چهاردهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران
دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران) ایران، تهران – ۲۰ و ۲۱ اسفندماه ۱۳۸۷

ترکیب مدل‌های زبانی برای درک زبان گفتار فارسی در بستر محاوره

فاطمه کاوه یزدی
آزمایشگاه پردازش گفتار
دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف
fkaveh@cs.sharif.edu

حسین صامتی
استادیار و عضو هیئت علمی
دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف
Sameti@sharif.edu

محمّد بحرانی
آزمایشگاه پردازش گفتار
دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف
Bahrani@ce.sharif.edu

چکیده: در این مقاله طراحی و پیاده‌سازی یک نمونه سیستم محاوره در زمینه اطلاع‌رسانی بانکی مورد توجه قرار گرفته؛ در همین راستا و برای توسعه واحد درک زبان گفتار، که با دریافت نتیجه بازشناسی گفتار، سعی در تشخیص منظور کاربر دارد، یک راه‌کار عملی مبتنی بر ترکیب مدل‌های زبانی ارائه شده است. این روش، ترکیبی از دو روش پایه است که برای درک زبان گفتار در یک بستر محاوره به کار گرفته شده و علاوه بر ارائه دقّت و انعطاف بالا در درک گفتار، دارای اثرات مثبتی در بهبود خروجی بازشناسی گفتار پیوسته نیز می‌باشد. این مدل دارای یک ابرساختار معنایی- نحوی سه‌جزئی و مبتنی بر گرامر مستقل از متن معنایی است، که جز مرکزی آن می‌تواند دارای زیرساختاری همانند خود می‌باشد. برای تکمیل این مدل از ترکیب آن با مدل آماری مبتنی بر n-gram بهره گرفته شده است و دقّت نهایی این روش در تشخیص هدف برابر ۹۶/۸ درصد تخمین زده شده است.

واژه های کلیدی: سیستم محاوره گفتاری، بازشناسی گفتار، درک زبان گفتار، مدل زبانی، لایه معنا، شکاف

۱- مقدمه

امروزه سیستم‌های محاوره گفتاری[1]، یکی از مهمترین دسته از سیستم‌های هوشمند مبتنی بر فناوری بازشناسی گفتار هستند که از دیدگاه‌های مختلف زبان‌شناسی، روان‌شناسی و علوم اعصاب‌شناختی مورد بررسی قرارگرفته‌اند. این سیستم‌ها را می‌توان نمونه‌های مبتنی بر دانشی دانست، که برای تعامل با کاربر انسانی با استفاده از گفتار و برای فراهم کردن اطّلاعات مورد نیاز کاربران توسعه‌یافته‌اند. از جمله وظایف این سیستم‌ها می‌توان به راهنمای اطلاعات مسافرتی[۱] و هواشناسی[۲]، کمک‌یار در فرایندهایی مانند یادگیری زبان [۳]، رانندگی[۴] و هدایت تماس در مراکز تماس [۵] اشاره کرد.

با گسترش روزافزون استفاده از اینترنت و وب جهانی[2] که هرکدام با سرعتی نجومی رشد می‌کنند، نیاز به استفاده از کامپیوتر نیز افزایش می‌یابد؛ درکنار این گسترش توجه و علاقه مردم برای دسترسی به این امکانات از طریق دستگاه‌های قابل حمل، می‌تواند به انتخاب گفتار به عنوان راه‌حل توصیه شده‌ای برای واسط کاربر این وسایل کوچک قابل حمل منجر شود. زبان گفتار به عنوان جایگزین مناسب برای واسط‌های کاربر سخت‌افزاری مانند ماوس، قلم نوری نه تنها به فضای عملکردی احتیاج ندارد بلکه به عنوان یک فنّاوری دست- آزاد[3] و چشم- آزاد[4] کار با رایانه‌های کوچک قابل حمل و دستی را در شرایط خاص نیز آسان می‌سازد.

با پیشرفت سیستم‌های بازشناسی گفتار و ارائه دقّت‌های بالا برای این سیستم‌ها، به مرور کاربردهای مبتنی بر بازشناسی گفتار تلفنی مورد توجه قرار گرفتند و از ابتدای دهه ۹۰ میلادی، سیستم‌های متفاوتی برای محاوره هدفمند با کاربردهایی با دامنه محدود بوجود آمدند. این سیستم‌ها با این مشّخصه شناخته می‌شوند که مقصود کاربر را به یک زبان طبیعی تشخیص داده و فرمانی مناسبِ آن، برای اجرای دستور کاربر صادر می‌نمایند[۶].

این سیستم‌ها در طول دوران متکامل شده و سه نسل از آنها به وجود آمده‌اند[۷] و با ادامه تحقیقات و تعریف پروژه‌های بزرگی مانند Olympus دانشگاه CMU، نسل چهارم از این سیستم‌ها در حال شکل‌گیری هستند. سیستم‌های محاوره از ابتدای شکل‌گیری تغییرات زیادی را پشت سر گذاشته‌اند، امّا دانستن این نکته ضروری است که بیشترین تغییرات در هر دوره- که منجر به ایجاد نسل جدیدی از آنها شده - در تعاریف آنها از فرایند درک گفتار کاربر تماس گیرنده تبلور یافته است. در ادامه و برای معرّفی این سیستم‌ها به بیان جزئیّات ساختاری از آنها و معرّفی واحدهای پایه در این سیستم‌ها می‌پردازیم.

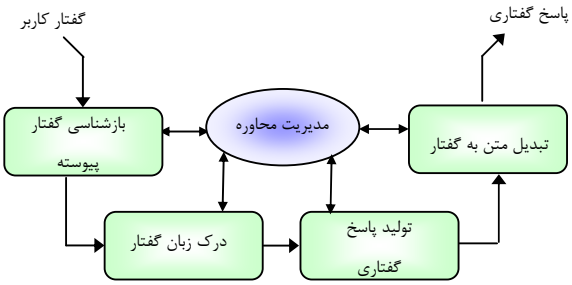
[1] Spoken Dialogue Systems (SDS)

[2] World Wide Web (WWW)

[3] Hands-free

[4] Eyes-free

هر سیستم محاوره از پنج جزء پایه با معماری مشابه شکل (۱) تشکیل شده‌است:



شکل (۱) : معماری کلی سیستم‌های محاوره گفتاری

با بررسی این معماری و وظیفه توصیف شده برای این سیستم‌ها مشخّص خواهد شد، مهمّترین و هم‌آوردجویانه‌ترین وظیفه در این زمینه درک زبان گفتار[1] است. درک زبان گفتار را می‌توان فرایند درک معنای[2] مورد نظر کاربر از بیان یک جمله یا عبارت دانست که مهمترین مشکل آن در زمینه درک یک معنا با توجه به شیوه نمایش آن، بروز می‌کند ؛ زیرا زبان‌های انسانی، معنا را در قالب فرم‌های ظاهری[3] مختلفی مانند لحن[4]، انتخاب کلمات[5] و نحو به کار می‌گیرند[۷].

بیشترین کاربرد سیستم‌های مزبور به هدایت تماس[6] مربوط می‌شود به خصوص با توّجه به این نکته که راهبری اتوماتیک تماس‌ها در مراکز تماس بزرگ می‌تواند در کاهش فشارکاری اپراتورهای انسانی مؤثّر بوده و سرعت پاسخگویی و تعداد خطوط تلفن مورد پشتیبانی را نیز افزایش دهد. در نهایت می‌توان گفت؛ کاربردهای متفاوتی که تنها وجه شباهت آنها طبیعت محاوره‌ای آنهاست، درکنار مشکلات نمایش معنا و ساختارهای غیرگرامری، مکث و قطع کردن جملات [۶] باعث می‌شود از بین فعّالیّت‌های پردازش زبان‌های گفتار یکی از پیچیده‌ترین فعّالیّت‌ها در همین سیستم‌ها صورت گیرد.

از دیگر مسائلی که روند توسعه این سیستم‌ها را کند می‌کند، نیاز به طراحی مجدد در هر سیستم با کاربرد متفاوت است،[۸ و۹] بدین معنا که طراحی سیستم برای یک کاربرد جدید باید از ابتدا شروع شده و تجربیّات پیشین ممکن است هیچ اثر مثبتی نداشته باشند و این مشکل در مهّمترین قسمت این سیستم‌ها – یعنی بخش درک زبان گفتار که طراحی کاملاً وابسته به بستر کاری دارد - بروز و ظهور می‌یابد.

در این مقاله یک نمونه از تلاش‌های انجام شده برای طراحی واحد درک زبان گفتار و بهبود توأم بازشناسی گفتار پیوسته فارسی در یک بستر محاوره مورد توجه قرار گرفته و معرّفی شده است.

از آنجا که نحوه به کارگیری واحد درک زبان و امکانات بازشناسی گفتار در طراحی یک سیستم تلفنی به زمینه کاری سیستم وابسته است در اوّلین بخش آتی به معرفی بستر کاری این پروژه پرداخته می‌شود و در ادامه روش ارائه شده از میان روش‌های متنوعی که برای ترکیب مدل‌های زبانی وجود دارد معرّفی می‌گردد و در نهایت در آخرین بخش نتایج ارزیابی و آزمون سیستم گنجانده شده و برداشت این گروه از این نتایج در بخش نتیجه‌گیری آورده شده است.

۲- بستر کاری

بستر محاوره در نظر گرفته شده در این فعّالیّت پژوهشی شامل یک بستر محاوره تلفنی با کاربرد هدایت تماس و اطّلاع‌رسانی تلفنی است. سابقه آماده سازی این سیستم‌ها نیز به ابتدای دهه ۹۰ میلادی باز می‌گردد که بانک‌های بزرگ بین‌المللی به علت گستردگی طیف خدمات خود در مناطق مختلف و علاقه به کاهش هزینه و کمبود تعداد شعب فعّال به ارائه خدمات تلفنی روی آوردند امّا تراکم بالای تماس‌های تلفنی در ابتدای این دهه، این غول‌های اقتصادی را برآن داشت تا با سرمایه‌گذاری در زمینه سیستم‌های خودکار تلفنی با کاهش تعداد تماس‌های وابسته به اپراتور انسانی خدمات خود را با سرعت بیشتر به کاربران بیشتری ارائه دهند. از نمونه‌های موفّق در این زمینه می‌توان به سیستم‌های تلفنی بانک‌های Fidelity Investment [۱۰] و E* Trade [۱۱] اشاره کرد.

سیستم مورد نظر در این مقاله نسل دوم از سیستم تلفنی فعلی – و در حال کار- است که قابلیّت اتوماسیون کلیه خدمات این سیستم را دارا می‌باشد. در سیستم اوّلیّه موجود کاربر در بدو تماس با منوهای سطح اوّل این سیستم آشنا شده و می‌تواند مرحله به مرحله و بدنبال اعلام منوهای هر سطح توسط سیستم با انتخاب شماره منو بر روی صفحه کلید به هر منو وارد شود؛ امّا مهمترین مشکلات در این سیستم‌ها عبارتند از :[۱۲]

- محدودیت برای افراد با معلولیت بینایی
- هزینه نگهداری بالا
- امکان به تله افتادن کاربران در منوهای تودرتو
- محدودیت تعداد انتخاب

که هر یک به نوبه خود امکان استفاده از آنها را محدود نموده است. امّا سیستم جایگزین با قابلیّت بازشناسی و درک گفتار می‌تواند علاوه بر رفع معضلات بالا، از تعداد تماس‌های منتقل شونده به اپراتور کاسته و امکان خدمات‌رسانی بهتر در زمان کوتاه‌تر و با نیروی کمتر را فراهم

---

[1] Spoken Language Understanding (SLU)
[2] Meaning
[3] Surface forms
[4] Prosody
[5] Lexical Choice
[6] Call Routing

کند؛ به علاوه در این سیستم کاربر با بیان مقصود خود در اوّلین قدم، گیج نخواهد شد.

سیستم جدید با دریافت جمله یا عبارت کاربر در بدو تماس با تعیین مقصود کاربر فرمان مناسبی را برای پاسخ به کاربر صادر خواهد کرد، از آنجا که این سیستم بیشتر نقش اطلاع‌رسانی را داشته و نقش هدایت تماس تنها به ارجاع به اپراتور و یا اشخاص با پست‌های خاص محدود می‌شود فرمان‌های این سیستم شامل بیان بخشی از اطلاعات مورد نیاز کاربر و یا هدایت به اپراتور (در صورت درخواست کاربر و یا عدم درک در تلاش مجدد) خواهد بود. با توجه به اینکه سیستم جدید باید به لحاظ ارائه خدمات، کاملاً مبتنی بر سیستم قبلی باشد، فعّالیّت‌های سیستم جدید نیز با همان ساختار طبقه‌بندی شده و شامل فرمان، درخواست تماس با اشخاص و انتخاب یک قسمت خاص از سیستم می‌باشد.

۳- طراحی واحد درک گفتار

برای توسعه بخش درک زبان گفتار دو رویکرد کلی وجود دارد که عبارتند از:

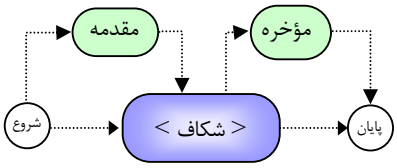
- روش مبتنی بر دانش
- روش مبتنی بر داده

روش مبتنی بر دانش، سعی در مدل‌سازی فرایندها با استفاده از دانش کارشناس یا کارشناسان خبره دارد، البته seneff در [۱۳] فرایند پیچیده تألیف گرامر را به استفاده مجدد بخشی از گرامر که مستقل از بستر و زمینه کاری است و تألیف بخش جدید تقلیل داده است. این تقلیل با فرض عدم تغییر ساختار نحوی در زمینه های مختلف استوار است، در این راستا می‌توان گفت یک گرامر مستقل از متن معنایی می‌تواند در سطح بالاتر برای ساختارهای نحوی متفاوت یکسان ظاهر شود.

در روش مبتنی بر داده اکثر سیستم‌های درک گفتار پیوسته از یک مدل مبتنی بر کانال بهره می‌گیرند، در این سیستم‌ها واحد درک زبان گفتار با استفاده از داده‌های برچسب خورده آموزش داده شده و مورد استفاده قرار می‌گیرند. اما از آنجا که سیستم مورد معرفی در این مقاله اوّلین نمونه از نوع خود بوده و تجربه و داده مورد نیاز برای اختیار هر یک از روشها به تنهایی وجود نداشته است؛ به همین دلیل از یک روش ترکیبی بهره گرفته‌ایم، بر اساس این روش هسته اولیه الگوهای مورد نیاز از داده‌های جمع‌آوری شده از یک سیستم Wizard-of-OZ شکل گرفته است که شامل بیش از ۳۵۰ نمونه می‌باشد. برای گسترش این الگوها از مهارتهای زبانی پایه‌ای بهره گرفته شده، تا الگوهای مرتبط مشابهی را به مجموعه اضافه نمایند.

۳-۱ ساختار پایه واحد درک زبان گفتار

الگوی مورد استفاده در درک زبان گفتار در این سیستم از ترکیب روش مبتنی بر مدل‌های جایگزین Yu [1] و همکاران در [۱۴] و روش wang همکاران در [۱۵] بدست آمده است. این روش نه تنها دقّت مناسبی برای بازشناسی گفتار در صورت ترکیب مدل‌های آکوستیکی با سایر مدل‌ها به دست می‌دهد بلکه دقّت قابل توجّهی در فرایند درک گفتار نیز ارائه می‌دهد. به علاوه اینکه این روش با داشتن مدل‌های جایگزین به عنوان انتخاب‌های موازی از انعطاف بالایی برخوردار است. در این روش برای هر یک از فرمان‌ها، دستورات و یا موارد درخواست تماس یک الگوی سه قسمتی ایجاد می‌شود که صورت کلی آن در شکل (۲) قابل مشاهده است، براساس این الگو هر فرمان دارای سه قسمت اصلی است که عبارتند از مقدمه [2]، شکاف [3] و مؤخره [4]؛ قسمت مرکزی مدل را شکاف تشکیل می‌دهد که مشخّص کننده یک فرمان و یا درخواست است.



شکل (۲) : الگوی سه جزئی پایه مدل پیشنهادی

از محاسن این مدل می‌توان به موارد زیر اشاره کرد:

- ساختار منعطف
- ساختار قابل تبدیل به گرامر مستقل از متن، با پارس ساده
- مقاوم در برابر خطاهای سطح نحو
- نرخ پس زدن پایین

از آنجا که ورودی واحد درک زبان گفتار از خروجی واحد بازشناسی گفتار پیوسته تأمین می‌شود معمولاً دارای خطاهای نحوی بوده و فرایند درک را با مشکل مواجه می‌کند. در چنین مواردی استفاده از یک گرامر منعطف می‌تواند این مشکل را کاهش دهد؛ بدین معنا که مواردی که از لحاظ معنایی درست بوده و تنها دارای اشکالات جزیی می‌باشند را می‌توان با فرض اشتباه در بازشناسی در اوّلین مرحله پذیرفت و تعیین عبارت دقیق را به مرحله بعدی سپرد. الگوی سه‌تایی معرّفی شده در بالا یک الگوی نحوی- معنایی است. بدین معنا که تقدّم اجزا براساس نحو فارسی تعیین شده و با کلیه کلماتی که از نظر معنایی در یک کلاس جای می‌گیرند برخورد یکسانی خواهد شد.

به عنوان مثال کلیه کلمات فرمان سیستمی و تقاضای اتصال دارای همین ساختار بوده و برای آنها تنها باید در قسمت‌های مقدمه و مؤخره

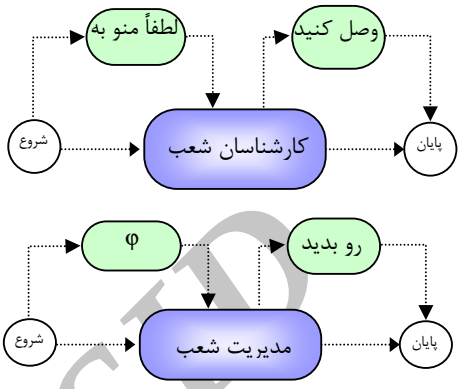
[1] Filler Model
[2] Preamble
[3] Slot
[4] Postamble

تغییراتی ایجاد کرد. زیرا نقش و ارزش آنها در جمله یکسان است و برخورد با آنها در لایه معنا تفاوتی ندارد.

به عنوان مثال دو جمله زیر را می‌توان در الگوی شکل (۳) گنجاند:

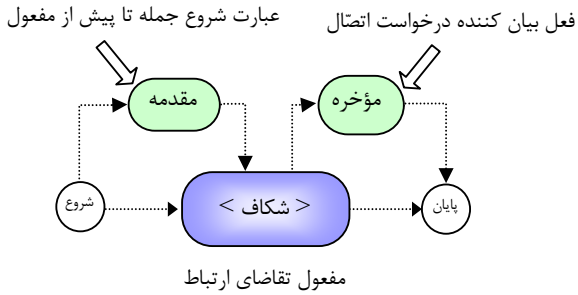
۱- لطفاً منو به کارشناسان شعب وصل کنید.

۲- مدیریت شعب رو بدید.



شکل (۳) : جاسازی دو جمله "*لطفاً منو به کارشناسان شعب وصل کنید*" و "*مدیریت شعب رو بدید*" در قالب کلی مدل پیشنهادی

این الگو کاملاً کلّی بوده و تمام مواردی را که دارای پیکربندی مشابه شکل (۴) باشند، خواهد پذیرفت:



شکل (۴) پیکربندی معنایی فرمان‌های درخواست اتصال

نمونه‌هایی از عبارات قابل درک با پیکربندی مشابه شکل (۴) عبارتند از:

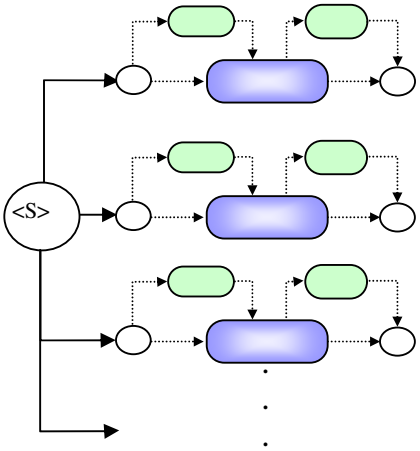
لطفاً آقای حسینی رو بدید.

می‌خواستم به سیستم تلفن گویا وصل بشم.

اطلاعات شعب مرکزی بانک رو میخوام.

کارشناسان شعب رو بدید.

از آنجا که در این سیستم کلیه وظایف از نظر اهمّیّت در یک سطح فرض شده‌اند بنابراین نسبت به هم اولویت نداشته و الگوهای سه‌تایی همه موارد به صورت موازی قرار می‌گیرند. براساس این ترکیب‌بندی برای هر یک از فرمان‌ها، الگوی مربوط به هر فرمان پایه پس از طراحی، در یک ساختار مبتنی بر مدل‌های جایگزین مانند شکل (۵) قرار می‌گیرد. یعنی اگر در کل سیستم بیش از سه دسته فرمان پایه وجود داشته باشند نتیجه پیاده‌سازی آنها به شکل زیر خواهد بود.



شکل (۵): الگوی تشکیل مدل جایگزین برای درک با استفاده از مدل‌های سه‌جزئی هر فرمان

البتّه ذکر این نکته ضروری است که این ترکیب سه‌تایی کاملاً عمومی بوده و برای جلوگیری از ابهام باید برای تمام کاربردها به صورت اختصاصی طراحی شود؛که این نیاز با یک طراحی جاسازی شده[1] در درون هر یک از شکاف‌ها انجام می‌گیرد که نمونه‌ای از آن در شکل (۶) ترسیم شده است. بر اساس این طراحی جاسازی شده، ممکن است هر شکاف خود شامل اجزاء مشابهی باشد؛ بدین معنا که هر شکاف در هر سطح ممکن است شامل یک الگوی سه‌جزئی با شکاف، مقدمه و مؤخره جدید باشد که در سطح بعدی از پارس قرار می‌گیرند. برای پارس بر اساس این پیکربندی، حرکت از گره شروع در شبکه سطح صفر شروع شده و گره‌ها یکی پس از دیگری پیموده می‌شوند تا اوّلین شکاف نیز پیموده شود، در این نقطه، در صورت وجود زیر شبکه در داخل شکاف و در صورت عدم اتمام جمله از شکاف در سطح صفر به شبکه جاسازی شده در آن در سطح یک پرشی انجام می‌شود. برای پرش از هر سطح به سطح بعد آدرس آخرین نقطه از شبکه پارس در سطح جاری در داخل پشته قرار می‌گیرد و پس از ورود به شبکه سطح پایین‌تر و حرکت در آن در بازگشت، آدرس محل قبلی در سطح پایه از داخل پشته خارج شده و ادامه عملیات در سطح صفر انجام می‌گیرد. نمونه‌ای از این فرایند به انضمام احتمالات حرکت، حاصله از ترکیب با مدل n-gram در تصویر (۶) گنجانده شده است، که در مورد نحوه ترکیب و جزئیّات بیشتر در بخش‌های آتی توضیحاتی گنجانده شده است.

[1] Embedded

۲-۳ ترکیب مدل مبتنی بر CFG با مدل N-gram

مدل‌های آماری در فرایند بازشناسی گفتار نقش اساسی داشته و در بیشتر موتورهای بازشناسی گفتار به کار گرفته می‌شوند، امّا مهمّترین مشکل در بکارگیری این مدل‌های آماری مانند مدل n-gram، کمبود داده آموزشی مورد نیاز برای یادگیری الگوهای زبانی است، به خصوص در مواردی که داده‌های مورد نیاز باید به یک زمینه خاص اختصاص داشته باشند و استفاده از داده‌های زبانی که برای کاربردهای عمومی جمع‌آوری شده‌اند، مفید واقع نمی‌شوند. در این پروژه سعی شده است تا با استفاده از یک ترکیب مناسب از مدل‌های زبانی نه تنها از مزایای مدل‌های مبتنی بر گرامر مستقل از متن در سطح معنا بهره گرفته شود، بلکه با استفاده از مدل‌های آماری از انعطاف‌پذیری بیشتری برخوردار شود ولی از آنجا که گرامر تعریف شده به عنوان ساختارِ زیربنایی بر کلاس‌های معنایی استوار است، اجزاء پایه مدل n-gram نیز از نوع کلاس هستند. که با استفاده از نتایج پروژه نمایشی[1] کلاس‌های مورد نیاز استخراج شده و کلمات دیگری – بر اساس تجربه متخصّصین و به عنوان کلمات مترادف، به لحاظ معنایی- به هر مجموعه اضافه شد. لازم به ذکر است تعریف مترادف معنایی در این پروژه مطابق با تعریف Wang در [۱۵] یکسان بوده و کلمات مترادف از نظر نحوی نیز باید از یک خانواده بوده و به عبارت بهتر، قابل جایگزینی با یکدیگر باشند. بر اساس این ساختار، جابجایی در وضعیت‌های معنایی تا انتهای عبارت، متضمن ضرب احتمالات جابجایی مارکف کلاس‌ها در سطوح متفاوت مدل می‌باشد. توصیف آماری این جابجایی را می‌توان به شکل آتی بیان کرد:

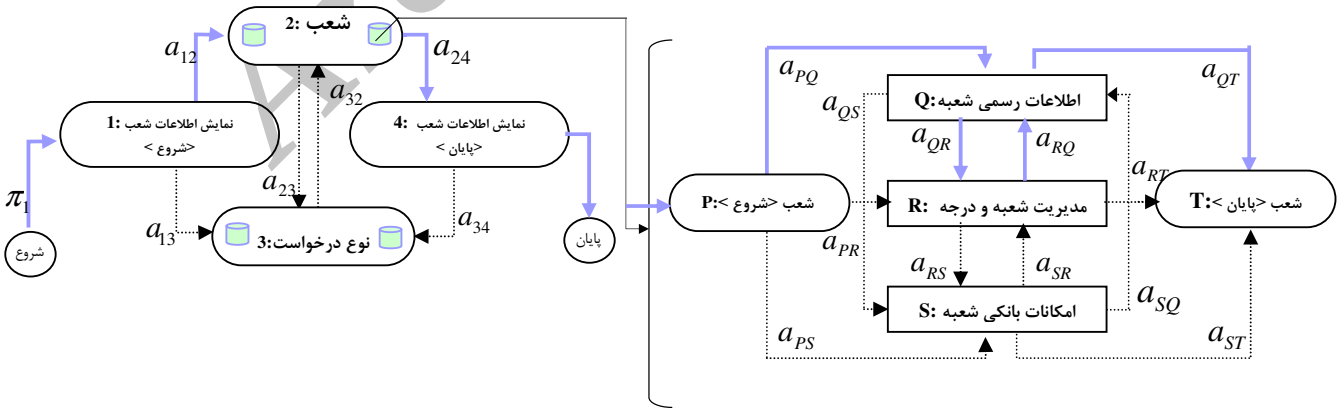
$$\Pr(M)=\prod_{i=1}^{|M|+1}\Pr(C_M(i)\mid C_M(i-1))\times\Pr(M(i)) \qquad (۱)$$

در این رابطه $|M|$ برابر تعداد شکاف‌های موجود در M بوده و $C_M(i)$ نام iامین شکاف در عبارت M و $M(i)$ زیرساختار پرکننده iامین شکاف در عبارت M است (اقتباس از [۱۵]) .

این رابطه، احتمال ساختار معنایی مورد استفاده در یک عبارت را بصورت بازگشتی محاسبه نموده و می‌تواند بدون در نظر گرفتن یک ساختار مجزا در رابطه، احتمال حرکت در زیرساختارهای شکاف را نیز در محاسبات دخیل کند. بخشی از ساختار طراحی شده برای کاربرد دریافت اطلاعات شعب در شکل (۶) ترسیم شده است. براساس ساختار موصوف در شکل (۶) مقدار احتمال حرکت در این شبکه برای پارس عبارت " لطفاً تلفن شعبه مرکزی استان تهران رو بدید" در رابطه (۲) محاسبه شده است.

$$\begin{aligned}\Pr(M)&=\Pr(Branch\ Info.)\\&\times\Pr(Branches\mid<S>;Branch\ Info.)\\&\times\Pr(O.\ Info\mid<S>;Branches)\\&\times\Pr(Br.\ Rank\mid O.\ Info.;Branches)\\&\times\Pr(O.\ Info.\mid Br.\ Rank;Branches)\\&\times\Pr(<\backslash S>\mid O.\ Info;Branches)\\&\times\Pr(<\backslash S>\mid Branches;Branch\ Info.)\\&=\pi_1\ a_{12}\ a_{PQ}\ a_{QR}\ a_{RQ}\ a_{QT}\ a_{24}\end{aligned} \qquad (۲)$$

برای محاسبه این احتمالات و حرکت بر روی شبکه تشکیل شده از مدل‌های جایگزین با دریافت جمله " لطفاً تلفن شعبه مرکزی استان تهران رو بدید" به ترتیب هر یک از فرمان‌های سطح صفر مورد بررسی قرار می‌گیرند و تلاش برای پارس جمله با ساختار آن فرمان ادامه می‌یابد، در این صورت، با بروز یک نقطه کور و یا تلّه، پارس متوقف شده و فرمان بعدی در سطح صفر انتخاب می‌شود. در شکل (۶) فرایند انتخاب فرمان سطح صفر به صورت مجزا به نمایش درنیامده و پارس در فرمان مناسب که "دریافت اطلاعات شعب " با شکاف شعب بوده، نمایش داده شده است.



شکل (۶): مدل ترکیبی برای فرمان دریافت اطلاعات شعب و مسیر حرکت مشخص شده در مدل برای جمله "لطفاً تلفن شعبه مرکزی استان تهران رو بدید."

[1] Demo

پس از عبور از عبارت "لطفاً" که می‌تواند در عبارت مقدمه جای گیرد جریان پارس به شکاف "شعب" می‌رسد، از آنجا که شکاف شعب دارای زیرشبکه جاسازی شده است، آدرس شکاف شعب در پشته قرار گرفته و کنترل به داخل زیرشبکه انتقال می‌یابد. تلفن به عنوان بخش بعدی در داخل جزء "اطلاعات رسمی شعبه" جای می‌گیرد و در ادامه کلمه مرکزی در داخل کلاس "مدیریت و درجه شعبه". کلاس مدیریت و درجه شعبه، شامل اطّلاعات مدیریتی شعبه و درجه اهمّیّت آن است. امّا استان محّل استقرار شعبه جزء خصوصیّات رسمی و اداری شعبه است که در کلاس اطلاعات رسمی شعبه جای می‌گیرد و کنترل را به این جزء برمی‌گرداند. در نهایت جز آخر که شامل عبارت "رو بدید" است در این زیرشبکه پارس نمی‌شود و نیاز به بازگشت را در ذهن متواتر می‌سازد. در بازگشت عبارت "رو بدید" در قالب درخواست اتصّال و به عنوان مؤخره شبکه "دریافت اطلاعات شعب" مورد پذیرش قرار می‌گیرد.

۴- پیاده‌سازی سیستم

با توّجه به اینکه پیش از این نمونه سیستم محاوره فعّالی برای زبان فارسی آماده نشده بود و سیستم فوق الذکر (که واحد درک زبان گفتار آن در این مقاله مورد بررسی قرار گرفته است) اوّلین نمونه آماده شده می‌باشد مشکلات زیادی در مسیر آماده سازی آن وجود داشت که بزرگترین آنها عدم وجود داده‌های آماده برای یک سیستم تلفنی با بستر محاوره بانکی بود. به همین دلیل در ابتدا یک سیستم نمایشی برای این منظور آماده و مراحل بازشناسی و درک گفتار در آن به ناظر انسانی سپرده شد؛ با راه‌اندازی این سیستم مکالمات در طول اجرای سیستم ضبط و بخش‌های مورد نیاز استخراج می‌شدند.

مهّمترین اطلاعاتی که از اجرای این سیستم بدست آمد عبارت بودند از :

- لیست کلمات مورد استفاده کاربران
- قالب گرامر معنایی
- اولویت کاربری منوهای مختلف
- اشتباهات و ابهامات در طراحی و راه‌اندازی

در ادامه با گسترش کلاس‌های معنایی، مجموعه کلمات استخراج شده به تمام موارد احتمالی تعمیم یافتند.

در دومین مرحله شبکه‌های گرامری مورد نیاز به صورت دستی استخراج شده و در یک روند بازگشتی در مقایسه با کلاس‌های معنایی تکامل یافته تا به وضعیّتی ارتقا یافتند که در ارزیابی دستی با تعداد ۳۵ قاعده گرامری کلی و تعداد ۲۰ زیرقاعده، برای مدلسازی زیرساختارهای داخل شکاف‌ها، قابلیت پاسخگویی به ۹۷ درصد موارد را در یک مجموعه ۱۵۰ موردی داشتند. این قواعد در قالب یک شبکه بزرگ با معماری معرفی شده در بخش ۳-۱ جای گرفتند که وظیفه آنها بهبود خطای خروجی بازشناسی گفتار و درک گفتار و همچنین تعیین هدف بود. در ادامه این مدل باید

با مدل مبتنی بر n-gram ترکیب شود، این ترکیب می‌توانست در افزایش قابلیّت انعطاف و قدرت رفع ابهام سیستم مؤثّر باشد.

به همین منظور در مرحله پیاده‌سازی از یک شبکه مبدّل بازگشتی[1] با یال‌های دارای ارزش احتمالاتی استفاده شد. این شبکه به عنوان یک ساختار پایه ای بر اساس ماشین متناهی الحالت که می تواند یک گرامر مستقل از متن را مدل نماید؛ [۱۶] بهترین ساختار برای این پیاده‌سازی بود. استفاده از شبکه‌های مبدل بازگشتی در درک گفتار دارای سابقه طولانی بوده و ساختار پایه در واحد phoenix به عنوان واحد درک گفتار در پروژه Communicator دانشگاه‌های CMU و Colorado می‌باشد[۱۷].

سیستم فوق در اوّلین مرحله در آزمایشگاه پردازش گفتار دانشگاه صنعتی شریف نصب و راه‌اندازی شد تا با دریافت مجموعه بهترین فرضیات بازشناسی شده توسط موتور بازشناسی گفتار نویسا تا حداکثر ۱۰۰ مورد، و عبور کلیه فرضیّات از ساختار شبکه‌ای خود، سه نمونه از بهترین موارد را انتخاب و برای پارس جزئی و تعیین هدف مورد استفاده قرار دهد.

برای ارزیابی این سیستم در تعیین اهداف کاربران از ۳ مجموعه مکالمه ۳۶ تایی با گوینده متفاوت استفاده شد که نتایج حاکی از تعیین هدف با دقت ۹۶/۸ درصد بود.

در ادامه برای بررسی میزان تأثّیر مثبت این مدل در بازشناسی گفتار پیوسته با جایگزین نمودن این مدل با مدل‌های مورد استفاده در موتور بازشناسی گفتار پیوسته میزان بهبود بازشناسی در مقایسه با حالتی که هیچ یک از مدل‌ها در بازشناسی مورد استفاده قرار نگرفتند مورد بررسی قرار گرفت که حاکی از نرخ بهبودی، بیش از ۲۰ درصد بود، که در جدول (۱) آورده شده است.

جدول(۱): نتایج ارزیابی مدل پیشنهادی در بهبود خروجی بازشناسی گفتار پیوسته در مقایسه با عدم استفاده از هیچ مدل زبانی به استثنای مدل آکوستیک

	مجموعه داده ۱	مجموعه داده ۲	مجموعه داده ۳	بدون مدل زبانی
Accuracy	۷۸،۸۲	۷۵،۳۰	۷۹،۵۱	۵۵،۳۳
Correctness	۸۰،۶۷	۷۷،۹۶	۸۰،۴۶	۶۱،۶۹

۵- نتیجه‌گیری

سیستم‌های بازشناسی گفتار پیوسته معمولاً از مدل‌های زبانی برای بهبود خروجی خود بهره می‌گیرند امّا بزرگترین مشکل در این زمینه حجم کم داده‌ها و یا عدم وجود آنها برای یک زمینه کاری است که

[1] Recursive Transition Network

روش‌های صرفاً مبتنی بر داده را با مشکل مواجه می‌کند؛ از طرفی زمانبر بودن فرایند تولید قواعد گرامری مناسب در روش‌های مبتنی بر دانش و نیاز به تجربه برای یک بستر خاص، این دسته از روش‌ها را با مشکل مواجه کرده است. با توجّه به مشکلات بیان شده برای روش‌های مبتنی بر داده و دانش ارائه یک مدل ترکیبی مانند آنچه در این مقاله آورده شده، می‌تواند راه‌کار مفیدی در مواجه با این مشکلات باشد. ترکیب مدل‌های زبانی در سیستم‌های بازشناسی گفتار معمولاً تا لایه نحو صورت می‌گیرد، امّا از ویژگی‌های خاص مدل ارائه شده می‌توان به این نکته اشاره کرد که ترکیب مدل‌ها تا سطح معنا در این مورد توانسته کاملاً کارا باشد. ارزیابی این ترکیب در مقایسه با حالت بدون مدل نشان‌دهنده پیشرفت شگرفی در این عرصه است که می‌تواند باب جدیدی را در زمینه ترکیب مدل‌های زبانی با کاربرد خاص باز کند. البتّه ذکر این نکته ضروری است که ترکیب مدل‌ها تا سطح معنا بیشتر برای کاربردهایی میّسر است که زمینه و دامنه فعّالیّت سیستم بازشناسی گفتار محدود و معیّن است؛ مانند بستر محاوره.

مسئله دیگری که بیش از سایرین رخ می‌نماید میزان انعطاف و مقاومت مدل پیشنهادی است که می‌تواند در مقابل اشتباهات نحوی و فی‌البداهه‌گویی کاربران به سهولت هدف مورد نظر را تعیین کند؛ چرا که حرکت به سطوح بالاتر از نحو، زمینه عمومی‌سازی در درک را فراهم می‌آورد همچنین بستر پیاده‌سازی مبتنی بر گرامر مستقل از متن که برای پیاده‌سازی‌های سریع تجاری [۱۸] مناسب است زمینه بروزسازی سریع را فراهم می‌آورد.

در این مدل برای جلوگیری از عمومی‌سازی بیش از حد از ترکیب با مدل‌های آماری استفاده می‌شود تا موارد نادرستی که از پالایش سطح اوّل عبور کرده‌اند در این مرحله پالایش شده و در عین حال نزدیکترین ورودی به عبارت بیان شده توسط کاربر با وجود یک اشتباه جزئی رد نخواهد شد و تنها در میان انتخاب‌ها جابجا خواهد شد.

۶- پیشنهاداتی برای فعّالیّت آتی

در جهت ادامه فعّالیّت در این زمینه و زمینه‌های مرتبط موارد زیر پیشنهاد می‌گردد:

- فعّالیّت بر روی بسترهای محاوره‌ای در زبان فارسی برای کاربردهای مختلف
- جمع‌آوری داده وتهیّه پیکره‌های گفتار محاوره‌ای
- ترکیب مدل‌های زبانی در سطوح بالاتر از نحو برای موارد استفاده متفاوت
- تهیّه سیستم‌های استخراج گرامر مستقل از متن معنایی برای کاربرد محاوره.
- مناسب‌سازی نمونه سیستم بازشناسی گفتار با اختصاص ساختار زبانی متناسب با کاربرد.

مراجع

[1] Seneff, S., Polifroni, J., "*Dialogue Management in the mercury flight reservation system*", In Proc. Of ANLP_NAACL Workshop on Conversational Systems, Seattle, Washington, 2000

[2] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T., Hetherington, L., "*Jupiter: a Telephone-based Conversational Interface for Weather Information*", IEEE Transaction on Speech and Audio Processing, 8 (1), pp. 85-96, 2006

[3] Ehsani, F., Bernstein, J., Najmi, A., "*An interactive dialogue system for learning Japanese*", Speech Communication, Vol. 30, pp. 167-177, 2000

[4] Bernsen, N. O., "*On-line user modeling in a mobile spoken dialogue system*", In Proc. Of Eurospeech, Geneva, Switzerland, pp. 737-740, 2003

[5] Huang, Q., Cox, S., "*Automatic Call-Routing without Transcriptions*", In Proc. Of Eurospeech, Geneva, Switzerland, pp. 1909-1912, 2003

[6] Zue, V., Glass, J. R., "*Conversational Interfaces: Advances and Challenges*", In Proc. Of The IEEE, 88 (8), pp. 1166-1180, 2000

[7] Editorial Board, "Introduction to the Special Issues on Spoken Language Understanding in Conversational Systems", Speech Communication, Vol. 48, pp. 233-238, 2006

[8] Hampel, T., "*Usability of Speech Dialogue Systems*", Springer-Verlag series on Signals and Communication Technology, Berlin, 2008

[9] Möller, S., "*Quality of Telephone-based Spoken Dialogue Systems*", Springer, Berlin, 2005

[10] Corporate Communications ,"Technology and Innovation at Fidelity Investments®", Access Date: 11/19/2007, http://personal.fidelity.com/myfidelity/InsideFidelity/NewsCenter/mediadocs/techback.html

[11] PRNNEWS, "AIR Tran Airways selects Speech to Automate customer flight information ", Access Date: 8/28/2008, http://www.prnewswire.com/airtran/20000508a.shtml

[12] Kotelly, B., "*The Art and Business of Speech Recognition: Creating the Noble Voice*", Addison-Wesley, 2003

[13] Seneff, S., "*TINA: A Natural Language System for Spoken Language Applications*", Computational Linguistics, 18(1), pp. 61-86, 1992

[14] Yu, D., Cheng Ju, Y., Wang, Y., Acero, A., "*N-Gram Based Filler Model for Robust Grammar Authoring*", In Proc. Of ICASSP 2006, Toulouse, France, 2006

[15] Wang, Y., Acero, A., "*Rapid development of spoken language understanding grammars*", Speech Communication, Vol. 48, pp. 390-416, 2006

[16] Jurafsky, D., Martin, H. , "*Speech and Language Processing: An introduction to Natural Language Processing, Computational linguistics and speech Recognition*", Prentice-Hall, pp. 294-299, 2000

[17] Ward, W., "*Recent improvements in the CMU Spoken Language Understanding System*", Human Language Technology Workshop, Plainsboro, NJ, USA, 1994

[18] Dev Ault, D., Traum, D., Arstein, R. "*Making Grammar Based Generation Easier to Deploy in Dialogue Systems*", In Proc. Of The 9th SIGdial Workshop on Discourse and Dialogue, pp. 198-207, 2008