Technical Disclosure Commons

Defensive Publications Series

August 2023

AI-BASED ADAPTATION OF VIDEO SETTINGS BASED ON LEARNED USER BEHAVIOR

Alan Gatzke

Andi Wilson

Arsenii Astashkin

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Gatzke, Alan; Wilson, Andi; and Astashkin, Arsenii, "AI-BASED ADAPTATION OF VIDEO SETTINGS BASED ON LEARNED USER BEHAVIOR", Technical Disclosure Commons, (August 17, 2023)
https://www.tdcommons.org/dpubs_series/6150

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

AI-BASED ADAPTATION OF VIDEO SETTINGS BASED ON LEARNED USER BEHAVIOR

AUTHORS:
Alan Gatzke
Andi Wilson
Arsenii Astashkin

ABSTRACT

Techniques are presented herein that allow an asynchronous video messaging platform or an online communication and collaboration system to learn, over time, from a user's behavior to determine when to automatically adjust a video recording's playback speed and/or volume whenever a speaker (in such a recording) says something that the user would find interesting. Aspects of the presented techniques encompass a gathering of data (regarding, for example, the user's manipulation of playback speed and volume settings and actions like rewinding or pausing), an analysis of such data (leveraging natural language processing (NLP) techniques to examine video content, such as transcripts or captions, to identify keywords, topics, or sentiments), a modeling of the user's behavior (including individual preferences, interaction patterns, content preferences, etc.), an offering of specific playback speed and volume setting suggestions, reinforcement learning algorithms that dynamically adjust all of the above, and a collection of operational metrics (regarding accuracy, precision, recall, etc.).

DETAILED DESCRIPTION

When a user listens to either a video recording through an asynchronous video messaging platform or a recording of a meeting through an online communication and collaboration system, they often elect to play that recording at a faster-than-normal speed (such as, for example, 1.5 times faster than normal, two times faster than normal, etc.). For example, such a viewer will often listen to a recording at twice the normal speed with the player window minimized and, in the background, while they are multitasking (i.e., while they are working on something else in the foreground). The user may also have the volume set at a lower level while they are multitasking. Then, when the person who is speaking in the video says something that grabs the listener's attention (e.g., "Wait, what did my manager just say about using machine learning in an online communication and collaboration system?") the user will need to locate and select (e.g., click on) the video player to bring it to the foreground and then find the speed adjustment control to slow down the recording's playback to normal speed. In some cases, they may also increase the volume level.

By the time user has fumbled around getting to the playback speed control and the volume setting, they may have missed some of what was said in the recording. As a result, they must pause the video and then jump back some amount to replay the interesting part at a normal speed with, perhaps, a higher volume level.

Collaboration systems should be able to learn, over time, from each user's behavior when to automatically adjust a recording's playback speed and volume setting whenever a speaker says something that an individual listener would find interesting. Techniques are presented herein that support such a capability.

Turning to the presented techniques, whenever an asynchronous video messaging platform or online communication and collaboration system user manually changes the playback speed or volume while watching a video, the techniques may store the transcript text of the words that were spoken in the video X number of seconds before and Y number of seconds after the timestamp where the user made the adjustment. The techniques may also store the speed adjustment details (e.g., from two times faster than normal to one-half of normal speed, from 1.5 times faster than normal to normal speed, etc.) along with any volume adjustment details. The transcript text and adjustment details that were described above may be stored for each user.

Under aspects of the presented techniques, a machine learning module (within an asynchronous video messaging platform or online communication and collaboration system) may learn, over time, which words, phrases, subject matter, presenters, etc. trigger a speed adjustment or volume adjustment for a specific user based on the stored data that was described above.

According to further aspects of the presented techniques, when an asynchronous video messaging platform or online communication and collaboration system user plays a video recording, that video's transcript text may be analyzed to look for content that would typically trigger that listener to make a speed or volume adjustment. The techniques may then automatically make the appropriate adjustments (based on that user's past history) when that section of the video is played.

Under still further aspects of the presented techniques, in an application's user settings area a user may enable or disable any automatic speed and volume adjustments. Such an option may provide the user with the flexibility to only enable one of the settings. For example, some users may choose to enable automatic speed adjustments but disable automatic volume adjustments.

The presented techniques encompass a number of different elements, several of which will be described below.

A first element encompasses a gathering of data. The techniques may collect information on how a user interacts with a video, including, for example, playback speed, volume changes, and actions like rewinding or pausing. The techniques may also analyze how engaged the user is with the video, including, for example, whether they like, comment on, or share the video and how long they spend on certain parts of the video.

A second element encompasses an analysis of content. Natural language processing (NLP) techniques may be employed to examine a video's content (including, for example, transcripts or captions) and identify keywords, topics, or sentiments. Such an activity may help determine which parts of a video people find more interesting.

A third element encompasses a modeling of user behavior. In support of that activity a user profile may comprise individual preferences, interaction patterns, and the kind of content that a user is interested in. Among other things, clustering techniques may be used to group users based on their patterns and preferences.

A fourth element encompasses an offering of suggestions for different settings. A recommendation system may predict what playback speed and volume level a user might prefer, based on how that user has previously interacted with different videos and what content they like. Collaborative filtering and content-based filtering may be employed in support of this functionality.

A fifth element encompasses a 'learn as we go' capability. A reinforcement learning algorithm may be employed to adjust a recording's playback speed and volume in real time based on user feedback. Such an algorithm may learn from a user (who has elected to optin to such an offering) and keep updating its suggestions to continually improve the user's viewing experience.

And finally, a sixth element encompasses checking on how the presented techniques are working. Different metrics (regarding, for example, accuracy, precision, recall, and an F1 score) may be used to monitor the performance of the presented technique's various (machine learning, etc.) components. The instant model may be continually revised, through adjustments to the hyperparameters and updates to the training data, in support of ongoing improvements.

Figure 1, below, presents elements of an exemplary system that is possible according to the presented techniques and which is reflective of the above discussion.

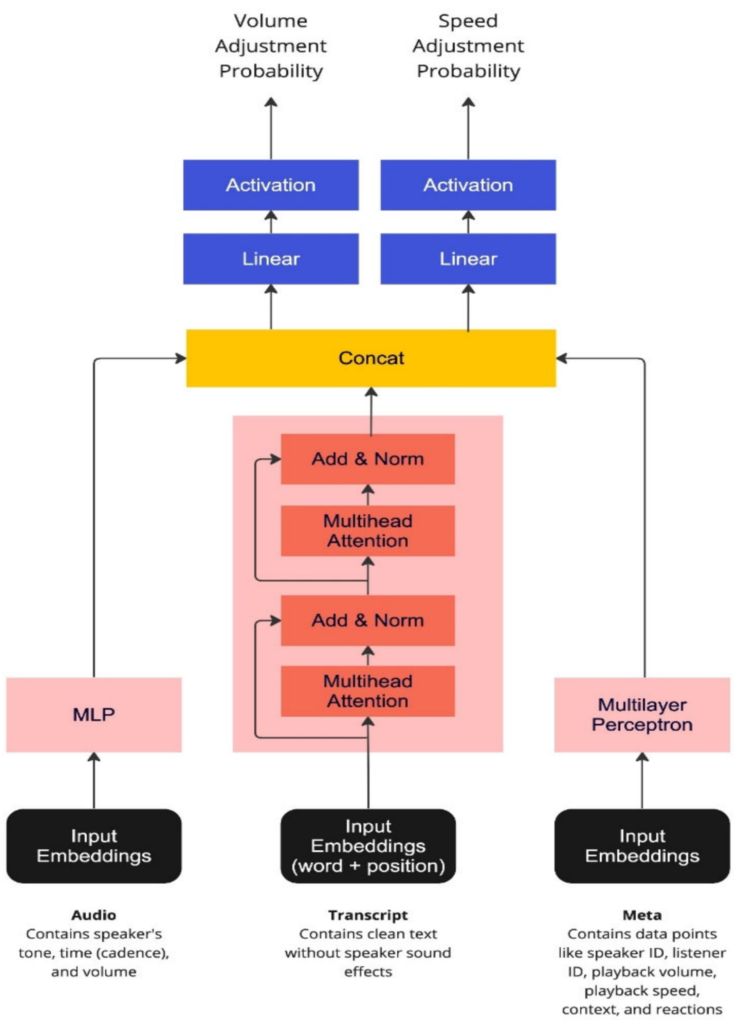

Figure 1: Exemplary System

Aspects of the presented techniques may be further understood through an illustrative example that highlights various initial assumptions that may be drawn based on the group in which a user resides.

Under the example, Chuck is delivering a presentation, and, during that delivery, he calls out some specific details regarding cost cutting in a particular group. The presented techniques may at that point develop an assumption regarding the need to slow down the recording playback speed and increase the volume if the instant viewer is part of the mentioned group. Such an action is an example of the presented techniques not leveraging past behavior but, instead, making playback adjustments based on a user's group membership.

The above-described example may be extended to encompass more general terms that a system, according to the presented techniques, may look for through, among other things, crowdsourcing. For example, if during his presentation Chuck speaks the word 'layoff' or the phrase 'limited restructuring,' the techniques could learn that all of the presentation's users may be interested in those topics. The techniques could then apply playback speed and volume changes even for a new user whose individual behavior has not yet been observed.

In summary, techniques have been presented herein that allow an asynchronous video messaging platform or an online communication and collaboration system to learn, over time, from a user's behavior to determine when to automatically adjust a video recording's playback speed and/or volume whenever a speaker (in such a recording) says something that the user would find interesting. Aspects of the presented techniques encompass a gathering of data (regarding, for example, the user's manipulation of playback speed and volume settings and actions like rewinding or pausing), an analysis of such data (leveraging NLP techniques to examine video content, such as transcripts or captions, to identify keywords, topics, or sentiments), a modeling of the user's behavior (including individual preferences, interaction patterns, content preferences, etc.), an offering of specific playback speed and volume setting suggestions, reinforcement learning algorithms that dynamically adjust all of the above, and a collection of operational metrics (regarding accuracy, precision, recall, etc.).