# Semi-supervised Projected Clustering for Classifying GABAergic Interneurons

Luis Guerra[1], Ruth Benavides-Piccione[2], Concha Bielza[1], Víctor Robles[3], Javier DeFelipe[2], and Pedro Larrañaga[1]

[1] Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid (UPM), Boadilla del Monte, Madrid, Spain
l.guerra@upm.es
[2] Laboratorio Cajal de Circuitos, Centro de Tecnología Biomédica, UPM and Instituto Cajal, CSIC
[3] Departamento de Arquitectura y Tecnología de Sistemas Informáticos, UPM, Boadilla del Monte, Madrid, Spain

**Abstract.** A systematic classification of neuron types is a critical topic of debate in neuroscience. In this study, we propose a semi-supervised projected clustering algorithm based on finite mixture models and the expectation-maximization (EM) algorithm, that is useful for classifying neuron types. Specifically, we analyzed cortical GABAergic interneurons from different animals and cortical layers. The new algorithm, called SeSProC, is a probabilistic approach for classifying known classes and for discovering possible new groups of interneurons. Basic morphological features containing information about axonal and dendritic arborization sizes and orientations are used to characterize the interneurons. SeSProC also identifies the relevance of each feature and group separately. This article aims to present the methodological approach, reporting results for known classes and possible new groups of interneurons.

**Keywords:** Clustering, semi-supervised, finite mixture model, EM, projected, cortical interneurons.

## 1 Introduction

Neuroscience is perhaps the field of science with most interdisciplinary research approaches due to the complexity of the nervous system. In recent years, mathematical and statistical methods, and machine learning techniques have been proved to be excellent tools for analyzing different aspects of the anatomical and functional organization of the brain. A problem, which remains unsolved since the early days of the study of brain structure, is the classification of neurons. Although efforts [1] have been made in order to produce an accepted classification and terminology, experts still have differences of opinion. Solving this problem is a key milestone, not only for organizing the vast amount of data that neuroscience produces, which is fundamental for a better understanding of the structure and functions of cortical circuits, but also for helping researchers communicate with each other.

N. Peek, R. Marín Morales, and M. Peleg (Eds.): AIME 2013, LNAI 7885, pp. 156–165, 2013.

Researchers have already attempted to quantitatively classify cortical neurons using machine learning techniques. Although some adopted a supervised approach to perform this task [2, 3], most reported research was based on clustering to discern between types of neuronal data. For example, hierarchical clustering has been widely used to discover groups[1] of pyramidal cells m[4, 5] and interneurons [6, 7, 8], the main two accepted morphological types of neurons [9]. Recently, a novel, web-based interactive experiment enabled 48 worldwide experts in neuroscience to classify interneurons by visual inspection according to pre-determined criteria [10]. Thanks to this new approach, researchers were able to investigate the suitability of several anatomical terms and neuron names and concluded that supervised classification models could automatically categorize some types of interneurons in conformity with expert assignments. However, although there has for the first time been some advance in neuron naming, characterization, and classification based on community consensus, the global problem remains unsolved since experts did not reach agreement on the classification of most terms, as discussed in [10].

Thus, in this study, we propose a novel semi-supervised projected clustering method that relies on model-based clustering [11] to classify interneurons. Our classification takes basic morphological features and retrieves the known information, in the shape of data labels, from expert opinions given in [10]. Our method is able to discover possible new groups of interneurons on which the scientific community largely agrees and also identifies the relevance of each feature and group separately. Therefore, our method differs from previous approaches to this task as regards both the classification approach and how feature relevance is identified. For further details about semi-supervised learning, see [12, 13]. Different approaches related to the localized manner for identifying interesting subsets of features are reviewed in [14, 15]. More specifically, model-based clustering with embedded search of feature-relevance factors was introduced in [16] and applied to magnetic resonance spectra within a medical context in [17]. Here we present some significant results about classes of interneurons on which agreement was high in [10], and the discovery of possible new groups.

## 2 Materials and Methods

### 2.1 Data

We selected 241 three-dimensional (3D) reconstructions of interneurons from several areas and layers of the cerebral cortex of different experimental animals (mouse, rat, and monkey) and humans, from [10]. All these reconstructions were extracted from NeuroMorpho.Org [18]. From this database, we selected labeled data depending on the number of equal votes (threshold) assigned by experts in [10]. Specifically, we selected three thresholds -18, 22, and 26- used to build

[1] Throughout the text, group is used for clustering approaches, whereas class refers to a label in a supervised approach.

three databases: *th18*, *th22*, and *th26*. A higher threshold is assumed to mean that confidence in the labeled cells is greater.

The labeled neurons belong to four different classes: `Common basket` (`CB`), `Horse-tail` (`HT`), `Large basket` (`LB`), and `Martinotti` (`MT`). Agreement on `Chandelier` cells was also high, although not enough 3D reconstructed cells were available for inclusion in the analysis. Thus, there are 118 labeled cells in *th18*, distributed as 49 `CB`, 9 `HT`, 27 `LB`, and 33 `MT`; 83 labeled cells in *th22*, distributed as 24 `CB`, 5 `HT`, 29 `LB`, and 25 `MT`; and finally, 47 labeled cells in *th26*, distributed as 9 `CB`, 4 `HT`, 12 `LB`, and 22 `MT`.

We described each neuron using nine basic morphological features related to axonal and dendritic arborizations. These features were measured using Neurolucida Explorer. The specific features are: $X_1$ = axonal arbors (Aa) at $(0, \pi]$ (over the soma), $X_2$ = Aa at $(\pi, 2\pi]$ (under the soma), $X_3$ = dendritic arbors (Da) at $(0, \pi]$, $X_4$ = Da at $(\pi, 2\pi]$, $X_5$ = Aa $< 300$µm from the soma, $X_6$ = Aa $[300\text{µm}, 600\text{µm}]$ from the soma, $X_7$ = Aa $> 300$µm from the soma, $X_8$ = Da $\leq 180$µm from the soma, and $X_9$ = Da $> 180$µm from the soma. The aim was to simulate expert interpretation at an early stage of a visual examination, i.e. the orientation and the size of each neuron.

### 2.2 Method

We have created a method called semi-supervised projected model-based clustering (SeSProC) [19]. SeSProC is based on Gaussian finite mixture models; however, its input data are both labeled and unlabeled instances. It is able to classify the unlabeled instances into either known or newly discovered groups. Besides, each feature is weighted to indicate its relevance for each group.

Let the observable data $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a set of instances, with $\mathbf{x}_i \in \Re^F, \forall i \in \{1, \ldots, N\}$. In a typical clustering problem, data are assumed to be generated from a probabilistic model given by a finite mixture of distributions with $K$ components, and the clustering solution is gathered in the mixture using a latent variable $\mathcal{Z}$. The basic density function for an instance $\mathbf{x}_i$ is

$$p(\mathbf{x}_i \mid \Theta) = \sum_{m=1}^{K} \pi_m p(\mathbf{x}_i \mid \boldsymbol{\theta}_m),$$

where $\pi_m$ is known as the mixing proportion and $\boldsymbol{\theta}_m$ is the parameter set of each component. The full parameter set of the mixture is $\Theta = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \pi_1, \ldots, \pi_K\}$. This set would be easy to find using the maximum likelihood method, if the complete-data, i.e. $\mathcal{X}$ and $\mathcal{Z}$, were known. However, $\mathcal{Z}$ is unknown and must be estimated together with the parameter set. We use the expectation-maximization (EM) [20] algorithm to calculate the expectation of the log-likelihood function with respect to the posterior distribution of the latent variable.

As SeSProC is a projected algorithm, the density function changes because the relevance of each feature for each component is also estimated to find the interesting subspaces. This information is gathered in a new latent variable $\mathcal{V}$. Defining $\rho_{mj} = p(v_{mj} = 1)$, i.e. the probability that feature $j$ is relevant to

component $m$, and assuming that features are conditionally independent given the component label, the new density function is

$$p(\mathbf{x}_i \mid \Theta) = \sum_{m=1}^{K} \pi_m \prod_{j=1}^{F} \Big( \rho_{mj} p(x_{ij} \mid \theta_{mj}) + (1 - \rho_{mj}) p(x_{ij} \mid \lambda_{mj}) \Big),$$

where $\theta_{mj}$ and $\lambda_{mj}$ indicate the parameters for the density function if feature $j$ is relevant and irrelevant, respectively, to component $m$. As before, if $\mathcal{Z}$ and $\mathcal{V}$ were known, the new complete-data log-likelihood function, with $z_{im}$ indicating instance $i$'s membership of component $m$, and $v_{mj}$ indicating feature $j$'s relevance to component $m$, would be

$$\begin{aligned} \log L(\Theta \mid \mathcal{X}, \mathcal{Z}, \mathcal{V}) = \sum_{i=1}^{N} \sum_{m=1}^{K} \Big( & z_{im} \log \pi_m \\ & + \sum_{j=1}^{F} \big( z_{im} \left[ v_{mj} (\log \rho_{mj} + \log p(x_{ij} \mid \theta_{mj})) \right. \\ & + \left. (1 - v_{mj}) (\log(1 - \rho_{mj}) + \log p(x_{ij} \mid \lambda_{mj})) \right] \big) \Big). \end{aligned}$$

As SeSProC is a semi-supervised clustering algorithm and its input contains some labeled data, $\mathcal{Z}$ is partially known. However, the unknown part of $\mathcal{Z}$ and $\mathcal{V}$ is estimated at iteration $t$, after the parameters from the previous iteration $t-1$ have been fixed, by calculating the expectation of the complete-data log likelihood function using the EM algorithm, as

$$\begin{aligned} & \mathbb{E}_{\mathcal{Z},\mathcal{V}|\mathcal{X},\Theta^{t-1}} [\log L(\Theta^{t-1} \mid \mathcal{X}, \mathcal{Z}, \mathcal{V})] \\ & = \sum_{i=1}^{N} \sum_{m=1}^{K} \gamma(z_{im}) \log \pi_m \\ & + \sum_{i=1}^{N} \sum_{m=1}^{K} \sum_{j=1}^{F} \gamma(u_{imj}) (\log \rho_{mj} + \log p(x_{ij} \mid \theta_{mj})) \\ & + \sum_{i=1}^{N} \sum_{m=1}^{K} \sum_{j=1}^{F} \gamma(w_{imj}) (\log(1 - \rho_{mj}) + \log p(x_{ij} \mid \lambda_{mj})), \end{aligned}$$

where $\gamma()$ is the expectation of each specific variable, with $\gamma(u_{imj}) = \gamma(z_{im})$ $\gamma(v_{mj})$ and $\gamma(w_{imj}) = \gamma(z_{im})(1 - \gamma(v_{mj}))$.

**Number of Clusters Estimation.** SeSProc also estimates the final number of clusters using a greedy forward search. The Schwartz criterion [21], also known as Bayesian information criterion (BIC), is used to compare models with different numbers of components. In the first step of the search ($s = 0$), a model

with $C$ components ($\mathcal{M}^0$) is built, $C$ being the known labels. Then $\mathcal{M}^1$ is built with $C+1$ components at $s=1$ and compared with $\mathcal{M}^0$ using the BIC. The process continues, adding a component at each new step, until the convergence criterion is reached, i.e. $\mathcal{M}^s$ is better than $\mathcal{M}^{s+1}$, returning $\mathcal{M}^s$. Note that labeled instances can only belong to the $C$ known components, whereas unlabeled instances can be members of any component $C+s$ at step $s$. A key aspect of this process is related to initialization. Labeled instances are used to initialize the known components, but the added components also have to be initialized. We assume that the instances that fit the components at step $s$ of the process worst are candidates for membership of the new component added at $s+1$, and are then used to initialize this component.

### 2.3 Empirical Setup

The process for obtaining input data with labeled and unlabeled instances for each class and data set was as follows: all labels of one of the known four classes were hidden to SeSProC, whereas the labels of the other three classes were unchanged. This process was designed to discriminate between unlabeled and labeled instances, since they belong to different classes of interneurons. Besides, more than one group could be found for the unlabeled instances, leading to the discovery of new groups that were unknown to the algorithm input[2].

Results were then evaluated in two respects. First, they were analyzed in terms of correctly and misclassified instances (see Section 3.1). We considered that a cell was misclassified (*mc*) if that cell was grouped into one of the known classes according to the labeled data. On the contrary, an instance was correctly classified (*cc*) if it was grouped into a completely new group, regardless of the number of new groups that were identified. We then defined accuracy as $\frac{cc}{mc+cc}$, ranging from 0 to 1. Accuracy was 1 when all unlabeled instances were grouped into new groups. We then evaluated the identified groups in terms of the newly discovered knowledge (see Section 3.2). We checked the results against a visual examination and expert opinion.

## 3 Results and Discussion

### 3.1 Discriminating Classes

Results for the discrimination of classes of interneurons on which the scientific community largely agrees are shown in Fig. 1. The evolution of the results depended on the threshold, and we find that, generally, SeSProC performance improved with a higher threshold. A higher threshold means that more experts agreed with the labels, and neurons were easier to classify.

[2] We base our model on the *cluster assumption* [12], which states that instances that belong to the same cluster are likely to be of the same class, whereas a class may be represented by several clusters.

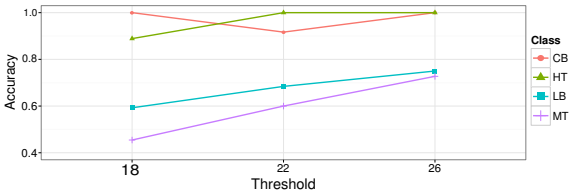


**Fig. 1.** Accuracy values depending on the class and the selected threshold

Results for discriminating CB cells from other classes were very accurate. For example, the 49 CB cells of *th18* were correctly discriminated from the other classes. All cells were again discriminated with *th26*, and only two (out of 24) cells were misclassified and grouped into the LB group with *th22*.

HT cells were the most distinct class of interneurons reported in this research. This was demonstrated by the good discrimination rate of HT cells from other classes, since only one out of nine neurons was misclassified with *th18*. There were no misclassified cells with *th22* and *th26*.

Discrimination of LB cells was worse than for CB and HT cells. All misclassified LB cells were confused with the CB class. This was anticipated because the shapes of some of these cells are very similar. However, many other LB cells have very different shapes, and this was identified by assigning these cells to new groups.

The discrimination of MT cells from other classes was the least accurate according to our data. With *th18*, 18 out of 33 MT cells were confused with other classes, even with the HT class. The *th18* group was conformed by a heterogeneous group of cells, which likely resemble other morphologies, but that still were considered as MT cells by the experts. However, only 6 out of 22 MT cells were misclassified with *th26*. These results revealed that the *th26* group was likely composed of cells that are morphologically distinct, i.e. those considered as representative MT cells. The discrimination rate for this class improved more than any other in the study.

Although the overall results showed an acceptable discrimination rate between the four classes of interneurons, there were some misclassified instances that were grouped into the wrong clusters. As the results of [10] show, the agreement among expert neuroscientists was rather limited. Therefore, it is far from easy to automatically discriminate cells perfectly. Regarding this point, Fig 2 shows four cells, two labeled, with at least 18 votes in [10], as CB (neurons A and C) and two as LB (neurons B and D). The discrimination between neurons A and B is visually very clear. However, differences between neurons C and D are less clear. The variability of shapes, sizes, and orientations when dealing with different populations of neurons is very high, which makes then hard to discriminate automatically based on morphological features.

To illustrate the problems related to the automatic discrimination of these classes, we performed experiments using a supervised classification approach to classify the instances. We used the naïve Bayes (NB) algorithm [22]. Although these results are out of the scope of this paper, the estimated mean accuracy

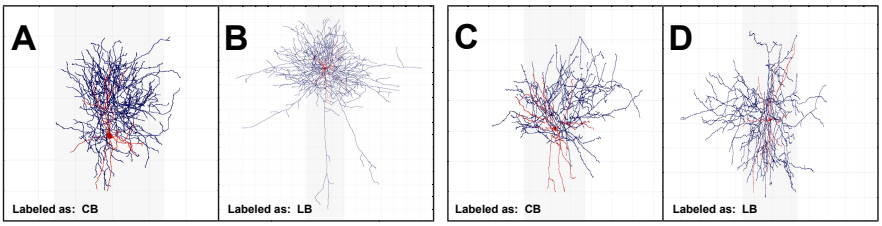


**Fig. 2.** Neurons A and C were labeled as CB and neurons B and D as LB by at least 18 experts in [10]. However, A and B are easier to discriminate than C and D. Note the different scales for each cell. Each square or rectangle in all figures represents 100µm.

values (using 10-fold cross-validation) ranged from 0.68 to 0.76 depending on the threshold and whether a feature subset selection process was performed before building the model. Although the approaches are not comparable, these values were lower than for our approach using SeSProC for averaged class results.

### 3.2 Discovering New Groups

Here we present some results illustrating the discovery of possible new groups of interneurons. Regarding CB cells, three groups were identified with *th18*. One of these groups contained 41 CB cells (see cell A in Fig. 3). The second group had three cells. The main features of these cells were that their axonal arborizations were not as dense and they had one or two descending long axonal colaterals (see cell B in Fig. 3). Finally, there were five CB cells in a third group that had very dense axonal arborizations (see cell C in Fig. 3).

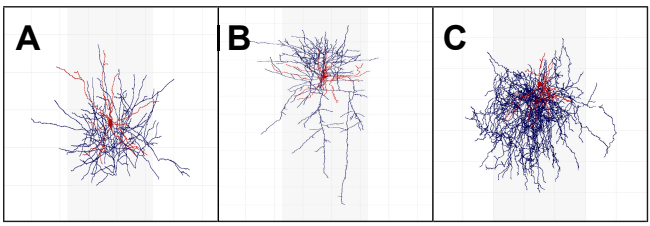


**Fig. 3.** Representative CB cells from each group identified by SeSProc with *th18*

Note that the population of cells changed when a different threshold was used. The eight CB cells that were grouped into B and C (see Fig. 3) with *th18* did not receive enough votes for inclusion in *th22* or *th26*. Therefore, these groups could not be identified. However, SeSProC did identify two new groups with *th26*, with five and four cells, which did not appear previously. The main difference between these groups was the size of the axonal arborizations of the neurons.

Fig. 4 shows an example of the relevance of each feature $j$ and group $m$ (value of $\rho_{mj}$, see Section 2.2) for CB groups with *th26* depending on the hidden class.

It is shown that feature relevance of the CB groups when hiding HT, LB, and MT was very similar, where features $X_2$ and $X_8$ were considered highly relevant. When CB labels were hidden in the input, SeSProC identified two previously commented groups. Features $X_2$, $X_3$, $X_5$, and $X_6$ were highly relevant for the first group, whereas features $X_8$ and $X_9$ were more relevant for the second group.

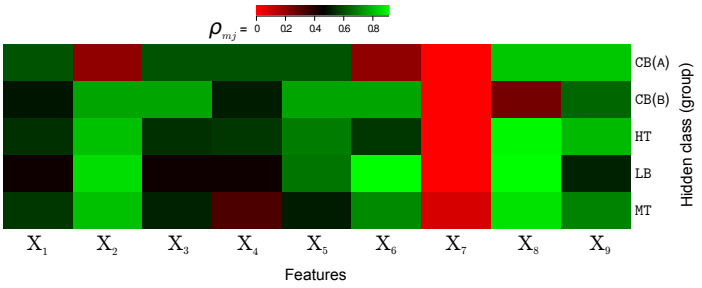


**Fig. 4.** Heatmap indicating the relevance of each feature ($X_1$-$X_9$) for each group (A and B) of CB cells when CB labels were hidden, and also for each group of CB cells when HT, LB, and MT labels were hidden (results for *th26*)

Regarding HT cells, only one group was identified for this type of cells regardless of the threshold. It shows the descending tight axonal arborizations that characterize this cell type.

SeSProC identified four, five, and three new groups of LB cells with *th18*, *th22*, and *th26*, respectively, revealing some interesting features after visual inspection. For example, regarding *th18*, one group contained LB cells with horizontally distributed axonal ramifications (see cell A in Fig. 5). Another group contained LB cells with a dense axonal arborization near the soma and a few descending long axonal colaterals (see cell B in Fig. 5). Finally, another two groups contained cells with sparse axonal arborizations distributed in several directions (see cells C and D in Fig. 5).

SeSProC only identified one group of MT cells with *th18* and *th22*. Two groups were identified with *th26*. The first group was mainly characterized by features

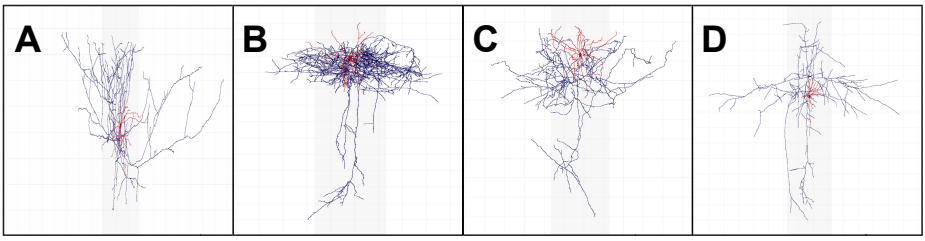


**Fig. 5.** Representative LB cells from each group identified by SeSProc with *th18*

describing the total cell size ($X_5$ to $X_9$), whereas features related to orientation ($X_1$ to $X_4$) were more relevant for the second group. Further analyses are necessary in order to obtain a more accurate classification for `MT` cells.

In summary, the fact that new groups were identified demonstrates that there is a lack in the homogeneity of the types that are defined to date. Thus, the present kind of analysis would help to advance in the understanding of the classification and characterization of neurons.

## 4 Conclusions

The classification of neurons is considered as one of the most challenging problems related to the study of neuronal circuits because data are scarce, experts disagree, and cells are morphologically, molecularly, and physiologically variable. We present a novel semi-supervised approach for classifying morphological neuron data, leading to the discovery of possible groups of neurons, which takes advantage of previous knowledge in the shape of data labels, and also identifies the relevance of each feature for each group.

We obtained a preliminary distinction among different classes of interneurons according to simple morphological features characterizing the size and the orientation of axonal and dendritic arbors of cells. We tackled this problem from a simple perspective regarding the morphological features since experts classify cells by visual examination. Although SeSProC outperformed a supervised classification approach, the most interesting output of our approach is related to the identification of new groups. Although preliminary results look interesting, especially for `CB` and `LB` cells, further analyses using different morphological features and labels are necessary to confirm these results.

SeSProC is open to further improvements, like the inclusion of uncertainty into labels. Instead of considering different thresholds to retrieve expert knowledge, it would be interesting to include information gathered from many experts in the shape of labels with some probability. Regarding the data, although it is generally thought that the same morphological types of neurons are found in all species, we cannot discard the possibility of inter-species variability, and further analyses are necessary in order to find representative types of neurons of particular species.

**Acknowledgements.** This research is partially supported by the Cajal Blue Brain Project (Spanish partner of the Blue Brain Project initiative from EPFL), the Spanish Ministry of Economy and Competitiveness TIN2010-20900-C04-04 and TIN2010-21289-C02-02 projects, and Consolider Ingenio 2010-CSD2007-00018.

## References

[1] Petilla Interneuron Nomenclature Group: Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. Nat. Rev. Neurosci. 9, 557–568 (2008)

[2] Marin, E.C., Jefferys, G., Komiyama, T., Zhu, H., Luo, L.: Representation of the glomerular olfactory map in the drosophila brain. Cell 149, 243–255 (2002)
[3] Guerra, L., McGarry, L.M., Robles, V., Bielza, C., Larrañaga, P., Yuste, R.: Comparison between supervised and unsupervised classifications of neuronal cell types: A case study. Dev. Neurobiol. 71(1), 71–82 (2011)
[4] Tsiola, A., Hamzei-Sichani, F., Peterlin, Z., Yuste, R.: Quantitative morphological classification of layer 5 neurons from mouse primary visual cortex. J. Compar. Neurol. 461, 415–428 (2003)
[5] Benavides-Piccione, R., Sichani, F.H., Yaez, I.B., DeFelipe, J., Yuste, R.: Dendritic size of pyramidal neurons differs among mouse cortical regions. Cereb. Cortex. 16, 990–1001 (2005)
[6] Cauli, B., Porter, J.T., Tsuzuki, K., Lambolez, B., Rossier, J., Quenet, B., Audinat, E.: Classification of fusiform neocortical interneurons based on unsupervised clustering. Proc. Natl. Acad. Sci. 97(11), 6144–6149 (2000)
[7] Karagiannis, A., Gallopin, T., Csaba, D., Battaglia, D., Geoffroy, H., Rossier, J., Hillman, E., et al.: Classification of NPY-expressing neocortical interneurons. J. Neurosci. 29, 3642–3659 (2009)
[8] McGarry, L.M., Packer, A., Fino, E., Nikolenko, V., Sippy, T., Yuste, R.: Quantitative classification of somatostatin-positive neocortical interneurons identifies three interneuron subtypes. Front. Neural Circuits. 4(12), 1–19 (2010)
[9] DeFelipe, J.: Cortical interneurons: From Cajal to 2001. Prog. Brain Res. 136, 215–238 (2002)
[10] DeFelipe, J., et al.: New insights in the classification and nomenclature of cortical GABAergic interneurons. Nat. Rev. Neurosci. 14(3), 202–216 (2013)
[11] McLachlan, G., Basford, K.: Mixture Models: Inference and Applications to Clustering. Marcel Dekker (1988)
[12] Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-Supervised Learning. MIT Press (2006)
[13] Zhu, X., Goldberg, A.: Introduction to Semi-Supervised Learning. Morgan & Claypool Publishers (2009)
[14] Parsons, L., Haque, E., Liu, H.: Subspace clustering for high dimensional data: A review. ACM SIGKDD Explorations Newsletter - Special Issue on Learning From Imbalanced Datasets 6(1), 90–105 (2004)
[15] Kriegel, H., Kröger, P., Zimek, A.: Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering and correlation clustering. ACM Trans. Knowl. Discov. Data. 3(1), 1–58 (2009)
[16] Law, M.H.C., Figueiredo, M.A.T., Jain, A.K.: Simultaneous feature selection and clustering using mixture models. IEEE T. Pattern Anal. 26(9), 1154–1166 (2004)
[17] Vellido, A., Lisboa, P.J.G., Vicente, D.: Robust analysis of MRS brain tumour data using t-GTM. Neurocomputing 69(79), 754–768 (2006)
[18] Ascoli, G.A., Donohue, D.E., Halavi, M.: NeuroMorpho. Org: A central resource for neuronal morphologies. J. Neurosci. 27(35), 9247–9251 (2007)
[19] Guerra, L., Bielza, C., Robles, V., Larrañaga, P.: Semi-supervised projected model-based clustering. Data Min. Knowl. Disc. (2012) (submitted)
[20] Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. 39(1), 1–38 (1977)
[21] Schwarz, G.: Estimating the dimension of a model. Ann. Stat. 6(2), 461–464 (1978)
[22] Minsky, M.: Steps toward artificial intelligence. In: Computers and Thought, pp. 406–450. McGraw-Hill (1961)