تصدیق مستقل از متن هویت گوینده با مدل آمیزه‌های گوسی تطبیق‌یافته‌ی ساختاری

رحیم سعیدی و حمیدرضا صادق‌محمدی
پژوهشکده‌ی برق جهاد دانشگاهی
نویسنده‌ی عهده‌دار مکاتبات: رحیم سعیدی

چکیده:

در این مقاله، روش مدل آمیزه‌ی گوسی ساختار یافته[1] (SGMM) که به‌منظور سرعت بخشی الگوریتم[2] GMM-UBM در سیستم تصدیق هویت گوینده پیشنهاد شده است، مورد بررسی قرار می گیرد. تأثیر برخی پارامترها در ساخت مدل پس‌زمینه‌ی ساختاری[3] (SBM) به‌تفصیل مورد بررسی و مقادیر بهینه در ساخت مدل مورد استفاده قرار می گیرد. همچنین برای پردازش امتیازهای خروجی ساختار SBM-SGMM ساختاری با عنوان شناسایی کننده‌ی GMM پیشنهاد می‌شود. شبیه‌سازی‌های انجام یافته نشان می‌دهد، ساختار بهینه‌ی SBM در ترکیب با پردازنده‌ی امتیازهای پیشنهاد شده، عملکرد بهتری نسبت به سیستم پایه در ترکیب با شبکه‌ی عصبی دارد و این در حالتی است که پیچیدگی سیستم پیشنهاد شده، پایین‌تر از شبکه‌ی عصبی می باشد. با استفاده از سیستم پیشنهادی، نرخ سرعت بخشی برابر با ۲/۷ حاصل گردید و در عین حال عملکرد سیستم نیز نسبت به سیستم GMM-UBM بهبود نشان می‌دهد. در بهترین حالت می‌توان به نرخ خطای برابر معادل ۰/۳۵٪ دست یافت که نسبت به عملکرد سیستم GMM-UBM با نرخ خطای برابر ۱/۷۱٪ بهبود قابل ملاحظه‌ای حاصل می شود.

واژه‌های کلیدی: تصدیق هویت مستقل از متن گوینده، GMM، UBM، SBM، SGMM.

۱- مقدمه

سیگنال‌ها و معیارهای مختلفی برای سیستم‌های شناسایی افراد از روی مشخصه‌های حیاتی آن‌ها تا به حال مورد مطالعه قرار گرفته‌اند. از این میان معروف‌ترین مشخصه‌ها، اثر انگشت، چهره و صدای افراد می‌باشد. با وجود این که هر کدام از این مشخصه‌ها مزایا و معایب خاص خود را با توجّه به دقّت و کاربرد دارد، عواملی وجود دارد که صدای افراد در هنگام صحبت را از این مشخصه‌ها متمایز می‌نماید. اوّل این‌که گفتار یک سیگنال طبیعی می‌باشد و تولید یک تکه گفتار مربوط به یک فرد برای افراد دیگر ممکن نیست. در بسیاری از کاربردها گفتار عمده‌ترین و یا تنها راه ممکن برای دسترسی به یک فرد می‌باشد (مانند ارتباطات تلفنی). دوم اینکه شبکه‌ی تلفن یک شبکه‌ی رایج و مطمئن برای به‌دست آوردن و تحویل سیگنال گفتار به‌شمار می‌رود و در کاربردهای مبتنی بر تلفن، هیچ نیازی به مبدل‌های خاص سیگنال یا نصب یک شبکه‌ی ویژه در پایانه‌های دسترسی نمی‌باشد. علاوه بر این، به‌وسیله یک تلفن همراه، تقریباً همه جا می‌توان به نمونه گفتار فرد دسترسی پیدا کرد. حتی در کاربردهایی غیر از کاربردهای مبتنی بر تلفن نیز، کارت‌های صوتی و میکروفن‌ها ابزارهایی هستند که هم قیمت آن‌ها پایین است و هم به‌راحتی در دسترس قرار دارند.

سیگنال گفتار حامل مراتب مختلفی از اطلاعات است. ابتدا سیگنال گفتار حامل کلمات یا پیغام ادا شده است، امّا

[1] Structural Gaussian Mixture Models
[2] Gaussian Mixture Model-Universal Background Model
[3] Structural Background Model

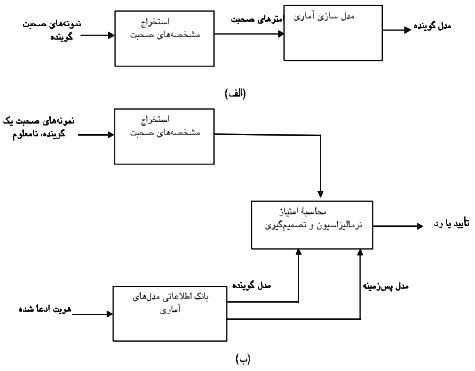
در مرتبه‌ی بعدی سیگنال گفتار حامل اطلاعاتی راجع به‌هویت گوینده می‌باشد. تشخیص گوینده به دو مقوله‌ی تعیین هویت گوینده و تصدیق هویت گوینده تقسیم می‌شود. در تصدیق هویت گوینده در مورد تأیید ادعای گوینده مبنی بر داشتن یک هویت، تصمیم‌گیری می‌شود. ولی در تعیین هویت گوینده، هویت از میان مجموعه‌ای از گوینده‌ها تعیین می‌گردد. در تصدیق هویت گوینده ابتدا فرد، مدعی یک هویت می‌شود این امر به‌عنوان نمونه به‌وسیله‌ی وارد کردن یک شماره‌ی کارمندی یا وارد کردن یک کارت هوشمند صورت می‌پذیرد. سپس سیستم نمونه‌های گفتار فرد را گرفته و پس از پردازش آنها تصمیمی را مبنی بر قبول یا رد ادعای فرد اتخاذ کرده و یا اعلام می‌نماید که اطلاعات ورودی جهت تصمیم‌گیری کافی نمی‌باشد. تشخیص گوینده می‌تواند به‌صورت وابسته به متن یا مستقل از متن باشد. در سیستم‌های وابسته به متن، عبارت مورد نظر برای سیستم شناخته شده است و می‌تواند به‌صورت یک جمله‌ی ثابت باشد و یا توسط سیستم به گوینده اعلان شود. اغلب، این عبارت به‌صورت رشته‌ی متوالی از اعداد می‌باشد که به‌عنوان رمز عبور شناخته می‌شود. ولی در سیستم‌های مستقل از متن، محدودیتی روی محتوای متنی گفتار ادا شده وجود ندارد. از سیستم‌های تصدیق هویت گوینده می‌توان در کنترل دسترسی به مکان‌های امنیتی، کنترل دسترسی به کامپیوترها، معاملات تلفنی که با کارت اعتباری انجام می‌شوند و همچنین در معاملات انجام شده از طریق شبکه‌ی اینترنت استفاده کرد.

هر سیستم تصدیق هویت گوینده، دارای دو مرحله‌ی مجزا از هم می باشد، مرحله‌ی آموزش و فاز آزمون. هر کدام از این فرآیندها را می‌توان به‌عنوان یک بخش مجزا تصور کرد. در شکل(۱الف) نمودار کلی جعبه‌ای فرآیند آموزش و در شکل۱(ب) نمودار کلی جعبه‌ای فرآیند آزمون نشان داده شده است. اولین قدم در فرآیند آموزش سیستم، همان‌گونه که در شکل(۱الف) مشاهده می‌شود استخراج پارامترهای مناسبی از سیگنال گفتار است که قابل استفاده برای مرحله‌ی مدل‌سازی آماری باشد . همان روالی که برای آموزش مدل یک گوینده به‌کار رفته است برای آموزش مدل پس‌زمینه نیز طی می‌شود. آن‌گونه که در شکل۱(ب) نشان داده شده است ورودی‌های سیستم در مرحله‌ی آزمون شامل یک هویت ادعا شده و نمونه‌های گفتار یک فرد ناشناس می‌باشد، که مدعی آن هویت به‌خصوص شده است . هدف سیستم در مرحله‌ی آزمون این است که تأیید کند آیا نمونه‌های گفتار رسیده به سیستم، متعلق به همان هویت ادعا شده است یا نه . برای انجام این کار ابتدا پارامترهای سیگنال گفتار، همانند روال مورد استفاده در مرحله‌ی آموزش، استخراج می‌شوند . سپس مدل گوینده‌ی مورد ادعا و نیز یک مدل پس‌زمینه که هر دو در مرحله‌ی آموزش محاسبه شده‌اند از بانک اطلاعاتی سیستم استخراج می‌شود. درنهایت سیستم با استفاده از پارامترهای گفتار استخراج شده‌ی فرد مدعّی و دو مدل آماری، امتیازهایی را محاسبه کرده ، آنها را هنجارسازی نموده و تصمیمی مبنی بر قبول یا رد فرد مدعی اتخاذ می کند.

در زمینه‌ی تشخیص گوینده در زبان فارسی کارهایی انجام شده است که به‌اختصار به آن می‌پردازیم. در سال ۱۳۷۳، آقایان مندولکانی و لطفی‌زاد [۱]، با استفاده از روش DTW[1] برای جمعیت ۱۰ نفری و به‌ازای ۱۰ جمله برای آموزش و ۱۰ جمله آزمون به کارآیی ۹۸٪ برای تعیین هویت گوینده وابسته به متن، دست یافته‌اند. بازهم در سال ۱۳۷۳ آقایان حدائق و لطفی‌زاد [۲]، با استفاده از روش DTW و بر روی جمعیت ۱۰ نفری و به‌ازای ۱۰ تکرار جمله‌ی خاص برای آموزش و ۱۰ تکرار همان جمله برای آزمون به کارآیی ۱۰۰٪ برای تصدیق هویت گوینده وابسته به متن، دست یافته‌اند. در سال ۱۳۷۴، آقایان صیادیان و غفوری‌فرد [۳]، با استفاده از کوانتیزاسیون برداری و بر روی جمعیت ۵۰ نفری گویندگان، به‌ازای ۱۰ جمله برای آموزش و یک جمله برای آزمون به کارآیی متوسط ۹۸/۰۳٪ برای تعیین هویت گوینده رسیده‌اند. در همان سال، آقایان مقصودلو، نخعی و تبیانی [۴]، با استفاده از کوانتیزاسیون برداری و بر روی جمعیت ۱۰ نفری مردان، به ازای ۸ کد پنج رقمی برای آموزش و کدهای سه رقمی برای آزمون، جهت تصدیق هویت گوینده به کارآیی ۹۹/۸۳٪ رسیده‌اند. در بهمن ماه همین سال آقای شیخ زادگان [۵] تعیین هویت گوینده، بصورت مستقل از متن را به طور جدّی مورد بررسی قرار دادند. در سال ۱۳۷۷، آقایان فیض‌آبادی و صدوقی [۶]، با استفاده از کوانتیزاسیون برداری و بر روی جمعیت ۳۰ نفری گویندگان و به ازای ۲۰ جمله و ۲۰ رقم برای آموزش و یک جمله برای آزمون به کارآیی ۱۰۰٪ برای تصدیق هویت گوینده رسیده‌اند. در سال ۱۳۷۸، آقایان نجاری و همایون‌پور [۷]، با استفاده از دو روش شبکه‌های عصبی و الگوریتم‌های ژنتیک و کوانتیزاسیون برداری بر روی جمعیت ۵۸ نفری (۳۶ مرد و ۲۲ زن)

[1] Dynamic Time Warping

گویندگان و به‌ازای ۵۰ رقم برای آموزش و ۷ رقم برای آزمایش در محیط تلفنی به کارآیی ۹۷/۸٪ رسیده اند.



شکل (۱): نمودار جعبه ای(الف) فاز آموزش و(ب) آزمون یک سیستم تصدیق هویت گوینده

در دی‌ماه همان سال آقای نوری‌وند [۸] عملکرد شبکه‌های عصبی را بر روی گفتار تلفنی به‌منظور بازشناسی گوینده، مورد بررسی قرار داده‌اند. در سال ۱۳۷۹، آقایان صیادیان، بدیع، حکاک و بیک‌زاده [۹]، با استفاده از مدل‌آمیزه‌های گوسی در سطح واج و یک مدل به‌ازای هر واج برای هر گوینده، بر روی جمعیت ۶۰ نفری (۴۰ مرد و ۲۰ زن) و به ازای ۱۰۰۰ جمله در دوره‌ی آموزش ـ که به‌صورت دستی واج‌نگاری می‌شود ـ و به‌ازای ۳ ثانیه گویش در دوره‌ی آزمون، به کارآیی ۱۰۰٪ برای تعیین هویت گوینده رسیده‌اند. در سال ۱۳۸۲، آقایان معین و بوستانی [۱۰]، با استفاده از روش‌های ماشین بردار مرزی، SVM[1]، مدل آمیزه‌ی گوسی GMM، و مدل مارکف نهانی، HMM[2] بر روی جمعیت ۴۰ نفری گویندگان و رشته‌های صفر تا ۹ برای آموزش و آزمون به نرخ خطای برابر ۲٪، ۵٪ و ۸٪ به ترتیب روش‌های ذکر شده دست یافته‌اند. در سال ۱۳۸۳، آقایان همایون‌پور و کبودیان [۱۱]، با استفاده از ترکیب روش‌های GMM و HMM بر روی جمعیت ۱۰۰ نفری گویندگان (۶۱ مرد و ۳۹ زن) و بر روی پایگاه داده‌ی تلفنی FARSDIGITS1 به‌ازای داده‌های آزمایشی، به کارآیی ۹۵/۵۱٪ در تعیین هویت گوینده و نرخ خطایی برابر ۰/۳۳٪ در تصدیق هویت گوینده دست یافته‌اند.

در راستای آزمون الگوریتم‌های ارایه شده در این مقاله، ابتدا یک پایگاه اطلاعاتی گفتار متشکل از ۱۹۰ گوینده تهیه شده است که محدودیتی روی محتوای گفتار وجود نداشته است. هدف اصلی در تحقیق فعلی، شبیه‌سازی و بررسی عملکرد دو نوع سیستم تصدیق مستقل از متن هویت گوینده است که در طی چند سال اخیر توسعه یافته‌اند. این دو روش عبارتند از: مدل آمیزه‌ی گوسی تطبیق‌یافته[3] و مدل آمیزه‌ی گوسی ساختار یافته در ترکیب با شبکه‌ی عصبی. پایه و اساس این روش‌ها کار ارایه شده توسط رینولدز[4] است[۱۲]. وی در تحقیق خود استفاده از مدل آمیزه‌ی گوسی را به‌منظور مدل‌سازی مدل آماری گویندگان در فضای مشخصه‌ها در شناسایی مستقل از متن گوینده، را تشریح نمود. ما نیز در این مقاله بنای اصلی کار را بر اساس پروژه‌ی گزارش شده در مرجع [۱۲] قرار داده و کارهای انجام شده در آن را با دید یک سیستم تصدیق هویت گوینده شبیه‌سازی نمودیم. معرفی مدل آمیزه‌ی گوسی توسط رینولدز تحول شگرفی در عملکرد سیستم‌های تصدیق هویت گوینده به‌وجود آورد که باعث شد روند اصلی تحقیقات در این زمینه بدین سو متمایل شود. تحقیقات انجام شده روی مدل آمیزه‌ی گوسی باعث بهبود عملکرد سیستم‌های استفاده کننده از این مدل گردید که این بهبود هم از نظر نقطه‌ی‌کار و هم از نظر بار محاسباتی عملیات حاصل شد؛ تا این‌که بار دیگر رینولدز و همکاران، تغییری بنیادین را در مدل آمیزه‌ی گوسی مطرح کردند که باعث بهبود بیشتر کارآیی این مدل گردید[۱۳]. این تغییر که هم مدل‌سازی مدل پس‌زمینه و هم در مدل‌سازی مدل گویندگان انجام شده بود و به نام سیستم GMM-UBM مشهور است، از آن زمان تاکنون جزء تفکیک ناپذیر بیشتر سیستم‌های تشخیص گوینده به‌شمار می‌رود.

کارهای مختلفی از سال ۲۰۰۰ به بعد در جهت بهبود عملکرد سیستم‌های مبتنی بر GMM-UBM انجام شده است. در سال ۲۰۰۱، آکنتالر[5] [۱۴]، تشریح کرد که ۹۰٪ وقت پردازش هر سیستم مبتنی بر GMM UBM صرف محاسبات امتیازدهی آمیزه‌های گوسی می‌شود. کار انجام شده در پروژه‌ی دکتری بینگ شیانگ[6] که در سال ۲۰۰۳ میلادی در دانشگاه کرنل ایالات متحده امریکا انجام شده بود، به علت نتایج خوب و پس‌زمینه‌ی ریاضی قوی برای تسریع سیستم GMM UBM انتخاب گردید[۱۵]. در این

[1] Support Vector Machine
[2] Hidden Markov Model

[3] Adapted Gaussian Mixture Models
[4] Reynolds
[5] Auckenthaler
[6] Bing Xiang

سیستم ابتدا یک ساختار درختی بر مبنای مدل UBM بنا می‌شود که SBM[1] نام می‌گیرد و سپس یک مدل ساختاری برای گویندگان از روی SBM تطبیق داده می‌شود که SGMM نامیده می‌شود. هدف از ساخت مدل درختی، کاهش بار محاسباتی سیستم تأیید هویت گوینده است که به قیمت از دست دادن مقدار کمی از دقّت تشخیص سیستم تمام می‌شود. برای جبران این کاهش دقّت، از یک شبکه‌ی عصبی برای ترکیب امتیازهای حاصله از لایه‌های مختلف ساختار درختی استفاده شده است. این سیستم ترکیبی، هم دارای دقت بیشتری نسبت به سیستم پایه‌های GMM-UBM و هم دارای بار محاسباتی کمتری نسبت به آن می‌باشد. در این مقاله، سیستم SBM-SGMM به‌همراه یک شبکه‌ی عصبی MLP به‌عنوان پردازش‌گر پسین شبیه‌سازی شده و با تغییر پارامترهای سیستم، عملکرد سیستم ترکیبی مورد بررسی قرار می‌گیرد. همچنین یک پردازش‌گر پسین جدید که نام آن را شناسایی کننده‌ی GMM گذاشته‌ایم، پیشنهاد می‌شود. این شناسایی‌کننده‌ی پیشنهادی از توزیع گوسی امتیازهای خروجی سیستم SBM-SGMM استفاده می‌کند. عملکرد شناسایی کننده‌ی GMM پیشنهادی توسط شبیه‌سازی کامپیوتری با عملکرد شبکه‌ی عصبی مقایسه می شود.

۲- مدل‌های GMM

در این بخش، مدل‌های GMM که با هدف بهبود کارآیی سیستم تصدیق هویت گوینده و کاهش بار محاسباتی سیستم، توسعه یافته‌اند، معرفی می‌شود. ابتدا مدل GMM-UBM که کارآیی آن در سال ۲۰۰۰ به اثبات رسید، شرح داده می شود [۱۲] و امتیازدهی سریع که از مزایای GMM UBM می‌باشد بحث می‌گردد. در ادامه، مدل SBM-SGMM که به‌منظور کاهش بار محاسباتی سیستم تصدیق هویت گوینده در سال ۲۰۰۳ معرفی شده است، تشریح می‌شود [۱۵]؛ سپس طریقه‌ی ساختن مدل SBM مورد بررسی قرار می‌گیرد. متناظر با کاهش بار محاسباتی سیستم، عملکرد سیستم تنزل می‌یابد که به‌خاطر جبران این تنزل عملکرد، از یک شبکه‌ی عصبی استفاده شده است. با استفاده از فرض توزیع گوسی امتیازها که در هنجارسازی امتیازها استفاده می‌شود، سیستمی با عنوان شناسایی کننده‌ی GMM پیشنهاد شده است که به‌جای شبکه‌ی عصبی قرار می گیرد.

۲-۱- مدل آمیزه‌های گوسی تطبیق یافته

یکی از راه‌کارهای به‌دست‌آوردن مدل پس‌زمینه، استفاده از یک مدل پس‌زمینه‌ی مستقل از گوینده، تحت عنوان UBM می‌باشد. در حقیقت UBM یک GMM است که برای بیان توزیع مستقل از گوینده‌ی مشخصه‌ها به‌کار می‌رود. به‌طور خاص در ساخت UBM به‌دنبال انتخاب سیگنال‌های گفتاری هستیم که مشخصات سیگنال‌های گفتاری را که انتظار می‌رود در حین تشخیص گوینده با آنها برخورد شود، در خود داشته باشند. این مشخصات می‌تواند در قالب نوع و کیفیت گفتار و نیز ساختار مجموعه‌ی گویندگان باشد. در سیستم GMM UBM مدل گویندگان بوسیله‌ی تطبیق پارامترهای مدل UBM با استفاده از گفتارهای آموزشی گویندگان و شکلی از قاعدهٔ آموزشی بیز یا همان تخمین MAP ساخته می‌شود. برعکس راه‌کار قبلی، آموزش ML که در آن، مدل هر گوینده، مستقل از UBM ساخته می‌شد [۱۶]، ایده‌ی اصلی در اینجا تطبیق مدل گویندگان نسبت به‌مدل UBM به‌وسیله‌ی به‌روزرسانی پارامترهای UBM می‌باشد. با انجام این کار ارتباط متقابل بین مدل گویندگان و مدل UBM برقرار می‌شود که نه تنها باعث بهبود عملکرد نسبت به مدل‌های آموزش دیده‌ی مستقل از هم می‌شود، بلکه همان‌گونه که در ادامه ذکر خواهد شد، این اجازه را می‌دهد تا از یک روش امتیازدهی سریع نیز استفاده کرد. ایدهٔ اصلی تخمین MAP برای GMM توسط گاوین[2] و لی[3] [۱۷] ارایه شد و سپس توسط لی، چن[4] و هو[5] [۱۸] تکمیل گردیده و بسط داده شد. معادلات مربوط به آموزش مدل UBM و چگونگی تطبیق مدل GMM از UBM به‌تفصیل در مراجع[۱۳و۱۶] بحث شده‌اند.

۲-۲- محاسبه‌ی نسبت درست‌نمایی لگاریتمی در سیستم GMM-UBM

نسبت درست‌نمایی لگاریتمی، LLR[6]، برای یک دنباله‌ی آزمون از بردارهای مشخصه‌ی X، به‌صورت درست‌نمایی

[1] Structural Background Model

[2] Gauvain
[3] Huo
[4] Chan
[5] Lee
[6] Log Likelihood Ratio

$\Lambda(x) = Log\left[p\left(x\middle|\lambda_{hyp}\right)\right] - Log\left[p\left(x\middle|\lambda_{\overline{hyp}}\right)\right]$ محاسبه می‌گردد که از این نسبت به‌عنوان امتیاز گوینده در سیستم تصدیق هویت گوینده استفاده می‌شود. در این رابطه λ_{hyp} مدل گوینده‌ی ادعا شده و $\lambda_{\overline{hyp}}$ مدل پس‌زمینه است. از این موضوع که مدل گویندگان، به‌وسیله‌ی تطبیق مدل UBM به‌دست آمده است، می‌توان روش امتیازدهی سریعی را استفاده کرد که دیگر نیاز به مقداردهی کامل هر دو GMM نباشد [۱۳]. این راه‌کار بر پایه‌ی دو اثر مشاهده شده در آزمایش‌ها به شرح زیر می‌باشد: اوّل این‌که هنگامی که یک GMM بزرگ به‌ازای یک بردار مشخصه مقداردهی می‌شود، تعداد کمی از آمیزه‌ها سهم عمده‌ای در مقدار درست‌نمایی دارند، این امر بدین علّت است که GMM نشان دهنده‌ی توزیعی در یک فضای بزرگ می‌باشد ولی یک بردار مشخصه تنها می‌تواند در نزدیکی تعداد کمی از اجزاء GMM واقع شود. بنابراین مقادیر درست‌نمایی را می‌توان با استفاده از مقداردهی تعداد C آمیزه‌ی جزء که بالاترین امتیاز را در بین بقیه‌ی اجزای GMM دارند به‌خوبی تقریب زد. دومین اثر مشاهده شده بدین قرار است که اجزای یک GMM تطبیق شده، تناظر خود را با آمیزه‌های UBM حفظ می‌کنند. به این معنی، بردارهایی که به یک آمیزه‌ی خاص در UBM نزدیک باشند، به آمیزه‌ی متناظر آن در مدل GMM گوینده نیز نزدیک خواهند بود. بدین ترتیب، راه‌کار امتیازدهی سریع را می‌توان به‌صورت زیر بیان کرد.

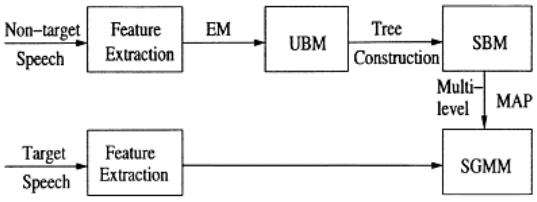
برای هر بردار مشخصه، C آمیزه‌ی جزء را که بالاترین امتیاز را در UBM آورده‌اند، پیدا کرده و مقدار درست‌نمایی مدل UBM به‌ازای بردار مشخصه‌ی داده شده را تنها با این C آمیزه محاسبه می‌کنند؛ سپس مقدار درست‌نمایی مدل GMM گوینده را نیز تنها به‌ازای C آمیزه‌ی متناظر آن در مدل گوینده به‌دست می‌آورند. در این حالت اگر UBM حاوی M آمیزه باشد، این راه‌کار امتیازدهی تنها به $M+C$ محاسبات گوسی برای تعیین LLR نیاز خواهد داشت. درصورتی‌که در راه‌کار اولیه‌ی $2M$ محاسبات گوسی مورد نیاز خواهد بود.

۲-۳- مدل آمیزه‌های گوسی ساختاری

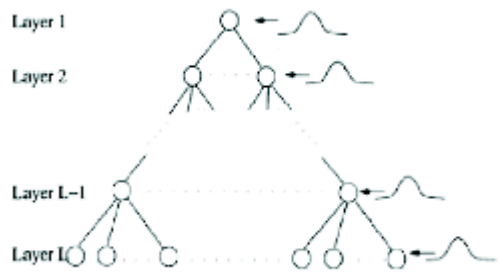
در مرجع [۱۵] روشی به نام مدل آمیزه‌های گوسی ساختاری SGMM برای کاهش بار محاسباتی پیشنهاد گردید که این شیوه در تحقیق حاضر نیز استفاده شده است. در این راه‌کار ابتدا یک مدل پس‌زمینه‌ی ساختاری، SBM ، براساس یک UBM به صورت سلسله مراتبی، خوشه‌بندی می‌شود. با این کار فضای آکوستیکی به نواحی مختلف، با مراتب تفکیک‌پذیری متفاوت، تقسیم می‌شود. هر گره در ساختار درختی SBM توسط یک آمیزه‌ی گوسی نمایش داده می‌شود. برای هر گوینده یک SGMM توسط الگوریتم تطبیق MAP چندمرحله‌ای از SBM تولید می‌شود. در هنگام آزمون می‌توان بار محاسباتی را به‌طرز مؤثری کاهش داد. این کار به‌وسیله‌ی جستجوی از بالا به پایین ساختار درختی SBM و ارزیابی تنها زیر مجموعه‌ی کوچکی از آمیزه‌های گوسی جزء در SBM و SGMM صورت می‌گیرد. در ادامه، طرز ساختن SBM به‌تفصیل مورد بررسی قرارگرفته است.

۲-۴- نحوه ساختن SBM

نمودار جعبه‌ای آموزش SBM و SGMM در شکل (۲) نشان داده شده است. همان‌طور که در شکل (۳) دیده می‌شود، بر پایه‌ی UBM می‌توان یک ساختار درختی SBM با L لایه را تولید کرد تا ساختار فضای آکوستیکی را مدل نمود. از طریق یک خوشه‌بندی سلسله مراتبی از بالا به پایین هر گره در L-1 لایه‌ی بالایی نمایش‌گر خوشه‌ای از آمیزه‌های گوسی جزء در UBM می‌باشد و خود توسط یک تابع چگالی احتمال گوسی مدل می‌شود. هر گره در لایه‌ی تحتانی مطابق با یک آمیزه‌ی گوسی جزء در UBM خواهد بود.



شکل (۲) : نمودار جعبه‌ای فرآیند آموزش در سیستم SBM – SGMM [۱۵].



شکل (۳): شمایی از یک ساختار درختی با L لایه [۱۵].

قبل از ساختن SBM، ابتدا باید یک معیار فاصله بین دو آمیزه‌ی گوسی را تعریف کرد. معیارهای فاصله‌ی مختلفی در مقالات مختلف مطرح شده است که معیارهای واگرایی KL[1] [۱۹] معیار فاصله‌ای است که در این تحقیق استفاده شده است. ساخت SBM بدین‌گونه می باشد:

۱- ابتدا ساختار درخت طراحی و تعداد لایه‌ها و تعداد شاخه‌های منشعب از هر گره در هر لایه معین می‌شود. هیچ راه‌کار مشخصی برای طراحی خودکار ساختار درختی وجود ندارد؛ زیرا ساختار بهینه ممکن است با توجه به اندازه‌ی مدل‌ها تغییر کند.

۲- گره‌ی ریشه[2] را به‌عنوان گره‌ی kام و مجموعهٔ G (که در مرحله‌ی اوّل تمام آمیزه‌های گوسی موجود در UBM می‌باشد.) را به‌عنوان G_{now} تعیین می‌کنیم. تابع چگالی احتمال گره را با استفاده از کمینه‌کردن فاصله KL (معادلات مربوط به تخمین ML گره‌ی ریشه برای کمینه‌کردن معیار فاصله در مرجع [۱۵] موجود است) برای گره‌ی ریشه محاسبه می‌کنیم.

۳- اگر گره‌ی kام دارای هیچ گره‌ی انشعابی[3] نباشد، خوشه‌بندی متوقف می‌شود؛ در غیراین‌صورت تابع چگالی احتمال اولیه‌ی هر گره‌ی انشعابی توسط روش minimax که در زیر توضیح داده می‌شود، برآورد می‌گردد. در اینجا $g^k(.)$ تابع چگالی احتمال گره برای گره‌ی kام، P_k تعداد گره‌های انشعابی از گره‌ی kام و $g^{(c_p)}(.)$ تابع چگالی احتمال گره‌ی انشعابی c_p می‌باشد که $p=1,...,P_k$.

i. از بین مجموعه‌ی G_{now} آمیزه‌ی گوسی $\hat{m}$ را به‌گونه‌ای انتخاب می‌کنیم که دارای بیشترین فاصله با $g^k(.)$ باشد. این آمیزه‌ی گوسی را به‌عنوان تابع چگالی احتمال گره‌انشعابی c_1 در نظر می‌گیریم. یعنی $g^{(c1)}(.) = g_{\hat{m}}(.)$.

ii. آمیزه‌های گوسی برای c_p از $p=2$ تا $p=P_k$ به‌طور متوالی به‌وسیله‌ی قاعده‌ی زیر انتخاب و به‌عنوان تابع چگالی احتمال گره‌ی انشعابی c_p منظور می‌شوند:

$$\hat{m} = \arg\max_{m} \min_{1\le i\le p-1} d(m, c_i)$$
$$g^{(c_p)}(.) = g_{\hat{m}}(.) \qquad (۱)$$

iii. در روابط فوق، $\hat{m}$ از بین بقیه‌ی آمیزه‌های انتخاب می‌شود که متعلق به گره‌ی kام بوده و هنوز به هیچ گره‌ی انشعابی اختصاص نیافته‌اند.

iv. تابع چگالی احتمال هر کدام از گره‌های انشعابی c_p و تابع چگالی احتمال گره‌ی kام درون‌یابی شده و تابع چگالی احتمال حاصله به‌عنوان تابع چگالی احتمال گره‌ی c_p منظور می‌شود. درون‌یابی به‌صورت زیر انجام می‌شود: (در روابط زیر $0 \le \alpha \le 1$ ضریب درون یابی است.)

$$\mu'_{c_p} = (1-\alpha)\mu_k(i) + \alpha\,\mu_{c_p}$$
$$\sigma^{2'}_{c_p}(i) = (1-\alpha)\left(\sigma^2_k(i)+\mu^2_k(i)\right) + \alpha\left(\sigma^2_{c_p}(i)+\mu^2_{c_p}(i)\right) - \mu'_{c_p} \qquad (۲)$$

۴- الگوریتم Kmeans را به‌صورت زیر تکرار می‌کنیم تا زمانی که مجموع فواصل کل هم‌گرا شود:

الف) به‌ازای هر آمیزه‌ی گوسی جزء موجود در G_{now}، فاصلهٔ آن را با هر کدام از توابع چگالی گره‌های انشعابی محاسبه کرده و هر آمیزه‌ی گوسی را به نزدیک‌ترین گره‌ی انشعابی نسبت می‌دهیم.

ب) تابع چگالی احتمال گره‌های انشعابی را توسط روش ML-KL محاسبه می‌کنیم.

ج) مجموع فواصل هر گره‌ی انشعابی را از آمیزه‌های گوسی متعلق به آن محاسبه کرده و سپس این مجموع فواصل به‌دست آمده به ازای هر گره‌ی انشعابی را نیز با هم جمع می‌کنیم تا مجموع فواصل کل به‌ازای تمام گره‌های انشعابی حاصل شود.

۵- هر کدام از گره‌های انشعابی را به‌عنوان گره‌ی kام قرار داده و مجموعه‌ی آمیزه‌های گوسی آن را به‌عنوان G_{now} در نظر می‌گیریم و به مرحله‌ی ۳ باز می‌گردیم.

[1] Kullback-Leibler Divergence
[2] Root Node
[3] Child Node

۶- مراحل ۳ تا ۵ برای لایه‌های دیگر نیز تکرار می‌شود تا خوشه‌های لایه‌ی *L-1* ایجاد شوند. هر کدام از آمیزه‌های گوسی موجود در هر کدام از خوشه‌های لایه‌ی *L-1* به عنوان یک گره‌ی انشعابی مجزا در نظر گرفته می‌شوند. این گره‌های انشعابی همان گره‌های لایه‌ی تحتانی می‌باشند.

به‌علّت اندازه‌ی متغیر خوشه‌ها، تعداد شاخه‌های گره‌هایی که در لایه‌ی *L-1* قرار دارند، از یک گره به گره دیگر تغییر می‌کند. تعداد این شاخه‌ها وابسته به راه‌کارهای مختلفی است که برای ساختن درخت اتخاذ می‌شود. هر چند تعداد کل گره‌های لایه‌ی تحتانی، همیشه برابر تعداد آمیزه‌های موجود در UBM می‌باشد.

۲-۵- شبکه‌ی عصبی

از شبکه‌ی عصبی به‌عنوان پردازش‌گر پسین در حوزه‌ی امتیازهای خروجی سیستم SBM-SGMM استفاده شده است[۱۵]. می‌توان به‌سادگی امتیازهای چندگانه‌ی حاصل از SBM , SGMM را به‌وسیله‌ی ترکیب خطی با هم ترکیب کرده و یا بیشینه‌ی مقادیر چندگانه‌ی LLR حاصله از لایه‌های متفاوت را انتخاب کرد. ولی همان‌گونه که در مرجع [۱۵] نشان داده شده است، با این تدابیر نمی‌توان بهبود عملکرد روشنی نسبت بـــه حالتی که فقط تنها از امتیازهای لایه‌ی تحتانی استفاده می‌شود به‌دست آورد. عملکرد سیستم تصدیق هویت گوینده با امتیازهای حاصله از لایه‌های پایین‌تر بهتر از عملکرد مشابه در لایه‌های بالاتر است؛ زیرا لایه‌های پایین‌تر فضای آکوستیکی را با تفکیک‌پذیری بیشتری مدل می‌کنند. هر چند عقیده بر این است که امتیازهای حاصله از لایه‌های بالاتر نیز حاوی اطلاعات مفیدی برای تأیید هویت گوینده است، به‌خصوص درحالتی‌که داده‌های آموزشی کم باشند، لایه‌های بالاتر با تعداد کمتر آمیزه‌های گوسی می‌تواند به‌صورت کامل‌تری نسبت به لایه‌های پایین‌تر آموزش داده شوند.

برای منعکس‌کردن وابستگی‌های غیرخطی که بین امتیازهای حاصله از لایه‌های مختلف موجود دارد، مناسب است که از یک شبکه‌ی عصبی چند لایه‌ی پرسپترون MLP، برای ترکیب امتیازها استفاده شود [۲۰]. نظریه‌ی شبکه‌های عصبی نشان می‌دهد که یک MLP استاندارد با یک لایه‌ی مخفی که از توابع فعال‌ساز غیرخطی هموار استفاده می‌کند، قابلیت تحقق‌پذیری بسیاری از مرزبندی‌های تصمیم‌گیری غیرخطی اختیاری را دارد [۲۱]. با توجه به این امر از یک شبکه‌ی عصبی سه‌لایه برای ترکیب امتیازهای حاصله از لایه‌های مختلف SBM , SGMM استفاده شده است تا امتیاز نهایی در هر آزمون حاصل شود. در شکل (۴) نمودار جعبه‌ای مرحله‌ی آزمون سیستم نمایش شده است که در آن امتیازهای لایه‌های مختلف به‌عنوان ورودی‌های MLP می‌باشند.

خروجی y_j در گره‌ی jام لایه‌ی مخفی با استفاده از تابع سیگموئید به‌صورت زیر محاسبه می‌شود:

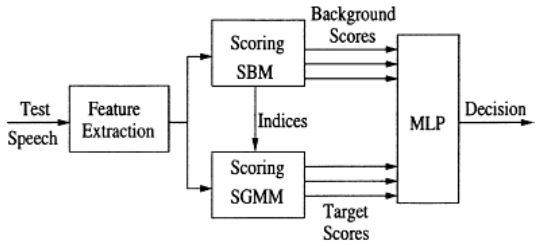
$$y_j = \frac{1}{1+e^{-\left(\vec{w}_{1,j}\vec{s}-\delta_{1,j}\right)}} \qquad (۳)$$

در رابطه‌ی بالا $d_{1,j}$ ،$\vec{w}_{1,j}$ ضرایب وزنی هستند که ورودی را به لایه‌ی مخفی متصل می‌کنند. عناصر بردار $\vec{s}$ امتیازهای مجزایی از لایه‌های مختلف SBM , SGMM است که در هر آزمون محاسبه می‌شوند و به‌عنوان ورودی MLP عمل می‌کنند. خروجی نهایی Z را می‌توان به‌صورت زیر محاسبه کرد:

$$Z = \frac{1}{1+e^{-\left(\vec{w}_2\vec{y}-\delta_2\right)}} \qquad (۴)$$

در رابطه‌ی فوق δ_2 ،$\vec{w}_2$ ضرایب وزنی هستند که لایه‌ی مخفی را به تنها گره‌ی خروجی متصل می‌کنند؛ این نوع MLP دارای *2(L-1)* گره‌ی ورودی خواهد بود که امتیازهای پس‌زمینه و امتیازهای گویندگان به‌عنوان ورودی‌های مجزا عمل خواهند کرد. بنابراین می‌توان Z را توسط تابع $f_1(.)$ به‌صورت زیر به بردار امتیازهای $\vec{s}$ مربوط کرد:

$$Z = f_1(S_{t_2}, S_{b_2}, ..., S_{t_L}, S_{b_L}) \qquad (۵)$$



شکل (۴): نمودار جعبه‌ای مرحله‌ی آزمون سیستم ترکیبی SBM-SGMM با شبکه عصبی [۱۵].

نوع دیگری از MLP نیز مورد بررسی قرار گرفته است که در آن MLP شامل *L-1* گره ورودی است که از *L-1* ورودی LLR حاصل از لایه‌های ۲ تا L مدل‌های SBM , SGMM

استفاده می‌کند. در این حالت z را می‌توان به‌صورت زیر نوشت:

$$Z = f_2(S_{t2} - S_{b2}, \ldots, S_{tL} - S_{bL}) \quad (۶)$$

دو نوع MLP دیگر نیز که عملکردشان مورد آزمایش قرار گرفته‌اند دارای دو گروه ورودی می‌باشند. این ورودی‌ها می‌توانند امتیازهای حاصل از لایه‌ی تحتانی SBM-SGMM و یا امتیازهای GMM-UBM باشند. در این حالت خروجی را می‌توان به‌صورت زیر نوشت:

$$\begin{aligned} Z &= f_3(S_{tL}, S_{bL}) \\ Z &= f_4(S_{GMM}, S_{UBM}) \end{aligned} \quad (۷)$$

تصمیم‌گیری نهایی براساس مقدار خروجی MLP انجام می‌گیرد. مرسوم‌ترین روش برای آموزش MLP؛ روش پس‌انتشار خطا، [1](BP)، است [۲۱]. با استفاده از امتیازهای به‌دست آمده در آزمون‌های با گویندگان واقعی و گویندگان وانمود کننده، خروجی مطلوب شبکه‌ی عصبی برابر مقدار ۱ برای گویندگان واقعی و به‌مقدار صفر برای گویندگان وانمود کننده، قرار داده می‌شود. با استفاده از این راه‌کار آموزش همراه با نظارت، MLP توانایی تمایز بین گویندگان واقعی و وانمود کننده را افزایش می‌دهد. در مرحله‌ی آموزش وقتی که مجموعه امتیازهای سیستم SBM-SGMM حاصل شده باشد، با قرار دادن آنها در MLP می‌توان بخشی از خطاها را تصحیح کرد (به این معنی که مقادیر LLR پایین حاصله از آزمون‌های گویندگان حقیقی را به خروجی بالاتری تبدیل کرد و یا مقادیر LLR بالای حاصل از آزمون‌های گویندگان وانمود کننده را به مقادیر خروجی پایین‌تری تبدیل کرد).

۲-۶- سیستم پیشنهادی برای تلفیق امتیازها

به‌منظور بهبود عملکرد سیستم و تمایز بیشتر گویندگان در حوزه‌ی امتیازها از روش هنجارسازی Znorm استفاده می‌شود [۲۲]. اساس روش Znorm بر پایه فرض گوسی بودن توزیع امتیازهای مربوط به گویندگان هدف و گویندگان وانمودکننده می‌باشد. اگر در روال آزمون سیستم GMM-UBM، امتیازهای حاصل از مدل گویندگان را از امتیازهای حاصل از مدل UBM کم نکنیم، می‌توان برای هر آزمایش به جای یک امتیاز، برداری دوبعدی از امتیازهای مدل GMM و مدل UBM به‌دست آورد. با رسم امتیازهای به دست آمده از مدل GMM برای گویندگان هدف و گویندگان وانمود کننده برحسب امتیازهای به دست آمده از مدل UBM می‌توان یک دید شهودی از توزیع امتیازها در فضایی دوبعدی به‌دست آورد. نقاط این فضای دو بعدی به شکل زیر به دست می‌آیند:

$$\vec{S}_{impostor} = \begin{bmatrix} S_{impostor-GMM} \\ S_{impostor-UBM} \end{bmatrix} \quad (۸)$$

$$\vec{S}_{target} = \begin{bmatrix} S_{target-GMM} \\ S_{target-UBM} \end{bmatrix} \quad (۹)$$

با الهام گرفتن از فرض توزیع گوسی امتیازها که در روش Znorm به آن اشاره شده بود، سیستمی مطابق شکل (۵) پیشنهاد می‌شود. در مورد سیستم پیشنهادی توضیحات زیر ضروری است:

شکل (۵): سیستم پیشنهادی به عنوان پردازشگر حوزه‌ی امتیازها

الف) سیستم پیشنهاد شده در ادامه‌ی سیستم GMM-UBM و به‌عنوان پردازش‌گر پسین در حوزه‌ی امتیازها عمل می‌کند.

ب) ساختار سیستم پیشنهادی مشابه یک سیستم تصدیق هویت گوینده است، ولی تفاوت اینجاست که در سیستم تصدیق هویت گوینده در مرحله‌ی آموزش هر گوینده، یک مدل GMM از روی مدل UBM تطبیق داده می‌شود ولی در پردازش‌گر پسین پیشنهادی، ابتدا سیستم GMM-UBM با یک‌سری گویندگان هدف و گویندگان وانمودکننده مورد آزمون قرار می‌گیرد تا امتیازهای $\vec{s}$ تولید شود. سپس برای تمامی امتیازهای به‌دست آمده از گویندگان هدف یک مدل GMM به نام Γ_{target} و از روی تمامی امتیازهای به دست آمده از گویندگان وانمودکننده، یک مدل GMM دیگر به نام $\Gamma_{impostor}$ آموزش داده می‌شود. هنگام آزمون سیستم مجتمع حاصل از سری کردن سیستم GMM-UBM و پردازش‌گر پسین پیشنهادی، که از این پس آن را «شناسایی

[1] Back-Propagation

کنندهٔ GMM»[1] می‌نامیم، امتیاز Λ' معیار تصمیم‌گیری برای تصدیق هویت گوینده واقع می‌شود.

ج) سیستم شناسایی کننده‌ی GMM پیشنهادی می‌تواند به جای شبکه‌ی عصبی در ادامه‌ی سیستم SBM-SGMM قرار گیرد و در نقش پردازش‌گر پسین انجام وظیفه کند. تفاوت این وضعیت با حالت (ب) در این است که ابعاد بردار امتیازهای $\vec{s}$ در اینجا می‌تواند بیشتر از ۲ باشد. اگر ساختار درختی با L لایه را مورد نظر بگیریم، بردار $\vec{s}$ می‌تواند دارای طولی برابر $2(L\text{-}1)$ باشد.

د) در وهله‌ی اول ممکن است این سؤال مطرح شود که چگونه برای کاهش بار محاسباتی سراغ SBM-SGMM می‌رویم ولی در مرحله‌ی تلفیق امتیازها دوباره سراغ GMM و مقداردهی آن می‌آییم. در پاسخ به این پرسش باید گفت که در سیستم GMM UBM یا SBM-SGMM به‌ازای یک تکه گفتار آزمون که T بردار مشخصه از آن استخراج شده باشد، T بار باید سیستم، آمیزه‌های گوسی را ارزیابی و سپس حول امتیازهای حاصله متوسط‌گیری کند. ولی هنگامی که نوبت به‌پردازش‌گر پسین شناسایی کننده GMM می‌رسد، به‌ازای T بردار مشخصه حاصل از تکه گفتار آزمون، یک بردار امتیازهای $\vec{s}$ به‌دست آمده است که GMMهای مدل‌های Γ_{target} و $\Gamma_{impostor}$ فقط یک بار لازم است ارزیابی شوند تا امتیاز Λ' محاسبه شود. نکته‌ی دیگری که باید در نظر داشت آن است که طول بردارهای مشخصه‌ی ورودی سیستم GMM-GUBM یاSBM-SGMM در اندازه‌ی حداقل ۱۲ تا حداکثر ۵۷ است (متناظر با ۱۲ تا ۱۹ ضریب MFCC که در مقالات مشاهده شده است) ولی طول بردارهای امتیاز ورودی شناسایی کننده‌ی GMM در اندازه‌ی ۲ تا ۶ و یا حداکثر ۸ می‌تواند باشد (متناظر با ساختار درختی ۵ لایه). این موضوع، بار دیگر کم بودن حجم محاسبات شناسایی کننده‌ی GMM را که به‌عنوان پردازش‌گر پسین استفاده می‌شود. توجیه می‌نماید.

۳- توصیف سیستم

۳-۱- بانک اطلاعاتی گفتار

بررسی عملکرد سیستم‌های تصدیق هویت گوینده نیاز به یک بانک اطلاعاتی گفتار دارد. جمع‌آوری نمونه‌های گفتار افراد از یک کارت تلویزیون رایانه‌ای با علامت تجاری Win Fast PVR انجام شده است. روال کار بدین‌گونه بود که وقتی یک گوینده در محیط استودیو (بدون حضور نویز خارجی) و با نرخ معمول، گفتار می‌کرد، صدای وی توسط خط داخلی رایانه به کارت صوتی منتقل و با نرم‌افزار ضبط‌کننده‌ی صدای تحت سیستم عامل ویندوز نمونه‌های گفتار وی ذخیره شد. گفتارهای هر گوینده با فرکانس نمونه‌برداری ۱۱۰۲۵ هرتز به‌صورت ۱۶ بیتی Mono و با پسوند PCM ذخیره شده است. هنگام ذخیره کردن نمونه‌های گفتار افراد، دقت شده است که فرد در حالت عادی و بدون هیجان یا استرس صحبت کند و تداخلی از گفتار دیگر گویندگان روی گفتار وی وجود نداشته باشد. محدوده‌ی سنّی افرادی که گفتار آن‌ها ضبط شده است، در محدوده‌ی تقریبی ۲۵ تا ۵۰ سال قرار دارد و همگی مرد هستند.

اولین مجموعه‌ی ذخیره شده از بانک اطلاعاتی گفتار، شامل تقریباً ۱/۵ ساعت گفتار از ۱۰۰ نفر گوینده می‌باشد که طول گفتار هر گوینده بین ۱۵ تا ۶۰ ثانیه است. از این مجموعه برای آموزش مدل UBM استفاده شده است. هر چقدر تعداد گویندگان و زمان گفتار آنها بیشتر باشد، مدل UBM رخدادهای فونتیکی موجود را بهتر می‌تواند پوشش دهد. مدل UBM با تمامی گفتار متعلق به مجموعه‌ی اول دادگان اطلاعاتی گفتار، آموزش داده شده است. دومین مجموعه‌ی ذخیره شده از دادگان اطلاعاتی گفتار شامل تقریباً ۳/۵ ساعت گفتار از ۶۰ نفر گوینده می‌باشد که طول گفتار هر گوینده بین ۳ تا ۴ دقیقه است. در مجموعه‌ی دوم دادگان اطلاعاتی گفتار تهیه شده برای هر گوینده ۳ تا ۴ دقیقه گفتار وجود دارد که از این مقدار، یک دقیقه‌ی اول برای آموزش مدل گویندگان، توسط تطبیق MAP نسبت به مدل UBM استفاده شده است. آموزش مدل گویندگان با زمان‌های ۱۵ و ۴۵ ثانیه به‌صورت مجزا انجام شده است تا بتوان تأثیر زمان آموزش بر عملکرد سیستم تصدیق هویت گوینده را بررسی کرد. از باقیمانده‌ی گفتار گوینده که بین ۲ تا ۳ دقیقه است برای آزمون استفاده شده است. سومین مجموعه‌ی ذخیره شده از دادگان اطلاعاتی گفتار شامل تقریباً ۱۰۰ دقیقه گفتار از ۳۰ نفر گوینده می‌باشد که طول گفتار هر گوینده بین ۳ تا ۴ دقیقه است. از این مجموعه نیز همانند مجموعه‌ی دوم برای آموزش و آزمون استفاده شده است ولی از امتیازهای حاصل از لایه‌های مختلف مدل SBM-SGMM این گویندگان برای آموزش پردازش‌گر

[1] GMM Identifier

پسین که شبکه‌ی عصبی، یا شناسایی‌کننده GMM می‌باشد استفاده شده است.

۳-۲- پیکربندی سیستم

بردارهای مشخصه‌ی استفاده شده در این پروژه شامل بردارهای ضرایب MFCC ۱۹ بعدی هستند که ۱۹ ضریب ΔMFCC نیز به خاطر لحاظ کردن تغییرات دینامیک ضرایب به بردار مشخصه اضافه شده‌اند و تشکیل یک بردار مشخصه‌ی ۳۸ بعدی را می‌دهند. فیلتر پیش تأکید، یک فیلتر FIR بالاگذر مرتبه‌ی ۱ با تابع تبدیل $H(z)=1-az^{-1}$ است که پارامتر a در آن مقدار ۰/۹۷ قرار داده شده است. طول قاب برابر ۲۳/۲ میلی ثانیه و گام قاب نیز برابر نصف طول قاب در نظر گرفته شده است. حذف سکوت به‌منظور حذف اطلاعات زاید و نگهداری اطلاعات مربوط به قسمت‌های واکدار گفتار گوینده انجام می‌شود. روش به‌کار رفته برای حذف سکوت، محاسبه‌ی یک مدل Bi-Gaussian از توزیع انرژی قاب‌های گفتار می‌باشد [۲۳]. در این حالت، برای نمونه‌های گفتار بدون نویز یا نویز پس‌زمینه‌ی کم، توزیع گوسی با میانگین پایین‌تر متناظر با سکوت و توزیع گوسی با میانگین بالاتر متناظر با گفتار می‌باشد. برای حذف سکوت، انرژی هر بردار به‌ازای دو توزیع گوسی حاصله، ارزیابی می‌شود و بردارهای مشخصه‌ای که انرژی متناظر با آنها دارای درست‌نمایی بالاتری در توزیع گوسی مربوط به سکوت باشد. به‌عنوان سکوت تلقی و حذف می‌شوند.

مقدار حداقل واریانس برای الگوریتم EM در آموزش مدل GMM در کار فعلی برابر ۰/۰۱ منظور شده است. شرط همگرائی الگوریتم EM، انجام ۵۰ تکرار از این الگوریتم و یا افزایش مقدار تابع درست‌نمایی در دو تکرار متوالی با مقداری کمتر از ۰/۰۱ تعیین شده است. در آموزش مدل GMM گویندگان از روی مدل UBM فقط میانگین‌ها تطبیق داده شده‌اند؛ زیرا طبق نتایج بررسی‌های انجام شده در [۱۱]، عملکرد کلی سیستم در حالت تطبیق میانگین‌ها بهتر از حالتی است که هر کدام از پارامترهای ضرایب وزنی، میانگین‌ها و واریانس‌ها به تنهایی یا با هم تطبیق داده شوند. مقدار ضریب وابستگی r برابر ۱۶ بوده و در مرحله‌ی آزمون تعداد C آمیزه‌ی جزء که بالاترین امتیاز را در UBM آورده‌اند برابر ۵ منظور شده است.

۳-۳- نحوه‌ی ارزیابی سیستم تصدیق هویت گوینده

برای ارزیابی عملکرد سیستم‌های تصدیق هویت گوینده یک تابع هزینه‌ی آشکارسازی، DCF[1]، توسط NIST، معرفی شده است که یک میانگین ریاضی وزن‌دار از نرخ‌های خطاهای از دست دادن، E_{miss} و آژیر اشتباه، E_{fa} می‌باشد[۲۴]. معیار کنونی این مزیت را دارد که کاربرد مربوطه را مدل کرده و عددی را تولید می‌کند که مستقیماً برای آن کاربرد خاص دارای مفهوم می‌باشد. هزینه‌ی تشخیص، $C_{\det}$، به‌صورت مجموعی وزن دار از احتمال پسین از دست دادن و آژیر اشتباه به‌صورت زیر مدل می‌شود:

$$DCF=C_{miss}E_{miss}P_{target}+C_{fa}E_{fa}(1-P_{target}) \qquad (۱۰)$$

در روابط فوق C_{miss} و C_{fa} هزینه‌های از دست دادن و آژیر اشتباه بوده و P_{target} احتمال پیشین گویندگان واقعی می‌باشد. همان‌گونه که از تعریف تابع هزینه‌ی تشخیص مشخص است این تابع علاوه بر استفاده از نرخ خطاهای از دست دادن و آژیر اشتباه دارای پارامترهایی چون C_{miss} و C_{fa} و P_{target} می‌باشد که قابلیت فوق‌العاده‌ای را به آن می‌بخشد تا برای کاربردهای مختلف و متناسب با شرایط، بتوان تابع هزینه را تعیین کرد. در کارهای تحقیقاتی طبق آنچه که NIST پیشنهاد کرده و در ارزیابی‌های خود آن را ملاک قرار می‌دهد، این پارامترها به ترتیب برابر $P_{target}=0.01$، $C_{miss}=10$، $C_{fa}=1$ مقداردهی می‌شوند.

نمودار DET[2] که به‌منظور نشان دادن عملکرد سیستم با مقادیر آستانه‌ی مختلف استفاده می‌شود، با استفاده از ترسیم انحراف‌های گوسی[3] متناظر با احتمالات از دست دادن و آژیر اشتباه به‌جای ترسیم خود احتمالات از دست دادن و آژیر اشتباه، نمایش شهودی بهتری را نسبت به نمودارهای ROC[4] فراهم می‌کند [۲۵]. نمودار DET دارای مقیاس احتمال غیرخطی است و درصورتی‌که توزیع احتمالات خطاها به‌صورت گوسی باشد، منحنی‌های به‌دست‌آمده حاصله، به‌صورت خطوط راست خواهند بود. مزیت مهم استفاده از نمودار DET در این است که فاصله‌ی بین منحنی‌ها، تفاوت عملکردها را به‌صورت مؤثرتری نسبت به

[1] Detection Cost Function
[2] Detection Error Trade-off
[3] Gaussian Deviates
[4] Receiver operating characteristics

نمودار ROC نشان می‌دهد. نرخ خطای برابر (EER[1]) برای یک سیستم، نقطه‌ی کاری است که در آن با تنظیم مقدار آستانه‌ی تصمیم‌گیری نرخ خطای از دست دادن با نرخ خطای آژیر اشتباه برابر شود.

۴- شبیه‌سازی سیستم تصدیق هویت گوینده SBM-SGMM در ترکیب با شبکه‌ی عصبی

در این بخش عملکرد سیستم تصدیق هویت گوینده که بر پایه‌ی مدل آمیزه‌های گوسی ساختاری و شبکه‌های عصبی شبیه‌سازی شده است، بررسی می‌شود. قابل ذکر است که بخشی از نتایج کارهای انجام شده توسط مؤلفان در مراجع [۲۶-۲۸] آورده شده است که در این مقاله، نتایج جمع بندی می‌شود. همان‌گونه که اشاره شد، هدف از استفاده‌ی مدل آمیزه‌های گوسی ساختاری، بالا بردن سرعت محاسبات (کاهش بار محاسباتی) در مرحله‌ی آزمون می‌باشد. در مرجع [۱۵] برای اوّلین‌بار مدل آمیزه‌های گوسی ساختاری برای استفاده در تصدیق هویت گوینده توسط بینگ شیانگ[2] معرفی شده است، امّا تأثیر دو عامل مشخص در ساخت مدل SBM بررسی نشده است که یکی از این عوامل ضریب درون‌یابی α می‌باشد. شیانگ در این مقاله مقادیر مربوط به ساخت مدل را به مرجع [۱۹] ارجاع کرده است، که در آن مقاله بدون بیان علت، *شینودا*[3] متذکر شده است که ضریب α در این کار برابر ۰/۱ قرار داده شده است. ضریب درون‌یابی α در حقیقت تعیین‌کننده‌ی سهم مشخصات آماری هر کدام از آمارگان گره‌ی فعلی و گره‌ی انشعابی از آن که به طریقه‌ی minimax به‌دست آمده است؛ در تعیین مشخصات آماری فعلی گره‌ی انشعابی می‌باشد. این کار در حقیقت تعیین نقطه‌ی شروع اولیه برای یک فرآیند K-mean می‌باشد که در طی آن فاصله‌ی بین مرکز خوشه‌ها با آمیزه‌های گوسی متعلق به آن خوشه‌ها، حداقل می‌شود. با تغییر ضریب درون‌یابی α، نقطه‌ی شروع فرآیند Kmeans تغییر کرده و در نهایت به مدل‌های مختلفی همگرا می‌شود. در تحقیق حاضر ضریب α از ۰/۱ تا ۱ با پله ۰/۱ تغییر داده شده و اثر آن در ساخت مدل آمیزه‌های گوسی ساختاری بررسی شده است. عامل دیگری که شیانگ به آن اشاره‌ای نکرده بود، نحوه‌ی انتساب آمیزه‌های گوسی به گره‌های انشعابی در طریقه‌ی minimax می‌باشد. این موضوع میزان نامتقارن بودن ساختار درست شده را تعیین می‌کند. شیانگ اشاره کرده است که تعداد گره‌های لایه‌ی تحتانی به‌ازای هر گره در لایه‌ی بالاتر متفاوت است، ولی میزانی برای متفاوت‌بودن ارایه نکرده است. در تحقیق حاضر دو روش ترتیبی و ترتیبی ـ ظرفیتی برای انتساب آمیزه‌های گوسی به گره‌های انشعابی ارایه شده و با تغییر ظرفیت هر گره‌ی انشعابی برای اختصاص آمیزه‌های گوسی به خود، ساختارهای با میزان‌های متفاوت نامتقارنی ساخته شده است. تأثیر این عامل بر عملکرد سیستم SBM-SGMM تحت عنوان متغیری به نام ضریب ظرفیت بیشینه بررسی شده است.

SBM بر مبنای UBM ساخته می‌شود و سپس بر مبنای ساختار به دست آمده؛ مدل SGMM گویندگان به‌صورت لایه‌به‌لایه توسط تطبیق MAP از مدل SBM حاصل می‌شود. به علت این‌که لایه‌ی تحتانی SBM همان UBM است، تطبیق لایه‌به‌لایه‌ی SBM برای حصول مدل SGMM گویندگان فقط در *L-1* لایه‌ی بالاتر انجام می‌شود. با توجّه به نتایج به‌دست آمده در بخش قبل، مدل ۶۴تایی برای دادگان اطلاعاتی گفتار موجود می‌تواند مدل بهینه‌ای باشد که توزیع آماری کلاس‌های صوتی موجود در گفتار گویندگان را به‌خوبی مدل می‌کند. به همین علت از مدل ۶۴ تایی برای ساخت ساختار درختی استفاده شده است. در طی آزمون‌های متفاوت زمان آموزش و آزمون بهینه برای مدل ۶۴ تایی به‌ترتیب برابر ۴۵ ثانیه‌ی و ۷ ثانیه به دست آمدند. روی مدل ۶۴ تایی ساختاری یک آزمایش ۴۵ ثانیه آموزش با ۷ ثانیه آزمون و یک آزمایش ۱۵ ثانیه‌ی آموزش با ۳ ثانیه آزمون انجام شده است، تا عملکرد سیستم روی داده‌های مختلف بررسی شود. بر مبنای سیستم ارائه شده‌ی شیانگ، شبکه‌ی عصبی استفاده شده به‌صورت MLP سه‌لایه می‌باشد که دارای تعداد گره‌های مختلف در لایه‌ی ورودی (بسته به تعداد لایه‌های مدل آمیزه‌های گوسی ساختاری)، تعدادی گره در لایه‌ی مخفی و یک گره در لایه‌ی خروجی می‌باشد. عملکرد سیستم تصدیق هویت گوینده نسبت به تعداد گره‌های لایه‌ی ورودی و تعداد گره‌های لایه‌ی مخفی نیز بررسی شده است. معیار فاصله برای انجام شبیه‌سازی‌ها، معیار فاصله‌ی واگرایی KL در نظر گرفته شده است. به هنگام تطبیق مدل SGMM گویندگان از روی مدل SBM در *L-1* لایه بالای لایه‌ی تحتانی هم میانگین‌ها و هم

[1] Equal Error Rate
[2] Bing Xiang
[3] Shinoda

واریانس‌ها تطبیق داده شدند. زیرا آن‌گونه که در [۱۵] بدان اشاره شده است؛ به علت اینکه تعداد آمیزه‌های گوسی به نسبت داده‌های آموزشی لایه‌های بالاتر کمتر است، عملکرد کلی سیستم در حالت تطبیق میانگین‌ها و واریانس‌ها بهتر از حالتی است که تنها میانگین‌ها و یا هر سه پارامتر ضرایب وزنی، میانگین‌ها و واریانس‌ها تطبیق شوند.

۴-۱- ضریب کاهش بار محاسباتی

برای نمایش بار محاسباتی مورد نیاز سیستم GMM UBM نسبت به سیستم SBM-SGMM یک ضریب کاهش بار محاسباتی به‌صورت زیر تعریف می‌شود [۱۵]:

$$F = \frac{M+C}{M_S+C_S} \qquad (۱۱)$$

در رابطه‌ی بالا M، C و C_s تعداد آمیزه‌های گوسی است که باید به ترتیب در UBM، مدل GMM گوینده و مدل SGMM گوینده ارزیابی شوند. M_S متوسط تعداد آمیزه‌های گوسی است که باید در SBM مورد ارزیابی قرار گیرد. در تمامی آزمایش‌های انجام شده در این بخش M برابر ۶۴ و C برابر ۵ می‌باشد. C_s را می‌توان به‌صورت $C_S=C+L-2$ محاسبه کرد. در این رابطهٔ L تعداد لایه‌های ساختار SBM-SGMM می‌باشد. علّت محاسبه‌ی C_s از رابطهٔ فوق این است که از لایه‌ی دوم تا لایه‌ی *L-1* تنها یک گره (معادل یک آمیزه‌ی گوسی) مورد ارزیابی قرار می‌گیرد و در لایه‌ی Lام (لایه‌ی تحتانی) تعداد C آمیزه‌ای که بالاترین امتیاز را در لایه‌ی تحتانی SBM کسب کرده باشند، ارزیابی می‌شوند. مقدار M_s را می‌توان به وسیله‌ی مجموع میانگین زیر تقریب زد:

$$M_S = \frac{1}{T}\sum_{t=1}^{T}\sum_{l=2}^{L} n_{l,t} \qquad (۱۲)$$

در رابطه‌ی فوق $n_{\ell,t}$ تعداد گره‌هایی است که در لایه‌ی ℓام، SBM به ازای بردار ورودی $\vec{x}_t$ مورد ارزیابی واقع می‌شوند. شیانگ در [۱۵] به علّت این‌که حالت نامتقارن بودن ساختار درختی را مورد بررسی قرار نداده بود، به حالتی هم برخورد نکرده بود که تعداد آمیزه‌های گوسی متعلق به گرهی در لایه‌ی *L-1* کمتر از C باشد. به همین علّت رابطه‌ی محاسبه‌ی C_s در کار حاضر به‌صورت زیر تصحیح می‌شود:

$$C_S = L - 2 + \frac{1}{T}\sum_{t=1}^{T} C_t \qquad (۱۳)$$

۴-۲- ساخت مدل *SBM*

همان‌گونه که اشاره شد، تأثیر دو عامل در ساخت مدل SBM تا به حال بررسی نشده است. هدف از آزمایش‌های انجام شده در این بخش، بررسی تأثیر ضریب درون‌یابی و ضریب ظرفیت بیشینه بر ساختار SBM و انتخاب مدل SBM بهینه برای ادامه‌ی کار می‌باشد. همان‌طور که در بخش‌های قبل اشاره شده بود، راه‌کار مشخصی برای تعیین تعداد لایه‌های ساختار درختی و تعداد گره‌های موجود در هر لایه وجود ندارد و ساختار بهینه ممکن است وابسته به نوع داده‌ها تغییر کند. در [۱۵] تأثیر انتخاب ساختارهای مختلف بر عملکرد سیستم و نیز بده‌بستان بین ضریب کاهش بار محاسباتی و حداقل تابع هزینه، بررسی و گزارش شده است. در اینجا از تکرار این امر خودداری شده و با توجه به UBM با ۶۴ آمیزه‌ی گوسی، ساختاری با یک گره در لایه‌ی اوّل، ۴ گره در لایه‌ی دوم و ۶۴ گره در لایه‌ی تحتانی مورد بررسی قرار گرفت. در صورتی که ساختار درختی به صورت متقارن ساخته شود، هر گره در لایه‌ی دوم، ۱۶ گره در لایه‌ی تحتانی را در زیر مجموعه‌ی خود خواهد داشت. ضریب درون‌یابی در رابطه‌ی (۲) ظاهر شده است که با تغییر این ضریب در بازه‌ی ۰/۱ تا ۱ با فواصل ۰/۱ می‌توان مدل‌های SBM مختلفی را به‌دست آورد. برای انتساب آمیزه‌های گوسی به گره‌های یک لایه از ساختار SBM، دو روش به صورت زیر پیشنهاد می‌شود:

- **انتخاب نوبتی:** انتخاب نوبتی بدین معنی است که پس از محاسبه‌ی فاصله‌ی واگرایی KL بین آمیزه‌های گوسی موجود با هر کدام از گره‌های لایه، هر کدام از گره‌ها به نوبت یکی از آمیزه‌های گوسی را که به خود نزدیکتر است انتخاب کرده و زیر مجموعه‌ی خود درآورد. بـــه این ترتیب ساختار درختی ایجاد شده به صورت کاملاً متقارن خواهد بود.
- **انتخاب نوبتی ـ ظرفیتی:** این روش مشابه روش انتخاب نوبتی می‌باشد ولی این تفاوت که ابتدا هر آمیزه‌ی گوسی به نزدیکترین گره نسبت داده می‌شود. درصورتی‌که تعداد آمیزه‌های گوسی زیر مجموعه‌ی یک گره از ظرفیت نامی آن گره بیشتر باشد، آمیزه‌هایی از این زیر مجموعه که فاصله‌ی آنها به نسبت بقیه‌ی آمیزه‌ها از گره بیشتر است، از زیر

مجموعه‌ی گره خارج شده و در معرض انتخاب بقیه‌ی گره‌ها قرار می‌گیرد.

ظرفیت نامی یک گره توسط ضریبی به نام ضریب ظرفیت بیشینه که در ظرفیت متقارن یک گره ضرب می‌شود، محاسبه می‌شود. ظرفیت متقارن، تعداد آمیزه‌های زیر مجموعه‌ی یک گره در حالت متقارن می‌باشد و ضرایب ظرفیت بیشینه از یک شروع شده و با فواصل ۰/۱ افزایش می‌یابد. به‌عنوان مثال در ساختار آزمایش شده در این بخش که به‌صورت ۶۴-۴-۱ می‌باشد، تعداد آمیزه‌های گوسی زیرمجموعه‌ی هر گره در لایه‌ی دوم در حالت متقارن برابر ۱۶ می‌باشد (ظرفیت متقارن). برای به‌دست آوردن ظرفیت نامتقارن هر گره، ضریب ظرفیت بیشینه در ظرفیت متقارن ضرب می‌شود: ۱۶×۱، ۱۶×۱/۱، ۱۶×۱/۲ و ... درصورتی‌که عدد حاصل شده برای ظرفیت نامتقارن، عددی اعشاری باشد به سمت بالا گرد می‌شود.

همان‌گونه که می‌توان از رفتار روش ترتیبی، نسبت به روش ترتیبی ـ ظرفیتی انتظار داشت، عملکرد آن نسبت به روش ترتیبی ـ ظرفیتی بسیار نازل‌تر بود. این امر را می‌توان بدین‌گونه توضیح داد که درصورتی تعداد بیشتری از آمیزه‌های گوسی پیرامون یکی از گره‌ها متمرکز باشند، با این روش به‌اجبار تعدادی از آمیزه‌های گوسی که نسبت به این گره دورتر از بقیه‌ی آمیزه‌ها هستند (ولی نسبت به بقیه‌ی گره‌ها به این گره نزدیک‌ترند)، در تصرف گره‌های دیگر قرار می‌گیرند. این نقیصه در روش نوبتی ـ ظرفیتی با افزایش ظرفیت گره‌ها جبران می‌شود. در روش نوبتی ـ ظرفیتی ضریب ظرفیت بیشینه از ۱ شروع می‌شود (ضریب ظرفیت بیشینه ۱ به معنی ساختار متقارن است)، و با فواصل ۰/۱ افزایش می‌یابد. افزایش ضریب ظرفیت را در این آزمایش تا جایی ادامه می‌دهیم که یکی از گره‌ها دیگر قادر به تصرف حتی یک آمیزه‌ی گوسی نباشد. در این حالت باید از مجموعه‌ی گره‌های لایه‌ی یک عدد کم کرد و این کار چون با فرض اولیه در ساختن ساختار درختی در تناقض است، افزایش ظرفیت متوقف شده است. در جدول (۱)، تأثیر تغییر ضریب درون‌یابی بین ۰/۱ تا ۱ با فواصل ۰/۱ و تغییر ضریب ظرفیت بیشینه شروع از ۱و افزایش‌های ۰/۱ بررسی شده است. دو معیار مقایسه بین ساختارهای متفاوت تولید شده در نظر گرفته شده است: ۱- مجموع کل فواصل واگرایی KL بین آمیزه‌های گوسی متعلق به هر گره با آن گره. ۲- امتیاز حاصل از ساختار SBM تولید شده به‌ازای ۵۰۰ بردار مشخصه که از مرحله‌ی آخر ساخت UBM به‌صورت تصادفی انتخاب شده‌اند. امتیازهای به‌دست آمده در قالب دو امتیاز می‌باشد که یکی امتیاز حاصل از لایه‌ی دوم و دیگری امتیاز حاصل از لایه‌ی تحتانی می‌باشد.

مجموع کل فواصل داخل خوشه‌ها، معیاری از میزان برازش آمارگان گره‌ها بر روی آمارگان آمیزه‌های گوسی متعلق به آن گره‌ها می‌باشد. این مجموع هر چقدر کمتر باشد، نشان می‌دهد که نقطه‌ی شروع اولیه به‌طور مناسب انتخاب شده است؛ زیرا الگوریتم K-mean در هر تکرار خود سعی در کم‌کردن این فواصل می‌نماید. آن‌گونه که در آزمایش‌های انجام گرفته مشاهده شده است، عموماً ۵ تا ۱۰ تکرار الگوریتم K-mean برای همگرایی الگوریتم همگرایی کامل را نتیجه داده است. امتیاز UBM به‌صورت مجموع $C=5$ آمیزه‌ی گوسی که در بین ۶۴ آمیزه‌ی گوسی UBM حداکثر امتیاز را آورده‌اند، محاسبه شده است. هر چقدر که امتیاز لایه‌ی تحتانی SBM به امتیاز UBM نزدیک‌تر باشد، ساختار SBM در روال جستجو به‌طور متوسط بهتر عمل کرده است. تعداد آمیزه‌های گوسی موجود در هر گره وابسته به استفاده‌ی از یکی از دو روش انتخاب نوبتی یا نوبتی- ظرفیتی می‌باشد و حداکثر تعداد آمیزه‌های گوسی موجود در یک گره، برابر ظرفیت بیشینه آن گره می‌باشد که در قسمت قبلی توضیح داده شده است. لایه‌های مختلف SBM، فضای آکوستیکی را با مراتب تفکیک‌پذیری متفاوتی مدل می‌کنند. این تفکیک‌پذیری از بالا به پایین بیشتر می‌شود به‌گونه‌ای که انتظار می‌رود لایه‌ی تحتانی حداکثر تفکیک‌پذیری را با بیشترین تعداد آمیزه‌های گوسی دارد، بهترین امتیاز را به دست آورد. بهترین امتیاز در اینجا امتیازی است که نزدیک‌تر به امتیاز UBM باشد، همان‌گونه که در جدول (۱) مشاهده می‌شود، نمی‌توان تأثیر ضریب درون‌یابی بر ساختار SBM را به‌صورت یک قاعده‌ی کلّی بیان کرد. زیرا روال مشخصی در رفتار مجموع کل فواصل داخل خوشه‌ها و امتیازهای لایه‌های دوم و سوم SBM با افزایش ضریب درون‌یابی وجود ندارد. نکته‌ی آشکار این است که مقدار بهینه‌ی ضریب درون‌یابی به‌ازای $\alpha=0.1$ به دست نمی‌آید و برای تعیین مقدار بهینه‌ی ضریب درون‌یابی در ساخت SBM باید بین مقادیر مختلف α جستجو کرده و مقداری را که باعث به‌وجود آمدن SBM مناسب‌تر می‌شود، برگزید.

جدول (۱): بررسی تأثیر تغییر ضریب درون‌یابی و ضریب ظرفیت بیشینه بر کیفیت تولید SBM

امتیاز UBM به ازای بردارهای مشخصه – ۵/۳۸۸۳-										
ضریب درونیابی / ضریب ظرفیت ماکزیمم	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
	مجموع کل فواصل داخل خوشه ها									
۱	۲۳۰۱/۹	۲۵۱۴/۵	۲۶۴۲	۲۴۹۶/۵	۲۵۱۶/۶	۲۴۴۳	۲۴۰۷/۱	۲۳۸۰/۸	۲۳۵۸/۸	۲۴۶۰
۱/۱	۲۱۵۶/۴	۲۳۱۵/۱	۲۰۴۰/۳	۲۲۴۲/۹	۲۳۵۶/۴	۲۳۰۵/۹	۲۳۰۹/۳	۲۲۶۹/۹	۲۳۵۳	۲۲۹۳/۲
۱/۲	۱۸۸۸/۳	۱۶۰۸/۸	۱۸۷۴/۵	۱۹۰۶/۲	۱۸۸۴/۶	۱۹۲۸/۱	۱۸۶۱/۹	۱۹۲۰	۱۸۹۸/۳	۱۸۳۸/۴
۱/۳	۱۳۰۵/۹	۱۳۰۵/۹	۱۳۵۲/۱	۱۲۸۱	۱۲۸۸/۲	۱۲۸۶/۴	۱۳۰۰/۱	۱۲۸۲	۱۲۸۶/۵	۱۲۹۱/۷
۱/۴	۱۲۹۳	۱۳۷۳/۵	۱۲۴۸/۳	۱۱۹۴/۳	۱۲۱۰/۹	۱۳۰۱/۸	۱۲۰۳/۳	۱۳۰۵/۳	۱۳۰۲/۸	۱۳۴۸/۶
۱/۵	۱۱۹۱/۷	۱۱۸۹/۴	۱۲۷۹/۸	۱۱۹۱/۲	۱۲۵۸/۶	۱۲۱۸/۵	۱۲۲۰/۷	۱۲۹۳/۹	۱۳۱۱	۱۲۶۸/۸
امتیاز لایهٔ دوم SBM										
۱	-۵/۷۲۲۷	-۵/۷۷۰۸	-۵/۸۶۸۶	-۵/۸۵	-۵/۹۶۷۵	-۵/۷۳۳۳	-۵/۸۱۹۴	-۵/۸۰۱۸	-۵/۹۱۴۵	-۵/۸۷۳۸
۱/۱	-۵/۷۸۲۷	-۵/۷۷۶	-۵/۸۰۰۶	-۵/۸۴۱۹	-۵/۷۱۷۴	-۵/۸۴۱۳	-۵/۷۳۴۵	-۵/۷۳۶۲	-۵/۷۱۵۳	-۵/۸۰۷۴
۱/۲	-۵/۷۰۴۶	-۵/۷۰۴۶	-۵/۸۵۸۷	-۵/۷۴۸۵	-۵/۷۰۷۹	-۵/۶۹۷۷	-۵/۸۳۰۷	-۵/۶۷۸۶	-۵/۷۱۰۸	-۵/۷۵۴۵
۱/۳	-۵/۷۶۶۸	-۵/۸۸۶۳	-۵/۷۰۳۵	-۵/۵۹۳۲	-۵/۶۴۳۸	-۵/۸۱۹۸	-۵/۶۳۶۷	-۵/۸۰۱۶	-۵/۷۶۶	-۵/۹۰۸۷
۱/۴	-۵/۶۴۸۸	-۵/۶۳۳۳	-۵/۷۴۹۵	-۵/۶۶۵۱	-۵/۶۷۸۹	-۵/۶۲۳	-۵/۶۷۴۳	-۵/۸۲۶۲	-۵/۷۸۴۳	-۵/۷۳۴۲
۱/۵	-۵/۹۲۷۳	-۵/۹۳۶۸	-۵/۹۴۶۶	-۵/۹۲۵۱	-۵/۸۵۴۷	-۵/۸۳۵۹	-۵/۷۸۹۸	-۵/۹۳۹۳	-۵/۹۳۹	-۵/۹۰۹۷
امتیاز لایهٔ تحتانی SBM										
۱	-۵/۸۳۲۸	-۵/۷۶۳۲	-۵/۹۰۰۱	-۵/۸۳۸۸	-۵/۹۷۷۱	-۵/۸۵۸۱	-۵/۸۵۹۵	-۵/۹۲۶۱	-۵/۹۰۴۲	-۵/۸۹۵۵
۱/۱	-۵/۶۹۳۴	-۵/۷۱۲	-۵/۷۸۱۸	-۵/۸۰۳	-۵/۷۳۷۸	-۵/۷۹۱۷	-۵/۷۶۴۷	-۵/۶۹۸۳	-۵/۷۳۲۴	-۵/۷۴۱۹
۱/۲	-۵/۷۰۸	-۵/۷۰۸	-۵/۷۰۱۳	-۵/۶۹۸۸	-۵/۷۱۳	-۵/۷۱۷	-۵/۷۴	-۵/۷۱۴۳	-۵/۷۲۳۹	-۵/۷۱۸۱
۱/۳	-۵/۶۵۹۴	-۵/۷۳۴۱	-۵/۶۲۳۴	-۵/۶۳۱۷	-۵/۶۵۳۲	-۵/۷۴۹	-۵/۶۵۵۵	-۵/۷۱۴۸	-۵/۶۹۵۱	-۵/۷۷۴۴
۱/۴	-۵/۶۲۹۴	-۵/۶۵۵۱	-۵/۶۸۶۵	-۵/۶۴۲۲	-۵/۶۸۳۳	-۵/۶۵۹۳	-۵/۶۶۴۱	-۵/۷۶۵۹	-۵/۷۴۵۳	-۵/۶۵۰۵
۱/۵	-۶/۰۰۸۴	-۵/۹۵۴۲	-۵/۹۸	-۶/۰۸۶۷	-۵/۹۴۴۶	-۵/۹۸۷۵	-۶/۱	-۶/۰۲۲۹	-۶/۰۴۱	-۶/۰۷۷۱
تعداد آمیزه‌های گوسی موجود در هر گروه										
۱	۱۶ ۱۶ ۱۶ ۱۶									
۱/۱	۱۸ ۱۸ ۱۸ ۱۰									
۱/۲	۲۰ ۲۰ ۲۰ ۴									
۱/۳	۲۱ ۲۱ ۲۱ ۱									
۱/۴	۱۳ ۱۹ ۲۱ ۱									
۱/۵	۲۴ ۲۱ ۱۸ ۱									

تأثیر ضریب ظرفیت بیشینه بر ساخت SBM را می‌توان با وضوح بیشتری بیان نمود. هر چند، باز هم نمی‌توان یک قاعده‌ی کلی را در رابطه با تأثیر ضریب ظرفیت بیشینه بر ساخت SBM استنتاج کرد. همان‌گونه که انتظار می‌رفت، با افزایش ضریب ظرفیت بیشینه مجموع کل فواصل داخل خوشه‌ها کمتر می‌شود؛ این امر در اغلب موارد مشاهده می‌گردد، هر چند همواره چنین نیست. در مورد امتیازهای لایه‌ی دوم و لایه‌ی تحتانی نیز بیان به همین شیوه است. تعداد آمیزه‌های گوسی موجود در هر گره، نشان می‌دهد توزیع آمیزه‌های گوسی پیرامون گره‌ی چهارم، بیشینه و پیرامون گره‌ی اوّل کمینه می‌باشد. با توجه به تحلیل‌های ارایه شده و در نظر گرفتن این نکته که در تصمیم‌گیری نهایی سیستم تصدیق هویت گوینده امتیازها مهم می‌باشند و نه مجموع کل فواصل داخل خوشه‌ها، ساختاری که نزدیک‌ترین امتیاز به UBM را در لایه‌ی تحتانی خود به‌دست آورده باشد. به‌عنوان معیار ساختار بهینه در نظر گرفته شده است. با این معیار ساختار SBMی که به ازای پارامترهای ضریب درون‌یابی ۰/۳ و ضریب ظرفیت بیشینه ۱/۳ حاصل گردیده و امتیاز ۵/۶۲۳۴- را در لایه‌ی تحتانی کسب کرده است، به عنوان ساختار SBM بهینه در ادامه‌ی این تحقیق استفاده شده است. مقدار ضریب کاهش بار محاسباتی حاصل شده برای این ساختار در طول آزمایش‌ها به‌طور متوسط برابر ۲/۷۴ بود.

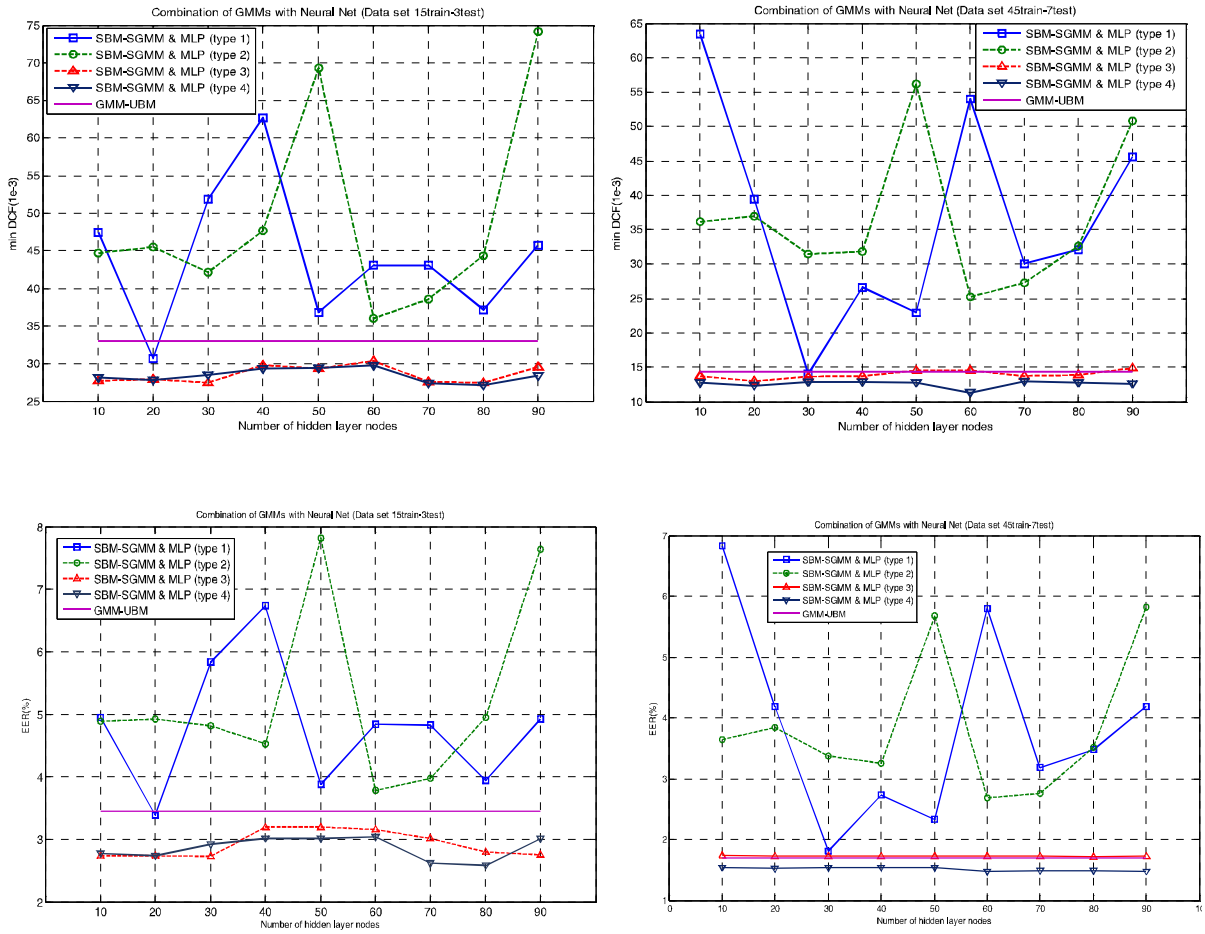
۴-۳- ترکیب امتیازها با شبکه‌ی عصبی

پس از تولید SBM، مدل گویندگان نسبت به این مدل تطبیق داده شده و مدل SGMM گویندگان به دست آمد. به این منظور از ۹۰ گوینده استفاده شده است که ۶۰ نفر از گویندگان همان گویندگانی هستند که در آزمایشهای سیستم GMM-UBM از آنها استفاده شده بود. از ۳۰ گوینده‌ی دیگر به‌منظور آموزش شبکه‌ی عصبی استفاده شده است. در ابتدای کار ۱۵۰۰ آزمایش گویندگان هدف و ۱۵۰۰۰ آزمایش گویندگان وانمودکننده با استفاده از ۳۰ گوینده روی ساختار SBM-SGMM انجام شد و امتیازهای حاصل از این گویندگان در قالب بردار امتیازهای $\bar{s}$ به‌علاوه‌ی برچسب گویندگان (مبنی بر اینکه امتیاز حاصله متعلق به گوینده‌ی هدف است یا متعلق به گوینده‌ی وانمودکننده) در اختیار شبکه‌ی عصبی قرار گرفت. برای انجام آزمایش‌ها، چهار نوع شبکه‌ی عصبی مورد آموزش قرار گرفت:

- شبکه‌ی عصبی با ۴ گره‌ی ورودی که ۲ امتیاز حاصل از SGMM و دو امتیاز حاصل از SBM بعنوان ورودی قبول می‌کرد.
- شبکه‌ی عصبی با ۲ گره‌ی ورودی که ۲ امتیاز حاصل از تفاضل امتیازهای دو لایه‌ی متناظر SGMM و SBM را بعنوان ورودی قبول می‌کرد.
- شبکه‌ی عصبی با ۲ گره‌ی ورودی که ۲ امتیاز حاصل از لایه‌های تحتانی SGMM و SBM را به‌عنوان ورودی قبول می‌کرد.
- شبکه‌ی عصبی با ۲ گره‌ی ورودی که ۲ امتیاز حاصل از GMM و UBM را به‌عنوان ورودی قبول می‌کرد.

شبکه‌ی عصبی نوع چهارم به‌منظور بررسی عملکرد سیستم ترکیبی GMM UBM با شبکه‌های عصبی نسبت به سیستم GMM-UBM مورد آزمایش قرار گرفته است. در تمامی آزمایش‌ها تعداد گره‌های لایه‌ی مخفی شبکه‌ی عصبی بین ۱۰ تا ۹۰ با فواصل ۱۰تای تغییر داده شده است تا بتوان تعداد بهینه‌ی گره‌های لایه‌ی مخفی را استنتاج کرد. آزمایش‌ها در دو گروه انجام شده است. گروه اوّل که از ۱۵ ثانیه گفتار برای آموزش SGMM و ۳ ثانیه گفتار برای آزمون SBM-SGMM استفاده کردند و گروه دوم که از ۴۵ ثانیه گفتار برای آموزش SGMM و ۷ ثانیه گفتار برای آزمون SBM-SGMM سود می‌برند. هدف از انجام آزمایش‌ها در دو گروه این بوده است که وابستگی نتایج به‌دست آمده به نوع امتیازهای حاصل از ساختار SBM-SGMM بررسی شود. نتایج به‌دست آمده از انجام آزمایش‌های گروه اوّل در شکل (۶) آورده شده است.

با بررسی نتایج حاصل از آزمایش‌های انجام شده روی داده‌های گروه ۱ و داده‌های گروه ۲ می‌توان چنین استنتاج کرد که با وجود اینکه ساختار SBM-SGMM امتیازها را با ابعاد بیشتری جهت تجزیه و تحلیل تولید می‌کند، ولی چون ابعاد بیشتر، از امتیازها لایه‌های بالاتر SBM-SGMM حاصل می‌شوند و دقت این لایه‌ها در مدل‌سازی گویندگان پائین است، این امتیازها دارای پراکندگی بالایی هستند. احتمالاً به همین علت عملکرد سیستم مجتمع با شبکه‌ی عصبی در

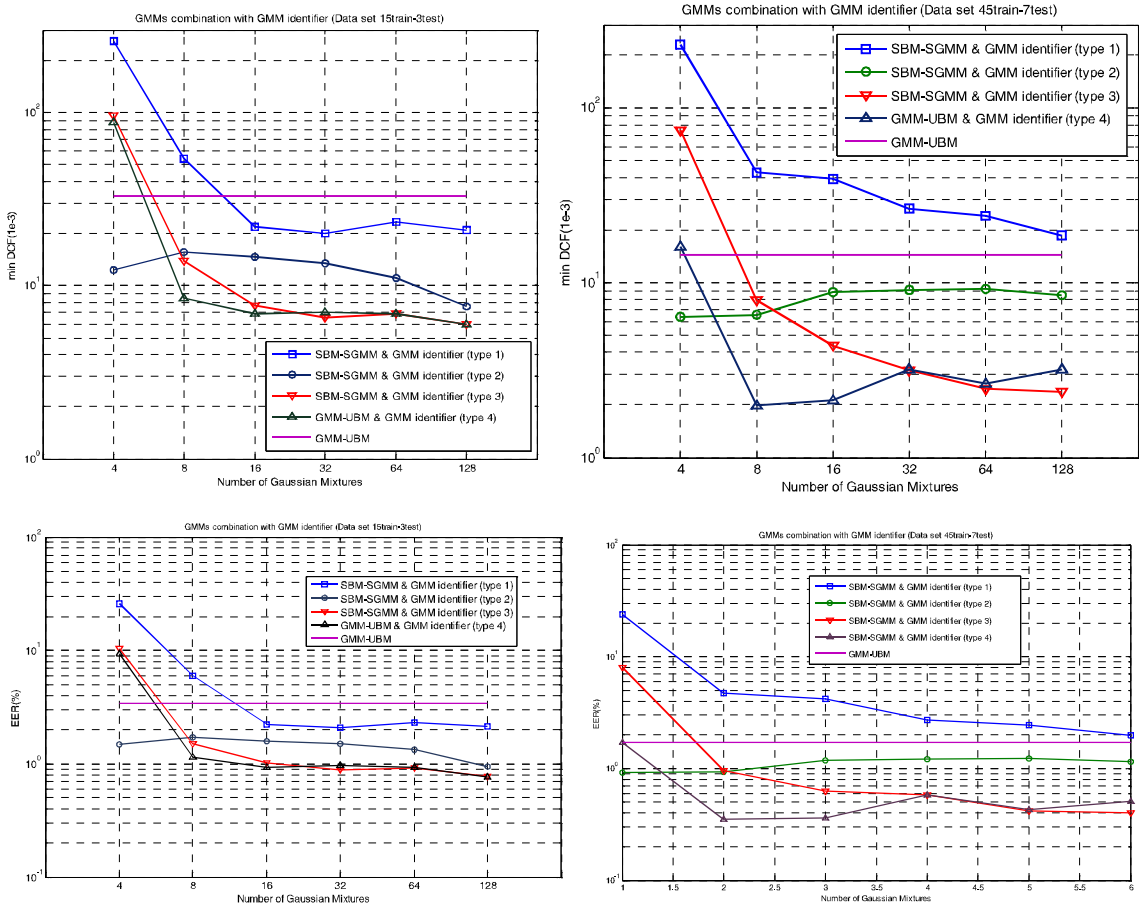


(ب) **(الف)**

شکل (۶): عملکرد شبکه‌ی عصبی به عنوان پردازشگر پسین در حوزه‌ی امتیازها الف) ۴۵ ثانیه آموزش،۷ ثانیه آزمون ب)۱۵ ثانیه آموزش، ۳ ثانیه آزمون (بالا نمودار min. DCF و پایین نمودار EER)

ترکیب‌های نوع اول و دوم چندان مطلوب نمی‌باشد. هر چند که با تعداد گره‌های لایه‌ی مخفی ۲۰ و ۳۰ به‌ترتیب برای داده‌های گروه اول و دوم، سیستم مجتمع SBM-SGMM-NN بـــه عملکردی بهتر از سیستم GMM-UBM دست پیدا می‌کند. اگر فقط از امتیازهای لایه‌ی تحتانی SBM و SGMM به‌عنوان ورودی شبکه‌ی عصبی استفاده شود، (ترکیب نوع سوم) عملکرد سیستم مجتمع در حالت کلی نسبت به GMM UBM بهتر شده و این امر در محدوده‌ی مورد آزمایش، تقریباً وابسته به تعداد گره‌های لایه‌ی مخفی نمی‌باشد. در این حالت سیستم با تعداد گره‌های لایه‌ی مخفی ۳۰ و ۲۰ به‌ترتیب برای داده‌های گروه اوّل و دوم به عملکرد بهینه نایل گردید. ترکیب امتیازهای GMM UBM با شبکه‌ی عصبی MLP (ترکیب نوع چهارم) در حالت کلی نسبت به GMM-UBM تنها، دارای عملکرد بهتری است.

این موضوع بدین علت است که در GMM-UBM فقط یک آستانه‌ی تصمیم‌گیری، ملاک تصمیم‌گیری نهایی است ولی در ترکیب با شبکه‌ی عصبی، شبکه‌ی عصبی با مدل کردن مرز بین امتیازها در فضای دو بعدی، یک مرز تصمیم‌گیری غیرخطی را پیدا کرده و ملاک تصمیم‌گیری قرار می‌دهد. در حالت کلی عملکرد سیستم ترکیبی نوع چهارم نسبت به سیستم ترکیبی سوم بهتر است، عکس این رفتار در داده‌های گروه اوّل با تعداد گره‌های لایه‌ی مخفی ۱۰ و ۳۰ عدد را می‌توان احتمالاً ناشی از کم‌بودن تعداد گره‌های لایه‌ی مخفی برای مدل کردن مرز تصمیم‌گیری و در نتیجه عدم بلوغ کامل سیستم در یادگیری مرز تصمیم دانست. نکته‌ی حایز اهمّیت، آن است که با افزایش تعداد گره‌های لایه‌ی مخفی شبکه‌ی عصبی از رفتار مشخصی پیروی نمی کند.



(ب) (الف)

شکل (۷): عملکرد شناسایی کننده GMM به عنوان پردازشگر پسین در حوزه‌ی امتیازها الف) ۴۵ ثانیه آموزش، ۷ ثانیه آزمون ب)۱۵ ثانیه آموزش، ۳ ثانیه آزمون (بالا نمودار min. DCF و پایین نمودار EER)

۴-۴- ترکیب امتیازها با شناسایی کننده‌ی *GMM*

سیستم پیشنهادی شناسایی کننده GMM سعی می‌کند، توزیع امتیازهای حاصل از GMM UBM یا SBM-SGMM را توسط GMM مدل کند. مشابه آزمایش‌های انجام شده در بخش قبل، چهارنوع شناسایی کننده‌ی GMM مورد آزمایش قرار گرفت که توصیف ورودی‌های آنها همانند ورودی‌های شبکه‌ی عصبی MLP می‌باشد. نتایج به‌دست آمده از انجام آزمایش‌ها روی گروه داده‌های اوّل و دوم در شکل (۷) آمده است. نتایج حاصل از آزمایش‌های انجام شده با سیستم پیشنهادی روی گروه داده‌های ۱ و گروه داده‌های ۲، عملکرد فوق‌العاده مناسب این سیستم را نسبت به شبکه‌ی عصبی در ترکیب با SBM-SGMM نشان می‌دهد. همان‌گونه که در نمودارها نشان داده شده است. با شناسایی کننده‌ی GMM مرتبه‌ی کمینه ۸ در اکثر سیستم‌های ترکیبی (به جز سیستم ترکیبی نوع اوّل برای داده‌های گروه دوم) عملکرد سیستم پیشنهادی بهتر از GMM UBM خواهد بود. در مقایسه با شبکه‌ی عصبی می‌توان گفت که با این شرایط بار محاسباتی سیستم پیشنهادی از سیستم ترکیبی SBM-SGMM-NN که توسط شیانگ پیشنهاد شده بود کمتر است. با بالا بردن مرتبه‌ی شناسایی کننده‌ی GMM و قبول بار محاسباتی بالاتر (قابل مقایسه با شبکه‌ی عصبی) عملکرد سیستم پیشنهادی نیز نسبت به سیستم ترکیبی SBM-SGMM و شبکه عصبی بهبود بیشتری می یابد.

۵- نتیجه‌گیری

در این مقاله ابتدا مدل GMM برای مدلسازی هویت گویندگان از روی نمونه‌های گفتار آنها معرفی شد. کارکرد GMM به‌عنوان یک مدل پارامتری که قابلیت مدل کردن هر توزیع اختیاری را دارد بررسی شد. روش امتیازدهی سریع

در مدل آمیزه‌های گوسی تطبیق یافته بحث شد و در ادامه مدل آمیزه‌های گوسی ساختاری که در سال ۲۰۰۳ کاربرد آنها در تصدیق هویت گوینده برای اولین بار ارایه شده بود، به‌عنوان روشی برای کاهش بار محاسباتی آمیزه‌های گوسی مورد بررسی قرار گفت. روش ساخت درخت برای تولید ساختار SBM، چگونگی تطبیق مدل SGMM گویندگان نسبت به SBM توسط الگوریتم MAP چندلایه‌ای مورد بحث قرار گرفت و ساختار بهینه استخراج گردید. کارکرد شبکه‌ی عصبی به‌عنوان یک ترکیب‌کننده‌ی امتیازهای خروجی سیستم‌های GMM-UBM و SBM-SGM تحلیل شد و درنهایت یک پردازش‌گر پسین جدید به‌منظور ترکیب امتیازها پیشنهاد گردید. سیستم پیشنهادی از لحاظ سرعت، قابل مقایسه با شبکه‌ی عصبی بوده و از لحاظ عملکرد بهتر از شبکه‌ی عصبی MLP عمل می‌نماید.

۶- مراجع

[۱] م. مندولکانی؛ م. لطفی‌زاد؛ «تشخیص هویت گوینده توسط کامپیوتر»، دومین کنفرانس مهندسی برق ایران؛ دانشگاه تربیت مدرس؛ تهران، ایران؛ ۱۳۷۳؛ صص ۳۵۳-۳۶۰.

[۲] م. ص. حدائق؛ م. لطفی‌زاد؛ «تصدیق هویت گوینده توسط کامپیوتر»، دومین کنفرانس مهندس برق ایران؛ دانشگاه تربیت مدرس؛ تهران، ایران؛ ۱۳۷۳؛ صص ۲۱۲-۲۲۱.

[۳] ا. صیادیان؛ ح. غفوری‌فرد؛ «استفاده از تغییرات دینامیکی ضرایب LSPF جهت کاهش خطای سیستمهای بازشناسی گوینده»؛ سومین کنفرانس مهندس برق ایران؛ دانشگاه علم و صنعت ایران؛ تهران ۱۳۷۴؛ صص ۲۰۷-۲۱۲

[۴] ح. مقصودلو؛ م. ر. نخعی؛ م. تبیانی؛ «سیستم تأیید هویت گوینده وابسته به متن با استفاده از روش کوانتیزاسیون برداری»؛ سومین کنفرانس مهندسی برق ایران؛ دانشگاه علم و صنعت ایران؛ تهران، ایران؛ ۱۳۷۴؛ صص. ۱۵۵-۱۶۲.

[۵] ج. شیخ زادگان ؛ "تعیین هویت گوینده به‌صورت مستقل از متن"؛ رساله دکترای مهندسی برق مخابرات؛ دانشکده فنی و مهندسی دانشگاه تربیت مدرس؛ بهمن ۱۳۷۴.

[۶] س. د. فیض‌آبادی؛ س. صدوقی؛ «سیستم تشخیص گوینده»؛ ششمین کنفرانس مهندسی برق ایران؛ دانشگاه صنعتی خواجه نصیرالدین طوسی؛ تهران، ایران؛ ۱۳۷۷؛ صص. ۳۶۹-۳۷۲.

[۷] م. م. همایون‌پور؛ ا. نجاری؛ «تصدیق هویت گوینده توسط تلفیق شبکه‌های عصبی و الگوریتم‌های ژنتیکی»؛ پنجمین کنفرانس بین‌المللی سالانه‌ی انجمن کامپیوتر ایران؛ دانشگاه شهید بهشتی؛ تهران، ایران، ۱۳۷۸؛ صص. ۲۵۷-۲۶۴.

[۸] ا. نوری وند ؛ "بازشناسی گوینده مستقل از متن بر اساس گفتار تلفنی توسط شبکه های عصبی"؛ پایان نامه کارشناسی ارشد مهندسی برق مخابرات؛ دانشکده فنی دانشگاه تهران؛ دی ۱۳۷۸.

[۹]. صیادیان؛ ک. بدیع؛ م. حکاک؛ م. ر. بیک‌زاده؛ «ارایه روش TSD-PGMM در بازشناسی گوینده مستقل از متن»؛ هشتمین کنفرانس مهندسی برق ایران؛ دانشگاه صنعتی اصفهان؛ اصفهان، ایران؛ ۱۳۷۹؛ صص. ۳۷۶-۳۸۲.

[۱۰]. م. ش. معین؛ ر. بوستانی؛ «مقایسه روش‌های، GMM و HMM, SVM به منظور بررسی هویت گوینده»؛ یازدهمین کنفرانس مهندسی برق ایران؛ ۱۳۸۳؛ دانشگاه شیراز؛ شیراز، ایران؛ صص. ۳۱۳-۳۱۹

[۱۱]. م. م. همایون‌پور؛ ج. کبودیان؛ به «تعیین و تصدیق هویت گوینده بر روی خط تلفن به کمک یک سیستم هیبرید مقاوم در برابر نویز و اثر انتقال کانال همراه با نرمالیزاسیون امتیازها»؛ مجله فنی و مهندسی مدرس؛ دانشکده فنی و مهندسی دانشگاه تربیت مدرس؛ ویژه‌نامه مهندسی برق ـ شماره شانزدهم، به ۱۳۸۳؛ صص. ۳۳-۴۸.

[12] D. A. Reynolds, A Gaussian mixture modeling approach to text independent speaker identification, Ph.D. thesis, Georgia Institute of Technology, Atlanta, Ga, USA, September 1992.

[13] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, “Speaker verification using adapted Gaussian mixture models,” Digital Signal Processing, vol. 10, pp. 19–41, 2000.

[14] R. Auckenthaler and J. Mason, "Gaussian selection applied to text-independent speaker verification," in *Proc. A Speaker Odyssey—Speaker Recognition Workshop*, 2001.

[15] B. Xiang and Toby Berger, "Efficient text-independent speaker verification with structural Gaussian mixture models and neural network," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 447-456, September 2003.

[16] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Transactions on Speech Audio Processing vol. 3, pp. 72-83, 1995.

[17] J. L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 291–298, Apr. 1994.

[18] Huo Q., Chan C., Lee C., "Bayesian Adaptive Learning of the Parameters of Hidden Markov Model for Speech Recognition", *IEEE Trans. Speech Audio Processing,* vol. 3, no. 5, pp. 334-345, 1995.

[19] K. Shinoda and C. H. Lee, "A structural Bayes approach to speaker adaptation", *IEEE Trans. Speech Audio Processing*, vol. 9, no. 3, pp. 276-287, March 2001.

[20] J. Navrátil, U. V. Chaudhari, and G. N. Ramaswamy, "Speaker verification using target and background dependent linear transforms and multi-system fusion," in *Proc. Eurospeech*, pp. 1389-1392, 2001.

[21] S. Raudys, *Statistical and Neural Classifiers: An Integrated Approach to Design*. New York: Springer, 2001.

[22] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," in *Proc. 5th European Conference on Speech Communication and Technology (Eurospeech '97)*, vol. 2, pp. 963–966, Rhodes, Greece, September 1997.

[23] F. Bimbot, J. Bonastre, C. Fredouille, G. Gravier, I. Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska, and D. A. Reynolds, "A Tutorial on Text-Independent Speaker Verification," *EURASIP Journal on Applied Signal Processing,* vol. 4, pp. 430–451, 2004.

[24] [Online] Available: http://www.nist.gov/speech/tests/spk/index.htm.

[25] A.Martin, G.Doddington, "the DET Curve in Assessment of Detection Task Performance", in *Proc. Eurospeech*, pp.1895-1898, 1997.

[26] R. Saeidi, H. R. Sadegh Mohammadi, and M. Khalaj Amir-Hosseini, "An efficient GMM classification post-processing method for structural Gaussian mixture model based speaker verification," in *Proc.ICASSP'06,* vol. 1, pp. 909-912, , Toulouse, France, May 2006.

[27] R. Saeidi, H. R. Sadegh Mohammadi, and M. Khalaj Amirhosseini "Study of model parameters effects in adapted Gaussian mixture models based text independent speaker verification", in *Proc. International Symposium of Telecommunications*, IST2005, vol. 1, pp. 387-392, Shiraz, Iran, 2005.

[28] R. Saeidi, H. R. Sadegh Mohammadi, and M. Khalaj Amirhosseini, "Efficient GMM-UBM system in text independent speaker verification using structural Gaussian mixture models," in *Proc. International Symp. of Telecommunications, IST2005,* vol. 1, pp. 39-44, Shiraz, Iran, Sept. 2005.

رحیم سعیدی مدرک کارشناسی خود را در رشته‌ی مهندسی برق-قدرت در سال ۱۳۸۱ از دانشگاه آزاد واحد ساوه و مدرک کارشناسی ارشد در رشته‌ی مهندسی برق - مخابرات گرایش سیستم را در سال ۱۳۸۴ از دانشگاه علم و صنعت ایران اخذ نموده است. وی در حال حاضر در پژوهشکده‌ی برق جهاد دانشگاهی به تحقیق در زمینه‌ی تشخیص مستقل از متن گوینده اشتغال دارد. زمینه‌های تحقیقاتی مورد علاقه‌ی وی پردازش سیگنال صحبت، تشخیص الگو و علوم اعصاب شناختی می باشد.

نشانی رایانامک (پست الکترونیکی) ایشان عبارت است از:
rahim.saeidi@gmail.com

حمیدرضا صادق‌محمّدی متولّد ۱۳۳۸ در تهران بوده و تحصیلات خود را در مقاطع کارشناسی مهندسی برق با گرایش مخابرات و کارشناسی ارشد مهندسی الکترونیک به‌ترتیب در سال‌های ۱۳۶۲ و ۱۳۶۶ از دانشگاه علم و صنعت ایران و در مقطع دکتری در رشته‌ی مهندسی برق با گرایش مخابرات را در سال ۱۳۷۵ از دانشگاه نیوساوت‌ولز استرالیا به پایان رساند. وی در سال ۱۳۶۰ به جهاد دانشگاهی پیوست و تاکنون مسئولیت‌های مختلفی را در این جهاد بر عهده داشته است و در حال حاضر به‌عنوان عضو هیئت علمی و رییس پژوهشکده‌ی برق جهاد دانشگاهی به فعالیت اشتغال دارد. زمینه‌های تحقیقاتی مورد علاقه‌ی ایشان عبارتند از: پردازش سیگنال، پردازش صحبت، شناسایی گوینده و الگوریتم‌های بهینه‌سازی. از دکتر صادق‌محمدی تاکنون بیش از ۳۰ مقاله در مجلات و همایش‌های معتبر داخلی و خارجی انتشار یافته است. ایشان از بدو تأسیس عهده‌دار مسئولیت سردبیری نشریه‌ی مهندسی برق و مهندسی کامپیوتر ایران از انتشارات پژوهشکده‌ی برق جهاد دانشگاهی بوده است.

نشانی رایانامک (پست الکترونیکی) ایشان عبارت است از:
h.sadegh@ijece.org