بهبود روش MLLR با استفاده از MCE و دانش‌آوائی برای تطبیق گوینده

رضا صحرائیان[۱]،بهزاد زمانی[۲]، احمد اکبری[۳]، احمد آیت‌اللهی[۴]

[۱] دانشکده مهندسی برق، دانشگاه علم و صنعت ایران، تهران،
Rsahraian@ee.iust.ac.ir
[۲] دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران،
Bzamani@iust.ac.ir
[۲] دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران،
Akbari@iust.ac.ir
[۱] دانشکده مهندسی برق، دانشگاه علم و صنعت ایران، تهران،
Ayatollahi@iust.ac.ir

چکیده

روش رگرسیون خطی با بیشینه درست‌نمایی (MLLR) یکی از روش‌های تطبیق گوینده است که با خوشه‌بندی پارامترهای مدل و تخمین ماتریس تبدیل برای هر خوشه و اعمال تبدیل بر پارامترهای مدل ، باعث بهبود عملکرد تطبیق و در نتیجه افزایش نرخ بازشناسی برای گوینده جدید می‌شود. خوشه‌بندی پارامترهای مدل معمولاً به دو صورت انجام می‌شود: بر اساس دانش آوایی و نزدیکی ویژگی‌ها در فضای آکوستیکی. در این مقاله با استفاده از روش متمایزسازی خطای کمینه کلاس‌بندی با متمایز کردن ویژگی‌های سیگنال گفتار با بهبود کلاس‌بندی، کارآیی روش MLLR بهبود یافته است. این بهبود در حالت اول با متمایزتر کردن کلاس‌های رگرسیون و در حالت دوم با یکنواخت کردن توزیع مخلوط‌های گوسی در کلاس‌های رگرسیون ایجاد می‌شود. روش پیشنهادی یعنی بکارگیری خطای کمینه کلاس‌بندی نسبت به روش‌های معادل بدون خطای کمینه کلاس‌بندی، به ترتیب باعث افزایش نسبی ۰.۴۲% تا ۰.۵۸% و ۰.۱۲% تا ۰.۷۲% در نرخ بازشناسی واج به ترتیب برای دادگان TIMIT و WSJ می گردد. در حالتی است که درخت کلاس رگرسیون بر اساس نزدیکی ویژگی‌ها در فضای آکوستیکی ساخته شده است ، نرخ بازشناسی واج بر روی دادگان TIMIT ۰.۲۵% افزایش نسبی یافته است.

کلمات کلیدی

بیشینه درست‌نمائی با رگرسیون خطی، تطبیق گوینده،کمترین خطای رده‌بندی.

۱- مقدمه

سیستم‌های بازشناسی گفتار مستقل از گوینده عملکرد خوبی برای گوینده‌هایی که در مرحله آموزش بوده‌اند، از خود نشان می‌دهند. با این وجود نرخ بازشناسی زمانی که شرایط آموزش و آزمایش متفاوت باشد، به خصوص زمانی که گوینده جدیدی مورد ارزیابی قرار گیرد، کاهش محسوسی می‌یابد. در عمل معمولاً داده‌های زیاد از گوینده جدید در دست نیست. روش‌های تطبیق سیستم مرجع به سیستم مربوط به گوینده جدید برای رفع این نقیصه مطرح شده‌اند که با وجود داده‌های کم گوینده جدید، عملکرد خوبی از خود نشان می‌دهند.

تطبیق با روش رگرسیون خطی با بیشینه درست‌نمایی (Maximum Likelihood Linear Regression (MLLR) به عنوان یکی از روش‌های تطبیق گوینده با داده‌های تطبیقی کم شناخته شده است[۱]. در این روش برای پارامترهای مخلوط‌های گوسی مدلها (میانگین‌ها و واریانس‌ها) در سیستم مدل مخفی مارکوف (HMM) مجموعه‌ای از تبدیل‌های خطی در نظر گرفته می‌شود. توابع تبدیل به گونه‌ای تخمین زده می‌شوند که احتمال درست‌نمایی داده‌های تطبیقی بیشینه شود.

برای تخمین ماتریس تبدیل از درخت رگرسیون خطی استفاده می‌کنیم. در این درخت اجزای مشابه در یک کلاس قرار و همگی تحت اعمال یک ماتریس تبدیل قرار می‌گیرند. تلاش‌های زیادی در جهت بهبود درخت رگرسیون مورد استفاده در روش MLLR برای بهبود تطبیق صورت گرفته است. در [۲] نشان داده شده است که همبستگی

معیار خوبی برای ساخت درخت رگرسیون می‌باشد. دو رویکرد در طراحی درخت‌های رگرسیون مورد توجه است[۳]:

دانش آوایی: در این روش برای تصمیم‌گیری اینکه کدام اجزا با یکدیگر تشکیل یک کلاس دهند، از دانش آوایی استفاده می‌شود. اجزا بر اساس کلاس‌های آوایی متنوع (مثل خیشومی ها و انفجاریها) یا در سطح پایین‌تر بر اساس واج‌ها از هم جدا می‌شوند.

فضای آکوستیکی: در این روش اجزا بدون در نظر گرفتن اینکه به کدام واج تعلق دارند، بر اساس نزدیکی‌شان در فضای آکوستیکی خوشه‌بندی می‌شوند. مزیت این روش عدم نیاز به دانش آوایی و وابستگی به داده-های اعمالی می‌باشد، با این حال هیچ اطلاعاتی در مورد هویت آواها در کلاس‌ها وجود ندارد.

روش خطای کلاس‌بند کمینه (Minimum Classification Error (MCE))، از روش‌های متمایزسازی است که هم برای تبدیل ویژگی و هم برای آموزش کلاس‌بند بکار گرفته می‌شود[۵]. استفاده از MCE برای تطبیق برای نخستین بار توسط Lin و همکارانش در [۶] گزارش شده است. در سایر مطالعاتی که در این زمینه انجام شده است از MCE عنوان راه‌حلی در مقایسه با بیشینه درست‌نمایی برای بهبود یادگیری پارامترهای مدلهای مخفی مارکوف ارایه شده است[۷].

روش پیشنهادی مقاله بکارگیری MCE به عنوان متمایزساز در تطبیق MLLR با دو نگرش ساخت درخت رگرسیون بر اساس دانش آوائی و فضای آکوستیکی می‌باشد. در [۸] اثر اعمال MCE در یکنواختی توزیع مخلوط‌های گوسی در کلاس‌های رگرسیون گزارش شده است. در این مقاله نتایج آزمایش برای این حالت روی کل گوینده‌های مجموعه تست TIMIT آورده شده است. همچنین MCE برای درخت رگرسیون ساخته شده بر اساس دانش آوایی درجهت افزایش تفکیک-پذیری بکارگرفته شده است و باعث می‌شود تا خوشه‌های تشکیل شده در درخت رگرسیون MLLR بیشترین تمایز را داشته باشند. از آنجائی که عملکرد MLLR رابطه مستقیمی با درخت رگرسیون دارد لذا عملکرد سیستم تطبیق بالا می‌رود.

ساختار ادامه مقاله به این صورت خواهد بود که در بخش دوم روش رگرسیون خطی با بیشینه درستنمایی شرح داده می‌شود. بخش سوم به روش خطای کمینه کلاسبندی می‌پردازد. بخش چهارم حاوی روش پیشنهادی است. نتایج آزمایشات در بخش پنجم بیان شده است. و بخش ششم در برگیرنده نتیجه‌گیری است.

۲- بیشینه احتمال درستنمایی رگرسیون خطی (MLLR)

روش MLLR با بکارگیری تبدیل اشتراکی قادر است تبدیل تطبیقی مقاومی را حتی برای آن دسته از پارامترهای مدل که در داده‌های تطبیقی نیستند فراهم سازد. معمولاً تطبیق تنها برای میانگین صورت می‌گیرد، چون فرض بر این است که تفاوت میان گوینده‌ها ناشی از متوسط موقعیت آواها در فضای آکوستیکی می‌باشد. از تطبیق کوواریانس کمتر استفاده می‌شود و اثر آن نیز غالباً کمتر از تطبیق میانگین است[۱].

تخمین پارامترهای مدل مربوط به گوینده جدید به کمک یک تابع شباهت مانند احتمال درستنمائی انجام می‌شود. بنابراین هدف تخمین پارامترهای مدل است به گونه‌ای که پارامترهای جدید باعث افزایش تابع احتمال درستنمایی شوند. فرض کنید m_k بردار میانگین مخلوط گوسی با n بعد باشد، می‌توان نوشت:

$$\hat{m}_k = W_k x_k \qquad (۱)$$

که W_k ماتریس تبدیل با ابعاد $n\times(n+1)$ و $x_k = [w, m_{k_1}, ..., m_{k_n}]^T$ بردار میانگین گسترش یافته است. w نشانگر افست است. فرض کنید داده‌های تطبیقی، مجموعه‌ای از T مشاهده باشد، $O=\{o_1,o_2,...,o_T\}$ و W_k برای R حالت $\{1,2,\ldots,r,\ldots,R\}$ تسهیم شده باشد، شکل عمومی رابطه تخمین تابع تبدیل MLLR برای کلاس kام به صورت رابطه (۲) است:

$$\sum_{r=1}^{R}\sum_{t=1}^{T} g_{kr}(t) C_{kr}^{-1} o_t x_{kr}^t = \sum_{r=1}^{R}\sum_{t=1}^{T} g_{kr}(t) C_{kr}^{-1} W_k x_{kr} x_{kr}^t \qquad (۲)$$

که C ماتریس کوواریانس و $g_r(t)$ احتمال پیشین می‌باشد.

۳- خطای کمینه کلاس‌بندی

ایده اصلی الگوریتم MCE، بهینه‌سازی نرخ خطای تجربی با استفاده از یک مجموعه داده آموزشی، جهت بهبود نرخ خطای بازشناسی می‌باشد. پس از آنکه نرخ خطای تجربی بوسیله یک کلاس‌بند کمینه شد، تخمینی بایاس شده از نرخ خطای واقعی بدست می‌آید. یک راه مؤثر جهت کاهش این نرخ و بهبود کارآیی کلی سیستم، افزایش مرز میان کلاس‌ها در داده‌های آموزشی است. به این منظور از گرادیان کاهشی یا ترکیب امتیازهای حاشیه و نرخ خطای تجربی استفاده شده است. در این روش از تابع هزینه‌ای بصورت رابطه (۳) استفاده شده است.

$$d_{k,Add}(O,F) = -g_k(O,F) + \frac{1}{h}\log\left[\frac{1}{M-1}\sum_{i=1,i\neq k}^{M}\exp(g_i(O,F)h)\right] \qquad (۳)$$

که در آن M تعداد کلاس‌ها، η مقداری مثبت، F مجموعه مدلها و $g_i(O,F)$ لگاریتم احتمال تولید شده برای مشاهده O، به شرط مدل λ_i بوسیله HMM می‌باشد. در روش MCE ، هدف کمینه نمودن تابع هزینه (۳) است. اما از آنجا که این توابع مشتق‌پذیر نیستند، از یک تابع پیوسته نرم مانند تابع سیگموید مطابق رابطه (۴) بعنوان تابع هزینه استفاده می‌کنیم.

$$l_k(O,F) = \frac{1}{1+\exp(-a d_k(O,F))} \qquad (۴)$$

که $d_k(O,F)$ تابع هزینه (۳) می‌باشد و a پارامتر تنظیم بزرگتر از یک است. بنابراین اگر $d_k(O,F)$ خیلی کوچکتر از صفر باشد، در واقع کلاس‌بندی درست رخ داده و $l_k(O,F)$ به صفر نزدیک می‌شود. از طرفی مقدار مثبت برای $d_k(O,F)$ نشان دهنده جریمه برای کلاس-بندی نادرست است و در اینصورت $l_k(O,F)$ به ۱ متمایل می‌شود. کل خطای کلاس‌بندی برای مشاهده O از رابطه (۵) بدست می‌آید:

$$L=\sum_{k=1}^{M} l_k(O,F) \qquad (۵)$$

با استفاده از معیار تابع هزینه سعی می‌شود ماتریس تبدیل بهینه بدست آید که با اعمال این تبدیل بر ویژگی‌ها تمایز بین کلاس‌ها بیشتر می‌شود. مقادیر درایه‌های ماتریس تبدیل به روش تکراری با استفاده از گرادیان کاهشی برای تابع L، با رابطه (۶) قابل محاسبه خواهد بود.

$$w_{n,iter}=w_{n,iter-1}-b\,\frac{\partial L}{\partial w_n} \qquad (۶)$$

که w_n، nامین درایه در قطر اصلی ماتریس تبدیل W؛ $iter$ اندیس تکرار الگوریتم و b پارامتر یادگیری می‌باشد. رابطه (۶) نشان دهنده یک رویه تکراری است. این رویه زمانی متوقف می‌شود که مقدار تابع هزینه از یک حد آستانه کمتر شود.

۴- روش پیشنهادی

فرض کنید X رشته‌ای از بردارهای ویژگی سیگنال گفتار باشد.

$$X=\{\vec{x}_n,\vec{x}_{n-1},\ldots,\vec{x}_1\} \qquad (۷)$$

درستنمایی مشاهده شدن بردارهای ویژگی در X با داشتن رشته آوایی U به صورت P(X|U) نمایش داده و به عنوان مدل آکوستیکی شناخته می‌شود. مدل آکوستیکی را می‌توان به صورت رابطه (۸) بسط داد.

$$P(X|U)=P(\vec{x}_n,\vec{x}_{n-1},\ldots,\vec{x}_1|U)=\prod_{n=1}^{N}P(\vec{x}_n|\vec{x}_{n-1},\ldots,\vec{x}_1,U) \qquad (۸)$$

با اعمال قانون بیز به عبارت درستنمایی $\vec{x}_n$ داریم.

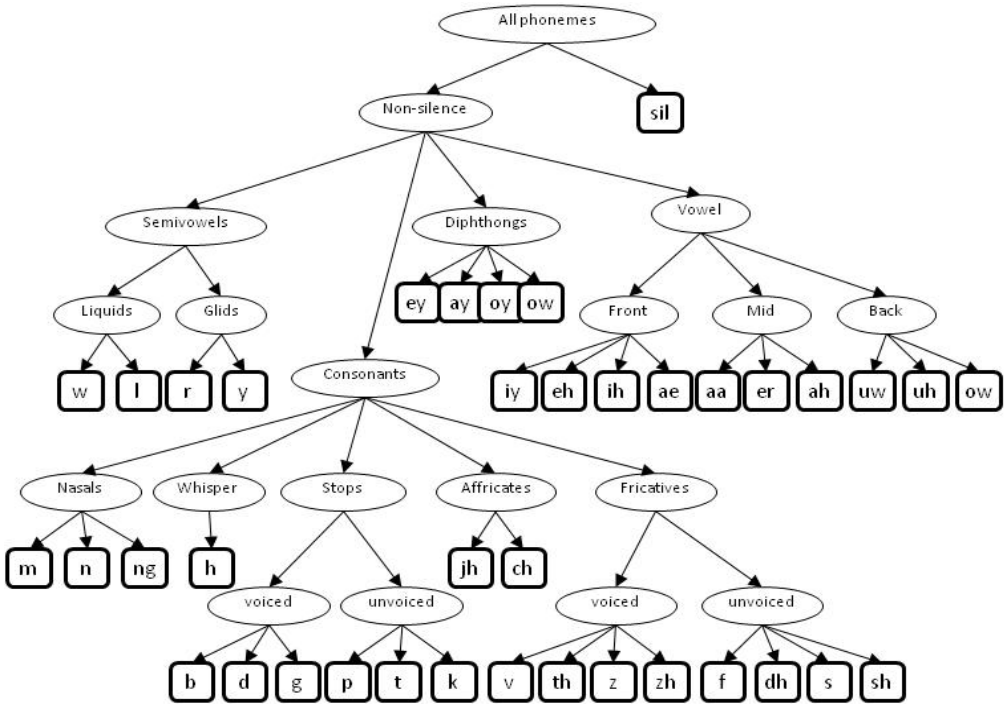
$$P(\vec{x}_n|\vec{x}_{n-1},\ldots,\vec{x}_1,U)=P(\vec{x}_n|U)\frac{P(\vec{x}_{n-1},\ldots,\vec{x}_1|\vec{x}_n,U)}{P(\vec{x}_{n-1},\ldots,\vec{x}_1|U)} \qquad (۹)$$

همانطور که می‌بینیم این عبارت به صورت حاصلضرب مدل آکوستیکی استاندارد $\vec{x}_n$ در یک نسبت احتمالی که تحت عنوان نسبت ثبات (Consistency Ratio) خوانده می‌شود، نوشته می‌شود. این نسبت نشان دهنده وابسته بودن بردارهای ویژگی می‌باشد که در آموزش اکثر سیستم‌های غیروابسته به گوینده در نظر گرفته نمی‌شود. با صرف‌نظر کردن همبستگی موجود میان مشاهدات مختلف، این سیستم‌ها قادر به نشان دادن محدودیت‌های گوینده نمی‌باشند[۹]. استفاده از MCE برای متمایز کردن مدل‌های مربوط به واج‌های مختلف باعث کاهش این همبستگی می‌شود و در نتیجه مدل‌ها به نحو بهتری آموزش می-یابند[۸].

کلاس رگرسیون بر اساس دانش آوایی استفاده شده در آزمایشات مربوط به این مقاله در شکل ۱ نشان داده شده است.کلاس‌های رگرسیون نهایی در این درخت واج‌های موجود در مدل مرجع می-باشند. در نظر نگرفتن همبستگی موجود در بردارهای ویژگی که در بالا توضیح داده شد باعث می‌شود تخمین ماتریس تبدیل برای این کلاس-ها به کمک داده‌های تطبیقی نیز تحت تأثیر قرار گیرد. به عبارتی بنا به رابطه (۲) مجزاتر شدن واج‌ها باعث می‌شود داده‌های تطبیقی اثر بهتری برای تخمین ماتریس MLLR داشته باشند، چرا که در این حالت مدل‌ها مجزاتر شده‌اند و اثر داده‌های تطبیقی مربوط به یک واج روی مدل‌های مربوط به واج‌های دیگر کاهش یافته و می‌توان گفت ماتریسی که برای یک کلاس تخمین زده می‌شود تأثیر بیشتری از داده‌های تطبیقی مربوط به همان کلاس داشته است.

بنا به رابطه (۲) علاوه بر میزان داده‌های تطبیقی هر چه تعداد گوسی-های هر کلاس بیشتر باشد تخمین مقاوم‌تری برای آن کلاس انجام می‌شود[۱۰]. بنابراین یکی از عواملی که می‌تواند باعث کاهش دقت تخمین ماتریس MLLR برای یک کلاس شود، کم بودن تعداد گوسی‌های آن کلاس است. با توجه به ثابت بودن تعداد کل گوسی‌ها، زیاد شدن تعداد گوسی‌ها در یک کلاس معادل کم شدن آن‌ها در کلاس‌های دیگر می‌شود. به طور کلی تجمع بیش از حد گوسی‌ها در یک کلاس نه تنها باعث کاهش دقت تخمین برای کلاس‌های دیگر می‌شود، بلکه باعث شبیه شدن تطبیق چند کلاسی به تطبیق تک کلاسی نیز می‌گردد. برای جلوگیری از تجمع مخلوطهای گوسی در یک کلاس باید به معیار خوشه‌بندی گوسی‌ها مراجعه کرد. همانطور که اشاره شد، این خوشه‌بندی معمولاً به دو صورت انجام می‌شود. این کلاس‌بند با در نظر گرفتن نزدیکی میانگین گوسی‌ها آن‌ها را کلاس-بندی می‌کند. بنابراین برای این که در خوشه‌بندی، گوسی‌ها در یک کلاس متمرکز نشوند می‌بایست فاصله بین میانگین‌های گوسی‌ها افزایش داده شود.

در روش پیشنهادی این مقاله ابتدا MCE به عنوان مرحله پیش-پردازش قبل از خوشه‌بندی MLLR بکار رفته سپس خوشه‌بندی انجام می‌شود. این عمل باعث کلاس بندی بهتر پارامترهای مدل می-گردد. انتظار می‌رود که با بکارگیری یک روش متمایزساز قبل از هر کلاس‌بندی کارائی آنرا بالا برد.



شکل (۱) – درخت رگرسیون MLLR مبتنی بر دانش آوائی

۵- نتایج آزمایشات

به منظور بررسی عملکرد روش پیشنهادی، نرخ بازشناسی در سیستم بازشناسی غیروابسته به گوینده و سیستم تطبیق یافته با MLLR در دو حالت با MCE و بدون MCE مقایسه شده است. برای انجام آزمایشات از دادگان TIMIT و WSJ استفاده شده است. در این بخش نتایج آزمایشات در سه قسمت بیان می‌شود. در ابتدا اثر MCE در خوشه‌بندی مبتنی بر دانش آوایی (مطابق شکل (۱)) روی دادگان TIMIT بررسی می‌شود. در ادامه آزمایشات برای ارزیابی اثر MCE روی درخت رگرسیون MLLR بدست آمده بر اساس فضای آکوستیکی روی کل دادگان TIMIT و WSJ گزارش شده است. برای این منظور سیستم بازشناسی غیروابسته به گوینده با ۳۹ مدل برای واجها آموزش داده شده است. در این سیستم هر مدل مخفی مارکوف دارای ۳ حالت و هر حالت دارای ۱۶ گوسی می‌باشد. همچنین خوشه بندی مخلوطهای گوسی را برای ۲، ۴، ۸ و ۱۶ کلاس انجام داده شد.

جدول (۱) نتایج نرخ بازشناسی واج را برای روش پیشنهادی روی دادگان TIMIT نشان می‌دهد. با توجه به نتایج مندرج در جدول (۱)، بکارگیری MCE در سیستم بازشناسی گفتار مستقل از گوینده (SI-MCE) نسبت به سیستم بازشناسی گفتار مستقل از گوینده (SI) باعث افزایش نرخ بازشناسی واج می شود. همچنین روش پیشنهادی یعنی بکارگیری MCE در MLLR باعث بهبود عملکرد سیستم می شود. همانطور که می‌توان دید استفاده از MCE باعث ۰.۲۵ بهبود نسبی در نرخ بازشناسی واج در حالت تطبیق به کمک درخت رگرسیون ساخته شده به کمک دانش آوایی شده است.

جدول (۱) - نتایج نرخ بازشناسی واج روی دادگان TIMIT برای روش تطبیق با درخت رگرسیون ساخته شده با دانش آوایی

روش	نرخ بازشناسی واج
SI	80.77
SI-MCE	80.91
MLLR	88.55
MLLR-MCE	88.9

با توجه به مدل استفاده شده، تعداد کل مخلوطهای گوسی ۱۸۷۲ می-باشد. همان گونه که گفته شد مطلوب این است که توزیع این مخلوطهای گوسی در کلاسها بطور یکنواخت انجام گیرد. آزمایشات نشان می‌دهد روش پیشنهادی یعنی بکارگیری MCE قبل از MLLR باعث می‌شود که مخلوطهای گوسی بطور یکنواخت‌تری بین کلاسها توزیع شوند[۸]. همچنین با افزایش تعداد کلاسها عملکرد روش پیشنهادی بهتر می‌شود و این امر به این دلیل است که با افزایش تعداد کلاس‌ها احتمال تداخل نمونه‌های کلاس‌ها و پیچیدگی مساله بیشتر می‌شود بنابراین اثر MCE به عنوان متمایزساز بیشتر دیده. در ادامه نتایج بازشناسی واج برای روشهای مختلف روی دادگان TIMIT و WSJ آورده شده است.

سیستم بازشناسی با استفاده از دادگان بخش آموزش TIMIT آموزش داده شده است و برای تطبیق هر گوینده از ۱۰ جمله در بخش تست استفاده شده است. نتایج بازشناسی روی ۱۴۰ گوینده میانگین گرفته شده است. با توجه به اینکه دادگان تطبیق برای خوشه بندی MLLR به ۸ و ۱۶ کلاس کم می‌باشد، نتایج بازشناسی فقط برای خوشه‌بندی MLLR با ۲ و ۴ کلاس بیان شده است.

جدول (۲) نتایج نرخ بازشناسی واج را برای روشهای روی دادگان TIMIT نشان می‌دهد. همانطور که می توان دید با افزایش تعداد

کلاس‌ها در MLLR عمل تطبیق بهتر انجام می‌گیرد. این امر را نیز در جدول (۲) می توان دنبال نمود، بگونه‌ای که نرخ بازشناسی واج در حالت ۲ کلاس حدود ۰.۴۲ درصد افزایش نسبی و در حالت ۴ کلاس حدود ۰.۵۸ درصد افزایش نسبی یافته است.

جدول (۲) - نتایج نرخ بازشناسی واج روی دادگان TIMIT برای روشهای مختلف

روش	نرخ بازشناسی واج
SI	80.77
SI-MCE	80.91
MLLR-2	87.57
MLLR-MCE-2	88.06
MLLR-4	88.89
MLLR-MCE-4	89.51

برای داشتن کلاسهای بیشتر در خوشه‌بندی MLLR نیاز به دادگان تطبیقی بیشتری برای هرگوینده می‌باشد. به همین دلیل آزمایشات را بر روی دادگان WSJ نیز انجام داده که نتایج آن در ادامه آورده شده است.

در این آزمایش از مدل‌های آموزش داده شده با دادگان آموزشی TIMIT به عنوان مدل‌های مرجع استفاده شده است. نتایج مرحله تست بر اساس میانگین‌گیری روی ۲۰ گوینده از دادگان WSJ گزارش شده است، که برای هر گوینده ۴۰ جمله برای تطبیق و ۷۰ جمله برای تست استفاده شده است.

جدول (۳) نرخ بازشناسی واج روی دادگان WSJ برای روشهای مختلف نشان می‌دهد. چنانچه دیده می‌شود با افزایش تعداد کلاس در خوشه‌بندی MLLR نرخ بازشناسی واج افزایش می‌یابد. مطابق مقادیر نشان داده شده در جدول (۳) با بکارگیری MCE در روش MLLR در خوشه‌بندی ۲ کلاسی ۰.۱۲ درصد و با افزایش تعداد کلاسها به ۱۶ کلاس میزان بهبود به ۰.۷۴ درصد می‌رسد.

جدول (۳) - نتایج نرخ بازشناسی واج روی دادگان WSJ برای روشهای مختلف

روش	نرخ بازشناسی واج
SI	45.29
SI-MCE	45.35
MLLR-2	50.18
MLLR-MCE-2	50.3
MLLR-4	51.24
MLLR-MCE-4	51.6
MLLR-8	52.89
MLLR-MCE-8	53.5
MLLR-16	55.02
MLLR-MCE-16	55.76

همانطور که نتایج آزمایشات انجام گرفته روی دادگان TIMIT و WSJ نشان دادند، روش پیشنهادی باعث افزایش نرخ بازشناسی در سیستم تطبیق گوینده شده است.

۶- نتیجه‌گیری

در این مقاله از روش متمایز ساز خطای کمینه کلاس‌بندی برای تفکیک ویژگی‌های سیگنال گفتار بر اساس مدلهای مخفی مارکف مرجع استفاده می‌شود. این امر باعث بهبود تطبیق MLLR در هر دو حالت تشکیل کلاس رگرسیون چه به کمک دانش آوایی و چه بر اساس نزدیکی در فضای آکوستیکی می‌شود. در حالت اول MCE با کاهش وابستگی آماری بین واج‌های مختلف در حوزه ویژگی‌ها باعث بهبود تطبیق می‌شود و در حالت دوم ضعف روش MLLR در خوشه بندی را برطرف کرده و موجب خوشه‌بندی یکنواخت می‌گردد. روش پیشنهادی باعث افزایش نسبی ۰.۴۲% تا ۰.۵۸% و ۰.۱۲% تا ۰.۷۲% در نرخ بازشناسی واج برای دادگان TIMIT و WSJ می گردد. این افزایش در حالتی است که درخت کلاس رگرسیون بر اساس نزدیکی ویژگی‌ها در فضای آکوستیکی ساخته شده است، در جایی که درخت رگرسیون بر اساس دانش آوایی ساخته شده است، نرخ بازشناسی واج بر روی دادگان TIMIT ۰.۲۵% افزایش یافته است.

مراجع

[1] C. J. Leggetter and P. C. Woodland, "*Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models*" J. Comput. Speech Lang., vol. 9, pp. 171–185, 1995.

[2] R. Haeb-Umbach, "*Automatic generation of Phonetic regression class trees for MLLR adaptation*", IEEE Trans. Speech Audio Process., vol. 9, no. 3, 2001.

[3] M. J. F. Gales, "*The generation and use of regression class trees for MLLR adaptation*," Cambridge Univ., Cambridge, U.K., Tech. Rep. CUED/F-INFENG/TR263, 1996.

[4] A. Mandal, M. Ostendorf, A. Stolcke, "*Improving robustness of MLLR adaptation with speaker-clustered regression class trees*", computer speech and language 23, 176-199, 2009.

[5] B. Zhang, and S., Matsoukas, "*Minimum phoneme error based heteroscedastic linear discriminant analysis for speech recognition*". In: Proceedings of ICASSP, vol. 1. Philadelphia, PA, pp. 925-928, 2005.

[6] C.-H. Lin, C.-H. Wu, and P.-C. Chang, "*A study on speaker adaptation for Mandarin syllable recognition with minimum error discriminative training*", IEICE Trans. Inf. & Syst., Vol.E78-D, No.6, pp.712-718, 1995.

[7] J. Wu and Q, Huo, "*A study of Minimum Classification error (MCE) Linear regression for supervised adaptation of MCE-trained Continues-Density Hidden Markov Model* ", IEEE Trans. On Audio, Speech and Language Processing, vol. 15, no. 2, Feb. 2007.

[۸] رضا صحرائیان، بهزاد زمانی، بابک ناصر شریف، احمد اکبری، احمد آیت‌اللهی، "**بهبود روش تطبیق گوینده MLLR در بازشناسی گفتار غیروابسته با کمک خطای کمینه کلاس‌بندی**"، پانزدهمین کنفرانس کامپیوتر ایران، متن، اسفند ۱۳۸۸.

[9] Timothy J. Hazen, "*A comparison of novel techniques for rapid speaker adaptation*", Speech Communication, vol. 31, no. 1, pp. 15-33, May 2000.

[10] Jonathan E. Hamaker, "*MLLR: A Speaker Adaptation Technique for LVCSR*", Lecture for a course at ISIP - Institute for Signal and Information Processing, Department of Electrical and Computer Engineering, Mississippi State University, 1999.