
بسم الله الرحمن الرحیم

دانشکده مهندسی کامپیوتر

رساله دکتری مهندسی هوش مصنوعی

یک مدل مبتنی بر یادگیری عمیق برای تحلیل هم‌ارجاعی عبارات اسمی

نگارنده: سمیرا حورعلی

استاد راهنما

دکتر مرتضی زاهدی

استاد مشاور

دکتر منصور فاتح

آذر ۱۳۹۹

صورت جلسه نهایی دفاع از رساله دکتری (Ph.D)

شماره: ۶۷۱۳/ف ت
تاریخ: ۹۹/۹/۱۵
ویرایش:

باسمه تعالی

مدیریت تحصیلات تکمیلی

فرم شماره ۱۱: صورت جلسه نهایی دفاع از رساله دکتری (Ph.D)

بدینوسیله گواهی می شود آقای/خانم **سمیرا حورعلی** دانشجوی دکتری رشته **مهندسی کامپیوتر** به شماره دانشجویی **۹۵۱۶۸۹۵** ورودی مهر ماه سال ۱۳۹۵ در تاریخ ۱۳۹۹/۹/۵ از رساله نظری■ / عملی☐ خود با عنوان: **یک مدل مبتنی بر یادگیری عمیق برای تحلیل هم‌ارجاعی عبارات اسمی** دفاع و با اخذ نمره ......۱۹....... به درجه : ...عالی...... نائل گردید.

جدول تعیین درجه نمره رساله برای ورودیهای ۹۴ و ماقبل

الف) درجه عالی: نمره ۲۰-۱۹ ☒	ب) درجه خیلی خوب: نمره ۱۸/۹۹ – ۱۷☐
ج) درجه خوب: نمره۱۶/۹۹– ۱۵☐	د) مردود: کمتر از ۱۵ ☐

جدول تعیین درجه نمره رساله برای ورودیهای ۹۵ و مابعد

الف) درجه عالی: نمره ۲۰-۱۹ ☒	ب) درجه خیلی خوب: نمره ۱۸/۹۹ – ۱۸☐
ج) درجه خوب: نمره۱۷/۹۹– ۱۶☐	د) مردود: کمتر از ۱۶ ☐

ردیف	هیئت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱	**دکتر مرتضی زاهدی**	استاد/ اساتید راهنما	**استادیار**	
۲	**دکتر منصور فاتح**	مشاور/ مشاورین	**استادیار**	
۳	**دکتر بهروز مینایی**	استاد مدعو خارجی	**دانشیار**	
۴	**دکتر حسین مروی**	استاد مدعو داخلی	**دانشیار**	
۵	**دکتر مرضیه رحیمی**	استاد مدعو داخلی	**استادیار**	
	دکتر علیرضا تجری	سرپرست (نماینده) تحصیلات تکمیلی دانشکده	**استادیار**	

مدیر محترم تحصیلات تکمیلی دانشگاه:

ضمن تأیید مراتب فوق مقرر فرمائید اقدامات لازم در خصوص انجام مراحل دانش آموختگی خانم **سمیرا حورعلی** بعمل آید.

وزارت علوم، تحقیقات و فناوری - دانشکده فناوری اطلاعات و کامپیوتر

تقدیم به روح پاک پدرم، مادر عزیزم و همسر عزیزم

که همواره مایه امید و دلگرمی برای من بوده اند

تشکر و قدردانی

سپاس و ستایش مر خدای را جل و جلاله که آثار قدرت او بر چهره روز روشن، تابان است و انوار حکمت او در دل شب تار، درفشان .آفریدگاری که خویشتن را به ما شناساند و درهای علم را بر ما گشود و عمری و فرصتی عطا فرمود تا بدان، بنده ضعیف خویش را در طریق علم و معرفت بیازماید.

و استادانی را راهنمایم قرار داد تا چشمم را به زیباییهای دنیا باز نمایند

بر خود لازم می‌دانم که از زحمات اساتید محترم جناب آقای دکتر مرتضی زاهدی و جناب آقای دکتر منصور فاتح که در این سال‌ها دلسوزانه مرا در نیل به این هدف یاری کردند تقدیر و تشکر به عمل آورم.

تعهدنامه

اینجانب **سمیرا حورعلی** دانشجوی دوره دکتری رشته **مهندسی کامپیوتر** دانشکده مهندسی کامپیوتر دانشگاه صنعتی شاهرود نویسنده پایان‌نامه **یک مدل مبتنی بر یادگیری عمیق برای تحلیل هم‌ارجاعی عبارات اسمی** تحت راهنمائی **دکتر مرتضی زاهدی** متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه ، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

چکیده

تحلیل هم‌ارجاعی یکی از گام‌های مهم در پردازش متن و یکی از راهکارهای اساسی برای ارتقاء عملکرد سیستم‌های مبتنی بر متن مانند استخراج اطلاعات، خلاصه‌سازی متون، ترجمه ماشینی و پرسش و پاسخ است. هدف از تحلیل هم‌ارجاعی شناسایی نامبری‌ها یا عباراتی در متن است که به یک موجودیت اشاره می‌کنند. منظور از نامبری، ضمایر، عبارات اسمی و موجودیت‌های نامدار است. هرچند طی چهار دهه گذشته تحلیل هم‌ارجاعی همواره یک موضوع تحقیقاتی فعال بوده است، اما برای استفاده در کاربردهای درک متن و در زمینه شناسایی نامبری‌ها، شناسایی مراجع کاندید و شناسایی زنجیره‌های هم‌مرجع، دقت آن در حد قابل‌قبول نیست. یکی از دلایل دشواری این مسئله نیاز آن به منابع دانش مختلف ازجمله دانش لغوی، دانش نحوی، دانش جهانی، ساختار گفتمان و دانش معنایی است. به‌گونه‌ای که در یک متن ممکن است، هر یک از موارد هم‌ارجاعی نیازمند یک یا چند منبع دانش برای تحلیل باشد؛ همچنین، انتخاب مرجع مناسب برای نامبری‌های موجود در متن تابع قواعد مختلفی است که تطابق جنسیت، تعداد، ساختار رشته‌ای، معنایی و سایر پارامترهای هم‌ارجاعی را تضمین کند. در این رساله، با در نظر گرفتن این ویژگی‌ها در راستای افزایش دقت، بازخوانی و سایر اهداف تحلیل هم‌ارجاعی مسئله در قالب پارامترهای مؤثر در کارایی مدل شده است. برای این منظور، با به‌کارگیری جدیدترین روش‌های تعبیه واژگان، شبکه‌های عصبی بازگشتی و مکانیزم توجه، بازنمایش دنباله کلمات و ویژگی‌های آن‌ها استخراج شد. سپس با استفاده از بازنمایش منابع دانش، انتخاب دانش مناسب و اطلاعات خوشه‌ها و موجودیت‌ها، نامبری‌ها و مراجع کاندید موجود در متن با دقت بالا استخراج شدند. همچنین با استفاده از مدل رتبه-بندی چندشاخصه مبتنی بر شبکه‌های عصبی پرسپترون و کوهونن مراجع کاندید رتبه‌بندی شده و زنجیره‌های هم‌ارجاعی استخراج شدند. طبق نتایج شبیه‌سازی روی پیکره انگلیسی CoNLL-2012 shared task و پیکره پزشکی i2b2، به‌ترتیب به Avg. F1، ۸۳/۹ درصد و ۹۷/۳ درصد دست یافتیم. علاوه بر این، مقدار F1 در تشخیص موجودیت‌های نامدار روی پیکره English Gigaword به‌میزان ۷/۰۴ درصد بهبود داده شده است.

کلمات کلیدی: پردازش زبان طبیعی، تحلیل هم‌ارجاعی، شبکه عصبی بازگشتی، رتبه‌بندی چند شاخصه، منبع دانش.

لیست مقالات مستخرج از پایان نامه

1- S. Hourali, M. Zahedi, and M. Fateh, “Coreference Resolution Using Neural MCDM and Fuzzy Weighting Technique,” International Journal of Computational Intelligence Systems, Vol. 13(1), 2020, pp. 56–65.

2- S. Hourali, M. Zahedi, and M. Fateh, “A New Model for Coreference Resolution Based on Knowledge Representation and Multi-criteria Ranking,” Journal of Intelligent & Fuzzy Systems, vol. 40, 2020, pp. 1-16.

۳. سمیرا حورعلی، مرتضی زاهدی، منصور فاتح، تحلیل هم‌ارجاعی مبتنی بر شبکه‌های عصبی عمیق و خوشه-بندی فازی در متن، ارسال شده به مجله مهندسی برق و کامپیوتر جهاد دانشگاهی (تحت داوری)

۴. سمیرا حورعلی، مرتضی زاهدی، منصور فاتح، تحلیل هم‌ارجاعی با استفاده از یادگیری عمیق و تکنیک تصمیم گیری چندمعیاره، ارسال شده به مجله پردازش علائم و داده‌ها (در حال انجام اصلاحات)

فهرست مطالب

فهرست جداول

فهرست اشکال

فصل ۱ : تعریف مسئله و مفاهیم اولیه

۱-۱ مقدمه

تحلیل هم‌ارجاعی یکی از گام‌های پردازش معنایی متن است که هدف آن شناسایی مراجع ضمایر و نامبری[1]های هم‌مرجع درون متن است. استفاده از این گام پردازشی در کاربردهای مختلف درک متن از جمله سیستم‌های استخراج اطلاعات، خلاصه‌سازی، پرسش و پاسخ، تشخیص احساسات و ترجمه ماشینی ضروری است. هرچند طی چند دهه گذشته تحلیل هم‌ارجاعی همواره یک موضوع تحقیقاتی فعال بوده است اما هنوز هم دقت روش‌های موجود آن در حد قابل‌قبول برای استفاده در کاربردهای درک متن نیست. در این فصل ابتدا مسئله تحلیل هم‌ارجاعی به‌طور کامل تعریف و بررسی می‌شود سپس دلایل دشواری این مسئله و اصطلاحات تخصصی موجود در این زمینه بیان می‌شود، در انتهای فصل نیز گام‌های انجام تحلیل هم‌ارجاعی مورد بررسی قرار می‌گیرد.

۱-۲ تعریف مسئله و چالش‌های موجود

در حال حاضر با توجه به کثرت شبکه‌های اجتماعی، شبکه‌های خبری تلویزیونی، رادیویی، اینترنتی و غیره، خواندن متون مختلف و به‌تبع آن تحلیل آن‌ها و دست‌یابی به ارتباطات در این متون نیازمند صرف هزینه زمانی و انسانی بسیار بالا است. در عصر کنونی با استفاده از فن‌های مختلف پردازش زبان طبیعی صورت می‌گیرد، یکی از چالش‌های موجود در این زمینه پایین‌بودن دقت سامانه‌های تحلیل مرجع است که سبب کشف روابط ناصحیح و یا عدم کشف روابط صحیح می‌شود [۱]. معمولاً حین نوشتن یک متن، برای اشاره به یک موجودیت (یک شخص، سازمان، محل و غیره) تنها از نام آن استفاده نمی‌کنیم بلکه بسته به شرایط و به دلیل جلوگیری از تکرار و بیان اطلاعات بیشتری در مورد آن موجودیت یا تأکید بر یک ویژگی خاص، از عبارات توصیفی مختلف و گاهی ضمایر برای اشاره به آن موجودیت استفاده می‌کنیم.

[1] mention

مثلاً ممکن است برای اشاره یک شخص نام کاملش بیان شود (عادل فردوسی پور)، یا تنها نام خانوادگی (فردوسی پور) او و در موارد غیررسمی‌تر تنها از نام کوچک (عادل) استفاده می‌کنیم. حتی ممکن است شخص یا اشیا را با ویژگی‌ها و یا کاربردشان توصیف کنیم (مجری برنامه نود یا گزارشگر خوب فوتبال) و یا با استفاده از یک ضمیر (او، ش و غیره) [۲].

به چنین عباراتی که برای اشاره به یک موجودیت استفاده می‌شوند نامبری[1] گوییم. بنابراین می‌توان گفت همه نامبری‌هایی که به یک موجودیت یکسان اشاره می‌کنند با یکدیگر هم‌مرجع[2] هستند و با هر یک از دیگر نامبری‌ها که به موجودیت دیگری اشاره می‌کنند ناهم‌مرجع هستند. مثلاً عادل فردوسی پور با مجری برنامه نود هم‌مرجع است اما با علی لاریجانی ناهم‌مرجع است.

با توجه به توضیحات فوق می‌توان گفت در هر متن تعدادی از نامبری‌ها به یک موجودیت مشترک اشاره می‌کنند. بنابراین از یک دیدگاه کاربردی هدف تحلیل هم‌ارجاعی[3] یافتن همه عباراتی است که دارای مرجع یکسان در یک سند هستند.

تحلیل هم‌ارجاعی یکی از چالش‌برانگیزترین مسائل حوزه پردازش زبان طبیعی است که عبارت است از تشخیص اینکه چه عبارات اسمی[4] یا نامبری‌هایی در متن به یک موجودیت مشترک در جهان واقع اشاره دارند [۳]. یا به‌طور مشابه این مسئله عبارت است از گروه‌بندی همه نامبری‌های سند، درون دسته‌های مشابه به‌گونه‌ای که نامبری‌های درون هر دسته به یک موجودیت مشترک در بحث اشاره کنند.

انسان‌ها که کاربران اصلی زبان هستند می‌توانند به‌طور سریع و ناخودآگاه مراجع همه عبارات زبانی را تشخیص دهند و اطلاعاتی که توسط این مراجع ارائه می‌شوند را به موجودیت‌های اصلی متصل کنند. به‌هرحال فرآیندی که باعث انجام این کار در انسان‌ها می‌شود هنوز دقیقاً مشخص نیست.

[1] markables, mentions, REs (referring expressions) or CEs (coreference elements)
[2] coreferent
[3] coreference resolution
[4] noun phrase

به‌عبارت‌دیگر، تصریح دانش ضمنی که در پشت پرده این فرآیند درک انسان‌ها قرار دارد کار دشواری است و این چالشی است که مسئله تحلیل هم‌ارجاعی به پردازش زبان طبیعی وارد می‌کند.

مسئله تحلیل هم‌ارجاعی یک مسئله جالب در پردازش زبان طبیعی و یک نکته کلیدی در درک متن است و در مسائلی مثل استخراج اطلاعات، خلاصه‌سازی، پاسخ به سؤالات، و ترجمه ماشینی که در آن‌ها درک متن از اهمیت بالایی برخوردار است، کاربرد فراوان است. به‌عنوان مثال در کاربرد پاسخ به سؤالات، استفاده از تحلیل هم‌ارجاعی یکی از مراحل پیش‌پردازشی است که باعث بهبود بازخوانی می شود[۴].

در خلاصه‌سازی، تحلیل هم‌ارجاعی تأثیر زیادی در یافتن جملات پراهمیت در متن دارد و اطلاعاتی را در رابطه با موضوع اصلی بحث فراهم می‌آورد. در ترجمه ماشینی، جایگزینی ضمایر با مراجع آن‌ها بسیار مهم است.

هرچند مسئله تحلیل هم‌ارجاعی یک موضوع تحقیقاتی فعال طی چهار دهه گذشته بوده است اما هنوز راه طولانی تا حل شدن کامل در پیش دارد. لبه دانش فعلی در زمینه این مسئله دارای دو نقطه ضعف است.

اول اینکه هیچ مدل محاسباتی دقیقی برای این مسئله وجود ندارد [۵]. بهترین رویکرد فعلی موجود برای تحلیل عبارات اسمی، رویکرد یادگیری ماشین با ناظر است که در آن یک رده‌بند دودویی با استفاده از داده‌های آموزشی آموزش داده شده و تعیین می‌کند که آیا دو عبارت اسمی هم‌مرجع هستند یا نه.

دومین نقطه ضعف عدم دسترسی اکثریت سیستم‌های تحلیل هم‌ارجاعی کنونی به دانش پیچیده مورد نیاز برای تحلیل درست مرجع‌هاست، زیرا بسیاری از موارد هم‌ارجاعی تنها با کمک دانش جهانی قابل حل هستند. بطور خلاصه چالش‌های موجود در زمینه تحلیل هم‌ارجاعی را می‌توان به صورت زیر بیان کرد:

- نیاز به منابع دانش مختلف ازجمله دانش لغوی، دانش نحوی، دانش جهانی، ساختار گفتمان و دانش معنایی
- نیاز به یک یا چند منبع دانش برای تحلیل هر یک از موارد هم‌ارجاعی
- متنوع بودن دلایل هم‌ارجاعی
- تطابق جنسیت، تعداد، ساختار رشته ای، معنایی و سایر پارامترهای هم‌ارجاعی جهت انتخاب مرجع
- مناسب برای نامبری‌های موجود در متن
- دشواری فرآیند تشخیص نامبری های موجود در متن

۱-۳ دلایل دشواری مسئله تحلیل هم‌ارجاعی

علیرغم تلاش‌های زیاد طی چهار دهه اخیر روی مسئله تحلیل هم‌ارجاعی، کارایی روش‌های حل این مسئله هنوز به حد قابل‌قبولی نرسیده است به گونه‌های که این مسئله هنوز یک موضوع تحقیقاتی فعال می‌باشد. برخی دلایل دشواری این مسئله در ذیل آمده‌اند.

- **متنوع بودن دلایل هم‌ارجاعی**

بسیاری از منابع اطلاعاتی در تحلیل هم‌ارجاعی دخیل هستند. به‌بیان‌دیگر ممکن است هرکدام از چندین نامبری درون یک سند با قانون متفاوتی از دیگری به مرجع خود تحلیل شوند. برخی از این قواعد عبارت‌اند از نزدیکی فاصله، تطبیق کامل رشته‌ای، تطبیق کلمه سر، رابطه بدل، اهمیت[1] نقش فاعل، اهمیت[2] نامبری تکراری، توازی نحوی، شباهت معنایی هستند.

- **نیاز به دانش فراوان (منابع اطلاعاتی مختلف)**

با در نظر گرفتن همه انواع موارد هم‌ارجاعی متوجه می‌شویم که هر نوع از هم‌ارجاعی نیاز به دانش و اطلاعات مخصوص به خود برای حل شدن دارد. برخی از انواع اطلاعات مورد نیاز برای حل این مسئله عبارتند از:

[1] salience

- اطلاعات لغوی: تطبیق رشته‌ای، تطبیق کلمه سر
- اطلاعات نحوی: نقش نحوی، جنسیت، تعداد، دسته معنایی
- دانش جهانی: دریاچه ارومیه همان نگین فیروزه‌ای ایران است.
- ساختار گفتمان: تمرکز بحث
- **موارد دشواری از هم‌ارجاعی** [۲]

[تیم ملی ایران]۱ روز گذشته در تهران به مصاف [ژاپن]۲ رفت. [تیم میزبان]مقدم=۱ در این بازی سه گل را وارد دروازه [حریف]مقدم=۲ کرد.

[تیم ملی ایران]۱ روز گذشته در توکیو به مصاف [ژاپن]۲ رفت. [تیم میزبان]مقدم=۲ در این بازی سه گل را وارد دروازه [حریف]مقدم=۱ کرد.

[علی]۱ روی [رضا]۲ داد زد زیرا [او]مقدم=۲ ماشین را خراب کرده بود

[علی]۱ روی [رضا]۲ داد زد زیرا [او]مقدم=۱ عصبانی بود

فاطمه [کیک]۱ را از روی [میز]۲ برداشت و [آن]مقدم=۲ را شست.

فاطمه [کیک]۱ را از روی [میز]۲ برداشت و [آن]مقدم=۱ را خورد.

هیچ منبع دانشی به‌تنهایی کاملاً قابل اطمینان نیست. مثلاً مقدم نزدیک‌تر به یک نامبری با احتمال بیشتری مرجع آن نامبری است تا یک مقدم دورتر. اما مثال‌های نقض زیادی برای این قانون وجود دارد.

۱-۴ برخی اصطلاحات و مفاهیم موجود

در ادامه اصطلاحات و مفاهیمی که در حوزه تحلیل هم‌ارجاعی وجود دارد بررسی می‌شود.

- **تحلیل هم‌ارجاعی در مقابل تحلیل ضمیر**[۱]

[1] anaphora resolution

مسئله تحلیل هم‌ارجاعی ممکن است با مسئله تحلیل ضمیر اشتباه گرفته شود. هدف تحلیل ضمیر یافتن مرجع یک ضمیر وابسته[۱] است. به بیان ساده‌تر مسئله تحلیل ضمیر عبارت است از شناسایی مرجع ضمایر در متن. هر چند که مسئله تحلیل هم‌ارجاعی به مسئله تحلیل ضمیر شبیه است اما یک گام از آن فراتر رفته و نیازمند بدست آوردن مرجع همه نامبری‌های (شامل ضمایر، اسامی خاص، عبارات اسمی معرفه و عبارات اسمی نکره) در متن است، یعنی حتی نامبری‌های غیر وابسته (که برای تفسیر نیاز به عبارت دیگری در متن ندارند) را نیز شامل می‌شود.

- **موجودیت[۲] و نامبری[۳]**

یک موجودیت، شیء یا مجموعه‌ای از اشیای موجود در جهان واقع است. موجودیت‌ها به چند نمونه محدود می‌شوند که عبارتند از: شخص، سازمان، و محل. یک نامبری یک ارجاع متنی به یک موجودیت است. به بیان دیگر یک موجودیت با مجموعه‌ای از نامبری‌هایی به یک شیء ثابت دلالت دارند منطبق است.

- **نامبری وابسته[۴] و مقدم[۵]**

هر نامبری وابسته ، نماینده یک عبارت زبانی است که برای تفسیر، به مقدم‌هایش وابسته است. مثلاً در جمله زیر ضمیر او یک نامبری وابسته است:

علی برادر من است. او از من بزرگ‌تر است.

- **موجودیت تک نامبری[۶] و موجودیت چندین نامبری[۷]**

بسته به اینکه در متن به یک موجودیت تنها یک‌بار اشاره شده است یا دو یا چند بار، موجودیت‌ها را به دو دسته موجودیت تک نامبری [۱] و موجودیت چندین نامبری[۲] تقسیم می‌کنند. به موجودیت تک

[1] anaphor
[2] entity
[3] mention
[4] anaphor
[5] antecedent
[6] Singleton entity
[7] multi-mention entity

نامبری، نامبری ایزوله[3] هم گفته می‌شود. در نتیجه حل مسئله تحلیل هم‌ارجاعی موجودیت‌ها تنها برای موجودیت با چندین نامبری صدق می‌کند، زیرا برای موجودیت با نامبری تکی هیچ مورد هم-ارجاعی وجود ندارد.

- **کاپولا[4]**

کاپولا حالتی است که در آن با کمک افعال اسمیه (am, is , are) یک نامبری با نامبری دیگر مترادف و هم‌مرجع فرض می‌شود.

[The Eyjafjallaj¨okull volcano] is [one of Iceland’s largest].

- **بدل**

بدل حالتی است که در آن دو نامبری بلافاصله پس از یکدیگر می‌آیند و معمولاً بین آن‌ها کاما قرار می‌گیرد. در این حالت دو نامبری هم‌مرجع هستند. در ساختار بدل معمولاً نامبری اول یک موجودیت نامدار و دومی یک عبارت اسمی است.

[The Eyjafjallaj¨okull volcano], [one of Iceland’s largest], had been dormant for nearly two centuries.

- **آنافورای زمانی**

مواردی از هم‌ارجاعی که به زمان مربوط می‌شوند. در این موارد عباراتی مثل then, when, before/after, until, as, while, since, immediately thereafter در زبان انگلیسی استفاده می شوند.

The mail arrived this morning. I was at home then (=this morning)

---

[1] singleton entities
[2] multi-mention entities
[3] isolated mentions
[4] predicative یا capula

- **مدل زوج-نامبری[1]، مدل موجودیت-نامبری[2] و مدل رتبه‌ای[3]**

سه مدل کلی برای مسئله تحلیل هم‌ارجاعی وجود دارد. مدل اول که به مدل زوج-نامبری[4] موسوم است بر مبنای زوج نامبری‌ها کار می‌کند، یعنی دو نامبری را یا به‌عنوان هم‌مرجع یا به‌عنوان غیرهم مرجع رده‌بندی می‌کند و سپس با ترکیب همه نامبری‌های هم‌مرجع زنجیره‌های هم‌مرجع موجود در سند را شناسایی می‌کند. مدل دیگر مدل موجودیت-نامبری است که بجای رده‌بندی هر نامبری با یک نامبری قبلی، آن را با یک موجودیت قبلی رده‌بندی می‌کند، که نتیجه آن به وجود آمدن مجموعه‌ای از نامبری‌ها برای هر موجودیت است. رویکرد دوم معمولاً از روش خوشه‌بندی استفاده می کند. مدل سوم مدل رتبه‌ای است که بجای مقایسه تنها دو نامبری با یکدیگر، مجموعه‌ای از نامبری‌ها را مقایسه کرده و نتیجه مقایسه تعیین رتبه هر نامبری است.

- **سنجه مبتنی برپیونددهی[5] و سنجه مبتنی بر دسته[6]**

در راستای دو رویکرد حل مسئله تحلیل هم‌ارجاعی (رویکرد زوج-نامبری و رویکرد موجودیت-نامبری)، می‌توان این راه‌حل‌ها را با دو مقیاس متفاوت نیز ارزیابی کرد: سنجه مبتنی بر پیونددهی[7] تعداد ربط‌های (موارد هم‌ارجاعی) درست، مفقود و اشتباه را شمارش می‌کند، در حالیکه سنجه مبتنی بر کلاس[8] موجودیت‌ها را به‌عنوان خوشه در نظر گرفته و نه تنها موجودیت‌های با چندین نامبری بلکه موجودیت‌های تک نامبری را نیز در نظر می‌گیرد.

[1] pairwise/mention-pair model
[2] entity-based/entity-mention model
[3] ranking model
[4] pairwise (or mention-pair) models
[5] link-based measure
[6] class-based measure
[7] link-based performance metrics
[8] class-based performance metrics

- **نامبری‌های سیستمی[۱] و نامبری‌های استاندارد طلایی[۲]**

مجموعه نامبری‌های موجود در استاندارد طلایی که توسط افراد متخصص به وجود آمده‌اند را نامبری های درست یا نامبری‌های طلایی گویند. اما نامبری‌های سیستمی یا نامبری‌هایی که توسط یک سیستم end-to-end به وجود می‌آید لزوماً با این نامبری‌های همخوانی ندارد.

- **اشاره به جلو[۳]**

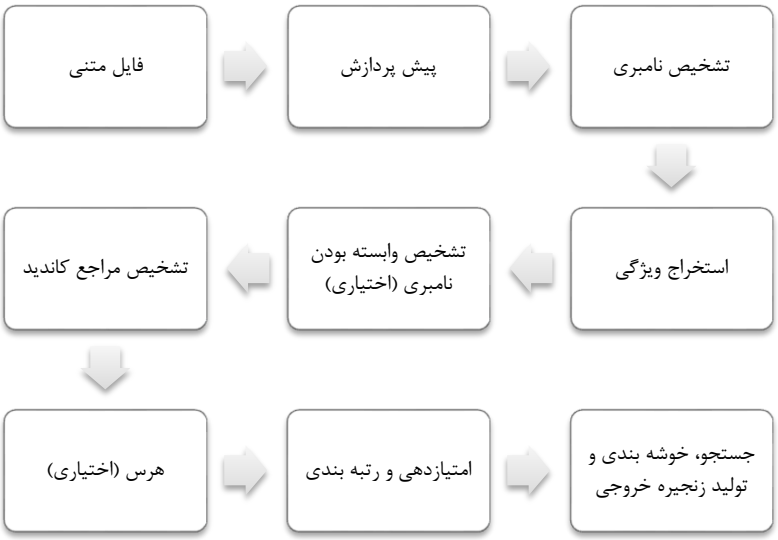
نامبری است که مرجع آن بعد از آن می‌آید. مثلاً در جمله زیر مرجع ضمیر her پس از آن آمده است.

Near [her], [Jill] saw a snake.

۱-۵ گام‌های مورد نیاز برای حل مسئله تحلیل هم‌ارجاعی

ورودی مسئله تحلیل هم‌ارجاعی یک متن خام مثلاً یک متن خبری و خروجی آن زنجیره‌های هم-ارجاعی است. علیرغم اختلاف‌های مهم، بیشتر سیستم‌های تحلیل هم‌ارجاعی موجود (مبتنی بر قواعد دست‌نویس یا مبتنی بر پیکره) را می‌توان به‌عنوان نمونه‌ای از الگوریتم‌های عمومی در نظر گرفت. این الگوریتم‌ها در ابتدا یک سند خام D را دریافت کرده و مجموعه‌ای از پیوندهای هم مرجعی L_D را برای آن محاسبه می‌کند. این پیوندها زوج‌هایی از موجودیت‌های نامدار هستند که در مرحله بعدی زنجیره های هم مرجعی را تولید می‌کنند. به‌منظور اینکه یک سیستم بتواند یک متن خام را گرفته و زنجیره-های هم‌ارجاعی آن را مشخص کند بایستی گام‌های شکل ۱-۱را انجام دهد.

[1] true/gold mentions
[2] system mentions
[3] cataphor



شکل ۰-۱. گام‌های مختلف یک سیستم تحلیل هم‌ارجاعی [۶]

۱- پیش‌پردازش

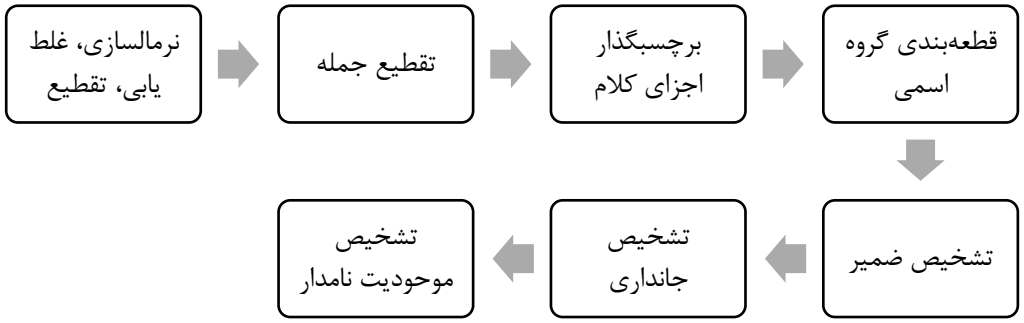
به‌منظور تولید ورودی‌های یک الگوریتم رده‌بند نامبری‌های هم‌مرجع، بایستی نامبری‌های موجود در متن شناسایی و استخراج شوند. مراحل پیش‌پردازش مورد نیاز برای این کار عبارتند از تک‌واژه سازی[۱]، ریشه‌یابی لغات[۲]، جداساز جملات[۳]، پردازش‌های ریخت‌شناسی[۴]، برچسب‌زنی اجزای کلام[۵]، شناسایی عبارات اسمی[۶]، شناسایی موجودیت‌های نامدار[۷]، استخراج عبارات اسمی تودرتو و تعیین دسته معنایی و در صورت امکان ایجاد درخت تجزیه وابستگی[۸] و درخت تجزیه سازه‌ای[۹].

برای این منظور بایستی یک سری عملیات پیش‌پردازش روی متن انجام شود. در شکل ۱-۲ پیمانه[۱۰] های پردازش زبان طبیعی زیر برای این کار پیشنهاد شد. خروجی این مراحل مرزهای خوش‌تعریف

[1] tokenization
[2] stemmer
[3] sentence segmentation
[4] morphological process
[5] pos tagging
[6] np extractor
[7] ner
[8] dependency parse tree
[9] constituent parse tree
[10] module

برای نامبری‌ها و اطلاعاتی در مورد نامبری‌هایی است که قرار است در مرحله استخراج ویژگی‌ها استفاده شوند.

سرانجام مواردی که مجموعه نامبری‌ها را تشکیل می‌دهند عبارتند از: عبارات اسمی استخراج‌شده، موجودیت‌های نامدار و عبارات اسمی تودرتو. برای موارد غیر نامدار لازم است دسته معنایی مشخص شود. مشکلات زیادی بر سر راه شناسایی نامبری‌ها وجود دارد، به‌عنوان مثال، نامبری‌ها اغلب تودرتو هستند، مرز آن‌ها با موارد موجود در استاندارد طلایی متمایز است و ممکن است مواردی از آن‌ها یافته نشده یا مواردی یافته شود که در استاندارد طلایی وجود ندارد. در یادگیری ماشین روش‌هایی وجود دارد که با به‌کارگیری آن‌ها، بازنمایشی با کیفیت برای داده بدست می‌آید، در این روش‌ها از داده‌ی خام در ورودی الگوریتم استفاده می‌گردد. به مجموعه‌ی این روش‌ها، یادگیری بازنمایش می‌گویند.



شکل ۰-۲. رویه‌های پردازش زبان طبیعی استفاده شده برای شناسایی و استخراج نامبری‌ها [۷]

همان‌طور که گفته شد، در سیستم‌های هم‌ارجاعیend-to-end هدف نهایی، پردازش متن‌های حاشیه‌نویسی نشده است. در این سیستم‌ها عموماً در فاز پیش‌پردازش بردار کلمات و بردارهای کاراکترهای موجود در متن استخراج شده، سپس با دنباله کلمات موجود در متن ورودی و نامبری‌ها با استفاده از شبکه‌های عصبی عمیق تشخیص داده می‌شوند. روش‌های جدید استخراج بردار کلمات

عبارتند از ELMo[1] [۸]، BERT [۹]، RoBERTa[2] [۱۰] و XLNet [۱۱]. از آنجا که در روش پیشنهادی از تعبیه واژگان RoBERTa استفاده شده است. این روش، روش BERT را توسعه و بهبود داده است. روش BERT و نحوه استخراج تعبیه واژگان توسط آن در زیر شرح داده می‌شود.

شبکه BERT در دو اندازه متفاوت آموزش داده شده است.BERT پایه شامل ۱۲ لایه رمزگذار[3] (که در مقاله اصلی Transformer Blocks نامیده می‌شوند) و شبکه بزرگ‌تر شامل ۲۴ لایه رمزگذار است. هر رمزگذار از یک لایه self-attention، یک لایه dense و یک لایه dropout تشکیل شده است. قبل از رمزگذارها لایه تعبیه و بعد از آن‌ها لایه‌های خروجی قرار دارد. شبکه پایه در مجموع ۱۱۰ میلیون پارامتر و شبکه بزرگ ۳۴۵ میلیون پارامتر دارد. برای آموزش BERT از دو روش زیر استفاده شده است:

الف) مدل زبانی پوششی (MLM[4])

در این روش ۱۵ درصد لغات متن با توکن [MASK] جایگزین شده و به ورودی BERT داده می‌شوند. یک لایه کلاس‌بندی به اندازه تعداد لغات با تابع فعال‌سازی softmax به خروجی رمزگذار اضافه شده است. در این روش باید لغات حذف شده توسط شبکه حدس زده شود. آموزش این شبکه دو سویه[5] است و به کلمات قبلی و بعدی حساس می‌باشد.

ب) پیش‌بینی جمله بعدی (NSP[6])

در این روش دو جمله با توکن [SEP] از هم جدا می‌شوند. وظیفه شبکه این بار این است که یاد بگیرد این دو جمله، دو جمله متوالی هستند یا خیر.

۲- استخراج نامبری‌ها

[1] embeddings from language models
[2] robustly Optimized BERT Pretraining Approach
[3] encoder
[4] Masked Language Model
[5] Bidirectional
[6] Next Sentence Prediction

دومین گام در حل مسئله تحلیل هم‌ارجاعی یافتن نامبری‌هایی است که بایستی مراجع آن‌ها مشخص شود. نوع عباراتی که به‌عنوان نامبری شناسایی می‌شوند به چندین عامل بستگی دارد از جمله: کاربرد مورد نظر، زبان متن، نوع و دامنه متن. مثلاً در یک سیستم که هدف آن تحلیل ضمیرها است، نامبری ها عبارتند از انواع ضمایر ازجمله: ضمیر شخص سوم، ضمیر ملکی، ضمیر انعکاسی و برای سیستمی که هدف آن تحلیل هم‌ارجاعی است نامبری‌ها عبارتند از موجودیت‌های نامدار، عبارات اسمی و ضمایر [۷].

در حالت کلی هر نامبری می‌تواند یکی از موارد زیر باشد:

- اسامی خاص و دیگر موجودیت‌های نامدار شامل اشخاص (دکتر حسن روحانی)، سازمان‌ها (مجلس شورای اسلامی)، و مکان‌ها (استان همدان، ساختمان چهار طبقه).
- عبارات اسمی[1]: عباراتی مثل رئیس جمهور منتخب که سرِ عبارت، یک اسم عام است.
- ضمایر مثل، آن‌ها، وی، ایشان و غیره.

۳- نمایش دانش، تابع ویژگی

در گام سوم با استفاده از منابع دانش مختلف استخراج ویژگی نامبری‌ها انجام می‌شود. روش‌های مختلف در این زمینه با یکدیگر متفاوت هستند، از این جهت که برخی فرآیند استخراج ویژگی موجودیت‌های نامدار را (با تکیه بر پیمانه‌های پیش‌پردازشی مثل برچسب‌زن اجزای کلام، استخراج گر موجودیت‌های نامدار، پارسرها) کاملاً خودکار انجام می‌دهند و برخی دیگر از اطلاعات استاندارد طلایی استفاده می‌کنند [۷].

الگوریتم‌های یادگیری ماشینی در حالت کلی دانش را با کمک بردارهای ویژگی نمایش می‌دهند. در مسئله تحلیل هم‌ارجاعی، یک تابع ویژگی، مجموعه‌ای از نامبری‌ها را با توجه به برخی سنجه‌ها ارزیابی می‌کند، مثلاً اینکه آیا یک مجموعه از نامبری‌ها از لحاظ جنسیت با یکدیگر مطابقت دارند. در

[1] common noun NPs or nominal mentions

حقیقت گام تحلیل، با استفاده از مقدار برگشت داده شده توسط مجموعه‌ای از توابع ویژگی انجام می شود. در ادامه، توابع ویژگی رایج و همچنین پیشرفت‌های انجام‌شده در زمینه به کار بردن توابع ویژگی جدید که از دانش جهانی و دانش معنایی استفاده می‌کنند، بررسی خواهد شد.

در سیستم‌های یادگیری با ناظر قدیمی، نمونه‌های یادگیری با جفت کردن نامبری‌های m_i و m_j و برچسب‌گذاری آن‌ها با برچسب‌های درست (هم‌مرجع)[1] یا نادرست (ناهم‌مرجع)[2] ایجاد می‌شدند. مثلاً نمونه آموزشی $\langle m_i, m_j, boolean \rangle$ درست است اگر و تنها اگر m_i و m_j هم‌مرجع باشند. زوج‌های $\langle m_i, m_j \rangle$ با بردارهای ویژگی که شامل ویژگی‌های تکی (اطلاعاتی در مورد یکی از نامبری‌ها، مثلاً تعداد آن) یا زوج ویژگی‌ها (اطلاعاتی در مورد ارتباط بین دو نامبری، مثلاً همخوانی در تعداد) است نمایش داده می‌شوند [۶].

۴- تشخیص وابسته بودن نامبری

در گام چهارم که اختیاری است، تشخیص وابسته بودن نامبری بررسی می‌شود که شامل جدا کردن آن بخش از نامبری‌هایی است که وابسته نیستند. این نامبری‌ها دارای هیچ مقدمی نیستند و نیازی به تحلیل ندارند. این گام برای مسئله تحلیل ضمیر انجام می‌شود و در مسئله تحلیل هم‌ارجاعی همواره انجام نمی‌شود چرا که همه نامبری‌ها وابسته هستند.

درحالی‌که سه گام قبلی به‌صورت پیش‌پردازش برای کل سند انجام می‌شوند، چهار گام بعدی به‌منظور تحلیل انجام شده و برای هر نامبری m_j موجود در A انجام می‌شود.

۵- تشخیص مراجع کاندید

گام پنجم وظیفه تولید مقدم‌های کاندید برای هر نامبری را بر عهده دارد. به‌صورت پیش‌فرض مجموعه نامبری‌های کاندید C_j نامبری‌هایی هستند که به‌صورت خطی قبل از نامبری m_j قرار دارند.

[1] positive instance
[2] negative instance

۶- هرس

گام پنجم وظیفه جداسازی مقدم‌های ناسازگار و درنتیجه کاهش تعداد مقدم‌های کاندید را بر عهده دارد. در این گام از C_j برخی مقدم‌های غیرمحتمل با توجه به مجموعه‌ای از محدودیت‌ها (مثل عدم سازگاری در جنسیت، تعداد، یا اصول قید گذاری) حذف می‌شوند.

۷- امتیازدهی و رتبه‌بندی

در گام ششم، مقدم‌های کاندیدی که در مراحل قبل استخراج شده‌اند، مرتب‌سازی و امتیازدهی می‌شوند. رتبه‌بندی بر مبنای مجموعه‌ای از قواعد یا محدودیت‌ها به دست می‌آید. در برخی رویکردها هر نامبری m_j یک امتیاز عددی را به‌صورت جداگانه دریافت می‌کند (مثلاً در رویکردهای احتمالی یا آماری) که احتمال اینکه m_i و m_j هم مرجع باشند را تعیین می‌کند. می‌توان از نمره به‌دست‌آمده برای هر نامبری برای رتبه‌بندی آن استفاده کرد. در برخی رویکردهای دیگر کاندیدها مستقیماً توسط قواعد اولویت‌دار رتبه‌بندی می‌شوند.

۸- جستجو، خوشه‌بندی و تشکیل زنجیره‌های نهایی

در گام هفتم که آخرین گام است، خوشه‌بندی نامبری‌های رتبه‌بندی شده‌ی گام قبل انجام می‌شود و زنجیره‌های هم‌ارجاعی نهایی بدست می‌آیند.

۱-۶ ساختار پایان‌نامه

ادامه رسالهبه این شکل سازمان‌دهی شده است. ابتدا در فصل ۲ به بیان مفاهیم و مراحل اساسی موضوع می‌پردازیم و برخی از یافته‌های محققان در این حیطه را طبقه‌بندی نموده و مورد بررسی قرار خواهیم داد. سپس در فصل ۳ مراحل مختلف روش پیشنهادی را توضیح خواهیم داد و در فصل ۴ به ارزیابی روش پیشنهادی خواهیم پرداخت. در فصل ۵ دستاوردهای پژوهش و پیشنهاداتی برای کارهای آتی را شرح خواهیم داد.

فصل ۲ : پیشینه پژوهش

۲-۱ مقدمه

در این فصل مروری بر کارهای قبلی مرتبط با این رساله خواهیم داشت. این فصل از دو بخش اصلی تشکیل شده است، در بخش اول، روش‌های شناسایی نامبری‌های موجود در متن بررسی می‌شود و در بخش دوم به بررسی انواع روش‌های تحلیل هم‌ارجاعی می‌پردازیم. سعی شده است پژوهش‌های گذشته مرتبط با موضوع هر دو قسمت بررسی شود و راه‌حل‌ها و روش‌هایی که تا کنون برای حل این مسئله به کار رفته است معرفی و رده‌بندی شود.

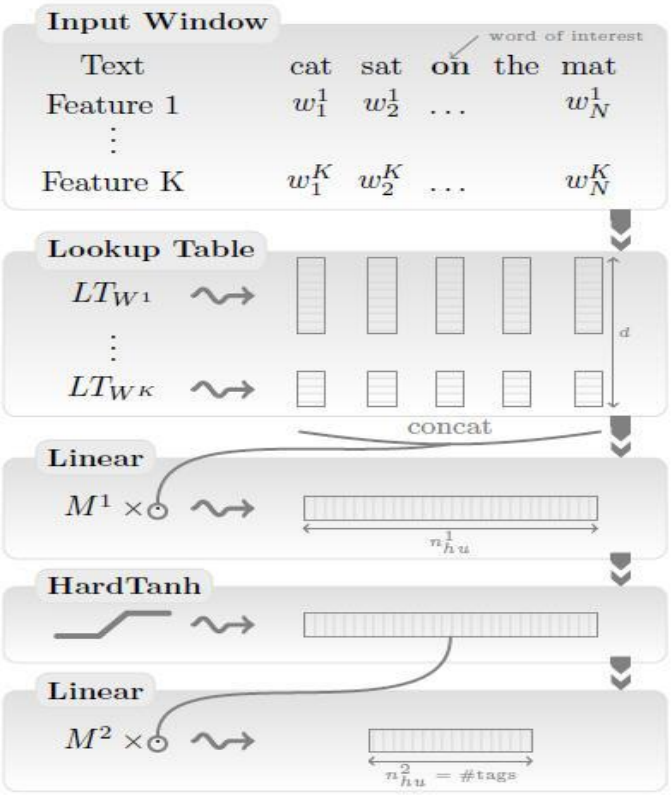
۲-۲ روش‌های تشخیص نامبری

در مقالات قدیمی، برای تشخیص موجودیت‌های نامدار در یک متن، ویژگی‌ها به‌صورت دستی، توسط اشخاص خبره و در واقع به‌صورت آزمون و خطا استخراج می‌گردید. هدف اصلی از به کار بردن شبکه-های عمیق در تشخیص موجودیت‌های نامدار، خودکار شدن استخراج ویژگی‌ها توسط شبکه و در نهایت، بهبود نتیجه‌ی تشخیص بود. این سیستم‌ها اصطلاحاً به‌صورت سرتاسر[1] آموزش می‌بینند. یعنی ورودی به‌صورت متن به شبکه وارد می‌شود و ویژگی‌ها در لایه‌های عمیق شبکه به‌صورت خودکار با استفاده از انتشار به عقب یاد گرفته می‌شوند [۱۲].

کولوبرت و همکارانش، با معرفی معماری یک شبکه‌ی عصبی و همچنین آموزش آن، راه‌حلی برای بسیاری از مسائل کاربردی پردازش زبان طبیعی همچون برچسب‌زنی ادات سخن، قطعه‌بندی[2]، تشخیص موجودیت‌های نامدار و برچسب‌زنی نقش معنایی[3] ارائه نمودند. در این معماری پنجره‌ای از کلمات در ورودی دریافت می‌شود و در چند لایه، ویژگی‌ها به‌صورت خودکار و با آموزش انتشار به عقب شبکه بدست می‌آیند. در لایه‌ی اول، استخراج ویژگی هر کلمه و در لایه‌ی دوم، استخراج ویژگی از پنجره‌ای از کلمات صورت می‌گیرد که با روش سبد واژگان تفاوت دارد و درواقع توالی واژگان در

[1] end to end
[2] chunking
[3] semantic role labeling

این بخش حائز اهمیت است، در لایه‌های بعدی برچسب کلمات تخصیص داده می‌شود [۱۳]. در شکل ۲-۱ معماری پیشنهادی کولوبرت برای پنجره‌ای از کلمات نمایش داده شده است.



شکل ۲-۱. معماری پیشنهادی کولوبرت [۱۳]

در تمامی مقالات تلاش اصلی در بدست آوردن بردار ویژگی برای کلمات است، بردارهای ویژگی مختلفی برای کلمات بدست می‌آید و سپس با پیونددهی این بردارها به یکدیگر و استفاده از شبکه‌ی عصبی، بردار بازنمایش جدیدی که حاوی ویژگی‌های سطح بالایی از زبان است، برای مجموعه‌ای از کلمات بدست می‌آید. سپس از این بردار بازنمایش جدید جهت شناسایی موجودیت‌های نامدار استفاده می‌شود. الکساندر پاسوس و همکارانش به‌جای استفاده از بردار تعبیه‌ی واژگان از بردار تعبیه-ی عبارات[1] استفاده کرده‌اند [۱۴].

[1] phrase embedding

دوسانتوس و گیمارس به‌جای اینکه صرفاً از بردار تعبیه‌ی واژه‌ها استفاده کنند، از بردار تعبیه‌ی واژه‌ها و بردار تعبیه‌ی حروف به‌طور همزمان استفاده کرده‌اند یعنی علاوه بر یادگیری بردار تعبیه‌ی واژه‌ها، با اســتفاده از یک لایــه‌ی پیچشی بردار تعبیــه‌ی حروف هم یاد گرفته می‌شــود [۱۵].

هوانگ و همکارانش بعد از شبکه‌ی عصبی بازگشتی(RNN) از یک لایه، میدان تصادفی شرطی (CRF) استفاده کردند. همچنین در این روش به‌جای استفاده از نورون‌های ساده در لایه‌های برگشتی از گیت‌های LSTM استفاده شده است.شبکه‌ی مطرح‌شده در این مقاله به‌صورت سراسری آموزش می‌بیند به این معنا که همه‌ی لایه‌ها با هم آموزش می‌بینند [۱۶].

چیو و نکولز، برای یادگیری بردار تعبیه‌ی حروف از لایه‌های پیچشی استفاده کردند. در واقع می‌توان گفت این روش همانند روش دوسانتوس و گیمارس است، با این تفاوت که در این روش، به‌جای استفاده از RNN [۱۷] ساده از شبکه‌ی عصبی LSTM [۱۸] دوجهته استفاده شده است. نام این مدل را شبکه‌ی برگشتی پیچشی (CRNN) نامیدند [۱۹].

لامپل و همکارانش دو مدل را معرفی کردند. مدلی که حاوی شبکه‌ی برگشتی دوجهته با گیت‌های LSTM به همراه یک لایه CRF بود نتیجه‌ی بهتری را حاصل کرد. در این مدل از بازنمایش واژگان با استفاده از بردار تعبیه‌ی حروف و بازنمایش واژگان به روش Word2Vec، توأمان استفاده شده است. در این مقاله بردار تعبیه‌ی حروف با استفاده از یک شبکه‌ی LSTM دو جهته بدست می‌آید [۲۰].

ما و هوی از مدلی مشابه با مدل لامپل و همکارانش استفاده کردند ولی در بدست آوردن بردار تعبیه-ی حروف به‌جای استفاده از یک شبکه‌ی LSTM دوجهته، از یک شبکه‌ی عصبی پیچشی(CNN) استفاده کرده‌اند [۲۱].

لین و همکارانش با اضافه کردن ویژگی‌های نحوی به بردار تعبیه‌ی واژگان و بردار تعبیه‌ی حروف به نتایجی بهتر، نسبت به مدل‌هایی که از ویژگی‌های نحوی کلمــات استفاده نمی‌کنند، دست یافتند. بهترین نتیجه‌ی گزارش شده مربوط به مدلی است که معماری مشابه با معماری لامپل دارد. مدل پیشنهادی بر روی متن‌های رسانه‌های اجتماعی آموزش و تست شده است [۲۲].

لی و همکارانش [۲۳] با استفاده از تعبیه واژگان ELMo و شبکه عصبی Bi-LSTM نامبری‌های موجود در متن را همزمان با تشکیل زنجیره‌های هم مرجع شناسایی کردند. تعبیه واژگان ELMo به دلیل استفاده از ساختار شبکه عصبی عمیق اطلاعات معنایی مفیدی جهت شناسایی دقیق نامبری‌ها در اختیار سیستم قرار می‌دهد.

جوشی و همکارانش [۲۴] از تعبیه واژگان BERT و شبکه عصبی Bi-LSTM نامبری‌های جهت موجود در متن استفاده کردند. آن‌ها با در نظر گرفتن مدل‌های زبانی از پیش آموزش‌دیده در دو سطح پاراگراف و سند موفق به شناسایی نامبری‌های موجود در متن با دقتی بالاتر از سایر روش‌های پیشین شدند.

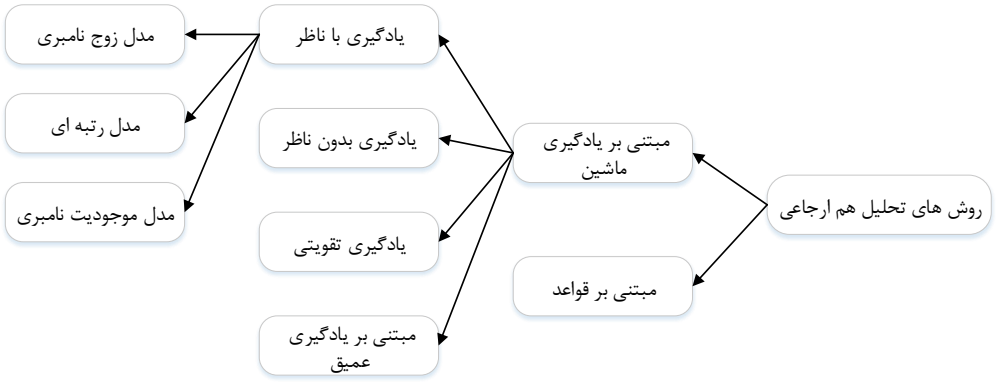
۲-۳ روش‌های تحلیل هم‌ارجاعی

در حالت کلی می‌توان روش‌های تحلیل هم‌ارجاعی را به دو دسته مبتنی بر قواعد زبانی[۱] و مبتنی بر یادگیری ماشین[۲] یا روش‌های مبتنی بر داده تقسیم کرد. در روش‌های مبتنی بر قاعده، مجموعه‌ای از قواعد زبانی که به‌صورت دست‌نویس توسط افراد خبره نوشته شده‌اند، به ترتیب اجرا می‌شوند تا موارد هم‌ارجاعی درون متن مشخص شوند. از مزایای این روش می‌توان به‌دقت بالا و سادگی طراحی اشاره کرد، اما قابلیت انعطاف این روش پایین است و لازم است برای هر زبان طبیعی مجزا، سیستم مجدداً از ابتدا توسط افراد خبره طراحی شود.

روش‌های مبتنی بر یادگیری ماشین[۳] نیز به چهار دسته با ناظر[۴]، بی‌ناظر[۵]، مبتنی بر یادگیری عمیق و مبتنی بر یادگیری تقویتی تقسیم می‌شوند. در روش‌های با ناظر، لازم است داده‌های آموزشی قبلاً توسط افراد حاشیه‌نویسی شده باشد، که این داده‌ها برای یادگیری سیستم مورد استفاده قرار می

[1] linguistics-based methods
[2] machine learning-based methods
[3] data driven approaches
[4] supervised
[5] unsupervised

گیرند. از طرف دیگر، در روش‌های بی‌ناظر نیازی به داده‌های آموزشی نیست (یا داده‌ها آموزشی بسیار کمی مورد نیاز است) اما دقت این روش هنوز برای حل مسئله تحلیل هم‌ارجاعی پایین است. شکل ۲-۲ رده‌بندی روش‌های مختلف را نمایش می‌دهد. در ادامه توضیحات دقیق‌تر هر کدام از روش‌ها به همراه نمونه‌هایی از آن‌ها آورده می‌شود.



شکل ۲-۲. روش‌های مختلف تحلیل هم‌ارجاعی

۲-۳-۱ رویکردهای مبتنی بر قواعد

به سیستم‌هایی که از قواعد و مکاشفات زبانی برای حل مسئله تحلیل هم‌ارجاعی استفاده می‌کنند سیستم‌های مبتنی بر قواعد گفته می‌شود. در این رویکردها، برای هر نامبری وابسته[1]، مجموعه‌ای از مقدم‌های کاندید انتخاب شده و سپس با استفاده از محدودیت‌ها و تقدم‌ها، محتمل‌ترین مقدم برای هر نامبری از این مجموعه انتخاب می‌شود.

روش‌هابز [۲۵] یکی از قدیمی‌ترین روش‌های مبتنی بر قواعد برای حل مسئله تحلیل ضمیر است. مبنای کار این الگوریتم عمدتاً درخت تجزیه جمله ورودی است. به بیان دیگر این الگوریتم از محدودیت‌های نحوی برای تحلیل ضمیر به مرجع درست آن بهره می‌برد.

[1] anaphor

تئوری مرکزیت[1] [۲۶] به‌منظور مدل کردن تمرکز بحث در یک گفتمان[2] و پیوستگی گفته‌ها[3] در یک گفتمان مطرح شد. یک گفتمان می‌تواند هر نوع متن از جمله مقالات خبری، متون ادبی، دیالوگ‌ها و امثالهم باشد. از جمله اهداف اصلی تئوری مرکزیت ردگیری موجودیت اصلی در یک گفتمان است. این دانش نیز می‌تواند در روش‌های مبتنی بر قواعد تحلیل هم‌ارجاعی بکار رود.

همچنین می‌توان از مجموعه‌ای از روش‌ها و مکاشفه‌ها برای حل مسئله تحلیل هم‌ارجاعی بهره برد. شاید مهم‌ترین کاری که از این روش استفاده کرده است، سیستمی است که برای تحلیل مراجع زبان انگلیسی توسط [۲۷] طراحی شده است. این سیستم از تعدادی غربال تشکیل شده است که از دقیق‌ترین به کم‌دقت‌ترین مرتب شده‌اند. پس از انتخاب مقدم‌های کاندید برای هر نامبری، این غربال ها به ترتیب روی مقدم‌ها اعمال می‌شوند تا مرجع هر نامبری تعیین شود.

۲-۳-۲ رویکردهای مبتنی بر یادگیری ماشین

روش‌های داده‌گرا یا روش‌های یادگیری ماشین قابلیت بسیار خوبی در کشف الگوها و تعمیم‌هایی دارند که به‌راحتی توسط چشم انسان قابل شناسایی نیستند. همچنین از میانه دهه ۹۰ میلادی روش های یادگیری ماشین رویکرد غالب به کار رفته برای حل مسئله تحلیل هم‌ارجاعی بوده است. روش های داده‌گرا نیاز به داده‌هایی دارند که با اطلاعات حاشیه‌نویسی شده هم‌ارجاعی، ارتباطات بین عبارات اسمی را مشخص کرده باشند. پروژه‌هایی که تأثیر زیادی در حل مسئله تحلیل هم‌ارجاعی داشته‌اند، سری‌های ششم و هفتم MUC[4] و ACE[5] بوده است. این ابتکارات موجب ایجاد ابزارهایی برای حاشیه‌نویسی داده، بهبود روش‌های حاشیه‌نویسی و ایجاد داده‌هایی برای توسعه سیستم‌های بعدی و همچنین بهبود روش‌های ارزیابی سیستم‌های تحلیل هم‌ارجاعی شد.

[1] centering theory
[2] discourse
[3] utterances
[4] message understanding conference
[5] automatic content extraction

می‌توان فرایند تحلیل هم‌ارجاعی به روش یادگیری ماشین را به دو زیر فرایند تقسیم کرد: رده‌بندی[1] و پیوندددهی[2]. رده‌بندی، فرایند بررسی وجود یا عدم وجود سازگاری بین المان‌هاست، که المان‌ها می توانند نامبری‌ها یا موجودیت‌هایی جزئی (گروهی از نامبری‌هایی که در مراحل قبل هم‌مرجع تشخیص داده شده‌اند) باشند. فرایند پیوندددهی، از اطلاعات مرحله رده‌بندی استفاده می‌کند تا المان‌ها را درون گروه‌هایی قرار دهد که موجودیت‌هایی نهایی را ایجاد می‌کنند.

سیستم‌های تحلیل مرجع موجود را با توجه به نحوه استفاده از فرایندهای رده‌بندی و پیوندددهی می توان به سه نوع زیر تقسیم کرد:

۱- رویکردهای جستجوی پسگرد

برای رده‌بندی نامبری‌ها با موارد قبلی، فضای پشت سر را به‌منظور یافتن بهترین مقدم جستجو می کنند. در این حالت، فرایند پیوندددهی معمولاً مکاشفه‌ای است که زوج-نامبری‌ها را به‌عنوان نمونه‌های مثبت رده‌بندی می‌کند (single-link).

۲- رویکرد دو-مرحله‌ای

فرایندهای رده‌بندی و پیوندددهی را به‌صورت پشت سر هم اجرا می‌کند. در مرحله اول یعنی رده‌بندی، همه زوج نامبری‌ها بررسی شده تا نمونه‌های مثبت (و مقدار اطمینان هر کدام) مشخص شود و در مرحله دوم، یعنی مرحله پیوندددهی، از الگوریتم‌هایی مثل بخش‌بندی گراف، خوشه‌بندی و برنامه‌نویسی صحیح خطی[3] استفاده می‌شود تا با کمک خروجی مرحله قبل نتیجه نهایی، یعنی مجموعه نامبری‌های مربوط به هر موجودیت بدست آید.

۳- رویکرد تک-مرحله‌ای

[1] classification
[2] linking
integer linear programming (ILP) [3]

مستقیماً مرحله پیونددهی را اجرا می‌کند، درحالی‌که رده‌بندی به‌صورت برخط اجرا می‌شود. به بیان دیگر عملیات رده‌بندی و شناسایی موجودیت‌ها به‌صورت همزمان با هم و در یک مرحله انجام می-شود. در این راستا، می‌توان از مدل‌های نامبری-موجودیت استفاده کرد.

یک نکته ضعف رویکردهای جستجوی پسگرد آنست که همه نامبری‌ها را به عنوان وابسته در نظر می‌گیرند[1]. تنها در حالتی که هیچ مقدمی برای یک نامبری یافت نشود، آن نامبری تکی در نظر شده یا اولین المان یک خوشه جدید در نظر گرفته می‌شود. هر چه تعداد کاندیدهای یک نامبری وابسته بیشتر باشد، احتمال اینکه این نامبری به اشتباه وابسته فرض شود بیشتر است. برای حل این مشکل حداقل دو راه‌حل پیشنهاد شده است.

ساده ترین راه‌حل، محـدود کردن گستره جستجوی پســگرد است. بسیاری از سیستم‌ها، جستجو را به دو یا سه جمله محدود می‌کنند. هرچند بسیاری از مراجع معمولا در همان جمله یا جمله قبلی نامبری فعال قرار دارند، اما یقینا محدود کردن گستره جستجو باعث کاهش فراخوانی خواهد شد.

راه‌حل دیگر، زیر سیستمی است که وابسته بودن یا نبودن یک نامبری را تشخیص می‌دهد. بنابراین، اگر یک نامبری بعنوان غیر وابسته تشخیص داده شود، هیچ جستجویی برای یافتن مقدم‌های آن انجام نخواهد شد. دنیس و بالدریج [۲۸] و بوگل و فرنک [۲۹] یک رده‌بند دودویی را آموزش داده‌اند که تعیین می‌کند آیا یک نامبری تکی وابسته هست یا نه و به عنوان یک فیلتر قبل از اینکه رده‌بندی هم‌ارجاعی اعمال شود مورد استفاده قرار می‌گیرد. به این نوع فیلتر وابستگی معمولا تشخیص دهنده گفتمان جدید هم گفته می‌شود که در دیگر کارها نیز مطالعه شده است [۳۰] [۳۱]. بهبود دقت سیستم، بدون تاثیر منفی روی بازخوانی، یک هدف مهم برای هر سیستم است. بهرحال این موضوع برای ساختار خط لوله در جستجوی پسگرد مطابق انتظار عمل نمی‌کند. لو [۳۲] سعی کرده است دلیل کم شدن بازخوانی را حین اعمال خط لوله فیلتر و رده‌بند به جستجوی پسگرد را تشخیص دهد.

[1] نامبری وابسته نامبری است که برای تفسیر، به مقدم هایش وابسته است

دلیل این مشکل آن است که کارایی کل سیستم به شدت وابسته به دقت فیلتر است. از طرف دیگر تشخیص اشتباه یک مورد مثبت[۱] باعث الحاق نادرست یک زنجیره هم‌ارجاعی به زنجیره دیگر خواهد شد و همچنین تشخیص اشتباه یک مورد منفی[۲] باعث می‌شود که یک نامبری وابسته به عنوان شروع یک زنجیره جدید در نظر گرفته شود و زنجیره درست هم‌ارجاعی قطع شود. در نتیجه یک بهبود در دقت فیلتر وابستگی باعث بهبود قابل توجهی در نتایج خواهد شد.

لو [۳۲] مشابه با کار قبلی خود [۳۳]، مدل دوقلویی معرفی شده که از مفهوم وابستگی نامبری استفاده می‌کند. این رویکرد از دو مدل تشکیل شده است. مدل اول خاصیت هم‌مرجع بودن نامبری وابسته با هر کدام از موجودیت‌های جزیی که تا کنون ایجاد شده‌اند را بررسی می‌کند. اگر نامبری وابسته با هرکدام از این موجودیت‌ها هم‌مرجع تشخیص داده شود، به صورت احتمالی به آن موجودیت متصل خواهد شد. در غیر اینصورت، مدل دومی وجود دارد که وابستگی نامبری را بررسی می‌کند که در آن مدل، احتمال ایجاد یک موجودیت جدید توسط نامبری را بررسی می‌کند (در صورت تشخیص عدم وابستگی نامبری). برای این منظور مدل، نامبری را با همه موجودیت‌هایی که تا کنون ایجاد شده‌اند به صورت یکجا مقایسه می‌کند و برای این کار از مجموعه ویژگی‌های متفاوتی استفاده می‌کند.

در روشی که برای تعیین وابستگی نامبری بر مبنای برش گراف پیشنهاد شده است [۳۴]، سیستم از اطلاعاتی که توسط رده‌بند هم‌مرجع فراهم می‌شود، استفاده می‌کند و بطور همزمان از فرایند تحلیل مرجع نیز استفاده می‌کند. هر نامبری توسط یک گره گراف نمایش داده می‌شود و دو یال، گره را به یک گروه وابسته و یک گروه غیر وابسته متصل می‌کنند. وزن یال‌ها با احتمالاتی که توسط فیلتر وابستگی نامبری بدست آمده است (و با همکاری احتمالات هم‌ارجاعی) مشخص می‌شود. گراف با توجه به وزن یال‌ها، برش داده می‌شود تا محتمل ترین بخش‌بندی پیدا شود. بخش‌بندی نهایی

---

[۱] false positive

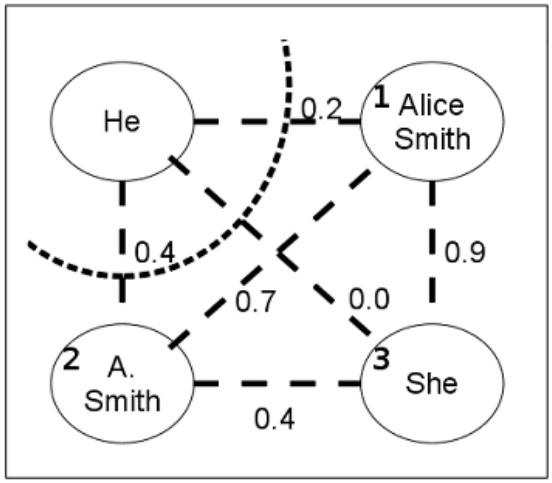
[۲] false negative

مشخص می‌کند که کدام نامبری‌ها وابسته هستند که این اطلاعات برای بهبود کارایی سیستم تحلیل مرجع مفید خواهند بود.

این مدل بهینه‌سازی سراسری همچنین می‌تواند به منظور به کار بستن دیگر انوع دانش سراسری مورد استفاده قرار گیرد. مثلاً، دنیس و بالدریج [۲۸] اطلاعاتی در رابطه با موجودیت‌های نامدار وابسته بودن[۱] را به منظور فراهم کردن سرنخ‌های بیشتر برای تعیین شروع زنجیره هم‌ارجاعی بکار بستند.

بخش‌بندی گراف، نوع دیگری از فرایند پیونددهی است. معمولاً این کار روی یک گراف بدون جهت انجام می‌شود که در آن راس‌ها نامبری‌ها، و یال‌های وزن دار روابط هم‌ارجاعی هستند که وزن آن‌ها بیانگر احتمال هم‌ارجاعی است. معمولاً وزن یال‌ها به صورت فاصله یا هزینه تصور می‌شود و الگوریتم یال‌هایی که مقدار آنها از یک مقدار آستانه r بیشتر است را به منظور جداکردن گروه‌هایی که موجودیت‌های مجزا را نشان می‌دهند قطع می‌کند. مقدار آستانه r یا هر پارامتر دیگری را می‌توان توسط داده‌های آموزشی فراگرفت. شکل ۲-۳ که از مدل ساپنا و همکاران [۶] اقتباس شده است یک مثال از نحوه نمایش نامبری‌ها و ارتباطاتشان توسط گراف را نشان می‌دهد. در این مثال یک الگوریتم تحلیل تصمیم به قطع یال‌های نامبری He و در نتیجه ایجاد زنجیره‌ای حاوی نامبری‌های A. Smith، Alice Smith، و She است. بدین ترتیب [۳۵] یک الگوریتم تجزیه گراف را روی یک گراف وزن دار بدون جهت اعمال کردند.

[۱] anaphoricity



شکل ۲-۳. مثالی از افراز گراف که بر مبنای وزن یال‌ها و خروجی احتمالی رده‌بند [۶]

نیکول و نیکول الگوریتم BestCut را برای بخش‌بندی گراف ارائه کردند [۳۶]. در این گراف، راس‌ها بیانگر نامبری‌ها بوده و وزن یال‌ها توسط یک الگوریتم رده‌بندی دودویی آنتروپی بیشینه محاسبه می شود که احتمال هم‌ارجاعی زوج نامبری‌ها را نشان می‌دهد. الگوریتم BestCut یال‌های با حداقل وزن برش را قطع می‌کند. برای این منظور، وزن برش یک گراف به دو زیرگراف برابر مجموع یال‌هایی بود که برش آنها را قطع می‌کرد. این فرایند تا زمانی ادامه پیدا می‌کند که یک شرط توقف، که توسط الگوریتم رده‌بندی فراگرفته می‌شود، برقرار شود. در این کار برای نمایش نامبری‌های موجود در سند، بجای یک گراف از مجموعه‌ای از گراف‌ها استفاده می‌شود. مثلاً موجودیت‌های نامدار و عبارات اسمی توسط نوعشان (شخص، سازمان، مکان، تسهیلات، و GPE) از یکدیگر جدا می‌شود. ضمایر، تا پایان عملیات تقسیم گراف به فرایند تحلیل اضافه نمی‌شود.

رویکرد مشابهی اما بر پایه خوشه‌بندی، برای بخش‌بندی گراف توسط کلنر و آیلود [۳۷] معرفی شد. همه زوج‌هایی که توسط الگوریتم رده‌بندی، هم‌مرجع تشخیص داده شده می‌شوند به الگوریتم خوشه بندی تحویل داده می‌شوند و از احتمال هم‌ارجاعی آن‌ها برای محاسبه هزینه‌ها استفاده می‌شود. الگوریتم خوشه‌بندی، زنجیره‌های هم‌ارجاعی را با حداقل هزینه تعیین می‌کند. در این فرایند، همه

نامبری‌ها بدون جداسازی آن‌ها توسط کلاس معنایی لحاظ شده‌اند که این باعث ساده شدن فرایند می‌شود.

کاردی و واگستاف نیز یک رویکرد خوشه‌بندی را برای مساله تحلیل مرجع معرفی کرده‌اند که از نوع یادگیری بی‌ناظر است [۳۸]. ترتیب تحلیل برعکس ترتیب یافته شدن نامبری‌ها درون سند است و تصمیمات به صورت حریصانه اتخاذ می‌شود. برای هر نامبری و با توجه به نامبری‌های قبلی آن یک فاصله محاسبه می‌شود. این فاصله به صورت حاصل جمع وزن دار مجموعه‌ای از ویژگی‌ها تعریف می شود. اگر فاصله از یک مقدار آستانه کمتر باشد، آن نامبری به خوشه اضافه خواهد شد. وزن توابع ویژگی به صورت دستی انتخاب می‌شود و مقدار آستانه بگونه‌ای انتخاب می‌شود که مقدار F1 را برای داده‌های توسعه بیشینه کند.

روش معرفی شده توسط ان‌جی یک مدل خوشه‌بندی مبتنی بر بیشینه‌سازی مورد انتظار[۱] (EM) بدون ناظر است، هرچند در آزمایشات به صورت باناظرِ ضعیف استفاده شده است [۳۹]. این روش، مجموعه‌ای از توابع ویژگی سازگاری هر زوج نامبری را ارزیابی کرده و خوشه‌بندی یک ماتریس مربعی بولی است که در آن $C_{i,j}$ پیونددهی زوج نامبری m_i و m_j را نشان می‌دهد. یعنی اگر m_i و m_j هم مرجع باشند $C_{i,j}$ برابر یک و در غیر اینصورت برابر صفر است.

طبق نتایج منتشر شده، نتایج بدست آمده توسط سیستم قابل قبول است اما هنوز با مقادیر بدست آمده توسط سیستم‌های باناظر قابل مقایسه نیست. بطور خلاصه می‌توان گفت، رویکردهای دو مرحله ای نتایج منسجم‌تری را در مقایسه با سیستم‌های جستجوی پسگرد تولید می‌کنند. دلیل اصلی این موضوع آنست که فرایند پیونددهی با اعمال تراگذاری می‌تواند از بروز تناقضات جلوگیری کند و بطور هم‌زمان با مقایسه هر چه بیشتر احتمالات هم‌ارجاعی با یکدیگر نتایج بهینه‌تری را تولید کند. بهرحال مشکل فقدان اطلاعات بافتاری در ارزیابی جداگانه زوج نامبری‌ها هنوز پابرجاست.

---

[۱] expectation maximization

- **روش‌های مبتنی بر یادگیری با ناظر**

روش‌های با ناظر معمولاً در دو مرحله انجام می‌شوند که در مرحله اول مجموعه‌ای از مقدم‌های کاندید برای هر نامبری وابسته انتخاب می‌شود و برای تعیین هم‌مرجع بودن هر یک از این زوج‌ها، از یک رده‌بند[1] استفاده می‌شود. در مرحله بعد زوج نامبری-مقدم‌هایی که به‌عنوان هم‌مرجع رده-بندی‌شده‌اند به یکدیگر متصل شده تا زنجیره‌های هم‌ارجاعی به وجود آید. رده‌بندی بر مبنای بردارهای ویژگی انجام می‌شود که برای هر زوج ساخته شده است. به چنین رویکردهایی با ناظر گفته می‌شود، زیرا برای آموزش مدل رده‌بندی زوج‌های نامبری-مرجع نیاز به داده‌های حاشیه‌نویسی شده وجود دارد.

هر چند که مسئله تحلیل هم‌ارجاعی، فرآیند تعیین تعلق هر عبارت اسمی به یک زنجیره از موارد هم مرجع است، اما به‌منظور اعمال یادگیری ماشین به این فرآیند، این مسئله به‌صورت یک مسئله رده-بندی تعریف می‌شود که ورودی آن زوج‌هایی از نامبری‌ها و مراجع کاندید است و خروجی آن برچسب های هم‌مرجع (و احتمالاً درجه اطمینان هم‌ارجاعی) و ناهم‌مرجع به این زوج‌هاست.

روش‌های پیش‌پردازشی که به‌منظور شناسایی و انتخاب زوج‌های نامبری-مقدم و ساخت مجموعه ویژگی‌هایی که آن‌ها را توصیف می‌کند، از ساده تا پیشرفته با یکدیگر متفاوت است، اما بیشتر این رویکردها، حاوی مراحل پیش‌پردازشی تک‌واژه سازی، برچسب‌زنی اجزای کلام، تشخیص عبارات اسمی و جداسازی عبارات اسمی هستند. هر کدام از این مراحل حاوی خطاهایی است که به مراحل بعدی گسترش می‌یابد و مثل هر پردازش سطح بالای زبان طبیعی، این نوع زنجیره خطا در تحلیل هم‌ارجاعی نیز اجتناب‌ناپذیر است.

پس از پیش‌پردازش‌های اولیه، می‌توان اطلاعات بیشتری را با توجه به ویژگی‌های زبان مورد استفاده و منابع موجود اضافه کرد، مثل WordNet [۴۰]، و نمونه‌های آن در دیگر زبان‌ها (فارس‌نت) [۴۱]

[1] classifier

به‌منظور اضافه کردن اطلاعات معنایی. هدف از انجام عملیات پیش‌پردازش استخراج مجموعه‌ای از نامبری‌ها و مجموعه‌ای از مراجع برای آن نامبری‌هاست. می‌توان این گام‌های پیش‌پردازش را به‌عنوان بخشی از مسئله تحلیل هم‌ارجاعی در نظر گرفت یا به‌عنوان اولین گام از یک فرآیند دو مرحله‌ای که مرحله اول آن شناسایی موجودیت‌ها و مرحله دوم شناسایی موارد هم‌ارجاعی بین این موجودیت‌ها است. با تقسیم مسئله به دو زیر مسئله فوق، محققان می‌توانند بدون درگیر شدن با مسئله اول (شناسایی نامبری‌ها) تمرکز خود را روی حل مسئله هم‌ارجاعی متمرکز کنند. مراحل لازم برای ایجاد رده‌بند برای روش یادگیری با ناظر به‌صورت زیر است:

۱. شناسایی و انتخاب مجموعه‌ای از مقدم‌های کاندید: معمولاً همه عبارات اسمی قبلی که از نظر فاصله مکانی درون یک آستانه خاص قرار دارند (معمولاً سه جمله قبل) برای هر نامبری وابسته انتخاب می‌شوند. داده‌های آموزشی با جفت کردن هر نامبری با مقدم‌های کاندید قبل آن ایجاد می‌شود. این روش معمولاً منجر به یک رده‌بندی نامتوازن می‌شود که بیشتر نمونه ها، نمونه‌های منفی هستند و نمونه‌های مثبت کمی درون این داده‌ها وجود دارد. به‌منظور متوازن کردن این داده‌ها، برخی روش‌ها برای انتخاب زیرمجموعه‌ای از نمونه‌ها استفاده می شود.

۲. ایجاد مجموعه‌ای از ویژگی‌ها: هر ویژگی می‌تواند توصیفی باشد، مثلاً توصیف یک نامبری یا مرجع کاندید با برشمردن خواصی مثل دسته کلمات و جنسیت آن، یا مقایسه‌ای باشد مثلاً با مقایسه نامبری با مقدم آن توسط برچسب‌هایی از قبیل سازگار بودن آن‌ها از لحاظ جنسیت یا تعداد و غیره.

۳. انتخاب یک الگوریتم یادگیری ماشین: الگوریتم‌های زیادی برای این کار تا کنون استفاده شده‌اند مثل درخت تصمیم‌گیری [۷]. سه مدل معروف رده‌بندی یعنی، زوج-نامبری، رتبه‌ای و موجودیت-نامبری بررسی خواهند شد و مزایا و معایب هر کدام بیان می‌شود.

الف) مدل زوج-نامبری

مبنای کار این مدل که به مدل محلی نیز موسوم است، یک رده‌بند دوتایی است که مشخص می‌کند آیا دو نامبری هم‌مرجع هستند یا نه. برای انجام این کار، یک بردار ویژگی برای هر زوج نامبری بر طبق ویژگی‌های بیان‌شده در فصل قبل ایجاد می‌شود.

به بیان دقیق‌تر، رده‌بند دو بردار ویژگی را دریافت کرده و تعیین می‌کند زوج-نامبری‌های متناظر آن‌ها هم‌مرجع هستند یا نیستند. در بسیاری از موارد، رده‌بند علاوه بر تعیین هم‌مرجع بودن، درجه اطمینانی را نیز برای این رابطه مشخص می‌کند. سپس و در مرحله بعدی، از این اطلاعات برای فرآیند پیونددهی استفاده می‌شود. مدل زوج-نامبری دارای دو نقطه ضعف است که عبارتند از فقدان اطلاعات و وجود تناقضات در رده‌بندی.

ب) مدل رتبه‌ای

مدل‌های رتبه‌ای سعی دارند بر مشکل فقدان اطلاعات موجود در مدل زوج-نامبری غلبه کنند. برای این منظور بجای بررسی مستقیم اینکه، آیا نامبری‌های m_i و m_j هم‌مرجع هستند یا نه، با جستجو در میان گروهی از نامبری‌ها، بهترین کاندید را برای هم‌ارجاعی با نامبری مورد نظر می‌یابند. در این حالت، خروجی رده‌بندی کننده، برچسب‌های هم‌مرجع و ناهم‌مرجع نیست بلکه رتبه‌های یک و دو است. فرآیند پیونددهی نیز با استفاده از این اطلاعات از بروز خطا و تناقض جلوگیری می‌کند. در مدل زوج نامبری، دو نامبری با یکدیگر مقایسه می‌شوند، اما در مدل رتبه‌ای تعداد نامبری‌های مقایسه شونده بیشتر است. مدل‌های رتبه‌ای ارتباط تنگاتنگی با رویکرد جستجوی پسگرد دارند. هر دوی آن‌ها بهترین کاندید ممکن را برای هم‌مرجع شدن با نامبری وابسته برمی‌گزینند.

ج) مدل موجودیت-نامبری

این مدل به مدل سراسری نیز موسوم است. در مدل موجودیت-نامبری، مهم‌ترین تغییر این است که بجای نامبری مفهوم موجودیت در مقایسه‌ها بکار می‌رود. در مدل موجودیت-نامبری، مقایسه بین یک

نامبری و یک موجودیت جزئی، یا بین دو موجودیت جزئی با یکدیگر انجام می‌شود و نتیجه مقایسه هم‌مرجع بودن یا نبودن آن‌ها است. در برخی از مدل‌ها، یک موجودیت جزئی ممکن است دارای ویژگی‌های خاصی برای خود باشد. به دلیل اطلاعاتی که موجودیت‌های جزئی به رده‌بند می‌دهند، مدل‌های موجودیت-نامبری در بیشتر مواقع بر مشکلات کمبود اطلاعات و تناقضات فائق می‌آیند.

یکی از مزایای این مدل، حل مشکل کافی نبودن قابلیت بیان مدل زوج-نامبری است. در دو مدل قبلی، رده‌بندی بر مبنای نامبری‌ها انجام می‌شد. تنها تفاوت این بود که در مدل زوج-نامبری رویکرد های بی‌ناظر نتیجه نهایی فرآیند تحلیل هم‌ارجاعی، ایجاد زنجیره‌های مجزایی از نامبری‌های هم‌مرجع است. با توجه به این موضوع، ایده استفاده از یک الگوریتم خوشه‌بندی که همه نامبری‌های هم‌مرجع را بدون استفاده از ویژگی‌های جفتی و رده‌بند، درون یک خوشه قرار دهد منطقی و مناسب به نظر می‌رسد.

رویکردهایی که پیش‌تر بررسی شد را با توجه به نیاز آن‌ها به داده‌های حاشیه‌نویسی شده برای روابط هم‌ارجاعی، رویکردهای با ناظر گوییم. اما ایجاد و بدست آوردن داده‌های حاشیه‌نویسی شده پرهزینه است و برای برخی زبان‌ها چنین داده‌هایی وجود ندارد. بنابراین با استفاده از رویکردهای بی‌ناظر (یا نیمه ناظر) می‌توان بر مشکل دستیابی به داده‌های حاشیه‌نویسی شده غلبه کرد.

- **روش‌های مبتنی بر یادگیری بی‌ناظر**

مبنای کار رویکردهای بی‌ناظر مسئله تحلیل هم‌ارجاعی، آن است که از آنجا که روابط هم‌ارجاعی دارای خواص متقارن، تعدی، و بازتابی است و هر زنجیره از عبارات اسمی هم‌مرجع یک دسته یکسان را می‌سازد، می‌توان مسئله تحلیل هم‌ارجاعی را به‌عنوان یک مسئله خوشه‌بندی در نظر گرفت [۴۲].

در رویکرد بی‌ناظر، فرآیند یافتن موارد هم‌مرجع تبدیل به عملیات خوشه‌بندی می‌شود که در آن عبارات اسمی موجود در اسناد با توجه به شباهت بین بردارهای ویژگی که عبارات اسمی را توصیف کند، به زنجیره‌هایی تقسیم می‌شوند. الگوریتم ارائه شده توسط [۴۲] بر مبنای این ابتکار است که همه عبارات اسمی که به یک موجودیت موجود در بحث اشاره دارند، بایستی به‌نوعی مرتبط یا مشابه

باشند و فاصله معنایی آن‌ها کم باشد؛ با وجود یک توصیف (مثلاً مجموعه‌ای از ویژگی‌ها) برای هر موجودیت اسمی و روشی برای اندازه‌گیری فاصله معنایی بین دو عبارت اسمی، عبارات اسمی هم مرجع بایستی توسط یک الگوریتم خوشه‌بندی درون یک گروه قرار گیرند.

مزایای چنین رویکردی عبارتند از:

- با توجه به بی‌ناظر بودن روش خوشه‌بندی، نیازی به داده‌های حاشیه‌نویسی شده آموزشی نیست
- این روش مستقل از دامنه است.
- این مدل امکان ترکیب محدودیت‌های محلی و سراسری را فراهم می‌آورد.
- محدودیت محلی: این نوع محدودیت تنها بین یک زوج نامبری نزدیک به هم رخ می‌دهد.
- محدودیت سراسری: این نوع محدودیت نه‌تنها موارد نزدیک، بلکه دیگر نامبری‌های موجود در بحث را هم در نظر می‌گیرد. پون و دمینگوس [۴۳] یک الگوریتم یادگیری بی‌ناظر مبتنی بر منطق مارکف معرفی کردند. این سیستم تعداد خوشه‌ها را از قبل مشخص کرده است در حالیکه سیستم مولد [۴۴] این‌چنین نیست. این سیستم با اصلاح الگوریتم EM به‌گونه‌ای که لازم نباشد تعداد خوشه‌ها از قبل مشخص شوند و با تعریف مجدد گام E-step که با کمک یک درخت Bell محتمل‌ترین n ناحیه هم‌ارجاع را بدست می‌آورد. سیستم قدیمی کاردیه و واگستف [۴۲] نیز بین سیستم‌های با ناظر و بی‌ناظر رده‌بندی می‌شود. این سیستم با اعمال الگوریتم خوشه‌بندی روی بردار ویژگی‌هایی که نامبری‌ها را نمایش می‌دهد، قصد دارد برای هر موجودیت یک خوشه ایجاد کند. اما به دلیل استفاده از وزن‌های ثابتی که توسط مکاشفه‌هایی به‌عنوان سنجه تعیین فاصله برای مقایسه بین خوشه‌ها به کار می‌روند، این سیستم جز سیستم‌های بی‌ناظر رده‌بندی نمی‌شود.
- می‌توان گفت در کارهای قبلی ایده نو یا روی الگوریتم و منطق استنتاج سیستم تحلیل هم-ارجاعی داده شده است یا روی ویژگی‌ها و دانش بکار رفته در سیستم. به نظر می‌رسد با

توجه به حجم انبوه کارهایی که روی الگوریتم‌ها و روش‌های تحلیل هم‌ارجاعی انجام‌شده و اینکه این مسئله در بسیاری از موارد حتی با بهترین الگوریتم موجود هم نیازمند دانش برای انجام تحلیل مورد نظر است، بهتر است تمرکز را بجای الگوریتم جدید روی اضافه کردن دانش جدید به این مسئله داشته باشیم. با پیشرفت‌های اخیر در معانی لغات و واژگان و توسعه پایگاه‌های دانش مقیاس بزرگ، محققین شروع به استفاده از دانش جهانی در مسئله تحلیل هم‌ارجاعی کرده‌اند [۴۵].

- **رویکردهای مبتنی بر یادگیری عمیق**

روش‌های که اخیراً در حیطه تحلیل هم‌ارجاعی ارائه شده‌اند، از شبکه‌های عمیق استفاده شده می-کنند. یکی از مزیت‌های اصلی به‌کارگیری شبکه‌های عمیق استفاده از متن خام در ورودی مدل است، بدین صورت که ویژگی‌های مفید، در داخل خود شبکه و بدون دخالت انسان از متن استخراج می‌گردد. شبکه‌های بازگشتی ساده‌ترین نوع شبکه برای استخراج اطلاعات از داده‌هایی هستند که ماهیت دنباله-ای دارند، این شبکه‌ها با اینکه وابستگی‌های موجود در دنباله‌هایی از واژگان در یک متن را استخراج می‌کنند ولی قابلیت استخراج وابستگی‌های طولانی را ندارند، درحالی‌که شبکه‌های LSTM به دلیل وجود واحد حافظه در خود، وابستگی‌های طولانی‌تری را از دنباله‌ی واژگان یک متن استخراج می‌کند. شبکه‌های عمیق بر اساس نوع معماری، به انواع مختلفی تقسیم می‌شوند برای مثالRNN, CNN LSTM, GAN REZNET, DBN و غیره.

شبکه‌های CNN در مسائلی که در زمینه‌ی تصویر و استخراج بردار کاراکترها[1] مطرح می‌شوند، عملکرد خوبی از خود نشان داده‌اند. در مسائلی که داده‌ها ماهیت ترتیبی دارند همانند مسائلی که بر روی متن و گفتار و سری‌های زمانی مطرح می‌شود، شبکه‌های RNN به‌خوبی عمل می‌کنند.

[1] character embedding

اولین مدل غیرخطی رتبه‌بندی نامبری برای تشخیص هم‌ارجاعی و یادگیری خودکار ویژگی‌ها بر اساس یادگیری عمیق در مرجع [۴۶] ارائه شده است. این رویکرد دو مسئله عمده را در تحلیل هم-ارجاعی بررسی کرده است. ۱) شناسایی موارد غیرهم‌مرجع موجود در متن، ۲) استخراج ویژگی‌های پیچیده در مدل‌های خطی برای ایجاد یک مرز شفاف بین نامبری‌های هم‌مرجع و غیر هم‌مرجع. این مدل با معرفی یک مدل شبکه عصبی جدید، تنها ویژگی‌های خام نامرتبط را به‌عنوان ورودی دریافت می‌کند. استخراج نامبری‌ها توسط سیستم تحلیل هم‌ارجاعی برکلی[1] انجام می‌شود. تابع امتیاز[2] شبکه عصبی برای مدل رتبه‌بندی نامبری به‌صورت زیر رابطه (۲-۱) شده است.

$$s(x,y) \triangleq \begin{cases} u^T g\left(\begin{bmatrix} h_a(x) \\ h_p(x,y) \end{bmatrix}\right) + u_0 & \text{if } y \neq \epsilon \\ u^T h_a(x) + v_0 & \text{if } y = \epsilon \end{cases} \quad (۱-۲)$$

$$h_a(x) \triangleq \tanh(W_a \theta_a(x) + b_a) \quad (۲-۲)$$

$$h_p(x,y) \triangleq \tanh(W_p \theta_p(x,y) + b_p) \quad (۳-۲)$$

در روابط (۲-۱)، (۲-۲) و (۲-۳) h_a و h_p به‌ترتیب توابع غیرخطی تعریف‌شده روی ویژگی نامبری (θ_a) و ویژگی جفت نامبری (θ_p) هستند. g تابع خطی g_1 یا تابع غیرخطی g_2 (tanh) و C'(x) خوشه‌ای که نامبری x به آن تعلق دارد. اگر x هم مرجعی نداشته باشد برابر ϵ است. y_n^l بالاترین امتیاز مرجع نامبری در خوشه C'(x) است و اگر x هم مرجعی نداشته باشد، مقدار آن برابر ϵ است. شبکه عصبی برای کاهش تغییرات متغیر نهفته توسط تابع هزینه[3] رابطه (۲-۴) آموزش می‌بینند.

$$L(\theta) = \sum_{n=1}^{N} \max_{\hat{y} \in \gamma(x_n)} \Delta(x_n, \hat{y})(1 + s(x_n, \hat{y}) - s(x_n, y_n^l)) \quad (۴-۲)$$

در رابطه (۲-۴)، Δ خطای تابع هزینه‌ی تعریف‌شده و $W, u, v, W_a, W_p, b_a, b_p$ مجموعه پارامترهایی هستند که باید بهینه شوند. Δ بر اساس انواع خطاهای ممکن در مسئله هم‌ارجاعی مقادیر متفاوتی

[1] berkeley coreference resolution
[2] scoring function
[3] loss function

دارد. رتبه‌بندی مراجع نامبری‌ها با فیلتر شدن نامبری‌های غیر هم‌مرجع صورت می‌گیرد. وایزمن و همکاران [۴۷] اولین مدل هم‌ارجاعی غیرخطی را ارائه دادند که ثابت کرد، مسئله هم‌ارجاعی می‌تواند، مدل‌سازی ویژگی‌های کلیدی روی خوشه‌های موجودیت مدل رتبه‌بندی نامبری مبتنی بر شبکه عصبی را تکمیل کرده و با ترکیب اطلاعات سطح موجودیت و شبکه عصبی بازگشتی (RNN) روی خوشه مراجع کاندید حل شود. از طریق وزن‌های محاسبه‌شده درRNN ها، نامبری‌ها به‌صورت دنباله-ای به هر خوشه تخصیص داده می‌شوند. ایده استفاده از شبکه‌های بازگشتی این بود که تصمیمات قبلی نیز در طول فرایند بررسی سازگاری نامبری- مرجع کاندید حفظ شود.

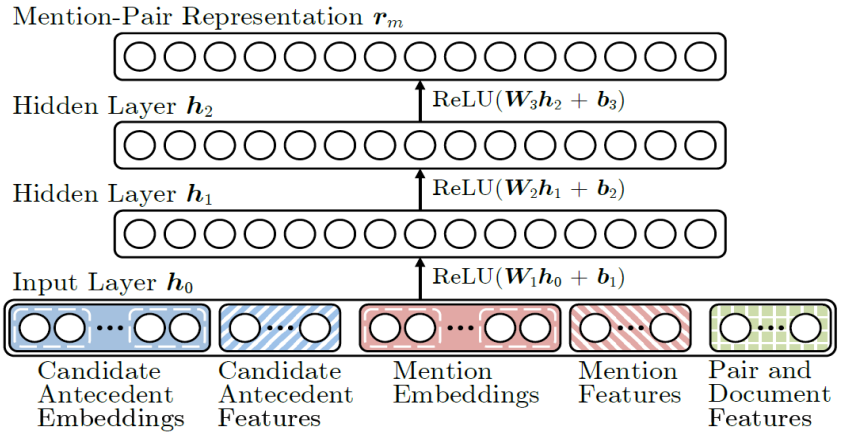
کلارک و منینگ [۴۸] تابع امتیاز[۱] مدل رتبه‌بندی مراجع را با اضافه کردن امتیازدهی جهانی[۲] تغییر دادند. ایده این روش بر مبنای ترکیب اطلاعات در سطح موجودیت و ویژگی‌های تعریف‌شده روی خوشه‌های زوج نامبری است. معماری این شبکه عصبی از چهار قسمت اصلی تشکیل شده است:

۱. رمزگذار زوج نامبری[۳] برای نمایش توزیع‌شده نامبری‌ها، ویژگی‌ها و تعبیه نامبری‌ها را از یک شبکه عصبی FFNN عبور می‌دهد (شکل ۲-۴).
۲. رمزگذار زوج خوشه[۴] که برای نمایش توزیع‌شده زوج خوشه‌ها، زوج نامبری را با هم ترکیب می‌کند (شکل ۲-۵).
۳. مدل رتبه‌بندی نامبری[۵] برای مقداردهی اولیه به وزن‌ها و بدست آوردن امتیازاتی که بعداً در رتبه‌بندی خوشه‌ها استفاده می‌شود.
۴. پیمانه رتبه‌بندی خوشه‌ها برای امتیازدهی به زوج خوشه‌ها با عبور آن‌ها از یک شبکه عصبی یک لایه‌ای.

[1] scoring function
[2] global scoring
[3] mention pair encoder
[4] cluster pair encoder
[5] mention ranking model

ویژگی‌های استفاده شده در مدل داخلی (رمزگذار زوج نامبری) عبارتند از: میانگین بردار کلمات هر نامبری، فاصله بین نامبری‌ها، تعبیه کلمه سر[1]، وابستگی‌ها، بردار اولین کلمه، آخرین کلمه، دو کلمه ماقبل، نوع نامبری، مکان نامبری، طول نامبری، تطابق رشته‌ها و غیره . این ویژگی‌ها در قالب یک بردار ورودی به یک شبکه FFNN که شامل سه لایه مخفی خطی تمام متصل داده می‌شوند. خروجی آخرین لایه نمایش برداری زوج نامبری‌ها است.

رمزگذار جفت خوشه، دو خوشه از نامبری‌ها $c_i = m_1^i, m_2^i, \dots, m_{|c_i|}^i$ و $c_j = m_1^j, m_2^j, \dots, m_{|c_j|}^j$ را گرفته و یک نمایش توزیع‌شده[2] $r_c(c_i + c_j) \in R^{2d}$ را تولید می‌کند. این ماتریس با استفاده از مقادیر ماکزیمم و میانگین، زوج نامبری‌ها را با هم ترکیب می‌کند. سپس مدل زوج نامبری (رتبه‌بندی نامبری) توسط خروجی رمزگذار زوج نامبری که شامل وزن‌های از پیش تعیین‌شده برای رتبه‌بندی خوشه‌ها و اتخاذ تصمیم هم‌مرجعی است، آموزش می‌بینند.



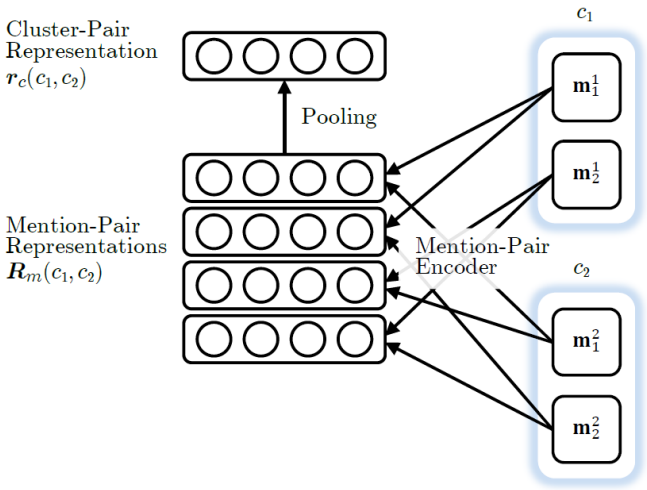
شکل ۲-۴. رمزگذار زوج نامبری[۴۸]

این مدل رتبه‌بندی توسط تابع هزینه آموزش می‌بینند. آخرین مرحله رتبه‌بندی خوشه‌ها است که از وزن‌های از پیش آموزش‌دیده‌ی مدل رتبه‌بندی نامبری‌ها برای بدست آوردن امتیاز برای تزریق خوشه‌های رمزگذار خوشه به یک شبکه عصبی یک لایه‌ی تمام متصل استفاده می‌کند. دو عمل بر

[1] head word embedding
[2] distributed representation

اساس امتیاز انجام می‌شود: ترکیب دو خوشه، و رد (هیچ عملی). در طول استنتاج بالاترین امتیاز (بیشترین احتمال) عمل در مرحله انتخاب می‌شود. این ایده جدید رتبه‌بندی خوشه‌ها، به F1 = ۶۵/۳۹ روی داده‌های انگلیسی CoNLL و ۶۳/۶۶ روی داده‌های چینی دست یافت.

کلارک و منینگ [۴۹] با تغییر تابع هزینه ابتکاری[1] ارائه‌شده در مقاله قبلی که آموزش پیچیده‌ای داشت، به سیاست یادگیری تقویتی مبتنی بر گرادیان و تغییر مقیاس پاداش‌دهی مرز بیشینه[2] ، الگوریتم قبلی را بهبود بخشید. این روش از اهمیت بالای عمل[3]های مستقل استفاده می‌کند. مستقل بودن عمل‌ها نشان می‌دهد که تأثیر هر عمل روی نتیجه نهایی متفاوت است، بنابراین این شرایط، کاندیدای مناسبی برای یادگیری تقویتی است. این مدل از مدل رتبه‌بندی نامبری مبتنی بر شبکه عصبی که در بالا راجع به آن توضیح داده شد، استفاده کرده که در آن تابع هزینه با یک تابع هزینه مبتنی بر یادگیری تقویتی جایگزین شده است.



شکل ۲-۵. رمزگذار زوج خوشه [۴۸]

[1] heuristic
[2] reward-rescale max-margin objective
[3] action

جدول ۲-۱. رده‌بندی سیستم‌های تحلیل مرجع مختلف

پژوهش	رده-بندی	سال	الگوریتم	یادگیری	پیکره استفاده شده
[۵۱]		۱۹۹۵	C4.5		Annotated news article
[۵۲]		۱۹۹۵	C4.5		MUC-5 English Joint Venture (EJV) corpus
[۷]		۲۰۰۱	C4.5		MUC-6 training documents
[۵۳]		۲۰۰۶	Maximum Entropy		MUC-6 andMUC-7 + ACE 2003 Data
[۵۴]		۲۰۰۶	SVM		ACE-2 V1.0 corpus
[۳۰]		۲۰۰۲	C4.5		MUC-6andMUC-7datasets + WordNet
[۵۵]	زوج نامبری	۲۰۰۵	C4.5+SVM		ACE coreference corpus
[۵۶]		۲۰۰۷	C4.5		unannotated BLLIP corpus+ ACE data sets
[۵۷]		۲۰۰۵	maxent classification		the ACE 2004 relation ontology +ACE 2002
[۵۸]		۲۰۰۸	Perceptron	با ناظر	ACE training data
[۵۹]		۲۰۰۹	Averaged Perceptron		MUC and ACE corpora
[۳۴]		۲۰۰۹	MaxEnt		ACE Phase II coreference corpus
[۶۰]		۲۰۱۱	maximum entropy classifier		Wikipedia and Yago + ACE-02 corpus
[۶۱]	رتبه ای	۲۰۰۳	C5.0		MUC-6 and MUC-7 data set
[۶۲]		۲۰۰۸			ACEcorpus
[۶۳]	موجودیت نامبری	۲۰۰۸	ILP		ACE-2003 corpus
[۴۵]		۲۰۱۱	SVM		YAGO, FrameNet, annotated+ unannotated data
[۳۳]		۲۰۰۴	Bell tree+ MaximumEntropy		MUC6 data
[۶۴]		۲۰۰۷	MaxEnt		ACE 2005 data

پژوهش	رده-بندی	سال	الگوریتم		پیکره استفاده شده
[۶۵]		۲۰۰۸	TiMBL		TuBa-D/Z coreference corpus: German newspaper Treebank
[۳۶]	زوج نامبری	۲۰۰۶	maximum entropy		ACE and MUC6 corpora
[۶۶]		۲۰۰۷	ILP		ACE datasets
[۶۷]		۲۰۰۷	ILP		MUC-6,MUC-7andACE data
[۶۸]		۲۰۰۸	ILP		MUC-6 and ACE data set
[۶۹]	زوج نامبری	۲۰۰۴	Dempster-Shafer	بی‌ناظر	Raw text as KB
[۴۲]		۱۹۹۹	clustering algorithm	نظارت ضعیف	Simple features
[۴۴]		۲۰۰۸	EM clustering		Simple features
[۷۰]	موجودیت نامبری	۲۰۰۷	First-Order Logic Model	باناظر	ACE 2004 data
[۷۱]		۲۰۰۵	SVM		MUC-6
[۷۲]	موجودیت نامبری	۲۰۱۰	spectral clustering		MUC6 and MUC7 data+ ACE 2004
[۷۳]		۲۰۰۴	C5.0		Handcrafted annotated coreference corpus
[۳۵]		۲۰۰۵	conditional random fields		MUC-6 and ACE data set
[۷۴]		۲۰۱۰	generative graphical model	نظارت ضعیف	ACE 2005 corpora
[۷۵]		۲۰۰۸	MLN		ACE-2004
[۷۶]		۲۰۱۲	linear regression + EM	باناظر	EventCorefBank (ECB) corpus
[۷۷]	رتبه ای	۲۰۱۳	SVM		Wikipedia + ACE 2004
[۷۸]	زوج نامبری	۲۰۱۲	C4.5		ACE 2004 and ACE 2005+ Webn-gram
[۷۹]		۲۰۱۵	ILP		+ACE-2004 OntoNotes-5.0 dataset
[۸۰]	موجودیت	۲۰۱۵	ILP	بی ناظر	Gigaword data

پژوهش	نامبری	سال	الگوریتم		پیکره استفاده شده
[۲۹]		۲۰۱۳	MLN	باناظر	CoNLL 2012 Shared Task data
[۸۱]	رتبه‌ای	۲۰۱۷	FFNN + CNN + Neural Attention Marginal Log-likelihood	عمیق	CoNLL 2012 Shared Task data
[۴۹]		۲۰۱۶	Heuristic Max-margin objective, REINFORCE policy gradient + Reward Rescaling Algorithm	تقویتی	CoNLL 2012 Shared Task data
[۸۲]		۲۰۱۸	coarse-to-fine approach + Heuristic antecedent pruning	عمیق	CoNLL 2012 Shared Task data
[۸۳]		۲۰۱۹	maximum entropy + regularization gradient model	تقویتی	CoNLL 2012 Shared Task data
[۸۴]		۲۰۱۹	Paragraph level and document Level resolution+ higher-order entity-level representations	عمیق	CoNLL 2012 Shared GAP Task data +
[۱]	زوج نامبری	۲۰۲۰	هرس گراف و بازنمایش مبتنی بر شبکه عصبی جفت نامبری	عمیق	پیکره اپسالا [۸۵]
[۳]	زوج نامبری	۲۰۲۰	قطعه‌بند مبتنی بر تجزیه‌گر وابستگی، انتخاب نمونه‌های مثبت و منفی	باناظر	پیکره اپسالا

یادگیری تقویتی با گرفتن بازخورد از مجموعه‌های متفاوت از عمل‌ها و ارتباطات انجام‌شده روی مدل رتبه‌بندی نامبری مورد استفاده قرار گرفته است. مدل ارائه‌شده بر اساس ماکزیمم پاداش، این عمل‌ها را بهینه می‌کند (الگوریتم تغییر مقیاس پاداش‌دهی). الگوریتم تغییر مقیاس پاداش‌دهی به امتیاز F1

به ترتیب ۶۵/۷۳ و ۶۳/۴ تحت معیارهای $CEAF\phi_4$ و B^3 روی پایگاه داده انگلیسی CONLL2012 دست یافت. چالش اصلی این روش در رابطه با نحوه تخصیص پاداش و جریمه است.

مدل جدیدی که با حداقل تعداد ویژگی‌ها توانست نسبت به مدل‌های دیگر بهتر عمل کند مدل مبتنی بر شبکه عصبی مربوط به لی و همکارانش [۵۰] است. این مدل اسناد را به‌صورت بردار کلمات با ابعاد بالا[1] در نظر می‌گیرد. بردار کلمات از طریق الحاق Glove، Turian و بردار کاراکترها بدست می‌آید. بردار کاراکترها با استفاده از شبکه عصبی پیچشی (CNN) در سطح کاراکتر با سه سایز پنجره مختلف بدست می‌آید. جملات برداری اسناد برای یادگیری نمایش کلمات، به یک شبکه عمیق وارد می‌شوند. سپس تمام نامبری‌های ممکن در سند متنی تحت یک بردار یک بعدی استخراج می‌شوند. این نمایش نامبری ترکیبی از بردار شروع کلمات[2]، بردار پایان کلمات[3]، بردار کلمات ابتدا[4] و ویژگی-های دیگر در سطح نامبری[5] است.

ایده این روش، هرس حریصانه کاندیدها برای آموزش و ارزیابی است و دنباله کلمات را تا ماکزیمم سایز ۱۰ در نظر می‌گیرد. نامبری‌ها با استفاده از FFNN امتیازدهی می‌شوند و فقط بخشی از دنباله کلمات که امتیاز بالایی دارند برای تشخیص هم‌مرجعی در نظر گرفته می‌شوند. ۲۵۰ نامبری قبل به‌عنوان مراجع کاندید در نظر گرفته می‌شوند و امتیاز زوج نامبری-مرجع[6] از طریق شبکه عصبی محاسبه می‌شود. تابع هزینه استفاده شده در مدل، لگاریتم احتمال مرزی[7] مراجع صحیح بر اساس خوشه‌بندی طلایی[8] است.

[1] high-dimensional
[2] start word embedding
[3] end word embedding
[4] head word embedding
[5] mention-level
[6] mention-antecedent pair
[7] marginal log-likelihood
[8] gold-clustering

در طول استنتاج مرجع دارای بهترین امتیاز به‌عنوان محتمل‌ترین مرجع انتخاب شده و زنجیره هم-مرجعی با استفاده از ویژگی انتقال‌پذیری[1] شکل می‌گیرد. نویسندگان آزمایش را تحت پنج مدل با ارزیابی‌های مختلف گزارش کردند و در هر مدل دنباله کلمات با در نظر گرفتن میانگین امتیاز نامبری هرس شده است. ایده پیشنهادی بر اساس بازخوانی، صحت و F1 روی معیارهای MUC، B^3 و CEAF بررسی شده است. نویسندگان همچنین تحلیل کمی و کیفی مدل را برای تفسیر بهتر ارائه داده‌اند. جنبه چالش برانگیز این روش زمان محاسبات بالا و تعداد زیاد پارامترهای قابل آموزش است. این مدل از یک شبکه عصبی عمیق بسیار بزرگ استفاده می‌کند، از این رو نگهداری آن بسیار سخت است.

مدل‌های هم‌ارجاعی مبتنی بر یادگیری عمیق کلمات را با استفاده از بردارها نمایش می‌دهند و روابط معنایی بین کلمات را شرح می‌دهند. این مدل‌ها، از ویژگی‌های کمتری نسبت به مدل‌های یادگیری ماشین استفاده می‌کنند. این سیستم‌ها به‌طور ضمنی وابستگی بین نامبری‌ها را با استفاده از شبکه-های RNN و LSTM در نظر می‌گیرند. یکی از معایب این سیستم‌ها این است که فهم آن‌ها سخت است و اغلب قبل از استفاده، نیاز به تعدیل خاص دسته[2] یا دامنه دارند. در الگوریتم‌های مبتنی بر یادگیری عمیق وابستگی به ویژگی‌ها در طول زمان کاهش می‌یابد، زیرا بردار کلمات از پیش آموزش‌دیده[3] و شباهت معنایی بین کلمات را تا حدودی در نظر می‌گیرند. برخلاف سیستم تحلیل هم‌ارجاعی استنفورد [۴۸] سیستم جدید تحلیل هم‌ارجاعی [۵۰] کمترین ویژگی‌های سطح نامبری و زوج نامبری-مرجع را استفاده می‌کند. این سیستم همچنین از هیچ‌یک از ابزارهای استخراج نامبری استفاده نمی‌کند و به‌طور حریصانه تمام دنباله‌های ممکن از کلمات را تا یک طول ثابت استخراج می-کند. مزیت دیگر این سیستم این است که برخلاف سایر مدل‌های عمیق از هیچ تابع هزینه ابتکاری استفاده نمی‌کند، بلکه از تابع هزینه ساده لگاریتم احتمال استفاده می‌کند. مدل‌های قبلی از توابع

[1] property of transitivity
[2] genre
[3] pre-trained word embeddings

هزینه ابتکاری استفاده می‌کردند که وابسته به یک تابع هزینه اشتباه بود که مقادیر آن بعد از آزمایش‌های زیاد مشخص می‌شد. اگرچه نگهداری این سیستم عمدتاً به علت ابعاد بزرگ آن سخت است، اما شواهد زیادی در رابطه با قابلیت اطمینان LSTMها و توانایی آن‌ها در جذب وابستگی‌های طولانی‌مدت وجود دارد. یکی دیگر از ضعف‌های احتمالی این مدل این است که رتبه‌بندی نامبری و انتخاب بالاترین امتیاز مرجع نامبری بدون استفاده از اطلاعات سطح خوشه انجام می‌شود. پیشرفت‌های آینده در زمینه سیستم تحلیل هم‌ارجاعی عمیق، با استفاده از تعریف ویژگی‌های بهتر برای یادگیری یا با ارائه معماری‌های بهتر ممکن است. در جدول ۲-۱ روش‌های مختلف تحلیل مرجع بررسی شده اند.

- **روش‌های مبتنی بر یادگیری تقویتی**

در روش‌های مبتنی بر یادگیری تقویتی [۴۹] [۸۳] عموماً، نامبری‌ها به‌عنوان حالت‌ها[1] و پیوندهای بین نامبری‌ها به‌عنوان عمل‌ها[2] در نظر گرفته شده‌اند؛ سپس طبق سیاست یادگیری تقویتی و تعریف توابع پاداش[3] و خطا[4] پیوندهای هم‌ارجاعی تشخیص داده شده است. در مرجع [۴۹] مجموعه آموزشی[5] در نظر گرفته شده شامل n واژه به‌صورت $m_1, m_2, \ldots, m_n$ است. $C(m_i)$ مجموعه مراجع کاندید برای هم مرجع بودن با واژه m_i و $T(m_i)$ مجموعه واژه‌هایی است، که با m_i هم مرجع هستند. برای عمل انجام شده توسط عامل یعنی پیونددهی واژه m_i به کلمه کاندید شده $c \in C(m_i)$، خطا[6] به‌صورت رابطه (۲-۵) تعریف شده است.

[1] States
[2] Actions
[3] Reward
[4] Punishment
[5] Training
[6] Error

$$\Delta_h(c,m_i)=\begin{cases}\alpha_{FN} & if \quad c=NA \ \wedge \ T(m_i)\neq\{NA\}\\ \alpha_{FA} & if \quad c\neq NA \ \wedge \ T(m_i)=\{NA\}\\ \alpha_{WL} & if \quad c\neq NA \ \wedge \ c\notin\{NA\}\\ 0 & if \quad c=NA \ \wedge \ T(m_i)=\{NA\}\end{cases} \qquad (۵-۲)$$

به‌عنوان مثال، اگر کلمه کاندید شده هم‌مرجعی در متن نداشته باشد ولی مجموعه هم‌مرجع‌های درست واژه m_i خالی نباشد و عامل به اشتباه واژه‌های c و m_i را هم‌مرجع تشخیص دهد، آنگاه خطای α_{FN} را دریافت می‌کند.

برای بهبود روش پیشنهادی هر عمل a_i را به یک عمل دیگر مثل $a'_i \in A$ تغییر داده شده تا مشخص شود که مدل به‌ازای عمل جدید چه پاداشی دریافت می‌کند $R(a_{1,\dots,}\ a_{i+1},\ a'_{i,}\ a'_{i+1},$ $\dots, a_T)$ پاداش نهایی دریافت شده برای عمل به‌صورت زیر محاسبه می‌شود که در آن $a_{1:T}$ دنباله‌ای از عمل‌ها است که طبق پارامترهای جاری مدل بالاترین امتیاز را دریافت کرده‌اند.

$$\Delta_r(c,m_i)=-R(a_1,\dots,(c,m_i),\dots,a_T)+\max_{a_i\in A_i}R(a_1,\dots,a_{i-1},a_i^{'},a_{i+1},\dots,a_T) \qquad (۶-۲)$$

مقدار Q[1] برای هر عمل a_i در حالت s بر اساس روابط (۲-۷) و (۲-۸) محاسبه می‌شود.

$$Q(s,a)=\sum_i W_i X_i(s,a) \qquad (۷-۲)$$

$$P_\theta(a)\ \alpha\, e^{s(c,m)} \qquad (۸-۲)$$

به این ترتیب الگوریتم تقویتی طبق رابطه (۲-۹) به بیشترین پاداش مورد انتظار می‌رسد.

$$J(\theta)=E_{[a_{1T}\square p_\theta]R(a_{1T})} \qquad (۹-۲)$$

از آنجا که باید تمامی دنباله‌های ممکن از عمل‌ها در نظر گرفته شود، محاسبه گرادیان زمان‌بر است و به‌صورت نمایی تحت این دنباله رشد می‌کند. بنابراین طبق روابط (۲-۱۰) و (۲-۱۱) از دنباله $a_{1:T}$

[1] Q-Value

نمونه‌گیری[1] کرده و گرادیان آن را محاسبه شده است. پارامتر b_i برای کاهش پراکندگی[2] استفاده می‌شود [۱۸].

$$\nabla_\theta j(\theta) = \sum_{i=1}^{T} \sum_{a_i' \in A_i} [\nabla_\theta P_\theta(a_i')](R(a_1,...,a_i',...,a_T) - b_i) \quad (۱۰-۲)$$

$$E_{a_i' \in A_i P_\theta} R(a_1,...,a_i',...,a_T) \quad (۱۱-۲)$$

۲-۴ نتیجه‌گیری

مسئله تحلیل مرجع یک فیلد تحقیقاتی فعال طی چند دهه گذشته بوده است که راه‌حل‌های ارائه‌شده برای آن هنوز پاسخگوی نیازهای پردازشی موجود نیست. حل کامل این مسئله نیازمند چند گام مختلف است از جمله استخراج نامبری‌ها، استخراج ویژگی‌های نامبری‌ها، بکار بستن یک الگوریتم به‌منظور استفاده از ویژگی‌ها برای شناسایی موارد هم‌مرجع و پیونددهی موارد هم‌مرجع به‌منظور تولید زنجیره‌های هم‌ارجاعی نهایی. در مرحله تولید ویژگی‌ها ممکن است یک گام اضافی نیز با هدف فراهم آوردن دانش مورد نیاز برای پردازش برداشته شود. همان‌طور که حجم بالایی از کارهای انجام‌شده پیشین در گذشته‌ی نزدیک نشان می‌دهد حل مسئله تحلیل مرجع مستلزم به کار بستن دانش از حوزه‌های مختلف است. در این فصل برخی کارهایی که با این منظور انجام شده بودند بررسی شد. در حالت کلی روش‌های تحلیل هم‌ارجاعی به دو دسته مبتنی بر قاعده و مبتنی بر یادگیری ماشین تقسیم می‌شوند. در روش‌های مبتنی بر قاعده موارد هم‌مرجع درون متن توسط مجموعه‌ای از قواعد دست‌نویس که توسط افراد خبره نوشته شده‌اند، استخراج می‌شوند. از مزایای این روش می‌توان به دقت بالا و سادگی طراحی اشاره کرد. اما قابلیت انعطاف این روش پایین است و باید برای هر زبان طبیعی، سیستم از ابتدا توسط افراد خبره طراحی شود.

[1] sampling
[2] variance

روش‌های مبتنی بر یادگیری ماشین نیز به سه دسته روش‌های مبتنی بر یادگیری تقویتی، آماری و مبتنی بر یادگیری عمیق تقسیم می‌شوند. در روش‌های آماری با ناظر به‌صورت دستی یا اتوماتیک داده‌های آموزشی برچسب‌گذاری شده‌اند. این داده‌ها برای یادگیری سیستم مورد استفاده قرار می-گیرند. از طرف دیگر، در روش‌های بدون ناظر نیازی به داده‌های آموزشی نیست (یا داده‌ها آموزشی بسیار کمی مورد نیاز است) اما دقت این روش هنوز برای حل مسئله تحلیل مرجع پایین است. در روش‌های مبتنی بر یادگیری تقویتی عموماً، نامبری‌ها به‌عنوان حالت‌ها و پیوندهای بین نامبری‌ها به‌عنوان عمل‌ها در نظر گرفته شده‌اند؛ سپس طبق سیاست یادگیری تقویتی و تعریف توابع پاداش و خطا پیوندهای هم‌ارجاعی تشخیص داده شده است. جدیدترین روش‌های ارائه‌شده در حیطه تحلیل هم‌ارجاعی، روش‌های مبتنی بر یادگیری عمیق هستند. در این روش‌ها نامبری‌های موجود در متن توسط ساختار یادگیری عمیق استخراج شده‌اند، سپس با تعریف توابع هزینه مختلف سیستم در جهت شناسایی درست زنجیره‌های هم‌مرجع، آموزش می‌بینند. در فصل آتی روش پیشنهادی که مبتنی بر یادگیری عمیق، بازنمایش دانش و تصمیم‌گیری چندشاخصه است، توضیح داده خواهد شد.

فصل ۳ : مدل مبتنی بر یادگیری عمیق برای تحلیل هم ارجاعی عبارات اسمی

۳-۱ مقدمه

مسئله تحلیل مرجع که یکی از مهم‌ترین مسائل حوزه پردازش زبان طبیعی است، طی چند دهه گذشته همواره یک موضوع تحقیقاتی فعال در این حوزه بوده است. دلیل این موضوع آن است که این مسئله به‌عنوان یک مسئله اصلی در وظایف پردازش زبان طبیعی از جمله ترجمه ماشینی، خلاصه-سازی متن، پاسخ به سؤالات و مشابه آن کاربرد دارد. از طرف دیگر کارایی روش‌های قبلی تحلیل مرجع هنوز پاسخگوی نیاز کاربردهای امروزی نیست. همان‌طور که پیش‌تر توضیح داده شد، یک دلیل مهم برای این ناکارآمدی، دشواری مسئله تحلیل مرجع و نیاز آن به دانش فراوان است.

اعتقاد بر این است که سطح کارایی بهتر در سیستم‌های تحلیل مرجع بدست نمی‌آید، مگر با تزریق دانش به این سیستم‌ها. با توجه به این نیاز در این فصل روش پیشنهادی معرفی می‌شود که هدف آن بهبود کارایی مسئله تحلیل مرجع با استفاده از شبکه عصبی عمیق، فراهم آوردن منابع دانش، بازنمایش آن‌ها و معرفی روش‌های رتبه‌بندی چند شاخصه مبتنی بر شبکه‌های عصبی است.

این فصل از دو بخش اصلی تشکیل شده است. در بخش اول نحوه استخراج ویژگی‌ها، نامبری‌ها و مراجع کاندید توسط تعبیه واژگان، شبکه GRU دوجهته و بازنمایش دانش توضیح داده می‌شود. در بخش دوم نحوه استفاده از اطلاعات سطح خوشه، نحوه امتیازدهی به مراجع کاندید و تشکیل زنجیره-های هم‌مرجع شرح داده می‌شود.

۳-۲ روش پیشنهادی

برای دستیابی به‌دقت بالاتر در تحلیل هم ارجاعی، با استفاده از شبکه عصبی عمیق و منابع دانش، می‌توان ویژگی‌های بدست آمده توسط شبکه را با ضرایب مناسبی با هم ترکیب کرد و به ویژگی‌های کاراتری دست یافت. نتایج حاصل از این ترکیب و استفاده از اطلاعات سطح موجودیت و منابع دانش مختلف در بالا بردن دقت الگوریتم پیشنهادی نقش به‌سزایی دارد. روش پیشنهادی شامل مراحل زیر است.

۳-۲-۱ گام اول: تعبیه واژگان

استخراج و انتخاب ویژگی‌های مناسب از یک مجموعه داده نقش حیاتی در بهبود کیفیت و کارایی روش‌های یادگیری ماشین دارد. خصوصاً در داده‌های با تعداد ابعاد بالا مانند متون، تصویر، صوت، ویدئو و غیره انتخاب ویژگی امری ضروری است. اغلب روش‌های یادگیری ماشین بر روی داده‌های عددی قابل اجرا هستند و برای استفاده و اجرای آنها روی داده‌های متنی نیاز به تبدیل متون به مجموعه اعداد است. در واقع، هدف رویکردهای مختلف تبدیل متن به بردارهای عددی، استخراج و انتخاب مجموعه‌ای از ویژگی‌های مناسب از متون زبان طبیعی است که در مرحله بعد بوسیله روش‌های یادگیری ماشین استفاده می‌شوند. بطور کلی، استخراج ویژگی از مجموعه داده‌ها با دو هدف انجام می‌شود:

۱- افزایش کارایی و سرعت روش‌های رده‌بندی با کاهش ابعاد و اندازه داده‌ها: بخصوص جهت بکارگیری برخی روش‌های رده‌بندی که فاز آموزش آنها هزینه و سربار زمانی یا حافظه‌ای بالایی دارند، این امر ضروری است.

۲- افزایش دقت روش‌های رده‌بندی: با حذف ویژگی‌های نویزی (که وجود آنها باعث افزایش خطای رده‌بندی برای داده‌های جدید می‌شوند) و استخراج ویژگی‌های مناسب (که باعث نزدیک شدن داده‌های درون دسته‌ها و تمایز بیشتر بین داده‌های دسته‌های مختلف می‌شوند).

رویکردهای مختلفی برای بردارسازی متون زبان طبیعی وجود دارند. در این رساله برای استخراج ویژگی و تعبیه کلمات از روش RoBERTa[1] استفاده می‌شود. در ساختار این روش از شبکه عصبی عمیق استفاده شده است. این شبکه‌ها به دلیل قابلیت یادگیری ویژگی‌های ترکیبی و جدید در لایه‌ها،

[1] Robustly Optimized BERT Pretraining Approach

به‌طور گسترده در همه زمینه‌های رده‌بندی مورد استفاده قرار گرفته‌اند. شبکه‌های عمیق توسعه‌یافته، مدل‌هایی برای یادگیری تبدیل غیرخطی، روی داده‌ها هستند.

در این شبکه‌ها، یادگیری بازنمایش داده‌ها توسط خود شبکه صورت می‌گیرد، این عمل در چندین لایه انجام می‌شود و ساختارهای پیچیده با استفاده از لایه‌های متعدد به شبکه آموزش داده می‌شود [۸۴]. مهم‌ترین قابلیت موجود در شبکه‌های عمیق عدم استفاده از ویژگی‌های دستی است. در واقع مهندسی ویژگی در شبکه‌های عمیق به‌صورت خودکار انجام می‌گیرد [۸۵]. دو مزیت این شیوه یادگیری در زیر آمده است.

۱. استخراج ویژگی‌های مناسب: نیاز اصلی هر الگوریتم یادگیری ویژگی‌هایی است که از ورودی‌ها استخراج می‌شود. ممکن است، این ویژگی‌ها از پیش به‌صورت دستی تهیه شده و به الگوریتم خورانده شود که این روش در الگوریتم‌های با ناظر به کار می‌رود. در مقابل روش‌های بدون نظارت خواهد بود که خود اقدام به استخراج ویژگی‌ها از ورودی خواهد نمود. استخراج دستی ویژگی‌ها علاوه بر اینکه زمان‌بر است، معمولاً هم ناقص است.

۲. یادگیری ویژگی‌های سطح بالا با استفاده از قابلیت لایه لایه بودن: یادگیری عمقی برای ما این امکان را به وجود می‌آورد که بتوانیم مفاهیم با سطح انتزاع بالا را با استفاده از یادگیری چند لایه از پایین به بالا بسازیم.

- **تعبیه واژگان RoBERTa**

روش RoBERTa جدیدترین روش بازنمایش یا تعبیه واژگان و جملات است. این روش، بهینه شده‌ی روش BERT است، که با بهبود روش آموزش، روش BERT را بازآموزی کرده است. BERT یک مبدل[1] دوجهته برای پیش‌آموزش روی داده‌های متنی بدون برچسب جهت یادگیری بازنمایش یک

[1] transformer

زبان است که از MLM و NSP استفاده می‌کند. RoBERTa برای بهبود روند آموزش، وظیفه NSP را از پیش‌آموزش BERT حذف کرده است.

همچنین این روش از آموزش دسته‌ای بزرگ‌تر استفاده کرده است. RoBERTa برای پیش‌آموزش از ۱۶۰ گیگابایت داده‌ی متنی استفاده می‌کند، از جمله ۱۶ گیگابایت متون کتاب‌ها و ویکی‌پدیا انگلیسی که در BERT استفاده می‌شود. داده‌های اضافی شامل مجموعه داده‌های پیکره CommonCrawl News (۶۳ میلیون مقاله، ۷۶ گیگابایت)، متون وب (۳۸ گیگابایت) و داستان‌های Common Crawl (۳۱ گیگابایت) است. با استفاده از این پیکره‌ها و اجرای آن‌ها، روند پیش‌آموزش RoBERTa انجام شده است. عملکرد RoBERTa در استخراج ویژگی و تعبیه واژگان از BERT و XLNet بهتر است. از طرف دیگر، برای کاهش زمان‌های محاسباتی (آموزش، پیش‌بینی) BERT یا مدل‌های مرتبط، از روش هرس جهت ایجاد یک شبکه کوچک‌تر برای بهبود عملکرد استفاده می‌کند.

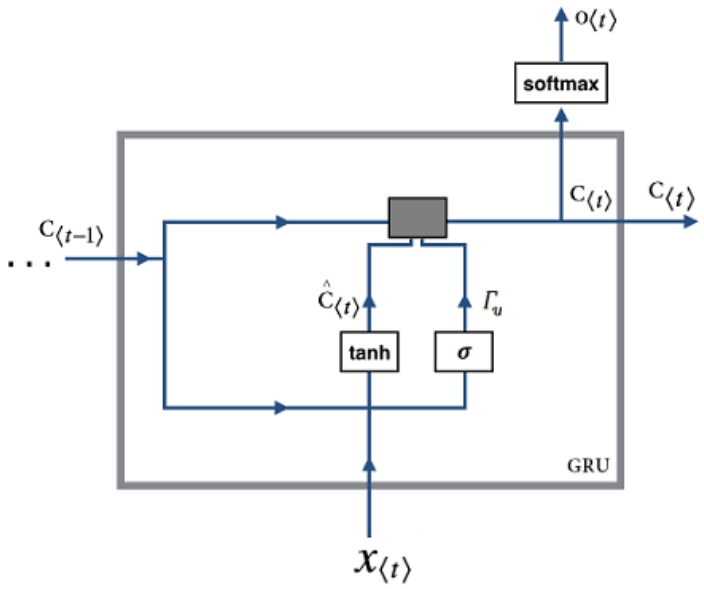
۳-۲-۲ گام دوم: استخراج دنباله کلمات

بردار کلمات و ویژگی‌های استخراج شده از گام قبل به‌عنوان ورودی به یک شبکه عصبی عمیق بازگشتیGRU[1] دوجهته (Bi-GRU) [۸۸] داده می‌شود. از شبکه GRU جهت مدل‌سازی و پیش-بینی داده‌های متوالی استفاده می‌شود. این شبکه مشکل فراموشی دنباله‌های طولانی را حل کرده است، زیرا ورود و خروج داده در واحدهای لایه میانی شبکه کنترل می‌شود. خروجی این شبکه‌ها به خروجی‌های لایه مخفی به ازای همه ورودی‌ها وابسته است و نورون‌های لایه پنهان با بلوک‌های حافظه جایگزین شده‌اند. توسط این شبکه تعبیه کلمات به‌هم الحاق شده، دنباله کلمات موجود در متن و ویژگی‌های آن‌ها شناسایی شده و ویژگی‌های دیگری نیز توسط بازنمایش منابع دانش مختلف به‌عنوان ورودی به مدل تزریق می‌شود.

[1] Gated Recurrent Unit

• شبکه GRU

معماری GRU [۸۸] به‌منظور برطرف کردن کاستی‌های شبکه عصبی بازگشتی سنتی نظیر مشکل محوشدگی گرادیان و کاهش سربار موجود در معماری LSTM ارائه شده است. GRU عموماً به‌عنوان نسخه‌ای تغییریافته از LSTM در نظر گرفته می‌شود، چرا که این معماری‌ها از طراحی مشابهی بهره می‌برند و در بعضی از موارد به‌صورت یکسان نتایج عالی بدست می‌دهند. این نوع معماری از مفاهیمی بنام دروازه به‌روزرسانی[1] و دروازه بازنشانی[2] استفاده می‌کند. این دو دروازه یا gate در اصل دو بردار هستند که با استفاده از آن‌ها تصمیم گرفته می‌شود، چه اطلاعاتی به خروجی منتقل شده و چه اطلاعاتی منتقل نشود. نکته خاص درباره این دروازه‌ها این است که این دروازه‌ها را می‌توان آموزش داد، تا اطلاعات مربوط به گام‌های زمانی بسیار قبل را بدون آنکه در حین گذر زمان (طی گام‌های زمانی مختلف) دستخوش تغییر شوند حفظ کند. ساختار یک سلول GRU به‌صورت شکل۳-۱ است.



شکل ۳-۱. ساختار یک سلول GRU [۸۸]

[1] update gate
[2] reset gate

$$\hat{C}_t = tanh(W_1[\Gamma_r * \mathrm{C}_{t-1}, \mathrm{x}_t] + \mathrm{b}_c) \quad (۱-۳)$$

$$\Gamma_u = \sigma(W_2[\mathrm{C}_{t-1}, \mathrm{x}_t] + \mathrm{b}_u) \quad (۲-۳)$$

$$\Gamma_r = \sigma(W_3[\mathrm{C}_{t-1}, \mathrm{x}_t] + \mathrm{b}_u) \quad (۳-۳)$$

$$C_t = \Gamma_u * \hat{C}_t + (1 - \Gamma_u) * \mathrm{C}_{t-1} \quad (۴-۳)$$

$$o_t = \mathrm{softmax}(W_4\mathrm{C}_t + \mathrm{b}_0) \quad (۵-۳)$$

در رابطه (۳-۴) C_t همان سلول حافظه[1] یا h_t در شبکه عصبی بازگشتی RNN است. یکی از مشکلات شبکه‌های عصبی بازگشتی سنتی عدم توانایی آن‌ها در بهره‌وری از دنباله‌های طولانی است. به‌عنوان مثال در جمله " دانش آموزان منتخب ایران به مرحله نهایی المپیاد ریاضی رسیده و به مقام نخست دست پیدا کردند" شبکه برای اینکه بتواند تشخیص صحیحی در انتخاب فعل انتهایی جمله داشته باشد، نیازمند به خاطر سپردن "دانش آموزان" در ابتدای جمله است. در شبکه عصبی بازگشتی مکانیزم مناسبی برای این مهم وجود ندارد و در بهترین حالت در یک شبکه عصبی بازگشتی سنتی چند گام محدود محلی دارای قدرت تأثیرگذاری در انتخاب‌های شبکه‌اند (چرا که یکی از مشکلات اصلی در RNN سنتی این است که در این شبکه‌ها محتویات گام زمانی فعلی همیشه با یک مقدار جدید از ورودی و حالت قبلی جایگزین می‌شوند و این‌طور اساساً قابلیت حفظ یک ویژگی از چندین گام زمانی خیلی عقب‌تر ممکن نیست) در شبکه عصبی بازگشتی GRU برای رفع این معضل از یک مفهوم تازه به‌نام دروازه به‌روزرسانی استفاده شده است.

به‌طور خیلی ساده دروازه به‌روزرسانی که با $Gamma_u$ نمایش داده می‌شود، اصطلاحاً سوئیچی است، که مشخص می‌کند، در یک گام زمانی حالت قبلی مورد استفاده قرار گیرد یا ورودی (و یا ترکیبی از هر دو). با استفاده از این قابلیت جدید شبکه قادر است، در دنباله‌های طولانی به‌راحتی یک حالت از چندین گام زمانی قبل را در چند گام زمانی بعدی اثر دهد، به‌عبارت‌دیگر شبکه قادر خواهد بود، تا المان‌هایی را از گذشته دور در حافظه خود نگهداشته و از آن بهره‌برداری کند. این کار از طریق رابطه (۳-۴) انجام می‌شود.

[1] memory cell

ایده کلی در اینجا به این صورت است که در یک گام زمانی می‌توان با ۱ یا ۰ کردن گاما مشخص کرد از حالت مربوط به گام زمانی قبل استفاده شود یا حالت جدید با توجه به ورودی تازه و حالت قبلی محاسبه شود. به‌طور دقیق‌تر در اینجا شبکه می‌تواند با میل دادن W_2 به سمت صفر نتیجه عبارت Γ_u را به سمت صفر برده و این‌طور بخش ابتدای عبارت را بی‌اثر و در نتیجه حالت قبلی مورد استفاده قرار گیرد و به گام زمانی بعدی منتقل می‌گردد. به همین صورت شبکه می‌تواند به حالت مابینی و حتی به‌عکس این مهم دست پیدا کند (با فاصله گرفتن از صفر و گرایش به سمت یک). پس دروازه به‌روزرسانی در اینجا حکم سوئیچی را دارد که به شبکه این امکان را می‌دهد تا مشخص کند، چه مقدار اطلاعات از گام‌های زمانی قبلی برای گام زمانی فعلی مورد نیاز است. این قابلیت اضافه‌شده در GRU نسبت به همتای سنتی خود دارای دو فایده است:

۱- هر واحد می‌تواند وجود یک ویژگی خاص در جریان ورودی را برای گام‌های زمانی طولانی بعدی داشته باشد. هر ویژگی‌ای که توسط دروازه به‌روزرسانی GRU مهم تشخیص داده شود، می‌تواند بدون اینکه رونویسی شده و از دست رود، حفظ شود.

۲- این قابلیت جدید در عمل مسیرهای میانبری ایجاد می‌کند که چندین گام زمانی را نادیده گرفته و پشت سر می‌گذارد (از روی چندین گام زمانی می‌پرند) این میانبرها به همین صورت به خطای تولیدی اجازه می‌دهد تا بدون آنکه خیلی سریع محو شود، به‌راحتی در فاز پس انتشار منتقل گردد و این‌طور معضلات مرتبط با گرادیان‌های محو شونده کاهش می‌یابد.

دروازه بازنشانی همانند سوئیچی عمل می‌کند که شبکه با کمک آن می‌تواند مشخص کند، چه میزان از اطلاعات گذشته در گام فعلی مورد نیاز نیست (فراموش شود) و در گام فعلی از چه میزان از اطلاعات گام قبل استفاده شود. به‌طور دقیق‌تر با صفر بودن این سویچ این دروازه در عمل شبکه را وادار می‌کند به‌گونه‌ای عمل کند که گویا در حال خواندن اولین بخش از دنباله ورودی است و این‌طور شبکه را قادر به فراموشی حالت محاسبه شده قبلی می‌کند و به همین صورت می‌تواند با فاصله گرفتن از صفر حالت مابینی را فراهم آورد. با کمک این دو قابلیت جدید شبکه عصبی GRU به‌راحتی

می‌تواند نسبت به ذخیره‌سازی و یا فیلتر کردن اطلاعات از گام‌های زمانی قبلی اقدام کرده و از دنباله‌های طولانی که شبکه عصبی بازگشتی سنتی در مواجه با آن با مشکلات عدیده‌ای مواجه بود بهره‌برداری کند و کاستی‌های شبکه عصبی بازگشتی سنتی را مرتفع سازد. این نوع شبکه‌ها بسیار قدرتمند بوده و به‌واسطه سربار کمتر آن‌ها نسبت به LSTM و قدرت بیشتر آن‌ها نسبت به RNN سنتی در روش پیشنهادی مورد استفاده گسترده قرار گرفته‌اند.

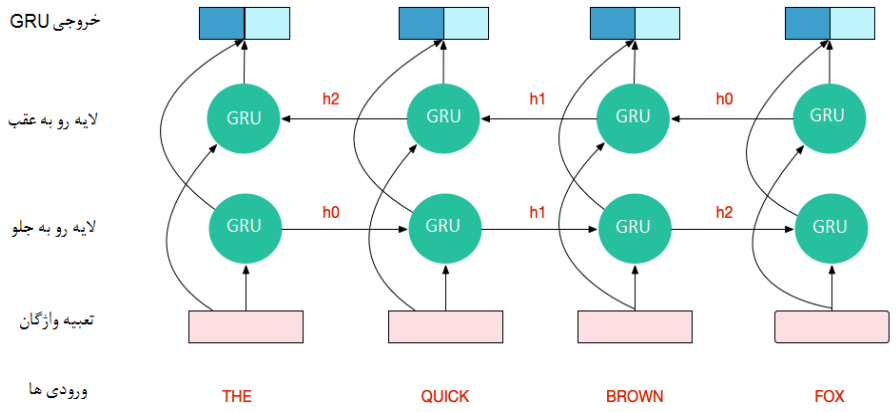
- **استخراج بازنمایش دنباله کلمات**

شبکه‌های Bi-GRU[1] یا GRU دو جهته، نسبت به شبکه‌های GRU معمولی، مفهوم را بهتر درک می‌کنند. این شبکه‌ها شامل دو لایه پنهان بازگشتی هستند. هر لایه پنهان شامل بلوک‌های GRU می‌باشد. در این شبکه‌ها، ورودی‌ها به دو لایه پنهان مجزا، که هر دو به یک لایه‌ی خروجی متصل هستند، داده می‌شوند. هر دنباله ورودی در دو جهت زمانی رو به جلو و از انتها به دو لایه پنهان بازگشتی مجزا داده شود. مقدار خالص رسیده به لایه خروجی جمع وزن‌دار مقدار خالص دو لایه‌ی روبه‌جلو و رو به عقب است. شبکه دو جهته، زمینه‌ی متن را بهتر درک می‌کند و به‌راحتی کلمه‌ی بعدی را پیش‌بینی می‌کند. در روش پیشنهادی تعبیه واژگان وارد این شبکه شده و الحاق شده‌ی آن-ها استخراج می‌شود. پس از Bi-GRU از چندین لایه دیگر استفاده شود، تا اطلاعات و ویژگی‌های بیشتر و مفیدتری از بردارهای بازنمایش استخراج شود، استفاده از این تکنیک در مسئله‌ی شناسایی نامبری‌ها مفید است.

با فرض اینکه سند D شامل T کلمه باشد، تعبیه واژگان استخراج‌شده توسط RoBERTa را به‌صورت $\{x_1, \ldots, x_T\}$ در نظر می‌گیریم. مطابق شکل ۳-۲ برای محاسبه بازنمایش (نمایش برداری) دنباله کلمات[2]، تعبیه واژگان را طبق روابط (۳-۶ -۳-۱۰) به شبکه بازگشتی Bi-GRU تزریق می‌کنیم. در رابطه (۳-۶)، Wها وزن‌های مدل برای هر واحد هستند. $\tilde{h}_t$ حالت پنهان، σ تابع منطقی سیگموئید

[1] Bidirectional Gated Recurrent Unit
[2] span representation

است، که به‌صورت $\sigma(x) = 1/(1 + e^{-x})$ تعریف می‌شود. r_t دروازه بازنشانی، z_t دوازه به‌روزرسانی و x_t ورودی شبکه است. $\sigma\epsilon[-1,1]$ جهت هر GRU را نشان می‌دهد و $\odot$ ضرب عنصر به عنصر دو بردار است. x_t^* خروجی الحاق شده‌ی Bi-GRU است. برای هر جمله از یک GRU مستقل استفاده می‌شود.



شکل ۳-۲. شبکه‌ی Bi-GRU

همان‌طور که گفته شد، هر واحد GRU در مکان t شامل دو گیت یا دروازه به‌روزرسانی (z_t) و بازنشانی (r_t) است.

$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1}) \quad (۶-۳)$$

$$z_t = \sigma(W_{zx}x_t + W_{zh}h_{t-1}) \quad (۷-۳)$$

$$\tilde{h}_t = tanh(W_{hx}x_t + W_{hh}(r_t \odot h_{t-1})) \quad (۸-۳)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t, \quad (۹-۳)$$

$$x_t^* = [h_{t+1}, h_{t-1}] \quad (۱۰-۳)$$

کلمه سر[1] (کلمه مهم و کلیدی) عبارات اسمی را می‌توان با کمک تجزیه‌گر نحوی[2] بدست آورد. مثلاً در نامبری Former Barcelona player کلمه سر player است، که پس از تعمیم به دسته معنایی person تبدیل می‌شود. به‌جای استفاده از تجزیه‌گر نحوی، مدل پیشنهادی با استفاده از مکانیزم

[1] head
[2] syntactic parses

توجه[1] [۸۹] وظیفه یافتن کلمات سر را طبق روابط (۳-۱۱الی۳-۱۳) می‌آموزد، که در آن FFNN شبکه عصبی روبه‌جلو است. توجه در واقع خروجی یک لایه تمام‌متصل با تابع فعال‌ساز Softmax است. این تکنیک باعث می‌شود، رمزگشا[2] بتواند در هر قدم فقط روی کلمات مهم تمرکز و توجه کند.

$$\alpha_t = w_\alpha . FFNN_\alpha(x_t^*) \quad (۱۱-۳)$$

$$a_{i,t} = \frac{exp(\alpha_t)}{\sum_{k=START(i)}^{END(i)} exp(\alpha_k)} \quad (۱۲-۳)$$

$$\hat{x}_i = \sum_{t=START(i)}^{END(i)} a_{i,t} . x_t \quad (۱۳-۳)$$

$\hat{x}_i$ مجموع وزنی بردار کلمات در دنباله کلمات i است. وزن‌های $a_{i,t}$ به‌صورت خودکار و توسط شبکه یاد گرفته می‌شوند و با کلمات سر مرتبط هستند. در نهایت برای دنباله کلمات i بازنمایش رابطه (۳-۱۴) استخراج می‌شود، که در آن $x^*_{START(i)}$ و $x^*_{END(i)}$ بازنمایش‌های مرزی ابتدا و انتهای دنباله کلمات، $\emptyset(i)$ بردار ویژگی و $\hat{x}_i$ بازنمایش برداری و مبتنی بر مکانیزم توجه کلمه سر است.

$$g_i = [x^*_{START(i)}, x^*_{END(i)}, \hat{x}_i, \emptyset(i)] \quad (۱۴-۳)$$

۳-۲-۳ گام سوم: بازنمایش دانش

هدف از تحلیل هم‌ارجاعی، تخصیص مراجع (y_i) به دنباله کلمات i در سند است. مجموعه مقادیر ممکن برای مراجع به‌صورت $\gamma(i) = \{\epsilon, 1, \dots, i-1\}$ تعریف می‌شود. درایه اول نشان می‌دهد که دنباله کلمات مدنظر هیچ مرجعی در متن ندارد، درایه دوم نشان می‌دهد که نامبری مدنظر یک مرجع در متن دارد و به همین ترتیب. برای دنباله کلمات i توزیع $p(y_i)$ به‌صورت رابطه (۳-۱۵) روی مراجع تعریف می‌شود، در این رابطه، F تابع امتیاز است. که نحوه محاسبه آن در ادامه شرح داده می‌شود.

[1] attention mechanism
[2] decoder

$$p(y_i) = \frac{e^{F(i,y_i)}}{\sum_{y \in \gamma(i)} e^{F(i,y)}} \quad (۱۵-۳)$$

به‌منظور شناسایی نامبری‌ها و مراجع آن‌ها، به دانش‌های مختلفی برای توصیف ویژگی‌های نامبری‌ها و مراجع کاندید نیاز داریم. این دانش می‌تواند دانش پایه‌ای باشد، که با روش‌های سطحی مثل مقایسه رشته‌ها به دست می‌آید، یا دانش عمقی باشد که با تحلیل ارتباط معنایی و ساختار گفتمان بدست می‌آید.

برخی از دانش‌های لازم عبارتند از:

- دانش زبان‌شناسی: به‌منظور بررسی توافق بین نامبری و مرجع در مواردی مثل جنسیت و تعداد و جاندار بودن.
- دانش معنایی: به‌منظور یا دیگر ارتباطات معنایی مثل روابط والد/فرزند و مترادف بین نامبری و مرجع.
- دانش نحوی: رفتار ضمایر توسط قواعد دستور زبان مشخص می‌شود، مثلاً ضمایر انعکاسی معمولاً به فاعل جمله اشاره می‌کنند. همچنین مثلاً در زبان انگلیسی اهمیت مفهوم توسط سلسله مراتب قواعد گرامری تعیین می‌شود.
- دانش مربوط به گفتمان: اطلاعاتی در مورد اینکه کدام بخش گفتمان از بقیه قسمت‌ها مهم‌تر است.
- نزدیکی اطلاعات: تازگی یا نزدیک بودن اطلاعات (از نظر مکان) در متن فاکتور مهمی است زیرا در ضمایر و دیگر نامبری‌ها مثل عبارات اسمی، مرجع نامبری معمولاً نزدیک به نامبری قرار دارد.
- دانش جهانی: به‌منظور حل برخی موارد هم‌ارجاعی استفاده از دانش زبانی به‌تنهایی کافی نخواهد بود بلکه لازم است از نوعی دانش که به آن دانش جهانی گفته می‌شود نیز بهره جست.

بیشتر دانش مورد استفاده برای تحلیل هم‌ارجاعی، از پیش‌پردازش‌های مرتبط با پردازش زبان طبیعی بدست می‌آید که عمدتاً در لایه‌های لغوی و نحوی انجام می‌گیرد. بسته به زبانی که با آن کار می‌کنیم (انگلیسی، فارسی و غیره) و همچنین دامنه مورد استفاده، می‌توان از منابع دانشی متفاوتی برای ساخت مجموعه ویژگی‌ها استفاده کرد. برای زبان انگلیسی، منابع اطلاعاتی رایج عبارتند از برچسب‌زنی اجزای کلام، تحلیل ساختواژی لغت، تجزیه‌کننده و رده‌بندی معنایی بر پایه اطلاعاتی مثل والد/ فرزند و مترادف از WordNet.

در سطح لغوی اطلاعاتی وجود دارد که می‌تواند در حل مسئله تحلیل هم‌ارجاعی مفید باشد، مثل تطبیق رشته‌ای از طریق تغییر ظاهر رشته (حروف کوچک و بزرگ، استفاده یا عدم استفاده از حروف اضافه و تعریف، و غیره)، و نام‌های مستعار (مخفف‌ها، سرنام‌ها، کنیه و القاب و غیره).

ویژگی همخوانی جنسیت و تعداد می‌تواند توسط نشانه‌های مختلفی مشخص شود، مثل پیشوندهایی شبیه خانم و آقا و یا لیستی از اسامی خانم‌ها و آقایان برای موجودیت‌های نامدار و همچنین وجود s جمع در انتهای اسم و تعریف‌کننده‌های قبل از اسم. می‌توان از تعیین دسته معنایی کلمه سر، برای تعیین جنسیت یک عبارت اسمی استفاده کرد: جنسیت مواردی که شیء هستند، به‌عنوان خنثی معرفی می‌شود و دیگر موارد یا به‌عنوان شخص معرفی می‌شوند یا اگر نتوان نوع آن‌ها را تشخیص داد به‌عنوان نامشخص معرفی می‌شوند. هنگام ساخت ویژگی‌های همخوانی جنسیت برای یک جفت نامبری-مقدم، از سه مقدار استفاده می‌شود: نامشخص، برای مواردی که جنسیت اقلاً یک مورد نامشخص است، سازگار، برای مواردی که جنسیت‌ها همخوانی دارد و ناسازگار، برای مواردی که جنسیت‌ها با یکدیگر متفاوت است.

برای تشخیص دقیق نامبری‌ها و مراجع کاندید از بازنمایش دانش[1] استفاده می‌کنیم. هدف از انجام بازنمایش دانش، غنی‌سازی دانش سیستم و سپس انتخاب مناسب‌ترین دانش برای حل هر یک از

[1] Knowledge representation

موارد هم‌ارجاعی است. بازنمایش دانش، حوزه‌ای از هوش مصنوعی است که به مناسب‌ترین اشکال و روش‌های ذخیره دانش در سیستم‌های کامپیوتری می‌پردازد. زمانی که رایانه‌ها بخواهند، در کنار انسان یا به‌جای انسان به استدلال بپردازند، نقش بازنمایش دانش به سبب دشواری‌های ناشی از مقیاس‌پذیری، حیاتی‌تر و اجتناب‌ناپذیر می‌گردد. این فرآیند باعث می‌شود، سیستم‌ها بتوانند وظایف پیچیده‌ای نظیر تشخیص یک بیماری پزشکی یا مکالماتشان با انسان‌ها را به‌راحتی انجام دهند. بازنمایش دانش نقش مهمی در بهبود عملکرد سیستم‌های مبتنی بر پردازش زبان طبیعی مانند تحلیل هم‌ارجاعی، پرسش و پاسخ، ترجمه ماشینی و غیره دارد. دلیل اینکه ما در این پایان‌نامه از بازنمایش دانش با استفاده از منابع لغوی، نحوی، معنایی و جهانی استفاده کردیم این است که بیشتر مشکلات تحلیل هم‌ارجاعی فقط با استفاده از این منابع دانش قابل حل است. . در واقع، هر یک از موارد هم‌ارجاعی به یک یا چند منبع دانش برای حل شدن نیاز دارد. علاوه بر این، با استفاده از منابع دانش یادگیری سیستم بهبودیافته و قدرت آن برای شناسایی زنجیره‌های هم‌ارجاعی در متن بالاتر می‌رود. برای این هدف از سه‌تایی‌های "ابتدا[1]، رابطه[2]، انتها[3]" استفاده کردیم. سپس، برای ایجاد منبع دانش کلی یا گراف دانش ($\mathcal{G}$)، تمامی آن‌ها را با هم ترکیب کردیم.

برای هر دنباله مدنظر، جهت نامبری بودن یا هم‌مرجع بودن با نامبری دیگر، دانش‌های مختلفی از منابع دانش و با روش‌های مختلف می‌توان استخراج کرد. برای سادگی و تعمیم در مدل پیشنهادی از تطابق ساختار رشته‌ای[4] برای استخراج دانش مرتبط استفاده نمودیم. به‌طور خاص، برای هر سه‌تایی $t \in \mathcal{G}$ که ابتدا و انتهای آن لیستی از کلمات است، اگر ابتدای آن مشابه با رشته‌ای در دنباله کلمات s باشد، آن را یک سه‌تایی مرتبط می‌دانیم. بنابراین، اطلاعات t را با میانگین‌گیری تعبیه تمام کلمات انتهای آن، کدگذاری می‌کنیم. برای مثال، اگر دنباله موردنظر ’the apple‘ باشد و سه‌تایی ’the)

[1] head
[2] relation
[3] tail
[4] string match

('apple', IsA, 'healthy food') با جستجو در منبع دانش یافت شود، این رابطه را توسط میانگین تعبیه کلمات 'healthy' و 'food' نمایش می‌دهیم. برای دنباله کلمات i مجموعه دانش بازیابی شده را با $\mathcal{K}_{s_i}$ نشان می‌دهیم، که حاوی n بازنمایش دانش مرتبط به‌صورت $k_{1,s_i}, k_{2,s_i}, \dots, k_{n,s_i}$ می‌باشد. برای ادغام دانش فوق‌الذکر در مدل پیشنهادی، با این چالش روبرو هستیم که تعداد زیادی تعبیه وجود دارد، درحالی‌که بیشتر آن‌ها در زمینه‌های خاص بی‌فایده هستند. برای حل این مشکل، از مکانیزم توجه دانش[1] جهت انتخاب مناسب‌ترین دانش استفاده کردیم. برای دنباله کلمات i، وزن هر $k_x \in \mathcal{K}_{s_i}$ توسط روابط (۳-۱۶ و ۳-۱۷) محاسبه می‌شود. برای هر جفت دنباله کلمات i و j، بازنمایش آن‌ها به هم الحاق شده و بازنمایش کلی $g_{i,j}$ حاصل می‌شود، تا دانش مناسب انتخاب شود. دانش کلی توسط رابطه (۳-۱۸) بدست می‌آید.

$$w_x = \frac{e^{\beta_{k_x}}}{\sum_{k_y \in \mathcal{K}_{s_i}} e^{\beta_{k_y}}} \quad (۳-۱۶)$$

$$\beta_k = NN_\beta([g_{i,j}, k]) \quad (۳-۱۷)$$

$$o_{s_i} = \sum_{k_x \in \mathcal{K}_{s_i}} w_x . k_x \quad (۳-۱۸)$$

جهت بررسی اعتبار دنباله کلمات از جهت نامبری بودن، تعبیه دنباله به همراه دانش مربوط به آن را که توسط رابطه (۳-۱۸) استخراج شد، طبق رابطه (۳-۲۰) به یک شبکه عصبی روبه‌جلو می‌دهیم، تا بر اساس ویژگی‌هایی که از مرحله قبل استخراج شده و نامبری‌هایی که سیستم آموزش دیده، نامبری‌بودن یا نبودن آن مشخص شود. در رابطه (۳-۲۰) $f_m(s_i)$ تابع امتیازدهی به دنباله کلمات s_i است که مشخص شود یک نامبری معتبر است یا خیر. در رابطه (۳-۲۱) $f_c(s_i, s_j)$ نیز تابع امتیازدهی به دو دنباله s_i و s_j است که مشخص شود آیا دو دنباله هم‌مرجع هستند یا خیر. پس از محاسبه مقادیر برای تابع F طبق رابطه (۳-۱۹)، خوشه‌های هم‌ارجاعی اولیه طبق رابطه (۳-۱۵) تشکیل می‌شود.

[1] knowledge attention module

$$F(s_i, s_j) = f_m(s_i) + f_m(s_j) + f_c(s_i, s_j) \quad (۱۹-۳)$$

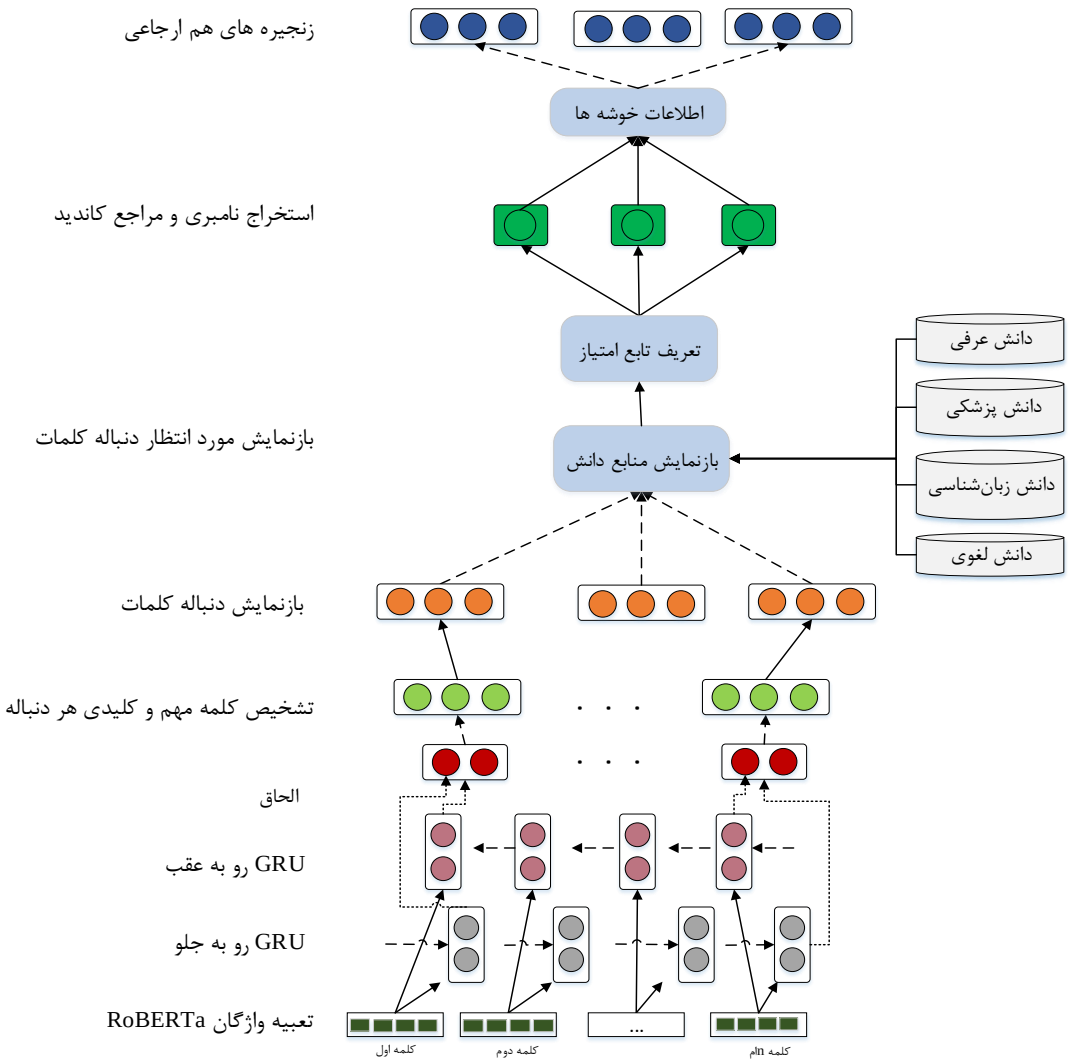
$$f_m(s_i) = NN_m([g_i, o_{s_i}]) \quad (۲۰-۳)$$

$$f_c(s_i, s_j) = NN_c([g_i, o_{s_i}, g_j, o_{s_j}, g_i \odot g_j, o_{s_i} \odot o_{s_j}]) \quad (۲۱-۳)$$

در داده‌های آموزش، فقط اطلاعات خوشه‌ها وجود دارد و اطلاعات مراجع پنهان است. بنابراین با استفاده از خوشه‌بندی درست و مراجع درست، شبکه توسط رابطه (۳-۲۲) آموزش می‌بیند.

$$\log \Pi_{i=1}^{N} {}_{\hat{y} \in \gamma(i) \cap GOLD(i)} P(\hat{y}) \quad (۲۲-۳)$$

در رابطه (۳-۲۲)، $GOLD(i)$ مجموعه دنباله کلمات موجود در خوشه درست شامل دنباله کلمه i است. اگر دنباله کلمه i به یک خوشه درست تعلق نداشته باشد و تمام مراجع درست، هرس شده باشند، آنگاه $GOLD(i) = \{\epsilon\}$. در اینصورت، مدل به‌طور طبیعی یاد می‌گیرد که دنباله‌ها را به‌طور دقیق هرس کند. ساختار روش پیشنهادی در شکل ۳-۳ نمایش داده شده است.



شکل ۳-۳. نحوه استخراج دنباله کلمات و نامبری‌ها

۳-۲-۴ گام چهارم: تشکیل زنجیره‌های هم‌ارجاعی نهایی

جهت تولید ویژگی‌هایی که به بازنمایش جهانی و سطح خوشه نزدیک‌ترند و برای دستیابی به بازنمایش دقیق‌تر دنباله کلمات از اطلاعات خوشه‌های هم‌مرجع استفاده می‌کنیم. به این صورت که طبق رابطه (۳-۲۳) بازنمایش هر نامبری را توسط مجموع بازنمایش نامبری‌هایی که در حال حاضر در خوشه هم‌ارجاعی هستند، در نظر می‌گیریم. در اینصورت به ویژگی‌های نزدیک‌تر به بازنمایش سطح خوشه و جهانی دست پیدا می‌کنیم.

$$g'_i = \sum_{j \in C(i)} g_j \qquad (۲۳-۳)$$

در نتیجه دنباله کلمات شامل اطلاعاتی درباره خوشه‌های هم‌مرجع خود می‌شود. جهت همسان‌سازی بازنمایش دنباله کلماتی که متعلق به یک خوشه هستند، فرض می‌کنیم اگر m نامبری در متن موردنظر وجود داشته باشند، حداکثر m موجودیت نیز وجود دارد. احتمال تناظر نامبری i با موجودیت j بر اساس توزیع مراجع کاندید ($p(y_i)$) و به‌صورت بازگشتی طبق رابطه (۳-۲۴) محاسبه می‌شود. در این رابطه $e_i(t)$ بازنمایش موجودیت در زمان t است.

$$Q(i \in E_j) = \begin{cases} \sum_{k=j}^{i=1} p(y_i = k).\, Q(k \in E_j) & j < i \\ p(y_i = \epsilon) & j = i \\ 0 & j > i \end{cases} \qquad (۲۴-۳)$$

جهت همسان‌سازی بازنمایش نامبری‌ها و استفاده از بازنمایش موجودیت‌ها طبق رابطه (۳-۲۵) به بازنمایش موجودیت E_i در زمان t نیاز داریم.

$$e_i(t) = \sum_{j=1}^{t} Q(j \in E_i).\, g_j \qquad (۲۵-۳)$$

بازنمایش نهایی نامبری i طبق رابطه (۳-۲۶) با استفاده از توزیع موجودیت‌ها و بازنمایش جهانی محاسبه می‌شود و مجدداً امتیاز نامبری توسط تابع امتیاز محاسبه می‌شود.

$$g'_i = \sum_{j=1}^{i} Q(i \in E_j).\, e_j^{(i)} \qquad (۲۶-۳)$$

پس از محاسبه بازنمایش جدید برای دنباله کلمات (g'_i) امتیاز نامبری توسط رابطه (۳-۳۱) محاسبه می‌شود. در واقع، توسط اطلاعات مبتنی بر موجودیت گام قبل، موجودیت‌های نامدار و سایر نامبری-های موجود در متن استخراج می‌شوند. امتیاز هم‌ارجاعی دنباله کلمات i و j نیز توسط رابطه (۳-۳۲) محاسبه می‌شود. در نهایت توسط رابطه (۳-۳۳) زنجیره‌های هم‌ارجاعی نهایی استخراج می‌شود.

$$w_x = \frac{e^{\beta k_x}}{\sum_{k_y \in \mathcal{K}_{s_i}} e^{\beta k_y}} \qquad (۲۷-۳)$$

$$\beta_k = NN_\beta([g'_{i,j}, k]) \quad (۲۸-۳)$$

$$o'_{s_i} = \sum_{k_x \in \mathcal{K}_{s_i}} w_x . k_x \quad (۲۹-۳)$$

$$F'(s_i, s_j) = f'_m(s_i) + f'_m(s_j) + f'_c(s_i, s_j) \quad (۳۰-۳)$$

$$f'_m(s_i) = NN_m([g'_i, o'_{s_i}]) \quad (۳۱-۳)$$

$$f'_c(s_i, s_j) = NN_c([g'_i, o'_{s_i}, g'_j, o'_{s_j}, g'_i \odot g'_j, o'_{s_i} \odot o'_{s_j}]) \quad (۳۲-۳)$$

$$p'(y_i) = \frac{e^{F'(g'_i, g'_{y_i})}}{\sum_{y \in \gamma(i)} e^{F'(g'_i, g'_y)}} \quad (۳۳-۳)$$

۳-۲-۵ رتبه‌بندی مراجع کاندید توسط شبکه‌های عصبی

جهت رتبه‌بندی مراجع کاندید برای هر نامبری مدنظر روش‌های رتبه‌بندی چند شاخصه نیز ارائه شده است. دیاگرام روش پیشنهادی در شکل ۳-۴ نمایش داده شده است. مراحل انجام این روش به‌صورت زیر است.

۱- مدلسازی مسئله و تشکیل ماتریس تصمیم

پس از شناسایی و استخراج نامبری‌ها، برای مراجع کاندید ماتریسی به‌صورت زیر در نظر گرفته می‌شود. در رابطه (۳-۳۴) r_{ij} نشان‌دهنده‌ی مقدار ویژگی jام برای مرجع کاندید iام است.

$$M = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1j} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2j} & \dots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{i1} & r_{i2} & \dots & r_{ij} & \dots & r_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nj} & \dots & r_{nm} \end{bmatrix} \quad (۳۴-۳)$$

برای تعیین وزن ویژگی‌ها مقدار آنتروپی E_j توسط روابط (۳-۳۵ و ۳-۳۶) محاسبه می‌شود. همچنین درجه انحراف (d) و وزن اولیه شاخص‌ها (w') به‌ترتیب توسط روابط (۳-۳۷) و (۳-۳۸) محاسبه می‌شوند.

$$k = \frac{1}{ln(n)} \quad (۳۵-۳)$$

$$E_j = -k\sum_{i=1}^{n}[r_{ij} ln r_{ij}] \quad (۳۶-۳)$$

$$d_j = 1 - E_j; \forall j \quad (۳۷-۳)$$

$$w_j' = \frac{d_j}{\sum_{j=1}^{m} d_j} \quad (۳۸-۳)$$

از طریق تلفیق بردار $m = [r_1 \quad r_2 \quad \dots \quad r_m]$ با وزن‌های ویژگی‌های مراجع کاندید، طبق رابطه (۳-۳۹) وزن ویژگی‌ها برای نامبری مدنظر محاسبه می‌شود.

$$w_j = \frac{r_j . w_j'}{\sum_{j=1}^{m} r_j . w_j'}, \sum w_j = 1 \quad (۳۹-۳)$$

۲- وزن‌دار کردن ماتریس تصمیم‌گیری

جهت تأثیر وزن‌های معیارها در ماتریس تصمیم‌گیری، طبق رابطه (۳-۴۰) هر ستون ماتریس در اهمیت مربوط به آن معیار ضرب می‌شود.

$$V = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \dots & w_m r_{1m} \\ w_1 r_{21} & w_2 r_{22} & \dots & w_m r_{2m} \\ \dots & \dots & \dots & \dots \\ w_1 r_{n1} & w_2 r_{n2} & \dots & w_m r_{nm} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \dots & \dots & \dots & \dots \\ v_{n1} & v_{n2} & \dots & v_{nm} \end{bmatrix} \quad (۴۰-۳)$$

۳- نرمال‌سازی ماتریس M

جهت اینکه ماتریس را به‌عنوان ورودی شبکه بتوان در نظر گرفت، لازم به یک مرحله نرمال‌سازی می‌باشد. رابطه (۳-۴۱) جهت نرمال‌سازی ماتریس وزن‌دار به کار گرفته شده است. بعد از انجام این عملیات، ماتریس نرمال وزن‌دار حاصل می‌شود و به‌عنوان ورودی به شبکه معرفی می‌شود.

$$a_{ij} = \frac{v_{ij}}{\max_j\{v_{ij}\}} \quad (۴۱-۳)$$

۴- رتبه‌بندی مراجع کاندید توسط شبکه‌های عصبی

الف) شبکه عصبی پرسپترون

جهت به‌کارگیری شبکه پرسپترون [۹۰] لازم است، یک سری نمونه ورودی و خروجی‌های متناظر با آن‌ها مشخص گردد و جهت آموزش به شبکه معرفی شود. در شبکه‌های با ناظر، مسئله اصلی تعیین نمونه‌های مجموعه آموزش می‌باشد. برای حل مسائل رتبه‌بندی، نمونه‌های مجموعه آموزش باید طوری باشد که همه حالت‌های مختلف گزینه‌ها را در بر گیرد. خروجی‌های شبکه پرسپترون، متناظر با لیست مرتب گزینه‌ها می‌باشد، بنابراین تعداد نورون‌ها در لایه خروجی شبکه برابر با تعداد حالت‌هایی است که گزینه‌ها نسبت به هم خواهند داشت.

شبکه پیشنهادی شامل n گزینه و m معیار، شبکه‌ای با nm ورودی و $n!$ خروجی می‌باشد. تابع خروجی به‌صورت تابع رقابتی می‌باشد. یعنی خروجی بیشترین مقدار نورون خروجی برابر با یک و بقیه مقادیر، صفر در نظر گرفته می‌شوند. تابع خروجی $\emptyset(V(t))$ طبق رابطه (۳-۴۲) می‌باشد.

$$\emptyset\big(v(t)\big) = \begin{cases} 1: max(v(t)) \\ 0: otherwise \end{cases} \quad (۴۲-۳)$$

نمونه‌های آموزش با استفاده از روابط زیر، حاصل می‌گردند.

$$f_{min}^{j} = \min_{i} a_{ij} \quad (۴۳-۳)$$

$$f_{max}^{j} = \max_{i} a_{ij} \quad (۴۴-۳)$$

$$T_i = \begin{bmatrix} t_i^{11} & t_i^{12} & \dots & t_i^{1m} \\ t_i^{21} & t_i^{22} & \dots & t_i^{2m} \\ \dots & \dots & \dots & \dots \\ t_i^{n1} & t_i^{n2} & \dots & t_i^{nm} \end{bmatrix} \quad (۴۵-۳)$$

$$t_i^{jk} = \frac{f_{max}^{j} - f_{min}^{j}}{m-1}(j-1) + f_{min}^{j} \quad (۴۶-۳)$$

با جابجایی سطرهای مختلف ماتریس T_i، نمونه‌های دیگر حاصل می‌شود. تعداد حالت‌هایی که n گزینه نسبت به هم می‌توانند داشته باشند، برابر $n!$ می‌باشد. بنابراین جابجایی n سطر مختلف ماتریس T_i کلیه حالت‌های ممکن جهت آموزش شبکه را تولید می‌نماید. این نمونه‌ها به همراه خروجی متناظر جهت آموزش شبکه معرفی می‌شوند. بعد از آموزش، ماتریس تصمیم به شبکه معرفی

می‌شود و با توجه به خروجی شبکه، اولویت گزینه‌ها نسبت به هم تعیین می‌شود و مراجع کاندید شده (گزینه‌ها) برای نامبری موردنظر رتبه‌بندی می‌شوند.

ب) شبکه رقابتی کوهونن

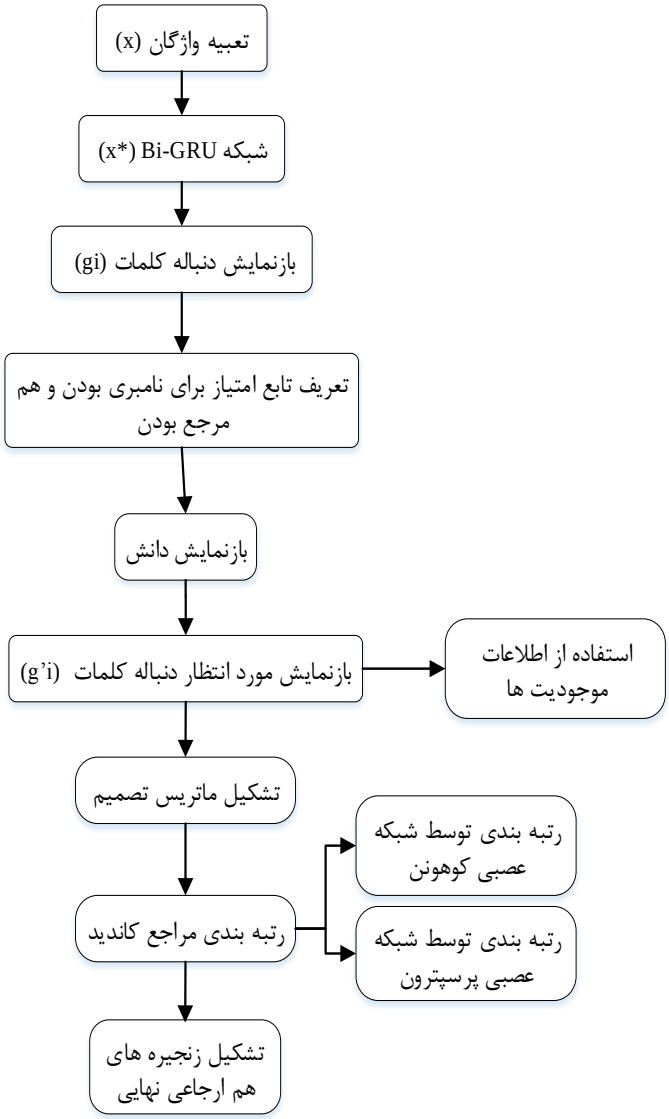
یکی از مشکلات شبکه پرسپترون، تعیین نمونه‌های آموزش می‌باشد. نتایج حاصل از شبکه پرسپترون کاملاً وابسته به تعیین نمونه‌های مجموعه آموزش می‌باشد. از میان شبکه‌های بدون ناظر، شبکه رقابتی کوهونن [۹۱] را به دلیل خاصیت خود ساماندهی مورد استفاده قرار دادیم. شبکه خود سامانده کوهونن به دلایل زیر مورد استفاده قرار می‌گیرد.

۱. تعریف نمونه‌های آموزش برای معرفی به شبکه نیاز به اطلاعات قبلی از مسئله ندارد.

۲. تعداد معیارها و گزینه‌های هر مسئله تصمیم‌گیری متفاوت است، بنابراین برای اینکه شبکه آموزش داده شده برای هر تعداد گزینه قابل‌قبول باشد، لازم است شبکه‌هایی مورد استفاده قرار گیرد که نیاز به اطلاعات مسئله در مورد تعداد گزینه‌ها نداشته باشد.

۳. شبکه رقابتی کوهونن، در حقیقت رقابت بین گزینه‌ها را برای انتخاب شدن مدل‌سازی می‌نماید.

۴. شبکه‌ای که در این روش به کار می‌بریم، دارای n ورودی و تعدادی ثابت نورون در لایه خروجی می‌باشد. تعداد نورون‌های لایه خروجی را می‌توان ثابت یا متغیر در نظر گرفت. تعداد نمونه‌های مجموعه آموزش با توجه به تعداد نورون‌های خروجی حاصل می‌گردد. نمونه‌های آموزش، نماینده گزینه‌ها به شمار می‌روند و بنابراین باید طوری انتخاب گردند که همه حالت‌های ممکن را در بر گیرند. از آنجائیکه ماتریس ورودی به شبکه، نرمال می‌باشد، لذا مقادیر نمونه‌های آموزش باید طوری باشد که بازه [۱و۰] را به‌طور کامل پوشش دهد.



شکل ۳-۴. دیاگرام روش پیشنهادی

در اینجا، ماتریس تصمیم‌گیری به‌عنوان ورودی به شبکه معرفی می‌شود، و با توجه به خروجی شبکه لیست مرتب گزینه‌ها حاصل می‌گردد. برای انجام این کار، نمونه‌های T_i جهت آموزش به شبکه معرفی می‌شوند. خروجی شبکه کوهونن از میان نورون‌های خروجی، نورونی خواهد بود که بیشترین میزان شباهت یا کمترین فاصله اقلیدسی را با نمونه ورودی داشته باشد. فواصل، برای هریک از نورون‌های لایه خروجی طبق رابطه (۳-۴۷) محاسبه می‌شود.

$$d_{min} = min\{d_j = \sum_{i=1}^{n} \left(g'_i - w_{ij}\right)^2, j = 1, \dots, m \qquad (۴۷-۳)$$

نورون خروجی برنده، مشخص می‌شود و با به‌کارگیری تابع همسایگی، طبق رابطه (۳-۴۸) وزن‌ها اصلاح می‌شوند.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)N(t)(g'_i - w_{ij}(t)) \quad (۴۸-۳)$$

مقدار پارامتر آموزش و تابع همسایگی طبق رابطه (۳-۴۹) تعریف می‌شود.

$$N(t) = \begin{cases} \frac{g'}{\alpha} + 1 & x > \alpha \\ -\frac{g'}{\alpha} + 1 & x < \alpha \\ 0 & otherwise \end{cases} \quad (۴۹-۳)$$

مقدار a با توجه به نمونه‌های آموزش تغییر می‌کند. مقادیر مربوط به هر گزینه از ماتریس تصمیم-گیری به شبکه معرفی می‌شود و با توجه به خروجی شبکه، هریک از گزینه‌ها به یکی از مجموعه‌های مربوط به خروجی، تعلق می‌یابند. بدین ترتیب، مرتب‌سازی کلی انجام می‌شود. جهت مرتب‌سازی گزینه‌ها، مقادیر حقیقی خروجی شبکه، مورد تحلیل قرار می‌گیرد. شاخص تصمیم‌گیری در این مرحله از مرتب‌سازی به‌صورت زیر در نظر گرفته می‌شود:

$$b_i = p^i_{j-1} - p^i_{j+1} \quad (۵۰-۳)$$

p^i_{j-1} و p^i_{j+1} به‌ترتیب، فواصل اقلیدسی نمونه ورودی با نورون قبلی و نورون بعدی متناظر با نورون خروجی در شبکه می‌باشد. مشخصه b_i جهت مرتب‌سازی گزینه‌ها در هر کلاس به کار می‌رود. در هر کلاس خواهیم داشت:

$$g'_i > g'_j \qquad if \qquad b_i > b_j \quad (۵۱-۳)$$

استفاده از شبکه عصبی، برای رتبه‌بندی مراجع کاندید دارای مزیت اجرای موازی می‌باشد، بنابراین تأثیر زیادی در کاهش زمان انجام محاسبات دارد. در به‌کارگیری شبکه‌های با ناظر مانند شبکه پرسپترون، مسئله اصلی تعیین نمونه‌های مجموعه آموزش می‌باشد. نتایج حاصل از این نوع شبکه‌ها، وابسته به تعیین نمونه‌های آموزش می‌باشد. چنانچه نمونه‌های آموزش به‌طور مناسب انتخاب شوند، نتایج حاصله مناسب و قابل‌قبول می‌باشد. در تعیین نمونه‌های آموزش در شبکه‌های با ناظر، هر چه بیشتر از اطلاعات مسئله استفاده کنیم، نتایج بهتری حاصل می‌گردد.

در به‌کارگیری شبکه بدون ناظر مانند شبکه کوهونن، خصوصیات شبکه مستقل از مسئله بوده و تعداد نورون‌های لایه خروجی ثابت در نظر گرفته می‌شود. همچنین، مجموعه نقاط ثابت جهت آموزش شبکه استفاده می‌شود. این مجموعه، مستقل از مسئله بوده و به تعداد گزینه‌ها بستگی ندارد. بنابراین می‌توان یک شبکه آموزش داده شده را چندین بار و برای مسائل مختلف به کار گرفت. پس از محاسبه رتبه و امتیاز هر یک از مراجع کاندید برای نامبری مدنظر، توزیع نهایی مراجع کاندید محاسبه می‌شود و زنجیره‌های هم‌ارجاعی نهایی تشکیل می‌شود.

۳-۳ نتیجه‌گیری

در این فصل روش پیشنهادی ارائه شد. روش پیشنهادی از سه مرحله اصلی تشکیل شده است. مراحل اصلی روش پیشنهادی عبارتند از: ۱) بازنمایش دنباله کلمات ۲) بازنمایش دانش و استخراج نامبری‌ها ۳) تشکیل زنجیره‌های هم‌ارجاعی

در گام اول که هدف آن بازنمایش دقیق دنباله کلمات است، از تعبیه واژگان RoBERTa، شبکه Bi-GRU استفاده شد. مهم‌ترین مسئله در پایین بودن دقت سایر روش‌ها در این زمینه، فقدان دانش است. از آنجا که مسئله تحلیل مرجع مشترک دارای ابهامات زیادی است و هر مورد هم‌ارجاعی با دانشی خاص قابل حل است، در روش پیشنهادی سعی شد تا با منابع دانش مختلف نظیر دانش زبان-شناسی، دانش پزشکی، دانش عرفی و غیره، دقت این مسئله بهبود داده شود. در گام دوم با استفاده از بازنمایش منابع دانش مختلف و شبکه عصبی امتیاز نامبری و امتیاز هم‌ارجاعی محاسبه شد و خوشه-های هم مرجع استخراج شدند. در گام سوم از اطلاعات خوشه‌ها و اطلاعات موجودیت‌ها جهت بازنمایش دقیق‌تر دنباله کلمات استفاده شد. در گام چهارم توسط بازنمایش جدید بدست آمده امتیازات هم ارجاعی محاسبه شد و زنجیره‌های هم ارجاعی استخراج شدند. همچنین توسط شبکه-های عصبی کوهونن و پرسپترون رتبه‌بندی مراجع کاندید انجام شد. با استفاده از این شبکه‌ها، بازنمایش دقیق مراجع کاندید و دخیل کردن تمام ویژگی‌های استخراج‌شده در تصمیم‌گیری، رتبه-

بندی مراجع کاندید انجام شد. در فصل بعد کارایی مراحل مختلف روش پیشنهادی بررسی خواهد شد.

فصل ۴ : ارزیابی

۴-۱ مقدمه

فاز ارزیابی پس از فاز یادگیری انجام می‌شود که در آن رده‌بندی کننده‌ای که با داده‌های آموزشی آموزش داده شده است با داده‌های آزمایشی و با انجام مراحل زیر مورد ارزیابی قرار می‌گیرد:

۱. شناسایی و انتخاب مجموعه‌ای از نامبری‌ها.

۲. اعمال سیستم مرجع‌گزینی موردنظر به داده‌های آزمایشی (خروجی این قسمت مجموعه‌ای از عبارات اسمی است که به‌عنوان هم مرجع یا غیر هم مرجع برچسب خورده‌اند)

۳. انجام ارزیابی روی این زنجیره‌ها

لازم به ذکر است که مشابه دیگر مباحث موجود در پردازش زبان طبیعی، ارزیابی سیستم تحلیل هم ارجاعی هم باید بیان‌کننده کارایی آن سیستم باشد و هم در مقایسه با سایر نمونه‌ها کارآمد در این حوزه باید از عملکرد قابل‌قبولی برخوردار باشد. سنجش میزان کارایی یک سیستم کار آسانی نیست و در هنگام ارزیابی یک مدل باید به موارد و پارامترهایی که برای مقایسه استفاده خواهند شد، دقت شود.

۴-۲ تحلیل ارزیابی‌ها و خطاها

یکی از ویژگی‌های مهمی که باعث کاهش کارایی سیستم تعیین مرجع می‌شود، عدم وجود مکانیزم مناسب برای شناسایی موجودیت‌های نامدار است که نتیجه این امر نیز انتشار خطای قسمت‌های پیش‌پردازشی در هم مرجعی می‌گردد. با در نظر گرفتن موارد مطروحه می‌توان گفت: در سیستم‌های تحلیل هم‌ارجاعی، مشکل اصلی در تعریف یک سنجه کارایی مناسب، از مشخص نبودن تعداد کامل موجودیت‌هایی که تا کنون با آن‌ها مواجه شده‌ایم ناشی می‌شود. این مشکل زمانی شدیدتر می‌شود

که نامبری‌های بدست آمده توسط سیستم[1] با نامبری‌های مشخص‌شده توسط استاندارد طلایی [2]منطبق نباشند (اگر نامبری‌ها به‌طور خودکار شناسایی شده باشند). بعلاوه نامبری‌های متفاوتی که توسط شماهای تفسیر متفاوت در نظر گرفته شده‌اند (مثلاً تنها نامبری‌های موجودیت‌های چند نامبری در MUC یا تنها نامبری‌های انواع معنایی خاصی در ACE) تأثیر مستقیمی روی پیچیده شدن ارزیابی یک متن خاص خواهند داشت. بنابراین، نتایج بدست آمده در پیکره‌های مختلف تفاوت زیادی با یکدیگر خواهد داشت.

۴-۳ ارزیابی

مشابه دیگر مباحث موجود در پردازش زبان طبیعی، ارزیابی سیستم تحلیل هم‌ارجاعی نیز علاوه بر ارزیابی کارایی آن می‌بایست مزایای کلی آن را در مقایسه با دیگر نمونه‌های کارآمد آن حوزه مقایسه کند. سنجش میزان کارایی یک سیستم کار آسانی نیست. کوبات [۹۲] و میتکو و هالت [۹۰] ناسازگاری‌هایی که حین ارزیابی نتایج سیستم‌های تحلیل ضمیر به وجود آمده است را گزارش کرده‌اند. آن‌ها مشخص کردند که انواع ضمیرهای مورد بحث در مطالعات مختلف با یکدیگر متفاوت بوده است و همچنین بیشتر الگوریتم‌ها از ویرایش پس از اجرای نتایج بهره برده‌اند که هر یک از این‌ها تأثیرات بسزایی روی نتایج دقت و بازخوانی بدست آمده خواهند داشت. مشکلات مشابه و حتی بدتری (در مقایسه با سیستم تحلیل ضمیر) حین ارزیابی سیستم‌های هم‌ارجاعی وجود خواهد داشت.

۴-۴ سنجه‌های ارزیابی سیستم تحلیل هم‌ارجاعی

خروجی سیستم تحلیل هم‌ارجاعی زنجیره‌های هم‌ارجاعی است که هر زنجیره هم‌ارجاعی حاوی نامبری‌هایی است که با یکدیگر هم‌مرجع هستند. در حقیقت هر زنجیره هم‌ارجاعی معادل یک

[1] system mentions (response mentions)
[2] true mentions

موجودیت درون متن است. با توجه به این نکته طبیعتاً سیستم تحلیل هم‌ارجاعی دارای دقت و کارایی بهتری خواهد بود که بیشترین تعداد زنجیره‌های هم‌ارجاعی را به‌درستی شناسایی کند. متأسفانه تعریف سنجه‌ای که بتواند به سهولت زنجیره‌های شناسایی شده توسط سیستم را با زنجیره-های واقعیِ درست مقایسه کند کار ساده‌ای نیست. به‌عبارت‌دیگر سیستم‌های تحلیل هم‌ارجاعی احتمالاً نمی‌توانند این زنجیره‌ها را به‌صورت صددرصد درست شناسایی کنند و دارای خطاهایی هستند. مثلاً ممکن است دو زنجیره متفاوت را یکی فرض کنند و یا نتوانند یک یا دو یا چند نامبری درون یک زنجیره را با آن اضافه کنند و یا یکی از نامبری‌های یک زنجیره را درون زنجیره دیگری قرار دهند. حال تشخیص اینکه کدام‌یک از این خطاها امتیاز منفی کمتری دارد کار ساده‌ای نیست. سنجه‌هایی که در ادامه می‌آیند، سعی کرده‌اند به روش‌های مختلف سیستم تحلیل هم‌ارجاعی را مورد ارزیابی قرار دهند.

۱-۴-۴ سنجه MUC[1]

قدیمی‌ترین سنجه ارزیابی موجود برای سیستم‌های تحلیل هم‌ارجاعی سنجه MUC [۹۴] است که به‌طور گسترده‌ای توسط سیستم‌های تحلیل هم‌ارجاعی تا کنون استفاده شده است و توسط ویلیان و همکاران معرفی شده است. این سنجه پیونددهی دو نامبری هم‌مرجع به یکدیگر را با یک ربط مشخص کرده و بر مبنای این ربط‌ها کار می‌کند. یعنی ربط‌های خروجی سیستم مورد ارزیابی را در نظر گرفته و مشخص می‌کند کدام درست و کدام نادرست است و بر مبنای آن‌ها محاسبات خود را انجام می‌دهد. به بیان دقیق‌تر با توجه به خروجی سیستم مورد ارزیابی[2] و خروجی سنجه[3] برای محاسبه دقت، تعداد ربط‌های مشترک بین خروجی سیستم مورد ارزیابی و خروجی سنجه را بر تعداد ربط‌های موجود در خروجی سیستم مورد ارزیابی تقسیم می‌کند. برای محاسبه بازخوانی، تعداد ربط-

[1] Message Understanding Conference
[2] response
[3] reference

های مشترک بین خروجی سیستم مورد ارزیابی و خروجی سنجه را بر تعداد ربط‌های موجود در خروجی سنجه تقسیم می‌کند.

علیرغم استفاده گسترده از این سنجه، MUC در سیستم‌هایی با تعداد ربط‌های زیاد نتایج بهتری را گزارش می‌کند. دقت[1] و بازخوانی[2] طبق روابط (۴-۲) و (۴-۳) محاسبه می‌شوند. T (مجموعه درست) مجموعه‌ای از لینک‌های ارجاعی است که به‌صورت دستی مشخص می‌شوند. R (مجموعه پاسخ) نشان‌دهنده خوشه‌های مشخص‌شده توسط سیستم تحلیل هم‌ارجاعی است. دقت با توجه به معادله (۴-۲) زیر محاسبه می‌شود.

$$partition\ (x, y) = \{y | y \in Y \& y \in x \neq \phi\} \quad (۱-۴)$$

$$Precision = \sum_{r \in R} \frac{|r| - |partition(r, T)|}{|r| - 1} \quad (۲-۴)$$

در رابطه (۴-۲)، $|partition(r, T)|$ تعداد خوشه‌هایی از T است که با R مشابه نیست. $|r| - |partition(r, T)|$ تعداد ربط‌های مشترک بین T و R است. $|r| - 1$ تعداد ربط‌های پاسخ است. بازخوانی طبق رابطه (۴-۳) محاسبه می‌شود.

$$Recall = \sum_{t \in T} \frac{|t| - |partition(t, R)|}{|t| - 1} \quad (۳-۴)$$

در رابطه (۴-۳)، $|partition(t, R)|$ تعداد خوشه‌هایی از R را نشان می‌دهد که با | t | مشابه نیست. $|t| - |partition(t, R)$ تعداد ربط‌های مشترک بین T و R است. $|t| - 1$ تعداد ربط‌های T است. به زبان ساده، دقت و بازخوانی طبق روابط (۴-۴) و (۴-۵) محاسبه می‌شوند.

$$\text{دقت} = \frac{\text{تعداد ربط‌های مشترک در پاسخ و مرجع}}{\text{تعداد ربط‌های موجود در پاسخ}} \quad (۴-۴)$$

$$\text{بازخوانی} = \frac{\text{تعداد ربط‌های مشترک در پاسخ و مرجع}}{\text{تعداد ربط‌های موجود در مرجع}} \quad (۵-۴)$$

[1] precision
[2] recall

۴-۴-۲ سنجه BCUB

سنجه B^3 [۹۵] دقت و بازخوانی را برای هر نامبری درون هر موجودیت به‌صورت مجزا محاسبه می‌کند و به همین دلیل تا حد زیادی مشکل سنجه MUC در تمایل به سمت موجودیت‌های حاوی نامبری‌های زیاد را حل کرده است. برای هر نامبری m در یک موجودیت، سنجه B^3 مقدار بازخوانی را با تقسیم تعداد نامبری‌های درست در آن موجودیت (با توجه به داده‌های طلایی) بر تعداد کل نامبری‌ها بدست می‌آورد. برای محاسبه دقت به‌طور مشابه تعداد نامبری‌های مشترک بین موجودیت‌های درست و موجودیت‌های سیستمی در صورت تقسیم قرار می‌گیرد با این تفاوت که این بار در مخرج کسر تعداد نامبری‌های موجودیت پاسخ قرار می‌گیرد. دقت و بازخوانی طبق سنجه B^3 به‌صورت روابط (۴-۶ الی ۴-۹) محاسبه می‌شوند. که در آن‌ها NTR_{e_i} تعداد عناصر درست در زنجیره خروجی موجودیت i، NR_{e_i} تعداد عناصر در زنجیره خروجی موجودیت i و NT_{e_i} تعداد عناصر در زنجیره طلایی موجودیت i است. w_i وزن اختصاص‌یافته به موجودیت i در سند است. معمولاً وزن‌ها $\frac{1}{N}$ در نظر گرفته می‌شوند.

$$Precision_i = \frac{NTR_{e_i}}{NR_{e_i}} \quad (۴-۶)$$

$$Recall_i = \frac{NTR_{e_i}}{NT_{e_i}} \quad (۴-۷)$$

$$FinalPrecision = \sum_{i=1}^{N} w_i * Precision \quad (۴-۸)$$

$$FinalRecall = \sum_{i=1}^{N} w_i * Recall \quad (۴-۹)$$

۴-۴-۳ سنجه CEAF

سنجه CEAF [۹۶] ارزیابی را با ایجاد تناظر یک‌به‌یک بین موجودیت‌های خروجی سیستم و استاندارد طلایی انجام می‌دهد. به‌منظور ایجاد بهترین تناظر از سنجه‌های تشابه بین موجودیت‌ها استفاده می‌شود. دو سنجه مورد استفاده عبارتند از سنجه مبتنی بر نامبری (CEALM) و سنجه

مبتنی بر موجودیت (CEALE). در سنجه CEALM تعداد نامبری‌های مشترک بین موجودیت را محاسبه می‌کند. در CEALE تعداد نسبی نامبری‌های مشترک درون موجودیت را محاسبه می‌کند. برای محاسبه بازخوانی در این سنجه مجموع شباهت‌ها بر تعداد نامبری‌های داده‌های طلایی تقسیم می‌شود. برای محاسبه دقت مجموع شباهت‌ها بر تعداد نامبری‌ها در خروجی سیستم تقسیم می‌شود. دقت و بازخوانی طبق سنجه CEAF به‌صورت زیر محاسبه می‌شوند. این معیار از سنجه‌های شباهت برای ایجاد یک نگاشت بین موجودیت‌های خروجی سیستم (خوشه‌های (R) شامل موجودیت‌های پاسخ) و موجودیت‌های طلایی (خوشه‌های (T) شامل موجودیت‌های درست) استفاده می‌کند. دقت و بازخوانی بر اساس این ایده محاسبه می‌شود. چهار تابع مختلف اندازه‌گیری شباهت طبق روابط (۴-۱۰ الی ۴-۱۳) تعریف می‌شود. $\emptyset_1(T,R)$ اگر نامبری‌های دو موجودیت یکسان باشند، آن‌ها را یکسان در نظر می‌گیرد، $\emptyset_2(T,R)$ در صورت وجود حداقل یک نامبری مشترک، دو موجودیت را مشابه در نظر می‌گیرد. $\emptyset_3(T,R)$ شباهت را بر اساس تعداد نامبری‌های مشترک بین دو موجودیت محاسبه می‌کند و $\emptyset_4(T,R)$ مقدار F بین T (موجودیت‌های درست) و R (موجودیت‌های پاسخ) است.

$$\emptyset_1(T,R) = \begin{cases} 1, & if\ R = T \\ 0, & otherwise \end{cases} \quad (۱۰\text{-}۴)$$

$$\emptyset_2(T,R) = \begin{cases} 1, & if\ R \cap T \neq \emptyset \\ 0, & otherwise \end{cases} \quad (۱۱\text{-}۴)$$

$$\emptyset_3(T,R) = |R \cap T| \quad (۱۲\text{-}۴)$$

$$\emptyset_4(T,R) = 2.\frac{|R \cap T|}{|R| + |T|} \quad (۱۳\text{-}۴)$$

دقت و بازخوانی طبق روابط (۴-۱۴ و ۴-۱۵) محاسبه می‌شوند. تابع $m(r)$ خوشه r را گرفته و خوشه t را برمی گرداند.

$$CEAF_{\emptyset_i}Precision = max_m \frac{\sum_{r \in R} \emptyset_i(r, m(r))}{\sum_{r \in R} \emptyset_i(r,r)} \quad (۱۴\text{-}۴)$$

$$CEAF_{\emptyset_i}Recall = max_m \frac{\sum_{r \in R} \emptyset_i(r, m(r))}{\sum_{r \in R} \emptyset_i(t,t)} \quad (۱۵\text{-}۴)$$

۴-۴-۴ سنجه BLANC

این سنجه [۹۷] از ایده Rand index که سنجه‌ای برای ارزیابی روش‌های خوشه‌بندی است، استفاده می‌کند و دقت، بازخوانی و سنجه F را برای هر ربط هم‌ارجاعی و غیر هم‌ارجاعی به‌طور جداگانه محاسبه می‌کند. ربط‌های غیر هم‌ارجاعی ربط‌های بین نامبری‌های غیر هم‌مرجع هستند. دقت و بازخوانی به ترتیب به‌صورت تعداد ربط‌های درست نسبی نسبت به کل ربط‌های خروجی سیستم و ربط‌های داده‌های استاندارد طلایی محاسبه می‌شوند. دقت و بازخوانی نهایی طبق روابط (۴-۲۰ و ۴-۲۱) با میانگین گرفتن روی همه دقت و بازخوانی محاسبه‌شده بدست می‌آید. در این روابط C_k مجموعه موجودیت طلایی (درست) و C_r مجموعه موجودیت پاسخ (خروجی سیستم) است. R_c و R_n به‌ترتیب مربوط به بازخوانی پیوندهای هم‌ارجاعی و غیر هم‌ارجاعی هستند. دقت نیز مشابه همین علائم تعریف می‌شود.

$$P_n = \frac{|N_k \cap N_r|}{|N_r|} \quad (۱۶-۴)$$

$$R_c = \frac{|C_k \cap C_r|}{|C_k|} \quad (۱۷-۴)$$

$$P_c = \frac{|C_k \cap C_r|}{|C_r|} \quad (۱۸-۴)$$

$$R_n = \frac{|N_k \cap N_r|}{|N_k|} \quad (۱۹-۴)$$

$$Recall = \frac{R_c + R_n}{2} \quad (۲۰-۴)$$

$$Precision = \frac{P_c + P_n}{2} \quad (۲۱-۴)$$

۴-۴-۵ سنجه LEA[1]

این سنجه [۹۸] به دو اصطلاح "اهمیت[2]" و "امتیاز تحلیل[3]" وابسته است. "اهمیت" به اندازه موجودیت بستگی دارد و "امتیاز تحلیل" بر اساس شباهت پیونددهی محاسبه می‌شود. تابع پیونددهی، تعداد کل پیوندهای ممکن بین n نامبری و موجودیت e را برمی‌گرداند.

$$\text{importance}(e_i) = |e_i| \quad (۲۲-۴)$$

$$\text{resoution} - \text{score}(k_i) = \sum_{r_j \in R} \frac{\text{link}(k_i \cap r_j)}{\text{link}(k_i)} \quad (۲۳-۴)$$

$$\text{Recall} = \frac{\sum_{k_i \in K} \text{importance}(k_i) * \sum_{r_j \in R} \frac{\text{link}(k_i \cap r_j)}{\text{link}(k_i)}}{\sum_{k_z \in K} \text{importance}(k_z)} \quad (۲۴-۴)$$

$$\text{Precision} = \frac{\sum_{k_i \in R} \text{importance}(r_i) * \sum_{k_j \in R} \frac{\text{link}(r_i \cap k_j)}{\text{link}(r_i)}}{\sum_{r_z \in K} \text{importance}(r_z)} \quad (۲۵-۴)$$

در معادلات بالا r_i مجموعه پاسخ و k_i مجموعه طلایی است.

۴-۵ پیکره‌های در نظر گرفته‌شده

۱. پیکره CoNLL-2012 shared task [۹۹]: این پیکره از جمله پیکره‌های استاندارد تحلیل هم‌ارجاعی برای چند زبان (انگلیسی، چینی و عربی) است که در بسیاری از مقالات و کارهای تحقیقاتی مورد استفاده قرار می‌گیرد. همچنین وظیفه شناسایی همه نامبری‌ها از رخدادها و موجودیت‌ها در متن و خوشه‌بندی آن‌ها در کلاس‌های هم‌ارز را نیز بر عهده دارد و به‌خوبی در جامعه پردازش زبان طبیعی به رسمیت شناخته شده است. در این مجموعه اسناد، موارد هم‌ارجاعی به‌صورت دستی مشخص شده‌اند. پیکره انگلیسی شامل ۳۴۹۳ سند در سه قسمت Train (۲۸۰۲ سند، هر سند شامل حداکثر ۴۰۰۹ کلمه)، Test (۳۴۸ سند)

[1] link-based Entity-Aware
[2] importance
[3] resolution-score

و Dev (۳۴۳ سند) است که از پیکره متنی بزرگ OntoNotes 5.0 شامل ۲ میلیون و ۹۰۰ کلمه، اطلاعات مربوط به اخبار، مکالمات، گفتگوهای تلفنی، اطلاعات وب، گروه‌های خبری و مصاحبه‌های تلویزیونی را در نظر گرفته و سبب گردیده نتایج حاصل از استفاده این پیکره دارای استاندارد قابل‌قبولی باشد.

۲. پیکره i2b2 [۱۰۰]: این پیکره شامل مجموعه داده‌های پرونده پزشکی الکترونیکی از دو سازمان مختلف Partners HealthCare (Part) و مرکز پزشکی Deaconess Beth Israel (Beth) است. تمام رکوردها سوابق به‌طور کامل بازنویسی شده و موارد هم‌ارجاعی به‌صورت دستی حاشیه‌نویسی شده‌اند.

۳. پیکره MUC6 [۱۰۱]: این پیکره توسط کمیته اطلاعات زبان‌شناسی (LDC) تهیه شده است و حاوی ۳۱۸ مقاله از مجله وال‌استریت، نرم‌افزار امتیازدهی و مستندات مرتبط با آن به زبان انگلیسی است.

۴. پیکره English Gigaword [۱۰۲]: این پیکره برای ارزیابی قدرت تشخیص موجودیت‌های نامدار در نظر گرفته شده است. یک آرشیو جامع از متون خبری انگلیسی می‌باشد و طی چندین سال در دانشگاه پنسیلوانیا به دست آمده است.

۶-۴ منابع دانش استفاده شده

همان‌طور که گفته شد، رویکرد اصلی روش پیشنهادی استفاده از منابع دانش تحت فرمت سه‌تایی ("ابتدا" (لیستی از کلمات)، رابطه بین ابتدا و انتها، "انتها" (لیستی از کلمات)) است. منابع دانش استفاده شده برای آموزش سیستم عبارتند از:

۱. منبع دانش عرفی (حس عام) OMCS[1] [۱۰۳]: از اولین سال‌های ظهور هوش مصنوعی، متخصصان این حوزه بر این مهم واقف بودند که هر سیستم هوشمند، نیازمند دسترسی به

[1] commonsense knowledge graph

دانش عرفی است. برای ما انسان‌ها دانش عرفی تداعی‌کننده قضاوت صحیح است. درحالی‌که "دانش عرفی "در حوزه زبان‌شناسی رایانشی اشاره به میلیون‌ها حقیقت کوچک و واضح دارد که برای انسان موضوعاتی پیش پا افتاده به‌حساب می‌آیند. جملاتی مانند "سوزن تیز است"، "برای باز کردن در ابتدا باید دستگیره را چرخاند "و"اگر روز تولد کسی را فراموش کنید، از دستتان ناراحت خواهد شد" نمونه‌های این دانش هستند. دانش عرفی، گستره وسیعی از دانش دنیای پیرامون را شامل می‌شود و جوانبی از دانش فضایی، فیزیکی، اجتماعی را پوشش می‌دهد. این نوع دانش معمولاً در ارتباطات اجتماعی محذوف است. برای داشتن درکی عمیق از اطلاعات متنی، رایانه‌ها نیازمند حجم گسترده‌ای از این نوع دانش هستند که در حال حاضر فقط در اختیار انسان است. هدف ما استخراج و تزریق این نوع دانش به سیستم است.

پروژه سینق [۱۰۳] یک پروژه بزرگ در زمینه دانش عرفی است که برخلاف Cyc که بانک اطلاعاتی آن توسط مهندسان دانش جمع‌آوری گشته، به کاربران معمولی اینترنت متکی بوده و این کاربران بودند که این نوع جملات را به پایگاه داده‌ی آن افزوده‌اند. این پروژه ۱/۶ میلیون عبارت عرفی جمع‌آوری کرده است، که ۷۰۰۰۰۰ جمله آن از طریق پایگاه داده Conceptnet قابل دسترسی است. در Conceptnet دانش عرفی در قالب جملات زبان طبیعی ذخیره می‌شود، از این رو بازنمایی روابط در قالبی صوری کار مشکلی خواهد بود. روش بهینه دیگر استفاده از ابزارهای پردازش زبان طبیعی جهت استخراج چنین حقایقی از منابعی همچون ویکی پدیاست که پایگاهی جامع از اطلاعات دنیای پیرامون ماست و به‌طور منظم توسط کاربرانش به‌روز رسانی و کنترل می‌شود. این منبع شامل ۶۰۰ هزار سه‌تایی است، به‌عنوان مثال (food, UsedFor, eat). تمام روابط در این منبع به‌صورت دستی مشخص شده است. در این پژوهش سه‌تایی‌های با امتیاز بالا (بالاتر از ۳) انتخاب شده است که تعداد آن‌ها ۶۱۶۷۳ سه‌تایی است.

۲. منبع دانش پزشکی i2b2 [۱۰۰]: این منبع شامل اطلاعات پزشکی است به‌عنوان مثال سه-

تایی (the CT scan, is, test). این منبع شامل ۲۲۲۳۴ سه‌تایی است.

۳. ویژگی‌های زبان‌شناسی: علاوه بر منابع دانش دست‌نویس که به آن‌ها اشاره شد، ویژگی‌های زبان‌شناسی نظیر جاندار بودن، جنسیت نیز در نظر گرفته شده است. تجزیه‌کننده[1] استنفورد برای تولید تعداد، جانداری و جنسیت برای همه عبارات اسمی به کار می‌رود، که می‌تواند به‌طور خودکار دانش زبانی (به‌صورت سه‌تایی) برای داده‌های ما تولید کند. ویژگی تعداد شامل مقدار s برای مفرد و مقدار p برای جمع است. ویژگی‌های جانداری نشان‌دهنده جانداری به‌صورت s یا p است که در صورت جانداری، مقدار مرد، زن و خنثی بودن برای ویژگی جنسیت استفاده می‌شود. برای مثال نامبری ‘the boys’ به‌عنوان جمع و مرد برچسب‌گذاری می‌شود که برای نمایش از سه‌تایی‌های (‘the boys’, plurarity, Plural) و (‘the boys’, AG, male) استفاده می‌شود. درنتیجه ۴۰۱۴۹ سه‌تایی برای جمع بودن و ۴۰۴۶۲ سه‌تایی برای جنسیت و جانداری در نظر گرفته می‌شود.

۴. دانش لغوی [۱۰۴]: این نوع دانش می‌تواند، در بسیاری از کارهای پردازش زبان طبیعی مانند تشخیص خطای معنایی، تشخیص استعاره، ابهام زدایی از حس کلمه، تجزیه نحوی و برچسب زن نقش معنایی به کار رود. برای این نوع دانش، ابتدا با استفاده از تجزیه‌کننده استنفورد اطلاعات پایگاه ویکی‌پدیا انگلیسی را تجزیه نمودیم و تمام لبه‌های وابستگی را با فرمت (گزاره[2]، آرگومان[3]، رابطه، شماره) استخراج شد. در تجزیه‌کننده استنفورد، هنگامی‌که فعل یک فعل ارتباطی (مانند am، is) است، بین پیش‌بینی‌کننده (گزاره) و موضوع، لبه ”nsubj“ ایجاد می‌شود. هر جفت در این منبع دانش با احتمال رابطه (۴-۱۹) اندازه-گیری می‌شود.

[1] Parser
[2] predicate
[3] argument

$$P_r(a|p) = \frac{Count_r(p,a)}{Count_r(p)} \qquad (۲۶-۴)$$

در رابطه (۴-۲۶)، $Count_r(p)$ و $Count_r(p,a)$ به‌ترتیب نشان‌دهنده این هستند که چند بار p و جفت گزاره-آرگومان (p,a) در رابطه r وجود داشته‌اند. در آزمایش‌ها در صورت $P_r(a|p) > 0.1$ و $Count_r(p,a) > 10$، سه‌تایی (p,r,a) را در نظر گرفتیم. به‌عنوان مثال، (‘cat’, nsubj, ‘meow’) یک رابطه معتبر است. سرانجام، دو رابطه nsubj و dobj برای این منبع دانش انتخاب شدند که به‌ترتیب شامل ۱۷۰۷۴ و ۴۵۳۶ جفت گزاره- آرگومان برای nsubj و dobj می‌باشد.

۴-۷ نتایج

در این بخش روش پیشنهادی از جنبه‌های مختلف مورد بررسی قرار گرفته و با سایر روش‌ها مقایسه شده است.

۴-۷-۱ دقت، بازخوانی و F1

در جدول ۴-۱ روش پیشنهادی طبق معیارهای MUC، B^3 و $CEAF_{\phi 4}$ توسط مقادیر دقت، بازخوانی و Avg.F1 (میانگین خروجی سه معیار MUC، B^3 و $CEAF_{\phi 4}$) با روش‌های مطرح و جدید تحلیل هم‌ارجاعی مقایسه شده است. همان‌طور که در جدول ۴-۱ مشاهده می‌شود، روش پیشنهادی طبق همه معیارها بر روش‌های قبلی برتری دارد و Avg.F1 به‌میزان ۳/۷ درصد بهبود یافته است. دلیل این برتری، تعبیه واژگان با استفاده از شبکه از پیش آموزش‌دیده RoBERTa و استفاده از منابع دانش مختلف است. با در نظر گرفتن تعبیه واژگان RoBERTa و اطلاعات موجودیت‌ها اطلاعات معنایی بهتری در اختیار سیستم قرار گرفته و سیستم توانایی بالاتری در تشخیص درست نامبری‌ها پیدا می‌کند. همچنین توسط ساختار شبکه عصبی بازگشتی دوجهته ویژگی‌های مفیدتری توسط شبکه برای نامبری‌ها استخراج شده و باعث می‌شود نامبری‌هایی که در روش‌های دیگر نادیده گرفته

شده‌اند، در روش پیشنهادی در نظر گرفته شوند. زیرا در استخراج کلمه مهم در دنباله کلمات و استخراج دانش مناسب برای هر یک از موارد هم‌ارجاعی از مکانیزم توجه استفاده شده است. علاوه بر این، استفاده از منابع دانش مختلف باعث شده است دانش سیستم نسبت به موارد مختلف هم‌ارجاعی بالا رود و توانایی بیشتری در یافتن زنجیره‌های هم‌مرجع پیدا کند.

جدول ۴-۱. نتایج ارزیابی روش پیشنهادی روی مجموعه تست پیکره انگلیسی CoNLL-2012 shared task

روش	MUC			B^3			$CEAF_{\emptyset 4}$			
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Avg.F1
[۸۲]	۸۱/۴	۷۹/۵	۸۰/۴	۷۲/۲	۶۹/۵	۷۰/۸	۶۸/۲	۶۷/۱	۶۷/۶	۷۳/۰
[۸۳]	۸۵/۴	۷۷/۹	۸۱/۴	۷۷/۹	۶۶/۴	۷۷/۷	۷۰/۶	۶۶/۳	۶۸/۴	۷۳/۸
[۱۰۵]	۸۲/۶	۸۳/۴	۸۳/۰	۷۳/۳	۷۶/۱	۷۴/۷	۷۲/۳	۷۱/۱	۷۱/۷	۷۶/۶
[۸۴]	۸۴/۷	۸۲/۴	۸۳/۵	۷۶/۵	۷۴/۰	۷۵/۳	۷۴/۱	۶۹/۸	۷۱/۹	۷۶/۹
[۱۰۶]	۸۵/۸	۸۴/۸	۸۵/۳	۷۸/۳	۷۷/۹	۷۸/۱	۷۶/۴	۷۴/۲	۷۵/۳	۷۹/۶
[۱۰۷]	۸۵/۹	۸۵/۵	۸۵/۷	۷۹/۰	۷۸/۹	۷۹/۰	۷۶/۷	۷۵/۲	۷۵/۹	۸۰/۲
روش پیشنهادی	۹۰/۵	۸۷/۹	۸۹/۲	۸۴/۶	۸۱/۷	۸۳/۱	۷۸/۲	۷۷/۱	۷۹/۳	۸۳/۹

۴-۷-۲ مقایسه دقت روش‌های رتبه‌بندی

در جدول ۴-۲ روش‌های رتبه‌بندی ارائه‌شده با سایر روش‌های رتبه‌بندی [۴۶، ۴۷، ۴۹ و ۸۳]، مقایسه شده است. وایزمن و همکاران [۴۶] سعی کرده‌اند با تزریق ویژگی‌های مختلف به سیستم دقت رتبه‌بندی را افزایش دهند. این ویژگی‌ها باید وزن‌دهی شوند، تا اهمیت هر یک از آن‌ها بر اساس نامبری مدنظر مشخص شود. روش‌های رتبه‌بندی مبتنی بر یادگیری عمیق [۴۶ و ۴۷] از توابع هزینه ابتکاری برای رتبه‌بندی مراجع کاندید استفاده می‌کنند که آموزش را پیچیده می‌کند. این روش‌ها دارای پارامترهای بسیاری برای آموزش هستند. روش‌های رتبه‌بندی مبتنی بر یادگیری تقویتی [۴۹ و

۸۳] نیز معایبی دارند. این روش‌ها به‌شدت به پارامترها وابسته هستند و تنظیم پارامترها فرآیندی دشوار است. از طرف دیگر، چون این پارامترها با آزمون و خطا بدست می‌آیند، ممکن است مقادیر بهینه برای آن‌ها بدست نیاید. بهتر است این مقادیر به‌طور خودکار و با استفاده از یک منطق خاص بدست آیند. همچنین، نادیده گرفتن رابطه بین حالت‌ها باعث از دست رفتن اطلاعات می‌شود. هر حالت نیاز به یادگیری به‌صورت جداگانه دارد و در نتیجه، زمان زیادی برای یادگیری صرف می‌شود. مدل رتبه‌بندی مبتنی بر شبکه عصبی کوهونن دقت عملکرد بهتری نسبت به شبکه عصبی پرسپترون جهت رتبه‌بندی دقیق‌تر مراجع کاندید دارد. استفاده از شبکه عصبی، برای رتبه‌بندی مراجع کاندید دارای مزیت اجرای موازی می‌باشد، بنابراین تأثیر زیادی در کاهش زمان انجام محاسبات دارد.

جدول ۴-۲. مقایسه روش‌های رتبه‌بندی روی مجموعه تست پیکره انگلیسی CoNLL-2012 shared task

Method	**MUC**			**B^3**			**$CEAF_{\emptyset 4}$**			
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Avg.F1
رتبه‌بندی مبتنی بر شبکه کوهونن	۸۹/۳	۸۵/۱	۸۷/۱	۸۱/۴	۷۸/۳	۷۹/۸	۷۶/۱	۷۶/۵	۷۶/۳	۸۱/۰۶
رتبه‌بندی مبتنی بر شبکه پرسپترون	۸۸/۲	۸۴/۸	۸۶/۵	۸۰/۹	۷۷/۹	۷۹/۴	۷۶/۸	۷۵/۷	۷۶/۲	۸۰/۷
[۴۶]	۷۷/۵	۶۹/۷	۷۳/۴	۶۶/۸	۵۶/۹	۶۱/۵	۶۲/۱	۵۳/۸	۵۷/۷	۶۴/۲
[۴۷]	۷۶/۲	۶۹/۳	۷۲/۶	۶۶/۱	۵۵/۸	۶۰/۵	۵۹/۴	۵۴/۹	۵۷/۱	۶۳/۴
[۸۳]	۸۵/۴	۷۷/۹	۸۱/۴	۷۷/۹	۶۶/۴	۷۷/۷	۷۰/۶	۶۶/۳	۶۸/۴	۷۳/۸
[۴۹]	۷۹/۲	۷۰/۴	۷۴/۶	۶۹/۹	۵۸/۰	۶۳/۴	۶۳/۵	۵۵/۵	۵۹/۲	۶۵/۷

در به‌کارگیری شبکه بدون ناظر مانند شبکه کوهونن، خصوصیات شبکه مستقل از مسئله بوده و تعداد نورون‌های لایه خروجی ثابت در نظر گرفته می‌شود. همچنین، مجموعه نقاط ثابت جهت آموزش شبکه استفاده می‌شود. این مجموعه، مستقل از مسئله بوده و به تعداد گزینه‌ها بستگی ندارد. بنابراین می‌توان یک شبکه آموزش داده شده را چندین بار و برای مسائل مختلف به کار گرفت. در به‌کارگیری شبکه-

های با ناظر مانند شبکه پرسپترون، مسئله اصلی تعیین نمونه‌های مجموعه آموزش می‌باشد. نتایج حاصل از این نوع شبکه‌ها، وابسته به تعیین نمونه‌های آموزش می‌باشد. چنانچه نمونه‌های آموزش به‌طور مناسب انتخاب شوند، نتایج حاصله مناسب و قابل‌قبول می‌باشد. در تعیین نمونه‌های آموزش در شبکه‌های با ناظر، هر چه بیشتر از اطلاعات مسئله استفاده کنیم، نتایج بهتری حاصل می‌گردد.

۴-۷-۳ بررسی تأثیر روش تعبیه واژگان

برای بررسی میزان تأثیر روش‌های مختلف تعبیه واژگان نظیر RoBERTa، ELMo، BERT و XLNet بر عملکرد روش پیشنهادی و مقدار Avg.F1، این روش‌ها در جدول ۴-۳ روی مجموعه توسعه پیکره CoNLL-2012 shared task با یکدیگر مقایسه شده‌اند. طول بردارهای استخراج شده و درنظر گرفته شده برای روش‌های ELMo، BERT، XLNet و RoBERTa به ترتیب برابر ۱۰۲۴، ۳۰۷۲، ۱۵۳۶ و ۷۶۸ می‌باشد. همان‌طور که مشخص است کارکرد روش BERT از روش ELMo بهتر است. روش ELMo در ساختار خود از مدل‌های زبانی دوجهته و شبکه BiLSTM استفاده کرده و ویژگی‌های نحوی و معنایی و زبان‌شناسی را در اختیار سیستم قرار می‌دهد. BERT در دو مرحله کار می‌کند، ابتدا، از مقدار زیادی از داده‌های بدون برچسب استفاده می‌کند تا بازنمایش یک زبان را به روشی بدون نظارت به نام پیش آموزش، یاد بگیرد. سپس، مدل از قبل آموزش‌دیده می‌تواند با استفاده از مقدار کمی از داده‌های آموزش‌دیده برچسب زده، به‌صورت دقیق برای انجام کارهای مختلف تنظیم شود. عملکرد بهتر BERT دو دلیل دارد. اولاً، یک مدل پیش‌آموزش جدید به نام مدل پوشش زبان (MLM[1]) و مکانیزم پیش‌بینی جمله بعدی (NSP[2]) ارائه داده است. دوماً، برای آموزش BERT از تعداد زیادی داده و قدرت محاسباتی بالا استفاده شده است. MLM امکان یادگیری دو طرفه از متن را ممکن می‌سازد، یعنی به مدل اجازه می‌دهد تا مفهوم هر کلمه را از کلماتی که قبل و بعد از آن ظاهر می‌شود، بیاموزد. این کار در روش ELMo امکان‌پذیر نیست، زیرا از مدل دو طرفه

[1] Masked Language Model
[2] Next Sentense Prediction

کم‌عمق استفاده می‌کند. استفاده از تعبیه واژگان XLNet عملکرد بهتری نسبت به روش BERT دارد. XLNet یک مبدل دوطرفه بزرگ است که از روش آموزش پیشرفته، داده‌های بزرگ‌تر و قدرت محاسباتی بیشتر برای دستیابی بهتر از به معیارهای پیش‌بینی BERT در بیست وظیفه زبانی استفاده می‌کند. برای بهبود آموزش، XLNet مدل‌سازی زبانی جایگشتی[1] را معرفی می‌کند، که در آن همه توکن‌ها به‌صورت تصادفی پیش‌بینی می‌شوند. این برخلاف مدل MLM در BERT است، که در آن فقط پانزده درصد توکن‌های پوشش‌دار پیش‌بینی شده است. همچنین برخلاف مدل‌های زبانی سنتی است که در آن‌ها همه توکن‌ها به‌جای روش تصادفی، به‌صورت ترتیبی پیش‌بینی می‌شدند. این روش به مدل کمک می‌کند تا روابط دوطرفه را بیاموزد، بنابراین وابستگی و روابط بین کلمات بهتر درک می‌کند. علاوه بر این، مبدل XL به‌عنوان معماری پایه مورد استفاده قرار گرفته است، که حتی در صورت عدم استفاده از آموزش مبتنی بر جایگشت، عملکرد خوبی نیز دارد. در نهایت روش RoBERTa بهتر از تمام روش‌های تعبیه واژگان عمل می‌کند، زیرا RoBERTa برای بهبود روند آموزش، وظیفه NSP را از پیش‌آموزش BERT حذف کرده است. همچنین این روش از آموزش دسته‌ای بزرگ‌تر استفاده کرده است. علاوه بر این، روش RoBERTa معماری روش XLNet را بهبود داده و توانسته است ویژگی‌های بهتری نسبت به این روش ارائه دهد. این روش در اکثر زمینه‌های پردازش متن نتایج قابل‌قبولی ارائه می‌کند.

جدول ۴-۳. نتایج ارزیابی تعبیه واژگان روی مجموعه توسعه پیکره انگلیسی CoNLL-2012 shared task

	Avg. F1	**Δ**
مدل پیشنهادی (تعبیه واژه RoBERTa)	۸۳/۶	
تعبیه واژه XLNet	۸۰/۱	-۳/۵
تعبیه واژه BERT	۷۷/۳	-۶/۳
تعبیه واژه ELMo	۷۵/۲	-۸/۴

[1] permutation language modeling

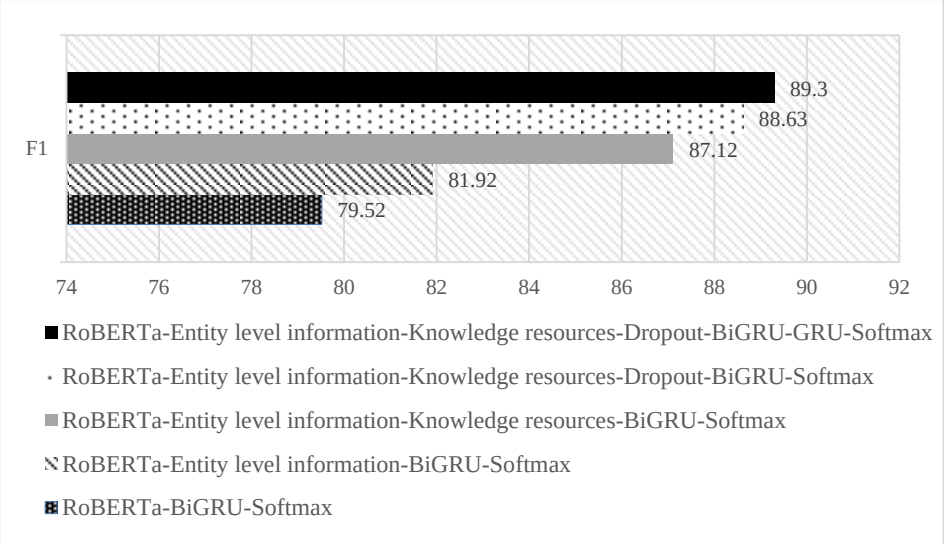
۴-۷-۴ قدرت تشخیص موجودیت‌های نامدار

در شکل ۴-۱ جهت سنجش عملکرد افزودن قابلیت‌های مختلف به سیستم در زمینه تشخیص موجودیت‌های نامدار، امتیاز F1 روی پیکره توسعه CoNLL-2012 shared task نمایش داده شده است. دقت و بازخوانی طبق روابط (۴-۲۷) و (۴-۲۸) محاسبه می‌شود. با توجه به شکل ۴-۱ مشخص است، که با افزودن بردار تعبیه‌ی واژگان، اطلاعات موجودیت‌ها، منابع دانش، بیش‌برازش و یک لایه اضافه GRU دقت تشخیص موجودیت‌ها و در نتیجه F1 افزایش یافته است.

$$\text{دقت} = \frac{\text{تعداد موجودیت‌های درست استخراج شده}}{\text{تعداد کل موجودیت‌های استخراج شده}} \quad (۴-۲۷)$$

$$\text{بازخوانی} = \frac{\text{تعداد موجودیت‌های درست استخراج شده}}{\text{تعداد کل موجودیت‌های درست}} \quad (۴-۲۸)$$

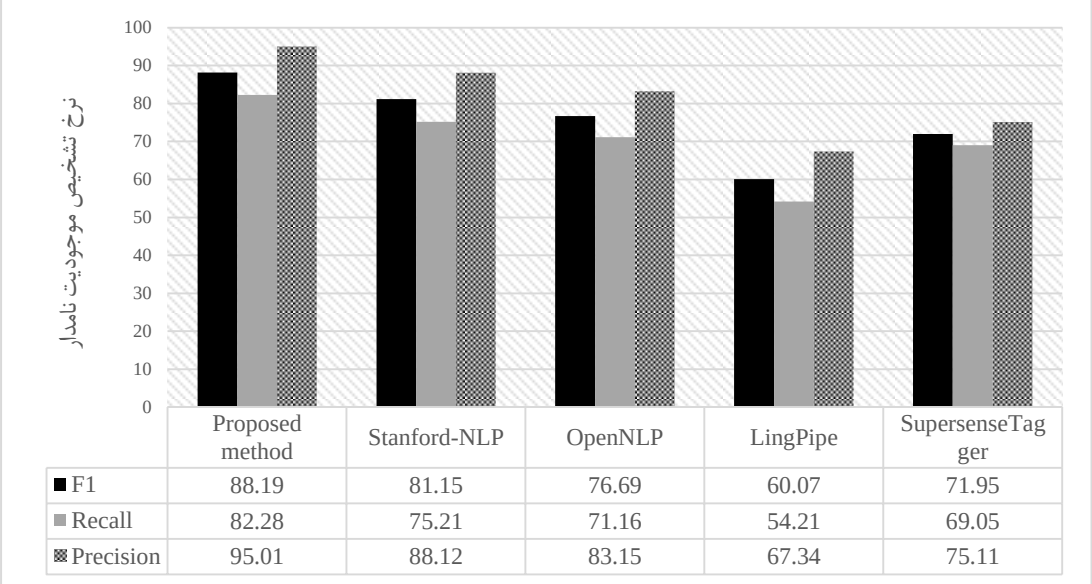
هرچه میزان اطلاعاتی که از ظاهر، معنا و ساختار واژگان در قالب بردارهای بازنمایش به شبکه وارد می‌شود، افزایش یابد، دقت نیز افزایش پیدا می‌کند و به‌تبع آن F1 نیز بیشتر می‌گردد. افزایش میزان اطلاعات از محتوای متنی که واژگان در آن قرار گرفته است با استفاده از Bi-GRU و یک لایه GRU بر روی خروجی آن، موجب افزایش بازخوانی می‌گردد و به‌تبع آن F1 نیز افزایش می‌یابد.



شکل ۴-۱. F1 در حالت‌های مختلف روش پیشنهادی روی پیکره توسعه CoNLL-2012 shared task

جهت ارزیابی عملکرد روش پیشنهادی در شناسایی موجودیت‌های نامدار، آن را با روش‌های موجود[1] در این زمینه نیز مقایسه کردیم. دقت و بازخوانی توسط روابط (۴-۲۷) و (۴-۲۸) محاسبه شده است. از آنجا که در روش پیشنهادی از اطلاعات موجودیت‌ها و منابع دانش مختلف تحت ساختار شبکه عصبی از پیش آموزش‌دیده در شناسایی نامبری‌ها استفاده شده و این نوع شبکه قابلیت استخراج خودکار ویژگی‌ها از موجودیت‌ها را دارد؛ عملکرد بهتری در مقایسه با سایر روش‌ها دارد.

همان‌گونه که در شکل ۴-۲ نشان داده شده است، Avg.F1 برای روش پیشنهادی در مقایسه با روش‌های قبلی، ۷/۰۴ درصد بهبود یافته است؛ زیرا استفاده از مدل‌های زبانی توسط روش پیش-آموزش استفاده شده و تعبیه درست واژگان، اطلاعات معنایی خوبی رو در اختیار سیستم قرار داده است.



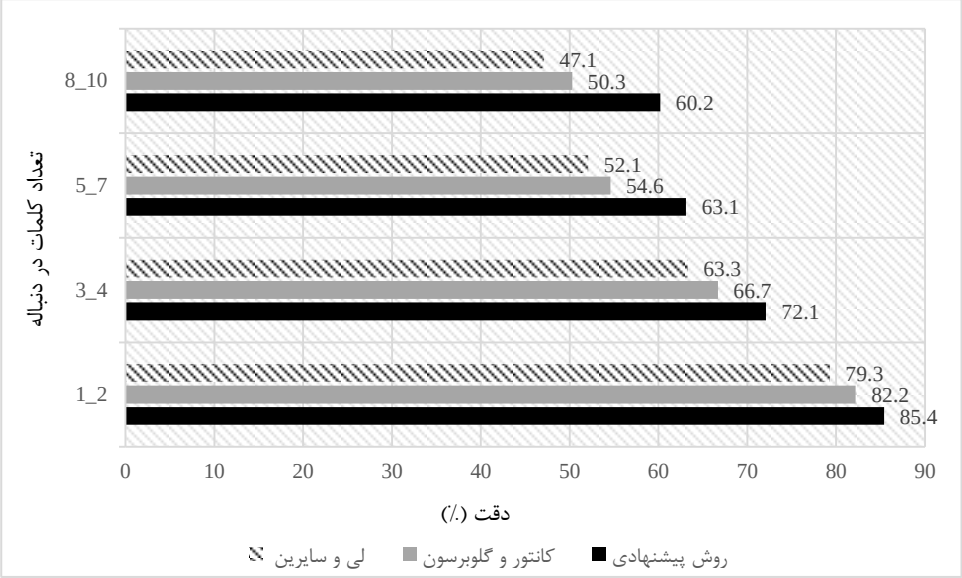
	Proposed method	Stanford-NLP	OpenNLP	LingPipe	SupersenseTagger
F1	88.19	81.15	76.69	60.07	71.95
Recall	82.28	75.21	71.16	54.21	69.05
Precision	95.01	88.12	83.15	67.34	75.11

شکل ۴-۲. دقت و بازخوانی در تشخیص موجودیت‌های نامدار روی پیکره English Gigaword

[1] http://nlp.stanford.edu/, http://opennlp.apache.org/, http://alias-i.com/lingpipe/, http://sites.google.com/site/massiciara/, http://afner.sourceforge.net/afner.html, http://www.alchemyapi.com

۴-۷-۵ قدرت تشخیص نامبری بر اساس طول دنباله کلمات

در شکل ۴-۳ دقت تشخیص نامبری روش پیشنهادی از میان دنباله کلمات و بر اساس تعداد کلمات موجود در دنباله با سایر روش‌های کانتور و گلوبرسون [۱۰۵] و لی و سایرین [۸۲] روی پیکره توسعه CONLL-2012 shared task مقایسه شده است. همان‌طور که مشاهده می‌شود، دقت روش پیشنهادی از سایر روش‌ها بالاتر است.



شکل ۴-۳. نرخ تشخیص نامبری برحسب طول دنباله کلمات روی پیکره توسعه CONLL-2012 shared task

در سایر روش‌ها دقت تشخیص نامبری با افزایش طول دنباله کلمات کاهش چشمگیری دارد اما در روش پیشنهادی دقت تشخیص نامبری کاهش اندکی دارد و برای دنباله‌هایی با طول بیش از پنج کلمه کاهش دقت ناچیز است. روش کانتور و گلوبرسون عملکرد بهتری نسبت به روش لی و سایرین دارد، زیرا از تعبیه واژگان BERT به‌جای ELMo استفاده کرده است و بازنمایش دقیق‌تری برای دنباله کلمات ارائه داده است. در روش لی و سایرین نیز، همپوشانی زیادی بین نامبری‌هایی طلایی و مجموعه توسعه وجود دارد. مدل پیشنهادی قادر است ۱۰۵۹ نامبری (۳۹۴ نامبری در مجموعه آموزش و ۶۶۵ نامبری که در مجموعه آموزش وجود ندارند.) را به‌طور درست شناسایی کند که روش-های کانتور و گلوبرسون و لی و سایرین قادر به این میزان تشخیص درست نیستند. همچنین طبق

ساختار ارائه‌شده نامبری‌های دیده نشده در مجموعه آموزش نیز قابل تشخیص هستند. علاوه بر این، تأثیر افزودن نامبری‌های طلایی نیز بر نرخ تشخیص نامبری بررسی شد و به این نتیجه رسیدیم که روش پیشنهادی عملکرد بهتری نسبت به دو روش دیگر دارد. علت برتری روش پیشنهادی استفاده از شبکه عمیق GRU دوجهته جهت استخراج دنباله کلمات است. همچنین تعبیه واژگان RoBERTa اطلاعات معنایی دقیقی در اختیار سیستم قرار می‌دهد.

۴-۷-۶ بررسی تأثیر منابع دانش مختلف در پیکره‌های CONLL-2012 shared task و i2b2

جهت بررسی میزان تأثیر هریک از منابع دانش بر عملکرد مدل پیشنهادی، تک‌تک این منابع را از مدل حذف کردیم و بررسی که کردیم که با حذف هریک از آن‌ها دقت روش پیشنهادی روی دو پایگاه داده CONLL-2012 shared task و i2b2 چه تغییری می‌کند. همان‌طور که در جدول ۴-۴ مشخص است، هر نوع دانش روی پیکره‌ای خاص عملکرد متغیری دارد. برای مثال، وجود دانش زبان-شناسی در پیکره CONLL-2012 shared task بالاترین تأثیر را نسب به بقیه دانش‌ها دارد. درحالی‌که در پیکره پزشکی i2b2 حذف دانش پزشکی بیشترین تأثیر منفی را بر عملکرد مدل پیشنهادی می‌گذارد.

جدول ۴-۴. نتایج حذف منابع دانش روی پیکره CoNLL-2012 shared task و i2b2

	CONLL-2012		**i2b2**	
	F1	$\Delta F1$	F1	$\Delta F1$
مدل به‌طور کامل	۸۳/۹	-	۹۷/۳	-
دانش OMCS	۸۰/۱	-۳/۸	۹۷/۱	-۰/۲
دانش پزشکی	۷۹/۱	-۴/۸	۹۶/۵	-۰/۸
دانش زبان‌شناسی	۷۸/۳	-۵/۶	۹۷/۱	-۰/۲
دانش لغوی	۷۸/۸	-۵/۱	۹۶/۸	-۰/۵

۴-۷-۷ توانایی شبکه‌های عصبی عمیق بازگشتی

همان‌طور که در بخش ۳-۲-۲ اشاره شد، جهت استخراج دنباله کلمات از تعبیه واژگان ورودی از شبکه GRU دوجهته استفاده شده است، جهت بررسی کارایی این شبکه عملکرد شبکه‌های RNN، LSTM و LSTM دوجهته نیز در این زمینه بررسی شد. در جدول ۴-۵ قدرت این شبکه‌های بازگشتی در استخراج دنباله کلمات نشان داده شده است. همان‌طور که مشخص است شبکه GRU دوجهته دارای عملکرد بهتر است و توانایی بالاتری در استخراج وابستگی‌های طولانی دارد. در بعضی زمینه‌های پردازش متن نیز شبکه‌های GRU توانسته‌اند نسبت به شبکه‌های LSTM نتایج بهتری تولید کنند، دلیل این برتری این است که در این شبکه‌ها گیت‌های فراموشی و ورودی در گیت به‌روزرسانی ادغام شده‌اند، که باعث بالاتر رفتن سرعت عملکرد آن‌ها می‌شود. همچنین استفاده از تعبیه واژگان RoBERTa باعث شده است که اطلاعات معنایی بهتری در اختیار این شبکه قرار گیرد و به‌دلیل داشتن سرعت بالای این شبکه‌ها دنباله کلمات با سرعت بالاتری استخراج شود؛ در نتیجه نامبری‌ها با دقت بالاتر استخراج شده و عملکرد سیستم بهبود داده می‌شود.

جدول ۴-۵. توانایی استخراج دنباله کلمات توسط شبکه‌های عصبی عمیق بازگشتی

شبکه بازگشتی	Prec.	Rec.	F1
RNN	۷۹/۲۲	۸۳/۱۳	۸۱/۱۲
LSTM	۸۱/۴۹	۸۵/۹۹	۸۳/۶۷
Bidirectional LSTM	۸۵/۳۱	۹۳/۹۰	۸۹/۳۹
GRU	۸۴/۱۹	۹۰/۸۲	۸۷/۳۷
Bidirectional GRU	۸۹/۹۳	۹۷/۹۵	۹۳/۷۶

۴-۷-۸ ارزیابی عملکرد روش پیشنهادی با استفاده از داده‌های آموزشی مختلف

در جدول ۴-۶ عملکرد روش پیشنهادی و روش لی و همکاران [۸۲] توسط داده‌های مختلف آموزش و تست نمایش داده شده است. همان‌طور که مشخص است هر دو روش زمانی‌که داده‌های آموزش از

پیکره i2b2 و داده‌های تست از پیکره CoNLL-2012 shared task است، عملکرد ضعیفی دارند. این نشان می‌دهد که دانش پزشکی تأثیر کمی در پیکره CoNLL-2012 shared task دارد. زمانی‌که داده‌های آموزش از پیکره CoNLL-2012 shared task و داده‌های تست از پیکره i2b2 باشند، روش پیشنهادی عملکرد قابل‌قبولی (۹/۲+) در مقایسه با روش لی و همکاران دارد که نشان‌دهنده توانایی منابع دانش در این پیکره است.

جدول ۴-۶. ارزیابی عملکرد روش لی و همکاران و روش پیشنهادی توسط داده‌های آموزشی مختلف

روش	داده‌های آموزش	داده‌های آزمایش	
		CoNLL-2012	**i2b2**
روش لی و همکاران	CoNLL-2012	۷۳/۰	۷۶/۷
	i2b2	۲۱/۴	۹۳/۸
روش پیشنهادی	CoNLL-2012	۸۳/۹	۸۵/۹
	i2b2	۴۶/۸	۹۷/۳

۴-۸ نتیجه‌گیری

همان‌طور که در ارزیابی‌ها و نتایج مشخص است، روش‌های پیشین نسبت به روش پیشنهادی دقت کمتری دارند. درک درست مفاهیم و تشخیص درست زمجیره‌های هم‌مرجع بودن کمک زیادی به بهبود عملکرد سیستم‌هایی نظیر خلاصه‌سازی [۱۰۸]، پرسش و پاسخ [۱۰۹]، ترجمه ماشینی [۱۱۰] و غیره می‌کند [۱۱۱ و ۱۱۲]. به‌عنوان مثال در ترجمه ماشینی، جایگزینی ضمایر با مراجع آن‌ها بسیار مهم است.

فصل ۵ : دستاوردهای پژوهش و پیشنهاداتی برای تحقیقات آتی

۵-۱ دستاوردهای پژوهش

در فصل ۴، مقایسه‌ی مناسبی بین روش پیشنهادی و روش‌های جدید تحلیل هم‌ارجاعی در چند بعد نظیر بهبود Avg.F1 در تشخیص زنجیره‌های هم‌ارجاعی، دقت روش‌های رتبه‌بندی چند شاخصه ارائه‌شده، میزان تأثیر روش‌های مختلف تعبیه واژگان، قدرت تشخیص موجودیت‌های نامدار و غیره صورت گرفت. همانطور که نشان داده شد، روش پیشنهادی با کمترین میزان خطا نسبت به سایر روش‌ها، مسئله تحلیل هم‌ارجاعی را به‌خوبی مدیریت می‌کند. در دقت تشخیص زنجیره‌های هم مرجع روش پیشنهادی، معیار Avg.F1 را به میزان ۳/۷ درصد بهبود داد. همچنین روش رتبه‌بندی مبتنی بر شبکه کوهونن عملکرد بهتری نسبت به روش رتبه‌بندی مبتنی بر شبکه پرسپترون داشت و روش-های رتبه‌بندی ارائه شده نسبت به سایر روش‌های رتبه‌بندی عملکرد بهتری داشتند. در تشخیص موجودیت‌های نامدار، معیار Avg.F1 روش پیشنهادی در مقایسه با روش‌های قبلی ۷/۰۴ درصد بهبود یافت. در تشخیص نامبری‌ها مشخص شد که افزایش طول دنباله کلمات تأثیر چندانی در کاهش دقت تشخیص ندارد. در بررسی میزان تأثیر منابع دانش مختلف مشخص شد که دانش زبان-شناسی و دانش لغوی، نقش مهم‌تری در بهبود عملکرد روش پیشنهادی روی پیکره CONLL-2012 shared task دارند. با توجه به دقت بالای مدل ارائه‌شده می‌توان از آن برای ارتقاء عملکرد سیستم-های مبتنی بر متن نظیر استخراج اطلاعات، خلاصه‌سازی متون، ترجمه ماشینی و پرسش و پاسخ استفاده کرد. همچنین به‌دلیل تجهیز مدل پیشنهادی به دانش پزشکی، می‌توان از آن در جهت بهبود کاربردهای پزشکی مبتنی بر متن استفاده کرد و در مواردی نظیر تشخیص بیماری، تداخل داروها و غیره از آن استفاده کرد.

۵-۲ پیشنهاداتی برای تحقیقات آتی

با توجه به اینکه لازم است برای دستیابی به سطح کارایی بیشتر در سیستم‌های تحلیل مرجع از دانش معنایی و دانش جهانی استفاده شود، روش پیشنهادی توانست تا اندازه‌ای این دانش را به وجود آورده و آن را در سیستم تحلیل مرجع به کار ببندد. با وجود بهبودی که روش پیشنهادی در کارایی سیستم تحلیل مرجع به وجود آورده است، اما این مسیر تحقیقاتی باز هم جای کار و بهبود دارد. مواردی که به‌عنوان کارهای آتی این پژوهش متصور هستند، عبارتند از:

- پیاده‌سازی این روش روی زبان فارسی: با توجه به حجم کمتر کار انجام‌شده روی زبان فارسی یک ایده خوب اعمال این روش به زبان فارسی می‌باشد.
- بهبود پایگاه دانش: می‌توان از منابع دانش تخصصی‌تر جهت استفاده از مدل پیشنهادی در زمینه‌های خاص استفاده کرد.
- استفاده از مدل موجودیت-نامبری بجای مدل زوج-نامبری و مدل رتبه‌ای
- استفاده از تکنیک‌های وزن‌دهی مختلف جهت وزن‌دهی دقیق به ویژگی‌ها: تکنیک‌های مختلفی نظیر روش وزن‌دهی بردار ویژه، روش فازی و غیره جهت وزن‌دهی به ویژگی‌هایی که به‌صورت دستی به سیستم اضافه می‌شوند، وجود دارد. می‌توان از هر یک از این تکنیک‌ها استفاده کرد و عملکرد سیستم را در حالت‌های مختلف مقایسه کرد.
- استفاده از تکنیک‌های تصمیم‌گیری چندشاخصه فازی
- ارائه مدل‌های ترکیبی جهت عملکرد سیستم تحلیل هم‌ارجاعی
- استفاده از دیگر شبکه‌های عصبی عمیق بازگشتی و تکنیک‌های جدید یادگیری عمیق
- استفاده از پایگاه دانش زوج افعال مرتبط جهت بهبود دقت شناسایی زنجیره‌های هم‌مرجع

پیوست

واژه‌نامه فارسی به انگلیسی

A	Anaphor ..... نامبری وابسته	
	anaphora resolution ..... تحلیل ضمیر	
	anaphoric filter ..... فیلتر شناسایی نامبری وابسته	
	anaphoricity ..... وابسته بودن نامبری	
	animacy ..... جاندار بودن	
	animacy agreement ..... توافق جانداری	
	annotation ..... حاشیه‌نویسی	
	antecedant ..... مقدم	
	appositive ..... بدل	
B	bootstrap ..... خودراه‌انداز	
	bridging ..... مجاز مرسل	
C	cataphor ..... اشاره به جلو	
	causative ..... علیت	
	centering theory ..... تئوری مرکزیت	
	chunking ..... قطعه‌بندی	
	classification ..... رده‌بندی	
	clause ..... بستار	
	cluster-level features ..... ویژگی‌های سطح خوشه	
	coherent ..... انسجام	
	concept ..... مفهوم	
	constituent ..... جزء	
	constituent parse ..... تجزیه سازه‌ای	
	constituent parse tree ..... درخت تجزیه سازه‌ای	
	constraints and preferences ..... محدودیت‌ها و تقدم ها	
	context ..... بافتار	
	coreference ..... هم‌ارجاعی	
	coreference module ..... پیمانه هم‌ارجاعی	
	coreference resolution ..... تحلیل مرجع	

coreferent ..... هم‌مرجع		
corpora/ corpus ..... پیکره		
data driven approaches ..... رویکرد داده گرا		D
dependency parse ..... تجزیه وابستگی		
dependency parse tree ..... درخت تجزیه وابستگی		
description logic ..... منطق توصیفی		
determiner ..... تعریف کننده		
deverbal nouns ..... گزاره‌های اسمی		
different surface realizations ..... صورت ظاهری متفاوت		
discourse ..... گفتمان		
discourse connectives ..... ربط گفتمان		
discourse knowledge ..... دانش گفتمان		
discourse structure ..... ساختار گفتمان		
discriminative ..... متمایز کننده		
disjoint ..... متمایز		
distributional Hypothesis ..... فرضیه توزیعی		
distributional Memory ..... حافظه توزیعی		
distributional semantic models ..... مدل معنایی توزیعی		
eager ..... مشتاق		E
end-to-end system ..... سیستم هم‌ارجاعی کامل		
entity ..... موجودیت		
entity-mention model ..... مدل موجودیت نامبری		
evidence ..... شاهد		
expectation ..... انتظار		
expletive pronouns ..... ضمیر جاپرکن		
expressive ..... مبیّن		
fact ..... واقعیت		F
false negative ..... منفی نادرست		
false positive ..... مثبت نادرست		
feature ..... ویژگی		
feature selection ..... گزینش ویژگی		
fiber ..... فایبر		

منطق مرتبه‌ی اول	first order logic	
تئوری تمرکز	focusing theory	
قاب	frame	
اطلس	gazeeteer	**G**
شرح	gloss	
استاندارد طلایی	gold-standard	
کلمه سر	head	**H**
کل‌نام	holonym	
هم‌نامی	homonymy	
نمونه	individual	**I**
استنتاج	inference	
بازیابی اطلاعات	information retrieval	
بررسی نمونه	instance checking	
بین جمله ای	intersentential	
درون جمله ای	intrasentential	
نامبری‌های منزوی / تکی	isolated mentions	
توامان	joint	**J**
استنتاج توامان	joint inference	
لغوی نحوی	lexico-syntactic	**L**
روش‌های زبانی	linguistics-based methods	
مقیاس مبتنی بر ربط	link-based measure	
سنجه مبتنی بر ربط	link-based performance metrics	
داده‌های متصل	linked data	
اتصال	linking	
اتصال‌دهنده‌ی منطقی	logical connective	
روش‌های یادگیری ماشینی	machine learning-based methods	**M**
نامبری	markable	
نامبری	mention	
شبکه منطق مارکف	markov logic network	
ماتریسی کردن	matricization	
بازنمایی معنا	meaning representation	
مدل زوج-نامبری	mention-pair (pairwise) model	

meronym ..... زیرنوع		
modifier ..... تعریف کننده		
morphological process ..... فرایندهای ساختواژی		
multi-mention entity ..... موجودیت دارای چند نامبری		
narrative schema ..... شمای روایی		**N**
negative instance ..... نمونه‌های منفی		
NER ..... تشخیص دهنده موجودیت		
nominal mention ..... نامبری عبارت اسمی		
non-anaphoric ..... غیر وابسته		
noun phrase ..... عبارت اسمی		
NP extractor ..... استخراجگر عبارات اسمی		
null anaphora or ellipsis ..... حذف به قرینه معنوی		
on demand ..... بر تقاضا		**O**
ontology ..... هستان‌نگار		
ontology population ..... مقداردهی به هستان‌نگار		
pairwise/mention-pair model ..... مدل زوج نامبری		**P**
paraphrase ..... بازگویی		
part of speech ..... اجزای کلام		
participant ..... شرکت‌کننده		
passive voice ..... مجهول		
penn treebank ..... بانک درخت پن		
phrase ..... عبارت		
polysemy ..... چندمعنایی		
POS tagging ..... برچسب زنی اجزای کلام		
positive instance ..... نمونه‌های مثبت		
predicate-Argument Structure (PAS) ..... ساختار گزاره-آرگومان		
predicative adjectives ..... گزاره‌های صفتی		
predicate nominative یا capula ..... کاپولا		
Query ..... پرس‌وجو		**Q**
ranking model ..... مدل رتبه‌ای		**R**
recency ..... نزدیکی مکانی		
referring expression ..... عبارت ارجاعی یا نامبری		

recency information	.....	اطلاعات نزدیکی مکانی
reflexive	.....	بازتابی
relation	.....	رابطه
relation discovery	.....	کشف رابطه
rhetorical and discourse coherence	.....	پیوستگی گفتمان
role	.....	نقش
rule based	.....	مبتنی بر قاعده
seed	.....	بذر، هسته
selectional preference	.....	ترجیح انتخابی
selectional restriction	.....	محدودیت انتخابی
self-supervised	.....	خوداظر
semantic analysis	.....	تحلیل معنایی
semantic role labeling	.....	برچسب زنی نقش معنایی
sense	.....	واژه‌بار
sentence segmentation	.....	تفکیک جملات
sentence splitter	.....	جدا کننده جمله
sequence labeling	.....	برچسب زنی دنباله
singleton entity	.....	موجودیت تکی
slot	.....	شکاف
sparse	.....	خلوت
SRL	.....	برچسب نقش معنایی
stemmer	.....	ریشه یاب
subsequent mention	.....	نامبری‌های بعدی
subsumption	.....	شمول
supervised	.....	با ناظر
symmetric	.....	متقارن
syntactic frames	.....	قاب نحوی
system mention	.....	نامبری سیستمی
template	.....	قالب
temporal direction	.....	جهت زمانی
term	.....	اصطلاح
theme	.....	موضوع
tokenizer	.....	تک‌واژ ساز

tokenization ..... تک‌واژ سازی transitive ..... تعدی true mentions ..... نامبری دقیق true/gold mentions ..... نامبری‌های طلایی	
unsupervised ..... بی‌ناظر	**U**
WSD-Word sense disambiguation ..... رفع ابهام واژه‌بار کلمات	**W**

مراجع

[۱] سهلانی ح، ۱۳۹۹، "مرجع‌گزینی در زبان فارسی با استفاده از شبکه عصبی عمیق" مجله پردازش علائم و داده‌ها، شماره ۲، دوره ۱۷: ص ۱۳۸-۱۲۱.

[۲] ظفری، ح، ۱۳۹۶، رساله دکتری، تحلیل هم‌ارجاعی با کمک الگوریتم های یادگیری ماشینی، مجتمع دانشگاهی فن‌آوری اطلاعات، ارتباطات و امنیت، دانشگاه صنعتی مالک اشتر.

[۳] رحیمی ز. و حسین نژاد ش. (۱۳۹۹) "هم‌مرجع‌یابی مبتنی بر پیکره در متون فارسی،" مجله پردازش علائم و داده ها، شماره ۱، دوره ۱۷: ص ۹۸-۷۹.

[4] Morton T. S. (1999) "Using coreference for question answering", In Proceedings of the Workshop on Coreference and its Applications, P85–89, United States.

[5] Haghighi, A. and Klein, D., (2009) "Simple coreference resolution with rich syntactic and semantic features", In Proceedings of the 2009 conference on empirical methods in natural language processing, P1152-1161, Singapore.

[6] Sapena E, Padró L, and Turmo J. (**2013**) "A Constraint-Based Hypergraph Partitioning Approach to Coreference Resolution" **Comput. Linguist**., **39**, pp**847–884**.

[7] W. M. Soon, H. T. Ng, and D. C. Y. Lim, (**2001**) "A Machine Learning Approach to Coreference Resolution of Noun Phrases" **Comput. Linguist.**, **27**, **4**, pp**521–544**.

[8] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L., (2018) "Deep Contextualized Word Representations", In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp 2227–2237, New Orleans, Louisiana.

[9] Devlin J., Chang M. W., Lee K., and Toutanova K., (2019) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp4171–4186, Minneapolis, Minnesota.

[10] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., (2019) "RoBERTa: A Robustly Optimized BERT Pretraining Approach" ArXiv Prepr. ArXiv190711692.

[11] Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R. R., and Le Q. V., (2019) "Xlnet: Generalized autoregressive pretraining for language understanding", in Advances in neural information processing systems, P5754–5764.

[12] Collobert R. and Weston J., (2008) "A unified architecture for natural language processing: Deep neural networks with multitask learning", In Proceedings of the 25th international conference on Machine learning, P160–167.

[13] Collobert R., Weston J., Bottou L., Karlen M., Kavukcuoglu K., and Kuksa P., (2011) "Natural language processing (almost) from scratch" **J. Mach. Learn. Res**, **12**, pp**2493–2537**.

[14] Passos A., Kumar V., and McCallum A., (2014) "Lexicon Infused Phrase Embeddings for Named Entity Resolution", In Proceedings of the Eighteenth

Conference on Computational Natural Language Learning, P78–86, Ann Arbor, Michigan.
[15] dos Santos C. and Guimarães V., (2015) "Boosting Named Entity Recognition with Neural Character Embeddings", In Proceedings of the Fifth Named Entity Workshop, P25–33, Beijing, China.
[16] Panchendrarajan R. and Amaresan A., (2018) "Bidirectional LSTM-CRF for Named Entity Recognition", PACLIC 2018, Hong Kong.
[17] Dupond S., (**2019**) "A thorough review on the current advance of neural network structures" **Annu. Rev. Control**, **14**, pp**200–230**.
[18] Hochreiter S., (**1997**), "Long Short-Term Memory" **Neural Comput**., **9**, **8**.
[19] Chiu J. P. and Nichols E., (**2016**) "Named entity recognition with bidirectional LSTM-CNNs" **Trans. Assoc. Comput. Linguist**., **4**, pp**357–370**.
[20] Knight K., Nenkova A., and Rambow O., (2016) "Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies", In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies San Diego, California.
[21] Ma X. and Hovy E., (2016) "End-to-end sequence labeling via bi-directional lstm-cnns-crf", in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, P1064–1074, Berlin, Germany.
[22] Lin B. Y., Xu F., Luo Z., and Zhu K., (2017) "Multi-channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media", In Proceedings of the 3rd Workshop on Noisy User-generated Text, P 160–165, Copenhagen, Denmark.
[23] Lee K., He L., and Zettlemoyer L., (2018) "Higher-Order Coreference Resolution with Coarse-to-Fine Inference", In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, P687–692, New Orleans, Louisiana.
[24] Joshi M., Levy O., Zettlemoyer L., and Weld D., (2019) "BERT for Coreference Resolution: Baselines and Analysis", In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), P5803–5808, Hong Kong, China.
[25] Hobbs J. R., (**1978**) "Resolving pronoun references" **Lingua**, **44**, **4**, pp**311–338**.
[26] Jenkins D., (**1995**) "Centering: A Framework for Modelling the Local Coherence of Discourse," **Comput. Linguist**., **21**, **2**, pp**203–225**.
[27] Lee H., Peirsman Y., Chang A., Chambers N., Surdeanu M., and Jurafsky D., (2011) "Stanford ' s Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task",In Proceedings of the 15th conference on computational natural language learning: Shared task. Association for Computational Linguistics, Portland, P28–34, Oregon, USA.
[28] Denis P. and Baldridge J., (2007) "Joint determination of anaphoricity and coreference resolution using integer programming", in Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, P236–243, Rochester, New York.
[29] Bögel, T. and Frank, A (2013). "A joint inference architecture for global coreference clustering with anaphoricity", In Language Processing and Knowledge in the Web. Springer, Berlin, Heidelberg, pp. 35-46.

[30] Ng V. and Cardie C., (2002) "Improving Machine Learning Approaches to Coreference Resolution", In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, Philadelphia, P104–111, Pennsylvania, USA.
[31] Uryupina O., (2003) "High-precision identification of discourse new and unique noun phrases", in The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics, P80–86.
[32] Luo X., (2007) "Coreference or not: A twin model for coreference resolution", in Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, Rochester, P73–80, New York.
[33] Luo X., Ittycheriah A., Jing H., Kambhatla N., and Roukos S., (2004) "A mention-synchronous coreference resolution algorithm based on the bell tree", In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), P135, Barcelona, Spain.
[34] Ng V., (**2009**) "Graph-cut-based anaphoricity determination for coreference resolution" **Comput. Linguist**., pp**575–583**.
[35] McCallum A. and Wellner B., (**2005**) "Conditional Models of Identity Uncertainty with Application to Noun Coreference" **Adv. Neural Inf. Process. Syst. 17**, pp **905–912**.
[36] Nicolae C. and Nicolae G., (**2006**) "BEST CUT : A Graph Algorithm for Coreference Resolution" **Comput. Linguist**., **2015**, pp**275–283**.
[37] Klenner M. and Ailloud E., (2008) "Enhancing coreference clustering", Proceedings of the Second Workshop on Anaphora Resolution (WAR II), P31–40, Bergen, Norway.
[38] Cardie C. and Wagstaff K (1999) "Noun phrase coreference as clustering", EMNLP.
[39] Ng V., (2008) "Unsupervised models for coreference resolution", In Proceedings of the Conference on Empirical Methods in Natural Language Processing, P640–649, United States.
[40] Miller G. A., (**1995)** "WordNet: a lexical database for English" **Commun. ACM, 38**, **11**, PP**39-41.**
[41] Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Famian, A., Bagherbeigi, S., Fekri, E., Monshizadeh, M. and Assi, S.M., (2010) "Semi automatic development of farsnet; the persian wordnet", In Proceedings of 5th global WordNet conference, Mumbai, India.
[42] Cardie C. and Wagstaff K., (1999) "Noun Phrase Coreference as Clustering", In 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, P82–89, USA.
[43] Poon H. and Domingos P., (2008) "Joint Unsupervised Coreference Resolution with Markov logic", In Proceedings of the 2008 conference on empirical methods in natural language processing, P650-659, Honolulu, Hawaii.
[44] Ng V., (2008) "Unsupervised Models for Coreference Resolution", In Proceedings of the 2008 conference on empirical methods in natural language processing, P640–649, Honolulu.
[45] Rahman A. and Ng V., (2011) "Coreference Resolution with World Knowledge", In Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, P814–824, United States.

[46] Wiseman S. J., Rush A. M., Shieber S. M., and Weston J., (2015) "Learning anaphoricity and antecedent ranking features for coreference resolution", In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, P1416–1426, Beijing, China.
[47] Wiseman S., Rush A. M., and Shieber S. M., (2016) "Learning Global Features for Coreference Resolution", In Proceedings of NAACL-HLT, P994–1004, San Diego, California.
[48] Clark K. and Manning C. D., (2016) "Improving Coreference Resolution by Learning Entity-Level Distributed Representations", In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, P643–653.
[49] Clark K. and Manning C. D., (2016) "Deep Reinforcement Learning for Mention-Ranking Coreference Models", In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, P2256–2262, Austin, Texas.
[50] Lee K., He L., Lewis M., and Zettlemoyer L., (2017) "End-to-end neural coreference resolution", Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, P188–197, Copenhagen, Denmark.
[51] C. Aone and S. Bennett, (1995) "Evaluating automated and manual acquisition of anaphora resolution strategies", In 33rd Annual Meeting of the Association for Computational Linguistics , P122–129.
[52] McCarthy J. F. and Lehnert W. G., (**1995**) "Using decision trees for coreference resolution," **IJCAI**, pp**1050–1055**.
[53] Ponzetto S. P. and Strube M., (2006) "Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution," Proceedings of the Human Language Technology Conference of the NAACL, Main Conference, P192–199, New York City, USA.
[54] Yang X., Su J., and Tan C. L., (2006) "Kernel-based pronoun resolution with structured syntactic knowledge", Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, P41–48, Sydney, Australia.
[55] Ng V., (2005) "Machine Learning for Coreference Resolution: From Local Classification to Global Ranking", ACL 2005, 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, P157–164, University of Michigan, USA.
[56] Ng V., (**2007**) "Shallow semantics for coreference resolution" **IJCAI Int. Jt. Conf. Artif. Intell**., **7**, pp**1689–1694**.
[57] Ji H., Westbrook D., and Grishman R., (2005) "Using semantic relations to refine coreference decisions", EMNLP, pp17–24.
[58] Bengtson E. and Roth D., (2008) "Understanding the value of features for coreference resolution", In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, P294-303, Honolulu, Hawaii.
[59] Stoyanov V., Gilbert N., Cardie C., and Riloff E., (2009) "Conundrums in Noun Phrase Coreference Resolution : Making Sense of the State-of-the-Art", Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, P656–664, Suntec, Singapore.
[60] Uryupina O., Poesio M., Giuliano C., and Tymoshenko K., (2011) "Disambiguation and Filtering Methods in Using Web Knowledge for Coreference

Resolution", In Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches, P317–322.
[61] Yang X., Zhou G., Su J., and Tan C. L., (2003) "Coreference resolution using competition learning approach", In Proceedings of the 41st annual meeting of the association for computational linguistics, P176–183, Sapporo, Japan.
[62] Denis P. and Baldridge J., (2008) "Specialized models and ranking for coreference resolution", Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, P660–669, Honolulu, Hawaii.
[63] Yang X., Tan C. L., Su J., Liu T., Lang J., and Li S., (2008) "An Entity-Mention Model for Coreference Resolution with Inductive Logic Programming", In Proceedings of ACL-08: Hlt, P843–851, Columbus, Ohio.
[64] Luo X., (**2007**) "Coreference or Not : A Twin Model for Coreference Resolution 1101 Kitchawan Road" **Comput. Linguist**., pp**73–80**.
[65] Klenner M. and Ailloud É., (2008) "Enhancing coreference clustering", In Proceedings of Second Bergen Workshop Anaphora Resolution. WAR II, P31–40, Bergen, Norway.
[66] Denis P. and Baldridge J., (**2007**) "Joint Determination of Anaphoricity and Coreference Resolution using Integer Programming" **Comput. Linguist**., pp**236–243**.
[67] Klenner M., (2007) "Enforcing consistency on coreference sets", In Recent Advances in Natural Language Processing (RANLP), P323–328 Borovets, Bulgaria.
[68] Finkel J. R. and Manning C. D., (2008) "Enforcing transitivity in coreference resolution", In Proceedings of ACL-08: HLT, Short Papers, P45-48, Columbus, Ohio.
[69] Bean D. and Riloff E., (2004) "Unsupervised Learning of Contextual Role Knowledge for Coreference Resolution", Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004, P297–304, Boston, Massachusetts, USA.
[70] Culotta A., Wick M., Hall R., and McCallum A., (2007) "First-order probabilistic models for coreference resolution", In Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, P81–88, Rochester, New York.
[71] Finley T. and Joachims T., (2005) "Supervised clustering with support vector machines", In Proceedings of the 22nd international conference on Machine learning, pp. 217–224, New York, United States.
[72] Cai J. and Strube M., (2010) "End-to-End Coreference Resolution via Hypergraph Partitioning," In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pp. 143–151, Beijing, China.
[73] Yang X., Su J., Zhou G., and Tan C. L., (2004) "An NP-cluster based approach to coreference resolution," In COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics, pp. 226-232, Geneva, Switzerland.
[74] Haghighi A. and Klein D., (2010) "Coreference resolution in a modular, entity-centered model", In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 385–393, Los Angeles, California.

[75] Poon H. and Domingos P., (2008) “Joint Unsupervised Coreference Resolution with Markov logic”, In Proceedings of the 2008 conference on empirical methods in natural language processing, P650-659, Honolulu, Hawaii.
[76] Lee H., Recasens M., Chang A., and Surdeanu M., (2012) “Joint entity and event coreference resolution across documents”, In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, P489-500, Jeju Island, Korea.
[77] Zheng J., Vilnis L., Singh S., Choi J., and McCallum A., (2013) “Dynamic Knowledge-Base Alignment for Coreference Resolution”, In Proceedings of the Seventeenth Conference on Computational Natural Language Learning, P153–162, Sofia, Bulgaria.
[78] Bansal M. and Klein D., (2012) “Coreference Semantics from Web Features”, In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, P389–398, Jeju Island, Korea.
[79] Peng H., Chang K.-W., and Roth D., (2015) “A Joint Framework for Coreference Resolution and Mention Head Detection”, In Proceedings of the Nineteenth Conference on Computational Natural Language Learning, P12–2, Beijing, China.
[80] Peng H., Khashabi D., and Roth D., (2015) “Solving Hard Coreference Problems”, NAACL HLT 2015 - 2015 Conference of the North American Chapter of the Association for Computational Linguistics, P809–819, Denver, United States.
[81] Lee K., He L., Lewis M., and Zettlemoyer L., (2017) “End-to-end Neural Coreference Resolution”, In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, P188–197, Copenhagen, Denmark.
[82] Lee K., He L., and Zettlemoyer L., (2018) “Higher-order coreference resolution with coarse-to-fine inference”, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, P687–692, New Orleans, Louisiana.
[83] Fei H., Li X., Li D., and Li P., (2019) “End-to-end Deep Reinforcement Learning Based Coreference Resolution”, In Proceedings of the 57th Conference of the Association for Computational Linguistics, P660–665, Florence, Italy.
[84] Joshi M., Levy O., Weld D. S., and Zettlemoyer L., (2019) “BERT for Coreference Resolution: Baselines and Analysis”, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), P5803–5808, Hong Kong, China.
[85] Seraji M., Jahani C., Megyesi B., and Nivre J., (2013), Technical report, “Uppsala Persian dependency treebank annotation guidelines”, Uppsala University.
[86] LeCun Y., Bengio Y., and Hinton G., (**2015**) “Deep learning” **nature**, **521**, **7553**, pp**436-444**.
[87] Deng L. and Yu D., (**2014**) “Deep learning: methods and applications” **Found. Trends® Signal Process.**, **7**, **3–4**, pp. **197–387**.
[88] Chung J., Gulcehre C., Cho K., and Bengio Y., (2014) “Empirical evaluation of gated recurrent neural networks on sequence modeling,” NIPS 2014 Workshop Deep Learning.
[89] Bahdanau D., Cho K., and Bengio Y., (2014) “Neural machine translation by jointly learning to align and translate”, 3rd International Conference on Learning Representations, ICLR 2015 - San Diego, United States.
[90] Kubat M., (**1999**) “Neural networks: a comprehensive foundation by Simon Haykin, Macmillan” **Knowl. Eng. Rev.**, **13**, **4**, pp**409–412**.

[91] Kohonen T., (2012) "Self-organization and associative memory", Vol. 8. Springer Science & Business Media.
[92] Mitkov R. and Hallett C., (**2007**) “Comparing pronoun resolution algorithms” **Comput. Intell**., **23**, **2**, pp**262–297**.
[93] D. K. Byron, (**2001**) “The uncommon denominator: A proposal for consistent reporting of pronoun resolution results” **Comput. Linguist.**, **27**, **4**, pp. **569–577**.
[94] Vilain M., Burger J., Aberdeen J., Connolly D., and Hirschman L., (1995) “A Model-Theoretic Coreference Scoring Scheme”, In Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, P45–52.
[95] Bagga A. and Baldwin B., (1998) “Algorithms for scoring coreference chains”, in The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference, P563–566, Granada, Spain.
[96] Luo X., (2005) “On coreference resolution performance metrics”, Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, P25–32, Vancouver, British Columbia, Canada.
[97] Marta R., and Eduard H., (**2011**) “BLANC: Implementing the Rand index for coreference evaluation” **Nat. Lang. Eng.**, **17**, **4**, pp**485–510**.
[98] Moosavi N. S. and Strube M., (2016) “Which coreference evaluation metric do you trust? a proposal for a link-based entity aware metric”, In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, P632–642, Berlin, Germany.
[99] Pradhan S., Moschitti A., Xue N., Uryupina O., and Zhang Y., (2012) “CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes”, in Joint Conference on EMNLP and CoNLL-Shared Task, , P1–40, Jeju Island, Korea.
[100] Uzuner O., Bodnari A., Shen S., Forbush T., Pestian J., and South B. R., (**2012**) “Evaluating the state of the art in coreference resolution for electronic medical records” **J. Am. Med. Inform. Assoc.**, **19**, **5**, pp**786–791**.
[101] Grishman R. and Sundheim B., (1996) “Message understanding conference-6: A brief history”, in COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics.
[102] Napoles C., Gormley M., and Van Durme B., (2012) “Annotated gigaword”, In Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, P95–100, Montréal, Canada.
[103] Singh P., Lin T., Mueller E. T., Lim G., Perkins T., and Zhu W. L., (2002) “Open Mind Common Sense: Knowledge acquisition from the general public”, In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", P1223–1237, Rhodes, Greece.
[104] Hobbs J. R., (**1978**) “Resolving pronoun references,” **Lingua**, **44**, **4**, pp**311–338**.
[105] Kantor B. and Globerson A., (2019) “Coreference Resolution with Entity Equalization”, In Proceedings of the 57th Conference of the Association for Computational Linguistics, P673–677, Florence, Italy.
[106] Joshi M., Chen D., Liu Y., Weld D. S., Zettlemoyer L., and Levy O., (**2020**) “Spanbert: Improving pre-training by representing and predicting spans” **Trans. Assoc. Comput. Linguist**., **8**, pp**64–77**.
[107] Xu L. and Choi J. D., (2020) “Revealing the Myth of Higher-Order Inference in Coreference Resolution”, In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), P8527, USA.

[108] Yuan L. ke and Hoey M., (**2014**) "Strategies of writing summaries for hard news texts: A text analysis approach" **Discourse Stud**., **16**, **1**, pp**89–105**.
[109] Ferret O. *et al.*, (**2002**) "How NLP can improve question answering" **KO Knowl. Organ**., **29**, **3–4**, pp**135–155**.
[110] Tripathi S. and Kansal V., (**2019**) "Machine Translation Evaluation: Unveiling the Role of Dense Sentence Vector Embedding for Morphologically Rich Language" **Int. J. Pattern Recognit. Artif. Intell., 34, 01**, p**2059001**.
[111] Hourali S., Zahedi M., and Fateh M., (**2020**) "Coreference Resolution Using Neural MCDM and Fuzzy Weighting Technique," **Int. J. Comput. Intell. Syst**, **13, 1,** pp**56–65.**
[112] Hourali S., Zahedi M., and Fateh M., (**2020**) "A New Model for Coreference Resolution Based on Knowledge Representation and Multi-criteria Ranking," **J. Intell. Fuzzy Syst.**, **40**, pp**1-16**.

Abstract

Coreference resolution is one of the important steps in text processing and one of the basic strategies for improving the performance of text-based systems such as information extraction, summarization, machine translation, and question answering. Coreference resolution deals with identifying all mentions or expressions that refer to the same real-world entity. Mentions may be pronouns, noun phrases and named entities. Numerous studies have focused on the issue of coreference resolution over the past four decades but its accuracy is still not acceptable for using in text comprehension applications, mention detection, candidate antecedent detection, and identifying coreference chains. One of the reasons for difficulty of this issue is requires various knowledge resources, including lexical knowledge, syntactic knowledge, world knowledge, discourse structure, and semantic knowledge. In fact, each type of coreference case requires one or more knowledge resources to be solved. Also, selecting the appropriate reference for existing mentions in the text is a function of the various rules that ensure the matching of gender, number, string structure, and other coreference parameters. In this thesis the problem is modeled in the form of effective parameters of the performance in order to improve precision, recall, and other coreference resolution goals using these features. For this purpose, using newest word embedding methods, recurrent neural network, and attention mechanism, span representation and their features were extracted. Then, mentions and candidate antecedent were extracted with high accuracy using knowledge resource representation, choosing appropriate knowledge and, entities and cluster level information. In addition, candidate antecedents were ranked and coreference chains were extracted using multi-criteria ranking based on Perceptron and Kohonen neural networks. According to the simulation results, we achieved 83.9% and 97.3% Avg. F1 on the English CoNLL-2012 shared task and i2b2 medical dataset respectively. Also, the named entity recognition rate was improved 7.04% on English Gigaword dataset.

Keywords: Natural language processing, Coreference resolution, Recurrent neural network, Multi-criteria ranking, Knowledge resource.

Shahrood University of Technology

Faculty of Computer Engineering

Ph.D Thesis in Artificial Intelligence Engineering

A Deep Learning Based Model for Coreference Resolution of Noun Phrases

By:Samira Hourali

Supervisor:
Dr. Morteza Zahedi

Advisor:
Dr. Mansoor Fateh

November, 2020