

Prepoznavanje pola govornika

Marko Vikić, BI 47/2017, vikic@uns.ac.rs

Ksenija Stanojević, BI 32/2017, ksenijastanojevic@uns.ac.rs

I. UVOD

Osnovna uloga govora je da ljudima obezbedi razmenu informacija. Ljudi putem govora izražavaju emocije, stavove, stupaju u kontakt sa drugim ljudima. Razvojem tehnologije, komunikacija se ne više ne vrši samo interakcijom između ljudi, već i ljudi i mašina.

Prepoznavanje govornika, kao i automatsko prepoznavanje govora, ima široku primenu: call centri, glasovno biranje, govorna pretraga, unos podataka. Pol govornika je korisna informacija u ovakvoj komunikaciji, jer nam daje više informacija o poruci koja se prenosi. Muškarci i žene imaju različite karakteristike glasa, različite frekvencije i boju.

Cilj rada je da obučimo model da prepozna kog je pola govornik iz baze snimljenih glasovnih poruka govornika različitog pola.

II. BAZA PODATAKA

Ova baza sadrži 55 govornika iz SEAC baze (Serbian Emotional Amateur Corpus). Svaki od govornika ima nekoliko desetina rečenica, predstavljenih akustičkim obeležjima, u obliku .csv fajla.

Obučavanje klasifikatora je izvršeno na bazi koja sadrži 3372 uzorka, od toga 1398 u klasi „Male“ i 1300 uzoraka u klasi „Female“.

Baza podataka je podeljena na dva dela, 80% podataka (44 govornika) čini trening podatke, a 20% (11 govornika) test podatke.

Podaci su podeljeni kako se sistem ne bi obučio na konkretan glas govornika. Prema zahtevima zadatka, izbačena su sledeća obeležja: „class“ - kako sistem ne bi imao informaciju o tome o kom se govorniku radi; „name“ – obeležje je jedinstveno za svaki uzorak i nije od značaja; „age“ – pod pretpostavkom da ne daje značajne informacije koje bi bile korisne za obuku modela.

Nakon izbacivanja ova dva obeležja iz baze podataka koja je prethodno imala 388 obeležja, model je obučen na 384 obeležja.

Uzorci koji su izdvojeni iz obeležja „class“ su ['p0006', 'p0005', 'p0001', 'p0002', 'p1007', 'p1008', 'p1009', 'p1060', 'p1010', 'p1011', 'p1012'] i oni predstavljaju test podatke. Baza test podataka sadrži 674 uzorka.

III. ANALIZA PODATAKA

U ovom radu, za obučavanje modela, korišćeni su klasifikatori: stabla odluke (DT), k-najbližih suseda (kNN) i mašina na bazi vektora nosača (SVM), a krosvalidacijom i redukcijom obeležja baze traženi su najpovoljniji parametri za izabrane algoritme i upoređene su njihove performanse.

Čest problem koji se susreće u mašinskom učenju je da za dat broj obeležja, tj. dimenzija, broj uzoraka potreban za popunjavanje prostora obeležja do zadovoljavajuće gustine je retko kada dostupan. Shodno tome, za dat broj trening uzoraka, postoji optimalan broj obeležja, ukoliko se taj broj prekorači dolazi do pada performansi modela. Iz ovog razloga, radi se redukcija dimenzionalnosti, koja dovodi i do bržeg izvršavanja programa.

Korišćene su dve metode smanjenja dimenzionalnosti: razlaganje na glavne komponente (PCA) i linearna diskriminantna analiza (LDA). Glavna razlika između ove dve metode je to što PCA koristi kriterijum reprezentacije, čiji je cilj da predstavi originalni skup uzoraka što tačnije, u prostoru sa manjim brojem dimenzija. LDA koristi kriterijum klasifikacije da bi sačuvao što više diskriminativnih informacija, u prostoru sa manjim brojem dimenzija.

Analizom baze utvrđeno je da nema nepotpunih vrednosti. Implementacija je odrađena u okviru biblioteke Scikit-learn u programskom jeziku Python.

1) Stabla odluke – DT

Stablo odluke spada u često korišćene metode nadgledanog učenja. Koren stabla (engl. *root node*) predstavlja čvor koji sadrži skup svih uzoraka, i od njega se stablo grana, odnosno, vrši se sukcesivna particija skupa uzoraka na dva disjunktna podskupa. Podela se vrši prema odgovoru na postavljeno pitanje (da ili ne), odnosno prema vrednosti datog obeležja za svaki pojedinačni uzorak, skup se deli na dva podskupa i formiraju se dva nova čvora.

2) Mašina na bazi vektora nosača – SVM

Model vektora nosača (SVM) je metod namenjen rešavanju problema binarne klasifikacije, čiji je cilj da identifikuje hiperpovrš koja treba da razdvoji uzorke u prostoru obeležja.

SVM se zasniva na klasifikatoru maksimalne margine koji optimalno deli prostor na delove koji sadrže uzorke iz jedne od klasa. Vektori nosači su oni uzorci koji se nalaze na ivicama margine.

3) *kNN*

KNN algoritam (algoritam K najbližih suseda) spada u grupu metoda kasnog učenja (engl: lazy learning) koja podrazumeva odlaganje obrade uzoraka za obuku do trenutka kada se javi zahtev za klasifikacijom nepoznatog uzorka. U pitanju je intuitivan algoritam koji vrši klasifikaciju nepoznatog uzorka na osnovu klasne pripadnosti bliskih uzoraka iz skupa za obuku.

4) *PCA metoda redukcije dimenzionalnosti*

Cilj PCA metode je predstavljanje uzoraka iz visokodimenzionalnog prostora u prostoru sa manjim brojem dimenzija što vernije, odnosno, zadržavajući u što većoj meri varijansu sadržanu u podacima. PCA predstavlja metodu nenadgledanog učenja, jer ne koristi informaciju o vrednosti izlaza za uzorke iz skupa za obuku. Polazeći od pretpostavke da su raspoloživa obeležja korelisana i da postoji redundantnost u višedimenzionalnoj reprezentaciji podataka, cilj PCA metode je određivanje novih nekorelisanih obeležja (PCA komponente) linearnim kombinacijama postojećih obeležja uz očuvanje varijanse u podacima. Izborom najinformativnijih PCA komponenata obrazuje se novi prostor sa manjim brojem dimenzija i uzorci se projektuju na njega. Najinformativnije PCA komponente biće one koje odgovaraju pravcima najvećeg rasipanja uzoraka.

5) *LDA metoda redukcije dimenzionalnosti*

Cilj linearne diskriminantne analize jeste da se dimenzionalnost prostora smanji, a da se pritom očuva što više diskriminatornih informacija. Formiranjem linearne kombinacije postojećih obeležja LDA pronalazi pravac w takav da se prilikom projekcije uzoraka na njega maksimizuje separabilnost između klasa. Uzimanje u obzir klasnih labela svrstava ovu metodu među metode nadgledanog učenja. Pобољшanje razdvajivosti između klasa postiže se maksimizacijom rastojanja između srednjih vrednosti LDA projekcija uzoraka iz različitih klasa i istovremenom minimizacijom varijanse projekcije uzoraka unutar svake klase pojedinačno.

Ustanovljeno je da je najveći broj komponenti (*n_components*) jednak 1, jer se za broj komponenti LDA redukcije uzima manji od dve vrednosti, a to su klasa – 1 i broj obeležja. Preuzeto sa veb stranice scikit-learn biblioteke.

IV. REZULTATI

Unakrsnom validacijom sa 10 podskupova obučen je klasifikator, na trening podacima. Kada je u pitanju SVM klasifikator iterativnom metodom na trening skupu dobijena je najveća tačnost (u iteraciji broj 6), te iz te iteracije uzeti su najbolji parametri za ovaj klasifikator.

Najbolji rezultati postignuti su koristeći linear kernel, regularizacioni parametar C postavljen na 10 i primenom metode OVR (jedan protiv ostalih). Validacijom na test skupu dobija se najbolja tačnost u kombinaciji sa LDA redukcijom dimenzionalnosti čiji je broj dimenzija postavljen na 1, što predstavlja teoretski maksimum koji dozvoljava LDA, o čemu je rečeno u III poglavlju ovog rada.

<i>Tačnost u %</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>
DT	93.58	89.84
SVM	97.55	97.10
kNN	98.92	95.44

Tabela 1. Upoređene vrednosti tačnosti obuke modela na trening uzorcima sa i bez redukcije

Kada je u pitanju kNN klasifikator iterativnom metodom na trening skupu, dobijena je najveća tačnost (u iteraciji broj 18), a time i najbolje vrednosti parametara za ovaj model. Najbolji rezultati postignuti su koristeći Euklidsku metriku za računanje udaljenosti između uzoraka, i parametrom k postavljenim na 15 suseda.

Validacijom na test skupu dobija se najbolja tačnost u kombinaciji sa LDA redukcijom dimenzionalnosti čiji je broj dimenzija postavljen na 1. Može se primetiti da su nešto manje tačnosti i međusobno slične dobijene bez redukcije i sa PCA redukcijom dimenzionalnosti. Za PCA iterativnom metodom dobijen je optimalan broj parametara koji iznosi 125.

<i>Tačnost u %</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>	<i>LDA redukcija</i>
DT	77.74	77.59	92.73
SVM	85.75	83.82	93.47
kNN	90.20	90.94	93.62

Tabela 2. Upoređene vrednosti tačnosti obuke modela na test uzorcima sa i bez redukcije

Za klasifikator na bazi stabala odluke iterativnom metodom na trening dobijena je najveća tačnost (u iteraciji broj 19), te iz te iteracije uzeti su najbolji parametri za ovaj klasifikator.

Validacijom na test skupu ustanovljeno je da ovaj klasifikator u poređenju sa PCA redukcijom i slučajem bez redukcije, ima dosta bolju tačnost u kombinaciji sa LDA redukcijom dimenzionalnosti, i to za maksimalnu dubinu od 35 i uz upotrebu entropije.

<i>Matrica konfuzije</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>
DT	1306. 92. 81. 1219.	1259. 139. 135. 1165.
SVM	1361. 37. 29. 1271.	1356. 42. 36. 1264.
kNN	1382. 16. 13. 1287.	1333. 65. 58. 1242.

Tabela 3. Matrice konfuzije modela na trening uzorcima sa i bez redukcije

<i>Matrica konfuzije</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>	<i>LDA redukcija</i>
DT	176 8 142 348	172 12 139 351	167 17 32 458
SVM	184 0 96 394	184 0 109 381	172 12 32 458
kNN	180 4 62 428	184 0 61 429	170. 14. 29. 461.

Tabela 4. Matrice konfuzije modela na test uzorcima sa i bez redukcije

<i>DT %</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>	<i>LDA redukcija</i>
Preciznost	97.75	96.70	96.42
Tačnost	77.74	77.60	92.72
Osetljivost	71.02	71.63	93.47
Specifičnost	95.65	93.48	90.76
F-mera	82.26	82.30	94.92

Tabela 5. Mere uspešnosti klasifikatora na test podacima

<i>SVM %</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>	<i>LDA redukcija</i>
Preciznost	100.00	100.00	97.44
Tačnost	85.76	83.83	93.47
Osetljivost	80.40	77.75	93.47
Specifičnost	100.00	100.00	93.48
F-mera	89.14	87.49	95.42

Tabela 6. Mere uspešnosti klasifikatora na test podacima

<i>kNN %</i>	<i>Bez redukcije</i>	<i>PCA redukcija</i>	<i>LDA redukcija</i>
Preciznost	99.07	100.00	97.05
Tačnost	90.20	90.95	93.62
Osetljivost	87.35	87.55	94.08
Specifičnost	97.83	100.00	92.40
F-mera	92.84	93.37	95.54

Tabela 7. Mere uspešnosti klasifikatora na test podacima

V. ZAKLJUČAK

Nakon upoređivanja različitih pristupa i različitih algoritama mašinskog učenja pri zadatku klasifikacije govornika po polu upotrebom akustičkih obeležja, zaključeno je da kNN klasifikator u kombinaciji sa LDA redukcijom dimenzionalnosti daje najbolje rezultate. Problem je kNN klasifikatorom analiziran koristeći Euklidsku metriku za računanje udaljenosti između uzoraka, i parametrom k postavljenim na 15 suseda, pri tome da je broj dimenzija postavljen na 1, što predstavlja teoretski maksimum koji dozvoljava LDA.

VI. LITERATURA

- [1] Tijana Nosek, Branko Brkljač, Danica Despotović, Milan Sečujski, Tatjana Lončar-Turukalo – „Praktikum iz mašinskog učenja”.
- [2] T. Đurkić, A. Lojanić, S. Suzić, B. Popović, M. Sečujski, T. Nosek - Emotion recognition from speech based on ML algorithms applied on two Serbian datasets
- [3] https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html?fbclid=IwAR0nCgq1-D3X2ymEWEGbFSIqULLT5709UXv7dXjT-a0Wy-WUQCYdrBQqkIs

