Novelty Detection in Large Environments

Stephen Marsland, Ulrich Nehmzow and Jonathan Shapiro
Department of Computer Science
University of Manchester
Oxford Road
Manchester M13 9PL
E-mail: {smarsland, ulrich, jls}@cs.man.ac.uk

5.4.2001

Abstract

Novelty detection systems detect inputs that do not conform to an acquired model of normality. In general this is done by training a neural network on a training set that is known not to contain any examples of classes that should be considered novel. Then, inputs to the filter are compared to the model, and those that do not match are rejected. Typically, such systems are used when there are many examples of one class, but relatively few of an important class. This is true, for example, in many health applications where there are many more examples of healthy test results than unhealthy results.

This paper describes a novelty filter that can operate online, so that the filter can assess the novelty of perceptions as they appear with respect to the model learnt so far, and then add the current perception into the model. A new neural network is described that adds new nodes into the map space in an ordered manner in order to maintain the topology of the input space. This forms the basis for the novelty filter. The resulting novelty filter is applied to a robot inspection task. A mobile robot explores a corridor environment using its sonar sensors to perceive the environment, which consists of 300 m of corridor. The novelty filter created using the new network is compared to the same filter based on the Self-Organising Map. It is shown that the new network does not have difficulties adapting to large environments.

1 Introduction

Novelty detection is concerned with recognising features that deviate from an acquired model of normality. That is, a feature is considered to be novel if it differs in some significant way from previous perceptions. The ability to detect novelty is an important survival trait in animals (the unexpected could be a sign of a potential predator), and it is also very useful for artificial learning systems.

Artificial neural networks perform well on classification tasks, provided that the data that they are trained on contains examples of all the possible different classes that will be seen in operation. Unfortunately, this is not always possible. For test data that are markedly different from any elements of the training set the network will still give an output, but this output is unlikely to be accurate. In many real world datasets there are lots of examples of one class, but relatively few of another. This means that the network cannot be trained to reliably detect examples of the second class. Examples include health monitoring, where there are almost always more examples of healthy test results than unhealthy ones, and fault detection in machines.

A number of artificial neural networks that perform novelty detection have been proposed in the literature. In general, these require a training set that is known to contain no feature that should be considered novel. The network learns to classify inputs in the training set and then, in normal operation the network classifies each input as either normal (i.e., similar to those in the training set), or novel. This can be done using kernel density estimation, where inputs are only classified as normal if they are in a

Proc. TIMR 01 - Towards Intelligent Mobile Robots, Manchester 2001.
Technical Report Series, Department of Computer Science, Manchester University, ISSN 1361 - 6161.
Report number UMCS-01-4-1. http://www.cs.man.ac.uk/cstechrep/titles01.html

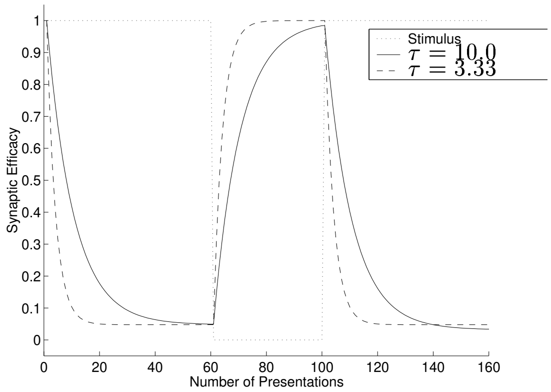

Figure 1: An example of how the synaptic efficacy drops when habituation occurs. In both curves, a constant stimulus $S(t) = 1$ is presented, causing the efficacy to fall. The stimulus is reduced to $S(t) = 0$ at time $t = 60$ where the graphs rise again, and becomes $S(t) = 1$ again at $t = 100$, causing another drop. The two curves show the effects of varying τ in equation 1. It can be seen that a smaller value of τ causes both the learning and forgetting to occur faster. The other variables were the same for both curves, $\alpha = 1.05$ and $y_0 = 1.0$.

part of the space that is covered by the training set, or by comparing the current input to the member of the training set that is most similar to it. Further details are given in section 4.

One problem with the methods described above is that the training has to be performed off-line. This means that the filter cannot classify a perception as novel and then learn to recognise that perception. Therefore, the training set has to be created very carefully, ensuring that it contains examples of all data that should be found to be normal, and no examples of novelty.

Novelty detection also enables a mobile robot to be used as an inspection agent. A robot that was trained to recognise all the features of some environment could then be used to explore other environments and highlight features of the second environment that were not in the first. This paper describes a novelty detection algorithm that is suitable for online use. The filter is tested on an inspection task, by using it to process the perceptions of a robot's sonar sensors as the robot explores an environment. The filter learns as the robot explores, and highlights features that are not in the model acquired by the robot during the exploration that has occurred so far. In this paper it is demonstrated that the implementation of the novelty filter proposed in previous work (Marsland et al., 2000b) does not scale well. An alternative is proposed, and it is demonstrated that the new novelty filter scales well.

2 The Novelty Filter

2.1 Habituation

Habituation is a reversible reduction in the behavioural response to a stimulus when it is presented repeatedly. It enables the animal to ignore stimuli that are seen often, so that it can concentrate on other, potentially more important, stimuli. Habituation is thought to be one of the fundamental mechanisms of adaptive behaviour, and can be seen in animals as diverse as the sea slug *Aplysia* (Bailey and Chen, 1983), cats (Thompson and Spencer, 1966), toads (Wang and Arbib, 1992) and humans (O'Keefe and Nadel, 1978). In contrast to other forms of behavioural decrement, such as fatigue, the response can be restored to its original level without the organism resting by introducing a change in the stimulus. An overview of the effects and causes of habituation can be found in Thompson and Spencer (1966).

There have been several attempts to model the effects of habituation computationally and to explain the interaction between habituation and dishabituation, the process whereby an habituated response returns to its original strength. Groves and Thompson (1970) suggested that dishabituation was an instance of sensitization and therefore an independent construct that interacts with habituation to produce a net response. Their model was used by Stanley (1976) to simulate habituation data from experiments of the spinal cord of a cat. He described the decrease of synaptic efficacy y by the first-order differential equation

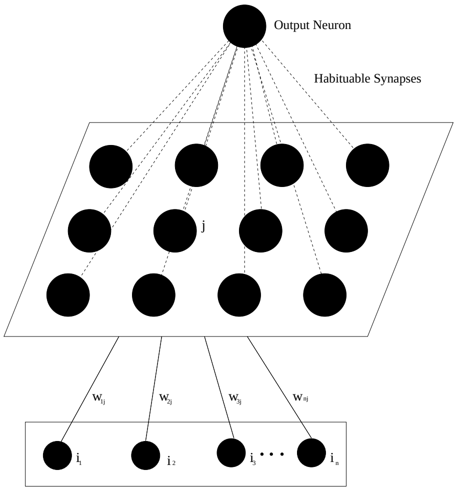

Figure 2: The novelty filter. The input layer connects to a clustering layer that represents the feature space. The winning neuron (i.e., the one 'closest' to the input) passes its output along a habituable synapse to the output neuron so that the output received from a neuron reduces with the number of times it fires.

$$\tau \frac{dy(t)}{dt} = \alpha \left[y_0 - y(t) \right] - S(t) \tag{1}$$

where y_0 is the initial value of y, $S(t)$ is the external stimulation and τ is a time constant governing the rate of habituation, while α controls the recovery rate. A graph showing the effects of this equation is given in figure 1. The values of the variables given in the caption of figure 1 are the ones that were used.

2.2 Using Habituation in a Novelty Filter

The effect of habituation is to filter out those perceptions that have been seen before, meaning that only novel stimuli are noticed. This is exactly the behaviour that a novelty filter should demonstrate. Before it is known whether a particular input is novel, it needs to be recognised. The novelty filter that is described in this paper uses a clustering network to perform this recognition. Each node of the clustering network is attached to an output neuron by a synapse that habituates with use (controlled by equation 1). The novelty filter is shown in figure 2.

The clustering network classifies the input, selecting the node whose weights best match the current input vector. The output of this winning node is propagated along the habituable synapse, which modifies the strength of the signal. If that node has not fired before, or fired only rarely, then the signal is passed on without attenuation, and the activity at the output neuron is high. If that node has fired often, meaning that the perception is not novel, then the efficacy of the synapse is low and so the output neuron receives a very weak signal. As well as the synapse connecting the winning node to the output habituating, the synapses of nodes that are neighbours of the winning node in the map space also habituate, although to a lesser extent.

This abstract representation of the novelty filter does not require that any particular clustering network be used. Previous work (Marsland et al., 2000a) tested several networks, in a set of experiments where a robot turned towards the most novel stimulus in its perceptual range. It was discovered that the Self-Organising Map of Kohonen (1982) was suitable for use on this task. This network is described in the next section. Then, in section 2.4, the new growing network is motivated and described.

2.3 The Habituating Self-Organising Map (HSOM)

The Self-Organising Map (Kohonen, 1982, 1993) is one of the most popular artificial neural networks in use. The network consists of a lattice of map nodes (neurons) that are each connected via adaptable weight vectors to the elements of an input vector. The SOM algorithm performs competitive learning, but instead of just the winning node being adapted, nodes that are close to the winner in the map space (neighbours) are also adapted, usually by a smaller amount.

This means that, over time, nodes that are close together in the lattice respond to similar inputs, so that the set of local neighbourhoods self-organise to produce a global ordering. The dimensionality of the map space and the number of nodes in the network are chosen in advance, typically the map is one- or two-dimensional.

For a given input, the distance between the input and each of the nodes in the map field is calculated, usually as the Euclidean distance

$$d_j = \sum_{i=0}^{N-1} (v_i(t) - w_{ij}(t))^2 , \tag{2}$$

where $v_i(t)$ is the input to node i at time t and w_{ij} is the the weight between input i and neuron j.

The node with the minimum d_j is selected as the winner, and the weight for that node and its neighbours in the map field are updated using:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)\,(v_i(t) - w_{ij}(t)) , \tag{3}$$

where j is the index of a node in the neighbourhood, and η is the learning rate ($0 \leq \eta(t) \leq 1$). To make the novelty filter from the SOM an habituation synapse with dynamics controlled by equation 1 was added, connecting each node in the map space to the output neuron.

2.4 The Grow When Required (GWR) network

When the SOM is used, the size of the network is defined in advance. If the network is trained in batch (i.e., off-line) then this is not a problem, as the network nodes spread out to represent the whole of the input space. However, when the network is used online, as is done in the experiments reported in this paper, there is no guarantee that this will happen. Instead, as the number of perceptions approaches the order of the number of nodes, the network will saturate, so that several different perceptions map onto the same node. The novelty filter habituates to the firing of a particular node, not directly to the perceptions that caused that node to fire. Therefore, if the filter has habituated to one perception and another causes the same node to fire, then the second perception will be ignored too. This is clearly not desirable.

The network that is described in this section is a proposed solution to the problem. The algorithm starts with only a small number of nodes, and adds new nodes when the activity of the best matching node is not sufficiently high, so that the match is poor. To allow for the fact that new nodes may not yet have been trained to match their intended input correctly, each node is equipped with a way of measuring how often the node has fired. This could be done using a simple counter, which is incremented whenever a node is the best-matching node. However, each node already has an habituating synapse attached, which measures how often that node has fired. This enables the growing process to be built into the novelty detection for which the GWR network was designed. It also means that as nodes are trained, which affects the weights of neighbouring nodes, so the habituating counter for the neighbouring nodes also decreases in value a little.

The basic idea behind the algorithm goes like this:
If a node matches an input well, then the activity of that node is close to 1. There are two reasons why the activity of the node could be low - either the node is still being positioned by the learning rule, having been added recently, or there is a mismatch, and that node is already representing another feature. If the node is a new one then it will not have fired often, so the habituation counter will be high. In that case the node should be trained to be a better match to the input.

So, if the habituation value is close to 0, then the node should be well placed in the map field and hence the activity should be high. If this is not the case then we need to introduce an extra node to match the input better. This node is added between the winning node, which caused the problem, and the input, with the weights of the new node being initialised to be the mean average of the weights for the best matching node and the input. If the node is well placed then a neighbourhood connection is put between the winning node and the second best matching node if it does not already exist, and all

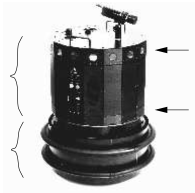

Figure 3: The Nomad 200 robot used in the experiments.

nodes in the neighbourhood of the winning node are updated using the same learning rate as for the SOM (equation 3).

Thus, two thresholds are needed to decide whether or not to insert a node on the current iteration: a minimum activity, a_T, below which the current node is not considered to be a sufficiently good match, and a maximum habituation value, above which the current node is not considered to have learnt sufficiently well. In practise, the value of the habituation threshold does not seem to matter very much, and is usually set to 5 presentations. The value of a_T does make a considerable difference. If the value is set very close to 1 then more nodes will be produced to make a better match with the inputs.

Further details of the GWR algorithm, together with examples of the network learning other datasets and comparisons with other networks will be given in a future paper.

3 Experimental Results

3.1 Description of the Experiments

Previous work (Marsland et al., 2000b) has investigated the novelty filter based on the HSOM on simple robot inspection tasks. In those experiments the robot made several trips along 10 m section of corridor, taking sonar scans as it travelled, and presenting an average of the scans over the last 10 cm of travel every 10 cm along the way, classifying the perceptions according to the model acquired so far before adding the current perception to the model. The Nomad 200 robot shown in figure 3 was used. The experiments showed that the novelty filter learnt an accurate model of the environment over two or three trips through an environment.

The experiments reported in this paper follow the same lines. The robot used a wall-following behaviour to travel through an environment, taking sonar scans as it travelled and produced an input vector by taking the average of these readings over the last 10 cm of travel, as in the previous experiments. The difference is that the environment in this case was a 300 m long circuit of corridor. The corridor was part of the 2nd floor of the Computer Science Department at the University of Manchester. It consists of a loop of corridor approximately two metres wide with breezeblock walls and wooden doors on both sides. There are several fire doors along the corridor. These were all kept open to allow the robot to travel along the corridors unimpeded. The environment was fairly static, the experiments were performed at a time when the building was mostly empty, although people did walk past the robot occasionally.

The aim of the experiments was to test how well the two networks described in the previous section, the HSOM and GWR networks, performed on this large environment. It was expected that a small HSOM network would saturate so that no perception would be considered to be novel. This was tested by making a significant change to the environment at the end of the last run, so that the novelty filter would detect these very last perceptions as being novel. This change was that a large number of cardboard boxes were placed around the robot so that it reached a dead end in the corridor.

While the robot explored, three different novelty filters were receiving the averaged sonar scans as input, and producing a value for the novelty of each perception. Two of the novelty filters were HSOM-based. The first had 49 nodes arranged in a 7×7 network, while the second used a 23×23 network of 529 nodes. The final network was a GWR network with an insertion threshold of 0.75, an habituation threshold of 0.1 and two nodes initially. The data was also stored so that other networks could be tested off-line.

The results of the experiments using the three networks described are shown in figures 4, 5 and 6. Each of the figures shows the activity of the output neuron for each perception of the robot. The spikes

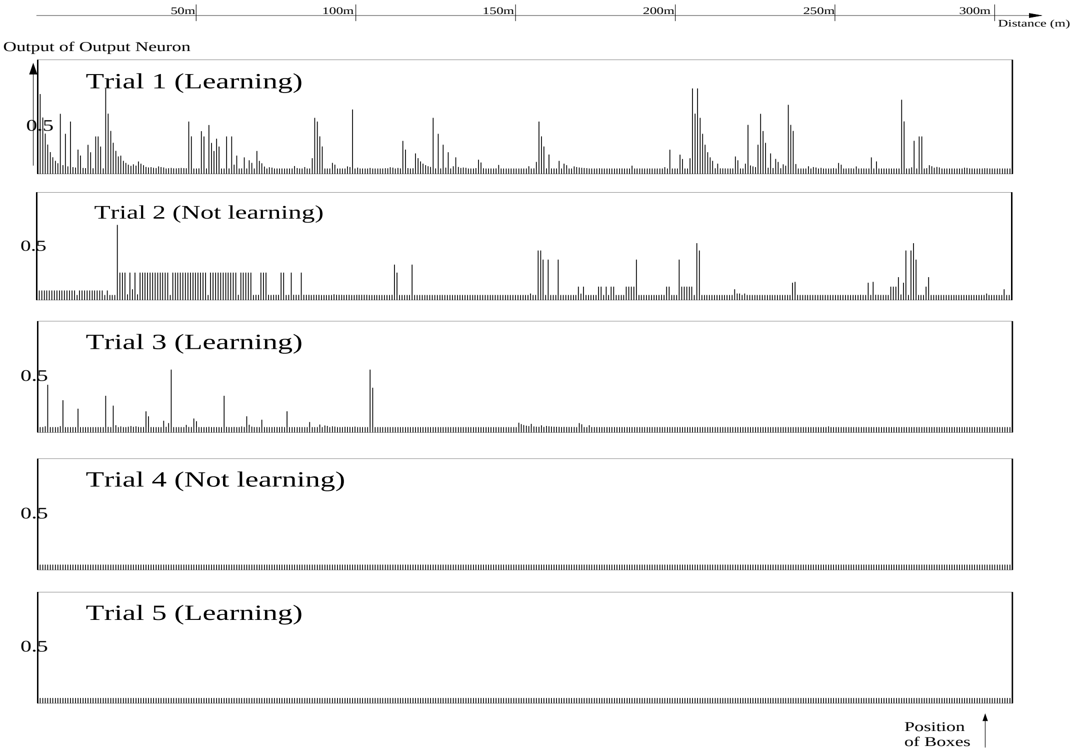

Figure 4: The output of the novelty filter using the small HSOM. The network saturates so that after partway through Pass 3 nothing is found to be novel.

show the maximum activity of the output neuron of the novelty filter over 1 m intervals (i.e., every 10 inputs). In all, five complete laps of the environment were performed. During the first run the novelty filters updated their model of the environment in addition to evaluating each perception. To test how well this learning was performed, during the second run the networks did not learn, but produced a novelty value for each perception. During the third run the learning was switched back on, but for the fourth it was again switched off. During the final run the network learnt again.

3.2 Results With the HSOM

3.2.1 A Small Network

The small HSOM comprised 49 neurons arranged in a 7×7 grid. A learning rate of $\eta = 0.25$ was used and was kept constant, so that the learning behaviour did not change. The neighbourhood size also remained unchanged at the 8 immediate neighbours of each neuron. The first two passes through the environment that are shown in figure 4 show the kind of results that are expected. The algorithm finds many of the perceptions to be novel, but also learns to recognise those parts of the environment that consist mostly of wall on both sides of the robot. However, on the third run, about halfway around the circuit, the novelty filter stops finding anything to be novel, and continues this way throughout the rest of that lap of the environment, and for the next two laps. Nor does the novelty filter find the perceptions of the boxes at the end of the last run to be novel. This is in contrast to the other networks, and demonstrates that the novelty filter based on a small SOM network can indeed saturate.

3.2.2 A Large Network

The large HSOM, which consisted of 529 neurons (23×23), also used a learning rate of $\eta = 0.25$ and a neighbourhood of the eight immediate neighbours. The results of this larger network are interesting in relation to the previous results for the small HSOM. It can be seen that the novelty filter does find the perceptions at the end of the environment particularly novel, which is good. However, even on the third

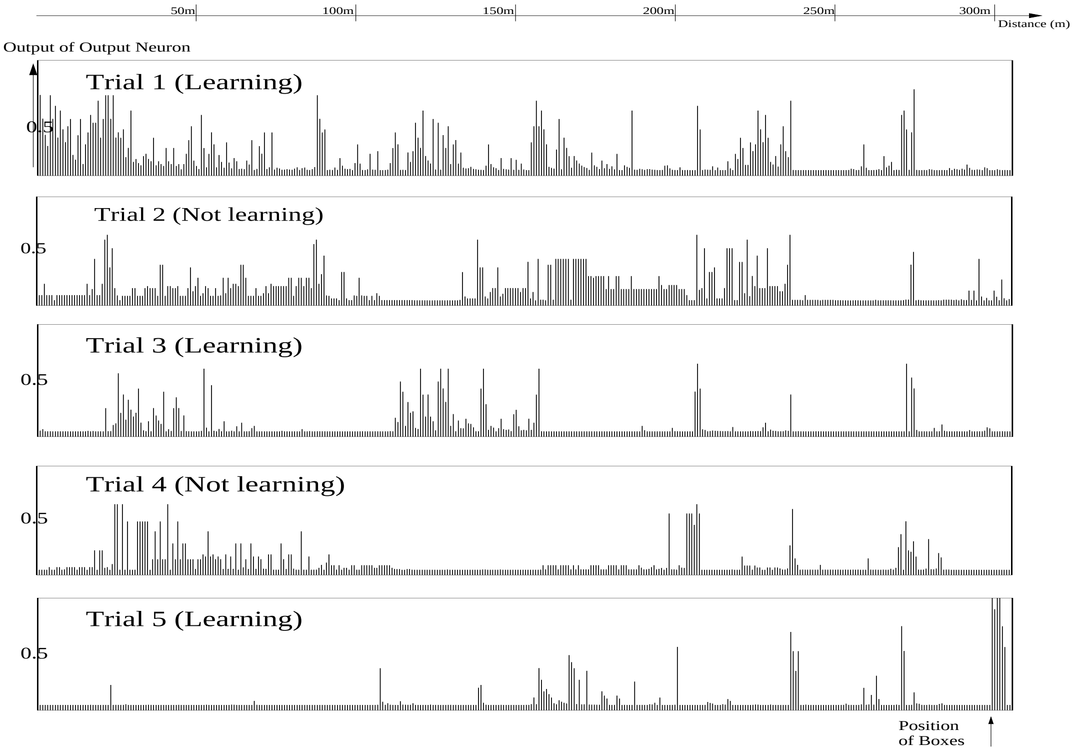

Figure 5: Using the large HSOM as the basis for the novelty filter means that the network does not saturate. However, the filter does not appear to generalise well.

training run around the environment the robot is still finding a fairly large number of perceptions novel. Many of these perceptions are where the robot is turning a corner. This is because the wall-following control program allows the robot to take a variety of paths around a corner, and so that same perceptions are not seen every time.

However, even taking this into consideration, the large HSOM network does seem to have some problems with noise in the sensors. This is partially to be expected since the network is so large, so that there is enough space in the network for each perception to be stored separately. This means that the network does not generalise. Using the data collected during the online experiments, a number of differently sized networks were tried. It was found that a 15×15 network performed well on this dataset, finding the boxes at the end of the run novel (i.e., not saturating), but not finding any other perceptions on the last run novel. Unfortunately there does not seem to be any way of predicting how large the network should be before the data is collected. This means that the filter cannot be used online as is intended.

3.3 Results With the GWR

Figure 6 shows the output of the novelty filter when the GWR network is used. The shape of the habituation curve can be seen clearly in Pass 1. These results are very promising - even in the second pass through the environment very little is found to be novel, and in the final pass only the novel features at the end of the run are highlighted. The data collected was used to experiment off-line with the insertion threshold, which was 0.75 for the results shown in figure 6. It was found that for values between 0.65 and 0.85 the results were very similar to those shown in the figure.

The number of nodes that the GWR network added during each of the learning runs is also given on the right of each graph. It can be seen that even at the end of the five passes through the environment the network had only 37 nodes, considerably less than even the smaller of the two HSOM networks. This is because the GWR network only adds nodes into those parts of the space from which inputs have been generated, while the HSOM attempts to cover the whole of the space. This means that the GWR network can generate considerably more parsimonious representations of the space, resulting in better

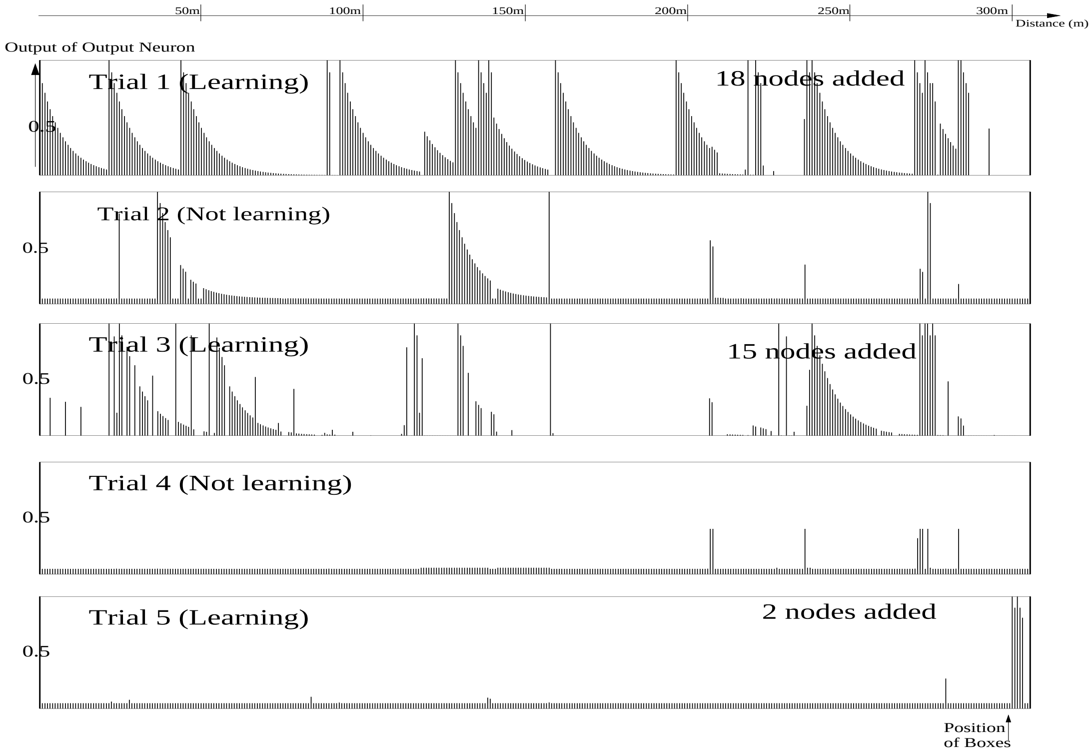

Figure 6: The output of the novelty filter using the GWR network. The filter learns quickly and correctly finds the changed perceptions at the end of Pass 5 to be novel.

tuned networks.

3.4 Discussion

The results of the previous sections have shown very clearly that the required preselection of the network size for the HSOM is a limitation for the purposes of the novelty filter described in this paper. However, by using the GWR network it has been shown that the filter can learn a reliable and accurate model of the environment and learn to detect deviations from that model.

The number of nodes that the GWR network uses to represent the robot's perceptions of the environment is considerably fewer than the HSOM. This has benefits for the storage of the network and also for the reliability of the method.

A number of other networks have also been used as the basis for the novelty filter and compared with the GWR network. These include the Growing Neural Gas (GNG), a growing network proposed by Fritzke (1995) and an unsupervised form of the Reduced Coulomb Energy (RCE) network (Reilly et al., 1982). The GNG network is very similar to the GWR network, except that the GNG adds new nodes to support the node in the network that currently has the highest error. This means that the network adds a new node only when it has had time to accumulate an estimate of the error at each node. This is not suitable for novelty detection, as any input that is novel that occurs when the network is not about to add a node will be ignored. Further details will be provided in a future paper.

4 Related Work

There have been a number of novelty detection techniques proposed in the literature. The first was Kohonen's Novelty Filter (Kohonen and Oja, 1976; Kohonen, 1993). This is an autoencoder network that is trained using back-propagation of error, so that the network extracts the principal components of the input. The network is trained on a dataset and, after training, any input presented to the network produces one of the learnt outputs, and the bitwise difference between input and output highlights novel components of the input.

Ypma and Duin (1997) proposed a novelty detection mechanism based on the self-organising map. They describe a number of measures by which the goodness of a SOM with respect to a particular dataset can be evaluated. In particular, they measure the average quantisation error over the dataset, and also measure how far away from each other map units that respond to similar inputs are. By training the SOM on data that is known to be normal, and then evaluating the measures on a new dataset, it can be seen whether or not the new dataset fits the same distribution as the data that generated the SOM.

Another approach using the SOM is to calculate the distance of the winning neuron from neighbourhoods that fired when training data known to be normal was introduced, and counting as novel those inputs where the distance is beyond a certain threshold. This method was used by Taylor and MacIntyre (1998) to detect faults when monitoring machines. The network was trained on data taken from machines operating normally, and data deviating from this pattern was taken as novel. This is a common technique when faced with a problem for which there is very little data in one class, relative to others. Examples include machine breakdowns (Nairac et al., 1999; Worden et al., 2000) and mammogram scans (Tarassenko et al., 1995). Often supervised techniques such as Gaussian Mixture Models or Parzen Windows are used, and the problem reduces to attempting to recognise when inputs do not belong to the distribution which generates the normal data (Bishop, 1994). This is the problem of kernel density estimation. The method proposed by Taylor and MacIntyre (1998) relies very strongly on the choice of threshold and on the properties of the data presented to the network, which must form strictly segmented neighbourhood clusters without much spread.

Growing networks such as Adaptive Resonance Theory (ART) (Carpenter and Grossberg, 1988) can be used to define as novel those things which have never been seen before, by using a new, uncommitted node to represent them. The use of habituation allows novelty to be defined more specifically as those things which have not been seen in the current context. The filter does this by learning an online, adaptive representation of the current environment. Habituation also allows the novelty of a stimulus to be evaluated, so that the novelty reduces with perception over time.

5 Summary and Conclusions

This paper has described a novelty filter that is capable of learning online. It has been demonstrated that the filter learns a model of an environment as perceived through a robot's sonar sensors, and detects deviations from the model.

The filter can be based on any clustering algorithm. This paper describes two networks that are suitable for use, the Self-Organising Map and a new network designed especially for the task. The new network is called the Grow When Required (GWR) network because of the way that it adds nodes into the map space whenever the current input is not sufficiently matched by the nodes already in the map.

The performance of the two networks are compared on a simple robot inspection task. A mobile robot explored a 300 m long loop of corridor, presenting its perceptions to the different novelty filters as it travelled. The novelty filters classified the perception with respect to the model learnt so far, and then added the current perception into the model. The SOM-based novelty filter requires that the number of nodes in the network is specified in advance. As it is not known how many perceptions will be seen by the network before it is used, this is a problem. We demonstrate that the SOM-based filter can saturate, so that every perception that is seen is considered to be normal, no matter what it is. This is never a problem for the GWR-based filter, which grows until the input space is well represented. This also means that the network that is produced is more parsimonious.

Future work will examine the behaviour of the GWR network in more detail, particularly the topology preservation of the input space and the quality of the mapping learnt. In addition, the inputs of different sensors such as a camera will be added to the novelty filter. This will demonstrate that the novelty filter is capable of online use as an inspection agent.

Acknowledgements

This work is supported by an EPSRC studentship.

References

Craig Bailey and M.C. Chen. Morphological basis of long-term habituation and sensitization in *aplysia*. *Science*, 220:91–93, 1983.

Christopher M. Bishop. Novelty detection and neural network validation. *IEEE Proceedings on Vision, Image and Signal Processing*, 141(4):217–222, 1994.

Gail A. Carpenter and Stephen Grossberg. The ART of adaptive pattern recognition by a self–organising neural network. *IEEE Computer*, 21:77 – 88, 1988.

Bernd Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7 (NIPS'94)*, pages 625–632, Cambridge, MA, 1995. MIT Press.

P.M. Groves and R.F. Thompson. Habituation: A dual-process theory. *Psychological Review*, 77(5):419–450, 1970.

Teuvo Kohonen. Self-organised formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

Teuvo Kohonen. *Self-Organization and Associative Memory, 3rd ed.* Springer, Berlin, 1993.

Teuvo Kohonen and E. Oja. Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements. *Biological Cybernetics*, 25:85–95, 1976.

Stephen Marsland, Ulrich Nehmzow, and Jonathan Shapiro. Novelty detection for robot neotaxis. In *Proceedings of the 2nd International Symposium on Neural Computation (NC'2000)*, pages 554 – 559, 2000a.

Stephen Marsland, Ulrich Nehmzow, and Jonathan Shapiro. Novelty detection on a mobile robot using habituation. In *From Animals to Animats: Proceedings of the 6th International Conference on Simulation of Adaptive Behaviour (SAB'00)*, pages 189 – 198. MIT Press, 2000b.

Alexandre Nairac, Neil Townsend, Roy Carr, Steve King, Peter Cowley, and Lionel Tarassenko. A system for the analysis of jet system vibration data. *Integrated Computer-Aided Engineering*, 6(1):53 – 65, 1999.

John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map.* Oxford University Press, Oxford, England, 1978.

Douglas L. Reilly, Leon N. Cooper, and Charles Erlbaum. A neural model for category learning. *Biological Cybernetics*, 45:35 – 41, 1982.

James C. Stanley. Computer simulation of a model of habituation. *Nature*, 261:146–148, 1976.

L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady. Novelty detection for the identification of masses in mammograms. In *Proceedings of the 4th IEE International Conference on Artificial Neural Networks (ICANN'95)*, pages 442 – 447, 1995.

Odin Taylor and John MacIntyre. Adaptive local fusion systems for novelty detection and diagnostics in condition monitoring. In *SPIE International Symposium on Aerospace/Defense Sensing*, 1998.

R.F. Thompson and W.A. Spencer. Habituation: A model phenomenon for the study of neuronal substrates of behaviour. *Psychological Review*, 73(1):16–43, 1966.

DeLiang Wang and Michael A. Arbib. Modelling the dishabituation hierarchy: The role of the primordial hippocampus. *Biological Cybernetics*, 76:535–544, 1992.

K. Worden, S.G. Pierce, G. Manson, W.R. Philp, W.J. Staszewski, and B. Culshaw. Detection of defects in composite plates using lamp waves and novelty detection. *International Journal of Systems Science*, 31(11): 1397 – 1409, 2000.

Alexander Ypma and Robert P. Duin. Novelty detection using self-organizing maps. In *Proceedings of International Conference on Neural Information Processing and Intelligent Information Systems (ICONIP'97)*, pages 1322 – 1325, 1997.