



VIZUELINI DTW KAO NASTAVNO SREDSTVO ZA POREĐENJE GOVORNIH UZORAKA

Gordana Marković¹, Branko Marković²

Rezime: Algoritam dinamičkog usklađivanja u vremenu (DTW-Dynamic Time Warping) omogućava da se nađe mera sličnosti između uzoraka i donese odgovarajuća odluka. Vizualizacija ovog algoritma, kao i uzoraka koji se porede, računanje njihove distance i odlučivanje koji su najsličniji tema je ovog rada.

Ključne reči: Dinamičko usklađivanje u vremenu, wave fajlovi, obrazovanje, govor.

THE VISUAL DTW AS AN EDUCATIONAL TOOL FOR SPEECH PATTERNS COMPARISON

Summary: The Algorithm of Dynamic Time Warping allows to find out a measure between patterns and to make the right decision. Visualizing this algorithm and patterns, calculating a distance and making decision who matches the best is explained in this paper.

Key words: Dynamic Time Warping, wave files, education, speech.

1. UVOD

U praksi postoji vrlo česta potreba da se određeni uzorci podataka porede. Cilj ovog poređenja je da se nađe mera njihove sličnosti i često da se od više ponuđenih uzoraka pronađe onaj koji najviše liči originalu. Ovaj proces posebno je prisutan u primeni sistema za automatsko prepoznavanje govora (ASR – Automatic Speech Recognition) gde se koriste govorni uzorci koje odlikuje promena u vremenu. Jedan od najčešće korišćenih algoritama za poredjenje govornih uzoraka je DTW. On omogućava da se izvrši vremensko uravnavanje uzoraka i njihova normalizacija, a na osnovu toga i poređenje. Softverski alat razvijen za svrhu obrazovanja kako ovaj algoritam funkcioniše omogućava da se vidi fizički izgled uzoraka koji se porede kao i formirana putanja po kojoj se računa minimalna distanca za DTW algoritam. Takođe ovaj softver daje prikaz izračunate distance između uzoraka i daje mogućnost da se više uzoraka poredi sa referentnim i pronađe koji je od uzorka najsličniji referentnom (na bazi minimalne distance). Softverski paket je razvijen u programskom jeziku Visual Basic 6.0, a uzorci su snimani korišćenjem softverskog paketa WiseEdit.

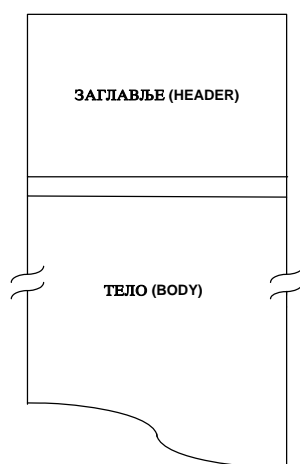
¹ Gordana Marković, Prehrambeno-ugostiteljska škola, Čačak, E-mail: branko333@nadlanu.yu

² Branko Marković, Visoka škola strukovnih i tehničkih studija, Čačak, E-mail: branko333@ptt.yu

2. GOVORNI UZORCI

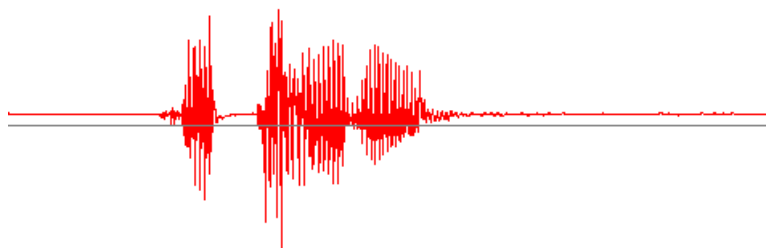
Osnovna potreba za poređenjem govornih uzoraka javlja se u slučajevima kada se želi verifikovati govornik. Naime često je potrebno da na osnovu izgovorene ključne reči, npr. lozinke, sistem može da odredi da li određeni govornik ima pravo da pristupi sistemu ili ne. Ukoliko neautorizovani govornik pokuša da uđe u sistem treba da bude odbijen. U tom smislu vrši se poređenje referentnog i ostalih govornih uzoraka.

Govorni uzorci koji se koriste u ovom nastavnom paketu imaju Microsoft-ov standard za snimanje i manipulaciju, a to je «wave» format. Izgled govornog uzorka «wave» formata je prikazan kao na slici 1.



Slika 1: Microsoft-ov «wave» fajl

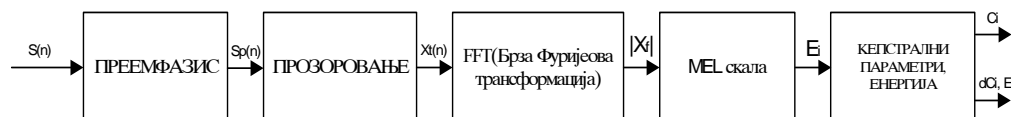
Ovaj fajl predstavlja binarni zapis snimljenog govornog uzorka i sastoji se iz dva dela: zaglavlja i tela. Zaglavlje nosi informacije o tome koja je frekvencija odmeravanja, koliki je broj bita po odmerku, broj bajta po sekundi, broj kanala koji se koristi, dužina fajla itd., dok telo nosi korisne informacije, tj. sam zapis govornog uzorka. Izgled snimljenog govornog fajla koji predstavlja reč «otvori» prikazan je na slici 2.



Slika 2: Prikaz govornog fajla za reč «otvori»

Na osnovu snimljenih govornih fajlova kreira se skup koji se smešta u bazu podataka, tj. apsolutni linkovi do fajlova smeštaju se u bazu, dok sami «wave» fajlovi se mogu smestiti na određeni direktorijum na koji baza pokazuje.

Posle odgovarajućeg snimanja potrebno je izvršiti predobradu ovih signala odnosno poželjno je dobiti skup višedimenzionih vektora koji služe za pređenje. Za ovaj projekat korišćeni su skupovi keprstralnih koeficijenata koji se dobijaju pošto se signal predobradi prolazeći kroz module za preemfazis, prozorovanje, brzu Furijerovu transformaciju i dobijanje keprstralnih koeficijenata otežanih na mel skali, kao što je prikazano na slici 3.



Slika 3: Predobrada govornog signala

3. DTW ALGORITAM

Da bi se poredili govorni uzorci potrebno ih je najpre pretvoriti u odgovajuće višedimenzione vektore. Ako se za dva vektora x i y iz vektorskog prostora Z želi naći mera sličnosti onda se definiše funkcija distance d koja je nenegativna i ima osobinu:

$$d(x, y) \geq 0 \quad (1)$$

Postoje različite vrste distanci kao mere između vektora govornih uzoraka, a najčešće korišćene su: logaritamsko keprstralno rastojanje, keprstralno rastojanje, otežano keprstralno rastojanje, rastojanje maksimalne verodostojnosti (ML), spektralno rastojanje na frekvencijskim skalama[1] itd.

Otežano keprstralno rastojanje se može odrediti na osnovu formule:

$$d_{cpw}^2(K) = \sum_{n=1}^K (w(n)C_n - w(n)C'_n)^2 \quad (2)$$

gde je $w(n)$ otežanje, C_n i C'_n keprstralni koeficijenti, a K broj uzetih koeficijenata.

Da bi se govorni uzorci mogli da porede potrebno ih je na neki način ujednačiti (npr. da im se poklapaju počeci i završeci), kao i da se druge fluktuacije (npr. promena energije) dovedu u određene granice. Zbog toga se vrši normalizacija i ujednačavanje posmatranih signala, a kao jedan od najefikasnijih algoritama za sam proces poređenje koristi se DTW, tj. dinamičko usklađivanje u vremenu koji je baziran na tehnici dinamičkog programiranja. Tehnika dinamičkog programiranja podrazumeva da se za dve tačke u prostoru Q koji ima n tačaka (npr. za tačke i i j) pronadje optimalna staza od i do j kroz potencijalno veći broj tačaka (staza) koje ove tačke razdajaju i da ta staza bude optimalna (sa stanovišta koštanja).

Ako se dva govorna uzorka X i Y predstavljaju sa spektralnim sekvencama tj. vektorima $(x_1, x_2, \dots, x_{T_x})$ i $(y_1, y_2, \dots, y_{T_y})$ onda je potrebno naći rastojanje između ova govorna uzorka.

Ako se sa $d(x_i, y_j)$ obeleži kratkovremena spektralna distorzija između dva elementa i i j od dva različita vektora X i Y ili uprošćenije sa $d(i, j)$ tada korišćenjem linearne vremenske normalizacije rastojanje između vektora X i Y dobija oblik:

$$d(X, Y) = \sum_{i=1}^{T_x} d(i, j) \quad (3)$$

pri čemu kao uslov linearne normalizacije mora važiti uslov:

$$j = \frac{T_y}{T_x} i \quad (4)$$

Da bi uskladjivanje bilo što realnije uvode se i funkcije razvlačenja u vremenu (warping functions) φ_x i φ_y , kao i otežanje puta $w(k)$ i faktor normalizacije staze M_φ tako da formula (4) dobija oblik:

$$d_\varphi(X, Y) = \sum_{k=1}^T w(k) d(\varphi_x(k), \varphi_y(k)) \frac{1}{M_\varphi} \quad (5)$$

Dalje je potrebno za sve moguće skupove $\varphi = (\varphi_x, \varphi_y)$ naći optimalnu stazu što se predstavlja formulom:

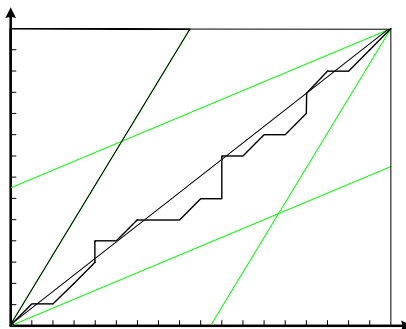
$$d(X, Y) \equiv \min d_\varphi(X, Y) \quad (6)$$

i što je osnovni zadatak DTW algoritma. Da bi se smanjio potencijalni broj staza po kojima bi se tražile optimalne staze uvode se i odgovarajuća ograničenja kao što su: ograničenja krajeva reči, ograničenje monotonosti, ograničenje lokalnog kontinuiteta, ograničenje opšteg puta i ograničenje otežanog nagiba. Pod pretpostavkom da se početna tačka labelira sa $(1, 1)$, a krajnja sa (T_x, T_y) , da se sa $c((i', j'), (i, j))$ označi otežana akumulirana razlika između tačaka (i', j') i (i, j) onda DTW algoritam ima sledeće korake:

$$1) \text{ Inicijalizacija: } D(1, 1) = d(1, 1)w(1) \quad (7)$$

$$2) \text{ Rekurzija: } D(i, j) = \min_{(i', j')} [D(i', j') + c((i', j'), (i, j))] \quad (8)$$

$$3) \text{ Završetak: } d(X, Y) = D(T_x, T_y) / M_\Phi \quad (9)$$



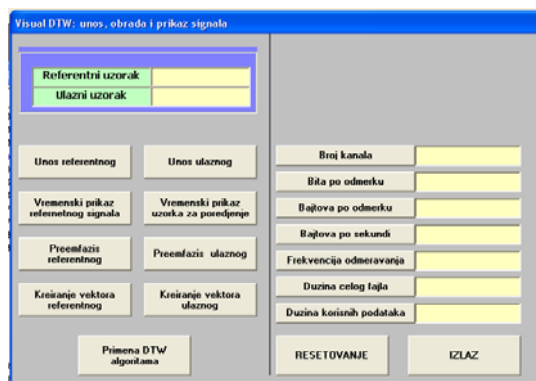
Slika 4: Poređenje uzoraka DTW algoritmom

Slika 4 daje grafički prikaz poređenja ulazne i referentne reči sa labeliranim tačkama $(1, 1)$

i (T_x, T_y) , ograničenjima, kao i stazom po kojoj su uzorci zadovoljavali uslov minimalnog rastojanja.

4. VIZUALIZACIJA DTW-A

Softverski paket koji je razvijen u Visual Basic-u 6.0 namenjen je bližem upoznavanju studenata sa načinom poređenja govornih uzoraka DTW algoritmom kao i za vizualizaciju uzoraka koji se porede. Sastoji se iz određenog broja formi koje su međusobno povezane linkovima i na bazi komandnih dugmića omogućene su sve aktivnosti potrebne za demonstraciju ovog softvera. Osnovna forma koja se dobija posle otvaranja tzv. «splash» prozora je kao na slici 5.



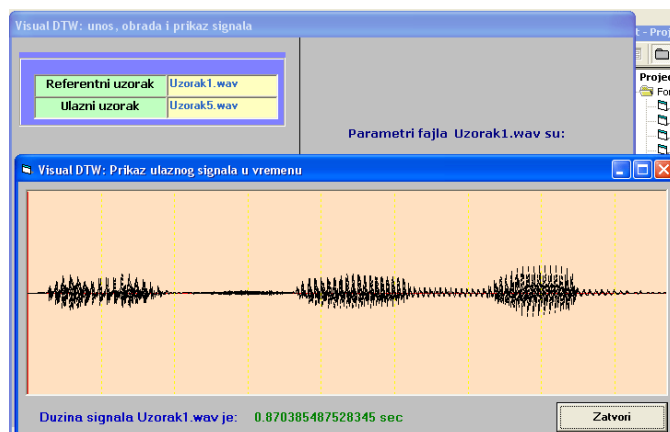
Slika 5: Osnovni meni za vizuelni DTW

Potrebno je najpre uneti nazive fajlova («wav» fajlovi) koji se porede u odgovarajuća polja «Referentni uzorak» i «Ulazni uzorak», a potom korišćenjem odgovarajućih komandnih dugmića («Unos referentnog», «Vremenski prikaz...», «Preemfazis...» i «Kreiranje vektora...») treba formirati vektore za referentni, a zatim istim postupkom i vektore za ulazni, tj. signal koji se porede. Tokom ovih aktivnosti program nam daje rezultate obrade i prikaze signala. Tako npr. posle pritiska na komandno dugme «Unos referentnog» na desnoj strani ekrana dobiće se rezultat analize tog signala kao na slici 6



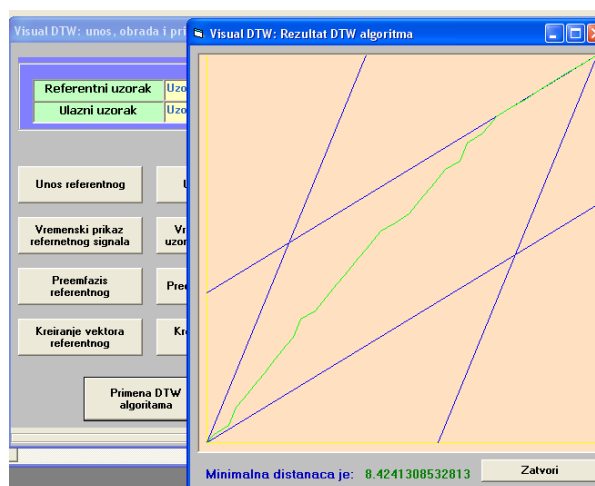
Slika 6: Prikaz analize signala «Uzorak1.wav»

Na sličan način ako se pritisne komandno dugme «Vremenski prikaz..» dobiće se izgled datog signala u vremenu kao i njegova dužina u sekundama (slika 7).



Slika 7: Prikaz vremenskog izgleda signala «Uzorak1.wav»

Pošto se postupak ponovi za oba uneta signala (tj. fajlove koji reprezentuju ove signale) potrebno je pritisnuti na komandno dugme «Primena DTW algoritma». Ova komanda će prouzrokovati akciju da se dva uneta uzorka uporede DTW metodom, odnosno da se vektori koji reprezentuju ove uzorke i koji su dobijeni na ranije opisani način (predobradom ulaznih signala) uporede. Mera njihove sličnosti kao i optimalna staza po kojoj su se poredili daje se pritiskom na ovu komandu, a za konkretan primer to bi izgledalo kao na slici 8.



Slika 8: Rezultat DTW poređenja, grafički i numerički

Posle dobijanja ovog rezultata korisnici nastavnog sredstva mogu upisati nazive uzoraka i rezultat njihovog poređenja i pritiskom na dugme «RESETOVANJE» (slika 6) mogu nastaviti druga poređenja uzoraka ili mogu pritiskom na «IZLAZ» izaći iz programa.

5. ZAKLJUČAK

Radi lakšeg i bržeg upoznavanja sa novim algoritmima i metodama za obučavanje studenata sve više se poseže za softverskim alatima koji imaju mogućnost grafičkog prikaza (GUI). Ovi alati su vrlo praktični jer kao što kaže izreka «Slika vredi hiljadu reči!» to korišćenje slikovitih prikaza pojedinih postupaka daje studentima mogućnost jasnijeg uvida u nekad složene algoritme i procedure. Poređenja govornih uzoraka DTW metodom je jedan od takvih algoritama. Kod njega se složeni matematički pristup predobrade i prikaza poređenja vektora može približiti kroz praktične vežbe i vizualizaciju najvažnijeg dela, a to je optimalna staza po kojoj se vektori porede. On takođe pruža i vizualizaciju signala koji se porede kao i analitičku vrednost minimalne distance.

Dalji pravci istraživanja i usavršavanja ovog softverskog paketa bi se odnosili na ispitivanje mogućnosti korišćenja i druge vrste vektora (koji sadrže delta kepstralne koeficijente, koeficijente bazirani na LPC analizi, otežane na Bark skali i sl.) kao i usavršavanje postojećih mehanizama i sistema ograničenja (izborom drugih Sakoe-Chiba-ovih ograničenja[3]), a takođe i proširivanje softverskog paketa na formiranje klastera gde bi se od skupa zadatih vektora (signala) formirali klasteri prema unapred zadatom kriterijumu maksimalnog dozvoljenog odstupanja.

6. LITERATURA

- [1] L. Rabiner, B-H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993.
- [2] L. Fissore, M. Codogno, G. Pirani, "Isolated word recognition in the mobile-radio system: Experiments and Results", pp. 1207-1210, Signal Processing V, 1990.
- [3] H. Sakoe and S. Chiba, "Dynamic programming optimization for spoken word recognition", IEEE Trans. Acoustics, Speech, Signal Proc., pp. 43-49, 1978.
- [4] D. Sankoff, J.B. Kruskall, "Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison", Addison-Wesley, 1983