December 2022

Dynamic Format Selection of Media Content

Kevin Chen

Anil Vijendran

Jeremy Zhou

Rob Gaunt

Pawel Jurczyk

See next page for additional authors

Inventor(s)
Kevin Chen, Anil Vijendran, Jeremy Zhou, Rob Gaunt, Pawel Jurczyk, Chubing Zeng, Michael Smouha, and Colin Whittaker

DYNAMIC FORMAT SELECTION OF MEDIA CONTENT

ABSTRACT

Disclosed herein is an improved mechanism for dynamic format selection of media content. The mechanism can use a machine learning model to select a format (e.g., a playback resolution) for providing media content based on information available at the time when the request for providing the media content is received. Such information can include, for example, location information (e.g., the geographic location of a user device that is presenting the media content), account information (e.g., previously selected media content items, watch time statistical information, subscription information, etc.), and/or network condition information (e.g., current egress traffic from one or more servers associated with the media streaming service).

BACKGROUND

Users frequently stream video content (e.g., videos, movies, television programs, music videos, etc.) from media content streaming services. A user device may use adaptive bitrate streaming, which can allow the user device to request different qualities of a video content item as the video content item is streamed from a server, thereby allowing the user device to continue presenting the video content item even as a quality of a network connection used to stream the video content item changes. For example, the user device may begin presenting segments of video content item that have a first, relatively high resolution, and, subsequently, in response to determining that the network connection has deteriorated, the user device can request segments of the video content item that have a second, lower resolution from the server. It can, however, be resource-intensive for a user device to determine an optimal format to be requested from the server. Additionally, the user device may request a particular format without regard to resources available to the server.

Accordingly, it is desirable to provide new systems and methods for dynamic format selection of media content.

DESCRIPTION

The systems and techniques described in this disclosure relate to dynamic format selection of media content.

Turning to FIG. 1, an illustrative example of a process for selecting a format of a video content item is shown in accordance with some embodiments of the disclosed subject matter. In some embodiments, blocks of process 100 can be executed by any suitable device. For example, in some embodiments, blocks of process 100 can be executed by a server associated with a video content streaming service.

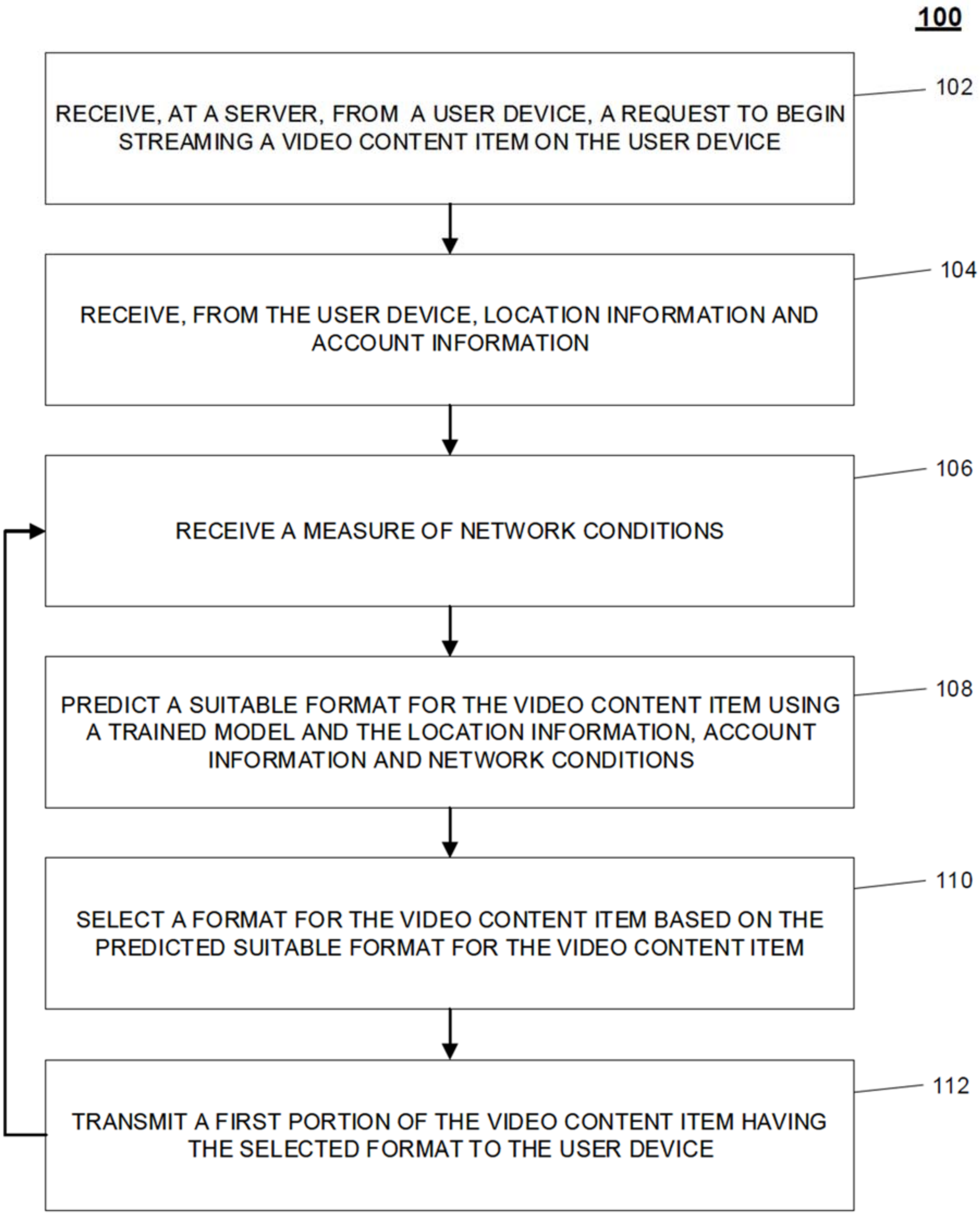
100
RECEIVE, AT A SERVER, FROM A USER DEVICE, A REQUEST TO BEGIN STREAMING A VIDEO CONTENT ITEM ON THE USER DEVICE
102
RECEIVE, FROM THE USER DEVICE, LOCATION INFORMATION AND ACCOUNT INFORMATION
104
RECEIVE A MEASURE OF NETWORK CONDITIONS
106
PREDICT A SUITABLE FORMAT FOR THE VIDEO CONTENT ITEM USING A TRAINED MODEL AND THE LOCATION INFORMATION, ACCOUNT INFORMATION AND NETWORK CONDITIONS
108
SELECT A FORMAT FOR THE VIDEO CONTENT ITEM BASED ON THE PREDICTED SUITABLE FORMAT FOR THE VIDEO CONTENT ITEM
110
TRANSMIT A FIRST PORTION OF THE VIDEO CONTENT ITEM HAVING THE SELECTED FORMAT TO THE USER DEVICE
112

FIG. 1

Process 100 can begin, at 102, by receiving, at a server from a user device, a request to begin streaming a video content item on the user device. As described above, in some embodiments, the server can be associated with any suitable entity or service, such as a video content streaming service, a social networking service, and/or any other suitable entity or service. In some embodiments, the server can receive the request to begin streaming the video content item in any suitable manner. For example, in some embodiments, the server can receive the request from the user device in response to determining that a user of the user device has selected an indication of the video content item (e.g., an indication presented in a page for browsing video content items, and/or selected in any other suitable manner). As another example, in some embodiments, the server can receive an indication that the video content item is a subsequent video content item in a playlist of video content items that are to be presented sequentially on the user device.

At 104, process 100 can receive, from the user device, location information and account information. In some embodiments, location information can include a current geographic location of the user device. For example, in some embodiments, the geographic location can be indicated by an IP address associated with the user device. As another example, in some embodiments, the geographic location can be indicated by current GPS coordinates associated with the user device. In a more particular example, in some embodiments, location information can include the time zone and local time associated with the current geographic location of the user device. In some embodiments, the account information can include any suitable information about the account being used to access the streaming service. For example, in some embodiments, the account information can include subscription information to a video content streaming service. As another example, in some embodiments, the account information can include previous user actions (e.g., previously selected video content item formats, amount of watch time associated with the previously selected video content item formats, and/or any other suitable user actions).

As another example, in some embodiments, the account information can include watch time statistics (e.g., total watch time of the account, amount of time watched per day, most likely time of day for activity, and/or any other suitable watch time statistics) and viewing history of the user account. As another example, in some embodiments, the account information can include information related to a billing cycle for payment of a subscription to a video content streaming service, and/or any other suitable information.

At 106, process 100 can receive a measure of network conditions. In some embodiments, the measure of network conditions can include current egress traffic from servers associated with the video content streaming service to internet service providers (ISPs). For example, in some embodiments, egress can be measured as gigabytes (GB) per unit time (e.g., hour) of content. As another example, in some embodiments, egress can be further associated with one or more suitable categories. In a more particular example, in some embodiments, egress can be measured for current traffic being served to ISPs in a particular metro area. In another more particular example, current egress can be measured according to specific attributes of the servers providing content for the streaming service (e.g., serving location, rack type, traffic quality of service). In another more particular example, egress can be designated as "peak" or "off-peak" based on the time of serving, and/or according to any suitable designation related to available network capacity. Continuing this particular example, peak egress can be identified as the maximum amount of content (GB) served between the streaming service and ISPs, averaged over a particular time window. As a particular numeric example, in some embodiments, peak egress can be quantified as the maximum egress (bytes) in a 5 minute window, divided by 300 seconds.

In some embodiments, the measure of network conditions can include a metric accounting for current egress traffic in relation to peak egress traffic. In some embodiments, the metric can be known as a peak multiplier. For example, in some embodiments, the peak multiplier can be a direct ratio of current egress and peak egress (e.g., current egress as a percentage of peak egress). As another example, in some embodiments, a mathematical function can be used to scale the current egress to the peak egress. As a more particular example, a Sigmoid function, or any other suitable mathematical function, can take, as inputs, the current egress and peak egress, and return a peak multiplier. As another example, in some embodiments, a series of predefined levels can be used to relate the current egress, peak egress and the peak multiplier. As a particular numerical example, in some instances, if the current egress is 40% of the peak egress, a peak multiplier of 0.1, or "off-peak" can be used. Continuing this example, for the current egress to peak egress percentages of (> 40% and </=60%) , (<60% and </=70%) , (<70% and </=100%), the corresponding peak multipliers can be 0.3, 0.7, and 1.0, respectively. In some embodiments, any suitable arrangement can be used to relate the peak multiplier to the current egress and peak egress.

At 108, process 100 can predict a suitable format for the video content item using a trained model and the location information and account information, both received at block 104, and network conditions received at block 106. In some embodiments, the format can include any suitable information, such as a resolution of a frame of the video content item (e.g., 144 pixels x 256 pixels, 240 pixels x 426 pixels, 360 pixels x 640 pixels, 480 pixels x 854 pixels, 720 pixels x 1280 pixels, 1080 pixels x 1920 pixels, and/or any other suitable resolution).

In some embodiments, process 100 can predict the suitable format for the video content item in any suitable manner. For example, in some embodiments, process 100 can predict the suitable format for the video content item using a model that was trained using training data that indicates a quality score associated with streaming video content items using particular formats for different user locations and/or under different network conditions (as shown in and described below in connection with FIG. 2).

As shown in and described below in connection with FIG. 2, in some embodiments, such a model can take, as inputs, user location information, user account information, network conditions, and format, and can predict, as an output, a predicted quality score associated with streaming a video content item with the video content item format to a user device associated with the device information and network information.

In some such embodiments, the model can then be used to predict a format that is likely to maximize the quality score. Note that, in some embodiments, the quality score can include any suitable metrics that indicate a quality of streaming of the video content item. For example, as described below in more detail in connection with FIG. 2, in some embodiments, the quality score can include a watch time metric, which can indicate an average duration of time a video content item is watched before presentation of the video content item is stopped. As another example, as described below in more detail in connection with FIG. 2, in some embodiments, the quality score can include a network traffic metric. The network traffic metric can, in some embodiments, indicate an egress cost associated with streaming the video content item to the viewer at the location indicated by the location information and with the predicted format of video content item.

Note that, in some embodiments, as described below in more detail in connection with FIG. 2, the trained model can use any suitable features other than those related to location information, account information, or network information to predict the suitable format for the video content item. For example, in some embodiments, the trained model can use input features related to the video content item (e.g., a genre or topic of the video content item, a duration of the video content item, a popularity of the video content item, and/or any other suitable video content item information), and/or any other suitable information.

At 110, process 100 can select a format for the video content item based on the predicted suitable format for the video content item. In some embodiments, process 100 can select the format for the video content item in any suitable manner. For example, in some embodiments, process 100 can select the format for the video content item to be the same as the predicted suitable format. As a more particular example, in some embodiments, in an instance in which process 100 predicts a suitable format as a particular resolution, process 100 can select the format as the particular resolution. As another example, in some embodiments, process 100 can select the format for the video content item based on the predicted suitable format and subject to any suitable rules or criteria. As a particular example, in some embodiments, process 100 can select the format for the video content item subject to rules that indicate a maximum resolution for particular types of video content (e.g., music videos, lectures, documentaries, etc.). As a specific example, in an instance in which the predicted suitable format is a particular resolution (e.g., 720 pixels x 1280 pixels), and in which the video content item is determined to be a music video, process 100 can select the format as a resolution that is lower than the resolution associated with the predicted suitable format (e.g., 360 pixels x 640 pixels, 240 pixels x 426 pixels, and/or any other suitable lower resolution). Note that, in some embodiments, process 100 can select the format for the video content item based on any suitable rules that indicate, for example, a maximum or minimum resolution associated with any conditions (e.g., network conditions, device conditions, user subscription conditions, conditions related to different types of video content, and/or any other suitable types of conditions).

At 112, process 100 can transmit a first portion of the video content item having the selected format to the user device. Note that, in some embodiments, the first portion of the video content item can have any suitable size (e.g., a particular number of kilobytes or megabytes, and/or any other suitable size) and/or can be associated with any suitable duration of the video content item (e.g., five seconds, ten seconds, and/or any other suitable duration). In some embodiments, process 100 can calculate a size or duration of the first portion of the video content item in any suitable manner, such as based on a buffer size used by the user device to store the video content item during streaming of the video content item. In some embodiments, process 100 can transmit the first portion of the video content item in any suitable manner. For example, in some embodiments, process 100 can use any suitable streaming protocol to transmit the first portion of the video content item. As another example, in some embodiments, process 100 can transmit the first portion of the video content item in connection with an indication of a condition under which the user device is to request an additional portion of the video content item from the server. As a more particular example, in some embodiments, process 100 can transmit the first portion of the video content item in connection with a minimum buffer amount at which the user device is to request additional portions of the video content item.

In some embodiments, process 100 can return to block 106 and can receive, from the user device, updated network conditions (e.g., that a serving attribute has changed, or that the metrics accounting for current egress traffic in relation to peak egress traffic have been updated), and/or updated location and/or account information (e.g., that an IP address has changed). In some such embodiments, process 100 can loop through blocks 106 – 112 and can predict an updated suitable format based on the updated network conditions, updated location information, and/or updated account information, select an updated format based on the predicted updated suitable format, and can transmit a second portion of the video content item having the selected updated format to the user device. In some embodiments, process 100 can loop back to block 106 in response to any suitable information and/or at any suitable frequency. For example, in some embodiments, process 100 can loop through blocks 106 – 112 at any suitable set frequency such that additional portions of the video content item are transmitted to the user device at a predetermined frequency. As another example, in some embodiments, process 100 can loop back to block 106 in response to receiving a request from the user device for an additional portion of the video content item (e.g., a request transmitted from the user device in response to the user device determining that an amount of the video content item remaining in a buffer of the user device is below a predetermined threshold, and/or a request transmitted from the user device in response to any other suitable information). As yet another example, in some embodiments, process 100 can loop back to block 106 in response to receiving information from the user device that indicates a change in a state of streaming the video content item. Note that, in some embodiments, by looping through blocks 106 – 112, process 100 can cause the format of the video content item to be changed multiple times during presentation of the video content item by the user device.

Turning to FIG. 2, an illustrative example 200 of a process for training a model to predict a quality score associated with streaming a video content item with a particular format is shown in accordance with some embodiments of the disclosed subject matter. In some embodiments, blocks of process 200 can be executed by any suitable device, such as a server that stores and/or streams video content items to user devices (e.g., a server associated with a video content sharing service, a server associated with a social networking service, and/or any other suitable server).

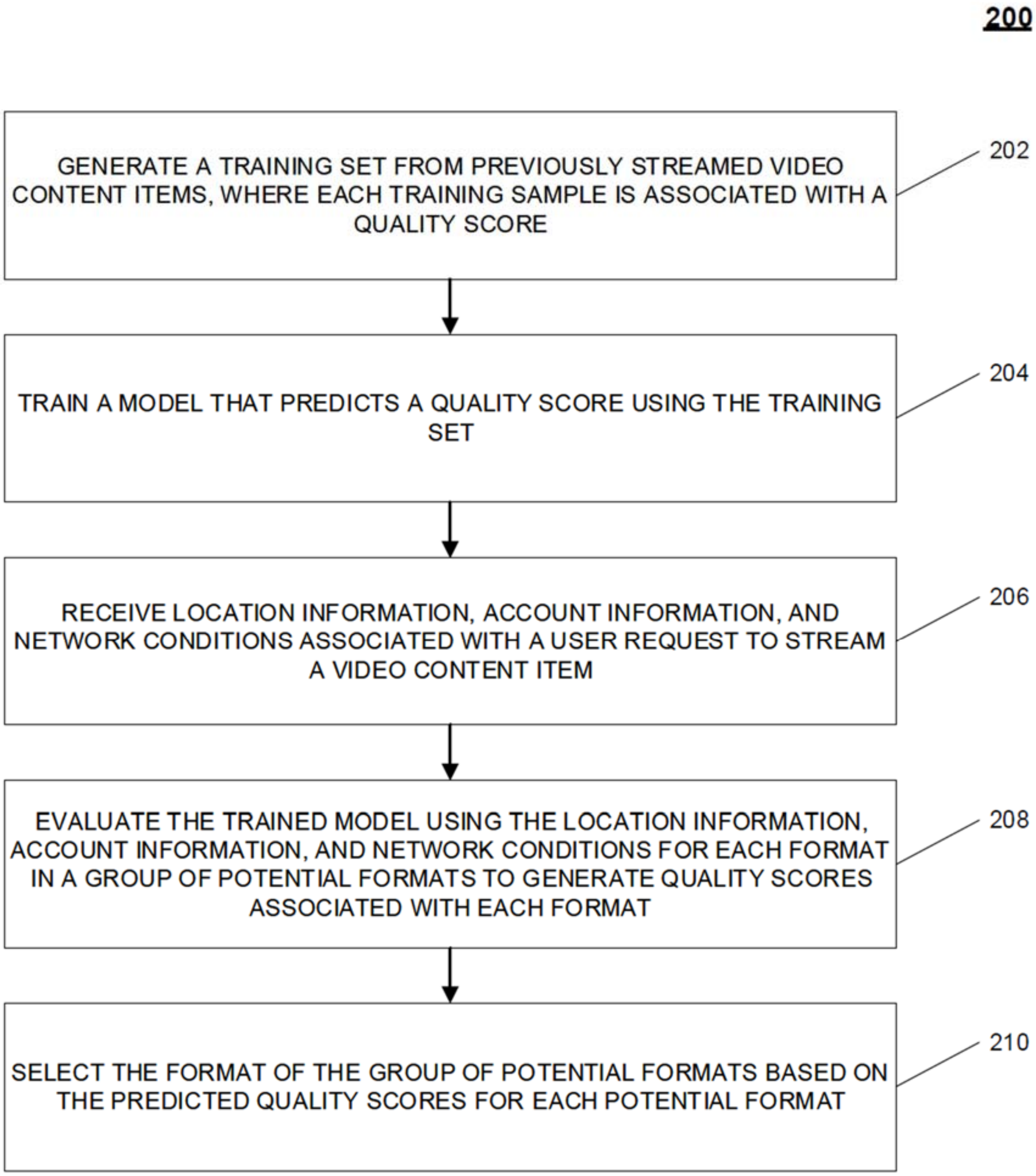

FIG. 2

Process 200 can begin, at 202, by generating a training set from previously streamed video content items, where each training sample is associated with a quality score that represents a quality of streaming of the video content item with a particular format. In some embodiments, each training sample can correspond to a video content item streamed to any device of any user of the streaming service. In some embodiments, each training sample can include any suitable information. For example, in some embodiments, a training sample can include information indicating location information to which the video content item was streamed to the user device (e.g., a GPS coordinate, IP address, time zone, local time). As another example, in some embodiments, a training sample can include account information, including previous user actions, associated with the user device used to stream the media content item (e.g., previously selected video content item formats, amount of watch time associated with the previously selected video content item formats, viewing history, information related to a billing cycle for payment of a subscription to a video content streaming service, and/or any other suitable account information). As yet another example, in some embodiments, a training sample can include information indicating the network conditions associated with the video content at the time it was previously streamed (e.g., egress traffic from servers associated with the video content streaming service to ISPs, serving attributes, designator for "peak"/"off-peak", metrics accounting for egress traffic in relation to peak egress traffic, and/or any other suitable measure of network conditions).

In some embodiments, each training sample can include a format at which the video content item was streamed to the user device. For example, in some embodiments, the format can include a resolution of the video content item at which the video content item was streamed to the user device. Note that, in instances in which the resolution of the video content item was changed during presentation of the video content item, the resolution can be a weighted resolution that corresponds to an average of the different resolutions of the video content item weighted by a duration of time each resolution was used.

In some embodiments, each training sample can be associated with a corresponding quality score that indicates a watch time metric and egress cost metric associated with streaming the video content item at the resolution (or weighted resolution) indicated in the training sample. In some embodiments, the quality score can be calculated in any suitable manner. For example, in some embodiments, the quality score can be calculated based on a watch time metric that indicates a duration of time video content items were watched. As a more particular example, in some embodiments, the quality score can be based on a total duration of time video content items were watched during a video content streaming session in which the video content item of the training sample was streamed. As a specific example, in an instance in which the video content item of the training sample was streamed during a video content streaming session that lasted for thirty minutes, the quality score can be based on a session watch time duration of thirty minutes. As another more particular example, in some embodiments, the quality score can be based on a duration of time the video content item of the training sample was watched or a percentage of the video content item that was watched prior to stopping presentation of the video content item. As a specific example, in an instance in which 50% of the video content item of the training sample was watched by a user of the user device prior to stopping presentation of the video content item, the quality score can be based on 50% of the video content item being watched.

Note that, in some embodiments, the quality score can be a function of a watch time metric (e.g., a duration of time the video content item was watched, a percentage of the video content item that was watched, a duration of time video content items were watched in a video content streaming session, and/or any other suitable watch time metric) or a function of an occupancy metric. In some embodiments, the function can be any suitable function. For example, in some embodiments, the function can be any suitable saturation function (e.g., a sigmoid function, a logistic function, and/or any other suitable saturation function.

In some embodiments, the quality score can be calculated based on an egress cost metric that indicates cost of serving the video content item at the specified resolution. In some embodiments, the egress cost metric can be calculated in any suitable manner. As a particular example, in some embodiments, the egress cost metric can be related to a historical peak egress, such as the peak multiplier as described above in connection with block 106 of FIG. 1. As a more particular example, the egress cost metric can be a dollar value based on the current egress as a percentage of the maximum peak egress within a prior amount of time (e.g., 24 hours).

As another more particular example, the egress cost metric can be calculated by a price per peak bytes based on serving attributes (e.g., rack type, quality of service, metro of request), a total amount of bytes requested, and an indication of the current network traffic as relating to peak traffic (e.g., peak multiplier based on time of day). As a specific example, in some embodiments, a video content item served to a viewer in a given metro location could be assigned an egress cost metric of 1.0 based on the viewer requesting the video content item at a given time of day with high network traffic, and an egress cost metric of 0.2 based on the viewer requesting the same video content item at a time of day with low network traffic.

At 204, process 200 can train a model that predicts a quality score based on user location information, user account information, network conditions, and/or video content information using the training set. In some embodiments, process 200 can train a model with any suitable architecture. For example, in some embodiments, process 200 can train a decision tree. As a more particular example, in some embodiments, process 200 can train a classification tree to generate a classification that a quality score is within a particular range (e.g., within 0 – 0.2, within 0.21 – 0.4, and/or any other suitable ranges). As another example, in some embodiments, process 200 can train a regression tree that generates continuous values that indicate a predicted quality score. In some embodiments, process 200 can generate a decision tree model that partitions the training set data based on any suitable features using any suitable technique or combination of techniques. For example, in some embodiments, process 200 can use any suitable algorithm(s) that identify input features along which branches of the tree are to be formed based on information gain, Gini impurity, and/or based on any other suitable metrics.

As another example, in some embodiments, process 200 can train a neural network. As a more particular example, in some embodiments, process 200 can train a multi-class perceptron that generates a classification of a quality score as being within a particular range, as described above. As another more particular example, in some embodiments, process 200 can train a neural network to output continuous values of a predicted quality score. Note that, in some such embodiments, any suitable parameters can be used by process 200 to train the model, such as any suitable learning rate, etc. Additionally, note that, in some embodiments, the training set generated at block 202 can be split into a training set and a validation set, which can be used to refine the model.

At 206, process 200 can receive location information, account information, and network conditions associated with a user request to stream a video content item. As described above in connection with block 104 of FIG. 1, in some embodiments the location information can include any suitable information related to a geographic location associated with the user device requesting the video content item, such as an IP address, GPS coordinates, time zone, and local time. As described above in connection with block 104 of FIG. 1, in some embodiments, the account information can include any suitable information related to the account being used to access the streaming service, such as previous user actions, previously selected video content item formats, amount of watch time associated with the previously selected formats, watch time statistics, viewing history, information related to a billing cycle for payment of a subscription to a video content streaming service, and/or any other suitable account information. As described above in connection with block 106 of FIG. 1, in some embodiments the network conditions can include any suitable information related to current egress traffic from servers associated with the video content streaming service to internet service providers (ISPs), such as current egress associated with a particular metro area, or measured according to specific attributes of the servers providing content (e.g., serving location, rack type, traffic quality of service), or further designated as "peak" or "off-peak" based on the time of serving and/or according to any suitable designation related to available network capacity. Additionally, in some embodiments, process 200 can receive any suitable information related to the video content item (e.g., a topic or genre associated with the video content item, a duration of the video content item, a popularity of the video content item, and/or any other suitable video content item information).

At 208, process 200 can evaluate the trained model using the location information, account information, network conditions, and/or video content item information to calculate a group of predicted quality scores corresponding to different formats in a group of potential formats. Note that, in some embodiments, the group of potential formats can include any suitable formats, such as different available resolutions of the video content item (e.g., 144 pixels x 256 pixels, 240 pixels x 426 pixels, 360 pixels x 640 pixels, 480 pixels x 854 pixels, 720 pixels x 1280 pixels, 1080 pixels x 1920 pixels, and/or any other suitable resolutions). Note that, in some embodiments, the group of potential formats can include any suitable number (e.g., one, two, three, five, ten, and/or any other suitable number) of potential formats. A specific example of a group of predicted quality scores corresponding to a group of potential formats can be: [144 pixels x 256 pixels, 0.2; 240 pixels x 426 pixels, 0.4; 360 pixels x 640 pixels, 0.7; 480 pixels x 854 pixels, 0.5; 720 pixels x 1280 pixels, 0.2; 1080 pixels x 1920 pixels, 0.1], indicating a highest predicted quality score for the format corresponding to a resolution of 360 pixels x 640 pixels.

Note that, in some embodiments, process 200 can evaluate the trained model using any suitable combination of input features. For example, in some embodiments, the trained model can use any combination of input features including any suitable location information, account information, network conditions, and/or video content item information (e.g., a genre or topic of the video content item, a duration of the video content item, a popularity of the video content item, and/or any other suitable video content item information). In some embodiments, in instances in which the trained model is a decision tree, process 200 can evaluate the trained model using input features selected during training of the decision tree.

At 210, process 200 can select the format of the group of potential formats corresponding based on the predicted quality score for each potential format. For example, in some embodiments, process 200 can select the format of the group of potential formats corresponding to the highest predicted quality score. As a more particular example, continuing with the example given above of potential formats and corresponding predicted quality scores of: [144 pixels x 256 pixels, 0.2; 240 pixels x 426 pixels, 0.4; 360 pixels x 640 pixels, 0.7; 480 pixels x 854 pixels, 0.5; 720 pixels x 1280 pixels, 0.2; 1080 pixels x 1920 pixels, 0.1], process 200 can select the format corresponding to the resolution of 360 pixels x 640 pixels.

Accordingly, mechanisms for dynamic format selection of media content are provided.