

# Tag Image Re-Ranking for Social Dataset

Miss. Radhika K. Rathi<sup>1</sup> & Prof. Karuna G. Bagde<sup>2</sup>

<sup>1</sup>Student, Dept of Computer Science and Information Technology, H.V.P.M.'s College of Engineering & Technology, Amravati.

<sup>2</sup>Asso.Professor, Dept of Computer Science and Information Technology, H.V.P.M.'s College of Engineering & Technology, Amravati.

---

**Abstract:** *Tag-based image search is an important method to process images contributed by social users in social media sharing websites like Flickr. Since user tagging is known to be uncontrolled, ambiguous, and overly personalized, a fundamental problem is how to interpret the relevance of a user-contributed tag with respect to the visual content the tag is describing. However, existing ranking methods for tag-based image search frequently return results that are irrelevant, low-diversity or time-consuming. In this work, it proposes a social re-ranking system for tag-based image retrieval with the consideration of image's relevance and diversity. We aim at re-ranking images according to their visual information, semantic information and social clues. The initial results include images contributed by different social users. Usually each user contributes several images. First we sort these images by inter-user re-ranking. Users that have higher contribution to the given query rank higher. Then we sequentially check title and time stamp ranking in which the desired output will get on the basis of title information and the recent time stamp which enhance the diversity performance of image ranking system also it count number of views has been utilized to improve the relevance performance of the image retrieval results. These selected images compose the final retrieved results. We build an identify keyword relevancy match the data is retrieved for the social image dataset to accelerate the searching process. Experimental results on Flickr dataset show that our tag image re-ranking method is effective and efficient.*

**Keywords** — Social Media, Tag-based Image Retrieval, Image Search, Title Information Re-Ranking, Time-Stamp Re-Ranking.

*Introduction*

## 1. Introduction

Social media sharing web sites like Flickr allow users to annotate images with free tags, which greatly facilitate social image search and browsing. However, currently tag-based image search on Flickr does not provide the option of relevance-based

ranking, i.e., the search results cannot be ranked according to their relevance levels with respect to the query tag, and this has limited the effectiveness of tag based search. Tag-based search, which returns images annotated with a specific query tag is an important way of searching or browsing images on social dataset. This image search method, to some extent, has achieved some success on exploiting the associated tags for indexing and searching large-scale web images compared with text-based image search and content-based image search [1]. Tag-based image search is more commonly used in social media than content based image retrieval [2] and context-and-content based image retrieval [3]. In recent years, the re-ranking problem in the tag-based image retrieval has gained researchers' wide attention.

Starting from this intuition and above analysis, it proposes a social re-ranking algorithm which user information is firstly introduced into the traditional ranking method considering the semantics, social clues and visual information of images. The contributions of this paper can be described as follows:

- 1) A tag-based image search approach with social re-ranking. We systematically fuse the visual information, social user's information and image view times to boost the diversity performance of the search result.
- 2) Inter user ranking is applied to rank users images according to query given. With this ranking the system achieve the good tradeoff between the diversity and relevance performance which also effectively eliminate the similar images from the same user in a ranked result.
- 3) Title and time stamp ranking in which the desired output will get on the basis of title information and the recent time stamp which enhance the diversity performance of image ranking system.
- 4) The view of an image in social media community is an important feature which indicates the click count of this image. The number of click count has been utilized to improve the relevance performance of the image retrieval results.

We take Flickr as an example to study the characteristics of social tagging. Flickr is one of the earliest and most popular social media sharing web sites and it has been intensively studied in recent years, especially on tagging characteristic, tag recommendation, etc. A recent study in reveals that users do annotate their photos with the motivation to make them better accessible to the general public. However, the tags provided by Flickr users are highly noisy and there are only around 50% tags actually related to the image. Fig. 1 illustrates an exemplary image from Flickr and its tags. From the figure we can see that only “sky” and “cloud” correctly describe the content of the given image, and the other tags are imprecise(e.g., dog, girl, etc.) or subjective(e.g., family, city, etc.). Meanwhile, several other tags that can be useful, such as “tree” and “grass”, have not been provided. The imprecise and incomplete tagging characteristics have significantly limited the access of social media. The imprecise tags will introduce false positives into user’s search result and incomplete tags will make the actually related images inaccessible. Therefore, it would be advantageous if a dedicated approach can be developed to improve the tags associated with social images such that they can better describe the content of the images.



Fig. 1. An exemplar image from Flickr and its associated tags.

## 2. Aim & Objectives

### Aim:

The aim of the system is to re-rank images according to their visual information, semantic information and offline dataset information which we will make by Crowd-Sourcing. The initial results include images contributed by different social users. Usually each user contributes several images. First we sort these images by inter-user re-ranking. Users that have higher contribution to the given query rank higher and proposed system also solve the problem of query ambiguity.

### Objective:

- To reduce time to search images according to image tag.
- To Solve the Query mismatch and Query ambiguity problem.
- To rank the images by considering their visual information, semantic information, uploaded time of image and no. of view of image.
- To maintain relevancy and accuracy of the system i.e. Tag Based image search.
- To get visual information and semantic information of image by crowd-sourcing.

## 3. Related work

In author [1], [4] propose our key idea is to learn the relevance of a tag with respect to an image from tagging behaviors of visual neighbors of that image. In particular, our algorithm estimates tag relevance by counting neighbor votes on tags and the tag refinement technique is able to improve the effectiveness of image tag recommendation for non-tagged images.

In author [5] are the major approaches in settling the diversity problem. However, the essence of social images is ignored. The social images uploaded and tagged by users are user-oriented. These user-oriented images which share the same user and tagged with same query are always taken in a fixed time interval at a specific spot. It is well-known that, images taken in the same time interval and fixed spot are fairly similar. To diversify the top ranked search results, it's better to re-rank the results by removing the duplicate images from the same user.

In author [6] propose a relevance-quality ranking method considering both image relevance and image quality. First, a relevance-based ranking scheme is utilized to automatically rank images according to their relevance to the query tag, which reckons the relevance scores based on both the visual similarity of images and the semantic consistency of associated tags. Then, quality scores are added to the candidate ranking list to accomplish the relevance-quality based ranking.

In author [7], [8], [9], [10] proposed a two-step similarity ranking solution for interactive image retrieval. It first propose a self-tune MR solution that focuses on the visual-based similarity ranking, and then develop a semantic-oriented similarity re-ranking method to address the dislocation problem.

Social image websites allow users to annotate their images with a set of descriptors such as tags. Thus, the tag-based image search can be easily accomplished by using the tags as query terms. Different from traditional web image websites, social media websites allow users to annotate social images

with tags for tag being the effective approach for social image search. Most of the literatures regarding the re-ranking of the tag-based image retrieval focus on tag processing, image relevance ranking and diversity enhancement of the retrieval results. The following parts present the existing works related to the above three aspects respectively.

#### A. Tag Processing Strategy

It has been long acknowledged that tag ranking and refinement play an important role in the re-ranking of tag-based image retrieval, for they lay a firm foundation on the development of re-ranking in tag based image retrieval (TBIR). Author in [1] proposed to learn the relevance of tags by visually weighted neighbor voting, a variant of the popular baseline neighbor voting algorithm. Author in [11] proposed a relevance tag ranking algorithm, which can automatically rank tags according to their relevance with the image content and presented a tag fusion method for tag relevance estimation to solve the limitations of a single measurement on tag relevance. Author in [12] raised a tag completion algorithm to fill in the missing tags and correct the erroneous tags for the given image.

#### B. Relevance Ranking Approach

To directly rank the raw photos without undergoing any intermediate tag processing, Author in [7] utilized an optimization framework to automatically rank images based on their relevance to a given tag. Visual consistency between images and semantic information of tags are both considered. Author in [10] proposed an image ranking method which represent images by sets of regions and apply these representations to the multiple-instance learning based on the max margin framework.

#### C. Diversity Enhancement

Many images on social media websites are actually close to each other. For example, several users used to upload continuously captured images in batch, and many of them will be visually and semantically close. When these images appear simultaneously in the top results, users will get only limited information. Therefore, a ranking scheme that can simultaneously generate relevant and diverse results is highly desired. The relevance based image retrieval approaches can boost the relevance performance; however the diversity performance of searching are often ignored. Many researchers dedicated their extensive efforts to solve this problem. Author proposed a hierarchical clustering method to cluster the search results into different semantic clusters by using visual, textual and link analysis. Author in [13] proposed a duplicate detection algorithm to represent images with hash code, so that large image database with similar hash

codes can be grouped quickly. We first get the initial results by keyword matching process. Then the inter-user and intra-user re-ranking are introduced to re-rank the initial results. Inter-user re-ranking algorithm is applied to rank users according to their contribution to the given query. After the inter-user re-ranking, we further introduce intra-user re-ranking to sequentially select the most relevant image from each image dataset of the ranked users.

## 4. Problem Definition

The following challenges block the path for the development of re-ranking technologies in tag-based image retrieval.

- 1) Tag mismatch
- 2) Query ambiguity

Thus, a fundamental problem in the re-ranking of the tag-based social image retrieval is how to reliably solve above problems. As far as the “tag mismatch” problem is concerned, tag refinement, tag relevance ranking and image relevance ranking approach have been dedicated to overcome above problems. As for the “query ambiguity” problem, an effective approach is to provide diverse retrieval results that cover multiple topics underlying a query. To diversify the top ranked search results, it’s better to re-rank the results by removing the duplicate images from the same user.

Disadvantages of Existing System:

- The same image can be interpreted in several ways with several different tags according to the background behind the image. Thus, many seemingly irrelevant tags are introduced.
- Existing system cannot search the tags which is synonyms of the searched query.
- In existing system only user’s contribution is considered and the similarity among users is ignored. In addition to this, much information in Flickr dataset are still ignored, such as title information, time stamp and so on.

## 5. Proposed system

Proposed system i.e. Tag-Based social re-ranking system includes two main sections: online and offline. In ofsfline section we provide the images to the water-armies and ask them to tag the images and construct offline dataset, we extract the visual feature, semantic feature and views for the images dataset. Semantic feature refers to the co-occurrence word set of query tags and the tags of the images.

Online parts consist of the following two steps: 1) Keyword matching. For an input query, proposed system will return the initial retrieval results by

keyword matching. 2) Inter-user re-ranking. The inter-user re-ranking is applied to rank the corresponding users with the consideration of their contributions to the given query. A regularization framework is proposed to determine the relevance level of each image by fusing the visual, semantic, time stamp information and views information into a unified system. Then we sequentially select the most relevant image in each ranked user's image set. These selected images constitute our re-ranking results. Hereinafter the details are displayed.

Online system:

Our online system carries out the following four steps to obtain the ranked images for the query tag  $q$ :

- 1) Keyword matching,
- 2) Inter-user re-ranking,
- 3) Title and time-stamp re-ranking, and
- 4) No. of view Count re-ranking.

## 6. Proposed Methodology

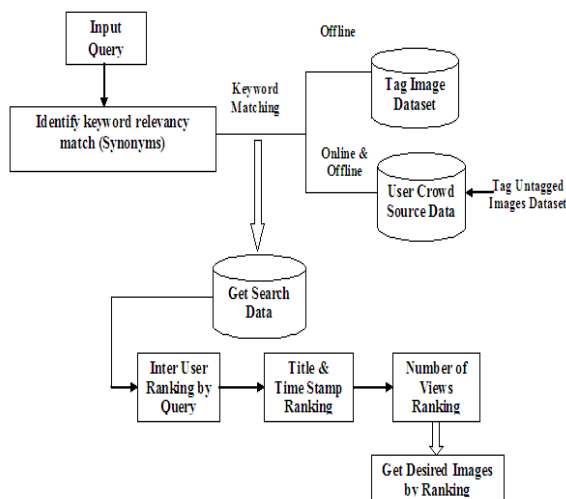


Fig 2. System Architecture

In order to achieve the tag based image re-ranking approach we proposed above methodology, explained as follows:

First system takes initial query as input. Proposed system then parse that query by removing stop words or by applying stemming on the query and pass the words to next step. System search for synonyms, annotations of each word in semantic dictionary and match each keyword in offline as well as online dataset as shown in above fig.. If match is found then system selects all the visual information, semantic information from the dataset and integrates as intermediate result. On the intermediate result system re-rank the images by using proposed methods such as inter-user re-ranking, title & time stamp re-ranking and no. of view re-ranking. In the

inter-user re-ranking, system re-ranks the images in descending order according to contribution of each user to the given query. The larger the contribution of user, the higher the corresponding user ranks. In next step system sort the images in descending order according to an uploaded date of the image. Then system sorts the images according to no. of view count of the image. Finally system rank images.

Following are steps of the proposed algorithm

- Step 1: Input Query  $q$ .
- Step 2: Parse the Keyword i.e. remove stop words and apply stemming.
- Step 3: Match the keyword with dataset.
- Step 4: Store the result in Set  $S_1$ .
- Step 5: Take the resulted tags in  $S_1$  & Get the annotations from dataset.
- Step 6: Repeat Step 3.
- Step 7: Store the result in Set  $S_2$ .
- Step 8: Merge the Set  $S_1$  and Set  $S_2$  i.e.  $S = S_1 + S_2$ .
- Step 9: Apply Sorting on Set  $S$  according to user contribution count, date of upload and View count of image.
- Step 10: Final Result as ranked images.

## 7. Advantages

1. Image Searching becomes very easy.
2. Accurate and relevant content retrieval system saves the time of user.
3. Duplication of tag is reduced.
4. The main problem of tag mismatching and query ambiguity is resolved.
5. Because of using time stamp and view count ranking of images user get latest images on the top of their search results.

## 8. Conclusion

We propose a social re-ranking method for tag-based image retrieval. In this social re-ranking method, inter-user re-ranking is carried out to obtain the retrieved results. In order to enhance the diversity performance, user information is firstly introduced into our proposed approach and obtains satisfactory results. Besides, views count of social image is also firstly fused into a traditional regularization framework to enhance the relevance performance of retrieved results. Our proposed method is effective and time-saving.

## References

- [1] X. Li, C. Snoek, and M. Worring. Learning tag relevance by neighbor voting for social image retrieval. Proceedings of the ACM International Conference on Multimedia information retrieval, 2008: 180-187.

- [2] X. Yang, X. Qian, Y. Xue. Scalable Mobile Image Retrieval by Exploring Contextual Saliency. *IEEE Transactions on Image Processing* 24(6): 1709-1721 (2015).
- [3] G. Qi, C. Aggarwal, Q. Tian, et al. Exploring Context and Content Links in Social Media: A Latent Space Method, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 2012.
- [4] S. Lee, W. D. Neve. Visually weighted neighbor voting for image tag relevance learning. *Multimedia Tools and Applications*, 1-24, 2013.
- [5] K. Yang, M. Wang, X. Hua, and H. Zhang. Social Image Search with Diverse Relevance Ranking. *Proceedings of the IEEE International Conference on Magnetism and Magnetic Materials*, 2010:174-184.
- [6] L. Chen, S. Zhu, Z. Li. Image retrieval via improved relevance ranking. In *Control Conference*, pp. 4620-4625, IEEE, 2014.
- [7] D. Liu, X. Hua, M. Wang, and H. Zhang. Boost Search Relevance For Tag-Based Social Image Retrieval. *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2009:1636-1639.
- [8] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao. Tag-based social image search with visual-text joint hypergraph learning. *Proceedings of the ACM International Conference on Multimedia information retrieval*, 2011:1517-1520.
- [9] D. Wu, J. Wu, M. Lu. A Two-Step Similarity Ranking Scheme for Image Retrieval. In *Parallel Architectures, Algorithms and Programming*, pp. 191-196, IEEE, 2014.
- [10] Y. Hu, M. Li. Multiple-instance ranking: Learning to rank images for image retrieval. In *Computer Vision and Pattern Recognition, CVPR 2008. IEEE Conference on* (pp. 1-8).
- [11] G. Agrawal, R. Chaudhary. Relevancy tag ranking. In *Computer and Communication Technology*, pp. 169-173, IEEE, 2011.
- [12] L. Wu, R. Jin. Tag completion for image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3), 716-727, 2013.
- [13] B. Wang, Z. Li, M. Li. Large-scale duplicate detection for web image search. In *Multimedia and Expo, 2006 IEEE International Conference on* (pp. 353-356).
- [14] Xueming Qian, Dan Lu, and Xiaoxiao Liu. Tag Based Image Search by Social Re-ranking. *IEEE TRANSACTIONS ON MULTIMEDIA*, 2016.