Volume 4, Issue 9, September 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Tag Completion using Annotation for Reliable Image Retrieval

Wakchaure Sujit R*
Dept. of Computer Engg, SPCOE
India

Prof. Chandre Pankaj R.
HOD, Dept. of Computer Engg, SPCOE
India

Abstract— Image sharing is much popular in worldwide, so it is vast to distinguish such images in efficient way based on their tags. User given tags are not relative, consistent and complete. They are noisy, missing and dissimilar to image content. User-provided textual tags of web images are widely utilized for facilitating image management and retrieval. It results into performance degradations. The goal of this work is to develop a probabilistic model to take into account the dependencies between keywords so as to provide superior image search, retrieval, more precise, corrected annotations. The main findings suggest that, under certain conditions, taking into account keyword completion, coupled with an efficient method to search over sets of tags is an effective method to increase annotation accuracy. A schematic is developed to illustrate the dynamic relationship between the motivations and their contexts. The schema will make an important contribution by annotation and completing tags what is an otherwise completely unknown area of image retrieval. This study based 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— TBIR, Tag completion, Matrix completion, Image annotation, Image retrieval

I. INTRODUCTION

Image retrieval typically refers to the automated process of retrieval of images from a database by low-level image attributes such as colour, shape, or texture. TBIR [2] 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.

The present study investigates amateur users' motivations to interact with digital images in online environments and situates the user in the contexts of their behaviour and interactions with images. Image retrieval scholarship in Library and Information Science (LIS) has proceeded over the past three decades by focusing on the attributes of the image, the user's image queries, and less frequently until recently, aspects of the user and her behaviour. This study takes a position different from the dominant stance in the field in that it questions whether amateur users actually may have reasons to pursue images online for more nuanced purposes other than 'retrieval.'

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.

This work proposes the following goals within the context of online user image behaviour, to: 1) understand the general contours of human motivations to interact with images; 2) Addressing all incomplete, noisy, missing tags and; 3) evaluate how users categorize their tags for working with the image. These study ultimate purposes to chart a new framework for understanding missing and noisy tags.

We use matrix completion [5], in that we represent the relation between tags and images by a tag matrix, row represents image and each column represents to a tag. Each entry in the tag matrix is a real number that represents the relevance of a tag to an image. Similarly, we represent the partially and noisy tagged images by an observed tag matrix, where an entry (i,j) is marked as 1 if and only if image i is annotated by keyword/ tag j. On other hand the tag information, we also compute the visual similarity between images based on the extracted visual features. We search optimal tag matrix that is consistent with the observed tag matrix and the visual similarity between images. Our proposed algorithm is effective in comparison to the state-of-the-art algorithms for automatic image annotation.

The rest of the paper is organized as follows. Section 2 reviews the related work on automatic image annotation. In Section 3, we describes problem of tag completion, algorithm. Section 4 describes automatic image annotation and tag-based search. 5. Finally, we conclude the paper with future work discussion.

II. REALATED WORK

As tag completion is to add tags with higher relevance to a given image, it would be natural to compare it with image auto-annotation and tag recommendation. Image auto-annotation [4–10] is to automatically and objectively associate unlabeled images with semantically related tags. Feng et al. [4] proposed a generative learning approach for auto-annotation based on multiple Bernoulli relevance model. Liu et al. [6] built multiple graphical models with various correlations between images and tags, and then performed auto-annotation with manifold learning processes.

In recent work shows, the major difference between our work and aforementioned efforts is that a textual document contains much redundancy of words to conveys its semantic whereas images are usually associated with only few tags. Furthermore, redundancy of tags is minimal in many social image tagging systems. Particularly, in Flickr, a tag cannot be assigned more than once to the same image. Moreover, the tags are assigned by different users with different motivations and different criteria for determining the degree of relatedness of a tag to an image. All these differences demand systematic investigation of the impact of different formulations on image search ranking.

More recent work shows improvement of automatic image annotation in considering visual features. Other than guessing annotated tags for the image, many algorithms [5] have been developed to predict annotations for individual areas within an image. So from this, performance of automatic image annotation is far from being satisfactory. Many algorithms have been proposed for automatic image annotation. They can roughly be grouped into two major categories, depending on the type of image representations used. The first group of approaches is based upon global image features, such as color moment, texture histogram, etc. The second group of approaches adopts the local visual features.

Several recent works on image annotation are based on distance metric learning. Monay and Gatica-Perez proposed annotating the image in a latent semantic space. Zhuang and Hoi proposed a two-view learning algorithm for tag reranking. Li et al. proposed a neighbour voting method for social tagging. Similarly to the classification-based approaches, these methods require clean and complete image tags, making them unsuitable for the tag completion problem.

Finally, our work is closely related to image tag refinement scheme. In this work, the authors formulated image tag refinement as an optimization framework based on the consistency between visual similarity and semantic similarity in social images. Unlike the proposed work that tries to complete the missing tags and correct the noisy tags, tag refinement is only designed to remove noisy tags that do not reflect the visual content of images.

III. TAG COMPLETION

We first present a framework for tag completion, and then describe an efficient algorithm for solving the optimization problem related to the proposed framework.

A. A Framework for Tag Completion

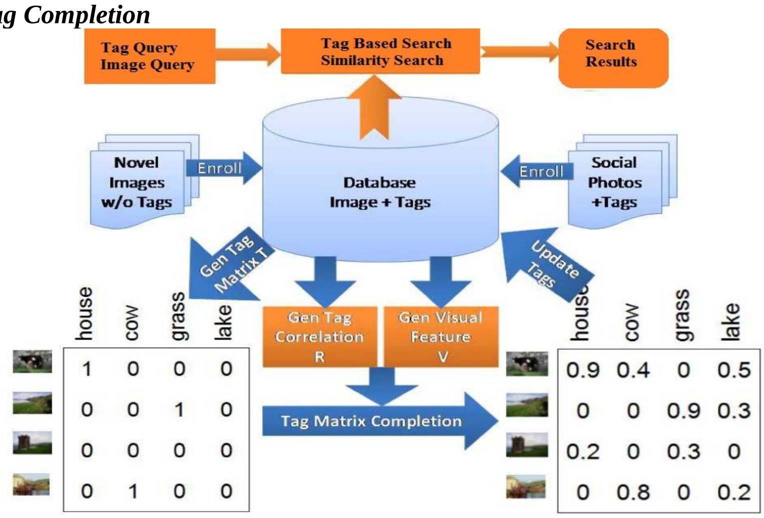

Fig. 1.Tag matrix completion for image search. Assigned tags for the given database, our algorithm creates a tag matrix states relation between images an d initially assigned tags. It completes tag matrix automatically with the help of updating significant score of tags to all images. The Completed tag matrix will be used for TBIR.

Fig. 1 describes the tag completion task. Given a binary image-tag matrix, our aim is to complete the tag matrix automatically with real numbers that shows the probability of assigning the tags to the images. Given the completed tag matrix, we can run TBIR to efficiently and accurately detect the relevant images for textual query.

B. Optimization

Prior to finding optimization problem, we develop a subgradient descent-based approach (Algorithm 1). As comparing to the other optimization approaches such as Newton's method and interior point methods, the subgradient descent approach is advantageous in that its computational complexity per iteration is significantly lower, making it suitable for large image datasets.

Algorithm 1. Tag Completion Algorithm (TMC)

1: INPUT:

- Observed tag matrix: $\hat{T}$ ε IRn x m
- Parameters: γ, η, λ, and μ
- Convergence threshold: ε

2: OUTPUT: the complete tag matrix T
3: Compute the tag correlation matrix R = $\hat{T}T\hat{T}$
4: Initialize w1 = 1d, T1 = $\hat{T}$, and t = 0
5: repeat
6: Set t = t + 1 and stepsize ηt = 1/t
7: Compute $\hat{T}$t+1 and ŵ t+1 according to (8)
8: Update the solutions Tt+1 and ŵ t+1 according to (9) and (10)
9: until convergence: ||L(Tt, wt)- L(Tt+1, wt+1)||≤ ε|| L(Tt, wt)||

C. Discussion

The objective of this work is to complete the tag matrix for all the images, it belongs to the category of transductive learning. In order to turn a transductive learning method into an inductive one, one common approach is to retrain a prediction model based on outputs from the transduction method [2]. A similar approach can be used for the proposed approach to make predictions for out-of-samples.

IV. EXPERIMENTS

Automatic image annotation and tag-based image retrieval are the two tasks for tag matrix. Three benchmark datasets are used in this study.

- Corel dataset [7]. It consists of 4,993 images, each image has 5 tags and 260 unique keywords used in this dataset.
- Labelme photo collection. It consists of 2,900 online photos, annotated by 495 nonabstract noun tags. The maximum number of annotated tags per image is 48.
- Flickr photo collection. It consists of one million images that are annotated by more than 10,000 tags. The maximum number of annotated tags per image is 76. Since most of the tags are only used by a small number of images, we reduce the vocabulary to the first 1,000 most popular tags used in this dataset, which reduces the database to 897,500 images.

A. Automatic Image Annotation

After evaluating the proposed algorithm for tag completion by automatic image annotation, We randomly separate each dataset into two group collections. One group consisting of 80 percent of images is used as training data, and the other collection consisting of 20 percent of images is used as testing data. We repeat this 20 times. Each run creates a new separation of the collections. We report the result based on the average over the 20 trials. To run the proposed algorithm for automatic image annotation, we simply view test images as special cases of partially tagged images, i.e., no tag is observed for test images. We apply algorithm for completion matrix for training and test images. We then rank the tags for test images in the descending orderbased on their relevance scores in the completed tag matrix, and return the top ranked tags as the annotations for the test images.

We compare the proposed tag matrix completion (TMC) algorithm to the following six state-of-the-art algorithms for automatic image annotation:

1. Multiple bernoulli relevance models (MBRMs) [6] that models the joint distribution of annotation tags and visual features by a mixture distribution.
2. Joint equal contribution method (JEC) [4] that finds appropriate annotation words for a test image by a k nearest neighbor classifier that combines multiple distance measures derived from different visual features.
3. Inference network method that applies the Bayesian network to model the relationship between visualfeatures and annotation words.
4. Large scale max-marin multilabel classification (LM3L) [8] that overcomes the training bias by incorporating correlation prior.
5. Tag Propagation method (TagProp) that propagates the label information from the labeled instances to the unlabeled instances via a weighted nearest neighbor graph.
6. Social tag relevance by neighbor voting (TagRel) that explores the tag relevance based on a neighbourhood voting approach.

B. Tag-Based Image Retrieval

Unlike the experiments for image annotation where each dataset is divided into a training set and a testing set, for the experiment of tag-based image retrieval, we include all the images from the dataset except the queries as the gallery images for retrieval. Similarly to the previous experiments, we vary the number of observed tags from 1 to 4. Similarly to the previous experiments, we only compare the proposed algorithm to TagProp and TagRel because the other approaches were unable to handle the partially tagged images. Below, we first present the results for queries with single tag, and then the results for queries consisting of multiple tags.

IV.B.1 *Results for Single-Tag Queries*

In this experiment, we restrict ourselves to the queries that consist of a single tag. Since every tag can be used as a query, we have in total 260 queries for the Corel5k dataset, 495 queries for Labelme dataset, and 1,000 queries for the

Fig. 2. Convergence of the proposed tag matrix completion method on the flickr dataset

Flickr and Tiny Image datasets. Besides the TagProp and TagRel methods, we also introduce a reference method that returns a gallery image if its observed tags include the query word. By comparing to the reference method, we will be able to determine the improvement made by the proposed matrix completion method.

IV.B.2 *Experimental Results for Queries with Multiple Tags*

Similarly to the previous experiment, we vary the number of observed tags from 1 to 4 for this experiment. To generate queries with multiple tags, we randomly select 200 images from the Flickr dataset, and use the annotated tags of the randomly selected images as the queries. To determine the relevance of returned images, we manually check for every query the first 20 retrieved images by the proposed method and by the TagProp method. For all the methods in comparison, we follow the method presented in Section 3.3 for calculating the tag-based similarity between the textual query and the completed tags of gallery images.

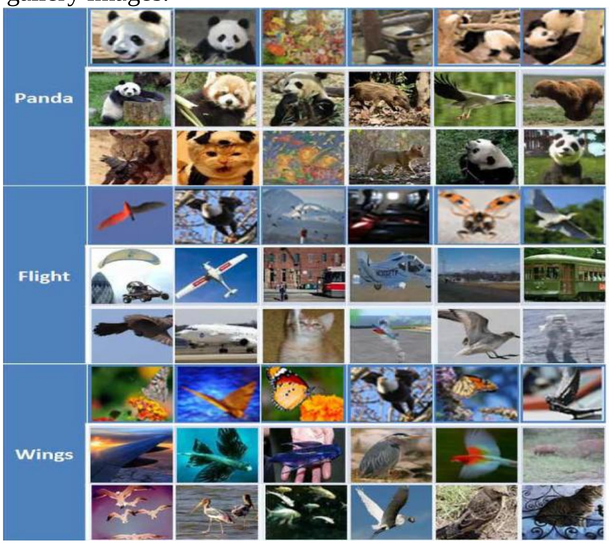

Fig. 2. Illustration of the single-tag-based image search. The word on the left is the query and images on its right are the search results. The images displayed in the three rows are the results returned by the proposed TMC method, the TagProp method, and the TagRel method, respectively. The blue outlines are the results for the proposed methods, the white lines are the results for the baseline methods.

For the TagProp method, we fill in the tag matrix T by applying the label propagation method in TagProp before computing the tag-based similarity. MAP scores for the first 5, 10, 15, and 20 images that are returned for each query. For complete comparison, we include two additional baseline methods.

- The reference method that computes the similarity between a gallery image and a query based on the occurrence of query tags in the observed annotation of the gallery image, and rank images in the descending order of their similarities.

The content-based image retrieval method that represents images by a bag-of-words model using a vocabulary of 10,000 visual words generated by the k-means clustering algorithm, and computes the similarity between a query image and a gallery image using the well-known TF/IDF weighting in text retrieval.

Fig. 3. Illustration of the similarity retrieval result based on multiple tag querying. Each image on the left is the query image whose annotated tags are used as the query word. The images to its right are, from top to bottom, the results returned by the proposed TMC method, the TagProp method, and the TagRel method, respectively.

V. CONCLUSION

This work proposed a tag matrix completion method for image tagging and image retrieval. We optimize the relation of tag matrix by minimizing the difference between tag-based similarity and visual content-based similarity. The proposed method falls into the category of semi-supervised learning in that both tagged images and untagged images are exploited to find the optimal tag matrix. We create two sets for tag completion in this work, i.e., automatic image annotation and tag-based image retrieval. Unreliable and inconsistent manual tags are efficiently overcome with the help of these techniques. Operational and experimental results on three open benchmark datasets show that the proposed method significantly outperforms several state-of-the-art methods for automatic image annotation.

ACKNOWLEDGMENT

In this context, I would like to thank respected and helpful guide Prof. Shamkuwar D.O. and Prof. Chandre P.R. of SPCOE, Otur institute who had been an incessant source of inspiration and help.

REFERENCES

[1] A.B. Goldberg, X. Zhu, B. Recht, J. Xu, and R.D. Nowak, "Transduction with Matrix Completion: Three Birds with One Stone," *Proc. Neural Information Processing Systems Foundation Conf.*, pp. 757-765, 2010.

[2] A. Gammerman, V. Vovk, and V. Vapnik, "Learning by Transduction," *Proc. Conf. Uncertainty in Artificial Intelligence,* pp. 148-155, 1998.

[3] A. Ioffe, "Composite Optimization: Second Order Conditions, Value Functions and Sensityvity," *Analysis and Optimization of Systems*, vol. 144, pp. 442-451, 1990.

[4] A. Makadia, V. Pavlovic, and S. Kumar, "A New Baseline for Image Annotation," *Proc. 10th European Conf. Computer Vision,* pp. 316-329, 2008.

[5] A. Rakhlin, "Applications of Empirical Processes in Learning Theory: Algorithmic Stability and Generalization Bounds," PhD thesis, MIT, 2006.

[6] S.L. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli Relevance Models for Image and Video Annotation," Proc. *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1002-1009, 2004.

[7] P. Duygulu, K. Barnard, J.F.G. de Freitas, and D.A. Forsyth, "Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary," *Proc. Seventh European Conf. Computer Vision*, pp. 97-112, 2002.

[8] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation," *Proc. 12th IEEE Int'l Conf. Computer Vision*, pp. 309-316, 2009.

[9] L. Wu, L. Yang, N. Yu, and X.-S. Hua, "Learning to Tag," *Proc. 18th Int'l Conf. World Wide Web*, pp. 361-361, 2009.

[10] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S.C.H. Hoi, and M. Satyanarayanan, "A Boosting Framework for Visuality-Preserving Distance Metric Learning and Its Application to Medical Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 30-44, Jan. 2010.

