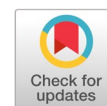


Multi-granularity active learning based on the three-way decision



Wu Xiaogang^{a,b,1,*}, Thitipong^{a,2}

^a Vincent Mary School of Science & Technology, Assumption University, Bangkok, Thailand

^b School of Information Technology, Xingyi Normal University for Nationalities, Xingyi, China

¹ wxg817@163.com; ² thitipong@scitech.au.edu

* corresponding author

ARTICLE INFO

Article history

Received February 21, 2023

Revised March 25, 2023

Accepted April 7, 2023

Available online May 4, 2023

Keywords

Three-way decision

Multi-grained features

Active learning

Unlabeled samples

Classification algorithm

ABSTRACT

The reliance on data and the high cost of data labeling are the main problems facing deep learning today. Active learning aims to make the best model with as few training samples as possible. Previous query strategies for active learning have mainly used the uncertainty and diversity criteria, and have not considered the data distribution's multi-granularity. To extract more valid information from the samples, we use three-way decisions to select uncertain samples and propose a multi-granularity active learning method (MGAL). The model divides the unlabeled samples into three parts: positive, negative, and boundary region. Through active iterative training samples, the decision delay of the boundary domain can reduce the decision cost. We validated the model on five UCI datasets and the CIFAR10 dataset. The experimental results show that the cost of three-way decisions is lower than that of two-way decisions. The multi-granularity active learning achieves good classification results, which validates the model. In this case study, the reader can learn about the ideas and methods of the three-way decision theory applied to deep learning.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Deep learning networks have made extraordinary progress in artificial intelligence [1]–[4]. Multi-granularity structure is an important feature of deep networks, whether it is image, text, or speech data, where feature representations can be extracted at different granularities [5], [6]. Deep networks extract finer-grained features by processing data from low to high and building multi-level, multi-granular semantics [7]–[9]. Deep networks consist of multiple layers of small groups of neurons (e.g. convolutional kernels, etc.), each of which processes only a portion of the input image. In general, the lower the underlying layer, the smaller the receptive field and the finer the granularity of the features. The features of each layer are combined into a new receptive field, usually more significant than the receptive field of the previous layer, and this is the granularity of the features.

As in Fig. 1, the VGGNet16 [10] network goes through multiple layers of feature extraction to obtain a high-level representation of an image. Res2Net [11] is an improved multi-level, multi-granularity residual network. Fig. 2 shows the structure of each residual unit to obtain different features f_i ($i=1,2,\dots,s$) by fusion to increase the multi-granularity feature representation capability of the network. The hierarchical pyramid structure of the FPN network [12] mines multi-grained features of images and fuses the information of multi-grained features to obtain rich and powerful high-level features (Fig. 3). Training deep learning networks requires a lot of data and tag costs [13]. Active learning can reduce labeling costs by selecting the most informative samples of the dataset [14], [15].

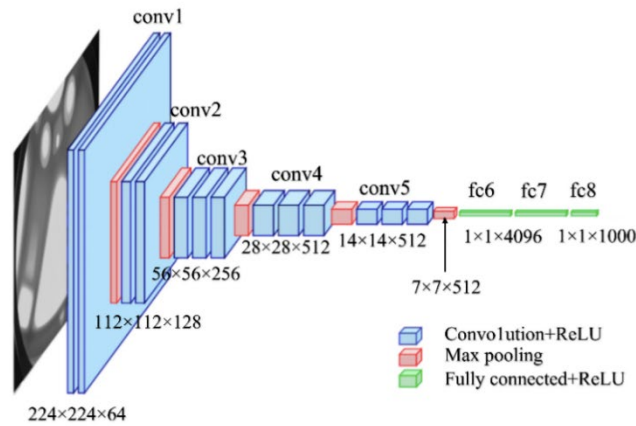


Fig. 1. Multi-grain convolutional features of VGG networks

Unlike traditional passive supervised learning, active learning methods train on labeled samples to obtain a priori knowledge, which is then used to evaluate the value of unlabeled samples. Therefore, how to efficiently select unclassified labeled samples with high classification contributions for annotation and add them to the existing training set is a key issue in active learning.

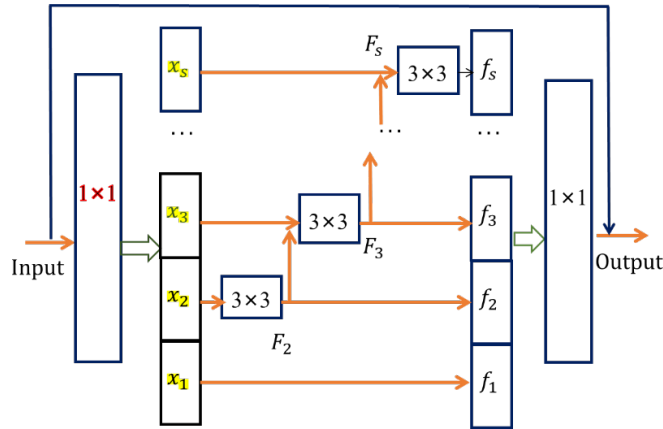


Fig. 2. Multi-granularity feature extraction with residual module

However, in active learning, the arbitrary uncertainty of the recognition samples may also lead to higher computational costs in iterative retraining [16]. To speed up training and extract features more efficiently, we introduce a three-way decision approach [17] in active learning networks, where each batch of new labels is trained incrementally and multi-granularity feature extraction of uncertain samples is performed using three-way decisions, which greatly reduces the computational demand of active learning methods.

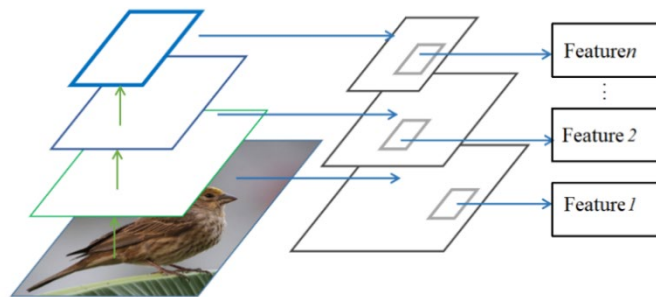


Fig. 3. Pyramidal multi-grain feature extraction for FPN networks

Three-way decision theory [18] is a multi-granularity approach to solving complex problems. It complements and extends classical two-way decision-making; It simulates solving problems for human minds by determining the acceptance and rejection domains and taking delayed confirmation decisions for inaccurate or uncertain information. By gathering more granular data, the uncertainty part is transformed into the realm of certainty (the domain and rejection), thereby reducing its inconsistency and improving the accuracy of decisions.

In this paper, our contributions are mainly as follows:

- A novel model combining deep and active learning is proposed to achieve dynamic incremental image recognition and classification of data.
- The sample selection strategy of active learning is improved to reduce the cost of iterative training.
- The hierarchical and multi-granularity characteristics of decision queries are achieved. Using three-way decision-making, representative samples from uncertain samples can be effectively extracted, improving the performance of the classifier.

Following the Introduction section, Section 2 describes the three-way decision and active learning algorithms. In Section 3, we discuss the network model parameters and experimental results, and Section 4 concludes the paper.

2. Method

2.1. Three-way decision

The three-way decision is a way to model a multi-level decision by dividing the whole into three distinct and related parts [19]. Humans can immediately judge entirely accepted and rejected; research and learning are required to make final judgments for uncertain things. The inaccuracy of an indefinite item is due to its different granularity. The solution to this uncertainty can be a granularity transformation of its attributes to refine the coarse granularity, thus making the uncertain object a definite object.

To implement the three-way decision, it is first necessary to introduce the decision function $f(x)$ of the entity, also called the evaluation function; then, a pair of threshold values α and β is introduced, and the event objects in the argument domain U are divided into positive $POS(U)$, boundary $BND(U)$ and negative domains $NEG(U)$ according to the decision state value and threshold value, corresponding to the acceptance, deferral and rejection rules of the three-way decision. (α, β) denotes the upper and lower approximation. The decision rules are defined as follows:

$$\begin{aligned} & \text{if } f(x) \geq \alpha, \text{ then } x \in POS(U) \\ & \text{if } \alpha < f(x) < \beta, \text{ then } x \in BND(U) \\ & \text{if } f(x) \leq \beta, \text{ then } x \in NEG(U) \end{aligned} \quad (1)$$

For a sample set U , C represents the description of the state U . For each sample $x \in U$, $f(x)$ represents the evaluation of the current definition of $Des(x)$. The three decision steps are: targets that satisfy C are assigned to $POS(U)$, those that do not are assigned to $NEG(U)$ and those that are difficult to judge are assigned to $BND(U)$. The cost of classifying a sample into the three domains is different. Assuming that S_P represents a sample in the positive domain and S_N represents a sample in the negative field, these two fields are divided into three regions, resulting in six different classification cost functions, as in (2).

$$\begin{aligned} \lambda_{PP} &= \lambda\{POS|L(x) = S_P\}, \lambda_{PN} = \lambda\{POS|L(x) = S_N\}, \\ \lambda_{NP} &= \lambda\{NEG|L(x) = S_P\}, \lambda_{NN} = \lambda\{NEG|L(x) = S_N\}, \end{aligned}$$

$$\lambda_{BP} = \lambda\{BND|L(x) = S_p\}, \lambda_{BN} = \lambda\{BND|L(x) = S_N\} \quad (2)$$

Where $L(x)$ is the conditional probability of classifying sample x into the positive domain, and the expected risk of classifying sample x into the three domains can be calculated using (3).

$$\begin{aligned} C(POS|x) &= \lambda_{PP}f(x) + \lambda_{PN}(1-f(x)) \\ C(NEG|x) &= \lambda_{NP}f(x) + \lambda_{NN}(1-f(x)) \\ C(BND|x) &= \lambda_{BP}f(x) + \lambda_{BN}(1-f(x)) \end{aligned} \quad (3)$$

In general, the cost of correctly classifying a positive domain sample into its region is small or even negligible, and the cost of delaying a decision on a sample should be lower than the cost of misclassification, i.e. :

$$\lambda_{PP} < \lambda_{BP} < \lambda_{NP}, \lambda_{NN} < \lambda_{BN} < \lambda_{PN} \quad (4)$$

Furthermore, the cost of incorrectly accepting a negative domain should be much higher than the cost of incorrectly rejecting a positive example, and based on the above equation, the decision threshold can be obtained as follows :

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{\lambda_{PN} - \lambda_{BN} + \lambda_{BP} - \lambda_{PP}}, \beta = \frac{\lambda_{BN} - \lambda_{NN}}{\lambda_{BN} - \lambda_{BP} + \lambda_{NP} - \lambda_{NN}} \quad (5)$$

The cost of partitioning the sample into boundary regions varies with the number of three-way decision steps. The richer the information, the less acceptable the delayed decision and the higher the cost of making the delayed decision. As the cost of delayed choices increases, the extent of the boundary region gradually shrinks or even disappears, and the three-way decision degenerates into a two-way decision.

2.2. Multi-granularity three-way decision model

The sequential three-way decision is a multi-granular decision model (Fig. 4) dealing with dynamic decision problems [20], [21].

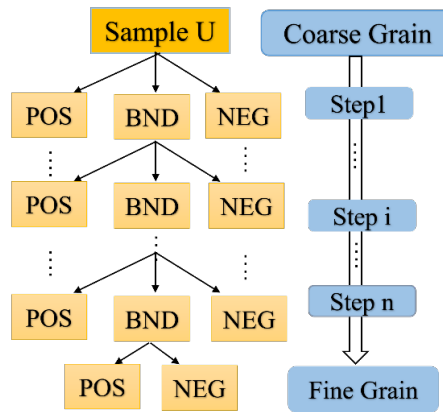


Fig. 4. Multi-grained decision model with sequential three-way decision

Assuming that the n -step decisions constitute n levels of granularity, $Des_i(x)$, $f_i(x)$ represent the state description of sample x at step i and the evaluation of the decision maker, respectively, and the increasing accuracy of these evaluations as in (6).

$$(Des_1(x), f_1(x)) \preceq (Des_2(x), f_2(x)) \preceq \dots \preceq (Des_n(x), f_n(x)) \quad (6)$$

This constitutes a step-by-step cognitive process of moving from coarse to fine granularity of the target.

For objects divided into boundary regions, only the last layer uses a two-way decision, and the other granular layers use a three-way decision. Therefore, a reasonable decision threshold is set on each grain layer. It is a multi-granularity decision approach where the cost of each delayed decision step increases as i increases. The threshold value for each decision step is given in (7).

$$\alpha_i = \frac{\lambda_{PN}^i - \lambda_{BN}^i}{(\lambda_{PN}^i - \lambda_{BN}^i) + (\lambda_{BP}^i - \lambda_{PP}^i)}$$

$$\beta_i = \frac{\lambda_{BP}^i - \lambda_{NN}^i}{(\lambda_{BN}^i - \lambda_{NN}^i) + (\lambda_{NP}^i - \lambda_{BP}^i)} \quad (7)$$

The final step of the three-way decision has $\alpha_n = \beta_n$, and assuming that $\lambda_{BN}^i / \lambda_{BP}^i = C$, the final threshold obtained can be calculated as follows:

$$\lambda_{BP}^n = \frac{\lambda_{PN} \lambda_{NP} - \lambda_{PP} \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + C(\lambda_{NP} - \lambda_{PP})}$$

$$\lambda_{BN}^n = C \frac{\lambda_{PN} \lambda_{NP} - \lambda_{PP} \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + C(\lambda_{NP} - \lambda_{PP})} \quad (8)$$

Fig. 5 shows the multi-granularity three-way decisions Classification algorithm. The Input was imbalanced dataset, attributes of multi-granularity information, decision cost thresholds of different granularity (α_i, β_i).

```

1   Initialize U1=U; POS=NEG=BND=∅
2   for i=1,2,...,n do
3        $P(X|[x_i]) = P(X)P([x_i]|X)/P([x_i])$ ;
4       if( $P(X|[x_i]) \geq \alpha_i$ )
5           POS=POS ∪  $POS_i$ 
6       else if( $P(X|[x_i]) \leq \beta_i$ )
7           NEG=NEG ∪  $NEG_i$ 
8       else
9            $U_{i+1} = BND_i$ 
10      i=i+1
11  end for
12  if  $U_n \neq \phi$  then
13       $POS_i = \{x \in U_n | P(X|[x_i]) \geq \gamma_i\}$ ;
14       $NEG_i = \{x \in U_n | P(X|[x_i]) < \gamma_i\}$ 
15      POS=POS ∪  $POS_i$ 
16      NEG=NEG ∪  $NEG_i$ 
17  End if

```

Fig. 5. Multi-granularity three-way decisions Classification algorithm

2.3. Active learning

The active learning method interacts with the classifier and the expert to obtain high-value samples to improve the classifier's performance. It avoids redundancy and unnecessary additions to data and reduces the cost of tagging large amounts of data [22]. Active learning is divided into two steps: 1) Sample selection: select the most valuable sample marks and then add them to the training set; 2) Model training: supervised learning is carried out to measure the classification performance of the classifier. In the active learning process, the two steps are executed alternately in an iterative. The algorithm terminates when a preset number of iterations or a preset classification performance is reached, as shown in Fig. 6.

In Fig. 6, an initial classifier must be trained by obtaining a portion of the used samples at the beginning. The selected examples will significantly influence the evaluation of the initial classifier.

Suppose the initial classifier has a certain level of accuracy. Suppose the initial classifier has a sure accuracy. In that case, it can minimize inefficient markers, speed up the active learning process to a certain extent, and achieve the same performance with fewer samples.

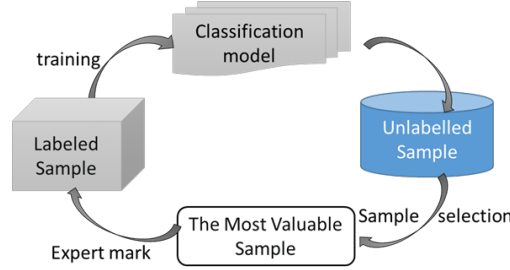


Fig. 6. Multi-granularity decision model with sequential three-way decision

The sample selection strategy determines the amount of information in the training set. The most straightforward method is random sampling, but more is needed to improve the generalization performance of the classifier. In contrast, generating high-entropy samples is a strategy for selecting the most valuable samples [23]. Based on the information entropy definition of the probability distribution, chosen examples by interval sampling are:

$$x^* = \operatorname{argmax}_{x \in U} (-\sum_i^n P(y = j_i | x) \log(P(y = j_i | x))), \quad (9)$$

where j_i denotes the most likely category i .

Fig. 7 shows the uncertainty entropy sampling algorithm. The Input was unlabeled samples U , sample queries: n , similarity threshold: s , and Output: L (set of samples to be labeled).

```

1  Initialization:  $t=0$ ;  $L=\emptyset$ 
2  While  $t < n$  do
3    Maximum similarity  $m=0$ 
4    Select sample:  $x_i \in U$ 
5    for each  $x_j \in L$  do
6      Computational similarity:  $\text{Sim}(x_i, x_j)$ 
7       $m = \max(m, \text{Sim}(x_i, x_j))$ 
8    end for
9    if  $(m < s)$  then
10      $L = L \cup x_i$ 
11      $t = t + 1$ 
12   End if
13    $U \leftarrow U - x_i$ 
14 End while
15 Return  $L$ 

```

Fig. 7. Uncertainty entropy sampling algorithm

We use the idea of non-maximum suppression in our uncertainty sampling strategy [24], where the first step of the algorithm selects the most informative sample of the current model and then determines whether the similarity between this sample and any sample in the iso be labeled higher than a set threshold. If it is higher than this threshold, piecemeal is suppressed in this. Otherwise, the sample is placed in the set to be labeled L .

3. Results and Discussion

3.1. Multi-granularity active learning model and parameters

To reduce overfitting and training time, we start training by selecting a portion of the samples from the sample. We use a multi-granularity active learning network (MGAL) to extract features for image

classification. In Fig. 8, feature extraction of image information is used as image description $Des(x)$, and then the probability result $f(x)$ of classification is obtained using the softmax function. After training with partially labeled samples, the unlabeled samples are evaluated and decisions are made by Algorithm 1, and then Algorithm 2 is used to select the most useful samples.

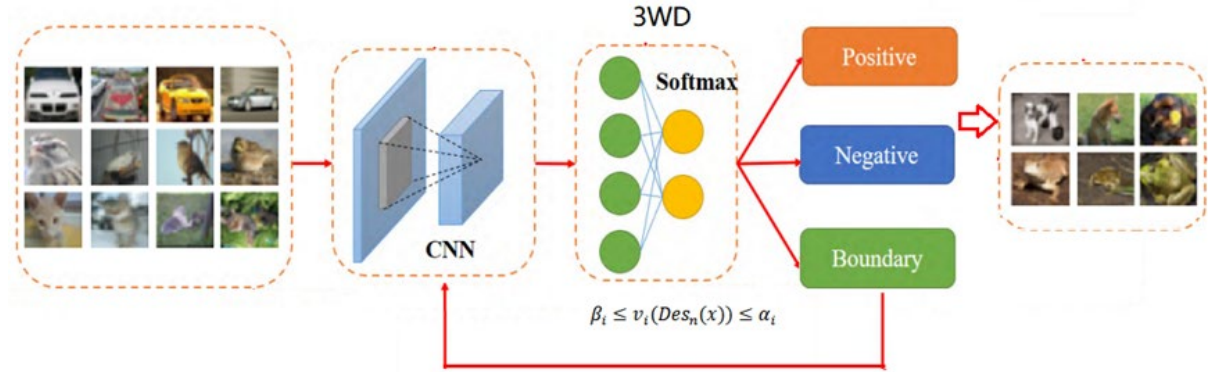


Fig. 8. Multi-granularity active learning model based on the three-way decision

The model makes a three-way decision on the classified sample set and calculates the classification cost for the three domains. The result with the lowest price is the decision result of the sample. So we add labeled samples to the previous step of the model. In each step model, these samples bring more information and make the model more accurately describe the sample set. The hyperparameters of the cost loss function in three-way decisions are shown in Table 1.

Table 1. Three-way decision cost

Decision	$L(x)=SP$	$L(x)=SN$
copy	$\lambda_{PP} = 0$	$\lambda_{PN} = 5$
Refuse	$\lambda_{NP} = 4$	$\lambda_{NN} = 0$
Delay decision	$\lambda_{BP}^i = 1 + \frac{i}{n}$	$\lambda_{BN} = 2 + \frac{i}{n}$

Confusion Matrix was used to analyse the classification results of model prediction, where TP represented the positive samples predicted into positive ones. TN is a positive class expected to be the negative class; FP is the negative samples indicated as positive samples; FN is the negative samples shown as negative ones and $TP+FP+TN+FN$ = Total number. The evaluation metrics are defined as follows.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

The following experiments compare the accuracy of the experiments when 10%, 20%, 30%, 40%, 60% and 80% of the unlabeled sample set are added.

The algorithm divides the data set into the training set(10%), the unlabeled sample set(60%) and the test set (30%). The threshold value of redundant information reduction was set at 99.9%. The CNN network calculated the output value of the posterior probability of the unlabeled sample work—the number of expected labels per iteration to 10% of the initial sample set.

3.2. Experimental result on the UCI dataset

The five publicly available datasets from the UCI dataset [25], both binary multi-classified fields, were selected for part 1 of the experiment, as shown in Table 2. Four other active learning methods were selected for comparison experiments to validate the three-way decision-based active learning method (MGAL) proposed in this paper.

- Random selection [26] (Random).
- Representation Samples [27] (RA).

- Group-based approach [28] (GA)
- Cost-effective Active Learning [29] (CEA)

Table 2. Examples of UCI public datasets

No.	Datasets	Features	Samples	Categories
1	Pima	8	768	2
2	Sonar	59	208	2
3	CMC	9	1473	3
4	Vehicle	18	846	4
5	Yeast	7	1484	10

As can be seen from Table 3, Our proposed MGAL model performs relatively well on most of the sample sets and outperforms other active learning algorithms many times with the same proportion of unlabeled samples added. Fig. 9 shows the average results of the performance metrics on different datasets, and the MGAL model achieves significant improvement for the same increment.

Table 3. Accuracy of active learning models with different increments of unmarked samples

Datasets	Algorithm	Increment for unmarked samples(%)				
	10%	20%	40%	60%	80%	
Pima	Random	0.7705	0.7769	0.789	0.7965	
	RA	0.7774	0.7876	0.7831	0.7903	
	GA	0.7745	0.7796	0.7914	0.7943	
	CEA	0.7861	0.7913	0.8033	0.8062	
	MGAL	0.8078	0.8091	0.8193	0.8357	
Sonar	Random	0.7415	0.7745	0.8321	0.8512	
	RA	0.7771	0.8109	0.828	0.8282	
	GA	0.7323	0.7732	0.8207	0.8697	
	CEA	0.7892	0.8126	0.8626	0.9141	
	MGAL	0.8022	0.8237	0.8716	0.9172	
CMC	Random	0.6509	0.6737	0.6862	0.6905	
	RA	0.6658	0.6795	0.6997	0.6822	
	GA	0.6579	0.6567	0.6688	0.671	
	CEA	0.6872	0.6867	0.7148	0.7251	
	MGAL	0.6956	0.7077	0.7297	0.7334	
Vehicle	Random	0.7874	0.7947	0.8226	0.8362	
	RA	0.7713	0.8001	0.8128	0.8325	
	GA	0.7778	0.8093	0.812	0.8343	
	CEA	0.7932	0.8193	0.8256	0.8446	
	MGAL	0.8026	0.8231	0.8313	0.8522	
Yeast	Random	0.7105	0.7251	0.7654	0.7806	
	RA	0.7348	0.7577	0.7739	0.7841	
	GA	0.7159	0.7255	0.7498	0.7889	
	CEA	0.7481	0.7581	0.7835	0.8244	
	MGAL	0.8073	0.8259	0.8379	0.8398	

The experimental results show that the evaluation results of the classification algorithm gradually become better and close to a fixed value as unlabeled samples are added. The random active learning method, which selects unlabeled selections randomly, may cause inconsistent results. The Representative active learning method, which can choose the most representative samples, also suffers from the situation that the density of samples is high, but the value is not high; the CEA algorithm takes into account the cost of different labels and is second only to our algorithm in terms of effectiveness.

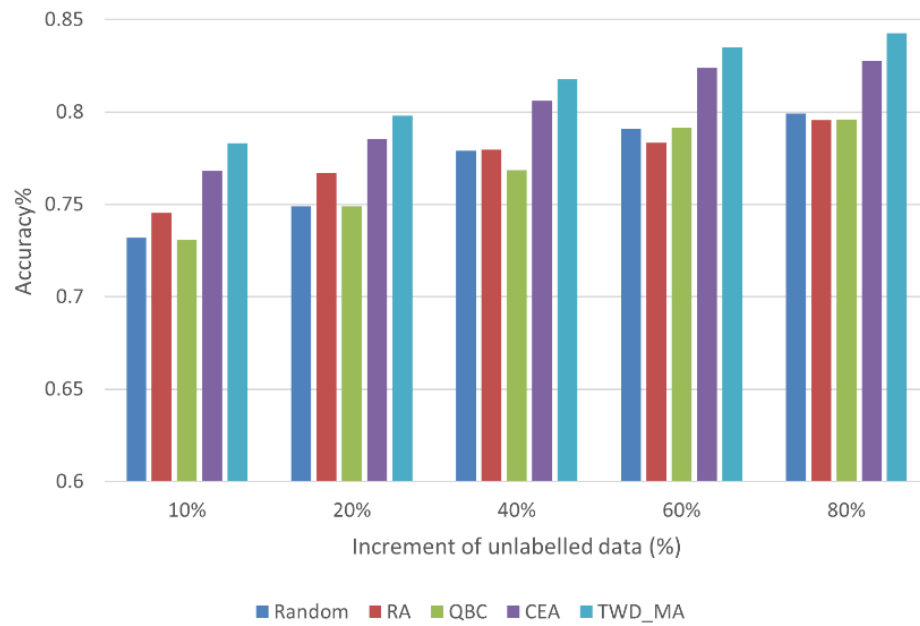


Fig. 9. Average classification accuracy on different data sets

3.3. Experiments on the dataset CIFAR10

Part 2 of the experiments compared the classification accuracy and decision cost of the three-way decision and with the classical two-way decision on the CIFAR10 dataset, which consists of 10 classes and 60,000 images, of which 80% were used for the training set, and 20% were used as the test set [30]. And a small portion of data containing labels was randomly selected as the initial training set. The similarity parameter $S1 = 0.70$ charged location for the data. The decision cost loss function parameters are shown in Table 1, and the classification accuracy and decision cost of the two-way and three-way decisions under different selection strategies were compared, as shown in Fig. 10, Fig. 11, and Fig. 12.

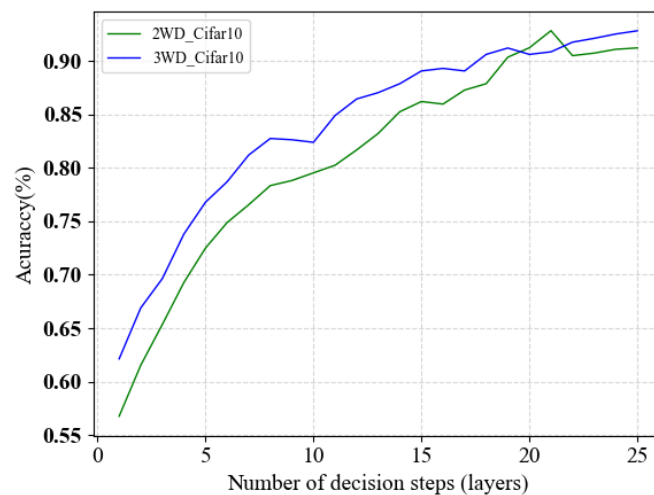


Fig. 10. Accuracy of three-way and two-way decisions

It is clear from the graph that the overall trend of the decision cost is the same for the three-way and two-way decisions. As the number of labeled samples increases and more time is spent on training, the network extracts more features, resulting in a more accurate description of the decision object and improving classification performance. For the same active learning strategy, the classification accuracy of the three-way decision is higher than that of the two-branch decision. At the same time, the cost is lower, which is the advantage of making delayed decisions with insufficient information.

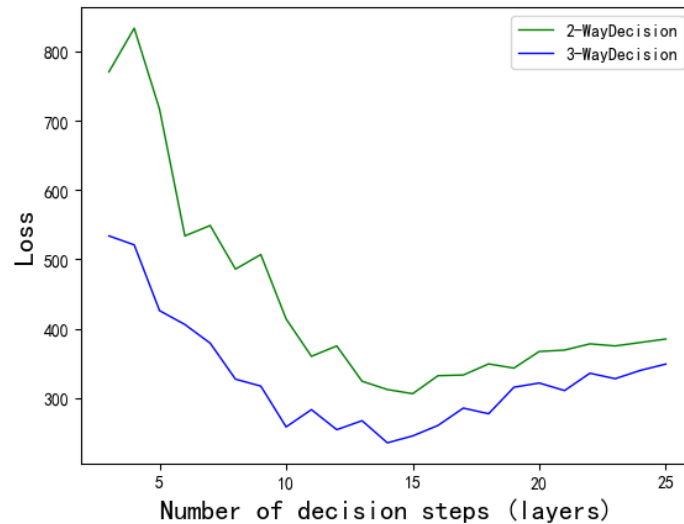


Fig. 11. Loss cost of three-way and two-way decisions

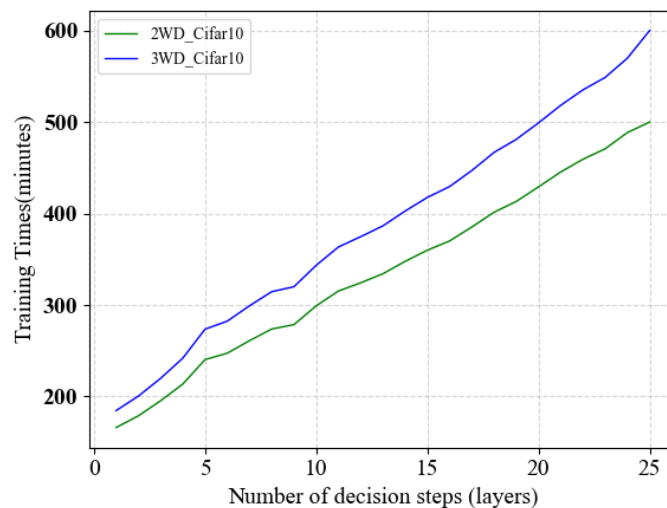


Fig. 12. Training time for three and two-way decisions

4. Conclusion

This paper proposes a new multi-granularity active learning classification model based on three-way decision theory. It can be used to label unlabeled samples and classify multi-granularity images. The three-way decision algorithm is used to divide the uncertain sample space. The suspicious samples are decomposed into three subdomains for three-way decision-making to select a representative and low-cost example and improve the decision efficiency. Compared with the two-way decision, the three-way decision can better handle the uncertainty and multi-granularity features in the samples. Experimental results on the UCI and CIFAR10 datasets show that the method significantly improves classification performance compared to other active learning methods. However, Our active learning algorithms can also increase computational complexity when dealing with high feature dimensions or large sample sizes. Future work will introduce downscaling and attribute reduction methods for complex datasets to further improve the efficiency of this model.

Acknowledgement

The author(s) thank Science Foundation of Minzu Normal University of Xingyi, China for supporting this research work.

Declarations

Author contribution. The 1st author's contribution to the paper is comprehensive and primary, with the 2nd author responsible for proofreading and advice.

Funding statement. This work was supported by the Science Foundation of Minzu Normal University of Xingyi (No. 20XYJS01).

Conflict of interest. The authors declare no conflict of interest.

Additional information. All information is private for this paper.

References

- [1] S. Liu, Y. Wang, Q. Yu, H. Liu, and Z. Peng, "CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection," *IEEE Access*, vol. 10, pp. 129116–129124, 2022, doi: [10.1109/ACCESS.2022.3228331](https://doi.org/10.1109/ACCESS.2022.3228331).
- [2] W. Wang *et al.*, "InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions," *arXiv*, pp.1-19, Nov. 2022, doi : [10.48550/arXiv.2211.05778](https://doi.org/10.48550/arXiv.2211.05778).
- [3] D. Su and P. Fung, "Improving Spoken Question Answering Using Contextualized Word Representation," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2020-May, pp. 8004–8008, May 2020, doi: [10.1109/ICASSP40776.2020.9053979](https://doi.org/10.1109/ICASSP40776.2020.9053979).
- [4] K. H. Choi and J. E. Ha, "Random Swin Transformer," *Int. Conf. Control. Autom. Syst.*, vol. 2022-November, pp. 1611–1614, 2022, doi: [10.23919/ICCAS55662.2022.10003789](https://doi.org/10.23919/ICCAS55662.2022.10003789).
- [5] W. Wang *et al.*, "mmLayout: Multi-grained MultiModal Transformer for Document Understanding," *Association for Computing Machinery (ACM).*, vol. 2022-Oct, pp. 4877–4886, doi: [10.48550/arXiv.2209.08569](https://doi.org/10.48550/arXiv.2209.08569).
- [6] J. Li, M. Wang, and X. Gong, "Transformer Based Multi-Grained Features for Unsupervised Person Re-Identification," *Proc. - 2023 IEEE/CVF Winter Conf. Appl. Comput. Vis. Work. WACVW 2023*, pp. 42–50, 2023, doi: [10.1109/WACVW58289.2023.00009](https://doi.org/10.1109/WACVW58289.2023.00009).
- [7] W. Guoyin, Y. Hong, W. Guoyin, and Y. Hong, "Multi-Granularity Cognitive Computing—A New Model for Big Data Intelligent Computing," *Front. Data Domputing*, vol. 1, no. 2, pp. 75–85, Jan. 2020, doi: [10.11871/jfdc.issn.2096-742X.2019.02.007](https://doi.org/10.11871/jfdc.issn.2096-742X.2019.02.007).
- [8] J. Chen, Z. Du, X. Sun, S. Zhao, and Y. Zhang, "A multi-granular network representation learning method," *Granul. Comput.*, vol. 6, no. 1, pp. 59–68, Jan. 2021, doi: [10.1007/s41066-019-00194-2](https://doi.org/10.1007/s41066-019-00194-2).
- [9] J. Chen, P. Wang, J. Liu, and Y. Qian, "Label Relation Graphs Enhanced Hierarchical Residual Network for Hierarchical Multi-Granularity Classification," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 4848–4857, 2022, doi: [10.1109/CVPR52688.2022.00481](https://doi.org/10.1109/CVPR52688.2022.00481).
- [10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1-14, Sep. 2015, doi: [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
- [11] S. H. Gao, M. M. Cheng, K. Zhao, X. Y. Zhang, M. H. Yang, and P. Torr, "Res2Net: A New Multi-Scale Backbone Architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021, doi: [10.1109/TPAMI.2019.2938758](https://doi.org/10.1109/TPAMI.2019.2938758).
- [12] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 936–944, Nov. 2017, doi: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
- [13] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling Vision Transformers," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 12094–12103, 2022, doi: [10.1109/CVPR52688.2022.01179](https://doi.org/10.1109/CVPR52688.2022.01179).
- [14] D. Yoo and I. S. Kweon, "Learning loss for active learning," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 93–102, Jun. 2019, doi: [10.1109/CVPR.2019.00018](https://doi.org/10.1109/CVPR.2019.00018).

- [15] J. Choi, I. Elezi, H. J. Lee, C. Farabet, and J. M. Alvarez, "Active Learning for Deep Object Detection via Probabilistic Modeling," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 10244–10253, 2021, doi: [10.1109/ICCV48922.2021.01010](https://doi.org/10.1109/ICCV48922.2021.01010).
- [16] J. Shao, Q. Wang, and F. Liu, "Learning to sample: An active learning framework," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2019–November, pp. 538–547, Nov. 2019, doi: [10.1109/ICDM.2019.00064](https://doi.org/10.1109/ICDM.2019.00064).
- [17] Y. Yao, "Three-way decision and granular computing," *Int. J. Approximate Reasoning*, vol. 103, pp. 107–123, Dec. 2018, doi: [10.1016/J.IJAR.2018.09.005](https://doi.org/10.1016/J.IJAR.2018.09.005).
- [18] A. Campagner, F. Cabitza, and D. Ciucci, "Three-Way Decision for Handling Uncertainty in Machine Learning: A Narrative Review," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12179 LNAI, pp. 137–152, 2020, doi: [10.1007/978-3-030-52705-1_10](https://doi.org/10.1007/978-3-030-52705-1_10).
- [19] Y. Yao, "Tri-level thinking: models of three-way decision," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 5, pp. 947–959, May 2020, doi: [10.1007/S13042-019-01040-2](https://doi.org/10.1007/S13042-019-01040-2).
- [20] J. Qian, D. W. Tang, Y. Yu, X. B. Yang, and S. Gao, "Hierarchical sequential three-way decision model," *Int. J. Approx. Reason.*, vol. 140, pp. 156–172, Jan. 2022, doi: [10.1016/j.ijar.2021.10.004](https://doi.org/10.1016/j.ijar.2021.10.004).
- [21] W. Qian, Y. Zhou, J. Qian, and Y. Wang, "Cost-sensitive sequential three-way decision for information system with fuzzy decision," *Int. J. Approx. Reason.*, vol. 149, pp. 85–103, Oct. 2022, doi: [10.1016/j.ijar.2022.07.006](https://doi.org/10.1016/j.ijar.2022.07.006).
- [22] P. Liu, L. Wang, R. Ranjan, G. He, and L. Zhao, "A Survey on Active Deep Learning: From Model Driven to Data Driven," *Association for Computing Machinery (ACM)*, vol. 54, no. 10, pp. 1–34, Sep. 2022, doi: [10.1145/3510414](https://doi.org/10.1145/3510414).
- [23] C. Mayer and R. Timofte, "Adversarial sampling for active learning," *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 3060–3068, Mar. 2020, doi: [10.1109/WACV45572.2020.9093556](https://doi.org/10.1109/WACV45572.2020.9093556).
- [24] Y. Yang and M. Loog, "A benchmark and comparison of active learning for logistic regression," *Pattern Recognit.*, vol. 83, pp. 401–415, Nov. 2018, doi: [10.1016/j.patcog.2018.06.004](https://doi.org/10.1016/j.patcog.2018.06.004).
- [25] Dua, D. and Graff, C. "UCI Machine Learning Repository". Irvine, CA: University of California, School of Information and Computer Science, 2019, Available at : archive.ics.uci.edu/ml.
- [26] P. Ren *et al.*, "A Survey of Deep Active Learning," *Association for Computing Machinery (ACM)*, vol. 54, no. 9, pp. 1–40, Oct. 2021, doi: [10.1145/3472291](https://doi.org/10.1145/3472291).
- [27] X. Yan *et al.*, "A clustering-based active learning method to query informative and representative samples," *Appl. Intell.*, vol. 52, no. 11, pp. 13250–13267, Sep. 2022, doi: [10.1007/S10489-021-03139-Y](https://doi.org/10.1007/S10489-021-03139-Y).
- [28] B. Settles and M. Craven, "An Analysis of Active Learning Strategies for Sequence Labeling Tasks," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ACL Press, 2008, pp. 1070–1079, doi: [10.5555/1613715.1613855](https://doi.org/10.5555/1613715.1613855).
- [29] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, "Cost-Effective Active Learning for Deep Image Classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2591–2600, Dec. 2017, doi: [10.1109/TCSVT.2016.2589879](https://doi.org/10.1109/TCSVT.2016.2589879).
- [30] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," *Computer Science University Of Toronto*, pp. 1–60, 2009. Available at : <http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf>.