MULTI-LABEL SVM ACTIVE LEARNING FOR IMAGE CLASSIFICATION

Xuchun Li, Lei Wang, Eric Sung

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798
pg03454644, elwang, eericsung@ntu.edu.sg

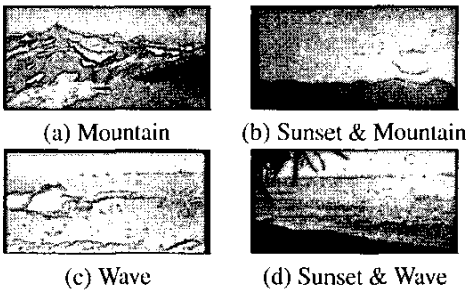
ABSTRACT

Image classification is an important task in computer vision. However, how to assign suitable labels to images is a subjective matter, especially when some images can be categorized into multiple classes simultaneously. Multi-label image classification focuses on the problem that each image can have one or multiple labels. It is known that manually labelling images is time-consuming and expensive. In order to reduce the human effort of labelling images, especially multi-label images, we proposed a multi-label SVM active learning method. We also proposed two selection strategies: Max Loss strategy and Mean Max Loss strategy. Experimental results on both artificial data and real-world images demonstrated the advantage of proposed method.

1. INTRODUCTION

Image classification is an important task in computer vision. It can give much help to image indexing and retrieval, because both of them need correctly label images. In an image classification task, some images can easily be assigned with unique labels, such as Figure 1(a) and (c), they can be labelled as "Mountain" and "Wave", respectively. The classification over these images is known as multi-class image classification. However, for the images such as Figure 1(b) and (d), the case is different. How should we label them? "Mountain", "Sunset", or "Wave"? Due to the subjectivity of human perception, each of them may be a correct label. In this case, if multi-class classification method is still used, these images will be rigidly grouped into only one of the classes. This can be harmful to the applications such as image retrieval. Multi-label classification is a case in which the data can have one or more labels. Therefore, multi-class classification can also be viewed as a special case of multi-label classification. Multi-label classification has been applied to document classification [1] and functional gene classification [2] problems. In this paper, we focus on multi-label image classification problem or image annotation problem [3].

In real applications, labelling multi-label images is often a time-consuming and costly task. Fortunately, in many cases, we can carefully select images used to query the oracle[1] instead of randomly picking images. If the selection strategy is well-designed, this approach can give rise to better classification result under a given number of labelling chances. This procedure can be categorized into active learning. Active learning is a mechanism which aims to optimize the classification performance while minimizing the number of needed labelled data for training. The first active learning mechanism [4] was introduced by Cohn et al. Recently, Tong et al. [5] used version space to analyze SVM and gave an optimal selection strategy. In [6], Yan et al. proposed a multi-class active learning model for video labelling. Although these methods achieved success in some applications, none of them addressed multi-label classification problems.



(a) Mountain (b) Sunset & Mountain

(c) Wave (d) Sunset & Wave

Fig. 1. The images with multi-label

Therefore, in this paper, we proposed a multi-label SVM active learning method and used it to solve multi-label image classification problem. The key issue of active learning mechanism is the optimization of selection strategy for fastest learning rate. Hence, we proposed two selection strategies: Max Loss strategy and Mean Max Loss strategy. Experimental results on both artificial and real-world image databases demonstrated that the proposed method can effectively reduce the demands of labelled images while maintaining a good classification performance.

[1] Oracle knows the ground truth of the selected image. In practice, oracle can be a group of human observers or annotators.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. MULTI-LABEL SVM

Now the multi-label SVM is introduced, which will be used for active learning in this paper. Multi-label SVM uses the one-versus-all method to combine predictions of multiple binary SVM classifiers. Let $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{y}_1), ..., (\mathbf{x}_m, \mathbf{y}_m)\}$, where $\mathbf{x}_i (\mathbf{x}_i \in \mathbf{R}^n)$ denotes a n-dimensional feature vector, its p-dimensional label vector is $\mathbf{y}_i (\mathbf{y}_i \in \mathbf{Y}^p)$ which consists of {-1, +1}, where +1 means the data belongs to this class and -1 mean the data does not. p is the size of label set. The j-th component of $\mathbf{y}_i$ corresponds to the output of j-th binary SVM. Given $\mathcal{D}$, each class is seperated from all other classes by a binary SVM classifier. Then p binary SVM classifiers are obtained in total. For i-th ($i = 1, ...p$) classifier, it can be written as $\langle \mathbf{w}_i^*, \mathbf{x} \rangle + b_i^* = 0$. $\mathbf{w}_i^*$ and b_i^* can be found by:

$$\begin{aligned} &minimize: \quad \Phi(\mathbf{w}_i) = \tfrac{1}{2}\|\mathbf{w}_i\|^2 + C\sum_{j=1}^{m} \xi_j^p \\ &subject\ to: \quad y_k\left(\langle \mathbf{w}_i, \mathbf{x}_k \rangle + b_i\right) \geq 1 - \xi_k, \quad k = 1, \ldots, m \end{aligned}$$

where ξ_k ($\xi_k \geq 0$) is the k-th slack variable and C is the parameter controlling the trade-off between function complexity and training error. For each binary SVM classifier f_i, a threshold t_i will be set. This will be discussed in the next section.

3. MULTI-LABEL SVM ACTIVE LEARNING

Assume that $\mathcal{I}$, a large set of multi-label images consists of a small set of labelled images, $\mathcal{L}^0$, and an unlabelled image set, $\mathcal{U}^0$. Proposed method focuses on the problem that given a limited number of labelling chances, how to select the most informative images from $\mathcal{U}^0$ to be labelled to train a classifier which can give the best classification performance on $\mathcal{I}$. Its procedure is described as follows. At the beginning, a multi-label SVM classifier, f^0, is trained on $\mathcal{L}^0$. After that, the most informative unlabelled images are selected from $\mathcal{U}^0$ by using the proposed selection strategy. After being labelled, they will be added into $\mathcal{L}^0$ to form a new set, $\mathcal{L}^1$. Afterwards, a new training-selecting-labelling learning cycle will begin based on $\mathcal{L}^1$ until the labelling chances are used up.

The key issue of active learning mechanism is how to select the most informative data per learning cycle to obtain the maximal improvement of classification performance. It is known that, by minimizing the expected loss over the data distribution, SVM can obtain the best classification performance [7]. Hence, we could use the decrease of the expected loss between two contiguous learning cycles as an indicator of the improvement of classification performance. Therefore, the objective of the proposed selection strategy is to maximize the decrease of the expected loss on $\mathcal{I}$ at the i-th learning cycle. Let $\mathcal{U}_s^{i*}$ denote the set of selected images, and $\mathcal{L}^i$, $\mathcal{U}^i$ denote the labelled and unlabelled image sets at i-th learning cycle. Hence, $\mathcal{U}_s^{i*}$ should satisfy

$$\mathcal{U}_s^{i*} = \arg\max_{\mathcal{U}_s^i \subset \mathcal{U}^i} [\sum_{\mathbf{x} \in \mathcal{I}} L^i(\mathbf{x}) - \sum_{\mathbf{x} \in \mathcal{I}} L^{i+1}(\mathbf{x}, \mathcal{U}_s^i)] \quad (1)$$

where $L^i(\mathbf{x})$ is the loss value of image $\mathbf{x}$ at the i-th learning cycle. $L^{i+1}(\mathbf{x}, \mathcal{U}_s^i)$ is the loss value of $\mathbf{x}$ at $(i+1)$-th learning cycle. $L^{i+1}(\mathbf{x}, \mathcal{U}_s^i)$ is a function of selected image set $\mathcal{U}_s^i$ because different $\mathcal{U}_s^i$ leads to different multi-label SVM classifier, f^{i+1}, at $(i+1)$-th learning cycle and then leads to different loss value for $\mathbf{x}$. But for the present learning cycle i, the classifier, f^i, has been fixed before selecting $\mathcal{U}_s^i$. Hence, $L^i(\mathbf{x})$ can be calculated before selecting $\mathcal{U}_s^i$.

By modifying the kernel of SVM, the images in the induced kernel space can become linearly separable [5]. This guarantees that the training error is zero and then the loss values of the images used for traing are zero. Note that there is $\mathcal{I} = \mathcal{L}^i + \mathcal{U}^i \quad (i = 0, 1, 2, ...)$. Therefore, Equation (1) becomes

$$\begin{aligned} \mathcal{U}_s^{i*} &= \arg\max_{\mathcal{U}_s^i \subset \mathcal{U}^i} [(\sum_{\mathbf{x} \in \mathcal{L}^i} L^i(\mathbf{x}) + \sum_{\mathbf{x} \in \mathcal{U}^i} L^i(\mathbf{x})) \\ &\quad -(\sum_{\mathbf{x} \in \mathcal{L}^{i+1}} L^{i+1}(\mathbf{x}, \mathcal{U}_s^i) + \sum_{\mathbf{x} \in \mathcal{U}^{i+1}} L^{i+1}(\mathbf{x}, \mathcal{U}_s^i))] \\ &= \arg\max_{\mathcal{U}_s^i \subset \mathcal{U}^i} [\sum_{\mathbf{x} \in \mathcal{U}^i} L^i(\mathbf{x}) - \sum_{\mathbf{x} \in \mathcal{U}^{i+1}} L^{i+1}(\mathbf{x}, \mathcal{U}_s^i)] \quad (2) \end{aligned}$$

In the i-th learning cycle, $\mathcal{U}^i$ consists of selected image set, $\mathcal{U}_s^i$, and non-selected image set, $\mathcal{U}_{\bar{s}}^i$. It is known that $\mathcal{U}^{i+1}$ will be $\mathcal{U}_{\bar{s}}^i$ because $\mathcal{U}_s^i$ will be removed from $\mathcal{U}^i$ after selection. As in [8], we assume that all the expected losses of $\mathbf{x}$ in $\mathcal{U}_{\bar{s}}^i$ have an equal influence between two contiguous learning cycles. Thus, Equation (2) can be rewritten as:

$$\begin{aligned} \mathcal{U}_s^{i*} &= \arg\max_{\mathcal{U}_s^i \subset \mathcal{U}^i} [(\sum_{\mathbf{x} \in \mathcal{U}_s^i} L^i(\mathbf{x}) + \sum_{\mathbf{x} \in \mathcal{U}_{\bar{s}}^i} L^i(\mathbf{x})) \\ &\quad - \sum_{\mathbf{x} \in \mathcal{U}_{\bar{s}}^i} L^{i+1}(\mathbf{x}, \mathcal{U}_s^i)] \\ &= \arg\max_{\mathcal{U}_s^i \subset \mathcal{U}^i} \sum_{\mathbf{x} \in \mathcal{U}_s^i} L^i(\mathbf{x}) \quad (3) \end{aligned}$$

This function indicates that the selected image set, $\mathcal{U}_s^{i*}$, consists of the images the sum of whose expected loss values are largest. In practice, these images can be selected as those which have the larger $|\mathcal{U}_s^{i*}|$ loss values in $\mathcal{U}^i$. ($|\mathcal{U}_s^{i*}|$ is the size of $\mathcal{U}_s^{i*}$)

The distinctive feature of multi-label image classification lies in that each image, $\mathbf{x}$, can have one or multiple labels. To achieve an accurate estimation of the loss value, $L^i(\mathbf{x})$, of $\mathbf{x}$ in Equation (3), the multi-label information has to be incorporated. Hence, in this paper, multi-label classification is viewed as a combination of n multi-class classification, where n is the number of predicted classes of $\mathbf{x}$. $L^i(\mathbf{x})$ is averaged over the n predicted classes. Thus, we define the loss value of $\mathbf{x}$ as

$$L(\mathbf{x}) = \frac{1}{n}\sum_{y=l_1}^{l_n}\sum_{j=1}^{p} D(m_{yj}f_j(\mathbf{x})) = \frac{1}{n}\sum_{y=l_1}^{l_n}\sum_{j=1}^{p}\max[(1-m_{yj}f_j(\mathbf{x})),0] \quad (4)$$

where $\{l_1,...,l_n\}$ are the indexes of predicted classes of $\mathbf{x}$. p is the total number of classes. m_{yj} is a component of a coding matrix M [7] which is a $p \times p$ matrix with diagonal components 1 and others -1. $f_j(\mathbf{x})$ is the output of $\mathbf{x}$ on the j-th binary SVM classifier. $D(m_{yj}f_j(\mathbf{x})) = \max\{(1 - m_{yj}f_j(\mathbf{x})), 0\}$ is the loss function of $\mathbf{x}$ classified by $f_j(\mathbf{x})$ and m_{yj} is used as the true label of $\mathbf{x}$. This follows the assumption in [8] that the prediction of $f_j(\mathbf{x})$ is positively correlated to the true label of $\mathbf{x}$.

In order to get the loss value of $\mathbf{x}$ on the predicted n classes, we need to decide the n predicted classes of $\mathbf{x}$. To achieve this, we should estimate the threshold of loss value on each binary SVM classifier. If the loss value of $\mathbf{x}$ on the J-th class is smaller than the threshold on the J-th binary SVM classifier, $\mathbf{x}$ will belong to class J. We estimate the J-th threshold t_J as:

$$t_J = \arg\min_{\mathbf{x}} \sum_{j=1}^{p} D(m_{Jj}f_j(\mathbf{x})), \ \forall \mathbf{x} \in \{\mathbf{x}|\mathbf{y}(J) = 1\} \quad (5)$$

where $\mathbf{y}$ is the label vector of $\mathbf{x}$ and $\mathbf{y}(J)$ is $\mathbf{y}$'s J-th component. $\mathbf{y}(J) = 1$ means $\mathbf{x}$ belongs to class J.

Hence, by substituting the Equation (4) into (3), the proposed Mean Max Loss(MML) Selection Strategy will select the image set $\mathcal{U}_s^{i*}$ satisfying:

$$\mathcal{U}_s^{i*} = \arg\max_{\mathcal{U}_s^i \in \mathcal{U}^i} \sum_{\mathbf{x}\in\mathcal{U}_s^i} \frac{1}{n}\{\sum_{y=l_1}^{l_n}\sum_{j=1}^{p}\max[(1-m_{yj}f_j(\mathbf{x})),0]\} \quad (6)$$

Although in multi-label image classification task, one image may have more than one labels, we could get a fairly simple selection strategy if we calculate the loss value only on the most certainly predicted class of this image. Hence, the proposed Max Loss(ML) Selection Strategy will select the image set $\mathcal{U}_s^{i*}$ as:

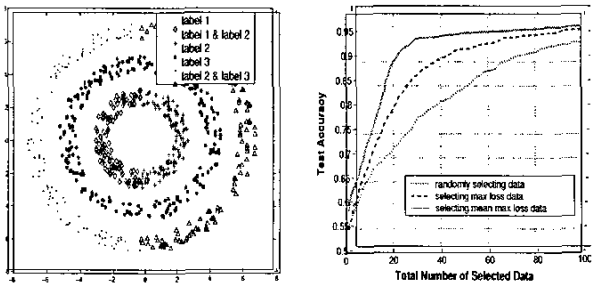
$$\mathcal{U}_s^{i*} = \arg\max_{\mathcal{U}_s^i \in \mathcal{U}^i} \sum_{\mathbf{x}\in\mathcal{U}_s^i}\sum_{j=1}^{p}\max\{(1-m_{yj}f_j(\mathbf{x})),0\} \quad (7)$$

where y is the index of the most certainly predicted class of $\mathbf{x}$. It can be found that the selection strategy proposed in [6] for multi-class active learning is similar to the proposed ML selection strategy and it is only a special case of the proposed MML selection strategy when the most certainly predicted class is involved only.

4. EXPERIMENT

4.1. Artificial Database

Figure 2(a) shows the artificial database consisting of three 2D circles with the same center. The data's multi-label information can be found from the legend of Figure 2(a). Figure 2(b) shows the comparison result of multi-label SVM among random selection, the proposed ML and MML selection strategies on the artificial database. From this figure, it can be seen that the two proposed strategies have a clear advantage over random selection. To achieve the same accuracy, for example, 90% test accuracy, nearly 35 data are saved by using the ML selection strategy and nearly 50 data are saved by using the MML selection strategy than random selection. Based on the same number of labelled training data, maximal 10% and 17% improvement are obtained by the two proposed strategies, respectively. Obviously, MML strategy is more effective than ML strategy. This is because we consider the multi-label information in MML strategy.



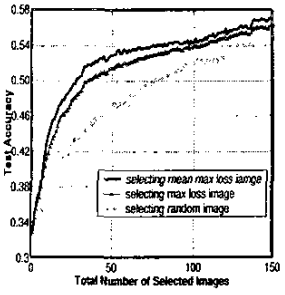
(a) Artificial Database (b) Comparison Result

Fig. 2. Comparison Result of Artificial Data Classification

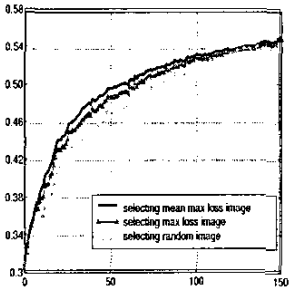
4.2. Real Color Image Database

The real color image database includes 400 general color images composed from *VisTex* of MIT and *Corel* Stock Photos. Four image classes are defined based on high-level semantics (i.e. defined by several human observers): "People", "Ocean", "Mountain", "Sunset". Each image belongs to one or several classes. For instance, if an image is a sunset over ocean as Figure 1(d), then this image belongs to both "Sunset" and "Ocean" classes. This database provides the ground truth for evaluation. A perceptually uniform color space, $CIE - Lab$, is used to represent general color images. Based on this color space, a feature vector of color Moments [9] is defined for each image. This feature vector consists of the mean, variance, and skewness of the pixel values of an image along L, a, and b axes, respectively. The dimensions of this feature vector is 3×3 (9 in total). In Figure 3, we plot the experimental result on the real color

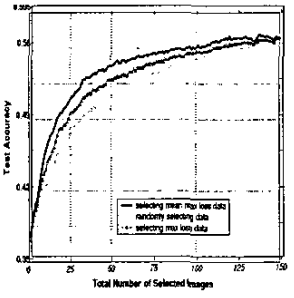
image database. Here, the numbers of initial training images are 20, 20 and 30, respectively. At each active learning cycle, the numbers of selected unlabelled images are 1, 5 and 1, respectively. It can be seen that the proposed MML selection strategy still shows the best performance.



(a)Initial 20, 1 Each



(b)Initial 20, 5 Each



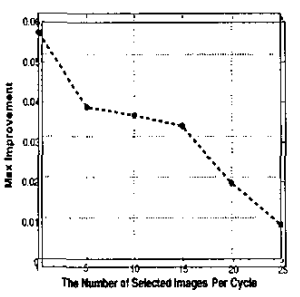
(c)Initial 30, 1 Each

Fig. 3. Comparison result of real image classification.

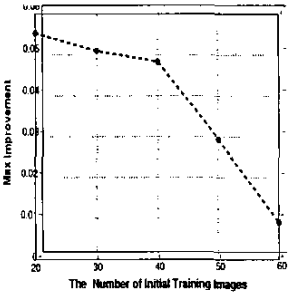
The relationship between the number of selected images per learning cycle and the maximal improvement achieved by the proposed MML selection strategy is shown in Figure 4 (a). It shows that the less the selected images per learning cycle, the higher the improvement of the MML strategy and that the MML strategy performs best when only one image is selected per learning cycle. This is because that selecting only one image per learning cycle gives the proposed active learning method more chances to interact with the oracle. This makes the selected images more informative for the subsequent classification. In Figure 4 (b), we show the relationship between the number of initial training images and the maximal improvement achieved by the MML selection strategy. It can be found that within the range from 20 to 50, the less the number of initial training images, the higher the improvement of the proposed strategy.

5. CONCLUSION

In this paper, a multi-label SVM active learning method was proposed for multi-label image classification. Based on theoretic analysis, we proposed two selection strategies: Max Loss strategy and Mean Max Loss strategy. To our best knowledge, we are the first one who use active learning to help multi-label classification. The experimental results on both artificial and real color image databases demonstrated that the two proposed methods can provide better classification performance than original multi-label SVM.



(a)



(b)

Fig. 4.

6. REFERENCES

[1] A. K. McCallum, "Multi-label text classification with a mixture model trained by em," in *AAAI'99 Workshop on Text Learning*, 1999.

[2] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *15th Annual Conference on Neural Information Processing Systems*, 2001, pp. 681–687.

[3] V.Lavrenko, R.Manmatha, and J.Jeon, "A model for learning the semantics of pictures," in *17th Annual Conference on Neural Information Processing Systems*, 2003.

[4] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.

[5] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2001.

[6] R. Yan, J. Yang, and A. Hauptmann, "Automatically labeling video data using multi-class active learning," in *The 9th International Conference on Computer Vision*, 2003, pp. 516–523.

[7] Erin L. Allwein, Robert E. Schapire, and Yoram Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *Journal of Machine Learning Research*, vol. 1, pp. 113–141, Dec 2000.

[8] C. Campbell, N. Cristianini, and A. Smola, "Query learning with large margin classifiers," in *Proceeding of the 17th International Conference on Machine Learning*, 2000, pp. 111–118.

[9] M. Stricker and M. Orengo, "Similarity of color images," in *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, 1995, pp. 381–392.