foundations of Speech and Language Processing*, stranice 115–133, New York, NY, 2004. Springer New York.

[32] L. Deng, A. Acero, G. Dahl, i D. Yu. Context-dependent pre-trained deep neural networks for large vocabulary speech recognition. svezak 20, stranice 30–42, 2012.

[33] J. Duchi, E. Hazan, i Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, 2011.

[34] R. O. Duda i P. E. Hart. *Pattern classification and scene analysis*. John Wiley, New York, 1973.

[35] David E. Rumelhart, Geoffrey E. Hinton, i Ronald J. Williams. Learning representations by back propagating errors. 323:533–536, 10 1986.

[36] Carl E. Williams i Kenneth N. Stevens. Emotions and speech: Some acoustical correlates. 52:1238–50, 11 1972.

[37] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, i Samy Bengio. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.*, 11:625–660, 2010.

[38] F. Eyben, A. Batliner, B. Schuller, D. Seppi, i S. Steidl. Cross-Corpus Classification of Realistic Emotions - Some Pilot Experiments. U LREC, urednik, *Proc. of theThird International Workshop on EMOTION (satellite of LREC): CORPORA FOR RESEARCH ON EMOTION AND AFFECT*, stranice 77–82, 2010.

[39] F. Eyben, F. Weninger, F. Gross, i B. Schuller. Recent developments in opensmile, the munich open-source multimedia feature extractor. U *Proceedings of the 21st ACM International Conference on Multimedia*, stranice 835–838, New York, NY, USA, 2013. ACM.

[40] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, i K. P. Truong. The geneva

minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2):190–202, 2016.

[41] R. Fernandez. *A Computational Model for the Automatic Recognition of Affect in Speech*. Doktorska disertacija, Cambridge, MA, USA, 2004.

[42] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, i M. Wilkes. Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Biomedical Engineering*, 47(7):829–837, 2000.

[43] X. Glorot, A. Bordes, i Y. Bengio. Deep sparse rectifier neural networks. U Geoffrey Gordon, David Dunson, i Miroslav Dudík, urednici, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, svezak 15 od *Proceedings of Machine Learning Research*, stranice 315–323, Fort Lauderdale, FL, USA, 2011. PMLR.

[44] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, i Y. Bengio. Generative adversarial nets. U Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, i K. Q. Weinberger, urednici, *Advances in Neural Information Processing Systems 27*, stranice 2672–2680. Curran Associates, Inc., 2014.

[45] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, i B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[46] G.E. Hinton i R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. 313:504–7, 08 2006.

[47] Geoffrey E. Hinton, Simon Osindero, i Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, Srpanj 2006.

[48] S. Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 6 (2):107–116, 1998.

[49] S. Hochreiter i J. Schmidhuber. Feature extraction through LOCOCODE. *Neural Computation*, 11(3):679–714, 1999.

[50] P. Indyk i R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. U *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, stranice 604–613, New York, NY, USA, 1998. ACM.

[51] S. Ioffe i C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[52] N. Japkowicz, S. Jose Hanson, i M. A. Gluck. Nonlinear autoassociation is not equivalent to pca. *Neural Comput.*, 12(3):531–545, Ožujak 2000.

[53] S. Khalid, T. Khalil, i S. Nasreen. A survey of feature selection and feature extraction techniques in machine learning. U *2014 Science and Information Conference*, stranice 372–378, 2014.

[54] Diederik P. Kingma i Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[55] Paul R. Kleinginna i Anne M. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5 (4):345–379, 1981.

[56] M. Kramer. Nonlinear principal component analysis using auto-associative neural networks. 37:233 – 243, 02 1991.

[57] S. Kullback i R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[58] Y. Lecun. A theoretical framework for back-propagation. U D. Touretzky, G. Hinton, i T. Sejnowski, urednici, *Proceedings of the 1988 Connectionist Models Summer School, CMU, Pittsburg, PA*, stranice 21–28. Morgan Kaufmann, 1988.

[59] Y. Lecun, L. Bottou, Y. Bengio, i P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[60] Chul Min Lee i S. S. Narayanan. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, 13(2):293–303, 2005.

[61] Chul Min Lee, S. S. Narayanan, i R. Pieraccini. Classifying emotions in human-machine spoken dialogs. U *Proceedings. IEEE International Conference on Multimedia and Expo*, svezak 1, stranice 737–740, 2002.

[62] Chul Min Lee, Shrikanth Narayanan, i Roberto Pieraccini. Combining acoustic and language information for emotion recognition. U *INTERSPEECH*, 2002.

[63] Chul Min Lee, Serdar Yildirim, Murtaza Bulut, Abe Kazemzadeh, Carlos Busso, Zhigang Deng, Sungbok Lee, i Shrikanth Narayanan. Emotion recognition based on phoneme classes. U *Proc. ICSLP'04*, stranice 889–892, 2004.

[64] Honglak Lee, Chaitanya Ekanadham, i Andrew Y. Ng. Sparse deep belief net model for visual area v2. U J. C. Platt, D. Koller, Y. Singer, i S. T. Roweis, urednici, *Advances in Neural Information Processing Systems 20*, stranice 873–880. Curran Associates, Inc., 2008.

[65] K. Leidelmeyer. *Emotions: an experimental approach*. Tilburg University Press, 1991.

[66] Lawrence I-Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45 1:255–68, 1989.

[67] Jackson J. Liscombe. *Prosody and Speaker State: Paralinguistics, Pragmatics, and Proficiency*. Doktorska disertacija, New York, NY, USA, 2007.

[68] H. Liu i H. Motoda. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[69] Andrew L. Maas. Rectifier nonlinearities improve neural network acoustic models. 2013.

[70] J. Masci, U. Meier, D. Cireşan, i J. Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. U T. Honkela, W. Duch, M. Girolami, i S. Kaski, urednici, *Artificial Neural Networks and Machine Learning – ICANN 2011*, stranice 52–59, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[71] D. Meyer, F. Leisch, i K. Hornik. Benchmarking support vector machines. 03 2003.

[72] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 izdanju, 1997.

[73] Donn Morrison, Ruili Wang, W. L. Xu, i Liyanage C. De Silva. Incremental learning for spoken affect classification and its application in call-centres. U *ALaRT*, 2005.

[74] J. Nicholson, K. Takahashi, i R. Nakatsu. Emotion recognition in speech using neural networks. U *Neural Information Processing, 1999. Proceedings. ICONIP '99. 6th International Conference on*, svezak 2, stranice 495–501, 1999.

[75] E. Oja. Data compression, feature extraction, and autoassociation in feedforward neural networks. U T. Kohonen, K. Mäkisara, O. Simula, i J. Kangas, urednici, *Artificial Neural Networks*, svezak 1, stranice 737–745. Elsevier Science Publishers B.V., North-Holland, 1991.

[76] S. J. Pan i Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[77] Rosalind W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, USA, 1997.

[78] O. Pierre-Yves. The production and recognition of emotions in speech: Features and algorithms. *Int. J. Hum.-Comput. Stud.*, 59(1-2):157–183, 2003.

[79] R. Plutchik. *A general psychoevolutionary theory of emotion*, stranice 3–33. Academic press, New York, 1980.

[80] Marc' Aurelio Ranzato, Y-Lan Boureau, i Yann LeCun. Sparse feature learning for deep belief networks. U *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS'07, stranice 1185–1192, USA, 2007. Curran Associates Inc.

[81] Marc'Aurelio Ranzato, Y-Lan Boureau, Sumit Chopra, i Yann LeCun. A unified energy-based framework for unsupervised learning. U Marina Meila i Xiaotong Shen, urednici, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, svezak 2 od *Proceedings of Machine Learning Research*, stranice 371–379, San Juan, Puerto Rico, 21–24 Mar 2007. PMLR.

[82] F. Ringeval, A. Sonderegger, J. Sauer, i D. Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. U *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, stranice 1–8, 2013.

[83] David E. Rumelhart, Geoffrey E. Hinton, i Ronald J. Williams. Neurocomputing: Foundations of research. poglavlje Learning Representations by Back-propagating Errors, stranice 696–699. MIT Press, Cambridge, MA, USA, 1988.

[84] Nicol N. Schraudolph. Centering neural network gradient factors. U *Neural Networks: Tricks of the Trade, volume 1524 of Lecture Notes in Computer Science*, stranice 207–226. Springer Verlag, 1997.

[85] B. Schuller i A. Batliner. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Wiley, 2013.

[86] B. Schuller, G. Rigoll, i M. Lang. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. U *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, svezak 1, stranice I–577–80, 2004.

[87] B. Schuller, R. Müller, M. Lang, G. Rigoll, i Technische Universität München. Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles. U *Proc. INTERSPEECH 2005, ISCA*, stranice 805–809, 2005.

[88] B. Schuller, S. Steidl, i A. Batliner. The interspeech 2009 emotion challenge. U *INTERSPEECH*, stranice 312–315. ISCA, 2009.

[89] B. Schuller, A. Batliner, S. Steidl, i D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Commun.*, 53(9-10):1062–1087, 2011.

[90] B. Schuller, Z. Zhang, F. Weninger, i G. Rigoll. Selecting training data for cross-corpus speech emotion recognition: Prototypicality vs . generalization. 2011.

[91] B. Schuller, M. Valstar, R. Cowie, i M. Pantic. Avec 2012: The continuous audio/visual emotion challenge - an introduction. U *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ICMI '12, stranice 361–362, New York, NY, USA, 2012. ACM.

[92] B. Schuller, Z. Zhang, F. Weninger, i F. Burkhardt. Synthesized speech for model training in cross-corpus recognition of human emotion. 15(3):313–323, 09 2012.

[93] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, H. Salamin, A. Polychroniou, F. Valente, i S. Kim. The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism, 2013.

[94] B. Schuller, S. Steidl, A. Batliner, J. Krajewski, J. Epps, F. Eyben, F. Ringeval, E. Marchi, i S. Schnieder. The interspeech 2014 computational paralinguistics challenge: Cognitive & physical load. U *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 09 2014.

[95] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, i G. Rigoll. Cross-corpus acoustic emotion recognition: Variances and strategies (extended abstract). U *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, stranice 470–476, 2015.

[96] B. W. Schuller. The computational paralinguistics challenge [social sciences]. *IEEE Signal Processing Magazine*, 29(4):97–101, 2012.

[97] Dino Seppi, Matteo Gerosa, Björn Schuller, Anton Batliner, i Stefan Steidl. Detecting Problems in Spoken Child-Computer-Interaction. U K. Berkling, D. Giuliani, i A. Potamianos, urednici, *Proceedings of the 1st Workshop on Child, Computer and Interaction*, 2008.

[98] V. Sethu, E. Ambikairajah, i J. Epps. Phonetic and speaker variations in automatic emotion classification. U *Proc. Interspeech 2008*, stranice 617–620, 01 2008.

[99] V. Sethu, J. Epps, i E. Ambikairajah. Speaker variability in speech based emotion models - analysis and normalisation. U *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, stranice 7522–7526, 2013.

[100] M. Shami i W. Verhelst. *Automatic Classification of Expressiveness in Speech: A Multi-corpus Study*, stranice 43–56. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[101] L. Shen, M. Wang, i R. Shen. Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. 12:176–189, 04 2009.

[102] J. Sjöberg i L. Ljung. Overtraining, regularization, and searching for minimum in neural networks. *IFAC Proceedings Volumes*, 25(14):73 – 78, 1992. 4th IFAC Symposium on Adaptive Systems in Control and Signal Processing 1992, Grenoble, France, 1-3 July.

[103] S. Steidl. *Automatic classification of emotion related user states in spontaneous children's speech.* Doktorska disertacija, University of Erlangen-Nuremberg, 2009.

[104] Charles W. Therrien. *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*. John Wiley & Sons, Inc., New York, NY, USA, 1989.

[105] T. Tieleman i G. Hinton. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.

[106] L. Torrey i J. Shavlik. Transfer learning. *Handbook of Research on Machine Learning Applications. IGI Global*, 3:17–35, 2009.

[107] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, i S. Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1301–1309, 2017.

[108] S. Usui, S. Nakauchi, i M. Nakano. Internal color representation acquired by a five-layer neural network. *Proc. Int. Conf. Artificial Neural Networks, 1991*, 1: 867–872, 1991.

[109] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, i Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. U *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, stranice 1096–1103, New York, NY, USA, 2008. ACM.

[110] F. Wang i J. Sun. Survey on distance metric learning and dimensionality reduction in data mining. *Data Mining and Knowledge Discovery*, 29(2):534–564, 2015.

[111] Rui Xia i Yang Liu. Using denoising autoencoder for emotion recognition. U *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*, stranice 2886–2889, 2013.

[112] Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, i Xuejie Zhang. Building chinese affective resources in valence-arousal dimensions. U *Proceedings of the 2016 Conference of the North Ameri-*

can Chapter of the Association for Computational Linguistics: Human Language Technologies, stranice 540–545. Association for Computational Linguistics, 2016.

[113] Z. Zhang, F. Weninger, M. Wöllmer, i B. Schuller. Unsupervised learning in cross-corpus acoustic emotion recognition. U *2011 IEEE Workshop on Automatic Speech Recognition Understanding*, stranice 523–528, 2011.

NENADZIRANO UČENJE ZNAČAJKI GOVORA KORIŠTENJEM NEURONSKIH MREŽA BAZIRANIH NA AUTOENKODERSKIM ARHITEKTURAMA

Sažetak

Nenadzirano učenje značajki i niždimenzionalnih reprezentacija podataka nalazi primjene u strojnom učenju, kompresiji s gubitkom i sl. Postojeće metode poput analize osnovnih komponenata (PCA) se često baziraju na linearnim transformacijama podataka, dok su modernije metode bazirane na neuronskim mrežama bolje opremljene za prepoznavanje nelinearnih odnosa u podacima. U okviru diplomskog rada implementiran je sustav za nenadzirano učenje značajki govora korištenjem rijetke autoenkoderske arhitekture neuronskih mreža. Dodatno, evaluirana je implementacija nad klasičnim problemom afektivnog računarstva (prepoznavanje emocija), uz korištenje učenja s prijenosom značajki (engl. *feature transfer learning*). Diskutirana je optimalna arhitektura mreže s obzirom na aktivacijsku funkciju i broj trening epoha i uspoređeni su rezultati s GeMAPS značajkama iz openSMILE alata.

Ključne riječi: govor, glas, učenje značajki, nenadzirano učenje, neuronske mreže, autoenkoderi

UNSUPERVISED SPEECH FEATURE LEARNING USING AUTOENCODER-BASED NEURAL NETWORK ARCHITECTURES

Abstract

Unsupervised feature learning and lower dimensional data representation finds application in machine learning, lossy compression, etc. Existing methods such as Principal Component Analysis (PCA) are often based on linear data transformations, while more modern methods based on neural networks are better equipped for detecting non-linear relationships in datasets. In this graduate thesis, a system for unsupervised learning of speech features was implemented, using a neural network based on a sparse autoencoder architecture. In addition, the implementation of the classical problem of emotion recognition in affective computing was evaluated, along with the use of feature transfer learning. The optimal network architecture was discussed with regards to the activation functions and number of training epochs and the results were compared with GeMAPS features from the openSMILE toolkit.

Keywords: speech, voice, feature learning, unsupervised learning, neural networks, autoencoders

Dodatak A
Ocjena klasifikacije po govornicima za autoenkoderske mreže

	UAR(%) za aktivacijsku funkciju i broj epoha				
Redni broj govornika	ELU-500	ReLU-500	LReLU-500	tanh-500	sigmoid-500
1.	29.95	29.52	29.56	29.81	29.93
2.	69.82	68.85	69.18	69.10	69.05
3.	93.04	94.74	93.32	95.63	93.79
4.	55.68	55.33	55.84	55.51	55.82
5.	27.92	27.69	27.89	27.88	27.64
6.	57.25	56.40	57.29	57.25	57.16
7.	70.97	70.94	70.31	70.66	70.85
8.	61.22	60.66	60.88	61.10	61.33
9.	72.40	72.48	72.28	72.96	72.82
10.	42.74	42.78	42.02	42.57	41.82
11.	62.47	62.12	62.78	62.33	62.61
12.	60.08	60.06	59.37	59.48	59.47
13.	15.34	14.83	14.88	14.93	14.89
14.	35.13	34.82	34.69	35.37	35.25
15.	51.07	50.11	50.67	50.53	50.58
16.	29.09	29.79	29.87	29.80	29.87
17.	70.61	70.91	70.75	70.78	71.32
18.	31.19	31.36	30.92	30.64	31.56

Dodatak B
Ocjena klasifikacije po govornicima za GeMAPS značajke

Redni broj govornika	UAR(%)
1.	29.99
2.	69.29
3.	100
4.	56.59
5.	26.90
6.	56.47
7.	72.70
8.	62.92
9.	70.85
10.	42.21
11.	58.40
12.	61.98
13.	15.99
14.	36.42
15.	51.10
16.	25.65
17.	76.33
18.	33.50