УДК 004.8:004.912

Крикунов Данила Анатольевич, магистрант, Тюменский Индустриальный Университет, г. Тюмень

Машаров Михаил Максимович, магистрант, Тюменский Индустриальный Университет, г. Тюмень

МЕТОДЫ ФОРМИРОВАНИЯ АДАПТИВНОГО КОНТЕКСТА ДЛЯ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В ЗАДАЧАХ СЕМАНТИЧЕСКОЙ СУММАРИЗАЦИИ ДИАЛОГОВЫХ СИСТЕМ

Аннотация

В статье рассматриваются методы формирования адаптивного контекста для больших языковых моделей в задачах семантической суммаризации диалоговых систем. Актуальность исследования обусловлена ростом объема неструктурированных коммуникационных данных и ограничениями современных LLM при обработке длинных транскриптов. Основное внимание уделяется проблемам потери смысловой связности, деградации качества суммаризации при увеличении объема контекста и снижению точности выделения ключевых событий диалога.

Предлагается подход к адаптивному формированию контекста, основанный на предварительной сегментации диалога, semantic chunking, ранжировании реплик и выделении turning points. Рассматриваются методы учета эмоциональной интенсивности, изменения тональности, interruption patterns и semantic relevance при построении prompt context. Отдельное внимание уделяется влиянию prompt engineering на качество semantic summarization и интерпретацию поведенческих сценариев.

Показано, что использование адаптивного контекстного отбора позволяет повысить качество генерации summary, снизить вероятность потери значимых фрагментов коммуникации и улучшить выделение поворотных точек диалога. Практическая значимость исследования связана с применением

предложенного подхода в системах анализа переговоров, клиентских коммуникаций и интеллектуальных диалоговых платформах.

Annotation

The article discusses methods for adaptive context formation for large language models in dialogue system semantic summarization tasks. The relevance of the study is determined by the growing volume of unstructured communication data and the limitations of modern LLMs when processing long transcripts. Particular attention is paid to the problems of semantic coherence loss, degradation of summarization quality with increasing context size and reduced accuracy of turning point extraction in dialogues.

The paper proposes an adaptive context formation approach based on dialogue segmentation, semantic chunking, utterance ranking and turning point extraction. Methods for considering emotional intensity, sentiment changes, interruption patterns and semantic relevance in prompt context construction are examined. Special attention is paid to the influence of prompt engineering on semantic summarization quality and behavioral scenario interpretation.

It is shown that adaptive context selection improves summary generation quality, reduces the probability of losing significant communication fragments and increases the accuracy of dialogue turning point extraction. The practical significance of the study is related to the application of the proposed approach in negotiation analysis systems, client communication analytics and intelligent dialogue platforms.

Ключевые слова: большие языковые модели, семантический анализ диалогов, адаптивное формирование контекста, интеллектуальная обработка текстов, обработка естественного языка, контекстная релевантность, суммаризация диалогов, промпт-инжиниринг

Keywords: large language models, dialogue semantic analysis, adaptive context formation, intelligent text processing, natural language processing, context relevance, dialogue summarization, prompt engineering

Современные системы анализа коммуникаций increasingly ориентируются на использование больших языковых моделей для автоматизированной обработки диалогов, построения semantic summary и выявления ключевых событий взаимодействия. Рост объема цифровых коммуникаций приводит к необходимости обработки длинных транскриптов, содержащих большое количество реплик, эмоциональных переходов и контекстных зависимостей. Исследования в области semantic summarization показывают, что увеличение объема входного контекста приводит к деградации качества генерации и снижению semantic coherence итогового ответа [1]. В связи с этим особую актуальность приобретает задача формирования адаптивного контекста для LLM в системах анализа диалогов.

Одной из ключевых проблем современных больших языковых моделей является ограниченность context window и снижение эффективности обработки длинных последовательностей. При работе с объемными диалогами модель начинает терять ранний контекст, игнорировать отдельные реплики и пропускать важные turning points коммуникации. Под turning points понимаются ключевые изменения эмоционального состояния, смена стратегии общения, конфликтные эпизоды и изменения тональности разговора. Исследования в области dialogue summarization показывают, что потеря подобных событий существенно снижает качество интерпретации коммуникационного поведения [2].

Традиционные методы семантической суммаризации преимущественно используют линейную передачу транскрипта в LLM без предварительного анализа структуры диалога. Подобный подход приводит к возникновению context dilution effect, при котором значимые фрагменты коммуникации теряются среди большого количества низкоприоритетных реплик. Особенно критичной данная проблема становится при анализе переговоров, клиентских коммуникаций и многопользовательских диалогов с высокой эмоциональной динамикой. Современные исследования показывают, что качество

суммаризации напрямую зависит от корректного отбора семантически значимых фрагментов диалога [6].

В рамках исследования рассматривается архитектурный подход к построению adaptive context pipeline для задач semantic summarization. На первом этапе выполняется предварительная обработка transcript, включающая normalization, dialogue segmentation и timestamp alignment. После очистки и структурирования транскрипта система формирует отдельные utterances, содержащие текст реплики, идентификатор спикера и временные интервалы. Подобная структура позволяет перейти от линейного представления диалога к иерархическому контекстному анализу.

Следующим этапом является semantic chunking - разбиение длинного диалога на смысловые сегменты. В отличие от fixed-size chunking, адаптивное разбиение учитывает semantic continuity, topic shift и эмоциональную динамику коммуникации. Для каждого фрагмента вычисляются дополнительные признаки: эмоциональная интенсивность, изменение sentiment score, interruption patterns и semantic relevance. Использование semantic chunking позволяет снизить вероятность потери важных контекстных зависимостей при обработке длинных диалогов [3].

После сегментации выполняется процедура context ranking, предназначенная для ранжирования реплик и смысловых блоков по степени значимости. Приоритет получают фрагменты, содержащие эмоциональные переходы, конфликтные эпизоды и признаки эскалации. Низкоприоритетные сегменты агрегируются или локально суммаризируются, что позволяет уменьшить объем передаваемого контекста без существенной потери смысловой информации. Подобный подход соответствует современным стратегиям adaptive context selection для больших языковых моделей [4].

Особое внимание в исследовании уделяется задаче выделения turning points. Для определения ключевых событий диалога используются признаки эмоциональной интенсивности, semantic shift, interruption rate и изменение коммуникативной стратегии участников. Turning point extraction позволяет

выделять критические эпизоды коммуникации: начало конфликта, смену эмоционального состояния, потерю вовлеченности или изменение отношения собеседника. Исследования в области анализа эмоциональных состояний показывают, что корректное определение turning points существенно повышает качество интерпретации коммуникационного поведения [5].

Важную роль в повышении качества semantic summarization играет prompt engineering. В рамках исследования рассматриваются методы role prompting, hierarchical prompting и instruction-based prompting. Использование структурированных prompt templates позволяет уменьшить вероятность hallucination и повысить consistency generated summaries. Особенно важным является сохранение temporal coherence между отдельными сегментами диалога, поскольку нарушение временной последовательности приводит к искажению логики коммуникации.

В рассматриваемой архитектуре LLM не используется как самостоятельный механизм анализа диалога, а выступает в роли semantic interpretation layer поверх предварительно обработанного pipeline. На вход модели передаются transcript segments, speaker metadata, emotional markers и behavioral signals. Использование структурированного контекста позволяет существенно повысить качество semantic summarization и уменьшить вероятность потери ключевых событий диалога [7].

Таким образом, предложенный подход к формированию адаптивного контекста для больших языковых моделей позволяет повысить устойчивость semantic summarization и улучшить качество выделения turning points в длинных диалогах. Использование semantic chunking, context ranking, prompt engineering и adaptive context selection формирует основу для построения интеллектуальных систем анализа коммуникаций нового поколения.

Литература

1. Голубинский А.Н. Автоматическая генерация аннотаций научных статей на основе больших языковых моделей // Информатика и автоматизация. 2025.

2. Дудихин В.В. Методология использования больших языковых моделей, обучаемых с использованием креативного промтинга, для интеллектуального реферирования контента // Стратегии бизнеса и управления. 2024.
3. Мансур А.М. Разработка чат-бота для классификации и анализа текстов на естественном языке с использованием локальных больших языковых моделей // КиберЛенинка. 2025.
4. Федотова А.М., Романов А.С. Методика идентификации текстов, сгенерированных большими языковыми моделями // Информатика и автоматизация. 2025.
5. Чжэнь Н.Р.Л. Искусственный интеллект и технологии языкового анализа // КиберЛенинка. 2024.
6. Sahoo P. et al. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. 2024.
7. Chen B. et al. Unleashing the Potential of Prompt Engineering in Large Language Models: A Comprehensive Review. 2023.

Literature

1. Golubinskiy A.N., Tolstykh A.A., Tolstykh M.Y. Automatic Generation of Scientific Articles Abstracts Based on Large Language Models // Informatics and Automation. 2025. Vol. 24. No. 1. P. 275–301.
2. Dudikhin V.V. Methodology for Using Large Language Models Trained with Creative Prompting for Intelligent Content Summarization // Business and Management Strategies. 2024.
3. Mansur A.M. Development of a Chatbot for Classification and Analysis of Natural Language Texts Using Local Large Language Models // CyberLeninka. 2025.
4. Fedotova A.M., Romanov A.S. Methodology for Identifying Texts Generated by Large Language Models // Informatics and Automation. 2025.
5. Zhen N.R.L. Artificial Intelligence and Language Analysis Technologies // CyberLeninka. 2024.

6. Sahoo P., Singh A., Saha S., Jain V., Mondal S., Chadha A. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. 2024.
7. Chen B., Zhang Z., Langrené N., Zhu S. Unleashing the Potential of Prompt Engineering in Large Language Models: A Comprehensive Review. 2023.