Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 3, No. 8, 2024

Long-Range Dependency Modeling and Decision Point Summarization for Large Language Models in Dialogue and Meeting Scenarios

Chi Shing Lee
Hunter College, New York, USA
chislee0708@gmail.com

Abstract: In highly interactive text scenarios such as dialogues and meetings, maintaining long-term stability and accurately condensing key decision points are the core challenges in generating decision-oriented summaries. Meeting transcriptions typically feature long rounds, frequent topic shifts, omissions of references, and scattered key information, easily leading to omissions of key points, missing conditions, or drifting conclusions. To address these issues, a representation framework based on rounds is constructed. Contextual continuity is maintained through dialogue structuring and long text segmentation, and cross-round information is accumulated using global contextual memory to improve consistent understanding of key entities, constraint updates, and discussion flow. Building upon this, an explicit scoring and gating mechanism oriented towards decision points is introduced to focus on key rounds that carry out conclusion confirmation, solution selection, and action determination. Combined with attention selection, weighted aggregation of evidence fragments is achieved, making the summary output more focused on the complete organization of decision elements such as conclusions, conditions, and action items. To enhance interpretability and auditability, a retrospective evaluation approach is further adopted. This approach establishes a correspondence between the summary content and original evidence fragments, characterizing the concentration and dispersion of evidence alignment. The contribution distribution between summary sentences and dialogue rounds is then visualized. Comparative evaluation shows that this approach is competitive in overall metrics, helping to more stably aggregate and compress decision-related information in long meeting contexts, and improving the usability of decision-oriented summaries in organizational collaboration and knowledge accumulation.

Keywords: Long dialogue summarization, evidence alignment, gated memory, and decision element extraction.

1. Introduction

In an era where dialogue and meetings have become core vehicles for organizational collaboration and knowledge production, large language models are widely used for recording, retrieval, and decision support[1,2]. However, real-world meetings often span long periods, with frequent topic shifts, multiple iterations of viewpoints, and information being disseminated among different speakers. Key conclusions are often formed implicitly and gradually. If a model lacks robust modeling capabilities for long-range dependencies, it is prone to inconsistencies, omissions of key constraints, and confusion regarding responsibility, thereby weakening the credible transformation and usability of meeting information[3]. Therefore, research on long-range dependency modeling in dialogue and meeting scenarios is not only crucial for the model's consistency in understanding complex interactive contexts but also directly impacts the reliability and efficiency of organizational knowledge management and decision support.

The core challenge of dialogue and meeting texts lies in the fact that the importance of information does not simply correspond to its location or intensity of expression. Many key decision points are scattered throughout processes of clarification, questioning, concession, and confirmation, often accompanied by dynamic updates of conditions, exceptions, and back-to-do items[4]. Traditional summarization often leans towards thematic generalization, failing to accurately capture the turning points and constraints in the decision-making chain, leading to missing action items, misjudgments of priorities, or dilution of risk points. To clearly present the unique needs of this type of scenario, this paper cites Table 1 in the main text to summarize the typical challenges and impacts of long-range dependencies and decision summarization in dialogue and meeting scenarios, so as to highlight the practical urgency and systematic nature of the research problem.

Table 1. Overview of Long-Range Dependency and Decision Summarization Challenges in Dialogue and Meeting Scenarios

Key Challenge	Typical Manifestations	Impact on Summarization and Decision Support
Cross-turn dependency and coreference resolution	Frequent pronouns and ellipses; key entities become clear only after many turns.	Easy to confuse subjects and responsibilities, leading to incorrect attribution in conclusions
Topic shifts and parallel threads	Multi-thread progression with frequent	Loose summary structure; key branches may be omitted or

	interruptions and backtracking	mistakenly merged/misaligned
Gradual decision formation	Conclusions converge progressively; confirmations and negations alternate	Hard to locate the true decision point; outputs become unstable or internally inconsistent
Dynamic updates of conditions and constraints	Constraints such as budget, timeline, and scope are revised multiple times	Action items may miss prerequisites, increasing execution risk
Multi-role stances and conflict coordination	Different roles pursue different goals; compromise processes are implicit	Summaries fail to reflect trade-offs, harming retrospectives and accountability
Sparse and dispersed critical information	Key numbers and conclusions are scattered across multiple fragments	Higher retrieval cost slows understanding and follow-up tracking

The generation of summaries for key decision points essentially involves the structured compression of the evidence and conclusion chains within a long-term context. This ensures that the summaries faithfully reflect the key turning points and constraints of the discussion process while providing an actionable checklist of key points and risk warnings for subsequent implementation. Compared to general summaries, decision point summaries emphasize traceability and actionability, requiring the preservation of causal and conditional integrity while compressing information, clearly identifying who decided what, under what conditions, and what needs to be done next. Therefore, long-term dependency modeling and decision point summary generation are not two separate tasks but rather a mutually supportive whole. The former provides a foundation for consistent understanding across rounds, while the latter provides direct value for organizational collaboration and governance[5].

From an application perspective, accurate long-term dependency modeling can improve the credibility and consistency of meeting minutes, reduce manual review costs, and minimize communication rework caused by misinterpretation[6,7]. High-quality summaries of key decision points help transform fragmented dialogues into searchable, traceable, and auditable decision assets, supporting key aspects such as project management, compliance documentation, knowledge accumulation, and risk control. More importantly, in organizational environments characterized by high-frequency collaboration and rapid iteration, models that can reliably identify and express key decision points will become the crucial hub connecting discussion and execution, propelling dialogue systems from information generation to decision support and action coordination, thus demonstrating clear research value and practical significance.

2. Datasets and Dataset Preprocessing

2.1 Dataset

This study uses QMSum as the dataset. QMSum is an open-source annotated dataset for long conference texts, built around long-range contextual understanding and summary generation of conference dialogues. It provides complete transcriptions of conference content from multiple domains, along with manually written query-summary pairs, requiring the model to locate and integrate key information within long dialogues to generate the target summary. The dataset comprises 232 conferences and contains 1808 query-summary pairs, covering long rounds of discussion and the need for cross-paragraph information association.

QMSum aligns well with the themes of long-range dependency modeling and key decision point summary generation for large language models in dialogue and meeting scenarios presented in this paper. Its task inherently requires maintaining consistent contextual memory and referential resolution within long dialogues, and compressing key information scattered across multiple rounds of discussion into usable summaries. This provides a stable data foundation for identifying key decision points such as discussion transitions, constraint updates, and conclusion confirmations. Furthermore, QMSum's open-source release facilitates reproduction and expansion to more granular structured decision point summary settings. For example, queries can be focused on decision items and action items, generating decision point summaries that include conclusions, conditions, and pending tasks.

2.2 Dataset Preprocessing

(1) Data Cleaning and Format Standardization

First, the original conference transcript is cleaned to remove obvious noise markers and invalid segments, such as repeated system prompts, meaningless placeholders, consecutive blank lines, and abnormal characters. Then, encoding and punctuation are standardized, and transcript fields from different sources are mapped to a consistent structured format, including at least optional information such as dialogue round index, speaker identifier, speech text, and timestamp, ensuring stable alignment to specific rounds and contextual positions in subsequent processing.

(2) Dialogue Structuring and Long Text Segmentation Strategies

The conference transcript is organized into a chronological sequence of rounds, preserving speaker boundaries to maintain the interactive structure of multi-party dialogue. For long conference texts, sliding windows or segmented splicing are used for segmentation and reorganization, ensuring that the model input length is controlled while preserving cross-segment dependency information as much as possible. Segmentation prioritizes breaking at natural boundaries, such as topic transitions, moderator summaries, agenda changes, or long pauses, and the start and end round range of each segment in the original text is recorded for easy retrieval and location of summary results.

(3) Construction of Key Decision Points and Summary Objectives

To highlight the research objectives of key decision point summary generation, standardization is implemented on the summary objective side: reference summaries are uniformly converted into decision-oriented expressions, including decision conclusions, triggering grounds, or discussion focus, constraints, and clear action item information as much as possible. If the dataset provides both query and summary pairs, the query semantics are categorized and filtered, prioritizing

query types related to decision matters, and establishing a soft alignment relationship between queries and source dialogue segments. This allows the model to focus more on the location and compression of decision points during training and inference, rather than simply generalizing themes.

(4) Label Consistency Check and Sample Quality Control

Consistency checks are performed on the cleaned and constructed samples, focusing on four types of issues: whether dialogue turns are missing or out of order, whether speaker labels are confused, whether the summary is semantically consistent with the corresponding segment, and whether there are obvious omissions or conflicts in decision elements. Low-quality samples are subject to elimination or correction rules, such as deleting extremely sparse segments, merging excessively short and semantically incomplete turns, and filtering out irrelevant chatter segments. The final output is a standard sample unit for modeling. Each sample contains a normalized dialogue fragment, optional query information, and a corresponding decision point summary target, thus providing a clean and stable data foundation for long-range dependency modeling and decision summary generation.

3. Method

Long-range dependency modeling for dialogue and meeting scenarios hinges on enabling the model to continuously retain topic cues, entity references, and constraint updates across extended rounds, accurately pinpointing key decision points upon discussion convergence, and then compressing these decision points into actionable summary outputs. To this end, this paper employs a hierarchical representation and memory mechanism: first, the meeting transcript is organized into a sequence by rounds; then, each round is encoded into a semantic vector; and long-term information is aggregated through a global memory vector spanning multiple rounds, allowing the model to maintain a consistent contextual state during topic switching and backtracking. The overall model architecture is shown in Figure 1.

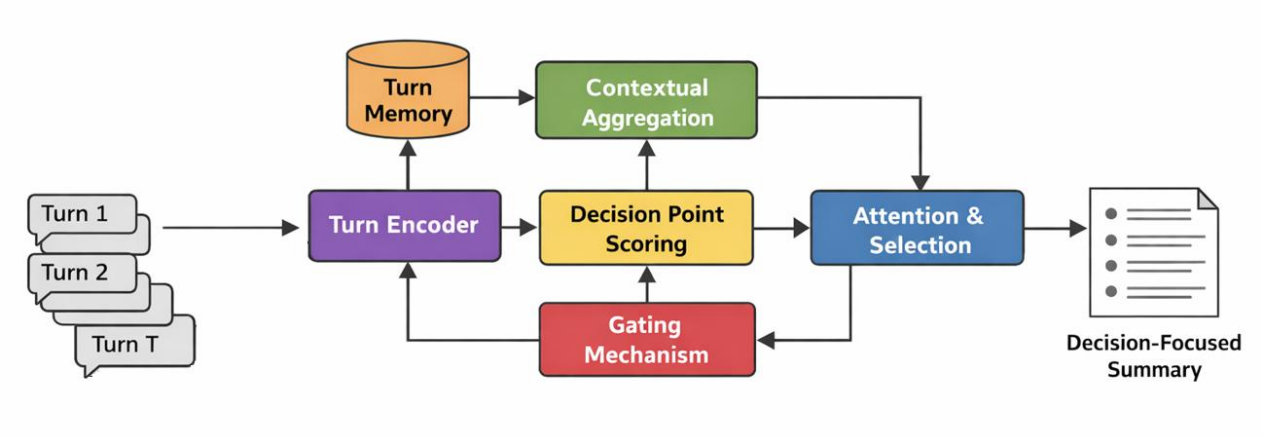

Figure 1. Overall model architecture diagram

Simultaneously, to avoid attention drift caused by information sparsity in long texts, the model introduces explicit scoring of decision points on top of the global memory, prioritizing the use of high-confidence decision fragments in subsequent summary generation stages, thereby preserving key elements such as conclusions, conditions, and actions during information compression.

$$h_t = f(x_t)$$

$$m_t = \alpha m_{t-1} + (1-\alpha)h_t$$

Where x_t represents the text input in round t, h_t is the semantic representation of that round, m_t is the global memory vector accumulated across rounds, and α is the memory retention coefficient. Based on this global memory, the model calculates its relevance to the current global context in each round, which is used to construct attention weights that are more sensitive to key decision points, so that rounds that are cited multiple times, confirmed, or carry constraint updates have a higher contribution in subsequent summarization stages.

$$a_t = \frac{exp(h_t^T m_T)}{\sum_{k=1}^{T} exp(h_k^T m_T)}$$

In terms of key decision point localization, this paper treats decision points as event-driven semantic peaks, typically corresponding to sets of statements such as conclusion confirmation, option selection, scope locking, or action item determination. The model outputs a decision score for each round, using this score as a filtering or weighting signal to compress long dialogues into a set of decision candidates. To maintain method simplicity and controllability, the decision score employs a linear mapping plus activation approach, while a soft threshold gating is introduced to suppress the influence of low-confidence rounds within a small range, thereby reducing interference from idle chatter or repetitive information on the decision summary.

$$p_t = \sigma(w^T h_t)$$

$$g_t = 1(p_t > \tau)$$

Where p_t represents the probability that round t belongs to a key decision point, $\sigma(\cdot)$ is the sigmoid function, w is a learnable parameter, τ is the threshold, and g_t is the gating indicator signal. Based on the gated round set, the summary generation module aggregates the selected representations and generates decision-oriented summary text. The training objective uses generative negative log-likelihood, and round attention and decision probability are used to softly weight the source information, making the model more inclined to cite decision-related content during generation, and output more focused conclusions and action points.

$$L = -\sum_{i=1}^{N} log P(y_i / y_{<i}, \sum_{t=1}^{T} a_t g_t h_t)$$

Here, y_i represents the i-th generated token of the summary sequence, and $\sum_{t=1}^{T} a_t g_t h_t$ is the context representation that integrates long-range dependency attention and decision gating. The overall approach ensures consistent understanding across rounds through hierarchical memory, achieves explicit focus on key decision points through decision scoring and gating, and then completes the generation of decision point summaries with a concise generation objective. Thus, it balances the ability to model long-range dependencies with the usability of decision information compression in dialogue and meeting scenarios.

4. Experimental Results and Analysis

4.1 Experimental setup

This study configured the model training and inference process within a unified hardware and software environment to ensure the stability and reproducibility of long text processing in dialogues and conferences. The hardware platform employed a single GPU for computing, coupled with a multi-core CPU and sufficient memory to support long-sequence batch processing and data loading. The software environment utilized mainstream deep learning frameworks and CUDA acceleration, including fixed random seeds, deterministic operator switches, and logging settings to reduce training fluctuations and facilitate subsequent reproduction of the experimental process and troubleshooting. The experimental setup is shown in Table 2.

Table 2: Experimental Setup: Hardware/Software Environment and Key Hyperparameter Configuration

Category	Item	Setting
Hardware	CPU	16 cores
Hardware	Memory	64 GB
Software	Operating System	Ubuntu 22.04 LTS
Software	Python	3.10
Software	Deep learning framework	PyTorch 2.2
Training	Batch size	8
Training	Optimizer	AdamW
Training	Learning rate	3e-4
Training	Weight decay	0.01
Training	Warmup ratio	0.10
Training	LR scheduler	Cosine annealing
Training	Dropout	0.10
Training	Gradient clipping	1.0
Training	Max epochs	30
Input	Max sequence length	2048
Inference	Decoding	Beam = 4
Inference	Temperature	0.3

In terms of hyperparameter settings, this paper uses AdamW for optimization, with learning rate and weight decay used to balance convergence speed and generalization ability, and warmup and cosine annealing to improve training stability. To suppress gradient explosion in long sequence training, a gradient pruning threshold is set, and appropriate dropout is used to alleviate overfitting. On the input side, the dialogue is organized by round, and the maximum sequence length is limited. For the excess part, a sliding window slice is used to preserve cross-round dependencies. In the inference stage, a conservative decoding temperature is adopted and combined with beam search or top-p sampling to minimize the fluctuations caused by randomness while ensuring the consistency and readability of the summary.

4.2 Experimental Results and Analysis

To evaluate the overall effectiveness of long-range dependency modeling and key decision point summarization methods for dialogue and meeting scenarios, we selected several representative long dialogue or meeting summarization methods as comparison objects and conducted quantitative comparisons under a unified evaluation index system. The table below shows the index configurations and comparisons of each method on the same public benchmark, used to characterize the differences between different methods in terms of information coverage, key content preservation, and semantic consistency.

Table 3: Comparative experimental results

Method	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore
Mihalcea et al.[8]	16.34	2.74	15.41	0.2552
See et al.[9]	28.71	6.08	25.11	0.4014
Liu et al.[10]	26.51	5.70	24.05	0.4705
Lewis et al.[11]	30.45	7.13	28.33	0.5214
Zhang et al.[12]	29.55	6.97	27.17	0.5173
Zhu et al.[13]	32.31	8.72	28.21	0.5332
Zhang et al.[14]	34.14	9.31	29.50	0.5401
Ours	35.62	9.98	30.47	0.5534

Overall, the results show a clear performance ladder: traditional graph ranking methods, as the baseline, are significantly weaker, while models incorporating neural generation and pre-training are significantly stronger overall. This indicates that in long dialogue and conference summarization tasks, the model's ability to model global context and generate semantics is the decisive factor. Furthermore, the improvements in the ROUGE series and BERTScore are basically synchronized, suggesting that stronger methods not only closely approximate the reference summary in n-gram coverage but also excel in semantic consistency and expression fit. This aligns with the frequent paraphrasing and information recombination characteristic of conference scenarios.

Comparing the differences between methods, pre-trained sequence-to-sequence models have a stable advantage over earlier generative models. ROUGE-1, ROUGE-2, ROUGE-L, and BERTScore all show varying degrees of improvement, reflecting stronger linguistic priors and better cross-sentence integration capabilities. On the other hand, the latter two types of methods, which are more geared towards long inputs and dialogue structures, performed better, especially in ROUGE-2. This usually means that the model is stronger in capturing the combination relationships of key information and phrase-level consistency, which is particularly important for identifying key decision points and their constraints.

Looking at Ours' performance, it achieved the highest scores in all four metrics, indicating that the method simultaneously improved in key information coverage, fine-grained content matching, and semantic consistency. Compared with the current best-performing method, Ours improved ROUGE-1, ROUGE-2, ROUGE-L, and BERTScore by 1.48, 0.67, 0.97, and 0.0133, respectively. While the improvements were not dramatic, they were relatively balanced, often indicating that the improvement did not rely on a single-dimensional bias but rather came from a comprehensive enhancement of long-range dependencies and the selection of decision-related content. This allows the model to capture the core points in long meeting texts while reducing the omission of key information and semantic drift.

To test the model's stability and sensitivity in evidence retrospective evaluation, we further categorized the key gating threshold τ and statistically analyzed the normalized distance distribution of evidence fragments corresponding to the summary statements under different τ values. This figure illustrates the overall trend and fluctuation range of evidence distance as a function of τ, where a smaller distance indicates that the model is more inclined to retrieve supporting evidence closer to the decision-making location, thus reflecting the impact of parameter settings on decision point localization, long-range dependency utilization, and the reliability of evidence alignment.

As shown in Figure 2, the evidence distance exhibits a significant non-linear change with the gating threshold τ: when τ is small, the point cloud and distribution are generally biased to the right, and the mean is also larger. This indicates that overly lenient gating introduces more weakly correlated evidence, making the backtracked evidence fragments more scattered and geographically distant, thus weakening the ability to accurately locate decision points. As τ increases to a medium range, the distribution shrinks to the left, the mean decreases significantly, and the shape becomes more concentrated. This indicates that at this point, the gating effectively filters out noisy rounds, making the model more inclined to select supporting information closer to the decision location, resulting in more compact evidence alignment and closer proximity to key discussion fragments.

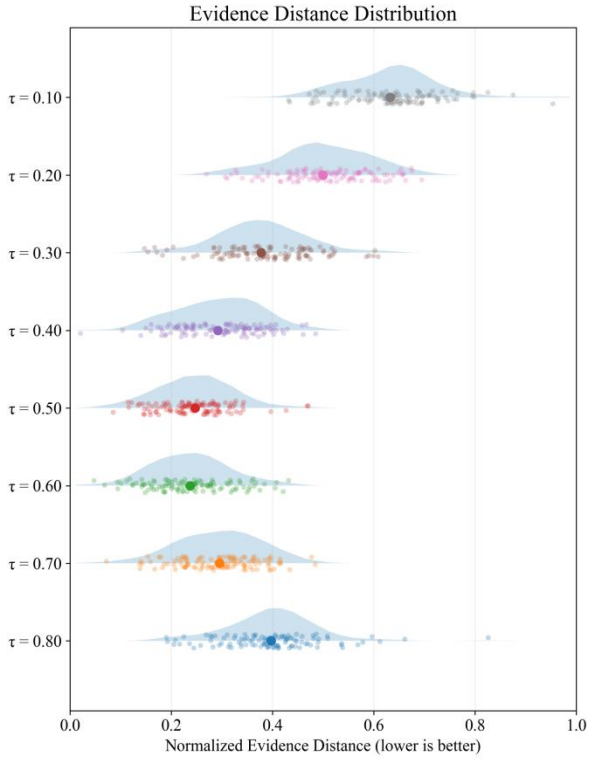

Figure 2. Sensitivity analysis of the gating threshold τ on summary-to-evidence backtracking, shown by the distribution of normalized evidence distance under different τ settings. Lower evidence distance indicates stronger decision-point grounding and more localized evidence alignment, while the spread reflects robustness to threshold changes.

When τ continues to increase, the distribution shifts to the right or widens to some extent, and more obvious outliers are visible. This reflects that overly strict gating leads to insufficient usable evidence, and the model may be forced to rely on a small number of fragments or indirect evidence, resulting in a rebound in evidence distance and a decrease in stability. Overall, these sensitivity results indicate that there is an optimal intermediate working range for the gating threshold, resulting in smaller evidence distances and lower fluctuations. They also suggest that parameter selection needs to strike a balance between covering more candidate evidence and suppressing irrelevant information to ensure that long-term dependency exploitation and evidence alignment reliability are improved simultaneously.

This heatmap expands the summaries at the sentence level and uses the dialogue rounds as the time axis to show the relative contribution intensity of different rounds to each summary sentence. This visualization can simultaneously present the alignment peak location and the degree of concentration or dispersion of the contribution distribution, thus enabling analysis of the evidence aggregation path, key decision point location characteristics, and evidence alignment stability under long-range dependencies. The experimental results are shown in Figure 3.

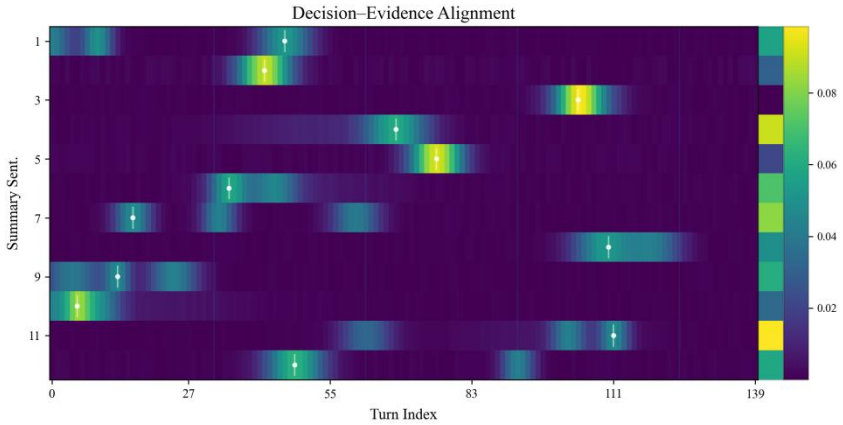

Figure 3. Decision-evidence alignment heatmap showing turn-level contribution patterns for each summary sentence over the full dialogue timeline.

The alignment pattern of the heatmap reveals that most summary sentences exhibit relatively clear high-response regions along the dialogue timeline. The white-marked points often correspond to local contribution peaks, indicating that each summary sentence tends to be dominated by a few key rounds, rather than relying on the entire dialogue on average. This peak-shaped structure typically means that the model can achieve relatively clear evidence localization in long meetings, compressing scattered discussions into more representative evidence segments. Simultaneously, the high-response regions of some summary sentences extend in a band-like pattern, suggesting that these sentences may integrate supplementary information or condition updates across multiple rounds, corresponding to the gradual clarification and multi-round confirmation processes commonly seen in meetings.

Furthermore, the high-response regions of different summary sentences are distributed at different locations along the timeline, showing significant clustering near several segment boundaries. This indicates that the model can form a segmented evidence aggregation path in multi-topic or multi-stage discussions, avoiding mixing information from different stages. The difference in the width of the aligned regions also reflects the dispersion of evidence: narrow and concentrated peaks correspond to more direct evidence chains, while wide and dispersed regions suggest that the summary sentence relies on multiple segments for support, potentially involving trade-offs, constraint supplementation, or action item splitting in the decision-making process. Overall, this visualization helps identify which summary sentences contain the most concentrated evidence and which rely more on cross-round integration, thus providing an intuitive basis for subsequent improvements to gating strategies, attention focus, or evidence screening.

5. Limitations

While this paper proposes a framework for long-range dependency modeling and key decision point summary generation for dialogue and meeting scenarios, certain limitations remain. First, in complex meetings, parallel agendas, implicit consensus, and cross-role positional games are common, and decision points often appear indirectly or through multiple rounds of confirmation. This can lead to ambiguity in decision boundaries and the scope of evidence, increasing uncertainty in decision point localization and element extraction. Second, transcribed texts may contain noise such as colloquialisms, omissions, unclear references, and grammatical errors. Furthermore, the expression habits of meetings across different domains vary significantly. These factors can affect the stable maintenance of long-range dependencies and the quality of information aggregation, potentially causing omissions or deviations in the summary regarding fine-grained conditions, exceptions, and action item details.

Furthermore, the proposed method incorporates explicit memory and gating mechanisms. While this helps focus on key information, it also introduces parameter sensitivity and threshold selection issues: overly loose gating may introduce redundant evidence, while overly strict gating may filter out crucial supplementary rounds for decision formation, thus affecting the completeness of evidence alignment and the executability of the summary. Meanwhile, the assessment of evidence traceability usually relies on the definition and matching rules of evidence fragments. Different annotation granularities or matching standards may lead to fluctuations in the assessment conclusions. In practical applications, it is also necessary to combine more detailed audit requirements and organizational norms to further improve the structured expression and traceability mechanism of decision-making elements in order to enhance the robustness and credibility when deployed across scenarios.

6. Conclusion

This paper addresses the challenges of long-range dependency modeling and key decision point summarization in large language models within dialogue and meeting scenarios. It systematically reviews the challenges posed by real-world meeting texts in terms of time span, topic shifts, ellipsis, and gradual decision formation, and proposes a solution centered on maintaining long-range context and focusing on decision-making. The research emphasizes achieving both a consistent understanding of the context and high-quality compression of decision information within long dialogues. This requires continuously tracking key entities, constraints, and discussion

threads, while also identifying segments that truly influence conclusions under conditions of uneven information density, thereby generating more usable summaries that more closely resemble the decision-making process. This goal aligns with the practical need for meeting minutes to evolve from content recording to decision support and provides a clear direction for building language model capabilities for organizational collaboration scenarios.

Methodologically, this paper treats long-range dependencies as accumulative contextual states and key decision points as semantic events that should be explicitly focused on, forming an end-to-end decision-oriented summary generation framework. This framework not only focuses on the readability of the summary text but also emphasizes the organization of decision conclusions, constraints, and subsequent actions, making the output more closely resemble the debriefing, tracking, and execution in actual workflows. Meanwhile, this paper demonstrates the differences in capabilities of various methods in long meeting summarization tasks through a unified comparative evaluation setting, providing a baseline reference and problem characterization method for subsequent research, and further highlighting the crucial role of long-term context preservation and decision element aggregation in this task.

From an application impact perspective, decision-oriented meeting summarization capabilities can significantly reduce the friction costs of information flow within organizations and improve the traceability and executability of communication results. In enterprise collaboration scenarios, it can be used to automatically generate well-structured decision minutes and to-do lists, assisting project management and cross-team alignment; in compliance and governance scenarios, it helps transform the discussion process into an auditable chain of decision evidence, serving risk control, internal audit, and policy implementation; in high-communication-density fields such as education, healthcare, and public services, decision-oriented summarization can also help to quickly summarize key points and arrange actions, reducing redundant communication caused by misunderstandings. Overall, this research promotes meeting understanding from information compression to decision empowerment, and has direct value for building reliable conversational intelligent assistants and organizational knowledge centers.

Future work can be further deepened in three directions. First, it can enhance the expressive power of decision-making structures at a finer granular level, such as more clearly distinguishing the relationships between conclusions, bases, constraints, and action items, making summaries not only readable but also easier to systematically integrate into task management and knowledge bases. Second, it can strengthen cross-domain and cross-organizational generalization capabilities, building more robust context preservation and evidence aggregation mechanisms for different meeting styles, technical terms, and transcription noise, thereby improving the reliability of real-world deployment. Third, it can explore closer integration with retrieval, knowledge graphs, or process systems, enabling models to continuously align external facts with organizational rules during long meetings and automatically map decision results to subsequent execution links, forming a closed-loop support from discussion to implementation. With the deepening application of dialogue systems in various industries, summary generation for long-range dependencies and key decision points will become an important foundational module for improving organizational efficiency and governance capabilities. This research provides a scalable framework and practical starting point for this direction.

References

[1] Zhong M, Liu Y, Xu Y, et al. Dialoglm: Pre-trained model for long dialogue understanding and summarization[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(10): 11765-11773.

[2] Nedoluzhko A, Singh M, Hledíková M, et al. ELITR minuting corpus: A novel dataset for automatic minuting from multi-party meetings in English and Czech[C]//Proceedings of the Thirteenth Language Resources and Evaluation Conference. 2022: 3174-3182.

[3] Ghosal T, Bojar O, Hledíková M, et al. Overview of the second shared task on automatic minuting (automin) at inlg 2023[C]//Proceedings of the 16th International Natural Language Generation Conference: Generation Challenges. 2023: 138-167.

[4] Y. Xue, "Enhancing Trustworthiness of Retrieval-Augmented Large Language Models via Confidence Calibration and Selective Rejection", 2024.

[5] H. Cui, "Task Graph Modeling and Constraint-Aware Scheduling for Reliable Large Language Model Agent Execution", 2024.

[6] Sun S, Yuan R, Li W, et al. Dialogue acts enhanced extract–abstract framework for meeting summarization[J]. Information Processing & Management, 2024, 61(3): 103635.

[7] J. Li, "LocateNet: Large Multimodal Models for Text-Guided Object Localization", 2024.

[8] Mihalcea R, Tarau P. Textrank: Bringing order into text[C]//Proceedings of the 2004 conference on empirical methods in natural language processing. 2004: 404-411.

[9] See A, Liu P J, Manning C D. Get to the point: Summarization with pointer-generator networks[J]. arXiv preprint arXiv:1704.04368, 2017.

[10] Liu Y. Fine-tune BERT for extractive summarization[J]. arXiv preprint arXiv:1903.10318, 2019.

[11] Lewis M, Liu Y, Goyal N, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. 2020: 7871-7880.

[12] Zhang J, Zhao Y, Saleh M, et al. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization[C]//International conference on machine learning. PMLR, 2020: 11328-11339.

[13] Zhu C, Xu R, Zeng M, et al. A hierarchical network for abstractive meeting summarization with cross-domain pretraining[J]. arXiv preprint arXiv:2004.02016, 2020.

[14] Zhang Y, Ni A, Mao Z, et al. Summn: A multi-stage summarization framework for long input dialogues and documents[C]//Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: long papers). 2022: 1592-1604.