# Online labelling strategies for growing neural gas

Oliver Beyer and Philipp Cimiano

Semantic Computing Group, CITEC, Bielefeld University,
obeyer@cit-ec.uni-bielefeld.de
http://www.sc.cit-ec.uni-bielefeld.de

**Abstract.** Growing neural gas (GNG) has been successfully applied to unsupervised learning problems. However, GNG-inspired approaches can also be applied to classification problems, provided they are extended with an appropriate labelling function. Most approaches along these lines have so far relied on strategies which label neurons a posteriori, after the training has been completed. As a consequence, such approaches require the training data to be stored until the labelling phase, which runs directly counter to the online nature of GNG. Thus, in order to restore the online property of classification approaches based on GNG, we present an approach in which the labelling is performed online. This online labelling strategy better matches the online nature of GNG where only neurons – but no explicit training examples – are stored. As the main contribution, we show that online labelling strategies do not deteriorate the performance compared to offline labelling strategies.

**Keywords:** artificial neural networks, growing neural gas, clustering, classification, labeling

## 1 Introduction

Self-organising approaches such as self-organising maps (SOM) [5], learning vector quantization (LVQ) [6], or neural gas (NG) [11] are often successfully used in unsupervised learning tasks, clustering in particular. The advantage of such neurally-inspired clustering approaches lies in their ability to learn the representation of a feature space without supervision. A further interesting property is the fact that they typically perform dimensionality reduction. Typical applications of these learning paradigms can be found in areas such as text mining [3, 12, 7] or pattern recognition [10]. Growing neural gas (GNG) [2] represents an extension of the NG algorithm in which the number of neurons is not fixed a priori as in NG, but grows over time. This is especially interesting in such clustering tasks where the number of clusters is previously unknown.

In order to apply self-organising approaches to classification problems, two extensions are necessary: i) a function which assigns labels to neurons and ii) a function that performs the prediction on unseen data points. So far, mainly offline labelling strategies which require the explicit storage of labelled training data have been considered for the first case [3, 12, 8–10]. They perform the assignment of labels to neurons a posteriori, after the training phase has been ended.

There are several disadvantages connected to these strategies. For example, offline labelling strategies assume that there is a definite end of the training phase after which the labelling can be performed. This is of course not given in life-long learning scenarios in which training and prediction are interleaved. Using offline labelling strategies, we are not able to perform predictions for unseen examples in every iteration step, which is the crucial characteristic of online classification approaches. In the area of cognitive systems engineering, the online nature of learning processes is crucial in order to learn complex behaviours incrementally and continuously [14]. Another disadvantage of offline labelling strategies is the fact that they directly run counter to the online nature of GNG, as training examples are stored explicitly.

In this paper we thus present and evaluate several strategies allowing to perform the labelling on-the-fly, thus extending GNG to an online classification algorithm. We compare these strategies to several offline strategies that have been proposed in the literature and examine in particular whether an online labelling strategy can compete in terms of performance, i.e. classification accuracy, with an a posteriori labelling strategy. In fact, we show on three data sets (one artificial and two standard data sets) that an online labelling strategy does not perform significantly worse compared to an offline labelling strategy.

We offer the following contributions in particular:

- We systematically evaluate different offline labelling strategies for GNG in a classification task.
- We extend GNG by an on-the-fly labelling step that allows us to extend GNG to an online classification algorithm.
- We present and systematically analyse various online labelling strategies and compare them to the offline labelling strategies, showing that they do not deteriorate the performance of a classification approach based on GNG.

The paper is structured as follows: in Section 2 we discuss how GNG can be extended to a classification algorithm and what labelling functions are typically used in the literature. In Section 3 we present our extension of GNG to an online classification algorithm, relying on a novel set of labelling functions which perform labelling on-the-fly. In Section 4, we present our experimental results obtained on the basis of three data sets. We discuss related work and conclude in Section 5.

## 2 Classification with GNG

In this article we rely on the original version of GNG introduced by Fritzke [2] which is an extension of NG [11]. The algorithm is depicted in Figure 1 (without step 4). The GNG algorithm initially starts with a small network of two neurons in step 1. In steps 2-8, a winner neuron and its neighbouring neurons are determined and adapted according to the presented input example (stimulus). In steps 9 and 10, neurons are removed and inserted into the network according to a fixed set of parameters. The algorithm stops when a predefined criterion is met, e.g. when the network has reached a certain size.

In order to apply GNG to a classification task, most approaches in the literature extend the algorithm by two functions. A neuron labelling function $l : N \rightarrow C$ where $C$ is the set of class labels, and a prediction function $l : D \rightarrow C$ where $D$ is the input space. We analyse the following offline neuron labelling functions as proposed by Lau et al. [9]. They are offline in the sense that they assume that the pairs $(d, l_d)$ with $d \in D_{train} \subseteq D$ and $l_d \in C$ seen in the training phase are explicitly stored:

- *Minimal-distance method (min-dist):* According to this strategy, neuron $n_i$ adopts the label $l_d$ of the closest data point $d \in D_{train}$:
$$l(n_i) = l_d = l(\arg \min_{d \in D_{train}} |n_i - d|^2)$$
- *Average-distance method (avg-dist):* According to this strategy, we assign to neuron $n_i$ the label of the category $c$ that minimises the average distance to all data points labelled with category $c$:
$$l(n_i) = \arg \min_c \sum_{k=1}^{|D(c)|} \frac{|n_i - d_k|^2}{|D(c)|}$$
where $D(c) = \{d \in D_{train} \mid l(d) = c\}$ is the set of all examples labelled with $c$.
- *Voronoi method (voronoi):* According to this strategy, we label neuron $n_i$ with that category $c$ having the highest overlap (in terms of data points labelled with $c$) with the data points in the voronoi cell for $n_i$. We denote the set of data points in the voronoi cell for $n_i$ as $v(n_i) = \{d \in D_{train} \mid \forall n_j, j \neq i : |n_j - d|^2 \geq |n_i - d|^2\}$ within the topological map.
$$l(n_i) = \arg \max_c |D(c) \cap v(n_i)|$$

In addition to the neuron labelling strategy, we need to define prediction functions that assign labels to unseen examples. These prediction strategies are inspired by linkage strategies typically used in cluster analysis (see [4, 1, 13]):

- *Single-linkage:* In this prediction strategy a new data point $d_{new}$ is labelled with category $c$ of the neuron $n$ that minimises the distance to this new example:
$$l(d_{new}) = \arg \min_c (\arg \min_{n \in N(c)} |n - d_{new}|^2)$$
where $N(c) = \{n \in N \mid l(n) = c\}$ is the set of all neurons labelled with category $c$ according to one of the above mentioned neuron labelling function.
- *Average-linkage:* In this strategy, example $d_{new}$ adopts the label of category $c$ having the minimal average distance to the example:
$$l(d_{new}) = \arg \min_c (\sum_{k=1}^{|N(c)|} \frac{|n_k - d_{new}|^2}{|N(c)|})$$
- *Complete-linkage:* In this prediction strategy a new data point $d_{new}$ is labelled with category $c$ of the neuron $n$ that minimises the maximal distance to this new example:
$$l(d_{new}) = \arg \min_c (\arg \max_{n \in N(c)} |n - d_{new}|^2)$$

## 3 Online labelling strategies for GNG

In order to extend GNG to an online classification algorithm, we extend the basic GNG by a step in which the label of the presented stimulus is assigned to the winner neuron during the learning process. We denote the winner neuron for data point $d$ by $w(d)$. All prediction strategies are local in the sense that they do not consider any neighbouring neurons besides the winner neuron $w(d)$. As the labelling is performed on-the-fly, the label assigned to a neuron can change over time, so that the labelling function is dependent on the number of examples the network has seen and has the following form: $l : N \times T \rightarrow C$. We will simply write $l_t(n_i)$ to denote the label assigned to neuron $n_i$ after having seen $t$ data points.

- *Relabelling method (relabel)*: According to this very simple strategy, the winner neuron $w(d)$ adopts the label of $d$:

$$l_t(n_i) = l_d, \ \textit{where } n_i = w(d)$$

- *Frequency-based method (freq)*: We assume that each neuron stores information about how often a data point of a certain category has been assigned to $n_i$ after $t$ examples have been presented to the network as $freq_t(c, n_i)$. A neuron is labelled by the category which maximises this frequency, i.e.

$$l_t(n_i) = \arg\max_c freq_t(c, n_i)$$

- *Limited-distance method (limit)*: According to this strategy, the winner neuron $n_i = w(d)$ adopts the category label $l_d$ of the data point $d$ if the distance between them is lower than the adaptive distance $\theta_t(n_i)$ of the neuron $n_i$. In case of a smaller distance, $\theta_t(n_i)$ will be updated with the new distance.

$$l_t(n_i) = \begin{cases} l_d, & \textit{if } |n_i - d|^2 \leq \theta_t(n_i) \\ l_{t-1}(n_i), & \textit{else} \end{cases}$$

As these labelling strategies do not guarantee that every neuron in the network is actually labelled, we need to extend the prediction strategy to handle unlabelled neurons. For the presented prediction strategies we simply ignore unlabelled neurons during the prediction state.

## 4 Experiments and results

We evaluate the labelling strategies in dependance of the mentioned prediction methods on three classification data sets: an artificial data set generated following a gaussian distribution, the ORL face database[1] and the image segmentation data set of the UCI machine learning database[2]:

[1] Samaria, F.S.: Parameterisation of a stochastic model for human face identification, *In Proc. of the IEEE Workshop on Applications of Computer Vision 1994*, 138-142

[2] Blake, C.L., Merz, C.J.: UCI repository of machine learning databases, 1998

```
Online labelling for growing neural gas (OGNG)

1. Start with two units $i$ and $j$ at random positions in the input space.
2. Present an input vector $x \in R^n$ from the input set or according to input distribution.
3. Find the nearest unit $s_1$ and the second nearest unit $s_2$.
4. Assign the label of $x$ to $s_1$ according to the present labelling strategy.
5. Increment the age of all edges emanating from $s_1$.
6. Update the local error variable by adding the squared distance between $w_{s_1}$ and $x$.
7. Move $s_1$ and all its topological neighbours (i.e. all the nodes connected to $s_1$ $\Delta error(s_1) = |w_{s_1} - x|^2$ by an edge) towards $x$ by fractions of $e_b$ and $e_n$ of the distance:
       $\Delta w_{s_1} = e_b(x - w_{s_1})$
       $\Delta w_n = e_n(x - w_n)$
   for all direct neighbours of $s_1$.
8. If $s_1$ and $s_2$ are connected by an edge, set the age of the edge to 0 (refresh). If there is no such edge, create one.
9. Remove edges with their age larger than $a_{max}$. If this results in nodes having no emanating edges, remove them as well.
10. If the number of input vectors presented or generated so far is an integer or multiple of a parameter $\lambda$, insert $a$ new node $r$ as follows:
    Determine unit $q$ with the largest error.
    Among the neighbours of $q$, find node $f$ with the largest error.
    Insert $a$ new node $r$ halfway between $q$ and $f$ as follows:
        $w_r = \frac{w_q + w_f}{2}$
    Create edges between $r$ and $q$, and $r$ and $f$. Remove the edge between $q$ and $f$.
    Decrease the error variable of $q$ and $f$ by multiplying them with a constant $\alpha$. Set the error $r$ with the new error variable of $q$.
11. Decrease all error variables of all nodes $i$ by a factor $\beta$.
12. If the stopping criterion is not met, go back to step (2). (For our experiments, the stopping criterion has been set to be the maximum network size.)
```

**Fig. 1.** GNG algorithm with extension for online labelling

*Artificial data set (ART):* The first data set is a two dimensional gaussian mixture distribution with 6 classes located at [0,6], [-2,2], [2,2], [0,-6], [-2,-2], [2,-2]. The data points of each class are gaussian distributed with the standard derivation of 1.
*ORL face database (ORL):* The second data set is the ORL face database which contains 400 frontal images of humans, performing different gestures. The data set consists of 40 individuals showing 10 gestures each. We downscaled each image from $92 \times 112$ to $46 \times 56$ and applied a principal component analysis (PCA) to reduce the number of dimensions from 2576 to 60, corresponding to 86.65% of the total variance.
*Image Segmentation data set (SEG):* The image segmentation data set consists of 2310 instances from 7 randomly selected outdoor images (brick-face, sky, foliage, cement, window, path, grass). Each instance includes 19 attributes that describe a $3 \times 3$ region within one of the images.

In order to compare the different labelling strategies to each other, we chose a fixed set of parameters for GNG and used this set in all of our experiments. This set was empirically determined on a trial-and-error basis through preliminary experiments. The GNG parameters are set as follows: insertion parameter $\lambda = 300$; maximum age $a_{max} = 120$; adaptation parameter for winner $e_b = 0.2$; adaptation parameter for neighbourhood $e_n = 0.006$; error variable decrease $\alpha = 0.5$; error variable decrease $\beta = 0.995$

For our experiments we randomly sampled 10 training/test sets out of our data and averaged the accuracy for each labelling and prediction setting. Thereby, we trained each GNG classifier with 4 labelled examples of each category. We validated the accuracy of the labelling methods and prediction strategies on test. Our results are shown in Tables 1 and 2. Both tables show the classification accuracy for various configurations of labelling methods (min-dist, avg-dist, voronoi, relabel, freq, limit) and prediction strategies (single-linkage, average-linkage and complete-linkage) averaged over the three different data sets using 4 training samples per class. We evaluated the accuracy of each labelling method combined with three prediction strategies (rows of the tables). Therefore, we consider the results of 54 experiments overall [3]. The results license the following conclusions:

- **Comparison of offline labelling strategies:** According to Table 1, there is no labelling method which significantly outperforms the others. Comparing the accuracy results averaged over all prediction strategies, the *voronoi method* is the most effective labelling method as it provides the highest accuracy with 77.59%, followed by the *min-dist method* with 76.27% and the *avg-dist method* with 74.28%. Concerning the prediction strategies, the *single-linkage prediction* strategy shows best results averaged over all methods with 81.41%, followed by the *average-linkage prediction* strategy with an accuracy of 77.65%. The *complete-linkage* yielded the worst results with an averaged accuracy of 69.07%.
- **Comparison of online labelling strategies:** According to Table 2, all three online labelling strategies are almost equal in their classification performance. The *limit method* performs slightly better than the other two methods and achieves an accuracy of 78.15%, followed by the *freq method* with an accuracy of 78.09% and the *relabel method* with an accuracy of 77.88%. As for the offline labelling strategies, here it is also the case that the *single-linkage prediction* is the best choice with an accuracy of 83.30%, followed by the *average-linkage prediction* with an accuracy of 80.90% and the *complete-linkage prediction* with an accuracy of 69.88%.
- **Online vs. offline labelling strategies:** Comparing the averaged accuracy of all labelling methods of Table 1 and 2, the results show that there is no significant difference between them in terms of performance. The online labelling methods even provide a slightly higher accuracy.
- **Impact of memory**: Strategies relying on some sort of memory (e.g. storing the frequency of seen labels as in the *freq method*), do not perform significantly better than a simple context-free (or memory-less) method (*relabel*

[3] The results of all 54 experiments can be found at http://www.sc.cit-ec.uni-bielefeld.de/people/obeyer

| Prediction strategies | Min-dist method | Avg-dist method | Voronoi method | *Average* |
|---|---|---|---|---|
| Single-linkage | 81.93 | 78.92 | 83.39 | **81.41** |
| Average-linkage | 77.15 | 76.35 | 79.46 | 77.65 |
| Complete-linkage | 69.72 | 67.56 | 69.92 | 69.07 |
| *Average* | 76.27 | 74.28 | **77.59** | |

**Table 1.** Classification accuracy for the offline labelling strategies (min-dist, avg-dist, voronoi) combined with the prediction strategies (single-linkage, average-linkage, complete-linkage) averaged over the three data sets (ART, ORL, SEG) trained with 4 labelled data points of each category (best prediction strategy and labelling method results are marked).

| Prediction strategies | Relabel method | Freq method | Limit method | *Average* |
|---|---|---|---|---|
| Single-Linkage | 83.25 | 83.25 | 83.39 | **83.30** |
| Average-Linkage | 80.46 | 81.12 | 81.13 | 80.90 |
| Complete-Linkage | 69.92 | 69.79 | 69.93 | 69.88 |
| *Average* | 77.88 | 78.05 | **78.15** | |

**Table 2.** Classification accuracy for the online labelling strategies (relabel, freq, limit) combined with the prediction strategies (single-linkage, average-linkage, complete-linkage) averaged over the three data sets (ART, ORL, SEG) trained with 4 labelled data points of each category (best prediction strategy and labelling method results are marked).

*method*) performing decisions on the basis of new data points only. This shows that the implementation of a label memory does not enhance the classifiers performance.

## 5 Related Work and Conclusion

In this paper we have presented, analysed and compared different online labelling strategies in order to extend the growing neural gas (GNG) algorithm to an online classification approach. While GNG is essentially an unsupervised algorithm, previous approaches have presented extensions of GNG for classification tasks. Such extensions typically rely on a suitable labelling function that assigns labels to neurons as well as a prediction function that assigns labels to unseen examples. In this line, we have experimentally compared different offline and online labelling strategies inspired by previous research. To our knowledge, there has been no systematic investigation and comparison of different offline strategies so far, a gap we have intended to fill. The question of how GNG can be extended to an online classification algorithm has also not been addressed previously. In most cases, offline strategies have been considered that perform the labelling after the training phase has ended and the network has stabilised to some extent as in the WEBSOM [7, 8] and LabelSOM [12] approaches. In both of these approaches, the label assignment is essentially determined by the distance of the labelled training data point to the neurons of the already trained network. Such offline labelling strategies contradict the online nature of GNG, whose interesting properties are that the network grows over time and only neurons, but no explicit examples, need to be stored in the network. In this sense, an important question we have addressed in this paper is whether GNG can be extended to a classification algorithm without affecting its online nature or degrading performance considerably. Our research has shown that this is indeed possible. Online labelling functions where the label of a neuron can change over time and is computed when a new example is assigned to the neuron in question do not perform worse than offline labelling strategies.

## Acknowledgement

This project has been funded by the Excellence Initiative of the Deutsche Forschungsgemeinschaft (DFG).

## References

1. Anderberg, M.R.: Cluster analysis for applications, *Academic Press, New York* (1973)
2. Fritzke, B.: A growing neural gas network learns topologies, *In Proceedings of the International Conference on Advances in Neural Information Processing Systems (NIPS) 1995*, 625-632 (1995)
3. Hyotyniemi, H.: Text document classification with self-organizing maps, *In Proc. of the Finnish Artificial Intelligence Conf. (STeP) 1996*, 64-72 (1996)
4. Johnson, S.C.: Hierarchical clustering schemes, *Psychometrika*, 32(3):241-254 (1967)
5. Kohonen, T.: Self-organized formation of topologically correct feature maps, *Biological cybernetics*, 43(1):59-69 (1982)
6. Kohonen, T.: Learning vector quantization. *The Handbook of Brain Theory and Neural Networks, MIT Press*, 537540 (1995)
7. Kohonen T., Kaski S., Lagus K.: Self organization of a massive document collection, *IEEE Transactions on Neural Networks*, 11(3):574-585 (2000)
8. Lagus K., Kaski S.: Keyword selection method for characterizing text document maps, *In Proc. of the Int. Conf. on Artificial Neural Networks (ICANN)*, 371-376 (1999)
9. Lau K., Yin H., Hubbard S.: Kernel self-organising maps for classification, *Neurocomputing*, 69(16-18):2033-2040 (2006)
10. Lefebvre G., Garcia C.: A probabilistic self-organizing map for facial recognition, *In Proc. of the Int. Conf. on Pattern Recognition (ICPR) 2008*, 1-4 (2008)
11. Martinetz, T., Schluten, K.: A "neural-gas" network learns topologies. *In Proc. of the Int. Conf. on Artificial Neural Networks (ICANN) 1991*, 397-402 (1991)
12. Rauber A.: LabelSOM: On the labeling of self-organizing maps, *In Proc. of the Int. Joint Conf. on Neural Networks (IJCNN) 1999*, 3527-3532 (1999)
13. Sneath, P.H.A., Sokal R.R.: Numerical taxonomy: The principles and practice of numerical classification, *W H Freeman & Co (Sd); First edition* (1973)
14. Steil J.J., Wersing H.: Recent trends in online learning for cognitive robotics, *In Proc. of the European Symposium on Artificial Neural Networks (ESANN) 2006*, 77-87 (2006)