Neural Network for Modeling Esthetic Selection

Tamás (Tom) Domonkos Gedeon

Department of Computer Science
Australian National University
Acton ACT 0200 Australia
tom@cs.anu.edu

Abstract. Some real world problems require significant human interaction for labeling the data, which is very expensive. Worse, in some cases, the exercise of human judgement is inherently subjective and contextual, and so the entire labeling must be done in one session, which may be too long. Our domain is the automatic generation of Mondrian-like images with an interactive interface for the user to select images. We use back-propagation neural networks to learn an approximation of a viewer's aesthetic using 2 category labelled data (images liked/disliked). We construct a data set for training in a sequential fashion related to the interactive art appreciation task, and produce an output profile which well approximates a regression task, even trained on classification data. Analysis of the learned network produces some surprises, with the discovery of some input contributions which are unexpected to the user.

Keywords: Neural networks, back-propagation, training set, incremental learning, artistic esthetic, art, Mondrian.

1 Introduction

We briefly introduce the work of the artist Piet Mondrian, discuss the computer generation and evaluation of art, propose our approach for almost real time learning and construction of a training set, describe our results and come to some conclusions.

1.1 Mondrian

Pieter Cornelis (Piet) Mondriaan, after 1912 Mondrian, (pronounced: Pete Mon-dree-on, IPA: [pit 'm_nd_i_n]) (b. Amersfoort, Netherlands, March 7, 1872 — d. New York City, February 1, 1944) was a Dutch painter.

He was an important contributor to the De Stijl art movement and group, which was founded by Theo van Doesburg. Despite being well-known, often-parodied and even trivialized, Mondriaan's paintings exhibit a complexity that belies their apparent simplicity. He is best known for his non-representational paintings that he called "compositions", consisting of rectangular forms of red, yellow, blue, white or black, separated by thick, black rectilinear lines. They are the result of a stylistic evolution that occurred over the course of nearly 30 years and continued beyond that point to the end of his life. [1].

M. Ishikawa et al. (Eds.): ICONIP 2007, Part II, LNCS 4985, pp. 666 – 674, 2008.

An example of one of Mondrian's abstract compositions is shown below:

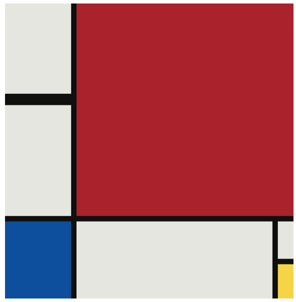

Fig. 1. Composition of Red, Blue and Yellow

There have been multiple attempts to perform mathematical analyses [2, 3] of the compositions by Piet Mondrian. None of them are successful in giving a convincing result revealing the "hidden math" within Mondrian's painting [1].

Hill [2] used "number math" for measuring the grid size, analysing the ratio between grids, and so on. One example of the conclusions reached is that some of Mondrian's compositions are triple connected, in that you can not separate the graph into two without cutting at least three lines. This applies to about half of Mondrian's work in the period 1918 – 1938.

More success was achieved by Reynolds [3] using "structural analysis" based on graph theory, which is consistent for many Mondrian works, though certainly not all.

1.2 Mondrian's Esthetic Choices

Mondrian's artistic role in esthetic choices in his compositions is still debated. Lee (2001) found that art students could not correctly identify genuine Mondrian compositions. Contrarily, McManus [4] found that the majority of subjects could distinguish between original and modified Mondrian compositions. Wolach [5] found that subjects could distinguish (preferred) Mondrian line spacings from divergent spacings. If subjects could select divergently spaced pictures that they preferred, then the preference for Mondrian spacings vanished. That is, Mondrian was good at producing compositions which appeal to many/most people, but compositions which appeal to any one person will be perceived as positively as Mondrian's by that person.

It appears then that while Mondrian could design his compositions to appeal to some communal esthetic appreciation. Where participant selection is possible, some individual esthetic appreciation is felt, which for an individual is as strong as the communal esthetic, since the preference for original Mondrian (spacings) vanish.

Our approach is to construct Mondrian-like images for individual users which they find esthetically pleasing. Our technique is described in the next sub-section.

Computer Generation of Art. We construct Mondrian-like images using a number of parameters describing a possible image. These parameters are chosen by a random process initially. Subsequently, we use an evolutionary algorithm to improve the images for each user [6].

In De Stijl, only vertical lines and horizontal lines are allowed in the graph, and all lines (almost always) terminate on other lines or the edge of the painting. The rectangle is also a basic element of Neo-Plasticism, but from a programming viewpoint, rectangles are a 'by-product' of horizontal and vertical lines. Thus a Mondrian-like graph could be deemed a collection of horizontal and vertical lines.

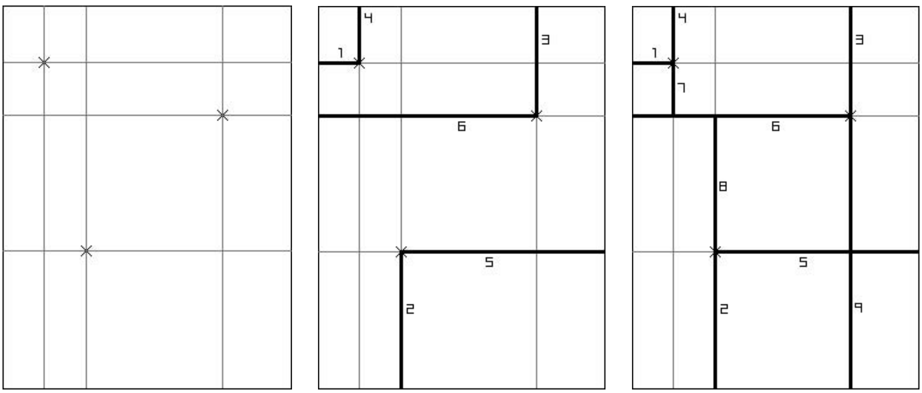

Fig. 2. Random initial points generated on a given canvas (x's on left), with imaginary lines drawn crossing the initial points (left). Draw lines emitted from each point in numbered sequence (middle). Skeleton complete (right). The final step is to randomly fill some rectangles with colour, being red, yellow, or blue (not shown above).

Clearly there are many potential choices here. How many initial points do we generate? How close may they be to each other? How far apart? How close to the edges can they get? And so on. We can deduce many properties which often hold for Mondrian's own compositions. We have already mentioned his spacing of lines.

Taylor [7] analysed the positions of 170 lines featured in 22 paintings, and found that Mondrian was twice as likely to position a line close to the canvas edge as he was to position it near the canvas centre. As we are interested in the development of individual esthetic choices, and we know [5] that the ability to chose can swamp the effect of Mondrian's own choices, we do not impose such conditions on our generation process.

In the centre of Figure 2 (above), we show a partially completed skeleton. The drawing of lines is probability based, so the same initial points can lead to different final images. Once all the lines are generated a remediation stage adds lines to eliminate remaining right angles, note this was unnecessary in the example above.

User's individual esthetic choices. To demonstrate briefly that individual esthetic choices lead to easily discernible differences in images, we provide the following examples for 4 subjects from our previous work [6].

In this paper we will concentrate on one subject as we will learn a single esthetic.

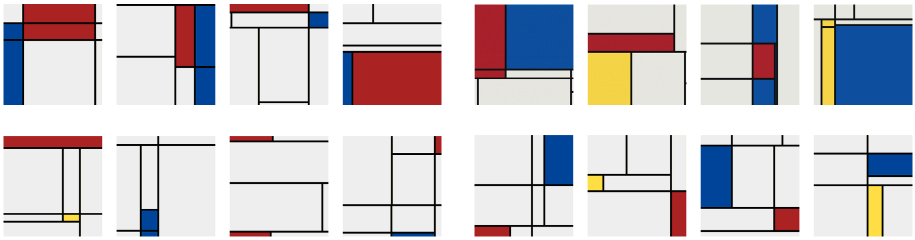

Fig. 3. Selected output for subjects (top left then right, bottom left then right) Z, P, J, T

Clearly, for Z's esthetic red and blue are present, and mostly touching along a rectangular edge (top left 4 images). Subject P seems to prefer 2 colours, particularly including blue, which usually touch, and extend across at least one of the horizontal or vertical dimensions (top right 4 images). Subject J clearly prefers small blocks of colours very close to the edges (bottom left 4 images). Finally, Subject T seems to prefer 'medium sized' blocks of colour separated by some space (bottom right 4 images). Subjects T and J were the authors of our previous work [6]. Subject T in the previous experiment will continue as the single subject for the experiments reported in this paper. This introduces an element of subjectivity which is considered in §2.

Computer Evaluation of Art. It has variously been suggested that it is worthwhile to consider trying to calculate the esthetic worth of images [8, 9].

We do not try to calculate the esthetic worth of images explicitly, rather we use neural network models to predict user preference. Hence the neural network weights arguably encode a representation of a user's esthetic preference.

Based on our previous work [6], we had some notions of the ways individual user esthetic preferences can vary. This was used to augment our search of the literature and wide internet browsing to identify image components which are likely to be significant. For example, line spacing [e.g., 5] has been extensively studied in Mondrian's work. It is also reported that Mondrian would sometimes spend days to decide on the placement of a single line. We generate Mondrian-like images from a vector based model, hence we have access to the exact point and line positionings at that stage, and hence the rectangles, and also their colours.

Table 1. Pre-processed vector image features

parameters	Description of image feature(s) encoded by parameters
1, 2	Number of lines in image; Sum of their lengths
3, 4, 5	Number of lines spanning entire image: horiz., vert. or in total
6, 7, 9, 10	Biggest/smallest distance between horizontal/vertical lines
8, 11	Length of smallest distance between lines vertically/horizontally
12, 13	Smallest/biggest distance between lines either horiz. or vert.
14, 15	Smallest horiz./vert. distance to and edge
16, 17, 18. 19	Proportion of image which is red/yellow/blue/coloured
20, 21, 22, 23, 24	Min distance between r&y/r&b/y&b/2 cols/2 cols Manhattan dist.
25, 26	Length 2 colours touch; Longest contiguous coloured areas.
27, 28	Longest parallel non-touching coloured areas; Distance between them.

2 Experimental Design and Subjectivity

The problem we are trying to solve in the experimental design is how to construct a meaningful scientific experiment in an inherently subjective domain. In many domains such as scientific and engineering, even legal, it is possible to construct data sets and run neural network experiments [9] which are straightforward to use. This is not easy in inherently subjective domains. For example, the TREC conferences provide some very large information retrieval data sets, which in essence assume that the overall ratings of search results are comparable from one expert to the next. This is likely to be a valid assumption in that domain. The equivalent in the generative art domain would be an average of the esthetic appreciation of the general population. One could argue that this is exactly what Mondrian has done in his compositions. But as we defined our focus earlier, with an emphasis on individual esthetic preference, it is not suitable for us here.

Our process using our evolutionary algorithm tool is that the user is presented with a number of generations of computer constructed Mondrian-like images at a time. The user is encouraged to indicate some active preference for, or dislike of, some images. While it is possible to rate every image, typically users will only select images for rating which they actively like or dislike. The first generation is completely at random, while subsequent generations use the rating information to improve images for that user. Ideally, the typical generation size would be 20, however the paucity of information available in that setting means improvement of images is small and likely indirectly reflects the esthetic preferences of the programmer (subject J in our previous work) implemented via the initial probabilistic settings the code uses. That is, the program clearly 'worked' when the images improved over time. This is based on observations from subject T. Note that the probabilistic settings are significant, because for a process to be seen to be generating art rather than 'just' images, some degree of randomness is probably essential [10], at least in so far as to ensure that the "result must not be precisely predictable" [11]. That is, some degree of surprise.

Hence our attempt to use neural networks to learn the user esthetic model. Neural local search methods require less training data than evolutionary global search, so the notion is to use the scarce data we have and train neural networks which can then act as the evaluation function for the evolutionary algorithm generating new images.

2.1 Experiment 1 – A Traditional Neural Network Data Set

In our first experiment, we use some labelled data to construct a training set of 100 patterns, being encodings of 50 images the user liked, and 50 images the user disliked. The test set of 44 patterns is similarly divided between liked/disliked image representations. This data set (training and test set) is based on an extremely long series of sessions viewing some 3,000 Mondrian-like images. This was impossible to get most users to commit to, hence this paper discusses the results on subject T only from our previous experiment.

Note that multiple subjects would only demonstrate that we can learn the esthetic preference for more than one person, as most of the inherent subjectivity would remain in the selection of liked/disliked images. Further, the elephant in the room would remain. The design of the experiment is not really independent of the measures used and subjects available. There is some tension here, which is usually expressed in the neural network community by the use of multiple test sets.

There is a training set to use to train a network, another set to decide when to stop training and another set ("which is not seen at all by the network") to actually calculate the usefulness of the network. The 2^{nd} and 3^{rd} sets are called test and validation sets, but some authors do train-validate-test while others do train-test-validate (the sequence we prefer). Arguably, the use of a 3^{rd} set is to prevent cheating.

A set of 1,000 Mondrian-like images was classified by subject T into like/unsure/dislike groups. This classification was used to calculate the mean square error as well as classification error. For neural network predictions in the range 0.4 images was classified by subject T into like/unsure/dislike groups. This classification was used to calculate the mean square error as well as classification error. For neural network predictions in the range $0.4 \leq p \leq 0.6$ where the label was 0.5, we record a correct classification (into the implicit 'unsure' category), while predictions above/below this are correct if the label is 1/0 respectively. The number of patters in each classification is as follows: yes 87, maybe 803, no 310.

2.2 Experiment 2 – Sequentially Trained Neural Networks

In experiment 1, we constructed a traditional type training and test set. In experiment 2, we follow a path more appropriate to the problem, following the sequence of presentation of results from a generation of Mondrian-like images to a user (again, for comparison, subject T), and so on.

The experiment was as follows. The user was presented with a sequence of 20 Mondrian-like images, from which the expectation was that 1 would be selected as liked, and one disliked. This sequence repeated in the first phase until 20 liked and 20 disliked images were collected. (For data set purposes, any episodes in which more or less liked/disliked images were collected are resorted so that it appears as if there is a guaranteed liked and disliked image in each cycle of presentations.)

For this phase all images are generated randomly. The 40 labelled images are used to construct the initial training set with 28 training and 12 test patterns, being alternating liked and disliked patterns. These sets are used to train neural networks (or varying topology described below), which are then used to select the next images to be presented in the next phase, with 10 of the images selected evenly distributed from the neural network results suggesting the user will like these, 2 dislike, and 8 at random. For reproducibility purposes, 1,000 images are constructed, and ranked by the trained neural network. Hence the "8 at random" represents about 1 in 4 of the remaining images once the liked and disliked are included.

For data set purposes, an entire phase of presentations of 20 images at a time is done before any further training, though otherwise we could use each extra image to retrain a neural net while the user is looking at the next set on screen. From the 20 liked and 20 disliked images collected, the training and test sets are increased in size, with 20 more training patterns and 20 more test patterns, again alternating liked and disliked patterns. A new network with the same topology is trained each time, with the previous learned weights ignored, to eliminate any extra effect of the initial set.

This overall process is repeated 4 times. Finally, the same 1,000 Mondrian-like images are used as the validation set as used in experiment 1.

3 Results

The results presented in this section are the validation results on a set of 1,000 Mondrian-like images (being their representation as described previously) which were fully classified by subject T into yes/maybe or unsure/no categories, which are represented by 1/0.5/0 values. These are used normally for mean squared error (MSE) calculations, and used to determine the number of correct classifications in §2.1.

Table 2. Performance of 3 trials, 2 neural network topologies for experiment 1, and one for 2

Expt 1, 28 x 5 x 1 net	Expt 1, 28 x 50 x 10 x 1 net	Expt 2, 28 x 5 x 1 net
TSS = 189.1	TSS = 190.3	TSS = 109.1
		START Train: 28 Test: 12
		THEN Train: 48 Test: 32
		THEN Train: 68 Test: 52
		THEN Train: 88 Test: 72
Train: 100 Test: 44	Train: 100 Test: 44	FINAL Train: 108 Test: 92

In experiment 1, two different initial topologies were used. The results were essentially identical, hence the simpler topology was used for experiment 2. Clearly, our experiment 2 performs significantly better than experiment 1. The number of training patterns is quite similar, though the test patterns are approximately double. While the number of test patterns is, we believe, of limited significance, we will test this in the future. In experiment 2, the correct classifications are 94%. This is a very high degree of prediction. At this point it is appropriate to remind the reader that there is significant subjectivity involved with the use of a single subject, and that being subject T, however there seems little reason to believe this would affect the success or otherwise of experiment 1 versus experiment 2.

Our view is that the difference in performance is due to much of the final data sets in Experiment 2 being in some sense related to and fine tuning the neural network representation of the user's esthetic preference, rather than 'randomly' adding to a training set in a very complex space. (Subject T has reported that during the extensive interaction with our Mondrian-like images, his perception of what he likes/dislikes has changed and evolved, and that there were images he liked "notwithstanding he should not like them"! We interpret this statement to mean that the images were ones that did not match his own view of the "kinds of images" he liked.) That is, if we consider the training set in experiment 1 being similar to experiment 2, then the training set in experiment 1 is like the initial training set of experiment 1 with 72 randomly chosen additions, while the training set of experiment 2 has the addition of 80 increasingly well chosen additions.

The difference between experiments 1 and 2 are very clearly illustrated in plotting the predicted values for patterns in the validation set.

The difference between prediction profiles is profound. Clearly the difference is significant, in that in experiment 1 the results are to produce classifications which are

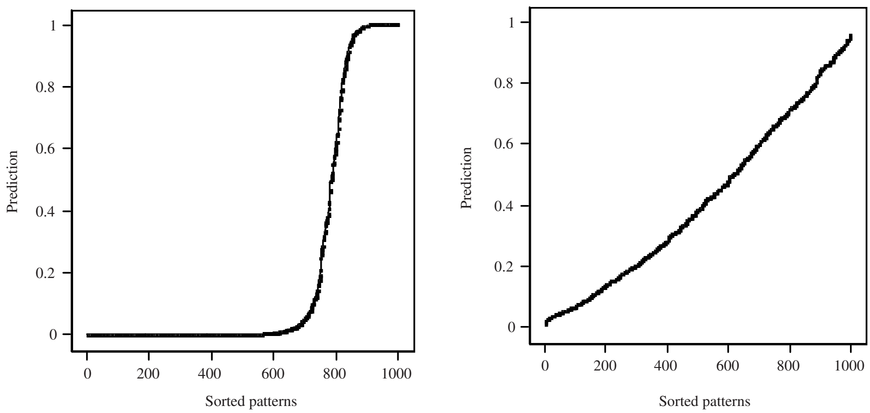

Fig. 4. Predicted output for each validation pattern, for subject T (experiment 1 on left, 2 right)

yes/no with a small transition, whereas in experiment 2 the results are a smooth transition from no via maybe/unsure to yes. The plots in the left and right of Figure 4 above are similar to what one would expect for a classification and a regression problem respectively. This suggests a possible explanation, in that the sequential, and contextual, construction of the training set has created (effectively) a regression style training set for a classification problem.

Significance of input parameters. We analysed the weight matrix (of experiment 2) to determine the significance of each of the input parameters [8]. The most significant were 19 9 24 22 17 15 20 25, in decreasing order of significance, representing approximately 10% down to 5% contribution each and overall accounting for over 50% of the contributions of the input to the hidden layer.

There are some surprises here. Parameter 19 is the proportion of coloured area is not surprising. Parameter 9 is the biggest vertical distance between lines. None of the rest of the related group (6 7 9 10) are similarly significant, with 7 (smallest horizontal distance between lines) having 1/2 the contribution while being double or more the contribution of the remaining parameters in that group. Parameter 24 is the Manhattan distance between 2 colours, and has double the contribution of Euclidean distance in parameter 23. Possibly this is due to the rectilinear nature of Mondrian's De Stijl compositions (the parameter was introduced to investigate whether different distance metrics make a difference, the result suggests further investigation).

Parameters 22, 20 and 17 were particularly surprising, being the distance between yellow and blue, distance between yellow and red, and proportion yellow. In none of our observations in this experiment or the prior work we reported [6] did we notice any relation or pattern in the esthetic choices of subject T which relate to yellow. This suggests that our technique is able to identify relationships which are not consciously available to users (or at least not available to subject T).

4 Conclusion

We have demonstrated that it is possible to use simple neural network models to learn the esthetic preferences of a subject when looking at Mondrian-like images generated by our computational process.

We found that the process of data set construction has a profound effect on the results achieved, with a hypothesis that an appropriate construction process can create a data set which approximates a regression problem even on a classification problem base. This would have huge significance as it would allow probabilistic or possibilistic conclusions to be made from categorical data.

We also found that users (or at least user T) are not completely aware of all of the significant features which can explain some of their choices.

References

1. Wikipedia, entries on de stijl, Piet Mondrian, `http://www.wikipedia.com`
2. Hill, A.: Art and Mathesis: Mondrian's Structures. Leonardo I, 233–234 (1968)
3. Reynolds, D.: Symbolist Aesthetics And Early Abstract Art, Cambridge UP, page 260 (1995)
4. McManus, I.C., Cheema, B., Stoker, J.: The aesthetics of composition: A study of Mondrian. Empirical Studies of the Arts 11(2), 83–94 (1993)
5. Wolach, A.H.: Line spacing in Mondrian paintings and computer-generated modifications. Journal of General Psychology (July 2005)
6. Gedeon, T.D., Shen, J.Y.: Making art using evolutionary algorithms and artificial AI. In: Proceedings BOOM 2007, p. 6 (2007)
7. Taylor, R.P.: Fractal expressionism-where art meets science, Art and Complexity. Elsevier Press, Amsterdam (2003)
8. Gedeon, T.D.: Data Mining of Inputs: Analysing Magnitude and Functional Measures. International Journal of Neural Systems 8(2), 209–218 (1997)
9. Brown, W.M., Gedeon, T.D., Groves, D.I.: Use of noise to augment training data to compensate for lack of deposit examples in training a neural network for mineral potential mapping. Natural Resources Research 12(2), 141–151 (2003)
10. Granger, M.J., Mazlack, L.J.: Representing Aesthetic Judgments. In: Proceedings of the International Conference on Cybernetics and Society, pp. 16–20 (1981)
11. Maiocchi, R.: Can you make a computer understand and produce art? AI & Society 5, 183–201 (1991)