

Systematic Review of Query-based Abstractive Summarization Studies

Neda Abbasi Dashtaki¹ | Mehrdad Cheshmehsohrabi² | Mitra Pashootanizadeh³

Hamidreza Baradaran Kashani⁴

Abstract

Purpose: Today, people are aware of the fact that knowledge is power. Therefore, there has been a shift from information retrieval to knowledge retrieval and knowledge discovery. On the other hand, the study of the vast volume of textual documents on the web has made access to and usability of knowledge challenging for them. One of the solutions to tackle this issue is query-based abstract summarization. Query-based abstract summarization is a fast and efficient approach for navigating texts and is considered a highly dynamic research area. In this study, a systematic review of the studies in this field has been conducted to identify and analyse the relevant research.

Method: In the present applied research, a systematic review was conducted using the PRISMA guidelines. This guideline is implemented in four steps: identification, screening, eligibility, and inclusion, utilizing an appropriate search strategy without time restrictions in the Scopus, Web of Science, IEEE Xplore, the ACM Digital Library, Google Scholar, ProQuest, Noor Mags, Mag Iran, SID, Civilica, Elm net, and Ganj databases. Ultimately, out of the 1,714 identified documents, 31 were found to be eligible and included in the systematic review.

Findings: The findings of the conducted review indicate that studies in this field are relatively recent and have been published with both upward and downward trends. Most of these studies are in the form of articles published in journals. Researchers have predominantly utilized a one-stage approach for the proposed summarization systems, with a greater focus on supervised and self-supervised learning. Additionally, they have employed methods based on rules, statistics, and machine learning. The models used are based on graphs, neural networks, and pre-trained architectures. The input type for the systems is mostly single-document, with Debatepedia identified as the most popular dataset. Among the seventeen of evaluation metrics, ROUGE has been the most widely used.

Conclusion: The reviews indicate how the synergies that have occurred in learning, models, methods used, and evaluation metrics have helped to mitigate challenges such as the mismatch between the generated summary and the query; the incongruity between the generated summary and the source text, the lack of labelled data for training models, redundancy, limited datasets, the absence of datasets specifically for this type of summarization, the lack of improved evaluation metrics for accurately assessing generated summaries, semantic ambiguity due to the lack of distinction between sentences with different meanings, and the absence of alignment between input and output sequences. Ultimately, these improvements have contributed to enhancing the overall performance of summarization systems and their development. However, the ability to understand semantic in these systems has not yet bridged the gap between system-generated summaries and human summaries. This is because the understood semantic remains superficial and shows a degree of reliance on the syntactical structures in the models. In fact, the ability to understand semantic can guarantee the creation of systems that recognize the deeper semantics and insights embedded in the text and apply them in their output based on the specified task. Accordingly, the presentation of innovations to address these inefficiencies is proposed as directions for future research. In this regard, semantic modelling and semantic understanding should be institutionalized within these summarization systems, contributing to the refinement and advancement of existing methodologies. Furthermore, it is essential to keep pace with the changing and evolving information sources as well as the developments in user requests and their knowledge domains. Additionally, there is a noticeable gap for these systems in non-English languages. This can be addressed by developing and strengthening natural language processing tools for non-English languages, enabling their practical implementation.

Keywords

Automatic Summarization, Query-based Summarization, Abstractive Approach, Systematic Review, PRISMA

Citation: Abbasi Dashtaki, N., Cheshmehsohrabi, M., Pashootanizadeh, M., & Baradaran Kashani, H. (2025). Systematic Review of Query-based Abstractive Summarization Studies. *Librarianship and Information Organization Studies*, 36 (3), 84-139

Doi: 10.30484/nastinfo.2025.3644.2295



1. PhD. Candidate, Knowledge and Information Science, University of Isfahan, Isfahan, Iran; nabbasi.d.69@gmail.com

2. Professor, Knowledge and Information Science, University of Isfahan, Isfahan, Iran (Corresponding Author); mo.sohrabi@edu.ui.ac.ir

3. Associate Professor, Knowledge and Information Science, University of Isfahan, Isfahan, Iran; m.pashootanizade@edu.ui.ac.ir

4. Assistant Professor, Artificial Intelligence, University of Isfahan, Isfahan, Iran; hrb.kashani@eng.ui.ac.ir

Article Type: Review Article

Article history: Received: 30 Aug. 2024;

Accepted: 10 Dec. 2024



Publisher: National Library and Archives of I.R. of Iran
© The Author(s).

مرور نظام‌مند مطالعات حوزه خلاصه‌سازی انتزاعی

مبتنی بر پرس‌وجو

ندا عباسی دشتکی^۱ | مهرداد چشمه‌سهرابی^۲ | میترا پشوتنی‌زاده^۳ | حمیدرضا برادران کاشانی^۴

چکیده

هدف: امروزه افراد به این حقیقت واقف‌اند که دانش قدرت است. لذا از بازیابی اطلاعات به سمت بازیابی دانش و کشف دانش سوق پیدا کرده‌اند. از طرفی، مطالعه حجم عظیم اسناد متنی وب، دسترس‌پذیری و کاربردپذیری دانش را برای آن‌ها دشوار نموده است. یکی از راهکارها جهت مواجهه با این مسئله، خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو است. خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو رویکردی سریع و کارآمد برای پیمایش متون است و یک حوزه پژوهشی بسیار پویا محسوب می‌شود. در این پژوهش با استفاده از مرور نظام‌مند، مطالعات پیرامون این حوزه شناسایی و تجزیه و تحلیل شده‌اند. **روش:** در پژوهش کاربردی حاضر با استفاده از دستورالعمل پریزما، یک مرور نظام‌مند انجام شده است. این دستورالعمل در قالب چهار گام شناسایی، غربالگری، شایستگی و شمول با استفاده از یک راهبرد جستجوی مناسب و بدون محدودیت زمانی در پایگاه‌های اسکوپوس، وب‌آوساینس، آی‌تریپل‌ای، پایگاه علمی کتابخانه دیجیتال ای‌سی‌ام، گوگل اسکالر، پروکوئست، نورمگز، مگیران، سید، سیویلیکا، علم نت و گنج اعمال شده است. درنهایت از ۱۷۱۴ مدرک شناسایی شده ۳۱ مورد واجد شرایط بوده و مشمول مرور نظام‌مند شده‌اند.

یافته‌ها: محاصل مرور انجام‌شده نشان می‌دهد که مطالعات این حوزه قدمت چندانی ندارند و با سیر نوأم صعودی و نزولی منتشر شده‌اند. اکثر این مطالعات از نوع مقاله منتشرشده در مجلات هستند. پژوهشگران برای سیستم‌های خلاصه‌سازی پیشنهادی بیشتر از رویکرد یک‌مرحله‌ای استفاده نموده‌اند و یادگیری‌های با نظارت و خودنظارتی بیشتر موردتوجه آن‌ها بوده است. همچنین، از روش‌های مبتنی بر قانون، آمار و یادگیری ماشین بهره گرفته‌اند. مدل‌های به‌کار گرفته‌شده مبتنی بر گراف، شبکه‌های عصبی و از پیش آموزش‌دیده است. نوع ورودی سیستم‌ها بیشتر تک‌سندی بوده و Debatepedia به‌عنوان محبوب‌ترین مجموعه داده شناسایی شده است. از میان هفده معیار ارزیابی ROUGE بیشترین کاربرد را داشته است.

نتیجه‌گیری: بررسی‌ها نشان داد که چگونه هم‌افزایی‌های اتفاق افتاده در یادگیری، مدل‌ها، روش‌های مورد استفاده و معیارهای ارزیابی کاهش چالش‌هایی از قبیل عدم تناسب خلاصه تولیدشده با پرس‌وجو، عدم تناسب خلاصه تولیدشده با متن منبع، فقدان داده‌های برچسب‌گذاری شده برای آموزش مدل‌ها، افزونگی، مجموعه داده‌های محدود، فقدان مجموعه داده مخصوص این نوع خلاصه‌سازی، عدم وجود معیارهای ارزیابی بهبودیافته برای ارزیابی دقیق خلاصه‌های تولیدشده، ابهام معنایی ناشی از عدم تمایز بین جملات با معنای متفاوت و عدم رابطه هم‌ترازی بین توالی‌های ورودی و خروجی را به دنبال داشته است و درنهایت به بهبود عملکرد کلی سیستم‌های خلاصه‌سازی و توسعه آن‌ها کمک نموده است. اما، توانایی درک معنا در سیستم‌ها هنوز فاصله میان خلاصه‌های سیستمی و خلاصه‌های انسانی را پر نکرده است. زیرا معنای درک شده هنوز سطحی بوده و تا حدی وابستگی به ساختارهای نحوی در مدل‌ها دیده می‌شود. درواقع، توانایی درک معنا می‌تواند ضامن ایجاد سیستم‌هایی باشد که معنا و بینش‌های عمیق نهفته در متن را تشخیص داده و براساس وظیفه مشخص شده آن‌ها را در خروجی خود اعمال می‌کنند. بر این قرار، ارائه نوآوری‌هایی جهت رفع این ناکارآمدی به‌عنوان جهت‌های پژوهشی آینده پیشنهاد می‌شود. در این مسیر باید مدل‌سازی‌های معنایی و درک معنا در این سیستم‌های خلاصه‌سازی نهادینه شود که به اصلاح و پیشرفت مسیر تکامل روش‌شناسی‌های موجود کمک می‌نماید. همچنین، بهتر است با تغییر و تکامل منابع اطلاعاتی و تحولات درخواست‌های کاربران و زمینه‌های دانشی آن‌ها نیز همگام شد. افزون بر این، خلأ این سیستم‌ها در زبان‌های غیرانگلیسی احساس می‌شود. این امر با ایجاد و تقویت ابزارهای پردازش زبان طبیعی برای زبان‌های غیر انگلیسی قابلیت عملیاتی‌سازی دارد.

کلیدواژه‌ها

خلاصه‌سازی خودکار (ماشینی)، خلاصه‌سازی مبتنی بر پرس‌وجو، رویکرد انتزاعی، مرور نظام‌مند، پریزما

۱. دانشجوی دکتری، علم اطلاعات و دانش‌شناسی، دانشگاه اصفهان، اصفهان، ایران؛

nabbasi.d.69@gmail.com

۲. استاد، علم اطلاعات و دانش‌شناسی، دانشگاه اصفهان، اصفهان، ایران (نویسنده مسئول)؛

mosohrabi@edu.ui.ac.ir

۳. دانشیار، علم اطلاعات و دانش‌شناسی، دانشگاه اصفهان، اصفهان، ایران؛

m.pashootanizade@edu.ui.ac.ir

۴. استادیار، هوش مصنوعی، دانشگاه اصفهان، اصفهان، ایران؛

hrbkashani@eng.ui.ac.ir

نوع مقاله: مروری

تاریخ دریافت: ۱۴۰۳/۰۶/۰۹

پذیرش: ۱۴۰۳/۰۹/۲۰



مقدمه

امروزه دانش در کنار نیروی کار، زمین و سرمایه، عنصر اساسی تولید است و ساختاری برای برآورد و ادغام تجربیات و اطلاعات جدید ارائه می دهد. از این رو، به عنوان منبع کلیدی دارای مزیت رقابتی، به رسمیت شناخته شده و از ارکان اصلی پیشرفت جوامع محسوب می شود. بر این قرار تولید، کسب، انتقال و کاربرد آن بسیار حیاتی است. در پی تداوم پیشرفت های بشر بسترهای ضبط، دسترسی و انتقال دانش نیز پیشرفت های چشمگیری داشته است. از جمله این پیشرفت ها ظهور وب در سال ۱۹۸۹ توسط برنرزی^۱ را می توان نام برد (Berners-Lee, 1998). با توجه به رشد روزافزون وب، بررسی حجم عظیمی از دانش موجود در اسناد متنی، کاری زمان بر و خسته کننده است. از سویی دیگر، افراد به این حقیقت واقف اند که دانش آن ها را قادر می سازد تفکر بهتر و توانمندی بیشتری برای زندگی با کیفیت تر و تأثیرگذارتر داشته باشند. این امر اشتیاق آن ها را به دسترسی و استفاده بیشتر از این عنصر قدرت بخش فزونی بخشیده و گرایش آن ها را از بازیابی اطلاعات به سمت بازیابی دانش و کشف دانش سوق داده است. بررسی ها نشان می دهد از جمله راهکارها جهت عملیاتی سازی این امور با صرف زمان، هزینه و انرژی کمتر، بهره گیری از خلاصه سازی انتزاعی مبتنی بر پرس و جو^۲ است.

خلاصه سازی انتزاعی با ساختن جملات جدید مانند یک انسان، خلاصه کلی کوتاه و مختصری از متن ایجاد می کند. این خلاصه که ممکن است شامل عبارات جدیدی باشد که در متن منبع موجود نیست (Andhale & Bewoor, 2016)؛ تمام نکات برجسته متن را به صورت فشرده و منسجم پوشش می دهد (Nema et al., 2017). علاوه بر این، به دنبال راه هایی برای تلفیق یا پارافریز^۳ جملات در اسناد به منظور تولید جملات انتزاعی از آن ها است (Alambo et al., 2020). خلاصه سازی مبتنی بر پرس و جو به منظور علاقه کاربر انجام می شود (Rahman & Borah, 2020) و یک سند یا اسناد با توجه به نیاز اطلاعاتی بیان شده در درخواست کاربر خلاصه می شوند (Nagalavi & Hanumanthappa, 2019). به طوری که کاربر زمان کمتری را صرف خواندن کل سند می کند و از این طریق به تصمیم گیری سریع وی کمک می شود (Rahman & Borah, 2020). با بهره گیری از این خلاصه سازی و خلاصه کردن چند سند به جای ارائه یک تکه از هر سند و ترکیب اطلاعات چندین سند با هم، می توان پاسخ منسجمی به سؤالات پیچیده داد (Nagalavi & Hanumanthappa, 2019).

¹. Berners-Lee

². Query-based abstractive summarization

³. Paraphrasing

(2019). خلاصه‌سازی مبتنی بر پرس‌وجو اولین بار در فراهمایی درک سند^۱ در سال ۲۰۰۴ ارائه شد. این فراهمایی از سال ۲۰۰۴ خلاصه‌سازی مذکور را با تمرکز بر پرس‌وجوهای پیچیده و پاسخ‌های خاص راه‌اندازی نموده است (Mehdad et al., 2014). اما، بررسی‌ها نشان می‌دهد که این نوع خلاصه‌سازی با تمرکز بر رویکرد انتزاعی قدمت چندانی ندارد. برخلاف رویکرد استخراجی که بر استخراج بخش‌هایی از متن مطابق با پرس‌وجوی اعلام‌شده تأکید دارد؛ رویکرد انتزاعی بر شناسایی اطلاعات برجسته موجود در متن و ایجاد متن جدیدی که مرتبط با پرس‌وجوی کاربر باشد متمرکز است.

خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو یک رویکرد سریع و کارآمد برای پیمایش و درک متون وب از جمله مقاله‌ها، اخبار، مطالب وب‌سایت‌ها، ویکی‌ها و وبلاگ‌ها است و طیف گسترده‌ای از کاربردها را دارد. لذا، یکی از دلایل کلیدی اهمیت این پژوهش، کاربرد گسترده آن در زمینه‌های مختلف است. از پژوهش‌های علمی و تولید محتوای خودکار گرفته تا بازاریابی، مدیریت اسناد سازمانی و پاسخگویی به نیازهای اطلاعاتی روزمره افراد. موتورهای جستجو که امروزه موارد استفاده زیادی برای کاربران وب دارند از این رویکرد استفاده کرده و از طریق درک سریع محتوای مرتبط با پرس‌وجو موجب صرفه‌جویی در وقت افراد می‌شوند. مدل‌های زبانی مختلف و ربات‌های چت که امروزه شاهد رشد روزافزون کمی و کیفی آن‌ها در عرصه‌های مختلف هستیم نیز با کمک خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو، به پرسش‌ها و درخواست‌های کاربران پاسخ می‌دهند. درواقع، این روش به موتورهای جستجو و ربات‌های چت توانایی می‌دهد تا اطلاعات مرتبطی را به شیوه‌ای مؤثر و سریع ارائه دهند. این ویژگی به‌ویژه در دنیای امروز که اطلاعات به‌سرعت در حال تغییر و گسترش است، حیاتی‌تر شده است. از این‌رو، خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در کانون توجه قرار گرفته و پژوهش در این حوزه از اهمیت بالایی برخوردار شده است. از این‌رو، پژوهشگران سعی کرده‌اند با استفاده از روش‌ها و مدل‌های مختلف متون را مطابق با پرس‌وجوی کاربر خلاصه کنند. در این میان تنها تعداد اندکی پژوهش به‌مرور اجمالی مطالعات این حوزه بدون تأکید خاص بر رویکرد انتزاعی آن پرداخته‌اند. این پژوهش‌ها با بررسی تنها ۱۰ مقاله در این حوزه، بحث پیرامون دو رویکرد مبتنی بر نمودار و مبتنی بر خوشه، بررسی اجمالی پیاده‌سازی، مزایا و چالش‌های موجود در حوزه خلاصه‌سازی مبتنی بر پرس‌وجو در پژوهش النزی و البالا^۲ (۲۰۲۳) ارائه معیارهای ارزیابی، مجموعه داده‌ها، تقسیم‌بندی کلی خلاصه‌سازی مبتنی بر پرس‌وجو با رویکرد استخراجی و انتزاعی، ارائه

^۱. Document Understanding Conference (DUC)

^۲. Alanzi & Alballa

روش‌های با نظارت و بدون نظارت و واکاوی چند معماری ارائه شده پیشین توسط یو^۱ (۲۰۲۲)؛ دسته‌بندی روش‌های خلاصه‌سازی، ارائه فنون مختلف مورد استفاده در هر دسته، گردآوری معیارهای ارزیابی و بحث پیرامون تحولات خلاصه‌سازی چندسندی در مطالعه روی و کندو^۲ (۲۰۲۳) و بررسی روش‌های مختلف خلاصه‌سازی و فنون کلی خلاصه‌سازی مبتنی بر پرس‌وجو توسط نیموات و جوشیارا^۳ (۲۰۱۷) انجام شد.

با آنکه النزی و البالا (۲۰۲۳)، انگیزه اصلی مرور انجام داده را ارائه یک بررسی جامع از مطالعات موجود در مورد خلاصه‌سازی چندسندی متمرکز بر پرس‌وجو، با هدف کمک به پژوهشگران در بهبود خلاصه‌های مبتنی بر پرس‌وجو دانسته‌اند. اما همه مدارک مربوط را مورد توجه قرار ندادند و تنها ۱۰ مدرک را با تمرکز بر روش‌های گراف و خوشه‌بندی مدنظر قرار دادند. همچنین، رویکرد انتخاب جملات مربوط به پرس‌وجو، فن مورد استفاده برای حذف افزونگی و معیارهای عملکرد در مجموعه داده‌ها را بررسی نمودند. اضافه بر این، در این پژوهش که در حقیقت مطالعه خلاصه‌سازی چندسندی مبتنی بر پرس‌وجو است، پیرامون مدل‌های مبتنی بر شبکه‌های عصبی، مدل‌های از پیش آموزش دیده شده، روش‌های مبتنی بر قاعده و روش‌های یادگیری ماشین بحث و بررسی مقایسه‌ای ارائه نشده است. از جمله پژوهش‌های دیگر، می‌توان به پژوهش روی و کندو (۲۰۲۳) اشاره کرد که روش‌شناسی واضحی ندارد و بر روش‌های استخراجی تأکید نموده و به اندازه کافی فنون خلاصه‌سازی انتزاعی را بررسی نکرده است. این پژوهش، بیشتر روش‌های یادگیری ماشین کلاسیک را مورد بحث قرار داده و عمیقاً به پیشرفت‌های فنون یادگیری عمیق که در کارهای خلاصه‌سازی امیدوارکننده بوده‌اند، نمی‌پردازد. حال آنکه این فنون، به‌ویژه مواردی که مبتنی بر ترنسفورمر^۴ هستند، نوید قابل توجهی در بهبود کیفیت خلاصه‌سازی نشان داده‌اند. به نظر می‌رسد عدم توجه ویژه به این پیشرفت‌ها به‌عنوان اصلی‌ترین و جدیدترین موارد، در پیش‌بینی روندهای پژوهشی آتی تأثیرگذار بوده و کاربرد یافته‌ها را در روندهای فعلی در این زمینه محدود نموده است. مطالعه نیموات و جوشیارا (۲۰۱۷) نیز در قالب یک مرور روایتی ساده و کوتاه به انواع خلاصه‌سازی و چند فن در این حوزه اشاره کرده است و فاقد نمونه‌های دقیق از پیاده‌سازی‌های عملی یا مطالعات موردی است که نشان دهد چگونه می‌توان این روش‌ها را در سناریوهای دنیای واقعی به کار برد. همچنین، یک خلأ بالقوه در بحث معیارهای ارزیابی به‌منظور ارزیابی روش‌های خلاصه‌سازی پیشنهادی وجود دارد که جهت اعتبارسنجی عملکرد آن‌ها در کاربردهای عملی بسیار مهم

¹. Yu

². Roy & Kundu

³. Nimavat & Joshiara

⁴. Transformer

است. پژوهش یو (۲۰۲۲) ضمن اینکه هفت سال از انتشار آن گذشته، یک مطالعه مروری بر روی خلاصه‌سازی انتزاعی و استخراجی با نظارت و بدون نظارت با محوریت پرس‌وجوی کاربر، بدون روش‌شناسی معتبر و بررسی نظام‌مند است. به نظر می‌رسد این امر شکافی در درک وضعیت فعلی و چشم‌اندازهای این حوزه محسوب می‌شود و نیاز به بررسی معتبر، جامع و جدیدی را می‌طلبد.

بررسی مرورهای مذکور که در خارج از ایران منتشر شده‌اند، نشان می‌دهد نه تنها به‌طور جزئی به رویکرد خلاصه‌سازی انتزاعی پرداخته‌اند، بلکه حوزه خلاصه‌سازی مبتنی بر پرس‌وجو را نیز بررسی نکرده‌اند. لذا به نظر می‌رسد مروری جامع، نظام‌مند و روشمند که بر اساس اصول موردنیاز یک مطالعه نظام‌مند باشد، انجام نشده است. از این رو، انجام مروری نظام‌مند به‌طور خاص پیرامون خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو، با جستجویی گسترده در پایگاه‌های اطلاعاتی مختلف و شناسایی همه مطالعات موجود نتایج معتبرتری از مرور ارائه می‌کند. از آنجاکه مرور نظام‌مند به‌عنوان ابزاری قوی به‌منظور ادغام یافته‌های پژوهشی متنوع عمل می‌کند و یافته‌های حاصل از آن از اعتبار قابل توجهی برخوردار است؛ در پژوهش حاضر تمامی ابعاد مطالعات حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در قالب یک بررسی نظام‌مند مورد واکاوی قرار گرفته‌اند. در این راستا، برخلاف پژوهش‌های پیشین (Yu, 2022; Nimavat & Joshia, 2017; Roy & Kundu, 2023; Alanzi & Alballa, 2023) که تک‌تک شاخص‌های مهم این حوزه از جمله روش‌ها و مدل‌های مورد استفاده را در یک سیر زمانی و پارادایمی نادیده گرفته‌اند و تنها مروری اجمالی در این حوزه انجام داده‌اند؛ در پژوهش حاضر در یک بازه زمانی از ابتدا تا زمان انتشار آخرین مطالعه در این حوزه، سیر انتشار و نوع مطالعات، رویکردها، روش‌ها، نوع یادگیری، مدل‌ها، مجموعه داده‌ها، نوع ورودی‌ها، معیارهای ارزیابی و چالش‌های موجود در این حوزه واکاوی و بررسی شده است. از این رو، برتری پژوهش حاضر نسبت به مطالعات گفته شده در بالا، علاوه بر داشتن یک روش‌شناسی مناسب و معتبر با بهره‌گیری از دستورالعمل معروف پریزما^۱، جامعیت مطالعات مورد بررسی است. با دقت زیاد سعی شد تمام مدارک معتبر منتشر شده در این حوزه را بدون محدودیت زمانی نسبت به مطالعات یادشده، مورد تحلیل و واکاوی قرار داده؛ ضعف‌های روش‌شناسی و پوشش‌دهی محدود آیت‌های آن‌ها را برطرف نموده؛ هم‌افزایی انجام‌شده را شناسایی کرده و چشم‌اندازهای آتی را بر این اساس پیش‌بینی نماید. این امر می‌تواند به‌عنوان نوآوری این مطالعه کیفی قلمداد شود. در این راستا، پژوهش حاضر نه تنها به تکمیل ادبیات موجود در این حوزه کمک شایانی می‌نماید، بلکه در راستای واکاوی همه‌جانبه سیستم‌های

^۱. PRISMA

خلاصه‌سازی پیشنهادی پژوهشگران نیز قابل توجه بوده؛ قضاوت دقیق سایر پژوهشگران درباره وضعیت موجود این حوزه را به دنبال داشته و به غنای علمی و نظری حوزه مورد مطالعه می‌افزاید. همچنین، با درک جامع از پژوهش‌های قبلی، پژوهشگران و توسعه‌دهندگان نرم‌افزارها می‌توانند تصمیمات بهتری بگیرند. زیرا کاربرد یافته‌ها و نتایج حاصل شده موجب شناساندن نقاط ضعف و قوت و شکاف‌های پژوهشی و اجرایی در ارائه مدل‌های کارآمد خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو همگام با پیشرفت فناوری‌ها و تغییرات دانش کاربران و رفتارهای اطلاع‌یابی آن‌ها، راهگشایی برای رفع چالش‌های این حوزه خواهد بود. به دنبال آن، نیل به راهکارهای بهینه‌سازی و فرایند توسعه آن‌ها تسهیل خواهد شد.

همچنین، زمینه آگاهی‌سازی و تشویق به طراحی سیستم‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجوی متون فارسی را هموار خواهد نمود. این پژوهش می‌تواند مشوقی جهت طراحی و توسعه سیستم‌های جدید، کارآمد و به‌روز برای خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در متون فارسی شده و منجر به ارتقای کیفیت ارائه اطلاعات موجود برای کاربران فارسی‌زبان و تسهیل فرآیندهای جستجو و درک محتوا شود که در نهایت به بهبود تجربه کاربری و آگاهی‌بخشی بیشتر منتهی خواهد شد.

روش پژوهش

پژوهش حاضر از نظر هدف کاربردی است. رویکرد گردآوری اطلاعات از نوع کیفی بوده و با فن مرور نظام‌مند^۱ انجام شده است. مرورهای نظام‌مند در حوزه‌ها، موضوع‌ها و مجله‌های مختلفی قابلیت انتشار دارند و استانداردها و دستورالعمل‌های مختلفی برای آماده‌سازی آن‌ها ارائه شده است. از جمله این دستورالعمل‌ها می‌توان پریزما را نام برد که در این پژوهش از آن استفاده شده است. پریزما شامل چند گام به شرح زیر است.

۱. در گام اول ابتدا کار شناسایی^۲ مطالعات با انتخاب پایگاه‌های اطلاعاتی مناسب، طراحی راهبرد جستجو و انتخاب بازه زمانی انتشار مطالعات انجام می‌شود.

۲. در گام دوم که غربالگری^۳ نام دارد، بر اساس معیارهای ورود از طریق بررسی نوع مطالعات، عناوین و چکیده‌های مدارک باقی‌مانده، موارد نامرتب و تکراری حذف می‌شوند.

۳. گام سوم مربوط به شایستگی^۴ مطالعات است که با مطالعه متن کامل هر مطالعه بالقوه مرتبط، واجد شرایط بودن آن بررسی می‌شود. هدف ارزیابی واجد شرایط بودن این است

1. Systematic Review

2. Identification

3. Screening

4. Eligibility

که اطمینان حاصل شود مطالعات منتخب با اهداف مرور نظام‌مند مطابقت دارد و از عمق و کیفیت لازم برای گنجاندن در آن بررسی برخوردار هستند. مدارکی که این گام ارزیابی را با موفقیت پشت سر می‌گذارند، پتانسیل خود را برای کمک به بینش‌های اساسی به پرسش‌های پژوهشی نشان داده و استحکام مرور نظام‌مند را افزایش می‌دهند. طراحی اشتباه مطالعه، مطالعه ناکافی، جامعه آماری، نمونه و روش نامناسب، یافته‌ها و نتایج غیر کاربردی برای مرور نظام‌مند از جمله مواردی هستند که عدم واجد شرایط بودن مطالعات را نشان می‌دهند.

۴. در گام چهارم که شمول^۱ نام دارد؛ مطالعات مناسب مشمول مرور نظام‌مند می‌شوند در اولین گام عملیاتی، پایگاه‌های اطلاعاتی اسکوپوس^۲، وب‌آوساینس^۳، آی‌تریپل‌ای^۴، کتابخانه دیجیتال ای‌سی‌ام^۵، گوگل اسکالر^۶، پروکوئست^۷، نورمگز^۸، مگیران^۹، سید^{۱۰}، سیویلیکا^{۱۱}، علم‌نت و گنج برای جستجو و شناسایی مدارک حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو برگزیده شدند. برای شناسایی کلیدواژه‌های مرتبط و مترادف با عبارت “summarization abstractive based Query” از وردنت^{۱۲} استفاده شد که نتیجه‌ای بازایی نشد. جهت بهبود جامعیت جستجو، کلیدواژه‌های موجود در مدارک مرتبط بررسی شد. جستجو‌ها نشان داد در بازایی با علامت - و بدون این علامت مطالعات مشابهی بازایی می‌شود. لذا از عبارت‌های دارای این علامت استفاده نشد. بدین ترتیب، راهبرد نهایی برای پایگاه‌های انگلیسی با استفاده از عملگر مناسب به صورت “Query oriented Abstractive summarization” OR “Query focused abstractive summarization” OR “Question focused abstractive summarization” OR “Question driven abstractive summarization” OR “Query based Abstractive summarization” برای جستجو در همه فیله‌های موجود طراحی شد.

از آنجاکه هنگام جستجو در پایگاه‌های اطلاعاتی فارسی علامت «» به منظور جستجوی عین عبارات وارد شده «خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو» یا «خلاصه‌سازی انتزاعی مبتنی بر پرسش» یا «خلاصه‌سازی انتزاعی متمرکز بر پرس‌وجو» یا «خلاصه‌سازی انتزاعی

¹. Included

². Scopus

³. Web of science

⁴. IEEE

⁵. ACM Digital Library

⁶. Google scholar

⁷. Proquest

⁸. Noormags

⁹. Magiran

¹⁰. Scientific Information Database (SID)

¹¹. Civilica

¹². Wordnet

متمركز بر پرسش» یا «خلاصه‌سازی انتزاعی پرس‌وجو محور» یا «خلاصه‌سازی انتزاعی پرسش محور» ترتیب اثر ندارد و کلیدواژه‌های موجود در این عبارات به‌صورت جدا در پایگاه مورد جستجو قرار می‌گیرند؛ کلمه عام خلاصه‌سازی برای بهبود دقت نتایج بازیابی شده مورد استفاده قرار گرفت. البته در استفاده از این کلیدواژه نیز «خلاصه» و «سازی» به‌صورت دو کلیدواژه مجزا توسط پایگاه‌ها تشخیص داده شده و مورد جستجو قرار گرفت.

با استفاده از راهبرد جستجوی مطالعات حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در پایگاه‌های مذکور و بدون محدودیت زمانی در انتشار مطالعات، در مجموع ۱۷۱۴ مطالعه بازیابی شد. در گام غربالگری پس از حذف ۱۷۴ مطالعه تکراری، معیارهای ورود برای ۱۵۴۰ مدرک باقیمانده بررسی شد. از جمله معیارهای ورود، مرتبط‌بودن مدارک و نوع مدارک بود. بررسی مرتبط‌بودن و نوع مطالعات به حذف ۱۵۰۷ مطالعه نامرتبط و مطالعاتی که از نوع گزارش، مرور، مرور فراهمایی و مطالعات سلب اعتبار شده بودند، منجر شد. در گام شایستگی، واجد شرایط بودن ۳۳ مطالعه باقیمانده با استفاده از معیارهایی به شرح زیر انجام شد که منجر به حذف ۴ مطالعه شد.

الف. در دسترس‌بودن متن کامل مدارک؛

ب. مشخص‌بودن هدف مطالعه مبنی بر ارائه مدل خلاصه‌سازی مبتنی بر پرس‌وجو به‌طور خاص با رویکرد انتزاعی؛

ج. مشخص‌بودن سهم اصلی مقاله در چهارچوب روش‌شناسی مشخص و تشریح واضح و کامل مدل ارائه‌شده؛

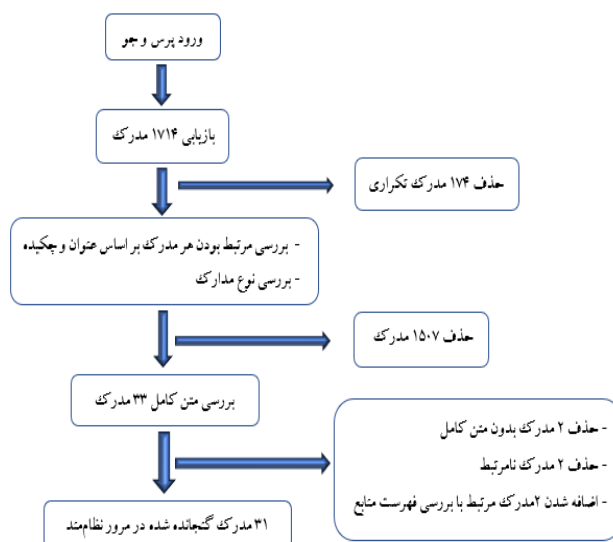
د. مشخص‌بودن معیارهای ارزیابی مدل‌ها؛

ه. اشاره به چالش‌های موجود.

بدین ترتیب، مطالعات پیرامون خلاصه‌سازی مبتنی بر پرس‌وجو با رویکرد استخراجی حذف شد. همچنین، مطالعاتی که روش‌شناسی مشخصی نداشتند و مدل پیشنهادی خود را به‌طور کامل و واضح توضیح ندادند و معیارهای ارزیابی و چالش‌های مواجه شده با آن در راستای انجام پژوهش خود را ارائه ندادند نیز از بررسی حذف شدند و باقی ماندن آن‌ها در جامعه مورد بررسی جهت مرور نظام‌مند حاضر امکان‌پذیر نشد.

در نهایت ۲۹ مدرک باقی ماند. ماحصل بررسی فهرست منابع این مطالعات به شناسایی دو مطالعه دیگر انجامید. در گام شمول، ۳۱ مطالعه به‌منظور انجام مرور نظام‌مند گنجانده شد. دلیل اینکه از ۱۷۱۴ مطالعه، ۳۱ مطالعه برای مرور برگزیده شد، اجرای دقیق چهار گام مذکور و رد شدن مدارک از صافی معیارهای ورود و خروج منتخب است که منجر به حذف ۱۵۰۷ مطالعه شد. در این بررسی تعداد زیادی مدرک نامرتبط و تعدادی مدارک از نوع گزارش،

مرور، مرور فراهمایی و مطالعات سلب اعتبار شده، مدارک با رویکرد استخراجی، مدارک با تشریح ناقص مدل و عدم ارائه معیارهای ارزیابی مورد استفاده و چالش‌های موجود از بررسی حذف شدند. حذف این تعداد مطالعه نشان‌دهنده تعهد به رعایت استانداردهای سخت‌گیرانه و انتخاب مطالعاتی است که به‌طور معناداری به اهداف کلی مطالعه کمک می‌کند. دقت‌نظر در انتخاب مطالعات تضمین می‌کند که تجزیه و تحلیل پیش‌رو بر اساس مطالعات مرتبط و با کیفیت بالا انجام شده است. در شکل ۱ گام‌های اجرا شده گفته شده در بالا، به‌منظور نیل به مدارک نهایی نشان داده شده است.



شکل ۱- فرایند اجرای دستورالعمل پریزما جهت نیل به مدارک نهایی

۳۱ مطالعه نهایی شده در محیط نرم‌افزار مکس کیودا^۱ مورد تجزیه و تحلیل قرار گرفت. مهم‌ترین مرحله از تحلیل داده‌های کیفی، مرحله کدگذاری/مقوله‌بندی است. در این مرحله پژوهشگر باید بتواند معنادارترین بخش‌های داده‌ها را سازمان‌دهی، مدیریت و بازیابی کند. درواقع آنچه در این مرحله انجام می‌گیرد فشرده ساختن داده‌ها در واحدهای قابل تحلیل از طریق ایجاد مقوله‌هایی از داده‌ها است (حریری، ۱۳۸۵). درواقع کدگذاری فرایند تحلیلی تبدیل داده‌ها به مفاهیم، فروگاهی، تبدیل مفاهیم به مقوله‌ها و تعیین روابط میان آن‌هاست. به‌طورکلی برای تحلیل داده‌های کیفی، دو شیوه تکوین استقرایی مقوله‌ها و به‌کارگیری قیاسی مقوله‌ها کاربرد دارد (بهشتی، ۱۳۹۵).

^۱ . Maxqda

در پژوهش حاضر از شیوه‌ی استقرایی استفاده شده است که شیوه‌ای جزء به کل محسوب می‌شود. بدین‌صورت که با تحلیل و شکستن متن هر مدرک، کدگذاری باز^۱ انجام گرفت. کدگذاری باز روند خردکردن، مقایسه کردن، مفهوم‌پردازی و مقوله‌بندی داده‌ها است (استراوس و کوربین^۲، ۱۳۹۰/۱۹۹۰) در این کدگذاری، مفاهیم آشکار و مستتر در متون شناسایی شد و داده‌هایی که بار معنایی یکسانی داشتند زیر یک کد قرار گرفتند. ازجمله مفاهیم مستتر می‌توان یادگیری خودنظارتی را به‌عنوان مثال نام برد که در هیچ‌کدام از متون از این نوع یادگیری به‌طور واضح یاد نشده است. اما، پژوهشگر واقف به این نکته بود که یادگیری خودنظارتی اصلی‌ترین روش در مدل‌های برت^۳، بارت^۴، پگاسوس^۵، جی‌پی‌تی^۶ و نظایر آن است و در ۱۵ مطالعه از این مدل‌ها بهره گرفته شده است.

فرایند کدگذاری برای همه موارد مرتبط با روش‌ها، انواع یادگیری و مدل‌های مورد‌استفاده در مطالعات انجام شد. درنهایت از طریق کدگذاری محوری^۷، روابط بین مفاهیم کدگذاری‌شده مشخص شده و طبقه‌بندی این کدها که به نظر می‌رسید به روش‌ها، فنون یادگیری و مدل‌های مشابه ربط دارند، صورت گرفت و کار مقوله‌بندی انجام شد. برای اطمینان از اعتبارپذیری^۷ که به واقعی‌بودن کدگذاری‌ها و مقوله‌بندی‌های انجام‌شده و اعتبار آن‌ها اشاره دارد؛ کدها و مقوله‌های یادشده در اختیار یک متخصص هوش مصنوعی قرار گرفت و بنابر نظر ایشان کار اصلاح و تکمیل انجام شد. ماحصل این یافته‌ها در جدول ۱ ارائه شده است و پاسخگویی به برخی پرسش‌های پژوهش در بخش یافته‌ها را تسهیل نمود. سایر موارد مثل سیر انتشار، نوع مدارک، نوع ورودی اسناد، مجموعه داده‌های مورد‌استفاده و چالش‌های موجود نیاز به مقوله‌بندی نداشتند و یافته‌های حاصل شده از بررسی متون پیرامون آن‌ها به‌طور مستقیم در بخش یافته‌ها ارائه شده و مورد تجزیه و تحلیل قرار گرفته است.

چنانچه در جدول ۱ آمده است، مقوله‌بندی کدهای مربوط به روش‌های مورد‌استفاده در مطالعات، به ایجاد سه مقوله منتهی شد. این مقوله‌ها عبارت‌اند از (۱) روش‌های مبتنی بر قاعده، (۲) روش‌های آماری و (۳) روش‌های مبتنی بر یادگیری ماشین. پیرامون نوع یادگیری که پژوهشگران در ارائه سیستم‌های خلاصه‌ساز خود از آن استفاده کرده‌اند نیز مقوله‌بندی کدهای شناسایی شده به ایجاد چهار مقوله به شرح (۱) یادگیری با نظارت، (۲) بدون نظارت،

۱. Free coding

۲. Bidirectional Encoder Representations from Transformers (BERT)

۳. Bidirectional and Auto-Regressive Transformers (BART)

۴. Pre-training with Extracted Gap-sentences for Abstractive Summarization Sequence-to-sequence (PEGASUS)

۵. GPT

۶. Axial coding

۷. Credibility

(۳) خودنظارتی و (۴) یادگیری انتقالی انجامید. در برخی موارد همانند یادگیری خودنظارتی، استنباط‌هایی نیاز بود تا به مقوله‌بندی درست دست پیدا کنیم. شناسایی و مقوله‌بندی کدهای پیرامون مدل‌های مورد استفاده در مطالعات نیز به ایجاد سه مقوله در این خصوص منجر شد. این مقوله‌ها عبارت‌اند از (۱) مدل‌های مبتنی بر گراف، (۲) مدل‌های مبتنی بر شبکه‌های عصبی و (۳) مدل‌های از پیش آموزش دیده. در مدل‌هایی که ما به عنوان مدل‌های مبتنی بر گراف از آن‌ها یاد کرده‌ایم؛ کدهایی شناسایی شد که مفهوم گراف با آن‌ها پیوند دارد. به‌طور مثال، کد clustering based-Graph را در نظر بگیرید که می‌گوید در مدرک x مدلی ارائه شده است که در آن از خوشه‌بندی مبتنی بر گراف استفاده شده است. در مورد هر کدام از این کدها و مقوله‌هایی که به آن‌ها اختصاص داده شده است در بخش یافته‌ها به‌طور مفصل بحث، بررسی و مقایسه انجام شد و وضعیت هر کدام را در حوزه خلاصه‌سازی واکاوی کرده‌ایم. همچنین پیرامون هم‌افزایی‌هایی که بین آن‌ها در نوآوری‌های پژوهشگران اتفاق افتاده است، صحبت شد.

جدول ۱- کدگذاری و مقوله‌بندی مدارک مورد بررسی

کد	مقوله	
Text Rank, latent semantic analysis, LSA	روش‌های مبتنی بر قاعده	روش‌های مورد استفاده
TF-IDF, TF-IDE, TF-ICE, LDA, Latent Query Distribution	روش‌های آماری	
Semantic Triples Clustering (STC), Chinese Whispers (CW) clustering, graph-based clustering, CNN, RNN, unidirectional RNN, BiLSTM, biLSTM, Gated Recurrent Units, GRU, BART, rectional RNN BERT, GPT-2, PEGASUS, T5, BERTSUM, BioBERT, QR-BERTSUM-TL, ChatGPT, mT5-small	روش‌های مبتنی بر یادگیری ماشین	
BART, BERT, GPT-2, PEGASUS, T5, BERTSUM, BioBERT, QR-BERTSUM-TL, ChatGPT, mT5-small	خودنظارتی	نوع یادگیری
weakly supervised Supervised	نظارت شده	
unsupervised	بدون نظارت	
transfer learning	یادگیری انتقالی	

کد	مقوله	
Textrank, Pagerank, Semantic Triples Clustering (STC), graph-based clustering	مدل های مبتنی بر گراف	مدل های مورد استفاده
CNN, RNN, unidirectional RNN, Bidirectional RNN, Gated Recurrent Units, GRU, LSTM, biLSTM	مدل های مبتنی بر شبکه های عصبی	
BART, BERT, GPT-2, PEGASUS, T5, BERTSUM, BioBERT, QRBERTSUM-TL, ChatGPT, mT5-small	مدل های از پیش آموزش دیده	

یافته ها

الف. سیر انتشار و نوع مطالعات

سیر انتشار مطالعات بررسی شده نشان می دهد این مطالعات که همه غیرفارسی هستند، فراوانی و قدمت زیادی ندارند و در بازه زمانی محدود ۱۰ سال اخیر منتشر شده اند. یک روند صعودی و نزولی در انتشار این مطالعات دیده می شود. به طور کلی، شاهد افزایش انتشار آن ها از سال ۲۰۱۵ تاکنون هستیم. این افزایش نمایانگر رشد تولید علم در این زمینه است. از سوی دیگر، نوسانات در این روند می تواند نشانه ای از تحولات فناورانه یا تغییرات در نیازهای پژوهشی باشد. البته با توجه به ظهور مدل های زبانی بزرگ که تحولی عظیم در پردازش زبان طبیعی و تعاملات انسان و ماشین ایجاد کرده اند، افزایش روزافزون کمی و کیفی آن ها در عرصه های مختلف و نقش تأثیرگذار این نوع خلاصه سازی در کنار سایر روش های به کاررفته در آن ها، انتظار افزایش انتشار مطالعات بیشتر در این حوزه منطقی به نظر می رسد. مطالعاتی مشابه با پژوهش اینیو و همکاران^۱ (۲۰۲۱) که در آن به نقش اساسی خلاصه سازی مذکور در سیستم پرسش و پاسخ پیشنهادی اشاره شده است.

بررسی نوع مطالعات نشان می دهد که علاوه بر پایان نامه، در نوع مقاله^۱ و مقاله فراهمایی^۲ که از انواع مدارک موجود در پایگاه های استنادی بوده اند، انتشار یافته اند. بخش اعظم مطالعات (۱۹ مدرک) به صورت مقاله منتشر شده اند. ۹ مقاله فراهمایی نیز انتشار یافته است و پایان نامه با فراوانی^۳ مطالعه کمترین توزیع را در این بررسی داشته است. به نظر می رسد این حوزه موضوعی در پژوهش های تحصیلات تکمیلی در قالب پژوهش های پایان نامه ای جایگاه درخوری ندارد و هنوز به اندازه کافی مورد توجه قرار نگرفته است. بر این اساس، می توان

^۱. Inoue et al.

^۲. Article

^۳. Proceeding paper

چنین نتیجه‌گیری کرد که این حوزه بیشتر در قالب مقالات علمی در حال توسعه است. این امر می‌تواند به دو دلیل عمده باشد: الف) نیاز به تأیید علمی توسط متخصصان این حوزه و انتشار در مجلات معتبر به‌عنوان یکی از معیارهای اصلی اعتبار علمی؛ و ب) توان بالای پژوهش‌ها در این حوزه برای تأثیرگذاری سریع و کاربردی در دنیای واقعی. البته با توجه به نقش پررنگ این حوزه در سیستم‌های نوظهور کنونی و پتانسیل‌های بالقوه فراوانی که می‌تواند در زمینه‌های مختلف علمی و صنعتی ارائه دهد، شایسته است در پایان‌نامه‌های تحصیلات تکمیلی گروه‌های آموزشی ذی‌ربط در دانشگاه‌ها و مراکز آموزش عالی توجه بیشتری به این حوزه شود تا زیر نظر گروه‌های پژوهشی متخصص به غنای ادبیات علمی آن کمک شده و زمینه پیشرفت‌های جدیدتر فراهم شود. در این راستا، ارائه حمایت‌های مالی و معنوی از پایان‌نامه‌هایی که در این حوزه انجام شود می‌تواند تا حدودی راهگشا باشد. همچنین، تقویت همکاری و ارتباطات دانشگاه یا صنعت و تعریف پروژه‌های مشترک در این حوزه می‌تواند مشوقی برای انجام پایان‌نامه‌های بیشتر و پیاده‌سازی سیستم‌های پیشنهادی با داده‌های واقعی موجود در بستر وب باشد که به کاربرپذیری آن‌ها کمک شایانی خواهد نمود.

ب. رویکردهای مورد استفاده

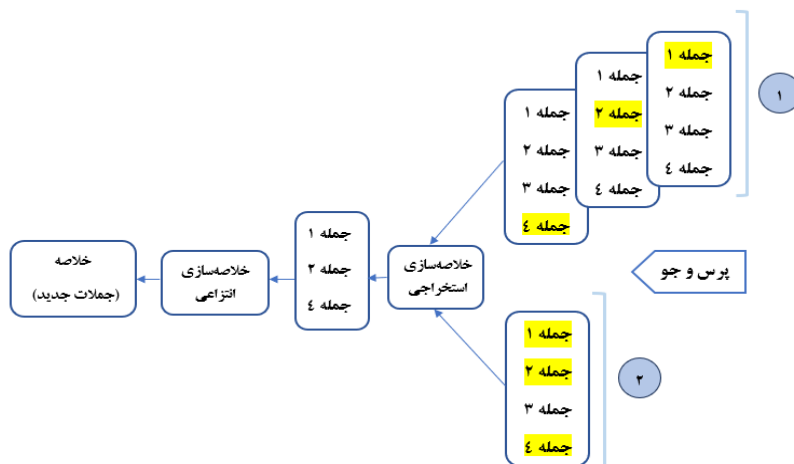
در این مطالعات از دو رویکرد انتزاعی و استخراجی-انتزاعی استفاده شده است. رویکرد انتزاعی که به آن رویکرد مستقیم نیز گفته می‌شود در ۲۸ مطالعه مورد استفاده قرار گرفته است. از این ۲۸ مطالعه نیز ۳ مورد علاوه بر رویکرد انتزاعی، یک رویکرد استخراجی نیز ارائه داده‌اند (Deng et al., 2023; Kumaravel & Sankaranarayanan, 2021; Girthana & Swamy-nathan, 2019).

هر دو رویکرد استخراجی و انتزاعی مزایا و معایب خود را دارند. اجرای خلاصه‌سازی استخراجی نسبتاً آسان‌تر از خلاصه‌سازی انتزاعی است. خلاصه‌سازی استخراجی که شامل انتخاب جملات یا بخش‌هایی از متن اصلی به‌منظور ایجاد خلاص‌های مختصر و مفید است، ممکن است قابلیت درک کاربر را از متن اصلی کاهش دهد. زیرا به‌اندازه درک یک انسان از متن و گزینش جملاتی از آن، کارآمد نیست. این نکته را می‌توان به کمبود غنای محتوایی آن نسبت داد. در مقابل، رویکرد انتزاعی نتایج غنی‌تر و متناسب با نیازهای کاربر ارائه می‌دهد. اما، اجرای آن پیچیدگی بیشتری دارد. این رویکرد از خلاصه‌سازی به‌ویژه زمانی که چندین سند متنی پیش‌روی کاربر است، می‌تواند به سرعت اطلاعات جدید و معناداری را که می‌تواند به‌سان برآیندی از مطالعه و تجزیه و تحلیل انسانی آن متون باشد ارائه دهد. به‌طوری‌که مفهوم و معنای کل متون را به شیوه‌ای مؤثر به کاربر منتقل نماید. البته این رویکرد

به الگوریتم‌های پیچیده و مدل‌های یادگیری عمیق نیاز دارد که فرایند طراحی و آموزش آن‌ها زمان‌بر و پرهزینه است. همچنین، برای تولید خروجی‌های با کیفیت به داده‌های آموزشی فراوان نیاز دارد.

انتخاب بین رویکردهای انتزاعی و استخراجی وابسته به نوع کاربرد و نیاز کاربر است. در شرایطی که سرعت در ارائه خلاصه اطلاعات مهم باشد، روش استخراجی ممکن است ترجیح داده شود. در حالی که در موقعیت‌هایی که درک عمیق و غنی از محتوا مورد نظر باشد رویکرد انتزاعی مورد استفاده قرار می‌گیرد.

ترکیب رویکردهای استخراجی و انتزاعی با تقویت مزایا و تضعیف معایب آن‌ها منجر به روش‌های ترکیبی برای خلاصه‌سازی متن می‌شود یاداو و همکاران^۱، (۲۰۲۲) که به آن رویکرد انتزاعی غیرمستقیم هم می‌گویند. برخلاف رویکرد مستقیم که یک مرحله‌ای است، رویکرد غیرمستقیم در دو مرحله انجام می‌گیرد، ابتدا خلاصه‌سازی استخراجی انجام می‌شود. خروجی حاصل از مرحله اول به عنوان ورودی مرحله دوم محسوب شده و روی آن خلاصه‌سازی انتزاعی صورت می‌پذیرد. رویکرد غیرمستقیم به دلیل داشتن دو مرحله، می‌تواند فرایند خلاصه‌سازی را کارآمدتر کند. مرحله اول به تمرکز بر اطلاعات کلیدی کمک می‌کند و مرحله دوم از طریق تجزیه و تحلیل و استنباط به غنی‌سازی اطلاعات می‌پردازد. این مراحل به صورت شماتیک بصری برای خلاصه‌سازی تک سندی و چند سندی در شکل ۲ نشان داده شده است.



شکل ۲- خلاصه‌سازی استخراجی- انتزاعی چندسندی و تک سندی

^۱. Yadav et al.

ترکیب دورویکرد از طریق خلاصه‌سازی استخراجی به‌عنوان مرحله اول و خلاصه‌سازی انتزاعی به‌عنوان مرحله دوم، نویدبخش بهینه‌سازی‌های قابل توجهی در فرایند خلاصه‌سازی اطلاعات است. این رویکرد غیرمستقیم با غنی‌سازی محتوا به تحلیل و برداشت‌های عمیق‌تر منجر می‌شود. چنین الگویی می‌تواند با وضوح بیشتری در درک کلان‌تر از موضوعات پیچیده کمک کند و زمینه‌ساز تولید نتایج قابل اعتمادتر در کاربردهای مختلف باشد. با پیشرفت‌های اخیر در حوزه هوش مصنوعی می‌توان انتظار داشت که در آینده شاهد توسعه کمی و کیفی این رویکرد و ظهور خلاصه‌سازهای تعاملی قدرتمندی باشیم.

ج. روش‌های مورد استفاده

مرور مطالعات نشان می‌دهد مسیر تکامل خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو روش‌های متفاوتی را طی کرده است. از روش‌های مبتنی بر قاعده^۱ گرفته تا روش‌های آماری^۲ و یادگیری ماشین^۳. ویشواکرما و همکاران^۴ (۲۰۲۱) می‌گویند روش‌های خلاصه‌سازی مبتنی بر قاعده در توسعه سیستم‌های خلاصه‌سازی، با استفاده از قواعد و الگوهای زبانی از پیش تعریف‌شده برای تقطیر جوهر محتوای متنی از جمله روش‌های بنیادی بوده‌اند ([Wibawa & Kurniawan, 2024](#)). یکی از فنون رایج مبتنی بر قاعده، روش امتیازدهی جمله است که بر اساس معیارهای مختلف مانند طول جمله، فراوانی کلمات کلیدی و وجود ویژگی‌های زبانی خاص به جملات امتیاز می‌دهد ([Allahyari et al., 2017](#)). به‌طور مثال، یک خلاصه‌کننده مبتنی بر قاعده، وظیفه جمع‌بندی مقالات حوزه بازیابی دانش را دارد. در این مورد، روش امتیازدهی جمله به جملاتی که حاوی اطلاعات بیشتر هستند، نمره بالاتری اختصاص می‌دهد. از جمله روش‌های مبتنی بر قاعده به‌کاررفته در مطالعات بررسی شده می‌توان تکست‌رنک^۵ و تحلیل معنایی پنهان^۶ را نام برد ([ShafieiBavani et al., 2016](#); [Kumaravel & Sankaranarayanan, 2021](#)). [Girithana & Swamynathan, 2019](#) تکست رنک مبتنی بر نظریه گراف است و بر اساس رتبه‌بندی که بر روی گره‌ها انجام می‌دهد جملات مهم متن را انتخاب نموده و تولید خلاصه را انجام می‌دهد. تکست‌رنک نسبت به تحلیل معنایی پنهان که جهت استخراج ساختارهای معنایی متن براساس کلمات مشاهده‌شده عمل می‌کند؛ قابلیت سادگی و انطباق‌پذیری بیشتری دارد و اهمیت جملات را از طریق روش‌های مبتنی بر گراف مشخص می‌کند.

¹. Rule-based

². Statistical

³. Machine learning

⁴. Vishwakarma et al.

⁵. Textrank

⁶. latent semantic analysis

درحالی که تحلیل معنایی پنهان در شناسایی روابط معنایی آن قدر قوی نیست. از سویی دیگر، با وجود پیچیدگی محاسباتی که دارد منجر به خلاصه سازی با انسجام کمتر می شود. مقادیر به دست آمده برای معیارهای ارزیابی $LCS, 2, 1, ROUGE$ برای تکست رنگ نسبت به تحلیل معنایی پنهان در مطالعه کومارول و سانکارانایاران^۱ (۲۰۲۱) گواهی بر این امر است. این مقایسه، نشان دهنده نیاز به توجه بیشتر به انتخاب روش های مناسب خلاصه سازی متناسب با نیازهای کاربران و نوع متون است. به طور کلی، در شرایطی که وضوح و دقت در تجزیه و تحلیل محتوا اهمیت دارد، روش های مبتنی بر قاعده، به ویژه تکست رنگ، می توانند به عنوان گزینه های کارآمد مورد استفاده قرار گیرند.

در عصر اطلاعات استفاده مؤثر از روش های آماری در پردازش زبان طبیعی اهمیت فراوانی پیدا کرده است. نظر به اینکه این روش ها به ویژه در داده های حجیم که شامل اطلاعات متنوع و پیچیده هستند، می توانند به شناسایی و استخراج محتوای کلیدی کمک کنند و نقش مهمی در حوزه خلاصه سازی متون ایفا نمایند (Giarelis et al., 2023). طبق بررسی ها از معیارهایی مانند فراوانی کلمات، جملات و غیره استفاده می کنند. این روش ها مستقل از هر زبانی هستند. به طوری که اگر خلاصه کننده با استفاده از این روش ها توسعه یابد، می تواند متون را به هر زبانی خلاصه کند (Gambhir & Gupta, 2017). از جمله این روش ها می توان به فرکانس معکوس فرکانس سند^۲ (Esteva et al., 2021; Deng et al., 2023; Israel et al., 2015; Abdullah, 2020; Baumel et al., 2018; Mehdad et al., 2014) اشاره کرد. فراوانی کلمه و معکوس فراوانی سند آمارهای عددی هستند که میزان اهمیت یک کلمه در یک سند را نشان می دهند. بدین صورت که فراوانی کلمه تعداد دفعاتی است که یک عبارت در سند رخ می دهد و معکوس فراوانی سند معیاری است که وزن عبارتی را که اغلب در مجموعه آمده اند کاهش داده و وزن اصطلاحاتی که به ندرت آمده اند را افزایش می دهد. سپس جملاتی که امتیاز بالایی دارند در خلاصه ذکر می شوند. یکی از مشکلات این روش این است که گاهی اوقات جملات طولانی تر به دلیل اینکه تعداد کلمات بیشتری دارند، امتیاز بالایی می گیرند که ممکن است منجر به اختلال در کیفیت خلاصه سازی شود. روش دیگر یک مدل احتمالاتی از یک مجموعه متن به نام تشخیص دیریکله پنهان کومارول و سانکارانایاران^۳ (۲۰۲۱) است. این مدل که برای تشخیص موضوعات پنهان در متون مورد استفاده قرار می گیرد؛ فرض می کند که هر سند می تواند از چندین موضوع تشکیل شده باشد و هر موضوع نیز دارای توزیع خاصی از کلمات است. این مدل سعی می کند این توزیع ها را شناسایی کند و تعیین کند هر

¹. Kumaravel & Sankaranarayanan

². TF-IDF

³. Latent Dirichlet Allocation (LDA)

کلمه به کدام موضوع تعلق دارد. پس از شناسایی موضوعات پنهان، با استفاده از جملات یا بخش‌هایی از متن که بیشتر به این موضوعات مربوط هستند، خلاصه متن تولید می‌شود. تشخیص دیریکله پنهان به دلیل توانایی که در شناسایی ساختارهای پنهان در داده‌های متنی دارد، یک راه‌حل مفید در زمینه پردازش زبان طبیعی به‌طور کلی و خلاصه‌سازی متون به‌طور خاص تبدیل شده است و چون می‌تواند جستجوهای عمیق‌تری در ساختار متون انجام دهد و توان تشخیص موضوعات پنهانی را دارد که فراوانی کلمه و معکوس فراوانی سند ناتوان از شناسایی آن‌ها است؛ نسبت به آن کارایی بیشتری در امر خلاصه‌سازی دارد. در مقابل، اگر هدف فقط شناسایی کلیدواژه‌ها در متون کوتاه باشد، فراوانی کلمه و معکوس فراوانی سند ممکن است مناسب‌تر و سریع‌تر باشد. به‌طور کلی، ترکیب این دو روش می‌تواند از نقاط ضعف هر یک بکاهد. آنجا که فراوانی کلمه و معکوس فراوانی سند ممکن است به‌خوبی نتواند موضوعات پنهان را شناسایی کند، تشخیص دیریکله پنهان در شناسایی روابط بین کلمات و توزیع موضوعات قوی‌تر است. توزیع پنهان پرسش^۱ نیز روشی است که در حوزه پردازش زبان طبیعی و به‌ویژه در مدل‌سازی جستجو و بازیابی اطلاعات به کار می‌رود. این مفهوم به تلاش برای شناسایی و مدل‌سازی توزیع‌های نهفته از درون پرسش‌ها و نیازهای اطلاعاتی کاربران اشاره دارد که در داده‌های جستجو و تعاملات کاربران قابل مشاهده نیستند، اما می‌توانند رفتار و جستجوی آن‌ها را به طرز معناداری توصیف کنند.

مرور انجام‌شده نشان می‌دهد در برخی مطالعات هم‌افزایی این روش‌ها دیده می‌شود که نشان‌دهنده تلاشی پویا و نوآورانه به‌منظور ارائه مدل‌های خلاصه‌سازی مؤثرتر و کارآمدتر است و در عصر اطلاعات که حجم زیادی از داده‌ها موجود است، انجام آن شایسته و بایسته می‌نماید. از آنجاکه هر روش به جنبه‌های خاص از متن توجه دارد؛ هم‌افزایی‌هایی انجام‌شده افزایش دقت را به دنبال داشته است. همچنین، ترکیب روش‌ها تأثیر خطاهای احتمالی هر روش به‌تنهایی را کاهش داده است. علاوه بر این، به دلیل اینکه هر روش می‌تواند به جنبه‌های مختلف اطلاعات توجه کند، استفاده هم‌زمان از آن‌ها تصویر جامع‌تری از متن نشان می‌دهد. به‌عبارت دیگر، این روش‌های ترکیبی می‌توانند ابعاد مختلف اطلاعات را از زوایای گوناگون تحلیل کنند و این خود به درک بهتر و دقیق‌تری از داده‌ها منجر می‌شود و با داشتن قابلیت تحلیل چندگانه، در سامانه‌های پیشرفته پتانسیل به‌کارگیری خواهد داشت. مطالعه گیرتانا و سامیناتان^۲ (۲۰۱۹) که در آن از تکست‌رنک و فراوانی کلمه و معکوس فراوانی سند استفاده شده است، نشان‌دهنده تلاش پژوهشگران در این خصوص است. جایی که پرس‌وجو از

^۱. Latent Query Distribution

^۲. Girthana & Swamynathan

جزئیات جدید ارائه شده توسط تحلیلگر ثبت اختراع و هستی شناسی دامنه گسترش می یابد از این روش ها استفاده شده است. همچنین برای پالایه و خلاصه کردن مجموعه اسناد بازیابی شده به صورت استخراجی و انتزاعی در سیستم پیشنهادی از این روش ها استفاده شد روش دیگر که استفاده از آن در خلاصه سازی انتزاعی مبتنی بر پرس و جو نسبت به دو روش گفته شده در بالا، قدمت کمتری دارد، یادگیری ماشین است و به عنوان یک روش نوآورانه و کارآمد، پیشرفت چشمگیری در این حوزه به شمار می آید. این تحول به وضوح نشان دهنده نیاز به ابزارهایی جدید برای پردازش حجم عظیم اطلاعات جهت غلبه بر محدودیت های روش های پیشین خلاصه سازی است. استفاده از روش های یادگیری ماشین به منظور خلاصه سازی متن نشان دهنده تغییر پارادایم در نحوه تولید خلاصه است. به اعتقاد کوهن^۱ (۱۹۶۲/۱۳۹۳) در کتاب ساختار انقلاب های علمی زمانی که وضعیت موجود پاسخگوی مسائل جامعه علمی نباشد اعوجاج هایی رخ می دهد که باعث انقلاب می شود. این انقلاب باعث تغییر پارادایم حاضر شده و پارادایم جدید با قابلیت حل مسائل نوظهور، یا جایگزین پارادایم قبل شده یا در کنار پارادایم قبل و در حالت غالب به حل مسائل می پردازد کوهن^۲ (۱۹۶۲/۱۳۹۳). در حوزه خلاصه سازی نیز روش های قبل قادر به خلاصه سازی سریع و با کیفیت بالا برای حجم عظیم متون نبوده و دقت و صحت رضایت بخشی را با تولید خلاصه ها ارائه نمی دادند. به طور مثال تحلیل معنایی پنهان که به عنوان یک روش مبتنی بر قانون، به منظور استخراج ساختارهای معنایی پنهان جملات و عبارات استفاده می شود، محدودیت هایی داشت. مانند عدم تجزیه و تحلیل و توجه به ترتیب کلمات، روابط نحوی و ریخت شناسی های^۲ متون. افزون بر این، تنها بر اطلاعات موجود در سند ورودی تکیه نموده تا دانش بیرونی (Mridha et al., 2021). از این رو انقلاب پیش آمده زمینه به کارگیری یادگیری ماشین در حوزه خلاصه سازی متن را فراهم نموده و تغییر پارادایم در این حوزه را به دنبال داشته است. این اقتباس از نظریه کوهن درباره تغییر پارادایم نه تنها در علم، بلکه در فناوری اطلاعات نشان می دهد که چگونه روش های یادگیری ماشین به دنبال رفع ناکارآمدی های روش های قدیمی در حوزه خلاصه سازی بوده اند. این روش به عنوان یک پارادایم جدید نه تنها قابلیت های بیشتری ایجاد کرده بلکه به بررسی و حل مسائل موجود به شکلی کارآمدتر کمک نموده است و می توان گفت پاسخگوی نیازهای فعلی و آینده نیز خواهد بود. زیرا نسبت به روش های پیشین، خلاصه های تولید شده توسط آن از انسجام و روان بودن بیشتری برخوردارند و به لحاظ زبانی طبیعی تر هستند. همچنین، در حفظ

¹. Kuhn

². Morphology

اطلاعات مهم متون موفق‌تر عمل می‌کنند و می‌توانند خلاصه‌هایی تولید کنند که به‌طور دقیق‌تر معنای اصلی متن را منتقل نمایند. یادگیری ماشین با تکیه بر الگوریتم‌های خود به ساختارهای زبانی و معنایی بیشتر توجه نموده و الگوها و روابط را از داده‌ها یاد می‌گیرد و خلاصه‌هایی با دقت و صحت رضایت‌بخش ارائه می‌دهد. همچنین، قابلیت به‌روزرسانی و بهبود مداوم دارد. این در حالی است که روش‌های قدیمی معمولاً ایستا و غیرقابل تطابق با تحولات جدید هستند. یادگیری ماشین به دلیل عملکرد خود در قابلیت انطباق و نوآوری مداوم، پتانسیل بالایی برای توسعه سیستم‌های خلاصه‌سازی متون دارد. برخلاف روش‌های مبتنی بر قانون و روش‌های آماری، یادگیری ماشین به‌ویژه معماری‌های یادگیری عمیق مانند ترنسفورمرها، می‌توانند الگوهای پیچیده و روابط معنایی را مستقیماً از داده‌ها یاد بگیرند و متن را در زبان‌ها و حوزه‌های مختلف با حداقل مداخله انسانی خلاصه کنند. علاوه بر این، این توانایی را دارد که با یادگرفتن از نمونه‌های آموزشی، قواعدی برای خلاصه‌سازی متن‌ها استخراج کنند. به‌عنوان مثال، می‌تواند تشخیص دهد که چه عبارات و جملاتی برای یک خلاصه مناسب‌تر هستند. لذا بررسی‌ها نشان می‌دهد در پژوهش‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو از آن بهره گرفته شده است. گواه این امر استفاده از ترنسفورمر در برخی از مطالعات بررسی شده در پژوهش حاضر (Arthur et al., 2023; Laskar et al., 2023; Deng et al., 2023; Abdullah & Chali, 2020; Laskar et al., 2020; Abdullah, 2020; Nishida et al., 2019; Wang et al., 2022; Ritharson et al., 2023; Laskar et al., 2023; Vig et al., 2021; Su et al., 2020; Laskar et al., 2021; Su et al., 2021; Xu & Lapata, 2021b; Polash, 2019; Xu & Lapata, 2021a; Pasunuru et al., 2021; Xu & Lapata, 2020).

ترنسفورمر مبتنی بر مکانیسم توجه^۱ است که در سال ۲۰۱۷ توسط وسوانی و همکاران^۲ معرفی شد. مکانیسم توجه را می‌توان به‌عنوان نگاشت یک پرس‌وجو و مجموعه‌ای از جفت‌های کلید-مقدار به یک خروجی توصیف کرد که در آن پرس‌وجو، کلیدها، مقادیر و خروجی همه بردار هستند. خروجی به‌عنوان مجموع وزنی مقادیر حساب شده و وزن اختصاص داده‌شده به هر مقدار توسط تابع سازگاری پرس‌وجو با کلید مربوطه محاسبه می‌گردد (Vaswani et al., 2017). ترنسفورمر به یک بلوک اساسی برای بسیاری از مدل‌های پیشرفته در پردازش زبان طبیعی تبدیل شده است (Abdullah, 2020). پژوهشگران زیادی در حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو از آن در پژوهش‌های خود استفاده کرده‌اند. در مطالعه لاسکار و همکاران^۳ (۲۰۲۳) از رمزگذار-رمزگشای مبتنی بر ترنسفورمر استفاده شده است. این پژوهشگران از مدل از پیش

^۱. Attention mechanism

^۲. Vaswani et al.

آموزش دیده بارت یا ترنسفورمر دو جهته و خودکار پس رونده استفاده کرده‌اند. بارت در تولید خلاصه‌ها از فنون مانند حذف نویز، رمز گذاری خودکار و غیره استفاده کرده و به خروجی‌های لایه نهایی رمز گذار توجه می‌کند (Vig et al., 2021) و باعث کیفیت بخشی به نتیجه خلاصه سازی می‌شود. وانگ و همکاران (۲۰۲۲)^۱ مکانیسم توجه دومر حلهای^۲ را برای وابستگی‌های ساختاری در بین جملات استفاده نمودند. در مدلی که نیشیدا و همکاران^۳ (۲۰۱۹) ارائه کردند بلوک رمز گذار ترنسفورمر^۴ باز نمایی سؤالات و متون را پردازش نموده و بلوک رمز گشای ترنسفورمر^۵ کلمات پاسخ را بر اساس ورودی تولید کرد. در مطالعه عبدالله^۶ (۲۰۲۰) نمایش‌های رمز گذار دوطرفه از ترنسفورمر (برت) برای بهبود خلاصه‌ها با ترکیب نمایش‌های کلمه و جمله در یک ترنسفورمر جامع استفاده شد. جانشانی کلمه^۷ در ترنسفورمر برای انتقال معنای هر نشانه استفاده شد. جانشانی بخش^۸ برای شناسایی جملات مورد استفاده قرار گرفت و جانشانی موقعیت^۹، موقعیت کلمه را تعیین نمود و به درک داده‌های ورودی کمک کرد. با گنجاندن جانشانی‌های مختلف، پیروی از چهار چوب رمز گذار رمز گشا و بهینه سازی مدل برای بهبود عملکرد در وظایف خلاصه سازی انتزاعی به عملکرد خوبی دست یافته است. در مطالعه سو و همکاران^{۱۰} (۲۰۲۱) رمز گذار ترنسفورمر با استفاده از لایه‌های توجه به خود^{۱۱} ساخته شد. لایه رمز گشای ترنسفورمر نیز ترکیبی از توجه به خود و توجه رمز گذار رمز گشا^{۱۲} است که به کلمات ورودی اجازه می‌دهد به یکدیگر توجه کرده و بهبود خلاصه سازی انتزاعی متمرکز بر پرس و جور تسهیل کنند. از ترنسفورمر در مطالعه پولاش^{۱۳} (۲۰۱۹) نیز برای ارائه مدلی بدون تکیه بر شبکه‌های عصبی باز گشتی^{۱۴} یا شبکه‌های عصبی پیچشی^{۱۵} استفاده شده است. یکی از اهداف اصلی استفاده از آن کاهش محاسبات در مدل است که از طریق مکانیسم توجه چند سر^{۱۶} انجام شد. ترنسفورمر مدل را قادر ساخت تا کلمات را به طور مؤثر در طول مراحل مختلف انتزاع کند و خلاصه باکیفیتی

¹. Wang et al.

². Two-step attention mechanism

³. Nishida et al.

⁴. Transformer encoder

⁵. Transformer decoder

⁶. Abdullah

⁷. Word embedding

⁸. Segmentation embedding

⁹. Position embedding

¹⁰. Su et al.

¹¹. Self-attention

¹². Encoder-decoder attention

¹³. Polash

¹⁴. Recurrent Neural Network (RNN)

¹⁵. Convolutional Neural Network (CNN)

¹⁶. Multi-head attention mechanism

تولید نماید. در سایر مطالعات مذکور نیز ترنسفورمر و مکانیسم توجه نقش بسزایی در تولید خلاصه‌های انتزاعی مبتنی بر پرس‌وجو ایفا نموده‌اند.

مرور انجام‌شده نشان می‌دهد که ترنسفورمر از جدیدترین معماری‌های موجود است که نقش حیاتی در پردازش زبان طبیعی ایفا نموده است و به‌عنوان نقطه عطفی در خلاصه‌سازی شناخته شود. مدل‌های یادگیری ماشین با استفاده از ترنسفورمها می‌توانند وابستگی‌های دوربرد و روابط معنایی درون‌متن را به تصویر بکشند و خلاصه‌های منسجم‌تر و مرتبط‌تری را تولید کنند. برخلاف مدل‌های دیگر مثل شبکه‌های عصبی بازگشتی، پیچشی و... که به ترتیب ورودی پردازش انجام می‌دهند؛ ترنسفورمر می‌تواند به‌طور هم‌زمان به همه کلمات ورودی توجه کند. این امر به‌ویژه در مورد متون بلند و پیچیده از اهمیت بیشتری برخوردار است، چراکه امکان تحلیل عمیق‌تری از نسبت‌ها و روابط معنایی بین اجزای مختلف متن را فراهم می‌آورد و می‌تواند به‌طور مؤثری اطلاعات را با استفاده از شناخت عمیق از ساختارهای زبانی و معنایی پردازش کند. لذا، استفاده از ترنسفورمر در خلاصه‌سازی به‌عنوان یک جایگزین برای شبکه‌های عصبی بازگشتی و پیچشی مزایای قابل‌توجهی دارد. به‌طور مؤثری اطلاعات را با استفاده از شناخت عمیق از ساختارهای زبانی و معنایی پردازش کنند. از آنجاکه ترنسفورمر می‌تواند به‌صورت موازی عمل کند، مواردی مانند سرعت پردازش و کاهش سختی محاسبات به طرز چشمگیری بهبود می‌یابد.

مدل‌هایی مانند بارت و برت که اشاره شده، عملاً به‌نوعی آموزش قبلی دیده‌اند که بر روی مجموعه‌های داده وسیع و متنوعی انجام شده است. این امر به مدل‌ها امکان می‌دهد که درک عمیق‌تری از ساختار زبان و معنای واژه‌ها پیدا کنند. به‌کارگیری این مدل‌ها در خلاصه‌سازی اجازه می‌دهد تا خروجی‌ها همچنان به لحاظ معنایی غنی و مرتبط با محتوای اصلی متن باشند. با به‌کارگیری مکانیسم توجه دو مرحله‌ای در مطالعه وانگ و همکاران (۲۰۲۲) مدل توانسته است با تحلیل وابستگی‌های ساختاری میان جملات به فهم بهتری از ساختار کلی متن برسد و درنهایت نکات کلیدی به‌طور مؤثرتری استخراج شوند و در تولید خلاصه موردنظر استفاده شوند. مکانیسم توجه به خود نیز به مدل‌های پیشنهادی اجازه داده است که کلمات و عبارات درون خود متن به یکدیگر توجه کنند. به این ترتیب روابط معنایی بیشتری برقرار شده و تحلیل عمیق‌تری اتفاق افتاده است.

بر اساس تجزیه و تحلیل و جمع‌بندی از مطالعات به نظر می‌رسد آینده پژوهش‌ها در این حوزه بیشتر بر پایه استفاده از ترنسفورمها استوار خواهد بود. این پژوهش‌ها نه تنها به بهبود کیفیت خروجی‌ها کمک کرده، بلکه زمینه‌ساز نوآوری‌های مداوم بیشتری در طراحی الگوریتم‌های هوشمند و نوین در پردازش زبان طبیعی خواهند بود. افزون بر آن، بهره‌گیری از

مکانیسم‌های مختلف توجه در خلاصه‌سازی انتزاعی به یک روند روبه‌رشد تبدیل شده است. بررسی موارد مختلف و نتایج مثبت حاکی از نگرش نظم یافته و هدفمند پژوهشگران در این حوزه است. به‌طور خاص، تأثیر مکانیسم‌های توجه بر کیفیت خلاصه‌های خروجی، یکی از زمینه‌های پژوهشی مهم است که انتظار می‌رود در آینده پیشرفت‌های خوبی داشته باشد. نکته حائز اهمیت اینکه از انواع مختلف مکانیسم‌های توجه در این مطالعات استفاده شده است. در شکل ۳ ابر عبارات پیرامون مکانیسم‌های توجه استفاده‌شده در مطالعات مورد بررسی نشان داده شده است.



شکل ۳- ابر عبارات پیرامون مکانیسم‌های توجه

همان‌طور که در شکل بالا مشهود است، ضمن پررنگ بودن ترنسفورمر و نقش محوری آن، بیشترین میزان استفاده از مکانیسم‌های توجه در مطالعات بررسی شده به ترتیب مربوط به بارت، توجه رمزگذار-رمزگشا و خودتوجهی است. از سایر مکانیسم‌های مورد استفاده در مطالعات می‌توان مکانیسم توجه مبتنی بر تنوع^۱، توجه نرم^۲، توجه افزایشی^۳، هم توجهی^۴ و خودتوجهی^۵ چند سر را نام برد. از پگاسوس و T5^۶ نیز در برخی مطالعات استفاده شده است که مبتنی بر ترنسفورمرها هستند. نتیجه این بررسی نشان‌دهنده تنوع و کارایی بالای مکانیسم‌های توجه در پیشرفت‌های اخیر حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو است و بر اهمیت انتخاب مناسب این مکانیسم‌ها در طراحی سیستم‌های خلاصه‌ساز نوین تأکید می‌کند.

¹. Diversity driven attention

2. Soft attention

³. Additive attention

4. Co-attention

⁵. Multi head self-attention

6. Text-to-Text Transfer Transformer

بررسی انجام‌شده گویای هم‌افزایی فنون مختلف از جمله مکانیسم‌های مختلف توجه در برخی مدل‌های ارائه شده است که با نوآوری‌های پژوهشگران اتفاق افتاده است. مدل‌ها می‌توانند توجهات مختلفی را به بخش‌های مختلف متون معطوف کنند. این موضوع بسیار اهمیت دارد، زیرا مدل باید بتواند با نهایت دقت بخش‌های کلیدی و اطلاعات مهم را شناسایی کند. به‌طور مثال، در مطالعه پولاش (۲۰۱۹) از چند مکانیسم استفاده شده است. از مکانیسم خودتوجهی برای تمرکز بر قسمت‌های مختلف جملات استفاده شده است. مکانیسم توجه در مدل رمزگذار-رمزگشا به رمزگشا اجازه می‌دهد تا با استفاده از پرس‌وجوهای از لایه رمزگشای قبلی، کلیدها و مقادیر خروجی رمزگذار به تمام کلمات موجود در دنباله‌های ورودی توجه نماید. اعمال یک لایه توجه نرم بر روی حالت‌های پنهان نیز به‌منظور تمرکز بر بخش‌های خاصی از سند ورودی انجام شد. این مکانیسم به مدل اجازه داد تا به‌طور انتخابی بر روی اطلاعات مربوطه از ورودی رمزگذاری‌شده در طول فرایند تولید خلاصه تمرکز کند و کیفیت خلاصه‌سازی را با توجه به عناصر مهم متن افزایش دهد. با استفاده از توجه نرم، مدل توانست به‌صورت پویا اهمیت بخش‌های مختلف متن ورودی را هنگام تولید خلاصه تنظیم کند و اطمینان حاصل نماید که خروجی تولیدشده منسجم و مرتبط با پرس‌وجوی ارائه شده است. وانگ و همکاران (۲۰۲۲) نیز مکانیسم هم‌توجهی را برای اعمال نفوذ اطلاعات متقابل بین سؤال و زمینه به کار بردند که به جمع‌بندی بهتر کمک نمود. از یک مکانیسم توجه افزایشی نیز برای به دست آوردن توزیع جملات سندهایی استفاده کردند که دقیقاً با سؤال همخوانی دارند. به‌طور کلی، این مکانیسم‌ها به مدل‌ها این فرصت را می‌دهند که به‌صورت پویا و انتخابی بر روی بخش‌های مهم متن ورودی تمرکز کنند که نتیجه آن تولید خروجی‌های باکیفیت‌تر و منسجم‌تر است. این ویژگی‌ها به‌ویژه در مواردی که کیفیت و دقت اطلاعات حیاتی است، اهمیت دوچندانی پیدا می‌کند. به‌جرئت می‌توان گفت توسعه و استفاده از مکانیسم‌های توجه در مدل‌های ترنسفورمر، انقلابی در زمینه خلاصه‌سازی متون به وجود آورده و اجازه داده مدل‌ها با دقت و کارایی بالاتری اطلاعات کلیدی را استخراج کنند. استفاده از مدل‌هایی نظیر بارت که به‌طور خاص برای این قبیل وظایف طراحی شده‌اند، به کاربران این امکان را می‌دهد که از اطلاعات عظیم و پیچیده به‌سادگی و با سرعت بیشتری بهره‌برداری نمایند به‌طور کلی، مرور انجام‌شده نشان داد که هر روش ویژگی‌های منحصر به فرد خود را دارد و شامل نقاط قوت و ضعفی است که چشم‌انداز خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو را شکل داده است. با این حال، این مرور گویای این است که در حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در سال‌های اخیر روش غالب مورد استفاده یادگیری ماشین است. یادگیری ماشین

توانایی شناسایی الگوها و روابط غیرخطی را دارد که ممکن است برای روش‌های آماری دشوار باشد. در این خصوص روش‌های مبتنی بر قاعده نیز محدودیت داشته و نیاز به تعریف قواعد از پیش تعریف‌شده دارند. همچنین، مدل‌های یادگیری ماشین برخلاف روش‌های آماری و روش‌های مبتنی بر قاعده، به راحتی می‌توانند با داده‌های جدید سازگار شده و به‌روزرسانی شوند. زیرا توانایی یادگیری و تطبیق با تغییرات در داده‌ها را دارند. این بدان معناست که با افزایش حجم داده‌ها یا تغییر در ساختار آن‌ها، روش‌های یادگیری ماشین می‌توانند به‌طور خودکار یاد بگیرند، بهبود یابند و نتایج دقیق‌تری ارائه دهند. اگر مدل با متون جدیدی که شامل اصطلاحات یا موضوعات جدید هستند آموزش ببیند، می‌تواند متون جدید دیگری که شامل این اصطلاحات هستند را بهتر خلاصه کند. در مقابل، روش‌های آماری توانایی یادگیری ندارند و نیازمند بازنگری و اصلاح مدل‌ها هستند. برخلاف روش‌های آماری، یادگیری ماشین می‌تواند بدون نیاز به فرضیه، به تحلیل داده‌ها بپردازد. این ویژگی باعث می‌شود که یادگیری ماشین در شرایطی که داده‌ها به‌طور کامل قابل کنترل نیستند، مؤثرتر عمل کند. به‌عنوان مثال، یک مدل ترنسفورمر می‌تواند با تحلیل داده‌های متنی بزرگ و بدون نیاز به فرضیه‌های خاص، خلاصه‌های دقیق از متن تولید کند. یادگیری ماشین به‌ویژه در پیش‌بینی نتایج جدید یا ناشناخته عملکرد بهتری دارد. این روش‌ها می‌توانند از الگوریتم‌های پیچیده‌ای مانند شبکه‌های عصبی استفاده کنند که برای شناسایی الگوهای پیچیده طراحی شده‌اند. این قابلیت باعث می‌شود که نتایج دقیق‌تری نسبت به روش‌های مبتنی بر قاعده و روش‌های آماری ارائه شود. افزون بر این، یادگیری ماشین می‌تواند فرایندهای خلاصه‌سازی متن را خودکار کند و زمان لازم برای پردازش اطلاعات را کاهش دهد. این ویژگی به‌ویژه در محیط‌هایی که نیاز به پردازش سریع اطلاعات وجود دارد، بسیار ارزشمند است. به‌طور کلی، روش‌های یادگیری ماشین با توانایی پردازش داده‌های پیچیده، کمترین نیاز به فرضیه‌ها، قابلیت یادگیری از داده‌ها، عملکرد بهتر در پیش‌بینی و امکان اتوماسیون فرایندها، مزایای قابل توجهی نسبت به روش‌های آماری و مبتنی بر قاعده دارند. پژوهشگران با کمک یادگیری ماشین به ارائه نوآوری در خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو اهتمام ورزیده‌اند و پیشنهادها و امیدوارکننده‌ای را برای تولید خلاصه‌های با کیفیت بالا ارائه داده که نیازهای در حال تغییر و تکامل کاربران را در دنیای داده‌محور کنونی تا حدی برآورده می‌کند. در کنار این مزایا، فرایند یادگیری ماشین ممکن است نیاز به زمان و منابع محاسباتی بیشتری داشته باشد. همچنین، نیاز به مجموعه‌ای از داده‌های آموزشی برای آموزش مدل دارد که ممکن است در دسترس نباشد. این مورد از جمله چالش‌هایی است که در این حوزه به‌طور قابل توجهی به چشم می‌خورد. علاوه بر این، در صورت عدم تنظیم صحیح، مدل‌ها ممکن است به داده‌های آموزشی بیش‌ازحد حساس شوند و کارایی کمتری در متون جدید داشته باشند. بر این اساس،

یک توجه برای هم‌افزایی اتفاق افتاده در مدل‌های پیشنهادی می‌تواند رفع کاستی‌های هر روش در کنار سایر روش‌ها باشد.

د. نوع یادگیری سیستم‌های خلاصه‌سازی پیشنهادی

بررسی انواع یادگیری به‌منظور خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو نشان‌دهنده تغییراتی در نحوه تولید خلاصه‌ها است. یکی از برجسته‌ترین نوع یادگیری در مطالعات بررسی شده، یادگیری نظارت‌شده^۱ است که در ۲۴ مطالعه از آن استفاده شده است. در برخی متون علاوه بر یادگیری نظارت‌شده از یادگیری‌های دیگری نیز در طول کار استفاده شده است. به‌طور مثال، لاسکار و همکاران^۲ (۲۰۲۱) از یادگیری انتقالی^۳ و یادگیری با نظارت ضعیف^۴ استفاده کرده‌اند که از آن‌ها با عنوان فنون تطبیق دامنه یاد شده است. ^۵ مطالعه از یادگیری انتقالی استفاده کرده‌اند. این نوع یادگیری که از یک مدل آموزش‌دیده بر روی یک مجموعه داده خاص به‌منظور انجام خلاصه‌سازی مشابه با مجموعه داده متفاوت استفاده می‌نماید؛ برای افزایش عملکرد مدل خلاصه‌سازی با استفاده از مدل‌های ترنسفورمر از پیش آموزش‌دیده استفاده شده است. مطالعه، لاسکار و همکاران (۲۰۲۱) بر روی یادگیری انتقالی، نشان‌دهنده توانایی مدل‌ها در بهره‌برداری از دانش به‌دست‌آمده از مجموعه داده‌های بزرگ‌تر است که این خود نشان‌دهنده پیشرفتی قابل توجه در میزان دقت خلاصه‌سازی است.

یادگیری با نظارت ضعیف که در ۲ مطالعه دیده شد؛ برای آموزش مدل با داده‌های برچسب‌گذاری شده محدود مورد استفاده قرار گرفت. این نوع یادگیری زمانی کارآمد است که برچسب‌های دقیق برای همه داده‌ها در دسترس نباشد. یافته‌های حاصل شده از ۲ مطالعه انجام‌شده با این نوع یادگیری نشان دادند استفاده از یادگیری با نظارت ضعیف می‌تواند کارایی مدل‌های خلاصه‌سازی را به‌طور چشمگیری در مقایسه با مدل‌های پیاده‌سازی شده که فقط یادگیری نظارت‌شده دارند، افزایش دهد.

در برخی پژوهش‌ها از دو نوع یادگیری با نظارت و بدون نظارت^۵ استفاده شد. مثال بارز پژوهش پولاش (۲۰۱۹) است که برای دست‌کاری مجموعه داده Debatepedia از این دو نوع یادگیری استفاده کرد. یادگیری بدون نظارت در ۶ مطالعه مورد استفاده قرار گرفته است. یکی از نمونه‌های یادگیری بدون نظارت در خلاصه‌سازی، استفاده از الگوریتم‌های خوشه‌بندی برای گروه‌بندی جملات یا اسناد مشابه باهم بر اساس معنای آن‌هاست.

¹. Supervised learning

². Laskar et al.

³. Transfer learning

⁴. Weakly supervised

⁵. Unsupervised

مورد دیگر یادگیری خودنظارتی^۱ است که اصلی ترین روش در مدل های برت، بارت، پگاسوس، T5 و جی پی تی است. این نوع یادگیری در ۱۵ مطالعه مورد استفاده بوده است که در آن مدل به جای یادگیری از داده های برچسب گذاری شده و خلاصه شده های آماده، از داده های خام برای شناسایی قسمت های مهم متن و تولید خلاصه های باکیفیت استفاده می کند. این امر در شرایطی که داده های برچسب گذاری شده زیادی در دسترس نیست، اهمیت بیشتری پیدا می کند. به نظر می رسد نوع یادگیری به مدل ها این امکان را می دهد که روابط پیچیده میان کلمات و جملات را درک کرده و خلاصه های با کیفیت بالایی تولید کنند بررسی ها نشان می دهد هم افزایی این فنون یادگیری مبتنی بر کارایی و مؤثر بودن مدل ها بوده و بر تولید خلاصه های دقیق تأکید می کند و نمایانگر یک رویکرد تکاملی در این حوزه است. پژوهش در این حوزه با کشف راه های جدید برای ترکیب و هماهنگ سازی این مدل ها می تواند به تولید خلاصه های باکیفیت و مفهومی تر منجر شود. این رویکرد نه تنها به دلیل افزایش دقت، بلکه به دلیل توانایی در پردازش داده های متنوع تر و به کارگیری دانش به دست آمده از زمینه های مختلف، حائز اهمیت است. واضح است که این پیشرفت ها حاکی از نیاز مداوم به سازگاری با چالش های نوظهور در جهان اطلاعات است و دریچه هایی را در مسیر پژوهش های آینده می گشاید.

ه. مدل های مورد استفاده

امر خلاصه سازی نشان دهنده تلاشی کارآمد در حوزه انتشار دانش است. در این مسیر، پیشرفت مدل های خلاصه سازی متون، تاریخچه پویای آن را آشکار می کند. هم افزایی مدل های مختلف، تحولات دگرگون کننده ای را در چشم انداز این حوزه و به طور خاص خلاصه سازی انتزاعی مبتنی بر پرس وجو نشان می دهد و گویای باز بودن مسیر پیشرفت های قابل توجه در آن است. از جمله این مدل ها که در مطالعات بررسی شده مورد استفاده قرار گرفته است می توان مدل های مبتنی بر گراف^۲، مدل های مبتنی بر شبکه های عصبی^۳ و مدل های از پیش آموزش دیده^۴ را نام برد.

به زعم ژو و همکاران^۵ (۲۰۲۳) مدل های مبتنی بر گراف یک نمایش گرافیکی از متن را اتخاذ می کنند. در این گراف، گره ها نشان دهنده جملات یا کلمات هستند و یال ها ارتباط بین آنها را نشان می دهند (Wibawa & Kurniawan, 2024). از میان مطالعات مورد بررسی، در

^۱. Self-supervised learning (SSL)

^۲. Graph-based model

^۳. Neural networks model

^۴. Pretrained model

^۵. Zhou et al.

پژوهش دنگ و همکاران^۱ (۲۰۲۳) از گراف استفاده شده است. گراف‌ها با در نظر گرفتن اطلاعات ساختاری و معنایی برای استدلال بر روی مسیرهای رابطه‌ای چندمرحله‌ای^۲ در روابط معنایی بین جملات سودمند واقع شد. سه نوع گراف با روابط معنایی متفاوت و تمرکز بر ارتباط معنایی، انسجام موضوعی و پیوند همبستگی برای شناسایی روابط متقابل سؤال-سند و جمله-جمله ساخته شده است. این مطالعه به‌طور مؤثر از مدل‌های گراف برای تقویت استدلال و استنتاج چندمرحله‌ای استفاده نموده است. در مطالعه شفيعی باوانی و همکاران^۳ (۲۰۱۶) از یک گراف برای خوشه‌بندی پرس‌وجو-جملات مرتبط استفاده شد. به‌منظور خوشه‌بندی، هرس گراف بر اساس نمرات شباهت معنایی اتفاق افتاد که منجر به حذف جملات نامرتب شد. همچنین، برای نمایش کلمات در جملات به‌عنوان دنباله‌ای از گره‌ها جهت کاهش افزونگی و درنهایت ایجاد خلاصه باکیفیت از گراف بهره گرفته شد. افزون بر آن، از گراف جهت ابهام‌زدایی از جفت‌های کلمه با استفاده از بافت یا فراوانی استفاده شد.

به نظر می‌رسد تمرکز مدل‌های مبتنی بر گراف در اخذ روابط غنی در ساختار و بافت متون است. در دو مطالعه بالا، کارآمدی گراف‌ها در تقطیر اطلاعات معنی‌دار و تولید خلاصه‌های غنی نشان داده شده است. این مطالعات با استفاده از گراف‌های مختلف و فنون خلاصه‌سازی انتزاعی، بهبود فرایند خلاصه‌سازی را با اطمینان از نمایش مختصر و مرتبط اطلاعات از متون به همراه داشتند. طبق نتایج حاصل‌شده از مطالعه دنگ و همکاران (۲۰۲۳) استفاده از گراف روشی مؤثر برای تقویت استدلال و درک روابط معنایی پیچیده در متون است که به تولید خلاصه‌های رضایت‌بخش می‌انجامد.

پژوهش‌های اخیر به‌وضوح نشان داده‌اند که مدل‌های مبتنی بر گراف با ترسیم روابط معنایی و ساختاری بین جملات و کلمات، می‌توانند منجر به تولید خلاصه‌های باکیفیت و مرتبط شوند. گراف‌ها با شناسایی و حفظ اطلاعات کلیدی، قادر به حذف جملات نامرتب و افزونگی‌هایی هستند که بر کیفیت نهایی خلاصه تأثیر مثبت می‌گذارد. اهمیت خاص این مدل‌ها در تبدیل داده‌های متنی به اطلاعات معنی‌دار، کاهش سطح ابهام و بهبود درک خواننده از متن اصلی نهفته است. همچنین با توجه به پتانسیل‌های گراف‌ها در تحلیل‌های چندمرحله‌ای و شناسایی روابط عمیق بین متون، می‌توانند به‌عنوان زیرساختی برای توسعه سیستم‌های هوش مصنوعی پیشرفته‌تر به کار روند. به نظر می‌رسد با ترکیب این مدل‌ها با فنون یادگیری ماشین و یادگیری عمیق، می‌توان به ابزارهای جامع‌تری دست یافت که نه تنها

¹. Deng et al.

². Multi-hop relational paths

³. ShafieiBavani et al.

خلاصه‌های دقیق‌تری تولید می‌کنند، بلکه قادر به استنتاج و تحلیل عمیق‌تر محتوای متون نیز خواهند بود. بهتر است پژوهش‌های آینده بر روی بهینه‌سازی این مدل‌ها و ادغام آن‌ها با فناوری‌های نوین تمرکز کنند تا قابلیت‌های آن‌ها در زمینه‌های مختلف به حداکثر برسد.

مدل‌های مبتنی بر شبکه‌های عصبی، نوع دیگری از مدل‌هایی هستند که در مطالعات مورد بررسی از آن‌ها استفاده شده است. شبکه عصبی یک سیستم پردازشی است که از مغز انسان الگوبرداری شده است. این سیستم مجموعه‌ای به هم پیوسته از نورون‌های مصنوعی است که از یک مدل عددی محاسباتی برای پردازش استفاده می‌کند (Andhale & Bewoor, 2016). شبکه‌های عصبی مبنای یادگیری عمیق هستند. از آنجاکه توانایی یادگیری ماشین در پردازش داده‌های طبیعی به شکل خام محدود است؛ برای رفع این محدودیت‌ها یادگیری عمیق به وجود آمد که برای حل مسائل پیچیده از شبکه‌های عصبی استفاده می‌کند.

فنون یادگیری عمیق^۱ روش‌های یادگیری بازنمایی^۲ با سطوح مختلف نمایش است که با ترکیب بخش‌های ساده اما غیرخطی به دست می‌آیند (LeCun et al., 2015). یادگیری عمیق مشکلات پیچیده را برای تسهیل فرایند تصمیم‌گیری تجزیه و تحلیل می‌کند و تلاش می‌نماید تا آنچه را که مغز انسان می‌تواند با استخراج ویژگی‌ها در سطوح مختلف انتزاع به دست آورد، تقلید کند (Suleiman & Awajan, 2020). لذا، در حل مشکلاتی که سال‌ها در برابر بهترین تلاش‌های جامعه هوش مصنوعی مقاومت کرده‌اند، پیشرفت‌های عمده‌ای داشته است. یادگیری عمیق در کشف ساختارهای پیچیده از درون داده‌های با ابعاد بسیار بالا عملکرد خوبی دارد و به مدل‌های محاسباتی که از لایه‌های پردازشی متعدد تشکیل شده‌اند اجازه می‌دهد تا نمایش داده‌ها را با سطوح انتزاعی متعدد بیاموزند. به همین دلیل در بسیاری از حوزه‌ها قابل استفاده است و به‌طور چشمگیری خلاصه‌سازی متون، تشخیص گفتار، تشخیص اشیاء و بسیاری از حوزه‌های دیگر مانند کشف دارو و نظایر آن را بهبود بخشیده است (LeCun et al., 2015).

در سال‌های اخیر از فنون یادگیری عمیق در خلاصه‌سازی انتزاعی متن استفاده‌های زیادی شده است. در این نوع خلاصه‌سازی از طریق شبکه عصبی در نگاه کلی و عام، الف) کلمات به وسیله یک جدول جستجو به بردارهای پیوسته تبدیل می‌شوند؛ ب) جملات/اسناد به عنوان بردارهای پیوسته با استفاده از جانشانی کلمه^۳ کدگذاری می‌شوند؛ ج) بازنمایی‌های جمله/سند به یک مدل برای انتخاب (خلاصه استخراجی) یا تولید (خلاصه انتزاعی) تبدیل

^۱. Deep learning

^۲. یادگیری بازنمایی (Learning Representation) مجموعه‌ای از روش‌هایی است که به ماشین اجازه می‌دهد با داده‌های خام تغذیه شود و به‌طور خودکار بازنمایی‌های مورد نیاز برای تشخیص یا طبقه‌بندی را کشف کند

^۳. Word Embedding

می‌شود. در هر یک از گام‌های گفته‌شده می‌توان از شبکه‌های عصبی بهره گرفت. در گام اول می‌توان از شبکه‌های عصبی برای به دست آوردن جداول جستجوی از پیش آموخته‌شده استفاده نمود. در گام دوم، شبکه‌های عصبی از قبیل شبکه عصبی پیچشی، شبکه عصبی بازگشتی و غیره می‌توانند به‌عنوان رمزگذار برای استخراج ویژگی‌های جمله/ سند استفاده شوند. در گام سوم، مدل‌های شبکه عصبی می‌توانند برای رتبه‌بندی/ انتخاب (استخراج) یا به‌عنوان رمزگشا برای تولید (انتزاع) خلاصه استفاده شوند (Dong, 2018).

از زمان شکوفایی یادگیری عمیق و در فاصله کمی از اولین استفاده از آن در خلاصه‌سازی انتزاعی متون در سال ۲۰۱۵، استفاده از شبکه‌های عصبی در این حوزه توجه زیادی را به خود جلب کرده است و این بهره‌گیری به‌طور چشمگیری فزونی یافته است، زیرا در صورت وجود داده‌های آموزشی زیاد، نسبت به مدل‌های سنتی عملکرد بسیار بهتری دارند. به نظر می‌رسد مدل‌های مختلف شبکه‌های عصبی با مهارت و توانایی در درک متون، انقلابی در خلاصه‌سازی ایجاد کرده‌اند.

این شبکه‌ها به‌منظور پیدا کردن الگوها و روابط پیچیده در متون اغلب از معماری‌های پیچیده‌ای مانند شبکه عصبی بازگشتی، شبکه عصبی پیچشی، شبکه عصبی گراف^۱، ترنسفورمرها و غیره استفاده می‌کنند. بررسی‌ها گویای این است که استفاده‌پذیری پژوهشگران از معماری‌های بالا، ایجاد خلاصه‌های رضایت‌بخشی را به دنبال دارد. ماحصل بررسی انجام‌شده بر روی مطالعات پیرامون خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در پژوهش حاضر گواه و تأییدی بر این امر است. شبکه عصبی بازگشتی در رمزگشایی مدل دنباله به دنباله به گرفتن ماهیت متوالی داده‌های متنی کمک کرده و به مدل اجازه داده است وابستگی‌های بین کلمات را درک کند. بدین ترتیب نقش مهمی در تولید خلاصه ایفا نمود. ماهیت ترتیبی این شبکه آن را قادر ساخت تا داده‌های متنی را به‌گونه‌ای پردازش کرده و جریان و انسجام ورودی اصلی را حفظ کند. لذا، ادغام این شبکه در رمزگشا، به عملکرد مدل‌های انتزاعی در درک و ترکیب مؤثر اطلاعات کمک کرد (Baumel et al., 2018). شبکه عصبی بازگشتی در یک مدل ترنسفورمر که با استفاده از مکانیسم توجه تقویت شده است؛ توانست داده‌های متوالی را با پردازش گام‌به‌گام ورودی مدیریت نماید. از طریق ماهیت ترتیبی شبکه، اطلاعات در هر مرحله زمانی جریان یافته و مدل را قادر ساخت تا بافت را حفظ کرده و خلاصه‌هایی تولید کند که محتوای اساسی اسناد ورودی را منعکس کند (Ritharson et al., 2023). از شبکه عصبی بازگشتی رمزگذار برای پرس‌وجو و برای سند استفاده شده است. این شبکه‌ها در رمزگذاری- رمزگشایی که اساس مدل پیشنهادی است، استفاده شده و

^۱. Graph Neural Network (GNN)

بهبود خلاصه‌سازی را به دنبال داشته‌اند (Nema et al., 2017). به نظر می‌رسد شبکه‌های عصبی بازگشتی با توجه به قابلیت‌های خود از جمله پردازش توالی، حفظ بافت و تقویت تعاملات معنایی، به ابزاری قدرتمند در عرصه خلاصه‌سازی تبدیل شده‌اند. با توجه به موفقیت‌های قابل توجه این مدل‌ها در تولید خلاصه‌های باکیفیت، پیشنهاد می‌شود پژوهش‌های آینده به بهینه‌سازی این ساختارها و بررسی ترکیب آن‌ها با فناوری‌های نوین بپردازد. این امر می‌تواند به توسعه سیستم‌های خلاصه‌سازی هوشمندتر منجر شود و به ارتقای کیفیت تعاملات انسانی با ماشین در این حوزه کمک نماید.

حافظه کوتاه‌مدت طولانی^۱ و واحد بازگشتی دروازه‌دار^۲ نیز در مدل‌سازی برخی از مطالعات استفاده شدند. لی و همکاران^۳ (۲۰۲۳) از حافظه کوتاه‌مدت طولانی در لایه‌های رمزگذار و رمزگشای مدل خود استفاده کردند تا اطلاعات متوالی را ضبط کند. لایه رمزگذار با حافظه کوتاه‌مدت طولانی دوطرفه وظیفه پردازش کلمات از پیش آموزش داده‌شده از سؤالات و جملات در پاسخ را بر عهده داشت و مدل را قادر ساخت تا داده‌های ورودی را بهتر درک کند. در لایه استنتاج نیز از حافظه کوتاه‌مدت طولانی دوطرفه برای اصلاح نمایش‌های سطح جمله استفاده شد که درک مدل از متن را بهبود بخشید. در مدل ارائه‌شده توسط آریال^۴ (۲۰۱۹) نیز حافظه کوتاه‌مدت طولانی به دلیل توانایی در حفظ و یادگیری اطلاعات در توالی‌های متن طولانی استفاده شد و شناسایی الگوهای پیچیده در داده‌های ورودی را تسهیل نمود. زیرا توانایی درک روابط بین کلمات و جملات یک سند را دارد. با ادغام آن مدل توانست بازنمایی و درک زمینه را در فرایند خلاصه‌سازی بهبود بخشد و خلاصه دقیق‌تر و مرتبط‌تری ارائه دهد. در رمزگذار و رمزگشای مدل ارائه‌شده توسط دنگ و همکاران^۵ (۲۰۲۰) نیز حافظه کوتاه‌مدت طولانی به کار گرفته شد. ادغام این شبکه در رمزگشا به یادگیری پاسخ‌های مختصر و آموزنده برای سؤالاتی که به سیستم داده شد کمک زیادی کرد و در ارتباط با اجزای دیگر مانند رمزگذار و مکانیسم‌های توجه، موجب شناسایی وابستگی‌های مهم بین بخش‌های مختلف متن ورودی شده و به تولید خلاصه‌های باکیفیت انجامید. در مدل ارائه‌شده توسط کو مارول و سانکارانایارانا (۲۰۲۱) نیز از حافظه کوتاه‌مدت طولانی به عنوان جزئی از رمزگشا استفاده شد و در کنار آن از حافظه کوتاه‌مدت طولانی دوطرفه در رمزگذار بهره گرفته شد. این شبکه توانایی حفظ اطلاعات در هر دو جهت عقب و جلو را با درک بهتر زمینه نشان داد و این درک دوسویه کیفیت

¹. Long-Short Term Memory (LSTM)

². Gated Recurrent Unit (GRU)

³. Li et al.

⁴. Aryal

⁵. Deng et al.

فرایند خلاصه‌سازی را فزونی بخشید. علاوه بر این، رمزگشای حافظه کوتاه‌مدت طولانی با ترکیب مکانیسم‌های توجه برای خلاصه‌سازی منسجم و دقیق کارایی خوبی نشان داد. مشابه با مطالعات پیشین، در مطالعه پولاش (۲۰۱۹) نیز یک حافظه کوتاه‌مدت طولانی در رمزگذار سند استفاده شد. این شبکه با حفظ اطلاعات در توالی‌های طولانی و تنظیم جریان آن، وابستگی‌ها را بین سند ورودی و پرس‌وجو برای ایجاد یک خلاصه منسجم ثبت نمود. در مطالعه دیگری که توسط نما و همکاران (۲۰۱۷) منتشر شد؛ حافظه کوتاه‌مدت طولانی با حفظ انسجام و جلوگیری از تکرار عبارات، کیفیت خلاصه‌ها را افزایش داده و در نهایت منجر به خلاصه‌های مختصر و منسجم‌تر شد. برای این منظور، مدل رمزگذار- رمزگشای ارائه‌شده برای متعادل کردن بردارهای زمینه متوالی به یکدیگر از حافظه کوتاه‌مدت طولانی استفاده نمود. در این مطالعه از واحد بازگشتی دروازه‌ای نیز در رمزگذارهای پرس‌وجو و هم در رمزگذارهای سند برای پردازش داده‌های ورودی استفاده شد. این واحد پرس‌وجو و سند را به‌صورت متوالی خوانده تا با پردازش کلمات در پرس‌وجو و سند به‌صورت متوالی از چپ به راست نمایش‌های پنهان را برای هر مرحله زمانی محاسبه نماید. در مدل پیشنهادی هاسلوویست و همکاران^۱ (۲۰۱۷) دروازه‌های واحد بازگشتی دروازه‌ای، جریان اطلاعات را تنظیم کرده و به مدل کمک نمودند تا تصمیم بگیرد چه اطلاعاتی را در تولید خلاصه‌های انتزاعی بر اساس پرس‌وجوها در اولویت قرار دهد.

مرور انجام‌شده نشان می‌دهد حافظه کوتاه‌مدت طولانی و واحد بازگشتی دروازه‌دار به خاطر توانایی‌شان در نگهداری و پردازش اطلاعات متوالی و پیچیده در حوزه خلاصه‌سازی مورد توجه قرار گرفته‌اند. عملکرد حافظه کوتاه‌مدت طولانی در رمزگذار و رمزگشا موجب درک بهتر از داده‌ها، شناسایی وابستگی‌های مهم بین بخش‌های مختلف متن ورودی و تشخیص الگوهای پیچیده شده است. دوطرفه بودن آن نیز امکان پردازش اطلاعات از هر دو سمت را میسر ساخته است و تأثیر مثبتی بر دقت در خلاصه‌سازی متن گذاشته است. واحد بازگشتی دروازه‌دار نیز به مدل‌ها این امکان را داد تا تصمیم بگیرند چه اطلاعاتی در تولید خلاصه‌های انتزاعی بر اساس پرسش ورودی باید در اولویت قرار گیرد. این دو شبکه در ترکیب با مکانیسم‌های توجه، کارایی بهتری در تولید خلاصه‌های منسجم و دقیق نشان داده‌اند. به‌نظر می‌رسد ادغام این شبکه‌ها با سایر روش‌های نوین دیگر می‌تواند در آینده راهکارهای نوآورانه‌تری را در زمینه خلاصه‌سازی به ارمغان آورد.

مدل‌های دیگری که در مطالعات بررسی‌شده مورد استفاده قرار گرفته‌اند مدل‌های از پیش آموزش‌دیده هستند. در سال‌های اخیر ظهور مدل‌های از پیش آموزش‌دیده خلاصه‌سازی

^۱. Hasselqvist et al.

متون را به‌عنوان یکی از کاربردهای پردازش زبان طبیعی وارد عصر جدیدی کرده است. مدل‌های از پیش آموزش‌دیده مورد استفاده در مطالعات مانند برت (Esteva et al., 2021; Laskar et al., 2020; Abdullah, 2020; Wang et al., 2022; Su et al., 2020) (جی‌پی‌تی (Esteva et al., 2021; Vig et al., 2021; بارت (et al., 2021; Laskar et al., 2023; Laskar et al., 2023) (Laskar et al., 2023; Su et al., 2020; Su et al., 2021; Xu & Lapata, 2021; Laskar et al., 2021))، برت‌سام^۱ به‌عنوان مدل برت مخصوص خلاصه‌سازی، پگاسوس و T5 به موفقیت‌های خوبی دست یافته‌اند. در مطالعه استوا و همکاران^۲ (۲۰۲۱)، برت نقش مهمی را به‌عنوان رمزگذار در مدل پیشنهادی ایفا نموده و به‌عنوان رمزگشا با مدل اصلاح‌شده جی‌پی‌تی-دو^۳ ترکیب شده است. نقش برت در این مدل، رمزگذاری مؤثر اطلاعات ورودی، گرفتن روابط بین کلمات و جملات جهت تولید خلاصه‌های دقیق و منسجم است. جی‌پی‌تی-دو یک مدل زبانی است که متن را با پیش‌بینی کلمه بعدی در یک دنباله تولید می‌کند. علی‌رغم اینکه جی‌پی‌تی-دو به‌طور خاص برای خلاصه‌سازی طراحی نشده است، به دلیل قابلیت‌های عمومی درک زبان، به‌عنوان یک پایه قوی عمل می‌کند (Liu et al., 2024). برت با استفاده از مدل‌سازی روی یک پیکره بزرگ از قبل آموزش‌دیده است. در این مدل، درصد مشخصی از کلمات در جملات ورودی به‌طور تصادفی انتخاب شده و با یک نشانه ویژه جایگزین می‌شوند. هدف این است که مدل کلمات ماسک شده پوشانده شده را بر اساس زمینه ارائه‌شده توسط کلمات اطراف پیش‌بینی کند. این رویکرد با در نظر گرفتن زمینه چپ و راست در طول آموزش، دوطرفه است. برای یک جمله برت با پیش‌بینی کلمات پوشانده شده، جاسازی‌های متنی را می‌آموزد. هدف، به حداکثر رساندن احتمال پیش‌بینی کلمات با توجه به زمینه است (Jeyakarthic & Le-oraj, 2024). گاولان و همکاران^۴ (۲۰۲۴) می‌گویند استفاده از مدل‌های مبتنی بر برت که بسیار قدرتمند هستند و در عین حال از منابع کمتری نسبت به مدل‌های زبان بزرگ معاصر استفاده می‌کنند، برای تبدیل متن به نمایش‌های عددی، دقت تشخیص بیماری‌ها از جمله افسردگی را به‌طور چشمگیری افزایش می‌دهد. این مدل‌ها به‌خوبی تفاوت‌های معنایی و نحوی پیچیده را به تصویر می‌کشند و دقت تشخیص علائم افسردگی را بهبود می‌بخشند (Gavalan et al., 2024). آن‌ها روش جدیدی را برای تشخیص افسردگی با استفاده از برت ارائه کردند که خلاصه‌سازی را بر روی مجموعه داده شامل مصاحبه با افراد مبتلا به افسردگی و افراد بدون علائم افسردگی انجام داده است. در مدلی که لاسکار و همکاران^۵ (۲۰۲۱) ارائه دادند، برت

^۱. BERTSUM

^۲. Esteva et al.

^۳. GPT-2

^۴. Gavalan et al.

^۵. Laskar et al.

برای رمزگذاری متن ورودی استفاده شد. با استفاده از آن، مدل توانست روابط بین کلمات را در متن درک کند و خلاصه دقیق‌تری تولید کند. افزون بر آن، استفاده از برت به کاهش مسائلی مانند تکرار کلمات در خلاصه‌های تولیدشده کمک نمود. رمزگذارهای ترنسفورمر از پیش آموزش‌دیده برت این مدل را قادر ساخت تا از مقادیر وسیعی از داده‌های متنی یاد بگیرد و توانایی آن را برای تولید خلاصه‌های انتزاعی باکیفیت افزایش دهد. در سایر مطالعات مذکور نیز برت برای وظایفی از قبیل افزایش تعامل بین سؤالات و متن‌ها، افزایش عملکرد مدل با رمزگذاری اطلاعات متن و پرس‌وجو، بهبود درک پرس‌وجو، رمزگذاری پارامترها در فرایند آموزش مدل، فراهم کردن جانشانی متن برای نمایش بهتر کلمات در خلاصه، بهبود تجزیه و تحلیل معنایی در مدل، بهبود مدل رمزگذار-رمزگشا و درنهایت ماحصل همه این کمک‌ها، به‌منظور تولید خلاصه‌های باکیفیت مورد استفاده قرار گرفت.

جی‌پی‌تی نیز یک ترنسفورمر از پیش آموزش‌دیده است که در آن مدل‌ها از قبل روی مجموعه‌های گسترده آموزش داده شده و برای کارهای خلاصه‌سازی خاص تنظیم شده‌اند. پیش‌آموزش مدل‌ها را قادر می‌سازد تا ظرافت‌های زبانی پیچیده را درک کنند و آن‌ها را در تولید خلاصه‌های منسجم ماهر سازد (Wibawa & Kurniawan, 2024). آموزش بر روی مجموعه داده‌های بزرگی از متن و کد به آن‌ها امکان می‌دهد تا روابط آماری بین کلمات و عبارات و الگوهای طبیعی زبان انسانی را بیاموزند. این دانش برای تولید خلاصه‌های روان و آموزنده به‌ویژه خلاصه‌های انتزاعی استفاده می‌شود. خلاصه‌نویسی انتزاعی شامل شناسایی ایده‌های اصلی در یک متن و بازنویسی آن‌ها به شیوه‌ای مختصر و آموزنده است. جی‌پی‌تی می‌تواند این کار را انجام دهد، زیرا قادر به یادگیری معنای اساسی متن است (Han & Choi, 2024). در دنیای واقعی جی‌پی‌تی به‌عنوان یک مثال عینی آشکار، موفقیت مدل‌های از پیش‌آموزش‌دیده را در خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو نمایان‌تر کرده است. اخیراً استفاده از نسخه‌های مختلف جی‌پی‌تی با طرح هرگونه پرس‌وجو و اخذ پاسخ‌های انتزاعی، کاربردپذیری آن را در کل جهان با سرعت بسیار چشمگیری فزونی داده و زمینه ایجاد سیستم‌هایی از این دست را در حوزه‌های مختلف فراهم نموده است. بنابراین گرفتن جی‌پی‌تی می‌تواند به دلیل موفقیت‌های آن در تولید باکیفیت خلاصه‌های انتزاعی مبتنی بر پرس‌وجو باشد که نشان‌دهنده کارآمدی این مدل زبانی در فهم و تحلیل متون است.

ماحصل بررسی مطالعات، گویای این است که مدل‌های مختلف، زمینه تحولات زیادی را در چشم‌انداز خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو فراهم کرده است. از مدل‌های مبتنی بر گراف گرفته تا شبکه‌های عصبی و مدل‌های از پیش آموزش‌دیده دستخوش پیشرفت‌های قابل‌توجهی بوده و کارایی سیستم‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو را به‌طور

چشمگیری بهبود بخشیده‌اند. مدل‌های مبتنی بر گراف در شناسایی جملات کلیدی بر اساس روابط درون‌متن مؤثر هستند و توانایی حفظ ساختار و انسجام اصلی سند را دارند. علاوه بر این، معمولاً تفسیرپذیری خوبی دارند، زیرا بر روابط ریاضی متکی هستند. مدل‌های مبتنی بر شبکه‌های عصبی نیز قادر به درک الگوهای زبانی و زمینه‌های پیچیده‌تر هستند. همچنین، توانایی مدیریت مجموعه داده‌های بزرگ و روابط پیچیده بین کلمات را دارند و با دریافت داده‌های بیشتر و آموزش بیشتر می‌توانند بهبود یابند و خلاصه‌های باکیفیتی ایجاد کنند. مدل‌های از پیش آموزش‌دیده به دلیل آموزش گسترده، توانایی بالایی در درک عمیق از ساختار و معانی جملات دارند و انعطاف‌پذیرند. بررسی‌ها نشان داده است که مدل‌های از پیش آموزش‌دیده معمولاً در حوزه خلاصه‌سازی عملکرد بهتری نسبت به مدل‌های مبتنی بر گراف و شبکه‌های عصبی دارند. این به دلیل توانایی‌شان در یادگیری معنا و ساختار زبان طبیعی، پردازش متن‌های بزرگ و پیچیده و نیز قابلیت تعمیم به دامنه‌های مختلف متنی است. مدل‌های مبتنی بر شبکه‌های عصبی نیز از مدل‌های مبتنی بر گراف بهتر عمل می‌کنند. هرچند که مدل‌های مبتنی بر گراف دارای مزایای خاص خود هستند، اما در برابر گنجایش و قدرت پردازش عمیق مدل‌های شبکه عصبی و از پیش آموزش‌دیده، در عمل ضعیف‌تر هستند. البته انتخاب مدل مناسب به نوع درخواست خلاصه، ویژگی‌های متن و داده‌های موجود بستگی دارد و ممکن است به‌کارگیری ترکیبی از این مدل‌ها کمک بسیار مؤثری در حصول بالاترین میزان دقت باشد. به همین دلیل است که هم‌افزایی مدل‌های مختلف در این امر به نتایج رضایت‌بخشی نائل شده است. به نظر می‌رسد این هم‌افزایی در شکل دادن به چشم‌انداز مدل‌های مربوط به این حوزه مؤثر است و نوید نوآوری‌های بیشتری را می‌دهد.

و. نوع ورودی سیستم‌های خلاصه‌سازی پیشنهادی

خلاصه‌سازی را می‌توان بر اساس نوع ورودی، به‌عنوان سیستم تک‌سندی^۱ یا چندسندی^۲ دسته‌بندی کرد. در خلاصه‌سازی تک‌سندی ورودی سیستم خلاصه‌ساز تنها یک سند است. این سند ضمن حفظ ضروری‌ترین اطلاعات، به شکل کوتاه‌تری ارائه می‌شود. هنگامی که این نوع خلاصه‌سازی جهت پرس‌وجوی کاربر انجام شود به آن خلاصه‌سازی تک‌سندی مبتنی بر پرس‌وجو گفته می‌شود. در خلاصه‌سازی چندسندی از چندین سند به‌عنوان ورودی استفاده می‌شود. با رشد اطلاعات ذخیره‌شده در وب از یک‌سو و حجم زیاد اطلاعات ارائه‌شده در هر سند از سوی دیگر، استفاده از خلاصه‌سازی چندسندی حیاتی‌تر شده است. این نوع خلاصه به

¹. Single document

². Multi document

کاربر کمک می‌کند به سرعت با اطلاعات موجود در یک مجموعه سند آشنا شود. خلاصه‌های چندسندی عموماً مختصر و جامع هستند. اما برخلاف همتای تک‌سندی، ساختن آن‌ها پیچیده‌تر و دشوارتر بوده و افزودن^۱ از جمله مشکلاتی است که با آن مواجه هستند. هنگامی که این نوع خلاصه‌سازی برای پاسخ به پرس‌وجوی کاربر انجام شود به آن خلاصه‌سازی چندسندی مبتنی بر پرس‌وجو گفته می‌شود. در چنین مواردی سیستم خلاصه‌سازی سعی می‌کند از طریق چندین سند به پرسش کاربر پاسخ دهد.

مطالعات با در نظر گرفتن تک‌سندی و چندسندی بودن سیستم خلاصه‌سازی پیشنهادی مورد بررسی قرار گرفتند. یافته‌ها گویای بیشتر بودن سیستم‌های تک‌سندی است که در ۱۴ مطالعه ارائه شده‌اند. در ۱۱ مطالعه نیز نوع ورودی خلاصه‌ساز پیشنهادی چندسندی است. در ۶ مطالعه نیز سیستم پیشنهادی توانایی ورودی تک‌سندی و چندسندی را دارد. خلاصه‌سازی تک‌سندی و چندسندی هر دو مهم و کاربردی هستند، اما به درک عمیق از ویژگی‌ها، چالش‌ها و قابلیت‌های هر روش برای انتخاب مناسب‌ترین رویکرد بر اساس نیازهای خاص کاربر و ویژگی‌های متون نیاز دارد.

بررسی مطالعات نشان داده است در خلاصه‌سازی چندسندی مبتنی بر پرس‌وجو علاوه بر مشکل افزودن^۱، مشکل برقراری ارتباط بین پرس‌وجو و جملات اسناد نیز دیده می‌شود. استفاده از حداکثر ارتباط حاشیه‌ای^۲، بازسازی مجموعه داده‌های به کار گرفته‌شده با هدف ایجاد سه تایی جدید جستجو-سند-خلاصه، خوشه‌بندی، روش‌های مبتنی بر گراف، ماشین بولتزمن محدود^۳ و غیره از جمله راه‌های پیشنهادی در این مطالعات به منظور کاهش افزودن^۱ بوده‌اند. انتخاب جملات مرتبط با پرس‌وجو نیز از طریق راه‌هایی مانند مکانیسم‌های مختلف توجه از جمله توجه چندسری، خودتوجهی، خودتوجهی چندسری، توجه نرم، توجه مبتنی بر تنوع، رمزگذاری سلسله مراتبی^۴، رمزگذاری موقعیتی^۵، امتیازدهی فرکانس معکوس فرکانس سند، برچسب‌گذاری بخشی از گفتار^۶، شباهت کسینوس^۷، ادغام چندسری^۸ در ترنسفورمر، جانشانی‌های مختلف کلمه، جمله و موقعیت، خوشه‌بندی سه‌گانه معنایی^۹، نمودار تشابه^{۱۰} و نظایر آن انجام شده است.

چالش‌های موجود به‌ویژه در زمینه درخواست‌های مبتنی بر پرس‌وجو، نیازمند

¹. Redundancy

². Maximum marginal relevance (MMR)

³. Restricted Boltzmann Machine (RBM)

⁴. Hierarchical encoding

⁵. Positional encoding

⁶. POS tagging

⁷. Cosine similarity

⁸. Multi-head pooling

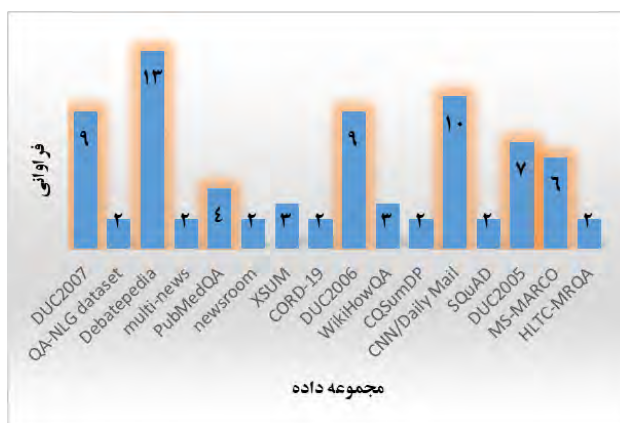
⁹. Semantic Triples Clustering (STC)

¹⁰. Similarity graph

راهکارهای تخصصی و نوآورانه‌ای است که استفاده از الگوریتم‌های پیچیده و فنون جدید را شامل می‌شود. این موضوع قطعاً زمینه‌های توسعه بیشتر و پژوهش‌های آینده را در این حوزه به‌ویژه در خلاصه‌سازی انتزاعی چندسندی مبتنی بر پرس و جو فراهم می‌نماید و همگام با پیشرفت‌های روز جهان، این مسیر توسعه به‌طور مداوم در حال تحول و بهبود خواهد بود.

ز. مجموعه داده‌های مورد استفاده

در طول سالیان، چندین فراهمی و کارگاه برای خلاصه‌سازی خودکار سازمان‌دهی شده است. این فراهمی‌ها مجموعه داده‌های مورد استفاده در آزمایش‌های پژوهشی گسترده را در دسترس قرار داده‌اند. روی این مجموعه داده‌ها عملیاتی جهت آماده‌سازی آن‌ها به‌عنوان یک متن استاندارد اعمال شده است تا برای ارزیابی روش‌شناسی‌های مختلف خلاصه‌سازی، عمل کنند. بررسی‌ها نشان می‌دهد مجموعه داده‌های مختلفی برای خلاصه‌سازی تک‌سندی و چندسندی وجود دارد. باین حال، تنها تعداد کمی مجموعه داده اختصاصی برای خلاصه‌سازی مبتنی بر پرس و جو در دسترس است. لذا از مجموعه داده‌های عمومی از جمله Dai-/CNN Mail ly برای اهداف خلاصه‌سازی انتزاعی مبتنی بر پرس و جو استفاده می‌شود (Pasunu-ru et al., 2021; Abdullah & Chali, 2020; Xu & Lapata, 2020). حاصل بررسی مطالعات این حوزه نشان‌دهنده استفاده از ۳۴ مجموعه داده توسط پژوهشگران است. این تنوع در مجموعه‌ها می‌تواند امکان مقایسه را برای پژوهشگران فراهم کند تا بتوانند روش‌های مختلف خلاصه‌سازی را در شرایط استاندارد ارزیابی نمایند. در نمودار ۱ پرکاربردترین این مجموعه داده‌ها ارائه شده است.



نمودار ۱- توزیع مجموعه داده‌های پراستفاده در مطالعات

طبق نمودار بالا، پرکاربردترین مجموعه داده در این مطالعات Debatepedia، Daily / CNN و Mail و DUC 2005-2007 است. Debatepedia که در ۱۳ مطالعه از آن بهره گرفته شده است؛ با استفاده از دایره‌المعارف Debatepedia ساخته شد. از این مجموعه داده ۸۰ درصد برای آموزش، ۱۰ درصد برای اعتبارسنجی و ۱۰ درصد برای آزمایش استفاده می‌شود. مجموعه داده Daily / CNN نیز حاوی بیش از ۳۰۰۰۰۰ مقالات خبری و خلاصه‌های چندجمله‌ای است. از این مجموعه داده ۹۲ درصد برای آموزش، ۴/۳ درصد برای اعتبارسنجی و ۳/۷ درصد برای آزمایش استفاده می‌شود (Giarelis et al. 2023). مجموعه داده‌های DUC 2005-2007 که در مجموع در ۲۶ مطالعه از آن‌ها استفاده شده است، برای خلاصه‌سازی چندسندی تدارک دیده شده‌اند. حاوی هیچ داده آموزشی برچسب‌گذاری شده‌ای نیستند و تنها داده‌های آزمایشی را ارائه می‌دهند (Baumel et al., 2018).

ازجمله مجموعه داده‌های محبوب دیگر که مختص خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو هستند می‌توان WikiHowQA، MARCO-MS و PubMedQA را نام برد که در این مطالعات از آن‌ها استفاده شده است. به‌زعم لاسکار و همکاران (۲۰۲۳) یکی از استثنای قابل توجه در میان این مجموعه داده‌ها، Debatepedia است. زیرا نیاز به تولید خلاصه‌های انتزاعی از یک سند کوتاه حاوی متن استدلالی دارد. هیچ‌یک از مجموعه داده‌های ذکر شده در بالا به موضوع ایجاد خلاصه‌های مبتنی بر پرس‌وجو از اسناد حاوی استدلال توجه نکردند. این موضوع، Debatepedia را به منبعی کارآمد برای پژوهشگران تبدیل کرده است تا روش‌هایی را به‌منظور خلاصه‌سازی یک سند کوتاه حاوی متن استدلالی برای پرسش داده‌شده ارائه کنند.

وجود مجموعه داده‌های متنوع به‌منظور خلاصه‌سازی تک‌سندی و چندسندی و کمبود مجموعه داده‌های اختصاصی برای خلاصه‌سازی مبتنی بر پرس‌وجو یک چالش بزرگ در این حوزه به‌شمار می‌رود. پژوهشگران در تلاش‌اند تا روش‌هایی پیشنهاد دهند که خلاصه‌ها را به شکل بهتری با توجه به سؤالات خاص و نیازهای کاربران تولید کنند. به‌عنوان یک نتیجه‌گیری کلی، نیاز به ایجاد و توسعه مجموعه داده‌های جدید و متناسب با خلاصه‌سازی مبتنی بر پرس‌وجو بیشتر احساس می‌شود تا این حوزه بتواند به‌طور مؤثرتری به نیازهای دنیای واقعی پاسخ دهد و بستر مناسبی برای پژوهش‌های آینده فراهم آورد.

ح. معیارهای ارزیابی مورد استفاده

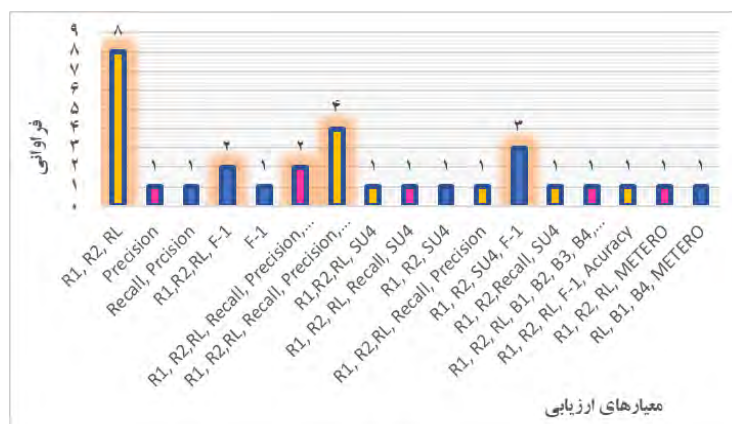
ارزیابی یک امر بسیار مهم در خلاصه‌سازی متون محسوب می‌شود. علاوه بر افزایش توسعه منابع و زیرساخت‌های قابل استفاده مجدد به مقایسه و تکرار نتایج کمک کرده و رقابت برای

بهبود نتایج را افزایش می دهد. باین حال، ارزیابی دستی چندین سند برای به دست آوردن یک دیدگاه بی طرفانه عملاً غیر ممکن است. لذا، معیارهای ارزیابی خودکار قابل اعتماد برای ارزیابی سریع و سازگار مورد نیاز است. ارزیابی خودکار خلاصه نیز یک کار چالش برانگیز است. زیرا برای ماشین آسان نیست که بداند چه نوع اطلاعاتی باید در خلاصه وجود داشته باشد. ارزیابی صحیح خلاصه ها نه تنها به تعیین کیفیت آن ها کمک می کند، بلکه می تواند به بهبود الگوریتم های خلاصه سازی نیز منجر شود. در این راستا، معیارهای سنجش متنوع به پژوهشگران امکان می دهند تا نقاط قوت و ضعف خلاصه های تولید شده را شناسایی کنند. از جمله معیارهای مورد استفاده به منظور ارزیابی خلاصه ها می توان ROUGE، بازیافت^۱، دقت^۲، F-1 و BLUE را نام برد. مرور مطالعات در پژوهش حاضر گویای این است که از معیارهای مختلفی برای خلاصه های انتزاعی مبتنی بر پرس و جوی تولید شده استفاده می شود. در این پژوهش هفده معیار شناسایی شده است. معیار F-1، ROUGE، بازیافت و دقت به ترتیب بیشترین استفاده را برای ارزیابی خلاصه های تولید شده در بین پژوهشگران داشته است.

معیار ROUGE به طور خودکار کیفیت خلاصه تولید شده توسط انسان را بر اساس آمار وقوع هم زمان gram-n اندازه گیری می کند (Kanapala et al., 2019). در مطالعات بررسی شده ROUGE-1 و ROUGE-2 هر کدام در ۲۷ مطالعه و ROUGE-L در ۲۳ مطالعه برای ارزیابی خلاصه های تولید شده مورد استفاده قرار گرفتند. معیار دقت، بازیافت و F-1 به ترتیب در ۹، ۱۰ و ۱۳ مطالعه استفاده شده اند. نکته حائز اهمیت این است که پژوهشگران برای ارزیابی تنها از یک معیار استفاده نکرده اند و خلاصه های تولید شده را از جهات مختلف مورد ارزیابی قرار داده اند تا به نتایج برتر دست پیدا کنند. لذا استفاده قابل توجهی از آمیخته ای از معیارهای ارزیابی توسط آن ها دیده می شود که در نمودار ۲ قابل مشاهده است. استفاده از چندین معیار برای ارزیابی کیفیت خلاصه ها خود نشان دهنده درک عمیق پژوهشگران از پیچیدگی های فرایند خلاصه سازی است. این امر نه تنها به ارائه نتایج دقیق تر و قابل اعتمادتر کمک می کند، بلکه امکان تطبیق و بهبود الگوریتم ها را نیز فراهم می آورد.

¹. Recall

². Precision



نمودار ۲- توزیع فراوانی آمیخته معیارهای ارزیابی در مطالعات

با توجه به این آمیختگی، شاید جای بررسی بیشتری در مورد تأثیر به کارگیری ترکیب‌های مختلف معیارها بر نتایج و عملکرد الگوریتم‌ها وجود داشته باشد. بهبود معیارهای موجود یا توسعه معیارهای جدید که جنبه‌های مهمی از کیفیت خلاصه‌ها را در نظر بگیرند، می‌تواند عنوان یک حوزه پژوهشی جدید در نظر گرفته شود.

نمودار ۲ نشان می‌دهد استفاده از معیارهای تک کلمه، جفت کلمه و طولانی‌ترین دنباله به‌عنوان زیر مجموعه‌های ROUGE در ۸ مطالعه به‌منظور ارزیابی خلاصه‌ها استفاده شده است. در ۴ مطالعه دیگر علاوه بر سه معیار پیش گفته، از دقت، بازیافت و F-1 نیز بهره گرفته شده است. آنچه مشهود است، تنها در دو مطالعه از معیار ROUGE استفاده نشده است. این نکته می‌تواند نشان‌دهنده اهمیت این معیار نسبت به سایر معیارها در امر ارزیابی خلاصه‌های تولیدشده توسط مدل‌های متنوع ارائه شده باشد. به نظر می‌رسد محبوبیت معیار ROUGE بین پژوهشگران این حوزه، به این دلیل است که بهتر از سایر معیارها می‌تواند مشخص کند خلاصه‌های سیستم تولیدشده با خلاصه‌های انسان قابل مقایسه است یا نه.

ط. چالش‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو

پرداختن به چالش‌های این حوزه به‌منظور اصلاح مدل‌ها و هموار کردن مسیر برای بهبود روش‌های موجود و ایجاد سیستم‌های مؤثرتر ضروری است و پژوهشگران را تشویق می‌کند. با زایش نوآوری‌های ثمربخش برای غلبه بر چالش‌های این حوزه تلاش کنند. ماحصل بررسی مطالعات این حوزه موضوعی بیانگر چالش‌های متنوعی است که در نمودار ۳ ارائه شده است.



نمودار ۳- چالش‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو

طبق نمودار بالا، عدم تناسب خلاصه تولید شده با پرس‌وجو از جمله چالش‌هایی است که در ۱۴ مطالعه به آن اشاره شده است. عدم انطباق میان هدف پرس‌وجو و اطلاعات استخراج شده می‌تواند به تولید نتایج ناکارآمد و گیج‌کننده منجر شود. بررسی عمیق‌تر علل این عدم تناسب می‌تواند شامل تحلیل فرایندهای انتخاب اطلاعات و قابلیت‌های الگوریتمی باشد که برای استخراج اطلاعات در نظر گرفته شده‌اند. برای حل این چالش، می‌توان به پیشنهاد تحلیل فرایندهای انتخاب اطلاعات دست زد. این تحلیل می‌تواند شامل استفاده از مدل‌های پیشرفته‌تر یادگیری عمیق باشد که قادر به درک بهتر نیازهای کاربر و تطبیق اطلاعات با پرسش‌ها هستند. چالش دیگر، فقدان داده‌های برچسب‌گذاری شده برای آموزش مدل‌ها است که در ۸ مطالعه دیده شده است. به‌طور مثال، 2005-2007 DUC به‌عنوان مجموعه داده‌های خلاصه‌سازی انتزاعی چندسندی متمرکز بر پرس‌وجو، داده آموزشی برچسب‌گذاری شده ندارند و فقط داده‌های آزمایشی ارائه می‌دهند. این امر می‌تواند کیفیت مدل‌های یادگیری ماشین را تحت تأثیر قرار دهد، زیرا بدون داده‌های برچسب‌گذاری شده مناسب، مدل‌ها نمی‌توانند به‌درستی ویژگی‌های کلیدی را یاد بگیرند. در این راستا، پیشنهاد بهبود روش‌های برچسب‌گذاری و جمع‌آوری داده‌ها می‌تواند به غلبه بر این مشکل کمک کند. مشکل افزونگی که در بخش‌های قبل گفته شد، بیشتر در خلاصه‌های چندسندی دیده می‌شود. عدم وجود معیارهای ارزیابی بهبودیافته برای ارزیابی دقیق خلاصه‌های تولید شده نیز چالش دیگری در این حوزه است. این کمبود می‌تواند بر تعیین استانداردهای عملکرد الگوریتم‌ها تأثیر بگذارد. پیشنهاد ایجاد معیارهای جدید که بتوانند به‌طور خاص کیفیت خلاصه‌های تولید شده را در زمینه پرس‌وجو ارزیابی کنند، می‌تواند به پیشرفت پژوهش‌های در این حوزه کمک کند. چالش دیگر مجموعه داده‌های محدود است. به‌طور مثال، Debate-

pedia که برای خلاصه‌سازی مبتنی بر پرس‌وجو پیشنهاد شده است مخصوص ورودی‌های تک‌سندی است و این محدودیت در استفاده برای ورودی‌های چندسندی چالش‌برانگیز شده است. در خصوص این چالش، کوچک بودن مجموعه داده‌های موجود برای خلاصه‌های مبتنی بر پرس‌وجوی تک‌سندی در مقایسه با مجموعه داده‌های خلاصه‌انتزاعی عمومی را می‌توان نام برد. نهایتاً چالش‌های مربوط به مجموعه داده‌ها به دنبال چالش فقدان مجموعه داده مخصوص خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو ایجاد شده است که ۵ پژوهشگر به آن اشاره کرده‌اند. به نظر می‌رسد اگر این چالش برطرف شود و مجموعه داده‌های مختص این نوع خلاصه‌سازی و برای هر دو نوع ورودی تک‌سندی و چندسندی ایجاد شود خودبه‌خود چالش‌های پیش‌گفته در مورد مجموعه داده‌ها نیز مرتفع خواهد شد. لذا، ایده ساخت این مجموعه داده می‌تواند پیشنهادی برای پژوهشگران جهت کمک به رفع برخی موانع تولید خلاصه‌های باکیفیت انتزاعی مبتنی بر پرس‌وجو باشد. سایر چالش‌ها عبارت‌اند از: ابهام معنایی ناشی از عدم تمایز بین جملات با معنای متفاوت، عدم تناسب خلاصه تولیدشده با متن منبع و عدم رابطه هم‌ترازی بین توالی‌های ورودی و خروجی که از کاستی‌های مکانیسم توجه به حساب می‌آید. تراز مطابقت بین خروجی و ورودی است. امتیاز تراز بالا نشان می‌دهد که رمزگشا برای تولید خروجی به‌شدت بر روی ورودی تمرکز کرده است. همچنین، نشان‌دهنده تمرکز مؤثر مدل بر ورودی و تولید خروجی متناسب است. برای بهبود این مشکلات، می‌توان از مکانیسم‌های متنوع توجه که تعامل و ارتباط بهتری بین بخش‌های مختلف متن برقرار می‌کنند، استفاده کرد.

نتیجه‌گیری

خلاصه‌سازی متون یک موضوع پژوهشی جالب و چالش‌برانگیز در پردازش زبان طبیعی است. این حوزه با وجود سابقه طولانی هنوز از اقبال خوبی در میان جامعه پژوهشی برخوردار است و یک حوزه پژوهشی بسیار پویا محسوب می‌شود. زیرا به‌طور پیوسته و گام‌به‌گام خود را با چالش‌ها و نیازهای جدید تطبیق می‌دهد. رشد سریع شبکه جهانی وب، افزایش اسناد الکترونیکی در بستر آن، اختصاص وقت برای مطالعه و پردازش این اسناد ارزش زمان را برای کاربران دوچندان کرده است. لذا نیاز به خلاصه‌سازی این اسناد را به‌منظور تشخیص سریع‌تر موضوع و محتوای متون افزایش داده است. از این رو، خلاصه‌سازی به‌عنوان یک حوزه پژوهشی حیاتی ظاهر شده است و پژوهشگران با اعمال نوآوری در مدل‌های ارائه‌شده سعی در پیشرفت این حوزه دارند. ثمره این تلاش‌ها پیشنهاد سیستم‌های خلاصه‌سازی مختلف است که هر کدام برای وظیفه خاصی در خلاصه‌سازی متون تدارک دیده شده‌اند. در

این میان، به نظر می‌رسد استفاده از پرس‌وجو برای خلاصه‌سازی می‌تواند بسیار ثمربخش باشد. خلاصه‌های که به این طریق حاصل می‌شود، خلاصه مبتنی بر پرس‌وجو نامیده می‌شود. در خلاصه‌سازی مبتنی بر پرس‌وجو، پرس‌وجو با محوریت آنچه مدنظر کاربر است اغلب یک کلمه یا عبارت است که به یک موجودیت خاص در سند یا اسناد اشاره دارد و از طریق آن نیاز اطلاعاتی کاربر مرتفع می‌شود. این نوع خلاصه‌سازی که با دو رویکرد انتزاعی و استخراجی قابل اجرا است؛ کاربردهای گسترده‌ای در دنیای واقعی دارد. در حال حاضر در ابزارهای مختلف از جمله موتورهای جستجو، مدل‌های زبانی و دستیارهای هوشمند به‌وفور در حال استفاده است و عملکرد این ابزارها تا حد زیادی به سیستمی بستگی دارد که دارای قابلیت پاسخگویی به پرس‌وجوی کاربران باشد. به خاطر پاسخگویی رضایت‌بخش این ابزارها در اکثر مواقع، کمتر دیده‌شده که کاربران وب از آن‌ها برای رفع نیازهای اطلاعاتی و نیازهای دانشی خود در حوزه‌های مختلف استفاده نکرده باشند. بخش عمده‌ای از این موفقیت ناشی از به‌کارگیری مدل‌ها و روش‌های متنوع در پیوند با خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو است که به کاربران امکان می‌دهد تا به اطلاعات دقیق و مرتبط دست یابند. به‌وضوح، این رویکردها نه تنها کارایی جستجوی اطلاعات را ارتقا می‌دهند، بلکه تجربه کاربری را نیز به نحو چشمگیری بهبود می‌بخشند. این تحولات نشان‌دهنده اهمیت و ضرورت استمرار پژوهش و توسعه در زمینه خلاصه‌سازی متون، به‌ویژه در ارتباط با فنون مبتنی بر پرس‌وجو است که می‌تواند به پاسخگویی بهتر به نیازهای کاربران در دنیای داده محور کنونی کمک کند.

با توجه به اهمیت حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو، سهم این پژوهش ارائه یک نمای کلی از مطالعات پیرامون این حوزه در قالب یک مرور نظام‌مند است که با هدف کمک به پژوهشگران در ارائه راهکارها و نوآوری‌هایی مؤثر به‌منظور سوق دادن روش‌شناسی این حوزه به سمت اثربخشی بیشتر و درنهایت بهبود خلاصه‌های انتزاعی مبتنی بر پرس‌وجو انجام شده است. نکته‌ای که از جستجوهای انجام‌شده در پایگاه‌های مختلف برای شناسایی مطالعات منتشرشده پیرامون این حوزه مشخص شد اقبال بیشتر خلاصه‌های استخراجی مبتنی بر پرس‌وجو است که موضوع گرایش‌های موردعلاقه فعلی بیشتر پژوهشگران حوزه خلاصه‌سازی است. از آنجاکه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو نیازمند استنباط، استنتاج و تحلیل بیشتر است و در این راه با چالش‌های زیاده‌تری روبه‌روست؛ شاید بتوان علت اقبال بیشتر همتای استخراجی آن را آسان بودن آن تلقی کرد. همچنین، می‌توان گفت شاید هنوز مهم‌ترین ویژگی‌ها برای تولید یک خلاصه رضایت‌بخش طول جمله، موقعیت جمله، فراوانی‌ها و کلیدواژه‌ها باشد که در خلاصه‌سازی استخراجی به آن‌ها توجه می‌شود.

پس از نهایی شدن مطالعات مشمول مرور نظام‌مند توسط دستورالعمل پریزما، اولین مورد بررسی شده در این پژوهش نوع مطالعات منتشر شده در این حوزه و سیر انتشار آن‌ها بود. در اکثر موارد، مطالعات از نوع مقالات منتشر شده در مجلات بوده و سیر انتشار آن‌ها علاوه بر اینکه گویای قدمت کم انتشار این مطالعات بوده، نشان‌دهنده حرکت سینوسی انتشار آن‌ها نیز است. با توجه به نقش بسزای این نوع خلاصه‌سازی در ابزارهای کنونی انتظار می‌رود مطالعات بیشتری در این حوزه منتشر شود و پیشینه آن از غنای بهتری برخوردار گردد. این امر می‌تواند به شناخت بهتر و عمیق‌تری از روش‌های جدید و فنون مؤثر کمک کند. رویکردهای مورد استفاده نیز در اکثر موارد رویکرد مستقیم و یک مرحله‌ای بوده و تعداد کمی با رویکرد دوم مرحله‌ای استخراجی-انتزاعی به ارائه سیستم خلاصه‌ساز پرداخته‌اند. علت بارز این امر می‌تواند پیچیدگی و مضاعف شدن چالش‌های موجود از جمله مجموعه داده‌های محدود، عدم تناسب خلاصه تولید شده با پرس و جو، عدم وجود معیارهای ارزیابی بهبود یافته، فقدان داده‌های برچسب‌گذاری شده برای آموزش مدل‌ها، عدم تناسب خلاصه تولید شده با متن منبع و غیره باشد که موجب عملکرد ضعیف سیستم می‌شود. با آنکه فقدان مجموعه داده و محدودیت مجموعه داده‌های حاضر از طریق به‌کارگیری فنون نظارت ضعیف، تطبیق دامنه، یادگیری بدون نظارت و غیره تا حدی قابل جبران است. اما پژوهشگران در اکثر موارد رویکرد یک مرحله‌ای را ترجیح داده‌اند. به نظر می‌رسد پتانسیل روش‌های دوم مرحله‌ای برای دستیابی به خلاصه‌هایی که بیشتر شبیه انسان است می‌تواند با مرتفع شدن بیشتر چالش‌های مذکور افزایش یافته و در رقابت با خلاصه‌های انسانی قرار گیرد. از سویی دیگر، ممکن است چالش‌های موجود، خود بازدارنده بوده و مانع از انجام رویکرد دوم مرحله‌ای شود. یک راهکار عملی می‌تواند توسعه و بهبود مجموعه داده‌ها باشد. این راهکار یکی از اولین و ضروری‌ترین مراحل ایجاد و گسترش مجموعه داده‌های برچسب‌گذاری شده با کیفیت است. برای عملیاتی‌سازی آن می‌توان از فنون جمع‌آوری داده‌های بهینه، تعامل با کاربران و استفاده از داده‌های عمومی واقعی برای تأمین داده‌ها بهره برد. البته با استفاده از فنون یادگیری بدون نظارت و یادگیری تقویتی، پژوهشگران می‌توانند بدون نیاز به داده‌های برچسب‌گذاری شده، سیستم‌های خود را بهبود دهند و طرح‌های نوآورانه‌تری را ایجاد کنند.

بررسی انواع یادگیری در مطالعات نشان داد، پژوهشگران بیشتر از یادگیری با نظارت و خودنظارتی استفاده کرده‌اند. یادگیری بدون نظارت و یادگیری انتقالی نیز در تعداد معدودی مورد استفاده بوده‌اند. اگرچه کاربرد این دو مورد کمتر است، اما می‌توانند به‌عنوان راهکار مکمل برای بهبود دقت و کیفیت خلاصه‌ها در برخی شرایط خاص مطرح شوند. به‌طور مثال، استفاده از یادگیری انتقالی برای به‌کارگیری دانش و ویژگی‌های یک مدل آموزش دیده در

یک دامنه دیگر می‌تواند پتانسیل‌های جدیدی را برای خلاصه‌سازی متون فراهم کند. مشاهده هم‌افزایی یادگیری‌ها در برخی مطالعات نشان‌دهنده تلاشی پویا به‌منظور خلاصه‌سازی مؤثرتر است که روش‌شناسی این حوزه را ارتقاء می‌بخشد. البته برای پردازش داده‌های بزرگ و اجرای مدل‌های پیچیده ناشی از هم‌افزایی‌ها، نیاز به زیرساخت‌های محاسباتی قدرتمند وجود دارد. سرمایه‌گذاری در این زمینه می‌تواند به افزایش کارایی مدل‌ها کمک کند و پژوهشگران را قادر سازد تا مدل‌های پیچیده اما کارآمدتری را پیاده‌سازی کنند. تحولات ایجادشده توسط مدل‌های پیچیده و کارآمد، می‌تواند تأثیرات اقتصادی و اجتماعی مثبتی را به همراه داشته باشد. این مهم می‌تواند شامل تسهیل دسترسی به اطلاعات، بهبود فرایندهای تصمیم‌گیری و افزایش بهره‌وری از متون در حوزه‌های مختلف باشد.

مسیر تکامل و پیشرفت خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو روش‌های متفاوتی از جمله روش‌های مبتنی بر قاعده، مبتنی بر آمار و یادگیری ماشین را طی کرده است. آنچه واضح است این نکته است که این مسیر، مسیری پویاست و گام‌های مهمی را در طول زمان نشان می‌دهد. روش‌های اولیه که در آغاز توسعه خلاصه‌سازی به کار گرفته شدند؛ بر اساس معیارهای مختلف مانند طول جمله، فراوانی کلمات کلیدی و وجود ویژگی‌های زبانی خاص به جملات امتیاز می‌دهد. این روش‌ها، به‌شدت وابسته به کیفیت قاعده‌ها و توانایی آن‌ها در درک پیچیدگی‌های زبانی بودند. بعد از آن‌ها روش‌های مبتنی بر آمار مورد استقبال قرار گرفتند که با استفاده از آمار توصیفی و تحلیل‌های کمی تلاش داشتند به شناسایی الگوهای عمومی در داده‌ها بپردازند. با ظهور یادگیری ماشین، سیستم‌های خلاصه‌سازی انتزاعی وارد مرحله‌ای جدید و پیشرفته‌تری شدند. مزیت اصلی این روش‌ها در توانایی یادگیری از حجم بالای داده و کشف الگوها و روابط پیچیده معنایی است. این روش‌ها با درک معنا سروکار دارند. توانایی درک این روش‌ها برای توسعه سیستم‌های خلاصه‌ساز کارآمدتر و مؤثرتر ضروری به نظر می‌رسد.

بررسی‌ها نشان می‌دهد روش‌های یادگیری ماشین به دلیل قابلیت یادگیری و نوآوری مداوم، بیشترین پتانسیل را برای توسعه و بهبود سیستم‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو دارند. اما، روش‌های مبتنی بر قانون و آمار نمی‌توانند الگوهای پیچیده و روابط معنایی را بیاموزند. با آنکه یادگیری ماشینی به دلیل عملکرد خوبی که دارد روش موردعلاقه پژوهشگران زیادی است؛ لیکن بهترین روش از همه جهات نیست. به‌طور مثال در صورت عدم وجود داده‌های کافی یادگیری آن با چالش‌هایی از جمله افت دقت و کاهش توانایی در تعمیم به موقعیت‌های جدید مواجه خواهد شد. البته نباید از اهمیت یادگیری غیرنظارتی و تأثیر آن در بهبود این نقطه‌ضعف غافل شد. با توجه به اینکه این روش به داده‌های

برچسب‌گذاری شده کمتری نیاز دارد، می‌تواند به‌طور قابل‌توجهی در شرایطی که دسترسی به داده‌های آموزشی محدود است مورد استفاده قرار بگیرد. لذا این راهکار می‌تواند در کنار قدرتمندی بسیار بیشتر یادگیری ماشین نسبت به روش‌های آماری و مبتنی بر قاعده، نقطه‌ضعف آن را هنگام کمبود داده‌های آموزشی مرتفع سازد. از این‌رو، استفاده بیشتر از آن در پژوهش‌های آتی پیشنهاد می‌شود. از طرف دیگر، روشی که به راحتی با بقیه روش‌ها ترکیب می‌شود، روش آماری است که هم‌افزایی آن با روش‌های دیگر در مطالعات مورد بررسی دیده شده است و یک ترکیب و تعامل مؤثر برای بهبود عملکرد سیستم‌ها را ارائه کرده است. خصلت کمی روش آماری این اعتقاد را تقویت می‌کند که با ترکیب شدن با سایر فنون به ایجاد سامانه‌های قدرتمندتر کمک می‌کند. بنابراین، استفاده از روش‌های آماری و یادگیری ماشین و شناسایی نقاط قوت هر یک در کنار یکدیگر می‌تواند منجر به توسعه مدل‌های ترکیبی شود که می‌توانند به‌صورت هم‌زمان از مزایای پیش‌بینی و تجزیه و تحلیل دقیق بهره‌برداری کنند. پژوهش‌های آینده می‌تواند به طراحی این مدل‌های ترکیبی متمرکز باشد و چشم‌انداز خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو در پیوند با این روش‌ها و ایجاد سیستم‌های نوآورانه جدید قرار دارد که منعکس‌کننده کاوش و نوآوری مداوم است. نکته حائز اهمیت اینکه، سیستم‌های خلاصه‌سازی که انعطاف‌پذیری بیشتری در برابر تغییرات و نیازهای جدید کاربران دارند، می‌تواند منجر به توسعه پایدارتری در این حوزه شود. فراهم کردن زیرساخت‌هایی که به سیستم‌ها اجازه می‌دهد به‌طور مداوم به‌روزرسانی شوند، اهمیت دارد. لذا، به‌کارگیری این قابلیت در سیستم‌های آتی با بهره‌گیری از یادگیری ماشین که به‌عنوان روش برتر ساخت خلاصه‌سازها شناسایی شده است، پیشنهاد می‌شود. علاوه بر این، تعامل با سیستم‌های هوش مصنوعی از جمله پردازش زبان طبیعی و تحلیل احساسات، قدرت تحلیل‌های متنی را افزایش داده و به درک بهتری از محتوای متنی منجر می‌شود، بنابراین برقراری پیوند با آن‌ها می‌تواند در بهبود سیستم‌های خلاصه‌ساز راهگشا باشد.

مرور مطالعات استفاده از مدل‌های مبتنی بر گراف و مدل‌های مبتنی بر شبکه‌های عصبی که در اکثر موارد از ترنسفورمرهای مبتنی بر مکانیسم‌های مختلف توجه بهره‌برده‌اند را نشان می‌دهد که بر اساس معماری رمزگذار-رمزگشا بوده و نوعی مدل دنباله به دنباله محسوب می‌شوند. با آنکه با پیشرفت شبکه‌های عصبی در استفاده از معماری رمزگذار-رمزگشا شاهد رشد کیفیت خلاصه‌های تولیدشده در مطالعات مورد بررسی هستیم؛ مدل‌های رمزگذار-رمزگشا نیز عاری از مشکل نبوده و در برخی موارد یک کلمه را چندین بار تکرار می‌کنند که باعث افزایش افزونگی و خدشه‌دار شدن انسجام خلاصه‌ها می‌شود. لذا، پیشنهاد می‌شود که پژوهشگران مدل‌های ترکیبی که از قابلیت‌های ترنسفورمر در کنار روش‌های کلاسیک بهره

می‌برند را توسعه دهند. همچنین، پژوهش در زمینه مقایسه عملکردهای مختلف مدل‌ها در شرایط واقعی و در زمینه‌های مختلف می‌تواند به بهبود کیفیت خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو کمک شایانی نماید.

بررسی مطالعات نشان می‌دهد علاوه بر مدل‌های بالا، توسعه مدل‌های از پیش آموزش دیده، پژوهشگران را به اصلاح و به‌کارگیری آن‌ها در زمینه تولید خلاصه‌های انتزاعی مبتنی بر پرس‌وجو برانگیخته است. یکی از مزیت‌های اصلی این مدل‌ها، کاهش زمان لازم برای آموزش از صفر و هزینه‌های مرتبط با آن است. استفاده از داده‌های بزرگ عمومی به عنوان ورودی، می‌تواند به مدل کمک کند تا به سرعت بتواند دانش عمومی را فراگیرد و سپس به کاربر نهایی ارائه دهد. همچنین، به واسطه به‌کارگیری داده‌های عمومی و بزرگ می‌توانند به‌سادگی به موضوعات جدید تعمیم یابند. تأکید بر این مزایای مدل‌های از پیش آموزش دیده می‌تواند به ایجاد مدل‌های اقتصادی و کارآمد برای پژوهشگران و توسعه‌دهندگان کمک کند. مدل‌های از پیش آموزش دیده مانند برت و جی‌پی‌تی اخیراً به نقطه عطفی در هوش مصنوعی تبدیل شده‌اند. از جمله مزایای برت، بهبود قابلیت تعامل معنایی و دقت در پردازش زبان طبیعی است. این مدل به راحتی می‌تواند با داده‌های جدید تنظیم شود. جی‌پی‌تی نیز به دلیل قدرت پردازش زبان طبیعی به صورت مولد، می‌تواند در موقعیت‌های مختلف بدون نیاز به آموزش مجدد عمیق عمل کند. از جمله وجوه تمایز این دو مدل این است که توجه در برت حالت دوطرفه دارد. در حالی که توجه جی‌پی‌تی یک‌طرفه است. برت بیشتر مناسب وظایف درک متن و تجزیه و تحلیل است. جی‌پی‌تی بیشتر برای تولید متن مورد استفاده قرار می‌گیرد. به‌طور کلی، این مدل‌ها با فنون خیلی پیشرفته‌تری طراحی شده‌اند و قادرند وابستگی‌های پیچیده زبانی و الگوهای معنایی را کشف کنند. این سیستم‌ها می‌توانند درک عمیق‌تری از متون و ارتباطات درون آن‌ها داشته باشند. به نظر می‌رسد تمایل جامعه هوش مصنوعی بر آن است که به جای یادگیری مدل‌ها از ابتدا، استفاده از این مدل‌ها به‌طور گسترده رواج پیدا کند. از این رو، استفاده بیشتری از این مدل‌ها در سیستم‌های خلاصه‌سازی پیشنهادی نیز دیده می‌شود.

نکته قابل توجه، چشمگیر بودن هم‌افزایی مدل‌های مذکور در برخی مطالعات است که هر کدام جنبه‌هایی از نوآوری در این حوزه موضوعی محسوب می‌شود. با این حال، با همه هم‌افزایی‌های انجام شده درک معنایی بالا در موارد مختلفی اتفاق نیفتاده است. این موضوع به‌ویژه در مواردی با محتوای پیچیده یا غنی از روابط معنایی بیشتر دیده می‌شود. درک معنایی در سیستم‌ها توانایی آن‌ها را از ساختارهای نحوی فراتر می‌برد و منجر به فهم عمیق متن

شده و پتانسیل واقعی سیستم‌ها را نمایان می‌کند. این امر در مطالعات مورد بررسی با وجود چالش‌های فراوان به‌طور بالفعل جامه عمل نپوشیده است و مدل‌ها اغلب قادر به شناسایی ساختارهای زبانی هستند؛ اما در درک و تجزیه و تحلیل عمیق معنایی و زمینه‌ای مشکل دارند. برای رویارویی با چالش‌های مرتبط با درک معنایی چندین راهبرد می‌تواند مفید باشد. از جمله این موارد، استفاده از داده‌های غنی و چندمنظوره است. بدین صورت که جمع‌آوری داده‌های آموزشی متنوع، شامل متن‌های با موضوعات مختلف و تنوع زبانی می‌تواند منجر به بهبود درک معنایی در مدل‌ها شود. همچنین، توسعه الگوریتم‌های پیشرفته با انجام پژوهش‌ها در زمینه الگوریتم‌های جدید یادگیری عمیق می‌تواند نشان‌دهنده روش‌های جدیدی برای بهبود درک معنایی و تحلیل عمیق‌تر متن باشد. آزمایش و به‌روزرسانی مداوم مدل‌ها می‌تواند راهکار دیگری در این خصوص باشد. در این راستا، بازنگری و به‌روزرسانی‌های منظم برای مدل‌ها بر اساس داده‌های جدید و تغییرات زبان ضروری است. همچنین، ایجاد ابزارهای تعاملی که کاربران می‌توانند به راحتی بازخورد دهند و مدل‌ها را تحت تأثیر قرار دهند؛ می‌تواند به مدل‌ها کمک کند تا یادگیری مستمر داشته باشند و در درک معنا موفق‌تر عمل کنند و در نهایت نسبت به نیازهای کاربران پاسخگوتر باشند.

از آنجاکه نوع ورودی تک‌سندی نسبت به ورودی چندسندی با چالش‌های کمتری از جمله افزونگی مواجه است؛ نوع ورودی مدل‌های پیشنهادی نیز در اکثر مطالعات تک‌سندی بوده است. در برخی موارد نیز از هر دو نوع ورودی بهره گرفته شده است. البته پژوهشگران از راه‌های مختلف سعی در کاهش یا برطرف کردن افزونگی حاصل از خلاصه‌سازی‌های چندسندی کرده‌اند. مثال بارز آن استفاده از الگوریتم حداقل ارتباط حاشی‌های، ایجاد سه‌تایی پرس‌وجو، سند، خلاصه؛ خوشه‌بندی، روش‌های مبتنی بر گراف و ماشین بولتزن محدود برای مقابله با این چالش است. بدین ترتیب که الگوریتم حداقل ارتباط حاشی‌های به‌منظور شناسایی نقاط مثبت و موارد تکراری در متون مورد استفاده قرار می‌گیرد. ایجاد سه‌تایی پرس‌وجو، سند و خلاصه به مدل‌ها کمک می‌کند تا رابطه بین پرسش، اطلاعات مربوطه و نتیجه قابل قبول را درک کنند. روش‌های خوشه‌بندی و مبتنی بر گراف می‌توانند به تجزیه و تحلیل و گروه‌بندی اطلاعات کمک کنند تا از تکرار و افزونگی جلوگیری شود. استفاده از ماشین بولتزن محدود نیز برای بهینه‌سازی و انتخاب بهترین خلاصه‌ها از میان محتوای چندگانه به کار می‌رود.

از جمله مجموعه داده‌های محبوب در حوزه خلاصه‌سازی مبتنی بر پرس‌وجو Debate-
DailyMail/CNN، pedia و DUC را می‌توان نام برد. Debatepedia مجموعه داده‌ای است که در سال ۲۰۱۷ به‌منظور خلاصه‌سازی مبتنی بر پرس‌وجو برای سیستم‌هایی با ورودی‌های

تک‌سندی پیشنهاد شد. لذا برای اهداف چندسندی با محدودیت رویه‌رو است. مجموعه داده DailyMail/CNN مخصوص خلاصه‌سازی مبتنی بر پرس‌وجو نیست، اما به خاطر چالش فقدان مجموعه داده‌های اختصاصی در این حوزه در مطالعات زیادی از آن استفاده شده است. ارزیابی یک خلاصه، چالش بزرگی است که سال‌ها با آن‌ها مقابله می‌شود. بررسی‌ها نشان داده است از میان معیارهای ارزیابی مورد استفاده معیارهای ROUGE با gram-n مختلف، دقت، بازیافت و F-1 از محبوبیت بیشتری برخوردار بوده‌اند. آنچه پیرامون چالش‌های موجود در این حوزه به دست آمد گویای کافی نبودن این معیارها برای ارزیابی دقیق‌تر خلاصه‌های تولید شده توسط سیستم‌های ارائه شده در مطالعات است. به طور مثال، معیار ROUGE وابستگی به شباهت‌های الگوی n-gram دارد و ممکن است نتواند به طور کامل ابعاد معنایی، ساختاری و سبک نوشتاری خلاصه‌ها را در نظر بگیرد. دو خلاصه ممکن است از نظر n-gram شباهت بالایی داشته باشند، ممکن است از نظر اطلاعات کلیدی و انتقال مفهوم، تفاوت‌های فاحشی داشته باشند. لذا با آنکه معیارهای پیش گفته در میان هفده معیار شناسایی شده استفاده‌پذیری بیشتری داشته‌اند و در اکثر موارد نیز در استفاده از سایر معیارها نیز هم‌افزایی‌های متنوعی در مطالعات دیده می‌شود؛ نمی‌توان گفت خلاصه‌های تولید شده سیستمی دقیقاً آن اندازه تفاوت با خلاصه‌های انسانی دارند که توسط این معیارها اعلام شده است. از این رو، بهبود این معیارها یا ارائه معیارهای جدید برای ارزیابی دقیق خلاصه‌های سیستمی مورد نیاز است. با توجه به این چالش‌ها، پیشنهاد می‌شود که پژوهشگران به سمت توسعه و استفاده از معیارهای جدیدی که به جنبه‌های معنایی و مفهومی توجه دارند حرکت کنند. معیارهایی مانند BLEU و ME-TEOR که بر روی شباهت‌های معنایی و نحوی تمرکز دارند، می‌توانند مکمل‌های مناسبی برای ROUGE باشند. علاوه بر این، استفاده از معیارهای مبتنی بر یادگیری عمیق می‌تواند به تدریج به ارزیابی‌های بهتری منجر شود. این ابزارها قادر به شناسایی و ارزیابی مفاهیم و روابط پیچیده‌تری هستند که معیارهای سنتی از آن‌ها غافل هستند.

به طور کلی، بررسی فنون، مدل‌ها و روش‌های مورد استفاده در مطالعات حوزه خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو نشان داد که چگونه هم‌افزایی‌های اتفاق افتاده تا حدی موجب بهبود عملکرد کلی سیستم‌های خلاصه‌سازی شده و به توسعه آن‌ها کمک نموده است. اما، آنچه مهم است توانایی درک معناست که در مطالعات مورد بررسی هنوز حالت بالفعل آن، فاصله میان خلاصه‌های سیستمی و خلاصه‌های انسانی را پر نکرده است. زیرا معنای درک شده توسط این سیستم‌ها هنوز سطحی بوده و تا حدی وابستگی به ساختارهای نحوی در آن‌ها دیده می‌شود. به نظر می‌رسد این امر ناشی از چالش‌های موجود در این حوزه است. از این رو، تشویق به ادامه پژوهش‌ها برای غلبه بر این چالش‌ها و ایجاد

نوآوری به‌منظور خلق سیستم‌های توانمند در درک معنا بسیار مهم است. درواقع، توانایی درک معنا ضامن ایجاد سیستم‌هایی خواهد بود که معنا و بینش‌های عمیق نهفته در متن را تشخیص داده و بر اساس وظیفه مشخص‌شده، آن‌ها را در خروجی خود اعمال می‌کنند. در این راستا، می‌توان راهکارهای عملی ازجمله تقویت پژوهش و توسعه را برای نهادینه کردن معنا در مدل‌ها و بهبود وضعیت فعلی در این حوزه موردتوجه قرار داد. برخی از این راهکارها شامل اقداماتی به شرح زیر است:

- سرمایه‌گذاری در پژوهش‌ها برای بهبود الگوریتم‌ها و مدل‌های خلاصه‌سازی انتزاعی
- برقراری همکاری میان پژوهشگران و مهندسان داده برای توسعه مدل‌های جدید و کارآمد
- آموزش و توانمندسازی پژوهشگران و تشویق آن‌ها به شرکت در پروژه‌های مشترک و تبادل تجربیات
- توسعه پایگاه‌های داده عمومی با مجموعه‌های متنوع متون و خلاصه‌ها به‌طوری‌که پژوهشگران بتوانند الگوریتم‌ها را با داده‌های واقعی مورد آزمایش قرار دهند
- ایجاد داده‌های آموزشی تخصصی با دامنه‌های مختلف؛
- بازنگری در معیارهای فعلی به‌منظور ارزیابی کیفیت خلاصه‌ها و توسعه معیارهای جدید
- ساخت تجهیزات و سیستم‌های قدرتمند برای یادگیری و آزمایش مدل‌ها با حجم عظیم داده
- بررسی آینده‌پژوهی حوزه خلاصه‌سازی انتزاعی با محوریت پرس‌وجوی کاربران؛
- افزودن قابلیت تحلیل احساسات به مدل‌های خلاصه‌سازی؛
- بررسی مفهوم شناسی مدل‌ها با قابلیت بازخورد گیری لحظه‌ای از کاربر؛
- ایجاد تعامل میان پژوهشگران و کاربران جهت بهبود سیستم‌های خلاصه‌سازی؛
- ایجاد امکان شخصی‌سازی مدل‌های خلاصه‌سازی؛
- ایجاد سیستم‌های خلاصه‌سازی سفارشی؛
- ایجاد امکان خلاصه‌سازی چندرسانه‌ای در مدل‌های خلاصه‌سازی با تمرکز بر دامنه‌های متنوع
- توجه بیشتر به یادگیری تقویتی در هم‌افزایی با سایر روش‌های یادگیری؛
- ایجاد و به‌کارگیری سیستم‌های ارزیابی خودکار هم‌زمان با ساخت خلاصه؛
- اعمال قابلیت تحلیل متن چندزبانه.

به نظر می‌رسد زایش نوآوری‌های پیشرفته از دل این راهکارها به‌منظور رفع چالش‌ها و بهبود وضعیت موجود، مدل‌سازی معنایی و درک معنا را در سیستم‌های خلاصه‌سازی انتزاعی مبتنی بر پرس‌وجو نهادینه کرده و به اصلاح و پیشرفت مسیر تکامل روش‌شناسی‌های موجود کمک خواهد نمود. نهاده‌سازی مدل‌سازی معنایی در سیستم‌های خلاصه‌سازی می‌تواند به دو جنبه معطوف شود: الف. توانایی درک عمیق‌تر متن و ارتباطات معنایی بین اجزای مختلف آن؛ و ب. قابلیت تولید خلاصه‌هایی که نه تنها به اطلاعات اصلی اشاره دارند، بلکه مفهوم و زمینه متن را نیز حفظ کنند. این موضوع به‌ویژه در خلاصه‌سازی انتزاعی که نیاز به تفسیر و تولید مجدد محتوا دارد، از اهمیت بالایی برخوردار است. نکته حائز اهمیت اینکه برای نیل به این امر، علاوه بر موارد بالا باید با تغییر و تکامل منابع اطلاعاتی و تحولات درخواست‌های کاربران و زمینه‌های دانشی آن‌ها همگام شد. زیرا به نظر می‌رسد کاربر فعلی با کاربر ۲ سال پیش متفاوت است. کاربر دو سال بعد نیز با کاربر فعلی متفاوت خواهد بود. این مسئله نشان‌دهنده اهمیت انعطاف‌پذیری سیستم‌ها و توانایی آن‌ها در انطباق با خواسته‌های متغیر کاربران است. پژوهش در زمینه‌شناسایی الگوهای تغییر نیازهای کاربران و چگونگی پاسخگویی سیستم‌های خلاصه‌سازی به این تغییرات می‌تواند به بهبود کارایی و مرتبط بودن آن‌ها منجر شود.

افزون بر این، سیستم‌های پیشنهادی در مطالعات بررسی شده به‌منظور خلاصه‌سازی متون انگلیسی ساخته شده‌اند. از این حیث خلأ ارائه سیستم‌هایی برای زبان‌های دیگر احساس می‌شود. به نظر می‌رسد این امر با ایجاد و تقویت ابزارهای پردازش زبان طبیعی مانند برچسب‌گذاری بخشی از گفتار، ریشه‌گیری، تشخیص موجودیت‌های نامدار و نظایر آن برای زبان‌های غیرانگلیسی قابلیت عملیاتی‌سازی دارد. سرمایه‌گذاری بر روی پروژه‌های مشارکتی بین کشورها می‌تواند به تبادل دانش و بهبود فناوری‌های پردازش زبان طبیعی در زبان‌های مختلف کمک کند. این امر همچنین می‌تواند به شناسایی و ارائه نیازهای خاص فرهنگی و زبانی هر کشور منجر شود.

منابع

- استراوس، آنسلم، و کوربین، جولیت (۱۳۹۰). اصول روش تحقیق کیفی: نظریه‌مبنایی، رویه‌ها و شیوه‌ها (بیوک محمدی، مترجم). تهران: پژوهشگاه علوم انسانی و مطالعات فرهنگی.
(اثر اصلی منتشر شده در سال ۱۹۹۰)
بهشتی، سید صمد (۱۳۹۵). تحلیل داده‌های کیفی با نرم‌افزار Maxqda (سید شاه‌رخ موسویان، ویراستار). تهران: روش‌شناسان.
حریری، نجلا (۱۳۸۵). اصول و روش‌های پژوهش کیفی. تهران: دانشگاه آزاد اسلامی واحد علوم و تحقیقات.

کوهن، تامس (۱۳۹۳). *ساختار انقلاب‌های علمی* (سعید زیباکلام، مترجم). تهران: سمت. (اثر اصلی منتشر شده در سال ۱۹۶۲)

References

- Abdullah, D. M. (2020). *Query focused abstractive summarization using BERTSUM model* [Master's thesis, University of Lethbridge]. URL: <https://opus.uleth.ca/handle/10133/5760>
- Abdullah, D. M., & Chali, Y. (2020). Towards generating query to perform query focused abstractive summarization using pre-trained model. In *Proceedings of the 13th International conference on natural language generation* (pp. 80-85). URL: <https://aclanthology.org/2020.inlg-1.11/>
- Alambo, A., Lohstroh, C., Madaus, E., Padhee, S., Foster, B., Banerjee, T., Thirunarayan, k., & Raymer, M. (2020, December). Topic-centric unsupervised multi-document summarization of scientific and news articles. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 591-596). URL: <https://arxiv.org/abs/2011.08072>
- Alanzi, E., & Alballaa, S. (2023). Query-Focused Multi-document Summarization Survey. *International Journal of Advanced Computer Science and Applications*, 14(6), 822-833. URL: <http://dx.doi.org/10.14569/IJACSA.2023.0140688>
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text summarization techniques: a brief survey. arXiv preprint. URL: <https://arxiv.org/abs/1707.02268>
- Andhale, N., & Bewoor, L. A. (2016). *An overview of text summarization techniques*. In 2016 international conference on computing communication control and automation (ICCUBE), Pune, India, 2016. URL: <https://ieeexplore.ieee.org/document/7860024>
- Arthur, M. P., Rameshchandra, T. J., & Dhanabalachandran, M. (2023). Abstractive Summarization Based Question-Answer System for Structural Information. In *International Conference on Applications of Natural Language to Information Systems* (pp. 416-427). Cham: Springer Nature Switzerland. URL: https://doi.org/10.1007/978-3-031-35320-8_30
- Aryal, C. (2019). Query-Focused Abstractive Summarization using Neural Networks [Master's thesis, University of Lethbridge] URL: <https://hdl.handle.net/10133/5400>
- Baumel, T., Eyal, M., & Elhadad, M. (2018). Query focused abstractive summarization: Incorporating query relevance, multi-document coverage, and summary length constraints into seq2seq models. arXiv preprint. URL: <https://arxiv.org/abs/1801.07704>
- Beheshti, S. S. (2016). Analysis of Qualitative Data with Maxqda Software (S. S. Mousavian, Ed). Tehran: Ravesh shenasan. (Original work published) [In Persian]
- Berners-Lee, T. (1998). *Semantic web road map*. URL: <https://www.w3.org/DesignIssues/Semantic.html>
- Deng, Y., Zhang, W., & Lam, W. (2020). Multi-hop inference for question-driven summarization. arXiv preprint. URL: <https://arxiv.org/abs/2010.03738>
- Deng, Y., Zhang, W., Xu, W., Shen, Y., & Lam, W. (2023). Nonfactoid question answering

- as query-focused summarization with graph-enhanced multi hop inference. *IEEE Transactions on Neural Networks and Learning Systems*, 35 (8), 1-14. URL: <https://ieeexplore.ieee.org/document/10083216>
- Dong, Y. (2018). A survey on neural network-based summarization methods. arXiv preprint arXiv. URL: <https://arxiv.org/abs/1804.04589>
- Esteva, A., Kale, A., Paulus, R., Hashimoto, K., Yin, W., Radev, D., & Socher, R. (2021). COVID-19 information retrieval with deep-learning based semantic search, *question answering, and abstractive summarization*. *NPJ digital medicine*, 4(1), 68. URL: <https://doi.org/10.1038/s41746-021-00437-0>
- Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1), 1-66. URL: <https://link.springer.com/article/10.1007/s10462-016-9475-9>
- Gavalan, H. S., Rastgoo, M. N., & Nakisa, B. (2024). A BERT-Based Summarization approach for depression detection. arXiv preprint. URL: <https://arxiv.org/pdf/2409.08483>
- Giarelis, N., Mastrokostas, C., & Karacapilidis, N. (2023). Abstractive vs. extractive summarization: *An experimental review*. *Applied Sciences*, 13(13), 7620. URL: <https://doi.org/10.3390/app13137620>
- Girithana, K., & Swamynathan, S. (2019). Query oriented extractive-abstractive summarization system (QEASS). In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data* (pp. 301-305). URL: <https://doi.org/10.1145/3297001.3297046>
- Hariri, N. (2002). *Principles and Methods of Qualitative Research*. Tehran: Islamic Azad University, Science and Research Branch [In Persian]
- Han, H., & Choi, J. (2024). Optimal path for Biomedical Text Summarization Using Pointer GPT. arXiv preprint. URL: <https://arxiv.org/pdf/2404.08654>
- Hasselqvist, J., Helmertz, N., & Kågebäck, M. (2017). Query-based abstractive summarization using neural networks. arXiv preprint. URL: <https://arxiv.org/abs/1712.06100>
- Inoue, N., Trivedi, H., Sinha, S., Balasubramanian, N., & Inui, K. (2021). Summarize-then-answer: Generating concise explanations for multi-hop reading comprehension. arXiv preprint. URL: <https://arxiv.org/abs/2109.06853>
- Israel, Q., Han, H., & Song, I. Y. (2015). Semantic analysis for focused multi-document summarization (fMDS) of text. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing* (pp. 339-344). URL: <https://doi.org/10.1145/2695664.2695672>
- Jeyakarthic, M., & Leoraj, A. (2024). *Knowledge-Infused Corpus Building for Context-Aware Summarization with Bert Model*. *Migration Letters*, 21(S4), 1681-1694. <https://migrationletters.com/index.php/ml/article/view/7588>
- Kanapala, A., Pal, S., & Pamula, R. (2019). Text summarization from legal documents: a survey. *Artificial Intelligence Review*, 51, 371-402. URL: <https://link.springer.com/article/10.1007/s10462-017-9566-2>
- Kuhn, T. (2014). *The Structure of Scientific Revolutions* (S. Zibakalam, Trans.). Tehran:

- Samt. (Original work published 1962) [In Persian]
- Kumaravel, G., & Sankaranarayanan, S. (2021). PQPS: Prior-Art Query-Based Patent Summarizer Using RBM and Bi-LSTM. *Mobile Information Systems*, 2021(1), 2497770. URL: <https://doi.org/10.1155/2021/2497770>
- Laskar, M. T. R., Hoque, E., & Huang, J. (2020). Query focused abstractive summarization via incorporating query relevance and transfer learning with transformer models. In *Advances in Artificial Intelligence: 33rd Canadian Conference on Artificial Intelligence, Canadian AI 2020, Ottawa, ON, Canada, May 13–15, 2020, Proceedings*. 33 (pp. 342-348). Springer International Publishing. URL: https://link.springer.com/chapter/10.1007/978-3-030-47358-7_35
- Laskar, M. T. R., Rahman, M., Jahan, I., Hoque, E., & Huang, J. (2023a). CQSumDP: a ChatGPT-annotated resource for query-focused abstractive summarization based on debatepedia. arXiv preprint. URL: <https://arxiv.org/abs/2305.06147>
- Laskar, M. T. R., Rahman, M., Jahan, I., Hoque, E., & Huang, J. (2023b). Can large language models fix data annotation errors? an empirical study using debatepedia for query-focused text summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 10245-10255). URL: <https://aclanthology.org/2023.findings-emnlp.686.pdf>
- Laskar, M., Hoque, E., & Huang, J. (2021). Domain Adaptation with Pre-trained Transformers for Query Focused Abstractive Text Summarization. arXiv e-prints. URL: <https://arxiv.org/abs/2112.11670>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444. URL: <https://www.nature.com/articles/nature14539>
- Li, B., Yang, P., Zhao, H., Zhang, P., & Liu, Z. (2023). Hierarchical sliding inference generator for question-driven abstractive answer summarization. *ACM Transactions on Information Systems*, 41(1), 1-27. URL: <https://doi.org/10.1145/3511891>
- Liu, M., Ma, Z., Li, J., Wu, Y. C., & Wang, X. (2024). Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. *IEEE Access*. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10584489>
- Mehdad, Y., Carenini, G., & Ng, R. (2014). Abstractive summarization of spoken and written conversations based on phrasal queries. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 1220-1230). URL: <https://aclanthology.org/P14-1115>
- Mridha, M. F., Lima, A. A., Nur, K., Das, S. C., Hasan, M., & Kabir, M. M. (2021). A survey of automatic text summarization: Progress, process and challenges. *IEEE Access*, 9, 156043-156070. URL: <https://ieeexplore.ieee.org/abstract/document/9623462/>
- Nagalavi, D., & Hanumanthappa, M. (2019). The NLP Techniques for Automatic Multi-Article News Summarization Based on Abstract Meaning Representation. In *Emerging Trends in Expert Applications and Security*, 253-260. Springer, Singapore. URL: https://link.springer.com/chapter/10.1007/978-981-13-2285-3_31
- Nema, P., Khapra, M., Laha, A., & Ravindran, B. (2017). Diversity driven attention model

- for query-based abstractive summarization. arXiv preprint. URL: <https://aclanthology.org/P17-1098>
- Nimavat, K., & Joshiara, H. A. (2017). query-based summarization methods for conversational agents: an overview. *International Journal of Advanced Research in Computer Science*, 8(8), 448-453. URL: <https://www.proquest.com/docview/1953785769>
- Nishida, K., Saito, I., Nishida, K., Shinoda, K., Otsuka, A., Asano, H., & Tomita, J. (2019). Multi-style generative reading comprehension. arXiv preprint. URL: <https://arxiv.org/abs/1901.02262>
- Pasunuru, R., Celikyilmaz, A., Galley, M., Xiong, C., Zhang, Y., Bansal, M., & Gao, J. (2021). Data augmentation for abstractive query-focused multi-document summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 35 (15), 13666-13674. URL: <https://arxiv.org/abs/2103.01863>
- Polash, M. M. H. (2019). Query-focused abstractive summarization using sequence-to-sequence and transformer models [Doctoral dissertation, University of Lethbridge]. URL: <https://hdl.handle.net/10133/5665>
- Rahman, N., & Borah, B. (2020). Improvement of query-based text summarization using word sense disambiguation. *Complex & Intelligent Systems*, 6, 75-85. URL: <https://link.springer.com/article/10.1007/s40747-019-0115-2>
- Ritharson, P. I., Juliet, D. S., Anitha, J., & Pandian, S. I. A. (2023). Multi-Document Summarization Made Easy: An Abstractive Query-Focused System Using Web Scraping and Transformer Models. In *2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India*, 2023. URL: <https://ieeexplore.ieee.org/document/10205946>
- Roy, P., & Kundu, S. (2023). Review on Query-focused Multi-document Summarization (QMDS) with Comparative Analysis. *ACM Computing Surveys*, 56(1), 1-38. URL: <https://doi.org/10.1145/3597299>
- ShafieiBavani, E., Ebrahimi, M., Wong, R., & Chen, F. (2016). A query-based summarization service from multiple news sources. In *2016 IEEE International Conference on Services Computing (SCC)* (pp. 42-49). IEEE. URL: <https://ieeexplore.ieee.org/document/7557434>
- Strauss, A., & Corbin, J. (2011). *Basics of qualitative research :grounded theory procedures and techniques* (B. Mohammadi, Trans.). Tehran: Humanities and Cultural Studies Research Institute. (Original work published 1990) [In Persian]
- Su, D., Xu, Y., Yu, T., Siddique, F. B., Barezi, E. J., & Fung, P. (2020). CAiRE-COVID: A question answering and query-focused multi-document summarization system for COVID-19 scholarly information management. arXiv preprint. URL: <https://aclanthology.org/2020.nlpCOVID19-2.14/>
- Su, D., Yu, T., & Fung, P. (2021). Improve query focused abstractive summarization by incorporating answer relevance. arXiv preprint. URL: <https://arxiv.org/abs/2105.12969>
- Suleiman, D., & Awajan, A. (2020). Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges. *Mathematical problems in*

- engineering, 2020, 1-29. URL: <https://doi.org/10.1155/2020/9365340>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. (pp. 6000-6010). URL: <https://dl.acm.org/doi/10.5555/3295222.3295349>
- Vig, J., Fabbri, A. R., Kryściński, W., Wu, C. S., & Liu, W. (2021). Exploring neural models for query-focused summarization. arXiv preprint. URL: <https://arxiv.org/abs/2112.07637>
- Wang, X., Wang, J., Xu, B., Lin, H., Zhang, B., & Yang, Z. (2022). Exploiting Intersentence Information for Better Question-Driven Abstractive Summarization: Algorithm Development and Validation. *JMIR Medical Informatics*, 10(8), e38052. URL: <https://medinform.jmir.org/2022/8/e38052>
- Wibawa, A. P., & Kurniawan, F. (2024). A survey of text summarization: Techniques, evaluation and challenges. *Natural Language Processing Journal*, 7, 100070. URL: <https://doi.org/10.1016/j.nlp.2024.100070>
- Xu, Y., & Lapata, M. (2020). Abstractive query focused summarization with query-free resources. arXiv preprint. URL: <https://arxiv.org/pdf/2012.14774v1>
- Xu, Y., & Lapata, M. (2021a). Generating query focused summaries from query-free resources. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers), 6096–6109. URL: <https://aclanthology.org/2021.acl-long.475>
- Xu, Y., & Lapata, M. (2021b). Text summarization with latent queries. arXiv preprint. URL: <https://arxiv.org/abs/2106.00104>
- Yadav, D., Desai, J., & Yadav, A. K. (2022). Automatic Text Summarization Methods: A Comprehensive Review. arXiv preprint. URL: <https://arxiv.org/abs/2204.01849>
- Yu, H. (2022). Survey of Query-based Text Summarization. arXiv preprint. URL: <https://arxiv.org/abs/2211.11548>