A Synapse Plasticity Model for Conceptual Drift Problems

Rudolph L Mappus IV (cmappus@cc.gatech.edu)
Ashwin Ram (ashwin@cc.gatech.edu)
College of Computing, Georgia Institute of Technology
Atlanta, GA 30332 USA

Abstract

Traditional supervised learning techniques do not address online learning problems such as concept drift, due to the fact that learning is offline when using these methods. Associative neural networks using Hebbian learning rules show robust performance in classification tasks involving concept drift. Biologically plausible neural networks represent a set of computational models designed to be more strongly related to biological neuron models. In this paper, we apply a biologically inspired plasticity model of synapse dynamics to a concept drift classification problem. The motivation for this method is to provide more biologically plausible networks in cognitive tasks.

Introduction

Concept drift is a real world problem associated with concept learning and classification: over time or as an agent gains more information about a concept and related concepts, the parameters of the features defining the concept change. Traditional supervised learners do not address the issue. Learning in problem domains with static concepts also avoids the problem, but it is a problem in dynamic learning problems and agents.

Intuitively, neural networks seem to have properties that would be helpful in concept drift problems. Particularly, associative neural networks seem appropriate to the problem. Traditional artificial neural networks however were designed as computational models first based on observed neuron behavior second.

Previous Work

Computational modeling of neuron processes has received a wealth of research (O'Reilly and Munakata, 2000). Some well researched models have been applied to learning problems such as concept drift. Biehl and Schwarze (Biehl and Schwarze, 1993) demonstrate a Hebbian learning model for handling random as well as correlated concept drift. Widmer and Kubat (Widmer and Kubat, 1996) show how latent variables can influence concept drift in the form of context. In this paper, we show results of timing dependent synapses, whose potentiation is governed by firing rate patterns of pre- and post-synaptic neurons. Schlimmer and Granger (Schlimmer and Granger, 1986) show how learning incrementally from noisy data can still generate classification results.

Consider the real world problem of classifying the working state of a power plant component (*e.g.* a turbine generator), given a set of sensor array readings for the component such as the RPM of component shafts or bearing temperatures. The goal of the learner in this domain is to classify the working state of the component in real time, given a stream of samples from the sensor array. Suppose that during the daytime hours, bearing temperatures rise from ambient heat, to levels that would, under other conditions, raise warning. In this case, the previous values used to classify nominal operation do not reflect the new operating parameters. A supervised learner that is not modified after training requires retraining on the new instances. We are interested in applying biologically plausible computational models to this problem.

Concept Drift

For the purpose of this discussion, we define conceptual drift to be the change in distribution parameters that define a classification. For example, suppose that the probability density of the an instance belonging to a classification was defined as Gaussian. The parameters defining that distribution would change over time, indicating a conceptual drift. This drift can be correlated, so that concepts at time t are related to concepts at time $t + n$. Clearly, even in a binary decision task, if drift occurs, the decision boundary generated for the concepts at t would not be the decision boundary generated at $t + n$.

Concept drift can increase without repetition (*i.e.* never returning to a previous distribution). Essentially, a concept's change is continual, and while the rate of change can slow, the change always leads to a new parameterization of the concept probability distribution in feature space. A good example of this might be bearing temperature whose value distribution permanently changes after seating.

Repeating and cyclic concept patterns There are some special cases worth mentioning. Concept drift can repeat. Suppose that a concept, after initial training, is explained by a parameterization b_1. In online learning, more instances arrive, inducing parameterization b_2. Still more instances arrive, driving the concept parameterization back to b_1. The learner, having learned parameterization b_1 would do best to take advantage of the previously learned concept. Concept drift can also be cyclic, where a concept repeats a set of parameterizations in an order, and this ordering repeats. In general, these special cases point to the fact that a latent variable affects the concept learning. All of these cases seem to require some memory, and the need for dynamic weighting of features.

Cognitive Motivation

There is growing evidence that synaptic behavior plays a fundamental role in many cognitive tasks. Hippocampal neurons responsible for working spatial memory seem to be highly plastic, relative to other cortical neurons. The implication is that a high level of plasticity is needed to handle dynamic spatial environments such as is needed to support navigation. Performance decay of learned processes that are not practiced seems to be attributable to weakening (depressed) potentiation between connections salient to the neuron processes related to the task. This point is noticed acutely again in spatial tasks, typically attributed to hippocampal areas of the brain. Aspects of motion detection also seem to apply to synaptic plasticity in the visual pathways. There has been some behavioral evidence that certain forms of dyslexia may be attributable to synapse plasticity in magnocellular visual pathway (layers I and II of the lateral geniculate nucleus) (Stein, 2001).

The human capacity to handle concept drift seems innate to many cognitive processes. The problem of classification in concept drift situations seems to share some of the dynamic demands required of plastic neural areas in the brain. We believe the use of stronger biological models in neural networks provides insight into cognitive computational phenomena and cognitive function.

Synaptic Plasticity

Synaptic plasticity has been used to refer to the relative tendency for a neuron (dendrite) to form new synapses as well as the potentiation of a synapse to propagate action potentials. In this paper we refer to plasticity as the latter. In Hebbian models of learning, the timing of action potential arrivals at the pre- and post- synaptic neurons governs the potentiation or depression of the synapse.

Computational model of plasticity

What properties of synapse plasticity are salient to a computational model? First, conductance of action potentials varies from region to region in the brain. At the neuron level, electrical properties vary from dendrite to dendrite, creating latencies in action potential propagation. However, the processes governing potentiation of synapses should be constant for all synapses. In other words, a constant factor is used to potentiate synapses when necessary. The same is true for annealing synapse potential due to a lack of arriving action potentials (*i.e.* use it or lose it policy). Synapse potentials are bounded in their values, where the minimum value represents no action potential propagation, and maximum potential means any arriving action potential is propagated.

Both pre- and post-synaptic neurons process action potentials S where $\{s_{pre}, s_{post}\} \in S$ (figure 1). Action potentials arrive at each synapse, creating a difference between pairs of action potentials $d = t_{s_{pre}} - t_{s_{post}}$. Potentiation is updated according to a window of positive difference (*i.e.* a pre-synaptic action potential precedes a post-synaptic action potential). A synapse potential is depressed if the difference is negative.

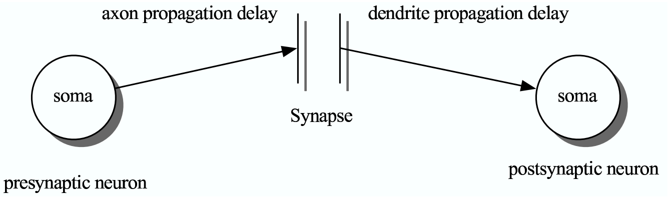


Figure 1: Plastic synapse neuron model (pre- and post-synaptic neurons).

Consider an integrate and fire neuron model, where the instantaneous sum of activation arriving from the dendrite is what is considered in action potential propagation. This decision function is modeled using a logistic transfer function $1/1 + e^{-u}$. In this case, a neuron soma represents the decision point for action potential propagation. Propagation of signals can be expressed in terms of the synapses and their relation to the soma. For each synapse, the distance from the presynaptic soma to the synapse represents the time for an action potential to propagate from the presynaptic soma to the synapse. Likewise, the latency from the synapse to the postsynaptic neuron soma represents the time for an action potential to propagate from the synapse to the soma as well as the variability in the conductance properties of individual connections. Here, we make the assumption that the conductance properties of the network are uniform. Suppose a fully connected network (i.e. a synapse exists between each neuron, but not recursive), where the latencies between synapses are chosen from a uniform random distribution. A synapse exists at each connection, and has a potential value. The potentiation of each synapse is initialized using a uniform random distribution. Neurons are organized in a laminar fashion, as in traditional networks, but latencies between neurons varies, so action potential propagation is not laminar. Inputs are delivered to the input layer of neurons at the soma, and arrive simultaneously. We define the resulting network by the set of synapse potential values, the connection latencies (weights), the synapse update function, using an integrate and fire neuron model.

Action potentials are propagated through the network at each timestep. The sigmoid decision function is used to propagate action potentials to the axon (i.e. integrate and fire). Each synapse's potentiation is updated, based on the pre- and postsynapse neurons' firing results. Qualitatively, neurons that fire in order (pre then post), increase potentiation, and lower the threshold for a signal to propagate at the synapse. Neurons that fire out of order depress the synapse potential, increasing the threshold signal required to propagate a signal. Synapses with potential 0 are not able to propagate signals. Synapses with max potential are able to propagate any arriving signal (signal with amplitude $a > 0$).

Network output

The network is allowed to propagate action potentials until all action potentials have propagated in the network (reached output neurons or were filtered). At each timestep, the values of the output nodes are noted, and once complete, the resulting "spike train" is analyzed for output. Biological networks of neurons output spike trains to indicate output from the network. In the simulated case, the spike train is analyzed for significant action potentials output within a window of time to determine positive output.

We define a positive window of activity in terms of the output population over all of the output windows. If any window shows activity above the mean activity, then we attribute a positive output by the network on that instance. We define a threshold decision rule for determining output activity of the network. If more than a threshold value of spikes are detected in the time window, then the network output is considered positive. The most sensitive decision rule would react to one action potential of minimal amplitude.

A network composed of these neurons behaves similarly to traditional networks: instances generate an activation pattern at the input nodes. Signal propagation through the network is governed by the potentiation of synapses and a propagation decision function at each neuron. The decision function is meant to model the fact that threshold signal must reach the soma before the neuron propagates an action potential (integrate and fire model). Training the network consists of passing the network positive instances of classification, so that common synapses salient to the task will potentiate as a result of the training input.

Processing action potentials in the network occurs as a discrete simulation of a neuron model, such that action potentials propagate through the network at each timestep. The maximum time for an input action potential to produce output is the maximum latency of a signal from input to output. In this case we assume action potentials propagate at a constant rate over the network (i.e. this does not model the chemical properties of action potential propagation).

Synapse potentiation A synapse potentiates (*lowers* the potential to propagate a signal across the synapse) based on firing pattern and rate of the pre- and post-synaptic neurons. Qualitatively, if the pre- and post-synaptic neurons fire sequentially, the affected synapse potentiates. If they fire out of sequence, the synapse is depressed. Synapse potential anneals from lack of activity as well. The potentiation window (Hebbian window) is modeled in the plastic synapse network using a quadratic equation.

Laminar networks Traditional neural networks are laminar in structure and processing. Nodes are assigned to layers: typically input, hidden and output and activation is computed per layer, for each node in a layer. The propagation of action potentials for use in tuning synapse potentiation is asynchronous with respect to the set of action potentials propagating in the network. While network components may be organized in a laminar fashion, action potential propagation is not.

Biological plausibility

An important aspect of the model is that it is biologically plausible in the sense that potentiation is a key feature of biological synapse dynamics. In a set of experiments meant to demonstrate the ability of TD learning to model the potentiation behavior of spike timing dependent plastic synapses, (Rao and Sejnowski, 2000) employ a neuron model meant to capture many biological behaviors. The experimental results support the use of a TD(0) function to capture the Hebbian activation window for a synapse. The attraction of using TD(0) is that it is computationally tractable. Shon et al. (Shon et al., 2004) employ a similar neuron model to model motion detection and prediction as observed in the visual cortex. Both experiments point out that it seems plausible the brain is adaptive for and reacting to statistical properties of the natural world.

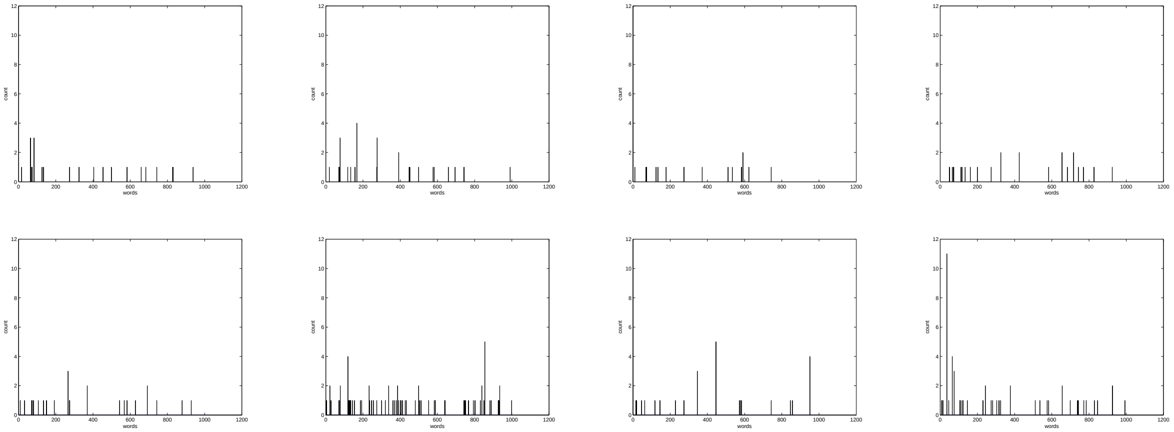
Another reason why this is a more biologically plausible network than a traditional network is it does not use a backpropagation function at the neuron level as the direct means to propagate error through the network. Backpropagating signals do so at the synapse level, and do not propagate beyond the synapse (i.e. action potentials do not flow backwards through the network). Recurrent connections, rather, are extensive in many areas of the brain, and seem to represent an important method that neurons use to receive feedback. The plastic model is more similar to this model of feedback updating than traditional networks.

Experiments

The biological and cognitive motivations for the model leads to a testable hypothesis: plasticity in terms of potentiation and Hebbian learning seems to be behind (at least short term) associative memory. If this is the case, then in concept drift problems where the drift is not random, but correlated to the original concept, then a plastic network (a network whose synapses are sensitive to activation patterns in the data) should perform better in classification accuracy than traditional networks.

Usenet Corpus

A Usenet corpus was used to test the network on real world data. The corpus consisted of samples of messages from twenty newsgroups (see table 1) ordered chronologically. A bag of words generator was used to represent each message (McCallum, 1996), and each document was normalized using the L_2 vector norm. A collection of 1000 words representing each message was used for the experiments. The vocabulary for the training set was ranked using infogain, and the top 1000 terms were selected, common to both training and test sets. Messages were classified by the news group it belonged to. The experimental networks were trained using a training set of instances, predetermined in the corpus body, and occurring chronologically before for test set. This arrangement of messages creates a distinct "drift" in the discrete word



comp.graphics *misc.forsale* *rec.autos* *sci.crypt*

Figure 2: Sample discrete category distributions for training (top) and testing (bottom) sets showing qualitative differences in distributions (drift). Word set is ranked infogain vocabulary of training set (common in testing set).

distributions of training and test sets for each group. A sample of group distributions for training and test sets can been seen in figure 2.

Neural network model

In order to facilitate discussion, we compare the results of the plastic synapse network to performance of a traditional artificial neural network (ANN) in the same task. The network was configured to the same structure as the plastic synapse network (n input nodes, n hidden nodes, 1 output node, learning rate $\alpha = 0.001$) using the Joone neural network toolkit (Marrone, 2005). The network was fully connected, composed of sigmoid nodes in the input and hidden layers. Output ω of the network was positive where $\omega >= 0.5$ and negative otherwise.

The network learning rate was tuned, using a range of values from 0.001 to 0.1. Over this range, the number of nodes in the hidden layer was tuned from its initial value n up to $n + 200$. The number of epochs was tuned using discrete values (100,1000,10000) to 100. In each of these cases for each of the Usenet groups, there was not a significant difference in performance between parameterizations. Two training strategies were used: train using the entire dataset, and train using only positive group examples.

Plastic Neural Network

The plastic neural network was initialized using the same neuron layer structure as the ANN, so that the number of neurons in both cases was the same. The TD window was set to unit time (maximum sensitivity to firing sequence); the synapse potential rate was initialized to 0.0001.

Action potentials were modeled by signal amplitude (strength of signal using the normalized input vector), and their arrival time at the soma or the synapse. Computation of action potential propagation produces a qualitative rule: presynaptic action potentials that produce propagated signals at a synapse cause synapse potentiation that lower the threshold amplitude of future arriving presynaptic action potentials. In these experiments, the networks were initialized with a fully connected hidden layer which fed to a set of output nodes. The networks were trained using only positive instances, in order to potentiate the appropriate synapses. During testing, synapse potentials were not changed in response to input.

Learning

The task in all of the experiments was a binary classification task. Instances were labeled according to the group being tested (members of the group in the training set were assigned a positive label, nonmembers a negative one). The task is one that is real in terms of concept drift: over time, the features describing documents that belong to a particular group change in parametrization. The task for the classifier is decide whether a new document belongs to the goal concept or not.

Results

Table 1 shows the training and test accuracy for the traditional and plastic synapse networks in each category of the Usenet corpus. The task was binary classification, where for each category, the category messages were positive instances, and all other instances were negative. Positive instance set sizes for training and test sets are shown for each category.

Plastic synapse neural network

The plastic synapse neural network was trained using only positive examples. The network used was feedforward, with recurrent connections (no reflective connections). A hidden layer was used that was fully connected to both the input and output nodes. While training accuracy was relatively high for all categories, test accuracy was insignificant, indicating overfitting and little

Table 1: Usenet corpus results. Size are the positive instance train and test set sizes, respectively. PSN training is the training accuracy of the plastic synapse network (* for all categories test accuracy was insignificant). ANN Positive is test set accuracy of the traditional neural network, using a training set of all positive instances. ANN Complete is the test set accuracy using the complete training set (positive and negative instances).

Newsgroup	Train	Test	PSN Training*	ANN Positive	ANN complete
alt.atheism	480	319	0.9962	0.4280	0.9747
comp.graphics	584	389	0.9537	0.4335	0.9692
comp.os.ms-windows.misc	591	394	0.9685	0.4339	0.9688
comp.sys.ibm.pc.hardware	590	392	0.9721	0.4339	0.9689
comp.sys.mac.hardware	578	385	0.9487	0.4332	0.9695
comp.windows.x	593	395	0.9518	0.4340	0.9687
misc.forsale	585	390	0.9647	0.4336	0.9691
rec.autos	594	396	0.9684	0.4340	0.9686
rec.motorcycles	598	398	0.9767	0.4342	0.9684
rec.sport.baseball	597	397	0.9973	0.4342	0.9685
rec.sport.hockey	600	399	0.9839	0.4343	0.9683
sci.crypt	595	280	0.9820	0.4341	0.9686
sci.electronics	591	393	0.9931	0.4339	0.9688
sci.med	594	396	0.9645	0.4340	0.9686
sci.space	593	394	0.9834	0.4340	0.9687
soc.religion.christian	599	398	0.9978	0.4343	0.9684
talk.politics.guns	546	364	0.9409	0.4315	0.9712
talk.politics.mideast	564	376	0.9763	0.4324	0.9702
talk.politics.misc	465	310	0.9713	0.4272	0.9754
talk.religion.misc	377	251	0.9566	0.4226	0.9801

generalization power in the classifier. As the distributions in figure 2 demonstrate, there does not seem to be significant overlap in terms between the training and test sets in many of the categories (see table 2). Qualitatively, less overlap produces fewer action potentials that are propagated in the network for positive instances, leading to outputs that are not detectable above negative instance input. Of course, allowing for online learning, even in the case where there is little or no feature overlap, produces high accuracy.

Traditional neural network

The traditional neural network did not appear to learn the binary classification task for any of the newsgroups when trained with either all positive instances or the complete dataset. High accuracy in the complete training set cases was due to the fact that the network performed as a majority classifier (100% false negative rate). To provide some insight into training the plastic synapse network, a traditional network was also trained using only positive instances. When trained with only positive examples, the network generated low classification accuracy, with similar responses for all document categories $SSD < 0.001$.

Conclusions

We have introduced a model biologically plausible model of synapse behavior and applied the model to a real world learning problem. In situations where real world learning problems display properties that mimic neuron processes, a neural network approach seems applicable.

The results indicate that potentiated synapses in neural networks seem to be able to classify a concept in a drift condition, but with little generalization power. In the constructive learning task, when training the learner with positive examples, the learner can only indicate how strongly a query instance matches (activates) the trained concept. We believe this explains the generalization performance of the network. One possible way to address this problem is to use the trained networks together to generate a classification. In the past, committee methods (e.g. bagging or boosting) have been used to increase performance of weak classifiers. In terms of the neural structure, an output neuron would be connected to the output of each individually trained network, and would serve as the voting mechanism (in the case of bagging) governing the output of the network. The implications for concept generalization would be that the classifier would be conservative in its estimates of new query points: only instances with small distance from the cluster would get a strong output from a particular network.

The traditional neural network solution is backpropagation of error through the network. Clearly, this method is less biologically plausible than an integrative approach. An important outcome of this work would be a biologically plausible model that can be applied to problem domains similar to the cognitive functions that mechanisms like synaptic plasticity support. The initial

Table 2: Category inner product angles between training and test sets for each Usenet category.

Newsgroup	Angle(degrees)
alt.atheism	84.18
comp.graphics	80.13
comp.os.ms-windows.misc	72.54
comp.sys.ibm.pc.hardware	85.71
comp.sys.mac.hardware	86.00
comp.windows.x	79.55
misc.forsale	83.06
rec.autos	84.93
rec.motorcycles	82.11
rec.sport.baseball	82.03
rec.sport.hockey	76.87
sci.crypt	84.10
sci.electronics	79.76
sci.med	74.01
sci.space	73.83
soc.religion.christian	65.21
talk.politics.guns	85.76
talk.politics.mideast	86.87
talk.politics.misc	81.89
talk.religion.misc	75.41

results shown here are encouraging for further experimentation. We want to emphasize this is not a cognitive model for concept drift, but rather through observing some likenesses of neural properties to the problem domain, we are generating a mechanism to handle concept drift. Using a model of synapse behavior in conjunction with a nonlaminar method of action potential propagation, the timing of action potentials (not rate coded) determines the potentiation of a synapse. Action potentials begin at the input nodes of the network, and propagate based on the latency of propagation in the axons and dendrites of the modeled network.

Future Work

To make a stronger claim about the importance of plasticity in dynamic learning problems, a well formed analysis is needed of the network behavior. A formal analysis of the network should provide better insight into the types of learning problems for which it is best suited. Incorporating backpropagation to the network would improve learning. While this seems clear, the solution is not straightforward and less biologically plausible. In a network where signals are time dependent, the instantaneous update based on error used in backpropagation is not easily addressed. Finally, generating neural structures that produce ensemble method behavior can improve classifier performance.

Acknowledgments

We would like to thank anonymous reviewers for helpful suggestions and comments.

References

Biehl, M. and Schwarze, H. (1993). Learning drifting concepts with neural networks. *Journal of Physics A: Math. Gen*, 26:2651–2665.

Marrone, P. (2005). Joone: Java object oriented neural engine the complete guide. http://www.joone.org.

McCallum, A. K. (1996). Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. http://www.cs.cmu.edu/ mccallum/bow.

O'Reilly, R. and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience*. MIT Press.

Rao, R. P. and Sejnowski, T. (2000). Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Computation*, 13:2221–2237.

Schlimmer, J. and Granger, R. (1986). Incremental learning from noisy data. *Machine Learning*, 1:317–354.

Shon, A. P., Rao, R. P., and Sejnowski, T. (2004). Motion detection and prediction through spike-timing dependent plasticity. *Network: Computational Neural Systems*, 15:179–198.

Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia*, 7(1):12–36.

Widmer, G. and Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1):69–101.