

Summarcha

Chat Summarizer using Machine Learning

Team members

Full name	Innopolis email	Tasks
Timofey Sedov	t.sedov@innopolis.university	Models research, ML, backend
Grigorii Fil	g.fil@innopolis.university	ML, Dataset search, backend, testing
Andrei Markov	a.markov@innopolis.university	Models training, full-stack, deploy

Project topic

The project is dedicated to the development of a tool for automatic chat summarization. It is based on the [T5](#) model which was finetuned on a diverse dataset comprising various chat logs from different platforms to ensure versatility and accuracy in summarization.

Link to the repository

<https://github.com/markovav-official/summarcha>

Solution building

Solution 1 - base Bart model with prefix tuning

This solution explores the use of the base [BART](#) (Bidirectional and Auto-Regressive Transformers) model, with a focus on prefix tuning. Prefix tuning involves pre-appending a learnable continuous task-specific vector to the input, allowing the model to adapt to summarization tasks with minimal changes to its parameters. This approach is efficient in terms of computation and retains the general knowledge of the pre-trained BART model.

Solution 2 - Bart-Summarize model with Telegram Translation API

In this approach, the BART model is further enhanced with a summarization-specific fine-tuning and integrated with the Telegram Translation API for real-time multi-language support. This allows the model to not only summarize chats but also translate and summarize conversations in various languages, making it highly effective in multilingual chat environments.

Solution 3 - ru-T5 model fine-tuned on the dataset

The ru-T5 model, a Russian variant of the T5 model, is fine-tuned on a curated dataset consisting of Russian language chat logs. This fine-tuning aims to adapt the model to understand and summarize conversations in Russian more accurately. The dataset includes a mix of formal and informal chat logs to ensure the model's effectiveness across different styles of communication.

Final solution and results

Our team's final solution was the development of a custom Telegram client featuring an innovative 'Summarize' button. This feature, powered by the ru-T5 model, allows users to effortlessly summarize chat conversations. It offers flexibility in summarization, enabling users to summarize either a specific number of messages or only the unread messages in a chat. This functionality significantly enhances user convenience by providing quick and efficient access to the gist of lengthy conversations or catching up on missed messages. The integration of this feature into the Telegram Web client has been well-received for its user-friendly interface and practicality, making it a valuable tool for both personal and professional communication within the Russian-speaking community.

Preview

<https://youtu.be/PhqPpuqfKOI>

Project timeline

Week 1-2: Initial research and selection of models (BART, T5). Development and testing of Solution 1.

Week 3-4: Integration of Telegram Translation API and testing of Solution 2.

Week 5-6: Fine-tuning ru-T5 model and testing of Solution 3.

Week 7-8: Choosing the best solution, modifying Telegram Web client.

Contributions

Andrei was mainly responsible for the frontend part of the project. He created a modified Telegram web client which allows users to send messages to the backend and receive the summary. He also created a dataset for the model finetuning.

Grigorii was responsible for the model finetuning. He performed EDA on the dataset and fine-tuned the model on the dataset.

Timofey was responsible for the backend part of the project. He created a backend which receives messages from the frontend and sends them to the model. He also performed research on the model and tested it.

References

- [1] [arXiv:1910.10683](#) [cs.LG]
- [2] dialogsum-ru dataset, <https://huggingface.co/datasets/d0rj/dialogsum-ru>
- [3] [arXiv:1910.13461](#) [cs.CL]