

Temporal Bayesian classifiers for modelling muscular dystrophy expression data

Allan Tucker^a, Peter A.C. 't Hoen^b, Veronica Vinciotti^a and Xiaohui Liu^{a,c}

^a*School of Information Systems Computing and Mathematics, Brunel University, Uxbridge UB8 3PH, UK*

^b*Center for Human and Clinical Genetics, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands*

^c*Leiden Institute of Advanced Computer Science, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands*

Abstract. The analysis of microarray data from time-series experiments requires specialised algorithms, which take the temporal ordering of the data into account. In this paper we explore a new architecture of Bayesian classifier that can be used to understand how biological mechanisms differ with respect to time. We show that this classifier improves the classification of microarray data and at the same time ensures that the models can easily be analysed by biologists by incorporating time transparently. In this paper we focus on data that has been generated to explore different types of muscular dystrophy.

1. Introduction

The analysis of microarray data has previously focussed on the clustering of genes into groups of similar expression profiles. This has amongst other things allowed biologists to infer the functions of previously unknown genes. More recently, methods to learn gene networks from such data have been explored with the aim of trying to investigate more than just pairwise relationships and understand the interactions between genes in more detail [7]. Another research problem that has arisen from microarray data is the classification of different samples of data into categories such as diseased and control groups. Many microarray datasets contain thousands of genes and the number of samples are usually very small. Therefore methods such as feature selection are required to prevent over-fitting. Previously we have developed a method for classifying this sort of data that uses simple models, sampling and global feature selection algorithms [17].

Microarray data from time-series experiments, where gene expression profiles are measured over the course of the experiment, require specialised algorithms. Recently, papers have documented using time-series models to capture the temporal relationships between genes [12,19]. Previously, we have investigated Bayesian classifiers for modelling relationships between different variables over time to classify visual field data [13]. This made use of dynamic links between genes. However, another method can be used to model temporal relationships which are relative to a fixed point in time, say the birth of an organism or the onset of a medical condition, where any feature over time is measured from that point. For example, if a variable can be used to determine medication for a condition that has been diagnosed previously then the decision is dependent on the time since the diagnosis.

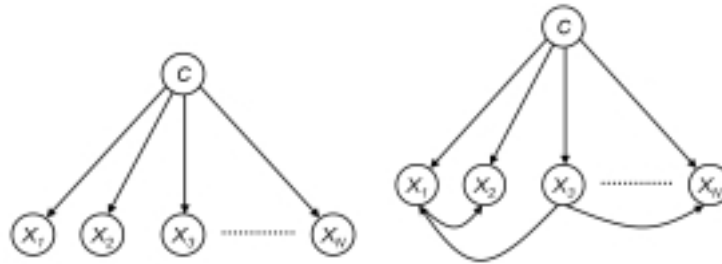


Fig. 1. Common Bayesian network classifier architectures: Naïve Bayes Classifier (NBC), Bayesian Network Classifier (BNC).

In this paper we introduce new models that capture temporal relationships and help to understand how biological mechanisms differ. We aim to improve the classification of time-series microarray data using new forms of Bayesian classifiers, whilst at the same time ensuring that the models can easily be analysed by biologists by using models that incorporate time transparently. These classifiers are described in Section 2. We focus on microarray data that has been generated in order to explore the different types of muscular dystrophy. The data and the experiments carried out are also described in Section 3. Section 4 documents the results and analyses the results whilst Section 5 concludes with implications and lines for future work.

2. Methods

2.1. Bayesian classifiers

Bayesian Networks (BNs) [11] are probabilistic models that can be used to model data transparently. This means that it is relatively easy to explain to non-statisticians how the data are being modelled unlike other ‘black box’ methods. A BN is a directed acyclic graph consisting of links between nodes that represent variables in the domain. Links are directed from a parent node to a child node, and with each node there is an associated set of conditional probability distributions.

Bayesian classifiers are a special form of Bayesian network where one node represents some classification of the data. The simplest Bayesian classifier is the Naïve Bayes Classifier (NBC). This classifier has been used with surprising success, given its simplicity, on a number of different applications. It consists of a set of probability distributions for each variable given the class. The assumption behind this is that each variable is independent of one another given the class. A more general Bayesian Network Classifier (BNC) includes links between the predictor variables [6]. This requires learning a network structure between variables using some scoring metric coupled with a heuristic search. An example of the structure of each classifier is illustrated in Fig. 1.

Many datasets involve measurements of variables over time and the dynamic Bayesian network (DBN) [5] is an extension of the BN to handle the sort of relationships found in time-series. A DBN is a BN where the N nodes represent variables at differing time slices. Therefore links occur between nodes over time and within the same time lag.

As opposed to DBNs, BNs can also incorporate time by including temporal nodes into the Bayesian network structures. For example, Friedman et al. [7] used this method when modelling yeast cell-cycle data where the temporal node was made the parent of every gene node.

In order to classify new data given a Bayesian classifier we need to perform inference and in this paper, we use a form of stochastic simulation called logic sampling [11] because of its speed and its intuitive

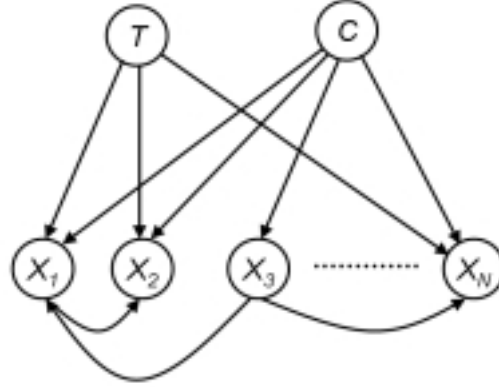


Fig. 2. Proposed architecture for incorporating time, the Temporal BNC (TBNC).

appeal. We now discuss how these methods of incorporating time can be used to analyse and classify gene expression data with respect to absolute and relative temporal relationships.

2.2. Incorporating time into Bayesian classifiers

Figure 2 illustrates the classifier architecture that we explore in this paper, the Temporal BN Classifier (TBNC). The TBNC allows genes to be conditioned upon a node that represents the time from some reference as well as other genes and the class node. Therefore the classification will take into account the time from the reference as well as the log ratio of the gene.

2.3. Learning the classifiers

For all classifiers the links between the class node and every gene are automatically inserted and fixed during the search. Links between genes and between the time node and genes are explored using a simulated annealing approach similar to one we used in [17] but that minimises the Minimum Description Length (MDL) of the network [9]. The main idea behind MDL is to compute the description length of the network as the sum of the description length of the model and the description length of the data given the model. For our problem, this is given by

$$\text{MDL-score} = \text{DL}_{\text{Model}} + \text{DL}_{\text{Data}} \quad (1)$$

$$\text{DL}_{\text{Model}} = \sum_i (1 + \pi_{X_i}) \log(N + 1) + \sum_i (q_i (s_i - 1) \frac{\log M}{2}) \quad (2)$$

$$\text{DL}_{\text{Data}} = M \sum_i H(X_i|C) \quad (3)$$

where s_i is the number of discretized states of the gene variable X_i , $n(x_{ij}|c)$ is the number of cases in the dataset where X_i takes on its j th unique state within the samples from class c , and $n(c) = \sum_{j=1}^s n(x_{ij}|c)$ is the total number of samples from class c . π_{X_i} is the parent set of X_i , q_i is the number of unique instantiations of the parents of node X_i . From $\hat{p}(x|c)$, an estimate of $p(c|x)$ is calculated using Bayes rule and the resulting classification rule assigns the sample x to the class associated to the highest estimated probability. $H(X_i|C) = -\sum_{j,c} n(x_{ij}, c) \log n(x_{ij}|c)$ and M is the number of biological

samples. More complex networks are expected to lead to a shorter description of the data given the model at the expense of a longer description needed to describe the model itself. In this way, the score in Eq. (1) provides a trade-off between complexity of the model and goodness-of-fit to the data. Furthermore, the factor $\frac{\log M}{2}$ in Eq. (2) is a penalty term based upon the sample size [8]. For datasets where the sample size is particularly small, this penalty will have a more pronounced effect.

This global optimisation search was chosen with the aim of avoiding local optima, which many greedy searches suffer from. The main idea behind our method is to make small changes to the classifier structure and then score the network. The changes involve using three operators, add, delete and swap to randomly add a link, remove a link and swap a link.

The optimisation algorithm is documented fully below, where D represents the input data, the initial annealing temperature is denoted by t_0 , the cooling parameter for the temperature by c , the maximum number of scoring function calls by maxfc and the score of a network by $\text{score}(\text{bn})$, computed by the MDL. $R(0, 1)$ is a uniform random number generator with limits 0 and 1. For all our experiments, we set t_0 to 1. This was based upon the initial scores when applied to the dataset investigated in this paper (we have generally found that a good starting temperature is similar to the changes in score in the early iterations). maxfc was set to 10000 as this was found through empirical analysis to ensure that convergence has occurred on the dataset explored. c was set to 0.999, calculated to ensure that the temperature after maxfc iterations was suitably close to zero.

```

Input:   $t_0, \text{maxfc}, D$ 
         $\text{fc} = 0, t = t_0$ 
        Initialise  $\text{bn}$  to a Bayesian
        classifier with no inter-gene links
         $\text{result} = \text{bn}$ 
         $\text{oldscore} = \text{score}(\text{bn})$ 
        While  $\text{fc} \leq \text{maxfc}$  do
            For each operator do
                Apply operator to  $\text{bn}$ 
                 $\text{newscore} = \text{score}(\text{bn})$ 
                 $\text{fc} = \text{fc} + 1$ 
                 $\text{dscore} = \text{newscore} - \text{oldscore}$ 
                If  $\text{dscore} < 0$  then
                     $\text{result} = \text{bn}$ 
                Else If  $R(0,1) < e^{\text{dscore}/t}$  Then
                    Undo the operator
                End If
            End For
             $t = t \times c$ 
        End While
Output:  $\text{result}$ 

```

Algorithm 1: Simulated annealing for building Bayesian networks

3. Muscular dystrophy data

Muscular dystrophies are a heterogeneous group of inherited disorders characterized by progressive muscle wasting and weakness. The genetic defects underlying many muscular dystrophies have been

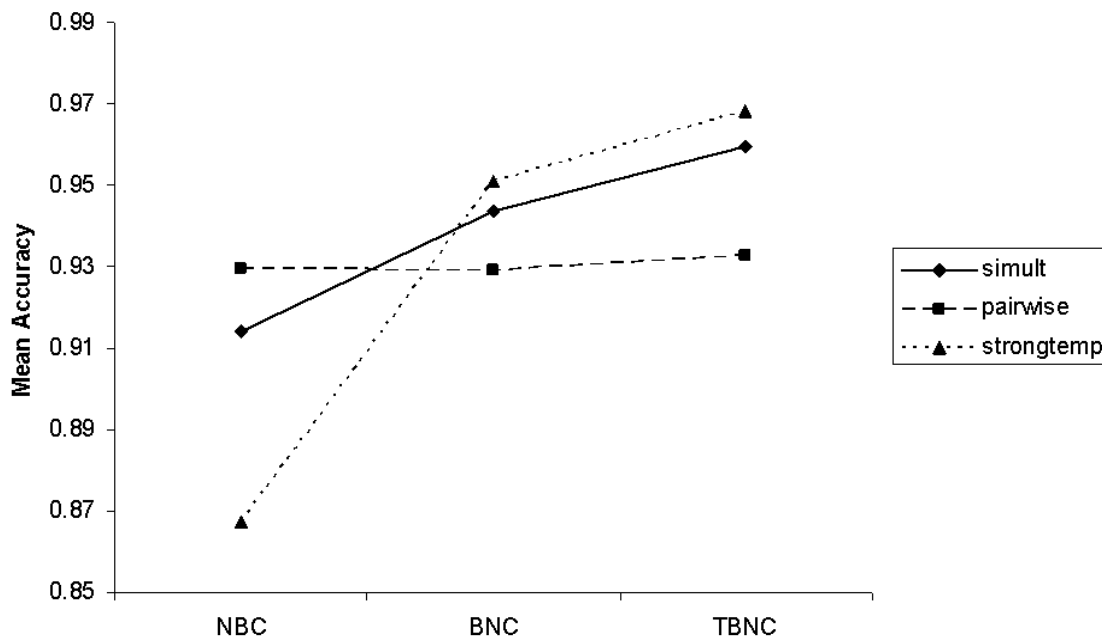


Fig. 3. Comparison of mean classification accuracy for each dataset.

elucidated [3,4]. A particular subset of muscular dystrophies is caused by mutations in genes coding for constituents of the dystrophin-associated glycoprotein complex (DGC). Mutations in the dystrophin gene cause Duchenne muscular dystrophy, whereas mutations in sarcoglycan genes are responsible for Limb-Girdle Muscular Dystrophies. Large-scale gene expression profiling of mouse models known to recapitulate different human muscular dystrophies was performed to delineate the molecular mechanisms underlying the shared and distinct phenotypic characteristics [14,15].

The MDX mouse is a mouse model for Duchenne muscular dystrophy, beta-sarcoglycan-deficient (BSG) and gamma-sarcoglycan-deficient (GSG) mice are mouse models for Limb-Girdle Muscular Dystrophies 2E and 2C. Expression profiles were generated from two individual mice (two biological replicates) at different ages: 1, 2.5, 4, 6, 8, 10, 12, 14, 20 weeks. There were four technical replicates in the experiment: the arrays were spotted in duplicate and the samples were hybridized twice (dye-swapped). The arrays used were spotted 7.5 K 65-mer oligonucleotide arrays (Sigma-Genosys mouse library). A temporal loop hybridization design was applied in which consecutive time points were hybridized to the same array. The data were normalised using the `all.norm` function from the `smida` R-library. The method essentially corrects for spatial, dye and across-array effects. These normalization procedures are applied in a sequential manner, starting with local corrections and proceeding towards more global corrections like across-array normalization. More details about the methods can be found in [18]. The log-ratios of the gene expression at a particular time point with respect to the first time point were then considered for the study. These were estimated from the raw log-ratios using a simple linear model [18]. The data were then discretised into four states using a frequency-based policy whereby the resultant genes appear in each state with equal probability.

For this paper, we focused on the analysis of three subsets of genes. These subsets were selected on the basis of a new statistical algorithm that we previously described. This algorithm involves the fit of a

second order polynomial through the temporal expression data of each class, followed by the application of a Hotelling T-test to test for differential expression between the classes [16]. The different subsets are:

- a) The 50 most significant genes from the hotelling T-test, testing differential expression between the four classes simultaneously [16]. From now on we refer to this as the *simultaneous-test* dataset.
- b) A combination of biomarker genes for WT, MDX, BSG and GSG. The biomarker list is generated by taking those genes that have adjusted p-value < 0.05 in three pairwise hotelling T-tests. For example, for MDX $p < 0.05$ for MDX vs WT, MDX vs BSG AND MDX vs GSG. We refer to the concatenated list for WT, MDX, BSG and GSG as the *pairwise-test* dataset. The individual lists of biomarker genes (used to create Fig. 9) are referred to as WT-PT, MDX-PT, BSG-PT, and GSG-PT, respectively.
- c) Those genes with adjusted $p < 0.05$ in simultaneous test *and* quadratic coefficients from fitted parabolic curve > 0.03 or < -0.03 in either WT, MDX, BSG or GSG. From now on we refer to this as the *strong-temporal-effect* dataset, because this selection of genes demonstrates most extensive temporal regulation, but are not necessarily the most discriminative genes between classes.

As there are two independent biological samples for each class of muscular dystrophy, we have decided to perform two-fold cross validation where one experiment involves training from data based solely on one biological sample and tested on the other. This approach avoids testing on data that are highly correlated with the training set (as a higher correlation is expected between technical repeats). Furthermore, for each fold, we repeat the network search 10 times, due to the stochastic nature of our simulated annealing algorithm. For each of these runs, the frequency count is maintained for each link in all networks generated on the training data for the corresponding fold and the classifiers tested on the portion of data taken out. In this way we are able to produce a confidence measure for each link in the network based on different training samples. This is similar to the method used by [7], where the confidence measure on links in a Bayesian network is achieved by bootstrapping the data.

4. Results

First of all we investigate the accuracy of the classifiers when determining whether the disease is not present (the wild-type) or, if it is, which form the disease takes – a four class classification problem. Figure 3 shows the accuracy of the classifiers when applied to each of the three datasets. It is evident that the simple NBC performs the worst with accuracies ranging from 87% to 93% (note that a totally random classification output would result in 25% accuracy). The BNC which models the dependencies between genes results in an increase in the mean accuracy for all datasets except pairwise-tests which shows a slight decrease. This could be due to spurious correlations being learnt due to temporal relationships not being taking into account and this is supported by the fact that the TBNC performs considerably better on the simultaneous-test and strong-temporal-effect datasets and slightly on the pairwise-tests, with accuracies ranging from 93% to 97%.

The results on the pairwise-tests dataset in general showed least improvement when taking time into account. Indeed, early experiments actually generated worse accuracies than for NBC. This was likely to be due to overfitting on data that did not exhibit much temporal variation. Therefore, the penalty term in Eq. (2) was doubled to $\log M$ for this dataset to generate less connected networks to minimise overfitting. The results shown here are those generated with the increased penalty term. It can be seen that there is still little increase over NBC. This highlights the risk of using more complex temporal

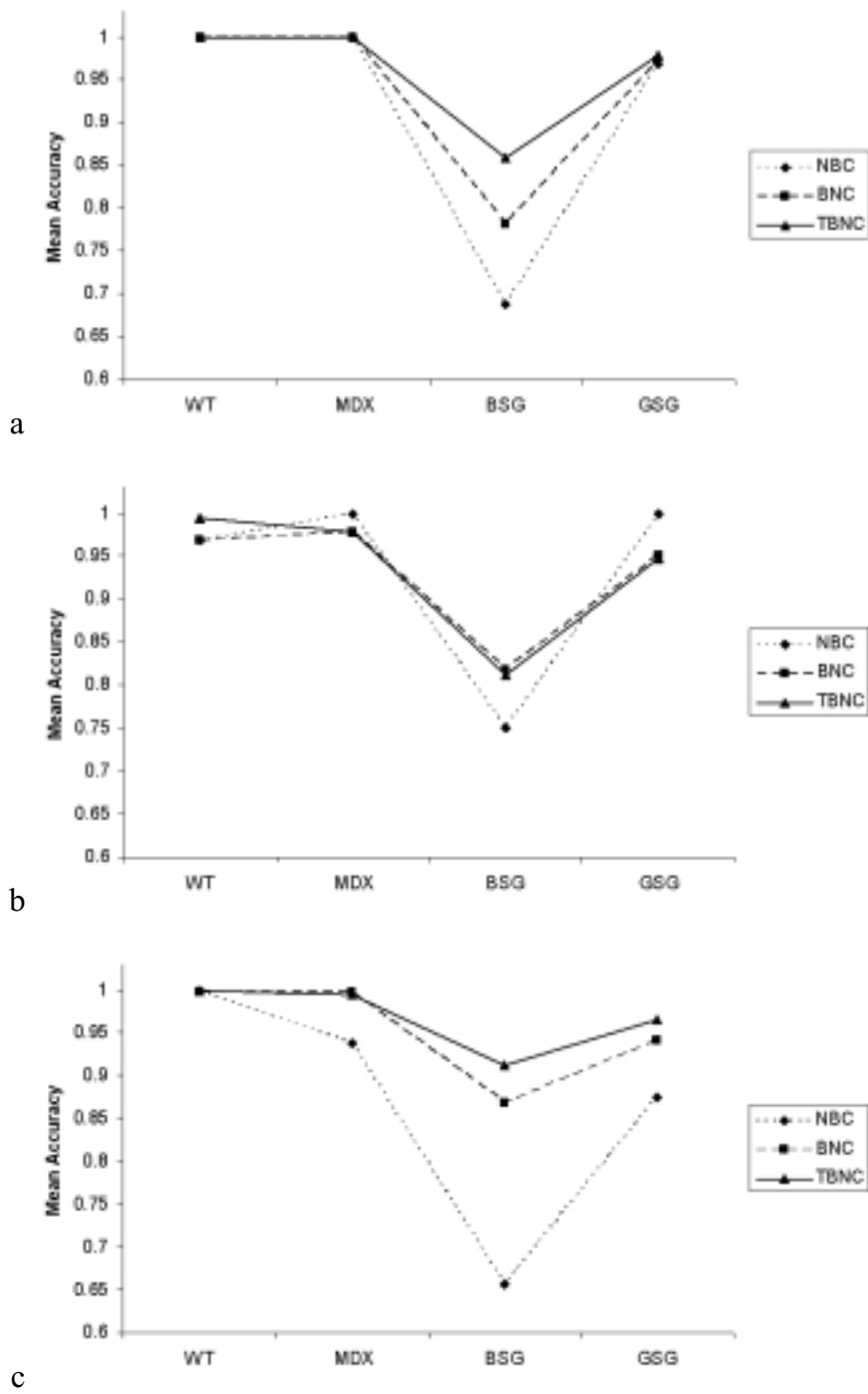


Fig. 4. Accuracy for each class type for (a) Simultaneous-Test dataset, (b) Pairwise-Tests dataset and (c) Strong-Temporal-Effect dataset.

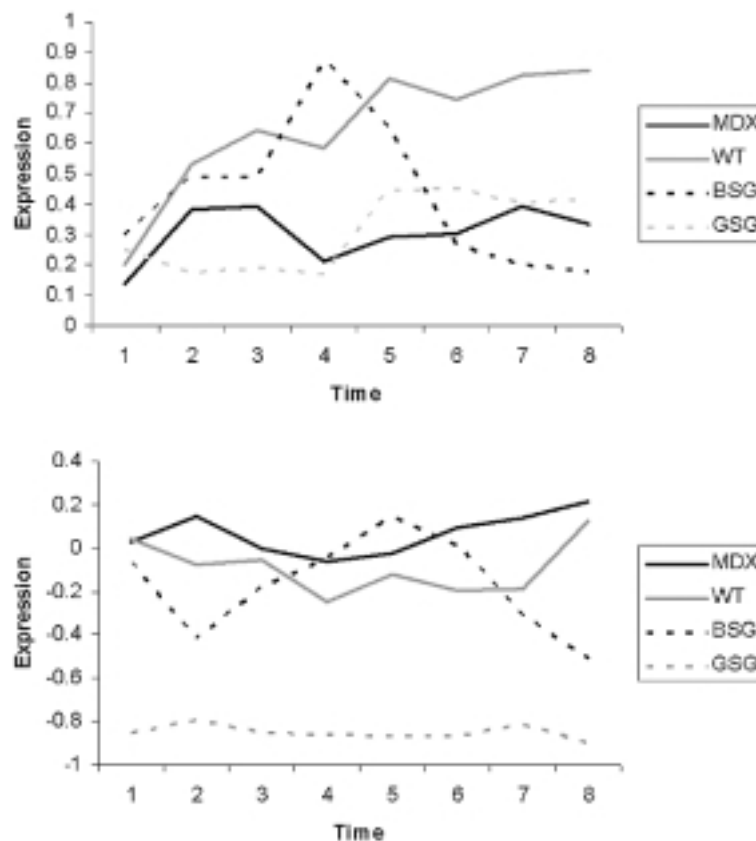


Fig. 5. Examples of genes found associated with the time node for the simultaneous-test dataset. *Dusp13* (top) and *Fpr-rs* (bottom) were found in 100% and 5% of runs, respectively.

models when learning classifiers, particularly when there is little temporal information exhibited in the data. However, if a strict enough penalty term is used to reduce overly complex networks then overfitting can be avoided. It should be noted that in contrast to the pairwise-test data, the results on the other datasets demonstrated considerable improvement without need for a stronger penalty term, in particular the strong-temporal-effect dataset, which improved from 85% to 96%.

We now look at the accuracy of the methods for each of the four classes in Fig. 4. It appears that the wild-type (WT) and the MDX form of muscular dystrophy are more easily classified than the BSG and GSG forms for all datasets. This is probably because the disease progression in the two sarcoglycanopathies (BSG and GSG mice) is similar and more progressive than in the dystrophinopathy (MDX), which makes it more difficult to discriminate between the two. BSG may be even harder to classify than GSG due to the fact that the GSG mice demonstrate a slightly more severe phenotype than the BSG mice. Notice, that the TBNC performs the best across all classes for both the simultaneous-test and strong-temporal-effect datasets but that the results are much less clear on the pairwise-tests dataset with NBC slightly outperforming the others for MDX and GSG. This implies that the genes contained within pairwise-tests classify the data without taking into account the temporal information, unlike the other two datasets. By looking at the coefficients of the fitted polynomial curves through the temporal expression profiles, it is clear that indeed the genes in the pairwise-tests demonstrate less extensive temporal regulation, but are rather differentially expressed over the whole time-course.

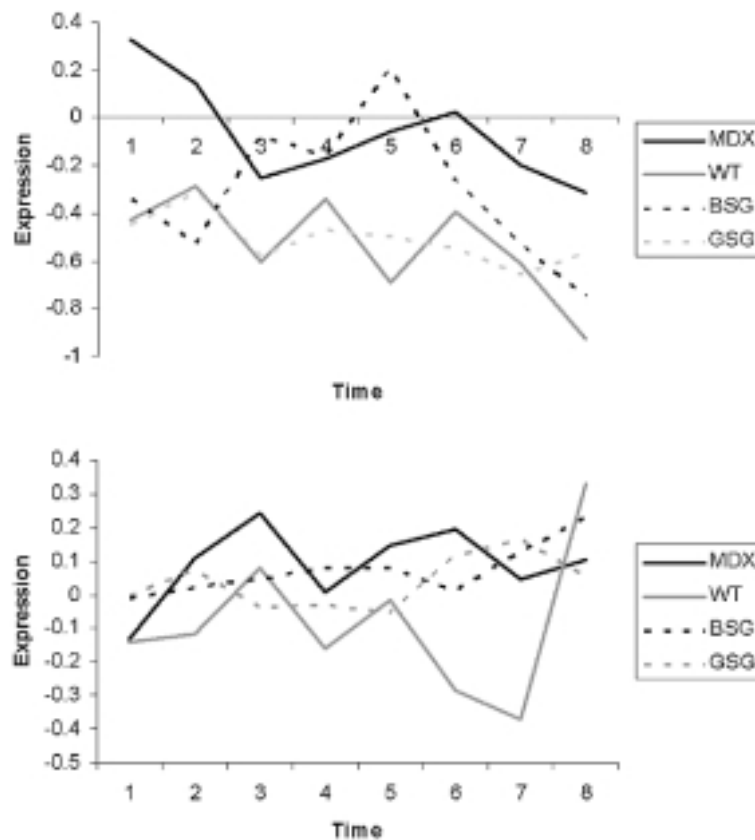


Fig. 6. Examples of genes found associated with the time node for the pairwise-tests dataset. Pdlim1 (top) and Acdc (bottom) were found in 40% and 0% of runs, respectively.

Also, notice that the TBNC improvement is more pronounced on the BSG classification for all the datasets, but also on the MDX and GSG classification on the strong-temporal-effect dataset. This implies that the relationship between the genes and the class should take into account the temporal information in order to generate reliable predictions when classifying such data.

Due to the transparent nature of Bayesian network classifiers we can explore some of the discovered relationships between genes including the temporal aspect in TBNC.

For the three MD datasets, we plot a sample of the gene that was discovered to be most commonly associated with the time node and that which was least. The plots in Figs 5–7 show the expression profiles of these genes. Notice that there is in general more overlap in expression values between the classes on the genes commonly associated with the time node: A gene may have high expression in the earlier time points for one class but low expression later on, whereas for another class the same gene may be lowly expressed early on and highly expressed later. If time is not taken into account then this relationship between gene and class is lost, and only genes that have consistently high or low expression for each class will assist classification.

For example, the genes found in the simultaneous-test dataset are shown in Fig. 5. Dusp13 is shown to have lower expression in the earlier timepoints for GSG than for BSG but this is not the case at later timepoints. For the pairwise-tests dataset (in Fig. 6), Pdlim1 appears to have numerous overlaps throughout the time course between the different classes. This is also the case in the strong-temporal-

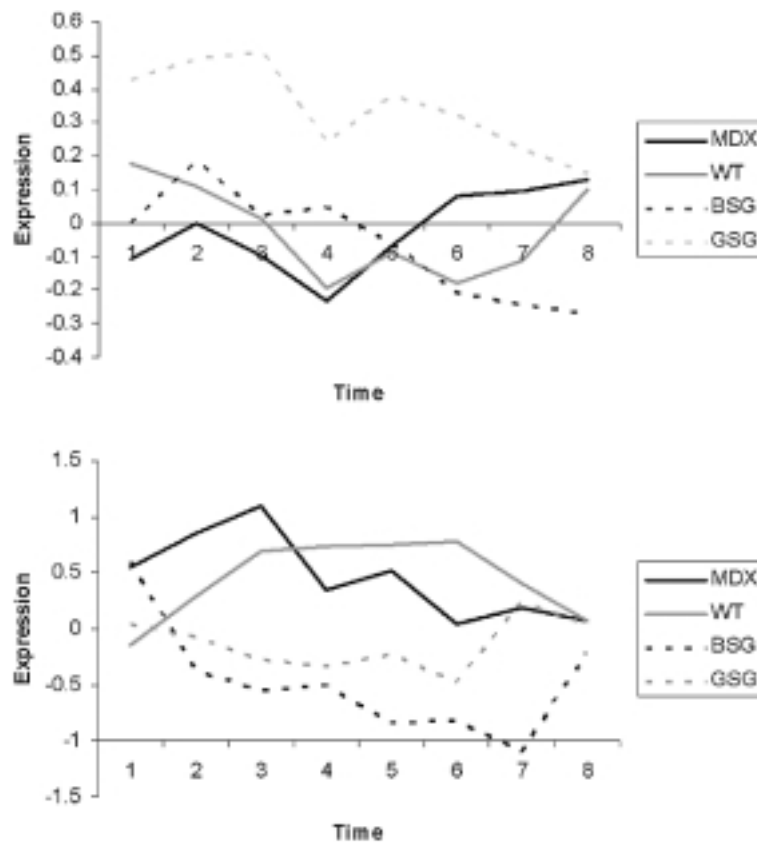


Fig. 7. Examples of genes found associated with the time node for the strong-temporal-effect dataset. *Errb2ip* (top) and *Ryr1* (bottom) were found in 45% and 0% of runs, respectively.

effect dataset where *Errb2ip* also contains numerous overlaps between the classes in Fig. 7.

In contrast to the genes found linked to the time node with high-frequency, the low-frequency genes in Figs 5–7 show significantly less change over time within each class, despite being good at differentiating the class in some cases. For example, *FPr-rs* in Fig. 5 shows that *GSG* is consistently lower than the other classes and *MDX* is nearly always higher than the others. This implies that the gene is very good at distinguishing the classes but that it does not require time to make the prediction. A similar observation can be made about *Ryr1* in Fig. 7. In Fig. 6, *Acdc* appears to contain some overlaps so it was surprising not to see this gene associated with the time node. However, there could easily be other genes that capture a similar relationship more consistently or robustly than this gene, so it could be that *Acdc* does not influence the network scoring metric enough to justify a link to the time node (recall that overly connected networks are penalised).

Tables 1–3 report the most commonly occurring links during learning the TBNCs for each MD dataset where *Time* denotes the temporal node. Notice that for the simultaneous-test dataset all of the links found with high frequency are associated with the time node. This is likely to be because this dataset contains the most significant differences in temporal behaviour between the classes. In contrast, the pairwise-tests dataset has many links between genes but with lower overall frequencies. This makes sense when looking at the accuracies in Figs 3 and 4 as the TBNC only showed a very small improvement over the NBC and BNC. The strong-temporal-effect dataset also has many links associated with the time node but the

Table 1

Table of the most frequently occurring genes found linked to the temporal node in TBNC for the Simultaneous-Test data

TBNC Link	Percent Fold 1	Percent Fold 2	Percent Total
Time → Dusp13	100	100	100
Time → Dlk1	80	90	85
Time → Casq2	80	80	80
Time → Pla2g7	60	90	75
Time → Mglap	70	80	75
Time → Myoc	100	50	75
Time → Pdlim1	30	90	60
Time → Ptpd	50	70	60
Time → Postn	60	60	60
Time → Gadd45a	60	60	60
Time → Rbpms	60	60	60
Time → Lgmn	80	40	60
Time → Nup210	40	70	55
Time → Peg3	50	60	55
Time → Actg1	50	60	55
Time → Lgals3	70	40	55
Time → Cri1	30	70	50
Time → Gtl2	40	60	50
Time → Ndn	40	60	50
Time → Gpiap1	40	60	50
Time → Dll1	60	40	50
Time → Ldb2	70	30	50

Table 2

Table of the most frequently occurring genes found linked to the temporal node in TBNC for the Pairwise-Tests data

TBNC Link	Percent Fold 1	Percent Fold 2	Percent Total
Opn1mw → Car11	40	70	55
Time → Pdlim1	0	80	40
Atp5j → Plagl1	0	80	40
Atp5j → Fbln5	70	0	35
Slc7a10 → Ndr2	40	30	35
Il6 → Atp5j	30	30	30
Plp1 → Matn2	60	0	30
NM010563 → Gtl2	60	0	30
Mga → J02644	60	0	30
Ndr2 → Pla2g7	0	60	30
Atp5j → Dpp4	50	10	30

frequencies are generally lower than for the simultaneous-test dataset, indicating less confidence in the genes being differentially expressed.

We also explore the relationships further using inference on a small subnetwork generated from the most interesting discovered links for the Simultaneous-Test data. Figure 8 shows such a network for three genes when (a) time is not observed and, (b) time is observed allowing the predicted class to be calculated with much higher confidence. In Fig. 8(a), class 0 and 3 are both potential classes (representing WT and GSG with probabilities of 0.662 and 0.287, respectively) whereas when Time is taken into account as in Fig. 8(b) the classification is resolved to class 3 – GSG – with a probability of 0.988. This nicely demonstrates how certain genes can be made to add useful information to a classifier but only when time is also a factor.

Dlk1 is highly expressed at the first time point and drops down quickly after that. However, the

Table 3
Table of the most frequently occurring genes found linked to the temporal node in TBNC for the Strong-Temporal-Effect data

TBNC Link	Percent Fold 1	Percent Fold 2	Percent Total
Time → Erbb2ip	40	50	45
Time → Snrpn	40	50	45
Time → Rpl38	40	40	80
Rpl38 → AJ251508	30	40	35
Ndr2 → Ryr1	30	40	35
Time → Pdlim5	40	20	30
Time → Il6	40	20	30
Time → Adh1	20	40	30
Time → Hk2	30	30	30
Time → Col1a1	10	50	30
Time → Psmc2	40	20	30
Hspa8 → Ndr2	20	40	30
Hspa8 → Eef1a1	50	10	30

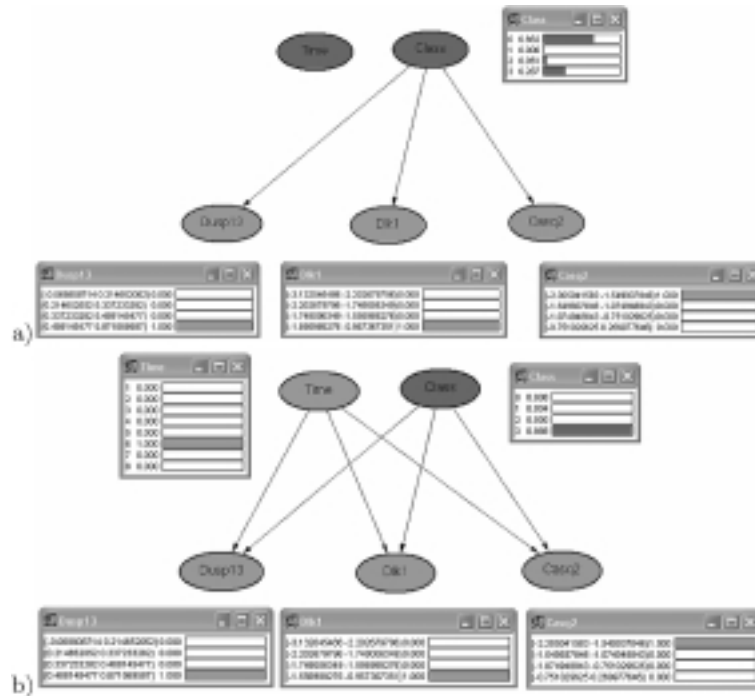


Fig. 8. Example of a network constructed with three genes associated with the time node for the simultaneous-test dataset when (a) time is not observed and, (b) time is observed. Notice how observing time in the network improves the confidence of the classification (shown by the distribution of the class node).

rate and the timing of the decrease in expression is very different for the four animal models. This was confirmed by an independent quantitative PCR assay [16]. The different temporal profiles for the different strains makes it logical that the time node is required here for correct and highly accurate classification. *Dlk1* is an interesting gene with respect to muscle disease, since mutations in sheep leading to prolonged postnatal expression give rise to the callipyge muscular hypertrophy phenotype [1, 10]. From the literature, *Dusp13* is a highly muscle- and testis-specific protein [2]. The expression in

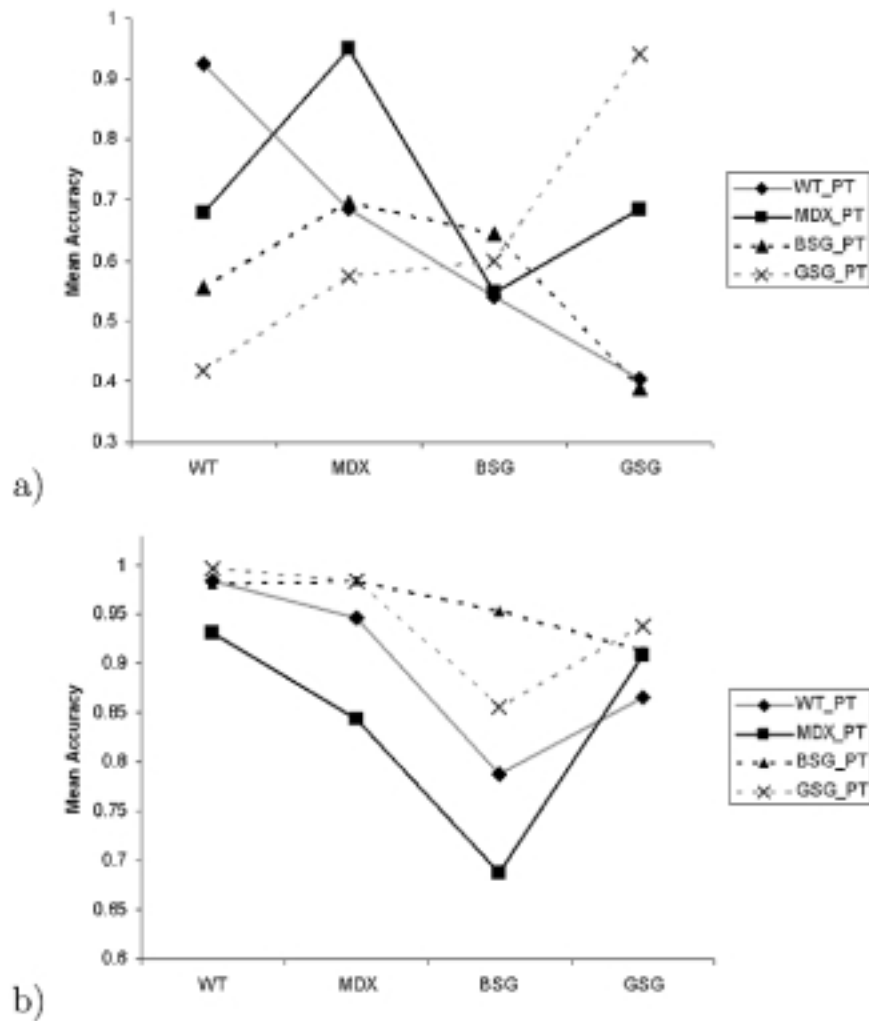


Fig. 9. The accuracy of the TBNC when applied to the subsets based upon the pairwise tests for the four different classes. a) using the pairwise-tests criteria and b) using the strong-temporal-effect criteria.

wild-type animals gradually increases over time (which is what is seen in wild-type animals, but not so much in the other strains). This would be an interesting protein for a follow-up study. Casq2 which is found with high frequency in the TBNC model is also supported in [16].

In addition to the three datasets that we have investigated here, we also explore how the TBNC performs on several other subsets of genes that are based on comparisons between each class: WT, MDX, BSG and GSG to the others for the pairwise-test and strong-temporal-effect criteria. Therefore, each respective subset is expected to be biased towards classifying that respective class. Figure 9 shows the results for the pairwise-tests and strong-temporal-effect criteria. It is clear on the pairwise-tests results that when the subset of genes are selected based upon a particular class, the classifier performs best at discriminating that class as would be expected, though this is not as clear on the BSG class. For the results using the strong-temporal-effect data, the effect is less evident. It is clear, however, that the classification of BSG, which is normally difficult, benefits greatly from the inclusion of some genes with strong temporal regulation in BSG.

5. Conclusions

In this paper we have investigated the use of Bayesian classifier architectures to classify time-series data with the aim of explaining the underlying structure of the data. The methods have been applied and tested on different subsets of real-world microarray data in order to classify forms of muscular dystrophy. In addition, the classifiers have enabled us to explore the interactions between genes responsible for differentiating between the classes. We have found that incorporating temporal information in the form of a time node results in more accurate classifiers that explicitly model time assisting in the interpretation of the models.

In [17], we explored the use of simple Bayesian classifiers for selecting genes that differentiate between different classes. Due to the small sample sizes of our datasets, we only explored classifiers that assumed independence between genes. However, more interesting features could be discovered if we took into account the relationships between genes including the temporal ones. We intend to combine our work presented in this paper with our feature selection methods in order to identify combinations of genes that work together over time that determine the class of the gene profile in question. We also intend to incorporate expert knowledge by hard-wiring certain key relationships into the networks and experiment with a number of different biological resources.

We would like to thank Professor Joost Kok for helping initiate the collaborations between Leiden and Brunel. This work was supported in part by the BBSRC in the UK (grants EGM17735 and BBC5062641).

References

- [1] C. Charlier, K. Segers, L. Karim, T. Shay, G. Gyapay, N. Cockett and M. Georges, The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting ehir imprinting nature, *Nature Genetics* **27**(4) (2001), 367–369.
- [2] H.-H. Chen, R. Luche, B. Wei and N.K. Tonks, Characterization of two distinct dual specificity phosphatases encoded in alternative open reading frames of a single gene located on human chromosome 10q22.2, *J. Biol. Chem.* **279**(40) (2004), 41404–41413.
- [3] R.D. Cohn and K.P. Campbell, Molecular basis of muscular dystrophies, *Muscle Nerve* **23** (2000), 1456–1471.
- [4] I. Dalkilic and L.M. Kunkel, Muscular dystrophies: genes to pathogenesis, *Curr. Opin. Genet. Dev.* **13** (2003), 213–238.
- [5] N. Friedman, *Learning the Structure of Dynamic Probabilistic Networks*, In Proceedings of the 14th Annual Conference on Uncertainty in AI, 1998, 139–147.
- [6] N. Friedman, D. Geiger and M. Goldszmidt, Bayesian network classifiers, *Machine Learning* **29** (1997), 131–163.
- [7] N. Friedman, M. Linial, I. Nachman and D. Pe'er, Using Bayesian networks to analyze expression data, *Journal of Computational Biology* **7** (2000), 601–620.
- [8] N. Friedman and Z. Yakhini, *On the Sample Complexity of Learning Bayesian Networks*, In Proceedings of the 12th conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1996, 274–282.
- [9] W. Lam and F. Bacchus, Learning Bayesian belief networks: an approach based on the MDL principle, *Computational Intelligence* **10**(4) (1994), 269–293.
- [10] S. Murphy, B. Freking, T. Smith, K. Leymaster, C. Nolan, A. Wylie, H. EVans and R. Jirtle, Abnormal postnatal maintenance of elevated dlk1 transcript levels in callipyge sheep, *Mamm Genome* **16**(3) (2005), 171–183.
- [11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.
- [12] M. Ramoni, P. Sebastiani and P. Cohen, Bayesian clustering by dynamics, *Machine Learning* **47**(1) (2002), 91–121.
- [13] A. Tucker, V. Vinciotti, D. Garway-Heath and X. Liu, A spatio-temporal bayesian network classifier for understanding visual field deterioration, *Artificial Intelligence in Medicine* **34** (2005), 163–177.
- [14] R. Turk, E. Sterrenburg, E.J. de Meijer, G.J.B. van Ommen, J.T. den Dunnen and P.A.C. 't Hoen, Muscle regeneration in dystrophin-deficient mdx mice studied by gene expression profiling, *BMC Genomics* **6**(98) (2005).
- [15] R. Turk, E. Sterrenburg, C.G.C. van der Wees, E.J. de Meijer, S. Groh, K. Campbell, S. Noguchi, G.J.B. van Ommen, J.T. den Dunnen and P.A.C. 't Hoen, Common pathological mechanisms in mouse models for muscular dystrophies, *FASEB J* **20** (2006), 127–129.

- [16] V. Vinciotti, X. Liu, R. Turk, E. de Meijer and Peter A.C. 't Hoen, Exploiting the full power of temporal gene expression profiling through a new statistical test: Application to the analysis of muscular dystrophy, *BMC Bioinformatics* (183) (2006), 17.
- [17] V. Vinciotti, A. Tucker, P. Kellam and X. Liu, Robust selection of predictive genes via a simple classifier, *Applied Bioinformatics* **5** (2006), 1–12.
- [18] E.C. Wit and J.D. McClure, *Statistics for Microarrays: Design, Analysis and Inference*. John Wiley & Sons, 2004.
- [19] M. Zou and S.D. Conzen, A new dynamic bayesian network approach for identifying gene regulatory networks from time course microarray data, *Bioinformatics* **21** (2005), 71–79.