Learning Tag Relevance by Context Analysis for Social Image Retrieval

Yong Cheng[1], Wenhui Mao[1], Cheng Jin[1], Yuejie Zhang[1], Xuanjing Huang[1], Tao Zhang[2]

[1]School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, P. R. China
[2]School of Information Management & Engineering, Shanghai University of Finance & Economics, Shanghai 200433, P. R. China
{13110240027, 13210240099, jc, yjzhang, xjhuang}@fudan.edu.cn, taozhang@mail.shfeu.edu.cn

Abstract. Tags associated with images significantly promote the development of social image retrieval. However, these user-annotated tags suffer the problems of noise and inconsistency, which limits the role they play in image retrieval. In this paper, we build a novel model to learn the tag relevance based on the context analysis for each tag. In our model, we firstly consider the user tagging habits and use a multi-model association network to capture the tag-tag relationship and tag-image relationship, and then accomplish the random-walk over the tag graph for each image to refine the tag relevance. Different from the earlier research work related to tag ranking, our contributions focuse on the globally-comparable tag relevance measure (i.e., can be compared across different images) and better tag relevance learning model by detailed context analysis for each tag. Our experiments on the public data from *Flickr* have obtained very positive results.

Keywords: Tag Relevance, Context Analysis, Tag Ranking, Image Retrieval.

1 Introduction

With the rapid growth of Internet and mobile devices, users are able to share social images more easily than before, and this leads to massive social images that are available online. Therefore, how to retrieval social images more efficiently and accurately has been an important research topic [1]. At present, the existing social image retrieval technology can be classified two categories, that is, content-based and annotation-based [2, 3]. Compare to the content-based approach, the annotation-based social image retrieval can supply better retrieval performance and more convenient user interface. However, this approach needs a lot of human-effort to annotate the images, which limits the application because of the high cost.

More recently, with the development of Web, there are some image sharing websites like *Flickr*, which allow users to upload images with some associated tags and these social tags can help solve the high-cost problem in annotation-based social image retrieval to some extent [4, 5]. Nevertheless, compared with the standard image annotations, the social image tags has some specific features and problems: 1) tagging accuracy, because the annotators are not experts, so inevitably there exists irrelevant tag, weakly-relevant tag, misspelling tags and so on. Fig. 1-*A* shows a social image of tiger with the associated tag set, obviously the tag "*tiger*" is the most relevant tag for this image, next "*wild*", and the other tags are either weakly relevant or irrelevant at all; 2) tag length and content, which means for the similar images, the length of tag sets and tag content may be very different. For example, when comparing Fig. 1-*A* with Fig. 1-*B*, we can see although these two images are similar, both the lengths and contents of these two tag sets are dissimilar. These characteristics make it difficult to apply traditional text retrieval model to image retrieval based on social tags. It can be also observed from Fig. 1 that for the query "*tiger*", if we utilize the classical tf-idf as the retrieval model, the final rank list will be <C, B, A>, which means for all the social image tag sets with "*tiger*", the less tags the image contains, the better rank the social image has, obviously this doesn't meet the actual situation. For example, although Fig. 1-*C* has the least tags, it is the least relevant image for "*tiger*".

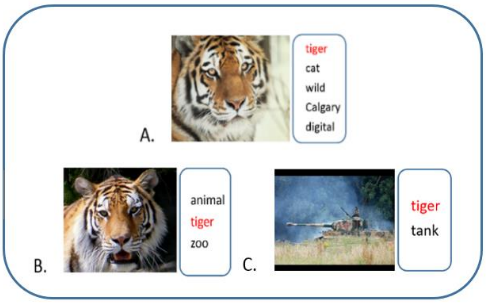

Fig. 1. An instantiation of social images from *Flickr*.

To solve all these problems above, the researchers have done a lot of related research work. Overall, these work mainly concerns with two aspects of tag set revise and tag ranking. Tag set revise contains tag enrichment and tag refinement, for example, Qian *et al*. built a graph for each image with its initial tags, and implemented the tag enrichment by a graph-cut approach [6]; and Xu *et al*. proposed to do tag refinement from the perspective of topic modeling and presented a graphical model called regularized Latent Dirichlet Allocation to exploit both the statistics of tags and visual affinities of images in the corpus [7]. Xia *et al*. and Sang *et al*. use concept ontology and user information to do the tag refinement [8, 9]. However, these approaches have the disadvantages that the noisy information may be introduced to tag set or remove useful information from tag set. And for tag ranking, which means rank the tag according their relevance to the image, for example, [10] is the first paper aiming to attack the tag ranking problem, they estimated initial relevance scores through probability density estimation, and then refined the initial scores by performing a random walk algorithm over a tag similarity graph. Zhuang *et al*. proposed a novel two-view learning approach to compute the tag relevance using both textual and visual contents of social images, and they mapped the learning into an optimization task and presented an efficient algorithm to solve it [11]. Xiao *et al*. also used both semantic and visual information to get the tag-tag similarity matrix, and then discovered the significance of each tag [12]. The existing tag ranking methods focused on the inner ranking in the tag sets, while for social image retrieval the emphasis should be put on the global comparison for the same query tag in different image tag sets. These methods only used rough ranking information for image retrieval. Obviously they ignored some useful information which can be used to improve retrieval performance.

Based on these observations above, a novel scheme is developed in this paper for facilitating social image retrieval by learning the relevance of the image tags. In our approach, a multi-model association network is constructed to model the relationship between tags; then we build a context-analysis model to compute the tag relevance and finally refine it using an approximate random-walk algorithm. Our scheme significantly differs from other work in:

1. our approach assume that a tag has high relevance when it has close relationship with other tags in the same image, and we make this assumption from the inspiration of the user tagging process for social images, generally users are less likely to tag two close-related tags as noises simultaneously, in other words, if user chose a tag for an image, the higher related this tag with other tags, the higher relevance this tag for the image. Our model well embodies this characteristic.
2. The relevance of tag can be compared globally, which means we use richer information to improve the performance of image retrieval and our experiments on a large number of public data from *Flickr* have obtained very positive results.
3. For multi-tag query, our model is more robust than other tag ranking approaches using only ranking information.

2 Multimodal Association Network Generation

For tags in the same tag set of a social image, the tag relevance is determined by the related context, that is, the relationship between a tag and the other tags in the same tag set and the relationship between a tag and the associated image. For the tag-tag inter-related relationship, we consider two aspects of the co-occurrence frequency and the visual similarity, because if two tags often co-occur in the same image tag set, or the visual feature of two tags are similar, we may infer the relationship between these two tags is close. For the tag-image inter-related relationship, we consider the visual similarity between the tag and the associated image. Thus a multi-model association network is utilized as an important knowledge source for tag relevance measurement.

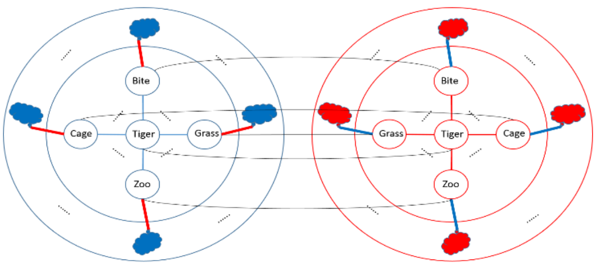

Fig. 2. The multimodal association network.

Our multimodal association network consists of two key components, i.e., the co-occurrence association network and the visual association network. As shown in Fig. 2, the left part is our co-occurrence association network, in which to better express what we mean, not all the tags are listed in the figure. Taking the tag "*tiger*" as an example, {"*bite*", "*cage*", "*grass*", "*zoo*" …} are all the tags co-occurring with "*tiger*" in at least one image, and all these tags and their neighbors constitute the entire association network. This network is formalized with a four-tuple <T, E, W_t, W_e>, where T represents a tag set that contains all the tags occur in the image database, and each tag can be seen as a node on the network, E represents a set of edges, which means if two tags co-occur in at least one image, we assume that there exist an edge between these two tag nodes. W_t represents the weight of the tag node t in T, and its corresponding value is set as the occurrence times of the tag *t*. Since each tag usually occurs only once in one image, this weight can be explained as the number of images which contain the tag t, as described in Formula (1). W_e represents the weight of the edge e in E, and its corresponding value can be set as the occurrence times of the edge e or the tag-tag pair. This weight can also be explained as the number of images that contain both of these two tags, as shown in Formula (2).

$$W_t = |IMG(t)| \tag{1}$$

$$W_e(t_i, t_j) = |IMG(e(t_i, t_j))| \tag{2}$$

The right part of Fig. 2 is our visual association network, which has the same structure with the co-occurrence association network, that is, these two networks have the same tag nodes and edges. The four-tuple <T', E', W'_t, W'_e> is aslo used to represent the visual association network, where T' and E' have the same meaning as in the co-occurrence association network and represent the tag set and the tag-tag pair set respectively; the value of W'_t is set as the visual feature vector of the tag t. which is calculted by using the mean histogram information of the tag; and the value of the W'_t is calculated by Formula (3).

$$W'_t = \frac{\sum_{img \in IMG(t)} HIST(img)}{|IMG(t)|} \tag{3}$$

where HIST(img) is the histogram feature of the image img; and IMG(t**)** is the image set, in which each image has the tag t. Here we do not use the complex visual feature, because the images from *Flickr* website are flexible and complex and the complex visual features may not be suitable. The value of W'_e is the Euclidean distance between two tags with the edge e, which is calculated by Formiula (4).

$$W'_e(t_i, t_j) = EU\_DIS\left(W'_{t_i}, W'_{t_j}\right) \tag{4}$$

where the function EU_DIS returns the Euclidean distance between the visual feature vectors of two tags of t_i and t_j.

3 Tag Relevance Learning

Our tag relevance measure model consists of two steps, firstly an initial relevance score is obtained through the context analysis for each tag on the basis of our multi-modal association network, and then a random-walk algorithm is adopted to make a refinement for the initial score of tag relevance.

3.1 Initial Measure based on Context Analysis

When we try to understand a certain word in a text, except for the original meaning of this word, the word's context is often utilized to help us and plays an important role for our understanding. Similarly, when we try to measure the tag relevance, the context around the tag can be utilized. For a given tag, the context includes the other tags in the same tag set for the corresponding social image and the image itself. Considering the first image in Fig. 1 as an example, the context of the tag "*tiger*" is shown in Fig. 3, thus we can calculate the initial tag relevance according to the context analysis. Here the "*object tag*" and "*context tags*" are used to represent the object tag and other tags in the same tag set for an image. Firstly, for the relationship between the object tag and the context tags, we assume that each tag of the context tags can provide its own contributions for the relevance of the object tag, and the contribution can be measured from two aspects of the co-occurrence frequency and the visual similarity. If two tags often co-occur in the same image, it means that users tend to annotate these two tags simultaneously for the same image. Generally users won't annotate two close-related tags as noises simultaneously, and in fact if the object tag has the higher co-occurrence frequency with the context tags, its relevance will be higher. For the object tag *otag*, the contribution of a tag *t* in the context tags can be calculated by Formlual (5).

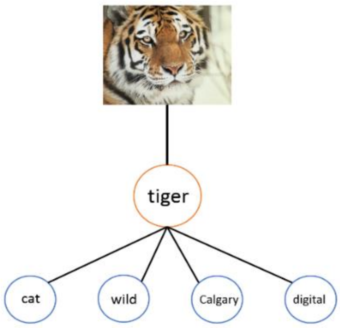

Fig. 3. The context of the tag "tiger" in the image Fig. 1-A.

$$\text{Contribute}(\text{t}, \text{otag}) = \frac{\text{W}_\text{e}(\text{t}, \text{otag})}{\text{W}_\text{otag}} \tag{5}$$

where W_{otag} and W_e are defined in our co-occurrence association network, and can be calculated by Formulae (1) and (2). Since the lengths of different "context tags" are different, we set a normalized weight for each tag in the context tags to avoid the influence of the length, and the value of weight can be calculated by Formula (6).

$$\text{Weight}(\text{t, otag})= \frac{W_e(\text{t, otag})/W_t}{\sum_{T\in \text{ContextTags}} W_e(\text{T, otag})/W_T} \quad (6)$$

Thus we can measure the relevance of the object tag from the view of co-occurrence association by Formula (7).

$$\text{TR}(\text{otag})= \sum_{t\in \text{ContextTags}} \text{Weight}(\text{t,otag})*\text{Contribute}(\text{t,otag}) \quad (7)$$

Similarly, we measure the contribution of context tags from the view of visual association in a similar manner. Something different is that we replace the co-occurrence property with the visual similarity, and the exponential function of the Euclidean distance is used to represent the visual similarity, as shown in Formula (8).

$$\text{VR}(\text{otag,img})= \exp(\text{-Dis}(\text{otag,img})) \quad (8)$$

$$\text{Dis}(\text{otag,img})= \sum_{t\in \text{ContextTags}} \text{Weight}(\text{t,otag})*\text{W}'_e(\text{t,otag})$$

where *img* means the corresponding image to the object tag. Expect the relationship between the object tag and context tags, the relevance of object tag is also affected by the relationship with the image itself, and the visual similarity between them can be calculated by Formula (9).

$$\text{VR}'(\text{otag,img})= \exp(\text{-Dis}'(\text{otag,img})) \quad (9)$$

$$\text{Dis}'(\text{otag,img})= \text{EU\_DIS}(\text{tag,img})$$

where *EU_DIS* is the Euclidean distance between the object tag and the image. Since Formulae (8) and (9) are all about the visual similarity, for simplifying our model we combine Formulae (8) and (9) as Formula (10).

$$\text{VR}_f(\text{otag,img})= \exp(\text{-Dis}_f(\text{otag,img}) \quad (10)$$

$$\text{Dis}_f(\text{otag,img})= 0.5*\sum_{t\in \text{ContextTags}} \text{Weight}(\text{t, otag})*\text{W}'\text{e}(\text{t, otag}) +0.5*\text{EU\_DIS}(\text{tag,img})$$

Finally we combine Formulae (7) and (10) and get the final initial relevance measure for the object tag, as shown in Formula (11).

$$\text{Relvance}(\text{otag, img})= \alpha\text{TR}(\text{otag})+(1\text{-}\alpha)\text{VR}_f(\text{otag,img}) \quad (11)$$

where α is a weight parameter ranging from 0 to 1.

3.2 Tag Relevance Refinement Through Random Walk

After acquiring the initial relevance score for each tag in the image annotation, we make a further refinement for the tag relevance using an approximate random-walk algorithm. Here, we refer the method implemented in [10], and accomplish some change in our work for the purpose of getting globally-comparable tag relevance. First, we see the tag set in each image as a full-connected graph, and the tag is the node in the graph, then a random-walk process is run on the graph based on the similarity between the tags, and finally more refined tag relevance can be obtained. We use the exponential function of the Google Distance [13] to compute the similarity between tags, as shown in Formula (12).

$$S(t_i, t_j) = \exp(-d(t_i, t_j)) \tag{12}$$

$$d(t_i, t_j) = \frac{\max\left(\log W_{t_i}, \log W_{t_j}\right) - \log W_e(t_i, t_j)}{\log G - \min(\log W_{t_i}, \log W_{t_j})}$$

where G is the total image numbers in the database. Based on this similarity measure, we can give the transition probabilities used in the process of random walk, as shown in Formula (13).

$$p_{ij} = \frac{S(t_i, t_j)}{\sum_k S(t_i, t_k)} \tag{13}$$

where p_{ij} is the transition probability from the tag t_i to the tag t_j; k is the total number of the neighbor tags of the tag t_i. Hence this formula can be seen a normalization process. We then run the random walk process on each image individually, as shown in Formula (14).

$$r_k(j) = \frac{1}{2m} \sum_i r_{k-1}(i) p_{ij} + \frac{1}{2} v_j \tag{14}$$

where $r_k(j)$ represents the relevance of the tag t_j in the k_{th} iteration; p_{ij} is the transition probability; V_j is the initial relevance of the tag t_j, and m is the total number of context tags for tag t_j. The reasons for using parameter m is that we want to exclude the impact of the tag set length, and it is fair to the images have different tag set length. For the same reason, we don’t consider the convergence of Formula (13), and just conducted a limited number of iterations. Another reason for this is that our initial relevance is relatively accurate, thus just a small number of iterations are needed to refine the relevance. On the contrary, too much iteration will reduce the accuracy.

4 Experiment and Analysis

4.1 Dataset and Query Definition

Our dataset is established based on *MIR Flickr Data* constructed by [14] and further development by [15] which have been used in *ImageCLEF*. It consists of 1,000,000 annotated images collected from *Flickr* with unconstrained contents, and these images have all the original user annotations. In our experiment, we use a subset dataset with about 200,000 images for retrieval just as in *ImageCLEF 2012*, which contains about 37863 tags after filtration. We choose 26 queries as the test query in our experiment, as shown in table 1, these queries cover various categories like *Animal, Building,* *Human, Event, Object,* and so on. Except 20 single-word queries, we also introduce 6 multi-word queries to see the performance of our model on different kinds of queries. As for ground truth collection, due to the large collection of 200 thousand images, we don't obtain all the ground truth relevance assessments for each query. We first get the retrieval results for each query using all the models we use in our experiment, and then we form a pool for each query by aggregating all these returned images and obtained the relevance assessments by annotating manually. Each image was labeled "relevant" or "irrelevant", which we then used to evaluate the results of each model.

Table 1. The test queries in our experiment

single-word query	*bride, church, fish, ring, piano, pizza, tiger, mother, bus, rainbow, football, couple, bedroom, winter, bike, sheep, lion, monkey, fire, reading*
multi-word query	*old man, apple cellphone phone, beach sunset sunrise, flower bee, girl reading book, horse riding*

4.2 Evaluation Metrics and Model Selection

In our work, we choose two evaluation metrics to assess the performance of social image retrieval via more effective tag relevance measurement.

- *AP@X*: This evaluation measure the average precision when a certain number of images are encountered. In our work, we use AP@5, AP@10, AP@50, AP@100 respectively to measure the experiment results.
- *MAP*: Here we focus on the non-interpolated MAP, that is, each time a relevant image is encountered from top to bottom, we compute the precision, and finally get the mean of these precision values.

For the model selection, there are two parameters we need to set: α in Formula (12) and the iteration numbers for Formula (13), When we set $\alpha = 0.7$ and iter = 5, the best retrieval performance can be acquired, as shown in Fig. 4.

4.3 Experimental Result

To investigate the effect of each part on the whole retrieval performance, we introduce four evaluation patterns.

- TEXT: use only textual information, which means we set the parameter $\alpha = 1$ in Formula (11).
- VISUAL: using only the visual information, which means we set the parameter $\alpha = 0$ in Formula (11).
- TEXT+VISUAL**:** using both textual and visual information, which means we choose the optimal value of α. As Fig. 4 shows, here we set $\alpha = 0.7$.
- TEXT+VISUAL+RANDOM**:** Making further refinement by using our random-walk algorithm based on the result of Formula (11).

Fig. 5 shows the results of above four patterns. It can be seen that the performance is 4) > 3) > 1) > 2) and this result can illustrate that each part in our model can contribute for the final performance. We can find that the performance is poor by only using the visual information. However, when combining with the textual information, the result is much better than only using the textual information. It can be also observed that the random-walk process is able to further improve the final result.

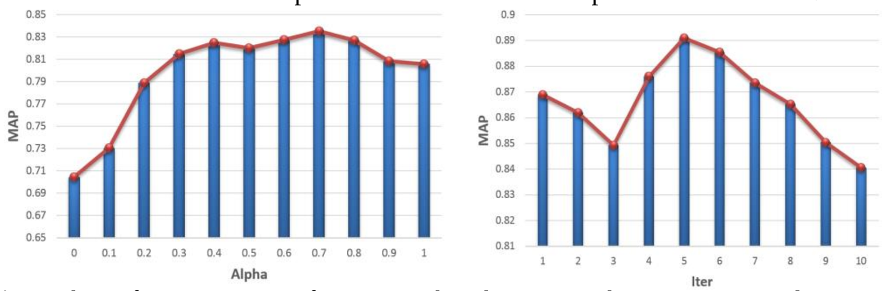

Fig. 4. The performance curves of our approach with respect to the parameters α and iterations.

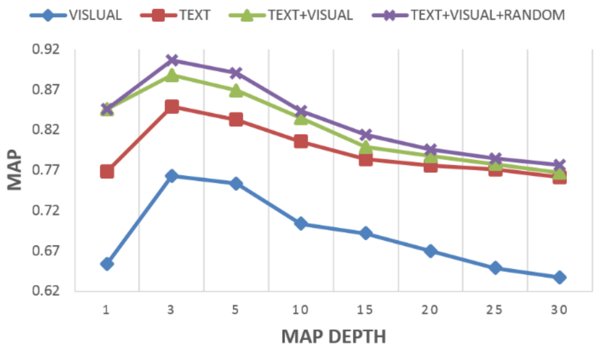

Fig. 5. The experiment results for the four cases in our approach.

To give full exhibition to the superiority of our alignment model, we have also performed a comparison between our and the other classical approaches in recent years. Since our work is mainly for social image retrieval, we choose two categories of existing approaches as our comparative experiments: one is traditional text retrieval model for annotation-based image retrieval, the other is tag ranking approaches for social image retrieval. For the first category, we choose two retrieval model, one is a language model implemented in [16], the other is a information-based model implemented in [17]. As for the tag ranking method, we also choose two methods, one is the tag ranking model in [10], for the other method, we combine our method with the method in [10].

- ***Language Model (LM):*** The tags for each image can be viewed as a document, the basic idea is to estimate a language model for each document, and documents are ranked by the likelihood of the query according to the estimated language model.
- ***Information-Based Model (IBM):*** This model is based on a hypothesis that the significance of a tag in the document can benefit from the difference of the behaviors of this tag at the document and collection levels.

- ***Tag Ranking Model (TRM):*** This model first gets local relevance scores using the probability density estimation and random walk algorithm and ranks the tags using these scores, and then use the rank information to get the global relevance for image retrieval by Formula (15).

$$\mathrm{relevance(t)=\text{-}rank_t + 1/N_t} \tag{15}$$

where $rank_t$ denotes the rank value of tag t, and N_t denotes the number of tags in the corresponding tag set.

- ***Joint Model (JM):*** We first get the tag relevance by using our method, however, just use these relevance values to rank the tags and then refer Formula (15) to compute the new relevance. We want to show the advantage of our approach in using more rich information to improve the retrieval performance.

In addition, method 3 and 4 both use the vector space model as the retrieval model, just as our approach. This means for single-word queries, we just rank the images according to the relevance of the word; for multi-word queries, we use the sum of the relevances to rank the images. Since the work in [Liu et al., 2009] used only single-word queries, for fairness, we first use the twenty single-word queries in Table 1 in our experiment, and the final results can be seen in Table 2. We can see that our approach outperforms other approaches obviously, and the comparison result between *JM* and our approach confirms the global comparability of our relevance and the advantage of our model in capturing the appropriate information to compute the relevance. We then implement the experiments on each query in Table 1. For the purpose of clarity, we just list the results of *LM*, *TRM* and our approach in Fig. 7, here we choose *AP@10* as our evaluation metrics since we usually pay attention to the top images of the return list. It can be seen that for single-word queries, our approach outperforms other approaches obviously, and our approach's performance is the best for almost all the queries. As for multi-word queries, our approach is more robust than tag ranking method *TRM* and has roughly the same results with *LM*. the reason for this is probably because that the presence of all query words in a tagged image largely guarantees the high relevance for the image, and the poor peformance of *TRM* once again verified the coarse of the rank information for image retrieval. At last, we show the top 5 retrieval results of these models in Fig. 6 for the query "*lion*", "*sheep*". It can be seen that generally our approach can exactly return the better retrieval results.

Table 2. The comparison between our and the existing methods.

	AP@5	AP@10	AP@50	AP@100	MAP(100)
LM	0.63	0.595	0.607	0.596	0.623
IBM	0.63	0.595	0.607	0.596	0.623
TRM	0.63	0.62	0.61	0.6	0.631
CM	0.61	0.63	0.646	0.605	0.654
Our Approach	**0.8**	**0.77**	**0.694**	**0.642**	**0.722**

Fig. 6. The top 5 images returned by different approaches for the queries "lion", "sheep".

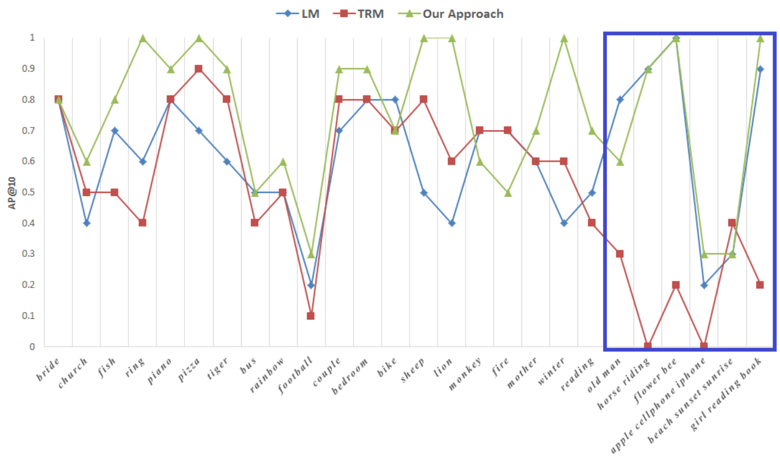

Fig. 7. The experiment results of different approaches on all the queries.

5 Conclusions

In this paper, we build a relevance learning model for the tags associated with the social images, and use the tag relevance to improve the performance of image retrieval. Our model considers the user tagging habits and try to compute the tag relevance by analyzing the context of the tag. For computing the relevance, a multi-model association network is constructed as our knowledge source and a random-walk algorithm is adopted to refine the relevance. The experiment results demonstrate that our approach can make use of rich information to compute the relevance and our tag relevance can help improving the retrieval performance significantly. For the future work, we will put more focus on multi-word queries, since different query words are not independent, simply consider the sum of the relevances of the query words are not enough, so our future work will consider the semantic relationships between them to further improve the performance.

Acknowledgments. This work is supported by National Science & Technology Pillar Program of China (No. 2012BAH59F04), National Natural Science Fund of China (No. 61170095; No. 71171126), and Shanghai Municipal R&D Foundation (No. 12dz1500203, 12511505300).

References

1. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys* (*CSUR*), 40(2), Article 5, 2008.
2. Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), 394-410, 2007.
3. Sun, A., Bhowmick, S.S., Tran, K., Nguyen, N., Bai, G.: Tag-Based social image retrieval: An empirical evaluation. *Journal of the American Society for Information Science and Technology*. 62(12), 2364-2381, 2011.
4. Goh, D.H., Ang, R.P., Lee, C.S., Chua, A.Y.K.: Fight or unite: Investigating game genres for image tagging. *Journal of the American Society for Information Science and Technology*. 62(7), 1311-1324, 2011.
5. Jain, R., Sinha, P.: Content without context is meaningless. In *Proceedings of MM 2010*. 1259-1268, 2010.
6. Qian, X., Hua, X.: Graph-cut based tag enrichment. In *Proceedings of SIGIR 2011*. 1111-1112, 2011.
7. Xu, H., Wang, J.D., Hua, X.S., Li, P.: Tag refinement by regularized LDA. In *Proceedings of MM 2009*. 573-576, 2009.
8. Xia, Z.Q., Peng, J.Y., Feng, X.Y., Fan, J.P.: Social Tag Enrichment via Automatic Abstract Tag Refinement. *Advances in Multimedia Information Processing*. 7674, 198-209, 2012.
9. Sang, J.T., Liu, J., Xu, C.S.: Exploiting user information for image tag refinement. In *Proceedings of MM 2011*, 1129-1132, 2011.
10. Liu, D., Hua, X.S., Yang, L.J., Wang, M., Zhang, H.J.: Tag ranking. In *Proceedings of WWW 2009*, 351-360, 2009.
11. Zhuang, J.F., Hoi, S.C.H.: A two-view learning approach for image tag ranking. In *Proceedings of WSDM 2011,* 625-634, 2011.
12. Xiao, J., Zhou, W.G., Tian, Q.: Exploring tag relevance for image tag re-ranking. In *Proceedings of SIGIR 2012*, 1069-1070, 2012.
13. Cilibrasi, R.L., Vitanyi, P.M.B.: The Google Similarity Distance. *IEEE Transactions on Knowledge and Data Engineering*. 19(3), 370-383, 2007.
14. Huiskes, M.J., Lew, M.S.: The MIR flickr retrieval evaluation. In *Proceedings of MIR 2008*, 39-43, 2008.
15. Huiskes, M.J., Thomee, B., and Lew, M.S.: New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative. In *Proceedings of MIR 2010*. 527-536, 2010.
16. Zhai, C.X., Lafferty, J.: A study of smoothing methods for language models applied to Ad Hoc Information retrieval. In *Proceedings of SIGIR 2001*. 334-342, 2001.
17. Clinchant, S., Gaussier, E.: Information-based models for ad hoc IR. In *Proceedings of SIGIR 2010*. 234-241, 2010.