Machine learning accelerates the technological development of brain science

Mengya Chen¹, *and Longfei Li²

¹School of Computer Engineering and Science, Shanghai University, Shanghai, China
²School of Computer Engineering and Science, Shanghai University, Shanghai, China
*Corresponding author: ashleychen@shu.edu.cn

Abstract. The development of machine learning provides a strong research foundation for the field of brain science. Compared with traditional research methods, the strategy based on machine learning is more suitable for the big data background of the current development of the field, and has a broad application prospect. This paper introduces the application status of machine learning in brain science research from three aspects: neuron reconstruction, neuron classification, neuron synthesis, and summarizes the development direction of machine learning technology in the field.

1. Introduction
The primary goal of neuroscience is to understand the function of the brain and the information processing process of the nervous system. In the past few decades, the progress of dye-labelling technology has provided a solid foundation for neuroscience [1]. The brain-like research has ushered in a new development opportunity. At present, the research focus in this brain science is divided into three categories: obtaining accurate and complete morphological structure data from the three-dimensional neuron images; morphological analysis of the reconstructed neurons to obtain the category information; synthesize virtual neurons with computational method [2]. As the focus of neuron data gradually turns to the whole brain level, the traditional research methods are no longer suitable for these complex and large-scale morphological data. More and more researches begin to use machine learning technology to solve these problems.

This paper introduces the application status of machine learning methods in the field of neuroscience, including some related morphological reconstruction methods, classification methods and generation methods of neurons, as shown in Figure 1.
2. Reconstruction method
Reconstruct morphological structure from three-dimensional image of neuron is one of the basic tasks of computational neurology, which is generally divided into image pre-processing step, cell body segmentation step, curve segmentation step and others. The main difficulty of reconstruction task lies in the huge scale of TB level image data and low signal-to-noise ratio. In addition, the overlapping structures in the whole-brain level data are difficult to be identified by naked eyes, which often leads to low efficiency and poor quality of manual reconstruction methods.

By using machine learning method, the initial reconstruction results of neurons can be classified to obtain the optimal path results. Therefore, ensemble learning and other methods can also be used to reconstruct the process.

Smart tracing [3] is an automated neural reconstruction framework without human intervention. According to the evaluation criteria based on the degree of confidence, the most reliable result is predicted from the existing automated reconstruction results.

In smart tracing method, neuron reconstruction is decomposed into several different segments, a reliability is calculated for each segment by judging whether it has an alternative path. That is, a fragment without an alternative path is more reliable than a fragment with an alternative path. For the fragment $S_{ij}$ between endpoints $i$ and $j$, first calculate the average brightness in the following way:

$$\bar{I}_{ij} = \frac{\int_{i}^{j} I(x)dx}{L_{ij}}$$  \hspace{1cm} (1)

Where $I(x)$ is the intensity value at point $x$, $L_{ij}$ is the path length of the fragment.

The confidence of the fragment is calculated by the average intensity $\bar{I}_{ij}$ of the shortest path $L_{ij}^{*}$ divided by the average intensity $\bar{I}_{ij}$ of the segment. When there is an alternative path, the confidence of the fragment may be close to 1, while if there is no alternative path, the shortest path may be a line with low brightness, which makes the confidence close to 0. According to the confidence value, the initial reconstruction result can be divided into three regions: foreground, uncertainty and background. Smart tracking extracts local sub regions based on MWR coding method to obtain raw feature values, then the final feature representation is obtained by feature selection method mRMR [4] (minimum redundancy and maximum correlation) and used for training a support vector machine model. Finally, the SVM model is used to determine the foreground region and background region and get a final reconstruction result. With the help of machine learning strategy, smart tracking greatly improves the effect of traditional reconstruction methods.
3. Classification method

The neural classification task is mainly based on the neural features extracted by L-measure [5] tool. Due to the different reconstruction standards of institutes, the feature set that can accurately describe a neuron type has not been unified in the academic community, thus the definition of neuron type information is still unclear.

In the task of neuron classification based on machine learning, feature selection strategy is often used to obtain a set of optimal feature subsets to describe the structure of neurons. Then, classifiers such as support vector machine or decision tree can be trained to obtain the feature space of the neuron, thus each neuron sample is classified into different subcategories according to this space, as shown in Figure 2.

Evelyn et al. [6] proposed a classification method based on neuron tree hierarchy, which decomposes neurons into different subparts according to hierarchy. Then they use rfecv (recurrent feature extraction and cross validated selection) method to get a subset of l-measure features. The result feature subset is used to represent a neuron sample to train a set of LIBSVM based e-support vector Classifier (SVC) models for 10 type of neurons. Finally, these models are applied to the classification task to study the relationship between neuron type information and neuron decompositions by depth, height or slices.

Xavier et al. [7] tested the classification effect of various supervised or unsupervised machine learning classification methods for 430 rat neuron samples belonging to 22 subcategories. The feature values of neuron samples are normalized to 0 and 1 intervals to calculate the classification accuracy. With 18 supervised algorithms and 8 unsupervised algorithms. With machine learning methods, the best classification accuracy can reach more 95%. Among the supervised learning methods, LDA algorithm achieves the best classification accuracy; among the unsupervised learning methods, Ward method and affinity propagation achieve better results.

4. Synthesis method

Virtual neuron synthesis technology solves the problem of tedious and time-consuming reconstruction process of real neurons. Due to the limitation of the experimental environment, part of the neuronal tissue is often lost in the process of sectioning or labelling, which easily leads to incomplete results of neuronal reconstruction. With the help of neural synthesis, more complete data of neuron morphology can be obtained. Therefore, the related morphological studies such as brain behaviour can be carried out. The existing virtual neuron synthesis methods are mainly divided into growth model and reconstruction model. Figure 3 shows a schematic of the neuronal synthesis process.
Figure 3. Figure of neuron synthesis.

In growth models, the growth behaviour of growth cones was simulated in different growth stages. The probability of neurite extension, bifurcation or termination is often calculated by various factors in the process of growth and development, such as molecular gradients [8] and neural tension [9]. Then growth models can simulate neuron growth by selecting a growth cone action with these factors at every time step.

In Lin's work [10], the development of neurons is described by local morphological variables, which is expressed as the replication and evolution of gene fragments by gene regulatory network (GRN) model. The input layer has two nodes to input the nutrition gradient information of neurons, and the output layer has 16 nodes to output the basic morphological variables in the development process. Through the nonnormalized sorting genetic algorithm II (NSGA-II), the network structure is optimized for the generation of virtual cat motor neurons.

On the other hand, reconstruction model measures the distribution of morphological variables of a group of real neurons. Then, the relevant calculation method is used to sample the parameters to obtain the virtual neuron structure. However, this kind of sampling process often assumes that the features are independent of each other, which is not true in real neuron data. In addition, the feature subset selection in different methods is often based on a priori perspective thus it lacks wide applicability in different neuron data sets.

In order to solve this problem, Pedro et al. [11] used Bayesian network to simulate the dendritic structure of neurons. By measuring the feature information of real neurons, the joint probability distribution is estimated to establish a Bayesian network, then the network is used to output a group of virtual dendritic structures. Among this method, Bayesian information criterion (BIC) based scoring equation is used in score plus search step to learn a Bayesian network. They tested the network to synthesis virtual pyramidal neurons from mouse neocortex.

5. Conclusion
This paper introduces the methods of neural reconstruction, neural classification and neural synthesis based on machine learning strategy in the field of brain science, and expounds the application status of related machine learning methods in the field of brain like research. We have reason to believe that with the gradual increase of the data scale in brain science research, powerful computing tools like machine learning and deep learning will finally replace the traditional research methods and provide new solutions to the related problems in the field.
References

[1] Parekh R and Ascoli G A 2013 Neuronal Morphology Goes Digital: A Research Hub for Cellular and System Neuroscience Neuron 77 1017–38

[2] Shillcock J C, Hawrylycz M, Hill S and Peng H 2016 Reconstructing the brain: from image stacks to neuron synthesis Brain Inform. 3 205–9

[3] Chen H, Xiao H, Liu T and Peng H 2015 SmartTracing: self-learning-based Neuron reconstruction Brain Inform. 2 135–44

[4] Ding C and Peng H 2005 MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA J. Bioinform. Comput. Biol. 03 185–205

[5] Scorcioni R, Polavaram S and Ascoli G A 2008 L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies Nat. Protoc. 3 866–76

[6] Cervantes E P, Comin C H, Junior R M C and Costa L da F 2019 Morphological Neuron Classification Based on Dendritic Tree Hierarchy Neuroinformatics 17 147–61

[7] Vasques X, Vanel L, Villette G and Cif L 2016 Morphological Neuron Classification Using Machine Learning Front. Neuroanat. 10

[8] Hentschel H G E and Ooyen A van 1999 Models of axon guidance and bundling during development Proc. R. Soc. Lond. B Biol. Sci. 266 2231–8

[9] Li G-H and Qin C-D 1996 A model for neurite growth and neuronal morphogenesis Math. Biosci. 132 97–110

[10] Lin X, Li Z, Ma H and Wang X 2018 An evolutionary developmental approach for generation of 3D neuronal morphologies using gene regulatory networks Neurocomputing 273 346–56

[11] López-Cruz P L, Bielza C, Larrañaga P, Benavides-Piccione R and DeFelipe J 2011 Models and Simulation of 3D Neuronal Dendritic Trees Using Bayesian Networks Neuroinformatics 9 347–69