بسم الله الرحمن الرحیم

دانشگاه بیرجند

دانشکده مهندسی برق و کامپیوتر

پایان‌نامه کارشناسی ارشد در رشته مهندسی برق (مخابرات)

تایید گوینده با استفاده از ماشین های بردار پشتیبان

استاد راهنما:

دکتر حسن فرسی

استاد مشاور:

دکتر حمید فرخی

نگارش:

زهرا محمدپور

بهمن ماه ۱۳۹۱

تأییدیه هیات داوران

(برای پایان‌نامه)

یک نسخه اصل فرم مربوطه

تقدیم به مادر عزیز ، مهربان و دلسوزم

و

تقدیم به خواهر خوبم

سپاسگزاری

پیش از هر چیز از خداوند به خاطر تمامی نعمت هایی که هدیه کرده سپاسگزارم. از استاد ارجمند، جناب آقای دکتر حسن فرسی که در طول انجام این پروژه با صبر و شکیبایی همواره مشوق و راهنمای من بوده اند، صمیمانه تشکر و قدردانی می کنم. همچنین از استاد مشاور گرامی جناب دکتر حمید فرخی کمال تشکر و امتنان را دارم.

در پایان از خانواده عزیزم بخاطر حمایت ها، دلگرمی ها و زحماتشان سپاسگزارم.

چکیده

تایید گوینده با استفاده از ماشین های بردار پشتیبان

به وسیله:

زهرا محمدپور

تعداد قابل توجهی از تحقیقات در استفاده از تکنیک های مدل های مخلوط گوسی (GMM) در سیستم های تایید گوینده مستقل از متن تمرکز دارد که جدیدترین عملکردها را ارائه می دهد. اخیرا استفاده از ماشین بردار پشتیبان (SVM) در سیستم های تایید گوینده، منجر به عملکرد بهتری نسبت به روش استاندارد GMM شده است. SVM یک طبقه بند تبعیضی است که نیاز به دو کلاس مثبت و منفی برای آموزش مدل گوینده دارد و بر اساس حداکثر حاشیه است که به موجب آن یک ابرصفحه تلاش می کند که کلاس ها را در فضایی با ابعاد بالا جدا سازد. ماشین های بردار پشتیبان استفاده شده در تایید گوینده پتانسیل بالایی را نشان داده است, به ویژه زمانی که در سیستم های ترکیبی برای تکمیل تکنیک های GMM استفاده شود. هدف از انجام این پایان نامه بهبود عملکرد سیستم تایید گوینده مستقل از متن با استفاده از روش جدید SVM است.

واژگان کلیدی: تایید گوینده، شناسایی گوینده، ماشین بردار پشتیبان، مدل مخلوط گوسی، مستقل از متن.

فهرست مطالب

عنوان **صفحه**

فهرست علایم و نشانه‌ها ..... ج

فهرست جدول‌ها ..... د

فهرست شکل‌ها ..... ه

فصل اول: مقدمه ۱

۱-۱- مقدمه ۱

۱-۲- شناسایی هویت گوینده ۲

۱-۳- سیستم های شناسایی گوینده ۳

۱-۴- کاربردهای شناسایی گوینده ۴

۱-۵- مراحل تشکیل سیستم های شناسایی گوینده ۵

۱-۶- تاریخچه مدل های شناسایی گوینده ۷

۱-۶-۱- مدل های قالبی ۸

۱-۶-۲- مدل های اتفاقی ۸

۱-۶-۳- مدل های جدا کننده ۹

۱-۷- هدف از این پایان نامه ۱۰

فصل دوم: استخراج ویژگی ۱۱

۲-۱- مقدمه ۱۱

۲-۲- فرآیند front-end ۱۲

۲-۲-۱- ضبط صدا و ذخیره سازی ۱۲

۲-۲-۲- پارامترسازی صدا و استخراج ویژگی ۱۳

۲-۲-۳- ضرایب مدل پیشگویی خطی به عنوان ویژگی ۱۵

۲-۳- تبدیل کپستروم و توانایی های آن ۱۶

۲-۳-۱- ضرایب کپستروم در مقیاس مل به عنوان ویژگی ۱۸

فصل سوم: مدل مخلوط گوسی و الگوریتم بیشینه سازی امید ۲۱

۳-۱- تخمین پارامتر ۲۱

۳-۲- الگوریتم بیشنیه سازی امید (EM) ۲۳

۳-۳- الگوریتم بیشینه سازی امید در تخمین حداکثر احتمال ۲۴
۳-۳-۱- الگوریتم بیشینه سازی امید برای مدل مخلوط گوسی ۲۵
۳-۴- کاربرد مدل مخلوط گوسی در بازشناسی گوینده ۲۷
۳-۴-۱- کاربرد در تعیین هویت گوینده ۲۷
۳-۴-۲- کاربرد در تایید هویت گوینده ۲۸

فصل چهارم: مدل ماشین بردار پشتیبان ۳۲

۴-۱- مقدمه ۳۲
۴-۲- ماشین بردار پشتیبان برای مسائل جدا پذیر خطی ۳۲
۴-۳- ماشین بردار پشتیبان برای مسائل جدا پذیر غیر خطی ۳۵
۴-۴- ماشین بردار پشتیبان غیر خطی ۳۶
۴-۵- کرنل ماشین بردار پشتیبان ۳۷
۴-۶- سوپر وکتور ۳۹
۴-۶-۱- کرنل خطی مدل های مخلوط گوسی ۳۹
۴-۶-۲- کرنل ضرب داخلی مدل های مخلوط گوسی ۴۱

فصل پنجم: پیاده سازی ،بحث و ارائه نتایج ۴۲

۵-۱- معیارهای ارزیابی سیستم بازشناسی گوینده ۴۲
۵-۱-۱- منحنی تغییرات خطای بازشناسی(DET) ۴۲
۵-۱-۲- تابع هزینه تشخیص (DCF) ۴۴
۵-۱-۳- نرخ خطای یکسان (EER) ۴۶
۵-۲- نتایج ۴۶

فصل ششم: نتیجه گیری ۴۹

۶-۱- نتیجه گیری ۴۹
۶-۲- پیشنهادات و کارهای آینده ۵۰

فهرست مراجع ۵۱

واژه نامه فارسی به انگلیسی ۵۴

واژه نامه انگلیسی به فارسی ۵۵

فهرست علایم و نشانه‌ها

عنوان	علامت اختصاری
تعداد کانال های فیلتر بانک	N
بردار میانگین	μ_i
ماتریس کواریانس	Σ_i
وزن های مخلوط	w_i
بعد داده ها	D
بردارهای آموزش	X
پارامترهای مدل چگالی	θ
مقدار آستانه	T
فاکتور ارتباط	r
ضریب انطباق وابسته به دیتا	α_k
ضریب انطباق وابسته به مخلوط	n_k
بردار عمود بر ابرصفحه	W
عرض از مبدا	b
ضریب جریمه	C
متغیر کمکی (Slack Variable)	ξ
بردار پشتیبان	SV
تابع کرنل	K

فهرست جدول‌ها

عنوان **صفحه**

جدول ۵-۱: مقادیر استاندارد پارامترهای DCF ۴۵
جدول ۵-۲: نرخ خطای یکسان و حداقل تابع هزینه تشخیص در دو روش با ۳۲ مخلوط گوسی ۴۸
جدول ۵-۳: نرخ خطای یکسان و حداقل تابع هزینه تشخیص در دو روش با ۶۴ مخلوط گوسی..... ۴۸

فهرست شکل‌ها

عنوان **صفحه**

شکل ۱-۱- شاخه های مختلف علم پردازش گفتار ۲
شکل ۱-۲- سیستم شناسایی گوینده و اجزای تشکیل دهنده ۵
شکل ۱-۳- مراحل ثبت نام و تصدیق سیستم های شناسایی گوینده ۶
شکل ۱-۴- سیر تکاملی مدل های سیستم شناسایی گوینده ۷
شکل ۲-۱- پنجره گذاری بر روی سیگنال صحبت ۱۴
شکل ۲-۲- تبدیلات کپستروم ۱۷
شکل ۲-۳- ساختار متعارف فیلتر بانک ۱۹
شکل ۳-۱- سیستم تعیین هویت گوینده با مدل مخلوط گوسی ۲۷
شکل ۳-۲- سیستم تایید هویت گوینده با آزمون نسبت احتمال ۲۹
شکل ۳-۳- ساختن مدل گوینده با وفق دادن مدل پس زمینه ۳۰
شکل ۴-۱- دسته بندی خطی نمونه ها ۳۳
شکل ۴-۲- دسته بندی نمونه های جدانشدنی ۳۵
شکل ۴-۳- روش محاسبه بردارهای سوپروکتور ۴۰
شکل ۵-۱- توابع چگالی احتمال دو کلاس با مرز جداساز ۴۳
شکل ۵-۲- منحنی DET ۴۳
شکل ۵-۳- نمودار تابع هزینه تشخیص(DCF) برای ویژگی های عملیاتی سیستم ۴۵
شکل ۵-۴- منحنی DET و نقاط EER برای هر دو روش با ۳۲ مخلوط گوسی ۴۷
شکل ۵-۵- منحنی DET و نقاط EER برای هر دو روش با ۶۴ مخلوط گوسی ۴۷

فصل ۱– مقدمه

۱–۱– مقدمه

پیشرفت مستمر فن آوری های روز دنیا نیاز به اقدامات امنیتی را برای این فن آوری ها حیاتی کرده است تا این اطمینان را بدهد که به دسترسی های غیر مجاز آسیب پذیر نیست. چنین اقدامات امنیتی نه تنها برای حفاظت از اطلاعات دیجیتالی ذخیره شده افراد لازم است , بلکه برای حفاظت از یک کشور و مردم آن نیز ضروری است. این نیاز به توسعه فن آوری های تایید هویت بیومتریک[۱] منجر شده است. بیومتریک یا علم زیست سنجی در واقع ویژگی های فیزیکی و رفتاری یک فرد هستند . از جمله ویژگی های بیومتریک می توان به تشخیص چهره و عنبیه، اثر انگشت، هندسه دست، شیوه راه رفتن و آنالیز صدا اشاره کرد.

صحبت و مکالمه به عنوان یک ویژگی بیومتریک غالب ترین راه ارتباط اجتماعی انسان به شمار می آید و به راحتی و بدون نیاز به فن آوری خاصی تولید می شود. هم اکنون کلمات از رسانه هایی مانند تلفن، اینترنت، تلویزیون و رادیو انتقال می یابند. این واقعیت اهمیت زبان و گفتار را مشخص می کند. از طرفی انسان برای برقراری ارتباط با سیستم های ساخته دست خود نیازمند برقراری ارتباطی از این نوع می باشد و در صورت برقراری چنین ارتباطی به راحتی میتواند دستورات خود را برای کنترل سیستم ارسال کند. علم پردازش گفتار به حل این مساله می پردازد.

پردازش گفتار را میتوان به شش بخش تشخیص گفتار[۲]، شناسایی گوینده[۳] ،کد کردن گفتار[۴]، تجزیه و تحلیل صدا[۵] ، ترکیب گفتار[۶] و افزایش کیفیت گفتار[۷] تقسیم بندی کرد [۱]. شکل زیر این تقسیم بندی را نشان می دهد.

---

[1] Biometric
[2] Speech Recognition
[3] Speaker Recognition
[4] Speech Coding
[5] Voice Analysis
[6] Speech Synthesis
[7] Speech Enhancement

شکل ۱-۱ شاخه های مختلف علم پردازش گفتار

سیستم های تشخیص گفتار یک سخن ادا شده را به حروف ،کلمات یا جملات تفکیک می کنند. کد کردن گفتار یک حالت به خصوص از فشرده سازی است و در بخش ارتباط تلفنی استفاده می شود. بخش آنالیز صدا به مطالعه گفتار برای مسایلی غیر از مفاهیم زبانی می پردازد. یک مثال از این مورد می تواند آنالیز صدای بیماری باشد که تارهای صوتی او مورد عمل جراحی قرار گرفته است. در سنتز گفتار به تولید مصنوعی صحبت از طریق کامپیوتر پرداخته می شود. سیستم های ادراک گفتار در حضور عواملی مانند نویز با مشکل مواجه می شوند، هدف سیستم های افزایش کیفیت گفتار کاهش اثرات مخرب چنین عواملی است.

۱-۲- شناسایی هویت گوینده

بطور کلی فرآیند شناسایی گوینده یا تشخیص صدا[1] به جداسازی افراد از روی صدای آنها گفته می شود. دو شخص مختلف به علت تفاوت در شکل حنجره، تارهای صوتی و بقیه اندامهای تولید صوت صدای یکسانی ندارند. علاوه بر این تفاوت های فیزیکی مسائلی مانند نحوه تلفظ، لهجه و ریتم ادای کلمات نیز باعث تمایز گفتار اشخاص می شود. عمل شناسایی گوینده در انسان به راحتی انجام می شود و این تشخیص به کمک تفاوت هایی صورت می گیرد. تفاوت های دسته اول مربوط به آناتومی حنجره و دستگاه صوتی هستند اما دسته دوم دارای منشا ذهنی و تفاوت در یادگیری هستند. به عنوان مثال تفاوت در منطقه سکونت و جغرافیای زندگی اشخاص موجب تفاوت در لهجه افراد می شود. در اینجا باید عنوان کرد که اگر چه دو فرد بنا به علت های عنوان شده صدای متفاوتی دارند اما صدای یک شخص نیز

[1] Voice Recognition

به نوبه خود در زمان های متفاوت ممکن است دچار تغییر شود. به عنوان مثال شخصی که دچار سرماخوردگی شده به علت تغییر شکل اندام های تولید صوت، صدای متفاوتی تولید می کند. اشتباه در تشخیص هویت افراد در هنگام برقراری ارتباط تلفنی مثال دیگری در این زمینه است. بسیاری از افراد تجربه چنین اشتباهی را داشته اند و در ابتدای تماس تلفنی در تعیین هویت طرف مقابل دچار اشتباه شده اند.

همانطورکه بیان شد هدف پردازش گفتار و شناسایی گوینده، هوشمند نمودن ماشین به منظور توانا کردن آن در درک گفتار و متمایز کردن گوینده های متفاوت است. بنابراین میتوان از همان ویژگی هایی که یک انسان برای تمایز افراد مختلف از آن بهره می برد برای توانایی ماشین از آنها استفاده کرد . برخلاف حیطه شناسایی گفتار پیوسته ، دقت سیستم های شناسایی گوینده با دقت انسان در تشخیص گوینده برابری می کند.

۱-۳- سیستم های شناسایی گوینده

به طور کلی تمایزهای موجود در صدای اشخاص به دو دسته تغییرات بین گویندگان [1] و تغییرات درون گوینده[2] تقسیم می شوند [۲]، تغییرات بین گویندگان به معنای تفاوت هایی است که در قطعات گفتاری گویندگان مختلف وجود دارد و دسته دوم تغییراتی است که در قطعات گفتاری مربوط به یک گوینده رخ می دهد. بدیهی است که ویژگیهای استخراج شده باید حاوی اطلاعات از دسته اول باشند و تا حد ممکن تاثیر پذیری کمتری از دسته دوم داشته باشند. تغییرات درون گوینده زیر مجموعه ای از تغییرات نشست[3] [۴،۳] بشمار می روند. به طور رسمی به تغییرات صورت گرفته در سیگنال صحبت یک نفر در دو زمان مختلف متغیر بودن کانال گفته می شود.

سیستم های شناسایی گوینده به لحاظ انجام وظیفه به سه دسته تایید هویت گوینده[4] ، تشخیص هویت گوینده[5] و قطعه بندی صدای گوینده های مختلف[6] تقسیم می شوند. تایید هویت گوینده فرآیندی است که تعیین می کند یک نمونه صحبت متعلق به گوینده مورد نظر است یا خیر. عمل تایید گوینده اغلب با تشخیص هویت گوینده اشتباه گرفته می شود، با این وجود تفاوت های ظریفی بین این دو وجود

[1] intra speaker variability
[2] inter speaker variability
[3] Session Variability
[4] Speaker Verification
[5] Speaker Identification
[6] Speaker Segmentation

دارد. در تشخیص گوینده، هویت فردی که نمونه صدایش تهیه شده از بین یک گروه بسته از گوینده ها تعیین می شود. ولی تایید شامل یک تصمیم دودویی است که مشخص می کند صحبت داده شده متعلق به گوینده ای که ادعا شده می باشد یا خیر. ماهیت باز بودن مجموعه فرآیند تایید، این عمل را ذاتا مشکل تر می کند. در قطعه بندی صدای گوینده های مختلف باید مشخص شود کدام گوینده چه زمانی در یک فایل صوتی صحبت می کند.

سیستم های شناسایی گوینده از دید دیگری به دو نوع وابسته به متن[۱] و مستقل از متن[۲] تقسیم بندی می شوند. در سیستم های وابسته به متن ،گوینده باید یک جمله معین شده توسط سیستم را ادا کند. این جمله می تواند یک جمله ثابت باشد تا به صورت تصادفی از یک مجموعه متشکل از کلمات انتخاب شود. اما در نوع مستقل از متن ،گوینده نسبت به جمله ای که ادا می کند مختار است. بنابراین داده های آموزش و تست ممکن است از نظر محتوا کاملا متفاوت باشند. با توجه به توانایی انسان در تشخیص مستقل از متن ، جهت گیری مقالات بیشتر به سمت این حالت از شناسایی گوینده است . به علت عدم وجود دانش پیشین [۳] از جمله ای که توسط یک گوینده بیان می شود تشخیص مستقل از متن به مراتب سخت تر از تشخیص وابسته به متن است.

۱-۴- کاربردهای شناسایی گوینده

تصدیق هویت گوینده می تواند به عنوان یکی از روشهای بیومتریک مورد استفاده قرار گیرد. مزیت سیستم هایی که از روشهای شناسایی گوینده بهره می برند نسبت به سایر روش ها در دسترس بودن تجهیزات و زیر ساخت های لازم برای ضبط و انتقال داده های صوتی است. امروزه با توجه به وسعت شبکه های تلفن و اینترنت به راحتی میتوان صدای اشخاص را به سرعت به مکان های مختلف ارسال نمود. در حالی که سیستم های بیومتریکی دیگری مانند تشخیص عنبیه یا اثر انگشت از چنین امکانی برخوردار نیستند. از طرفی هزینه ی تجهیزات جمع آوری و پردازش داده در شناسایی گوینده کمتر است. علیرغم چنین مزیت هایی، شناسایی گوینده در این حوزه مشکلاتی نیز پیش رو دارد که باید به آنها پرداخته شود. پخش صدای ضبط شده از یک شخص در جهت جعل هویت و به منظور کسب مجوز عبور از یک سیستم شناسایی هویت از جمله این مشکلات است [۵].

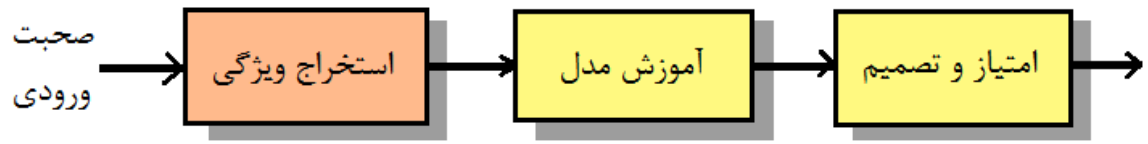
[1] Text-Dependent
[2] Text-Independent
[3] prior knowledge

حیطه دیگری که شناسایی گوینده در آن کاربرد دارد، جرم شناسی است که اخیراً بسیار مورد توجه قرار گرفته است. همانطور که گفته شد صدای انسان ها مانند اثر انگشت میتواند موجب تعیین هویت افراد گردد. بسیاری از صحبت ها و هماهنگی های افراد مجرم از طریق مکالمه های تلفنی صورت می گیرد. میزان شباهت صدای یک فرد مظنون و یک قطعه صدای ضبط شده می تواند برای اثبات جرم مورد استفاده قرار گیرد[۷،۶].

سیستم های امنیتی که تنها فرد یا افراد خاصی مجاز به ورود به آن سیستم هستند، دسته دیگری از کاربردهای شناسایی در محل[1] گوینده است. سیستم های تلفنبانک یک مثال بسیار مشهور در این مورد هستند.

۱–۵– مراحل تشکیل سیستم های شناسایی گوینده

شکل ۱-۲ یک سیستم شناسایی گوینده را در حالت کلی نشان می دهد. ورودی سیستم سیگنال صوت و خروجی آن عددی است که امتیازی را به این سیگنال نسبت می دهد. براساس این امتیاز است که تشخیص صورت می گیرد.



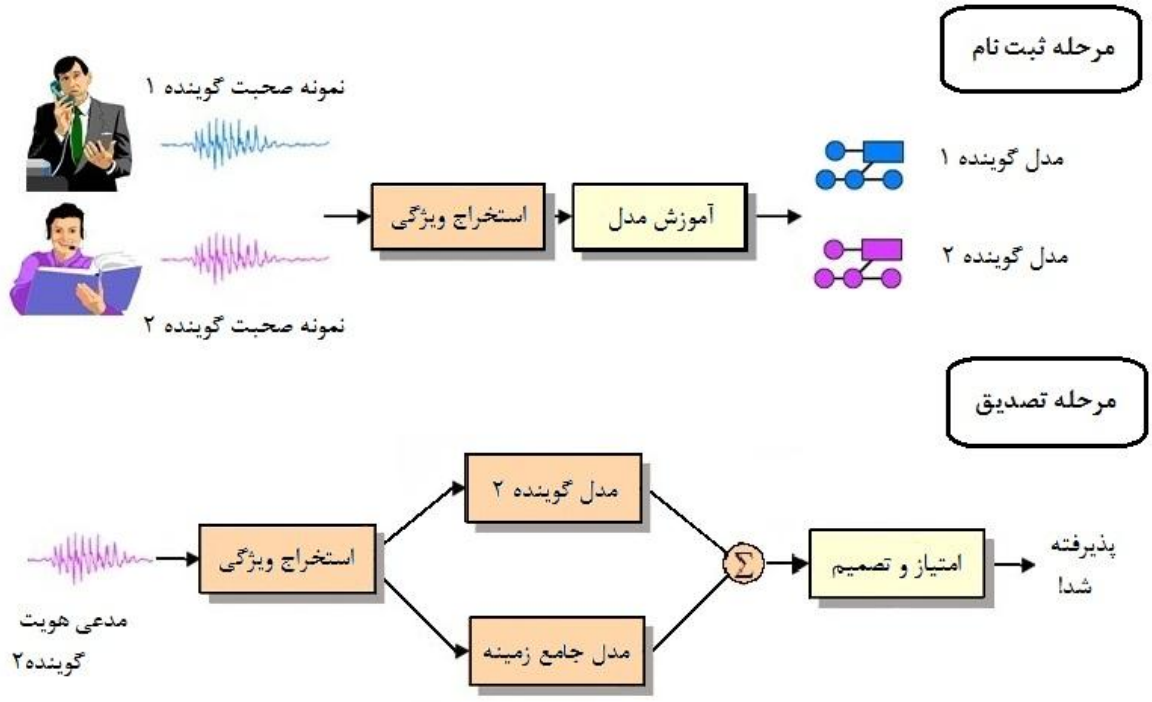
شکل ۱-۲ سیستم شناسایی گوینده و اجزای تشکیل دهنده

هر سیستم دارای دو مرحله ثبت نام [2] و تصدیق [3] است در مرحله ثبت نام یا جمع آوری[4] هر شخص یک عبارت یا کلمه بیان می کند . نمونه صدا از آنالوگ به دیجیتال تبدیل می شود، سپس ویژگی های صدای شخص استخراج می شود و یک مدل برای گوینده آموزش داده می شود. از این مدل در مرحله تصدیق استفاده می شود و امتیازی برای آن محاسبه می شود. با مقایسه امتیاز صدای ورودی و صدای

[1] On-site
[2] Enrollment
[3] Recognition
[4] Collection

ذخیره شده، سیستم عمل تصمیم گیری را انجام می دهد. این دو مرحله در شکل ۱-۳ نشان داده شده است.

هر قطعه از سیگنال صبحت از ابتدای ورود به سیستم تا محاسبه امتیاز ،چند مرحله را طی میکند. همانگونه که در شکل ۱-۳ ملاحظه می شود قدم ابتدایی در هر کدام از دو مرحله استخراج ویژگی از سیگنال صبحت است. در این مرحله ابتدا قطعه سیگنال ورودی به فریم هایی تقسیم بندی میشود. سپس از هر فریم اطلاعاتی استخراج میشود که معمولاً به صورت یک بردار است. این بردارها چنان بدست می آیند که حاوی اطلاعات جدا کننده مربوط به اشخاص است.



شکل ۱-۳ مراحل ثبت نام و تصدیق سیستم های شناسایی گوینده

مدل ها می توانند به دو دسته تولیدی[1] یا تبعیضی[2] تقسیم بندی شوند. مدل های نوع اول تابع توزیع احتمال ویژگی های یک گوینده را مدل می کنند، اما در نوع دوم مرز بین گوینده ها مدل می شود. مدل

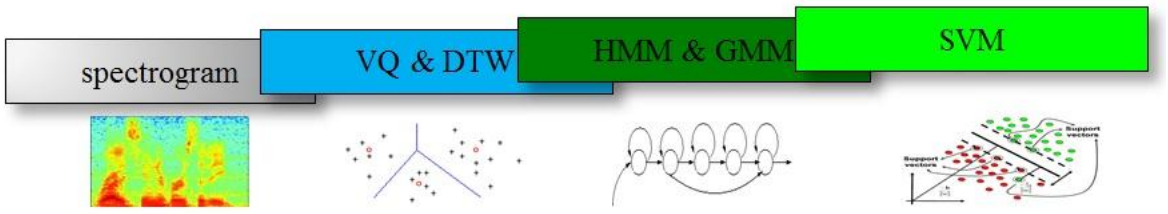
[1] Generative
[2] Discriminative

مخلوط گوسی, مدل های ماشین بردار پشتیبان و شبکه های عصبی نمونه هایی از مدل های نوع دوم می باشند.

تحقیقات شناسایی گوینده در هرکدام از سه بخش استخراج ویژگی، آموزش مدل و امتیاز صورت می گیرد. برخی پژوهش ها به ارائه شیوه های جدید در بخش مدل سازی و استخراج ویژگی یا روش های طبقه بندی[1] می پردازند و برخی افزایش کیفیت و مقاومت روش ها در برابر عامل تغییرات نشست را بررسی می کنند.

۱-۶- تاریخچه مدل های شناسایی گوینده

سابقه پژوهش های صورت گرفته در زمینه شناسایی گوینده به سالهای ۱۹۵۰ باز می گردد [۹،۸] و تاکنون در حدود ۳۵۰۰ مقاله در زمینه شناسایی گوینده نوشته شده است [۱۰]. بنابراین میتوان به اهمیت این شاخه از پردازش گفتار و سابقه آن پی برد. در این فصل مروری مختصر بر مدلهای متفاوت ارائه شده در زمینه شناسایی گوینده خواهیم داشت. مدلهای کلاسیک را میتوان به دو دسته مدلهای قالبی[2] و مدلهای اتفاقی[3] [۱۱] دسته بندی کرد. در مدل های نوع اول بردارهای ویژگی در حالت آموزش و تست مستقیماً با هم مقایسه می شوند. این مقایسه براساس این فرض صورت می گیرد که بردارهای تست یک تصویر غیر دقیق از بردارهای آموزشی هستند. در مدل های اتفاقی هر گوینده با یک مدل احتمالی مدل میشود. این مدل میتواند از نوع پارامتریک یا غیرپارامتریک باشد. در فاز آموزش پارامترهای مدل پارامتریک بدست می آید. پس از آموزش تشخیص با استفاده از احتمال مدل به شرط سخن ادا شده صورت می گیرد. مدل مخفی مارکوف و مدل مخلوط گوسی می توانند مثال هایی از اینگونه مدل ها باشند که به ترتیب برای سیستم های وابسته به متن و مستقل از متن استفاده می شوند. شکل زیر سیر تحول مدل در سیستم های شناسایی گوینده را نشان می دهد.



۱-۴ سیر تکاملی مدل های سیستم شناسایی گوینده

[1] Classification
[2] Template models
[3] Stochastic models

۱-۶-۱- مدل های قالبی

مدل کوانتیزه کردن بردار[1] یکی از ساده ترین مدل هایی است که در دهه ۸۰ میلادی با بهره گیری از روش های فشرده سازی داده ارایه شد [۱۲]. این مدل در دسته مدلهای قالبی قرار دارد. در این مدل مقدار کمینه فاصله بردارهای آموزش و تست محاسبه می گردد. هر چه این مقدار بین داده های آموزشی و داده های تست کمتر باشد، احتمال آنکه این داده ها توسط یک شخص بیان شده باشد بیشتر است. به جای استفاده از تمام بردارهای آموزشی میتوان این بردارها را توسط یک روش خوشه بندی به چند دسته تقسیم کرد و اختلاف فاصله نماینده های هر دسته کتاب کد [2]را با داده های تست مقایسه کرد.

۱-۶-۲- مدل های اتفاقی

مدل مخفی مارکف[3] (HMM) از دسته مدل های اتفاقی است که در بازشناسی گفتار و گوینده استفاده شده است. این مدل تغییرات دینامیکی و زمانی مربوط به ویژگیهای استخراج شده از سیگنال گفتار را مدل می کنند. مدل مخفی مارکف پارامترهایی دارد که در فاز آموزش و بوسیله الگوریتم هایی معین می گردند. در شناسایی گوینده برای هر شخص یک مدل آموزش داده می شود در بازشناسی وابسته به متن پارامترهای یک مدل مخفی مارکوف چند حالته مورد آموزش واقع می شوند. اما در حالت مستقل از متن یک مدل تک حالته که همان مدل مخلوط گوسی است، بردارهای ویژگی اتفاقی سیگنال گفتار را مدل می کند. در مراجع [۵] میتوان پژوهش هایی در این زمینه را ملاحظه کرد.

Rose , Reynolds در سال ۱۹۹۵ طی مقاله ای از مدل مخلوط گوسی برای هدف تعیین هویت گوینده استفاده کردند [۱۳] . در این مقاله با استفاده از ویژگی های طیفی زمان کوتاه به دست آمده از سیگنال صحبت و الگوریتم بیشینه سازی امید[4] پارامترهای مدلی که برای شخص در نظر گرفته می شود آموزش داده می شوند و این مدل به منظور تعیین هویت گوینده مورد استفاده قرار می گیرد. Reynolds سپس در سال ۲۰۰۰ با استفاده از مدل مخلوط گوسی و روش ارائه شده برای وفقی کردن آن توانست روشی موثر برای تایید هویت گوینده ارائه دهد [۱۴] . در این روش از دو مدل یکی برای گوینده ای که می خواهیم هویت او را مورد تایید قرار دهیم و دیگری برای سایر گوینده ها استفاده می کنیم؛ سپس با مقایسه تفاضل خروجی این دو مدل با یک مقدار آستانه، هویت گوینده تایید یا تکذیب میشود.

---

[1] Vector quantization
[2] Codebook
[3] Hidden Markov Model
[4] Expectation Maximization

۱-۶-۳- مدل های جدا کننده

ماشین بردار پشتیبان در سال های اخیر برای بهبود کیفیت سیستم های شناسایی گوینده به کار گرفته شده است. تعدادی از محققین از ماشین بردار پشتیبان به همراه مدل مخلوط گوسی استفاده کرده اند. Rendls , wan در سال ۲۰۰۵ در مقاله ای از ماشین بردار پشتیبان در شناسایی گوینده استفاده کردند [۱۵]. در این پژوهش ماشین بردار پشتیبان بر مبنای کرنل های فضای امتیاز آموزش داده می شود. این نوع کرنل ها حالت عمومی کرنل های فیشر هستند. در این تحقیق جدا سازی به ازای تمام بردارهای بدست آمده از سیگنال صحبت انجام میشود و در مقابل روشهایی است که جداسازی را در سطح فریم انجام می دهند.

از دیگر تحقیقات صورت گرفته در زمینه مدل ماشین بردار پشتیبان میتوان به پژوهش انجام شده توسط Campbell و گروه همکار او اشاره کرد. روش پیشنهادی Campbell و گروه همکار او بر مبنای استفاده از بردارهای میانگین مدل مخلوط گوسی است که از ترکیب این بردارهای میانگین برداری به نام سوپروکتور[1] به وجود می آید [۱۶]. سپس این بردار بزرگ در ماشین بردار پشتیبان استفاده میشود. کرنل های مورد استفاده در این مقاله ایده نوینی است که مطرح شده و مورد استفاده قرار گرفته اند. این کرنل ها با استفاده از معیار فاصله در مدل های مخلوط گوسی به دست می آیند. دقت روش مورد بررسی در این مقاله قابل مقایسه و در حد روش هایی است که از مدل مخلوط گوسی بهره می برند.

Campbell در مقاله دیگری از چند جمله ای ها برای مدل کردن مرز جدا ساز بین بردارهای ویژگی استخراج شده از صدای اشخاص استفاده کرد [۱۷]. ایده این مقاله استفاده از چند جمله ای ها و معیار مینیمم میانگین خطا برای پیدا کردن مرزی است که توسط این چند جمله ای ایجاد می شود. مزیت این روش نسبت به روشهای کلاسیک کاهش پیچیدگی محاسباتی، پیاده سازی آسان و کاهش حافظه مورد احتیاج آن است. در ادامه این پژوهش Campbell مدل چند جمله ای را مبنای معرفی یک کرنل قرار داد و روش نوینی برای ایجاد یک سیستم شناسایی گوینده توسط ماشین بردار پشتیبان پیشنهاد کرد. این پژوهش مبنای معرفی یک کرنل ماشین بردار پشتیبان با نام کرنل GLDS [۱۸] گردید.

[1] - Supervector