This is an author produced version of Using Epigenetic Networks for the Analysis of Movement Associated with Levodopa Therapy for Parkinson's Disease.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/104520/

**Article:**
Turner, Alexander Phillip, Lones, Michael Adam, Trefzer, Martin Albrecht orcid.org/0000-0002-6196-6832 et al. (5 more authors) (2016) Using Epigenetic Networks for the Analysis of Movement Associated with Levodopa Therapy for Parkinson's Disease. Biosystems. ISSN 0303-2647

https://doi.org/10.1016/j.biosystems.2016.05.005
Using epigenetic networks for the analysis of movement associated with levodopa therapy for Parkinson’s disease

Alexander P. Turner\textsuperscript{a,}\textsuperscript{*}, Michael A. Lones\textsuperscript{b}, Martin A. Trefzer\textsuperscript{a}, Stephen L. Smith\textsuperscript{a}, Stuart Jamieson\textsuperscript{c}, Jane E. Alty\textsuperscript{c}, Jeremy Cosgrove\textsuperscript{c}, Andy M. Tyrrell\textsuperscript{a}

\textsuperscript{a} Department of Electronics, University of York, Heslington, York YO10 5DD, UK
\textsuperscript{b} School of Mathematical and Computer Sciences, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, UK
\textsuperscript{c} Leeds Teaching Hospitals NHS Trust, Leeds, West Yorkshire LS1 3EX, UK

\section{1. Introduction}

Since the beginnings of computer science, there have been many successful attempts at capturing biological models within a computational framework. This field, known as bio-inspired computation, has given rise to algorithms which have been shown to out-perform humans on real world tasks, for instance in object recognition (He et al., 2015). Indeed, many of the computational algorithms that are now state of the art are built upon bio-inspired principles (Oquab et al., 2014; Hinton et al., 2012; Yang et al., 2013; Zhou et al., 2011). The design of bio-inspired algorithms falls on a spectrum. At one end of this spectrum are attempts to capture as much detail from the biological system as possible, to best promote the possibility of emergent behaviour. At the other are intentionally minimalist approaches, using pared-back models that capture only an abstract representation of the biological system. Each of these approaches has advantages, but often the simpler models are functionally complex and capable of capturing real-world biological dynamics (Bull, 2013; Wang et al., 2012).

In this work we use a novel bio-inspired architecture, termed epiNet, which is modelled upon the interactions between genetic and epigenetic processes within biological cells (Turner et al., 2015). Computational modelling of genetic processes, in the form of artificial gene regulatory networks, is nothing new (Lones, 2016). However, the inclusion of epigenetic elements allows for a connectionist architecture with a dynamical topological morphology – that is, the ability for a computational network to autonomously partition itself and select partitions based upon environmental or internal state. This serves two primary advantages. First, it supports task specialisation, where a partition assigned to a specific environmental or internal context is only used in that scenario. Secondly, with minimal analysis it provides a method of characterising a network’s behaviour from the ground up, by mapping the functionality of the individual partitions and building up an image of the system’s transitions between these partitions. This helps automate the process of model validation. Because of this, we propose that the properties of epiNet lend themselves well to the classification and analysis of real world time series data, where the underlying generative model is often poorly understood.

We describe the application of epiNet to the particular problem of understanding and classifying the movements associated with the neurodegenerative disorder Parkinson’s disease when a patient is undergoing treatment with the dopamine replacement drug levodopa.

\textsuperscript{*} Corresponding author.
E-mail addresses: alexander.turner@york.ac.uk (A.P. Turner), m.lones@hw.ac.uk (M.A. Lones), martin.trefzer@york.ac.uk (M.A. Trefzer), stephen.smith@york.ac.uk (S.L. Smith), stuart.jamieson1@nhs.net (S. Jamieson), Jane.Alty@hymns.ac.uk (J.E. Alty), jezcos@doctors.net.uk (J. Cosgrove), andy.tyrrell@york.ac.uk (A.M. Tyrrell).

\url{http://dx.doi.org/10.1016/j.biosystems.2016.05.005}

0303-2647/© 2016 The Author(s). Published by Elsevier Ireland Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
levodopa. Incorrect dosage of levodopa can have severe manifestations, notably the involuntary and violent muscle spasms known as levodopa-induced dyskinesia (LID). Correct dosage involves finding a balancing point between the removal of Parkinsonian symptoms and the onset of side-effects, providing a challenge to both patients and clinicians alike. In this work, we demonstrate the potential of epiNet by analysing the topological changes and transitions between dynamical states within the model when classifying movement data collected during levodopa therapy. We show that it is possible to provide clinically relevant information about LID, the underlying data and the processes underpinning why at any given point epiNet has made a certain decision.

2. Background

2.1. Epigenetics

Epigenetics refers to mechanisms which result in changes in gene expression without altering the underlying DNA (Bird, 2007; Turner, 2001). In Fig. 1, a general eukaryotic arrangement of DNA is illustrated. Within the cell nucleus, DNA is wrapped around a histone octamer over 1.67 turns. This combination of DNA and histones is referred to as chromatin, one of the major epigenetic structures. Chromatin has two prominent biological roles: compressing the size of a DNA strand, and controlling physical access to the DNA. The dynamically varying structure of chromatin allows DNA to move relative to it, allowing the cellular machinery (e.g. polymerase, transcription factors) to access it. Controlling this movement means controlling which genes are actually being transcribed at any given moment. This idea is central to the model used within this paper, where there exists a ‘bank’ of genes (the genome) which is inactive by default. Then, chromatin reorganises itself relative to the DNA in order to change which genes are accessible to the cellular machinery at any given time.

Additionally, there are other epigenetic marks. One of the most pervasive is DNA methylation, where a methyl group is added to either the cytosine and adenine nucleotides within DNA. In a similar vein to chromatin modifications, methylations typically prevent transcription by physically inhibiting the cellular machinery’s ability to straddle the DNA, preventing processes such as transcription (Bird, 2007; Turner, 2001). There are also other epigenetic mechanisms, such as micro RNAs, prions and SRNAs.

2.2. Epigenetic networks

The model used in this work (previously described in Turner et al. (2015)), builds upon a range of previous work from more statically derived genetic and epigenetic functionality (representing genetic networks and static epigenetic marks such as DNA methylation) (Lones et al., 2013; Turner et al., 2014; Reil, 1999; Bull, 2014) to more dynamic models which take inspiration from chromatin modifications (Turner et al., 2013, 2015, 2016). Indeed, the low-level element of epiNet, the artificial gene model, is derived from (Lones et al., 2010), and remains unchanged. This artificial gene is a computational element that takes inputs and processes them using a function, producing a single transformed output. In this paper, we use a sigmoid function, meaning that these low-level elements are similar to perceptrons in artificial neural networks. This function can be parameterised during evolution, allowing the gradient of the function and offset to change for each individual gene, allowing for functions ranging from the typical sigmoid to an approximate step function (Fig. 2). Many of these genes are linked together to form an artificial gene regulatory network, in which certain genes are mapped to external inputs and outputs (Lones et al., 2010).

We have explored several approaches to building epigenetic structures into, or on top of, existing artificial gene regulatory networks (Turner et al., 2013, 2014, 2015). In these models, genes are generally always active unless made inactive by an epigenetic analogue. Additionally, the epigenetic analogues are static and fixed in place and control a small part of the network. With epiNet, by comparison, genes are inactive until turned on by an epigenetic analogue. The epigenetic analogues are then regulated by genes, causing them to move around the genome, switching on and off different parts of the network over the course of time.

The main structure in this model is a genetic structure, consisting of a number of genes (30–100) which exist on a 1-dimensional linear scale (Fig. 3) between [0,1]. These genes are static and are not directly executed. Execution of the genes occurs when genes are copied from the genetic structure to the protein network. The protein network functions in a similar way to the networks in (Lones et al., 2010), where it is the structure which directly interacts with an external environment (task) and is executable. However, which
The components of epiNet. Genes that are covered by an epigenetic analogue become expressed, forming a structure akin to a protein network. The protein network then takes inputs from a task and is executed to produce output(s). The epigenetic analogue is regulated by the protein network; following execution, it reads expression levels in the protein network and passes them through a sigmoid function. The result then becomes the next position of the epigenetic analogue.

3. Methodology

The objectives of this work are:

- To understand the properties of epiNet as a predictive modelling tool for the analysis of levodopa-induced dyskinesia
- To understand how the analysis of epiNet's topological morphology when classifying the data can provide useful clinically relevant information about levodopa-induced dyskinesia

To achieve these objectives, we applied epiNet to the classification of real world medical data derived from accelerometers worn by Parkinson’s patients undergoing a levodopa based medication regime.

3.1. Clinical study data

The data was acquired from 25 patients who had confirmed symptomatic Parkinson’s disease and were being treated with levodopa therapy. The patients wore small lightweight sensing devices with integrated accelerometers and gyroscopes to monitor their movements at a sample rate of 100 Hz. These devices each stored 6 channels of information, corresponding to acceleration in the three translational and three rotational axes. The patients were unconstrained yet asked to complete various movement tests throughout a 2 h window at non-specific intervals. Whilst doing this, the patients were recorded using an infrared video camera. Trained clinicians then used these recordings to grade the data, using the Unified Dyskinesia Rating Scale (UDysRS) to mark up periods of LID. The data was cross-correlated between clinicians to ensure accuracy over the grading of the patients. The patients were rated on a scale of 0–4, where 0 corresponds to the patient displaying no signs of dyskinesia, and 4 corresponds to severe dyskinesia. Table 1 summarises the number of instances of each grade of dyskinesia within the data set.

3.2. Evolutionary algorithms

Evolutionary algorithms (EAs) are population-based metaheuristics which carry out optimisation procedures motivated by biological evolution. Evolution is the process of adapting an individual or population to environmental dynamics across successive
generations via the stochastic breeding of population members. Evolutionary algorithms model this principle by using the data of a system or model as a genome, and by crossing over and mutating solutions within a population according to an objective measure of their ability to solve a problem (known as their fitness). Over successive generations, the individuals’ fitness will improve as the population learns to solve a particular task.

EAs are relatively flexible in how they represent and evaluate solutions. Their appropriateness for optimising epigets is also suggested by the role of biological evolution in designing biological genetic regulatory systems. For these reasons, we used them to optimise both the architecture and parameters of the epinet model in order to solve the problem of predicting the degree of dyskinesia in segments of accelerometry time series data. We used the multi-objective evolutionary algorithm NSGA-II (Deb et al., 2002) with a population size of 400 and an optimisation period of 250 generations. We used tournament selection of size 4. Point mutation was used for all components of the genome and epigenome with a probability of 0.05. The crossover operator exchanged 30% of the data between individuals, and the crossover rate is 0.5. These parameters were selected as previous exploratory experimentation has shown that these allow for the evolution of complex behaviours which were not able to be learned by other algorithms within identical time frames (Turner, 2013; Turner et al., 2015).

We use a multi-objective fitness function in this work. The first objective is to correctly classify each time step within a data segment, with one point awarded for each correct classification. The second aggregates the points awarded for a data segment; if over 50% of the time points are correct, that data segment is considered to be correctly classified and a point is awarded. This allows the system to handle variable-sized data segments (204–16,001 time steps) without biasing search towards classifiers that perform well on only the shorter or longer samples, which tends to happen when only one of these objectives is used.

3.3. Data handling

The complete data set contained 9672 labelled sequences. To make the optimisation process computationally tractable, we uniformly sub-sampled (without replacement) 1000 instances for use as a training set. The remaining samples were used for testing, allowing a fairly robust estimate of a classifier’s generality.

The data was split into 2 classes of dyskinesia severity: high and low. High indicates an LCD rating of either 2, 3 or 4. These correspond to clinically significant levels of dyskinesia that require a change in levodopa dosage. Low indicates an LCD rating of 0 or 1, i.e. no, or very mild, dyskinesia. The training data was split accordingly; the low class comprising 500 data segments, with 250 rated 0 and 250 rated 1, the high class comprising 500 data segments with 166 rated 2, 167 rated 3, and 167 rated 4.

Each data segment comprises 6 times series, corresponding to changes of acceleration over time in the three translational and rotational axes. At each time step, this multivariate data is fed into an epinet classifier as a group of 6 values, each copied into the activation levels of input genes. After the epinet has processed the input, a single output is read from the expression level of an output gene. An output value below 0.5 is interpreted as indicating low dyskinesia; a value equal to or greater than 0.5 indicates high dyskinesia.

4. Results

We carried out 20 independent runs, during which each epinet classifier was evaluated on its ability to correctly predict the labels of each time step and each data segment within the training set. Following these 20 runs, all the final generation classifiers were re-evaluated on the test set. Fig. 4 shows the performance of these controllers on both the training and test sets, with each point showing the per-step and per-segment accuracy of a single controller. On the training set, classifiers were found with accuracies of ~80%. When re-evaluated on the test set, accuracies dropped to around ~70%; however, most showed an ability to generalise.

Fig. 5 summarises the distributions for both objectives. It is notable that test set accuracies were significantly higher for whole data segments (more important from a clinical viewpoint) rather than single time steps.

Please cite this article in press as: Turner, A.P., et al., Using epigenetic networks for the analysis of movement associated with levodopa therapy for Parkinson’s disease. BioSystems (2016), http://dx.doi.org/10.1016/j.biosystems.2016.05.005
than individual time steps. This is unsurprising, given the larger amount of historical time series data present in an entire data segment, and the relative stochasticity of movements at a single time step. However, it is also evident from Fig. 4 that the majority of the classifiers are fairly balanced in their responses at the per-step and per-segment level, with most clustered along the diagonal.

4.1. Analysis of network dynamics

The results show that trained instances of the epiNet model are capable of classifying movements associated with levodopa therapy. We next carried out post-hoc analysis of the dynamical behaviour of the epiNet classifiers, in order to gain some insight into how they achieve this.

Figs. 6 and 7 show the dynamical response of a single evolved epiNet classifier whilst processing two different acceleration time series data segments. First of all, it is evident that the dynamics of the network are adaptive, with the activation levels of the genes and the positions of the epigenetic analogues undergoing significant changes during the course of classification. This change in dynamics is caused by the autonomous topological changes of the network as a result of changing dynamics in the environmental input data. Analysis of a sample of the evolved classifiers suggests that this behaviour is typical, with <5% of the instances not undergoing topological change whilst processing the input data, and these topologically-invariant solutions having generally lower fitness values. This suggests that topological change is a useful mechanism, and it seems likely that topological changes are driven by, and hence are a signal of, dynamical changes in the patient’s movements.

For the network analysed in Figs. 6 and 7, some genes are almost always active (approximately 3–5), some genes are never active, and others are intermittently active. Additionally, the size of the expressed network varies significantly over time; sometimes no genes are active, sometimes upwards of 15 are active. It can also be noticed (particularly in Fig. 7) that some changes in the activity of the genes do not lead to changes in the overall dynamics of the network. For example, at time steps 220–800, gene 31 is constantly being activated and deactivated, yet this causes no discernible difference in the network’s output.

The evolutionary algorithm was given the freedom to evolve the architecture of the epiNets within bounds. The initial parameters for the network were 40 genes and 3 epigenetic molecules and most networks maintained a similar architecture throughout optimisation. The number of genes varied little, with the largest solutions containing 52 genes, the smallest containing 30. There appeared to be no discernible difference between the performance of networks based on the number of genes alone. Most functional solutions contained 3 epigenetic molecules. However, a significant number contained only 2, but these otherwise functioned similarly to the networks containing 3: suggesting that useful behaviour can be achieved with less molecules, but with some impairment to evolvability. Networks which contained 1 or 4 epigenetic molecules had poor functionality, with neither of them resulting being a solution in the top 20% of the population. A possibility, as seen is previous work is that too many epigenetic molecules leads to a greater likelihood of interference between regulatory regions.

4.2. Knowledge extraction

The data used to train the networks contains 6 input channels and a single output. EpiNet allows these inputs to be mapped onto genes multiple times (i.e. one input gets mapped onto multiple genes); it also allows particular inputs to be ignored by the network (Fig. 8). To understand the way in which epiNet classifiers use the inputs, we analysed the best evolved solutions.

A pervasive trend was for 3 of the 6 inputs to be used significantly more than the others, although none of the high fitness networks used a permanently static mapping. Inputs 2, 3 and 6 were a common choice; however, 4 and 5 were also used widely in other networks. It was also common for input 1 (acceleration in the x-plane) to be completely ignored by the networks. In many solutions, single inputs were mapped onto multiple genes (mostly 2, sometimes 3 or 4), which were then excited in tandem. Hence, a particular input could drive multiple regulatory circuits. This can be interpreted as an indication of the relative importance of the input variable. However, over the course of execution, not all inputs remained in use. This can be seen as a context sensitive switch, with input signals driving topological change that causes the network to

Please cite this article in press as: Turner, A.P., et al., Using epigenetic networks for the analysis of movement associated with levodopa therapy for Parkinson’s disease. BioSystems (2016), http://dx.doi.org/10.1016/j.biosystems.2016.05.005
focus on movement in particular axes. For instance, for over 75% of
the networks’ execution, half of the inputs are not used.

For further analysis, we looked at creating a minimum working
example (MWE) of an evolved solution, assuming that evolved
epiNets would be architecturally more complex than they needed
to be in order to solve the problem. This was done by iteratively
removing a gene from the network and then re-evaluating it. If the performance worsened, the gene was replaced; if not, it
was removed permanently. The epiNet architecture is particularly
robust to deleterious mutations as its genes are not all interconnected during execution, and it is therefore unlikely that all of
the dynamical regimes of the network would be affected by a single
gene deletion. We found that removing genes had a less damaging
effect to the networks’ performance than removing epigenetic
molecules, which almost always adversely affected performance.
The networks we considered initially started with 40 genes, and in
general could be reduced to around 5 genes (Fig. 9). Although this
had no effect on functionality, the networks in this state were more
sensitive to perturbations; because of this increased sensitivity, it
is probable that they lacked the ability to move through the search
space in as timely a manner as the original networks, as small per-
turbations would be likely to dramatically alter behaviour. Because
these networks were artificially created by removing genes, it is
difficult to fully assert what the networks structure would be if
they were evolved to be as small as possible. However, in previous
work (Turner et al., 2015) it was shown that doing such, removed
the robustness of the networks during execution, but the networks
were still able to find solutions to complex tasks. Hence it is likely
that artificially altering network structure has different effects than
evolving the networks to do so.
The functioning of an MWE version of the network analysed in
Fig. 9 is depicted in Fig. 10, showing the activity of each gene over
a transitional period. This network was capable of classifying both
time steps and data segments with over 69% accuracy. It can be seen
that genes 1 and 5 initially oscillate. Whilst in this dynamical phase,
the network is completely inactive for 50% of the time. Once the
transition occurs, genes 3 and 4 become permanently active. This
switching behaviour was common in the MWEs, and in general the
networks appeared to have two stable attractors corresponding to
movements in the high and low dyskinetic classes. The transition
between attractors was also, in most cases, permanent, suggesting
that once a significant dyskinetic signal was detected, the network
remained in the high dyskinesia state, increasing the likelihood of
classifying the whole segment as exhibiting dyskinesia. For most
MWEs, it was also evident that the transition between attractors
was governed by a single input. Although other inputs might be
mapped to genes after this transition, the permanent change in
attractor meant that they were not being used.
This suggests that it is possible to train effective classifiers using
only a single channel of the accelerometry data. This is significant,
as it could considerably reduce the training time of the networks. It

Please cite this article in press as: Turner, A.P., et al., Using epigenetic networks for the analysis of movement associated with levodopa
also means that patients could be fitted with simpler sensors that consume less energy and are therefore less bulky. Furthermore, the MWE equivalents of the evolved classifiers are quite small. In principle, this means they could be implemented in a relatively small digital (or potentially analogue) circuit deployed within a sensor platform. This could provide a single-device solution to monitoring dyskinesia in Parkinson’s patients.

4.3. Conclusion

In this paper we have shown that epiNet is able to classify levodopa-induced dyskinesia within patients suffering from Parkinson’s disease. We have also shown how analysis of an epiNet’s structure can give us insight into both how and why classification decisions have been made. In particular we showed that post-processing of the networks could often reduce the number of genes in the network from 40 to around 5 without altering functionality. This also showed that the networks could make decisions based on a single accelerometer input, significantly increasing the efficiency of execution. This opens up the opportunity for small, wearable diagnostic tools to be implemented using a hardware model of the MWEs.

The results are promising, however, there were specific caveats associated with the work in this paper. The first is that the combination of population based evolutionary algorithms and epiNet meant that run times were relatively high. This limited the scope of the networks ability to learn, but conversely, also allowed for a comprehensive method of testing because a large amount of data was not used in training. It would have also been beneficial to allow for a larger population size and more generations to gather a better understanding of epiNet’s potential. A further possibility to improve this is to migrate the software to a GPU implementation as a significant part of the processing within the networks is floating point arithmetic, hence, the performance increase should be substantial. Within this work there has not been a direct focus on objective performance, but a focus on the underlying dynamics of the networks when classifying high frequency movement data. Because of the novel dynamical properties of the network, we feel that direct comparison with other techniques with differing dynamical functionality would at present be of limited value as variations in performance could be attributed to the parameters of the respective techniques. The work in this paper is to be used as an underpinning of future research in which a robust analysis of epiNet performance will be carried out in reference to other models, such as those in Lones et al. (2014b).

Within this work we had to reclassify the data into either a high or low levodopa-induced dyskinesia rating. In addition, there was a class imbalance where there was an order of magnitude more data available data sets with no levodopa-induced dyskinesia, compared to those who had the most pronounced symptoms. Hence, it might be possible to look at a class weighting which is dependent on classification outcomes (and how incorrect classifications may effect patients) and the relative performance of the networks. Indeed, it may also be beneficial to look at re-balancing the classes by omitting certain classifications within the data (to only focus on the most severe symptoms) as seen in Lones et al. (2014a).

Overall, the results in this paper have shown that epiNet has significant potential as a classification tool, and that the analysis of the network structure and dynamics can provide clinically relevant information about the decisions made by epiNet which are important both for patient outcome and model validation.

Acknowledgements

The authors acknowledge the support of the EPSRC through the platform grant EP/K040820/1. Data created during this research is available at the following DOI: 10.15124/7eb2b24-89b6-4a8d-b7cb-621861e347b8.

We thank the patients with Parkinson’s who participated in this study. Without their support, this research would not have been possible.

References


Please cite this article in press as: Turner, A.P., et al., Using epigenetic networks for the analysis of movement associated with levodopa therapy for Parkinson’s disease. BioSystems (2016), http://dx.doi.org/10.1016/j.biosystems.2016.05.005


