

Neural Decoding Based on Active Learning for Intracortical Brain-Machine Interfaces

Weihua Wang, Jian Huang*, Dongrui Wu

School of Artificial Intelligence and Automation
Huazhong University of Science and Technology (HUST)
Wuhan, Hubei, China
huang_jian@hust.edu.cn

Peng Zhang

Wuhan National Laboratory for Optoelectronics
Huazhong University of Science and Technology (HUST)
Wuhan, Hubei, China

Abstract—In intracortical brain-machine interfaces (iBMIs), it is time-consuming and expensive to label the large number of unlabeled samples. In this paper, three greedy sampling active learning algorithms, named denoised greedy sampling on the inputs (DGSx), denoised greedy sampling on the outputs (DGSy) and denoised improved greedy sampling (DiGS), were proposed to solve the problem of labeling samples. In iBMIs, in order to reduce the influence of abnormal points in the raw data, One-class-SVM was assumed to denoise to improve the performance of original greedy sampling algorithms and achieve stable or robust decoding. Compared with Query by Committee (QBC) and Uncertainty Sampling (US), the proposed approaches achieved higher accuracies. The efficiency of proposed approaches was demonstrated by the experiment with rhesus' electrophysiological signals in iBMIs.

Index Terms—intracortical brain-machine interface, active learning, denoise, rhesus macaque

I. INTRODUCTION

Intracortical brain-machine interfaces (iBMIs), which establish a connection between brain and external devices, repair and expand motor functions in paralyzed patients and amputees [1], [2]. iBMIs can record the electrical signal of neural activities from the cerebral cortex through the microelectrode array, which has the advantages of high resolution and rich motion information, making it possible to control the external devices with multiple degrees of freedom [3]–[5]. This paper is about the intracortical brain-machine interfaces based on spike, and the research object is rhesus macaques. Because of their similar anatomical nerve structure of brain with human, the study of their brain nerve mechanism will accumulate enough practical experience and make contribution to the future of intracortical interfaces to clinical application.

Since Schmidt et al. showed that signals from neurons in the cortex of monkeys could be recorded over a long period of time, and external devices could be controlled by release rate of a single spike in the primary motor cortex [6], [7], the pioneering work in this field has begun. However, there are some challenges to choose the monkeys as our research object. If we attempt to monitor the motion of the monkeys, the most intuitive way is installing sensors on them and get the corresponding labels. But in practice, this approach will run into difficulties because it will cause monkeys to be emotionally unstable. Eventually, they may destroy the sensors in their

bodies. Thus, we end up with a lot of data without labels and have to label it. So, it is a significant work to improve the performance of classifiers in this field.

There are some works trying to improve the performance of classifiers. By the leave-one-subject-out cross-validation procedure and l_1 regularization, Fazli et al. proposed a transfer learning method in EEG classification tasks and found common brain activity patterns in classifier aggregation step [10]. Wang et al. proposed a series of reinforcement learning methods, attention gated reinforcement learning (AGREL) and quantized attention gated reinforcement learning (QAGKRL), to adaptively reconstruct 2D continuous trajectory in a center-out task, solving the problem that Q-learning is difficult to generalize [11]–[13]. On the one hand, when we monitor neural signals of monkeys, small size of data will easily miss key information. And on the other hand, to train a decoder, most of these approaches need a large number of labeled samples, which is time-consuming to be labeled.

Then, is there a method which can train the decoder using a small number of labeled samples while maintaining its performance? Active learning is a valuable direction for this question. It selects the most valuable unlabeled samples to label rather than random label.

Active learning is a subfield of machine learning. Its key hypothesis is that if the learning algorithm is allowed to choose the data actively, the classifier or regression model would perform better with less training. This property of active learning makes it easy to solve the problem of traditional supervised machine learning which is trained on hundreds of instances—it's expensive or time-consuming. Active learning makes it possible to train a better classifier using fewer labeled samples. A query function Q selects one or a batch of the most useful samples from unlabeled sample pool U , and asks the supervisor S to label them. One or a batch of the most useful samples from U were selected by the query function and labeled by S . Finally, the classifier C was trained by the samples in the set of labeled samples L , and next query was conducted. Q is the key point of active learning. For different query functions, active learning generally includes three learning scenarios, namely membership query synthesis, stream-based selective sampling and pool-based sampling [9]. The difference of these three scenarios is showed in Fig. 1.

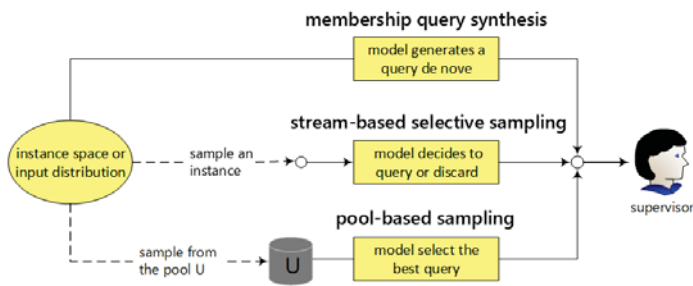


Fig. 1. Three kind of active learning scenarios

There are three widely used criteria in query function, informativeness, representativeness and diversity. Informativeness measures how well an unlabeled instance which helps reduce the uncertainty of a statistical model, whereas representativeness measures how well an instance which helps represent the structure of input patterns [18]. Diversity tends to select samples which are full of all the input space, instead of concentrating in a small corner. Query-by-committee (QBC) and uncertainty sampling are two famous and important approaches in active learning. Seung [17] proposed a query selection framework named query-by-committee (QBC) which involves maintaining a committee $\mathcal{C} = \{\theta^{(1)}, \dots, \theta^{(C)}\}$. The committee members vote for the samples, and the samples with the biggest vote entropy will be selected. Uncertainty sampling trains only one learner and selects the samples with least confidence. Wu proposed sequential pool-based active learning for regression (ALR) [8]. It is a greedy sampling algorithm since the selected samples are most representative and diverse. However, these strategies make it easy to select outliers meanwhile.

In this study, we bring the idea of ALR into iBMIs creatively, suit it to the classification problem and improved its performance. To solve the problem that it is easy to choose outliers, we proposed to denoise the raw data, which is benefit to improve classifier's accuracy. Finally, we compare the results with classical active learning methods (QBC and uncertainty sampling) and verify the validity of this method. This method selects the most representative or diverse samples in the data pool to label, thus reducing the workload of labeling those numerous samples, and allow using numerous samples and experts to label them by reviewing the video.

II. METHODS

This section introduce the experimental setup and the algorithm

A. Experimental setup and Electrophysiological recordings

All the experiments and surgical procedures relating to this study were approved by the Institutional Animal Care and Use Committee at Huazhong University of Science and Technology.

We use an adult rhesus macaque in the experiment. The monkey with a restricted left arm is guide by experimental apparatuses and will perform reaching and grasping tasks using its another hand. The experimental apparatus (showed in Fig. 2)

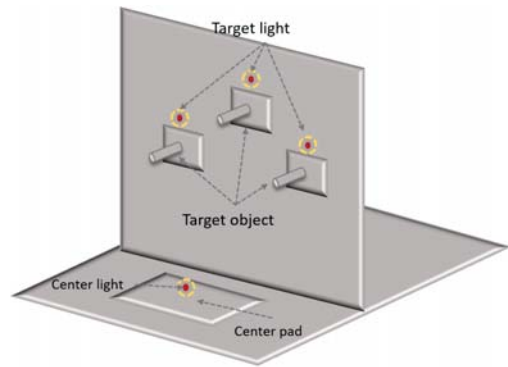


Fig. 2. The experiment set of the monkey

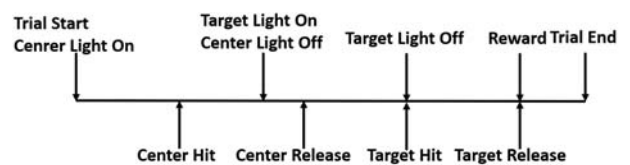


Fig. 3. The sequence of a trial for the monkey

consists of three same target objects of the front panel and a center pad of the bottom. The sequences of the experiment are showed in Fig. 3. For each trail, the monkey was noticed to look at the center pad at the time of the center light on. The center will hold for about 500ms, then the monkey will release its hand from the center pad (named 'center release', CR), and reach the corresponding target object and grasp it when one of the target lights goes up. We define the event when the monkey grasped the target object as 'target hit' (TH). In the end, the monkey got a few drops of water as a reward after a target holding time. The above description was a successful trial. Otherwise the trail will be aborted if the monkey made mistakes [14].

We surgically implanted a 16-channel FMA array (Microprobe Inc.) and two 32-channel Utah arrays (Blackrock Microsystems) into the somatosensory cortex (S1), primary motor cortex (M1), and posterior parietal cortex (PPC) respectively to a monkey. We record neural signals by all 80 channels of the three arrays and use a 128-channel OmniPlex system (Plexon, Inc.) to record monkey's behavioral data and neural signals and filtered the signals between 250 Hz and 6 kHz. Then the sampling rate was 40 kHz. After that, threshold detecting was used to collect spike counts. We choose time periods of 200ms before and after CR to decode motor intention and classify which position the monkey would grasp (three-category classification). In each trial, a neural activity vector (NAV) was extracted from neural signals as the features [15], as is shown in Fig. 4. A window size of 400ms was used with a bin size of 50ms and a total of 80 units were selected, and then compose the NAV with 640 features.

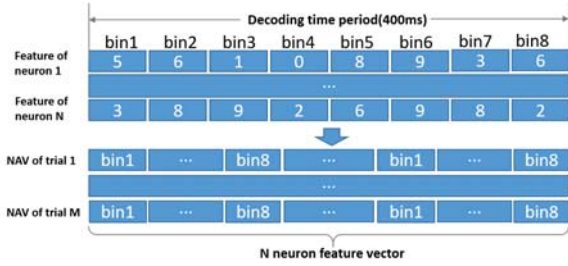


Fig. 4. The structure of neural data

B. Greedy sampling active learning

In iBMIs, it is not difficult to obtain a great deal of data, but relatively expensive to label them. In this case, active learning can proactively select the suitable samples to label. Yu and Kim [16] proposed a series passive sampling approaches for regression. In each iteration, these approaches require to update the regression model and predictions of unlabeled samples. Wu modified the Greedy Sampling (GS) slightly, and named it as GSx, where x means the input space [8]. Besides, he also proposed Greedy Sampling on output space (named GSy) and improved Greedy Sampling (named iGS). The core idea of these active learning methods is to select the most representative samples. In this subsection, we will briefly introduce these three algorithms and suit them to the classification problem in iBMIs.

1) *GSx*: The key idea of GSx is to select the most diverse samples in the input space to label. We want to select K samples from the unlabeled samples pool, which has N samples $\{\mathbf{x}_n\}_{n=1}^N$. We select the first sample which is closest to the centroid of all N samples, and label it. Then, we attempt to select the remaining $K-1$ samples incrementally. It is crucial to make sure the current selected sample the most diverse by computing its distance to the other samples. Finally, GSx trains the model by these selected K samples.

In general, the first k samples are assumed to have been selected. There is the remaining $N-k$ samples, $\{\mathbf{x}_n\}_{n=k+1}^N$ in the pool, and all of them is unlabeled. GSx decides the selected samples by computing its distance to each of the k labeled samples:

$$d_{nm}^x = \|\mathbf{x}_n - \mathbf{x}_m\|, \quad m = 1, \dots, k; n = k + 1, \dots, N \quad (1)$$

Choose the smallest from d_{nm}^x , named d_n^x , as the shortest distance from \mathbf{x}_n to the other labeled samples:

$$d_n^x = \min_m d_{nm}^x, \quad n = k + 1, \dots, N \quad (2)$$

Finally, select the sample with the maximum d_n^x to label.

In conclusion, the first selected sample is closest to the centroid to label. In each subsequent iteration, the most representative sample (farthest to the labeled samples in input space) is selected to label.

2) *GSy*: GSy pays attention to the diversity in the output space instead of the input space.

In order to select K samples to label, K_0 samples are selected by GSx firstly, where K_0 is the minimum number required to train a classifier or regression model. Next, assumed k ($k \geq K_0$) samples have been selected by GSy, and labeled with outputs $\{y_m\}_{m=1}^k$. Then, select the remaining $K-k$ samples:

$$d_{nm}^y = \|f(\mathbf{x}_n) - y_m\|, \quad m = 1, \dots, k; n = k + 1, \dots, N \quad (3)$$

$$d_n^y = \min_m d_{nm}^y, \quad n = k + 1, \dots, N \quad (4)$$

select the sample with the maximum d_n^y to label. In the above equation, different from regression problem in [8], $f(x)$, in the form of $[p_0, p_1, p_2]$, is the result of the a neural network classifier, which represents each probability of every class (the highest probability is then used as the category label). Similarly, y_m is also in the form of $[p_0, p_1, p_2]$ computing by the classifier. This intermediate value can well reflect the specificity of network output, so it is decided to use the norm of the intermediate values vector to measure the specificity of output value and measure the diversity of the output space.

In brief, GSy selects a few samples by GSx firstly. In every subsequent iteration, to achieve diversity, we select the sample which is furthest to all selected samples before in output space.

3) *iGS*: Sometimes, GSy can't guarantee the diversity of the most sensitive predictor as expected for it only consider diversity of output space. Thus, we improved its performance by combine input and output space, and named it as iGS.

Similarly, we select first K_0 samples by GSx, where K_0 is the minimum number required to train a classifier or regression model (sometimes set it to be the number of features in the input space). Next, assumed k ($k \geq K_0$) samples have been selected by iGS, and labeled with outputs $\{y_n\}_{n=k+1}^N$. To select the next $K-k$ samples, iGS computes first d_{nm}^x in Eq. 1 and d_{nm}^y in Eq. 3, and d_{nm}^{xy} by following equation:

$$d_{nm}^{xy} = \min_m d_{nm}^x d_{nm}^y, \quad n = k + 1, \dots, N \quad (5)$$

select the sample with the maximum d_n^{xy} to label.

Since $\|f(\mathbf{x}_n) - y_m\|$ cannot be directly used to represent the diversity of samples in output space like [8], we also use the intermediate output $[p_0, p_1, p_2]$ of neural network to represent each probability of every class.

In summary, iGS selects K_0 samples using GSx firstly. In order to achieve diversity among the selected samples, we selected the sample which is furthest to all selected samples before in the input and output space, in every subsequent iteration.

C. Denoise raw data

The GSx, GSy and iGS, these algorithms are based on greedy sampling to select samples to label. However, this idea guarantees the diversity of samples, while making it easy to select the abnormal points, especially when the sample scale is

relatively small. When there are "abnormal points" in the raw data, these data cannot represent the whole data set. However, because of the large difference from normal data, these points are easy to be selected by the active learning algorithm, which leads to the unstable prediction accuracy in the early and middle stages. In next section, it will be exhibited that the classification accuracy fluctuates greatly when the number of sample selection is small (less than 5%), even lower than the accuracy of random selection.

One-class-SVM is suitable for denoising in a certain proportion, or in cases where most of the known training samples are positive samples and few are negative samples. One-class-SVM has a good application effect in the problem of denoising only one kind of data. In this study, we used One-class-SVM to denoise the original data, and then applied three active learning algorithms to realize the classification of the electrophysiological data in iBMIs. This method is named denoised GSx (DGSx). For clarity, we have listed the pseudocode for DGSx in Algorithm 1:

Algorithm 1 The DGSx AL algorithm

Input: N unlabeled samples, $\{\mathbf{x}_n\}_{n=1}^N$
 K , the maximum number to query

Output: The classification model $f(\mathbf{x})$

- 1: // Raw data denoising
- 2: Set $Z = \{\mathbf{x}_n\}_{n=1}^N$, and $S = \emptyset$
- 3: Remove the abnormal points from Z by One-class-SVM, the sample set after denoising is $Z' = \{\mathbf{x}_n\}_{n=1}^{N'}$
- 4: // Initialize the first selection
- 5: Identify \mathbf{x}' , the closest sample to the centroid of Z'
- 6: Move \mathbf{x}' from Z' to S
- 7: Re-index the sample in S as \mathbf{x}_1 , and the samples in Z' as $\{\mathbf{x}_n\}_{n=2}^{N'}$;
- 8: // Select $K - 1$ more samples incrementally
- 9: **for** $k = 1, \dots, K - 1$ **do**
- 10: **for** $n = k + 1, \dots, N'$ **do**
- 11: Compute d_n^x in (2)
- 12: **end for**
- 13: Identify the \mathbf{x}' that has the largest d_n^x ;
- 14: Move \mathbf{x}' from Z' to S ;
- 15: Re-index the samples in S as $\{\mathbf{x}_m\}_{m=1}^{k+1}$, and the samples in Z' as $\{\mathbf{x}_n\}_{n=k+2}^{N'}$;
- 16: **end for**
- 17: Query to label all K samples in S ;
- 18: Construct the classification model $f(\mathbf{x})$ from S .

Similarly, the pseudocode of denoised GSy (DGSy) is shown in Algorithm 2:

Finally, the pseudocode of denoised iGS (DiGS) is in Algorithm 3:

III. RESULTS

A. Greedy sampling active learning in the iBMIs

This section will demonstrate the application of three active learning algorithms in iBMIs. Four days of monkey

Algorithm 2 The DGSy ALR approach.

Input: N unlabeled samples, $\{\mathbf{x}_n\}_{n=1}^N$
 K , the maximum number to query

Output: The classification model $f(\mathbf{x})$

- 1: // Raw data denoising
- 2: Set $Z = \{\mathbf{x}_n\}_{n=1}^N$, and $S = \emptyset$
- 3: Remove the abnormal points from Z by One-class-SVM, the sample set after denoising is $Z' = \{\mathbf{x}_n\}_{n=1}^{N'}$
- 4: // Initialize the first selection
- 5: Identify \mathbf{x}' , the closest sample to the centroid of Z'
- 6: Move \mathbf{x}' from Z' to S
- 7: Re-index the sample in S as \mathbf{x}_1 , and the samples in Z' as $\{\mathbf{x}_n\}_{n=2}^{N'}$;
- 8: // Select $K_0 - 1$ more samples incrementally using GSx
- 9: Identify K_0 , the minimum number of labeled samples required to construct $f(\mathbf{x})$;
- 10: **for** $k = 1, \dots, K_0 - 1$ **do**
- 11: **for** $n = k + 1, \dots, N'$ **do**
- 12: Compute d_n^x in (2)
- 13: **end for**
- 14: **end for**
- 15: Query to label the K_0 samples in S ;
- 16: Construct the classification model $f(\mathbf{x})$ from S ;
- 17: // Select $K - K_0$ more samples incrementally
- 18: **for** $k = K_0, \dots, K - 1$ **do**
- 19: **for** $n = k, \dots, N$ **do**
- 20: Compute d_n^y in (4);
- 21: **end for**
- 22: Identify the \mathbf{x}' that has the largest d_n^y ;
- 23: Move \mathbf{x}' from Z' to S ;
- 24: Re-index the samples in S as $\{\mathbf{x}_m\}_{m=1}^{k+1}$, and the samples in Z' as $\{\mathbf{x}_n\}_{n=k+2}^{N'}$;
- 25: Update the classification model $f(\mathbf{x})$ using S .
- 26: **end for**

electrophysiological signals, named D1/D2/D3/D4, were used in the experiment. As previously mentioned, each NAV has 640 features, and the electrophysiological data of monkey consists of 630 NAV. So, the input data is 630×640 dimensional. A part of the 630 samples is selected to label to train the classifier, and then the rest is used as the test set. The three algorithms proposed were compared in this section (in Fig. 5). Generally, as K increased, it's intuitive that all algorithms performed better, because more labeled training samples generally lead to a more reliable classifier. As is shown in Fig. 5 :

(1) Compared with random algorithm, GSx, GSy and iGS have advantage if the scale of selected data is large (more than 10%), and the accuracy is improved by 5% to 10%. But random selection performs better in the case of small data scale (less than 5%).

(2) All proposed algorithms perform better than random selection. This means GS active learning strategy will make it easy to label data in iBMIs.

(3) From the best to the worst, the average performances

Algorithm 3 The DiGS ALR approach.

Input: N unlabeled samples, $\{\mathbf{x}_n\}_{n=1}^N$
 K , the maximum number to query

Output: The classification model $f(\mathbf{x})$

- 1: // Raw data denoising
 - 2: Set $Z = \{\mathbf{x}_n\}_{n=1}^N$, and $S = \emptyset$
 - 3: Remove the abnormal points from Z by One-class-SVM, the sample set after denoising is $Z' = \{\mathbf{x}_n\}_{n=1}^{N'}$
 - 4: // Initialize the first selection
 - 5: Identify \mathbf{x}' , the closest sample to the centroid of Z'
 - 6: Move \mathbf{x}' from Z' to S
 - 7: Re-index the sample in S as \mathbf{x}_1 , and the samples in Z' as $\{\mathbf{x}_n\}_{n=2}^{N'}$;
 - 8: // Select $K_0 - 1$ more samples incrementally using GSx
 - 9: Identify K_0 , the minimum number of labeled samples required to construct $f(\mathbf{x})$;
 - 10: **for** $k = 1, \dots, K_0 - 1$ **do**
 - 11: **for** $n = k + 1, \dots, N'$ **do**
 - 12: Compute d_n^x in (2)
 - 13: **end for**
 - 14: **end for**
 - 15: Query to label the K_0 samples in S ;
 - 16: Construct the classification model $f(\mathbf{x})$ from S ;
 - 17: // Select $K - K_0$ more samples incrementally
 - 18: **for** $k = K_0, \dots, K - 1$ **do**
 - 19: **for** $n = k, \dots, N$ **do**
 - 20: Compute d_n^{xy} in (5);
 - 21: **end for**
 - 22: Identify the \mathbf{x}' that has the largest d_n^y ;
 - 23: Move \mathbf{x}' from Z' to S ;
 - 24: Re-index the samples in S as $\{\mathbf{x}_m\}_{m=1}^{k+1}$, and the samples in Z' as $\{\mathbf{x}_n\}_{n=k+2}^{N'}$;
 - 25: Update the classification model $f(\mathbf{x})$ using S .
 - 26: **end for**
-

of the four algorithms was $iGS > GSy > GSx > random$. iGS outperformed other algorithms with the stable accuracy over 98%.

B. Denoised greedy sampling active learning

The outliers in the original data will influence the selecting results of active learning algorithm, leading to accuracy fluctuations. It is especially obvious in the case of small size of data. Therefore, we adopted One-class-SVM to denoise raw data, and apply denoised greedy sampling algorithms. We removed the outliers, and projected it to 2D space by Principal Component Analysis (PCA). As is shown in Fig. 6:

After removing outliers, the accuracy of classifier went up and accuracy fluctuates went down obviously. As a contrast, classical active learning methods, QBC and US were also used to classify neural signals. The result is shown in Fig. 7, Fig. 8 and Fig. 9. We can conclude that:

(1) As was expected, the fluctuation of accuracy decreased after denoising, especially when selected data is less than 5%. The result is clearly shown in Fig. 7.

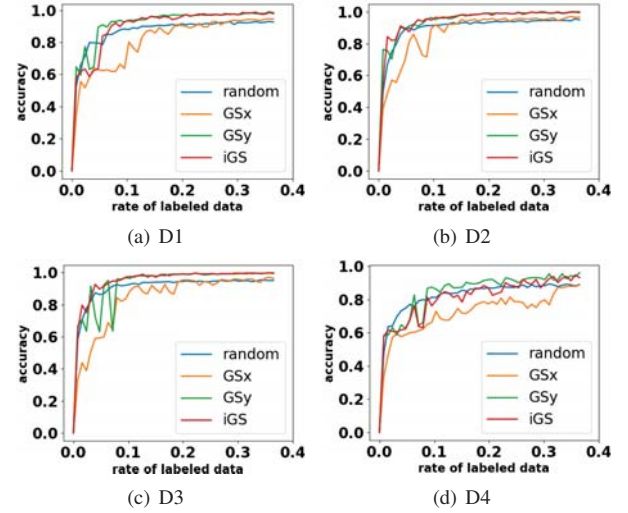


Fig. 5. The comparison accuracy of GSX/GSY/iGS/random (a)D1 (b)D2 (c)D3 (d)D4

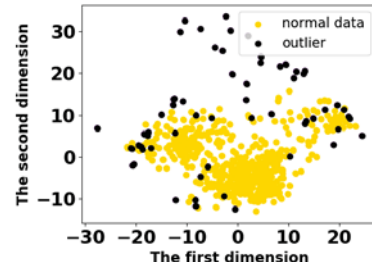


Fig. 6. Denoised original data and diminished by PCA to 2D space

(2) From best to worst, the ranking of the algorithms was $DiGS > DGSy > QBC > US > DGSx > random$. On average, the accuracy of $DGSy$ or $DiGS$ is 10% to 15% more accurate than QBC and US when selected data is less than 5%, as Fig. 8.

(3) Consistent with the above description, $DGSx/DGSy$ outperformed random selection and QBC/US . The average maximum accuracy of each algorithm is shown in Fig. 9.

(4) The standard deviations of $DGSx/DGSy/DiGS$ are smaller than classical active learning algorithms (QBC and US). Understandably, it is denoising raw data that results in stable performance of the classifiers.

After removing abnormal data, all of the three kinds of classification accuracy are improved, and the accuracy fluctuates decreased significantly at the same time. Through experiments, it is easy to see that the proposed algorithms outperformed classical active learning algorithms (QBC and US) in neural signal classification, verifying the validity of them.

IV. CONCLUSIONS

Beneficially, iBMIs translate brain signals into commands to control external devices. But in practice, it would be difficult to obtain an accurate classifier without a large number of samples. At the same time, it is time-consuming to label these considerable unlabeled samples. In this paper, we creatively

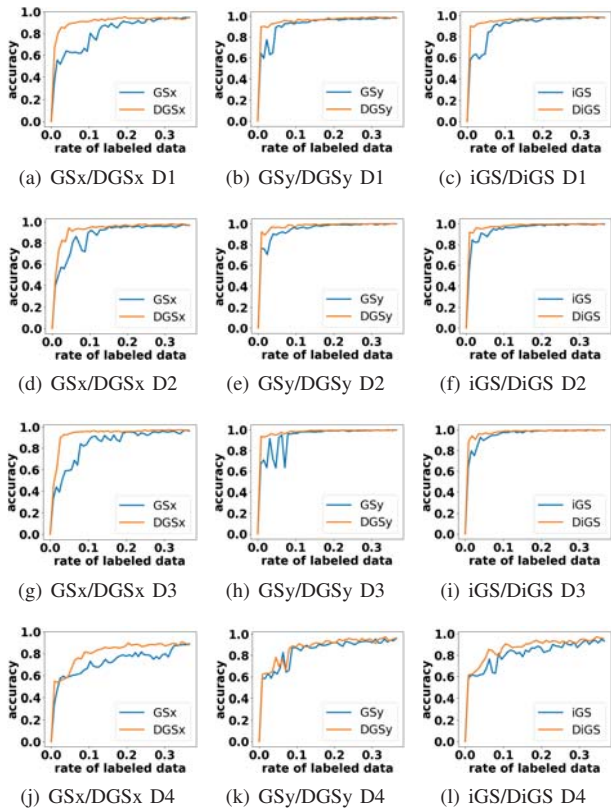


Fig. 7. The performances of DGSx/DGSy/DiGS

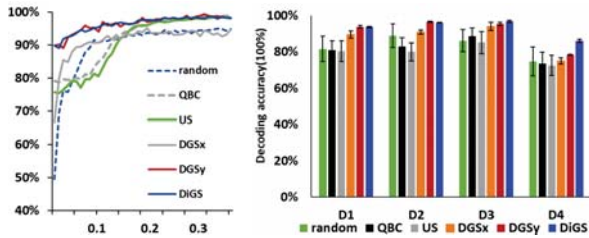


Fig. 8. Performance of six algorithms. (a) The accuracy of six algorithms. (b) Classification accuracy when the queries is 5%

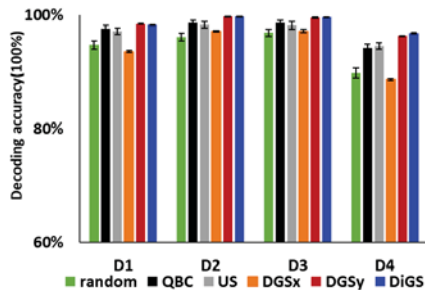


Fig. 9. Comparison of maximum accuracy

put forward the greedy sampling algorithm in iBMIs. However, they tend to select the abnormal points in the raw data, which will reduce the performance of classifier. We proposed three denoised greedy sampling algorithms (DGSx, DGSy and DiGS), improving the performance of original algorithms. Denoising ensures the accuracy and stability of decoding, especially when the labeled data is less than 5%, which is significant in practice. Compared with classical active learning approaches, QBC and US, the outstanding advantages of our algorithms are verified.

Besides classification problems, we will also try to combine active learning with transfer learning for regression problems.

ACKNOWLEDGMENT

The work was supported by the International Science and Technology Cooperation Program of China under Grant No. 2017YFE0128300 and the Technology Innovation Project of Hubei Province of China under Grant 2019AEA171.

REFERENCES

- [1] J. Vansteensel, J. Mariska, et al, "Fully Implanted Brain-Computer Interface in a Locked-In Patient with ALS.", *The New England journal of medicine*, vol.375, no.21, pp.2060-2066, 2016.
- [2] Hochberg, R. Leigh, et al, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia.", *Nature*, vol.442, no.7099, pp.164-271, 2006.
- [3] T. Pistohl, Schulze-Bonhage, et al, "Decoding natural grasp types from human ECoG.", *NeuroImage*, vol.59, no.1, pp.248-260, 2012.
- [4] J. Zhuang, W. Truccolo, et al, "Decoding 3-D reach and grasp kinematics from high-frequency local field potentials in primate primary motor cortex.", *IEEE transactions on bio-medical engineering* vol.57, no.7, pp.1774-1784, 2010
- [5] M. Velliste, S. Perel, et al, "Cortical control of a prosthetic arm for self-feeding.", *Nature*, vol.453, no.7198, pp.1098-1101, 2008
- [6] E. E. Fetz, D. V. Finocchio, "Correlations between activity of motor cortex cells and arm muscles during operantly conditioned response patterns.", *Experimental brain research*, vol.23, no.3, pp.217-240, 1975.
- [7] E. M. Schmidt, J. S. McIntosh, "Fine control of operantly conditioned firing patterns of cortical neurons.", *Experimental neurology*, vol.61, no.2, pp.349-369, 1978.
- [8] D. Wu, C.T. Lin, and J. Huang, "Active Learning for Regression Using Greedy Sampling.", *Information Sciences*, vol.474, pp.90-105, 2019.
- [9] B. Settles, "Active Learning Literature Survey.", University of Wisconsin Madison, Jan.26, 2010
- [10] S. Fazli, F. Popescu et al, "Subject-independent mental state classification in single trials.", *Neural Networks the Official Journal of the International Neural Network Society*, vol.22, no.9, pp:1305-12, 2009.
- [11] Y. Wang, F. Wang, et al, "Neural Control of a Tracking Task via Attention-Gated Reinforcement Learning for Brain-Machine Interfaces.", *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, 1-1, 2014.
- [12] F. Wang, Y. Wang, et al, "Quantized Attention-Gated Kernel Reinforcement Learning for Brain-Machine Interface Decoding.", *IEEE transactions on neural networks and learning systems*, vol.28, no.4, pp.873-886, 2015.
- [13] H. Li, F. Wang, Q. Zhang, et al, "Maximum correntropy based attention-gated reinforcement learning designed for brain machine interface.", *Conf Proc IEEE Eng Med Biol Soc*. pp:3056-3059, 2016.
- [14] P. Zhang, X. Ma, H. Huang H, J. He, "Predicting hand orientation in reach-to-grasp tasks using neural activities from primary motor cortex.", *Conf Proc IEEE Eng Med Biol Soc*. pp.1306-1309, 2014.
- [15] J. Hu, J. Si, B. P. Olson, J. He, "Feature detection in motor cortical spikes by principal component analysis.", *IEEE Trans Neural Syst Rehabil Eng*, vol.13, no.3, pp.256-262, 2005.
- [16] H. Yu and S. Kim, "Passive Sampling for Regression.", 2010 IEEE International Conference on Data Mining, pp.1151-1156
- [17] H. S. Seung, M. A. Opper, H. Sompolinsky, "Query by Committee.", *Proc of the Fith Workshop on Computational Learning Theory*, vol.284, pp.287-294, 1992.
- [18] Z. Zhou, "A brief introduction to weakly supervised learning.", *National Science Review*, vol.5, no.1, pp.44-53, 2018.