An Analysis of Character Representation from Print (Book) vs Visual (Movie) Media

M Saydur Rahman

Abstract

Our paper is about An Analysis of Character Representation from Print(Book) vs Visual(Film) Media. Our goal is Given an input video to select a bunch of frames to create a summary video that can capture the vital information of the input video. With many books getting adapted as movies, video summarization provides a useful tool that assists video analysis. In this paper, we formulate a video movie summarization to determine the importance of any character in movies with their book counterparts. We propose a fully convolutional sequence network model classifier to solve the video summarization problem, a search model to analyze book representation, and statistical analysis for comparison.

Introduction

How much a movie adaptation of a book gives characters true representation. We are trying to find the answer to that question via automation via the use of neural networks. We are going to analyze the video via our neural network while trying to collect statistical data. The text analysis will provide us data on a character representation from the book. We plan on using statistical analysis to find out the representation true to the book author's intention. More than 80% of internet traffic will be real data except in particular, it could be video data[3]. It gives us a certain intuitive sense of how much video data there and how many tools to manage the massive amount of video data. Video recorded in one day could have 5 minutes of video summarization instead of 24 hours of video. We can use video

summarization for thumbnail generation, if we put a mouse around the video on youtube, we will see that the thumbnail option for the summary of the video, which includes the main video is for. A fair bit of the summarization system can automatically generate those thumbnails. In video Customization, if we put a different video on social media on another platform, they show they have a limitation on how long that video can be. Somehow we need to adjust those accordingly. We can use Video Summarization for a better summary In highlight creation compare it with our text analyzer Before we watch any movie based on our favorite book.

Discussion

We try to solve the summary using the labeling problem. Our first approach will be before a fully convolutional sequence network for video summarization[1,6,7]. First, we try to understand how to extract frames from a raw video. Then we tried to drag our image by choice. We want to have machine learning to produce a sequence of 0 or 1 labeling on this video[1,6,7]. If the frame is labeled 0 that means that frame is not selected as part of the summary and marked as 1 that is chosen as part of the resume. Deep learning another model called a convolutional neural network. CNN, known to be computationally efficient because the computation involved model is easily findable. Another problem in computer vision is called semantic segmentation. In semantic segmentation, we have an input image[5]. Want to produce pixel leveling of that image. For every pixel, we are going to assign an integer, and the integer value of that pixel is going to tell us what object class that pixel is, whether it's people or whether it's the horse. Once we have this kind of pixel labeling, we will get semantic labeling of the image, which tells us what the object is, basically the item for the pixel[5]. Semantic segmentation is a very similar problem because of the video summarization of the sequence of frames[6,7]. Try labeling every frame rather than labeling every pixel. Once we are marking, we can gather the video summarization result. Semantic segmentation, we label every pixel. Semantic segmentation makes some

sense in two-dimensional data because the image has the width and height means a two-dimensional matrix[5]. Video summarization, we have one-dimensional data because it's thinner, that's we have a sequence of frames[6,7]. In Fully Convolutional Sequence Network(FCSN), Key subject is, we will preprocess the whole video[6,7]. We can use the original image for pre-processing the frame. Every frame represents a num vector. We imagine a 1D image, which is a much higher number of channel dimensions. We can represent 1*T*D tensor, where T= length of the video and D= dimension frame feature that we can extract the part of each frame. Each of the structures represented by a dimensional vector. We use a 1D image because we import a 1D image. Layer in the encoder 1D convolution operation. 1d convolutional network has a sliding window. This window extract features from those five frames that it covered. Slide access the time dimension. Every time dimension, we have to extract a part of a different location in a video. End of the encoder, we can look at some feature maps, and the length of the feature is shorter than the size of the original video. The decoder state, 1D convolutional, and 1D un-polling where 2D operation replaces by the 1D process. We also skip connecting the encoder layer directly to the other layer in the decoder layer. Finally, we got the labeling frame.

In our project, we analyze story character representation from a book and then we compare that with the same character representation in visual media. So for text analysis, first we use the word frequency technique. Word frequency is a text analysis technique that measures the most frequently occurring words or concepts in a given text using the numerical statistic TF-IDF (Term Frequency-Inverse Document Frequency). For text analysis, we tried to use the CNN model because we found some research papers where they prove that for text analysis CNN is appropriate. We faced some problems when we tried to analyze text character frequency by character name. In the story a character name was used one time in a situation then the author used a pronoun to represent that character. So when we tried to analyze the character timing, we faced problems for that reason. We are trying to solve the problem.

Conclusions

The movie vs book representation analyzer can be a bais free way of rating media contents. In the future, the system can be improved to analyze multi-part books to film adaptations with improvement even tv series can also be analyzed.

The movie analyzing can be improved using character time on screen (neural network based facial recognition) with improved accuracy but at a higher workload.

The book representation can be improved with reference recognition (how many times a character comes up in conversations).

Adding audio-based character recognition can also be accurate in finding out character representation in movies and books in the future to improve.

References

[1] Video Summarization Using Fully ConvolutionalSequence Networks. https://openaccess.thecvf.com/content_ECCV_2018/papers/Mrigank_Rochan_Video_Summarization_Using_ECCV_2018_paper.pdf

[2] Learning video summarization from unpaired data. https://openaccess.thecvf.com/content_CVPR_2019/papers/Rochan_Video_Summarization_by_Learning_From_Unpaired_Data_CVPR_2019_paper.pdf

[3] Cisco visual networking index: Forecast and methodology, 2016-2021. https://www.cisco.com/.

[4] Open video project.https://open-video.org/

[5] DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.https://ieeexplore.ieee.org/document/7913730.

[6] An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.https://arxiv.org/abs/1803.01271.

[7] Diverse Sequential Subset Selection for Supervised Video Summarization. https://papers.nips.cc/paper/5413-diverse-sequential-subset-selection-for-supervised-video-summarization.

[8]Statistical Methods for Comparison of Data Sets of Construction Methods and Building Evaluation https://www.sciencedirect.com/science/article/pii/S1877042813029686/pdf?md5=2ee7f259710d2a331e4a54e73e979d69&pid=1-s2.0-S1877042813029686-main.pdf