تطبیق گوینده در بازشناسی گفتار پیوسته براساس تخمین MAP مبتنی بر تبدیل MLLR

سعید شریفیان* و سید محمد احدی**

دانشکدهٔ مهندسی برق، دانشگاه صنعتی امیر کبیر

(دریافت مقاله: ۸۲/٤/٤ – دریافت نسخه نهایی: ۸۲/۱۱/٤)

چکیده ـ روشهای مختلفی برای تطبیق گوینده در سیستمهای بازشناسی گفتار معرفی گردیده‌اند. در برخی روشها نظیر تخمین MAP تنها مدلهایی که داده آموزشی متناظرشان موجود باشد تازه سازی می‌شوند و برای بهبود قابل توجه دقت بازشناسی، داده آموزشی نسبتاً زیادی مورد نیاز است. در برخی دیگر نظیر MLLR که تعدادی تبدیلات عمومی بر روی خوشه‌های مدلها اعمال می‌شود، برای دادگان کم آموزشی نتایج مطلوبی حاصل می‌شود، اما با افزایش دادگان، کارایی به حد اشباع می‌رسد. در این مقاله روش جدیدی مطرح می‌شود که از مزایای هر دو روش فوق برای دسترسی به کیفیت بالاتر بهره می‌برد. در این روش مدلهایی که داده آموزشی آنها موجود است به کمک تخمین MAP آموزش می‌بینند و برای مدلهایی که داده آموزشی (کافی) ندارند، با استفاده از روش MLLR مقادیر پیشینه مناسب برای تخمین MAP تأمین می‌شود. این روش، در عمل، بر روی یک سیستم آموزش دیده براساس دادگان فارس دات به نتایج بهتری نسبت به هر یک از دو روش MAP و MLLR دست یافته است.

واژگان کلیدی : مدلهای مارکوف پنهان (HMM)، بازشناسی گفتار پیوسته فارسی، تطبیق گوینده، تخمین MAP، تبدیل MLLR

Speaker Adaptation in Continuous Speech Recognition Using MLLR-Based MAP Estimation

S. Sharifian and S. M. Ahadi

Electrical Engineering Department, Amirkabir University, Hafez Avenue, Tehran 15914, Iran

Abstract: ***A variety of methods are used for speaker adaptation in speech recognition. In some techniques, such as MAP estimation, only the models with available training data are updated. Hence, large amounts of training data are required in order to have significant recognition improvements. In some others, such as MLLR, where several general transformations are applied to model clusters, the results are desirable for small training data, but with increasing training data, the performance improvement reaches the saturation lvel. In this paper, a new approach is introduced that makes use of the advantages of both mentioned techniques to improve the recognition rate. Here, the models with available training data are trained using MAP while***

---

* – کارشناس ارشد ** – استادیار

for those with insufficient training data, appropriate prior parameters for MAP estimation are found using MLLR. This technique has yielded better performance in comparison to either MAP or MLLR, in a system based on FARSDAT speech corpus.

Keywords: *Hidden Markov models (HMM), continuous Persian (Farsi) speech recognition, Speaker adaptation, MAP estimation, MLLR transformation.*

فهرست علائم

$C^{(r)}$	ماتریس قطری کوواریانس مقیاس شده گوسی rام در MLLR	φ	پارامتر رشد در MAP وزنی
$c_{ii}^{(r)}$	عنصر قطری iام از ماتریس قطری کوواریانس مقیاس شده گوسی r ام در MLLR	$\gamma(t)$	درستنمایی بودن در یک حالت خاص از مدل در لحظه t
$d(.,.)$	فاصله دو حالت از مدل	λ	مجموعه پارامترهای مدل
O_t	داده‌های تطبیق در لحظه t	μ	بردار میانگین توزیع احتمال چند متغیره
$P(.)$	چگالی احتمالی	$\mu_{..}$	عضو بردار میانگین توزیع احتمال چند متغیره
$P(.\|.)$	چگالی احتمالی شرطی	$\hat{\mu}_j^{(q)}$	میانگین مدل اولیه قبل از آموزش، در MAP وزنی
P_r	درستنمایی لگاریتمی میانگین برای مشاهده rام	μ'	میانگین مدل آموزش دیده‌ای در آن دسته که داده آموزشی متناظر آن موجود است(در MAP وزنی)
w_j	ماتریس تبدیل n × (n+1) در MLLR		
Z, G_i	پارامترهای کمکی تبدیل MLLR	ν_j	بردار میانگین گسترش یافته در MLLR
$\alpha^{(q)r}, \beta^{(q)r}$	پارامترهای آلگوریتم جلوروونده–عقب رونده برای مشاهده rام از مدل qام	Σ	ماتریس کوواریانس توزیع احتمال چند متغیره
$\mu_k, \tau_k, \alpha_k, \beta_k$	پارامترهای توزیع پیشینه	σ	عضو ماتریس کوواریانس توزیع احتمال‌چندمتغیره

۱- مقدمه

برقراری ارتباط طبیعی بین انسان و ماشین از جمله مواردی است که در سالهای اخیر مورد توجه بسیاری از محققان قرار گرفته است. با توجه به اینکه گفتار یکی از مهمترین روشهای برقراری ارتباط میان انسانهاست، تلاش بسیاری برای برقراری ارتباط گفتاری میان انسان و ماشین صورت گرفته است. بازشناسی گفتار، یکی از دو محور اصلی در این ارتباط است.

مدلهای مارکوف پنهان[1] (HMMs) امروزه به عنوان پرکاربردترین روش برای برآوردن این منظور مورد استفاده قرار می‌گیرند. یکی از مسائلی که همواره فراروی این‌گونه سیستمها (و به طور کلی هر سیستم بازشناسی گفتاری) بوده، مسئله تنوع گویندگان و ویژگیهای گفتاری آنها بوده است. این مطلب باعث شده که در بسیاری از موارد، محققان برای پرهیز از این مشکل، به ایجاد سیستمهای وابسته به گوینده[2] (SD) روی آورند. این‌گونه سیستمها، که با استفاده از گفتار یک گوینده خاص آموزش می‌بینند، از توانایی بالایی در بازشناخت گفتار گوینده مزبور برخوردارند. با این همه، چنانچه از چنین سیستمی برای بازشناسی گفتار گوینده دیگری استفاده شود توانایی سیستم به شدت افت می‌کند. در مقابل سیستمهای ناوابسته به گوینده[3] (SI)، به خاطر آموزش دیدن با گفتار گویندگان مختلف[4]، قادر به بازشناخت گفتار طیف وسیعی از گویندگان هستند. به همین دلیل این‌گونه سیستمها مورد توجه بیشتر محققان، در مقایسه با سیستمهای وابسته به گوینده، قرار گرفته‌اند. با این همه، توانایی اینگونه سیستمها در بازشناسی گفتار یک گوینده عموماً کمتر از توانایی یک سیستم وابسته به گوینده آموزش دیده با گفتار همان گوینده است. لازم به اشاره است که این مطلب، نه تنها می‌تواند ناشی از عدم تطابق میان مشخصات گویندگان در دادگان آموزشی و مشخصات گوینده جدید باشد، بلکه تفاوت در

میکروفون ، کانال انتقال و نویز محیط نیز می‌تواند مؤثر باشد.

بـرای جبـران ایـن عدم تطابقها و دستیابی به دقت بالاتر در بازشناسـی گفتار یک گوینده جدید، روشهایی مطرح گردیده‌اند که می‌توانند در دو گروه کلی دسته بندی شوند:

۱- جبران سازی ویژگیهای استخراج شده ازگفتار [۱]

۲- تطبـیق مـدل گویـنده [۲ و ۳] کـه در آن پارامتـرهای مـدل HMM تغییر می‌کنند.

هر چند که ترکیب این دو روش [٤] نیز می‌تواند به نتایج خوبی منتهی شود، دراین مقاله ما تنها به بررسی عملکرد تطبیق مدل می‌پردازیم.

یـک روش تطبـیق خوب روشی است که بتواند با در اختیار داشـتن داده آموزشی کمی از یک گوینده جدید دقت بازشناسی را (تـا حـد امکـان) افـزایش دهـد. از آن مهمتر اینکه با موجود بـودن داده آموزشـی کافـی (برای آموزش عادی سیستم با گفتار گویـنده جدیـد)، دقتـی بهتر یا معادل روشهای معمول آموزش، نظیـر درسـت نمایـی بیشـینه[۵] (ML)، ارائه دهد. روشهای کمی وجود دارند که هر دو مورد فوق را در برگیرند. عمومیترین این روشـها، روش تطبیقـی اسـت کـه براسـاس تئوری بِیز[٦] بنا شده اسـت و تخمـین MAP نامـیده مـی‌شـود[۲ و ۵]. در این روش پارامتـرهای مـدل بـه عـنوان متغیرهای تصادفی که تابع چگالی احـتمال آنها موجود است در نظر گرفته می‌شوند. در این حالت تخمـین MAP را مـی‌تـوان بـه دسـت آوردن مـد تابـع چگالی احـتمال پسینه[۷] با وجود داده آموزشی توصیف کرد. هنگامی که داده کمـی از گویـنده جدیـد بـرای آموزش وجود دارد نتایج به دسـت آمـده از تخمـین MAP مـی‌تـواند بسیار بهتر از نتایج به دسـت آمـده ازتخمـین ML باشد. با افزایش داده‌های آموزشی، کارایـی تخمـین MAP بـه سـمت کارایـی ML مـیل مـی‌کند و توانایی بازشناسی همانند یک سیستم SD را ارائه می‌کند. ضعف روش MAP در ایـن اسـت کـه تـنها مدلهایـی که داده آموزشی متناظر آنها وجود دارد تطبیق می‌شوند و این امر هنگامی که داده آموزشـی کمی (مثلاً چند جمله، در یک سیستم بازشناسی گفتار پیوسـته بـا دایـره لغـات متوسط) در اختیار است منجر به نتایج خوبی نمی‌شود.

روشـهای دیگری نیز وجود دارند که اساساً با تخمین MAP تفـاوت دارند. مهمترین این روشها، روشهای مبتنی بر تبدیل‌اند. از جملـه ایـن روشـها مـی‌تـوان به MLLR [۳] [۸]، VFS[۹] [٦] و SM[۱۰] [٤] اشـاره کـرد. همچـنان‌کـه از نام آنها برمی‌آید، در این روشها، برای دستیابی به تطبیق پارامترها، اقدام به یافتن یک تابع نگاشت یا تبدیل کلی یا جزیی بر اساس داده‌های تطبیق موجود می‌شود. در این خانواده از تکنیکها، برای رفع مشکل تعداد زیاد پارامتـرهای آزاد در مـدل، ایـن پارامترها به کمک گره زدن یا با اعمـال محدودیتهایـی برروی آنها کاهش می‌یابند. این امر باعث مـی‌شود که با افزایش دادگان آموزشی پارامترها به سمت مقادیر مورد انتظار از روش درستنمایی بیشینه(ML) میل نکنند.

هـیچ یک از روشهای مبتنی بر تبدیل و روش تخمین MAP به طور همزمان دو شرط بهبود بازشناسی با دادگان آموزشی کم و نـزدیک شدن به تقریب ML در صورت آموزش با داده کافی را در بـر نمـی‌گیـرند. بـه همـین دلـیل روشـهای دیگری سعی کـرده‌انـد همـزمان بـه ایـن دو ویژگی دست یابند. از جمله این روشـها، روشهای EMAP[۱۱] [۷]،[۸]، SMAP [۹]، RMP[۱۳] [۱۰] و روش Quasi-Bayes بـا همبستگی میان بردارهای میانگین [۱۱] هسـتند کـه از تخمـین MAP مشـتق شـده‌اند. در این روشها بر اسـاس همبستگی میان مدلها، همه مدلها تازه سازی می‌شوند به این ترتیب که اگر داده آموزشی متناظر با یک مدل موجود باشد تمـام مدلهایـی کـه بـا مدل فوق همبستگی دارند نیز با این داده آموزشـی، آمـوزش مـی‌بینـند. هر چند که این روشها در تئوری کـاملاً عمومی‌اند اما در عمل احتیاج به تقریب زیادی دارند زیرا یافـتن و اسـتفاده از همبسـتگی بین مدلها، به ویژه در زمانی که داده آموزشـی بسـیار کم است، کار مشکلی است. به عنوان مثال در مـراجع [۷ و ۸] مـدلها به چندین زیر مدل تقسیم شده‌اند که تعداد این زیر مدلها به میزان داده آموزشی بستگی دارد.

از سـوی دیگر توسعه برخی روشهای مبتنی بر تبدیل، مانند MLLR ، به نحوی که تخمین MAP را در برگیرند، نظیر روش MAPLR[۱٤] [۱۲ و ۱۳] ، نیـز بـه نتایج خوبی منتهی شده است. در این روش ماتریس تبدیل پیشینه از مدل قبلی برای آموزش به همراه

ماتریس تبدیل به دست آمده از روش MLLR به مدل اعمال می‌شود.

در مقاله حاضر روش جدیدی ارائه شده است که اساس آن نیز بر روشهای MLLR و MAP است و مزایای هر دو روش را در بر می‌گیرد. در این روش، دو شیوه MLLR و MAP به گونه دیگری تؤاماً مورد استفاده قرار می‌گیرند به این نحو که ابتدا به کمک خوشه بندی مدلهای HMM که به صورت سه آوایی[15] آموزش دیده‌اند مدلها دسته بندی می‌شوند. سپس به کمک داده‌های آموزشی و روش MLLR برای هر خوشه یک ماتریس تبدیل به دست می‌آید. یک تبدیل عمومی هم برای خوشه‌هایی که هیچ‌گونه بردار آموزشی ندارند به دست می‌آوریم. حال هر مدلی که داده آموزشی آن موجود باشد به کمک تخمین MAP تطبیق می‌یابد و پارامترهای مدلهایی که داده آموزشی ندارند، با استفاده از مدل تبدیل یافته توسط ماتریس تبدیل MLLR به عنوان مقادیر کمکی و به کمک تخمینی مشابه MAP ، تطبیق می‌شود. به این ترتیب، علاوه بر استفاده از مزایای هر دو روش، معایبی که هر یک از آنها به تنهایی دارند نیز پوشانده می‌شود. پایگاه داده مورد استفاده در این مقاله *فارس دات* است که یک دادگان با دایره کلمات متوسط است.

مقاله حاضر مشتمل بر شش بخش است که پس از مقدمه، بخش دوم به بررسی اجمالی روشهای مبتنی بر تخمین MAP اختصاص یافته است. در بخش سوم به روش تطبیق MLLR و مسائل مربوط به آن پرداخته شده و بخش چهارم به طرح روش پیشنهادی این مقاله می پردازد. در بخش پنجم به مسائل مربوط به پیاده سازی و بررسی نتایج می پردازیم و در بخش ششم نیز نتیجه گیری و پیشنهادات ارائه شده است.

۲- تطبیق بیزین

۲-۱- تخمین MAP

تخمین MAP یا بیزین[16] از جمله روشهایی است که برای تطبیق پارامترهای HMM به کار می‌رود. تفاوت این روش با روش تخمین درستنمائی بیشینه در این است که در روش ML فرض بر این است که پارامترهای مدل ثابت ولی ناشناخته‌اند حال آنکه در تخمین MAP این پارامترها متغیرهای تصادفی فرض می‌شوند که توزیع پیشینه[17] آنها مشخص است. این قابلیت روش MAP را در شرایط داده‌های آموزشی پراکنده[18] کاراتر می‌کند چرا که اطلاعات پیشینه را نیز در تطبیق دخالت می‌دهد.

برای تطبیق گوینده در روش MAP از دو منبع داده استفاده می‌شود؛ یکی چگالی توزیع پیشینه پارامترها $P(\lambda)$ و دیگری داده‌های تطبیقی O_t. روش MAP تلاش در به دست آوردن مدلی دارد که برای آن

$$\lambda_{MAP} \propto \arg\max_{\lambda} P(O_t \mid \lambda) \cdot P(\lambda) \qquad (۱)$$

معادلات مربوط به تخمین بردارهای میانگین و واریانس توسط روش MAP برای یک HMM دارای چگالی مشاهدات تک گوسین چند متغیره به صورت زیر است [۲].

$$\mu_k^{'} = \frac{\tau_k \mu_k + \sum_{t=1}^{n} \gamma_k(t) \cdot o_t}{\tau_k + \sum_{t=1}^{n} \gamma_k(t)} \qquad (۲)$$

$$\Sigma_k^{'} = \frac{2\beta_k + \sum_{t=1}^{n} \gamma_k(t)(o_t - \mu_k^{'})^2 + \tau_k(\mu_k - \mu_k^{'})^2}{2\alpha_k - 1 + \sum_{t=1}^{n} \gamma_k(t)} \qquad (۳)$$

که در این معادلات $\mu_k, \tau_k, \alpha_k, \beta_k$ پارامترهای توزیع پیشینه‌اند [۲].

علی‌رغم اینکه مزیت عمده روش MAP بر ML، استفاده از پارامترهای پیشینه در تخمین است، آنچه در تخمین به شیوهٔ MAP عمدتاً مشکل ساز است،تخمین مناسب پارامترهای پیشینه است، از آنجا که لزوم داشتن پارامترهای پیشینهٔ مناسب انکار ناپذیر است، تلاش فراوانی برای یافتن پارامترهای پیشینه مناسب صورت گرفته که عموماً به راه حلهای غیر تحلیلی منجر شده و هنوز یک راه تحلیلی مناسب برای حل این مسئله یافت نشده است [۱۴].

۲-۲- تخمین MAP وزنی

روش MAP وزنی روشی است که عمدتاً برای استفاده در کاربردهایی که دارای پارامتر آزاد زیاد بوده و از این رو تخمین

پارامترها در آنها دارای مشکل است طراحی شده است [۱۵]. مثالی از کاربرد این روش، تخمین پارامترها در یک سیستم وابسته به متن[۱۹]، نظیر یک سیستم سه آوایی است. در این روش هم مشابه روش گره زدن، که در آموزش این‌گونه مدلها کاربرد فراوانی دارد، حالتهای مشابه خوشه بندی می‌شوند. با مشاهده داده آموزشی متناظر با هر یک از مدلهای موجود در یک دسته، مدل فوق با داده آموزشی جدید به کمک تطبیق MAP معمولی آموزش می‌بیند، اما مدلهای دیگر هم دسته مدل فوق در صورت نداشتن داده آموزشی به کمک معادله زیر بازتخمین می‌شوند [۱۵]:

$$\mu_j^{(q)} = \frac{\tau_j((1-\varphi)\times\hat{\mu}_j^{(q)} + (\varphi\times\mu')) + \sum_{r=1}^{R}\frac{1}{P_r}\sum_{t=1}^{T_r}\alpha^{(q)r}(t,j)\beta^{(q)r}(t,j)O_j^r}{\tau_j + \sum_{r=1}^{R}\frac{1}{P_r}\sum_{t=1}^{T_r}\alpha^{(q)r}(t,j)\beta^{(q)r}(t,j)} \qquad (٤)$$

در معادله فوق $\hat{\mu}_j^{(q)}$ میانگین چگالی احتمال حالت مدل اولیه قبل از آموزش و μ' میانگین چگالی احتمال حالت مدل آموزش دیده‌ای در آن دسته است که داده آموزشی متناظر آن موجود است. φ نیز پارامتر رشد است که معمولا مقداری بین صفر و کمتر از یک دارد و از معادله زیر قابل محاسبه است:

$$\varphi = \left|\frac{\hat{\mu}_j^{(q)} - \mu'}{\hat{\mu}_j^{(q)}}\right| \qquad (٥)$$

مقدار پارامتر φ در حقیقت نسبت تغییرات میانگین آموزش دیده و میانگین مدل اولیه است و به نوعی می‌توان آن را یک پارامتر برای وزن دهی براساس میزان تغییرات مدل بیان کرد. در صورت صفر شدن پارامتر φ تخمین MAP وزنی به سمت تخمین MAP معمولی میل خواهد کرد.

تئوری استفاده از چنین روشی برای تطبیق بر این اساس است که در تخمینMAP صورت رابطه را می‌توان به دو قسمت تجزیه کرد. یکی از این دو قسمت میزان درستنمایی داده آموزشی در مدل فعلی است و دیگری $\tau_j\hat{\mu}_j^{(q)}$ است که بیانگر میزان مشارکت میانگین توزیع پیشینه یعنی $\hat{\mu}_j^{(q)}$ در تخمین MAP است. این بخش در حقیقت میزان استفاده از اطلاعات پیشینه را برای به دست آوردن میانگین جدید مشخص می‌کند. پارامتر $\hat{\mu}_j^{(q)}$ که در اینجا بیانگر اطلاعات پیشینه است، درMAP معمولی از میانگین چگالی احتمال حالت مدل SI قابل محاسبه است. بدیهی است که با نزدیکتر کردن این پارامتر به مدل متناظر SD می‌توان تخمین بهتری را به دست آورد. از آنجا که مدلهای دسته بندی شده در یک گروه دارای مقادیر نزدیک میانگین‌اند، می‌توان میزان تغییرات یک مدل را که برای آن داده آموزشی وجود دارد به سایر مدلهای هم گروه آن تعمیم داد و فرض کرد که تمام میانگینهای موجود در آن مدل به یک نسبت تغییر می‌کنند. با این فرض $\hat{\mu}_j^{(q)}$ موجود در تخمین MAP معمولی را به دو بخش میانگین قبلی مدل SI و میانگین جدید مدل آموزش دیده در گروه، که با نسبت وزنی خاصی که از میزان تغییرات مدل SI و SD ناشی شده است ادغام می‌شوند، تبدیل می‌کنیم.

۳- روش MLLR

۳-۱- الگوریتم خوشه بندی

پیش از آنکه به تبدیل MLLR بپردازیم باید به نحوی دسته‌های تبدیل را تعیین کرد. این عمل به این خاطر انجام می‌شود که پارامترهایی که شبیه به هم هستند با هم در یک گروه قرار گیرند و یک تبدیل یکتا برای همه آنها استفاده شود. روشهای زیادی برای خوشه بندی وجود دارند که می‌توانند بر اساس یک درخت تصمیم‌گیری ویا دوری و نزدیکی بردارهای مدل از یکدیگر بنا شوند. از آنجا که مدل مورد استفاده در این کاربرد سه آوایی است، برای کاهش پارامترها از گره زدن حالتها استفاده شده است. از آنجا که مدل هر آوا از سه حالت و هر حالت از یک تک گوسین تشکیل شده است خوشه بندی عملاً به خوشه کردن گوسینها محدود شده است. برای خوشه بندی می‌توان از چنین الگوریتمی استفاده کرد [۱۶].

ابتدا برای هر یک از حالتها یک دسته جدا در نظر می‌گیریم. سپس دو دسته‌ای که کمترین فاصله را دارند می‌یابیم و اگر این فاصله از مقدار از قبل تعیین شده‌ای کمتر بود آن دو دسته را

ادغـام می‌کنیم و محددا دو دسته دیگر انتخاب می‌کنیم. این کار را آن قدر تکرار می‌کنیم تا الگوریتم همگرا شود.

بـرای یافتن فاصله دسته‌هایی که دارای چندین عضو هستند فاصـله تمـام اعضـای هـر دسته را با تمام عضوهای دسته دیگر محاسـبه مـی‌کنیم و سـپس از آنهـا میانـگن می‌گیریم. در این الگوریـتم از رابطه دیورژانس، به شرح زیر برای به دست آوردن فاصله بین دسته‌ها استفاده کرده‌ایم[۱۷].

$$d(S_1,S_2)=\left(\frac{1}{N}\sum_{i=1}^{N}\frac{(\mu_{1i}-\mu_{2i})^2}{\sigma_{1i}\sigma_{2i}}\right)^{\frac{1}{2}} \qquad (٦)$$

۳-۲- تطبیق MLLR

در این بخش به بررسی روشMLLR و روابط مورد استفاده در آن مـی‌پردازیم. جزییات بیشتر در این خصوص در مرجع[۳] یافت می‌شود.

هرعنصـر مخلـوط گوسـین کـه نماینده یک حالت در مدل HMM تـک گوسـین اسـت توسـط بـردار مـیانه μ_j و ماتـریس کـــوواریانس Σ_j مشـــخص مـــی‌شـــود. در روش MLLR میانگیـنهای مدل SI به میانگینهای نا معلوم اما قابل تخمین مدل تطبـیق یافـته، یعنـی $\hat{\mu}_j$، توسـط یـک تـبدیل رگرسیون خطی نگاشت می‌شوند. این نگاشت از داده‌های آموزشی به طریق زیر به دست می‌آید:

$$\hat{\mu}_j = w_j.\nu_j \qquad (٧)$$

کـه در آن w_j یـک ماتـریس تبدیل$n \times (n+1)$ است که n بعد بـردارهای میانگـین بـوده و ν_j بـردار میانگین گسترش یافته به صورت $\nu_j=[1,\mu_{j1},...,\mu_{jn}]'$ است.

اگـر بـرای هـر مـدل یـک w به دست آید، آنگاه این روش تطبـیق بـه همان روش استاندارد Baum-Welch تبدیل می‌شود. امـا بـرای ایـنکه بـا اسـتفاده از داده کم تمام مدلها تطبیق یابند، کلاسهای رگرسیونی تعریف می‌شوند که شامل پارامترهای مدل گـره خـورده‌انـد. بـرای هر کلاس یک ماتریس تبدیل به دست مـی‌آیـد کـه از داده‌های آموزشی متعلق به آن کلاس مشتق شده اسـت. پـس از به دست آمدن میانگینهای جدید($\hat{\mu}_j$) عملیات بازشناسـی توسـط این میانگینها انجام می‌گیرد. تبدیلها به نحوی محاسـبه مـی‌شـوند کـه درسـتنمایی داده‌های آموزشی را بیشینه کنـند. اگر T فریم از داده آموزشی $O=o_1,o_2,...,o_T$ موجود باشـد احـتمال حضور در حالت j ام در زمان t در حالی که داده O تولید شده است را با $\gamma_j(t)$ نمایش می‌دهیم.

$$\gamma_j(t)=\frac{f(O,\theta_t=j|\lambda)}{f(O|\lambda)} \qquad (٨)$$

$\gamma_j(t)$ از الگوریـتم Forward-Backward بـه سـادگی بـرای داده‌هـای آموزشـی محاسـبه مـی‌شـود. بـا فرض اینکه ماتریس کـوواریانس قطـری باشـد و ماتـریس تبدیل w_j بین R گوسی، $j_1 \cdots j_R$، مشـترک باشـد، ماتـریس تـبدیل w_j سـتون به ستون از رابطـه $w_i=G_i^{-1}z_i$ بـه دسـت می‌آید. در این رابطه z_i ستون i ام از ماتریس Z است که به صورت

$$Z=\sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{jr}(t)\Sigma_{jr}^{-1}o_t\nu_{jr}' \qquad (٩)$$

است و G_i نیز به صورت

$$G_i=\sum_{r=1}^{R}c_{ii}^{(r)}\nu_{jr}\nu_{jr}' \qquad (١٠)$$

نوشـته مـی‌شـود کـه $c_{ii}^{(r)}$ عنصـر قطـری i ام از ماتریس قطری کـوواریانس مقـیاس شـده گوسـی r ام در کـلاس گوسیهای R است که از رابطه

$$C^{(r)}=\sum_{t=1}^{T}\gamma_{jr}(t)\Sigma_{jr}^{-1} \qquad (١١)$$

بـه دسـت می‌آید. تازه سازی پارامترها به کمک روابط فوق یک مـرحله از تطبیق MLLR است. اگر تغییرات در پارامترها با یک مـرحله تطبیق به نحوی باشد که احتمال پیشینه اشغال آن حالت افـزایش یابـد، آن گـاه با تکرار روش MLLR میزان درستنمایی مدل افزایش می‌یابد.

٤- بهبود مدلهای پیشینه در تخمین MAP مبتنی بر تبدیل MLLR (MPIM)[۲۰]

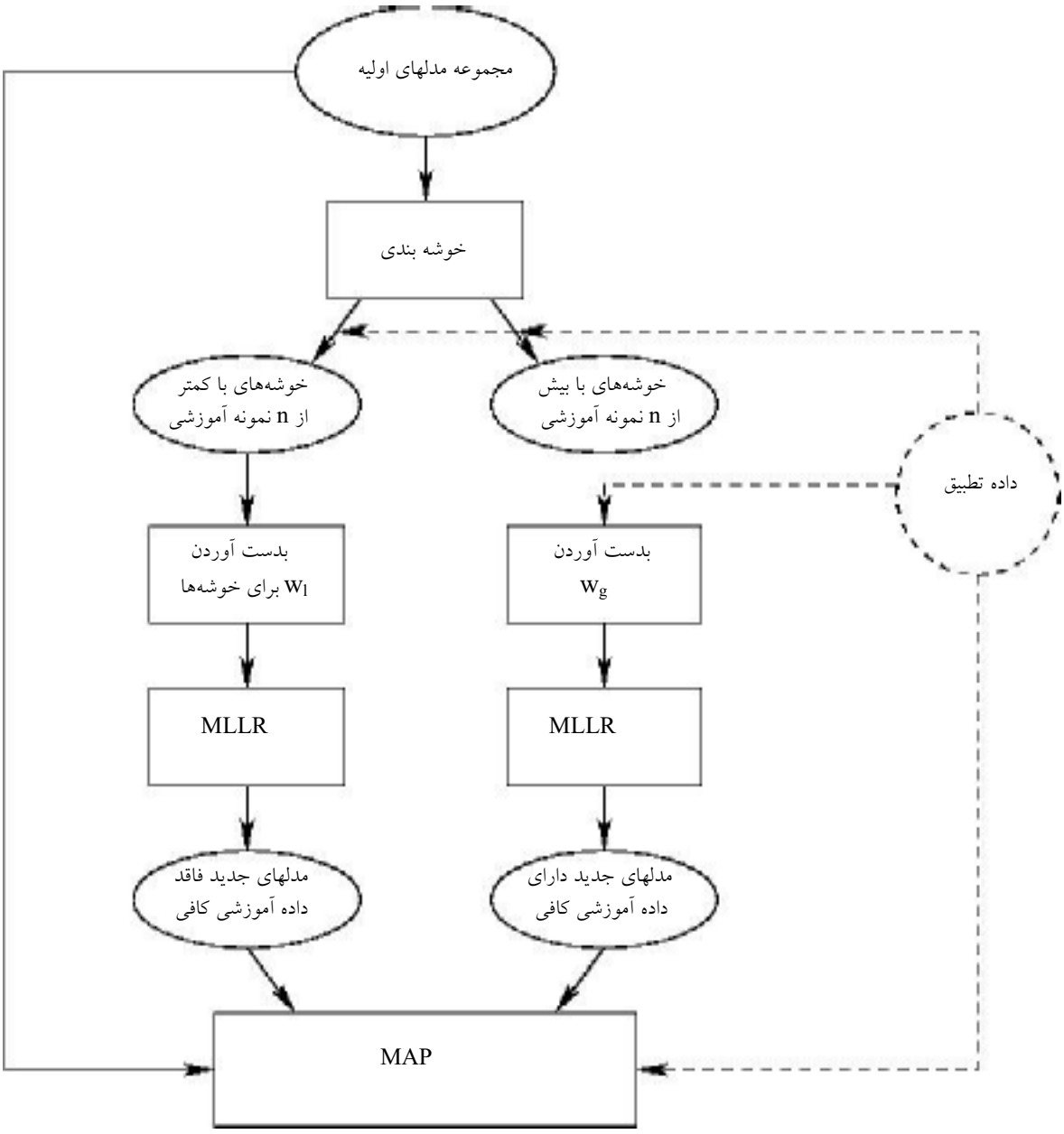
علی‌رغم استفاده از دو شیوه MAP و MLLR، روش پیشنهادی MPIM، از نظر نحوه استفاده از این دو شیوه، کاملاً با روش MAPLR [۱۸] متفاوت است. در روش MAPLR اطلاعات مربوط به توزیع احتمال پیشینه ماتریسهای تبدیل مورد استفاده قرار می‌گیرد، به این معنی که اقدام به تخمین ماتریسهای انتقال MLLR به همراه پارامترهای مدل HMM، در چارچوب تخمین MAP، می‌شود، حال آنکه در روش پیشنهادی MPIM از مدلهای تطبیق یافته به کمک ماتریس تبدیل MLLR برای بهبود مدلهای پیشینه در تخمین MAP استفاده می‌شود.

این روش تنها بر روی میانگینهای توزیعهای گوسین در مدلها اعمال می‌شود. شیوه کار به این صورت است که ابتدا مدلهای اولیه بر اساس شباهت و نزدیکی گوسین‌ها، که با معیار دیورژانس سنجیده می‌شود، خوشه بندی می‌شوند. سپس برای هر خوشه به دست آمده به شرط وجود حداقل سه نمونه داده آموزشی در آن خوشه یک ماتریس تبدیل w_l به کمک روش MLLR به دست می‌آید. یک ماتریس w_g (ماتریس تبدیل کلی) نیز با استفاده از تمام داده‌های آموزشی تشکیل می‌شود. حال با استفاده از ماتریسهای تبدیل w_l به دست آمده بردارهای میانگین جدید را برای توزیعهای همه مدلها حساب می‌کنیم. برای به دست آوردن میانگینهای جدید توزیعها برای مدلهای خوشه‌هایی که داده آموزشی کافی ندارند، از w_g استفاده می‌کنیم. با این حال، میانگینهای به دست آمده توسط ماتریسهای تبدیل مستقیما جایگزین میانگینهای قبلی نمی‌شوند. در اینجا به کمک تخمین MAP مدلهایی که داده آموزشی متناظر آنها وجود دارد توسط آن داده‌ها آموزش می‌بینند و مدلهایی که داده آموزشی متناظر ندارند از مدلهای تبدیل یافته برای بهبود مدلهای پیشینه در تخمین MAP استفاده می‌کنند.

به کمک این روش با هر تعداد داده آموزشی ورودی، هر چند کم، تمام مدلهای موجود به نوعی آموزش دیده و تغییر می‌یابند و چون از تخمین MAP برای تطبیق استفاده می‌کنند مدلهای به دست آمده هیچ گاه اشباع نمی‌شوند و همچنین، آنچنان که در بخش بعد خواهیم دید با دادگان کم، دقت بازشناسی به نحو چشمگیری افزایش می‌یابد. نمودار جعبه‌ای (۱) مراحل اجرای این الگوریتم را نشان می دهد.

اساس کار به این صورت است که در رابطه MAP وزنی، به جای استفاده از مدل آموزش دیده توسط روش MAP از مدل تبدیل یافته توسط ماتریس تبدیل به دست آمده از روش MLLR استفاده می‌گردد. بر این اساس، فرض می‌شود که $\hat{\mu}_j = w_j \times v_j$ بردار میانگین حالت jام تبدیل یافته توسط روش MLLR باشد که در آن w_j تبدیل متناظر با آن حالت و v_j بردار میانگین گسترش یافته مدل SI است. همچنین اگر میانگین مدل SI متناظر با آن حالت با $\hat{\mu}_j^{(q)}$ نمایش داده شود، آن گاه می‌توان روش وزنی MAP و MLLR را به صورت رابطه (٤) برای تطبیق میانگینهای حالت jام نوشت که در آن $\mu_j^{(q)}$ میانگین تطبیق یافته با روش MAP و MLLR است. در این حالت می‌توان پارامتر φ را مطابق با شیوه به کار رفته در روش MAP وزنی به صورت رابطه (٥) نوشت. در این رابطه وزن φ نشان دهنده میزان مشارکت هر یک از مدلهای SI و مدل تطبیق یافته توسط روش MLLR در تعیین اطلاعات پیشینه مدل است. در این رابطه هر چه مقدار φ به سمت یک نزدیکتر شود، مدل به سمت تطبیق MAP معمولی میل می‌کند و نقش مدل اولیه SI در تخمین MAP بارزتر می‌شود. اگر φ به سمت صفر میل کند، میانگین تبدیل یافته توسط روش MLLR به عنوان میانگین احتمال پیشینه مدل در تخمین MAP خودنمایی می‌کند. عملاً با زیاد شدن مقدار φ عملیات تطبیق کند می‌شود و نیاز به داده آموزشی بیشتری دارد ولی در عوض به اشباع نمی‌رود و با افزایش دادگان تطبیقی قابلیت بهینه شدن مدل را فراهم می آورد. این خصوصیت دقیقا معادل خصوصیت روش MAP است.

در مقابل با کاهش میزان φ می‌توانیم نقش تبدیل MLLR را در تطبیق افزایش دهیم. یعنی با میزان داده کمتر آموزشی سریعاً مدلها به سمت مدل SD میل می‌کنند و نرخ بازشناسی نسبت به تخمین MAP افزایش قابل ملاحظه‌ای می‌یابد اما سیستم به



شکل۱- روش تطبیق پیشنهادی (MPIM).

اشباع می‌رود. در عمل شروع به اشباع رفتن سیستم از صد جمله آموزشی به بعد است که برای عملیات تطبیق سریع بسیار مناسب است.

۵- پیاده سازی و نتایج

سیستم SI مورد استفاده برای تطبیق گوینده سیستم HMM سه آوایی است که با استفاده از دادگان *فارس دات*[۲۱] آموزش دیده است[۱۵]. برای آموزش این سیستم از حدود ۱۸۰۰ جمله که از گویندگان دارای لهجه تهرانی در این دادگان به دست آمده استفاده شده است. هر گوینده در حدود ۲۰ جمله را ادا کرده که دو جمله بین تمام گوینده‌ها مشترک است. در این سیستم، ابتدا سیگنال گفتار از نرخ اصلی نمونه برداری ٤٤/١ kHz به ١٦ kHz کاهش نرخ داده می‌شود. پس از انجام عملیات پردازشی ۱۲ ضریب LP کپستروم، ۱۲ ضریب دلتا

کپستروم و یک ضریب لگاریتم انرژی از هر فریم ۲۵ میلی ثانیه‌ای استخراج می‌شود. برای هر آوا از سه حالت استفاده شده است که هر حالت از یک گوسین تشکیل یافته است. لازم به اشاره است که سیستم فوق به علت قابل دسترس بودن مورد استفاده واقع شده و بدیهی است که بخش‌های مختلف آن و بویژه بخش استخراج ویژگی‌ها در این سیستم بازشناسی گفتار می‌توانند از شیوه‌های مناسبتری بهره ببرند.

برای آموزش این سیستم از گره زدن حالتها استفاده شده است. پس از اعمال الگوریتم خوشه بندی ۸۰ خوشه به دست آمد. برای آزمون الگوریتمها از سه گوینده مرد با لهجه معمولی استفاده شد. از هر گوینده ۳۰۰ جمله ضبط شد که در دو گروه آموزش و تطبیق (۲۰۰ جمله) و آزمایش (۱۰۰ جمله) طبقه‌بندی شد. سپس نتایج بازشناسی برای هر گوینده توسط سیستم SI و تطبیق با روشهای MAP وMLLR و روش تلفیقی پیشنهادی MPIM به ازای ۵ و۲۰ و۵۰ و۱۰۰ جمله آموزشی به دست آمد. لازم به اشاره است که استفاده از رابطه(۵) در تعیین پارامتر φ به جواب قابل قبولی منجر نگردید. علت این امر احتمالاً آن است که در روشMLLR مقادیر بدست آمده برای μ'_j ممکن است با مقدار قبلی مدلSI یعنی $\hat{\mu}_j^{(q)}$ تفاوت زیادی داشته باشند. در نتیجه استفاده از وزنهای ثابت به جای وزنهای متغیر آزمایش گردید که به نتایج بسیار خوبی منتهی شد. مقدار بهینه‌ای که برای φ بدست آمده و کلیه آزمایشات براساس آن می‌باشد مقدار $\varphi = 0.3$ است. در جدول(۱) متوسط نتایج بازشناسی برای سه گوینده مشاهده می‌شود. این نتایج با استفاده از گرامر زوج کلمه به دست آمده‌اند. این در حالی است که مدلSD دقت بازشناسی معادل۹۶٪ را داراست که با ۲۰۰ جمله آموزشی حاصل شده است.

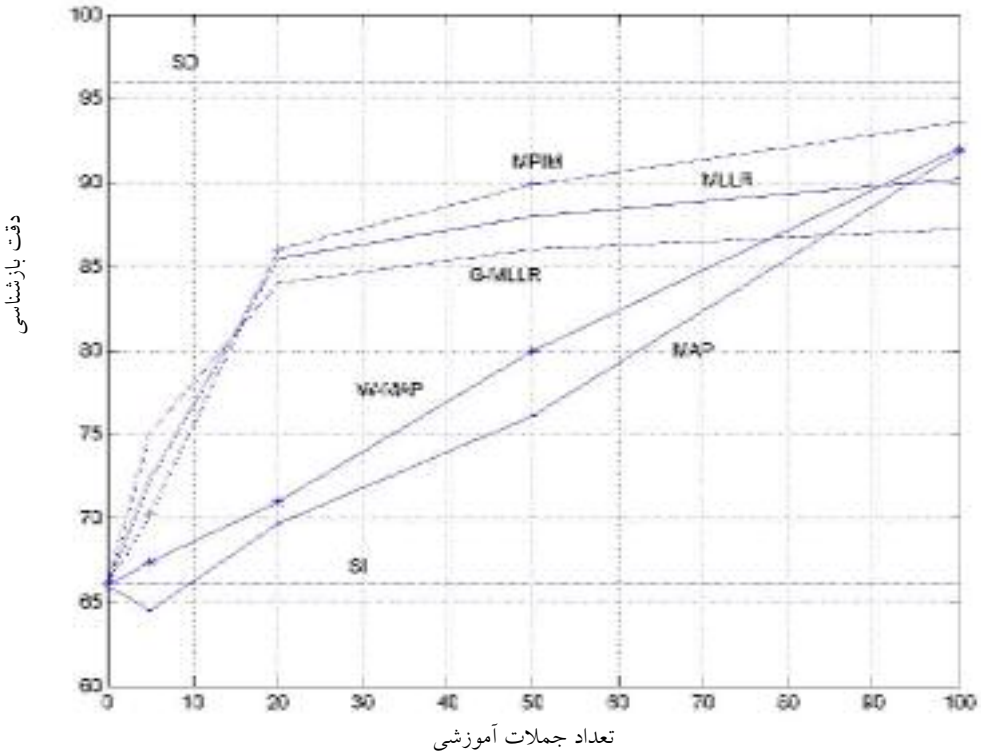
نتایج ارائه شده در جدول(۱) حاوی نکات جالبی‌اند. تطبیق MLLR به خوبی با میزان کم داده تطبیق توانسته است بهبودی مناسبی کسب کند، اگرچه با افزایش تعداد جملات، این میزان بهبودی نیز افزایش یافته ولی این افزایش چندان زیاد نیست. برایMAP با تعداد بسیار کم جملات تطبیق، حتی کاهش دقت بازشناسی نیز دیده می‌شود که ناشی از پراکندگی زیاد داده آموزشی و تخمین نادرست ناشی از آن برای برخی پارامترهاست. در مقابل با افزایش داده تطبیق، دقت بازشناسی افزایش می‌یابد و با ۱۰۰ جمله تطبیق، دقت نسبت بهMLLR نیز فزونی می‌گیرد که در واقع نشان دهنده خاصیتMAP در گرایش تدریجی تخمین به سمت تخمینML با افزایش داده است. نکته قابل توجه در این جدول آن است که روشMPIM با ۵۰ جمله تطبیق حدود۸۰٪ و با ۱۰۰ جمله تطبیق حدود ۹۲٪ فاصله بین دو سیستمSI وSD را پیموده، حال آنکه این مقادیر برای روشMAP به ترتیب حدود ۳٤٪ و ۸٦٪ و برای روش MLLR به ترتیب حدود ۷۳٪ و ۸۱٪ است.

شکل (۲) روند بهبود نرخ بازشناسی را برای چند روش تطبیق گوینده با افزایش تعداد جملات تطبیق نشان می دهد. در اینجا نتایج برای روشهای MAP، MAP وزنی، MLLR ، MLLR عمومیت یافته (G–MLLR)[۲۲] و MPIM نشان داده شده اند. همچنانکه ملاحظه می‌شود، روشهای مبتنی بر MLLR، در مقایسه با روشهای مبتنی بر MAP، با تعداد جملات کم، به بهبودی قابل ملاحظه ای در نرخ بازشناسی دست می یابند، حال آنکه این روشها، با افزایش تعداد جملات تطبیق، کارایی خود را از دست داده و نمی‌توانند همانند MAP مرتباً با افزایش داده آموزشی به سمت توانایی یک سیستم وابسته به گوینده گرایش یابند. در مقابل، روش MPIM از هر دو مزیت برخوردار بوده و در حجم کم داده از قابلیتهای MLLR و در حجم زیاد داده از تواناییهای MAP بهره می‌برد.

در مجموع می‌توان گفت که روش تلفیقی پیشنهادی با توجه به استفاده از مزیتهای MLLR در اعمال تابع انتقال عمومی و MAP درآموزش مناسب، به دقتی بیش از هر دو روش فوق در جملات تطبیقی متوسط تا زیاد دست می‌یابد. لازم به اشاره است که کاهش دقت MAP در ۵ جمله تطبیقی، تاثیر منفی خود را برروی روش تلفیقی پیشنهادی نیز گذاشته است. با اعمال شرط حداقل نمونه دیده شده برای تطبیق پارامترها در MAP می‌توان این مشکل را نیز تا حد زیادی مرتفع نمود.

جدول۱- میانگین نتایج به دست آمده از اجرای الگوریتمها

روش	۵ جمله	۲۰ جمله	۵۰ جمله	۱۰۰ جمله
SI	٦٦/٠			
SD	۹٦/٠			
MLLR	۷۲/۳	۸۵/۵	۸۸/٠	۹٠/۲
MAP	٦٤/۵	٦۹/۷	۷٦/۱	۹۱/۸
MPIM	۷٠/۲	۸٦/٠	۸۹/۹	۹۳/٦



شکل۲- مقایسه بهبودی در نرخ بازشناسی با افزایش داده تطبیق برای روش MPIM در مقایسه با چندین روش دیگر تطبیق گوینده.

مسئله پیچیدگی آلگوریتم تطبیق نیز در اینجا قابل اشاره است. بدست آوردن دو گونه ماتریس تبدیل و اعمال دو گانهMLLR ، اگرچه حجم محاسبات را به دقیقاً دو برابر تبدیل افزایش نمی‌دهد، ولی در مجموع باعث افزایش نسبی حجم محاسبات می‌گردد. افزوده شدنMAP به این مجموعه نیز به نوبه خود موجب افزایش میزان محاسبات می‌گردد. در مجموع می‌توان گفت که این آلگوریتم، نسبت به هر دو روشMAP وMLLR از حجم محاسبات بیشتری برخوردار است. با این همه، این مطلب نمی‌تواند در عمل چندان مشکل آفرین باشد چرا که پراستفاده‌ترین زمینه برای اینگونه آلگوریتم‌ها، تطبیق گوینده با نظارت یکباره است که عموماً به صورتoff-line انجام می‌گیرد و این افزایش حجم محاسبات مشکل عمده‌ای برای آن تلقی نمی‌گردد.

٦- نتیجه‌گیری و پیشنهادات

در این مقاله ضمن طرح روشهای مختلف تطبیق گوینده نظیر MAP و MLLR و روش تلفیقی پیشنهادی MPIM، نتایج پیاده سازی و اجرا ارائه گردید. نتایج به دست آمده حکایت از

افـزایش دقـت بازشناسـی با استفاده از این روش دارد به طوری کـه در تعـداد جمـلات متوسـط بـه بـالا نسبت به هر دو روش MLLR و MAP بهـبودی حاصل شده است. برای بهبود هر چه بیشـتر نـتایج و ادامـه کـار اسـتفاده از حالتهایـی با چند مخلوط گوسـین بـه جـای حالـتهای تک مخلوطی، پیاده‌سازی الگوریتم خوشـه‌بـندی دقیقتـر و اسـتفاده از سـاختار درختی در خوشه‌بندی همانند روشی که درSMAP پیاده‌سازی شده است پیشنهاد می‌شود.

واژه نامه

1. hidden Markov models
2. speaker dependent
3. speaker independent
4. speaker pooling
5. maximum likelihood
6. Bayes
7. posterior
8. maximum likelihood linear regression
9. vector field smoothing
10. stochastic matching
11. extended MAP
12. structural MAP
13. regression-based model prediction
14. maximum *a posteriori* linear regression
15. triphone
16. Bayesian
17. prior
18. sparse training data
19. context-dependent
20. MLLR-based prior improvement for MAP estimation
21. FarsDat
22. generalized MLLR

مراجع

1. Lee, C-H., "On Stochastic Feature and Model Compensation Approaches to Robust Speech Recognition," *Speech Communication*, vol. 25, pp. 29-47, 1998.
2. Gauvain, J-L. and Lee, C-H., "Maximum *a Posteriori* Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 291-298, 1994.
3. Leggetter, C.J. and Woodland, P.C., "Maximum Likehood Linear Regression for Speaker Adaptation of Continuous-Density Hidden Markov Models," *Comput. Speech Lang*., vol. 9, pp. 171-185, 1995.
4. Sankar, A. and Lee, C-H., "A maximum Likelihood Approach to Stochastic Matching for Robust Speech Recognition," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 190-202, May 1996.
5. Lee, C-H., Lin, C-H. and Juang, B-H., "A Study on Speaker Adaptation of the Parameters of Continuous-Density Hidden Markov Models," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-39, pp. 806-814, Apr. 1991.
6. Ohkura, K., Sugiyama, M., and Sagayama, S., "Speaker Adaptation Based on Transfer-Vector-Field Smoothing with Continuous Mixture Density HMMs," in *Proc. Int. Conf. Spoken Language Processing '92*, pp. 369-372, 1992.
7. Zavaliagkos, G., and Schwartz, R., "Maximum *a Posteriori* Adaptation for Large-Scale HMM Recognizers," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing'95*, pp.725-728, Detroit, 1995.
8. Stern, R.M. and Lasry, M.J., " Dynamic Speaker Adaptation for Feature-Based Isolated Word Recognition," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-35, pp.751-763, June 1997.
9. Shinoda, K., and Lee, C-H., "A Structural Bayes Approach to Speaker Adaptation," *IEEE Transactions on Speech and Audio Processing*, vol. 9, No. 3, March 2001.
10. Ahadi, S.M., and Woodland, P.C., "Combined Bayesian and Predictive Techniques for Rapid Speaker Adaptation of Continuous Density Hidden Markov Models," *Computer Speech and Language*, vol. 11, No. 3, July 1997.
11. Lee, C-H., "On-Line Adaptive Learning of the Correlated Continuous-Density Hidden Markov Model for Speech Recognition," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 386-397, July 1998.
12. Siohan, O., Chesta, C., and Lee, C-H., "Hidden Markov Model Adaptation Using Maximum *a Posteriori* Linear Regression," In *Proc. Workshop Robust Methods Speech Recognition in Adverse Conditions*, Tampere, Finland, pp. 147-150, May 1999.
13. Siohan, O., Chesta, C., and Lee, C-H., "Joint Maximum *A Posteriori* Adaptation of Transformation and HMM Parameters". *IEEE Transactions on Speech and Audio Processing*, vol. 9, No. 4, May 2001.
14. Huo, Q., and Chan, C., "Bayesian Adaptive Learning of Parameters of Hidden Markov Model for Speech Recognition," Technical Report, Department of Computer Science, University of Hong Kong, 1992.
15. شـمس، س. ح. و احدی، س. م. "روش آموزش وزنی وابسته

به متن در گفتار پیوسته فارسی"، *نشریه علمی پژوهشی /میرکبیر*، شماره ۵۳، سال ۱٤،ص ۹۲-۸۰، زمستان ۱۳۸۱.

16. Ahadi, S.M., "Reduced Context Sensitivity in Persian Speech Recognition via Syllable Modeling," in *Proc. SST-2000*, pp. 492-497, Canberra.
17. HTK Hidden Markov Model Toolkit ver. 3.1, Reference Manual, Cambridge University Engineering Department, p. 76, 2001.
18. Siohan, O., Myrvoll, T-A. and Lee, C-H., "Structural Maximum *a Posteriori* Linear Regression for Fast HMM Adaptation" In *Workshop on Automatic Speech Recognition: Challenges for the new Millenium*, Paris, France, Sept.2000, ISCA ITRW ASR2000.