فصلنامه پژوهش در علوم، مهندسی و فناوری
دوره ۴، شماره ۲، تابستان ۱۳۹۷
صفحات ۱۷-۲۹
ISSN: 2476-5082
www.uctjournals.com

ارائه و پیاده‌سازی یک سیستم تشخیص جنسیت گوینده مبتنی بر به‌کارگیری اطلاعات طیفی و نوایی گفتار با استفاده از آبَر بردارها

مریم نفری افشاری [۱*]، محمد ادبی تبار [۲]

۱. کارشناسی ارشد مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه آزاد اسلامی ، واحد ساری ، مازندران ، ایران.
۲. دانشکده مهندسی کامپیوتر، دانشگاه آزاد اسلامی ، واحد ساری ، مازندران ، ایران.
نویسنده مسئول : مریم نفری افشاری

چکیده

تشخیص جنسیت، بر اساس صدای گوینده، به مسئله‌ی تعیین اینکه جنسیت گوینده‌ی یک قطعه گفتاری مذکر و یا مونث است می‌پردازد. تشخیص جنسیت گوینده گامی مهم در سیستم‌های بازشناسی گوینده و گفتار است. در هر دوی این سیستم‌ها تشخیص جنسیت، می‌تواند مسئله را از یک مستقل از جنسیت به یک مسئله وابسته به جنسیت تبدیل کند و به این ترتیب، اندازه و پیچیدگی مسئله کاهش پیدا می‌کند. در این نوشتار به معرفی روشی بر مبنای ابربردارها برای تشخیص جنسیت گوینده می‌پردازیم. در ابتدا با استفاده از ویژگی‌های استخراج شده مدل‌های مخلوط گاوسی را آموزش می‌دهیم و سپس با استفاده از میانگین این مولفه‌های گاوسی اقدام به ایجاد ابربردارها می‌کنیم و در نهایت با استفاده از دسته‌بندی کننده‌ی ماشین بردار پشتیبان دسته‌بندی را انجام می‌دهیم. در این تحقیق تاثیر حجم داده‌ی آموزشی و تست، نوع ویژگی بکار رفته، میزان کاهش بعد ابربردارها، نوع کرنل SVM و مواردی از این قبیل مورد بررسی قرار گرفته است و در بهترین شرایط بر روی دادگان فارسدات میکروفنی، به راندمان 94٪ دست یافته ایم.

واژه‌های کلیدی: تشخیص جنسیت گوینده، مدل‌های مخلوط گاوسی، ابربردار، ماشین بردار پشتیبان

طرح مسئله و مرور سوابق

تشخیص جنسیت[1]، بر اساس صدای گوینده، به مسئله‌ی تعیین اینکه جنسیت گوینده‌ی یک قطعه گفتاری مذکر و یا مونث است می‌پردازد. تشخیص جنسیت گوینده گامی مهم در سیستم‌های بازشناسی گوینده و گفتار است [۱]و[۲] در هر دوی این سیستم‌ها تشخیص جنسیت، می‌تواند مسئله را از یک مستقل از جنسیت به یک مسئله وابسته به جنسیت تبدیل کند و به این ترتیب، اندازه و پیچیدگی مسئله کاهش پیدا می‌کند و کارایی آن به طرز قابل توجهی افزایش پیدا می‌کند [۳] و[۴] و از این رو تشخیص جنسیت می‌تواند نقش مهمی را در فرایند پردازش گفتار ایفا کند. البته چنین سیستمی به تنهایی نیز می‌تواند به خصوص ار لحاظ آماری مفید باشد، به عنوان مثال هنگامی‌که بخواهیم بدانیم در عرض یک ماه از تعداد تلفن‌هایی که به یک مرکز می‌شود چه تعداد آن مونث و چه تعداد آن مذکر بوده‌اند[۵].

برای تشخیص جنسیت گوینده یک گفتار می‌توان از پارامترهای مختلفی که از یک قطعه‌ی گفتاری استخراج می‌شود استفاده کرد. از مهمترین این روش‌ها و پارامترها که تا کنون در مقالات و تحقیقات مختلف برای تشخیص جنسیت گوینده از آنها استفاده شده است می‌توان به فرکانس پایه صدا یا گام[2] صدا، همچنین ضرایب کپستروم فرکانس مل، فرمنت صدا، ضرایب پیشگویی خطی، مدل پنهان مارکف و مدل مخلوط گاوسی و مواردی از این قبیل اشاره کرد.

در سال ۲۰۰۳ هرب و چن یک روش را بر اساس رویکرد طبقه‌بندی کننده‌ی صوتی کلی[3] ارائه کردند [۹]. طبقه‌بندی کننده‌ی صوتی سیگنال صوت را بر اساس اسپکتروم مرتبه‌ی اول در پنجره‌های یک ثانیه‌ای تقسیم‌بندی کرده و با استفاده از شبکه‌ها عصبی به طبقه‌بندی صوت می‌پردازد. سیستم ارائه شده توسط آنها این ویژگی را داشت که در محیط‌های مختلف کارایی خوبی از خود نشان می‌داد و این روش نسبت به نویز مقاوم بود. دقت گزارش شده حاصل از روش آنها در حدود ۹۲ درصد می‌باشد.

در سال ۲۰۰۲، ژانگ از روشی برای تعیین مقدار فرکانس گام که بر مبنای تابع میانگین تفاضل دامنه (AMDF) استفاده کرد که این روش را بر روی پیکره‌ی گفتاری DARPA TIMIT اجرا کرد و نتایج قابل قبولی به دست آورد[۱۰]. کوک در سال ۲۰۰۲ روشی را بر اساس طبقه‌بندی کننده‌ی صوتی‌ای که بر مبنای ویژگی‌های استخراج شده‌ی ضرایب کپسترال فرکانس مل یک طبقه‌بندی کننده با استفاده از GMM طراحی کرده است[۱۱]. البته آنها در این مقاله داده‌ها را به جای دو دسته به سه دسته (مرد، زن و اطلاعیه‌های ورزشی[4]) تقسیم بندی کردند که دقت به دست آمده حدود ۷۴٪ است.

در سال ۲۰۰۵ اسلومکا با استفاده از ترکیب رویکرد استفاده از گام و همچنین استفاده از یک طبقه‌بندی کننده‌ی صوتی که بر مبنای GMM کار می‌کند روشی را برای تشخیص جنسیت گوینده ارائه داد [۱۲]. این روش ترکیبی بر روی جملات ۷ ثانیه‌ای که از داده‌های تلفنی پیکره‌ی OGI با حذف سکوت استخراج شده بودند دقتی حدود ۹۴ درصد را ارائه می‌داد.

همانطور که مطرح شد، مدل پنهان مارکف نیز برای عمل تشخیص جنسیت گوینده مورد استفاده قرار گرفته است. در این روش برای هر جنسیت یک موتور بازشناسی گفتار HMM آموزش داده می‌شود و از این مدل‌های وابسته به جنسیت برای تشخیص جنسیت گوینده‌ی گفتار استفاده می‌شود. مدلی که بیشترین مقدار درست نمایی[5] را داشته باشد به عنوان جنسیت گوینده انتخاب می‌شود.

در سال ۲۰۰۶ تینگ و ینگچون روشی ترکیبی بر اساس استفاده از فرکانس گام و همچنین ضرایب کپستروم فرکانس مل ارائه دادند[۱۵]. در روش ارائه شده توسط آنها فاز آنالیز آکوستیکی با ایجاد دو مدل مخلوط گاوسی برای هر یک از جنسیت‌ها ساخته می‌شود و همچنین علاوه بر آن مقدار فرکانس گام نیز محاسبه می‌شود و با استفاده از تعیین یک سطح آستانه جنسیت گوینده تخمین زده می‌شود. سپس اطلاعات به دست آمده توسط دو قسمت یعنی آنالیز آکوستیکی توسط ضرایب MFCC و

1 Gender Identification
2 Pitch
3 General Audio Classifier
4 Sports Announcement
5 Likelihood

همچنین تخمین با استفاده از مقدار فرکانس گام صدا، توسط یک روش ترکیبی نرمال‌سازی خطی با یکدیگر ترکیب شده و برای تصمیم‌گیری نهایی استفاده می‌شود. دقت به دست آمده‌ی این روش بر روی پیکره‌ی SRMC حدود ۹۳ درصد گزارش شده است.

در سال ۲۰۱۲ یاکون و داپنگ روشی دو مرحله‌ای برای تشخیص جنسیت گوینده ارائه کردند [۱۶]. در روش ارائه شده توسط این دو، در مرحله‌ی اول تصمیم‌گیری بر اساس مقدار فرکانس گام صورت می‌پذیرد. یک سری از نمونه‌ها به سادگی بر اساس مقادیر فرکانس گام قابل تشخیص هستند و مقدار فرکانس گام آنها به گونه است که می‌توان تشخیص داد که آیا گوینده زن است و یا مرد. پیچیدگی این مرحله بسیار پایین است چرا که فقط کافی است مقدار گام را با یک حد آستانه مقایسه کرد.

این بررسی اجمالی روش‌های موجود نشان می‌دهد که دقت ارائه شده در این مقالات عموماً بر روی جملاتی با طول بین ۵ تا ۷ ثانیه که به صورت دستی جدا شده اند به دست آمده است. در چندین مورد از این مطالعات پیش پردازش‌های گفتاری متعددی نیز بر روی داده‌های آموزشی و آزمایشی صورت پذیرفته است. از این جمله می‌توان به حذف سکوت در برخی از روش‌ها اشاره کرد. از آنجا که تشخیص سکوت در گفتار تمیز به خوبی قابل انجام است این روش‌ها در مورد گفتاری که تمیز نباشد (نویز داشته باشد) به مشکل بر خواهد خورد.

در برخی از این روش‌ها نیز بازشناسی واج صورت می‌پذیرد که پس از بازشناسی واج و در سطح واج برای جنسیت گوینده تصمیم گیری می‌شود. مشخصا عمل بازشناسی واج، خود به پیچیدگی سیستم تشخیص جنسیت گوینده می‌افزاید چرا که بایستی یک لایه برای بازشناسی واج به سیستم اضافه شود.

تعریف پروژه و مراحل آن

هدف این پروژه، ارائه و پیاده‌سازی یک سیستم تشخیص جنسیت گوینده مبتنی بر به‌کارگیری اطلاعات طیفی و نوایی گفتار با استفاده از آبَر بردارها است. ایده‌ی ابر بردارها از استفاده روزافزون و موفق آن‌ها در پردازش سیگنال‌های صوتی و گفتاری به خصوص مبحث بازشناسی گوینده الهام گرفته شده است. ابر بردارها به عنوان یک روش مقاوم و پیشرو در نمایش نمونه‌های سیگنال در فضای با ابعاد بالا و به صورت یک بردار واحد مطرح هستند، که این مسئله هم ناشی از ویژگی‌های به‌کاررفته و متناسب با ابر بردارها و هم ناشی از ماهیت ابر بردار و نحوه نمایش نمونه‌های سیگنالی است [۱۷]، به طوری که استفاده از ابر بردارها به عنوان ورودی ماشین بردار پشتیبان منجر به توسعه کرنل‌های مختلف و افزایش کاربردهای آن شد [۱۸] [۱۹] [۲۰]. بعضی اوقات «ابر بردار» به ترکیب تعداد زیادی بردارهای با ابعاد کوچک با هم و حاصل شدن یک بردار با ابعاد بالاتر اطلاق می‌شود؛ همانند پشته‌سازی بردارهای میانگین d بعدی از یک مدل مخلوط گوسی با K مؤلفه و تبدیل به یک ابر بردار گوسی Kd بعدی [۱۸].

مراحل انجام این پروژه را می‌توان اینگونه بیان کرد که برای هر جنسیت (مرد و زن) از گفتار مربوط به آن جنسیت، پس از فریم‌بندی اقدام به استخراج ویژگی می‌کنیم و سپس با استفاده از این ویژگی‌ها اقدام به ساخت یک مدل مخلوط گاوسی می‌شود که در نهایت تعدادی مولفه گاوسی خواهیم داشت که هر یک شامل بردار میانگین، کواریانس و وزن مربوطه خواهد بود. مراحل انجام پروژه را در دیاگرام شکل (۱) مشاهده می‌کنید. همانطور که مشاهده می‌شود ابتدا با در اختیار داشتن داده‌های گفتاری آموزشی نسبت به استخراج ویژگی‌ها اقدام می‌کنیم، ویژگی‌های مورد استفاده – که پس از انجام آزمایشات مشخص شد که این ترکیب ویژگی‌ها بهترین نتیجه را ارائه می‌دهد- را در جدول (۱) مشاهده می‌کنید.

جدول ۰ ویژگی‌های استفاده شده

نوع ویژگی	ضرایب طیفی	مشتق اول ضرایب طیفی	مشتق دوم ضرایب طیفی	ضرایب MFCC	مشتق اول ضرایب MFCC	مشتق دوم ضرایب MFCC	فرکانس گام
تعداد	**۱۲**	**۱۲**	**۱۲**	**۱۲**	**۱۲**	**۱۲**	**۱**

پس از استخراج ویژگی‌های داده‌های آموزشی، با استفاده از بخشی از داده‌های آموزشی اقدام به ایجاد یک مدل جهانی مدل مخلوط گاوسی[۶] می‌کنیم. در مرحله‌ی بعد با استفاده از دیگر داده‌های آموزشی و همچنین مدل جهانی ایجاد شده در مرحله‌ی قبل، این مدل جهانی را برای هر یک از داده‌های آموزشی با استفاده از ویژگی‌های آنها تطبیق می‌دهیم سپس با ترکیب بردارهای میانگین مدل‌های ایجاد شده اقدام به ساخت ابر بردارها می‌کنیم و در نهایت با استفاده از روش دسته‌بندی ماشین بردار پشتیبان عمل مدل‌سازی و دسته‌بندی را انجام می‌دهیم.

ابربردار، مدل مخلوط گاوسی و ماشین بردار پشتیبان

ماشین بردار پشتیبان

ماشین بردار پشتیبانی[۷] یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. الگوریتم SVM اولیه در ۱۹۶۳ توسط وپنیک ابداع شد و در سال ۱۹۹۵ توسط وپنیک و کورتس برای حالت غیرخطی تعمیم داده شد [۲۲]. مبنای کاری دسته‌بندی کننده‌ی SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. حل معادله‌ی پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند صورت می‌گیرد. قبل از تقسیمِ خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها را به وسیله‌ی تابعِ کرنل به فضای با ابعاد خیلی بالاتر می‌بریم. برای اینکه بتوانیم مساله ابعاد خیلی بالا را با استفاده از این روش‌ها حل کنیم از قضیه دوگانی لاگرانژ برای تبدیلِ مساله‌ی مینیمم‌سازی مورد نظر به فرم دوگانی آن که در آن به جای تابع پیچیده‌ی کرنل که ما را به فضایی با ابعاد بالا می‌برد، تابعِ ساده‌تری به نامِ تابع هسته که ضرب برداری تابع کرنل است ظاهر می‌شود استفاده می‌کنیم. در ماشین بردار پشتیبان آموزش نسبتاً ساده است و برخلاف شبکه‌های عصبی در بیشینه‌های محلی گیر نمی‌افتد. برای داده‌های با ابعاد بالا تقریباً خوب جواب می‌دهد.

مدل مخلوط گاوسی

امروزه یکی از رایج‌ترین روش‌های مدل کردن گویندگان، مدل مخلوط گاوسی[۸] است که در آن توزیع بردارهای ویژگی یک گوینده، مدل می‌شوند. پارامترهای مدل در GMM عبارتند از بردارهای میانگین، ماتریسهای کواریانس و وزنهای مخلوط‌ها که در طی یک فرایند آموزشی تکراری مبتنی بر بیشینه‌سازی انتظار[۹] تخمین زده می‌شوند.

مدلهای مخلوط گاوسی یکی از مهم‌ترین و پرکاربردترین روش‌های پردازش گفتار به ویژه موارد مربوط به بازشناسی گوینده می‌باشند [۲۳]. در سیستم‌های مبتنی بر مدل مخلوط گاوسی توزیع احتمال بردارهای ویژگی یک سیگنال گفتار به صورت یک ترکیب خطی از K مخلوط گاوسی به صورت رابطه‌ی زیر بیان می‌شود:

$$P(X_t|\lambda) = \sum_{k=1}^{K} C_k N(X_t, \mu_k, \Sigma_k) \qquad (۱)$$

در رابطه‌ی بالا λ یک علامت اختصاری برای پارامترهای مدل مخلوط گاوسی است.

$$\lambda = \{C, \mu, \Sigma\} \qquad 1 \leq k \leq K \qquad (۲)$$

همچنین $N(X_t, \mu_k, \Sigma_k)$ یک تابع چگالی احتمال با بردار میانگین μ_k و ماتریس کواریانس Σ_k می‌باشد.

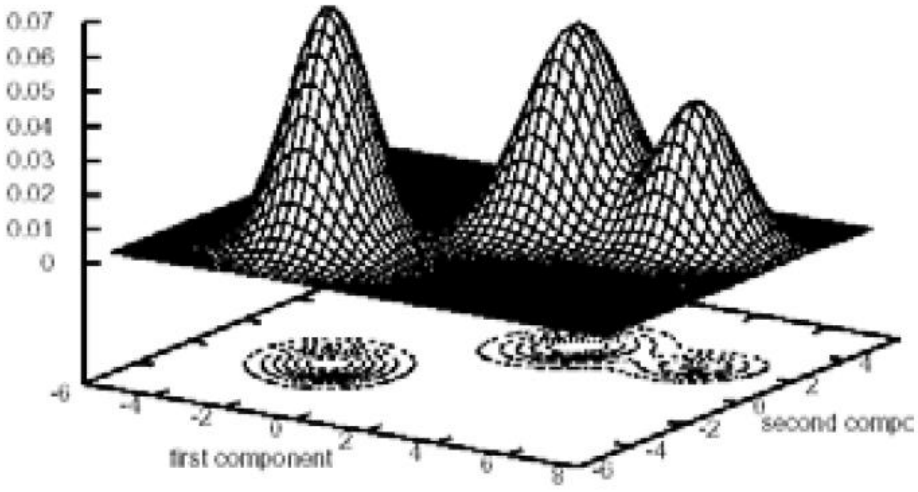
$$N(X_t, \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \times \exp[-\frac{1}{2}(x_t - \mu_k)' \Sigma_k^{-1} (x_t - \mu_k)] \qquad (۳)$$

6 Gaussian Mixture Model-Universal Background Model (GMM-UBM)
7 Support vector machines - SVMs
8 Gaussian Mixture Model
9 Expectation Maximization

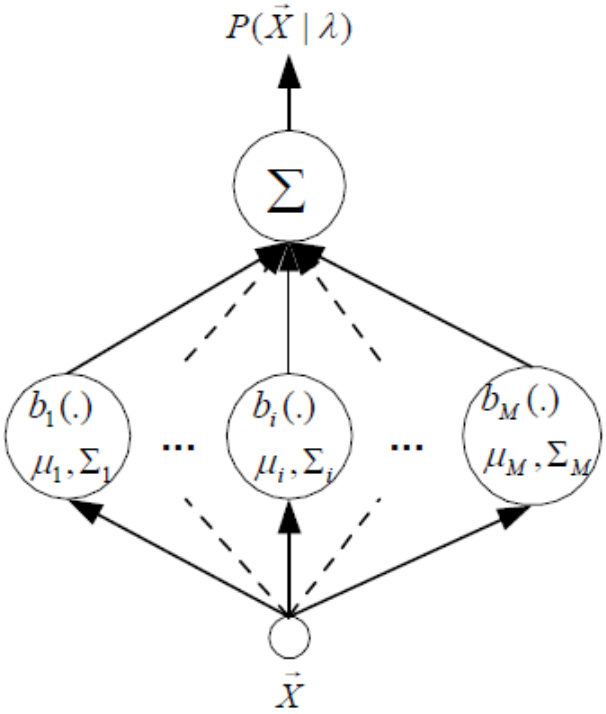
که در آن منظور از علامت (')، ترانهادهٔ یک ماتریس و D تعداد بعد بردار x_t است. معمولاً از ماتریس کواریانس قطری استفاده می‌شود. دلیل این امر را نیز در ویژگی‌های کپسترال دانسته‌اند. زیرا ویژگی‌های کپسترال تقریباً غیر همبسته هستند [۲۴]. انتخاب تعداد مخلوط‌ها وابسته به تعداد داده‌های آموزشی است. البته بایستی به اندازه‌ای باشند که بتوانند تغییرات صوتی گویندگان را مدل کند. به عبارت دیگر تعداد مخلوط‌های گاوسی بایستی به اندازه‌ای باشد که با تعداد موجود از داده‌های آموزشی بتوان پارامترهای آن را تخمین زد. از طرف دیگر وزن‌های مخلوط‌ها بایستی محدودیت زیر را ارضا کند.

$$\sum_{k=1}^{K} c_k = 1 \quad , \qquad c_k \geq 0 \tag{۴}$$

این مدل کردن را با ترکیب خطی تعدادی تابع گاوسی انجام می‌شود به همین دلیل به آن مدل مخلوط گاوسی گفته می‌شود . شکل (۲) این فرآیند را برای یک حالت ساده دو بعدی و با سه عنصر مخلوط نشان می‌دهد. یک مخلوط گاوسی از مجموع وزن دار M توزیع گاوسی مطابق شکل (۳) به دست می‌آید. مدل GMM می‌تواند فرم‌های مختلف داشته باشد . یک مدل می‌تواند به ازای هر عنصر مخلوط یک ماتریس کوواریانس داشته باشد که به آن کوواریانس نقطه‌ای گفته می‌شود. مدل دیگر این است که یک ماتریس کوواریانس برای کل گوینده‌ها وجود داشته باشد که به آن کوواریانس کلی گفته می‌شود. همچنین ماتریس کوواریانس می‌تواند کامل یا قطری باشد. ماتریس کوواریانس کامل، وابستگی اجزاء عنصر مخلوط را به همدیگر نشان می‌دهد ولی پارامترهای بسیار بیشتری برای آموزش دارد و به همین دلیل در تعداد زیادتر عناصر مخلوط، آموزش در بسیاری موارد دچار کمبود داده می‌گردد. در مقابل با در نظر گرفتن ماتریس کوواریانس به صورت قطری (که در آن صورت تبدیل به بردار واریانس می‌شود) و با در نظر گرفتن تعداد مخلوط بیشتر می‌توان پیچیدگی محاسبات و آموزش را پایین آورد و به نتایج مطلوب‌تری دست یافت.



شکل ۲. یک توزیع مخلوط گاوسی [۲۵]



شکل ۳- یک مدل مخلوط گاوسی [۲۵]

روش پیشنهادی

استخراج ویژگی ها

قدم اول در پیاده سازی پروژه‌ی انتخاب و استخراج ویژگی‌های مناسب برای عمل دسته‌بندی است. با بررسی‌های انجام شده، ویژگی‌های مناسب برای عمل دسته‌بندی جنسیت گوینده، ویژگی‌های MFCC[۱۰]، طیفی[۱۱] و فرکانس گام[۱۲] تشخیص داده شد.

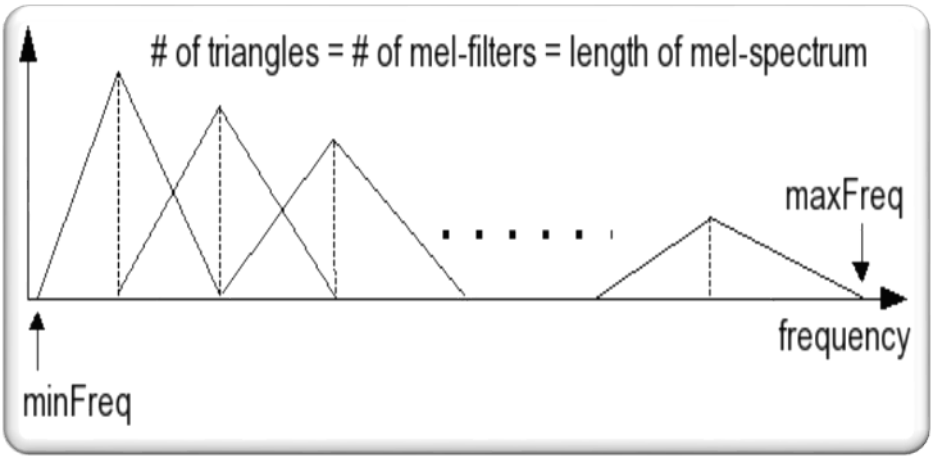
ویژگی MFCC

ضرایب MFCC الهام گرفته از خواص شنیداری گوش انسان در دریافت و فهم گفتار می‌باشد[۳۸]. جهت محاسبه ضرایب MFCCهر قاب، سیگنال مربوط به هر زبان را ابتدا در پنجره همینگ با معادله‌ی (۱) ضرب کرده و سپس از سیگنال بدست آمده تبدیل فوریه گسسته[۱۳] گرفته می‌شود. اندازه تبدیل فوریه گرفته شده، محاسبه شده و بر روی پوش طیف بدست آمده مراحل زیر برای استخراج ضرایب MFCC سیگنال انجام می‌شود.

$$w(k+1) = 0.54 - 0.46\cos\left(\frac{2k\pi}{n-1}\right) \qquad k = 0, 1, 2, \ldots, n-1 \qquad (۱)$$

طیف سیگنال از تعدادی فیلتر با عرض باند مقیاس مل عبور داده می‌شود. در شکل (۴) این فیلترها را مشاهده می‌کنید.

10 Mel-Frequency Cepstrum Coefficients
11 Spectral Coefficients
12 Pitch
13 Fast Fourier Transform



شکل ۴ فیلتر بانک مل [۳۸]

این فیلترها تفکیک فرکانسی سیستم ادراک گوش انسان را شبیه سازی می‌کنند. در فرکانس‌های بالا، پهنای باند فیلترها زیادتر است و این امر بدان معناست که حساسیت گوش انسان نسبت به تغییر فرکانس در فرکانس‌های بالا، کمتر از حساسیت آن در فرکانس‌های پایین است.

فیلترهای اولیه با تغییر فرکانس مرکزی به صورت خطی (فاصله فرکانسی بین مراکز آنها ۱۳۳/۳۳ هرتز است) و فیلترهای بعدی با تغییر فرکانس مرکزی به صورت لگاریتمی و فرکانس مرکزی هر فیلتر ۱/۰۸ برابر فرکانس مرکزی فیلتر قبلی است. فیلترها به شکل مثلثی بوده و شروع هر فیلتر از فرکانس مرکزی فیلتر قبلی و خاتمه آن در فرکانس مرکزی فیلتر بعدی است و ماکزیمم آن در فرکانس مرکزی خودش می‌باشد. سپس از خروجی فیلترها لگاریتم گرفته میشود. به منظور کاهش تعداد مولفه‌های بردار ویژگی، از مقادیر لگاریتم خروجی‌های فیلترهای فوق تبدیل کسینوسی گسسته[14] گرفته می‌شود. نتیجه به‌دست آمده برابر با ضرایب مورد نظر می‌باشند. رابطه‌ی (۲) تبدیل کسینوسی گسسته را روی خروجی‌های فیلترها نشان می‌دهد.

$$c(i) = \sum_{j=1}^{j=F} \log(X_j) \cos\left(\frac{\pi i\ (j + 0.55)}{F}\right) \quad 1 \le i \le F \qquad (۲)$$

که در آن F تعداد فیلترها، X_j خروجی حاصل از فیلتر j ام و $c(i)$ ضرایب MFCC حاصل می‌باشند. بردارهای ویژگی برای هرکلاس به تعداد قاب‌های آن کلاس می‌باشد.

در این پروژه هر فایل گفتاری به تعدادی پنجره همینگ با طول ۲۵ میلی ثانیه تقسیم شد که این پنجره‌ها به میزان ۱۰ میلی ثانیه با یکدیگر همپوشانی داشتند. تعداد ۱۲ ضریب MFCC برای هر فریم استخراج شد که با اعمال تعداد ۱۲ فیلتر بانک به دست آمد. البته پس از به دست آمدن این ضرایب MFCC از آنها دو بار مشتق گرفته شد تا مشتقات اول و دوم آنها نیز مورد محاسبه قرار گیرد.

ویژگی طیفی

در این پروژه برای استخراج ضرایب طیفی پس از انجام بررسی‌های لازم تصمیم گرفته شد که برای ویژگی‌های طیفی در روند استخراج ویژگی‌های MFCC از خروجی حاصل از اعمال فیلتر بانک مل استفاده شود. برای این منظور پس از فریم بندی داده‌های گفتاری (که بیان شد طول هر فریم ۲۵ میلی ثانیه در نظر گرفته شد و این فریم‌ها به میزان ۱۰ میلی ثانیه هم

[14] Discrete Cosine Transform

پوشانی دارند) از آنها تبدیل فوریه گسسته و سپس بزرگی[15] خروجی حاصله را محاسبه می‌کنیم. آنگاه بر روی خروجی به دست آمده فیلتر بانک مل را اعمال می‌کنیم و از خروجی حاصل به عنوان ویژگی‌های طیفی استفاده می‌شود.

در این پروژه تعداد ۱۲ ویژگی طیفی برای هر فریم گفتاری استخراج شد و سپس برای این مجموعه مقادیر مشتقات اول و دوم نیز محاسبه شد و در پیاده‌سازی نهایی از آن بهره برده شد.

ویژگی گام

یکی از ویژگی‌هایی که استفاده‌ی بسیار زیادی در تشخیص جنسیت گوینده داشته است پریود گام است. مهمترین دلیل استفاده از پریود گام این واقعیت است که میانگین فرکانس بنیادی صدای مردان عموما بین ۵۰ تا ۲۰۰ هرتز است در حالی که برای زنان این میانگین بین ۱۵۰ تا ۳۰۰ هرتز است. [۳۹]. اگر چه چالش‌های دیگری نیز در پیش روی تشخیص جنسیت از روی گام وجود دارد. اول اینکه تخمین صحیح و کارای پریود گام فقط از روی گفتار تمیز و بدون نویز قابل حصول است در صورتی که در حالت طبیعی همیشه گفتار تمیز نیست[۴۰]. از طرفی همپوشانی زیادی بین محدوده پریود گام صدای مردان و زنان وجود دارد که این واقعیت عمل تشخیص جنسیت از روی گام را دچار مشکل می‌کند. الگوریتم‌های ارائه شده برای این روش را می‌توان به دو بخش تقسیم کرد. برخی از این الگوریتم‌ها در حوزه‌ی زمان کار می‌کنند و محاسبات شان را برای تشخیص جنسیت گوینده در حوزه‌ی زمان بررسی می‌کنند هر چند اکثر الگوریتم‌های این روش به دسته‌ی دوم که در حوزه‌ی فرکانس محاسبات خود و تصمیم گیری‌های خود را انجام می‌دهند تعلق دارند. در پیاده‌سازی نهایی برای هر فریم یک ویژگی گام نیز استخراج شد. برای استخراج گام از الگوریتمی مبتنی بر روش نسبت زیرهارمونیک به هارمونیک[16] استفاده شد[۴۱]. این الگوریتم از مقیاس فرکانس لگاریتمی و یک تکنیک شیف طیفی برای به دست آوردن مجموع بزرگی هارمونیک‌ها و زیرهارمونیک‌ها استفاده می‌کند.

۲. ساخت مدل پس زمینه‌ای سراسری

پس از استخراج ویژگی‌ها از آنها برای مدل سازی استفاده می‌شود. اولین قدم ساخت یک مدل پس زمینه‌ای سراسری[17] برای مدل مخلوط گاوسی است. در فصل دوم راجع به ساخت چنین مدلی صحبت شد. برای ساخت این مدل ابتدا با استفاده از تعدادی از داده‌های گفتاری آموزشی که از آنها ویژگی‌های بیان شده در مرحله قبل استخراج شده است اقدام به ایجاد یک مدل سراسری می‌کنیم. در این پروژه برای مدل سازی مدل مخلوط گاوسی از ابزار NETLAB استفاده شد. نرم افزار شبیه سازی NETLAB برای پیاده سازی تکنیک‌های آنالیز داده‌ها استفاده می‌شود. این ابزار بر اساس موارد مطرح شده در کتاب بازشناسی الگو Bishop پیاده‌سازی شده است.

تعداد عناصر گاوسی برای ایجاد مدل‌های مخلوط گاوسی ۵۱۲ عدد در نظر گرفته شد. ابتدا برای ساخت مدل پس زمینه‌ای سراسری ابتدا یک معماری کلی برای مدل مخلوط گاوسی ایجاد می‌کنیم. سپس با استفاده از بخشی از داده‌ها (تعدادی داده‌ی گفتاری از مجموعه داده‌های پیکره‌ی فارسدات) برای مقدار دهی اولیه به این مدل استفاده می‌کنیم. سپس از الگوریتم بیشینه سازی انتظار[18] برای آموزش مدل مخلوط گاوسی استفاده می‌کنیم.

برای هر یک از جنسیت‌ها (مرد و زن) یک مدل پس زمینه‌ی سراسری جداگانه ایجاد شد. در مراحل بعد از این مدل سراسری برای ایجاد ابربردارها استفاده می‌شود.

۳. تطبیق مدل با داده‌های آموزشی و استخراج ابربردارها

پس از ایجاد مدل‌های پس زمینه‌ی سراسری برای هر یک از جنسیت‌ها، تعدادی داده‌ی آموزشی در نظر گرفته شد و پس از استخراج ویژگی آنها، از آنها برای تطبیق مدل‌های سراسری ایجاد شده برای هر یک از گویندگان استفاده گردید. به این صورت که برای هر گوینده یک مدل جدید ایجاد شد که حاصل تطبیق مدل سراسری با ویژگی‌های هر یک از داده‌های آموزشی

15 Magnitude
16 Subharmonic-to-Harmonic Ratio (SHR)
17 Universal Background Model (UBM)
18 Expectation–maximization algorithm

گوینده با جنسیت مورد نظر است. در این فرایند تطبیق نرخ تطبیق ۶۰ درصد در نظر گرفته شد. پس از پایان این مرحله به ازای هر یک از داده‌های آموزشی یک مدل مخلوط گاوسی داریم.

آنگاه با استفاده از مدل‌های مخلوط گاوسی به دست آمده در مرحله‌ی قبل برای استخراج ابربردارها اقدام می‌کنیم. در تحقیق بیان شد که ویژگی‌های هر یک از عناصر مدل مخلوط گاوسی عبارتند از بردارهای میانگین، ماتریسهای کواریانس و وزنهای مخلوط‌ها که در طی یک فرایند آموزشی تکراری مبتنی بر بیشینه‌سازی انتظار تخمین زده می‌شوند. در این مرحله برای ساخت ابر بردارها از پارامتر میانگین مدل‌ها استفاده می‌شود. با توجه به اینکه بعد ویژگی‌های استخراج شده برای هر فریم گفتاری ۷۳ می‌باشد (۳۶ ضریب طیفی و مشتقات، ۳۶ ضریب MFCC و مشتقات و یک ویژگی گام برای هر فریم) و از طرفی تعداد مولفه‌های گوسی هر یک از مدل‌ها ۵۱۲ در نظر گرفته شد بعد ابربردار برابر با ۳۷۳۷۶ در تعداد فریم هر یک از فایل‌های آموزشی خواهد بود.

۴. مدل سازی با استفاده از ماشین بردار پشتیبان

پس از استخراج ابر بردارها تمام ابربردارهای گویندگان مرد را در کنار هم در یک گروه قرار می‌دهیم و برای گویندگان زن نیز چنین کاری را انجام می‌دهیم و به ابربردارهای مردها برچسب ۱ و به ابربردارهای زن‌ها برچسب صفر را اختصاص می‌دهیم. در این روش، دسته‌بندی کننده ناحیه مرزی متعلق به دو کلاس مختلف را با استفاده از نگاشت نمونه‌های ورودی به فضایی با بعد بالاتر و سپس جستجو برای یافتن ابر صفحه‌ی جدا کننده‌ی دو کلاس در این فضا، یاد می‌گیرد. به طور معمول دو دید برای این هدف وجود دارد. یکی از آنها استراتژی «یک در مقابل همه» برای دسته‌بندی هر جفت کلاس و کلاس‌های باقی مانده است. دیگر استراتژی «یک در مقابل یک» برای دسته‌بندی هر جفت است. در شرایطی که دسته‌بندی اول به دسته‌بندی مبهم منجر می‌شود. برای مسائل چند کلاسی، رهیافت کلی کاهش مسئله‌ی چند کلاسی به چندین مسئله دودویی است. هریک از مسائل با یک جداکننده دودویی حل می‌شود. سپس خروجی جداکننده‌های دودویی ماشین بردار پشتیبان با هم ترکیب شده و به این ترتیب مسئله چند کلاس حل می‌شود. در این پروژه با توجه به اینکه دو کلاس بیشتر نداریم از استراتژی «یک در مقابل یک» استفاده شد.

برای مدل‌سازی و دسته‌بندی با استفاده از SVM در این پروژه از ابزار LibSVM استفاده شد. این ابزار، یک کتابخانه‌ی متن‌باز برای یادگیری ماشین است که توسط دانشگاه ملی تایوان نوشته شده است[۴۲].

برای ایجاد ماشین بردار پشتیبان می‌توان پارامترهای مختلفی را تنظیم کرد و از آنها استفاده کرد. اولین بخش مربوط به نوع تابع استفاده شده به عنوان کرنل در ماشین بردار پشتیبان است. انواع کرنل موجود به شرح زیر است:

- کرنل خطی (linear) به صورت

$$u' * v \qquad (۱)$$

- کرنل چند جمله‌ای (polynomial) به صورت

$$(\text{gamma} * u' * v + \text{coef0})\text{^degree} \qquad (۲)$$

- کرنل بر اساس توابع پایه شعاعی(radial basis function) به صورت

$$\exp(-\text{gamma} * |u - v|\text{^}2) \qquad (۳)$$

- کرنل سیگموئیدی (sigmoid) به صورت

$$\text{gamma} * u' * v + \text{coef0} \qquad (۴)$$

که پس از تست‌ها و آزمایشات انجام شده دو کرنل چند جمله‌ای و سیگموئیدی بهترین عملکرد را از خود نشان دادند که به عنوان دو گزینه‌ی قابل انتخاب در پروژه گنجانده شدند. همچنین یک گزینه‌ی دیگر برای ایجاد ماشین پشتیبان، نوع ماشین ایجاد شده است. اینکه این ماشین چه ماهیتی داشته باشد را با استفاده از این گزینه می‌توان تنظیم کرد. انواع حالتی که ماشین بردار پشتیبان می‌پذیرد عبارت است از:

- C-SVC (ماشین بردار پشتیبان معمولی با الگوریتم استاندارد)
- nu-SVC (تنظیم خودکار کلاسه‌بندی بردارهای پشتیبان)
- one-class SVM (انتخاب یک ابر کره برای حداکثر سازی چگالی)
- epsilon-SVR (رگرسیون مقاوم بردارهای پشتیبان بر اساس خطایی کوچک)
- nu-SVR (تنظیم خودکار کلاسه بندی بردارهای پشتیبان با مینیموم کردن اپسیلون)

که پس از آزمایشات انجام گرفته مشخص شد که دو حالت C-SVC و one-class SVM بیشترین کارایی را از خود نشان می‌دهند که در پیاده‌سازی نهایی پروژه از حالت C-SVC استفاده شد.

نتیجه گیری

در این تحقیق راجع به ابربردارها و اینکه چگونه از آنها در پیاده سازی پروژه استفاده می‌شود صحبت شد.

سپس راجع به مفاهیم، تکنیک‌ها و روش‌هایی که در انجام پروژه از آنها بهره برده شد –همچون ابربردارها، ماشین بردار پشتیبان و مدل مخلوط گاوسی- صحبت کردیم. بیان شد که ماشین بردار پشتیبانی یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. همچنین راجع به مدل‌های مخلوط گاوسی نیز در این فصل صحبت کردیم. بیان شد که امروزه یکی از رایج‌ترین روش‌های مدل کردن گویندگان، مدل مخلوط گاوسی است که در آن توزیع بردارهای ویژگی یک گوینده، مدل می‌شوند. پارامترهای مدل در GMM عبارتند از بردارهای میانگین، ماتریس‌های کواریانس و وزن‌های مخلوط‌ها که در طی یک فرایند آموزشی تکراری مبتنی بر بیشینه سازی انتظار تخمین زده می‌شوند.

همچنین راجع به ابربردارها نیز صحبت شد. بیان شد که ابربردار به ترکیب تعداد زیادی بردارهای با ابعاد کوچک با هم و حاصل شدن یک بردار با ابعاد بالاتر اطلاق می‌شود؛ برای مثال با پشته‌سازی بردارهای میانگین d بعدی از یک مدل مخلوط گاوسی با K مؤلفه و تبدیل به یک ابربردار گاوسی Kd بعدی به دست آورد. در بالا ترجمه شده

بعد از ان به بررسی روال و مراحل روش پیشنهادی برای دسته‌بندی جنسیت گوینده پرداختیم. بیان شد که قدم اول در پیاده سازی پروژه‌ی انتخاب و استخراج ویژگی‌های مناسب برای عمل دسته‌بندی است. با بررسی‌های انجام شده، ویژگی‌های مناسب برای عمل دسته‌بندی جنسیت گوینده، ویژگی‌های MFCC ، طیفی و فرکانس گام تشخیص داده شد. گفته شد که پس از استخراج ویژگی‌ها اولین قدم ساخت یک مدل پس زمینه‌ای سراسری برای مدل مخلوط گاوسی است. در این فصل همچنین راجع به تطبیق ویژگی‌های گویندگان صحبت شد و نحوه‌ی عمل دسته بندی با ماشین بردار پشتیبان توضیح داده شد.

نتایج به دست آمده در تحقیق که مربوط به آزمایش‌ها و نتایج بود نشان داد که بهتر است در پیاده‌سازی نهایی پروژه از ضرایب MFCC و مشتقات آن، ضرایب طیفی و مشتقات آن به همراه فرکانس گام صدا –بدون مشتقات آن- استفاده شود. همچنین طول گفتار مورد استفاده نیز مورد بررسی قرار گرفت و مشخص شد که داده هایی با طول ۸ ثانیه برای این منظور کافی است. در بخش دیگری از آزمایش‌ها طول فایل گفتاری نیز مورد بررسی قرار گرفت که نتایج نشان داد طول فایلی در حدود ۸ ثانیه می‌تواند برای انجام عمل دسته‌بندی مناسب باشد و نیازی به فایلی با طول بیشتر ن می‌باشد. در نهایت کرنل مورد استفاده برای ماشین بردار پشتیبان نیز مورد بررسی قرار گرفت که نتایج آزمایش ما را به این نتیجه رساند که بهتر است در پیاده سازی نهایی از کرنل چند جمله‌ای استفاده کنیم.

هر چند دسته‌بندی کننده‌ی ماشین بردار پشتیبان نتایج قابل قبولی را از خود نشان داد اما می‌توان دیگر دسته‌بندی کننده‌ها را نیز مورد بررسی قرار داد و یا کرنلی را برای ماشین بردار پشتیبان ارائه داد که برای داده هایی با طول کمتر از ۸ ثانیه بتوانند جنسیت گوینده را به درستی تشخیص دهند.

همچنین مشاهده شد که استفاده از روش کاهش بعد PCA به طرز قابل توجهی دقت دسته بندی را کاهش می‌دهد؛ می‌توان از روش‌های دیگر کاهش بعد برای عمل کاهش بعد استفاده کرد و نتایج آن را بررسی کرد.

منابع:

[1] H. Hanson, E. Chuang, "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," The Journalof the Acoustical Society of America, vol. 106, p. 1064, 1999.

[2] A. Kondoz, "Digital speech: coding for low bit rate communication systems" John Wiley and Sons Ltd, 2004.

[3] A. Acero, X. Huang, "Speaker and gender normalization for continuous-density hidden Markov models", IEEE international conference on acoustics speech and signal processing, vol. 1. Citeseer, 1996. pp. 2-4.

[4] C. Neti, S. Roukos, "Phone-context specific gender-dependent acoustic-models forcontinuous speech recognition", Automatic Speech Recognition and Understanding, 1997. Proceedings, 1997 IEEE Workshop on. pp. 192-198

[5] Dongdong Li, Yingchun Yang, Zhaohui Wu. "Add prior knowledge to speaker recognition", Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2005, part of the SPIE Defense and Security Symposium 2005. Vol. 5813.

[6] Y. Konig, N. Morgan, "GDNN a gender dependent neural network for continuous speech recognition", International Joint Conference on Neural Networks, 1992. IJCNN, Volume: 2, 7-11, pp. 332 -337 vol.2.

[7] V. Rivarol, A.Farhat, D. O'Shaughnessy, "Robust gender-dependent acoustic-phonetic modelling in continuous speech recognition based on a new automatic male female classification", Proceedings, Fourth International Conference on Spoken Language, 1996. ICSLP 96, Volume: 2, 3-6 Oct. 1996, pp. 1081 -1084 vol.2, 1996

[8] C. Neti, S. Roukos, "Phone-context specific gender-dependent acoustic-models for continuous speech recognition", Proceedings, 1998 IEEE Workshop on Automatic Speech Recognition and Understanding.

[9] H. Harb, L. Chen, "Gender Identification Using a General Audio Classifier", In Proc. IEEE International Conference on Multimedia and Expo, ICME03, 2, pp. 2003.

[10] E. Jung, A. Schwarzbacher, R. Lawlor, "Implementation of real-time AMDF pitch-detection for voice gender normalization", Proceedings of the 14th International Conference on Digital Signal Processing, 2002. DSP 2002, pp. 827 -830, vol.2.

[11] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals IEEE Transactions on Speech and Audio Processing", vol. 10, no. 5. 2002.

[12] S. Slomka, S. Sridharan, "Automatic Gender Identification Optimised For Language Independence", Proceeding of IEEE TENCON- Speech and Image Technologies for Computing and Telecommunications pp 145-148. 2005.

[13] S. Parris, M. Carey. "Language Independent Gender Identification", ICASSP, pp 685-688, 1997.

[14] P. Martland, S. Whiteside, "Analysis of ten vowel sounds across gender and regional cultural accent", Proceedings Fourth International Conference on Spoken Language, 1998. ICSLP 96, Volume: 4, 3-6, pp: 2231 -2234 vol.4.

[15] H. Ting, Y. Yingchun, W. Zhaohui, "Combining MFCC and Pitch to Enhance the Performance of the Gender Recognition", 0-7803-9737, Signal Processing, 2006 8th International Conference.
[16] Y. Hu, D. Wu, A. Nucci, "Pitch-based Gender Identification with Two-stage Classification" , Security and Communication Networks Volume 5, Issue 2, pp: 211–225, February 2012.
[17] T. Kinnunen, H. Li, "An overview of text-independent speaker recognition: From features to supervectors", Speech Communication, vol. 52, pp. 12-40, 2010.
[18] W. Campbell, D. Sturim, "Support vector machines using GMM supervectors for speaker verification", IEEE Signal Processing Letters, vol. 13, no. 5, pp. 308-311, 2006.
[19] C. Longworth, M. Gales, "Combining derivative and parametric kernels for speaker verification", IEEE Trans. Audio, Speech Language Process., vol. 6, no. 1, pp. 1-10, 2007.
[20] V. Wan, S. Renals, "Speaker verification using sequence discriminant support vector machines", IEEE Transactions on Speech and Audio Processing, vol. 13, no. 2, pp. 203–210, 2005.
[21] W.Campbell, J. Campbell, "Support vector machines for speaker and language recognition", Computer Speech and Language, vol. 20, no. 2-3 SPEC. ISS., pp. 210-229, 2006.
[22] Vladimir Vapnik. "The Nature of Statistical Learning Theory", Springer-Verlag, 1995. ISBN 0-387-98780-0.
[23] M. Arcienega, A. Drygajlo, "Pitch-dependent GMMs for text-independent speaker recognition system", Proc. Eurospeech, Aalborg, Denmark, pp. 2821-2824, Sept. 2001.
[24] Q.Y. Hong, S. Kwong, "A genetic classification method for speaker recognition", Engineering Applications of Artificial Intelligence, vol. 18, pp. 13–19, 2005.
[25] D.A. Reynolds, R.C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models", IEEE Transactions on speech and audio processing, Vol.3, No.1, January 1995.
[26] W.M. Campbell, J.P. Campbell, D.A. Reynolds, E. Singer, "Support vector machines for speaker and language recognition", Computer Speech and Language, vol. 20, no. 2-3 SPEC. ISS., pp. 210-229, 2006.
[27] Xiaodan Zhuang, Xi Zhou, Mark A. Hasegawa-Johnson, and Thomas S. Huang, "Real-world acoustic event detection", Pattern Recognition Letters, vol. 31, no. 12, pp. 1543-1551, Sept. 2010.
[28] Xiaodan Zhuang, Xi Zhou, Mark Hasegawa-Johnson, and Thomas Huang, "Face age estimation using patch-based hidden Markov model supervectors", in 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL, 2008, pp. 1-4.
[29] C. Longworth and M. Gales, "Combining derivative and parametric kernels for speaker verification", IEEE Trans. Audio, Speech Language Process., vol. 6, no. 1, pp. 1-10, 2007.
[30] W.M. Campbell, D.E. Sturim, and D.A. Reynolds, "Support vector machines using GMM supervectors for speaker verification", IEEE Signal Processing Letters, vol. 13, no. 5, pp. 308-311, 2006b.
[31] N. Dehak and G. Chollet, "Support vector GMMs for speaker verification", in Proc. IEEE Odyssey: The Speaker and Language Recognition Workshop, San Juan, Puerto Rico, 2006.
[32] K.A. Lee, C. You, H. Li, T. Kinnunen, and D. Zhu, "Characterizing speech utterances for speaker verification with sequence kernel SVM", in Proc. Ninth Interspeech (Interspeech 2008), Brisbane, Australia, 2008, pp. 1397–1400.

[33] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn, “Speaker verification using adapted Gaussian mixture models”, Digital Signal Processing: A Review Journal, vol. 10, no. 1, pp. 19-41, 2000.
[34] N. Dehak, R. Dehak, P. Kenny, and P. Dumouchel, “Comparison between factor analysis and GMM support vector machines for speaker verification”, in The Speaker and Language Recognition Workshop (Odyssey 2008), Stellenbosch, South Africa, 2008.
[35] C.H. You, K.A. Lee, and H. Li, “An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition”, IEEE Signal Process. Lett, vol. 16, no. 1, pp. 49-52, 2009.
[36] K.A. Lee, C. You, H. Li, and T. Kinnunen, “A GMM-based probabilistic sequence kernel for speaker recognition”, in Proc. Interspeech 2007 (ICSLP), Antwerp, Belgium, 2007, pp. 294-297.
[37] A. Reynolds Douglas and C. Rose Richard, “Robust text-independent speaker identification using Gaussian mixture speaker models”, IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72-83, 1995.
[38] K.K. Paliwal, “On The Use Of Fliter-Bank Energies As Fetures For Robust Speech Recognition”, ISSPA, 1999.
[39] M. Gelfer and V. Mikos, “The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels”, Journal of Voice, vol. 19, no. 4, pp. 544–554, 2005.
[40] H. Ting, Y. Yingchun, and W. Zhaohui, “Combining MFCC and pitch to enhance the performance of the gender recognition”, in Signal Processing, 2006 8th International Conference on, vol. 1, 2006.
[41] X. Sun, "A Pitch Determination Algorithm Based On Subharmonicto-Harmonic Ratio", Department of Communication Sciences and Disorders, Northwestern University2299 N. Campus Dr., Evanston, IL 60208, USA, 2008.
[42] Chang, Chih-Chung; Lin, Chih-Jen, “LIBSVM: A library for support vector machines”, ACM Transactions on Intelligent Systems and Technology 2,2011.
[43] M. Bijankhan, “Great Farsdat Database”, Technical report, Research center on Inteligent Signal Proccessing, 2002.
[44] Yiming Yang “An Evolution of statistical Approaches to Text Categorization”, Information Retrieval 1, pp.69-90 1999.
[45] Jolliffe I.T. “Principal Component Analysis, Series: Springer Series in Statistics”, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4