به نام خدا

پردازش گفتار (۱۴۸-۰۵-۸۳)

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

تاریخ تحویل: ۱۴۰۲/۰۲/۰۴

تمرین شماره ۲

۱. (۱۲٪) [ویژگی‌های سیستم‌ها]

خواص علی بودن، خطی بودن، تغییرپذیری با زمان، و حافظه‌دار بودن در مورد سیستم‌های زیر بررسی کنید.

آ- $y(t) = 2x(\frac{t}{2}) - 1$

ب- $y[n] = nx[n]$

پ- $y(t) = \int_{-\infty}^{t} x(\tau) d\tau$

ت- $y(t) = \begin{cases} x(t) + x(t-2) & t \geq 0 \\ 0 & o.w. \end{cases}$

ث- $y(t) = \sin(x(t))$

ج- $y(t) = \frac{1}{2M+1} \sum_{k=-M}^{+M} x[n-k]$

۲. (۱۸٪) [سیستم‌های LTI و کانولوشن]

آ- سیستم LTI، با پاسخ ضربه $h(t) = \begin{cases} 4|1-|t|| & |t| \leq 1 \\ 0 & o.w. \end{cases}$ و ورودی $x(t) = \begin{cases} 2 & -2 \leq t < 0 \\ 0 & o.w. \\ -2 & 0 \leq t \leq 1 \end{cases}$ را در نظر بگیرید. **بدون محاسبه انتگرال کانولوشن** به پرسش‌های زیر پاسخ دهید:

- خروجی در کدام بازه‌ها دقیقا برابر با صفر است؟
- خروجی در کدام t بیشینه است؟
- خروجی در $t=1$ چه مقداری دارد؟

ب- همانگونه که در شکل زیر (سمت چپ) نشان داده‌شده‌است، رابطه ورودی و خروجی یک سیستم علی از اتصال سری سه سیستم دیگر درست شده‌است. پاسخ ضربه کل سیستم، در سمت راست نشان داده شده است. اگر $h_2[n] = u[n] - u[n-2]$ باشد:

- پاسخ ضربه سیستم یکم ($h_1[n]$) چیست؟
- پاسخ کل سیستم به ورودی $x[n] = u[n] - 2u[n-1] + u[n-2]$ حساب کنید.

h.veisi@ut.ac.ir

به نام خدا

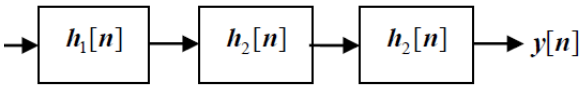
دانشکده علوم و فنون نوین

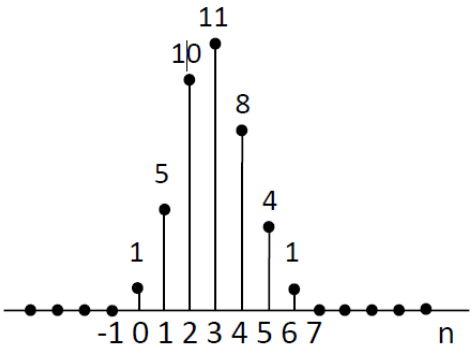
پردازش گفتار (۱۴۸-۰۵-۸۳)

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

تمرین شماره ۲

تاریخ تحویل: ۱۴۰۲/۰۲/۰۴





۳. (۳۵٪) [پیاده‌سازی: شباهت/فاصله بین فرمان‌های آوایی]

در این پرسش می‌خواهیم یک برنامه بنویسید که بتواند صداها را با استخراج ویژگی، شباهت/فاصله بین آواها محاسبه کند و نزدیک‌ترین آوا بازگردانده شود. برای انجام این بخش از این لینک زیر قابل دریافت است.

https://www.kaggle.com/datasets/jbuchner/synthetic-speech-commands-dataset?resource=download

در این دادگان، ۳۰ دسته فایل صدا با نرخ نمونه‌برداری 16KHz قرار دارد. هر دسته دارای تعداد متغییری فایل صدا به طول ۱ ثانیه است. مسئله دسته‌بندی صداها همانطور که در تمرین قبل گفته شد بیشتر با ویژگی از سیگنال گفتار بنام MFCC انجام می‌گیرد. در این تمرین تلاش می‌کنیم بیشتر با این ویژگی و سیگنال‌های آوایی آشنا شویم. برای بدست آوردن این ویژگی، کتابخانه librosa می‌تواند کمک‌کننده باشد (در مورد این ویژگی، سر کلاس TA، صحبت شد). حال بنا داریم برای آشنایی بیشتر با سیگنال آوایی با اعداد تولید شده بیشتر سرکار داشته باشیم.

در ادامه، می‌خواهیم تا با بکارگیری شباهت/فاصله بین سیگنال‌های آوایی، کار شباهت/فاصله بین فرمان‌های آوایی را انجام دهید. برای این بخش باید:

گام ۱- دادگان از تک تک فایل‌ها خوانده شود (۵۰۰ صوت اول خوانده شود).

گام ۲- تابعی برای استخراج این ویژگی فایل‌های صدا بسازید و دادگان را فراهم کنید.

گام ۳- به صورت تصادفی، ۸۰٪ از دادگان برای آموزش و ۲۰٪ باقیمانده برای آزمون مدل تقسیم

h.veisi@ut.ac.ir

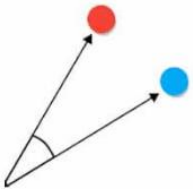
شود. دقت کنید که مدل در آموزش خود نباید از این ۲۰٪ داده را ببیند.

گام ۴- بردار بدست آمده از آواها را می‌خواهیم با سنجه‌های محاسبه نزدیکی و شباهت (شباهت/فاصله) بین فرمان‌های آوایی انجام شود.

برای این کار، از آوای ورودی ویژگی MFCC یافت می‌شود و با معیارهای شباهتی/فاصله که در ادامه تعریف شده است تعیین می‌گردد به کدام فرمان نزدیک است.

یادداشت:

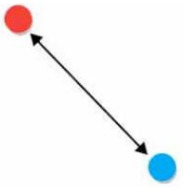
*۱. **شباهت کسینوسی**: شباهت کسینوسی یک معیار شباهت است که پایه آن محاسبه‌ی میزان کسینوس زاویه‌ی بین دو بردار است. در صورت انطباق دو بردار (در این معیار نشانه شباهت کامل است) که زاویه‌ی بین دو بردار صفر می‌باشد مقدار آن برابر ۱ خواهد شد و در کمترین میزان شباهت دو بردار یعنی اگر زاویه بین دو بردار ۱۸۰ درجه باشد نتیجه این معیار ۱- خواهد شد فرمول آن در شکل ۱ آورده شده است.*



$$D(x,y) = cos(\theta) = \frac{x \cdot y}{\|x\| \, \|y\|}$$

شکل ۱ فرمول شباهت کسینوسی

*۲. **فاصله اقلیدسی**: معیاری است که می‌توان آن را با محاسبه طول دو بردار متصل بهم توصیف کرد. معیاری که به بهترین وجه می‌تواند به عنوان طول یک قطعه متصل به دو بردارتوضیح داده شود. فرمول فاصله اقلیدسی بسیار ساده است و برای محاسبه فاصله دو نقطه در مختصات دکارتی از قضیه فیثاغورث استفاده می‌کند. فاصله اقلیدسی کوتاه ترین مسیر بین دو نقطه را به عنوان فاصله پیدا می‌کند فرمول آن در شکل ۲ آورده شده است.*



$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2)}$$

شکل ۲ فرمول فاصله اقلیدسی

*۳. **فاصله منهتن** : فاصله منهتن، اشاره به ساختار شبکه گونه خیابان‌ها و تقاطع‌های منطقه منهتن در*

h.veisi@ut.ac.ir

به نام خدا

پردازش گفتار (۱۴۸-۰۵-۸۳)

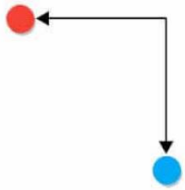
نیم‌سال دوم ۱۴۰۲-۱۴۰۱

تمرین شماره ۲

تاریخ تحویل: ۱۴۰۲/۰۲/۰۴

دانشکده علوم و فنون نوین

شهر نیویورک دارد. این مفهوم از قرن هجدهم در تحلیل‌های رگرسیون و لسو (LASSO) مورد استفاده بوده است. برای محاسبه فاصله بین دو برادری که مقادیر واقعی دارد استفاده می‌شود. فرمول آن در شکل ۳ قابل مشاهده است.



$$D(x,y) = \sum_{i=1}^{k} |x_i - y_i|$$

شکل ۳ فرمول فاصله منهتن

در این گام ابتدا توابعی برای معیارهای شباهت تعریف شود سپس داده‌های تست به عنوان ورودی داده شود و تک تک معیارهای شباهت (شباهت کسینوسی ، فاصله اقلیدسی، فاصله منهتن) برای تمام داده‌ها محاسبه شود و داده‌ای که کمترین فاصله را با داده‌ی ورودی داشت به عنوان خروجی بازگردانده شود.

برای تک تک معیارهای شباهت recall، precision، accuracy، f1 score گزارش شود.

۴. (۳۵٪) [پیاده‌سازی: سنتز گفتار فارسی]

در این سوال می‌خواهیم یک برنامه بنویسید که قادر باشد اعداد صفر تا یک میلیون را سنتز کند.

الف) ورودی این برنامه یک عدد (به صورت عددی یا حروفی) است (مانند ۱۴۰۹ یا دویست و دو هزار و ده) و خروجی آن فایل صوتی معادل است. برای این کار قطعاتی از کلمات لازم را ضبط کرده (نرخ نمونه‌برداری 16 KHz، چندی‌سازی 16 Bits، به صورت مونو و فرمت Wav) و با اتصال آنها به هم عدد درخواستی را بسازید. تعداد قطعات ضبط شما چند تاست؟ علاوه بر ارسال کد و قطعات ضبط شده، فایل‌های خروجی برنامه خود را برای دو عدد ۱۴۰۹ و دویست و دو هزار و ده نیز ارسال کنید.

ب) سیستم سنتز گفتار خود را به صورتی تغییر دهید که مبتنی بر واج باشد. برای این کار همه واج‌های فارسی را که در جدول زیر آورده شده، هر کدام را در یک فایل، ضبط کنید (نرخ نمونه‌برداری 16 KHz، چندی‌سازی 16 Bits، به صورت مونو و فرمت Wav). قسمت‌های سکوت ابتدا و انتهای هر کدام از فایل‌ها را به صورت دستی حذف کنید. با پشت سر هم قرار دادن این واج‌ها، سیگنال صوتی دو عدد

h.veisi@ut.ac.ir

به نام خدا

دانشکده علوم و فنون نوین

پردازش گفتار (۱۴۸-۰۵-۸۳)

نیم‌سال دوم ۱۴۰۲-۱۴۰۱

تاریخ تحویل: ۱۴۰۲/۰۲/۰۴

تمرین شماره ۲

۱۴۰۹ و دویست و دو هزار و ده را تولید کنید و به عنوان خروجی ارسال کنید. کیفیت این دو فایل چطور است؟! مزیت این روش این است که علاوه بر اعداد، می‌توانید هر کلمه دیگری در زبان فارسی را سنتز کنید، برای تست این موضوع فایل صوتی سه عبارت/کلمه "دانشگاه تهران"، "به کجا چنین شتابان" و اسم و فامیل خودتان را تولید کنید و ارسال کنید. برای افزایش قابل فهم بودن صداهای تولیدی در این حالت چه روشی را پیشنهاد می‌کنید؟

همه واج‌های ضبط شده و صداهای تولیدی اعداد و کلمات را به همراه پاسخ این تمرین ارسال کنید.

#	IPA	Char	Code	Farsi Letter	Phonetic Description
1	I	i	105	اي	high front unrounded
2	e	e	101	اِ	mid front unrounded
3	a	a	97	آ	low front unrounded
4	u	u	117	او	high back rounded
5	o	o	111	اُ	mid back rounded
6	/	/	47	آ	low back rounded
7	p	p	112	پ	unvoiced bilabial plosive
8	b	b	98	ب	voiced bilabial plosive
9	t	t	116	ت، ط	unvoiced dental plosive
10	d	d	100	د	voiced dental plosive
11	k	k	107	ک	unvoiced velar plosive
12	g	g	103	گ	voiced velar plosive
13	q	q	113	ق، غ	voiced uvular plosive
14	]	]	93	أ، ؤ ، ع	glottal stop
15	$	$	36	چ	unvoiced alveopalatal affricate closure
16	,	,	44	ج	voiced alveopalatal affricate
17	f	f	102	ف	unvoiced labiodental fricative
18	v	v	118	و	voiced labiodental fricative
19	s	s	115	س، ث، ص	unvoiced alveolar fricative
20	z	z	122	ز، ذ، ظ، ض	voiced alveolar fricative
21	.	.	46	ش	unvoiced alveopalatal fricative
22	[	[	91	ژ	voiced alveopalatal fricative
23	x	x	120	خ	unvoiced uvular fricative
24	h	h	104	ه، ح	unvoiced glottal fricative
25	l	l	108	ل	lateral alveolar
26	r	r	114	ر	trill alveolar
27	m	m	109	م	nasal bilabial
28	n	n	110	ن	nasal alveolar
29	y	y	121	ي	approximant palatal

h.veisi@ut.ac.ir