ISSN: 2321-7782 (Online)
Volume 3, Issue 11, November 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Improvised Tag Ranker for Tag Based Image Retrieval (TBIR)

Priyanka Jadhav
Computer Department
Jayawantrao Sawant College of Engineering
Pune, India

Abstract: Many social image search engines are based on keyword/tag matching. This 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. Recent studies have shown that manual tags are often unreliable and inconsistent. Most studies cast image annotation into a multi-label classification problem. The main shortcoming of this approach is that it requires a large number of training images with clean and complete annotations in order to learn a reliable model for tag prediction. We address this limitation by developing a novel approach that combines the strength of tag ranking with the power of matrix recovery. Instead of having to make a binary decision for each tag, our approach ranks tags in the descending order of their relevance to the given image, significantly simplifying the problem. In addition, the proposed method aggregates the prediction models for different tags into a matrix, and casts tag ranking into a matrix recovery problem. It introduces the matrix trace norm to explicitly control the model complexity so that a reliable prediction model can be learned for tag ranking even when the tag space is large and the number of training images is limited. Experiments on multiple well-known image datasets demonstrate the effectiveness of the proposed framework for tag ranking compared to the state-of-the-art approaches for image annotation and tag ranking.

I. INTRODUCTION

How to accurately retrieve images from enormous collections of digital photos has become an important research topic. Content-based image retrieval (CBIR) addresses this challenge by identifying the matched images based on their visual similarity to a query image. However due to the semantic gap between the low-level visual features used to represent images and the high-level semantic tags used to describe image content, limited performance is achieved by CBIR techniques. To address the limitation of CBIR, many algorithms have been developed for tag based image retrieval (TBIR) that represents images by manually assigned keywords/tags. It allows a user to present his/her information needs by textual information and find the relevant images based on the match between the textual query and the assigned image tags. Recent studies have shown that TBIR is usually more effective than CBIR in identifying the relevant images since it is time-consuming to manually label images, various algorithms have been developed for automatic image annotation in this work, we focus on the tag ranking approach for automatic image annotation. Instead of having to decide, for each tag, if it should be assigned to a given image, the tag ranking approach ranks tags in the descending order of their relevance to the given image. By avoiding making binary decision for each tag, the tag ranking approach significantly simplifies the problem, leading to a better performance than the traditional classification based approaches for image annotation. In addition, studies have shown that tag ranking approaches are more robust to noisy and missing tags than the classification approaches.

Rest of the paper is organized as follows. Section 2reviews the related work on automatic image annotation and tag ranking. In Section 3, we introduce the formulation details of the proposed framework and describe an efficient algorithm for computing the optimal solution. Experimental results on five different image data sets are reported and analyzed inspection 4. Finally, Section 5 concludes this work with future directions. Although multiple algorithms have been developed for tag ranking; they tend to perform poorly when the number of training images is limited compared to the number of tags, a scenario often encountered in real world applications. In this work, we address this limitation by casting tag ranking into a matrix recovery problem. The key idea is to aggregate the prediction models for different tags into a matrix. Instead of learning each prediction model independently, we propose to learn all the prediction models simultaneously by exploring the theory of matrix recovery, where trace norm regularization is introduced to capture the dependence among different tags and to control the model complexity. We show, both theoretically and empirically, that with the introduction of trace norm regularize, a reliable prediction model can be learned for tag ranking even when the tag space is large and the number of training images is small.

II. Literature Survey

In this section we review the related work on automatic image annotation and tag ranking.

a) Automatic Image Annotation

Automatic image annotation aims to find a subset of keywords/tags that describes the visual content of an image. It plays an important role in bridging the semantic gap between low-level features and high-level semantic content of images. Most automatic image annotation algorithms can be classified into three categories (i) generative models that model the joint distribution between tags and visual features, (ii) discriminative models that view image annotation as a classification problem, and (iii) search based approaches. Below, we will briefly review approaches in each category. Both mixture models and topic models, two well-known approaches in generative model, have been successfully applied to automatic image annotation. In, a Gaussian mixture model is used to model the dependence between keywords and visual features.

Since a large number of training examples are needed for estimating the joint probability distribution over both features and keywords, the generative models are unable to handle the challenge of large tag space with limited number of training images. Discriminative models, views image annotation as a multi-class classification problem, and learn one binary classification model for either one or multiple tags. A structured max-margin algorithm is developed in to exploit the dependence among tags. One problem with discriminative approaches for image annotation is imbalanced data distribution because each binary classifier is designed to distinguish image of one class from images of the other classes. It becomes more severe when the number of classes/tags is large

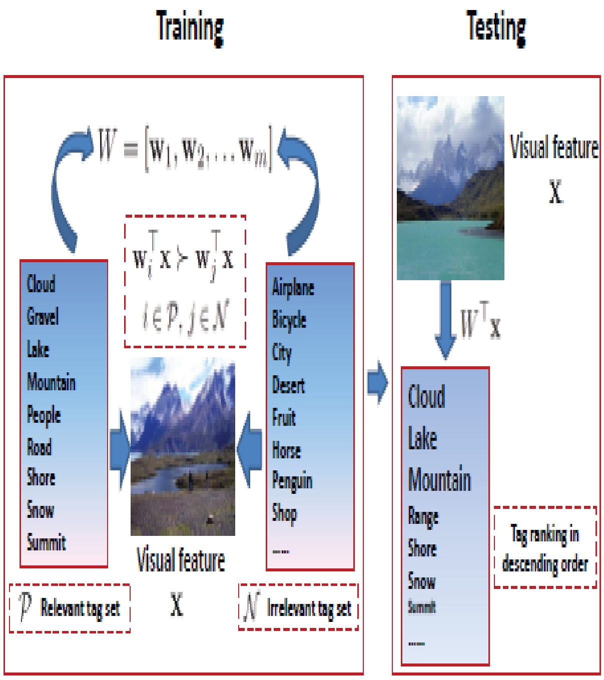
b) Tag Ranking

Tag ranking aims to learn a ranking function that puts relevant tags in front of the irrelevant ones. In the simplest form, it learns a scoring function that assigns larger values to the relevant tags than to those irrelevant ones. In the Authors develop a classification framework for tag ranking that computes tag scores for a test image based on the neighbor voting. It was extended in to the case where each image is represented by multiple sets of visual features. Liu et al. utilizes the Kernel Density Estimation (KDE) to calculate relevance scores for different tags, and performs a random walk to further improve the performance of tag ranking by exploring the correlation between tags. Similarly, Tang et al. proposed a two-stage graph-based relevance propagation approach. In a two-view tag weighting method is proposed to effectively exploit both the correlation among tags and the dependence between visual features and tags. In a max-margin riffled independence model is developed for tag ranking. As mentioned in the introduction section, most of the existing algorithms for tag ranking tend to perform poorly when the tag space is large and the number of training images is limited.

III. Implementation

We first describe our experimental setup, including image datasets, feature extraction, and evaluation measures. We then present three sets of experiments to verify the effectiveness of the proposed tag ranking approach, where

The first experiment evaluates the performance of image annotation with limited training examples.



The proposed method aggregates the prediction models for different tags into a matrix, and casts tag ranking into a matrix recovery problem. It introduces the matrix trace norm to explicitly control the model complexity so that a reliable prediction model can be learned for tag ranking even when the tag space is large and the number of training images is limited. Experiments on multiple well-known image datasets demonstrate the effectiveness of the proposed framework for tag ranking compared to the state-of the- art approaches for image annotation and tag ranking.

Automatic image Annotation.

Automatic image annotation aims to find a subset of keywords/ tags that describes the visual content of an image. It plays an important role in bridging the semantic gap between low-level features and high-level semantic content of images. Most automatic image annotation algorithms can be classified into three categories generative models that model the joint distribution between tags and visual features, discriminative models that view image annotation as a classification problem, and search based approaches. Below, we will briefly review approaches in each category. Both mixture models and topic models, two well-known approaches in generative model, have been successfully applied to automatic image annotation. In a Gaussian mixture model is used to model the dependence between keywords and visual features. In kernel density estimation is applied to model the distribution of visual features and to estimate the conditional probability of keyword assignments given the visual features. Topic models annotate images as samples from a specific mixture of topics, which each topic is a joint distribution between image features and annotation keywords. Various topic models have been developed for image annotation, including probabilistic latent semantic analysis (pLSA), latent Dirichlet allocation and hierarchical Dirichlet processes. Since a large number of training examples are needed for estimating the joint probability distribution over both features and keywords, the generative models are unable to handle the challenge of large tag space with limited number of training images

Discriminative models, views image annotation as a multi-class classification problem, and learn one binary classification model for either one or multiple tags. A 2D multiresolution hidden Markov model (MHMM) is proposed to model the relationship between tags and visual content .A structured max-margin algorithm is developed in to exploit the dependence among tags. One problem with discriminative approaches for image annotation is imbalanced data distribution because each binary classifier is designed to distinguish image of one class from images of the other classes. It becomes more severe when the number of classes/tags is large .Another limitation of these approaches is that they are unable to capture the correlation among classes, which is known to be important in multi-label learning. To overcome

Tag ranking.

Tag ranking aims to learn a ranking function that puts relevant tags in front of the irrelevant ones. In the simplest form, it learns a scoring function that assigns larger values to the relevant tags than to those irrelevant ones. In, the authors develop a classification framework for tag ranking that computes tag scores for a test image based on the neighbor voting. It was extended in to the case where each image is represented by multiple sets of visual features. Liu et al. Utilizes the Kernel Density Estimation (KDE) to calculate relevance scores for different tags, and performs a random walk to further improve the performance of tag ranking by exploring the correlation between tags. Similarly, Tang et al. Proposed a two-stage graph-based relevance propagation approach. In, a two-view tag weighting method is proposed to effectively exploit both the correlation among tags and the dependence between visual features and tags. In, a max-margin riffled independence model is developed for tag ranking. As mentioned in the introduction section, most of the existing algorithms for tag ranking tend to perform poorly when the tag space is large and the number of training images is limited.

Low-rank.

In mathematics, low-rank approximation is a minimization problem, in which the cost function measures the fit between a given matrix (the data) and an approximating matrix (the optimization variable), subject to a constraint that the approximating matrix has reduced rank. The problem is used for mathematical modeling and data compression. The rank constraint is related to a constraint on the complexity of a model that fits the data. In applications, often there are other constraints on the approximating matrix apart from the rank constraint, e.g., non-negativity and Henkel.

We study the rank, trace-norm and max-norm as complexity measures of matrices, focusing on the problem of fitting a matrix with matrices having low complexity. We present generalization error bounds for predicting unobserved entries that are based on these measures. We also consider the possible relations between these measures. We show gaps between them, and bounds on the extent of such gaps.

Matrix recovery.

A common modeling assumption in many engineering applications is that the underlying data lies (approximately) on a low-dimensional linear subspace. This property has been widely exploited by classical Principal Component Analysis (PCA) to achieve dimensionality reduction. However, real-life data is often corrupted with large errors or can even be incomplete. Although classical PCA is effective against the presence of small Gaussian noise in the data, it is highly sensitive to even sparse errors of very high magnitude. We propose powerful tools that exactly and efficiently correct large errors in such structured data. The basic idea is to formulate the problem as a matrix rank minimization problem and solve it efficiently by nuclear-norm minimization. Our algorithms achieve state-of-the-art performance in low-rank matrix recovery with theoretical guarantees. Please browse the links to the left for more information. The introduction section provides a brief overview of the low-rank matrix recovery problem and introduces state-of-the-art algorithms to solve. Please refer to our papers in the references section for complete technical details, and to the sample code section for MATLAB packages. The applications section showcases engineering problems where our techniques have been used to achieve state-of-the-art performance.

Trace norm.

Trace-norm and max-norm as complexity measures of matrices, focusing on the problem of fitting a matrix with matrices having low complexity. We present generalization error bounds for predicting unobserved entries that are based on these measures. We also consider the possible relations between these measures.

Accelerated Gradient Algorithm

Gradient descent is a first-order optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or of the approximate gradient) of the function at the current point. If instead one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent.

Gradient descent is also known as steepest descent, or the method of steepest descent. When known as the latter, gradient descent should not be confused with the method of steepest descent for approximating integrals.

a) ***Image Datasets***

To evaluate the proposed algorithm for image tagging, we conduct extensive experiments on five benchmark datasets for image annotation/tagging, including Corel5K, ESPGame, IAPRTC-12, Pascal VOC2007 and SUN Attribute. The first three image datasets are used to evaluate the performance of automatic image annotation, and the last two image datasets are used to evaluate tag ranking since a relevance scores provide for every assigned tag. Table I summarizes the statistics of the image datasets used in our study.

Corel5K. This dataset contains about 5, 000 images that are manually annotated with 1 to 5 keywords. The annotation vocabulary contains 260 keywords. A fixed set of 499 images are used as test and the rest images are used for training.

ESPGame. This dataset is obtained from an online game named ESP. We use a subset of around 20,000 images that are publicly available.

IAPRTC-12. This image collection is comprised of 19, 627images, each accompanied with descriptions in multiple languages that were initially published for cross-lingual retrieval. Nouns are extracted from the textual descriptions to form the keyword assignments to images. We use the annotation results provided in.

Pascal VOC2007. This dataset is comprised of 9, 963images. We use the tags provided in that are collected From 758 workers using Amazon Mechanical Turk. As a result, for each image, we compute the relevance score for each assigned tag based on its votes from different workers. This relevance score will be used to evaluate ranking performance. On average, each image in this dataset is annotated by 4.2 tags from a vocabulary of 399 tags.

SUN Attribute. The SUN Attribute dataset contains 14,340images and 102 scenes attributes spanning from materials, surface properties, lighting, functions and affordances, to spatial envelope properties. Similar to Pascal VOC2007, the annotated tags are collected from a large number of workers using the Amazon Mechanical Turk and therefore the votes from different workers can be used to compute the relevance score for different tags.

b) ***Automatic Image Annotation with Limited Number of Training Images***

In the first experiment, we evaluate the annotation performance of the proposed image tagging method with limited training images. To this end, we randomly sample only 10% of images for training and use the remaining 90% for testing. Each experiment is repeated 10 times, each with a different splitting of training and testing data. We report the result based on the average over the trials. The following state-of-the art approaches for image annotation are used as the baseline Approaches in our evaluation:

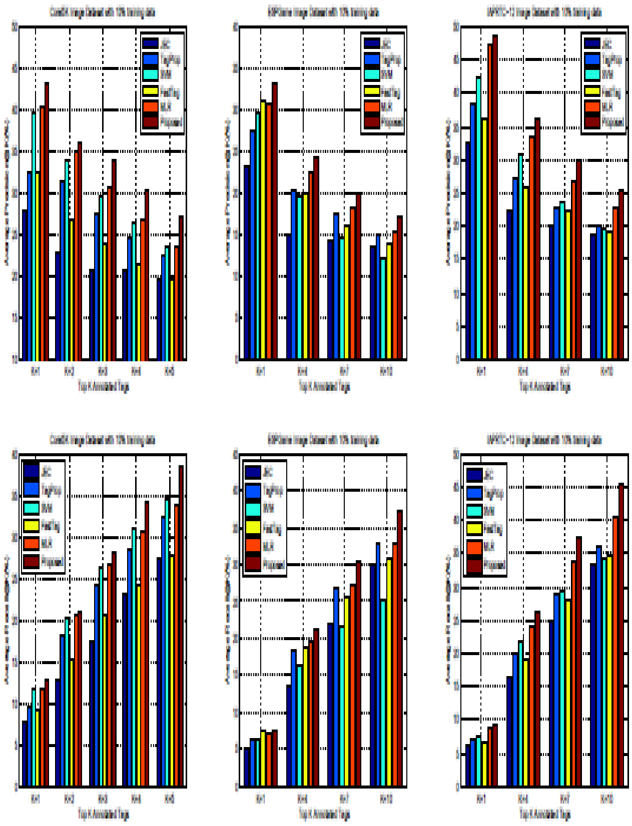
Joint equal contribution method (JEC). It finds appropriate annotation words for a test image based on a *k* nearest neighbor classifier that used a combined distance measure derived from multiple sets of visual features.

Tag Propagation method (Tag Prop). It propagates the tag information from the labeled images to the unlabeled ones via a weighted nearest neighbor graph, where RBF kernel function is used for computing weights between images.

Multi-Class SVM method (SVM). It simply implements One-versus-All (OvA) SVM classifier for each tag, and ranks the tags based on the output probability values.

Fast Image Tagging method (Fast Tag). It explores multi-view learning technique for multi-label learning. In particular, it defines two classifiers, one for each view of the data, and introduces a co-regularize in the objective Function to enforce that the predictions based on different views are consistent for most training examples.

Efficient Multi-Label Ranking method (MLR). This approach explores the group lasso technique in multi-label ranking to effectively handle the missing class the key parameter for Tag Prop is the number of nearest neighbors used to determine the nearest neighbor graph. We set it to be 200 as suggested by the original work. For both SVM and MLR methods, linear function instead of RBF kernel function is adopted here for fair comparison. The optimal value for penalty parameter *C* in both methods is found by cross validation. Note that although Fast Tag method also adopts linear image feature classifiers, it incorporates non-linearity into the feature space as a reprocessing step.



c) Tag Ranking

In this subsection, we evaluate the proposed algorithm for tag ranking. Given an image and a list of associated tags, the goal of tag ranking is to rank the tags according to their relevance to the image content. Both the Pascal VOC2007 and SUNAttribute datasets are used in this experiment since a relevance score is provided for each assigned tag. We randomly select 10% of images from each dataset for training, and use the remaining 90% for testing. We repeat the experiment 10 times and repeat the averaged NDCG.

IV. Conclusion

We have proposed a novel tag ranking scheme for automatic image annotation. The proposed scheme casts the tag ranking problem into a matrix recovery problem and introduces trace norm regularization to control the model complexity. A tag matrix completion method for image tagging and image retrieval. We consider the image-tag relation as a tag matrix, and aim to optimize the 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. Extensive experiments on image annotation and tag ranking have demonstrated that the proposed method significantly outperforms several state-of-the-art methods for image annotation especially when the number of training images is limited and when many of the assigned image tags are missing. In the future, we plan to apply the proposed framework to the image annotation problem when image tags are acquired by crowdsourcing that tend to be noisy and incomplete.

References

1. R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, 2008.
2. L. Wu, R. Jin, and A. Jain, "Tag completion for image retrieval," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 35, no. 3, pp.716–727, 2013.
3. A. Makadia, V. Pavlovic, and S. Kumar, "Baselines for image annotation," International Journal of Computer Vision, vol. 90, no. 1, pp.88–105, 2010.
4. M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation," in IEEE Int. Conf. on Computer Vision, 2009, pp. 309–316.
5. X. Li, C. Snoek, and M. Worring, "Learning social tag relevance by neighbor voting," IEEE Trans. on Multimedia, vol. 20, no. 11, pp. 1254–1259, 2009.
6. J. Zhuang and S. Hoi, "A two-view learning approach for image tag ranking," in ACM Int. Conf. on WSDM, 2011, pp. 625–634.
7. Z. Li, J. Liu, C. Xu, and H. Lu, "Mlrank: multi-correlation learning to rank for image annotation," Pattern Recognition, vol. 46, no. 10, pp. 2700–2710, 2013.

AUTHOR(S) PROFILE

Priyanka Jadhav received the B.E degree in Information Technology and pursuing M.E degree in Computer Engineering from Jayawantrao Sawant College of Engineering.