Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 12, Number 2, Summer 2023, pp. 74-88
Received: 02/04/2023, Revised: 08/06/2023, Accepted: 08/26/2023

Integrated Semantic Representation (ISR-Model): Syntax-Independent Model for Natural Language

Nasim Tohidi[1] and Chitra Dadkhah[2*]
1- Artificial Engineering Department, Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran.
2*- Artificial Engineering Department, Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran.
[1]n.tohidi@email.kntu.ac.ir and [2*]dadkhah@kntu.ac.ir

Corresponding author's address: Chitra Dadkhah, Seyed Khandan, Shariati Ave, Faculty of Computer Engineering, Room No 305, K. N. Toosi University of Technology, Tehran, Iran.

Abstract- Today, one of the most widely used natural language representation methods is meaning representation for text processing in different systems. Meaning representation methods have had many applications in the field of natural language processing in recent years, including automatic text summarization systems and question and answer systems, etc. Many text expressions may be different in terms of grammatical structure, but they are the same in terms of meaning, so how to apply a single and uniform meaning to them is one of the challenges of text processing. The main goal of this paper is to design an Integrated Semantic Representation (ISR) model for natural language. The proposed model, while maintaining its simplicity in annotation and understanding, does not depend on specific natural language features or on syntactic and lexical structure. In this regard, several examples in both English and Persian language, which have different in terms of written and grammatical structure, are presented in this paper. Moreover, by using the proposed model, the same representation is provided for texts with the same meaning and different grammatical structure. The proposed model is designed in graph and list format is recommended to annotate its corpuses. One of the main features of this model is that it can represent semantic relations at both sentence-level and document-level and is able to represent complex and important linguistic phenomena such as aspect, tense, and quantification. The simplicity of the proposed model helps to avoid making language processing slow or complicated in various applications, and the preparation of structures based on this model for different natural languages will not be too complicated, so that it can be used both for natural languages with low resources and for those with various resources. Further, features of the proposed model are compared with one of the most important related works.

Keywords- Meaning Representation, Natural Language Processing, Text, Graph, Corpus, Semantics.

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۲، شماره ۲، تابستان ۱۴۰۲ ، صفحه ۷۴-۸۸
دریافت: ۱۴۰۱/۱۱/۱۵، بازنگری: ۱۴۰۲/۰۵/۱۵، پذیرش: ۱۴۰۲/۰۶/۰۴

ISR-Model: مدل بازنمایی معنای یکپارچه زبان طبیعی مستقل از ساختار نحوی

نسیم توحیدی[۱]، چیترا دادخواه*[۲]
۱- گروه هوش‌مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.
*۲- گروه هوش‌مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران.
[1] n.tohidi@email.kntu.ac.ir, [2*] dadkhah@kntu.ac.ir

* نشانی نویسنده مسئول: چیترا دادخواه، تهران، خیابان شریعتی، ضلع جنوب شرقی پل سید خندان، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکده مهندسی کامپیوتر، اتاق ۳۰۵.

چکیده-امروزه یکی از پرکاربردترین روش‌های بازنمایی زبان طبیعی، بازنمایی معنا جهت پردازش متون در سامانه‌های مختلف است. روش‌های بازنمایی معنا در سال‌های اخیر کاربردهای متعددی در حوزه پردازش زبان طبیعی از جمله سامانه‌های خلاصه‌سازی خودکار متن و سامانه‌های پرسش و پاسخ و غیره داشته‌اند. بسیاری از عبارات متنی ممکن است از لحاظ ساختار گرامری متفاوت، اما از نظر معنایی یکسانی باشند، لذا چگونگی اطلاق یک معنای یکسان و واحد به آن‌ها یکی از چالش‌های پردازش متن است. هدف اصلی این مقاله طراحی مدلی برای بازنمایی معنایی یکپارچه و مستقل از ساختار نحوی برای زبان طبیعی است. مدل پیشنهاد شده با حفظ سادگی خود در حاشیه‌نویسی و درک، عدم وابستگی به ویژگی‌های زبان طبیعی خاص و همچنین عدم وابستگی به ساختار نحوی و لغوی را دارد. در این راستا، مثال‌های از جملات زبان فارسی و انگلیسی که از نظر ساختار گرامری وتفاوت هایی با یکدیگر دارند، در مقاله با روش بازنمایی پیشنهادی ارائه شده‌اند. به علاوه، با استفاده از مدل پیشنهادی، بازنمایی یکسانی برای متون با معنای یکسان و در عین حال ساختار گرامری متفاوت ارائه می‌گردد. مدل پیشنهادی در قالب گراف و لیست جهت تفسیر در پیکره طراحی شده است. یکی از ویژگی‌های اصلی مدل پیشنهادی، قابلیت بازنمایی روابط معنایی در سطح جمله و در سطح سند جهت بازنمایی پدیده‌های پیچیده و مهم زبانی مانند جنبه، زمان و سورها است. سادگی مدل پیشنهادی سبب می‌گردد که پردازش‌های زبانی در کاربردهای مختلف کند ویا پیچیده نشود، و تهیه پیکره‌های مبتنی بر این مدل در آینده برای زبان‌های طبیعی مختلف پیچیدگی زیادی نداشته باشد، تا هم برای زبان‌های طبیعی با منابع محدود و هم برای زبان‌های طبیعی با امکانات کامل قابل بهره‌برداری باشد. همچنین، ویژگی‌های مدل پیشنهادی با یکی از مهم‌ترین کارهای مرتبط پیشین مقایسه شده است.

واژه‌های کلیدی: بازنمایی معنا، پردازش زبان طبیعی، متن، گراف، پیکره، معنا

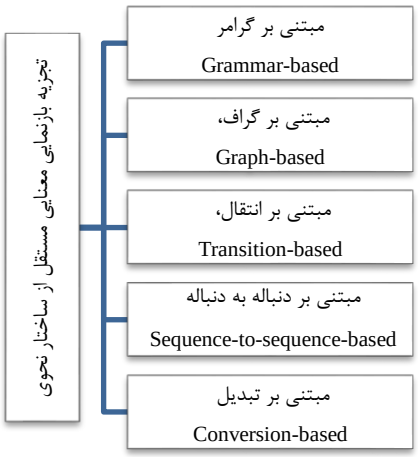
۱- مقدمه

در جهان امروز، پردازش زبان طبیعی اهمیت بسیار زیاد و کاربردهای فراوانی در سامانه‌های هوشمند دارد. بسیاری از کارهای روزمره ما به وسیله این سامانه‌ها تسهیل و تسریع شده‌اند تا جایی که حتی مسئله نگارش مقالات علمی به وسیله این سامانه‌ها بسیار داغ است[۱]. در حالت کلی می‌توان فرآیندهای پردازش زبان طبیعی را در چهار سطح لغوی[۲]، نحوی[۳]، معنایی[۴] و مفهومی[۵] تقسیم‌بندی نمود [۱].

یکی از سطوح پیچیده و بسیار مهم زبان طبیعی سطح معنا است، که پیشرفت‌های صورت گرفته در آن در مقایسه با دو سطح اول بسیار کم‌تر است. برای درک جمله‌های یک زبان طبیعی توسط ماشین، رویکردهای مختلفی را می‌توان در پیش گرفت. در این راستا، پس از برداشتن گام‌های گوناگون در زمینه پردازش سطح لغوی و سطح نحوی زبان و پیش از به کارگیری شیوه‌های استنتاج مفهومی، معمولا لازم است تا چارچوب مناسبی برای بازنمایی معنای جمله‌های زبان در نظر گرفته شود، تا طی فرآیند پردازش معنا،

جمله‌ها در قالب این چارچوب بازنمایی شوند. همچنین، در برخی موارد به جای انجام تحلیل‌های سطح نحوی و لغوی می‌توان از بازنمایی معنا استفاده نمود تا پیچیدگی‌های روش‌های مبتنی بر آن کم‌تر و ابزارهای با محدودیت کم‌تر تولید شوند. در نتیجه، این بازنمایی قابل استفاده و فهم برای ماشین خواهد بود و می‌تواند تا حد بسیار زیادی فرآیندهای پردازش معنای زبان طبیعی در کاربردهای مختلف مانند ترجمه ماشینی، خلاصه سازی متن، بازیابی اطلاعات و غیره را تسهیل نماید.

در این راستا، به طور خلاصه، کارهای صورت گرفته برای تجزیه بازنمایی معنایی مستقل از ساختار نحوی را می‌توان در ۵ گروه اصلی تقسیم‌بندی نمود (شکل ۱).



شکل ۱: طبقه‌بندی روش‌های تجزیه بازنمایی معنایی مستقل از ساختار نحوی

در روش مبتنی بر گرامر، یک گرامر برای محدود کردن گراف‌ها در فرآیند تجزیه مورد استفاده قرار می‌گیرد. سپس، گراف با بیشترین امتیاز توسط تابع امتیازدهی به عنوان خروجی انتخاب می‌شود. سیستم‌های بسیاری وجود دارند که از این روش برای تجزیه بازنمایی معنا استفاده نموده‌اند. مثلا، در دو پژوهش [۲, ۳] برای اولین‌بار از گرامر جایگزینی ابریال همزمان (SHRG[6]) برای تجزیه بازنمایی معنا استفاده نمودند. این تجزیه‌گرها رشته‌های زبان طبیعی را با استفاده از گرامر مستقل از متن (CFG[7]) و گراف‌های بازنمایی را با استفاده از گرامر جایگزینی ابر یال (HRG[8]) مدل کردند.

در روش مبتنی بر گراف تلاش بر این است تا گرافی که تابع امتیازدهی را بیشینه می‌نماید، ساخته شود. در اولین و مهم‌ترین پژوهشی که از روش مبتنی بر گراف استفاده نمود، تجزیه‌گر JAMR ارائه شد [۴]. در این پژوهش، الگوریتمی معرفی شد که ابتدا مفاهیم را با استفاده از یک مدل شبه‌مارکوف تشخیص می‌داد و سپس، ارتباطات بین آن‌ها را با جستجو برای زیرگراف متصل پوشا بیشینه (MSCG[9]) استخراج می‌نمود.

در روش مبتنی بر انتقال، یک گراف در قالب دنباله‌ای از عمل[۱۰]ها تولید می‌شود و زمانی که یک عمل انتخاب شد، جایگزین‌های آن عمل دیگر بازبینی نخواهند شد. در پژوهش صورت گرفته در [۵]، رویه تجزیه بازنمایی معنای مستقل از ساختار نحوی در قالب یک انتقال از درخت وابستگی به گراف بازنمایی معنایی فرمول‌بندی شده است. همچنین، این روش از الگوریتم Perception برای یادگیری مدل خطی استفاده می‌نماید. هم‌چنین، در پژوهشی دیگر، از این روش برای طراحی یک تجزیه‌گر افزایشی برای بازنمایی معنایی انتزاعی استفاده شد [۶]. ارزیابی‌های انجام شده در این پژوهش‌ها با استفاده از معیار Smatch انجام شده است. در یک پژوهش دیگر که اخیرا صورت گرفته است [۷]، یک مدل تجزیه بازنمایی معنایی مستقل از ساختار نحوی، طراحی شده که از یک الگوریتم مبتنی بر انتقال شامل دو پشته[۱۱] بهره می‌برد. این مدل روی دو مجموعه‌داده انگلیسی و چینی اعمال شده است و می‌تواند در یک زمان خطی جملات ورودی را در قالب گراف‌های بازنمایی معنایی تجزیه نماید.

در روش مبتنی بر دنباله به دنباله که به روش ترجمه ماشینی نیز معروف است، به طور مستقیم از ورودی سیستم و با استفاده از روش‌های شبکه عصبی دنباله به دنباله، قالب متنی مرتبط با یک گراف بازنمایی معنا تولید می‌شود. معمولا در این روش برای تسهیل نمودن فرآیند، نیاز به مجموعه‌ای از پیش‌پردازش‌ها خواهد بود. در یک پژوهش برای اولین‌بار از این روش برای ترجمه ماشینی مبتنی بر شبکه عصبی استفاده شد [۸] که در آن، پیمایش پیش‌ترتیب گراف بازنمایی معنا به عنوان رشته‌های زبان در نظر گرفته شد و به این صورت عمل تجزیه بازنمایی معنا با استفاده از نشانه‌گذاری PENMAN [۹]، انجام گرفت. در مقابل، در [۱۰]، برای حل مسئله رایج خلوتی[۱۲] (کمبود داده)، از یک رویکرد خودآموز (خودیادگیر) که از یک پیکره نشانه‌گذاری‌نشده بزرگ بهره می‌برد، استفاده شد.

در روش مبتنی بر تبدیل، درخت‌های وابستگی به گراف‌های بازنمایی معنا تبدیل می‌شوند که از آن برای طراحی تجزیه‌گرهای محدودی تاکنون استفاده شده است. به طور مثال، در یک پژوهش تلاش شد تا از ترجمه ماشینی مبتنی بر نحو برای تجزیه جملات زبان انگلیسی به بازنمایی معنایی انتزاعی استفاده گردد [۱۱]. به این منظور، یک مدل زبانی مختص بازنمایی معنایی انتزاعی معرفی شد و داده‌ها و ویژگی‌های برگرفته از منابع معنایی به آن اضافه گردید.

تا به امروز، بازنمایی معنای مستقل از ساختار نحوی در کاربردهای مختلف پردازش زبان طبیعی، بسیار مورد استفاده قرار گرفته است که آن‌ها را به ۶ دسته اصلی خلاصه‌سازی، پرسش و پاسخ، بازیابی اطلاعات، ترجمه ماشینی، درک ماشینی و ارتباط موجودیت‌ها و داده‌های مرتبط می‌توان تقسیم بندی نمود [۱۲].

هدف اصلی این مقاله، ارائه روش بازنمایی معنایی یکپارچه زبان طبیعی به صورت عمومی است. در واقع، روش بازنمایی معنای پیشنهادی قابل استفاده در کاربردهای مختلف پردازش زبان و بدون وابستگی به زبان طبیعی خاصی و مستقل از سطح نحوی و لغوی ارائه شده است تا برای برآوردن نیازهای طیف گسترده‌ای از کاربردهای پردازش زبان طبیعی مورد استفاده قرار گیرد.

نوآوری‌های اصلی در ارائه مدل پیشنهادی این مقاله عبارتند از:

- قابلیت بازنمایی متن در زبان‌های طبیعی مختلف
- قابلیت بازنمایی ارتباطات معنایی در سطح سند، علاوه بر بازنمایی ارتباطات معنایی در سطح جمله
- قابلیت بازنمایی معنایی طیف وسیعی از پدیده‌های زبانی

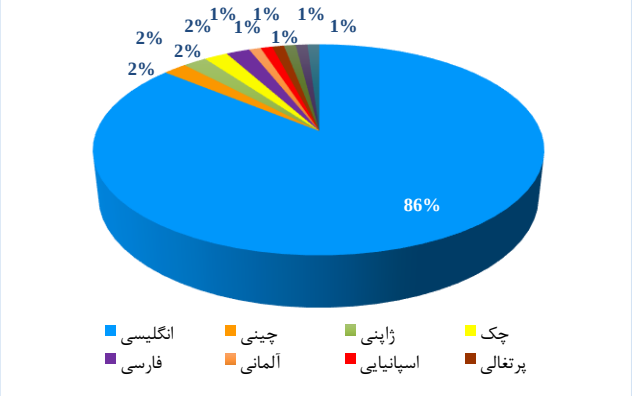
در ادامه، در بخش ۲، کارهای پیشین مرتبط مرور خواهند شد. در بخش ۳، جزئیات مدل پیشنهادی مطرح و مثال‌های متعددی بررسی خواهند شد و ساختار رسمی مدل پیشنهادی معرفی می‌شود. در بخش ۴، به ارزیابی مدل پیشنهادی می‌پردازیم و در بخش ۵ نیز به نتیجه‌گیری و اشاره به کارهای آتی خواهیم پرداخت.

۲- کارهای پیشین

توجه به این نکته حائز اهمیت است که فرآیند پردازش متن در زبان انگلیسی، به عنوان یک زبان بین‌المللی شناخته شده، با وجود منابع غنی شامل پیکره‌ها و ابزارهای مختلف و فراوان، نسبت به بسیاری از زبان‌ها راحت‌تر است. لذا در بیشتر حوزه‌ها و کاربردهای پردازش زبان طبیعی، اکثر تلاش‌های صورت گرفته در رابطه با بازنمایی معنا نیز مربوط به زبان انگلیسی بوده است [۱۲].

اخیرا، علی‌رغم ویژگی‌ها و چالش‌های خاص زبان فارسی، تلاش‌های بسیاری در راستای پیاده‌سازی سیستم‌های هوشمند با استفاده از تکنیک‌های پردازش زبان فارسی صورت گرفته است [۱۳, ۱۴]؛ و ما امروز شاهد رواج استفاده از سامانه‌های هوشمند برای خلاصه‌سازی، ویرایش متن، پرسش و پاسخ و ترجمه ماشینی برای زبان فارسی، که از مهم‌ترین کاربردهای پردازش زبان طبیعی محسوب می‌شوند، هستیم [۱۵, ۱۶]. اما همچنان در مقایسه با زبان‌های دیگر به خصوص زبان انگلیسی، کمبودهای بسیاری برای به‌کارگیری سیستم‌های هوشمند در زبان فارسی وجود دارد [۱۷, ۱۸]. بنابراین، نیاز به پژوهش‌های بیشتر برای زبان فارسی و سایر زبان‌های با شرایط مشابه کاملا وجود دارد و ارائه یک بازنمایی معنایی مستقل از ساختار نحوی برای زبان‌های طبیعی مختلف می‌تواند آغازگر یک مسیر مهم از پیشرفت‌های آتی در زمینه پردازش این زبان‌ها باشد.

در همین راستا و برای مطالعه کارهای مشابه پیشین انجام شده، ابتدا این کارها را از نظر زبان‌های طبیعی که هر مدل مبتنی بر آن طراحی شده است، بررسی نمودیم. شکل ۲ خروجی این بررسی را نشان می‌دهد که در آن توزیع زبان‌های طبیعی مختلفی که در گذشته پژوهش‌های مشابهی در زمینه بازنمایی معنای مستقل از ساختار نحوی از سال ۲۰۱۳ تاکنون برای آن‌ها صورت گرفته است، در قالب یک نمودار دایره‌ای قابل مشاهده است.



شکل ۲: توزیع زبان‌های مختلف در پژوهش‌های بازنمایی معنایی مستقل از ساختار نحوی

با توجه به شکل ۲، اکثر کارهای پیشین انجام شده (۸۶ درصد) در این حوزه متمرکز بر زبان انگلیسی بوده‌اند. با توجه به این نکته، نیاز به ارائه روشی که بتوان از آن در زبان‌های طبیعی مختلف استفاده نمود، هم از نظر بازنمایی اسناد چندزبانی و هم از نظر بازنمایی متون به زبان‌های طبیعی غیرانگلیسی کاملا محسوس است.

در ادامه به معرفی برخی از تجزیه‌گرها و تولیدکننده‌های مبتنی بر بازنمایی معنایی انتزاعی می‌پردازیم تا با کاربرد فراوان مدل‌های بازنمایی معنا در حوزه پردازش زبان طبیعی آشنا شوید.

۲-۱- تجزیه‌گر معنای زبان طبیعی

روش USP[13] به عنوان اولین تجزیه‌گر معنای بدون ناظر در سال ۲۰۰۹ توسط Poon و Pedro ارائه شد و از رویکرد مارکوف بهره می‌برد [۱۹]. در آن، درخت‌های تجزیه نحوی به شکل شبه‌منطقی تبدیل می‌شدند و فرم‌های لامبدا[14] به صورت بازگشتی از آن‌ها

استخراج می‌گردید. سپس، یک دسته‌بندی انجام می‌شد و فرم‌هایی که به صورت نحوی مشابه بودند در یک گروه قرار می‌گرفتند.

در سال ۲۰۱۰، سیستم SEMAFOR توسط Chen و همکارانش ارائه شد، که یک تجزیه‌گر متن‌باز برای بازنمایی معنای به روش قاب بود [۲۰]. این سیستم در رقابت‌های استخراج معنای SemEval سال ۲۰۰۷ توانست بهترین نتیجه را به دست آورد.

در سال ۲۰۱۳، Banarescu و همکارانش یک بازنمایی معنای مستقل از ساختار نحوی و لغوی برای زبان انگلیسی ارائه نمودند [۲۱]. این بازنمایی معنا، معنای یک جمله را به عنوان گره‌های برچسب زده شده در گراف و یال‌های جهت‌دار بین آن‌ها نشان می‌دهد. همچنین، از مجموعه مفاهیم زبان و پیکره PropBank استفاده می‌کند [۲۲]. در نهایت، این بازنمایی، ایده «چه کسی چه کاری را برای چه کسی انجام می‌دهد» را در یک منطق سبک گزاره‌ای به دست می‌آورد، که این فرآیند مجزا از تغییرات در سطح نحوی و لغوی صورت می‌گیرد [۲۳]. چند سال بعد، یک پژوهش تکمیلی نیز برای تهیه یک مجموعه‌داده شامل ۳۹۰۰۰ جمله علامت‌گذاری شده برای بازنمایی معنای مستقل از ساختار نحوی توسط Knight و همکارانش صورت گرفت، که یکی از مهم‌ترین اقدامات اخیر در این راستا است [۲۴]. در سال‌های بعد پژوهش‌های دیگری تلاش کردند تا کمبودهای روش اولیه را از جهات مختلف مورد بررسی قرار داده و آن‌ها را برطرف نمایند [۲۵].

در سال ۲۰۱۸، Damonte و Cohen در پژوهشی نشان دادند که می‌توان تجزیه‌گر بازنمایی معنایی زبان انگلیسی را برای زبان‌های ایتالیایی، اسپانیایی، آلمانی و چینی نیز گسترش داد [۲۶]. در واقع در این پژوهش، مدل پیشنهادی نویسندگان ابتدا جمله‌های ورودی را با روش‌های ترجمه ماشینی به زبان انگلیسی ترجمه می‌نمود و سپس بازنمایی معنایی معادل آن جمله در خروجی تولید می‌شد.

۲-۲- تولید کننده معنای زبان طبیعی

در پژوهش ارائه شده توسط Flanigan و همکارانش، اولین تولیدکننده بازنمایی مستقل از ساختار نحوی معرفی شد [۲۷]. این تولیدکننده برای بازنمای معنای عمیق از روش‌های یادگیری باناظر استفاده کرده است. در مجموع دو سیستم اصلی وجود دارند که از روش مبتنی بر مبدل درختی[۱۵]برای تولید بازنمایی معنای مستقل از ساختار نحوی استفاده کرده‌اند. سیستم اول در سال ۲۰۱۶ توسط Flanigan در [۲۷] و همکارانش ارائه شد و سیستم دوم در سال ۲۰۱۷ توسط Gruzitis و همکارانش در [۲۸] ارائه گردید. در این پژوهش‌ها، گراف‌های بازنمایی معنا ابتدا به درخت‌های نحوی مرتبط تبدیل شده و سپس، یک چارچوب گرامری برای تولید جمله‌های زبان انگلیسی از این درخت‌ها به کار گرفته می‌شود.

به عنوان مثالی از کاربرد رویکرد گرافی مبتنی بر گرامر[۱۶]، در سال ۲۰۱۷، Song و همکارانش در [۲۹] از یک گرامر جایگذاری گره همروند[۱۷] (SNRG) برای تولید بازنمایی معنایی استفاده نمودند. SNRGها مشابه گرامرهای مستقل از متن هستند، اما یک طرف گرامر روی گراف‌های گرامری تعریف شده است. در این پژوهش، SNRG را با یک مدل آماری مبتنی بر ویژگی ترکیب نمودند، که شامل یک مدل زبانی برای تولید زبان طبیعی بود.

Song و همکارانش در [۳۰] یک تولیدکننده بازنمایی معنای مستقل از ساختار نحوی را در قالب یک مسئله فروشنده دوره‌گرد غیرمتقارن، طراحی نمودند، که در آن گره‌ها محل به‌کارگیری قواعدی هستند که قطعات گراف بازنمایی معنای را به رشته‌ها نگاشت می‌کنند. دو تولیدکننده اصلی بازنمایی معنای مستقل از ساختار نحوی مبتنی بر قاعده وجود دارند. در سیستم معرفی شده در سال ۲۰۱۷ توسط Gruzitis و همکارانش در [۲۸]، گراف‌های بازنمایی معنا به درخت‌های نحوی تبدیل می‌شوند و از یک تولیدکننده چارچوب گرامری برای تولید جمله‌های زبان انگلیسی استفاده می‌شود. در این روش اگر این تولیدکننده نتواند خروجی تولید کند، از خروجی JAMR استفاده می‌شود.

در سیستم دیگر طراحی شده در سال ۲۰۱۷ توسط Mille و همکارانش در [۳۱]، گراف‌های بازنمایی معنای مستقل از ساختار نحوی با استفاده از یک دنباله از مبدل‌های گراف مبتنی بر قاعده به ساختارهای نحوی مسطح انتقال می‌یابند. در رویکرد مبتنی بر انتقال[۱۸]، تولید بازنمایی معنای مستقل از ساختار نحوی در قالب یک دنباله از عمل‌ها که توسط یک طبقه‌بند یادگیری‌شده انتخاب می‌شوند، انجام می‌شود.

در سال ۲۰۱۷، در دو مقاله که توسط Lampours و Vlachos در [۳۲] و همچنین Schick و همکارانش در [۳۳] صورت گرفته است، گراف‌های بازنمایی معنای مستقل از ساختار نحوی با استفاده از دنباله‌ای از عمل‌ها به (شبه)درخت‌های وابستگی نحوی انتقال می‌یابند و سپس در گام بعدی، در فرآیند تولید به صورت خطی مورد استفاده قرار می‌گیرند.

در سال ۲۰۱۸، پژوهشی توسط Gao صورت گرفت که در آن درک معنا با استفاده از بازنمایی معنایی انتزاعی و مدل فضای برداری[۱۹] بررسی شد [۳۴]. نتیجه این بررسی نشان داد که این دو روش می‌توانند کیفیت درک زبان انگلسی توسط ماشین را بسیار افزایش

دهد. به همین دلیل در کاربردهای متفاوت و بسیاری، از جمله بازیابی اطلاعات و ترجمه ماشینی از این دو روش استفاده شده است.

در سال ۲۰۲۰، Ribeiro و همکارانش یک مدل برای تولید متن از گراف پیشنهاد دادند که اطلاعات مختلف و مکمل موجود در گراف بازنمایی معنایی انتزاعی را رمزگذاری می‌کرد [۳۵]. این مدل با استفاده از روش‌های رمزگذاری مختلف، بازنمایی‌های پایین به بالا و بالا به پایین گره‌ها را به صورت موازی یاد می‌گرفت.

در سال ۲۰۲۰، پژوهشی دیگر توسط Lai و همکارانش صورت گرفت که در آن بازنمایی معنایی انتزاعی به منطق گزاره‌ای مرتبه اول[۲۰] ترجمه می‌شد [۳۶]. این روش می‌توانست زمینه معنایی یک عبارت را در قالب یک استدلال به دست آورد. هم‌چنین، با استفاده از اصول اولیه بازنمایی معنایی انتزاعی، این امکان را فراهم می‌کرد تا پدیده‌هایی مانند کمیت، نفی و متغیرهای دامنه‌دار مدل شوند.

یکی از جدیدترین پژوهش‌های صورت گرفته مربوط Ilmy و Khorda در سال ۲۰۲۱ است [۳۷]. آن‌ها تلاش کردند تا جملات زبان اندونزیایی را با استفاده از روش‌های یادگیری ماشین به بازنمایی معنایی انتزاعی تجزیه کنند. سیستم طراحی شده در این پژوهش شامل سه مرحله است: پیش‌بینی دوتایی[۲۱]، پیش‌بینی برچسب[۲۲] و ساخت گراف[۲۳]. مرحله اول از مولفه تجزیه وابستگی برای به دست آوردن لبه بین کلمات برای بازنمایی معنایی انتزاعی، استفاده می‌کرد. در مرحله دوم با توجه به خروجی مرحله قبل از الگوریتم یادگیری نظارت شده برای پیش‌بینی برچسب بین لبه‌های بازنمایی معنا استفاده می‌شد. در نهایت مرحله سوم گراف بازنمایی معنایی انتزاعی را تشکیل می‌دهد. در این پژوهش مانند بیشتر پژوهش‌های صورت گرفته در زمینه بازنمایی معنایی انتزاعی از معیار Smatch برای ارزیابی سیستم استفاده شده است، که توانسته است نتیجه قابل قبولی را با توجه به این معیار کسب نماید.

۲-۳- کاربرد بازنمایی معنای زبان فارسی

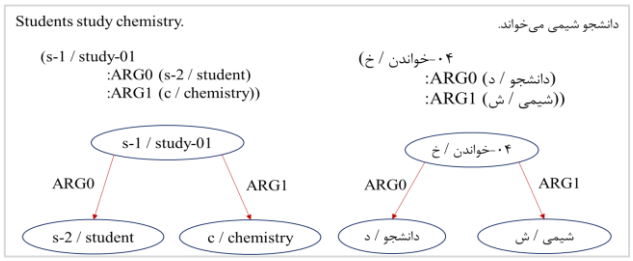
در سال ۲۰۱۷، حدادان و بحرانی در قالب یک پایان‌نامه کارشناسی ارشد در دانشگاه شریف، برای اولین بار از بازنمایی معنایی انتزاعی برای خلاصه‌سازی چکیده‌ای زبان فارسی استفاده نمودند [۳۸]. آن‌ها برای آموزش سیستم خود از ۲۰۰ جمله از کتاب «شازده کوچولو» استفاده نمودند و با استفاده از الگوریتم یادگیری ماشین SVM[۲۴] توانستند به نتایج قابل قبولی دست یابند. معیارهای ارزیابی مورد استفاده در این پژوهش معیارهای precision و recall بود.

در سال ۲۰۲۰، کاکائی و همکارانش در قالب یک پایان‌نامه کارشناسی ارشد دیگر در دانشگاه شریف یک روش تولید جملات زبان فارسی از بازنمایی معنایی انتزاعی را ارائه نمودند [۳۹]. در این روش از الگوریتم‌های یادگیری عمیق و شبکه‌های عصبی به این منظور استفاده شد. به علاوه، برای آموزش این الگوریتم‌ها، دو مجموعه‌داده به صورت خودکار تولید شدند و ارزیابی روش با استفاده از این دو مجموعه‌داده و معیار ارزیابی BLEU انجام شده است.

با بررسی کارهای پیشین ملاحظه می‌شود که اکثر این کارها قابلیت بازنمایی معنا تنها در یک زبان طبیعی خاص (اکثرا انگلیسی) را داشته‌اند [۴۰]. اما در این مقاله هدف ارائه مدل بازنمایی معنایی است که وابسته به ویژگی‌های یک زبان طبیعی خاص نباشد. همچنین، در ادامه اشاره خواهد شد که برخلاف کارهای گذشته، مدل پیشنهادی قادر به بازنمایی روابط معنایی در سطح سند می‌باشد و بسیاری از روابط معنایی که در مدل‌های پیشین قابل بازنمایی نبوده‌اند، مانند زمان و همبستگی، در مدل پیشنهادی قابلیت بازنمایی دارند.

۳- مدل پیشنهادی بازنمایی معنا

مدل پیشنهادی بازنمایی معنا (ISR-Model[۲۵]) به صورت گراف که یال‌ها و گره‌ها به ترتیب بیانگر روابط[۲۶] و مفاهیم[۲۷] یا کلاس‌ها هستند، ارائه شده است. هر مفهوم در جمله دارای یک یا چند نمونه[۲۸] است که قبل از علامت «/» نشان داده می‌شود. مفاهیم در مدل پیشنهادی به زبان طبیعی مورد پردازش و روابط به صورت کلمات کلیدی در زبان انگلیسی بیان شده‌اند. شکل ۳ مثالی از بازنمایی معنایی یک جمله به زبان انگلیسی و زبان فارسی را با مدل بازنمایی معنای پیشنهادی، نشان می‌دهد.

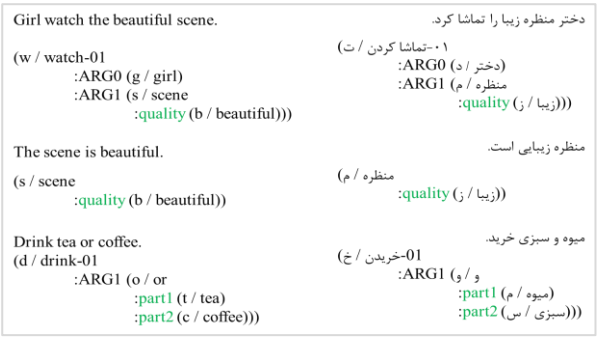


شکل ۳: مثالی از بازنمایی معنای یک جمله به زبان انگلیسی و فارسی

همان‌طور که در شکل ۳ ملاحظه می‌شود، برای بازنمایی مفاهیم از ریشه[۲۹] مفهوم و برای بازنمایی افعال برای زبان‌هایی که ابزار مربوط به ریشه‌یابی[۳۰] وجود دارد از lemma ویا مصدر استفاده می‌شود. همچنین، برای مفاهیمی که شامل بیش از یک معنی[۳۱] هستند، شماره مرتبط با معنای مورد نظر در پیکره‌ای[۳۲] مانند PropBank [۴۰][۳۳] برای زبان انگلیسی بعد از «-» در کنار مفاهیم نوشته می‌شود. اما در زبان‌هایی مانند فارسی که چنین منابعی در دسترس

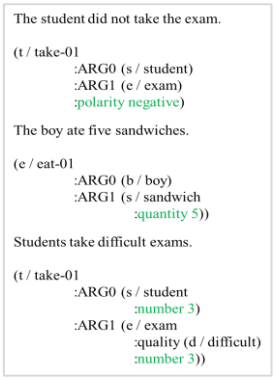
نیست، می‌توان ترتیب معناهای مرتبط با یک مفهوم در یک فرهنگ معتبر مانند فرهنگ معین[۳۴] را مورد نظر قرار داد. توجه به این نکته الزامی است که در نسخه فارسی PropBank [۴۱]، مانند نسخه انگلیسی، شماره معنای مفاهیم وجود ندارد.

در روش بازنمایی پیشنهادی، head جمله به عنوان رویداد[۳۵] و مفهوم اصلی در نظر گرفته شده و روابط در دو گروه کلی اصلی و فرعی تقسیم‌بندی شده‌اند. روابط اصلی، آرگومان‌های مفاهیم هستند، مانند ARG0 و ARG1 در شکل ۳، که با استفاده از پیکره‌ای مانند PropBank قابل استخراج هستند. روابط فرعی بیانگر سایر روابط معنایی موجود در متن هستند که هدف آن‌ها بازنمایی جزئیات معنایی مرتبط با موجودیت(های) درگیر در یک رخداد معنایی است. این روابط انواع مختلفی دارند، از جمله ویژگی‌های آرگومان‌ها، اطلاعات در مورد پدیده‌های زبانی (مانند زمان، مکان، کیفیت و تعداد) وغیره، موجود در متن که از پیش در مدل تعریف شده و برای بازنمایی متناسب با معنا مورد استفاده قرار می‌گیرند. نمونه‌هایی از روابط فرعی در شکل ۴ با رنگ سبز قابل ملاحظه است.



شکل ۴: مثال‌هایی از بازنمایی رابطه فرعی quality در مدل پیشنهادی

در شکل ۴، از رابطه فرعی quality برای بازنمایی معنایی جملاتی به زبان فارسی و انگلیسی استفاده شده است. این رابطه برای بازنمایی کیفیت یک مفهوم، که صریحا در متن بیان شده باشد، استفاده می‌شود. نوع دیگری از رابطه فرعی partn نام دارد که برای بازنمایی معنا زمانی که چندین کلمه با استفاده از «و» ویا «یا» (با یک نقش معنایی) در کنار دیگر قرار می‌گیرند، استفاده می‌شود. در این مثال‌ها مشاهده می‌شود که مفاهیم مرتبط با روابط فرعی، خود دارای نمونه‌ای هستند که در بازنمایی از آن استفاده می‌شود (مانند «س» نمونه «سبزی» یا t نمونه tea). علاوه بر رابطه‌های اینچنین، گروهی از روابط فرعی یا ویژگی‌ها هستند که با توجه به معنای جمله مقدار مشخصی می‌گیرند و نمونه‌ای ندارند. در این موارد، مقدار ذکر شده یا قابل استنتاج از معنای جمله به آن‌ها اختصاص داده می‌شود. شکل ۵ نمونه‌ای از کاربرد این روابط را نشان می‌دهد.



شکل ۵: مثال‌هایی از روابط فرعی با مقدار مشخص در مدل پیشنهادی

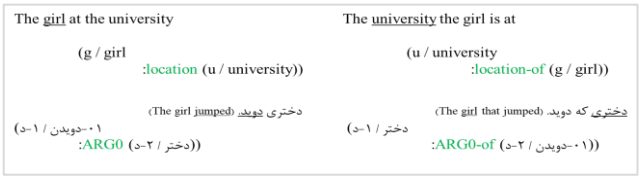
در شکل ۵، سه ویژگی polarity، quantity و number به ترتیب برای بازنمایی نفی، کمیت و تعداد در مدل پیشنهادی استفاده شده‌اند و به هر کدام با توجه به جمله مربوطه یک مقدار اختصاص داده شده است. ویژگی polarity برای بازنمایی معنای نفی رویداد در متن، ویژگی quantity کمیت یک مفهوم در زمانی که تعداد آن در متن صریحا ذکر شده باشد و ویژگی number بیانگر تعداد یک مفهوم، زمانی که اطلاعات صریحی در مورد آن در متن وجود ندارد، در نظر گرفته شده است. مثلا در جمله اول در شکل ۵ به دلیل وجود فعل نفی ویژگی polarity negative برای رویداد take در مدل بازنمایی ارائه شده است. همچنین برای جمله دوم در شکل ۵ به دلیل وجود کلمه five در مدل بازنمایی quantity five برای مفهوم sandwich ارائه شده و در انتها برای جمله سوم از ویژگی number جهت بازنمایی معنای جمع برای مفاهیم studentو exam استفاده شده است. همچنین در زبان‌هایی مانند فارسی و انگلیسی یک مفهوم می تواند مفرد یا جمع باشد اما در زبان عربی علاوه بر جمع و مفرد یک حالت دیگر وجود دارد که در آن پسوند «ان» نشان‌دهنده این است که دو نمونه از موجودیت در رویداد مشارکت داشته‌اند، مانند: «هناک شخصان». بدین منظور، در حالت کلی برای ویژگی number سه مقدار در مدل پیشنهادی در نظر گرفته شده است: مقدار ۱ به معنای مفرد، مقدار ۲ به معنای مشارکت دو نمونه از موجودیت و مقدار ۳ به معنای مشارکت دو نمونه یا بیشتر از موجودیت در رویداد که بیانگر همان مفهوم جمع است.

لازم به ذکر است که روابط فرعی محدود به موارد ذکر شده نیستند. در این راستا، جدول ۱ حاوی لیستی از روابط فرعی در نظر گرفته شده در مدل پیشنهادی است که به راحتی قابلیت توسعه و به‌کارگیری در مدل را دارد. در این لیست، کلمات به صورت کلیدی برای تمامی زبان‌های طبیعی به صورت یکسان در نظر گرفته شده است و نیاز به استفاده از معادل این کلمات در زبان پردازشی مورد نظر نیست (زیرا همان‌طور که اشاره شد روابط در مدل پیشنهادی به زبان انگلیسی هستند).

جدول۱: لیست روابط فرعی

accompanier	destination	mode	range
add-to	day-period	partn	similar-with
age	decade	number	source
coreference	degree	order	subset-of
aspect	name	part-of	time
beneficiary	direction	same-entity	time-after
cause	domain	path	time-before
century	duration	predicate	time-same
certainty	example	polarity	topic
compared-to	frequency	possession	unit
condition	instrument	purpose	value
concession	location	quality	weekday
consist-of	manner	quantifier	wiki
contrast-with	medium	quantity	

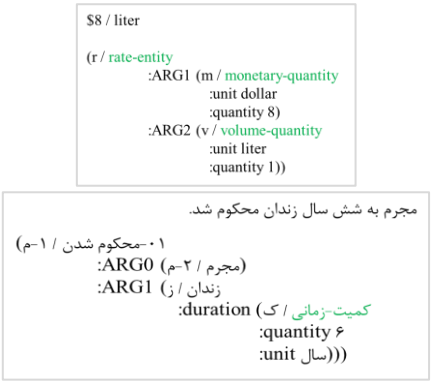
در زبان طبیعی گاهی جملات مشابه اما دارای تاکید متفاوت هستند که آن‌ها را، علیرغم شباهت، از لحاظ معنایی متمایز می‌نماید. روش بازنمایی پیشنهادی برای بازنمایی درست این پدیده، از روابط معکوس (با اضافه کردن of- در ادامه اسم رابطه) استفاده می‌کند. توجه به این نکته، به خصوص در کاربردهایی که نیاز به تولید متن[۳۶] دارند بسیار حائز اهمیت است. مثال‌هایی از بازنمایی روابط معکوس در شکل ۶ قابل ملاحظه هستند.



شکل ۶: مثال‌های بازنمایی روابط معکوس در مدل بازنمایی معنای پیشنهادی

در شکل ۶، دو جمله «دختری دوید» و «دختری که دوید» بسیار به هم شبیه هستند و در پیش‌پردازش بسیاری از کاربردهای زبان طبیعی ممکن است واژه «که» از جمله دوم حذف شود و دو جمله یکسان در نظر گرفته شوند. حال آنکه، این دو جمله از نظر معنایی از ابتدا با یکدیگر متفاوت بوده‌اند. به همین دلیل در مدل پیشنهادی با جابه‌جا کردن جای فرزند و والد در گراف بازنمایی و استفاده از رابطه معکوس، بازنمایی معنای این دو جمله متفاوت در نظر گرفته می‌شود تا بازنمایی با دقت بیشتری صورت گیرد.

در بازنمایی جمله فارسی و عبارت انگلیسی شکل ۷ ملاحظه می‌شود که برای بازنمایی یک مدت چند ساله و یک عبارت نرخی به ترتیب از مفهومی به نام «کمیت-زمانی» و مفهومی به نام rate-entity استفاده شده است. مفاهیمی از این دست که صریحا در متن بیان نشده‌اند و با استنتاج از معنای متن به دست می‌آیند، مفاهیم میانی نام دارند و برای بازنمایی معنای برخی جملات وجود مفاهیم میانی از پیش تعریف شده برای زبان طبیعی مورد پردازش الزامی است.



شکل ۷: مثال‌هایی از مفاهیم میانی در مدل بازنمایی معنای پیشنهادی

توجه به این نکته الزامی است که روابط مرتبط با یک مفهوم واحد (در یک سطح) مانند quantity و unit در مدل پیشنهادی مستقل از ترتیب بیان[۳۷] هستند، اما در این مقاله برای نگارش استاندارد ابتدا روابط اصلی و سپس روابط فرعی در گراف بازنمایی ذکر شده‌اند.

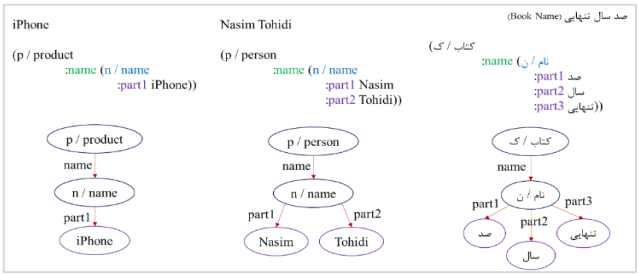
در جدول ۲ لیست مفاهیم میانی برای زبان فارسی و انگلیسی از پیش تعریف شده است. این لیست نیز قابلیت گسترش و تغییر را داراست و برای به‌کارگیری در زبان طبیعی مورد پردازش نیاز به ترجمه همین مفاهیم میانی به آن زبان طبیعی وجود دارد.

جدول ۲: لیست مفاهیم میانی

فارسی	انگلیسی
دارای-نقش-ارتباط	have-relation-role
دارای-هدف	have-purpose
کمیت-زمانی	temporal-quantity
کمیت-فاصله	distance-quantity
واقع-شدن-در	be-located-at
بازنمایی-ناشناخته	amr-unknown
توسط-خود	by-oneself
دارای-شرط	have-condition
دارای-نقش-سازمانی	have-organizational-role
دارای-مقدار	have-quant
دارای-قطبیت	have-polarity
دارای-درجه	have-degree
دارای-اذعان	have-concession
دارای-تکرار	have-frequency
به-جای	instead-of
درخواست-تایید	request-confirmation
کمیت-پولی	monetary-quantity
چند-جمله‌ای	multi-sentence
کشور-منطقه	country-region
موجودیت-تاریخ	date-entity
موجودیت-نرخ	rate-entity
موقعیت-نسبی	relative-position
تضاد-۰۰	contrast-00
شرط-۰۰	condition-00
دلیل-۰۰	cause-00

سه مفهوم میانی contrast-00، condition-00 و cause-00 در جدول ۲ در زمان‌هایی که در جمله از کلماتی مانند but، if و as ویا عبارات دیگر معادل آن‌ها استفاده شده است، بکار می‌روند. هر سه این مفاهیم بین دو عبارت ارتباط معنایی برقرار می‌کنند، بنابراین، هر کدام دو آرگومان به عنوان رابطه اصلی دارند تا ارتباط آن‌ها را با عبارت دو سمت آن‌ها به عنوان فرزندان برقرار نمایند. علت قرار دادن 00 به جای شماره معنی برای این مفاهیم این است تا در صورت استفاده صریح از این واژه‌ها، کاربرد آن‌ها به عنوان مفهوم میانی با کاربرد مفاهیم معمولی متمایز باشد.

بازنمایی موجودیت‌های نامدار[۳۸] نیز در مدل بازنمایی پیشنهادی در نظر گرفته شده است. در این مدل، لیستی از انواع موجودیت‌های نامدار ارائه شده است و برای بازنمایی در همه موارد، بعد از مشخص نمودن نوع موجودیت به عنوان یک مفهوم، از رابطه name به عنوان ویژگی آن مفهوم، که ممکن است از اجزای مختلفی تشکیل شده باشد، استفاده می‌گردد (شکل ۸).



شکل ۸: مثال‌هایی از بازنمایی موجودیت‌های نامدار در مدل بازنمایی پیشنهادی

همان‌طور که در شکل ۸ ملاحظه می‌شود، در گره ریشه ابتدا نوع موجودیت نامدار و سپس با رابطه فرعی name و مفهوم نام، اسم موجودیت مشخص می‌شود. اسم یک موجودیت نامدار خود شامل یک یا چند جزء است که با استفاده از part1، part2 و part3 بازنمایی می‌شوند. به این منظور از یک لیست از پیش تعریف شده مانند جدول ۳ برای موجودیت‌های نامدار به زبان انگلیسی و ترجمه آن به زبان طبیعی مد نظر می‌توان استفاده نمود که طبعا، این لیست نیز در آینده قابلیت تغییر و گسترش را داراست.

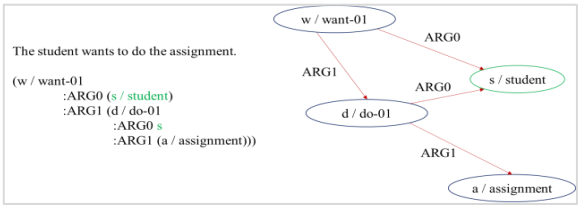
پدیده زبانی دیگر در مدل ارائه شده ساختار Reentrant است. زمانی که یک مفهوم چند نقش دارد (در جاهای مختلف به اشتراک گذاشته می‌شود و بیش از یک‌بار نیاز به ارجاع به آن مفهوم وجود دارد)، در قالب یک گره با چند والد در گراف نمایش داده می‌شود. مثال بازنمایی این پدیده در شکل ۹ قابل مشاهده است.

جدول ۳: لیست موجودیت‌های نامدار [۲۱]

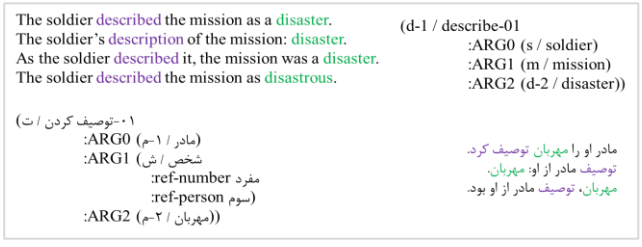
person	region	airport	game
family	continent	station	festival
animal	ocean	port	product
language	sea	tunnel	vehicle
nationality	lake	bridge	ship
group	river	road	aircraft
movement	gulf	railway	spaceship
organization	bay	canal	brand
company	strait	building	car
military	canal	theater	artwork
school	peninsula	museum	picture
university	mountain	palace	music
institute	volcano	hotel	show
team	valley	park	program
league	canyon	zoo	publication
location	island	amusement-park	book
city	desert	event	newspaper
district	forest	incident	magazine
county	planet	disaster	journal
state	star	earthquake	object
province	constellation	war	award
territory	facility	conference	law

در شکل ۹ ملاحظه می‌شود که نمونه s از مفهوم student هم آرگومان اول نمونه w از مفهوم want-01 و هم آرگومان اول نمونه d از مفهوم do-01 است. بنابراین، در روش بازنمایی، دو بار به آن ارجاع شده است. توجه به این نکته الزامی است که ارجاع به یک نمونه ممکن است هم در سطح جمله و هم در سطح سند صورت بگیرد.

یکی از اهداف اصلی، ارائه مدلی مستقل از ساختار نحوی زبان باشد. بنابراین، بازنمایی معنای جملاتی که ساختار نحوی و لغوی متفاوت، اما ساختار معنایی یکسان دارند باید یکسان باشد. مثال شکل ۱۰ نشان می‌دهد که روش پیشنهادی این قابلیت را داراست.



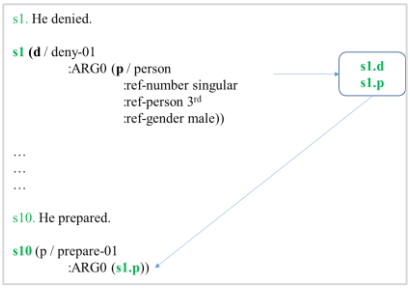
شکل ۹: مثالی از بازنمایی ساختار Reentrant در مدل بازنمایی پیشنهادی



شکل ۱۰: بازنمایی جملات هم‌معنا با ساختار متفاوت در مدل پیشنهادی

در شکل ۱۰، مشاهده می‌شود که برای بازنمایی ضمیر از روابط فرعی مختلف (ref-gender، ref-number و ref-person) استفاده می‌شود. این روابط فرعی با توجه به قواعد زبان‌های طبیعی مختلف ممکن است متفاوت باشند. به عنوان مثال، در زبان فارسی ضمیر

«او» اطلاعاتی در مورد جنسیت نمی‌دهند، اما در زبان انگلیسی ضمیر He رابطه فرعی ref-gender را در مدل پیشنهادی مانند شکل ۱۱ را دربردارد. همچنین امکان بازنمایی روابط بین‌جمله‌ای و در سطح سند با در نظر گرفتن شماره جمله (s1 در شکل ۱۱) در مدل پیشنهادی وجود دارد و براحتی می توان نمونه هایی که قبلا در جملات بیان شده اند با ارجاع به نمونه بیان شده در جمله مربوطه از بازنمایی مجدد نوع موجودیت خودداری کرد (s1.p).



شکل ۱۱: اضافه کردن شماره جمله در مدل بازنمایی معنای پیشنهادی

برای بازنمایی جنبه[۳۹] رویداد در مدل بازنمایی معنایی پیشنهادی، پنج دسته کلی زیر در نظر گرفته شده است که توسط رابطه aspect قابل بازنمایی هستند و در شکل ۱۲ قابل مشاهده است.

- ثابت[۴۰]: این جنبه مربوط به رویدادهای ثابت است که هیچ تغییری در طول رویداد رخ نمی‌دهد. رویدادهایی که در یک زمان مشخص یا در یک وضعیت[۴۱]، مفهومی را توصیف می‌کنند. این دسته از رویدادها نه می‌توانند به صورت استمراری و نه امری بیان شوند و همچنین قابلیت توصیف توسط قیود را هم ندارند. مانند: I know my departure time. یا من زمان عزیمت را می‌دانم.
- عادی[۴۲]: این جنبه مربوط به رویدادهای منظم و عادت در گذشته یا حال است. در زبان انگلیسی برای نشان دادن یک عادت در گذشته از عبارت used to استفاده می‌شود. مانند: نسیم سحرخیز است یا Nasim is an early bird یا I used to stay there frequently.
- فعال[۴۳]: این جنبه نشان می‌دهد که یک رویداد لزوما در زمان مشخص به پایان نرسیده و ممکن است در زمان ایجاد سند ادامه داشته باشد. این دسته از رویدادها به صورت استمرار و امری قابل بیان هستند. اما مثلا در زبان انگلیسی با حرف اضافه in جهت بیان مدت زمان رویداد ارایه نمی‌شوند و معمولا از حرف اضافه for استفاده می‌شود. مانند: من در حال سفر به آلمان هستم یا I am traveling to Germany.
- دستاورد[۴۴]: این جنبه برای رویداهایی است به صورت ذاتی در زمان نقطه پایان دارند و حاصل رویداد دیگری هستند که در زبان انگلیسی با حرف اضافه in جهت بیان مدت زمان رویداد ارایه می‌شوند، ولی از for برای آن‌ها استفاده نمی‌شود. همچنین، از کلمه stop می‌توان برای آن‌ها استفاده کرد. مانند: او پرواز را برای من رزرو کرد یا She booked me a flight.
- کامل[۴۵]: این جنبه برای رویدادهایی است که در یک وضعیت به نتیجه کامل می‌رسند. این دسته از رویدادها در زبان انگلیسی با کلمه stop و for قابل بیان نمی‌باشند. مانند: او درگاه ورودی پرواز خود را پیدا کرد یا She found her gate.

```
s1. He denied accusations.

s1 (d / deny-01
        :ARG0 (p / person
                :ref-number singular
                :ref-person 3rd
                :ref-gender male)
        :ARG1 (a / accusation
                :number 3)
        :aspect complete)
```

شکل ۱۲: بازنمایی جنبه در مدل بازنمایی معنای پیشنهادی

برخی کلمات مانند every و some در زبان انگلیسی و «بعضی» و «همه» در زبان فارسی بر معنای جمله تاثیرگذار بوده و تغییر این کلمات معنای جمله را تغییر می‌دهد، بنابراین بازنمایی آن‌ها بسیار مهم است. در مدل پیشنهادی جهت بازنمایی معنای این کلمات از ویژگی quantifier که بیانگر سورهای وجودی[۴۶] و جهانی[۴۷] استفاده شده است (شکل ۱۳). برای بازنمایی این پدیده زبانی مهم لازم است از قبل لیست کلمات مرتبط با این پدیده در زبان طبیعی مورد پردازش معین شده باشندکه این مهم با مراجعه به کارهای پیشین یا بررسی توسط زبان‌شناس مرتبط به راحتی قابل دستیابی است.

```
s1. Someone answered all questions.

s1 (a / answer-01
        :ARG0 (p / person
                :number 1
                :quantifier some)
        :ARG1 (q / question
                :number 3
                :quantifier all))

s2. Some questions were answered by everyone.

s2 (q / question
        :number 3
        :quantifier some
        :ARG0-of (a / answer-01
                :ARG1 (p / person
                        :quantifier every)))
```

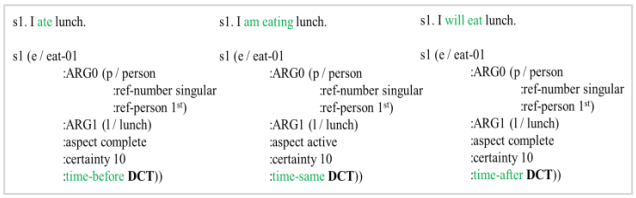
شکل ۱۳: بازنمایی سورها در مدل بازنمایی معنای پیشنهادی

برخی از کلمات مانند «احتمالا» و might به ترتیب در زبان فارسی و انگلیسی به‌طور ذاتی دربرگیرنده ابهام در معنای جمله بوده و سبب بروز عدم‌قطعیت[۴۸] در بیان معنای جمله می‌شوند. به منظور بازنمایی عدم‌قطعیت در مدل پیشنهادی از ویژگی certainty و مقادیر کمی از پیش تعریف شده، طبق جدول ۴، برای نشان دادن درجه قطعیت استفاده شده است. تبدیل مقادیر کیفی به کمی می‌تواند در هر بازه از مقادیر بنابر تنوع و تعدد این کلمات در زبان طبیعی مورد پردازش در نظر گرفته شود. انعطاف‌پذیری مدل از این حیث باعث می‌شود، علیرغم تفاوت‌ها، این پدیده به راحتی قابل بازنمایی در مدل باشد. در جدول ۴، بازه ۱ تا ۱۰ برای درجه قطعیت، که ۱۰ بیانگر اطمینان کامل در معنای جمله است، در نظر گرفته شده است.

جدول ۴: بازنمایی عدم‌قطعیت در مدل بازنمایی معنای پیشنهادی

مثال	توضیح	مقدار
The event held.	زیاد	۱۰
The event probably held.	متوسط	۵
The event might have held.	کم	۱

از آنجایی‌که بازنمایی پدیده زمان در جمله ممکن است تنها به یک رویداد در جمله مرتبط نباشد و هر رویداد خود مرجع رویداد دیگری باشد و ترتیب زمانی مشخصی بین رویدادها وجود داشته باشد، لذا در مدل بازنمایی معنای پیشنهادی از کلمه DCT[۴۹]، به معنای زمان ساخت سند، به عنوان یک مرجع زمانی و از روابط فرعی time-before، time-after و time-same برای بازنمایی تقدم و تاخر نسبت به این زمان مرجع استفاده شده است (شکل ۱۴). شکل ۱۵ نمونه‌ای از توالی رخدادها با زمان مشابه افعال در یک سند را نشان می‌دهد.

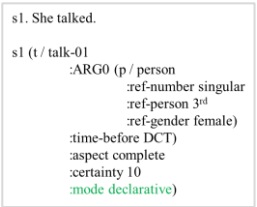


شکل ۱۴: مثال‌هایی از بازنمایی زمان رویداد در مدل بازنمایی معنای پیشنهادی

Ali went into the shop. He had asked Maryam what she needed. He bought several things.

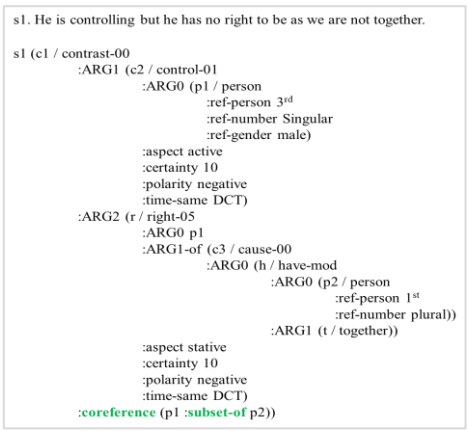
شکل ۱۵: بازنمایی زمان در سطح سند در مدل بازنمایی معنای پیشنهادی

برای بازنمایی نوع جمله در مدل بازنمایی معنای پیشنهادی از ویژگی mode با مقدارهای از پیش‌تعریف شده مانند imperative یا امری، interrogative یا پرسشی، exclamative یا تعجبی و declarative یا اخباری در شکل ۱۶ استفاده شده است.



شکل ۱۶: مثالی از بازنمایی نوع جمله در مدل بازنمایی پیشنهادی

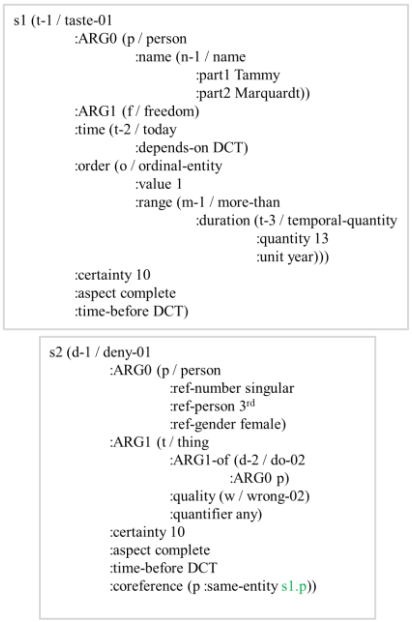
پدیده زبانی مهم دیگر همبستگی است که خود شامل دو نوع همبستگی موجودیت[۵۰] و همبستگی رویداد[۵۱] است [۴۲]. هر دو این موارد با تعریف رابطه فرعی coreference در مدل پیشنهادی قابل بازنمایی هستند (شکل ۱۷).



شکل ۱۷: مثالی از بازنمایی همبستگی در مدل بازنمایی معنای پیشنهادی

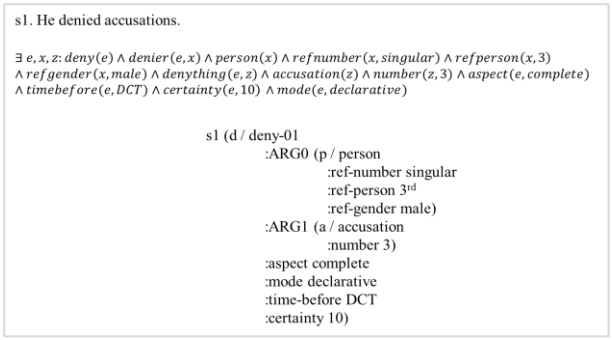
نمونه‌ای از بازنمایی معنای دو جمله با در نظر گرفتن روابط آن‌ها در سطح سند را در شکل ۱۸ مشاهده می‌نمایید.

s1. Tammy Marquardt tasted freedom today for the first time in more than 13 years.
s2. She denied any wrongdoing.



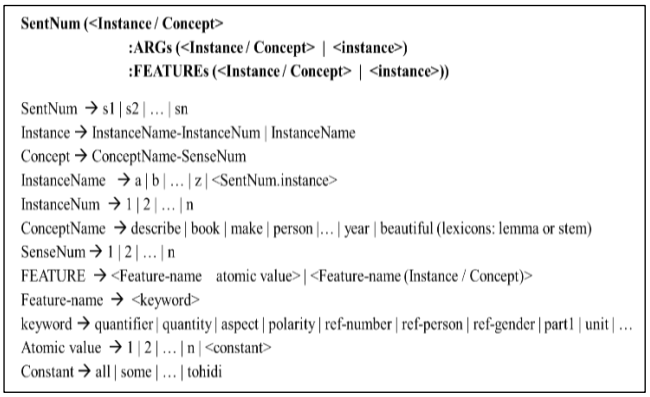
شکل ۱۸: مثال بازنمایی ارتباطات معنایی در سطح سند در مدل پیشنهادی

یکی از اهداف طراحی مدل بازنمایی معنای پیشنهادی قابلیت استنتاج منطقی است. تمامی پدیده‌های ارائه شده توسط مدل بازنمایی پیشنهادی به راحتی قابل بیان توسط منطق می‌باشند که در شکل ۱۹ نمونه‌ای از معادل بازنمایی منطق مدل پیشنهادی بر طبق روش neo-Davidsonian [۴۳] را مشاهده می‌نمایید.



شکل ۱۹: بازنمایی جمله در قالب منطق و مدل بازنمایی معنای پیشنهادی

ساختار رسمی[۵۲] مدل بازنمایی معنای پیشنهادی (ISR-Model) که در برگیرنده تمامی پدیده‌های ذکر شده در این مقاله هست را در شکل ۲۰ مشاهده می‌نمایید.



شکل ۲۰: ساختار رسمی مدل بازنمایی معنای پیشنهادی (ISR-Model)

۴- ارزیابی مدل پیشنهادی

در جدول ۵ مدل بازنمایی معنایی پیشنهادی (ISR-Model) با یکی از مدل‌های مرتبط پیشین از حیث توانایی بازنمایی برخی از مهم‌ترین پدیده‌های معنایی زبان مقایسه شده است.

جدول ۵: مقایسه مدل بازنمایی معنای پیشنهادی با مدل بازنمایی انتزاعی

قابلیت بازنمایی	بازنمایی معنایی انتزاعی [۲۱]	مدل پیشنهادی
نفی	دارد	دارد
موجودیت‌های نامدار	دارد	دارد
زمان	ندارد	دارد
عدم قطعیت	ندارد	دارد
جنبه	ندارد	دارد
ساختار Reentrant	دارد	دارد
روابط معکوس	دارد	دارد
روابط در سطح جمله	دارد	دارد
روابط در سطح سند	ندارد	دارد
همبستگی	ندارد	دارد
نوع جمله	ندارد	دارد
سورها	ندارد	دارد
زبان‌های طبیعی مختلف	ندارد	دارد

همان‌طور که در جدول ۵ مشاهده می‌شود، روش بازنمایی پیشنهادی در مقایسه با یکی از اصلی‌ترین کارهای مرتبط پیشین که در بسیاری از حوزه‌های پردازش زبان طبیعی کاربرد داشته است، برتری‌های فراوانی دارد. به عنوان مثال، امکان بازنمایی زمان در کاربردهایی مانند ترجمه ماشینی و امکان بازنمایی همبستگی در کاربردهایی مانند خلاصه‌سازی خودکار متن اهمیت بسیار داشته و مدل پیشنهادی قابلیت بازنمایی چنین مفاهیم معنایی را دارا است.

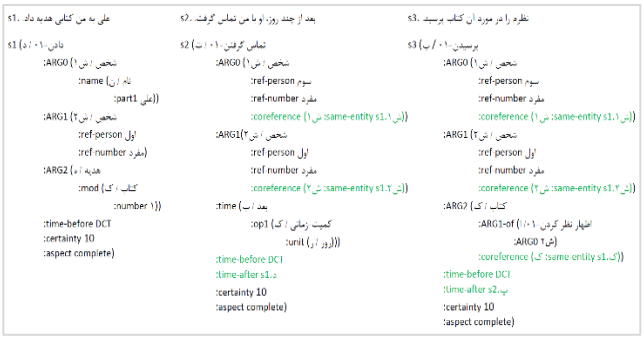
جمله «من شام خورده بودم.» را در نظر بگیرید. اگر مدل قابلیت بازنمایی زمان را نداشته باشد، در کاربردی مانند ترجمه ماشینی، این جمله به هر کدام از این جملات قابل ترجمه خواهد بود:

I eat dinner. یا I am eating dinner. یا I have eaten dinner.

یا I ate dinner. یا I had eaten dinner. یا I will eat dinner.

در مدل پیشنهادی، به دلیل بازنمایی زمان، جمله دقیقا به جمله معادل آن یعنی .I had eaten dinner ترجمه خواهد شد.

به عنوان نمونه‌ای دیگر، سند حاوی جملات «علی به من کتابی هدیه داد. بعد از چند روز، او با من تماس گرفت. نظرم را در مورد آن کتاب پرسید.» را در نظر بگیرید. در کاربردی مانند خلاصه‌سازی خودکار متن، تشخیص اینکه در جمله سوم منظور از «آن کتاب» همان کتاب جمله اول است و منظور از «او» در جمله دوم «علی» است، بسیار مهم خواهد بود. بنابراین، در صورتی که مدل بازنمایی امکان تشخیص روابط معنایی در سطح سند و پیوستگی زمانی وقایع و همچنین بازنمایی همبستگی را نداشته باشد، کیفیت سیستم خلاصه‌ساز مبتنی بر آن مدل بسیار پایین خواهد آمد. بازنمایی این سند کوتاه توسط روش پیشنهادی در شکل ۲۱ قابل مشاهده است.



شکل ۲۱: بازنمایی معنا توسط مدل پیشنهادی برای یک سند کوتاه

به همین دلیل در مدل بازنمایی معنایی پیشنهادی طیف وسیعی از پدیده‌های زبانی که در هر یک از کاربردهای پردازش زبان طبیعی مهم و تاثیرگذار هستند، در نظر گرفته شده است.

علاوه بر موارد ذکر شده، توجه به این نکته بسیار مهم است که اکثریت قریب به اتفاق کارهای گذشته انجام شده در رابطه با بازنمایی معنایی زبان، صرفا با یک زبان طبیعی کار می‌کردند. اما مدل پیشنهادی در این مقاله صرفا برای یک زبان طبیعی خاص طراحی نشده است. زیرا، در هیچ یک از بخش‌های ساختار روش پیشنهادی از ویژگی‌های یک زبان طبیعی خاص ویا ابزار موجود در یک زبان طبیعی مشخص استفاده نشده است. همچنین، در مورد منابع مورد نیاز و لیست‌های از پیش تعریف شده برای حاشیه‌نویسی پیکره مبتنی بر این مدل اشاره شد که چگونه می‌توان حاشیه‌نویسی را برای زبان‌های طبیعی مختلف، چه دارای منابع گسترده باشند چه محدود، به کار گرفت.

برای ارزیابی چگونگی استفاده از مدل پیشنهادی در کاربردهای عملی، در این پژوهش دو مجموعه‌داده به صورت دستی توسط عامل انسانی تولید شدند، که هر کدام شامل ۶۰۲ جمله از ۱۰ فصل اول کتاب «شازده کوچولو»[۵۳] به ترتیب به زبان انگلیسی و ترجمه فارسی آن‌ها به همراه مدل بازنمایی معنایی پیشنهادی است. اطلاعات این دو پیکره در جدول ۶ قابل ملاحظه است.

جدول ۶: خصوصیات مجموعه‌داده‌ها

تعداد جملات	۶۰۲
میانگین طول جملات	۱۱.۸۳
طول کوتاه‌ترین جمله	۱
طول بلندترین جمله	۵۷

بعد از تولید مجموعه‌داده‌های مذکور، تجزیه‌گری[۵۴]، مبتنی بر تجزیه‌گر ارائه شده در [۴۵]، بر روی این مجموعه‌داده‌ها اعمال شد. سپس، بازنمایی جملات تجزیه شده توسط این تجزیه‌گر با استفاده از معیار ارزیابی استاندارد Smatch [۴۶]، که فرمول آن در (۱) قابل مشاهده است، مورد ارزیابی قرار گرفت (جدول ۷).

$$Smatch = 2 * \frac{\frac{M}{T}*\frac{M}{G}}{\frac{M}{T}+\frac{M}{G}} \qquad (۱)$$

معیار Smatch، شبیه به F-score عمل کرده و شباهت دو گراف بازنمایی را بر حسب یال‌های منطبق با هر گره نگاشت شده محاسبه می‌کند. در فرمول (۱)، M تعداد یال‌های منطبق، T تعداد تمام یال‌ها در گراف بازنمایی اولیه موجود و G تعداد تمام یال‌ها در گراف بازنمایی مدل پیشنهادی است.

جهت ارزیابی از روش اعتبارسنجی متقابل k-fold بر روی دو مجموعه‌داده ذکر شده با مقدار ۱۰ برای k، که هر بار یک فصل از مجموعه‌داده به عنوان داده آزمون و ۹ فصل دیگر به عنوان داده آموزشی در نظر گرفته شده است و در نهایت میانگین مقادیر به دست آمده از این ارزیابی‌ها به عنوان نتیجه نهایی گزارش شده است. دلیل تقسیم‌بندی میزان داده آموزش و آزمون مبتنی بر فصول، و نه مبتنی بر تعداد جملات، این است که در مدل پیشنهادی یکی از اصلی‌ترین ویژگی‌ها مربوط به بازنمایی پدیده همبستگی است؛ و برای آموزش تجزیه‌گر در رابطه با این ویژگی لازم است تا جملات به هم مرتبط، به صورت یکجا و پیوسته، به عنوان داده آموزشی به تجزیه‌گر داده شوند. بنابراین، انتخاب جملات تصادفی از مجموعه‌داده‌ها به عنوان داده آزمون نتیجه مناسبی در خروجی تجزیه‌گر به همراه نخواهد داشت.

جدول ۷: نتایج ارزیابی تجزیه‌گر جملات مجموعه‌داده توسط مدل بازنمایی معنایی پیشنهادی

مجموعه‌داده	Smatch
فارسی	۶۰.۴
انگلیسی	۶۳.۰

همان‌طور که در جدول ۷ مشاهده می‌شود مدل بازنمایی معنایی پیشنهادی قابلیت تجزیه خودکار و استفاده در کاربردهای پردازش زبان طبیعی را با دقت قابل قبولی داراست.

۵- نتیجه‌گیری و کارهای آتی

در این مقاله یک روش بازنمایی معنای یکپارچه به نام ISR-Model که کاربردهای بسیاری در حوزه پردازش زبان طبیعی مانند خلاصه‌سازی، ترجمه ماشینی، پرسش و پاسخ و غیره دارد، ارائه شده است. مدل بازنمایی معنای پیشنهادی دارای ساختار گراف است که با نمایش ساختار لیست قابلیت ذخیره و پردازش توسط الگوریتم‌های مختلف را دارد.

مدل بازنمایی معنایی پیشنهادی برای زبان‌های طبیعی، بدون وابستگی به عناصر یک زبان طبیعی خاص، طراحی شده و از ویژگی‌های نحوی و لغوی زبان نیز مستقل است. این مدل ساده و قابل فهم بوده و کلیه پدیده‌های در نظر گرفته شده قابلیت تبدیل به روش بازنمایی منطق مرتبه اول (روشی انعطاف‌پذیر، اثبات‌پذیر و قابلیت استنتاج) را دارا می‌باشد. اجزای اصلی مدل پیشنهادی شامل مفاهیم (اصلی و میانی)، نمونه‌ها، آرگومان‌ها (روابط اصلی) و ویژگی‌ها (روابط فرعی) با امکان پدیده‌های زبانی در سطح جمله و در سطح سند است و قابلیت بازنمایی پدیده‌های مهم معنایی مانند موجودیت‌های نامدار، سورها، جنبه، نوع جمله، نفی، کمیت، تعداد، روابط معکوس، همبستگی، ساختار Reentrant و قطعیت را دارد.

همه این ویژگی‌ها مدل پیشنهادی را در مقایسه با مدل بازنمایی معنایی انتزاعی که در سال ۲۰۱۳ توسط Banarescu و همکاران معرفی شد و در سال‌های اخیر بسیار مورد توجه قرار گرفت، مدلی کاراتر و جامع‌تر نموده است.

در این مقاله مثال‌هایی از زبان انگلیسی و فارسی دال بر عدم وابستگی مدل پیشنهادی به زبان طبیعی خاصی ارائه شده است. منظور از عدم وابستگی، قابلیت استفاده از مدل پیشنهادی برای زبان‌های طبیعی مختلف و عدم تکیه بر ویژگی‌های یک زبان طبیعی خاص است. طبعا به‌کارگیری مدل پیشنهادی فاقد چالش نیست. یکی از مهم‌ترین این چالش‌ها نیاز به منابع زبانی مانند پیکره‌ها و برخی لیست‌های از پیش تعریف شده در زبان‌های طبیعی مختلف است. توجه به این نکته الزامی است که در مدل پیشنهادی تنها به چند منبع اندک و محدود برای استفاده در زبان‌های طبیعی مختلف نیاز است و چگونگی تهیه این منابع در زبان‌های طبیعی با ویژگی‌ها و میزان منابع متفاوت، همان‌طور که توضیح داده شد، بسیار ساده است. اصلی‌ترین پیکره مورد نیاز پیکره‌ای برای تشخیص معانی مختلف یک واژه در زبان و پیکره‌ای برای تعیین آرگومان‌های یک رویداد در جمله است. تاکنون نسخه اولیه لیست‌های مورد نیاز (روابط فرعی، مفاهیم میانی و موجودیت‌های نامدار) تهیه شده‌اند که در مقاله به آن‌ها اشاره شد. در مورد تشخیص معانی مختلف یک واژه در زبان به نظر می‌رسد این چالش تا حد زیادی با استفاده از فرهنگ‌هایی که در اکثر زبان‌های طبیعی وجود دارند قابل حل است.

همچنین، در این مقاله دو مجموعه داده کوچک به زبان‌های فارسی و انگلیسی برای مدل پیشنهادی تولید شدند و نتایج اعمال یک تجزیه‌گر بر روی این مجموعه‌داده‌ها گزارش شد.

یک گام مهم در ادامه این پژوهش به عنوان کار آتی تولید پیکره‌های استاندارد و بزرگ برای روش پیشنهادی در زبان‌های طبیعی متنوع است؛ تا بتوان از این پیکره‌ها برای طراحی، پیاده‌سازی و آموزش تجزیه‌گرها و تولیدکننده‌های متون زبان‌های طبیعی استفاده نمود. سپس، می‌توان از این ابزارها برای بهبود عملکرد کاربردهای مختلف پردازش زبان طبیعی مانند سامانه‌های خلاصه‌ساز خودکار متن و سامانه‌های پرسش و پاسخ که به زبان‌های طبیعی مختلف کار می‌کنند، استفاده نمود.

مراجع

[1] Tohidi, Nasim; Hasheminejad, Seyed Mohammad Hossein;, "MOQAS: Multi-objective question answering system," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 4, pp. 3495-3512, 2019.

[2] Peng, Xiaochang; Gildea, Daniel, "UofR at SemEval-2016 task 8: Learning synchronous hyperedge replacement grammar for AMR parsing," in *In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California, 2016.

[3] Peng, Xiaochang; Song, Linfeng; Gildea, Daniel, "A synchronous hyperedge replacement grammar based approach for AMR parsing," in *In Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, Beijing, China, 2015.

[4] Flanigan, Jeffrey; Thomson, Sam; Carbonell, Jaime; Dyer, Chris; Smith. Noah A., "A Discriminative Graph-Based Parser for the Abstract Meaning Representation," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, 2014.

[5] C. Wang, "Abstract Meaning Representation Parsing," PhD thesis, Brandeis University, 2018.

[6] Damonte, Marco; Cohen, Shay B.; Satta, Giorgio, "An Incremental Parser for Abstract Meaning Representation," in *Proceedings of EACL*, Valencia, Spain, 2017.

[7] Gu, Min; Gu, Yanhui; Luo, Weilan; Xu, Guandong; Yang, Zhenglu; Zhou, Junsheng; Qu, Weiguang, "From text to graph: a general transition-based AMR parsing using neural network," *Neural Computing and Applications*, 2020.

[8] Barzdins, Guntis; Gosko, Didzis, "RIGA at SemEval- 2016 Task 8: Impact of Smatch Extensions and Character-Level Neural Translation on AMR Parsing Accuracy," in *In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California, 2016.

[9] W. C. Mann, "An overview of the Penman text generation system," in *In Proceedings of the Third AAAI Conference on Artificial Intelligence*, 1983.

[10] Konstas, Ioannis; Iyer, Srinivasan; Yatskar, Mark; Choi, Yejin; Zettlemoyer, Luke, "Neural AMR: Sequence-to-Sequence Models for Parsing and Generation," in *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 2017.

[11] Pust, Michael; Hermjakob, Ulf; Knight, Kevin; Marcu, Daniel; May, Jonathan, "Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation," in *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015.

[12] Tohidi, Nasim; Dadkhah, Chitra, "A Study on Abstract Meaning Representation Applications," in *The first Conference on Artificial Intelligence and Smart Computing*, Online, 2022.

[13] Jahanbakhsh-Nagadeh, Zoleikha; Feizi-Derakhshi, Mohammad-Reza; Sharifi, Arash;, "A Speech Act Classifier for Persian Texts and its Application in Identifying Rumors," *Journal of Soft Computing and Information Technology*, vol. 9, no. 1, pp. 18-27, 2020.

[14] Miri; Ali; Razavi, Seyyed Mohammad; Miri, Ismail;, "Ensemble Recognition of Persian Typed Sub-word in limited Space Using Smart weighted voting," *Journal of Soft Computing and Information Technology (JSCIT)*, vol. 10, no. 3, pp. 60-72, 2021.

[15] Farahani, Mehrdad; Gharachorloo, Mohammad; Farahani, Marzieh; Manthouri, Mohammad, "ParsBERT: Transformer-based Model for Persian Language Understanding," *Neural Processing Letters*, vol. 53, p. 3831–3847, 2021.

[16] Tohidi, Nasim; Dadkhah, Chitra; Rustamov, Rustam B., "Optimizing Persian Multi-objective Question Answering System," *International Journal on Technical and Physical Problems of Engineering (IJTPE)*, vol. 13, no. 46, pp. 62-69, 2021.

[17] Kazemi, Arefeh; Mozafari, Jamshid; Nematbakhsh, Mohammad Ali, "PersianQuAD: The Native Question Answering Dataset for the Persian Language," *IEEE Access*, vol. 10, pp. 26045-26057, 2022.

[18] Tohidi, Nasim; Hasheminejad, Seyed Mohammad Hossein, "A Practice of Human-Machine Collaboration for Persian Text Summarization," in *The 27th International Computer Conference, the Computer Society of Iran*, Online, 2022.

[19] Poon, Hoifung; Domingos, Pedro, "Unsupervised Semantic Parsing," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapore, 2009.

[20] Chen, Desai; Schneider, Nathan; Das, Dipanjan; Smith, Noah A., "SEMAFOR: Frame Argument Resolution with Log-Linear Models," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, 2010.

[21] Banarescu, Laura; Bonial, Claire; Cai, Shu; Georgescu, Madalina; Griffitt, Kira; Hermjakob, Ulf; Knight, Kevin; Koehn, Philipp; Palmer, Martha; Schneider, Nathan, "Abstract Meaning Representation for Sembanking," in *In proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, Sofia, Bulgaria, 2013.

[22] Kingsbury, Paul; Palmer, Martha, "From TreeBank to PropBank," in *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain, 2002.

[23] K. Cao, "Learning meaning representations for text generation with deep generative models," PhD thesis, University of Cambridge, 2019.

[24] Knight, Kevin; Badarau, Bianca; Baranescu, Laura; Bonial, Claire; Bardocz, Madalina; Griffitt, Kira; Hermjakob, Ulf; Marcu, Daniel; Palmer, Martha; O'Gorman, Tim; Schneider, Nathan, "Abstract Meaning Representation (AMR) Annotation Release 2.0," Linguistic Data Consortium, Philadelphia, 2017.

[25] Van Gysel, Jens E.L.; Vigus, Meagan; Chun, Jayeol; Lai, Kenneth; Moeller, Sarah; Yao, Jiarui; O'Gorman, Tim; Cowell, Andrew; Croft, William; Huang, Chu-Ren; Hajič, Jan; Martin, James H.; Oepen, Stephan; Palmer, Martha; Pustejovsky, James; et al., "Designing a Uniform Meaning Representation for Natural Language Processing," *KI - Künstliche Intelligenz*, vol. 35, p. 343–360, 2021.

[26] Damonte, Marco; Cohen, Shay B., "Cross-Lingual Abstract Meaning Representation Parsing," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, Louisiana, 2018.

[27] Flanigan, Jeffrey; Dyer, Chris; Smith, Noah A.; Carbonell, Jaime, "Generation from abstract meaning representation using tree transducers," in *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, 2016b.

[28] Gruzitis, Normunds; Gosko, Didzis; Barzdins, Guntis, "RIGOTRIO at SemEval-2017 Task 9: Combining Machine Learning and Grammar Engineering for AMR Parsing and Generation.," in *In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017.

[29] Song, Linfeng; Peng, Xiaochang; Zhang, Yue; Wang, Zhiguo; Gildea, Daniel, "AMR-to-text Generation with Synchronous Node Replacement Grammar," in *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vancouver, Canada, 2017.

[30] Song, Linfeng; Zhang, Yue; Peng, Xiaochang; Wang, Zhiguo; Gildea, Daniel, "AMR-to-text generation as a Traveling Salesman Problem," in *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, 2016.

[31] Mille, Simon; Carlini, Roberto; Burga, Alicia; Wanner, Leo , "FORGe at SemEval-2017 Task 9: Deep sentence generation based on a sequence of graph transducers," in *In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017.

[32] Lampouras, Gerasimos; Vlachos, Andreas, "Sheffield at SemEval-2017 Task 9: Transition-based language generation from AMR," in *In Proceedings of the 11th InternationalWorkshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017.

[33] T. Schick, "Transition-Based Generation from Abstract Meaning Representations," Master's thesis, 2017.

[34] Y. H. Gao, "English Semantic Recognition Based on Abstract Meaning Representation and Vector Space Model," *Journal of Physics: Conference Series*, vol. 1087, no. 5, 2018.

[35] Ribeiro, Leonardo; Gardent, Claire; Gurevych, Iryna, "Enhancing AMR-to-Text Generation with Dual Graph Representations," in *Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, 2019.

[36] Lai, Kenneth; Donatelli, Lucia; Pustejovsky, James, "A Continuation Semantics for Abstract Meaning Representation," The Second International Workshop on Designing Meaning Representations, Barcelona, Spain, 2020.

[37] Ilmy, Adylan Roaffa; Khodra, Masayu Leylia, "Parsing Indonesian Sentence into Abstract Meaning Representation using Machine Learning Approach," *arXiv*, 2021.

[38] S. Haddadan and M. Bahrani, "Persian Abstractive Summarization using Graph-based Abstract Meaning Representation," Languages and Linguistics Center, Sharif University of Technology, Tehran, Iran, 2017.

[39] Kakaei, Farokh; Rahimi, Saeed; Bahrani, Mohammad, "Generating Text from Abstract Meaning Representation in Persian," Languages and Linguistics Center, Sharif University of Technology, Tehran, Iran, 2020.

[40] Tohidi, Nasim; Dadkhah, Chitra, "A Short Review of Abstract Meaning Representation Applications," *Journal of Modeling & Simulation in Electrical & Electronics Engineering*, vol. 2, no. 3, pp. 1-9, 2022.

[41] Palmer, Martha; Gildea, Daniel; Kingsbury, Paul, "The Proposition Bank: An Annotated Corpus of Semantic Roles," *Computational Linguistics*, vol. 31, no. 1, pp. 71-106, 2005.

[42] Mirzaei, Azadeh; Moloodi, Amirsaeid, "Persian Proposition Bank," in *10th edition of the Language Resources and Evaluation Conference*, 2016.

[43] Barhom, Shany; Shwartz, Vered; Eirew, Alon; Bugert, Michael; Reimers, Nils; Dagan, Ido;, "Revisiting Joint Modeling of Cross-document Entity and Event Coreference Resolution," in *The 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019.

[44] Jurafsky, Daniel; Martin, James H. , Speech and Language Processing, Prentice Hall, 2021.

[45] Guo, Zhijiang; Lu, Wei;, "Better Transition-Based AMR Parsing with a Refined Search Space," in *The 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018.

[46] Cai, Shu; Knight, Kevin, "Smatch: an Evaluation Metric for Semantic Feature Structures," in *The 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, Bulgaria, 2013.

پاورقی‌ها:

[1] https://www.nature.com/articles/d41586-023-00107-z

[2] Lexical

[3] Syntactic

[4] Semantic
[5] Contextual
[6] Synchronous Hyper-edge Replacement Grammar
[7] Context Free Grammar
[8] Hyper-edge Replacement Grammar
[9] Maximum Spanning Connected sub-Graph
[10] Action
[11] Stack
[12] Sparsity
[13] Unsupervised Semantic Parsing
[14] Lambda
[15] Tree-transducer-based
[16] Graph-grammar-based
[17] Synchronous node replacement grammar
[18] Transition-based
[19] Vector Space Model
[20] First Order Predicate Logic
[21] Pair prediction
[22] Label prediction
[23] Graph construction
[24] Support Vector Machine
[25] Integrated Semantic Representation-Model
[26] Relation
[27] Concept
[28] Instance
[29] stem
[30] lemmatization
[31] sense
[32] corpus
[33] https://propbank.github.io/
[34] https://abadis.ir/moeen/
[35] event
[36] Text Generation
[37] order-independent
[38] Named Entities
[39] Aspect
[40] Stative
[41] State
[42] Habitual
[43] Active
[44] Accomplishment
[45] Complete
[46] Existential Quantifier
[47] Universal Quantifier
[48] Uncertainty
[49] Document Creation Time
[50] Entity Coreference
[51] Event Coreference
[52] Formal
[53] https://amr.isi.edu/download/amr-bank-struct-v3.0.txt
[54] Parser