E-ISSN: | P-ISSN:

Management: Modeling, Analysis and Application

www.mma.reapress.com

Manag. Model. Anal. Appl. Vol. 1, No. 3 (2024) 187–194.

Paper Type: Original Article

Improving the Text Summarization Process Using the Neutrosophic Hypergraph Model

Atiyeh Galin Moghadam[1,*], Seyed Ahmad Edalatpanah[2], Hadi Roshan[1]

[1] *Department of Computer Engineering, Ayandegan University, Tonekabon, Iran; moghadam.at2023@gmail.com; roshan.hadi@gmail.com.*

[2] *Department of Industrial Engineering, Ayandegan University, Tonekabon, Iran; saedalatpanah@gmail.com.*

Citation:

Galin Moghadam, A., Edalatpanah, A., & Roshan, H. (2024). Improving the text summarization process using the Neutrosophic hypergraph model. *Management: Modeling, analysis and applications, 1*(3), 187-194.

Received: 07/02/2024 Reviewed: 12/04/2024 Revised: 09/05/2024 Accepted: 27/06/2024

Abstract

Purpose: The purpose of this research is to present a new model based on neutrosophic hypergraphs for automatic text summarization that can model complex relationships between sentences well and produce an accurate and readable summary.

Methodology: In the proposed model, first the sentences of the text are displayed as nodes in a hypergraph and the connections between them are defined as edges. Then, using neutrosophic logic, the more important sentences are identified and selected for summary generation. This model has the ability to process ambiguous and uncertain data, which helps to better understand the semantic structure of the text.

Findings: Testing the model on three text databases of different lengths (India Text, Samsung Text, and BBC News) showed that this method, with average values of F1_Score=0.86, Precision=0.81, and Recall=0.90, performs better than other similar models and provides more accurate and readable summaries.

Originality/Value: The main innovation of the research is the use of neutrosophic hypergraphs to process complex relationships between sentences and manage uncertain data, which increases the accuracy and quality of summarization. This approach is an important step in improving automatic text summarization methods.

Keywords: Text summarization, Neutrosophic hypergraph model, Natural language processing, Neutrosophic logic, Deep learning.

Corresponding Author: moghadam.at2023@gmail.com

مدیریت: مدلسازی، تحلیل و کاربرد

دوره ۱، شماره (۳)، (۱۴۰۳)، ۱۹۴-۱۸۷

www.mma.reapress.com

نوع مقاله: پژوهشی

بهبود فرآیند خلاصه‌سازی متن با بهره‌گیری از مدل ابرگراف نوتروسوفیک

عطیه گلین مقدم[1]*، احمد عدالت پناه[2]، هادی روشن[1]

[1]گروه مهندسی کامپیوتر، دانشگاه آیندگان، تنکابن، ایران.

[2]گروه مهندسی صنایع، دانشگاه آیندگان، تنکابن، ایران.

چکیده

هدف: هدف این پژوهش ارایه مدلی جدید مبتنی بر ابرگراف[1]های نوتروسوفیک[2] برای خلاصه‌سازی خودکار متن[3] است که بتواند روابط پیچیده بین جملات را به‌خوبی مدلسازی کرده و خلاصه‌ای دقیق و خوانا تولید کند.

روش‌شناسی پژوهش: در مدل پیشنهادی، ابتدا جملات متن به‌صورت گره‌هایی در یک ابرگراف نمایش داده می‌شوند و ارتباطات بین آن‌ها به‌صورت یال‌ها تعریف می‌گردد. سپس با بهره‌گیری از منطق نوتروسوفیک، جملات مهم‌تر شناسایی و برای تولید خلاصه انتخاب می‌شوند. این مدل قابلیت پردازش داده‌های مبهم و نامطمئن را دارد که به درک بهتر ساختار معنایی متن کمک می‌کند.

یافته‌ها: آزمایش مدل روی سه پایگاه داده متنی با طول‌های مختلف (*India Text, Samsung Text, and BBC News*) نشان داد که این روش با میانگین مقادیر *F1_Score=0.86* ، *Precision=0.81* و *Recall=0.90*، عملکرد بهتری نسبت به سایر مدل‌های مشابه دارد و خلاصه‌های دقیق‌تر و خواناتری ارایه می‌دهد.

اصالت/ارزش افزوده علمی: نوآوری اصلی پژوهش در به‌کارگیری ابرگراف‌های نوتروسوفیک برای پردازش روابط پیچیده بین جملات و مدیریت داده‌های نامطمئن است که باعث افزایش دقت و کیفیت خلاصه‌سازی می‌شود. این رویکرد، گامی مهم در بهبود روش‌های خلاصه‌سازی خودکار متن محسوب می‌شود.

کلیدواژه‌ها: خلاصه‌سازی متن، مدل ابرگراف نوتروسوفیک، پردازش زبان طبیعی، منطق نوتروسوفیک، یادگیری عمیق.

۱- مقدمه

خلاصه‌سازی خودکار متن فرآیندی است که شامل استفاده از الگوریتم‌هایی برای شناسایی و استخراج مهم‌ترین اطلاعات از یک سند متنی و ارایه آن به شکل کوتاه و مختصر می‌باشد. روش‌های متفاوتی برای خلاصه‌سازی متن[4] تاکنون ارایه شده‌اند که عموما یا در کیفیت خلاصه‌سازی و یا در حجم خلاصه‌سازی با چالش جدی مواجه بوده‌اند. لذا ارایه یک مدل کارآمد در این زمینه که هر دو جنبه مطرح‌شده را در نظر بگیرند حائز اهمیت است. ابرگراف نوتروسوفیک یک ساختار ریاضی قدرتمند است که قابلیت نمایش اطلاعات نامطمئن و مبهم را دارد. این ساختار، با ترکیب مفاهیم گراف و منطق نوتروسوفیک، ابزاری کارآمد برای مدلسازی پیچیدگی‌های موجود در داده‌های زبانی فراهم می‌کند. خلاصه‌سازی خودکار متن به‌عنوان

[1] Hyper graph
[2] Neutrosophic
[3] Automatic text summarization
[4] Text Summarization (TS)

یکی از چالش‌های اصلی در حوزه پردازش زبان طبیعی[1] مطرح است. با توجه به حجم عظیم اطلاعات موجود در دنیای امروز، نیاز به ابزارهای خودکار برای خلاصه‌سازی متون بیش از پیش احساس می‌شود. توسعه مدل ابرگراف نوتروسوفیک برای *TS*، رویکردی نوین و امیدوارکننده است که می‌تواند به بهبود کیفیت خلاصه متن‌های تولیدشده کمک کند. در این رویکرد، متن به‌صورت یک ابرگراف نوتروسوفیک مدلسازی می‌شود. گره‌های این ابرگراف کلمات یا عبارات متن هستند و یال‌های آن، روابط معنایی بین گره‌ها را نشان می‌دهند. وزن یال‌ها نیز با استفاده از منطق نوتروسوفیک، درجه اهمیت و قطعیت روابط را بیان می‌کند [1]-[3].

روش‌های سنتی خلاصه‌سازی اغلب با مشکلاتی مانند عدم دقت در انتخاب جملات کلیدی، از دست دادن اطلاعات مهم و عدم توانایی در درک پیچیدگی‌های معنایی متون مواجه بودند. در این پژوهش، با هدف بهبود کیفیت خلاصه‌های تولیدشده، یک چارچوب نوین مبتنی بر ابرگراف‌های نوتروسوفیک ارایه می‌شود. این ویژگی باعث می‌شود که ابرگراف‌های نوتروسوفیک ابزاری مناسب برای مدلسازی پیچیدگی‌های موجود در متون طبیعی باشد [2]-[5].

هدف از این پژوهش توسعه یک مدل *TS* مبتنی بر ابرگراف‌های نوتروسوفیک، طراحی و پیاده‌سازی مدل ابرگراف نوتروسوفیک برای نمایش ساختار معنایی متن، ارزیابی عملکرد مدل پیشنهادی بر روی مجموعه پایگاه داده‌های استاندارد معرفی‌شده در بخش‌های بعدی و مقایسه آن با سایر روش‌های *TS* در این حوزه می‌باشد.

در ادامه، مقاله‌ی پیش رو شامل بخش‌های زیر می‌باشد؛ در ابتدا مروری بر ادبیات تحقیق مرتبط با *TS* و ابرگراف‌های نوتروسوفیک ارایه شده، سپس یک معماری مناسبی برای مدل پیشنهادی جهت *TS* مبتنی بر ابرگراف‌های نوتروسوفیک مطرح گردیده و در بخش نتایج شبیه‌سازی به ارزیابی مدل پیشنهادی روی مجموعه داده‌های معرفی‌شده پرداخته و پس از مقایسه آن با سایر روش‌های مشابه، در نهایت، به جمع‌بندی و نتیجه‌گیری کلی پژوهش و ارایه پیشنهادهایی برای محققان در این حوزه ارایه شده است.

۲- مروری بر ادبیات تحقیق

TS در اواخر دهه‌ی پنجاه میلادی موردتوجه قرار گرفت، یک مساله‌ی ناخوشایند در این حوزه زمان‌بر بودن آن برای انسان و عدم دقت کافی در فرآیند *TS* بود، به همین دلیل محققان علاقه‌مند در حوزه‌ی *NLP* به‌دنبال ارایه راهکاری با بهره‌گیری از روش‌های مبتنی بر یادگیری ماشین بودند[1]-[3]. *NLP* و *TS* و تکنیک‌های مرتبط با آن با رویکرد الگوریتم‌های یادگیری ماشین اولین‌بار توسط لون ارایه شد که شامل تشخیص جملات از متن و تشخیص کلمات از هر جمله می‌باشد [2]-[4] خلاصه‌کننده‌های متن عموما از دو روش خلاصه‌کننده‌های انتزاعی و خلاصه‌کننده‌های استخراجی استفاده می‌کنند. خلاصه‌کننده‌های انتزاعی شامل متونی است که در متن عینا وجود ندارند و بر اساس دانش‌های تحلیل متن و بازنویسی متن ایجاد می‌گردند، شبیه درکی از متن به شکل خلاصه که هدف اصلی مقاله پیش رو می‌باشد. در مقابل روش دوم که خلاصه‌سازی استخراجی متن[2] می‌باشد مستقیما از جملات موجود در متن اقدام به استخراج کلمات مرتبط با آن جمله می‌کند و تغییر مفهومی و معنایی روی متن انجام نمی‌دهد [6]-[8] هنگامی‌که جملات یا عبارات کلیدی از یک متن استخراج می‌گردند، بر اساس ارتباط آن‌ها با ایده‌ی اصلی متن رتبه‌بندی می‌شوند. یکی از روش‌های رایج رتبه‌بندی که مبتنی بر آمار فرکانس کلمات در معکوس فرکانس کل کلمات متن می‌باشد روش فراوانی اصطلاح - فراوانی سند معکوس[3] می‌باشد. البته روش‌های دیگری همانند *Bert* و *Word Embedding* نیز در این حوزه حائز اهمیت بوده که نگارنده مقاله از روش *TF-IDF* در مقایسه با دو روش دیگر بهره برده است.

ونگ و همکاران [9] یک روش جدید برای رتبه‌بندی جملات در خلاصه‌سازی متنی ارایه داده‌اند که از ابرگراف‌ها برای مدلسازی روابط گروهی پیچیده بین جملات استفاده می‌کند. روش‌های سنتی مانند *LexRank* و *HITS*، اسناد را به‌صورت یک گراف متنی مدلسازی می‌کنند که در آن گره‌ها جملات و یال‌ها روابط زوجی بین جملات را نشان می‌دهند. این مدلسازی قادر به نمایش روابط گروهی پیچیده بین چندین جمله نیست. استفاده از یک ابر گراف متنی باعث می‌شود که گره‌ها همچنان جملات را نشان دهند، اما ابر یال‌ها می‌توانند بیش از دو جمله را به هم متصل کنند. با استفاده

[1] Natrual Language Processing (NLP)
[2] Extractive Text Summarization (ETS)
[3] Term Frequency-Inverse Document Frequency (TF-IDF)

از ابرگراف، روابط گروهی و روابط زوجی در یک چارچوب واحد ادغام می‌شوند. سپس، یک الگوریتم رتبه‌بندی نیمه‌نظارت‌شده مبتنی بر ابرگراف برای خلاصه‌سازی استخراج‌کننده مبتنی بر پرسش توسعه داده می‌شود. در این الگوریتم، تاثیر پرسش از طریق ساختار ابرگراف متنی به جملات منتقل می‌شود. این روش روی مجموعه داده‌های *DUC* ارزیابی شده و بهبود عملکردی نسبت به چندین سیستم پایه نشان داده است.

ون لیرده و همکاران [10] یک روش جدید خلاصه‌سازی مبتنی بر ابرگراف پیشنهاد داده‌اند که بر مشکل روش‌های قبلی در انتخاب مجموعه‌ای هماهنگ از جملات مرتبط با هم و اجتناب از خلاصه‌های تکراری که فاقد موضوعات مهم متن هستند، غلبه می‌کند. روش‌های قبلی مبتنی بر گراف و ابرگراف، جملات یک مجموعه متن را به‌صورت گره‌ها در نظر می‌گرفتند و لبه‌ها در گراف و ابر یال‌ها در ابرگراف نشان‌دهنده شباهت‌های واژگانی بین جملات بودند. سپس هر جمله به‌صورت جداگانه با استفاده از الگوریتم‌های رایج رتبه‌بندی گره‌ها، امتیاز می‌گرفت و خلاصه با انتخاب جملات با امتیاز بالا ساخته می‌شد. برای رفع این مشکل، ون لیرده و همکاران [10] در روش پیشنهادی خود، هر گره را یک جمله و هر ابر یال را یک موضوع در نظر می‌گیرند. به این معنا که هر ابر یال گروهی از جملات مرتبط با یک موضوع خاص را نشان می‌دهد. اهمیت موضوعات در متن و ارتباط آن‌ها با یک پرسش از پیش تعریف‌شده توسط کاربر، وزن‌دهی می‌شود. همچنین در این مقاله نشان داده شده است که شناسایی مجموعه‌ای از جملات که موضوعات مرتبط متن را پوشش می‌دهند، معادل یافتن یک «گذر ابر گراف»[1] در ابرگراف موضوع محور است. دو رویکرد جدید برای مفهوم «گذر ابرگراف» برای خلاصه‌سازی پیشنهاد شده و الگوریتم‌های زمان چندجمله‌ای بر اساس نظریه توابع زیرمجموعه‌های خوش‌رفتار[2] برای حل مسایل بهینه‌سازی گسسته مرتبط با این روش ارایه شده است.

تحلیل مقایسه‌ای دقیق با مدل‌های مرتبط روی مجموعه داده‌های ارزیابی *DUC*، اثربخشی روش پیشنهادی را نشان می‌دهد که حداقل ۶% در معیار ارزیابی *ROUGE-SU4* نسبت به روش‌های موجود مبتنی بر گراف یا ابرگراف عملکرد بهتری دارد.

ورما و همکاران [2] یک طرح متن‌کاوی جدید برای تحلیل داده‌های بزرگ و *NLP* ارایه کرده‌اند که قادر به تفسیر و تحلیل داده‌های متنی حجیم است. آن‌ها روی *TS* به‌عنوان یک ستون اساسی در موتورهای توصیه تمرکز کرده‌اند و یک تغییر پیش‌بینی شده به‌سمت یک طرح خلاصه‌سازی استخراجی مبتنی بر گراف *ETS* را مورد بررسی قرار داده‌اند. روش‌های *ETS*، اگرچه ساده‌تر و کم‌هزینه‌تر هستند، اما در ارزیابی نظرات و بازخورد در مورد محصولات یا خدمات کلیدی اهمیت دارند. با این حال، روش‌های فعلی نتوانسته‌اند نگرانی‌ها در مورد پیچیدگی، انعطاف‌پذیری و تقاضاهای محاسباتی را به‌طور کامل برطرف کنند. بنابراین، در این مقاله، ورما و همکاران [2] روش خلاصه‌سازی مبتنی بر گراف برای رویداد و متن[3] را با استفاده از یک مدل مبتنی بر گراف برای ایجاد ارتباط بین کلمات و جملات از طریق روش‌های آماری پیشنهاد کرده‌اند. این ساختار شامل یک مرحله پس‌پردازش است که شامل خوشه‌بندی جملات مبتنی بر گراف می‌شود. با استفاده از چارچوب *Apache Spark*، این طرح برای اجرای موازی طراحی شده است و آن را برای کاربردهای دنیای واقعی قابل انطباق می‌کند. برای ارزیابی، آن‌ها ۵۰۰ سند را از مجموعه داده‌های *WikiHow* و Opinosis انتخاب کردند، آن‌ها را به پنج کلاس طبقه‌بندی کردند و پارامترهای ارزیابی خلاصه‌سازی مبتنی بر بازخوانی *ROUGE* را برای مقایسه با معیارهای *ROUGE-1*، *ROUGE-2* و *ROUGE-L* اعمال کردند. نتایج شامل امتیازهای بازخوانی ۰/۳۹۴۲، ۰/۰۹۵۲ و ۰/۳۴۳۶ برای *ROUGE-1*، *ROUGE-2* و *ROUGE-L* به‌ترتیب (هنگام استفاده از روش خوشه‌بندی شده) است. از طریق مقایسه با مدل‌های موجود مانند *BERTEXT* با *3-Gram* و *4-Gram* و *MATCHSUM*، مدل مطرح‌شده بهبودهای قابل توجهی را نشان داده است که کاربردپذیری و اثربخشی آن را در سناریوهای دنیای واقعی تایید می‌کند.

دایانا و همکاران [1] به بررسی چالش تشخیص و حذف سخن‌چینی نفرت‌انگیز در پلتفرم‌های رسانه‌های اجتماعی پرداخته‌اند. آن‌ها اشاره کرده‌اند که با افزایش فعالیت این پلتفرم‌ها، کاربران فرصت بیشتری برای ارسال محتوای آنلاین دارند که برخی از آن‌ها ممکن است حاوی سخن‌چینی نفرت‌انگیز باشند. تشخیص و حذف این نوع محتوا از پلتفرم‌هایی مانند توییتر و فیس‌بوک دشوار است. روش‌های بسیاری در طول سال‌ها برای مقابله با این مشکل توسعه یافته‌اند که بیشتر آن‌ها روش‌های یادگیری ماشین زمان‌بر هستند. دایانا و همکاران [1] در این مقاله یک روش جدید برای تشخیص و حذف سخن‌چینی نفرت‌انگیز پیشنهاد کرده‌اند که عدم قطعیت را در سطح کلمه و جمله با استفاده از منطق فازی اعمال‌شده بر ابرگراف‌های نوتروسوفیک در نظر می‌گیرد. در این روش، اسناد متنی به ابرگراف‌های نوتروسوفیک مدلسازی می‌شوند که در آن هر گره و ابر یال دارای

---

[1] Hypergraph transversal

[2] submodular functions

[3] Graph-based Event and Text Summarization (GETS)

سه تابع عضویت مرتبط به نام‌های عضویت (درستی)، عدم عضویت (نادرستی) و عدم قطعیت است. سپس عملگرهای مورفولوژیکی مانند توسعه، فرسایش و ... بر روی این ابرگراف‌ها اعمال می‌شوند. با استفاده از این عملیات، عملگرهای دیگری مانند نازک‌سازی، ضخیم‌سازی، ضربه یا از دست دادن و اسکلت‌بندی نیز اعمال می‌شوند. در نهایت، سخن‌چینی نفرت‌انگیز شناسایی و حذف می‌شود. این روش جدید با توییت‌های توییتر آزمایش شده است و نتایج امیدوارکننده‌ای با دقت ۸۸% نشان داده است.

۳- مدل پیشنهادی پژوهش

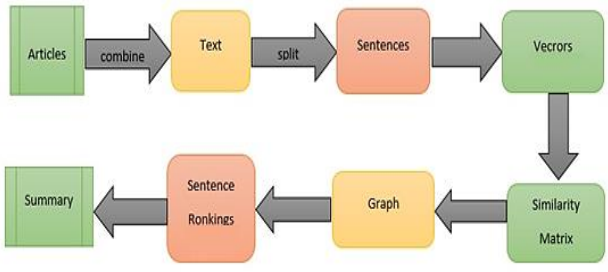
TS یکی از چالش‌های مهم در *NLP* است. هدف از خلاصه‌سازی، تولید یک نسخه کوتاه‌تر از یک متن است که حاوی اطلاعات اصلی و کلیدی آن باشد. در سال‌های اخیر، با پیشرفت‌های چشمگیر در حوزه یادگیری ماشین، روش‌های مختلفی برای *TS* ارایه شده است. یکی از این روش‌ها که به دلیل توانایی در مدلسازی روابط پیچیده و ابهام‌زدایی در متن موردتوجه قرار گرفته است، استفاده از مدل‌های ابرگراف نوتروسوفیک است. ابرگراف‌ها ساختارهای داده‌ای قدرتمندی هستند که می‌توانند روابط پیچیده بین موجودیت‌ها را مدلسازی کنند. در زمینه *TS*، هر جمله می‌تواند به‌عنوان یک گره در ابرگراف در نظر گرفته شود و روابط بین جملات (مانند روابط معنایی، زمانی و علیتی) به‌عنوان یال‌ها نمایش داده شوند. منطق نوتروسوفیک نیز به‌عنوان یک ابزار قدرتمند برای مدلسازی اطلاعات ناقص و مبهم در سیستم‌های اطلاعاتی مطرح شده است. در این منطق، هر گزاره می‌تواند سه مقدار عضویت (درستی)، عدم عضویت (نادرستی) و عدم قطعیت داشته باشد. این ویژگی، منطق نوتروسوفیک را برای مدلسازی ابهاماتی که در متن وجود دارد، بسیار مناسب می‌سازد. ترکیب ابرگراف‌ها و منطق نوتروسوفیک در *TS*، به ما اجازه می‌دهد تا روابط پیچیده بین جملات را با در نظر گرفتن ابهامات و عدم قطعیت‌های موجود در متن مدلسازی کنیم. در این روش، به هر جمله یک بردار نوتروسوفیک اختصاص داده می‌شود که نشان‌دهنده میزان عضویت، عدم عضویت و عدم قطعیت آن جمله در خلاصه نهایی است. سپس، با استفاده از الگوریتم‌های مناسب، جملاتی که بیشترین اهمیت را در خلاصه دارند، انتخاب می‌شوند [1]-[4].

مزایای استفاده از مدل‌های ابرگراف نوتروسوفیک در *TS* شامل موارد زیر می‌باشد:

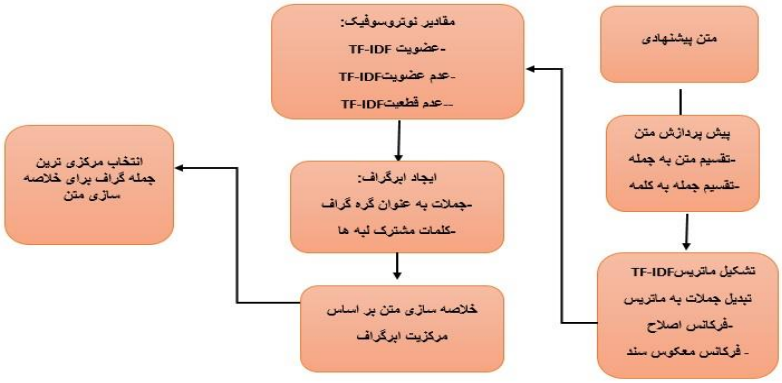
۱. مدلسازی روابط پیچیده: توانایی در مدلسازی روابط پیچیده بین جملات، مانند روابط معنایی و نحوی.
۲. مدیریت ابهام: توانایی در مدیریت ابهام و عدم قطعیت موجود در متن.
۳. انعطاف‌پذیری بالا: قابلیت تطبیق با انواع مختلف متن و سبک‌های نوشتاری.
۴. تولید خلاصه‌های با کیفیت بالا: تولید خلاصه‌هایی که هم محتوای اصلی متن را حفظ می‌کنند و هم خوانا هستند.

۳-۱- مراحل انجام کار در مدل پیشنهادی

تشکیل مدل ابرگراف در پژوهش پیش رو که بر پایه منطق نوتروسوفیک است، یک مرحله کلیدی و تعیین‌کننده در فرآیند *TS* به شمار می‌رود. ابرگراف‌ها، ساختارهای داده‌ای قدرتمندی هستند که قادر به نمایش روابط پیچیده بین موجودیت‌ها هستند. در زمینه *TS*، هر جمله یک گره در ابرگراف و روابط معنایی بین جملات نیز به‌صورت یال نمایش داده می‌شوند. ابتدا یک شماتیک کلی از فرآیند TS [1]-[4] در **شکل ۱** نمایش داده شده و در **شکل ۲** مراحل کامل و دقیق اجرای پژوهش پیشنهادی ارایه گردیده است.



شکل ۱- شماتیک کلی از فرآیند خلاصه‌سازی متن.
Figure 1- General schematic of text summarization.



شکل ۲- مراحل کامل اجرای مدل پیشنهادی برای خلاصه‌سازی متن.

Figure 2 - Complete steps of implementing the proposed model for text summarization.

در مدل پیشنهادی با ارایه‌ی یک راهکار جدید مبتنی بر منطق نوتروسوفیک با تبدیل بردارهای عددی از مقادیر توکن‌سازی به‌دست آمده با روش *TF-IDF* از روی ابرگراف‌ها اقدام به *TS* نموده‌ایم. مجموعه پایگاه داده‌های مورد استفاده برای تحلیل مدل پیشنهادی شامل متون دریافتی *BBC News* 2019 *Samsung Text* , *Summary 2021,India Text 2007* که از مجموعه پایگاه داده استاندارد *Kaggle* در سه قالب متن کوتاه، متن متوسط و متن طولانی شامل ۲۱۵۳۵ حرف برای متن کوتاه، ۶۴۵۵۶ حرف برای متن متوسط و ۱۳۲۴۳۵ حرف برای متن طولانی مورد استفاده قرار گرفته است. ابتدا متن را از منبع ورودی مشخصه دریافت می‌کنیم، سپس پاراگراف‌های موجود در متن را جدا کرده و در هر پاراگراف کلمات کلیدی را استخراج می‌کنیم. با استفاده از الگوریتم *TF_IDF* بردارسازی میان کلمات کلیدی را انجام می‌دهیم و سپس با توجه به تعداد تکرار هر کلمه و جایگاه آن در بردار به‌دست آمده اقدام به ساخت یک مدل ابرگراف می‌کنیم که ارتباط میان هر دو گره در گراف که همان یال گراف می‌باشد با مقادیر وزنی به‌دست آمده از مرحله‌ی قبل مقداردهی می‌شود. وجود مقدار بزرگتر بر روی هر یال در گراف بیان وجود ارتباط معنایی بیشتر میان گره‌های مرتبط با آن کلمات کلیدی می‌باشد، به این معنی که میان آن دو کلمه کلیدی ارتباط معنایی بیشتری وجود دارد که این مساله قطعا به *TS* با کیفیت بهتر منجر خواهد شد. استفده از منطق نوتروسوفیک (منطق سه مقداری با مقادیر عدم قطعیت، عضویت و عدم عضویت) برای تعیین میزان درجه عضویت یا تعلق ارتباطی میان دو گره یا همان عدد روی یال در مقایسه با ارتباط میان این گره با سایر گره‌های موجود در گرف می‌تواند منجر به تولید یک ابرگراف قطعیت گردیده که اعداد روی یال‌های آن ابرگراف قطعیت مقادیر بسیار دقیقی برای حفظ متن بر اساس آن کلمات مورد نظر می‌باشد. در حقیقت ارایه چنین مدلی علاوه‌بر بالا بردن میزان کیفیت *TS* در حجم خلاصه‌سازی و حذف کلمات با میزان عددی کمتر در درجه عضویت نوتروسوفیک می‌تواند مفید واقع گردد.

۲-۳- مثالی جهت درک مدل

My name is Atieh. My family is Galian. My name and my family are not similar together. But my name is so beautiful more than my family.

Vector TF-IDF: [my=5 name=3 atieh=1 family=4 galian=1 similar=1 beautiful=1]

پس از ایجاد مدل ابرگراف روی متن و سپس اجرای منطق نوتروسوفیک بر کلمات روی گره‌های گراف و وزنه‌ای میان گره‌ها که همان یال‌ها می‌باشد اعداد **جدول ۱** به‌دست خواهند آمد:

جدول ۱ - نمایش مقادیر نوتروسوفیک میان متن ارایه‌شده در مثال فوق.

Table 1– Display of neutrosophic values within the text presented in the example above.

Words	T	I	F
My	0.8	0.2	0.1
Name	0.9	0.1	0.05
Atieh	0.9	0.1	0.05
Family	0.7	0.15	0.1
Galian	0.8	0.1	0.05
Similar	0.7	0.3	0.2
Beautiful	0.8	0.1	0.05

هر یک از مقادیر نوتروسوفیک بیانگر یکی از معانی زیر می‌باشد:

T عضویت: درجه‌ای که رابطه میان کلمات را بیان می‌کند. مثلا ۰/۹ (زیاد)

I عدم قطعیت: درجه‌ای که در مورد این رابطه نامطمئن هستیم. مثلا ۰/۰۵ (کم)

F عدم عضویت: درجه‌ای که این رابطه وجود ندارد. مثلا ۰/۰۵ (کم)

برای نمایش میزان دقت مدل پیشنهادی نمونه‌ای از یک متن استاندارد خلاصه‌شده با مدل پیشنهادی نمایش داده شده است :

Sample Text:

"Neutrosophic logic is a general framework for unifying many existing logics and handles indeterminacy explicitly. It has applications in areas such as decision making, image processing, and data mining. Text summarization is the process of creating a concise version of a document by selecting important sentences or phrases. Hypergraphs can model complex relationships between sentences in a text document by representing each sentence as a node and using hyperedges to represent the relationships. This approach captures more intricate relationships than traditional graphs"

Summarized Text:

"Neutrosophic logic handles uncertainty and has applications in various fields. Hypergraphs enhance text summarization by modeling complex sentence relationships."

۴- نتایج شبیه‌سازی مدل پیشنهادی

مدل پیشنهادی مطرح‌شده توسط پژوهنده بر روی سه مجموعه پایگاه داده معرفی‌شده در بخش قبل اجرا گردیده و نتیجه‌ی اجرای مدل با سه روش فازی و فازی شهودی مقایسه شده که در بخش نتایج ارایه شده است. نتایج به‌دست آمده نشان می‌دهد که *TS* بر اساس توسعه‌ی ابرگراف‌ها مبتنی بر منطق نوتروسوفیک در مقایسه با منطق فازی و منطق فازی شهودی هم در کیفیت *TS* و هم در میزان یا حجم خلاصه‌سازی بهتر عمل نموده است. برای مقایسه‌ی مدل مطرح‌شده توسط پژوهشگر با دو روش دیگر از پارامترهای ارزیابی *Precision,Recall,F1-Score* استفاده نموده است. نتایج ارزیابی بر روی سه متن معرفی‌شده (متن کوتاه، متن متوسط و متن طولانی) با روش پژوهنده و در مقایسه با دو روش دیگر و با بررسی پارامترهای ارزیابی هم به‌صورت جداگانه و هم در قالب یک جدول مقایسه‌ای ارایه شده است.

۴-۱- معرفی متغیرهای مورد نیاز در ارزیابی روش پیشنهادی

در ارزیابی مدل بر اساس *Recall,Precision,F1-Score* متغیرهای زیر استفاده شده‌اند:

متغیر*TP*: کلماتی که هم در خلاصه تولیدشده و هم در خلاصه مرجع وجود دارند.

متغیر*FP*: کلماتی که در خلاصه تولیدشده وجود دارند اما در خلاصه مرجع وجود ندارند.

متغیر*TN*: کلماتی که در خلاصه تولیدشده وجود ندارند اما در خلاصه مرجع وجود دارند.

متغیر*FN*: کلماتی که هم در خلاصه تولیدشده وجود ندارند و هم در خلاصه مرجع وجود ندارند.

$$Recall = \frac{TP}{TP+FN}. \quad (۱)$$

$$Recall = \frac{TP}{TP + FP}. \quad (۲)$$

$$F1-SCORE = 2 * \frac{Precision * Recall}{Precision + Recall}. \quad (۳)$$

۴-۲- نتایج اجرای مدل روی مجموعه‌ی پایگاه داده‌ی BBC News Summary

این مجموعه داده برای *TS* استخراجی دارای ۴۰۷ مقاله خبری سیاسی *BBC* از سال ۲۰۰۴ تا ۲۰۰۵ در پوشه مقاله اخبار است. برای هر مقاله پنج خلاصه در پوشه ی خلاصه ارایه شده است. اولین ستون مربوط به متن مقالات بوده است. تعداد کاراکترهای بررسی‌شده در این متن شامل ۱۳۲۴۳۵ کاراکتر می‌باشد. نتایج خلاصه‌سازی به‌صورت زیر بوده است:

جدول ۲- نتایج مدل پیشنهادی روی پایگاه داده اول.
Table 2– Results of the proposed model on the first database.

Recall	Precision	F1-Score
0.80	0.94	0.86

۴-۳- نتایج اجرای مدل بر روی مجموعه ی پایگاه داده Samsun Dataset Text Summarization

این مجموعه پایگاه داده شامل تعداد زیادی دیالوگ یا متن رد و بدل شده میان تعدادی کاربران مختلف می‌باشد. در این پایگاه داده ستون اول مربوط به *ID* کاربران و ستون دوم مربوط به موضوع دیالوگ انجام‌شده بوده و ویژگی سوم مربوط به متن‌های تایپ‌شده میان کاربران بوده است. از این مجموعه داده متنی شامل ۶۴۵۵۶ کاراکتر استخراج شده است. نتایج خلاصه‌سازی به‌صورت زیر بوده است:

جدول۳ - نتایج مدل پیشنهادی بر روی پایگاه داده دوم.
Table 3– Results of the proposed model on the second database.

Recall	Precision	F1-Score
0.82	0.86	0.84

۴-۴- نتایج اجرای مدل بر روی مجموعه ی پایگاه داده India Text for Summarization

این مجموعه داده شامل یک متن استاندارد با تعداد ۲۱۵۳۵ کاراکتر بوده که برای *TS* استفاده می‌شود. این مجموعه نمونه داده در مقالات زیادی برای *TS* استفاده شده است. نتایج خلاصه‌سازی به‌صورت زیر بوده است.

جدول ۴- نتایج مدل پیشنهادی روی پایگاه داده سوم.
Table 4– Results of the proposed model on the third database.

Recall	Precision	F1-Score
0.95	0.76	0.85

در **جدول ۵** نتایج مقایسه‌ای مدل پیشنهادی با دو روش دیگر ارایه شده است :

جدول ۵- مقایسه نتایج مدل پیشنهادی با دو روش مشابه دیگر روی سه پایگاه داده مطرح شده.
Table 5 - Comparison of the results of the proposed model with two other similar methods on the three proposed databases.

	دیتاست ۱	دیتاست ۲	دیتاست ۳
مدل ابرگراف مبتنی بر منطق فازی	0.86	0.63	0.67 Recall
	0.69	0.69	0.64 Precision
	0.77	0.66	0.66 F1-score
مدل ابرگراف مبتنی بر منطق فازی شهودی	0.93	0.90	0.93 Recall
	0.73	0.82	0.79 Precision
	0.82	0.85	0.86 F1-score
مدل پیشنهادی ابرگراف مبتنی بر منطق نوتروسوفیک	0.95	0.82	0.94 Recall
	0.76	0.86	0.80 Precision
	0.85	0.84	0.86 F1-score

همان‌گونه که در **جدول ۵** مشخص است روش پیشنهادی در متن‌های طولانی خیلی بهتر از دو روش دیگر عمل نموده است. اما در متن متوسط روش فازی شهودی تاحدودی نسبت به روش پیشنهادی عملکرد بهتری داشته است. در متن کوتاه نیز مجددا روش پیشنهادی تاحدودی نسبت به دو روش دیگر عملکرد بهتری داشته است. به‌طور کل نتایج نشان می‌دهد که مدل توسعه ابرگراف مبتنی بر منطق نوتروسوفیک در مقایسه با مدل مبتنی بر روش فازی و یا فازی شهودی علمکرد دقیق‌تر و مناسب‌تری از خود داشته است.

۵- نتیجه‌گیری

ارزیابی خلاصه‌سازی مبتنی بر ابرگراف، یک موضوع پیچیده و چندجانبه است. انتخاب معیارهای مناسب و ترکیب آن‌ها با ارزیابی انسانی، می‌تواند به ما کمک کند تا عملکرد مدل‌های خلاصه‌سازی را به‌طور دقیق ارزیابی کرده و آن‌ها را بهبود بخشیم.

با توجه به نتایج به‌دست آمده از شبیه‌سازی‌های انجام‌شده با سه روش توسعه ابرگراف مبتنی بر منطق فازی، توسعه ابرگراف مبتنی بر منطق فازی شهودی و توسعه ابرگراف مبتنی بر منطق نوتروسوفیک (مدل پیشنهادی) گویای این بوده که در متون خلاصه‌سازی‌شده از پایگاه داده‌های استاندارد معرفی‌شده روش پیشنهادی توانسته نتایج قابل‌قبول‌تری ارایه نماید. مدل‌های ابرگراف نوتروسوفیک، با قابلیت مدیریت ابهام و عدم قطعیت در داده‌ها، ابزار قدرتمندی برای *TS* به شمار می‌آیند. با این حال، همچون هر مدل دیگری، با محدودیت‌هایی نیز همراه هستند که بهبود آن‌ها می‌تواند به ارتقای کارایی و دقت مدل بیانجامد.

پیشنهادات

استفاده از الگوریتم‌های موازی و توزیع‌شده برای تسریع محاسبات در گراف‌های بزرگ.

استفاده از تکنیک‌های افزایش داده برای تولید داده‌های مصنوعی و افزایش حجم داده‌های آموزشی.

استفاده از روش‌های بهینه‌سازی خودکار مانند الگوریتم‌های تکاملی و یا روش‌های مبتنی بر یادگیری ماشین برای تعیین بهترین پارامترها.

با توجه به پتانسیل بالای مدل‌های ابرگراف نوتروسوفیک در حوزه *TS*، تحقیقات بیشتر در این زمینه می‌تواند به پیشرفت قابل توجهی در این حوزه منجر شود.

منابع

[1] Dhanya, P. M., & Ramkumar, P. B. (2023). Text analysis using morphological operations on a neutrosophic text hypergraph. *Neutrosophic sets and systems*, *61*, 337–364. https://digitalrepository.unm.edu/nss_journal/vol61/iss1/19/

[2] Verma, J. P., Bhargav, S., Bhavsar, M., Bhattacharya, P., Bostani, A., Chowdhury, S., … ., & Mehbodniya, A. (2023). Graph-based extractive text summarization sentence scoring scheme for big data applications. *Information*, *14*(9). https://doi.org/10.3390/info14090472

[3] Akram, M., Shahzadi, S., Rasool, A., & Sarwar, M. (2022). Decision-making methods based on fuzzy soft competition hypergraphs. *Complex & intelligent systems*, *8*(3), 2325–2348. https://doi.org/10.1007/s40747-022-00646-4

[4] Peng, X., & Dai, J. (2020). A bibliometric analysis of neutrosophic set: two decades review from 1998 to 2017. *Artificial intelligence review*, *53*(1), 199–255. https://doi.org/10.1007/s10462-018-9652-0

[5] Smarandache, F., & Broumi, S. (2019). *Neutrosophic graph theory and algorithms*. IGI Global. https://B2n.ir/zr9533

[6] Quek, S. G., Selvachandran, G., Ajay, D., Chellamani, P., Taniar, D., Fujita, H., … ., & Giang, N. L. (2022). New concepts of pentapartitioned neutrosophic graphs and applications for determining safest paths and towns in response to COVID-19. *Computational and applied mathematics*, *41*(4), 151. https://doi.org/10.1007/s40314-022-01823-4

[7] Abonyi, J., & Czvetkó, T. (2022). Hypergraph and network flow-based quality function deployment. *Heliyon*, *8*(12), e12263. https://www.cell.com/heliyon/fulltext/S2405-8440(22)03551-4

[8] Ihsan, M., Saeed, M., Rahman, A. U., & Smarandache, F. (2022). Multi-attribute decision support model based on bijective hypersoft expert set. *Punjab university journal of mathematics*, *54*(1), 55-73. https://doi.org/10.52280/pujm.2021.540105

[9] Wang, W., Li, S., Li, J., Li, W., & Wei, F. (2013). Exploring hypergraph-based semi-supervised ranking for query-oriented summarization. *Information sciences*, *237*, 271–286. https://doi.org/10.1016/j.ins.2013.03.012

[10] Van Lierde, H., & Chow, T. W. S. (2019). Query-oriented text summarization based on hypergraph transversals. *Information processing & management*, *56*(4), 1317–1338. https://doi.org/10.1016/j.ipm.2019.03.003