November 2019

Detection of salient video segments and associated comments via reinforcement learning

Aditya Shah

Jared M. Zimmerman

Detection of salient video segments and associated comments via reinforcement learning

ABSTRACT

The application of computer vision and techniques that rely on machine intelligence has improved automated determination of notable segments within video content. However, further refinement in these techniques can improve the quality and accuracy of the marked segments. Also, it can be difficult for automated techniques to surface the most notable text comments associated with video content. This disclosure describes techniques that combines initial independent automated determination of notable video segments and text comments to refine and improve each aspect based on the other. The techniques leverage audience engagement with live and pre-recorded video content as captured via user comments entered while the video is being viewed. Machine learning models that identify salient video segments can be trained using the output of models that identify notable comments and vice versa, using reinforcement learning.

KEYWORDS

- Key moment
- Salient moment
- Video segment
- Comment feed
- Time-coded comments
- Sentiment detection
- Temporal pattern
- User engagement
- Reinforcement learning

BACKGROUND

Video content often includes segments that viewers consider more important or interesting than other parts of the video. For instance, video of a sporting event includes key moments during a game when the audience is more engaged. Determining segments of a video that contain important or interesting moments via automated techniques can be extremely difficult. While the application of computer vision and other related techniques that utilize machine intelligence has improved automated determination of notable segments within video content, further refinement in these techniques can improve the quality and accuracy of the segments detected as salient.

A variety of video content, such as live sports broadcasts, is associated with real-time text comments and reactions from the viewers. The comments vary in terms of quality, interest, relevance, etc. Similar to the case of video content, it can be difficult to surface the most notable comments automatically, especially when a large volume of comments are generated within a short duration, which is often the case for comments associated with video content such as sporting events, awards shows, game shows, etc.

DESCRIPTION

This disclosure describes techniques that combine initial independent automated determination of notable video segments and text comments to refine and improve each aspect based on the other. Thus, the output of a model that marks notable video segments is used as a training input for a model that identifies salient text comments associated with the video, and vice versa. Using reinforcement learning, each model can be strengthened based on the output of the other model.

Video content and user comments are accessed and used for training machine learning models with specific user permission. Model output includes identification of salient video segments and notable comments, which can be utilized to provide highlights of the video, to surface salient comments in a live streaming video, etc. If a user denies permission, their comments and video content are not used during model training or to identify salient video segments or comments.

The described techniques leverage audience engagement with live or pre-recorded video content as captured via user comments entered while the video is being viewed. The comments reflect viewer engagement and interest in the corresponding points of the video. As a result, interaction and sentiment patterns within the comments are indicative of the level of interest of the audience in the corresponding events in the video content. For instance, comments containing high levels of emotions, such as surprise, disappointment, excitement etc. can increase during notable moments within the video, such as a team scoring in a game, a contestant being eliminated from a competition, a pundit making a controversial statement during an interview, etc.

The techniques involve two initial parallel processes that operate in a quasi-independent manner: (i) a trained machine learning model is applied for automated analysis of video content to identify and mark segments that contain notable moments; and (ii) a trained machine learning model is employed to perform automated analysis of text comments associated with the video to identify notable and high quality comments and mark the corresponding time points in the video pertaining to the comments. When the users permit, the output of each of the two initial models is then provided as input to the other model to refine the determined notable video segments and text comments.

The initial automated process applied to video content results in timestamps that mark the beginning and end of notable segments within the video. Such segments can include key moments, notable events, interesting aspects, etc. For example, segments within an awards event can mark parts within the video that correspond to granting of each of the awards, segments within a sports feed can capture parts of the game that involve scoring plays, segments within a TV show can identify key plot developments, etc.

Similarly, with user permission, the initial automated process applied to the text comments associated with the video can yield a set of comments considered to be high in quality and relevance. If users permit, the process can further be used to identify notable temporal patterns within the comment feed such as intervals of increased comment volume, appearance of keywords within specific periods, times of high interaction between commenters, patterns of sentiment reflected in the text, etc. User permitted identification of comments of high quality and relevance and patterns of interesting interactions and sentiment can be based on analyzing the text of the comments and/or associated lightweight interactions such as emotion symbols, upvotes, downvotes, reactions, etc.

The video content and the associated comment feed can be synchronized based on timestamps. Once synchronized, the notable video segments and text comments/patterns flagged by the respective initial automated processes can be used to refine the initial output of each of the processes. To that end, initially flagged video segments can be used to examine the text comments associated with that segment to identify additional comments of high quality and relevance and additional notable patterns of interaction and engagement. Similarly, if the users permit, initially flagged comments and commenting patterns can inform analysis of

corresponding segments within the video content to surface additional video content of interest that was previously unmarked.

Application of the techniques to a diverse variety of video content over a period of time can serve to refine the models. For instance, with user permission, notable patterns within the comment feed associated with a video can be applied to train the visual perception models for improved prediction of notable segments within the video. For example, an increase in anticipation reflected within the text of the comments (e.g., ahead of an important passage of play in a game) can be indicative of an upcoming event of high interest and excitement within the video.

An example application of the described techniques is improved automated generation of short highlight reels that capture the most important moments within long videos, such as sporting events, award shows, etc. In addition to improving automatically generated highlight reels, the proposed techniques can leverage the feedback loop created by the engagement metrics of user comments and interactions with the highlights of the video to further improve the highlight generation process. For instance, engagement metrics can indicate the relative success of various strategies for creating highlights such as including only positive sentiment, incorporating a mix of positive and negative aspects, etc. Moreover, with permission from the users, the metrics can be analyzed to determine a customized optimal sentiment mix for a highlight reel of a given video for a given audience.

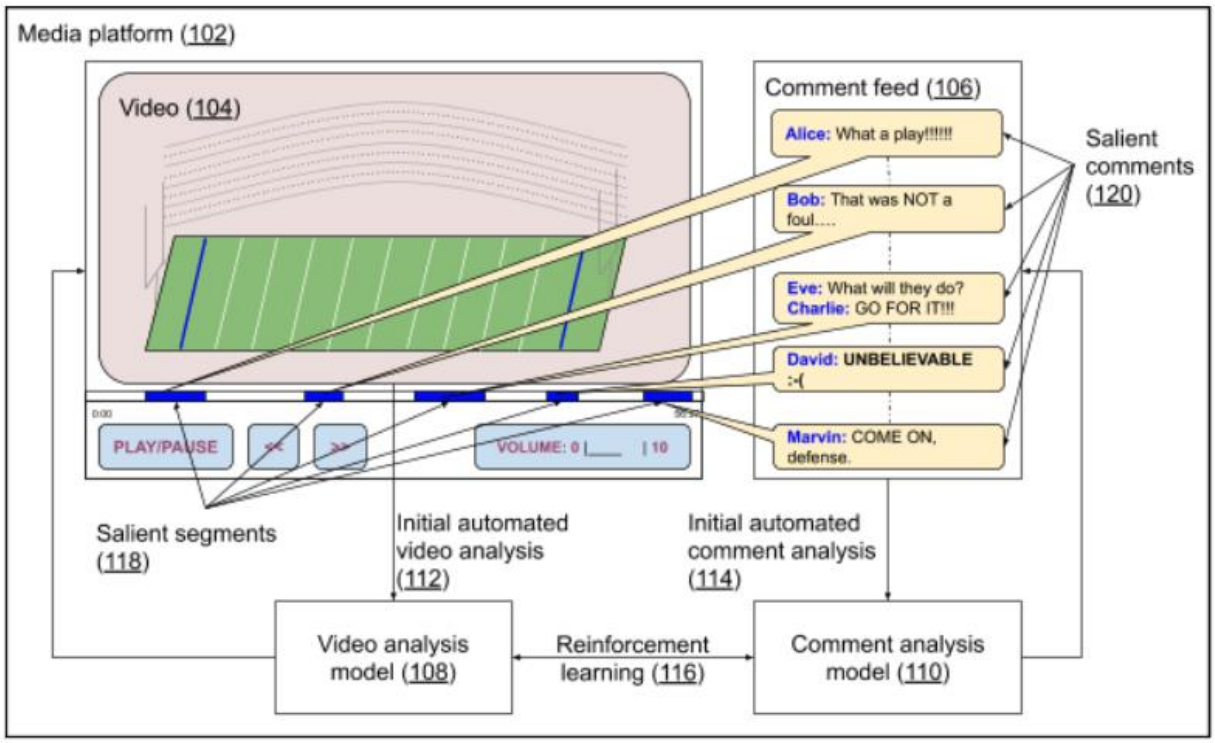

Fig. 1: Detecting and marking up salient segments within a video and associated comments

Fig. 1 shows an example operational implementation of techniques described in this disclosure. A media platform (102) is used to view video of a sporting event (104), e.g., a football game. The video is associated with a comment feed (106) that allows viewers to comment and interact with each other while watching the video. A video analysis model (108) is employed for initial automated analysis of the video (112) to determine segments within the video containing notable moments. Similarly, a comment analysis model (110) is employed for initial automated analysis of the comment feed (114) to detect salient interaction and sentiment patterns as well as high-quality and relevant comments. The output is identification of salient segments of the video (118) along with associated salient comments (120) within the comments feed.

The techniques described above can be employed by any media platform or service that includes user comments on video content. The techniques can be used, e.g., during live video

broadcasts of content such as sports, video game tournaments or other e-sports, award shows, personal events, etc. The techniques can be extended to commenting platforms used as a side channel for community engagement while watching video content elsewhere, such as on television. In such cases, even though the platform itself does not include video content, the techniques can serve to improve detection and flagging of salient comments and commenting patterns based on coupling with the corresponding video viewed on a separate platform. The video and comment analysis models can be a component of the media platform server. Alternatively, the models can operate locally on a user device. Further, video and comments analysis can be provided by the content provider or be implemented as a separate service.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that combines initial independent automated determination of notable video segments and text comments to refine and improve each aspect

based on the other. The techniques leverage audience engagement with live and pre-recorded video content as captured via user comments entered while the video is being viewed. Machine learning models that identify salient video segments can be trained using the output of models that identify notable comments and vice versa, using reinforcement learning. The described techniques can be employed by media platforms or services that enable users to comment on video content as well as those used as a side channel for community engagement while watching video content elsewhere.

REFERENCES

1. Zhang, Kun, and Yu He. "Segmentation of a video based on user engagement in respective segments of the video." U.S. Patent 9,465,435, issued October 11, 2016.