

Adaptive Search Based On User Tags in Social Networking

Ch.Priyanka¹, Dr. P. Govardhan²

¹(Computer Science and Engineering, Vardhaman College of Engineering, India)

²(Information Technology, Vardhaman College of Engineering, India)

Abstract: With the popularity of the network and development of multimedia retrieval and relevant technology, the traditional information retrieval techniques do not meet the users' demand. We empirically validate this approach on the social photo-sharing site Flickr, which allows users to annotate images with freely chosen tags and to search for images labeled with a certain tag. In the existing system we are going to use the content based image retrieval methodologies which were not sufficient to retrieve the required image so, we are going to propose a new methodology called as tagging. The results of our study show that it is possible to use social tags to improve the accessibility. We use metadata associated with images tagged with an ambiguous query term to identify topics corresponding to different senses of the term, and then personalize results of image search by displaying to the user only those images that are of his/her interest.

Keywords—Tagged image search, Topic model, Image retrieval, Tagging, Reranking.

I. Introduction

As human beings get image sound and any other information by seeing, hearing, perception and analysis. Human judge similarity of images and sounds according to their semantic contents, for instance the searching for a star's picture is based on his facial characters or other contents. So the retrieval methods based on text or keywords for the digital multimedia apparently can't meet the demand that human being get multimedia information exactly. With more and more multimedia information appear on the Internet and other digital multimedia as well as human beings thirst for exact and fast retrieval, we will go deep into this area in this paper which is based on contents multimedia information retrieval becoming the focus of the academic research, as well as image retrieval of contents is one of the important study aspect of multimedia information retrieval. Rather than simply searching for, and passively consuming information, users of blogs, wikis and social media sites like del.icio.us, Flickr and dig, are creating, evaluating, and distributing information. In the process of using these sites, users are generating not only content that could be of interest to other users, but also a large quantity of metadata in the form of tags and ratings, which can be used to improve Web search and personal.

Previous literature has considered image reranking methods aimed at dealing with noisy metadata with the goal of promoting relevant content to the top ranks. A common strategy is to select a group of relevant images from the original result set, and learn content-based models to select similar images Nevertheless, the increasing size of collections poses an additional challenge, when working at very large scale, and the chances of having too many assets similarly relevant to the original query grow. For instance, querying for "dog" would find thousands of relevant images in typical Web image datasets. Increasingly sophisticated ranking and reranking schemes solely based on relevance can deal with the problem only to a certain extent. When too many relevant resources exist in the dataset, additional parameters need to be considered for ranking search results.

The image retrieval that is based on artificial notes labels images by using text firstly, in fact it has already changed image retrieval into traditional keywords retrieval. There are two problems remain in this method. On the one hand, it brings too heavy workload. On the other hand, it still remains subjectivity and uncertainty. Because the image retrieval that is based on artificial notes still remains insufficiency, the further study that adapts vision image features has been come up and become the main study. The character of this method is image feature extraction impersonally, whether the retrieval is good or not depends on the accuracy of the features extraction. So the research based on vision features is becoming the focus in the academic community. The feature of vision can be classified by semantic hierarchy into middle level feature and low-level feature. Low-level feature includes color, texture and inflexion. Middle level involves shape description and object feature.

II. Framework

In this chapter we focus on tags, although the analysis can be easily expanded to include other types of metadata, including social networks (Lerman et al., 2007). Tags are freely-chosen keywords users associate with content. Tagging was introduced as a means for users to organize their own content in order to facilitate searching and browsing for relevant information. The distinguishing feature of tagging systems is that they use

an uncontrolled vocabulary, and that the user is free to highlight any one of the object's properties. The methods for personalizing results of image search on Flickr.

2.1 Reranking For Aesthetics

This paper studies the impact of aesthetic characteristics of images on the perceived quality of search results by users. To the end, we combine relevance scores obtained by relevance-oriented rank methods with aesthetic quality scores predicted for photographs. We call this combination of relevance and aesthetic scores as ranking aesthetic-aware reranking. Intuitively, relevance and aesthetic quality are orthogonal dimensions and therefore convey complementary information about documents being retrieved. In the simplest case scenario, we can think of aesthetic quality as a way to break relevance score ties to enhance results. In this section, we introduce and describe the main components of the reranking strategy adopted.

Most systems use the query by example approach, where the user selects one or several images, and the system returns the ones judged similar. An alternative way of querying the image database based on content is by allowing the user to sketch the desired image's color/texture layout, thus abstracting himself, the objects searched for. Other more targeted systems allow the user to specify spatial constraints on the dominant objects. All of these methods suffer some-what from the drawback that the system relies on the users abilities and does not adapt to his/her needs.

The personalization method described in this chapter will fail if a user makes a query in a domain in which he/she has not previously expressed any interest. For example, suppose that a child portrait photographer wants to find beautiful mountain scenery. If she has never created tags relating to mountains landscape photography in general, the personalization method described above will fail. However, the Flickr community as a whole has generated a significant amount of data about nature and landscape photography and mountains in particular. Analysis of community-generated data can help the user discover mountain imagery the community has identified as being good. We need algorithms to mine community-generated metadata and knowledge to identify community-specific topics of interest, vocabulary, authorities within the communities and community-vetted content.

When define a new primitive, a function for computing the similarity between two sets of feature data previously extracted must also be supplied by the developer. When comparing two images, for each primitive in the current query combination, a similarity score is computed using the distance function defined within the primitive. These individual scores are combined in an overall score using a set of weights in a way characteristic to the application. This score is then stored in a score structure, which contains also the individual similarity scores for each primitive. This allows a quick recompilation of the overall score for a new set of weights. In recent years, extensive efforts have been focusing on personalized search. Regarding the resources they leveraged, explicit user profile relevance feedback user history data (browsing log click-through data and social annotations context information and social network are exploited. For the implementation there are two primary strategies query refinement and result processing. In the following we review the related work by the strategy they used.

2.2 Automatic Web Image Retrieval

In this module, for each group of Flickr photos, we first construct a group-specific lexicon which contains only the tags of all photos within the same group. for each query tag (*e.g.* "dog"), we exploit the inverted file method to automatically find the positive training web im-ages that are related to the tag "dog" as well as negative training web images which do not contain the tag "dog" in the surrounding texts. Considering the total number of negative training web images (up to millions) is much larger than that of the positive training web images, we randomly sample a fixed number of negative web images and combine them with the positive web images to construct a smaller training set. To reduce the representation size, to facilitate the retrieval process while maintaining translation invariance, an alternative approach is to use an adaptive sampling scheme. Wavelet maxima have been shown to work well in detecting edges which are likely key features in a query. Moreover this method provides flexibility in choosing filters and the size of extracted data. By varying the applied filers, one could control the amount of data to be recorded. We are currently experimenting with this approach and results will be reported soon.

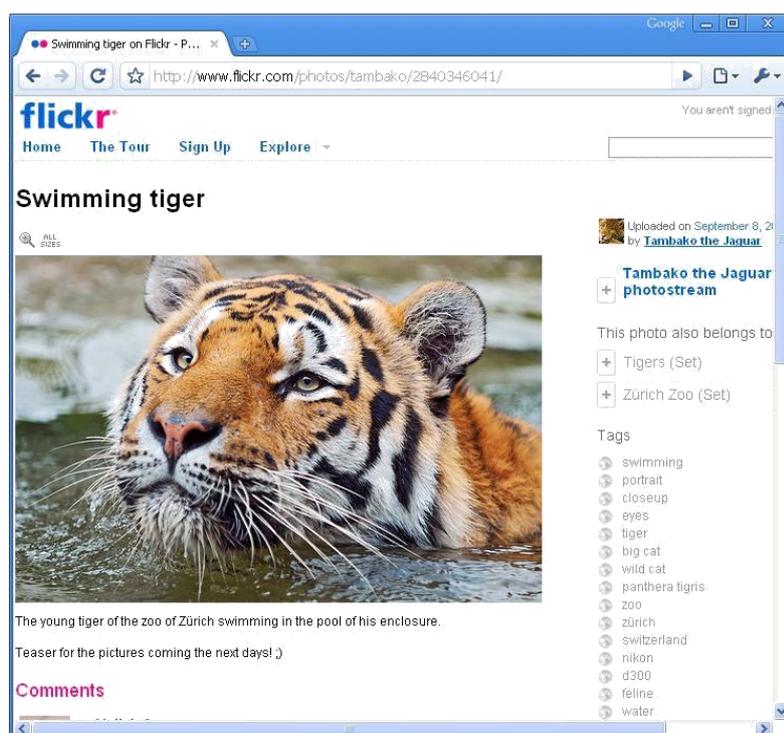


Figure 1: Screen shot of an image page of Flickr user Tambako the Jaguar showing the image and the tags he attached to the image

2.3 Tag-Based Personalization

We outline a probabilistic model that takes advantage of the images' tag and group information to discover latent topics contained in a set of images. If the dataset is a result of a search for images that have been tagged with the query term, the topics correspond to different senses of the query term. The users' interests can similarly be described by collections of tags they used to describe their own images. In the research community of personalized search, evaluation is not an easy task since relevance judgment can only be evaluated by the searchers themselves. The most widely accepted approach is user study where participants are asked to judge the search results. Obviously this approach is very costly. In addition, a common problem for user study is that the results are likely to be biased as the participants know that they are being tested. Another extensively used approach is by user query logs click or-through history. However, this needs a large-scale real search logs, which is not available for most of the researchers. Suppose a user is interested in wildlife photography and wants to see images of tigers on Flickr. The user can search for all public images tagged with the keyword "tiger". As of March 2007, such a search returns over 55, 500 results. When images are arranged by their "interestingness," the first page of results contains many images of tigers, but also of a tiger shark, cats, butterfly and a fish. Subsequent pages of search results show, in addition to tigers, children in striped suits, flowers (tiger lily), more cats, Mac OS X (tiger) screenshots, golfing pictures (Tiger Woods), etc. In other words, results include many false positives, images that are irrelevant to what the user had in mind when executing the search.

III. Future Research Directions

In this paper, we have shown that community feedback found in Web based social sharing systems can be used to improve the ranking of image search results. More specifically, we have leveraged user comments about photographs to create a comment-based feature representation of images conveying the opinion, positive or negative, of users about the images. We have used these features for building regression models aimed at Predicting the aesthetic quality of images, using ratings provided by users of the community as ground truth. Finally, we have studied how to combine relevance and aesthetic scores to rerank image search results. Our experiments have shown that context-based representations outperform visual-based in terms of prediction accuracy. We also conducted a user study to determine user satisfaction with aesthetic aware reranking of search results, which revealed a consistent preference of results reranked by the combination of aesthetic and relevance scores.

IV. Conclusion

We presented two methods for personalizing results of image search on Flickr. Both methods rely on the meta-data users create through their everyday activities on Flickr, namely user's contacts and the tags they used for annotating their images. We claim that this information captures user's tastes and preferences in photography and can be used to personalize search results to the individual user. In addition to creating content, users of Web 2.0 sites generate large quantities of metadata, or data about data, that describe their interests, tastes and preferences. These metadata, in the form of tags and social networks, are created mainly to help users organize and manage their own content. Our method relies on metadata created by users through their everyday activities on Flickr, namely the tags they used for annotating their images and the groups to which they submitted these images.

References

- [1] B. Smyth, "A community-based approach to personalizing web search," *Computer*, vol. 40, no. 8, pp. 42–50, 2007.
- [2] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu, "Exploring folksonomy for personalized search," in *SIGIR*, 2008, pp. 155–162.
- [3] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'El, I. Ronen, E. Uziel, S. Yogev, and S. Chernov, "Personalized social search based on the user's social network," in *CIKM*, 2009, pp. 1227–1236.
- [4] Y. Cai and Q. Li, "Personalized search by tag-based user profile and resource profile in collaborative tagging systems," in *CIKM*, 2010, pp. 969–978.
- [5] D. Lu and Q. Li, "Personalized search on flickr based on searcher's preference prediction," in *WWW (Companion Volume)*, 2011, pp. 81–82.
- [6] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search?" in *WSDM*, 2008, pp. 195–206.
- [7] S. Bao, G.-R. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su, "Optimizing web search using social annotations," in *WWW*, 2007, pp. 501–510.
- [8] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles, "Exploring social annotations for information retrieval," in *WWW*, 2008, pp. 715–724.