برچسب‌زنی نقش معنایی جملات فارسی با رویکرد یادگیری مبتنی بر حافظه

آزاده کامل قالیباف[1]،سعید راحتی قوچانی[1]، اعظم استاجی[2]
[1]دانشکدهٔ فنی و مهندسی، گروه هوش مصنوعی، دانشگاه آزاد اسلامی مشهد
[2]دانشکده ادبیات و علوم انسانی، گروه زبان‌شناسی همگانی، دانشگاه فردوسی مشهد

چکیده:

استخراج نقش‌های معنایی یکی از گام‌های اصلی در بازنمایی معنی متن است. نقش‌های معنایی، ارتباط معنایی بین فعل و آرگومان‌های آن در جمله را مشخص می‌کنند. در این مقاله یک سیستم برچسب‌زنی خودکار نقش معنایی برای متون فارسی با رویکرد یادگیری ماشین ارائه شده است. مجموعه داده‌های مورد نیاز سیستم بخشی از پیکرهٔ متنی زبان فارسی است که توسط پژوهشکدهٔ پردازش هوشمند علائم تهیّه و برچسب‌گذاری شده است. سیستم پیشنهادی از دو مرحلهٔ تشکیل شده؛ در مرحلهٔ اوّل با تجزیهٔ نحوی جمله، حد و مرز سازه و همچنین نوع گروه نحوی این اجزا در جمله مشخص می‌شود. این اطّلاعات به‌عنوان ورودی در مرحلهٔ دوم مورد استفاده قرار می‌گیرد. مرحلهٔ دوم سیستم مربوط به تخصیص نقش‌های معنایی مناسب به سازه‌های مشخص شده در مرحلهٔ قبل می‌باشد. برای این منظور از ویژگی‌های نحوی و ساختاری هر سازه، بهره گرفته می‌شود. نتایج به‌دست آمده نشان‌دهندهٔ ۸۱.۶٪ =F1 برای زیر سیستم تجزیهٔ نحوی، و ۸۷.۴٪ =F1 برای زیرسیستم برچسب‌زنی معنایی درحالتی‌که ورودی‌های سیستم به‌صورت دستی تصحیح شده باشند. همچنین کارآیی کل سیستم ۷۳.۸٪ =F1 را برای سیستم کامل برچسب‌زنی معنایی، یعنی تجزیهٔ نحوی و تخصیص نقش نشان می‌دهد. نتایج به‌دست آمده حاکی از آن است که می‌توان از یک پیکرهٔ آموزشی کوچک ۱۳۰۰ کلمه‌ای نتایج قابل قبولی به‌دست آورد.

واژه‌های کلیدی: پردازش زبان طبیعی، برچسب‌زنی معنایی، تجزیهٔ سطحی معنایی، تجزیهٔ سطحی نحوی، یادگیری مبتنی بر حافظه.

۱- مقدمه

در سال‌های اخیر تجزیهٔ معنایی زبان طبیعی بسیار مورد توجّه قرار گرفته و به یک بحث کلیدی در حوزهٔ استخراج اطلاعات، پرسش و پاسخ، خلاصه‌سازی و به‌طور کلی در تمام کاربردهای NLP که نیازمند نوعی تفسیر معنایی هستند، تبدیل شده است [۱].

برچسب‌زنی نقش معنایی[1] یا تجزیهٔ سطحی معنایی[2] را می‌توان تحلیل معنایی متن در سطح جمله دانست، که در آن فعل جمله، مشخص‌کنندهٔ رویداد واقع شده، و سایر اجزای جمله هر یک نقشی در ارتباط با این رویداد می‌پذیرند. به این ترتیب روابطی نظیر چه کسی، چه چیزی را برای چه کسی و در کجا و چه موقع انجام داده است در جمله مشخص می‌شود [۲].

الگوریتم‌های مختلفی برای تخصیص خودکار نقش‌های معنایی در جملات انگلیسی ارایه شده است. امّا در مورد چگونگی عملکرد این الگوریتم‌ها در سایر زبان‌ها و در مورد نحوهٔ آموزش این الگوریتم‌ها در مواردی که مجموعه‌های بزرگ برچسب‌گذاری شده در دسترس نباشد، صحبتی به میان نیامده است [۳].

به طور کلی روش‌های موجود را برای تعیین نقش‌های معنایی، می‌توان به دو دسته روش‌های مبتنی بر قواعد و روش‌های مبتنی بر یادگیری (آماری) تقسیم‌بندی کرد.

در روش‌های مبتنی بر قواعد، تحلیل معنایی متن به‌کمک لغت‌نامه‌ها، گرامرها و سایر منابع معنایی انجام می‌گیرد؛ این منابع بیشتر به‌صورت دستی تهیه می‌شوند[۱].

[1] Semantic Role Labeling
[2] Shallow Semantic Parsing

به‌کارگیری این رویکرد، مستلزم ایجاد گرامرهای دستی وسیع و فراگیر برای انواع جملات با ساختارهای دستوری متفاوت در زبان است که با توجّه به تنوع و پیچیدگی جملات و افعال به‌کار رفته در آن‌ها، به‌طور معمول این روش‌ها بخش محدودی از زبان را مورد توجّه قرار می‌دهند و برای کاربردهای خاص همچون تعیین نقش‌های معنایی در سیستم محاورهٔ فرودگاه مناسب‌اند [۳].

سیستم‌هایی که در همین اواخر به‌منظور برچسب‌زنی نقش‌های معنایی توسعه یافته‌اند، از روش‌های متنوع یادگیری ماشین استفاده می‌کنند[۴]،[۵] و[۶]. در این دسته روش‌ها از پیکره‌هایی که جملات آن‌ها به‌صورت دستی برچسب‌گذاری معنایی شده‌اند، جهت استخراج قواعد به‌صورت خودکار استفاده می‌شود.

مهم‌ترین پیکره‌هایی که در زبان انگلیسی برای این منظور ایجاد شده‌اند عبارتند از FrameNet [7] و PropBank [8]، که هر یک روش بازنمایی معنایی خاص خود را دارند. امّا متأسفّانه زبان فارسی فاقد چنین پیکره‌های معنایی است و به‌همین دلیل برخلاف توجّه زیاد به برچسب‌زنی خودکار نقش معنایی در سال‌های اخیر، کارهای انجام شده بیشتر بر روی پیکره‌های انگلیسی بوده است.

سیستم برچسب‌زنی معنایی پیشنهاد شده در این مقاله از دو مرحله تشکیل شده است. ابتدا آرگومان‌های فعل، به‌وسیلهٔ یک تجزیه‌گر نحوی سطحی یا قطعه‌بند، شناسایی می‌شود؛ سپس در مرحلهٔ بعد با توجّه به رابطهٔ معنایی که این آرگومان‌ها با فعل جمله دارند، نقش معنایی مناسب به آن‌ها تخصیص داده می‌شود. در پیاده‌سازی هر دو مرحلهٔ از دسته‌بند یادگیری مبتنی بر حافظه، با الگوریتم یادگیری با سرپرستی بر روی یک پیکرهٔ برچسب‌گذاری شده به‌صورت دستی، استفاده شده است.

از آن‌جا که تهیهٔ پیکره‌ای جامع برای زبان فارسی، که شامل انواع جملات با ساختارها و افعال مختلف باشد بسیار مشکل و زمان‌بر است. ناگزیر کار را بر روی پنجاه فعل سادهٔ پربسامد فارسی محدود کرده و منابع لازم را برای آن فراهم نموده‌ایم. این منابع شامل مجموعه‌ای از جملات برچسب‌گذاری شده با نقش‌های معنایی و مجموعهٔ افعال طبقه‌بندی شده براساس این نقش‌ها می‌باشد. ذکر این نکته حائز اهمیت است که با جمع‌آوری داده‌های بیشتر و تکمیل این مجموعه می‌توان این سیستم را برای کل افعال زبان تعمیم داد.

ویژگی اصلی دسته‌بند یادگیری مبتنی بر حافظه برای استفاده در مسائل NLP آن است که در این روش برخلاف روش‌های چکیده‌سازی، نمونه‌های آموزش عیناً در حافظه ذخیره‌سازی می‌شوند و به این ترتیب موارد استثنا و بی‌قاعدگی‌های زبان نیز در چرخهٔ یادگیری مورد توجّه قرار می‌گیرد که این نکته در مسائل زبانی از اهمّیت بالایی برخوردار است. این الگوریتم درحالتی‌که ویژگی‌ها به‌درستی انتخاب و وزن‌دهی شده باشند، بهترین عملکرد را خواهد داشت [۹]. ما در این سیستم از اطّلاعات نحوی آرگومان‌ها به‌عنوان مجموعه ویژگی استفاده کرده‌ایم. امّا از آن‌جایی‌که تا به حال هیچ تجزیه‌گر خودکاری برای تجزیهٔ نحوی جملات فارسی تولید نشده است، بر آن شدیم تا مرحلهٔ اوّل سیستم را به طراحی یک تجزیه‌گر نحوی سطحی اختصاص دهیم.

تجزیهٔ نحوی سطحی[1] یا جزئی[2] در سال‌های اخیر به‌عنوان جایگزین مناسبی برای تجزیهٔ نحوی کامل[3] شناخته شده است. یک تجزیه‌گر سطحی، سعی دارد تا جمله را به بخش‌هایی[4] تقسیم کند که هر بخش به یک واحد نحوی (مانند عبارت اسمی، فعلی، حرف اضافه‌ای) منسوب باشد. این عمل شامل شناسایی محدودهٔ واحدها و تعیین نوع گروه نحوی آن‌ها می‌باشد. تجزیهٔ سطحی، به‌سادگی آموزش می‌پذیرد، سریع، مقاوم و با ابهام کمتر است. چنین ویژگی‌هایی آن را تبدیل به یک انتخاب خوب در مقابل تجزیهٔ کامل کرده است [۱۰].

ادامهٔ مقاله به‌صورت زیر سازمان‌دهی شده است: ابتدا در بخش دو به بررسی مشخصات پیکرهٔ متنی و نحوهٔ برچسب‌گذاری آن با مجموعهٔ دوازده نقش معنایی می‌پردازیم. در بخش سه به توصیف ساختار کلی سیستم برچسب‌زنی نقش معنایی و بررسی اجزای این سیستم با جزئیات پیاده‌سازی آن اختصاص دارد. نتایج تجربی و تحلیل نتایج در بخش چهار آورده شده است. درنهایت نتیجه‌گیری حاصل از این مطالعه در بخش پنج ارائه شده است.

۲- پیکرهٔ متنی و برچسب‌گذاری معنایی

در این تحقیق بخشی از پیکرهٔ متنی زبان فارسی، که توسط پژوهشکدهٔ پردازش هوشمند علائم تهیهٔ و برچسب‌گذاری شده، استفاده می‌شود [۱۵]. این پیکره، شامل ۱۰ میلیون کلمه می‌باشد که با مجموعه غنی از برچسب‌ها، شامل ۱۸۶ برچسب مختلف، برچسب‌گذاری شده است. ازآنجایی‌که برچسب‌زنی خودکار بر اساس یادگیری ماشینی انجام می‌شود، در نظر گرفتن مجموعهٔ بزرگی از برچسب‌ها،

[1] Shallow Syntactic Parsing
[2] Partial Parsing
[3] Full Syntactic Parsing
[4] Chunck

نیازمند حجم دادهٔ بیشتری برای آموزش می‌باشد. به همین منظور در بخش ۳-۲ مجموعهٔ برچسب‌ها را با در نظر گرفتن شرایط و نیاز سیستم به چهارده مورد کاهش داده‌ایم.

همان‌طور که گفته شد ایجاد پیکرهٔ برچسب زده شدهٔ معنایی برای فارسی مورد بی‌توجهی واقع شده است. در اینجا ما بخشی از پیکرهٔ ساده شدهٔ زبان فارسی را انتخاب کرده و با برچسب‌های نحوی و معنایی مورد نیاز، آن را برچسب‌گذاری کرده‌ایم. پیکرهٔ تولیدشدهٔ کوچک شامل جملات با ساختارهای متنوع می‌باشد. در این بخش ابتدا به توصیف نقش‌های معنایی استفاده شده در سیستم می‌پردازیم و سپس داده‌های به‌کار رفته در آزمایش‌ها را معرفی می‌کنیم.

۲-۱ نقش‌های معنایی

نقش‌های معنایی بیان‌کنندهٔ روابط میان یک فعل و متمم‌های آن است. استخراج این نقش‌ها یکی از مراحل اصلی در بازنمایی معنایی متن به‌شمار می‌رود. از مشکلات عمدهٔ موجود بر سر راه بازنمایی دانش معنایی، با نقش‌های معنایی می‌توان به عدم وجود چهارچوب مشخصّی جهت تعیین مجموعه نقش‌های معنایی اشاره کرد. تعداد و نوع نقش‌های معنایی بسته به دیدگاه مفهومی نسبت به زبان و کاربرد آن متفاوت است، به‌طوری که در یک سمت طیف می‌توان به تئوری‌هایی اشاره کرد که وجود دو نقش معنایی Proto-Agent و Proto-Patient را کافی می‌داند [۱۲] و در سمت دیگر طیف می‌توان تئوری‌هایی را مّد نظر قرار داد که برای هر فعل، نقش معنایی خاصی را در نظر گرفته‌اند [۱۳]، در میانهٔ این دو طیف نیز تئوری‌هایی به چشم می‌خورد که یک مجموعه کوچک از نقش‌های انتزاعی، در حدود ده نقش را پیشنهاد کرده‌اند، مانند مجموعه نقش‌های Fillmore [۱۴].

هر کاربرد با توجّه به وسعت دامنه و اطلاعات پردازشی مورد نیاز خود، می‌تواند از نقش‌های معنایی متفاوتی جهت کسب نتیجهٔ بهتر استفاده نماید. به‌عنوان مثال در سیستم طراحی شده برای رزرو بلیط در یک آژانس مسافرتی از مجموعه نقش‌های مختص به فعل مانند From_City و To_City استفاده شده است[۳]. در چنین سیستم‌هایی با توجّه به محدودیت رویدادهای موجود در سیستم، استفاده از چنین نقش‌هایی منجر به نتایج مطلوبی خواهد شد.

در این مقاله با توجّه به نوع افعال و پیکرهٔ انتخاب شده، و همچنین نوع کاربردی که از سیستم در نظر داریم، یک مجموعهٔ دوازده‌تایی از نقش‌های معنایی تعریف کرده‌ایم. به‌طوری که داده‌های پیکره را پوشش داده، و به‌خوبی جواب‌گوی نیاز سیستم باشد. این مجموعه در (جدول ۱) نشان داده شده است.

جدول (۱) –نقش‌های معنایی

نقش معنایی	تعریف نقش
Agent	موجودیتی که انجام‌دهنده یک کار یا سبب یک اتفاق است.
Patient	موجودیتی که از وقوع فعل تاثیر پذیرفته باشد.
Source	محل یا موجودیتی عمل انتقال از سمت آن صورت گرفته باشد.
Goal	محل یا موجودیتی که عمل انتقال به سمت آن صورت گرفته باشد.
Topic	عبارتی که پیام منتقل شده بوسیله فعل را بیان می‌کند.
Percept	آنچه در افعال شناختی درک می‌شود.
Instrument	ماده و ابزار انجام فعل.
Beneficiary	موجود زنده ای که از عمل انجام شده به نحوی سود برده باشد.
Time	زمانی که وقوع اتفاق یا انجام عمل در آن زمان واقع شده است.
Location	جایی (فیزیکی یا غیر فیزیکی)که وقوع اتفاق یا انجام عمل در آنجا واقع شده است.
Manner	چگونگی وقوع یک اتفاق یا انجام عمل.
Reason	وقوع یک اتفاق که خود دلیل و هدف از وقوع یک اتفاق و یا انجام عمل باشد.

مجموعه نقش‌های به‌کار رفته را می‌توان به دو دسته نقش های اصلی و نقش‌های عمومی تقسیم کرد. نقش‌های اصلی نقش‌هایی هستند که با توجّه به معنا و ظرفیت افعال مشخص می‌شوند؛ درحالی‌که نقش‌های عمومی حالت قید دارند و به‌صورت اختیاری و برای ارائه توضیحات بیشتر با هر فعلی استفاده می‌شوند. برای مثال در جملهٔ "این خانه را مرحوم پدرش به‌مبلغ هفتصد تومان از تاجر یزدی خریده بود." نقش‌های اصلی و عمومی به‌صورت زیر خواهد بود.

این خانه را	Patient	اصلی
مرحوم پدرش	Agent	اصلی
به مبلغ ۷۰۰ تومان	manner	عمومی
از تاجر یزدی	Source	اصلی
خریده بود.	predicate	

۲-۲ مجموعه داده‌های آموزش و آزمایش

در سیستم حاضر تشخیص نقش‌های معنایی را به مجموعهٔ پنجاه فعل سادهٔ (غیر اسنادی و غیر سبک) متداول و پرتکرار محدود کرده‌ایم. از این رو پیکرهٔ متنی که برای آموزش سیستم استفاده شده تنها شامل جملاتی با این پنجاه فعل می‌باشد. این پیکره با انتخاب و استخراج جملاتی از پیکرهٔ متنی زبان فارسی که فعل اصلی آنها یکی از این پنجاه فعل بوده شکل گرفته است. فراوانی افعال انتخاب شده در پیکرهٔ جدید در بازهٔ ۱۰ تا ۶۰، با میانگین ۳۵ می‌باشد.

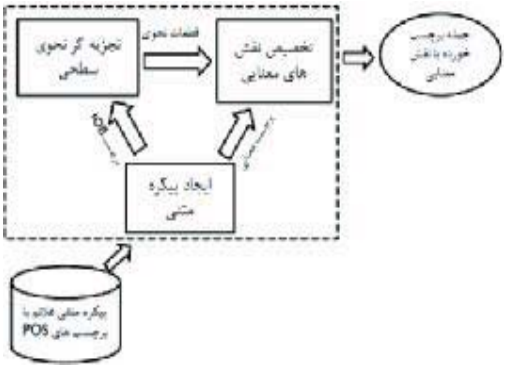
در انتخاب افعال سعی کرده‌ایم تنوع ساختارهای نحوی و ظرفیتی مانند افعال با یک، دو و سه آرگومان و همچنین افعال با الگوهای مختلف اتصال آرگومان درنظر گرفته شود. به این ترتیب پیکره‌ای متشکل از دوهزار جمله تهیه و با نظارت و راهنمایی یک زبان‌شناس خبره برچسب‌گذاری نقش‌های معنایی در آن انجام شد. از آن‌جا که در این پروژه بیشتر، سازه‌های اصلی جمله موردنظر است، در مواردی که جملات مرکب دارای بندهای تودرتو و پیچیده بوده‌اند، بندهای توصیفگر و غیرضروری برای کاهش پیچیدگی حذف شده‌اند. همچنین به دلیل استقلال نحوی و معنایی جملات مرکب همپایه، این جملات تفکیک و به‌صورت دوجملهٔ مستقل در نظر گرفته می‌شوند. اما در مورد جملات مرکب ناهم‌پایه که از یک جملهٔ اصلی و یک یا چند جملهٔ وابسته تشکیل می‌شوند، جملات وابسته به‌طور معمول یک نقش نحوی در ارتباط با جملهٔ اصلی دارد. سپس برای تشکیل مجموعه داده‌های آموزش و آزمایش، جملات مربوط به هر فعل را به دو بخش تقسیم کردیم: ۷۰٪ جملات برای آموزش و ۳۰٪ برای آزمایش در نظر گرفته شده‌اند. به این ترتیب درمجموع ۱۳۰۰ جمله برای آموزش و هفصد جمله برای آزمایش خواهیم داشت. فهرست افعال به‌همراه کلاس معنایی آن‌ها در بخش ۳-۲ آورده شده است. در بخش بعد به توصیف رویکرد SRL پیشنهاد شده در این مقاله خواهیم پرداخت.

۳-رویکرد پیشنهادی

نحوهٔ عملکرد سیستم برچسب‌زنی نقش معنایی در (شکل ۱) نشان داده شده است.

همان‌طور که در (شکل ۱) دیده می‌شود، عمل تخصیص خودکار نقش معنایی از دو زیر سیستم تشکیل شده است:

(۱) شناسایی محدوده و نوع گروه نحوی آرگومان‌ها در جمله، که به‌وسیلهٔ یک بازنمایی نحوی مسطح[1] که حاصل خروجی یک قطعه‌بند نحوی یا تجزیه‌گر سطحی است، انجام می‌شود.



شکل (۱) معماری سیستم برچسب‌زنی نقش معنایی

(۲) برچسب‌زنی این آرگومان‌ها با نقش‌های معنایی مناسب. برای این منظور از یک روش یادگیری ماشین برای تشخیص نقش‌های معنایی مختلف استفاده می‌شود.

با توجّه به این‌که در هر دو سیستم از روش یادگیری مبتنی بر حافظه به‌عنوان دسته‌بند استفاده می‌شود، در بخش بعد به بررسی این الگوریتم یادگیری می‌پردازیم و در ادامه نحوهٔ عملکرد هر یک از این زیرسیستم‌ها به‌طور دقیق تشریح خواهد شد.

۳-۱- یادگیری مبتنی بر حافظه

روش یادگیری مبتنی بر حافظه[2] یا MBL که به‌روش یادگیری مبتنی بر مشابهت[3] یا مبتنی بر نمونه[4] نیز معروف است، بر اساس ذخیره‌سازی مجموعه داده‌های آموزش در حافظه و محاسبه میزان مشابهت دادهٔ جدید با داده های ذخیره شده، عمل دسته‌بندی را انجام می‌دهد. الگوریتم‌های مختلفی برای یادگیری مبتنی بر حافظه وجود دارد که وجهه تمایز آن‌ها در نحوهٔ محاسبهٔ معیار مشابهت، روش ذخیره‌سازی نمونه‌های آموزش در حافظه، و روش جستجو در حافظه می‌باشد. درواقع اساس روش تمام الگوریتم‌های MBL از الگوریتم K نزدیکترین همسایه مشتق شده، که با بهنه‌سازی سرعت و تعمیم الگوریتم K-NN بر روی داده‌های غیرعددی، روش MBL شکل گرفته است [۱۶].

یک سیستم MBL از دو مؤلّفهٔ تشکیل شده است: یک مؤلّفهٔ یادگیری که مبتنی بر حافظه است و یک مؤلّفهٔ

[1] Flat Represenatation
[2] Memory-Based Learning(MBL)
[3] Similarity-Based
[4] Instance-Based

کارآیی[1] که مبتنی بر مشابهت می‌باشد. الگوریتم‌های یادگیری مبتنی بر MBL عبارتند از IB1 و IGTree که IB1 از همان الگوریتم K_NN استفاده می‌کند و در IGTree نمونه‌ها در یک ساختار درختی ذخیره می‌شود و با یک روش جستجوی بالا به پایین، نزدیک‌ترین همسایه‌ها تخمین زده می‌شود [۱۶].

ساده‌ترین معیار مشابهت، معیار همپوشانی است که در آن مقادیر ویژگی‌های متناظر بین داده ورودی و داده‌های ذخیره شده در سیستم، مقایسه شده و به‌ازای هر اختلاف یک واحد به نرخ شباهت اضافه می‌شود. در رابطه زیر $\Delta(X,Y)$ اختلاف بین دو دادهٔ X و Y است:

$$\Delta(X,Y)=\sum_{i=1}^{n}\delta(x_i,y_i) \qquad (۱)$$

و δ اختلاف بین هر ویژگی از این دو داده است که مقدار آن از رابطهٔ دو مشخص می‌شود :

(۲)

$$\delta(x_i,y_i)=\begin{cases}\text{abs}\left(\dfrac{x_i-y_i}{\max_i-\min_i}\right) & \text{if numeric}\\ 0 & \text{if } x_i=y_i\\ 1 & \text{if } x_i\neq y_i\end{cases}$$

نرخ شباهت برای هر داده ورودی مقداری بین ۰ و تعداد ویژگی‌ها است. مقدار ۰ برای تطابق محض و با بیشتر شدن این مقدار میزان شباهت کاهش می‌یابد [۲]. با افزودن وزن به هر ویژگی برحسب میزان تأثیرگذاری آن، دسته‌ای دیگر از روش‌ها با عنوان روش‌های هم‌پوشانی وزن‌دار[2] را خواهیم داشت که از روش‌های وزن‌دهی مختلفی مانند بهره اطلاعاتی[3]، نرخ بهره یا چی دو[4] و غیره استفاده می‌شود [۱۶].

۳-۲- مرحلهٔ۱ : تجزیهٔ سطحی متن

هدف اصلی تجزیهٔ سطحی، تقسیم جمله به بخشهایی است که متعلق به گروه‌های نحوی مشخّصی (گروه اسمی، فعلی، حرف اضافه) می‌باشند. این بخش‌ها همان آرگومان‌های معنایی گزارهٔ داده شده (فعل جمله) هستند. جهت تشخیص محدوده گروه‌ها روش‌های برچسب‌گذاری مختلفی وجود دارد که از متداول‌ترین آن‌ها می‌توان به برچسب‌گذاری با براکت و برچسب‌گذاری به فرمت IOB اشاره کرد. در این مقاله برای تشخیص مرز بین گروه‌ها از فرمت IOB استفاده می‌شود که برای اوّلین‌بار در سال ۱۹۹۸ توسط Ratnaprkhi[17] به کار برده شد، در این روش از سه نوع برچسب برای کلمات سازندهٔ گروه استفاده می‌شود، برچسب B^5 برای کلمه ابتدای هر گروه، برچسب I^6 برای کلماتی که داخل گروه قرار دارند و برچسب O^7 برای کاراکتر انتهای جمله (به‌طور معمول نقطه)، همچنین نوع گروه نحوی مربوطه در ادامهٔ هر یک از این برچسب ها مشخص می‌شود. به‌عنوان مثال برچسب IOB برای جمله "سلطان حسنک را بالای دار فرستاد. " به‌صورت زیر است:

سلطان	B-NP
حسنک	B-NP
را	I-NP
بالای	B-PP
دار	I-PP
فرستاد	B-VP
.	O

چرخهٔ پیاده‌سازی این زیرسیستم را می‌توان به مراحل زیر تقسیم کرد:

۱- برچسب‌گذاری دستی سه‌هزار جمله از پیکرهٔ متنی زبان فارسی به‌روش IOB جهت آموزش دسته‌بند.

۲- استفاده از برچسب‌های POS به‌عنوان مجموعه ویژگی.

همان‌طور که در بخش دو اشاره شد، پیکرهٔ علائم با مجموعهٔ ۱۸۶ برچسب نحوی، حاشیه نویسی شده است. این مجموعه برچسب‌ها دربردارندهٔ مشخصات نحوی دقیق کلمات می‌باشند. امّا با توجّه به حجم کم داده‌ها در این سیستم و همچنین عدم نیاز مسئله به چنین مجموعهٔ وسیعی، مجموعه برچسب‌های پیکره را به مجموعه‌ای متشکل از دوازده برچسب نحوی ساده‌سازی می‌کنیم. مجموعه برچسب‌های استفاده شده، در جدول دو مشخص شده است. این کاهش تعداد برچسب‌ها بر اساس فراوانی برچسب‌ها و همچنین دسته‌بندی برچسب‌های مشابه از نظر معنایی و افزایش عمومیت برچسب‌ها به نوع نحوی کلی‌تر صورت گرفته است. به این ترتیب برچسب‌های جزئی‌تر تحت عنوان برچسب نحوی کلی‌تر دسته‌بندی شده‌اند.

[1] Performance
[2] Weighted Overlap
[3] Information Gain
[4] Gain Ratio or chi-square

[5] Begin
[6] Inpute
[7] Output

جدول (۲) – مجموعهٔ کاهش یافتهٔ برچسب‌ها

حرف اضافه	P	اسم	N
حرف ربط	CON	صفت	ADJ
جداکننده	DELM	قید	ADV
حرف تعریف	DET	فعل خاص	V_CR
سور	QUA	فعل اسنادی	V_PRE
کیفیت نما	SPEC	فعل کمکی	V_AUX

با توجّه به (جدول۲) از آن‌جا که فعل جمله تعیین کنندهٔ تعداد سازه‌ها و ساخت جمله است، برچسب فعل به‌طور دقیق‌تر با سه برچسب V_CR، V_PRE و V_AUX درنظر گرفته شده است.

مجموعه ویژگی هر کلمه، با لغزاندن یک پنجره به‌اندازهٔ پنج بر روی متن حاصل می‌شود. به این ترتیب در هر زمانPOS کلمه مورد بررسی در مرکز پنجره (مکان ۳) واقع می‌شود و از POS دو کلمهٔ قبل و دو کلمهٔ بعدی آن برای تشخیص برچسب IOB استفاده می‌شود[1]. یعنی به‌دنبال پیدا کردن الگویی از روی ترتیب POS کلمات، برای تشخیص برچسب IOB کلمهٔ مرکزی هستیم. به‌منظور استخراج خودکار این مجموعه ویژگی، برنامه‌ای در محیط VB.NET تهٔه شده است.

مزیّت این مجموعه‌ویژگی آن است که کلمه را در بافت متن درنظر می‌گیرد. همچنین با توجه به POS کلمهٔ هسته، نوع گروه نحوی آن تشخیص داده می‌شود.

برای هر جزء در جمله، بردار ویژگی به‌دست آمده از مرحلهٔ قبل به‌همراه برچسب IOB آن جهت آموزش به الگوریتم یادگیری مبتنی بر حافظه داده می‌شود. نتایج بدست آمده در بخش ۵ بررسی شده است.

۳-۳- مرحله ۲ : برچسب‌زنی معنایی

پس از شناسایی محدودهٔ قطعات معنایی در جمله، نوبت به تخصیص نقش‌های معنایی مناسب می‌رسد.

سیستم برچسب‌زنی نقش معنایی نیز شامل مراحل استخراج ویژگی و اجرای الگوریتم دسته‌بندی می‌باشد.

مجموعه‌ویژگی درنظر گرفته شده در این سیستم بیشتر با مرور سیستم‌های قبلی و بررسی تأثیر ویژگی‌های به‌کار رفته در آن‌ها بر روی ساختار جملات فارسی انتخاب شده‌اند. مجموعه ویژگی‌های برگزیده عبارتند از:

- **نوع گروه نحوی جزء جاری**[2] (مانند گروه اسمی، فعلی، حرف اضافه‌ای، قیدی، بند)
- **نوع گروه نحوی یک جزء قبل در جمله**
- **نوع گروه نحوی یک جزء بعد در جمله**
- **موقعیت جزء نسبت به فعل** (این ویژگی سه مقدار ۱، -۱ و ۰ را می‌گیرد. ۱ درصورتی‌که جزء قبل از فعل در جمله آمده باشد، -۱ در صورتی که بعد از فعل واقع شده باشد و ۰ برای خود فعل). ترتیب اصلی کلمات در جملات سادهٔ فارسی SOV و در جملات مرکب ناهمپایه SVO می‌باشد. در مجموعه داده‌های سیستم ۷۵٪ آرگومان‌ها قبل از فعل و ۲۵٪ آن‌ها بعد از فعل جمله واقع شده‌اند. به این ترنیب در زبان فارسی نیز همچون انگلیسی موقعیت آرگومان نشانهٔ خوبی برای شناسایی نقش معنایی آن می‌باشد. به‌طور مثال در همهٔ جملات نقش عامل پیش از فعل بوده درحالی‌که در ۳۰٪ موارد نقش پذیرا پس از فعل واقع شده که بیشتر به‌صورت بند متممی بوده است.
- **وجه فعل** (این ویژگی دو مقدار معلوم و مجهول را می‌گیرد). فعل‌های معلوم و مجهول در فارسی ساختار گزاره- آرگومان یکسانی دارند؛ اما ممکن است توابع گرامری(فاعل، مفعول،...) به مجموعه نقش‌های معنایی متفاوتی نگاشته شوند. از مجموع پیکرهٔ دوهزار جمله‌ای، ۱۷۴۰ مورد دارای ساختار جمله‌ای معلوم (۸۷٪)، و ۲۶۰ مورد دارای ساختار مجهول (۱۳٪) هستند.
- **کلاس معنایی فعل**. این کلاس‌ها براساس نقش‌های معنایی است که هر فعل می‌تواند بپذیرد. در ادامه نحوهٔ کلاس‌بندی افعال با توضیحات لازم ذکر شده است.

علی‌رغم دسته‌بندی‌های معنایی متنوعی که برای افعال انگلیسی وجود دارد، تاکنون دسته‌بندی مشابهٔ بر روی افعال فارسی انجام نشده است. ما در این سیستم یک دسته‌بندی، شامل هجده کلاس برای پنجاه فعل سادهٔ فارسی انجام داده‌ایم که افعال را براساس ظرفیت‌های نحوی و معنایی گروه‌بندی می‌کند.

برای این منظور ابتدا افعال را بر اساس ظرفیت نحوی دسته‌بندی کرده و سپس این دسته‌ها را بر اساس نقش‌های معنایی که می‌گیرند به دسته‌های کوچکتر تقسیم کردیم. (نقشهای معنایی که در کنار آنها علامت + آمده اجباری و سایر نقش‌ها در جمله اختیاری هستند.)

فهرست کامل افعال به‌کار رفته در این سیستم و کلاس‌بندی آنها بر اساس نقش‌های معنایی که می‌پذیرند در (جدول ۳) مشخص شده است:

[1] در صورتی که کلمات قبل یا بعد از کلمه مورد نظر در محدودهٔ جمله وجود نداشته باشند، برچسب NULL درنظر گرفته می شود.

[2] Current Constituent

جدول (۳) – دسته‌بندی افعال در کلاسهای معنایی

مجموعه نقشهای معنایی	افعال
Agent,+ (topic or patient) , +] [+goal	اندیشیدن، آموختن، نوشتن
[Agent, location+]	ایستادن، خوابیدن، نشستن
[Agent,+patient+]	بوسیدن،پسندیدن،کشتن، آزمودن
[Agent,+patient, inst+]	پوشیدن،ساختن،شکستن، بریدن
Agent,+patient,goal, benf+] [	بافتن، فرستادن
[Agent,+patient, source+]	خریدن، دزدیدن، ربودن
[Agent,+patient, goal+]	فروختن،باختن،گنجاندن، انداختن، فشردن
Agent,+patient, goal, +] [source	پاشیدن، ریختن، پرداختن
[Agent, source,goal, inst+]	پریدن، گریختن
[Agent,goal, inst+]	دویدن،خندیدن،نگریستن، جنگیدن
[(Agant,+(patient or topic+]	پذیرفتن، دیدن، فهمیدن
Agent,+(patient or topic), +] [source,benf	شنیدن، خواندن، پرسیدن
[(Agent,+(topic or source+]	ترسیدن
[Agent,+topic, goal+]	فرمودن، کوشیدن، گفتن
Agent or patient), +goal,)+] [inst	چسبیدن
[Patient, inst+]	سوختن
[Agent,+patient,+percept+]	شناختن
Agent,+(patient or +] [topic), +percept	دانستن

به‌عنوان مثال فعل "گریختن" متعلق به کلاس ۹ است که به‌صورت زیر توصیف می‌شود:

[Agent, source,goal, inst+]

نقش‌های معنایی برای جمله "سه زندانی با هلیکوپتر از زندان گریختند. " به‌صورت زیر می‌باشد:

Agent	سه زندانی
Instrument	با هلیکوپتر
Time	روز شنبه
Source	از زندان
Predicate	گریختند.

همان‌طور که در مثال بالا دیده می‌شود، ترتیب ظهور آرگومان‌ها در جمله الزاماً با ترتیب مشخص شده در مجموعه نقش‌ها یکی نیست.

ذکر این نکته ضروری است که نقش‌های معنایی هر فعل، صرفاً با توجّه به نمونه جملات موجود از آن فعل در پیکره تعیین شده؛ به این معنی که ممکن است فعل نقش‌های معنایی (اختیاری) دیگری هم بپذیرد که در ساختار جملات پیکره نیامده است. همچنین در تعیین مجموعه نقش‌های معنایی، هر فعل در معنی اصلی آن مورد توجه بوده است به‌عنوان مثال فعل "خواندن" در بعضی جملات با معنی "نامیدن" یا "دانستن" آمده، که از این موارد صرف نظر شده است. مانند:

رهبر انقلاب فساد اداری را کشنده خواند.

اهل محل او را پنجه‌طلائی می‌خواندند.

۴- تحلیل نتایج

در اینجا از نرم‌افزارTiMBL برای پیاده‌سازی الگوریتم یادگیری مبتنی بر حافظه بهره گرفته‌ایم. TiMBL توسط یک گروه تحقیقاتی در دانشگاه Tilburg به‌منظور استفاده در کاربردهای پردازش زبان طبیعی تطبیق داده شده است[۱۸]. MBL برای بسیاری از مسائل NLP نسبت به سایر روش‌های یادگیری ماشین برتری دارد.

الگوریتم یادگیری که در این سیستم به‌کار برده‌ایم الگوریتم IB1 با روش معیار مشابهت همپوشانی وزن‌دار با وزن‌دهی نرخ بهره و تعداد همسایگی یک می‌باشد.

برای ارزیابی سیستم علاوه بر نرخ صحت (درصد داده‌های آزمایش که به‌درستی دسته‌بندی شده‌اند) که در رابطهٔ سه مشخص شده:

$$\text{Accuracy}=\frac{\#\text{of correctly tagged tokens}}{\#\text{of tokens}} \qquad (۳)$$

از تعدادی معیارهای ارزیابی متداول برای سیستم‌های پردازش زبان، مانند دقت، Recall و F1 نیز استفاده شده است که در ادامه به توضیح هر یک می‌پردازیم.

معیار دقت، نشان دهنده درصد آرگومان‌های درست تشخیص داده شده نسبت به تمام آرگومان‌های تشخیص داده شده توسط سیستم است:

$$\text{Precision} = \frac{\#\text{of correctly tagged tokens as phrase type X}}{\#\text{of detected tokens as phrase type X}} \qquad (۴)$$

و Recall را به صورت تعداد آرگومان‌های درست تشخیص داده شده نسبت به تمام آرگومانهای درست تعریف می‌کنیم:

$$\text{Recall} = \frac{\#\text{ of correctly tagged tokens as phrase type X}}{\#\text{of tokens as phrase type X}} \qquad (۵)$$

همان طور که در رابطهٔ شش مشخص است، معیار F1، میانگین هارمونیک Precision وRecall می‌باشد و مقایر این دو معیار را در یک واحد خلاصه می‌کند:

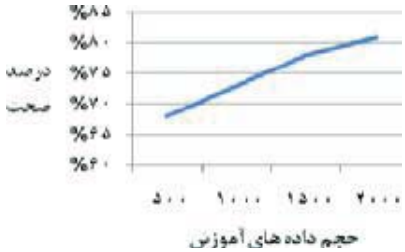
$$F\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (۶)$$

نتایج زیرسیستم تجزیهٔ سطحی، با آموزش برروی دوهزار جمله و آزمایش بر روی یک‌هزار جمله در (جدول ۴) نشان داده شده است.

جدول (۴)– کارآیی سیستم تجزیه گر سطحی

Total Accuracy	81.6
F1 (Micro-avg)	81.6
F1 (Macro-avg)	78.5

به‌منظور بررسی تأثیر افزایش حجم داده‌های آموزش بر روی نتایج تجزیه‌گر، سیستم را پنج مرتبه، با حجم داده‌های مختلف آموزش داده و هر بار بر روی یک مجموعهٔ ثابت آزمایش کرده‌ایم. همان‌طور که در (شکل ۲) مشاهده می‌شود افزایش حجم داده‌ها در بهبود عملکرد سیستم بسیار مؤثراند به‌همین سبب انتظار می‌رود با افزایش حجم داده‌ها نتایج بهتری برای سیستم تجزیهٔ سطحی نحوی به‌دست آید.



شکل (۲) : کارآیی سیستم تجزیه‌گر سطحی

(جدول ۵) کارآیی سیستم برچسب‌زنی معنایی را با آموزش بر روی ۱۳۰۰ جمله و آزمایش بر روی هفتصد جمله، برای هر نقش، به‌صورت مجزا نشان می‌دهد. این نتایج در شرایطی حاصل شده است که از خطای ناشی از مرحلهٔ قبل چشم‌پوشی و ورودی‌های صحیح به سیستم داده شده باشد.

جدول (۵) – کارآیی سیستم تخصیص نقش معنایی

نقش معنایی	دقت	بازخوانی
Agent	89.2	95.4
Predicate	1	1
Topic	93.6	99.2
#	1	1
Goal	73.7	85.3
Manner	47.1	34.8
Time	33.7	21.9
Reason	12.6	0.11
Location	51.4	58.5
Patient	86.2	87.8
Instrument	64.7	82.5
Source	62.1	70.5
Beneficiary	0	0
Percept	84.6	90.2

دقّت تشخیص پایین در برخی نقش‌ها را می‌توان ناشی از فراوانی کم این نقش‌ها در مجموعه داده دانست. (جدول ۶) آماری از فراوانی نقش‌ها در پیکرهٔ آموزش را نشان می‌دهد:

جدول (۶) – فرکانس نقش‌های معنایی در داده های آموزش

Label	Frq.	Label	Frq.
Agent	698	Location	188
Patient	505	Time	129
Goal	272	Instrument	31
Predicate	966	Source	101
Reason	53	Beneficiary	9
Topic	199	Percept	72
Manner	194		

هم‌چنین (جدول۷) نتایج کلی مرحلهٔ برچسب‌زنی را با ورودی‌های استاندارد نشان می‌دهد، که حاصل میانگین‌گیری مقادیر (جدول۵) می‌باشد.

دو روش برای میانگین‌گیری استفاده شده است: میانگین‌گیری میکرو[1]و میانگین‌گیری ماکرو[2]. در روش میکرو مقدار F1 هر کلاس با توجه به بسامد آن کلاس در داده‌های آزمایش وزن‌دهٔ می‌شود و در روش ماکرو مقادیر F1 تمام کلاس‌ها باهم جمع و بر تعداد کلاس‌ها تقسیم می‌شود.

جدول (۷) – کارآیی سیستم تخصیص نقش معنایی

Total Accuracy	87.4
F1 (Micro-avg)	87.4
F1 (Macro-avg)	70.6

همچنین میزان مشارکت هر ویژگی را در کارآیی سیستم، با حذف آن ویژگی و محاسبهٔ اختلاف کارآیی سیستم در حالت وجود و عدم وجود آن ویژگی در (جدول ۸) به‌دست آورده‌ایم. همان‌طور که از مقادیر (جدول ۸) مشخص است بیشترین کاهش کارآیی در حالتی که ویژگی‌های گروه نحوی آرگومان جاری و کلاس فعل حذف شده‌اند روی داده است که نشان‌دهندهٔ اهمیت و تأثیر این ویژگی‌ها در تشخیص برچسب مناسب می‌باشد.

[1] Micro-Averaging
[2] Macro-Averaging

جدول (۸) – ΔF1 ناشی از حذف ویژگی ها

Feature(s) removed	ΔF1
All features	0.8624
Current constituent's Phrase type	-0.0494
Previous constituent 's Phrase type	-0.0234
Next constituent 's Phrase type	-0.0266
Position	-0.0009
Voice	-0.0109
Verb Class	-0.0710

از آن‌جا که کارآیی سیستم تجزیهٔ نحوی بر نتایج سیستم برچسب‌زنی معنایی تأثیر مستقیم دارد، واضح است که معیارهای ارزیابی برای سیستم نهایی کمتر از مقادیر (جدول ۷) باشد. کارآیی سیستم نهایی در (جدول ۹) نشان داده شده است:

جدول (۹)– کارآیی کل سیستم برچسب‌زنی نقش معنایی

Total Accuracy	۷۳.۸
(F1 (Micro-avg	۷۳.۸
(F1 (Macro-avg	۷۰.۹

از مقایسهٔ نتایج دو (جدول ۷ و ۹)، مشخّص می‌شود که استفادهٔ مستقیم از خروجی سیستم تجزیه نحوی، کارآیی سیستم برچسب‌زنی معنایی را نسبت به حالتی که خروجی‌ها به‌صورت دستی تصحیح شده باشد، ۱۴٪ کاهش می‌دهد.

۵-نتیجه‌گیری و بحث

در این مقاله، به طراحی و پیاده‌سازی یک سیستم برچسب‌زنی معنایی مبتنی بر پیکرهٔ برای جملات فارسی پرداخته‌ایم. از آن‌جایی که هیچ پیکرهٔ برچسب‌گذاری شدهٔ معنایی برای زبان فارسی وجود ندارد؛ بخشی از پیاده‌سازی سیستم به تهیه و حاشیه‌نویسی مجموعه کوچکی متشکل از ۱۳۰۰ جمله برای آموزش سیستم اختصاص داده شد. نتیجهٔ به‌دست آمده در مرحلهٔ تجزیهٔ نحوی سطحی، ۸۱٪ و برای مرحلهٔ برچسب‌زنی معنایی، ۸۷٪ می‌باشد که با در نظر گرفتن حجم داده‌ها و سایر محدودیت‌ها در زبان فارسی قابل قبول است. این نتایج نشان می‌دهد که اعمال ویژگی‌های انتخاب شده از سیستم‌های انگلیسی برای داده‌های فارسی عملکرد خوبی دارد.

از آن‌جا که تاکنون سیستم مشابهٔ در زمینه برچسب‌زنی معنایی جملات فارسی با روش‌های یادگیری ماشین ارائه نشده است، امکان مقایسهٔ نتایج با کارهای قبلی وجود ندارد. همچنین با توجّه به این‌که حجم و فرمت داده‌های استفاده شده در این سیستم با سیستم‌های تهیه شده برای زبان انگلیسی متفاوت است، نمی‌توان به مقایسهٔ صحیحی در این زمینه دست یافت. تنها سیستمی که برای استخراج نقش معنایی در زبان فارسی کار شده است، مربوط به [۱۱] با نرخ صحت ۸۱.۶٪ می‌باشد که از استراتژی مبتنی بر قاعده استفاده کرده‌است. مقایسهٔ نتایج دو سیستم برتری سیستم پیشنهاد شده در این مقاله را نشان می‌دهد. همچنین نتیجهٔ گزارش‌شده برای سیستم مشابهٔ که در زبان آلمانی تهیه شده است [۲]، نرخ صحّت ۸۰٪ را نشان می‌دهد. بیشترین نرخ صحت گزارش‌شده در مقالات انگلیسی مشابه [۵] در حدود ۹۰٪ می‌باشد که با توجه به حجم داده و منابع دردسترس برای زبان انگلیسی این امر قابل توجیه می‌باشد.

از مزایای این سیستم می‌توان به‌کاربرد آن به‌عنوان جزئی از سیستم‌های بزرگ‌تر پردازش‌های زبان طبیعی اشاره کرد. با گسترش حجم پیکره و فهرست افعال می‌توان عملکرد سیستم را بهبود بخشید و آن را برای تجزیهٔ انواع جملات تعمیم داد.

۶-مراجع

[1] Marquez Lluís, Carreras Xavier, Litkowski Kenneth, Stevenson Suzanne, "Semantic Role Labeling: An Introduction to the Special Issue" Association For Computational Linguistic 2008.

[2] Gerwert Stevens , "Automatic semantic role labeling in a Dutch corpus" , master thesis, Universiteit Utrecht, Faculty of arts , September 2006

[3] Sun Honglin, Jurafsky Daniel. "Shallow Semantic Parsing of Chinese" . In Proceedings of NAACL 2004, Boston, USA

[4] Gilda Daniel, Jurafsky Daniel, "Automatic Labeling Of Semantic Role", Association of computer linguistic, 28(3). pp.245–288. 2002

[5] Pradhan Sameer, Jurafsky Daniel, "Support Vector Learning for Semantic Argument Classification", Springer Science, 2005

[6] Lim Joon-Ho, Hwang Young-sook, Park So-young, and Rim Hae-chang. "Semantic role labeling using maximum entropy model", In Proceedings of CoNLL-2004. 2004

[7] http://framenet.icsi.berkeley.edu/

[8] http://verbs.colorado.edu/~mpalmer/projects/ ace/PBguidelines.pdf

[9] Morante Roser, Busser Bertjan, "Role Labelling for Catalan and Spanish using TiMBL", Proceedings of the 4th International Workshop on

Semantic Evaluations (SemEval-2007), pages 183–186.
[10] Hammerton James , M. Osborne, S. Armstrong, W. Daelemans. 2002. "Introduction to Special Issue on Machine Learning Approaches to Shallow Parsing", Journal of Machine Learning Research , 551-558.
[11] Sadr-Mousavi Maryam, Shamsfard Mehrnoush, "Thematic Role Extraction Using Shallow Parsing", International journal of computational Intelligence Volume 4 , 2007
[12] Dowty David, "Thematic Proto-roles and Argument Selection", Language 67, pp.547–619,1991.
[13] Levin Beth, "English Verb Classes and Alternations", the University of Chicago Press, Chicago and London, 1993
[14] Fillmore Charles, "The case for case". Academic Press, New York, 1997
[15] http://www.rcisp.com
[16] Daelemans Walter, "TiMBL: Tilburg Memory-Based Learner", Tilburg University and CNTS Research Group, University of Antwerp 2006
[17] Ratnaparkhi Adwait, "A linear observed time statistical parser based on maximum entropy models", InEMNLP-97, The Second Conference on Empirical Methods in Natural Language Processing, 1997.
[18] ILK: Induction of Linguistic Knowledge, http://ilk.uvt.nl/

آزاده کامل قالی‌باف مدرک کارشناسی خود را در رشتهٔ مهندسی کامپیوتر-نرم‌افزار در سال ۱۳۸۴ از دانشگاه آزاد اسلامی واحد مشهد و مدرک کارشناسی ارشد در رشتهٔ مهندسی کامپیوتر-هوش مصنوعی را در سال ۱۳۸۸ از همان دانشگاه اخذ نموده است. زمینه‌های تحقیقاتی مورد علاقهٔ وی پردازش زبان طبیعی، معناشناسی و تجزیه تحلیل معنایی متن می‌باشد.

نشانی (رایانامک) پست الکترونیکی ایشان عبارت است از: azadeh_kamel@hotmail.com

سعید راحتی قوچانی متولد ۱۳۴۶ شهرستان قوچان، دانش‌آموختهٔ کارشناسی الکترونیک سال ۱۳۶۹ دانشکدهٔ فنی دانشگاه تهران و کارشناسی ارشد مخابرات ۱۳۷۲ دانشگاه آزاد اسلامی تهران جنوب و دکترای مخابرات ۱۳۷۷ دانشگاه آزاد اسلامی واحد علوم و تحقیقات می‌باشد. وی پژوهشگر برتر جشنوارهٔ فردوسی ۱۳۷۹، رتبهٔ اوّل پژوهش سال ۱۳۸۶ و رتبهٔ دوم پژوهش سال ۱۳۸۵ دانشگاه آزاد اسلامی مشهد شده است. او از سال ۱۳۷۸ به عنوان استادیار مخابرات دانشگاه آزاد اسلامی مشهد مشغول به کار می‌باشد. ایشان از ۱۳۷۸ به مدّت سه سال به‌عنوان معاون آموزشی پژوهشی و بعد از آن به‌مدت سه سال به‌عنوان رئیس دانشکدهٔ مهندسی دانشگاه آزاد اسلامی مشهد و بین سال‌های ۱۳۸۴ تا ۱۳۸۸ معاون آموزشی و پژوهشی دانشگاه امام رضا (ع) بود. ایشان تاکنون بیش از شصت مقاله در کنفرانس‌های داخلی و خارجی و نشریات به چاپ رسانده است. گرایش‌های تحقیقاتی ایشان پردازش گفتار و آموزش شبکه‌های عصبی و کاربرد آن در مدل‌سازی سیستم‌های بیولوژیک می‌باشد.

نشانی (رایانامک) پست الکترونیکی ایشان عبارت است از: rahati@mshdia.ac.ir

اعظم استاجی عضو هیأت علمی گروه زبان‌شناسی دانشگاه فردوسی مشهد است. وی تحصیلات کارشناسی خود را در رشتهٔ زبان و ادبیات انگلیسی، دانشگاه شهید بهشتی تهران به اتمام رساند. تحصیلات کارشناسی ارشد وی در رشتهٔ زبان‌شناسی همگانی، دانشگاه فردوسی مشهد در سال ۱۳۷۶ به‌پایان رسید. در سال ۱۳۷۷ تحصیلات دوره دکتری خود را در رشتهٔ زبان‌شناسی همگانی، دانشگاه فردوسی مشهد آغاز کرد و در سال۱۳۸۳ به‌عنوان نخستین فارغ‌التحصیل دورهٔ دکتری رشتهٔ زبان‌شناسی همگانی دانشگاه فردوسی مشهد، تحصیلات خود را به اتمام رساند. وی از سال ۱۳۸۳ تاکنون در دانشگاه فردوسی مشهد مشغول کار می‌باشد. زمینه‌های مورد علاقهٔ وی زبان‌شناسی تاریخی، واج‌شناسی، تحلیل گفتمان و زبان‌شناسی رایانه‌ای می‌باشد.

نشانی (رایانامک) پست الکترونیکی ایشان عبارت است از: estaji@um.ac.ir