بهبود سرعت و دقت در استفاده از برنامه‌نویسی ژنتیک برای تصدیق هویت گوینده

سعیده سادات سدیدپور، محمدمهدی همایونپور و مهدی فسنقری

چکیده: در تصدیق هویت گوینده، سیستم هویت شخصی را که با سیستم تماس برقرار کرده است، بررسی کرده و تعیین می‌کند که وی همان شخص مدعی است و یا دروغ می‌گوید. در این مقاله از برنامه‌نویسی ژنتیک به‌عنوان روشی برای مدل‌سازی گویندگان استفاده شد. با توجه به زمان زیاد آموزش مدل‌ها توسط برنامه‌نویسی ژنتیک، ایده بهره‌مندی از فشرده‌سازی داده‌های آموزشی، به‌منظور کاهش زمان آموزش مدل‌ها مطرح گردید و بدین ترتیب زمان لازم برای مدل‌سازی گویندگان با استفاده از برنامه‌نویسی ژنتیک در حدود ۲۰ برابر کاهش داده شد. آموزش چندین درخت برنامه‌نویسی ژنتیک به‌عنوان مدل هر گوینده، ایده دیگری است که به‌منظور بهبود دقت تصدیق هویت گوینده در این مقاله مطرح شده است. در این روش، داده‌های آموزشی به تعداد کمی خوشه تفکیک شده و به‌ازای هر خوشه، یک درخت برنامه‌نویسی ژنتیک آموزش داده می‌شود. بدین ترتیب یک گوینده با چندین درخت برنامه‌نویسی ژنتیک مدل می‌شود. با استفاده از روش پیشنهادی، کارایی برنامه‌نویسی ژنتیک برای تصدیق هویت گوینده از ۵۰٪ به حدود ۹۲٪ افزایش پیدا کرده است. نتایج حاصل از عملکرد برنامه‌نویسی ژنتیک با کارایی روش‌های تمایزی دیگری مثل شبکه‌های عصبی MLP و LVQ و نیز روش‌های غیر تمایزی مانند LBG، GMM، GMM - UBM و VQ - MAP مقایسه گردید و مشاهده شد که برنامه‌نویسی ژنتیک کارایی بهتری را نسبت به دیگر روش‌ها نتیجه می‌دهد.

کلید واژه: **بازشناسی گوینده، تصدیق هویت گوینده، برنامه‌نویسی ژنتیک، خوشه‌بندی، ویژگی MFCC، ویژگی PLP.**

۱- مقدمه

نیازی فزاینده برای شناسایی هویت افراد در دنیای اطلاعات امروز وجود دارد که کاربردهای آن از کارت‌های اعتباری گرفته تا جرم‌شناسی گسترش یافته است. در حالت کلی، شخص می‌تواند توسط چیزی که مالک آن است و یا اطلاعات مخصوص به خود، شناسایی شود. این ابزار تعیین هویت سنتی که در طول چندین قرن مورد استفاده قرار گرفته‌اند، می‌توانند گم، دزدیده، فراموش‌شده و یا مورد سوء استفاده قرار گیرند.

روش‌های جدیدتر شناسایی هویت مبتنی بر استفاده از خصوصیات فیزیکی و رفتاری افراد است که به روش‌های شناسایی هویت مبتنی بر زیست‌سنجی[1] معروف هستند. این روش‌ها بعضی از مشکلات ذکرشده در روش‌های سنتی را تا حدود زیادی برطرف می‌کنند. یکی از تکنیک‌های شناسایی هویت مبتنی بر زیست‌سنجی که به‌عنوان یک روش مناسب مورد توجه محققان بوده است، روش مبتنی بر استفاده از گفتار است.

به‌طور کلی سیگنال گفتار شامل دو بخش کلی مفهوم گفتار و هویت گوینده است. زمینه شناخت گوینده مربوط به تکنیک‌هایی برای استخراج مشخصات گوینده از روی سیگنال است. در مغز انسان به راحتی گوینده‌های آشنا از یکدیگر تفکیک و تشخیص داده می‌شوند، ولیکن آنچه که برای تشخیص گوینده در سیستم‌های مصنوعی لازم است، شناخت و استخراج مشخصات گوینده از روی سیگنال گفتار آنها است.

هر شخصی یک الگوی صدای منحصر به فرد دارد که او را از دیگران متمایز می‌سازد. در صورتی که دو شخص صداهای مشابه داشته باشند، شباهت معمولاً ظاهری است و از طریق آنالیز الگوهای صدا، تمایز بین آنها با یک درجه بالایی از دقت امکان‌پذیر است. با ظهور کامپیوترهای رقومی، آنالیز مشخصات مختلف صدای شخص با استفاده از روش‌های پردازش سیگنال رقومی (DSP) امکان‌پذیر شده است. الگوریتم‌های قدرتمند برای جداسازی مشخصات ویژه یک الگوی صدا و مقایسه این الگوها با یک نسخه ذخیره‌شده جهت پی‌بردن به این که شخص صحبت‌کننده چه کسی است در دسترس هستند. با این الگوریتم‌ها، شناسایی صحیح یک شخص در صورتی که گفتار او توسط عواملی چون نویز محیطی، نویز حاصل از انتقال گفتار بر روی کانال‌های مخابراتی و مانند آن مخدوش نشده باشد، با یک احتمال خطای کوچک قابل قبول امکان‌پذیر است [۱].

به‌منظور انجام پردازش گفتار، سیگنال‌های گفتار بایستی به پارامترهایی تبدیل شوند که قابلیت استفاده توسط روش‌های مختلف را داشته باشند. از جمله ضرایب مورد استفاده، می‌توان به ضرایب خودهمبستگی، ضرایب پیشگویی خطی، ضرایب کپسترال مبتنی بر معیار مل (MFCC)، ضرایب پیشگویی خطی ادراکی (PLP)[2] و پارامترهای نوایی مانند فرکانس گام اشاره کرد. در مرحله بعدی، مدل‌سازی گویندگان صورت می‌گیرد که برای این منظور از روش‌های گوناگونی از جمله مدل مخلوط گوسی (GMM) و مدل مخفی مارکوف (HMM) می‌توان استفاده کرد.

در ادامه مروری بر کارهای پیشین خواهیم داشت. سپس سبک ارائه مقاله بیان می‌شود و روش برنامه‌نویسی ژنتیک و خوشه‌بندی که در این مقاله مورد استفاده قرار گرفته‌اند، توضیح داده می‌شوند. بخش بعد، مروری کلی بر اجزای مختلف تشکیل‌دهنده یک سیستم تشخیص گوینده خواهد داشت. سپس روش پیشنهادی تشریح شده و نتایج آن ارائه می‌گردد. در نهایت بحث بر روی نتایج انجام شده و نتیجه‌گیری ارائه خواهد شد.

این مقاله در تاریخ ۹ آذر ماه ۱۳۸۸ دریافت و در تاریخ ۱۸ بهمن ماه ۱۳۸۹ بازنگری شد. این تحقیق توسط دانشگاه صنعتی امیرکبیر پشتیبانی شده است.

سعیده سادات سدیدپور، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، (email: sadidpur@aut.ac.ir).

محمدمهدی همایونپور، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، (email: homayoun@aut.ac.ir).

مهدی فسنقری، پژوهشکده فناوری اطلاعات، مرکز تحقیقات مخابرات، انتهای امیرآباد شمالی تهران، (email: fasanghari@gmail.com).

1. Biometric

2. Perceptual Linear Prediction

Archive of SID

www.SID.ir

۲- مروری بر کارهای پیشین

مروری جامع بر روی مطالعات انجام‌شده در تصدیق هویت گوینده نشان می‌دهد که می‌توان بخش‌های اصلی تشکیل‌دهنده یک سیستم تصدیق هویت گوینده را به‌صورت زیر تقسیم‌بندی کرد [۲]:

- استخراج ویژگی
- تطابق با الگو و مدل‌سازی
- تصمیم‌گیری

تحقیقات در زمینه طراحی سیستم شناسایی گوینده و شناسایی گفتار، یکی از موضوعات مورد تحقیق در طول پنج دهه گذشته بوده است. اولین تحقیقات در زمینه تشخیص گوینده، ده سال بعد از شروع تحقیقات در زمینه تشخیص گفتار و در سال ۱۹۶۰ میلادی، توسط پروزونسکی در آزمایشگاه بل با استفاده از بانک‌های فیلتر و طیف‌نگاشت[1] آغاز شد [۳]. در دهه ۱۹۷۰ به‌منظور استخراج ویژگی‌های گوینده مستقل از متن، پارامترهای ضرایب خودهمبستگی [۴]، روش ماتریس کواریانس طیفی [۵] و ضرایب پیشگویی خطی [۶] محاسبه شده‌اند. با ارائه روش خوشه‌بندی که یک روش غیر پارامتریک است، مجموعه بردارهای ویژگی آموزشی گویندگان در یک بسته کوچک به نام کتاب‌کد یا چندی‌سازی برداری (VQ) فشرده‌سازی شده است [۷]. در همان دهه، پریتس با استفاده از مدل مخفی مارکوف (که یک مدل پارامتریک است) بیان‌های مختلف گویندگان را مشخصه‌سازی کرد [۸].

مرحله استخراج ویژگی به دنبال دسته مشخصه‌ای است که بیشترین تمایز را بین گوینده‌ها فراهم آورد و همچنین در شرایط نامناسب (مثل کانال نویزی) اعوجاج کمتری داشته باشد. ویژگی‌های مختلفی در سال‌های اخیر به‌صورت گسترده در سیستم‌های تصدیق هویت گوینده مورد استفاده قرار گرفته است. از جمله این ویژگی‌ها، ضرایب کپسترال هستند که مقیاس فرکانسی را به نحوی تغییر می‌دهند که متناظر با دریافت غیر خطی انسان از فرکانس‌های صوت باشد [۹]. ضرایب کپسترال انواع مختلفی دارند. این ضرایب به طور عمده در تخمین ادراک غیر خطی انسان از فرکانس گام، طراحی بانک فیلتر و فشرده‌سازی خروجی بانک فیلتر با هم متفاوت هستند. یک ارزیابی مقایسه‌ای در [۱۰] برای مقایسه انواع ضرایب کپسترال از جمله ضرایب کپسترال مبتنی بر معیار مل (ضرایب MFCC) آورده شده است. تشخیص گوینده مبتنی بر موجک بسته‌ای، روش دیگری است که برای استخراج مشخصه‌های گفتار استفاده می‌شود [۱۱]. تشخیص گوینده اگر تنها با استفاده از فرکانس گام صورت گیرد، زمانی خوب عمل می‌کند که تعداد گویندگان موجود در پایگاه داده کم باشد. ترکیب فرکانس گام و MFCC ویژگی‌های مکملی هستند که عملکرد سیستم را در حالتی که تعداد گویندگان زیاد شود، ارتقا می‌دهند [۱۲].

مرحله مدل‌کردن، از بردارهای مشخصه یک گوینده خاص برای آموزش صحیح یک مدل عمومی برای همان گوینده استفاده می‌کند. از مدل مخلوط گوسی برای تعیین هویت گوینده مستقل از متن استفاده می‌شود. مؤلفه‌های گوسی موجود در GMM برای نمایش اشکال طیفی وابسته به گوینده مناسب می‌باشند [۱۳]. یک سیستم شناسایی گوینده جدید ارائه‌شده در [۱۴] از یک مجموعه ترکیب جدید که مشخصات طولانی مدت گوینده را شامل می‌شود، استفاده می‌کند. این سیستم شامل مجموعه‌ای از GMMهای مستقل است. خروجی این GMMها در مرحله امتیازدهی با استفاده از یک شبکه عصبی ترکیب می‌شوند.

در سیستم شناسایی گوینده با استفاده از کلمه عبور ویژه هر کاربر و مبتنی بر مدل HMM - GMM، مشتری می‌تواند کلمه عبور مربوط به خود را بدون هیچ محدودیت گفتاری انتخاب کند. کلمه عبور، چندین بار در طول مرحله یادگیری تلفظ می‌شود تا مدل وابسته به مشتری ساخته شود [۱۵]. سیستم‌های شناسایی گوینده با فناوری جدید نیز ایجاد شده‌اند که مبتنی بر مدل‌های مولد آموزش‌دیده به‌صورت تمایزی می‌باشند [۱۶].

ماتسوی عملکرد روش مبتنی بر خوشه‌بندی را با حالت‌های پیوسته و گسسته مدل مخفی مارکوف مقایسه نمود و دریافت که روش مدل مخفی مارکوف پیوسته بسیار بهتر از روش مدل مخفی مارکوف گسسته عمل می‌کند [۱۷]. در طی سال‌های ۱۹۹۸ تا ۲۰۰۲، لی و همکارانش رویکرد جدیدی به تصدیق هویت گوینده داشتند. آنها از اطلاعات مفهومی گفتار و مدل مخفی مارکوف و بردارهای ویژگی کپستروم مبتنی بر مل جهت تصدیق هویت گوینده استفاده کردند [۱۸] و [۱۹]. در سال ۲۰۰۳ نیز وان از ماشین بردار پشتیبان (SVM) و بردارهای ویژگی کپستروم مبتنی بر معیار مل جهت تصدیق هویت گوینده استفاده کرد [۲۰].

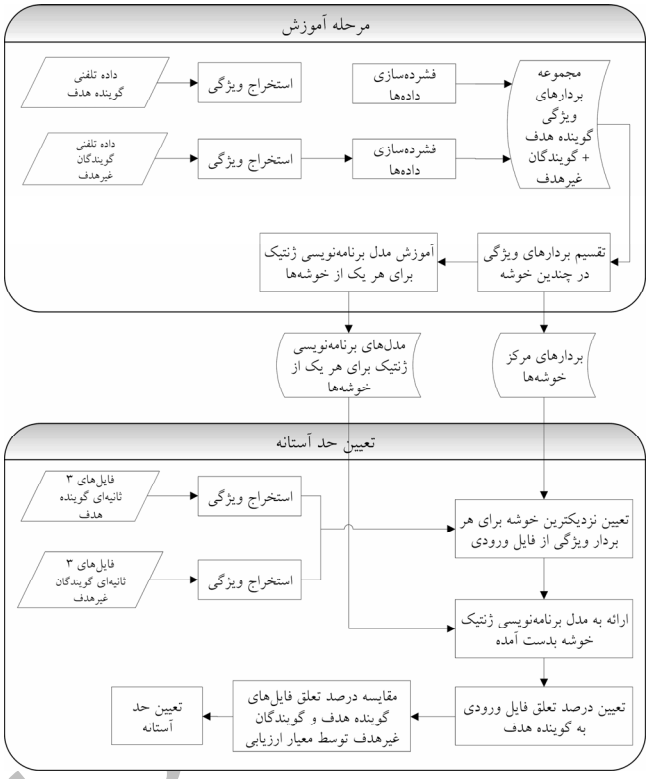
یکی از روش‌های کارا در کاربردهای تشخیص گوینده، مدل مخلوط گوسی است. رینولدز از روش توسعه‌یافته GMM - UBM در تصدیق هویت گوینده استفاده کرد [۲۱]. هونگ از مدل مخلوط گوسی و مخفی مارکوف، حداکثر درستنمایی و بهینه‌سازی پارامترها با استفاده از الگوریتم ژنتیک برای تشخیص گوینده استفاده نمود [۲۲]. هاتاماکی نیز روش VQ - MAP را به‌منظور تصدیق هویت گوینده به‌کار برد. این روش شبیه GMM - UBM است؛ با این تفاوت که تنها از مراکز گوسی‌ها استفاده کرده و در نتیجه پیاده‌سازی آن ساده‌تر می‌باشد [۲۳]. آوسی از شبکه عصبی موجک بسته‌ای و الگوریتم ژنتیک به‌منظور استخراج ویژگی و طبقه‌بندی برای تشخیص گوینده استفاده کرده است [۲۴]. فارائون با استفاده از الگوریتم انتخاب منفی سیستم‌های ایمنی مصنوعی و ضرایب MFCC، به ساخت اثر صوتی برای تشخیص گوینده وابسته به متن پرداخته است [۲۵]. دی تصدیق هویت گوینده مستقل از متن را با استفاده از برنامه‌نویسی ژنتیک و به همراه مهاجرت انجام داده است [۱].

ووهایبی شبکه‌های عصبی برای تشخیص گوینده را با یکدیگر مقایسه کرده است که بخش تعیین هویت، مبتنی بر یک شبکه عصبی است که می‌تواند در یک یا دو کلاس طبقه‌بندی شود. نتایج مقایسه چندین الگوریتم متفاوت نشان می‌دهد که کارایی شبکه عصبی برای طبقه‌بندی دو کلاسی افزایش می‌یابد [۲۶]. لانگ تشخیص گوینده مستقل از متن را به کمک انتخاب ویژگی موجک مبتنی بر شبکه عصبی چندلایه با استفاده از الگوریتم‌های یادگیری با نظارت انجام داده است [۲۷]. مارکوف تشخیص گوینده مستقل از متن را با استفاده از تبدیل غیر خطی درستنمایی فریم به‌گونه‌ای انجام داده است که درستنمایی حاصل از هر فریم بر طبق برخی توابع غیر خطی به امتیازات جدیدی تبدیل می‌شود [۲۸].

مرحله تصمیم‌گیری شامل استفاده از امتیازات تولیدشده توسط مدل‌های گوینده است که تطابق گوینده مدعی را با هویت گوینده مرجع مورد بررسی قرار می‌دهد. یک روش تصمیم‌گیری رایج در سیستم‌های تصدیق هویت گوینده، مبتنی بر مدل مخلوط گوسی است. این روش که نرخ لگاریتم درستنمایی نام دارد؛ نسبت احتمال مربوط به گوینده مدعی به گوینده مرجع و نیز گویندگان دروغگو را محاسبه و بر هم تقسیم کرده، حاصل را با یک مقدار آستانه مقایسه می‌کند [۲۱]. دسته‌بندی کننده GMM می‌تواند به‌عنوان یک روش پس‌پردازش بهینه برای افزایش

1. Spectrogram

Archive of SID
www.SID.ir



شکل ۱: مراحل انجام تصدیق هویت گوینده تلفنی توسط ایده‌های پیشنهادی مقاله.

کارایی سیستم‌های تصدیق هویت مبتنی بر مدل مخلوط گوسی مورد استفاده قرار گیرد [۲۹]. روش تصمیم‌گیری رایج نرخ احتمال لگاریتمی می‌تواند با برخی از توابع تصمیم‌گیری پیچیده‌تر و تفکیک‌کننده‌تر مانند توابع مبتنی بر رگرسیون خطی و یا ماشین بردارهای پشتیبان جایگزین شود [۳۰]. سطح آستانه تصمیم‌گیری برای رد یا قبول گوینده می‌تواند به صورت یک تابع از نرخ پذیرش اشتباه و تعداد فریم‌های آموزشی در مدل گوینده محاسبه شود [۳۱].

خطوط تلفن به دلیل محدودکردن پهنای باند و اثر نویز کانال، تأثیر منفی بر روی داده‌های صوتی گذاشته و تشخیص گوینده را مشکل می‌سازند. در اکثر مطالعات، تصدیق هویت گوینده بر روی داده‌های تمیز که به‌صورت مستقیم توسط میکروفن ضبط شده‌اند، انجام شده است. در صورتی که هدف مطالعه بررسی داده تلفنی و نویز کانال نیز بوده است، با استفاده از روش‌های مختلف داده تمیز را به داده نویزی تلفنی تبدیل کرده‌اند.

در این مقاله که به تصدیق هویت گوینده بر روی داده تلفنی اختصاص دارد، از داده ضبط‌شده پس از تماس تلفنی استفاده شده است که در نتیجه به روش‌های مقاوم نسبت به نویز کانال نیاز می‌باشد که برنامه‌نویسی ژنتیک یکی از آنها است. با توجه به داده‌های با حجم زیاد گفتاری، مدت زمان آموزش مدل‌های برنامه‌نویسی ژنتیک طولانی می‌باشد که در این مقاله تلاش شده است با کمک ایده‌های مختلف، مدت زمان آموزش کاهش یافته و نیز با ایده‌هایی دیگر، کاهش دقت آن جبران و هویت گوینده با دقت مناسبی تشخیص داده شود.

۳- سبک‌شناسی

هدف این مقاله، تحقیق و بهبود شیوه‌های کنونی مورد استفاده در زمینه شناسایی یک فرد با استفاده از سیگنال گفتار می‌باشد. برنامه‌نویسی ژنتیک، روشی است که آموزش آن زمان زیادی را می‌طلبد ولی در تحقیقات صورت‌گرفته توسط آقای دی، دقت بالایی برای تصدیق هویت گوینده با استفاده از برنامه‌نویسی ژنتیک که بعد از این به اختصار آن را GP می‌نامیم، گزارش شده است [۱]. در اغلب مواردی که از GP استفاده می‌شود، تنها یک بار آموزش مدل انجام می‌گیرد ولی در کاربردهای پردازش گفتار مثل شناسایی گوینده، برای هر گوینده، یک مدل آموزش داده می‌شود که زمان زیاد برای آموزش هر مدل، کاربرد GP در پردازش گفتار را محدود می‌نماید. در این مقاله ایده‌هایی برای کاهش زمان آموزش و بهبود عملکرد سیستم ارائه خواهد شد.

تعداد نسل، تعداد جمعیت اولیه و تعداد داده ورودی برای محاسبه برازندگی از جمله عوامل مؤثر در زمان آموزش برنامه‌نویسی ژنتیک هستند. دو مورد اول، مربوط به تنظیمات GP و مورد سوم ورودی سیستم آموزشی می‌باشد. یکی از روش‌هایی که می‌تواند به‌منظور کاهش زمان آموزش استفاده شود، کاهش حجم داده‌های ورودی می‌باشد. در داده‌های صوتی به‌ازای هر فریم، یک بردار ویژگی به‌دست می‌آید که یک داده ورودی برای سیستم آموزشی GP می‌باشد. اگر کاهش حجم داده ورودی، به‌صورت استفاده از بردارهای ویژگی کمتر و در نتیجه تعداد فریم‌های کمتر باشد، در صورتی که تعداد آنها از حد معینی کمتر باشد، آموزش کامل نخواهد بود و GP نمی‌تواند تمامی خصوصیات گفتار گوینده را ببیند و خود را با این خصوصیات تطبیق دهد. استفاده از حدود دو دقیقه گفتار گویندگان لازم است تا مدل بتواند آموزش مورد نیاز را ببیند که با تنظیمات در نظر گرفته شده برای جابه‌جایی فریم‌ها، تقریباً برابر با ۱۲۰۰۰ بردار ویژگی است. این تعداد، زمان آموزش را بسیار افزایش می‌دهد. پس حد تعادلی برای انتخاب تعداد بردارهای آموزشی لازم است. در این مقاله برای کاهش بردارهای ویژگی از روش‌های چندی‌سازی برداری و مدل مخلوط گوسی استفاده شده است.

در صورت استفاده از روش‌های چندی‌سازی برداری برای هر گوینده در فاز آموزش یک کتاب‌کد به‌عنوان مدل آن گوینده ساخته می‌شود. در مرحله تست فاصله بردارهای ویژگی داده تست از اعضای کتاب‌کد (مدل) گوینده به‌دست آمده و مجموع این فواصل در کار تصدیق یا تعیین هویت گوینده استفاده می‌شود. روش GP به دنبال ابرصفحه‌ای است که بتواند داده‌های دسته‌های مختلف را از یکدیگر تفکیک کند. این ابرصفحه در بهترین حالت، مرزی در فضای داده‌هاست که داده‌های گوینده خودی و گویندگان غیر خودی در دو سمت آن قرار گیرند. روش‌های چندی‌سازی برداری، به‌دلیل غیر تمایزی بودن، امکان تفکیک مرز بین داده‌های کلاس‌های خودی و غیر خودی را ندارند.

با توجه به نتایج به‌دست آمده، پراکندگی نمایندگان گویندگان خودی و غیر خودی به‌گونه‌ای می‌باشد که داده‌های خودی و غیر خودی، با یک ابرصفحه از یکدیگر جدا نمی‌شوند و همواره خطایی وجود خواهد داشت.

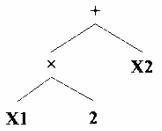
مراحل ایده‌های پیشنهادی مقاله برای تصدیق هویت گوینده توسط برنامه‌نویسی ژنتیک در شکل ۱ نمایش داده شده است. در مرحله تست از فایل‌های ۳ ثانیه‌ای متفاوتی استفاده شده و درصد تعلق فایل ورودی به گوینده هدف محاسبه می‌شود. در مرحله بعد این درصد تعلق با حد آستانه مقایسه شده و صحت ادعای گوینده تعیین می‌شود.

۳-۱ برنامه‌نویسی ژنتیک

برنامه‌نویسی ژنتیک توسط کرامر تعریف شد [۳۲] و بعدها توسط کزا گسترش یافت [۳۳]. این روش، زیرمجموعه‌ای از الگوریتم ژنتیک بوده و از عملگرهای آن از جمله بازترکیبی و جهش استفاده می‌کند. در این روش، درخت‌های عبارت تولید می‌شود که این درخت‌ها به‌عنوان مدل یا یک برنامه کامپیوتری استفاده می‌شوند. هر یک از این برنامه‌های

Archive of SID

www.SID.ir



شکل ۲: نمونه‌ای از یک فرد در برنامه‌نویسی ژنتیک (معادل $X+(2\times X1)$).

کامپیوتری یا مدل‌ها یک عضو از جمعیت می‌باشند. از قابلیت‌های برنامه‌نویسی ژنتیک این است که می‌تواند معادلات ریاضی را مدل کرده و در این راستا، خطوط و منحنی‌های مختلف را شبیه‌سازی کند.

تفاوت اصلی برنامه‌نویسی ژنتیک و الگوریتم ژنتیک در نمایش افراد جمعیت می‌باشد. در روش برنامه‌نویسی ژنتیک، یک فرد، هم‌زمان با ژنوتیپ و فنوتیپ ارائه می‌شود. افراد جمعیت، درخت‌های کامپیوتری (گراف‌های) ساخت‌یافته سلسله مراتبی با اندازه‌ها و شکل‌های متفاوت هستند. نمونه‌ای از یک فرد در برنامه‌نویسی ژنتیک در شکل ۲ نمایش داده شده است. گره‌های غیر پایانه‌ای، توابع و گره‌های پایانه‌ای یا برگ‌ها، متغیرها یا ثابت‌ها هستند. مجموعه توابع، متغیرها و ثابت‌ها، توسط کاربر مشخص می‌شود. آنگاه برنامه‌نویسی ژنتیک در فضای بسیار بزرگ معادلاتی که توسط این مقادیر اولیه قابل بیان هستند، یک عمل جستجوی تکاملی را انجام خوهد داد.

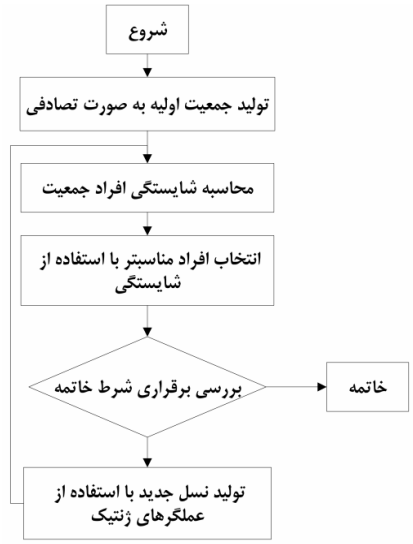
برنامه‌نویسی ژنتیک روشی است که یک جستجوی غیر مرتبط انجام می‌دهد و هیچ اندازه شیب و پیچیدگی ساختاری را برای راه حل، پیش‌بینی نمی‌کند. به همان اندازه که کاربر از تابع هدف اطلاعات دارد و گروه داده‌های مرتبطی که وارد می‌شوند، خروجی متناظری را دریافت کرده و مسئله پاسخ داده می‌شود [۳۴]. برنامه‌نویسی ژنتیک شامل مراحل زیر می‌باشد که در شکل ۳ نشان داده شده است:

۱) با تولید تصادفی جمعیت راه حل‌های با اندازه N شروع می‌شود.
۲) هر راه حل از جمعیت، ارزیابی شده و مقدار برازندگی آن به‌دست می‌آید.
۳) یک عملگر ژنتیک به‌صورت احتمالی انتخاب می‌شود.
- اگر عملگر انتخابی، تکثیر باشد، فردی از جمعیت جاری انتخاب و به جمعیت جدید کپی می‌شود. عملگر تکثیر، اصول انتخاب طبیعی و بقای شایسته‌ترین‌ها را تکرار می‌کند.
- اگر عملگر انتخابی، بازترکیبی باشد، دو فرد انتخاب می‌شوند. سپس به‌صورت تصادفی، زیردرختی در هر فرد انتخاب و این زیردرخت‌ها با یکدیگر جابه‌جا می‌شوند. دو فرزند تولیدشده در جمعیت جدید قرار می‌گیرند. بازترکیبی، نقش حیاتی در فرآیند تکامل ایفا می‌کند.
- اگر عملگر انتخابی، جهش باشد، فردی از جمعیت جاری انتخاب و سپس زیردرختی از آن انتخاب و عمل جهش بر روی آن انجام می‌شود. جهش تنوع را حفظ می‌کند.

۴) مرحله ۳ تا رسیدن جمعیت جدید به N عضو ادامه می‌یابد.
۵) بر خلاف GA، GP لزوماً همگرا نمی‌شود. بنابراین مراحل ۲ تا ۴، تا رسیدن به راه حل مطلوب ادامه می‌یابد و اگر به نتیجه مطلوب دست نیابد، بعد از تعداد تکرار معینی خاتمه می‌یابد.

برنامه‌نویسی ژنتیک با شبکه‌های عصبی رقابت می‌کند ولی مشکل آنها را ندارد. در شبکه عصبی امکان گرفتارشدن در مینیمم محلی وجود دارد ولی برای برنامه‌نویسی ژنتیک، این اتفاق نمی‌افتد. در برنامه نویسی ژنتیک ساختن مدل به دلیل نیاز به جمعیت زیاد، زمان‌بر است ولی به‌دلیل این که تنها یک بار مدل ساخته می‌شود و سپس از آن استفاده می‌شود، کندبودن آموزش آن مشکلی ایجاد نمی‌کند.

بازنمایی کروموزوم‌ها به‌صورت درخت و گراف می‌باشد. جمعیت اولیه،



شکل ۳: فرایند برنامه‌نویسی ژنتیک.

به‌صورت درخت کامل یا ناقص تولید می‌شود. برازندگی، مجموع قدر مطلق تفاضل بین نتایج به‌دست آمده و نتایج مطلوب و در اصل همان فاصله اقلیدسی می‌باشد. مقدار برازندگی کمتر، برتری فرد را نشان می‌دهد. در انتخاب افراد جمعیت برای انجام عملیات‌های مختلف، غیر از جهش که به‌صورت تصادفی می‌باشد، از روش‌های مختلف تورنمنت و شایسته‌سالاری استفاده می‌شود.

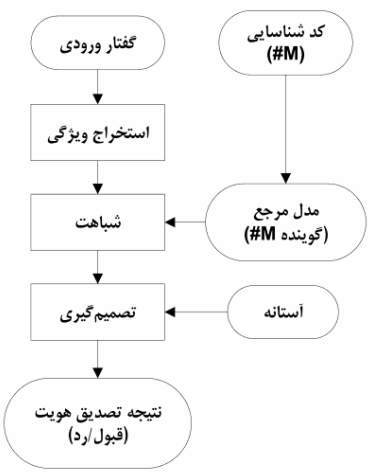
تولید نسل جدید از افراد، توسط اعمال عملگرهای ژنتیک بر روی جمعیت نسل قبل انجام می‌گیرد. دو عملگر ژنتیک، بازترکیبی درخت و جهش درخت می‌باشند که اجرا می‌شود. در بازترکیبی درخت، گره‌ای تصادفی از هر دو درخت والدین انتخاب می‌شود و شاخه‌های مربوطه نسبی خود را جابه‌جا می‌کنند تا دو فرزند جدید تولید کنند. در جهش، یک گره تصادفی از درخت والد انتخاب می‌شود. سپس زیردرخت انتخابی، حذف می‌شود و یا توسط یک درخت تصادفی جدید جایگزین می‌شود.

۳-۲ خوشه‌بندی

برای استفاده از چندی‌سازی برداری برای مسئله شناسایی گوینده، ابتدا به‌ازای هر گوینده تمام بردارهای ویژگی آن گوینده را کنار هم قرار داده و سپس با یک الگوریتم خوشه‌بندی، این بردارهای ویژگی را به M خوشه تقسیم می‌کنند. در مرحله آموزش به‌ازای هر گوینده، کتاب‌کد مخصوص به آن گوینده ایجاد می‌شود و در مرحله تست، سعی در یافتن نزدیک‌ترین بردارهای کتاب‌کد به بردار ورودی داریم. از جمله روش‌های چندی‌سازی برداری می‌توان به روش‌های Split LBG و K - means اشاره کرد. مدل مخلوط گوسی را نیز تقریباً می‌توان نوعی استفاده از چندی‌سازی برداری محسوب نمود.

در روش Split LBG که نوعی الگوریتم چندی‌سازی برداری است، تعداد اعضای کتاب‌کد حاصل از خوشه‌بندی همواره توانی از ۲ می‌باشد. لازم به‌ذکر است که مرکز ثقل هر خوشه به‌عنوان عضوی از کتاب‌کد حاصل از خوشه‌بندی در نظر گرفته می‌شود. در این الگوریتم ابتدا تمام بردارهای ویژگی در یک خوشه قرار گرفته و سپس مرکز ثقل این خوشه محاسبه می‌شود. سپس مرکز ثقل حاصل، با اضافه و کم‌کردن یک بردار به نام بردار ε حاوی مقادیر کوچک به دو بردار جدید تبدیل می‌شود. این دو بردار به‌عنوان مراکز ثقل خوشه‌های جدید در نظر گرفته می‌شوند. هر بار یکی از بردارهای آموزشی به مراکز ثقل دو خوشه جدید چندی شده و به نزدیک‌ترین خوشه تعلق می‌گیرد. پس از اختصاص کلیه بردارهای

Archive of SID
www.SID.ir



شکل ۴: ساختار یک سیستم تصدیق هویت گوینده.

آموزشی به این دو خوشه، در نهایت میانگین برداری بردارهای آموزشی اختصاص‌یافته به هر گروه محاسبه و به‌عنوان مراکز ثقل جدید در نظر گرفته می‌شوند. اختصاص بردارهای آموزشی به خوشه‌ها مجدداً صورت گرفته و هر بار مرکز ثقل جدید محاسبه می‌شود و این کار تا همگراشدن کامل خوشه‌بندی با این دو خوشه ادامه می‌یابد. در ادامه بردار ε مجدداً به مراکز ثقل این دو خوشه اضافه و کم شده و بردارهای حاصل به‌عنوان مراکز ثقل چهار خوشه جدید در نظر گرفته می‌شوند. اختصاص بردارهای آموزشی به این چهار خوشه و تکرار آن تا زمان همگراشدن چهار خوشه ادامه می‌یابد. تفکیک خوشه به خوشه‌های جدید با اضافه‌کردن بردار ε به تعداد دفعات لازم انجام می‌شود تا نهایتاً تعداد خوشه‌ها به تعداد مورد نظر که توانی از ۲ خواهد بود، برسد. [۲].

الگوریتم چندی‌سازی برداری K - Means، اشیای خوشه‌ای با صفات مشابه را به K قسمت تقسیم می‌کند. هدف آن تعیین K میانگین داده‌های تولیدشده از توزیع گوسی است که صفات اشیا را به شکل یک فضای برداری فرض می‌کند. این الگوریتم سعی می‌کند تا کمینه کلی واریانس داخلی خوشه‌ها را پیدا کند [۳۵] و نهایتاً همانند الگوریتم Split LBG یک کتاب‌کد حاوی مراکز ثقل خوشه‌ها ایجاد می‌نماید.

مدل مخلوط گوسی در واقع یک مدل مخفی مارکوف یک‌حالته است که تابع چگالی احتمال آن دارای چندین مؤلفه گوسی می‌باشد. بردارهای کتاب‌کد در روش VQ، متناظر با میانگین توابع گوسی در مدل GMM می‌باشند اما روش VQ به اندازه مدل GMM توانایی مدل‌نمودن تابع چگالی احتمال نمونه‌ها را ندارد [۳۶].

برای بهبود نتایج روش‌های خوشه‌بندی، مقالات استفاده از مدل پس‌زمینه جهانی و تطبیق آن برای هر گوینده را پیشنهاد می‌کنند. در روش GMM - UBM، ابتدا مدل پس‌زمینه‌ای برای تعدادی از گویندگان که گوینده اصلی جزء آنها نیست، آموزش داده می‌شود. سپس برای آموزش مدل هر گوینده، میانگین‌ها و واریانس‌های مدل UBM برای گوینده هدف به‌روز رسانی شده و مدل گوینده اصلی به‌دست می‌آید.

روش دیگری که با ایده از روش GMM - UBM پیشنهاد شده است، روش VQ - MAP می‌باشد که برای کاهش زمان محاسبات، تنها از میانگین‌های مؤلفه‌های گوسی استفاده می‌کند. در این روش نیز ابتدا مدل UBM برای تعدادی از گویندگان مرجع آموزش داده می‌شود. در مراحل بعد با استفاده از روش K - Means، میانگین‌ها به‌روز رسانی شده و مدل گوینده هدف آموزش داده می‌شود [۲۱]، [۲۳] و [۳۷].

٤- تشخیص گوینده

سیستم کامپیوتری که هویت یک فرد را بر اساس مشخصات صوتی او تشخیص می‌دهد، سیستم شناسایی گوینده خودکار نام دارد. در یک دسته‌بندی کلی برای سیستم‌های شناسایی گوینده، می‌توان دو زیرشاخه تصدیق هویت گوینده و تعیین هویت گوینده را در نظر گرفت. بر خلاف سیستم تعیین هویت گوینده که در آن هویت گوینده از میان کاربرهایی که قبلاً به سیستم معرفی شده‌اند تشخیص داده می‌شود، سیستم تصدیق هویت گوینده، گوینده مدعی را با هویت مرجع تطبیق داده و با توجه به میزان مطابقت، وی را قبول و یا به‌عنوان یک گوینده دروغگو رد می‌کند.

در یک سیستم تصدیق هویت گوینده، شخص با اعلام نام خود و یا با واردکردن یک شماره شناسایی شخصی، اعلام هویت می‌کند. فرض بر این است که این فرد قبلاً در مرحله ثبت‌نام، نمونه‌ای از گفتار خود را در اختیار گذاشته و سیستم تصدیق هویت نیز مدلی از صدای این فرد و نیز یک سطح آستانه تصمیم‌گیری برای رد یا قبول هویت او به‌دست آورده است. پس از اعلام هویت توسط کاربر و بیان یک عبارت به‌عنوان رمز عبور صوتی، سیستم ویژگی‌هایی را از گفتار کاربر استخراج و پس از مقایسه آن با مدل کاربر ادعاشده، اگر شباهت صدای گوینده ناشناس با مشخصه صوتی گوینده مرجع بیشتر از یک حد آستانه باشد، هویت ادعاشده توسط گوینده ناشناس مورد تأیید قرار می‌گیرد. شکل ۴ جزئیات بیشتری از یک سیستم تصدیق هویت را نشان می‌دهد [۳۸].

سیستم‌های شناسایی گوینده را می‌توان به دو گروه وابسته به متن و مستقل از متن دسته‌بندی کرد. منظور از سیستم‌های شناسایی گوینده وابسته به متن، حالتی است که در آن سیستم، در مورد محتوای زبانی داده‌های صوتی ورودی و یا متنی که باید خوانده شود اطلاعات قبلی دارد. این داده‌های صوتی معمولاً به شکل عبارات کلیدی (کلمه رمز) است که توسط سیستم از کاربر به‌هنگام ورود به سیستم خواسته می‌شود. در مقابل در سیستم‌های شناسایی گوینده مستقل از متن، هیچ اطلاعی راجع به متن گفته‌شده برای دسترسی و ورود به سیستم وجود ندارد. هر چند که انعطاف‌پذیری این سیستم‌ها به‌خاطر عدم نیاز کاربر به حفظ و به‌خاطر سپردن کلمات رمز بالا است؛ لیکن در مواردی که کاربر در مجاورت افراد دیگر در حال صحبت قرار دارد، دقت سیستم در تشخیص و ارزیابی گوینده پایین است. همچنین از آنجا که مشخص نیست چه عبارت و کلمه‌ای توسط کاربر بیان خواهد شد، باید سیستم با حجم داده‌ای وسیعی از دادگان کاربر آموزش ببیند تا برای هر ترکیب آوایی دلخواه آماده شود.

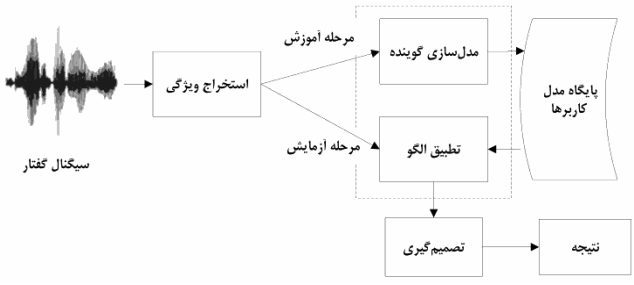
در سیستم‌های بازشناسی گوینده، عوامل مهمی وجود دارند که یک سیستم بازشناسی گوینده را به یک سیستم بازشناسی که بتواند در دنیای واقعی استفاده شود، نزدیک‌تر می‌کند.

در تحقیقات در زمینه بازشناسی گوینده بهتر است که:

- تعداد جلسات ضبط صدا از گوینده، خیلی کم نباشد تا نمونه‌برداری از گفتار گوینده، فقط به یک حالت خاص از او مربوط نشود.
- فاصله بین جلسات ضبط زیاد باشد تا تنوع داده‌های گفتاری موجود در پایگاه داده از هر گوینده به اندازه کافی باشد.
- پایگاه داده جمع‌آوری شده بهتر است به تعداد کافی گوینده زن داشته باشد، چون معمولاً شناسایی و تمایز زنان مشکل‌تر است [۳۹].
- پایگاه داده جمع‌آوری شده در محیط واقعی و نه در محیط کنترل‌شده ضبط و جمع‌آوری شود تا در کاربردهای واقعی مورد استفاده باشد.

Archive of SID

www.SID.ir



شکل ۵: ساختار یک سیستم پایه شناسایی گوینده.

٤-١ روش‌های مدل‌سازی

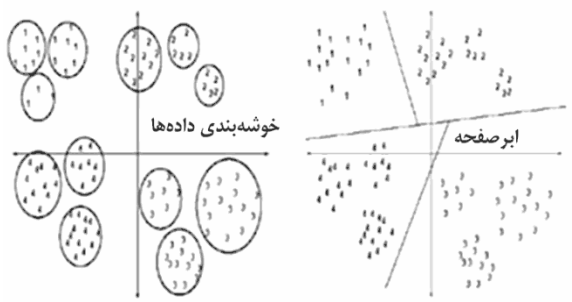
مدل‌سازی گوینده بخش حیاتی یک سیستم تشخیص هویت گوینده را تشکیل می‌دهد. هدف از مرحله مدل‌سازی، تولید یک الگوی مرجع، هنگام معرفی یک کاربر به سیستم و مشخص‌کردن شباهت بین الگوی مرجع و الگوی تست در طول مرحله تست است. با توجه به نمودار موجود در شکل ۵، دسته‌بندی کننده، متشکل از دو بخش مدل‌سازی و تطابق الگو است که هر کدام بسته به مرحله عملیاتی وظیفه خاص خود را انجام می‌دهند. یک مدل خوب از ویژگی‌های کاربر، باید بتواند تطابق الگوی مناسبی را با ویژگی‌های صحبت گوینده مدعی در مرحله تست داشته باشد.

در یک سیستم تشخیص گوینده، روش‌های مدل‌سازی مختلف را می‌توان به دو گروه مدل‌های مولد و مدل‌های تمایزی طبقه‌بندی کرد. مدل‌های تمایزی فقط مرز بین دسته‌ها را مدل می‌کنند و از نوسانات درون هر دسته صرف نظر می‌کنند و از جمله آنها می‌توان به شبکه عصبی چندلایه پرسپترونی (MLP)، دسته‌بندی کننده‌های چندجمله‌ای و ماشین بردار پشتیبان اشاره کرد. مدل‌های مولد با نمونه‌های یک گوینده، مستقل از گوینده دیگر برخورد می‌کنند و برای هر گوینده، یک مدل با استفاده از توابع چگالی احتمال، از روی نمونه‌های داده یادگیری گوینده ساخته می‌شود و مدل‌های مخلوط گوسی و مدل‌های مارکوف پنهان از جمله آنها هستند.

شکل ۶ سمت چپ، دسته‌بندی‌کننده مولد را نشان می‌دهد که در آن داده‌های یادگیری برای هر گوینده در داخل خوشه‌هایی گروه‌بندی شده‌اند که این کار مدل آن گوینده را تشکیل می‌دهد. در مرحله تست فاصله از داده‌های گوینده مدعی تا میانگین خوشه‌ها، نشان‌دهنده شباهت گفتار گوینده مدعی به مدل مربوط به هویت ادعا شده است. شکل ۶ سمت راست، یک دسته‌بندی‌کننده تمایزی را نشان می‌دهد. در اینجا فضای ویژگی به‌وسیله دسته‌ای از صفحات تمایزی به نواحی متمایز تقسیم می‌شوند. در مرحله تست، بردارهای ویژگی گوینده با نواحی مختلف مقایسه شده و در هر ناحیه‌ای که قرار گیرند، فرض می‌شود که به وسیله گوینده متناظر با آن ناحیه ایجاد گشته‌اند [۴۰].

مدل‌های ترکیبی از ترکیب مدل‌های تمایزی و مولد ایجاد می‌شود و سه نوع مختلف دارد [۴۱]:

- مدل‌های ترکیبی سری که یک یا چند مدل طبقه‌بندی تمایزی یا مولد به‌صورت سری در کنار هم قرار می‌گیرند.
- مدل‌های پیچشی که از دو یا چند مدل طبقه‌بندی مولد و یا تمایزی تشکیل شده‌اند و از بعضی از مدل‌ها برای تخمین پارامترهای مدل‌های دیگر استفاده می‌شود.
- مدل‌های ترکیبی بر پایه انتخاب/ هم‌جوشی که به معنای جمع‌کردن وزن‌دار دو یا چند نتیجه مشابه و تولید یک نتیجه مناسب‌تر، حاوی



شکل ۶: مقایسه دسته‌بندی‌کننده‌های مولد با دسته‌بندی‌کننده‌های تمایزی (از راست به چپ: دسته‌بندی‌کننده تمایزی، دسته‌بندی‌کننده مولد) [۴۰].

اطلاعات وزن داده‌شده تمامی نتیجه‌ها هستند.

در این مقاله تمرکز اصلی بر روی روش برنامه‌نویسی ژنتیک است که در زیرمجموعه مدل‌های تمایزی قرار می‌گیرد. از روش‌های ترکیبی سری برای بهبود کارایی روش برنامه‌نویسی ژنتیک استفاده شده است. به‌عنوان مثال، بردارهای آموزشی به یک روش چندی‌سازی برداری و یا یک مدل GMM داده شده و اعضای کتاب‌کد حاصل از روش چندی‌سازی و یا پارامترهای گوسی‌های حاصل از مدل GMM به‌عنوان داده‌های آموزشی به روش برنامه‌نویسی ژنتیک داده شده است.

در سیستم‌های مدل‌سازی تمایزی، نیاز به مجموعه داده‌ای از گویندگان غیر از گوینده هدف می‌باشد تا سیستم تمایز بین داده مربوط به گوینده هدف و دیگر گویندگان را پیدا کند. به این گویندگان غیر از گوینده هدف، غیر خودی‌ها گویند. معرفی مدل غیر خودی‌ها در سیستم‌های شناسایی گوینده به‌عنوان یک به‌هنجارکننده برای حداقل‌کردن تغییرات نامربوط به گوینده (مثل میکروفون، نویز و متن بیان‌شده) عمل می‌کند. دو روش غالب برای نشان‌دادن مدل غیر خودی‌ها، مدل پس‌زمینه جهانی و هم‌گروهی‌ها[1] هستند. سیستم‌های شناسایی گوینده بایستی به فریم‌های گفتاری متعلق به گویندگان مجاز، در مقایسه با فریم‌های گفتاری متعلق به سایر گویندگان امتیاز بیشتری بدهد.

انتخاب اندازه و ترکیب غیر خودی‌ها، موضوعی برای بسیاری از تحقیقات بوده است [۴۲]. در حالت کلی مشخص شده است که برای به‌دست آوردن بهترین کارایی، نیاز به استفاده از مجموعه گوینده‌های دیگر وابسته به گوینده است که گویندگان مشابه با گوینده هدف، اصطلاحاً هم‌گروهی‌های آن گوینده نامیده می‌شوند. عیب این روش این است که اگر سیستم از کاربران زیادی استفاده کند، هر کدام باید برای خود یک مجموعه هم‌گروهی‌ها داشته باشد.

روش دیگر برای مدل‌کردن غیر خودی‌ها، یک مدل جهانی است که با استفاده از سیگنال‌های گفتاری تعداد زیادی از گویندگان ساخته می‌شود. با استفاده از این روش می‌توان مدلی کلی از سیگنال گفتار گویندگان به‌دست آورد. ایده اصلی این دیدگاه، مدل‌کردن خواص گفتاری مشترک گویندگان می‌باشد.

٤-٢ معیارهای ارزیابی

سیستم تصدیق هویت گوینده، از امتیازهای به‌دست آمده برای بیان شباهت بین گوینده مدعی و مدل مربوط به هویت ادعاشده توسط او، در مرحله تطبیق الگو استفاده می‌نماید. نسبت تعداد الگوهای پذیرفته‌شده اشتباه به تعداد کل الگوهای برگرفته از غیر خودی‌ها، نرخ پذیرش اشتباه (FAR) و نسبت تعداد الگوهای مردودی متعلق به کاربر به تعداد کل

1. Cohorts

Archive of SID
www.SID.ir

الگوهای برگرفته از او، نرخ رد اشتباه (FRR) خوانده می‌شود [۴۳]. یک معیار متداول برای ارزیابی سیستم‌های تصدیق هویت گوینده، نصف نرخ خطای کلی (HTER) می‌باشد. HTER، میانگین ریاضی نرخ خطای پذیرش اشتباه و نرخ خطای رد اشتباه در مقدار آستانه خاص انتخابی می‌باشد و با (۱) تعریف می‌شود

$$HTER = \frac{1}{2}(FAR + FRR) \qquad (1)$$

۵- خوشه‌بندی- برنامه‌نویسی ژنتیک

ایده پیشنهادی در این مقاله برای بهبود بیشتر دقت تصدیق هویت به کمک روش برنامه‌نویسی ژنتیک، استفاده از روش ترکیبی خوشه‌بندی- برنامه‌نویسی ژنتیک می‌باشد. اولین گام پیشنهادی برای آموزش مدل GP هر گوینده، فشرده‌سازی داده‌ها می‌باشد. داده‌ها با حجم مناسبی که حداقل چند دقیقه را در بر بگیرد، از جلسات مختلف ضبط‌شده گوینده هدف انتخاب و بردارهای ویژگی MFCC یا PLP آنها استخراج می‌شود. سپس فشرده‌سازی با استفاده از روش‌های خوشه‌بندی مانند LBG یا GMM بر روی آنها انجام و ۱۰۲۴ نماینده حاصل می‌شود. این نمایندگان، مراکز ثقل خوشه‌ها در روش Split LBG یا Kmeans و یا میانگین‌ها در مدل مخلوط گوسی هستند. این گام برای داده گویندگان غیر خودی نیز انجام می‌شود. بنا به روش مورد استفاده برای انتخاب گویندگان غیر خودی، تعدادی گوینده غیر از گوینده هدف انتخاب می‌شوند. در این مقاله از مجموعه گویندگان مدل جهانی برای گویندگان غیر خودی استفاده شده است که مجموعه‌ای مجزا از گویندگان اصلی سیستم آموزش و تست هستند. پس از انتخاب گویندگان غیر خودی، چندین ثانیه از گفتار آنها انتخاب و پس از فشرده‌سازی، ۱۰۲۴ نماینده حاصل از داده‌های غیر خودی به‌دست می‌آید. بایستی تعداد بردارهای ویژگی گوینده هدف و گویندگان غیر خودی که به‌عنوان ورودی به برنامه‌نویسی ژنتیک ارائه می‌شوند، تقریباً برابر باشد زیرا تعداد بیشتر بردارهای هر یک از دو مجموعه، منجر می‌شود که منحنی به‌دست آمده به سمت آن مجموعه تمایل پیدا کند و در نتیجه، نتایج نامناسبی را ایجاد کند.

در اولین آزمایشاتی که در بخش نتایج ارائه شده‌اند، برای هر گوینده تنها یک مدل برنامه‌نویسی ژنتیک با استفاده از ویژگی MFCC آموزش داده می‌شد. نتایج نامناسب آنها نشان می‌دهد که نمایش داده‌های گفتاری گویندگان مختلف با استفاده از ویژگی MFCC تداخلی دارند که تنها با یک منحنی که برنامه‌نویسی ژنتیک به‌دست می‌آورد، قابل جداسازی نیستند و در نتیجه، ایده آموزش چندین مدل برای مدل‌کردن بخش‌های مختلف سیستم تولید گفتار هر گوینده مطرح شد. در گام بعدی، برچسب‌هایی برای نمایندگان به‌دست آمده در نظر گرفته می‌شود. برچسب بردارهای ویژگی گوینده هدف ۱ و برای گویندگان غیر خودی ۱- می‌باشد. سپس نمایندگان هر دو مجموعه روی هم ریخته شده و با استفاده از روش‌های خوشه‌بندی مثل Split LBG یا Kmeans، به چند خوشه تقسیم می‌شوند. لازم به ذکر است که تعلق هر یک از نمایندگان موجود در هر خوشه به گوینده هدف یا گویندگان غیر خودی با برچسب‌ها مشخص شده است. به‌دلیل شباهت‌هایی که در داده‌های گفتاری وجود دارد، در اغلب خوشه‌ها داده‌هایی از هر دو مجموعه وجود دارد.

در مرحله سوم با استفاده از داده‌های موجود در هر خوشه، یک درخت GP آموزش داده می‌شود. مجموعه مدل‌های آموزش داده شده برای هر گوینده به‌عنوان مدل گوینده هدف در نظر گرفته می‌شوند. آموزش چندین مدل به‌ازای هر گوینده، زمان آموزش مدل‌ها را افزایش می‌دهد. در صورت زیادبودن هم‌پوشانی توزیع داده‌ها و در نتیجه نواحی در فضای ویژگی‌ها، به خوشه‌های بیشتری برای جداسازی داده‌ها نیاز خواهد بود. در این صورت افزایش تعداد خوشه‌ها، منجر به بهبود دقت تصدیق هویت می‌شود و در مقابل، مدت آموزش را افزایش می‌دهد. در نتیجه انتخاب تعداد خوشه‌هایی که تعادلی بین این دو برقرار کند، مهم است که این تعداد با آزمون و خطا به‌دست می‌آید.

در مرحله تست، هر بردار ویژگی نمونه آزمایشی، با مراکز ثقل تمامی خوشه‌های به‌دست آمده در مرحله آموزش مقایسه و نزدیک‌ترین خوشه به داده‌های تست مشخص می‌گردد. آنگاه بردار ویژگی تست به مدل متعلق به آن خوشه اعمال و مدل برنامه‌نویسی ژنتیک، دسته آن بردار ویژگی (هدف یا غیر خودی) را مشخص می‌کند. این عمل برای تمام بردارهای ویژگی داده تست تکرار و تعداد دفعاتی که بردارهای ویژگی داده تست خودی تشخیص داده شده‌اند، شمارش می‌شود. سپس با سطح آستانه حداقل HTER به‌صورت پسین مقایسه و به‌ازای هر فایل ورودی ۳ ثانیه‌ای، گوینده مدعی قبول یا رد می‌شود. دقت برای هر گوینده، با استفاده از میانگین بر روی نرخ قبول و رد اشتباه محاسبه شده و در نهایت، دقت کلی، میانگین کل دقت‌ها است.

۵-۱ جهش

در این مقاله از چندین نوع جهش استفاده شده است که به تفصیل بررسی می‌شوند. در جهش، در ابتدا یک عدد تصادفی تولید می‌شود که انجام یا عدم انجام جهش را با توجه به مقایسه با نرخ جهش تعیین می‌کند. سپس در صورت انجام جهش، محل جهش به‌صورت تصادفی انتخاب می‌گردد.

- **جهش مقدار ساده:** اگر محل انتخاب‌شده برای جهش از برگ‌ها و پایانه‌هایی باشد که مقدار عددی دارند، می‌تواند با مقدار عددی دیگر که به‌صورت تصادفی انتخاب می‌شود، جایگزین شود. انتخاب مقادیر عددی ثابت، از مهم‌ترین مسائل در آموزش مدل GP هستند که برای برازش مدل‌ها در هر مسئله استفاده می‌شوند. برای یافتن اعداد جایگزین و به‌منظور پیداکردن مقادیر بهینه‌تر از استراتژی تکاملی (ES) استفاده شده است. به‌دلیل زمان‌بر بودن، استراتژی تکاملی، به‌عنوان جهش در نظر گرفته شده است تا ضمن استفاده از بهبود ایجادشده توسط آن، از افزایش بسیار زیاد زمان آموزش مدل‌ها جلوگیری شود.
- **جهش جایگزینی:** در این روش اگر درخت به حد اشباع نرسیده باشد، می‌توان یک گره را با یک درخت تصادفی جایگزین کرد. گره انتخاب‌شده برای جایگزینی، همان محل جهش است که به صورت تصادفی انتخاب شده است. زیردرخت برای جایگزینی نیز به صورت تصادفی تولید می‌شود. این جهش به‌منظور تکامل افراد انجام می‌گیرد.
- **جهش جایگشت:** در این روش عملوندهای یک گره که تابع می‌باشد، جایگزین یکدیگر می‌شود. این روش لزوماً بر روی تابعی انجام می‌شود که بیش از یک عملوند داشته باشد. توابعی که بیش از دو عملوند دارند، جابه‌جایی عملوندها به‌صورت تصادفی تعیین می‌شود. این جهش به‌منظور بهینه‌کردن مدل‌ها استفاده می‌شود.
- **جهش ویرایش:** این جهش به کل ژنوم اعمال می‌شود و اطلاعات اضافی و افزونه ژنوم را هرس می‌کند. از چند روشی که به‌منظور

Archive of SID

www.SID.ir

ویرایش به‌کار رفته است، یک نمونه با توجه به بررسی گره و نیاز آن انتخاب می‌شود. اگر گره پایانه باشد، عملی انجام نمی‌شود و اگر تابع با عملوندهای عدد صحیح یا اعشاری باشد، تابع به عملوندها اعمال شده و مقدار نتیجه، جایگزین زیردرخت مورد نظر می‌شود. اگر گره تابع ضرب یا تقسیم با یک عملوند صفر باشد که پاسخ را صفر می‌کند، زیردرخت با گره پایانی آن زیردرخت جایگزین می‌شود. اگر تابع جمع یا تفریق با عملوند صفر باشد که تأثیری در نتیجه ندارد، زیردرخت با ریشه عملوند غیر صفر، جایگزین درخت با ریشه گره تابع مورد نظر خواهد شد. مشکل چاقی یا رشد بیش از حد درخت‌ها، یکی از مشکلاتی است که برنامه‌نویسی ژنتیک با آن درگیر می‌باشد و از جمله روش‌های پیشنهادی برای کنترل آن این است که ابتدا درخت هرس شده و شاخ و برگ‌های زاید آن حذف شود. در نتیجه، این جهش به قصد رفع چاقی بوده و به دلیل این که زمان آموزش را افزایش داده و نیز ممکن است در بازترکیبی‌های مختلف، بخش‌هایی که با مدل‌های دیگر جابه‌جا می‌شوند، مفید باشند، به‌صورت جهش در نظر گرفته شده و برای تمامی افراد اعمال نمی‌شود.

- **جهش معاوضه:** در این روش یک گره با گره‌ای با تعداد پارامتر مشابه تعویض می‌شود؛ چنانچه گره پایانه باشد، با پایانه‌ای که می‌تواند عدد صحیح، اعشاری، توابع بدون عملوند مثل عدد π و یا پارامتر باشد، معاوضه می‌شود؛ و چنانچه گره از توابع باشد، تابعی با تعداد عملوند مشابه به‌صورت تصادفی تولید شده و با گره غیر پایانه‌ای محل جهش جایگزین می‌گردد. این جهش نیز برای بهبود دقت مدل آموزشی استفاده می‌شود.
- جهش تلنگری: در این روش دو گره مجاور که تعداد پارامترهای مشابهی دارند، تعویض می‌شوند. چنانچه گره‌های مجاور، پایانه‌ای (که می‌توانند عدد صحیح، اعشاری، توابع بدون عملوند مثل عدد π و یا پارامتر) باشند، با یکدیگر جابه‌جا می‌شوند. چنانچه گره‌های مجاور از توابع باشند، در صورتی که تعداد عملوندهای یکسانی داشته باشند، با هم جایگزین شده و در اصل، پارامترهای دو تابع با هم تعویض می‌شوند. این جهش به‌منظور انتخاب بهینه‌تر پارامتر و توابع استفاده می‌شود.

۲-۵ تابع برازندگی

در این مقاله از تابع برازندگی ترکیبی استفاده شده است که سعی شده برای مسئله شناسایی گوینده تطبیق یابد.

- **تابع حداقل‌کردن تعداد تشخیص اشتباه:** در این تابع مقدار خروجی مطلوب برای داده‌های خودی، مثبت و برای داده‌های گویندگان غیر خودی، منفی در نظر گرفته می‌شود. تابع برازندگی، کمتربودن خطا می‌باشد و معادله تابع برازندگی به‌صورت (۲) خواهد بود

$$\begin{aligned} &\text{if} \quad (x \in X_1 \ \ \& \ \ G(x) < 0) \,||\, (x \in X_2 \ \ \& \ \ G(x) \geq 0) \\ &\qquad err(i)++; \\ &best_fitness = \arg\min_i [err(i)] \end{aligned} \tag{۲}$$

x داده آموزشی، X_1 کلاس داده‌های گوینده خودی و X_2 کلاس داده‌های گویندگان غیر خودی است. $G(x)$ مقدار خروجی GP و err تعداد داده‌های آموزشی است که اشتباه شناسایی شده است. در این روش $best_fitness$ فردی است که خطای کمتری در تشخیص داشته است.

- **تابع حداکثرکردن حاشیه‌ها:** برای رفع مشکل تابع حداقل‌کردن فاصله، تابع جدیدی پیشنهاد می‌گردد که تعمیم در آن در نظر گرفته شود. تابع مذکور سعی می‌کرد تا منحنی جداکننده دسته‌ها به داده‌ها نزدیک شود، ولی در این تابع سعی می‌شود که بیشترین حاشیه برای هر دسته در نظر گرفته شود و منحنی به‌دست آمده، بیشترین فاصله را از داده‌های هر دسته داشته باشد. تابع برازندگی به‌صورت (۳) خواهد بود

$$\begin{aligned} &\text{if} \quad (x \in X_1 \ \ \& \ \ G(x) \geq 0) \\ &\qquad fitness2(i) = \min_{x \in X} |G(x)| \\ &\text{if} \quad (x \in X_2 \ \ \& \ \ G(x) < 0) \\ &\qquad fitness3(i) = \min_{x \in X} |G(x)| \\ &best_fitness = \arg\max_i [fitness2(i) + fitness3(i)] \end{aligned} \tag{۳}$$

متغیرهای این تابع هم مانند توابع قبلی می‌باشد. $best_fitness$ فردی از جمعیت است که بیشترین فاصله را از داده‌ها داشته باشد. مشکل این روش این است که منحنی به‌دست آمده به سمت بی‌نهایت میل می‌کند.

- **تابع حداقل‌کردن حاشیه خودی و حداکثرکردن حاشیه غیر خودی:** برای رفع مشکل تابع حداقل‌کردن فاصله، تابع حداکثرکردن حاشیه‌ها پیشنهاد شد ولی این تابع نیز مشکل را برطرف نساخت. این تابع که حد فاصل دو تابع قبل می‌باشد، سعی در نزدیک‌کردن منحنی به داده‌های خودی و دورکردن از داده‌های غیر خودی دارد. تابع برازندگی به‌صورت (۴) خواهد بود

$$\begin{aligned} &\text{if} \quad (x \in X_1 \ \ \& \ \ G(x) \geq 0) \\ &\qquad fitness4(i) = \min_{x \in X} |G(x)| \\ &\text{if} \quad (x \in X_2 \ \ \& \ \ G(x) < 0) \\ &\qquad fitness5(i) = \min_{x \in X} |G(x)| \\ &best_fitness = \arg\max_i [fitness5(i) - fitness4(i)] \end{aligned} \tag{۴}$$

متغیرهای این تابع هم مانند توابع قبلی می‌باشد. $best_fitness$ فردی از جمعیت است که بیشترین برازندگی را داشته و فاصله آن از داده‌های غیر خودی بیشتر و از داده‌های غیر خودی کمتر باشد. مشکل این روش این است که احتمال به سمت بی‌نهایت میل‌کردن منحنی وجود دارد.

- **تابع ترکیبی:** برای رسیدن به بهترین نتیجه بهتر است که ترکیبی از توابع را استفاده کرد. در این حالت، اولویت اول به تعداد خطا اختصاص داده می‌شود. هرچه خطا کمتر باشد، برازندگی بیشتر خواهد شد. در مرحله بعد به یکی از توابع حداکثرکردن حاشیه‌ها یا حداقل‌کردن حاشیه خودی و حداکثرکردن حاشیه غیر خودی اهمیت داده می‌شود. برای مواردی که دارای خطای مشابه می‌باشد، بیشتربودن تابع ثانویه، تعیین‌کننده فرد شایسته‌تر خواهد بود. روابط (۵) و (۶)، توابع ترکیبی مورد استفاده هستند

$$\begin{aligned} &fitnessT(i) = |\alpha.err(i) - \beta(fitness2(i) + fitness3(i))| \\ &best_fitness = \arg\min_i [fitnessT(i)] \end{aligned} \tag{۵}$$

Archive of SID

www.SID.ir

جدول ۱: مقادیر پارامترهای تولید جمعیت.

پارامتر GP	مقدار پارامتر
جمعیت اولیه	۱۰۰۰
تعداد نسل	۱۵۰
مجموعه توابع	$\{\cos, \sin, \tan, \log_{10}, \log, \text{sqrt}, /, *, -, +\}$
توابع مورد استفاده در برگ	π
محدوده عدد مورد استفاده در پایانه	۰–۱۰
روش بازترکیبی	یک نقطه‌ای یکنواخت
روش انتخاب	تورنمنت
اندازه تور در تورنمنت	۳

$$fitnessT(i) = \left|\alpha.err(i) - \beta(fitness_5(i) - fitness_4(i))\right|$$
$$best_fitness = \arg\min_i [fitnessT(i)] \quad (۶)$$

α و β ضرایبی هستند که تعیین‌کننده مقدار تأثیر هر یک از توابع برازندگی می‌باشند و تعیین مناسب مقادیر آنها، تأثیر به‌سزایی در نتایج نهایی خواهد داشت.

۵-۳ پارامترهای برنامه‌نویسی ژنتیک

به‌دلیل این که برنامه‌نویسی ژنتیک، روش مدل‌سازی پایه در این مقاله می‌باشد، با به‌دست آوردن بهترین تنظیمات می‌توان نتایج حاصل را بهبود داده و بهینه کرد. جداول ۱ و ۲ به‌ترتیب مقادیر بهینه پارامترهای به‌دست آمده در تولید جمعیت و مقادیر نرخ عملیات مختلف برای تولید نسل را نشان می‌دهند. جمعیت اولیه زیاد، امکان دیدن جواب‌های بیشتر را ممکن می‌سازد ولی مدت زمان محاسبه تابع برازندگی را افزایش می‌دهد که ۱۰۰۰ به‌عنوان جمعیت اولیه انتخاب می‌شود. برای دیدن حالت‌های مختلف با این تعداد جمعیت، تعداد نسل بیشتری باید در نظر گرفته شود که ۱۵۰ نسل می‌باشد.

۴ عمل اصلی حسابی، انتخاب پایه مجموعه توابع است. برای بهبود نتایج، لگاریتم‌ها، جذر، sin و cos موجب کوچک‌کردن و tan موجب کوچک یا بزرگ‌کردن خروجی می‌شود. با توجه به استفاده از توابع سینوسی، بهتر است عدد π نیز بتواند در برگ‌ها انتخاب شود. محدوده عددی ۰ تا ۱۰، برای پوشش محدوده اعداد مناسب است. بازترکیبی، از نوع یک‌نقطه‌ای یکنواخت انتخاب شده است که بیانگر جابه‌جایی والدین از یک نقطه تصادفی در هر یک از آنها است. انتخاب بازماندگان به روش تورنمنت با اندازه تور ۳ صورت می‌گیرد که به معنای انتخاب تصادفی ۳ فرد و برگزیدن یکی از آنها است.

نخبگان و بازماندگان افرادی هستند که بدون شرکت در بازترکیبی یا جهش به نسل بعد منتقل می‌شوند. با توجه به انتخاب یک فرد به‌عنوان بهترین فرد نهایی، تعیین ۲۰ نخبه که بهترین‌های یک نسل می‌باشند، انتخاب مناسبی است. بازماندگان لزوماً بهترین‌های یک نسل نیستند و از بین افراد انتخاب‌شده به اندازه تور انتخاب می‌شوند. جهش قوی به معنی اعمال ۳ بار جهش است. مجموع نرخ جهش و جهش قوی عدد بزرگی در نظر گرفته شده است تا جایگزین کاهش جمعیت شود و افراد با حالت‌های متنوع‌تری دیده شوند. بر خلاف برنامه‌نویسی ژنتیک پایه، تعدادی افراد جدید در هر نسل تولید می‌شوند تا افراد نامناسب‌تر هر نسل حذف شده و افرادی ایجاد شوند که شاید بتوانند جزو بهترین‌ها باشند.

۵-۴ دادگان

دادگان مورد استفاده، استخراج‌شده از دادگان گفتاری TPersianDat است. این دادگان در آزمایشگاه سیستم‌های هوشمند صوتی- گفتاری دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر طی دو سال ضبط شده و حاوی گفتار بیش از ۱۲۰ گوینده می‌باشد که هر گوینده، در ۵ جلسه با یک سیستم تلفنی ضبط صدا تماس و در جلسات ضبط شرکت کرده است. فرکانس نمونه‌برداری ۸ KHz می‌باشد و نمونه‌ها ۱۶بیتی هستند که در اصل، از فایل‌های PCM لگاریتمی با استاندارد A - Law به‌دست آمده‌اند. از بین ۱۲۰ گوینده اصلی این دادگان، ۶۰ گوینده به‌عنوان کاربران اصلی و ۶۰ گوینده به‌عنوان کاربران مدل جهانی در نظر گرفته شده‌اند.

جدول ۲: مقادیر نرخ عملیات مختلف برای تولید نسل.

پارامتر GP	مقدار پارامتر
انتخاب نخبگان	۲۰ تا از تعداد جمعیت
نرخ انتخاب بازماندگان و نخبگان	٪۱۵
نرخ بازترکیبی	٪۲۰
نرخ جهش	٪۳۵
نرخ جهش قوی	٪۱۵
نرخ تولید افراد جدید در هر نسل	٪۱۵

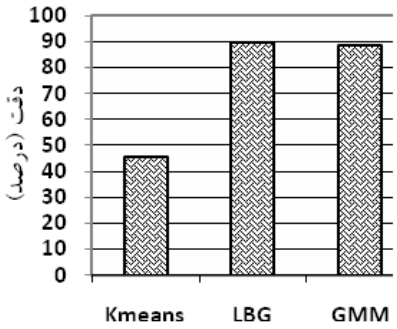
فایل‌های داده به‌صورت گفتارهای مختلف از گوینده می‌باشد که در جلسات مختلفی ضبط شده و داده‌های حذف سکوت شده هر جلسه، به هم متصل شده‌اند. سیگنال گفتار به فریم‌های ۳۰ ms با میزان جابه‌جایی ۱۰ ms تقسیم شده است. از ضرایب MFCC و PLP به‌عنوان ویژگی‌های مختص گوینده استفاده شده‌اند. علت اهمیت MFCC در این است که این ویژگی در کارهای پژوهشی مربوط به تشخیص خودکار هویت گوینده بسیار مفید بوده و نتایج قابل قبولی با استفاده از این ویژگی در این دسته از پژوهش‌ها گزارش شده است. ضرایب PLP نیز مبتنی بر بانک فیلتر فرکانسی مل است که در [۱] از آن استفاده شده است. نتایج به‌دست آمده در بخش آزمایشات نشان می‌دهند که تداخل داده‌های گفتاری گویندگان مختلف با استفاده از این ویژگی، بسیار کم بوده و نتایج بهتری را ارائه می‌کند.

برای هر یک از گویندگان، در مرحله آموزش از ۲ دقیقه گفتار از ۴ جلسه اول همان گوینده و ۱ دقیقه گفتار از هر یک از گویندگان غیر خودی و در مرحله تعیین آستانه از ۱ دقیقه گفتار از ۴ جلسه اول همان گوینده و ۱ دقیقه گفتار از هر یک از گویندگان غیر خودی استفاده شده است. در مرحله تست برای داده خودی هر گوینده، تنها از جلسه پنجم خود گوینده و برای داده غیر خودی، از داده‌هایی که به‌صورت تصادفی از جلسات پنجم گویندگان غیر خودی انتخاب می‌شود، استفاده شده است. در نتیجه دادگان مستقل از متن می‌باشد زیرا گویندگان مختلف، صحبت‌های یکسان و مختلفی را ادا کرده و هیچ محدودیتی در متون گفتار وجود ندارد. گفتار هر گوینده اصلی حاوی موارد زیر است:

- موارد ۱ و ۲ که حاوی شماره شناسایی و شماره فرم هستند.
- موارد ۳ الی ۱۷ که همگی گفتار خواندنی هستند.
- مورد ۱۸ یک جمله که به لحاظ آوایی غنی است، می‌باشد.
- مورد ۱۹ یک جمله کوتاه است که برای بازشناسی گوینده مستقل از متن با یک جمله کوتاه مناسب است.
- مورد ۲۰ یک جمله متوسط است.
- مورد ۲۱ یک جمله بلند است.
- موارد ۲۲ الی ۲۴ که همگی گفتار فی‌البداهه مربوط به بیان یک خبر و پاسخ فی‌البداهه به یک سؤال می‌باشند.

Archive of SID

www.SID.ir



شکل ۷: نتایج دقت با استفاده از روش‌های LBG، Kmeans و GMM با ۱۰۲۴ نماینده بر تصدیق هویت گوینده بر حسب درصد.

جدول ۳: نتایج تصدیق هویت توسط GP با استفاده از داده‌های به‌دست آمده از خوشه‌بندی و GMM بر حسب درصد.

روش استخراج ویژگی	خطای قبول اشتباه	خطای رد اشتباه	دقت
LBG	۲۴٫۸	۵۸٫۳	۵۹٫۵
KMeans	۳۸٫۱	۶۱٫۷	۵۰٫۱
GMM	۲۵٫۸	۴۹٫۰	۶۲٫۶

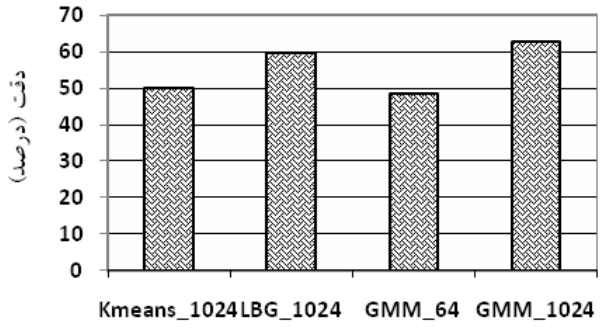
۱۲ ویژگی MFCC همراه با انرژی یا بدون انرژی به همراه مشتقات اول و دوم آنها و نیز ۱۲ ویژگی PLP استفاده می‌شوند. بردار ویژگی هر فریم در صورت استفاده از انرژی، ۳۹ و در صورت عدم استفاده از انرژی، ۳۶بعدی می‌باشد.

٦- نتایج

پیاده‌سازی با استفاده از یک کامپیوتر شخصی و شبیه‌سازی‌های لازم با استفاده از زبان برنامه‌نویسی C++، در محیط Visual Studio ۶ انجام گرفته است. آزمایشات ابتدایی از ضرایب MFCC استفاده می‌کنند. پس از به‌دست آوردن تنظیمات مناسب، نتایج با ضرایب PLP نیز مقایسه می‌شود.

در گام اول برای بیان تأثیر و انتخاب روش مناسب‌تر بین Kmeans، Split LBG و GMM که از جمله تکنیک‌های مؤثر در پردازش گفتار می‌باشد، تصدیق هویت گوینده با استفاده از این روش‌ها به‌عنوان روش‌های غیر تمایزی انجام شده است. مقایسه نتایج به‌دست آمده از این روش‌ها در شکل ۷ نشان داده شده است. نمودار شکل ۷ بیانگر برتری روش LBG نسبت به دیگر روش‌ها می‌باشد.

در ادامه از مراکز ثقل خوشه‌های حاصل از روش‌های Split LBG و Kmeans، به‌عنوان نمایندگان داده‌های آموزشی برای داده‌های گوینده خودی و داده‌های گویندگان غیر خودی استفاده شده است. برای هر گوینده مدل گویندگان غیر خودی با استفاده از ۱۵ کاربر هم‌گروهی برای گوینده خودی (شبیه‌ترین گوینده‌ها به گوینده مورد نظر) به‌دست آمده است. اگر از میانگین مؤلفه‌های گوسی به‌عنوان داده ورودی به GP استفاده شود، به‌دلیل زمان‌بر بودن آموزش GMMها، نمی‌توان برای کاربران هم‌گروهی هر گوینده (به‌عنوان داده‌های غیر خودی) مدل مخلوط گوسی را به‌دلیل زیادشدن داده‌های آموزشی، آموزش داد؛ در این حالت برای کلیه داده‌های کاربران غیر خودی هر گوینده، یک مدل مخلوط گوسی آموزش داده شده و از میانگین‌های مؤلفه‌های گوسی این مدل GMM، به‌عنوان داده‌های غیر خودی استفاده می‌شود. جدول ۳ نتایج برنامه‌نویسی ژنتیک با استفاده از داده‌های به‌دست آمده از Split LBG، Kmeans و GMM را مقایسه می‌کند.



شکل ۸: مقایسه دقت نتایج تصدیق هویت گوینده با استفاده از GP بر روی نمایندگان روش‌های مختلف.

جدول ۴: نتایج تصدیق هویت توسط GP با توابع برازندگی متفاوت بر حسب درصد.

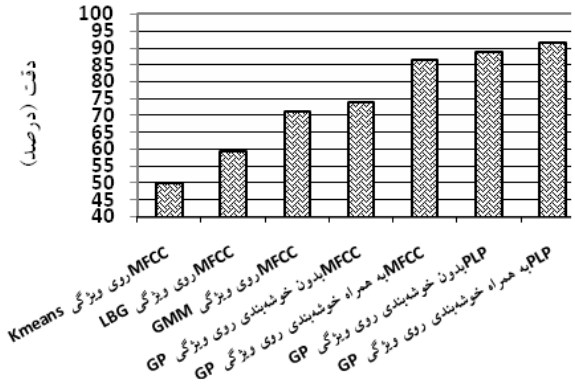
تابع برازندگی	خطای قبول اشتباه	خطای رد اشتباه	دقت
حداقل‌کردن تعداد تشخیص اشتباه	۵۰٫۸	۴۵٫۱	۵۲٫۰
ترکیب حداقل تعداد خطا و حداکثرشدن حاشیه‌ها	۲۴٫۸	۵۸٫۳	۵۹٫۵
ترکیب حداقل تعداد خطا و حداقل‌شدن حاشیه خودی و حداکثرشدن حاشیه غیر خودی	۴۹٫۹	۲۱٫۳	۶۴٫۴

شکل ۸ مقایسه دقت نتایج تصدیق هویت گوینده با استفاده از GP بر روی نمایندگان روش‌های مختلف را نشان می‌دهد. مشاهده می‌شود که روش GMM با تعداد مؤلفه‌های برابر با تعداد اعضای کتاب‌کد روش‌های چندی‌سازی برداری، نتایج مطلوب‌تری را ارائه می‌کند. برای گویندگان اصلی و هم‌گروهی، آموزش کتاب‌کدهای ۱۰۲۴ عضوی به روش KMeans و LBG به‌ترتیب به ۱ و ۶ ساعت زمان نیاز دارند که این تفاوت زمان با توجه به مقدار برتری دقت نتایج روش LBG نسبت به روش KMeans، استفاده از روش LBG را پیشنهاد می‌دهد. آموزش GMMهای ۶۴ و ۱۰۲۴ مؤلفه‌ای برای گویندگان خودی و مدل جهانی، به‌ترتیب ۱۲ ساعت و ۱۰ روز به طول انجامید. بار محاسباتی GMMهای ۱۰۲۴ مؤلفه‌ای در برابر بهبود ۳٪ در روش LBG در صورتی قابل صرف نظر است که نیازی به تغییرات و آموزش مجدد نباشد و در غیر این صورت استفاده از نمایندگان LBG پیشنهاد بهتری می‌باشد.

در بخش قبل، توابع برازندگی مورد استفاده در این مقاله تشریح شدند. در جدول ۴ نتایج استفاده از توابع مختلف بر روی نمایندگان روش LBG نشان داده شده است. α و β در توابع ترکیبی به‌ترتیب برابر ۱۰۰۰۰۰۰۰۰- و ۱ می‌باشند. نتایج بیانگر برتری استفاده از توابع ترکیبی می‌باشند. به‌دلیل تداخل داده‌های گفتاری خودی و غیر خودی، تابع حداکثرکردن حاشیه خودی و غیر خودی باعث می‌شود که تابع از هر دو دسته دور شده و دقت مناسبی را ایجاد نکند. با توجه به نتایج به‌دست آمده، ایده استفاده از تابعی که حاشیه خودی را حداقل و حاشیه غیر خودی را حداکثر کند، مطرح شد تا به غیر خودی‌ها امکان نوسان بیشتری داده و در نتیجه خطا کاهش یابد. نتایج نشان می‌دهد که ایده پیشنهادی مؤثر بوده است.

تا این مرحله در تمامی آزمایش‌ها، از بردارهای ویژگی ۳۹بعدی که شامل ضرایب MFCC، انرژی و مشتقات اول و دوم بوده‌اند، استفاده شده است. علت این بود که با توجه به نتایج ارائه‌شده در مقالات مختلف، مشتقات و انرژی نیز همانند خود ضرایب MFCC حاوی خصوصیات

Archive of SID
www.SID.ir



شکل ۹: مقایسه تأثیر ایده‌های پیشنهادی در تصدیق هویت گوینده با استفاده از برنامه‌نویسی ژنتیک.

جدول ۵: نتایج تصدیق هویت با استفاده از ضریب انرژی، ضرایب MFCC و مشتقات اول و دوم ضرایب MFCC بر حسب درصد.

بردار ویژگی	خطای قبول اشتباه	خطای رد اشتباه	دقت
۱۲ MFCC و انرژی + مشتقات اول و دوم آنها	۳۰/۶	۵۰/۵	۵۹/۴
۱۲ MFCC + مشتقات اول و دوم	۲۷/۴	۵۱/۲	۶۰/۷
۱۲ MFCC	۳۶/۲	۱۶/۲	۷۳/۸

گویندگان بوده و بایستی تأثیر مثبتی در نتایج داشته باشند. برنامه‌نویسی ژنتیک، منحنی‌هایی را پیدا می‌کند که پارامترها در برگ‌های درخت‌ها قرار می‌گیرند. یافتن منحنی مناسب‌تر، وابسته به پیداکردن پارامترهای بهتر می‌باشد که با توجه به این که انتخاب این پارامترها به‌صورت تصادفی می‌باشد، کاهش تعداد پارامترها در پیداکردن درخت‌های مناسب‌تر مؤثر است. نتایج نشان می‌دهد که استفاده از ضریب انرژی در بردارهای ویژگی، به‌جای بهبود نتایج، تأثیر نامطلوب در کارایی سیستم تصدیق هویت داشته و نتایج را بدتر می‌کند. البته این تأثیر، ناچیز و قابل صرف نظر می‌باشد و علت آن به‌جز افزایش تعداد پارامترها، عدم وابستگی خصوصیات گوینده به ویژگی انرژی می‌باشد. استفاده از مشتقات نیز تأثیر منفی دارد زیرا به شدت تعداد پارامترهای GP را افزایش می‌دهد. به نظر می‌رسد بر اساس نتایج ارائه‌شده در جدول ۵ استفاده از تنها ۱۲ ضریب MFCC کارایی بهتری داشته باشد و دقت نتایج تصدیق هویت را بهبود دهد.

ایده پیشنهادی این مقاله که روش ترکیبی خوشه‌بندی- برنامه‌نویسی ژنتیک می‌باشد، با استفاده از خوشه‌بندی به روش‌های Kmeans یا Split LBG و با تعداد خوشه ۵، ۸، ۱۰، ۱۶ و ۳۲ انجام گردیده است. انتخاب دسته‌هایی با تعداد غیر توان ۲، با روش Kmeans انجام می‌شود. جدول ۶ تأثیر خوشه‌بندی و ساخت یک درخت GP به‌ازای هر خوشه را بر نتایج تصدیق هویت به روش GP با استفاده از ویژگی‌های MFCC نمایش می‌دهد. مشاهده می‌شود که خوشه‌بندی به روش بیان‌شده، تأثیر بسیار زیادی در بهبود دقت GP در تصدیق هویت گوینده دارد. روش Split LBG با تعداد خوشه کمتر، پاسخی مشابه Kmeans با خوشه‌های بیشتر ایجاد می‌کند. افزایش خوشه‌ها برای داده‌های آموزشی MFCC گرچه کارایی را بهتر می‌کند، اما از طرفی زمان آموزش و تست را افزایش می‌دهد. بنابراین با در نظر گرفتن یک تعادل بین کارایی و سرعت آموزش، تعداد حداکثر خوشه‌ها ۳۲ انتخاب گردید. در این حالت برای هر گوینده، مدت زمان آموزش از ۱۵ دقیقه بدون استفاده از خوشه‌بندی به

جدول ۶: نتایج تصدیق هویت به روش GP با استفاده از خوشه‌بندی داده‌های آموزشی MFCC و ساخت یک درخت GP به‌ازای هر خوشه بر حسب درصد.

روش خوشه‌بندی	تعداد خوشه‌ها	خطای قبول اشتباه	خطای رد اشتباه	دقت
بدون استفاده از خوشه‌بندی		۳۶/۲	۱۶/۲	۷۳/۸
Kmeans	۵	۲۷/۶	۱۷/۵	۷۷/۵
Kmeans	۱۰	۱۴/۳	۱۷/۹	۸۴/۴
LBG	۸	۱۷/۰	۱۵/۰	۸۴/۰
LBG	۱۶	۱۵/۳	۱۴/۱	۸۵/۳
LBG	۳۲	۱۱/۹	۱۵/۱	۸۶/۵

جدول ۷: نتایج تصدیق هویت با استفاده از روش برنامه‌نویسی ژنتیک بر روی ویژگی‌های MFCC و PLP بر حسب درصد.

بردار ویژگی	خطای قبول اشتباه	خطای رد اشتباه	دقت
MFCC	۳۶/۲	۱۶/۲	۷۳/۸
PLP	۹/۵	۱۲/۷	۸۸/۹

جدول ۸: نتایج تصدیق هویت به روش GP با استفاده از خوشه‌بندی داده‌های آموزشی PLP و ساخت یک درخت GP به‌ازای هر خوشه بر حسب درصد.

روش خوشه‌بندی	تعداد خوشه‌ها	خطای قبول اشتباه	خطای رد اشتباه	دقت
بدون استفاده از خوشه‌بندی		۹/۵	۱۲/۷	۸۸/۹
Split LBG	۴	۸/۲	۸/۸	۹۱/۵
Split LBG	۸	۱۳/۳	۲۴/۳	۸۱/۲
Split LBG	۳۲	۱۰/۶	۱۳/۵	۸۸/۰

حدود ۲ ساعت برای ۳۲ خوشه افزایش می‌یابد. استفاده از خوشه‌بندی نمایندگان حاصل به ۳۲ خوشه به روش Split LBG، حدود ۱۳٪ بهبود در دقت تصدیق هویت ایجاد می‌کند.

جدول ۷ نتایج تصدیق هویت گوینده با استفاده از برنامه‌نویسی ژنتیک را بر روی داده‌های گفتاری با استفاده از ویژگی‌های MFCC و PLP با یکدیگر مقایسه می‌کند. نتایج بهتر تصدیق هویت به روش برنامه‌نویسی ژنتیک بر روی ویژگی‌های PLP نسبت به MFCC، تداخل کمتر داده‌های گویندگان در فضای ویژگی‌های PLP را نشان می‌دهد.

به‌منظور بهبود بیشتر در نتایج، خوشه‌بندی بر روی داده‌های به‌دست آمده از ویژگی PLP اعمال شده و در جدول ۸ نتایج آن ارائه شده است. نتایج حاصل نشان می‌دهد که خوشه‌بندی با تنظیمات مناسب، بهبود خوبی را در نتایج ایجاد می‌کند. مقایسه نتایج به‌دست آمده از اعمال خوشه‌بندی در برنامه‌نویسی ژنتیک نشان می‌دهد در صورتی که ویژگی‌های استخراج‌شده از داده‌های گفتاری گویندگان مختلف تداخل زیادی داشته باشند (مانند ویژگی‌های MFCC) و نتایج برنامه‌نویسی ژنتیک مطلوب نباشد، افزایش خوشه‌ها نتایج را بهبود می‌دهد. در صورتی که داده‌های گفتاری تداخل کمی داشته (مانند ویژگی‌های PLP) و نتایج برنامه‌نویسی ژنتیک مناسب باشند، خوشه‌بندی تأثیر مثبت دارد ولی افزایش خوشه‌ها لزوماً بهبود ایجاد نمی‌کند.

شکل ۹ مقایسه تأثیر ایده‌های پیشنهادی به‌منظور بهبود دقت تصدیق هویت گوینده توسط برنامه‌نویسی ژنتیک را نشان می‌دهد. مشاهده می‌شود که در گام‌های متوالی و با استفاده از ایده‌های مختلف، دقت نتایج افزایش یافته و در نهایت حدود ۴۰٪ بهبود حاصل گردیده است. بیشترین بهبود توسط خوشه‌بندی داده‌ها بر روی ویژگی‌های PLP و ساخت یک

Archive of SID

www.SID.ir

جدول ۹: مقایسه دقت و زمان مورد نیاز روش‌های مختلف در تصدیق هویت گوینده.

روش	دقت (٪)	زمان مورد نیاز (دقیقه برای هر گوینده)
Kmeans	۴۵٫۳	۲
LBG	۸۹٫۶	۳
GMM	۸۸٫۵	۵
LVQ	۷۲٫۷	۷
GP	۸۸٫۹	۱۵
GP با خوشه‌بندی	۹۱٫۵	۳۰
MLP	۸۱٫۶	۱۰
GMM - UBM	۷۵٫۲	۱۲
VQ - MAP	۸۵٫۴	۶

درخت GP به‌ازای هر خوشه به‌دست آمده است که دلیل آن، عدم وجود تداخل در داده‌های گفتاری گویندگان مختلف با استفاده از ویژگی‌های PLP است.

در مقایسه نتایج روش GP با دیگر روش‌ها، روش شبکه عصبی به‌دلیل تخمین منحنی، از لحاظ عملکرد شباهت بیشتری به برنامه‌نویسی ژنتیک داشته و بهترین گزینه برای نمایش کارایی روش GP می‌باشد. جدول ۹ زمان مورد نیاز برای آموزش مدل گویندگان و دقت نتایج تصدیق هویت گوینده با استفاده از روش‌های مختلف را نشان می‌دهد. بهترین نتیجه با استفاده از روش برنامه‌نویسی ژنتیک به‌دست آمده است که از نتایج روش‌های غیر تمایزی Split LBG و GMM نیز بهتر بوده و حدود ۲۰٪ نسبت به روش LVQ که با استفاده از نماینده‌ها و به‌صورت تمایزی عمل می‌کند، بهبود نشان می‌دهد. روش KMeans بدترین نتایج را ارائه می‌دهد که دلیل آن انتخاب نامناسب نمایندگان و عدم پوشش مطلوب فضای داده می‌باشد.

برای مقایسه با روش‌های مطرح دیگر برای تصدیق هویت گوینده، دو روش GMM - UBM و VQ - MAP نیز انتخاب شدند. برای پیاده‌سازی هر دوی این روش‌ها، مدل‌های با ۱۰۲۴ خوشه و مؤلفه گوسی استفاده و نتایج اعمال این دو روش بر روی دادگان در جدول ۹ نشان داده شده است. در ضمن برای آموزش UBM اولیه، از گویندگان مرجع موجود در پایگاه داده استفاده شده و سپس مدل برای هر گوینده به‌روز رسانی می‌شود. با توجه به این که یکی از مهم‌ترین اهداف در تصدیق هویت گوینده افزایش دقت سیستم می‌باشد، نتایج به‌دست آمده نشان می‌دهند که روش پیشنهادی مقاله که استفاده از برنامه‌نویسی ژنتیک همراه با خوشه‌بندی است، دقت بالاتری را ارائه داده و کارایی بهتری را حاصل می‌کند.

۷- نتیجه‌گیری

برنامه‌نویسی ژنتیک یکی از قدرتمندترین الگوریتم‌های تکاملی و روش‌های پردازش هوشمند است که توانایی زیادی نسبت به اغلب روش‌های یادگیری ماشین دارا می‌باشد و عملکرد آن با شبکه عصبی مقایسه می‌شود. هدف این مقاله استفاده از برنامه‌نویسی ژنتیک و تطبیق آن به‌منظور تصدیق هویت گوینده می‌باشد که بخشی از بازشناسی گوینده است.

مشکل اصلی برنامه‌نویسی ژنتیک زمان بسیار طولانی در فاز آموزش آن است. یکی از مواردی که در مدت زمان تأثیر دارد، زمان مورد نیاز برای ارزیابی داده‌های آموزشی با درخت‌های به‌دست آمده می‌باشد. در پردازش‌های گفتار، داده‌ها دارای بعد بالا و حجم زیاد می‌باشند زیرا به‌ازای هر ۲۰-۳۰ میلی‌ثانیه از گفتار، یک بردار ویژگی حاصل می‌شود. از طرفی چندین دقیقه گفتار برای آموزش سیستم مورد نیاز می‌باشد. برای کاهش زمان آموزش مدل‌ها با استفاده از روش GP، ایده استفاده از نمایندگانی به‌جای داده‌های اصلی مطرح شد. این نمایندگان با فشرده‌سازی داده‌های اصلی و با استفاده از روش‌های خوشه‌بندیِ داده‌های آموزشی مثل چندی‌سازی برداری و مدل مخلوط گوسی به‌دست می‌آیند. بنابراین نمایندگان، مراکز ثقل خوشه‌ها و میانگین‌های مؤلفه‌های گوسی که بیانگر توزیع آماری داده‌های آموزشی هستند می‌باشند که به‌جای استفاده مستقیم از خود داده‌ها استفاده می‌شوند.

در روش برنامه‌نویسی ژنتیک که روشی تمایزی است، درخت‌هایی به‌عنوان جمعیت آموزش داده می‌شوند که یک ابرصفحه ساده یا پیچیده را به‌عنوان خروجی برای جداسازی داده‌های خودی و غیر خودی تولید می‌کنند. ابرصفحه به‌دست آمده در این روش، بین داده گوینده خودی و غیر خودی تفاوت قائل می‌شود. به‌علت این که بخش‌هایی از داده‌های گفتاری و نمایندگان مورد استفاده با یکدیگر تداخل دارند، این ایده پیشنهاد شد که به‌جای آموزش تنها یک درخت GP برای هر گوینده، داده‌های آموزشی به تعداد کمی خوشه تفکیک شده و به‌ازای هر خوشه یک درخت GP آموزش داده شود. بدین ترتیب یک گوینده با چندین درخت GP مدل می‌گردد. این تدبیر، کارایی روش GP در تصدیق هویت را از نزدیک به ۷۴٪ به حدود ۸۶٪ افزایش می‌دهد. استفاده از ویژگی PLP به‌دلیل تداخل کمتر در داده‌ها، بهبود آشکاری را در نتایج ایجاد کرده است. این ایده‌ها مجموعاً منجر به بهبود عملکرد روش GP در تصدیق هویت گویندگان از حدود ۵۰٪ به حدود ۹۲٪ و نیز افزایش سرعتی در حد ۲۰ برابر در آموزش مدل‌های گویندگان به کمک روش برنامه‌نویسی ژنتیک گردید.

روش برنامه‌نویسی ژنتیک با روش‌های غیر تمایزی مانند چندی‌سازی برداری، مدل مخلوط گوسی، GMM - UBM و VQ - MAP و روش تمایزی با ارائه نمایندگان چندی‌سازی و نیز روش‌های تمایزی دیگر مثل شبکه عصبی که عملکردی مشابه GP دارند، مقایسه شد. نتایج به‌دست آمده از روش برنامه‌نویسی ژنتیک به همراه پردازش‌های بیان‌شده، همواره نسبت به دیگر روش‌ها بهتر عمل کرده و کارایی بهتری را نتیجه می‌دهد.

برای ادامه کار، جستجوی روش‌هایی برای تنظیم بهتر برنامه‌نویسی ژنتیک پیشنهاد می‌شود که بتوانند بهبود بیشتری را در نتایج تصدیق هویت گوینده ایجاد کنند.

مراجع

[1] P. Day and A. Nandi, "Robust text - independent speaker verification using genetic programming," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 285-295, Jan. 2007.

[2] J. Campbell, "Speaker recognition: a tutorial," in *Proc. IEEE*, vol. 85, no. 9, pp. 1437-1462, Sep. 1997.

[3] S. Pruzansky, "Pattern-matching procedure for automatic talker recognition," *The Journal of the Acoustical Society of America*, vol. 35, no. 3, pp. 354-358, 1963.

[4] P. Bricker, "Statistical techniques for talker identification," *Bell System Technical J.,* vol. 50, no. 4, pp. 1427-1454, Apr. 1971.

[5] K. Li and G. Hughs, "Talker identification as they appear in correlation matrics of continous speech spectral," *The J. of the Acoustical Society of America*, vol. 55, pp. 833-837, 1974.

[6] M. Sambur, "Speaker recognition and verification using linear prediction analysis," *The J. of the Acoustical Society of America,* vol. 53, p. 354, 1973.

[7] B. Juang and F. Soong, "Speaker recognition based on source coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, pp. 613-616, 3-6 Apr. 1990.

Archive of SID
www.SID.ir

[8] A. Poritz, "Linear predictive hidden markov models and the speech signal," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1291-1294, May 1982.
[9] B. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. 29, no. 2, pp. 254-272, Apr. 1981.
[10] T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative evaluation of various MFCC implementations on the speaker verification task," in *Proc. 10th Int. Conf. on Speech and Computer, SPECOM*, vol. 1, pp. 191-194,Oct. 2005.
[11] M. Siafarikas, T. Ganchev, and N. Fakotak, "Wavelet packet based speaker verification," in *Proc. ISCA-ICSLP*, vol. 1, pp. 257-264, Toledo, Spain, Jun. 2004.
[12] H. Ezzaidi, J. Rout, and D. Shaughnessy, *Combining Pitch and MFCC for Speaker Recognition Systems*, University Quebc, Ermetis, Canada, 2001.
[13] D. Reynolds and R. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, Jan. 1995.
[14] S. Stafford, A Gaussian Mixture Model Based Speaker Verification System That Captures Sequential Information, M. S. Thesis, 2005.
[15] M. Faouzi BenZeghiba and H. Bourlard, "User-customized password speaker verification using multiple reference and background models," *Speech Communication*, vol. 48, no. 9, pp. 1200-1213, 2006.
[16] S. Adinarayanan, *Text-Independent Speaker Verification Using Support Vector Machine*, M.Sc. Thesis, Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, 2005.
[17] T. Matsui and S. Furui, "Comparision of text independent Speaker recognition methods using VQ-distortion and discrete/continuous HMMs," in *Proc. ICSLP*, vol. 2, pp. 157-160, 1992.
[18] Q. Li and B. Juang, "Speaker verification using verbal information verification for automatic enrollment," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing,* vol. 1, pp. 650-658, 1998.
[19] X. Li and K. Chen, "Mandarin verbal information verification," in *Proc. IEEE Int Conf. on Acoust., Speech, and Signal Processing*, vol. 1, pp. 833-836, 13-17 May 2002.
[20] V. Wan, *Speaker Verification Using Support Vector Machines*, Ph. D. Thesis, University of Sheffield, 2003.
[21] D. Reynolds, T. Quatieri, and R. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19-41, Jan. 2000.
[22] Q. Hong and S. Kwong, "A discriminative training approach for text-independent speaker recognition," *Signal Processing*, vol. 85, no. 7, pp. 1449-1463, Jul. 2005.
[23] V. Hautamaki, T. Kinnunen, I. Karkkainen, J. Saastamoinen, M. Tuononen, and P. Franti, "Maximum a posteriori adaptation of the centroid model for speaker verification," *IEEE Signal Processing Letters*, vol. 15, pp. 162-165, 2008.
[24] E. Avci, "A new optimum feature extraction and classification method for speaker recognition: GWPNN," *Expert Systems with Applications*, vol. 32, no. 2, pp. 485-498, Feb. 2007.
[25] K. Faraoun and A. Boukelif, "Artificial immune systems for text-dependent speaker recognition," *J. of Computer Science*, vol. 5, no. 4, pp. 19-26, Dec. 2006.
[26] R. Wouhaybi and M. Al-Alaoui, "Comparison of neural networks for speaker recognition," in *Proc. Sixth IEEE Int. Conf. on Electronics, Circuits and Systems, ICECS'99*, vol.1, pp. 125-128, Pafos, Cyprus, Sep. 1999.
[27] S. Lung, "Efficient text independent speaker recognition with wavelet feature selection based multilayered neural network using supervised learning algorithm," *Pattern Recognition*, vol. 40, no. 10, pp. 3616-3620, Dec. 2007.
[28] K. Markov and S. Nakagawa, "Text-independent speaker recognition using non-linear frame likelihood transformation," *Speech Communication*, vol. 24, no. 3, pp. 193-209, Jun. 1998.
[29] R. Saeidi, H. Sadegh Mohammadi, and M. Khalaj Amirhosseini, "An efficient GMM classification post-processing method for structural gaussian mixture model based speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*,, vol. 6, pp. 909-912, May 2006.
[30] S. Bengio and J. Mariethoz, "Learning the decision function for speaker verification," in *Proc. IEEE Internat. Conf. on Acoust., Speech, and Signal Processing*, vol. 1, pp. 425-428, May. 2001.
[31] N. Mirghafori and M. Hebert, "Parameterization of the score the thereshold for s text-dependent adaptive speaker verification system," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, pp. 361-364, Montreal, Canada, May 2004.
[32] N. Cramer, "A representation for the adaptive generation of simple sequential programs," in *Proc. of the 1st Int. Conf. on Genetic Algorithms*, vol. 1, pp. 183-187, Jul. 1985.
[33] J. Koza, Genetic Programming: on the Programming of Computers by Means of Natural Selection: MIT Press, 1992.
[34] G. Liai, Z. Shuguang, Z. Yongjie, and L. Lihua, "A new codebook design method based on genetic programming," *in Proc. IEEE 8th Int. Conf. on Electronic Measurement and Instruments. ICEMI'2007*, vol. 3, pp. 250-253, Aug. 2007.
[35] K-Means Algorithm. Available:http://en.wikipedia.org/wiki/k-means_algorithm
[36] F. Soong, A. Rosenberg, L. Rabiner, and B. Juang, "A vector quantization approach to speaker recognition," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process,* vol. 1, pp. 387-390, Apr. 1985.
[37] T. Kinnunen, J. Saastamoinen, V. Hautamaki, M. Vinni, and P. Franti, "Comparative evaluation of maximum a posteriori vector quantization and gaussian mixture models in speaker verification," *Pattern Recognition Letters*, vol. 30, no. 4, pp. 341-347, Mar. 2009.
[38] S. Furui, "Section 1.7: Speaker Recognition," in Survey of the State of the Art in Human Language Technology, ed: Cambridge University Press, 1996.
[39] S. Furui, Digital Speech Processing Synthesis and Recognition, New York: Marcel Dekker Inc., 1989.
[40] R. Ramachandran, K. Farrell, R. Ramachandrana, and R. Mammone, "Speaker recognition-general classifier approaches and data fusion methods," *Pattern Recognition*, vol. 35, pp. 2801-2821, Dec. 2002.
[۴۱] ع. صادقی نائینی، *بازشناسی گوینده مبتنی بر هم‌جوشی و فضای حالت ساخته‌شده از گویندگان ویژه*، پایان‌نامه کارشناسی ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، ۱۳۸۵.
[42] A. Higgins, L. Bahler, and J. Porter, "Speaker verification using randomized phrase prompting," *Digital Signal Processing*, vol. 1, no. 2, pp. 89-106, 1991.
[43] About FAR, FRR and EER. Available:http://www.bioid.com/sdk/docs/about_eer.htm

سعیده سادات سدیدپور در سال ۱۳۸۵ مدرک کارشناسی مهندسی کامپیوتر خود را از دانشگاه شهید بهشتی و در سال ۱۳۸۸ مدرک کارشناسی ارشد مهندسی کامپیوتر خود را از دانشگاه صنعتی امیرکبیر دریافت نمود. نام‌برده کارشناسی دیگری را در رشته هومیوپاتی از دانشگاه بین‌المللی برکلی اخذ نموده است. ایشان از سال ۱۳۸۸ تا کنون به تدریس در دانشگاه‌های تهران و قم و نیز فعالیت تحقیقاتی در مرکز تحقیقات مخابرات ایران مشغول بوده و زمینه‌های علمی مورد علاقه نام‌برده شامل الگوریتم‌های هوشمند، تکاملی و محاسبات زیستی و پردازش‌های مختلف شامل گفتار و تصویر می‌باشد.

محمد مهدی همایونپور تحصیلات خود را در مقاطع کارشناسی در رشته الکترونیک در سال ۱۳۶۶ از دانشگاه صنعتی امیرکبیر و در مقطع کارشناسی ارشد در سال ۱۳۶۹ در رشته مخابرات دانشگاه خواجه نصیرالدین طوسی بپایان رسانیده است. نام‌برده کارشناسی ارشد دیگری را در رشته فونتیک از دانشگاه سوربون جدید و دکترای خود را در رشته مهندسی برق از دانشگاه پاریس ۱۱ در سال ۱۳۷۴ اخذ نموده است. ایشان از سال ۱۳۷۴ عضو هیأت علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر و در حال حاضر دانشیار این دانشکده هستند. زمینه‌های مورد علاقه نام‌برده پردازش گفتار، پردازش سیگنال دیجیتال، چند رسانه‌ای، اتوماسیون صنعتی و طراحی سخت افزار است.

مهدی فسنقری تحصیلات خود را در سال ۱۳۸۵ در مقطع کارشناسی ارشد مهندسی فناوری اطلاعات از دانشگاه تربیت مدرس به پایان رسانده و هم‌اکنون دانشجوی مقطع دکتری مهندسی صنایع دانشگاه تهران و عضو هیأت علمی مرکز تحقیقات مخابرات ایران می‌باشد. نام‌برده در دانشگاه‌های تهران در حوزهٔ مهندسی فناوری اطلاعات مشغول به تدریس بوده و زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: برنامه ریزی توسعه فناوری اطلاعات، معماری سازمانی، حاکمیت فناوری اطلاعات، تصمیم‌گیری‌های پیچیده در حوزه فناوری اطلاعات و نرم رایانش و کاربردهای آن در فناوری اطلاعات.

Archive of SID

www.SID.ir