Technical Disclosure Commons

Defensive Publications Series

June 20, 2018

Intelligent ranking for photo galleries using sharing intent

John Oberbeck

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

Recommended Citation

Oberbeck, John, "Intelligent ranking for photo galleries using sharing intent", Technical Disclosure Commons, (June 20, 2018)
https://www.tdcommons.org/dpubs_series/1261

Intelligent ranking for photo galleries using sharing intent

ABSTRACT

Users often share digital photographs over communication links, e.g., via SMS, chat, URLs, etc. This disclosure describes techniques to determine and present ranked candidate images for sharing to other users and user devices. The ranking is based on a variety of characteristics of the images, including time of capture of the images, as well as other user-permitted factors that can indicate user intent, such as history of previous sharing of images by the user, the current or recent device context of the user (chat or conversation, application being used, data on the user's screen, and so on), etc. By intelligently ranking images in sharing galleries, the user's time spent searching for and sharing images can be reduced, which can improve user satisfaction and sharing frequency.

KEYWORDS

- photo sharing

- image sharing

- sharing history

- image rank

- sharing context

- photo library

- user intent

BACKGROUND

Sharing of images such as photographs is commonly performed between user devices over communication links in two primary ways. In a first way, a user finds an image that they want to share, and inputs a sharing intent to the user device to share that image, e.g., by

selecting recipient user(s) and providing a send command. In a second way of sharing, the user is using the user device (e.g., using an application on the device) and a list of candidate images to share are displayed by the device within a context on the device, e.g., in response to the user opening an application, performing an action within the application, etc. The user can select images from the list to share. For example, such context can include a person-to-person device chat conversation via a messaging application, user creation of a post on a social media platform, etc. The second method of sharing photo selection is cumbersome, especially for sharing older images. For example, currently, a gallery of the user's images is presented, ranked based on their recency (time of capture). However, such a gallery does not take into account other indications of user intent that the device or application may have access to.

DESCRIPTION

This disclosure describes techniques that provide an intelligent ranking, and scores for such a ranking, of images from a user's collection for sharing to other users. Described techniques combine intrinsic, user-based, and/or contextual characteristics of device and images, as permitted by the user, to produce a ranking score, which is used to rank images and present images based on their ranks in image galleries used for sharing. Determined ranking scores can be used, for example, to order some or all of the images in a gallery for sharing purposes, or to suggest higher-ranking images for sharing by displaying such images in a displayed area or control of a user interface. A user interface that presents images based on the ranking scores allows users to select images from the gallery to share or upload from the user device to other devices over a communication network. The described ranking scores are more effective and relevant to users than ranking images to share based only on time scores (e.g., recency of time of capture).

```
┌─────────────────────────┐
│  Train models for evaluation │
│         functions          │
│             1              │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Display of gallery of    │
│ sharing images triggered   │
│             2              │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│  Select a particular image │
│       from gallery         │
│             3              │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│    Determine scores for    │
│ particular image based on  │
│    evaluation functions    │
│             4              │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│  Determine overall ranking │
│  score for particular image │
│             5              │
└─────────────────────────┘
            │
            ▼
          ◇ Another ◇
   YES ◇  image in  6  ◇
          ◇ gallery? ◇
            │ NO
            ▼
┌─────────────────────────┐
│   Order images based on    │
│ overall ranking scores and │
│  display in ranked order   │
│             7              │
└─────────────────────────┘
```
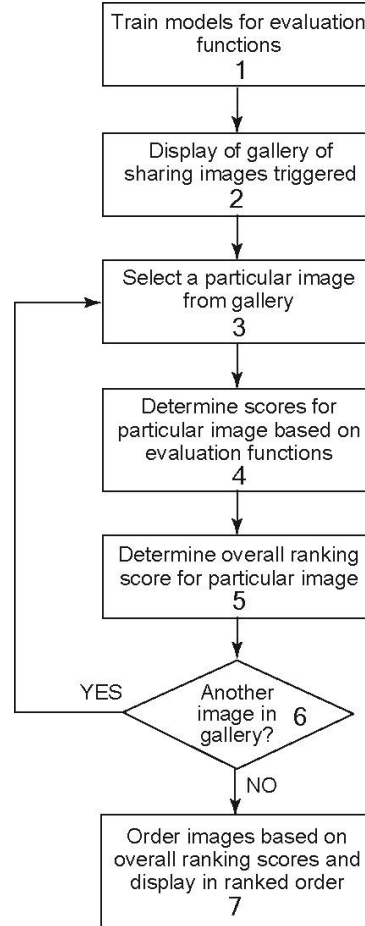
Fig. 1

Fig. 1 illustrates an example technique for determining a ranking order of images in a gallery and presenting the images based on the ranking order, to provide a selection of images to the user which can be shared to other users.

In step 1, one or more machine learning models are trained to provide evaluation functions used to determine rankings of images within a set of candidate sharing images. In the

described example, these functions can include a photo-based function, a user-based semantic function, and a context-based semantic function.

The models are trained and implemented upon user permission to access user data that serves as input to the models. Users are provided with options to indicate their permission or denial of permission for access to various data, e.g., photos and other content in the user's image library (which serve as input to the photo-based function), user-specific factors such as prior sharing history, previous device actions, etc. (which serve as input to the user-based semantic function), and contextual factors such as time, location, application in use, etc. (which server as input to the context-based semantic function). In using the models, use is made only of user-permitted data. One or more of the models are not implemented, if users deny permission. In various implementations, training can be performed based on generalized data that is not attributable to particular users, and/or training can be performed only locally on a user device with user data, e.g., a federated learning approach.

The functions are trained to indicate how likely a given image is to be shared by the user of the device, based on particular features including characteristics of the given image, the given device used by the user, and/or previous device actions of the user. Examples of features used to train each of the evaluation functions are described below.

Various different functions can be trained and used in different implementations. For example, two or more of these evaluation functions can be combined in to a single function, one or more of these evaluation functions can be divided into multiple functions, etc.

In step 2, an event occurs on a device that triggers a gallery of candidate sharing images to be displayed by the device to allow the user to select images to share. For example, this can occur at a time when the user is using a given device. Applications that provide such sharing

image galleries include applications providing user chat, social media, image viewing, messaging, videoconferences, voice calls and video calls, email, note taking, reminders, web browsers, etc. For example, the event may be that the user has commanded to create a post or message in a messaging application to send to another user, or the user has sent a message in a chat application indicating that the user may wish to share an image in the chat conversation.

In step 3, a particular image is selected from the gallery to process.

In step 4, various scores are determined for the particular image based on the evaluation functions and any other functions. For example, a respective score is provided by a time-based function, the photo-based function, the user-based semantic function, and the context-based semantic function, in an inference stage of the trained machine learning models. These functions are described in greater detail below.

The sub-scores from the photo-based function and the time-based function can be determined on a server or on the user device. The user-based and context-based semantic sub-scores can be determined by the user device, as such scores may rely on information that is only available locally.

In step 5, an overall ranking score for the particular image is created by combining the sub-scores together. The sub-scores can be combined into the ranking score in many different ways. The sub-scores can be combined linearly with weighting constants to determine the overall ranking score. In an example, the ranking score for the particular image I is calculated in the following manner:

```
RS(I) = w_p * PS(I) + w_c * CSS(I) + w_u * USS(I) + w_t * TS(I)
```

where the weighting constants sum such that: `w_p + w_c + w_u + w_t = 1.`

The values of the weighting constants can be determined experimentally by reviewing aggregate data of image sharing actions performed on devices by users (without determining

actions of individual users). For example, the values can be based on which values are correlated with the greatest number of image sharing actions performed per user per unit of time (e.g., per week). For example, the behavior of typical gallery sorting, based only on capture time, can be modelled as:

```
RS(I) = 0 * PS(I) + 0 * (CSS(I) + 0 * USS(I) + 1 * TS(I)
```

In step 6, it is determined whether there are any more images in the gallery to process for ranking scores. In various examples, all of the images in the gallery can be processed, or a subset of images can be processed. If another image is to be processed, then process flow returns to step 7, a different image is selected from the gallery to process and step 4 to determine sub-scores and determine an overall ranking score for the image.

If there are no more images in the gallery to process as determined in step 6, then in step 7, the images are ordered based on their overall ranking scores and are displayed in the gallery in the determined order. In various examples, all of the ranked images can be displayed, or a subset of the images can be displayed, e.g., the top N ranked images, or N requested images. In further examples, top ranked images can be provided as suggestions, e.g., displayed within user interface elements such as buttons.

In some examples, steps 3 to 7 can be pre-processed, at least in part, prior to a trigger of the display of ranked images as in step 2.

An operating system can utilize the ranking scores for presented image galleries on the device, allowing all applications to take advantage of improvements in the image sharing experience without individual developer effort on a per-application basis.

Different functions utilized for scoring are described in more detail below. Each of these functions is implemented only upon user permission, e.g., to access the input factors for the function.

Time-Based Function

The time-based function determines a time-based score using a time of capture of the particular image. This can be a recency function, where more recent images have a higher time-based score than images which are older. The time of capture of an image can be determined from image metadata stored with the given image or stored in association with the given image.

Photo-Based Function

The photo-based function determines a photo-based score based on both intrinsic qualities of a given image and the relation of the image to other images. The photo-based function measures how likely a given image is to be shared based on characteristics of the given image, including its relation to other images of the user's collection. In some examples, a general-purpose neural net model (machine learning model) can be used to implement the photo-based function. This model can be trained based on these characteristics (features). The training data for this model can be, for example, the data set of existing user images used in an application on user devices, and the rate at which these images are shared from the application by users of the application manually finding and selecting these images.

Some examples of features used to train the photo-based function to determine a photo-based score for a particular image are described below:

1) The number of images that are in a group G having visually similar images (i.e., |G|, the order of G). This value represents the number of images which are highly visually similar to a given image, and can be used as an indicator of the desirability of the image to the user. For example, a higher value of |G| represents a greater desire by the user to

capture a high-quality image within G, since the user captured multiple photos of a particular subject. For example, users commonly capture "bursts" of several images of a subject using a camera, e.g., in order to increase the probability that one of the captured images is a high-quality image. In some examples, images in the group G can be considered images a user has captured in a single session (e.g., captured within a short threshold time period of each other). Similarity of images can be determined using any of a variety of image similarity detection techniques.

2) A ranking of an image within the group G. The ranking is based on particular criteria. For example, the images in the group G can be ordered based on their visual quality scores as determined by a visual quality ranking function, and the rank of a given image within that order can be used to train the photo-based function. The visual quality ranking function can be used to determine image visual quality and provides a visual quality score or measure. For example, the visual quality ranking function can use particular image characteristics as parameters, such as the resolution of an image, presence of blur, color noise, image tilt or rotation, presence of high-dynamic range (HDR), presence of face(s) in the image, whether human faces have their eyes open, and/or other image characteristics considered relevant to image visual quality.

3) The location of image capture. For example, this characteristic can take the form of a specific geographic location (e.g. latitude / longitude coordinates), which can be provided in metadata of the image. This characteristic can also take the form of a categorization of the location, such as "park" or "beach," which can be determined using a machine learning model, coordinates and map data, object detection techniques to detect characteristic objects of a location, etc. (if user consent has been obtained).

4) The subject of the given image. The subject(s) of the image can be indicated by descriptors or tags as "person," "face," "sunset," "meal," etc. Image detection and recognition techniques can be used to determine the subjects of the image, if user consent has been obtained.

5) Time parameters based on the time of capture of the given image. For example, the time of capture can be converted into a form of the time of day at which the given image was captured, e.g., "morning" or "afternoon," where each time of day covers a particular time range within a day time period. In another example, the time of capture can be converted into a time of year at which the given image was captured, such as the season ("summer") or holiday ("Halloween"). For example, users may be more likely to share images captured at certain times of year. In some cases, particular anniversary dates specific to the user can be used to contribute to the photo-based score, e.g., a marriage anniversary date, if user consent has been obtained.

User-based Semantic Function

The user-based semantic function determines a user-based semantic score based on previous user interaction with images on the given device. The user-based semantic function measures how likely a particular image is to be shared based on characteristics including a user's previous actions on the device, such as previous sharing of images and/or views of images. In some examples, a variant of a Long Short-term Memory (LSTM) model (machine learning model) can be used to implement the user-based semantic function. This model can be trained on these characteristics (features) based on a stored history of user actions on the user device, if user consent has been obtained.

The features used to train the user-based semantic function reference particular constructs. These constructs include:

1) The current context of the device (at the time of sharing). This includes information such as the application being executed, the time, and other information associated with the application. For example, in chat-based applications, the device context includes information such as user names of the other users (e.g., contacts of the user) who are included in the chat, and the previous N messages sent by users in the chat conversation.

2) A first set of sharing actions previously performed by the user on the given device, that include the given image. These sharing actions can include previous commands causing the given image to be sent from the given device to other users and their devices.

3) A second set of sharing actions previously performed by the user on the given device. This includes all sharing actions for images, regardless of whether those sharing actions included the given image.

Some examples of features used to train the user-based semantic function to determine a user-based semantic score for a particular image are described below. In some examples, the features used to train the photo-based function can also be used in the training of the user-based semantic function.

1) The time of the most recent sharing action of the user. This characteristic is respectively determined for both first and second sets of sharing actions described above, e.g., for the set of sharing actions on the given device including the given image, and for the set of sharing actions on the given device including any image. For example, the time of previous sharing actions may be indicative of times when the user prefers to share images.

2) The number of previous sharing actions of the user. This characteristic is respectively determined for both first and second sets of sharing actions described above, e.g., for the set of sharing actions including the given image, and for the set of sharing actions including any image. For example, the number of previous sharing actions may be indicative of the frequency of user shares.

3) The time of the most recent view of the given image on the device. For example, this characteristic may be indicative of recent interest of the user in the given image.

4) The time of the most recent sharing action on the device of the given image within the first set of sharing actions on the device. This characteristic may be indicative of recent sharing activity of the user for the given image.

Context-based Semantic Function

The context-based semantic function determines a context-based semantic score for a particular image based on the user's current context on the device. The context-based semantic function measures how likely a particular image is to be shared based on characteristics including states of the user's device and user interactions on the device during those states.

In some examples, a model trained as a multi-loss model can be used to implement the context-based semantic function. In an example simple form, such a context-based semantic function can provide a binary 0 or 1 output value depending on whether the subject detected in the image (e.g., detected as a photo-based function characteristic described above) is included in the preceding N messages of the current context for chat applications, where N can be a number such as 2, 3, etc. For example, if a user contact sends a message to the given user, "how are you coping with the new dog?", a value of 1 can be returned as the context-based

semantic function score for images that have "dog" as a subject, and a value of 0 returned for all other images.

The features used to train the user-based semantic function reference particular constructs. These constructs include:

1) The current context of the device (at the time of sharing). This can be similar to the context described above for the user-based semantic function. Also, this context may differ from the context above in that it can include more transient information, such as current and/or recent screen content (e.g., messages, notifications, text, video, etc. that are displayed on the screen).

2) A first set of sharing actions previously performed by the user on the given device, within the application of the current context on the given device. This includes all sharing actions for all images. These sharing actions can be similar as described above for the user-based semantic function.

3) A second set of sharing actions previously performed by the user on the given device, that include the given image, and which are within the application of the current context on the given device.

The following set of features, as well as the features described above used to train the photo-based function, can be used to train a multi-loss model for the context-based semantic function:

1) The other user that the user is messaging within the current context (if the context application being used is a messaging application). If multiple other users are communicating, e.g., in a multi-user chat, then all of the other users can be included in

the feature. For example, the user may share images of different subjects with one user than with a different user.

2) The most recent message the user has sent in the current context (if the context application being used is a messaging application).

3) The most recent message the other user has sent in the current context (if the context application being used is a messaging application).

4) The second-most recent message the user has sent in the current context.

5) The number of previous sharing actions of the user, performed within the application of the current context on the given device. This feature is respectively determined for both first and second sets of sharing actions described above, e.g., for sharing actions including the given image, and for sharing actions including any image.

6) The time of the most recent sharing action of the user in the application of the current context.

Since the model for the context-based semantic function is trained both with these features and the features above for the photo-based function, the context-based semantic score output by the context-based semantic function may correlate characteristics of images such as subjects or location with device context such as other users and recent messages sent. For example, a user may be more likely to share images of certain image subjects with other users of that user's family in a chat conversation. In another example, a location feature can be correlated to the device context, where a user in a chat conversation mentions a location in the conversation, so an image captured at that location can be scored higher.

<u>Further Examples</u>

Applications can be developed to determine and present particular layouts of sharing image galleries using the ranking signal provided by the techniques herein. In some cases, ranking scores and time scores can both be used. Some potential layouts include: a) a gallery sorted by decreasing time of capture scores (most recent photos ranked highest); b) a gallery sorted by decreasing ranking scores; c) a gallery sorted by decreasing ranking scores with the option to toggle the sorted value between ranking score and time score; d) a multi-modal gallery with N images displayed with the highest ranking scores, followed by all images sorted by time scores; and e) a multi-modal gallery displaying all images with ranking scores above a threshold t, followed by all images sorted by time scores.

To increase efficiency, it is possible to "lazily evaluate" the scores by the evaluation functions, e.g., only use the evaluation functions to determine scores for images having a time-based score above a certain threshold. For example, if an image is too old, then it is not evaluated by these functions. Other thresholds can be similarly applied to use of the evaluation functions.

In some cases, a subset of the gallery images can be evaluated to obtain overall ranking scores for those images, and these images can be displayed first in the galley based on the overall ranking scores. The remaining gallery images can be displayed after this subset, e.g., based on different criteria such as time of capture (e.g., most recent of these remaining images displayed first).

Different ways of providing image ranking scores for candidate sharing images can be used. For example, a ranking technique can combine the above-described overall ranking score with a known recency ranking.

The features used to train the models for the evaluation functions can be used to train any of the models and/or other models. Furthermore, additional or different features can be utilized in the evaluation functions. For example, the photo-based function can determine a score for an image based on a particular type of subject (e.g., babies) present in the image, based on other visual characteristics (e.g., saturation, brightness, etc.), etc. The context-based semantic function can determine a score based on whether a respondent in a conversation has responded with an image of their own, etc.

Other approaches that those described above can be used to train or produce each of the ranking functions. For example, other approaches such as logistic regression can be used.

Additional evaluation functions can be used in addition to the evaluation functions described above. For example, a function can evaluate a corpus of shared images in which user-generated votes or ratings are associated with each image. A sub-score can be generated for each image in the gallery based on a similarity of the image to other images in the corpus having high amounts of votes, or high ratings.

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 for providing additional ranking scores for ranking images displayed in a gallery on a user's device. The combination of different types of ranking scores that are intrinsic to the image itself (such as time-based and photo-based scores), based on a user's previous behavior (such as user-based semantic scores), and based on the user's current context (such as context-based semantic scores), allows images to be ranked based on the content of the images and based on user behavior regarding sharing of images. Such ranking allows images to be displayed, e.g., at the top of a gallery of candidate images, which are more relevant to user intent for sharing. This eases and simplifies the process of selecting such images.

The additional described features used in determining the ranking scores of images to share can improve the sharing experience for users. For example, the improved ranking scores make it easier to find a desired image to share, decreasing the time-to-complete for the user action of sharing an image. Second-order effects of decreasing the time-to-complete include an increase in number of shared images and an increase in the satisfaction of users of a device or service as related to image sharing. Furthermore, the models for ranking a user's personal images within the context of a conversation can be used in other media types.