|  |  |
| --- | --- |
| Paper Title: | Toward Unifying Text Segmentation and Long Document Summarization |
| Paper DoI: | <https://preview.aclanthology.org/emnlp-22-ingestion/2022.emnlp-main.8/> |
| Research Question | Can the model designed by the authors which is a combination of text summarizer and section segmentation learn from written documents and transfer the knowledge to spoken transcripts? |
| Related Work | Most summarizers are enabled by transformer-based models that generally process long sequences. Other methods include content-based that builds representations from words to sentences, sentences to paragraphs, and eventually to documents. Some abstractive strategies can produce brief summaries, however, they are prone to factual errors which can mislead readers.  Kosorek et al.(2018) worked on designing the neural text segmentation model that predicts section boundaries. Lukasik et al. (2020) compared three architectures based on Transformers on a Wikipedia dataset. As both summarization and segmentations are done separately by the previous researchers, the authors of this paper have taken inspiration from them and worked in a combined framework of summarization and segmentation. |
| Experiment Design | **Approach:**  The authors have designed a system named “Lodoss” (Long document summarization with segmentation). It performs summarization and segmentation simultaneously. This system has a new regularizer drawing on determinantal point processes to measure the quality of all sentences in the summary ensuring that the summary is informative and sentences are not repeated.  They employ the Lonformer model from (Beltagy et al. 2020) with dilated window attention which allow each token to attend only to its local window. The summarizer is built on top of Longformer by stacking two layers of inter-sentence Transformers to it.  **Experiments:**   1. Dataset: Scientific articles from open-access repositories (arXiv.org, PubMed.com) as they follow logical document structure. For transcripts, lectures from VideoLectures.NET were used. 2. Experimental Settings: They used HuggingFace, PyTorch and PyTorch Lightning. Adam optimizer was also used. The training was performed on 8 NVIDIA Tesla P40GPUs.  * Learning rate: 3e-5 * Batch size = 8 * Epochs = 20 |
| Result | **Summarization Results**  **Results on Scientific Papers:**   * The authors compared three model variants: * Lodoss-base, using Lsum * Lodoss-joint, using Lsum + Lseg * Lodoss-full, using (Lsum + Lseg) + *βLDPP* * Their model strongly outperformed both extractive and abstractive baselines which suggested that the unification of summarization and segmentation is effective.   **Results on Lecture Transcripts:**   * They trained their model from scratch using the transcripts. * They observed that the model pretrained on written documents substantially better than training it from scratch. Additionally, they found that the knowledge gained from summarizing written document could be transferred to summarization of spoken transcripts. Hence, proving their hypothesis. |