



استخراج خودکار قواعد تلفظی و تولید گونه‌های آوایی کلمات برای بکارگیری در بازشناسی گفتار پیوسته فارسی

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

^۱ دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران
^۲ دانشکده ادبیات و علوم انسانی، دانشگاه تهران، تهران، ایران
^۳ پژوهشکده پردازش هوشمند علائم، تهران، ایران

چکیده

در گفتار پیوسته، کلمات به صورتهای گوناگونی تلفظ می‌شوند. وجود تنوعات تلفظی ایجاب می‌نماید واژگان سیستم بازشناسی گفتار پیوسته، حاوی گونه‌های مختلف تلفظی کلمات باشد. از این طریق می‌توان درصد دقت بازشناسی کلمات را افزایش داد. در این مقاله برآنیم روشی خودکار جهت تولید گونه‌های تلفظی کلمات ارائه نماییم. در این روش قواعد تلفظی از طریق مقایسه نسخه‌های همردیف‌سازی شده دنباله واجی مرجع کلمه و دنباله واجی بازشناسی شده از سیگنال گفتار، یادگیری می‌شوند. علاوه بر این، تعدادی قاعده که بر اساس دانش و اطلاعات آواشناسی زبان فارسی ایجاد شده، جهت تکمیل قواعد قبلی، به آنها اضافه می‌شوند. متعاقباً بعد از هرس آماری قواعد یادگیری شده، با اعمال آنها به دنباله واجی مرجع کلمات به تولید گونه‌های تلفظی آنها پرداخته می‌شود. این شیوه نسبت به روش اضافه نمودن دستی گونه‌های تلفظی کلمات، بر اساس اطلاعات آواشناسی، برتری دارد، چراکه در این روش گونه‌های تلفظی با توجه به خطاهای سیستم بازشناسی آواها به همراه احتمالات وقوع آنها و بصورت خودکار تولید می‌گردند. این روش با استفاده از دادگان گفتاری "فارس‌دات بزرگ"، پیاده‌سازی گردید و بکارگیری آن در تولید واژگان سیستم "شنوا" که یک سیستم بازشناسی گفتار پیوسته فارسی است، ۳/۴۷٪ افزایش درصد دقت بازشناسی کلمات را حاصل نمود.

کلمات کلیدی: مدلسازی تنوعات تلفظی، قواعد تلفظی، گونه‌های تلفظی، واژگان

۱- مقدمه

بازشناسی شده با ترکیب‌های آوایی کلمات موجود در واژگان سیستم می‌باشد. با اینحال ارائه بیش از حد گونه‌های تلفظی در واژگان باعث افزایش پیچیدگی^۲ واژگان سیستم خواهد شد و این مساله می‌تواند درصد دقت بازشناسی را کاهش دهد؛ بنابراین باید محدودیتهایی در پذیرش گونه‌های تلفظی کلمات اعمال نمود و تنها از گونه‌هایی که احتمال وقوع بالا دارند در تولید واژگان سیستم بازشناسی استفاده نمود [۱]. در تلاشهای اولیه، گونه‌های تلفظی^۳ کلمات، صرفاً بر اساس اطلاعات زبانشناسی ساخته می‌شد، اما در این روش ضعفهای بخش شناسایی واج سیستم بازشناسی گفتار، در نظر گرفته نمی‌شود؛ ضمن آنکه احتمال وقوع هر گونه قابل محاسبه نیست. جهت حل این مشکلات و همچنین خودکار نمودن تولید گونه‌های تلفظی کلمات که در روش دستی وقت زیادی از زبانشناسان خبره می‌گیرد، در سالهای اخیر محققین کوشش‌های فراوانی برای طراحی روش‌هایی برای استخراج خودکار تنوعات تلفظی از دادگان، نموده‌اند [۲،۳]. مدلسازی تنوعات تلفظی کلمات در تولید گفتار پیوسته نیز کاربرد دارد و اعمال مناسب این تنوعات

وجود تنوعات تلفظی در گفتار پیوسته یک مشکل اساسی در بازشناسی آن است. به بیان دیگر چنانچه کلمات همواره به یک نحو بیان می‌شدند، بازشناسی گفتار کار چندان مشکلی نبود. ولی تاثیر متقابل واجها بر یکدیگر^۱، تنوعات موضعی در ادای واجها و عوامل دیگر که تحت تاثیر سرعت گفتار، مشخصات گوینده، لهجه و حالات روحی گوینده‌ها هستند، باعث ایجاد گونه‌های مختلف تلفظی کلمات می‌گردند. برخی از این پدیده‌ها وابسته به گوینده و برخی از آنها مستقل از گوینده می‌باشند. اغلب سیستم‌های بازشناسی خودکار گفتار پیوسته بر پایه مدل‌های صوتی واج یا هجا شکل می‌گیرند، واژگان این سیستم‌ها باید حاوی گونه‌های تلفظی کلمات باشد تا گفتار ورودی به نحو مطلوب‌تری با کلمات موجود در آنها تطبیق یابد [۱]. استفاده از گونه‌های تلفظی در واژگان سیستم باعث افزایش درصد دقت بازشناسی کلمات در سیستم می‌گردد. این بهبود ناشی از تطبیق بهتر آواهای

با تخمین احتمالات قواعد تلفظی از دادگان آموزش حل نمود [۱۰]. در برخی از روش‌های مدلسازی، گونه‌های تلفظی مستقیماً از دادگان استخراج می‌گردند مانند هولتر^۸، اما روش‌های مبتنی بر قاعده، دارای مزایایی نسبت به این روش می‌باشند. اگر این قواعد به اندازه کافی عمومیت داشته باشند، حتی به کلماتی که در دادگان آموزش دیده نشده‌اند، قابل تعمیم خواهند بود، درحالی‌که این امر در روش استخراج گونه‌ها بصورت مستقیم قابل انجام نمی‌باشد. کار کرملی و مارتنز^۹ نیز بر همین اساس از قواعد استخراج شده برای تولید گونه‌های تمام کلمات استفاده نمودند [۷]. قواعد تلفظی می‌توانند مبتنی بر اطلاعات واج‌شناسی و یا بر اساس تغییرات واجی دیده‌شده پس از بازشناسی دادگان صوتی استخراج گردند. در روش دوم از مقایسه بین دنباله واجی مرجع و دنباله آوایی بازشناسی شده کلمات و استفاده از الگوریتم همدیف‌سازی این دو دنباله به استخراج قواعد پرداخته می‌شود. قواعد استخراج شده به شیوه دوم می‌توانند علاوه بر تنوعات آوایی دیده شده در سیگنال گفتار پیوسته، تنوعات ناشی از ضعفهای سیستم بازشناسی آوا را نیز مدل کنند. این قواعد با توجه به دادگان آموزش شکل می‌گیرند و تعمیم آنها به دادگان آزمون، مستلزم هرس نمودن آنها است. اینکار با توجه به معیارهای آماری و یا اعمال محدودیتهای مبتنی بر قواعد آواشناسی، قابل انجام می‌باشد [۱۱]. درخت تصمیم نیز می‌تواند ابزاری برای تولید گونه‌های تلفظی کلمات باشد. این کار را درخت می‌تواند، از طریق سؤالاتی که با توجه به دنباله واجی مرجع کلمه طرح می‌نماید، انجام دهد [۱۲]. درخت‌های تصمیم، ابزاری جهت طبقه‌بندی دادگان می‌باشند. در مرحله آموزش، درخت تصمیم ساختار خود را با استفاده از دادگان آموزش شکل می‌دهد [۱۳]. در ساخت یک درخت تصمیم چند نکته حائز اهمیت است: اول انتخاب سؤالاتی که به تفکیک اجزای دادگان آموزش می‌پردازند. برخی از الگوریتم‌ها می‌توانند بصورت خودکار سؤالات را طراحی نمایند. انتخاب سؤالات به گونه‌ای صورت می‌گیرد که حداکثر کاهش آنتروپی محقق شود. نکته بعدی نحوه توسعه و رشد درخت تصمیم به منظور نمایش بهتر دادگان آموزش است. توسعه درخت تصمیم تا جایی انجام می‌گیرد که شرط توقف تقسیم گره‌ها تحقق یابد. این شرط تعداد کمتر از حد آستانه عناصر موجود در آن گره و یا کاهش آنتروپی زیر حد آستانه موردنظر می‌باشد. گره‌هایی که دیگر قابل تفکیک نیستند به گره‌های نهایی موسومند. بعد از آموزش و ساخت درخت تصمیم، هرس آن به منظور تعمیم دادن آن به دادگان غیر از آموزش صورت می‌گیرد، چراکه ممکن است درخت تصمیم بیش از حد به دادگان آموزش تطابق یابد و خاصیت تعمیم نداشته باشد، که در جهت رفع این اشکال هرس درخت تصمیم در فاز نهایی صورت می‌گیرد. فوسلر^{۱۰} با استفاده از درخت‌های تصمیم به ارائه تنوعات تلفظی موضعی مبادرت ورزید [۱۴]. همچنین مدل‌های ایستا و پویا تلفظی را برای سیستم بازشناسی گفتار با استفاده از درخت تصمیم ارائه نمود [۱۵].

$$W^* = \arg \max_{[W \in L]} p(W | O) = \arg \max_{[W \in L]} p(O | W) p(W) \quad (1)$$

W^* کلمه انتخاب‌شده از میان واژگان است. $p(O | W)$ احتمال آن است که کلمه W متعلق به واژگان L ، به صورت دنباله O تلفظ گردد. $p(W)$ احتمال وقوع کلمه در حالت کلی است. محاسبه احتمال بیزین $p(O | W) p(W)$ در عمل با استفاده از مدل‌های مخفی مارکوف صورت می‌گیرد. مدل‌های مخفی مارکوف شامل یکسری گره‌ها که معرف واج‌ها و یکسری احتمالات گذر بین حالت‌ها می‌باشند. با توجه به مدل‌های مخفی مارکوف احتمالات $p(O | W) p(W)$ برای هر کلمه موجود در واژگان محاسبه می‌شود. محاسبه $p(O | W)$ برای هر کلمه با استفاده از

به طبیعی‌تر شدن گفتار تولیدی کمک می‌نماید. در نرخ بالای گفتار معمولاً گونه‌های تلفظی با طول واجی کمتر، توسط گوینده ادا می‌شوند، که ضرورت ارائه مدل‌های تلفظی در سیستم تولید گفتار را که به عواملی مثل نرخ تولید گفتار حساس باشد، نمایان می‌کند [۴]. مدلسازی تلفظ همچنین در سیستم‌های بازشناسی گوینده قابل استفاده است. در این کاربرد، سیستم نحوه تلفظ مختص هر گوینده را از طریق مقایسه نسخه‌های همدیف‌سازی شده^۴ واج‌های مرجع^۵ و واج‌های بازشناسی‌شده، یاد می‌گیرد، و به این ترتیب از طریق یادگیری چگونگی تلفظ کلمات توسط گوینده، به بازشناسی گوینده‌ها می‌پردازد [۵]. کاربرد دیگری از ارائه مدل‌های تلفظی در سیستم‌های تقطیع اتوماتیک گفتار می‌باشد. استفاده از قواعد تلفظی^۶ که یکی از روش‌های مدلسازی تنوعات تلفظی است، از جمله ابزارهای تقطیع گفتار است [۶].

این مقاله به ارائه روشی خودکار به منظور استخراج قواعد تلفظی از دادگان گفتار پیوسته می‌پردازد و نحوه بکارگیری این قواعد را در تولید گونه‌های آوایی کلمات تشریح می‌نماید. بکارگیری این قواعد در تولید واژگان سیستم بازشناسی گفتار پیوسته فارسی "شنوا" باعث افزایش درصد دقت بازشناسی کلمات گردیده است. بخش ۲ از این مقاله به مروری بر روش‌های مختلف مدلسازی تنوعات تلفظی، با توجه به کاربرد خاص آن در بازشناسی گفتار پیوسته اختصاص دارد. در بخش ۳ اصول روش پیاده‌سازی شده برای استخراج قواعد تلفظی بررسی می‌گردد. دادگان مورد استفاده برای استخراج تنوعات تلفظی در بخش ۴ معرفی می‌شود. در بخش ۵ نحوه استخراج قواعد تلفظی بطور دقیق تشریح می‌شود. روش‌های آماری استفاده شده جهت هرس قواعد تلفظی در بخش بعد ارائه خواهند شد. در بخش ۷ فرآیند تولید گونه‌های تلفظی، با استفاده از قواعد استخراج‌شده بیان می‌گردد. نتایج آزمون‌های انجام‌گرفته با استفاده از گونه‌های تلفظی تولید شده در بازشناسی گفتار پیوسته فارسی در بخش ۸ ارائه می‌شوند. در انتها به بحث و نتیجه‌گیری در ارتباط با روش پیاده‌سازی شده خواهیم پرداخت.

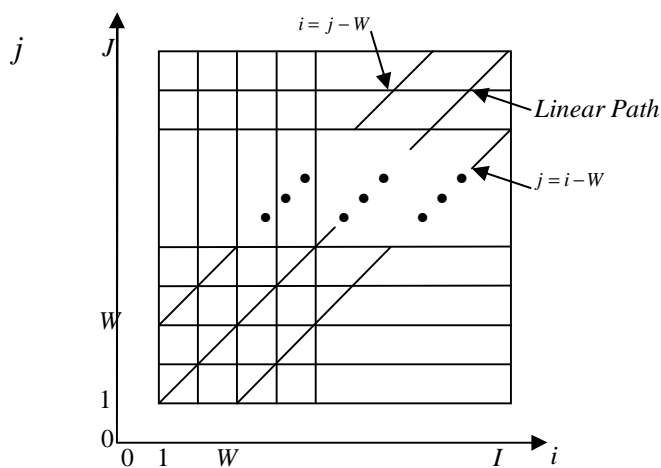
۲- مروری بر روش‌های مختلف مدلسازی تنوعات تلفظی

در این قسمت بررسی مختصری بر روش‌های مختلف مدلسازی تنوعات تلفظی خواهیم داشت. مدل‌های تلفظی می‌توانند شامل مدلسازی تنوعات تلفظی بین کلمه‌ای و یا منحصر به تنوعات داخل کلمه‌ای گردند. این مدل‌ها در سطح کلمه و یا عناصر زیر کلمه طراحی می‌شوند. آموزش مدل‌های تلفظی بصورت مبتنی بر دانش زبانشناسی یا بصورت خودکار و با استفاده از دادگان آموزش مناسب صورت می‌پذیرد. سؤالاتی نظیر اینکه: چه نوع تنوعات تلفظی را می‌خواهیم مدلسازی کنیم؟ داده‌های مورد نیاز چگونه مهیا می‌گردند؟ آیا این داده‌ها در غالب یک مدل ارائه می‌گردند؟ یا بصورت ذکر گونه‌ها در واژگان ارائه می‌شوند؟ و اینکه تنوعات مدل شده در کدام قسمت سیستم بازشناسی گفتار قرار خواهند گرفت، ویژگی‌های مدل‌های تلفظی را مشخص می‌نمایند. مدل‌های تلفظی به روش‌های مختلفی از جمله جمع‌آوری قواعد تلفظی، آموزش درخت‌های تصمیم، مدل‌های مخفی مارکوف و شبکه‌های عصبی مصنوعی قابل ارائه می‌باشند [۱].

در مدلسازی با استفاده از قواعد تلفظی، هدف تولید گونه‌های تلفظی کلمات با استفاده از اعمال قواعد تلفظی روی دنباله واجی مرجع کلمه می‌باشد. این روش مبتنی بر این باور است که تقریباً تمامی گونه‌های تلفظی کلمات از طریق اعمال قواعد تلفظی به دنباله واجی مرجع کلمه، قابل دستیابی می‌باشند. تعدادی زیادی از محققان قواعد تلفظی را در سیستم‌های بازشناسی خود استفاده نمودند تا مشکل تنوعات تلفظی را حل نمایند. از این طریق بهبودهای چشمگیری در خروجی سیستم‌ها گزارش شده‌اند [۷، ۸، ۹]. قواعد تلفظی غالباً فقط یک توصیف کیفی از یک تغییر مجاز واجی را ارائه می‌نمایند و فاقد احتمالات کمی هستند. این باعث می‌شود که اهمیت آنها نسبت به یکدیگر قابل مقایسه نباشد. تاجمن^۷ این مساله را

شده‌اند. به منظور جلوگیری از تولید گونه‌های نامتعارف، تمهیداتی بر پایه اطلاعات زبانشناسی در قسمت اعمال قواعد بر دنباله‌های واجی مرجع در نظر گرفته شده است از جمله اینکه اعمال دو قاعده درج در یک موضع از دنباله مرجع کلمه مجاز نمی‌باشد. محدودیت‌هایی از این دست که در فرآیند اعمال قواعد برای تولید گونه‌های تلفظی کلمات بکار رفته‌اند، مبتنی بر اطلاعات آواشناسی زبان فارسی بوده‌اند. استفاده از دانش زبانی برای تصحیح فرآیند خودکار تولید گونه‌های تلفظی در کارهای مشابه دیده نشده است؛ و برای اولین بار در این تحقیق به آن پرداخته شده است. احتمالات گونه‌های تلفظی کلمات با توجه به احتمال قواعدی که برای تولید آن گونه‌ها مورد استفاده قرار گرفته‌اند، محاسبه می‌شوند. محققان طول دنباله‌های واجی متنی را صفر تا دو واج از هر طرف در نظر گرفته‌اند و در مرحله تولید گونه‌های تلفظی با طبقه‌بندی قواعد بر اساس طول دنباله شرط آنها، برای اعمال آنها اولویت قائل شده‌اند [۷].

در مرحله اول، بطور خودکار از دادگان برچسب‌دار شده گفتار پیوسته فارسی دنباله‌های واجی مرجع کلمات به همراه دنباله‌های واجی زبانشناسی شده استخراج می‌شوند تا تعداد زیادی ورودی آموزش که هر کدام شامل دنباله مرجع و زبانشناسی شده یک کلمه هستند، جهت استخراج قواعد تلفظی آماده گردند. سپس یک الگوریتم هم‌ردیف‌سازی دنباله‌های مرجع و زبانشناسی شده موجود در هر ورودی را با توجه به فاصله واجها از یکدیگر، طوری هم‌ردیف‌سازی می‌نماید که حداقل فاصله مجموع محاسبه شده براساس اختلاف واجها، ایجاد شود. فاصله واج‌های زبان فارسی از یکدیگر بر اساس مشخصات ذاتی واجها و اختلاف بین آنها مشخص گردیده‌اند. به منظور بهینه‌سازی فرآیند هم‌ردیف‌سازی برای کاربرد خاصی که در این کار مورد نیاز بود، قیودی در فضای جستجوی الگوریتم هم‌ردیف‌سازی اعمال گردید. شکل ۲ محدوده مجاز اعمال شده را برای جستجو در الگوریتم هم‌ردیف‌سازی بین دو رشته واج مرجع و ورودی نشان می‌دهد. این محدوده جستجو، بهترین نتایج را در هم‌ردیف‌سازی ایجاد می‌نماید. محدودیت اعمال شده از نظر ریاضی بصورت $|i - j| < 3$ قابل بیان است. اگر i و j به ترتیب شماره واجهای موجود در دنباله‌های مرجع و زبانشناسی شده باشند. بنابراین هم‌ردیف‌سازی واجها فقط در محدوده فضای ذکر شده قابل انجام خواهد بود.



شکل ۲- محدوده جستجوی اعمال شده در هم‌ردیف‌سازی واجهای موجود در دنباله‌های مرجع و زبانشناسی شده کلمات [۱۸].

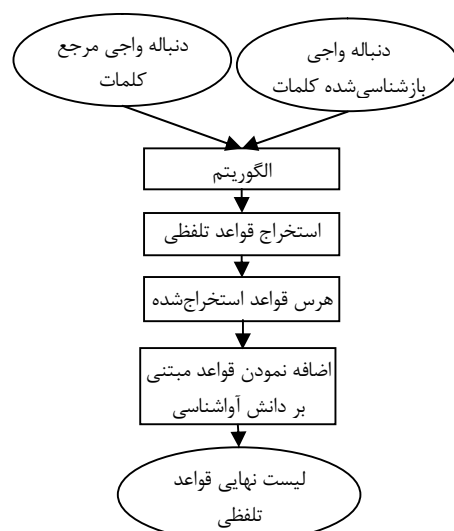
با مقایسه دنباله‌های واجی مرجع و زبانشناسی شده بعد از هم‌ردیف‌سازی می‌توان، نحوه تبدیل، حذف و درج واجها را بصورت قواعدی استخراج نمود. این قواعد تلفظی بصورت وابسته به متن^{۱۶} واجی استخراج می‌گردند [۱۸]. حذف و درج واجها در نسخه‌های هم‌ردیف‌سازی شده بصورت نمادین با علامت f در نسخه مرجع یا

الگوریتم‌های برنامه‌ریزی پویا و یا الگوریتم‌های ویتربی صورت می‌گیرد. وترز و استولک^{۱۱} با استفاده از مدل‌های مخفی مارکوف به مدلسازی واژگان چند تلفظی برای سیستم زبانشناسی گفتار مستقل از گوینده پرداختند [۱۶]. استفاده از مدل‌های مخفی مارکوف در سیستم‌هایی که طبقه شناسایی واج سیستم با واژگان سیستم ادغام شده باشند، مقدور است؛ ولی در سیستم‌هایی که این دو طبقه از یکدیگر جدا باشند رایج نیست.

شبکه‌های عصبی نیز روش دیگری برای مدلسازی تنوعات تلفظی می‌باشند که در برخی از منابع کارآیی آنها بهتر از روشهای آماری دیگر ذکر شده است. مساله موردنظر پیش‌بینی دنباله واجی زبانشناسی شده $\hat{P} = [\hat{p}_1, \dots, \hat{p}_N]$ با توجه به دنباله واجی مرجع $P = [p_1, \dots, p_N]$ می‌باشد. مشخصه‌ها به ارائه متن واجی در ورودی شبکه عصبی می‌پردازند، فوکودا و ساجیساکی^{۱۲} از گونه‌های تلفظی برای آموزش شبکه عصبی استفاده نمودند تا بتوانند گونه‌های تلفظی کلمات را پیش‌بینی نمایند [۱۷]. در کارهای جدیدتر از مشخصه‌های تمایزگر دیگری نظیر فرکانس پایه، جای تکیه کلمه و آهنگ بیان کلمه، بعنوان مشخصه‌های اضافی در ورودی شبکه عصبی استفاده می‌کنند، این مشخصه‌ها باعث پیش‌بینی بهتر نحوه تحقق واجها می‌شوند. چن و هازگاو^{۱۳} با تعریف چنین مشخصه‌هایی در ورودی شبکه عصبی، کارآیی سیستم خود را افزایش دادند [۹]. استفاده از شبکه‌های عصبی مستلزم داشتن دادگان آموزش با حجم بسیار زیاد برای همگرا نمودن شبکه عصبی و طراحی بهینه برای ساختار شبکه می‌باشد. لذا با حجم دادگان متوسط استفاده از این روش رایج نیست.

۳- اصول روش پیاده‌سازی شده

روش موردنظر در این مقاله استفاده از قواعد تلفظی برای تولید گونه‌های تلفظی کلمات است. این کار از طریق اعمال آنها به دنباله واجی مرجع کلمه انجام می‌شود. قواعد تلفظی و احتمال اعمال آنها بصورت اتوماتیک از دادگان استخراج می‌گردند، هر قاعده تلفظی شامل دو واج متنی^{۱۴} چپی و راستی و یک دنباله واج مرکزی^{۱۵} است که با توجه به آنچه قاعده بیان می‌کند، قابل تغییر است؛



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

این تغییرات شامل حذف، درج یا تبدیل واج می‌باشند. تعریف قواعد خاص با ساختاری مناسب برای زبان فارسی از نوآوری‌ها و ویژگی‌های جدید در کار انجام شده است. در این کار قواعد با توجه به خصوصیات زبان فارسی تعریف و استخراج

می شود. واج سمت چپ و واج سمت راست دنباله مرکزی به ترتیب به عنوان دنباله های L و R یا دنباله های متنی از نسخه مرجع استخراج می گردند. شرط اعمال قاعده تلفظی وقوع دنباله واج LFR در رشته واجی مرجع کلمه می باشد، به همین دلیل دنباله LFR را دنباله شرط^{۱۷} می نامند [۷]. در واقع واج های متنی باعث می شوند اعمال قاعده وابسته به شرایط متنی و واج های اطراف گردد. طول دنباله های واجی L و R از صفر تا دو واج متغیر در نظر گرفته می شوند اما با توجه به نظر زبانشناسان فارسی تصمیم گرفتیم طول این دنباله ها را بصورت ثابت و معادل یک واج از هر طرف واج مرکزی در نظر بگیریم. از نظر زبانشناسان در فارسی تاثیر متقابل واج ها بر یکدیگر برای واج های با فاصله بیشتر از یک واج چندان چشمگیر نمی باشد.

برای روشن تر شدن نحوه کار قواعد استخراج شده از دو دنباله واجی همردیف سازی شده را در مورد یک کلمه فرضی ذکر می نمایم.

| | | | | | | | | |
|-----|-----|-----|-----|------|------|-----|-----|---------------------------------|
| a | b | f | d | e | f | g | : | دنباله واجی مرجع کلمه : |
| a | b | c | d | e' | f' | g | : | دنباله واجی بازشناسی شده کلمه : |
| | | b | f | d | e | f | g | قاعده استخراج شده ۱: |
| | | d | e | f | g | | | قاعده استخراج شده ۲: |

f بصورت نمادین درج واج را در دنباله بازشناسی شده نشان می دهد. در قاعده اول واج c بین واج های b و d درج شده است. و در قاعده دوم دنباله واج ef به دنباله واج $e'f'$ تبدیل شده است.

۵-۱- تعریف قواعد خاص برای فارسی

به منظور بهینه سازی ساختار قواعد استخراج شده برای زبان فارسی، تصمیم گرفتیم با توجه به خصوصیات آوایی زبان فارسی، قواعد خاصی را برای این زبان طراحی نماییم. با توجه به نظر زبانشناسان و آنچه که در دادگان ملاحظه نمودیم، در زبان فارسی چنانچه یک واج همخوان رسا بین دو واکه قرار گیرد. کل این مجموعه سه واجی می تواند بصورت تنها یک واکه تلفظ گردد [۱]. به همین دلیل، قواعدی را برای مدلسازی این تغییرات واجی طراحی نمودیم. صورت کلی این قواعد بصورت زیر در نظر گرفته شد.

واکه ۲- همخوان رسا- واکه ۱ (R) واکه ۱ یا ۲

با توجه به ساختار این قواعد، الگوریتم استخراج قواعد را به نحوی تغییر دادیم که قادر باشد هر دو نوع قواعد را همزمان استخراج نماید.

۶- هرس قواعد تلفظی

قواعد استخراج شده باید با استفاده از معیارهای آماری و یا معیارهای مبتنی بر دانش آواشناسی هرس گردند تا قواعدی که بصورت اشتباه استخراج شده اند یا از نظر آماری کم ارزش هستند از لیست قواعد حذف گردند. چنین قواعدی معمولاً به دلایلی از قبیل اشتباه موردی گوینده در بیان خود و یا خطای الگوریتم همردیف سازی ایجاد می گردند. معیارهای آماری مورد استفاده شامل سه معیار پوشش^{۱۸} قاعده، میزان وقوع قاعده و درست نمایی اعمال^{۱۹} آن بوده است. برای هر قاعده مقادیر کمی سه معیار فوق محاسبه می شود و چنانچه مقدار هر کدام از این سه معیار از حد آستانه کمتر باشد، قاعده از لیست قواعد حذف می گردد. در ادامه نحوه استخراج معیارهای پوشش، میزان وقوع و درست نمایی اعمال قاعده ذکر می شود.

۶-۱ پوشش قاعده تلفظی

با توجه به قاعده تلفظی $LFR \rightarrow O$ ، بنا به تعریف، دنباله واجی LFR دنباله واجی شرط این قاعده نامیده می شود؛ چراکه شرط اعمال این قاعده به دنباله واجی مرجع کلمه وجود دنباله LFR در دنباله واجی مرجع می باشد. برای مثال دنباله

بازشناسی شده، نشان داده می شوند. قواعد استخراج شده در مرحله بعدی با استفاده از معیارهای آماری و یا اطلاعات آواشناسی هرس می گردند. در این مرحله احتمالات قواعد نیز با توجه به دادگان آموزش محاسبه می گردند. بعد از هرس این قواعد، لیستی از قواعد ارزشمند حفظ می شوند؛ لازم به ذکر است بعد از این مرحله، برخی از قواعد مبتنی بر دانش آواشناسی فارسی نیز به لیست قواعد استخراج شده از دادگان، اضافه می گردد. برای مثال چند نمونه از قواعد مبتنی بر دانش آواشناسی ذکر می شوند. در زبان فارسی عمدتاً بست چاکنایی بدون توجه به زمینه واجی حذف می شود. لذا قاعده JBf یکی از قواعد اضافه شده بر مبنای اطلاعات آواشناسی بوده است. همچنین معمولاً واج h و در بسیاری از موارد t در آخر کلمات حذف می شوند. این پدیده زبانی نیز به صورت یک قاعده مبتنی بر اطلاعات آواشناسی به لیست نهایی قواعد اضافه گردید. نمونه دیگر اینکه در زبان فارسی چنانچه سه همخوان در کنار یکدیگر قرار گیرند، یکی از آنها حذف می شود. مثلاً کلمه دستگاه با صورت واجی $dastg/h$ معمولاً بصورت $dasg/h$ تلفظ می شود. حذف یک همخوان در دنباله سه همخوان یک قاعده دیگر مبتنی بر اطلاعات آواشناسی بوده است، که بصورت قاعده در لیست نهایی قواعد اضافه گردید. مثال دیگر اینکه چنانچه بست چاکنایی یا $/v/$ بین دو واکه قرار گیرد، می تواند به یک واکه تبدیل گردد [۱۹].

بعد از اضافه نمودن قواعد مبتنی بر دانش آواشناسی، قواعد نهایی به منظور تولید گونه های تلفظی کلمات، بر روی دنباله واجی مرجع کلمات اعمال می گردند. درست نمایی گونه های تلفظی با توجه به احتمال قواعدی که در تولید آن گونه تلفظی دخیل بوده اند محاسبه می شود. در بخش ۷ به تشریح فرآیند تولید گونه های تلفظی و نحوه محاسبه درست نمایی گونه های تولید شده خواهیم پرداخت.

۴- دادگان آموزش

دادگان مورد استفاده در این تحقیق، بخشی از فارس دات بزرگ، تهیه شده در پژوهشکده پردازش هوشمند علائم، بزرگترین دادگان گفتار پیوسته فارسی است. این دادگان شامل مجموعه ای از گفتار ۱۰۰ گوینده است. که با توجه به تنوع سنی، جنسیت، سطح تحصیلات و تعلقشان به یکی از ۱۰ گویش رایج در ایران انتخاب شده اند. گویش های موجود در دادگان شامل تهرانی، ترکی، اصفهانی، جنوبی، شمالی، خراسانی، بلوچی، کردی، لری و یزدی می باشند. هر گوینده ۲۵ صفحه متن را در یک اتاق عادی خوانده است. متون شامل محدوده وسیعی از مطالب سیاسی، اقتصادی، فرهنگی، اجتماعی، ورزشی و ... می باشد. برچسب های واجی مرجع و فونتیک بر اساس کاراکترهای IPA با روشی مشابه دادگان فارس دات انجام شده است. حدود ۱۷۵۰۰۰ کلمه بازشناسی شده از این دادگان به منظور استخراج خودکار قواعد تلفظی مورد استفاده قرار گرفت. دادگان فارس دات بزرگ در پی نیاز به دادگان گفتاری با حجم زیاد طراحی شده است و از نظر ساختاری به دادگان فارس دات شباهت دارد [۲۰].

۵- قواعد تلفظی و نحوه استخراج آنها

هر قاعده تلفظی را به صورت زیر نمایش می دهیم.

$$LFR \rightarrow O$$

قاعده فوق بیان می کند، چنانچه رشته واج F در بین دنباله واجی سمت چپ L و دنباله واجی سمت راست R قرار گیرد می تواند بصورت دنباله واجی O تلفظ یا بازشناسی گردد [۷، ۸]. تاکید می کنیم L, F, R و O نمادهای معرف گروه های واجی نیستند بلکه دنباله های واجی می باشند. الگوریتم با مقایسه دنباله های همردیف سازی شده واجی مرجع و بازشناسی شده کلمات از چپ به راست، نقاط اختلاف بین این دو رشته واج را مشخص می نماید؛ و نواحی متفاوت بصورت قواعد تلفظی استخراج می شوند. واج های مرکزی با در نظر گرفتن تمام واج های متفاوت، از دنباله واجی مرجع استخراج می گردند. بعد از استخراج دنباله واج مرکزی، دنباله نظیر آن از دنباله واجی بازشناسی شده بعنوان دنباله خروجی یا O در نظر گرفته

۴-۶ هرس قواعد درج بر مبنای دانش آواشناسی زبان فارسی

سرعت ادای جملات موجود در دادگان گفتار پیوسته که در این مقاله مورد استفاده قرار گرفت نسبتاً زیاد است. در گفتار پیوسته با سرعت زیاد معمولاً درج واج کمتر صورت می‌گیرد و بیشتر پدیده حذف واج ملاحظه می‌گردد (۲۱). لذا جهت حفظ قواعد درج استخراج‌شده که تعداد آنها کم بود، برخوردی متفاوت در مرحله هرس صورت گرفت. برای این منظور هرس قواعد درج و تخصیص احتمال مناسب به آنها بصورت دستی و با استفاده از اطلاعات آواشناسی زبان فارسی انجام پذیرفت. برای مثال بر مبنای دانش آواشناسی چنانچه بعد از // همخوانهای /m/ یا // و یا /y/ قرار گیرند، واکه /i/ قابل درج است؛ و در غیر اینصورت درج آن ممکن نیست. نمونه دیگر اینکه اگر یک همخوان رسای واکدار قبل از یک واج انفجاری بیاید، احتمال درج واکه بین آنها وجود دارد. استفاده از اطلاعات آواشناسی زبان فارسی و ادغام آن با روش‌های استخراج آماری اطلاعات مبتنی بر دادگان از ویژگی‌های شاخص کار انجام گرفته بوده است. که در کلیه بخش‌های مقاله از جمله هرس قواعد انجام گرفته است.

۷- فرآیند تولید گونه‌های تلفظی

برای تولید گونه‌های تلفظی کلمات از لیست نهایی قواعد و دنباله واجی مرجع کلمه استفاده می‌شود. بدین ترتیب که دنباله واجی مرجع کلمه از چپ به راست و بصورت واج به واج مورد توجه قرار می‌گیرد، و در هر مکان بررسی می‌گردد که آیا دنباله شرط قواعد با واج‌های آن مکان تطابق دارد یا خیر و چنانچه واج‌ها با دنباله شرط مطابقت داشتند، قاعده به دنباله مرجع اعمال می‌گردد و یک گونه جزئی^{۲۰} تولید می‌شود. دلیل نامگذاری گونه جزئی این است که قسمتی از گونه که بعد از واج‌های تغییر یافته قرار دارد هنوز بررسی نشده و ممکن است در آن قسمت باز هم قواعدی قابل اعمال باشد. به هر گونه جزئی یک عدد به نام نشانگر اختصاص می‌یابد این عدد شماره واجی را نشان می‌دهد که گونه تلفظی کلمه تا آنجا بررسی شده است. نقش این عدد جلوگیری از اعمال قواعد دیگر، در مکانی است که قبلاً یکبار تغییر یافته است. گونه‌ای که نشانگر آن معادل شماره واج انتهایی کلمه باشد، گونه کامل نامیده می‌شود. فرآیند بارها تکرار می‌شود و هر بار گونه‌های جزئی جدیدی با اعمال قواعد بر قسمت‌های بررسی نشده گونه‌ها تولید می‌گردد. تکرار فرآیند اعمال قواعد به گونه‌های جزئی زمانی خاتمه می‌یابد که تمام گونه‌ها بصورت گونه‌های کامل درآیند [۱]. استفاده از قواعد عدم تغییر برای محاسبه احتمال تلفظ کلمه بصورت دنباله واجی مرجع آن از ویژگی‌های این پیاده‌سازی است که در بخش ۷-۲ به آن می‌پردازیم. اعمال محدودیت‌های دیگری در اعمال قواعد که مبتنی بر اطلاعات آواشناسی زبان فارسی بوده است، از دیگر نکات برجسته و نوین در این تحقیق بوده است.

۷-۱ اعمال قواعد تلفظی در مرز کلمات

قواعد تلفظی با استفاده از دادگان گفتار پیوسته استخراج گردیده‌اند، بنابراین قواعدی استخراج شدند که نحوه تغییرات بین کلمه‌ای را توصیف می‌نمودند. دنباله شرط قواعد بین کلمه‌ای شامل واج‌هایی از دو کلمه مجزا می‌باشند. از آنجا که در فرآیند تولید گونه هر بار گونه‌های یک کلمه تولید می‌گردند، باید راهکاری طراحی گردد که بتوان قواعد بین کلمه‌ای را نیز در تولید گونه‌ها بکار برد. برای این منظور "دنباله واجی مرجع تعمیم یافته"^{۲۱} تعریف می‌گردد، برای این کار دو نماد % به معنای مرز کلمه، و دو نماد * که با هر واجی در دنباله شرط قواعد قابل تطابق می‌باشند، در دو طرف دنباله واجی مرجع، قرار می‌گیرند. برای مثال دنباله واجی مرجع کلمه "دانش" بصورت زیر می‌باشد.

$$* \% d/ne.\% *$$

چنانچه یک قاعده تلفظی بین کلمه‌ای به دنباله واجی مرجع تعمیم یافته قابل اعمال باشد، فرآیند یک گونه تلفظی جدید تولید می‌کند و شماره قاعده تلفظی بین کلمه‌ای که اعمال گردیده به عنوان یک مشخصه اضافی به گونه تولید شده تخصیص می‌یابد، با استفاده از این راهکار، هنگام بازشناسی تنها گونه‌های سازگار

شرط قاعده تلفظی [nhi→nhi] بصورت /nhi/ می‌باشد. بنابراین، شرط اعمال این قاعده تلفظی به یک کلمه، وجود دنباله شرط /nhi/ در دنباله واجی مرجع آن کلمه است. تعداد وقوع دنباله شرط یک قاعده در تمام دنباله‌های مرجع واجی کلمات موجود در دادگان، بنا به تعریف پوشش آن قاعده نام دارد. پوشش یک قاعده بیانگر تعداد دفعاتی است که آن قاعده تلفظی می‌توانسته اتفاق بیافتد، صرفنظر از آنکه اتفاق افتاده باشد یا خیر. چنانچه پوشش قاعده‌ای کم باشد، به معنای آن است که شرایط لازم برای اعمال قاعده در کل دادگان کم رخ داده است. یا بطور کلی کم تحقق می‌گردد. چنین قواعدی عمدتاً به دلیل تلفظ اشتباه گوینده بصورت موردی و یا همردیف‌سازی اشتباه بین دنباله واجی مرجع و دنباله بازشناسی شده به غلط استخراج شده‌اند. بنابراین منطقی به نظر می‌آید که قواعدی که پوشش آنها کمتر از حد آستانه‌ای باشد، از لیست نهایی قواعد حذف شوند.

۶-۲ میزان وقوع قاعده تلفظی

میزان وقوع، معیار خوبی برای هرس نمودن قواعد می‌باشد. تعداد وقوع یک قاعده تلفظی بعد از مرحله تولید قواعد محاسبه می‌گردد، بدین ترتیب که تعداد استخراج یک قاعده تلفظی، 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 رخ داده است. باید توجه داشت میزان وقوع یک قاعده تلفظی تابعی از ابعاد دادگان نیز می‌باشد.

۶-۳ احتمال اعمال قاعده تلفظی

چنانچه در دنباله واجی مرجع کلمه دنباله شرط LFR وجود داشته باشد، احتمال آنکه قاعده $O \rightarrow LFR$ اعمال گردد، احتمال اعمال این قاعده تعریف می‌شود که رابطه ریاضی آن بصورت زیر است.

$$(2) \quad P(r_i) = \text{Pr}(O | \text{تلفظ } F \text{ بصورت } LFR \text{ در دنباله واجی مرجع کلمه})$$

در این رابطه، r_i نمایانگر قاعده نام از لیست قواعد است. این کمیت بیانگر آن است که با شرط وجود LFR در دنباله مرجع چقدر احتمال دارد F بصورت دنباله O تلفظ و یا بازشناسی گردد. مجموع احتمالات در فضای رخداد‌های ممکن مساوی یک است. و بنابراین این رابطه از نوع احتمال است. رخداد‌های ممکن شامل تلفظی F بصورت دیگر یا حذف و درج و یا بازشناسی آن بدون تغییر است. O می‌تواند بصورت دنباله تهی نیز باشد، که معادل عدم بازشناسی یا عدم تلفظ دنباله واجی F است. احتمال اعمال هر قاعده با استفاده از مقادیر پوشش و میزان وقوع قاعده که در دو بخش قبل ذکر شد، با استفاده از رابطه زیر تخمین زده می‌شود.

$$(3) \quad P(r_i) = \frac{\text{تعداد تلفظ تغییر یافته } F \text{ بصورت } O \text{ در دنباله‌های } LFR \text{ مرجع}}{\text{تعداد دنباله‌های } LFR \text{ دیده شده در دنباله‌های واجی کل دادگان}}$$

احتمالات اعمال قواعد ذخیره می‌شوند تا در مرحله تولید گونه جهت محاسبه احتمال گونه‌های تلفظی کلمات استفاده شوند. این سه معیار باید همزمان با هم برای هرس قواعد استفاده شوند، در غیر این صورت هرس به بهترین نحو انجام نمی‌گیرد. فرض کنید قاعده‌ای تنها یکبار در کل دادگان همردیف‌سازی شده، دیده شده باشد و پوشش قاعده نیز یک باشد، به این ترتیب احتمال وقوع آن نیز یک خواهد بود، اگر الگوریتم تنها براساس احتمال وقوع قاعده به هرس قواعد بپردازد، این قاعده با احتمال وقوع بالا (وقوع قطعی به شرط وجود دنباله شرط) حذف نمی‌شود، حال آنکه ممکن است این قاعده تنها بر اثر یک تلفظ نادرست اتفاقی گوینده ایجاد شده باشد. حدود آستانه معیارهای ذکر شده به منظور هرس قواعد باید به دقت انتخاب شوند، در این ارتباط عواملی مثل حجم دادگان آموزش باید مورد توجه قرار گیرند.

گونه بصورت $0.00045/0.01/0.05 \times 0.9 \times 0.1$ محاسبه خواهد شد. چنانچه قواعد زیادی در تولید یک گونه وارد شده باشند احتمال گونه کم خواهد شد، لذا می توان گونه هایی را قابل قبول دانست که احتمالشان از یک حد آستانه بالاتر باشد.

۸- نتایج تجربی

در این کار از تعداد ۲۲۵۰۰۰ ورودی شامل دنباله واجی مرجع و دنباله واجی باز شناسی شده کلمات استفاده گردید. از این تعداد کلمه ورودی، تعداد ۱۷۵۰۰۰ مورد که بعد از عمل همردیف سازی، هزینه همردیف سازی معقولی را نشان می دادند، جهت استخراج قواعد بکارگرفته شدند؛ و در نهایت تعداد ۴۷۰۰۰ قاعده از تطابق مجموعه دنباله های همردیف سازی شده مرجع و بازشناسی شده، استخراج گردیدند؛ از این تعداد، ۳۵۰ قاعده با ساختار مناسب برای زبان فارسی که در بخش ۵-۱ اشاره شد می باشند. این قواعد بعد از انجام مراحل هرس به ۹۸۰۰ قاعده کاهش یافتند. بهترین راه برای آزمایش نتایج حاصل از ساخت گونه های تلفظی، بکارگیری آنها در یک زمینه کاربردی مثل بازشناسی گفتار پیوسته است. به این منظور سیستم بازشناسی گفتار پیوسته فارسی "شنوا"، محصول پژوهشکده پردازش هوشمند علائم را انتخاب نمودیم. از مشخصات سیستم مورد استفاده، حجم واژگان سیستم است که در حد ۱۲۰۰ کلمه می باشد و اینکه برای محاسبه کلمات بازشناسی شده از مدل زبانی نحوی معنائی استفاده نمی کند. جهت انجام آزمایشات، از ۳۰ جمله که در دادگان آموزش قواعد تلفظی وجود نداشته اند بعنوان ورودی سیستم استفاده شد. جهت آزمایش کارائی قواعد تلفظی، ابتدا واژگان سیستم را با استفاده از تنها، دنباله واجی مرجع کلمات تشکیل دادیم و دقت بازشناسی کلمات را برابر با $46/3$ درصد محاسبه نمودیم. سپس نسخه واجی کلمات واژگان را به سیستم تولید گونه های تلفظی وارد نمودیم و خروجی آنها را حذف گونه های کم احتمال و اعمال برخی تصحیحات کلی مبتنی بر اطلاعات آواشناسی زبان فارسی، در واژگان سیستم قرار دادیم و مجدداً دقت بازشناسی کلمات را محاسبه نمودیم که مقدار جدید $49/77$ درصد حاصل شد. بنابراین مشاهده می گردد که با استفاده از گونه های تلفظی کلمات و احتمال آنها در فرآیند جستجوی کلمات سیستم بازشناسی گفتار پیوسته، به میزان $3/47$ درصد، کاهش خطای بازشناسی کلمات داریم. که معادل میزان $6/46$ درصد نرخ بهبود خطای بازشناسی کلمات بوده است؛ این نتیجه بخوبی عملکرد سیستم تولید گونه های تلفظی کلمات را به اثبات می رساند.

۹- نتیجه گیری

در این مقاله روشی خودکار برای تولید گونه های تلفظی قابل قبول کلمات ارائه گردید. در این راستا از قواعد تلفظی برای نمایش تنوعات تلفظی استفاده نمودیم. این قواعد از دادگان با حجم زیاد گفتار پیوسته فارسی استخراج گردیدند. این کار از طریق همردیف سازی نسخه های تلفظی مرجع کلمات با نسخه های تلفظی کلمات بازشناسی شده از گفته ها انجام گرفت. در انتها قواعد استخراج شده بعد از اینکه با استفاده از معیارهای آماری هرس گردیدند، جهت تولید گونه های تلفظی از نسخ مرجع کلمات استفاده شدند. سپس بعد از اضافه نمودن تعدادی قواعد براساس اطلاعات آواشناسی زبان فارسی، از مجموع آنها در جهت تولید گونه های تلفظی کلمات استفاده گردید. بکارگیری این گونه های تلفظی در واژگان سیستم "شنوا" که یک سیستم بازشناسی گفتار پیوسته فارسی است، $6/46$ درصد نرخ بهبود خطای بازشناسی کلمات نسبت به حالتی که تنها از گونه واجی مرجع کلمات در واژگان استفاده می شد، حاصل نمود.

مجاز به بازشناسی در کنار یکدیگر خواهند بود. در این گونه ها قاعده بین کلمه ای انتهای در گونه کلمه اول بازشناسی شده باید مطابق قاعده بین کلمه ای ابتدایی، در گونه کلمه بعدی باشد [۲۲].

۷-۲ الگوریتم تولید گونه

الگوریتم زیر فرآیند تولید گونه را نشان می دهد. در این الگوریتم لیستی گونه های تولید شده را ذخیره می نماید که با $CL^{۲۲}$ نشان داده شده است این لیست جهت جلوگیری از تولید مجدد یک گونه تلفظی ایجاد گردیده است، $IL^{۲۳}$ نیز حاوی گونه های جزئی است که منتظر پردازش بیشتر هستند و $PL^{۲۴}$ شامل گونه هایی است که در تکرار اخیر تولید شده اند. در این ساختار PL همواره زیرمجموعه CL می باشد. همانطور که در الگوریتم مشخص می باشد شماره قواعد اعمال شده به ابتدا و انتهای گونه ها که قواعد بین کلمه ای هستند به گونه های تولید شده الصاق می شوند. اما احتمالهای قواعد بین کلمه ای فعلاً در محاسبه احتمال وقوع گونه وارد نمی شوند و وقتی در نظر گرفته می شوند که مدلهای تلفظی در سیستم بازشناسی گرد هم آیند. ساختار اصلی الگوریتم تولید گونه های تلفظی بصورت زیر طراحی شد. در فرآیند تولید گونه های تلفظی باید ملاحظاتی را در نظر داشت تا از تولید گونه های نامتعارف جلوگیری شود. به عنوان مثال اعمال یک قاعده جدید در جایی که قبلاً قاعده ای اعمال شده و تغییر صورت گرفته، مجاز نیست. این محدودیتها بر اساس اطلاعات آواشناسی زبان فارسی طراحی شده اند.

برای محاسبه احتمال وقوع گونه های تلفظی، همانطور که در الگوریتم دیده می شود، احتمال قواعدی که اعمال آنها به تولید آن گونه منتج شده را در احتمال وقوع گونه تلفظی مرجع کلمه ضرب می نماییم. احتمال وقوع گونه تلفظی مرجع به معنای احتمال تلفظ یا بازشناسی کلمه به صورت واجی مرجع می باشد. برای محاسبه این احتمال که در الگوریتم فوق با احتمال V_1 نشان داده شده، نیاز به تعریف قواعد عدم تغییر^{۲۵}، می باشد. قواعد عدم تغییر را بصورت زیر نمایش می دهیم:

$$\underline{LFR} \rightarrow F$$

این قواعد باعث ایجاد تغییر نمی شوند و تنها به دلیل اهمیت درست نمایی آنها در محاسبه احتمال وقوع گونه تلفظی مرجع، تعریف می شوند. احتمال قاعده فوق به معنای احتمال بازشناسی دنباله F در متن واجی L و R بدون تغییر و بصورت F می باشد. برای محاسبه احتمال این قاعده از رابطه زیر استفاده کرده ایم:

$$P(r : LFR \rightarrow F) = 1 - \sum_{r_i \in R_c} P(r_i) \quad (4)$$

r_i قاعده i ام است و R_c مجموعه قواعدی است که دنباله های شرط همگی آنها، صرفنظر از دنباله خروجی شان، یکسان و بصورت LFR می باشند. محاسبه احتمال قواعد عدم تغییر، بعد از مرحله هرس قواعد تولید شده از دادگان و تهیه لیست نهایی قواعد تلفظی انجام می گیرد. به منظور محاسبه احتمال گونه مرجع، الگوریتمی مشابه الگوریتم فوق، قواعد عدم تغییر را به قسمت های مختلف نسخه مرجع کلمه که احتمال اولیه آن مساوی می باشد، اعمال می نماید. نکته مهم این است که مکان هایی از نسخه مرجع که قواعد عدم تغییر در آنها اعمال شده اند به همراه احتمال آنها در متغیری ذخیره می شوند، به این ترتیب در مرحله تولید گونه های تلفظی، اگر قواعد اصلی بر آن مناطق اعمال گردند، احتمال قواعد عدم تغییر اعمال شده در آن مناطق در محاسبه احتمال نهایی گونه خنثی می گردند. احتمال قاعده عدم تغییر و احتمال قاعده جدید اعمالی در یک منطقه، نباید همزمان با هم ملحوظ گردند. در اینجا برای واضح تر شدن نحوه محاسبه احتمال گونه های تلفظی مثالی ذکر می کنیم، فرض کنید از نسخه مرجع کلمه اطلاعات با دنباله واجی بصورت $ettel/t/$ و با احتمال $0/9$ با اعمال دو قاعده $/l/ \rightarrow /l/$ با احتمال $0/5$ و $t \rightarrow tt$ با احتمال $0/1$ گونه $etel/t/$ ایجاد گردد، احتمال این

۱- مقدار دهی اولیه:

$$CL = []; PL = []; IL = [];$$

[دنباله مرجع تعمیم یافته] = گونه اول یا $v_1 = 1$ نشانگر v_1 و (v_1) احتمال

و $= 0$ = قاعده ابتدایی v_1 و $= 0$ = قاعده انتهایی v_1

۲- مرحله تکرار:

تمام گونه‌های کامل را از IL بردار و به PL و CL اضافه کن.

آبه ازای تمام گونه‌های جزئی v_j در IL :

• به ازای تمام قواعد R_i موجود در لیست قواعد:

اگر دنباله شرط قاعده با واج‌های دنباله مرجع تعمیم یافته از قسمتی که نشانگر نشان داده‌الی 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 بین‌کلمه‌ای است:

آنگاه قاعده ابتدایی v_j یا قاعده انتهایی $v_j = i$ و (v_j) احتمال = (v_j) احتمال

در غیر این صورت: قاعده ابتدایی $v_j = v_j$ = قاعده ابتدایی v_j و قاعده انتهایی $v_j = v_j$ = قاعده انتهایی v_j

و (R_i) احتمال \times (v_j) 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: $IL, CL = []$; PL را به IL منتقل کن و سپس برو به ۲.

باید توجه داشت تخصیص احتمال به گونه‌ها نقش ویژه‌ای در کاهش نرخ خطای بازشناسی دارد. در روش تولید گونه مبتنی بر اطلاعات آواشناسی، امکان تخصیص احتمال به گونه‌ها وجود ندارد. از کارهایی که در ادامه اینکار می‌توان دنبال نمود، در نظر داشتن عواملی نظیر نرخ تولید گفتار در واحد زمان و موضع تکیه در کلمه است، که می‌تواند قابلیت فرآیند تولید گونه‌های تلفظی را افزایش دهد. با دخالت دادن این ویژگی در نحوه تولید گونه‌های تلفظی کلمات می‌توان گونه‌ها را مناسب با نرخ گفتار تولید نمود و به این ترتیب خاصیت تطبیق‌پذیری با نرخ گفتار را در سیستم ایجاد نمود. همچنین استخراج قواعدی که مختص به هجاهای تکیه‌دار در کلمات باشند می‌تواند کارآیی فرآیند تولید گونه‌های تلفظی را بهبود بخشد. راهکار دیگر در جهت تولید گونه‌های مفیدتر، اعمال هوشمندانه‌تر قواعد تلفظی به دنباله واجی کلمات است. برای این منظور باید ساختار واجی کلمه و عواملی نظیر طول کلمه، تعداد هجاها و نوع واج‌های موجود در کلمه را در نظر داشت. از این طریق می‌توان از تولید گونه‌های نامتعارف و مزاحم که باعث افزایش سرگشتگی در سیستم می‌شوند جلوگیری نمود.

تشکر و قدردانی

نویسندگان لازم می‌دانند، مراتب تشکر خود را از کلیه مسئولین و محققین پژوهشکده پردازش هوشمند علائم، به دلیل حمایت‌های مادی و معنوی از این تحقیق و در اختیار گذاردن امکانات لازم از جمله دادگان فارسی‌دات اعلام نمایند.

مراجع

- مهندسی پزشکی، دانشگاه صنعتی امیرکبیر، ۱۳۸۲
- [2] H. Strik and C. Cucchiaroni, "Modeling Pronunciation Variation for ASR: A Survey of the Literature," *Speech Communication*, vol. 29, no. 2, pp. 225-246, 1999.
- [3] B. Byrne, M. Finke and S. Khudanpur, "Pronunciation Modeling for Conversational Speech Recognition: A Status Report From WS97," *Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding*, Santa Barbara, California, 1997.
- [4] S. Werner, M. Wolff, M. Eichner and R. Hoffmann, "Modeling Pronunciation Variation for Spontaneous Speech Synthesis," *Proc. of ICASSP2004*, pp. 673-676, 2004.
- [5] D. Klusacek, J. Navratil, D. A. Reynolds and J. P. Campbell, "Conditional Pronunciation Modeling in Speaker Detection," *Proc. of ICASSP*, pp. 1-4, 2003.
- [6] N. Beringer, "Independent Automatic Segmentation by Self-learning Categorical Pronunciation Rules," *Proc. of Eurospeech*, Geneva, pp. 785-788, 2003.
- [7] N. Cremelie and J. P. Martens, "In Search of Better Pronunciation Models for Speech Recognition," *Speech Communication*, vol. 29, pp. 115-136, 1999.
- [8] N. Cremelie and J. Martens "Automatic Rule-Based Generation of Word Pronunciation Networks," *Proc. of Eurospeech*, pp. 2459-2462, 1997.

[۱] ب. وزیرنژاد، مدلسازی گونه‌های تلفظی کلمات در سیستم بازشناسی گفتار پیوسته فارسی، پایان‌نامه کارشناسی ارشد مهندسی پزشکی، دانشکده

- 8 Holter
- 9 Cremelie, Martens
- 10 Fosler
- 11 Wooters, Stolke
- 12 Fukada, Sagisaki
- 13 Chen, Hasegawa
- 14 Contextual
- 15 Focus
- 16 Context Dependent
- 17 Condition
- 18 Coverage
- 19 Application Likelihood
- 20 Partial Variant
- 21 Extended Reference Transcript
- 22 Consumed List
- 23 Input List
- 24 Produced List
- 25 Identity Rules



بهرام وزیرزاد در زمستان ۱۳۵۷ در تهران متولد شد. او در سال ۱۳۷۹ درجه کارشناسی خود را در رشته مهندسی پزشکی از دانشگاه علوم پزشکی شهیدبهشتی تهران اخذ نمود. متعاقب آن در سال ۱۳۸۲ موفق به اخذ درجه کارشناسی ارشد در رشته مهندسی پزشکی گرایش بیوالکتریک از دانشگاه صنعتی امیرکبیر گردید. او از سال ۱۳۸۲ دانشجوی مقطع دکتری تخصصی در رشته مهندسی پزشکی گرایش بیوالکتریک در دانشگاه صنعتی امیرکبیر می باشد. زمینه کارهای تحقیقاتی او تا کنون پردازش سیگنال گفتار، بازشناسی گفتار و روش های بازشناخت الگو بوده است. آدرس پست الکترونیکی نامبرده عبارتست از:

bvazirnezhad@aut.ac.ir



فرشاد الماس گنج در سال ۱۳۶۳ در رشته برق گرایش الکترونیک در مقطع کارشناسی از دانشگاه صنعتی امیرکبیر فارغ التحصیل گردیده است. سپس در همین گرایش دوره کارشناسی ارشد خود را در سال ۱۳۶۷ به پایان رسانید، و به سمت عضو هیئت علمی دانشگاه صنعتی امیرکبیر مشغول بکار گردید. بعد از تاخیری کوتاه، تحصیل خود را در رشته برق گرایش مهندسی پزشکی در دانشگاه تربیت مدرس ادامه داد؛ و در سال ۱۳۷۷ به درجه دکتری دست یافت. از آن زمان تا کنون در دانشکده مهندسی پزشکی دانشگاه صنعتی امیرکبیر با سمت استادیاری حضور دارد. زمینه تخصصی مورد علاقه او پردازش سیگنال و خصوصا پردازش انواع سیگنال های گفتاری است. آدرس پست الکترونیکی نامبرده عبارتست از:

almas@aut.ac.ir



محمود بیجن خان مدرک کارشناسی خود را در سال ۱۳۶۰ از دانشگاه ایالتی تگزاس در آرلینگتون ایالات متحده آمریکا دریافت نمود. او موفق به اخذ درجات کارشناسی ارشد و دکتری در زمینه زبانشناسی از دانشگاه تهران به ترتیب در سال های ۱۳۶۹ و ۱۳۷۴ گردید. بیجن خان از سال ۱۳۷۲ به هیئت علمی دانشکده ادبیات و علوم انسانی دانشگاه تهران پیوست. او در یازدهمین جشنواره بین المللی خوارزمی، بخاطر ارائه طرح دادگان گفتاری زبان فارسی، حائز رتبه سوم تحقیقات کاربردی شد. زمینه های تحقیقاتی مورد علاقه او آواشناسی، واج شناسی و زبانشناسی رایانه ای می باشد. آدرس پست الکترونیکی نامبرده عبارتست از:

mbjghan@chamran.ut.ac.ir

- [9] K. Chen and M. Hasegawa-Johnson, "Modeling Pronunciation Variation Using Artificial Neural Networks for English Spontaneous Speech," *Proc. of ICSLP*, 2004.
- [10] G. Tajchman, E. Fosler and D. Jurafsky, "Building Multiple Pronunciation Models for Novel Words Using Exploratory Computation Phonology," *Proc. of Eurospeech*, pp. 2247-2250, 1995.
- [11] M. Wester, "Pronunciation Modeling for ASR-Knowledge-based and Data-derived Methods," *Journal of Computer Speech and Language*, vol. 17, pp. 69-85, 2003.
- [12] P. A. Jande, "Pronunciation Variation Modeling using Decision Tree Induction from Multiple Linguistic Parameters." *Proc. of FONETIK*, Stockholm University, pp. 12-15, 2004.
- [13] H. Yu, and T.Schultz, "Enhanced Tree Clustering with Single Pronunciation Dictionary for Conversational Speech Recognition," *Proc. of Eurospeech*, Geneva, pp. 1869-2596, 2003.
- [14] E. Fosler-Lussier, "Contextual Word and Syllable Pronunciation Models," *Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding*, Keystone, Colorado, U.S.A, 1999.
- [15] J. E. Fosler-Lussier, *Dynamic Pronunciation Models for Automatic Speech Recognition*. Ph.D. thesis, University of California, Berkeley, 1999.
- [16] C. Wooters and A. Stolke, "Multiple-Pronunciation Lexical Modeling in a Speaker Independent Speech Understanding System," *Proc. Of ICSLP94*, pp. 1363-1366, 1994.
- [17] T. Fukada and Y. Sagisaki "Automatic Generation of a Pronunciation Dictionary Based on a Pronunciation Network," *Proc. of Eurospeech*, pp. 2471-2474, 1997.
- [18] J. Deller, J. Proakis and J. Hansen, *Discrete-Time Processing of Speech Signals*, Mac Milan Company, 1993.
- [19] ی. ثمره، *آواشناسی زبان فارسی*، ویرایش دوم، مرکز و نشر دانشگاهی، تهران، ۱۳۷۸.
- [20] M. Bijankhan and M. J. Sheikhzadegan, "FARSDAT-The Farsi Spoken Language Database", *proc. Of the 5th int. conf. on speech sciences and Technology*, vol. 2, pp. 826-829, Perth, 1994.
- [21] Q. Yang and J. P. Martens, "Data-Driven Lexical Modeling of Pronunciation Variation for ASR," *Proc. Of ICSLP*, China, Beijing, 2000.
- [22] Q. Yang and J. P. Martens, "On the Importance of Exception and Cross-Word Rules for the Data-Driven Creation of Lexica for ASR," *Proc. of 11th ProRisc Workshop*, Veldhoven, the Netherlands. pp. 589-593, 2000.

¹ Co articulation
² Confusion
³ Pronunciation Variants
⁴ Aligned
⁵ Reference
⁶ Pronunciation Rules
⁷ Tajchman