International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 28-30 April, 2016

A Review of Image Annotation Methods

Pallavi D. Chaudhari[1], Nitin N. Patil[2]
[1]*P.G. Student, Department of Computer Engineering,SES's R. C. Patel Institute of Technology, Shirpur,* *pallavidc14@gmail.com*
[2]*Head & Associate Professor, Department of Computer Engineering,SES's R. C. Patel Institute of Technology, Shirpur,er_nitinpatil@gmail.com*

Abstract-Image annotation has been an active research issue in the recent years due to its potentially large impact on both image understanding and image retrieval. Image annotation can be viewed as multi-label learning problem in most of the studies where each image could contain multiple objects and consequently could be associated with a set of labels. Large amount of research have been done on image retrieval in the past two decades. However, recently research is focusing to bridge the semantic gap between low level image features and high level semantic tags used to express image content. In this paper, we present acomprehensive survey on latest development in automatic image annotation and tag ranking.
Keywords-Automatic image annotation; Tag ranking; matrix recovery; image retrieval; multi label image

I. INTRODUCTION

All Now-a-days, the popularity of digital imaging devices increases the growth of images on the world wide web. Consequently, image retrieval has become challenging research issue. Most traditional techniques of image retrieval uses metadata such as keywords or captioning on the image to retrieve over the annotation words. The goal of Automatic image annotation is to automatically assign keywords or "tags" to the images. It is becomes an important research topic in the image retrieval & management systems. Image annotation can be viewed as multi-label learning problem in which a set of labels are associated with the images containing multiple objects [1]. Since, annotate the images manually is the most time consuming task, which creates yet more challenging to image annotation problem. Several machine learning models have been build up, to overcome the challenging problems of image annotation, that linking to low level features and annotations.
In the model-driven methods, the algorithms can be divided into two main groups: Probability-based methods and Classification-based methods. The Probability-based methods guess the correlations among images and annotation keywords [2]. The Classification-based methods use a class label to characterize annotation keywords and formulate image annotation to a classification problem [3].
Content-based image retrieval scheme need the user to retrieve images based on their visual match to a query image. Although the multiple researches have been performed on image retrieval. However, recent studies have shown that there is a semantic gap between low-level visual feature for representing images in Content-based image retrieval and the high-level semantic tags for describing image content. The semantic gap is linking through the automatic image annotation, that captures semantic features with machine learning techniques [4][5].
Many algorithms have been developed for tag based image retrieval to overcome the limitation of Content-based image retrieval. TBIR uses manually labeled keywords or tags to represent images and let a user to present required information in textual form. The relevant image is searched by similarities between the image tags and the textual query. However recent research shows that, retrieving the relevant images by TBIR is efficient than CBIR [6].
Tag ranking intends a ranking function that place relevant tags in front of the irrelevant ones. It is a scoring function which allocates greater values to the relevant tags than to those irrelevant ones [5][7].

The objective of this paper is to summarize the previous work done on image annotation and tag ranking.

II. RELATED WORK

Many research has been done in the field of image annotation and tag ranking. Many image annotation methods have been proposed in the literature.

In social image retrieval tags quality plays a very important role. Recently the focus of research have been done to deal with tag quality problem. Although tag refinement techniques able to improve tag quality and perform weakly to deal with which tag is more relevant than the others. Recently, Liu et. al. proposed the tag ranking scheme to address the social tag issue. The aim of social tag ranking is to automatically rank the tags associated with social image according to their relevance semantic image content [5].

J. Zhuanget. al. proposed a novel two-view learning approach that address the tag ranking problem. It is data-driven approach without assuming any parametric model on the importance between images and tags. Hence it has better flexibility to fit the varied data [8].

The limitation of multi-label classification problem is that it requires a large number of training images. To overcome this limitation, recently, S. Feng et. al. developed a novel tag ranking scheme for automatic image annotation. This scheme casts tag ranking problem into matrix recovery problem and introduce matrix trace norm regularization to control the complexity of model. This approach ranks tags in their descending order of their relevance to the given image. Experimental results of this tag ranking framework demonstrate the effectiveness as compared to the state-of-the-art approaches for tag ranking and image annotation [7].

Recently, more focus has been given for multi-label learning that reflect on the correlation among categories. Many advances have been developed for multi-label learning which imprison the dependency among classes. C. Wang et.al. proposed framework of multi-label sparse coding for extracting feature and classification within the image annotation. In this a label sparse coding algorithm is developed to efficiently control multi-label information for dimensionality reduction. This method is proposed to propagate the multi-labels of the training images to the query image with the sparse reconstruction coefficients. The experimental result shows that the superior performance of the proposed multi-label sparse coding framework over the state-of-the-art algorithms [1].

X. Cai et.al. developed new graph Structured Sparsity model for multi-label image annotation to include the topological constraints of relation graph. In this the label correlation is model using the relational graph in multi-label classifications. This method capture and utilize the hidden class structures in relational graph for improving the annotation results. For solving the problem they derived an efficient optimization algorithm. Empirical result of this method shows better annotation results than the state-of-the-art methods [9].

O. Yakhnenkoet. al. proposed a Hierarchical Dirichlet process model which is a nonparametric Bayesian model. This model solves the problem of training a probabilistic model from a dataset of images and their associated captions to predict the image annotation problem as well as the image object-label correspondence problem. It is important to find such correlation between images and text for many problems in domain where labeled data is not readily available. The comparison of MoM-HDA model and MoM-LDA model results that MoM-HDA perform just as well as or better than the MoM-LDA model on both the image object-label correspondence task and image annotation task [10].

J. Jeonet. al. proposed an automatic approach to annotating and retrieving images based on training set of images. In this they assume that regions in an image can be explained using a small glossary of blobs which are created from image features using clustering. Experiment results shows that the annotation performance of proposed cross-media relevance model is almost six times as good than a word-blob co-occurrence model and twice as good as state of the art model. This approach shows the usefulness of using formal information retrieval models for the image annotation and retrieval [2].

Although, Joint Equal Contribution model of Makadiaet. al. [11] and the TagProp model of Guillauminet. al. rely on local nearby neighborhoods and work unpredictably well even though their ease. The TagProp is the current state-of-the-art method for image annotation and keyword based image retrieval. In this the combination of metric learning permissible by directly exploiting the log-likelihood of the tag predictions in the training set. They also demonstrated a word specific sigmoidal modulation of the weighted neighbor tag prediction that increase the evoke of rare words [12].
Since, due to the limitation of manual annotation, images are not properly annotated with all the relevant labels. Y. Vermaet. al. proposed 2PKNN, a variant of the classical K-nearest neighbour algorithm that combined with metric learning and gives efficient performance than the previous methods. It deals with the "weak-labeling" and "class-imbalance" issues. They showed that a classification metric learning algorithm can be efficiently modified for the more complex multi-label classification problems such as annotation. The fig. 1 shows annotations for example images from the two datasets by this method [13]. The second row shows the ground-truth annotations and the third row shows the labels predicted using method 2PKNN. The labels in blue (bold) are those that match with ground truth. The labels in red (italics) are those that, though depicted in the corresponding images, are missing in their ground-truth annotations and are predicted by this method.

Corel 5K		ESP Game	
			
bear, reection, water, black	field, horses, mare, foals	green, phone, woman, hair	fight, grass, game
bear, **reec-tion**, **water**, **black**, *river*	**field**, **horses**, **mare**, **foals**, *tree*	**green**, **phone**, **woman**, **hair**, *suit*	**fight**, **grass**, **game**, *anime*, *man*

Figure. 1Annotations for Example Images from the Two Datasets.

Many kernel distance metric learning algorithms have been developed so as to capture the nonlinear relationships between semantics and visual features of image. Z. Feng proposed a robust kernel metric learning (RKML) algorithm that evidently deals with the problems arising from high dimensionality and limitations of binary constraints in tags. This algorithm directly utilizes the real valued similarity measure, based on image tags for learning a distance metric. Experimental result of proposed algorithm shows that efficiency and effectiveness of algorithm by comparing it to the state-of-the-art approaches for image annotation [14].

III. CLASSIFICATION OF AUTOMATIC IMAGE ANNOTATION METHOD

The purpose of automatic image annotation is to find a subset of keywords or tags that reflect the visual content of an image. Most automatic image annotation algorithms can be classified into three type as generative models, discriminative model and search based approaches. In this section, we will briefly review the approaches under each category.

3.1. Generative Model

It model the joint distribution between tags and visual features of image. Topic models and Mixture models are two well known approaches in generative model that have been effectively applied to automatic image annotation. Images are annotated by topic models as samples from a specific mixture of topics, in which each topic is a joint distribution between annotation keywords and image features. Various topic models have been proposed for image annotation such as probabilistic latent semantic analysis [15], latent Dirichlet allocation [16][17] and hierarchical Dirichlet processes [10].

G. Carneiro et.al. proposed a Gaussian mixture model is used to model the dependency between visual features and keywords [3].

3.2. Discriminative Model

Image annotation views as a multi-class classification problem by discriminative models. A 2D multi-resolution hidded Markov model (MHMM) is proposed by J. Li et. al. to model the relationship between tags and visual content of image [18]. Discriminative approaches causes a problem of imbalanced data distribution as each binary classifier is intended to differentiate image of one class from images of the other classes. The limitation of these approaches is that they are unable to capture the correlation among classes, which is key in multi-label learning. Although, to overcome these issues, several algorithms [19][20] are developed to control the keyword correlation as the additional information.

3.3. Search Based Model

Search based approaches developed for automatic image annotation have been verified to be relatively efficient, mainly for large image datasets including many keywords. In search based approach a test image is annotated with the common tags shared by the subset of training images. The core of search based annotation approaches is to efficiently measure the visual match betweenimages [11][12].

CONCLUSIONS

Although multiple algorithms have been proposed for tag ranking, they be likely to perform poorly when the number of training images is limited as compared to the number of tags. To address this limitation a novel tag ranking scheme is proposed that cast tag ranking to matrix recovery problem and introduces trance norm regulation to control the complexity of model. To overcome the tag ranking problem a tag ranking scheme and novel two-view learning approach are proposed.

To overcome the limitation of manual annotation the 2PKNN algorithm is developed that deals with "weak-labeling" and "class-imbalance" issue. In the future, the focus should be on image annotation topic to deals with image tags which are acquired by crowdsoursing that lean to be noisy and incomplete.

REFERENCES

[1] C. Wang, S. Yan, L. Zhang, and H. J. Zhang, "Multi-label sparse coding for automatic image annotation," in *Proc. IEEE Int. Conf. Comput. Vis.Pattern Recognit.*,, pp. 1643–1650, Jun. 2009.

[2] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," *inProc. 26th Annu.ACM Int. Conf. SIGIR*, pp. 119–126, 2003.

[3] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 394–410, Mar. 2007.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, 2008, Art. ID 5.

[5] J. Wu, H. Shen, Y. Li, Z.-B. Xiao, M.-Y. Lu, and C. L. Wang, "Learning a hybrid similarity measure for image retrieval," *Pattern Recognit.*, vol. 46, no. 11, pp. 2927–2939, 2013.

[6] L. Wu, R. Jin, and A. K. Jain, "Tag completion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 716–727, Mar. 2013.

[7] S. Feng, Z. Feng, and Rong Jin, "Learning to Rank Image Tags with limited training example, "*IEEE Trans. Image Processing,* VOL. 24, NO. 4, APRIL 2015

[8] J. Zhuang and S. C. H. Hoi, "A two-view learning approach for image tag ranking," in *Proc. 4th ACM Int. Conf. WSDM*, pp. 625–634, 2011.

[9] X. Cai, F. Nie, W. Cai, and H. Huang, "New graph structured sparsity model for multi-label image annotations," in *Proc. IEEE Int. Conf.Comput. Vis.*, pp. 801–808, Dec. 2013.

[10] O. Yakhnenko and V. Honavar, "Annotating images and image objects using a hierarchical Dirichlet process model," in *Proc. 9th Int. Workshop* Multimedia Data Mining, pp. 1–7, 2008.

[11] A. Makadia, V. Pavlovic, and S. Kumar, "Baselines for image annotation," *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 88–105, 2010.

[12] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation," in Proc. IEEE 12th Int. Conf. Comput. Vis., pp. 309–316, Sep./Oct. 2009.

[13] Y. Verma and C. V. Jawahar, "Image annotation using metric learning in semantic neighbourhoods," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 836–849, 2012.

[14] Z. Feng, R. Jin, and A. Jain, "Large-scale image annotation by efficient and robust kernel metric learning," in *Proc. IEEE Int. Conf. Comput.Vis.*, pp. 1609–1616, Dec. 2013.

[15] F. Monay and D. Gatica-Perez, "PLSA-based image auto-annotation: Constraining the latent space," in *Proc. 12th Annu. ACM Int. Conf.Multimedia*, pp. 348–351, 2004.

[16] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *J. Mach. Learn. Res.*, vol. 3, pp. 1107–1135, Mar. 2003.

[17] D. Putthividhya, H. T. Attias, and S. S. Nagarajan, "Topic regression multi-modal latent Dirichlet allocation for image annotation," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 3408–3415, Jun 2010.

[18] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 25, no. 9, pp. 1075–1088, Sep. 2003.

[19] H. Wang, H. Huang, and C. Ding, "Image annotation using multi-label correlated Green's function," in Proc. IEEE 12th Int. Conf. Comput. Vis., pp. 2029–2034, Sep./Oct. 2009.

[20] X. Cai, F. Nie, W. Cai, and H. Huang, "New graph structured sparsity model for multi-label image annotations," in *Proc. IEEE Int. Conf.Comput. Vis.*, pp. 801–808, Dec. 2013.