

Active learning SVM with regularization path for image classification

Fuming Sun · Yan Xu · Jun Zhou

Received: 27 December 2013 / Revised: 23 April 2014 / Accepted: 2 June 2014 /

Published online: 19 June 2014

© Springer Science+Business Media New York 2014

Abstract In classification problems, many different active learning techniques are often adopted to find the most informative samples for labeling in order to save human labors. Among them, active learning support vector machine (SVM) is one of the most representative approaches, in which model parameter is usually set as a fixed default value during the whole learning process. Note that model parameter is closely related to the training set. Hence dynamic parameter is desirable to make a satisfactory learning performance. To target this issue, we proposed a novel algorithm, called active learning SVM with regularization path, which can fit the entire solution path of SVM for every value of model parameters. In this algorithm, we first traced the entire solution path of the current classifier to find a series of candidate model parameters, and then used unlabeled samples to select the best model parameter. Besides, in the initial phase of training, we constructed a training sample sets by using an improved K-medoids cluster algorithm. Experimental results conducted from real-world data sets showed the effectiveness of the proposed algorithm for image classification problems.

Keywords Active Learning · regularization · SVM

1 Introduction

The rapid development of multimedia and network technologies has lead to an explosive growth of social media in recent years, such as Flickr and Youtube. Such media repositories encourage internet users to share, evaluate and distribute media information, as well as to annotate the photos with tags (i.e., keywords). By virtue of user-contributed contextual information, social media have produced many challenging research problems and many exciting real-world applications. As one of applications, image classification is to group images with similar objects that they contain, which is highly useful in everyday life. It is very easy for humans to identify complex natural images containing animals, buildings, forest, sun, river and etc. however, it is hard for computer to accurately and rapidly categorize images. To target this issue, some researchers explored a variety of supervised methods and unsupervised methods. Supervised learning consists of training and prediction phases. It requires a batch of training samples annotated by semantic labels to establish a learner of the satisfactory generalization capability. For learner, labeled data always seem very few,

F. Sun (✉) · Y. Xu · J. Zhou

School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, Liaoning, China

e-mail: sunwenfriend@hotmail.com

while unlabeled data may be abundant or publicly accessible. It is well known that annotating huge amount of unlabeled samples is a costly work. Therefore, how to obtain a good classifier without involving much more human labeling efforts has attracted much more attention of researchers.

Recently, active learning has been widely applied to machine translation [12], image retrieval [1], text classification [3], automatic multimedia tagging [15], etc. The reason is that it has the ability to automatically select the most informative samples for human labeling so that the labeling effort can be remarkably reduced. As one of the most representative methods, active learning support vector machine prefers to select the uncertain ones as the current classifier's training samples. In supervised learning, the initial training samples are given in advance. In this case, once a suitable model parameter is found, the parameter will be reused in the follow process of learning. However, the number of training samples continues to increase using active learning in the querying process. Correspondingly, the learning model should also be changed with respect to the data set. Unfortunately, it is often ignored or neglected. We address that dynamic model parameter are desirable to boost the performance of classification.

In this paper, we proposed an efficient algorithm, called active learning SVM with regularization path, in which we use unlabeled samples to choose the best regularization parameter. Although cross-validation is often adopted to find the proper regularization parameter for SVM model in supervised learning, it is not applicable to active learning [2]. The most likely reason is that the relationship between the queried samples and the original samples no longer meet the independent and identical distribution characteristics [16]. Note that the active learning algorithm requires a small number of samples for learning the initial model. To give reliable predictions on new unlabeled samples, the classifier must have a reasonable accuracy. Therefore, it is very important to make a decision which samples should be selected for labeling at initial stage [5]. In general, it does not guarantee that the samples in the initial training set, which are collected using the random sampling method, are the most representative ones [20, 21]. Therefore, we adopt an improved K-medoids cluster algorithm to choose the most representative samples as the initial training sets. To the best of our knowledge, this problem is rarely addressed in literatures.

In brief, our main contributions in this paper are summarized below.

- (1) We presented an efficient active learning SVM with regularization path, in which the model parameter is decided by using unlabeled samples. Active learning and SVM model are heavily correlated with each other. Therefore, it is necessary for SVM model to adjust the regularization parameter with the increase of the training samples.
- (2) We addressed the issue of constructing initial training sets by using an improved K-medoids cluster algorithm to choose the most representative samples. Using representative samples, it may significantly improve the performance of the classifier.

The rest of this paper is organized as follows. In Section 2, some related works are given. Besides, we briefly introduce active learning SVM and analyze it in Section 3. Furthermore, in Section 4, we present how to construct initial training set. Moreover, the description of active learning SVM regularization path algorithm is introduced in Section 5. In Section 6, we run experiments for empirical analysis. Finally, in Section 7 we make some conclusions and discussions for future work.

2 Related work

Formally, the typical active learning system contains two parts: sample selection and model learning. At initial stage, the learner first begins with a small number of labeled samples. Then,

unlabeled samples, which may be randomly chosen or active collected, are submitted for human labeling. Following this, the new labeled samples are added to the labeled sets. The two parts work alternately until they stop criteria. This process is shown in Fig. 1 (reader may also see Ref. [13]).

According to different ways of obtaining unlabeled samples, active learning methods are categorized into three types, which include member query learning, stream-based and pool-based methods, as shown in Fig. 2 (reader may also refer to Ref. [9]). In member query learning, the learner provides the samples for labeling. But in most cases, it is difficult for experts to annotate samples. In stream-based active learning (sometimes called sequential active learning), each unlabeled sample is obtained one at a time from data sets. Then the learner must determine whether or not to label it. In pool-based active learning, samples are selected from the pool for human labeling according to some certain criterions, which are used to evaluate samples in the pool. Since pool-based active learning has been widely and adequately studied in many applications, we focus on this method in this paper.

The central idea of active learning is to obtain satisfying performance with a small number of labeled training samples by choosing the important and representative ones. Obviously, the key of active learning is the sampling selection strategy. In many real-world situations, there are many unnecessary and redundant samples. So how to use sampling selection strategy to choose the important samples becomes a crucial issue. Many approaches have been proposed for sample selection. Zhu et al. [19] combined expected Error Reduction with graph-based semi-supervised learning strategy, resulting in a better performance than random or uncertainty sampling. Although uncertainty sampling is prone to select uncertain samples for current classifier, this approach may lead to selecting outliers without considering the underlying distribution of unlabeled samples. To tackle the outlier problem effectively, Zhu et al. also proposed a density-based re-ranking method [20]. In addition, there are other sample selecting strategies that consider the sample's diversity, density and relevant. As different performance and values are associated with special applications, it is difficult to directly compare the pros and cons of these sample selection strategies. In most cases, it is rational to combine these strategies together. Yuan et al. [17] adopted a combination of uncertainty and diversity sampling strategy in content-based image retrieval, and meanwhile they also proposed to get more relevant samples by moving the boundary. Zha et al. [18] adopted diversity, relevance and density criteria in video indexing and showed its better performance. Tian et al. [11] proposed active ranking information based sample selection strategy to reduce the user's

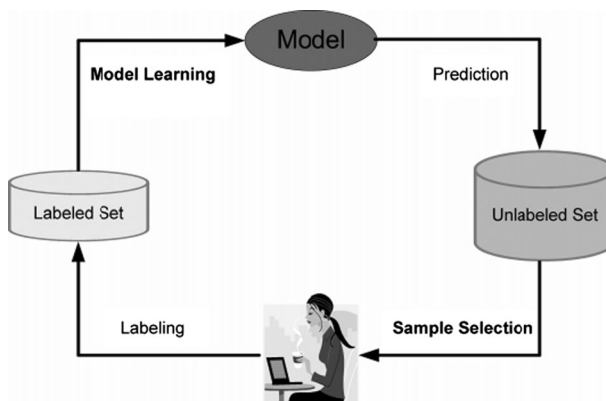


Fig. 1 An illustration of active learning

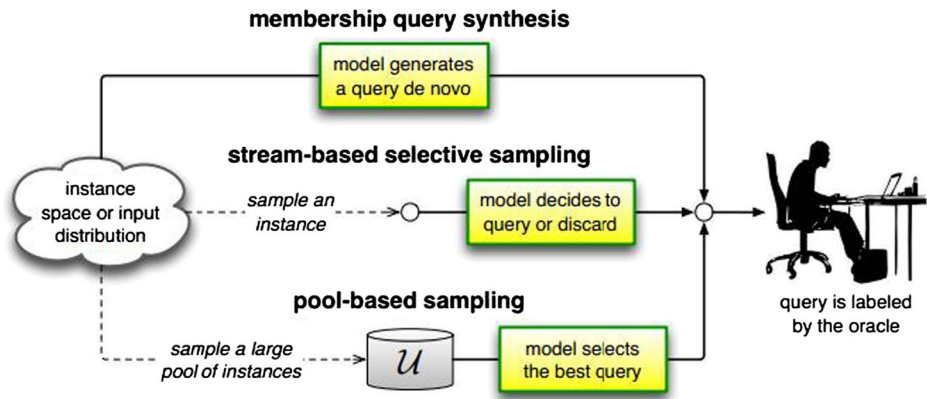


Fig. 2 Three methods of active learning

labeling efforts, and the idea of ranking information and low computational complexity are also reflected in [6, 7].

With different fields having its own characteristics, active learning has developed many novel algorithms, which differ from the traditional approach, such as semi-supervised learning with active learning, multi-label learning with active learning and incremental learning with active learning. These novel algorithms not only retain the advantage of the traditional method but make full use of other mature learning methods, thereby making active learning be a good solution to deal with various practical problems.

Both active learning and semi-supervised learning involve unlabeled samples. It is therefore natural to think about combining them with each other to enhance the sampling strategy so that much more help can provide to the learner. Presently, the new approach has been applied in various fields [14]. A direct approach to deal with multi-label active learning is to translate multiple labels into a set of independent binary problems [4]. In other words, a binary active learning algorithm is adopted to solve each label independently. But this approach does not take into account the correlations between the labels due to consider each label to be independent. Strong relationship usually occurs in multiple labels. To solve this problem, Qi et al. [8] proposed a two-dimensional active learning strategy, which selects sample-label pairs for labeling. For the selected samples, it is reasonable to label just several special concepts, and the others can be inferred through the relationship among multiple labels. For example, “tiger” and “grass”, these two labels often occur simultaneously, so labeling one of them will remarkably reduce the uncertainty of the other label. Multi-instance active learning (MIAL) is a novel approach that studies active learning algorithms based on multi-instance framework. A direct way to deal with MIAL problem is to adopt the existing criteria of supervised learning. Settles et al. [10] proposed to label samples in positive bags and meanwhile adopted uncertainty and expected gradient length sampling to select samples in positive bags. In MIAL, some bags are labeled and some are not, so the learner can query for labels of unlabeled bags, selected instances from positive bags and selected positive bags.

Up to now, there are still remaining problem in active learning to be solved, the description is summarized as following:

- 1) Usually labeling different samples with different concepts even with the same ones may be cost different efforts, so it is worthwhile to consider how the active learning method choose samples or annotators.

- 2) The efforts of labeling samples may be different due to the degree of recognizing objects in samples. Under this occasion, how to select samples with high recognizability is still to be handled.
- 3) In order to achieve a trade-off between labeling cost and computational efficiency, how many samples should be selected in iteration for labeling is another problem to be solved.

In some special fields, there are many samples with missing category and attribute, so it is another challenge to respond.

3 Active learning and SVM

3.1 SVM

For classification problems, the central idea of the soft margin SVM is to find the maximum margin hyperplane $f(x)=w^T\varphi(x)+b$ by solving optimization problem:

$$\begin{cases} \min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ y_i(w^T \cdot \varphi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, 0 < i < n \end{cases} \quad (1)$$

where C is a trade-off parameter that controls the amount of overlap, ξ_i is a slack variable allowing points to be on the wrong side of the margin, and $\varphi(x)$ is the mapping function. The above optimization problem can be formulated using the Loss and Penalty criterion:

$$\min \sum_{i=1}^n L(y_i f(x_i)) + \frac{\lambda}{2} \|w\|^2 \quad (2)$$

Here, regularization parameter λ corresponds to $\frac{1}{C}$, hinge loss function $L(y_i f(x_i)) = \max(1 - y_i f(x_i))$. The optimal solution of SVM can be expressed as

$$f^*(x) = \frac{1}{\lambda} \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b \quad (3)$$

In SVM, seeking the best classifier model can be converted into a regularized parameter optimization problem. The formulation (2) emphasizes the important role of regularization. In essence, model selection can be formulated as selection of a best value of λ . Selection of the regularization parameter is vital important in training a satisfying classifier model. That is to say, we should dynamically adjust the regularization parameter to guarantee a satisfactory learning result according to the current training dataset.

The role of regularization (reader may see Ref. [2]) is illustrated in Fig. 3. From Fig. 3, we can see the results of two SVM models on the same dataset. The radial kernel function $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ is used with $\gamma = 1$. The model with $C = 10000$ is less regularized than the left model with $C = 2$, and suffers a larger test error.

3.2 Active learning SVM

In active learning SVM, the goal of the active learner is to select the support vector, which is the most informative sample in unlabeled samples pool. The most informative and uncertain sample can be selected by

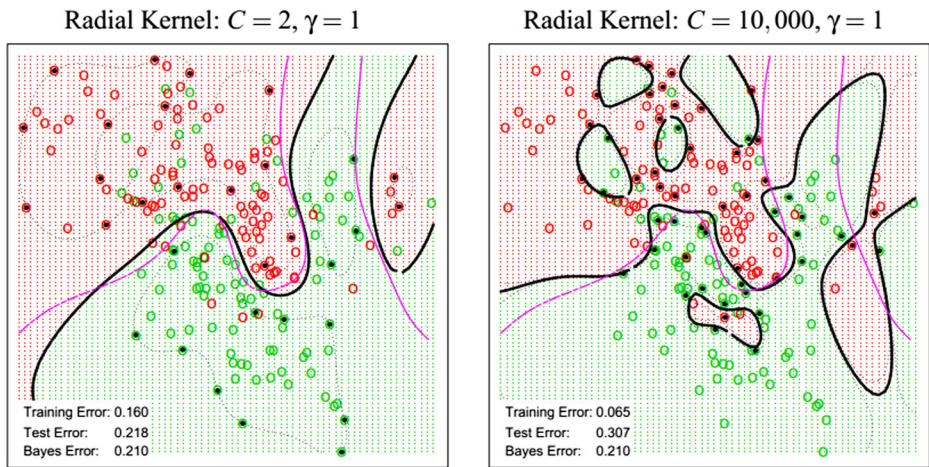


Fig. 3 Illustration of the need for regularization

$$i = \arg \min_{i \in U} |f^*(x_i)| \tag{4}$$

where $f^*(x_i)$ is decision function (namely classifier) in item (3). After a certain query, most of the support vectors can be selected. The role of training samples is to find the optimal classifier, which is expected to make accurate judgment on unlabeled samples. Under this circumstance, active learning and SVM model are heavily correlated with each other. Therefore, it is necessary for SVM model to adjust the regularization parameter with the increase of the training samples. SVM model selection problems are mainly affected by two factors: the training samples and the regularization parameter. Undesirable regularization parameter will lead to querying useless samples, which will affect classifier update and mislead next query for new samples.

3.3 Regularization path algorithm

Hastie et al. proposed a method called SVMpath algorithm to efficiently adjust SVM regularization parameter [2]. For every regularization parameter, we can trace all paths of SVM solution by SVMpath, with the same computational complexity as tracing one SVM model.

The optimization problem of the soft margin SVM in item (1) can be formulated within the regularization framework as item (2). Then based on (2), we construct the Lagrange primal function:

$$\begin{aligned}
 L_P &= \sum_{i=1}^n \xi_i + \frac{\lambda}{2} \langle \mathbf{w}, \mathbf{w} \rangle + \sum_{i=1}^n \alpha_i (1 - y_i f(\mathbf{x}_i) - \xi_i) - \sum_{i=1}^n \mu_i \xi_i \\
 \frac{\partial}{\partial \mathbf{w}} : \mathbf{w} &= \frac{1}{\lambda} \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i \\
 \frac{\partial}{\partial \xi_i} : \alpha_i &= 1 - \mu_i \\
 \frac{\partial}{\partial w_0} : \sum_{i=1}^n y_i \alpha_i &= 0
 \end{aligned} \tag{5}$$

along with the KKT conditions

$$\begin{aligned} \alpha_i [1 - y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) - \xi_i] &= 0 \quad i = 1..n \\ \mu_i \xi_i &= (1 - \alpha_i) \xi_i = 0 \quad i = 1..n \end{aligned} \tag{6}$$

According to different values of $y_i f(\mathbf{x}_i)$ (see Fig. 4 in Ref. [16]), we can define the following three sets:

$$E = \{i : y_i f(\mathbf{x}_i) = 1, 0 \leq \alpha_i \leq 1\}$$

$$L = \{i : y_i f(\mathbf{x}_i) < 1, \alpha_i = 1\}$$

$$R = \{i : y_i f(\mathbf{x}_i) > 1, \alpha_i = 0\}$$

Regularization paths algorithm takes advantage of the piecewise linear relationship between Lagrange Factor α and the regularization parameter λ . In this algorithm, we attempt to construct the entire regularization path after analyzing the change of the set E , L and R . The λ is updated when points move from L or R to E and vice versa. At initial stage, we set $\lambda = \lambda^{\max}$, and with change of the points, λ is estimated recursively until the stop criterion.

3.4 How to select the model parameter

As mentioned above, model selection is quite important for active learning. In supervised learning, cross-validation is often adopted to find the most suitable parameter for SVM model. Commonly, in cross-validation, we divide the labeled samples into a training set and a validation set. The training set is utilized to construct the SVM classifier and the validation set to select the most suitable model parameter.

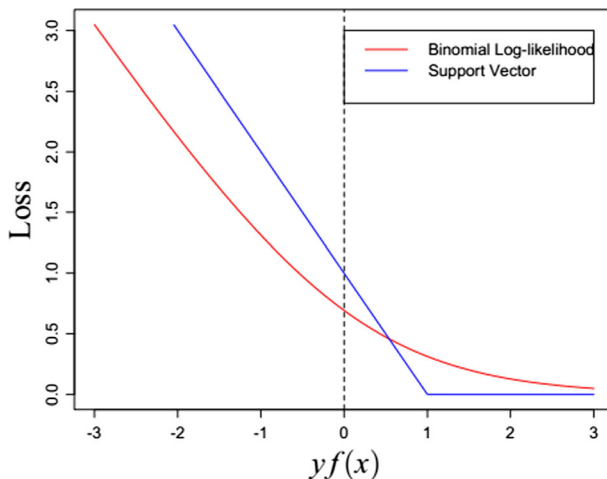


Fig. 4 Hinge loss and negative binomial log-likelihood

However, this issue is not suitable to active learning. The reason is that the query samples are not independently and identically distributed when selected from the original data distribution [16]. If we use such samples as validation set, it will result in severe over-fitting and affect the following query process. In active learning process, the labeled samples are scarce while the unlabeled samples are enriched. Therefore, we use unlabeled samples to build the validation set. Then the model selection criterion can be given as follows:

$$\lambda^* = \operatorname{argmax}_{\lambda \in \Lambda} \sum_{i=1}^u \max(0, 1 - |f(x_i)|) \tag{7}$$

where Λ is the candidate set of the model parameter, u represents the number of unlabeled samples, and $|f(x_i)|$ denotes the margin.

4 Construct the initial training set

A small number of samples are required to train the initial model in active learning algorithm. This is because the classifier needs to have a reasonable accuracy in order to give reliable predictions on new unlabeled samples and decide which samples ought to be labeled [5]. Generally, the initial training set is generated by random sampling. However, it cannot guarantee that the selected sample is more representative, while using representative samples may significantly improve the performance of the classifier. Therefore we take advantage of K-medoids cluster algorithm to collect training samples utilized to construct the initial training set.

The procedure of the K-medoids cluster algorithm can be described as follows:

Algorithm 1 K-medoids cluster algorithm

input Number of cluster

1. Choose k samples randomly as cluster centers $\varphi_j (j=1, 2, \dots, k)$ for initial clusters $\psi_j (j=1, 2, \dots, k)$
- repeat**
2. Choose the next sample and calculate the distance of the selected sample to k cluster centers by using the distance measure. Here, take Euclidean distance measure. Then distribute the samples $U = \{x_1, x_2, \dots, x_n\}$ into k cluster ψ_j , where $\psi_j = \{x_i | d(x_i, \phi_j) < d(x_i, \phi_{l \neq j})\}$
3. Re-calculate the cluster center φ_j for each cluster ψ_j according to the following item:

$$\varphi_j = \frac{1}{m} \sum_{i=1, x_i \in \psi_j}^m x_i, \text{ then replace the cluster center } \varphi_j \text{ with}$$

$$\phi_j' = \{x_i | d(x_i, \phi_j) < d(x_{l \neq i}, \phi_j)\}$$

until $\varphi_j \neq \phi_j'$

output Cluster result

Although the K-medoids cluster algorithm has been applied to many areas, it has some shortcomings. Adjacent samples can be easily selected as initial k cluster centers. In this situation, it requires much iteration to achieve high accurate cluster, reducing the efficiency of the clustering algorithm. We give an example to illustrate the above issue in Fig. 5.

Let $k=2$, and select two samples (A and B) randomly as the initial cluster centers. Then perform the K-medoids cluster algorithm. The results are as follows:

- (1) Cluster centers: $\{A, B\}$; cluster result: $\{A\}$ and $\{B, C, D, E, F, G\}$; new cluster centers: $\{A, D\}$
- (2) Cluster centers: $\{A, D\}$; cluster result: $\{A, B, C\}$ and $\{D, E, F, G\}$; new cluster centers: $\{B, F\}$
- (3) Cluster centers: $\{B, F\}$; cluster result: $\{A, B, C\}$ and $\{D, E, F, G\}$; new cluster centers: $\{B, F\}$
- (4) No change in cluster center and the algorithm stops. The cluster result is: $\{A, B, C\}$ and $\{D, E, F, G\}$

If we select sample A and G (sample G is the farthest one to sample A among all samples) as the initial cluster centers. The process of cluster is detailed as follows:

- (1) Cluster centers: $\{A, G\}$; cluster result: $\{A, B, C\}$ and $\{D, E, F, G\}$; new cluster centers: $\{B, F\}$
- (2) Cluster centers: $\{B, F\}$; cluster result: $\{A, B, C\}$ and $\{D, E, F, G\}$; new cluster centers: $\{B, F\}$
- (3) No change in cluster center and the algorithm stops. The cluster result is: $\{A, B, C\}$ and $\{D, E, F, G\}$

The computational complexity of the K-medoids cluster algorithm is $o(ndkt)$, where n denotes the number of samples, d is the number of features, k is the number of the clusters and t is the number of the iterations. From the above analysis, the second occasion requires a few iterations and reduces the computational complexity. Therefore the latter is adopted in this paper.

5 The description of active learning SVM regularization path algorithm

In order to dynamically adjust the regularization parameter to guarantee a satisfactory learning result, we propose an active learning SVM algorithm with regularization path that can fit the entire solution path of SVM for every value of model parameters. This algorithm traces the entire solution path of the current classifier and subsequent finds a series of candidate model parameters, then uses unlabeled samples to select the best model parameter.

Based on the analysis given in Section 3, we present the process of selecting regularization parameter and training model in Fig. 6, and we also summarize the active learning SVM regularization path algorithm as follows.

Algorithm 2 active learning SVM regularization path algorithm
repeat

1. Use K-medoids cluster algorithm to select m samples for constructing initial training set $S = \{x_1, x_2, \dots, x_m\}$.
2. Use labeled sample set $S = \{x_1, x_2, \dots, x_m\}$ to train SVM classifier and construct SVM regularization path for every candidate parameter λ .

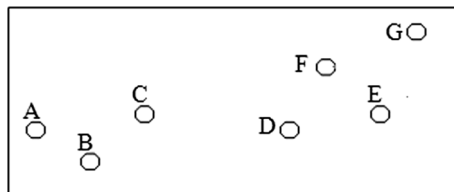


Fig. 5 Data points for cluster

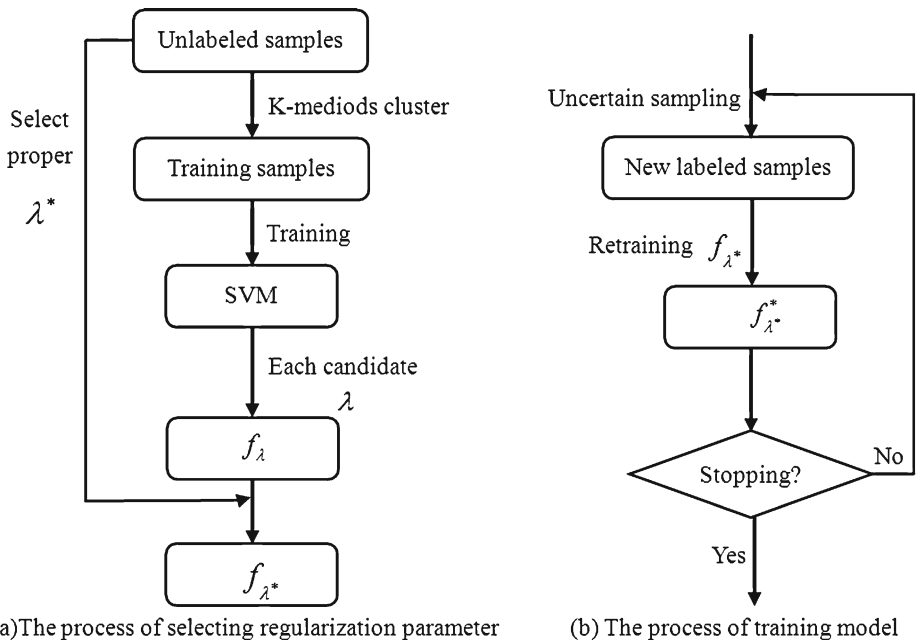


Fig. 6 The process of the active learning SVM regularization path algorithm. a. The process of selecting the regularization parameter. b. The process of training model

3. Use the unlabeled samples to select the most suitable regularization parameter utilizing Eq. 7, and train the current classifier.
4. Based on the current classifier, use Eq. 4 to select the most uncertain samples to re-train the current classifier.

until Stop criterion is met.

6 Experimental results

In this section, we first compare the active learning SVM regularization path algorithm with active learning SVM. And we also give results on random query (RSVM). Meanwhile uncertain sampling is implemented to query samples. Besides, at initial stage, K-medoids cluster algorithm is used to select samples for constructing training set. Then, we compare the K-medoids cluster sampling with random sampling at initial stage in order to evaluate our proposed K-medoids cluster algorithm.

Table 1 UCI benchmark

Data set	Instance	Attribute	Class
Heart	270	13	2
Pima	768	8	2
Ionosphere	351	34	2
Australian	690	14	2

Table 2 NUS-WIDE

Data set	Instance	Attribute	Class
A-G	400	73	2
S-H	400	144	2
G-P	400	64	2
M-W	400	73	2

6.1 Experiment setting

We use four data sets from the UCI benchmark to evaluate the SVM regularization path algorithm and the K-medoids cluster algorithm. Meanwhile, a real-world web image database from National University of Singapore (NUS-WIDE) is also adopted to evaluate the SVM regularization path algorithm.

The NUS-WIDE database includes 269,648 images and the associated tags, and six types of low-level features extracted from these image. In our experiments, we choose three types of features, including 64-D color histogram, 144-D color correlogram and 73-D edge direction histogram. NUS-WIDE contains three lite versions (including NUS-WIDE-LITE, NUS-WIDE-OBJECT and NUS-WIDE-SCENE) and each version cover a general sub-set of the whole dataset. We only focus on the NUS-WIDE-SCENE (covering 33 scene concepts). In our experiments, we use eight concepts for training, including airport, grass, sunset, mountain, window, house, plant, and glacier. The eight concepts consist of four datasets, including airport-glass (A-G), sunset-house(S-H), mountain-window (M-W) and glacier- plants (G-P).

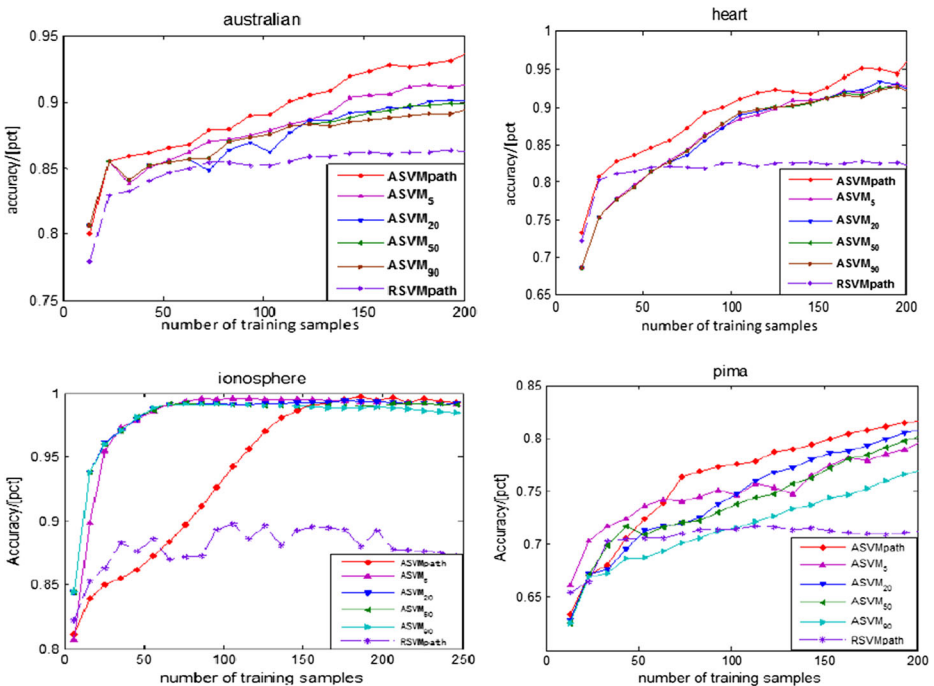


Fig. 7 The learning accuracy curves

The details of the data sets are listed in Table 1 for UCI benchmark and Table 2 for NUS-WIDE data sets. All algorithms run 10 times for each dataset, and then the average results are given.

In classification problems, a certain number of labeled samples should be used to construct a base classifier. So in this experiment, training samples are selected by using K-medoids to train a classifier. Here, we set $K=2$ (The number of clusters is equal to the classes of the dataset). In each iteration, 10 samples are selected (cluster center and its neighbor samples) as the initial labeled samples, and in each query, only one sample is chosen to be labeled. The process stops when there are a certain number of samples labeled.

In active learning SVM regularization path (ASVMpath), we use radius basis function as the kernel function. As for active learning SVM (ASVM), we provide proper values $C \in \{5, 20, 50, 90\}$ (on NUS-WIDE database, we set $C \in \{1, 5, 20, 50, 90\}$) for regularization parameter in advance.

In order to evaluate the effectiveness of the K-medoids, the cluster sampling is compared with random sampling. We first random select some labeled samples as the initial labeled samples, and other samples compose the unlabeled sample sets. Then random sampling and K-medoids cluster sampling are adopted to select the training samples. The process stops when there are 200 labeled samples.

6.2 Comparison with ASVM on UCI benchmarks

In this subsection, we compare the ASVM with our ASVMpath on UCI benchmarks. The learning accuracy curves during the active learning are shown in Fig. 7. From Fig. 7, we can see that ASVMpath performs better than ASVM in most cases. Meanwhile, the results show that the SVM regularization path algorithm is very effective to select the model parameter. But when we conduct experiments on the ionosphere data set, we see that ASVMpath performs

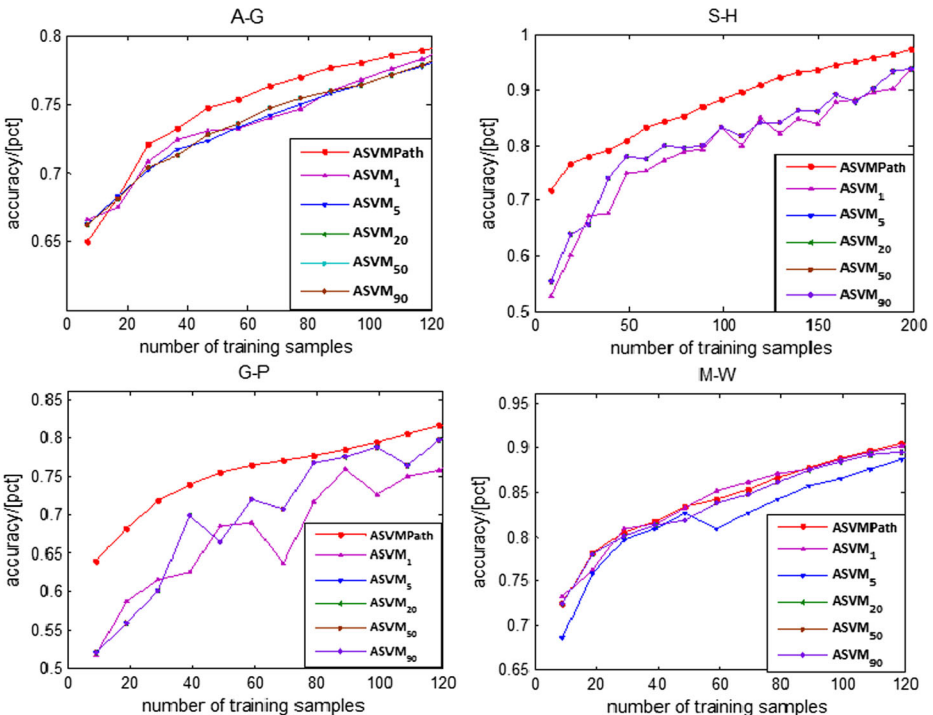


Fig. 8 The learning accuracy curves

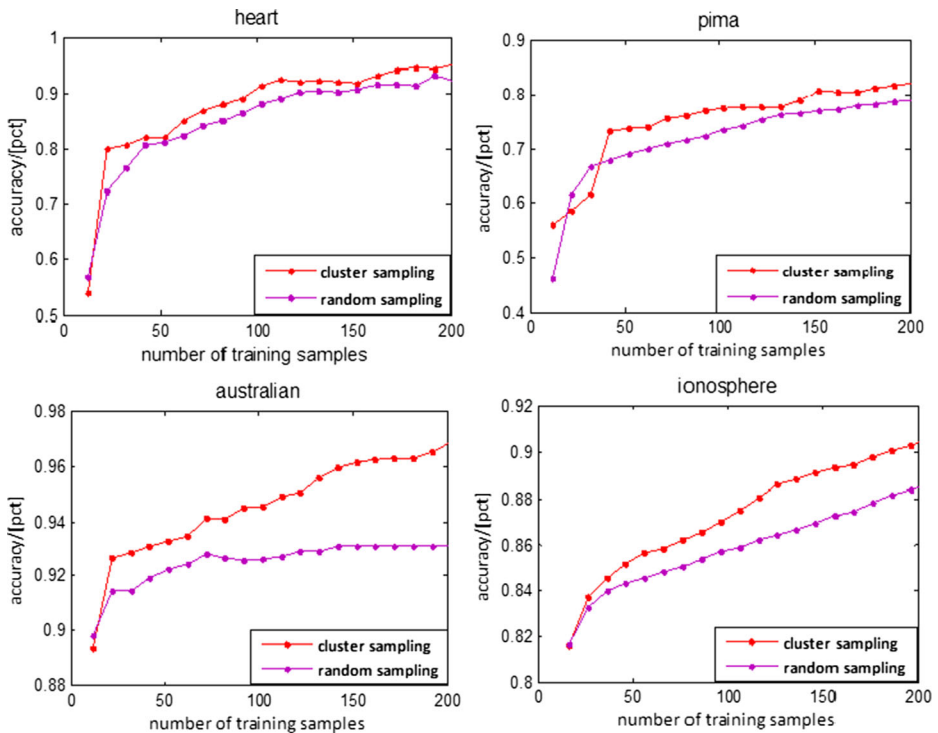


Fig. 9 The learning accuracy curves

worse than ASVM. The reason may be that it is not suitable to select the model parameter by using the unlabeled samples in the former algorithm.

6.3 Comparison with ASVM on NUS-WIDE

In this subsection, we compare the ASVM with our ASVMpath on NUS-WIDE database. The learning accuracy curves are illustrated in Fig. 8. We can see from Fig. 8 that ASVMpath performs better than ASVM in most cases. We also see that all methods get high accuracy with a small number of labeled samples. For the M-W (mountain-window) dataset, the two methods yield similar performance. The most likely reason is that the SVM methods are not very sensitive to the regularization parameter in this case. Note that some curves are not fully observed in Fig. 8 because the curves are covered by other curves.

6.4 Training samples selection from UCI benchmarks

In this subsection, we compare K-medoids cluster sampling with random sampling. Figure 9 compares the accuracy of the classifiers, which are obtained by using the training data collected by our K-medoids cluster method and by random selection method, respectively. K-medoids cluster method can clearly outperform random selection for most training data sets. But for pima dataset, the cluster sampling is less better random sampling. This may be due to the reason that the number of cluster predefined unreasonable.

7 Conclusions

In classification problems, model selection is a quite important problem. In this paper, regularization path algorithm is adopted to select the most suitable model parameter. At first, we find a series of candidate model parameters by tracing the entire solution path of the current classifier. Furthermore, we use the unlabeled samples to select the best model parameter. In addition, at initial stage of training, we select the training samples using the K-medoids cluster algorithm. Finally, we carry out experiments on UCI benchmark and NUS-WIDE dataset with the active learning SVM regularization path (ASVMpath) and the active learning SVM (ASVM) respectively, and then make a comparison between the two methods. The experiment results showed that the ASVMpath has an advantage over ASVM among most data sets. Besides, we also compare the K-medoids cluster sampling with random sampling, and the results show that K-medoids cluster method can clearly outperform random selection for most training data sets. In future work, we will focus on how to make full use of the misclassified samples to optimize the algorithm further.

Acknowledgement This work is supported by National Natural Science Foundation of China under Grant No. 61272214.

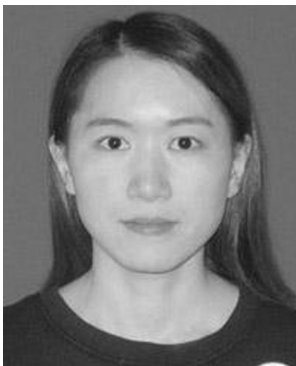
References

1. Gosselin PH, Cord M (2008) Active learning methods for interactive image retrieval. *IEEE Trans Image Process* 17(7):1200–1211
2. Hastie T, Rosset S, Tibshirani R, Zhu J (2004) The entire regularization path for the support vector machine. *J Mach Learn Res* 5:1391–1415
3. Hoi SC, Rong J, Lyu MR (2006) Large-scale text categorization by batch mode active learning. In: *Proceedings of the 15th International Conference on World Wide Web*. ACM 633–642
4. Li X, Wang L, Sung E (2004) Multi-label SVM active learning for image classification. In: *Proc Int Conf Image Process* 4:2207–2210
5. Lughofer E (2012) Hybrid active learning for reducing the annotation effort of operators in classification systems. *Pattern Recogn* 45(2):884–896
6. Pang Y, Ji Z, Jing P, Li X (2013) Ranking Graph embedding for learning to rerank. *IEEE Trans Neural Netw Learn Syst* 24(8):1292–1303
7. Pang Y, Zhang K, Yuan Y, Wang K. Distributed object detection with linear SVMs. *IEEE Transactions on Cybernetics*, in press. DOI: [10.1109/TCYB.2014.2301453](https://doi.org/10.1109/TCYB.2014.2301453)
8. Qi G-J, Hua X-S, Rui Y, Tang JH, Mei T, Zhang H-J (2007) Correlative multi-label video annotation. In: *Proceedings of ACM Multimedia*, pp. 17–26
9. Settles B (2010) Active learning literature survey. University of Wisconsin, Madison
10. Settles B, Craven M, Ray S (2007) Multiple-instance active learning. In: *Proc Neural Inf Process Syst* 20:1289–1296
11. Tian X, Tao D, Hua X-S, Wu X (2010) Active reranking for web image search. *IEEE Trans Image Process* 19(3):805–820
12. Tur G, Schapire RE, Hakkani-Tur D (2003) Active learning for spoken language understanding. In: *Proc IEEE Int Conf Acoust Speech Signal Process* 1:276–279
13. Wang M, Hua X-S (2011) Active learning in multimedia annotation and retrieval: A survey. *ACM Trans Intell Syst Technol (TIST)* 2(2):10
14. Wang M, Hua X-S, Mei T, Tang J, Qi G-J (2007) Interactive video annotation by multi-concept multimodality active learning. *Int J Semant Comput* 1(4):459–477
15. Wang M, Ni B, Hua X-S, Chua T-S (2012) Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Comput Surv (CSUR)* 44(4):25
16. Wang Z, Yan S, Zhang C (2011) Active learning with adaptive regularization. *Pattern Recogn* 44(10):2375–2383

17. Yuan J, Zhou X, Zhang J, Wang M, Zhang Q, Wang W, Shi B (2007) Positive sample enhanced angle-diversity active learning for SVM based image retrieval. In: IEEE International Conference on Multimedia and Expo 2202–2205
18. Zha Z-J, Wang M, Zheng Y-T, Yang Y, Hong R, Chua T.-S. Interactive video indexing with statistical active learning. *IEEE Trans Multimedia* 14:(1)17–2, 20127
19. Zhu X, Lafferty J, Ghahramani Z (2003) Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In: ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining 58–65
20. Zhu J, Wang H, Tsou BK, Ma M (2010) Active learning with sampling by uncertainty and density for data annotations. *IEEE Trans Audio Speech Lang Process* 18(6):1323–1331
21. Zhu J, Wang H, Yao T, Tsou BK (2008) Active learning with sampling by uncertainty and density for word sense disambiguation and text classification. In: Proc 22nd Int Conf Comput Linguist 1:1137–1144



Fuming Sun is currently a Professor of School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, China. He received his Ph.D. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2007. From September 2012 to July 2013, he worked as a Visiting Scholar in Department of Automation, Tsinghua University. His current research interests include content-based image retrieval, image content analysis, and pattern recognition.



Yan Xu received her B.S. degree in 2010 from Xingtian College, Qufu Normal University and received her Master degree from Liaoning University of Technology, China, in 2014.



Jun Zhou is currently a Professor of School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, China. She received her Ph.D. degree from the Northeastern University, Shenyang, China, in 2005. Her current research interests include data mining, pattern recognition and image processing.