ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

Volume 3, Special Issue 3, March 2014

2014 International Conference on Innovations in Engineering and Technology (ICIET'14)
On 21st & 22nd March Organized by

K.L.N. College of Engineering, Madurai, Tamil Nadu, India

Study On Tag Refinement And Tag Completion For Effective Image Retrieval

Saambavi.M , Azarudeen.K

Student, Dept of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, India.

Professor, Dept of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, India.

Abstract: Online sharing of images is increasingly becoming popular, resulting in the availability of vast collections of user contributed images that have been annotated with user supplied tags. Many social image search engines are based on keyword/tag matching. It is because tag-based image retrieval (TBIR) is not only efficient but also effective. The performance of TBIR is highly dependent on the availability and quality of manual tags. Since many users tend to choose general and ambiguous tags in order to minimize their efforts in choosing appropriate tags, they are usually incomplete and insufficient to describe the whole semantic content of corresponding images resulting in unsatisfactory performances in tag related applications. This is a study on various techniques which are used to complete the missing tags and correct the noisy tags for given images thereby improving the retrieval performance.

Keywords: Tag completion, Tag refinement, Tag

I. INTRODUCTION

With the rapid advance in the technology of digital imaging, there is an explosive growth in the amount of available image data in our daily lives. These images can be retrieved in two different ways :content-based image retrieval and keyword/tag-based image retrieval[8].

CBIR takes an image as a query and identifies the matched images based on the visual similarity between the query image and the images in the gallery[7]. Various visual features, including both global features and local features are used in CBIR. Despite the significant efforts, there exist three limitations which restrict its practicability.
Firstly, the precision of CBIR is usually unsatisfactory because of the semantic gap between low-level visual features and high-level semantic concepts. Secondly, the efficiency of CBIR is usually low due to the high dimensionality of visual features. Thirdly, the query form of CBIR is unnatural for image search owing to the possible absence of appropriate example images.

TBIR is a straightforward solution to conquer the disadvantages of CBIR.TBIR allows a user to present his/her information need as a textual query and find the relevant images based on the match between the textual query and the manual annotation of images. The text information used can be acquired from image title, surrounding text and user tag.

There into, user tags are more consistent with semantic concepts and effective to describe image contents. Especially with the prevalence of photo sharing websites such as Flickr and Picasa, which host vast of digital images with user provided tags, tag based image retrieval has become potentially popular and practical in extensive applications. Nevertheless, the performance of tag based image retrieval is still far from satisfactory suffering from the inferior quality of image tags.

In most cases, the tags are provided by the users who upload their images to the social media sites (e.g., Flickr). These tags are often incomplete and noisy. For example, many users make slips in spelling when entering tags. The noisy tags are tags that are irrelevant to the image content. This includes meaningless tags, tags that contain typographical errors and tags that are not visually related to the image content. This noisy tags will introduce false positives into user's search result and incomplete tags will make the actually related images inaccessible. The presence of noisy tags can be due to several reasons. Most people interpret an image using their imagination and sometimes same tags are used to annotate different images that have been captured during the same event.

In addition, user-provided tags are often biased towards personal perspectives and thus there is a gap between these tags and the content of the images that common users are interested in. For example, an image uploader may tag his dog photos with "bomb", and it may make these photos appear in the search results of query "bomb". On the other hand, many potentially useful tags may be missed, as it is impractical for average users to annotate the images comprehensively.

To summarize, user-provided tags are often *imprecise*, *biased* and *incomplete* for describing the content of the images. Thus there is a gap between these tags and the actual content of the images. This leads to performance degradations of various tag-dependent applications like tag based image retrieval. The most straightforward approach to tackle the difficulty is to ask humans to check the tags associated with each image, but obviously this way is infeasible considering the large number of images and tags. Therefore, refining tags is thus highly desirable for tag based image retrieval and other related applications.

II. REVIEW WORKS

A. Tag Refinement

1) Neighbor Voting Algorithm: Xirong Li and Cees G. M. Snoek uses Neighbor voting algorithm in order to learn the relevance of a tag to an image. This algorithm learns tag relevance accurately and efficiently by accumulating votes from visual neighbors[3]. It consist of two steps retrieval of visual neighbors of a given input image and tag relevance learning.

Intuitively, if different persons label visually similar images using the same tags, these tags are likely to reflect objective aspects of the visual content. The intuition implies that the relevance of a tag with respect to an image might be inferred from tagging behavior of visual neighbors of that image: the more frequent the tag occurs in the neighbor set, the more relevant it might be.

The key idea is, by propagating common tags through visual links introduced by visual similarity, each tag accumulates its relevance credit by receiving neighbor votes. This algorithm is a good measure of both image ranking and tag ranking. Hence a good tag relevance measurement should take into account the distribution of a tag in the neighbor set and in the entire collection, simultaneously.

In the Figure 1 since four neighbour images are labeled with 'bridge', the tag relevance value of 'bridge' with respect to the input image is 4. Since the algorithm does not requires any model training for any visual concept , it is efficient in handling large scale image dataset. This algorithm predicts more relevant tags even when the visual search is unsatisfactory.

2) Tag Quality Improvement: Dong Liu, Meng Wang, Linjun Yang, Xian-Sheng Hua and HongJiang Zhang proposed a scheme to improve poorly annotated tags associated with social images[1]. This scheme is based on the consistency of visual similarity and semantic similarity of images. Here the semantic similarity of two images is defined as the similarities of their tag sets. In addition the improved tags provided by this scheme must not change too much from the initial tags since the initial user provided tags carry valuable information. It handles uncontrolled vocabulary of tags and diversity of social images.

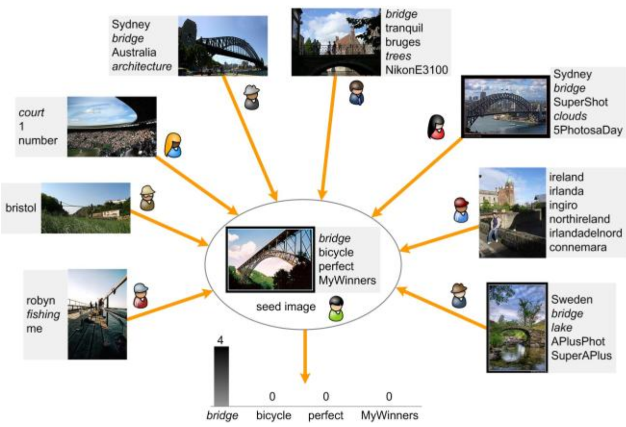

Fig 1. Learning Tag Relevance by Neighbor Voting

3) Tag Categorization and Neighbor Voting: Sihyoung Lee, Wesley De Neve and Yong Man Ro proposed a modular approach towards tag refinement, taking into account the nature of tags. They utilized neighbor voting to learn the relevance of each tag, and then differentiated noisy tags from correct ones[2]. First, tags are automatically categorized in five categories using WordNet: *'where', 'when', 'who', 'what',* and *'how'*. Next, neighbor voting algorithm is used to learn the relevance of tags along the '*what*' dimension. To automatically categorize the tags WordNet is used.

Tags with a relevance value lower than a particular threshold value are considered to be noisy, and these tags are removed. This technique is able to successfully differentiate correct tags from noisy tags along the '*what*' dimension. In addition, this technique is able to improve the effectiveness of image tag recommendation for non-tagged images. This is the first attempt to make use of tag categorization.

This technique is more efficient than neighbor voting algorithm as it reduces more noisy tags and it improves the effectiveness of image tag recommendation for non-tagged images.

Table I. Mapping of WordNet Categories onto Sihyoung Lee Categories

Sihyoung Lee categories	WordNet semantic noun categories

what	animal, artifact, attribute, body, food, object, phenomenon, plant, shape, something, substance
where	location, space
who	person, group
when	time, event
how	act, cognition, communication, feeling, motive, possession, process, quantity, relation, state

.

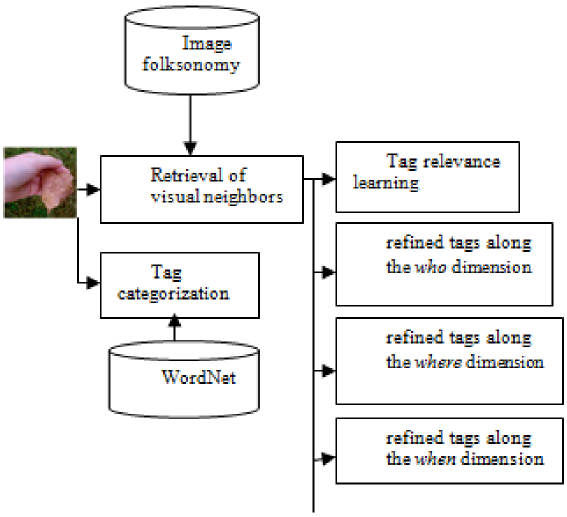

Fig 2.The Architecture of the Tag Refinement System

4) Retagging: Dong Liu, Xian-Sheng Hua, Meng Wang and Hong Jiang Zhang proposed a retagging scheme that assigns images with better content descriptors. The goal of this scheme is to bridge the gap between the tags and the content of images. This scheme considers both "visual similarity" and "semantic similarity" in social images[4]. Due to this consideration the improved tags will not be much deviated from the initially user-provided tags.

As many tags are intrinsically not closely related to the visual content of the images, a knowledge based method to differentiate visual content related tags from unrelated ones is employed. In order to improve the coverage of the tags, the tag set is enriched with appropriate synonyms and hypernyms based on external knowledge base.

This scheme consists of three stages : tag filtering, tag refinement and further enrichment. Tag filtering eliminates content-unrelated tags based on a vocabulary that is built with both WordNet lexical knowledge base and domain knowledge in vision field. With the obtained content related tags, refinement is performed on them by modeling the consistency of visual and semantic similarities between images. Finally, the further enrichment component expands each tag with appropriate synonyms and hypernyms by mining the WordNet lexical knowledge base as well as the statistic information on social image website.

5) Tag refinement by pursuing the low-rank, content consistency, tag correlation and error sparsity:

Guangyu Zhu, Shuicheng Yan and Yi Ma formulated the tag refinement problem as a decomposition of the user-provided tag matrix D into a low-rank refined matrix A and a sparse error matrix E, namely $D = A + E$[5] . The optimality is measured by four aspects : 1) *low-rank*: A is of low-rank owing to the semantic correlations among the tags; 2) *content consistency*: if two images are visually similar, their tag vectors (*i.e.*, column vectors of A) should also be similar; 3) *tag correlation*: if two tags co-occur with high frequency in general images, their co-occurrence frequency (described by two row vectors of A) should also be high; and 4) *error sparsity*: the matrix E is sparse since the tag matrix D is sparse and also humans can provide reasonably accurate tags.

The low-rank and error sparsity is integrated into the optimization procedure for image tag refinement. With the assistance of constraints of content consistency and tag correlation, the approach is capable of correcting imprecise tags and enriching the incomplete ones. An accelerated proximal gradient method is proposed to speed up the optimization, which facilitates approach to be workable on large-scale image datasets.

B. Tag Completion

1) Linear Sparse Reconstruction: Zijia Lin, Guiguang Ding, Mingqing Hu, Jianmin Wang and Xiaojun Ye proposed a novel scheme denoted as LSR for automatic image tag completion via image-specific and tag-specific **L**inear **S**parse **R**econstructions. Specifically, each image and tag is optimally reconstructed with remaining ones under constraints of sparsity, and then the reconstruction values from both perspectives are normalized and merged for predicting the relevance of unlabelled tags[6].

The sparsity constraints are attributed to the observation that generally an image contains a few objects and a tag connotes a few levels of meaning, and usually corresponding objects or levels of meaning are redundantly contained or implied in the context. Regarding the image specific reconstruction, both low-level image features and high-level tagging row vectors were considered.

As for the tag-specific reconstruction, the corresponding column vectors in the initial tagging matrix, which essentially mines their concurrence for seeking unlabelled high-confidence tags with initially labelled ones within an image is considered. LSR is a unified framework merging image-image similarity, image-tag association and tag-tag concurrence for tag completion. This algorithm performs tag completion for each row and column separately, instead of performing global refinement for the tagging matrix

III. COMPARISON

Table II. Comparison of Various Tag Refinement and Completion Methods

Method	**Technique Used**	**Usage**	**Advantage**

Neighbor voting algorithm	Tag relevance is learned by accumulating votes from visual neighbors	Tag Refinement	Predicts more relevant tags when the visual search is unsatisfactory
Tag quality improvement	Visual and semantic similarity consistency	Tag Refinement	The improved tags didn't change from the initial tags
Tag Categorization and Neighbor Voting	Tags are categorized and neighbor voting algorithm is used	Tag Refinement	Successfully differentiates correct tags from noisy tags along the '*what*' dimension
Retagging	Visual and semantic similarity consistency	Tag Refinement	Used in tag based image search and tag based automatic annotation
Tag refinement by pursuing the low-rank, content consistency, tag correlation and error sparsity	The user provided tag matrix D is divided into a low-rank refined matrix A and a sparse error matrix E, namely $D = A + E$.	Tag Refinement	Works in large scale image datasets
Linear Sparse Reconstruction	Each image and tag is optimally reconstructed and then the reconstruction values are normalized and merged	Tag Completion	Tag completion is performed for each row and column separately

IV.CONCLUSION

The performance of TBIR depends mainly on the tags provided by the users. Most of the user supplied tags are not relevant to the images and many are incomplete. This paper presents study on various tag refinement and completion techniques designed for effective image retrieval. Many concentrate only on tag refinement. These techniques aims at improving the quality of the tags associated with social images. A comparison of these techniques are also given.

REFERENCES

[1] Dong Liu, Meng Wang , Linjun Yang , Xian-Sheng Hua , HongJiang Zhang." Tag Quality Improvement for Social Images". In ICME, pages 350-353, 2009.

[2] S. Lee, W. Neve, and Y. Ro. Image tag refinement along the 'what' dimension using tag categorization and neighbor voting. In ICME '10.

[3] X. Li, C. G. M. Snoek, M. Worring, "Learning Social Tag Relevance by Neighbor Voting," IEEE Trans. Multimedia, vol. 11, no. 7, pp. 1310-1322, Aug. 2009.

[4] D.Liu, X. Hua, M. Wang, and H. Zhang. "Image retagging" In MM'10.

[5] G. Zhu, S. Yan, and Y. Ma, "Image Tag Refinement towards Low Rank, Content Tag Prior and Error Sparsity," Proc. ACM Int'l Conf. Multimedia, pp. 461-470, 2010.

[6] Zijia Lin, Guiguang Ding, Mingqing Hu, Jianmin Wang, Xiaojun Ye." Image Tag Completion via Image-Specific and Tag-Specific Linear Sparse Reconstructions". In CVPR, pages1618-1625,2013

[7] M.S. Lew, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," ACM Trans. Multimedia Computing, Comm. and Applications, vol. 2, pp. 1-19, 2006

[8] L. Chen, D. Xu, I.W. Tsang, and J. Luo, "Tag-Based Web Photo Retrieval Improved by Batch Mode Re-Tagging," Proc.IEEEConf. ComputerVision and Pattern Recognition, pp. 3440-3446, 2010

[9] K. Yanai and K. Barnard. Image region entropy: A measure of "visualness" of web images associated with one concept. In MM, pages 419-422, 2005.

[10] Y. Lu, L. Zhang, Q. Tian, and W. Ma. "What are the high-level concepts with small semantic gaps?" In CVPR, pages 1-8, 2008.

[11] S. Overell, B. SigurbjÄornsson, and R. Zwol. "Classifying tags using open content resources". In WSDM, pages 64-73, 2009.

[12] M. J. Huiskes and M. S. Lew, "The MIR Flickr Retrieval Evaluation," Proc. ACM International Conference on Multimedia Information Retrieval, pp. 39-43, Oct. 2008.