Natalija Todorčević

Analiza karakterističnih obeležja pri klasifikaciji muzike

U ovom radu izvšena je klasifikacija muzike u šest različitih muzičkih žanrova: klasična muzika, folk, house, RnB, rock i jazz. Klasifikacija je vršena pomoću neuronskih mreža. Testirana su tri tipa vektora karakteristika: spektar signala, logaritam spektra signala i kepstar signala. Formirane su baze podataka segmenata pesama u trajanju od 1.5, 2, 5 i 10 sekundi, i to po 300 uzoraka iz svakog žanra. Na osnovu dužine trajanja uzoraka i kombinacije ulaznih vektora testirano je 28 neuronskih mreža.

Uvod

Analiza, klasifikacija i proučavanje audio sadržaja ima širok spektar primene u industriji zabave, upravljanju audio arhivama, filtriranju, obradi i skladištenju dolaznih podataka. Muzika se može klasifikovati na više različitih načina, prema izvođaču, žanru muzike itd. Većina sistema za klasifikaciju muzike kombinuje dve faze u obradi podataka: izdvajanje vektora karakteristika i klasifikaciju. Prilikom klasifikacije muzike do sada su korišćene kombinacije različitih vektora karakteristika kao što su zero-crosing rate, širina frekvencionog opsega, spectral centroid, energija signala, Mel-frequency cepstral koeficienti itd. (Toonen Dekkers i Aarts 1995; Scheirer i Slaney 1997; Lu i Hankinson 1998).

Muzički žanrovi ne predstavljaju granice koje su jasno određene, ali veliki broj pesama se može svrstati prema određenim kriterijumima u neke osnovne muzičke žanrove radi lakšeg snalaženja. U ovom radu je ispitana preciznost sistema za klasifikaciju muzike preko neuronskih mreža. Neuronske mreže su se jako dobro pokazale u oblasti prepoznavanja oblika. Mreža je trenirana više puta na uzorcima različitih dužina, čiji su rezultati na kraju upoređivani. Određena su tri tipa karakteristika na osnovu kojih je klasifikacija izvršena: spektar signala, logaritam spektra signala i kepstar signala. Korišćeno je 6 žanrova muzike: klasična, jazz, rock, folk, house i RnB. Formirane su baze sa uzorcima u trajanju od 1.5, 2, 5 i 10 sekundi. Izdvajano je 300 uzoraka iz svakog žanra i to po 5 uzoraka iz svake pesme.

Izdvajanje vektora karakteristika

Svi obrađivani podaci su u WAV (Waveform Audio File) formatu. Za razliku od ostalih audio formata, WAV format nije kompresovan, što znači da sadrži veći broj informacija, kao i to da zauzima veći memorijski prostor. Frekvencija odabiranja iznosi 44.1 kHz i jedna sekunda audio snimka može pružiti zadovoljavajući broj informacija koje bi se prosledile neuronskoj mreži. Kod prepoznavanja oblika i mašinskog učenja, vektori karakteristika pretstavljaju n-dimenzione vektore koji nose informacije o nekom objektu, odnosno signalu. Formirane su baze uzoraka pesama iz kojih su izdvajani vektori karakteristika.

Izdvajana su tri tipa vektora karakteristika (tabela 1):

Tabela 1. Vektori karakteristika

Tip vektora	Dužina vektora
Spektar signala	32
Logaritam spektra signala	32
Kepstar signala	12

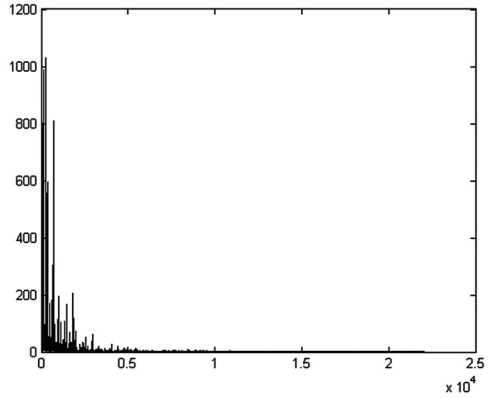
Spektar signala

Spektar signala se dobija pomoću Furijeove transformacije. Diskretna Furijeova transformacija je specifična vrsta diskretne transformacije, koja se koristi u okviru Furijeove analize. DFT najčešće vrši

Natalija Todorčević (1994), Kragujevac, Ivana Cankara 11, učenica 3. razreda Prve kragujevačke gimnazije

MENTOR: Marko Bežulj, MDCS, Beograd

transformaciju funkcije iz vremenskog u frekvencijski domen. U frekvenciskom domenu dobijamo zavisnost amplitude od frekvencija prostih signala od kojih se sastoji složeni signal, tj. zvuk. DFT zahteva unos diskretne funkcije, tj. signala, jer je kontinualni signal sastavljen od beskonačnog broja uzoraka te nije pogodan za kompjutersku obradu. Diskretan signal se može generisati na osnovu odgovarajućeg kontinualnog signala. Vrši se odabiranje (semplovanje) sa određenim periodom T, koji mora biti dovoljno mali kako bi diskretan signal predstavljao zadovoljavajuće dobru aproksimaciju polaznog kontinualnog signala.



Slika 1. Diskretna Furijeova transformacija (DFT) od amplitudske vrednosti

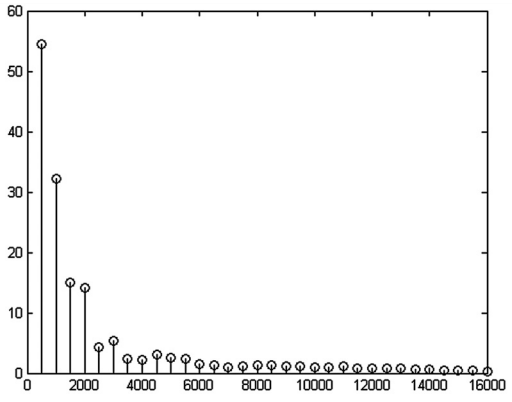
Figure 1. Discrete Fourier Transformation (DFT) of audio signal

Ulaz DFT-a je konačan niz realnih ili kompleksih brojeva, što DFT čini idealnim za obradu informacija preko računara. DFT ima široku primenu u obradi signala i srodnim oblastima u kojima se analizira frekvencija, pri rešavanju parcijalnih diferencijalnih jednačina, konvulaciji, kompresiji podataka u računarima itd. U praksi se koriste algoritmi brze Furijeove transformacije (FFT), te se termin DFT često može zameniti sa FFT.

Na osnovu DFT-a izdvojena su 32 karakteristična obeležja tako što je izršeno usrednjavanje DFT-a ulaznog signala (slika 1) na 32 segmenta (slika 2). Na osnovu Nikvist-Šenonove teoreme odabiranja sa frekvencijom odabiranja 44.1 kHz se mogu očuvati informacije do frekvencije do 22 kHz, što spada u opseg čujnosti ljudskog uha. Međutim na grafiku se može uočiti da su amplitude frekvencija većih od 16 kHz veoma niske, te su frekvencije veće od 16 kHz odbačene pri projektovanju klasifikatora.

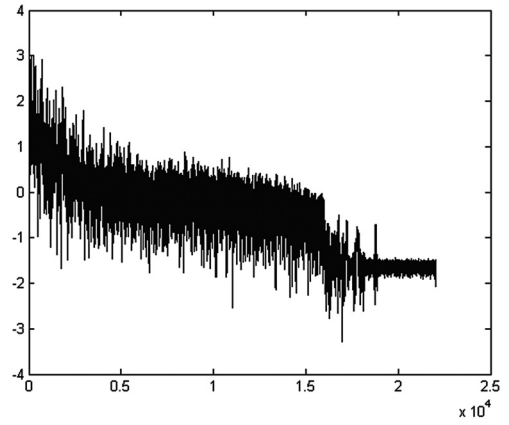
Logaritam spektra signala

Drugi tip vektora karakteristika jeste logaritam spektra signala (slika 3). Usrednjavanjem se izdvajaju 32 karakteristična obeležja, uzimajući vrednosti do 16 kHz (slika 4). Ove vrednosti ističu razliku amplituda veoma malih vrednosti koje se uglavnom nalaze na višim frekvencijama.



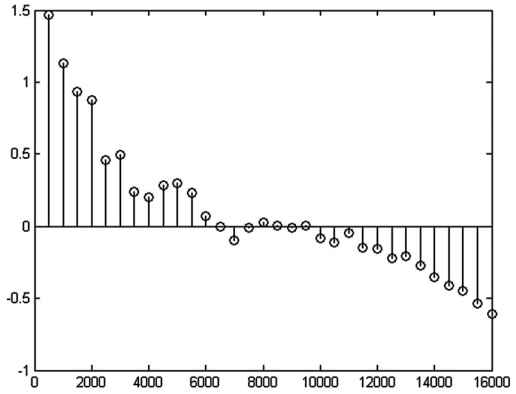
Slika 2. Usrednjena vrednost DFT-a ulaznog signala na 32 segmenta

Figure 2. Average DFT of input signal over 32 bins



Slika 3. Logaritam spektra signala

Figure 3. Logarithm of the signal spectrum

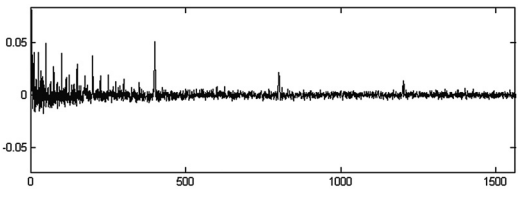


Slika 4. Usrednjena vrednost logaritma spektra signala

Figure 4. Average logarithm of the signal spectrum

Kepstar signala

Kepstar se definiše kao inverzna Furijeova transformacija logaritamske vrednosti spektra signala. Ako za kratko logaritam spektra signala posmatramo kao običan vremenski oblik signala, tada možemo uočiti sporo promenljive komponente i brzo oscilujuće komponente. Drugim rečima, filtriranjem ovog signala bilo bi moguće razdvojiti ta dva dela. Dodatno, ako bismo izračunali Furijeovu transformaciju tog signala, sporo promenljive komponente nalazile bi se na frekvencijama blizu nule. Pojam kepstra upravo proizlazi iz činjenice da se ovde radi o svojevrsnom spektru od spektra. Zbog jako malih vrednosti, vektor karakteristika kepstra signala određuju prvih 12 vrednosti kepstra.



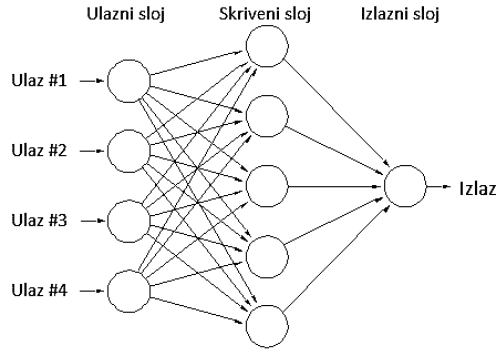
Slika 5. Kepstar signala

Figure 5. Signal cepstrum

Klasifikacija

Neuronska mreža je jedan oblik implementacije sistema veštačke inteligencije, nastale kao proizvod pokušaja da se simuliraju pravi biološki neuronski sistemi. Model se sastoji od čvorova koji mogu biti:

- ulazni čvorovi, koji predstavljaju ulazne atribute,
- skriveni čvorovi, koji utiču na kompleksnost mreže, i
- izlazni čvorovi, koji služe za prikazivanje rezultata (slika 6).



Slika 6. Osnovna struktura neuronske mreže

Figure 6. Basic structure of a neural network

Sami čvorovi se nazivaju neuroni. Svaki ulazni čvor je povezan sa izlaznim kroz vezu kojoj je dodeljena određena težina. Treniranje neuronske mreže se sastoji od toga da se ti težinski koeficijenti međusobno adaptiraju tako da oslikavaju realnu vezu između ulaznih atributa i izlazne vrednosti.

Korišćena je neuronska mreža koja sadrži 12 skrivenih neurona i 6 izlaznih. Izlazni neuroni predstavljaju 6 žanrova muzike. Broj ulaznih neurona zavisi od kombinacije tri tipa vektora karakteristika. Na taj način formirano je 7 neuronskih mreža.

Broj ulaznih neurona je iznosio:

- 76 kada su korišćena sva tri tipa vektora karakteristika
- 64 kada su korišćeni spektar signala i logaritam od spektra signala
- 44 kada su korišćeni spektar signala i kepstar signala ili logaritam od spektra signala i kepstar signala

– 32 kada se korišćeni spektar signala ili logaritam od spektra signala
– 12 kada je korišćen kepstar signala.

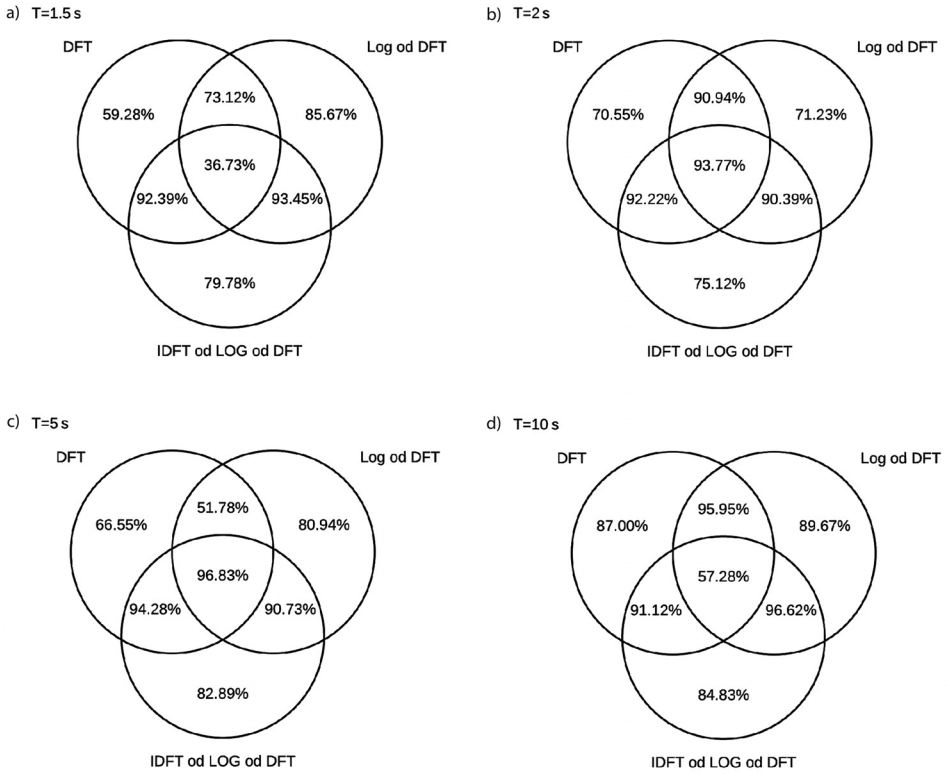
Rezultati i diskusija

Od ukupnog broja podataka 70% je iskorišćeno za treniranje mreže, 15% za verifikaciju, dok je na preostalih 15% podataka testirana mreža. Venovi dijagrami na slici 7. predstavljaju dobijene preciznosti neuronskih mreža pri različitim karakterističnim obeležjima, kao i pri različitim vremenskim intervalima na kojima se izračunavaju karakteristična obeležja.

Na sledećim tabelama su prikazane matrice konfuzije. Na tabelama 2–7 su prikazani rezultati klasifikacije za uzorke u trajanju od 2 s u zavisnosti od tipa vektora karakteristika koji učestvuju u klasifikaciji. Na tabelama 8, 9, 10 i 11 prikazani su rezultati za sva tri tipa vektora karakteristika u zavisnosti od dužine trajanja uzorka.

Tabela 2. Matrica konfuzije za neuronske mreže sa 32 ulazna neurona (sprektar signala) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	240	13	9	32	2	4
Folk	29	191	10	23	32	15
House	44	18	198	30	2	8
Jazz	21	11	4	250	4	10
RnB	23	10	13	16	226	12
Rock	64	16	7	26	22	165



Slika 7. Rezultati klasifikacije neuronskim mrežama, projektovanim za različita karakteristična obeležja, kao i pri različitim vremenskim intervalima na kojima se izračunavaju karakteristična obeležja.

Figure 7. Neural network classification results, with different feature vectors and different time slices for calculating feature vectors.

Tabela 2 predstavlja matricu konfuzije neuronske mreže kod koje ulazni neuroni uključuju samo spektar signala uzoraka koji traju 2 s. Kod ovog tipa neuronske mreže dolazi do svrstavanja većeg broja uzoraka rock, house, jazz i RnB u klasičnu muziku, folk u RnB, dok se kod jazz muzike najveći broj uzoraka tačno svrstao.

Tabela 3 je matrica konfuzije neuronske mreže koja sadrži vektor karakteristika tipa logaritma spektra signala. Kod ovakve neuronske mreže mali broj uzoraka drugih tipova muzike meša sa klasičnom muzikom, dok se folk, house i rock sa najvećom greškom svrstavaju u RnB. Uzorci jazz muzike su u najvećem broju pravilno klasifikovani.

Tabela 3. Matrica konfuzije za neuronske mreže sa 32 ulazna neurona (logaritam spektra signala) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	238	12	17	21	7	5
Folk	2	188	28	19	53	10
House	5	32	185	17	53	8
Jazz	5	5	15	261	6	8
RnB	13	38	15	18	207	9
Rock	2	16	28	15	36	203

Matrica konfuzije neuronske mreže kod koje ulazni neuroni nose podatke o kepstru signala je prikazana u tabeli 4. Najbolje je klasifikovana rock muzika, a najveća greška se pravi prilikom svrstavanja folk muzike kao house.

Tabela 4. Matrica konfuzije za neuronske mreže sa 12 ulaznih neurona (kepstar) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	248	2	3	31	15	1
Folk	2	190	54	21	13	20
House	7	25	209	29	16	14
Jazz	11	11	14	249	9	6
RnB	17	16	29	31	205	2
Rock	0	19	18	11	1	251

Iz tabele 5 može se videti matrica konfuzije koja uključuje dva tipa vektora karakteristika, i to spektar signala i logaritam spektra signala. Veliki broj uzoraka je precizno klasifikovan i ne dolazi do velikog mešanja žanrova. Razultati su zadovoljavajući kao i kod tipova mreže koje kombinuju spektar i kepstar, logaritam od spektra i kepstar (tabele 6 i 7), kao i mreže koja uključuje sva tri tipa vektora za ulazne neurone (tabela 8).

Tabela 5. Matrica konfuzije za neuronske mreže sa 64 ulazna neurona (spektar signala i logaritam spektra) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	273	9	7	9	1	1
Folk	3	267	6	6	4	14
House	3	4	283	7	3	0
Jazz	1	7	2	280	5	5
RnB	5	11	8	2	266	8
Rock	1	10	13	4	4	268

Tabela 6. Matrica konfuzije za neuronske mreže sa 44 ulazna neurona (spektrar signala i kepstar) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	265	0	3	1	31	0
Folk	2	265	12	2	12	7
House	0	3	284	1	4	8
Jazz	6	7	2	274	4	7
RnB	1	2	2	1	292	2
Rock	1	6	5	3	5	280

Tabela 7. Matrica konfuzije za neuronske mreže sa 44 ulazna neurona (logaritam spektra signala i kepstar) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	279	0	1	8	10	2
Folk	2	249	24	6	7	12
House	4	6	283	1	3	3
Jazz	6	9	2	278	4	1
RnB	11	6	10	4	266	3
Rock	5	7	10	3	3	272

Upoređivanjem tabela 8, 9, 10 i 11, gde su uključena sva tri tipa vektora, može se videti zavisnost ovih podataka od trajanja odabiraka.

Tabela 8. Matrica konfuzije za neuronske mreže sa 76 ulaznih neurona (uključuje sva tri tipa) testirane na uzorcima u trajanju od 2 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	297	0	0	3	0	0
Folk	1	284	7	2	2	4
House	7	8	279	2	1	3
Jazz	2	1	0	292	0	5
RnB	11	6	9	12	261	1
Rock	11	7	5	2	0	275

Tabela 9. Matrica konfuzije za neuronske mreže sa 76 ulaznih neurona (uključuje sva tri tipa) testirane na uzorcima u trajanju od 1.5 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	194	1	14	91	0	0
Folk	4	24	168	62	42	0
House	2	7	181	70	40	0
Jazz	20	1	78	197	4	0
RnB	2	17	190	26	65	0
Rock	12	0	184	87	17	0

Tabela 10. Matrica konfuzije za neuronske mreže sa 76 ulaznih neurona (uključuje sva tri tipa) testirane na uzorcima u trajanju od 5 s

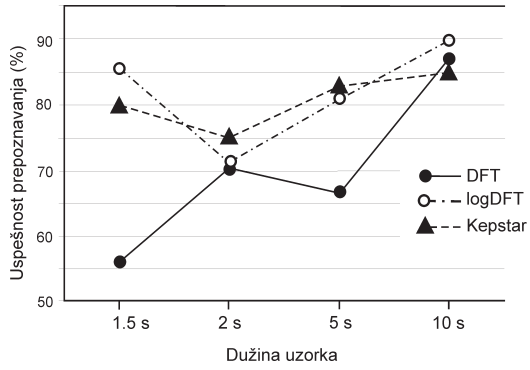
	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	290	1	0	5	1	3
Folk	1	279	5	1	5	9
House	0	4	294	0	0	2
Jazz	2	1	0	296	1	0
RnB	1	1	2	1	293	2
Rock	1	5	3	0	0	291

Najbolje rezultate možemo videti u tabeli 10, gde je dužina trajanja uzorka 5 s. U tom slučaju dolazi do najmanjeg mešanja žanrova. Uzorci u trajanju od 10 i 1.5 daju veoma loše rezultate (slika 7, a i d). Iz tabele 9, kod uzorka u trajanju od 1.5 s se može videti da čak ni jedna pesma iz Rock žanra nije pravilno klasifikovana, kao i klasifikacije skoro svih uzoraka Rock, Folk i RnB muzike u House muziku. Možemo zaključiti da ne postoji linearna zavisnost performansi od dužine trajanja uzoraka, već da postoji optimalna dužina, što je prilikom ovog testiranja 5 s (slika 7 d, tabela 10).

Tabela 11. Matrica konfuzije za neuronske mreže sa 76 ulaznih neurona (uključuje sva tri tipa) testirane na uzorcima u trajanju od 10 s

	Klasika	Folk	House	Jazz	RnB	Rock
Klasika	252	0	0	38	6	4
Folk	10	140	28	39	36	47
House	17	33	100	34	85	31
Jazz	83	0	7	164	17	29
RnB	17	8	16	36	216	7
Rock	17	37	33	41	13	159

Na grafiku na slici 8 prikazana je zavisnost sva tri tipa vektora karakteristika od dužine uzoraka. Može se uočiti da pojedinačno ovi vektori daju najbolje rezultate za dužinu uzorka od 10 s, međutim već smo pokazali da se kombinacijom ovih vektora dobijaju bolji rezultati. Stoga se može zaključiti da ovi vektori pojedinačno ne nose dovoljno informacija za klasifikaciju.



Slika 8. Preciznost klasifikatora čiji su vektori karakteristika DFT, LogDFT i Kepstar za različite dužine uzorka

Figure 8. Neural network precision, with feature vector containing only DFT (circle/solid), LogDFT (ring/dash-dot) and Cepstrum (triangle/dash) with different time slices

Zaključak

U analizi različitih parametara sistema za klasifikaciju, najveća preciznost je postignuta sa uzorcima od 5 s i vektorom karakteristika koji obuhvata DFT, Log(DFT) i Kepstralne koeficijente, i iznosi 96.83%. Slične preciznosti, 96.62% i 95.95%, postignute su sa uzorcima od 10 s i vektorom karakteristika koji obuhvata Log(DFT) i kepstar, odnosno DFT i Log (DFT).

Analizom klasifikacije na osnovu zasebnih parametara mogu se doneti sledeći zaključci:

- Uspešnost prepoznavanja korišćenjem DFT signala kao vektora karakteristika raste sa dužinom uzorka na kojem se izračunava DFT.
- Uspešnost prepoznavanja pomoću Log(DFT) signala kao vektora karakteristika ima složenu zavisnost od dužine uzorka. Bitno je naglasiti da lokalni minimum nije dodatno istražen.
- Uspešnost prepoznavanja korišćenjem kepstralnih koeficijenata signala kao vektora karakteristika je invarijantna u odnosu na vreme dužine uzorka.

Dodatno je potrebno ispitati kombinovanje vektora karakteristika različitih dužina. Na primer, da li DFT 2 s, LogDFT 10 s i kepstar 10 s (što su njihovi zasebni maksimumi) donose unapređenje preciznosti sistema.

U literaturi se često preporučuje da broj neurona u skrivenom sloju bude dva puta veći od broja ulaznih neurona, međutim u tom istraživanju kombinovani vektor karakteristika ima veliki broj koeficijenata. To znači da preporuka za skriveni sloj čini neuronsku mrežu veoma velikom, što donosi probleme u samom algoritmu treniranja (potrebna memorija, trajanje treniranja i broj potrebnih odabiraka u bazi usled većeg broja) zbog toga je broj neurona bio konstantan za sva merenja. Potrebno je dodatno istražiti kako broj neurona u skrivenom sloju utiče na uspešnost prepoznavanja muzičkog žanra.

Literatura

Lu G., Hankinson T. 1998. A technique towards automatic audio classification and retrieval. *Signal Processing Proceedings, 1998. ICSP'98., 1998 IEEE Fourth International Conference on*, **2**: 1142-1145.

Scheirer E., Slaney M. 1997. Construction and evaluation of a robust multifeature speech/music discriminator. *Acoustics, Speech and Signal processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, **2**: 1331–1334.

Toonen Dekkers R. T. J., Aarts R. M. 1995. On a very low-cost speech-music discriminator. Tech. Rep. No. 124/95. Tehnička napomena. Philips Research Nat. Lab, Eindhoven, Netherlands

---

Natalija Todorčević

Music Genre Classification

In this paper music is classified into six different music genres: classical music, folk, house, RnB, Rock and Jazz music. The classification has been done using neural networks. Three types of features vectors have been used: spectrum of a signal, logarithm of the signal spectrum and cepstrum of signal. The feature vector was calculated on 4 different time slices (1.5, 2, 5, and 10 seconds). Based on time slices and the feature vectors, 28 different neural network classification have been implemented and tested. A brief classification of results, with different feature vectors and different time slices for calculating feature vectors, is shown in Figure 7.