December 17, 2018

DETERMINING BREAKPOINTS IN MEDIA CONTENT

Migle Padegimaite

Sammy El Ghazzal

DETERMINING BREAKPOINTS IN MEDIA CONTENT

Background

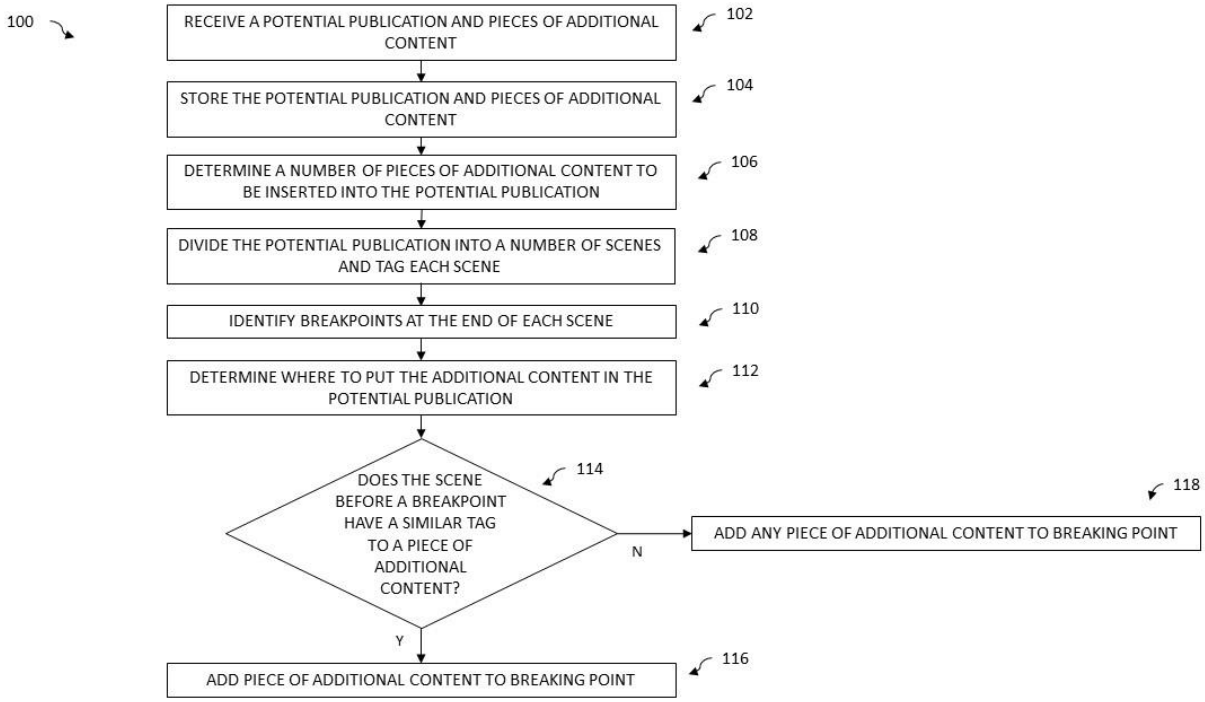
In recent years, more and more people are watching media content hosted by online media providers, such as YouTube, and straying away from watching media content on televisions as their mode of leisure and entertainment. With less people watching television, advertisements shown on television are getting less viewers, setting the stage for advertisers to focus on advertisements shown by online media providers.

Unlike media shown on television, which is typically divided into multiple segments with planned breaks for advertisements, media content hosted by online media providers generally does not account for breaks for advertisements. It is up to the media provider to break the media into segments and place those advertisements between the segments. Unfortunately, online media providers host hundreds of thousands of media content, with content being added each day, so it is impractical to manually determine the best place to put breaks in the content and add advertisements. Consequently, media providers often put breaks at arbitrary points in the media content, regardless of the context of the media content immediately before or after the break.

Summary

The system and method discussed herein provides a method of determining breakpoints in media content via a machine learning architecture by accounting for the context of the media content and placing separate pieces of additional content, such as advertisements, into the breakpoints. The machine learning architecture determines the number of pieces of additional content to insert into the media content, divides the media content into a number of scenes and tags each scene, identifies break points at the end of each scene, identifies which break point to insert pieces of additional content into, and then inserts the additional content. The machine learning architecture places a break point at the end of each scene and determines which break point to insert additional content into by determining the content and context of the scene or scenes before and/or after the break point.

A flow chart of the process of the system and method, according to some implementations, is shown below:



Adding Break Points to the Potential Publication

A flowchart of a process 100 for determining break points in a potential publication for inserting additional content into the potential publication, according to some implementations, is shown. Process 100 may be performed by a break point identifier which may be associated with a publication management system. The break point identifier could be a data processor associated with a backend processing system within a break point identifier environment. The break point identifier can include any number of processors. The break point identifier may learn how to autonomously conduct some of the operations listed below, so the system can run without any human or other entity input. Further, the break point identifier may continuously improve so break points are found more accurately within a potential publication and additional content can be inserted into break points that more closely match the context of the scenes surrounding the break points.

In an operation 102, a media provider, such as YouTube, receives both a potential publication and one or more pieces of additional content at the break point identifier. Such receipt may be simultaneous or not (e.g. additional content may be pre-loaded). In many implementations, the media provider receives both the potential publications and one or more pieces of additional content over the network. The potential publication and the one or more pieces of additional content may be locally loaded, however. The potential publications are often user video uploads created so third party viewers can watch them, but can be any form of media created for any reason. The potential publications are uploaded to the media provider's break point identifier. The potential publications can be any length and can be about any subject. Additional content may be any type and form of publication or video, including television programs, video logs (vlogs), documentaries, movies, news segments, or any other type and form of publication. Through the break point identifier, the media provider may compress the potential publication and/or the additional content once they are added. Additional content may be added in conjunction with a potential publication, but additional content is often uploaded separately from potential publications so the additional content can be paired with any number of potential publications. There may be times where a potential publication is uploaded to the media provider without any corresponding additional content.

In an operation 104, the potential publication and the additional content, the break point identifier may store the publications within the break point identifier. Both the additional content and the potential publication may be stored in a repository within the break point identifier. Additional content may be inserted into multiples types of media, including videos, sound recordings, radio segments, video games, and other forms of media. In many implementations, additional content may be received and stored in advance of receipt of potential publications. In these implementations, the additional content may be inserted into multiple potential publications.

In an operation 106, the break point identifier determines the minimum number of break points that need to be identified in a stored potential publication so the potential publication can include each piece of additional content that is allocated to be inserted into the potential publication. If the media provider wants to include two advertisements in a video, then the break point identifier will include at least two break points in the video, for example. The number of pieces of additional content that will be added to the potential publication is generally

predetermined by the media provider. The media provider may determine the amount of additional content it wants to insert into the potential publication based on the content and characteristics of each potential publication. Example characteristics include the length of the potential publication, the popularity of the potential publication, the subject matter of the potential publication, and other characteristics of the potential content. Additionally, the provider may determine a rule set that governs how many pieces of additional content should be inserted into potential publications of a given length. Potential publications that are eight minutes long may need a minimum of 3 break points, while potential publications that are seven minutes long may need a minimum of 2 break points for additional content. The media provider may make any number of rules concerning how many break points and additional content to include for potential publications of any length.

In an operation 108, the break point identifier divides the potential publication into a number of scenes based on the content of each scene. To break the potential publication into scenes, the break point identifier may take the entire potential publication as an input and analyze the potential publication to determine the types of content it includes. The break point identifier may analyze a potential publication's visual and audio content when determining the types of content it includes. Different types of visual content the potential publication may include are, but are not limited to, dialogue, landscape, romance, comedy, length of a continuous camera shot, and whether a picture fades to or from black. Different types of audio content the potential publication may include are dialog, music, sound effects, and other such parameters. The break point identifier may use a rolling window to determine the types of content of the potential publication based on aggregated data that the rolling window collects to create scenes. The rolling window collects data from a pre-selected number of frames within the potential publication at one time and moves across the timeline of the potential publication, collecting data from each frame in the potential publication and within the context of other frames within the rolling window. The pre-selected length of the rolling window can include any number of frames. A rolling window may be used in some implementations to collect data instead of just collecting data at each isolated frame so the break point identifier can more accurately determine the context of the frames in the potential publication (e.g. avoiding mischaracterizing action sequences with multiple cuts or fades as separate scenes in which content may be inserted). Scenes may be any length and may include consecutive frames of visual content and the

associated audio of the visual content of the potential publication. Scenes may also contain one of visual content and audio content depending on the format of the potential publication. There can be any number of scenes in a potential publication.

To learn how to detect different types of content in a potential publication, the break point identifier may receive training data gathered from reference publications that have been manually broken up and tagged by human raters and/or publishers or producers. The reference publications may be manually segmented or divided into different scenes based on the content and context of each scene, creating boundaries for each scene. The reviewer or publisher may choose scenes to be as short as possible so the break point identifier can determine different scenes as finely as possible. When segmenting the reference publication into scenes, the human raters may also tag each scene with tags that describe the content or characteristics of each scene. Different tags may be, but are not limited to, dialogue, landscape, romance, action, comedy, length of a continuous camera shot, and whether a picture fades to or from black. Different types of audio content the potential publication may include are dialog, music, sound effects, and other such parameters. The break point identifier processes the reference publications and, in some implementations, uses a pre-selected threshold percentage to determine which actions or pieces of content are associated with specific scenes. In some such implementations, if the break point identifier determines that certain actions or pieces of content are tagged with the same tag in a percentage of reference publications that is above a pre-selected threshold percentage, the break point identifier may associate the action or content with the tag in potential publications. The training data may also include other tags humans have given the same scene as metadata. Further, scenes may be tagged multiple times, a scene may be tagged as action and dialog, for example, within the same reference publication, and the extra tags are included as metadata. For example, if the threshold percentage is 75 percent and if an explosion is shown in a video in multiple reference publications and is tagged with the tag “action” in 80 percent of the reference publications, the break point identifier may determine that the explosions in potential publications should be tagged with the “action” tag. If the explosion is only tagged with the action tag in 70 percent of the reference publications, however, then the break point identifier may not identify explosions with action tags in potential publications. Identifications of such explosions may be done visually (e.g. detecting flashes, colors corresponding to flames, etc.) and/or acoustically (e.g. detecting sound effects matching reference explosions). The break point

identifier may also learn the boundaries of the explosion by analyzing the human rater set boundaries of the explosions in the reference publications. Further, the break point identifier may determine metadata for the explosion scene by seeing the other tags human raters gave the explosion. The break point identifier may also learn to apply time stamps at the boundaries at the end of each scene.

In some implementations, the break point identifier may comprise a machine learning system, such as a neural network or support vector machine or other classifier system, and may be trained using the manually tagged reference publications. In one such implementation, the machine learning system may classify each frame within a publication as including a break point or not including a break point (e.g. two outputs from a neural network), with inputs from a plurality of frames (e.g. a rolling or sliding window of frames). For example, in one such implementation, a sliding window may comprise a target frame (e.g. frame to be classified as being a break point or not) along with a number of frames before and after the target frame (e.g. 90 frames, or 3 seconds at 30 frames per second; 150 frames or 5 seconds; or any other such time duration). The number of frames before and after the target frame may be the same or different. In some implementations, the frames provided to the break point identifier may be filtered or decimated to reduce processing and/or storage requirements (e.g. one frame out of every 10, one frame per second, or any other rate). In some implementations, the frames may also be compressed, reduced in resolution and/or bit depth, or otherwise processed to reduce analysis processing requirements. Audio may also be provided for analysis, including dialog, music, or sound effects, within the same window of time. In some implementations, the audio may also be reduced or filtered to reduce processing and/or storage requirements (e.g. reducing a sample rate or bit depth, compressing the audio via a lossy compression algorithm such as MP3, filtering low and/or high frequencies, mixing stereo audio to mono, etc.). In some implementations, additional metadata may also be provided for analysis, such as the genre, type, or other identifiers or characteristics discussed above.

Reference publications may be manually tagged with break points, and may similarly be provided to the machine learning system in a training process. The machine learning system may adjust parameter or bias weights (e.g. in hidden layers of a neural network) such that the frames of the reference publications are accurately identified in accordance with the manual tags. Once trained, the system may be applied to new publications to identify break points. In some

implementations, the machine learning system may include additional outputs, such as tagging target frames or scenes between adjacent break points according to genre, content, or other characteristics as discussed above. The manually added tags from reference publications may similarly be provided as outputs of the network for supervised learning, as discussed above.

Once the break point identifier has learned how to segment and tag scenes in potential publications, the break point identifier can scan a potential publication to find scenes within the potential publication that match scenes from the reference publication training data. The break point identifier can set the boundaries of the scenes, label the scenes with appropriate tags, and create a time stamp at the end of each scene.

In an operation 110, the break point identifier identifies break points in the potential publication after segmenting the potential publication into scenes. Break points are points in a potential publication where additional content may be inserted. To identify the break points, the break point identifier analyzes the content of each scene along with each scene's boundaries and identifies break points at the end of each scene along with the content of the scene. For example, after determining a set of frames make up a scene about comedy and tagging the scene with the comedy tag, the break point identifier identifies a break point at the end of the scene and includes information in the break point that it follows a comedy scene. The break point identifier scans the entire potential publication for scenes and identifies break points at the end of each scene. The break point identifier may also place break points at the beginning of each scene and include information about that scene. Further, the break point identifier can place break points at the beginning of the potential publication.

In operation 112, the break point identifier determines which break points to insert additional content into in the potential publication. To do so, the break point identifier first identifies the number of pieces of additional content that will be inserted into break points and, thus, the minimum number of break points that are needed in the potential publication. This number can be determined by the break point identifier as described above or determined by the media provider. The break point identifier then places a piece of additional content in break points based on a pre-selected set of rules, such as trying to maximizing the spacing of the ads. For example, if the potential publication is five minutes long and two pieces of additional content need to be placed in two break points, the break point identifier may place the additional content at the beginning of the potential publication and at the break point that is closest to the 2.5

minute mark. The break point identifier can use different sets of rules to identify different break points to insert the additional content. Over time, using a reinforcement learning approach, the break point identifier can use metrics such as user engagement to determine how to maximize the effectiveness of each piece of additional content that is inserted into the potential publication. If there are not enough break points in the potential publication to accommodate the additional content that needs to be included with the potential publication, the break point identifier may add break points at places in the potential publication based on other sets of rules, such as evenly throughout the potential publication, or at the beginning or end (e.g. pre-roll or post-roll content).

In operation 114, the break point identifier determines which piece of additional content to insert into specific break points in the potential publication. To do so, the break point identifier analyzes the break points it has determined additional content can be inserted into and determines the tag of the scene that immediately precedes the break point. The break point identifier may also determine the tag of the scene that immediately comes after the break point. If the scene's tag matches a piece of additional content's tag, then, in operation 116, the break point identifier may place the piece additional content with the same tag into the break point. If there is not a piece of additional content with a tag that matches a scene's tag, the break point identifier may place additional content into break points regardless of the additional content's tag. The additional content can be tagged by the content provider that provides the content or it can be tagged by the break point identifier in a manner similar to how the break point identifier tagged the potential publication.

Abstract

Systems and methods for determining breakpoints in media content via a machine learning architecture by accounting for the context of the media content, and for placing separate pieces of additional content into the breakpoints, are described. The machine learning architecture determines the number of pieces of additional content to insert into the media content, divides the media content into a number of scenes and tags each scene with appropriate identifiers based on comparison to reference content, identifies break points at the end of each scene, identifies which break point to insert pieces of additional content into, and then inserts the additional content. The machine learning architecture places a break point at the end of each scene and determines which break point to insert additional content into by determining the content and context of the scene or scenes before and/or after the break point.