طبقه‌بندی شورایی تطبیقی برای تصدیق گوینده مستقل از متن

محمد هاشمی‌نژاد[۱]، دانشجوی دکتری؛ حسن فرسی[۲]، دانشیار؛ ناصر مهرشاد[۳]، دانشیار

۱- دانشکده مهندسی برق و کامپیوتر - دانشگاه بیرجند - خراسان جنوبی - ایران - mhashemi@birjand.ac.ir
۲- دانشکده مهندسی برق و کامپیوتر - دانشگاه بیرجند - خراسان جنوبی - ایران - hfarsi@birjand.ac.ir
۳- دانشکده مهندسی برق و کامپیوتر - دانشگاه بیرجند - خراسان جنوبی - ایران - nmehrshad@birjand.ac.ir

چکیده: این مقاله مسئله طبقه‌بندی شورایی را برای تصدیق گوینده مستقل از متن بررسی می‌کند. ازآنجاکه ممکن است یک طبقه‌بند از اطلاعات مختلف سیگنال گفتار بهره نبرد، استفاده از یک طبقه‌بند برای تصدیق گوینده ممکن است منجر به تصمیم قابل‌اعتمادی نشود. بنابراین بهترین سامانه‌های تصدیق گوینده از مجموعه‌ای از طبقه‌بندهای مکمل برای رسیدن به تصمیمات قابل‌اعتماد استفاده می‌کنند. در اکثر مطالعات اخیر که روی ترکیب طبقه‌بندها برای تصدیق گوینده انجام شده است، ترکیب خطی وزن‌داری از امتیاز طبقه‌بندهای خبره پایه برای رسیدن به امتیاز نهایی تصدیق استفاده می‌شود که وزن‌های این ترکیب با استفاده از روشی مانند رگرسیون لجستیک و در زمان آموزش به دست می‌آیند. در این تحقیقات مسائلی از قبیل همبستگی بین طبقه‌بندها و برتری برخی طبقه‌بندها برای برخی داده آزمون به‌خوبی در نظر گرفته نشده است. در این مقاله با استفاده از فرایند طراحی شورا و قاعده ترکیب بر اساس داده آزمون برای هر دو مسئله راه‌حلی ارائه می‌شود. بررسی‌های انجام‌شده ما روی دادگان ارزیابی تصدیق گوینده NIST 2004 نشان می‌دهد روش پیشنهادی در مقایسه با روش مبنای ترکیب تنک طبقه‌بندها کارایی مناسبی را دارد.

واژه‌های کلیدی: بازشناسی گوینده، تصدیق گوینده، طبقه‌بندی شورایی، طبقه‌بندی شورایی تطبیقی، رگرسیون لجستیک.

Adaptive Ensemble Classification for Speaker Verification

M. Hasheminejad[1], PhD Student; H. Farsi[2], Associate Professor; N. Mehrshad[3], Associate Professor

1- Faculty of Electrical and Computer Engineering, University of Birjand, Khorasan Jonoobi, Iran, Email: mhashemi@birand.ac.ir
2- Faculty of Electrical and Computer Engineering, University of Birjand, Khorasan Jonoobi, Iran, Email: hfarsi@birand.ac.ir
3- Faculty of Electrical and Computer Engineering, University of Birjand, Khorasan Jonoobi, Iran, Email: nmehrshad@birand.ac.ir

Abstract: This paper considers the ensemble classification for the text independent speaker verification issue. Using one classifier for the speaker verification may not result in dependable decision, because it may not exploit different characteristics of speech signal. Therefore, state-of-the-art speaker verification systems use an ensemble of classifiers for the verification. Most of the ensemble speaker verification systems use a weighted summation of the score of the individual expert classifiers to calculate the final score of the verification. The weights of this score fusion is obtained using a method, e.g. logistic regression, in the training phase. These works do not efficiently take into account issues such as correlation of classifiers and instance specific behavior of the base classifiers into account. In this paper a new solution is proposed for these two issues by using the process of ensemble design and combination rule based on training data. The obtained results on NIST 2004 speaker evaluation corpus show the effectiveness of the proposed methods in comparison to the sparse classifier fusion, as a baseline method.

Keywords: Speaker recognition, speaker verification, ensemble classification, adaptive ensemble classification, logistic regression.

تاریخ ارسال مقاله: ۱۳۹۴/۱۰/۱۲
تاریخ اصلاح مقاله: ۱۳۹۴/۱۲/۱۴
تاریخ پذیرش مقاله: ۱۳۹۵/۰۲/۰۹
نام نویسنده مسئول: حسن فرسی
نشانی نویسنده مسئول: ایران - بیرجند - انتهای بلوار دانشگاه - دانشگاه بیرجند - دانشکده مهندسی برق و کامپیوتر.

۱- مقدمه

مطالعات علمی نشان داده است که اطلاعات مختلفی در سیگنال صحبت وجود دارد که می‌تواند به بازشناسی گوینده کمک کند. بازشناسی گوینده یک فرایند تصمیم‌گیری در مورد هویت گوینده با استفاده از سیگنال گفتار فرد است. حوزه بازشناسی گوینده دارای دو شاخه اصلیِ تصدیق گوینده و شناسایی گوینده است. در تصدیق گوینده، ابتدا ادعا می‌شود که یک فایل صحبت ورودی متعلق به یکی از کلاس‌های ثبت‌شده است. سپس درست یا نادرست بودن این ادعا بررسی می‌شود. از سوی دیگر یک سامانه شناسایی گوینده، در ابتدا تعدادی گویندگان را به‌عنوان هدف ثبت می‌کند و در زمان آزمون هویت صاحب سیگنال صحبت ورودی را مشخص می‌کند. در تصدیق گوینده به شباهت گفتار ورودی و کلاس ثبت‌شده یک امتیاز داده می‌شود. درجایی که چند کلاس ثبت‌شده است، گفتار ورودی با همه کلاس‌ها مقایسه می‌شود و به تعداد کلاس‌ها امتیاز شباهت به دست می‌آید. در نتیجه یک سامانه تصدیق گوینده می‌تواند با مقایسه این امتیازها و تعریف یک حد آستانه برای آن‌ها، شناسایی گوینده را انجام دهد. این حد آستانه از روی داده‌های آموزشی به دست می‌آید. به‌علاوه معیارهای پیشرفته‌تری برای ارزیابی کارایی سامانه‌های تصدیق گوینده وجود دارد، اکثر تحقیقات این حوزه به تصدیق گوینده پرداخته‌اند.

برای استفاده از اطلاعات مختلف موجود در سیگنال صحبت در فرایند تصدیق گوینده، می‌توان از شورای طبقه‌بندها استفاده کرد. تحقیقات حوزه‌های مختلف بازشناسی نشان داده است که استفاده از ترکیب طبقه‌بندها کارایی این سامانه‌ها را افزایش می‌دهد. به‌عنوان‌مثال می‌توان به ترکیب روش‌های مختلف برای رسیدن به یک مدل بهینه توصیه‌گر برای تجارت الکترونیک در [۱] اشاره کرد. ترکیب طبقه‌بندها در مسائل دیگری از قبیل افزایش کارایی مدل‌های پیش‌بین جریان ترافیک [۲] و موارد دیگر نیز کارایی دارد. طبقه‌بندی شورایی از اهمیت زیادی در تصدیق گوینده برخوردار است. در طبقه‌بندی شورایی تعدادی طبقه‌بند پایه با هم ترکیب می‌شوند. این ترکیب در سطوح ویژگی، امتیاز و تصمیم انجام می‌شود [۳]. در ترکیب در سطح ویژگی، بردارهای مختلف ویژگی در کنار هم قرار داده می‌شوند و یک بردار ویژگی با طول بیش‌تر تشکیل می‌شود. ترکیب در سطح امتیاز شامل به دست آوردن امتیاز شباهت از طبقه‌بندهای پایه و ترکیب آن‌ها با استفاده از قواعد مناسب می‌شود. در ترکیب در سطح تصمیم، هرکدام از طبقه‌بندهای پایه تصمیم مجزایی در مورد تصدیق صاحب سیگنال صحبت ورودی می‌گیرند، سپس با استفاده از تلفیق منطقی تصمیمات، رأی اکثریت، یا روش مناسب دیگر تصمیم تلفیقی نهایی گرفته می‌شود. در این مقاله تمرکز ما بر روی ترکیب در سطح امتیاز است که در آن امتیاز نهایی جمع وزن‌داری از امتیازهای پایه است.

برخلاف اکثر کارهای طبقه‌بندی شورایی که از وزن‌های ثابتی برای ترکیب امتیازها استفاده می‌کنند [۴]، یا امتیاز نهایی میانگین ساده ریاضی امتیازهای پایه است [۵]، ما از ترکیبی استفاده می‌کنیم که در آن وزن‌های ترکیب برای هر داده آزمون به‌طور اختصاصی تعیین می‌شوند. علیرغم این‌که استفاده از وزن‌های دائمی برای همه داده‌های آزمون ممکن است در برخی مواقع اثربخش باشد، به دست آوردن وزن‌هایی بهینه که برای همه داده‌های آزمون مؤثر باشند، کاری بسیار مشکل است. در این روش‌ها، در مرحله آموزش، مجموعه‌ای از وزن‌های منحصربه‌فرد با استفاده از داده‌های آموزش به دست می‌آیند. در مراحل بعدی از این وزن‌ها برای تصدیق گوینده استفاده می‌شود. با توجه به‌این‌که ممکن است برخی طبقه‌بندها برای بعضی داده‌های آزمون مؤثر باشند و برای بعضی دیگر نباشند، احتمالاً وزن‌های به‌دست‌آمده قابل‌تعمیم و مؤثر برای همه داده‌های آزمون نخواهند بود. در این مقاله ما ضمن بررسی این مشکل، راه‌حلی برای در نظر گرفتن رفتار وابسته به نمونه طبقه‌بندها پیشنهاد خواهیم کرد. ما با الهام از مرجع‌های[۴، ۶ و ۷] به موارد زیر عمل خواهیم کرد:

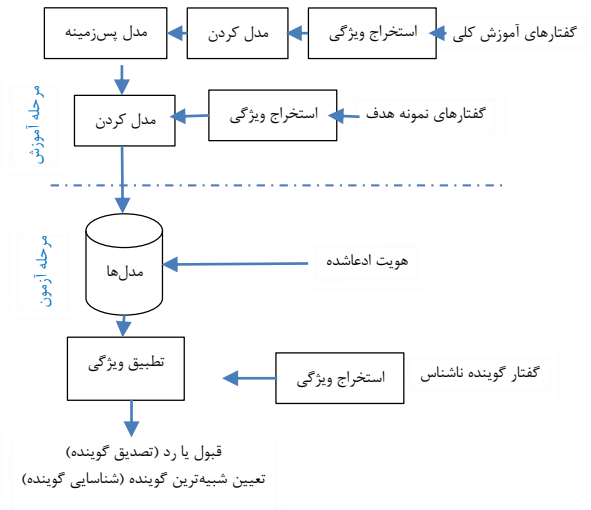
- وزن هرکدام از طبقه‌بندها را بر اساس نمونه‌های آزمون تعیین می‌کنیم. سپس با استفاده از وزن‌های به‌دست‌آمده امتیاز نهایی را از مجموع وزن‌دار همه امتیازهای پایه به دست خواهیم آورد.
- با استفاده از رفتار وابسته به نمونه آزمون یک شورای تنک از طبقه‌بندها را به دست خواهیم آورد؛ بنابراین، امتیاز نهایی جمع وزن‌داری از تعداد کمی از طبقه‌بندها خواهد بود.
- فرمول جدیدی را برای تعیین امتیاز نهایی معرفی خواهیم کرد.

در آزمایش‌های انجام‌شده رگرسیون لجستیک با عبارت تنظیم شبکه کشسان[۱] (ترکیب تنک) به‌عنوان روش مبنا، اثربخشی و قابلیت تعمیم روش پیشنهادی را نشان خواهیم داد.

۲- طبقه‌بندهای پایه

یک سامانه تصدیق گوینده شامل دو بخش آموزش و آزمون می‌شود. در بخش آموزش گویندگان هدف (سوژه‌ها) را به سامانه معرفی می‌کنیم. در بخش آزمون، گفتاری به‌عنوان ورودی به سامانه وارد می‌شود و ادعا می‌شود که به هرکدام از گویندگان هدف تعلق دارند و درستی این ادعا بررسی می‌شود. شکل ۱ نحوه عملکرد چنین سامانه‌ای را نشان می‌دهد.

روش‌های استخراج ویژگی گوینده، سیگنال صحبت اولیه را به یک نمایش فشرده تبدیل می‌کند. این روش‌ها تلاش دارند ویژگی‌هایی از صحبت را که مخصوص گوینده هستند در این نمایش فشرده، نگه دارند. دو نوع شناخته‌شده از این ویژگی‌ها، ویژگی‌های صوتی زمان کوتاه و ویژگی‌های عروضی[۲] [۸] هستند. از جمله ویژگی‌های زمان کوتاه صوتی می‌توان به ضرایب کپسترال مِل فرکانس MFCC[۳]، ضرایب پیش‌بینی خطی ادراکی PLP[۴]، ضرایب پیش‌بینی خطی وزن‌دار تثبیت‌شده SWLP[۵] [۹] و ضرایب کپسترال پیش‌بینی خطی LPCC[۶] [۸] اشاره کرد. ویژگی‌های MFCC یکی از پراستفاده‌ترین ویژگی‌ها در بازشناسی گوینده هستند. در [۱۰] ویژگی‌های MFCC و PLP توضیح داده شده است. در بخش تطبیق ویژگی (شکل ۱)، سامانه میزان شباهت ویژگی اهداف ثبت‌شده (الگوها) و ویژگی‌های داده آزمون را



شکل ۱: شمای کلی یک سامانه تصدیق گوینده

برحسب امتیاز تصدیق به دست می‌آورد. در انتها، امتیاز به‌دست‌آمده با یک حد آستانه مقایسه می‌شود. اگر این امتیاز بیش‌ازحد آستانه باشد جواب تصدیق، مثبت و در غیر این صورت منفی خواهد بود.

دو شرط اساسی برای کارایی خوب طبقه‌بندی شورایی، گوناگونی طبقه‌بندهای پایه و کارایی قابل‌قبول هرکدام از آن‌ها است. ما در آزمایش‌های انجام‌شده از سه طبقه‌بند شناخته‌شده با کارایی قابل‌قبول استفاده کردیم. این روش‌ها، GMM-SVM-KL [۱۱]، GMM-SVM-BHAT[۱۲] و ivector-PLAD [۱۳] هستند. روش‌های مبتنی بر بردارهای ماشین پشتیبان (SVM) در تصدیق گوینده مستقل از متن موفق بوده‌اند. در این روش‌ها، SVM روی ابر بردار مخلوط گوسی عمل می‌کند. این روش‌ها یک تابع کرنل را استفاده می‌کنند (دیورژانس کالبک-لیبلر[۷] (KL) و فاصله باچاریا[۸] (BHAT) که در این مقاله استفاده شده، جزو موفق‌ترین کرنل‌های SVM در تصدیق گوینده مستقل از متن بوده‌اند). سپس الگوریتم تصویر خصوصیات نوفه[۹] (NAP) به آن‌ها اعمال می‌شود.

روش‌های مبتنی بر بردار شناسایی[۱۰] (ivector) را تقریباً می‌توان موفق‌ترین روش‌ها در بازشناسی گوینده مستقل از متن نامید. این روش‌ها تعدادی فریم پشت سر هم را به یک بردار با طول ثابت تبدیل می‌کنند. این بردار نماینده کل صحبت یا به عبارتی گوینده است و می‌توان از آن به‌عنوان ورودی یک سامانه بازشناسی الگو استفاده کرد. ما برای امتیازدهی به شباهت بردارهای شناسایی از تحلیل تفکیک خطی آماری استفاده می‌کنیم.

۳- ترکیب امتیاز در تصدیق گوینده

به‌طورکلی روش‌های ترکیب طبقه‌بندها به سه گروه اصلی تقسیم‌بندی می‌شوند: ترکیب در سطح ویژگی، ترکیب در سطح امتیاز و ترکیب در سطح تصمیم. ترکیب در سطح ویژگی قبل از به‌کارگیری قالب تطبیق (شکل ۱) اتفاق می‌افتد. در این فرایند یک بردار ویژگی جدید تشکیل می‌شود. این بردار ویژگی از کنار هم قرار گرفتن بردارهای ویژگی که قبلاً استخراج شده شکل می‌گیرد. سپس تطبیق روی بردار ویژگی جدید انجام می‌شود. در ترکیب در سطح امتیاز که موضوع اصلی ما است، ابتدا تعدادی طبقه‌بند خبره برای به دست آوردن امتیازهای پایه شباهت الگوی از پیش ذخیره شده و گفتار ورودی به کار گرفته می‌شوند. تحقیقات نشان داده است که ترکیب در سطح امتیاز کارایی بهتری از ترکیب در سطح ویژگی دارد [۷]. در ترکیب در سطح تصمیم، تصمیم هر رد یا قبول هر خبره به‌عنوان ورودی به بخش تصمیم‌گیری نهایی داده می‌شود [۱۴].

در ترکیب در سطح امتیاز، امتیاز نهایی از ترکیب خطی یا به عبارتی جمع وزن‌دار امتیازهای پایه به دست می‌آید. علاوه بر آموزش طبقه‌بندهای پایه لازم است فرایند ترکیب امتیاز نیز آموزش ببینند. یکی از اولین راه‌هایی که برای این کار به ذهن می‌رسد این است که با استفاده از امتیازهای برچسب دار آموزشی، وزن ثابتی برای هر طبقه‌بند به دست آورد. برچسب امتیازهایی که در واقع مربوط به دو گفتار متعلق به یک گوینده هستند «صحیح» و آن‌هایی که در واقع مربوط به یک نفر نیستند «غیرصحیح» است. به‌طور قراردادی می‌توان «صحیح» را به‌صورت ۰ و «غیرصحیح» را به‌صورت ۱ نمایش داد.

اگر تعداد امتیازهای آموزشی n_{dev} باشد، امتیازها را با s_i و برچسب آن‌ها را با y_i نمایش بدهیم، نمایش ریاضی امتیازهای آموزشی به‌صورت: $\boldsymbol{D} = \{(s_i, y_i, i = 1, \dots, n_{dev})\}$ خواهد بود. چنین سامانه آموزشی نیاز به هیچ‌گونه اطلاعاتی در مورد نحوه آموزش طبقه‌بندهای پایه یا ویژگی‌های سیگنال صوتی ندارد. پس از انتخاب امتیازهای آموزشی یک روش مناسب آموزش برای کمینه کردن یک معیار خطا یا بیشینه کردن معیار کارایی به کار گرفته می‌شود. این بهینه‌سازی را می‌توان مستقیماً با استفاده از شبکه عصبی [۱۵]، الگوریتم‌های ابری [۱۶] یا رگرسیون لجستی [۴]، که کاربرد زیادی دارد به دست آورد.

۳-۱- ترکیب بر اساس رگرسیون لجستیک

بهترین سامانه‌های بازشناسی گوینده از تعدادی طبقه‌بندی برای رسیدن به نتایج قابل‌اعتماد بهره می‌گیرند. رگرسیون لجستیک یک روش تفکیک‌کننده[۱۱] [۱۷] است که معمولاً برای ترکیب امتیازها در تصدیق گوینده استفاده می‌شود. در این بخش به نحوه کارکرد این روش خواهیم پرداخت و دلیل استفاده زیاد آن در تصدیق گوینده و چگونگی ارتقاء آن در سال‌های اخیر را بررسی خواهیم کرد.

در مرحله آزمون یک سیستم تصدیق گوینده شورایی، ابتدا فرضیه تعلق سیگنال صوتی ورودی به هر کلاس از پیش ثبت‌شده به‌طور جداگانه توسط هرکدام از طبقه‌بندهای پایه بررسی می‌شود. خروجی هرکدام از طبقه‌بندهای پایه، یک امتیاز شباهت است. پس‌ازآن سامانه نیاز به یک روش ترکیب امتیاز دارد تا تصمیم نهایی تصدیق را بگیرد. در حقیقت کار بخش ترکیب امتیاز، یک نگاشت از فضای n بعدی به فضای دودویی یا $\{۰ ، ۱\}$ است. می‌توان این مسئله را به‌عنوان یک مسئله طبقه‌بندی دو کلاسی با بردار ورودی n بعدی در نظر گرفت.

برای استفاده از این بردار در بسیاری از روش‌ها، لازم است عناصر بردار از یک نوع باشند.

می‌توان بهترین وزن‌ها را با استفاده از روش brute force روی امتیازهای آموزشی به دست آورد. در این صورت ممکن است وزن‌های به‌دست‌آمده صرفاً برای امتیازهای آموزش معتبر باشند و قابلیت تعمیم نداشته باشند. ازآنجاکه ممکن است تفاوت زیادی بین امتیازهای آموزش و آزمون وجود داشته باشد، در گزارش‌های علمی توصیه شده است که از تخمین احتمال‌ها به‌عنوان امتیاز شباهت استفاده شود [۱۸]. برای رسیدن به این احتمال‌ها می‌توان از قانون بیز استفاده کرد [۱۷]. با استفاده از قانون بیز می‌توان با حداقل احتمال خطای طبقه‌بندی به این احتمال‌ها دست یافت. در ادامه توضیحات کلی در مورد تصمیم‌گیری بر اساس قانون بیز آمده است.

فرض کنید دو کلاس هدف[۱۲] (T) و غیرهدف[۱۳] (I) داشته باشیم. برای یک بردار تصادفی $\mathbf{X}$ که ممکن است متعلق به هرکدام از این کلاس‌ها باشد، هزینه طبقه‌بندی نمونه‌ای که در اصل متعلق به کلاس i است، به کلاس j می‌تواند یک تابع اتلاف ۰-۱ باشد (معادله (۱)).

$$C_{ij} = \begin{cases} 0 \;\; if \; i = j \\ 1 \;\; if \; i \neq j \end{cases} \tag{۱}$$

این معادله به طبقه‌بندی صحیح هزینه صفر و به طبقه‌بندی غلط هزینه یک را نسبت می‌دهد. با این فرضیه‌ها قانون بیز احتمال پسین کلاس i را به‌صورت معادله (۲) تعریف می‌کند.

$$P(c_i|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|c_i).P(c_i)}{P(\boldsymbol{X})} \tag{۲}$$

در این معادله $P(\boldsymbol{X})$ احتمال پیشین $\mathbf{X}$ و $P(c_i)$ احتمال پیشین c_i است. برای رسیدن به امتیازهای دقیق لازم است توزیع احتمال‌ها را به‌طور دقیق به دست آوریم. با استفاده از حدس‌هایی در مورد احتمال‌ها پیشین می‌توان معادله (۲) را به $P(\boldsymbol{X}|c_i)$ ساده کرد [۱۸]. اگر فرض کنیم طبقه‌بندهای مختلف مستقل از یکدیگر هستند $P(\boldsymbol{X}|c_i)$ منجر به معادله (۳) می‌شود.

$$P(\boldsymbol{X}|c_i) = P(s_1, \dots, s_K|c_i) = \prod_{k=1}^{K} P(s_k|c_i) \tag{۳}$$

در این معادله K تعداد طبقه‌بندهای پایه است.

ازآنجاکه امتیازهای کلاس T همبستگی دارند، به نظر می‌رسد فرض اخیر ما درباره استقلال طبقه‌بندها چندان درست نباشد. به این دلیل که اگر یک داده آزمون متعلق به یک کلاس باشد، امتیازهای کلاس T یا همان هدف برای همه طبقه‌بندهای پایه قوی، نزدیک به یک خواهد بود. منطقی‌تر آن است که فرض کنیم امتیاز s_k برای کلاس I و امتیاز $(1 - s_k)$ برای کلاس T مستقل هستند. احتمال پسین کلاس T را می‌توان از معادله (۴) به دست آورد [۱۸].

$$P(T|s_1, \dots, s_k) = \frac{1}{1 + e^{-\{(\sum_{k=1}^{K} x_k) + x_0\}}} \tag{۴}$$

در این معادله $x_0 = ln\frac{P(T)}{P(I)}$ و $x_k = ln\frac{P(s_k|T)}{P(s_k|I)}$ است. اگر فرض کنیم احتمال‌ها عضو خانواده نمایی هستند (معادلات (۵) و (۶)).

$$P(s_k|T) = f(s_k).e^{(C_k.s_k + C_{k0})} \tag{۵}$$

$$P(s_k|I) = f(s_k).e^{(C_k.s_k + C_{k0})} \tag{۶}$$

بنابراین معادله (۴) به رگرسیون لجستیک یا تابع توزیع لجستیک ساده می‌شود (معادله (۷)).

$$P(T|s_1, \dots, s_k) = \frac{1}{1 + e^{-g(s)}} = \pi \tag{۷}$$

که در آن $g(s)$ برابر معادله (۸) است.

$$g(s) = \beta_0 + \beta_1.s_1 + \cdots + \beta_K.s_K \tag{۸}$$

،

$$\beta_0 = \sum_{k=1}^{K} (C_{k0} - I_{k0}) + ln\frac{P(T)}{P(I)} \tag{۹}$$

و

$$\beta_K = C_k + I_k \tag{۱۰}$$

یک نوع خاص خانواده نمایی، توزیع گوسی است. اگر فرض کنیم توزیع کلاس‌ها گوسی است، معادلات (۹) و (۱۰) معادل معادلات (۱۱) و (۱۲) می‌شوند.

$$\beta_0 = \sum_{k=1}^{K} \frac{\left(\mu_k^I\right)^2 - \left(\mu_k^T\right)^2}{2\sigma_k^2} + ln\frac{P(T)}{P(I)} \tag{۱۱}$$

$$\beta_K = \frac{\mu_k^T - \mu_k^I}{\sigma_k^2} \tag{۱۲}$$

که در آن‌ها μ_k^T و μ_k^I به ترتیب میانگین توزیع کلاس‌های هدف و غیرهدف هستند و σ_k^2 واریانس مشترک است. نکته قابل‌توجه این روش، متناسب بودن وزن طبقه‌بند kام، β_K، با اختلاف میانگین توزیع هدف و غیرهدف است. همچنین اگر امتیازهای هدف برای یک طبقه‌بند پراکنده باشند، آن طبقه‌بند قابل‌اعتماد نیست و وزن کم‌تری دارد.

تا اینجا هدف ما این است که وزن‌های بهینه (β_K در معادله (۸)) را طوری پیدا کنیم که $P(T|s_1, \dots, s_k)$ مربوط به معادله (۷) بیشینه شود. برای حل این مسئله محققین با در نظر گرفتن نکات مختلف، تابع هزینه‌های مختلفی را معرفی کردند. یکی از جدیدترین توابع هزینه‌ای که اخیراً به این منظور معرفی‌شده $C_{wlr}(w, D)$ است [۴]. معادله (۱۳) این تابع هزینه را نشان می‌دهد.

$$C_{wlr}(w) = \frac{P_{eff}}{N_t} \sum_{i=1}^{N_t} log\left(1 + e^{-w^T s_i - logit\, P_{eff}}\right) + \frac{1 - P_{eff}}{N_f} \sum_{j=1}^{N_f} log\left(1 + e^{w^T s_j + logit\, P_{eff}}\right) \tag{۱۳}$$

در این معادله $P_{eff} = logit^{-1}\left(logit(P_{tar}) + log(C_{miss}/C_{fa})\right)$، به‌احتمال پیشین گوینده هدف (P_{tar})، هزینه طبقه‌بندی غلط (C_{miss}) و هزینه قبول غلط (C_{fa}) بستگی دارد. همان‌گونه که گفته شد، هدف از تعریف چنین تابع هزینه‌ای پیدا کردن وزن‌های بهینه‌ای

است که این تابع هزینه را کمینه کنند. معادله (۱۴) این مسئله بهینه‌سازی را به‌صورت فرمولی نمایش می‌دهد.

$$\boldsymbol{W}^* = \underset{\boldsymbol{W}}{argmin}\, C_{wlr}(\boldsymbol{W}) \quad (۱۴)$$

هاوتاماکی و همکارانش در [۴] نشان دادند که نوع تنظیم‌شده معادله (۱۴) که تعداد تُنُکی از طبقه‌بندها را انتخاب می‌کند، عملکرد بهتری دارد و مسئله را به‌صورت معادله (۱۵) معرفی کردند.

$$\boldsymbol{W}^* = \underset{\boldsymbol{W}}{argmin}\{C_{wlr}(\boldsymbol{W}) + \lambda(\alpha\|\boldsymbol{W}\|_1 + (1-\alpha)\|\boldsymbol{W}\|_2^2)\} \quad (۱۵)$$

در این معادله λ که ضریب لاگرانژ است، میزان کم شدن تعداد طبقه‌بندها را مشخص می‌کند. محدودیت $\|\boldsymbol{W}\|_1$ در این معادله به‌عنوان لسو[۱۴] شناخته می‌شود و $\|\boldsymbol{W}\|_2^2$ مربوط به رگرسیون ریج[۱۵] است. ترکیب این دو شبکه منعطف نام دارد. بخش لسو در شبکه منعطف باعث می‌شود اکثر وزن‌ها نزدیک صفر باشند. به عبارتی تنک بودن وزن‌ها را افزایش می‌دهد. بخش ریج باعث می‌شود وزن‌ها به‌شدت حالت صرفاً محدودیت لسو، به سمت صفر نروند. ضریب α میزان مشارکت لسو و ریج را در این معادله مشخص می‌کنند. این مسئله را می‌توان توسط الگوریتم آماده ProjectL1 [۱۹] حل رد.

اگرچه هدف این روش افزایش قابلیت تعمیم طبقه‌بندی شورایی است، برای هر طبقه‌بند یک وزن ثابت در نظر می‌گیرد. این وزن‌ها که از داده‌های آموزشی به دست می‌آیند، برای همه داده‌های آزمون یکسان هستند. این روش در مرحله آموزش طبقه‌بندی شورایی تعداد کمی از طبقه‌بندها را برای شورا انتخاب می‌کند و طبقه‌بندهای انتخاب‌نشده را کاملاً حذف می‌کند. ما در آزمایش‌های پیش رو نشان خواهیم داد برخی طبقه‌بندها که با این روش در مرحله آموزش حذف می‌شوند، برای برخی داده‌های آزمون عملکرد بهتری نسبت به برخی طبقه‌بندهای باقی‌مانده در شورا خواهند داشت. تحقیقات اخیر نشان داده است که طراحی طبقه‌بند شورایی با در نظر گرفتن عملکرد وابسته به نمونه آزمون طبقه‌بندهای پایه قابلیت تعمیم طبقه‌بند شورایی را افزایش می‌دهد [۶ و ۷].

۳-۲- وزن‌دهی بر اساس نمونه آزمون

وزن‌دهی بر اساس یک نمونه آزمون باید متناسب با توانایی تخمین امتیاز هر طبقه‌بند روی آن نمونه انجام شود. اگر در یک مقایسه در حقیقت دو طرف به یک گروه تعلق داشته باشند، امتیاز طبقه‌بند باید بالا باشد و اگر متعلق به یک گروه نباشند، امتیاز باید پایین باشد. اگر طبقه‌بندی در حالت تعلق دو طرف مقایسه به یک گروه امتیاز بالایی بدهد پس قابلیت اطمینان بالایی دارد و اگر امتیاز پایینی داشته باشد، قابلیت اطمینان کمی دارد. از سوی دیگر، برای یک مقایسه که دو طرف به یک گروه تعلق ندارد، اگر طبقه‌بندی امتیاز بالایی داشته باشد، قابلیت اطمینان کم و در صورت داشتن امتیاز کم قابلیت اطمینان بالایی دارد.

به دست آوردن وزن‌های خاص برای هر نمونه آزمون کار مشکلی است. با توجه به این‌که در حالت آزمون برچسب کلاس هر نمونه مشخص نیست، نمی‌دانیم تصمیم طبقه‌بند درست هست یا نه و درنتیجه نمی‌توانیم وزن‌های خاص هر نمونه آزمون را به دست آوریم.

در [۶] معیاری به نام شاخص وضوح[۱۶] تعریف شده که با استفاده از آن می‌توان وزن‌های مخصوصی برای هر طبقه‌بند به دست آورد. این معیار با استفاده از امتیاز آزمون به‌دست‌آمده و تعدادی از امتیازهای از پیش‌ذخیره‌شده برای هر طبقه‌بند محاسبه می‌شود. هر طبقه‌بند تعداد n_0 امتیاز هدف و n_1 امتیاز غیرهدف دارد که در زمان آموزش ذخیره شده‌اند. به امتیازهای هدف اصطلاحاً امتیازهای مثبت و به امتیازهای غیرهدف، منفی می‌گویند. شاخص وضوح به دو عامل بستگی دارد. عامل اول تلف مرتبط[۱۷] (RL) است و عامل دوم عامل تلف نامرتبط [۱۸] (IL). عامل RL موقعیت یک بردار امتیاز آزمون، S^{ts}، را نسبت به امتیازهای آموزشی منفی S_i^{ntr} مشخص می‌کند. معادله (۱۶) که در آن W بردار وزن، S^{ts} بردار امتیاز آزمون، n_n تعداد امتیازهای آموزشی منفی و U تابع پله واحد است، این عامل را تعریف می‌کند.

$$RL(\boldsymbol{S}^{ts}, \boldsymbol{W}) = \frac{1}{n_n}\sum_{i=1}^{n_n} U(\boldsymbol{W}^T\boldsymbol{S}_i^{ntr} - \boldsymbol{W}^T\boldsymbol{S}^{ts}) \quad (۱۶)$$

در حقیقت RL نسبت تقسیم تعدادی از امتیازهای آموزشی منفی به کل امتیازهای آموزشی منفی است. در نتیجه، این مقدار در بازه [۰ ، ۱] است. برای یک مقایسه که دو طرف به یک کلاس تعلق دارند، حالت ایده‌آل این است که امتیاز به‌دست‌آمده از همه امتیازهای آموزشی منفی بالاتر باشد. درنتیجه در حالت ایده‌آل این عامل بهتر است نزدیک به صفر باشد. برای حالتی که دو طرف مقایسه متعلق به یک کلاس نباشند، حالت ایده‌آل این است که امتیاز آزمون از همه امتیازهای آموزشی منفی کم‌تر باشد. درنتیجه این مقدار نزدیک به یک خواهد بود. عامل دوم یا IL، موقعیت بردار امتیاز آزمون را نسبت به امتیازهای آموزشی مثبت (S_i^{ptr}) مشخص می‌کند. معادله (۱۷) این عامل را تعریف می‌کند.

$$IL(\boldsymbol{S}^{ts}, \boldsymbol{W}) = \frac{1}{n_p}\sum_{j=1}^{n_p} U\left(\boldsymbol{W}^T\boldsymbol{S}^{ts} - \boldsymbol{W}^T\boldsymbol{S}_j^{ptr}\right) \quad (۱۷)$$

در این معادله n_p تعداد امتیازهای آموزشی مثبت است. برای مقایسه‌ای که دو طرف به یک کلاس تعلق دارند، در حالت ایده‌آل IL نزدیک یک است و برای حالتی که به یک کلاس تعلق ندارند، نزدیک به صفر است.

شاخص وضوح خام به‌صورت معادله (۱۸) تعریف می‌شود.

$$RCL(\boldsymbol{S}^{ts}, \boldsymbol{W}) = RL(\boldsymbol{S}^{ts}, \boldsymbol{W}) - IL(\boldsymbol{S}^{ts}, \boldsymbol{W}) \quad (۱۸)$$

قدر مطلق RCL شاخص وضوح نامیده می‌شود (معادله (۱۹)).

$$CL(\boldsymbol{S}^{ts}, \boldsymbol{W}) = |RL(\boldsymbol{S}^{ts}, \boldsymbol{W}) - IL(\boldsymbol{S}^{ts}, \boldsymbol{W})| \quad (۱۹)$$

همان‌گونه که گفته شد در صورت تعلق دو طرف مقایسه به یک گروه، در حالت ایده‌آل RL برابر با صفر و IL برابر یک است. درنتیجه CL برابر با یک می‌شود. برای حالتی که دو طرف به یک گروه تعلق ندارند هم CL برابر یک می‌شود. درنتیجه هر چه مقدار CL بیش‌تر

باشد طبقه‌بند مطمئن‌تر خواهد بود. بنابراین ما از این شاخص برای انتخاب اعضای شورای طبقه‌بندی و قاعده ترکیب استفاده می‌کنیم.

با استفاده از شاخص وضوح، سه مسئله حل خواهد شد. اولین مسئله کارایی متفاوت هر طبقه‌بند در برابر داده‌های متفاوت آزمون است. این مسئله بر هر دو موضوع انتخاب اعضای شورا و تعیین وزن اثر خواهد داشت. در این مقاله یک راه‌حل جدید برای این مسئله ارائه شده است. دومین مسئله انتخاب تعداد مناسب طبقه‌بند است و مسئله سوم همبستگی بین طبقه‌بندها. ممکن است برای نمونه‌های متفاوت آزمون تعداد متفاوتی از طبقه‌بندها مناسب باشد و ممکن است در شرایط مختلف همبستگی داشته یا نداشته باشند. ارائه یک راه‌حل برای این مسئله یکی دیگر از نوآوری‌های این مقاله است. با استفاده از شاخص وضوح و یک حد آستانه مناسب، ما می‌توانیم تعداد متغیر یا به عبارتی تطبیقی از طبقه‌بندها را انتخاب کنیم. درحالتی‌که CL برای همه طبقه‌بندها از مقدار آستانه پایین‌تر باشد، از تعداد کمینه از پیش تعریف‌شده‌ای استفاده می‌کنیم (به‌عنوان‌مثال ۸ طبقه‌بند برتر). مقدار حد آستانه و تعداد از پیش‌تعریف‌شده را می‌توان از روی امتیازهای آموزشی به دست آورد، به‌طوری‌که بهترین کارایی طبقه‌بندی، برای امتیازهای آموزشی به دست آید.

در زمان انتخاب اعضای شورا، برای محاسبه CL به بردار وزن نیاز داریم که به آن دسترسی نداریم. درنتیجه در این حالت مقدار یک را برای همه وزن‌ها در نظر می‌گیریم. ما برای انتخاب اعضای شورا دو راهکار در نظر گرفتیم. در راهکار اول، از یک مقدار ثابت حد آستانه برای CL استفاده می‌کنیم. طبقه‌بندهای با CL بالاتر از حد آستانه به‌عنوان اعضای شورا استفاده می‌شوند. در این راهکار تعداد اعضای شورا برای نمونه‌های مختلف آزمون متفاوت خواهد بود. درحالتی‌که همه طبقه‌بندها مقداری زیر حد آستانه دارند همه طبقه‌بندها برای عضویت در شورا انتخاب می‌شوند. در راهکار دوم، حد آستانه ثابت نیست و متناسب با مقدار CL تغییر خواهد کرد. لازم به ذکر است قبل از محاسبه CL باید مطمئن شد که کلیه امتیازهای از نوع میزان درست‌نمایی لگاریتمی [۲۰] هستند.

برای در نظر گرفتن کارایی وابسته به نمونه آزمون و قابلیت تعمیم رگرسیون لجستیک، ما استفاده از معادله (۲۰) را پیشنهاد می‌کنیم که در آن $\mathcal{F}$ تابعی بر اساس نمونه آزمون، وزن طبقه‌بندها و نمونه‌های آموزشی مثبت و منفی است.

$$\boldsymbol{W}^{*}=\underset{\boldsymbol{W}}{\operatorname{argmin}}\{C_{wlr}(\boldsymbol{W})+\lambda\mathcal{F}(\boldsymbol{S}^{ts},\boldsymbol{S}^{tr},\boldsymbol{W})\} \qquad (۲۰)$$

با توجه به این‌که RL و IL گسسته هستند CL مشتق‌پذیر نیست و اگر مستقیماً از CL در معادله (۲۰) استفاده کنیم، حل معادله با مشکل مواجه می‌شود. پس به‌جای فرمول اصلی RL و IL از مشابه سیگموئید آن‌ها استفاده می‌کنیم (معادله‌های (۲۱) و (۲۲)).

$$RL(\boldsymbol{S}^{ts},\boldsymbol{W})=\frac{1}{n_n}\sum_{i=1}^{n_n}\frac{1}{1+e^{-\alpha\boldsymbol{W}^{T}(\boldsymbol{S}_i^{ntr}-\boldsymbol{S}^{ts})}} \qquad (۲۱)$$

$$IL(\boldsymbol{S}^{ts},\boldsymbol{W})=\frac{1}{n_p}\sum_{j=1}^{n_p}\frac{1}{1+e^{-\beta\boldsymbol{W}^{T}(\boldsymbol{S}^{ts}-\boldsymbol{S}_j^{ptr})}} \qquad (۲۲)$$

با استفاده از مقدار مناسب α و β، این معادلات می‌توانند به مقدار اولیه RL و IL نزدیک باشند. انتخاب مقدار بزرگ برای این دو پارامتر، باعث تخمین دقیق‌تر RL و IL، درعین‌حال باعث ایجاد کمینه‌های محلی زیاد برای CL می‌شود. از سوی دیگر انتخاب مقدار کم برای این پارامترها باعث تخمین ضعیف CL می‌شود. درنتیجه این دو مقدار تأثیر قابل‌توجهی روی کارایی طبقه‌بندی دارند. حتی با این تغییر هم نمی‌توانیم CL را در معادله جایگزین کنیم، چون در صفر مشتق‌پذیر نیست. برای حل این مشکل از رگرسیون ریج [۱۷] شاخص وضوح خام (RCL) استفاده می‌کنیم (معادله (۲۳)).

$$\boldsymbol{W}^{*}=\underset{\boldsymbol{W}}{\operatorname{argmin}}\{C_{wlr}(\boldsymbol{W})-\lambda\|RCL\|_2^2\} \qquad (۲۳)$$

گرچه در اینجا معادله بهینه‌سازی در مرحله آزمون انجام می‌شود، با استفاده منابع و کدهای آزاد موجود در کسری از ثانیه می‌توان به نقاط بهینه دست یافت. این مسئله را می‌توان با استفاده از بسته‌های نرم‌افزاری آماده [۱۷] حل کرد. برای بهینه‌سازی سریع‌تر می‌توان در زمان آموزش ابتدا وزن‌های بهینه را برای همه ترکیب‌های اعضای شورا به دست آورد. سپس در زمان آزمون، پس از انتخاب اعضای شورا با استفاده از CL، وزن‌های بهینه مربوط به آن ترکیب اعضا، استفاده شوند. اگر تعداد k طبقه‌بند داشته باشیم، تعداد حالت‌های ممکن برابر $\sum_{i=1}^{k}\frac{k!}{(k-i)!i!}$ خواهند بود. این تعداد برای ۱۲ طبقه‌بند برابر با ۴۰۹۶ حالت خواهد بود. سپس با حل معادله (۲۴) برای هرکدام از حالت‌ها، بردار وزن ذخیره به دست خواهد آمد.

$$\boldsymbol{W}^{*}=\underset{\boldsymbol{W}}{\operatorname{argmin}}\{C_{wlr}(\boldsymbol{W})-\lambda\|\boldsymbol{W}\|_2^2\} \qquad (۲۴)$$

تنظیم ریج در این معادله، اندازه وزن‌ها را کوچک نگه می‌دارد. درنهایت پس از انتخاب اعضای قابل‌اعتماد و وزن‌های مناسب، با ترکیب امتیازهای قابل‌اعتماد پایه، امتیاز نهایی به دست می‌آید.

۴- آزمایش‌ها

۴-۱- پایگاه داده

در این تحقیق از پایگاه داده NIST SRE 2004 و Switchboard II [۲۱] برای بررسی روش پیشنهادی استفاده شد. ازآنجاکه در این تحقیق از طبقه‌بندهای زیادی با ویژگی‌های متفاوتی استفاده شده و هر طبقه‌بند قادر است ویژگی‌های خاصی از سیگنال صحبت را آشکار کند، ما ترجیح دادیم داده‌ها را تقسیم‌بندی نکنیم و از کل پایگاه داده (زبان‌های مختلف، کانال‌های مختلف و موارد دیگر از قبیل جنسیت جداسازی نشدند). در بخش‌های آموزش و آزمون استفاده کنیم. پایگاه داده NIST 2004 شامل ۶۲۴۴ مکالمه آموزشی می‌شود. مدل پس‌زمینه کلی[۱۹] (UBM) با استفاده از این داده‌ها آموزش داده شد. این پایگاه داده در بخش آزمون شامل ۶۶۰ گوینده زن و مرد و به‌طورکلی ۴۶۲۳ فایل صوتی می‌شود. صحبت‌های موجود در این پایگاه داده شامل پنج زبان عربی، انگلیسی، چینی، روسی و اسپانیولی می‌شود. با

توجه به این‌که ما برای آموزش هر گوینده هدف به یک فایل صوتی، تعدادی فایل برای رسیدن به امتیازهای مثبت و منفی نیاز داشتیم گوینده‌هایی را برای بخش آموزش طبقه‌بندها استفاده کردیم که حداقل ۱۱ فایل صوتی داشته باشند. به‌علاوه برای آموزش NAP، PLDA و λ از پایگاه داده Switchboard II استفاده کردیم.

۴-۲- طراحی آزمایش‌ها

در طبقه‌بندی شورایی اعتقاد بر این است که گوناگونی طبقه‌بندهای پایه کارایی طبقه‌بندی شورایی را افزایش می‌دهد [۲۲]. به‌علاوه طبقه‌بندهای پایه باید به‌اندازه کافی کارا باشند. درنتیجه آزمایش‌های ما با استفاده از سه روش معروف در بازشناسی گوینده مستقل از متن و چهار نوع متفاوت ویژگی انجام شد. ویژگی‌های استفاده‌شده MFCC، PLCC، PLP و SWLP هستند. نتایج به‌دست‌آمده با روش مبنای ترکیب تنک طبقه‌بندها که در [۴] معرفی شده، مقایسه شده است.

معمولاً در سیگنال صحبت بخش‌های سکوت وجود دارد که ارتباطی با ویژگی‌های گوینده ندارند. در نتیجه، در ابتدا برای حذف بخش‌های سکوت، یک الگوریتم تشخیص صحبت با استفاده از انرژی فریم اعمال شد. این باعث افزایش کارایی (یا به عبارتی کاهش زمان محاسبات) طبقه‌بندی می‌شود. سپس ویژگی‌ها با استفاده از پنجره همینگ ۲۵ میلی‌ثانیه‌ای با ۵۰٪ همپوشانی (۱۲/۵ میلی‌ثانیه) استخراج شدند. برای استخراج ویژگی‌های MFCC، PLCC و PLP از بسته نرم‌افزاری HTK [۲۳] استخراج شد. همچنین برای استخراج ویژگی‌های SWLP که در [۲۴] معرفی‌شده‌اند از کدهای MATLAB قابل‌دسترسی در اینترنت[۲۰] استفاده شدند. کلیه ویژگی‌ها با ابعاد ۱۴ استخراج شد. همچنین ابعاد UBM استفاده‌شده در این آزمایش‌های ۲۰۴۸ انتخاب شد. برای استخراج مدل ivector و GMM از بسته نرم‌افزار MSR Toolkit [۲۵] استفاده شد. ابعاد هر ivector ۴۰۰ است که با استفاده از LDA به ۲۰۰ کاهش می‌یابد. یک نکته مهم در ترکیب امتیاز این است که با توجه به تفاوت ویژگی‌ها و روش‌های استفاده‌شده، ممکن است امتیازهای بسیار متفاوت باشند. با استفاده از نتایج تحقیقات انجام‌شده [۴] ما برای پیش تنظیم امتیازها از z-cal(Clipped) استفاده کردیم. کارایی طبقه‌بندهای پایه و شورای طبقه‌بندی با استفاده از بسته نرم‌افزاری BOSARIS Toolkit [۲۶] و معیارهای EER، minDCF و CLLR بررسی شدند.

۴-۳- نتایج

در این بخش ما از سه روش برای ترکیب طبقه‌بندها استفاده کردیم. در اولین روش از تابع هزینه C_{wlr} تنظیم‌شده با شبکه کشسان ($\alpha =$۰/۱) [۴] استفاده کردیم. در روش دوم (ترکیب تطبیقی ۱) عبارت تنظیم را با عبارت پیشنهادی خود در معادله (۲۳) جایگزین کردیم. در این روش بهینه‌سازی روی هر داده آزمون انجام می‌شود. در سومین روش (ترکیب تطبیقی ۲) بردار وزن برای همه حالت‌های ممکن برای ترکیب اعضای شورا محاسبه شد. به‌صورت تجربی ما به این نتیجه رسیدیم که اگر تعداد اعضای شورا بین چهار تا هشت طبقه‌بند محدود شود، بهترین نتایج حاصل می‌شود. در این آزمایش، مناسب‌ترین طبقه‌بندها برای هر داده آزمون انتخاب می‌شود و بهینه‌سازی وزن‌ها در مرحله آزمون انجام نمی‌شود. جدول ۱ نتایج اعمال ۱۲ طبقه‌بند پایه روی داده‌های آموزشی را نشان می‌دهد. این کار برای به دست آوردن امتیازهای آموزشی انجام شده است.

پس از به دست آوردن امتیازهای آموزشی داده‌های آزمون برای طبقه‌بندی شورایی استفاده شدند که نتایج آن در جدول ۲ قابل‌مشاهده است. تفاوت عملکرد طبقه‌بندها در دو جدول ۱ و ۲ نشان‌دهنده وابستگی عملکرد طبقه‌بندها به نمونه آزمون است. چراکه تنها تفاوت آزمایش‌های مرتبط با این دو جدول در داده‌ها است. به‌عنوان‌مثال ivector-PLDA که از ویژگی‌های PLP استفاده می‌کند دارای بهترین EER در جدول ۱ است، درحالی‌که این وضعیت برای جدول ۲ صادق نیست. دلیل بعدی عملکرد کلی GMM-SVM-KL است که در جدول ۱ خوب است ولی در جدول ۲ به‌جز برای ویژگی SWLP بدترین عملکرد را از نظر EER دارد. همچنین مقایسه عملکرد GMM-SVM-KL با ویژگی LPCC و ivector با ویژگی‌های MFCC نیز نشان‌دهنده همین نکته است.

به‌عنوان مثالی از تقویت بازشناسی با استفاده از روش پیشنهادی می‌توان به امتیاز مثبت طبقه‌بندی فایل صحبت '*xalm.sph*' در پایگاه داده NIST SRE 2004 اشاره کرد. امتیاز نهایی این طبقه‌بندی با استفاده از روش مبنای ترکیب تنک [۴] برابر با ۵/۱۴۵۸ و با استفاده از روش پیشنهادی دوم که در آن شش طبقه‌بند اول، سوم، هفتم، دهم، یازدهم و دوازدهم انتخاب می‌شوند ۷/۲۷۰۶ به دست می‌آید. چون در اینجا دو طرف مقایسه به یک گروه تعلق دارند، امتیاز شباهت بیش‌تر نشان از عملکرد بهتر در این داده است.

جدول ۱: نتایج پیاده‌سازی ۱۲ طبقه‌بند پایه برای به دست آوردن امتیازهای پایه

ردیف	طبقه‌بند	ویژگی	EER (%)	MinDCF ×100
۱	Ivector-PLDA	MFCC	۶/۶۸	۴/۶۸
۲	Ivector-PLDA	LPCC	۷/۲۵	۵/۷۴
۳	Ivector-PLDA	PLP	۶/۵۹	۵/۰۶
۴	Ivector-PLDA	SWLP	۹/۱۲	۸/۵۹
۵	GMM-SVM-BHAT	MFCC	۷/۲۳	۷/۲۳
۶	GMM-SVM-BHAT	LPCC	۸/۳۵	۶/۰۱
۷	GMM-SVM-BHAT	PLP	۸/۱۵	۶/۶۷
۸	GMM-SVM-BHAT	SWLP	۱۰/۵۴	۸/۱۹
۹	GMM-SVM-KL	MFCC	۷/۴۴	۵/۵۳
۱۰	GMM-SVM-KL	LPCC	۶/۶۶	۴/۶۹
۱۱	GMM-SVM-KL	PLP	۷/۴۵	۶/۴۲
۱۲	GMM-SVM-KL	SWLP	۷/۸۸	۵/۵۶

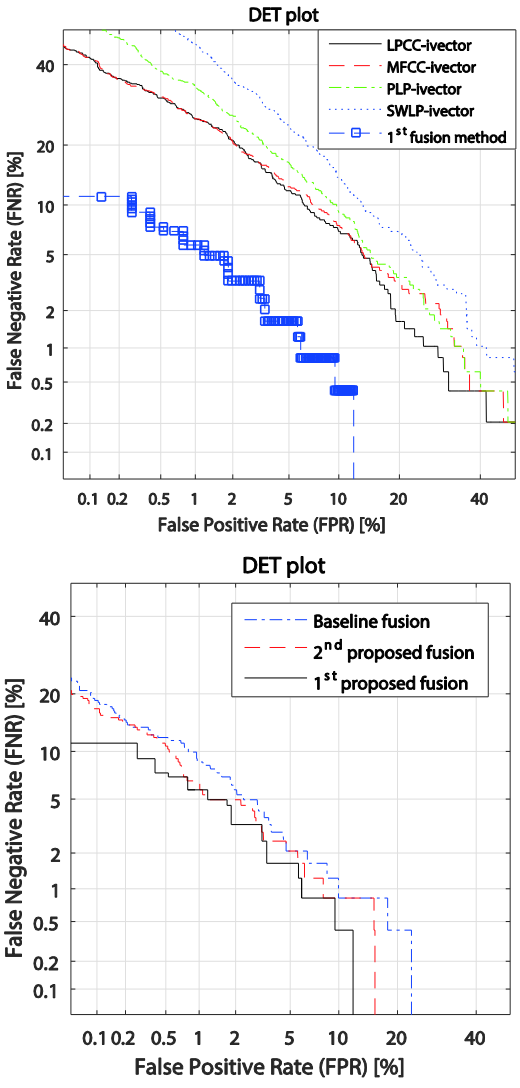
همان‌گونه که گفته شد، برای استفاده از CL باید از یک حد آستانه استفاده شود. مقادیر بالاتر از این حد، به‌عنوان اعضای شورا انتخاب می‌شوند. این حد از داده‌های آموزشی انتخاب و برای طبقه‌بندی داده‌های آزمون استفاده می‌شود. در شکل ۲ روش‌های پیشنهادی ترکیب طبقه‌بند با روش ترکیب تنک طبقه‌بندها مقایسه شده است. این منحنی‌ها با استفاده از بسته نرم‌افزاری MSR Toolbox به‌دست‌آمده است. همان‌گونه که از شکل ۲ مشخص است، روش اول پیشنهادی (ترکیب تطبیقی ۱) تقریباً در همه‌جا خطای کم‌تری دارد.

جدول ۲: نتیجه پیاده‌سازی طبقه‌بندهای پایه و ترکیب طبقه‌بندی روی داده‌های آزمون

ردیف	طبقه‌بند	ویژگی	EER (%)	MinDCF ×100
۱	Ivector-PLDA	MFCC	۸/۰۱	۵/۵۹
۲	Ivector-PLDA	LPCC	۵/۷۹	۴/۸۵
۳	Ivector-PLDA	PLP	۷/۷۶	۵/۲۷
۴	Ivector-PLDA	SWLP	۱۰/۶۵	۸/۳۷
۵	GMM-SVM-BHAT	MFCC	۸/۱۲	۷/۷۶
۶	GMM-SVM-BHAT	LPCC	۷/۵۶	۶/۱۶
۷	GMM-SVM-BHAT	PLP	۷/۷۱	۶/۳۰
۸	GMM-SVM-BHAT	SWLP	۱۱/۰۵	۸/۴۸
۹	GMM-SVM-KL	MFCC	۹/۱۲	۶/۷۱
۱۰	GMM-SVM-KL	LPCC	۸/۴۱	۵/۶۷
۱۱	GMM-SVM-KL	PLP	۸/۳۷	۶/۹۱
۱۲	GMM-SVM-KL	SWLP	۹/۴۷	۶/۵۷
۱۳	ترکیب تنک	_	۳/۳۷	۳/۰۲
۱۴	ترکیب تطبیقی ۱	_	۲/۵۶	۲/۲۶
۱۵	ترکیب تطبیقی ۲	_	۲/۸۹	۲/۲۵

۴-۴- همبستگی طبقه‌بندها

گوناگونی طبقه‌بندها یک نکته مهم در طبقه‌بندی شورایی است. گوناگونی بیش‌تر طبقه‌بندها باعث می‌شود جنبه‌های مختلف سیگنال در طبقه‌بندی در نظر گرفته شود. همبستگی نقطه مقابل گوناگونی است. اگر دو طبقه‌بند دارای همبستگی بالایی باشند، می‌توان بدون از دست دادن میزان قابل‌توجهی اطلاعات به‌جای هر دو آن‌ها از یکی استفاده کرد. در روش پیشنهادی، اثر تضعیف‌کننده همبستگی طبقه‌بندها به‌طور خودکار حذف می‌شود. در این روش با انتخاب حد آستانه مناسب، صرفاً طبقه‌بندهایی که غیرقابل‌اعتماد هستند از مجموعه حذف می‌شوند. بااین‌حال، طبقه‌بندهای باقی‌مانده آن‌قدر قابل‌اعتماد هستند که در مجموعه شورا باقی بمانند. سؤالی که اینجا پیش می‌آید این است که اگر دو طبقه‌بند همبستگی زیادی داشته باشند چه اتفاقی می‌افتد. در صورت وجود چنین حالتی ممکن است وجود دو طبقه‌بند مشابه سهم مشارکت بقیه طبقه‌بندها را در مجموعه تحت تأثیر قرار دهند. این تأثیر منفی وقتی اتفاق می‌افتد که وزن طبقه‌بندها در همه حال ثابت باشند. این اثر مادامی‌که آموزش وزن‌ها در زمان آموزش انجام شود باقی خواهد ماند. ولی زمانی که آموزش وزن‌ها در زمان آزمون و متناسب با نمونه آزمون انجام شود، فرایند آموزش وزن‌ها مؤثرترین وزن‌ها را خواهد داد.



شکل ۲: بالا) مقایسه روش‌های پایه بر اساس ivector با ویژگی‌های MFCC، LPCC، PLP، SWLP و روش ترکیب پیشنهادی اول. پایین) مقایسه دو روش پیشنهادی با روش Sparse fusion (روش مبنا)

۵- نتیجه‌گیری

در این تحقیق روشی پیشنهاد گردید که در آن اعضای شورای طبقه‌بندی گوینده و قاعده ترکیب آن‌ها متناسب با نمونه آزمون انجام می‌شود. در این روش تعداد تطبیقی از بهترین طبقه‌بندها انتخاب

می‌شود و وزن آن‌ها متناسب با نمونه آزمون محاسبه می‌شود. این روش از تعدادی طبقه‌بند شناخته‌شده با کارایی نسبتاً خوب برای تصدیق گوینده استفاده می‌کند و آن‌ها را با استفاده از امتیازهای برچسب دار از پیش‌ذخیره شده و تک امتیاز آزمون رتبه‌بندی می‌کند. فرایند محاسبه وزن از رگرسیون لجستیک استفاده می‌کند و مسئله بهینه‌سازی با یک عبارت وابسته به نمونه آزمون محدود می‌شود.

آزمایش‌ها بر روی داده‌های گوناگون NIST 2004 اثربخشی این روش را نشان می‌دهد. به‌عنوان ادامه این تحقیق، کار روی مقدار ضریب لاگرانژ معادله (۲۳) (λ) و زمان‌بندی آموزش وزن‌ها پیشنهاد می‌شود.

مراجع

[۱] مصطفی رجب‌زاده و رضا رافع، «ارائه یک سیستم توصیه‌گر ترکیبی برای تجارت الکترونیک،» *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۵، شماره ۴، صفحات ۸۵-۹۱، ۱۳۹۴.

[۲] سیدهادی حسینی، بابک نجار اعرابی، بهزاد مشیری و اشکان رحیمی‌کیان، «الگوریتم ترکیب فازی مدل‌های پیش‌بین جریان ترافیک در حضور داده‌های اغتشاشی،» *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۶، شماره ۱، ۱۳۹۵.

[3] X. Zhou, A. S. d'Avila Garcez, H. Ali, S. N. Tran, and K. Iqbal, "Unimodal late fusion for NIST i-vector challenge on speaker detection," *Electronics Letters*, vol. 50, no. 15, pp. 1098–1100, 2014.

[4] V. Hautamäki, T. Kinnunen, F. Sedlák, K. A. Lee, B. Ma, and H. Li., "Sparse classifier fusion for speaker verification," *IEEE Transactions on Audio, Speech Language Processing*, vol. 21, no. 8, pp. 1622–1631, 2013.

[5] Z. Lei, Y. Yang, and Z. Wu, "Ensemble of support vector machine for text-independent speaker recognition," *International Journal of Computer Science and Network Security*, vol. 6, no. 5, pp. 163–167, 2006.

[6] A. Kumar, and B. Raj, "Unsupervised fusion weight learning in multiple classifier systems," *arXiv preprint arXiv: 1502.01823*, 2015.

[7] K. Lai, D. Liu, S. Chang, and M. Chen, "Learning sample specific weights for late fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2772–2783, 2015.

[8] Farrús Cabeceran, *Fusing Prosodic and Acoustic Information for Robust Speaker Recognition*, Ph.D. Thesis, University of Politècnica de Catalunya, 2008.

[9] R. Saeidi, J. Pohjalainen, T. Kinnunen, and P. Alku, "Temporally weighted linear prediction features for tackling additive noise in speaker verification," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 599–602, 2010.

[10] M. J. Alam, T. Kinnunen, P. Kenny, P. Ouellet, and D. O'Shaughnessy, "Multitaper MFCC and PLP features for speaker verification using i-vectors," *Speech Communication*, vol. 55, no. 2, pp. 237–251, 2013.

[11] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," *IEEE International Conference on. Acoustic, Speech Signal Processing*, vol. 1, p. I, 2006.

[12] C. You, K. A. Lee, and H. Li, "GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1300–1312, 2010.

[13] P. Mat, M. Kara, and P. Kenny, "Full-covariance UBM and heavy-tailed plda in i-vector speaker verification," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4516–4519, 2011.

[14] P. Emerson, *Designing an All-Inclusive Democracy: Consensual Voting Procedures for Use in Parliaments, Councils and Committees*, Springer Science & Business Media, 2007.

[15] W. M. Campbell, D. E. Sturim, W. Shen, D. A. Reynolds, and J. Navrátil, "The MIT-LL/IBM 2006 speaker recognition system: High-performance reduced-complexity recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, 2007.

[16] S. Chernbumroong, S. Cang, and H. Yu, "Genetic algorithm-based classifiers fusion for multi-sensor activity recognition of elderly people," *IEEE J. Biomedical Health Informatics*, vol. 19, no. 1, pp. 282–289, 2014.

[17] C. M. Bishop, *Pattern recognition and machine learning (Information Science and Statistics)*, Springer, 2007.

[18] S. Pigeon, P. Druyts, and P. Verlinde, "Applying logistic regression to the fusion of the NIST'99 1-speaker submissions," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 237–248, 2000.

[19] M. Schmidt, G. Fung, and R. Rosales, "Fast optimization methods for L1 regularization: A comparative study and two new approaches," *Lecture Notes in Computer Science*, vol. 4701, pp. 286–297, 2007.

[20] N. Brümmer, *Measuring, refining and calibrating speaker and language information extracted from speech*, Ph.D. Thesis, University of Stellenbosch, 2010.

[21] National Institute of Standards and Technology, *The evaluation plan of NIST 2004 Speaker Recognition evaluation campaign*, [Online], Available online at: http://www.nist.gov/speech/tests/spk/2004/SRE04evalplan-v1a.pdf, 2004.

[22] S. Wang, and X. Yao, "Relationships between diversity of classification ensembles and single-class performance measures," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 206–219, 2013.

[23] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, "Hidden markov model toolkit (htk) version 3.4 user's guide," *Cambridge University Engineering Department*, Cambridge, MA, 2002.

[24] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilised weighted linear prediction," *Speech Communication*, vol. 51, no. 5, pp. 401–411, 2009.

[25] S. O. Sadjadi, M. Slaney, and L. Heck, "MSR identity toolbox v1. 0: A matlab toolbox for speaker-recognition research," *Speech and Language Processing Technical Committee Newsletter*, 2013.

[26] N. Brümmer, and E. de Villiers, "Bosaris toolkit [software package]," [Online], Available online at: https://sites.google.com/site/bosaristoolkit, 2011.

زیرنویس‌ها

[1] Elastic-net
[2] Prosodic
[3] Mel-Frequency Cepstral Coefficients
[4] Perceptual linear predictive
[5] Stabilized Weighted Linear Predictive
[6] Linear Predictive Cepstral Coefficients
[7] Kullback-Leibler
[8] Bhattacharya
[9] Nuisance attribute projection
[10] Identity-vector
[11] Discriminating
[12] Target
[13] Impostor
[14] LASSO
[15] Ridge
[16] Clarity index
[17] Relevance loss
[18] Irrelevance loss
[19] Universal background model
[20] http://users.spa.aalto.fi/jpohjala/xlp/