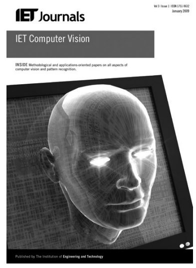
www.ietdl.org

Published in IET Computer Vision
Received on 5th June 2014
Revised on 18th August 2014
Accepted on 16th September 2014
doi: 10.1049/iet-cvi.2014.0140



ISSN 1751-9632

Active learning combining uncertainty and diversity for multi-class image classification

Yingjie Gu[1]*, Zhong Jin*[1]*, Steve C. Chiu*[2]

[1]*Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China*
[2]*Department of Electrical Engineering, Idaho State University, Pocatello 83209-8060, USA*
E-mail: csyjgu@gmail.com

Abstract: In computer vision and pattern recognition applications, there are usually a vast number of unlabelled data whereas the labelled data are very limited. Active learning is a kind of method that selects the most representative or informative examples for labelling and training; thus, the best prediction accuracy can be achieved. A novel active learning algorithm is proposed here based on one-versus-one strategy support vector machine (SVM) to solve multi-class image classification. A new uncertainty measure is proposed based on some binary SVM classifiers and some of the most uncertain examples are selected from SVM output. To ensure that the selected examples are diverse from each other, Gaussian kernel is adopted to measure the similarity between any two examples. From the previous selected examples, a batch of diverse and uncertain examples are selected by the dynamic programming method for labelling. The experimental results on two datasets demonstrate the effectiveness of the proposed algorithm.

1 Introduction

Image classification is a significant problem in computer vision and pattern recognition. Many supervised learning algorithms have been proposed to solve this problem. However, the classification results rely heavily on the quality of the labelled data. In real-world applications, there are a large number of unlabelled data, whereas the labelled data are expensive to obtain. Moreover, redundant data in the training set slow down the training process without improving prediction accuracy. To improve classification performance, informative or representative examples should be used for learning and the redundant examples must be removed from the training set. Active learning [1, 2] is a kind of approach that selects the most informative examples for labelling and training a classifier.

The most important problem in active learning is how to select the most valuable examples so that the maximum prediction accuracy can be achieved. There are some criteria that have been proposed to direct example selection. Uncertainty sampling, which queries the examples whose prediction labels are most uncertain, is the most popular criterion. The typical uncertainty sampling is support vector machine (SVM)-based margin sampling [3] that selects the examples nearest to the hyperplane. Moreover, criteria such as variance reduction [4], diversity [5] and optimal experimental design [6] have also been widely explored in active learning.

According to the number of examples selected at each time, active learning can be classified as single-mode active learning and batch-mode active learning. Traditional margin sampling is single-mode active learning, which selects only one example at each time. Now, more and more batch-mode active learning algorithms have been proposed, since they more efficiently improve the classifier's performance. Fu *et al.* [7] proposed an active learning algorithm to find an instance subset with a maximum utility value. To achieve this goal, the following are simultaneously considered: (i) the importance of individual instances and (ii) the disparity between instances, to build an instance-correlation matrix.

Many active learning methods [3, 8] were proposed to solve binary classification problems. However, multi-class classification is a more practical and significant problem in real-world applications. Demir *et al.* [5] proposed an active learning method with one-versus-all (OVA) strategy SVM, to perform classification of remote-sensing images. Both uncertainty and diversity criteria were combined in the example selection process. Joshi *et al.* [9] developed a value-of-information algorithm that chooses informative examples while also considering users' annotation cost.

Recently, Kang and Xu [10] proposed an active learning method combining uncertainty and diverse criteria. They first selected h most uncertain examples. Then, k most diverse examples are selected from the h examples. These two criteria are used in two phases of example selection, respectively, while they are not combined together. Guo and Schuurmans [11] proposed discriminative active learning that can simultaneously maximise the likelihood of labelled instances and minimise the entropy of the labels for unlabelled instances. However, this algorithm was only applied for two-class classification problems and the computational complexity is very high. Ebert *et al.* [12] combined exploration and exploitation for example selection. Specifically, two uncertainty measure methods and three diversity measure methods were explored and compared. The combination method only computed the sum of criteria scores.

Although these works explored more than one criterion in active learning, the criteria's combination methods are usually too simple, for example, computing the sum or product of the criteria's scores. Several active learning algorithms based on OVA strategy SVM [5] have been developed, while there is little work based on one-versus-one (OVO) SVM [13]. In this paper, a novel batch-mode active learning algorithm is proposed, which is based on OVO SVM and combines uncertainty and diversity criteria. Inspired by [5, 7, 13], we use dynamic programming to combine different criteria in active learning. The example selection process can be divided into two phases: the first phase is selecting a relative large batch of uncertain examples; the other phase is selecting a batch of uncertain and diverse examples from the examples previously selected. This is the first work applying dynamic programming to the combination of different criteria.

The rest of this paper is organised as follows: a short introduction of related work is provided in Section 2. Section 3 elaborates the proposed active learning approach. The experiment settings and results of image classification are presented in Section 4. Section 5 discusses the conclusion and future work.

2 Related work

In this section, we introduce some work which is related to our proposed approach. In Section 2.1, a brief introduction of the general active learning problem is presented. An active learning algorithm based on OVO SVM is introduced in Section 2.2.

2.1 Active learning problem

Given a labelled dataset $\mathcal{L} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$, where each $\boldsymbol{x}_i$ is an instance of d-dimensional feature vector and has a label $y_i \in \{1, 2, \ldots, C\}$. The unlabelled dataset is $\mathcal{U} = \{\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, \ldots, \boldsymbol{x}_{n+m}\}$, where the label of $\boldsymbol{x}_i (i = n+1, \ldots, n+m)$ is unknown and $n \ll m$. Generally, a model can be trained from the initial labelled dataset $\mathcal{L}$ with classifier $\mathcal{C}$. The task of active learning is to find a subset $\mathcal{Z} = \{\boldsymbol{x}_{s_1}, \boldsymbol{x}_{s_2}, \ldots, \boldsymbol{x}_{s_k}\} \subseteq \mathcal{U}$ to improve the classification performance most on testing set $\mathcal{D}_{\text{test}}$. The process of example selection usually repeats several times until the number of selected examples or required accuracy is reached. The process of the active learning method is shown in Fig. 1.

Require:

$\mathcal{L}$: Initial labeled data set

$\mathcal{U}$: Initial unlabeled data set

Repeat:

Train a model $f(\mathcal{L})$ from $\mathcal{L}$ with classifier $\mathcal{C}$

Calculate $f(\mathcal{U})$

Select $k (k \geq 1)$ examples $\mathcal{Z} = \{\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \ldots, \mathbf{x}_{s_k}\}$ from $\mathcal{U}$ with some criteria

Query and label $\mathcal{Z}$

$\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{Z}$

$\mathcal{U} \leftarrow \mathcal{U} - \mathcal{Z}$

Until the number of selected points or the required accuracy is reached

Fig. 1 *General process of active learning*

2.2 Multi-class active learning with OVO SVM

The OVO strategy for multi-class SVM is computationally efficient and shows good classification performance. Recently, Joshi *et al.* [13] proposed two active learning algorithms called entropy measure (EP) and best-versus-second best (BvSB) based on OVO SVM.

2.2.1 Entropy measure (EP): Entropy is an effective method to measure examples' uncertainty and relies on probability estimates of class membership for all the examples. For the multi-class case, Joshi *et al.* adopted the pairwise coupling method [14] to obtain probability estimates.

Suppose that the probability of unlabelled example $\boldsymbol{x}_i$ belongs to class j $(j = 1, \ldots, C)$ is p_j. The discrete entropy of $\boldsymbol{x}_i$ can be computed as

$$H(\boldsymbol{x}_i) = -\sum_{j=1}^{C} p_j \times \log(p_j) \tag{1}$$

If the entropy of an example is large, the classifier is uncertain about its label prediction.

At each round of active learning, examples with the highest estimated value of discrete entropy are selected to query the user. The entropy-based active learning outperforms random selection in some cases.

2.2.2 Best-versus-second best (BvSB): In EP-based active learning, the entropy value is closely related to the probability values of unimportant classes. Instead of using the entropy score, Joshi *et al.* proposed a more greedy approach to measure examples' uncertainty. Specifically, the difference between the probability values of the two classes having the highest estimated probability value is computed as a measure of uncertainty. This method is called the best-versus-second-best (BvSB) approach and is a more direct way to estimate the uncertainty of the prediction results.

Both EP and BvSB methods are based on the uncertainty criterion. They estimate the uncertainty of the classification results with different methods while other criteria of active learning are neglected. In Section 3, a new method combining uncertainty and diversity criteria is proposed for example selection.

3 Proposed active learning algorithm

Our proposed active learning is based on uncertainty and diversity criteria. The example selection process can be divided into two phases: one is selecting h most uncertain examples from all unlabelled examples; the other is selecting k diverse and uncertain examples using dynamic programming from h uncertain examples. In Sections 3.1 and 3.2, the method to estimate examples' uncertainty and diversity is introduced. The proposed algorithm is described in Section 3.3. Section 3.4 introduces a dynamic programming approach to solve the optimisation problem.

doi: 10.1049/iet-cvi.2014.0140

3.1 Uncertainty measure

Original SVM is a classifier for binary classification problems. However, it can also be applied to multi-class problem by OVO strategy. Assume that the labelled dataset is $\mathcal{L} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ while the unlabelled dataset is $\mathcal{U} = \{\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, \ldots, \boldsymbol{x}_{n+m}\}$, and $n \ll m$. The label of $\boldsymbol{x}_i (i = 1, \ldots, n)$ is $y_i \in \{1, 2, \ldots, C\}$. In OVO strategy, a binary classifier is created between any two classes, so that there are $C(C-1)/2$ binary classifiers. Suppose $f_{ij}(*)$ $(i = 1, \ldots, C-1, j = i+1, \ldots, C)$ is the decision function to classify examples from class i and class j, $f_{ji}(*)$ can be defined as $f_{ji}(*) = -f_{ij}(*)$. For an example x, $f_{ij}(\boldsymbol{x}) > 0$ means classifier $f_{ij}(*)$ prefers to predict $\boldsymbol{x}$ as label i. Larger $f_{ij}(\boldsymbol{x})$ indicates that the classification result is more certain.

The final label of $\boldsymbol{x}$ is predicted as the class which gets the maximum votes. If we define

$$v_{ij} = \begin{cases} 1, & \text{if } f_{ij}(\boldsymbol{x}) \geq 0 \\ 0, & \text{else} \end{cases} \tag{2}$$

The final label of $\boldsymbol{x}$ is predicted as

$$p = \underset{i=1,\ldots,C}{\operatorname{argmax}} \sum_{j=1}^{C} v_{ij} \tag{3}$$

To estimate the uncertainty of classification results on unlabelled data, we compute the difference between the decision values of the two classes having the max votes. As defined above, p is the class having the most votes. Suppose q is the class having the second most votes. Then

$$q = \underset{i=1,\ldots,C, i\neq p}{\operatorname{argmax}} \sum_{j=1}^{C} v_{ij} \tag{4}$$

The uncertainty of prediction on unlabelled example $\boldsymbol{x}$ can be measured as

$$\operatorname{unc}(\boldsymbol{x}) = \sum_{j=1}^{C} f_{pj}(\boldsymbol{x}) - \sum_{j=1}^{C} f_{qj}(\boldsymbol{x}) \tag{5}$$

Smaller $\operatorname{unc}(\boldsymbol{x})$ means that the prediction result on $\boldsymbol{x}$ is more uncertain. This uncertainty measure shares a similar idea with BvSB method. The difference from BvSB is that this method only selects uncertain examples for manual labelling, so we select h most uncertain examples for further selection.

3.2 Diversity measure

Uncertainty sampling is a strong criterion for example selection in active learning, but it has a drawback. It may select some uncertain examples that have much redundant information. For example, some selected examples are very close and similar to each other. Thus, the overall information of the selected examples is not much.

In this section, the Gaussian kernel is adopted as a similarity measure between two examples. The similarity of examples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined as

$$S(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right) \tag{6}$$

As can be seen from (6), the similarity $S(\boldsymbol{x}_i, \boldsymbol{x}_j)$ between two examples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is small if these two examples are far from each other, and vice versa.

Suppose a batch of examples $\mathcal{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_h\}$ have been selected. If we hope that the new example $\boldsymbol{x}$ has small similarity with examples in $\mathcal{S}$, we only need to make sure that the value of $\max_{i=1,\ldots,h} S(x, s_i)$ is small. Therefore, the similarity between a new example $\boldsymbol{x}$ and the selected examples S is defined as

$$\operatorname{sim}(\boldsymbol{x}, \mathcal{S}) = \max_{\boldsymbol{s}_j \in \mathcal{S}} \mathcal{S}(\boldsymbol{x}, \boldsymbol{s}_j) \tag{7}$$

Smaller $\operatorname{sim}(\boldsymbol{x}, \mathcal{S})$ means less similarity between $\boldsymbol{x}$ and $\mathcal{S}$.

3.3 Proposed algorithm

In active learning, we aim to select uncertain examples that are also diverse from each other. First, h most uncertain examples are selected from all the unlabelled examples. Then we try to select k examples that are most diverse and uncertain from the selected h examples. Suppose the h most uncertain examples are $\mathcal{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_h\} \subseteq \mathcal{U}$. $\mathcal{Z}_k$ is an arbitrary subset of S that contains k examples and $\mathcal{Z}_k = \{\boldsymbol{s}_{z_1}, \ldots, \boldsymbol{s}_{z_k}\}$. The final solution $\mathcal{Z}$ can be obtained by solving the following problem

$$\mathcal{Z} = \underset{\mathcal{Z}_k \subseteq \mathcal{S}}{\operatorname{argmin}} \sum_{i=1}^{k} \left(\lambda \operatorname{unc}\left(\boldsymbol{s}_{z_i}\right) + (1-\lambda) \operatorname{sim}\left(\boldsymbol{s}_{z_i}, \mathcal{Z}_k\right)\right) \tag{8}$$

where λ is the tradeoff parameter that can determine the importance of uncertainty and diversity. By solving (8), the examples with most uncertainty and diversity can be selected for active learning.

Unfortunately, the optimisation problem (8) is a highly complicated problem. To obtain the optimal subset $\mathcal{Z}$, we would have to search over all possible sets. It is impossible to finish the process in a short time with the increase of examples.

It can be noted that $\operatorname{unc}(\boldsymbol{s}_{z_i})$ is only dependent on s_{z_i} while $\operatorname{sim}(\boldsymbol{s}_{z_i}, \mathcal{Z}_k)$ is relevant to $\{\boldsymbol{s}_{z_1}, \ldots, \boldsymbol{s}_{z_k}\}$. Suppose $\mathcal{S}_u = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_u\} (u \leq h)$, and $Z(u, v) (v \leq u)$ is the optimal solution of selecting v examples from $\mathcal{S}_u$. We transform problem (8) into a relatively simple form

$$\mathcal{Z} = \underset{\mathcal{Z}_k \subseteq \mathcal{S}}{\operatorname{argmin}} \sum_{i=1}^{k} (\lambda \operatorname{unc}(\boldsymbol{s}_{z_i}) + (1-\lambda) \operatorname{sim}(\boldsymbol{s}_{z_i}, Z(z_i - 1, i - 1))) \tag{9}$$

where $\operatorname{sim}(\boldsymbol{s}_{z_i}, Z(z_i - 1, i - 1))$ is the similarity between $\boldsymbol{s}_{z_i}$ and the selected $i-1$ examples from $S_{z_i - 1} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{z_i - 1}\}$.

Obviously, $\operatorname{sim}(\boldsymbol{s}_{z_i}, Z(z_i - 1, i - 1))$ is relevant to $\{z_1, \ldots, z_i\}$ but irrelevant to $\{z_i + 1, \ldots, z_k\}$. It means that when we select the ith example, it is required to be diverse from the selected examples $\{\boldsymbol{s}_{z_1}, \ldots, f\boldsymbol{s}_{z_{i-1}}\}$. This guarantees that the next selected example must be different from the previous selected examples and the global similarity is small.

3.4 Dynamic programming approach

The optimal $\mathcal{Z}$ in problem (9) can be obtained by dynamic programming that breaks it down into simpler subproblems.

doi: 10.1049/iet-cvi.2014.0140

As previously defined, $\mathcal{S}_u = \{s_1, s_2, \ldots, s_u\} (u \le h)$, and $Z(u, v)$ is the optimal solution of selecting v examples from $\mathcal{S}_u$.

Specifically, $\mathcal{Z} = Z(h, k)$. Now, we define $F(u, v)$ and $Z(u, v)$ $(u \ge v)$ as follows

$$F(u, v) = \min_{\mathcal{Z}_v \subseteq \mathcal{S}_u} \sum_{i=1}^{v} \Big(\lambda \mathrm{unc}\big(s_{z_i}\big) + (1-\lambda)\mathrm{sim}\big(s_{z_i}, Z(z_i - 1, i-1)\big)\Big) \quad (10)$$

$$Z(u, v) = \operatorname*{argmin}_{\mathcal{Z}_v \subseteq \mathcal{S}_u} \sum_{i=1}^{v} (\lambda \mathrm{unc}(s_{z_i}) + (1-\lambda)\mathrm{sim}(s_{z_i}, Z(z_i - 1, i-1))) \quad (11)$$

where $u \in \{1, 2, \ldots, h\}$, $v \in \{1, 2, \ldots, k\}$, and $u \ge v$.

Our final goal is to find $Z(h, k)$ that decides which k examples should be selected from the h most uncertain examples.

There are two special situations that should be noted: $v = 1$ and $u = v$. If $v = 1$, there is no redundancy since only one example is selected. Therefore, the example with maximum uncertainty should be selected. If $u = v$, obviously, all of the examples in $\mathcal{S}_u$ should be selected. So

$$F(u, v) = \begin{cases} \min\limits_{s_i \in \mathcal{S}_u} \mathrm{unc}(s_i) & \text{if } v = 1 \\ \sum_{i=1}^{u} \lambda \mathrm{unc}(s_i) + (1-\lambda)\mathrm{sim}(s_i, \mathcal{S}_{i-1}) & \text{if } u = v \end{cases} \quad (12)$$

Suppose we have already obtained the optimal solution of selecting $v-1$, and v examples from $\mathcal{S}_{u-1}$, now we consider how to select v examples from $\mathcal{S}_u$. If the prediction of example s_u is certain and s_u is similar with previous selected examples $Z(u-1, v-1)$, obviously we will not select s_u. Hence, the optimal solution of selecting v examples from $\mathcal{S}_u$ should be the same as selecting v examples from $\mathcal{S}_{u-1}$. On the contrary, if the prediction of s_u is uncertain and s_u is diverse from $Z(u-1, v-1)$, we prefer to select it for labelling. In this situation, since $Z(u-1, v-1)$ is the optimal solution of selecting $v-1$ examples from $\mathcal{S}_{u-1}$, the optimal solution of selecting v examples from $\mathcal{S}_u$ is $Z(u, v) = Z(u-1, v-1) \cup s_u$.

From the above analysis, the relationships between $F(u-1, v-1)$, $F(u-1, v)$ and $F(u, v)$ can be concluded as (13).

$$\begin{aligned} C(u) &= \lambda \mathrm{unc}(s_u) + (1-\lambda)\mathrm{sim}(s_u, Z(u-1, v-1)) \\ F(u, v) &= \min(F(u-1, v), F(u-1, v-1) + C(u)) \end{aligned} \quad (13)$$

Input:

Initial labeled data set $\mathcal{L} = \{\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_n\}$

Initial unlabeled data set $\mathcal{U} = \{\mathbf{x}_{n+1}, \mathbf{x}_{n+2}, .., \mathbf{x}_{n+m}\}$

the gaussian parameter (σ), the tradeoff parameter (λ)

the number of uncertain examples need to select (h), the number of final selected examples at each round (k)

the number of times of the example selection (t)

Procedure:

For j = 1 to t
  Train $C(C-1)/2$ binary classifiers using labeled data set $\mathcal{L}$
  Perform classification on unlabeled data set $\mathcal{U}$
  Compute uncertainty of each unlabeled example using (5), select h most uncertain examples $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_h\}$
  Compute similarity between any two examples in $\mathcal{S}$
  Initilise $\mathcal{S}_u = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_u\}$, $F = \mathbf{0}_{nk}$, $Z(u, v) = \emptyset$
  For $u = 1 : h$
    $F(u, 1) = \min\limits_{\mathbf{s}_i \in S_u} \; unc(\mathbf{s}_i)$
    $Z(u, 1) = \arg\min\limits_{\mathbf{s}_i \in S_u} \; unc(\mathbf{s}_i)$
  End
  For $u = 2 : h$
    For $v = 2 : k$
      $C(u) = \lambda unc(\mathbf{s}_u) + (1-\lambda) sim(\mathbf{s}_u, Z(u-1, v-1))$
      $F(u, v) = \min(F(u-1, v), F(u-1, v-1) + C(u))$
      $Z(u, v) = \begin{cases} Z(u-1, v) & if \;\; F(u, v) = F(u-1, v) \\ Z(u-1, v-1) \cup \mathbf{s}_u & else \end{cases}$
    End
  End
  $\mathcal{L} \leftarrow \mathcal{L} \cup Z(n, k)$
  $\mathcal{U} \leftarrow \mathcal{U} - Z(n, k)$
End

Fig. 2 *Active learning with uncertainty and diversity sampling*

doi: 10.1049/iet-cvi.2014.0140

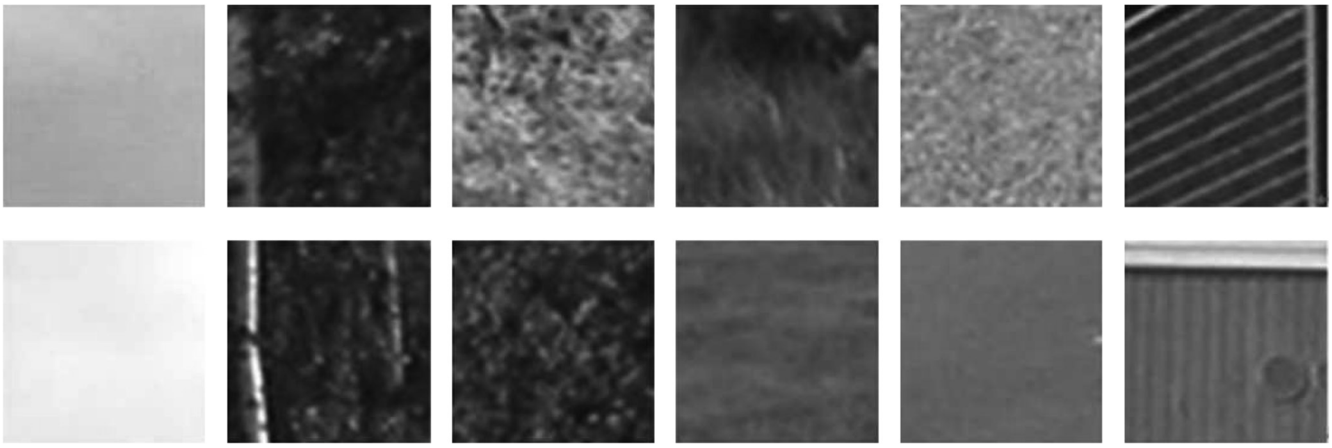


Fig. 3 *Examples of Outex from categories: sky, tree, bush, grass, road and building*

where $2 \le u \le h$, $2 \le v \le k$ and $v \le u$.

The global optimal solution $Z(h, k)$ can be obtained by iteration of (14).

$$Z(u, v) = \begin{cases} Z(u-1, v), & \text{if } F(u, v) = F(u-1, v) \\ Z(u-1, v-1) \cup \boldsymbol{s}_u, & \text{else} \end{cases} \quad (14)$$

In this way, the selected examples are uncertain and diverse from each other.

The proposed active learning algorithm is summarised in Fig. 2. As can be seen from Fig. 2, the proposed active learning algorithm is easy to perform and the computational cost is low.

4 Experiments

This section reports experimental results of the proposed active learning algorithm. To assess the effectiveness of the proposed active learning technique, it is compared with other methods:

- Random sampling (RS) method, which selects examples randomly from unlabelled dataset.
- EP method [13], which queries examples with maximum entropy.
- BvSB algorithm [13], which computes the difference between the probability values of the two classes having the highest estimated probability value as a measure of uncertainty. The most uncertain examples are selected for labelling.
- Multiclass-level uncertainty with angle-based diversity (MCLU-ABD) [5], which combines uncertainty and diversity criteria and uses cosine angle distance to measure the examples' similarity.
- Active learning combining uncertainty and diversity, which is proposed in this paper.

In this section, the experiments are carried out on terrain classification with the Outex Database [15] and scene recognition with scene 13 dataset [16]. Sections 4.1 and 4.2 introduce the settings and results of these two experiments, respectively. We apply OVO SVM with radial basis function kernel as the baseline classifier in our experiments. The optimal parameters C and γ are found by grid search on the parameter space.

There are some parameters in our algorithm, namely, the Gaussian kernel parameter (σ), the tradeoff parameter (λ) and the number of most uncertain examples selected each time (m). σ, λ and m are empirically set to 1, 0.6 and $10 \times k$, respectively. k is the number of final selected examples for active learning and is decided based on the size of dataset.

4.1 Terrain classification

The terrain image dataset used in the experiment was constructed from the Outex Database [15], which consists of two datasets: Outex-0 and Outex-1. Each of them is composed of 20 outdoor scene images and the image size is 2272×1704 pixels. The images are marked as one type of bush, grass, tree, sky, road and building. We cut the marked area of each image into 64×64-pixel patches and each patch is regarded as an example in the experiments. The examples in Outex-0 have six categories, including bush, grass, tree, sky, road and building. Outex-1 has only five categories, since it does not have the bush category.

In the experiments, 50 patches of each class are extracted to construct a dataset. Five examples from each class are randomly selected for initial labelling to construct $\mathcal{L}$. The rest is composed as unlabelled dataset $\mathcal{U}$. The testing dataset, which is predicted to evaluate active learning algorithms' performance, also includes 50 patches from each class. At each round, k examples from $\mathcal{U}$ are selected with different active learning methods for labelling and adding to $\mathcal{L}$. We set k to be 5 in Outex-0 and 3 in Outex-1. The classifiers are retrained from $\mathcal{L}$ and the classification is performed on testing dataset. The correct classification rate is used as accuracy to evaluate the performance of active learning methods.

Two examples of each class are shown in Fig. 3. To achieve better classification performance, both colour and texture features are extracted. For the colour feature, we extract the colour histogram feature proposed in [17]. For the texture feature, the popular rotation-invariant operators $\mathrm{LBP}_{8,1+16,3}^{\mathrm{riu2}}$ [18] are adopted. Lastly, each example is represented by a 43-dimensional feature vector.

In this experiment, k examples are selected at each round and the iteration will repeat 20 times. Thus, $20 \times k$ examples are selected in total. At each round, $C(C-1)/2$ binary classifiers are trained from the labelled set. We perform classification on the testing set and compute the classification accuracy. The experiments are repeated 20 times and the average accuracy is computed as the final result.

doi: 10.1049/iet-cvi.2014.0140

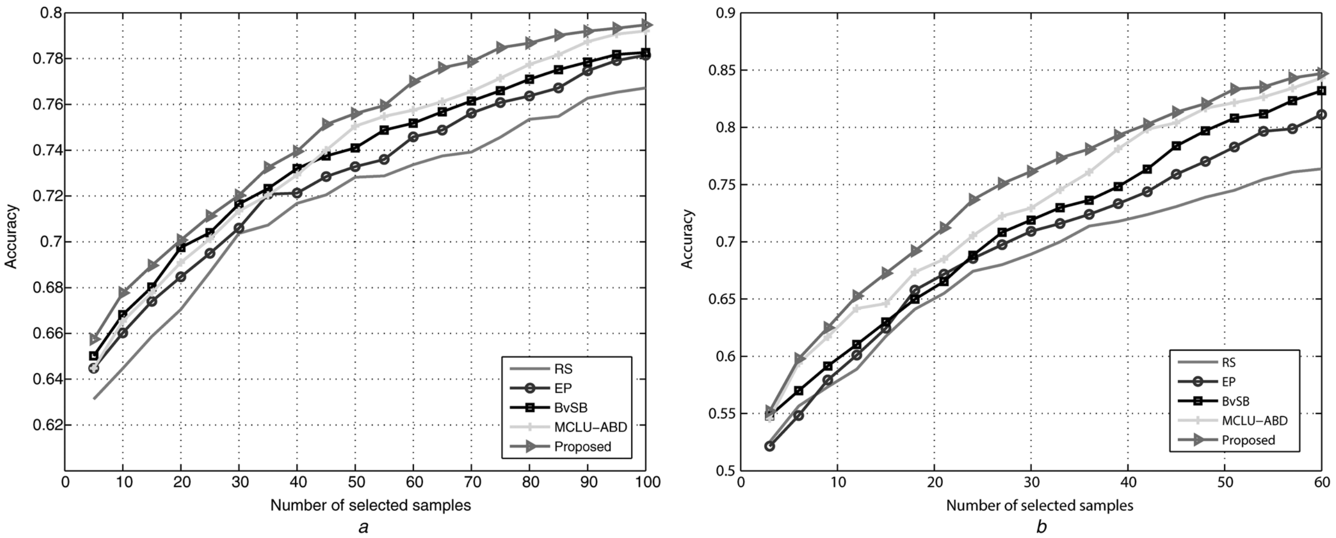


Fig. 4 *Classification performance on Outex-0 and Outex-1 dataset with different active learning algorithms*
a Results on Outex-0
b Results on Outex-1

Fig. 4 shows the average classification accuracy against the number of selected examples by different active learning algorithms. It can be seen that the proposed algorithm significantly outperforms RS, EP, BvSB and MCLU-ABD methods in most cases. MCLU-ABD performs better than RS, EP and MCLU-ABD, while performing worse than the proposed method. Both EP and BvSB outperform RS, and BvSB is slightly better than EP. As can be seen from Fig. 4*a*, with only 60 selected examples, the proposed algorithm performs even better than the other algorithms with 70 selected examples.

4.2 Scene recognition

Scene recognition is a vital important problem in computer vision. In this section, we perform scene recognition with the scene 13 dataset [16] using different active learning methods. The image dataset we used in this experiment consists of 2600 images of 13 natural scene categories (i.e. 200 images of each category). 40% of the dataset constructs the testing set to evaluate the classification performance. Six images from each category are randomly selected to construct the initial labelled set $\mathcal{L}$. The rest of the images are considered as unlabelled dataset $\mathcal{U}$ for active learning selection. The GIST features [19] are extracted since they have good discriminatory power in classification. Fig. 5 shows some sample images from the categories of forest, mountain, suburb, office, highway, kitchen, bedroom and living room.

At each round, we apply each active learning algorithm to select $k(=10)$ examples for manual labelling and adding to $\mathcal{L}$ for retraining classifiers. This selection process will repeat 20 times, thus 200 examples are selected in total. The experiments are repeated 20 times and the average accuracy is computed as the final result.

Fig. 6 shows the average accuracy of different active learning approaches. Again, our proposed active learning algorithm outperforms the other algorithms in most of the cases. MCLU-ABD performs the second best and it is close to the proposed method in some cases. BvSB performs slight better than EP. RS performs worst in most of cases since it selects examples without any criteria.

4.3 Time consumption analysis

Time consumption of machine learning algorithms is a significant problem in real-world applications. In this section, the time cost of different active learning algorithms is analysed.

Fig. 5 *Sample images from categories: forest, mountain, suburb, office, highway, kitchen, bedroom and living room*

doi: 10.1049/iet-cvi.2014.0140

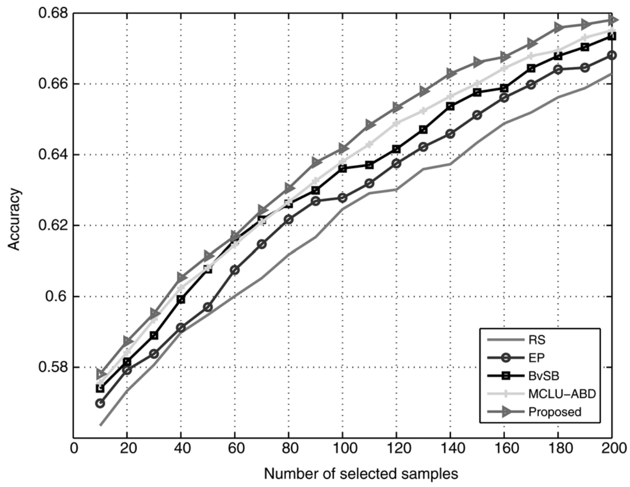


Fig. 6 *Classification results on scene 13 dataset*

A Macbook with 2.4 GHz Intel Core i5 and 8 GB RAM with OS X 10.9 and MATLAB 2013a was used as the experimental platform. The time consumption of two experiments is shown in Tables 1 and 2. Obviously, RS costs the least time among all the algorithms. The time consumption of EP, BvSB and MCLU-ABD is comparable to each other. Our proposed method costs a little more time than EP, BvSB and MCLU-ABD, since it computes both the uncertainty measure and the diversity measure of the examples. Considering the improvement of classification performance, this little extra time is worth while. Consequently, it is not a very time-consuming algorithm.

4.4 Parameter selection

There are two essential parameters in our proposed algorithm: the number of selected most uncertain examples h and the tradeoff parameter λ in (9). These two parameters are empirically set to $10 \times k$ and 0.6 in previous experiments. In this section, we examine the impacts of these two parameters on the experimental performance.

The impacts of the parameters on the two datasets are very similar. Therefore, only the results on the scene-13 dataset are shown here. For comparison, we let each active learning algorithm select $k = 10$ examples for adding to the training set with initial 100 labelled examples. As before, the evaluations are conducted with 20 randomly generated subsets and each subset contains 2600 samples. Figs. 7 and 8 show the average classification accuracy against different parameters h and λ. In Fig. 7, λ is fixed at 0.6, and in Fig. 8, the value of h is fixed at 100. As can be seen from the figures, the proposed method can achieve better performance than other methods over a large range of h and λ. For example, in Fig. 8, if the value of λ is in [0.3, 0.8], the performance is better than other algorithms. If λ is close to 0 or 1, the proposed method degenerates into uncertainty sampling or diversity sampling where only one criterion works. Therefore, the performance is worse than MCLU-ABD. In conclusion, the experimental results are not very sensitive to the parameters. Thus, the parameter

Table 1 Time consumption of terrain classification (Outex-0) experiment (seconds)

Algorithm	RS	EP	BvSB	MCLU-ABD	Proposed
time-consuming	2.61	8.65	8.30	8.02	9.55

Table 2 Time consumption of scene recognition experiment (seconds)

Algorithm	RS	EP	BvSB	MCLU-ABD	Proposed
time-consuming	20.49	70.97	70.70	74.89	77.46

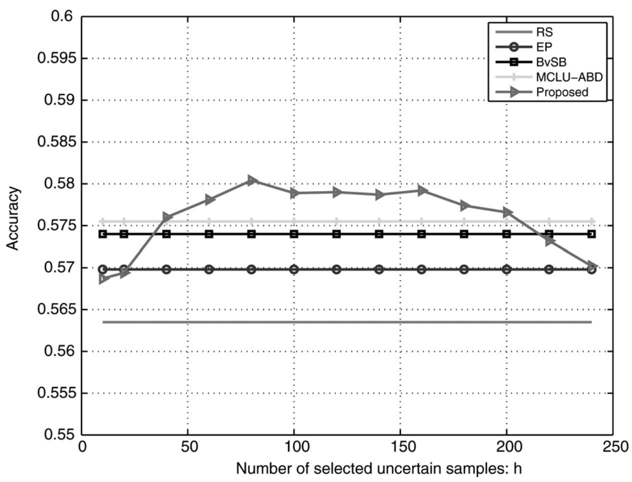


Fig. 7 *Impacts of the parameter h on the performance of the proposed active learning algorithm*

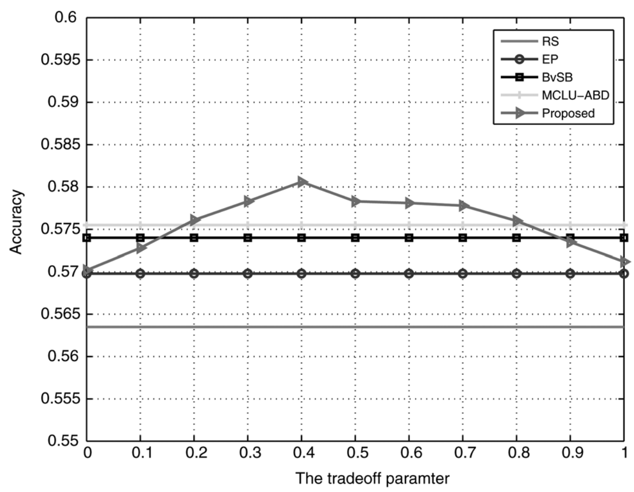


Fig. 8 *Impacts of the tradeoff parameter λ on the performance of the proposed active learning algorithm*

selection is not a crucial problem in our algorithm. It is not a difficult job to find appropriate parameters for the experiment.

5 Conclusions

In this paper, we proposed a novel active learning algorithm to select the most uncertain and diverse examples. One-vs-one strategy has been adopted for SVM to solve multi-class classification. A new uncertain measure is developed from the multiple binary classifiers and the Gaussian kernel is applied for similarity measure. During the example selection process, first h most uncertain examples are selected for further selecting. Then k most uncertain and diverse examples are selected by dynamic programming algorithm. The experimental results on two real-world applications demonstrate the effectiveness of our approach.

Both uncertainty and diversity are very popular criteria in active learning. How to measure the uncertainty and diversity of examples has been explored a lot, but there is not an accepted best method. In addition, how to combine different criteria is a critical and difficult problem in active learning. In the future, we will explore better methods to measure examples' uncertainty and diversity and develop an advanced combination approach.

6 Acknowledgment

This work is partially supported by the National Natural Science Foundation of China under grant nos. 61373063, 61233011, 61125305, 61375007, 61220301, and by the National Basic Research Program of China under grant no. 2014CB349303.

7 References

1 Settles, B.: 'Active learning literature survey', University of Wisconsin, Madison, 2010
2 Fu, Y., Zhu, X., Li, B.: 'A survey on instance selection for active learning', *Knowledge Inf. Syst.*, 2013, **35**, (2), pp. 249–283
3 Tong, S., Chang, E.: 'Support vector machine active learning for image retrieval'. Proc. Ninth ACM Int. Conf. Multimedia ACM, 2001, pp. 107–118
4 Ji, M., Han, J.: 'A variance minimization criterion to active learning on graphs'. Int. Conf. Artificial Intelligence and Statistics, 2012, pp. 556–564
5 Demir, B., Persello, C., Bruzzone, L.: 'Batch-mode active-learning methods for the interactive classification of remote sensing images', *IEEE Trans. Geosci. Remote Sens.*, 2011, **49**, (3), pp. 1014–1031
6 Gu, Y., Jin, Z.: 'Neighborhood preserving d-optimal design for active learning and its application to terrain classification', *Neural Comput. Appl.*, 2013, **23**, (7–8), pp. 2085–2092
7 Yifan, F., Xingquan, Z., Elmagarmid, A.K.: 'Active learning with optimal instance subset selection', *IEEE Trans. Cybern.*, 2013, **43**, (2), pp. 464–475
8 Hoi, S.C., Jin, R., Zhu, J., Lyu, M.R.: 'Semi-supervised SVM batch mode active learning for image retrieval'. IEEE Conf. Computer Vision and Pattern Recognition, 2008 (CVPR 2008), 2008, pp. 1–7
9 Joshi, A.J., Porikli, F., Papanikolopoulos, N.P.: 'Scalable active learning for multiclass image classification', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (11), pp. 2259–2273
10 Kang, L., Xu, Q.: 'A novel batch-mode active learning method for SVM classifier', *J. Inf. Comput. Sci.*, 2012, **9**, (16), pp. 5077–5084
11 Guo, Y., Schuurmans, D.: 'Discriminative batch mode active learning', Proc. of Adv. Neural Inf. Process. Syst., 2008, pp. 593–600
12 Ebert, S., Fritz, M., Schiele, B.: 'Ralf: A reinforced active learning formulation for object class recognition'. 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2012, pp. 3626–3633
13 Joshi, A.J., Porikli, F., Papanikolopoulos, N.: 'Multi-class active learning for image classification'. IEEE Conf. Computer Vision and Pattern Recognition, 2009 (CVPR 2009), 2009, pp. 2372–2379
14 Wu, T.-F., Lin, C.-J., Weng, R.C.: 'Probability estimates for multi-class classification by pairwise coupling', *J. Mach. Learn. Res.*, 2004, **5**, (975–1005), p. 4
15 University of Oulu texture database. [Online]. Available at: http://www.outex.oulu.fi/temp/
16 Fei-Fei, L., Perona, P.: 'A Bayesian hierarchical model for learning natural scene categories'. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2005 (CVPR 2005), 2005, vol. 2, pp. 524–531
17 Procopio, M.J., Mulligan, J., Grudic, G.: 'Learning terrain segmentation with classifier ensembles for autonomous robot navigation in unstructured environments', *J. Field Robot.*, 2009, **26**, (2), pp. 145–175
18 Ojala, T., Pietikainen, M., Maenpaa, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (7), pp. 971–987
19 Oliva, A., Torralba, A.: 'Modeling the shape of the scene: A holistic representation of the spatial envelope', *Int. J. Comput. Vis.*, 2001, **42**, (3), pp. 145–175

doi: 10.1049/iet-cvi.2014.0140