جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فصل اول مقدمه:

صحت و گفتار نقش اساسی در ارتباط انسانها ایفا می‌کنند و یکی از دلایل پیشرفت انسانها است.

برای برقراری ارتباط کامپیوتر با انسان بوسیلهٔ گفتار در کار لازم است انجام شود. یکی سنتزل گفتار است. و دیگری بازشناسی گفتار، سنتز گفتار بیان گفتار بوسیلهٔ کامپیوتر می‌باشد و بازشناسی یعنی فهمیدن گفتار در بازشناسی گفتار. هدف بدست آوردن دنبالهٔ آوایی یک گفتار می‌باشد و این دنبالهٔ آوایی می‌تواند بر اساس واج، سیلاب، کلمه، جمله و ... باشد. بازشناسی گفتار عکس عمل سنتز است و گفتار را به متن تبدیل می‌کند. اما انجام بازشناسی گفتار به دلیل خاصیت صدای انسانها، دارای پیچیدگی‌های زیادی است. اما اغلب بازشناسی کامل و درست غیر ممکن است. حتی خود انسانها هم نمی‌توانند به طور کامل همه صداها را بفهمند و میزان، درک گوش انسانها حدود 70٪ می‌باشد. شکل 1-1 ارتباط گفتاری بین انسانها و کامپیوتر را نشان می‌دهد. به دلیل نقش مهم و کاربردهای فراوانی که بازشناسی

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

گفتار دارد، تحقیقات و مقاله‌های زیادی در این زمینه انجام شده و راه حل‌های متفاوتی پیشنهاد شده است، ولی بازشناسی گفتار کاملاً درست هنوز امکان‌پذیر نمی‌باشد.

بازشناسی گفتار دارای کاربردهای زیادی است. از جمله کاربردهای بازشناسی گفتار، حل مشکل تایپ است، با کمک بازشناسی گفتار می‌توان جمله‌ها را یکی پس از دیگری خواند و کامپیوتر آنها را تایپ کند. یکی دیگر از کاربردهای بازشناسی گفتار، حل مشکل صحبت دو فرد مختلف همزبان است. یکی از مشکلات انسانها ارتباط با افرادی است که با زبانهای متفاوت صحبت می‌کنند. ارتباط بدون دانستن زبان مشکل است. و یادگیری یک زبان دیگر کار وقت گیر و پر زحمتی است ولی به کمک بازشناسی گفتار به یادگیری زبانهای مختلف احتیاجی نخواهد بود و می‌توان با یک دستگاه کوچک با فردی که با زبان دیگری صحبت می‌کند، صحبت نمود. یک کامپیوتر کوچک صدای شما را گرفته و به تعدادی از کلمات تبدیل می‌نماید. سپس این کلمات به زبان دیگر ترجمه شده و در نهایت با زبان جدید گفته

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

می‌شوند. دو مرحلهٔ آخر این سیستم جزو مسایل انجام شده گفتار هستند و با کامل نمودن مسئله بازشناسی گفتار بدون دانستن زبان‌های دیگر به آنها تکلم نمود.

شکل 1-2 نشان دهنده ارتباط دو فرد با زبان‌های مختلف است. یکی از کاربردهای دیگر بازشناسی گفتار، برقراری ارتباط با کامپیوتر است. همان گونه که به انسانهای دیگر دسترس می‌دهید، به کامپیوتر هم می‌توان دستور داد و با آن صحبت کرد. یا حتی می‌توانید از او بخواهید کاری برایتان انجام دهد.

حتی با کمک بازشناسی گفتار می‌توان به انسانهای نابینا و ناشنوا کمک کرد. به طور مثال نابینایان می‌توانند با صحبت کردن و دادن دستور به کامپیوتر با آن کار کنند.

از دستاوردهای جدید بازشناسی گفتار و پردازش مدت کاربرد آن در آموزش‌های زبان دوم می‌باشد. بدین ترتیب که با ایجاد سیستمی‌که قابلیت آشکارسازی خطای تلفظ بین لهجه‌های زبان

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اصلی و لهجهٔ یک فردی که به زبان دوم فرد سخن می‌گوید، وی را در یادگیری و تصحیح تلفظ و لهجه آن زبان کمک نمی‌کنند.

بازشناسی گفتار می‌تواند برای شرایط مختلفی انجام گیرد. هر کدام از این شرایط می‌توانند باعث مشکل شدن، پیچیدگی بازشناسی شوند. یکی از این شرایط، وابسته بودن بازشناسی به یک گوینده یا مستقل بودن آن از گوینده است. مستقل بودن از گوینده به معنای آن است که بتوان کار بازشناسی را برای هر فرد انجام داد. از دیگر شرایط بازشناسی گسسته یا پیوسته بودن گفتار است. راحتی بازشناسی گفتار گسسته، داشتن ابتدا و انتهای عصر کلمه یا اساساً خود کلمه یا همان واحد آوایی می‌باشد. همچنین از دیگر شرایطی که در بازشناسی مطرح است،تعداد واژگان می‌باشد. یعنی بازشناسی گفتار برای چه تعداد کلمه‌ای صورت می‌پذیرد.

هدف از انجام پروژهٔ فوق در ابتدا بازشناسی کلمات گسسته قرانی و در مرحلهٔ دوم ارزیابی نحوهٔ بیان و تلفظ کلمات قرآنی می‌باشد. از آنجائیکه برای مقایسه بین کلمهٔ ادا شده توسط کاربر و صدای استاد

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

باید یک سیستم بهینه وجود داشته باشد. در مرحلهٔ اول سعی می‌کنیم، سیستم را به حالت بهینه خود برسانیم و سپس پارامترهای این سیستم جهت انجام مرحلهٔ دوم استفاده کنیم.

اما چون در هنگام ارزیابی نحوهٔ بیان کلمه قرآنی، کلمه مورد نظر از قبل مشخص است، بنابراین در مرحله دوم احتیاجی به بازشناسی گفتار نمی‌باشد.

در بخش اول برای بهتر درک کردن مفهوم بازشناسی به بررسی سیستم تولید صوت و شنوایی انسان می‌پردازیم. سپس وارد مفاهیم بازشناسی گفتار خواهیم شد. در این مرحله روشهای جداسازی سیگنال زمینه از روی سیگنال صحبت مورد بررسی قرار می‌گیرد. سپس نحوهٔ استخراج ماتری ضرائب کپستروم و در نهایت بازشناسی گفتار بوسیلهٔ الگوریتم انحراف زمانی پویا (DTU) و مدل مخفی مارکوف مورد بررسی قرار می‌گیرد.

پس از آشنایی با ابزارهای بازشناسی گفتار، نحوهٔ پیاده سازی الگوریتم‌های فوق ذکر خواهد شد. بعد از راه‌اندازی سیستم

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بازشناسی گفتار کلمات مقطع، بوسیلهٔ الگوریتم DTN مشاهده شد نرخ بازشناسی گفتار پائینی است و حدود ٪47 می‌باشد. از این رو در جهت بهبود پارامترهای سیستم و بهینه کردن آن در مراحل بازشناسی و پارامترهای آن تغییراتی داده شد، که به ذکر آنها پرداخته خواهد شد.

پس از بهینه کردن پارامترهای سیستم بازشناسی گفتار و رسانیدن نرخ بازشناسی گفتار به ٪99 برای 20 کلمه قرآنی الگوریتم‌های ارزیابی نحوهٔ بیان بوسیلهٔ روش DTA بحث خواهد شد.

در بخش انتهایی به بررسی مدل مخفی مارکوف خواهیم پرداخت. سپس مراحل پیاده سازی الگوریتم فوق بوسیلهٔ نرم افزار و نکات عملی آن گفته خواهد شد. در نهایت سیستم بازشناسی گفتار کلمات مقطع قرآنی و نحوهٔ پیاده سازی آن مورد بررسی قرار خواهد گرفت و در مرحلهٔ بعدی الگوریتم ارزیابی نحوهٔ بیان بوسیلهٔ ذکر خواهد شد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تغییر محیط اکوستیکی روی نتیجه بازشناسی اثر خواهد گذاشت. از آنجائیکه سیستم فوق برای نمونه‌های آزمایشگاهی آموزش داده شده با تغییر محیط اکوستیکی مطمئناً نتایج بازشناسی تغییر خواهد کرد و نمونه‌های واقعی دارای نوین میکروفن، محیط و همچنین برگشت صدا خواهند بود.

در انتها به بررسی سیستم‌های بهبود گفتار خواهیم پرداخت، هدف از این بخش حذف هزینه ورودی از طریق میکروفن و از بین بردن تأثیرهای محیط بر روی سیگنال صدا می‌باشد.

در این بخش به دو روش اشاره خواهیم: ابتدا روش spectral subtraction که به میزان یک روش عمدی برای حذف نویز می‌رود ذکر خواهد شد.

سپس به معرفی یک الگوریتم جدید در حذف نویز میکروفن خواهیم پرداخت.

مدل اعضای صوتی انسان

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در شکل (1-2) یک دیاگرام شماتیک از مکانیزم تولید صحبت انسان نشان داده شده است. هنگام صحبت معمولی، قفسه سینه با فشار وارد کردن به ششها باعث می‌شود که هوای فشرده از ششها از طریق حنجره بیرون رانده شود. تارهای صوتی که درست در پشت غدهٔ تیروئید قرار گرفته‌اند، اگر تحت تنش قرار گیرند، با عبور هوا مرتعش می‌شوند و بدین ترتیب هوا نیز متناسب با فرکانس ارتعاش تارهای صوتی مرتعش شده و در این حالت حروف صدادار تولید می‌گردند.

اگر تارهای صوتی از هم جدا شوند، جریان هوا از درون فاصلهٔ بین تارهای صوتی عبور می‌کند و تأثیر آن ایجاد نمی‌شود. جریان هوا سپس از فضای حلق عبور نموده و بسته به موقعیت دریچهٔ تنظیم عبور هوا از دهان یا بینی از فضای این دو عبور می‌نماید. جریان هوا از طریق دهان و بینی یا هر دو مشترکاً به بیرون داده می‌شود و هنگام صحبت این کاملاً قابل حس کردن است.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در حالت تولید حرف بی صدا مانند «س» یا «پ» تارهای صوتی در هم باز می‌شوند و یکی از دو حالت زیر غالب است. یا یک جریان مغشوش هوا تولید می‌شود، هنگامی‌که هوا از درون فضای نیمه بسته باریک در نقطه‌ای از اعضای صوتی عبور می‌کند (مانری) و یا یک تحریک گذری مختصر بدنبال ایجاد فشار پشت یک نقطهٔ کاملاً بسته در اعضای صوتی انسان اتفاق می‌افتد (مانند p).

وقتیکه جز جز کننده‌های مختلف مانند زبان، لبها، آرواره‌ها و پردهٔ تفکیک بینی و دهان در حین صحبت مدام حالتشان عوض می‌شود. شکل قسمتهای مختلف فضای داخل ناخیه صوتی تغییر می‌کند. ناحیه صوتی از حنجره تا لبها مانند یک حفرهٔ تشدید کننده عمل می‌کند که فرکانسهای معینی را تقویت و بقیه فرکانسها را تضعیف می‌نماید. اعضای صوتی انسان مثل یک لولهٔ صوتی غیر یکنواخت است که از تارهای صوتی تا لبها ادامه دارد و طول آن در افراد مانع حدود 17cm می‌باشد. بنابراین اولین فرکانس تشدید آن در فرکانس زیر اتفاق می‌افتد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

سطح مقطع غیر یکنواخت این لوله – مقدار زیادی متکی به وضعیت جز جز کننده‌ها است. و از صفر تا نزدیک 20cm متغیر است.

عضو صوتی مدهای تشدید یعنی از ارتعاش را داراست که فرمنت نامیده می‌شود که به مقدار زیادی به موقعیت دقیق جزءجزءکننده‌ها بستگی دارد.

شکل (2–2) تصویر شماتیک نیم رخ ناحیه صوتی را برای چند حرف صدادار نشان داده است و مقادیر نمونه فرکانسها نیز ذیل آن برای سه فرمنت اول بر حسب Hz داده شده است.

شکل 3–2 مشخصه‌های فرکانسی انتقالی این حروف را نشان می‌دهد، وضعیت تشدیدها به روشنی در این منحنی‌ها دیده می‌شود. خوبست که بدانیم بطور قابل ملاحظه‌ای در فهم صحبتها، فقط 3 فرمنت اول در تعیین صدایی که شنیده می‌شود مهم هستند. اگرچه برای تولید اصوات با کیفیت قابل قبول و بهتر فرمنت‌های بالا نیز مورد نیاز می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل موج صدای تولید شده بوسیلهٔ حنجرهٔ در هر حال یک سینوسی معمولی نیست. اگر اینطور بود ناحیه صوتی تشدید کننده، در خروجی فقط یک سیگنال سینوس می‌داد که بسته به میزان دور یا نزدیک بودن آن به فرکانس تشدید، تقویت یا تضعیف شده بود.

حنجره دارای دو لبه چین خوردهٔ پوستی بنام تارهای صوتی است که در هر سیکل از پریود فرکانس گام یکبار از هم باز شده و دوباره بهم می‌آید. فرکانس هیچ در مکالکات انسان مذکر از 50 الی 250 هرتز متغیر است که بطور متوسط حدود 100Hzاست. برای انسان مؤنث این فرکانس در رنج بالاتر تا حدود 500Hz قرار دارد. در آواز خواندن این فرکانس بالاتر نیز هست. بعضی آوازخوانهای اپرا، فرکانس گام خود را تا 1000Hz می‌توانند برسانند.

حرکت نوسانی تارهای صوتی شکل موجی تولید می‌کند که می‌توان آن را با یک پالس مثلثی تقریب زد. این شکل موج دارای طیف فرکانسی غنی است که با شیب 12db/ocdao می‌افتد و

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

همهٔ‌هارمونیکها نیز تحت تأثیر نواحی تشدید اعضای صوتی قرار می‌گیرند. (شکل 4-2)

شکل 4-2 بالائی مربوط است به مدل فیلتر منبع که مشخصات فیلتر و طیف است. شکل سمت راست تحریک دهانهٔ حنجره در گفتار طبیعی است و بالاخره شکل سمت چپ تقویت در تحریک دهانه حنجره است.

ناحیه صوتی وقتی که به وسیلهٔ یک شکل موج با طیف‌هارمونیکی گسترده قرار می‌گیرد. نقاط موجی در طیف انرژی شکل موج مکالمات تولید می‌کند که همان فرمنت‌ها هستند. پائین ترین فرمنت که اولین فرمنت نامیده می‌بود از حدود .200Hz تا 100Hz در حین صحبت متغیر است. و مقدار دقیق آن متکی به ابعاد ناحیه صوتی می‌باشد.

فرمنت دوم از حدود 500Hz تا 9500Hz متغیر است و فرمنت سوم از حدود 1500Hz تا 3500Hz. البته گفتار یک پدیدهٔ استاتیک و ثابت نیست. مدل لولهٔ صوتی می‌تواند نمایشگر طیف گفتار در مدتی که یک حرف صدادار بطور ممتد کشیده می‌شود و دهان در حالت ثابت باقی

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

می‌ماند (مانند آآآ) باشد. اما در گفتار واقعی زبان و لبها در تحریک دائم هستند و شکل ناحیه صوتی را مرتباً تغییر می‌دهند و نتیجتاً موقعیت فرکانسهای تشدید را عوض می‌کنند. این مشابه یک لوله صوتی است که بطور مداوم از قسمتهای مختلف فشرده و منبسط می‌شود.

بعنوان مثال در هنگام بیان کلمهٔ «میز» احساس می‌کنیم که چطور زبانتان به سقف دهان نزدیک می‌شود و باعث ایجاد یک حالت عبوری نیمه بسته در نزدیک جلوی حفرهٔ صوتی می‌شود.

در طیف یک حرف صدادار که بطور مداوم ادامه داده شود، بصورت مداوم ادامه داده شود، بصورت یک طیف انرژی ثابت می‌آید. اما باید توجه داشت که منظور از حروف صدادار در اینجا با آنچه معمولاً تصور می‌شود متفاوت است. بگوئید «I» و ببینید زبان هنگام بیان به آهستگی تغییر موقعیت می‌دهد. از نظر تکنیکی این تها یک حرف صدادار نیست و یک لغزش بین دو موقعیت مربوط به دو حرف صدادار است.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تفاوتهای شنوایی اصلی بین حروف صدادار مختلف و فرکانسهای دو فرمنت اول آنهاست. دیدیم که صحبت کردن، محدود کردن صوت است بعد از آنکه بوسیلهٔ نوسانات در حنجره تولید شده است. وقتی که با حالت نجوا و زمزمه صحبت می‌کنیم، تارهای صوتی در حنجره کمی‌از هم جدا نگاه داشته شده‌اند و هوای عبوری از آنها بصورت مغشوش در می‌آید و باعث تحریک حفرهٔ تشدید کننده (اعضای صوتی) بوسیلهٔ یک نویز می‌گردد.

فرمنتها در اینجا نیز حضور دارند و روی نویز سوار شده‌اند. برای حروف صدادار ریشهٔ حروف در تارهای صوتی است و صدا حاوی فرت‌های شبه پریودیک با باند عریض است که توسط مرتعش شدن تارهای صوتی ایجاد گردیده‌اند.

برای حروف بی صدا مانند «س» صدا در نقطه نیمه بسته تحت فشار در عضو صوتی قرار دارد و شامل جریان هوای شبه رندوم مغشوش می‌باشد. برای حروف بی صدا مانند p (مثل pop) ریشهٔ حرف در نقطه مسدود قرار دارد و بوسیلهٔ آزاد شدن هوای فشرده یکه

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پشت نقطهٔ کاملاً مسدود ایجاد گردیده است، تولید می‌گردد. از نوع اخیر که صداهای تنفسی نامیده می‌شود، حرف H مثل کلمه Hello را نیز می‌توان ذکر نمود. بدین ترتیب حروف مکالمات را به 3 دسته می‌توان تقسیم نمود:

1- حروف صدادار

2- حروف بی صدا سایشی مثل س ر ش ف

3- حروف بی صدای تنفسی هـ پ

تولید حروف بی صدا از نوع سایشی نیز میسر است که مثلاً حروف ز ژ – و که آنها را صدادار سایشی می‌نامیم. نمونهٔ حروف بی صدا سایشی س – ش – ف هستند.

2-2 مدل منبع – فیلتر گفتار

فرض اساسی در تقریباً تمامی‌سیستمهای پردازش گفتار این است که منبع تحریک و سیستم اعضای صوتی مستقل از هم هستند. این موضوع به ما اجازه می‌دهد که در مورد تابع تبدیل عضو صوتی

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بحث کنیم و این امکان را می‌دهد که این سیستم را با هر منبع ممکن دیگر تحریک نمائیم.

فرض فوق در مورد اکثر حالات مورد نظر ما به خوبی معتبر می‌باشد. البته حالاتی نیز وجود دارد که فرض فوق معتبر نمی‌باشد و مدل اساسی می‌شکند (مانند حرف p در po). برای بیشتر قسمتها ما معتبر بودن آن را فرض می‌نمائیم. بر این اساس یک مدل دیجیتالی ساده تولید گفتار را در شکل (5–2) مشاهده می‌کنیم.

منابع تحریک عبارتند از یک مولد پالس که فرکانس آن همان فرکانس گام می‌باشد و یک مولد نویز رندوم.

مولد پالس در هر تعداد از نمونه و مرتبط با شروع عبور یک حجم از هوا از تارهای صوتی، یک پالس تولید می‌کند که طول آن متناسب با پریود گام می‌باشد. خروجی نویز رندوم مشابه اغتشاش شبه رندوم برار حروف بی صدا می‌باشد. هر کدام یا هر دو این منابع ممکن است بعنتوان ورودی برای یک فیلتر دیجیتال خطی و متغیر با زبان بکار روند. این فیلتر، عضو صوتی (ناحیه صوتی) را مشابه سازی

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

می‌نماید و ندا ضرایب فیلتر تعیین کننده ناحیه صوتی بعنوان یک تابع متغیر نسبت به زمان در حین گفتار می‌باشند.

بطور متوسط در هر 10 میلی ثانیه یکبار، ضرایب فیلتر عوض می‌شوند که نشانگر مشخصات ناحیه صوتی جدید هستند، کنترل بهره

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فصل دوم

Speech analysist

مقدمه:

در این بخش در مورد تجزیه و تحلیل سیگنال صوت بحث خواهد شد و مراحل پردازش روی سیگنال صحبت جهت آمادگی آن برای بازشناسی مورد بررسی قرار خواهد گرفت.

در این بخش اطلاعاتی در مورد نحوهٔ فریم بندی، اعمال پنجره، عملیات جداسازی سیگنال صحبت از روی زمینه، voice Decision ، فرکانس فرمنت و ضرایب LPC ، کپستروم بحث خواهد شد.

فریم بندی سیگنال صحبت

دنبال نمونه‌های از سیگنال صحبت در شکل نشان داده شده است. همان طور که از شکل پیدا است، خواص سیگنال با گذشت زمان تغییر می‌کند. مثلاً در بعضی از زمانتها سیگنال واکه دار یا بی واکه است یا نقاط ماکزیمم دامنه بسیار تغییر می‌کند و همچنین در نقاطی که سیگنال صحبت واکه دار است فرکانس گام عوض می‌شود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در تمام کارهای پردازش سیگنال فرض بر این است که خواص و ویژگی سیگنال صورت در طول زمان به آرامی تغییر می‌کند.

در طول یک دوره کوتاه از زمان تقریباً ثابت است. با فرضهای فوق ما به روشی دست پیدا می‌کنیم که در آن به پردازش زمان کوتاه یک قسمت از سیگنال صحبت می‌پردازد.

اغلب این بخش‌های کوتاه سیگنال صحبت که به آن analysis frame نیز می‌گویند. با یکدیگر هم پوشانی دارند.

اگر بخواهیم یک بخش یا قسمت از سیگنال صحبت را نشان بدهیم بصورت ریاضی به فرم زیر می‌باشد.

که در آن m طول فریم می‌باشد.

برای بدست آوردن N ، نمونه فدیک سیگنال صحبت باید آن را فریم بندی کنیم.

اما برای از بین بردن تأثیر لبه‌ها باید از پنجره استفاده نمود. استفاده از پنجره دو مزیت دارد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

1- پنجره با تضعیف سیگنال در ابتدا و انتهای پنجره اثر تغییر ناگهانی دامنه را در ابتدا و انتهای پنجره یا فریم کاهش می‌دهد.

2- با ضرب کردن پنجره در یک سیگنال صحبت در زمان، موجب ایجاد کانولوتن طیف پنجره و سیگنال صحبت در محور فرکانس خواهیم شد. در حقیقت ما با این یک عمل Weignted moving avarage در محور فرکانس انجام داده ایم.

این کار باعث از بین رفتن اعوجاج حاصل از فریم بندی سیگنال صحبت می‌شود.

پنجره بکار برده شده باید دارای دو خاصیت باشد: اول دقت فرکانسی بالا یعنی، robe اصلی بسیار باریک و کوتاه باشد. 2- فرکانس کوچک نسبت به سایر مؤلفه‌های طیف ایجاد شده بوسیلهٔ کانولوتن. به عبارت دیگر تضعیف بسیار زیاد در robe اصلی.

پنجره Hamming دارای خاصیت‌های فوق بوده

با معلوم کردن میزان هم پوشانی و طول پنجرهٔ Haming می‌توان سیگنال صحبت را به بخش‌هایی به طول مساوی تقسیم نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فیلتر پیش تأکید

ممکن است محدودهٔ دینامیک طیف صحبت بسیار وسیع باشد. این باعث می‌شود که در هنگام محاسبهٔ ماتریس مشخصه سیگنال دچار مشکل شویم و همچنین این فیلتر پیش تأکید باعث یکنواخت تر کردن طیف فرکانسی خواهد شد. برای این فیلتر پیش تأکید از یک فیلتر FIR درجهٔ اول استفاده می‌کنیم.

می‌توان مقدار بهینه را بدست آورد ولی بسته به گوینده‌های مختلف فرق می‌کند ولی مقدار آن زیاد در نتایج تأثیر ندارد.

جداسازی سیگنال صحبت از روی سیگنال زمینه

شکل اساسی در پردازش صوت، تشخیص سیگنال صحبت از سیگنال نویز زمینه می‌باشد. از این مسأله اغلب بعنوان مسأله

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تشخیص ابتدا و انتهای صوت نام برده می‌شود. بوسیله تشخیص درست ابتدا و انتهای یک سیگنال صحبت، هم میزان پردازش سیگنال پائین می‌آید، هم نرخ بازشناسی بالا می‌رود.

الگوریتم‌های مختلفی برای تشخیص و جداسازی سیگنال صحبت از روی سیگنال زمینه وجود دارند. در این پروژه دو روش و پیاده سازی شده است. در روش اول از پارامترهای میزان عبور از صفر و انرژی هر فریم برای پیدا کردن ابتدا و انتهای سیگنال صحبت استفاده می‌شود.

این الگوریتم به طور قابل ملاحظه‌ای می‌تواند در محیطهای اکوستیکی که دارای سیگنال به نویز 30d هستند. با دقت بالا کار کند. الگوریتم اول برای گوینده‌های مختلف و شرایط مختلف، قسمت شد و نتایج خوبی بدست آمد.

الگوریتم دوم تقریباً شبیه الگوریتم اول است، و فقط کمی‌تفاوت با آن در نحوهٔ استفاده از پارامتر انرژی دارد.

الگوریتم دوم نیز از پارامترهای انرژی و استفاده می‌کند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای دستیابی به یک الگوریتم که بتواند سیگنال صوت را غیر از صوت جدا کند، ابتدا لازم است محیط صوتی را که در آن صدا ضبط شده است مشخص کنیم، عموماً در این پروژه‌ها دارای دو نوع محیط صوتی می‌باشیم. در حالت اول صدای کاربرد در یک محیط آزمایشگاهی بدون حضور، هیچ نیز اکوستیک ضبط شده است. در حالت دوم، صدای کاربرد بوسیلهٔ یک میکروفن معمولی از طریق کامپیوتر ضبط می‌شود که به همراه آن نویز وجود دارد.

در شکل (1) سکوت زمینه در هنگام ضبط صدا در محیط اول و دوم آورده شده است.

همان طور که در شکلها دیده می‌شود، سکوتی که در محیط اکوستیک باشد، دارای یک مؤلفه فرکانس پائین قبلی (با پریود 8ms) می‌باشد. اما سکوتی که در محیط معمولی و از طریق کامپیوتر ضبط شده دارای یک طیف وسیعی از فرکانسها می‌باشد.

شکل (2) طیف فرکانسی این دو سکوت زمینه را نشان می‌دهد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این طیفهای فرکانسی از یک پنجرهٔ Hamming، که دارای 512 نقطه است بدست آمده، دانهٔ آن به صورت لگاریتمی‌می‌باشد. غیر از مؤلفه فرکانس پائینی تقریباً هر دو طیف شبیه به هم هستند.

مسأله اساسی در پیدا کردن ابتدا و انتهای سیگنال صحبت، نویزهای موجود در سیگنال صحبت می‌باشد.

یک راه ساده جهت جدا کردن سیگنال صحبت از روی تغییرات سریع انرژی سیگنال صحبت در هنگام اول سیگنال و سکوت زمینه است.

در هنگامی‌که در حالت اول یک صدا ضبط می‌شود می‌توان حتی از طریق چشم نیز تفاوت بین سیگنال زمینه و سیگنال صورت را به دلیل پائین بودن سطح نویز و یا در حقیقت عدم وجود نویز تشخیص داد. در حقیقت چشم ما از طریق مشاهدهٔ تغییرات، ناگهانی شکل موج یا همان تغزیت ناگهانی انرژی قادر به تشخیص ابتدا و انتهای سیگنال صوت می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

همان طور که در بخش قبلی در مورد سیگنال بی صدا بحث کردیم، تمام این صوتها دارای ماهیت نویز گونه می‌باشند، بنابراین با افزایش سطح نویز سیگنال زمینه، اگر صوت با یک حرف سایشی مثل «ف» شروع شود دیگر چشم قابلیت تشخیص ابتدای سیگنال را از نویز ندارد. همچنین به دلیل پائین بودن انرژی سیگنال صوت بی صدا پیدا کردن یک آستانه خوب برای جدا کردن ابتدا و انتهای سیگنال صوت فقط با پارامتر انرژی مشکل می‌باشد.

همان طور که گفته شد به کمک پارامتر انرژی نمی‌توان ابتدا و انتهای سیگنال صوت را معین نمود. پارامتر دیگری که در الگوریتم استفاده خواهد شد پارامتر میزان عبور از صفر هم فریم می‌باشد.

این پارامتر بیان می‌کند، سیگنال صوت در هر فریم چند بار به سطح مثبت و سپس در نمونه بعدی به سطح منفی رفته است، یعنی در حقیقت از صفر عبور کرده است.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

میزان این پارامتر ارتباط مستقیمی با فرکانس سیگنال دارد. هر چقدر فرکانس سیگنال بیشتر باشد نرخ عبور از صفر آن نیز بیشتر خواهد بود.

همان طور که قبلاً گفته شد، صوتهای بی صدا ماهیت نویز گونه دارند ولی فرکانس عبور از صفر آنها کمتر است از نویز سفید یا نویز زمینه می‌باشد. یعنی نویز زمینه دارای خاصیت پراکندگی بیشتری است. سپس با کمک این پارامتر می‌توان به راحتی اصوات بی صدا را از روی سیگنال زمینه جدا نمود.

به طور کلی مشکل جداسازی سیگنال‌های صوتی از روی زمینه را می‌توان به سیگنال‌هایی محدود کرد که اصوات زیر ختم شوند:

ا) صوتهای سایشی ضعیف مثل «ف»

ب) صوتهای انفجاری مثل «پ، ک و ت»

ج) کلماتی که به حروفی ختم می‌شوند که از طریق بینی ادا می‌شوند مثل «م، ن»

د) حروف صدادار سایشی در انتهای کلمه

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

هـ) کم شدن اثر حرف صدادار در انتهای کلمه

با توجه به مسائل مطرح نشده می‌توان با کمک پارامترهـای انـرژی و ZCR الگوریتمی‌را طراحی نمود که قابلیت حل مسائل فـوق را داشـته باشد.

- الگوریتم تشخیص ابتدا و انتهای سیگنال با کمک انرژی و ZCR

طبق بحث‌های گذشته هدف از این الگوریتم

1- سادگی، کارآمدی بالا در هنگام پردازش

2- پیدا کردن یک نقطه ابتدا و انتها با اطمینان بالا

3- قابلیت به کار بردن الگوریتم در مورد سیگنالهای با زمینهٔ متفاوت

همان طور که گفته شد با کمک پارامترهـای انـرژی و میـزان عبـور از صفر به همراه یک سری تصمیمات منطقـی در مرحلـهٔ آخـر مـی‌تـوان الگوریتم با قابلیتهای فوق را پیاده سازی کرد.

هـر دو پـارامتر انـرژی و میـزان عبـور از صـفر، بسـیار سـاده قابـل محاسبه هستند. برای پیدا کـردن انـرژی هـر فـریم مـی‌تـوان از جمـع مقدار دامنه به توان دو استفاده نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

n شمارهٔ هر فریم می‌باشد و M طول پنجره می‌باشد.

برای محاسبهٔ مقدار عبور از صفر ابتدا مقدار DC سیگنال را از آن کم می‌کنیم، سپس آن را از یک فیلتر به 11 گذر عبور می‌دهیم. این دو کار را برای هر فریم انجام داد، سپس مقدار دفعاتی را که سیگنال از سطح مثبت منفی رفته و یا بالعکس را طبق فرمول زیر حساب می‌کنیم.

پس از پیدا کردن مقدار انرژی و میزان عبور از صفر برای هر فریم طبق الگوریتم و با پیدا کردن نقاط آستانه می‌توان به جداسازی سیگنال صوت از روی زمینه پرداخت.

قبل از توضیح الگوریتم در بعضی از مقالات مشاهده شده که ، توصیه می‌شود قبل از فریم بندی و پردازش سیگنال صوت، سیگنال را از یک فیلتر پائین گذر با فرکانس 10Hz و یک فیلتر بالاگذر 100Hz عبور دهیم. با انجام عملیات فوق و پیاده سازی روش مذکور مشاهده می‌شود کاملاً کیفیت شنیداری سیگنال پائین می‌آید، ثانیاً نرخ

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بازشناسی کاهش خواهد یافت. لذا از انجام عمل فیتر کردن خودداری شده است و در مرحلهٔ اول از یک فیلتر بالاگذر FIR جهت حذف DC استفاده شده است.

فرض بر این است که در حدود بین 100ms تا 200ms اول سیگنال هیچ نوع صدایی وجود ندارد و فقط سیگنال زمینه خالص وجود دارد. بنابراین در این محدوده می‌توان ویژگی‌های آماری سیگنال زمینه را پیدا نمود. این ویژگیها تا میانگین و انحراف معیار و مقدار انرژی و میزان عبور از صفر سیگنال سکوت می‌باشد.

برای پیدا کردن مقادیر آستانه برای میزان عبور از صفر از فرمول زیر استفاده می‌کنیم.

یعنی میانگین مقدار ZCR

با در برابر انحراف میعار آن صحیح می‌کنیم.

و بدین ترتیب از طریق این مقدار آستانه می‌توان صوت بی صدا را از روی سیگنال زمینه جدا نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای پیدا کردن مقادیر آستانه از انرژی به این ترتیب عمل می‌کنیم.

مقدار ماکزیم انرژی فریمها را بدست می‌آوریم و همچنین میانگین انرژی سکوت زمینه را بدست می‌آوریم.

سپس از طریق فرمول زیر مقادیر آستانه را بدست می‌آوریم.

فرمول 2 نشان می‌دهد، مقدار برابر با 3 درصد ماکزیمم انرژی (که برای مقدار سکوت نرمالیزه شده) می‌باشد و فرمول (3) بیان می‌کند مقدار 4 برابر انرژی سکوت می‌باشد.

مقدار آستانه پائین مقدار این دو عدد یعنی و می‌باشد، مقدار آستانه بالایی 5 برابر مقدار آستانه پائین است.

در شکل 1 فلوچارت مربوط به الگوریتم برای حدس اولیه نشان داده شده است. در ابتدا الگوریتم از اولین فریم شروع به جستجو برای یافتن نقطه‌ای می‌کند مقدار انرژی آن فریم بیشتر از حد آستانهٔ پائینی باشد. بعد از یافتن اولین فریم که مقدار انرژی آن از حد آستانه پایینی گذشت، آن فریم را به عنوان نقطه شروع اولیه می‌نامیم. البته

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این اتفاق به شرطی می‌افتد که بعد از چند فریم مقدار انرژی از حد آستانه بالایی نیز عبور کند. همچنین نباید میزان انرژی قبل از رسیدن به ITW از ITL کمتر باشد.

دلیل قرار دادن مقادیر آستانه بالایی جهت مطمئن شدن از حضور سیگنال صوتی در فایل ضبط شده است.

الگوریتم مشابه‌ای برای پیدا کردن نقطهٔ انتهایی به کار می‌رود. بدین ترتیب که الگوریتم از آخرین فریم به صورت معکوس شروع به یافتن نقطه‌ای یا فریمی‌می‌کند که مقدار انرژی آن بیشتر از سطح ITL باشد.

با پیدا کردن نقاط اولیه ابتدایی و انتهایی ما این نقاط را می‌نامیم. تا این زمان ما تنها از پارامتر انرژی استفاده نموده ایم که بتوانیم نقاط ابتدا و انتها را مشخص کنیم. این نقاط ابتدا و انتها به طور کامل بیان گر وجود نقاط کاملی که سیگنال صوت در آن شروع و خاتمه یافته نمی‌باشد. دلیل این موضوع را قبلاً گفته ایم و باید در این مرحله بگوییم قسمتی از سیگنال صوت خارج از می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پس از یافتن نقاط با الگوریتم شروع به چک کردن مقدار میزان عبور از صفر برای نقاط یعنی حدود 250ms قبل می‌کند. اگر تعداد زمانهایی که میزان عبور از صفر هر فریم از مقدار آستانه IZCT کمتر باشد. در حدود 2 یا 3 بیشتر بود. نقطه انتهایی به همان آخرین نقطه که از حد آستانه کمتر شد، منتقل می‌گردد. در صورتیکه در این 250ms هیچ فریمی‌یافت نشود که مقدار میزان عبور از صفر آن کمتر از حد آستانه باشد. همان نقطه به عنوان اول فریم شناخته خواهد شد.

الگوریتم مشابه‌ای برای پیدا کردن نقاط انتهایی به کار می‌رود. این بار نقاط برای پیدا کردن فریم‌هایی که دارای میزان عبور از صفر زیر مقدار آستانه هستند جستجو خواهد شد.

Fast End point Dection algorithm in office EnviROMENT

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این الگوریتم شامل 4 مرحله می‌باشد. در مرحلهٔ اول سیگنال صوت یک کلمه، پیش پردازش شده و نویز زمینه تخمین زده می‌شود و از آن جهت وفق دادن الگوریتم در مراحل بعدی استفاده خواهد شد. در مرحلهٔ دوم اولین و آخرین نقطه فریم واکه‌دار به عنوان مبنای جستجو معین خواهند شد.

در مرحلهٔ سوم با قرار دادن یک سطح انرژی پائین در اطراف ناحیه ابتدا و انتها می‌توان در مرحلهٔ چهارم نقاط ابتدایی و انتهایی را مشخص نمود.

تخمین اولیه نویز زمینه:

برای حذف DC ، و تقویت جزءهای فرکانس بالا، ابتدا سیگنال را با فیلتر درجهٔ اول FIR ، پیش تأکید می‌کنیم.

با بدست آوردن نمونه‌هایی از ابتدا و انتهای سیگنال می‌توان نویز زمینه یا (نویز محیط اکوستیکی) را حدس زد. با کمک رابط (2) انرژی نویز را در دو فریم اول و آخر که طول آنها زیاد است و همپوشانی هم با هم ندارند حساب می‌کنیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

که در آن طول پنجره یا طول فریم می‌باشد (حدود 80ms)

میزان نویز در ابتدای سیگنال زمینه با کمک فرمول (3) محاسبه خواهد شد.

اگر میزان تفاوت انرژی دو فریم کمتر از دو برابر یکی انرژیها باشد، انرژی نویز برابر با میانگین دو انرژی است، در غیر این صورت انرژی نویز برابر مینیمم این دو انرژی است.

نویز تخمین زده شده در انتهای سیگنال هم به همان صورت تخمین زده خواهد شد که از دو مقدار انرژی فریم‌های آخری استفاده خواهد شد.

در نهایت مقدار انرژی نویز در کل سیگنال با کمک میزان نویز در ابتدا و انتهای سیگنال تخمین زده خواهد شد.

اگر اختلاف بین دو مقدار کمتر یا مساوی دو برابر یکی از مقدارها باشد، نویز زمینه برابر با میانگین دو مقدار خواهد بود. در غیر این

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

صورت نویز زمینه قابل تشخیص نخواهد بود و سیگنال ورودی برگشت داده خواهد شد و خط آشکار می‌شود.

با این وجود، سطح انرژی نویز بدست آمده، باید درحد دو آستانه قرار گیرد. در غیر این صورت سیگنال ورودی غیر قابل قبول می‌باشد و به عنوان کاملاً نویزی یا بسیار ضعیف شناخته خواهد شد.

TN مقدار قابل قبول انرژی نویز برای محیطهای اکوستیکی می‌باشد و TS به عنوان مقدار انرژی می‌نیمم سکوت برای تشخیص قطعی یا عدم وجود سیگنال می‌باشد.

مقدار TL و TN به نوع میکروفن و خطای کواتیزیشن بستگی دارد. می‌توان به طور حدودی و در نظر گرفت.

پیدا کردن اولین و آخرین فریم واکه‌دار

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مکان شروع اولین فریم واکه‌دار صحبت ورودی و مکان آخرین فریم واکه‌دار صحبت ورودی به عنوان مبنا برای جستجو مشخص می‌شوند.

برای مشخص کردن واکه دار بودن یا نبودن فریم به جستجوی دامنه در زمان می‌پردازیم. اولین فریمی‌که دارای N قله بالای حد آستانه TA باشد به عنوان اولین فریم voice ورودی شناخته خواهد شد.

مقدار N به طور تجربی بدست می‌آید.

بنابراین مقدار

به عنوان اولین فریم واکه‌دار بدست می‌آید.

مقدار آستانه برای دامنه (TA) به طور تجربی از طریق فرمول زیر بدست می‌آید.

که در آن

و یک ثابت است که به طور تجربی بدست می‌آید.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

همان طریق که گفته شد، الگوریتم مشابهی در حوزهٔ زمان با چک کردن دامنه به صورت معکوس از آخرین فریم شروع به پردازش می‌کند و اولین فریمی‌که واکه‌دار بود به عنوان معلوم می‌شود.

تفاضل بین باید از حد یک آستانه بیشتر باشد تا مشخص شود سیگنالی وجود داشته است و یا حداقل سیگنال موجود دارای معنا می‌باشد. این مقدار حدود 20ms می‌باشد.

در غیر این صورت الگوریتم تشخیص خط می‌دهد.

مکان ناحیهٔ دارای سطح انرژی پائینی

در ابتدای سیگنال یک محدودهٔ کم انرژی قرار داده می‌شود که فرض می‌شود، نقطه شروع در آنجا قرار دارد.

همچنین در انتهای سیگنال ورودی یک محدوده کم انرژی قرار داده می‌شود، که فرض می‌شود نقطه انتهایی درون آن قرار دارد. در محدودهٔ این نقاط، الگوریتم جستجو برای پیدا کردن نقاط نهایی شروع و پایان سیگنال صحبت بسیار سریع‌تر عمل خواهد کرد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

یک فریم 80ms از نقطه ابتدایی اولیه به سمت عقب برگردانده می‌شود، و منحنی انرژی سیگنال را رسم می‌کند. این مقادیر انرژی با دو مقدار آستانه جهت پیدا کردن نواحی کم انرژی مقایسه خواهند شد.

نتایج تحلیلی برای نواحی از طریق فرمول زیر بدست می‌آید.

مقادیر به طور تجربی پیدا خواهند شد.

شکل (2) مقادیر ، و زمانهای ، را نشان می‌دهد.

در انتهای سیگنال ورودی یک فریم 80ms در نقطه انتهایی اولیه به سمت جلو حرکت داده می‌شود و منحنی انرژی سیگنال رسم خواهد شد. این مقادیر انرژی با دو مقدار آستانه جهت پیدا کردن نواحی کم انرژی مقایسه خواهد شد.

، یک مقادیر انرژی هستند که به طور تجربی بدست می‌آیند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

قابل توجه است که مقادیر انرژی آستانه انتهایی بیشتر از نقطه اولیه می‌باشد. این به دلیل این است که ناحیه انتهایی سیگنال صحبت دارای محدودهٔ نویز تنفس است.

مرحله 4 پیدا کردن نقطهٔ انتهایی و ابتدایی

در محدودهٔ نواحی کم انرژی که در بخش قبل حدس زده شد، نقطه واقعی ابتدایی و انتهایی جستجو خواهد شد. در بین محدودهٔ ، سیگنال به پنجره‌هایی بدون همپوشانی با طول 30ms تقسیم شده و مقادیر انرژی برای آن محاسبه خواهد شد.

نقطه شروع واقعی سیگنال، متناسب است با میزان ماکزییم مقدار منحنی انرژی. فرمول تحلیلی جهت پیدا کردن نقطهٔ ابتدایی به شرح ذیل است.

به همان روش، نقاط بین ، جهت پیدا کردن نقطه انتهایی جستجو خواهند شد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پیاده سازی الگوریتم‌ها

هر دو الگوریتم فوق پیاده سازی شده‌اند. الگوریتم نهایی که جهت بکارگیری در سیستم پیاده سازی شد، مخلوطی از دو الگوریتم فوق می‌باشد.

در الگوریتم نهایی، روش پیدا کردن انرژی نویز و تخمین مقدار آن مانند روش دوم می‌باشد، و جهت پیدا کردن مقادیر آستانه از این مقدار طبق روابط الگوریتم اول استفاده خواهیم کرد. برای پیدا کردن حد آستانه تعداد عبور از صفر از رابطهٔ

استفاده می‌کنیم.

دلیل عدم استفاده از ساختار کلی الگوریتم دوم و پیاده سازی الگوریتم اول، وابستگی بسیار شدید الگوریتم دوم به پارامترهای تجربی بود. همان طور که در الگوریتم دوم مشاهده می‌کنیم، ما در این الگوریتم دارای حدود 7 پارامتر هستیم که به طور تجربی و به

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

روش آزمایش و خطا بدست می‌آید. ولی در الگوریتم اول تنها یک پارامتر است که به روش تجربی بدست می‌آید.

همچنین الگوریتم دوم شدیداً وابسته به طول پنجره است و برای پیدا کردن طول پنجرهٔ بهینه باید تمام مقادیر پارامترها را تغییر داد.

بنابراین پس از پیاده سازی روش‌های فوق تصمیم گرفته شد از ایده‌های الگوریتم دوم در جهت پیاده سازی الگوریتم اول استفاده کنیم.

با پیاده سازی الگوریتم اول نتایج خوبی بدست آوردیم. البته در این مرحله آزادی عمل در انتخاب طول پنجره وجود داشت و همچنین بازشناسی گفتار وجود نداشت. بنابراین در این مرحلهٔ هدف فقط جداسازی سیگنال صحبت در زمینه بوده که به خوبی انجام پذیرفت.

استخراج ضرائب کپزرم

مدل فیلتر منبع که در فصل اول در مورد مسیر صوتی انسان معرفی کردیم بیان می‌کند، سیگنال صوت حاصل ضرب یک سیگنال تحریک و یک فیلتر خطی در فضای فرکانسی می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

که در این صورت باید خلاص طیف قدرت یک فریم از سیگنال صوت را بتواند نشان دهد. همچنین نیز بتواند بخوبی جزئیات سیگنال تحریک را نشان دهد.

با یک نگاه دقیق به معادلهٔ (1) می‌توان فهمید که از طریق تبدیل ضرب به جمع و سپس فیلتر کردن نتیجه می‌توان توابع ، را بدست آورد. برای تبدیل ضرب به جمع می‌توان از خواص لگاریتم استفاده نمود.

برای بیشتر فعالیت‌های مربوط به صوت نا بخش حقیقی دامنه را احتیاج داریم پس معادلهٔ را می‌توان بصورت نوشت.

به صورت خیلی آرام تغییر می‌کند و دارای دو مؤلفه فرکانس بالا و یک مؤلفه فرکانس پائین می‌باشد. بنابراین با یک تبدیل دیگر می‌توان این مؤلفه‌ها را به صورت طبیعی از هم جدا نمود به راحتی می‌توان را بدست آورد. به این روند، تحلیل کپستروم می‌گویند.

همان طور که در اشکال موجود دیده می‌شود، بیشتر جزئیات در نزدیکی نقطه شروع، در اوایل سیگنال اتفاق می‌افتد. بنابراین ضرائب

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مرتبهٔ پائین دارای جزئیاتی راجع به خواص فرکانسی می‌باشند. ضرایب بعدی شامل و پیکهایی هستند که در صورت واکه‌دار بودن فریم می‌توانند بیان‌گر فرکانس گام باشند.

ضرائب مرتبه پائیین کپستروم نسبت به شیب طیف فرکانسی حساسیت دارند، همچنین نوع پالی خروجی حنجره و تارهای صوتی نیز روی آنها تأثیر می‌گذارد. ضرایب مرتبهٔ بالای کپستروم نسبت به مکان پنجره و طول آن و مقدار هم‌پوشان و سایر عوامل موقتی تأثیرپذیر هستند. همچنین در تمام سیستمهای پردازش صوت – صورت مستقل گوینده باید تمام اطلاعات مربوط به یک گویندهٔ خاص را از ضرائب مشخصه حذف نمود.

جهت از بین بردن تغییرات بحث شده و جداسازی ، ، از یک پنجره استفاده می‌کنیم. که به صورت یک سینوسی عمل می‌کند. این پنجره مقدارهای واقع شده در وسط پنجره را تقویت می‌کند و مقدارهای ابتدایی و انتهایی را کمی‌تضعیف می‌کند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

که در آن L طول پنجره ، یا طول مورد دلخواه ضرائب کپستروم می‌باشد.

برای هر فریم از سیگنال صحبت می‌توان مقدارهای ضرائب کپستروم (معمولاً) را استخراج نمود، و ماتریس بدست آمده را به عنوان ماتریس ضرائب ویژگی یا مشخصه آن سیگنال صحرت معرفی نمود.

محاسبهٔ ضرایب دلتاکپترال

همانطور که می‌دانیم ضرائب پیشگویی خطی و یا ضرای کپترال مربوط به یک قطعه تحلیلی از سیگنال صحبت می‌باشند و عمل استخراج این ضرایب بدون در نظر گرفتن قطعات قبلی یا بعدی صورت می‌گیرد. بالطبع ضرایب مشخصه بدست آمده تنها نماینده خصوصیات همان قطعه خاص از سیگنال صحبت خواهند بود.

در حقیقت سیگنال صحبت غیر ایستا می‌باشد و در نتیجه مشخصه‌های استخراجی باید بازگو کنندهٔ تغییرات دینامیک سیگنال صحبت نیز باشند.

لذا استفاده از ضرایب دلتا کپسترال پیشنهاد می‌گردد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در این صورت مجموع مشخصه‌های K قطعه قبل و بعد از یک قطعه به همراه ضرایب کپترال همان قطعه به عنوان مشخصهٔ آن فریم در نظر گرفته خواهد شد.

پردازش روی فریم‌های واکه دار:

همان طور که در فصل اول سخن گفتیم، اختلاف انسان به دو دستهٔ واکه دار یا بی واکه تقسیم می‌شوند. همچنین در مورد حروف واکه دار می‌توان گفت بعضی از آنها صدادار هستند. در فارسی دارای 1 حرف صدادار هستیم.

از آنجا که بیشتر اطلاعات شنیداری توسط حروف صدادار منتقل خواهند شد. بنابراین این حروف دارای اهمیت زیادی هستند. از این رو ما احتیاج به شناسایی حروف صدادار در یک کلمه هستیم. علت این امر را در بخش ارزیابی صدای گوینده بیان خواهیم کرد.

روشهای پیدا کردن فریم واکه‌در گوناگون هستند و از پارامترهای مختلفی می‌توان استفاده نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

همچنین اطلاعات دیگری که در این فریم‌ها موجود است. فرکانس گام شخص گوینده است. دربارهٔ نحوهٔ استفاده از فرکانسی گام بعداً صحبت خواهیم کرد.

در این پروژه ما دو روش را جهت شناسایی فریم واکه‌دار پیاده سازی کردیم. همچنین از هر دو روش فرکانی گام را نیز استخراج نمودیم.

روش اول، روش خود همبستگی می‌باشد. در روش دوم از ضرایب کپستروم جهت استخراج فریم واکه‌دار و فرکانی استفاده خواهد شد.

روش اول: استفاده از autocorrelation

تابع خود همبستگی روش ساده‌ای را برای نمایش پریود یک شکل در حوزهٔ زمان فراهم می‌آورد. در این روش‌ها به بررسی روشهای پیاده سازی تشخیص واکه‌دار بودن فریم و سپس فرکانی گام آن از طریق تابع خود همبستگی خواهیم پرداخت.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

یکی از محدودیتهای استفاده از تابع خود همبستگی این است که اطلاعات زیادی را از سیگنال در خود نگه می‌دارد. برای جلوگیری و از بین بردن مسئله فوق بهترین راه حل این است که در هنگام پردازش سیگنال، ورودی را طوری جلو ببریم که، خاصیت پریودیک بودن سیگنال بر سایر خواص و ویژگیهای سیگنال غالب باشد.

از تکنیکهایی که این چنین عملیاتی را روی سیگنال انجام می‌دهند. بعنوان «صاف کنندهٔ طیف فرکانسی» یاد می‌شود. این تکنیکها کارشان حذف اطلاعات مربوط به فیلتر صوتی می‌باشد. با این کار، می‌توان میزان دامنهٔ هرهارمونیک را به شکل همان قطار پالی پریودیکی در آورد.

روش‌های مختلفی برای هموار کردن طیف فرکانسی وجود دارد، اما بهترین روش و ساده ترین آنها، بنام «برش مرکزی» مشهور است.

در روشی که توسط Jsondhi (نام net) ارائه شد، سیگنالی که برش مرکزی داده شده، توسط یک تابع غیر خطی بدست می‌آید.

که در آن در شکل نشان داده شده است.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

یک قسمت از سیگنال صحبت که می‌خواهیم از آن برای ورودی جهت تابع خود همسبتگی استفاده کنیم در شکل نشان داده شد.

برای این فریم، مقدار ماکزیمم دامنه Amax پیدا شده است و دارای آن می‌توان مقدار CL را بدست آورد.

جهت پیدا کردن مقادیر CL راه‌های مختلفی وجود دارد. مثلاً در مقاله Sondhi

مقدار CL از این فرمول بدست می‌آید.

همان طور که در شکل دیده می‌شود، مقادیر نمونه‌هایی که بیشتر از CL هستند، برابر است با مقدار ورودی منهای مقدار سطح برش (CL)، و برای نمونه‌هایی که پائین تر از سطح CL هستند. این مقادیر صفر می‌باشند.

شکل خروجی سیگنال صحبت پس از انجام برش مرکزی را نشان می‌دهد.

در این شکل مشاهده می‌کنید، نقاط قله تبدیل به پالس‌هایی شده‌اند که مانند پالس‌های حنجره عمل خواهند کرد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در شکل تأثیر برش مرکزی در روی محاسبهٔ تابع خود همبستگی نشان داده شده است. شکل 3-a)

همان طور که مشاهده می‌کنید در نقطه پریود فرکانس گام یک قله بسیار قوی مقدار زیاد در تابع خود همبستگی وجود دارد. همچنین پیکهایی وجود دارد که می‌توان از آنها به عنوان نوسانهای ضعیف شده، فیلتر صوتی یاد کرد.

در شکل [3-b] مقدار سیگنال برش داده شده پس از انجام عمل با سطح معین نشان داده شده است. این سطح برابر است با 68٪ ماکزیم مقدار 100 نمونهٔ اول توجه کنید. تمام شکل موج باقی مانده پس از برش، یک سری پالس هستند که در محدودهٔ فرکانس تمام قرار دارند. بنابراین تابع خود همبستگی موج‌ها دارای پیکهایی بمراتب کمتر از حالت قبلی است و بنابراین تصمیم‌گیری بهتر خواهد بود و امکان اشتباه پائین‌تر خواهد آمد.

با نگاه به شکل می‌توان تأثیر سطح برش را مشاهده نمود. به طور خیلی واضح می‌توان فهمید با افزایش سطح برش، تعداد نقاط قله که

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

از سطح برش بیشتر هستند، کاهش خواهد یافت. پالس کمتری در شکل موج خروجی ظاهر خواهد شد. بنابراین تعداد نقاط قلهٔ کمتری در تابع خود همبستگی ظاهر خواهد شد.

این پدیده در شکل 4 نشان داده شده است که در آن می‌توان سطوح برش را مقادیر مختلفی انتخاب نمود.

بطور خیلی واضح مشخص می‌شود، به محض اینکه سطح برش کاهش می‌یابد، تعداد پیکهای بیشتری از سطح برش عبور خواهند کرد و بنابراین تابع خودهمبستگی پیچیده تر خواهد شد.

این مثالها بیان می‌کنند برای پیدا کردن فرکانس گام و یا واکه دار بودن فریم، به صورت دقیق باید سطح برش بالا باشد. همچنین اگر سطح برش خیلی بالا باشد مشکلات زیادی به وجود خواهد آمد.

این امکان وجود دارد که دامنهٔ سیگنال صحبت به طور خیلی زیادی در طول فریم تغییر (به طور مثال در ابتدا یا انتهای سیگنال صحبت واکه‌دار) کند. بنابراین اگر میزان سطح برش بر اساس بیشترین مقدار

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فریم، معین شود، این امکان وجود دارد که قسمتهای زیادی از سیگنال پائین سطح برش بیفتد و از بین برود.

جهت کم کردن اثر دامنه روی سطح برش، sondhi ، پیشنهاد کرد. مقدار CL حدود 30٪ مقدار ماکزیمم فریم باشد. این کار یک مشکل دارد به دلیل پائین بودن سطح CL امکان ایجاد خطا زیاد خواهد شد. ولی باز به دلیل همین پائین سطح برش، امکان از بین رفتن سیگنال‌های ضعیف در هنگام برش کم خواهد شد.

روش دیگر برای افزایش سطح برش بدون از دست دادن اطلاعات زیر، این است که مقدار ماکزیمم را در ابتدا و انتهای فریم بدست بیاوریم. مقدار CL می‌تواند بین 60٪ تا 80٪ مقدار می‌نیمم این دو مقدار باشد.

به طور کلی مراحل الگوریتم را به صورت زیر خلاصه نمود:

1- پیدا کردن مقدار سطح برش با اندازه‌گیری مقدار ماکزیمم در ابتدا و انتهای فریم

2- انجام تابع برش مرکزی

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

3- انجام تابع خودهمبستگی

4- پیدا کردن مقدار ماکزیمم در محدودهٔ فرکانس گام

پس از پیدا کردن ماکزیمم در محدودهٔ فرکانس گام باید تشخیص بدهیم، که آیا فریم واکه‌دار یا خیر.

در صورت واکه دار بودن فریم فرکانس گام تخمین زده شده اعتبار خواهد داشت. در غیر این صورت فرکانس گام معنایی ندارد.

برای تشخیص واکه دار بودن فریم از میزان انرژی آن فریم استفاده می‌کنیم در صورتیکه مقدار ماکزیمم پیک در محدودهٔ فرکانس گام از یک مقدار آستانه بیشتر شد، این فریم واکه‌دار است. این مقدار آستانه از طریق درصدی از انرژی بدست می‌آید.

از آنجا که محدودهٔ فرکانس گام در حدود 50-500Hz می‌باشد. برای دقت بیشتر و از بین بردن تأثیر فرکانسهای‌هارمونیک در هنگام تخمین فرکانس گام می‌توان در مرحلهٔ اول قبل از فریم بندی سیگنال را از یک فیلتر پائین گذر با پهنای باند 900Hz عبور داد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

روش دوم: استفاده از ضرائب کپستروم

در قسمتهای قبل با نحوهٔ استخراج ضرائب کپستروم و نحوهٔ محاسبه آنها آشنا شدیم. همان طور که در شکل مشاهده می‌کنیم در محدودهٔ فرکانس گام در ضرائب کپستروم یک قله وجود دارد و دارای یک مقدار ماکزیمم هستیم. البته این اتفاق فقط برای فریم‌های واکه دار اتفاق می‌افتد. در صورت واکه دار بودن فریم، یک مقدار ماکزیمم به صورت قله در فرکانس گام وجود دارد. از این خاصیت ضرائب کپستروم می‌توان برای بدست آوردن، واکه‌دار یا بی‌واکه بودن فریم و همچنین تعیین فرکانس گام استفاده نمود.

پیدا کردم فرکانس گام و تعیین واکه دار بودن یا بی واکه بودن فریم بر اساس ضرائب کپستروم بسیار آسان می‌باشد. الگوریتم بدین صورت عمل می‌کند. پس از محاسبهٔ ضرائب کپستروم به روش توضیح داده شده توسط FFT، مقدار ماکزیمم این ضرائب در محدودهٔ فرکانس گام پیدا خواهد شد. اگر این مقدار ماکزیمم از یک

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

حد آستانه بیشتر باشد. این فریم، یک فریم واکه‌دار است و نقطه ماکزیمم، یک تخمین خوبی از فرکانس گام را می‌دهد.

در غیر این صورت فریم مذکور بی واکه می‌باشد. به طور معمول اندازهٔ فریم‌ها برای محاسبهٔ کپستروم حدود می‌باشد که در این صورت پارامترهای تحریک زیاد عوض نخواهند شد.

ممکن است در مرحلهٔ اول این فکر به ذهن خطور کند که روش فوق روش بسیار ساده و دقیقی خواهد بود. ولی همیشه مسأله پیدا کردن فریم واکه‌دار به این راحتی نخواهد بود.

وجود یک نقطه ماکزیمم قوی در محدودهٔ 3ms تا 2ms از ضرائب کپستروم بیان‌گر واکه دار بودن فریم است، اما عدم وجود این مقدار یا پائین بودن این مقدار به معنی بی‌واکه بودن فریم همواره نمی‌باشد. بدلیل اینکه وجود نقطه ماکزیمم در محدودهٔ فرکانس گام به عوامل زیادی بستگی دارد، که شامل مقدار فرکانسهای فرمنت و همچنین طول پنجره م یباشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

به دلیل استفاده از پنجرهٔ Hamming و داشتن هم‌پوشانی بین فریم یا، طول پنجره و نقطهٔ قرار گرفتن آن، تأثیر زیادی در دامنهٔ قلهٔ موجود در محدوده فرکانس گام دارد.

برای مثال فرض کنید، مقدار طول پنجره از اندازهٔ دو پریود کمتر باشد، بنابراین نمی‌توان انتظار وجود هیچ نقطهٔ ماکزیمم را در تابع خود همبستگی داشت. بنابراین طول پنجره را همواره طوری تعیین می‌کنند که بتوانند حداقل دو پریود از فرکانس گام را در خود جای دهد. برای فرکانس‌های گام یک پنجره حدوداً 40ms لازم است.

اما همان طور که قبلاً گفتیم جهت ثابت نگه داشتن خاصیت سیگنال صحبت در پنجره، طول پنجره باید تا آنجا که امان دارد کوچک باشد. با افزایش طول پنجره مقدار تغییرات سیگنال صحبت از ابتدا تا انتهای فریم زیاد خواهد شد ممکن است دیگر نتوان آن توسط مدل نمایش داد.

یکی از راههای بهینه کردن طول پنجره، قرار دادن طول پنجره به اندازهٔ پریود فرکانس گام آخرین فریم واکه‌دار قبلی می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پیاده سازی الگوریتم‌ها

در مورد پیاده سازی روش اول نکته قابل اهمیت میزان CL می‌باشد. جهت پیدا کردن مقدار بهینه، ابتدا سعی شد فریمها به صورت دسته به واکه‌دار و بی واکه تقسیم شوند. سپس به کمک مقدار CL سعی شد. تعدا فریمهایی که واکه‌دار بودن داربست شناخته شدند را به ماکزیمم برسانیم.

از طرفی تعداد فریمهایی که به طور اشتباه تعیین شده بودند را کمتر می‌نماییم. بدین ترتیب مقدار بهینه CL را بدست آوریم. پارامتر دیگری که در پیاده سازی الگوریتم فوق مؤثر است.

پیدا کردن یک مقدار آستانه خوب برای انرژی جهت جدا کردن فریم واکه‌دار از بی واکه می‌باشد.

برای پیدا کردن این مقدار به صورت تجربی عمل کرده و مقدار بهینه را بدست آوریم.

در مورد پیاده سازی الگوریتم دوم، باز هم جهت پیدا کردن یک مقدار بهینه میزان آستانه برای مقایسه با ماکزیمم پیک در محدودهٔ فرکانس

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

گام باز هم به صورت تجربی عمل کرده و این مقدار را بهینه بدست آوردیم.

در روش تجربی برای هر دو الگوریتم، صحبت جداسازی فریم واکه‌دار را به کمک چشم و دیدن شکل موج آن فریم و همچنین شنیدن صدای آن فریم مورد بررسی قرار دادیم.

به دلیلی که گوش انسان قادر به شنیدن یک صدا در حدود 20ms نمی‌باشد. به وسیلهٔ یک برنامه، فریم فوق را حدود 2s تکرار نمودیم تا صدای آن فریم واضح و قابل شنیدن باشد. بدین ترتیب به کمک چشم و گوش و از طریق تجربی می‌توان واکه‌دار بودن فریمها را تشخیص داد.

فرکانس فرمنت

فرکانسهای فرمنت اول و دوم و سوم، حاوی اطلاعات زیادی در مورد اصوات صدادار هستند. بطور کلی از طریق فرکانسهای فرمنت می‌توان حروف صدادار را بازشناسی نمود. در حقیقت، فرکانس

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فرمنت فقط برای فریم‌های واکه‌دار معنا پیدا می‌کند. البته در مورد حروف بی صدا نیز فرکانس‌های فرمنت نیز کاربرد دارند.

ما در این پروژه از فرکانسهای فرمنت برای بازشناسی حروف صدادار استفاده می‌کنیم. برای این کار ابتدا توسط الگوریتم‌های توضیح داده شده فریمهای واکه‌دار را پیدا کرده و سپس فرکانسهای فرمنت را پیدا می‌کنیم.

برای پیدا کردن فرکانسهای فرمنت دو روش را امتحان کردیم، در روش اول از حل ریشهٔ معادلات بوجود آمده توسط ضرایب LPC استفاده کرده و فرکانسهای فرمنت را پیدا نمودیم. در روش دوم از طریق جستجوی نقاط ماکزیمم در روی طیف فرکانسی و بدست آوردن فرکانسهای نظیر آن نقاط به این امر پرداختیم.

استخراج فرکانس فرمنت از طریق حل ریشه‌های LPC

در پردازش صوت، فرکانسهای فرمنت، فرکانس رزوناس فیلتر صوتی هستند. تخمین مکان و پهنای باند فرکانس فرمنت کاربرد زیادی دارد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

رایج ترین تکنیک برای استخراج فرکانس فرمنت استفاده از ضرایب فیلتر بدست آمده از تحلیل LPC یک فریم صوتی می‌باشد. وقتی چند جمله‌ای پیشگویی کننده بدست آمد. فرکانی فرمنت را از دو طریق می‌توان بدست آورد. اول از طریق جستجو نقطه ماکزیمم (peak picking) در روی منحنی پاسخ فرکانسی فیلتر، دوم – حل معادله که در آن هر زوج مختلط که ریشه‌های معادله هستند جهت پیدا کردن فرکانس فرمنت و مقداری پهنای باند آن استفاده خواهد شد.

تعدادی از منابع معتبر در پردازش صوت، روش تبدیل یک ریشه مختلط – همراه فرکانس نمونه‌برداری به فرکانس فرمنت و پهنای باند به طریق زیر مطرح کرده‌اند.

با فرض اینکه چند جمله‌ای کد کننده به مانند یک فیلتر درجه دوم مقام قطب باشد فرکانس فرمنت و پهنای باند آن از طریق بدست می‌آید.

آنالیز پیشگویی خطی (LPC)

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

آنالیز LPC روی قطعاتی از سیگنال صحبت انجام می‌گردد. اساس LPC بر پیشگویی خطی است. در مدل پیشگویی خطی، فرض می‌گردد که سیگنال صحبت یک فرایند بازگشتی است.

در رابطه ، تولید شده توسط مدل، سیگنال تحریک، پارامترهای پیشگویی و P مرتبه پیشگوکننده می‌باشد. در این عبارت، G پارامتر بهره می‌باشد که جهت انطباق انرژی صحبت تولید شده با صحبت اولیه بکار می‌رود. در حوزه Z ، تابع تبدیل فیلتر LPC بصورت زیر تعریف می‌گردد.

و A(z)، فیلتر پیشگویی می‌باشد.

و نیز برابر است با

در تحلیل LPC صحبت، پارامترهای مربوط به هر دو مدل تحریک و مدل تولید صحبت توسط سیگنال ورودی تخمین زده می‌شود. رابطه (4. 2) نشان می‌دهد، تبدیلات مربوط به تابع تبدیل فیلتر مدل تولید

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

صحبت و مدل تحریک در یکدیگر ضرب می‌شوند. از دیدگاه حوزه فرکانس، این بدان معنی است که مدل تولید صحبت اطلاعات مربوط به پوش طیف و مدل تحریک نیز اطلاعات مربوط به جزئیات طیف صحبت را در اختیار ما می‌گذارد.

مدل تولید صحبت

در آنالیز LPC، مدل تولید صحبت توسط یک فیلتر تمام قطب H(z) نمایش داده می‌شود. بدلیل آنکه صحبت یک فرآیند متغیر با زمان است. H(z) باید یک فیلتر متغیر با زمان باشد که ضرایب آن با زمان تغییر می‌کنند. و بدلیل آنکه صوت تولید شده توسط مدل صحبت دارای تغییرات ملایمی‌است، می‌توان صحبت را فرایند اتفاقی در نظر گرفت که خصوصیات آن به کندی تغییر می‌یابند. این فرض باعث می‌شود که از فرضیه اساسی ایستایی زمان – کوتاه در تحلیل LPC استفاده گردد. این فرضیه بیان می‌دارد که سیگنال صحبت را می‌توان در داخل یک پنجره به طول L که L بقدر کافی کوچک باشد، ایستا تصور نمود. این فرضیه امکان مدل کردن صحبت، توسط یک سری

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فیلترهای ثابت H(z)، که ضرایب آنها در داخل پنجره ثابت باشد را فراهم می‌نماید. ضرایب A(Z) ، یعنی به ازای با تحلیل پیشگویی خطی سیگنال صحبت حاصل می‌شود. از طرق مختلفی می‌توان به تحلیل پیشگویی خطی نگریست، یکی از آموزنده ترین این مدلها شکل 3-2 می‌باشد.

با توجه به این شکل که تخمینی از سیگنال صحبت می‌باشد توسط پیشگوی خطی A(Z) و سیگنال صحبت ورودی S[n] حاصل می‌گردد. با کم کردن سیگنال حاصل از سیگنال صحبت اولیه، e[n] سیگنال خطا، که سیگنال مانده پیشگویی خوانده می‌شود حاصل می‌گردد. سیگنال خطا را می‌توان با فیلتر معکوس زیر بدست آورد.

ضرایب پیشگو بطور تقریبی با کمینه کردن انرژی مانده پیشگویی، E، حاصل می‌گردد. در این عبارت e[n] خروجی فیلتر معکوس است.

روشهای متفاوتی جهت بدست آوردن ضرایب پیشگو وجود دارد از جمله مشهورترین این روشها، روش همبستگی و روش کواریانس

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

است که هر دو از تکنیکهای پایه در حوزه زمان می‌باشند و بر اساس معیار کمترین مجذور عمل می‌نمایند.

روش همبستگی

در روش همبستگی، یک پنجره متحرک، صحبت را به قطعاتی تقسیم می‌نماید. این فرایند در شکل 4-2 نشان داده شده است. در هر محل از پنجره، معمولاً قسمتی به طول 10 تا 30 میلی ثانیه از سیگنال صحبت جدا می‌گردد تا یک قطعه تحلیلی از سیگنال ایجاد گردد. طول سیگنال حاصل نامحدود می‌باشد ولی در هر جا به غیر از محل پنجره صفر خواهد بود. پس امکان محاسبه تابع همبستگی صحیح برای تمام سیگنال وجود دارد. i امین قطعه تحلیلی بصورت زیر بدست می‌آید.

امین پنجره تحلیلی می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

I فاصله بین هر دو قطعه تحلیلی است. همبستگی قطعه تحلیلی بصورت زیر تعریف می‌گردد.

تابع پنجره w[n] ، بگونه‌ای انتخاب می‌شود که دارای کاهش تدریجی در لبه‌ها باشد (مانند پنجره همینگ) طول پنجره، L، در نظر گرفته می‌شود.

کمینه نمودن متوسط انرژی مانده، معادله نرمال ماتریسی بدست خواهد داد.

که بردار ضرایب LPC می‌باشند، و R ماتریس ضرایب همبستگی می‌باشد.

و ماتریس R، ماتریس توپلیتز متقارن می‌باشد که می‌توان آنرا بطور کارا توسط الگوریتم داربین حل نمود. این الگوریتم بازگشتی می‌باشد و از ساختار توپلیتز ماتریس R جهت حل کارای ضرایب LPC استفاده می‌نماید. این الگوریتم را می‌توان در مجموعه معادلات زیر خلاصه نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

معادلات ( ) تا () بطور بازگشتی برای حل می‌گردد. ضرایب برای دارای اطلاعاتی مشابه با ضرایب LPC می‌باشند و ضرایب انعکاسی یا ضرایب همبستگی جزئی نامیده می‌شوند. همانطور که در شکل 5-2 نشان داده شده، امکان پیاده سازی مستقیم فیلتر مدل تولید صحبت با عبارات پارکور وجود دارد.

کمیت انرژی خطای پیشگویی با پیشگویی مرتبه i می‌باشد. از طرفی یک کمیت مثبت است. معادله (2 .13 .5) نشان می‌دهد که تمامی‌ضرایب پارکو دارای مقدار کمتر از یک می‌باشند. یعنی،

از آنجایی که فیلتر مدل تولید صحبت بازگشتی است، پایداری یک مسئله مهم است. واضح است که شرط معادله (2 .14) شرط لازم و کافی برای پایداری فیلتر مدل تولید صحبت می‌باشد.

روش کوواریانس

در روش کواواریانس سیگنال صحبت به تنهایی توسط پنجره تقسیم نمی‌گردد. بلکه دنباله خطای پیشگویی e[n] (شکل 3-2) توسط پنجره

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تقسیم می‌گردد و انرژی آن کمینه می‌شود. بنابراین کمیت تعریف شده

متناظر با ضرایب پیشگویی کمینه می‌گردد. نتایج کمینه سازی توسط تساوی ماتریس زیر بیان می‌گردد.

که بردار ضرای پیشگو می‌باشد و ماتریس متفاوتی است که بصورت زیر تعریف می‌شود.

و بدلیل اینکه یک ماتریس توپلیتز نمی‌باشد، نمی‌تواند بصورت کارایی همانند معادلات نرمال در روش همبستگی حل گردد، ولی نسبتاً معادلات کارایی جهت حل معادلات متقارن وجود دارند.

مقایسه روش‌های همبستگی و کوواریانس

هر دو روش همبستگی و کوواریانس مجموعه ضرایب پیشگوی مربوط به فیلتر مدل تولید صحبت را تولید می‌کنند، و هر دو نیز دارای الگوریتم‌های کارایی می‌باشند. تفاوت اصلی دو روش در طریقه اعمال پنجره بر روی سیگنال صحبت می‌باشد. در روش همبستگی، ابتدا پنجره روی سیگنال صحبت اعمال می‌گردد و سپس

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تخمین زمان کوتاه (همراه با کمی تغییر) از صحبت اولیه صورت می‌گیرد. بنابراین ابتدا همبستگی واقعی محاسبه می‌گردد و سپس ضرایب بهینه LPC از این قطعه صحبت تغییر یافته محاسبه می‌گردد. از لحاظ ریاضی، این یک فرایند مهارشدنی است، که تضمین می‌کند فیلترهای تولید صحبت، خوش رفتار (پایدار) باشند. در کل، روش همبستگی از کارایی خوبی برخوردار است.

از طرف دیگر، روش کوواریانس، بطور مستقیم پنجره را بر روی سیگنال صحبت اعمال نمی‌کند لذا تغییر و یا انحرافی در سیگنال صحبت قبل از شروع پردازش ایجاد نمی‌گردد. بلکه در این روش پنجره بر روی خط اعمال می‌شود. این روش بطور بالقوه کارایی بالایی را باید بدست آورد زیرا پردازش بر روی سیگنال تغییر یافته صورت نمی‌گیرد ولی اعمال پنجره روی سیگنال تغییر یافته صورت نمی‌گیرد ولی اعمال پنجره روی سیگنال خطا، خود می‌تواند انحراف و تغییرات جزئی ناخواسته‌ای ایجاد کند. لذا ممکن است فیلترهای تولید شده برای مدل تولید صحبت، ناپایدار باشند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مرتبه پیشگو

از آنجایی که انرژی مانده با هر بار تکرار الگوریتم بازگشتی داربین کاهش می‌یابد پس انرژی خطای پیشگویی با افزایش قطبهای فیلتر، یعنی P، کاهش خواهد یافت و مقدار محاسبات لازم نیز افزایش می‌یابد، لذا ایجاد تعادل بین این دو معیار ضروری می‌باشد. یک راه جهت تعیین P تعریف یک حد آستانه مانند است بطوریکه

مقدار باید بگونه‌ای تعیین گردد که پس از برقرار گشتن شرط رابطه (2 .18)، تغییرات خطا محسوس نباشد. در اینصورت P=pیک انتخاب خوب خواهد بود. در صحبت، از دو قطب (یک زوج قطب) برای هر فرمنت استفاده می‌شود. سیگنال صحبت دارای صفر نیز می‌باشد. اما تأثیر چندانی ندارد لذا معمولاً در تابع تبدیل مدل تولید صحبت از آن صرفنظر می‌شود. در عمل، برای صحبت در محدوده 8KHZ، از پیشگوهایی با مرتبه بین 10 تا 16 استفاده می‌گردد.

کوانتیزاسیون برداری

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در فصلهای قبل در مورد چگونگی محاسبه پارامترهای کپسترال و دلتا کپسترال بحث نمودیم. از مزیتهای مهم استفاده از این پارامترها به عنوان مشخصه یک قطعه تحلیلی، علاوه بر حساسیت کم پارامترها نسبت به تغییرات ناخواسته، فشرده سازی اطالعات نیز می‌باشد. بعنوان مثال اگر یک قطعه تحلیلی شامل 240 نمونه و طول بردار مشخصه نیز 24 باشد (تشکیل یافته از 12 پارامتر کپسترال و 12 دلتا پارامتر کپسترال) آنگاه حجم فشرده سازی اطالعات به نسبت یک به ده خواهد بود.

هر چند که این مقدار از فشرده سازی جهت بسیاری از روشهای درک صحبت مناسب می‌باشد اما برای روشهای آماری همچون مدل مارکف مخفی (HMM) که مستلزم تخمین تابع چگالی احتمال می‌باشد منجر به افزایش هزینه محاسبات در مرحله یادگیری و تشخیص می‌شود. لذا جهت کاهش هزینه محاسبات، می‌توان بجای استفاده از بردار مشخصه با مقادیر پیوسته پارامترهای کپسترال و دلتا کپسترال، از بردار مشخصه با مقادیر گسسته این پارامترها

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

استفاده نمود که در نتیجه هزینه محاسبه تابع چگالی احتمال را به محاسبه احتمال هر یک از بردارهای گسسته کاهش خواهد داد.

از طرفی کوانتیزه نمودن جداگانه هر یک از پارامترها، راه حل خوبی نخواهد بود. زیرا که اگر تعداد متوسط نقاط کوانتیزاسیون برای هر پارامتر را هشت نقطه فرض می‌کنیم تعداد کل بردارهای کوانتیزه به 824 خواهد رسید. استفاده از این تعداد بردار، مشکلاتی را به وجود خواهد آورد. اولین مشکل مربوط به حجم حافظه مورد نیاز می‌باشد و دوم اینکه احتمال وقوع بعضی از بردارها صفر خواهد بود. جهت غلبه بر این مشکلات، کوانیزاسیون برداری پیشنهاد می‌گردد. در کوانتیزاسیون برداری، بجای کوانتیزه نمودن مجزای‌های عنصر از بردار، تمامی‌عناصر بردار با هم و در یک زمان کوانتیزه می‌گردند. عبارت بهتر بجای انتخاب نقاط کوانتیزاسیون در فضای یک بعدی و یافتن نقاط متناظر با عناصر حاصله از کوانتیزاسیون در فضای n بعدی (برای بردار مشخصه با n عنصر)، از همان بدو امر، نقاط کوانتیزه از فضای n بعدی انتخاب می‌گردند. روشهای متعددی جهت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

انجام کوانتیزاسیون برداری وجود دارد. از میان روشهای موجود، روش LBC را بر می‌گزینیم.

روشهای متعددی جهت انجام عملیات کوانتیزاسیون برداری وجود دارند. ولی آنچه که در تمامی‌این روشها مشترک می‌باشد، یافتن نقاط کوانتیزاسیون توسط یک مجموعه داده آموزشی می‌باشد. این عمل با حداقل نمودن خطای کونتیزاسیون به ازای بردار ورودی متعلق به مجموعه آموزش و بردار کوانتیزه مربوط به آن صورت می‌گیرد. خطای کوانتیزاسیون را می‌توان فاصله اقلیدسی بین دو بردار Y و X تعریف نمود.

از میان روشهای موجود، از روش LBG در این پروژه استفاده شده است. این روش سعی در یافتن L بردار کوانتیزه بهینه دارد. اگر از میان تمام مجموعه‌های L تایی از بردارهای کوانتیزه بتوان مجموعه‌ای یافت که خطای حاصل از کوانتیزاسیون مجموعه بردارهای آموزشی توسط آن حداقل باشد آنگاه L بردار مجموعه مذکور بهینه خواهد بود. در عمل جهت یافتن L بردار کوانتیزه بهینه

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

از مقدار متوسط خطای کوانتیزاسیون استفاده می‌شود اگر تعداد بردارهای آموزشی به اندازه کافی باشد می‌توان امید داشت که حداقل نمودن متوسط خطای کوانتیزاسیون، معادل با حداقل نمودن خطای کوانتیزاسیون باشد. جهت رسیدن به یک مجموعه L تایی از بردارهای کوانتیزه بهینه، که دارای حداقل متوسط خطای کوانتیزاسیون، باشد باید از دو شرط لازم زیر استفاده نمود:

1- کوانتیزاه کننده برداری که حداقل خطای کوانتیزاسیون را دارد بعنوان بردار کوانتیزاسیون استفاده گردد.

2- کوانتیزه کننده برداری انتخاب شده جهت یک زیر مجموعه از مجموعه بردارهای آموزشی باید دارای حداقل متوسط اعوجاج باشد. بعنوان مثال اگر کوانتیزه کننده برداری منتخب برای زیر مجموعه باشد و از فاصله اقلیدسی بعنوان معیار اعوجاج استفاده شود آنگاه از رابطه زیر حاصل می‌گردد.

نماد نشان دهنده تعداد اعضای زیر مجموعه می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

الگوریتم LBG بر اساس دو شرط مذکور طراحی شده است. مراحل این الگوریتم بصورت زیر می‌باشد.

الگوریتم LBG

1- بطور تصادفی و یا با یک روش مناسب مقادیر اولیه بردارهای کوانتیزه را تعیین نمایید.

2- مجموعه بردارهای آموزشی را با توجه به شرط اول به زیر مجموعه‌های تقسیم نمایید.

3- ، با توجه به شرط دوم، بردارهای کوانتیزه جدید را بدست آورید.

4- اگر نسبت مقدار متوسط خطای حاصل از مرحله m به مرحله m-1 کمتر از یک حد آستانه از پیش تعیین شده بود، آنگاه الگوریتم خاتمه یافته است در غیر اینصورت الگوریتم را از مرحله 2 تکرار کنید.

کاهش مقدار متوسط خطا در هر بار تکرار الگوریتم امری واضح است زیرا با شروع مجدد الگوریتم از مرحله دوم، بردارهای مجموعه آموزشی بگونه‌ای در زیر مجموعه‌های متناظر با هر بردار کوانتیزه

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

جابجا می‌گردند که حداقل خطا را داشته باشند. با این تقسیم‌بندی جدید، خطای کوانتیزاسیون حاصل از بردار موجود در مجموعه آموزشی یا تغییر نمی‌کند و یا اینکه کاهش می‌یابد در نتیجه مقدار متوسط خطا نیز متناظر با خطای کوانتیزاسیون تغییر خواهد یافت. با اجرای مرحله سوم نیز بردارهای کوانتیزه مربوط به هر زیر مجموعه، برابر متوسط بردارهای آن زیر مجموعه قرار داده می‌شود که در نتیجه باعث حداقل شدن خطا در آن زیر مجموعه و بدنبال آن باعث حداقل شدن متوسط خطا خواهد گردید. اگرچه الگوریتم LBG یک الگوریتم موفق جهت انجام کوانتیزاسیون‌برداری می‌باشد ولی باید توجه داشت که این الگوریتم مانند تمام الگوریتم‌های جستجوی محلی، تنها قادر به یافتن بهینه محلی می‌باشد و قابلیت جستجوی کلی و یافتن بهینه مطلق را ندارد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فصل سوم: انحراف پویای زمانی (Dynamic time warpping)

مقدمه:

بازشناسی گفتار به صورت خودکار دارای جنبه‌های کاربردی فراوانی است و سالها روی آن تحقیق شده است. یک روش شناخته شده خوب در حوزهٔ بازشناسی گفتار بر مبنای این اصل کار می‌کند که برای هر کلمه یک یا دو الگوی اکوستیکی ذخیره می‌شود. روند بازشناسی گفتار به این صورت است که صدای ورودی با الگوهای از پیش ذخیره شده، تطبیق داده می‌شود، الگویی که دارای کمترین فاصلهٔ اندازه‌گیری شده با صدای ورودی باشد به عنوان کلمهٔ تشخیص داده شده، معین می‌گردد. یک الگوریتم که برای پیدا کردن بهترین تطابق، یا همان کوچکترین فاصله بکار می‌رود انحراف پویای زمانی می‌باشد. این الگوریتم بر مبنای برنامه ریزی پویا کار می‌کند.

هدف این فصل معرفی الگوریتم انحراف پویای زمانی، ویژگی‌ها و انواع آن می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

صوت یک پدیدهٔ وابسته به زمان می‌باشد. ممکن است چندین کلمه بیان شده توسط یک نفر یا نفرات مختلف طول‌های مختلفی داشته باشند و گفتارهای مربوط به یک کلمه با طول مساوی ممکن است در وسط کلمه با هم فرق کنند که این به دلیل تلفظ بخشهای مختلف کلمه با سرعت‌های متفاوت می‌باشد.

برای بدست آوردن فاصلهٔ نهایی بین دو الگوی صحبت (که هر کدام شامل یک رشته بردار ویژگی می‌باشند)، باید هم ترازی زمانی صورت پذیرد.

- هم ترازی‌سازی زمانی و نرمالیزاسیون

دو الگوی X و Y را در نظر بگیرید، که هر کدام از این الگوها شامل بردارهای و می‌باشند که در آن نشان دهندهٔ پارامترهای زمان کوتاه اکوستیکی صوت می‌باشند. در این جا ماتریس ویژگی هیچ تفاوتی نمی‌کند، و می‌تواند هر مشخصهٔ طیف فرکانسی از سیگنال صحبت باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

ما از و برای نشـان دادن انـدیکی مربـوط بـه زمـان X و Y اسـتفاده می‌کنیم. نیازی به دانستن نمی‌باشد. عـدم تشـابه بـین X و Y توسـط تابع انـدازه‌گیـری اعوجـاج طیـف یـک فـریم زمـان کوتـاه مـدت بیـان می‌شود.

این تابع را با نشان می‌دهیم که برای سادگی به صوت بیان می‌کنیم.

که در آن می‌باشد.

شاید، ساده ترین روش برای حـل مسـأله هـم تـراز سـازی زمـانی و نرمالیزاسـیون، بکـار بـردن روش نرمالیزاسـیون خطـی باشـد. در نرمالیزاسیون خطی، عدم تشـابه و اخـتلاف بـین X و Y بـه صـورت زیر به سادگی بیان می‌شود.

(به دلیل اینکه و هر دو عدد صحیح هسـتند بایـد عملیـات round-off صورت پذیرد)

عملیات جمع می‌تواند از تا صورت بپذیرد کـه ایـن بسـتگی بـه جهـت دلخواه در نرمالیزاسیون دارد. در رابطهٔ بالا ما از d برای نشـان دادن تابع اعوجاج بین دو جملهٔ طیف فرکانسی استفاده نمودیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

نرمالیزاسیون خطی، الزاماً فرض می‌کنید تغییرات نرخ گفتن متناسب است با زمان و طول کلمه، و از خود صوت مستقل است. بنابراین برای اندازه‌گیری اعوجاج، فقط تفاوت بین نقاطی که روی خط راست قطری مستطیل قرار دارند محاسبه خواهد شد. (شکل). هر نقطه روی قطر مستطیل بیان گر d(ix,iy) می‌باشد که این مقدار فاصلهٔ طیفی بین X و Y در آن فریم بیان می‌کند.

مدل بیان شده در هم ترازسازی بین دو الگو زیاد واقعی نمی‌باشد، و برای الگوهای دلخواه جهت نرمالیزاسیون و هم ترازسازی باید راههای واقعی تری را جستجو نمود.

یک راه عمومی‌تر هم ترازسازی زمانی و نرمالیزاسیون استفاده از دو تابع انحراف دهنده ، که هر کدام از آنها به ترتیب می‌توانند اندیکس دو الگوی صوتی ix، iy را به یک محور زمانی نرمال k نسبت دهند.

مقدار نهایی عدم تشابه (X,Y) می‌تواند براساس تابع انحراف دهنده بیان شود. این تابع حاصل جمع اعوجاج در کل کلمه می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

که در آن دوباره بیان کننده میزان اعوجاج زمان کوتاه طیف ، می‌باشد. m(k) یک عدد مثبت که نشان دهندهٔ ضرایب وزن دهی در مسیر می‌باشد و فاکتور نرمالیزه کنندهٔ مسیر است. در شکل [1] به طور کلی نرمالیزاسیون زمانی نشان داده شده است شکل [1-b] نشان دهندهٔ مسیری است که در آن مقدار d(x,y) محاسبه شده است.

نقاط روی محور افقی در شکل [1-a] نشان دهندهٔ میزان k هستند که از 1 شروع می‌شود و تا T افزایش می‌یابد، در اینجا مقدار نرمال شده طول کلمه می‌باشد.

در شکل [1-a] مقدارهای ix و iy به عنوان تابعی از مقدار k در واحد زمان نشان داده شده‌اند.

برای کامل کردن مفهوم میزان عدم تشابه بین دو الگوی (x,y) ، ما باید مفهوم راه را که در معادلهٔ (1) نشان داده بیان کنیم. می‌توان گفت تعداد زیادی تابع انحراف دهنده وجود دارد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اما مسئله مهم در اینجا این است که کدام مسیر را انتخاب نمود که بتوان میزان عدم تشابه را در کل مسیر با یک دقت خوبی اندازه‌گیری کرد.

یک راه ساده و خوب ، تعریف کردن میزان عدم تشابه $d(x,y)$ بر مبنای کوچکترین مقدار در بین مسیرهای مختلف می‌باشد.

که در آن مسیر باید یک سری شرایط را دارا باشد که دربارهٔ آن بحث خواهد شد. بطور کلی تعریف معادلهٔ (2) وقتی درست خواهد بود که X و Y هر دو یک کلمه باشند. به خاطر اینکه انتخاب بهترین مسیر، یعنی کم کردن میزان اعوجاج عدم تشابه در طول مسیریکه هم ترازسازی زمانی شده است می‌باشد. این کار جهت از بین بردن تأثیر نرخ سرعت گفتار گوینده‌های مختلف می‌باشد.

نکته اساسی در پیدا کردن بهترین هم ترازسازی بین دو الگو این است که این کار برابر است با پیدا کردن بهترین مسیر در بین grid mapping بردارهای مشخصهٔ اکوستیکی الگوی اول با بردارهای مشخصه اکوستیکی الگوی دوم می‌باشد. برای پیدا کردن این مسیر،

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

باید مسأله محاسبهٔ کمترین عدم تشابه بین دو الگوی صحبت را حل نمود.

شکل معادلهٔ اول بیان می‌کند که در آن میزان اعوجاج در هر فریم جمع خواهد شد. بیان می‌کند، که یک راه حل خوب برای مسأله پیدا کردن کمترین اعوجاج می‌باشد.

مروری بر Dynamic programing

DP یکی از پرکاربردترین ابزارها برای فعالیتهای جستجو جهت حل مسائل تصمیم‌گیری ترتیبی می‌باشد. برای نشان دادن قابلیت‌های آن در تشخیص گفتار و علی الخصوص در مسأله هم ترازسازی زمانی و نرمالیزاسیون، ما دربارهٔ دو مسأله مهم که الگوریتم DP از آن استفاده می‌کند بحث خواهیم کرد.

اولین مسأله، مسأله پیدا کردن مسیر بهینه می‌باشد که بدین صورت می‌توان آن را بیان کرد. N نقطه را در نظر بگیرید. برای هر دو نقطه یک عدد مثبت بنام وجود دارد که بیان کننده میزان هزینه حرکت از نقطه به نقطه در یک مرحله می‌باشد. مسأله در اینجا پیدا کردن

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

کمترین هزینه برای رشته‌ای از حرکات از نقطه به نقطه است. در اینجا محدودیت مرحله وجود ندارد. در شکل (3) این مسأله نشان داده شده است.

بدلیل اینکه در این نوع مسأله تعداد حرکات گفته نشده و محدودیتی در تعداد مراحل وجودندارد. ما این نوع مسأله را مسأله تصمیم آسنکرون ترتیبی می‌نامیم.

در اینجا مسأله این است که کدام روش باعث پیدا کردان کمترین هزینه در حرکت از نقطه به نقطه خواهد شد. اصول اولیه بهینه سازی که بر مبنای الگوریتم‌های محاسباتی برای بهینه کردن مسائل فوق می‌باشد، مطابق با گفته‌های Bellman خواهد بود.

اصول بهینه کردن دارای ویژگی است که می‌گوید، مهم نیست حالت اولیه کجا است و تصمیم‌ها برای حرکت چیست. اما تصمیمات باقی مانده و بعدی باید با توجه به حالت و تصمیم اولیه با تشکیل حالت بهینه را بدهند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای اینکه اصول Bellman را به صورت یک سری معادلات قابل محاسبه برای الگوریتم نوشتن تبدیل کنیم. در نظر بگیرید ما یک حرکت را از نقطه به نقطه میانی در یک یا چند مرحله طی می‌کنیم. کمترین هزینه همان طور که تعریف شده می‌باشد از آنجائیکه حرکت از نقطه به در یک مرحله هزینهٔ را در برخواهد داشت. مسیر بهینه‌ای که می‌گوید از کدام نقطه میانی می‌توان کمترین هزینه را در حرکت به سمت j پیمود به صورت زیر بیان خواهد شد.

برای عمومی‌کردن معادله (2) برای همهٔ حالات و جهت داشتن یک رشتهٔ از حرکات از نقطه i به j می‌توان معادله را به صورت زیر بیان نمود:

که در آن کمترین هزینه در حرکت از نقطه i به j بدون محدودیت مراحل می‌باشد. معادلهٔ (3) بیان می‌کند، هر حرکت جزئی در حرکت از نقطه i به j و یا هر نقطه میانی، باید اصول بهینه بودن را رعایت کند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای معلوم کردن کمترین هزینه در بین مسیرهای بین نقطه i و j بدون محدودیت مراحل، از برنامه DP ساده زیر می‌توان استفاده نمود.

در اینجا بهترین مسیر یا مرحله حرکت برای حرکت از نقطه i به L می‌باشد. تعداد ماکزیمم حرکتها در مسیر می‌باشد.

مسأله دوم DP، تصمیم سنکرون ترتیبی می‌باشد که با مدل آسنکرون در تعداد مراحل پردازش فرق می‌کند. در این روش برای پیدا کردن بهترین مسیر در حرکت از نقطه i به j محدودیت مراحل وجود دارد و باید در تعداد معینی از مراحل مثلاً M حرکت از نقطه i به j رسید و هزینهٔ حرکت فوق را می‌نامیم. (توجه کنید اگر M به اندازهٔ کافی بزرگ باشد، رشتهٔ حرکات می‌تواند پریودیک باشد).

مفهوم الگوریتم به سادگی توسط شکل (4) قابل توضیح است. N نقطه به صورت عمودی قرار دارند و M حرکت به صورت افقی در شکل کشیده شده است. چون در هر نقطه تعداد N مسیر برای حرکت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

وجوددارد بنابراین تعداد کل حرکات در یک مرحله می‌باشد. بنابراین تعداد نقاطی که تشکیل یک مسیر را می‌دهند و نقطه i توسط M حرکت به نقطه j وصل می‌کنند می‌باشد.

قواعد بهینه بودن مسیر در اینجا نیز صادق است و می‌تواند کاربرد داشته باشد، بعد از m حرکت m<M، مسیر در نقطه L به پایان می‌رسد، که L می‌تواند باشد، که مقدار هزینه حرکت آن.

فرض کنید حرکت به نقطه ختم خواهد شد. بنابراین مشابه معادله (2) داریم.

معادلهٔ 5 یک حالت بازگشتی را توصیف می‌کند که امکان جستجوی بهترین مسیر را به صورت پیوسته و در حالت پیشرو می‌دهد. درست است که N مسیر برای رسیدن به نقطه 1 وجود دارد. اما قانون بهینه بودن مسیر بیان می‌کند، تنها باید بهترین مسیر طبق معادلهٔ در نظر گرفته شود.

الگوریتم می‌تواند به صورت زیر خلاصه شود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

به دلیل بهینه سازی، الگوریتم فقط N مسیر را دنبال می‌کند. برای پیدا کردن هم ترازسازی و نرمالیزاسیون، چون در آنها مسأله عدم تشابه الگوها مطرح است. پیدا کردن بهترین مسیر طبق (3) مورد نیاز می‌باشد.

کاربردهای تصمیم‌گیری سنکرون و آسنکرون ترتیبی کاربردهای فراوانی در پردازش گفتار و تشخیص گفتار دارد.

محدودیتهای نرمالیزاسیون زمانی

چند محدودیت در هنگام تابع انحراف دهنده لازم است. عدم وجود محدودیت در معادله ممکن است به تطابق غلطی بین دو کلمه که از لحاظ معنایی شبیه هم نیستند، منجر شود. مثلاً دو کلمه "we" و "you" را در نظر بگیرید، اگر هیچ گونه محدودیتی در روی و اعمال نشود و برگشت در زمان به عقب امکان داشته باشد، میزان عدم تشابه اندازه‌گیری شده برای این دو کلمه می‌تواند بسیار کوچک باشد، و مفهوم بازشناسی گفتار را دچار مشکل کند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

محدودیتهای تابع انحراف دهنده که برای هم ترازسازی زمانی بین دو کلمه معقول و لازم به نظر می‌آید شامل:

محدودیت در نقطه انتهایی

یکنواخت بودن و هموار بودن مسیر

پیوستگی محلی

مسیر سراسری

وزن دهی شیب منحنی مسیر

در قسمتهای آینده ما دربارهٔ این محدودیتها بحث خواهیم کرد.

محدودیتهای نقطه انتها

وقتی که الگوی صحبتی‌هایی که می‌می‌خواهیم مقایسه کنیم کلمات مقطع باشند، معمولاً آنها دارای تعریف مشخصی از ابتدا و انتها می‌باشند و نقاط ابتدا و انتهای آنها معلوم است. این نقاط طبق الگوریتم‌های توضیح داده شده در فصل قبل می‌تواند بدست آید.

برای نرمالیزاسیون بنابراین محدودهٔ کلمه معلوم است. محدودیتهای که برای تابع انحراف دهنده از لحاظ ابتدا و انتها وجود دارد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

ما در اینجا فرض می‌کنیم اولین فریم صحبت به عنوان یک (1) شناخته می‌شود. در مواقعی که به دلیل وجود نویز انتها و ابتدا را نمی‌توان بدست آورد. محدودیتهای گفته شده در بالا باید تغییر داده شود تا این عدم اطمینان از مکان ابتدا و انتها در نظر گرفته شود.

شرایط یکنواختی

همان طور که قبلاً بحث شده نحوهٔ قرار گرفتن اطلاعات طیف فرکانسی در الگوی سیگنال صحبت، از لحاظ معنای زبان‌شناسی بسیار مهم می‌باشد. برای حفظ ترتیب قرار گرفتن فریمها که حاوی اطلاعات مربوط به طیف هستند در هنگام عملیات نرمالیزاسیون لازم است که محدودیتهایی راجع به یکنواخت کردن مسیر اعمال کنیم.

همان طور که در شکل 2 نشان داده شده، محدودیت یکنواخت بودن مسیر می‌گوید در تمام مسیرهایی که مقدار محاسبه می‌شوند، هیچ مسیری نباید وجود داشته باشد که در آن شیب منحنی آن منفی باشد. این محدودیت باعث حذف کردن امکان بازگشت به عقب در محور زمان در هنگام عملیات انحراف دهی می‌شود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

محدودیت پیوستگی محلی

در سیگنال صحبت، تنها وجود بعضی از است که موجبات تشخیص درست کلمه را از سایر کلمات مهیا می‌سازد. از آنجا که هدف در نرمالیزاسیون زمانی پیدا کردن بهترین نقاط بین دو کلمه می‌باشد. بنابراین نباید این کار منجر به حذف هیچ یک از اطلاعات صوتی شود. برای اطمینان حاصل کردن از صحبت انجام عملیات زمانی و جلوگیری از بین رفتن اطلاعات، ما عموماً یک سری قوانین مربوط به محدودیتهای پیوستگی محلی را اعمال می‌کنیم، محدودیتهای مربوط به پیوستگی محلی می‌تواند اشکال مختلفی داشته باشد. یک نمونه که توسط sakoe pchiba ارائه شد بدین فرم است.

مشخص کردن چنین محدودیتهایی اغلب بسیار پیچیده می‌باشد، و بنابراین بسیار ساده است که جهت مشخص کردن آنها را به شکل مقداری که در مسیر افزایش می‌یابند، بیان کنیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بنابراین ما مسیر را که شامل تعدادی از حرکت‌ها از نقطه مبدأ به مقصد است چنین بیان می‌کنیم که در آنها هر حرکت توسط مقدار افزایش مختصاتی بیان شده.

در شکل ، 3 مسیر ، ، نشان داده شده است که آنها را می‌توان بصورت

را مشخص نمود. مسیر نشان داده شده در شکل را می‌توان به صورت زیر نشان داد.

برای مسیری که از نقطه (1 و 1) شروع می‌شود ما به آن k=1اختصاص می‌دهیم (طبق معادله 4.136) ما معمولاً برای نقطه شروع مقدار p=q-1 قرار می‌دهیم.

بنابراین اگر مسیر در نقطه (Tx,Ty) پایان یابد داریم:

با استفاده از علامت گذاری بالا، می‌توان انواع مختلف محدودیتهای پیوستگی محلی را بسادگی تعریف نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

جدول تعداد زیادی از محدودیت‌های محلی را نشان می‌دهد. نحوهٔ بیان این محدودیتهای محلی بصورت مسیرهای مجاز برای رسیدن به نقطه (ix,iy) می‌باشد.

نوع محدودیتها درست مانند حالتی است که ما در معادله بیان کردیم. (بطور خیلی واضح و) هیچ وقت همزمان اتفاق نمی‌افتند). نوع II و III نیز این مسأله صادق است ولی در نوع دوم و هر دو، دارای دو حرکت می‌باشند و بنابراین دو دفعه بر مقدار آنها افزوده شده. در حالیکه و در حالت یک مسیر تک حرکتی هستند.

دیگر محدودیتها نیز می‌تواند به همین طریق مورد بررسی قرار گیرد، تنها یک استثناء وجود دارد و آن هم بوسیلهٔ Itakura بیان شده و آن هم آخرین محدودیت در پائین جدول می‌باشد که دارای یک قانون غیر معمول است که از امکان بوجود آمدن حرکت (0 و 1) و (0 و 1) جلوگیری می‌کند.

محدودیت در مسیر سراسری

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

به دلیل وجود محدودیتهای پیوستگی محلی، قسمتهایی از منحنی (ix,iy)، از ناحیه‌ای که مسیر بهینه در آن قرار می‌گیرد خارج خواهند و جز این ناحیه نمی‌شوند. نقاط مجازی که تابع انحراف دهندهٔ بهینه مسیر در آنجا قرار می‌گیرد می‌توان به صورت زیر بیان کرد.

که در این معادلات L نشان دهندهٔ اندیکس مسیرهای مجاز می‌باشد (PL) و همچنین (TL) تعداد حرکات کلی در مسیر PL می‌باشد. برای مثال در نوع ، میزان ، برای ترتیب حرکات ، ، می‌باشد. این پارامتر ، به ترتیب میزان انبساط را در تابع انحراف دهنده مشخص می‌کنند.

به طور معمول ، جدول میزان Qmax و Qmin را برای هر نوع از محدودیتها نشان می‌دهد.

با استفاده از مقادیر ماکزیمم و می‌نیمم انبساط مسیر ما می‌توانیم مقدار محدودیت سراسری را بطور زیر تعریف کنیم:

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

معادلات(4.145 (محدودهٔ نقاطی از صفحه (ix,iy) را نشان می‌دهند که می‌توانند از نقطه (1 و 1) از طریق مسیر مجاز به آنها دسترسی پیدا کرد. معادلهٔ محدودهٔ نقاطی از صفحه را نشان می‌دهد که دارای مسیر مجاز به نقاط انتهایی (Tx,Ty) دارد.

شکل(4.4 (تأثیر محدودیتهای مسیر سراسری را بوسیلهٔ معادلات (4.146, 4.145)

برای حالت نشان می‌دهد. محدودهٔ مسیرهای مجاز در شکل توسط خطهایی با شیب نشان داده شده. همان طور که در معادله (4.146) و شکل 4.41 نمایش داده شده مدت زمان نقش مهمی‌در مشخص کردن نقاط مجاز دارند.

توجه کنید اگر ، نامساوی دوم در معادله (4.146) تبدیل می‌شود به که وقتی با نامساوی اول معادلهٔ (4.145) ترکیب شود، به رابطهٔ زیر خواهیم رسید.

این محدودیت مسیر سراسری یک خط راست است که نقطه (1 و 1) را به (Tx,Ty) وصل می‌کند، و تنها مسیر مجازی است که در شکل 4.42 نشان داده شده است، در این حالت بدلیل تفاوت بسیار زیاد در

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

طول زمانی دو الگو حتی با حداکثر انبساط (2) ، تنها یک مسیر ممکن خواهد شد.

یک شرایط مشابه زمانی اتفاق خواهد افتا که . در این مدت نیز فقط نرمالیزاسیون زمانی خطی امکان پذیر می‌باشد. بنابراین در واقعی که و یا زمانیکه ، عملیات همترازسازی زمانی متوقف خواهد شد، و هیچ گونه عملیات انحراف دهندگی زمانی اتفاق نخواهد افتا. این نکته بسیار مهمی‌است که در هنگام پیاده سازی باید مورد توجه قرار گیرد.

همچنین یک محدودیت مسیر سراسری نیز توسط Sakoe & chiba معرفی شد که به قرار ذیل است.

که در آن مقدار مجاز ماکزیمم انحراف در زمان بین دو الگو در هر فریم می‌باشد. این محدودیت مسیر سراسری اضافه شده باعث حذف کلیه مسیرهایی می‌شود باعث فشردگی یا انبساط زیاد در واحد زمان شده و همان طور که در شکل 4.41 نشان داده شده با بریدن گوشه‌های منحنی بطور مؤثری محدودهٔ مجاز را کاهش می‌دهد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

محدود کننده‌هایی شبیه به معادله (4.14a) بنام محدودیت‌های Rang-limiting شناخته می‌شوند، بدلیل اینکه آنها میزان قدر مطلق تفاوت در محور زمان را در هنگام انحراف دادن محدود می‌کنند.

وزن دهی شیب منحنی

وزن دهی شیب در طول مسیر نیز تا امروز یک بعد دیگر مطالعاتی و تحقیقاتی در جهت بدست آوردن، مسیر بهینه برای انحراف دادن می‌باشد. همان طور که درمعادله بیان شد،تابع وزن دهنده ، نقش هر فریم زمان کوتاه را در میزان اعوجاج ، نشان میدهد. باید دید کلی تر (یعنی تمام کلمه) تابع وزن دهی می‌تواند طوری طراحی شود که تأثیر قابل قبولی در میزان بازشناسی داشته باشد. اما در اندازهٔ کوچکتر و درحالت جزئی تر ، می‌توان مقدار وزن‌ها را بر طبق محدودیتهای مسیرهای توضیح داده شده معین نمود. این توابع محلی وزن دهنده بنام توابع SIOP weighting شناخته خواهند شد. دلیل این امر آن است که آنها معمولاً به شیب مسیرها سر و کار دارند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

با توجه به محدودیت‌های پیوستگی محلی،توابع وزن‌دهنده بسیاری قابل تعریف هستند. روابط زیر 4 نوع تابع وزن‌دهنده هستند که توسط Chibat sakoe [20] توصیف شده‌اند.

در روابط بالا جهت مقدار دهی اولیه فرض شده که مقدار می‌باشد.

برای نشان دادن تأثیر تابع وزن دهنده، شکل 4.43 تأثیر بکار بردن 4 نوع تابع وزن دهنده را روی نوع از محدودیت‌های پیوستگی محلی را نشان می‌دهد. عددی که به هر مسیر نسبت داده شده، میزان وزن آن مسیر می‌باشد. بدلیل اینکه اعوجاج بالا بیانگر تطابق کمتر می‌باشد، از مقدار وزن دهنده بیشتری برای مسیرهایی که زیاد دلخواه نمی‌باشد، استفاده شده است.

همان طور که در شکل 4.43 نشان داده شده است، برای مثال برای نوع (b) بیشتر سعی می‌کند به صورت بایاس باقی بماند تا بصورت قطری حرکت کند.

توابع وزن دهندهٔ بالا وقتی با سایر فریم‌های جدول 4.5 بکار برده می‌شوند. ممکن است در بعضی از مواقع منجر به وجود آمدن

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

وضعیت در طول مسیر بشود. بعنوان مثال، با ترکیب و بکار بردن تابع وزن دهنده آن با نوع دوم محدودیت پیوستگی محلی باعث ایجاد یک موقعیت بدون وزن در حرکت و خواهد شد.

این حالت‌ها در شکل 4.44 نشان داده شده است. حرکت افقی در و حرکت عمودی در هر کدام دارای مقدار وزن صفر خواهند بود. این بدان معنی است که فقط اعوجاج بدست آمده از حرکت قطری در معادلات 4.125 و 4.124 مورد نظر قرار خواهد گرفت. استفاده از این نوع توابع وزن دهی موجب بوجود آمدن عدم پیوستگی در نرمالیزاسیون زمانی خواهد شد.

یک راه برای کم کردن این مشکل توزیع دوباره ضرایب یا روان تر کردن وزن‌ها موقعی که هر مقدار به طور ناگهانی در یکی مسیرها تغییر می‌کند.

همان طور که در شکل 4.44 نشان داده شده برای نوع دوم محدودیت پیوستگی محلی، توضیح دوباره وزن شیبها موجب تقسیم شدن مساوی وزن‌ها در طول مسیر خواهد شد در حالیکه در حالت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اول مشاهده می‌کنیم وزن یک مسیر، به طور ناگهانی تغییر خواهد کرد.

مجموع اعوجاج نهایی نیازمند به نرمالیزاسیون کلی می‌باشد (همان طور که در معادله 4.124) نشان داده شده. هدف از نرمالیزاسیون سراسری پیدا کردن مقدار متوسط اعوجاج در طول مسیر است که این مقدار مستقل از طول دو الگویی است که در حال مقایسه می‌باشند.

مقدار فاکتور نرمالیزاسیون برای یک جمله وزن دهی شده به طور معمول برابر است با مجموع مولفه‌های وزن دهنده هر جمله

برای تابع‌های وزن دهنده نوع (c) و نوع (d) مقدار و فاکتور نرمالیزاسیون به صورت زیر محاسبه خواهد شد.

که این مقادیر مستقل از تابع انحراف دهنده و نوع محدودیت می‌باشند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای تابع وزن دهنده شیب از نوع (a) و (b) با این وجود مقدار فاکتور نرمالیزاسیون که توسط (4.154) توصیف شده تابعی از طول مسیر می‌باشد.

همان طور که پیدا کردن فاکتور نرمالیزاسیون برای یک تابع انحراف ساده می‌باشد، در صورتیکه مسأله منیمم‌سازی اعوجاج در رابطهٔ (4.125) از طریق الگوریتم DP حل شود، بسیار مشکل خواهد شد. برای حل این مسأله از یک مقدار فاکتور نرمالیزاسیون اختیاری اما منطقی استفاده می‌کنیم که مستقل از نوع تابع انحراف دهنده باشد و الگوریتم DP بتواند به کمک آن مسیر بهینه را پیدا کند.

به طور معمول برای تابع‌های وزن دهنده نوع و مقدار وزن دهنده شیب به صورت اختیاری

قرار داده می‌شوند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

انحراف زمانی پویا (dynamic - lime warping)

در این زمان می‌توان نشان چگونه با از الگوریتم DP در حل مسأله منیمم کردن اعوجاج برای پیدا کردن میزان عدم تشابه دو الگوی صحبت در حالیکه نرمالیزاسیون زمانی و هم ترازسازی انجام می‌دهیم استفاده کرد. با وجودیکه دربارهٔ محدودیت‌های پیوستگی محلی و وزن دهی شیب که در بخش قبلی دربارهٔ آنها سخن گفتیم باید با الگوریتم اصلی منطبق باشند. اصول اولیه DP و بهینه سازی بخصوص معادله (4.128)

بطور مستقیم دارای کاربرد در این الگوریتم می‌باشد.

بدلیل محدودیت انتها ما رابطهٔ 4.125 را با کمک Tx و Ty می‌نویسیم.

X و Y به نقاط TX و TY ختم می‌شوند، بطور مشابه مقدار جزئی مجموع اعوجاج در طول مسیر که نقطه (1 و 1) را به (ix,iy) وصل می‌کند.

که در آن

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بنابراین الگوریتم بازگشتی DP با وجود محدودیت می‌شود.

که در آن مقدار مجموع اعوجاج وزن دهی شده (فاصلهٔ محلی) بین دو نقطه و نقطه می‌باشد

که در آن تعداد حرکت‌های یک مسیر جهت حرکت در نقطه به می‌باشد.

باعث فاکتور نرمالیزاسیون را به دلیل مستقل بودن از نوع تابع انحراف دهنده از الگوریتم حذف نموده ایم. در این مرحله می‌توان مراحل الگوریتم برای پیدا کردن بهترین مسیر بین نقطه (1 و 1) و نقطه به طور زیر خلاصه نمود.

1) مقدار دهی

2) بازگشتی

ایده الگوریتم آن است که مرحلهٔ بازگشتی برای تمام مسیرهایی که از نقطه به نقطه می‌رسد، محاسبه خواهد شد. البته در این محاسبهٔ محدودیتهای مربوط به پیوستگی محلی نیز در نظر گرفته خواهد شد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تنها مقدارهایی از (ix,iy) که می‌توان از نقطه (1 و 1) به آنها رسید و از آنجا به نقطه (Tx,Ty) رفت در این محاسبه منظور خواهند شد. شکل 4.40 یک شبکه نقطه‌ای را نشان می‌دهد که نقاط مشخص شده در مرحلهٔ دوم و بازگشتی محاسبه شده‌اند. در این شکل از 2-to-1 برای گسترش و فشردگی استفاده شده و شیب منحنی‌ها می‌باشد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فصل چهارم

استفاده از مدلهای مارکف مخفی

در تشخیص گفتار

روشهای تشخیص گفتار به سه دسته کلی تقسیم می گردند

1- روشهای مبتنی بر مقایسه با الگوهای ذخیره شده

2- روشهای تشخیص گفتار با استفاده از شبکه های عصبی

3- روشهای آماری تشخیص گفتار

از میان روشهای مذکور، روش آماری تشخیص گفتار با استفاده از مدل مارکف مخفی مد نظر ما می باشد.

از این به بعد تنها فرایندهایی مد نظر خواهند بود که در آنها ظرف راست رابطه (1.3) مستقل از زمان باشند. درنتیجه به تعریف احتمالات انتقال حالت به شکل زیر خواهیم رسید.

ضرایب انتقال حالت باید در قیود زیر صدق کنند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فرایند اتفاقی بالا را می توان یک مدل مارکف مشاهده پذیر نامید زیرا خروجی فرایند، در هر لحظه از زمان، مجموعه ای از حالات می باشد که هر حال متناظر با یک پدیده فیزیکی (مشاهده پذیر) می باشد. جهت درک مطلب، یک مدل مارکف سه حالته ساده از هوا را درنظر می گیریم. فرض می کنیم که هوای مشاهده شده در یک روز (مثلاً هنگام ظهر)، یکی از وضعیتهای زیر باشد

حالت 1: بارانی یا (برفی)

حالت 2: ابری

حالت 3: آفتابی

حال ادعا می کنیم که وضعیت هوا در روز t، توسط یکی از سه حالت بالا مشخص می شود، و ماتریس احتمال انتقالات حالت A به صورت زیر می باشد

فرض می کنیم که وضعیت هوا در روز اول (t=1) آفتابی است. سوال ما اینست که (مطابق با مدل) احتمال اینکه وضعیت هوا در 7 روز بعد بترتیب «آفتابی – آفتابی – بارانی – بارانی – آفتابی – ابری –

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

آفتابی» باشد چیست؟ بعبارت دیگر ما دنباله مشاهدات O را بصورت متناظر با $t=1,2,\ldots 8$, تعریف کرده ایم و حال می خواهیم احتمال مشاهده O را با مدل داده شده، بدست آوریم. مقدار احتمال را می توان بصورت زیر محاسبه نمود

د ر بالا ما از عبارت

جهت نمایش احتمالات حالت شروع استفاده نموده ایم. سؤال جال دیگری که می توان مطرح نمود (و با استفاده از مدل جواب گرفت) این است که اگر فرض کنیم مدل در یک حالت مشخص باشد، احتمال اینکه در d روز در همان حالت بماند چیست؟ این احتمال را میتوان با محاسبه دنباله مشاهده زیر بدست آورد. یعنی

که با مدل داده شده،خواهیم داشت

2-3- مدل مارکف مخفی (HMM) [21]

پیش تر، مدل مارکفی را بررسی نمودیم که در آن هر حالت مطابق با یک پدیده (فیزیکی) قابل مشاهده بود. این مدل بسیار محدود می باشد

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

و کاربرد وسیعی ندارد. در اینجا ما مفهوم مارکف را بگونه ای گسترش خواهیم داد که در آن، مشاهده، خود تابعی از حالت باشد، بعبارت دیگر، در مدل حاصله حالات مدل مشاهده نمی شود بلکه تنها یک متغیر تصادفی وابسته به مشاهده می باشند. جهت درک بهتر مدل مربوط به آزمایشات کاسه و گلوله را در زیر می آوریم.

مدل کاسه و گلوله

سیستم کاسه و گلوله در شکل زیر درنظر می گیریم

فرض کنیم که N کاسه شیشه ای در یک اتاق وجود دارد. و درون هر کاسه تعداد زیادی گلوله رنگی با تعداد M رنگ مجزا برای گلوله ها وجود داشته باشد. فرآیند فیزیکی با یم فرآیند تصادفی، یک کاسه را انتخاب می کند. و از آن کاسه گلوله ای بطور تصادفی برمی دارد. کاسه های بعدی نیز طبق فرآیند انتخاب تصادفی مربوط به کاسه فعلی انتخاب می شوند، و فرآیند انتخاب گلوله نیز برای هر کدام در آن کاسه ها تکرار می شود. در آخر، این فرآیند منجر به تولید یک دنباله محدود مشاهده از رنگها خواهد شد. واضح است که یک مدل

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

HMM ساده و متناظر با فرآیند کاسه و گلوله مدلیست که در آن هر حالت متناظر با یک کاسه مشخص، و هر مشاهده (رنگ گلوله) تابع احتمالی از آن حالت باشد.

1-2-3- اجزای یک HMM

مثال بالا به ما یک ایده خوبی در مورد چیستی HMM و چگونگی کاربرد آن داد. حال عناصر یک HMM، و چگونگی تولید دنباله مشاهدات توسط مدل را بررسی می کنیم.

هر HMM توسط صفات ذیل مشخص می گردد

1) N، تعداد حالات مدل. اگرچه حالات مدل مخفی هستند، ولی اغلب در بیشتر کاربردهای عملی، شی یا پدیده فیزیکی مهمی را به حالت یا مجموعه ای از حالتها منسوب می کنند. در مدل کاسه و گلوله، حالتها متناظر با کاسه ها بودند. معمولاً حالتها بگونه ای به یکدیگر متصلند که از هر حالت بتوان به حالت دیگری رفت. کل حالات منحصر بفرد را

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بصورت عناصر یک مجموعه S={S1,S2,S3,...,SN} و حالت در زمان t را بصورت qt نمایش می دهیم.

2) M، تعداد نمادهای مشاهده منحصر بفرد برای هر حالت. نمادهای مشاهدات متناظر با خروجی فیزیکی سیستم مدل می شوند. در آزمایش مربوط به کاسه و گلوله، نمادها، رنگ گلوله های انتخابی از کاسه ها بودند. نمادهای منحصر بفرد را بصورت عناصر یک مجموعه V={V1,V2,...,VM} نمایش می دهیم.

3) توزیع احتمال انتقال حالت را بصورت ماتریس A={aij} نمایش می دهیم که در آن می باشد. در حالت خاص وقتی که بتوان از هر حالت در یک قدم به حالت دیگر رسید، باید به ازای تمام I,j ها باشد. در سایر حالات به ازای بعضی زوجهای (I,j) می تواند صفر باشد.

4) توزیع احتمال مشاهده نماد در حالت j را بصورت ماتریس B={bj(k)+} نمایش می دهیم که در آن

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

5) توزیع حالت شروع را بصورت مجموعه نمـــایش می دهیم که در آ"

با داشتن مقادیر مناسبN، M، A، B و می تـوان از HMM بعنـوان یک مولد جهت بدست آوردن دنباله مشاهده

(که در آن هر مشاهده یـک نمـاد از V اسـت، و T تعـداد مشـاهدات درون دنباله می باشـد) اسـتفاده نمـود. الگـوریتم زیـر را درنظـر مـی گیریم

1) یک حالت شروع مانند متناظر با توزیع حالت شـروع انتخـاب کنید.

2) t را برابر 1 قرار دهید.

3) متناظر با توزیع احتمال مشاهده نماد در حالت I یعنی را برابر با مقداردهی کنید.

4) متناظر با توزیع احتمال انتقال برای حالت یعنی بـه حالـت جدیـد بروید.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

5) t را برابر t+1قرار داده و در صورتیکه t<T می باشد به مرحله 3 بروید، در غیر اینصورت الگوریتم پایان یافته است.

از الگوریتم بالا می توان بعنوان یک مولد مشاهدات و همچنین بعنوان یک مدل جهت درک چگونگی تولید یک دنباله مشاهده، توسط HMM مناسب استفاده نمود. همانطور که دیدیم جهت تعیین کامل یک HMM، لازمست که دو پارامتر مدل (N,M) نمادهای مشاهده و مقادیر توزیع احتمالی B,A, تعیین شوند. جهت سادگی، از نمایش اختصاری جهت نمایش کامل مجموعه پارامتر مدل استفاده می شود.

2-2-3- سه مسئله اساسی در HMM

برای اینکه یک HMM، در کاربردهای دنیای واقعی مفید واقع شود باید بتوان سه مسئله مهم را حل نمود. این مسائل بترتیب عبارتند از

مسئله 1: دنباله مشاهدات و مدل مفروضات مسئله می باشندؤ مطلوبست احتمال تولید مشاهدات توسط مدل، یعنی

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مسئله 2: دنباله مشاهدات و مدل مفروضات مسئله می باشند، مطلوبست دنباله حالتهایی که بصورت بهینه مشاهدات را تولید کنند؟

مسئله 3: باز هم مفروضات مسئله دنباله مشاهدات و مدل می باشد، مطلوبست نحوه تطبیق و تغییر پارامترهای مدل، بگونه ای که ماکزیمم گردد؟

مسئله 1 در حقیقت مسئله ارزیابی است، یعنی یک مدل و دنباله ای از مشاهدات داریم، چگونه باید احتمال تولید دنباله مشاهده توسط مدل را محاسبه نماییم. مسئله 1 از دیدگاه امتحان مدل جهت درجه تطابق مدل با دنباله مشاهده نیز مهم است، زیرا از این طریق می توان از بین چندین مدل، بهترین مدل که بیشترین تطابق با مشاهدات را دارد بدست آورد.

مسئله 2 در حقیقت مسئله آشکارسازی می باشد، زیرا سعی در یافتن قسمت مخفی مدل را دارد یعنی یافتن حالت صحیح. واضح است که تمامی مدلهای غیرمولد دنباله مشاهده نمی توانند و نباید دنباله حالت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

صحیح یافته شده را در خود داشته باشند. بنابراین باید با توجه به کاربرد، بهترین ضابطه ممکن را جهت حل این مسئله برگزینند.

مسئله 3، مسئله یادگیری می باشد، یعنی سعی بر این است که با بهینه کردن پارامترهای مدل بتوان توصیف دقیقتری از دنباله مشاهده ارائه داد. دنباله مشاهداتی که جهت تطبیق پارامترهای مدل بکار می روند دنباله های آموزشی نام دارند زیرا جهت آموزش HMM استفاده می شوند.

جهت درک بهتر، یک سیستم تشخیص صحبت با کلمات منفرد را درنظر می گیریم. قصد داریم که برای هر کلمه از w کلمه موجود در فرهنگ لغات، یک مدل HMM مجزای N حالته بسازیم. سیگنال صحبت مربوط به هر کلمه را بصورت دنباله زمانی از بردارهای طیف کد شده نمایش می دهیم. فرض می کنیم که عمل کد توسط یک کتاب کد طیف با M بردار طیف منحصر بفرد صورت گیرد، لذا هر مشاهده شاخصی از نزدیکترین بردار طیف (در محدوده طیف) به سیگنال صحبت مبدا خواهد بود. بنابراین، برای هر کلمه از فرهنگ

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

لغات، یک دنباله آموزشی داریم که شامل یک تعداد تکرار از دنباله های شاخصهای کتاب کد مربوط به آن کلمه می باشد (که توسط یک یا چند گوینده بیان شده است). اولین عمل ساختن مدلهای منحصر بفرد کلمه می باشد. این عمل با استفاده از حل مسئله 3، یعنی محاسبه بهینه پارامترهای مدل برای هر مدل کلمه امکان پذیر است. جهت فهم بیشتر معنای فیزیکی حالتهای مدل، با استفاده از حل مسئله 2 دنباله های آموزشی کلمه را به حالتها می شکنیم و سپس صفات هر بردار طیف که منجر به وقوع یک مشاهده در یک حالت شده است را بررسی می کنیم. هدف در اینجا بهبود بخشیدن مدل است (بعنوان مثال، افزایش تعداد حالات،استفاده از کتابهای کد با اندازه های متفاوت و ...) بگونه ای که بتوان قابلیت مول کردن دنباله های کلمه ادا شده را بالا برد. بالاخره، وقتی که مجموعه wتایی HMM طراحی و بهینه سازی شد و مورد مطالعه قرار گرفت، تشخیص یک کلمه ناشناخته با استفاده از حل مسئله 1 بدین صورت انجام می گیرد که امتیاز هر مدل کلمه، برای دنباله مشاهده آزمایش می شود، و سپس

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

کلمه ای که مدل آن بالاترین امتیاز (بیشترین شباهت) را دارا است انتخاب می گردد.

در فصل بعد به حل سه مسئله اساسی HMM خواهیم پرداخت و خواهیم دید که این سه مسئله اساسی در چهارچوب احتمالات کاملا به یکدیگر وابسته می باشند.

1-2-2-3- حل مسئله 1

مطلوب ما در اینجا محاسبه احتمال دنباله مشاهده بشرط مدل داده شده می باشد، یعنی، . ساده ترین راهی که بنظر می رسد درنظر گرفتن تمام حالات بطول T (تعداد مشاهدات) می باشد. ابتدا یک مسیر معین مانند Q را در نظر می گیریم.

که در آن ، حاللت شروع می باشدذ. احتمال مشاهده دنباله O بشرط دنباله حالت Q بصورت زیر محاسبه می گردد.

توجه داشته باشیم که در محاسبه عبارت (3.14.1) ما مشاهدات را مستقل از یکدیگر فرض کرده ایم. لذا خواهیم داشت

احتمال وقوع دنباله حالت Q بشرط مدل بصورت زیر خواهد بود

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

احتمال وقوع O و Q هم زمان، ضرب دو عبارت بالا خواهد بود

احتمال دنباله مشاهده O بشرط مدل عبارآزمون از جمع بالا به ازای تمام Qهای ممکن، یعنی

با کمی دقت درخواهیم یافت که محاسبات (18. 3) از نوع محاسبات با درجه می باشد زیرا به ازای هر حالت جهت انتخاب وجود دارد (بعلارت دیگر تعداد دنباله حالت ممکن N می باشد) و جهت هر دنباله حالت ممکن 2T محاسبه لازم است. (بطور دقیق، ضرب و جمع لازم است). پس با توجه به حجم محاسبات زیاد، عملا پیاده سازی روش مستقیم، غیرممکن است حتی به ازای مقدار کوچک N و T، بعنوان مثال برای N=5 (تعداد حالات) و T=100 (تعداد مشاهدات)، تعداد محاسبات لازم می باشد. واضح است که باید یک روند بهینه جهت حل مسئله 1 یافت. خوشبختانه این الگوریتم وجود دارد و نام آن الگوریتم پیشرو- پسرو می باشد.

1-1-2-2-3- الگوریتم پیشرو – پسرو

متغیر پیشروی را بصورت زیر تعریف می کنیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

که در واقع، احتمال قسمتی از دنباله مشاهده یعنی (تا زمان t) و بودن در حالت I در زمان t بشرط مدل مفروض می باشد. می توانیم را با روش استقرایی، بصورت زیر حل نمود.

1)مقدار دهی اولیه

3)استقرا

3)پایان

مرحله 1 و 3 کاملاً واضح می باشندذ و برای اینکه درستی مرحله 2 را نیز تحقیق کنیم، با توجه به شکل 4-3- می توان روابط زیر را نوشت

شکل3-3. تعداد حالات موجود در زمانt جهت انتقال به حالت در زمان t+1

شکل 3-3 نشان می دهد که از N حالت ممکن در زمان t می توان به حالت در زمان t+1 رسید. از طرفی ، احتمال وقوع مشاهدات و بودن در وضعیت در زمان t می باشد پس حاصلضرب احتمال وقوع مشاهدات و رسیدن به حالت در زمان t+1 از طریق در

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

زمان t می باشد. اگر حاصلضرب بالا را به ازای تمامی N حالت ممکن محاسبه نموده و با هم جمع کنیم، نتیجه احتمال بودن در حالت به ازای دیدن مشاهدات می باشد. که اگر این مقدار را با احتمال وقوع مشاهده در حالت یعنی ضرب کنیم، بدست خواهد آمد. عبارت (3.21) به ازای تمام حالات j_t در زمان مفروض t محاسبه می شوند که این محاسبه باید به ازای تکرار گردد. در نهایت، در مرحله 3، از جمع متغیرهای پایانی پیشرو بدست می آید که تعریف بصورت زیر خواهد بود

مقدار محاسبات لازم جهت بدست آوردن که از درجه خواهد بود که در مقایسه با خیلی کمتر می باشد (بطور دقیق، $N(N+1)(T-1)+N$ ضرب و $N(N-1)(T-1)$ جمع لازم است) پس برای $N=5$ و $T=100$ حدود 3000 محاسبه در روش پیشرو لازم است که در مقایسه با روش مستقیم که محاسبه لازم داشت توان مقدار محاسبات حدود 69 بار کوچکتر می گردد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

شکل 4-3. پیاده سازی محاسبات بصورت عبارتی از مشاهدات t، و حالات I از توروراره.

از طریق ساختار تو.رواره ای شکل 4-3 نیز می توان عمل محاسبه احتمال متغیر پیشرو را انجام داد. نکته اصلی در این است که در اینجا فقط N حالت (گره های موجود در شبکه در هر قطعه زمانی) موجود است، ولی می توان تمام دنباله حالتها را با ترکیب این N گره موجود ساخت و بزرگی طول دنباله مشاهده نیز مهم نخواهد بود. در زمان t=I (اولین قطعه زمانی)، باید مقدار محسبه گردد. در زمانهای t=2,3,...,T تنها باید مقادیر محاسبه گردد. جهت عملیات محاسبه، فقط به N مقدار قبلی یعنی احتیاج می باشد. زیرا در هر گره از شبکه در یک قطعه زمانی به N گره موجود در قطعه زمانی قبل متصل می باشد. بطریق مشابه، می توان متغیر پسرو ، که در حل مسئله 3 بکار می رود را تعریف نمود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بعبارت دیگر، احتمال قسمتی از دنباله مشاهده در زمان t-1 تا به انتها، بشرط بودن در حالت i در زمان t و مدل . دوباره از روش استقرا جهت حل کمک می گیریم.

1) مقداردهی اولیه

3) استقرار

بررسی درستی روابط را دنبال نخواهیم کرد ولی در اینجا نیز جهت محاسبه به محاسبات از درجه نیاز می باشد که می تواند از روی ساختار شبکه ای مانند شکل 4-3 بدست آورد.

2-2-2-3- حل مسئله 2

برخلاف مسئله 1، که تنها یک راه حل داشت، در اینجا چند روش جهت حل مسئله 2 وجود دارد. لذا باید حل بهینه را جهت یافتن دنباله حالت متناظر با دنباله مشاهده مفروض پیدا کرد مشکل در تعریف دنباله حالت بهینه می باشد، یعنی چندین ضابطه بهینگی می توان تعریف نمود. بعنوان مثال، یک ضابطه بهینگی می تواند انتخاب حالتهای ای باشد که بصورت منحصر بفرد بیشترین شباهت را

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

دارند. این ضابطه بهینگی مقدار میانگین حالتهای منحصر بفرد صحیح را بیشینه می کند. جهت پیاده سازی این حل برای مسئله 2، متغیر زیر را تعریف می کنیم.

که عبارتست از، احتمال بودن در حالت در زمان t با شرط مفروض بودن دنباله مشاهده O و مدل . معادله (3 .28) را می توان بصورت عباراتی از متغیرهای پیشرو – پسرو بیان نمود.

با استفاده از می توانیم حالتن در زمان t که بصورت منحصر بفرد بیشترین شباهت را داراست بصورت زیر بدست آورد.

اگرچه معادله مقدار متوسط حالتهای صحیح را بیشینه می کند (با انتخاب حالت با بیشترین شباهت در هر زمان t)، اما ممکن است مشکلاتی در دنباله حالت نتیجه وجود داشته باشد. بعنوان مثال، هنگامی که در HMM بعضی از انتقال حالتها دارای احتمال صفر باشند (به ازای بعضی از اوزها، صفر باشد) ممکن است دنباله حالت بهینه، دنباله حالت معتبری نباشد. این مشکل از اینجا نشات میگیرد که معادله (3 .30) شبیه ترین حالت در هر لحظه از زمان را انتخاب

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

می کند و احتمال وقوع حالتهای دنباله در این معادله درنظر گرفته نشده است.

یک طریقه حل این مشکل می تواند تغییر ضابطه بهینگی باشد. بعنوان مثال، ضابطه بهینگی می تواند بیشینه کردن مقدار متوسط حالتهای دوتایی صحیح و یا حالتهای سه تایی باشد و ... . اگرچه این ضوابط می تواند برای بعضی از کاربردها جوابگو باشند، اما ضابطه ای که بهترین دنباله حالت (مسیر) را بیابد بیشترین کاربرد را دارد، بعبارت دیگر، ضابطه ای که را بیشینه کند و معادل است با بیشینه کردن . الگوریتمی جهت یافتن بهترین دنباله حالت وجود دارد،که براساس روشهای برنامه نویسی پویا بنا نهاده شده است، نام آن الگوریتم ویتربی[1] می باشد.

الگوریتم ویتربی: جهت یافتن بهترین دنباله برای دنباله مشاهده داده شده ، لازم است که کمیت زیر را تعریف نماییم.

[1] Viterbi Algorithm

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بعبارت دیگر، بهترین امتیاز (بیشترین احتمال) در طول یک مسیر منفرد می باشد که در زمان t، برای t مشاهده اول وجود دارد بگونه ای که حالت پایانی باشد. با استقرار خواهیم داشت.

برای اینکه بتوانیم دنباله حالت را بدرتی بازیابی کنیم لازم است که به ازای هر مقدار t، j، آرگومانی که عبارت (22. 3) را بیشینه می کند نگهداری کنیم. عمل نگداری را از طریق آرایه انجام می دهیم. بعبارت دیگر، محتوای ، حالتی می باشد در زمان t-1، که بهترین مسیر موجود به حالت j در زمان t را می دهد. الگوریتم کامل جهت یافتن بهرتین دنباله حالت در زیر آمده است.

1) مقداردهی اولیه

2) تکرار

3) پایان

4) برگشت به عقب از روی مسیر (دنباله حالت)

ملاحظه می شود که الگوریتم ویتربی از لحاظ پیاده سازی، عملیاتی شبیه عملیات محاسبه متغیر پیشرو دارد. تفاوت اصلی این دو گرفتن

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بیشینه برروی حالتهای قبلی در معادله (3 .34) بجای عملیات جمع در معادله (3 .22) می باشد. واضح می باشد که ساختار شبکه ای، کارایی بالایی در پیاده سازی الگوریتم ویتربی دارد.

3-2-2-3- حل مسئله 3

سومین و مشکلترین مسئله، تعیین روشی جهت بهینه نمودن پارامترهای HMM می باشد. روش تحلیلی و یا راه حل بهینه ای جهت تخمین پارامترهای مدل به ازای داده های آموزشی محدود وجود ندارد. با این حال، الگوریتمهای تکراری متعددی جهت یافتن بهینه محلی پارامترهای مدل وجود دارد. نکته مهم در این نوع الگوریتمها، ضابطه و یا معیاری است که بر اساس آن، پارامترهای مدل بهینه می گردند و یا بعبارت دیگر، مدل آموزش داده می شود. در این پروژه، از دو معیار بیشترین شباهت (ML) و بیشترین اطلاعات متقابل (MMI) استفاده شده است. در فصلهای چهارم و پنجم در مورد هرکدام از معیارها و نحوه آموزش توسط آنها، بتفصیل شرح خواهیم داد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

هدف از آموزش با ضابطه ML، تغییر پارامترهای مدل، در جهت بیشینه نمودن به ازای مشاهده مفروض O و مدل می باشد. اگرچه راه حل تحلیلی جهت آن وجود ندارد، ولی الگوریتمهایی نظیر الگوریتم بام-ولچ و یا الگوریتم های مبتنی بر گرادیان، جهت تخمین پارامترهای مدل بروش ML وجود دارند که هر دو از الگوریتمهای تکراری می باشند. در این فصل قصد داریم تا نحوه آموزش HMM به روش ML را با استفاده از الگوریتم بام-ولچ شرح دهیم.

هدف از آموزش با ضابطه ML، تغییر پارامترهای مدل،در جهت بیشینه نمودن به ازای مشاهده مفروض O و مدل می باشد. اگرچه راه تحلیلی جهت آن وجود ندارد، ولی الگوریتمهایی نظیر الگوریتم بام – ولچ و یا الگوریتم های مبتنی بر گرادیان، جهت تخمین پارامترهای مدل بروش ML وجود دارند که هر دو از الگوریتمهای تکراری می باشند. در این فصل قصد داریم تا نحوه آموزش HMM به روش ML را با استفاده از الگوریتم بام – ولچ شرح دهیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

1-4- الگوریتم بام – ولچ [21] و [11]

جهت توصیف الگوریتم تکراری بهبود پارامترهای HMM، ابتدا ، که احتمال بودن در حالت در زمان t و بودن در حالت در زمان t+1، بشرط مدل و دنباله مشاهده مفروض می باشد را تعریف می کنیم.

که شکل زیر نیز، همین مطلب را نشان می دهد

شکل 1-4. نمایش سلسله عملیات لازم جهت محاسبه احتمال بودن در حالت در زمان t و حالت در زمان t+1

با استفاده از متغیرهای پیشرو و پسرو خواهیم داشت:

با تعریفی که قبلاً از داشتیم (احتمال بودن در حالت در زمان t بشرط دنباله مشاهده و مدل) می توان رابطه زیر را بین و نوشت.

اگر را برروی شاخص زمان t جمع ببندیم، کمیتی بدست خواهیمن آورد که می توان آنرا تعداد متوسط زمانهای مشاهده حالت و یا معادل آن، تعداد متوسط انتقالات انجام گرفته از حالت (البته قطعه زمانی t=T را در جمع در نظر نگیریم) تفسیر نمود. بطور مشابه،

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

جمع بروی t (از t=1 تا t=T-1) را می توان تعداد متوسط انتقالات از حالت به حالت تفسیر نمود. پس داریم

با استفاده از فرمولهای بالا می توان روشی جهت بهبود پارامترهای HMM ارائه نمود. فرمولهای بهبود پارامترهای و A ،B در این روش بصورت زیر خواهد بود.

اگر مدل فعلی را بنامیم و با استفاده از محاسبه طرف راست معادلات

(4 .7) مدل بهبود یافته را بدست آوریم، رابطه زیر برای مدل فعلی و مدل بهبود یافته همیشه برقرار خواهد بود.

حال اگر به اساس الگوریتم بالا، بصورت تکراری را بدست آورده و جایگزین نماییم و ددوبره عملیات بهبود را تا رسیدن به یک نقطه حدی ادامه دهیم، احتمال مشاهده شدن O از روی مدل افزایش خواهد یافت. این الگوریتم بهبود را الگوریتم بیشترین شباهت نامیده اند. باید این نکته را ذکر کرد که الگوریتم پیشرو – پسرو فقط نقاط بیشینه محلی را بدست می دهند، و در بسیاری از مسائل، فضای

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بهینه سازی بسیار پیچیده و دارای نقاط بیشینه محلی فراوانی می باشند. فرمولهای بهبود (4. 7. 1) تا (4. 7. 2) را می توان با بیشینه کردن تابع کمکی بام

به ازای بدست آورد. بام و همکارانش ثابت نموده اند که بیشینه نمودن باعث افزایش شباهت خواهد شد، یعنی

در زیر رابطه (4. 10) را با استفاده از تابع کمکی بام ثابت می کنیم.

فرض کنید که به ازای ماکزیمم گردد

با استفاده از رابطه کمکی بام خواهیم داشت

2-4- مسائل مربوط به پیاده سازی روش ML

در این بخش قصد داریم تا به مسائلی مربوط به پیاده سازی عملی همچون مقیاس بندی، دنباله های مشاهده چندتایی، تخمین اولیه پارامترها و انتخاب اندازه و نوع مدل بپردازیم. برای بعضی از مسائل پیاده سازی می توان راه حلهایی تحلیلی ارائه داد ولی برای بعضی دیگر تنها می توان از تجربیات حاصل از کار با HMM در چندین سال اخیر بهره گرفت.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

الف – مقیاس بندی

برای اینکه بدانیم که چرا انجام مقیاس بندی در پیاده سازی اولیه آموزش HMMها ضروری است، تعریف و فرمول شماره (3. 19) را درنظر می گیریم. همانطور که می توان دید ، از تعداد زیادی جمع عبارت تشکیل شده، که هر عبارت به شکل

می باشد از طرفی a و b از یک کوچکترند (معمولاً بسیار کوچکتر از یک)، لذا می توان انتظار داشت که با بزرگ شدن t (بعنوان مثال 10 یا بیشتر)، هر عبارت از برسعت بطور نمایی به صفر نزدیک شود و به ازای مقادیر خیلی بزرگ t (بعنوان مثال 100 یا بیشتر) بازه دینامیک نحاسبات از بازه دقت هر ماشینی (حتی با دقت مضاعف) فراتر خواهد رفت. لذا تنها راه انجام محاسبات، استفاده از مقیاس بندی می باشد.

در الگوریتم مقیاس بندی پایه که مورد استفاده قرار می گیرد را در ضریب مقیاس بندی که مستقل از i می باشد (بعبارتی تنها به t وابسته است) ضرب می کنیم، با این هدف که تغییر یافته در بازه

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

دینامیک به ازای باقی بماند. از یک الگوریتم مقیاس بندی مشابه برای ها نیز استفاده می شود (زیرا که ها نیز بسرعت نمایی بسمت صفر میل می کنند). در پایان عمل محاسبات، ضرایب مقیاس بندی را از محاسبات حذف می نماییم. جهت درک بهتر الگوریتم مقیاس بندی، معادلات مربوط به تخمین ضرایب انتقال حالت را درنظر بگیرید. اگر معادله (7. 4) را بصورت عباراتی از متغیرهای پیشرو – پسرو بنویسیم خواهیم داشت

محاسبات مربوط به را درنظر می گیریم. به ازای هر t، ابتدا را مطابق با فرمول استقرا (21. 3) محاسبه نموده، و سپس آنرا در ضریب مقیاس بندی ضرب می نماییم.

بنابراین به ازای یک t معین، ابتدا را بصورت زیر محاسبه مب نماییم.

سپس مجموعه که مقادیر مقیاس داده شده می باشد را بصورت زیر بدست می آوریم.

با استفاده از استقرا، را می توان بصورت زیر نوشت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بنابراین می توانیم را بصورت زیر بنویسیم.

بعبارتی، هر توسط حاصلجمع برروی تمامی حالتها بطور موثری تغییر مقیاس داده می شود. سپس ما عبارات را بصورت بازگشت به عقب محاسبه می نماییم. تنها تفاوت در اینجا اینست که ما از فاکتورهای مقیاس بندی مشابه ای در هر زمان t برای ها استفاده خواهیم نمود به همان طریق که برای ها استفاده نمودیم. بنابراین ای که تغییر مقیاس دادن شده باشد بدین شکل خواهد بود.

بنابراین هر فاکتور مقیاس بندی مقدار را بصورت قابل ملاحظه ای به 1 نزدیک می کند، و از طرفی مقادیر و قابل مقایسه می باشند. و استفاده از فاکتورهای مشابه جهت همانطور که برای ها استفاده شده اند می تواند محاسبات را در محدوده قابل قبولی نگه دارد. لذا با توجه به متغیرهای مقیاس داده شده، معادله تخمین (14. 4) بصورت زیر درخواهد آمد

اما هر را می توان بصورت زیر نوشت

و هر را می توان بدین صورت نوشت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

لذا معادله (4 .21) را می توان بدین صورت نوشت

و بالاخره عبارت از صورت و مخرج معادله بالا حذف می شوند و ما به معادله اصلی تخمین خواهیم رسید. واضح است که الگوریتم مقیاس بندی بالا را جهت تخمین مقادیر و B نیز می توان بکار برد. و نیز کاملاً مشهود است کهع الگوریتم مقیاس بندی معادله (4 .16) را لازم نیست در هر لحظه از زمان بکار برد بلکه باید آنرا در مواقع مطلوب و ضروری (بعنوان مثال جهت جلوگیری از ته ریز شدن)[2] بکار برد. اگر نقیاس بندی در یک لحظه t از زمان انجام نگیرد بدین معناست که ضرایب دارای مقدار یک می باشند و تمامی شرایط بحث شده در بالا معتبر خواهند بود. تنها تغییر واقعی که در روند HMM بعلت مقیاس بندی رخ خواهد داد مربوط به روند محاسبه خواهد بود. زیرا ما نمی توانیم مقادیر تغییر مقیاس داده شده را جمع ببندیم. اما می توانیم از این خاصیت استفاده نماییم که

بنابراین خواهیم داشت

[2] Underflow

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

یعنی ما الگوریتم p را می توانیم محاسبه کنیم نه خود p را، زیرا دقت لازم جهت محاسبه p از محدوده دقت هر ماشینی خارج است.

توجه داشته باشیم که به هنگام استفاده از الگوریتم ویتربی جهت بدست اوردن دنباله حالت با بیشترین شباهت احتیاجی به انجام مقیاس بندی نخواهد بود اگر که الگوریتم را به روش زیر استفاده نماییم. تعریف زیر را درنظر می گیریم.

- مرحله مقداردهی اولیه
- مرحله بازگشتی
- مرحله پایانی

دوباره به جای p به مقدار رسیدیم ولی با محاسبات کمتر و بدون مشکل محاسباتی. (توجه داشته باشیم که عبارات از معادله بازشگتی (31. 4) را می توان قبلاً محاسبه نمود. بنابراین هیچ هزینه محاسباتی دربرنخواهد داشت و علاوه بر آن، عبارات نیز بهنگام استفاده از آنالیز نماد مشاهده محدود (بعنوان مثال، یک کتاب کد دنباله مشاهدات) می تواند قبلاً محاسبه گردد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

ب. دنباله های مشاهده چندتایی[3]

در مدل مارکف مخفی چپ به راست یا مدل بیکیس[4] پردازش حالتها از حالت 1 در t=1 تا حالت N در زمان t=T به صورت متوالی انجام می گیرد. در روش چپ به راست می توان یکسری محدودیتهایی روی ماتریس انتقال حالت، و احتمالات حالت اولیه اعمال نمود. اما، مشکل اصلی در مدلهای چپ به راست این است که نمی توان با استفاده از فقط یک دنباله حالت مدل را آموزش داد (بعلارت دیگر، پارامترهایی مدل را تخمین زد). غلت این امر اینست که طبیعت ناپایدار حالتها در مدل، تنها امکان تعداد مشاهدات کمی را برای هر حالت می دهند (تا اینکه یک انتقال به حالت بعدی صورت بگیرد). لذا، برای اینکه داده کافی جهت تخمینی مطمئن از همه پارامترهای مدا داشته باشیم، مجبور به استفاده از دنباله های مشاهده چندتایی می باشیم.

[3] Multi Observation Sequences

[4] Bakis

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تغییرات رویه تخمین بسیار ساده و بصورت زیر است. دنباله های مشاهده kتایی را بصورت زیر بیان می کنیم.

که kامین دنباله مشاهده می باشد. ما فرض می کنیم که هر دنباله مشاهده از سایر دنباله های مشاهده مستقل می باشد، و هدف ما تطبیق و یا تغییر پارامترهای مدل یعنی می باشد بگونه ای که باعث بیشینه شدن با تعریف زیر بشود.

از طرفی فرمولهای تخمین براساس تعداد تکرار رخدادهای متفاوت بود، در اینجا فرمولهای تخمین برای دنباله های مشاهده چندتایی بعلت جمع بستن تعداد تکرارهای منحصر به هر دنباله مشاهده چندتایی بعلت جمع بستن تعداد تکرارهای منحصر به هر دنباله با یکدیگر تغییر یم یابد. بنابراین فرمولهای تخمین تغییر یافته برای و بصورت زیر می باشند.

و تخمین زده نمی شود زیرا 1= و 0= به ازای .

مقیاس بندی متناسب عبارات (4 .35) و (4 .36) ساده می باشد زیرا هر دنباله مشاهده فاکتور مقیاس بندی خودش را دارا می باشد. ایده

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اصلی، از بین فکتور مقیاس بندی در هر عبارت قبل از جمع می باشد. این عمل با نوشتن معادلات تخمین بصورت عباراتی از متغیرهای مقیاس یافته ممکن است یعنی

با این راه، برای هر دنباله مشاهده ، فاکتورهای مقیاس بندی مشابه در هر عبارت جمع روی t ظاهر خواهند شد. همانگونه که برای عبارت ظاهر شده است، و بنابراین می توان آنار حذف نمود. لذا با استفاده از متغیرهای مقیاس یافته به بدون تغییر مقیاس دست خواهیم یافت. برای عبارت تیز به همین نتیجه مشابه خواهیم رسید.

پ. تخمین اولیه پارامترهای HMM

در تئوری، معادلات تخمین باید مقادیری از پارامترهای مدل را بدست دهند که متناظر با بیشینه محلی تابع شباهت باشد. سؤال اصلی در اینجا این است که تخمین اولیه پارامترهای HMM باید چگونه باشد تا بیشینه محلی، بیشینه عمومی تابع شباهت باشد. اساسا هیچ جواب ساده و مستقیمی جهت پاسخ به سوال بالا وجود ندارد. ولی در عوض، تجارب نشان داده اند که تخمینهای اولیه تصادفی (بصورت

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

تصادفی و با محدودیت غیرصفر بودن مقدار) و یا یکنواخت پارامترهای و A، اغلب جهت تخمین این پارامترها مناسب بوده اند. همچنین، برای پارامترهای B، تجربه نشان داده است که تخمین اولیه خوب در حالت نماد گسسته[5] مفید، و در حالت تنوزیع پیوسته ضروری می باشد. اینچنین تخمینهای اولیه ای می تواند از چندین راه حاصل شود،مانند خوشه بندی دستی دنباله (های) مشاهده به حالتها با متوسط گیری مشاهداتن داخل حالتها، خوشه بندی با بیشترین شباهت برای مشاهدات با متوسط گیری، و خوشه بندی به روش k-میانگین[6] و غیره.

5- انتخاب مدل

مسئله دیگر در پیاده سازی HMM انتخاب نوع مدل، انتخاب اندازه مدل (تعداد حالتها) و انتخاب نمادهای مشاهده (گسسته یا پیوسته، منفرد یا مختلط) می باشد. متاسفانه هیچ روش ساده و تئوری

[5] Discreate Symbol

[6] K-means

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

صحیحی جهت انجام این انتخابها وجود ندارد. این انتخابها باید با توجه به سیگنالی که قرار است مدل شود انجام گیرد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

فصل 5: بازشناسی و ارزیابی نحوه بیان کلمات مقطع قرآنی

در این فصل به بررسی راه اندازی سیستم بازشناسی گفتار مقطع کلمات قرآنی و همچنین چگونگی ارزیابی نحوه بیان این کلمات خواهیم پرداخت.

فلوچارت مربوط به سیستم بازشناسی گفتار کلمات مقطع بصورت زیر می باشد.

در مرحله اول باید سیگنال صحبت جهت عمل باشناسی آماده کنیم. این مرحله شامل استخراج خواص اکوستیکی صحبت می باشد. در اولین قدم سیگنال صحبت پس از پیش تاکید توسط یک فیلتر FIR درجه اول – فریم هایی با طول 25ms – همراه 10ms هم پوشانی تقسیم شده سپس برای هر فریم مقادیر انرژی و میزان عبور صفر مربوط به آن محاسبه شده و این مقادیر – الگوریتم تشخیص ابتدا و انتهای صحبت داده می شود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پس از تشخیص ابتدا و انتهای سیگنال صحبت، در سیگنال خاصی صحبت مقدار ضرائب کپتروم استخراج خواهد شد. جهت استخراج ضرائی کپستروم از روش FFT استفاده شده است.

پس از آن نوبت به ایتخراج پارامترهای گذری که حاوی اطلاعات مهمی از سیگنال هستند، به همراه ضرایب کپستروم مشتقات اول و دوم – به نحوی که در فصل های گذشته شد، استخراج خواهد شد.

برای بدست‌آوردن ماتریس ویژگی و مشخصه سیگنال، این ماتریس را بدین گونه تعریف می کنیم.

Feature Matrix fF.M=[ceps Δceps $\Delta\Delta$ceps]

در این حالت تعداد شرایب کپستروم می تواندمتغیر باشد. و بسته به تعداد ضرایب کپستروم اندازهٔ ماتریس فرق می کند. تعداد سطرهای ماتریس بیان گر تعداد فریم ها هستند.

پس از بدست آوردن ماتریس ضرایب، نوبت به انتخاب ابزار بازشناسی می رسد. در بخش اول به بررسی سیستم بازشناسی گفتار بوسیلهٔ الگوریتم DTW خواهیم پرداخت و در بخش بعدی سیستم بازشناسی گفتار HMM مورد بررسی قرار خواهد گرفت.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بازشناسی گفتار بوسیله الگوریتم DTW

□□□□ □□□□ □□□□□□ □ □□□□ □□□□□ □□□□ □□□□□ □□□□□ □□□□ □□ □□ □□□□ □□□□□ □□□□□□□□ □□ □□□□ □□□□□□□ □□□□□ □□□□□ □□□□□ □□□□ □□ □□

.□□□□

فلوچارت الگوریتم شناسایی DTW بصورت زیر می باشد.

در این مرحله تعداد ضرایب کپستروم و طول پنجره نقش مهمی در میزان دقت سیستم خواهد داشت. همچنین در مرحلهٔ اول برای هر کلمه یک کلمهٔ مبنا که صدای استاد می باشد انتخاب شده است.

جهت تست سیستم مورد نظر از 5 فایل صوتی که توسط افراد مختلف کلمات قرآنی را ادا کرده اند استفاده کردیم. با دادن این فایلها به سیستم و مقایسهٔ آنها با صدای استاد،دقت سیستم را اندازه گیری کردیم.

همان طور که گفتیم تعداد ضرایب کپستروم و طول پنجره و میزان همپوشانی در تعیین نرخ بازشناسی نقش مهمی دارند. بنابراین در مرحلهٔ دوم سعی در بالا بردن این ضریب تا حد امکان نمودیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

جهت پیدا کردن مقادیر بهینه شروع به تغییر دادن هرکدام از پارامترهای مذکور کردیم. این کار را تا زمانی ادامه می دادیم که مطمئن می شدیم سیستم به حالت اشباع خود رسیده و بیشتر از این دیگر توسط تغییر این پارامتر، بهبودی در روند بازشناسی حاصل نخواهد شد.

همان طور که مشاهده می کنیم در مرحلهٔ اول با تغییر تمامی پارامترها نتوانستیم بیشتر از برنامه در بازشناسی گفتار دسترسی پیدا کنیم.

پس از این مرحله سعی در بهبود بخشیدن سیستم با تغییر در ساختار اصلی روند الگوریتم بازشناسی گفتار نمودیم.

[illegible]

در فصل دوم نحوهٔ جداسازی سیگنال صحبت از سیگنال زمینه را دیدیم. همچنین در سیستم مقدماتی بازشناسی گفتار در قسمت اول سیگنال صحبت پنجره هایی با طول و هم پوشانی معین تقسیم خواهد شد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

سپس برای هر فریم و میزان عبور از صفر آن محاسبه شده و توسط الگوریتم EPD جدا خواهد شد.

در فصل دوم هنگام پیاده سازی الگوریتم فوق گفتیم که در آنجا هیچ گونه محدودیتی در مورد طول پنجره نبوده، بنابراین در آن قسمت که هدف صرفاً جداسازی سیگنال صحبت از روی زمینه بود، با انتخاب طول پنجره کوچکتر این امر میسر شد و الگوریتم به خوبی جواب داد.

اما در این مرحله در مورد طول پنجره محدودیت وجود دارد و دیگر نمی توان آن رابه اندازهٔ دلخواه گرفت.

به دلیل اینکه می خواهیم در اینجا فرکانس Pitch را محاسبه کنیم و تقریباً فرکانس Pitch دارای محدودهٔ 6.00HZ است پس نمی توانیم طول پنجره را کوچکتر از 25ms انتخاب نمائیم. با افزایش طول پنجره مشاهده شد، کیفیت سیگنال جداشده از زمینه پائین آمد، و قسمتهایی از آن بریده شده است.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

دلیل این امر این است که با زیاد شدن طول پنجره اولاً مقدار عبور از صفر زیاد می شود و بنابراین پیدا کردن یک میزان آستانه خوب جهت جدا کردن سیگنال صحبت از زمینه سخت می شود.

این مشکل بیشتر در مورد کلمات شروع می شود که با حروف سایشی مثل «ک، ف، ه» شروع می شوند. بدلیل اینکه ایم کلمات ماهیت نویزگونه ای دارند، دارای انرژی کمی یم باشند، پس الگوریتم انرژی قادر به شناسایی آنها در اول فریم نمی باشد.

از آنجائیکه طول پنجره زیا است ممکن است مقدار زیادی در آن سکوت قرار بگیرد. حدود 10ms اول سکوت باشد و بعد از آن سیگنال صحبت با یک حرف سایشی مثل ه، یا ف، یا ک شروع شود.

مشاهده شد در این صورت بدلیل بالا بودن میزان عبور از صفر فریم، الگوریتم 2CR فریم را به عنوان زمینه شناخته و آن را حذف می نماید. بدین ترتیب مشاهده شد سیگنال صحبت استخراج شده بریده می باشد و صدای اول آن واضح نمی باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مشکل فرق در مورد بعضی از کلمات که به حرف «م و ن» ختم می شوند نیز صادق بود، و مشاهده می شد، کلمات فرق در آخر کلمه دچار مشکل خواهند بود، صدای حرف آخر به خوبی شنیده نمی شد.

مشکل بعدی، اضافه شدن مقداری سکوت و زمینه در اول کلمه هنگام جداسازی بود. بازهم این مشکل بدیلیل بالا بودن طول پنجره اتفاق می افتد. یعنی اگر بازهم مثلاً 10ms اول فریم سکوت باشد و 10ms دو یک حرف صدادار باشد بدلیل بالا بودن انرژی فریم الگوریتم انرژی این فریم را به عنوان اول سیگنال شناخته و بنابراین مقداری سکوت به اول سیگنال اضافه خواهد شد.

علاوه بر مشکلات فوق نتایج نشان می داد که سیستم برای پنجره های با طول بالا بعتر عمل می کرد. ولی از طرفی برای برخی از کلنات دچار مشکل می شد و هیچ گاه درست آن ها را تشخیص نمی دادد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای حل این مشکل تصمیم گرفتیم قسمت فریم بندی سیگنال را در دو مرحله انجام بدهیم. مرحله اول سیگنال را با طوب پنجره های کوچک و بدون هم پوشانی (هم پوشانی نقش زیادی ندارد)

- فریم هایی با طول مساوی تقسیم می کنیم. و در مرحله دوم سیگنال صحبت خالص جدا شده در زمینه را با طول پنجره های بزرگ تر و یا بهینه فریم بندی می کنیم.
- این کار دارای دو مزیت می باشد: اول اینکه بدلیل کوچک بودن پنجره ها و بالا رفتن دقت و کوچک بودن مقدار 2CR ، دقت الگوریتم EPD فوق العاده بالا می رود، و سیگنال جداسازی شده در زمینه هیچ مشکلی از لحاظ بریده شدن کلمات ندارد و ضمناً بدلیل کوچک بودن طول پنجره ها میزان سکون اضافه شده به اول فریم تقریباً صفر می باشد و یا بسیار کم می باشد.
- مزیت دوم این کار: تسریع در هبنه سازی طول پنجره هاست چون که طول پنجره ها هم در دقت بازشناسی گفتار

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

موثر است و هم در نحوهٔ جداسازی سیگنال از زمینه، بنابراین بدلیل مربوط بودن این پدیده به همدیگر پیدا کردن طول پنجره بسیار سخت می باشد.

با این کار بدلیل جداسازی ایزولاسیون طول پنجره ها مربوط به هر پدیده در مراحل بهینه سازی خیلی راحت می توان هر دو پارامتر را به مقدار بهسنه رسانید.

در مرحلهٔ بعدی کار دوباره سعی کردیم این دو پارامتر را با روش سعی و خطا به مقدار بهینه برسانیم و این کار را تا حدی که سیستم به حالت اشباع خود برسد ادامه دادیم.

تطابق چندالگویی [7]

تعبیر خواص اکوستیکی از گوینده ای به گویندهٔ دیگر بسیار پیچیده می باشد. یکی از اشکالات عمده الگو، سیستمهای تطابق الگو در همین مساله مستقل از گوینده بودن می باشد. بنابراین ممکن است مقداری خطا در هنگام تطبیق دو کلمه مشابه که توسط گوینده های متفاوت گفته باشد، بوجود آید. در بعضی از مواقع همین مقدار خطا بسیار

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

سرنوشت ساز می باشد و ممکن است باعث خطای بازشناسی شود. مثلاً ممکن است در هنگام تطابق دو کلمه مشابه که یک توسط یک مرد گفته شده باشد و دیگری توسط یک بچه یا زن – دلیل تفاوت در ساختار صوتی این دو بوجود آید. البته در هنگام طراحی سیستم های مستقل از گوینده باید تا حد امکان اطلاعات مربوط به گوشنده را حذف نمود، ولی باز هم جنسیت و خود گوینده در نرخ بازشناسی تاثیر دارد.

یک مشکل دیگر که در هنگام بازشناسی گفتار کلمات گفتار قرآنی بوجود می آیدؤ مساله تفاوت در تلفظ در بین افراد مختلف است. به دلیل تلفظ عربی و فارسی کلمات قرآنی ممکن است در هنگام بازشناسی مشکلاتی وجود آید. بطور مثال در کلمه «بک» در هنگام تلفظ این کلمه می توان آن را به همان صورت فارسی بیان کرد. و یا آن را بصورت عربی و تبدیل « ر » به «ی» تلفظ نمود و آن را «بیک» گفت.

---

[7] - multi template matching

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در مورد کلمه «حسر» که می توان آن را بصورت عربی «حاسر» تلفظ نمود و یا می توان آن را به صورت فارسی و به همان صورت خسر بیان نمود. در تمامی این موارد کلمه بیان شده درست می باشد و سیستم باید قادر به شناسایی هر دو کلمه باشد. اما اگر ما به عنوان کلمهٔ مبنا فقط یکی از این دو کلمات را قرار دهیم، در تطابق کلمه مشابه با تلفظ متفاوت دچار مشکل خواهیم شد، و این مشکل البته فقط در هنگام سر و کار داشتن با لغات قرآنی پیش می آید.

یکی از دلایل پایین بودن نرخ بازشناسی گفتار در سیستم اولیه، همین مسائل فوق بود، یعنی اولاً مدل بلحاظ مستقل از گوینده بودن زیاد قوی نبود، ثانیاً به دلیل متفاوت بودن تلفظ بعضی از نمونه ها، احتمال می رفت بدلیل تفاوت آن ها با کلمهٔ مبنا دچار خطای بازشناسی شویم.

برای حل مشکل دو راه حل را امتحان نمودیم.

روش اول: استفاده در میانگین الگوها:

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

یک روش بسیار ساده برای تطابق چندالگویی استفاده از میانگین الگوها می باشد. اما مشکل زمانی آغاز می شود که اولاً طول کلمات مشابه ای توسط گوینده های مختلف بیان شده اند یکی نمی باشد، ثانیاً هیچ دلیلی برای مشابه بودن وطلاعات طیف فرکانس هر فریم یک کلمه با همان فریم سایر کلمات نمی باشد. پس بنابراین عمل جمع کردن میانگین بین هر فریم کار بسیار اشتباهی است.

جهت حل این مشکل از روش زیر استفاده کردیم فرض کنید دارای 6 کلمه مشابه می باشیم هرکدام توسط گوینده های مختلف بیان شده اند و می خواهیم یک کلمه مبنا برای آن کلمع انتخاب کنیم.

برای این کار یکی از این 6 الگوریتم را به عنوان مبنا درنظر می گیریم به صورت کاملاً اختیاری، سپس به مقایسه این کلمه با 5 کلمه دیگر از طریق الگوریتم DTW می پردازیم. یعنی کلمه مبنا را با هر کدام از کلمات تطابق داده و منحنی های انحراف دهنده آنها را برای هر یک از کلمه بدست می آوریم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

نام تابع انحراف دهنده کلمه مبنا را R(K) و نام تابع تنحراف دهنده تست را T(K) می نامیم.

این توابع انحراف دهنده دو مزیت دارند: اولاً طول دو کلمه مبنا و تست یکی شده است، ثانیاً ما می توانستیم تشخیص بدهیم که کدام فریم از کلمهٔ مبنا با کدام فریم از کلمهٔ تست از لحاظ خواص طیف فرکانس مشابه است.

با داشتن این اطلاعات در مورد همهٔ کلمات سرانجام به کمک منحنی های R(K) و T(K) خواهمی فهمید که کدام فریم از کلمهٔ مبنا با چه فرمهایی از سایر کلمات مشابه است. بدین ترتیب به هر دو مشکل فوق فائق آمدیم و حالا می توانیم از طریق میانگین گیری از فریم های مشابه، فرم مبنای جدید را بدست بیاوریم.

نکته ای که در اینجا باقی می ماند انتخاب خود کلمهٔ مبنا می باشد. می توان کلمه را انتخاب نمود که دارای بیشترین تنطابق با سیار کلمات باشد. این کار را با تکرار الگوریتم DTW برای هر کلمه مبنا با سیار

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

کلمات انجام داد. کلمه مبنایی بیشترین تطابق دارد که مجموع خطایش با سایر کلمات می نیمم باشد.

با پیاده سازی الگوریتم و اجرای آن توسط سیستم مشاهده شد نتایج خوبی بدست نیامد.

الگوریتم پیدا کردن کلمهٔ مناسب

روش k – نزدیک ترین همسایگی[8]

روش k- نزدیکترین همسایگی دو روش ساده و خوبی در تطابق چندالگویی است. این روش بخوبی نیازهای ما را در مورد قوی کردن در مقابل تغییرات خواص اگوستیکی از گوینده ای به گویندهٔ دیگر مشکل تفاوت تلفظ این کلمه تست و مبنا را حل می کند.

برای آشنایی با الگوریتم فرض کنید دوباره برای یک کلمه دارای 6 صدا هستیم که هرکدام توسط افراد مختلف گفته شده اند تمامی این 6 کلمه را به عنوان مبنا درنظر می گیریم و کلمهٔ تست را بوسیله الگوریتم DTW بصورت جداگانه با این 6 کلمه مبنا مقایسه خواهیم کرد. حال ما دارای 6 عدد هستیم که هرکدام متناظر هستند میزان

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

عدم تشابه آن کلمه تست با کلمهٔ مبنا می باشد. در مرحله بعدی این 6 عدد را به ترتیب از کوچک به بزرگ مرتب می کنیم. در نهایت برای پیدا کردن میزان عدم تشابه بین کلمهٔ تست و این شاخه[9] از میانگین k عنصر اولین مجموعه استفاده می کنیم.

الگوریتم بسادگی بیان می کند، میزان عدم تشابه بین کلمهٔ تست قرآن شاخه برابر است با میانگین مقدار عدم تشابه بین کلمهٔ تست و کلماتی که بهترین تطابق با آن کلمه را دارند می باشند. همان طور که می توان دید کلمه ای که بدلیل تفاوت تلفظ و یا تفاوت در نوع گوینده بات کلمه تست دارا یعدم تشابه زیادی است براحتی از روند تصمیم گیری در مورد تعیین میزان عدم تشابه کلی خارج می شود. به طور معمول می توان k را برابر 2 یا 3 گرفت.

با پیاده سازی الگوریتم نتایجج بسیار خوبی بدست آمد و نرخ بازشناسی حدود ٪25 افزایش یافت.

فلوچارت مربوط به الگوریتم knn

---

[8] - K-nearest neighbour hood
[9] - Cluster

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

حذف DC

[illegible]

وجود مقدار کمی DC در سیگنال باعث ایجاد مشکلاتی در زمینه بازشناسی و همچنین در نحوهٔ جداسازی سیگنال صحبت از زمینه می شود.

دلیل خراب شدن نرخ بازشناسی به وسیلهٔ مقدار DC وابسته بودن ضریب اول کپستروم انرژی فریم است، و میزان انرژی فریم هر رابطه مستقیم با میزان DC آن دارد. بنابراین برای از بین بردن تاثیر DC روی ضریب کپستروم، بوسیله میانگین گیری و پیدا کردن مقدار آن، با کم کردن مقدار DC از فریم این مشکل را تا حدودی حل می کنیم.

روش دیگر برای حذف DC استفاده از یک فیلتر بالاگذر می باشد. این فیلتر FIR بوده و دارای فرم زیر می باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این فیلتر علاوه بر حذف DC بسیاری از نویزهـای فرکـانس پـائین را حذف می کند. تاثیر دیگر مقدار DC برروی دقت جداسـازی سـیگنال صحبت از روی زمینه می باشـد. بـا بـالا بـودن میـزان DC سـیگنال تقارن حول محور زمان برای سیگنال بهم مـی خـورد و ممکـن اسـت، سیگنال اصلاً به حد منفی نرسد و درنتیجه اصلاً عبور از صـفر نکنـد. حذف DC یک فریم باعث ایجاد تقارت حول محور زمان شده و دقـت اندازه گیری میزان عبور از صفر را افزایش می دهد.

با حذف مقدار DC طی دو مرحله میـزان نـرخ بازشناسـی 2٪ بهبـود یافت.

□□□ □□□□□ □□□□ □□□□□□□□

حذف مقدار میانگین ضرایب کپستروم نشان دادن که یک راه سـاده و خوب برای از بین بردن تاثیر کانال انتقـال ماننـد تلفـن میکـروفن مـی باشد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این دو روش بدین صورت عمل می کند که مقدار میانگین بردار مشخصه را برای هر گورنده بدست می آورد. سپس این مقدار را از تمامی بردارهای مشخصه کم می کند.

این روش، روش بسیار خوبی می باشد، و بسیار مفید برای نرمالیزه کردن خواص طیفی نسبت به محیط می باشد با انجام این کار روند بازشناسی بسیار افزایش پیدا کرد و نرخ بازشناسی به حدود 20٪ رسید.

MEI Scald-MFCC

مطالعه روی نحوه شنیدن انسان و مدل گوش نشان داده است که درک انسان از محتوای فرکانسی یک صوت خواه بصورت یک تن تک صدا باشد یا یک سیگنال صحبت از یک مقیاس خطی پیروی نمی کند. گوش انسان بصورت غیرخطی در محور فرکانس عمل می کند و دستاوردهای تجربی نشان می دهد که با طراحی یک سیستم در قسمت پیش پردازش که مانند روش فوق در گوش عمل می کند، کارآیی بازشناسی گفتار را بهبود می بخشد. با مطالعاتی که دانشمندان روی ساختمان گوش انسان انجام دادند، و به کمک چندین

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

آزمایش، توانستند نحوهٔ نگاشت بین مقیاس فرکانسی حقیقی _(HZ) و مقیاس دریافت شده گوش (Mels) را پیدا کنند، این تابع بصورت زیر عمل کند.

همان طور که در شکل مشاهده می کنیم مقیاس MEI در محدودهٔ زیر 1KHZ به صورت خطی عمل می کند و در محدودهٔ بالای 1KHZ بصورت لگاریتمی، عمل خواهد کرد.

برای شبیه سازی رفتار غیرخطی گوش در مقابل فرکانسها از فیلتر استفاده می کنیم. همان طور که گفتیم این فیلتر بانک در محدودهٔ زیر 1KH بصورت خطی عمل می کند و در بالای محدودهٔ بصورت لگاریتمی عمل خواهد کرد.

همان طور که در شکل دیده می شود، درمحدودهٔ زیر lKHZ تعداد بیشتری فیلتر قرار دارد، و این امر بیان می کند که بیشتر اطلاعات طیف فرکانسی صورت در این بخش محدود شده.

نوع پنجره ها در این پیاده سازی مثلثی انتخاب شد. برای بهبود کارآیی ضرایب کسپتروم از فیلتر بانک استفاده می کنیم. به این

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

ترتیب که با محاسبهٔ مقدار FFT فریم آن را به فضای فرکانسی می بریم و پس از آن با مقیاس MEL نمونه برداری می کنیم. و عکس تبدیل فوریه می گوییم.

با محاسبهٔ میزان FFT یک فریم، مقدار دانه را در هر فرکانس بدست می آوریم. سپس با ضرب کردن گین هر فیلتر و حساب کردن مجموع دانه هایی که فرکانس آن ها در محدودهٔ آن فیلتر می باشد. میزان انرژی در آن فیلتر را بدست می آوریم و بدین ترتیب با این کارها پوشش انرژی سیگنال را در محدودهٔ فرکانس بدست آورده ایم.

سپس از طریق تبدیل DCT مقدار ضرایب کپستروم را حساب می کنیم. البته قبل از اعمال تبدیل از مقادیر لگاریتم گرفته و سپس تبدیل را انجام می دهیم.

که در آن mj میزان انرژی هر فیلتر می باشد و N تعداد فیلترهاست

الگوریتم فوق پیاده سازی شد و تاثیر خوبی در نتایج بازشناسی گفتار داشت و نرخ بازشناسی را حدود 10٪ افزایش داد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این آخرین مرحله از بهبود سیستم بود و درصد بازشناسی گفتار برای 20 کلمه قرآنی هرکدام توسط 4 نفر گفته شد تست گردید. این 4 نفر از سنین مختلف بودند. در نهایت میزان بازشناسی گفتار به دقت 99٪ رسید و سیستم به حالت اشباع خود رسید.

بازشناسی گفتار بوسیله مدل مخفی مارکوف

پیاده سازی

الگوریتم سیستم Hmm بسیار پیچیده و مشکل می باشد. بدلیل حجم محاسبات بالا و سر و کار داشتن با اعداد زیر 1 در بعضی از مواقع موجب Overflow شدن سریع سیستم خواهد شد.

جهت پیاده سازی باید همواره این مساله را درنظر داشته باشیم و همواره باید مقیاس بندی را رعایت کنیم. نکته دیگری که در مورد پیاده سازی Hmm باید مورد توجه قرار بگیرد، مقدار واریانس می باشد. اگر مقدار واریانس از یک حد کوچک باشد. به دلیل تقسیم بر عدد صفر شدن در هنگام گرفتن دترمینال باعث ایجاد خط شود. پس

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

حتماً باید در هنگام بروز چنین حالتهایی مقدار احتمال را یک مقدار کوچک درنظر گرفت.

[illegible]

[illegible] Hmm [illegible] Hmm [illegible] Matlab [illegible] Hmm [illegible].

مقدار دهی اولیه Hmm

پس از مشخص نمودن ساختار اولیه یک مدل Hmm نوبت به مقداردهی آن می رسد.

برای مقداردهی اولیه یک مدل Hmm به طریق زیر عمل می کنیم. ابتدا یک کلمه را از بین کلمه هایی که برای آموزش در نظر گرفته ایم انتخاب می نماییم. سپس ا" را به تعداد N قسمت تقسیم می کنیم. در

[10] - Scaulhg

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اینجا N تعداد حالات می باشد و قبل از مقداردهی باید میزان N و M ، مشخص باشد.

پس از تقسیم ماتریس مشخصه به N قسمت، هر کدام از قسمتها را می توان بعنوان یک حالت در نظر گرفت که باید توسط M ترکیب مدل شود.

در این قسمت از بردارهایی که همگی در یک حالت قرار گرفته اند، استفاده کرده و آنها را به عنوان دیتای ورودی به الگوریتم داده و به M، بردار جدید کوانتیزه می کنیم. بدین ترتیب می توان مقادیر میانگین هر حالت و ترکیب را بدست آورد.

برای بدست آوردن واریانس باید از تعداد بیشتری دیتا استفاده کرد:

مقادیر ماتریس انتقال حالت را می توان بصورت رندم مقداردهی نمود یا طبق الگوریتم زیر عمل نمود.

حال برای محاسبهٔ واریانس باید تعداد دیتای بیشتری داشته باشیم.

2 راه داریم: راه اول بصورت غیر خطی عمل کنیم و به جای تقسیم یک کلمه به N قسمت مثلاً 10 کلمه را به N قسمت تقسیم کنیم و

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

سپس تمام قسمتهای اول را به عنوان حالت اول در نظر بگیریم و همهٔ قسمتهای Nام را به عنوان حالت N ام.

راه دوم: استفاده از الگوریتم riterbi می‌باشد. با استفاده از الگوریتم می‌توان فهمید هر کدام از اعضای ماتریس مشخصه متعلق به کدام حالت هستند. برای پیدا کردن واریانس بر طبق الگوریتم riterbi به این صورت عمل می‌کنیم که با اعمال مدل اولیه که در آن مقدارهای واریانس یک عدد ثابت قرار داده شده بر روی کلماتی که توسط گوینده‌های مختلفی بیان شده. بردارهایی از ماتریس مشخصه که در یک حالت قرار می‌گیرند را پیدا کرده با استفاده از الگوریتم میزان میانگین و واریانس را محاسبه می‌کنیم.

تخمین بیشترین شباهت

پس از بدست آوردن مدل اولیه نوبت: بهبود و ضرایب می‌رسد. این کار را توسط الگوریتم Baum-welch انجام می‌دهیم. پس از 10 الی 20 بار، تکرار الگوریتم سرانجام همگرا خواهد شد.

بازشناسی گفتار

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پس از بدست آوردن مدول اولیه برای هر کلمه نوبت به مرحلهٔ تست می‌رسد. برای بوسیلهٔ الگوریتم Forward می‌توان برای هر یک از کلمات تست بدست آورد و نرخ بازشناسی گفتار را بدست آورد.

محدودیتهای روش Hmm

بدلیل طولانی بودن روند آموزش شبکه بهینه کردن پارامترهای آن بسیار دشوار می‌باشد. این پارامترها می‌تواند شامل، طول پنجره و میزان هم‌پوشانی، تعداد ضرایب کپستروم، تعداد حالتها و ترکیب‌ها و ...

کم بودن دیتای آموزش نیز مشکل دیگری است که باعث ضعیف شدن مدل خواهد شد. در صورتیکه نتوانیم تعداد زیادی دیتا از گوینده‌های مختلف جمع‌آوری کنیم. مدل از لحاظ عدم وابستگی به گوینده بسیار ضعیف خواهد شد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بخش دوم لرزیابی نحوۀ بیان گفتار قرآنی

در بخش دوم پس از پایان یافتن مراحل بازشناسی گفتار قرآنی به بررسی الگوریتم‌های ارائه شده در ارزیابی نحوۀ بیان گفتار قرآنی خواهیم پرداخت. در ارزیابی نحوۀ بیان گفتار قرآنی، سیستم به این صورت عمل می‌کند که بسته به نوع تلفظ در مورد یک کلمه یک نمره بین 0 تا 100 به آن می‌دهد.

برای ارزیابی یک کلمه راههای مختلفی وجود دارد. در اینجا هدف ما شناسایی تلفظ درست کلمه و ادای کلمه مانند صدای استاد می‌باشد.

بدلیل اینکه در این مرحله کلمه مورد نظر از قبل معین می‌باشد، احتیاجی به سیستم بازشناسی گفتار نخواهد بود.

در بخش اول، ارزیابی نحوۀ بیان گفتار قرآنی را توسط الگوریتم و تطابق دهندگی الگو بررسی کرده و در بخش دوم از مدل برای ارزیابی صدا استفاده می‌کنیم.

ارزیابی نحوه بیان گفتار قرآنی بوسیله و تطابق الگو

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در بخش بازشناسی گفتار دیدیم که چگونه پارامترهای سیستم بازشناسی به حالت بهینه رسید. در هنگام بهینه شدن پارامترهای سیستم، مطمئن هستیم که اختلاف مشاهده شده بین دو الگوی صحبت در کمترین حد ممکن خواهد بود.

جهت ارزیابی نحوهٔ تلفظ کلمات قرآنی باید صدای کاربر را با صدای استاد مقایسه کنیم. در اینجا نیز همان مطالب گفته شده در مورد هم ترازسازی و نرمالیزاسیون صادق می‌باشد. همچنین باز هم مطالبی را که در مورد استفاده از الگوریتم KNN جهت پیدا کردن کلمهٔ مبنا مقایسه با صدای کاربر گفتیم نیز به قوت خویش باقی است.

در قسمت اول توسط الگوریتم و پارامترهای سیستم بهینه بازشناسی گفتار، به مقایسه صدای کاربر با صدای استاد می‌پردازیم.

ابتدا سیگنال صحبت کاربر را به پنجره‌هایی با طول و هم‌پوشانی معین تقسیم نموده و پنجره همینگ را در آن ضرب می‌کنیم و آن را در یک متغیر ذخیره می‌کنیم.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

سپس از هر فریم ضرایب کپستروم مربوط به آن را استخراج نموده و ماتریس مشخصه را تشکیل داده و توسط الگوریتم و فاصلهٔ اقلیدسی بین هر دو فریم، طول کلمهٔ گفته شده توسط کاربر را با صدای استاد یکی می‌کنیم.

پس از انجام مرحلهٔ هم ترازسازی و پیدا کردن فریمهای معادل اکنون نوبت به مقایسه خواص طیفی دو کلمه خواهیم پرداخت.

قبل از بیان نحوهٔ مقایسه به بررسی روشهای مقایسه خواص طیفی دو کلمه می‌پردازیم.

Log spectral Distance

فرض کنید ما دارای دو طیف فرکانسی ، هستیم. تفاوت بین دو طیف در مقیاس لگاریتم و در محور فرکانس برابر است با

یک راه ساده برای تعریف میزان اعوجاج بین و بیان آن به صورت زیر می‌باشد.

که برای حالت گسسته آن

که در آن ، تبدیل‌های گسسته طیف طیف فوریه می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

برای مقدار P=1 روابط بالا بیان کنندهٔ مقدار متوسط قدر مطلق لگاریتم اعوجاج طیف می‌باشد و برای P=2 این روابط میزان rms ، قدر مطلق لگاریتم اعوجاج را نشان می‌دهند که کاربردهای وسیعی در بازشناسی گفتار داد.

Cepstral Distance

ضریب کپستروم مختلط سیگنال بر مبنای عکس فوریه لگاریتم طیف فرکانسی سیگنال بیان می‌شود.

که در آن ضرایب حقیقی هستند و اغلب از آنها به عنوان ضرایب کپسترول یاد می‌شود. برای ، با بکار بردن قوانین بار مدال می‌توان (یعنی در رابط) را به ضرایب کپسترول ربط داد.

Weignted cepstral Distance

دلیل استفاده زیاد از فاصلهٔ ضرایب کپستروم این امر است، که از راه ضرایب کپستروم می‌توان میزان قدر مطلق لگاریتم فاصلهٔ فرکانسی را تخمین زد.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

می‌توان نشان داد تحت شرایط خاص، ضرایب کپستروم به غیر از ، دارای مقدار متوسط صفر و ورایانس متناسب با عکس مجذور اندیکس ضریب می‌باشند.

برای نرمالیزه کردن ضرایب کپستروم بوسیلهٔ واریانس می‌توان نشان داد.

Distance base on LPC

از آنجائیکه LPC نشان دادن روشی بسیار مناسب در پردازش سیگنال می‌باشد. می‌توان از آن برای نشان دادن طیف فرکانس فریم نیز استفاده نمود.

که در آن و دو تابع هستند که به بررسی خواص طیف فرکانسی می‌پردازند. در حقیقت همان پوش می‌باشد.

از این روش ارائه پاسخ فرکانسی بر مبنای LPC می‌توان روشهای مختلفی را برای بیان اختلاف دو فریم

MLSP بعنوان روش جهت می‌نیمم کردن تفاوت خواص طیفی زمان کوتاه یک فریم ارائه شد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پیدا کردن منحنی فاصله‌ها

در بخش‌های قبل با روشهای پیدا کردن فاصله آشنا شدیم. پس از هم ترازسازی دو کلمه کاربر و استاد با یکدیگر توسط الگوریتم اکنون نوبت به پیدا کردن منحنی فاصلهٔ بین هر فریم صدای استاد و صدای کاربر بر اساس روشهای گفته می‌شود.

از روی این منحنی‌ها 3 اطلاعات را استخراج می‌کنیم:

1- میزان اعوجاج متوسط: این مقدار بیان می‌کند که مقدار متوسط فاصلهٔ بین دو الگو چقدر می‌باشد.

2- نقاط ماکزیمم: مقدار این نقطه ماکزیمم بیان می‌کند. فاصلهٔ دو فریم در آن نقطه خیلی زیاد بوده و یا به عبارتی در آن فریم صدای کاربر و استاد اصلاً شبیه هم نمی‌باشد.

3- تعداد نقاطی که از یک حد آستانه اعوجاج بیشتر باشند: این نقاط فریم‌هایی را نشان می‌دهند که به طور کلی دارای تعداد غیر قابل قبولی اختلاف با فریم صدای استاد می‌باشد. هر چه تعداد این فریم‌ها

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

بیشتر باشد، این موضوع ثابت می‌شود که صدای کاربر اصلاً شبیه صدای استاد نبود.

همان طور که گفتیم برای انتخاب کلمهٔ مبنا از الگوریتم KNN استفاده می‌کنیم. برای انتخاب بهترین تطابق تنها از مقدار متوسط اعوجاج استفاده خواهیم کرد. قابل ذکر است مقدار متوسط اعوجاج به طریقی توصیف کنندهٔ دو مقدار دیگر هم می‌باشد. هر چقدر مقدار ماکزیمم و تعدا نقاط عبور از آستانه بیشتر باشد مقدار متوسط هم افزایش می‌یابد و بالعکس.

امتیازدهی به گوینده

برای امتیازدهی به گوینده احتیاج به یک سری دیتا آموزش دادیم. این دیتاها باید از لحاظ تلفظ کاملاً رعایت قوانین را کرده باشند و مطابق با صدای استاد باشند.

با مقایسه این کلمات می‌توان مقادیر متوسط اعوجاج، ماکزیمم، و حد آستانه برای صداهای قابل قبول را بدست آورد و از این مقادیر برای امتیازدهی به صدای سایر گویندگان پرداخت.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

در این مرحله مقادیر متوسط و واریانس 3 پارامتر یاد شده پیدا خواهد شد.

استفاده از فرکانس فرمنت در ارزیابی نحوهٔ بیان گفتار قرآنی

در فصل دوم با مفهوم حروف صدادار و فرکانسهای فرمنت متناظر آنها آشنا شویم. همچنین در بخش دوم، دربارهٔ چگونگی استخراج فرکانس فرمنت بحث کردیم.

بیشتر افراد در خواندن کلمات قرآنی در هنگام تلظ وزن کلمه دچار مشکل خواهند شد. وزن کلمه در حقیقت همان حروف صدادار کلمه می‌باشد. پس در هنگام ارزیابی تلفظ اگر بتوانیم بنحوی وزن کلمه یا همان حروف صدادار را شناسایی کنیم. در مورد کلمات قرآنی، قدم خوبی را برای این کار برداشته‌ایم.

اما مشکل اساسی در شناسایی حروف صدادار می‌باشد. یکی از روش‌های شناسایی حروف صدادار استفاده از فرکانس فرمنت می‌باشد. اما فرکانس‌های فرمنت زیاد قادر به جداسازی حروف صدا نزدیک به هم مثل «ر – ی» نیستند.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

[illegible]

در هنگام ارزیابی تلفظ درست است که بین «بک» و بیک» بسیار تفاوت است، اما بخش مهم این است که تقریب اگر قانون بالا را در نظر بگیریم هر دو کلمه را درست بیان کرده‌اند. ولی اگر کاربری می‌گفت بک یا باگ قضیه فرق می‌کرد و باید نمرهٔ صفر به آن داده می‌شد.

برای همین یک دسته بندی در فرکانس فرمنت و حروف صدادار انجام دادیم.

با این گروه بندی می‌توان براحتی از طریق فرکانس فرمنت می‌توان به شناسایی حروف صدادار متعلق به هر گروه پرداخت.

باز هم در اینجا تأکید می‌کنیم در اینجا هدف شناسایی حروف صدادار نمی‌باشد بلکه هدف تشخیص گروه مربوط به آن فریم می‌باشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

روش کار به این صورت است که ابتدا فریم‌های واکه دار معین می‌شود، سپس فرکانس فرمنت آن استخراج شده و به یک الگوریتم ساده داده می‌شود که نوع گروه آن معین می‌شود.

الگوریتم خیلی ساده عمل می‌کند.

همان طور که مشاهده می‌کنید. گروه دوم دارای بیشترین اختلاف در فرکانس‌های اول و دوم است و گروه اول کمترین اختلاف را دارست.

همچنین برای گرو محدودهٔ فرکانس مجاز برای فرمنت اول و دوم معلوم است.

این محدوده‌ها برای هر گروه جوری است که دارای هم‌پوشانی نمی‌باشند و براحتی از هم قابل تفکیک است.

یک الگوریتم ساده با مقایسه فرکانس فرمنت با محدودهٔ مجاز گروه، گروه مربوط به آن فرکانس را پیدا می‌کند. در صورتیکه، فرکانس فرمنت متعلق به هیچ گروهی نباشد از آن صرفنظر شده و در ارزیابی مورد استفاده قرار نمی‌گیرد.

استفاده از هر ارزیابی نحوهٔ بیان

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

کوانتزاسیون برداری روش ساده ای برای مدل کردن یک کلمه می‌باشد. همچنین در کوانتزاسیون برداری می‌توان از کلماتی استفاده کرد که توسط گوینده‌های مختلف بیان شده است. بدین ترتیب برداری که بدست می‌آید می‌تواند متوسط بردارهای مشخصه کلمات با گوینده‌های متفاوت باشد و این یعنی مستقل از گوینده بودن.

دلیل استفاده از کوانتزاسیون برداری اول با کوانتیزه کردن یک کلمه در حقیقت ما بخش‌های مختلف آوایی یک کلمه را مدل می‌کنیم، یعنی هر کدام از بردارهای نمونه یک نماینده از خواص طیفی مربوط به آن بخشی از کلمه می‌باشد. 2- با بکار بردن گوینده‌های مختلف می‌توان بردارهایی را بدست آورد که در خود خاصیت مستقل از گوینده بودن را بصورت نهفته دارند.

با ذکر دلایل بالا، کوانتزاسیون برداری می‌تواند کمک بسیار زیادی در حوزهٔ بازشناسی گفتار انجام دهد. ولی ما از خاصیت در ارزیابی گفتار استفاده می‌کنیم. در روش تطابق اگر گفتیم می‌خواهیم دو طیف فرکانسی مربوط به یک فریم را از هم کم کنیم. در این روش همواره

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

خطا وجود دارد. چون هیچ وقت طیف فرکانسی دو سیگنال مثل هم نخواهد بود. ولی اگر ما بیاییم عمل کوانتزاسیون انجام دهیم. بردارهای مشخصه که مشابه هستند به یک بردار ثابت کوانتیزه کنیم. آنوقت می‌توان اختلاف بین سیگنال را به حداقل رسانید.

کوانتیزاسیون برداری فقط برای بردارهای مشخصه کپسترال و فاصلهٔ ضرایب کپستروم قابل پیاده سازی است. در این، تعداد بردارها نقش مهمی در دقت سیستم خواهند داشت.

در روش کوانتزاسیون برداری دو روش برای بدست آوردن بردارهای نمونه وجود دارد: 1- روش اول استفاده از یک کلمه برای بدست آوردن برچسب‌های آوایی مربوط به آن کلمه است. یعنی فقط یک کلمه را مدل می‌کنیم.

در این صورت تعداد بردارها بسته به طول کلمه می‌تواند بین 6 تا 10 عدد انتخاب شود.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

2- استفاده از کلمات مختلف برای بدست آوردن بردارهایی کوانیزه:

در این صورت ما دارای تعداد زیادی بردار خواهیم بود که بیانگر خواص طیفی مربوط به واج‌های مختلف می‌باشند.

هر دو روش فوق پیاده سازی شد

در پیاده سازی روش اول، بسته به طول کلمه هر کلمه را با یک کتاب کد 10 تا 6 تایی مدل کردیم. در این روش ابتدا ضرایب کپستروم هر فریم با کتاب کد مقایسه می‌شدند و نزدیکترین بردار به آن ضرایب در کتاب کد پیدا می‌شود. در این مرحله به جای جایگزین کردن اندیکس کتاب کد در فریم. بردار نمونهٔ آن اندیکس به جای ضرایب فریم قبلی جای‌گزین می‌شود.

یعنی ضرایب کپستروم جدید از جایگزینی بردار نمونه کتاب کد بدست می‌آید. در این مرحله عمل کوانتیزاسیون بدون تغییر سایز انجام داده ایم.

www.kandoocn.com

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

پس از بدست آوردن ماتریس مشخصه جدید برای صدای استاد و کاربر همان مراحل قبل را در هم ترازسازی و پیداکردن 3 پارامتر طی می‌کنیم.

قابل ذکر به دلیل عوض شدن ماهیت موضوع، کلیه پارامترهای مربوط به مقادیر متوسط قابل قبول و حدود آستانه عوض می‌شوند و باید به شیوهٔ جدید دو بار مرحلهٔ آموزش طی شود.

این الگوریتم پس از پیاده سازی نتایج خوبی داشت.

با پیاده سازی روش دوم دو مشکل اساسی ایجاد شد:

پیدا کردن یک کتاب کد با تعداد codeهای بالا احتیاج به تعداد زیادی کلمه دارد و هر کلمه نیز باید توسط گوینده‌های مختلف بیان شده باشد.

به دلیل بالا بودن حجم کار زمان محاسبات بالا می‌رود.

مشاهده شد با بالارفتن اندازهٔ کتاب کد. دقت سیستم پائین می‌آید و پیدا کردن مقادیر آستانه و قابل قبول بسیار سخت تر شد. بنابراین روش دوم بکار گرفته نشد.

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

این روش، روش بسیار خوبی است. اما مشکل اصلی آن پیدا کردن درست و دقیق فرکانس فرمنت و فریم‌های واکه‌دار می‌باشد.

نمره دهی نهایی

پس از بدست آوردن مقادیر قابل قبول برای 3 پارامتر گفته شده در مرحلهٔ آزمایش، در مرحلهٔ تست اکنون نوبت به نمره دهی نهایی می‌رسد.

برای اینکه یک کلمه که توسط کاربر گفته شده، شرایط لازم را برای نمره دادن کسب کند باید از 3 پارامتر آن حداقل 2 پارامتر آن که یکی از آن‌ها مقدار متوسط اعوجاج می‌باشد بالای مقادیر قابل قبول باشد در غیر این صورت برای برای آن نمرهٔ صفر در نظر گرفته می‌شود.

در قسمت آزمایش در مرحلهٔ اول علاوه بر مقادیر آستانه قابل قبول برای 3 پارامتر، مقادیر متوسط این 3 پارامتر نیز استخراج شدند، این 3 پارامتر به ما می‌گویند کلمه ای که درست بیان شده به طور معمول دارای این مقادیر است، اما پارامترهای مجاز بیان می‌کنند که برای

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

اینکه کلمه ای درست بیان شده باشد باید مقدار پارامترهـایش از ایـن حد تجاوز نکند.

از مقدارهای متوسط می‌توان برای مـدل کـردن کلمـهٔ خـوب ادا شـده استفاده کرد. برای مدل کردن می‌توان از تابع توزیـع نرمـال اسـتفاده کرد. ما فقط از طرف راست منحنی استفاده می‌کنیم. این نقـاط نقـاطی هستند که مقدار آنها از حد متوسط بزرگتر است.

برای صداهایی که 3 پارامتر آنها از مقدار متوسط مربوطه‌شان کمتـر است نمره 100 منظور خواهد شد.

سرانجام با کمک یک رابطهٔ نهایی بـه امتیـازدهی بـه صـدای شـخص طبق پارامترهای مربوط به آن پرداخته خواهد شد.

در اینجا و پارامترهای وزن دهنده برای مدل محسـوب مـی‌شـوند. از آنجا که این 3 پارامتر هر 3 دارای یک اهمیت یکسان نمی‌باشـند، پـس نقش آن هم در نمـره‌دهـی نبایـد یکسـان باشـد. طبیعـی اسـت مقـدار متوسط اعوجاج پارامتر قدرتمندی نسبت به مـاکزیمم اخـتلاف اسـت. بنابراین تغییرات سریع در اختلاف بین مقدار متوسط اعوجاج مـدل و

جهت خرید فایل به سایت www.kandoocn.com مراجعه نمایید یا با شماره های ۰۹۳۶۶۰۲۷۴۱۷ و ۰۹۳۰۳۵۲۲۸۸۶ تماس حاصل نمایید

مقدار متوسط اعوجاج صدای شخص باید بسیار تأثیرگذارتر در نحوهٔ نمره‌دهی باشد جدول مربوط به و بصورت زیر است.

استفاده از Hmm در ارزیابی نحوهٔ بیان

در بخش ارزیابی نحوهٔ بیان بوسیله Hmm نیز دچار همان مشکلات بخش بازشناسی هستیم، یعنی کمبود دیتای آموزشی. بنابراین بصورت عملی نتوانیم از مدل Hmm در ارزیابی نحوهٔ بیان خیلی خوب استفاده کنیم. در این بخش ما فقط توانستیم بگوئیم کلمه رد است یا قبول. ولی نمی‌توانستیم به آن نمره دهیم. بعلاوه مشکل مستقل از گوینده بودن هم زیاد قبول نشد.

برای رد کردن کلمه مقدار یا نمره ای که مدل کلمه به دنبالهٔ آوایی دادن استفاده می‌کنیم.

روش کار به این صورت است که ابتدا در مرحلهٔ آموزش مقادیر را برای چند کلمه که تلفظ درستی دارند پیدا کردن و یک مقدار آستانه را بطور تجربی بدست می‌آوریم.

در مرحلهٔ تست با مقایسه کلمهٔ کاربر می‌توان آن را رد یا قبول کرد.