

SID



ابزارهای
پژوهش



سرویس ترجمه
تخصصی



کارگاه های
آموزشی



بلاگ
مرکز اطلاعات علمی



سامانه ویراستاری
STES



فیلم های
آموزشی

کارگاه های آموزشی مرکز اطلاعات علمی



آموزش مهارت های کاربردی در تدوین و چاپ مقالات ISI

آموزش مهارت های کاربردی
در تدوین و چاپ مقالات ISI



روش تحقیق کمی

روش تحقیق کمی



آموزش نرم افزار Word برای پژوهشگران

آموزش نرم افزار Word
برای پژوهشگران

طبقه بندی سیگنالهای شنیداری با استفاده از شبکه های عصبی همراه با ترکیب بازنمایی ادراکی و کیسترال

محمد علی مرادمند^۱، سید علی سید صالحی^۲، فرشاد الماس گنج^۳

دانشگاه صنعتی امیرکبیر - دانشکده مهندسی پزشکی

m8033276@aut.ac.ir

چکیده

با توجه به اهمیت روزافزون پردازش سیگنال های شنیداری، ضرورت طبقه بندی این سیگنالها در مراحل اولیه و قبل از انجام پردازشهای پیشرفته تر ضروری می باشد. با مشخص شدن نوع سیگنال شنیداری و اینکه آیا سیگنال گفتاری، موسیقی و یا ... است می توان نسبت به نوع پردازش های بعدی که باید روی آن انجام بگیرد تصمیم گیری نمود. در اینجا نیز سعی می شود با توجه به سیستم شنوایی و درک انسان یک الگوریتم مقاوم برای طبقه بندی سیگنالهای به دو طبقه گفتاری و غیر گفتاری ارائه شود. پس از استخراج ویژگیهای شنیداری و تعلیم شبکه عصبی جلوسو بر روی دادگان تعلیم در طولهای متفاوت، سیستم بر روی دادگان آزمون دو طبقه آزمایش و با سایر روشهای طبقه بندی مقایسه می گردد.

واژه های کلیدی: طبقه بندی سیگنالهای شنیداری، شبکه های عصبی، گفتاری، غیر گفتاری

سراغ تجزیه و تحلیل آن می رود. به طور کلی می توان به دلایل زیر برای اهمیت طبقه بندی سیگنالهای شنیداری اشاره نمود:

(۱) انواع مختلف سیگنالهای شنیداری نیاز به پردازشهای مختلفی دارند. به عنوان مثال در سیستمهای باز شناسی گفتار اگر سیگنالهای غیر گفتاری، مورد پردازش قرار گیرند ممکن است موتور باز شناخت دچار اشتباه شود، حال آنکه با طبقه بندی سیگنال شنیداری می توان فقط سیگنال گفتار را در این سیستم وارد نمود.

(۲) در بسیاری از کاربردها نوع سیگنال مهم می باشد.

مقدمه

تحقیق در مورد طبقه بندی و دوباره بدست آوردن تصاویر عمر طولانی دارد. با گسترش روز افزون صوت در اینترنت و دیگر وسایل ارتباطی و شبکه های سوئیچ کننده با فرامین صوتی، این تحقیقات بر روی سیگنالهای شنیداری نیز متمرکز شده است. به بیان دیگر پیشرفتهای اخیر در مورد باز شناخت گفتار و صوت، طبقه بندی این سیگنالها را بسیار مهم جلوه می کند. سیستم شنوایی انسان قابلیت بسیار بالایی در طبقه بندی سیگنالهای شنیداری دارد. بدین ترتیب که با شنیدن هر صدا پس از تعیین نوع آن، به

^۱ - دانشجوی مقطع کارشناسی ارشد بیوالکترونیک، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر

^۲ - استادیار دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر

^۳ - استادیار دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر

گسسته^۴ تقسیم می شوند و در نهایت طبقه بندی بر اساس فاصله کسینوسی و قوانین نزدیکترین همسایگی انجام می گیرد.

یک روش سلسله مراتبی نیز برای طبقه بندی سیگنالهای شنیداری تا جزئی ترین طبقات ارائه شده است [7]. در این روش ابتدا با استفاده از ویژگیهای آماری زمان کوتاه، سیگنالهای شنیداری را در یک سطح کلی به طبقات گفتار، موسیقی و اصوات محیطی تقسیم می کند و سپس در یک سطح جزئی با استفاده از مدل مارکوف پنهان طبقه مربوط به اصوات محیطی را به زیر طبقات جزئی تر تقسیم می کند. در دومقاله دیگر از همین نویسنده [9],[8]، سیگنالهای مخلوط صوتی تصویری فقط بر اساس سیگنالهای صوتی طبقه بندی می شوند که در مقایسه با طبقه بندی سیگنالهای تصویری بسیار راحتتر و مقرون به صرفه تر می باشد.

سیستمی برای طبقه بندی فایلهای صوتی جمع آوری شده در وب ارائه گردید [13]، که این روش سیگنالها را به سه دسته گفتار، موسیقی و غیره طبقه بندی می کند. در یک کار جامع از ویژگیهای سیگنالهای شنیداری در چهار حوزه زمان، فرکانس، زمان-فرکانس و حوزه ضرایب استفاده گردید و با توجه به اهمیت هر ویژگی برای هر طبقه از الگوریتم انتخاب ویژگی ترتیبی (SFS) برای بردار ویژگی هر طبقه استفاده شد. در نهایت برای طبقه بندی روشهای مختلف طبقه بندی بر روی دادگان آزمایش گردید و در بهترین شرایط دقتی حدود ۹۳٪ بدست آمد. [3]

در چند کار اخیر از تبدیل ویولت برای استخراج بردار ویژگی استفاده می شود. [14],[1],[2],[3]. به دلیل رزولوشن زمانی-فرکانسی تبدیل ویولت و نزدیکی به سیستم درک صوت انسان، بردارهای ویژگی از ضرایب در هر زیر باند تبدیل ویولت گسسته استخراج می گردند در اینجا، یک الگوریتم با دقت بالا برای طبقه بندی و

قطعه بندی سیکتالهای شنیداری مطرح می شود، که در آن هدف نهایی طبقه بندی و تمایز بین طبقات گفتار و غیر گفتاری در پنجره های زمانی با طولهای متفاوت از ۳۲

۳ برای سیگنالهای تصویری شنیداری می توان کل سیگنال را فقط بر اساس سیگنال شنیداری طبقه بندی نمود. که در بسیاری از اوقات طبقه بندی سیگنال تصویری صوتی بر حسب صوت بسیار ساده تر از طبقه بندی تصویری می باشد.

۴ فضای جستجو در سیگنال شنیداری کاهش می یابد.

۵ سیستمهای سوئیچ شونده با صوت

در کارهای ارائه شده برای طبقه بندی سیگنالهای شنیداری، بحث بر روی تفاوت مشخصات گفتار، موسیقی و دیگر اصوات می باشد بدین ترتیب که پس از استخراج بردارهای ویژگی برای هر طبقه، با استفاده از روشهای مختلف طبقه بندی، سیگنالهای شنیداری طبقه بندی می گردند. به طور کلی طبقه بندی سیگنالهای شنیداری نیز همانند بازشناسی الگو دارای دو بعد انتخاب ویژگی و طبقه بندی بر اساس ویژگیهای انتخاب شده می باشد. با توجه به مطالب گفته شده یک بازنمایی مؤثر باید بتواند مهمترین خصوصیات اصوات را برای طبقه بندی ارائه کند، به نحوی که تحت شرایط مختلف انعطاف خوبی داشته باشد و توانایی طبقه بندی اصوات مختلف را داشته باشد. بعد از بازنمایی، انتخاب معیار فاصله و قوانین طبقه بندی کننده نکته اساسی دیگر می باشد.

سیستم باز شناخت و طبقه بندی کننده ماسل-فیش توسط آقای ارلینگ وود در سال ۱۹۹۶ میلادی ارائه شد. این کار نسبت به کارهای کوچک قبلی بسیار متمایز و قابل قبول بود. در این روش برای تحلیل و طبقه بندی سیگنالهای شنیداری از چهار ویژگی اکوستیکی صوت استفاده می شود که عبارتند از: پیچ صوتی، دامنه، روشنایی و پهنای باند. با استفاده از یک فاصله اقلیدسی نرمالیزه شده و قوانین طبقه بندی نزدیکترین همسایگی، سیگنالهای شنیداری مختلف به طبقه های مربوطه نسبت داده می شوند [4],[5]

در کار دیگری [6]، از ویژگی ضرایب کپسترال در مقیاس مل^۳ با ساختار درختی استفاده شده است، برای هر نمونه صوت با توجه به فرکانسهای آن، یک هیستوگرام ساخته می شود و از آن بعنوان یک بردار ویژگی استفاده می شود. سپس فضای بردارهای ویژگی، به تعدادی منطقه

^۳ mel frequency cepsteal coefficients

^۴ bin

در حالیکه ویژگیهای ادراکی مثل روشنایی، پهنای باند و انرژی زیر باندها خصوصیات مختلف طیفی صوت را نشان می دهند، اما برخی از ویژگیهای سیگنال از دست می روند. ضرایب کیسترال شکل طیف فرکانسی صوت را ارائه می کنند، که از روی آن می توان اغلب سیگنالهای اصلی را بازسازی نمود، در نتیجه یک مکمل برای ویژگیهای ادراکی می باشند. در توضیح این ویژگیها که در ادامه آمده است، ضرایب تبدیل فوریه کوتاه مدت $F(w)$ تحت فریمهای ۳۲ میلی ثانیه ای محاسبه می شود. شرح ویژگیهای مورد نظر در ادامه آمده است.

ویژگیهای ادراکی

(۱) توان کل طیف برای هر فریم به صورت زیر محاسبه می شود:

$$P = \log \left(\int_0^{w_0} |F(w)|^2 dw \right)$$

که در آن $|F(w)|^2$ توان در فرکانس w و $w_0=4000$ نصف فرکانس نمونه برداری می باشند.

(۲) طیف فرکانسی سیگنالهای شنیداری به چهار زیر باند $[0, w_0/8], [w_0/8, w_0/4], \dots, [w_0/2, w_0]$ تقسیم می شود. از لگاریتم انرژی هر زیر باند به صورت زیر استفاده می شود:

$$P_j = \log \left(\int_{L_j}^{H_j} |F(w)|^2 dw \right)$$

که در آن L_j و H_j مرزهای پائین و بالای زیر باند j ام هستند. بدین ترتیب برای چهار زیر باند، چهار توان بعنوان ویژگی استخراج می گردد.

(۳) روشنایی سیگنالهای شنیداری بعنوان مرکز فرکانسی طیف به صورت زیر تعریف می شود

$$W_c = \frac{\int_0^{w_0} w |F(w)|^2 dw}{\int_0^{w_0} |F(w)|^2 dw}$$

(۴) پهنای باند به صورت زیر بدست می آید

$$B = \sqrt{\frac{\int_0^{w_0} (w - W_c)^2 |F(w)|^2 dw}{\int_0^{w_0} |F(w)|^2 dw}}$$

که حاصل از مجذور توان تفاضل اجزاء طیفی و مرکز فرکانسی می باشد.

میلی ثانیه (یعنی برابر طول یک فریم) الی یک ثانیه می باشد. ویژگیهای مختلفی جهت تمایز بین سیگنالهای شنیداری می توان معرفی و ارائه نمود. برای طبقه بندی علاوه بر استفاده از ویژگیهای متداول از چندین ویژگی جدید نیز استفاده می شود.

در این مقاله یک الگوریتم جدید با دقت بالا برای طبقه بندی سیگنالهای شنیداری ارائه می گردد، که در آن هدف نهایی تمایز بین سیگنالهای گفتاری، و غیر گفتاری در زمانهای مختلف سیگنال (از ۳۲ میلی ثانیه الی ۱ ثانیه) می باشد. نکته مهم انتخاب ویژگیهای ادراکی بعلاوه ویژگیهای کیسترال در مقیاس مل و ترکیب این دو مجموعه ویژگیها برای یک بازنمایی مناسب از سیگنال و استفاده از شبکه عصبی جلوسو بعنوان طبقه بندی کننده می باشد. پس از تعلیم شبکه عصبی بر روی دادگان دو طبقه گفتاری و غیر گفتاری، شبکه های تعلیم دیده برای بازه های زمانی مختلف بر روی دادگان تست آزمایش و با سایر روشهای طبقه بندی کننده آماری مقایسه می شوند.

استخراج ویژگیهای شنیداری

قبل از استخراج ویژگیها، هر سیگنال شنیداری (در حالت PCM هشت بیتی) به نرخ نمونه برداری ۸۰۰۰ نمونه در ثانیه نمونه برداری مجدد می شود. هر فریم ۳۲ میلی ثانیه ای تحت پنجره های زمانی همینگ با همپوشانی ۲۵ درصد بدست می آیند. هر فریم اگر شرط

$$\sum_{i=1}^{256} (w_i s_i)^2 < 400^2$$

سکوت برچسب دهی می شود، که در آن S دامنه سیگنال در i و 400 سطح آستانه سکوت می باشد. ویژگیهای شنیداری تحت هر فریم غیر سکوت استخراج می گردند. متوسط و انحراف معیار مسیر بازنمایی تحت هر clips برای تصمیم گیری نهایی برای طبقه هر نمونه محاسبه می شوند.

دو نوع ویژگی تحت هر فریم استخراج می گردد (۱) ویژگیهای ادراکی، شامل توان کل، توان زیرباندها، روشنایی، پهنای باند و (۲) ضرایب کیسترال در مقیاس فرکانسی مل. نکته مهم انتخاب ویژگیها بر اساس ویژگیهای ادراکی، MFCC's و ترکیب این دو می باشد.

ضرایب کپسترال در مقیاس مل

در روشهای طیفی استخراج پارامترهای بازنمائی بخصوص در روشهایی که از تحلیل فوریه جهت بدست آوردن طیف سیگنال استفاده می شود، عموماً از بانک فیلتر جهت محاسبه انرژی طیف حول فرکانسهایی مشخص بعنوان پارامترهای بازنمائی استفاده می گردد. تعداد این فیلترهای میانگذر در سیستمهای مختلف متفاوت است، ولی معمولاً بین ۱۶ الی ۲۰ فیلتر مورد استفاده قرار می گیرند. افزایش فیلترها معمولاً موجب بهبود کیفیت طبقه بندی می شود ولی در صورت کاهش پهنای باند فیلترها تا حد کمتر از فرکانس واک سیگنال شنیداری، کیفیت طبقه بندی افت می کند. تنظیم فواصل مابین فیلترها به صورت غیر خطی و در مقیاس مل یا بارک که مقیاسهای الهام گرفته از سیستم شنوایی انسان می باشند، انجام می گیرد. روابط تبدیل مقیاس هر تیز به این دو مقیاس عبارتند از:

$$f_{mel} = 2595 \log \left[1 + \frac{f_{HZ}}{700} \right]$$

$$f_{bark} = 6 \ln \left[\frac{f_{HZ}}{600} + \sqrt{\left(\frac{f_{HZ}}{600} \right)^2 + 1} \right]$$

هر دو مقیاس تقریباً شبیه به هم بوده و تا فرکانس یک کیلو هر تیز به صورت تقریباً خطی و بالاتر از این فرکانس به صورت لگاریتمی می باشند. در اینجا از پارامترهای ضرایب کپسترال با مقیاس مل (MFCC) برای بازنمائی استفاده می شود. این ضرایب از روی توان FFT هر فریم محاسبه می شوند. ضرایب توان از فیلتر بانکهایی مثلثی که شامل ۱۹ فیلتر میان گذر مثلثی هستند، گذرانده می شوند. این ساختار فیلتر بانکی توسط بازه هایی با طول ثابت در مقیاس مل، محدوده فرکانسی صفر الی ۴۰۰۰ هر تیز را پوشش می دهند. با در نظر گرفتن خروجی هر فیلتر بانک به صورت Sk ضرایب MFCC به صورت زیر محاسبه می گردند:

$$c_n = \sqrt{\frac{2}{k} \sum_{k=1}^K (\log S_k) \cos[n(k - .05)\pi / k]}$$

$$n=1,2,3,\dots,L$$

که در آن L مرتبه کپستروم می باشد.

هنجارسازی بردارهای ویژگی

برای هر فریم غیر سکوت از هر نمونه صوت، هشت ویژگی ادراکی استخراج می شود که این ویژگیهای ادراکی عبارتند از: توان طیف، توان زیر باندها، روشنایی، پهنای باند و فرکانس پیچ. از این هشت ویژگی روی کل فریمهای غیر سکوت هر نمونه میانگین و انحراف معیار گرفته می شود. در نهایت برای هر نمونه تعلیم (و یا تست) یک بردار ۱۶ بعدی از ویژگیهای ادراکی استخراج می شوند. با افزودن نرخ سکوت (نسبت تعداد فریمهای سکوت به کل فریمهای یک نمونه) و نرخ پیچ (نسبت تعداد فریمهای دارای پیچ به کل فریمهای نمونه) یک بردار ویژگی ۱۶ بعدی از ویژگیهای ادراکی برای هر نمونه بدست می آید. که این ویژگی با "perc" نشان داده می شوند. هر ویژگی xi از اجزاء این بردار perc به صورت زیر نرمالیزه می شوند

$$x'_i = (x_i - \mu_i) / \delta_i$$

که در آن همبستگی بین ویژگیهای متفاوت صرفنظر شده است. و در آن μ_i متوسط و δ_i انحراف معیار کل مجموعه های تعلیم می باشند. در اینجا بردار ویژگی نهایی پس از نرمالیزه شدن به صورت "Perc" نشان داده می شوند. برای ضرایب MFCC's نیز به همین صورت عمل می شود که پس از محاسبه ضرایب MFCC's از درجه L بر روی کل فریمهای غیر سکوت یک نمونه، میانگین و انحراف معیار آنها حساب می شوند. پس یک بردار ویژگی با بعد 2L از روی ضرایب MFCC's بدست می آید که با CepsL نشان داده می شوند. که برای این بردار ویژگی از حالت غیر نرمالیزه استفاده می شود چون در حالت نرمالیزه دقت طبقه بندی کاهش می یابد. به منظور انتخاب مناسب ترین بردار ویژگی از تلفیق ویژگیهای Perc و CepsL استفاده می شود، در نهایت نیز بردار ویژگی PercCeps تشکیل داده می شود. بعد این بردار ویژگی برابر ۱۶ (حاصل از متوسط و انحراف از معیار ویژگیهای ادراکی و دو ویژگی نرخ سکوت و نرخ پیچ) بعلاوه 2L مربوط به متوسط و واریانس L ضرایب MFCC's یعنی 16+2L می باشد. که با توجه به آزمایشات انجام شده در [5] از کپستروم مرتبه ۸ استفاده می شود. در نتیجه یک مجموعه ویژگی ۳۲ بعدی برای یک قطعه یک ثانیه ای بدست می آید. برای ترکیب ویژگیهای ادراکی با ویژگیهای

مختلف برای طبقه غیر گفتاری می باشند. در دادگان تست نیز سعی شده است از انواع مختلف سیگنالهای گفتاری و غیر گفتاری که مشابه آنها در دادگان تعلیم وجود ندارد استفاده شود. در جدول ۱ صحت طبقه بندی دادگان تست دو طبقه گفتاری و غیر گفتاری با استفاده از طبقه بندی کننده شبکه عصبی و در جدول ۲ برای طبقه بندی کننده آماری k_{nn} آمده است. همانطور که ملاحظه می شود نتایج بدست آمده با استفاده از شبکه های عصبی نسبت به روش طبقه بندی آماری در تمام طولهای زمانی سیگنال دارای دقت بالاتری می باشد. مهمترین ویژگی شبکه های عصبی نسبت به سایر روشهای طبقه بندی، سرعت بالای آن در طبقه بندی و همچنین تعمیم نمونه های تعلیم به تنوعات خارج از دادگان تعلیم می باشد.

نتیجه گیری

در این مقاله با استفاده از ساختار شبکه های عصبی و مقایسه آن با یک روش طبقه بندی آماری، سیگنالهای شنیداری به دو طبقه گفتاری و غیر گفتاری طبقه بندی می شوند. برای طبقه بندی یک نمونه سیگنال شنیداری به یکی از این دو طبقه، ابتدا بردار ویژگی برای فریمهای ۳۲ میلی ثانیه ای از سیگنال مورد نظر استخراج می گردد، که این ویژگیها شامل ترکیب ویژگیهای ادراکی و کپستروال می باشند و از ساختار سیستم شنوائی انسان الهام گرفته اند، سپس بر اساس این بردارهای ویژگی برای طولهای مختلف از سیگنال (۳۲ الی ۱۰۰۰ میلی ثانیه) شبکه های عصبی جلوسوی متناظر طراحی و تعلیم داده می شوند. در آزمایش بر روی دادگان تست ملاحظه گردید که شبکه عصبی با قدرت بالا قادر به طبقه بندی سیگنالهای شنیداری به دو طبقه گفتاری و غیر گفتاری می باشد که در مقایسه با یک روش طبقه بندی آماری نیز این مطلب نشان داده شده است.

جدول ۱. صحت طبقه بندی برای دو طبقه گفتاری

و غیر گفتاری با استفاده از شبکه عصبی

| طول بازه سیگنال تحت طبقه بندی | صحت طبقه بندی برای دادگان تست | صحت طبقه بندی برای دادگان تست |
|-------------------------------|-------------------------------|-------------------------------|
| | صحت طبقه بندی برای دادگان تست | صحت طبقه بندی برای دادگان تست |

کپستروم با توجه به عدم نرمالیزه بودن ویژگیهای کپستروم و نرمالزه بودن ویژگیهای ادراکی، هر یک به واریانس دیگر ویژگیها تقسیم می شوند. بدین ترتیب تمام ۱۶ ویژگی ادراکی دارای انحراف معیار ۱ می باشند، بدین ترتیب مجموع انحراف معیار برای کل ویژگیهای ادراکی برابر ۱۶ می باشد ($S1=16*1$). با توجه به غیر نرمالیزه بودن ویژگیهای کپستروم انحراف معیار کل این مجموعه ویژگی برابر $S2 = \sum_{i=1}^{2L} \partial_i$ می باشد که در آن ∂_i انحراف معیار ویژگی i ام از L ویژگی کپستروم می باشد. و در نهایت ترکیب وزن دار این دو ویژگی به صورت زیر بدست می آید.

$$\text{PercCepsL} = (\text{Perc}/s2) \oplus (\text{CepsL}/s1)$$

در نهایت با توجه به مرتبه ۸ ضرایب کپستروم بردار ویژگی نرمالیزه شده ۳۲ بعدی بدست می آید.

طبقه بندی کننده

برای تمایز بین دادگان دو طبقه گفتاری و غیر گفتاری، پس از استخراج بردار ویژگی، از ساختار شبکه های عصبی مصنوعی جلوسو به عنوان سیستم طبقه بندی کننده استفاده می شود. برای طبقه بندی هر طول از سیگنال (۳۲ میلی ثانیه الی ۱ ثانیه) یک شبکه عصبی جلوسو با ۳۲ نورون ورودی، ۲۰ نورون در لایه پنهان و یک نورون خروجی طراحی می گردد. شبکه طراحی شده با استفاده از دادگان تعلیم دو طبقه گفتاری و غیر گفتاری، توسط قواعد پس انتشار خطا تعلیم می بیند، در مرحله تعلیم با توجه به قوانین یادگیری پس انتشار خطا، ملاحظه گردید که تعلیم شبکه به مینیمم های محلی منتهی می گردد، برای رفع این مسئله از تکنیکهای همگرایی شامل انتخاب ضرایب یادگیری و مومنت متغیر به همراه اضافه نمودن نویز رندم به وزنه های شبکه در حال تعلیم استفاده شد و شبکه پس از بیرون آمدن از مینیممهای محلی در مسیر مناسب همگرا گردید. در نهایت پس از تعلیم شبکه عصبی بر روی دادگان دو طبقه گفتاری و غیر گفتاری با طولهای متفاوت، دادگان تست این دو طبقه آزمایش می گردند. دادگان تعلیم شامل حدود ۱۰۰ قطعه یک ثانیه ای متنوع از گفتار چند مرد و زن برای طبقه گفتار و همچنین حدود ۱۵۰ قطعه یک ثانیه ای شامل انواع موسیقی و اصوات محیطی

on database engineering & application, 2001 ,pp.339-345

[4] Erling Wold and et al “Content-Based Classification, Search, and Retrieval of Audio “IEEE Multimedia ,fall 1996 ,pp.27-36

[5] Stan Z. Li “Content-Based Audio Classification and Retrieval using the Nearest Feature Line Method” IEEE Transaction on Speech and Audio Processing vol8,no.5,September 2000 ,page 619-625

[6] Janathan Foote “A Similarity Measure for Automatic Audio classification “IEEE international conference on multimedia and expo, vol.1, pp.452-455, 1997

[7] Tong Zhang and et al “Hierarchical Classification of Audio Data for Archiving and Retrieval” IEEE proceeding on acoustic ,speech and signal processing ,1999 ,vol6 ,pp.3001-3004

[8] Tong Zhang and et al “Classification and retrieval of sound effects in audiovisual data management” IEEE thirty-third Asilomab conference on signal ,system ,and computer ,1999 ,pp 730-734

[9] Tong Zhang and et al “Audio Content Analysis for Online Audiovisual Data Segmentation and Classification “ IEEE Transaction on speech and audio processing vol.9 no.4 May 2001 ,pp.705-710

[10] George Tzanetakis and et al “Audio Analysis using the Discrete Wavelet Transform” EUROMICRO conference proceedings ,25th ,1999 ,vol.2 ,pp.61-69

[11] Lie Lu, Hao jiang and Hong Juang Zhang “A Robust Audio Classification and Segmentation Method” IEEE 2001

[12] Yao Wang , Zhu Liu ,and Jin-cheng Huang “Multimedia Content Analysis : Using both Audio and visual Clues” IEEE signal processing magazine 2000 pp:12-36

[13] Pedro J. Moreno “Using the Fisher Kernel Method for Web Audio Classification “ IEEE International conference on acoustic ,speech

| غیرگفتاری | گفتاری | |
|-----------|---------|--------|
| ٪.۸۸ | ٪. ۹۱ | 32ms |
| ٪. ۸۲ | ٪. ۹۶ | 100ms |
| ٪. ۸۴ | ٪. ۹۸ | 300ms |
| ٪. ۸۷.۶ | ٪. ۹۹ | 500ms |
| ٪.۹۳ | ٪. ۹۹.۵ | 1000ms |

جدول ۲. صحت طبقه بندی برای دو طبقه گفتاری و غیر گفتاری با استفاده از طبقه بندی کننده آماری

| صحت طبقه بندی برای دادگان تست طبقه غیرگفتاری | صحت طبقه بندی برای دادگان تست طبقه گفتاری | طول بازه های سیگنال تحت طبقه بندی K_NN |
|--|---|--|
| ٪. ۶۹.۵ | ٪. ۷۷ | 32ms |
| ٪. ۷۳.۵ | ٪. ۸۰ | 100ms |
| ٪. ۷۵ | ٪. ۸۵ | 300ms |
| ٪. ۷۷ | ٪. ۹۰ | 500ms |
| ٪.۸۳.۵ | ٪. ۹۴ | 1000ms |

فهرست مراجع

[1] Guo Li and Ashfaq A. Khokhar ”Content-based Indexing and Retrieval of Audio Data using Wavelets” IEEE international conference on multimedia and expo. ,2000 ,vol.2 ,pp884-888

[2] S.R Suramanya and A. Youssef “Wavelet Indexing of Audio Data in Audio/Multimedia Databases” IEEE proceeding of the international workshop on multimedia database management ,1998 ,pp869-878

[3] Mingchun Liu and Ch. Wan “A study no Content-based Classification and Retrieval of Audio Database” IEEE international symposium

,and signal processing 2000 ,ICASSP'00 ,vol.4
,pp.2417-2420

[۱۴] محمد علی مرادمنند ، فرشاد الماس گنج “ طبقه بندی سیگنالهای شنیداری با استفاده از ضرایب تبدیل ویولت “ مجموعه مقالات هشتمین کنفرانس بین المللی کامپیوتر ایران (CSICC'2003)، صفحه ۵۲۱-۵۱۶ ، دانشگاه فردوسی مشهد

، اسفند ۱۳۸۱

Archive of SID

SID



ابزارهای
پژوهش



سرویس ترجمه
تخصصی



کارگاه های
آموزشی



بلاگ
مرکز اطلاعات علمی



سامانه ویراستاری
STES



فیلم های
آموزشی

کارگاه های آموزشی مرکز اطلاعات علمی



تازه های آموزش
آموزش مهارت های کاربردی در تدوین و چاپ مقالات ISI

آموزش مهارت های کاربردی
در تدوین و چاپ مقالات ISI



تازه های آموزش
روش تحقیق کمی

روش تحقیق کمی



تازه های آموزش
آموزش نرم افزار Word برای پژوهشگران

آموزش نرم افزار Word
برای پژوهشگران