



یکپارچه سازی یادگیری چندکاره در مدل های زبانی بر پایه یادگیری عمیق برای بهبود درک زبان طبیعی: مروری بر مقالات

علی رباطی جوزم

مریی فنی، گروه برق و کامپیوتر، دانشگاه ملی مهارت تهران

چکیده

پیشرفت های اخیر در حوزه پردازش زبان طبیعی (NLP)، به ویژه با ظهور مدل های یادگیری عمیق، به بهبود قابل توجه عملکرد در طیف وسیعی از وظایف منجر شده است. با این حال، تعمیم این مدل ها به همه وظایف NLP همچنان چالش برانگیز است. درک زبان طبیعی (NLU) به عنوان زیرمجموعه ای از NLP، بر درک معنایی متن تمرکز دارد و شامل وظایفی مانند ترجمه ماشینی و تحلیل احساسات است. یادگیری چند وظیفه ای (MTL) به عنوان روشی برای آموزش همزمان چندین مدل، پتانسیل بهبود عملکرد در NLU را نشان داده است. این پژوهش با انجام یک مرور سیستماتیک بر مقالات، مدل های زبانی مورد استفاده در NLU را بررسی کرده و نقاط قوت و ضعف آن ها را تحلیل می کند. نتایج این پژوهش نشان می دهد که مدل های زبانی مبتنی بر روش های یادگیری بدون نظارت، پتانسیل بالایی برای بهبود عملکرد دارند. با این حال، طراحی یک چارچوب عمومی برای مدل های زبانی که بتواند به طور همزمان چندین وظیفه NLU را با عملکرد بالا انجام دهد، همچنان یک چالش اساسی است. این پژوهش با ارائه یک چارچوب مفهومی، جهت گیری های آینده تحقیقات در این حوزه را مشخص می کند.

واژگان کلیدی: یادگیری عمیق، نمایش دانش، NLU چند وظیفه ای، یادگیری بدون نظارت



مقدمه

درک زبان طبیعی (NLU) شاخه‌ای از هوش مصنوعی (AI) است که به طور فزاینده‌ای در حال رشد و پیشرفت است. NLU بر توانایی کامپیوتر برای درک و تفسیر زبان انسان تمرکز دارد. در واقع، NLU به کامپیوترها کمک می‌کند تا معنای متن یا گفتار را بفهمند، درست مانند انسان‌ها. این امر شامل تجزیه و تحلیل ساختار جملات، درک معنی کلمات و اصطلاحات، و استخراج اطلاعات کلیدی از متن است. NLU یک موضوع تحقیقاتی نسبتاً جدید است که در آن یک کامپیوتر قبل از انجام وظایف استاندارد NLU مانند بازیابی اطلاعات، پاسخ به سوالات، ترجمه زبان، خلاصه‌سازی متن، طبقه‌بندی اخبار و غیره، اطلاعات را از متن زبان طبیعی تجزیه و تحلیل و استخراج می‌کند. روندهای اخیر در متن‌کاوی نیاز فزاینده استخراج اطلاعات با کیفیت بالا از متن ساختاریافته و بدون ساختار را برآورده می‌کند. برعکس، روندهای اخیر در درک سیستماتیک زبان (SLU) در جهت درک اهداف عملی در متن ورودی، همراه با صحت ساختاری دستوری زبان ورودی است. NLU در حال حاضر در زمینه‌های مختلفی مانند ربات‌های گفتگو، دستیارهای مجازی، جستجوی اطلاعات، ترجمه ماشینی و خلاصه‌سازی متن کاربرد دارد. با پیشرفت تکنولوژی، انتظار می‌رود که کاربردهای NLU در آینده به طور فزاینده‌ای گسترش یابد و نقش مهمی در تعاملات انسان و ماشین ایفا کند.

یکی از کاربردهای محبوب NLU، چت‌بات‌ها هستند. چت‌بات‌ها برنامه‌های کامپیوتری هستند که می‌توانند با انسان‌ها به صورت متنی یا صوتی گفتگو کنند. چت‌بات‌ها به طور فزاینده‌ای در حال تبدیل شدن به ابزاری مهم برای ارائه خدمات به مشتریان هستند. آنها می‌توانند به سوالات متداول پاسخ دهند، اطلاعات مربوط به محصولات و خدمات را ارائه دهند، و حتی مشکلات ساده را حل کنند. طبق آمار خدمات مشتری هوش مصنوعی گartner، چت‌بات‌ها تا سال ۲۰۲۰ مسئول ۸۵٪ از خدمات مشتری خواهند بود. طبق آمار هوش مصنوعی کرانچ بیس، بیش از ۱۰۰۰۰ توسعه‌دهنده اکنون چت‌بات‌هایی را برای فیس‌بوک مسنجر می‌سازند. چت‌بات‌ها همچنین در زمینه‌های دیگر مانند آموزش، سلامت، و سرگرمی نیز کاربرد دارند. به عنوان مثال، می‌توان از چت‌بات‌ها برای آموزش زبان‌های خارجی، ارائه مشاوره پزشکی، یا به عنوان یک همراه مجازی برای افراد سالمند استفاده کرد. با پیشرفت تکنولوژی، چت‌بات‌ها روز به روز هوشمندتر و شبیه‌تر به انسان می‌شوند و انتظار می‌رود که در آینده نقش مهمی در زندگی ما ایفا کنند.

استفاده از چت‌بات‌ها و NLU مزایای زیادی برای کسب‌وکارها دارد. اولاً، چت‌بات‌ها می‌توانند به طور ۲۴/۷ در دسترس باشند و به مشتریان در هر زمان از شبانه‌روز خدمات ارائه دهند. ثانیاً، چت‌بات‌ها می‌توانند به طور همزمان به چندین مشتری پاسخ دهند، که این امر می‌تواند به طور قابل توجهی زمان انتظار برای دریافت خدمات را کاهش دهد. ثالثاً، چت‌بات‌ها می‌توانند با جمع‌آوری اطلاعات از مشتریان، به کسب‌وکارها در درک بهتر نیازها و خواسته‌های آنها کمک کنند. طبق آمار انتظار می‌رود چت‌بات‌ها هزینه‌های تجاری را تا ۸ میلیارد دلار کاهش دهند. این آمارها ظهور چت‌بات‌های مبتنی بر هوش مصنوعی (AI مکالمه‌ای) را در بسیاری از بخش‌های تجاری مانند بانکداری، آموزش، گردشگری، حقوقی و دولتی پیش‌بینی می‌کند که در آن تعامل با مشتری و تجربه مشتری می‌تواند با استفاده از تکنیک‌های مبتنی بر هوش مصنوعی طراحی شود. علاوه بر این، چت‌بات‌ها می‌توانند به طور خودکار وظایف ساده و تکراری را انجام دهند، که این امر می‌تواند به آزادسازی زمان کارکنان برای تمرکز بر وظایف پیچیده‌تر و خلاقانه‌تر کمک کند.

در مجموع، NLU و چت‌بات‌ها در حال تبدیل شدن به ابزارهای ضروری برای کسب‌وکارها در دنیای امروز هستند. با پیشرفت تکنولوژی، انتظار می‌رود که کاربردهای NLU و چت‌بات‌ها در آینده به طور فزاینده‌ای گسترش یابد و نقش مهمی در تعاملات انسان و ماشین ایفا کند. NLU به طور فزاینده‌ای در حال پیشرفت است و به زودی شاهد خواهیم بود که کامپیوترها می‌توانند زبان انسان را به طور کامل درک و تفسیر کنند. این امر می‌تواند به طور قابل توجهی نحوه تعامل ما با کامپیوترها را تغییر دهد و به ما کمک کند تا وظایف خود را به طور کارآمدتر و خلاقانه‌تر انجام دهیم.



جدول ۱. دامنه‌های کاربردی برای NLU.

کاربردها	حوزه
IBM Watson	ترجمه دستگاهی
رزرو بلیط، تنظیم قرار با استفاده از دستیار گوگل	سیستم های کارمحور مبتنی بر مکالمه
خلاصه کننده ها	تحلیل محتوای بزرگ مقیاس
تشخیص طعنه، تحلیل عواطف	دسته بندی متن
Alexa, Cortana	فعال سازی صوتی

هوش مصنوعی به دنبال ساخت ماشین‌هایی است که قادر به درک زبان، تفکر، حل مسئله و خلاقیت باشند [۸]. در این میان، تحقیقات در حوزه NLU (درک زبان طبیعی) نقشی حیاتی ایفا می‌کند. وظایف NLU شامل خواندن، تفسیر و طبقه‌بندی متن است و به سیستم‌هایی نیاز دارد که بتوانند پس از خواندن یک متن، به سوالات پاسخ دهند. این امر مستلزم درک زبان در سطح انسان است. کاربردهای NLU شامل دستیارهای مجازی است که با خواندن اسناد می‌توانند به سوالات مشتریان پاسخ دهند. همچنین NLU در محیط‌های کاری برای پردازش ایمیل‌ها و اسناد و خلاصه‌سازی اطلاعات مفید است. دستیارهای صوتی در منازل نیز می‌توانند به کاربران در تعامل با لوازم خانگی کمک کنند. به طور کلی، NLU به عنوان ابزاری قدرتمند در جهت ارتقای تعامل انسان و ماشین و ارائه خدمات هوشمند به کار می‌رود [۱۰-۱۲].

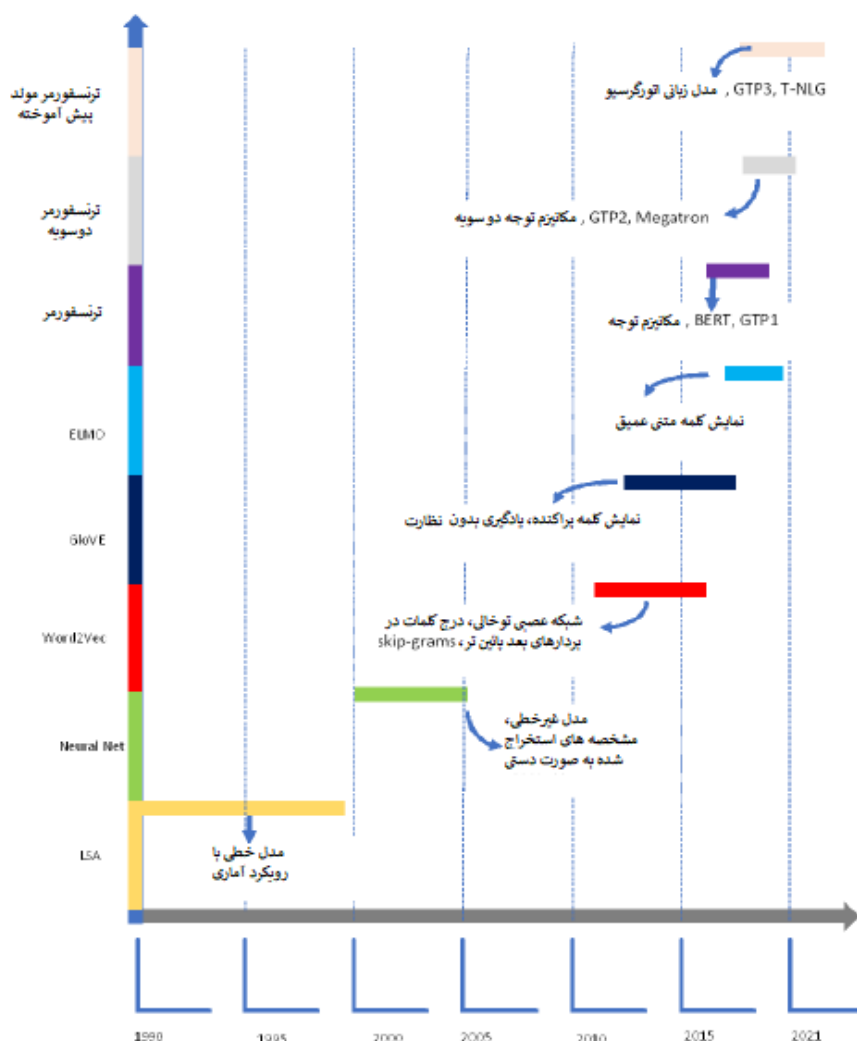
وظیفه اصلی در NLU طبقه‌بندی متن (TC) است. داده‌های متنی از منابع ناهمگنی مانند رسانه‌های اجتماعی و ایمیل‌ها جمع‌آوری می‌شوند. با وجود حجم عظیم داده‌های متنی، استخراج اطلاعات مفید به دلیل ساختار نامنظم متن دشوار است. طبقه‌بندی خودکار متن به دو دسته مبتنی بر قوانین و مبتنی بر هوش مصنوعی تقسیم می‌شود. روش‌های مبتنی بر هوش مصنوعی با استفاده از نمونه‌های از پیش برچسب‌گذاری شده، متن را طبقه‌بندی می‌کنند. در روش‌های کلاسیک یادگیری ماشین، ابتدا ویژگی‌هایی از متن استخراج شده و سپس با استفاده از الگوریتم‌هایی مانند Naive Bayes و SVM متن طبقه‌بندی می‌شود. این روش‌ها به دلیل وابستگی به دانش تخصصی و عدم استفاده از داده‌های آموزشی حجیم، محدودیت‌هایی دارند. روش‌های مبتنی بر شبکه‌های عصبی با نگاشت متن به یک بردار ویژگی پیوسته، نیاز به استخراج دستی ویژگی‌ها را از بین می‌برند. مدل‌های LSA و Word²vec از جمله اولین مدل‌های تعبیه بودند. با پیشرفت تکنولوژی، مدل‌های پیچیده‌تری مانند Elmo و GPT با استفاده از شبکه‌های LSTM و Transformer توسعه یافتند. BERT با ۳۴۰ میلیون پارامتر و آموزش با ۳.۳ میلیارد کلمه، یکی از مدل‌های قدرتمند مبتنی بر Transformer است [۱۴-۱۹]. مدل‌های GPT-۳ و Gshard با میلیاردها پارامتر، جدیدترین نمونه‌های مدل‌های زبانی بزرگ هستند. با وجود عملکرد چشمگیر این مدل‌ها در برخی وظایف NLU، برخی محققان معتقدند که این مدل‌ها فاقد درک واقعی زبان هستند و برای کاربردهای حساس مناسب نیستند [۲۰].

پژوهش‌های پیشین

هدف اصلی این بررسی سیستماتیک ادبیات (SLR) ارزیابی مدل‌های یادگیری عمیق موجود در حوزه NLU برای وظایف چندگانه است. کمبود SLR در موضوع تحقیقاتی NLU مشهود است. بررسی [۱۵] به طور خاص به مدل‌های یادگیری عمیق مبتنی بر ترانسفورمر می‌پردازد و بر تکنیک‌های رمزگذاری دانش و چالش‌های وابستگی زمینه و زبان تمرکز دارد. این مطالعه، مدل BERT را به



عنوان هسته اصلی این نوع مدل‌ها معرفی می‌کند. مطالعه دیگری [۱۵] مروری بر یادگیری چند وظیفه‌ای (MTL) ارائه می‌دهد و رویکردهای عملی برای تنظیمات مختلف یادگیری مانند نظارت‌شده، بدون نظارت و نیمه‌نظارت‌شده را معرفی می‌کند.



شکل ۱. تکامل مدل‌های یادگیری عمیق در NLU.

همچنین یادگیری موازی و توزیع‌شده MTL را برای افزایش سرعت و کارایی پیشنهاد می‌کند. بررسی [۱۶] به بررسی وضعیت فعلی (SOTA) مدل‌های NLP برای وظایف استاندارد NLP می‌پردازد و معماری‌ها، طبقه‌بندی‌ها و ارزیابی‌های مقایسه‌ای NLP، NLU و NLG را ارائه می‌دهد. بررسی [۱۷] به درستی نقش مکانیسم خود توجه و معماری‌های مبتنی بر ترانسفورمر را در بهبود عملکرد مدل‌های زبانی به صورت تصاعدی مورد تأکید قرار می‌دهد. تحقیقات قبلی دارای محدودیت‌هایی مانند تمرکز بر وظایف و معماری‌های خاص، عدم بررسی تعمیم مدل‌های زبانی برای همه وظایف NLU، کمبود بحث در مورد روش‌های نمایش دانش برای وظایف چندگانه و بررسی ابزارهای آنلاین برای ساخت چارچوب عمومی برای NLU چند وظیفه‌ای است. این SLR به بررسی روندها و چالش‌های فعلی در ساخت چارچوب عمومی برای NLU چند وظیفه‌ای، کیفیت مجموعه داده‌های مقایسه‌ای و تکنیک‌های ساخت چنین چارچوبی می‌پردازد.



هیچ SLR جامعی که مدل‌های زبان عمومی برای NLU چند وظیفه‌ای را به طور کامل بررسی کند و مزایا، تجزیه و تحلیل مقایسه‌ای، طبقه‌بندی‌ها و چالش‌ها را پوشش دهد، وجود ندارد. NLU وظایفی مانند استنباط، استنتاج متن، تحلیل احساسات و تشخیص موجودیت نام‌گذاری شده را شامل می‌شود. هدف تحقیقات NLU دستیابی به تسلط بر این وظایف در مجموعه داده‌های مقایسه‌ای استاندارد مانند GLUE و superGLUE است. این تحقیق به دنبال شناسایی و بررسی انتقادی مقالات و خروجی‌های آنها با توجه به سوالات تحقیقاتی است. این SLR، ۹۳ مطالعه اولیه در مورد مدل‌های زبانی در NLU از سال ۲۰۱۱ تا ۲۰۲۱ را شناسایی کرده است. مطالعه‌ای دقیق از مجموعه داده‌های مقایسه‌ای در حوزه عمومی انجام شده و معیاری مناسب برای یک چارچوب عمومی مدل‌های زبانی برای NLU چند وظیفه‌ای ارائه شده است. خلاصه‌ای از ابزارهای آنلاین موجود برای ساخت چنین چارچوبی ارائه شده و شکاف‌های تحقیقاتی شناسایی شده‌اند. این شکاف‌ها مسیرهای آینده تحقیقات NLU را روشن می‌کنند. در نهایت، یک چارچوب مفهومی برای مدل‌های زبان عمومی با رمزگذاری ترانسفورمر بهبود یافته با یادگیری فعال برای NLU چند وظیفه‌ای به عنوان خروجی این SLR پیشنهاد شده است.

جدول ۲. پرسش‌های پژوهشی

پرسش پژوهشی ۱. رویکردهای الگوریتمی متفاوت موجود در ترکیب MTL برای بهبود دقت چارچوب پیشنه‌ای برای NLU کدامند؟
پرسش پژوهشی ۲. کدام یک مجموعه داده‌های آزمایشی استاندارد برای ارزیابی وظایف MTL-NLU هستند؟
پرسش پژوهشی ۳. کدام روش‌های یادگیری برای بهبود عملکرد یادگیری NLU در ترکیب با وظایف مختلف وجود دارند؟
پرسش پژوهشی ۴. کدام روش‌ها برای کاهش نیاز به نمونه‌های بزرگ داده تفسیر شده کارآمد هستند؟
پرسش پژوهشی ۵. کدام روش‌های ارائه دانش رایج برای MTL-NLU وجود دارند؟

روش تحقیق

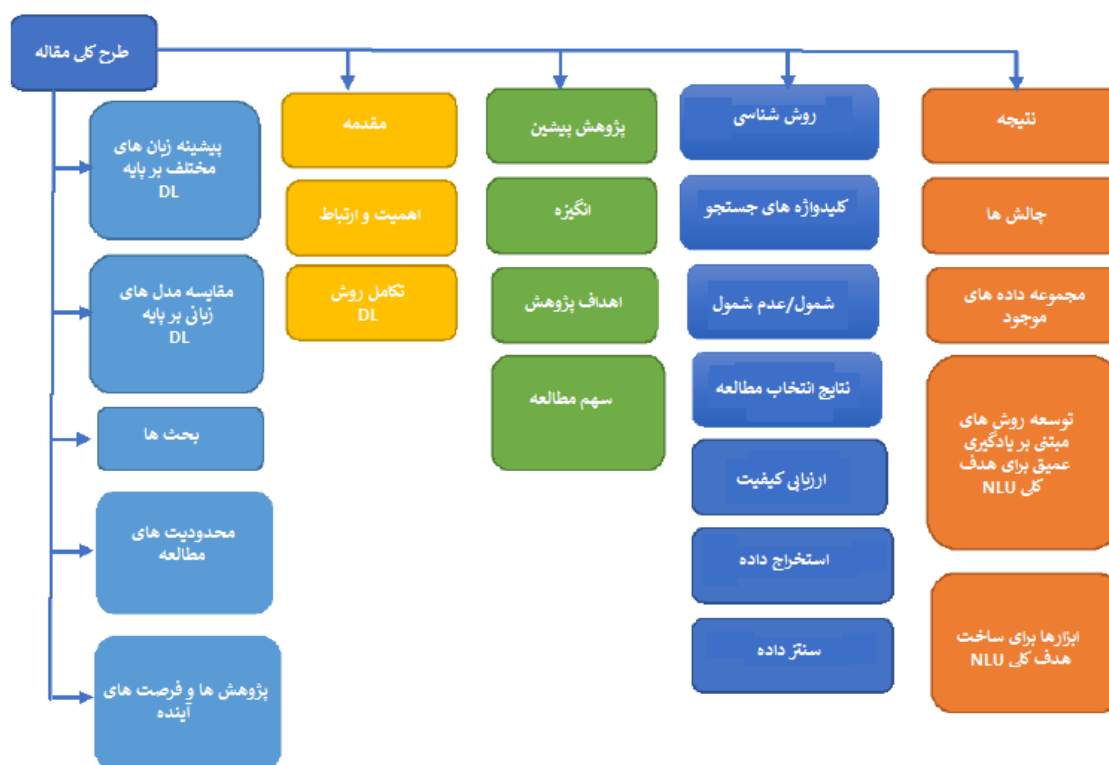
برای تهیه این SLR، از اصول راهنما معرفی شده توسط کیچنهام و چارترز [۱۸] پیروی شد. جدول ۵ تکنیک‌های PIOC (جمعیت، مداخله، نتیجه، زمینه) مورد استفاده برای محصور کردن سوالات تحقیق را نشان می‌دهد. نمودار جریان فرآیند برای این فرآیند در شکل ۳ نشان داده شده است.

معیارهای انتخاب مطالعات تحقیقاتی

عبارات کلیدی برای به دست آوردن نتایج جستجوی مورد نیاز برای پرس و جو در مورد پرسش‌های پژوهشی حوزه انتخاب شدند. رشته جستجو در زیر نشان داده شده است:

(multi-task nlu" OR multi-task nlu framework" OR" natural language understanding" AND learning" OR active learning" OR deep learning" AND unsupervised ("attention model"

نتایج جستجو در جدول ۳ نمایش داده شده است. اگرچه این حوزه از سال ۲۰۰۰ مورد مطالعه قرار گرفته است، اما تمرکز بر روی مقالات از سال ۲۰۱۱ تا ۲۰۲۱ برای نشان دادن توسعه فعلی در این حوزه قرار دارد.



شکل ۲. طرح کلی مقاله.

جدول ۳. نتیجه جستجوی پایگاه داده های ادبیات.

پایگاه داده منبع	تعداد کل	تعداد مرتبط پس از بکارگیری معیار شمول/عدم شمول و حذف تکراری ها
Scopus	۱۹۷	۴۸
Web of science	۷۱	۳۳
Emerald	۷۲	۰۲
IEEE Xplore	۰۳	۰۱
Google Scholar	۴۳۸	۲۲
مجموع	۷۸۱	۱۰۶

معیارهای شمول و عدم شمول

مقالات تحقیقاتی مورد بررسی برای این SLR باید مرتبط باشند. مطالعات از بهبود تکنیک ها، ساخت چارچوب ها برای NLU و پاسخگویی به حوزه های کاربردی مختلف متغیر بودند. زبان انتخابی انگلیسی است و ممکن است بررسی همتا شده باشد. معیارهای شمول و حذف برای انتخاب مقالات به شرح زیر است:

نتایج انتخاب مطالعه



نمودار جریان برای انتخاب مقالات مرتبط برای این SLR در شکل ۳ نشان داده شده است. رشته جستجو برای انتخاب ۷۸۱ مقاله از پایگاه‌های داده مختلف ذکر شده در جدول ۴ انتخاب شد. پس از حذف موارد تکراری و اعمال معیارهای شمول و حذف، ۱۰۶ مقاله تحقیقاتی برای این SLR در نظر گرفته شد. با استفاده از تکنیک‌های گلوله برفی برای درج سهم‌های قابل توجه، مجموع به ۱۱۵ افزایش یافت. در نهایت، پس از اعمال معیارهای ارزیابی کیفیت، ۱۰۲ مطالعه برای تهیه یک مرور سیستماتیک ادبیات انتخاب شدند.

جدول ۴. معیار شمول و عدم شمول

شماره معیار	موضوع	معیار شمول	معیار عدم شمول
۱	مدل‌های زبانی	تمرکز بر کارهای NLU	تمرکز صرف بر کارهای NLU
۲	یادگیری چندکاره	تمرکز بر ترکیبی از کارها	تمرکز صرف بر یک کار
۳	بازه سال	۲۰۲۱-۲۰۱۱	پیش از سال ۲۰۱۱
۴	مرتبط بودن پرسش پژوهشی	مرتبط با حداقل یک پرسش پژوهشی	نامرتبط با پرسش‌های پژوهشی

معیارهای ارزیابی کیفیت برای انتخاب مطالعه

معیارهای ارزیابی کیفیت، ارتباط مقالات تحقیقاتی را برای پاسخگویی به پرسش‌های پژوهش تضمین می‌کنند. مطالعات تحقیقاتی بر اساس معیارهای ذکر شده در جدول ۶ به عنوان ۱ یا ۰ درجه بندی شدند.

امتیاز ۴ کیفیت برای انجام این SLR در نظر گرفته شده است. جدول ۹ معیارهای ارزیابی کیفیت را فهرست می‌کند.

استخراج داده‌ها

بررسی اجمالی فرآیند استخراج داده‌ها برای پاسخ به سوالات تحقیقاتی از مطالعات در جدول ۷ فهرست شده است.

سنتز داده‌ها

جدول ۷ سنتز داده‌ها را برای پاسخگویی مفصل به سوالات تحقیقاتی نشان می‌دهد.

پیشینه تحقیق

درک زبان طبیعی شامل ساخت مدل‌های زبانی، آموزش این مدل‌ها و آزمایش دقت آنها است. وظایف مختلف NLU، مانند پاسخ به سوالات و NLI، را می‌توان به عنوان یک مسئله طبقه‌بندی مطرح کرد. این بخش وظایف TC ارائه شده در این مطالعه را ارائه می‌کند. تحلیل احساسات روشی برای استخراج قطبیت و چشم‌انداز نظرات مشتریان است. این مشکل را می‌توان به عنوان یک مسئله دو کلاسه یا چند کلاسه بیان کرد. یک سیستم طبقه‌بندی اخبار می‌تواند با تشخیص موضوعات نوظهور یا ارائه پیشنهادات خبری مناسب بسته به ترجیحات خواننده، به مصرف‌کنندگان در انتخاب اخبار مرتبط در حال حاضر کمک کند. طبقه‌بندی موضوع یک کار است که شامل تعیین موضوع کلی یا عنوان یک سند است، خواه بررسی فیلم در مورد رتبه‌بندی بیننده باشد یا درآمد ناخالص در دوره مشخص. پرسش و پاسخ (QA) وظایف QA استخراجی و مولد انواع وظایف QA هستند. استخراج Take extents در یک سند در [۱۹] SQUAD به عنوان نمونه‌ای از یک کار طبقه‌بندی متن برای یک سوال و مجموعه‌ای از یک پاسخ مناسب است. الگوریتم پاسخ هر نامزد را به عنوان



صحیح یا نادرست طبقه‌بندی می‌کند. QA-NLI پیش‌بینی می‌کند که آیا می‌توان معنای یک متن را از معنای متن دیگر پیش‌بینی کرد یا خیر. یک برچسب متعلق به تناقض استنباط است و توسط یک سیستم NLI به یک جفت واحد متن تعلق ندارد [۲۰].

وظایف QA استخراجی و QA مولد دو نوع وظیفه QA هستند. وظایف استخراجی در طول طول سند در [۲۳] SQUAD به عنوان نمونه‌ای از یک کار TC با توجه به یک سوال و مجموعه‌ای از پاسخ‌های احتمالی پخش می‌شوند. الگوریتم پاسخ هر نامزد را به عنوان صحیح یا نادرست طبقه‌بندی می‌کند. این مطالعه فقط به QA استخراجی، یک وظیفه ایجاد متن که پاسخ‌ها را در پرواز تولید می‌کند، می‌پردازد.

NLI پیش‌بینی می‌کند که آیا می‌توان معنای یک متن را از معنای متن دیگر پیش‌بینی کرد. مقایسه جفت متن نوعی تعمیم‌یافته از NLI است که پارافریز نامیده می‌شود. مسئله تعیین اینکه چقدر احتمال دارد یک جمله از جمله دیگر پارافریز شود با مقایسه شباهت معنایی دو جمله است.

جدول ۵. اطلاعات PIOC (جمعیت، مداخله، نتیجه، زمینه).

پارامتر	معنا	کلیدواژه‌ها
جمعیت	حوزه‌های کاربرد	«درک زبان طبیعی» یا «NLU»
مداخله	روش‌ها/چارچوب نرم افزاری	«یادگیری فعال» یا «یادگیری نظارت نشده»
خروجی	جزء اهمیت	«چارچوب NLU چندکاره»
زمینه	زمینه مداخله	مدل توجه

جدول ۶. معیارهای ارزیابی کیفیت.

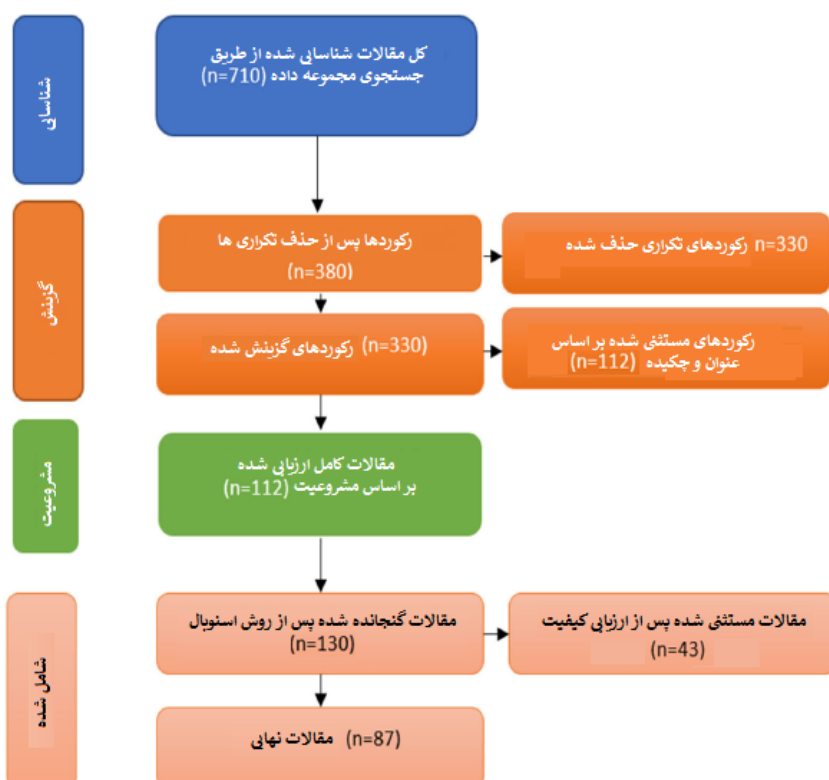
معیار	امتیاز
مطالعه نتیجه می‌دهد	۱ اگر معیار صحیح باشد، در غیر این صورت ۰
مطالعه اثبات تجربی ارائه می‌کند	۱ اگر معیار صحیح باشد، در غیر این صورت ۰
مطالعه شامل اهداف و نتایج می‌شود	۱ اگر معیار صحیح باشد، در غیر این صورت ۰
مطالعه مراجع مناسبی ارائه می‌کند	۱ اگر معیار صحیح باشد، در غیر این صورت ۰

جدول ۷. طبقه‌بندی مطالعات انتخاب‌شده برای پاسخ به سوالات تحقیقاتی.

دسته بندی	داده‌های مرتبط استخراج شده از مطالعات انتخابی
روش‌های مورد استفاده برای ترکیب کارهای مرتبط با متن در NLU چندکاره	چالش‌ها برای کار با توازن در عین ترکیب کارها در NLU چندکاره



Fastformers, BERT, ALBERT, RoBERTa, OpenGTP, ToyBERT	روش های ارتقا یافته کدگذاری ترنسفورمر
	مجموعه داده و روش های اعتبارسنجی
MT-DN, Emie, Xlnet, UniLM	ترکیب روش های برای ترجمه دستگاہی چندکاره و رمزگذاری ترنسفورمر
مجموعه داده های آزمایشی در دسترس عموم، مدل های توسعه داده شده توسط پژوهشگران حوزه NLU	مجموعه داده های آزمایش و مدل های توسعه داده شده توسط پژوهشگران حوزه NLU



شکل ۳. نمودار جریان برای انتخاب مقالات مرتبط. ترجمه ماشینی RNN یک مدل اساسی حیاتی است و پیشرفت های زیادی در توپولوژی های پیشرفته شبکه و روش های آموزشی مدل های منحصر به فرد وجود داشته است.

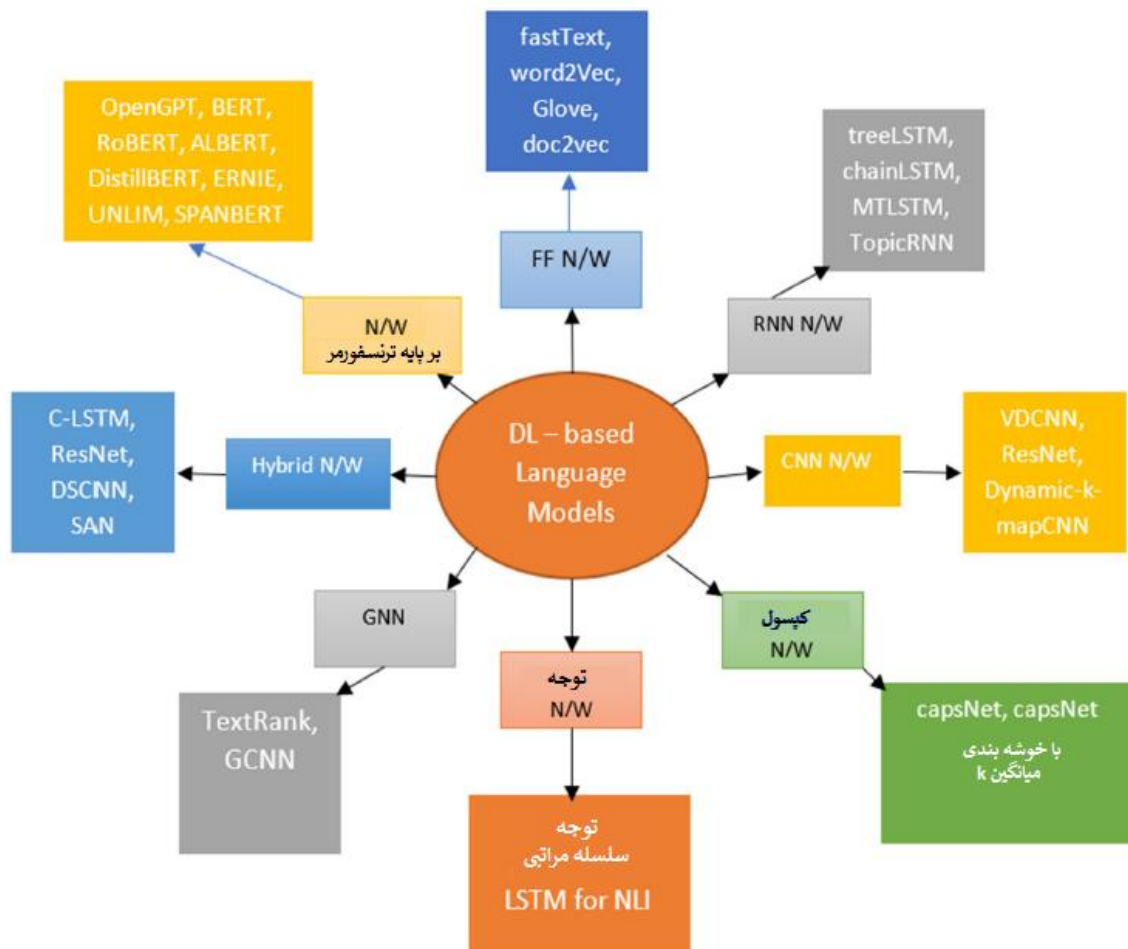
ترجمه ماشینی عصبی - هدف ترجمه ماشینی عصبی ترجمه متن با شبیه سازی قابلیت های مغز انسان است. هدف ترجمه یک زبان منبع معین به یک زبان هدف با حفظ معنای آن و هدف آن است. هنگام ترجمه، مغز انسان ابتدا جمله را درک می کند، سپس یک نمایش ذهنی از جمله می سازد و در نهایت این نمایش ذهنی را به جمله ای در زبان دیگر تبدیل می کند.

ترجمه ماشینی عصبی با دو فرآیند مدولار رمزگذاری و رمزگشایی، فرآیند ترجمه انسانی را تقلید می کند. رمزگذار عبارات زبان منبع را به نمایش های برداری فضای معنایی تبدیل می کند. بسته به بردارهای معنایی

درک خواندن ماشین وظیفه خواندن و درک متن توسط یک برنامه کامپیوتری است. این تلاش نیازمند ایجاد سیستم هایی است که بتوانند پس از خواندن یک سند به پرسش ها پاسخ دهند. این نیاز کاربردهای زیادی دارد، از جمله امکان ارائه پاسخ های هوشمند و



صحیح به پرسش‌های زبان طبیعی توسط موتورهای جستجو با خواندن متن‌های مرتبط. علاوه بر این، ماشین‌های درک خواندن را می‌توان در دستیاران مجازی برای پاسخگویی به سوالات پشتیبانی مشتری پس از خواندن اسناد استفاده کرد. همچنین می‌توان از آن در محیط کار برای کمک به کاربران در خواندن و پردازش ایمیل‌ها یا اسناد تجاری و همچنین خلاصه کردن اطلاعات مرتبط استفاده کرد. استفاده از مجموعه داده‌های بزرگ مقیاس و با برچسب‌گذاری دستی به بهبودهای اخیر در درک خواندن ماشین کمک کرده است. این بخش به بررسی مدل‌های مختلف یادگیری عمیقی که برای مشکلات طبقه‌بندی متن ارائه شده‌اند می‌پردازد. بر اساس ساختار مدل‌های خود، این مدل‌ها در دسته‌های زیر قرار می‌گیرند: (۱) شبکه‌های پیش‌خور: متن به عنوان یک کیسه کلمات در نظر گرفته می‌شود (بخش A). (۲) مدل‌های مبتنی بر RNN: برای حدس وابستگی‌های کلمات با مشاهده متن به عنوان کلماتی در یک ترتیب خاص استفاده می‌شوند. (بخش B). (۳) مدل‌های مبتنی بر CNN: برای شناسایی سبک‌های متن برای طبقه‌بندی متن آموزش داده می‌شوند. (بخش C). (۴) شبکه‌های کپسولی: با مشکل از دست رفتن اطلاعات در عملیات ترکیب CNS مقابله می‌کنند و برای طبقه‌بندی متن استفاده شده‌اند. (بخش D). (۵) مکانیسم توجه: در ساخت مدل‌های Dell مفید است زیرا در شناسایی کلمات مرتبط در متن موثر است. (بخش E). (۶) شبکه‌های عصبی گراف: برای نمایش ساختار گراف ذاتی زبان طبیعی طراحی شده‌اند. (بخش F). (۷) مدل‌های ترکیبی: توجه و متون را برای ضبط ویژگی‌های محلی و جهانی ترکیب می‌کنند. (بخش G). (۸) ترانسفورمرها: عمدتاً برای موزایی‌سازی بسیار بیشتر از RNN ها استفاده می‌شوند و امکان آموزش مدل‌های زبانی عظیم مبتنی بر GPU را فراهم می‌کنند. (بخش H). این انواع مختلف مدل‌ها بر اساس تکنیک‌های یادگیری عمیق در شکل ۴ نشان داده شده‌اند.



شکل ۴. انواع مدل‌های زبانی مبتنی بر یادگیری عمیق.



مدل‌های مبتنی بر شبکه‌های پیش‌خور

مدل‌های ساده DL برای نمایش متن شامل شبکه‌های پیش‌خور هستند. با وجود این، آنها سطح خوبی از دقت را در چندین معیار مقایسه‌ای TC دارند. متن در این مدل‌ها به عنوان مجموعه‌ای از کلمات مشاهده می‌شود. این مدل‌ها با استفاده از [۲۱] word2vec یا GloVe [۲۲] یک نمایش برداری برای هر کلمه به دست می‌آورند. اینها مدل‌های تعبیه محبوب هستند. جولین و همکاران [۲۳] یک طبقه‌بندی‌کننده دیگر به نام fastText معرفی کردند. کارآمد و ساده است. مجموعه‌ای از n-gram ها به عنوان یک ویژگی مکمل در fastText برای شناخت اطلاعات در مورد ترتیب کلمات محلی استفاده می‌شود. این تکنیک با گزارش نتایج قابل مقایسه با روش‌هایی که از ترتیب کلمات استفاده می‌کنند، کارآمد ثابت شده است. لی و میکولوف [۲۵] doc2vec را معرفی کردند، روشی برای یادگیری نمایش‌های ویژگی با طول ثابت از متن با طول متغیر، با استفاده از یک الگوریتم بدون نظارت.

مدل‌های مبتنی بر RNN

معمولاً متن به عنوان ترتیبی از کلمات در مدل‌های مبتنی بر RNN درمان می‌شود. هدف اصلی یک مدل مبتنی بر RNN برای طبقه‌بندی متن، ضبط روابط کلمات بین جملات و ساختار متن است. از سوی دیگر، مدل‌های ساده مبتنی بر RNN به خوبی شبکه‌های عصبی پیش‌خور استاندارد عمل نمی‌کنند. مدل حافظه کوتاه بلند مدت (LSTM) یکی از بسیاری از تغییرات RNN است که برای به دست آوردن وابستگی‌های بلندمدت کلمات جملات در نظر گرفته شده است.

یک سلول حافظه با دروازه‌های ورودی، خروجی و فراموشی اختراع شده است تا مقادیر را در یک بازه زمانی مشخص به خاطر بسپارد. مدل‌های LSTM با استفاده از این سلول حافظه، مشکلات گرایان ناپدید شونده و انفجار گرایان را که RNN‌های ساده از آن رنج می‌برند، برطرف می‌کنند. تای و همکارانش [۲۶] برای یادگیری نمایش‌های معنایی غنی، یک مدل درخت-LSTM ساختند که تعمیمی از توپولوژی‌های شبکه ساختارمند LSTM است. با توجه به ساختارهای نحوی زبان طبیعی، این مدل کلمات را برای تشکیل عبارات ترکیب می‌کند و بنابراین، درخت-LSTM برای وظایف NLP نسبت به زنجیره-LSTM کارآمدتر است. آنها عملکرد درخت-LSTM‌ها را در دو کار نشان می‌دهند: تحلیل احساسات و پیش‌بینی رابطه معنایی بین دو جمله. در مرجع [۲۷] ژو و همکارانش عملکرد زنجیره-LSTM ساختارمند را نسبت به پیشینیان خود با ذخیره بسیاری از سلول‌های جانشین با استفاده از یک فرآیند بازگشتی در سلول‌های حافظه بهبود بخشیدند.

با ضبط اطلاعات مفید با بازه‌های زمانی مختلف، شبکه عصبی LSTM چند مقیاسی زمانی [۲۸] (MT-LSTM) برای نمایش متون معناداری مانند جملات و اسناد ساخته شده است. لایه‌های مخفی یک مدل LSTM معمولی توسط MT-LSTM به گروه‌های متعددی تقسیم می‌شوند. در زمان‌های مختلف، هر گروه فعال و به‌روز می‌شود. MT-LSTM می‌تواند با موفقیت اسناد متعددی را در مدل جای دهد. گزارش شده است که MT-LSTM بر مجموعه‌ای از خطوط پایه، از جمله مدل‌های طبقه‌بندی متن مبتنی بر LSTM و RNN، غلبه می‌کند. RNN‌ها در حفظ ساختار محلی ترتیب کلمات بهتر عمل می‌کنند، اما در وابستگی‌های نوع بلندمدت دچار مشکل می‌شوند. برای تحلیل احساسات، گفته می‌شود که TopicRNN از خط پایه RNN عملکرد بهتری دارد. مدل‌های مبتنی بر RNN دیگری نیز جذاب هستند. لیو و همکارانش [۲۹] از یادگیری چند وظیفه‌ای برای آموزش RNN‌ها استفاده می‌کنند تا از داده‌های برجسب‌گذاری شده برای آموزش از وظایف مختلف مرتبط استفاده کنند.

مدل‌های مبتنی بر CNN

شبکه‌های عصبی کانولوشن (CNN) برای شناسایی الگوها در فضا و شبکه‌های عصبی بازگشتی (RNN) برای تشخیص الگوها در طول زمان آموزش داده می‌شوند [۳۰]. RNN‌ها در وظایف پردازش زبان طبیعی (NLP) مانند برجسب‌گذاری (POS) که نیاز به درک معنایی بلندمدت دارند، عملکرد خوبی دارند. اما CNN در مواردی که تشخیص الگوهای محلی و مستقل از مکان در سند حیاتی است، مانند تشخیص جملات معناداری که احساس خاصی را منتقل می‌کنند، عملکرد بهتری دارد. به همین دلیل، CNN به انتخاب محبوب

برای طراحی مدل‌های رایج طبقه‌بندی متن تبدیل شده است. کالچرر و همکاران [۳۱] یک الگوریتم طبقه‌بندی متن بر اساس CNN به نام Dynamic K-Max Pooling ارائه کردند که ابتدا معیارهای جمله را تولید می‌کند و سپس با استفاده از ساختار کانولوشنال و لایه‌های استخراجی پویا K-Max، نقشه‌ای از ویژگی‌ها در سراسر جمله ایجاد می‌کند که درجات مختلف ارتباط بین اصطلاحات را ثبت می‌کند. کیم [۳۲] یک مدل ساده‌تر مبتنی بر CNN برای طبقه‌بندی متن پیشنهاد کرد و چهار روش مختلف برای یادگیری تعبیه کلمات (CNN-static، CNN-non-static، CNN-multi-channel و CNN-rand) را با هم مقایسه کرد. این مدل‌ها در بررسی احساسات و طبقه‌بندی سوالات عملکرد بهتری نسبت به مدل‌های قبلی دارند. لیو و همکاران [۳۳] یک مدل جدید مبتنی بر CNN ارائه کردند که با استفاده از رویکرد استخراجی پویا حداکثر و لایه محدودکننده پنهان، دقت را بهبود می‌بخشد. محققان به جای استفاده از بردارهای کلمه با ابعاد پایین، از متن با ابعاد بالا برای شناخت تعبیه‌های نواحی کوتاه متن در طبقه‌بندی استفاده می‌کنند. پروسا و خوش‌گفتار [۳۴] روشی برای رمزگذاری متن ورودی با استفاده از CNN‌ها ارائه کردند که حافظه و زمان آموزش را به طور قابل توجهی کاهش می‌دهد. مطالعاتی مانند کونه و همکاران [۳۵] که مدل VDCNN را معرفی کردند، نشان دادند که عملکرد با افزایش عمق مدل بهبود می‌یابد. دک و همکاران [۳۷] با تغییر ساختار VDCNN، اندازه مدل را تا ۲۰ برابر کاهش دادند. گو و همکاران [۳۸] استفاده از تعبیه‌های کلمات وزنی در مدل CNN با کانال‌های چندگانه را توصیه کردند. ژانگ و والاس [۳۹] به این نتیجه رسیدند که GloVe و word²vec عملکرد بهتری نسبت به بردارهای یک‌گرم دارند و استخراج حداکثر بهترین روش در بین رویکردهای استخراجی موجود است.

مدل‌های مبتنی بر شبکه‌های عصبی کپسولی (Capsule Neural Network)

CNN‌ها از چندین لایه کانولوشن و استخراجی برای طبقه‌بندی تصاویر یا متن استفاده می‌کنند. عملیات استخراجی ویژگی‌های مهم را تشخیص می‌دهند و پیچیدگی محاسباتی فرآیندهای کانولوشن را کاهش می‌دهند، اما آنها اطلاعات فضایی را از دست می‌دهند و ممکن است بسته به جهت گیری یا نسبت آنها، اقلام را اشتباه طبقه‌بندی کنند. هینتون و همکاران [۴۰] یک تکنیک جدید به نام شبکه‌های کپسولی را برای غلبه بر چالش‌های استخراجی ارائه کردند. یک کپسول مجموعه‌ای از نورون‌ها است که بردار فعالیت آن بسیاری از ویژگی‌های یک بلوک یا بخشی از بلوک را نشان می‌دهد. طول بردار احتمال وجود بلوک را نشان می‌دهد و جهت بردار ویژگی‌های بلوک را نشان می‌دهد. کپسول‌ها هر کپسول را از لایه پایین به بهترین کپسول والد مناسب خود در لایه بالا هدایت می‌کنند، با اطلاعات موجود در شبکه تا آخرین لایه برای طبقه‌بندی متوقف خواهد شد. این کار را می‌توان با استفاده از الگوریتم‌های مختلفی مانند مسیریابی پویا-توسط-توافق [۴۱] یا الگوریتم انتظار-حداکثر [۴۲] انجام داد. این شبکه‌ها اخیراً برای طبقه‌بندی متن استفاده شده‌اند، با کپسول‌هایی که با نمایش یک خط در سند یا کل سند به شکل یک بردار سفارشی شده‌اند. کیم و همکاران [۴۳] یک مدل مبتنی بر CapsNet با معماری قابل مقایسه توسعه دادند. این مدل از چهار لایه تشکیل شده است: ۱) اسناد به عنوان یک سری از تعبیه‌های کلمات توسط لایه ورودی پذیرفته می‌شوند. ۲) نقشه ویژگی توسط یک لایه کانولوشن ساخته می‌شود و از یک واحد خطی دروازه‌دار برای حفظ اطلاعات خاص استفاده می‌کند. ۳) ویژگی‌های محلی توسط یک لایه کپسول کانولوشن جمع‌آوری می‌شوند. ۴) و در نهایت، یک لایه کپسول متن که برچسب‌های کلاس را پیش‌بینی می‌کند. طبق گفته نویسندگان، اشیاء را می‌توان آزادتر در متن نسبت به تصاویر ساخت. در مقابل موقعیت‌ها، معنای یک سند می‌تواند بدون تغییر باقی بماند، اگرچه ترتیب برخی از خطوط سند تغییر می‌کند. رن و لو [۴۴] یک تغییر جدید از CapsNets را ارائه کردند که از یک سبک معماری ترکیبی بین کپسول‌ها و یک الگوریتم مسیریابی مبتنی بر خوشه‌بندی k-means جدید استفاده می‌کند. ابتدا، تمام بردارهای کلمه کد در کتاب‌های کد برای ایجاد تعبیه‌های کلمه استفاده می‌شوند. سپس ویژگی‌های کپسول‌های سطح پایین با استفاده از مسیریابی k-means در کپسول‌های سطح بالا ادغام می‌شوند.

مدل‌های با مکانیسم توجه

نحوه توجه یک فرد به بخش‌های مختلف یک عکس یا کلمات مرتبط یک جمله، توجه را انگیزه می‌دهد. توجه به یک مفهوم و ابزار مرکزی در توسعه مدل‌های DL برای NLP تبدیل شده است [۴۵]. می‌توان بردار وزن‌های معنادار را به عنوان یک بردار از وزن‌های معنادار در یک مغز آجیل در نظر گرفت. برای حدس زدن یک کلمه در یک جمله، با استفاده از بردار توجه، تخمین زده می‌شود که چقدر قویاً با کلمات دیگر مرتبط است و با اضافه کردن مقادیر وزنی بردار توجه، مقدار هدف پیش‌بینی می‌شود. این بخش برخی از گسترده‌ترین مدل‌های توجه را بررسی می‌کند که به ایجاد یک چارچوب ذهنی جدید کمک می‌کنند. یانگ و همکاران [۴۶] یک شبکه مبتنی بر مکانیسم توجه سلسله مراتبی برای طبقه‌بندی متن پیشنهاد کردند. این مدل دارای دو ویژگی متمایز است: یک ساختار سلسله مراتبی که ساختار سلسله مراتبی سند را تقلید می‌کند و دو سطح از مکانیسم‌های توجه که هم در سطح کلمه و هم در سطح جمله اعمال می‌شوند و به آن اجازه می‌دهند تا در حین ساخت نمایش سند، توجه متفاوتی به قسمت‌های بیشتر و کمتر معنادار داشته باشد. این مدل در شش آزمایش طبقه‌بندی متن، از روش‌های قبلی با حاشیه قابل توجهی پیشی می‌گیرد. رویکرد توجه سلسله مراتبی توسط ژو و همکاران [۴۷] به طبقه‌بندی احساسات بین زبانی گسترش یافت. سند با استفاده از یک شبکه LSTM در هر زبان مدل‌سازی می‌شود. پس از آن، تحلیل نهایی احساسات با استفاده از توجه سلسله مراتبی در مدل سطح جمله انجام می‌شود. از سوی دیگر، مدل‌های توجهی که در سطح کلمه طراحی شده‌اند، کلمات مهم را در هر مرحله درک می‌کنند. شن و همکاران [۴۸] مدلی را کشف کردند که دارای مکانیسم خود توجه برای NLP است، با توجه جهت‌دار و چند بعدی بین عناصر از توالی‌های ورودی. برای یادگیری تعبیه جمله، یک شبکه عصبی با وزن بالا استفاده می‌شود که منحصراً به نوع توجه بستگی دارد و شامل CNNها یا RNNها نیست. لیو و همکاران [۴۹] یک مدل LSTM برای NLI ارائه کردند که شامل توجه داخلی است. این مدل برای رمزگذاری یک متن از یک روش دو مرحله‌ای استفاده می‌کند. در مرحله اول، نمایش جمله با استفاده از استخراجی میانگین در سطح کلمه Bi-LSTM تولید می‌شود. پس از آن، از یک مکانیسم توجه برای تعویض استخراجی میانگین بر روی همان عبارت با نمایش‌های برتر استفاده می‌شود. این تکنیک نمایش جمله را در مرحله اول افزایش داد و بیشتر از توجه برای پردازش متن استفاده کرد.

مدل‌های مبتنی بر شبکه‌های عصبی گراف

اگرچه متون معمولی دارای ترتیب سریالی هستند، اما همچنین شامل ساختارهای گراف نهفته‌ای مانند درخت‌های تجزیه هستند که روابط را بر اساس نحو و معنای جملات حدس می‌زنند. [۵۰] TextRank یک مدل NLP مبتنی بر گراف اصلی است که توسعه یافته است. نویسندگان پیشنهاد می‌کنند که یک متن زبان طبیعی به عنوان یک گراف $G(V, E)$ نشان داده شود، که در آن V نشان‌دهنده مجموعه‌ای از گره‌ها و E نشان‌دهنده مجموعه‌ای از لبه‌ها بین گره‌ها است. گره‌ها واحدهای متنی مختلفی را مشخص می‌کنند که جملات را تکمیل می‌کنند. بسته به نوع کاربردها، لبه‌ها نیز می‌توانند برای بیان اشکال مختلف روابط بین گره‌ها استفاده شوند. شبکه‌های عصبی گراف (GNNs) معاصر با اقتباس از روش‌های DL ایجاد می‌شوند. TextRank یکی از این روش‌ها برای داده‌های گراف است. در طول چند سال گذشته، مدل‌های DNN مبتنی بر CNNها، RNNها و خودرمزگذارها برای رسیدگی به پیچیدگی داده‌های گراف اقتباس شده‌اند [۵۱]. [۵۲] GCNs و مشتقات آنها در بین انواع مختلف GNNs بسیار رایج هستند زیرا عملیاتی هستند و به راحتی با شبکه‌های دیگر ترکیب می‌شوند و به نتایج مطلوب در کاربردهای مختلف رسیده‌اند. GCNs در گراف‌ها نوع کارآمدتری از CNNها هستند. برای یادگیری نمایش‌های گراف، GCNs لایه‌های فیلترهای طیف مرتبه اول دنبال شده توسط یک تابع فعال‌سازی غیرخطی را انباشته می‌کنند. TC یک کاربرد رایج GNNs در NLP است. برای استنباط برچسب‌های سند، GNNs از روابط متقابل اسناد یا کلمات استفاده می‌کنند [۵۳]. یائو، و همکاران [۵۴] از یک مدل GCNN استفاده کردند که از CNN با شبکه‌های گراف استفاده می‌کند. آنها پس از ایجاد یک گراف متن واحد، یک شبکه کانولوشنال متن گراف (Text GCN) را برای مجموعه داده یاد می‌گیرند. این بر اساس کلماتی است که با هم رخ می‌دهند و روابط کلمات بین سند است. Text GCN با یک نمایش یک-گرم از هر کلمه در سند شروع به یادگیری می‌کند و سپس تعبیه‌هایی را برای اسناد و اسناد بیشتر یاد می‌گیرد. این روش توسط حاشیه‌نویسی‌های کلاس سند شناخته شده نظارت می‌شود. آموزش GNNs برای یک مجموعه داده متن عظیم گران است. تلاش‌هایی برای کاهش هزینه مدل‌سازی با حذف پیچیدگی مدل یا تطبیق تکنیک‌های آموزش مدل انجام شده است.



مدل‌های با تکنیک‌های ترکیبی

بسیاری از مدل‌های ترکیبی برای تشخیص اسناد جهانی و محلی با ترکیب معماری‌های LSTM و CNN ساخته شده‌اند. ژو و همکاران [۵۵] یک شبکه کانولوشنال LSTM (C-LSTM) را پیشنهاد کردند. LSTM-C از یک CNN برای استخراج یک آرایش از نمایش‌های عبارت (n-gram) استفاده می‌کند و سپس ورودی را به یک شبکه LSTM برای تولید نمایش سطح جمله می‌دهد. ژانگ و همکاران [۵۶] برای مدل‌سازی سند، یک CNN حساس به وابستگی (DSCNN) را پیشنهاد می‌کنند. چن و همکاران [۵۷] از یک مدل CNN-RNN برای انجام TC چند برچسب استفاده کردند. تانگ و همکاران [۵۸] از یک CNN برای درک نمایش‌های جمله استفاده می‌کنند که روابط ذاتی بین جملات را رمزگذاری می‌کند. ژیاو و چو [۵۹] یک سند را در ترتیب خاص کاراکترها به جای کلمات در نظر گرفتند و رمزگذاری اسناد را با استفاده از کانولوشن مبتنی بر الفبا با لایه‌های بازگشتی پیشنهاد کردند. در مقایسه با مدل‌های سطح کلمه، مدل پیشنهادی با پارامترهای بسیار کمتر به نتایج قابل مقایسه‌ای دست یافت.

Recurrent CNN [۶۰] برای یادگیری نمایش‌های کلمات از یک ساختار بازگشتی برای تشخیص وابستگی زمینه ای بلندمدت استفاده می‌کند. برای کاهش شلوغی، استخراج حداکثر برای انتخاب خودکار مهم‌ترین اصطلاحات برای کار طبقه‌بندی متن استفاده می‌شود. در درک خواندن ماشین، لیو و همکاران [۶۱] یک شبکه پاسخ استوکاستیک مقاوم (SAN) برای استدلال با استفاده از یک رویکرد چند مرحله‌ای پیشنهاد کردند. SAN انواع مختلفی از شبکه‌های عصبی را در بر می‌گیرد، مانند شبکه‌هایی با حافظه، LSTM-Bi، CNN، توجه و ترانسفورم‌ها همگی نمونه‌هایی از تبدیل‌ها هستند. نمایش‌های زمینه برای سیستم پاسخ به سوالات برای گذرگاه‌ها با استفاده از مؤلفه Bi-LSTM به دست می‌آیند. روش توجه آن یک نمایش گذرگاه را تولید می‌کند که آگاهانه به سوال است. سپس گذرگاه در یک حافظه کاری که توسط یک واحد بازگشتی دروازه‌دار ایجاد شده است، ذخیره می‌شود. در نهایت، پیش‌بینی‌ها توسط یک ماژول پاسخ که از یک واحد بازگشتی دروازه‌دار استفاده می‌کند، تولید می‌شوند. ترکیب شبکه‌های بزرگراهی با RNN و CNN‌ها موضوع چندین مطالعه بوده است. اطلاعات لایه به لایه در شبکه‌های عصبی چند لایه معمولی حرکت می‌کنند. با افزایش عمق، آموزش DNN مبتنی بر گرادین پیچیده‌تر می‌شود. شبکه‌های بزرگراهی [۶۲] برای ساده‌تر کردن آموزش شبکه‌های عصبی عمیق طراحی شده‌اند. آنها اجازه می‌دهند جریان آزاد اطلاعات در سطوح مختلف از طریق بزرگراه‌های داده، مشابه اتصالات سریع ResNet [۳۶]، جریان یابد.

مدل‌های مبتنی بر ترانسفورم‌ها

پردازش متوالی متن یکی از موانع محاسباتی است که RNN‌ها با آن مواجه هستند. اگرچه RNN‌ها نسبت به CNN‌ها متوالی‌تر هستند، اما هزینه محاسباتی ضبط ارتباطات بین کلمات در یک عبارت با افزایش طول جمله، بسیار شبیه به CNN‌ها، افزایش می‌یابد. مدل‌های مبتنی بر ترانسفورم‌ها [۷] با استفاده از خود توجه برای محاسبه یک "امتیاز توجه" برای هر کلمه در یک جمله از سند به طور همزمان، این محدودیت را برطرف می‌کنند و تأثیر هر کلمه بر کلمات دیگر را مدل‌سازی می‌کنند. برخلاف CNN‌ها و RNN‌ها، ترانسفورم‌ها امکان موازی‌سازی بسیار بیشتری را فراهم می‌کنند و امکان آموزش کارآمد مدل‌های عظیم را بر روی حجم عظیمی از داده‌ها بر روی GPU‌ها فراهم می‌کنند.

از سال ۲۰۱۸، چندین مدل LM پیش‌آموزشی بزرگ مقیاس مبتنی بر ترانسفورم‌ها ظهور کرده‌اند. مدل‌های مبتنی بر ترانسفورم‌ها از معماری‌های شبکه بسیار عمیق‌تر (مثلاً، ترانسفورم‌های ۴۸ لایه [۶۳]) استفاده می‌کنند. پیش‌آموزی این مدل‌ها نیز بر روی مقادیر بیشتری از متن برای ضبط زمینه نمایش‌های متن انجام می‌شود.

مدل‌های زبان از پیش آموزش دیده (PLM) محبوب بر اساس شکل نمایش، سبک مدل، وظایف پیش‌آموزی و وظایف مرتبط، مطابق با آخرین بررسی کیو و همکاران [۶۴] طبقه‌بندی می‌شوند. PLM‌های خودبازگشتی و PLM‌های خودرمزگذاری دو نوع از PLM‌ها هستند. یکی از PLM‌های خودبازگشتی اولیه است که یک مدل یک سو است که یک متن را کلمه به کلمه از چپ به راست یا



برعکس پیش‌بینی می‌کند، با هر پیش‌بینی کلمه بر اساس پیش‌بینی‌های قبلی. با اضافه کردن طبقه‌بندی‌کننده‌های خطی برای وظایف مرتبط و تنظیم برچسب‌های مرتبط با وظایف، OpenGPT را می‌توان برای وظایف پایین‌دستی مانند TC تنظیم کرد. یکی از PLM‌های خودرمزگذاری خط پایه است. در تضاد با OpenGPT که کلمات را بر اساس پیش‌بینی‌های گذشته پیش‌بینی می‌کند، BERT معمولاً با استفاده از کار مدل‌سازی زبان ماسک شده (MLM) آموزش داده می‌شود که به طور دلخواه برخی از قسمت‌ها را در یک دنباله متنی ماسک می‌کند و سپس آنها را به طور مستقل با استفاده از بردارهای رمزگذاری تولید شده توسط یک ترانسفورمر که متن را در هر دو جهت پردازش می‌کند، بهبود می‌بخشد. [۶۵] RoBERTa یک نسخه قوی‌تر از BERT است که با داده‌های بیشتر آموزش داده شده است. [۶۶] ALBERT استفاده از حافظه BERT را کاهش می‌دهد در حالی که سرعت آموزش آن را افزایش می‌دهد. DistillBERT [۶۷] از تکنیک تقطیر دانش در طول پیش‌آموزی برای کاهش اندازه BERT تقریباً به نصف در حالی که قابلیت‌های اصلی آن را حفظ می‌کند و استنتاج را با ضریب دو سرعت می‌دهد، استفاده می‌کند. [۶۸] SpanBERT یک افزونه BERT است که نمایش و پیش‌بینی بازه‌های متن را بهبود می‌بخشد. [۶۹] ERNIE پایگاه‌های دانش خارجی را برای بهبود عملکرد ادغام می‌کند. [۷۰] XLNet مفاهیم مدل‌های خودبازگشتی مانند OpenGPT و BERT را ترکیب می‌کند.

همانطور که قبلاً گفته شد، OpenGPT با استفاده از یک ترانسفورمر چپ به راست، نمایش متن را برای ایجاد زبان طبیعی یاد می‌گیرد، در حالی که BERT از یک ترانسفورمر استفاده می‌کند که می‌تواند متن را در هر دو جهت برای تفسیر زبان طبیعی پردازش کند. مدل زبان یکپارچه [۷۱] (UniLM) مدلی برای درک و ایجاد زبان طبیعی است. UniLM بر روی انواع مختلف وظایف مدل‌سازی زبان که به جهت تجزیه و تحلیل مربوط نمی‌شوند، از پیش آموزش دیده است. یک معماری ترانسفورمر مشترک با ماسک‌های خود توجه خاص برای ایجاد مدل‌سازی یکپارچه استفاده می‌شود. [۷۱] UniLM با استفاده از وظایف مختلف مدل‌سازی زبان مانند پیش‌بینی کلمه بعدی، پر کردن شکاف و تولید متن، آموزش داده می‌شود. [۷۱] UniLM به عملکرد SOTA جدیدی برای درک زبان طبیعی و ایجاد وظایف دست یافته است و از PLM‌های قبلی پیشی گرفته است. تحلیل عملکرد مدل‌های زبان معنادار، همراه با تکنیک‌های بکار رفته، در جدول ۸ خلاصه شده است.

جدول ۸. تحلیل عملکرد مدل‌های مهم زبان

مرجع	مجموعه داده	روش‌ها	کاربرد	عملکرد (دقت)
[۹]	AGNews	BERT	دسته بندی متن	۹۲.۷
[۲۳]	AGNews	FastText	دسته بندی متن	۹۲.۵
[۲۶]	SNLI	LSTM درختی	استنباط	۸۵.۷
[۲۷]	Amazon-۲, Yelp-۲	LSTM زنجیری	بررسی محصول، تحلیل عواطف	۹۵.۱۲, ۹۴.۳۹
[۳۵]	SST-۲	VDCNN	تحلیل عواطف	۸۴.۵
[۴۰]	AGNews و SougeNews	CasuleNetB	دسته بندی متن	۹۷.۲۵, ۹۲.۳۹
[۵۱]	۲۰News	GNN	دسته بندی متن	۸۸.۵



[۶۱]	SNLI	SAN	استنباط	۸۵.۶
[۶۵]	SNLI, SST-۲, SQuAD ۲.۰	RoBERTa	استنباط، تحلیل عواطف، سیستم پرسش پاسخ	۸۲.۳، ۹۲.۶، ۹۵.۶
[۶۶]	SST-۲	ALBERT	تحلیل عواطف	۹۵.۲
[۶۸]	SQuAD ۱.۱, SQuAD ۲.۰	SPANBERT	سیستم پرسش پاسخ	۷۳.۷، ۹۴.۶
[۷۰]	AGNews	XLNet	دسته بندی متن	۹۵.۵
[۷۱]	AGNews, IMDB	ULMFiT	بررسی فیلم ها	۹۵.۴، ۸۲.۱

نتایج

این بخش به بررسی چالش‌های موجود در مدل‌های زبان یادگیری عمیق مبتنی بر چندکاره (MTL) می‌پردازد و به دنبال پاسخ به این سوال است که چه رویکردهای الگوریتمی مختلفی برای ترکیب MTL جهت بهبود دقت چارچوب پیشنهادی برای NLU در دسترس هستند. چالش اصلی در تحقیقات NLU، نمایش زبان وظایف مختلف برای آموزش است. روش‌های سنتی مبتنی بر قوانین، به ویژگی‌های تولید شده توسط انسان متکی بودند و قادر به تشخیص کلمات غیرمنتظره نبوده و نیازمند زمان قابل توجهی برای مهندسی ویژگی‌ها بودند. از طرف دیگر، NLU مبتنی بر یادگیری عمیق، ویژگی‌ها را استخراج کرده و نمایش‌های زبان را به طور خودکار می‌آموزد. مدل‌های MTL که از پایه یادگیری تک‌وظیفه‌ای ایجاد شده‌اند و نمایش تعمیم‌یافته‌ای را در حالی که از بیش‌برازش یک کار فردی معین جلوگیری می‌کنند، یاد می‌گیرند، در سال‌های اخیر توجه زیادی را به خود جلب کرده‌اند. مدل [۷۲] MT-DNN که بسیاری از وظایف را به طور همزمان می‌آموزد، نمونه‌ای از این مدل‌ها است. این رویکرد از روشی برای تقطیر دانش استفاده می‌کند که یک مدل چندکاره کوچکتر را با استفاده از مدل‌های تک‌وظیفه بزرگتر آموزش می‌دهد. با این حال، مدل‌های MTL با مشکلاتی مانند تأثیرپذیری دقت مدل به دلیل آموزش مشترک با وزن‌های برابر و عدم تسلط کامل بر کار اصلی در طول MTL مواجه هستند. برای رفع این معایب، طرح SIWLM (مدل‌سازی زبان وزنی متوالی و فشرده) با الهام از یانگ و همکاران [۷۳] ارائه شده است. SIWLM شامل دو نوع یادگیری وزنی متوالی (SWL) و وزنی فشرده (IWL) است که در آن وظایف اصلی و فرعی دارای وزن اولیه هستند و تمام وظایف در طول آموزش به طور مستقل تغییر می‌کنند. MTDNN-SIWLM با اعمال وزن ایده‌آل برای هر کار در مجموعه داده‌های GLUE، عملکرد معادل یا بهتری را در تمام این مجموعه داده‌ها به دست می‌آورد.

یکی دیگر از چالش‌های مهم در یادگیری چندوظیفه‌ای، تداخل وظایف است. تداخل وظیفه به عنوان پارادوکسی بین ثبات و حساسیت مشاهده می‌شود، به طوری که اطلاعات ضروری برای یک کار ممکن است برای کار دیگر بیهوده باشد و منجر به اهداف متناقض در حین آموزش شبکه‌های چندوظیفه‌ای و در نتیجه عملکرد ضعیف شود. برای حل این مشکل، مطالعات [۷۶-۷۷] دو استراتژی توجه وظیفه جداگانه را پیشنهاد می‌کنند. اولین استراتژی شامل سیگنال‌های مدولاسیون وابسته به داده‌های خاص وظیفه است که فعالیت عصبی را تقویت یا کاهش می‌دهند. دومین استراتژی، آداپتور باقیمانده خاص وظیفه است که اطلاعات خاص وظیفه را با نمادهای ایجاد شده توسط یک ساختار وظیفه مشترک ترکیب می‌کند. این روش به ما امکان می‌دهد یک سیستم نماد مشترک را یاد بگیریم که به همه فعالیت‌ها خدمت می‌کند و در عین حال با یک پردازش مرتبط با وظیفه برای توسعه ساختارهای پیچیده‌تر مرتبط با وظیفه همکاری می‌کند. همچنین، [۷۶] روشی را برای یادگیری نمایش کم‌بعدی مورد استفاده در کاربردهای مختلف توصیف می‌کند که از یک منظم‌کننده جدید برای مدیریت تعداد ویژگی‌های یادگیری مشترک برای همه وظایف استفاده می‌کند. این رویکرد دارای تفسیر



مستقیمی است و بین مراحل نظارت شده و بدون نظارت جایگزین می شود. در حالی که یادگیری عمیق و یادگیری تقویتی عمیق (RL) نویدبخش توانایی رایانه ها در انجام وظایف پیچیده هستند، نیازهای داده ای رویکردهای موجود، دستیابی به طیف گسترده ای از قابلیت ها را چالش برانگیز می کند، به خصوص زمانی که هر کار به طور مستقل از ابتدا یاد گرفته شود. برای حل مشکلات یادگیری چندوظیفه ای، یک استراتژی طبیعی آموزش یک شبکه بر روی چندین کار به طور همزمان برای کشف ساختار مشترک در بین وظایف به روشی کارآمدتر و مؤثرتر از مقابله با وظایف به صورت جداگانه است. [۷۷] ممکن است مزایای فرض شده یادگیری چندوظیفه ای را بدون هزینه عملکرد نهایی به دست آورد اگر مسائل بهینه سازی برای یادگیری چندوظیفه ای به درستی برطرف شوند.

تمام وظایف NLP، به طور معمول QA، خلاصه سازی محتوا، NLI، وظایف کاندید را به عنوان یک معیار برای دکلون زبان طبیعی (decaNLP) در نظر می گیرند [۷۲]. این مطالعه رویکردی را برای بازسازی تمام مشاغل به عنوان مجموعه ای از سؤالاتی که باید در یک زمینه خاص پاسخ داده شوند، پیشنهاد می کند. یک شبکه جدید پاسخ به سؤالات چندوظیفه ای (MQAN) پیشنهاد شده است که وظایف decaNLP را بدون تنظیم دقیق هیچ شبکه ای یاد می گیرد [۷۲]. نمایش دانش به غیر از جفت QA باید برای MTL-NLU ارزیابی شود. رویکرد پیشنهادی در [۷۲] MT-DNN به دلیل ماهیت جامع آن به نظر عملی می رسد. اما دقت مدل به استراتژی های اتخاذ شده برای رسیدگی مؤثر به تداخل وظیفه حساس است و مقدار بالاتر تداخل وظیفه می تواند عملکرد مدل را به طور نامطلوبی تحت تأثیر قرار دهد. پرسش پژوهشی دوم به دنبال یافتن مجموعه داده های مقایسه ای استاندارد برای ارزیابی وظایف MTL-NLU است. [۷۸] GLUE و جانشین آن [۷۹] SuperGLUE معیارهای استاندارد برای ارزیابی عملکرد یک مدل بر روی مجموعه ای از وظایف به جای یک کار واحد هستند. GLUE شامل مجموعه داده های معیار و مجموعه داده تحلیلی برای ارزیابی و تجزیه و تحلیل دقت چارچوب در زبان طبیعی مربوط به طیف گسترده ای از پدیده های زبانی و یک هیئت رهبری عمومی برای ردیابی عملکرد است. در سال های اخیر، مدل ها و رویکردهای جدید پیش آموزی و یادگیری انتقال منجر به دستاوردهای قابل توجهی در زمینه های مختلف درک زبان شده اند. با این حال، عملکرد روی معیار GLUE به عملکرد انسان های غیر متخصص نزدیک شده است، که نشان می دهد پتانسیل محدودی برای توسعه بیشتر وجود دارد. در [۷۹]، سوپرگل (SuperGLUE)، یک معیار جدید الهام گرفته از گل (GLUE) که شامل مجموعه جدیدی از مشکلات پیچیده تر درک زبان، منابع بهبود یافته و یک جدول امتیازات عمومی جدید است، به عنوان یک استاندارد معیار بندی واقعی برای MTL-NLU ایجاد شده است.

جدول ۹. اطلاعات دقیق در مورد GLUE و منبع [۸۰، ۸۱] SuperGLUE.

مجموعه داده	بدنه	وظیفه	معیار	حوزه
GLUE	CoLA	تک جمله ای	ضریب ماتیوس	ترکیبی
	SST-۲	تحلیل عاطفه	دقت	بررسی ها
	MRPC	کارهای شباهت و پاداش دهی	امتیاز F_1 ، دقت	اقدام جدید
	STS-B	شباهت جمله	ضریب پیرسون یا اسپیرمن	ترکیبی
	QQP	پاداش دادن	امتیاز F_1 ، دقت	پرسش و پاسخ
	MNLI	استنباط	دقت	ترکیبی



ویکی پدیا	دقت	استنباط	QNLI	
ویکی پدیا، موارد جدید	دقت	استنباط	RTE	
رمان ها	امتیاز F ^۱	استنباط	WNLI	
جستجوهای گوگل، ویکی پدیا	دقت	پرسش و پاسخ	BoolQ	superGLUE
ترکیبی	دقت، F ^۱	NLI	CB	
بلاگ ها، دایره المعارف عکاسی	دقت	پرسش و پاسخ	COPA	
ترکیبی	F ^۱	پرسش و پاسخ	MultiRC	
اخبار (CNN, Daily Mail)	F ^۱ /EM	پرسش و پاسخ	ReCoRD	
اخبار، ویکی پدیا	دقت	NLI	RTE	
VerbNet, Wikipedia	دقت	WSD	WiC	
داستان ها	دقت	مرجع مشترک	WSC	

زمینه مجموعه داده‌های معیار برای درک زبان طبیعی (NLU) به سرعت در حال تکامل است و قطعاً امکان افزودن وظایف بیشتر مرتبط با NLU در آینده وجود دارد. یکی از پرسش‌های پژوهشی مهم، بررسی روش‌های یادگیری برای بهبود عملکرد یادگیری NLU در ترکیب وظایف چندگانه است. بازی‌های متنی که از زبان طبیعی برای بازسازی جهان‌ها و تعامل با بازیکنان استفاده می‌کنند، به عنوان آزمایشگاهی برای ربات‌های مستقل درک زبان مورد استفاده قرار گرفته‌اند [۸۲-۸۳]. با این حال، مشخص نیست که عوامل مصنوعی تا چه حد از درک معنایی متن استفاده می‌کنند. مطالعات انجام شده در زمینه ZORK-I، یک بازی ماجراجویی متنی، شواهد متناقضی را با این اصل اساسی نشان داده و بهبودی در ساختار واحد NLU پیشنهاد کرده‌اند. این مطالعات، تکنیک‌های تنظیم معنایی وارد شده به سیستم را نیز نشان می‌دهند. در آموزش مشترک [۸۲]، یک RNN عمیق با حافظه نسبی جزئی گفتگو با آموزش متقابل واحد NLU و واحد پیش‌بینی اقدام سیستم پیشنهاد شده است. مطالعه [۸۳]، خواننده توجه دروازه‌ای را به عنوان یک مدل یکپارچه با مکانیسم توجه جدید مبتنی بر تعامل چندگانه معرفی می‌کند. [۸۴] UMLFiT، تکنیکی برای تنظیم دقیق هر مدل زبانی است که به طور مؤثری از مدل‌های موجود در کارهای مختلف TC پیشی می‌گیرد. در مطالعه [۸۵]، از یک تکنیک آموزشی جدید برای آموزش CNN در آزمایش‌های درک خبر استفاده می‌شود. مطالعه [۸۶] یک نظریه جدید معنایی قاب برای زبان انگلیسی پیشنهاد می‌کند. در مطالعه [۸۷]، دستور زبان ساختاری (CG) به همراه هوش مصنوعی برای نمایش دانش برای درک عمیق متن استفاده می‌شود. مطالعه [۸۸] مدل‌های مبتنی بر ترانسفورمر جدیدترین مربوط به مدل‌های NLP را خلاصه می‌کند. مطالعه [۸۹]، روشی با نگرش پیشگیرانه برای هدایت گفتگو به سمت دستیابی به اهداف مختلف کوچینگ برای کاربران مسن با استفاده از تکنیک پشتیبانی ارتباطی مبتنی بر مدل GROW پیشنهاد می‌کند.

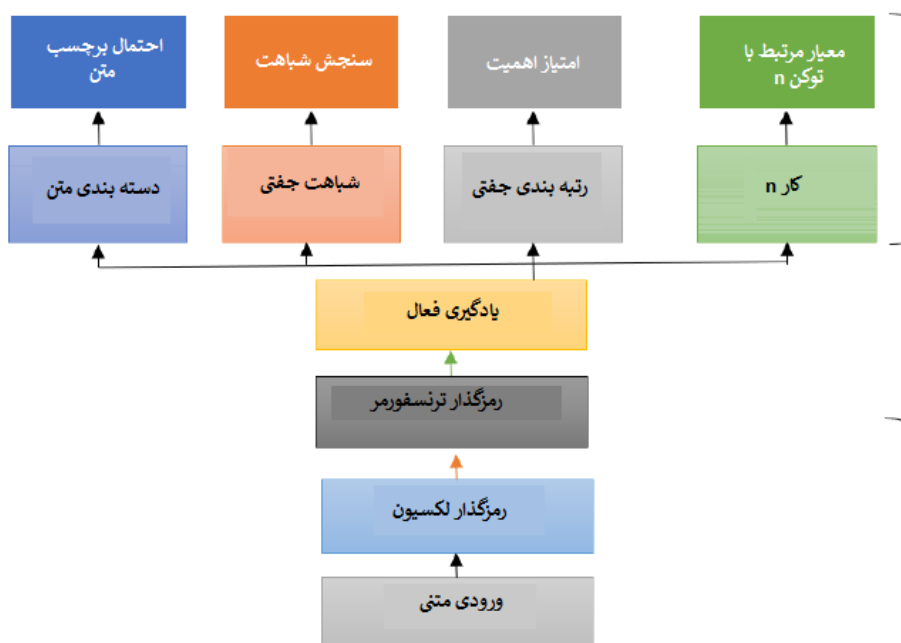
مطالعه [۹۰] تکنیک‌های توضیح کلمات را برای حوزه مراقبت پرستاری ارزیابی کرده و فست‌تکست را به عنوان یک مدل توضیح کلمات بهتر معرفی می‌کند. مطالعه [۹۱] عملکرد سرویس‌های NLU و نقش آن‌ها در معماری کلی یک چت‌بات را توصیف می‌کند. مطالعه [۹۲] بر زبان ایتالیایی به عنوان یک زبان غیر انگلیسی برای استفاده از موتور NLU تمرکز دارد. در مطالعه [۹۳]، یک چارچوب عملکردی پیشنهاد شده و اصول RASA NLU معرفی شده‌اند. مطالعه [۹۴] رویکردی را برای ترکیب انواع مختلف معنایی و مدل‌های زبانی در حین پیش‌آموزش و مرحله تنظیم دقیق برای بهبود دقت پیش‌بینی برجسته می‌کند. مطالعه [۸۲] به مشکل ورودی خارج از دامنه (OOD) می‌پردازد و روشی جدید برای تولید نمونه‌های شبه OOD با کیفیت بالا پیشنهاد می‌کند. در مطالعه [۷۵]، یادگیری بردار معنایی برای NLU مورد بررسی قرار می‌گیرد. مطالعه [۹۵] تلاش می‌کند تا مشکل آموزش خاص دامنه را حل کند. مطالعه [۹۶] طراحی رابط کاربری مؤثری برای سیستم تحلیل سهام مبتنی بر NLU را نشان می‌دهد. مطالعه [۹۷] تجزیه و تحلیل نظری NLU را به عنوان یک مشکل روش‌شناختی تأیید می‌کند. مطالعه [۹۸] بر نامناسب بودن رویکردهای NLP برای سیستم‌های NLU تأکید می‌کند. در مطالعه [۹۹]، مشکل اضافه کردن وظایف جدید با داده‌های آموزشی موجود مورد بررسی قرار می‌گیرد. مطالعه [۱۰۰] به مشکل جمله‌های مسیر باغ (GP) می‌پردازد. مطالعه [۱۰۱] یک معماری جدید برای ادغام فعال دانش پس‌زمینه صریح در مدل NLU معرفی می‌کند. مطالعه [۱۰۲] روش‌های مختلفی را برای تغییر داده‌های مجموعه داده استاندارد طلایی توصیف و ارزیابی می‌کند. روش توجه دروازه‌ای به دلیل کارایی آن در ادبیات فعلی رایج است. پرسش پژوهشی دیگر بررسی می‌کند که کدام تکنیک‌ها برای کاهش نیاز به نمونه‌های داده‌های دارای حاشیه‌نویسی عظیم مؤثر هستند. [۱۰۵] VIRAAL رویکرد جدیدی برای کاهش تلاش‌های حاشیه‌نویسی در سیستم‌های NLU پیشنهاد می‌کند. مطالعه [۱۰۶] تلاش می‌کند مشکلات دامنه‌های کوچک را که به مقدار زیادی داده‌های آموزشی مرتبط با دامنه نیاز دارند، برطرف کند. [۱۰۷] HUMOD، یک مجموعه داده گفتگوهای فیلم با کیفیت بالا با حاشیه‌نویسی انسانی است. مطالعه [۱۰۸] یک تکنیک یادگیری دوگانه برای NLU نیمه‌نظارتی پیشنهاد می‌کند. در مطالعه [۱۰۹]، رویکرد جدیدی برای استفاده از وظایف کمکی برای ارائه نظارت اضافی بر کار اصلی برای جبران کمبود داده پیشنهاد شده است. روش [۷۲] VIRAAL در کاهش نیاز به نمونه‌های حاشیه‌نویسی عظیم برای آموزش مؤثرتر است. پرسش پژوهشی آخر به دنبال یافتن تکنیک‌های نمایش دانش رایج برای MTL- NLU است. ژانگ و والاس [۳۹] نشان دادند که word²vec و GloVe بهتر از بردارهای یک-داغ هستند. مدل [۶] Elmo اطلاعات زمینه‌ای را در نمایش دانش در نظر می‌گیرد. [۱۱۰] EntityNLM می‌تواند موجودیت‌های پویا را مدل‌سازی کند. مطالعه [۱۱۱] تکنیکی برای فشرده‌سازی مدل‌های زبانی حجیم پیشنهاد می‌کند. تکنیک هرس مدل [۱۱۱] روشی کارآمدتر و پرکاربردتر در فشرده‌سازی نمایش دانش است.

بحث و بررسی

این مطالعه با بررسی ۱۱۶ مقاله علمی، به کندوکاو در جنبه‌های مختلف ساخت مدل‌های زبانی با استفاده از رویکردهای یادگیری عمیق می‌پردازد. در این بررسی، چندین سوال کلیدی در مورد مدل‌های زبانی مورد بحث قرار گرفته است. شبکه‌های عصبی پیش‌خور (FFNN) متن ورودی را به عنوان کیسه‌ای از کلمات در نظر می‌گیرند، در حالی که RNN می‌تواند ترتیب کلمات را درک کنند و CNN‌ها در تشخیص الگوهایی مانند کلمات معنی‌دار قوی هستند. مکانیسم‌های توجه در شناسایی کلمات مرتبط در متن ورودی بسیار عالی عمل می‌کنند. شبکه‌های عصبی Siamese برای وظایف شباهت متن و GNN‌ها در صورت مفید بودن ساختارهای زبان طبیعی برای کار مورد نظر، عملکرد بهتری دارند. با این حال، مدل‌های مبتنی بر ترانسفورمر از همه مدل‌های دیگر عملکرد بهتری دارند. معماری چارچوب توسط کار مورد نظر، وجود نمونه‌های دارای حاشیه‌نویسی و محدودیت‌های دامنه کاربرد تعیین می‌شود و این الزامات فقط می‌توانند توسط مدل‌های زبانی معماری ترکیبی برآورده شوند. مطالعات موجود، مدل‌های زبانی مبتنی بر MTL با الگوی یادگیری نظارت شده را بررسی می‌کنند، اما مطالعات کمی یادگیری بدون نظارت را با MTL ترکیب می‌کنند. ترکیب نمایش‌های مبتنی بر ترانسفورمر با MTL می‌تواند کارایی و دقت مدل چارچوب زبان را بهبود بخشد. با استفاده از تکنیک یادگیری فعال در MTL می‌توان به طور کارآمد به مسئله تشخیص وظیفه خارج از دامنه و تداخل وظیفه رسیدگی کرد. به دلیل کمبود نمونه‌های دارای حاشیه‌نویسی در



وظایف NLU، توسعه رویکردهایی برای کاهش نیاز به این نمونه‌ها با استفاده از تکنیک‌های یادگیری فعال حائز اهمیت است. مدل‌های زبان از پیش آموزش دیده (PLM) عملکرد را برای تمام وظایف مرتبط با متن ارتقا داده‌اند و PLM‌های خودرمزگذار به دلیل کارایی، رایج‌ترین مدل‌های زبانی هستند. لایه‌های مرتبط با کار خاص می‌توانند به‌طور مستقل یا به‌طور مشترک با PLM آموزش داده شوند که این امر به وجود برچسب‌های مرتبط با دامنه بستگی دارد. تنظیم دقیق چندوظیفه‌ای جایگزینی عالی برای استفاده از داده‌های دارای برچسب از دامنه‌های مرتبط است. اندازه مدل‌های مبتنی بر ترانسفورمر موجود بسیار زیاد است و رویکردهای متعددی برای فشرده‌سازی نمایش مبتنی بر ترانسفورمر در دسترس است، اما استراتژی بهینه برای رمزگذاری متن برای NLU در ادبیات فعلی مشخص نیست. ادغام دانش بدیهی در مدل‌های DL می‌تواند به‌طور چشمگیری احتمال افزایش کارایی مدل را افزایش دهد، همانطور که انسان‌ها از درک بدیهی برای انجام کارهای مختلف استفاده می‌کنند. به عنوان مثال، یک سیستم سؤال و پاسخ عمومی می‌تواند به سؤالات مربوط به دنیای واقعی پاسخ دهد و در صورت عدم وجود اطلاعات کافی، دانش بدیهی می‌تواند کمک کند. سیستم‌های مبتنی بر یادگیری ماشین می‌توانند با استفاده از دانش بدیهی و قراردادهای «پیش‌فرض»، در مورد ناشناخته‌ها استدلال کنند. با وجود پتانسیل بالای مدل‌های DL در معیارهای چالش‌برانگیز، عملکرد اکثر این مدل‌ها قابل تفسیر نیست و مشخص نیست که چرا برخی از مدل‌ها در یک مجموعه داده عملکرد بهتری دارند در حالی که در مجموعه داده‌های دیگر عملکرد ضعیفی دارند. مکانیسم‌های توجه تا حدی به این ابهامات پاسخ می‌دهند، اما تجزیه و تحلیل جامع رفتار و پویایی‌های اساسی این مدل‌ها هنوز وجود ندارد. دانش بیشتر در مورد عناصر نظری این مدل‌ها می‌تواند به توسعه مدل‌های متناسب با تنظیمات مختلف تجزیه و تحلیل متن کمک کند.



شکل ۵. چارچوب پیشنهادی برای NLU چندوظیفه‌ای.

چارچوب پیشنهادی

این مطالعه SLR را برای پاسخ به پنج سوال تحقیقاتی آغاز کرد تا زمینه را برای مطالعات آینده در حوزه NLU فراهم کند. این پرسش‌های پژوهش به عنوان سنگ بنای چارچوب پیشنهادی NLU چندوظیفه‌ای عمل کردند. شکل ۵ چارچوب پیشنهادی را نشان می‌دهد.

همانطور که در شکل ۵ نشان داده شده است، چارچوب پیشنهادی از BERT مشتق شده است و دارای سه لایه پایین است که برای همه وظایف مشترک هستند و لایه‌های انتهایی نمایش و خروجی مرتبط با وظیفه را نشان می‌دهند. ورودی متن است که به عنوان توضیح کلمات تبدیل شده است. سپس رمزگذار ترانسفورمر از خودتوجهی برای دریافت اطلاعات زمینه ای برای هر کلمه استفاده می‌کند و در مرحله بعدی رشته ای از توضیح های زمینه ای ایجاد می‌کند. نمایش معنایی مبادله شده برای دستیابی به اهداف مورد نیاز در وظایف چندگانه استفاده می‌شود. تکنیک یادگیری فعال بیشتر برای به دست آوردن اطلاعات بیشتر از نمونه‌های کمتر اما آموزنده‌تر استفاده می‌شود و در نتیجه تقاضا برای نمونه‌های بیشتر کاهش می‌یابد.

آموزش چارچوب پیشنهادی شامل مراحل مانند پیش‌آموزش و یادگیری چندوظیفه‌ای است. مرحله پیش‌آموزش مشابه مدل BERT است. مدل‌سازی زبان ماسک شده و پیش‌بینی جمله بعدی برای یادگیری پارامترهای هر دو رمزگذار استفاده می‌شوند. پارامترهای مدل در مرحله نهایی با استفاده از روش نزول گرادیان تصادفی (SGD) یاد گرفته می‌شوند.

انتظار می‌رود این چارچوب با فرض ترکیب تکنیک‌های متعدد، عملکرد خوبی داشته باشد.

محدودیت‌های مطالعه

در این SLR، مدل‌های زبان DL فعلی مورد استفاده در وظایف طبقه‌بندی متن مورد بررسی و تجزیه و تحلیل انتقادی قرار گرفتند. این بررسی مقایسه‌ها و چالش‌هایی را برای توسعه مدل‌های DL برای NLU چندوظیفه‌ای ارائه می‌دهد. با این حال، به دلیل کمبود تحقیقات ادبیات و کار در این زمینه، و همچنین طیف گسترده‌ای از مدل‌های DL، یافتن و انتخاب ادبیات مرتبط کاری پر زحمت، سخت و دشوار است. برای برآورده کردن معیارهای شمول و حذف مورد نیاز، کلمات کلیدی مورد استفاده برای جستجوی انتشارات و روش‌های ارزشمند ممکن است متفاوت یا تغییر کنند.

یکی از محدودیت‌های اصلی SLR برای این حوزه این است که حتی اگر یک فرآیند بررسی سیستماتیک دنبال شود، نمی‌توان تضمین کرد که تمام آثار مرتبط این حوزه استخراج شده‌اند. معتبرترین پایگاه‌های داده الکترونیکی در علوم کامپیوتر در پایگاه‌های داده جستجو گنجانده شدند.

محدودیت دیگر، تصورات از پیش تعیین‌شده نویسندگان در مورد روش NLU چندوظیفه‌ای به عنوان یک کل است. چارچوب پیشنهادی در مراحل طراحی است و اعتبارسنجی تجربی آن خارج از محدوده این SLR است، اما اهداف تحقیقاتی بلندمدت را نشان می‌دهد.

کارهای آینده و چالش‌ها

چندین نگرانی باید در تحقیقات آینده مورد توجه قرار گیرد.

- وظایف پرت: برای شروع، شناخته شده است که وظایف پرت که با سایر وظایف ارتباطی ندارند، هنگام یادگیری مشترک، عملکرد تمام فعالیت‌ها را مختل می‌کنند. چندین استراتژی برای کاهش عواقب مضر وظایف پرت وجود دارد. با این حال، هیچ رویکرد یا ارزیابی نظری مشخصی برای بررسی عواقب زیانبار وجود ندارد. این مسئله یک موضوع مهم است که نیاز به تحقیقات بیشتری برای ایمن‌سازی MTL برای استفاده انسان دارد.
- روش‌های یادگیری: یادگیری عمیق به عنوان یک استراتژی برجسته در زمینه‌های مختلف ظهور کرده است، با ارائه مدل‌های مختلف یادگیری عمیق MTL در روش‌های تغییر ویژگی، رتبه پایین، دسته‌بندی وظایف و یادگیری مرتبط با وظیفه. همانطور که قبلاً گفته شد، اکثر آنها فقط دارای لایه‌های مشترک پنهان هستند. این روش زمانی مؤثر است که تمام وظایف مرتبط باشند، اما مستعد وظایف پرت است و عملکرد را به میزان قابل توجهی کاهش می‌دهد.
- امنیت: مقاومت این چارچوب DNN چندوظیفه‌ای در برابر انواع مختلف حملات باید بررسی شود.



- توسعه: در نهایت، اکثر مطالعات تا به امروز بر روی وظایف یادگیری نظارت شده متمرکز شده‌اند و تنها تعداد کمی بر روی انواع یادگیری مانند بدون نظارت، نیمه‌نظارتی، فعال و تقویت برای آن مشکلات یادگیری بدون نظارت تمرکز کرده‌اند. منطقی است که روش‌های مختلف چندوظیفه‌ای را اصلاح یا گسترش دهیم. اعتقاد بر این است که چنین سازگاری و گسترش‌هایی نیاز به تلاش بیشتری در توسعه مدل‌های مرتبط خواهند داشت.

نتیجه‌گیری

این مطالعه به بررسی دقیق مدل‌های زبانی مبتنی بر یادگیری عمیق (DL) برای درک زبان طبیعی (NLU) چندوظیفه‌ای می‌پردازد. هدف از این بررسی سیستماتیک ادبیات (SLR)، شناسایی زمینه‌هایی برای پیشرفت در مدل‌های زبانی معاصر است. با وجود اینکه به طور کلی پذیرفته شده است که ارتباط وظایف در یادگیری چندوظیفه‌ای (MTL) منجر به جهت‌گیری صحیح می‌شود، کاروانا [۱۱۲] نشان داد که برخی از جهت‌گیری‌های القایی می‌توانند مضر باشند و هیچ مفهوم جهانی قوی برای اندازه‌گیری دشواری تداخل وظایف وجود ندارد. MTL با اهداف پیچیده و رقابتی متعددی روبرو است و همچنان به استراتژی‌های عمیق‌تر و کلی‌تر برای انتخاب و ارزیابی وظایف نیاز دارد. برای درک بهتر ارتباط و انتخاب وظیفه، انجام مطالعات بیشتر در مورد پیامدهای MTL ضروری است. این SLR به بررسی و تجزیه و تحلیل مدل‌های زبانی مبتنی بر DL برای تحقیقات NLU چندوظیفه‌ای می‌پردازد و مشکلات مدل‌های زبانی موجود برای NLU چندوظیفه‌ای را شناسایی می‌کند. این مطالعه همچنین ضرورت ترکیب الگوهای یادگیری نظارت شده و بدون نظارت را تشخیص می‌دهد و نیاز به ترکیب نمایش مبتنی بر ترانسفورمر با نمایش توضیح کلمات برای متن را بررسی می‌کند. علاوه بر این، استراتژی‌های فشرده‌سازی مدل که می‌توانند برای کاهش اندازه مجموعه داده استفاده شوند، مورد شناسایی قرار می‌گیرند. این حوزه برای کاربردهای مختلف، از جمله هوش مصنوعی مکالمه‌ای، سیستم‌های مبتنی بر چت‌بات برای آموزش، حقوق، بازار سهام و خدمات مشتری، مورد بررسی قرار می‌گیرد. یافته‌ها نشان می‌دهند که مدل ترکیبی، که استراتژی‌های متنوعی مانند MTL و یادگیری فعال را ادغام می‌کند، به دلیل اثربخشی در مدیریت وظایف مرتبط با متن برای NLU، مورد توجه بیشتری قرار می‌گیرد. همچنین مشخص شده است که ادغام الگوهای نظارت شده و بدون نظارت کمتر مورد توجه قرار گرفته است. بنابراین، ساخت یک مدل ترکیبی NLU چندوظیفه‌ای چشم‌اندازی امیدوارکننده است و همچنان تحقیقات برای ساخت مدل‌های MTL مقاوم و توانمند برای نسل بعدی هوش مصنوعی عمومی ادامه دارد.

منابع

- Kommandotech. (۲۰۲۱). Astounding Artificial Intelligence Statistics for ۲۰۲۰. Accessed: Oct. ۲۲, ۲۰۲۱. [Online]. Available: <https://kommandotech.com/statistics/artificial-intelligence-statistics/>
- Botcore. (۲۰۱۸). Chatbots: The Past, Present, and Future. Accessed: Oct. ۲۲, ۲۰۲۱. [Online]. Available: <https://botcore.ai/blog/chatbots-the-past-present-and-future>
- S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," J. Amer. Soc. Inf. Sci., vol. ۴۱, no. ۶, pp. ۳۹۱-۴۰۷, ۱۹۹۰.
- Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," J. Mach. Learn. Res., vol. ۳, pp. ۱۱۳۷-۱۱۵۵, Nov. ۱۹۸۵.
- T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Proc. Adv. Neural Inf. Process. Syst., ۲۰۱۳, pp. ۳۱۱۱-۳۱۱۹.
- M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," ۲۰۱۸, arXiv:۱۸۰۲.۰۵۳۶۵.
- A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., ۲۰۱۷, pp. ۵۹۹۸-۶۰۰۸.
- A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. Improving Language Understanding by Generative Pre-Training. Accessed: Oct. ۲۲, ۲۰۲۱. [Online]. Available: <https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/languageunderstandingpaper.pdf>
- J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," ۲۰۱۸, arXiv:۱۸۱۰.۰۴۸۰۵.



- T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, and P. Dhariwal, "Language models are few-shot learners," 2020, arXiv:2005.14165.
- D. Lepikhin, H. Lee, Y. Xu, D. Chen, O. Firat, Y. Huang, M. Krikun, N. Shazeer, and Z. Chen, "GShard: Scaling giant models with conditional computation and automatic sharding," 2020, arXiv:2006.16668.
- Developer. (2021). Announcing Megatron or Training Trillion Parameter Models Rivavailability. Accessed: Oct. 22, 2021. [Online]. Available: <https://developer.nvidia.com/blog/announcing-megatron-for-training-trillion-parameter-models-riva-availability/>
- G. Marcus, "The next decade in AI: Four steps towards robust artificial intelligence," 2020, arXiv:2002.06177.
- Y. Nie, A. Williams, E. Dinan, M. Bansal, J. Weston, and D. Kiela, "Adversarial NLI: A new benchmark for natural language understanding," 2019, arXiv:1910.14599.
- E. S. Dos Reis, C. A. Da Costa, D. E. Da Silveira, and R. S. Bavaresco, "Transformer aftermath review," Commun. ACM, vol. 64, no. 4, pp. 154-163, Apr. 2021.
- Y. Zhang and Q. Yang, An Overview of Multi-Task Learning, vol. 5. London, U.K.: Oxford Univ. Press, 2018, pp. 30-43.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, and A. N. Gomez, "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., Dec. 2017, pp. 5999-6009.
- B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in SE," Guidel. Perform. Syst. Lit. Rev., vol. 3, pp. 1-44, Oct. 2007.
- P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," 2016, arXiv:1606.05250.
- M. Marelli, L. Bentivogli, M. Baroni, R. Bernardi, S. Menini, and R. Zamparelli, "SemEval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), 2014, pp. 1-8.
- T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, arXiv:1301.3781.
- J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), vol. 2014, pp. 1532-1543.
- A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "FastText.Zip: Compressing text classification models," 2016, arXiv:1612.03251.
- S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2012, pp. 90-94.
- Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in Proc. Int. Conf. Mach. Learn., 2014, pp. 1188-1196.
- K. Sheng Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," 2015, arXiv:1503.00075.
- X. Zhu, P. Sobihani, and H. Guo, "Long short-term memory over recursive structures," in Proc. Int. Conf. Mach. Learn., 2015, pp. 1604-1612.
- P. Liu, X. Qiu, X. Chen, S. Wu, and X.-J. Huang, "Multi-timescale long short-term memory neural network for modelling sentences and documents," in Proc. Conf. Empirical Methods Natural Lang. Process., 2015, pp. 2326-2335.
- P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," 2016, arXiv:1605.05011.
- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
- N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, 2014, pp. 1-11.
- Y. Kim, "Convolutional neural networks for sentence classification," in Proc. Conf. Empirical Methods Natural Lang. Process., 2014, pp. 1-12.
- J. Liu, W.-C. Chang, Y. Wu, and Y. Yang, "Deep learning for extreme multi-label text classification," in Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2017, pp. 115-124.
- J. D. Prusa and T. M. Khoshgoftaar, "Designing a better data representation for deep neural networks and text classification," in Proc. IEEE 17th Int. Conf. Inf. Reuse Integr. (IRI), Jul. 2016, pp. 411-416.
- A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," 2016, arXiv:1606.01781.
- K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770-778.



- A. B. Duque, L. L. J. Santos, D. Macédo, and C. Zanchettin, "Squeezed very deep convolutional neural networks for text classification," in *Artificial Neural Networks and Machine Learning (Lecture Notes in Computer Science)*. Munich, Germany: European Neural Network Society, ۲۰۱۹.
- B. Guo, C. Zhang, J. Liu, and X. Ma, "Improving text classification with weighted word embeddings via a multi-channel TextCNN model," *Neurocomputing*, vol. ۳۶۳, pp. ۳۶۶-۳۷۴, Oct. ۲۰۱۹.
- Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," ۲۰۱۵, arXiv:۱۵۱۰.۰۳۸۲۰.
- G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming autoencoders," in *Proc. Int. Conf. Artif. Neural Netw.* Espoo, Finland: Springer, ۲۰۱۱, pp. ۴۴-۵۱.
- S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, ۲۰۱۷, pp. ۳۸۵۶-۳۸۶۶.
- S. Sabour, N. Frosst, and G. Hinton, "Matrix capsules with em routing," in *Proc. ۳th Int. Conf. Learn. Represent.*, ۲۰۱۸, pp. ۱-۱۵.
- J. Kim, S. Jang, and S. Choi, "Text classification using capsules," *Neurocomputing*, vol. ۳۷۶, no. ۱, pp. ۲۱۴-۲۲۱, Feb. ۲۰۱۸.
- H. Ren and H. Lu, "Compositional coding capsule network with K-means routing for text classification," ۲۰۱۸, arXiv:۱۸۱۰.۰۹۱۷۷.
- D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," ۲۰۱۴, arXiv:۱۴۰۹.۰۴۷۳.
- Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, "Hierarchical attention networks for document classification," in *Proc. NAACL*, ۲۰۱۶, pp. ۱۴۸۰-۱۴۸۹.
- X. Zhou, X. Wan, and J. Xiao, "Attention-based LSTM network for cross-lingual sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, ۲۰۱۶, pp. ۲۴۷-۲۵۶.
- T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "DiSAN: Directional self-attention network for RNN/CNN-free language understanding," in *Proc. ۳rd AAAI Conf. Artif. Intell.*, ۲۰۱۸, pp. ۵۴۴۶-۵۴۵۵.
- Y. Liu, C. Sun, L. Lin, and X. Wang, "Learning natural language inference using bidirectional LSTM model and inner-attention," ۲۰۱۶, arXiv:۱۶۰۵.۰۹۰۹۰.
- R. Mihalcea and P. Tarau, "TextRank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, ۲۰۰۴, pp. ۴۰۴-۴۱۱.
- Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," ۲۰۱۹, arXiv:۱۹۰۱.۰۵۵۹۶.
- T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," ۲۰۱۶, arXiv:۱۶۰۹.۰۲۹۰۷.
- G. Cucurull, A. Casanova, A. Romero, P. Liü, and Y. Bengio, "Graph attention networks," ۲۰۱۷, arXiv:۱۷۱۰.۱۰۹۰۳.
- L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. ۳۳, ۲۰۱۹, pp. ۷۳۷۰-۷۳۷۷.
- C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," ۲۰۱۵, arXiv:۱۵۱۱.۰۸۶۳۰.
- R. Zhang, H. Lee, and D. R. Radev, "Dependency sensitive convolutional neural networks for modeling sentences and documents," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, ۲۰۱۶, pp. ۱-۱۱.
- G. Chen, D. Ye, E. Cambria, J. Chen, and Z. Xing, "Ensemble application of convolutional and recurrent neural networks for multi-label text categorization," in *Proc. IJCNN*, ۲۰۱۷, pp. ۲۳۷۷-۲۳۸۳.
- D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, ۲۰۱۵, pp. ۱۴۲۲-۱۴۳۲.
- Y. Xiao and K. Cho, "Efficient character-level document classification by combining convolution and recurrent layers," ۲۰۱۶, arXiv:۱۶۰۲.۰۰۳۶۷.
- R. Wang, Z. Li, J. Cao, T. Chen, and L. Wang, "Convolutional recurrent neural networks for text classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. ۲۰۱۹, pp. ۱-۶.
- X. Liu, Y. Shen, K. Duh, and J. Gao, "Stochastic answer networks for machine reading comprehension," ۲۰۱۷, arXiv:۱۷۱۲.۰۳۵۵۶.
- R. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, ۲۰۱۵, pp. ۱-۹.
- A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multi-task learners," *OpenAI Blog*, vol. ۱, no. ۸, p. ۹, ۲۰۱۹.
- X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," ۲۰۲۰, arXiv:۲۰۰۳.۰۸۲۷۱.



- Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, arXiv:1907.11692.
- Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," 2019, arXiv:1909.11942.
- V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, arXiv:1910.01108.
- M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, and O. Levy, "SpanBERT: Improving pre-training by representing and predicting spans," 2019, arXiv:1907.10529.
- Y. Sun, S. Wang, Y. Li, S. Feng, X. Chen, H. Zhang, X. Tian, D. Zhu, H. Tian, and H. Wu, "ERNIE: Enhanced representation through knowledge integration," 2019, arXiv:1904.09223.
- Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pre-training for language understanding," in Proc. Adv. Neural Inf. Process. Syst., 2019, pp. 5754-5764.
- L. Dong, N. Yang, W. Wang, F. Wei, X. Liu, Y. Wang, J. Gao, M. Zhou, and H.-W. Hon, "Unified language model pre-training for natural language understanding and generation," in Proc. Adv. Neural Inf. Process. Syst., 2019, pp. 13042-13054.
- B. McCann, N. S. Keskar, C. Xiong, and R. Socher, "The natural language decathlon: Multitask learning as question answering," 2018, arXiv:1806.08730.
- S. Son, S. Hwang, S. Bae, S. J. Park, and J.-H. Choi, "A sequential and intensive weighted language modeling scheme for multi-task learning based natural language understanding," Appl. Sci., vol. 11, no. 9, p. 3905, Mar. 2021.
- M. McShane, "Natural language understanding (NLU, not NLP) in cognitive systems," AI Mag., vol. 38, no. 4, pp. 43-56, 2017.
- S. Jung, "Semantic vector learning for natural language understanding," Comput. Speech Lang., vol. 56, pp. 130-145, Jul. 2019.
- T. Yu, "Gradient surgery for multi-task learning," in Proc. 34th Conf. Neural Inf. Process. Syst., Vancouver, BC, Canada, 2020, pp. 5824-5836.
- A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," 2018, arXiv:1804.07461.
- A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "GLUE: A multi-task benchmark and analysis platform for natural language understanding," 2018, arXiv:1804.07461.
- A. Wang, "SuperGLUE: A stickier benchmark for general-purpose language understanding systems," in Proc. 33rd Conf. Neural Inf. Process. Syst., Vancouver, BC, Canada, 2019, pp. 1-15.
- K. Narasimhan and M. Hausknecht, "Reading and acting while blindfolded: The need for semantics in text game agents," 2021, arXiv:2103.13002.
- A. Zweig and D. Weinshall, "Hierarchical regularization cascade for joint learning," in Proc. 30th Int. Conf. Mach. Learn., vol. 2, 2013, pp. 1074-1082.
- G. Chen and M. Huang, "Out-of-domain detection for natural language understanding in dialog systems," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 28, pp. 1198-1209, 2020.
- B. Dhingra, "Gated-attention readers for text comprehension," in Proc. 50th Annu. Meeting ACL, Vancouver, BC, Canada, 2017, pp. 1832-1846.
- J. Howard, "Universal language model fine-tuning for text classification," in Proc. 56th Annu. Meeting ACL, Melbourne, VIC, Australia, 2018, pp. 328-339.
- D. Chen, "A thorough examination of the CNN/daily mail reading comprehension task," in Proc. 54th Annu. Meeting ACL, Berlin, Germany, 2018, pp. 2358-2370.
- W. Che, Y. Liu, Y. Wang, B. Zheng, and T. Liu, "Towards better UD parsing," in Proc. CoNLL, 2018, pp. 55-64.
- D. Kiselev, "An AI using construction grammar to understand text: Parsing improvements," Int. J. Cogn. Inform. Natural Intell., vol. 15, no. 2, pp. 47-61, Apr. 2021.
- S. Singh and A. Mahmood, "The NLP cookbook: Modern recipes for transformer based deep learning architectures," IEEE Access, vol. 9, pp. 78675-78702, 2021.
- C. Montenegro, A. L. Zorrilla, J. M. Olaso, R. Santana, R. Justo, J. A. Lozano, and M. I. Torres, "A dialogue-act taxonomy for a virtual coach designed to improve the life of elderly," Multimodal Technol. Interact., vol. 3, no. 3, p. 52, Jul. 2019.
- T. Mairitha, N. Mairitha, and S. Inoue, "Automatic labelled dialogue generation for nursing record systems," J. Pers. Med., vol. 10, no. 3, pp. 1-24, Sep. 2020.
- D. Braun, "Evaluating natural language understanding services for conversational question answering systems," in Proc. SIGDIAL Conf., Saarbrücken, Germany, 2017, pp. 174-185.
- M. Zubani, L. Sigalini, I. Serina, and A. E. Gerevini, "Evaluating different natural language understanding services in a real business case for the Italian language," Proc. Comput. Sci., vol. 176, pp. 995-1004, Oct. 2020.

- A. Jiao, "An intelligent chatbot system based on entity extraction using RASA NLU and neural network," J. Phys. Conf. Ser., vol. 1487, Mar. 2020, Art. no. 012014, doi: 10.1088/1742-6596/1487/1/012014.
- T. Mayer, "Enriching language models with semantics," in Proc. 24th Eur. Conf. Artif. Intell., Santiago, Spain, Aug. 2020, pp. 1-3.
- G. B. Jensen and B. McGillivray, "Enhancing domain-specific supervised natural language intent classification with a top-down selective ensemble model," Mach. Learn. Knowl. Extraction, vol. 1, no. 2, pp. 630-640, Apr. 2019.
- P. Lauren, "A conversational user interface for stock analysis," in Proc. Int. Conf. Big Data, 2019, pp. 5298-5305.
- V. Shymko, "Natural language understanding: Methodological conceptualization," Psycholinguistics, vol. 25, no. 1, pp. 431-443, Apr. 2019.
- E. Tomal and K. D. Forbus, "EA NLU: Practical language understanding for cognitive modelling," in Proc. 22nd Int. Florida Artif. Intell. Res. Soc. Conf., vol. 22, 2009, pp. 117-122.
- Z. Li and D. Hoiem, "Learning without forgetting," 2016, arXiv:1606.09282.
- J.-L. Du, "Predicting garden path sentences based on natural language understanding system," Int. J. Adv. Comput. Sci. Appl., vol. 3, no. 11, 2012, doi: 10.14569/IJACSA.2012.031101.
- D. Weissenborn, T. Kořinský, and C. Dyer, "Dynamic integration of background knowledge in neural NLU systems," 2017, arXiv:1706.02096.
- S. Peng and A. Zeldes, "All roads lead to UD: Converting stanford and penn parses to english universal dependencies with multi-layer annotations," in Proc. Joint Workshop Linguistic Annotation, Multiword Expressions, Construct. Workshop, 2018, pp. 167-177.
- G. Senay, B. Y. Idrissi, and M. Haziza, "VirAAL: Virtual adversarial active learning for NLU," Tech. Rep., May 2020. [Online]. Available: <https://arxiv.org/abs/2005.07287v4>
- B. Wu, B. Wei, J. Liu, K. Wu, and M. Wang, "Faceted text segmentation via multitask learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 9, pp. 3846-3857, Sep. 2021, doi: 10.1109/TNNLS.2020.3010996.
- M. Dimovski, C. Musat, V. Ilievski, A. Hossmann, and M. Baeriswyl, "Submodularity-inspired data selection for goal-oriented chatbot training based on sentence embeddings," 2018, arXiv:1802.00707.
- A. Tiwari, T. Saha, S. Saha, S. Sengupta, A. Maitra, R. Ramnani, and P. Bhattacharyya, "A dynamic goal adapted task-oriented dialogue agent," PLoS ONE, vol. 16, no. 4, Apr. 2021, Art. no. e0249030.
- E. Merdivan, D. Singh, S. Hanke, J. Kropf, A. Holzinger, and M. Geist, "Human annotated dialogues dataset for natural conversational agents," Appl. Sci., vol. 10, no. 3, p. 762, Jan. 2020.
- S. Zhu, R. Cao, and K. Yu, "Dual learning for semi-supervised natural language understanding," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 28, pp. 1936-1947, 2020.
- B. Shi, "Auxiliary task reweighting for minimum-data learning," in Proc. 34th Conf. Neural Inf. Process. Syst., Vancouver, BC, Canada, 2020, pp. 7148-7160.
- Y. Ji, C. Tan, S. Martschat, Y. Choi, and N. A. Smith, "Dynamic entity representations in neural language models," 2017, arXiv:1708.00781.
- L. Liu, X. Ren, J. Shang, X. Gu, J. Peng, and J. Han, "Efficient contextualized representation: Language model pruning for sequence labelling," in Proc. Conf. Empirical Methods Natural Lang. Process., 2018, pp. 1215-1225.
- R. Caruana, "Multitask learning," Mach. Learn. J., vol. 28, no. 1, pp. 41-75, 1997.



Integration of Multi-Task Learning in Deep Learning-Based Language Models for Improving Natural Language Understanding: A Literature Review

Ali Robati Javazm

Technical Instructor, Department of Electrical and Computer Engineering, National Skills University of Tehran

Abstract

Recent advances in the field of Natural Language Processing (NLP), particularly with the advent of deep learning models, have led to significant performance improvements across a wide range of tasks. However, generalizing these models to all NLP tasks remains a challenge. Natural Language Understanding (NLU), as a subset of NLP, focuses on the semantic understanding of text and includes tasks such as machine translation and sentiment analysis. Multi-Task Learning (MTL), as a method for simultaneously training multiple models, has shown potential for improving performance in NLU. This research, through a systematic review of the literature, examines the language models used in NLU and analyzes their strengths and weaknesses. The results of this research indicate that language models based on unsupervised learning methods have high potential for performance improvement. However, designing a general framework for language models that can simultaneously perform multiple NLU tasks with high performance remains a major challenge. This research, by presenting a conceptual framework, outlines the future directions of research in this area.

Keywords: Deep Learning, Knowledge Representation, Multi-Task NLU, Unsupervised Learning