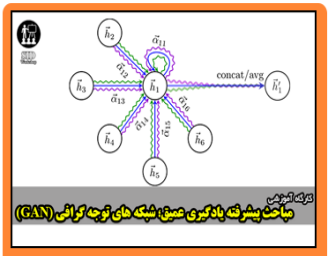
کارگاه های آموزشی مرکز اطلاعات علمی جهاد دانشگاهی

کارگاه آنلاین مکالمه روزمره انگلیسی

کارگاه آنلاین آموزش استفاده از وب آو ساینس



مباحث پیشرفته یادگیری عمیق؛
شبکه های توجه گرافی
(Graph Attention Networks)

طبقه بندی سیگنالهای شنیداری با استفاده از نزدیکترین خط ویژگی و مقایسهٔ آن با سایر روشهای طبقه بندی آماری

محمد علی مرادمند [1] دانشجوی کارشناسی ارشد مهندسی پزشکی- بیوالکتریک ، محمد حسن مرادی [2] و فرشاد الماس گنج [3]

استادیار ، دانشکدهٔ مهندسی پزشکی دانشگاه صنعتی امیر کبیر

[1] m8033276@aut.ac.ir [2] mhmoradi@aut.ac.ir [3] almas@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 یک الگوریتم مقاوم برای طبقه بندی سیگنالهای شنیداری ارائه شود ، به طوری که قادر به طبقه بندی و قطعه بندی هر جریان شنیداری [1] به دو طبقهٔ گفتاری و غیر گفتاری باشد. بعد از استخراج ویژگیهای زمان کوتاه ، روشهای مختلف طبقه بندی آماری بر روی طولهای متفاوت از دادگان این دو طبقه آزمایش می شوند.

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

1) 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 بازشناخت دچار اشتباه شود ، حال آنکه با طبقه بندی سیگنال شنیداری می توان فقط سیگنال گفتار را در این سیستم وارد نمود.

2) در بسیاری از کاربردها نوع سیگنال مهم می باشد .

3) 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 تصویر می باشد .

4) فضای جستجو در سیگنال شنیداری کاهش می یابد .

5) 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

[1] audio stream

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 بازنمائی 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 بازنمائی، انتخاب معیار فاصله و قوانین طبقه بندی کننده نکتهٔ اساسی دیگر می باشد .

سیستم باز شناخت و طبقه بندی کنندهٔ ماسل-فیش توسط آقای ارلینگ وود در سال 1996 میلادی ارائهٔ 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 فاصلهٔ اقلیدسی نرمالیزه شده و قوانین طبقه بندی نزدیکترین همسایگی ، سیگنالهای شنیداری مختلف به طبقه های مربوطه نسبت داده می شوند [5],[4]

در کار دیگری [6] ، از ویژگی ضرایب کپسترال در مقیاس مل[3] 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 منطقهٔ گسسته[4] تقسیم می شوند و در نهایت طبقه بندی بر اساس فاصلهٔ 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 طبقهٔ مربوط به اصوات محیطی را به زیر طبقات جزئی تر تقسیم می کند. در دومقالهٔ 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 ارئه گردید [13] ،که این روش سیگنالها را به سه دستهٔ گفتار ، موسیقی و غیره طبقه بندی می کند. در یک کار جامع از ویژگیهای سیگنالهای شنیداری در چهار حوزهٔ زمان ، فرکانس ، زمان-فرکانس و حوزهٔ 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 ٪93 بدست آمد . [3]

در چند کار اخیر از تبدیل ویولت برای استخراج بردار ویژگی استفاده می شود . [14], [3],[2],[1] . به دلیل رزولوشن زمانی-فرکانسی تبدیل ویولت و نزدیکی به سیستم درک صوت انسان ، بردارهای ویژگی از ضرایب در هرزیر 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 32 میلی ثانیه(یعنی برابر طول یک فریم) الی یک ثانیه می باشد . ویژگیهای مختلفی جهت تمائز 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. به طور معمول ویژگیهای شنیداری در دو سطح استخراج می شوند . در سطح فریم کوتاه مدت و در سطح CLIP بلند مدت . فریم عبارت است از یک سری نمونه های پشت سر هم که دارای طول بین 20 الی 40 میلی ثانیه

---

3 mel frequency cepsteal coefficients
4 bin

هستند . که در این سطح با فرض ایستان بودن سیگنال می شود ویژگیهای مورد نظر از قبیل تبدیل فوریه ، بلندی و... را از روی آن فریم استخراج نمود . در حالیکه برای رسیدن به یک مفهوم معنایی و قابل تمائز از روی یک ویژگی برای سیگنالهای شنیداری نیاز به تحلیل روی یک مدت زمان طولانی تر از سیگنال می باشد . که طول این پنجره می تواند بین چند میلی ثانیه (حداقل طول یک فریم) تا چند ثانیه باشد ، به این بازهٔ زمانی در اصطلاح CLIPS گفته می شود . یک CLIPS شامل چندین فریم پشت سر هم و دارای همپوشانی می باشد .

پس از بازنمائی سیگنالهای شنیداری طبقه بندی کننده های آماری مختلف(از جمله k-nn ، k-nc ، k-nfl و nfl) در فضای بردارهای ویژگی بر روی طولهای مختلف ازدادگان تعلیم وتست آزمایش می گردند. در روش نزدیکترین خط ویژگی از اطلاعات مربوط به نقاط ویژگی هر زوج نمونهٔ تعلیم استفاده می شود ، علیرغم روشهای طبقه بندی آماری دیگر که از اطلاعات هر تک نمونهٔ تعلیم استفاده می کنند، در آزمایش دادگان تست دو طبقهٔ گفتاری و غیر گفتاری، نیز برتری روش نزدیکترین خط ویژگی نسبت به سایر روشهای طبقه بندی آماری نشان داده می شوند.

2. استخراج ویژگیهای شنیداری

قبل از استخراج ویژگیها ، هر سیگنال شنیداری (در حالت PCM هشت بیتی) به نرخ نمونه برداری 8000 نمونه در ثانیه نمونه برداری مجدد می شود . هر فریم 32 میلی ثانیه ای تحت پنجره های زمانی همینگ با همپوشانی 25 در صد بدست می آیند . هر فریم اگرشرط $\sum_{i=1}^{256}(w_i s_i)^2 < 400^2$ برقرار شود ، بعنواان فریم سکوت برچسب دهی می شود ، که در آن S دامنهٔ سیگنال در i و 400سطح آستانهٔ سکوت می باشد. ویژگیهای شنیداری تحت هر فریم غیر سکوت استخراج می گردند .

متوسط و انحراف معیار مسیر بازنمائی تحت هر clips برای تصمیم گیری نهایی برای طبقهٔ هر نمونه محاسبه می شوند .

دو نوع ویژگی تحت هر فریم استخراج می گردد 1) ویژگیهای ادراکی ، شامل توان کل ، توان زیرباندها ، روشنائی ، پهنای باند و 2) ضرایب کپسترال در مقیاس فرکانسی مل . نکته مهم انتخاب ویژگیها بر اساس ویژگیهای ادراکی ، MFCC's و ترکیب این دو می باشد . در حالیکه ویژگیهای ادراکی مثل روشنائی ، پهنای باند و انرژی زیر باندها خصوصیات مختلف طیفی صوت را نشان می دهند ، اما برخی از ویژگیهای سیگنال از دست می روند. ضرایب کپسترال 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 فوریهٔ کوتاه مدت F(w) تحت فریمهای 32 میلی ثانیه ای محاسبه می شود .شرح ویژگیهای مورد نظر درادامه آمده است.

2-1 ویژگیهای ادراکی

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

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

که در آن $|F(w)|^2$ توان در فرکانس w و w0=4000 نصف فرکانس نمونه برداری می باشند .

2) طیف فرکانسی سیگنالهای شنیداری به چهار زیر باند $[0,\omega_0/8],[\omega_0/8,\omega_0/4],...,[\omega_0/2,\omega_0]$ تقسیم می شود . از لگاریتم انرژی هر زیر باند به صورت زیر استفاده می شود :

$$P_j = \log(\int_{L_j}^{H_j} |F(\omega)|^2 d\omega)$$

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

3) روشنائی سیگنالهای شنیداری بعنوان مرکز فرکانسی طیف به صورت زیر تعریف می شود

$$W_c = \frac{\int_0^{\omega_0} \omega |F(\omega)|^2 d\omega}{\int_0^{\omega_0} |F(\omega)|^2 d\omega}$$

4) پهنای باند به صورت زیر بدست می آید

$$B = \sqrt{\frac{\int_0^{\omega_0} (\omega - W_c)\ |F(\omega)|^2 d\omega}{\int_0^{\omega_0} |F(\omega)|^2 d\omega}}$$

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

2-2 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 محاسبهٔ انرژی طیف حول فرکانسهایی مشخص بعنوان پارامترهای بازنمایی استفاده می گردد . تعداد این فیلترهای میانگذر در سیتمهای مختلف متفاوت است ، ولی معمولاً بین 16 الی 20 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 هرتز به این دو مقیاس عبارتند از :

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

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

هر دو مقیاس تقریباً شبیه به هم بوده و تا فرکانس یک کیلو هرتز 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 بانکهای مثلثی که شامل 19 فیلتر میان گذر مثلثی هستند ، گذرانده می شوند . این ساختار فیلتر بانکی توسط بازه هایی با طول ثابت در مقیاس مل ، محدودهٔ فرکانسی صفر الی 4000 هرتز 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,…,L

که در آن L مرتبهٔ کپستروم می باشد .

3-2 هنجارسازی بردارهای ویژگی

برای هر فریم غیر سکوت از هر نمونهٔ صوت ، هشت ویژگی ادراکی استخراج می شود که این ویژگیهای ادراکی عبارتند از : توان طیف ، توان زیر باندها ، روشنائی ، 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 نمونهٔ تعلیم (و یا تست) یک بردار 16 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 16 بعدی از ویژگیهای ادراکی برای هر نمونه بدست می آید . که این ویژگیه 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 محاسبهٔ ضرایب MFCC's از درجهٔ 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 16 (حاصل از متوسط و انحراف از معیار ویژگیهای ادراکی و دو ویژگی نرخ سکوت و نرخ پیچ) بعلاوهٔ 2L مربوط به متوسط و واریانس L ضریب MFCC's یعنی 16+2L می باشد . که با توجه به آزمایشات انجام شده در [5] از کپستروم مرتبهٔ 8 استفاده می شود . در نتیجه یک مجموعهٔ ویژگی 32 بعدی برای یک قطعهٔ یک ثانیه ای بدست می آید . برای ترکیب ویژگیهای ادراکی با ویژگیهای 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 16 ویژگی ادراکی دارای انحراف معیار 1 می باشند ، بدین ترتیب مجموع انحراف معیار برای کل ویژگیهای ادراکی برابر 16 می باشد (S1=16*1) . با توجه به غیر نرمالیزه بودن ویژگیهای کپستروم انحراف معیار کل این مجموعهٔ ویژگی برابر $S2 = \sum_{i=1}^{2L} \partial_i$ می باشد که در آن ∂_i انحراف معیار ویژگی iام از L ویژگی کپستروم می باشد . و در نهایت ترکیب وزن دار این دو ویژگی به صورت زیر بدست می آ یند .

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

در نهایت با توجه به مرتبهٔ 8 ضرایب کپستروم بردار ویژگی نرمالیزه شدهٔ 32 بعدی بدست می آید .

3. طبقه بندی کننده های آماری

همانطور که گفته شد ، طبقه بندی سیگنالهای شنیداری نیز همانند بازشناسی الگو دارای دو بعد انتخاب ویژگی و طبقه بندی بر اساس ویژگیهای انتخاب شده می باشد. یک بازنمائی مؤثر باید قادر به ارائهٔ خصوصیات متمایز سیگنالهای شنیداری باشد ، به نحوی که تحت شرایط مختلف انعطاف خوبی داشته باشند و توانایی طبقه بندی اصوات به طبقات مختلف را داشته باشد. بعد از انتخاب بردار ویژگی ، مسئله بر سر روشهای طبقه بندی مبتنی براین ویژگیها می باشد . بنابراین دو مطلب انتخاب ویژگیها و طبقه بندی بر اساس این ویژگیها به طور توأم در طراحی سیستم طبقه بندی کننده دخیل می باشند . در ادامه چند روش طبقه بندی کنندهٔ آماری مورد بررسی قرار می گیرند و عملکردآنها بر روی طبقه بندی سیگنالهای شنیداری با هم مقایسه می شوند .[5] , [4] , [1]

3-1 طبقه بندی به روشهای k-NN و k-NC

این روشها تغییر یافتهٔ روش نزدیکترین همسایگی می باشد [1]. K نمونه از یک مجموعهٔ تعلیم که نزدیکترین فاصله را به بردار ویژگی P دارند ، مشخص می گردند . در روش NC مرکزهر طبقه بعنوان نقطهٔ نشان دهندهٔ آن طبقه مطرح می شود ، بدین ترتیب فاصلهٔ بردار ویژگی نمونهٔ تست تا مرکز هر طبقه محاسبه و بعنوان معیاری برای طبقه بندی بکار برده می شود . برای روش **k-NC** در ابتدا نزدیکترین همسایهٔ K از هر کلاس را بجای کل مجموعهٔ تعلیم در نظر می گیریم . در روش K-NC برای هر نمونهٔ تست ، فاصلهٔ بردار ویژگی آن نمونه تا مرکز این K نمونه بعنوان معیاری برای طبقه بندی نمونه تست بکار برده می شود. در روش K-NN فاصلهٔ نمونهٔ تست تا تمام نمونه های کل دادگان تعلیم محاسبه و سپس k نمونه از کل دادگان را که نسبت به نمونهٔ تست کمترین فاصله را دارند انتخاب می شوند ، در میان این k نمونه تعداد نمونه های هر طبقه که حائز اکثریت باشد ، نمونهٔ تست به آن طبقه تعلق دارد .

2-3 روش طبقه بندی نزدیکترین خط ویژگی[1]

یک روش جدید طبقه بندی و بازیابی الگو، طبقه بندی مبتنی بر نزدیکترین خط ویژگی (NFL) می باشد. یک فرض اولیه و معقول برای روش NFL این است که مجموعهٔ تعلیم اصوات موجود شامل حداقل بیش از دو نقطهٔ ویژگی (نمونهٔ تعلیم) برای هر طبقه باشد. روش NFL از اطلاعات بدست آمده توسط چند مجموعهٔ تعلیم برای هر طبقه استفاده می کند. در مقایسه با روشهای NN که برای هر نمونهٔ تست با تمام مجموعه های تعلیم به طور جداگانه محاسبه می شوند، در اینجا فاصلهٔ نمونهٔ تست با خط واصل بین دو نقطهٔ ویژگی (مجموعهٔ تعلیم)حساب می شود.

در روش NFL بین کل زوج نمونه های تعلیم از هر طبقه خطوط ویژگی رسم می شوند ، این خطوط نمایانگر هر طبقه می باشند و تغییرات اطلاعات شنیداری بین نمونه های تعلیم در هر طبقه را پوشش می دهند . بدین ترتیب ظرفیت دادگان تعلیم هر طبقه افزایش می یابد . طبقه بندی بر اساس کمترین فاصله برای نمونهٔ تست تا خطوط ویژگی هر طبقه انجام می گیرد . به طور خلاصه می توان مراحل زیر را برای این روش بیان نمود : در ابتدا یک صوت به یک نقطه ویژگی (بردار ویژگی) نسبت داده می شود . از تغییرات یک صوت به طور پیوسته در فضای ویژگی یک منحنی ایجاد می شود . و این منحنی که ناشی از تغییرات بین نمونه های تعلیم مختلف یک طبقه می باشد ، تشکیل یک زیر فضا از فضای ویژگیهای آن طبقه را می دهد . پس یک نمونهٔ تست باید به این زیر فضا نزدیک باشد نه اینکه صرفاً به نقاط ویژگی آن طبقه که محدود هستند نزدیک باشد .

در روش NFL ، هر جفت از نقاط ویژگی مربوط به یک طبقه در فضای ویژگیها با یک مدل خطی درون و برون یابی می شوند و این توسط یک خط واصل بین دو نقطهٔ ویژگی حاصل می شود . این خط واصل بین دو نقطهٔ ویژگی تشکیل یک خط ویژگی را می دهند . خط ویژگی ، اطلاعاتی مربوط به تغییرات بین دو صوت را تهیه می کند . پس با یک خط ویژگی که توسط دو نقطهٔ ویژگی تشکیل می شود ، می توان چندین نقطهٔ ویژگی مربوط به آن طبقه را ایجاد نمود

4. آزمایش

هدف از انجام آزمایشات متعدد مقایسهٔ روشهای مختلف طبقه بندی شامل , NFL , K-NFL , k-NN K-NC بر روی دادگان دو طبقهٔ گفتاری و غیر گفتاری با استفاده از مجموعهٔ ویژگیها شامل تلفیق ویژگیهای ادراکی و ویژگیهای MFFC's می باشد . آزمایشات برای هر روش بر روی طولهای متفاوت از دادگان تکرار می گردد . روش k-NN یک سری قوانین تصمیم فقط برای طبقه بندی می باشد ، در حالیکه NC می تواند برای هر دو منظور طبقه بندی و بازیابی بکار رود. در NC ، یک طبقه بوسیلهٔ مرکز مجموعه های تعلیم مربوط به آن طبقه معرفی می گردد . اما در k-nc بجای مرکز کل نمونه های تعلیم هر طبقه ، برای طبقه بندی نمونهٔ تست جدید از مرکز k نمونهٔ تعلیم که نزدیکترین فاصله به نمونهٔ تست را دارند استفاده می گردد . k-NN و k-nc از این لحاظ که از اطلاعات مربوط به چند مجموعهٔ تعلیم در هر طبقه استفاده می کنند مشابه NFL می باشند. در k-nfl نیز از ترکیب k-nn و nfl استفاده می شود ، بدین ترتیب که از مجموعه خطوط واصل بین k نمونه از دادگان تعلیم هر طبقه که کمترین فاصله را به نمونهٔ تست دارند ، بعنوان نمایندهٔ هر طبقه استفاده می شود .

روشهای طبقه بندی مختلف آماری برای طبقه بندی دو طبقهٔ کلی گفتار و غیر گفتار با استفاده از دادگان تعلیم روی دادگان تست آزمایش گردید . دادگان تعلیم شامل حدود 100 قطعهٔ یک ثانیه ای متنوع از گفتار چند مرد و زن برای طبقهٔ گفتار و همچنین حدود 150 قطعهٔ یک ثانیه ای شامل انواع موسیقی و اصوات محیطی مختلف برای طبقهٔ غیر گفتاری می باشند . همانطور که در جدول 1 نتایج حاصل از آزمایش روشهای مختلف طبقه بندی بر روی طولهای متفاوت سیگنالهای شنیداری ملاحظه می گردد ، برای

[1] nearest feature line

طولهای کوچکتر از سیگنال دقت طبقه بندی کاهش می یابد و سیستم برای طول بزرگتر از سیگنال طبقه بندی را با دقت بسیار بالایی انجام می دهد . در مجموع می توان نسبت به برتری روش n-fl در طبقه بندی دادگان گفتاری و غیر گفتاری پی برد .

جدول 1. صحت طبقه بندی برای دو طبقهٔ گفتاری و غیر گفتاری در حالات مختلف

نوع طبقه بندی و طول بازه های سیگنال تحت طبقه بندی	صحت طبقه بندی برای دادگان تست طبقهٔ گفتاری	صحت طبقه بندی برای دادگان تست طبقهٔ غیرگفتاری
K-NN -1 برای طول 32ms	٪ 77	٪ 69.5
K-NN برای طول 100ms	٪ 80	٪ 73.5
K-NN برای طول 300ms	٪ 85	٪ 75
K-NN برای طول 500ms	٪ 90	٪ 77
K-NN برای طول 1000ms	٪ 94	٪83
K-NC -2 برای طول 32ms	٪ 59.2	٪ 80
K-NC برای طول 100ms	٪ 70	٪ 86
K-NC برای طول 300ms	٪ 77.5	٪ 89
K-NC برای طول 500ms	٪ 81	٪ 90
K-NC برای طول 1000ms	٪ 83	٪ 92
K-NFL -3 برای طول 32ms	٪ 60	٪ 83
K-NFL برای طول 100ms	٪ 78	٪ 87
K-NFL برای طول 300ms	٪ 85	٪ 89
K-NFL برای طول 500ms	٪ 87	٪ 94
K-NFL برای طول 1000ms	٪ 88	٪95.5
NFL -4 برای طول 32ms	٪ 78	٪86
NFL برای طول 100ms	٪ 80	٪88
NFL برای طول 300ms	٪ 84	٪ 89
NFL برای طول 500ms	٪ 89	٪ 92
NFL برای طول 1000ms	٪ 93	٪ 98

5. نتیجه گیری

در این مقاله روشهای مختلف طبقه بندی آماری بر روی دو طبقهٔ سیگنالهای شنیداری گفتاری و غیر گفتاری آزمایش شدند . برای طبقه بندی یک نمونه سیگنال شنیداری به یکی ازاین دو طبقه ، ابتدا بردار ویژگی برای فریمهای 32 میلی ثانیه ای از سیگنال مورد نظر استخراج می گردد ، که این ویژگیها شامل ترکیب ویژگیهای ادراکی و کپسترال می باشند ، سپس بر اساس این بردارهای ویژگی بر روی طولهای مختلف از سیگنال (از 32میلی ثانیه الی 1 ثانیه) طبقه بندی کننده های مختلف آماری از جمله k-nc ، k-nn ، k-nfl و nfl آزمایش و با هم در طولهای مختلف از سیگنال مقایسه گردیدند . ملاحظه شد در مجموع روش nfl قادر به طبقه بندی دقیقتر دو طبقهٔ گفتار و غیر گفتار در طولهای مختلف از سیگنال می باشد .

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 Retrievalof Audio Database“IEEE international symposium 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

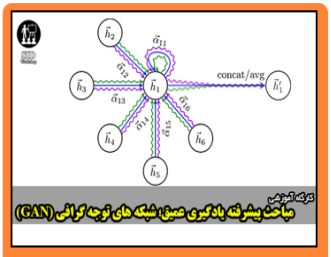
[13] Pedro J. Moreno “Using the Fisher Kernel Method for Web Audio Classification “ IEEE International conference on acoustic ,speech ,and signal proceesing 2000 ,ICASSP’00 ,vol.4 ,pp.2417-2420

[14] محمد علی مرادمند ، فرشاد الماس گنج “ طبقه بندی سیگنالهای شنیداری با استفاده از ضرایب تبدیل ویولت “ مجموعه مقالات هشتمین کنفرانس بین المللی کامپیوتر ایران (CSICC’2003)، صفحهٔ 521-516 ، دانشگاه فردوسی مشهد ، اسفند 1381

کارگاه های آموزشی مرکز اطلاعات علمی جهاد دانشگاهی

کارگاه آنلاین مکالمه روزمره انگلیسی

کارگاه آنلاین آموزش استفاده از وب آو ساینس



مباحث پیشرفته یادگیری عمیق؛ شبکه های توجه گرافی (Graph Attention Networks)