Tag Relevance for Social Image Retrieval in Accordance with Neighbor Voting Algorithm

Deepshikha Mishra[1], Uday Prtap Singh[2] Vineet Richhariya[3]
[1] M.Tech Scholar, Dept. of Information Technology, LNCT, Bhopal, India
[2] Professor, Department of Computer Science and Engineering, LNCT, Bhopal, India
3 Prof. & HOD Dept. of Computer Science, LNCT, Bhopal,India

Abstract
Social image retrieval is important for exploiting the increasing amounts of amateur-tagged multimedia such as Flickr images. Intuitively, if different persons label similar images using the same tags, these tags are likely to reflect objective aspects of the visual content. Interpreting the relevance of a user-contributed tag with respect to the visual content of an image is an emerging problem in social image retrieval. An algorithm is proposed that scalably and reliably learns tag relevance by accumulating votes from visually similar neighbours. Treated as tag frequency, learned tag relevance is seamlessly embedded into current tag-based social image retrieval paradigms. Preliminary experiments on two thousand Flickr images demonstrate the potential of the proposed algorithm. The tag relevance learning algorithm substantially improves upon baselines for all the experiments. The results suggest that the proposed algorithm is promising for real-world applications.

Keywords
neighbour voting, tag relevance, user contributed tag, social image tagging.

1. INTRODUCTION

Image sharing websites such as Flickr and Facebook are hosting billions of personal photos. Tagging is a significant feature of social bookmarking systems which enables users to add, annotate, edit and share bookmarks of a web documents. Social image tagging, assigning tags to images by common users, is reshaping the way people manage and access such large-scale visual content. Image tagging basically refers to a process of categorizing or mapping of images on the basis of their contents either visual or context. Along with the rapid growth of personal albums in social networking sites, it has been seen that tagging is the most promising and practical way to facilitate the huge photos database semantically searchable. To tag an image firstly the training set is manually tagged and then the tags of the testing set are automatically predicted.

In a social tagging environment with large and diverse visual content, a light weight or unsupervised learning method which effectively and efficiently estimates tag relevance is required. Two simplest and easiest ways for multi feature tag relevance learning are-the classical borda count and uniform tagger. Image tagging can be done in two ways:-

1. Manual Image Tagging.
2. Automatic Image Tagging.

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images.

Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools. There are various image retrieval techniques. These techniques can be categorised according to whether they are based on text, content, multimodal fusion, or semantic concepts. We differentiate these techniques by the type of features that are used to represent the images as well as the approaches that are used to retrieve similar images.

The text-based image retrieval techniques use keywords, the CBIR techniques use low-level image features, the multimodal fusion techniques use a combination of various image representative features, and the semantic-based techniques use concepts.

2. RELATED TERMS & WORK

Text Based Image Retrieval-Text-based image retrieval is also called description-based image retrieval. Text-based image retrieval is used to retrieve the XML documents containing the images based on the textual information for a specific multimedia query. To overcome the limitations of CBIR, TBIR represents the visual content of images by manually assigned keywords/tags. It allows a user to present his/her information need as a textual query, and find the relevant images based on the

Manuscript received July 5, 2014
Manuscript revised July 20, 2014

match between the textual query and the manual annotations of images

Content Based Image Retrieval- In content based image retrieval, images are searched and retrieved on the basis of similarity of their visual contents to a query image using features of the image. A feature extraction module is used to extract low-level image features from the images in the collection. Commonly extracted image features include color, texture and shape.

Multimodal Fusion Image Retrieval- Multimodal fusion image retrieval involves data fusion and machine learning algorithms. Data fusion, also known as combination of evidence, is a technique of merging multiple sources of evidence. By using multiple modalities, we can learn the skimming effect, chorus effect and dark horse effect

Semantic Based Image Retrieval- Image retrieval based on the semantic meaning of the images is currently being explored by many researchers. This is one of the efforts to close the semantic gap problem. In this context, there are two main approaches: Annotating images or image segments with keywords through automatic image annotation or adopting the semantic web initiatives.

Improving Image Tagging- Image tagging can be improved by tagging the images on the basis of their features and tags should be relevant to the image and with the help of which image can be retrieved from pool of the databases.Retrieval of image can also be done by multiple features together and is efficient for both unlabelled and labelled images. For labelled images tags are predicted on the behalf of the features, characteristics, colour, texture etc and for unlabelled images tagging is done when we load a query image and get it neighbour images and tags are then predicted on the basis of the retrieved images and the features being exhibited by same tagged image. Depending on whether a target image is labelled, we can categorize existing methods into two main scenarios, explicitly improving image tagging for labelled images and automated image tagging for unlabeled images. In the first scenario, given an image labelled with some tags, one tries to improve image tagging by removing noisy tags [22], recommending new tags relevant to existing ones [23], or reducing tag ambiguity [4]. In [22] for instance, the authors assume that the majority of existing tags are relevant with respect to the image. They then measure the relevance of a tag by computing word similarity between the tag and other tags. While in [23], the authors find new tags relevant with respect to the original ones by exploiting tag co-occurrence in a large user-tagged image database. To be exact, by using each of the original tags as a seed, they find a list of candidate tags having the largest co-occurrence with the seed tag. These lists are later aggregated into a single list and the top ranked tags are selected as the final suggestion..

Methods in the second scenario try to predict relevant tags for unlabeled images. We can categories these methods according to their model-dependence into model-based and model-free approaches. The model-based approaches, often conducted in a supervised learning framework, focus on learning a mapping or projection between low-level visual features and high-level semantic concepts given a number of training [24],[25],[26]. Moreover, the approaches are often computationally costly, making them difficult to scale up. In addition, the rapid growth of new multimedia data makes the trained models outdated quickly. To tackle these difficulties, a lightweight meta-learning algorithm is proposed in [27].

The general idea of the algorithm is to progressively improve tagging accuracy by taking into account both the tags automatically predicted by an existing model and the tags provided by a user as implicit relevance feedback. In disparity to the model-based approaches, the model-free approaches attempt to predict relevant tags for an image by utilizing images on the Internet [5], [28]. These approaches imagine there exist a large well-labelled database such that one can find a visual duplicate for the unlabeled image. Then, automatic tagging is done by simply propagating tags from the duplicate to that image. In reality, however, the database is of limited scale with noisy annotations. Hence, neighbour search is first conducted to find visual neighbours. Disambiguation methods are then used to select relevant tags out of the raw annotations of the neighbours.

To review, the existing methods for image tagging try to rank relevant tags ahead of irrelevant ones in terms of the tags relevance value with respect to an image. However, since the tag ranking criterion is not directly related to the performance of image retrieval using the tagging results, optimizing image tagging does not necessarily yield good image rankings [12].

Improving Image Retrieval- Image retrieval can be improved on the basis of the content as well as the features, characteristics, colour etc of the image. First of all the query image is loaded and then its neighbour images are retrieved on the basis of features it could be text based image retrieval or content based image retrieval in which retrieval of images is done on the basis of text or content of the image it can be anything like colour, feature, characteristics. Retrieval of images can be done for labelled as well as for unlabelled images. In labelled image retrieval images are retrieved on the behalf of tags which differentiate each group from other. Image with similar features are grouped together and will be retrieved in a group only whenever a feature of the grouped is being called for images containing that feature, the whole group will be retrieved. While for unlabelled images retrieval is done again on the basis of grouping of images and tags are being predicted on the behalf of the characteristics of similar group images, tag prediction for unlabelled images is done with the help of the features of all the pictures which are being retrieved and the features which are

present in all the retrieved images are considered as tags and are being predicted for the whole group together.

Given insufficient image tagging results, one might expect to improve image retrieval directly. Quite a few methods follow this research line, either by re-ranking search results in light of visual consistency. Re-ranking methods assume that the majority of search results are relevant with respect to the query and relevant examples tend to have similar visual patterns such as colour and texture. To find the dominant visual patterns, density estimation methods are often used, typically in the form of clustering [8] and random walk [10]. In [10] for instance, the authors pull a random walk model to find visually representative images in a search result list obtained by text-based retrieval. To be specific, first an adjacent graph is constructed wherein each node corresponds to a result image and the edge between two nodes are weighted in terms of the visual similarity between the two corresponding images. A random walk is then replicated on the graph to estimate the probability that each node is visited. Since images in dense regions are more likely to be visited, the above probability is used to compute the representativeness of an image in the visual feature space and accordingly re-rank the search results. The obscurity in density estimation and the associated computational expense put the utility of re-ranking methods for social image retrieval into question. Re-ranking of images is done on the basis of the features being called up and the retrieval of images is done on the basis of priority. Priority list is being set on the basis of the images constituting the features which are being called by the query image, and then images are retrieved on the basis of priority and the image with the most similar feature is being retrieved first and so on. The image which matches the most requirement of the query image is being retrieved first and then at last with least priority. Ranking is done on the behalf of priority list.

3. LEARNING TAG RELEVANCE BY NEIGHBOR VOTING

For accomplishment of image retrieval, a tag relevance measurement is seeked such that images relevant with respect to a tag are ranked at the forefront of images irrelevant with respect to the tag. In the same time, to fulfil image tagging, the measurement should rank tags relevant with respect to an image ahead of tags irrelevant with respect to the image. From our earlier discussions we know that if different persons label visually similar images using the same tags, these tags are most probable to reflect objective aspects of the visual content. This suggests that the relevance of a tag given an image might be inferred from how visual neighbours of that image are tagged: the more regular the tag occurs in the neighbor set, the more relevant it might be, to the query image. Thus, a good tag relevance measurement should take into account the distribution of a tag in the neighbor set and in the entire collection, at the same time. Motivated by the informal analysis above, I propose a neighbor voting algorithm for learning tag relevance. Though the proposed algorithm is simple, I deem it important to gain insight into the rationale for the algorithm.

The Goal Of Tag Relevance Learning- Some notation for the ease of explanation are described. A collection of user-tagged images is denoted as Ψ and a vocabulary of tags used in Ψ as W. For an image I $\in \Psi$ and a tag $t \in$ W, let r*(t, I) : {W,Ψ} $\mapsto$ R be a tag relevance measurement. It is called r*(t, I) an ideal measurement for image and tag ranking if it satisfies the following two conditions:

Condition 1: Image ranking. Given two images $I_1, I_2 \in \Psi$ and tag $t \in$ W, if t is relevant to I_1 but irrelevant to I_2, then r*(t, I_1) > r*(t, I_2).....(i)

Condition 2: Tag ranking. Given two tags $t_1, t_2 \in$W and image I $\in \Psi$, if I is relevant to t_1 but irrelevant to t_2,then r*(t_1, I) > r*(t_2, I).....(ii)

The goal is to find a tag relevance measurement satisfying the above two conditions.

Learning Tag Relevance From Visual Neighbors- Given an image I labeled with a tag t, the occurrence frequency of t in visual neighbors of I to some extent reflects the relevance of t with respect to I. It is to be noted that the neighbors can be decomposed into two parts according to their relevance to t, i.e., images relevant and irrelevant to t. If we know how relevant and irrelevant images are labelled with t and how they are distributed in the neighbor set, the tag's distribution in the neighbors can be estimated.

To formalize the above notions, first of all I have defined a few notations as listed in Table I. No one can study how images relevant and irrelevant to a tag are labelled with that tag. In a large user-tagged image database, it is plausible that for a specific tag t, the number of images irrelevant to the tag is significantly larger than the number of relevant images, i.e., $|R_t^c| \gg |R_t|$, where $|\bullet|$ is the cardinality operator on image sets. Also, one might expect that user tagging is better than tagging at random such that relevant images are more likely to be labelled, meaning $|L_t \cap R_t| > |L_t \cap R_t^c|$.

Table 1

Major notations which are used in the algorithm

Notations	Definition
Ψ	A collection of user-tagged images
L_t	$L_t \subset \Psi$, all images labeled with tag t in the collection.
R_t	$R_t \subset \Psi$, all images relevant with respect to tag t in the collection.
R_t^c	$R_t^c = \Psi / R_t$, all images irrelevant with

	respect to tag *t* in the collection.
P($t \mid R_t$)	Probability of correct tagging, i.e., an image randomly selected from R_t is labelled with tag *t*.
P($t \mid R_t^c$)	Probability of incorrect tagging, i.e., an image randomly selected from R_t^c is labeled with tag *t*.
P(R_t)	Probability that an image randomly selected from the entire collection is relevant to tag *t*.
P(R_t^c)	Probability that an image randomly selected from the entire collection is irrelevant to tag *t*.
f	A similarity function between two images, measured on low-level visual features.
$N_{f\,(I,k)}$	$N_{f(I,k)} \subset \Psi$, k nearest neighbors (k-nn) of an image I found in the collection by f.
N_{rank} (k)	N_{rank} (k) $\subset \Psi$, k images randomly selected from the collection.
$n_{t\,[.]}$	An operator counting the number of tag w in any subset of the collection.

By approximating the probability of correct tagging P($t \mid R_t$) using $|L_t \cap R_t| / |R_t|$ and the probability of incorrect tagging P($t|R_t^c$) using $|Lt \cap R_t^c| / |R_t^c|$, we have P($t|R_t$) > P($t|R_t^c$). Thus, we make a statement on user tagging behaviour, which is,

Statement 1: User tagging. In a large user-tagged image database, the probability of correct tagging is larger than the probability of incorrect tagging.

Next, the distribution of images relevant and irrelevant with respect to tag *t* is analyed in the k nearest neighbor set of image I. In comparison to random sampling, a content-based visual search defined by a similarity function f can be viewed as a sampling process biased by the query image. Two situations are considered with respect to the visual search accuracy, that is, equal to and better than random sampling. In the first case where the visual search is equal to random sampling, the number of relevant images in the neighbor set is the same as the number of relevant images in a set of k images randomly selected from the collection. Whereas in the second case where the visual search is better than random sampling, given two images I1 relevant to tag *t* and I2 irrelevant to *t*, we expect to have

$|N_f$ (I1, k) ∩ R*t*| > $|N_{rank}$(k) ∩ R*t*| > $|N_f$ (I2, k) ∩ R*t*|.

For example, consider *t* to be 'flower', I1 a flower image and I2 a non-flower image. In this example, N_f (I1, k) should contain more flower images than N_{rank}(k), while N_f (I2, k) should contain less flower images than N_{rank}(k). Viewing random sampling as a baseline, we introduce an offset variable $^{\varepsilon}$I,*t* to indicate the visual search accuracy. In particular, (P(R*t*) +$^{\varepsilon}$ I, *t*) is used to represent the probability that an image randomly selected from the neighbor set NN_f (I, k) is relevant with respect to *t*. Since an image is either relevant or irrelevant to *t*, we use (1 − (P(R*t*) + $^{\varepsilon}$ I, *t*)), namely (P(R_t^c) −$^{\varepsilon}$ I, *t*), to represent the probability that an image randomly selected from NN_f (I, k) is irrelevant with respect to *t*. Then, the number of relevant images in the neighbor set can expressed as

$|N_f$ (I, k) ∩ R_t| = k.(P(R_t) + $^{\varepsilon}$ I, *t*),.....(iii)

and the number of irrelevant images in the neighbor set as

$|N_f$ (I, k) ∩R_t^c| = k. (P(R_t^c) − $^{\varepsilon}$ I, *t*).....(iv)

It is to be mentioned that the variable $^{\varepsilon}$ I,*t* is introduced to help derive important properties of the proposed algorithm. I have not relied on $^{\varepsilon}$ I,*t* for implementing the algorithm.

Based on this discussion, if the visual search is equal to random sampling, we have $^{\varepsilon}$ I,*t* = 0. If the visual is better than random sampling, we have

$^{\varepsilon}$ I_1,*t* > 0 > $^{\varepsilon}$ I_2,*t*, for $I_1 \in R_t$ and $I_2 \in R_t^c$.....(v)

We then make our second statement as

Statement 2: Visual search. A content-based visual search is better than random sampling.

Keeping in mind the analysis of user tagging and visual search, now considering the distribution of tag *t* within the neighbor set of image I. Since the neighbor set can be divided into two distinct subsets N_f(I, k) ∩ R*t* andN_f (I, k) ∩ R_t^c, the number of w in the two subsets are counted, separately. That is,

n_t [N_f (I, k)]= n_t [N_f (I, k) ∩R_t] + n_t [N_f (I, k) ∩R_t^c]= k. (P (R_t) + $^{\varepsilon}$ I, *t*) P (t|R_t) + k. (P(R_t^c) – $^{\varepsilon}$ I, *t*) P (*t*|R_t^c).....(vi)

Similarly, we derive

n_t [N_{rank} (k)] = k. (P(R_t) P (*t*|R_t) + P (R_t^c) P (*t*|R_t^c)).....(vii)

Since n_t[N_{rank} (k)] reflects the occurrence frequency of *t* in the entire collection, it is denoted as *Prior (t,* k*)*, By substituting (vii) into (vi), we get

n_t [N_f (I, k)]- *Prior* (*t,* k)= k. (P(*t*|R_t) – P(*t*|R_t^c)) $^{\varepsilon}$ I, *t*.....(viii)

Further, by defining

Tag relevance (*t,* I, k):= n_t [N_f (I, k)]- *Prior* (*t,* k).....(ix)

The following two theorems can be concluded:

Theorem 1: Image Ranking: Given statement 1 and statement2, tag relevance yields an ideal image ranking for tag *t*, that is for $I_1 \in R_t$ and $I_2 \in R_t^c$,we have tag relevance (*t*, I_1) > (*t*, I_2).

Theorem 2: Tag Ranking: Given statement 1 and statement 2, tag relevance yields an ideal tag ranking for image I, that is for two tags t_1 and t_2, if I ∈ R_{t1} and I ∈ R_{t2} we have tag relevance (t_1, I) > tag relevance (t_2, I).

The appendix can be referred for detailed proofs of the two theorems. Note that in the proof of theorem 1, statement 2 (Eq.(v)) can be relaxed as ($^{\varepsilon}I_1$, *t*) which I call relaxed as statement 2. Since the relaxed statement is more likely to hold than its origin, this indicates that image ranking is relatively easier than tag ranking.

The tag relevance function in Eq. (ix) consists of two components which represents the distribution of the tag in the local neighborhood and in the entire collection, respectively. This observation confirms our assumption made in the beginning of Section 4 that a good tag relevance measurement should take into account both distributions.

Neighbor Voting Algoritm-Input: A user tagged image.

Output: Tag relevance (*t*, I, k) ,that is the tag relevance value of each tag *t* in I.

Find the k-nearest visual neighbors of I from the collection with the unique user constraint that is a user has at most one image in the neighbor set.

for tag *t* in tags of I do
    tag relevance (*t*, I, k) = 0
end for
for image J in the neighbor set of I do
    for tag *t* in (tags_of_J ∩ tags_of_I)do
    tag relevance (*t*, I, k)= tag relevance (*t*, I, k)+1
  end for
end for
tag relevance (*t*, I, k)= tag relevance (*t*, I, k)- *Prior* (*t*, k)
tag relevance (*t* ,I, k)= max (tag relevance (*t*, I, k) .1)

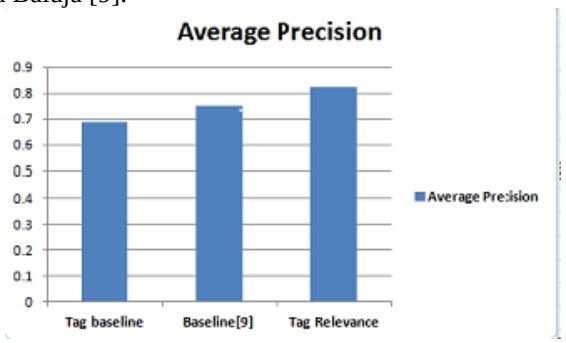
Experimental Setup Experiments-We evaluate our tag relevance learning algorithm in both an image ranking scenario and a tag ranking scenario. For image ranking, we compare three tag-based image retrieval methods with and without tag relevance learning. For tag ranking, we demonstrate the potential of our algorithm in helping user tagging in two settings, namely, tag suggestion for labelled images and tag suggestion for unlabeled images. Specifically, we design the following three experiments.

Experiment 1: Tag-Based Image Retrieval: A general tag-based retrieval framework widely used in existing systems such as Flickr and YouTube is employed. OKAPI-BM25, a well founded ranking function for text retrieval [14] as a baseline is adopted. Given a query q containing keywords {t1. . . tn}, the relevance score of an image I is computed as

$$\text{score}(q, I) = \sum_{w \in q} qt \quad f(t) \quad idf(t) \frac{f(t).(k1+1)}{f(t)+k1.(1-b+\frac{bll}{lavrg})} \quad \text{.....(xi)}$$

where qt$f_{(t)}$ if the frequency of tag *t*, $f_{(t)}$ the frequency of *t* in the tags of I, l_I the total number of tags of I, and l_{avrg} the average value of l_I over the entire collection. The function $idf_{(t)}$ is calculated as log(N-|L_t| + 0.5) / (|L_t| + 0.5), where N is the number of images in the collection and |L_t| is the number of images labelled with *t*. By using learned tag relevance value as updated tag frequency in the ranking function namely substituting tag relevance (*t*, I, k) for $f_{(t)}$ in eq.(xi), we investigated how the algorithm improves upon the baseline. The performance of the baseline method and our method has been studied, given various combinations of parameters. In total, there are three parameters to be optimized. One is k, the number of neighbors for learning tag relevance. k is chosen from{10; 20; 30; 40; 50; 100; 120; 150; 200}.The other two are b and k1 in OKAPI-BM25. The parameter b ($0 \leq b \leq 1$) controls the normalization effect of document length. The document length is the number of tags in a labelled image. Let b range from 0 to 1 with interval 0.1. The variable k1 is a positive parameter for regularizing the impact of tag frequency. Since k1 does not affect ranking for common choice in text retrieval [10].

Considering that the OKAPI-BM25 ranking function originally aims for text retrieval and thus might not be optimal for tag-based image retrieval, further comparison with a recent achievement in web image retrieval by Jing and Baluja [9].



Fiq. Shows the average precision value when applied different algorithms.

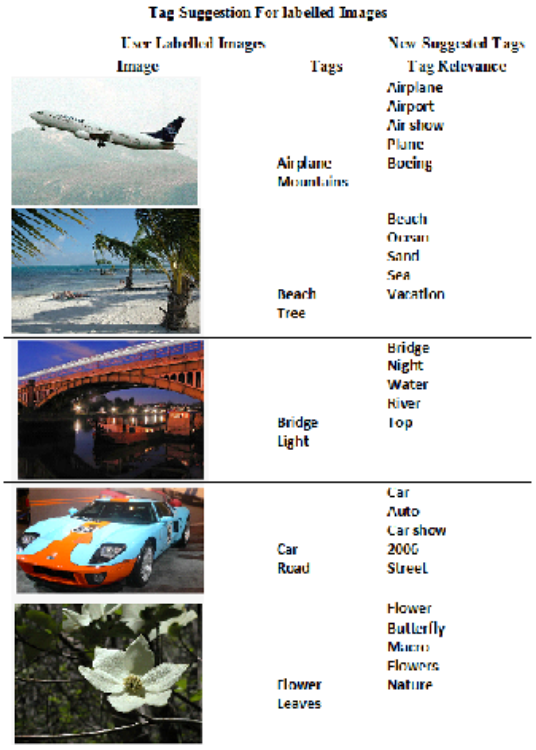
Query	Tag baseline	Tag baseline[28]	Tag Relevance
Airplane	0.450	0.750	0.600
Beach	0.433	0.350	0.800
Boat	0.500	0.383	0.850
Bridge	0.966	0.866	0.966
Bus	0.833	0.883	0.950
Butterfly	0.766	0.800	0.750
Car	0.816	0.900	0.900
Cityscape	0.300	0.916	0.610
Classroom	0.566	0.933	0.850
Dog	0.916	0.866	0.816
Flower	0.916	0.983	0.866
Harbour	0.700	0.583	0.683
Horse	0.683	0.983	1.000
Kitchen	0.800	0.900	0.800
Lion	0.866	0.750	1.000
Mountain	0.600	0.450	0.866
Rhino	0.916	0.900	0.716
Sheep	0.850	0.800	0.850
Street	0.350	0.400	0.666
Tiger	0.516	0.650	0.966
Average	0.687	0.752	0.825

Experiment 2: Tag suggestion for labelled images. Given an image labelled with some tags, it is aimed for automated methods that accurately suggest new tags relevant to the image. We investigate how our algorithm

improves upon a recent method by Sigurbjornsson and Van Zwol [38] by introducing visual content information into the tag suggestion process. Similar to [38], first x is computed, candidate tags having the highest co-occurrence with the initial tags. For each candidate tag, then compute its relevance score with respect to the image as,

score(c, I) = score (c,t_I) . $\frac{\gamma}{\gamma+(rank-1)}$..... (xii)

where c is the candidate tag, I the image, and t_I the set of initial tags. The function score(c, t_I) computes a relevance score between the candidate tag and the initial tags. V ote+ is adopted, the best method in [38], as an implementation of the score function. The input $rank_c$ is the position of tag c in the candidate tag list ranked by tag relevance in descending order. The variable γ is a positive parameter for regularizing the effect of tag relevance learning. By optimizing the algorithm on the same training set as used in [38], the optimized setting of the two parameters x and γ are determined as 17 and 20 respectively.



Experiment 3: Tag suggestion for unlabeled images. Compare with two model-free approaches: a tag frequency (tf) approach and an approach by Wang *et al.* [30] which re-weights the frequency of a tag by its inverse document frequency (tf-idf). For our algorithm, since no user-defined tags are available, all tags in the vocabulary are considered as candidates. Tag relevance for each candidate tag is estimated with respect to the unlabeled image, and then rank the tags in descending order by tag relevance. Care is taken to make the comparison fair. First, since the baselines do not consider user information, the unique-user constraint is removed from our algorithm. Second, for all methods the numbers of the visual neighbors are fixed to 100, as suggested in [30]. Finally, for each method, the top 5 tags are selected as a final suggestion for each test image.

In all the three experiments, baseline is used to represent the baseline methods, and tag relevance for our method.

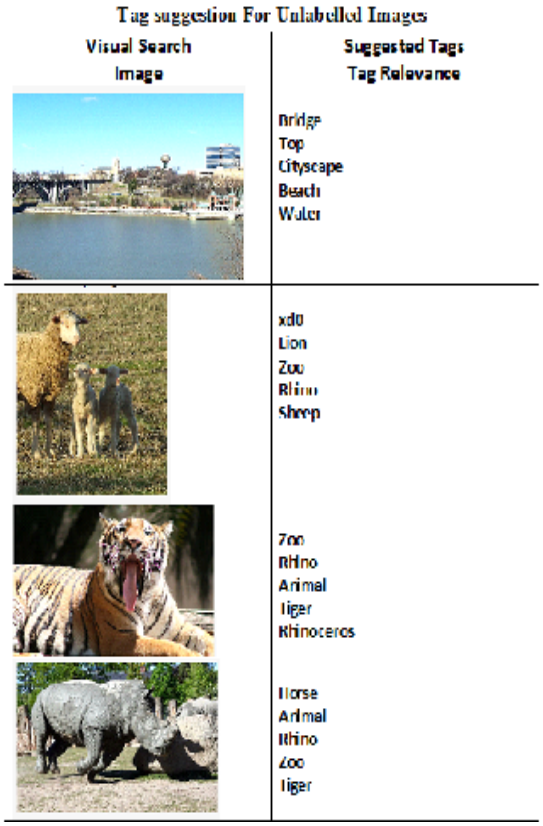
Evaluation set for tag suggestion. For evaluation of the performance of tag suggestion for labelled and unlabeled images, a ground truth set is adopted from [23], which is created by manually assessing the relevance of tags with respect to images. The set consists of 2000 Flickr images collection. Note that these tags might be predicted by tag suggestion methods. In that case, the tags are considered irrelevant. The number of tags per image in the evaluation set varies from 1 to 5. Examples of the ground truth are shown in Figure 5.

Evaluation Criteria-For image retrieval, images relevant with respect to user queries should be ranked as high as possible. Meanwhile, ranking quality of the whole list is important not only for user browsing, but also for applications using search results as a starting point. For tag suggestion, tags relevant with respect to user images should be ranked as high as possible. Also, the candidate tag list should be short such that users pick out relevant tags easily and efficiently. Thus, the following two standard criteria are adopted to measure the different aspects of the performance. Given a ranked list of L instances where an instance is an image for image retrieval and a tag for tag suggestion, we measure **precision at *n* (P@n)**: The proportion of relevant instances in the top n retrieved results, where n ≤ l. For image retrieval, we report P@1, P@5,P@10,P@15 and P@20 for each query. For tag suggestion, we report P@1 and P@5, averaged over all test images, as used in [30]. We consider a predicted tag relevant with respect to a test image if the tag is from the ground truth tags of the image. The Porter stemming is done before tag matching. Since we always predict 5 tags for each image, for those images having less than 5 ground truth tags, their P@5 will be smaller than 1.

Average precision (AP): AP measures ranking quality of the whole list. Since it is an approximation of the area under the precision-recall curve [43], AP is commonly considered as a good combination of precision and recall, e.g., [7]. The AP value is calculated as (1/R) $\sum_{i=1}^{l}\left(\frac{Ri}{i}\right)\delta i$ where R is the number of relevant instances in the list, Ri the number of relevant instances in the top i ranked instances, δi=1 if the i-th instance is relevant and 0 otherwise. For evaluation of the overall performance, we use mean average precision abbreviated as MAP, a common measurement in information retrieval.

MAP is the mean value of the AP over all queries in the image retrieval experiment and all test images in the tag suggestion experiments.

Large-scale Content-based Visual Search-To implement the neighbor voting algorithm, there is a need to define visual similarity between images and then search visual neighbors in our 2000 Flickr photo database. Visual similarity between two images is measured using corresponding visual features. To search millions of images by content, efficient indexing methods are imperative for speed up. A K-means clustering based method is adopted for its empirical success in largescale content-based image retrieval [32]. The search space is thus reduced. Since the search operation in individual subsets can be executed in parallel, neighbor search is executed in a distributed super computer.



4. CONCLUSION AND FUTURE WORK

Since user tagging is known to be subjective and overly personalized, a fundamental problem in social image analysis and retrieval is how to accurately interpret the relevance of a tag with respect to the visual content the tag is describing. In this paper, we propose a neighbor voting algorithm as an initial step towards conquering the problem. Our key idea is to learn the relevance of a tag with respect to an image from tagging behaviours of visual neighbors of that image. In particular, our algorithm estimates tag relevance by counting neighbor votes on tags. We show that when

1) The probability of correct user tagging is larger than the probability of incorrect user tagging and

2) Content-based visual search is better than random sampling.

The algorithm produces a good tag relevance measurement for both image ranking and tag ranking. Also, since the proposed algorithm does not require any model training for any visual concept, it is efficient in handling large-scale image data sets.

To verify the algorithm, three experiments were conducted on two thousand Flickr photos: one image ranking experiment and two tag ranking experiments. For the image ranking experiment, social image retrieval is improved by using learned tag relevance as updated tag frequency in a general tag-based retrieval framework.For the tag ranking experiments, two settings are considered, i.e., tag suggestion for labelled images and tag suggestion for unlabeled images.In the tag suggestion experiment for unlabeled images, the algorithm compares favourably against two baselines. Specifically, we effectively restrain high frequency tags without overweighting rare tags. This study demonstrates that the proposed algorithm predicts more relevant tags even when the visual search is unsatisfactory. In short, all the three experiments show the general applicability of tag relevance learning for both image ranking and tag ranking. The results suggest a large potential of our algorithm for real-world applications.

REFERENCES

[1] Rafael C. Gonzalez and Richard E. Woods, “Digital Image Processing”,Pearson Edu 2nd edition, 2005.

[2] Anil K. Jain, “Fundamental Of Digital Image Processing”, Pearson Education 1st edition, 1989.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” *IEEE Trans. PAMI*, vol. 22, no. 12, pp. 1349–1380, 2000.

[4] K. Weinberger, M. Slaney, and R. van Zwol, “Resolving tag ambiguity,” in *Proc. ACM Multimedia*, 2008, pp. 111–119.

[5] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, and T. Tuytelaars, “A thousand words in a scene,” *IEEE Trans. PAMI*, vol. 29, no. 9, pp. 1575–1589, 2007.

[6] X.-J. Wang, L. Zhang, X. Li, and W.-Y. Ma, “Annotating images by mining image search results,” *IEEE Trans. PAMI*, vol. 30, no. 11, pp. 1919–1932, 2008.

[7] D. Grangier and S. Bengio, "A discriminative kernel-based approach to rank images from text queries," *IEEE Trans. PAMI*, vol. 30, no. 8, pp. 1371–1384, 2008.
[8] G. Park, Y. Baek, and H.-K. Lee, "Majority based ranking approach in web image retrieval," in *Proc. CIVR*, 2003, pp. 499–504.
[9] Y. Jing and S. Baluja, "Visual Rank: Applying page rank to large-scale image search," IEEE Trans. PAMI, vol. 30, no. 11, pp. 1877–1890, 2008.
[10] K. S. Jones, S. Walker, and S. E. Robertson, "A probabilistic model of information retrieval: development and comparative experiments – part 2," Jour. Information Processing and Management, vol. 36, no. 6, pp.809–840, 2000.
[11] X. Li, L. Chen, L. Zhang, F. Lin, and W.-Y. Ma, "Image annotation by large-scale content-based image retrieval," in *Proc. ACM Multimedia*, 2006, pp. 607–610.
[12] X. Li, C. G. M. Snoek, and M. Worring, "Annotating images by harnessing worldwide user-tagged photos," in *Proc. ICASSP*, 2009 (in press).
[13] Yong Rui and Thomas S. Huang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", *Journal VCIR* 10, pp. 39–62, 1999.
[14] Lin Chen, Dong Xu, Ivor W. Tsang, and Jiebo Luo, "Tag-Based Image Retrieval Improved by Augmented Features and Group-Based Refinement" *IEEE Trans. Multimedia*,vol. 14, no. 4, Aug 2012.
[15] John Eakins, Margaret Graham, "Content-based Image Retrieval", Oct 1999.
[16] J. Zobel and A. Moffat. Inverted files for text search engines,*ACM Computing Surveys*, 2006.
[17] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low rank, content-tag prior and error sparsity. ACM Multimedia, pages 461–470, 2010.
[18] R. Datta, W. Ge, J. Li, and J. Z. Wang. Toward bridging the annotation-retrieval gap in image search by a generative modeling approach. *ACM Multimedia*, pages 977–986, 2006.
[19] M. E. I. Kipp and G. D. Campbell. Patterns and Inconsistencies in Collaborative Tagging Systems : An Examination of Tagging Practices. Annual General Meeting of the American Society for Information Science and Technology, 2006.
[20] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. Proceedings of the 16th International Conference on World Wide Web, pages 211–220, 2007.
[21] P. Heymann, G. Koutrika, and H. Garcia-Molina. Can social bookmarking improve web search? *Proceedings of the International Conference on Web Search and Web Data Mining*, pages 195–206, 2008.
[22] Y. Jin, L. Khan, L. Wang, and M. Awad, "Image annotations by combining multiple evidence & Wordnet," in *Proc. ACM Multimedia*, 2005, pp. 706–715.
[23] B. Sigurbj ¨ornsson and R. van Zwol, "Flickr tag recommendation based on collective knowledge," in *Proc. WWW*, 2008, pp. 327–336.
[24] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," Jour. Machine Learning Research, vol. 3, no. 6, pp. 1107–1135, 2003.
[25] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," *IEEE Trans. PAMI*, vol. 30, no. 6, pp. 985–1002, 2008.
[26] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, and T. Tuytelaars, "A thousand words in a scene," *IEEE Trans. PAMI*, vol. 29, no. 9, pp. 1575–1589, 2007.
[27] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Tagging over time: realworld image annotation by lightweight meta-learning," in *Proc. ACM Mutlimedia*, 2007, pp. 393–402.
[28] C. Wang, F. Jing, L. Zhang, and H.-J. Zhang, "Scalable search-based image annotation," *Multimedia Systems*, vol. 14, no. 4, pp. 205–220, 2008.
[29] X. Li, C. G. M. Snoek, and M. Worring, "Learning tag relevance by neighbor voting for social image retrieval," in *Proc. ACM MIR*, 2008, pp. 180–187.
[30] C. Wang, F. Jing, L. Zhang, and H.-J. Zhang, "Scalable search-based image annotation," Multimedia Systems, vol. 14, no. 4, pp. 205–220, 2008
[31] J. Huang, S. Kumar, M. Mitra, W. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. CVPR*, 1997, pp. 762–768.
[32] E. H¨orster, R. Lienhart, and M. Slaney, "Image retrieval on large-scale image databases," in Proc. CIVR, 2007, pp. 17–24.