مقاله پژوهشی

کاربست انواع جانمایی کلمات پیش‌آموزش‌داده‌شده در مدل‌های یادگیری عمیق برای تولید عنوان از متون فارسی

محمدابراهیم شناسا و بهروز مینایی بیدگلی

چکیده: با پیدایش روش‌های یادگیری عمیق، مدل‌های دنباله به دنباله با هدف ترجمه ماشینی یک جمله مبدأ به یک جمله مقصد ارائه شدند و از همین ایده برای ترجمه یا تبدیل یک متن به شکل خلاصه‌شده آن استفاده گردیده است. خلاصه‌هایی که به این روش تولید می‌شوند از نظر ساختاری خواناتر بوده و معمولاً معنای کاملی را به خواننده منتقل می‌کنند. در چنین ساختارهایی برای بازنمایی معنایی واژه‌ها از بردارهای جانمایی کلمات استفاده می‌شود که در آن، وزن هر کلمه با توجه به کلمات مجاور آن از یک پیکره بزرگ آموزش داده می‌شود. در حالت کلی وزن این بردارها با انتخاب یک پنجره مجاورت برای هر کلمه به‌دست می‌آید؛ اما در مدل‌های زبانی بافتاری مانند برت برای محاسبه وزن این کلمات از مبدل‌های چندلایه استفاده می‌شود که به تمامی کلمات موجود در متن توجه می‌کنند. تاکنون مقالات متعددی نشان داده‌اند که مدل‌های زبانی بافتاری به‌دلیل قابلیت ریزتنظیم وزن‌ها برای انجام یک وظیفه پردازش زبان طبیعی خاص، موفق‌تر از سایر روش‌های جانمایی کلمات عمل می‌کنند؛ اما بررسی عملکرد وزن اولیه این مدل‌ها برای کاربست در تولید عنوان در زبان فارسی مورد توجه قرار نگرفته است.

در این مقاله به بررسی رفتار جانمایی کلمات به‌صورت پیش‌آموزش‌داده‌شده و بدون ریزتنظیم آنها در تولید عنوان از متون فارسی می‌پردازیم. برای یادگیری مدل از «علم‌نت» که یک پیکره فارسی شامل حدود ۳۵۰ هزار جفت چکیده و عنوان مقالات علمی می‌باشد، استفاده شده است. نتایج نشان می‌دهند استفاده از مدل برت حتی بدون ریزتنظیم وزن‌های آن در بهبود کیفیت عناوین فارسی تولیدشده تأثیرگذار بوده و معیار ROUGE-۱ را در فارسی به ۴۲٪ می‌رساند که بهتر از سایر مدل‌های جانمایی است.

کلیدواژه: **یادگیری عمیق، مدل دنباله به دنباله مبتنی بر برت، خلاصه‌سازی چکیده‌ای، تولید عنوان، دادگان مقالات فارسی.**

۱- مقدمه

امروزه به‌دلیل وجود حجم بالایی از متون، افراد نیاز دارند بدون صرف وقت و تنها با داشتن عنوانی که موضوع اصلی متن را برساند، آن را بیابند. از این رو تولید خودکار عنوان برای متون، یکی از دغدغه‌هایی است که به آن پرداخته می‌شود. عنوان متن، نمایشی خلاصه‌شده از متن و با طولی کمتر از یک جمله است که به کمک آن می‌توان جان کلام یک متن را بدون خواندن تمام آن فهمید.

روش‌های سنتی برای تولید عنوان را می‌توان به دو دسته کلی تقسیم نمود؛ روش‌های پایین به بالا[1] که در آن مجموعه‌ای از مفاهیم، رویدادها و عبارات منتخب متن با استفاده از روش‌های ترکیب جمله در کنار هم قرار می‌گیرند؛ اما معمولاً عناوین تولیدشده به این شکل، ساختار نحوی صحیح و کاملی ندارند. روش‌های بالا به پایین[2] که با انتخاب جملات مهم، سعی در حذف اجزای زاید آن با حفظ ساختار نحوی جمله دارند که در این حالت نیز جمله کاهش‌یافته بعضاً از نظر معنایی کامل نیست [۱].

موفقیت‌های اخیر روش‌های یادگیری عمیق در تولید متون روان و پرمحتوا سبب شده که محققان به استفاده از این روش‌ها در خلاصه‌سازی متن روی بیاورند. از ویژگی مهم این روش‌ها می‌توان به نمایش معنایی کلمات در قالب بردارهای جانمایی اشاره نمود. در یک تقسیم‌بندی، بردارهای جانمایی به دو دسته پیش‌آموزش‌دیده[3] و بافتاری[4] تقسیم می‌شوند. در روش پیش‌آموزش‌دیده، وزن‌های بردار جانمایی هر کلمه با توجه به یک پنجره همسایگی محدود از کلمات مجاور در یک پیکره بسیار بزرگ به‌دست می‌آید. این وزن‌ها طی فرایند آموزش و استفاده در یک وظیفه پردازش متن، ثابت بوده و دستخوش تغییر نمی‌شوند. در عوض، وزن بردارهای بافتاری با آموزش یک شبکه مبدل[5] چندلایه به‌دست می‌آید که برای تخمین وزن‌های بردار کلمه به تمامی کلمات موجود در متن توجه می‌کند. با توجه به قابلیت مبدل‌ها، این وزن‌ها طی فرایند آموزش متناسب با یک وظیفه پردازشی خاص می‌توانند ریزتنظیم[6] شوند. از کاربردی‌ترین مدل‌های بافتاری می‌توان به مدل برت (BERT)[7] اشاره نمود که در ادامه معرفی می‌گردد. مقالات متعددی روی قابلیت ریزتنظیم مدل‌های بافتاری تمرکز کرده و نشان داده‌اند این مدل‌ها به‌دلیل استفاده از این ویژگی، نسبت به سایر مدل‌های جانمایی موفق‌تر عمل می‌کنند [۲]؛ اما کمتر مقاله‌ای می‌توان یافت که به ارزیابی تأثیر وزن‌های اولیه این مدل‌ها صرف نظر از قابلیت ریزتنظیم آنها بپردازد.

طبق بررسی ما مقالات ارائه‌شده در زمینه تولید خودکار عنوان از متون فارسی بسیار نادر است که تنها می‌توان به [۳] اشاره نمود که دو روش پایه مبتنی بر مدل برت را بر روی یک دادگان متشکل از حدود ۹۳ هزار متن خبری ارزیابی کرده است؛ در حالی که مقالات ارزیابی معتبر از نمونه دادگان‌های استانداردی مانند Gigaword و CNN-DailyMail استفاده

این مقاله در تاریخ ۲۵ مرداد ماه ۱۴۰۱ دریافت و در تاریخ ۳۱ مرداد ماه ۱۴۰۲ بازنگری شد.

محمد ابراهیم شناسا، دانشکده برق و کامپیوتر، واحد علوم و تحقیقات دانشگاه آزاد اسلامی، تهران، ایران، (email: me.shenasa@iau-tnb.ac.ir).

بهروز مینایی بیدگلی (نویسنده مسئول)، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران، (email: b_minaei@iust.ac.ir).

1. Top-Down
2. Bottom-Up
3. Pre-Trained
4. Contextualized
5. Transformer
6. Fine-Tune
7. Bidirectional Encoder Representations from Transformers

می‌کنند که در حدود چند صد هزار متن خبری دارند.

مقاله [۴] با معرفی مجموعه دادگان «علم‌نت» حاوی ۳۵۰۰۰۰ نمونه چکیده- عنوان از مقالات علمی و ارزیابی روش‌های پایه یادگیری عمیق روی آن نشان داده که می‌توان از این مجموعه به‌عنوان یک دادگان معیار برای ارزیابی روش‌های خلاصه‌سازی در فارسی استفاده نمود.

در این مقاله مدل‌های مختلف جانمایی را برای تولید عنوان از مجموعه مقالات علمی علم‌نت به‌کار می‌گیریم. به عبارت دیگر نوآوری این مقاله، ارائه چارچوبی استاندارد و جامع‌نگر برای ارزیابی عملکرد طیف متعددی از مدل‌های جانمایی کلمات و روش‌های یادگیری عمیق برای تولید عنوان از متون فارسی می‌باشد. نتایج به‌دست‌آمده نشان می‌دهند استفاده از مدل بافتاری حتی در حالت استفاده از وزن‌های اولیه آن، سبب بهبود کیفیت خلاصه تولیدشده نسبت به سایر مدل‌های جانمایی می‌شود؛ به‌طوری که استفاده از مدل مبدل با به‌کارگیری مدل برت چندزبانه[1] بهترین نتیجه را می‌دهد. در ادامه مقاله ابتدا کارهای انجام‌شده در زمینه تولید خودکار عنوان مورد بررسی قرار می‌گیرد؛ پس از آن مجموعه داده و روش‌های به‌کارگرفته‌شده معرفی می‌گردند و نهایتاً به انجام آزمایش‌ها و بحث بر روی نتایج پرداخته می‌شود.

۲- پیشینه تحقیق

روش‌های خلاصه‌سازی به دو دسته استخراجی[2] و چکیده‌ای[3] تقسیم می‌شوند. روش‌های استخراجی چند جمله مهم از متن را به‌عنوان خلاصه انتخاب می‌کنند؛ اما روش‌های چکیده‌ای با کاهش و بازنویسی عبارات، سعی در تولید یک خلاصه جدید و بامعنا دارند. از آنجا که معمولاً عنوان، خلاصه‌ای کوتاه‌تر از یک جمله است، روش‌های استخراجی در این زمینه چندان موفق عمل نمی‌کنند. روش‌های چکیده‌ای به‌کارگرفته‌شده برای تولید عنوان به روش‌های بالا به پایین و پایین به بالا تقسیم می‌شوند.

از روش‌های بالا به پایین می‌توان به انتخاب تک‌جمله مهم و کاهش آن با استفاده از مجموعه‌ای از قوانین دستی و ازپیش‌تعریف‌شده اشاره کرد [۵] تا [۷]. در برخی تحقیقات، روش‌های احتمالاتی و مدل‌های یادگیری ماشین برای انتخاب عبارات مهم جمله به‌کار گرفته می‌شوند که در اغلب آنها از گرامرهای احتمالاتی برای کاهش و حذف عبارات استفاده شده است [۸] تا [۱۰]. برخی از روش‌ها با درنظرگرفتن این نکته که عبارات عنوان در جملات متعددی از متن پخش شده‌اند، به کاهش چندجمله‌ای متن برای رسیدن به یک عنوان یکپارچه روی آورده‌اند [۱۱]. در تحقیق دیگری برای کاهش جمله از مدل LSTM[4] که مبتنی بر یادگیری عمیق است، استفاده می‌گردد [۱۲] که در آن، واژه‌های جمله به دنباله‌ای از صفر و یک‌ها ترجمه می‌شوند که نشانه حذف یا باقی‌ماندن واژه است. این مدل، واژه‌های جمله را به‌عنوان ورودی دریافت کرده و بهترین زیردنباله ممکن را به‌ازای تابع احتمالاتی برمی‌گرداند. در برخی از روش‌ها برای کاهش جملات موجود در توئیت‌های کاربران از تحلیل احساسات جهت کشف کلماتی که قطبیت بیشتری دارند استفاده شده و به این ترتیب، قسمت‌های غیرضروری جملات شناسایی و حذف می‌گردند [۱۳] و [۱۴].

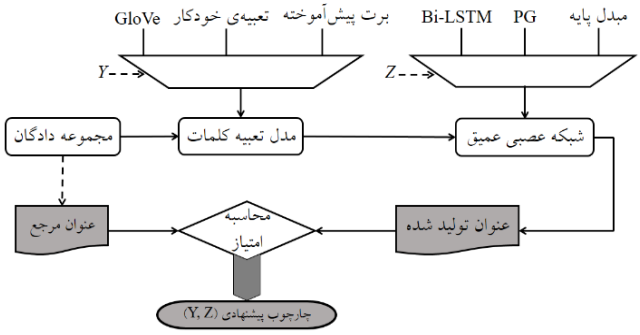
از روش‌های پایین به بالا می‌توان به استفاده از مدل‌های آماری ترجمه ماشینی مبتنی بر کانال نویزی برای نگاشت متن به‌عنوان اشاره نمود [۱۵]. در برخی از تحقیقات پس از انتخاب عبارات مهم جمله با استفاده از گرامر شبه هم‌زمانی[5]، عبارت مورد نظر کاهش داده شده یا بازنویسی می‌گردد [۱]. در یکی از تحقیقات با استفاده از یک مدل مبتنی بر رویداد، گرافی از روابط سه‌تایی موجود در متن به‌همراه واژه‌های آنها ساخته شده و سپس از یک الگوریتم فشرده‌سازی چندجمله‌ای برای ادغام رویدادها جهت تولید عنوان استفاده می‌گردد [۱۶].

اخیراً به‌دلیل موفقیت مدل‌های یادگیری عمیق در خلاصه‌سازی، کارهای زیادی به تولید عنوان با استفاده از روش‌های یادگیری عمیق پرداخته‌اند. از نخستین تحقیقات انجام‌شده در این زمینه می‌توان به فشرده‌سازی جمله از طریق حذف یا باقی‌گذاشتن کلمات جمله با استفاده از مدل شبکه عصبی بازگشتی RNN)[6]) اشاره نمود [۱۲]. در این تحقیق برای یادگیری مدل از دو میلیون جفت جمله و فشرده آن استفاده گردیده و نشان داده شده که این مدل حتی بدون درنظرگرفتن ویژگی‌های نحوی و معنایی کلمات، بهتر از یادگیری سنتی عمل می‌کند. معماری حاکم بر اکثر روش‌های لبه دانش استفاده از مدل کدگذار- کدگشاست که معمولاً یک سازوکار یادآوری[7] نیز به آن افزوده می‌شود.

روش‌های اولیه از یک شبکه عصبی بازگشتی و به‌طور خاص LSTM به‌عنوان کدگذار و کدگشا استفاده می‌کردند [۱۲]، [۱۷] و [۱۸]. به‌دلیل تمایل این مدل‌ها به تکرار کلمات در خروجی- به‌خصوص در حالت چندجمله‌ای- [۱۹] از یک شبکه PG[8] استفاده کرد که هم‌زمان امکان تولید کلمه جدید در خروجی را علاوه بر کپی از متن اصلی فراهم می‌کرد و در عین حال یک سازوکار همگرایی هم برای جلوگیری از تکرار به‌کار گرفته شد. مقاله [۲۰] معماری دنباله به دنباله مبتنی بر LSTM را با لایه‌های پیچشی[9] ترکیب کرده و نشان داده که به‌دلیل توانایی شبکه پیچشی در تشخیص الگوها و ویژگی‌های مختلف در داده‌های دنباله‌ای، نتایج بهتری نسبت به مدل‌های مبتنی بر LSTM تولید می‌شود. مقاله [۲۱] با عمومی‌سازی واژه‌های متن به مفاهیم بالادستی، برداری از مفاهیم سطح بالا را به‌عنوان ورودی به مدل می‌دهد و خلاصه تولیدشده را پس از یک مرحله پس‌پردازش به مفاهیم اصلی بازمی‌گرداند. مقاله [۲۲] از شبکه‌های مبدل چندلایه به‌عنوان کدگذار و کدگشا استفاده کرده و نشان داده که نتایج نسبت به مدل مبتنی بر LSTM به‌نحو قابل توجهی بهبود می‌یابد.

مقاله [۲] از یک معماری مبتنی بر BERT برای خلاصه‌سازی استفاده کرده که در این معماری از مدل BERT پیش‌آموخته به‌عنوان رمزگذار و یک شبکه مبدل شش‌لایه به‌عنوان رمزگشا استفاده شده است. نویسندگان نشان دادند که استفاده از مدل‌های مبتنی بر BERT به‌صورت قابل توجهی نتایج را نسبت به سایر مدل‌های دنباله به دنباله افزایش می‌دهد. نمونه‌های بهبودیافته‌ای از BERT مانند RoBERTa ارائه شده‌اند که با ایجاد تغییراتی از قبیل استفاده از داده‌های آموزشی بیشتر، مدت زمان آموزش طولانی‌تر و استفاده از جعبه لغت بزرگ‌تر در برخی از حوزه‌های پردازش زبان طبیعی موفق‌تر عمل کرده‌اند [۲۳]؛ اما به‌دلیل دسترسی و مقبولیت بیشتر BERT، اکثر روش‌های خلاصه‌سازی از همین مدل استفاده کرده‌اند. در [۲۴] به بررسی و ارتقای مفهوم انتقال یادگیری[10] با

1. Multilingual-BERT
2. Extractive
3. Abstractive
4. Long Short Term Memories
5. Quasi-Synchronous Grammar
6. Recurrent Neural Network
7. Attention Mechanism
8. Pointer-Generator
9. Convolutional
10. Transfer Learning



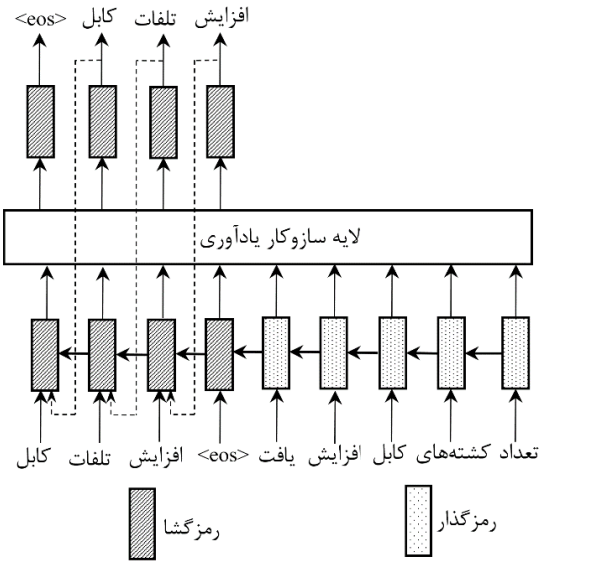
شکل ۱: چارچوب پیشنهادی.

استفاده از معماری متن‌به‌متن تبدیلگر پرداخته شده است. ایده اصلی این است که می‌توان اکثریت مسائل زبانی را به یک قالب مشترک تبدیل کرده و از یک مدل بزرگ انتقال یادگیری در تمام این مسائل استفاده کرد. این مقاله با استفاده از مکانیزم توجه مشترک و تعبیه‌گرها، معماری متن‌به‌متن تبدیلگر را برای مسائل ترجمه ماشینی، تولید متن و پرسش و پاسخ به‌کار گرفته و به نتایج بهتری نسبت به مدل‌های یادگیری پایه رسیده است. از آنجا که BERT با هدف پیش‌بینی کلمات پوشانده‌شده در متن (و نه خلاصه‌سازی) پیش‌آموزش دیده است و صرفاً یک مؤلفه رمزگذار وزن‌دهی‌شده می‌باشد، [۲۵] یک مدل رمزگذار- رمزگشا را با هدف خلاصه‌سازی متن پیش‌آموزش و نشان داده که کیفیت خلاصه‌های تولیدشده به‌مراتب بهتر از به‌کارگیری وزن‌های اولیه BERT در فرایند خلاصه‌سازی است. مقاله [۲۶] برای رفع محدودیت مدل BERT که صرفاً به پیش‌بینی کلمات پوشانده‌شده می‌پردازد از یک مدل تولیدگر روبه‌جلو برای یادگیری اصلاح متنی که کلمات آن به هم ریخته است، به شکل صحیح آن استفاده می‌کند و پس از مرحله پیش‌آموزش، این مدل برای تولید خلاصه از متن استفاده می‌شود. مقاله [۲۷] از یک مبدل چندلایه به‌عنوان رمزگذار و رمزگشا استفاده کرده و ویژگی این روش آن است که از یک مدل زبانی مبتنی بر پوشانه[1] یاد می‌گیرد که دنباله کلماتی کلماتی را که از متن اصلی به خروجی کپی می‌شود کنترل کند. نتایج به‌دست‌آمده نشان می‌دهند که اگر روش صرفاً بر روی کلمات دیده‌شده در متن اصلی آموزش داده شود به‌جای تولید کلمات جدید در خروجی، تمایل به استفاده از کلمات دیده‌شده در خروجی پیدا می‌کند. در عوض هرچه کلمات نادیده بیشتری در فرایند یادگیری ظاهر شوند، مدل به‌جای اینکه صرفاً کلمات را کپی کند به تولید کلمات جدید روی می‌آورد. در ادامه، چارچوب پیشنهادی را برای ارزیابی مدل‌های متنوع یادگیری عمیق برای تولید عنوان از متون فارسی ارائه می‌کنیم.

۳- چارچوب پیشنهادی

برای ارزیابی جامع کارایی مدل‌های یادگیری عمیق جهت تولید عنوان، چارچوبی به‌صورت شکل ۱ پیشنهاد می‌شود که طبق آن، هر مدل از دو پارامتر اصلی Y و Z تشکیل شده که Y نوع مدل جانمایی کلمات و Z شبکه عصبی عمیق به‌کارگرفته‌شده را نمایش می‌دهد. برای ارزیابی هر چارچوب، شبکه عصبی عمیق آن با دادگان و مدل جانمایی منتخب آموزش می‌بینند؛ مدلی برای خلاصه‌سازی متن، ایجاد و سپس کارایی مدل با مجموعه داده آزمون سنجیده می‌شود. بدین ترتیب که عنوان تولیدشده توسط مدل با عنوان مرجع انسانی مقایسه گردیده و هر مدلی که عناوین باکیفیت‌تری تولید کند به‌عنوان مدل پیشنهادی معرفی می‌شود.

1. Masked Language Model



شکل ۲: نمایی از معماری رمزگذار- رمزگشا.

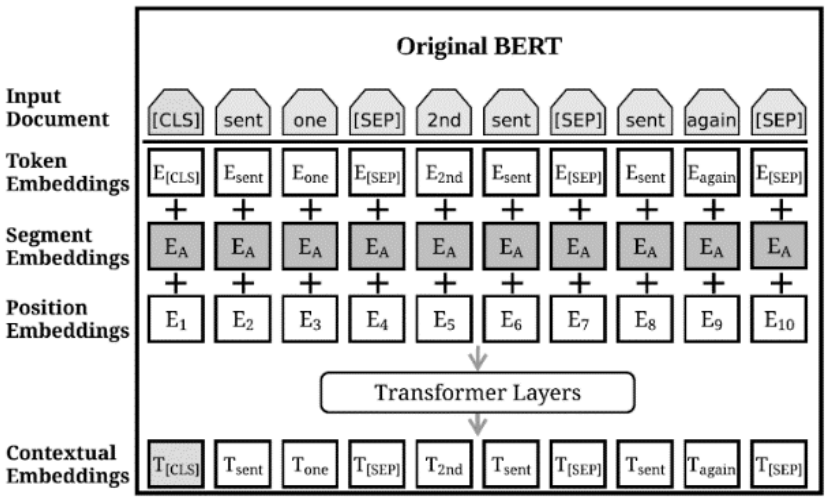
اکثریت قریب به اتفاق روش‌های خلاصه‌سازی مبتنی بر یادگیری عمیق از مدل یادگیری دنباله به دنباله استفاده می‌کنند. این مدل ابتدا در ترجمه ماشینی مصنوعی استفاده شد و معماری آن به‌صورت رمزگذار- رمزگشا[2] می‌باشد. نمایی کلی از این معماری در شکل ۲ مشاهده می‌شود. در این معماری کلمات ورودی یا متن اصلی یک‌به‌یک و به‌صورت برداری به رمزگذار داده می‌شود تا به دنباله‌ای از وضعیت‌ها تبدیل گردند. سپس در هر گام از یادگیری، کلمه قبلی خروجی یا خلاصه (در زمان آزمون، این کلمه حذف می‌شود) به رمزگشا داده می‌شود تا وضعیتی را تولید کند. در ادامه، لایه سازوکار یادآوری سعی می‌کند بردار وزنی از توجه به کلمات را طوری بسازد که رمزگشا با استفاده از آن و وضعیت فعلی خود بتواند کلمه بعدی خروجی را حدس بزند.

۳-۱ مدل زبانی برت پیش‌آموخته

مدل برت یک مدل نمایش زبان مبتنی بر بافت است که بردارهای جانمایی واژگان را با استفاده از مدل زبانی مبتنی بر پوشانه ایجاد می‌کند و بعداً می‌تواند با استفاده از بردارهای کلمات اطراف، پوشانه را پیش‌بینی کند. پیش‌آموزش[3] آن با استفاده از یک مدل مبدل چندلایه و به‌صورت بی‌ناظر بر روی حجم بسیار زیادی از متون انجام می‌شود و در ادامه می‌تواند بسته به کاربرد مورد نظر، ریزتنظیم شود.

معماری کلی برت در شکل ۳ آمده است. ابتدا علائم پردازشی به ابتدای متن و هر یک از جملات ورودی افزوده شده و ورودی به‌صورت دنباله‌ای از واژه‌ها مانند $X=[w_1,\ldots,w_n]$ به مدل داده می‌شود. برای هر واژه w_i سه نوع جانمایی در نظر گرفته می‌شود: ۱) جانمایی واژه که نمایانگر معنای واژه است؛ ۲) جانمایی قطعه‌بندی[4] که برای تمایز میان دو جمله می‌باشد و ۳) جانمایی جایگاه واژه که بیانگر مکان آن واژه در متن است. بردار تلفیق‌شده این سه بردار جانمایی به‌عنوان ورودی به یک مبدل چندلایه داده می‌شود که هر لایه از سازوکار یادآوری چندسَره[5] استفاده می‌کند. مدل برت برای هر واژه در بالاترین لایه، یک بردار خروجی t_i حاوی اطلاعات مبتنی بر بافت تولید خواهد کرد.

2. Encoder-Decoder
3. Pre-Training
4. Segmentation Embedding
5. Multi-Head Attention



شکل ۳: معماری مبتنی بر مدل برت برگرفته از [۲].

برت چندزبانه مانند مدل پایه آن، یک مبدل دوازده‌لایه شامل ۷۶۸ واحد پنهان می‌باشد که بر روی متون ویکی‌پدیای حدود ۱۰۰ زبان از جمله فارسی آموزش دیده شده است. برت فارسی نیز بر اساس ساختار برت چندزبانه می‌باشد و بر روی حدود ۹۵ میلیون کلمه آموزش دیده شده است.

۳-۲ شبکه‌های عمیق عصبی تولید عنوان

در این بخش به بررسی و تحلیل رفتار شبکه‌های عصبی عمیق مورد استفاده و معرفی نقاط ضعف و قوت آنها می‌پردازیم.

۳-۲-۱ روش LSTM دوجهته

در این روش از یک LSTM دوجهته به‌عنوان رمزگذار و یک LSTM یک‌جهته به‌عنوان رمزگشا و همچنین سازوکار یادآوری برگرفته از [۲۸] برای تولید خلاصه از متن استفاده می‌شود. همچنین هر یک از کلمات متن با یک مدل جانمایی پیش‌آموخته (مانند GloVe[1] [۲۹]) می‌تواند نمایش داده شود.

دنباله کلمات ورودی در فرایند یادگیری به شکل $X=(x_1,x_2,\ldots,x_n)$ و در قالب بردارهای جانمایی، یکی‌یکی و به‌صورت روبه‌جلو و روبه‌عقب به رمزگذار داده می‌شود که منجر به تولید وضعیت پنهان به شکل $h_i = bi_lstm(x_i, h_{i-1})$ برای بردار جانمایی x_i می‌شود. بردار بافت متن، یک توزیع یادآوری روی کلمات متن ورودی است که به رمزگشا می‌گوید در لحظه تولید خروجی به کدام کلمات توجه بیشتری داشته باشد. به عبارت دقیق‌تر، هر یک از عددهای این بردار، جمع وزنی وضعیت‌های رمزگذار است که توسط (۱) تا (۳) محاسبه می‌شود

$$c_t = \sum_{i=1}^{|X|} a_i^t h_i \qquad (۱)$$

$$a_i^t = \mathrm{softmax}(e_i^t) \qquad (۲)$$

$$e_i^t = \tanh(w_h h_i + w_s s_{t-1} + b) \qquad (۳)$$

که a_i^t میزان توجهی است که در هر لحظه از زمان مانند t به وضعیت h_i از رمزگذار داده می‌شود. پارامتر e_i^t میزان شباهت کلمه ورودی x_i با خروجی را در لحظه t نشان می‌دهد. W_h، W_s و b پارامترهایی هستند که باید آموزش داده شوند و s_{t-1} وضعیت قبلی رمزگشاست. از طرف دیگر دنباله کلمات خروجی به شکل $Y=(y_1,y_2,\ldots,y_n)$ به رمزگشا داده می‌شود تا با دیدن کلمه y_{t-1}، بردار بافت[2] متن c_t و وضعیت رمزگشا که از رابطه $s_t = lstm(s_{t-1}, y_{t-1}, c_t)$ به‌دست می‌آید، یاد بگیرد کلمه y_t از خلاصه را حدس بزند. احتمال حضور یک کلمه در خلاصه با استفاده از یک توزیع احتمالاتی روی جعبه لغات به‌دست می‌آید که توسط (۴) محاسبه می‌شود

$$P_{vocab} = \mathrm{softmax}(V'(V[s_t, c_t]+b)+b') \qquad (۴)$$

که در این رابطه V، V'، b و b' پارامترهایی هستند که آموزش داده می‌شوند. با ایده از [۱۹] نمای کلی فرایند یادگیری در شکل ۴ مشاهده می‌گردد.

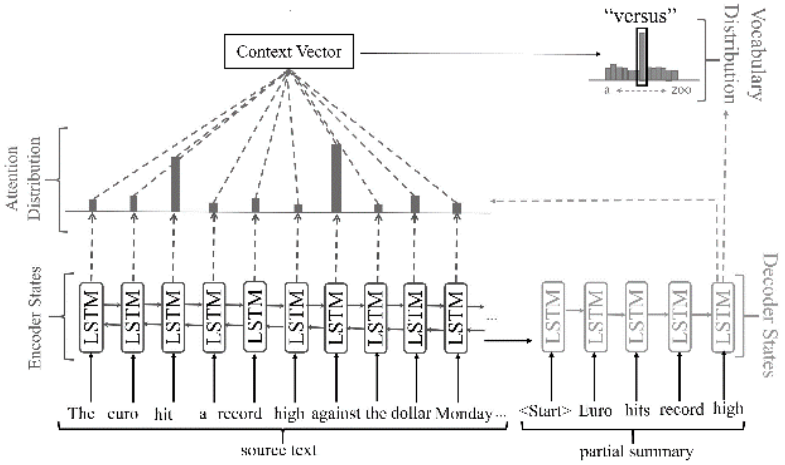
۳-۲-۲ شبکه‌های PG

از مشکلات روش مبتنی بر LSTM دوجهته آن است که تمایل به تکرار کلمات تولیدشده در خروجی دارد؛ به همین جهت شبکه PG ارائه شد که برای تولید خلاصه، هم امکان کپی‌کردن کلمه از متن اصلی توسط توزیع یادآوری و هم امکان تولید کلمه جدید با توزیع روی جعبه لغات را به‌وجود می‌آورد. همچنین برای جلوگیری از تکرار کلمات در خلاصه از برداری به نام بردار پوشش[3] استفاده می‌کند که میزان استفاده کلمات توسط توزیع یادآوری را نشان می‌دهد. در حقیقت می‌توان گفت احتمال نهایی حضور یک کلمه در خلاصه، تابعی از توزیع احتمالاتی سازوکار یادآوری کلمات، توزیع بردار پوشش و توزیع روی کل جعبه لغات است.

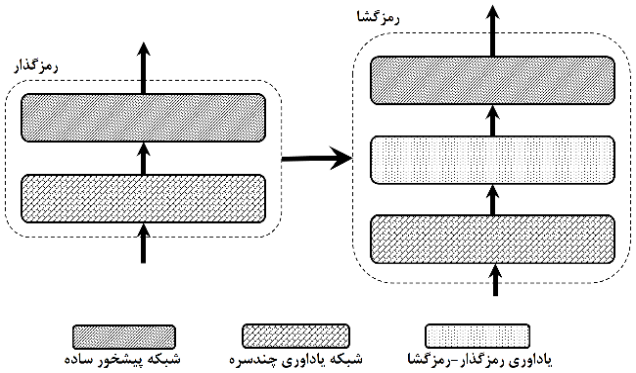
۳-۲-۳ مبدل چندلایه

روش LSTM دوجهته صرفاً از سازوکار یادآوری به‌عنوان پلی برای ارتباط میان رمزگذار و رمزگشا استفاده می‌کند؛ اما در این روش رمزگذار و رمزگشا از جنس مبدل‌های چندلایه هستند و هر یک به تنهایی مجهز به سازوکار یادآوری می‌باشند. هر لایه از مبدل، خود از دو زیرلایه تشکیل شده که زیرلایه اول، سازوکار یادآوری چندسره است. در سازوکار یادآوری چندسره، هر یک از بردارهای یادآوری کلمات از ترکیب خطی بردارهای سایر کلمات متن تشکیل می‌گردد. زیرلایه دوم، یک شبکه پیشخور[4] تمام‌متصل ساده بر اساس مکان کلمات می‌باشد. نمای کلی معماری یک شبکه مبدل دولایه در شکل ۵ مشاهده می‌گردد.

1. Global Vectors for Word Representation
2. Context Vector
3. Coverage Vector
4. Feed-Forward



شکل ۴: معماری مبتنی بر LSTM دوجهته برگرفته از [۱۹].



شکل ۵: نمایی از معماری رمزگذار- رمزگشا بر مبنای مبدل.

لازم به ذکر است که سازوکار یادآوری چندسره در رمزگشا از یک پوشانه برای یادآوری بر اساس کلمات قبلی که در خلاصه تولید شده است، استفاده می‌کند. علاوه بر این در رمزگشا لایه دیگری به نام یادآوری رمزگذار- رمزگشا[1] وجود دارد که به رمزگشا کمک می‌کند برای تولید کلمه جدید به وضعیت رمزگذارها نیز توجه کند [۳۰].

۳-۳ بررسی نظری کارایی مدل‌ها

همان طور که پیش‌تر اشاره شد، دو مدل LSTM و مبدل به‌عنوان دو شبکه کارا و رایج برای یادگیری داده‌های متنی دنباله‌دار مورد استفاده قرار می‌گیرند. مقاله شاخص [۳۰] در این میان به بررسی شبکه مبدل پرداخته و نشان داده که معماری آن از نظر تئوری، دو امتیاز اساسی نسبت به مدل LSTM دارد. امتیاز اول استفاده از مکانیزم خودتوجهی[2] است که سبب می‌شود مدل مبدل، بردار توجه به هر کلمه را به‌صورت ترکیبی خطی از بردار سایر کلمات متن محاسبه کند؛ برعکس مدل LSTM که توجه به هر کلمه را به‌صورت ترتیبی انجام می‌دهد و در به‌یادآوری اطلاعات توجه به کلمات دورتر در دنباله‌های خیلی طولانی دچار مشکل می‌شود. امتیاز دیگر، استفاده از مکانیزم توجه چندسره است که برای تمرکز هم‌زمان بر روی زیرمجموعه‌های مختلف داده‌های ورودی استفاده می‌شود و به مدل اجازه می‌دهد تا چندین نمایش بافتاری از داده‌ها را به‌صورت موازی بیاموزد و در پیش‌بینی کلمات و تولید دنباله خروجی بهتر عمل کند. بنابراین می‌توان انتظار داشت خلاصه‌های حاصل از شبکه مبدل نسبت به سایر شبکه‌های عصبی، کیفیت بهتری داشته باشند.

۴- آزمایش‌ها و تحلیل‌ها

جهت تحلیل رفتار مدل‌های جانمایی پیش‌آموخته، کارایی آنها را در روش‌های یادگیری عمیق در هر دو زبان فارسی و انگلیسی مورد تحلیل و بررسی قرار می‌دهیم.

۴-۱ تنظیمات آزمایشگاهی

در ادامه به معرفی دادگان و تنظیمات انجام‌شده بر روی شبکه‌های یادگیری عمیق معرفی‌شده اشاره خواهیم کرد.

۴-۱-۱ دادگان

دادگان خلاصه‌سازی مورد استفاده (دادگان علم‌نت [۴])، مجموعه‌ای حاوی ۳۵۰ هزار جفت چکیده مقاله و عنوان به زبان فارسی است که از سایت‌های نمایه‌کننده اسناد علمی مانند مگیران، نورمگز، ensani.ir، سیویلیکا و ایرانداک خزش[3] شده است. بر اساس یک آمارگیری اولیه از دادگان مورد نظر مشخص گردید که حدود ۹۰٪ کلمات خلاصه در ۴ جمله اول متن ظاهر می‌شوند؛ بنابراین ۴ جمله اول متن برای فرایند آموزش و آزمون انتخاب گردید. جهت به‌کارگیری روش‌های ارائه‌شده بر روی این دادگان، آماده‌سازی‌هایی متناسب با هر یک از روش‌ها بر روی دادگان صورت گرفت. علاوه بر این پیش‌پردازش‌هایی نظیر نرمال‌سازی متن و جداسازی واژه‌ها با استفاده از ابزار پردازش فارسی هضم [۳۱] بر روی داده‌ها انجام شد.

۴-۱-۲ روش پایه

طبق بررسی ما، تحقیق شاخصی در خصوص خلاصه‌سازی چکیده‌ای متون فارسی با روش‌های باناظر وجود ندارد و آخرین مقاله ارائه‌شده در زمینه خلاصه‌سازی چکیده‌ای [۳۲] با روش بی‌ناظر و مبتنی بر گراف است؛ به همین دلیل از یک روش پایه ابداعی استفاده کردیم. در این روش، عبارات متن با استفاده از جداساز عبارت[4] موجود در ابزار هضم، استخراج گردید و به یک رده‌بند مبتنی بر SVM[5] برای امتیازدهی داده شد. ویژگی‌های آموزش‌داده‌شده به این رده‌بند، مکان عبارت در جمله، برچسب‌های صرفی، موجودیت نامدار و ویژگی‌های آماری نظیر

1. Encoder-Decoder Attention
2. Self-Attention
3. Scrawl
4. Chunker
5. Support Vector Machine

جدول ۱: نتایج خلاصه‌سازی روی داده آزمون علم‌نت.

روش‌های پایه و آزمایش‌شده خلاصه‌سازی	ROUGE -۱	ROUGE -۲	ROUGE-L
جمله اول	۲۰	۱۰	۱۵
جمله دوم	۱۵	۷	۱۲
جمله سوم	۱۵	۶	۱۱
مبتنی بر عبارت	۲۹	۱۴	۲۵
LSTM دوجهته با جانمایی خودکار	۴۰	۲۶	۳۷
LSTM دوجهته با بردار GloVe	۴۱	۲۷	۳۸
LSTM دوجهته با بردار ML-BERT	۳۸	۲۵	۳۵
شبکه PG با جانمایی خودکار	۳۹	۲۵	۳۵
شبکه PG با بردار GloVe	۴۰	۲۶	۳۶
شبکه PG با بردار ML-BERT	۳۸	۲۶	۳۵
مبدل با جانمایی خودکار	۴۰	۲۶	۳۶
مبدل با بردار GloVe	۴۱	۲۶	۳۷
مبدل با بردار ML-BERT	**۴۲**	**۲۸**	**۳۸**

TFIDF[1] و فیلد هدف آن، وجود یا عدم وجود عبارت در خلاصه می‌باشد. سپس این عبارات بر اساس امتیاز مرتب گردید و در کنار هم قرار گرفت تا خلاصه را تشکیل دهد. چون میانگین طول خلاصه دادگان ۱۴ کلمه بود، در این روش که به نام روش «مبتنی بر عبارت» در جدول ۱ دیده می‌شود، حداکثر طول خلاصه ۱۴ کلمه در نظر گرفته شد. همچنین ۱۴ کلمه ابتدایی جمله اول، دوم و سوم هر متن نیز به‌عنوان یک روش پایه به‌کار گرفته شد.

٤-۱-۳ تنظیم پارامترها

برای پیاده‌سازی و آزمایش روش‌ها مستقیماً از ابزار open-NMT استفاده گردیده است [۲۲]. تعداد مستندات مورد استفاده برای آموزش ۳۰۰۰۰۰ نمونه، برای آزمون ۳۰۰۰۰ نمونه و برای ارزیابی ۵۰۰۰ نمونه است. در همه روش‌ها آموزش طی ۱۰۰۰۰۰ گام صورت می‌گیرد که هر ۱۰۰۰۰ گام، مدل ذخیره و ارزیابی می‌شود. جعبه لغت مورد استفاده در متن و خلاصه مشترک بوده و به تعداد ۵۰۰۰۰ لغت است.

هر یک از سه روش با سه نوع بردار جانمایی مورد آزمون قرار گرفت. در نوع اول از یک لایه جانمایی خودکار درون مدل استفاده گردید که در جدول ۱ به نام «جانمایی خودکار» مشخص شده است. در نوع دوم از بردارهای پیش‌یادگیری‌شده GloVe با ۳۰۰ بعد که شامل حدود ۴۰۰۰۰۰ کلمه است، استفاده شد [۲۹]. لازم به ذکر است که بردارهای پیش‌آموزش‌دیده متنوعی برای جانمایی کلمات ارائه شده که معروف‌ترین آنها GloVe و word2vec هستند. مدل word2vec صرفاً با استفاده از پنجره محدودی از همسایگی کلمات ایجاد می‌گردد؛ در حالی که در ساخت GloVe به بافت سراسری کلمات متن نیز توجه می‌شود. به عبارت بهتر بردار GloVe، یک ماتریس هم‌رخدادی از کل کلمات متن به‌وجود می‌آورد و این خاصیت سبب می‌شود برای خلاصه‌سازی متن مناسب‌تر باشد. این موضوع علی‌الخصوص در مورد متون تخصصی، بیشتر خود را نشان می‌دهد و در مقالات شاخصی به آن اشاره شده [۳۳] و [۳۴] و به همین دلیل از بردارهای GloVe استفاده گردید. در نوع سوم از بردارهای پیش‌آموزش‌دیده برت چندزبانه استفاده گردید که در ادامه توضیح داده شده است.

در مدل LSTM دوجهته، طول بردار کلمات ۱۲۸ در نظر گرفته شد و از سازوکار یادآوری مطابق شکل ۱ استفاده گردید. LSTM مورد استفاده برای رمزگذار، تک‌لایه و تعداد وضعیت‌های نهان آن ۵۱۲ می‌باشد. در زمان آزمون و تولید خلاصه از جستجوی مبتنی بر پرتو با اندازه ۵ استفاده شد و رمزگشا تا جایی که به نشانه انتهای دنباله برسد به تولید کلمات خلاصه ادامه داد.

در شبکه PG نرخ یادگیری برابر با ۰/۱۵ تعیین و روش بهینه‌سازی adagrad استفاده شده است. همچنین به‌دلیل امکان هم‌زمان کپی کلمات از متن و تولید کلمه جدید، خطای ازدست‌دادن کپی[2] بر اساس طول دنباله خروجی نرمال‌سازی می‌شود.

در مدل مبدل چندلایه، رمزگذار و رمزگشا ۴ لایه تعریف شده‌اند؛ طول بردار کلمات ۵۱۲ در نظر گرفته شده و از بهینه‌ساز adam استفاده گردیده است. روش‌های مبتنی بر مدل برت با استفاده از مدل برت پایه چندزبانه [۳۵] آزمایش شدند.

٤-۲ نتایج آزمایش‌ها

برای ارزیابی خلاصه‌های تولیدشده توسط مدل‌ها از معیار ROUGE استفاده گردیده است [۳۶] که معروف‌ترین اقسام آن ROUGE -۱، ROUGE -۲ و ROUGE-L هستند و به‌ترتیب به‌صورت نسبت یک گرمی‌ها، دو گرمی‌ها و طولانی‌ترین دنباله ترتیبی مشترک میان خلاصه ماشینی و انسانی، تقسیم بر طول خلاصه انسانی تعریف می‌شوند. جدول ۱ نتایج ارزیابی مدل‌های ارائه‌شده را نشان می‌دهد. روش‌هایی که بیشترین امتیاز را کسب کرده‌اند به‌صورت پررنگ در جدول مشخص شده‌اند.

جدول ۲ خروجی تولیدشده توسط برخی از روش‌ها را برای متن نمونه سند به‌صورت «پژوهش حاضر با هدف بررسی تأثیر قابلیت‌های بازاریابی بر استراتژی‌های رقابتی با نقش تعدیلگر شدت رقابت صورت گرفت. این پژوهش از لحاظ هدف، کاربردی و از نظر روش در زمره تحقیقات توصیفی- پیمایشی قرار می‌گیرد. جامعه تحقیق حاضر، کلیه کارکنان شرکت توکافولاد اصفهان به تعداد ۷۴۱ نفر است. برای تعیین حجم نمونه از فرمول کوکران استفاده شده که با توجه به جامعه، حجم نمونه مورد نیاز پژوهش ۲۵۳ نفر است» و عنوان اصلی به شکل «بررسی تأثیر قابلیت‌های بازاریابی بر استراتژی‌های رقابتی با نقش تعدیلگر شدت رقابت (مطالعه موردی: توکافولاد اصفهان)» را نمایش داده است.

در جدول ۳، خروجی تولیدشده برای نمونه متن سند به شکل «ارزیابی روند ورود و خروج انرژی در سامانه‌های کشاورزی»، یکی از روش‌های تعیین سطح پایداری در این سامانه‌هاست. از این رو در تحقیق حاضر، میزان انرژی‌های ورودی و خروجی و شاخص‌های مربوط به آن در مزارع آفتابگردان روغنی شهرستان خوی در سال زراعی ۹۶-۹۷ مورد مطالعه قرار گرفت. بر اساس نتایج حاصل از این پژوهش، بیشترین سهم از کل انرژی ورودی مربوط به نهاده‌های مصرفی شامل کود شیمیایی نیتروژن (۴۳٪/۹۸)، سوخت مصرفی (۲۵٪/۷۴) و ماشین‌آلات (۸٪/۴۲) بود. همچنین میزان اتکا به منابع تجدیدناپذیر انرژی برای تولید آفتابگردان در این منطقه زیاد است و باید برای جایگزین‌کردن منابع انرژی تجدیدپذیر به‌جای منابع تجدیدناپذیر تلاش گردد» و عنوان اصلی به‌صورت «ارزیابی انرژی مصرفی در مزارع آفتابگردان روغنی (مطالعه موردی: شهرستان خوی)» نمایش داده شده است.

1. Term Frequency-Inverse Document Frequency

2. Copy Loss

جدول ۲: نمونه اول خروجی خلاصه‌سازها.

روش	عنوان تولیدشده
LSTM با جانمایی خودکار	تأثیر قابلیت‌های بازاریابی بر استراتژی‌های رقابتی با نقش تعدیلگر شدت رقابت
LSTM با GloVe	تأثیر قابلیت‌های بازاریابی بر استراتژی‌های رقابتی با نقش تعدیلگر شدت رقابت
مبدل با جانمایی خودکار	بررسی تأثیر قابلیت‌های بازاریابی بر استراتژی‌های رقابتی با نقش رقابت
مبدل با GloVe	تأثیر قابلیت‌های بازاریابی بر استراتژی‌های رقابتی با نقش تعدیلگر رقابت

جدول ۳: نمونه دوم خروجی خلاصه‌سازها.

روش	عنوان تولیدشده
LSTM دوجهته	ارزیابی روند ورود و خروج انرژی در مزارع آفتابگردان روغنی
LSTM با GloVe	بررسی میزان انرژی‌های ورودی و خروجی و شاخص‌های مربوط به آن در مزارع آفتابگردان روغنی
مبدل با جانمایی خودکار	ارزیابی روند ورود و خروج انرژی در مزارع آفتابگردان روغنی
مبدل با GloVe	ارزیابی انرژی تجدیدپذیر برای تولید آفتابگردان روغنی در غرب کشور

٤-٣ تحلیل نتایج

از داده‌های جدول ۱ مشخص است که روش‌های مبتنی بر یادگیری عمیق نسبت به روش پایه (مبتنی بر عبارت) به نتایج بسیار بهتری دست یافته‌اند. همچنین دیده می‌شود که روش‌های یادگیری عمیق علی‌رغم تولید عناوینی با طول کمتر از یک جمله نسبت به هر سه جمله اول متن به امتیاز ROUGE بالاتری رسیده‌اند.

شکل ۶ امتیازات ROUGE به‌دست‌آمده توسط سه روش مورد نظر را بر اساس سه نوع جانمایی کلمه نشان می‌دهد. با نگاه کلی به نمودارها می‌توان ملاحظه کرد که بهترین نتیجه‌ها در حالت استفاده از شبکه مبدل با جانمایی برت چندزبانه به‌دست آمده است. این موضوع نشان می‌دهد استفاده از مدل برت حتی در حالت پیش‌آموخته آن از سایر انواع جانمایی بهتر عمل می‌کند. همچنین شبکه مبدل به‌دلیل سازوکار توجه چندسره نسبت به دو روش دیگر موفق‌تر بوده است.

مشاهده می‌شود روش‌ها در حالت استفاده از بردارهای GloVe، نتایج بهتری نسبت به لایه جانمایی خودکار داشته‌اند. چون در حالت جانمایی خودکار، لایه جانمایی به‌تدریج و با دیدن داده‌های جدید ساخته می‌شود و اندازه جعبه لغات در نظر گرفته شده در این حالت، حدود یک‌هشتم لغات GloVe می‌باشد، می‌توان انتظار داشت نتایج GloVe نسبت به جانمایی خودکار بهبود یابند.

با تحلیل نتایج به‌دست‌آمده، برخی ویژگی‌های خاص زبان فارسی وجود دارد که باید در فرایند تولید عنوان مد نظر قرار گیرد. یکی از این ویژگی‌ها ترکیب‌پذیری کلمات است که در زبان فارسی، کلمات و جملات از ترکیب‌پذیری بیشتری نسبت به سایر زبان‌ها برخوردارند و لازم است برای تولید عنوان از انتخاب و ترکیب صحیح کلمات استفاده شود؛ یا برای مثال، ساختار جملات فارسی به‌طور کلی از ساختار جملات زبان‌های اروپایی متفاوت است و در تولید عنوان، ساختار جملات صحیح و مناسب با زبان فارسی باید مد نظر قرار گیرد. همچنین استفاده از اصطلاحات و عبارات خاص زبان فارسی می‌تواند در تولید عنوان باکیفیت نقش مؤثری داشته باشد. با توجه به به این موارد، استفاده از مدل‌های جانمایی بافتاری (که در آنها به ترکیب و موقعیت کلمات و جملات توجه بیشتری می‌شود) مانند برت به‌جای سایر مدل‌های جانمایی توصیه می‌شود. همچنین استفاده از ساختار مبدل به‌دلیل توجه هم‌زمان به تمام کلمات می‌تواند موجب یادگیری بهتر ویژگی‌های مورد نظر زبان فارسی شود.

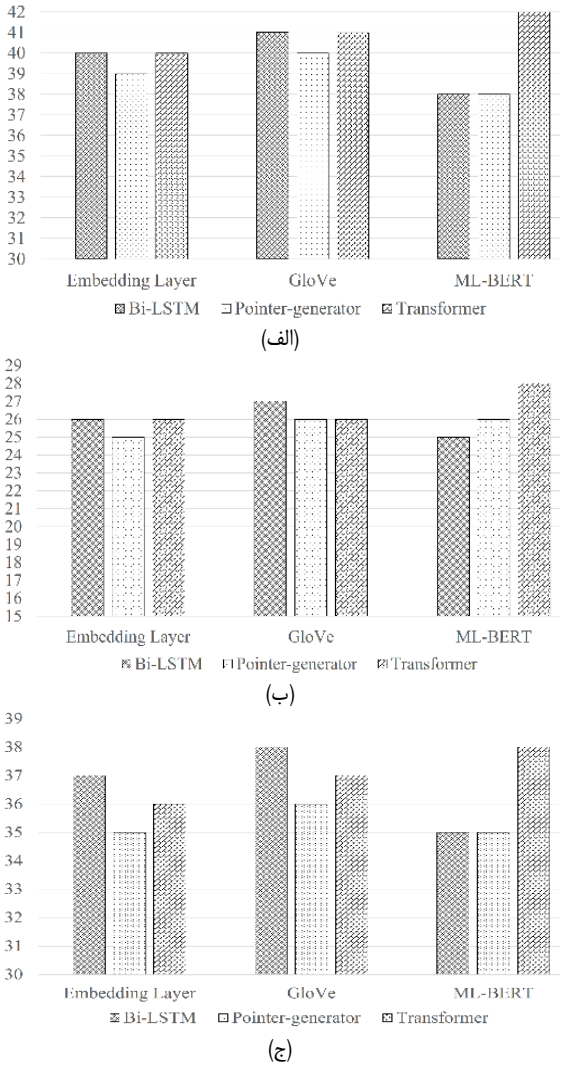
٤-٤ بررسی و تحلیل نتایج مدل‌ها در زبان انگلیسی

مشابه نتایجی که در بخش قبل به آنها اشاره شد، در آزمایش‌های متعدد بر روی زبان انگلیسی نیز به‌دست آمده و مقالات مختلف به آن اشاره کرده‌اند. به‌عنوان مثال، دسته‌ای از مقالات به این موضوع اشاره کرده‌اند که استفاده از بردارهای جانمایی پیش‌آموخته سبب بهبود کیفیت خلاصه‌ها نسبت به استفاده از جانمایی خودکار می‌شود [۳۷] تا [۳۹]. همچنین مقاله شاخصی نشان داده که مدل مبدل در مجموع نسبت به سایر مدل‌ها موفق‌تر عمل می‌کند [۴۰] که این نتیجه در مقاله ما نیز نشان داده شده است. همچنین طی تحقیقاتی که در مقاله شاخص [۴۱] به آن اشاره شده است، مبدل‌ها به‌دلیل استفاده از سازوکار یادآوری چندسره، توجه به هر کلمه را به‌صورت یک ترکیب خطی از توجه به همه کلمات موجود در متن بیان می‌کنند و همین امر سبب می‌شود که برایند دقیق‌تری از تولید کلمات خروجی داشته باشد. همچنین در مبدل‌ها میان سازوکار یادآوری رمزگذار و رمزگشا پلی برقرار شده که دقت وابستگی میان آنها را افزایش می‌دهد. همان طور که پیش‌تر گفته شد، وزن‌های اولیه مدلی مانند برت پایه از یک مبدل دوازده‌لایه به همراه سازوکار توجه درونی بر روی مجموعه بسیار بزرگی از متون پیش‌آموزش داده شده که این خود سبب می‌شود وزن‌های دقیق‌تری نسبت به بردارهای جانمایی ازپیش‌آموخته داشته باشد.

طی تحقیق جامعی که بر روی لایه‌های مختلف مدل‌های جانمایی بافتاری در [۴۱] انجام گردیده است، نشان داده شده که تمایز دو کلمه یکسان مانند «شیر» جنگل و «شیر» خوراکی که در دو بافت متنی متفاوت ظاهر می‌شوند در لایه‌های بالایی مدل بافتاری پایین است و هرچه به سمت لایه‌های پایین‌تر می‌رویم، این تمایز و تفاوت معنایی، خود را بیشتر نشان می‌دهد. این مقاله بردارهای جانمایی پیش‌آموخته و بافتاری را در چهار حوزه پردازش زبان طبیعی ارزیابی کرده و نشان داده که بردارهای جانمایی که حتی در لایه اول از مدل برت ایجاد می‌شوند، کارایی بهتری نسبت به سایر مدل‌های پیش‌آموخته دارند که بر یافته‌های ما در این مقاله مهر تأیید می‌زند.

۵- نتیجه‌گیری و کارهای آتی

در این مقاله، مدل‌های بافتاری را برای تولید عنوان بر روی دادگان علم‌نت به‌کار گرفتیم. مجموعاً نتایج به‌دست‌آمده گویای آن است که روش‌های خلاصه‌سازی مبتنی بر برت، بهبود چشم‌گیری نسبت به سایر



شکل ۶: نمودار امتیازات روش‌ها، (الف) نمودار امتیازات ROUGE-۱، (ب) نمودار امتیازات ROUGE-۲ و (ج) نمودار امتیازات ROUGE-L.

روش‌های پایه یادگیری عمیق دارند. استفاده از مدل برت در بسیاری از حوزه‌های کاربردی پردازش زبان طبیعی، علی‌الخصوص زبان انگلیسی منجر به بهبود نتایج گردیده و نتایج به‌دست‌آمده در این تحقیق نیز مؤیدی بر این ادعاست.

عدم وجود یک مدل برت خاص خلاصه‌سازی، نیاز به مدل‌هایی را که خاص یک حوزه پردازش زبان طبیعی باشد نشان می‌دهد و باید با اختصاص سخت‌افزار مناسب و جمع‌آوری منابع متنی کافی به ساخت این مدل‌ها اقدام نمود. از آنجا که طول متن ورودی و خلاصه‌های دادگان بر خلاصه‌سازی تأثیر دارد، زمانی که طول متن ورودی بسیار طولانی مثلاً یک کتاب باشد، استفاده از مدل‌های پیشنهادی نمی‌تواند موفق عمل کند. در این حالت کار بر روی تولید خلاصه طولانی از متن‌های طولانی‌تر با مبدل‌های پایه مانند LongFormer [۴۲] پیشنهاد داده می‌شود.

چنانچه مجموعه متون کافی در یک شاخه موضوعی خاص وجود داشته باشد، پیش‌آموزش مدل برت روی خلاصه‌سازی در یک موضوع مثلاً پزشکی سبب بهبود نتایج خواهد شد. یا تمام متن‌ها برای ساخت یک مدل برت خلاصه‌سازی عمومی می‌تواند مورد استفاده قرار بگیرد.

مراجع

[1] K. Woodsend, Y. Feng, and M. Lapata, "Title generation with quasi-synchronous grammar," in *Proc. EMNLP Conf. Empir. Methods Nat. Lang. Process*, pp. 513-523, Cambridge, MA, USA, 9-11 Oct. 2010.

[2] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," in Proc. of the Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing, EMNLP-IJCNLP'19, pp. 3728-3738, Hong Kong, China, 3-7 Nov. 2019.

[3] M. Farahani, M. Gharachorloo, and M. Manthouri, "Leveraging parsBERT and pretrained mT5 for persian abstractive text summarization," in *Proc. 26th Int. Comput. Conf. Comput. Soc. Iran, CSICC'21*, 6 pp., Tehran, Iran, 3-4 Mar. 2021.

[4] M. E. Shenassa and B. Minaei-Bidgoli, "ElmNet: a benchmark dataset for generating headlines from Persian papers," *Multimed. Tools Appl.*, vol. 81, no. 2, pp. 1853-1866, Jan. 2022.

[5] B. Dorr, D. Zajic, and R. Schwartz, "Hedge trimmer," in *Proc. of the HLT-NAACL 03 on Text Summarization Workshop*, vol. 5, 8 pp., Stroudsburg, PA, USA, 31-31 May 2003.

[6] L. Vanderwende, H. Suzuki, and C. Brockett, "Microsoft research at DUC2006: task-focused summarization with sentence simplification and lexical expansion," in *Proc. of Document Understanding Workshop, DUC'06*, pp. 70-77, New York, NY, USA, 6-8 Jun. 2006.

[7] J. M. Conroy, J. D. Schlesinger, D. P. O'leary, and J. Goldstein, "Back to basics: CLASSY 2006," in *Proc. of Document Understanding Workshop, DUC'06*, pp. 150-158, New York, NY, USA, 6-8 Jun. 2006.

[8] K. Knight and D. Marcu, "Summarization beyond sentence extraction: a probabilistic approach to sentence compression," *Artif. Intell.*, vol. 139, no. 1, pp. 91-107, Jul. 2002.

[9] M. Galley and K. McKeown, "Lexicalized markov grammars for sentence compression," in Proc. The Conf. of the North American Chapter of the Association for Computational Linguistics, Hlt-Naacl'07, pp. 180-187, 2007.

[10] J. Turner and E. Charniak, "Supervised and unsupervised learning for sentence compression," in *Proc. of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 290-297, Ann Arbor, MI, USA, Jun. 2005.

[11] E. Alfonseca, D. Pighin, and G. Garrido, "Heady: news headline abstraction through event pattern clustering," in *Proc. 51st Annu. Meet. Assoc. Comput. Linguist. Conf.*, *ACL'13,* vol. 1, pp. 1243-1253, Sofia, Bulgaria, 4-9 Aug. 2013.

[12] K. Filippova, E. Alfonseca, C. A. Colmenares, L. Kaiser, and O. Vinyals, "Sentence compression by deletion with LSTMs," in *Proc. Conf. on Empirical Methods in Natural Language Processing*, *EMNLP'15,* pp. 360-368, Lisbon, Portugal, 17-21 Sept.2015.

[13] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, "Sentence compression for aspect-based sentiment analysis," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2111-2124, Dec. 2015.

[14] Z. Wei, Y. Liu, C. Li, and W. Gao, "Using tweets to help sentence compression for news highlights generation," *Social Media Content Analysis: Natural Language Processing and Beyond*, vol. 3, pp. 309-320, Nov. 2017.

[15] M. Banko, V. O. Mittal, and M. J. Witbrock, "Headline generation based on statistical translation," in *Proc. of the 38th Annual Meeting on Association for Computational Linguistics*, pp. 318-325, Hong Kong, China, 3-6 Oct. 2000.

[16] R. Sun, Y. Zhang, M. Zhang, and D. Ji, "Event-driven headline generation," in Proc. ACL-IJCNLP 2015-53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int.Joint Conf. on Natural Language Processing of the Asian Federation of Natural Language Processing, pp. 462-472, Beijing, China, 26-31 Jul. 2015.

[17] S. Chopra, M. Auli, and A. M. Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in *Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics*, pp. 93-98, San Diego, CA, USA, 12-17 Jun. 2016.

[18] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence RNNs and Beyond," in *Proc. 20th SIGNLL Conf. Comput. Nat. Lang. Learn.*, pp. 280-290, Berlin, Germany, 11-12 Aug. 2016.

[19] A. See, P. J. Liu, and C. D. Manning, "Get to the point: summarization with pointer-generator networks," in *Proc. 55th Annu. Meet. Assoc. for Comput. Linguist.*, vol. 1, pp. 1073-1083, Vancouver, Canada, 30 Jul.-4 Aug. 2017.

[20] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *Proc. of the 34th Int. Conf. on Machine Learning, ICM'17L*, pp. 1243-1252, Sydney, Australia, 6-11 Aug. 2017.

[21] P. Kouris, G. Alexandridis, and A. Stafylopatis, "Abstractive text summarization based on deep learning and semantic content generalization," in *Proc. 57th Annual Meeting of the Association for*

Computational Linguistics, Conf., ACL'19, pp. 5082-5092, Florence, Italy, 28 Jul.-2 Aug. 2020.
[22] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush, "OpenNMT: open-source toolkit for neural machine translation," in *Proc. 55th Annual Meeting of the Association for Computational Linguistics, Proc. of System Demonstrations*, *ACL'17,* pp. 67-72, Vancouver, Canada, 30 Jul.-4 Aug. 2017.
[23] Y. Liu, et al., RoBERTa: A Robustly Optimized BERT Pretraining Approach, http://arxiv.org/abs/1907.11692, 2019.
[24] C. Raffel, *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, pp. 1-67, 2020.
[25] M. Lewis, *et al.*, "BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proc. of the Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 7871-7880, 5-10 Jul. 2020.
[26] Z. Yang, *et al.*, "XLNet: generalized autoregressive pretraining for language understanding," in *Proc. 33rd Conference on Neural Information Processing Systems*, pp. 5753-5763, Vancouver, Canada, 8-14 Dec. 2019.
[27] K. Song, B. Wang, Z. Feng, L. Ren, and F. Liu, "Controlling the amount of verbatim copying in abstractive summarization," in *Proc. 34th AAAI Conf. on Artificial Intelligence*, *AAAI'20,* pp. 8902-8909, New York, NY, USA, 7-12 Feb. 2020.
[28] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. on Learning Representations, ICLR'15.*, 15 pp., San Diego, CA, USA, 7-9 May 2015.
[29] J. Pennington, R. Socher, and C. D. Manning, "GloVe: global vectors for word representation," in *Proc. of the Conf. on Empirical Methods in Natural Language Processing, EMNLP'14,* pp. 1532-1543, Doha, Qatar.25-29 Oct. 2014.
[30] A. Vaswani, et al., "Attention is all you need," in Proc. of the 31st Int. Conf. on Neural Information Processing Systems, NIPS'27, pp. 5999-6009, Long Beach, CA, USA, 4-9 Dec. 2017.
[31] HAZM, "Python library for digesting Persian text," Sobhe, https://github.com/sobhe/hazm%0Ahttps://github.com/sobhe/hazm, 2014.
[۳۲] ن. غنی و ن. ریاحی، "خلاصه‌سازی چکیده‌ای متون فارسی با رویکرد مبتنی بر گراف،" مجموعه مقالات *سیزدهمین کنفرانس بین‌المللی فناوری اطلاعات،کامپیوتر و مخابرات*، ۲۲ ص.، تفلیس،گرجستان، ۱۴ آبان ۱۴۰۰.
[33] M. Moradi, M. Dashti, and M. Samwald, "Summarization of biomedical articles using domain-specific word embeddings and graph ranking," *J. Biomed. Inform.*, vol. 107, Article ID: 103452. Jul. 2020.
[34] D. Anand and R. Wagh, "Effective deep learning approaches for summarization of legal texts," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2141-2150, May 2022.
[35] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *NAACL-HLT'19,* pp. 4171-4186, Minneapolis, MN, USA3-7 Jun. 2019.
[36] C. Y. Lin, " ROUGE: a package for automatic evaluation of summaries," in *Proc. of the Workshop on Text Summarization Branches Out,* pp. 74–81, Barcelona, Spain, 25–26 July 2004.
[37] T. A. Dang and N. T. T. Nguyen, "Abstractive text summarization using pointer-generator networks with pre-trained word embedding," in *Proc. ACM Int. Conf. Proc. Series*, pp. 473-478, Hanoi, Viet Nam4-6 Dec. 2019.
[38] D. Nam, J. Yasmin, and F. Zulkernine, "Effects of pre-trained word embeddings on text-based deception detection," in Proc. IEEE 18th Int. Conf. on Dependable, Autonomic and Secure Computing, IEEE 18th Int. Conf. on Pervasive Intelligence and Computing, IEEE 6th Int. Conf. on Cloud and Big Data Computing and IEEE 5th Int. Conf. on Cybe Conf on Cyber Science and Technology Congress, pp. 437-443, Calgary, Canada, 17-22 Aug. 2020.
[39] R. Weng, H. Yu, S. Huang, S. Cheng, and W. Luo, "Acquiring knowledge from pre-trained model to neural machine translation," in *Proc. 34th AAAI Conf. on Artificial Intelligence*, *AAAI'20,* pp. 9266-9273, New York, NY, USA, 7-12 Feb. 2020.
[40] S. Gehrmann, Y. Deng, and A. M. Rush, "Bottom-up abstractive summarization," in *Proc. of the Conf. on Empirical Methods in Natural Language Processing, EMNLP'18*, pp. 4098-4109, Brussels, Belgium, 31 Oct.-4 Nov. 2018.
[41] K. Ethayarajh, "How contextual are contextualized word representations? comparing the geometry of BERT, ELMO, and GPT-2 embeddings," in *Proc. Conf. on Empirical Methods in Natural Language Processing and 9th Inte. Joint Conf. on Natural Language Processing, EMNLP-IJCNLP'19,* pp. 55-65, Hong Kong, China, 3-7 Nov. 2019.
[42] I. Beltagy, M. E. Peters, and A. Cohan, *Longformer: The Long-Document Transformer*, [Online] Available: http://arxiv.org/abs/2004.05150, 2020.

محمدابراهیم شناسا مدرک کارشناسی را از دانشگاه علم و صنعت ایران و کارشناسی ارشد را از واحد علوم و تحقیقات در رشته مهندسی کامپیوتر دریافت نمود. وی در حال حاضر دانشجوی دکترای هوش مصنوعی و عضو هیئت علمی دانشگاه آزاد واحد تهران- شمال است. حوزه تحقیقاتی مورد علاقه وی پردازش زبان طبیعی و داده‌کاوی می‌باشد.

بهروز مینایی بیدگلی دکترای خود را از دانشکده مهندسی و علوم کامپیوتر دانشگاه میشیگان آمریکا دریافت کرد. او در حال حاضر دانشیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران است و هدایت دو گروه تحقیقاتی داده‌کاوی و بازی‌های رایانه‌ای را بر عهده دارد. حوزه‌های تحقیقاتی ایشان متن‌کاوی و پردازش زبان طبیعی‌است. اطلاعات بیشتر از ایشان در سایت http://minaei.iust.ac.ir قابل دسترس‌است.