Research on Multiple Instance Active Learning Algorithm Based on Support Vector Machine

Respondent: Sudong Chen

Instructor: Ran Wang

CONTENTS

Introduction

1

In traditional machine learning, sample : instance = 1 : 1.

But in practical problems, sample : instances = 1 : N.

So Multiple-Instance Learning(MIL) is proposed to solve such problems.

2

Multiple Instance Active Learning mainly studies the problem of Active Learning in multiple-instance framework.
It can collect enough training bags for multiple-instance learning problems by iteratively selecting and querying the most valuable unlabeled bags.

3

In this paper, we compare the test accuracy and AUC values of two multiple-instance learning methods (mi-SVM and MI-SVM) on five evaluation criteria through experiments on MNIST handwritten digital image recognition problem.

Background Knowledge

In **MIL**, training samples are called bags. A bag consists of many instances.

If all the instances of a bag are negative, the bag can be defined as a **negative bag**.

If there is at least one positive instance in this bag, it can be defined as a **positive bag**.

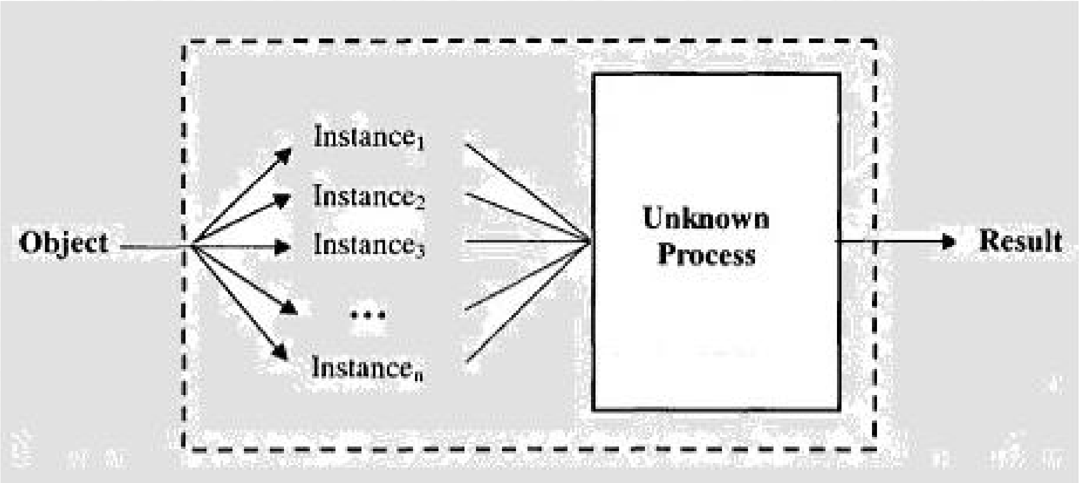
1

Multiple Instance Learning (MIL) ——

1）MI Algorithm based on instance level space

2）MI Algorithm based on bag level

3）MI Algorithm based on embedded space



1

Multiple Instance Learning (MIL) ——

1） MI Algorithm based on instance level space

2） MI Algorithm based on bag level

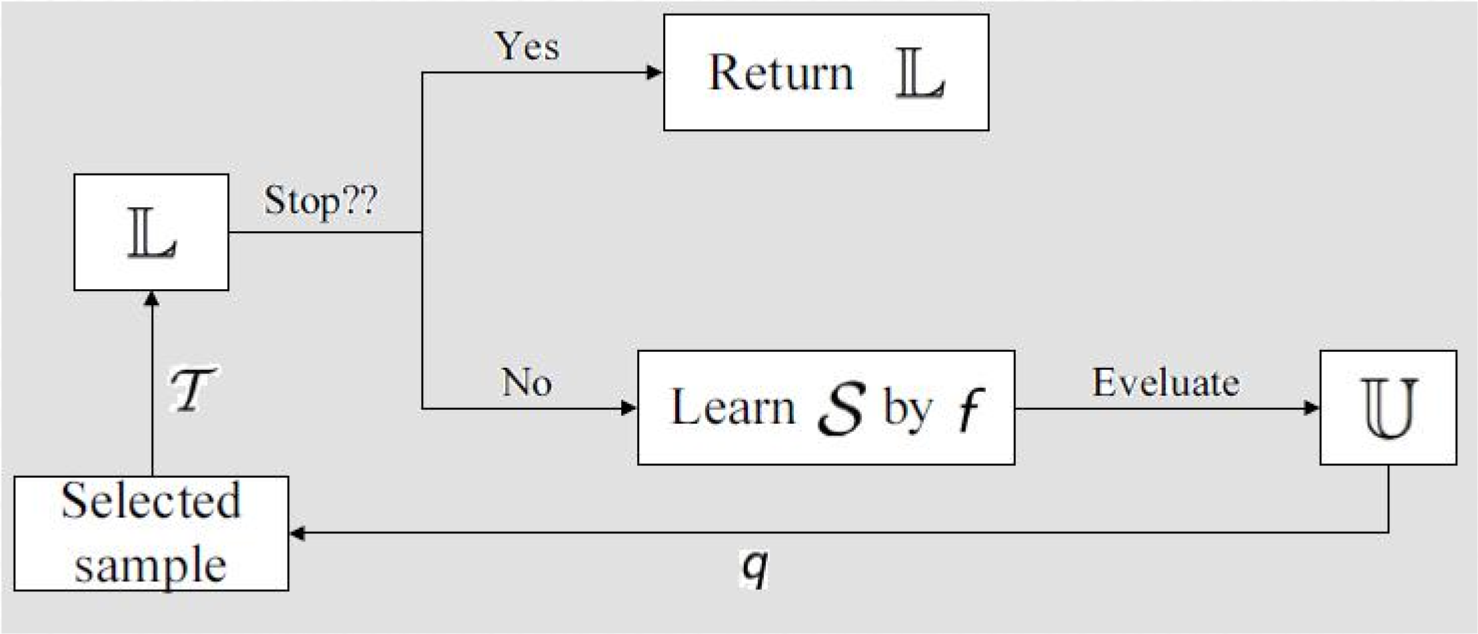
3） MI Algorithm based on embedded space

2

Active Learning (AL) ——

1） Stream-based AL

2） Pool-based AL



1 Multiple Instance Learning (MIL) ——

1） MI Algorithm based on instance level space

2） MI Algorithm based on bag level

3） MI Algorithm based on embedded space

2 Active Learning (AL) ——

1） Stream-based AL

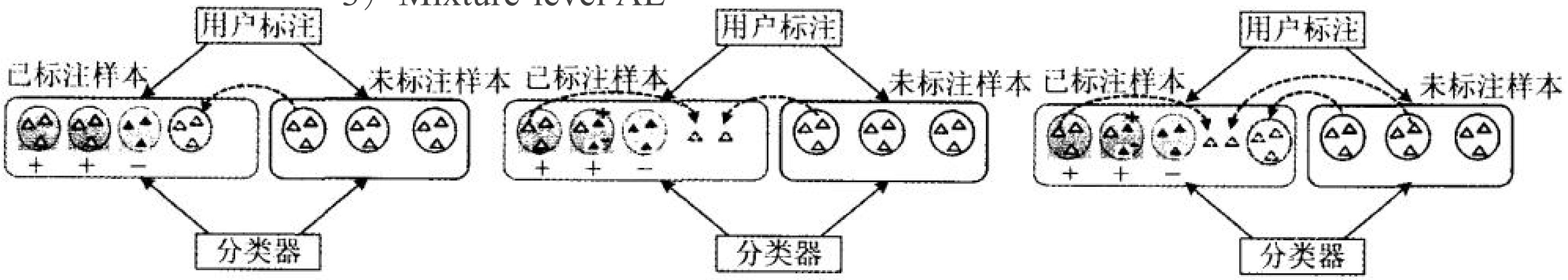
2） Pool-based AL

3 Multiple Instance Active Learning (MIAL) ——

1） Bag-level AL

2） Instance-level AL

3） Mixture-level AL



(a) 包层主动学习　(b) 示例层主动学习　(c) 混合层次主动学习

Methods

Support Vector Machine

A method of machine learning. Its goal is to find the optimal classification hyperplane that maximizes the margin of training data sets.

A

Hyperplane

A plane that can be used to divide data.

Margin

Double the distance from the data point to the hyperplane.

B

Linear situation

$$\max_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(\mathbf{x}_i\cdot\mathbf{x}_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i=0 \quad \alpha_i\geq 0, \quad i=1,2,\ldots,N$$

C

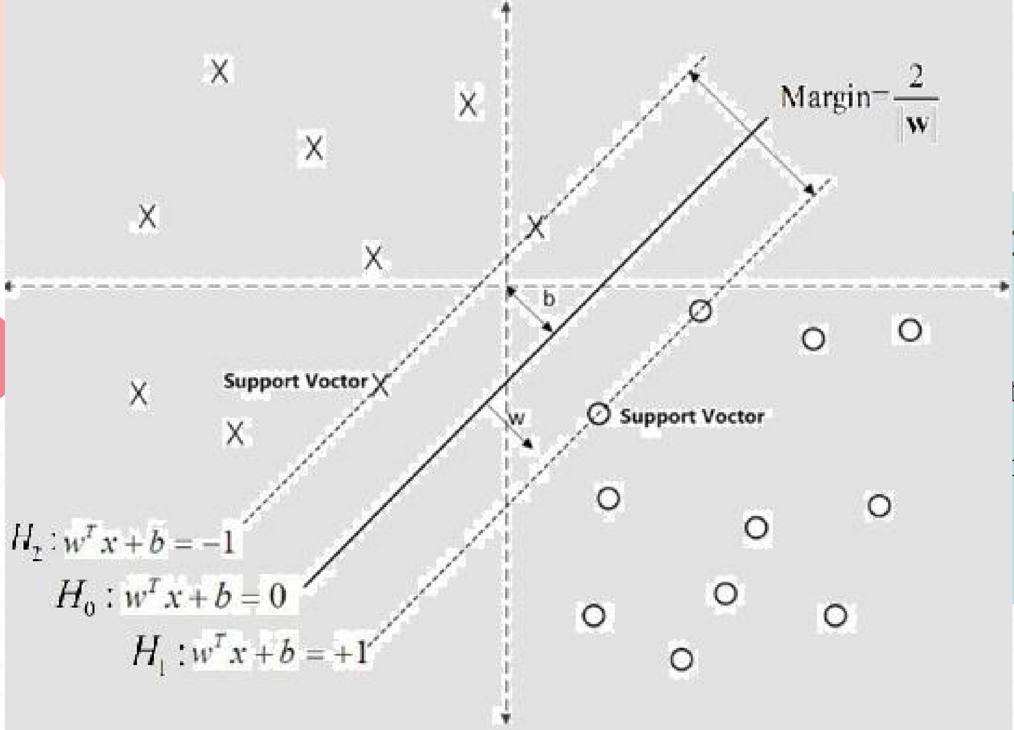
Nonlinear situation

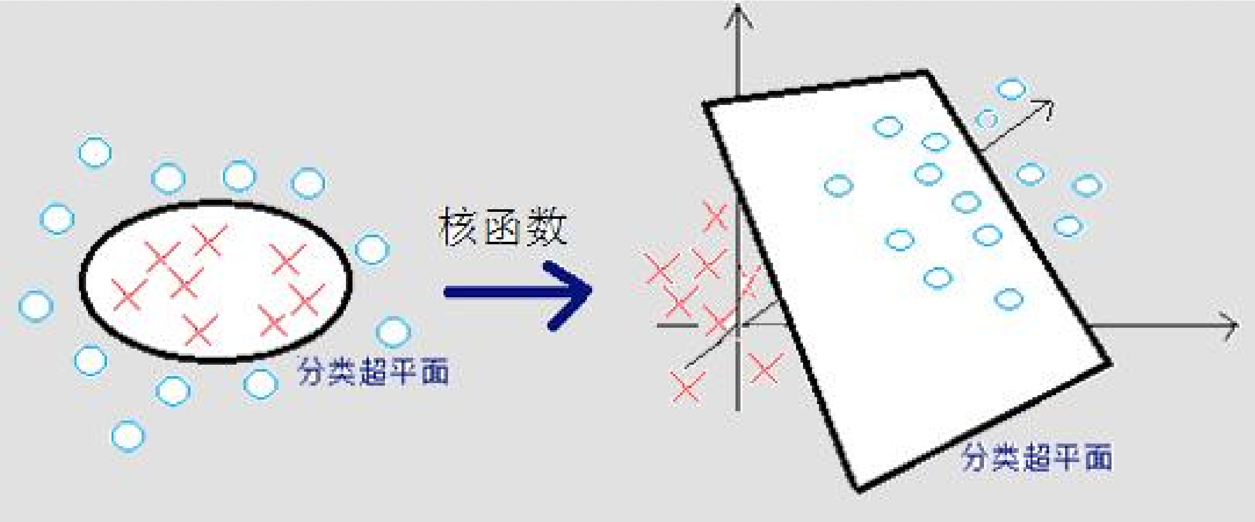
$$\max_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i\cdot\mathbf{x}_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i=0$$

$$0\leq\alpha_i\leq C, \quad i=1,2,\ldots,N$$

D





Linear situation

$$\max_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \quad \alpha_i \geq 0, \quad i = 1,2,\ldots,N$$

C

Nonlinear situation

$$\max_{\boldsymbol{\alpha}} \quad Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, \quad i = 1,2,\ldots,N$$

D

mi-SVM Model

$$\min_{\{y_i\}} \min_{\mathbf{w},b,\xi} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i$$

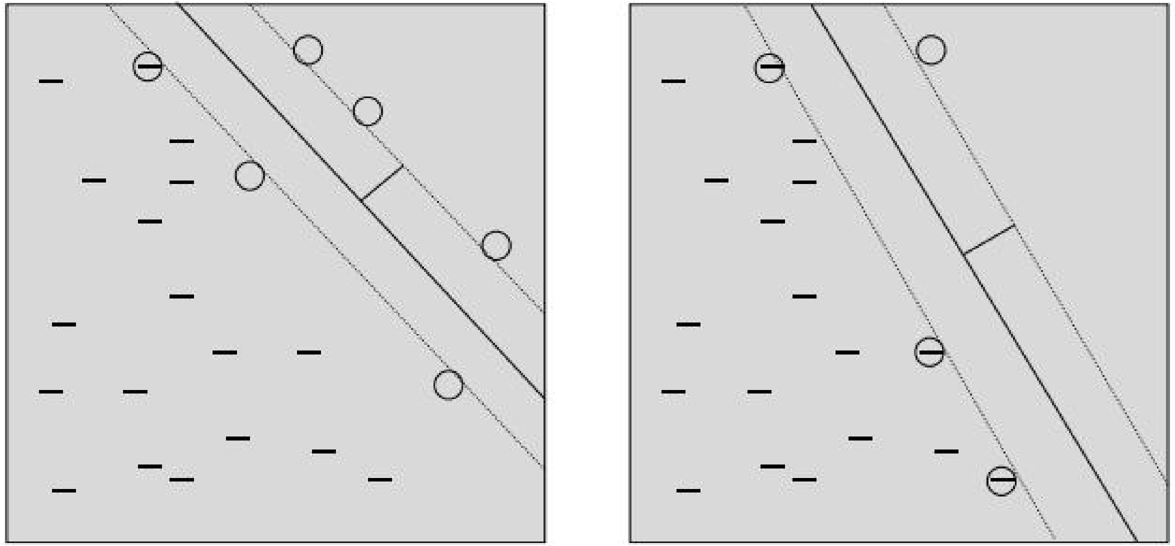
$$\text{s.t.} \quad \forall i:\ y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i,\ \ \xi_i \geq 0,\ y_i \in \{-1, 1\}$$

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1,\ \ \forall I \text{ s.t. } Y_I = 1,\ \ \text{以及}\ y_i = -1, \forall I \text{ s.t. } Y_I = -1$$

MI-SVM Model

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_I \xi_I$$

$$\text{s.t.} \quad \forall I:\ Y_I \max_{i \in I}(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_I,\ \ \xi_I \geq 0$$



MIL大间隔分类器：左图为mi-SVM解决方案，右图为MI-SVM解决方案

Evaluation criteria for MIAL instance bags

1．Bag Margin

$$\begin{cases} I_m(B_i) = 1 / \min_{B_{ij} \in B_i} | f(B_{ij}) | \\ I_\alpha(B_i) = \sqrt{n_i / \sum_{B_{ij} \in B_i} | f(B_{ij}) |^2} \end{cases}$$

2．Softmax Model

$$\begin{cases} P(y_{ij} = +1 | B_i) = \text{sotfmax}_\alpha (P(y_{i1} = +1 | B_{i1}), .., P(y_{in_i} = +1 | B_{in_i})) \\ P(y_i = -1 | B_i) = 1 - P(y_{ij} = +1 | B_i) \end{cases}$$

$$I(B_i) = -\sum_{y_i = \pm 1} P(y_i | B_i) \log P(y_i | B_i)$$

Evaluation criteria for MIAL instance bags

3. CombineU Model

$$\text{sotfmax}_{\alpha}(x_1,\ldots,x_n)=\sum_{i=1}^{n}x_i\cdot e^{\alpha\cdot x_i}\ /\sum_{i=1}^{n}e^{\alpha\cdot x_i}$$

$$I(B_i)=\text{sotfmax}_{\alpha}(u(B_{i1}),\ldots,u(B_{in_i}))$$

4. Noisy—Or Model

$$\begin{cases}P(y_i=+1\,|\,B_i)=1-\prod_{B_{ij}\in B_i}(1-P(y_{ij}=+1\,|\,B_{ij}))\\P(y_i=-1\,|\,B_{ij})=\prod_{B_{ij}\in B_i}(1-P(y_{ij}=+1\,|\,B_{ij}))\end{cases}$$

Evaluation criteria for MIAL instance bags

5. Fisher Information

$$I_{q(x)}(\theta) = -\int q(x)dx \int p(y \mid x,\theta) \frac{\partial^2}{\partial\theta^2} \log p(y \mid x,\theta) dx$$

$$I(B_i) = -tr\left(\sum_{y_i=\pm 1} p(y_i \mid B_i,\theta) \frac{\partial^2}{\partial\theta^2} \log p(y_i \mid B_i,\theta)\right)$$

$$= (P_i^- / P_i^+) \times \sum_{j\in B_i} (P_{ij}^+ \times \phi(B_{ij}))^T \sum_{j\in B_i} (P_{ij}^+ \times \phi(B_{ij})))$$

$$= (P_i^- / P_i^+) \times \sum_{j\in B_i} \sum_{q\in B_i} P_{ij}^+ \times P_{iq}^+ \times K(B_{ij}, B_{iq})$$

Results

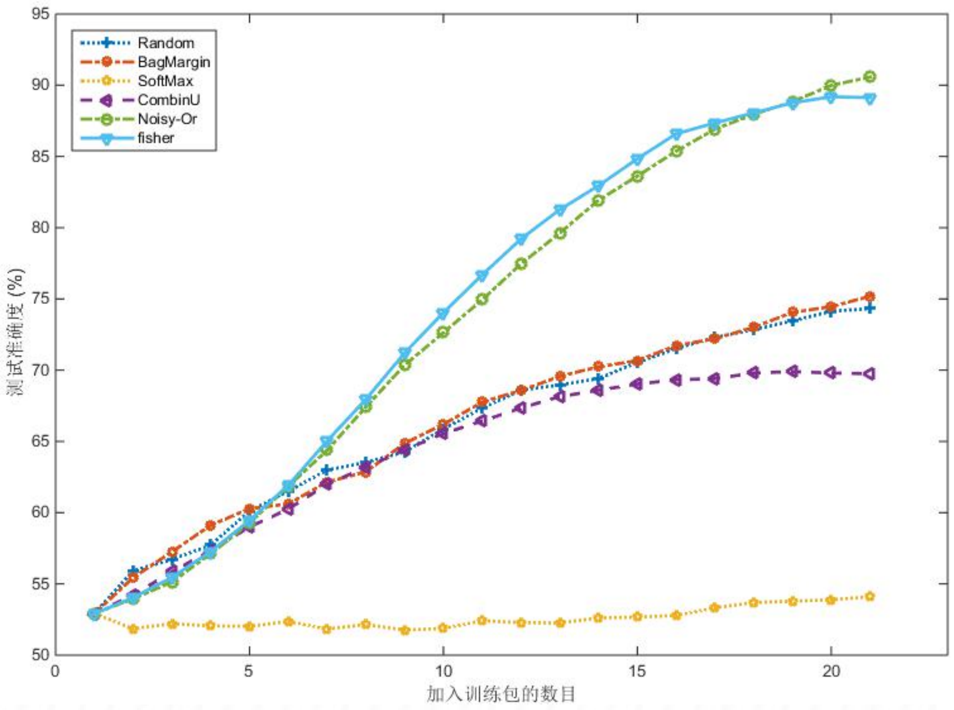
1.The data set is a newly generated MIL data set from MNIST handwritten digital image recognition problem. The original problem contains 60,000 training samples and 10,000 testing samples.

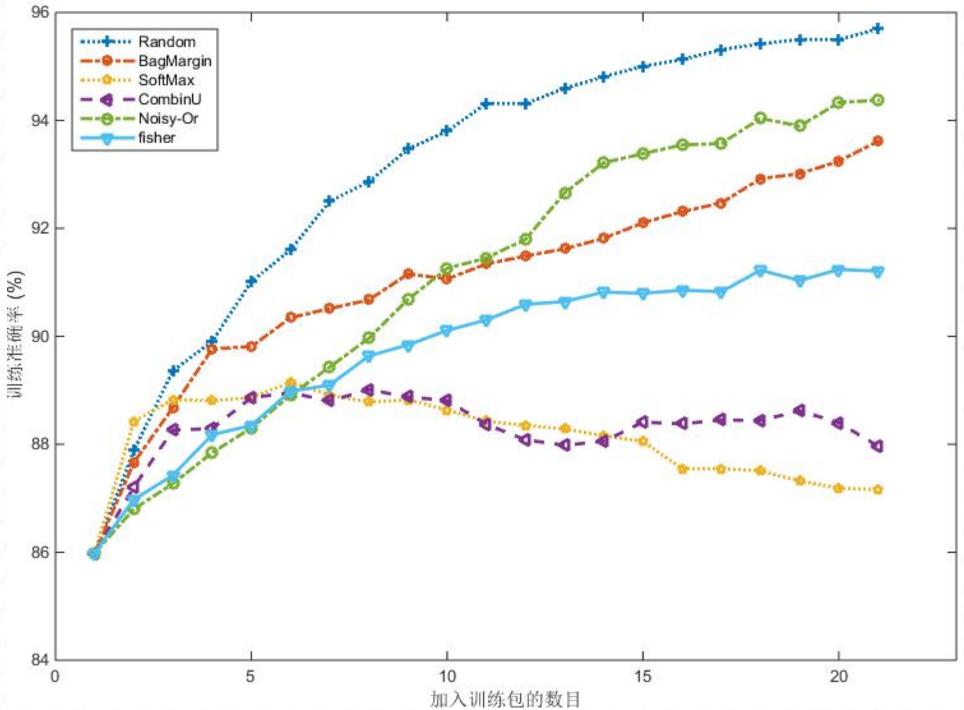
2.For each class, we manually generate a single digital MIL data set containing 100 positive and 100 negative bags. We randomly allocate the number of instances in a bag from [10,40].

3. For each handwritten digital image set, we randomly select 50% of the bags as the training set and the remaining 50% of the bags as the testing set.

The following figure shows the average test accuracy of the data set.（The left is mi-SVM and the right one is MI-SVM）

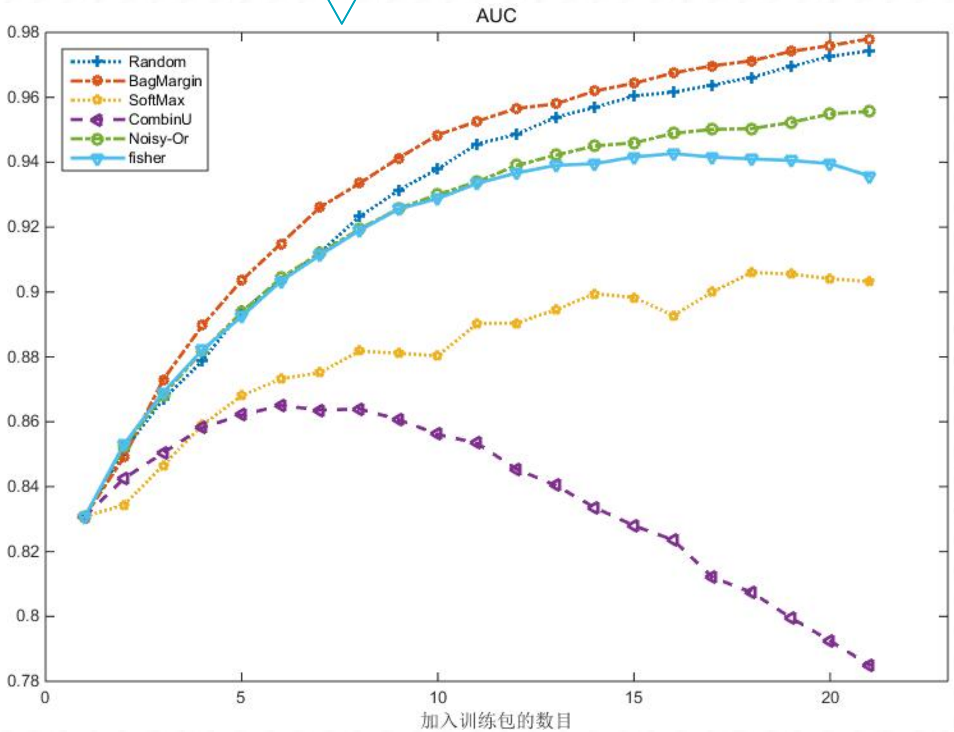
1. The learning performance of Noisy-Or and Fisher Information in **mi-SVM** model is the best among all algorithms, while Softmax is the worst.
2. The accuracies of **MI-SVM** model are all more than 86%, and the learning effect of other algorithms is not as good as Random, so the learning effect is not obvious.

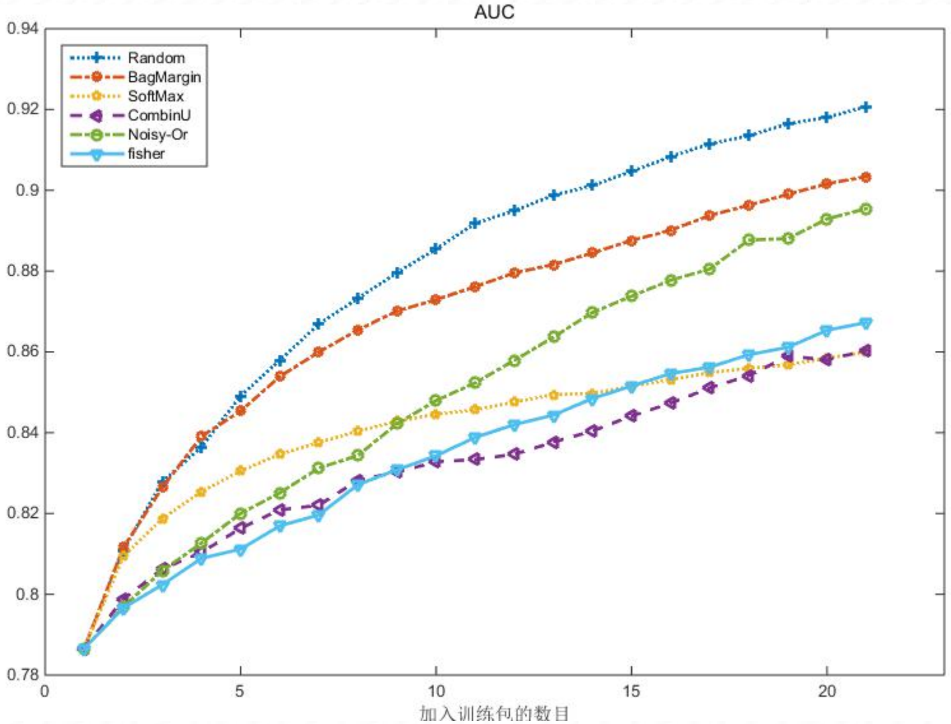




The following figure shows the average AUC of the data set.（The left one is mi-SVM and the right is MI-SVM）

1. In **mi-SVM** model, the AUC values of the Softmax and the CombinU are lower than other algorithms, which means the effect is not so good as others.
2. In **MI-SVM** model, the AUC of all algorithms is getting larger and larger, which shows that the classification effect is getting better and better.





Thank You