Morphology cluster and prediction of growth of human brain pyramidal neurons**★**

Chao Yu[1], Zengxin Han[1], Wencong Zeng[1], Shenquan Liu[2]

1*School of Computer, China University of Geosciences, Wuhan 430074, Hubei Province, China*
2*School of Science, South China University of Technology, Guangzhou 510640, Guangdong Province, China*

Abstract

Predicting neuron growth is valuable to understand the morphology of neurons, thus it is helpful in the research of neuron classification. This study sought to propose a new method of predicting the growth of human neurons using 1 907 sets of data in human brain pyramidal neurons obtained from the website of NeuroMorpho.Org. First, we analyzed neurons in a morphology field and used an expectation-maximization algorithm to specify the neurons into six clusters. Second, naive Bayes classifier was used to verify the accuracy of the expectation-maximization algorithm. Experiment results proved that the cluster groups here were efficient and feasible. Finally, a new method to rank the six expectation-maximization algorithm clustered classes was used in predicting the growth of human pyramidal neurons.
Key Words: neurons; morphological cluster; expectation-maximization; naive Bayes; 10-fold cross validation; neural regeneration

Chao Yu★, Master, School of Computer, China University of Geosciences, Wuhan 430074, Hubei Province, China

INTRODUCTION

The Human Brain Project mainly researches neuroscience and informatics, with an aim to build a database of human brain neurons and to accelerate the cognitive, multidisciplinary database for analysis[1-2]. Geometrical characteristics and physical characteristics of electricity are the two most important aspects in neuronal structure and function. Nowadays, the research of neuronal morphology attracts much attention[3-11]. Ristanovic *et al*[12] studied morphology analysis and classification of large neurons in human dentate nucleus. Alavi *et al* [13] used three popular classification methods to classify rodent brain neurons, and helped to research brain function and neurological disorders. Narihisa *et al*[14] applied a Bayes algorithm on neuron data to perform clustering, and gained a good clustering result. However, he did not predict the growing tendency of human neurons. In medical research fields, predicting growing tendency of a certain type of human brain neuron can be used to check if neurons are lesions or benign in their developmental stages. It is difficult to record the growing tendency of a neuron, but we can cluster human pyramidal neurons into several classes first, and then use the class weight to rank the different class, hoping to predict growing tendency by the order.

RESULTS

Human pyramidal neuron clusters

A total of 1 907 groups of human pyramidal neuron data from NeuroMorpho.Org were used to mimic the growing tendency. The expectation-maximization (EM) algorithm was used to divide human pyramidal neurons data into several clusters. We set maximum iterations as 100, minimum standard deviation as 1×10^{-6}. We gained six classes of the datasets, presenting the different growing stages of neurons (Table 1).

Table 1 The distribution of human pyramidal neurons

Human pyramidal neurons	Sample number	Percentage (%)
Cluster 0	388	20
Cluster 1	394	21
Cluster 2	219	11
Cluster 3	163	9
Cluster 4	446	23
Cluster 5	297	16

The data of human pyramid neurons came from website http://neuromorpho.org/neuroMorpho/ index.jsp, and the result was calculated by expectation-maximization algorithm using Weka Release 3-4-12 software (Machine learning group, Waikato, New Zealand), which is an open source software for free use.

The number of instances in a cluster is close to another, so that a cluster can be treated as a dependent growing stage of neurons.

We have to verify the result calculated by the cluster algorithm, making sure of its correction. Then we rank the disordered clusters in a new way to get the growing tendency of human pyramidal neurons.

Verification of neuron clusters and growing tendency prediction of human pyramidal neurons

We used naive Bayes to identify our clusters and the results are shown in Table 2. Validation of cluster accuracy by naive Bayes is shown in Figure 1.

We ranked the 6 classes by weight, presenting the growing tendency of pyramidal neurons in continuous growth stages. The clusters were ordered by weight from minimum to maximum. We got the order of the clusters as cluster 4, cluster 1, cluster 5, cluster 0, cluster 2 and cluster 3 (Table 3).

Table 2 Detailed accuracy of human pyramidal neurons by class

TP rate	FP rate	Precision	Recall	F-measure	ROC area	Cluster
0.946	0.020	0.924	0.946	0.935	0.996	0
0.937	0.019	0.929	0.937	0.933	0.990	1
0.995	0.003	0.978	0.995	0.986	1.000	2
0.969	0.003	0.963	0.969	0.966	0.994	3
0.960	0.005	0.982	0.960	0.971	0.991	4
0.923	0.010	0.945	0.923	0.934	0.996	5

Six evaluation methods are taken to illustrate the accuracy; they are true positive rate (TP rate), false positive rate (FP rate), precision, recall, harmonic mean of precision and recall (F-measure), and receiver operating characteristic area (ROC area)[15]. All these results were calculated automatically by software Weka using naïve Bayes method. It is better to approach zero for false positive rate, while for the other five evaluation methods, the values are better if they are close to one.
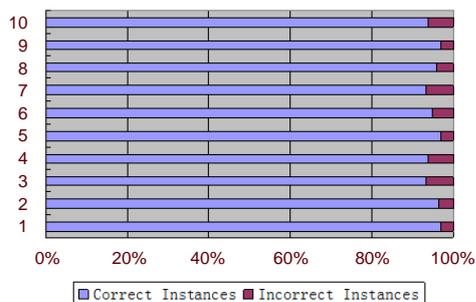
Figure 1   Validation of accuracy in cluster of human pyramidal neurons by naive Bayes for ten times. Abscissa stands for the accuracy each time, and ordinate means the number of times. This figure was outputted by Microsoft Excel.

We concluded from Table 3 that if a pyramidal neuron stays in cluster 1, then it will grow to cluster 5 in the future. The accuracy of the cluster is 95.12%, which means most of the neurons are clustered correctly, but it is really difficult to distinguish some of the human pyramidal neurons. For example, if a neuron ought to be in stage 4, it may be divided into stage 3 or stage 5 because geometry characteristics of neurons in their neighboring stage are so similar. Figure 2 describes the six different stages of human pyramidal neurons.

Table 3   Ranking human pyramidal neurons by weight

Cluster	Weight	Stage
0	0.361 563	4
1	0.296 557	2
2	0.420 220	5
3	0.775 171	6
4	0.219 133	1
5	0.323 701	3

The six clusters were divided by expectation-maximization algorithm and the weight values were calculated by formula (5) according to the attributes of each cluster. By ranking the weights, neuron clusters represented six continuous stages.

DISCUSSION

It has been proposed[16] that a novel method is required to improve the line-pixel detection technique to detect the curvilinear structure more accurately. In this paper, we used the data from the website NeuroMorpho.Org, which is the largest collection of publicly accessible three-dimensional neuronal metadata which. We extracted the main attributes and analyzed the morphology of human pyramidal neurons, and figured out a way to predict the growing tendency of neurons. After describing the morphology of neurons, neuron clustering was also studied. The study regarding the diversity of ganglion cells in the mouse retina analyzed the neurons quantitatively[17] and they used k-means methods, automatically clustered a series of cells and judged them by silhouette analysis. Our paper used another cluster method named the EM method to research human pyramidal neurons. Generally speaking, EM method is more accurate and more widely used than k-means methods. However, the k-means method and the EM method are comparable when the input data are exactly the same. Nevertheless, the morphology of clustering neurons has been found to vary with different types of neurons. Molnar *et al*[18] described morphology and clustering of pyramidal neurons. Although they analyzed the same kind of cells and discussed the significance of their achievement, they failed to give further research on prediction of neuron growing tendency of the same type.

Technically speaking, the data gained by staining the neurons results in the death of the neurons due to the staining process. Thus, no neurons will be counted twice, while a lot of neuron data are used to predict the different periods of growing tendency of pyramidal neurons by clustering. Moreover, original data are in high dimension, so it is necessary to use classic classification algorithms, such as naïve Bayes algorithm to verify the accuracy of cluster identification. Once the clustering algorithm was verified, we could predict neuron growing tendency by dividing them into several smaller clusters. Different neuron clusters have different geometry features that can be treated as disordered periods in neuronal growth.
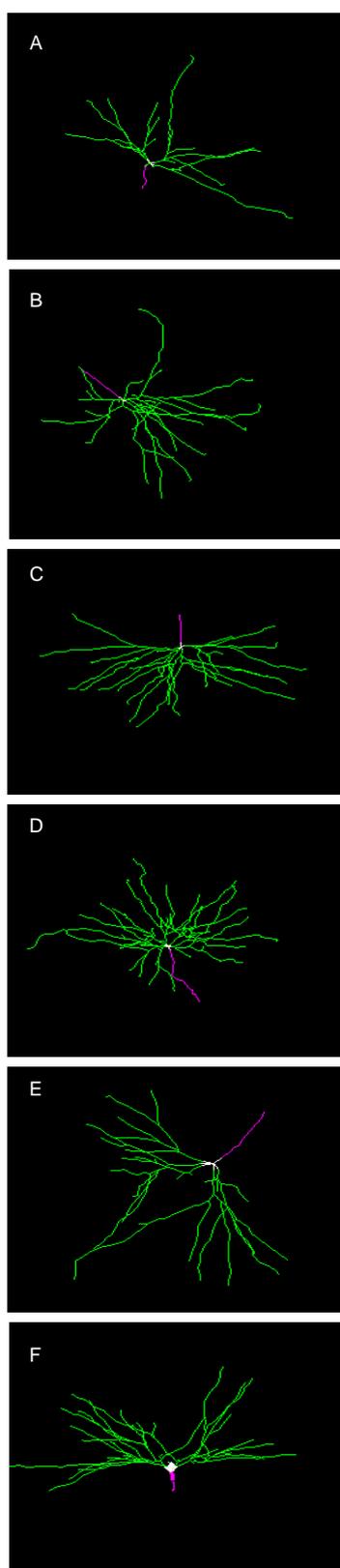
Figure 2 Human pyramidal neurons ordered by neuron weights. The picture in stage 1 was extracted from cluster 4 randomly. Similarly, the picture in stage 2 was a random neuron in cluster 1. All the pictures came from the website http://neuromorpho.org/neuroMorpho/index.jsp and could be downloaded freely. (A–F) Stages 1–6.

In this paper, EM algorithm was chosen to divide the same type neurons into 6 classes. A 10-fold cross validation of naïve Bayes, the most frequently used method, was used to verify the accuracy of each cluster. It divides the dataset into 10 equal pieces, and uses 9 pieces of the dataset as a standard distribution sample, whereas the remaining piece is used to verify accuracy. The validation was repeated ten times with each piece used individually as the test set: the final result was the average accuracy registered by each of these ten separate validations. Meanwhile, we proposed a new method presented in the last part of this paper to predict the growing tendency of human pyramidal neurons by calculating the classes' weight, and ranked the classes by their weight. It is important to separate human pyramidal neurons into several clusters, presenting different growing stages. Our novel method is easy to understand and can be used effectively.

We can infer from the ranking that if a neuron belongs to the one class, it will mature to the next class; if it belongs to the last class, then it will most likely die by its continuous growth. Although we predicted the growing tendency of human pyramidal neurons by six classes, the growth of the neuron is continuous and can be divided into more than six classes as well. In future research, we hope to figure out a way to predict neuron growth tendency more accurately rather than distributing them into six classes. Also, as for a different dataset, the accompanying parameters are not always the same. Self-adaptive parameters could be considered in the EM algorithm to increase the cluster accuracy. This method is widely used because it allows choosing of the number of clusters that can greatly reduce human subjective factors.

MATERIALS AND METHODS

Design
A neuronal cluster analysis.
Time and setting
All experiments were performed at China University of Geosciences (Wuhan, China) from September to December 2010.
Materials
There are a lot of neuron data presented on website http://neuromorpho.org/neuroMorpho/index.jsp and they are listed by several categories. It is easy to find out all of the human pyramidal neurons by browsing all files by animal species. A total of 1 907 groups of pyramidal neurons of human beings were used in this study to make the result more objective.
Methods
Establishment of morphological cluster models
We identified 78 attributes for each neuron using the Neuron software downloaded from website http://fourcoffees.com/project/neuron/ to extract the geometry feature of neurons. For more information about the definitions of attributes, please visit the web site of

Neuron software. We observed that four of the attributes were zero for all the neurons, namely Minimum branch order, Minimum path distance, Minimum Euclidean distance, and Minimum compartments length. As a result, we deleted these four attributes and used the remaining 74 as our base criteria.

EM algorithm[19] divided these 1 907 data into several clusters and the number of clusters was automatically chosen. The EM algorithm has the following two steps[20]:
(1) Speculating initial parameters to calculate cluster probabilities.
(2) Calculation of the distribution parameters to re-estimate the parameters, then repeat.

Variable w_i denotes the probability of instance x_i belongs to cluster A, then the mean μ_A and standard deviation σ_A for A are calculated by (1) and (2).

$$\mu_A = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_i x_i + \cdots w_n x_n}{w_1 + w_2 + \cdots + w_i + \cdots + w_n} \tag{1}$$

$$\sigma_A^2 = \frac{w_1(x_1 - \mu)^2 + w_2(x_2 - \mu)^2 + \cdots + w_n(x_n - \mu)^2}{w_1 + w_2 + \cdots + w_n} \tag{2}$$

Here x_i can represent any instances, but not only the instance belongs to cluster A. EM algorithm stops when it converges toward a fixed point, that is, the increase of overall log-likelihood becomes negligible.

Neuronal validations by naive Bayes classifiers

In this paper, the neuronal dataset had 74 attributes and contained a large number of samples, so we chose naive Bayes classifiers to validate cluster accuracy. Naive Bayes classifiers can be very efficiently worked into many complex situations. Naive Bayes classifiers assumed independent variables and did not work well with relevant or dependent variables. The probability of an instance belongs to a class C is calculated by (3).

$$p(C \mid F_1, F_2, \cdots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^{n} p(F_i \mid C) \tag{3}$$

In formula (3), Z is a scaling factor dependent only on the value of features, where $F_1, \cdots, F_n$. $p(C)$ is the possibility of C appears in the training set, and $p(F_i \mid C)$ is the independent probability distributions[21]. All the data of neurons are non-nominal and continuous, so it is possible to apply naive Bayes on these data. Naive Bayes theory assumes that the values associated with each class are distributed according to Gaussian distribution[22] and the probability of some values given a class can be computed by (4).

$$P(x = v \mid c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v - \mu_c)^2}{2\sigma_c^2}} \tag{4}$$

Where μ_c and σ_c^2 are normal distribution parameters and stands for average and standard deviation values in x associated with class c[23]. We can compute the probability of some value given a class by plugging v into the equation.

Pyramidal neurons growing tendency prediction

Since the pyramidal neurons of human brain were divided into 6 types and the correction of EM algorithm has been proved in this paper, the next step was to figure out the orders. Since attribute values are widely ranged, from 10^{-4} to 10^{6}, SPSS Release 13 (SPSS Inc., Chicago, IL, USA) is used to normalize these values and reduce them to smaller ranges and calculate the weight to 6 different clusters, by which to order the growing tendency of neurons. The detail is as formula (5) shown below.

$$weight = \sum_{i=1}^{74} (|\mu_i| \sigma_i / \sum_{i=1}^{74} \sigma_i) \tag{5}$$

The *weight* value denotes the weight of a specific cluster, and μ_i and σ_i are the average arithmetic mean and standard deviation of ith attribute in the certain cluster.

When neurons are growing, their dendrites and axon grow as well. It can infer that the neuron's volume and area increases when it is growing, while minimum breach order decreases as well. According to formula (5), if neurons are growing bigger, their absolute μ_i are larger. Meanwhile, neurons contain more information when they are growing bigger, so there is more dispersion, which leads to a higher σ^2 value. We use $|\mu_i|$ multiply $\sigma_i / \sum_{i=1}^{74} \sigma_i$ to get a relatively higher value when the neuron is bigger, and assume that neurons increase exponentially when they are growing up, and then die without getting smaller.

REFERENCES

[1] Koslow SH, Huerta MF. Neuroinformatics: an Overview of the Human Brain Project. Mahwah: Lawrence Erlbaum Associates. 1997.

[2] Shepherd GM, Mirsky JS, Healy MD, et al. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. Trends Neurosci. 1998;21:460-468.

[3] Russo D, Bombardi C, Castellani G, et al. Characterization of spinal ganglion neurons in horse (equus caballus). A morphometric, neurochemical and tracing study. Neuroscience. 2011;176:53-71.

[4]   Mendoza KC, McLane VD, Kim S, et al. In vitro application of gold nanoprobes in live neurons for phenotypical classification, connectivity assessment, and electrophysiological recording. Brain Res. 2010;1325:19-27.

[5]   Pohl K, Grimson WE, Bouix S, et al. Anatomical guided segmentation with non-stationary tissue class distributions in an expectation-maximization framework. In ISBI. Arlington, VA, USA: IEEE. 2004.

[6]   Buchser WJ, Slepak TI, Arenas OG, et al. Kinase/phosphatase overexpression reveals pathways regulating hippocampal neuron morphology. Mol Syst Biol. 2010;6:1-16.

[7]   Baharnoori M, Brake WG, Srivastava LK. Prenatal immune challenge induces developmental changes in the morphology fo pyramidal neurons of the prefrontal cortex and hippocampus in rats. Schizophr Res. 2009;107:99-109.

[8]   Gelfo F, Bartolo PD, Giovine A, et al. Layer and regional effects of environmental enrichment on the pyramidal neuron morphology of the rat. Neurobiol Learn Mem. 2009;91(4):353-365.

[9]   Jaworshi J, Kapitein LC, Gouveia SM. Dynamic microtubules regulate dendritic spine morphology and synaptic plasticity. Neuron. 2009;61:85-100.

[10] Thuma JB, White WE, Hobbs KH, et al. Pyloric neuron morphology in the stomatogastric ganglion of the lobster, Panulirus interruptus. Brain Behav Evol. 2009;73(1):26-42.

[11] Jiang R, Liu Q, Liu Q, et al. A proposal for the morphological classification and nomenclature of neurons. Neural Regen Res. 2011;6:1925-1930.

[12] Ristanovic D, Milosevic NT, Stefanovic BD, et al. Morphology and classification of large neurons in the adult human dentate nucleus: A qualitative and quantitative analysis of 2D images. Neurosci Res. 2010;67:1-7.

[13] Alavi A, Cavanagh B, Tuxworth G, et al. Automated classification of dopaminergic neurons in the rodent brain. Proceedings of International Joint Conference on Neural Netw. Atlanta, Georgia, USA. 2009.

[14] Narihisa M, Shotaro A, Yasuko SM, et al. Visualization of multi-neuron activity by simultaneous optimization of clustering and dimension reduction. Neural Netw. 2010;23:743-751.

[15] Kumar K, Andrews M, Jayashankar V, et al. Improvement in diagnosis of breast tumour using ultrasound elastography and echography: A phantom based analysis. Biomed Imaging Interv J. 2009;5(4):e30.

[16] Losavio BE, Liang Y, Santamaria-Pang A, et al. Live neuron morphology automatically reconstructed from multiphoton and confocal imaging data. J Neurophysiol. 2008;100:2422-2429.

[17] Kong J, Fish DR, Rockhill RL, et al. Diversity of ganglion cells in the mouse retina: Unsupervised morphological classification and its limits. J Comp Neurol. 2005;489:293-310.

[18] Molnar Z, Cheung FP. Towards the classification of subpopulations of layer V pyramidal projection neurons. Neurosci Res. 2006;55:105-115.

[19] Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. IEEE T Med Imaging. 2001;20:45-57.

[20] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Diego, CA: Academic Press. 2000.

[21] John GH, Langley P. Estimating Continuous Distributions in Bayesian Classifiers. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. 1995.

[22] McCallum A, Nigam K. A comparison of event models for naive bayes text classification. AAAI Technical Report WS-98-05. 1998.

[23] Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques. San Fransisco, CA: Morgan Kaufmann Publisher. 2011.

(Edited by Li AP, Wang X/Yang Y/Wang L)