Enhancing Clinical Trial Summarization: Leveraging Large Language Models and Knowledge Graphs for Entity Preservation

Pouyan Nahed, Mina Esmail Zadeh Nojoo Kambar, and Kazem Taghva

Abstract ClinicalTrials.gov is an accessible online medical resource for researchers, healthcare professionals, and policy designers seeking detailed information on clinical trials. Summarizing these long clinical records can significantly reduce the time needed for the database users as the process transforms comprehensive information into concise synopses, preserving the essential meaning and facilitating understanding. In this paper, we employ the Bidirectional and Auto-Regressive Transformers model to generate the trials' brief summaries. Our contributions provide new preprocessing techniques for model training, which leads to a robust summarization model. The fine-tuned model significantly enhanced ROUGE-1, ROUGE-2, and ROUGE-L F1-scores by 14%, 23%, and 20%, respectively, compared to previous studies. Additionally, we present an innovative knowledge graph based on entity classes to assess the generated summaries. This graph not only quantifies the essential entities transformed from the original text to the summaries but also provides insights into their specific order and arrangement in sentences.

Keywords Large language models · Clinical data · Summarization · Named entity preservation · Knowledge graph

1 Introduction

ClinicalTrials.gov is a database of medical documents that offers comprehensive and publicly accessible records of registered clinical trials worldwide. This extensive repository contains many fields including detailed trial descriptions, study objec-

P. Nahed (✉) · M. E. Z. N. Kambar · K. Taghva
Computer Science Department, University of Nevada Las Vegas, Las Vegas, USA
e-mail: Nahed@unlv.nevada.edu

M. E. Z. N. Kambar
e-mail: Esmailza@unlv.nevada.edu

K. Taghva
e-mail: Kazem.Taghva@unlv.edu

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1003,
https://doi.org/10.1007/978-981-97-3302-6_26

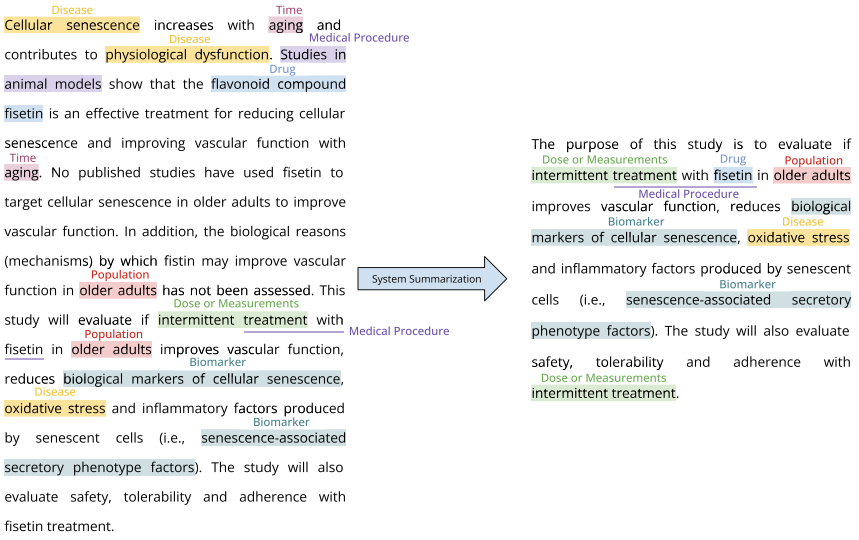

Fig. 1 An example of summary generated by our system. Entities within the text are annotated to demonstrate our model is able to summarize the descriptions to a shorter text while preserving all the main entities of the text. Some entities are shown by underscore due to overlap with other entities

tives, eligibility criteria, locations, intervention details, and outcomes. The database is a resource for researchers, healthcare professionals, and policy designers, which provides a wealth of information necessary to understand the breadth and depth of failed, ongoing, and completed clinical research. However, the database extensive and intricate nature poses a challenge regarding efficient record retrieval and comprehension, underscoring the need for an effective data summarization [1] (Fig. 1).

This work is driven by three primary goals. First, observing that many trials within the database lack concise and informative summaries which creates a gap in accessibility and understanding. Second, making trial information easier to understand especially for complicated areas like different treatment methods can make the database more helpful for different user groups. Third, summarizing this data effectively will uncover new opportunities for advanced data analysis and text mining which can lead to novel insights and contribution to the broader field of medical research and practice. In addition to these goals, we want to include all the principal entities from the original text in the summary to maximize the knowledge transfer. To achieve our objectives, we conducted a graph-based entity analysis to address our model performance on knowledge transfer and text summarization. We show that our model is cable of preserving most of the principal entities from the description in the generated summary. Our study employs the Bidirectional and Auto-Regressive Transformers (BART) model, a sophisticated Natural Language Processing (NLP) tool [2]. Our contributions are as follows:

1. We undertake meticulous cleaning and database preparation, focusing on diverse fields to train a robust and effective summarization model. This alignment is particularly challenging given the varied nature of data fields such as study design, participant details, and endpoints.
2. We fine-tune an abstractive text summarization model, resulting in improved quality of summaries as evidenced by enhanced ROUGE-1, ROUGE-2, and ROUGE-L F1-scores, which increased by 14%, 23%, and 20%, respectively, compared to the results reported by [3].
3. We introduce a novel entity class-based heterogeneous knowledge graph that evaluates the generated summaries. This graph incorporates two types of nodes: entity types and sentence numbers. Nodes from these classes are connected if the corresponding entity class is present in the associated sentence. For each trial record, we create two graphs for both the original and BART-generated summaries. The proximity of these graphs calculated by Jaccard similarity indicates the preservation of entity classes and their sentence-wise direction in the summaries.

Our model generates concise clinical trials summaries, offering compact abstracts that encapsulate essential study details. Despite their brevity, these summaries are information-rich, encompassing key elements such as study objectives, methodologies, interventions, and outcomes. Our innovative sentence-wise summary evaluation graphs provide valuable insights into the preserving of principal entities and the clinical trial summary structure.

2 Background

Over the past few years, technological advancements have resulted in a surge of textual data in the biomedical field. Automatic text summarization systems are pivotal in this context, as they play a significant role in streamlining physicians' time and pinpointing relevant information [4]. Text summarization aims to produce a more concise passage from a document, ensuring grammatical and logical coherence while retaining essential information. Most of these studies in this domain falls into *abstractive* or *extractive* summaries [5].

The extractive approach initially identifies noteworthy sentences from the source document and subsequently organizes them to create a summary without altering the original text. Studies [3, 6, 7] present state-of-the-art extractive methods on biomedical datasets. Research conducted by [6] introduced a method using Ontology and Graph-Based techniques, surpassing baseline approaches with a ROUGE-L F1-score of 0.29 on the PubMed Central dataset. In [7], a model named BioBERTSum is introduced. This model employed a domain-aware pre-trained language model as its encoder, subsequently fine-tuning it for the specific biomedical extractive summarization task. The approach demonstrated superior performance compared to previous Bert-based methods on the PubMed dataset, achieving a ROUGE-L score of 0.37. The potential of text summarization is important in the context of clinical

trials, offering an efficient means to comprehend subjects and investigate interventions concisely. The clinicaltrials.gov database, a comprehensive and openly accessible resource, features two primary data fields: detailed description and brief summary. These fields outline a trial's main goals and methods, providing an excellent information at a glance. The detailed description field typically tends to be longer than the brief summary.

In a study by Gulden [3], various text summarization algorithms, including LexRank and SumBasic were employed to condense detailed descriptions into brief summaries. Standard ROUGE metrics were then calculated, achieving a F1-score of 0.35 for ROUGE-1.

Abstractive summarization goes beyond extracting sentences by aiming to understand the text's meaning. Unlike extractive summarization, it generates concise summaries using its own language and style, often introducing new elements for a more human-like text. Reference [8] provides a comprehensive review of pre-trained language models in biomedical text summarization. It emphasizes that pre-trained language models equipped with decoders, such as the GPT series, T5, and BART are particularly well-suited for abstractive synopses [9].

Research conducted in [10] explores the BART model's application in summarizing multiple medical documents. The findings reveal that this model can produce cohesive synopses that align with the reference summaries in evidence direction approximately 50% of the time. Reference [11] utilizes the BART model to generate the biomedical evidence summaries of multiple clinical trials. Their data is 4528 systematic reviews composed by members of the Cochrane Collaboration (https://www.cochrane.org/). They suggest new ways to improve summaries using unique models for specific fields. For example, they highlight important parts of the information and focus more on reports from large and high-quality trials. These methods make the summaries more accurate. Lastly, they suggest a new approach to check if the summaries' information is correct by utilizing models that can infer the direction of reported findings.

In this study, we employ BART to generate summaries of clinical trials using *Brief Summary* and *Detailed Descriptions* sourced from clinicaltrials.gov. Our primary objective is to develop a model capable of enhancing brief summaries by inferring from the detailed trial descriptions, thereby resolving the issue of low-quality summaries that are either blank or need more comprehensive information about trial features. Our approach closely aligns with the methodology outlined in a previous paper [3]. However, we fine-tune BART, a well-suited auto-regressive model for summarization. We propose a novel knowledge graph that assesses generated summaries, which includes entity types and sentence numbers as nodes,that are connected when the associated sentence contains the corresponding entity class. We create two graphs for each trial record, one for the original and one for the BART-generated summary. The graphs' similarities indicate how well entity classes and their sentence structure are maintained in the summaries.

3 Methods

In the field of NLP, summarization tasks play a crucial role in distilling extensive texts into concise, informative summaries. Generally, there are two primary types of summarization methods: extractive and abstractive. Formally, given a document $D = \{s_1, s_2, \ldots, s_n\}$, an abstractive summary S is created such that $S = \{s'_1, s'_2, \ldots, s'_m\}$, where $m \leq n$, and each s'_i is a newly generated sentence encapsulating the document's core information. However, extractive summarization selects and compiles the most relevant and significant sentences from the original document without altering their form. In this approach, for the same document D, the extractive summary d is a subset of D, defined as $d = \{s_{i1}, s_{i2}, \ldots, s_{ik}\}$, where each s_{ij} is directly taken from D and $k \leq n$. This method maintains the integrity of the original text's structure and content.

In this research, we utilize the BART model which is an Encoder-Decoder transformers architecture to generate the trials' abstractive summaries. Three predominant training strategies are delineated in contemporary literature. The first, feature-based methods, use Pre-trained Language Models (PLMs) for contextual representations without altering their pre-trained parameters. The second, which we adopt, is the fine-tuning-based approach, where PLMs are fine-tuned as text encoders on a task-specific basis, enhancing their performance for specific tasks like summarization. The third, domain-adaptation-with-fine-tuning, involves initially adapting PLMs to a specific domain before fine-tuning them to task-specific data, thus blending broad and domain-specific knowledge [8].

3.1 Dataset

Clinicaltrials.gov frequently publishes their database's XML archives for content analysis. It contains all key information about trials such as: interventions, conditions, descriptions of the trial, and study arms. There are two primary columns in this database labeled as *Detailed Description* and *Brief Summary* which are focus of our research.

The Detailed Description field on ClinicalTrials.gov, written by human experts, provides a comprehensive overview of a clinical study. It includes the study's objectives, design, participant eligibility criteria, intervention details, and outcome measures. Additionally, it outlines the study's duration, locations, and contact information. This section is essential for conveying the study's purpose, methodology, and other key aspects to researchers, healthcare professionals, and potential participants. Alongside the Detailed Description, ClinicalTrials.gov also features a Brief Summary section. This section offers a concise overview of the clinical study, presenting key information in a readable format. It typically includes a succinct explanation of the study's purpose, the type of research being conducted, and basic information about the study design and interventions. The Brief Summary is designed to provide

a snapshot of the study, making it easier for the general public, patients, and healthcare professionals to understand the essential aspects of the research without delving into the technical details found in the Detailed Description.

3.2 *Data Processing*

Data quality is a crucial part of any Machine Learning pipeline [12]. The raw data in its original form is often too cluttered with noise to be effectively utilized for fine-tuning transformer models. Given these models' robust memory capabilities, they benefit significantly from a smaller and high-quality dataset rather than a larger noisy one. To achieve this higher quality, we implement several preprocessing steps. These steps are designed to sift through the data, removing low-quality entries and refining the dataset for training purposes.

$$\text{TF}(b, d) = \frac{\text{Frequency of bigram } b \text{ in document } d}{\text{Total number of bigrams in document } d} \tag{1}$$

$$\text{IDF}(b, D) = \log\left(\frac{N}{\text{Number of documents with bigram } b}\right) \tag{2}$$

$$\text{TF-IDF}(b, d, D) = \text{TF}(b, d) \times \text{IDF}(b, D) \tag{3}$$

Figure 2 details our study's preprocessing steps. One of the key motivations for our research stems from the observation that a significant number of trials lack descriptions and summaries. Given our model's reliance on pairs of detailed descriptions and brief summaries, the initial phase of preprocessing involves filtering out trials that lack either of these elements. Following this, we set aside approximately 5% of the data, equating to about 15,000 trials, as a held-out test set to evaluate model performance. This test set was selected prior to further data cleaning to maintain a

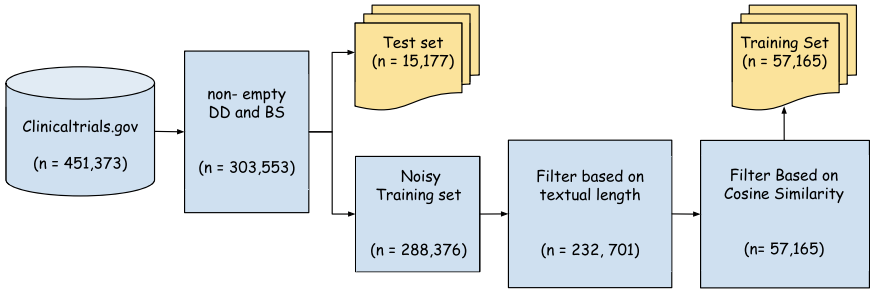

Fig. 2 Preprocessing steps flowchart. Initial filtering removes trials without descriptions or summaries, followed by setting aside a 5% test set. Further steps include truncating long summaries and constructing bigram TF-IDF vectors for similarity analysis. The process yields 57,165 quality pairs for fine-tuning the BART model. Detailed Description (DD) and Brief Summary (BS)

representation of the original data distribution, thereby ensuring a fair comparison with baseline models that were tested on a similar dataset.

In the subsequent stages of preprocessing, after segregating the test set, we begin by removing summaries that are over 70% as long as their corresponding detailed descriptions. This approach helps the model learn to generate more concise summaries, rather than reproducing summaries of similar length to the original descriptions. Following this, we construct bigram TF-IDF vectors for both source and target columns of the remaining data, as outlined in Eqs. 1–3. These vectors play a critical role in evaluating the similarity between the columns. To quantify this similarity, we use the Cosine Similarity score, which ranges from -1 to 1. Scores closer to 1 indicate a strong correlation, those nearing -1 suggest diametric opposition, and a score of 0 indicates no similarity. In our analysis, we focus on trials exhibiting high similarity, selecting those with a Cosine Similarity score above a certain threshold, which we set at 0.3 in our experiment. The preprocessing steps results in 57,165 high-quality pairs of detailed descriptions and brief summaries used to fine-tune our BART model.

3.3 Model Fine-Tune

In our approach, the BART model, which incorporates an encoder-decoder structure, is used to transform the textual content from the 'Detailed Description' column into a summarized form represented in the 'Brief Summary.' Let's consider $x = \{x_1, x_2, ..., x_n\}$ as the sequence of tokens from the 'Detailed Description', and $y = \{y_1, y_2, ..., y_m\}$ as the corresponding token sequence in the 'Brief Summary'. The encoder part of the BART model, denoted as $\text{Enc}_{\theta_{enc}}$, converts the input sequence x into a continuous latent representation z. This encoding process is mathematically expressed as $z = \text{Enc}_{\theta_{\text{enc}}}(x)$, where z symbolizes the encoded form of x, and θ_{enc} refers to the encoder's parameters. Following the encoding, the decoder, represented as $\text{Dec}_{\theta_{\text{dec}}}$, takes over to produce the output summary sequence $\hat{y}$.

The decoding process is captured by the equation $\hat{y} = \text{Dec}_{\theta_{\text{dec}}}(z)$, where the decoder aims to generate a summary that approximates the target sequence y, with θ_{dec} as the decoder's parameters. Through this iterative training process, the model's ability to generate accurate and coherent summaries from the input text is enhanced, leading to improved performance in summarizing the 'Detailed Description' column into the 'Brief Summary'.

4 Results

This section evaluates our model, divided into three key components. Firstly, we introduce the metrics employed for assessment. Then, we delve into the BART's implementation details, shedding light on its performance in the clinical trials'

summarization task. Lastly, we present comprehensive insights into generating knowledge graphs, outlining the process that facilitates of entity preservation evaluation within the summarized text.

4.1 Model Evaluation

$$\text{Precision} = \frac{\sum_{S \in \{\text{Reference Summaries}\}} \sum_{\text{gram}_n \in \text{System Summary}} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{\text{gram}_n \in \text{System Summary}} \text{Count}(\text{gram}_n)} \tag{4}$$

$$\text{Recall} = \frac{\sum_{S \in \{\text{Reference Summaries}\}} \sum_{\text{gram}_n \in \text{System Summary}} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{S \in \{\text{Reference Summaries}\}} \sum_{\text{gram}_n \in S} \text{Count}(gram_n)} \tag{5}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision}_{\text{ROUGE-N}} \times \text{Recall}_{\text{ROUGE-N}}}{\text{Precision}_{\text{ROUGE-N}} + \text{Recall}_{\text{ROUGE-N}}} \tag{6}$$

In the initial step of model evaluation, we measured the system performance using the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) score [13], a prevalent metric to assess sequence-to-sequence systems such as summarization and translation. The ROUGE metric encompasses several variants, each focusing on different aspects of the text. ROUGE-1 and ROUGE-2 measure the overlap of unigrams and bigrams, respectively, between the model predictions and reference texts, providing insights into lexical similarity. ROUGE-L, on the other hand, evaluates the longest common subsequence, which is crucial to understand sentence-level structure and fluency. These metrics, as detailed in Eqs. 4–7, collectively offer a comprehensive evaluation of our model's performance. Our model, employing an abstractive summarization approach, outperformed the baseline models by a significant margin. This was evident in the results of Table 1, which shows marked improvements across all three ROUGE-1, ROUGE-2, and ROUGE-L scores, indicating not only lexical alignment with reference texts but also structural and contextual coherence.

$$\begin{aligned} R_{\text{lcs}} &= \frac{\text{LCS}(X, Y)}{m} \\ P_{\text{lcs}} &= \frac{\text{LCS}(X, Y)}{n} \\ F_{\text{lcs}} &= \frac{(1 + \beta^2) \cdot R_{\text{lcs}} \cdot P_{\text{lcs}}}{R_{\text{lcs}} + \beta^2 \cdot P_{\text{lcs}}} \end{aligned} \tag{7}$$

Table 1 The performance metrics, including ROUGE-1, ROUGE-2, and ROUGE-L, across various methods for summarizing clinicaltrial.gov descriptions

	ROUGE-1			ROUGE-2			ROUGE-L		
	F1	R	P	F1	R	P	F1	R	P
Random	0.316	0.300	0.297	0.130	0.128	0.125	0.279	0.266	0.250
LexRank	0.359	0.359	0.348	0.171	0.176	0.168	0.319	0.320	0.297
TextRank	0.348	0.380	0.353	0.167	0.189	0.172	0.309	0.338	0.300
LSA	0.337	0.368	0.343	0.161	0.176	0.164	0.299	0.328	0.293
Luhn	0.344	0.371	0.346	0.164	0.183	0.168	0.306	0.331	0.296
SumBasic	0.336	0.296	0.302	0.134	0.123	0.123	0.296	0.263	0.253
KLSUm	0.326	0.317	0.312	0.143	0.140	0.137	0.288	0.281	0.265
BART	**0.402**	**0.450**	**0.409**	**0.213**	**0.241**	**0.223**	**0.369**	**0.412**	**0.369**

The best scores highlighted in bold text

4.2 Implementation Details

In our experiments, we employed the *"facebook/bart-large-cnn"* [2] model from the Hugging Face transformers library for sequence-to-sequence language processing tasks. The learning rate was set at 5×10^{-5}, and the training duration was extended over 5 epochs. We opted for the AdamW optimizer to facilitate the training process. These specific details, including the model choice, learning rate, number of epochs, and optimizer, are provided to ensure reproducibility of our experimental setup, allowing others to replicate and validate the results.

4.3 Graph-Based Evaluation of Named Entity Order Preservation in Generated Summaries

In addition to employing the ROUGE metric, we conducted a supplementary assessment of the generated summaries using an innovative methodology centered around a bipartite knowledge graph. This approach examines the named entity class types and their corresponding sentence numbers. The objective is to verify that the generated summaries maintain the fidelity of both the entity class types and their coherent sequential arrangement within the overall structure of the summarized texts. This multifaceted evaluation provides a comprehensive perspective on the summarization process' effectiveness, ensuring linguistic coherence and semantic integrity in the representation of information.

Nodes Representation Within the graphs, our nodes are organized into two distinct classes. The first class indicates the named entity types, while the second class corresponds to the text's sentence numbers. We are doing the named entity type extraction in two passes.

- **First Pass**: We leverage *llama 2-70 billion* parameters [14] served by vLLM [15] to extract named entities from the brief summary derived from a randomly sampled set of 1000 records. We employ a frequency-based criterion to discern the ten most commonly occurring entity types within the extracted text. These are 'disease,' 'medical condition,' 'drug,' 'device,' 'dose or measurements,' 'clinical trial phase,' 'population,' 'time,' 'medical procedure,' and 'biomarker.' We address this node type as $E = \{e_1, e_2, \ldots, e_{10}\}$.
- **Second Pass**: After identifying the most prevalent entity types, we select the summaries generated by BART for the initial 1000 records. In this phase, we disassemble the summaries into individual sentences. Subsequently, both the generated and original summaries undergo analysis using *llama 2 70 billion*. Listing 1.1 indicates the prompt we have used. This involves applying a specific prompt to ascertain the presence or absence of a particular entity type within each sentence. The outcome of this pass for each trial d_i results in two binary matrices. The first matrix corresponds to the original brief summary as ground truth which has dimensions $m_i \times 10$, where m_i represents the number of sentences in the document i. The second matrix pertains to the generated summary and possesses dimensions $n_i \times 10$; with n_i denotes the number of sentences in the generated summary text i. In this case $i \in \mathbb{N}$, $1 \leq i \leq D$.

Graph Representation We employ binary matrices to construct a bipartite graph that captures the relationships between entity types and sentence numbers for both the original and generated records. We designate the graph corresponding to the i_{th} original records as G_{1i}, and the graph for the model-generated summaries as G_{2i} (Fig. 3).

Graph-Based Evaluation Results In our study, we focus on documents with an equal number of sentences in both their original content and the BART-generated Brief Summaries ($n = m$). The assessment of these bipartite graphs relies on utilizing Jaccard similarity, as outlined in Formula 9. Conducting D trials (in our case, $D = 1000$), we calculate the average Jaccard similarity, yielding 0.71. This result signifies that approximately **71%** of entity classes and their sentence-wise positional relationships within the sentences are retained in the summarized text.

Listing 1.1 Llama 2 prompt for extracting entity classes

```
quest_list = ["disease","medical condition","drug","device",
"Dose or measurements","clinical trial phase","population",
"Time","Medical Procedure","Biomarker"]


question = f""" Does the following sentence includes any
   named entities with {{quest_list[i]}} type?:
   '''{{sentence}}'' """
```

Proposition 1 *Assume the number of sentences in generated and original summaries is identical, denoted as $n = m$. In this context, $E(G)$ represents the edges of the graph G.*

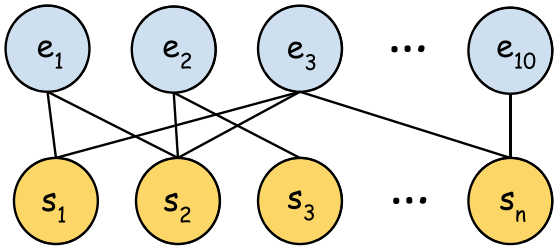

Fig. 3 Overview of a bipartite graph illustrating connections between entity class nodes ($e_1, e_2, ..., e_{10}$) and sentences on the record ($s_1, s_2, ..., s_n$)

$$\frac{\sum_{i=1}^{D} J(G_{i1}, G_{i2})}{D} \quad \text{for } 1 \leq i \leq D \tag{8}$$

$$J(G_1, G_2) = \frac{|E(G_1) \cap E(G_2)|}{|E(G_1) \cup E(G_2)|} \tag{9}$$

5 Conclusion

Our research marks a significant stride in NLP's application to medical informatics, leveraging the capabilities of the BART model to generate structured summaries of clinical trials. Our fine-tuned BART model notably outperforms previous baseline models, distinguishing itself with its abstractive approach that enables it to generate more coherent and contextually rich summaries. This advancement in summarization technology is pivotal, as it transforms detailed and often complex trial information into concise, comprehensible formats. The enhanced accessibility and utility of trial data, as facilitated by our model, underscore its potential to significantly impact the way clinical information is consumed and utilized.

Looking ahead, the research reveals two critical areas for future exploration. The first involves a deeper comparative analysis with expert-generated summaries. This comparison would assess the AI-generated summaries against those created by human experts, providing a nuanced understanding of the model's accuracy and areas for improvement. Secondly, there is an exciting opportunity to explore generating summaries based on inherent trial properties, such as their goals, methodologies, and outcomes, independent of their detailed descriptions. Such an approach promises to refine the summarization process, making it more efficient and possibly unveiling new perspectives in trial classification and analysis. These future research paths hold the promise of not only extending the current work's utility but also of contributing significantly to NLP's evolving landscape in the realm of medical data analysis [16–19]. Our previous efforts in the areas of information retrieval and document analysis may also significantly affect our future work [20–22].

Acknowledgements This research paper is based on works that are supported by the National Science Foundation Grant No. 2117941

References

1. Ramprasad S, Marshall IJ, McInerney DJ, Wallace BC (2023) Proceedings of the conference. Association for Computational Linguistics Meeting, vol 2023. NIH Public Access, p 236
2. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, Stoyanov V, Zettlemoyer L (2020) Jurafsky D, Chai J, Schluter N, Tetreault J (eds) Proceedings of the 58th Annual meeting of the association for computational linguistics. Association for Computational Linguistics, Online, pp 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703. URL https://aclanthology.org/2020.acl-main.703
3. Gulden C, Kirchner M, Schüttler C, Hinderer M, Kampf M, Prokosch HU, Toddenroth D (2019) Int J Med Inf 129:114
4. Chaves A, Kesiku C, Garcia-Zapirain B (2022) Information 13(8):393
5. Chen Y, Ma Y, Mao X, Li Q (2019) Data Sci Eng 4:14
6. Yongkiatpanich C, Wichadakul D (2019) 2019 IEEE 4th International conference on computer and communication systems (ICCCS). IEEE, pp 105–110
7. Du Y, Li Q, Wang L, He Y (2020) Knowl-Based Syst 199:105964
8. Xie Q, Luo Z, Wang B, Ananiadou S (2023) https://api.semanticscholar.org/CorpusID:259924436
9. Akdemir A, Shibuya T (2020) CLEF working notes
10. DeYoung J, Beltagy I, van Zuylen M, Kuehl B, Wang LL (2021) arXiv preprint arXiv:2104.06486
11. Wallace BC, Saha S, Soboczenski F, Marshall IJ (2021) AMIA Summits Translat Sci Proc 2021:605
12. Jain A, Patel H, Nagalapatti L, Gupta N, Mehta S, Guttula S, Mujumdar S, Afzal S, Sharma Mittal R, Munigala V (2020) Proceedings of the 26th ACM SIGKDD International conference on knowledge discovery & data mining. Association for Computing Machinery, New York, NY, USA, KDD '20, pp 3561–3562. https://doi.org/10.1145/3394486.3406477
13. Lin CY (2004) Annual meeting of the association for computational linguistics. https://api.semanticscholar.org/CorpusID:964287
14. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, Bashlykov N, Batra S, Bhargava P, Bhosale S et al (2023) arXiv preprint arXiv:2307.09288
15. Kwon W, Li Z, Zhuang S, Sheng Y, Zheng L, Yu CH, Gonzalez JE, Zhang H, Stoica I (2023) Proceedings of the ACM SIGOPS 29th symposium on operating systems principles
16. Cummings J, Lee G, Nahed P, Kambar MEZN, Zhong K, Fonseca J, Taghva K (2022) Alzheimer's & Dementia: Translat Res Clin Interventions 8(1):e12295
17. Kambar MEZN, Nahed P, Cacho JRF, Lee G, Cummings J, Taghva K (2022) Proceedings of Sixth International congress on information and communication technology: ICICT 2021, London, vol 1. Springer, pp 513–521
18. Nahed P, Kambar MEZN, Cacho JRF, Lee G, Cummings J, Taghva K (2023) Intelligent sustainable systems: selected papers of WorldS4 2022, volume 1. Springer, pp 63–73
19. Esmaeilzadeh A, Taghva K (2021) In Arai K (ed) Intelligent systems and applications—proceedings of the 2021 intelligent systems conference, IntelliSys 2021, vol 3. Amsterdam, The Netherlands, 2-3 September, 2021. Lecture notes in networks and systems, vol 296. Springer, pp 175–189
20. Taghva K, Nartker TA, Borsack J, Condit A (2000) Document recognition and retrieval VII, San Jose, CA, USA, January 22. In: Lopresti DP, Zhou J (eds) SPIE proceedings, vol 3967, pp 157–164
21. Taghva K, Veni R (2010) In Latifi S (ed) Seventh International conference on information technology: new generations, ITNG 2010. Las Vegas, Nevada, USA, 12-14 April 2010. IEEE Computer Society, pp 222–226
22. Taghva K, Borsack J, Nartker TA, Condit A (2004) Inf Process Manag 40(3):441