

تعیین مرد یا زن بودن گوینده از روی صدا با استفاده از ماشین بردار پشتیبان

محمد مهدی همایون پور، مریم خراشادیزاده

آزمایشگاه سیستم‌های هوشمند صوتی و گفتاری

دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیر کبیر

mkhorashadizadeh@yahoo.com, homayoun@ce.aut.ac.ir

چکیده

تمایل به ایجاد ارتباط گفتاری انسان با کامپیوتر و همین‌طور گسترش استفاده از اینترنت و امکان دریافت سرویسهای مختلف اطلاعاتی و خدمات از راه دور، باعث اهمیت روزافزون تکنیکهای مختلف پردازش گفتار شده است. یکی از تکنیکهای پردازش گفتار، تشخیص اتوماتیک زن یا مرد بودن یک فرد از روی صدا است. در این مقاله مسأله تشخیص جنسیت با استفاده از ماشین بردار پشتیبان (SVM) مطرح شده است. تشخیص جنسیت با استفاده از ضرایب کپسترال مبتنی بر آنالیز پیشگوئی خطی و ضرایب کپسترال مبتنی بر معیار مل انجام شده است. در این روش از SVM با هسته RBF و همچنین SVM با هسته چندجمله‌ای استفاده شده است و نتایج برتری هسته RBF را بر هسته چندجمله‌ای در این کاربرد خاص نشان می‌دهد. آزمایشات مختلف حاکی از برتری ضرایب MFCC نسبت به ضرایب LPCC در تعیین جنسیت گوینده می‌باشد، همچنین افزودن مشتق اول به ویژگی‌ها باعث بهبود کارایی سیستم می‌شود.

واژه های کلیدی: تشخیص جنسیت- ماشین بردار پشتیبان- ضرایب کپسترال مبتنی بر پیشگوئی خطی و معیار مل

مقدمه

یا حتی بدون اطلاع کاربر، با استفاده از گفتار شخص جنسیت او را تشخیص داد. اطلاع از زن یا مرد بودن کاربر می‌تواند در نوع و چگونگی اطلاعات و خدماتی که در اختیار فرد مورد نظر قرار می‌گیرد، مؤثر باشد. علاوه بر این در سالهای اخیر که استفاده از سیستمهای تشخیص هویت بیومتریک مثل اثر انگشت، قرنيه چشم و همچنین صدا مطرح شده است، تشخیص زن یا مرد بودن گوینده می‌تواند در این سیستمها نیز مفید بوده و باعث تسريع عمل آنها شود. به عبارت دیگر می‌توان ابتدا زن یا مرد بودن گوینده را تشخیص داد و سپس در گروه زنان و یا مردان، فرد را تعیین هویت کرد. تشخیص جنسیت اتوماتیک با روشهای مختلفی مثل شبکه‌های عصبی پرسپترون چند لایه (MLP) و مدل مخلوط گوسی (GMM) انجام شده است [۵].

در این مقاله، مسأله تشخیص جنسیت از روی صدا با

یکی از مهمترین راههای برقراری ارتباط طبیعی بین انسانها، ارتباط از طریق گفتار است. تمایل به ایجاد ارتباط گفتاری، با کامپیوتر سبب ایجاد یک بستر تحقیقاتی گسترده گردیده است. این تحقیقات در زمینه‌های مختلف مثل بازشناسی گفتار، سنتز و تولید گفتار، تصدیق یا تعیین هویت گوینده و ... انجام گرفته و پیشرفتهای قابل توجهی در این زمینه‌ها به دست آمده است.

امروزه با گسترش فناوری اطلاعات و همگانی شدن استفاده از کامپیوتر، تلفن، شبکه‌های محلی و خصوصاً اینترنت به دلیل دریافت سرویسهای مختلف اطلاعاتی و خدمات از راه دور، بر اهمیت تکنیکهای مختلف پردازش گفتار افزوده شده است. یکی از تکنیکهای پردازش گفتار، تشخیص اتوماتیک جنسیت یا به عبارتی زن یا مرد بودن کاربری است که می‌خواهد به اطلاعات یا خدماتی دسترسی داشته باشد. در این سیستمها می‌توان بدون پرسش صریح از کاربر

که اغلب برای دسته‌بندی باینری مورد استفاده قرار می‌گیرد. ماشین بردار پشتیبان با معرفی یک فضای ویژگی منتج از بکارگیری توابع هسته^۵، داده‌های ورودی را به فضایی با ابعاد بالاتر برده و قابلیت جدایی‌پذیری داده‌هایی که در حالت عادی به صورت خطی جدایی‌پذیر نیستند را افزایش می‌دهد. نوعاً این کار، یعنی بردن بردارهای ورودی به فضاهای ابعاد بالاتر، با افزایش پیچیدگی محاسباتی و مشکل همپوشانی همراه است. اما ماشین‌های بردار پشتیبان با فضای ابعاد بالاتر به صورت مستقیم ارتباط ندارند و تنها به روابط ضرب داخلی در این فضا نیاز دارند. SVM ها دارای خواص زیر هستند:

۱. طراحی دسته‌بندی کننده با حداکثر تعمیم
۲. رسیدن به بهینه سراسری تابع هزینه
۳. تعیین خودکار ساختار و توپولوژی بهینه برای دسته‌بندی کننده
۴. مدل کردن توابع تمایز غیرخطی با استفاده از هسته‌های غیرخطی و مفهوم ضرب داخلی

ماشین بردار پشتیبان خطی

فرض کنید تعدادی از بردارهای ویژگی یا الگوهای آموزشی به صورت $\{x_1, x_2, \dots, x_i, \dots, x_N\}$ داریم، که هر کدام یک بردار ویژگی d بعدی بوده و دارای برچسب y_i است و $y_i \in \{-1, +1\}$. هدف، حل یک مسأله دسته‌بندی دو کلاسه به صورت بهینه است. فرض کنید، بخواهیم این دو کلاس را با تابع تمایز $f(x)$ و با یک ابرصفحه H با معادلات زیر از هم جدا کنیم:

$$H: wx + b = 0 \quad (1)$$

$$f(x) = \text{sign}(wx + b) \quad (2)$$

بردار w ، بردار عمود بر ابرصفحه جدا کننده و b بایاس

است (شکل ۱). $\frac{|b|}{\|w\|}$ فاصله مبدأ تا ابرصفحه است و $\|w\|$ نرم بردار w است.

$$\|w\| = \left(\sum_{i=1}^d w_i^2 \right)^{1/2} \quad (3)$$

استفاده از ماشین بردار پشتیبان^۱ (SVM) مطرح شده است. ماشین بردار پشتیبان یکی از ایده‌های جدید در شناسایی و دسته بندی الگو است. این روش، دارای خواص بسیار ارزشمندی است که آن را برای شناسایی الگو و همچنین این کاربرد خاص مناسب می‌سازد. یکی از خواص مهم SVM این است که یک طبقه‌بندی کننده با حداکثر تعمیم ایجاد می‌کند، در صورتی که در هیچیک از طبقه‌بندی کننده‌های الگو، خاصیت تعمیم^۲ طبقه‌بندی کننده به طور مستقیم در تابع هزینه دخالت داده نشده است. خاصیت دیگر آن، این است که بر خلاف بسیاری از روشها، SVM با مشکل بهینه‌های محلی مواجه نیست و نکته قابل ذکر دیگر، تعیین ساختار و توپولوژی بهینه توسط SVM است، در صورتی که تعیین ساختار و توپولوژی طبقه‌بندی کننده یکی از معضلات مطرح در طراحی طبقه‌بندی کننده‌ها است. به عنوان مثال تعداد بهینه نودهای لایه مخفی در شبکه عصبی MLP، تعداد توابع گوسی در شبکه RBF یا تعداد بهینه حالتها و توابع گوسی در مدل مخفی مارکف از مشکلات این روشها است. در ضمن در این مقاله تشخیص جنسیت با کمک صدا و ویژگی‌های LPCC و MFCC انجام شده و تأثیر افزودن مشتقات آنها نیز مورد بررسی قرار گرفته است.

در بخش ۲ این مقاله به طور مختصر ماشین بردار پشتیبان توضیح داده خواهد شد. بخش ۳ به دادگان گفتاری و استخراج ویژگی اختصاص دارد. در بخش ۴ جزئیات پیاده سازی، آموزش و بازشناسی و همچنین نتایج آزمایشات انجام شده بیان می‌شود و در پایان نیز یک جمع بندی از مطالب و نتایج به دست آمده ارائه می‌گردد.

ماشین بردار پشتیبان

یک محقق روسی به نام ولادیمیر وپنیک^۳ در سال ۱۹۶۵ گامی مهم در طراحی دسته بندی کننده‌ها برداشت و نظریه آماری یادگیری^۴ را به صورت مستحکمتری بنا نهاد و ماشین بردار پشتیبان را بر این اساس ارائه داد[۴]. ماشین بردار پشتیبان یک روش یادگیری نسبتاً جدید است

¹ Support Vector Machine

² Generalization

³ Vladimir Vapnik

⁴ Statistical Learning Theory

⁵ kernels

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^N \alpha_i (y_i (w x_i + b) - 1) \quad (10)$$

برای اینکه (w, b, α) جواب مسأله باشد، این جواب باید در شرایط KKT^2 صدق کند و در نقطه جواب، مشتق L نسبت به w ، b و α برابر صفر باشد. با مساوی قرار دادن مشتق برابر با صفر به معادلات زیر خواهیم رسید:

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (11)$$

$$\sum_{i=1}^N \alpha_i y_i = 0$$

با قرار دادن مقدار w از رابطه فوق در $L(w, b, \alpha)$ به مسأله دوگان^۳ برای بهینه‌سازی مقید خواهیم رسید:

$$\begin{cases} \text{Maximize}_w & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^N \alpha_i \\ \text{Subject to} & \alpha_i \geq 0 \text{ for } i=1, 2, \dots, N \\ & \sum_{i=1}^N \alpha_i y_i = 0 \end{cases} \quad (12)$$

پس از حل این مسأله دوگان بهینه‌سازی، ضرایب لاگرانژ $\alpha_i \geq 0$ به دست می‌آیند. در واقع هر کدام از ضرایب لاگرانژ α_i متناظر با یکی از الگوهای x_i می‌باشند، الگوهای x_i که متناظر با ضرایب $\alpha_i > 0$ (مثبت) هستند، بردارهای پشتیبان^۴ (sv_i) نامیده می‌شوند. مقدار بردار وزن و بایاس از روابط زیر به دست می‌آید:

$$w = \sum_{i=1}^{N_{sv}} \alpha_i y_i sv_i$$

$$b_j = y_j - \sum_{i=1}^{N_{sv}} \alpha_i y_i sv_i sv_j \quad (13)$$

$$b = \frac{1}{N_{sv}} \sum_{j=1}^{N_{sv}} b_j$$

تابع تمایز برای طبقه‌بندی یک الگوی x به صورت زیر خواهد بود:

$$f(x) = \text{sign} \left(\sum_{i=1}^{N_{sv}} \alpha_i y_i x sv_i + b \right) \quad (14)$$

تا به حال مسأله SVM برای حالت جدایی‌پذیر بررسی شده است، حال آنکه بیشتر مسائل به صورت جدایی‌ناپذیر

هر ابرصفحه (w, b) را می‌توان بدون هیچ تفاوتی به صورت $(\lambda w, \lambda b)$ برای هر مقدار حقیقی مثبت λ نشان داد. بنابراین ابرصفحه‌های مبنایی تعریف می‌شود که نمونه‌ها را با یک حداقل فاصله از ابرصفحه متمایز کننده جدا کند. به این ترتیب به صورت تلویحی مقدار λ برای ابرصفحه متمایز کننده تنظیم می‌شود. قیود فوق را می‌توان به صورت زیر فرموله کرد:

$$x_i w + b \geq +1 \quad \text{for } y_i = +1 \quad (4)$$

$$x_i w + b \leq -1 \quad \text{for } y_i = -1 \quad (5)$$

این دو نیز در هم ادغام می‌شوند:

$$y_i (x_i w + b) - 1 \geq 0 \quad (6)$$

نقاطی که حالت تساوی معادلات (۴) و (۵) را ارضا می‌کنند، به ترتیب روی ابرصفحه‌های H_1 و H_2 قرار می‌گیرند که معادلات زیر را دارا هستند:

$$H_1: w x_i + b = +1 \quad (7)$$

$$H_2: w x_i + b = -1 \quad (8)$$

ناحیه بین دو ابرصفحه H_1 و H_2 را حاشیه یا ناحیه مرزی^۱

گویند. فاصله بین دو ابرصفحه H_1 و H_2 برابر $\frac{2}{\|w\|}$ خواهد

بود. هدف ماشین بردار پشتیبان این است که ابرصفحه متمایز کننده‌ای با بزرگترین مقدار حاشیه را پیدا کند. طراحی ابرصفحه با بیشترین عرض ناحیه مرزی یا ناحیه مرزی بهینه بدان معناست که با شرط درست دسته بندی شدن الگوها، عرض ناحیه مرزی حداکثر شود. به عبارت

دیگر باید $\frac{2}{\|w\|}$ ماکزیمم شود و $\|w\|$ مینیمم گردد. پس در

واقع طراحی یک طبقه‌بندی کننده ابرصفحه‌ای با ناحیه مرزی بهینه به صورت زیر خواهد بود:

$$\begin{cases} \text{Minimize}_w & \frac{1}{2} \|w\|^2 \\ \text{Subject to} & w_i (w x_i + b) \geq 1 \\ & \text{for } i = 1, 2, \dots, N \end{cases} \quad (9)$$

مسأله فوق یک مسأله بهینه‌سازی مقید است. برای حل این مسأله، تابع لاگرانژی زیر را تشکیل داده و ضرایب لاگرانژ α_i را به دست می‌آورند:

² Karush-Kuhn-Tucker Conditions

³ Dual Problem

⁴ Support Vectors

¹ Margin

$$k(x_a, x_b) = \phi(x_a) \cdot \phi(x_b) \quad (17)$$

به این توابع، توابع هسته می‌گویند. در این حالت فرم تابع ابر صفحه متمایز کننده به صورت زیر خواهد بود:

$$f(x) = \text{sign} \left(\sum_{i=1}^{N_{sv}} \alpha_i y_i k(x_i, x) + b \right) \quad (18)$$

توابع هسته متعددی موجود است. در زیر تعدادی از این توابع آورده شده است:

$$1. \quad k(x, y) = (x \cdot y + 1)^P \quad \text{چندجمله‌ای از درجه } P$$

$$2. \quad \text{تابع پایه شعاعی گوسی (RBF)}$$

$$k(x, y) = e^{-\frac{(x-y)^2}{2\sigma^2}}$$

$$3. \quad \text{تانژانت هیپربولیک}$$

$$k(x, y) = \tanh(x \cdot y + \theta)$$

دادگان گفتاری و استخراج ویژگی

در این تحقیق از دادگان فارسی فارس دات استفاده گردیده است [۷]. داده‌های صوتی ۴۱ مرد و ۳۶ زن با لهجه تهرانی استفاده شده است. هر گوینده ۱۰ جمله را که به صورت تصادفی از بین مجموعه‌ای از جملات انتخاب شده اند، بیان می‌کنند. جملات در دو جلسه متفاوت بیان شده‌اند. در پیاده سازی انجام شده از گفتار ۲۱ مرد و ۱۶ زن برای آموزش و از داده‌های گفتاری ۲۰ مرد و ۲۰ زن باقیمانده برای آزمایش استفاده شده است. ضرایب کپسترال حاصل از آنالیز پیشگونی خطی (LPCC) و ضرایب کپسترال حاصل از طیف فوریه مبتنی بر معیار مل (MFCC)، یا مستقیماً یا به همراه مشتق اولشان، به عنوان ویژگی‌های استخراج شده از فایل‌های صوتی در نظر گرفته شده و برای دسته‌بندی استفاده شده‌اند.

در بین کلمات بیان شده توسط گوینده معمولاً سکوت وجود دارد. چون این سکوت در بین گویندگان زن و مرد یکسان است، لذا برای تمایز بین آنها مفید نبوده و بنابراین سکوت بین جملات حذف شده و از گفتار بدون سکوت در آموزش و آزمایش استفاده می‌شود.

برای به دست آوردن ویژگی‌ها از پنجره‌های همینگ به طول ۳۰ میلی ثانیه استفاده شده است. مقدار پیش تأکید برابر ۰/۹۷۵ و میزان همپوشانی فریمها ۱۰ میلی ثانیه فرض شده است. تعداد ضرایب استخراج شده از هر فریم

هستند. برای حالت جدایی‌ناپذیر یک دسته متغیر به نام متغیر کمبود^۱ تعریف می‌شود، به طوریکه شرایط زیر برقرار باشد:

$$y_i (wx_i + b) \geq 1 - \zeta_i \quad \text{for } i = 1, 2, \dots, N \quad (15)$$

واضح است که هر چقدر مجموع مقادیر متغیرهای کمبود بیشتر شود، از حالت بهینه دورتر شده و خطا بیشتر خواهد شد. اگر مسأله بهینه‌سازی مقید با توجه به رابطه فوق حل شود، روابطی مشابه حالت جدایی‌پذیر به دست می‌آید، با این تفاوت که محدوده تغییرات ضرایب لا گرانش و شرط انتخاب بردارهای پشتیبان تفاوت می‌کند (برای اطلاعات بیشتر به مقالات ۳ و ۴ رجوع کنید).

ماشین بردار پشتیبان غیرخطی

در حالت غیرخطی، می‌توان با اعمال پیش پردازش روی داده‌ها، مسأله را به فضایی برد که در آنجا با یک ابرصفحه ساده قابل حل باشد. برای این منظور یک نگاشت $\phi(x)$ تعریف می‌شود که بردار ورودی d بعدی x را به بردار d' بعدی z تبدیل کند:

$$\phi: \mathcal{R}^d \rightarrow \mathcal{R}^{d'} \quad (16)$$

$\phi(x)$ باید به گونه‌ای انتخاب شود، که بردارهای ویژگی جدید جدایی‌پذیر باشند. در حالت کلی می‌توان گفت که اگر $\phi(x)$ بردارهای ورودی را به فضایی ببرد، که تعداد ابعاد آن به اندازه کافی بزرگ باشد ($d' \gg d$)، بردارها جدایی‌پذیر خواهند شد. اما انتقال به فضای با ابعاد بسیار بزرگ، عملاً می‌تواند به لحاظ بار محاسباتی و فضای مورد نیاز غیر عملی باشد. همچنین امکان ایجاد مشکل همپوشانی نیز افزایش می‌یابد.

ماشین بردار پشتیبان در فضای جدید برای حل مسأله بهینه‌سازی به سادگی به جای x مقدار $\phi(x)$ را قرار می‌دهد. با این جایگذاری می‌توان دید که در روابط همیشه ترکیبی از دو $\phi(x)$ به صورت ضرب داخلی در کنار هم دیده می‌شود. پس اگر بتوان تابعی برای ضرب داخلی در فضای ویژگی جدید با ابعاد بالاتر پیدا کرد، دیگر مستقیماً نیازی به نگاشت $\phi(x)$ نخواهد بود.

¹ Slack Variables

۱۲ تا است که در صورت استفاده از مشتق اول، طول بردار ویژگی به ۲۴ عنصر افزایش می‌یابد [۱].

پیاده سازی و نتایج آزمایشات

برای ارزیابی کارایی روش ماشین بردار پشتیبان در تعیین جنسیت افراد، اقدام به پیاده سازی این روش نموده و آزمایشاتی به شرح زیر انجام شده است:

ابتدا ویژگی‌های موردنظر مطابق با آنچه در بخش ۳ بیان شد، استخراج شدند و این ویژگی‌ها برحسب زن یا مرد بودن به صورت صحیح بر چسب گذاری شدند (۱ : مرد و ۰ : زن). این بردارها برای آموزش SVM مورد استفاده قرار گرفتند. در این تحقیق ابتدا از SVM با تابع هسته RBF استفاده شده است و نتایج گزارش شده به ازاء بهترین σ به دست آمده برای تابع هسته RBF می‌باشد. برای بررسی کارایی SVM داده‌های گفتاری با طول زمانی تقریباً ۴ ثانیه (۲۰۰ بردار ویژگی) به کار برده شده است. یک داده صوتی وقتی مرد(زن) تشخیص داده می‌شود، که در صورت اعمال آن به ماشین بردار پشتیبان بیش از نیمی از بردارهای ویژگی مربوط به آن فایل صوتی بر چسب مرد(زن) را دریافت کنند. نسبت بردارهای ویژگی با برچسب صحیح به کل بردارهای ویژگی هر داده صوتی نیز به عنوان دقت تشخیص جنسیت، محاسبه شده است و میانگین دقت داده‌های آزمایشی به عنوان یک معیار دقیقتر برای ارزیابی طبقه‌بندی کننده ارائه شده است. به تعبیر دیگر میانگین محاسبه شده، نسبت بردارهای ویژگی با برچسب صحیح به کل بردارهای ویژگی یک فایل صوتی بزرگ است که از دنبال هم قرار دادن همه داده‌های آزمایشی به دست می‌آید. این میانگین نشان دهنده دقت SVM در تشخیص جنسیت هر بردار ویژگی تنهاست.

ماشین‌های بردار پشتیبان با مجموعه‌های آموزشی مختلف (از نظر حجم داده‌های آموزشی) تحت آموزش قرار گرفتند. برای آموزش SVM از مجموعه‌های آموزشی تقریباً ۴، ۵، ۶، ۸، ۱۶ دقیقه (۱۲۰۰۰، ۱۵۰۰۰، ۱۸۰۰۰، ۲۴۰۰۰، ۴۸۰۰۰ بردار ویژگی) استفاده شده و نتایج حاصل از آنها با هم مقایسه گردیده است. در ابتدا از ویژگی LPCC و همین طور LPCC و مشتق اول آن برای آموزش یک SVM با تابع هسته RBF استفاده شده است. آزمایشات با σ های

مختلف انجام شده است. نتایج به دست آمده حاکی از این مسأله بود که با $\sigma \approx 2/67$ راندمان بالاتری نسبت به مقادیر دیگر σ به دست می‌آید. جدول ۱ و ۲ نتایج به دست آمده از بردارهای ویژگی‌های LPCC و LPCC و مشتق اول آنها را با استفاده از SVM با تابع هسته RBF و $\sigma \approx 2/67$ نشان می‌دهد.

در ادامه، آزمایشات فوق با استفاده از ویژگی‌های MFCC و MFCC و مشتق اول آنها دوباره تکرار شده است. در اینجا، $\sigma \approx 8/45$ سبب ایجاد راندمان بهتری نسبت به مقادیر دیگر می‌شود. نتایج به دست آمده در جدول ۳ و ۴ نشان داده شده است.

با مقایسه جداول ۱، ۲، ۳ و ۴ می‌توان گفت که بطور کلی ضرایب LPCC و MFCC هر دو ویژگی‌های مناسبی در تشخیص جنسیت با روش ماشین بردار پشتیبان هستند، ولی به طور نسبی ضرایب MFCC راندمان بالاتری نسبت به ضرایب LPCC دارند و همچنین استفاده از این ویژگی‌ها به همراه مشتق اول‌شان نتایج بهتری از استفاده این ویژگی‌ها به تنهایی دارد.

نکته دیگری که با مقایسه جداول می‌توان استنباط کرد، این است که بطور کلی با افزایش داده‌های آموزشی راندمان سیستم بالا می‌رود. البته در بعضی از جداول با افزایش داده‌های آموزشی راندمان بازشناسی کاهش می‌یابد، ولی اگر میانگین دقت بازشناسی مردان و زنان در نظر گرفته شود، با افزایش داده‌های آموزشی یک روند افزایشی خواهیم داشت. ولی همان‌طور که جداول نشان می‌دهد راندمان به دست آمده از مجموعه‌های آموزشی با حجم کم نیز نسبت به راندمان مجموعه‌های بزرگتر قابل قبول است.

بعد از انجام آزمایش با تابع هسته RBF، از تابع هسته چندجمله‌ای استفاده شده است. در این آزمایشات از ۶ دقیقه (۱۸۰۰ بردار ویژگی) گفتار برای آموزش SVM استفاده شده و مطابق حالت قبل از ۴ ثانیه گفتار برای آزمایش استفاده شده است. جداول ۵ و ۶ نتایج به دست آمده از تابع هسته چندجمله‌ای با درجات مختلف را نشان می‌دهد.

با دقت در جداول ۵ و ۶ می‌توان دید که به طور کلی هسته RBF نتایج بهتری نسبت به هسته چندجمله‌ای داشته است. مطلب دیگری که می‌توان از جداول فهمید این مسأله

راندمان بالاتری را نسبت به هسته چندجمله‌ای نشان می‌دهد ولی در صورت استفاده از تابع هسته چندجمله‌ای می‌توان گفت که چندجمله‌ای با درجه ۳ به نتایج بهتری منجر می‌شود. در این پژوهش تأثیرات حجم داده‌های آزمایشی هم مورد توجه قرار گرفته است که نتایج گویای تغییرات ناچیز در کارایی سیستم با افزایش حجم داده‌های آزمایشی می‌باشد. در انتها می‌توان گفت ماشین بردار پشتیبان به عنوان یک روش جدید دارای قابلیت‌های ارزشمندی است و می‌تواند در بسیاری از مسائل پردازش گفتار از جمله تعیین جنسیت در کاربردهای عملی مورد استفاده قرار گیرد.

مراجع

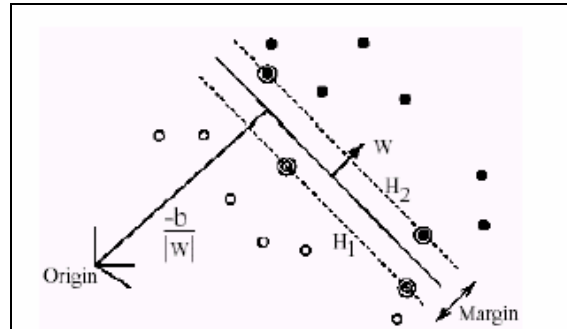
- 1- S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchene, P. Woodland, The HTK Book (HTK Version 3.0), July 2000.
- 2- J. K. Deller, J. G. Proakis, J. H. Hansen, "Discrete-Time Processing of Speech Signals", Macmillan Publishing company, 1993.
- 3- C. J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2, pp. 121-167, 1998.
- ۴- جهان‌شاه کبودیان، محمد رحمتی، محمدمهدی همایون‌پور، "ماشین بردار پشتیبان: روشی برای دسته‌بندی الگوها"، اولین کنفرانس فناوری اطلاعات و دانش، ص. ۴۸۱-۴۸۹، ۱۳۸۲.
- ۵- محمد مهدی همایون‌پور، نرگس احمیدی، بهاره مبارک‌آبادی، "تعیین محدوده سنی و زن یا مرد بودن به کمک صدا با استفاده از مدل مخلوط گوسی و شبکه عصبی"، یازدهمین کنفرانس مهندسی برق ایران، ۱۶-۱۸ اردیبهشت ۱۳۸۲.
- ۶- محمدمهدی همایون‌پور، گفتارپردازی رقمی، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، زمستان ۸۱.
- ۷- دادگان صوتی فارس دات، مرکز پردازش علائم هوشمند، تهران.

است که در صورت استفاده از ضرایب LPCC و مشتق اول آنها چندجمله‌ای با درجه ۱ قابلیت جداسازی نمونه‌های صوتی را ندارد و به نظر می‌رسد که چندجمله‌ای با درجه ۳ برای هر دو نوع ضرایب نتایج بهتری داشته است. در ضمن با توجه به آزمایشات باز هم می‌توان برتری ضرایب MFCC بر ضرایب LPCC را در بخش تشخیص جنسیت گوینده دید.

در ادامه آزمایشات، SVM با هسته RBF به ازاء حجم‌های مختلف داده‌های آزمایشی مورد بررسی قرار گرفته است. در این آزمایشات بعد از آموزش SVM با ۶ دقیقه گفتار، تشخیص جنسیت روی داده‌های صوتی ۲، ۴، ۶، ۸ ثانیه انجام شده و نتایج آن در جدول ۷ نشان داده شده است. با بررسی جدول ۷ به این نتیجه می‌توان رسید که افزایش حجم داده‌های صوتی آزمایش تأثیر زیادی در افزایش کارایی ندارد، بلکه در مواردی سبب کاهش کارایی نیز شده است.

نتیجه گیری

در این تحقیق، ماشین بردار پشتیبان برای تشخیص جنسیت افراد بکار گرفته شده است. راندمان بالای به دست آمده با استفاده از حجم کمی داده‌های آموزشی و آزمایشی نشان دهنده مناسب بودن این روش برای این کاربرد است. آزمایشات انجام شده نشان می‌دهد، که افزایش حجم داده‌های آموزش سبب افزایش راندمان سیستم می‌شود، ولی با در نظر گرفتن حجم داده آموزشی نسبتاً کم هم نتایج قابل قبولی به دست می‌آید. نکته دیگری که از مجموع نتایج برمی‌آید، این است که به طور کلی ویژگی‌های MFCC و LPCC برای این کاربرد مناسب هستند، ولی ویژگی‌های MFCC نتایج بهتری نسبت به LPCC دارند و همچنین استفاده از مشتق اول این ویژگی‌ها در بهبود کارایی تأثیر دارد. در این تحقیق برای این کاربرد خاص، تابع هسته RBF با هسته چندجمله‌ای مقایسه شده است و نتایج به دست آمده از هسته RBF



شکل ۱- ابرصفحات جدا کننده خطی برای موارد جدایی پذیر.
نمونه‌های دایره دار بردارهای پشتیبان هستند.

جدول ۱- نتایج حاصل از تعیین جنسیت توسط SVM و با استفاده از ضرایب LPCC

$$\sigma \approx 2/67$$

حجم داده‌های آموزشی (دقیقه)	۴	۵	۶	۸	۱۶
راندمن بازشناسی مردان (%)	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (%)	۷۵/۷	۷۷/۲	۷۷/۳	۷۸/۲	۷۷/۷
راندمن بازشناسی زنان (%)	۹۰	۸۵	۹۰	۹۰	۹۰
میانگین دقت بازشناسی زنان (%)	۶۵/۸	۶۴/۸	۶۶/۲	۶۶/۵	۶۹/۳

جدول ۲- نتایج حاصل از تعیین جنسیت توسط SVM و با استفاده از ضرایب LPCC و مشتق اول آنها

$$\sigma \approx 2/67$$

حجم داده‌های آموزشی (دقیقه)	۴	۵	۶	۸	۱۶
راندمن بازشناسی مردان (%)	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (%)	۷۶/۲	۷۷	۷۷/۶	۷۸/۳	۷۸/۷
راندمن بازشناسی زنان (%)	۹۰	۹۰	۹۰	۹۰	۹۰
میانگین دقت بازشناسی زنان (%)	۶۶/۹	۶۷/۶	۶۸/۸	۶۹/۷	۷۲/۵

جدول ۳- نتایج حاصل از تعیین جنسیت توسط SVM و با استفاده از ضرایب MFCC

$$\sigma \approx 8/45$$

حجم داده‌های آموزشی (دقیقه)	۴	۵	۶	۸	۱۶
راندمن بازشناسی مردان (%)	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (%)	۸۵/۷	۸۹/۹	۸۹/۸	۹۰/۱	۸۹/۴
راندمن بازشناسی زنان (%)	۹۵	۹۵	۹۰	۹۵	۹۵
میانگین دقت بازشناسی زنان (%)	۸۱/۱	۷۷/۳	۷۷/۸	۷۷/۷	۷۸/۶

جدول ۴- نتایج حاصل از تعیین جنسیت توسط SVM و با استفاده از ضرایب MFCC و مشتق اول آنها

$$\sigma \approx 8/45$$

حجم داده‌های آموزشی (دقیقه)	۴	۵	۶	۸	۱۶
راندمن بازشناسی مردان (/.)	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (/.)	۸۵/۵	۸۹/۰	۸۹/۰۵	۸۹/۸	۸۹/۰۸
راندمن بازشناسی زنان (/.)	۹۵	۹۰	۱۰۰	۱۰۰	۹۵
میانگین دقت بازشناسی زنان (/.)	۸۱/۵	۷۸/۸	۷۹/۵۸	۷۹/۳۵	۸۰/۳۵

جدول ۵- نتایج حاصل از تعیین جنسیت توسط SVM با هسته چندجمله‌ای و با استفاده از ضرایب LPCC

و مشتق اول آنها

نوع هسته	RBF	چندجمله‌ای			
	$\sigma \approx 2/67$	P=۱	P=۲	P=۳	P=۴
راندمن بازشناسی مردان (/.)	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (/.)	۷۷/۶	۱۰۰	۸۱/۰	۸۱/۳	۸۲/۵
راندمن بازشناسی زنان (/.)	۹۰	۰	۷۵	۸۵	۵۰
میانگین دقت بازشناسی زنان (/.)	۶۸/۸	۰	۶۰/۱	۵۹/۰	۵۲/۹

جدول ۶- نتایج حاصل از تعیین جنسیت توسط SVM با هسته چندجمله‌ای و با استفاده از ضرایب MFCC

و مشتق اول آنها

نوع هسته	RBF	چندجمله‌ای			
	$\sigma \approx 8/45$	P=۱	P=۲	P=۳	P=۴
راندمن بازشناسی مردان (/.)	۱۰۰	۹۵	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (/.)	۸۹/۰۵	۹۰/۲	۹۲/۱	۹۳/۰۵	۹۳/۸
راندمن بازشناسی زنان (/.)	۱۰۰	۹۰	۸۵	۸۵	۸۵
میانگین دقت بازشناسی زنان (/.)	۷۹/۶	۷۴/۱	۷۴/۳	۷۴/۰	۷۲/۷

جدول ۷. نتایج حاصل از تعیین جنسیت توسط SVM و با استفاده از ضرایب LPCC و مشتق اول آنها

$$\sigma \approx 2/68$$

حجم داده‌های آزمایشی (ثانیه)	۲	۴	۶	۸
راندمن بازشناسی مردان (/.)	۹۵	۱۰۰	۱۰۰	۱۰۰
میانگین دقت بازشناسی مردان (/.)	۷۷/۵۵	۷۷/۶۵	۷۷/۶۴	۷۷/۹۸
راندمن بازشناسی زنان (/.)	۹۰	۹۰	۹۰	۹۰
میانگین دقت بازشناسی زنان (/.)	۷۰/۹	۶۸/۸	۶۹/۵	۶۸/۴

