July 14, 2016

Automatic Detection of Optimal Ad Insertion Points in Video

Ji Chen

Yu Chen

Automatic Detection of Optimal Ad Insertion Points in Video

Abstract

This paper describes a system employing supervised machine learning algorithms to automatically detect optimal ad insertion points in an item of video content. The system can learn from data obtained from manual ad insertion. Manual ad insertion examples can be obtained from broadcast television programming containing ads, or from manually applied labels obtained through crowd sourcing. The system can process the raw audio and video signals to obtain a set of features that best describe points of manual video ad insertion, and feed them into a machine learning model. The machine learning model can include a deep neural network or other supervised classification algorithms. This system can learn from human inputs, and produce results that best fit in human perception of optimal ad insertion points. The system can thus improve user experience on video ads, drive higher ad revenue for content publishers, and reduce the burden on content providers.

Background

Current online video ads are inserted into video content in pre-set, arbitrary time slots, without consideration of the video context. Sometimes ads are inserted at points which negatively impact user experience, such as in the middle of a sentence or song. This can make the video ads annoying relative to manually placed TV or radio, ads which are placed in appropriate slots between relatively independent segments of content.

Discussion

Figure 1 illustrates an example system for automatic detection of optimal ad insertion points in video. The system can include a data processing system 110 in communication with a content publisher 150, a content provider 155, and a client 160 via a network 105. The system can include more than one of each component acting in concert or individually. The data processing system 110 can include a training model data collector 120, a signal extractor 125, a modeler 130, a model validator 135, a content item inserter 140, and a database 145.

The content publisher 150 may host content and resources for viewing and access by one or more clients 160. The hosted content can include web pages, web applications, images, audio, and video. The content publisher 150 can receive the hosted content from the content provider

155. The content provider 155 can upload the web pages, web applications, images, audio, and video to the content publisher 150 for hosting. The content publisher 150 may include third-party content items such as new bulletins, web widgets, and advertisements into the uploaded content prior to serving the content to one or more viewers at one or more clients 160. The third-party content items can be provided by the data processing system 110. The content publisher 150 may receive remuneration or compensation from the data processing system 110 for each view, click, or other interaction between the client 160 and a content item. The content publisher 150 may in turn incentivize the content provider 155 to upload content and include third-party content items for display with the uploaded content.

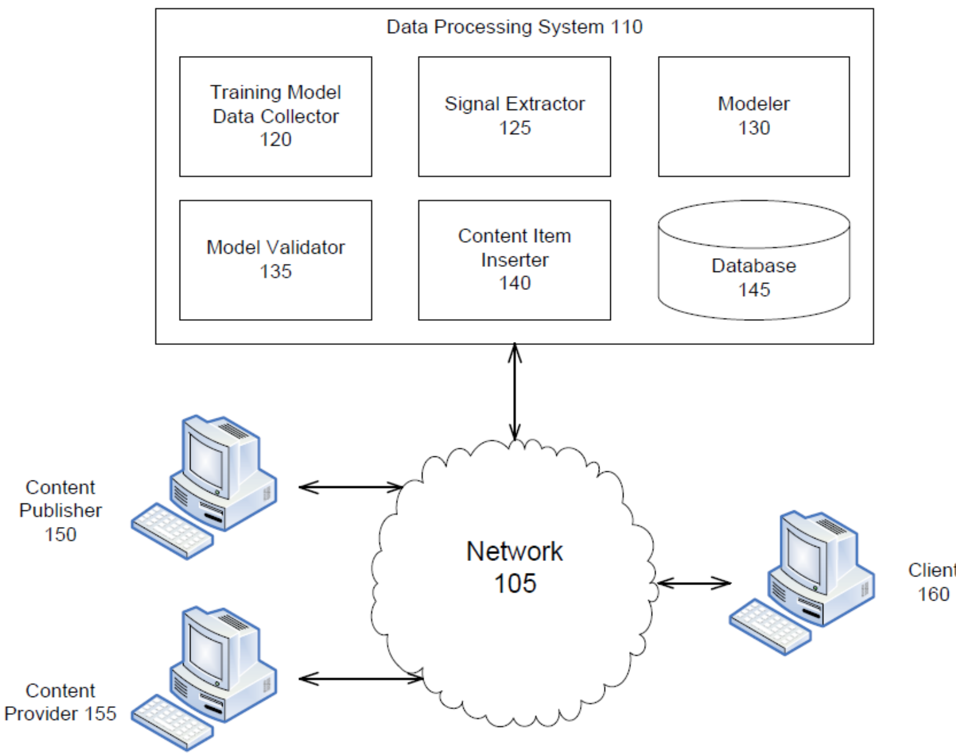

Figure 1: Example System for Automatic Detection of Optimal Ad Insertion Point in Video

The data processing system 110 can work in conjunction with the content publisher 150 to automatically detect insertion points for third-party content items in content hosted by the content publisher 150. This allows the content provider 155 to upload content and consent to the insertion of third-party content items, without having to manually set each insertion point for the content items. The data processing system 110 can include several components, modules, or subsystems for automatically detecting ad insertion points in the hosted content.

The data processing system 110 can include a training model data collector 120. The training model data collector 120 can receive training data from content that has already had content items inserted or from crowd sourcing. For example, most radio and television shows are broadcast with commercials added. The training model data collector 120 can receive copies of broadcast content with commercials. The broadcast content can thus provide data of examples of manually placed content items. In another example, the training data can be gathered from crowd sourcing techniques. In crowd sourcing examples, the data processing system 110 can insert content items into hosted content at randomly or quasi-randomly chosen points, and use the crowd sourcing service for feedback regarding the appropriateness of the insertion point. Alternatively, the data processing system 110 can provide the raw (content item free) content to the crowd sourcing system, which can in turn allow human operators to manually select insertion points. The data collected by the training model data collector 120 can be fed to the modeler 130 for processing.

The data processing system 110 can include a signal extractor 125. The signal extractor 125 can apply an algorithm to parse audio to determine words, sentences, or context of dialogue; music; or ambient noise. The signal extractor 125 can process the parsed content to determine points of change or inflection in the dialogue, music, or ambient noise. The signal extractor 125 can set markers for possible insertion points at the determined points of change. During the building of the model, the signal extractor 125 can extract signals from the training data. During deployment, the signal extractor 125 can extract signals from content uploaded by the content provider 155, and set markers for candidate insertion points.

The data processing system 110 can include a modeler 130. The modeler 130 can receive the data collected by the training model data collector 120 as well as data from the signal extractor 125, and use them to build a model. The modeler 130 can employ a machine learning algorithm to automatically select content item insertion points in hosted content. The machine learning algorithm can include a deep neural network (DNN), long short-term memory (LSTM), or other supervised classification algorithms. The modeler 130 can build a model based on features extracted from content in the training data, and from the content item insertion points in that content. The model can then be applied to raw content to automatically select content item insertion points.

The data processing system 110 can include a model validator 135. After the modeler 130 builds the model, the model validator 135 can gauge the accuracy of the model. The model validator 135 can validate the model by applying it to known-good data that was not used to build the model. Model validation can verify that the model was not over fit to the training data, and can therefore produce useful results from real data.

The data processing system 110 can include a content item inserter 140. The content item inserter 140 can apply the model to uploaded content to set insertion points for third-party content items. The content item inserter 140 can set insertion points based on markers set by the signal extractor 125. The content item inserter 140 can insert content items into the insertions points. Alternatively, the content item inserter 140 can place cues in the content. The cues can tell the data processing system 110 when during playback of the hosted content to insert content items. The cues may include metadata describing context of the insertion point that can be used for selecting relevant content items for insertion.

The data processing system 110 can include a database 145. The database 145 can store all of the instructions and/or data for carrying out the operations of the system.

Figure 2 illustrates an example method for automatic detection of optimal ad insertion point in video. The example method can include collecting training data (Step 210), extracting signals (Step 220), building a model (Step 230), validating the model (Step 240), deploying the system (Step 250), and refining the model (Step 260). Each Step of the example method is described more fully below.

The example method can include collecting training data (Step 210). The training model data collector 120 can collect training data from audio and video clips with insertion points placed manually; for example, a television or radio broadcast with commercials. In some implementations, the training model data collector 120 can collect training data using crowd sourcing. In one possible crowd sourcing technique, the data processing system 110 can place insertion points into a piece of hosted content, and employ one or more human viewers to rate the appropriateness of the insertion points. In another possible crowd sourcing technique, the data processing system 110 can play the piece of hosted content for one or more human viewers, and allow the viewers to affirmatively indicate appropriate insertion points.

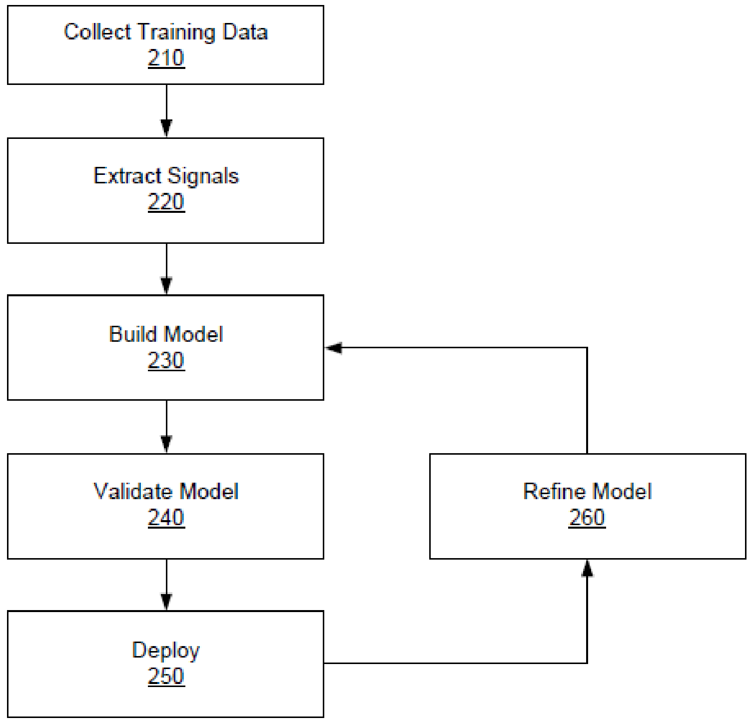

Figure 2: Example Method for Automatic Detection of Optimal Ad Insertion Point in Video

The example method can include extracting signals (Step 220). The signal extractor 125 can extract signals from the training data. The signal extractor 125 can parse audio to determine words, sentences, or context, and set markers at points of change or inflection. The signal extractor 125 can similarly parse music or ambient noise of the uploaded content to determine other points of change.

The example method can include building a model (Step 230). The modeler 130 can use the training data, and the markers set by the signal extractor 125, to build a model for selecting insertion points in content. The modeler 130 can look at the markers set in the training data by the signal extractor 125 against the actual placement of insertion points in the training data, and use both to train the model on the placement of optimal insertion points, using the human selected insertion points as a guide.

The example method can include validating the model (Step 240). The model validator 135 can validate the model by applying it to known-good data that was not used by the modeler 130 in building the model. The model validator 135 can compare insertion points set by the model to the human selected insertion points. If the validator 135 determines that a similarity in the respective insertion points exceeds a certain threshold, it can conclude that the model is valid and will likely set appropriate insertion points in newly presented pieces of hosted content.

The example method can include deploying the system (Step 250). The data processing system 110 can deploy the model once it has been validated. The signal extractor 125 can set markers at points of change in the newly presented pieces of hosted content. The content item inserter 140 can apply the model to content and to the markers set by the signal extractor 125 to set insertion points in the content. The content item inserter 140 may insert content items at the insertion points. Alternatively, the content item inserter 140 may place cues at the insertion points, allowing the data processing system 110 to select third-party content items for insertion during playback, rather than in advance.

The example method can include refining the model (Step 260). The data processing system 110 can continually refine the model based on feedback to the system. The feedback may be in the form of adjustment of insertion points performed by the content provider 155 who uploaded the content. That is, the content provider 155 may preview their uploaded content with content items added, and decide to move one or more insertion points to a more appropriate time. The feedback may be in the form of crowd sourcing techniques similar to those employed for collecting training data. For example, individuals can view or listen to the content and provide feedback regarding the appropriateness of each insertion point. The feedback can be used by the modeler 130 to refine the model. The model validator 135 can revalidate the model periodically or after significant changes.

Conclusion

The above-discussed automatic ad insertion system can improve the viewer experience by inserting ads at appropriate points in hosted content. The system can also be provided as a service to content providers whereby ads are automatically placed into uploaded audio and video clips at appropriate times. This can save both the uploader and the host the labor of selecting insertion points manually. Furthermore, both parties are more likely to insert an optimal number of ads into an audio or video clip if there is no marginal increase work required to do so. For example, a video uploader may insert fewer than an optimal number of ads if each ad insertion requires a distinct affirmative act. The system can thus improve user experience on video ads, drive higher ad revenue for content publishers, and reduce the burden on content providers.