هفتمین کنفرانس ملی مهندسی برق و الکترونیک ایران
دانشگاه آزاد اسلامی گناباد
۲۸ و ۲۹ مرداد ماه ۹۴

تجزیه و تحلیل ویژگی های انتخاب شده بر اساس روش های فیلتر در تشخیص احساسات از سیگنال‌های صحبت

نرجس یزدانیان، حمید محمودیان
دانشکده مهندسی برق
دانشگاه آزاد اسلامی واحد نجف آباد
اصفهان، ایران

چکیده – امروزه یکی از مسائلی که در تعامل انسان و ماشین نقش مهمی ایفا می نماید تشخیص احساس از روی سیگنال گفتار می‌باشد. در این مقاله به منظور بازشناسی احساسات گوینده ویژگی‌های پرکاربردی چون ضرایب کپسترال فرکانس مل، ضرایب کپسترال پیشگویی خطی، ضرایب پیشگویی خطی ادراکی، فرکانس فرمنت، نرخ عبور از صفر و فرکانس گام به عنوان بردار ویژگی از نمونه‌های صوتی استخراج شده‌است و پس از رتبه‌بندی این ویژگی‌ها به کمک روش‌های فیلتر چون همبستگی پیرسون، آزمون t و آنتروپی به طبقه بندی هر حالت احساسی از حالت طبیعی با کمک طبقه بند ماشین بردار پشتیبان پرداخته شده است که در این مقاله سعی بر آن است که با انتخاب ویژگی‌های منفرد از گروه‌های مختلف ویژگی‌های سیگنال گفتار به بالاترین راندمان در تشخیص احساسات از گفتار رسید که با توجه به مقایسات انجام گرفته برروی روش-های انتخاب ویژگی، روش آزمونt دارای میزان درستی بیشتر با تعداد ویژگی کمتر می‌باشد.

واژه‌های کلیدی – بازشناسی احساسات؛ آنتروپی؛ همبستگی پیرسون؛ آزمون t؛ ماشین بردار پشتیبان

۱. مقدمه

پردازش گفتار به عنوان یکی از زیرشاخه‌های پردازش سیگنال، به سرعت در حال گسترش است. علی‌رغم پیشرفت‌های گسترده در حوزهٔ پردازش گفتار، استخراج و درک احساس پنهان در گفتار انسان (مانند خشم، شادی و غیره) از یک سو و تولید گفتار احساسی مناسب از سوی دیگر، همچنان یکی از چالش‌های مهم برای ساخت ماشین‌های هوشمند محسوب می‌شود [**1**,**2**].

بطور کلی سیستم تشخیص احساسات از گفتار شامل سه مرحله اصلی استخراج ویژگی، انتخاب ویژگی و طبقه بندی می باشد. ویژگی‌های صوتی استخراج شده می‌توانند کارآیی سیستم‌های بازشناسی گفتار را به میزان زیادی تحت تاثیر قرار دهند به همین دلیل در مرحله اول باید اطلاعاتی از سیگنال استخراج شود که حداکثر همبستگی را با احساسات داشته باشد و در عین حال به سایر عوامل از جمله محتوای گفتار وابسته نباشد [3].

فعالیت‌های زیادی در زمینه تشخیص احساسات از گفتار تاکنون انجام شده است که در آنها از ویژگی‌های پرکاربردی چون ضرایب کپسترال فرکانسی مل (MFCC) و مشتقات آنها [4]، ضرایب کپسترال پیشگویی خطی (LPC) [**5**]، ضرایب پیشگویی خطی ادراکی (PLP)، فرکانس فرمنت، نرخ عبور از

صفر، ضرایب کپسترال [6] و فرکانس گام [7] استفاده شده است.

پتروشین در سال ۱۹۹۹ با ترکیب ویژگی‌های فرکانس پیچ و انرژی فرمنت و با بکارگیری شبکه عصبی توانست به طبقه بندی چهار احساس عصبابیت، شادی، ترس و طبیعی با میزان درستی ۷۰٪ برسد [8]. هچنین Pao و همکارانش در سال ۲۰۰۵ و ۲۰۰۷ با استفاده از ویژگی‌های MFCC LPC,PLP, موفق به طبقه بندی چهار حالت احساسات شامل عصبابیت، شادی، ناراحت و طبیعی شدند [9,10] و درسال ۲۰۰۹ Bozkurt و همکارانش به طبقه بندی پنج احساس مختلف با بکارگیری ویژگی‌های اسپکترال، عروضی و ویژگی‌هایی براساس مدل مخفی مارکف به میزان درستی ۶۳٪ رسید [11].

پس از انتخاب ویژگی‌های تاثیرگذار از ماشین بردار پشتیبان به عنوان طبقه بند دو کلاسه استفاده شده است. John در سال ۱۹۹۴ دو روش برای انتخاب ویژگی بیان کرده که عبارتند از روش فیلتر و لفاف بسته بندی [12] که در این مقاله از روش‌های فیتر مانند همبستگی پیرسون، آزمون t و آنتروپی استفاده شده است.

در این مقاله سعی بر آن است که با انتخاب ویژگی های منفرد از گروه‌های مختلف ویژگی‌های سیگنال گفتار با استفاده از روش‌های فیلتر و ترکیب این ویژگی های تاثیرگذار در تشخیص احساسات به بالاترین راندمان رسید.

در ادامه مقاله در فصل دوم به بررسی سیستم تشخیص احساسات از گفتار و در فصل سوم به آزمایش‌ها و نتایج پرداخته شده است .

۲. سیستم تشخیص احساسات

سیستم‌های تشخیص احساس از روی گفتار معمولاً از دیدگاه تشخیص الگو، در سه بخش استخراج ویژگی، انتخاب ویژگی و طبقه بندی مورد بررسی قرار می‌گیرند.

۲,۱. استخراج ویژگی

به عنوان اولین گام در تشخیص احساسات بایستی در مورد ویژگی‌هایی که در جهت تفکیک احساسات مختلف مناسب می‌باشند تصمیم‌گیری کنیم. در این مقاله ویژگی‌های صوتی پرکاربرد از جمله MFCC ، LPC، PLP فرمنت‌ها، جیتر، شیمر، عبور از صفر، انرژی و فرکانس گام مورد بررسی قرار گرفته‌اند که در جدول ۱ تعداد هر ویژگی و نحوه نمایش ویژگی‌ها مشخص شده است.

جدول ۱- ویژگی‌های مورد استفاده در مقاله و نحوه نمایش ویژگی‌ها

ویژگی ها	تعداد ویژگی ها	نحوه نمایش
Minimum	1	min
Maximum	1	max
Mean	1	mean
Zero Crossing	1	zc
Standard Deviation	1	std
Energy	1	en
Feb	12	f_1 تا f_{12}
Formant	3	fo1 تا fo_3
PLP	9	plp_1 تا plp_9
LPC	11	lpc_1 تا lpc_{11}
MFCC	13	mf_1 تا mf_{13}
Amplitude	1	amp
Pitch	1	p
Cepstrum	12	cep_1 تا cep_{12}
Shimmer	1	sh
Jitter	1	ji

۲,۲. انتخاب ویژگی

مرحله بعدی مرحله انتخاب ویژگی است، باید از بین مجموعه ویژگی‌های مناسب احتمالی با استفاده از یک معیار تفکیک کننده، بهترین ویژگی‌ها را انتخاب کنیم. روش‌های انتخاب ویژگی به دو دسته فیلتر و لفاف بسته بندی تقسیم

میشوند [13]. در روش لفاف بسته بندی از یک تابع دسته بندی برای ارزیابی شایستگی زیرمجموعه‌های ویژگی استفاده می شود، این روش از فیدبکی از الگوریتم یادگیری اعمال شده استفاده می‌کند [15,14]. در روش فیلتر از هیچ تابع دسته بندی استفاده نمی‌شود. روش فیلتر یک روش از پیش انتخاب شده‌ای است که مستقل از الگوریتم یادگیری ماشین اعمال شده می‌باشد. روش فیلتر به این صورت عمل می‌کند که برای هر ویژگی یک رتبه اعتبار محاسبه می شود، در این مقاله از معیارهای همبستگی پیرسون، آزمون t و آنتروپی برای رتبه بندی ویژگی های استفاده شده است. سپس این رتبه‌ها مرتب شده و ویژگی هایی با کمترین رتبه حذف می‌شوند (از یک آستانه برای رتبه های ویژگی استفاده می‌شود). سپس نتایج زیرمجموعه‌ای از ویژگی ها به عنوان ورودی به یک سیستم طبقه‌بند داده می شود.

۲,۲,۱. ضریب همبستگی پیرسون

ضریب همبستگی پیرسون، توسط سرکارل پیرسون معرفی شده‌است [16]. این ضریب میزان همبستگی بین دو متغیر فاصله‌ای یا نسبی را محاسبه کرده مقدار آن بین ۱+ و ۱− می باشد اگر مقدار بدست آمده مثبت باشد به معنی این است که تغییرات دو متغیر به طور هم جهت اتفاق می‌افتد و برعکس اگر مقدار r منفی شد یعنی اینکه دو متغیر در جهت عکس هم عمل می کنند. اگر مقدار بدست آمده صفر شد نشان میدهد که هیچ رابطه‌ای بین دو متغیر وجود ندارد و اگر ۱+ شد همبستگی مثبت کامل و اگر ۱− شد همبستگی کامل و منفی است.

$$corr(X,Y) = \frac{\mathrm{cov}(X,Y)}{\delta_X \delta_Y} \qquad (۱)$$

در معادله (۱)، X بیانگر نمونه‌های آموزشی و Y بیانگر برچسب کلاس‌های مربوطه میباشد و δ_X و δ_Y به ترتیب انحراف معیار نمونه‌های آموزشی و برچسب کلاس های مربوطه را نمایش میدهد.

۲,۲,۲. آزمون t (T-test)

یکی دیگر از روش‌های انتخاب ویژگی آزمون t (تی استودنت) می‌باشد. این روش ویژگی ها را براساس معیاری رتبه بندی میکند که نمونه های واقع در یک کلاس دارای کمترین پراکندگی (واریانس) با نمونه های همان کلاس و بیشترین پراکندگی با نمونه‌های سایر کلاس‌ها داشته باشد [17].

$$t = \frac{X - \mu}{\frac{S}{\sqrt{n-1}}} \qquad (۲)$$

در معادله (۲) ، X بیانگر نمونه های آموزشی، μ بیانگر میانگین، S واریانس نمونه‌ها آموزشی و n تعداد ویژگی‌ها می-باشد.

۲,۲,۳. آنتروپی

آنتروپی نیز یکی از روش‌های پرکاربرد انتخاب ویژگی می‌باشد که میزان اطلاعاتی را که بوسیله یک ویژگی بدست می‌آید را در نظر می‌گیرد. ویژگی X در این روش بر ویژگی Y اولویت دارد، اگر اطلاعات بدست آمده از ویژگی X بیشتر از اطلاعاتی باشد، که از ویژگی Y بدست می‌آید [18].

$$\mathrm{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \qquad (۳)$$

در معادله (۳) ، p_i احتمال وقوع ویژگی i ام می‌باشد.

۲,۳. طبقه بندی

در این مرحله از طبقه بندی کننده ماشین بردارهای پشتیبان (SVM) برای طبقه بندی استفاده شده است. این طبقه بند یکی از روش‌های نوین طبقه بندی می باشد که در حالت پایه تنها برای جداکردن دو کلاس از یکدیگر طراحی شده است،

ولی از آن برای طبقه بندی چندکلاسه نیز استفاده می‌کنند. ایده اولیه این روش برای اولین بار در سال ۱۹۷۹ توسط وپنیک (Vapinik) مطرح شده است. اساس کار SVM بر مبنای حداقل سازی خطای ساختاری (SRM: Structural Risk Minimization) بنا شده است [17]. در تقسیم خطی داده ها، خطی برای جداسازی دادگان انتخاب میشود که حاشیه اطمینان بیشتری داشته باشد. برای اینکه ماشین بتواند دادگانی با پیچیدگی بالا را دسته بندی کند باید دادگان توسط کرنل مناسب به فضای با ابعاد بالاتر منتقل شوند. فرض کنید n نمونه آموزشی در یک فضای حقیقی با ابعاد P و دو کلاس داریم می خواهیم صفحه ای را پیدا کنیم، که نقاط کلاس $c_i=1$ را از نقاط کلاس $c_i= -1$ جدا کند، معادله ی این صفحه به صورت $wx-b=0$ نوشته میشود، که w بردار نرمال و عمود بر صفحه است و پارامتر b/w فاصله ی صفحه از مبدا بردار نرمال، را مشخص می‌کند. برای هر نمونه اگر $wxi-b \geq 1$ آنگاه نمونه xi متعلق به کلاس اول خواهد بود، در غیر اینصورت اگر $wxi-b \leq -1$ آنگاه نمونه xi متعلق به کلاس دوم خواهد بود. مقادیر b و w باید به نحوی انتخاب شوند تا شرط ماکزیمم بودن فاصله تامین شود.

تابع کرنل ، تابع وزنی است که در تکنیک‌های پیش بینی غیر پارامتریک استفاده شده و دارای دو شرط است [18]:

$K(u) = K(-u)$ و $\int_{+\infty}^{-\infty} K(u)du = ۱$ برای تمام مقادیر u. اگر دو کلاس به نحوی باشند که نتوان آنها را با یک صفحه جدا کرد، از روش های غیر خطی و تعریف تابع تصمیم گیری و توابع کرنل استفاده می کنیم.

در آموزش SVM کرنل ها و پارامترهای آن نقش مهمی دارند. بنابراین باید به درستی انتخاب شوند تا دقت طبقه بندی بهبود یابد. در این مقاله کارآیی کرنل‌های مختلف در دسته بندی نیز مورد بررسی قرار گرفته است. به همین منظور کرنل-های چند جمله ای و تابع پایه شعاعی با فاکتورهای مختلف مورد استفاده قرار گرفته اند تا بهترین نتایج برای هر ترکیب از بردار ویژگی به دست بیاید.

۳. آزمایش‌ها و نتایج

نمونه های صوتی مورد استفاده در این مطالعه از پایگاه داده زبان آلمانی دانشگاه برلین (Emo-DB) تهیه شده است که این پایگاه داده شامل ۵۰۰ نمونه سیگنال صوتی از ۱۰ نفر هنرپیشه مرد و زن در حالات مختلف طبیعی، ترس، خشم، شادی، کسالت، تنفر و ناراحتی می‌باشد [**13**]. در این مقاله آزمایش‌ها برروی ۱۳ نمونه سیگنال صوتی انتخاب شده از این مجموعه می‌باشد که برای شبیه سازی از نرم افزار MATLAB **2013**a استفاده شده است که شامل ۲ جمله از هریک از حالت‌های ترس، شادی، ناراحتی، تنفر و عصبانیت و ۳ جمله از حالت طبیعی میباشد. در مرحله استخراج ویژگی سیگنال به پنجره‌هایی به طول ۴۰ میلی ثانیه با هم پوشانی ۷۵٪ و فرکانس نمونه‌برداری ۱۶ کیلوهرتز تفکیک شده است و هر سیگنال صوتی را به ۲۵۶ فریم تقسیم نموده ایم و از هر فریم ۷۰ ویژگی را استخراج نمودیم که شامل ویژگی‌های عبور از صفر، فرکانس فرمنت، PLP، ضرایب پیشگویی خطی، ضرایب مل کپستروم، فرکانس پیچ، ضرایب کپستروم جیتر و فیشر می باشد.

در مجموع ۷۵٪ از ماتریس ویژگی استخراج شده برای آموزش و بقیه برای آزمایش استفاده شده اند. هر حالت احساسی بطور جداگانه با حالت طبیعی مقایسه شده است و سپس به کمک روش‌های فیلتر همچون آنتروپی، آزمون t و ضرایب همبستگی پیرسون به رتبه‌بندی ویژگی‌های با ارزش پرداختیم و در نهایت برای تشخیص احساسات نمونه‌های آزمایش از طبقه بند ماشین بردار پشتیبان دودویی با کرنل تابع پایه شعایی استفاده شده است و میزان درستی هر حالت در «جدول ۲» تا «جدول ۶» آمده‌است که میزان درستی با نماد

ACC و تعداد ویژگی های انتخاب ویژگی با NF مشخص شده است.

۱۵	۹۷,۳۱	همبستگی پیرسون
۸	**۱۰۰**	آزمون t
۱۰	۹۹,۴۰	آنتروپی

جدول۲ – نتایج حاصل از طبقه بند ماشین بردار پشتیبان برای حالت طبیعی و ترس

طبیعی-ترس		حالات احساسی / روش های انتخاب ویژگی
NF	ACC (٪)	
۸	**۹۵,۶۵**	همبستگی پیرسون
۱۵	۹۲,۷۵	آزمون t
۱۷	۹۴,۲۰	آنتروپی

جدول۳ – نتایج حاصل از طبقه بند ماشین بردار پشتیبان برای حالت طبیعی و شادی

طبیعی-شادی		حالات احساسی / روش های انتخاب ویژگی
NF	ACC (٪)	
۱۴	۹۷,۵۵	همبستگی پیرسون
۱۳	**۹۹,۰۸**	آزمون t
۱۶	۹۶,۰۲	آنتروپی

.

جدول۴ – نتایج حاصل از طبقه بند ماشین بردار پشتیبان برای حالت طبیعی و ناراحتی

طبیعی-ناراحتی		حالات احساسی / روش های انتخاب ویژگی
NF	ACC (٪)	

جدول۵ – نتایج حاصل از طبقه بند ماشین بردار پشتیبان برای حالت طبیعی و تنفر

طبیعی-تنفر		حالات احساسی / روش های انتخاب ویژگی
NF	ACC (٪)	
۱۳	۹۸,۲۰	همبستگی پیرسون
۱۱	**۱۰۰**	آزمون t
۱۳	۹۶,۷۰	آنتروپی

جدول۶ – نتایج حاصل از طبقه بند ماشین بردار پشتیبان برای حالت طبیعی و عصبانیت

طبیعی-عصبانیت		حالات احساسی / روش های انتخاب ویژگی
NF	ACC (٪)	
۱۲	۹۴,۹۴	همبستگی پیرسون
۱۲	**۹۷,۶۱**	آزمون t
۱۴	۹۴,۹۴	آنتروپی

نتایج بدست آمده از جدول‌های ۲ تا ۶ نشان میدهد که با استفاده از طبقه بند ماشین بردار پشتیبان برای مقایسه حالت ترس از طبیعی روش انتخاب ویژگی همبستگی پیرسون با ۸ ویژگی بالاترین میزان درستی ٪۹۵,۶۵ و برای مقایسه حالت-های شادی، ناراحتی، تنفر و عصبانیت از حالت طبیعی روش

انتخاب ویژگی آزمون t با کمترین تعداد ویژگی بترتیب ۱۳ ،۸ ،۱۱ و ۱۲ بالاترین میزان درستی حاصل شد.

۱. بحث و نتیجه‌گیری

در این مقاله برای تشخیص احساسات از گفتار ابتدا روش‌های استخراج ویژگی اعمال و پس از آن برای داشتن عملکرد بهتر سیستم، روش‌های رتبه بندی انتخاب ویژگی مورد بررسی قرار گرفته و در نهایت میزان درستی هر روش توسط طبقه‌بند ماشین بردار پشتیبان حاصل گردید.

با توجه به نتایج حاصل شده روش آزمون t برای تشخیص حالات احساسی نسبت به سایر روش‌های انتخاب ویژگی شامل همبستگی پیرسون و آنتروپی، دارای عملکرد بهتر (بیشترین میزان درستی و کمترین تعداد ویژگی انتخابی) می-باشد. با توجه به اینکه روش آزمون t ویژگی ها را براساس معیاری رتبه بندی میکند که نمونه های واقع در یک کلاس دارای کمترین پراکندگی (واریانس) با نمونه های همان کلاس و بیشترین پراکندگی با نمونه‌های سایر کلاس‌ها باشد، بنابراین برتری این روش نسبت به سایر روش‌های انتخاب ویژگی دور از ذهن نمی‌باشد.

منابع

[1] O'Shaughnessy, "*Speech communication human and machine*, " Reading: Addison-Wesley,1987.

[2] France, D. J. et. al.,"*Acoustical properties of speech as indicators ofdepression and suicidal risk*," IEEE Trans. Biomedical Engineering, Vol. 47, No. 7, pp. 829-837, 2000.

[3] T. Pao and C. Wang, "*A study on the search of the most discriminative speech features in the speaker dependent speech emotion recognition*", Proc. IEEE Fifth Int. Sym. Parallel Architectures, Algoritm and Programming,157-162, 2012.

[4] Shikha Gupta, Jafreezal Jaafar , "*FEATURE EXTRACTION USING MFCC*", An International Journal (SIPIJ) Vol.4, No.4, August 2013

[5] Makhoul, J. "Linear prediction: A tutorial review". *Proceedings of the IEEE*, Vol.*63*, No.4, pp.561–580, 1975.

[6] Rabiner, L. R., & Juang, B. H. "*Fundamentals of speech recognition*." Englewood Cliffs: Prentice-Hall, 1993.

[7] Banziger, T., & Scherer, K. R, "The role of intonation in emotional expressions. *Speech Communication*, " Vol46, pp.252–267, 2005.

[8] Petrushin, V. "*Emotion in speech: recognition and application to call centres*. *Artificial neural networks in engineering (ANNIE)* ". 1999.

[9] T. L., Chen, Y. T., Yeh, J. H., & Liao, W. Y, "*Combining acoustic features for improved emotion recognition in Mandarin speech.* " In J. Tao, T. Tan, & R. Picard (Eds.), *ACII. LNCS* pp.279–285). Berlin: Springer 2005.

[10] Pao, T. L., Chen, Y. T., Yeh, J. H., & Chien, C. S."*Feature combination for better differentiating anger from neutral in Mandarin emotional speech*. " *LNCS: Vol. 4738*, 2007.

[11] Bozkurt, E., Erzin, E., Erdem, C. E., & Erdem, A. T. "*Improving automatic emotion recognition from speech signals.* " In *10th* annual conference of the international speech communication association (interspeech), Brighton, UK, Sept. 6–10, pp. 324– 327, 2009.

[12] J. Jarmulak; S. Craw, "*Genetic Algorithms for Feature Selection and Weighting*", Appears in Proceedings of the IJCAI'99 workshop on Automating the Construction of Case Based Reasoners, 1999

[13] J. Jarmulak; S. Craw, "*Genetic Algorithms for Feature Selection and Weighting*", Appears in Proceedings of the IJCAI'99 workshop on Automating the Construction of Case Based Reasoners, 1999.

[14] Kabir, Md. M., Islam, Md. M., Murase, K., "*A new wrapper feature selection approach using neural network*", Neurocomputing, pp. 3273-3283, 2010.

[15] Kohavi, R., John, G., "*Wrappers for Feature Subset Selection.* " Artificial Intelligence, pp.273-324, 1997.

[16] K.Pearson "*Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia.*" *Phil Trans R Soc Lond Series A*, pp. 253-318.

[17] Markowski, Carol A.; Markowski, Edward P. (1990). "Conditions for the Effectiveness of a Preliminary Test of Variance". *The American Statistician* **44 (4): 322–326**.

[18] K.Neshatian. and M.Zhang "*Genetic programing for performance improvement and dimensionality reduction of classification problem*" IEEE World Congress on Computational Intelligence(CEC), Hong Kong, 2008.