

Generating short description of videos

Team members

Karthikk Tamil mani, mani, B00838575, kr630601@dal.ca

Srikrishna Sasidharan, sasidharan, B00835818, sr950863@dal.ca

Thanigaiselvan Senthil Shanmugam, tshanmugam, B00798284, th876217@dal.ca

Muthukumar Rajendran, mrajendran, B00835091, mt911642@dal.ca

Problem statement

The idea of this project is to create a short description of any videos available on the Internet. The range of videos can widely vary or change. By creating a short description, the user or the person who's willing to view the video can go through the gist or short description of the video and can know whether they are relevant or not rather than going through the entire video and looking for specific texts. Abstractive type of text summarization is used to generate descriptions. In this way, the model is not entirely dependent upon the video's vocabulary.

Possible approaches

The initial process involves converting the audio data into a corresponding text. Various approaches explored to achieve the desired result.

1. Google's speech to text API was used to convert the audio data by running a program in the browser. However, the results obtained were not complete and had some lost data.
2. We used the voice typing feature in the google docs, which helped us to achieve the complete result. The drawback of this method is that it is time-consuming as it requires the whole video to be played from the beginning. Another issue is that the output obtained does not contain any punctuation.
3. To overcome the time-consuming issue, we tried to retrieve the subtitles that are generated by Youtube. This approach was successful in overcoming the time-consuming issue. However, the earlier issue of punctuation not available still exists in the achieved result.

So, we found a dataset that has information about various news articles with the punctuations. We would like to proceed with the current dataset to build our model. If possible, we will try to find a way to embed the punctuation in the video output as well. Once the text is obtained from the video, we will be using any one of the following approaches to summarise:

1. One method is to use the Document context vector and RNNs to generate summaries. This method uses document-context based Seq2Seq learning rather than the traditional Seq2Seq along with RNN's. The neural net is trained with the already available human-generated golden-standards of summaries, and in this way, evaluation is done. Using the document-context Seq2Seq model, better human-like context-based summaries are generated rather than the usual generic summaries. In this way, the user can better understand the context of the video or the text. Also, this model uses semi-supervised learning[1].

Generating short description of videos

2. The approach is to combine deep learning techniques with semantic data transformation to generate a short summary of a text. Deep learning encoder and decoder are used in addition to the already known semantic data transformation to generate a summary retaining the essential aspects of the original text. This method addresses the problem of out-of-vocabulary or rare words[2].
3. The approach is based on a model that uses the abstractive text summarization technique. In this technique, Attentional Encoder-Decoder Recurrent Neural Networks is used to achieve better results for corpora [3].
4. Using skip-thought vectors and word embeddings to predict the next word and the next sentences using encoder and decoder RNN's. Once this is done, we will cluster the sentences that are closest to the main context. To achieve unsupervised clustering, K-means clustering will be used to obtain abstractive summarization[4].

Process For the rest of the term:

1. Week 1: Fetching the required dataset - Until the end of this week

As of now, we do have the BBC dataset that contains the transcriptions and the short descriptions of the videos. Still, we would be searching for other datasets until the end of this week.

2. Week 2: Preprocessing.

3. Parallely, a couple of people will be working on finalizing the approach to be used

4. Week 3 -end of term: Roughly we are expecting 2 weeks to build the Once done, we will evaluate the model

References:

- [1] C. Khatri, G. Singh, and N. Parikh, "Abstractive and Extractive Text Summarization using Document Context Vector and Recurrent Neural Networks," *arXiv.org*, 2018. [Online]. Available: <https://arxiv.org/abs/1807.08000v2>. [Accessed: 03-Mar-2020].
- [2] P. Kouris, G. Alexandridis, and A. Stafylopatis, "Abstractive Text Summarization Based on Deep Learning and Semantic Content Generalization," *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.
- [3] R. Nallapati, B. Zhou, Santos Cicero, C. Gulcehre, and B. Xiang, "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond," *arXiv.org*, 2016. [Online]. Available: <https://arxiv.org/abs/1602.06023>. [Accessed: 03-Mar-2020].
- [4] Madhav Thaker, "Comparing Text Summarization Techniques," *Medium*, 25-Mar-2019. [Online]. Available: <https://towardsdatascience.com/comparing-text-summarization-techniques-d1e2e465584e>. [Accessed: 03-Mar-2020].