A team of researchers at the University of Virginia has recently carried out a large-scale analysis aimed at identifying features in film trailers that best predict a movie's genre and estimated budget. In their study, outlined in a paper pre-published on arXiv, the researchers specifically compared the effectiveness of visual, audio, text, and metadata-based features.

"Video understanding is the next frontier after image understanding," Vicente Ordonez, one of the researchers who carried out the study, told TechXplore. "However, much work on video understanding has so far focused on short clips with a human performing a single action. We wanted something longer, but there is also the issue of computational power. Video trailers seemed like an intermediate compromise, as they display a multitude of things, from scary to funny."

Movie trailers are short and can easily be paired with movie descriptions. Ordonez and his colleagues realized that these characteristics make them ideal to investigate parallels between video and language.

"We found that combining all the modalities (i.e. video, text, audio and metadata), we were able to gather valuable insights on expected correlations between specific genres and a particular modality, for example, that visual features are more valuable when predicting a movie as animated or not," Paola Cascante-Bonilla, another researcher involved in the study, told TechXplore. "Moreover, we found that including the audio in our experiments significantly boosts the genre prediction performance in comparison to only using the video, text and metadata."

The researchers observed that while analyzing movie posters led to unsatisfactory results, focusing on all movie features presented in a trailer (i.e. video, text, audio and metadata) led to significant improvements. These findings are particularly noteworthy, as they could help to develop more effective tools to analyze movies and serve as a basis for future research studies.

Interestingly, when focusing on video, text and audio data extracted from trailers, Ordonez, Cascante-Bonilla and their colleagues were able to estimate a movie's genre with an accuracy comparable to that achieved by analyzing the movie's metadata (i.e. information about its actors, director, etc.). The techniques used by the researchers in their study, which combine different features/modalities, could therefore be used to...
analyze a wider range of movies.

In their study, the team also introduced a new dataset for training and evaluating tools to analyze movies. This dataset, called Moviescope, includes 5,000 movies, along with their corresponding trailers, movie posters, movie plots and associated metadata.

“Our findings suggest that just a movie’s textual summary is not enough to differentiate between an animated movie and a movie of another genre,” said Siva Sivaraman, another researcher involved in the study who now works at Microsoft. “You need to 'see' the trailer to be able to decide if a given movie is animated or not. The modal attention technique we used allows us to identify and analyze the features that the model pays closer attention to when predicting a particular genre. As we predicted, the model learns to weigh the visual feature over other features while making predictions for the animation genre.”

The findings gathered by this team of researchers could have important implications both for the analysis of movies and for movie advertising. In the future, other research groups could use these observations to develop more effective tools for predicting specific aspects of movies. In addition, the techniques used by the Ordonez and his colleagues could inform the advertising industry on how to create more impactful trailers.

"We are now planning to use movie plots and posters to analyze the way movies are advertised and make recommendations about maximizing the effectiveness of movie advertising from both the perspective of consumers and distributors," Ordonez said.

https://arxiv.org/abs/1908.03180

© 2019 Science X Network