

مدلسازی گونه‌های تلفظی کلمات در سیستم بازشناسی اتوماتیک گفتار پیوسته فارسی

بهرام وزیرنژاد^۱، فرشاد الماس گنج^۲، محمود بیجن خان^۳

۱-۲- دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی

۳- دانشگاه تهران، دانشکده ادبیات و علوم انسانی

bahram_vazirnezhad@yahoo.com

چکیده

واژگان یک سیستم بازشناسی گفتار باید حاوی مدل‌های تلفظی از کلمات باشد تا بتواند شکلهای مختلف تحقق کلمات بر اساس ترتیب عناصر زیرکلمه که معمولاً واجها میباشند را، توصیف نماید. در این ارتباط روشی را ارائه می‌نماییم تا مدل‌های تلفظی فراتر از ذکر تلفظ‌های ساده مرجع و اولیه باشند و نیز مدل‌های جدیدی که قابلیت توصیف چند گونه تلفظی از هر کلمه را دارند ارائه می‌گردند. از آنجاییکه این روش قابلیت تولید گونه‌های تلفظی کلمات را بصورت خودکار و همراه با احتمال وقوع آنها دارد، بر روش اضافه کردن دستی گونه‌های تلفظی در واژگان برتری دارد. این روش، قواعد تلفظی را از ورودیهایی شامل فرم‌های واجی مرجع و بازشناسی شده یاد می‌گیرد و متعاقباً قواعد یاد گرفته شده را بعد از انجام مراحل هرس قواعد، جهت تولید گونه‌های تلفظی رایج هر کلمه بکار می‌برد. با یادگیری قواعد تلفظی بجای گونه‌های تلفظی از دادگان تعلیم می‌توان مزایای روشهای مبتنی بر داده و مبتنی بر قاعده را با هم ترکیب نمود. از خصوصیات بارز روش بکار گرفته شده این است که احتمال گونه‌های تلفظی تولید شده بر اساس دادگان تعلیم و بصورت واقعی قابل محاسبه می‌باشد حتی اگر گونه تلفظی در تمام دادگان دیده نشده باشد. کلمات کلیدی: مدلسازی تغییرات تلفظی- واژگان- قواعد تلفظی- گونه‌های تلفظی

۱- مقدمه

تقریباً تمامی سیستم‌های بازشناسی اتوماتیک گفتار بر پایه مدل‌های آکوستیکی زیر کلمه مثل: واج یا هجا شکل می‌گیرند، بنابراین سیستم و بطور اختصاصی واژگان سیستم باید توانائی توصیف ورودی‌ها که کلمات یا جملات می‌باشند را به تمام صورت‌های قابل دریافت، داشته باشد. در سالهای اخیر محققین کوشش فراوانی در جهت ایجاد مدل‌های تلفظی که شامل گونه‌های مختلف تلفظ کلمات می‌باشند را به انجام رسانده‌اند، وقوع انواع تلفظها در گفتار پیوسته یک پدیده شناخته شده آوایی می‌باشد. تأثیر متقابل کلمات و تغییرات موضعی و مکانیزمهای دیگر همگی باعث ایجاد گونه‌های مختلف تلفظ یک کلمه می‌گردند. بعضی از این پدیده‌ها وابسته به گوینده می‌باشند در حالیکه برخی دیگر به گوینده وابسته نیستند [۷].

وقتی گونه‌های تلفظی کلمات توسط مدل‌های تلفظی پوشش داده شوند صحت بازشناسی افزایش می‌یابد، در واقع در حین بازشناسی ورودی‌ها بهتر با تلفظ اصلی کلمات تطبیق می‌یابند و بنابراین فرآیند بازشناسی تسهیل می‌شود. مدل‌های تلفظی در واژگان، به منظور متمایز نمودن ورودی‌ها در سطح کلمات ایجاد می‌شوند. با اینحال معرفی گونه‌های تلفظی زیاد در واژگان افزایش سردرگمی در سیستم را نیز به دنبال خواهد داشت و این مسأله می‌تواند صحت بازشناسی را کاهش دهد [۱]. بنابراین

^۱ دانشجوی دکتری مهندسی پزشکی، گرایش بیوالکتریک

^۲ استادیار دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر

^۳ استادیار دانشکده ادبیات و علوم انسانی، دانشگاه تهران

باید محدودیتهایی در پذیرش گونه‌ها در نظر بگیریم این مسأله در مورد گونه‌های کلمات کوتاه بیشتر اهمیت پیدا می‌کند. در این مقاله روشی برای مدلسازی تغییرات تلفظی ارائه می‌شود، هدف از اینکار ایجاد یک الگوریتم اتوماتیک جهت ایجاد واژگانی با تلفظهای معمول هر کلمه، بعلاوه احتماله‌های وقوع آنها می‌باشد. این روش مبتنی بر قواعد تلفظی و با استفاده از دادگان گفتاری فارس دات بزرگ انجام می‌گیرد. اساس این روش بر این باور استوار می‌باشد که تقریباً تمام گونه‌های تلفظی کلمات با اعمال قواعد تلفظی به فرم مرجع تلفظی کلمه قابل دستیابی می‌باشند. بعد از یادگیری قواعد از دادگان، می‌توان برای کاربردهای دیگر از آنها استفاده نمود و نیازی به تولید مجدد قواعد نیست.

۲- روشهای مختلف مدلسازی تغییرات تلفظی

در این قسمت به بررسی خصوصیات روشهای مختلف مدلسازی تغییرات تلفظی می‌پردازیم. روشهای گوناگون جهت مدلسازی تغییرات تلفظی را به سختی می‌توان براساس نوع تغییرات تلفظی طبقه‌بندی کرد. بنابراین چارچوبی که برای طبقه‌بندی روشهای مختلف مدلسازی استفاده می‌کنیم براساس سئوالات مطرح شده در ادامه می‌باشد:

چه نوع تغییرات تلفظی را می‌خواهیم مدلسازی کنیم؟، داده‌های مورد نیاز چگونه مهیا می‌گردند؟، آیا این داده‌ها بعد از پردازش باید فرمالیزه گردد یا خیر؟، تغییرات مدل شده در کدام قسمت سیستم بازشناسی گفتار قرار خواهند گرفت؟

نتایج به روشهای مختلفی از جمله استخراج قواعد تلفظی، تعلیم درختهای تصمیم، تعلیم شبکه‌های عصبی مصنوعی و محاسبه ماتریس ابهام واج قابل ارائه می‌باشند.

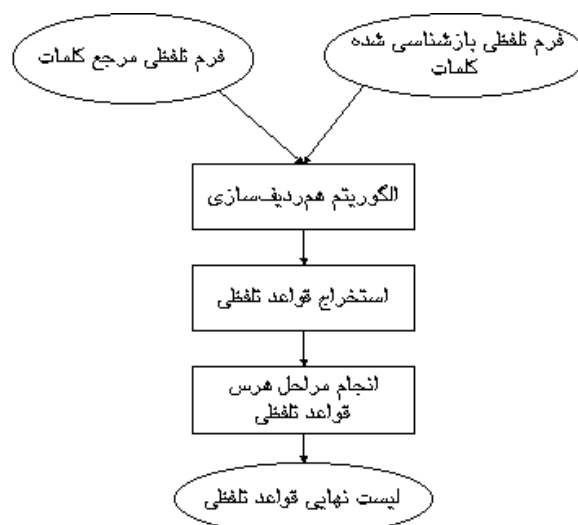
اکثر سیستم‌های بازشناسی اتوماتیک گفتار شامل سه قسمت اصلی جهت بازشناسی می‌باشند که مدلسازی تغییرات تلفظی در آنها قابل انجام است [۴].

۱- واژگان، ۲- مدل‌های آکوستیکی، ۳- مدل زبانی

برای دستیابی به یک سیستم بازشناسی گفتار مناسب لازم است تا مدلسازی گونه‌ها در سه سطح انجام گیرد.

۳- اصول روش پیاده‌سازی شده

به منظور تولید گونه‌های تلفظی کلمات در ابتدا باید قواعد تلفظی را از دادگان استخراج و در مرحله بعد، از این قواعد در جهت تولید گونه‌ها استفاده نمود شکل (۱) مراحل مختلف روش بکارگرفته شده را تا مرحله تولید قواعد تلفظی نشان می‌دهد، در قسمت‌های بعد به توضیح مراحل اصلی این الگوریتم خواهیم پرداخت.



شکل ۱- مراحل استخراج قواعد تلفظی واجی که در مراحل بعد جهت تولید گونه‌های تلفظی کلمات مورد استفاده قرار می‌گیرند [۵].

۳-۱- تولید ورودی‌های تعلیم از دادگان

در این مرحله یک الگوریتم اتوماتیک از دادگان برچسب‌دار شده گفتار پیوسته فارسی تلفظ‌های مرجع کلمات را به همراه آنچه واقعا بیان شده استخراج می‌نماید. در انتها تعداد زیادی ورودی تعلیم که هر کدام شامل یک تلفظ مرجع و بیان شده هستند، جهت استخراج قواعد تلفظی تولید می‌شود.

۳-۲- الگوریتم هم‌ردیف‌سازی^۱

بعد از آنکه تعداد زیادی ورودی دوتایی تولید شد یک الگوریتم پیچش زمانی پویا^۲ برای تمام ورودی‌ها، فرم مرجع و فرم بازشناسی شده را با توجه به فاصله واجها از یکدیگر طوری هم‌ردیف‌سازی می‌نماید که حداقل فاصله براساس اختلاف واجها بین آنها ایجاد شود.

۳-۳- الگوریتم استخراج قواعد تلفظی

این الگوریتم اختلاف‌های بین فرم مرجع و فرم بازشناسی شده موجود در هر ورودی را بعد از انجام هم‌ردیف‌سازی، بصورت قواعد تلفظی که توصیف‌کننده این تغییرات هستند، استخراج می‌نماید.

۳-۴- هرس قواعد^۳

در این مرحله تعداد دفعاتی که یک قاعده تلفظی از کل دادگان استخراج شده است و نیز تعداد دفعاتی که آن قاعده می‌توانسته اعمال گردد (صرف‌نظر از اینکه اعمال شده یا نه) شمارش می‌گردد، بر اساس این دو شمارش احتمال وقوع قاعده محاسبه می‌شود، در نهایت براساس این احتمال و شمارش‌های انجام شده برای هر قاعده، قواعد استخراج شده طوری هرس می‌گردند که قواعد ارزشمند حفظ شوند و قواعدی که بصورت اشتباه استخراج شده‌اند حذف گردند.

۳-۵- الگوریتم تولید گونه تلفظی^۴

در این الگوریتم قواعد تلفظی باقی‌مانده بعد از مرحله هرس، جهت تولید گونه‌های مختلف تلفظی کلمات مورد استفاده قرار می‌گیرند. فرم مرجع کلمه یا عبارت به‌عنوان ورودی الگوریتم می‌باشد. در این فرآیند قواعد تلفظی چنانچه قابل اعمال به متن واجی موجود در فرم مرجع کلمه باشند اعمال می‌گردند که به تولید گونه‌های مختلف تلفظی آن منجر می‌شود با توجه به احتمال وقوع قواعد تلفظی احتمال گونه‌های تلفظی نیز محاسبه خواهند شد.

۴- دادگان

جهت استخراج قواعد تلفظی به دادگانی شامل فرم‌های مرجع و فرم‌های بازشناسی شده کلمات بصورت واج نویسی شده احتیاج است برای این منظور قسمتی از دادگان گفتاری فارس‌دات بزرگ گردآوری شده توسط مرکز تحقیقات پردازش هوشمند علائم مورد استفاده قرار گرفت. بخش استفاده شده شامل ۹۰۰ فایل wave از گفتار پیوسته فارسی می‌باشد، و محدوده وسیعی از گوینده‌ها شامل مرد و زن با سنین متفاوت و تحصیلات مختلف و لهجه‌های گوناگون را پوشش می‌دهد. با استفاده از این مجموعه، ۱۷۵۰۰۰ کلمه در دو فرم مرجع و بازشناسی شده استخراج گردید. در مراحل بعدی، فرم‌های مرجع و بازشناسی شده کلمات بعد از انجام هم‌ردیف‌سازی جهت تولید قواعد تلفظی مورد استفاده قرار می‌گیرند.

¹ Allignment

² Dynamic Time Warping

³ Rule Pruning

⁴ Variant Generation

۵- الگوریتم استخراج قواعد

۵-۱- قواعد تلفظی وابسته به متن واجی

قاعده تلفظ را به صورت زیر نمایش می‌دهیم:

$$L \underline{F} R \rightarrow O$$

رابطه بالا به این مفهوم است که چنانچه دنباله واجی F در بین دنباله واجی سمت چپ یا L و دنباله واجی سمت راست یا R قرار بگیرد می‌تواند بصورت دنباله واجی O تلفظ گردد [۳].

الگوریتم در سه مرحله جهت تولید قواعد تلفظی اقدام می‌نماید. که این مراحل در زیر توضیح داده می‌شوند.

مرحله اول: استخراج واجهای مرکزی^۱ یا F : الگوریتم با اسکن هر ورودی از چپ به راست بعد از انجام هم‌ردیف‌سازی نقاط اختلاف بین فرم مرجع تلفظی و فرم بازشناسی شده را مشخص می‌نماید هر اختلاف بصورت یک قاعده تلفظی استخراج می‌شود و واجهای مرکزی با در نظر گرفتن حداکثر واجهای متفاوت، از فرم مرجع استخراج می‌گردند.

مرحله دوم: استخراج دنباله خروجی^۲: بعد از تعیین دنباله واج مرکزی یا F دنباله نظیر در صورت بازشناسی شده کلمه بعنوان دنباله خروجی یا O در نظر گرفته می‌شود.

مرحله سوم: استخراج دنباله های چپی و راستی^۳: واج سمت چپ و واج سمت راست دنباله مرکزی به ترتیب به عنوان دنباله‌های L و R یا دنباله‌های متنی از فرم تلفظی مرجع استخراج می‌گردند. واجهای متنی به منظور وابسته کردن قاعده به شرایط متنی استخراج می‌شوند. طول این دنباله‌ها در کارهای انجام شده تا کنون از صفر تا دو واج در نظر گرفته شده‌اند. اما در پیاده‌سازی فعلی طول این دنباله‌ها را بصورت ثابت معادل یک واج از هر طرف واج مرکزی در نظر گرفته‌ایم.

بعنوان مثال در مورد کلمه فرضی زیر قواعد تلفظی را بصورتی که می‌بینید استخراج می‌نماییم:

a	b	ϕ	d	e	f	g	فرم تلفظی مرجع کلمه :
a	b	c	d	e'	f'	g	فرم تلفظی بازشناسی شده کلمه :
$b \underline{\phi} d \rightarrow c$							قاعده استخراج شده ۱ :
$d \underline{e f} g \rightarrow e' f'$							قاعده استخراج شده ۲ :

۵-۲- قواعد تلفظی تعریف شده خاص زبان فارسی

جزئیات موجود در فضای آوایی زبان فارسی، نیاز به تعریف قواعدی مختص این زبان را، علاوه بر قواعد ذکر شده در بالا ایجاب می‌نماید. در زبان فارسی هرگاه یک واج همخوان رسا بین دو واکه قرار گیرد به احتمال زیاد در حین تلفظ واج همخوان و یکی از واکه‌ها حذف می‌گردند. این قواعد را بصورت زیر تعریف می‌نماییم:

$$\text{واکه ۱ یا ۲} \rightarrow \text{واکه ۲} - \text{همخوان رسا} - \text{واکه ۱}$$

این قواعد بر خلاف قواعد قبلی برای اعمال به فرم مرجع تلفظی، وابسته به متن واجی نمی‌باشند. به منظور استخراج صحیح این قواعد الگوریتم تولید قواعد اصلاح گردید، به این ترتیب که هر جا اختلافی بین فرم مرجع و بازشناسی شده مشاهده شود، ابتدا واجهای قبلی و بعدی بررسی می‌گردند و چنانچه شرایط استخراج قاعده بصورت خاص بیان شده وجود داشته باشد، قاعده به این صورت تولید می‌گردد، در غیر این صورت قاعده بصورت وابسته به متن واجی ایجاد خواهد شد.

¹ Focus

² Output

³ Left and Right Context

۶- الگوریتم هرس قواعد تلفظی

۶-۱- پوشش قاعده تلفظی

هر قاعده تلفظی به شکل $LFR \rightarrow O$ بنا به تعریف یک دنباله شرط به صورت LFR دارد مثلاً قاعده تلفظی $nh i \rightarrow ?$ دارای دنباله شرط $/nh i/$ می‌باشد. بنا به تعریف تعداد وقوع دنباله شرط را در فرمهای مرجع واج نویسی شده کلمات موجود در دادگان مورد استفاده، پوشش قاعده تلفظی می‌گوییم. چنانچه پوشش یک قاعده تلفظی کم باشد بدان معنی است که شرایط اعمال قاعده بسیار کم رخ می‌دهد. بدین ترتیب منطقی بنظر می‌رسد که قواعدی را که پوشش آنها کمتر از یک حد آستانه باشد حذف نماییم. این مرحله از هرس، قواعد بسیار خاصی را حذف میکند که عمدتاً به سه دلیل به وجود آمده‌اند:

۱- اشتباه در واج نویسی فرم مرجع تلفظی

۲- تلفظ اشتباه گوینده بصورت موردی

۳- عدم هم‌ردیف‌سازی صحیح بین فرم مرجع و فرم بازشناسی شده

خیلی از قواعد اشتباه که به این دلیل ایجاد می‌شوند دارای دنباله مرکزی F با طول زیاد می‌باشند که وقتی از محک پوشش برای هرس کردن قواعد استفاده نماییم از لیست قواعد حذف خواهند شد [۶].

۶-۲- میزان وقوع قاعده تلفظی

میزان وقوع، یک محک نسبتاً مناسب برای هرس کردن قواعد می‌باشد. تعداد وقوع یک قاعده تلفظی بعد از مرحله تولید قواعد محاسبه می‌گردد، بدین ترتیب که تعداد استخراج یک قاعده تلفظی مشخص، از کل دادگان شمارش می‌شود، چنانچه این شمارش ناچیز باشد به این معناست که احتمال رخ دادن آن قاعده تلفظی کم است یا حداقل با توجه به دادگان مورد استفاده آن قاعده تلفظی بسیار کم دیده شده است. بنابراین میزان وقوع یک قاعده تلفظی به ابعاد دادگان نیز مربوط می‌باشد.

۶-۳- احتمال وقوع قاعده تلفظی

چنانچه در یک فرم مرجع تلفظی دنباله شرط LFR قاعده $LFR \rightarrow O$ وجود داشته باشد احتمال آنکه دنباله F بصورت دنباله O تلفظ گردد، احتمال وقوع این قاعده تعریف می‌گردد و رابطه آن بصورت زیر می‌باشد [۲].

$$P_{AL}(r_i) = \Pr(\text{output}(r_i) | \text{condition}(r_i)) \quad (1)$$

برای محاسبه احتمالهای وقوع باید دو شمارش انجام گیرد، یکی شمارش تعداد دفعاتی که یک قاعده تلفظی می‌توانسته اتفاق بیافتد صرف نظر از اینکه اتفاق افتاده باشد یا خیر این شمارش معادل میزان پوشش قاعده تلفظی است که در مرحله قبل به تشریح آن پرداختیم و دیگری شمارش میزان وقوع قاعده تلفظی می‌باشد. براین اساس احتمال وقوع بصورت زیر محاسبه می‌گردد.

$$P_{AL}(r_i) = \frac{\text{شمارش پوشش قاعده}}{\text{شمارش میزان وقوع}} \quad (2)$$

۶-۴- نحوه پیاده‌سازی الگوریتم هرس قواعد

به‌منظور هرس قواعد و حذف قواعد بی‌ارزش از سه محک پوشش قاعده، میزان وقوع قاعده و احتمال وقوع آن استفاده شد، به اینصورت که برای هر قاعده سه پارامتر فوق محاسبه می‌گردد و چنانچه در مورد هر پارامتر مقدار محاسبه شده کمتر از یک حد آستانه باشد، قاعده از لیست نهایی قواعد حذف می‌گردد. استفاده هم‌زمان از این سه محک قابلیت خوبی به الگوریتم هرس می‌دهد. فرض کنید قاعده تنها یکبار در کل دادگان بعد از انجام هم‌ردیف‌سازی دیده شده باشد و پوشش قاعده نیز یک باشد. به این ترتیب هر سه پارامتر فوق مساوی یک خواهند بود اگر الگوریتم هرس تنها براساس احتمال وقوع قاعده به هرس قواعد

پپردازد این قاعده با احتمال وقوع بالای یک (وقوع قطعی به شرط وجود دنباله شرط) هرگز دورریخته نمی‌شود حال آنکه ممکن است این قاعده تنها در اثر یک تلفظ نادرست اتفاقی گوینده ایجاد شده باشد.

با توجه به بررسی‌های انجام گرفته مشخص شد تنظیم دقیق حدود آستانه به منظور هرس براساس این سه پارامتر باید به دقت انجام گیرد و باید عواملی مثل حجم دادگان تعلیم مورد توجه قرار گیرد در مورد قواعد درج واج چون دنباله F نداریم دنباله LFR شامل تنها دو واج خواهد شد و بنابراین پوشش قاعده بسیار زیاد خواهد بود، بنابراین تصمیم گرفته شد این قواعد را بصورت جداگانه با حدود آستانه دیگری هرس نماییم. اگر برخورد مشابهی با این قواعد شود تعداد زیادی از قواعد درج از دست می‌روند. راه دیگر در هرس کردن قواعد درج این است که این قواعد بصورت دستی هرس شوند به این ترتیب که با توجه به اطلاعات زبانشناسی اگر یک قاعده نادرست باشد از لیست حذف می‌گردد احتمال قواعد درج را نیز می‌توانیم بصورت دستی وارد نماییم [۸].

۷- فرآیند تولید گونه‌های تلفظی

۷-۱- تأثیر متقابل کلمات در گفتار پیوسته

با توجه به اینکه از دادگان گفتار پیوسته به منظور استخراج قواعد تلفظی استفاده گردیده است، بنابراین قواعدی قابل استخراج می‌باشند که نحوه تغییرات بین کلمه‌ای را توصیف نمایند. قبل از اینکه فرآیند تولید گونه‌ها را با جزئیات بیشتر بررسی نماییم، به بررسی دقیق‌تری از قواعد بین کلمه‌ای خواهیم پرداخت.

دنباله شرط قواعد بین کلمه‌ای شامل قسمتهایی از فرم مرجع دو کلمه مجزا می‌باشد. از آنجا که در فرآیند تولید گونه در هر اجرا گونه‌های یک کلمه تولید می‌گردند، باید مکانیزمی طراحی شود که بتوانیم قواعد بین کلمه‌ای را نیز در تولید گونه بکار ببریم. یک راه حل ساده با تعریف فرم مرجع تعمیم یافته^۱ قابل دستیابی است در این حالت در اطراف فرم مرجع دو سمبل % به معنای مرز کلمه و یک * که با هر واجی در دنباله شرط قاعده قابل تطابق می‌باشد قرار می‌گیرند. به عنوان مثال فرم مرجع کلمه دانش بصورت زیر می‌باشد:

* % donef % *

چنانچه یک قاعده تلفظی بین کلمه‌ای قابل اعمال به فرم مرجع تعمیم یافته باشد. فرآیند تولید گونه یک گونه تلفظی تولید می‌کند و شماره قاعده تلفظی بین کلمه‌ای که باعث این تغییر شده است. به عنوان یک پارامتر اضافی به گونه تولید شده الصاق می‌گردد، با این راهکار هنگام بازشناسی تنها گونه‌های سازگار^۲ کنار هم قابل بازشناسی هستند در این گونه‌ها قاعده انتهایی اعمال شده به گونه اول با قاعده ابتدایی اعمال شده به گونه دوم یکی است [۱۳].

۷-۲- اصول کلی فرآیند تولید گونه‌های تلفظی

با توجه به آنچه ذکر شد اکنون می‌توان اصول کلی فرآیند تولید گونه را توضیح داد. همانطور که گفته شد این فرآیند در هر لحظه روی یک کلمه کار می‌کند و فرم مرجع تعمیم یافته را از چپ به راست می‌روبد^۳، در هر مکان فرآیند بررسی می‌کند که آیا دنباله شرط قواعد با واجهای آن مکان تطابق دارد یا خیر و چنانچه تطابق بین دنباله شرط قاعده و کلمه تشخیص داده شد یک گونه جزئی^۴ با اعمال قاعده به فرم مرجع ساخته می‌شود. دلیل نامگذاری گونه‌های جزئی این است که قسمتی از گونه که بعد از قسمت تغییر یافته گونه قرار دارد هنوز بررسی نشده و ممکن است در آن قسمت باز هم قاعده‌ای قابل اعمال باشد. به هر گونه جزئی یک عدد به نام نشانگر^۵ اختصاص می‌یابد این عدد مکان واجی که گونه تا آنجا بررسی شده را نشان می‌دهد و نقش

¹ Extended Reference form

² Compatible

³ Scan

⁴ Partial Variant

⁵ Marker

آن جلوگیری از تغییر مجدد در یک مکان از گونه جزئی می‌باشد. گونه‌ای که نشانگر آن تا انتهای کلمه قبل از مرز یا روی مرز رسیده باشد یک گونه کامل نامیده می‌شود. هر بار که فرآیند تولید گونه اجرا می‌شود گونه‌های جزئی جدیدی از اعمال قواعد بر قسمتهای بررسی نشده گونه‌ها تولید می‌شود. فرآیند زمانی خاتمه می‌یابد که تمام گونه‌ها بصورت گونه‌های کامل درآیند [۹]

۷-۳- الگوریتم تولید گونه

الگوریتم زیر فرآیند تولید گونه را بطور کامل بررسی می‌کند. در این الگوریتم لیستی گونه‌های تولید شده را ذخیره می‌نماید که با CL نشان داده شده است این لیست جهت جلوگیری از تولید مجدد یک گونه تلفظی ایجاد گردیده است، IL نیز حاوی گونه‌هایی است که برای پردازش بیشتر منتظر هستند و PL شامل گونه‌هایی است که اخیراً تولید شده‌اند. در این ساختار PL همواره زیرمجموعه CL می‌باشد.

همانطور که در الگوریتم مشخص می‌باشد شماره قواعد اعمال شده به ابتدا و انتهای گونه‌ها که قواعد بین کلمه‌ای هستند به گونه‌های تولید شده الصاق می‌شوند. اما احتمالهای این قواعد فعلاً در محاسبه احتمال وقوع گونه وارد نمی‌شوند و وقتی در نظر گرفته می‌شوند که مدلهای تلفظی در سیستم بازشناسی گرد هم آیند [۶].

الگوریتم تولید گونه‌های تلفظی بصورت زیر می‌باشد:

۱- مقدار دهی اولیه:

$il = []; pl = []; cl = [];$ و [فرم مرجع تعمیم یافته] = گونه اول یا v_1 و $v_1 = 1$ = نشانگر و $(v_1) = 1$ = احتمال و 0 = قاعده ابتدایی و 0 = قاعده انتهایی

۲- مرحله تکرار:

□ تمام گونه‌های کامل را از il بردار و به pl و cl اضافه کن.

□ به ازای تمام گونه‌های جزئی v_j در il :

● به ازای تمام قواعد R_i موجود در لیست قواعد:

اگر دنباله شرط قاعده با فرم مرجع تعمیم یافته از قسمت شروع شده با نشانگر الی آخر تطبیق یافت:

آنگاه قاعده r_i را به گونه v_j اعمال کن ← گونه v'_j

اگر $v'_j \notin cl$

آنگاه (F_i) طول + (v_j) نشانگر = (v'_j) نشانگر

اگر قاعده r_i یک قاعده بین کلمه‌ای است:

آنگاه قاعده ابتدایی یا قاعده انتهایی = i و (v_j) = احتمال (v'_j) = احتمال

در غیر این صورت (v'_j) = قاعده ابتدایی = (v_j) = قاعده انتهایی = (v_j) = قاعده انتهایی

و (r_i) = احتمال $(v_j) \times$ احتمال (v'_j) = احتمال

v'_j را به cl و pl اضافه کن.

● اگر هیچ قاعده‌ای به v_j اعمال نشد نشانگر (v_j) را یک واحد افزایش بده و v_j را به pl و cl اضافه کن.

۳- شرط توقف:

اگر تمام v_j ها در pl کامل هستند آنگاه توقف کن.

در غیر این صورت $cl = []; il = []; pl$ را به il منتقل کن و سپس برو به ۲.

¹ Consumed List

² Input List

³ Produced List

۸- محدودیت‌های تولید گونه

در فرآیند تولید گونه‌های تلفظی باید ملاحظات را در نظر گرفت تا از تولید گونه‌های نادرست جلوگیری شود. این موارد عبارتند از: ۱- دو بار تغییر در یک مکان مجاز نیست، ۲- درج دو واج کنار هم مجاز نیست، ۳- حذف دو واج اصلی در کنار هم مجاز نیست [۱۲]، [۱۱].

۹- محاسبه احتمال وقوع گونه‌های تلفظی

برای محاسبه احتمال وقوع گونه‌های تلفظی، همانطور که در الگوریتم نیز ملاحظه نمودید، احتمال قواعدی که اعمال آنها به فرم مرجع به تولید آن گونه منتج شده را در احتمال وقوع فرم تلفظی مرجع، ضرب نماییم.

۹-۱- احتمال وقوع گونه تلفظی مرجع

احتمال وقوع گونه تلفظی مرجع به معنای احتمال تلفظ کلمه به فرم مرجع و بدون تغییر می‌باشد. برای محاسبه این احتمال که در الگوریتم فوق با احتمال (v_1) نشان داده شده، نیاز به تعریف قواعد عدم تغییر^۱، می‌باشد. قواعد عدم تغییر را بصورت زیر نمایش می‌دهیم:

$$LFR \rightarrow F$$

این قواعد باعث ایجاد تغییر نمی‌شوند و تنها به دلیل اهمیت احتمال آنها در محاسبه احتمال وقوع گونه تلفظی مرجع، تعریف می‌شوند. احتمال قاعده فوق به معنای احتمال تلفظ دنباله F در متن واجی L و R ، بدون تغییر می‌باشد. برای محاسبه احتمال این قاعده از رابطه زیر استفاده کرده‌ایم:

$$P(r : LFR \rightarrow F) = 1 - \sum_{r_i \in R_C} P(r_i) \quad (3)$$

R_C مجموعه قواعدی است که دنباله‌های شرط همگی آنها، صرفنظر از دنباله خروجی‌شان، یکسان و بصورت LFR می‌باشند. محاسبه احتمال قواعد عدم تغییر، بعد از مرحله هرس قواعد تولید شده از دادگان و تهیه لیست نهایی قواعد تلفظی انجام می‌گیرد. به منظور محاسبه احتمال گونه مرجع، الگوریتمی مشابه الگوریتم فوق، قواعد عدم تغییر را به قسمتهای مختلف فرم مرجع کلمه که احتمال اولیه آن مساوی ۱ می‌باشد، اعمال می‌نماید. نکته مهم این است که مکان‌هایی از فرم مرجع که قواعد عدم تغییر در آنها اعمال شده‌اند به همراه احتمال آنها در متغیری ذخیره می‌شوند، به این ترتیب در مرحله تولید گونه‌های تلفظی، اگر قواعد اصلی بر آن مناطق اعمال گردند، احتمال قواعد عدم تغییر اعمال شده در آن مناطق در محاسبه احتمال نهایی گونه خنثی می‌گردند. چون احتمال قاعده عدم تغییر و احتمال قاعده جدید اعمالی در یک منطقه، نباید همزمان با هم ملحوظ گردند.

۹-۲- محاسبه احتمال گونه‌های تولیدی

برای محاسبه احتمال گونه‌های تلفظی، همانطور که در الگوریتم نیز ملاحظه نموده‌اید احتمال قواعد اعمال شده به فرم مرجع برای تولید آن گونه خاص را، در هم ضرب می‌نماییم. فرض کنید از فرم مرجع کلمه اطلاعات با فرم واجی بصورت $/ete/ɔ?ɔt/$ و با احتمال $0/9$ با اعمال دو قاعده $ɔ?ɔ \rightarrow ɔ$ با احتمال $0/05$ و $tte \rightarrow \phi$ با احتمال $0/01$ گونه $/ete/ɔt/$ ایجاد گردد، احتمال این گونه بصورت $0/9 \times 0/05 \times 0/01 = 0/0045$ محاسبه خواهد شد. چنانچه قواعد زیادی در تولید یک گونه وارد شده باشند احتمال گونه کم خواهد شد، شاید این گونه‌ها منطقی بنظر نرسند، بنابراین می‌توانیم تنها گونه‌هایی را نشان دهیم که احتمال آنها از حدی بالاتر باشند [۱۰].

¹Identity Rules

۱۰- بحث و نتایج

در این مقاله روشی برای ساختن گونه‌های تلفظی قابل قبول کلمات ارائه گردید. در این راستا از قواعد تلفظی به عنوان شکل فرمالیزه نمایش تغییرات تلفظی استفاده نمودیم، این قواعد از دادگان با حجم زیاد گفتار پیوسته فارسی استخراج گردیدند. این کار از طریق یک الگوریتم هم‌ردیف‌سازی فرم‌های تلفظی مرجع با فرم‌های تلفظی بازشناسی شده جملات انجام گرفت. در انتها قواعد استخراج شده بعد از اینکه با استفاده از محک‌های آماری هرس گردیدند جهت تولید گونه‌های تلفظی از یک فرم مرجع بکار گرفته شدند. در این کار از تعداد ۲۲۵۰۰۰ ورودی شامل فرم تلفظی مرجع و باز شناسی شده کلمه استفاده گردید. این ورودی‌ها هم‌ردیف‌سازی شدند اما بعد از انجام هم‌ردیف‌سازی تنها ورودی‌هایی را جهت استخراج قواعد استفاده کردیم که نسبتاً خوب بازشناسی شده بودند اینکار با استفاده از معیار فاصله و ایجاد یک حد آستانه در الگوریتم هم‌ردیف‌سازی صورت گرفت. و ۱۷۵۰۰۰ ورودی جهت استخراج قواعد باقی ماندند. از این تعداد ۴۷۰۰۰ قاعده استخراج گردید. این قواعد بعد از انجام مراحل هرس به ۹۸۰۰ قاعده کاهش یافتند. در انتها از این قواعد جهت تولید گونه‌های تلفظی کلمات استفاده گردید، که با توجه به بررسی‌های فراوان انجام شده در سیستم بازشناسی گفتار پیوسته فارسی، باعث افزایش کارایی سیستم نسبت به کارایی سیستم در روش تولید گونه‌های تلفظی توسط افراد زبان‌شناس گردیده است.

باید توجه داشت، تخصیص احتمال‌های صحیح به گونه‌ها نقش ویژه‌ای در کاهش نرخ خطای بازشناسی دارد. مزیت اساسی این روش، محاسبه احتمال وقوع گونه‌های تلفظی بر اساس دادگان است، که در روش مبتنی بر اطلاعات افراد زبان‌شناس، این امکان بصورت کمی وجود ندارد.

مراجع

- [1] B. Gold, N. Morgan "Speech and Audio Signal Processing" 2000, John Wiley & Sons
- [2] J. Deller, J. Proakis, J. Hansen "Discrete-Time Processing of Speech Signals" 1993, Mac Millan Company
- [3] Ch. Wooters, A. Stolcke "Multiple Pronunciation Lexical Modeling in a Speaker Independent Understanding System" ICSLP 94
- [4] M. Saraclar, H. Nock, S. Kudanpur "Pronunciation Modeling By Sharing Gaussian Densities Across Phonetic Models" Eurospeech 99
- [5] N. Cremelie, J. Martens "In Search of Better Pronunciation Models for Speech Recognition" Speech Communication 29-99, pp. 115-136
- [6] N. Cremelie, J. Martens "Automatic Rule-Based Generation of Word Pronunciation Networks" Eurospeech 97, pp. 2459-2462
- [7] H. Strik, C. Cucchiarini "Modeling Pronunciation Variation for ASR: A Survey of the Literature" Speech Communication 29-99, pp. 225-246
- [8] B. Byrne, M. Finke, S. Khudanpur "Pronunciation Modeling for Conversational Speech Recognition: A Status Report From WS97" IEEE 97
- [9] G. Tajchman, E. Fosler, D. Jurafsky "Building Multiple Pronunciation Models for Novel Words Using Exploratory Computation Phonology" Eurospeech 95
- [10] J. Fosler "Dynamic Pronunciation Models for Automatic Speech Recognition" ICSI 99
- [11] F. Zheng, Z. Song, P. Fung, W. Byrne "Modeling Pronunciation Variation Using Context-Dependent Weighting and B/S Refined Acoustic Modeling" Eurospeech 2001
- [12] Q. Yang, J. P. Martens "Data-Driven Lexical Modeling of Pronunciation Variation for ASR" ICSLP 2000
- [13] Q. Yang, J. P. Martens "On the Importance of Exception and Cross-Word Rules for the Data-Driven Creation of Lexica for ASR" ELIS 2000