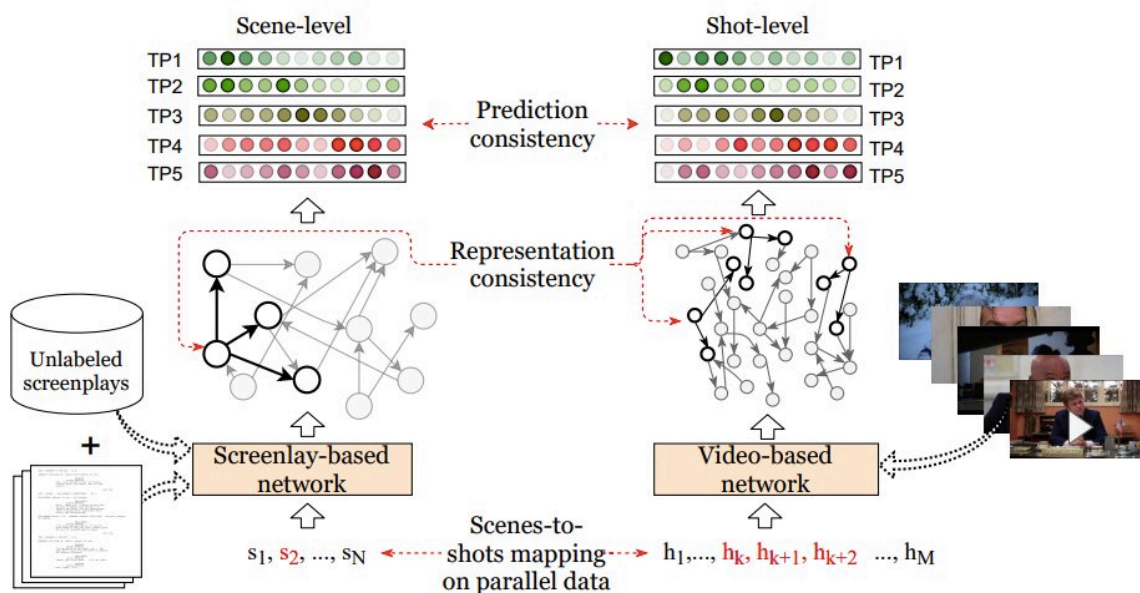# A new model that automatically generates movie trailers

November 29 2021, by Ingrid Fadelli

Overview of the architecture created by the researchers. Two networks process different views of the movie with different degrees of granularity. The video-based network takes as input multimodal fine-grained shot representations based on the movie's video stream. The screenplay-based network processes textual scene representations which are coarse-grained and based on the movie's screenplay. The networks are trained jointly on TP identification with losses enforcing prediction and representation consistency between them. Credit: Papalampidi, Keller & Lapata.

Trailers, short video clips that introduce new movies, are often crucial elements in the promotional strategies employed by film production companies. To be most effective, trailers should briefly summarize a movie's plot, conveying its artistic style and overall mood in appealing ways.

So far, movie trailers have been primarily created by humans. Recently, however, some computer scientists started exploring the possibility that these promotional clips could also be automatically generated by machines.

Researchers at University of Edinburgh developed an artificial neural network-based model that can automatically generate film trailers. This model, presented in a paper pre-published on arXiv, is based on an unsupervised, graph-based machine-learning algorithm.

To best tackle the task of automatic film trailer generation, the researchers decomposed it into two sub-tasks, namely the identification of the movie's narrative structure and the prediction of the sentiment (i.e., mood and feeling) conveyed by it. The technique they created thus processes both parts of the movie (i.e., videos) and text extracts from a movie's screenplay.

"We model movies as graphs, where nodes are shots and edges denote semantic relations between them," Pinelopi Papalamidi, Frank Keller and Mirella Lapata, the three researchers who carried out the study, wrote in their paper. "We learn these relations using joint contrastive training, which leverages privileged textual information (e.g., characters, actions, situations), from screenplays. An unsupervised algorithm then traverses the graph and generates trailers."

Essentially, the film trailer generation method they created is composed of two neural networks. While one of these networks processes

multimodal shot representations derived from the movie's video stream, the other analyzes textual scene representations that are based on the movie's screenplay.

Combined, the two neural networks can identify turning points in the movie, which are parts of the movie that are particularly salient and that should be featured in trailers. Turning points in movies typically include an opportunity, a change of plan, the point of no return, a major setback and a climax.

Papalampidi, Keller and Lapata evaluated their technique for generating film trailers in a series of tests. Remarkably, they found that it could identify turning points in movies with a significantly greater accuracy than other baseline methods for the generation of film trailers.

In addition, the researchers used their model to create trailers for 41 different movies. They then compared the quality of the trailers it produced to that of trailers generated by techniques trained with supervised learning by asking human viewers recruited on Amazon Mechanical Turk (AMT) which ones they preferred. Interestingly, most of the respondents preferred the trailers created by their technique to those produced by supervised models.

While the model created by Papalampidi, Keller and Lapata might not yet create perfect trailers, it could eventually be used by film production companies to facilitate and speed up the production of trailers. Meanwhile, the team plans to continue working on their technique, to improve the quality of the trailers it produces further.

"In the future, we would like to focus on methods for predicting fine-grained emotions (e.g., grief, loathing, terror, joy) in movies," the researchers added in their paper. "In this work, we consider positive/negative sentiment as a stand-in for emotions, due to the

absence of in-domain labeled datasets. Avenues for future work include new emotion datasets for movies, as well as emotion detection models based on textual and audiovisual cues."

More information: Pinelopi Papalampidi, Frank Keller, Mirella Lapata, Film trailer generation via task decomposition. arXiv:2111.08774v1 [cs.CV], arxiv.org/abs/2111.08774

© 2021 Science X Network