
بسم الله الرحمن الرحيم

دانشگاه آزاد اسلامی

واحد تهران مرکز

موضوع:

پردازش زبان فارسی

فهرست مطالب

فهرست شکل ها

فهرست جداول

۱ بازشناسی گفتار گسسته و پیوستهٔ فارسی

۱ـ ۱ مقدمه

هدف در بخش بازشناسی گفتار گسسته و پیوستهٔ فارسی، دستیابی به سیستم‌های بازشناسی گفتارفارسی با توانائی بازشناسی گفتار گسسته در چارچوب دایرهٔ کلمات تعیین شده و نرخ بازشناسی قابل‌قبول، بر روی دادگان تعریف شده و نیز بازشناسی گفتار پیوستهٔ فارسی با دایرهٔ کلمات مشخص و درچارچوب دادگان اصلاح شدهٔ پیوسته و با نرخ بازشناسی موردنظر می‌باشد.

در هر یک از این دو راستا، اقدامات صورت گرفته شامل مطالعه، بررسی و پیاده‌سازی روش‌های‌مطرح در بازشناسی جهت دستیابی به نتایج مطروحه بوده است. تهیهٔ دادگان‌های گفتاری گسسته وپیوسته با مشخصات ازپیش تعیین شده، پیاده‌سازی روش‌های مبتنی برHMMهای با چگالی‌مشاهدات پیوسته و نیز شبکه‌های عصبی، ارزیابی و بهبود روش‌های پیاده‌سازی شده و احیاناً اصلاح یاپیشنهاداتی جهت اصلاح این روش‌ها می‌باشد.

۱ـ ۲ بازشناسی گفتار گسستهٔ فارسی

بازشناسی گفتار گسسته از جمله بخش‌های مهمّ در بحث بازشناسی گفتار می‌باشد که در کاربردهای‌مختلف ازجمله فرمان و کنترل موارد استفادهٔ فراوانی دارد.

۱ـ ۲ـ ۱ دایرهٔ کلمات و دادگان

دایرهٔ کلمات از جمله عوامل مهمّ در تأمین بازشناسی با کیفیت مناسب گفتار گسسته می‌باشد. هدف‌در این طرح پژوهشی دستیابی به بازشناسی در محدودهٔ کلمات صد کلمه بوده است. با این همه، از آنجاکه پیش از آغاز این طرح، یک دادگان گفتاری از کلمات گسسته (ارقام فارسی) با تعداد ده کلمه و بصورت ناوابسته به گوینده موجود بوده است، مرحلهٔ ابتدائی کار بر روی این دادگان به انجام رسید. لازم به اشاره‌است که این دادگان از مجموعه‌ای از دانشجویان در محدودهٔ سنی ۲۰ تا ۳۰ سال از هر دو جنسیت ضبط‌گردیده و مشتمل بر دو بخش آموزشی و آزمایشی بوده است. شرایط ضبط، محیط با نویز زمینهٔ کم بوده‌است.

دادگان دیگری برای دستیابی به اهداف این طرح پژوهشی طراحی و ضبط گردید. در طراحی این‌دادگان تلاش شد کلمات مورد استفاده، ضمن پوشش یک محدودهٔ صدکلمه‌ای، از ویژگی‌های خاصی‌برخوردار باشند. به همین جهت، مجموعهٔ کلمات شامل ۳۸ کلمهٔ اصلی تشکیل‌دهندهٔ کلیهٔ اعداد تاشش رقمی در زبان فارسی (صفر تا نوزده، مضارب ده، مضارب صد و مضارب هزار) بهمراه اسامی‌شهرهای مهمّ ایران می‌باشد. درعین حال ضبط دادگان بصورت مستقیم به داخل کامپیوتر از طریق یک‌میکروفون دینامیک مناسب(Sony F-VK98) و در محیط اداری (دفتر کار) با نویز محیط معمولی‌صورت گرفت. در مجموع از ۱۰۲ گوینده (نیمی مرد و نیمی زن)، هر یک از ۱۰۰ کلمهٔ تعریف شده، ۵ بارضبط گردیدند. البته باتوجه به ضبط گفتار در محیط دانشگاهی و از افرادی که دارای مدرک تحصیلی‌حداقل دیپلم متوسطه بوده‌اند، دسترسی به افراد با سنّ بالای ٤٠ سال عملاً مقدور نبوده و لذا سنّ افرادشرکت‌کننده در تهیهٔ این دادگان محدودهٔ وسیعی را شامل نمی‌شود [۱][۲].

پس از جمع‌آوری دادگان، اقدام به بررسی تک‌تک حدود ۵۱/۰۰۰ بیان گفتاری موجود گردید واشکالات مختلف موجود در آنها از قبیل صداهای اضافی ضبط شده، بریدگی کلمات در آغاز یا پایان،برخی لهجه‌ها و گویش‌های نامعمول، صدای دهان در آغاز ادای کلمات و ... که علیرغم پیش‌بینی‌های‌لازم در ضبط برخی کلمات وجود داشتند، مشخص و بیان‌های دارای مشکل حذف گردیدند [٤]. سپس‌دادگان به دو بخش آموزشی و آزمایشی تقسیم گردید. این دادگان به عنوان دادگان اصلی در این بخش ازطرح پژوهشی پردازش گفتار فارسی مورد استفاده واقع گردید.

درعین حال در بعضی مراحل، ازجمله در بازشناسی به کمک شبکه‌های عصبی نیز یک مجموعهٔ ۱۲کلمه‌ای از کلمات برای کنترل یک یونیت دندانپزشکی مورد استفاده قرار گرفت که بنابه عللی که بعداًاشاره خواهد گردید مورد استفادهٔ بیشتری نیافت [۲].

۱ـ ۲ـ ۲ سیستم‌های مورد استفاده

جهت دستیابی به بازشناسی کلمات گسسته، دو شیوه مورد توجه قرار گرفت که ذیلاً به این دو شیوه‌می‌پردازیم.

۱ـ ۲ـ ۲ـ ۱ مدل‌های مارکوف پنهان با چگالی مشاهدات پیوسته (CDHMMs)

مدل‌های مارکوف پنهان به عنوان یکی از شیوه‌های موفق، امروزه در کاربردهای بازشناسی گفتارپیوسته و گسسته کاربردهای فراوانی یافته‌اند. این مُدل‌ها باتوجه به توانائی بالائی که در مُدل‌نمودن‌ویژگی‌های گفتار و بخصوص ویژگی دینامیک گفتار دارند، مورد بررسی و استفادهٔ فراوانی در این زمینه‌قرار گرفته‌اند. مناسب‌ترین این مُدل‌ها جهت اینگونه مدل‌سازی، از نقطه‌نظر

چگالی احتمالات‌خروجی، انواع با چگالی مشاهدات پیوسته می‌باشند. این‌گونه ازHMM ها به دلیل دقت بالائی که درمُدل‌سازی ارائه می‌دهند مناسب تشخیص داده شده‌اند [۵][٦]. به همین دلیل، در این پیاده‌سازی نیز ازاینگونه مُدل‌های آماری استفاده گردید.

۱ـ۲ـ۲ـ۱ـ۱ توپولوژی مُدل‌ها

پیاده‌سازی اوّلیه بر روی دادگان گفتاری ده‌کلمه‌ای براساسCDHMMs دارای ٦ حالت صورت‌گرفته بود. در کاربرد ۱۰۰ کلمه‌ای، باتوجه به اینکه طول برخی کلمات طولانی‌تر بوده، ازCDHMMs با۸ حالت استفاده گردید. اینHMM ها از نوع چپ ـ راست و بدون انتقال جهشی درنظر گرفته شدند [٤].اشاره به این نکته نیز در اینجا خالی از اهمیت نیست که باتوجه به توانائیHMM ها در مُدل‌سازی‌دینامیک سیگنال گفتاری، مُدل‌نمودن سکوت ابتدا و انتهای کلمه نیز برعهدهٔ حالت‌های آغازین و پایانیHMM ها گذاشته شد و از آشکارسازی نقاط ابتدا و انتهای گفتار پرهیز گردید.

۱ـ۲ـ۲ـ۱ـ۲ استخراج ویژگی‌های طیفی

ویژگی‌های طیفی مورد استفاده در این بخش از طرح پژوهشی، ضرائب کپسترال LP انتخاب گردیدند.برای استخراج این ضرائب از سیگنال گفتاری رقمی‌شده، مراحل زیر به ترتیب به انجام رسید [۱][٤].

۱ـ پیش‌تأکید[1] . وظیفهٔ این بخش مسطح‌کردن طیف فرکانسی سیگنال گفتار برای پیشگیری از مشکلات‌احتمالی ناشی از محدودیت طول لغت دیجیتال در پردازش‌های بعدی می‌باشد.

۲ـ تقسیم به قاب‌ها[2] . سیگنال گفتار در این بخش به مجموعه‌ای از بلوک‌های متداخل با طول زمانی‌حدود ۲۰ تا ۲۵ میلی‌ثانیه و فاصلهٔ فریم ۱۰ تا ۱۵ میلی‌ثانیه تقسیم می‌گردد. هر قاب سپس مورد پردازش‌مستقل قرار گرفته و ویژگی‌های سیگنال گفتاری در داخل آن پایدار فرض می‌شود.

۳ـ اعمال پنجرهٔ همینگ[3] که باعث کاهش دامنهٔ نمونه‌های کناری قاب می‌گردد.

٤ـ استخراج ضرائب LPC[4].

۵ـ بدست‌آوردن ضرایب کپسترال[5] با استفاده از ضرائب LPC .

٦ـ اعمال لیفترجوانگ[6] جهت وزن‌دهی ضرائب کپسترال.

[1] Pre-emphasis
[2] Frame Blocking
[3] Hamming Window
[4] Linear Predictive Coding
[5] Cepstral Coefficients
[6] Juang Lifter

۷ـ بدست‌آوردن ضرائب دلتا و دلتا ـ دلتا برای ضرائب کپسترال و انرژی لگاریتمی.

بردارهای ویژگی بدست آمده از مراحل فوق به عنوان ویژگی‌های طیفی نمایندهٔ گفتار در مراحل‌بعدی پردازش مورد استفاده قرار می‌گیرند.

۱ـ۲ـ۲ـ۱ـ۳ آموزش مُدل‌ها

برای بدست آوردن یک سیستم بازشناسی گفتار مناسب، لازم است ابتدا مُدل‌ها به‌نحو مطلوبی مورد آموزش قرار گیرند. یکی از پراستفاده‌ترین روش ها در آموزش مدل‌هایHMM ، روشی مبتنی بر تکنیک‌درستنمائی بیشینه (ML[۷]) است، موسوم به Baum –Welch. این آلگوریتم یک روش تکراری است ورسیدن آن به یک ماکزیمم محلّی برای درستنمائی اثبات شده است [۷]. با استفاده از این آلگوریتم و درشرایط دنباله‌های مشاهدهٔ چندگانه، روابط باز تخمین برای پارامترهای اصلی در یکCDHMM به قرارزیر می‌باشند [۱][٦]:

$$\hat{c}_{jk} = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^{(l)}(j,k)}{\sum_{l=1}^{L}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_t^{(l)}(j,m)} \qquad (۱-۱)$$

$$\hat{\boldsymbol{\mu}}_{jk} = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^{(l)}(j,k)\cdot \boldsymbol{O}_t^{(l)}}{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^{(l)}(j,k)} \qquad (۱-۲)$$

$$\hat{\boldsymbol{U}}_{jk} = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^{(l)}(j,k)\cdot(\boldsymbol{O}_t^{(l)}-\boldsymbol{\mu}_{jk})(\boldsymbol{O}_t^{(l)}-\boldsymbol{\mu}_{jk})^T}{\sum_{l=1}^{L}\sum_{t=1}^{T}\gamma_t^{(l)}(j,k)} \qquad (۱-۳)$$

در این روابط، c_{jk} وزن مخلوط kاُم از حالت jاُم، $\boldsymbol{\mu}_{jk}$ بردار میانگین آن و $\boldsymbol{U}_{jk}$ ماتریس کوواریانس آن می‌باشند و روابط برای $\boldsymbol{\mu}_{jk}$ و $\boldsymbol{U}_{jk}$ به صورت برداری (یا ماتریسی) نوشته شده‌اند. همچنین،$\gamma_t(j,k)$ احتمال بودن در مخلوط kام از حالت jام در لحظهٔ t و مشاهدهٔ $\boldsymbol{O}_t$ است.

پیش از اینکه مدل‌هایCDHMM بتوانند آموزش ببینند، لازم است مُدل‌های اوّلیه با مقادیر مناسبی‌مقداردهی شوند تا دستیابی به ماکزیمم محلّی بتواند مترادف با رسیدن به ماکزیمم کلّی تلقی شود. ازاین‌رو مقداردهی اوّلیهٔ پارامترها از اهمیت فوق‌العاده‌ای برخوردار است. برای این منظور در دو مرحله به‌شرح زیر عمل می‌شود:

الف: کلیهٔ مقادیر دنباله‌های مشاهدهٔ آموزشی موجود برای هر مُدل بطور یکنواخت به تعدادحالت‌های آن مُدل تقسیم شده و بر روی تمامی بردارهای بدست آمده برای هر حالت مقادیر

[7] Maximum Likelihood

بردارهای‌میانگین و واریانس بدست می‌آیند و به عنوان مقادیر ابتدائی برای پارامترهای آن حالت مورد استفاده‌قرار می‌گیرند.

ب: با استفاده از مقادیر ابتدائی فوق، تمامی دنباله‌های مشاهدهٔ موجود با استفاده از آلگوریتم ویتربی‌با حالات مدل مربوطه همردیف می‌شوند. سپس مقادیر همردیف شده بند الف فوق برای بدست‌آوردن پارامترهای هر حالت مورد استفاده قرار می‌گیرند. این عمل تا رسیدن به یک معیار همگرائی و یارسیدن به تعداد موردنظر از دفعات تکرار ادامه می‌یابد.

پارامترهای بدست آمده از مراحل فوق، سپس به عنوان پارامترهای اوّلیه در آلگوریتم Baum-Welch مورد استفاده قرار می‌گیرند.

درخصوص آموزش مدل‌های دارای چگالی‌های مشاهدهٔ مخلوط گوسین[8] ، اگرچه استفاده از روش‌فوق نیز جایز می‌باشد ولی ترجیح داده می‌شود که در مراحل بعدی نسبت به افزایش تعداد عناصرمخلوط اقدام نمود. بنابراین مراحل ابتدائی آموزش عموماً با چگالی‌های مشاهدهٔ تک‌گوسین صورت‌می‌پذیرد و سپس برای افزایش تعداد عناصر مخلوط از آلگوریتمی نظیر روش شکافت مخلوط[9] که دربحث بازشناسی گفتار پیوسته به تفصیل بیشتری مورد اشاره قرار خواهد گرفت، استفاده می‌شود.

نکتهٔ دیگری که در اینجا حائز اهمیت است، استفاده از نمایش درستنمائی بصورت لگاریتمی است که لازمهٔ پیاده‌سازی آلگوریتم‌هائی نظیرBaum-Welch می‌باشد چرا که مقادیردرستنمایی بعد از چند لحظهٔ زمانی به علّت ضرب‌شدن در مقادیر احتمالی بسیار کوچک، به سمت‌صفر میل خواهد نمود که عمل آموزش را غیرممکن می‌سازد. بهمین جهت لازم است که در پیاده‌سازی‌از نمایش لگاریتمی احتمالات و درستنمائی استفاده نمود. در این صورت ضرب‌ها هم تبدیل به جمع شده و بنابراین مشکل میل‌نمودن سریع مقادیر حاصلضرب به سمت صفر نیز از بین خواهد رفت.

۱ـ۲ـ۲ـ۱ـ٤ بازشناسی کلمات جدا

مرحلهٔ بازشناسی کلمات جدا با استفاده از HMMها عموماً به کمک آلگوریتم ویتربی[10] صورت‌می‌گیرد [۵][٦]. این آلگوریتم که براساس شیوهٔ برنامه‌نویسی پویا طراحی گردیده از مزایای چندی به‌شرح زیر برخوردار است:

ـ یافتن مسیر بهینهٔ حالت براساس اصل بهینگی Bellman .

ـ کاهش فراوان حجم محاسبات به دلیل استفاده از اصول برنامه‌نویسی پویا.

[8] Mixture-Gaussian Observation Densities
[9] Mixture Splitting
[10] Viterbi Algorithm

ـ یافتن همزمان احتمال دنبالهٔ مشاهدات به شرط وجود مدل($P(O|\lambda)$) که در بازشناسی گفتار کاربرددارد.

ـ عدم استفاده از جمع (برخلاف آلگوریتم جلورونده[11]) و درمقابل استفاده از بیشینه‌سازی که باعث‌ساده‌شدن پیاده‌سازی آن بصورت لگاریتمی می‌شود.

باتوجه به موارد فوق، آلگوریتم ویتربی یکی از روش‌های مناسب و مطلوب جهت انجام بازشناسی‌گفتار تلقی شده و لذا در این بخش از طرح پژوهشی جاری نیز مورد استفاده قرار گرفته است.

۱ـ ۲ـ ۲ـ ۲ شبکه‌های عصبی با تأخیر زمانی[12]

یکی دیگر از روش‌های مطرح در بازشناسی گفتار، استفاده از شبکه‌های عصبی با تأخیر زمانی(TDNN) می‌باشد. این نوع شبکه به عنوان گونه‌ای تغییریافته از[13] MLP جهت کاربرد در بازشناسی‌گفتار مطرح گردیده است. از آنجائی که شبکه‌های عصبی توانائی بالائی در طبقه‌بندی الگوهای استاتیک‌دارند، جهت اعمال آنها در این زمینه، نیاز به افزودن ویژگی توانائی برخورد با دینامیک الگوهای گفتاری‌به آنها می‌باشد. در TDNN ، اطلاعات چند فریم گفتاری، به همین منظور، بطور همزمان به سیستم‌اعمال می‌شود. در عمل ملاحظه گردیده است که توانائی نسبتاً قابل قبولی را می‌توان در این زمینه ازاین‌گونه شبکه‌ها انتظار داشت [۸].

۱ـ ۲ـ ۲ـ ۲ـ ۱ دادگان به‌کار رفته

برای این مرحله از پیاده‌سازی، از دادگانی که با استفاده از دوازده کلمهٔ کنترلی اشاره‌شدهٔ قبلی تشکیل‌گردیده بود استفاده گردید. این دادگان با استفاده از گفتار ۷۰ گوینده که به تساوی از مردان و زنان تشکیل‌شده بودند و با سه بار تکرار هر کلمه تشکیل گردیده بود. علاوه بر این یک بخش وابسته به گوینده نیز دراین دادگان پیش‌بینی شده بود تا بتوان توانائی سیستم بازشناسی را روی اینگونه کاربرد نیز آزمود. [۱]

۱ـ ۲ـ ۲ـ ۲ـ ۲ پیاده‌سازی

در مرحلهٔ پیاده‌سازی این سیستم نیز همانند سیستم مبنی برCDHMM ، نیاز به مراحل استخراج‌ویژگی‌های گفتاری، آموزش سیستم بازشناسی و بالاخره شبیه‌سازی و بازشناسی می‌باشد. موارد فوق برروی یک شبکهٔ عصبی از نوعTDNN پیاده‌سازی گردیده و شبکه با دو مجموعهٔ دادگان،

[11] Forward Algorithm
[12] Time Delay Neural Networks
[13] Multi-Layer Perceptron

یکی دادگان‌مبتنی بر مجموعهٔ ارقام اشاره شده در بخش ۱ـ ۲ـ ۱ و دیگری مجموعهٔ اشاره شده در بخش۱ـ ۲ـ ۲ـ ۱ آزمایش گردید [۲]. اگرچه نتایج بدست آمده از این مرحله نسبتاً مناسب می‌باشند، ولی‌کیفیت بالاتر سیستمCDHMM در بازشناسی گفتار گسسته کاملاً آشکار بوده است. بررسی دقیق‌ترنشان می‌دهد که شبکهTDNN ، علیرغم تأخیر زمانی ایجاد شده، قادر به دخالت دادن مناسب ومطلوب دینامیک گفتار در طبقه‌بندی نمی‌باشد. یکی از روش‌های پیشنهاد شده استفاده از نوعی پیچش‌زمانی به عنوان پردازش اوّلیه جهت انطباق مناسب زمانی سیگنال گفتار با الگوهای استفاده شده برای‌آموزش شبکه می‌باشد. از آنجا که روش‌هائی نظیر پیچش زمانی پویا(DTW[14]) خود به عنوان‌آلگوریتم‌های بازشناسی گفتار مستقیماً مورد استفاده می‌باشند، تصمیم گرفته شد تحقیق بیشتر در این‌زمینه متوقف شده و پیاده‌سازی تنها براساسCDHMMs ادامه یابد.

۱ـ ۲ـ ۳ مشکلات در بازشناسی و اصلاحات مربوطه

سیستم بازشناسی ایجاد شده براساس CDHMMs ، علیرغم توانائی‌های خوبی که از خود نشان‌می‌دهد، از دو مشکل اساسی رنج می‌برد: محدودیت سرعت و کاهش کیفیت در شرایط مختلف‌محیطی (نظیر وجود نویز زمینه). لذا دنبالهٔ کار در این بخش بر روی رفع این دو مشکل و یافتن‌راه‌حل‌هائی برای آنها متمرکز گردید.

۱ـ ۲ـ ۳ـ ۱ سرعت بازشناسی

مسئلهٔ سرعت در بازشناسی در مراحل مختلف کار قابل طرح می‌باشد. پیاده‌سازی بهینهٔ آلگوریتم‌هاموجب می‌گردد که با سرعت بالاتری به انجام برسند. چه در مرحلهٔ آموزش و چه در مرحلهٔ بازشناسی،این فاکتور دارای اهمیت زیادی می‌باشد. داشتن سرعت بالا در آموزش، درعین حال دارای اهمیت سرعت در بازشناسی نمی‌باشد چرا که آموزش عموماً در شرایطoff-line صورت می‌گیرد درحالیکه‌در کاربردهای واقعی، بازشناسی اغلبon-line می‌باشد. به همین جهت اهمیت سرعت در بازشناسی‌بیش از آموزش می‌باشد. البته باید توجه نمود که بهر صورت کندبودن آموزش خود باعث تأخیر درپیاده‌سازی موارد آزمایشی سیستم می‌گردد.

علاوه بر موارد فوق باید توجه نمود که چه در آموزش و چه در بازشناسی، از آنجا که نیاز به استخراج‌ویژگی‌های گفتاری می‌باشد، سرعت استخراج ویژگی‌ها دارای اهمیت می‌باشد. با این همه، در این‌خصوص نمی‌توان بهبودی چندانی، جز از طریق بهینه‌سازی برنامه‌نویسی و یا برخی تغییرات جزئی‌بدست آورد[۱]. درخصوص آلگوریتم آموزش نیز باتوجه به اینکه در این طرح پژوهشی،

[14] Dynamic Time Warping

آموزش‌بصورتoff-line صورت می‌گرفته است، افزایش سرعت اجرای الگوریتم چندان مورد توجه قرارنگرفته است.

آلگوریتم بازشناسی، باتوجه به چند مورد، نیاز به افزایش سرعت دارد. اوّل اینکه این آلگوریتم‌می‌تواند به صورتon-line مورد استفاده واقع شود. علاوه بر این، باتوجه به استفاده ازHMM های باچگالی پیوسته، عمدهٔ زمان صرف محاسبهٔ تابع چگالی احتمال مخلوط پیوسته (گوسین‌ها) می‌گردد که‌شامل محاسبهٔ مقادیر نمائی و ضرب‌های برداری می‌باشد. باتوجه به زمان‌گیربودن این محاسبات واینکه باید به‌ازاء هر بردار ورودی این کار تکرار گردد، زمان صرف‌شدهٔ کلّی به‌ازاء هر دنبالهٔ بردارهای‌ورودی (بیان گفتاری) قابل توجه می‌باشد. از آنجا که در بازشناسی با دایرهٔ کلمات ۱۰۰ کلمه، تعدادمُدل‌هائی که باید مورد بررسی قرار گیرند، زیاد است (۱۰۰ عدد)، این محاسبات باید ۱۰۰ بار تکرارگردد تا مُدل مناسب بتواند بررسی و یافت شود. بنابراین زمان زیادی باید صرف شود تا بازشناسی‌بتواند صورت گیرد.

شیوه‌های چندی برای بهبود سرعت در آلگوریتم بازشناسی به‌کار گرفته شده‌اند. آنچه در این طرح‌پژوهشی برای این منظور مورد توجه قرار گرفته از دو جنبه قابل طرح است. جنبهٔ اوّل بهبود پیاده‌سازی‌آلگوریتم جهت دستیابی به سرعت‌های بازشناسی قابل قبول می‌باشد. دوّمین جنبه از کار، توجه به‌روش‌های جستجوی بهبودیافته جهت کاهش فضای جستجو در هنگام کار با تعداد کلمات بالا می‌باشد.

در پیاده‌سازی آلگوریتم، توجه به ماتریس انتقال می‌تواند از جهت افزایش سرعت راهگشا باشد. توجه به غیرارگودیک[۱۵] بودنHMM و نیز محدودنمودن اختصاص فریم‌ها به حالت‌ها از نقطه‌نظر زمانی، باتوجه به محدودیت‌های طبیعی در گفتار، می‌توانند باعث افزایش نسبی سرعت اجرای الگوریتم شوند [۱][۲].

از جنبهٔ کاهش فضای جستجو، می‌توان بجای اقدام به اعمال تک‌تک مُدل‌ها (مثلاً ۱۰۰ مدل) به دادهٔورودی جهت محاسبهٔ درستنمائی مربوطه و سپس انتخاب بالاترین درستنمائی جهت بدست‌آوردن‌کاندیدای مناسب، به شیوهٔ دیگری عمل نمود [٤]. در شیوهٔ اخیر، اقدام به اعمال همزمان آلگوریتم‌ویتربی جهت همردیف‌سازی و بدست آوردن درستنمائی نهائی بین گفتار ورودی و تمامی ۱۰۰ مدل‌گفتاری موجود می‌گردد. این شیوه، اگرچه در ظاهر امر مانند شیوهٔ قبلی است، با این تفاوت که تمام‌محاسبات همزمان و بطور موازی صورت می‌گیرند، امّا دارای این ویژگی مطلوب است که می‌توان بااستفاده از تکنیک جستجوی شعاعی[۱٦]، در مقاطعی از عمل، فضای جستجو را با استفاده از یک شعاع‌جستجوی محدود که ازطریق اعمال یک سطح آستانهٔ نسبی (نسبت به

[15] Non-Ergodic
[16] Beam Search

درستنمائی بیشینه در هر مقطع)ایجاد می‌شود، کاهش داد. اینگونه جستجوها در موارد مختلف و با درجات موفقیت قابل قبولی درکاربردهای مختلف و بویژه کاربردهای بازشناسی گفتار مورد استفاده قرار گرفته‌اند [۹][۱۰]. به این‌ترتیب و با اعمال سطح آستانهٔ مناسب می‌توان ضمن داشتن نرخ خطای قابل قبول در بازشناسی،سرعت اجرای آلگوریتم بازشناسی را نیز افزایش داد.

۱ـ ۲ـ ۳ـ ۲ مقابله با اثرات شرایط مختلف محیطی

شرایط محیطی مختلف می‌توانند به شدت کیفیت یک سیستم بازشناسی گفتار را تحت تأثیر قراردهند. این شرایط شامل انواع نویزهای جمع‌شونده با سیگنال[۱۷] ناشی از منابع مختلف آلایندهٔ فضای‌صوتی می‌باشند. ازجملهٔ این موارد می‌توان به انواع نویز ناشی از کار دستگاههای مختلف برقی،موتورهای احتراقی و نظایر آنها، صدای سایر افراد که به عنوان نویز زمینه سیگنال گفتار را تحت تأثیر قرارمی‌دهد و یا هرگونه صدای دیگری نظیر صدای باد، ضربات و اصطکاک، جریان آب و نظائر آنها باشد. این منابع می‌توانند ازنظر طیف فرکانسی، انرژی سیگنال جمع‌شونده، و پایداری زمانی، کاملاًویژگی‌های متفاوتی داشته باشند.

منابع دیگری نیز که می‌توانند بر سیگنال گفتار موردنظر تأثیر بگذارند، شامل منابع نویز ضرب‌شونده[18] یا Convolutional است که عموماً مشخصات فرکانسی سیگنال گفتاری را تحت تأثیر قرار می‌دهند. این‌منابع عموماً شامل کانال‌های انتقال می‌باشند نظیر خط تلفن، میکروفون، ضبط گفتار، کانال رادیوئی وامثال آنها.

منابع نویز، تأثیر آنها بر کیفیت بازشناسی و روش‌های معمول در مقابله با آنها در [۲] مورد بررسی قرارگرفتند. یکی از روش‌هائی که درمقابله با نویز کارآئی خوبی از خود نشان داده است "معیار تصویروزن‌دهی شده[۱۹] " (WPM) نام دارد. در این روش، باتوجه به اثر نویز سفید بر روی بردارهای کپسترال یک‌معیار تصویر معرفی می‌شود که به کمک آن می‌توان برای هر بردار درستنمائی را به شکل جدیدی در هرحالت ازHMM محاسبه نمود [۱]. در بررسی‌های به‌عمل آمده مزایای چندی به شرح زیر برای روشWPM بدست آمد:

الف ـ این روش علاوه بر نویز سفید، برای نویزهای رنگی نیز تاحد زیادی مفید می‌باشد.

ب ـ دقّت بازشناسی در شرایط نویزی با استفاده از این روش افزایش قابل ملاحظه‌ای می‌یابد.

پ ـ آموزش مدل‌ها برای اعمال این روش دستخوش تغییر نشده وWPM تنها در مرحلهٔ بازشناسی‌قابل اعمال است.

[17] Additive Noise
[18] Multiplicative
[19] Weighted Projection Measure