Leveraging Large Language Model to Generate Multi-modal Timeline Summarization

Zheng Liu[1,3] and Chaomurilige Wang[2,3](✉)

[1] School of Chinese Ethnic Minority Languages and Literatures, Minzu University of China, Beijing 100081, China
liuzheng@muc.edu.cn

[2] School of Information Engineering, Minzu University of China, Beijing 100081, China
chaomurilige@muc.edu.cn

[3] Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing 100081, China

Abstract. In the current era of abundant information, the skill to extract crucial events and present them in a succinct manner holds significant importance. Multi-modal timeline summarization (MTS) is especially pivotal across various domains such as news reporting, historical analysis, and social media monitoring. The recent progress in natural language processing has led to the emergence of large language models as potent tools for generating accurate and coherent summaries. However, the application of Large Language Models (LLMs) in multimodal summarization faces challenges, including the scarcity of extensive multimodal datasets for training and evaluation, as well as the requirement for more sophisticated evaluation metrics that can accommodate the multi-modal nature of the task, in contrast to current metrics designed for text summarization. Addressing these challenges necessitates the development of larger and more diverse datasets, along with the creation of novel evaluation metrics tailored to multi-modal summarization. This study aims to investigate the utilization of large language models in generating timeline summarizations, elucidating their capabilities, challenges, and potential applications.

Keywords: Multimodal Learning · Timeline Summarization · Large Language Model

1 Introduction

1.1 A Subsection Sample

In recent years, the field of natural language processing (NLP) has witnessed groundbreaking advancements, primarily attributed to the development of large language models. These models, such as OpenAI's GPT (Generative Pre-trained Transformer), have demonstrated impressive capabilities in various NLP tasks, including machine translation, sentiment analysis, and text generation. One particularly intriguing application of these models is timeline summarization generation, where they have shown promising potential [1].

W. Yu and L. Xuan (Eds.): MLICOM 2023, LNICST 518, pp. 157–168, 2024.
https://doi.org/10.1007/978-3-031-71716-1_13

Timeline summarization involves the extraction of critical events from a given set of documents or sources and presenting them in a compressed and cohesive manner. It serves as a valuable tool for condensing vast amounts of information into a concise and easily digestible format. News agencies can employ timeline summarization techniques to provide users with an overview of important events, while historians can use it to analyze historical timelines more efficiently [2].

Before the advent of large language models, traditional approaches to timeline summarization relied on rule-based systems, machine learning algorithms, and information retrieval techniques. These methods often struggled to capture the diverse and contextual nature of language, leading to sub-optimal summarization quality [3]. However, the emergence of large language models has introduced a paradigm shift in this domain.

Large language models, trained on massive amounts of text data, possess the ability to learn complex patterns, dependencies, and semantic relationships within language. This makes them well-suited for timeline summarization generation. By utilizing pretrained models and fine-tuning them on specific summarization tasks, these models can effectively summarize timelines by understanding the temporal ordering of events, their significance, and their relationships [4].

While large language models offer promising results in timeline summarization, several challenges and limitations need to be addressed. One significant challenge is the risk of generating biased or inaccurate summaries due to the biases inherent in the training data [5]. Additionally, large language models often lack control and can generate summaries that are excessively detailed or miss important events [6]. Overcoming these challenges requires careful fine-tuning, dataset curation, and evaluation methodologies.

The main contributions in this paper can be summarized as following:

(1) In this paper, we purpose a novel and meaningful task: multi-modal timeline summarization. We purpose a comprehensive problem definition for the new task, in which text is treated as the central modality while others as adjacent modalities.
(2) Besides, we purpose a simple framework and a baseline model to solve MTS task. Researchers can conduct further research based on this baseline model. Our baseline generates smooth content for each daily content for timeline via LLM, attempting to explore paradigms between LLM and multimodal summarization.
(3) Aiming to solve the problem of the lack of evaluating methods, we purpose a batch of metrics for quantifying the generated multi-modal timeline summarization.

2 Relate Works

2.1 Multi-modal Event Detection

Multi-modal event detection has emerged as a noteworthy and rapidly developing research area, with the primary objective of autonomously detecting and comprehending events through the integration of diverse modalities, including text, images, audio, and video. The synergistic analysis of information from these multiple modalities enables a more comprehensive understanding of complex real-world events, and facilitates the creation of intelligent systems with advanced event comprehension capabilities [7].

In recent years, substantial advancements have propelled the field of multi-modal event detection. These advancements can be attributed to several key factors, including the availability of large-scale multi-modal datasets, progress in deep learning algorithms, and the growing demand for intelligent systems with the capacity to comprehend and interpret real-world events [8].

Furthermore, the evolution of deep learning algorithms has significantly enhanced the capacity of multi-modal event detection systems to extract and analyze complex patterns and correlations across different modalities [9]. The application of deep learning techniques, such as multi-modal fusion and cross-modal attention mechanisms, has facilitated the seamless integration of information from text, images, audio, and video, leading to improved event detection and comprehension [10].

The increasing demand for intelligent systems capable of comprehending real-world events has been a driving force behind the rapid progress in multi-modal event detection. With the proliferation of multimedia content across various platforms, there is a growing need for automated systems that can effectively process and make sense of diverse types of information to extract meaningful events and insights.

2.2 Multimodal Summarization

Multi-modal summarization is a rapidly developing research area in the field of natural language processing (NLP). It focuses on generating concise and coherent summaries from multiple modes of information, including text, images, audio, and video [11–13]. While traditional text-based summarization techniques have been successful in extracting important information from single textual documents, they are limited in their ability to handle the diverse content available in real-world scenarios.

To achieve effective multi-modal summarization, researchers have explored different approaches. One common approach is to extract salient information from each modality separately and then combine them to create a coherent summary. For instance, relevant textual information can be extracted using text-based NLP techniques, while visual or auditory cues can be used to identify important details in images or audio. These extracted pieces of information are then fused together to form a concise summary that captures the essence of the multi-modal content [14–16].

Another approach in multi-modal summarization is to employ techniques that can directly process and understand the content across different modalities. This involves developing models that can simultaneously analyze textual, visual, and auditory information, considering the interdependencies and interactions between them [17]. These models can use advanced deep learning techniques, such as neural networks with attention mechanisms, to capture the relationships between the different modalities and generate cohesive summaries. By leveraging the strengths of different modalities, multi-modal summarization enables a more comprehensive understanding of complex content [18].

2.3 Large Language Models

The potential applications of large language models in timeline summarization generation are vast. Beyond news agencies and historians, these models can be integrated into various domains that require efficient information extraction, such as social media monitoring, legal document analysis, and medical record summarization. Additionally, advancements in large language models can lead to the development of user-interactive systems, allowing users to customize the level of summary detail and specify their preferences [19, 20].

3 Methodology

3.1 Problem Definition

According to [2], the timeline summarization (TLS) can be defined as a task that generates a sequence of dates and their associated daily natural language descriptions from a collection of topic-related documents. These descriptions summarize the most important content related to the topic. The multi-modal timeline summarization (MTS) task can be defined as a TLS task that takes more than one mode of information representation (termed as modality) as input, and depends on information sharing across different modalities to generate the final timeline. In MTS task, natural language is treated as the central modality, while other modalities are treated as adjacent modalities to aid the central modality for enriching the output information.

Thus, we formulate MTS task as follows:

Input: A time-stamped news article collection $D = \{doc_1, doc_2, \ldots, doc_{|D|}\}$. The collection D contains event-related news information with more than one mode of information representation.

Output: A multi-modal timeline summarization MTS is generated by integrating multiple types of information of D so that MTS includes a sequence of time/date and summary pairs $(t_1, s_1), \ldots, (t_k, s_k)$ where s_k are the summary sentences for the time t_k. Each summary in MTS should be informative, consistent and coherent.

3.2 Framework

Multi-modal timeline summarization is an important technique in the field of information retrieval and multimedia analysis. It involves the integration of multiple modalities, such as images, videos, and audio, into a cohesive and comprehensive text-based summary of events. The goal is to provide users with a condensed and coherent representation of a timeline, which includes various types of media.

To achieve multi-modal timeline summarization, a common framework is to transform all modalities into natural language, enabling a seamless integration of different media types. This text-central framework, as shown in Fig. 1, allows for the extraction and fusion of information from various modalities, leading to a more effective summary of events.

The first step in the text-central framework is to process and analyze each modality separately. Textual information can be extracted using natural language processing techniques, while images and videos can be analyzed using computer vision algorithms, and audio can be transcribed and analyzed using speech recognition technologies. These individual analyses provide important information from each modality.

Next, the extracted information from different modalities is integrated into a unified representation. This can be achieved through various fusion techniques, such as feature-level fusion, where features from different modalities are combined, or decision-level fusion, where decisions from different modalities are aggregated.

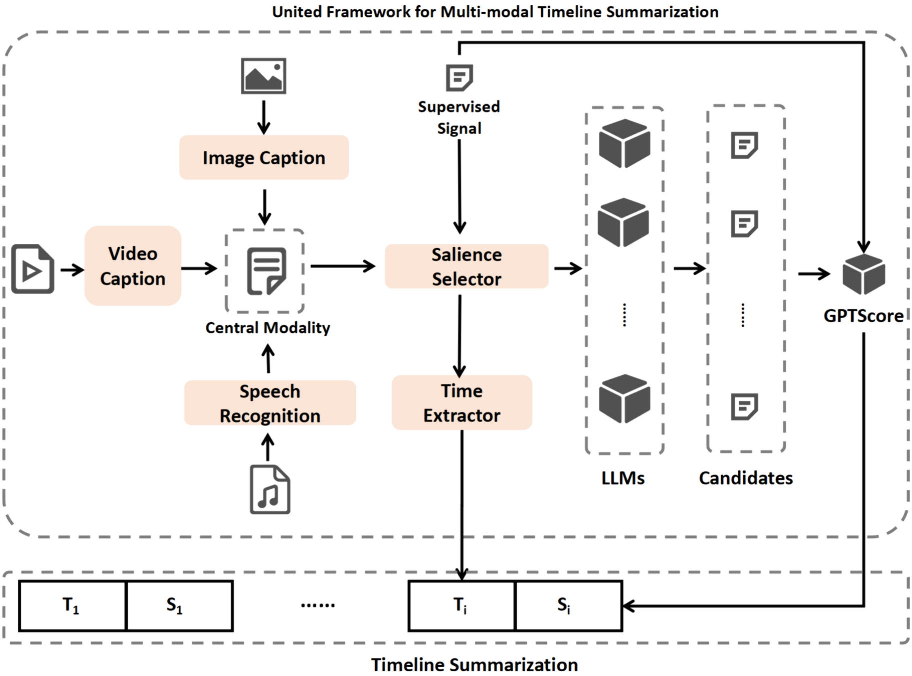

Fig. 1. The text-central united framework for multi-modal timeline summarization via LLM.

Once the information is integrated, the final step is to generate a text-based summary that represents the timeline of events. This summary should be concise, coherent, and capture the key aspects of the events. Natural language generation techniques, such as text summarization algorithms, can be applied to transform the integrated information into a textual summary.

The resulting multi-modal timeline summary provides users with a comprehensive and coherent representation of the events, incorporating various media types. This has important applications in fields such as news summarization, video summarization, and event understanding, where users can quickly grasp the key information from a large volume of multimedia data.

3.3 Multi-modal Information Representation

As the descriptions of event news, all the multi-modal information can be integrated to text format. This integration involves transforming images, videos, and audio into textual representations using specialized models such as image captioning, video captioning, and speech recognition. By leveraging these models, the diverse components of an event news article can be seamlessly combined into a cohesive and coherent textual representation. This approach ensures that all relevant information is conveyed in a unified and accessible manner, enhancing the overall quality and comprehensibility of the news content.

Thus, the input of the text-central framework can be described as following:

$$Input = text + IC(img) + VC(vi) + SR(au) \tag{1}$$

where *IC, VC, SR* represent image captioning, video captioning, and speech recognition operation. *Img, vi, au* denote the modalities of image, video and audio.

3.4 Salience Selector Trained with Supervised Signal

In the context of transforming information into text mode, the limitation of input length can restrict the use of large language models (LLMs). To address this limitation, researchers have proposed the use of a salience selector to identify and prioritize important information that can provide a better context for an event.

The salience selector approach involves training a neural network model to rank the importance of text segments, typically using supervised signals provided by reference summaries in the datasets. By leveraging these reference summaries, the model can effectively learn to identify salient information and improve the quality of the generated context.

The objective of the loss function in this approach is twofold. First, the model aims to minimize the similarity between the topic of an event and the text within a paragraph. This ensures that the generated context aligns well with the specific event being summarized. Second, the model also seeks to minimize the discrepancy in similarity between the reference summary and the paragraph text. This encourages the model to generate summaries that are not only contextually relevant but also align closely with the reference summaries, thus maintaining coherence.

By minimizing these differences, the model learns to capture the relationship between event topics and paragraph content more effectively. It balances the need to provide accurate and informative summaries while considering the alignment with the reference summaries. This is particularly relevant in an academic context, where summarization requires precision and coherence.

The use of a salience selector and the training of a neural network model with supervised signals help address the challenge of input length limitations in generating high-quality summaries. By prioritizing salient information and optimizing the model's performance in capturing event-topic relationships, this approach enhances the model's ability to generate accurate and coherent summaries in an academic setting.

In conclusion, the incorporation of a salience selector and the training of a neural network model with supervised signals improve the quality of the generated context by selecting major information relevant to an event. The objective of the loss function is to minimize differences in similarity, both between the event topic and paragraph text and between the reference summary and paragraph text. This approach enables the model to generate accurate and coherent summaries, which is crucial in an academic context.

3.5 Time Extraction via LLM

Time selection involves two main components: extracting text timestamps and selecting significant time points. Our framework leverages the zero-shot ability of Large

Language Models (LLMs) to extract temporal information from news articles. In addition, we employ a prompt learning Question-and-Answer (Q&A) method to identify the occurrence time within each paragraph. By appropriately organizing the extracted time information, we group together sentences that share the same temporal context, thereby forming coherent and meaningful paragraphs.

The importance of publishing time and extraction time needs to be assessed individually. Publishing time refers to the time at which a document is made available to the public, while extraction time represents the time required to extract relevant information from the document. Both of these factors can be evaluated by analyzing their frequency of occurrence and obtaining their respective distributions. To determine the significance of publishing time and extraction time, a weighted and biased summation of their frequency distributions can be performed. This approach involves assigning weights to each distribution based on their significance in the context of timeline summarization. By establishing a threshold, we can identify time points above this threshold as crucial for effectively summarizing the timeline. Through a comprehensive analysis of the frequency distributions of publishing time and extraction time, their importance can be evaluated individually, leading to a more accurate and comprehensive timeline summarization. Thus, the distribution score can be calculated as following:

$$score = \alpha * Pt + \beta * Et + bias \tag{2}$$

where α, β are the weights to each distribution and Pt, Et denote the publish time and extracted timestamps. These salient paragraphs serve as the foundation for generating timeline summaries in the next steps.

3.6 Multi-modal Summarization via LLM

With the advent of advanced natural language processing techniques, large language models have emerged as powerful tools for automating the process of summarizing text documents. Inspired by GPTScore [3], our framework utilizes several LLMs to generate different candidate summaries and select the final summarization by GPT-3.5. The target of the GPTScore can be defined as following:

$$Target = argmax(P(Candidate,\ Salience\,Content,\ GPT{-}3.5)) \tag{3}$$

GPTScore utilizes the emergent abilities of LLM to score generated texts [3] and select the final summaries based on the source text. The final summaries would be the most proper content for describing the source text.

3.7 Timeline Assembly

Understanding and analyzing the information presented in a timeline requires a strategic approach to organize the data effectively. This typically involves arranging all given paragraphs in ascending chronological order, while simultaneously disregarding non-critical time points. Such a method allows us to pinpoint pivotal moments in time and make connections with corresponding multi-modal summaries.

Multi-modal technology, encompassing multiple types of input and output, could be indispensable for this task. For instance, if a particular model requires a multi-modal output, the calculation of modal similarity becomes crucial. In addition, multi-modal retrieval technology could be employed to match the modal information in tandem with any abstract text. As part of a larger system, these tools can assist in synthesizing and processing data in a more holistic manner.

The aforementioned approach strives to enhance our comprehension of complex datasets by structuring them in a logical and coherent manner. It aims to unearth patterns and relationships which might otherwise remain concealed when using traditional analytical methods. By utilizing multiple modes of information processing, we can delve deeper and uncover hidden relationships within the data.

Furthermore, the nuanced understanding gained from this approach can give rise to valuable insights and novel discoveries. The benefits of this methodology are far-reaching, with potential applications across a myriad of fields and industries. It could contribute to innovation in academic research, business strategy, historical studies, and many more domains, by providing a more comprehensive view of the data under analysis.

4 Experiment Settings

4.1 Datasets

In this study, we have constructed the mm-disaster dataset with the aim of evaluating the efficacy of our proposed framework. The mm-disaster dataset comprises four clusters of event documents. On average, each cluster contains approximately 120 documents. Within the mm-disaster dataset, each document consists of an average of 15.25 paragraphs. Furthermore, each paragraph contains an average of 33.54 tokens.

To ensure the accuracy of our dataset, we have established a reference timeline by integrating information from reputable sources such as Wikipedia and Baidu Baike Encyclopedia. This timeline provides a detailed account of each event, with temporal information accurate down to the minute. By utilizing this meticulously curated dataset, we are able to thoroughly assess the performance of our framework.

4.2 Evaluation Metrics

Salient Time Selection Quality (STSQ). Assessing the effectiveness of selecting significant time points primarily entails the computation of precision (P) and recall (R) measures. However, in order to evaluate the results in a more rigorous and meaningful manner, it is crucial to incorporate the F1 score as well.

F1 score can be defined as following:

$$F1 = \frac{2 * P * R}{P + R} \tag{4}$$

Text Summarization Quality (TSQ). Numerous studies have been conducted to assess the quality of generated text, employing various evaluation methods. Among the commonly utilized approaches are n-gram-based metrics like BLUE and ROUGE [5], as well as more recent techniques such as BERTScore [6] and MoverScore [7], which leverage pre-trained models. These evaluation methods serve the crucial purpose of assessing the linguistic excellence and coherence of generated text, contributing to the advancement of the field of natural language generation.

Timeline Assembly Quality (TAQ). Evaluation metrics in terms of details include the quality of time and text. It is crucial to consider the timeline assembly quality of the generated results. To effectively evaluate the entire timeline, two comprehensive approaches are adopted. The first method (TAQ-1) involves concatenating the entire timeline into a cohesive text by TSQ method. The second method (TAQ-2) subsequently calculates the summary at each time point after the TSQ assessment. Various techniques are employed in this process, including handling text at non-salient time points. To evaluate the quality of the entire timeline, we have implemented an evaluation method based on a reward and punishment mechanism (RP). This ensures a logical and academic evaluation of the entire timeline's quality.

The final TAQ score can be defined as following:

$$TAQ - 1 = TSQ(concat(S_i)) \tag{5}$$

$$TAQ - 2 = \sum\nolimits_{k}^{1} RP(t_i) * TSQ(S_i) \tag{6}$$

where $RP(t_i)$ denotes the result of time selection, which is 1 if correct and -1 if incorrect.

TAQ-1 considers the timeline as a long single text and gauges the quality of outputs by measuring the similarity between the evaluation results and reference answers. In contrast, TAQ-2 provides a more nuanced evaluation by taking into account the correct timing of the context and assessing the quality of the context text. Ultimately, these evaluation methods contribute to the advancement of timeline generation techniques and facilitate the development of more accurate and reliable models.

4.3 Baseline Implementation

To evaluate the usability of the framework, this article employs the latest relevant research to implement various models. For image captioning, Clip is utilized, while for video captioning, ClipCap is employed. Speech recognition is accomplished through the use of Whisper. Furthermore, the salience selector utilizes a BERT-sentence model based on Salience Networks. To achieve optimal results, several large language models (LLMs) are chosen, including LLaMA-7B, LLaMA-13B, Vicuna-13B, Alpaca-7B, and Alpaca-13B. The combination of these models provides a comprehensive and effective framework for various natural language processing tasks. The selection of these models is based on their performance and accuracy in previous studies, ensuring the reliability of the framework.

4.4 Results

To prove the effectiveness of our framework, we have successfully implemented it and conducted a MTS generation experiment using the mm-disaster dataset. The generated results were analyzed using a set of evaluation metrics proposed in this article. These metrics were used to organize the evaluation results of the generated context across three dimensions: STSQ, TSQ, and TAQ. To gauge the significance of each component in the framework, we performed ablation experiments, and the comprehensive results are presented in Table 1. For STSQ, we considered accuracy, recall, and their respective F1 scores. TSQ was evaluated using the ROUGE series as the chosen metrics, where R-1, R-2, and R-L represent their F1 scores. TAQ encompassed two values, TAQ-1 and TAQ-2, as reported in Sect. 4.2.

Table 1. The ablation results of our method.

Models/Metrics	STSQ			TSQ			TAQ	
	P	R	F1	R-1	R-2	R-L	TAQ-1	TAQ-2
MTS	0.413	0.397	0.405	31.25	25.21	24.98	22.47	9.48
-w/o MMIR	0.262	0.254	0.258	25.55	19.73	17.54	17.58	5.64
-w/o SS	0.394	0.362	0.377	27.75	21.88	20.64	19.87	8.64
-w/o GPTS	0.418	0.393	0.405	28.25	23.97	22.99	20.11	9.11

The primary intent of this research article is to provide empirical evidence for the viability of our proposed framework rather than to introduce a State-of-the-Art (SOTA) model. The experiments were meticulously designed to not only test the framework's capabilities but to also rigorously assess the quality of the generated context. The results, as presented in the article, substantiate the framework's competency in producing summaries that are semantically and temporally coherent. In essence, the generated context was found to be aligned with the semantic and temporal contours of the source data, thus underscoring the practical utility of the framework.

The implications of this research are manifold. By demonstrating the framework's feasibility, we lay the groundwork for further exploration and refinement within this sphere. Future investigations could, for instance, focus on scaling the framework to accommodate larger and more complex datasets, integrating advanced natural language processing techniques to enhance semantic analysis, or developing more sophisticated temporal alignment algorithms.

5 Conclusion

This paper proposes a new and significant task, namely multi-modal timeline summarization (MTS), which aims to condense multiple modalities into a concise summary in chronological order. The task is defined comprehensively, and text is considered as the primary modality while others serve as secondary ones. To address this task, a simple framework and baseline model are proposed, which can serve as a starting point for further research in this area. The proposed baseline model generates smooth content for daily timelines using LLM, which attempts to explore the relationship between LLM and multimodal summarization. To overcome the challenge of the lack of evaluating methods for MTS, a set of metrics is proposed to quantify the quality of the generated multi-modal timeline summarization. Overall, this work contributes to the field of multi-modal summarization and provides a new research direction for future studies.

As large language models continue to evolve, their impact on timeline summarization generation is becoming increasingly significant. The ability to generate coherent and accurate summaries of events from large amounts of textual data has the potential to revolutionize the way we consume and analyze information. While challenges remain, ongoing research and technological advancements provide a promising path forward. By addressing these challenges, we can harness the power of large language models to unlock new possibilities in timeline summarization generation.

References

1. Hai, C., Lee, Y.: Query-based event extraction along a timeline. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 425–432. ACM (2004)
2. Yi, Y., Adam, J., Antoine, D.T., Kazunari, S., Masatoshi, Y.: Multi-timeline summarization (MTLS): improving timeline summarization by generating multiple summaries. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, vol. 1, pp 377–387. ACL (2021)
3. Fu, J., et al.: Gptscore: evaluate as you desire. arXiv preprint arXiv:2302.04166 (2023)
4. Julius, S., Katja, M.: Abstractive timeline summarization. In: Proceedings of the 2nd Workshop on New Frontiers in Summarization, pp. 21–31 (2019)
5. Lin, C.-Y.: Rouge: a package for automatic evaluation of summaries. In: Text Summarization Branches Out, pp. 74–81 (2004)
6. Vetrov, A.A., Gorn, E.: A new approach to calculating BERTScore for automatic assessment of translation quality. arXiv, abs/2203.05598 (2022)
7. Wei, Z., et al.: MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance (2019). https://doi.org/10.18653/v1/D19-1053
8. Anubhav, J., et al.: A survey on multi-modal summarization. ACM Comput. Surv. **55**(13s), 1–36 (2023)
9. Liao, Y., et al.: WILSON: a divide and conquer approach for fast and effective news timeline summarization. In: International Conference on Extending Database Technology (2021)
10. Chen, X., et al.: Learning towards abstractive timeline summarization. In: International Joint Conference on Artificial Intelligence (2019)
11. Tran, G., Alrifai, M., Herder, E.: Timeline summarization from relevant headlines. In: Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (eds.) ECIR 2015. LNCS, vol. 9022, pp. 245–256. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16354-3_26
12. Rémy, K., Xavier, T., et al.: Finding salient dates for building thematic timelines. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Jeju Island, Korea, pp. 730–739. Association for Computational Linguistics (2012)
13. Leo, B., Maximilian, B., Katja, M.: Dataset reproducibility and IR methods in timeline summarization. In: Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), pp. 1763–1771 (2020)

14. Yijun, D., et al.: Comparative timeline summarization via dynamic affinity-preserving random walk. In: Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020), pp. 1778–1785 (2020)
15. Demian, G., Georgiana, I.: Examining the state-of-the-art in news timeline summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), pp. 1322–1334 (2020)
16. Martschat, S., Markert, K.: Improving rouge for timeline summarization. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017), pp. 285–290 (2017)
17. Martschat, S., Markert, K.: A temporally sensitive submodularity framework for timeline summarization. In: Proceedings of the 22nd Conference on Computational Natural Language Learning (CONLL 2018), pp. 230–240 (2018)
18. Pasquali, A., Campos, R., Ribeiro, A., Santana, B., Jorge, A., Jatowt, A.:TLS-Covid19: a new annotated corpus for timeline summarization. In: Proceedings of the 43rd European Conference on Information Retrieval (ECIR 2021), pp. 497–512 (2021)
19. Pasquali, A., Mangaravite, V., Campos, R., Jorge, A.M., Jatowt, A.: Interactive system for automatically generating temporal narratives. In: Proceedings of the 41st EuropeanConference on Information Retrieval (ECIR 2019), pp. 251–255 (2019)
20. Chen, X., Chan, Z., Gao, S., Yu, M.H., Zhao, D., Yan, R.: Learning towards abstractive timeline summarization. In: Proceedings of the 28th InternationalJoint Conference on Artificial Intelligence (IJCAI 2019), pp. 4939–4945 (2019)