

The function of machine learning BCI classifiers

Bowei Tian

School of Cyber Science and Engineering, Wuhan University, China

2020302192183@whu.edu.cn

Abstract. Recently, machine learning(ML) has become a hot issue, and most brain-computer interface(BCI) systems contains machine learning structured classifiers. The classifiers are designed to process the selected feature and send the most possible signal user generates to the reception device. However, it's not easy to realize without some advanced ML methods. This survey mainly focuses on some effective ML means to classify features, including state-of-the-art invention: MDM classifier. Moreover, we propose some promising directions to further research this topic.

Keywords: ML, BCI, classifiers.

1. Introduction

Nowadays, the brain-computer interface has developed enormously due to its powerful connections between human and machine. The purpose of BCI is to allow users with serious motor disabilities to have effective and automatic control over devices such as cleaners, computers, neural prostheses and furniture. Such an interface will increase an individual's independence, upgrade the quality of life and reduce social costs[1].

It's widely known that BCI devices has two functions. One function is recording data on the electrode, the other is for encoding and decoding neural signals[2]. And the way we classify signals into decisions are necessary in the process. In recent years, as deep learning algorithms evolved rapidly[18], huge amounts of methods have sprung like mushrooms, so as classifiers.

Classifiers have been researched by scholars for decades, especially in the post-deep-learning era. Multi-layer perceptron(MLP) [3, 4, 7, 17] is a forward structure artificial neural network, including input layer, output layer and several hidden layers; k-Means [5, 8, 9] is the most commonly used clustering algorithm based on Euclidean distance, which believes that the closer the distance between two targets, the greater the similarity. However, these traditional methods can't perform well in the BCI systems. And here are some reasons for this:

1.As for traditional supervised learning, it is sometimes hard for users to gain such a huge amount of labeled data[13], especially in the BCI settings. Due to the limitation of labeled data, it is a better choice to use unsupervised learning methods;

2.k-means algorithm and some basic unsupervised learning methods are hard to fit by gradient descent. Specifically, only convex datasets works;

3.if the datasets is not balanced, i.e. positive data or negative data is in the majority, the classification result won't be as good as expected.

According to these dilemmas, we put effort into these problems by first introducing some basic features of the classification in BCI in the second section. After that, we explore deeper into two classification methods separately in the third and fourth section. Finally, we'll draw some conclusions and analyze the future directions according to state-of-the-art methods.

2. Feature classification overview

2.1. Processing schemes in BCI systems

We summarize the processing schemes that have been used in BCI systems. It mainly focuses on the following signal processing components of a BCI: the preprocessing method, feature extraction, feature selection, dimensionality reduction, feature classification and post-processing blocks. The feature classification step begins after the processing of feature selection step, in which states we ought to map the weights on the features and decide one gaining the highest score. And the entire process is shown in Figure 1.

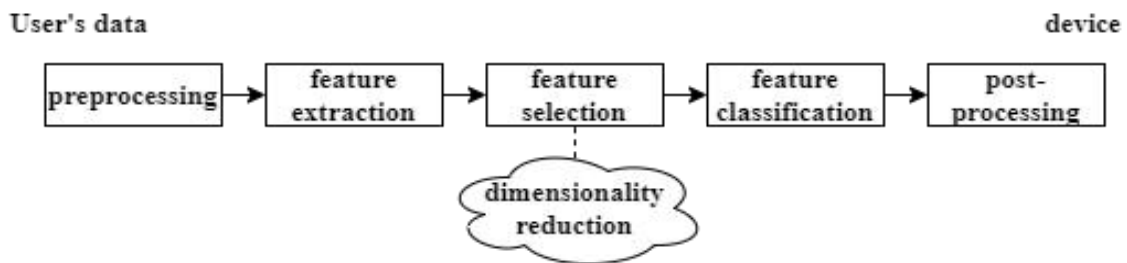


Figure 1. Functional model of a BCI system[12], note that this figure concentrates on the main process: preprocessing, feature extraction, feature selection, feature classification and post-processing.

2.2. Characters of the features

In BCI, the characters of the features are as follows:

- 1.frequency band power feature;
- 2.checkpoint feature.

These two kinds of features are two domains of the Fourier Transform, and besides there are a lot of other ways such as connection feature[19] and covariance matrix[23]. It is an important information when we consider which unsupervised model to use.

2.3. Ways to measure features

In recent 10 years, scientists have invented powerful tools to handle the classification problem. Here we list a few:

1. Adaptive classifier: The efficiency of adaptive classifier is greater than static classifier[14], which can be considered in the unsupervised learning.
2. Due to the incompatibility of DNN, shallow neural network is considered more useful in the future[6].
3. sLDA was discovered to be more efficient and robust than the traditional LDA, in terms of limited training datasets[10].
4. RGC can be seen as a very promising region[21], MDM classifier based on RGC is viewed as the state-of-the-art technique to solve multiple BCI problems, especially sport imagination, P300 and SSVEP classification.
5. Tensor method[15, 16] is still a prosperous research direction.

In the following two parts, we'll explore two different classification methods, i.e. the adaptive classifier and the MDM classifier.

3. Adaptive boost classifier

In summary, the adaptive boost classifier (Adaboost) is trying to combine several expert's opinions altogether to get the best choice [20].

3.1. Boosting definition

Boosting is Adaboost's first trick, which aims to construct and synthesize multiple machine learning methods to complete the task, called "weak learner". Boosting algorithm's working mechanism is using initial weights to train weak learner 1, after an update, we compute the weights again, making sure the training samples that have higher learning loss will be trained more attentively. We thus deploy this strategy to train weak learner 2. After several rounds entitled T , we packed those steps altogether and a "strong learner" is completed. The Process of Boosting is interpreted in Figure 2.

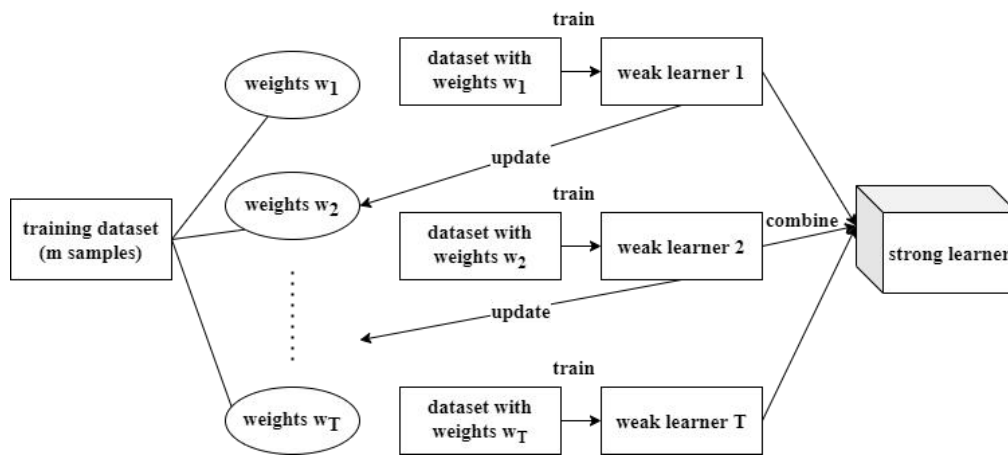


Figure 2. Boosting strategy in the Adaboost. The picture shows that several weak learners can update altogether to generate the final strong learner using the weight updating strategy.

3.2. Adaptive definition

Adaptive express that each weak classifier will strengthen the mistaken samples, and those samples will inevitably be repeatedly trained in the following weak classifiers. This strategy helps the strong learner to be more adaptive in all kinds of samples.

3.3. Algorithms

Based on the information above, we construct the Adaboost algorithms as follows:

1. Initialize the weight distribution of the training data. If there are N examples, each training example is initially assigned the same weight $1/N$.
2. Train the weak classifier. During training, if a point has been correctly classified, its weight will be reduced in the construction of the next training set. On the contrary, if a sample point is not classified accurately, then its weight is improved. The updated set of samples is used to train the next classifier, and the training process continues iteratively.
3. Multiple weak classifiers are combined to form a strong classifier. After the training process of each weak classifier, the weight of the weak classifier with small classification error rate was increased to make it play a greater role in the final classification function. And the weight of the weak classifier with large classification error rate is reduced, so that it plays a small decisive role in the final classification function.

4. MDM classifier

4.1. Reasons for the robustness

It is widely known that a great challenge of BCI is the brain-electric data's vulnerability to manual factors, such as environment, biological pollution and equipment. One way to solve this problem is experimenting in the lab, but it can't help in the long run.

MDM classifier, based on Riemannian geometry, is proven to be more robust to abnormal data due to its geometric mean method.

When an average abnormal data is found, compared to arithmetic mean, bias of geometric mean(to the center of distribution) is smaller. That's because data distribution of the two situations is different(one is χ^2 , the other is N). As for χ^2 distribution, bias is larger than normal distribution with the corresponding parameters[11].

5. Conclusions

From the very first paper to declare neural network can classify latent space into valid information, to the state-of-the-art MDM classifier solving multiple BCI problems, many more methods can be examined in BCI dimension. Here we come up with several possible directions in the future:

Sequence-to-sequence tasks: BCI is expert at observing users' decision, i.e. classifying tasks. According to a famous paper[22], the paper describes a technique breakthrough in Natural Language Processing(NLP), which aims to run a sequence-to-sequence model in parallel. We can conclude that BCI in the near future can also realize such a seq2seq task, able to translate language or analyzing puzzles for users, which is difficult for BCI in the current stage.

Another interesting direction is Internet of Things(IoT). It may be soon possible for BCI systems to contact not only with users or creatures, but also about other furniture and facilities.

In conclusion, BCI research stimulates our hope and expectation of the reasons and structures of brain. And we expect that classification work will also be much more exciting in the near future.

References

- [1] Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering* 4(2), R32 (2007)
- [2] Birbaumer, N.: Breaking the silence: brain-computer interfaces (bci) for communication and motor control. *Psychophysiology* 43(6), 517–532 (2006)
- [3] Anderson, C.W., Stolz, E.A., Shamsunder, S.: Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Transactions on Biomedical Engineering* 45(3), 277–286 (1998)
- [4] Anderson, C.W., Devulapalli, S.V., Stolz, E.A.: Eeg signal classification with different signal representations. In: *Proceedings of 1995 IEEE Workshop on Neural Networks for Signal Processing*. pp. 475–483. IEEE (1995)
- [5] Bashashati, A., Ward, R.K., Birch, G.E.: A new design of the asynchronous brain computer interface using the knowledge of the path of features. In: *Conference Proceedings. 2nd International IEEE EMBS Conference on Neural Engineering*, 2005. pp. 101–104. IEEE (2005)
- [6] Bianchini, M., Scarselli, F.: On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *IEEE transactions on neural networks and learning systems* 25(8), 1553–1565 (2014)
- [7] Anderson, C.W., Stolz, E.A., Shamsunder, S.: Discriminating mental tasks using eeg represented by ar models. In: *Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society*. vol. 2, pp. 875–876. IEEE (1995)
- [8] Birch, G.E., Bozorgzadeh, Z., Mason, S.G.: Initial on-line evaluations of the lf-asd brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary

- motor potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10(4), 219–224 (2002)
- [9] Birch, G.E., Mason, S.G., Borisoff, J.F.: Current trends in brain-computer interface research at the neil squire foundation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11(2), 123–126 (2003)
 - [10] Bulgac, A., Yu, Y.: Superfluid lda (slda): Local density approximation for systems with superfluid correlations. *International Journal of Modern Physics E* 13(01), 147–156 (2004)
 - [11] Congedo, M., Barachant, A., Bhatia, R.: Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces* 4(3), 155–174 (2017)
 - [12] Fatourechi, M., Ward, R.K., Mason, S.G., Huggins, J., Schlögl, A., Birch, G.E.: Comparison of evaluation metrics in classification applications with imbalanced datasets. In: 2008 seventh international conference on machine learning and applications. pp. 777–782. IEEE (2008)
 - [13] Hastie, T., Tibshirani, R., Friedman, J.: Overview of supervised learning. In: *The elements of statistical learning*, pp. 9–41. Springer (2009)
 - [14] Lai, N., Kan, M., Han, C., Song, X., Shan, S.: Learning to learn adaptive classifier–predictor for few-shot learning. *IEEE transactions on neural networks and learning systems* 32(8), 3458–3470 (2020)
 - [15] Li, J., Zhang, L.: Regularized tensor discriminant analysis for single trial eeg classification in bci. *Pattern Recognition Letters* 31(7), 619–628 (2010)
 - [16] Liu, Y., Li, M., Zhang, H., Wang, H., Li, J., Jia, J., Wu, Y., Zhang, L.: A tensor-based scheme for stroke patients’ motor imagery eeg analysis in bcifes rehabilitation training. *Journal of neuroscience methods* 222, 238–249 (2014)
 - [17] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering* 15(3), 031005 (2018)
 - [18] Mansoor, A., Usman, M.W., Jamil, N., Naeem, M.A.: Deep learning algorithm for brain-computer interface. *Scientific Programming* 2020 (2020)
 - [19] Moon, S.E., Chen, C.J., Hsieh, C.J., Wang, J.L., Lee, J.S.: Emotional eeg classification using connectivity features and convolutional neural networks. *Neural Networks* 132, 96–107 (2020)
 - [20] Schapire, R.E.: Explaining adaboost. In: *Empirical inference*, pp. 37–52. Springer (2013)
 - [21] Singhal, S., Bhatia, B., Jayaram, H., Becker, S., Jones, M.F., Cottrill, P.B., Khaw, P.T., Salt, T.E., Limb, G.A.: Human müller glia with stem cell characteristics differentiate into retinal ganglion cell (rgc) precursors in vitro and partially restore rgc function in vivo following transplantation. *Stem cells translational medicine* 1(3), 188–199 (2012)
 - [22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* 30 (2017)
 - [23] Wang, R., Guo, H., Davis, L.S., Dai, Q.: Covariance discriminative learning: A natural and efficient approach to image set classification. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 2496–2503. IEEE (2012)