A Comparison of the Self-Organizing Map and Growing Neural Gas Network in the Context of Optical Character Recognition

Kaleigh Clary
Department of Computer Science & Mathematics
Hendrix College
claryka@hendrix.edu

1 Problem and Motivation

With the advent of tablet and touch screen computing, the number of applications that utilize written text recognition technologies is rapidly increasing. These applications rely on fast and accurate optical character recognition (OCR) algorithms. The self-organizing map (SOM) is an unsupervised algorithm capable of great performance in machine learning. As such, it has become a benchmark in OCR. In spite of its popularity, it does have some significant shortcomings. Its training algorithm is inefficient, its accuracy is highly dependent on a classification algorithm, and it generates unnecessary hybrid output nodes. This research investigated alternative data structures for application in optical character recognition. The growing neural gas network (GNG) is an alteration of the SOM that was developed to address these shortcomings. The purpose of this research was to assess whether the GNG network compensates for the SOM's shortcomings in the context of OCR.

2 Background and Related Work

The self-organizing map is an unsupervised machine learning algorithm, meaning it organizes data without the use of labels. The SOM is an algorithm that constructs an array of ideal representation output nodes based on vector inputs. Each node is connected to every other node in the network, but only one node responds to a particular input. Response positions are ordered according to the sum of squared differences between vector inputs, so similar nodes cluster together [1]. The network forms an ideal representation of output for a particular input.

Figure 1 is an example representation of a SOM of colors. Each node is a color, and each vector input is an RGB value.

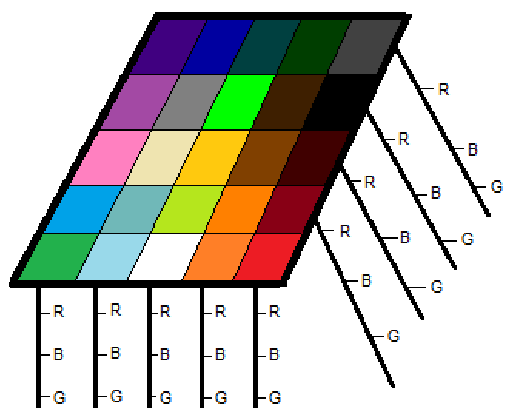

Figure 1: Example of a self-organizing map of colors. There is an RGB value associated with each node in the map. Note that like colors are organized into clusters.

In OCR applications, each node is a representation of a letter, and each vector input is a letter drawing. During the training algorithm, the map organizes representations of letters according to closeness between those representations. Closeness here is measured by the sum of squared differences.

The map is of fixed size, so its dimension must be assigned a priori. The most appropriate map size is determined by testing the performance of maps varying in size. Since the SOM training algorithm runs in quadratic time in the size of the map [1], any testing to determine the optimal size of the map is of limited efficiency. The SOM also generates unnecessary output nodes. Because of the way nodes are trained, hybrid letter outputs are formed in the network. These are nodes that represent a combination of letters. Hybrid outputs can be helpful in other applications where labels may allow overlapping classification. This is not often an issue in optical character recognition because letters have distinct representations. Since the SOM is an unsupervised machine learning algorithm, the recognition accuracy is highly dependent on the classification algorithm used to assign labels to various output nodes.

The GNG is an unsupervised machine learning algorithm made up of nodes and their connections. The arrangement of the nodes is determined by the input data. When applied to OCR, nodes represent the placement of valid pixels. The network creates a neighborhood of nodes to model the topological relationship of the data inputs; nodes with similar input values are connected [2]. The network is then a Voronoi diagram of vector inputs [3]. A node's neighborhood is simply the collection of the other nodes to which it is connected. A neighborhood connection indicates that either node has influence on the other during training. Each connection has an age, which monitors whether the connection is still valid. Relevant connections are refreshed every iteration of the training algorithm. One image at a time is used for training.

Figure 2 shows a visual representation of a growing neural gas network trained on C-labeled drawings. Each blue circle represents a node, and the lines between nodes indicate that the two connected nodes are neighbors.

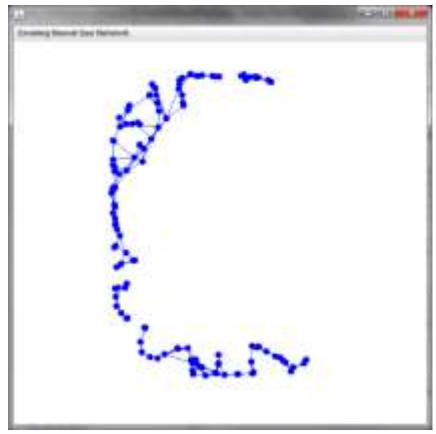

Figure 2: Example GNG trained on a set of C-labeled drawings.

The GNG is an iterative alteration of the neural gas network. There has been some discussion of the SOM's performance in OCR in comparison to the neural gas network [3]. As yet, no formal comparison between the SOM and GNG has been conducted.

3 Approach and Uniqueness

Results from [3] indicate that the neural gas network is more accurate in OCR than the SOM, and results from [2] indicate that the GNG is more accurate than the neural gas network. This research represents the first direct comparison between the SOM and GNG.

The design of the GNG eliminates the problems with the SOM in optical character recognition. In contrast to the SOM, the GNG network grows to a predetermined number of nodes, so the dimensions of the map need not be considered, and it trains in linear time. Given these advantages, we hypothesized that the GNG would more accurately identify letter drawings in OCR.

In this application, the GNG network models the shape of a letter based on the placement of valid pixels in a drawing, and each network represents one letter. We trained a distinct network for each letter label in the training sample. This addresses the hybrid letter issue with using the SOM and removes the need for a classification algorithm. Creating this array of networks corresponding to different labels makes this algorithm supervised since a priori knowledge of labels is required to generate one network per label.

Because the GNG network is an alteration of the SOM designed to address problems that arise in optical character recognition applications, we hypothesized that it is more accurate and more efficient than the SOM in letter identification. We developed implementations of the SOM and the GNG and compared their performance in training time and recognition accuracy. The two networks were trained using the same set of sample letter drawings, and recognition accuracy was tested using a separate set of letter drawings. In each sample, the letter was drawn at a standard size and appeared in the center of the image. Both algorithms were tested with the same set of training and testing samples. The data set is organized into labels based on letter. The networks were tested with various numbers of labels ranging 2-9.

Several classification algorithms were developed for the SOM, and the results reported are those for the classficiation algorithm with the highest percentage of corrently identified images from the test set. This algorithm determined, for each node of the trained map, the label with the smallest distance from the node's output. A drawing was classified by finding the node with the smallest distance from the drawing and return the label previously determined for that node. Distance is again defined as the sum of squared differences.

A similar classification algorithm was used for the GNG. A drawing was classified by determining which network had the smallest distance between itself and the drawing and returning the label of that network.

Each test set is made of up at least 10 binary images of letter characters. Binary images were selected so that image noise would not influence results. Each pixel in the image is either on or off; if it is on, it is part of the character. Otherwise, the pixel is ignored in training and identification. The modified networks were tested using 9 different character image sets, each with 10 or more images.

Images were generated in a Java application. Each character set (both training and test sets) was drawn by one person using a mouse in the center of a JPanel canvas. An example from the C-labeled drawing set is shown in Figure 3.

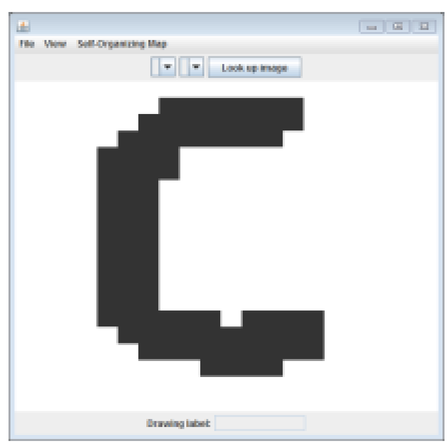

Figure 3: Example C-labeled drawing from the training set. It is a binary image.

4 Results and Contributions

The results of our experimentation appear in Figures 4 and 5.

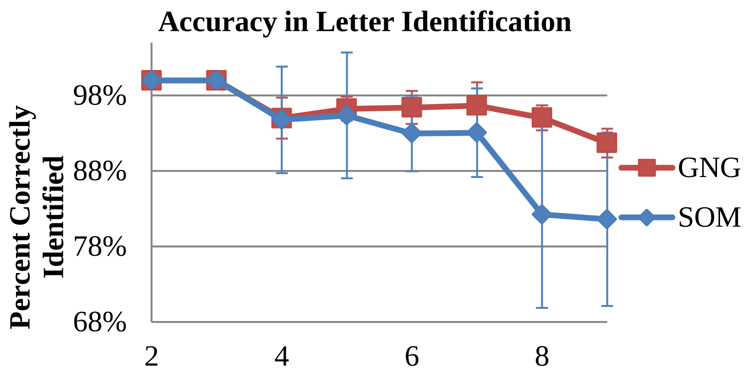

Figure 4: Graph of recognition accuracy in letter identification for SOMs and GNGs with 2-9 labels. There is a 95% confidence interval.

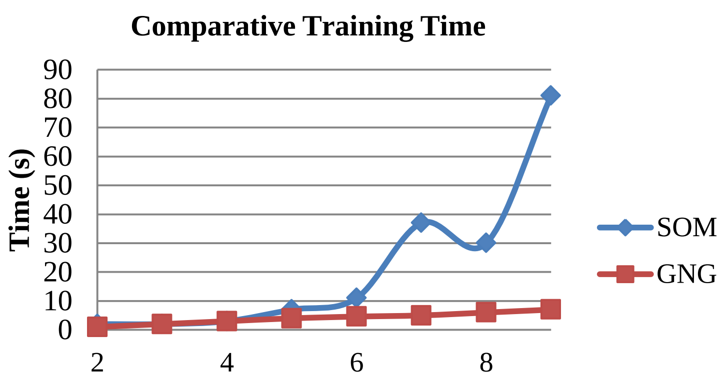

Figure 5: Graph of training time in seconds for SOMs and GNGs with 2-9 labels. The SOM has quadratic training time increase as the number of labels increases. GNG network training time increases linearly.

Both networks correctly identified 100% of drawings with a small number of labels. As the number of labels increased, however, the GNG outperformed the SOM. At nine labels, the SOM correctly identified 81.6% of drawings. The GNG network correctly identified 91.7%.

The GNG also completed training much faster than the SOM. With a small (2-3 labels) data set, the SOM and GNG are both quick to train, at about 6 seconds. With a larger data set (9 labels), the SOM takes 81s to train, while the GNG networks takes 7s to train.

5 Conclusions and Future Work

This work shows that use of the growing neural gas network should be preferred over the self-organizing map in optical character recognition.

The results clearly indicate that the growing neural gas algorithm is the superior method for classifying letter drawings. It outperforms the SOM in both speed and accuracy. As the number of labels increases, the recognition accuracy of the SOM declines at a much faster rate than that of the GNG. The different complexity classes of the two algorithms are also evident from the results. The linear training time of the GNG is preferable to the quadratic training time of the SOM, especially as the number of labels increases.

These results suggest several opportunities for future work. First, this work could be repeated using a much larger data set with more drawings per label and more labels in the set.

Other future research could investigate identification of characters when there may be some variation of letter style in a single letter label set. In this work, label sets were limited to a single style by having all examples in a label drawn by the same person in both the training and testing sets. Other OCR label sets may not be segregated in this way. In that case, we predict recognition accuracy of either the SOM or

GNG would be significantly lower because the average of the set would not represent a single style.
This issue can be addressed with the GNG by modifying the training algorithm so that when a network is trained for a label, the distance between the network and each drawing in the set is calculated. If the distance is over some threshold, a new network would be generated from drawings closer to this significantly different drawing. Then each label would have a set of networks, each corresponding to a different style for that label. This mimics the clustering in an hierarchical SOM, and the performance of the two networks could be compared.

A third direction for future work is identification of transformed characters. We determined that the GNG network is preferable to the SOM in letter identification using samples of centered, correctly oriented letters. Because letters have a standard shape, relationships between nodes are preserved if the image is not in a standard orientation. Thus, we hypothesize that a network made up of nodes and their connections, like the GNG, should be able to accurately identify letters that have been transformed from the standard orientation.

Networks created from images of transformed characters are not as accurate in identifying characters because the underlying representation of the letters is not consistent. We predict that the recognition accuracy of characters that may have been distorted by various transformations can be increased by using the GNG to create a representation of the image and manipulating pixel locations through affine transformations to reduce distortion by these distortions.

Finally, performance of both networks could be compared using a label set with noisy images. It is possible that the presence of hybrid values in the SOM network could be useful for identifying noisy images.

6 References

1. Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464-1480.
2. Hölmstrom, J. (2002). *Growing neural gas: Experiments with GNG, GNG with utility and supervised GNG*. (Master's thesis, Uppsala University, Uppsala, Sweden).
3. Martinez, T. and Schulten, K. (1991). *A "neural-gas" network learns topologies.* In Kohonen, T., K. Makisara, O. Simula and J. Kangas (Eds.), *Artificial neural networks: proceedings of the 1991 International Conference on Artificial Neural Networks* (pp. 397-402). North-Holland.