Known Unknowns: Novelty Detection in Condition Monitoring

John A. Quinn[1,2] and Christopher K.I. Williams[1]

[1] School of Informatics, [2] Simpson Centre for Reproductive Health
University of Edinburgh, United Kingdom
[john.quinn, c.k.i.williams]@ed.ac.uk

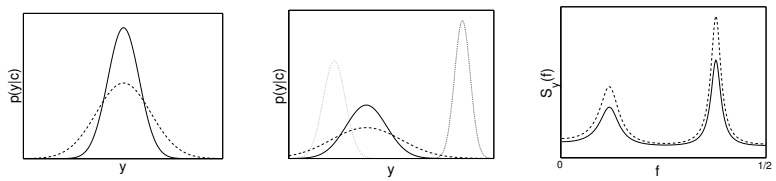
Abstract. In time-series analysis it is often assumed that observed data can be modelled as being derived from a number of regimes of dynamics, as e.g. in a Switching Kalman Filter (SKF) [1, 2]. However, it may not be possible to model all of the regimes, and in this case it can be useful to represent explicitly a 'novel' regime. We apply this idea to the Factorial Switching Kalman Filter (FSKF) by introducing an extra factor (the 'X-factor') to account for the unmodelled variation. We apply our method to physiological monitoring data from premature infants receiving intensive care, and demonstrate that the model is effective in detecting abnormal sequences of observations that are not modelled by the known regimes.

1 Introduction

In time-series analysis it is often assumed that observed data can be modelled as being derived from a number of regimes of dynamics, as e.g. in a Switching Kalman Filter (SKF) [1, 2]. However, in complex, real-world data (as found e.g. in medicine, engineering or finance) it may be that there are a very large number of possible regimes, and that a model may only have knowledge of commonly occurring ones. In this case it can be useful to represent explicitly a 'novel' regime, in order to model observations that do not correspond to any of the known regimes. The inclusion of this extra regime gives a condition monitoring system two potential benefits. Firstly, it is useful to know when novel regimes are being followed, e.g. in order to raise an alarm. Secondly, the new class provides a measure of confidence for the system. That is, by confidently classifying a regime as 'none of the above' we know that there is some structure in the data which is missing in the model.

We use the Factorial Switching Kalman Filter (FSKF), an extension of the SKF, as a general framework for condition monitoring. The FSKF has a number of factors which affect the dynamics of the observations; conditional on a particular combination of factor settings, the model is equivalent to a Kalman filter. In section 2 we extend the model of Williams et al. [3] by adding an extra factor, referred to here as the 'X-factor', representing all variation which is not normal and not similar to any of the known regimes.

In section 3 we apply our method to physiological monitoring data from premature infants receiving intensive care. This data typically has a number of



(a) (b) (c)

Fig. 1. (a) Class conditional likelihoods in a static 1D model, for the normal class (solid) and the X-factor (dashed). (b) Likelihoods of the normal class and X-factor in conjunction with other known, abnormal regimes (shown dotted). (c) The power spectral density of a latent AR(5) process with white observation noise (solid), and that of a corresponding X-factor process (dashed).

common regimes—artifactual and basic physiological patterns—as well as some uncommon regimes. Examples of the causes of uncommon regimes might be neurological problems or sepsis, or even the baby's reaction to a linen change or the flash of a camera, and include so many possibilities that it would be very difficult to model them all explicity. The model is shown to be successful in identifying clinically significant novelty in complex multivariate data.

2 Model description

As a general condition monitoring framework we use the FSKF [3, 4]. In this model, M discrete factor settings $f_t^{(1)} \dots f_t^{(M)}$ affect the hidden continuous state $\mathbf{x}_t$ and the observations $\mathbf{y}_t$. The system dynamics and observation process are taken to be dependent on the variable s_t, an index which is a cross-product of the factor settings,

$$\mathbf{x}_t \sim \mathcal{N}\left(\mathbf{A}^{(s_t)}\mathbf{x}_{t-1}, \mathbf{Q}^{(s_t)}\right), \qquad \mathbf{y}_t \sim \mathcal{N}\left(\mathbf{C}^{(s_t)}\mathbf{x}_t, \mathbf{R}^{(s_t)}\right), \tag{1}$$

so that s_t effectively 'switches' the model in and out of different dynamical regimes. Conditioned on s_t, the model is equivalent to a linear Gaussian state-space (Kalman filter). The factor settings are taken to be a priori independent and first-order Markovian.

2.1 Novel dynamics

First imagine that we have independent, one-dimensional observations which are conditionally Gaussian, $\mathbf{y}|s \sim \mathcal{N}\left(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}\right)$. For condition monitoring we are interested in problems where we assume that the possible settings of s represent a 'normal' mode and a number of known additional modes. We assume here that the normal regime is indexed by $s = 1$, and the additional known modes by

$s = 2, \ldots, K$. In this static case, we can construct a new model for unexpected data points by inflating the covariance of the normal mode, so that

$$\boldsymbol{\Sigma}^{(*)} = \xi \boldsymbol{\Sigma}^{(1)}, \qquad \boldsymbol{\mu}^{(*)} = \boldsymbol{\mu}^{(1)} \ . \tag{2}$$

where normally $\xi > 1$. This type of construction for unexpected observations is referred to as an 'X-factor'[1].

The likelihood functions for a normal class and a corresponding X-factor are shown in Figure 1(a). Clearly, data points that are far away from the normal range are more likely to be classified as belonging to the X-factor. For condition monitoring this can be used in conjunction with a number of known classes, as shown in 1(b). Here, the X-factor has the highest likelihood for regions which are far away from any known modes, as well as far away from normality.

We can generalise this approach to novelty detection by adding a new factor to a trained FSKF, with parameters based on those of the learnt normal dynamics as follows:

$$\mathbf{Q}^{(*)} = \xi \mathbf{Q}^{(1)} \ , \tag{3}$$

$$\left\{\mathbf{A}^{(*)}, \mathbf{C}^{(*)}, \mathbf{R}^{(*)}\right\} = \left\{\mathbf{A}^{(1)}, \mathbf{C}^{(1)}, \mathbf{R}^{(1)}\right\} \ , \tag{4}$$

where the switch setting $s_t = 1$ again represents normal dynamics, where no factor is active. To see why (3) and (4) are a dynamic generalisation of (2), consider the specific case of a hidden scalar AR(p) process,

$$x_t \sim \mathcal{N}\left(\sum_{k=1}^{p} \alpha_k x_{t-k}, \sigma_q^2\right), \quad y_t \sim \mathcal{N}(x_t, \sigma_r^2) \ . \tag{5}$$

The power spectral density for the hidden process x_t at frequency f is given by

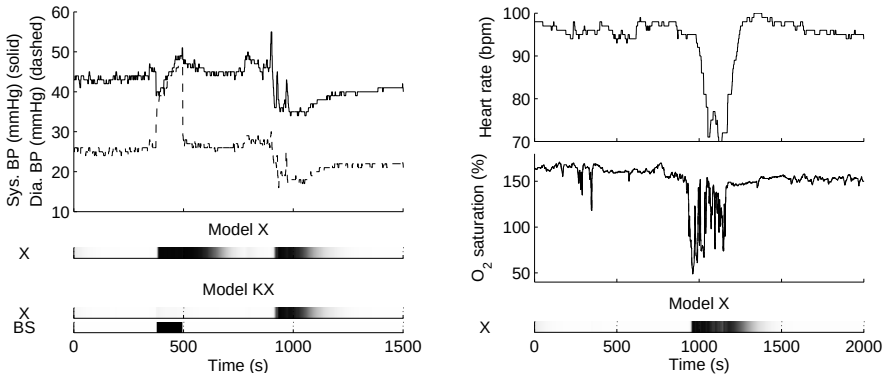
$$S_x(f) = \frac{\sigma_q^2}{\left|1 - \sum_{k=1}^{p} \alpha_k e^{-2\pi i f k}\right|^2} \ , \tag{6}$$

where $-\frac{1}{2} \leq f \leq \frac{1}{2}$, assuming one observed value per unit of time. By inflating σ_q^2 (as specified in (3)) we observe that the power is increased at each frequency. The observed process has the spectrum $S_y(f) = S_x(f) + \sigma_r^2$. As the scale of $S_y(f)$ is determined by the magnitudes of the two noise variances, inflating σ_q^2 will have the effect of increasing the power at every frequency, as illustrated in Figure 1(c).

Under an AR(p) model driven by Gaussian noise, any sequence of x's (and also the y's) are jointly Gaussian. The eigenfunctions are sinusoids and the eigenvalues are given by the power spectrum. Hence inflating the system noise has created a dynamical analogue of the static construction given above.

Preliminary experiments with the data described below showed that $\xi = 2$ was a suitable setting. It is possible to learn ξ using EM, see [5].

[1] The notation ξ is chosen by association with the word $\xi\epsilon\nu o\varsigma$, or *xenos*, meaning 'strange'.



(a) (b)

Fig. 2. Inferred durations for known factors and the X-factor. Panel (a) shows systolic and diastolic blood pressure readings containing a blood sample (BS) artifact between times 300 and 500, and physiological disturbance between times 800 and 1100. The **X** model picks up both these periods, while adding the known factors causes the artifact to be correctly reclassified. In panel (b), the X-factor picks up another significant and unusual period of variation. The minor variation around time 300 was not judged to be clinically significant, and was not picked up by the X-factor.

3 Experiments

24 hour periods of monitoring data were collected from 13 premature babies receiving intensive care in the neonatal unit at the Royal Infirmary of Edinburgh. Each period comprised of readings taken once per second: heart rate, blood pressures, core and peripheral temperatures, oxygen saturation and environmental temperature and humidity. This data was annotated by clinical experts with the times during which four common physiological and artifactual patterns occurred: bradycardia, a slowing of the heart; changes related to handling the baby; core temperature probe disconnection; blood sample artifacts; also the times in which other clinically significant patterns were apparent which were did not match any of the previous categories. FSKF models were set up with linear Gaussian state-space dynamics representing each of these patterns (for details see [5]). The thirteen data periods were split into four training cases and nine test cases. Normal dynamics were trained separately for each baby, and for this a 30 minute period for each baby was annotated as 'normal'. Approximate inference was performed in this model using the Gaussian sum method [6], to find the filtered estimates of the factor settings and hidden continuous state $p(s_t, \mathbf{x}_t|\mathbf{y}_{1:t})$. In all the following experiments, the setting $\xi = 2$ was used.

We consider three models with three different sets of factors. Model **K** contains factors representing known patterns (blood sample, bradycardia etc) only. Model **KX** contains the known factors and the X-factor. Model **X** contains one factor that switches between normality and the X-factor. Examples of the op-

Model		X-factor	Bradycardia	Core temp.	Blood sample	Handling
K	AUC	-	0.96	0.90	0.93	0.75
	EER	-	0.06	0.20	0.18	0.32
KX	AUC	0.81	0.97	0.92	0.95	0.74
	EER	0.22	0.06	0.19	0.16	0.33

Table 1. Summary statistics of performance. AUC denotes area under ROC curve and EER denotes the equal error rate.

eration of these models are shown in Figure 2. In panel (a), a period of blood pressure measurements contains a known artifactual pattern, caused by taking a blood sample, and a novel pattern of physiological disturbance. In model **X**, both patterns are picked up by the X-factor. Adding the known patterns causes the artifactual period to be correctly claimed by the appropriate factor. In panel (b), another period of physiological instability is picked up by the X-factor.

Quantitative results for the **K** and **KX** models are given in Table 1. Recall that the test data has annotations for each of the five factors (the four specific ones and the X-factor). We compare the inferred filtered probability of each factor to the relevant gold standard binary annotation. For each factor a ROC curve is plotted, and summary statistics of the area under curve (AUC) and equal error rate[2] (EER) are computed. Note that because of the periods of novel dynamics, model **K** has an incomplete factor set. The table shows that adding the X-factor allows these periods of novel dynamics to be classified while maintaining the accuracy of the known patterns. For the **X** model, any non-normal episode was annotated as belonging to the X-factor. An AUC of 0.73 and EER of 0.29 was obtained in this case.

3.1 Relation to previous work

There is a large body of work on statistical approaches to novelty detection, reviewed in [7]. In general the goal is to learn the density of training data and to raise an alarm for new data points which fall in low density areas. In a time-series context this involves modelling the next observation $p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})$ based on the earlier observations, and detecting observations that have low probability. This method is used, for example, by Ma and Perkins [8]. Such approaches define a model of normality, and look for deviations from it, e.g. by setting a threshold.

A somewhat different take is to define a broad 'outlier' distribution as well as normality, and carry out probabilistic inference to assign patterns to the normal or outlier components. For time-series data this approach was followed by Smyth [9], who considered the use of an unknown state when using a HMM for condition monitoring. This uses a similar idea to ours but in a simpler context, as in his work there is no factorial state structure and no explicit temporal model.

[2] We give error rates, so smaller numbers are better. Often 1 - error rate is given in EER tables.

Acknowledgements The authors thank Neil McIntosh and Birgit Wefers for providing expert annotation of the monitoring data. Author JQ was supported by a grant from the premature baby charity BLISS. The work was also supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. The title is inspired by a quotation from Donald Rumsfeld.

References

1. M. West and P. J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, 1997. Second edition.
2. Z. Ghahramani and G. E. Hinton. Variational Learning for Switching State-Space Models. *Neural Computation*, 12(4):963–996, 1998.
3. C.K.I. Williams, J. Quinn, and N. McIntosh. Factorial Switching Kalman Filters for Condition Monitoring in Neonatal Intensive Care. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*. MIT Press, 2006.
4. J. Ma and L. Deng. A mixed level switching dynamic system for continuous speech recognition. *Computer Speech and Language*, 18:49–65, 2004.
5. J.A. Quinn. *Bayesian Condition Monitoring in Neonatal Intensive Care*. PhD thesis, School of Informatics, University of Edinburgh, 2007.
6. D. L. Alspach and H. W. Sorenson. Nonlinear Bayesian Estimation Using Gaussian Sum Approximations. *IEEE Transactions on Automatic Control*, 17(4):439–448, 1972.
7. M. Markou and Singh S. Novelty detection: a review - part 1: statistical approaches. *Signal Processing*, 83:2481–2497, 2003.
8. J. Ma and S. Perkins. Online Novelty Detection on Temporal Sequences. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 613–618, 2003.
9. P. Smyth. Markov monitoring with unknown states. *IEEE Journal on Selected Areas in Communications*, 12(9):1600–1612, 1994.