Book2movie: Aligning Video Scenes With Book Chapters

Makarand Tapaswi, Martin Bauml, Rainer Stiefelhagen, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1827-1835

Abstract

Film adaptations of novels often visually display in a few shots what is described in many pages of the source novel. In this paper we present a new problem: to align book chapters with video scenes. Such an alignment facilitates finding differences between the adaptation and the original source, and also acts as a basis for deriving rich descriptions from the novel for the video clips. We propose an efficient method to compute an alignment between book chapters and video scenes using matching dialogs and character identities as cues. A major consideration is to allow the alignment to be non-sequential. Our suggested shortest path based approach deals with the non-sequential alignments and can be used to determine whether a video scene was part of the original book. We create a new data set involving two popular novel-to-film adaptations with widely varying properties and compare our method against other text-to-video alignment baselines. Using the alignment, we present a qualitative analysis of describing the video through rich narratives obtained from the novel.

Related Material

[pdf]  
[bibtex]

@InProceedings{Tapaswi_2015_CVPR,  
author = {Tapaswi, Makarand and Bauml, Martin and Stiefelhagen, Rainer},  
title = {Book2Movie: Aligning Video Scenes With Book Chapters},  
booktitle = {The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)},  
month = {June},  
year = {2015}}

To align movies and books we exploit a neural sentence embedding that is trained in an unsupervised way from a large corpus of books, as well as a video-text neural embedding for computing similarities between movie clips and sentences in the book. We propose a context-aware CNN to combine information from multiple sources. We demonstrate good quantitative performance for movie/book alignment and show several qualitative examples that showcase the diversity of tasks our model can be used for. In this paper we present a new problem: to align book chapters with video scenes. Such an alignment facilitates finding differences between the adaptation and the original source, and also acts as a basis for deriving rich descriptions from the novel for the video clips.