Poređenje sistema za prepoznavanje govora na srpskom jeziku baziranih na punim i dijagonalnim kovarijansnim matricama

Marko B. Janev, Nikša Jakovljević, Darko Pekar

***Sadržaj* — U ovom radu su upoređene perfromanse dva sistema za prepoznavanje govora baziranih na skrivenim Markovljevim modelima i Gausovim raspodelama od kojih jedan koristi punu a drugi dijagonalnu kovarijansnu matricu pri prepoznavanju. Za obuku sistema je korišćen metod baziran na Gausovom kalasifikatoru.**

***Ključne reči* — Gausov klasifikator, skriveni Markovljevi modeli, puna kovarijansna matrica, regularizacija**

I. UVOD

SVRHA ovog rada je da se opiše primena metode Gausove klasifikacije (GC) sa punim kovarijansnim matricama na obuku sistema za automatsko prepoznavanje govora zasnovanog na skrivenim Markovljevim modelima (HMM) i Gausovim raspodelama kao emitujućim raspodelama i da se prikaže uočena prednost pomenutog algoritma u odnosu na varijantu GC algoritma sa dijagonalnim kovarijansnim matricama koja se ogleda u značajnom smanjenju greške prepoznavanja, izražene preko greške prepoznavanja na nivou reči (GNR).

Karakteristično za problematiku prepoznavanja govora je izražena korelacija između obeležja, a klasteri koje čine opservacije iz skupa za obuku koje težimo da opišemo Gausovim raspodelama, različito su orijentisani. Potpuno zanemarivanje korelisanosti između obeležja i korišćenje dijagonalnih kovarijansnih matrica u velikoj meri degradira performanse sistema za prepoznavanje govora. U radu je predstavljeno rešenje koje datu problematiku rešava korišćenjem GC algoritma klasterizacije sa punim kovarijansnim matricama za obuku modela, tj. procenu parametara emitujućih raspodela HMM modela.

U poglavlju II dat je kratak pregled osnovnih definicija koje opisuju HMM modele. Poglavlje III daje kratak opis Gausovog klasifikatora, dok poglavlje IV opisuje rešenje za procenu parametara emitujućih raspodela HMM modela realizovano u ovom radu, koje koristi GC klasifikator sa punim kovarijansnim matricama. U poglavlju V dati su rezultati izvršenih eksperimenata.

Rad je realizovan u okviru projekta "Razvoj govornih tehnologija za srpski jezik i primena u 'Telekomu Srbija'" (TR-6144A) Ministarstva za nauku države Srbije.

M. B. Janev, Alfanum Speech Tehnologies ltd., Trg D. Obradovića 6, 21000 Novi Sad, Serbia. e-mail: marko.janev @alfanum.co.yu

N. Jakovljević, FTN. Trg D. Obradovića 6, 21000 Novi Sad, Serbia.

D. Pekar, Alfanum Speech Tehnologies ltd., Trg D. Obradovića 6, 21000 Novi Sad, Serbia, e-mail: darko.pekar@alfanum.co.yu

II. SKRIVENI MARKOVLJEVI MODELI SA PONDERISANIM GAUSOVIM PASPODELAMA

Kao osnova sistema za prepoznavanje koristi se skriveni Markovljev model (HMM) sa konačnim brojem stanja $\{S_1,..., S_M\}$. Model se može opisati uređenom četvorkom $\lambda = (A, b, \pi, M)$ [1], gde je M broj stanja modela, $A = [a_{ij}]$ matrica verovatnoća prelaza između stanja Markovevljevog modela, π vektor inicijalnih verovatnoća stanja, $b = [b_1...b_M]$ vektor uslovnih gustina raspodela $b_j(o)$ opservacija po stanjima $S_j, j \in \{1,...,M\}$ (emitujuće gustine raspodela stanja). Matrica verovatnoće prelaza i vektor inicijalnih verovatnoća stanja zadovoljavaju ograničenja:

$$\sum_{j=1}^{M} a_{ij} = 1, \forall i \in \{1,..., M\}, \sum_{j=1}^{M} \pi_j = 1 \qquad (1)$$

Korišćeni su HMM modeli sa ponderisanim smešama, tako da se emitujuća gustina raspodele za stanje j može predstaviti preko izraza:

$$b_j(o) = \sum_{k=1}^{G^{(j)}} c_k^{(j)} p(o \mid \theta_k^{(j)}) \qquad (2)$$

gde je za $p(o \mid \theta_k^{(j)})$ korišćena normalna raspodela $N(\mu, \Sigma)$. Težinski koeficijenti zadovoljavaju ograničenja:

$$\sum_{k=1}^{G^{(j)}} c_k^{(j)} = 1, \forall j \in \{1,..., M\} \qquad (3)$$

III. GAUSOV KLASIFIKATOR

Problem klasifikacije svodi se na sledeće: Ako se posmatra C klasa $\{\omega_1,..., \omega_C\}$ sa poznatim apriori verovatnoćama $p(\omega_j), j \in \{1,..., C\}$, cilj je da se raspoložive opservacije svrstaju (klasifikuju) u klase, uz minimizaciju verovatnoće greške klasifikacije. Ako nemamo druge informacije o opservacijama osim pretpostavljenih raspodela, tada je pravilo odluke kojim opservaciju $x \in R^p$ dodeljujemo nekoj klasi (zasnovano na verovatnoćama), da x dodelimo onoj klasi ω_j za koju je maksimalna aposteriori verovatnoća $p(\omega_j \mid x)$ pripadanja datoj klasi.

Primenom Bajesove teoreme, može se formulisati pravilo po kom će opservacija x biti dodeljena onoj klasi

ω_j, za koju je $p(\omega_j)p(x|\omega_j)$ najveće, gde je $p(x|\omega_j)$ uslovna gustina raspodele opservacija po klasama.

Ako se pretpostave normalne raspodele za $p(x|\omega_j)$, klasifikator se naziva Gausov. Korišćenjem prirodnog logaritma $\ln(.)$, gornje pravilo se svodi na: $x \in \omega_i$ akko je: $g_i(x) > g_j(x), j \neq i$, gde je:

$$g_i(x) = \ln(p(\omega_i)) - \frac{1}{2}\ln(|\Sigma_i|) - \frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) \quad (4)$$

Parametri pretpostavljenih raspodela su: μ_i, Σ_i dok je apriori verovatnoće klasa moguće proceniti kao $p(\omega_i) = n_i/n$, gde je n_i broj pripadajućih opservacija klasi ω_i, a n ukupan broj opservacija.

Prethodno opisano pravilo dodeljivanja opservacija klasterima se u GC algoritmu iterativno primenjuje, dok broj prelazaka opservacija između klasa u dva uzastopna koraka ne padne ispod zadatog praga.

Procene očekivanja i kovarijansne matrice po kriterijumu maksimalne verodostojnosti (ML) za slučaj jedne Gausove smeše su:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^T \quad (5)$$

Iako to nije ML procena očekivanja i kovarijansnih matrica za slučaj sa više od jedne klase, iste procene (gde su opservacije uzete po pripadajućim klasama) se mogu uzeti [2] i za procenu očekivanja i kovarijansnih matrica $(\hat{\mu}_j, \hat{\Sigma}_j)$ za raspodele $p(x|\omega_j)$, što je korišćeno u ovom radu. Ovi procenjeni parametri na izlazu iz GC algoritma, korišćeni su kao procene parametara smeša korišćenih HMM modela.

IV. Poređenje Gausovog klasifikatora sa punim i dijagonalnim kovarijansnim matricama

U problemima gde se tretiraju podaci za koje se pretpostavlja da su generisani od strane različitih Gausovih smeša, vrlo često ne postoji dovoljan broj podataka za dobru procenu kovarijansne matrice [2]. Zato može da se javi problem da su procene nekih kovarijansnih matrica singularne (ili blizu singularne). To automatski znači, ne samo da će doći do numeričke nestabilnosti prilikom pokušaja inverzije, već i da je procena loša, što ima negativan uticaj na prepoznavanje.

Postoji nekoliko alternativa za rešavanje tog problema. Jedna (najgrublja) je da se izvrši aproksimacija kovarijansnih matrica dijagonalnim (i eventualno izvrši dijagonalno loudovanje). Time se osigurava njihova regularnost (kao i jači ali potreban uslov dobre pozicioniranosti). Iako se time značajno smanjuje računska složenost pri obuci i prepoznavanju, vrši se potpuno zanemarivanje korelisanosti izmedju obeležja. Ovo predstavlja varijantu GC klasifikatora sa dijagonalnim kovarijansnim matricama.

Drugi način je da se koristi PCA analiza i da se kovarijansne matrice projektuju na prostor u kojima su projekcije regularne i dobro pozicionirane. To podrazumeva korišćenje GC klasifikatora u prostoru redukovanih dimenzija (što bi bilo poželjno), ali zbog problematike izraženo različite orjentacije klastera u prostoru obeležja nekog modela, kao i zbog toga što čak i ako to nije izraženo na nivou jednog modela, u problematici prepoznavanja govora, kod različitih modela male karakteristične vrednosti koje izazivaju singularnost kod jednog modela, mogu da se pojave po sasvim različitim obeležjima kod drugog modela, metoda nije primerena pomenutoj problematici.

Ako bi se pretpostavile jednake kovarijansne matrice po klasterima to bi regularizovalo problem [2], ali bi se kvadratna diskriminativna funkcija GC klasifikatora degenerisala u linearnu i to bi predstavljalo dosta grubu aproksimaciju. Zato se u radu pribeglo trećoj alternativi, da za svaki klaster postoji prag za minimalan broj pripadajućih opservacija, koji bi obezbeđivao dovoljan broj opservacija za korektnu procenu kovarijansnih matrica. To je impliciralo znatno manji broj pretpostavljenih smeša po modelu nego u slučaju dijagonalnog GC klasifikatora koji u sebi već uključuje regularizaciju. Ako se ipak javi blizu singularna matrica (što se ispostavilo kao neredak slučaj), vrši se regularizacija date matrice ograničavanjem sa donje strane suviše malih karakterističnih vrednosti, tako da kondicioni broj rezultujuće matrice (količnik između najveće i najmanje karakteristične vrednosti) $\chi(\Sigma) = \lambda_{max}/\lambda_{min}$ bude manji od nekog zadatog praga. Opisanim postupkom se takođe postigla regularnost i dobra pozicioniranost u ovom slučaju punih kovarijansnih matrica uz bolju procenu parametara raspodela nego u slučaju GC klasifikatora sa dijagonalnim kovarijansnim matricama.

U eksperimentima je potvrđena prethodno navedena teza da GC klasifikatoru sa punim kovarijansnim matricama pogoduje znatno manji broj smeša po modelu koji se obučava u odnosu na GC klasifikator sa dijagonalnim kovarijansnim matricama. Prosečan broj smeša po stanju koji daje najbolji razultat za GC sa punim kovarijansnim matricama je 2,39 što je manje od onog za dijagonalnu varijantu koji iznosi 6,12. To ima i još jednu interpretaciju: Pošto GC klasifikator sa punim kovarijansnim matricama ima mogućnost rotiranja svake smeše tj. klastera posebno u prostoru obeležja, jer tretira korelisanosti obeležja nad opservacijama u pripadajućem klasteru (znači na nivou jednog klastera), on samim tim ima i mogućnost boljeg pokrivanja prostora obeležja, što dijagonalni GC klasifikator nije u stanju, jer u potpunosti zanemaruje korelisanosti. Zato veći broj smeša koji je poželjan kod dijagonalnog GC klasifikatora, šteti proceni parametara (i kasnijem prepoznavanju sa tako procenjenim parametrima) kod GC klasifikatora sa punim kovarijansnim matricama.

Pošto se broj potrebnih opservacija za dobru procenu pune kovarijansne matrice posmatra u odosu na dimenziju prostora obeležja [2], u sistemu je izvršena i redukcija skupa obeležja, pri čemu su dobijeni znatno bolji rezultati

prepoznavanja (poglavlje V), što takođe ide u prilog prethodno iznetim tvrdnjama.

U sistemu su korišćene heurističke, apriori procene potrebnog broja smeša po modelu, zasnovane na fonetskom predznanju, pri čemu se vodilo računa o prethodno navedenoj razlici izmedju GC klasifikatora sa punom i dijagonalnom kovarijansnom matricom.

U poglavlju V prikazani su rezultati na osnovu kojih se zaključuje da ni primena DCT (*Discrete Cosine Transform*) transformacije nad korišćenim obeležjima, u cilju bar delimičnog dekorelisanja obeležja, uz korišćenje GC klasifikatora sa dijagonalnim kovarijansnim matricama, ne daje značajnija poboljšanja.

V. Eksperimentalni rezultati

A. Opis baze i gramatike

Za potrebe obuke i testiranja korišćena je govorna baza snimljena na Fakultetu tehničkih nauka u Novom Sadu. Baza sadrži snimke oko 1000 govornika (oko 500 muških i 500 ženskih) snimljenih preko javne telfonske mreže. Svi fajlovi su snimani u A-low formatu sa učestalošću odabiranja 8 kHz. Detalji o ovoj bazi su dati u [3].

Baza je podeljena na dve disjunktne celine, deo baze namenjen obuci (koji uzima i njen najveći deo) i deo baze namenjen testiranju. Da bi se smanjile akustičke varijacije u izgovoru koje su posledica razlika u boji i visini glasa kod muškaraca i žena za svaki pol se obučava zaseban sistem za prepoznavanje. Oba sistema koriste ista obeležja i topologiju HMM, ali se obučavaju i testiraju na odvojenim celinama. Test baza za muškarce sadrži 964 snimka odnosno 1547 reči, a za žene 848 snimaka i 1351 reč.

B. Opis obeležja

U ovom radu su vršeni i eksperimenti sa obeležjima jer se pokazalo, da sama promena načina opisivanja izlazne verovatnoće stanja ne mora da rezultira očekivanim poboljšanjem performansi. Ovde će biti predstavljene dve varijante koje su dale najbolje rezultate.

U prvoj varijanti obeležja su grupisana u dva strima. Prvi strim čine obeležja koja opisuju energiju (normalizovana energija, logaritam energije i njihovi prvi i drugi izvodi). Drugi strim čine obeležja koja opisuju obvojnicu spektra (36 obeležja koji predstavljaju nagibe obvojnice spektra kao i njihve prve i druge izvode). Detaljan opis ovih obeležja je dat u [4]. Na dalje, u tekstu, ova obeležja će biti referencirana kao stara.

Posmatranjem obeležja koja opisuju spektralnu obvojnicu u vremenu ustanovljeno je da postoje neočekivano velike varijacije između susednih segmenata koji su u spektrogramu slični. Ovakva obeležja su u okviru jednog stanja generisala raspodelu koja je u velikoj meri odstupala od Gausove. Da bi se smanjile ove varijacije filtar banke koje se koriste pri izdvajanju obeležja su proširene i delimično preklopljene. Na normalizovanu energiju koja po svojoj prirodi nema Gausovu raspodelu primenjena je transformacija po obliku slična tangensnoj funkciji, čime je na neki način postignuta njena Gausianizacija. Pokazalo se i da je prvi izvod više nego dovoljan za adekvatno opisivanje govornog signala. Drugu varijantu predstavlja 26 obeležja grupisanih u jedan strim od kojih 24 obeležja opisuju obvojnicu spektra (12 statičkih i njihovi prvi izvodi) i 2 energiju (transformisana normalizovana energija i prvi izvod normalizovane energije). Na dalje u tekstu će ova obeležja biti referencirana kao nova.

C. Rezultati prepoznavanja

U tabeli 1 je dat prikaz nekoliko sistema za prepoznavanje govora. Prvo slovo u oznaci sistema upućuje na to kojoj grupi govornika je sistem prilagođen ('m' muškarcima, a 'f' ženama). Broj u oznaci ukazuje na to koja vrsta obeležja je korišćena i na koji način su modelovane izlazne verovatnoće (01 označava sistem koji koristi stara obeležja a u Gausovim raspodelama dijagonalne kovarijansne matrice; 02 označava sistem koji koristi nova obeležja, a u Gausovim raspodelama dijagonalne kovarijansne matrice; 03 označava sistem koji koristi nova obeležja a u Gausovim raspodelama pune kovarijansne matrice).

Kao što se iz priloženih rezultata može videti, korišćenje pune kovarijansne matrice je rezultovalo drastičnim smanjenjem greške na nivou reči (GNR) kod sistema prilagođenih ženama (GNR se sa 10.27 % smanjio na 3.33 %). Kod sistema prilagođenih muškracima postoji poboljšanje performansi, ali ne u tako velikoj meri (GNR se sa 7.24 % smanjio na 5.17 %). Ova razlika se može protumačiti da nova obeležja ne opisuju na adekvatan način govor muškaraca. U prilog ovom tvrđenju ide i činjenica da je razlika između performansi sistema koji se razlikuju po vrsti obeležja, a u Gausovim raspodelama koriste samo dijagonalne elemente kovarijansne matrice neznatna kod sistema prilagođenih ženama i prilično velika kod sistema namenjenih muškarcima. Eksperimenti koji treba da daju potpun odgovor na ovo pitanje su u toku.

Realni sistem u toku prepoznavanje istovremeno koristi sisteme prilagođene muškarcima i ženama a performanse koje se pri tome postižu izražene preko GNR-a se nalaze u intervalu:

$$C = \max(\mathrm{GNR}_m, \mathrm{GNR}_f)$$
$$0.9C \le \mathrm{GNR}_{\mathrm{res}} \le C \qquad (6)$$

gde indeks 'm' ukazuje na sistem prilagođen muškarcima, 'f' sistem prilagođen ženama, a 'res' na rezultujući sistem.

Tabela 1: Poređenje rezultata prepoznavanja

Oznaka	br. zamena	br. umetanja	br. brisanja	GNR [%]
m_01	69	36	7	7.24
m_02	127	25	102	16.42
m_03	58	17	5	5.17
f_01	82	52	9	10.27
f_02	88	43	9	10.36
f_03	22	22	1	3.33

Eksperimenti su pokazali da efikasnost GC algoritma zavisi od izbora vrste i dimenzije obeležja. Kao ilustracija ove tvrdnje u nastavku će biti navedeni rezultati dobijeni u eksperimentima sa sistemima prilagođenim muškim govornicima. Na primer za slučaj starih obeležja prilikom korišćenja punih kovarijansnih matrica povećava se GNR sa 7.24 % na 8.92 %. Ovakav rezultat je donekle posledica dimenzionalnosti prostora kao što je objašnjeno u odeljku IV ovog rada. Dimenzionalana zavisnost efikasnosti primene GC algoritma se ogleda i u činjenici da se GNR povećao sa 5.17 % na 9.50 % kada se kod novih obeležja pored prvog izvoda koristi i njihov drugi izvod. Priroda obeležja je bitna za efikasnost algoritma. U slučaju novih obeležja pri čemu ne postoji preklapanje između susednih filtar banki GNR je porastao na 6.85 % u odnosu na 5.17 %. Ovo ukazuje na činjenicu da je za efikasnost algoritma bitno obezbediti kontinualnost obeležja između susednih sličnih segmenata.

TABELA 2: EFEKAT DCT TRANSFORMACIJE NA DEKORELACIJU OBELEŽJA IZRAŽEN PREKO PERFORMANSI SISTEMA

oznaka	br. zamena	br. umetanja	br. brisanja	GNR [%]
m_full	75	43	3	7.82
m_full_dct	77	45	3	8.08
m_diag	161	70	26	16.61
m_diag_dct	146	69	15	14.87
f_full	25	20	2	3.48
f_full_dct	25	20	2	3.48
f_diag	84	37	14	9.99
f_diag_dct	81	34	17	9.77

Standardni i najjednostavniji način smanjena stepena korelacije između obeležja je primena diskretne kosinusne transformacije (DCT). Izvršen je set eksperimenta koji su imali za cilj utvrđivanje efekta dekorelacije koeficijenta primenom DCT-a a čiji su rezultati dati u tabeli 2. Da bi se eliminisao uticaj optimalnog izbora broja parametara, svi eksperimenti su vršeni sa modelima sa po jednom Gausovom raspodelom po stanju. Prvo slovo u oznaci sistema upućuje na to kojoj grupi govornika je prilagođen sistem ('m' muškarcima, 'f' ženama), 'full' ukazuje da su u sistemu Gausove raspodele modelovane punim a 'diag' dijagonalnim kovarijansnim matricama. Sistemi koji u svojoj oznaci imaju 'dct' predstavljaju sisteme kod kojih je nad obeležjima primenjena i DCT.

Kao što se vidi iz tabele 2, kod sistema u kojima se koriste dijagonalne kovarijansne matrice, nešto manji GNR se dobija kada se primeni DCT. Ovo ukazuje da DCT u izvesnoj meri vrši dekorelaciju obeležja, ali ne u dovoljnoj.

Interesantno je zapaziti zanemarljive razlike u performansama kada se koristi veći broj Gausovih raspodela po stanju (rezultati u tabeli 1) i svega jedna Gausova raspodela po stanju (rezultati u tabeli 2). Kod sistema prilagođenog ženama, ovi prvi su za nijansu bolji, ali u nivou statističkih greška.

ZAKLJUČAK

U radu su predstavljeni analiza i rezultati koji demonstriraju prednost u pogledu performansi GC klasifikatora sa punim kovarijansnim matricama u odnosu na GC klasifikatore sa dijagonalnim kovarijansnim matricama, koja se ogleda u značajno manjoj grešci prepoznavanja. Eksperimentalno je potvrđeno da je uzrok tome uzimanje u obzir korelisanosti na nivou klastera od strane GC klasifikatora sa punom kovarijansnom matricom. Uspešnost GC klasifikatora zavisi od dimenzionalnosti prostora obeležja, što eksperimentalni rezultati i potvrđuju. Pored ovoga neophodno obezbediti kontinualnost obeležja između susednih sličnih segmenata.

Pokazano je da primena DCT transformacije na obeležja vrši dekorelaciju obeležja, ali ne u meri kojom bi se dostigle performanse sistema koji je dobijen GC algoritmom sa punim kovarijansnim matricama.

LITERATURA

[1] L.R Rabiner, “A tuuorial on Hidden Markov Models and Selected Applications in Speech Recognition”, .Proceedings of the IEEE, vol. 77, no. 2, February 1989.

[2] A. Webb, “Statistical Pattern Recognition”, London, Arnold, 1999, pp. 29-42.

[3] N. Đurić, D. Pekar, Lj. Jovanov, DOGS 2002, “Structure of SpeechDat(E) Database for Serbian, Recorded over the Public Telephone Network”, Bečej, 2002.

[4] N. Jakovljević, D. Mišković, M. Sečujski, D. Pekar, “Vocal Tract Normalization Based on Formant Positions”, IS-LTC, Ljubljana, 2006.

ABSTRACT

In this paper performance of the two different speech recognition systems based on the Hidden Markov Models and Gaussian Mixture Model are compared. The first system uses full covariance matrix and the second one uses diagonal approximation of covariance matrix in recognition. Both systems are trained by Gaussian classification process.

COMPARATION OF SPEECH RECOGNITION SYSTEMS FOR SERBIAN LENGUAGE BASED ON FULL AND DIAGONAL COVARIANSE MATRICES

Marko.B.Janev, Nikša Jakovljević, Darko Pekar