دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

دانشکده مهندسی کامپیوتر

پروژه تحقیقاتی درس سمینار

روش‌های خلاصه‌سازی انتزاعی

نگارش

زهرا زنجانی

استاد درس

دکتر رضا صفابخش

۱۴۰۲- ۱۴۰۱


بسم الله الرحمن الرحيم

چکیده

خلاصه‌سازی نقش مهمی در علم اطلاعات و بازیابی دارد، زیرا ارتباط نزدیکی با فشرده‌سازی داده‌ها و درک اطلاعات دارد. توانایی تولید خلاصه‌های مناسب می‌تواند موجب بهبود کارآمدی سیستم‌های استخراج اطلاعات و صرفه جویی در وقت انسان‌ها شود. خلاصه‌سازی خودکار به عنوان یک کار برجسته در پردازش زبان طبیعی [1] ظاهر شده است. با این حال، علیرغم اهمیت آن، چالش‌های خلاصه‌سازی خودکار تا حد زیادی حل نشده باقی مانده است. این گزارش مروری جامع از وضعیت فعلی خلاصه‌سازی خودکار ارائه می‌کند و رویکردها، تکنیک ها و معیارهای ارزیابی مختلف به کار گرفته شده در این زمینه را بررسی می‌کند.

واژه‌های کلیدی:

خلاصه‌سازی متن، پردازش زبان طبیعی، یادگیری عمیق،یادگیری تقویتی

[1] natural language processing (NLP)

فهرست مطالب

عنوان صفحه

فهرست اشکال

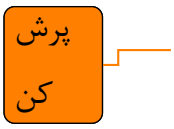
شکل صفحه

فهرست جداول

جدول صفحه

فهرست علائم و اختصارات

عنوان اختصاری	عنوان کامل



فصل اول

مقدمه

با رشد روزافزون اینترنت، محتوای متنی در اینترنت (به عنوان مثال وب سایت‌ها، نظرات کاربران، اخبار، وبلاگ‌ها، شبکه‌های رسانه‌های اجتماعی و غیره)به صورت تصاعدی افزایش می‌یابد. در نتیجه، کاربران زمان زیادی را صرف یافتن اطلاعات مورد نظر خود می کنند و حتی نمی توانند تمام محتوای متنی نتایج جستجو را بخوانند و درک کنند. خلاصه‌سازی خودکار اسناد می‌تواند به شناسایی مهم‌ترین اطلاعات، ارائه خلاصه‌ای جامع و صرفه‌جویی در وقت خوانندگان کمک کند. خلاصه‌سازی خودکار متن فرآیند تولید یک متن کوتاه است که بخش‌های اصلی یک سند طولانی‌تر را پوشش می‌دهد. یک خلاصه خوب جنبه‌های مهمی مانند خوانایی، انسجام، نحو، غیر زائد بودن، ترتیب جملات، مختصر بودن، تنوع اطلاعات و پوشش اطلاعات را در نظر می‌گیرد [۵] .

در سال‌های گذشته تلاش‌های زیادی برای تولید خلاصه‌سازی خودکار قابل قبول و خوانا صورت گرفته است. پژوهش‌های مرتبط با عمل خلاصه‌سازی خودکار متن در دهه ۵۰ میلادی شکل گرفتند. در یکی از این پژوهش‌ها لوهن و همکاران روشی برای خلاصه‌سازی اسناد علمی ارائه دادند که در آن تابعی بر اساس فرکانس تکرار کلمات یا عبارات به عنوان ویژگی تعریف می‌شود و با یادگیری وزن‌های مرتبط با این ویژگی ها خلاصه استخراج می‌شود [۱۶]. در کارهای تحقیقاتی اولیه، مدل‌های غیرعصبی مبتنی بر ساختار برای تولید خلاصه‌سازی خودکار مورد استفاده قرار گرفتند. با شروع دوره‌ی شبکه‌های عصبی عمیق پژوهش‌ها بر روی خلاصه‌سازی بیشتر شد. رویکردهای نوین خلاصه‌سازی شامل شبکه‌های عصبی عمیق دنباله به دنباله [۱] ، روش‌های بر پایه‌ی مدل تبدیل‌کننده [۲] و مدل‌های زبانی از پیش آموزش دیده [۳] می‌باشد. همچنین برخی از پژوهش های اخیر نشان داده‌اند، استفاده از رویکردهای مبتنی بر یادگیری تقویتی [۴] می‌تواند موجب بهبود معیارهای مختلف، از جمله امتیازات روژ، کیفیت کلی، خوانایی، انسجام، نحو، غیر افزونگی، ترتیب جملات، مختصر بودن، تنوع اطلاعات، پوشش اطلاعات شود.

[1] Deep neural sequence to sequence models

[2] transformer

[3] Pretrained language models (PTLMs)

[4] reinforcement learning (RL)

فصل دوم

روش‌های مبتنی بر ساختار

روش‌های خلاصه‌سازی مبتنی بر ساختار شامل رویکردهایی است که از ویژگی‌های ساختاری متن ورودی برای تولید خلاصه‌های مختصر و منسجم استفاده می‌کنند. در این رویکرد اطلاعات مهم متن به یک ساختار از پیش تعریف شده داده می‌شود و خلاصه با توجه به ساختار ایجاد می‌شود. در این فصل روش‌های مبتنی بر درخت [1]، مبتنی بر قالب[2]، مبتنی بر هستان‌شناسی[3]، عبارت مقدمه و بدنه[4]، مبتنی بر گراف[5] و مبتنی بر قانون[6] مورد بررسی قرار می‌گیرد.

۲-۱ روش مبتنی بر درخت

روش مبتنی بر درخت در خلاصه‌سازی متن شامل استفاده از درخت‌های وابستگی برای نمایش سند متنی است. متن مبدأ ابتدا به درخت‌های وابستگی تبدیل می‌شود، سپس این درخت‌ها در یک درخت واحد ادغام می‌شوند. در نهایت درخت وابستگی ادغام شده به جمله تبدیل می‌شود. فرآیند تبدیل درخت وابستگی به رشته‌ی کلمات را خطی سازی درخت می‌گویند. عملکرد این روش به انتخاب تجزیه‌کننده و حفظ وابستگی بین کلمات بستگی دارد و این باعث محدود شدن کارایی می‌شود. تکنیک‌های پیشنهاد شده شامل استفاده از تجزیه‌کننده‌های کم عمق برای ترکیب جملات مشابه، فشرده‌سازی با استفاده از حذف زیردرخت‌ها، تولید درخت‌های تودرتو با استفاده از ساختارهای بلاغی و تجزیه وابستگی است.[۲].

۲-۲ روش مبتنی بر قالب

روش‌های مبتنی بر الگو در خلاصه‌سازی متن شامل استفاده از قالب‌های از پیش تعریف شده برای نمایش سند است. این قالب‌ها برای مطابقت با الگوها و قوانین خاص در محتوای متنی طراحی شده‌اند و امکان استخراج اطلاعات مرتبط را فراهم می‌کنند که می‌توان آن‌ها را در فضای قالب ترسیم کرد. این فرآیند شامل تطبیق متن با این الگوها و قوانین برای شناسایی محتوای متناسب با الگو است. این روش بسیار منسجم است زیرا خلاصه‌هایی تولید می‌کند که به ساختار و قالب قالب‌ها پایبند هستند. یکی از چالش‌های روش مبتنی بر الگو، نیاز به تجزیه و تحلیل معنایی دقیق است، زیرا قالب‌ها نیاز به محتوای خاص و مرتبط برای پر شدن دارند [۲].

[1] tree-based
[2] template-based
[3] ontology-based
[4] lead-and-body phrase
[5] graph-based
[6] rule-based

۲-۳ روش مبتنی بر هستان شناسی

روش مبتنی بر هستی‌شناسی در خلاصه‌سازی متن شامل استفاده از پایگاه دانش یا هستی‌شناسی برای بهبود فرآیند خلاصه‌سازی است. بسیاری از اسناد موجود در اینترنت به حوزه‌های خاصی با واژگان محدود مرتبط هستند که می‌توانند توسط هستی‌شناسی بهتر نمایش داده شوند. هستی‌شناسی نامگذاری و تعریف رسمی‌انواع موجودیت مربوط به یک دامنه خاص را ارائه می‌دهد که به عنوان پایگاه دانش عمل می‌کند. با استفاده از هستی‌شناسی، سیستم خلاصه‌سازی می‌تواند نمایش معنایی محتوای اطلاعات را بهبود بخشد. تکنیک‌های این روش شامل استفاده از هستی‌شناسی برای ساخت یک مدل معنایی، نگاشت جملات به گره‌های هستی‌شناسی، و محاسبه امتیاز مربوط به موجودیت برای رتبه بندی جملات است. به طور کلی، روش مبتنی بر هستی‌شناسی از دانش خاص دامنه برای ایجاد خلاصه‌های دقیق‌تر و آموزنده‌تر استفاده می‌کند [۲]. لی و همکاران یک سیستم فازی را ارائه کرد که از هستی‌شناسی طراحی شده توسط متخصص حوزه اخبار استفاده می‌کند. جملات بر اساس طبقه بندی کننده اصطلاحی مبتنی بر هستی شناسی طبقه بندی می‌شوند. مکانیزم استنتاج فازی درجه عضویت برای هر جمله را با توجه به طبقه‌بندی کننده محاسبه می‌کند [۱۲].

۲-۴ روش عبارت مقدمه و بدنه

روش عبارت مقدمه و بدنه یک رویکرد خلاصه‌سازی متن است که بر شناسایی و بازنگری جملات اصلی، معروف به جملات کلیدی، در یک سند تمرکز دارد. جملات کلیدی معمولا حاوی اطلاعات مفید هستند و خلاصه خوبی از محتوا ارائه می‌دهند. این روش شامل درج و جایگزینی عبارات در جمله اصلی برای ایجاد تکرار مناسب با بازبینی‌های معنی‌دار است. محدودیت‌های این روش شامل عدم وجود مدل تعمیم یافته برای خلاصه‌سازی و تاثیر منفی مدل تجزیه دستوری است. [۲]. ایشیکاوا و همکاران روش خلاصه‌سازی ترکیبی مبتنی بر روش فرکانس عبارت [۷] و عبارت مقدمه و بدنه پیشنهاد کردند. تابع توزیع زاویه‌ای ضربدر بسامد عبارت میزان اهمیت هر جمله را مشخص می‌کند. دستورها براساس اهمیت برای نوشتن خلاصه رتبه بندی می‌شوند [۹].

۲-۵ روش مبتنی بر گراف

یکی دیگر از رویکردهای خلاصه‌سازی روش مبتنی بر نمودار است که هر جمله در یک سند را به عنوان یک راس در یک نمودار نشان می‌دهد. جملات بر اساس روابط معنایی با یال‌ها به هم متصل می‌شوند و وزن یال‌ها نشان دهنده قدرت رابطه است. سپس از یک الگوریتم رتبه بندی نمودار برای تعیین اهمیت هر جمله استفاده می‌شود. جملات با اهمیت بالاتر در خلاصه گنجانده می‌شوند. این روش نیازی به

[7] Term frequency (TF)

دانش عمیق زبانی یا حوزه‌ای ندارد و می‌تواند با انتخاب جملاتی با اهمیت بالا، خلاصه‌های مختصر و منسجم ایجاد کند [۲]. مالیروس و اسکیانیس از مرکزیت گره برای نشان دادن اهمیت یک اصطلاح در سند استفاده می‌کنند. مرکزیت‌های گره محلی و جهانی برای وزن‌دهی عبارت در نظر گرفته می‌شوند تا خلاصه را شکل دهند [۱۸].

۲-۶ روش مبتنی بر قانون

در روش خلاصه سازی مبتنی بر قاعده، اسناد را به دسته بندی ها و جنبه ها تقسیم می کنیم. سپس، ماژولی به نام ماژول انتخاب محتوا، بهترین اطلاعات را براساس قوانین از پیش تعریف شده انتخاب می کند تا به جنبه های هر دسته پاسخ دهد. در نهایت، ما از الگوهای تولید برای ایجاد جملات خلاصه مختصر استفاده می کنیم. بنابراین، اساسا، ما قوانینی داریم که به ما کمک می کنند تا مهم ترین اطلاعات را برای هر جنبه انتخاب کنیم، و سپس از آن قوانین برای تولید یک خلاصه استفاده می کنیم. [۱۹].

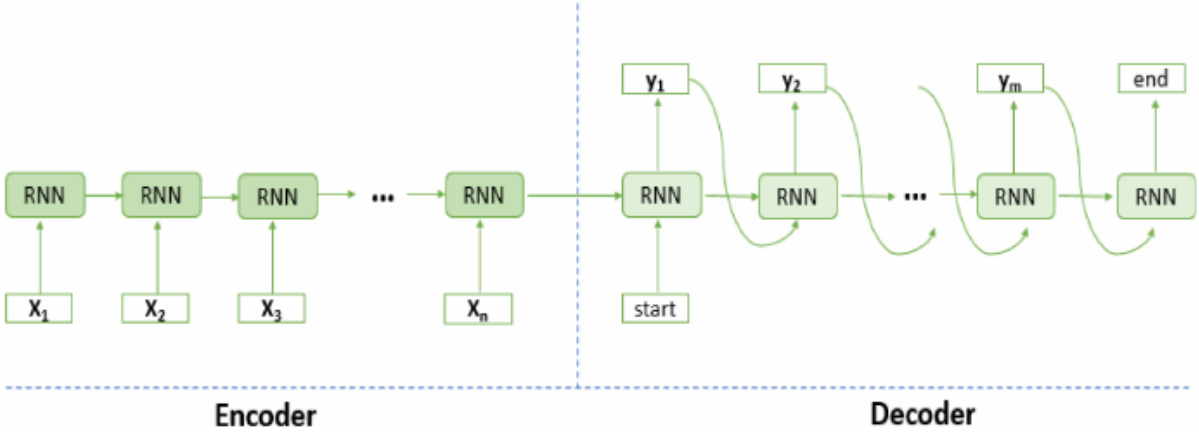
فصل سوم

روش‌های مبتنی بر شبکه‌ی عصبی

مقدمه؟

۳-۱ روش‌های مبتنی بر مدل کدگذار-کدگشا

قبل از ظهور ترنسفورمرها ، مدل‌های شبکه عصبی عمیق دنباله به دنباله بهترین مدل‌ها برای وظایف تولید متن از جمله ترجمه‌ی ماشینی و خلاصه‌سازی متن بوده‌اند. این مدل‌ها ورودی را از یک فرم به فرم دیگر نگاشت می‌کنند تا نتایج مورد نظر را تولید کنند. معماری کدگذار-کدگشا رویکرد اصلی برای مدل‌سازی مدل‌های دنباله به دنباله است. شکل ۳-۱ معماری پایه‌ی مدل کدگذار-کدگشا را شرح می‌دهد. شبکه‌های بازگشتی [۱] [۶] و حافظه‌های کوتاه مدت طولانی [۸] برای توالی طراحی شده‌اند و مناسب‌ترین معماری‌های یادگیری عمیق برای کدگذاری و پردازش داده‌های دنباله‌ای مانند متن هستند. اما این شبکه‌ها در مدیریت حافظه‌ی بلند مدت طولانی [۲] مشکل دارند. یائو [۳] و همکاران مدل کدگذاری

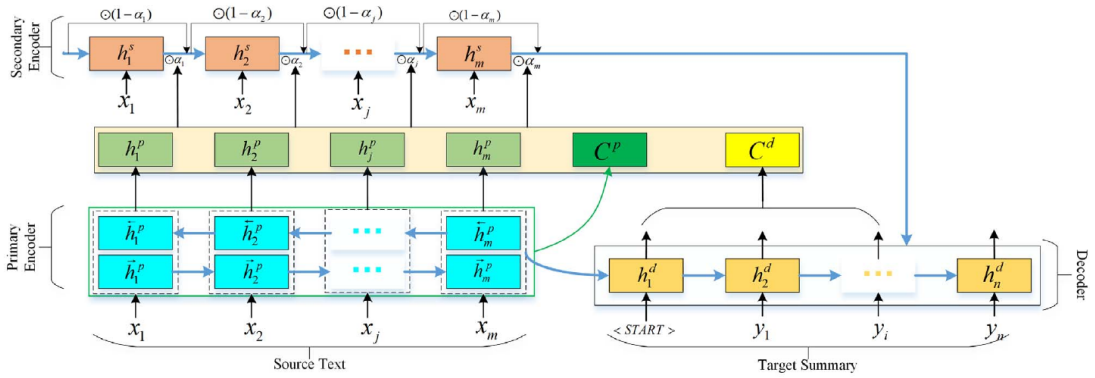


شکل ۳-۱: معماری پایه‌ی مدل کدگذار-کدگشا[۱]

دوگانه را برای خلاصه‌سازی انتزاعی پیشنهاد داده‌اند. کدگذاری دوگانه به مدل اجازه می دهد تا دو نمایش متفاوت از متن را بیاموزد: نمایش متن ورودی و نمایش خلاصه مرجع. این به مدل اجازه می‌دهد تا روابط بین متن ورودی و خلاصه مرجع را بهتر درک کند، که می‌تواند منجر به خلاصه‌های دقیق‌تر و آموزنده‌تر شود. این مدل از یک کدگذار اولیه، یک کدگذار ثانویه و یک کدگشا مجهز به مکانیزم توجه تشکیل شده است و هر سه ماژول فوق از واحد بازگشتی دروازه‌ای [۴] استفاده می‌کنند. کدگذار اولیه بردارهای معنایی هر کلمه را در ترتیب ورودی محاسبه می کند. و کدگذار ثانویه ابتدا وزن اهمیت هر کلمه را در ترتیب ورودی محاسبه می کند و سپس بردارهای معنایی مربوطه را دوباره محاسبه می

[1] recurrent neural network (RNN)
[2] long short-term memory networks(LSTM)
[3] Yao
[4] gated recurrent unit (GRU)

کند.در نهایت کدگشا با مکانیسم توجه به صورت مرحله ای کدگشایی می کند و در هر مرحله یک توالی خروجی با طول ثابت جزئی ایجاد می کند(شکل ۳-۲). در این مدل کدگذار ثانویه یک عملیات کدگذاری را براساس ورودی هر مرحله و خروجی مرحله‌ی قبل انجام می دهد. بنابراین کیفیت متون قبلی تولید شده توسط کدگشا بر خروجی های جدید تاثیر می‌گذارد [۲۷].



شکل ۳-۲: معماری پایه‌ی مدل دوگانه‌ی کدگذار [۲۷]

مزایای مدل کدگذاری دوگانه به شرح زیر می‌باشد.

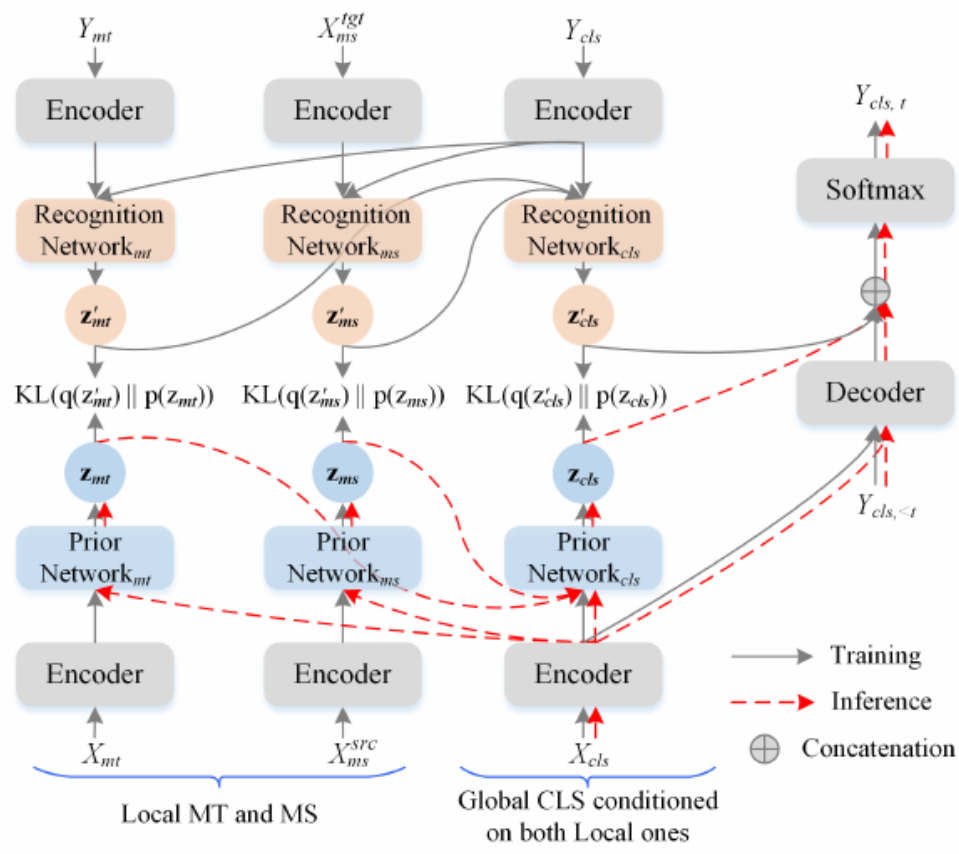
- کدگذاری دوگانه به مدل اجازه می‌دهد دو نمایش متفاوت از متن را یاد بگیرد.
- کدگذاری دوگانه باعث می‌شود تا مدل روابط بین متن ورودی و خلاصه مرجع را بهتر درک کند.
- مدل ارایه شده خلاصه های دقیق و حاوی اطلاعات مفید تولید می‌کند.

محدودیت های مدل ارائه شده به شرح زیر می‌باشد.

- این مدل نسبت به سایر رویکردهای خلاصه‌سازی انتزاعی پیچیده‌تر است.
- آموزش و توسعه‌ی مدل دشوارتر است.
- مدل به حجم زیادی از داده های آموزشی نیاز دارد.

یکی از مدل‌های ارایه شده براساس مدل کدگذاری دوگانه مدل سلسله مراتبی متغیر برای خلاصه‌سازی متقاطع زبانی [۵] می باشد. مدل پیشنهادی شامل دو متغیر نهفته محلی، یکی برای ترجمه و دیگری برای خلاصه سازی، و یک متغیر نهفته جهانی برای خلاصه سازی بین زبانی است. متغیرهای نهفته محلی به ترتیب برای بازسازی ترجمه و خلاصه زبان مبدأ محدود می شوند. سپس از متغیر نهفته سراسری برای تولید خلاصه بین زبانی استفاده می شود. قسمت کد گذار دو بخش دارد که هر بخش وظیفه‌ی تولید یکی از متغیرهای نهفته محلی را دارد. بخش کدگشا با استفاده از نمایش‌های نهفته‌ی محلی خلاصه‌ی نهایی را تولید می‌کند. ساختار سلسله مراتبی مدل به آن اجازه می دهد تا رابطه سلسله مراتبی بین ترجمه، خلاصه سازی و خلاصه‌سازی بین زبانی را بیاموزد[۱۴].

[5] cross-lingual



شکل ۳-۳: معماری پایه‌ی مدل سلسله مراتبی متغیر برای خلاصه‌سازی متقابل زبانی[۱۴]

متغیرهای محلی z_mt و z_ms به ترتیب برای ترجمه و خلاصه سازی طراحی شده اند. سپس z_cls جهانی برای خلاصه سازی بین زبانی است، خطوط خاکستری ن نشان‌دهنده فرآیند آموزشی است که مسئول تولید (z'_{mt}، z'_{ms}، z'_{cls}) از توزیع پسین متناظر پیش‌بینی‌شده توسط شبکه‌های شناسایی است که یادگیری شبکه‌های قبلی را هدایت می‌کند.خطوط قرمز چین نشان دهنده فرآیند استنتاج برای تولید نمایش‌های نهفته (z_{mt}، z_{ms}، z_{cls}) از توزیع های قبلی مربوطه پیش بینی شده توسط شبکه های قبلی است.

۳-۲ روش های مبتنی بر مدل ترنسفورمر ها

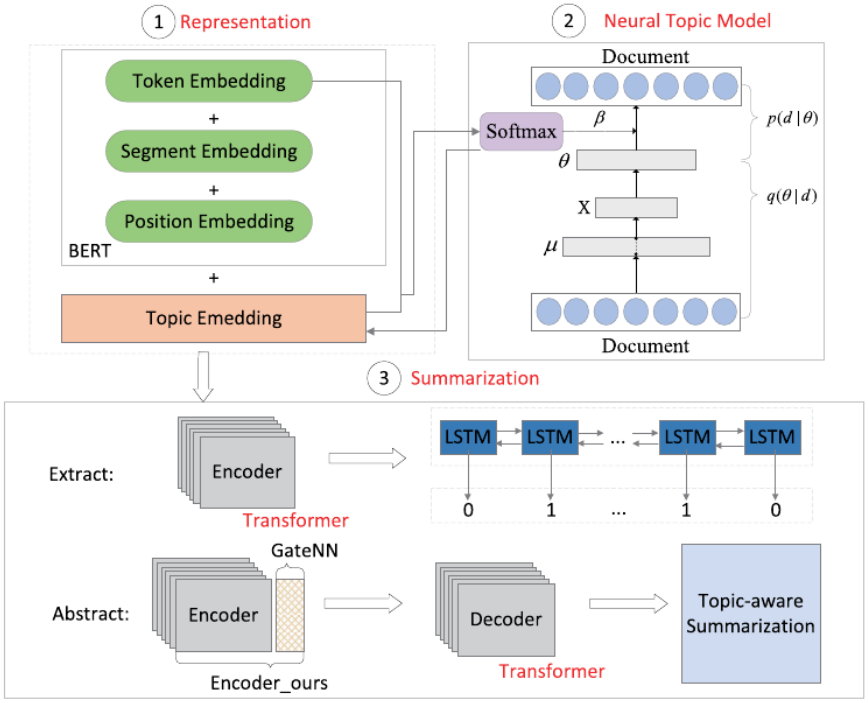
با ظهور ترنسفورمرها [۶] ، بهبودهای قابل توجهی در کیفیت نتایج خلاصه‌سازی خودکار به وجود آمد. ترنسفومرها با استفاده از مکانیزم توجه به خود [۷] شباهت بین ورودی‌ها را بدون توجه به موقعیت موازی آن‌ها با حضور مستقل هر توکن در توالی ورودی مدل می‌کنند و به طور مؤثر مشکلات شبکه‌های بازگشتی را حل می‌کنند [۲۴] . یکی از جهت‌گیری‌های رایج پژوهشی، اصلاح یا تطبیق ترنسفورمرها و مدل‌های زبانی از پیش آموزش دیده با وظایف مختلف مانند خلاصه‌سازی است. مدل‌های مبتنی بر مدل‌های زبانی از پیش آموزش دیده که با هدف خلاصه‌سازی انتزاعی طراحی شده‌اند از ویژگی‌های معنایی و متنی غنی

[6] transformers

[7] Self-attention

بازنمایی‌های زبان برای بهبود کیفیت و دقت خلاصه‌ها استفاده می‌کنند.

پان و همکاران رویکرد جدید بر اساس مدل برت را برای خلاصه سازی متن پیشنهاد کرده‌اند. نویسندگان استدلال می‌کنند که مدل‌های خلاصه‌سازی متن موجود، موضوع متن را در نظر نمی‌گیرند، که می‌تواند منجر به خلاصه‌هایی شود که آموزنده یا مرتبط نیستند. مدل ارائه شده که تی‌برت‌سام [۸] نامیده می‌شود از سه بخش ایجاد بازنمایی، مدل موضوعی عصبی [۹] و مدل خلاصه‌سازی تشکیل شده است. ساختار مدل را در شکل ۳-۴ نشان داده شده است.



شکل ۳-۴: معماری مدل تی‌برت‌سام [۱۷]

همانطور که در شکل ۳-۵ نشان داده شده است، بازنمایی ایجاد شده برای هر جمله ورودی، با استفاده از یک شبکه‌ی ترنسفورمر دوسویه‌ی [۱۰] چند لایه و حاصل جمع چهار نوع تعبیه (تعبیه نشانه [۱۱] ، تعبیه قطعه [۱۲] ، تعبیه موقعیت و تعبیه موضوع) به دست می‌آید که تعبیه موضوع در این مقاله معرفی شده و سه تعبیه دیگر مشابه مدل برت هستند. وجود تعبیه موضوع در تولید بازنمایی هر کلمه یا جمله موجب افزودن اطلاعات پیش زمینه‌ای به هر کلمه و حل مشکل چند معنایی می‌شود.

مدل موضوعی عصبی وظیفه‌ی ایجاد تعبیه موضوعی را دارد. این مدل دارای دو جزء است: یک شبکه

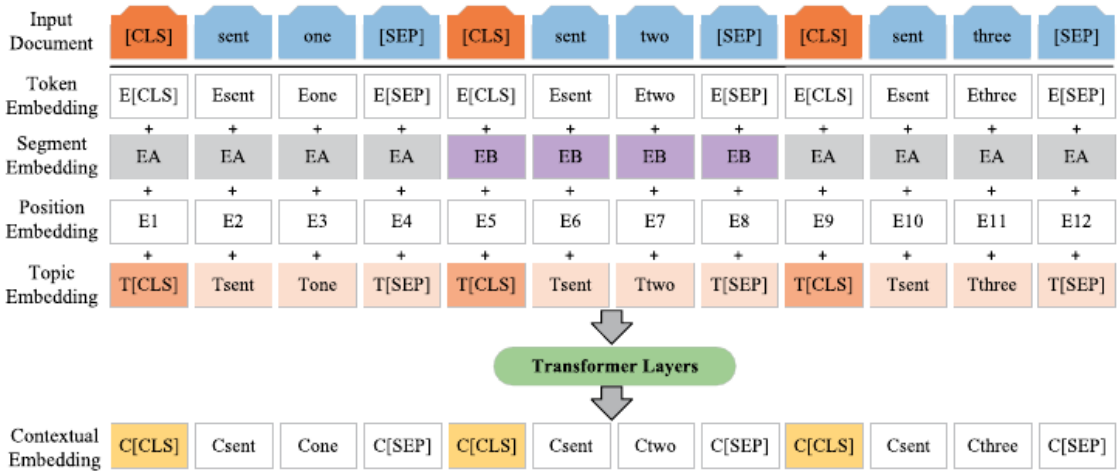
[8] T-BERTSum

[9] Neural Topic Model(NTM)

[10] bidirectional

[11] token embedding

[12] segment embedding,



شکل ۳-۵: تعبیه مدل تی‌برت‌سام [۱۷]

مولد و یک شبکه استنتاج. شبکه مولد یک سند را به عنوان ورودی می گیرد و یک توزیع موضوعی را بر روی کلمات موجود در سند خروجی می دهد. شبکه استنتاج یک سند را به عنوان ورودی می گیرد و خروجی آن پارامترهای توزیع موضوع است. مدل خلاصه سازی مبتنی بر معماری کدگذار - کدگشای ترنسفورمر است. کدگذار بازنمایی ایجاد شده را به عنوان ورودی می‌گیرد و دنباله‌ای از حالت‌های پنهان را تولید می‌کند. سپس کدگشا با استفاده از این حالت های پنهان و متن خلاصه را خروجی می‌کند. همانطور که در شکل ۳-۶ نشان داده شده است، به منظور فیلتر کردن اطلاعات کلیدی توالی ورودی، شبکه دروازه ای قبل از کدگشا اضافه می شود. این شبکه برای کنترل جریان اطلاعات از دنباله ورودی به دنباله خروجی افزوده شده است و باعث می‌شود کدگشا بر روی تولید خلاصه از اطلاعات کلیدی و حذف اطلاعات غیرضروری تمرکز کند. این مدل می تواند خلاصه‌هایی تولید کند که هم آموزنده و هم مرتبط با موضوع متن باشد و قابلیت تطبیق با وظایف و حوزه های مختلف را دارد.

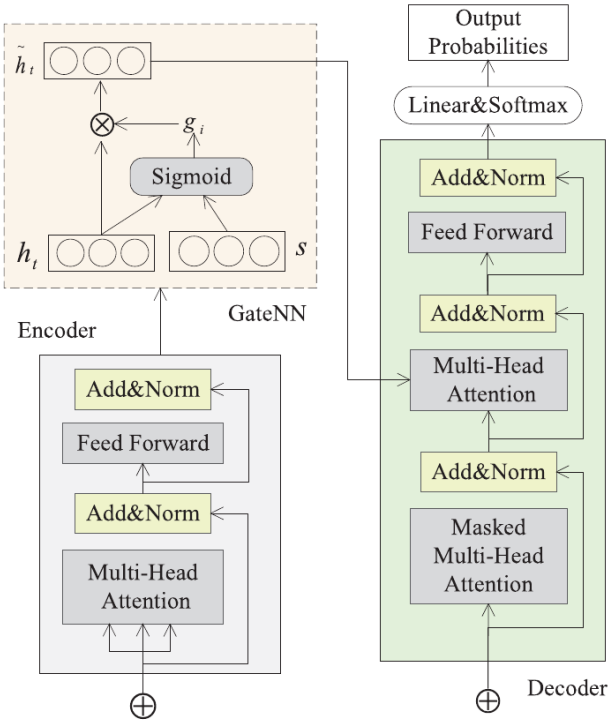
ادیت

بیشتر مدل‌های خلاصه‌سازی انتزاعی موجود برای تولید خلاصه‌های با طول ثابت به جای خلاصه‌های با طول متغیر طراحی شده‌اند، بنابراین سو [۱۳] و همکاران یک مدل مبتنی بر ترنسفورمر ارائه دادند که خلاصه‌های انتزاعی با طول متغیر را با توجه به تقاضای کاربر تولید کند.

مدل خلاصه‌سازی دو مرحله‌ای در رویکرد پیشنهادی با تقسیم متن ورودی به بخش‌ها و تولید خلاصه‌هایی برای هر بخش، به خلاصه‌سازی انتزاعی با طول متغیر دست می‌یابد. در اینجا نحوه کار آن آمده است:

بخش بندی متن: این مرحله متن ورودی را به تعداد قسمت های از پیش تعیین شده تقسیم می کند. تعداد بخش ها را می توان توسط کاربر مشخص کرد یا با توجه به نسبت دلخواه طول ورودی تنظیم کرد. مدل تقسیم‌بندی متن از مدل $BERT - biLSTM$ برای شناسایی مرزهای بین بخش‌ها استفاده

[13] Ming-Hsiang Su



شکل ۳-۶: معماری ترنسفورمر تی‌برت‌سام [۱۷]

این مدل شامل شبکه‌ی دروازه‌ای و کدگذار-کدگشا با توجه چند سر می‌باشد[۱۷]

می‌کند. این مرحله تضمین می کند که مرحله خلاصه سازی انتزاعی بر روی بخش های منسجم متن انجام می شود. این به بهبود کیفیت خلاصه های تولید شده کمک می کند.. هدف یافتن نقاط تقسیم بندی است که نشان دهنده تغییر موضوع در متن است.

خلاصه سازی استخراجی: پس از تقسیم بندی متن، یک مدل استخراجی بر اساس مدل خلاصه سازی مبتنی بر برت‌سام [۱۴] ساخته می شود. این مدل مهم‌ترین جمله را از هر بخش استخراج می‌کند. این جملات استخراج شده به عنوان ورودی برای مرحله دوم مدل خلاصه سازی عمل می‌کنند.

خلاصه‌سازی اسناد: در مرحله دوم از جملات استخراج شده برای آموزش ماژول خلاصه سازی اسناد استفاده می شود. این ماژول یک خلاصه سرفصل از کل ورودی متن ایجاد می کند. پارامترهای این ماژول با در نظر گرفتن امتیازات ضرر ماژول خلاصه سازی اسناد و ماژول خلاصه سازی بخش به روز می شود.

خلاصه‌سازی بخش‌ها: بخش های به دست آمده از مرحله تقسیم بندی متن برای آموزش ماژول خلاصه سازی در مرحله اول استفاده می شود. این ماژول یک خلاصه بر اساس جمله برای هر بخش تولید می کند. امتیازات ضرر ماژول خلاصه سازی سند و ماژول خلاصه سازی بخش برای به روز رسانی پارامترهای ماژول خلاصه سازی بخش در نظر گرفته می شود.

[14] BertSum

آموزش مشارکتی: آموزش مشارکتی برای آموزش متناوب ماژول خلاصه سازی بخش و ماژول خلاصه سازی اسناد تا زمان همگرایی اعمال می‌شود. این فرآیند به بهینه سازی عملکرد هر دو ماژول کمک می کند.

خلاصه‌سازی با طول متغیر: در طول آزمایش، خروجی‌های ماژول خلاصه‌سازی بخش به هم متصل می‌شوند تا نتیجه خلاصه‌سازی انتزاعی با طول متغیر ارائه شود. تعداد بخش ها را می توان توسط کاربر مشخص کرد یا با توجه به نسبت دلخواه طول ورودی تنظیم کرد.

ادیتش کن

با ترکیب روش‌های استخراجی و انتزاعی در مدل خلاصه‌سازی دو مرحله‌ای، رویکرد پیشنهادی می‌تواند خلاصه‌های انتزاعی روان و با طول متغیر را با توجه به خواسته‌های کاربر تولید کند. این مدل می تواند خلاصه های انتزاعی با طول متغیر را با توجه به خواسته های کاربر ایجاد کند. این یک پیشرفت نسبت به مدل های قبلی است زیرا می تواند به طور همزمان به خلاصه سازی انتزاعی روان و با طول متغیر دست یابد.

لوئیس و همکاران مدلی با نام بارت [۱۵] ارائه دادند. این مدل مشابه با مدل اصلی تبدیل‌کننده، ساختاری کدگذار-کدگشا دارد. بر خلاف سادگی این مدل این مدل را می‌توان نسخه عمومی‌تری از برت و جی‌پی‌تی [۱۶] (به دلیل داشتن کدگذار دو طرفه و کدگشای چپ به راست) دانست. (۳-۷)این مدل در عملیات تولید متن، مانند ترجمه ماشینی یا خالصه‌سازی انتزاعی متن، و همچنین در فهم متن کاربرد دارد. بارت را می‌توان با استفاده از اهداف رمزگذاری خودکار حذف نویز آموزش داد. در ابتدا، توالی ورودی با استفاده از یک تابع نویز دلخواه خراب می‌شود. سپس ورودی خراب توسط یک شبکه ترنسفورمر بازسازی می‌شود. این مدل طیف گسترده ای از نویز ها از جمله پوشاندن توکن، حذف توکن، پر کردن متن، چرخش سند، به هم ریختن جمله (به هم زدن تصادفی ترتیب کلمه یک جمله) را ارزیابی می‌کند[۱۵، ۱۳].

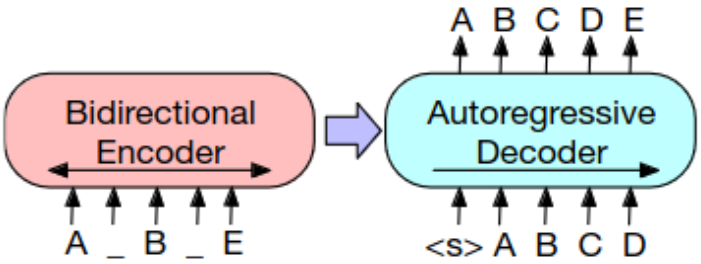
یک کم دیگه اضافه کن

با این که بارت دقت خلاصه‌سازی انتزاعی متن را بهبود بخشید، ولی عمل‌های تعریف شده در مرحله پیش‌آموزش آن، مختص خلاصه‌سازی انتزاعی متن نبودند، در نتیجه در سال ۲۰۲۰ مدلی تحت عنوان پگاسوس [۱۷] توسط ژنگ و همکاران ارائه شد که معماری مشابه با بارت داشت ولی پیش‌آموزش آن مختص خلاصه‌سازی انتزاعی متن بود. مدل پگاسوس یک مدل دنباله به دنباله کدگذار کدگشا مبتنی بر ترنسفورمر است که بر روی مجموعه‌های متنی بدون نظارت با هدف تولید جملات فاصله‌افتاده [۱۸] از قبل آموزش داده شده است [۲۹].

این مدل دو عمل پیش‌آموزش معرفی کرده است که در ادامه به شرح آنها می‌پردازیم:

۱. تولید جملات فاصله‌افتاده : این فرض مطرح شده است که اگر عمل پیش‌آموزش مدل به عمل پایین‌دست [۱۹] نزدیک‌تر باشد، نتیجه نهایی بهتر و همچنین تنظیم دقیق پارامترها [۲۰] سریع‌تر

[15] BART
[16] GPT
[17] PEGASUS
[18] gap sentences generation
[19] downstream task
[20] fine-tuning



شکل ۳-۷: ساختار مدل بارت [۱۳]

ورودی‌های کدگذار نیازی به همسویی با خروجی‌های کدگشا ندارند، که امکان تبدیل نویز دلخواه را فراهم می‌کند. در اینجا، یک سند با جایگزین کردن دهانه‌های متن با نمادهای ماسک خراب شده است. سند خراب (سمت چپ) با یک مدل دو طرفه کدگذاری می‌شود و سپس احتمال سند اصلی (سمت راست) با کدگشای خودبازگشتی محاسبه می‌شود. برای تنظیم دقیق، یک سند خراب به رمزگذار و رمزگشا وارد می‌شود و ما از نمایش‌هایی از حالت پنهان نهایی کدگشا استفاده می‌کنیم[۱۳].

خواهد بود. با توجه به این که این مدل قرار است فقط برای خلاصه‌سازی انتزاعی متن استفاده شود، عمل پیش‌آموزش تولید متن‌های مشابه با خلاصه از یک سند ورودی تعریف شده‌است. بر اساس یک متغیر که درصد جملات پنهان‌شده را مشخص می‌شود، تعدادی از جملات انتخاب شده و هر جمله به طور کامل با توکن $[MASK1]$ جایگزین میشود. برای انتخاب این جملات، سه راه پیشنهاد شده است.

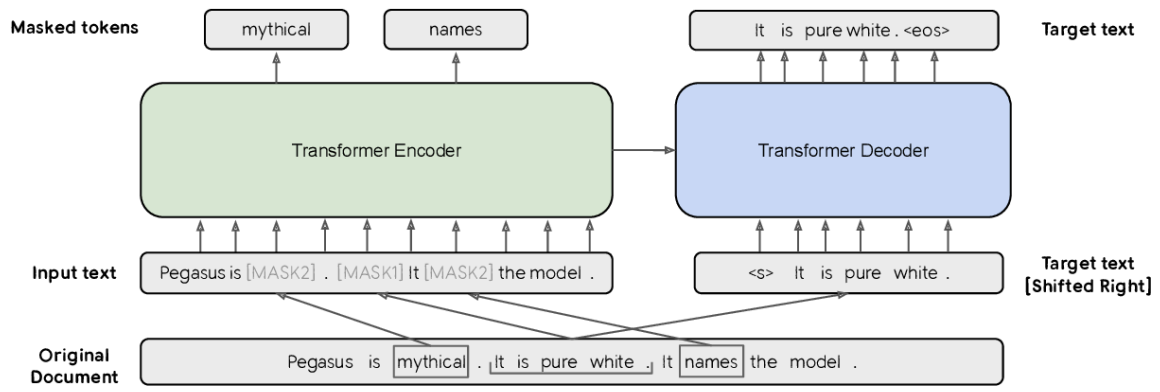
- انتخاب تصادفی: mجمله به صورت تصادفی از متن انتخاب شده و پنهان می‌شوند.
- انتخاب جملات اول متن: mجمله اول متن پنهان می‌شوند. دلیل این کار، فرض مهم‌تر بودن جملات ابتدای متن نسبت به جملات بعدی است.
- انتخاب جملات مهم متن: برای انتخاب m جمله مهم متن از تقریب معیار ارزیابی روژ-۱ استفاده می‌شود. به ازای هر جمله از متن، یک دوتایی از آن جمله و کل متن سند فاقد آن جمله ساخته شده و ارزیابی می‌شود که چقدر ممکن است این جمله، خلاصه کل سند فاقد آن جمله باشد. جملاتی که امتیاز بالاتر گرفته‌اند از نظر خلاصه بودن مهم‌تر هستند و پنهان می‌شوند.

مدل زبانی پوشیده شده: مشابه مدل برت ۵۱ درصد از توکن‌های متن ورودی انتخاب می‌شوند و سپس ۸۰ درصد از این توکن‌ها، با توکن $[MASK2]$ و۱۰ درصد توکن ها با یک توکن تصادفی جایگزین می‌شوند. ۱۰ درصد دیگر بدون تغییر باقی می‌ماند. شکل ۳-۸ اعمال همزمان این دو عمل، یعنی تولید جمالت فاصله افتاده و مدل زبانی پوشیده شده را بر روی یک مثال نشان می‌دهد.

کدیا [۲۱] و همکاران الگوریتم حداکثر سازی نقطه-محصول فرا یادگیری (امدات) [۲۲] را پیشنهاد دادند.

[21] Kedia

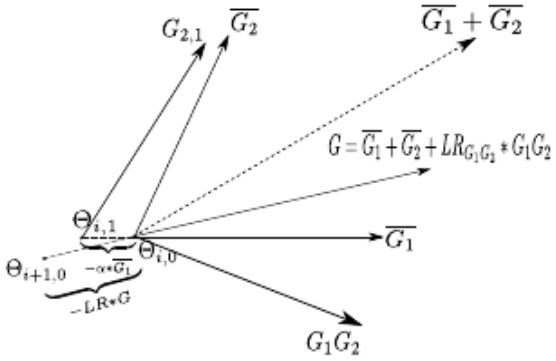
[22] Meta-Learned Dot-Product Maximization (MDot)



شکل ۳-۸: ساختار مدل پگاسوس [۲۹]

معماری پایه پگاسوس یک کدگذار-کدگشا ترنسفورمر استاندارد است. جملات فاصله‌افتاده و مدل زبانی پوشیده شده به طور همزمان در این مثال به عنوان اهداف پیش‌آموزش اعمال می شوند. در اصل سه جمله وجود دارد. یک جمله با $[MASK1]$ پوشانده شده و به عنوان متن تولید هدف جملات فاصله‌افتاده استفاده می‌شود. دو جمله دیگر در ورودی باقی می‌مانند و برخی از نشانه ها به طور تصادفی توسط $[MASK2]$پوشانده می‌شوند [۲۹].

این الگوریتم بر اساس ایده به حداکثر رساندن حاصل‌ضرب نقطه‌ای بین گرادیان های مدل در نقاط مختلف آموزش با استفاده از تکنیکی به نام تفاوت های محدود [۲۳] است. این الگوریتم از نظر محاسباتی کارآمد است و می تواند برای مدل های بزرگ مانند برت اعمال شود و سربار محاسباتی را کاهش بدهد [۲۳]. عملکرد مناسب مدل پگاسوس در خلاصه‌سازی متون باعث شده بهترین مدل خلاصه‌سازی متون کوتاه مبتنی بر مدل پگاسوس و تکنیک تنظیم [۲۴] امدات باشد.



شکل ۳-۹: الگوریتم امدات [۲۳]

محاسبه گرادیان برای به حداکثر رساندن محصول نقطه‌ای با استفاده از تقریب تفاضل محدود، و استفاده از آن برای تنظیم گرادیان استاندارد [۲۳].

[23] finite differences

[24] regularization

۳-۲-۱ ایده های ارایه شده بهبود خلاصه سازی متون طولانی

یکی از مشکلات مدل ترنسفومر در خلاصه‌سازی متون طولانی حافظه‌ی درجه دوم پیچیدگی‌های محاسباتی و تعداد زیاد عملیات می‌باشد. برای حل این مشکلات کارهای مختلفی انجام شده است. به عنوان مثال شبکه‌ی ریفورمر [۲۵] برای حل چالش های محاسباتی مرتبط با پردازش دنباله های طولانی متن ارائه شده است. لایه‌های برگشت‌پذیر [۲۶] معرفی شده در این مقاله امکان بازسازی ورودی از خروجی را در طول گذر به عقب، کاهش نیازهای حافظه و امکان پردازش کارآمد دنباله‌های طولانی را فراهم می‌کنند. . علاوه بر این، ریفورمر از تکه تکه کردن برای پردازش بخش های کوچک تر ورودی به طور مستقل استفاده می‌کند که موازی‌سازی را ممکن می‌کند و مصرف حافظه را کاهش می‌دهد. یکی از کمک های کلیدی آن استفاده از درهم‌سازی حساس به مکان در مکانیسم توجه [۲۷] است. درهم‌سازی حساس به مکان با توجه به زیرمجموعه ای از نشانه ها بر اساس مقادیر هش آنها، محاسبه توجه کامل را تقریب می زند، که منجر به محاسبه توجه کارآمدتر می‌شود. علاوه بر این، ریفورمر از کدگذاری های موقعیت محوری برای کدگذاری اطلاعات موقعیت توکن ها به صورت فشرده استفاده می‌کند. این تکنیک‌ها مجموعاً مدل ریفورمر را بسیار مقیاس‌پذیر و کارآمد در حافظه می‌سازد، و آن را قادر می‌سازد تا دنباله‌های طولانی متن را مدیریت کند و در عین حال عملکرد رقابتی را در وظایف مختلف پردازش زبان طبیعی حفظ کند [۱۱]

ادیت

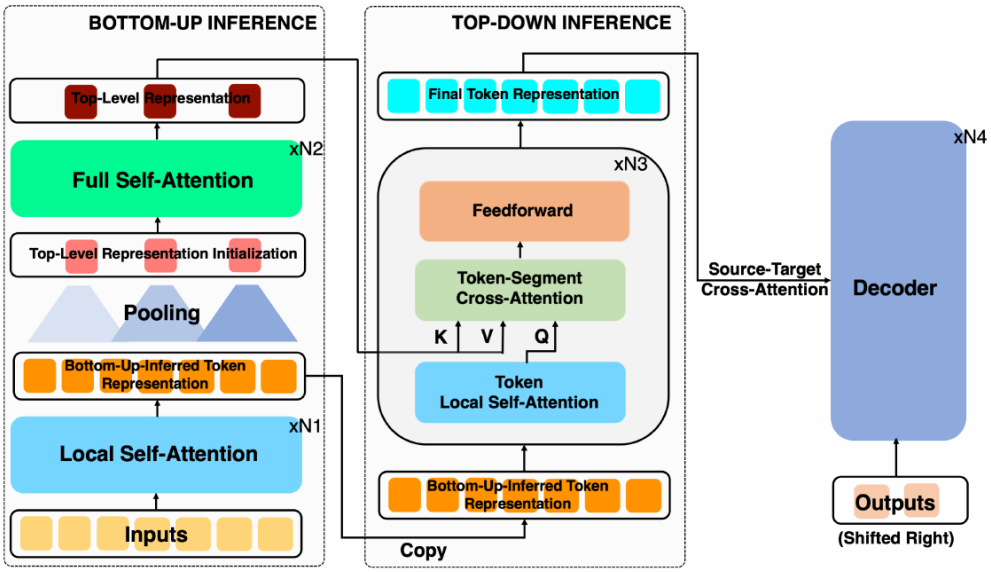
. همچنین شبکه‌ی ترنسفورمر پراکنده [۲۸] با معرفی فاکتورسازی ماتریس پراکنده‌ی توجه، زمان و حافظه مورد نیاز را به کاهش می‌دهد. با استفاده از پراکندگی، مدل می‌تواند تنها به زیرمجموعه‌ای از نشانه‌های ورودی توجه کند و روی مرتبط‌ترین اطلاعات تمرکز کند و بقیه را نادیده بگیرد. این رویکرد پیچیدگی محاسباتی را کاهش می‌دهد و مدل می‌تواند توالی‌های طولانی‌تر را مدیریت کند [۴] . مشابه شبکه‌ی ترنسفورمر پراکنده مدل بیگ‌برد نیز [۲۹] با استفاده از مکانیزم توجه پراکنده [۳۰] که وابستگی را به خطی کاهش می‌دهد و عملکرد ترنسفورمر را در مواجه با توالی [۳۱] طولانی بهبود می‌بخشد. مدل بیگ‌برد نوآوری‌های دیگری مانند توجه جهانی [۳۲] را معرفی می‌کند، که در آن توکن های خاص به تمام توکن های دیگر در دنباله توجه می کنند و وابستگی های دوربرد را به طور موثرتری به دست می آورند. همچنین شامل یک فرآیند پالایش تکراری است که وزن های توجه را برای بهبود عملکرد مدل اصلاح می کند [۲۸].

ادیت

در سال‌های اخیر مدل‌های مختلفی برای بهبود کیفیت خروجی مدل خلاصه‌سازی خودکار اسناد

[25] Reformer
[26] reversible layers
[27] locality-sensitive hashing (LSH)
[28] sparse
[29] Big Bird
[30] Sparse attention
[31] sequence
[32] global attention

بلند ارائه شده است. به عنوان مثال پایل [۳۳] و همکاران که برای بهبود خلاصه انتزاعی نهایی متون طولانی از رویکرد ترکیبی استخراجی-انتزاعی با استفاده از مدل زبانی از پیش آموزش دیده جی‌پی‌تی-دو [۳۴] استفاده می‌کنند. در این مدل مرحله استخراج ساده قبل از تولید خلاصه انجام می‌شود، سپس برای شرطی کردن مدل زبانی ترنسفورمر بر روی اطلاعات مربوط قبل از تولید خلاصه استفاده می‌شود. این رویکرد در مقایسه با کارهای قبلی که از مکانیزم کپی استفاده می‌کنند، خلاصه‌های انتزاعی بیشتری تولید می‌کند [۲۲]. پانگ [۳۵] و همکاران یک ساختار سلسله مراتبی برای اسناد طولانی فرض کرده‌اند. در این ساختار سطح بالا بر وابستگی دوربرد تمرکز می‌کند و سطح پایین جزئیات را حفظ می‌کند. در استنتاج از پایین به بالا، تعبیه‌های متنی نشانه‌ها با استفاده از توجه محلی محاسبه می‌شوند و برای دریافت وابستگی های دوربرد و زمینه جهانی، استنتاج از بالا به پایین برای نمایش های توکن اعمال می شود. یک ساختار پنهان چند مقیاسی دو سطحی استفاده می‌شود، که در آن سطح پایین شامل نمایش‌های نشانه‌ای است که توسط استنتاج پایین به بالا محاسبه می‌شود، سپس با اعمال مکانیزم توجه به سطوح بزرگ‌تر روابط بین بخش‌های مختلف سند را بدست می‌آورد. ساختار مدل را در شکل ۳-۱۰ نشان داده شده است. روش پیشنهادی یک رویکرد جدید امیدوارکننده برای خلاصه‌سازی اسناد طولانی است و نسبت به روش های قبلی کارآمدتر و موثرتر است[۲۰].



شکل ۳-۱۰: معماری مدل ترنسفورمر از بالا به پایین[۲۰]

ادیت

جیدیوتیس و همکاران شیوه‌ی تقسیم و غلبه (دنسر) [۳۶] را برای بهبود خلاصه سازی اسناد طولانی

[33] Pilault
[34] GPT-2
[35] Pang
[36] Divide-and-ConquER (DANCER)

پیشنهاد کرده اند.این روش به طور خودکار خلاصه یک سند را به چند بخش تقسیم می‌کند و هر یک از این بخش‌ها را به بخش مناسب سند جفت می‌کند تا خلاصه‌های هدف متمایز ایجاد کند. شیوه‌ی معرفی شده در نظر می‌گیرد که متون طولانی به صورت بخش‌های گسسته ساختاربندی شده‌اند.

برای مطابقت هر قسمت از خلاصه با بخشی از سند در دنسر از معیار روژ [۳۷] استفاده می‌شود. در این روش معیار روژ-ال بین هر یک از جملات خلاصه و تمام جملات سند محاسبه می‌شود و هر جمله ی خلاصه هدف به بخش حاوی جمله با بیشترین روژ- ال نسبت داده می‌شود. سپس تمام جملات خلاصه‌ی هدف مربوط به هر بخش را به هم الحاق می کنیم تا خلاصه‌ی هدف برای هر بخش ایجاد شود. در طول آموزش هر بخش از سند به همراه جمله‌ی خلاصه‌ی مربوط به آن به عنوان متن ورودی و خلاصه‌ی هدف استفاده می‌شود. مزایای این روش آموزش:

۱. تقسیم مساله به چند زیر مساله باعث کاهش پیچیدگی و ساده‌سازی مساله می‌شود.

۲. انتخاب خلاصه‌های هدف برای هر بخش بر اساس امتیازات روژ-ال هر جمله باعث تطابق بهتر و متمرکزتر بین دنباله‌های منبع و هدف ایجاد می‌شود.

۳. تقسیم هر سند آموزشی به چند جفت ورودی-هدف، نمونه های آموزشی بسیار بیشتری ایجاد می کند. این کار برای مدل‌های خلاصه‌سازی عصبی مفید است.

۴. این روش می‌تواند از مدل‌های خلاصه‌سازی مختلف از جمله شبکه‌ی عصبی بازگشتی و ترنسفورمرها استفاده کند.

هنگام کار با اسناد ساختاریافته طولانی، معمولاً همه بخش‌های سند کلیدی برای سند نیستند. اگر یک مقاله آکادمیک را به عنوان مثال در نظر بگیریم، بخش هایی مانند مرور ادبیات یا پیشینه در تلاش برای خلاصه کردن نکات اصلی مقاله ضروری نیستند و باعث افزودن نویز می‌شوند. بنابراین از بخش مرور ادبیات صرف نظر می‌شود و تمرکز سیستم خلاصه‌سازی فقط روی بخش‌های مقدمه، روش‌ها، نتایج و نتیجه‌گیری می‌باشد.

میتونه حذف بشه

این مدل قابل ترکیب با پگاسوس یا مدل مولد نقطه‌ای [۳۸] می‌باشد. بخش کدگشا مدل مولد نقطه‌ای با ایجاد جملات تکراری مقابله می‌کند. هرچند ممکن است به خاطر تکرار اطلاعات در بخش های مختلف بازهم خلاصه‌ی تکراری ایجاد شود.

شیونگ و همکاران با اصلاح هدف بهینه‌سازی، معماری مدل‌های از پیش آموزش دیده و مجموعه‌ی دادگان پیش‌آموزش [۳۹] روشی را برای ساخت مدل‌های مناسب متون طولانی پیشنهاد می‌کنند. مدل‌های پیش‌آموزش دیده متن به متن، مانند برت و بارت، معمولاً بر روی دنباله‌های متن کوتاه، مانند جملات یا پاراگراف‌ها آموزش داده می‌شوند. در حالی که بسیاری از وظایف پردازش زبان طبیعی، مانند پاسخگویی به سؤال و خلاصه کردن، به توانایی پردازش توالی متن طولانی نیاز دارند این مقاله تعدادی از تکنیک‌ها

[37] ROUGE

[38] Pointer-Generator model

[39] pretraining corpus

را برای تطبیق مدل‌های متن به متن از پیش آموزش دیده برای دنباله‌های متن طولانی پیشنهاد می‌کند. این تکنیک ها عبارتند از:

- ارائه‌ی مدل براساس یک ترنسفورمر با مکانیزم توجه به خود پراکنده‌ی بلوکی [۴۰] در قسمت کدگذار است. این مکانیزم امکان استفاده‌ی مجدد از وزن‌های مدل‌های از پیش آموزش دیده را فراهم می‌کند.
- مکانیرم توکن سراسری [۴۱]: در این مکانیزم یک مجموعه‌ی کوچک از توکن‌های سراسری به کل توالی توجه می‌کنند و امکان تعاملات دوربرد در کدگذار فراهم می‌شود.
- هم‌پوشانی بلوک‌های توجه [۴۲] : توجه لغزشی با همپوشانی یک راه ساده برای معرفی اتصالات دوربرد در مدل های توجه محلی است. در این رویکرد، توکن‌های درون هر بلوک به تمام توکن‌های درون خود بلوک و همچنین نیمی از توکن‌های بلوک‌های چپ و راست مجاور نزدیک می‌شوند. این نسخه بلوکی از پنجره‌های توجه همپوشانی، راه ساده‌تر و کارآمدتری را برای معرفی اتصالات دوربرد ارائه می‌کند و در عین حال موازی‌سازی را در پیاده‌سازی مدل تسهیل می‌کند.
- لایه‌ی خود توجه مبتنی بر ادغام بلوکی تقویت شده [۴۳]: این لایه به عنوان جایگزین لایه خود توجهی برای اتصالات دوربرد معرفی شده است. این رویکرد به واحدهای توجه درون بلوک ها اجازه می دهد تا به جای توجه به همسایگان بلافصل خود، بر خلاصه ای از اطلاعات کلی در بلوک ها تمرکز کنند. این لایه در تصویر ۳-۱۱نشان داده شده است. این مدل را قادر می سازد تا از اطلاعات گسترده تری در سراسر سند برای تصمیم گیری استفاده کند و وابستگی های دوربرد را در نظر بگیرد. با بکارگیری عملیات ادغام، ابعاد و نمایش بردارهای توجه کاهش می‌یابد که منجر به افزایش سرعت و کارایی مدل می‌شود.

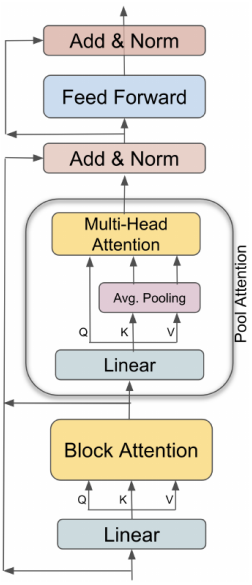
نویسندگان تکنیک‌های پیشنهادی را در تعدادی از وظایف توالی متن طولانی، از جمله پاسخ‌گویی به سؤال و خلاصه‌نویسی، ارزیابی کرده‌اند. نتایج نشان می‌دهد که مدل‌های اقتباس‌شده در تمامی وظایف از مدل‌های پایه بهتر عمل می‌کنند. این تکنیک‌ها استفاده از مدل های متنی از پیش آموزش دیده را برای طیف وسیعی از وظایف پردازش زبان طبیعی ممکن می سازد [۲۶].

[40] Block-sparse self-attention

[41] Global-token mechanism

[42] Overlapping attention windows

[43] Pooling-augmented blockwise attention



شکل ۳-۱۱: لایه خودتوجهی تقویت شده ادغام شده [۲۶]

فصل چهارم

روش های مبتنی بر یادگیری تقویتی

کارهای تحقیقاتی در زمینه ی یادگیری تقویتی [1] و پردازش زبان طبیعی در سال های اخیر رشد کرده است. در یادگیری تقویتی یک عامل با محیط تعامل می کند و با آزمون و خطا، خط مشی بهینه را برای تصمیم گیری متوالی برای به حداکثر رساندن پاداش تجمعی آینده می آموزد. این پاداش می تواند یک معیار تعریف شده توسط توسعه دهنده بر اساس کار در حال حل باشد. در خلاصه سازی خودکار انتزاعی متن، نمونه‌هایی از چنین پاداش‌هایی ممکن است شامل حفظ برجستگی، مستلزم منطقی هدایت‌شده، و غیر افزونگی باشد.

به طور کلی، یادگیری تقویتی در سه حوزه مختلف برای بهبود خلاصه سازی خودکار استفاده می‌شود:

۴-۱ یادگیری تقویتی برای حل مسائل عمیق توالی به دنباله

استفاده از یادگیری تقویتی به منظور حل مسائل گوناگونی که مدل‌های دنباله به دنباله عمیق قادر به حل آن‌ها نیستند، امکانات بیشتری را فراهم می‌کند. به عنوان مثال، مشکلاتی مانند کمبود نوآوری در ایجاد خلاصه‌های خلاقانه و آموزنده و کاهش کیفیت خلاصه ها در صورت افزایش طول مقالات منبع، با استفاده از سیستم‌های یادگیری تقویتی و یادگیری خط‌مشی [2] بهبود یافته است. علاوه بر این مدل‌های دنباله به دنباله عمیق را نمی‌توان برای خلاصه کردن طیف گسترده ای از اسناد استفاده کرد، زیرا مدلی که بر روی یک مجموعه داده آموزش داده می‌شود، در یک مجموعه داده دیگر به خوبی عمل نمی‌کند و قابلیت تعمیم ندارد. رویکردهای مبتنی بر یادگیری تقویتی می‌تواند این مشکل را با استفاده ازگرادیان خط مشی انتقادی [3] و ترکیب آن با یادگیری انتقالی [4] برای انتقال دانش از یک مجموعه داده به مجموعه دیگر برطرف کنند [۱۰].

۴-۲ یادگیری تقویتی برای ترکیب خلاصه های استخراجی و انتزاعی

از یادگیری تقویتی برای ترکیب ویژگی های استخراجی با خلاصه انتزاعی برای استفاده از هر دو نوع خلاصه ی خودکار با الهام از رفتار انسان استفاده می‌شود. این مدل‌ها ابتدا برجسته‌ترین جملات را از سند ورودی استخراج می‌کنند، سپس با استفاده از دو شبکه: شبکه‌های استخراج‌کننده و انتزاعی، آنها را انتزاع می‌کنند. به عنوان مثال لیو و همکاران یک چارچوب متخاصم را پیشنهاد می‌کنند که مدل‌های انتزاعی و استخراجی را همزمان با استفاده از گرادیان خط مشی برای بهینه‌سازی مدل انتزاعی برای خلاصه‌ای با پاداش بالا، آموزش می‌دهد که منجر به خلاصه‌ای منسجم‌تر می‌شود [۱]. همچنین چن و بانسال یک مدل خلاصه سازی سریع پیشنهاد کردند که جملات برجسته را استخراج می کرد و سپس با استفاده از گرادیان خط مشی سطح جمله مبتنی بر یادگیری‌تقویتی بازنویسی می‌کرد [۳].

[1] reinforcement learning

[2] policy learning

[3] self-critic policy gradient

[4] Transfer Learning (TL)

۴-۳ یادگیری تقویتی برای ایجاد معیارها و پاداش های جدید

خلاصه‌سازی اسناد، مانند سایر وظایف مولد زبان، اغلب به دلیل استفاده از اهداف آموزشی مبتنی بر درست‌نمایی بیشینه [۵] مورد انتقاد قرار گرفته است. درست‌نمایی بیشینه کیفیت خلاصه‌ی تولید شده را در نظر نمی‌گیرد و ممکن است خلاصه‌هایی تولید کند که فقط یک کپی از اسناد ورودی هستند، یا می‌توانند خلاصه‌هایی را بیاموزند که پر از کلمات بی‌معنی هستند. به همین دلیل، یادگیری تقویتی به عنوان جایگزینی برای بهینه‌سازی مستقیم مدل‌ها بر روی معیارهای ارزیابی و پاداش صریح به کیفیت پیش‌بینی‌های مدل استفاده شده است [۲۱]. معیارهای مختلفی مانند روژ-۱ [۶]، روژ-۲ [۷]، روژ-ال [۸]، امتیاز اف-۱ [۹] و امتیازبرت [۱۰] به عنوان پاداش در رویکردهای یادگیری تقویتی استفاده شده است. با این حال، پارنل و همکاران استدلال می کند که استفاده از امتیازات روژ به عنوان پاداش، جنبه های مهم خلاصه سازی، مانند خوانایی، روان بودن و اشتراک اطلاعات بین اسنادی در خلاصه سازی چند سندی را نادیده می گیرد و یک پاداش پوشش اصلاح شده همراه با یک برآوردگر گرادیان سیاست مبتنی بر اصول (ریلکس) [۱۱] را پیشنهاد می‌دهند [۲۱، ۱]. ریلکس یک برآوردگر گرادیان خط مشی است که دارای واریانس کم و بی طرفانه است. برای مسائل یادگیری تقویتی با فضاهای کنش مداوم، مانند خلاصه سازی متن، مناسب است [۷].

در تابع ضرر بر حسب ریلکس ۴-۱ بخش اول عبارت سیاست را تشویق می کند تا خروجی هایی تولید کند که پاداش مورد انتظار بالایی دارند و بخش دوم سیاست را تشویق می کند تا خروجی هایی مشابه خروجی های تولید شده در گذشته تولید کند همچنین

r نشان دهنده‌ی پاداش $c_\phi(\tilde{z})$ یک متغیر کنترلی از پارامترهای است که انتظار می رود به شدت با پاداش کاهش واریانس همبستگی داشته باشد. $p(ys)$ احتمال دنباله مشاهده شده خروجی y_s است. z دنباله نمونه های $Gumbel - Softmax$ است. $\tilde{z}$ دنباله ای از نمونه ها از یک توزیع $Gumbel - Softmax$ مشروط بر y_s است.

$$L_{RELAX} = -[r - c_\phi(\tilde{z})] \log p(y^s) \quad + c_\phi(z) - c_\phi(\tilde{z}) \qquad (۱-۴)$$

[5] maximum likelihood
[6] ROUGE-1
[7] ROUGE-2
[8] ROUGE-L
[9] F1-score
[10] BERTScore
[11] modified coverage reward along with a principled policy gradient estimator (RELAX)

فصل پنجم

نتایج

فصل ششم

جمع‌بندی

منابع و مراجع

[1] Alomari, Ayham, Idris, Norisma, Sabri, Aznul Qalid Md, and Alsmadi, Izzat. Deep reinforcement and transfer learning for abstractive text summarization: A review. Computer Speech and Language, 71:101276, 2022.

[2] Andhale, Narendra and Bewoor, Laxmi A. An overview of text summarization techniques. In 2016 international conference on computing communication control and automation (ICCUBEA), pages 1–7. IEEE, 2016.

[3] Chen, Yen-Chun and Bansal, Mohit. Fast abstractive summarization with reinforce-selected sentence rewriting. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 675–686, 2018.

[4] Child, Rewon, Gray, Scott, Radford, Alec, and Sutskever, Ilya. Generating long sequences with sparse transformers, 2019.

[5] El-Kassas, Wafaa S., Salama, Cherif R., Rafea, Ahmed A., and Mohamed, Hoda K. Automatic text summarization: A comprehensive survey. Expert Systems with Applications, 165:113679, 2021.

[6] Elman, Jeffrey L. Finding structure in time. Cognitive science, 14(2):179–211, 1990.

[7] Grathwohl, Will, Choi, Dami, Wu, Yuhuai, Roeder, Geoffrey, and Duvenaud, David Kristjanson. Backpropagation through the void: Optimizing control variates for black-box gradient estimation. ArXiv, abs/1711.00123, 2017.

[8] Hochreiter, Sepp and Schmidhuber, Jürgen. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[9] Ishikawa, Kai, Ando, Shinichi, and Okumura, Akitoshi. Hybrid text summarization method based on the tf method and the lead method. In NTCIR Conference on Evaluation of Information Access Technologies, 2001.

[10] Keneshloo, Yaser, Ramakrishnan, Naren, and Reddy, Chandan K. Deep transfer reinforcement learning for text summarization. ArXiv, abs/1810.06667, 2018.

[11] Kitaev, Nikita, Kaiser, Lukasz, and Levskaya, Anselm. Reformer: The efficient transformer. In International Conference on Learning Representations, 2019.

[12] Lee, Chang-Shing, Jian, Zhi-Wei, and Huang, Lin-Kai. A fuzzy ontology and its application to news summarization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 35(5):859–880, 2005.

[13] Lewis, Mike, Liu, Yinhan, Goyal, Naman, Ghazvininejad, Marjan, Mohamed, Abdelrahman, Levy, Omer, Stoyanov, Veselin, and Zettlemoyer, Luke. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7871–7880, Online, July 2020. Association for Computational Linguistics.

[14] Liang, Yunlong, Meng, Fandong, Zhou, Chulun, Xu, Jinan, Chen, Yufeng, Su, Jinsong, and Zhou, Jie. A variational hierarchical model for neural cross-lingual summarization. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2088–2099, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[15] Liu, Qi, Kusner, Matt J, and Blunsom, Phil. A survey on contextual embeddings. arXiv preprint arXiv:2003.07278, 2020.

[16] Luhn, Hans Peter. The automatic creation of literature abstracts. IBM Journal of research and development, 2(2):159–165, 1958.

[17] Ma, Tinghuai, Pan, Qian, Rong, Huan, Qian, Yurong, Tian, Yuan, and Al-Nabhan, Najla Abdulrahman. T-bertsum: Topic-aware text summarization based on bert. IEEE Transactions on Computational Social Systems, 9:879–890, 2022.

[18] Malliaros, Fragkiskos D. and Skianis, Konstantinos. Graph-based term weighting for text categorization. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ASONAM '15, page 1473–1479, New York, NY, USA, 2015. Association for Computing Machinery.

[19] Moratanch, N. and Chitrakala, S. A survey on abstractive text summarization. In 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), pages 1–7, 2016.

[20] Pang, Bo, Nijkamp, Erik, Kryscinski, Wojciech, Savarese, Silvio, Zhou, Yingbo, and Xiong, Caiming. Long document summarization with top-down and bottom-up inference. In Findings of the Association for Computational Linguistics: EACL 2023, pages 1237–1254, 2023.

[21] Parnell, Jacob, Unanue, Inigo Jauregi, and Piccardi, Massimo. A multi-document coverage reward for relaxed multi-document summarization. In Annual Meeting of the Association for Computational Linguistics, 2022.

[22] Pilault, Jonathan, Li, Raymond, Subramanian, Sandeep, and Pal, Christopher. On extractive and abstractive neural document summarization with transformer language

models. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 9308–9319, 2020.

[23] Sherborne, Tom and Lapata, Mirella. Meta-learning a cross-lingual manifold for semantic parsing. Transactions of the Association for Computational Linguistics, 11:49–67, 2023.

[24] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, ukasz, and Polosukhin, Illia. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[25] Xiao, Wen and Carenini, Giuseppe. Systematically exploring redundancy reduction in summarizing long documents. In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, pages 516–528, 2020.

[26] Xiong, Wenhan, Gupta, Anchit, Toshniwal, Shubham, Mehdad, Yashar, and tau Yih, Wen. Adapting pretrained text-to-text models for long text sequences. ArXiv, abs/2209.10052, 2022.

[27] Yao, Kaichun, Zhang, Libo, Du, Dawei, Luo, Tiejian, Tao, Lili, and Wu, Yanjun. Dual encoding for abstractive text summarization. IEEE transactions on cybernetics, 50(3):985–996, 2018.

[28] Zaheer, Manzil, Guruganesh, Guru, Dubey, Kumar Avinava, Ainslie, Joshua, Alberti, Chris, Ontanon, Santiago, Pham, Philip, Ravula, Anirudh, Wang, Qifan, Yang, Li, et al. Big bird: Transformers for longer sequences. Advances in neural information processing systems, 33:17283–17297, 2020.

[29] Zhang, Jingqing, Zhao, Yao, Saleh, Mohammad, and Liu, Peter. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In International Conference on Machine Learning, pages 11328–11339. PMLR, 2020.

واژه‌نامه‌ی فارسی به انگلیسی

ا

ابر تکامل عصبی توپولوژی تقویت‌کننده Hyper Neuroevolution of Augmenting Topologies

ازدحام ذرات . . . . . Particle Swarm

استنتاج زبان طبیعی Natural Language Inference

الگوریتم تکاملی . . . . . Evolutionary Algorithm

انتروپی . . . . . Entropy

انتشار به عقب . . . . . Backpropagation

ب

برپایه داده . . . . . Data Driven

بردار ماشین پشتیبان . . Support Vector Machine

بستر . . . . . Substrate

بیز ساده‌لوحانه . . . . . Naive Bayes

بیل . . . . . Beel

پ

پاسخ دوره‌ای به سوالات . . . . . Episodic Question-Answering

پیوسته . . . . . Continues

ت

تابع قدم . . . . . Step Function

تارشدن . . . . . Blurring

تعویض . . . . . Switch

تقارن . . . . . Symmetry

تکامل عصبی توپولوژی‌های تقویت‌کننده Neuroevolution of Augmenting Topologies

تکرار . . . . . Repetition

تکرار با تغییر Repetition with Variation

توجه چندسر . Multi Headed Attention

توجه سخت . . . . . Hard Attention

توجه نرم . . . . . Soft Attention

تیزکردن . . . . . Sharpen

ج

جابجایی . . . . . Shift

جمعیت . . . . . Population

جهش . . . . . Mutation

خ

خطای هوبر . . . . . Hubber Loss

خطای ریشه مربعات خطا . Rooted Mean Square Error

د

دانش پیشین . . . . . Prior Knowledge

دروازه . . . . . Gate

درون‌یابی . . . . . Interpolation

دسته . . . . . Batch

ر

ردیابی دانش . . . . Knowledge Tracing

رونوشت‌گیری . . . . . Copy

رونوشت‌گیری تکرارشونده . Repeat Copy

ژ

ژائو . . . . . Zhao

س

سر . . . . . Head

سری‌های زمانی . . . . . Time Series

ش

شباهت کسینوسی . . Cosine Similarity

شبکه پرسپترونی چندلایه . Multi Layer Perceptran

شبکه تولید الگوی ترکیبی Compositional Pattern Producing Networks

شبکه عصبی . . . . . Neural Network

شبکه عصبی بازگشتی Recurrent Neural Network

شبکه عصبی بازگشتی دروازه‌دار . . Gated Recurrent Neural Network

شبکه عصبی جلورو Feedforward Neural Network

شبکه حافظه‌ای . . . . Memory Network

شروع سرد . . . . . Cold Start

ص

صحت . . . . . Accuracy

ف

فاکتور تیزی . . . . . Sharpness Factor

فالکن . . . . . Falcon

ک

کامپیوتر عصبی متمایز . . Differentiable Neural Computer

کدگذاری . . . . . Encoding

کنترل‌گر . . . . . Controller

کولیر . . . . . Collier

گ

گالچره . . . . . Gulcehre

گره . . . . . Node

گریوز . . . . . Graves

گسسته . . . . . Discrete

ل

لاجیتس . . . . . Logits

م

ماشین تورینگ عصبی . . Neural Turing Machine

ماشین تورینگ عصبی ابرتکاملی . . Hyper Evolvable Neural Turing Machine

ماشین تورینگ عصبی پویا . . . Dynamic Neural Turing Machine

ماشین تورینگ عصبی تکاملی Evolvable Neural Turing Machine

معیار شباهت . . . . . Similarity Metric

مقداردهی اولیه . . . . . Initialization

مقیاس‌پذیری . . . . . Scalability

مکانیسم توجه . . Attention Mechanism

منبع‌باز . . . . . Open Source

منظم‌سازی . . . . . Regularity

ن

ناحیه زیرنمودار . . . . Area Under Curve

نسل . . . . . Generation

نوار . . . . . Tape

و

وضعیت مخفی . . . . . Hidden State

ه

هسته . . . . . Kernel

ی

یادآوری انجمنی . . . Association Recall

یادگیری ترتیبی . . . Sequence Learning

یال . . . . . Edge

واژه‌نامه‌ی انگلیسی به فارسی

A

Accuracy . . . . . صحت

Area Under Curve . . . . ناحیه زیرنمودار

Association Recall . . . یادآوری انجمنی

Attention Mechanism . . مکانیسم توجه

B

Backpropagation . . . . . انتشار به عقب

Batch . . . . . دسته

Beel . . . . . بیل

Blurring . . . . . تارشدن

C

Cold Start . . . . . شروع سرد

Collier . . . . . کولیر

Compositional شبکه تولید الگوی ترکیبی
Pattern Producing Networks

Continues . . . . . پیوسته

Controller . . . . . کنترل‌گر

Copy . . . . . رونوشت‌گیری

Cosine Similarity . . شباهت کسینوسی

D

Data Driven . . . . . برپایه داده

Differentiable . . کامپیوتر عصبی متمایز
Neural Computer

Discrete . . . . . گسسته

Dynamic . . . ماشین تورینگ عصبی پویا
Neural Turing Machine

E

Edge . . . . . یال

Episodic . . . . . پاسخ دوره‌ای به سوالات
Question-Answering

Encoding . . . . . کدگذاری

Entropy . . . . . انتروپی

Evolutionary . . . . . الگوریتم تکاملی
Algorithm

مقداردهی اولیه . . . . . Initialization

درون‌یابی . . . . . Interpolation

K

هسته . . . . . Kernel

ردیابی دانش . . . . Knowledge Tracing

k-نزدیک‌ترین همسایه . . . . . K-Nearest Neighbours

L

لاجیتس . . . . . Logits

M

شبکه حافظه‌ای . . . . Memory Network

توجه چندسر . Multi Headed Attention

شبکه پرسپترونی چندلایه . Multi Layer Perceptran

جهش . . . . . Mutation

N

بیز ساده‌لوحانه . . . . . Naive Bayes

استنتاج زبان طبیعی Natural Language Inference

شبکه عصبی . . . . . Neural Network

ماشین تورینگ عصبی . . Neural Turing Machine

تکامل عصبی توپولوژی‌های تقویت‌کننده Neuroevolution of Augmenting Topologies

گره . . . . . Node

O

ماشین تورینگ عصبی تکاملی Evolvable Neural Turing Machine

F

فالکن . . . . . Falcon

شبکه عصبی جلورو Feedforward Neural Network

G

دروازه . . . . . Gate

شبکه عصبی بازگشتی دروازه‌دار . . Gated Recurrent Neural Network

نسل . . . . . Generation

گریوز . . . . . Graves

گالچره . . . . . Gulcehre

H

توجه سخت . . . . . Hard Attention

سر . . . . . Head

وضعیت مخفی . . . . . Hidden State

خطای هوبر . . . . . Hubber Loss

ماشین تورینگ عصبی ابرتکاملی . . Hyper Evolvable Neural Turing Machine

ابر تکامل عصبی توپولوژی تقویت‌کننده Hyper Neuroevolution of Augmenting Topologies

I

یادگیری ترتیبی . . . Sequence Learning

تیزکردن . . . . . Sharpen

فاکتور تیزی . . . . . Sharpness Factor

جابجایی . . . . . Shift

معیار شباهت . . . . . Similarity Metric

توجه نرم . . . . . Soft Attention

بستر . . . . . Substrate

بردار ماشین پشتیبان . . Support Vector Machine

تابع قدم . . . . . Step Function

تعویض . . . . . Switch

تقارن . . . . . Symmetry

T

نوار . . . . . Tape

سری‌های زمانی . . . . . Time Series

Z

ژائو . . . . . Zhao

تک‌روشن . . . . . One-hot

منبع‌باز . . . . . Open Source

P

ازدحام ذرات . . . . . Particle Swarm

جمعیت . . . . . Population

دانش پیشین . . . . . Prior Knowledge

R

شبکه عصبی بازگشتی Recurrent Neural Network

منظم‌سازی . . . . . Regularity

رونوشت‌گیری تکرارشونده . Repeat Copy

تکرار . . . . . Repetition

تکرار با تغییر Repetition with Variation

خطای ریشه مربعات خطا . Rooted Mean Square Error

S

مقیاس‌پذیری . . . . . Scalability