Summarization of Online Image Collections via Implicit Feedback

Shane Ahern, Simon King, Mor Naaman, Rahul Nair
Yahoo! Research Berkeley
Berkeley, CA, USA
{sahern, simonk, mor, rnair}@yahoo-inc.com

ABSTRACT

The availability of map interfaces and location-aware devices makes a growing amount of unstructured, geo-referenced information available on the Web. In particular, over twelve million geo-referenced photos are now available on Flickr, a popular photo-sharing website. We show a method to analyze the Flickr data and generate aggregate knowledge in the form of "representative tags" for arbitrary areas in the world. We display these tags on a map interface in an interactive web application along with images associated with each tag. We then use the implicit feedback of the aggregate user interactions with the tags and images to learn which images best describe the area shown on the map.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

algorithms, human factors

Keywords

image summarization, implicit feedback, geo-referenced data, photographs, tagging, geotagged, visualization

1. INTRODUCTION

The amount of geographically-annotated material available on the Web is constantly growing. The existence of location data allows the generation of interesting location-driven *aggregate* knowledge from these unstructured text-based collections: when enough information is available, systems can identify useful location-driven trends and patterns in the text data.

More than any other resource, geo-referenced (or "geo-tagged") photographs are commonplace on the Web today. Our focus in this work is Flickr [1], a popular photo-sharing website. Flickr allows users to associate photos with multiple tags (unstructured textual labels) and with a location (most commonly done by users "dragging" photos onto map locations where the photos were taken).

Our application, World Explorer, considers all geotagged photos on Flickr and generates an aggregate representation that allows navigation, exploration and understanding of the underlying data (and the world). The algorithm is based on the work of Jaffe et al. [2] and uses a TF-IDF approach to determine the most "descriptive" tags in a given geographical area. The application was designed as an interactive prototype that would encourage people to freely explore any place in the world and to serendipitously discover interesting locations and photographs. The World Explorer allows users to examine both familiar and unfamiliar locations.

The main navigational component of the visualization is a map. The user can pan, zoom (both in and out), or directly enter a location in a search box to navigate the map. The principal elements of the visualization are the "primary tags" (Figure 1) which are overlaid by the application on top of the viewed map area. The primary tags are meant to give the user an idea regarding the landmarks, points of interest and other items available in the viewed area. Whenever the user changes the focus of the map, the application retrieves the relevant primary tags for the new area and zoom level from the server. These primary tags are the top-scoring tags that represent the viewed area according to our analysis; the font size is indicative of the tag's relative importance.

The principal interaction method in the application is clicking on a tag. When a user clicks one of the tags shown on the map, the application loads 20 public Flickr photos that were annotated with that tag from the geographic region where the tag appears; i.e., photos that visually explain and extend the tag information. In Figure 1, the user clicked on the `Golden Gate Bridge` tag to see related photos. The photos are laid out in a random layout and ordering that provides an aesthetically pleasing view while intentionally obscuring the details of some of the images. Once the photos are displayed, any photo can be expanded to be examined in more detail by double-clicking on it. When expanded, the image is shown in correct rotation, together with additional metadata such as the photo title and the name of the user who took that photo. Users can also close a photograph by clicking on a red "X" icon on the top right corner of the photograph.

2. ANALYSIS

Traditional image summarization algorithms rely on feature extraction to examine the low level features of the image in question. We instead analyzed the pattern of user interaction with the World Explorer to see if we could extract information about the images associated with various tags. We based our analysis on two heuristics: 1) If a user examines a photograph in more detail, it is more likely that that photograph is relevant to the tag/location being examined. 2) If a user 'closes' a photograph there is a greater proba-

WWW2007, May 8–12, 2007, Banff, Canada.
ACM 978-1-59593-654-7/07/0005.

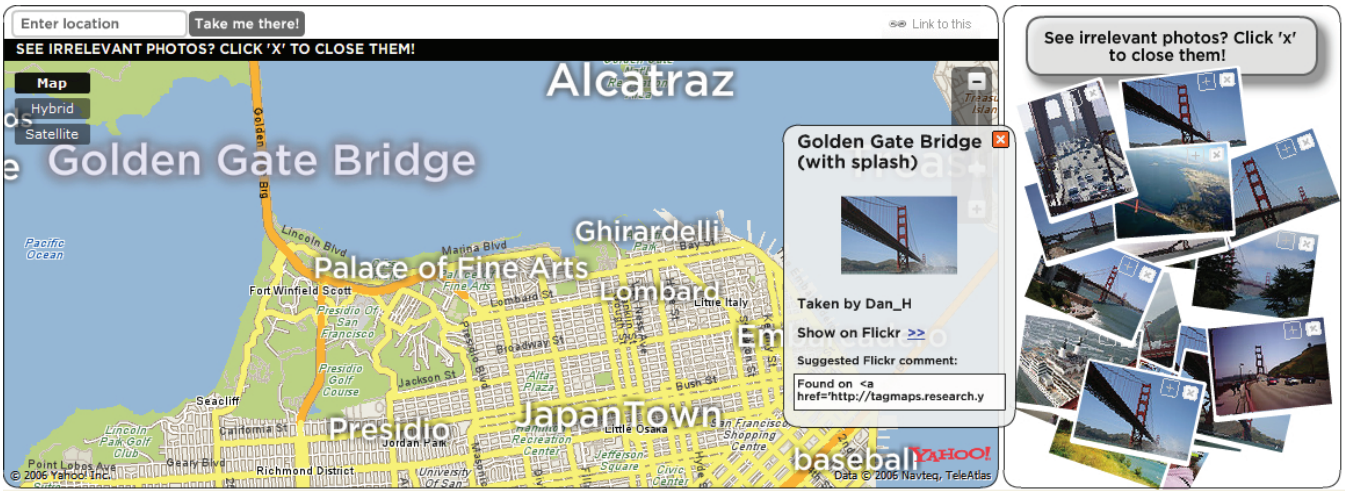

Figure 1: A screenshot from the World Explorer visualization, showing parts of San Francisco; the user highlighted the tag Golden Gate Bridge to bring up photos with that tag from that area, and then selected one of the photos to get an expanded view.

bility that the photograph is not representative of the tag in question. Implicit feedback from user interaction is not a new topic; other systems have utilized such data before, e.g. using clickthrough data in web search [3]. However, since our interaction is constrained by both a location and a tag, the feedback is more relevant than in web search, for example, where user intentions are not always known. While our approach is somewhat similar to the work of von Ahn et al. in the ESP Game [4], where user interaction is captured and utilized, we capture interaction that is rooted in the actual application, and does not require a game settings.

A total of 2405 people have interacted with the World Explorer application over 21 days. For the analysis here, we concentrate on the individual photos that users interacted with the most: photos that were expanded at least 4 times, and photos that were closed at least three times. The photos were coded into 4 categories: 1) *Representative*: photos that accurately represent the tag. 2) *Performers*: Candid photos of professional artists that are related to the tag (dancers, music bands, street actors, etc...). 3) *Portrait*: Photos of people posing for the camera. 4) *Non-representative*: Photos that are not representative of the tag. Of the 73 photographs that have been examined by at least 4 users in the detail view (i.e., providing positive feedback), we found that 60 (82%) of the images were representative of the tag in question. In this set there were 4 (5%) images of performers as well as 3 portraits and 6 non-representative images (4% and 8% respectively). On the other hand, examining the 46 photographs that have been closed by at least 3 users shows 8 portraits (17 %) along with 7 (15%) other non representative images. The number of performers also rises to 8% (6 images) while the number of shots that are deemed "representative" drops to 57% (26 images).

These numbers initially suggest that users are more likely to examine images that are representative of the tag in question, and will usually ignore portrait images and other photos that are not representative of the tag in question. Also, the numbers indicate that users are more inclined to close portrait and non representative images than they are to examine them in detail. The number of "performer" images suggests a split between users – there is no clear indication of positive or negative feedback for this type of images. Indeed, "performer" images could be, sometimes, representative of the examined tag.

3. CONCLUSION

We have outlined a way to utilize user interaction with a map based image browsing tool to generate image summaries of geographical areas. By utilizing the implicit positive and negative feedback (rooted in the application itself), we can eventually reach a community consensus as to which images best describe a tag and a location. While initial result indicate that the system is likely to identify representative photos, we have not yet tested an additional assumption: that better-quality images can be identified in a similar manner. This hypothesis requires a more careful annotation of the dataset, and a robust way to manually decide on a photo's quality measure.

4. REFERENCES

[1] Flickr.com. http://www.flickr.com.

[2] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia Information Retrieval*, pages 89–98, New York, NY, USA, 2006. ACM Press.

[3] T. Joachims. Optimizing search engines using clickthrough data. In *KDD '02: Proceedings of the 8th ACM SIGKDD international conference on Knowledge Discovery and Data mining*, pages 133–142, New York, NY, USA, 2002. ACM Press.

[4] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 319–326, New York, NY, USA, 2004. ACM Press.