



دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

پروژه تحقیقاتی درس سمینار

روش‌های خلاصه‌سازی انتزاعی

نگارش

زهرا زنجانی

استاد درس

دکتر رضا صفابخش

۱۴۰۱ - ۱۴۰۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

## 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<sup>۱</sup> 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، تکنیک‌ها 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

---

<sup>1</sup> natural language processing (NLP)

# فهرست مطالب

صفحه

عنوان

۵	فهرست نمادها
۱	۱ مقدمه
۳	۲ روش‌های مبتنی بر ساختار
۴	۱-۲ روش مبتنی بر درخت
۴	۲-۲ روش مبتنی بر قالب
۵	۳-۲ روش مبتنی بر هستان شناسی
۵	۴-۲ روش عبارت مقدمه و بدنه
۵	۵-۲ روش مبتنی بر گراف
۶	۶-۲ روش مبتنی بر قانون
۷	۳ روش‌های مبتنی بر شبکه‌ی عصبی
۸	۱-۳ روش‌های مبتنی بر مدل کدگذار-کدگشا
۹	۲-۳ روش‌های مبتنی بر مدل ترنسفورمرها
۱۷	۱-۲-۳ ایده‌های ارائه شده بهبود خلاصه‌سازی متون طولانی
۲۲	۴ روش‌های مبتنی بر یادگیری تقویتی
۲۳	۱-۴ یادگیری تقویتی برای حل مسائل عمیق توالی به دنباله
۲۳	۲-۴ یادگیری تقویتی برای ترکیب خلاصه‌های استخراجی و انتزاعی
۲۴	۳-۴ یادگیری تقویتی برای ایجاد معیارها و پاداش‌های جدید
۲۵	۵ نتایج
۲۶	۶ جمع‌بندی
۲۷	منابع و مراجع
۳۱	واژه‌نامه‌ی فارسی به انگلیسی
۳۴	واژه‌نامه‌ی انگلیسی به فارسی

شکل	فهرست اشکال	صفحه
۱-۳	معماری پایه‌ی مدل کدگذار-کدگشا [۱]	۸
۲-۳	معماری پایه‌ی مدل دوگانه‌ی کدگذار [۲۶]	۹
۳-۳	معماری پایه‌ی مدل سلسله‌مراتبی متغیر برای خلاصه‌سازی متقابل زبانی [۱۴]	۱۰
۴-۳	معماری مدل تی‌برت‌سام [۱۷]	۱۱
۵-۳	تعبیه مدل تی‌برت‌سام [۱۷]	۱۲
۶-۳	معماری ترنسفورمر تی‌برت‌سام [۱۷]	۱۲
۷-۳	چهارچوب ایجاد خلاصه با طول متغیر [۲۳]	۱۳
۸-۳	ساختار مدل بارت [۱۳]	۱۵
۹-۳	ساختار مدل پگاسوس [۲۸]	۱۶
۱۰-۳	الگوریتم امدات [۲۲]	۱۷
۱۱-۳	معماری مدل ترنسفورمر از بالا به پایین [؟]	۱۹
۱۲-۳	لایه خودتوجهی تقویت شده ادغام شده [۲۵]	۲۱

صفحه

فهرست جداول

جدول

## فهرست علائم و اختصارات

عنوان اختصاری      عنوان کامل

ریلکسی

پاداش همراه با برآوردگر گرادیان سیاست مبتنی بر اصول

پرش

کن

# 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 می‌کنند و حتی نمی‌توانند 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 ویژگی‌ها 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<sup>۱</sup>، روش‌های بر پایه‌ی مدل تبدیل‌کننده<sup>۲</sup> و مدل‌های زبانی از پیش آموزش دیده<sup>۳</sup> می‌باشد. همچنین برخی از پژوهش‌های اخیر نشان داده‌اند، استفاده از رویکردهای مبتنی بر یادگیری تقویتی<sup>۴</sup> 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.

<sup>1</sup>Deep neural sequence to sequence models

<sup>2</sup>transformer

<sup>3</sup>Pretrained language models (PTLMs)

<sup>4</sup> 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<sup>۱</sup>، مبتنی بر قالب<sup>۲</sup>، مبتنی بر هستان‌شناسی<sup>۳</sup>، عبارت مقدمه و بدنه<sup>۴</sup>، مبتنی بر گراف<sup>۵</sup> و مبتنی بر قانون<sup>۶</sup> مورد بررسی قرار می‌گیرد.

## ۱-۲ 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].

<sup>1</sup>tree-based

<sup>2</sup>template-based

<sup>3</sup>ontology-based

<sup>4</sup>lead-and-body phrase

<sup>5</sup>graph-based

<sup>6</sup>rule-based

## ۳-۲ 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 رسمی انواع 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 هستی‌شناسی طبقه بندی می‌شوند. مکانیزم استنتاج فازی درجه عضویت برای هر جمله را با توجه به طبقه‌بندی کننده محاسبه می‌کند [۱۲].

## ۴-۲ 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 معمولاً 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<sup>۷</sup> 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].

## ۵-۲ 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

<sup>7</sup>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 اسکینیس 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].

## ۶-۲ 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، اساساً، 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. \_\_\_\_\_

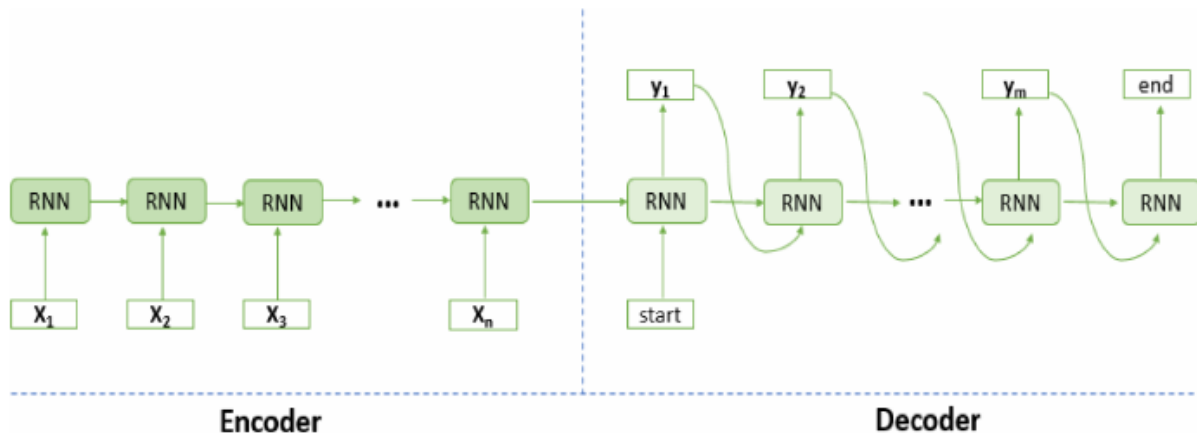
[۱۹].

## فصل سوم

### روش‌های مبتنی بر شبکه‌ی عصبی

### ۱-۳ روش‌های مبتنی بر مدل کدگذار-کدگشا

قبل از ظهور ترنسفورمرها، مدل‌های شبکه عصبی عمیق دنباله به دنباله<sup>۱</sup> بهترین مدل برای کارهای 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 ۱-۳ معماری پایه‌ی مدل کدگذار-کدگشا را شرح می‌دهد. شبکه‌های بازگشتی<sup>۲</sup> [۶] و حافظه‌های کوتاه مدت طولانی<sup>۳</sup> [۸] برای توالی طراحی شده‌اند و برای کدگذاری و پردازش داده‌های دنباله‌ای مانند متن مناسب هستند اما در مدیریت حافظه‌ی بلند مدت مشکل دارند.



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

یائو<sup>۴</sup> و همکاران مدل کدگذاری دوگانه را برای خلاصه‌سازی انتزاعی پیشنهاد داده‌اند. این مدل برای درک بهتر روابط بین متن ورودی و خلاصه مرجع بازنمایی متن ورودی و بازنمایی خلاصه‌ی مرجع را می‌آموزد. همانطور که در شکل ۲-۳ نشان داده شده است. 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<sup>۵</sup> استفاده می‌کنند. کدگذار اولیه بردارهای معنایی هر کلمه در ترتیب ورودی را محاسبه می‌کند. کدگذار ثانویه وزن اهمیت هر کلمه در ترتیب ورودی و بردارهای معنایی مربوطه را دوباره

<sup>۱</sup>sequence to sequence

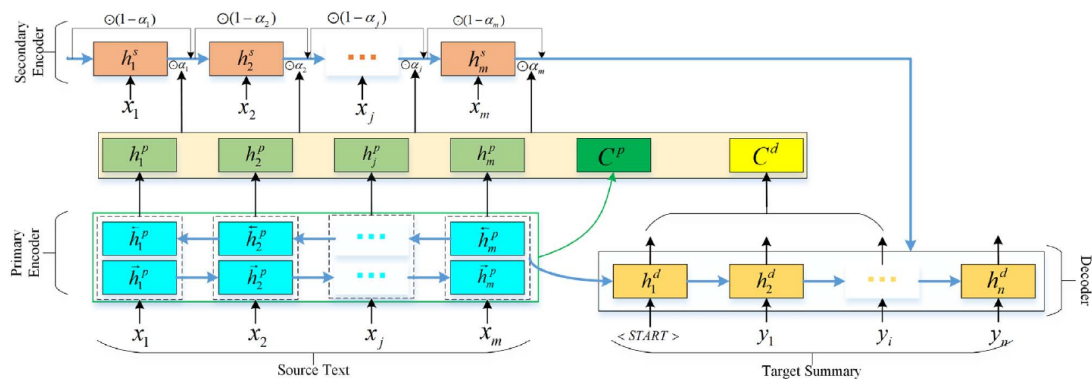
<sup>۲</sup>recurrent neural network (RNN)

<sup>۳</sup>long short-term memory networks(LSTM)

<sup>۴</sup>Yao

<sup>۵</sup>gated recurrent unit (GRU)

محاسبه می‌کند. در نهایت کدگشا با مکانیسم توجه به صورت مرحله‌ای کدگشایی می‌کند و در هر مرحله یک توالی خروجی با طول ثابت جزئی ایجاد می‌کند. در این مدل کدگذار ثانویه عملیات کدگذاری را براساس ورودی هر مرحله و خروجی مرحله‌ی قبل انجام می‌دهد بنابراین کیفیت متون قبلی تولید شده توسط کدگشا بر خروجی‌های جدید تاثیر می‌گذارد [۲۶].



شکل ۳-۲: معماری پایه‌ی مدل دوگانه‌ی کدگذار [۲۶]

مدل کدگذاری دوگانه مدل سلسله مراتبی متغیر بر اساس مدل کدگذاری دوگانه برای خلاصه‌سازی متقاطع زبانی<sup>۶</sup> پیشنهاد شده است. این مدل شامل دو متغیر نهفته محلی و یک متغیر نهفته جامع است. از متغیرهای نهفته محلی برای بازسازی ترجمه و خلاصه زبان مبدأ و از متغیر نهفته سراسری برای تولید خلاصه بین زبانی استفاده می‌شود. قسمت کدگذار این مدل دو بخش دارد که هر بخش وظیفه‌ی تولید یکی از متغیرهای نهفته محلی را دارد و بخش کدگشا با استفاده از نمایش‌های نهفته‌ی محلی خلاصه‌ی نهایی را تولید می‌کند. ساختار سلسله مراتبی این مدل به آن اجازه می‌دهد تا رابطه سلسله مراتبی بین ترجمه، خلاصه‌سازی و خلاصه‌سازی بین زبانی را بیاموزد [۱۴].

## ۳-۲ روش‌های مبتنی بر مدل ترنسفورمرها

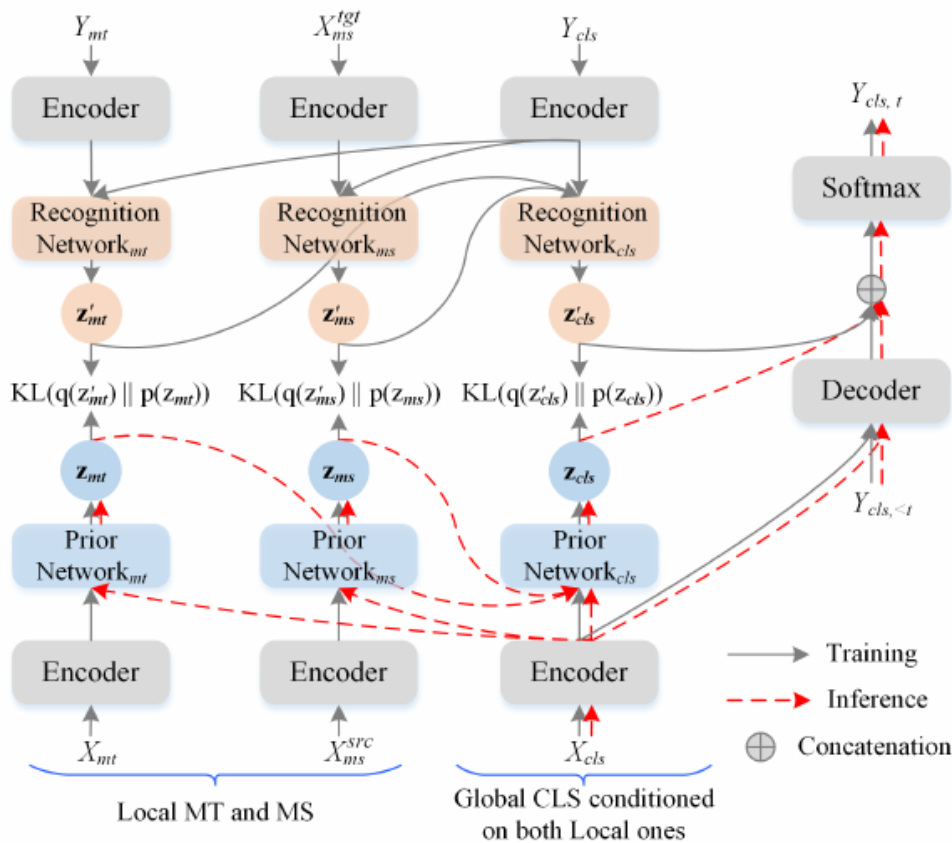
با ظهور ترنسفورمرها<sup>۷</sup>، بهبودهای قابل توجهی در کیفیت نتایج خلاصه‌سازی خودکار به وجود آمد. ترنسفورمرها با استفاده از مکانیزم توجه به خود<sup>۸</sup> شباهت بین ورودی‌ها را بدون توجه به موقعیت موازی آن‌ها با حضور مستقل هر توکن در توالی ورودی مدل می‌کنند و به طور مؤثر مشکلات شبکه‌های بازگشتی را 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

<sup>۶</sup>cross-lingual

<sup>۷</sup>transformers

<sup>۸</sup>Self-attention





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

متغیرهای محلی  $z_{ms}$  و  $z_{mt}$  به ترتیب برای ترجمه و خلاصه‌سازی و متغیر جامع  $z_{cls}$  برای خلاصه‌سازی بین زبانی طراحی شده‌اند. خطوط خاکستری نشان‌دهنده فرآیند آموزشی است که مسئول تولید  $(z'_{cls}, z'_{ms}, z'_{mt})$  از توزیع پسین متناظر پیش‌بینی شده توسط شبکه است. خطوط قرمز خط چین نشان‌دهنده فرآیند استنتاج برای تولید نمایش‌های نهفته  $(z_{cls}, z_{ms}, z_{mt})$  از توزیع‌های پیش‌بینی شده توسط شبکه‌های قبلی است.

بازنمایی‌های زبان برای بهبود کیفیت و دقت خلاصه‌ها استفاده می‌کنند.

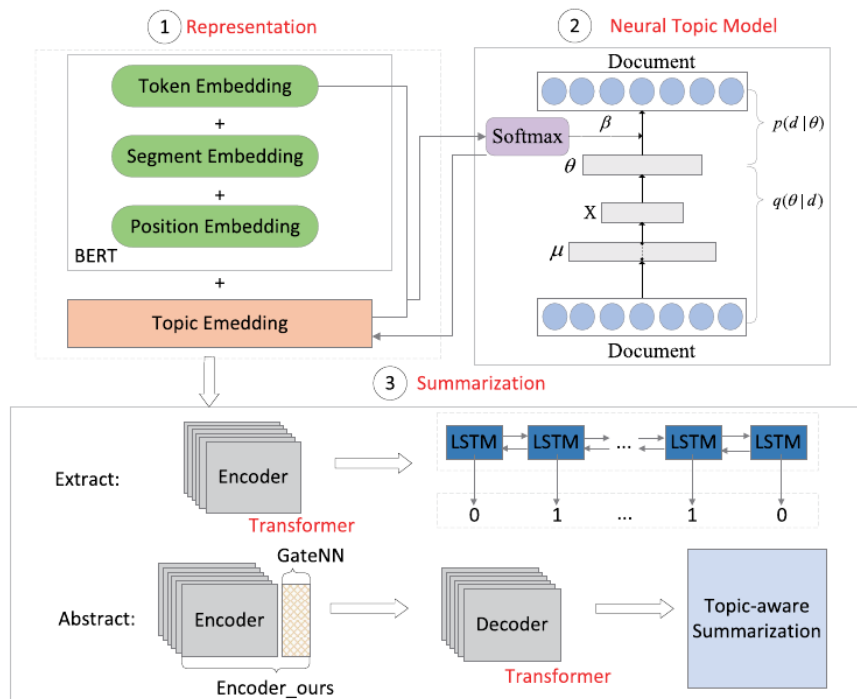
پان<sup>۹</sup> و همکاران یک مدل خلاصه‌سازی بر اساس مدل برت را پیشنهاد کرده‌اند. نویسندگان استدلال می‌کنند که خلاصه‌های تولید شده توسط مدل‌های خلاصه‌سازی متن موجود که موضوع متن را در نظر نمی‌گیرند، مرتبط یا حاوی اطلاعات مفید نیستند. مدل ارائه شده که تی‌برت‌سام<sup>۱۰</sup> نامیده می‌شود از سه بخش ایجاد بازنمایی، مدل موضوعی عصبی<sup>۱۱</sup> و مدل خلاصه‌سازی تشکیل شده است. ساختار مدل را در شکل ۴-۳ نشان داده شده است.

همانطور که در شکل ۵-۳ نشان داده شده است، بازنمایی ایجاد شده برای هر جمله ورودی، با استفاده

<sup>۹</sup>Pan

<sup>۱۰</sup>T-BERTSum

<sup>۱۱</sup>Neural Topic Model (NTM)



شکل ۳-۴: معماری مدل تی‌برت‌سام [۱۷]

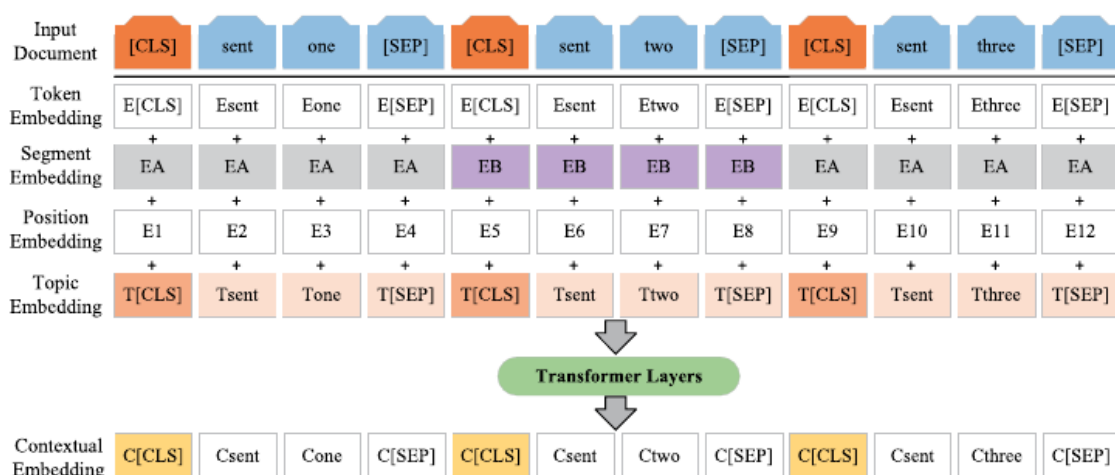
از یک شبکه‌ی ترنسفورمر دوسویه‌ی<sup>۱۲</sup> چند لایه و حاصل جمع چهار نوع تعبیه (تعبیه نشانه<sup>۱۳</sup>، تعبیه قطعه<sup>۱۴</sup>، تعبیه موقعیت و تعبیه موضوع) به دست می‌آید که تعبیه موضوع در این مقاله معرفی شده و سه تعبیه دیگر مشابه مدل برت هستند. وجود تعبیه موضوع در تولید بازنمایی هر کلمه یا جمله موجب 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 مولد و یک شبکه استنتاج. شبکه مولد یک سند را به عنوان ورودی می‌گیرد و یک توزیع موضوعی را بر روی کلمات موجود در سند خروجی می‌دهد. شبکه استنتاج یک سند را به عنوان ورودی می‌گیرد و خروجی آن پارامترهای توزیع موضوع است. بخش خلاصه‌سازی مدل 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 حالت‌های پنهان و متن خلاصه را تولید می‌کند. همانطور که در شکل ۳-۶ نشان داده شده است، به منظور فیلتر کردن اطلاعات کلیدی توالی ورودی، شبکه دروازه‌ای قبل از کدگشا اضافه می‌شود. 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

<sup>12</sup>bidirectional

<sup>13</sup>token embedding

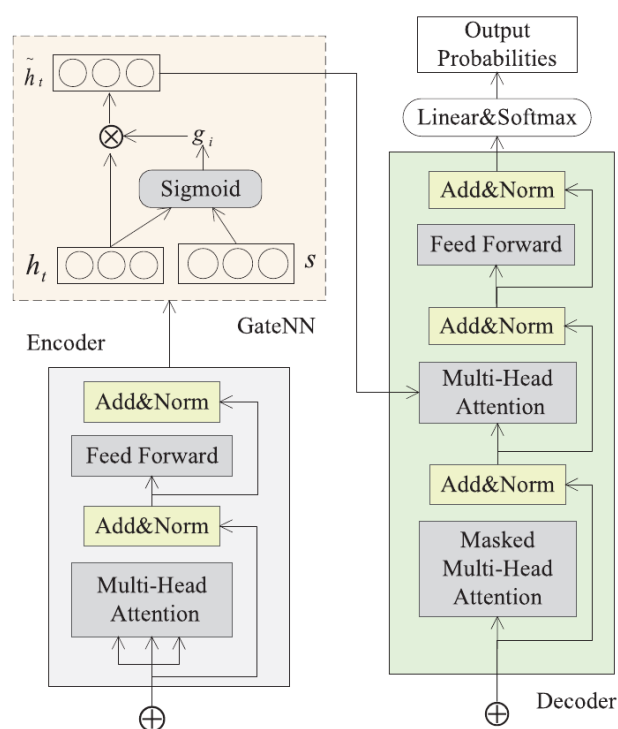
<sup>14</sup>segment embedding



شکل ۳-۵: تعبیه مدل تی‌برت‌سام [۱۷]

کلیدی و حذف اطلاعات غیرضروری تمرکز کند. این مدل می‌تواند خلاصه‌هایی تولید کند که مرتبط با موضوع متن و حاوی اطلاعات مفید باشد و قابلیت تطبیق با حوزه‌های مختلف را دارد.

جمله‌ی  
آخر  
تکراری  
هست

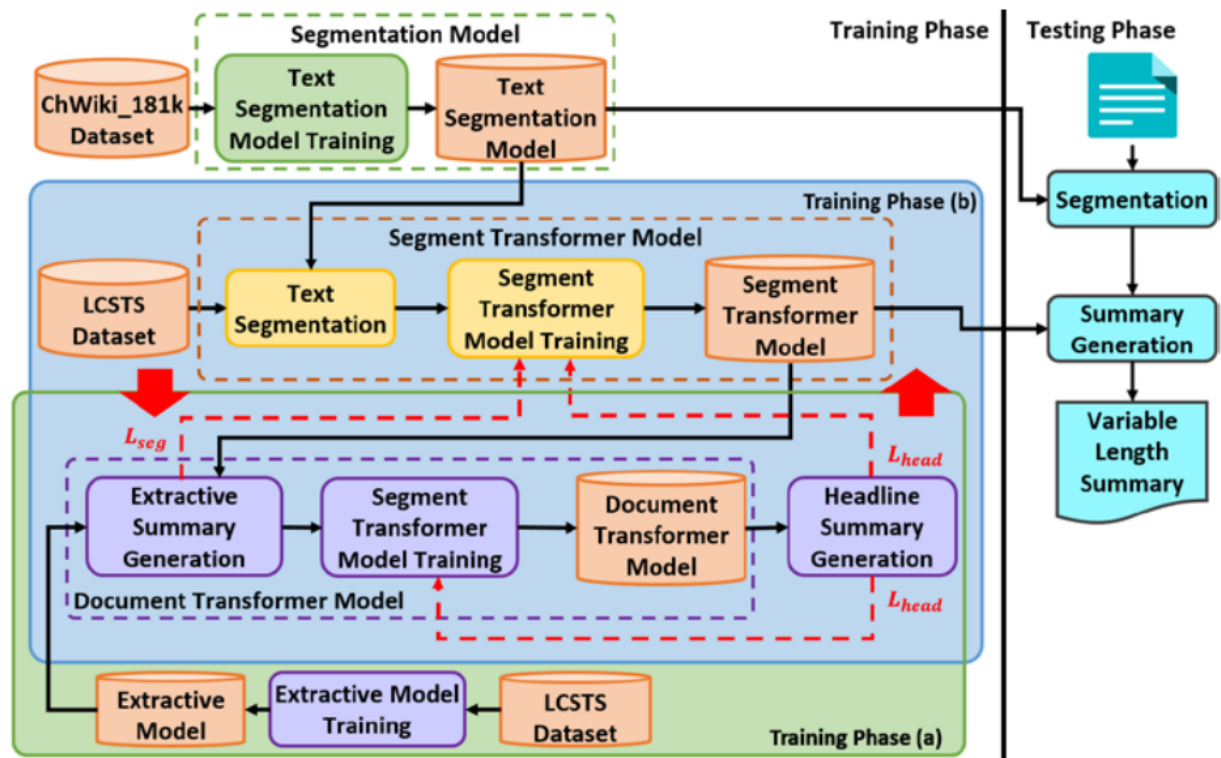


شکل ۳-۶: معماری ترنسفورمر تی‌برت‌سام [۱۷]

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

## ادیت

اکثر مدل‌های خلاصه‌سازی انتزاعی موجود برای تولید خلاصه‌های با طول ثابت طراحی شده‌اند، بنابراین سو<sup>۱۵</sup> و همکاران یک مدل دو مرحله‌ای مبتنی بر ترنسفورمر ارائه دادند که خلاصه‌های انتزاعی با طول متغیر را با توجه به تقاضای کاربر تولید کند. مطابق شکل ۷-۳ مدل پیشنهادی با تقسیم متن ورودی به بخش‌ها و تولید خلاصه‌ی هر بخش، خلاصه‌ی انتزاعی با طول متغیر تولید می‌کند [۲۳].



شکل ۷-۳: چهارچوب ایجاد خلاصه با طول متغیر [۲۳]

بخش‌های مدل ارائه شده به شرح زیر است.

- بخش بندی متن: این مرحله متن ورودی را به تعدادی قسمت از پیش تعیین شده تقسیم می‌کند. تعداد بخش‌ها را می‌تواند توسط کاربر مشخص شود یا با توجه به نسبت دلخواه طول ورودی تنظیم کرد. برای شناسایی مرزهای بین بخش‌ها از مدل  $BERT - biLSTM$  استفاده می‌شود. این مرحله تضمین می‌کند که مرحله خلاصه‌سازی انتزاعی بر روی بخش‌های منسجم متن انجام می‌شود. هدف این بخش یافتن نقاط تقسیم‌بندی است که نشان دهنده تغییر موضوع در متن است و به بهبود کیفیت خلاصه‌های تولید شده کمک می‌کند.

<sup>15</sup>Ming-Hsiang Su

- خلاصه‌سازی استخراجی: پس از تقسیم‌بندی متن، با استفاده از یک مدل خلاصه‌سازی استخراجی مبتنی بر برت‌سام<sup>۱۶</sup> مهم‌ترین جمله را از هر بخش استخراج می‌شود.

•

خلاصه‌سازی اسناد: در مرحله دوم از جملات استخراج شده برای آموزش ماژول خلاصه‌سازی اسناد استفاده می‌شود. این ماژول یک خلاصه سرفصل از کل ورودی متن ایجاد می‌کند. پارامترهای این ماژول با در نظر گرفتن امتیازات ضرر ماژول خلاصه‌سازی اسناد و ماژول خلاصه‌سازی بخش به روز می‌شود. خلاصه‌سازی بخش‌ها: بخش‌های به دست آمده از مرحله تقسیم‌بندی متن برای آموزش ماژول خلاصه‌سازی در مرحله اول استفاده می‌شود. این ماژول یک خلاصه بر اساس جمله برای هر بخش تولید می‌کند. امتیازات ضرر ماژول خلاصه‌سازی سند و ماژول خلاصه‌سازی بخش برای به روز رسانی پارامترهای ماژول خلاصه‌سازی بخش در نظر گرفته می‌شود.

آموزش مشارکتی: آموزش مشارکتی برای آموزش متناوب ماژول خلاصه‌سازی بخش و ماژول خلاصه‌سازی اسناد تا زمان همگرایی اعمال می‌شود. این فرآیند به بهینه سازی عملکرد هر دو ماژول کمک می‌کند. 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 بخش‌ها را می‌توان توسط کاربر مشخص کرد یا با توجه به نسبت دلخواه طول ورودی تنظیم کرد.

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 می‌تواند خلاصه‌های انتزاعی با طول متغیر را با توجه به خواسته‌های کاربر ایجاد کند. این یک پیشرفت نسبت به مدل‌های قبلی است زیرا می‌تواند به طور همزمان به خلاصه‌سازی انتزاعی روان و با طول متغیر دست یابد.

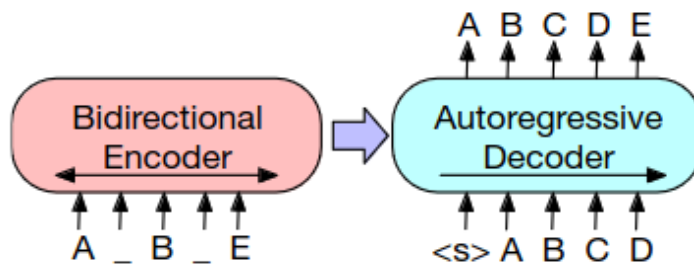
لونیس و همکاران مدلی با نام بارت<sup>۱۷</sup> ارائه دادند. این مدل مشابه با مدل اصلی تبدیل‌کننده، ساختاری کدگذار-کدگشا دارد. بر خلاف سادگی این مدل این مدل را می‌توان نسخه عمومی‌تری از برت و جی‌پی‌تی<sup>۱۸</sup> (به دلیل داشتن کدگذار دو طرفه و کدگشای چپ به راست) دانست. (۳-۸) این مدل در عملیات تولید متن، مانند ترجمه ماشینی یا خلاصه‌سازی 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 گسترده‌ای از نویزها 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 [۱۳، ۱۵].

با این که بارت دقت خلاصه‌سازی انتزاعی متن را بهبود بخشید، ولی عمل‌های تعریف شده در مرحله

<sup>16</sup>BertSum

<sup>17</sup>BART

<sup>18</sup>GPT



شکل ۳-۸: 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].

پیش‌آموزش آن، مختص خلاصه‌سازی انتزاعی متن نبودند، در نتیجه در سال ۲۰۲۰ مدلی تحت عنوان پگاسوس<sup>۱۹</sup> 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<sup>۲۰</sup> از قبل آموزش داده شده است [۲۸].

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<sup>۲۱</sup> نزدیک‌تر باشد، نتیجه نهایی بهتر و همچنین تنظیم دقیق پارامترها<sup>۲۲</sup> سریع‌تر 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 شده است. بر اساس یک متغیر که درصد جملات پنهان شده 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: جمله به صورت تصادفی از متن انتخاب شده و پنهان می‌شوند.
- انتخاب جملات اول متن: جمله اول متن پنهان می‌شوند. دلیل این کار، فرض مهم‌تر بودن جملات ابتدای متن نسبت به جملات بعدی است.

<sup>19</sup>PEGASUS

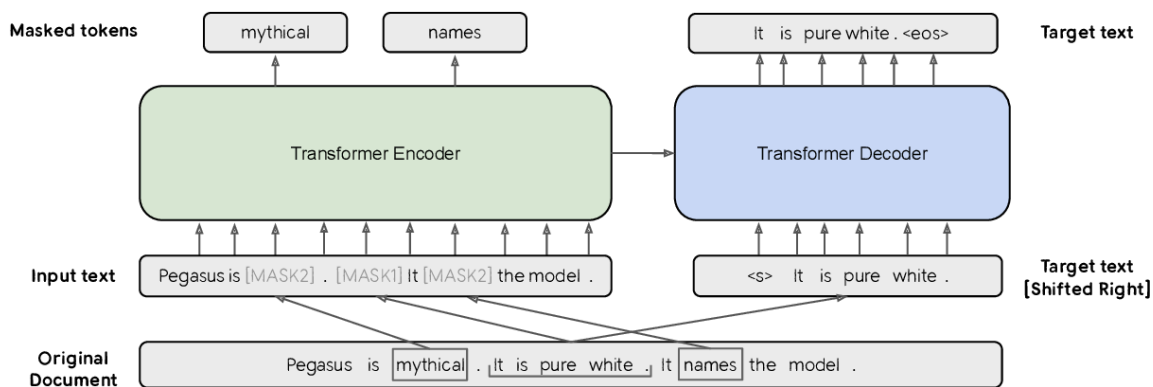
<sup>20</sup>gap sentences generation

<sup>21</sup>downstream task

<sup>22</sup>fine-tuning

- 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  $[MASK2]$  و ۱۰ درصد توکن‌ها با یک توکن تصادفی جایگزین می‌شوند. ۱۰ درصد دیگر بدون تغییر باقی می‌ماند. شکل ۳-۹ اعمال همزمان این دو عمل، یعنی تولید جمالت فاصله افتاده و مدل زبانی پوشیده شده را بر روی یک مثال نشان می‌دهد.



شکل ۳-۹: ساختار مدل پگاسوس [۲۸]

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 می‌شوند. 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 نشانه‌ها به طور تصادفی توسط  $[MASK2]$  پوشانده می‌شوند [۲۸].

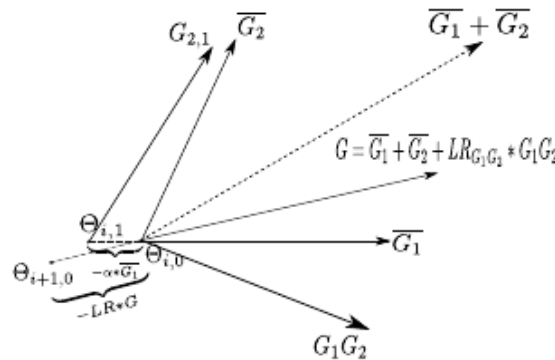
کدیا<sup>۲۳</sup> و همکاران الگوریتم حداکثر سازی نقطه-محصول فرا یادگیری (ام‌دات)<sup>۲۴</sup> را پیشنهاد دادند. این الگوریتم بر اساس ایده به حداکثر رساندن حاصل ضرب نقطه‌ای بین گرادین‌های مدل در نقاط مختلف آموزش با استفاده از تکنیکی به نام تفاوت‌های محدود<sup>۲۵</sup> است. این الگوریتم از نظر محاسباتی کارآمد است و می‌تواند برای مدل‌های بزرگ مانند برت اعمال شود و سربار محاسباتی را کاهش بدهد [۲۲]. عملکرد مناسب مدل پگاسوس در خلاصه‌سازی متون باعث شده بهترین مدل خلاصه‌سازی متون کوتاه مبتنی بر مدل پگاسوس و تکنیک تنظیم<sup>۲۶</sup> ام‌دات باشد.

<sup>23</sup>Kedia

<sup>24</sup>Meta-Learned Dot-Product Maximization (MDot)

<sup>25</sup>finite differences

<sup>26</sup>regularization



شکل ۳-۱۰: الگوریتم امدات [۲۲]

محاسبه گرادیان برای به حداکثر رساندن محصول نقطه‌ای با استفاده از تقریب تفاضل محدود، و استفاده از آن برای تنظیم گرادیان استاندارد [۲۲].

### ۳-۲-۱ ایده‌های ارائه شده بهبود خلاصه‌سازی متون طولانی

یکی از مشکلات مدل ترنسفورمر 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<sup>۲۷</sup> برای حل چالش‌های محاسباتی مرتبط با پردازش دنباله‌های طولانی متن ارائه شده است. لایه‌های برگشت‌پذیر<sup>۲۸</sup> 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 بخش‌های کوچک تر ورودی به طور مستقل استفاده می‌کند که موازی‌سازی را ممکن می‌کند و مصرف حافظه را کاهش می‌دهد. یکی از کمک‌های کلیدی آن استفاده از درهم‌سازی حساس به مکان در مکانیسم توجه<sup>۲۹</sup> است. درهم‌سازی حساس به مکان با توجه به زیرمجموعه‌ای از نشانه‌ها بر اساس مقادیر هش آنها، محاسبه توجه کامل را تقریب می‌زند، که منجر به محاسبه توجه کارآمدتر می‌شود. علاوه بر این، ریفورمر از کدگذاری‌های موقعیت محوری برای کدگذاری اطلاعات موقعیت توکن‌ها 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<sup>۳۰</sup> 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

<sup>۲۷</sup>Reformer

<sup>۲۸</sup>reversible layers

<sup>۲۹</sup>locality-sensitive hashing (LSH)

<sup>۳۰</sup>sparse



پیچیدگی محاسباتی را کاهش می‌دهد و مدل می‌تواند توالی‌های طولانی‌تر را مدیریت کند [۴]. مشابه شبکه‌ی ترنسفورمر پراکنده مدل بیگ‌برد نیز <sup>۳۱</sup> با استفاده از مکانیزم توجه پراکنده <sup>۳۲</sup> که وابستگی را به خطی کاهش می‌دهد و عملکرد ترنسفورمر را در مواجهه با توالی <sup>۳۳</sup> طولانی بهبود می‌بخشد. مدل بیگ‌برد نوآوری‌های دیگری مانند توجه جامع <sup>۳۴</sup> را معرفی می‌کند، که در آن توکن‌های خاص به تمام توکن‌های دیگر در دنباله توجه می‌کنند و وابستگی‌های دوربرد را به طور موثرتری به دست می‌آورند. همچنین شامل یک فرآیند پالایش تکراری است که وزن‌های توجه را برای بهبود عملکرد مدل اصلاح می‌کند [۲۷].

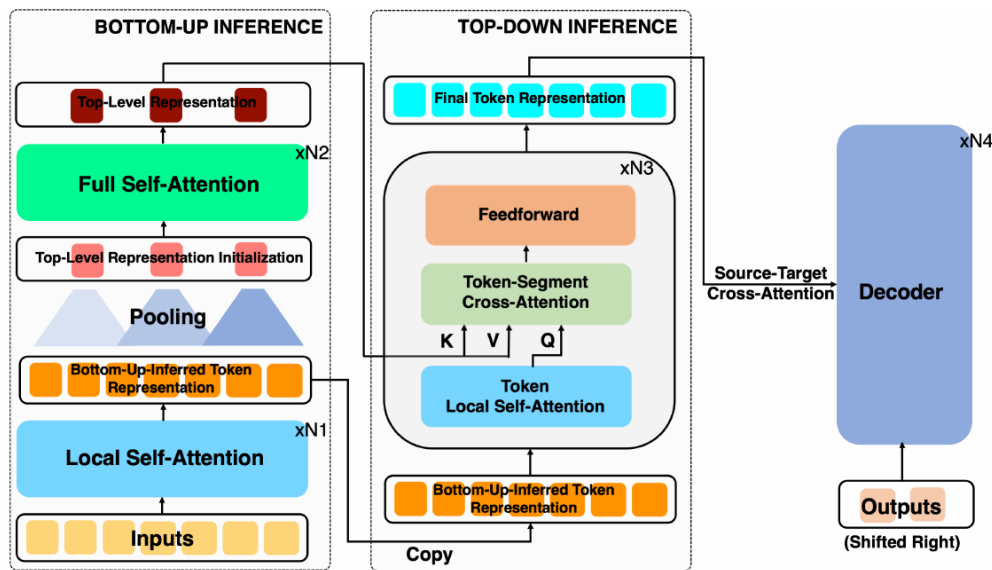
ادیت

در سال‌های اخیر مدل‌های مختلفی برای بهبود کیفیت خروجی مدل خلاصه‌سازی خودکار اسناد بلند ارائه شده است. به عنوان مثال پایل <sup>۳۵</sup> 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 <sup>۳۶</sup> 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 [۲۱]. پانگ <sup>۳۷</sup> 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 وابستگی‌های دوربرد و زمینه جامع، استنتاج از بالا به پایین برای نمایش‌های توکن اعمال می‌شود. 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 ۳-۱۱ نشان داده شده است. روش پیشنهادی یک رویکرد جدید امیدوارکننده برای خلاصه‌سازی اسناد طولانی است و نسبت به روش‌های قبلی کارآمدتر و موثرتر است [۹].

ادیت

جیدیوتیس و همکاران شیوه‌ی تقسیم و غلبه (دنسر) <sup>۳۸</sup> را برای بهبود خلاصه‌سازی اسناد طولانی پیشنهاد کرده‌اند. 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 ساختار بندی شده‌اند.

<sup>31</sup> Big Bird<sup>32</sup> Sparse attention<sup>33</sup> sequence<sup>34</sup> global attention<sup>35</sup> Pilault<sup>36</sup> GPT-2<sup>37</sup> Pang<sup>38</sup> Divide-and-Conquer (DANCER)



شکل ۳-۱۱: معماری مدل ترنسفورمر از بالا به پایین [۹]

برای مطابقت هر قسمت از خلاصه با بخشی از سند در دنسر از معیار روژ<sup>۳۹</sup> استفاده می‌شود. در این روش معیار روژ-ال بین هر یک از جملات خلاصه و تمام جملات سند محاسبه می‌شود و هر جمله‌ی 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 می‌کنیم 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، نمونه‌های آموزشی بسیار بیشتری ایجاد می‌کند. 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، بخش‌هایی مانند مرور ادبیات یا پیشینه در تلاش برای خلاصه کردن نکات اصلی مقاله ضروری نیستند و باعث افزودن نویز می‌شوند. بنابراین از بخش مرور

<sup>39</sup>ROUGE

ادبیات صرف نظر می‌شود و تمرکز سیستم خلاصه‌سازی فقط روی بخش‌های مقدمه، روش‌ها، نتایج و نتیجه‌گیری می‌باشد.

این مدل قابل ترکیب با پگاسوس یا مدل مولد نقطه‌ای<sup>۴۰</sup> می‌باشد. بخش کدگشا مدل مولد نقطه‌ای با ایجاد جملات تکراری مقابله می‌کند. هرچند ممکن است به خاطر تکرار اطلاعات در بخش‌های مختلف بازهم خلاصه‌ی تکراری ایجاد شود.

شیونگ و همکاران با اصلاح هدف بهینه‌سازی، معماری مدل‌های از پیش آموزش دیده و مجموعه‌ی دادگان پیش‌آموزش<sup>۴۱</sup> 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 را برای تطبیق مدل‌های متن به متن از پیش آموزش دیده برای دنباله‌های متن طولانی پیشنهاد می‌کند. این تکنیک‌ها عبارتند از:

- ارائه‌ی مدل براساس یک ترنسفورمر با مکانیزم توجه به خود پراکنده‌ی بلوکی<sup>۴۲</sup> در قسمت کدگذار است. این مکانیزم امکان استفاده‌ی مجدد از وزن‌های مدل‌های از پیش آموزش دیده را فراهم می‌کند.
- مکانیزم توکن سراسری<sup>۴۳</sup>: 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<sup>۴۴</sup>: توجه لغزشی با همپوشانی یک راه ساده برای معرفی اتصالات دوربرد در مدل‌های 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<sup>۴۵</sup>: این لایه به عنوان جایگزین لایه خود توجهی برای اتصالات دوربرد معرفی شده است. این رویکرد به واحدهای توجه درون بلوک‌ها اجازه می‌دهد تا به جای توجه به همسایگان بلافاصل خود، بر خلاصه‌ای از اطلاعات کلی در بلوک‌ها تمرکز کنند. این لایه در تصویر ۳-۱۲ نشان داده شده است. این مدل را قادر می‌سازد تا از اطلاعات گسترده تری در سراسر سند برای تصمیم‌گیری استفاده کند و وابستگی‌های دوربرد را در نظر

<sup>40</sup>Pointer-Generator model

<sup>41</sup>pretraining corpus

<sup>42</sup>Block-sparse self-attention

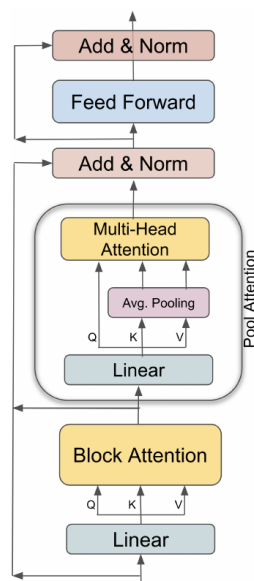
<sup>43</sup>Global-token mechanism

<sup>44</sup> Overlapping attention windows

<sup>45</sup>Pooling-augmented blockwise attention

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 مدل‌های متنی از پیش آموزش دیده را برای طیف وسیعی از وظایف پردازش زبان طبیعی ممکن می‌سازد [۲۵].



شکل ۳-۱۲: لایه خودتوجهی تقویت شده ادغام شده [۲۵]

## فصل چهارم

### روش‌های مبتنی بر یادگیری تقویتی

کارهای تحقیقاتی در زمینه یادگیری تقویتی<sup>۱</sup> و پردازش زبان طبیعی در سال‌های اخیر رشد کرده است. در یادگیری تقویتی یک عامل با محیط تعامل می‌کند و با آزمون و خطا، خط مشی بهینه را برای تصمیم‌گیری متوالی برای به حداکثر رساندن پاداش تجمعی آینده می‌آموزد. این پاداش می‌تواند یک معیار تعریف شده توسط توسعه دهنده بر اساس کار در حال حل باشد. در خلاصه‌سازی 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 خلاصه‌سازی 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 خلاصه‌ها در صورت افزایش طول مقالات منبع، با استفاده از سیستم‌های یادگیری تقویتی و یادگیری خط‌مشی<sup>۲</sup> بهبود یافته است. علاوه بر این مدل‌های دنباله به دنباله عمیق را نمی‌توان برای خلاصه کردن طیف گسترده‌ای 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 از گرادیان خط‌مشی انتقادی<sup>۳</sup> و ترکیب آن با یادگیری انتقالی<sup>۴</sup> برای انتقال دانش از یک مجموعه داده به مجموعه دیگر برطرف کنند [۱۰].

## ۴-۲ یادگیری تقویتی برای ترکیب خلاصه‌های استخراجی و انتزاعی

از یادگیری تقویتی برای ترکیب ویژگی‌های استخراجی با خلاصه انتزاعی برای استفاده از هر دو نوع خلاصه‌ی 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 ليو و همکاران یک چارچوب متخصص را پیشنهاد می‌کنند که مدل‌های انتزاعی و استخراجی را همزمان با استفاده از گرادیان خط‌مشی برای بهینه‌سازی مدل انتزاعی برای خلاصه‌ای با پاداش بالا، آموزش می‌دهد که منجر به خلاصه‌ای منسجم‌تر می‌شود [۱]. همچنین چن و بانسال یک مدل خلاصه‌سازی سریع پیشنهاد کردند که جملات برجسته را استخراج می‌کرد و سپس با استفاده از گرادیان خط‌مشی سطح جمله مبتنی بر یادگیری تقویتی بازنویسی می‌کرد [۳].

<sup>1</sup>reinforcement learning

<sup>2</sup> policy learning

<sup>3</sup>self-critic policy gradient

<sup>4</sup>Transfer Learning (TL)

## ۳-۴ یادگیری تقویتی برای ایجاد معیارها و پاداش‌های جدید

خلاصه‌سازی اسناد، مانند سایر کارهای مولد زبان، اغلب به دلیل استفاده از اهداف آموزشی مبتنی بر درست‌نمایی بیشینه<sup>۵</sup> 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 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 [۲۰]. معیارهای ارزیابی خلاصه‌سازی مانند روژ-۱<sup>۶</sup>، روژ-۲<sup>۷</sup>، روژ-ال<sup>۸</sup> و امتیازبرت<sup>۹</sup> به عنوان پاداش در رویکردهای یادگیری تقویتی استفاده شده است. با این حال، پارنل و همکاران استدلال می‌کنند که استفاده از امتیازات روژ به عنوان پاداش، جنبه‌های مهم خلاصه‌سازی، مانند خوانایی، روان بودن و اشتراک اطلاعات بین اسنادی در خلاصه‌سازی چند سندی را نادیده می‌گیرد و یک پاداش پوشش اصلاح شده همراه با یک برآوردگر گرادیان سیاست مبتنی بر اصول (ریلکس)<sup>۱۰</sup> را پیشنهاد می‌دهند [۱، ۲۰]. ریلکس یک برآوردگر گرادیان خط مشی<sup>۱۱</sup> با واریانس کم و بدون سوگیری<sup>۱۲</sup> است که برای مسائل یادگیری تقویتی با فضاهای کنش مداوم، مانند خلاصه‌سازی متن، مناسب است [۷].

در عبارت ۱-۴ تابع ضرر ارائه شده برحسب ریلکس نمایش داده شده است. بخش اول این عبارت سیاست را به تولید خروجی‌هایی با پاداش مورد انتظار بالا و بخش دوم به تولید خروجی‌های مشابه خروجی‌های قبلی تشویق می‌کند. در این عبارت  $r$  نشان دهنده‌ی پاداش  $c_\phi(\tilde{z})$  یک متغیر کنترلی از پارامترهای است که انتظار می‌رود با پاداش کاهش واریانس همبستگی داشته باشد.  $p(y_s)$  احتمال دنباله مشاهده شده خروجی  $y_s$  است.  $z$  دنباله نمونه‌های  $Gumbel - Softmax$  است.  $\tilde{z}$  دنباله ای از نمونه‌ها از یک توزیع  $Gumbel - Softmax$  مشروط بر  $y_s$  است.

$$L_{RELAX} = -[r - c_\phi(\tilde{z})] \log p(y^s) + c_\phi(z) - c_\phi(\tilde{z}) \quad (1-4)$$

یک

بار

دیگه

متغیرها

رو

چک

کن

<sup>5</sup>maximum likelihood<sup>6</sup>ROUGE-1<sup>7</sup>ROUGE-2<sup>8</sup>ROUGE-L<sup>9</sup>BERTScore<sup>10</sup>modified coverage reward along with a principled policy gradient estimator (RELAX)<sup>11</sup>policy<sup>12</sup>bias

## فصل پنجم

### نتایج



## فصل ششم

### جمع بندی

## 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 (ICCUBE)*, pages 1–7. IEEE, 2016.
- [3] Chen, Yen-Chun and Bansal, Mohit. Fast abstractive summarization with reinforcement 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] 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.
- [21] 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.
- [22] 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.

- [23] Su, Ming-Hsiang, Wu, Chung-Hsien, and Cheng, Hao-Tse. A two-stage transformer-based approach for variable-length abstractive summarization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2061–2072, 2020.
- [24] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, and Polosukhin, Illia. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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

Beel . . . . . بیل	۱
پ	
Episodic . . . . . پاسخ دوره‌ای به سوالات Question-Answering	ابر تکامل عصبی توپولوژی تقویت‌کننده Hyper Neuroevolution of Augmenting Topologies
Continues . . . . . پیوسته	ازدحام ذرات . . . . . Particle Swarm
ت	
Step Function . . . . . تابع قدم	Natural Language استنتاج زبان طبیعی Inference
Blurring . . . . . تارشدن	Evolutionary . . . . . الگوریتم تکاملی Algorithm
Switch . . . . . تعویض	Entropy . . . . . انتروپی
Symmetry . . . . . تقارن	Backpropagation . . . . . انتشار به عقب
تکامل عصبی توپولوژی‌های تقویت‌کننده Neuroevolution of Augmenting Topologies	ب
Repetition . . . . . تکرار	Data Driven . . . . . برپایه داده
Repetition with Variation تکرار با تغییر	Support Vector . . بردار ماشین پشتیبان Machine
Multi Headed Attention . توجه چندسر	Substrate . . . . . بستر
Hard Attention . . . . . توجه سخت	Naive Bayes . . . . . بیز ساده‌لوحانه
Soft Attention . . . . . توجه نرم	

Time Series . . . . . سری‌های زمانی	Sharpen . . . . . تیز کردن
ش	ج
Cosine Similarity . . شباهت کسینوسی	Shift . . . . . جابجایی
Multi Layer . شبکه پرسپترون چندلایه	Population . . . . . جمعیت
Perceptron	Mutation . . . . . جهش
Compositional شبکه تولید الگوی ترکیبی	خ
Pattern Producing Networks	Hubber Loss . . . . . خطای هوبر
Neural Network . . . . . شبکه عصبی	Rooted Mean . خطای ریشه مربعات خطا
Recurrent Neural شبکه عصبی بازگشتی	Square Error
Network	د
Gated . شبکه عصبی بازگشتی دروازه‌دار	Prior Knowledge . . . . . دانش پیشین
Recurrent Neural Network	Gate . . . . . دروازه
Feedforward Neural شبکه عصبی جلورو	Interpolation . . . . . درون‌یابی
Network	Batch . . . . . دسته
Memory Network . . . شبکه حافظه‌ای	ر
Cold Start . . . . . شروع سرد	Knowledge Tracing . . . . . ردیابی دانش
ص	رونوشت‌گیری
Accuracy . . . . . صحت	Copy . . . . . رونوشت‌گیری
ف	Repeat Copy . تکرار شونده
Sharpness Factor . . . . . فاکتور تیزی	ژ
Falcon . . . . . فالکن	ژائو Zhao . . . . .
ک	س
Differentiable . . کامپیوتر عصبی متمایز	سر Head . . . . .
Neural Computer	
Encoding . . . . . کدگذاری	
Controllor . . . . . کنترل‌گر	

Area Under Curve . . . . ناحیه زیرنمودار	Collier . . . . . کولیر
Generation . . . . . نسل	گ
Tape . . . . . نوار	Gulcehre . . . . . گالچره
و	Node . . . . . گره
Hidden State . . . . . وضعیت مخفی	Graves . . . . . گریوز
ه	Discrete . . . . . گسسته
Kernel . . . . . هسته	ل
ی	Logits . . . . . لاجیتس
Association Recall . . . یادآوری انجمنی	م
Sequence Learning . . . یادگیری ترتیبی	ماشین تورینگ عصبی . . Neural Turing Machine
Edge . . . . . یال	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



# واژه‌نامه‌ی انگلیسی به فارسی

A	رونوشت‌گیری . . . . . Copy
Accuracy . . . . . صحت	Cosine Similarity . . شباهت کسینوسی
Area Under Curve . . . . ناحیه زیرمنمودار	D
Association Recall . . . یادآوری انجمنی	Data Driven . . . . . برپایه داده
Attention Mechanism . . مکانیسم توجه	Differentiable . . کامپیوتر عصبی متمایز
B	Neural Computer
Backpropagation . . . . . انتشار به عقب	Discrete . . . . . گسسته
Batch . . . . . دسته	Dynamic . . . ماشین تورینگ عصبی پویا
Beel . . . . . بیل	Neural Turing Machine
Blurring . . . . . تارشدن	E
C	Edge . . . . . یال
Cold Start . . . . . شروع سرد	Episodic . . . . . پاسخ دوره‌ای به سوالات
Collier . . . . . کولیر	Question-Answering
Compositional شبکه تولید الگوی ترکیبی	Encoding . . . . . کدگذاری
Pattern Producing Networks	Entropy . . . . . انتروپی
Continues . . . . . پیوسته	Evolutionary . . . . . الگوریتم تکاملی
Controller . . . . . کنترل‌گر	Algorithm

ماشین تورینگ عصبی تکاملی Evolvable Neural Turing Machine	مقداردهی اولیه Initialization . . . . .
F	درون‌یابی Interpolation . . . . .
Falcon . . . . .	K
شبکه عصبی جلورو Feedforward Neural Network	هسته Kernel . . . . .
G	ردیابی دانش Knowledge Tracing . . . .
Gate . . . . .	$k$ -نزدیک‌ترین همسایه K-Nearest Neighbours
دروازه Gate . . . . .	L
شبکه عصبی بازگشتی دروازه‌دار Gated . . Recurrent Neural Network	لاجیتس Logits . . . . .
نسل Generation . . . . .	M
گریوز Graves . . . . .	شبکه حافظه‌ای Memory Network . . . .
گالچره Gulcehre . . . . .	توجه چندسر Multi Headed Attention .
H	شبکه پرسپترونی چندلایه Multi Layer . Perceptron
توجه سخت Hard Attention . . . . .	جهش Mutation . . . . .
سر Head . . . . .	N
وضعیت مخفی Hidden State . . . . .	بیز ساده‌لوحانه Naive Bayes . . . . .
خطای هوبر Hubber Loss . . . . .	استنتاج زبان طبیعی Natural Language Inference
ماشین تورینگ عصبی ابر تکاملی Hyper . . Evolvable Neural Turing Machine	شبکه عصبی Neural Network . . . . .
ابر تکامل عصبی توپولوژی تقویت‌کننده Hyper Neuroevolution of Augmenting Topologies	ماشین تورینگ عصبی Neural Turing . . Machine
I	تکامل عصبی توپولوژی‌های تقویت‌کننده Neuroevolution of Augmenting Topologies
	گره Node . . . . .
	O

One-hot . . . . . تک‌روشن	Sequence Learning . . . یادگیری ترتیبی
Open Source . . . . . منبع‌باز	Sharpen . . . . . تیزکردن
P	Sharpness Factor . . . . . فاکتور تیزی
Particle Swarm . . . . . ازدحام ذرات	Shift . . . . . جابجایی
Population . . . . . جمعیت	Similarity Metric . . . . . معیار شباهت
Prior Knowledge . . . . . دانش پیشین	Soft Attention . . . . . توجه نرم
R	Substrate . . . . . بستر
Recurrent Neural Network شبکه عصبی بازگشتی	Support Vector Machine بردار ماشین پشتیبان
Regularity . . . . . منظم‌سازی	Step Function . . . . . تابع قدم
Repeat Copy . رونوشت‌گیری تکرارشونده	Switch . . . . . تعویض
Repetition . . . . . تکرار	Symmetry . . . . . تقارن
Repetition with Variation تکرار با تغییر	T
Rooted Mean Square Error خطای ریشه مربعات خطا	Tape . . . . . نوار
S	Time Series . . . . . سری‌های زمانی
Scalability . . . . . مقیاس‌پذیری	Z
	Zhao . . . . . ژائو