بهبود دقت مدل GMM با استفاده از کرنل PSK در کاربرد تشخیص زبان گفتاری

فهیمه قاسمیان و محمدمهدی همایون‌پور

چکیده: مدل مخلوط گاوسی [1](GMM)، روشی ساده و مؤثر برای مدل‌کردن آماری فضای ویژگی‌هاست که به‌طور گسترده در کاربرد تشخیص زبان مورد استفاده قرار گرفته و از الگوریتم بیشینه‌سازی امید ریاضی برای آموزش پارامترهای این مدل استفاده می‌شود. در این مقاله با توجه به مشکلی که در آموزش مدل GMM وجود دارد، مدلی جدید با نام PAW - GMM ارائه شده است. در این مدل، قدرت هر مؤلفه از مدل GMM در تمایز یک زبان از سایر زبان‌ها، برای تعیین وزن هر مؤلفه در نظر گرفته می‌شود. مدل PAW - GMM به‌دلیل در نظر گرفتن خواص تمایزی مؤلفه‌های مخلوط گاوسی، سبب افزایش دقت سیستم‌های تشخیص زبانی می‌شود که از این مدل به‌عنوان جایگزین مدل GMM استفاده می‌کنند. همچنین یکی از مشکلاتی که در سیستم GMM - PSK - SVM که یکی از بهترین سیستم‌های تشخیص زبان است وجود دارد، پیچیدگی محاسباتی بالا خصوصاً با اضافه‌شدن تعداد زبان‌هاست. از این رو سیستم UBM - PSK - SVM ارائه شده است که با ثابت نگه داشتن دقت سیستم GMM - PSK - SVM، سبب کاهش پیچیدگی محاسباتی آن شده و در نتیجه قدرت تعمیم به زبان‌های بالاتر را افزایش می‌دهد. آزمایش‌های صورت‌گرفته بر روی ۴ سیستم تشخیص زبان مختلف با استفاده از داده‌های مربوط به ۴ زبان انگلیسی، فارسی، فرانسوی و آلمانی دادگان OGI، کارایی تکنیک‌های ارائه‌شده را نشان می‌دهد.

کلید واژه: **تشخیص زبان، کرنل دنباله‌ای PSK، ماشین بردار پشتیبان (SVM)، مدل مخلوط گاوسی (GMM).**

۱- مقدمه

تشخیص خودکار زبان فرایندی است که طی آن، کامپیوتر زبان مربوط به قطعه گفتار ورودی را تشخیص می‌دهد. به‌طور کلی کاربردهای تشخیص زبان در ۲ دسته اصلی قرار می‌گیرند: پیش‌پردازش برای سیستم‌های ماشینی و پیش‌پردازش برای شنوندگان انسانی. به‌عنوان نمونه در صورتی که تنها از گفتار به‌عنوان ورودی سیستم بازیابی اطلاعات چندزبانه استفاده شود، سیستم باید قادر به تشخیص زبان گوینده قبل یا حین تشخیص دستورات گفتاری باشد. تعیین زبان در حین تشخیص نیاز به تعداد زیادی شناساگر گفتار دارد (برای هر زبان یک شناساگر) که به‌طور موازی اجرا شوند. از آنجایی که ممکن است نیاز باشد ده‌ها یا حتی صدها زبان ورودی پشتیبانی شوند، هزینه تهیه سخت‌افزار چنین سیستمی بسیار بالاست. به‌جای این کار می‌توان از یک سیستم که قبل از شناساگر گفتار، عمل تشخیص زبان را انجام دهد، استفاده شود. در این صورت سیستم تشخیص زبان به سرعت لیست زبان‌هایی که احتمال بالاتری نسبت به زبان‌های دیگر برای دستور گفته‌شده دارند را فراهم می‌کند و مدل‌های شناساگر گفتار وابسته به زبان تعداد کمی از زبان‌ها که دارای احتمال بالاتری هستند، بر روی سخت‌افزار بار شده و اجرا می‌شوند.

تاکنون هدف از تحقیقات در زمینه تشخیص زبان گفتاری، توسعه سیستم‌هایی بوده است که تا حد ممکن وابسته به دانش سطح بالا برای تمایز میان زبان‌ها نباشند. اکثر روش‌های موفق در این زمینه بر مبنای ویژگی‌های اکوستیکی، نوایی یا واج‌آرایی (قوانین حاکم بر توالی مجاز واج‌ها) عمل تشخیص را انجام می‌دهند. گرچه واج‌ها به‌طور قابل ملاحظه‌ای میان زبان‌ها مشترک است اما فرکانس تکرار و توالی چندتایی آنها می‌تواند از یک زبان به زبان دیگر تفاوت زیادی داشته باشد [۱]. پس از ارائه سیستم تشخیص زبان PPRLM[2] و موفقیت این سیستم در تشخیص زبان، تحقیقات بیشتری در زمینه اطلاعات واج‌آرایی صورت گرفت. آزمایش‌های صورت‌گرفته بر روی شنوندگان انسانی نشان داده است که شنوندگان چندزبانه قدرت بالاتری در تشخیص زبان‌ها نسبت به شنوندگان تک‌زبانه دارند [۲]. PPR[3] که از مجموعه موازی از شناساگرهای واجی تشکیل شده است، یک روش مؤثر در تبدیل قطعات گفتار ورودی به توالی‌های واجی است که می‌تواند به‌عنوان قسمت ابتدایی[4] سیستم‌های تشخیص زبان به کار رود. PPR - VSM[5] [۳]، نمونه موفق دیگری از سیستم‌های تشخیص زبان است که از PPR در ساختار آن استفاده شده است. برای بهبود این سیستم روش TOPT برای ساخت شناساگرهای آوایی جدید ارائه شد و سبب افزایش دقت سیستم PPR - VSM گردید [۴]. یکی از مشکلاتی که در PPR وجود دارد، این است که برای آموزش به داده‌های آموزشی نیاز دارد که در سطح واج برای بیش از یک زبان، برچسب‌گذاری شده باشند. عمل برچسب‌گذاری در سطح واج فرایندی وقت‌گیر و هزینه‌بر است. برای رفع این مشکل روشی ارائه شده است که در آن از مجموعه‌ای از مدل‌های GMM به‌عنوان شناساگرهای واجی استفاده می‌شود که در این صورت برای آموزش به داده‌ای که در سطح واج برچسب خورده باشد، احتیاج نخواهیم داشت [۵].

دسته دیگر از روش‌ها، روش‌های تمایزی هستند که از ماشین بردار پشتیبان جهت تشخیص زبان استفاده می‌کنند. دسته دیگر از روش‌ها، روش‌های تمایزی هستند که از ماشین بردار پشتیبان جهت تشخیص زبان استفاده می‌کنند و جزء بهترین سیستم‌های تشخیص زبان هستند. در

این مقاله در تاریخ ۱۲ اسفند ماه ۱۳۸۹ دریافت و در تاریخ ۲۹ آذر ماه ۱۳۹۰ بازنگری شد. این تحقیق توسط مرکز تحقیقات مخابرات ایران بر اساس قرارداد شماره T/۵۰۰/۱۴۹۳۹ پشتیبانی شده است.

فهیمه قاسمیان، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران (email: f_ghasemian@aut.ac.ir).

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

1. Gaussian Mixture Model
2. Parallel Phone Recognizers Followed by Language Model
3. Parallel Phone Recognizer
4. Front-End
5. Parallel Phone Recognizer-Vector Space Model

Archive of SID
www.SID.ir

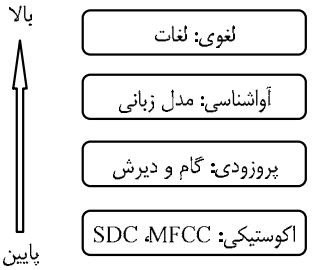
تحقیقات اخیر، کرنل‌هایی موسوم به کرنل دنباله‌ای در ماشین بردار پشتیبان، مورد استفاده قرار گرفته است. این کرنل‌ها برای تبدیل دنباله‌ی بردارهای ویژگی که از قطعه گفتاری با طول متغیر استخراج شده است، به یک بردار ویژگی با ابعاد بالا، مورد استفاده قرار می‌گیرد. سیستم تشخیص زبان تمایزی که از کرنل دنباله‌ای PSK[1] استفاده می‌کند [۶] و در این مقاله GMM - PSK - SVM نامیده شده است، جزء این دسته از روش‌هاست. تاکنون کرنل‌های دنباله‌ای مختلفی ارائه شده است که از جمله می‌توان به کرنل‌های دنباله‌ای GLDS[2] [۷]، کرنل GSV[3] [۸] و MLLR [۹] اشاره نمود. اخیراً نیز سیستم‌های تشخیص زبان اکوستیکی مبتنی بر JFA [۱۰] و i - vector [۱۱] ارائه شده است که از دقت بالایی برخوردارند.

با نگاهی به تاریخچه سیستم‌های تشخیص زبان گفتاری می‌توان مشاهده کرد که مدل GMM در اکثر سیستم‌های موفق مورد استفاده قرار گرفته است. مرجع [۱۲] برای افزایش دقت مدل GMM از همه داده‌های آموزشی استفاده نکرد، بلکه ابتدا از یک مجموعه آموزشی کوچک شروع کرده و مدل‌ها را آموزش داد. سپس کل نمونه‌های مجموعه آموزشی را مورد ارزیابی قرار داده و آنهایی را که بر اساس میزان عدم قطعیت که با توجه به معیار آنتروپی یا احتمال نسبی تعیین می‌شود، مرتب کرده و درصدی از این لیست را به مجموعه آموزشی قبلی اضافه نمود و مدل‌ها را بر اساس این مجموعه آموزشی جدید آموزش داد و این کار را تا زمانی که دقت مجموعه ارزیابی افزایش یابد، ادامه داد.

الگوریتم رایج برای آموزش پارامترهای مدل GMM، الگوریتم بیشینه‌سازی امید ریاضی (EM)[4] است. در این الگوریتم هر مدل مستقل از سایر مدل‌ها با استفاده از داده‌های مربوط به آن زبان آموزش داده می‌شود و در آن از هیچ‌گونه اطلاعات تمایزی برای آموزش مدل‌ها استفاده نمی‌شود. بنابراین یک راه برای افزایش دقت مدل‌های GMM می‌تواند تغییر مکانیزم وزن‌دهی مؤلفه‌ها بر اساس اطلاعات تمایزکنندگی آنها باشد. بر این اساس در این مقاله مدلی جدید با نام PAW[5] - GMM ارائه شده است که مشابه مدل GMM با استفاده از الگوریتم EM آموزش داده می‌شود اما پس از آموزش مدل‌های GMM، وزن مؤلفه‌های هر مدل بر اساس قدرت تمایزکنندگی هر مؤلفه و در قالب سیستم GMM - PSK - SVM تطبیق داده می‌شود.

یکی از مشکلات سیستم GMM - PSK - SVM به‌عنوان یکی از بهترین سیستم‌های تشخیص زبان، پیچیدگی محاسباتی بالا خصوصاً با افزایش تعداد زبان‌هاست. یک راه حل موجود این است که زیرمجموعه‌ای از زبان‌ها را برای آموزش مدل‌های GMM مورد استفاده قرار دهیم و از همه زبان‌ها برای مدل‌کردن فضای آوایی استفاده نکنیم. این راه حل سبب کاهش دقت سیستم می‌شود و انتخاب مناسب زیرمجموعه‌ای از زبان‌ها که بیشترین پوشش را برای فضای آوایی کلیه زبان‌های هدف داشته باشد، خود مسئله دیگری است. در این مقاله روش دیگری برای حل این مسئله ارائه شده است. در این روش از کلیه زبان‌ها برای ساخت فضای آوایی استفاده می‌شود اما در این فضای آوایی که هر جزء تشکیل‌دهنده آن را یک مؤلفه از مدل‌های GMM تشکیل می‌دهد، تنها آواها یا به عبارت دیگر مؤلفه‌هایی را انتخاب می‌کنیم

1. Probabilistic Sequence Kernel
2. Generalized Discernment Sequence Kernel
3. Gaussian Supervector
4. Expectation Maximization
5. PSK Adapted Weight



شکل ۱: معیارهای تشخیص زبان از نقطه نظر سطح انتزاع [۱۳].

که از قدرت تمایزکنندگی بالاتری در تشخیص زبان هدف از سایر زبان‌ها برخوردارند. از کنار هم قراردادن مؤلفه‌های انتخابی یک مدل GMM واحد (UBM)[6] حاصل می‌شود که تعداد مؤلفه‌های آن از مجموع تعداد مؤلفه‌های تک‌تک مدل‌ها کمتر است و به این ترتیب سبب کاهش پیچیدگی محاسباتی سیستم می‌شود. در صورتی که به‌جای مدل‌های GMM از مدل‌های PAW-GMM ارائه‌شده استفاده شود، مدل PAW - UBM حاصل می‌شود.

در بخش ۲ سیستم‌های تشخیص زبان گفتاری که اخیراً ارائه شده و مبنای تکنیک‌های ارائه‌شده در این مقاله است، شرح داده شده و سپس در بخش ۳ به بیان تکنیک‌های ارائه‌شده جهت بهبود دقت تشخیص زبان خواهیم پرداخت. نتایج آزمایش‌ها و ارزیابی‌های صورت‌گرفته نیز در بخش ۴ بیان گردیده و در خاتمه در بخش ۵ به نتیجه‌گیری پرداخته شده است.

۲- معرفی سیستم‌های تشخیص زبان گفتاری

۲-۱ مقدمه

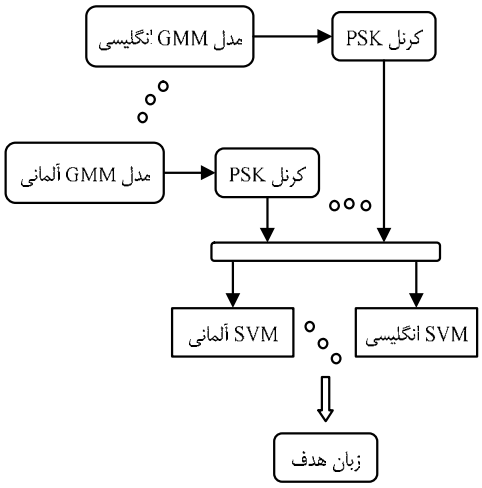
در تشخیص خودکار زبان‌ها، یافتن مشخصه‌های مؤثر برای جداسازی زبان‌ها از اهمیت بالایی برخوردار است [۱۳]. انسان و ماشین می‌توانند معیارهای مختلفی را برای تشخیص زبان مورد استفاده قرار دهند که عبارتند از:

- آواشناسی[7]: مجموعه آواها و فرکانس تکرار آنها از یک زبان به زبان دیگر متفاوت است. برای مثال یک آوا ممکن است در ۲ زبان وجود داشته باشد، اما در یکی از زبان‌ها فرکانس تکرار بیشتری نسبت به دیگری داشته باشد. همچنین واج‌آرایی[8] که قوانین ناظر بر توالی‌های مجاز آواها در یک زبان را مشخص می‌کند، از یک زبان به زبان دیگر متفاوت است.
- ریخت‌شناسی[9]: واژگان مورد استفاده، ریشه لغات، نحوه اشتقاق لغات و تولید لغات جدید از یک زبان به زبان دیگر متفاوت است.
- نحو[10]: الگوی جملات از یک زبان به زبان دیگر متفاوت است و ممکن است یک کلمه در ۲ زبان مشترک باشد (برای مثال کلمه bin در زبان آلمانی و انگلیسی)، اما مجموعه لغاتی که قبل و بعد از این لغات می‌آیند، ممکن است متفاوت باشند.

در شکل ۱ معیارهای تشخیص زبان از نقطه نظر سطح انتزاع نشان داده شده است. سیستم‌هایی که از سطوح پایین‌تر استفاده می‌کنند (ویژگی‌های اکوستیکی و پروزودی)، ساده‌تر بوده و نیازی به داده‌های

6. Universal Background Model
7. Phonology
8. Phonotactics
9. Morphology
10. Syntax

Archive of SID
www.SID.ir



شکل ۲: سیستم تشخیص زبان GMM - PSK - SVM.

آموزشی که در سطح واج برچسب‌گذاری شده باشند، ندارند. سیستم‌هایی که سطوح بالاتر را مورد استفاده قرار می‌دهند، پیچیده‌تر و قدرت تعمیم آنها به تعداد زبان‌های بیشتر، پایین‌تر است اما از دقت بالاتری برخوردارند.

۲-۲ سیستم تشخیص زبان بر مبنای مدل مخلوط گاوسی (GMM)

این روش یک روش ساده و آماری برای تشخیص زبان است و سعی بر جداسازی زبان‌ها بر مبنای تفاوت‌هایی که میان آواها و فرکانس تکرار آنهاست، دارد. GMM فرض می‌کند که هر بردار ویژگی V_t که مربوط به فریم زمانی t است، توسط یک تابع چگالی احتمالاتی تولید می‌شود. به عبارت دیگر بردارهای ویژگی (ویژگی MFCC) استخراج‌شده از فریم‌ها، دارای یک تابع توزیع احتمالاتی است که این تابع توزیع را می‌توان به‌صورت مجموع توابع توزیع نرمال چندمتغیره[1] به صورت (۱) نوشت [۲]

$$p\langle V_t|\lambda_t\rangle = \sum_{k=1}^{K} p_k b_k(V_t) \qquad (۱)$$

$$\lambda = \{p_k, \mu_k, \sum_k\} \qquad (۲)$$

که در آن λ مجموعه پارامترهای مدل است و k نشان‌دهنده شماره مخلوط، p_k وزن مخلوط با محدودیت $\sum_{k=1}^{N} p_k = 1$ و b_k تابع توزیع گاوسی است که با استفاده از دو پارامتر μ_k و $\sum_k$ مشخص می‌شود [۲].

به‌ازای هر زبان هدف یک مدل GMM با استفاده از الگوریتم EM و داده‌های آموزشی مربوط به آن زبان، آموزش داده می‌شود. در طول فاز تشخیص نیز ابتدا بردارهای ویژگی از سیگنال ورودی استخراج شده و سپس لگاریتم احتمال تعلق قطعه گفتاری به تک‌تک مدل‌های زبانی با استفاده از (۳) محاسبه می‌شود

$$L(\{x_t\}|\lambda_l) = \sum_{t=1}^{T} [\log\langle x_t|\lambda_l\rangle] \qquad (۳)$$

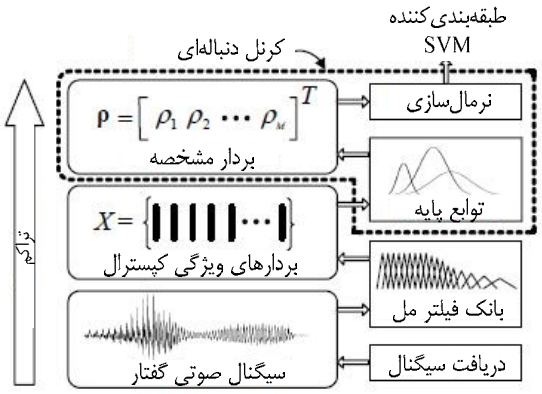
که در آن λ_l مدل GMM زبان l، $\{x_t\}$ توالی بردارهای ویژگی قطعه گفتاری ورودی و T طول زمانی آن است. در نهایت زبان مربوط به قطعه گفتاری با استفاده از تساوی (۴) تعیین می‌شود [۲]

$$l' = \arg\max_l L(\{x_t\}|\lambda_l) \qquad (۴)$$

۲-۳ سیستم تشخیص زبان GMM - PSK - SVM

همان‌طور که در شکل ۲ مشاهده می‌شود این سیستم از ۳ قسمت

1. Multivariate Gaussian Densities



شکل ۳: نمایش قطعات گفتاری برای کاربرد تأیید هویت گوینده (با حرکت از سمت پایین به بالا نمایش سیگنال فشرده‌تر است) [۶].

تشکیل شده است: مدل‌های GMM برای مدل‌کردن فضای آوایی، کرنل دنباله‌ای PSK برای تبدیل توالی بردارهای ویژگی استخراج‌شده از قطعه گفتار ورودی به برداری با بعد ثابت و طبقه‌بندی‌کننده‌های ماشین بردار پشتیبان[2] (SVM). به‌منظور استفاده از SVM برای تشخیص زبان، ۲ موضوع بایستی حل شود. ابتدا بایستی یک قالب کاری که بتوان از SVM به‌عنوان یک طبقه‌بندی‌کننده چندکلاسه استفاده نمود، فراهم شود. دوم این که SVM باید قادر به بررسی توالی از بردارهای ویژگی (به‌عنوان نمونه توالی از بردارهای ویژگی MFCC مربوط به یک قطعه گفتاری) به‌منظور عمل طبقه‌بندی باشد. موضوع اول با استفاده از استراتژی یکی در مقابل همه[3] حل می‌شود [۱۴]. در این استراتژی به‌ازای هر زبان یک مدل SVM با در نظر گرفتن نمونه‌های آموزشی مربوط به آن زبان به‌عنوان نمونه‌های مثبت (نمونه‌های با برچسب ۱+) و نمونه‌های مربوط به سایر زبان‌ها به‌عنوان نمونه‌های منفی (نمونه‌های با برچسب ۱−)، آموزش داده می‌شود. راه حل موضوع دوم نیز استفاده از کرنل دنباله‌ای است. کرنل‌های دنباله‌ای متعددی ارائه شده است که از میان این کرنل‌ها، کرنل PSK کارایی بالایی در کاربرد تشخیص زبان از خود نشان داده است [۶]. نحوه ساخت این کرنل در قسمت بعد توضیح داده شده است.

۲-۴ کرنل دنباله‌ای

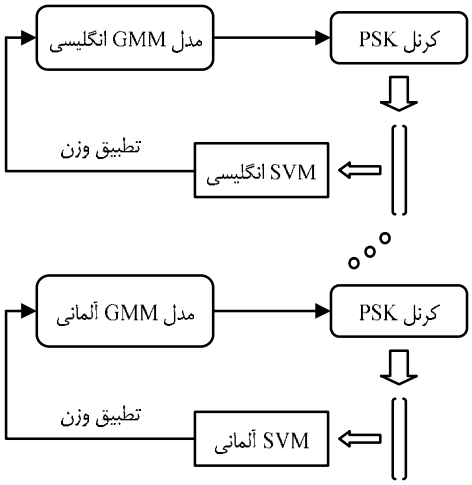
با استفاده از کرنل‌های دنباله‌ای می‌توان قطعات گفتاری با طول متغیر را به بردارهای ویژگی با ابعاد ثابت تبدیل کرد. کرنل‌های دنباله‌ای مختلفی وجود دارند. این کرنل‌ها در نحوه توسعه ویژگی‌ها (انتخاب توابع پایه)، تبدیل ویژگی و نرمال‌سازی ویژگی‌ها با یکدیگر متفاوتند. از کرنل‌های دنباله‌ای می‌توان کرنل‌های GLDS، GMM Supervector، PSK و Fisher را نام برد [۶]. سیگنال گفتار را می‌توان بسته به کاربرد به صورت‌های مختلفی نشان داد. برای مثال برای کاربرد تعیین هویت گوینده در سطح اول سیگنال گفتار اکوستیکی را داریم که نشان‌دهنده تغییراتی است که در سطح فشار صدای تولیدشده در اثر حرکات مجرای گفتار به‌وجود می‌آید. در سطح بعدی بردارهای ویژگی کپسترال هستند که یک نمایش فشرده از طیف زمان کوتاه که وابسته به تغییرات تارهای صوتی در طول زمان است را ارائه می‌دهند. بنابراین ویژگی‌های کپسترال اطلاعاتی را راجع به گوینده و همچنین اطلاعات زبانی در بردارند.

در کاربرد تأیید هویت گوینده ما تنها به مؤلفه اطلاعات گوینده نیاز داریم. بنابراین می‌خواهیم اثر سایر اطلاعات تأثیرگذار موجود در ویژگی

2. Support Vector Machine
3. One-vs-All

Archive of SID

www.SID.ir



شکل ۴: تکنیک تطبیق وزن ارائه‌شده.

کپسترال را خنثی کنیم. همان‌طور که در شکل ۳ نشان داده شده است، از یک سری توابع پایه برای آنالیز و خلاصه‌سازی اطلاعات خاص گوینده به یک بردار ρ که حاوی اطلاعات گوینده است، استفاده شده است. نشان‌دادن قطعات گفتاری با بردارهای با ابعاد ثابت، مدل‌سازی گوینده را نسبت به زمانی که قطعات گفتاری را با دنباله‌ای با طول متغیر از بردارها نشان دهیم، ساده‌تر می‌سازد [۶]. در مورد تشخیص زبان نیز به همین صورت است و این کار سبب ساده‌سازی تشخیص زبان شده و آموزش بهتر صورت می‌گیرد. همچنین هدف در تشخیص زبان بر خلاف تأیید هویت گوینده، پررنگ کردن اطلاعات زبانی است.

در صورتی که $\varphi_j(x)\ ,\ j=1,2,\ldots,M$ مجموعه‌ای از M تابع پایه باشد، توالی $X=\{x_1,x_2,\ldots,x_N\}$ از N بردار ویژگی کپسترال با ابعاد D با استفاده از مراحل زیر به بردار مشخصه ρ تبدیل می‌شود [۶]:

۱) توسعه ویژگی

$$x \to \varphi(x)=[\varphi_1(x),\varphi_2(x),\ldots,\varphi_M(x)]^T$$

۲) آنالیز آماری

$$\rho = f\{\varphi(x); x\in X\}$$

که در آن علامت T نشان‌دهنده ترانهاده ماتریس و $f\{.\}$ تابعی از بردارهای ویژگی توسعه‌یافته $\varphi(x)$ (به‌عنوان نمونه میانگین ریاضی) است. با استفاده از این ۲ مرحله، تعداد ضرایب ویژگی از $N\times D$ ضریب X به M ضریب ρ تبدیل می‌شود. در صورتی که $M \gg D$ باشد، احتمال این که مسئله به‌صورت خطی حل شود، بیشتر می‌شود.

در این صورت می‌توان از SVM با کرنل خطی برای دسته‌بندی بردارهای ویژگی به‌دست آمده استفاده نمود [۶].

حال به توضیح کرنل دنباله‌ای PSK می‌پردازیم. در کرنل PSK عمل توسعه به‌صورت زیر انجام می‌شود [۶]

$$\varphi_{PSK}(x)=[p(j=1|x),p(j=2|x),\ldots,p(j=M|x)]^T \quad (۵)$$

که در آن $p(j=1|x)$ نشان‌دهنده احتمال ثانویه j امین مخلوط گاوسی است و از (۶) محاسبه می‌شود. هر کدام از مؤلفه‌های این بردار توسعه احتمال رخداد j امین کلاس صوتی را در صورت مشاهده بردار x نشان می‌دهد. بردار شاخص ρ با استفاده از (۷) محاسبه می‌شود

$$p(j|x)=\frac{p(x|j)p(j)}{\sum_{i=1}^{M}p(x|i)p(i)} \quad (۶)$$

$$\rho_{PSK}(x)=\frac{1}{N}\sum_{x\in X}\varphi_{PSK}(x) \quad (۷)$$

می‌توان ρ_{PSK} را به‌صورت یک هیستوگرام M سلولی تعبیر کرد که هر سلول، احتمال رخداد هر مؤلفه از مدل‌های GMM که مدل‌کننده یک یا بخشی از یک آواست را در قطعه گفتاری X نشان می‌دهد [۶].

۳- مدل‌های ارائه‌شده

۳-۱ مدل PAW - GMM

در آموزش مدل GMM، الگوریتم EM به‌ازای هر مؤلفه، یک بردار میانگین، واریانس و وزن را محاسبه می‌کند. وزن هر مؤلفه بر اساس احتمال رخداد آن مؤلفه به‌ازای داده‌های آموزشی تعیین می‌شود، به‌عبارت دیگر وزن هر مؤلفه، متناسب با فرکانس تکرار آن مؤلفه در فضای آواهای آن زبان است و آموزش درست این وزن‌ها تأثیر زیادی در دقت تشخیص زبان دارد. ایرادی که در روش مدل‌کردن GMM وجود دارد این است که هر زبان مستقل از زبان دیگر مدل می‌شود. به‌عنوان نمونه ممکن است یک یا بخشی از یک آوا فرکانس تکرار کمی در یک زبان داشته باشد و در مقابل آوای دیگر فرکانس تکرار زیادتری داشته باشد اما تأثیر آوای اول در تشخیص این زبان از سایر زبان‌ها بیشتر از آوای دوم باشد. بنابراین یک راه برای افزایش دقت مدل‌های GMM می‌تواند تغییر مکانیزم وزن‌دهی مؤلفه‌ها بر اساس اطلاعات تمایزدهندگی آنها باشد. به این منظور در مقاله قبل همین نویسندگان [۱۵]، روشی ارائه شد که تطبیق وزن مؤلفه‌های GMM در قالب سیستم تشخیص زبان GMM-VSM صورت گرفت. از آنجایی که سیستم GMM-PSK-SVM دقت بالاتری در تشخیص زبان دارد، با استفاده از این سیستم مدل PAW-GMM ارائه شده است. نحوه آموزش این مدل‌ها در شکل ۴ قابل مشاهده است.

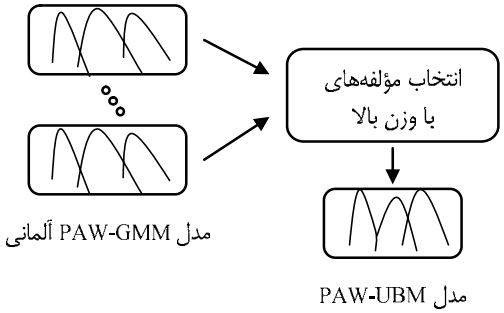
در این سیستم در فاز آموزش، پس از استخراج بردارهای ویژگی از داده‌های آموزشی، توالی بردارهای ویژگی مربوط به هر قطعه گفتاری، با استفاده از کرنل PSK به برداری با طول ثابت (برابر با تعداد مؤلفه‌های مدل گاوسی) تبدیل می‌شود. سپس این بردارهای ویژگی برای آموزش مدل‌های SVM با کرنل خطی مورد استفاده قرار می‌گیرند که به‌ازای هر شناساگر GMM یک مدل SVM با استفاده از بردارهای ویژگی مربوط به آن شناساگر و با در نظر گرفتن بردارهای متعلق به زبان آن شناساگر به‌عنوان نمونه‌های مثبت (نمونه‌های با برچسب ۱+) و بردارهای ویژگی متعلق به سایر زبان‌ها به‌عنوان نمونه‌های منفی (نمونه‌های با برچسب ۱−)، با استفاده از تمامی داده‌های آموزشی، آموزش داده می‌شود.

پس از آموزش، از هر SVM یک بردار وزن به‌دست می‌آید که از (۸) محاسبه می‌شود

$$w=\sum_{i=1}^{N}\alpha_i b(x_i)+d \quad (۸)$$

که در آن $b(x_i)$ بردارهای پشتیبان، α_i ضرایب مربوط به بردارهای پشتیبان و N تعداد بردارهای پشتیبان است. هر وزن $|wi|$ نشان‌دهنده سهم هر بعد از بردارهای ویژگی در ساخت ابرصفحه جداکننده حاصل از مدل SVM است. به عبارت دیگر هر چقدر $|wi|$ بیشتر باشد، بعد iام از بردارهای ویژگی تأثیر بیشتری در متمایزکردن زبان هدف از سایر زبان‌ها دارد [۱۶]. به این ترتیب می‌توان از بردارهای وزن حاصل از مدل‌های SVM به‌عنوان وزن‌های جدید مؤلفه‌های GMM استفاده نمود. از آنجایی که این وزن‌ها بر اساس قدرت تمایزکنندگی هر مؤلفه در تشخیص یک زبان از سایر زبان‌ها تعیین شده‌اند، انتظار می‌رود که دقت سیستم‌های

Archive of SID
www.SID.ir



شکل ۵: نحوه ساخت مدل PAW - UBM.

تشخیص زبانی که از مدل GMM با وزن‌های تطبیق‌یافته استفاده می‌کنند، افزایش یابد. در [۱۷] نیز روشی برای متمایزکردن مدل‌های GMM با استفاده از SVM ارائه شده است که با روش ارائه‌شده متفاوت است. در این روش بردارهای میانگین و واریانس مدل‌های GMM به‌عنوان ورودی به مدل SVM داده شده است و از SVM برای ایجاد مدل‌های GMM با بردارهای میانگین و واریانس جدید استفاده شده است.

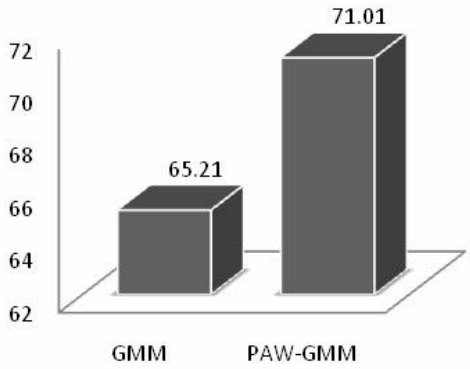
۳-۲ مدل PAW - UBM

همان‌طور که قبلاً به آن اشاره کردیم، در سیستم تشخیص زبان GMM - PSK - SVM که در آن از مجموعه‌ای از مدل‌های GMM برای مدل‌کردن فضای آوایی استفاده می‌شود، ابعاد بردارهای ویژگی خصوصاً زمانی که از تعداد بیشتری مدل GMM برای افزایش دقت سیستم استفاده شود، بالاست که سبب افزایش هزینه محاسباتی می‌شود. یک راه حل می‌تواند بر این اساس باشد که همه فضای آوایی متعلق به یک زبان در تشخیص آن زبان از سایر زبان‌ها مؤثر نیست از این رو می‌توان با تنها در نظر گرفتن مؤلفه‌هایی از هر مدل GMM که قدرت تمایزکنندگی بالایی دارند، بعد بردارهای ویژگی را کم کرد و از کنار هم قرار دادن مؤلفه‌های انتخابی از هر مدل GMM، یک مدل GMM واحد (مدل UBM) را ایجاد نمود. قدرت تمایزکنندگی هر مؤلفه را می‌توان بر اساس بردار وزنی حاصل از هر طبقه‌بندی‌کننده SVM مشخص نمود، به‌عبارت دیگر از PAW - GMM جهت ساخت مدل UBM که PAW - UBM نام‌گذاری شده است، استفاده نمود. نحوه ساخت این مدل در شکل ۵ نشان داده شده است.

۴- آزمایش‌ها

تمامی آزمایش‌ها در این مقاله با استفاده از دادگان تلفنی چندزبانه OGI [۱۸] صورت گرفته است. این دادگان شامل ۱۰ زبان انگلیسی، فارسی، فرانسوی، آلمانی، ژاپنی، کره‌ای، اسپانیایی، ماندارین، تامیل و ویتنامی است. از بین این زبان‌ها، ۴ زبان انگلیسی، فرانسه، فارسی و آلمانی جهت آموزش و ارزیابی سیستم‌ها انتخاب شده و آموزش سیستم‌ها با استفاده ازکلیه قطعات موجود در قسمت آموزش این دادگان و ارزیابی با استفاده از قطعات ۴۵ثانیه‌ای قسمت ارزیابی صورت گرفت.

ابتدا قطعات گفتاری به فریم‌های ۳۰ میلی‌ثانیه با ۱۰ میلی‌ثانیه هم‌پوشانی تبدیل و بخش‌های سکوت از آنها حذف گردید. سپس از هر فریم بردار ویژگی MFCC شامل ۱۲ ضریب اول و انرژی به همراه مشتق اول و دوم این ضرایب استخراج شد. پس از استخراج بردارهای ویژگی، از CMS جهت نرمال‌سازی و حذف اثرات کانال استفاده شد. در مرحله بعد، ۴ مدل GMM (مدل GMM انگلیسی، فارسی، فرانسوی و



شکل ۶: تأثیر استفاده از مدل PAW - GMM بر دقت تشخیص زبان.

آلمانی) با ۲۵۶ مؤلفه با استفاده از الگوریتم EM و ابزار HTK[1] و داده‌های آموزشی زبان هدف، آموزش داده شد. در مورد هر مدل، الگوریتم EM، ۱۰ مرتبه تکرار شد. پس از آموزش، این مدل‌ها به‌عنوان مدل‌های آوایی در سیستم GMM - PSK - SVM (شکل ۳) مورد استفاده قرار گرفتند و توالی بردارهای ویژگی مربوط به قطعه گفتار ورودی، پس از اعمال کرنل PSK به بردارهای با طول ثابت ۲۵۶ بعد نگاشت داده شد. پس از تبدیل تمامی قطعات گفتاری آموزشی به بردارهای با طول ثابت (۲۵۶ بعد)، این بردارها جهت آموزش طبقه‌بندی‌کننده‌های SVM با کرنل خطی مورد استفاده قرار گرفتند. به این ترتیب که به‌ازای هر مدل GMM، یک طبقه‌بندی‌کننده SVM با استفاده از بردارهای ویژگی حاصل از این شناساگر و در نظر گرفتن بردارهای ویژگی مربوط به زبان مدل GMM به‌عنوان نمونه‌های با برچسب ۱+ و بردارهای ویژگی مربوط به سایر زبان‌ها به‌عنوان بردارهای ویژگی با برچسب ۱− آموزش داده شد. از هر مدل SVM پس از آموزش، یک بردار وزن حاصل می‌شود که هر بعد آن نشان‌دهنده تأثیر مؤلفه گاوسی متناظر در جداکردن زبان مربوطه از سایر زبان‌هاست. پس از آموزش سیستم، ۴ بردار وزن با بعد ۲۵۶ حاصل شد که این بردارها به‌عنوان وزن‌های جدید مدل‌های GMM در نظر گرفته شدند. به این ترتیب ۴ مدل PAW - GMM حاصل شد.

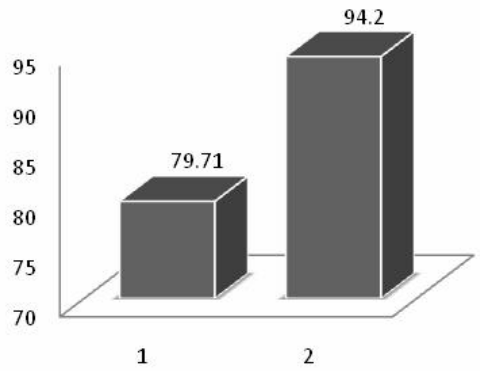
در آزمایش اول تأثیر استفاده از مدل‌های GMM ارائه شده (PAW - GMM)، به‌عنوان جایگزین مدل‌های GMM بر دقت تشخیص زبان سیستمی که در قسمت ۲-۲ توصیف شد، مورد بررسی قرار گرفته است. در این سیستم از ۴ مدل PAW - GMM (به‌ازای هر زبان یک مدل) برای ساخت سیستم استفاده شده و زبان هدف بر اساس ماکسیمم‌گیری روی مقادیر احتمالی خروجی حاصل از هر مدل به‌ازای قطعه گفتار ورودی تعیین شد. دقت این سیستم تشخیص زبان در حالتی که از مدل‌های GMM برای مدل‌کردن فضای آوایی استفاده شود و حالتی که از مدل‌های PAW-GMM ارائه شده استفاده شود، در شکل ۶ نشان داده شده است. همان‌طور که در این شکل مشاهده می‌شود، دقت سیستم تشخیص زبان در حالتی که از مدل‌های PAW - GMM ارائه شده استفاده شود، افزایش چشم‌گیری پیدا می‌کند (افزایش از ۶۵/۲۱ درصد به ۷۱/۰۱ درصد).

در آزمایش دوم تأثیر استفاده از مدل‌های ارائه‌شده بر دقت سیستم تشخیص زبان GMM - PSK - SVM مورد بررسی قرار گرفته است. به این منظور تأثیر این مدل در قالب ۲ سیستم تشخیص زبان که همگی بر مبنای GMM - PSK - SVM شنوندگان انسانی را در حالات مختلف مدل می‌کنند، مورد مطالعه قرار گرفته است. سیستم اول مدل‌کننده

1. Available at: http://htk.eng.cam.ac.uk/

Archive of SID

www.SID.ir



شکل ۷: دقت تشخیص زبان سیستم GMM - PSK - SVM در دو حالت مدل‌کننده شنونده تک‌زبانه و مسلط بر چند زبان.

جدول ۱: تأثیر استفاده از مدل PAW - GMM بر دقت تشخیص زبان.

PAW - GMM استفاده از وزن‌های حاصل از سیستم GMM - PSK - SVM (شنونده چندزبانه)	PAW - GMM استفاده از وزن‌های حاصل از سیستم GMM - PSK - SVM (شنونده تک‌زبانه)	GMM
٪۷۸٫۲۶	٪۷۱٫۰۱	٪۶۵٫۲۱

جدول ۲: تأثیر استفاده از مدل PAW - GMM بر دقت تشخیص زبان سیستم GMM - PSK - SVM.

سیستم PAW-GMM-PSK-SVM شنونده مسلط بر ۴ زبان	سیستم GMM-PSK-SVM شنونده مسلط بر ۴ زبان	سیستم PAW-GMM-PSK-SVM شنونده تک‌زبانه	سیستم GMM-PSK-SVM شنونده تک‌زبانه
٪۹۴٫۲۰	٪۹۲٫۷۵	٪۷۹٫۷۱	٪۷۸٫۲۶

SVM شنونده انسانی است که تنها بر یک زبان مسلط است و سیستم دوم مدل‌کننده شنونده انسانی است که بر هر ۴ زبان مسلط است. تفاوت این ۲ سیستم در نحوه شکل‌دادن بردارهای ویژگی برای آموزش مدل‌های است. در سیستم اول از ۴ مدل GMM (انگلیسی، فارسی، فرانسوی و آلمانی) استفاده شده است و بردار حاصل از هر مدل GMM برای آموزش مدل SVM مربوط به آن زبان، مورد استفاده قرار می‌گیرد.

سیستم دوم مانند سیستم اول از ۴ مدل GMM استفاده می‌کند، اما هر مدل SVM با استفاده از بردارهای ویژگی با بعد (۴×۲۵۶ = ۱۰۲۴) که از کنار هم قرار دادن بردارهای حاصل از تبدیل‌یافته خروجی همه مدل‌ها حاصل می‌شوند، آموزش داده می‌شود. دقت تشخیص زبان این ۲ سیستم در حالاتی که از مدل‌های GMM و مدل‌های PAW - GMM استفاده شود، در شکل ۷ نشان داده شده است.

همان‌طور که در نمودار شکل ۷ قابل مشاهده است، دقت سیستم تشخیص زبان دوم که مدل‌کننده شنونده انسانی است که بر هر ۴ زبان مسلط است، بالاتر از سیستم اول است که این نتیجه با نتایج تجربی نیز تطبیق دارد. چرا که آزمایش‌های صورت‌گرفته بر روی شنوندگان انسانی نشان داده است که افرادی که چندزبانه هستند در تفکیک زبان‌ها نسبت به افرادی که تنها یک زبان را می‌دانند بهتر عمل می‌کنند.

در سیستم تشخیص زبان اول هر مدل SVM به‌عنوان نمونه مدل SVM فارسی، مانند شنونده فارسی زبانی است که تنها زبان فارسی را می‌داند و از هیچ زبان دیگری اطلاعی ندارد. این شنونده هر سیگنال گفتاری را که می‌شنود با توجه به فضای آوایی که از زبان فارسی (مدل GMM زبان فارسی) یاد گرفته است، عمل تشخیص زبان فارسی از سایر زبان‌ها را انجام می‌دهد اما در سیستم تشخیص زبان دوم، هر مدل SVM مانند یک شنونده انسانی چندزبانه عمل می‌کند چرا که از اطلاعات آوایی مربوط به هر ۴ زبان انگلیسی، فارسی، فرانسه و آلمانی (۴ مدل آوایی GMM) بهره می‌گیرد و عمل تشخیص را انجام می‌دهد.

در آزمایش سوم تأثیر استفاده از مدل PAW - GMM بر دقت سیستم‌های اول و دوم مورد بررسی قرار گرفته است. در سیستم اول پس از آموزش از آنجایی که بعد بردارهای ویژگی ورودی مربوط به هر SVM برابر ۲۵۶ است، هر SVM بردار وزنی با بعد ۲۵۶ را به‌عنوان خروجی مرحله آموزش برمی‌گرداند که این بردارهای وزنی به‌جای وزن‌هایی که از الگوریتم EM حاصل شده است، در GMM متناظر مورد استفاده قرار می‌گیرد و به این ترتیب مدل‌های PAW - GMM حاصل می‌شود. نحوه استفاده از وزن‌های حاصل از SVM در مورد سیستم دوم کمی متفاوت است. چرا که ابعاد بردارهای ورودی SVM برابر ۱۰۲۴ است، به همین جهت از هر SVM بردارهای وزنی با بعد ۱۰۲۴ به‌دست می‌آید. در این حالت ۲۵۶ بعد اول از بردار وزنی حاصل از SVM مربوط به زبان انگلیسی به‌عنوان وزن‌های جدید مدل GMM انگلیسی، ۲۵۶ بعد دوم از بردار وزنی حاصل از SVM مربوط به زبان فارسی به‌عنوان وزن‌های جدید مدل GMM فارسی و به همین ترتیب وزن‌های مربوط به GMM های فرانسوی و آلمانی تطبیق داده می‌شود و به این ترتیب مدل‌های PAW - GMM با تطبیق وزن مدل‌های GMM حاصل می‌شوند. نتایج حاصل در ۲ حالت مورد بررسی قرار گرفته است:

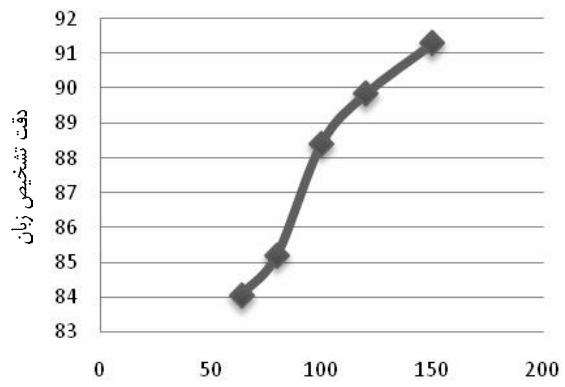
- تأثیر مدل‌های PAW - GMM بر دقت تشخیص زبان سیستمی که از GMM مستقیماً جهت تشخیص زبان استفاده می‌نماید (سیستمی که در قسمت ۲-۲ توصیف شد).
- تأثیر مدل‌های PAW - GMM بر دقت تشخیص زبان سیستمی که از GMM به‌صورت غیر مستقیم استفاده می‌نماید (سیستم‌های GMM - PSK - SVM اول و دوم).

نتایج حاصل در جدول ۱ آمده است. همان‌طور که مشاهده می‌شود در حالتی که از مدل GMM مستقیماً جهت تشخیص زبان استفاده شود، استفاده از مدل PAW - GMM به‌عنوان مدل جایگزین سبب افزایش دقت تشخیص زبان می‌شود. این افزایش هم در حالتی که مدل‌های PAW - GMM از تطبیق وزن در قالب سیستم GMM - PSK - SVM مدل‌کننده شنونده تک‌زبانه حاصل شود و هم در حالتی که از سیستم مدل‌کننده شنونده مسلط بر ۴ زبان حاصل شود، قابل مشاهده است. اما در حالت مدل‌کننده شنونده مسلط بر ۴ زبان این افزایش بیشتر است.

نتایج حاصل از استفاده از مدل PAW-GMM به‌عنوان مدل جایگزین GMM بر دقت تشخیص زبان سیستمی که از GMM به‌صورت غیر مستقیم استفاده می‌نماید (سیستم‌های GMM - PSK - SVM اول و دوم) در جدول ۲ آورده شده است. همان‌طور که مشاهده می‌شود در مورد هر ۲ سیستم، استفاده از PAW - GMM به‌عنوان مدل جایگزین، سبب افزایش دقت تشخیص زبان شده است.

Archive of SID

www.SID.ir



شکل ۸: دقت سیستم PAW - UBM - PSK - SVM به‌ازای تعداد مؤلفه‌های انتخابی مختلف از هر مدل GMM.

در آزمایش چهارم دقت تشخیص زبان مدل UBM ارائه‌شده مورد ارزیابی قرار گرفته است. پس از آموزش سیستم GMM - PSK - SVM در حالت مدل‌شنونده مسلط بر ۴ زبان، بر اساس بردارهای وزن حاصل از هر SVM، مؤلفه‌هایی که قدرت تمایز بالاتری در هر مدل GMM دارند، مشخص شد و از کنار هم قرار گرفتن آنها یک مدل UBM ساخته شد. سپس سیستم GMM - PSK - SVMای آموزش داده شد که از این مدل به‌عنوان مدل‌کننده فضای آوایی زبان‌ها استفاده می‌کند و دقت آن با استفاده از داده‌های ارزیابی اندازه‌گیری شد. نتایج حاصل در نمودار شکل ۸ به‌ازای تعداد مؤلفه‌های انتخابی مختلف از هر مدل GMM، نشان داده شده است. همان‌طور که مشاهده می‌شود در حالتی که از هر مدل GMM، ۶۴ مؤلفه انتخاب شود (تعداد مؤلفه‌های UBM برابر ۲۵۶ باشد)، دقت تشخیص زبان برابر ۸۴٫۰۶ درصد است که نسبت به سیستمی که مدل‌کننده شنونده تک‌زبانه است، دقت بالاتر است (افزایش از ۷۸٫۲۶ درصد به ۸۴٫۰۶ درصد). در صورتی که تعداد مؤلفه‌های انتخابی از هر مدل را افزایش دهیم، دقت بالاتر می‌رود و در حالتی که ۱۵۰ مؤلفه از هر مدل GMM انتخاب شود، دقت از ۸۴٫۰۶ درصد به ۹۱٫۳۰ درصد افزایش پیدا می‌کند.

برای بهبود بیشتر دقت سیستم تشخیص زبان، با ثابت نگه داشتن تعداد مؤلفه‌های مدل UBM بر روی ۶۰۰ از تکنیک تطبیق وزن ارائه‌شده استفاده شده است. به این صورت که پس از آموزش سیستم PAW - UBM - PSK - SVM از وزن‌های حاصل از مدل‌های SVM به‌عنوان وزن‌های جدید مدل UBM استفاده شده است. دقت سیستم با استفاده از این تکنیک افزایش یافته و به دقت سیستم تشخیص زبان GMM - PSK - SVM در حالتی که مدل‌کننده شنونده مسلط بر ۴ زبان است، رسید. در حالی که ابعاد بردارهای ویژگی از ۱۰۲۴ به ۶۰۰ کاهش یافته است.

برای مقایسه سیستم PAW - UBM - PSK - SVM و سیستم PAW - GMM - PSK - SVM از نقطه نظر سرعت تشخیص زبان مربوط به یک قطعه گفتاری از معیار xRT استفاده شده است. این معیار توسط NIST برای مقایسه سرعت سیستم‌ها مورد استفاده قرار می‌گیرد و از تقسیم متوسط زمان شناسایی قطعه گفتاری (TPT)[1] بر طول قطعه (SSD)[2] حاصل می‌شود. هرچه این معیار کوچک‌تر باشد، سیستم از سرعت بالاتری برخوردار است. مقادیر این معیار برای هر یک از این

1. Total Processing Time
2. Source Signal Duration

جدول ۳: مقایسه سرعت سیستم‌ها بر اساس معیار xRT.

PAW - UBM - PSK - SVM	PAW - GMM - PSK - SVM
۰٫۳۴۲	۰٫۴۴۶

سیستم‌ها در جدول ۳ نشان داده شده است. همان‌طور که مشاهده می‌شود، مقدار مربوط به سیستم PAW - UBM - PSK - SVM کمتر است که نشان‌دهنده سرعت بالاتر این سیستم است.

۵- جمع‌بندی و نتیجه‌گیری

در این مقاله با در نظر گرفتن مشکلی که در روش مدل‌کردن به‌وسیله مدل GMM در کاربرد تشخیص زبان وجود داشت، مدلی جدید با نام PAW - GMM ارائه گردید. در این مدل وزن مؤلفه‌های GMM بر اساس قدرت آنها در تمایز یک زبان از سایر زبان‌ها تعیین می‌شود. ارزیابی‌های صورت‌گرفته با استفاده از داده‌های ۴ زبان انگلیسی، فارسی، فرانسوی و آلمانی از دادگان OGI نشان داد که استفاده از مدل ارائه‌شده به‌عنوان جایگزین مدل GMM در سیستم تشخیص زبان، سبب افزایش دقت تشخیص زبان می‌شود. این افزایش هم در سیستم‌هایی که از مدل GMM مستقیماً و هم در سیستم‌هایی که از GMM به‌صورت غیر مستقیم استفاده می‌کنند، مشاهده گردید. در بهترین حالت دقت سیستم نوع اول از ۶۵٫۲۱ درصد به ۷۸٫۲۶ درصد افزایش و دقت سیستم نوع دوم از ۹۲٫۷۵ به ۹۴٫۲۰ درصد افزایش یافت. همچنین با توجه به پیچیدگی محاسباتی بالایی که در سیستم GMM-PSK-SVM در حالت مدل‌کننده چندزبانه وجود دارد، سیستم PAW - UBM - PSK - SVM ارائه شد. این سیستم با ثابت نگه داشتن دقت تشخیص زبان، سبب کاهش پیچیدگی محاسباتی گردید.

مراجع

[1] A. Ziaei, S. M. Ahadi, S. M. Mirrezaie, and H. Yeganeh, "Spoken language identification using a new sequence kernel - based SVM back - end classifier," in *Proc. IEEE Int. Symp. on Signal Processing and Information Technology*, pp. 324-329, Dec. 2008.

[2] M. A. Zissman, "Comparision of four approaches to automatic language identification of telephone speech," *IEEE Trans. on Speech and Audio Processing*, vol. 4, no. 1, pp. 31-44, Jan. 1996

[3] H. Li, B. Ma, and C. H. Lee, "A vector space modeling approach to spoken language identification," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 1, pp. 271-284, Jan. 2007.

[4] R. Tong, B. Ma, H. Li, and E. S. Chng, "A target oriented phonotactic front-end for spoken language recognition," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 17, no. 7, pp. 1335-1347, Sep. 2009.

[5] P. A. Torres-Carrasquillo, E. Singer, M. A. Kohler, R. J. Greene, D. A. Reynolds, and J. A. Deller, "Approaches to language identification using gaussian mixture models and shifted delta cepstral features," in *Proc. ICSLP*, pp. 89-92, 2002.

[6] K. -A. Lee, C. You, H. Li, T. Kinnunen, and D. Zhu, "Characterizing speech utterances for speaker verification with sequence kernel SVM," in *Proc. InterSpeech 2008*, pp. 1397-1400, 2008.

[7] W. Campbell, J. Campbell, D. Reynolds, E. Singer, and P. Torres - Carrasquillo, "Support vector machines for speaker and language recognition," *Comput. Speech Lang.*, vol. 20, no. 2-3, pp. 210-229, Jul.. 2006.

[8] W. Campbell, D. Sturim, and D. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308-311, May 2006.

[9] Z. Karam and W. Campbell, "A new kernel for SVM MLLR based speaker recognition," in *Proc. InterSpeech 2007*, pp. 290-293, Antwerp, Belgium, Aug. 2007.

[10] N. Dehak, P. A. Torres, D. Reynolds, and R. Dehak, "Language recognition via i-vectors and dimensionality reduction," in *Proc. InterSpeech 2011*, pp. 857-860, 2011.

Archive of SID

www.SID.ir

[11] D. Matrouf, F. Verdet, M. Rouvier, J. F. Bonastre, and G. Linarès, "Modeling nuisance variabilities with factor analysis for GMM - based audio pattern classification," *Computer Speech and Language*, vol. 25, no. 3, pp. 481-498, Jul. 2011.
[12] J. Farinas and F. Pellegrino, "Automatic rhythm modeling for language identification," in *Proc. Eurospeech*, vol. 4, pp. 2539-2542, 2001.
[13] R. Tong, M. Bin, D. Zhu, H. Li, and E. S. Chng, "Integrating acoustic, prosodic and phonotactic features for spoken language identification," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, vol. 1, pp. 205-208, 14-19 May 2006.
[14] W. M. Campbell, E. Singer, P. A. Torres - Carrasquillo, and D. A. Reynolds, "Language recognition with support vector machines," in *Proc. Odyssey: the Speaker and Language Recognition Workshop*, pp. 41-44, 2004.
[15] F. Ghasemian and M. M. Homayounpour, "Towards better GMM-based acoustic modeling for spoken language identification," *in Proc. ICEE*, 4 pp., 2011.
[16] R. Tong, B. Ma, H. Li, and E. S. Chng, "Target-oriented phone tokenizers for spoken language recognition," in *Proc. InterSpeech 2009*, pp. 200-203, Apr. 2009.
[17] W. M. Campbell, "A covariance kernel for SVM language recognition," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, pp. 4141-4144, Apr. 2008.
[18] Y. K. Muthusamy, R. A. Cole, and B. T. Oshika, "The OGI multi - language telephone speech corpus," in *Proc. ICSLP 92*, pp. 895-898, 1992.

فهیمه قاسمیان تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد به ترتیب در رشته‌های مهندسی کامپیوتر (نرم افزار) و مهندسی کامپیوتر (هوش مصنوعی) در سال‌های ۱۳۸۷ و ۱۳۸۹ از دانشگاه صنعتی امیرکبیر به پایان رسانده و در حال حاضر دانشجوی دکترای نرم افزار دانشگاه اصفهان است. زمینه‌های مورد علاقه ایشان عبارتند از پردازش گفتار، پردازش متن، وب معنایی، مهندسی نرم افزار مدل رانده.

محمد مهدی همایون‌پور تحصیلات خود در مقطع کارشناسی در رشته مهندسی برق (الکترونیک) در دانشگاه صنعتی امیرکبیر (سال ۱۳۶۶)، کارشناسی ارشد را در رشته برق (مخابرات)، از دانشگاه خواجه نصیرالدین طوسی (سال ۱۳۶۹)، کارشناسی ارشد دوم خود را در زمینه فونتیک (۱۳۷۴) در دانشگاه سوربون جدید در فرانسه و هم‌زمان دورهٔ دکترای خود را در دانشگاه پاریس ۱۱ در زمینه مهندسی برق (۱۳۷۴) به‌پایان رسانید. نام‌برده از سال ۱۳۷۴ در سمت عضو هیأت علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر به تدریس و تحقیق مشغول می‌باشد. زمینه‌های تخصصی مورد علاقه ایشان شامل پردازش سیگنال‌های دیجیتال، بازشناسی گفتار و گوینده, تبدیل متن به گفتار, کدینگ گفتار، پردازش زبان طبیعی, تشخیص نفوذ در سیستم‌ها و شبکه‌های کامپیوتری، اتوماسیون صنعتی، چند رسانه‌ای و طراحی سخت افزار می‌باشد.

Archive of SID

www.SID.ir