

مدل سازی سه‌واجی به عنوان واحد آوایی در سیستم‌های بازشناسی گفتار پیوسته فارسی مبتنی بر مدل مخفی مارکوف

باقر باباعلی	محمد بحرانی	حسین صامتی	نسیبه نصیری
آزمایشگاه پردازش گفتار	آزمایشگاه پردازش گفتار	استادیار و عضو هیئت علمی	آزمایشگاه پردازش گفتار
دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف	دانشگاه صنعتی شریف	دانشگاه صنعتی شریف	دانشگاه صنعتی شریف
babaali@ce.sharif.edu	bahrani@ce.sharif.edu	sameti@sharif.edu	n_nasiri@ce.sharif.edu

آوایی وابسته به متن عملاً غیر ممکن است.

به منظور فائق آمدن بر مشکل کمبود داده آموزشی برای مدل‌های وابسته به متن، معمولاً سه‌واجی‌های مشابه (یا حالت‌های مربوط به مدل آن‌ها) به هم گره زده می‌شوند. دو روش مشهور برای گره‌زدن حالت‌ها در مدل‌های مبتنی بر سه‌واجی، روش مشتق شده از داده (دسته‌بندی پایین به بالا) [۲]، [۸] و روش مبتنی بر درخت تصمیم‌گیری (دسته‌بندی بالا به پایین) [۲]، [۳]، [۴] می‌باشد که ما در این مقاله، از روش دوم که رایج‌تر است استفاده کرده‌ایم. مشکل ذاتی اعمال راهکارهای مشتق شده از داده به متن‌های جدید باعث شده است که روش‌های مبتنی بر درخت تصمیم‌گیری مورد توجه قرار گیرند و کاربرد بیشتری داشته باشند.

در این مقاله ابتدا راجع به گره زدن حالت‌ها به دو روش مشتق شده از داده و مبتنی بر درخت تصمیم‌گیری بحث خواهد شد سپس دادگان‌های گفتاری به کار برده شده برای آموزش و ارزیابی معرفی خواهند شد و در ادامه آزمایش‌های انجام شده و نتایج حاصل از آنها آورده می‌شود.

۲- روش دسته‌بندی مشتق شده از داده و مبتنی بر درخت تصمیم‌گیری

رویه کلی آموزش مدل‌های مبتنی بر سه‌واجی بدین صورت است که ابتدا همه سه‌واجی‌های موجود در دادگان آموزشی ابتدا با استفاده از همان تعداد داده آموزشی موجود، آموزش داده می‌شوند سپس حالت‌های مربوط به سه‌واجی‌های با بافت مشابه مورد دسته‌بندی قرار می‌گیرند و حالت‌های شبیه‌تر در یک دسته قرار می‌گیرند. در نهایت حالت‌های موجود در یک دسته به عنوان یک حالت واحد در نظر گرفته می‌شوند و به اصطلاح به هم گره زده می‌شوند. در مرحله بعد حالت‌های گره زده شده مورد آموزش مجدد قرار می‌گیرند.

اگر حالت‌های مربوط به سه‌واجی‌های مشابه بر اساس شباهت‌هایی با هم ترکیب^۱ شوند، به این روش دسته‌بندی پایین به بالا گفته می‌شود. اگر همه^۲ سه‌واجی‌ها ابتدا در یک دسته قرار داشته باشند و بر اساس شباهت‌هایی به صورت تکرارپذیر شکسته^۳ شوند و در دسته‌های مختلفی قرار بگیرند، به این روش دسته‌بندی بالا به پایین گفته می‌شود.

چکیده: برای سیستم‌های بازشناسی گفتار پیوسته معمولاً استفاده از واحد آوایی سه‌واجی جهت نیل به عملکرد بهتر غیر قابل اجتناب است. مشکل واحدهای آوایی سه‌واجی تعداد بالای آنها و در نتیجه نیاز به دادگان آموزشی بسیار بزرگ برای آموزش تعداد زیاد حالت‌های مدل می‌باشد. برای حل این مشکل و اختصاص داده کافی برای آموزش هر حالت از مدل مخفی مارکوف، روش مرسوم استفاده از متد گره‌زدن حالت‌ها است. در این مقاله، مراحل گره‌زدن حالت‌های مربوط به سه‌واجی‌ها در روند آموزش یک سیستم بازشناسی گفتار پیوسته فارسی طراحی و اجرا شده است و براساس نتایج و عملکرد سیستم بازشناسی، تعداد بهینه حالت‌های گره زده شده (زنون‌ها) و پارامترهایی مانند تعداد گوسی‌های هر حالت برای مجموعه دادگان‌های فارسی‌دات بزرگ و کوچک گزارش شده است. میزان کاهش خطا در سیستم بازشناسی گفتار فارسی پس از به‌کارگیری سه‌واجی‌ها حدود ۱۱٪ بوده است.

واژه‌های کلیدی: سه‌واجی، گره زدن حالت‌ها، درخت تصمیم‌گیری، بازشناسی گفتار پیوسته

۱- مقدمه

در بازشناسی گفتار با واژگان بزرگ، مدل‌سازی آواها به صورت مستقل از متن معمولاً دقت و کارایی پایینی را نتیجه می‌دهد زیرا نمی‌تواند اثرات هم آوایی^۱ را در بازشناسی گفتار وارد کند. به منظور فائق آمدن بر این مشکل، مدل‌های آوایی وابسته به متن^۲ در بازشناسی گفتار پیوسته به کار برده می‌شود. در مدل‌سازی وابسته به متن دو واحد آوایی مشابه در گفتار اگر دارای همسایه‌های متفاوت باشند، هر کدام یک مدل جداگانه را تشکیل می‌دهند بنابراین مدل‌سازی دقیق و جزئی و انعکاس اثر هم آوایی در این روش امکان‌پذیر می‌باشد.

در این مقاله ما از سه‌واجی‌ها برای مدل‌سازی وابسته به متن استفاده کرده‌ایم. به علت زیاد بودن تعداد انواع سه‌واجی‌ها در یک زبان، برای آموزش مدل‌های مبتنی بر سه‌واجی به مقدار زیادی داده^۳ گفتاری برای آموزش نیاز داریم. در واقع آموزش مدل‌های وابسته به متن قابل اطمینان مشکل می‌باشد، زیرا تضمین دسترسی به چنین داده^۴ آموزشی بسیار سخت است. به همین دلیل تخمین قطعی پارامترهای مدل‌های

۳-۱- دو کلاس i و j را که فاصله میان گروهی آن‌ها $d(i,j)$ مینیمم باشد، پیدا می‌کنیم.

۳-۲- تازمانی که $d(i,j) < TC$ باشد:

۳-۲-۱- کلاس‌های i و j را با هم ترکیب می‌کنیم.

۳-۲-۲- دو کلاس i و j را که فاصله میان گروهی آن‌ها

$d(i,j)$ مینیمم باشد، پیدا می‌کنیم.

۳-۳- برای کلاس‌های به دست آمده تعداد داده‌های آموزشی را محاسبه کرده و اگر این تعداد برای یک کلاس از حد خاصی

کمتر بود آن را با نزدیک‌ترین کلاس ترکیب می‌کنیم.

۴- حالت‌های موجود در کلاس‌های نهایی را با هم گره می‌زنیم.

۵- به طور متوالی تعداد گوسی‌ها را در هر حالت افزایش می‌دهیم و حالت‌ها را دوباره آموزش می‌دهیم تا زمانی که به تعداد گوسی

دلخواه برسیم.

در مرحله ۵ از الگوریتم بالا به صورت زیر عمل می‌کنیم:

با داشتن m گوسی در هر حالت، توزیع گوسی با بزرگترین وزن را پیدا کرده و آن را به دو گوسی تقسیم می‌کنیم. وزن هر گوسی جدید برابر نصف وزن قبلی می‌باشد. میانگین گوسی‌های جدید با اضافه و کم کردن ۰،۲ از انحراف معیار گوسی اولیه به میانگین آن گوسی به دست می‌آید. سپس همه پارامترهای مدل با دو تکرار الگوریتم بام-ولج به روز می‌شوند و مراحل بالا تکرار می‌گردد.

در روش دوم که در این مقاله نیز استفاده شده است، رویه دسته‌بندی حالت‌ها با استفاده از درخت تصمیم‌گیری صورت می‌گیرد. در این روش تمام حالت‌های مربوط به بافت‌های مشابه و همچنین مکان مشابه در مدل، در ریشه درخت قرار می‌گیرند (منظور از بافت مشابه، بافت‌هایی است که در آن واج میانی سه واجی یکسان می‌باشد). سپس یک سری پرسش در مورد بافت چپ و راست حالت‌ها مطرح می‌شود با توجه به پرسش مطرح شده، حالت‌ها به دو دسته تقسیم می‌شوند که این دو دسته دو گره جدید در درخت ایجاد می‌کنند. از بین پرسش‌های مطرح شده پرسشی در نهایت انتخاب می‌شود که دو دسته حاصل از آن بیشترین افزایش در میزان شباهت^۱ را نسبت به دسته اولیه نتیجه بدهند. این روند ادامه پیدا می‌کند و درخت به تدریج رشد می‌کند تا در نهایت به جایی برسیم که افزایش در میزان شباهت از حد خاصی کمتر شود و یا داده آموزشی برای حالت‌های موجود در گره‌های درخت به حد کافی نباشد. پس از تشکیل درخت، هر برگ از درخت نشان‌دهنده یک کلاس از حالت‌ها می‌باشد. حالت‌های موجود در هر کلاس به هم گره زده می‌شوند و یک حالت را تشکیل می‌دهند.

رویه کلی آموزش سه‌واجی‌ها با استفاده از درخت تصمیم‌گیری همانند رویه ذکر شده در روش مشتق شده از داده می‌باشد با این تفاوت که به جای مرحله ۳، مراحل زیر را خواهیم داشت:

۳-۱- برای هر سه واجی با واج میانی مشابه، حالت‌های متناظر را

در ریشه درخت تصمیم‌گیری قرار می‌دهیم.

۳-۲- لگاریتم شباهت داده‌های آموزشی را در ریشه درخت

در روش پایین به بالا، حالت‌ها با استفاده از الگوریتم‌های خوشه‌بندی، مورد دسته‌بندی قرار می‌گیرند. برای دسته‌بندی به روش زیر عمل می‌کنیم: ابتدا هر حالت، جداگانه به عنوان یک کلاس در نظر گرفته می‌شود. سپس همه کلاس‌ها دو به دو امتحان می‌شوند و دو کلاسی که مینیمم فاصله میان گروهی^۲ را داشته باشند با هم ترکیب می‌شوند. ترکیب کردن تازمانی انجام می‌گیرد که مینیمم فاصله میان گروهی بین کلاس‌ها از حد خاصی بزرگتر شود و همچنین داده آموزشی کافی برای هر کلاس موجود باشد.

فاصله میان گروهی دو کلاس در واقع نشان‌دهنده جذر واگرایی بین توزیع‌های گوسی مربوط به دو کلاس می‌باشد. برای حالتی که ماتریس کوواریانس به صورت قطری است این فاصله با روش زیر محاسبه می‌گردد [۸]:

$$d(i,j) = \left[\frac{1}{V} \sum_{k=1}^V \left(\frac{\sigma_{ik}^2}{\sigma_{jk}^2} + \frac{\sigma_{jk}^2}{\sigma_{ik}^2} - 2 + \left(\frac{1}{\sigma_{ik}^2} + \frac{1}{\sigma_{jk}^2} \right) (\mu_{ik} - \mu_{jk})^2 \right) \right]^{1/2}$$

این فاصله را می‌توان با رابطه ساده زیر نیز محاسبه نمود [۸]:

$$d(i,j) = \left[\frac{1}{V} \sum_{k=1}^V \left(\frac{(\mu_{ik} - \mu_{jk})^2}{\sqrt{\sigma_{ik}^2 \sigma_{jk}^2}} \right) \right]^{1/2}$$

در روابط بالا μ_i و σ_i به ترتیب بردار میانگین و قطر ماتریس کوواریانس مربوط به توزیع گوسی کلاس i می‌باشند. V نیز بُعد بردارها می‌باشد. توزیع گوسی هر کلاس با توجه به توزیع‌های گوسی حالت‌های موجود در آن کلاس محاسبه می‌شود. برای هر کلاس با مجموعه حالت‌های S داریم [۳]:

$$\mu(S) = \frac{\sum_s \gamma_s \mu_s}{\sum_s \gamma_s}$$

$$\Sigma(S) = \frac{\sum_s [\gamma_s (\Sigma_s + \mu_s \mu_s^T)]}{\sum_s \gamma_s} - \left(\frac{\sum_s \gamma_s \mu_s}{\sum_s \gamma_s} \right) \left(\frac{\sum_s \gamma_s \mu_s}{\sum_s \gamma_s} \right)^T$$

در روابط بالا $\mu(S)$ و $\Sigma(S)$ به ترتیب میانگین و کوواریانس کل کلاس می‌باشد و μ_s و Σ_s میانگین و کوواریانس حالت s از کلاس مورد نظر می‌باشند. γ_s هم تعداد داده‌های آموزشی^۳ مربوط به حالت s می‌باشد. پس از انجام رویه دسته‌بندی، حالت‌های موجود در هر کلاس به هم گره زده می‌شوند و یک حالت را تشکیل می‌دهند. رویه کلی دسته‌بندی و گره‌زدن حالت‌ها در این روش به شرح زیر می‌باشد:

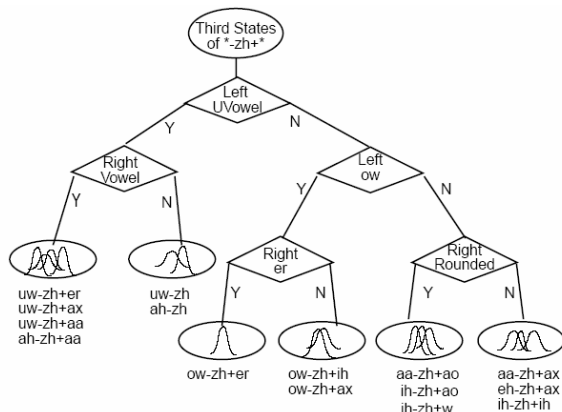
۱- مدل‌های مبتنی بر واج را (با یک گوسی در هر حالت) به عنوان مدل اولیه آموزش می‌دهیم.

۲- از مدل‌های مبتنی بر واج به عنوان مدل اولیه استفاده کرده و مدل‌های مبتنی بر سه واجی را با الگوریتم بام-ولج [۱] آموزش می‌دهیم (واضح است که فقط سه‌واجی‌های دیده شده آموزش داده می‌شوند). در حین آموزش تعداد داده‌های آموزشی برای هر حالت ذخیره می‌شود.

۳- برای هر سه واجی با واج میانی مشابه، حالت‌های متناظر (از نظر مکان آن حالت در مدل HMM) را برای دسته‌بندی در نظر می‌گیریم (هر حالت به عنوان یک دسته در نظر گرفته می‌شود).

کوچک برای آموزش استفاده شده است. مجموعه دادگان فارسیات کوچک شامل ۶۰۸۰ جمله فارسی از ۳۰۴ گوینده می باشد که هر یک ۲۰ جمله را ادا نموده اند که مجموعاً معادل ۵ ساعت گفتار می باشد. مجموعه دادگان فارسیات بزرگ شامل ۱۰۰ گوینده با گویش های مختلف می باشد که در کل معادل ۷۰ ساعت گفتار می باشد.

در این مقاله ۵۹۴۰ جمله از دادگان فارسیات کوچک برای آموزش و ۱۴۰ جمله باقی مانده از دادگان برای تست به کار برده شده است. برای مجموعه مدل های ساخته شده با دادگان فارسیات بزرگ، کل دادگان برای آموزش به کار برده شده است و ۱۴۰ جمله از مجموعه دادگان فارسیات کوچک برای تست به کار گرفته شده است. با به کار بردن مجموعه تست فارسیات کوچک برای مدل های ساخته شده از دادگان فارسیات بزرگ، می توانیم در نتایج حاصل از دادگان فارسیات بزرگ، تغییرات در شرایط ضبط دادگان را نیز لحاظ کنیم.



شکل ۱- گره زدن حالت ها با استفاده از درخت تصمیم گیری برای واج انگلیسی zh [۴]

۴- آزمایش ها و نتایج

مدل های HMM به کار برده شده در این مقاله از نوع چگالی پیوسته و با تلفیق های گوسی می باشند و پرش بین حالت های آن ها فقط به صورت چپ به راست مجاز می باشد. آموزش HMM ها با استفاده از الگوریتم بام-ولج انجام می شود.

در مرحله آموزش برای ساختن مدل های ترکیبی^{۱۳} مربوط به داده های آموزشی از یک فرهنگ لغت استفاده می شود. فرهنگ لغت به کار برده شده برای دادگان فارسیات کوچک شامل ۱۱۴۷ کلمه و برای فارسیات بزرگ شامل ۴۷۸۰۲ کلمه می باشد. همچنین با استفاده از فرهنگ لغات می توان تمام انواع سه واجی هایی دیده شده در دادگان آموزشی را مشخص کرد. با توجه به این فرهنگ لغات، در مجموع تعداد سه واجی ها درون کلمه ای^{۱۴} در مجموعه دادگان فارسیات کوچک ۴۳۶۲ و در مجموعه دادگان فارسیات بزرگ ۱۶۳۶۴ می باشد.

در روند آموزش اولیه، تعداد حالت ها برای هر مدل ۵ و تعداد گوسی ها در هر حالت ۱ در نظر گرفته شد. در نتیجه تعداد کل حالت ها در مجموعه دادگان فارسیات کوچک ۲۱۸۱۰ و در مجموعه دادگان فارسیات بزرگ ۸۱۸۲۰ می باشد. در مجموع ۱۳۰ پرسش خودکار با استفاده از روش جایگشت و بازگشت، برای ساخته شدن درخت

محاسبه می کنیم. اگر S مجموعه حالت های موجود در یک گره باشد، لگاریتم شباهت برابر است با:

$$L(S) = -\frac{1}{2} (\log((2\pi)^n |\Sigma(S)| + n) + \sum_{s \in S} \gamma_s)$$

که γ_s ، تعداد داده های آموزشی مربوط به حالت s بعد بردارهای آموزشی و $\Sigma(S)$ کواریانس داده های آموزشی مربوط به حالت های موجود در S می باشد [۴]، [۹].

۳-۳ پرسش های مختلف در مورد بافت چپ و راست حالت های موجود در ریشه درخت، مطرح شده و بر اساس جواب آن ها، دو گره فرزند چپ و راست برای درخت ایجاد می شود.

۳-۴ لگاریتم شباهت برای هر فرزند محاسبه شده و میزان افزایش در شباهت برآورد می شود [۴]، [۹]:

$$\Delta L = L(S_y) + L(S_n) - L(S)$$

که S_y و S_n به ترتیب مجموعه حالت ها در زیردرخت های چپ و راست می باشند.

۳-۵ پرسشی که به ازای آن ΔL ماکزیمم است انتخاب شده و به شرطی که ΔL و همچنین تعداد داده های آموزشی برای هر گره فرزند از یک حد آستانه بیشتر باشد، تقسیم بندی حاصل از آن پرسش به عنوان کلاس های جدید نگه داشته می شود (در غیر این صورت از تقسیم کردن گره اولیه صرف نظر می شود).

۳-۶ مراحل ۳-۳ تا ۳-۵ را برای گره های جدید تکرار می کنیم تا زمانی که دیگر هیچ گرهی را نتوان (با توجه به شرایط بالا) به دو گره جدید تقسیم کرد.

از مزایای استفاده از این روش کنترل میزان داده آموزشی برای حالت ها است؛ یعنی اگر مطرح کردن یک پرسش بر روی یک گره از درخت باعث شود که برای گره های فرزند داده آموزشی کافی وجود نداشته باشد تقسیم بندی آن گره متوقف می شود؛ در نتیجه گره های انتهایی (برگ ها) نشان دهنده کلاس هایی از حالت ها هستند که برای آموزش آنها داده کافی وجود دارد. مزیت دیگر این روش، توانایی سروکار داشتن با سه واجی های دیده نشده^{۱۱} است. معمولاً در داده های آزمون سه واجی هایی مشاهده می شود که در دادگان آموزشی موجود نبوده اند که این مساله در هنگام بازشناسی مشکل ساز می شود. با استفاده از درخت های تصمیم گیری تشکیل شده در مرحله آموزش، سه واجی های دیده نشده را می توان به یکی از دسته های موجود نسبت داد. شکل ۵ درخت تصمیم گیری را برای واج انگلیسی zh نشان می دهد.

باین حال، این روش در بازتاب تغییرات وابسته به زبان شناسی^{۱۲} بسیار ضعیف عمل می کند، زیرا این روش دسته بندی اش را بر اساس قوانین زبانی انجام می دهد. بنابراین در این روش طرح کردن دقیق مجموعه سوال ها بسیار مهم می باشد. طرح سوال ها می تواند خودکار [۵] یا دستی باشد. در این مقاله طراحی سوال ها به طور خودکار و با استفاده از روش جایگشت و بازگشت که در [۷] آمده است، انجام گرفته است.

۳- دادگان

در این مقاله، از مجموعه دادگان های فارسیات بزرگ و فارسیات

به وسیله درخت تصمیم‌گیری پیاده سازی و ارزیابی گردید. مدل‌های سه‌واچی با مجموعه دادگان‌های فارسی‌دات بزرگ و کوچک آموزش داده شدند و برای هر دادگان، مدل‌های مختلف با تعداد زنون‌ها و تعداد تلفیق‌های مختلف مورد ارزیابی قرار گرفتند. در هر دادگان با تعداد مشخصی زنون به بیشترین میزان افزایش دقت رسیدیم. در مجموعه دادگان فارسی‌دات کوچک به بهترین تعداد زنون ۵۰۰ و در مجموعه دادگان فارسی‌دات بزرگ به بهترین تعداد زنون ۴۰۰۰ رسیدیم. با به کار بردن مدل‌های آوایی سه‌واچی، دقت سیستم بازنشاسی نسبت به مدل‌های مبتنی بر واج افزایش قابل توجهی یافت. از آنجایی که طرح سوال دستی برای درخت تصمیم‌گیری بسیار وقت‌گیر و هزینه‌بر می‌باشد، به کار بردن روش‌های مختلف تولید خودکار پرسش‌ها برای درخت تصمیم‌گیری از کارهای آینده می‌باشد.

۶- منابع

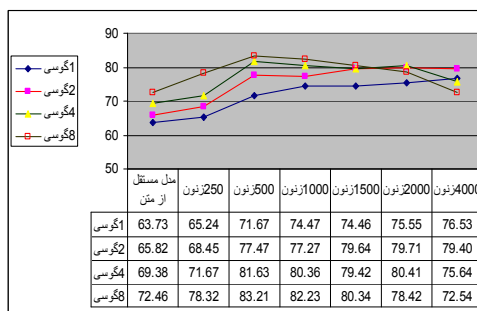
- [1] L. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, New Jersey: Prentice Hall, 1993.
- [2] K. Beulen, E. Bransch, H. Ney, state tying for context dependent phoneme models, *In Proc. Eurospeech97*, pp. 1179-1182, 1997.
- [3] J. Zhao, X. Zhang, A. Ganapathiraju, N. Deshmukh, and J. Picone, *tutorial for decision tree-based state tying for acoustic modeling*, Institute for Signal and Information Processing, Mississippi State University, 1999.
- [4] J.J. Odell, *the Use of Context in Large Vocabulary Speech Recognition*, Ph.D. Thesis, Cambridge University, 1995.
- [5] K. Beulen and H. Ney, Automatic Question Generation for Decision Tree Based State Tying, *Proceedings of ICASSP98*, pp. 805-808, Seattle, USA, May 1998.
- [6] M.Y. Hwang, F. Alleva, X.D. Huang, Senones, multipass search and unified stochastic modeling in Sphinx-II, *Eurospeech 93*, pp. 2143-2146, 1993.
- [7] Singh, R., Raj, B., Stern, R. M. Automatic Clustering and Generation of Contextual Questions for Tied States in Hidden Markov Models. *In Proc. ICSLP 99*, Vol. 1117-1202, 1999
- [8] S.J. Young, P.C. Woodland, The Use of State Tying in Continuous Speech Recognition, *Proc. Europe Conf. on Speech Communication and Technology*, pp. 2203-2206, Sept. 1993.
- [9] S.J. Young, J.J. Odell, P.C. Woodland, Tree-Based State Tying for High Accuracy Acoustic Modeling, *Proc. ARPA Human Language Technology Workshop*, Plainsboro, NJ, pp. 405-410, Morgan Kaufmann, March 1994.

زیر نویس‌ها

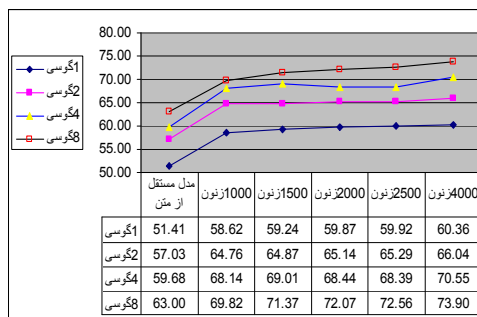
- ¹ Co-articulation
- ² Context Dependent Modeling
- ³ tie
- ⁴ data-driven clustering
- ⁵ decision tree-based state tying
- ⁶ merge
- ⁷ split
- ⁸ inter-group distance
- ⁹ occupation count
- ¹⁰ likelihood
- ¹¹ unseen triphones
- ¹² linguistic variation
- ¹³ composite models
- ¹⁴ within-word triphones
- ¹⁵ senone

تصمیم‌گیری به کار برده شده است که ۶۵ عدد از این پرسش‌ها در مورد بافت چپ و ۶۵ عدد بقیه در مورد بافت راست مطرح می‌شوند. با توجه به تعداد واج‌ها (۲۹ واج) و تعداد حالت‌ها در هر مدل (۵ حالت)، تعداد ۲۹×۵ درخت تصمیم‌گیری تشکیل می‌شود. بعد از انجام رویه دست‌بندی و پیدا شدن حالت‌های مشابه، داده‌های آموزشی مربوط به همه این حالت‌ها برای آموزش یک حالت عمومی (گره زده شده) به کار برده می‌شود؛ به این حالت عمومی زنون^{۱۵} گفته می‌شود [۶]. تعداد زنون‌ها را می‌توان با تغییر مقادیر آستانه در الگوریتم دست‌بندی کنترل کرد. در آزمایش‌های ما زنون‌ها مورد آموزش مجدد قرار گرفتند (با ۲، ۴ و ۸ گوسی) و در نهایت مدل حاصل برای بازنشاسی به کار گرفته شد. با مقایسه مدل‌های مبتنی بر سه‌واچی با مدل‌های مبتنی بر واج مشاهده می‌کنیم که دقت سیستم بازنشاسی گفتار ۱۰ الی ۱۱ درصد افزایش یافته است.

نتایج حاصل از بازنشاسی با تعداد زنون‌ها و تعداد گوسی‌های مختلف برای مجموعه دادگان فارسی‌دات کوچک و بزرگ در شکل‌های ۲ و ۳ نشان شده است. همان‌طور که مشاهده می‌شود در مجموعه دادگان فارسی‌دات کوچک به ازای ۵۰۰ زنون به بهترین نتیجه رسیده‌ایم ولی در مجموعه دادگان فارسی‌دات بزرگ با افزایش تعداد زنون‌ها دقت بازنشاسی همچنان افزایش یافته است. البته به علت طولانی شدن بیش از حد زمان آموزش، آزمایش‌ها را با تعداد زنون‌های بیشتر از ۴۰۰۰ تکرار نکردیم.



شکل ۲- دقت سیستم بازنشاسی گفتار با استفاده از مدل‌های مستقل از متن (واج) و وابسته به متن (سه‌واچی) با تعداد زنون‌های متفاوت در دادگان فارسی‌دات کوچک



شکل ۳- دقت سیستم بازنشاسی گفتار با استفاده از مدل‌های مستقل از متن (واج) و وابسته به متن (سه‌واچی) با تعداد زنون‌های متفاوت در دادگان فارسی‌دات بزرگ

۵- جمع‌بندی و نتیجه‌گیری

در این مقاله دو روش رایج گره‌زدن حالت‌ها بررسی شد و روش گره‌زدن