International Journal of Emerging Technologies in Engineering Research (IJETER)
Volume 4, Issue 5, May (2016) www.ijeter.everscience.org

An Efficient Multi-Label Image Classification Method for Image Annotation

Apurva Dhurandher
CSE, SSTC, CSVTU, Bhilai, Chhattisgarh, India.

Shreya Jain
CSE, SSTC, CSVTU, Bhilai, Chhattisgarh, India.

Abstract – With growth in the volume of visual data, there is a growing interest in efficient categorization of visual data for better retrieval and browsing of such data using semantic keywords. Recently, multi-label classification is also an active research area due to its applicability to real-life data. In this paper we formulated multi-label classification of images as an image annotation tasks by using the output from the multi-labeled classifiers as the semantic keywords for annotation. We used convolutional neural network for feature extraction and principal component analysis for dimensionality reduction. We also proposed a mutual information based two stage multi-label classification method for efficient classification by taking advantage of label correlation among different class labels. Experimental results on Pascal VOC datasets shows our method outperforms the binary relevance method.

Index Terms – Visual Data, Classification, Neural Network.

1. INTRODUCTION

With advancement in digital technologies and internet the number visual data such as images and videos have increased manifold. For efficient image retrieval and processing it is required to annotate the given image with some keywords and concepts by identifying a set of objects present in each image. Man- ual text based image annotation was used in earlier days but it was very time consuming and expensive and it becomes infeasible with such exponential increase in visual data. The image annotation problem for images contain- ing more than one class labels can be cast as a multi-label classification problem. Multi-label classification generalizes the standard multi- class classification by allowing each instance to be simultaneously assigned into multiple la- bel categories. A key challenge for multi-label classification is label sparsity that is many la- bels lacks sufficient training instances for build- ing efficient classifiers. Also many labels are correlated with each other that is occurrence of one label influence occurrence other labels too. For example, presence of airplane in the image increases the chance of sky in the back- ground. Similarly, presence of seasons label summer makes mutually exclusive labels like winter improbable. Hence,exploiting label de- pendency can significantly boost classification performance. Most of multi-label method uses binary decomposition of multi-label datasets but uses the same features for training the classifiers which may contain redundant fea- tures. Hence using useful and discriminative features will also enhance the classification abil- ity. We perform following tasks in our pro- posed method.

We used convolutional neural network for feature extraction

• We performed principal component analy- sis for dimensionality reduction

• We used two stage binary relevance method to make our method scalable

• We used mutual information based label dependency modelling to enhance the clas- sification ability.

The rest of the paper is organized as follows. We firstly review some related work on image classification using multi-label learning and various methods to improve its classification ability in section 2. In section 3 we present our proposed work. Section 4 summarizes our experiments and results. Finally, we conclude this paper in section 5.

2. RELATED WORKS

Multi-label classification is gaining widespread attention nowadays due to its applicability to real life problems. In [1] proposes a convo- lutional neural network based model where shared CNN weights is connected with object segment hypotheses. In [2] author proposes to use label-specific features to exploit the benefit of discrimination of different class labels. He proposed to use k-means clustering followed by data transformation based on the cluster centres. Various multilabel classification meth- ods [3] are proposed which fall into two cate- gories 1) Problem transformation where data is transformed to be used 2) Algorithm adap- tation where single-label methods are adapted to be used on multi-label datasets. In [4] uses a similar two stage classification procedure for exploiting co-occurrence of classes in label sets of documents using iterative SVMs and a gen- eral kernel function for heterogeneous features. In [7] author uses a deterministic Bayesian net- work for capturing deterministic relationships and in [8] author uses a markov network based model for dependency modelling as well as for efficient inference. In [9] author proposed a hy- brid method using k-NN and SVM for image classification and annotation.

3. PROPOSED METHOD

In this section we will give a detailed explana- tion of our proposed method. As mentioned in previous section the current research is focused in different aspects of multi-label learning such as improving classification ability of the classi- fiers by modelling label dependency, by extract- ing good features, selecting relevant features for the classifiers and making the algorithm scalable and efficient. We tried to combined all good features from the previous work and to improve the classification ability by utiliz- ing the label correlation among different class labels.

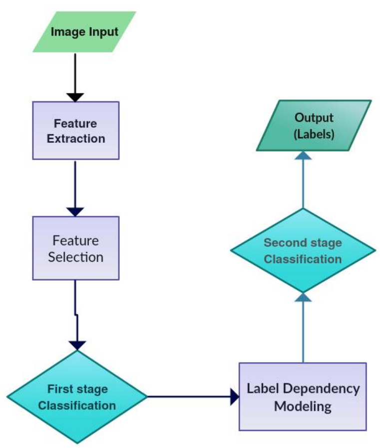

Figure 1: Flow diagram of our proposed method

3.1. Feature extraction

The classification ability of the classifier is greatly influenced by the features extracted. Hence to get better features for our classifiers we employed the convolutional neural network (CNN). CNN is multi-layered perceptron in-spired from biological process of vision system in animals. A stack of different layers are used to form the architecture of CNN namely 1) Convolutional layer 2) Pooling layer 3) ReLU layer 4) Fully connected layer and 5)Loss layer

Convolutional network consist of a set of filters whose activation maps can be learned. The parameters learned are shared and it is also invariant to the spatial arrangement of the ac- tive features. Pooling layer uses non-linear function to down-sample the output obtained from convolutional network. The most popular function for pooling is max-pooling. Rectified Linear Units (ReLU) is a layer of neurons that applies the non-saturating activation function by making all the negative values in its to 0. Fully connected layers do the modeling of the classifier of the neural network. The last layer is loss layer and it specifies how the network training penalizes the difference between the output and ground truth. After training the CNN the features can be extracted from the desired layer by doing forward pass of the raw input.

3.2. Dimensionality reduction

The features extracted also contains correlated as well as redundant data and since the num- ber of available training data is less compared to the dimensionality of input features it is de- sired to transform the input data to a smaller and uncorrelated space which helps in efficient modelling of the classifiers. So, we used prin- cipal component analysis (PCA) for this task. PCA is an orthogonal linear transformation that transforms the data to a new coordinate system where coordinates are organised in the descending order of the variance. After per- forming PCA we selected top m ($m << M$) features and give these features as the input to our classification model.

3.3. Binary relevance

Since in our problem setting also known as multi-label learning, each data instance is asso- ciated with more than one class label. Binary relevance is a popular method in the multi- label learning which decomposes the multi- label dataset in binary datasets corresponding to each label. Classifiers are built for respec- tive datasets and prediction is done by taking into account all the output from these binary datasets. This method is simple, efficient, scal- able and also give us the flexibility to use any off-the-shelf classifier as our base classifiers. Though it is criticized for ignoring label depen- dencies among different class labels.

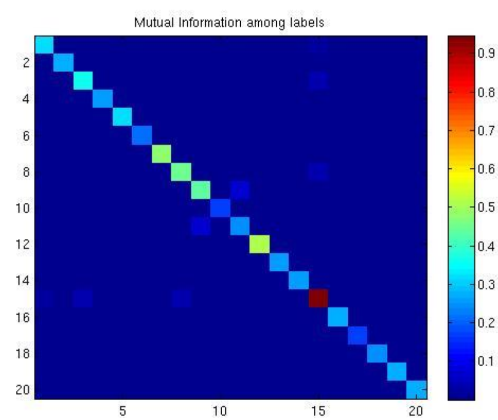

Figure 2: Measure of mutual information among labels

3.4. Label dependency modelling

The class labels in multi-label datasets are usu- ally correlated and utilising this information during classification will improve our classifi- cation ability but as mentioned above binary relevance method ignores the label dependen- cies and hence it will useful to use the label correlation information during our classifica- tion stage. We used mutual information(MI) which is a measure of the mutual dependence between the two variables and based on the MI values we selected the labels from previ- ous classification stage and augmented their output to the first stage feature vector to form new augmented feature vector for second stage classification. Figure 2 shows the measure of mutual information between labels and it is apparent from the figure that all labels are not independent of each other.

3.5. Second stage classification

The second stage classifiers are modelled using the new input features which also contains the output of classifiers from first stage classification whose labels are strongly dependent on this label. We used support vector machines (SVMs) as our base classifiers. The prediction from all other classifiers was concatenated to form the final prediction.

4. EXPERIMENTS

4.1. Data description

We used Pascal VOC2012 dataset for our ex- periments. It is a benchmark in visual object category recognition and detection tasks. It consists of 20 classes namely person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa and tv/monitor We used 10,000 images out of 11540 to train the model and remaining 1540 was used for testing.

4.2. Evaluation measures

For performance evaluation we used the evalu- ation metrics of multi-label classification meth- ods as the prediction may be partially correct and single-label evaluation metric may not be able to capture this notion. The metrics we used are summarised below. Yi and Zi are given and predicted label sets, respectively and n is the number of instances.

1. Accuracy(A): Accuracy is defined as av- erage proportion of the predicted correct labels to the total number (predicted and actual) of labels.

$$\text{Accuracy}, A = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \quad (1)$$

2. Recall (R): Recall is the proportion of pre- dicted correct labels to the total number of predicted labels, averaged over all in- stances.

$$\text{Recall}, R = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i|} \quad (2)$$

3. F1-Measure (F): F1-Measure is the har- monic mean of precision and recall, as followed from single-label classification.

$$\text{F1} - \text{Measure}, F = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i| + |Z_i|} \quad (3)$$

4.Hamming Loss (HL): Hamming loss eval- uates the fraction of misclassified instance- label pairs, i.e. a relevant label is missed or an irrelevant is predicted.

$$\text{Hammingloss}, HL = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}|Z_i \Delta Y_i| \quad (4)$$

4.3. Experimental Setups

We used Caffe [6] framework along with matlab to extract features. The Caffenet from model zoo was used to extract fea- tures. The fc7 features were extracted by chopping off other top layers from the caffenet model and employing forward pass method from caffenet. Matlab tool- box function pca was used to perform principal component analysis. We var- ied the number of principal components and finally selected 200 from 4096 dimen- sional space for further experiments. The multi-label dataset was decomposed into 20 binary datasets and SVMs with linear kernels are used as base classifiers. We used LIBSVM [5] library for implementing SVMs. To model the label dependencies, we first found the pairwise mutual information of all the classes and then using are dependent on the given label. We ex- perimented with various threshold value and finally selected 0.002 and 0.01 for la- bel selection. For second stage of classi- fication, we augmented the input feature space with the outputs of the classifiers whose labels with given label had high mutual information value. The results are reported in the next section.

	BR	brPCA (50)	brPCA(200)	PM(t=0 .01)	PM(t=0. 002)
Accuracy	0.31 87	0.3180	0.3498	0.3683	0.4030
Recall	0.33 82	0.3355	0.3645	0.3831	0.4320
F1	0.35 77	0.3595	0.3907	0.4118	0.4514
Hammin g loss	0.06 30	0.0624	0.0598	0.0591	0.0423

Table 1: Evaluation results for Pascal VOC dataset using different methods

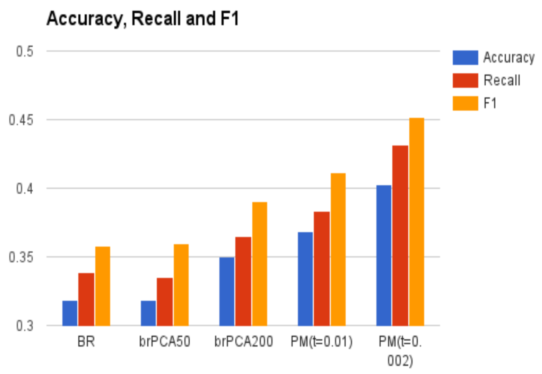

Figure 3: Evalution Scores for different methods on pascal dataset

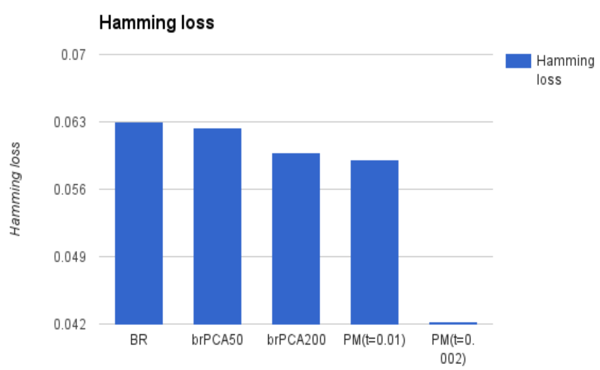

Figure 4: Hamming loss for different methods on pascal dataset

5. RESULTS

In first experiment, we only used binary relevance method of multi-label classifica- tion on the extracted features which we called BR. Then we performed PCA for dimensionality reduction as well as select- ing features with high variance. We re- ported the results with 50 and 200 dimen- sional input space after performing PCA, which we called brPCA50 and brPCA200 respectively. Finally we used mutual in- formation to exploit label correlation by selecting some labels as dependent based on some threshold value. We reported two of such experiment with threshold value(MI score) = 0.01 and 0.002 respec- tively. We called these proposed methods as PM(t=0.01) and PM(t=0.002).

The results are also plotted as bar graphs. Some of the sample test images(Fig 5,6 and 7) are also shown with their Ground truth(GT) and predicted output.

Figure 5: GT: {dog, so f a}, Output:{dog}

Figure 6: GT: {bottle, dinning − table, person}, Output:{dinning − table, person}

Figure 7: GT: {bird}, Output:{bird}

6. CONCLUSION

In this paper we performed multi-label classi- fication of images and proposed two methods to improve the classification ability. First by using efficient features for decomposed binary datasets and the second by modelling label de- pendencies by using pairwise mutual informa- tion among the class labels. We experimented with the various dimensionality reduction size and feature selection thresholds. For efficient classification we employed convolutional neu- ral network for feature extraction. Also by using binary relevance we ensured the scala- bility of our method For future we would like to explore better methods for label specific fea- ture selection as well graphical model based label dependencies modelling. Also we would like to extend our classification method to vari- ous other datasets such as dress classification, hand-gesture classification, text classification and other multi-labeled datasets.

REFERENCES

[1] Yunchao Wei, Wei Xia, Junshi Huang, Bingbing Ni, Jian Dong, Yao Zhao, Shuicheng Yan CNN: Single-label to Multi-label arXiv : 1406.5726, 2014
[2] M. L. Zhang and L. Wu, LIFT: Multi-label learning with label-specific features IEEE Transactions on Pattern Analysis and Ma- chine Intelligencevol. 37, no. 1, pp. 107- 120, Jan. 1 2015.
[3] M. L. Zhang and Z. H. Zhou A Review on Multi-Label Learning Algorithms IEEE Transactions on Knowledge and Data En- gineering, 2014vol. 26, no. 8, pp. 1819- 1837, Aug. 2014
[4] S. Godbole and S. Sarawagi Discrimina- tive methods for multi-labeled classifica- tion Proc. of Pacific-Asia Conf. on Know. Disc. and Data Mining (PAKDD), 2004
[5] Chang, Chih-Chung and Lin, Chih-Jen LIBSVM: A library for support vector ma- chines ACM Transactions on Intelligent Systems and Technology, 2011
[6] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor Caffe: Convolutional Architecture for Fast Feature Embedding arXiv preprint arXiv:1408.5093, 2014
[7] Papagiannopoulou, Christina and Tsoumakas, Grigorios and Tsamardinos, Ioannis Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2015 KDD-2015
[8] Arias, Jacinto and Gamez, Jose A. and Nielsen, Thomas D. and Puerta, Jose M A Scalable Pairwise Class Interaction Frame- work for Multidimensional Classification Int. J. Approx. Reasoning, 2016
[9] Guanglei Chu and Kai Niu and Baoyu Tian Automatic image annotation combin- ing SVMS and KNN algorithm IEEE 3rd International Conference on Cloud Com- puting and Intelligence Systems, 2014