International Conference On Emanations in Modern Technology and Engineering (ICEMTE-2017) ISSN: 2321-8169
Volume: 5 Issue: 3 13 - 17

Dictionary Learning Classification for Multi-Label Image Annotation: A Survey

Yogini B. Bhadane
PG Student
Department of Computer Engineering
R. C. Patel Institute of Technology
Shirpur, India.
yoginibhadane@gmail.com

Nitin N. Patil
Head & Associate Professor
Department of Computer Engineering
R. C. Patel Institute of Technology
Shirpur, India.
er_nitinpatil@rediffmail.com

Abstract **-** Many of the clients have the ability create and store digital images. Peoples do not spend their time for organizing and grouping their special image collections. So it's difficult for normal users to finding particular images. Image annotation contains a number of methods that goals to find the link between words and image features just like color, shape and texture to provide correct annotation words to images. Here, not only presents a (SLED) semantic label embedding dictionary representation which solves the problem of inconsistent label combinations among training data and testing data also solves the difficulty of image annotation under weakly supervised setting. The whole system divided into two parts, i.e. the training data and the testing data. The training data is divided into exclusive groups. In it Fisher discrimination principle used for the train the label. Then co-occurrence labels would provide the context information. This context information adds into the original dictionary label. In testing part, use reconstruction coefficient and label propagation to find the score of each label.

Index Terms - *Image Annotation, Dictionary Learning, Sparse Coding, Reconstruction Coefficient.*

*****

I. INTRODUCTION

Today's humanity and daily life are extremely changed. The major contribution is the internet and digital devices. Digital data such as audio, video and images can be easily stored and transmitted with the help of devices and the internet. The popularity of photo and video allocation websites (Flicker and YouTube, etc.), a large number of sensitively labelled or unlabelled visual data is available. Several numbers of consumers have the ability to create and store digital images. Peoples did not spend their time for organizing and grouping (labelling) their personal image collections. So it's difficult for a common user to find exact images. The search engines regain related to images still mainly based on textual queries instead of raw images. The automatic image annotation, that link images with human-provided keywords or labels have received much research interest.

A. Automatic Image Annotation

Annotation is nothing but estimation note or text. Image annotation is a process in which human provide a key in form of words (text) to the system, then system find out those images, words that express the image? i.e. Sky, Jets, Smoke.

Fig. 1: Simple Image [1].

Image Retrieval is the method in which choose particular images from a large collection of the database using some techniques. The supervised training process using the semantic modules gives better outcomes in image annotation and retrieval process. Given a database of images and a query (e.g. Image), what are the images are described by the words? As shown in Fig. 2, correct image retrieval way and in Fig. 3 incorrect image retrieval way using query.

Query: Bear

Fig. 2: Correct Image Retrieval [1].

Fig. 3: Fault in Image Retrieval [1].

Some image search engines are stood on a keyword or tag equivalent. Label based image retrieval (TBIR) process is one of the image retrieval techniques, which are not only professional but also effective in the image annotation task [2].

An automatic image annotation aim that was allocates images with human predefined labels, which have a typical multi-label learning problem. Image annotation is a complicated job for two main reasons: First are the well famous pixels to predicate or semantic gap difficulty. Second, complexity arises due to lack of communication between the keywords (tags) and image regions in the training data. The supervised knowledge used to generate codebook, in some cases information is gotten wasted. To minimize this information supervised learning methods used. These algorithms can be divided into two schemes, i.e. the parametric and non-parametric schemes [3, 4, 5].

1) *Parametric scheme:* These schemes help to the solve multi-label categorization problems in image annotation. This scheme is based on bag-of-words (BoW) model. The bag-of-words model is an image representation formula for image categorization and annotation tasks. K-means or sparse coding play an imperative role to construct the dictionary. Multi-label sparse coding configuration used for feature extraction and grouping within the situation of automatic image annotation. Multi-label sparse coding contains three mechanisms, i.e. feature demonstration based on probabilistic area modeling, label sparse coding for feature mining and image annotation. Some multi-label visual sorting problems are solved using the label exclusive context with linear demonstration and sorting process [6, 7, 8].

2) *Non-Parametric schemes:* These schemes help to solve the sparse reconstruction problem in image annotation technique. Nonparametric schemes are based on the reconstruction coefficient. The reconstruction constants located on the semantic dictionary [9].

B. Dictionary learning

It is an important component for construction, effective and well-organized demonstration for sparse coding and bag-of-word, etc. The K-SVD and K-means clustering techniques are used to study an over-complete dictionary from image patches. New algorithms used for adapting dictionaries in order to achieve sparse signal representations. The dictionaries learning problems as the smallest amount square's problem and solves it by iterative algorithms to diminish the reconstruction error. Dictionary learning also useful in some cases i.e. reconstruction, sparse coding and classification for related to images. M. Aharon *et al.* solved dictionary size problems. In this case, K-SVD algorithm is an iterative approach which used for image re-establishment and compression process. The dictionary learning process can be classified into two categories, i.e. unsupervised dictionary learning and supervised dictionary learning [9, 10].

1) Unsupervised dictionary learning: In unsupervised dictionary learning category, given only unlabeled input data. This learning category captures higher-level features in that data. Unsupervised dictionary learning has trained a dictionary for the duration of the renovation optimization which minimizes the remaining errors of reconstructing the original signals. The k-means clustering process and the K-SVD algorithm to learn over-complete dictionary from image patches. Unsupervised learning is best for reconstruction process but not for classification [11, 12].

2) Supervised dictionary learning: The supervised dictionary learning developed for discriminative representation by inserting information of labels into the dictionary. The obtainable supervised dictionary learning method can be nearly divided into three categories based on the structure of dictionaries, i.e. learning numerous dictionaries, learn a compact and discriminative dictionary, adding the label information into the objective function. In this learning class only labeled input data is given, which is weakly or strongly labeled data. This learning class is useful for object recognition process [13].

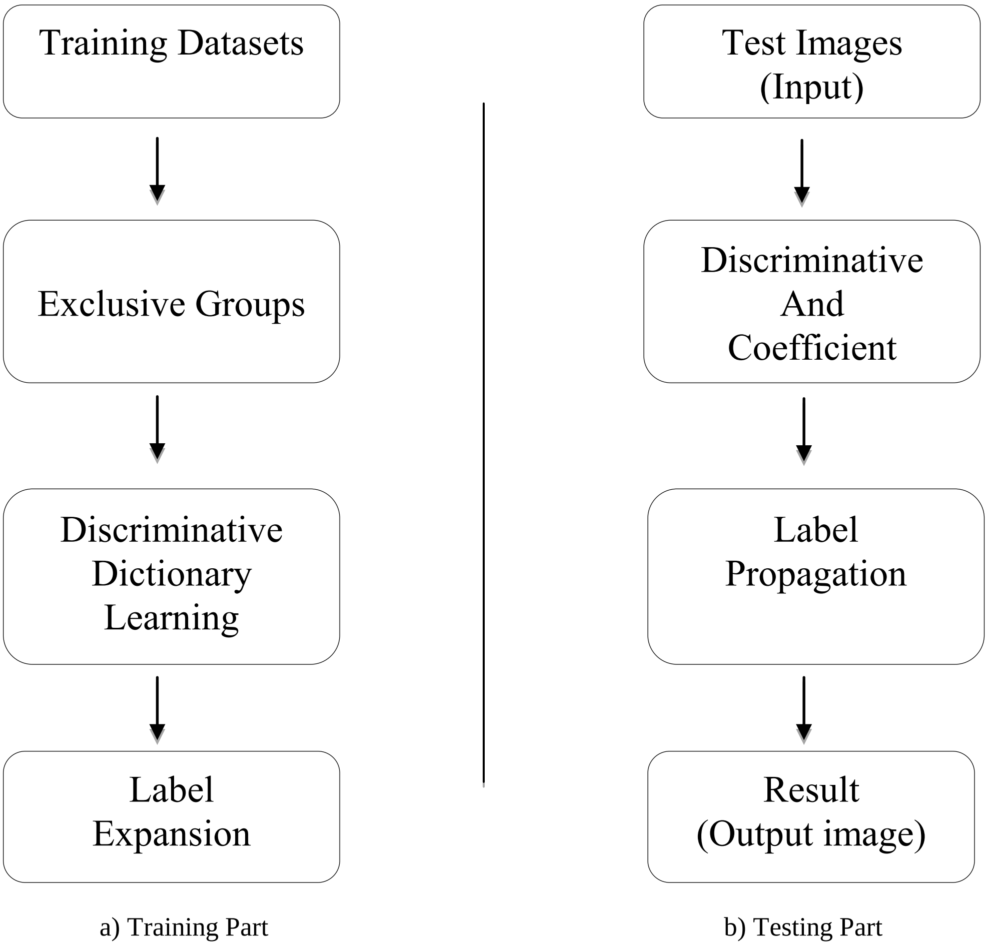

Fig. 4: The Survey of Image Annotation System.

C. Survey on Image Annotation System

Image annotation system divided into two parts, i.e. training part and testing part that is shown in Fig. 4 (a) and (b).

1) Training Part: This part called as offline part. In this part large datasets available. Those datasets related to images, i.e. NUS-WIDE-LITE, Corel5K, etc. Which is goes through many processes as shown in Fig. 4 (a). Here, train the images using a variety of processes. Fisher discrimination dictionary learning used for object categorization in an image also dictionary size problems solved [14].

2) Testing Part: This part also called as online part. In this part, test the images, which are going through many processes as shown in Fig. 4 (b). Here, test the input image using several methods also using some algorithms to find the reconstruction coefficient to increased accuracy rate of images [15].

II. LITERATURE SURVEY

G. Carneiro *et al.* have suggested the supervised knowledge class for image annotation and retrieval task. The actual problems explained for image annotation and retrieval process. The supervised method is shown to accomplish superior accuracy than various in the past available methods. Here, construction of a model of semantic classes which solved the image retrieval problems. Supervised multiclass labelling (SML) process used to solve multiclass image classification problems. There are differences between SML and other algorithms. In this paper, some description about human annotation and SML annotation words using images [1].

Lei Wu *et al.* have demonstrated label completion for image recovery in which image recovery methods can be classified into groups, i.e. content-based image retrieval (CBIR) and keyword or tag-based image retrieval (TBIR) methods. Here, a whole determination about CBIR and TBIR. There is an outline for tag matrix completion and its relevance to image search. In this case, some tag completion problem solved using different methods. Here, we study the problem of tag completion where the ambition is to involuntarily fill in the absent tags as well as accurate noisy tags for particular images [2].

A. Makadia *et al.* have established a new standard method for image annotation that treats image retrieval problems. Here, solve the image annotation problem, i.e. pixel to predicate or semantic gap difficulty, the lack of communication between the keywords and image regions in the training data. Here, automatically assigning keywords to images is of great attention as it allows one to index recovered and recognized large collections of images data. In the process of image annotation, some difficult task occurred in the annotation process that solved by using some baseline methods. In, the parametric model to capture the relationship between image features and keywords. Also some variances between human and actual (predicted) annotation words [3].

V. Lavrenko *et al.* have developed the models for learning semantics of images. The semantics of images which means that allows us to automatically annotate an image with keywords and to recover images based on text queries. These processes have done by formalism that model's generation of annotated images. In this case, image annotation and image retrieval methods explained. Some models which used in the generation of annotated images, i.e. cross-media relevance model (CMRM) and continuous-space relevance model (CRM) this model work on the continuous features with better outcomes [4].

S. Lazebnik and M. Raginsky have developed supervised knowledge to codebooks for information loss minimization case. In this case, some methods which are authenticated on artificial and real datasets also useful to two different problems i.e. learning discriminative visual vocabularies for bag-of-features in image organization and image separation. Here, illustrated methods, which are used to create codebooks for the bag-of-features process in image classification. In this process, original images go through the segmentation process then there is a calculation done for how much average information is a loss or not after this case images are rearranged [5].

Lei Wu *et al.* have introduced a new agenda of semantics-preserving bag-of-words (SPBoW) for object representation. Here, described an actual concept of BoW and SPBoW. The SPBoW model tries to learn a codebook by reducing the semantic gap. The SPBoW method is effective and shows potential for object illustration process. The bag-of-words (BoW) models that usually suffer from the semantic loss in the codebook generation process, some new technique overcomes this drawback by learning an effective distance metric that aims to bridge the semantic gap between low-level features and high-level semantics. They proposed a novel measurement of semantic gap and then try to diminish the gap via distance metric learning [6].

C. Wang *et al.* have described sparse coding structure for feature mining and classification done in the background of automatic image annotation. Also some outline described for multi-label information in feature extraction, data sparse coding for multi-label data to broadcast the multi-labels of the training images to the query image with the thin reconstruction coefficients. Multi-label sparse coding procedure described three components. In the semantic recovery, outcomes showed using some datasets which are related to images. Here, some differences between human annotation words and MSC annotation words are shown using some algorithms [7].

X. Chen *et al.* have demonstrated a new approach to multi-label image classification, which incorporates a new type of context, i.e. label exclusive context which is used for linear representation and classification. To Label exclusive linear representation (LELR) model is used to connect label exclusive context into a multi-label linear representation framework for visual classification. In the challenging real-world visual classification tasks validate, that LELR is a powerful model to improve the performance of linear representation and classification. Determination of LELR smooth minimization algorithm used with a kernel-view. Here, the main aim is to use of multi-label visual classification to solve the problem where the image sample can be assigned with multiple class labels at a time [8].

X. Cao *et al.* have recommended suitable methods to build a dictionary using dictionary learning methods, i.e. supervised and unsupervised learning methods. Here, introduced a new semantic label embedding dictionary (SLED) representation for multi-label image annotation. SLED method in image annotation gives better outcomes as comparing with the state-of-the-art algorithms. SLED method used to solve dictionary size problems. The whole system divided into two parts, i.e. training and testing parts. In this method, used some datasets i.e. NUS-WIDE-LITE, Corel5k and IAPR-TC12 [9].

M. Aharon *et al.* have suggested K-SVD and K-means (clustering process) algorithms for solving dictionary learning problems. These algorithms renew the dictionary atoms to better in shape, the data which means increases the dictionary size using some algorithms, i.e. K-SVD and K-means. There is much application that uses sparse representation, i.e. compression, regularization in inverse problems, feature extraction, etc**.** There is a big description for K-SVD and K-means algorithms, which used to build a dictionary for images [10].

Z. Jiang *et al.* have implemented discriminative dictionary knowledge for sparse coding using labels consistent K-SVD. Dictionary learning process useful in some cases, i.e. reconstruction, sparse coding and classification, etc. To learn their constructive and discriminative dictionary using the label consistent K-SVD algorithm with the supervised information of input signal, i.e. LC-KSVD1 and LC-KSVD2 [11].

Z. Jiang *et al.* have described label consist of K-SVD procedure for learning a discriminative dictionary (sparse coding) which is used for recognition. Discriminative sparse code error technique used to combine the reconstruction error and the classification error. In this process, K-SVD algorithm gives the optimal solution. The K-SVD algorithm is used in case of sparse coding method to build a dictionary. The optimization processes of the objective functions used for increment dictionary learning task [12].

N. Zhou *et al.* have expressed a narrative joint dictionary learning algorithm (JDL) to exploit the visual correlation within a group of visually similar object categories for dictionary learning. This is helpful for supervised dictionary learning by embedding the information of labels into the dictionary. A generally shared dictionary and multiple category specific dictionaries have been learned in JDL model for a group of visually connected object classes. Dictionary learning algorithm solved the object categorization problem [13].

M. Yang *et al.* have illustrated sparse representation process using Fisher discrimination dictionary learning (FDDL) to image annotation. Dictionary plays an important job in sparse representation or sparse coding based on image reconstruction and classification. The used of FDDL in multi-class object categorization. Here, we describe the problem related to dictionary size and the classification process [14].

J. Winn *et al.* have introduced object categorization by the visual dictionary. Here, described new algorithms for automatic recognition of object classes from images. The object categorization used for image retrieval, web search and interactive image editing. Also, high classification accuracy is demonstrated for object classes. A new supervised learning algorithm used for estimating appearance-based models from images. Clustering process used for object categorization. Histogram process shows the accuracy of the images and reduced the dictionary by combining visual words. In this method, how automatically recognition of objects from images with help of some algorithms are Shown [15].

CONCLUSIONS AND FUTURE WORK

In this paper, we have studied multi-label image annotation with dictionary learning methods. Under weakly supervised setting, semantic label embedding dictionary (SLED) solves the image annotation problems. This is new embedding dictionary representation technique. Some optimization algorithms solved the classification problems. Automatic image annotation system divided into two parts, i.e. training and testing parts. This paper gives the solution of inconsistent label combinations between the train and test data. By using three well-known datasets gives good outcomes. In image annotation, solved the huge gap in between the train and test data. Some app estimation gives the accuracy rate of image labels. Here, we state the dictionary size problem and find the reconstruction coefficients of images.

In Future work, the system performance will be increased by selecting the discriminative dictionary items for different labels. Also, the pre-processing method will be used to expand the images.

REFERENCES

[1] 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.

[2] 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.

[3] A. Makadia, V. Pavlovic and S. Kumar, “A new baseline for image annotation,” in *Proc. ECCV*, pp. 316-329, 2008.

[4] V. Lavrenko, R. Manmatha and J. Jeon, “A model for learning the semantics of pictures,” in *Proc. NIPS*, pp. 553-560, 2003.

[5] S. Lazebnik and M. Raginsky, “Supervised learning of quantizer codebooks by information loss minimization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1294–1309, Jul. 2009.

[6] L. Wu, S. C. H. Hoi and N. Yu, "Semantics-preserving bag-of-words models and applications," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1908-1920, Jul. 2010.

[7] C. Wang, S. Yan, L. Zhang and H.-J. Zhang, “Multi-label sparse coding for automatic image annotation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1643-1650, Jun. 2009.

[8] X. Chen, X.-T. Yuan, Q. Chen, S. Yan and T.-S. Chua, “Multi-label visual classification with label exclusive context,” in *Proc. ICCV*, pp. 834-841, Nov. 2011.

[9] X. Cao, H. Zhang, X. Guo, Si Liu and Dan Meng, “SLED: Semantic label embedding dictionary representation for multi-label image annotation,” *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2746-2759, Sep. 2015.

[10] M. Aharon, M. Elad and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311-4322, Nov. 2006.

[11] Z. Jiang, Z. Lin and L. S. Davis, “Learning a discriminative dictionary for sparse coding via label consistent K-SVD,” in *Proc. IEEE Conf. CVPR*, pp. 1697-1704, Jun. 2011.

[12] Z. Jiang, Z. Lin and L. S. Davis, “Label consistent K-SVD: Learning a discriminative dictionary for recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651-2664, Nov. 2013.

[13] N. Zhou, Y. Shen, J. Peng and J. Fan, “Learning inter-related visual dictionary for object recognition,” in *Proc. IEEE Conf. CVPR*, pp. 3490-3497, Jun. 2012.

[14] M. Yang, L. Zhang, X. Feng and D. Zhang, “Sparse representation based Fisher discrimination dictionary learning for image classification,” *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209-232, 2014.

[15] J. Winn, A. Criminisi and T. Minka, “Object categorization by the learned universal visual dictionary,” in *Proc. ICCV*, pp. 1800-1807, Oct. 2005.