IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 1, January 2016.
www.ijiset.com
ISSN 2348 – 7968

A Survey on Novel Dictionary Learning Method for Multi-label Image Annotation

Vaishali S. Garud
Department of Computer Engineering, Matoshri College of Engineering and Research Centre, Nashik, Maharashtra, India

Abstract

The automatic image annotation is relatively new research topic or area for researcher. The task of automatic image annotation is uses to associates the subset of word from given dictionary to images or assign the human provided keyword to the images. Previous methods of weakly supervised multi-label image annotation are relying on the unsupervised feature representation. The components of unsupervised feature representation are not directly correlated with specific labels. In practical cases, there is a big gap between training data (offline part) and the testing data (online part), say the label combination of the testing data and training data is not always consistent. When the features and labels are extracted from the training images, we first construct the label matrix to explore the exclusive label group. After that, we follow the dictionary learning framework to learn the discriminative dictionary representation. Then we measure the relevance between the visual words and their corresponding labels. After that, a multi-task learning framework is used to expand the labels of visual words. Finally the label propagation is obtained, by using the Rank SVM algorithm to computes the score for image annotation. The main aim of this paper is to study the various techniques used for automatic image annotation and dictionary learning for multi-label image annotation.

Keywords: **Exclusive label group, dictionary representation, label expansion, label propagation.**

1. Introduction

1.1 Automatic image annotation

The automatic image annotation is relatively new research topic. Automatic image annotation methods try to answer the growing requirements for processing huge collection of image data available both in internet and large multimedia databases. The task of automatic image annotation is uses to assign subset of word from given dictionary to a previously unseen image on the basis of weakly (incompletely) annotated training set of images without knowledge which part of an image lead to which words. This method can be regarded as a type of multi-class image classification with a very large number of classes as large as the vocabulary size. Typically, image analysis in the form of extracted feature vectors and the training annotation words are used by machine learning techniques to attempt to automatically apply annotations to new images. The first methods learned the correlations between image features and training annotations, then techniques were developed using machine translation to try to translate the textual vocabulary with the 'visual vocabulary'.

1.2 Methods for image annotation

A variety of algorithms have been proposed for this task with promising results, which can be roughly divided into two groups, i.e. the parametric and non-parametric methods.

1.2.1 Parametric methods

The parametric methods [1], [2], [14] treat the image annotation as a multi-label classification problem. These methods are mostly based on the BoW paradigm [15] that is composed of local feature extraction, dictionary learning, feature encoding, pooling and the classifier training. In general, visual low-level features are initially extracted from each image. Then the dictionary is constructed by either k-means or sparse coding [10], [14], [16]. To capture the final representations of the images, a max or mean pooling strategy is employed to cluster the low-level features. Based on the representations, several classifiers are created using machine learning techniques such as SVM [17] for annotating the test images. The performances of created classifiers based image annotation methods are highly depend on the training data. It would be sharply degraded when the testing data has inconsistent label combinations with the training data.

1.2.2 Non-parametric methods

The non-parametric methods [5], [6] consider the image annotation as a sparse reconstruction problem, which is based on the reconstruction coefficients. In this scheme, the feature representations of training images are directly selected as the basis term. Then the labels of training data are propagated to the test data according to the reconstruction coefficients. However, the main limitation of such kind of methods is that they assume all the labels are independent of each other. Moreover, if the ground truth labels of the training data are in-complete, these reconstruction based methods become sensitive to the training labels. In addition, the label propagation may fail as similar representations would refer to different semantic objects.

ISSN 2348 – 7968

2. RELATED WORK

Abundant literature has been dedicated to the image annotation, dictionary learning and tremendous progress has been made ranging from efficient and scalable algorithm for different datasets. In related work, we briefly present a review on existing image annotation methods.

2.1 Image Annotation

Automatic image annotation aims to assign images with human predefined labels, which is usually viewed as a typical multi-label learning problem. The existing methods can be roughly divided into three categories including 1) Classification based [1], [2]. 2) Probabilistic modelling based [3], [4]. And 3) Reconstruction-based methods [5], [6].

2.1.1 Classification based methods

The classification based scheme annotates the images by training the semantic label classifiers.

In [2], Cusano et al. conquer the annotation problem by solving a multi-classification problem. The paper describes an innovative image annotation tool for classifying image regions in one of seven classes sky, skin, vegetation, snow, water, ground, and buildings or as unknown. This tool could be productively applied in the management of large image and video databases where a considerable volume of images/frames there must be automatically indexed. The annotation is performed by a classification system based on a multi-class Support Vector Machine.

Similarly, Carneiro et al. [1] propose to annotate the images without prior segmentation for each class by estimating the corresponding semantic class probability map. A probabilistic formulation for semantic image annotation and retrieval is proposed. Annotation and retrieval are posed as classification problems where each class is defined as the group of database images labelled with a common semantic label. The [18] main limitation of classification based method is that they need the supervised label information of the training images to train the Classification model.

2.1.2 Probabilistic modelling based methods

The probabilistic modelling based methods focus on modelling the distribution each label based on the visual features and attempt to infer the correlation or joint probability between images and annotation keywords.

Ueda and Saito [3] propose a generative model for multi-label leaning that explicitly incorporates the pair wise correlation between any two class labels. Conventionally, the binary classification approach has been employed, in which whether or not text belongs to a category is judged by the binary classifier for every category. In contrast, our approach can simultaneously detect multiple categories of text using PMMs. We derive efficient learning and prediction algorithms for PMMs. We also empirically show that our method could significantly outperform the conventional binary methods when applied to multi-labelled text categorization using real World Wide Web pages. However[18], the computation efficient and accuracy are two limitations for modelling the semantic distributions.

2.1.3 Reconstruction-based methods

The third category i.e. reconstruction based methods makes use of the sparse reconstruction framework to accomplish this task.

In [5], Wang et al. employ the sparse coding framework to get the coefficients of reconstructions. Based on the non-zero coefficients, the labels of reference images can thus be transferred to the test image.

In this paper, author presents a multi-label sparse coding framework for feature extraction and classification within the context of automatic image annotation. First, each image is encoded into a so-called super vector, derived from the universal Gaussian Mixture Models on order less image patches. Then, a label sparse coding based subspace learning algorithm is derived to effectively harness multi label information for dimensionality reduction. Finally, the sparse coding method for multi-label data is proposed to propagate the multi-labels of the training images to the query image with the sparse l1 reconstruction coefficients.

Furthermore, Chen et al. [6] introduce a prior about the exclusive labels to automatically annotate the images. In this paper a novel approach to multi labels image classification which incorporates a new type of context label exclusive context with linear representation and classification. Given a set of exclusive label groups that describe the negative relationship among class labels, our method, namely LELR for Label Exclusive Linear Representation, enforces repulsive assignment of the labels from each group to a query image. The problem can be formulated as an exclusive Lasso (eLasso) model with group overlaps and transformation. Since existing eLasso solvers are not directly applicable to solving such an variant of eLasso in their setting, It propose a Nesterovs smoothing approximation algorithm for efficient optimization.

One [18] of the drawbacks is that the existed methods are based on the raw low-level features, which cause the annotation results to be sensitive to the noisy feature representation.

ISSN 2348 – 7968

2.2 Dictionary learning

Recent work on dictionary learning can be generally classified into two categories, say unsupervised and supervised dictionary learning.

2.2.1 Unsupervised dictionary learning

The former one focuses on simultaneously minimizing the reconstruction error and the residual error.
In [7], the K-SVD and K-means clustering techniques are used to learn an over-complete dictionary from image patches. In this paper author proposes a novel algorithm for adapting dictionaries in order to achieve sparse signal representations. Given a set of training signals, we seek the dictionary that leads to the best representation for each member in this set, under strict sparsely constraints. We present a new method the K-SVD algorithm generalizing the K-means clustering process. K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better at the data. The update of the dictionary columns is combined with an update of the sparse representations, there by accelerating convergence. The K-SVD algorithm is flexible and can work with any pursuit method (e.g., basis pursuit, FOCUSS, or matching pursuit). We analyse this algorithm and demonstrate its results both on synthetic tests and in applications on real image data.

Lee et al. [8] treat the dictionary learning problem as a least squares problem, and solve it by an iterative algorithm to minimize the reconstruction error. Sparse coding provides a class of algorithms for finding succinct representations of stimuli; given only unlabeled input data, it discovers basis functions that capture higher-level features in the data. However, finding sparse codes remains a very difficult computational problem. In this paper, it presents efficient sparse coding algorithms that are based on iteratively solving two convex optimization problems: an L1-regularized least squares problem and an L2-constrained least squares problem. We propose novel algorithms to solve both of these optimization problems. This algorithms result in a significant speedup for sparse coding, allowing us to learn larger sparse codes than possible with previously described algorithms. We apply these algorithms to natural images and demonstrate that the inferred sparse codes exhibit end-stopping and non-classical receptive field surround
Suppression and, therefore, may provide a partial explanation for these two phenomena in V1 neurons.
Wright et al. [9] directly employ the training samples of the whole training set as the dictionary, and then they compute the least residual errors for face recognition. In this project, it implements a robust face recognition system via sparse representation and convex optimization. it treat each test sample as sparse linear combination of training samples, and get the sparse solution via L1-minimization. We also explore the group sparseness (L2-norm) as well as normal L1-norm regularization. We discuss the role of feature extraction and classification robustness to occlusion or pixel corruption of face recognition system. The dictionaries generated via unsupervised learning are expected to be of low reconstruction error and better generalization. Due to the lack of the label information, these methods have limited discriminative power.

2.2.2 Supervised dictionary learning

The supervised dictionary learning aims for discriminative representation by embedding the information of labels into the dictionary. The existing supervised dictionary learning approaches can be roughly divided into three main categories regarding the structure of dictionaries.

2.2.2.1 Learning multiple dictionaries

The first category is based on learning multiple dictionaries to promote the discrimination among different categories.

Zhou et al. [10] learns multiple dictionaries for visually correlated object category based on the Fisher discrimination criterion. Object recognition is challenging especially when the objects from different categories are visually similar to each other. In this paper, it presents a novel joint dictionary learning (JDL) algorithm to exploit the visual correlation within a group of visually similar object categories for dictionary learning where a commonly shared dictionary and multiple
Category specific dictionaries are accordingly modelled. To enhance the discrimination of the dictionaries, the dictionary learning problem is formulated as a joint optimization by adding a discriminative term on the principle of the Fisher discrimination criterion. As well as presenting the JDL model, a classification scheme is developed to better take advantage of the multiple dictionaries that have been trained. The effectiveness of the proposed algorithm has been evaluated on popular visual benchmarks. One [18] drawback of this kind of category is it's computational in efficiency, especially when the number of classes is relatively large.

2.2.2.2 Learn a compact and discriminative dictionary

ISSN 2348 – 7968

The thought of the second category is to learn a compact and discriminative dictionary by merging or selecting informative visual words from a larger dictionary.

In [11], they develop to merge the visual dictionary visual words by considering the trade off between the intra-class compactness and interclass discrimination power. The method is simple and extremely fast, making it suitable for many applications such as semantic image retrieval, web search, and interactive image editing. It classifies a region according to the proportions of different visual words (clusters in feature space). The specific visual words and the typical proportions in each object are learned from a segmented training set. The main contribution of this paper is twofold I) an optimally compact visual dictionary is learned by pair-wise merging of visual words from an initially large dictionary. The final visual words are described by GMMs. II) a novel Statistical measure of discrimination is proposed which is optimized by each merge operation.

In [12] paper, Qiu et al. present an approach for dictionary learning of action attributes via information maximization based on Gaussian Process. The [18] shortcoming of this category is that the results are easily influenced by the noise and the intra-class visual variation.

2.2.2.3 Incorporates the label information into objective function

The last kind of category incorporates the label information as a regularization term into the objective function.

In [13] paper, Jiang et al. simultaneously integrate the label consistent constraint, the reconstruction error and the classification error into one single objective function to achieve the goal. A label consistent K-SVD (LC- KSVD) algorithm to learn a discriminative dictionary for sparse coding is presented. In addition to using class labels of training data, we also associate label information with each dictionary item (columns of the dictionary matrix) to enforce discriminability in sparse codes during the dictionary learning process. More specifically, we introduce a new label consistency constraint called discriminative sparse-code error and combine it with the reconstruction error and the classification error to form a unified objective function. The optimal solution is efficiently obtained using the K-SVD algorithm. Our algorithm learns a single over complete dictionary and an optimal linear classifier jointly. The incremental dictionary learning algorithm is presented for the situation of limited memory resources. It yields dictionaries so that feature points with the same class labels have similar sparse codes. The [18] numerical evaluations of the supervised dictionary learning demonstrate the advantage of using the label training data. But it costs a high price to collect the fully supervised data, which is obviously impractical in real cases.

4. System Architecture

The system architecture is explained as follows [18]:

1] Firstly given a set of weakly label training images as shown in Fig. 1(a), we compute a non-occurrence label matrix based on the annotations of each single training image.

2] The computed matrix is then used to discover the exclusive label groups as shown in Fig. 1(b). In each exclusive group, we train the label-specific dictionaries by introducing the Fisher discriminates criterion as a regularization term.

3] Afterwards, we obtain the discriminative dictionaries and the corresponding label of each dictionary item as shown in Fig. 1(c). Inspired by the observation that co-occurrence labels would provide the context information, Author proposes a multi-task learning based method to exploit the semantic correlation between the dictionary items and the labels.

4] Then, we add the context information into the original dictionary labels based on the correlations as shown in Fig. 1(d). So far, each visual word is related with at least one label.

5]In the test time, since each dictionary visual word has the corresponding labels, a group sparse reconstruction based framework is developed to obtain the reconstruction coefficients based on the semantic dictionary as shown in Fig. 1(e).

6]Finally, with the reconstruction coefficients and dictionary labels, the test image is annotated through a robust label propagation method based on the Rank SVM as shown in Fig. 1(f) and (g).

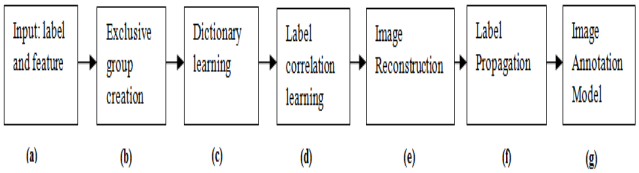

Fig.1 Block Diagram of System Architecture

ISSN 2348 – 7968

4. Conclusions

In this paper, a novel dictionary learning method for multi image annotation proposed, to solve the problem of inconsistent label combination. The main contribution is that, embedding the label information into the discriminative dictionary. Also the exclusive label discovery and the dictionary expansion on the basis of semantic correlation between the labels help to solve the problem of image annotation for weakly labeled data. The system consists of training part and testing part.

The given framework has three main limitations. First, dictionary size cannot be automatically determined, second unbalanced training data distribution, third incomplete labeled training data which influence the performance. In this paper I presented the different methods used for image annotation and dictionary learning. Each method applies its own methods and approaches to improve image annotation.

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.394410, Mar. 2007.

[2] C. Cusano, G. Ciocca, and R. Schettini, Image annotation using SVM, Proc. SPIE, vol. 5304, pp. 330338, 2004.

[3] N. Ueda and K. Saito, Parametric mixture models for multi-labeled text, in Proc. NIPS, 2002, pp. 721728.

[4] T. Gri_ths and Z. Ghahramani, Infinite latent feature models and the Indian buffet process, in Proc. NIPS, 2005, pp. 475482.

[5] 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), Jun. 2009, pp. 16431650.

[6] X. Chen, X.-T. Yuan, Q. Chen, S. Yan, and T.-S. Chua, Multi-label visual classification with label exclusive context, in Proc. ICCV,
Nov. 2011, pp. 834841.

[7] 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. 43114322, Nov.2006.37

[8] H. Lee, A. Battle, R. Raina, and A. Y. Ng, E_cient sparse coding algorithms, in Proc. NIPS, 2006, pp. 801808.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern
Anal. Mach. Intell., vol. 31, no. 2, pp. 210227, Feb. 2009.

[10] N. Zhou, Y. Shen, J. Peng, and J. Fan, Learning inter-related visual dictionary for object recognition, in Proc. IEEE Conf. CVPR, Jun.
2012, pp. 34903497.

[11] J. Winn, A. Criminisi, and T. Minka, Object categorization by learned universal visual dictionary, in Proc. ICCV, Oct. 2005, pp.
18001807.

[12] Q. Qiu, Z. Jiang, and R. Chellappa, Sparse dictionary-based representation and recognition of action attributes, in Proc. ICCV,
Nov. 2011, pp. 707714.

[13] 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. 26512664, Nov. 2013.

[14] F. Perronnin, J. Snchez, and T. Mensink, Improving the Fisher kernel for large-scale image classi_cation, in Proc. ECCV, 2010, pp. 143156.

[15] L. Fei-Fei and P. Perona, A Bayesian hierarchical model for learning natural scene categories, in Proc. IEEE Conf. CVPR, Jun. 2005, pp. 524531.

[16] M. Zheng et al., Graph regularized sparse coding for image representation, IEEE Trans. Image Process., vol. 20, no. 5, pp. 13271336, May 2011.

[17] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, p. 27, Apr. 2011.

[18] Xiaochun Cao, Senior Member, IEEE, Hua Zhang, Xiaojie Guo,Member, IEEE, Si Liu, and Dan Meng, Senior Member, IEEE "SLED: Semantic Label Embedding Dictionary Representation for Multilabel Image Annotation" IEEE Transactions On Image Processing, Vol. 24, No. 9, September 2015.