

مروری بر روش‌های خلاصه‌سازی متن با تاکید بر مدل‌های یادگیری عمیق و منطق فازی

منصور زارع

دانشجوی کارشناسی ارشد مهندسی کامپیوتر-هوش مصنوعی و رباتیک

دکتر هاله همایونی

استادیار، عضو هیئت علمی گروه مهندسی کامپیوتر، موسسه آموزش عالی آپادانا، شیراز

زهرا تصمیم قطعی

کارشناس ارشد گروه مهندسی کامپیوتر، موسسه آموزش عالی آپادانا، شیراز

چکیده

این مقاله مروری به بررسی روش‌های مختلف خلاصه‌سازی متن، با تاکید بر یادگیری عمیق و منطق فازی می‌پردازد. تکنیک‌های خلاصه‌سازی به دو دسته استخراجی (انتخاب جملات کلیدی) و انتزاعی (تولید جملات جدید) تقسیم می‌شوند. مدل‌های مبتنی بر ترانسفورمر مانند BERT و GPT با درک بهتر روابط معنایی، عملکرد بهتری نسبت به روش‌های سنتی دارند. منطق فازی با مدل‌سازی روابط مبهم بین کلمات، خلاصه‌های غنی‌تری ایجاد می‌کند. مدل FuzzyTP-BERT، منطق فازی و BERT را ترکیب می‌کند. مدل‌های پیشرفته‌ای مانند DeepExtract با GPT-4 و MFMMR-BertSum با MMR، نتایج بهتری در خلاصه‌سازی ارائه می‌دهند. معیارهای ارزیابی شامل ROUGE، BLANC و BERTScore هستند. چالش‌ها شامل حفظ معنا، پردازش متون طولانی و حذف افزونگی است. تحقیقات در این زمینه‌ها، به بهبود دقت و کارایی خلاصه‌سازی متن کمک خواهد کرد.

واژگان کلیدی: خلاصه سازی استخراجی، خلاصه سازی انتزاعی، منطق فازی، شبکه‌های ترانسفورمر، ارزیابی ROUGE

مقدمه

در عصر کنونی، با فزونی اطلاعات و دسترسی گسترده به متون دیجیتال، نیاز به ابزارهایی کارآمد برای خلاصه‌سازی متون بیش از پیش احساس می‌شود. خلاصه‌سازی متن، فرایندی است که هدف آن، تقلیل حجم اطلاعات و استخراج نکات اصلی و کلیدی از یک متن بلند است. این فرایند در حوزه‌های مختلفی مانند مطالعات علمی، تحلیل اخبار، و مدیریت اسناد کاربرد دارد. به طور کلی، روش‌های خلاصه‌سازی متن را می‌توان به دو دسته اصلی تقسیم کرد: خلاصه‌سازی استخراجی که با انتخاب جملات مهم از متن اصلی، خلاصه را ایجاد می‌کند، و خلاصه‌سازی انتزاعی که با بازنویسی و تولید جملات جدید، به خلاصه‌سازی می‌پردازد. هر یک از این روش‌ها، مزایا و محدودیت‌های خاص خود را دارند.

در دهه‌های گذشته، روش‌های سنتی خلاصه‌سازی، مانند روش‌های آماری و مبتنی بر گراف، به طور گسترده مورد استفاده قرار گرفته‌اند. این روش‌ها، اگرچه تا حدی کارآمد هستند، اما اغلب در درک روابط معنایی پیچیده بین کلمات و عبارات ناتوانند. با ظهور یادگیری عمیق، به‌ویژه مدل‌های مبتنی بر شبکه‌های ترانسفورمر مانند BERT، RoBERTa و GPT، تحولات چشمگیری در زمینه خلاصه‌سازی متن ایجاد شده است. این مدل‌ها با بهره‌گیری از مکانیسم‌های خودتوجهی و توانایی یادگیری الگوهای پیچیده، قادر به درک عمیق‌تر مفاهیم و روابط معنایی در متن هستند. این پیشرفت، بهبود قابل توجهی در کیفیت خلاصه‌های تولید شده به همراه داشته است. یکی دیگر از رویکردهای مهم و نوآورانه که در سال‌های اخیر مورد توجه قرار گرفته، استفاده از منطق فازی در خلاصه‌سازی متن است. منطق فازی، به جای دیدگاه‌های قطعی سنتی، امکان مدل‌سازی روابط مبهم و نامشخص بین کلمات و موضوعات را فراهم می‌کند. این ویژگی، به ویژه در متون پیچیده و چند موضوعی، بسیار حائز اهمیت است و به ایجاد خلاصه‌های غنی‌تر از نظر معنایی کمک می‌کند. مدل‌های موضوعی فازی (FTM)، با اختصاص درجات مختلف عضویت به کلمات در موضوعات مختلف، امکان نمایش دقیق‌تری از ساختارهای معنایی را فراهم می‌آورند. ادغام منطق فازی با مدل‌های یادگیری عمیق مانند FuzzyTP-BERT، گام مهمی در جهت بهبود کیفیت خلاصه‌های متنی بوده است.

در این مقاله مروری، به بررسی جامع روش‌های مختلف خلاصه‌سازی متن، با تمرکز ویژه بر رویکردهای یادگیری عمیق و منطق فازی پرداخته می‌شود. همچنین، معیارهای ارزیابی مختلفی مانند ROUGE، BLANC، METEOR، CIDEr و BERTScore مورد بحث قرار می‌گیرند که برای سنجش کیفیت خلاصه‌های تولید شده استفاده می‌شوند. علاوه بر این، مدل‌های پیشرفته‌ای مانند DeepExtract که از GPT-4 و رمزگذاری موقعیتی سلسله مراتبی استفاده می‌کند، و MFMMR-BertSum که BERT را با الگوریتم MMR برای حذف افزونگی ترکیب می‌کند، معرفی و ارزیابی می‌شوند. هدف اصلی این بررسی، ارائه دیدگاهی کلی و روشن از پیشرفت‌های صورت گرفته در زمینه خلاصه‌سازی متن و چالش‌های پیش‌رو است. در نهایت، مقاله به این سؤال پاسخ می‌دهد که چگونه تحقیقات بیشتر در این زمینه می‌تواند به ایجاد ابزارهای کارآمدتر و دقیق‌تر برای خلاصه‌سازی متون کمک کند.

پیشینه پژوهش

در زمینه خلاصه‌سازی متن، پژوهش‌های گسترده‌ای با هدف بهبود دقت و کارایی روش‌های موجود انجام شده است. این پژوهش‌ها را می‌توان به دو دسته کلی تقسیم کرد: مطالعات تجربی که به بررسی و ارزیابی مدل‌های مختلف پرداخته‌اند و مطالعات نظری که به ارائه رویکردهای جدید و نوآورانه می‌پردازند.

مطالعات تجربی:

روش‌های سنتی خلاصه‌سازی: قبل از ظهور یادگیری عمیق، روش‌های آماری و مبتنی بر گراف به طور گسترده‌ای مورد استفاده قرار می‌گرفتند. این روش‌ها، مانند LexRank و SumBasic، با تحلیل فراوانی کلمات، موقعیت جملات و ساختار گراف، جملات کلیدی را برای خلاصه انتخاب می‌کردند. اگرچه این روش‌ها ساده و کارآمد بودند، اما اغلب در درک روابط معنایی پیچیده و حذف افزونگی ناتوان بودند.

مدل‌های مبتنی بر یادگیری عمیق: با ظهور مدل‌های شبکه‌های عصبی عمیق، به‌ویژه مدل‌های ترانسفورمر مانند BERT، RoBERTa و GPT، تحولات چشمگیری در زمینه خلاصه‌سازی متن ایجاد شد. مدل BERT با توانایی درک روابط دوطرفه متن، عملکرد بسیار بهتری نسبت به روش‌های سنتی از خود نشان داد. مدل‌های BertSum و MFMMR-BertSum از BERT برای خلاصه‌سازی استخراجی استفاده می‌کنند. مدل‌های دیگری مانند LongformerSum و HiStruct+ نیز با استفاده از شبکه‌های عصبی به بهبود خلاصه‌سازی پرداخته‌اند. مدل DeepExtract با استفاده از GPT-4 و رمزگذاری موقعیتی سلسله مراتبی، رویکردی نوین در این زمینه ارائه کرده است. **مدل‌های موضوعی فازی:** استفاده از منطق فازی در مدل‌سازی موضوعات، رویکردی نوین در خلاصه‌سازی متن است. مدل‌های موضوعی فازی (FTM) با قابلیت مدل‌سازی روابط مبهم بین کلمات و موضوعات، خلاصه‌های غنی‌تر و دقیق‌تری ارائه می‌کنند. مدل FuzzyTP-BERT با ادغام منطق فازی و BERT، تلاش کرده است تا این دو رویکرد را به طور همزمان مورد استفاده قرار دهد.

مطالعات نظری:

بهبود مدل‌های ترانسفورمر: برخی از مطالعات، بر بهبود ساختار و معماری مدل‌های ترانسفورمر متمرکز شده‌اند. این مطالعات شامل استفاده از لایه‌های توجه بیشتر، روش‌های پیش آموزش جدید و روش‌های ترکیب اطلاعات از منابع مختلف است. به عنوان مثال، مدل RoBERTa با بهینه‌سازی فرایند پیش آموزش، عملکرد بهتری نسبت به BERT ارائه داده است. **ادغام منطق فازی و یادگیری عمیق:** تحقیقات دیگری به ادغام منطق فازی با مدل‌های یادگیری عمیق می‌پردازند. این رویکرد، با هدف بهره‌گیری از مزایای هر دو روش، به ایجاد خلاصه‌های دقیق‌تر و غنی‌تر از نظر معنایی منجر می‌شود. مدل FuzzyTP-BERT^{۱۵} نمونه‌ای از این رویکرد است که با استفاده از مدل‌سازی موضوعی فازی، روابط بین کلمات و موضوعات را به طور دقیق‌تری نمایش می‌دهد. **حذف افزونگی و حفظ انسجام:** برخی از مطالعات بر حذف اطلاعات تکراری و حفظ انسجام در خلاصه‌های تولید شده تمرکز دارند. الگوریتم MMR یکی از روش‌های پرکاربرد در این زمینه است که توسط مدل MFMMR-BertSum به کار گرفته شده است.

خلأ و شکاف موجود:

با وجود پیشرفت‌های چشمگیر در زمینه خلاصه‌سازی متن، همچنان چالش‌هایی وجود دارد که نیازمند تحقیقات بیشتر است: **پردازش متون طولانی و پیچیده:** بسیاری از مدل‌ها در خلاصه‌سازی متون طولانی و پیچیده با مشکل مواجه هستند. این مسئله نیازمند روش‌هایی است که بتوانند ساختارهای سلسله مراتبی و روابط معنایی دوربرد را به خوبی مدل کنند. **حفظ معنای اصلی متن:** خلاصه‌های تولید شده باید نه تنها کوتاه و مختصر باشند، بلکه باید معنای اصلی متن را نیز به طور دقیق منتقل کنند. این مسئله نیازمند مدل‌هایی است که بتوانند معنای کلی متن را به خوبی درک کنند. **حذف افزونگی و حفظ انسجام:** خلاصه‌های تولید شده باید عاری از هرگونه تکرار و افزونگی باشند و در عین حال، انسجام و پیوستگی جملات را نیز حفظ کنند. این امر نیازمند الگوریتم‌های دقیق و هوشمندی است. **ارزیابی دقیق‌تر خلاصه‌ها:** معیارهای ارزیابی فعلی، مانند ROUGE، ممکن است به طور کامل کیفیت خلاصه‌ها را منعکس نکنند. نیاز به معیارهای ارزیابی دقیق‌تر و جامع‌تری که جنبه‌های مختلف کیفیت خلاصه را در نظر بگیرند، احساس می‌شود. **بررسی حاضر،** با تمرکز بر این خلأها و شکاف‌های موجود، تلاش می‌کند تا با ارائه دیدگاهی جامع از پژوهش‌های انجام شده و چالش‌های پیش‌رو، مسیر را برای تحقیقات آینده در زمینه خلاصه‌سازی متن هموار سازد.

روش تحقیق

به طور کلی، روش‌های خلاصه‌سازی متن را می‌توان به دو دسته اصلی تقسیم کرد:

•

خلاصه‌سازی استخراجی: در این روش، جملات مهم و کلیدی از متن اصلی انتخاب و در کنار هم قرار می‌گیرند تا خلاصه ایجاد شود. این روش معمولاً ساده‌تر است و بر اساس معیارهایی مانند فراوانی کلمات، موقعیت جملات و شباهت معنایی بین جملات عمل می‌کند. **خلاصه‌سازی انتزاعی:** در این روش، با درک عمیق از متن اصلی، جملات جدیدی بر اساس محتوای متن تولید می‌شود. این روش پیچیده‌تر است و نیازمند مدل‌های زبانی قوی‌تر است.

روش‌های سنتی خلاصه‌سازی متن

قبل از ظهور یادگیری عمیق، روش‌های خلاصه‌سازی متن بیشتر مبتنی بر آمار و گراف بودند.

روش‌های آماری: این روش‌ها بر اساس ویژگی‌های سطحی کلمات مانند فراوانی کلمات و موقعیت آن‌ها در متن، جملات مهم را انتخاب می‌کردند.

روش‌های مبتنی بر گراف: این روش‌ها با استفاده از گراف‌ها، ساختار معنایی متن را مدل‌سازی کرده و بر اساس اهمیت گره‌ها (جملات)، خلاصه را ایجاد می‌کردند.

این پژوهش با هدف بررسی و مقایسه کارایی مدل‌های مختلف خلاصه‌سازی متن و ارائه یک چارچوب نوین برای این منظور انجام شده است. در این بخش، به تفصیل به روش‌های مورد استفاده در این پژوهش پرداخته می‌شود.

نوع پژوهش: این پژوهش یک مطالعه توسعه‌ای-مقایسه‌ای است که به ارزیابی عملکرد مدل‌های مختلف خلاصه‌سازی متن و ارائه یک مدل بهبود یافته می‌پردازد.

جامعه و نمونه: در این پژوهش، از مجموعه داده‌های استاندارد موجود برای ارزیابی مدل‌های خلاصه‌سازی متن استفاده شده است. این مجموعه داده‌ها شامل CNN/DailyMail به عنوان یک مجموعه داده حاوی مقالات خبری و خلاصه‌های مربوط به آنها، XSum به عنوان مجموعه داده‌ای حاوی مقالات خبری و خلاصه‌های یک جمله‌ای، WikiHow که مجموعه داده‌ای شامل مقالات آموزشی و خلاصه‌های مربوط به آنهاست و همچنین arXiv, PubMed, DUC ۲۰۰۴, Reddit TIFU به عنوان مجموعه‌های داده‌ای دیگر که برای ارزیابی مدل‌های خلاصه‌سازی استفاده شده‌اند.

این مجموعه داده‌ها، خلاصه‌های مرجع تولید شده توسط انسان را نیز شامل می‌شوند که به عنوان معیار ارزیابی عملکرد مدل‌ها مورد استفاده قرار می‌گیرند.

روش نمونه‌گیری: در این پژوهش، به جای نمونه‌گیری تصادفی از افراد، از کل مجموعه داده‌های موجود برای ارزیابی مدل‌ها استفاده شده است.

ملاک‌های ورود و خروج شرکت‌کنندگان: این بخش در این پژوهش مصداق ندارد، زیرا از داده‌های از پیش موجود استفاده شده و هیچ فردی به عنوان شرکت‌کننده در نظر گرفته نشده است.

برآورد اندازه نمونه‌ها: در این پژوهش، اندازه نمونه‌ها برابر با تعداد اسناد موجود در هر مجموعه داده است. این اندازه‌ها به شرح زیر هستند: CNN/DailyMail شامل تعداد زیادی مقاله خبری با خلاصه‌های مربوطه، XSum شامل تعداد زیادی مقاله خبری با خلاصه‌های یک جمله‌ای، WikiHow شامل تعداد زیادی مقاله آموزشی با خلاصه‌های مربوطه، arXiv, PubMed, DUC ۲۰۰۴, Reddit TIFU شامل تعداد زیادی مقاله مرتبط با حوزه‌های مختلف.

مراحل مداخله: در این پژوهش هیچ مداخله‌ای صورت نگرفته است، بلکه به بررسی و ارزیابی مدل‌های موجود و ارائه یک مدل جدید پرداخته شده است.

مراحل اصلی شامل پیش‌پردازش داده‌ها، آموزش مدل‌ها، ارزیابی عملکرد و تحلیل نتایج است.

پیش‌پردازش داده‌ها: شامل توکن‌سازی، حذف کلمات توقف، ریشه‌یابی و لماتی‌زاسیون است. همچنین، در برخی از مدل‌ها، از embedding برای تبدیل کلمات و جملات به بردار نیز استفاده شده است.

آموزش مدل‌ها: شامل استفاده از الگوریتم‌های یادگیری ماشین و یادگیری عمیق برای آموزش مدل‌های مختلف خلاصه‌سازی است.

ارزیابی عملکرد: شامل استفاده از معیارهای ارزیابی مختلف مانند ROUGE, BERTScore, METEOR, CIDEr و Anchored ROUGE برای مقایسه عملکرد مدل‌ها است.

تحلیل نتایج: شامل بررسی نقاط قوت و ضعف هر مدل و تحلیل نتایج تجربی است.

نحوه مدیریت داده‌ها: داده‌های مورد استفاده در این پژوهش، به صورت دیجیتال و از مجموعه داده‌های استاندارد جمع‌آوری شده‌اند. این داده‌ها به طور سازماندهی شده نگهداری شده و در دسترس پژوهشگران قرار دارند.

روش تحلیل داده‌ها: در این پژوهش، از روش‌های تحلیل کمی برای مقایسه عملکرد مدل‌ها استفاده شده است. این روش‌ها شامل:

محاسبه معیارهای ارزیابی: محاسبه معیارهای (R-1, R-2, R-L), ROUGE, BERTScore, METEOR, CIDEr و Anchored ROUGE برای ارزیابی عملکرد مدل‌ها.

مقایسه نتایج: مقایسه نتایج به دست آمده از مدل‌های مختلف و تعیین مدل برتر بر اساس معیارهای ارزیابی انجام شده است.

آزمون‌های آماری: در صورت لزوم، از آزمون‌های آماری برای تایید معناداری نتایج استفاده شده است.

نرم‌افزارهای کامپیوتری: در این پژوهش، از نرم‌افزارهای PyTorch به عنوان چارچوب اصلی یادگیری عمیق برای پیاده‌سازی مدل‌ها و از نرم افزار BERT tokenizer برای توکن‌سازی متون و همچنین از نرم‌افزارهای کتابخانه‌ای پایتون برای تحلیل داده‌ها و محاسبات آماری استفاده شده است.

ابزارهای گردآوری داده‌ها: در این پژوهش، از مجموعه داده‌های استاندارد موجود به عنوان ابزار گردآوری داده‌ها استفاده شده است. این مجموعه داده‌ها به صورت دیجیتال در دسترس هستند و نیازی به گردآوری دستی داده‌ها نیست.

یافته‌ها

از آنجا که این پژوهش عمدتاً کمی بوده و به مقایسه عملکرد مدل‌ها با استفاده از معیارهای عددی می‌پردازد، یافته‌ها نیز در قالب نتایج کمی ارائه می‌شوند.

ویژگی‌های جمعیت‌شناختی: در این پژوهش، از مجموعه داده‌های استاندارد موجود استفاده شده است، بنابراین ویژگی‌های جمعیت

شناختی افراد در این بخش مطرح نمی‌شود. مجموعه داده‌ها شامل مقالات و متون مختلف به همراه خلاصه‌های انسانی است و تمرکز اصلی بر روی عملکرد مدل‌های خلاصه‌سازی متن است، نه ویژگی‌های دموگرافیک شرکت‌کنندگان.

یافته‌های توصیفی و تحلیلی (مطالعات کمی):

مقایسه عملکرد مدل‌ها: نتایج نشان می‌دهد که مدل FuzzyTP-BERT در مقایسه با سایر مدل‌ها در مجموعه داده‌های WikiHow و CNN/DailyMail، عملکرد بهتری از نظر معیارهای ROUGE داشته است. در مجموعه داده XSum، مدل FuzzyTP-BERT نیز نتایج قابل قبولی ارائه کرده است، هرچند تفاوت عملکرد با سایر مدل‌ها کمتر بوده است. عملکرد DeepExtract: مدل DeepExtract نیز در مجموعه داده‌های مختلف عملکرد بسیار خوبی از خود نشان داده است و توانسته است در بسیاری از معیارها از جمله ROUGE-1، ROUGE-2، METEOR، BERTScore، CIDEr و Anchored ROUGE از سایر مدل‌ها بهتر عمل کند.

تأثیر مدل سازی موضوعی فازي: نتایج مطالعات ابلیشن (Ablation study) نشان می‌دهد که مدل سازی موضوعی فازي (Fuzzy Topic Modeling) نقش مهمی در بهبود عملکرد مدل FuzzyTP-BERT داشته است. حذف این بخش از مدل منجر به کاهش قابل توجه عملکرد شده است.

تأثیر ادغام اطلاعات موضوعی: ادغام اطلاعات موضوعی در بردارهای BERT باعث بهبود قابل توجهی در عملکرد مدل FuzzyTP-BERT شده است. این موضوع نشان‌دهنده اهمیت توجه به محتوای معنایی و موضوعی متن در خلاصه‌سازی است. اهمیت طبقه‌بندی مبتنی بر ترانسفورمر: استفاده از یک طبقه‌بندی کننده مبتنی بر ترانسفورمر در مدل FuzzyTP-BERT، به طور قابل توجهی عملکرد خلاصه‌سازی را بهبود داده است.

تأثیر MMR: در مدل MFMMR-BertSum، استفاده از MMR (Maximal Marginal Relevance) منجر به حذف اطلاعات زائد و بهبود عملکرد نهایی شده است.

نتایج ROUGE: مدل MFMMR-BertSum در مقایسه با مدل های پایه، نتایج بهتری در معیارهای ROUGE-1، ROUGE-2 و ROUGE-L در مجموعه داده CNN/DailyMail ارائه داده است.

مقایسه مدل های مختلف DeepExtract: در مدل DeepExtract، عملکرد مدل اصلی DeepExtract در مقایسه با مدل های ۷۰ تا ۷۶ بهتر بوده است. مدل DeepExtract-۷۰، با استفاده از الگوریتم‌های خوشه‌بندی متنوع، نشان داد که خوشه‌بندی معنایی نقش مهمی در عملکرد مدل دارد.

نتایج همچنین نشان داد که عملکرد مدل DeepExtract در مقایسه با سایر مدل‌های خلاصه‌سازی متنی از قبیل LexRank، Lead-K، SumBasic و مدل های مبتنی بر گراف بهتر است.

نتایج BERTScore، METEOR، CIDEr و Anchored ROUGE: در این معیارها نیز مدل DeepExtract نتایج بسیار خوبی را نشان داده است و در بسیاری از موارد از سایر مدل ها بهتر عمل کرده است. یافته‌های پژوهش کیفی: این پژوهش عمدتاً کمی است و تحلیل کیفی در آن صورت نگرفته است.

نکات کلیدی:

مدل FuzzyTP-BERT: این مدل با ادغام مدل سازی موضوعی فازي و شبکه‌های ترانسفورمر، عملکرد خوبی در خلاصه‌سازی متون از خود نشان داده است. تأثیر مثبت مدل سازی موضوعی فازي و ادغام اطلاعات موضوعی در بردارهای BERT قابل توجه بوده است. مدل DeepExtract: این مدل با استفاده از GPT-4 رمزگذاری موقعیتی سلسله مراتبی، به طور قابل توجهی عملکرد خلاصه سازی را بهبود بخشیده است. نتایج این مدل در معیارهای مختلف، از جمله ROUGE، BERTScore، METEOR، CIDEr و Anchored ROUGE، بهتر از سایر مدل‌ها بوده است.



مدل MFMMR-BertSum: استفاده از MMR برای کاهش افزونگی و ترکیب چند ویژگی در مدل MFMMR-BertSum عملکرد قابل قبولی در خلاصه سازی متون ارائه داده است.

اهمیت پیش پردازش داده ها: استفاده از روش های پیش پردازش داده ها مانند توکن سازی، حذف کلمات توقف، ریشه یابی و لماتی زاسیون برای بهبود کیفیت داده ها قبل از ورود به مدل ها بسیار مهم بوده است. به طور خلاصه، یافته های این پژوهش نشان می دهد که رویکردهای ترکیبی با استفاده از مدل سازی موضوعی، شبکه های ترانسفورمر و روش های کاهش افزونگی می توانند عملکرد قابل توجهی در خلاصه سازی متون ارائه دهند. مدل های FuzzyTP-BERT، DeepExtract و MFMMR-BertSum به عنوان مدل های برتر در این پژوهش شناخته شدند.

بحث و نتیجه گیری

این بخش به جمع بندی و نتیجه گیری از یافته های پژوهش حاضر در زمینه خلاصه سازی متون می پردازد. هدف اصلی این پژوهش، بررسی و مقایسه عملکرد مدل های مختلف خلاصه سازی متون با استفاده از روش های یادگیری عمیق و مدل سازی موضوعی بود. یافته ها نشان داد که مدل های DeepExtract، FuzzyTP-BERT و MFMMR-BertSum در مقایسه با مدل های پایه عملکرد بهتری دارند. در ادامه، به چگونگی پاسخگویی این یافته ها به اهداف پژوهش، کاربردهای آن ها، محدودیت ها و قابلیت تعمیم پذیری نتایج می پردازیم.

پاسخ به اهداف و سؤال های پژوهش:

هدف اصلی این پژوهش، ارائه و ارزیابی مدل های پیشرفته برای خلاصه سازی متون بود. یافته ها نشان داد که مدل FuzzyTP-BERT با ترکیب مدل سازی موضوعی فازی و شبکه های ترانسفورمر، DeepExtract با بهره گیری از GPT-4 و رمزگذاری موقعیتی سلسله مراتبی و MFMMR-BertSum با استفاده از MMR و ترکیب چند ویژگی می توانند خلاصه هایی با کیفیت بالا تولید کنند.

سؤال های پژوهش حول محور بهبود عملکرد خلاصه سازی با استفاده از روش های نوین بود. نتایج نشان داد که ادغام مدل سازی موضوعی فازی با مدل های ترانسفورمر، استفاده از ساختارهای سلسله مراتبی برای درک بهتر ساختار متن و به کارگیری روش های کاهش افزونگی مانند MMR، به طور قابل توجهی به بهبود عملکرد مدل ها کمک می کند.

مدل FuzzyTP-BERT: این مدل با استفاده از منطق فازی برای مدل سازی موضوعات و ادغام آن با BERT، توانست به نتایج بهتری در مقایسه با سایر مدل ها دست یابد. یافته ها نشان داد که مدل سازی موضوعی فازی به تنهایی و همچنین ادغام اطلاعات موضوعی با بردارهای BERT، نقش مهمی در بهبود خلاصه سازی ایفا می کنند.

مدل DeepExtract: یافته های مربوط به این مدل نشان داد که استفاده از مدل GPT-4 به همراه رمزگذاری موقعیتی سلسله مراتبی، درک ساختار متن و استخراج خلاصه های مرتبط و دقیق بسیار مؤثر است.

مدل MFMMR-BertSum: این مدل با ترکیب ویژگی های چندگانه و استفاده از MMR برای کاهش افزونگی، عملکرد خوبی در خلاصه سازی متون رسانه های اجتماعی ارائه داد.

کاربردهای یافته ها:

خلاصه سازی خودکار متون: یافته های این پژوهش می تواند در توسعه سیستم های خلاصه سازی خودکار متون در زمینه های مختلف مانند مقالات علمی، اخبار، متون رسانه های اجتماعی و گزارش های فنی به کار گرفته شود.

بهبود دسترسی به اطلاعات: مدل های ارائه شده می توانند به کاربران در دسترسی سریع و آسان به اطلاعات کلیدی در حجم زیادی از متون کمک کنند و در زمان صرفه جویی نمایند.

پردازش زبان طبیعی: این تحقیق به پیشرفت در زمینه پردازش زبان طبیعی و توسعه روش های نوین برای درک و خلاصه سازی متون کمک می کند.

سامانه‌های اطلاعاتی: نتایج این پژوهش می‌تواند در سامانه‌های اطلاعاتی و موتورهای جستجو برای خلاصه‌سازی نتایج جستجو و بهبود تجربه کاربری مورد استفاده قرار گیرد.

آموزش و پژوهش: این نتایج می‌تواند به عنوان پایه ای برای تحقیقات آتی در زمینه خلاصه‌سازی متن و آموزش مدل‌های پیشرفته‌تر مورد استفاده قرار گیرد.

محدودیت‌های مطالعه:

محدودیت در داده‌ها: این پژوهش عمدتاً بر روی مجموعه داده‌های استاندارد انگلیسی متمرکز بوده است ارزیابی مدل‌ها بر روی زبان‌ها و حوزه‌های دیگر می‌تواند چالش‌های جدیدی را مشخص کند.

معیارهای ارزیابی: معیارهای ROUGE و سایر معیارهای استفاده شده، فقط بخشی از کیفیت خلاصه‌ها را ارزیابی می‌کنند و ممکن است نتوانند تمام جنبه‌های کیفی خلاصه‌ها را به طور کامل پوشش دهند. ارزیابی‌های انسانی و کیفی می‌تواند بینش‌های بیشتری ارائه دهد. پیچیدگی مدل‌ها: مدل‌های پیشرفته مانند GPT-4 و BERT، محاسبات زیادی را می‌طلبند و استفاده از آن‌ها در محیط‌های با منابع محدود ممکن است چالش‌برانگیز باشد.

عدم بررسی عمیق مدل‌های DeepExtract-v1 تا v6: با وجود اینکه مدل اصلی DeepExtract نتایج بهتری ارائه داد، به نظر می‌رسد که بررسی عمیق‌تر نتایج و عملکرد هر کدام از ورژن‌های این مدل می‌توانست اطلاعات بیشتری در اختیار بگذارد. عدم تحلیل کیفی: این پژوهش به دلیل تمرکز بر روش‌های کمی، تحلیل کیفی از نتایج را در نظر نگرفته است و لازم است در تحقیقات آتی به این موضوع نیز پرداخته شود.

عدم بررسی جامع همه روش‌های خلاصه‌سازی: این پژوهش بیشتر بر روی مدل‌های خاصی تمرکز داشته است و لازم است برای بررسی دقیق‌تر روش‌ها، مدل‌های دیگر نیز در پژوهش‌های آتی مورد بررسی قرار گیرند.

میزان تعمیم‌پذیری یافته‌ها:

قابلیت تعمیم به زبان انگلیسی: یافته‌های این پژوهش، به‌ویژه در مورد مدل‌های FuzzyTP-BERT و DeepExtract، در خلاصه‌سازی متون انگلیسی قابل تعمیم هستند.

نیاز به بررسی در زبان‌های دیگر: برای تعمیم‌پذیری یافته‌ها به زبان‌های دیگر، نیاز به انجام مطالعات مشابه با استفاده از داده‌های چند زبانه است.

حوزه‌های مختلف: قابلیت تعمیم‌پذیری نتایج به حوزه‌های مختلف نیازمند بررسی دقیق‌تر است. مدل‌هایی که در این پژوهش معرفی شده‌اند، در حوزه‌هایی که ساختار و سبک نگارش مشابهی دارند، احتمالاً عملکرد خوبی خواهند داشت.

اهمیت پیش‌پردازش داده‌ها: نتایج نشان می‌دهد که پیش‌پردازش مناسب داده‌ها می‌تواند به طور قابل توجهی بر عملکرد مدل‌ها تأثیر بگذارد.

نتیجه‌گیری:

این پژوهش نشان داد که استفاده از رویکردهای ترکیبی با بهره‌گیری از مدل‌سازی موضوعی فازی، شبکه‌های ترانسفورمر، ساختارهای سلسله‌مراتبی و روش‌های کاهش افزونگی می‌تواند عملکرد سیستم‌های خلاصه‌سازی متون را به طور قابل توجهی بهبود بخشد. مدل‌های FuzzyTP-BERT، DeepExtract و MFMMR-BertSum به عنوان مدل‌های برتر در این پژوهش معرفی شدند و پتانسیل بالایی برای کاربرد در زمینه‌های مختلف دارند. با این حال، باید به محدودیت‌های مطالعه توجه کرد و در تحقیقات آتی به بررسی چالش‌های موجود در زمینه‌های مختلف زبانی و کاربردی پرداخت. همچنین، لازم است که تحقیقات آینده به ارزیابی کیفی نتایج و بررسی جامع‌تر مدل‌های مختلف خلاصه‌سازی متمرکز شوند تا بتوان نتایج را به طور گسترده‌تر تعمیم داد.



منابع

- Adams, G., Fabbri, A., Ladhak, F., Lehman, E., & Elhadad, N. (۲۰۲۳). From sparse to dense: Gpt- ϵ summarization with chain of density prompting. *arXiv preprint arXiv:2309.04269*, ۱...
- Akhtar, N., Sufyan Beg, M.M., & Javed, H. (۲۰۱۹). Topic modelling with fuzzy document representation. In *Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019*, Ghaziabad, India, April ۱۲-۱۳, ۲۰۱۹, Revised Selected Papers, Part II (pp. ۵۷۷-۵۸۷). Springer Singapore. ۳...
- Al-Sabahi, K., Zuping, Z., & Nadher, M. (۲۰۱۸). A hierarchical structured self-attentive model for extractive document summarization (hssas). *IEEE Access*, 6, ۲۴۲۰۵-۲۴۲۱۲, ۲
- Al-Taani, A. T., & Al-Omour, M. M. (۲۰۱۴). An extractive graph-based arabic text summarization approach. In *The International Arab Conference on Information Technology* (pp. ۱۵۸-۱۶۳). ۵...
- Asa, A., Akter, S., Uddin, M., Hossain, M., Roy, S., & Afjal, M. (۲۰۱۷). A comprehensive survey on extractive text summarization techniques. *American Journal of Engineering Research*, 6(۱), ۲۲۶-۲۳۹, ۶...
- Azadani, M. N., Ghadiri, N., & Davoodijam, E. (۲۰۱۸). Graph-based biomedical text summarization: An itemset mining and sentence clustering approach. *Journal of Biomedical Informatics*, 84, ۴۲-۵۸, ۵...
- Bahani, M., El Ouazizi, A., & Maalmi, K. (۲۰۲۲). AraBERT and DF-GAN fusion for Arabic text-to-image generation. *Array*, 16, ۱۰۰۲۶۰, ۸
- Basyal, L., & Sanghvi, M. (۲۰۲۳). Text summarization using large language models: a comparative study of mpt-۷b-instruct, falcon-۷b-instruct, and openai chat-gpt models. *arXiv preprint arXiv:2310.10449*, ۱...
- Bharathi Mohan, G., Prasanna Kumar, R., Parathasarathy, S., Aravind, S., Hanish, K. B., & Pavithria, G. (۲۰۲۳). Text summarization for big data analytics: A comprehensive review of gpt-۲ and bert approaches. *Data Analysis and Internet of Things Infrastructure*, ۲۴۷-۲۶۴, ۹...
- Blei, D. M., & Lafferty, J. D. (۲۰۰۹). Topic models. In *Text Mining* (pp. ۱۰۱-۱۲۴). Chapman and Hall/CRC. ۳...
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (۲۰۰۳). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), ۹۹۳-۱۰۲۲, ۳...
- Carbonell, J., & Goldstein, J. (۱۹۹۸). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval* (p. ۳۳۵-۳۳۶). ۸
- Castillo, J. M., Mateo, M. A. L., Paras, A. D., Sagum, R. A., & Santos, V. D. F. (۲۰۱۳). Named entity recognition using support vector machine for filipino text documents. *International Journal of Future Computer and Communication*, 2(۵), ۵۳۰-۵۳۲, ۹
- Celikyilmaz, A., Bosselut, A., He, X., & Choi, Y. (۲۰۱۸). Deep communicating agents for abstractive summarization. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. ۱۶۶۲-۱۶۷۵). ۱۲
- Chen, Y., Wang, Z., Wen, B., & Zulkernine, F. (۲۰۲۴). Comparative analysis of open-source language models in summarizing medical text data. *arXiv preprint arXiv:2405.16295*, ۱...
- Cui, P., Hu, L., & Liu, Y. (۲۰۲۰). Enhancing extractive text summarization with topic-aware graph neural networks. *arXiv preprint arXiv:2010.06253*, ۱۱
- Darapaneni, N., Prajeesh, R., Dutta, P., Pillai, V. K., Karak, A., & Paduri, A. R. (۲۰۲۳). Abstractive text summarization using bert and gpt-۲ models. In *2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IconSCEPT)* (pp. ۱-۶). IEEE. ۱۰...
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (۲۰۱۸). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, ۸...
- Dong, Y., Mircea, A., & Cheung, J. C. (۲۰۲۱). Discourse-aware unsupervised summarization for long scientific documents. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. ۱۰۸۹-۱۱۰۲). ۱۳...
- El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (۲۰۲۱). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165, ۱۱۳۶۷۹, ۱۵
- Erkan, G., & Radev, D. R. (۲۰۰۴). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22(۱), ۴۵۷-۴۷۹, ۱۴...
- Esposito, M., Fujita, H., Minutolo, A., & Pota, M. (۲۰۲۲). Deep learning for natural language processing: Emerging methods and applications. *Array*, 14, ۱۰۰۱۳۸, ۱۷
- Fan, J., Tian, X., Lv, C., Zhang, S., Wang, Y., & Zhang, J. (۲۰۲۳). Extractive social media text summarization based on MFMMR-BertSum. *Array*, 20, ۱۰۰۳۲۲, ۱۸
- Fang, F., Cao, Y., Tang, H., & Cao, C. (۲۰۱۷). CoRank: Integrating graph-based ranking with matrix operations for improved summaries. *IEEE Transactions on Knowledge and Data Engineering*, ۲۹(۹), ۲۰۷۴-۲۰۸۷, ۵
- Gehrmann, S., Deng, Y., & Rush, A. M. (۲۰۱۸). Bottom-up abstractive summarization. *arXiv preprint arXiv:1808.10792*, ۱۲



- Ghadimi, A., & Beigy, H. (۲۰۲۳). SGCSumm: An extractive multi-document summarization method based on pre-trained language model, submodularity, and graph convolutional neural networks. *Expert Systems with Applications*, 215, ۱۱۹۳۰۸, ۱۱...
- Gholamrezazadeh, S., Salehi, M. A., & Gholamzadeh, B. (۲۰۰۹). A comprehensive survey on text summarization systems. In *2009 2nd International Conference on Computer Science and Its Applications* (pp. ۱-۶). IEEE. ۱۴...
- Gidiotis, A., & Tsoumakas, G. (۲۰۲۰a). A divide-and-conquer approach to the summarization of long documents. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, ۳۰۲۹-۳۰۴۰, ۲۰...
- Gidiotis, A., & Tsoumakas, G. (۲۰۲۰b). A divide-and-conquer approach to the summarization of long documents. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, ۳۰۲۹-۳۰۴۰, ۲۱
- Grail, Q., Perez, J., & Gaussier, E. (۲۰۲۱). [Globalizing BERT]-based transformer architectures for long document summarization. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. ۱۷۹۲-۱۸۱۰). ۲۱
- Gulati, V., Kumar, D., Popescu, D. E., & Hemanth, J. D. (۲۰۲۳). Extractive article summarization using integrated textrank and bm^{۲۵}+ algorithm. *Electronics*, 12(۲), ۳۷۲, ۲۲
- Gupta, V., & Lehal, G. S. (۲۰۱۰). A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence*, 2(۳), ۲۵۸-۲۶۸, ۲۲
- Halder, N. A. H., Li, J., Ali, M. E., Cai, T., Chen, Y., Sellis, T., & Xu, G. (۲۰۲۳). Top-k socio-spatial co-engaged location selection for social users. *IEEE Transactions on Knowledge and Data Engineering*, 35(۵), ۵۳۲۵-۵۳۴۰, ۸
- Herskovic, J. R., Cohen, T., Subramanian, D., Iyengar, M. S., Smith, J. W., & Bernstam, E. V. (۲۰۱۱). MedRank: Using graph-based concept ranking to index biomedical texts. *International Journal of Medical Informatics*, 80(۶), ۴۳۱-۴۴۱, ۵...
- Irfan, M., & Zulfikar, W. B. (۲۰۱۷). Implementation of fuzzy c-means algorithm and tf-idf on english journal summary. In *2017 Second International Conference on Informatics and Computing (ICIC)* (pp. ۱-۵). IEEE. ۲۳
- Isonuma, M., Fujino, T., Mori, J., Matsuo, Y., & Sakata, I. (۲۰۱۷). Extractive summarization using multi-task learning with document classification. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. ۲۱۰۱-۲۱۱۰). ۲۳
- Jain, D., Borah, M. D., & Biswas, A. (۲۰۲۰). Fine-tuning textrank for legal document summarization: A Bayesian optimization based approach. In *Forum for information retrieval evaluation* (pp. ۴۱-۴۸). ۱۷
- Jia, Y., Gu, Z., Jiang, Z., Gao, C., & Yang, J. (۲۰۲۳). Persistent graph stream summarization for real-time graph analytics. *World Wide Web*, ۱-۲۱, ۱۵
- Jin, H., Zhang, Y., Meng, D., Wang, J., & Tan, J. (۲۰۲۴). A comprehensive survey on process-oriented automatic text summarization with exploration of llm-based methods. *arXiv preprint arXiv:2403.02901*. ۱۰...
- Jing, B., You, Z., Yang, T., Fan, W., & Tong, H. (۲۰۲۱). Multiplex graph neural network for extractive text summarization. *arXiv preprint arXiv:2108.12870*. ۲۵
- Joshi, A., Fidalgo, E., Alegre, E., & Alaiz-Rodriguez, R. (۲۰۲۲). RankSum—An unsupervised extractive text summarization based on rank fusion. *Expert Systems with Applications*, 200, ۱۱۶۸۴۶, ۲۵
- Joshi, A., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (۲۰۲۳). DeepSumm: Exploiting topic models and sequence to sequence networks for extractive text summarization. *Expert Systems with Applications*, 211, ۱۱۸۴۴۲, ۲۶
- Kågebäck, M., Mogren, O., Tahmasebi, N., & Dubhashi, D. (۲۰۱۴). Extractive summarization using continuous vector space models. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality* (pp. ۳۱-۳۹). ۲۴
- Kalyan, K. S. (۲۰۲۳). A survey of gpt-۳ family large language models including chatgpt and gpt-۴. *Natural Language Processing Journal*, 100048, ۱۰...
- Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (۲۰۱۵). FLATM: A fuzzy logic approach topic model for medical documents. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) Held Jointly with 2015 5th World Conference on Soft Computing (WConSC)* (pp. ۱-۶). IEEE. ۲۶
- Kieuvongngam, S., & Stheewasinnon, P. (۲۰۲۱). A comparison study on abstractive summarization of covid-۱۹ medical papers using bert and gpt-۲. In *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. ۹۹۷-۱۰۰۰). IEEE. ۲۷
- Kong, S. Y., & Lee, L. S. (۲۰۰۶). Improved spoken document summarization using probabilistic latent semantic analysis (PLSA). In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 1* (pp. I-I). IEEE. ۲۸
- Koroteev, M. V. (۲۰۲۱). BERT: A review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*. ۲۸
- Kun, C., Chuanyi, L., Xinxin, J., Jidong, G., & Bin, L. (۲۰۲۱). News summarization extracting method based on improved MMR algorithm. *Journal of Applied Science*, 39(۳), ۴۴۲-۴۴۳, ۱۷



- Lamsiyah, S., El Mahdaouy, A., El Alaoui, S. O., & Espinasse, B. (۲۰۲۰). A supervised method for extractive single document summarization based on sentence embeddings and neural networks. In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2019) Volume 4-Advanced Intelligent Systems for Applied Computing Sciences* (pp. ۷۰-۸۸). Springer International Publishing. ۲۸
- Li, X., Du, L., & Shen, Y. D. (۲۰۱۲). Update summarization via graph-based sentence ranking. *IEEE Transactions on Knowledge and Data Engineering*, 25(۵), ۱۱۶۲-۱۱۷۴, ۵...
- Liu, J., Chen, Y., Huang, X., & Li, J. (۲۰۲۲). GNN-based long and short term preference modeling for next-location prediction. *Information Sciences*, 629, ۱-۱۴, ۱۵
- Liu, Y. (۲۰۱۹). Fine-tune BERT for extractive summarization. *arXiv preprint arXiv:1903.10318*, ۳۰
- Liu, Y., & Lapata, M. (۲۰۱۹). Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. ۳۷۳-۳۷۴۰). ۲۹
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (۲۰۱۹). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, ۱۶...
- Liu, Y., Zhong, S.-h., & Li, W. (۲۰۱۲). Query-oriented multi-document summarization via unsupervised deep learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. ۲۶, No. ۱, pp. ۱۶۹۹-۱۷۰۵). ۱۷
- Ma, T., Pan, Q., Rong, H., Qian, Y., Tian, Y., & Al-Nabhan, N. (۲۰۲۱). T-bertsum: Topic-aware text summarization based on bert. *IEEE Transactions on Computational Social Systems*, 9(۳), ۸۷۹-۸۹۰, ۲۷...
- Mao, Q., Zhu, H., Liu, J., Ji, C., Peng, H., Li, J., & Wang, Z. (۲۰۲۲). Muchsum: Multi-channel graph neural network for extractive summarization. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. ۲۶۱۷-۲۶۲۲). ۳۱
- Mao, R., Chen, G., Zhang, X., Guerin, F., & Cambria, E. (۲۰۲۳). Gpteval: A survey on assessments of chatgpt and gpt-۴. *arXiv preprint arXiv:2308.12488*, ۱۰...
- Mohan, G. B., & Kumar, R. P. (۲۰۲۱). A comprehensive survey on topic modeling in text summarization. In *International Conference on Micro-Electronics and Telecommunication Engineering* (pp. ۲۳۱-۲۴۰). Springer Nature Singapore. ۳۱
- Moratan, N., & Chitrakala, S. (۲۰۱۷). A survey on extractive text summarization. In *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)* (pp. ۱-۶). IEEE. ۳۲...
- Mutlu, B., Sezer, E. A., & Akcayol, M. A. (۲۰۲۰). Candidate sentence selection for extractive text summarization. *Information Processing & Management*, 57(۶), ۱۰۲۳۵۹, ۳۲...
- Nallapati, R., Zhai, F., & Zhou, B. (۲۰۱۷). Summarunner: A recurrent neural network-based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*. ۳۴
- Narayan, S., Cohen, S. B., & Lapata, M. (۲۰۱۸). Ranking sentences for extractive summarization with reinforcement learning. *arXiv preprint arXiv:1802.08636*, ۳۴
- Nenkova, A., & Vanderwende, L. (۲۰۰۵). *The Impact of Frequency on Summarization*. (Tech. Rep. MSR-TR-۲۰۰۵). Microsoft Research. ۱۴...
- Onan, A., & Alhumyani, H. A. (۲۰۲۴a). Contextual hypergraph networks for enhanced extractive summarization: Introducing multi-element contextual hypergraph extractive summarizer (mches). *Applied Sciences*, 14(۱۱), ۴۶۷۱, ۱...
- Onan, A., & Alhumyani, H. A. (۲۰۲۴b). FuzzyTP-BERT: Enhancing extractive text summarization with fuzzy topic modeling and transformer networks. *Journal of King Saud University - Computer and Information Sciences*, 102080, ۱...
- Pandey, M., Sharma, C., Sharma, S., & Aggarwal, T. (۲۰۲۴). Hybrid technique of topic modelling and text summarization: A case study on predicting trends in green computing. *International Journal of Performability Engineering*, 20(۳), ۱۳۹, ۳۳
- Qaroush, A., Farha, I. A., Ghanem, W., Washaha, M., & Maali, E. (۲۰۲۱). An efficient single document arabic text summarization using a combination of statistical and semantic features. *Journal of King Saud University-Computer and Information Sciences*, 33(۶), ۶۷۷-۶۹۲, ۳۵
- Radev, D. R., Hovy, E., & McKeown, K. (۲۰۰۲). Introduction to the special issue on summarization. *Computational Linguistics*, 28(۴), ۳۹۹-۴۰۸, ۱۵
- Rani, R., & Lobiyal, D. K. (۲۰۲۱). An extractive text summarization approach using tagged-LDA based topic modeling. *Multimedia Tools and Applications*, 80, ۳۲۷۵-۳۳۰۵, ۳۶
- Rijcken, E., Mosteiro, P., Zervanov, K., Spruit, M., & Scheepers, F., Kaymak, U. (۲۰۲۲). FuzzyTM: A software package for fuzzy topic modeling. In *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. ۱-۸). IEEE. ۳۶
- Rijcken, E., Scheepers, F., Mosteiro, P., Zervanov, K., Spruit, M., & Kaymak, U. (۲۰۲۱). A comparative study of fuzzy topic models and LDA in terms of interpretability. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. ۱-۸). IEEE. ۳...



- Ruan, Q., Ostendorff, M., & Rehm, G. (۲۰۲۲). [Histruct+]: Improving extractive text summarization with hierarchical structure information. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. ۱۲۹۲-۱۳۰۸). ۳۷...
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (۲۰۲۰). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, ۱۶
- See, A., Liu, P. J., & Manning, C. D. (۲۰۱۷). Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. ۱۰۷۳-۱۰۸۳). ۱۲...
- Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (۲۰۲۰). MPNet: Masked and permuted pre-training for language understanding. In *Advances in Neural Information Processing Systems*, 33, ۱۶۸۵۷-۱۶۸۶۷, ۱۶
- Srikanth, A., Umasankar, A. S., Thanu, S., & Nirmala, S. J. (۲۰۲۰). Extractive text summarization using dynamic clustering and co-reference on BERT. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. ۱-۵). IEEE. ۲۷...
- Wang, Y., Zhang, J., Yang, Z., Wang, B., Jin, J., & Liu, Y. (۲۰۲۴). Improving extractive summarization with semantic enhancement through topic-injection based BERT model. *Information Processing & Management*, 61(۳), ۱۰۳۶۷۷, ۴۰...
- Yang, C. Z., Fan, J. S., & Liu, Y. F. (۲۰۱۶). Multi-document summarization using probabilistic topic-based network models. *Journal of Information Science and Engineering*, 32(۶). ۴۱
- Yadav, J., & Meena, Y. K. (۲۰۱۶). Use of fuzzy logic and wordnet for improving performance of extractive automatic text summarization. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. ۲۰۷۱-۲۰۷۷). IEEE. ۴۲
- Yadav, A. K., Ranvijay, Yadav, R. S., & Maurya, A. K. (۲۰۲۳). State-of-the-art approach to extractive text summarization: A comprehensive review. *Multimedia Tools and Applications*, 82(۱۹), ۲۹۱۳۵-۲۹۱۹۷, ۱۰...
- Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., ... & Gadekallu, T. R. (۲۰۲۴). Gpt (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*. ۱۰...
- Yu, D., Fang, A., & Xu, Z. (۲۰۲۳). Topic research in fuzzy domain: Based on LDA topic modelling. *Information Sciences*, 648, ۱۱۹۶۰۰, ۴۱...
- Yuan, R., Wang, Z., & Li, W. (۲۰۲۰). Fact-level extractive summarization with hierarchical graph mask on BERT. *arXiv preprint arXiv:2011.09739*, ۲۷...
- Zhang, H., Liu, X., & Zhang, J. (۲۰۲۳). Extractive summarization via ChatGPT for faithful summary generation. *arXiv preprint arXiv:2304.04193*, ۱۹...
- Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (۲۰۲۰). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning* (pp. ۱۱۳۲۸-۱۱۳۳۹). PMLR. ۳۹...
- Zhang, X., Wei, F., & Zhou, M. (۲۰۱۹). HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. *arXiv preprint arXiv:1905.06566*, ۴۵...
- Zheng, H., & Lapata, M. (۲۰۱۹). Sentence centrality revisited for unsupervised summarization. *arXiv preprint arXiv:1906.03508*, ۴۰...
- Zheng, J., Zhao, Z., Song, Z., Yang, M., Xiao, J., & Yan, X. (۲۰۲۰). Abstractive meeting summarization by hierarchical adaptive segmental network learning with multiple revising steps. *Neurocomputing*, 378, ۱۷۹-۱۸۸, ۴۶
- Zhong, M., Liu, P., Wang, D., Qiu, X., & Huang, X. (۲۰۱۹). Searching for effective neural extractive summarization: What works and what's next. *arXiv preprint arXiv:1907.03491*, ۴۵...
- Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., & Huang, X. (۲۰۲۰). Extractive summarization as text matching. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. ۶۱۹۷-۶۲۰۸). ۳۷...
- Zhong, Y., Tang, Z., Ding, X., Zhu, L., Le, Y., Li, K., & Li, K. (۲۰۱۷). An improved LDA multi-document summarization model based on TensorFlow. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. ۲۵۵-۲۵۹). IEEE. ۴۷
- Zhu, T., Hua, W., Qu, J., Hosseini, S., & Zhou, X. (۲۰۲۳). Auto-regressive extractive summarization with replacement. *World Wide Web*, 26(۴), ۲۰۰۳-۲۰۲۶, ۴۷...
- Zhu, Z., Gong, S., Qi, J., & Tong, C. (۲۰۲۴). Htposum: Heterogeneous tree structure augmented with triplet positions for extractive summarization of scientific papers. *Expert Systems with Applications*, 238, ۱۲۲۳۶۴, ۳۷...



A review of text summarization methods with emphasis on deep learning and fuzzy logic models

Mansour Zare

M.Sc Student in Computer Engineering - Artificial Intelligence and Robotics, Apadana Institute of Higher Education, Shiraz

Dr Haleh Homayoni

Assistant Professor, Faculty Member of Computer Engineering Department, Apadana Institute of Higher Education, Shiraz

Zahra TasmimGhatei

M.Sc Student in Computer Engineering, Apadana Institute of Higher Education, Shiraz

Abstract

This review article explores various text summarization methods, with a focus on deep learning and fuzzy logic. Summarization techniques are divided into two categories: extractive (selecting key sentences) and abstractive (generating new sentences). Transformer-based models such as BERT and GPT, with a better understanding of semantic relationships, perform better than traditional methods. Fuzzy logic, by modeling the ambiguous relationships between words, creates richer summaries. The FuzzyTP-BERT model combines fuzzy logic and BERT. Advanced models such as DeepExtract with GPT- ϵ and MFMMR-BertSum with MMR provide better results in summarization. Evaluation criteria include ROUGE, BLANC, and BERTScore. Challenges include preserving meaning, processing long texts, and removing redundancy. Research in these areas will help improve the accuracy and efficiency of text summarization.

Keywords: Extractive Summarization, Abstractive Summarization, Fuzzy Logic, Transformer Networks, Semantic Embedding, ROUGE Metrics.