

Technical Disclosure Commons

Defensive Publications Series

August 2020

Improving Object Recognition Using Post-Recognition Actions

Jared M. Zimmerman

Mai Tran

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

Recommended Citation

Zimmerman, Jared M. and Tran, Mai, "Improving Object Recognition Using Post-Recognition Actions", Technical Disclosure Commons, (August 28, 2020)
https://www.tdcommons.org/dpubs_series/3555



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Improving Object Recognition Using Post-Recognition Actions

ABSTRACT

This disclosure describes techniques that, with user permission, utilize user actions as implicit feedback for image recognition. With user permission and express consent, post-recognition user actions are utilized to adjust image recognition model(s). An image recognition request and one or more image(s) are received. Image recognition techniques are applied to the received image(s) and results are provided to the user. Based on actions performed by the user post-recognition, various factors are inferred and utilized to train image recognition model(s) to improve the performance of the model for future searches. With permission, user affinity model(s) are trained that can provide customization of future search results for the particular user.

KEYWORDS

- Visual search
- Image search
- Image recognition
- Object recognition
- Search engine
- Affinity score
- Affinity model
- Image-based shopping

BACKGROUND

Image recognition techniques are used in various contexts such as for standalone image-based search as well as to supplement voice or text based search. Text-based searches may be performed by a user on any device such as a smartphone, computer, or other device that supports text input, e.g., by entering a search query in a search bar user interface. Voice-based searches may be performed via devices such as smart speakers/ appliances, smartphones, computers, or other devices that support voice input.

Receiving user feedback for voice or text-based search results helps improve search performance, including creating user-specific customizations. Analysis of a user journey is a common technique for evaluation of typed or spoken user queries. For example, a user provided search term “jaguar” may refer to a member of the cat family just as well as a car brand. In such situations, user responses to search results provided to the user, e.g., the user repeating the search using a modified search term (e.g., “jaguar cars”), clicking on a particular search result (e.g., the image of a jaguar), etc. are utilized to provide feedback to a user affinity model which can be utilized to adjust the model for subsequent queries, e.g. in ranking of search results, etc.

By contrast, there are no easy mechanisms for incorporating user feedback for image-based searches where the search query includes an image, and where object recognition techniques are utilized to identify search results. Users cannot easily provide feedback for such search results, e.g., since query modification is typically not possible. While mechanisms for explicit user feedback can be utilized, e.g. by querying the user to submit an evaluation response to a performed image recognition task, such mechanisms are burdensome and infrequently used.

DESCRIPTION

This disclosure describes techniques that, with user permission, utilize user actions after image recognition results are delivered as implicit feedback. User actions are utilized as implicit feedback to adjust image recognition model(s) and/or user preferences. Based on the specific permission for use of implicit feedback, model updates are made to improve image recognition in future image recognition tasks across users and/or to personalize image recognition results for future queries from the particular user.

The described techniques can be utilized in a variety of applications that perform object recognition. For example, some applications enable users to point their camera at a particular object, e.g., restaurant or movie theater signage, and based on the image in the camera field of view, obtain reviews or information about timings, make reservations, etc. Some applications may provide image-based search and/or shopping features whereby a user can provide an image of an object of interest, and is provided with links to shopping site(s) where the object of interest can be purchased.

Implicit feedback derived from the user's post recognition actions can be utilized in addition to explicit feedback mechanisms that may be in place that enable a user to provide direct feedback on the quality/accuracy of a completed image recognition task. To obtain explicit feedback, feedback mechanisms are provided to users upon completion of an image/object recognition workflow. Additionally, various implicit factors such as improved time to a successful query (indicative of a user learning new terms for an object/image), successful purchase of an item (indicative of accurate image recognition), user submission/request for a new similar image analysis (indicative of inaccurate image recognition), etc. may be inferred from post-recognition user actions. The factors are utilized to train image recognition models for

future recognition tasks. The implicit factors can be determined based on user permitted data such as browsing patterns, user purchases, search terms in future queries, etc. that are indicative of success in image identification.

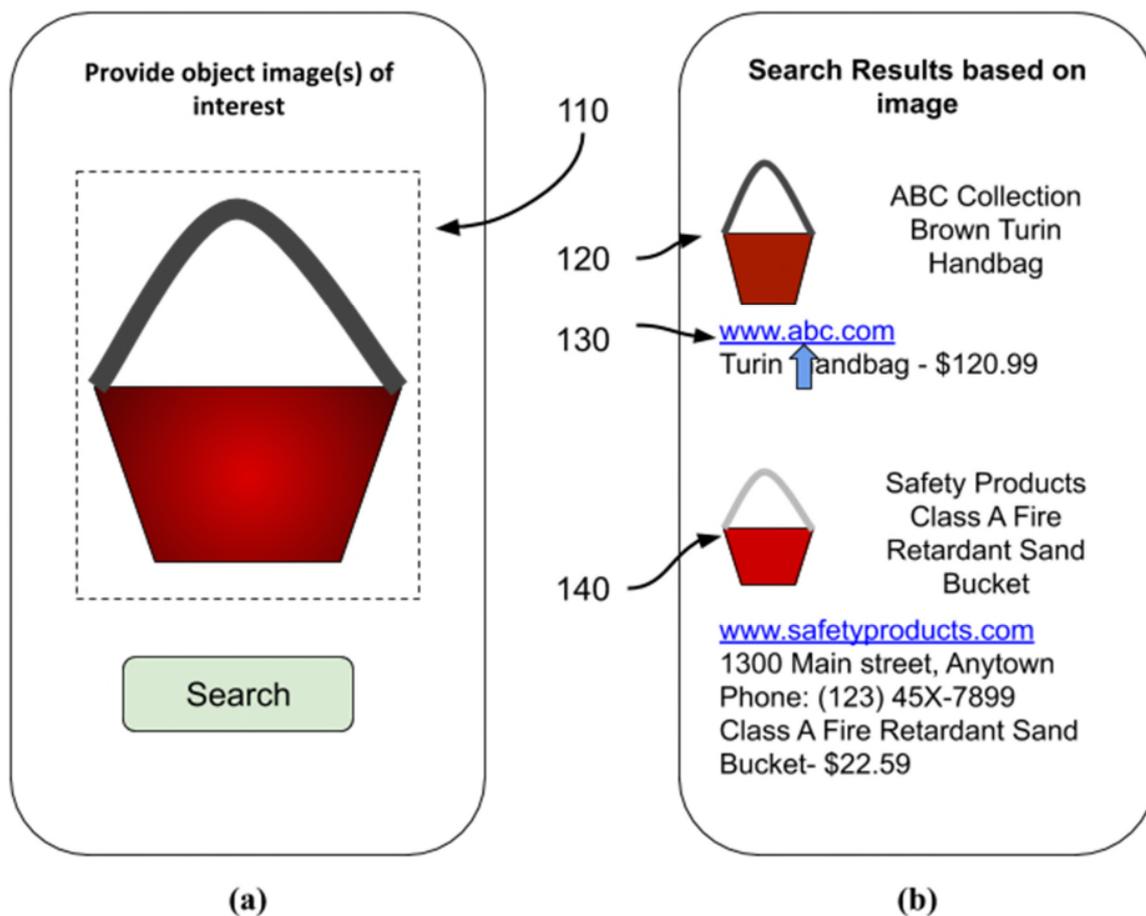


Fig. 1: User responses to image recognition tasks are utilized as implicit feedback

Fig. 1 depicts an example application that performs an image recognition task in service of a user goal. In this illustrative example, the user has provided an image (110) of an object of interest. For example, the image can be a live view from a camera of a user device, a stored image, or any other image. Object recognition techniques are applied to recognize the image and corresponding search results are provided to the user via a user interface (UI). The image may be

an image of an object that the user sees in real life, an image from an article that the user is reading on their user device, or any other image.

In this illustrative example, as depicted in Fig. 1(a), a user provides (e.g., a stored image on their device, a live camera view, an image found on the internet, etc.) the image of an object of interest. Upon completion of object recognition, search results are identified and displayed on the user device, as depicted in Fig. 1(b). The search results include a first result (120) corresponding to a handbag (“ABC Collection Brown Turin Handbag”) and a second result (140) of a fire retardant sand bucket (“Safety Products Class A Fire Retardant Sand Bucket”).

In this illustrative example, the user intent is to search for a handbag (as indicated by image 110) and not a sand bucket. Thus, upon display of the search results, the user clicks on the provided link (130) to obtain additional information about the handbag, indicating that the first search result matched the user intent. The result can include additional information, e.g., a link to purchase the handbag, reviews for the handbag, etc.

Per techniques of this disclosure, the user’s action of clicking on the link associated with the handbag is utilized as an implicit (positive) feedback regarding the image recognition result returned based on the provided image. A further action, e.g., the user making a purchase of the handbag, can be utilized as a still stronger positive feedback for the returned image recognition result. The feedback can be utilized to adjust an image recognition model as well as adjust a user affinity model associated with the particular user. Such adjustments can provide improved results for future searches by the particular user as well as other users.

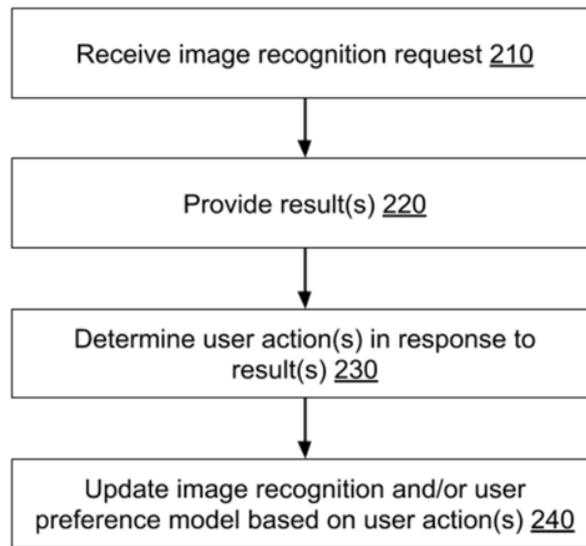


Fig. 2: Utilization of post image-recognition user action(s) as feedback

Fig. 2 depicts an example process for utilizing post image recognition user action(s) as implicit feedback, per techniques of this disclosure. With user permission, the process may be utilized in any recognition based application, e.g. image-based shopping, facial recognition based digital photo gallery, image-based search, etc.

The process commences with an image recognition request and one or more image(s) being received (210), e.g. from a user via their user device. Image recognition techniques (e.g., that utilize image recognition models) are applied to the received image(s) and corresponding results provided (220) to the user. Depending on the user context and permissions, image recognition may be performed locally on the user device, and/or on a server to which the image(s) are transmitted.

With user permission, user actions performed by the user in response to being provided results from the image recognition are determined (230). For example, it may be determined that the user submitted a fresh query, e.g. image, voice, or text query, in response to viewing the provided results. In another example, it may be determined that the user clicked on a link

provided in the results, e.g., a link that identifies an object detected in an image using image recognition. In another example, it may be determined that the user clicked on a link provided along with the results, and subsequently made a purchase of an item associated with the image.

The feedback thus obtained is utilized to train (adjust) image recognition model(s) that can improve the performance of future searches. If the user permits use of feedback data for personalization, the feedback can also be used to train (adjust) user affinity model(s) that provide personalization/customization of future search results for the particular user. For example, a preferred class of objects associated with the user based on previous image-based searches can be utilized to determine a likely class of an object in a current image search performed by the user. This can improve the quality of search results provided to the user, particularly for images that are associated with a low confidence level result.

If the user permits, implicit user feedback can be utilized to improve generalized user models. For example, previously obtained user feedback can be utilized to infer that users have a general preference for a certain class of goods, e.g. home goods, natural world, etc., when performing an image-based search. Such inferences can be utilized to rank and improve provided search results.

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 that, with user permission, utilize user actions as implicit feedback for image recognition. With user permission and express consent, post-recognition user actions are utilized to adjust image recognition model(s). An image recognition request and one or more image(s) are received. Image recognition techniques are applied to the received image(s) and results are provided to the user. Based on actions performed by the user post-recognition, various factors are inferred and utilized to train image recognition model(s) to improve the performance of the model for future searches. With permission, user affinity model(s) are trained that can provide customization of future search results for the particular user.