Change detection with Kalman Filter and CUSUM

Milton Severo[1,2] and João Gama[1,3]

[1] Fac. of Economy, University of Porto, Portugal
[2] Department of Hygiene and Epidemiology, Fac. of Medicine, University of Porto, Portugal
[3] LIACC, University of Porto
`milton@med.up.pt,jgama@liacc.up.pt`

Abstract. Knowledge discovery systems are constrained by three main limited resources: time, memory and sample size. Sample size is traditionally the dominant limitation, but in many present-day data-mining applications the time and memory are the major limitations [6]. Several incremental learning algorithms have been proposed to deal with this limitations (e.g., [5, 12, 6]). However most learning algorithms, including the incremental, make the assumption that the examples are draw from stationary distribution [13]. The aim of this study is to present a detection system (DSKC) for regression problems. The system is modular and works as a post-processor of a regressor. It is composed by a regression predictor, a Kalman filter and a Cumulative Sum of Recursive Residual (CUSUM) change detector. The system continuously monitors the error of the regression model. A significant increase of the error is interpreted as a change in the distribution that generates the examples over time. When a change is detected, the actual regression model is deleted and a new one is constructed. In this paper we tested DSKC with a set of three artificial experiments, and two real-world datasets: a Physiological dataset and a clinic dataset of Sleep Apnoea. Sleep Apnoea is a common disorder characterized by periods of breathing cessation (apnoea) and periods of reduced breathing (hypopnea) [7]. This is a real-application where the goal is to detect changes in the signals that monitor breathing. The experimental results showed that the system detected changes fast and with high probability. The results also showed that the system is robust to false alarms and can be applied with efficiency to problems where the information is available over time.

1 Introduction

Knowledge discovery systems are constrained by three main limited resources: time, memory and sample size. Sample size is traditionally the dominant limitation, but in many present-day data-mining applications, time and memory are the major limitations [6]. Several incremental learning algorithms have been proposed to deal with these limitations (e.g., [5, 12, 6]). However most learning algorithms, includeding the incremental, make the assumption that the examples are drawn from stationary distribution [13].

In many practical problems arising in quality control, signal processing, monitoring in industrial plants or biomedical, the target concept may change rapidly [2]. For this reason, it is essential to construct algorithms with the purpose of detecting changes in the target concept. If we can identify abrupt changes of target concept, we can re-learn the concept using only the relevant information. There are two types of approaches to this problem: methods where the learning algorithm includes the detection mechanism, and approaches where the detection mechanism is outside (working as a wrapper) of the learning algorithm. The second approach has the advantage of being independent of the learning algorithm used. There are also several methods for solving change detection problems: time windows, weighting examples according their utility or age, etc [9]. In the machine learning community few works address this problem. In [15] a method for structural break detection is presented. The method is an intensive-computing algorithm not applicable for our proposes of processing large datasets.

The work presented here follows a time-window approach. Our focus is determining the appropriate size of the time window. We use a Kalman filter [14, 18] that smooths regression model residuals associated with a change detection CUSUM method [2, 4, 10]. The Kalman filter is widely used in aeronautics and engineering for two main purposes: for combining measurements of the same variables but from different sensors, and for combining an inexact forecast of system's state with an inexact measurement of the state [17]. When dealing with a time series of data points $x_1, x_2, ..., x_n$ a filter computes the best guess for the point x_{n+1} taking into account all previous points and provides a correction using an inexact measurement of x_{n+1}.

The next section explains the method structure of the proposed system. The experimental evaluation is presented in section 3. In this section we apply our system to estimate the airflow of a person with Sleep Apnoea. We use the on-line change detection algorithm to detect changes in the airflow. Last section presents the conclusions and lessons learned.

2 Detection System in Regression Models with Kalman Filter and CUSUM

In this paper we propose a modular detection system (DSKC) for regression problems. The general framework is shown in figure 2. The system is composed by three components: a regression learning algorithm, a Kalman filter [14] and a CUSUM [2, 4]. At each iteration, the system first component, the learning algorithm, receives one unlabeled example, x_i, and then the actual model predicts, $\hat{y}_i$. After the model forecast, it receives an input from the environment, y_i and calculates the residual $r_i = |y_i - \hat{y}_i|$. The system uses r_i and the Kalman filter error estimate of the actual model, $\hat{r}_{i-1}$, to compute a residual for the dispersion, $rd_i = |r_i - \hat{r}_{i-1}|$. The Kalman filter, the system second component, receives both residuals and updates the learning algorithm state estimate. The state estimate

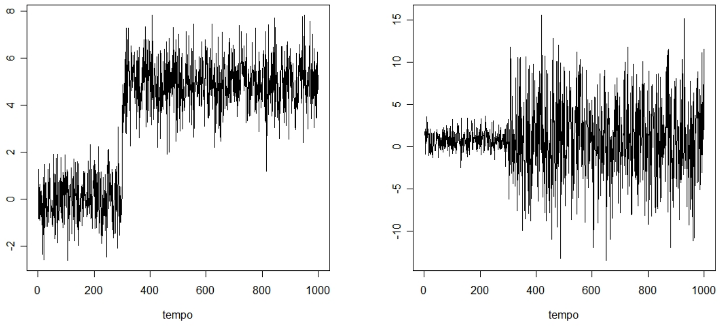

Fig. 1. Two types of changes: changes in the mean and changes in dispersion.

is form by mean error, $\hat{r}_i$, and the dispersion error, $\hat{r}d_i$. Normally, a learning algorithm will improve the predictions with the arrival of new examples, mainly in the initial learning stage. For that reason, it is very important to provide a run-time estimation of the residuals. In general, run-time estimation is provided by simple mechanism, such as auto regressive or auto regressive moving average or Kalman filter. The advantages of the last filter are: allows to adaptively tune the filter memory to faster track variations in the estimation and allows to improve the accuracy of the estimation by exploiting the state update laws and variance of the estimation [14].

The proposed system detects changes in mean error of the actual model and changes in the respective dispersion 1. The pair $(r_i, \hat{r}_i)$ is transmitted to the mean CUSUM and the pair $(rd_i, \hat{r}d_i)$ is transmitted to the dispersion CUSUM. Both CUSUM's compare the values they receive. A change occurs if significant differences between both values received or significant differences between consecutive residuals are found. If the change is an increase in the error mean or an increase in the dispersion, the system gives an order to erase the actual learning model and start to construct a new model using the new examples. If the change is a decrease in the error mean or a decrease in the dispersion error, then the system gives an order to the Kalman filter to weight the new residuals heavier, thus the filter can follow the mean error and dispersion error faster. If no significant differences are found, the new example is incorporated into the learning model.

The proposed architecture is general. It can work with any regression learning algorithm and any loss function. The main assumption is that a change is reflected in the distribution of the examples, leading to an increase of the error of the actual regression model.

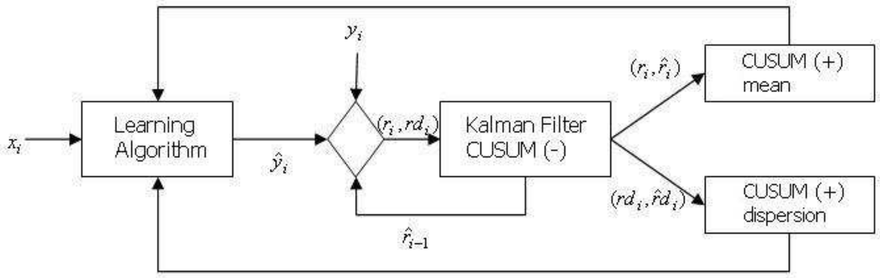

Fig. 2. Detection System with Kalman Filter and CUSUM Framework

3 Experimental Evaluation

In this section we describe the evaluation of the proposed system. We used three artificial datasets and two real-world datasets. A real-world physiological [1] dataset was used to evaluate if our DSKC provides a reliable estimate of the learning model error. We used three artificial datasets and one real-world dataset of Sleep Apnoea to evaluate the efficiency of the DSKC. An artificial data allows us to perform controlled experiments. The experiments were designed in a way that we know when the change occurs. To test the generality of the proposed methodology, we used two distinct learning algorithms: a regression tree and a linear regression model [4]. Four performance measures were used to evaluate the efficiency of the DSKC: number of false alarms (FA) and true alarms (TA), mean number of examples for detection of the changes (MNE) and normalized mean absolute error (NMAE). The median test [3] was used to compare the normalized mean absolute error with and without change detection for each dataset and learning algorithm.

3.1 Physiological Dataset

In this section we study the impact in boosting the discriminative power of a signal given by our system components: the Kalman filter and the CUSUM. We use the physiological dataset [1] that was collected using BodyMedia wearable body

---

[4] In the set of experiments reported here, we use batch versions of both algorithms that train a new model at every iteration using all examples from the last change detected or since the beginning until that moment. The focus of the paper is change detection in regression problems. It is expected that the main conclusions apply to incremental versions of the algorithms.

area under the ROC curve							
		sensor 7			**sensor 9**		
ID	age	KFC	KF	sensor	KFC	KF	sensor
1	37	0.880	0.927	0.928	0.779	0.819	0.736
2	31	0.952	0.952	0.874	0.808	0.761	0.659
2	32	0.965	0.964	0.892	0.828	0.771	0.670
4	34	0.954	0.973	0.923	0.645	0.603	0.582
4	35	0.966	0.963	0.918	0.715	0.667	0.598
5	36	0.921	0.943	0.911	0.846	0.829	0.755
5	37	0.922	0.924	0.896	0.851	0.811	0.734
6	30	0.907	0.899	0.875	0.600	0.579	0.504
6	31	0.883	0.891	0.862	0.684	0.658	0.565
9	43	0.807	0.934	0.859	0.590	0.588	0.559
14	30	0.942	0.994	0.994	0.622	0.631	0.573
19	24	0.955	0.976	0.979	0.684	0.639	0.562
20	29	0.944	0.972	0.932	0.854	0.836	0.835
25	40	0.928	0.977	0.972	0.678	0.587	0.537
26	32	0.917	0.922	0.855	0.615	0.591	0.536
32	24	0.952	0.970	0.946	0.531	0.500	0.547
Median	32	0.935	0.957	0.915	0.708	0.649	0.622

Table 1. Area under the ROC curve for the sensor, the Kalman filter estimate (KF) and Kalman filter with CUSUM estimate (KFC)

monitors. These continuous data are measures of 9 sensors and an indication of the physical activities of the user. The dataset comprises several months of data from 18 subjects. We divided the dataset by subject and age and used only the sets related to changes between sleep and awake. We measured the discrimination between sleep and awake phases using the original sensors, using only the Kalman filter estimate of a sensor, and using the estimate of the sensor from Kalman filter with CUSUM. We used the sensors 7 and 9 because they were, respectively, the sensors with the larger and the smaller discriminative power. The discrimination power was measured using the area under the ROC curve. The results show (table 1) that the discrimination power increases when we applied the Kalman filter or the Kalman filter with CUSUM to the original sensors. This fact is more evident when the sensor exhibit less discriminative power. The less discriminative sensor is sensor 9. In that case, the improvement verified with the Kalman filter plus the CUSUM is 5.9% with a p-value $p < 0.002$. The improvement decreases for sensor 7, where the Kalman filter alone has better results $(2.2\%, p < 0,018)$. These results suggest that the use of the Kalman filter with CUSUM provides a reliable estimate of the learning model error.

3.2 Artificial Datasets

The three artificial datasets used were composed by 3000 random examples. Two random changes were generated between the 30th and the 2700th examples.

1. The first dataset has five normally distributed attributes with mean 0 and standard deviation 50. The dependent variable is a linear combination of the five attributes with an white noisy with standard deviation 10. New coefficients of linear combination where built at every change. The five coefficients of the linear combination where generated by a uniform distribution over [0, 10].
2. The second dataset has two uniformly distributed attributes over [0, 2]. The dependent variable is a linear combination of the first attribute sine, and the cosine of the second attribute. We add, to each attribute, white noise with standard deviation 1. As in the previous artificial dataset, new coefficients were built at every change.
3. The third dataset is a modified version of the dataset that appear in the MARS paper [8]. This dataset has 10 independent predictor variables $x_1, ..., x_{10}$ each of which is uniformly distributed over [0,1]. The response is given by

   $$y = 10 \sin(\pi x_1 x_2) + 20 (x_3 - 0,5)^2 + 10 x_4 + 5 x_5 + \epsilon \cdot \qquad (1)$$

   A permutation of the predictor variables was made at each change.

For each type of dataset we randomly generated 10 datasets.

3.3 Results on Artificial Datasets

Tables 2, 3 and 4 show the results for respectively, dataset one, two and three. We can observe that DSKC is effective with all learning algorithms. Overall we detected 73% ($CI95\% = [64\% - 81\%]$) of all true changes with no false alarms ($CI95\% = [0\% - 6\%]$). The results show that the proportion of true changes varies between 50% for the third dataset with linear regression and 90% for the same dataset but with regression trees; the mean number examples needed for detection varies from 8.3 to 42.13. We found significant differences ($p < 0.05$) between the use and not use of our detection system for the normalized mean absolute error, except for the 2nd dataset with linear regression model. The mean normalized error decreased for all datasets with the use of our DSKC. We observed that when the second change occurs relatively closed to the first change or when the first change occurs relatively closed to the beginning of the experience, the change was not detected. As we can see in table 5, the proportion of changes detected was 25%, when the number of examples between changes is less than 332, against 89%, when there are more than 532 examples. The association between the number of examples required by the learning algorithm and detection or not detection of the change is significant ($p < 0.001$).

3.4 Sleep Apnoea Dataset

After measuring the performance of our detection system in the artificial datasets, we evaluate the performance in a real problem where change points and change rates are not known. For such we applied our system to a dataset from patients

	regression trees					**linear models**				
	without det.	**with det.**				**without det.**	**with det.**			
	NMAE	NMAE	TA	FA	MNE	NMAE	NMAE	TA	FA	MNE
1	0.75	0.71	1	0	45.0	0.30	0.10	2	0	3.0
2	0.73	0.62	2	0	38.0	0.25	0.11	1	0	9.0
3	0.85	0.66	2	0	27.5	0.51	0.11	2	0	5.0
4	0.68	0.67	1	0	16.0	0.29	0.12	1	0	4.0
5	0.66	0.63	2	0	45.0	0.40	0.13	2	0	19.5
6	0.68	0.64	2	0	40.5	0.31	0.10	2	0	2.0
7	0.79	0.57	2	0	9.5	0.30	0.21	1	0	6.0
8	0.73	0.59	1	0	51.0	0.28	0.08	2	0	26.5
9	0.73	0.69	1	0	43.0	0.22	0.08	2	0	4.5
10	0.84	0.76	1	0	10.0	0.38	0.09	2	0	3.5
η	0.73	0.65	2	0	39.2	0.30	0.11	2	0	4.8
$\bar{x}$	0.74	0.65	1.5	0	32.6	0.32	0.11	1.7	0	8.3

Table 2. Results for the 1st artificial dataset.

	regression trees					**linear models**				
	without det.	**with det.**				**without det.**	**with det.**			
	NMAE	NMAE	TA	FA	MNE	NMAE	NMAE	TA	FA	MNE
1	0.57	0.56	1	0	8.0	0.85	0.82	1	0	23.0
2	0.45	0.33	1	0	5.0	0.88	0.88	2	0	48.0
3	0.45	0.42	1	0	78.0	0.89	0.89	1	0	35.0
4	0.53	0.47	2	0	10.5	0.81	0.82	2	0	95.0
5	0.39	0.37	1	0	13.0	0.90	0.90	2	0	84.5
6	0.61	0.36	2	0	9.0	0.83	0.83	0	0	–
7	0.48	0.39	1	0	39.0	1.00	1.00	1	0	5.0
8	0.54	0.45	2	0	8.0	1.00	1.00	2	0	89.5
9	0.45	0.41	2	0	10.0	0.84	0.85	1	0	27.0
10	0.45	0.41	2	1	33.5	0.87	0.87	1	0	83.0
η	0.47	0.41	1.5	0	10.2	0.87	0.87	1.0	0	48.0
$\bar{x}$	0.49	0.42	1.5	0	21.4	0.89	0.88	1.3	0	54.4

Table 3. Results for the 2nd artificial dataset.

with Sleep Apnoea. Sleep Apnoea is a common disorder characterized by periods of breathing cessation (apnoea) and periods of reduced breathing (hyponea). The standard approach to diagnoses apnoea consists of monitoring a wide range of signals (airflow, snoring, oxygen saturation, heart rate...) during patient sleep. There are several methods for quantifying the severity of the disorder, such as measuring the number of Apnoeas and Hypopnoea per hour of sleep or measuring the number of breathing events per hour. There is a heterogeneity of methods for defining abnormal breathing events, such as reduction in airflow or oxygen saturation or snoring [7]. It can be seen as pathological, when the number of

	regression trees					**linear models**				
	without det.	**with det.**				**without det.**	**with det.**			
	NMAE	NMAE	VA	FA	MNE	NMAE	NMAE	VA	FA	MNE
1	0.80	0.67	2	0	21.5	0.71	0.59	1	0	38.0
2	0.73	0.69	1	0	33.0	0.73	0.73	0	0	–
3	0.84	0.66	2	0	23.0	0.68	0.57	1	0	65.0
4	0.86	0.67	2	0	28.0	0.71	0.58	2	0	55.0
5	0.82	0.66	2	0	33.0	0.68	0.57	1	0	19.0
6	0.71	0.68	1	0	14.0	0.75	0.59	2	0	54.5
7	0.86	0.68	2	0	39.0	0.69	0.59	1	0	25.0
8	0.80	0.66	2	0	50.5	0.63	0.60	1	0	41.0
9	0.87	0.68	2	0	20.5	0.88	0.88	0	0	–
10	0.82	0.68	2	0	17.5	0.67	0.59	1	0	39.0
η	0.82	0.67	2	0	25.5	0.70	0.59	1.0	0	40.0
$\bar{x}$	0.81	0.67	1.8	0	28.0	0.63	0.62	1.0	0	42.1

Table 4. Results for the 3rd artificial dataset.

number e.g.	not detected	detected	p
$[1 - 332[$	9 (75.0%)	3 (25.0%)	< 0.001
$[332 - 532[$	11 (45.8%)	13 (54.2%)	
$[532 - \infty[$	8 (11.1%)	64 (88.9%)	

Table 5. Association between the number the examples read and the detection or not of the change

Apnoeas and Hypopnoea/hour is larger then 20 events per hour [11]. Our goal in this experiment was to evaluate if our detection system could detect abnormal breathing events.

The real-world dataset was a set of sleep signals from a patient with Sleep Apnoea. Three of the 7 signals (airflow, abdominal movement signals and snoring) had 16Hz frequency and the other 4 signals (heart rate, light, oxygen saturation and body) had 1Hz frequency. All signal with 16Hz were transform in 1Hz using the mean for each second. The dataset contained 26102 records from 7 signals, which is approximated 7.5 hours of sleep.

3.5 Results on Sleep Apnoea Dataset

Taking into consideration the problem, we built a model to predict the airflow using all other signals as predictor variables. In this dataset, the regression model is evaluated using the normalized mean absolute error statistic. We did not have any indication where the abnormal breathing event would occur. For this reason, we could not evaluate the number of false alarms and true alarms and the mean number of examples for detection of the changes.

We used two regression models: a generalised linear regression and a regression tree[5]. Both learning algorithms employed exhibited slight better results using the detection mechanism than without (table 6). The total number of alarms (TA) were 18 and 17, respectively, for the regression tree and linear regression models. In order to evaluate the agreement between both learning algorithms to detected abnormal breathing events, we compared the alarms proximity between them. We detected 13 pairs of alarms, specifically, 13 alarms detected by the regression tree model occured in the proximity of 13 alarms detected by the linear regression model (table 7). We carried out a second experience off-line, where:

	regression trees			**linear model**		
	without det.	**with det.**		**without det.**	**with det.**	
	NMAE	NMAE	TA	NMAE	NMAE	TA
1	0.940	0.923	18	0.981	0.974	17

Table 6. Results for dataset of Sleep Apnoea.

1. A regression tree is applied to predict airflow filter by the Kalman Filter with CUSUM.
2. A regression tree is applied to predict airflow dispersion filter by the Kalman Filter with CUSUM.
3. The cut points of the regressions trees leafs and the alarms detected by our DSKC were compared (figure 3).

As shown in table 7, there were 7 increases detected in time in the airflow dispersion and all of them were detected at least by one of the learning models applied to Sleep Apnoea Dataset. There were 4 increases detected in time in the airflow, and 2 of them by one of the learning models applied. Despite both learning algorithms investigated exhibited slight better results using the detection mechanism than without, the alarms detected by both models seems to show agreement, which may imply that we have detected true changes in the distribution of examples in the Sleep Apnoea Dataset.

4 Conclusions

In this paper we discussed the problem of maintaining accurate regression models in dynamic, non-stationary environments. The system continuously monitors the residual of the regression algorithm, looking for changes in the mean and changes in the variance. The proposed method maintains a regression model where residuals are filtered by a Kalman filter. A CUSUM algorithm continuously monitors

---

[5] The GLM and CART versions implemented in [16].

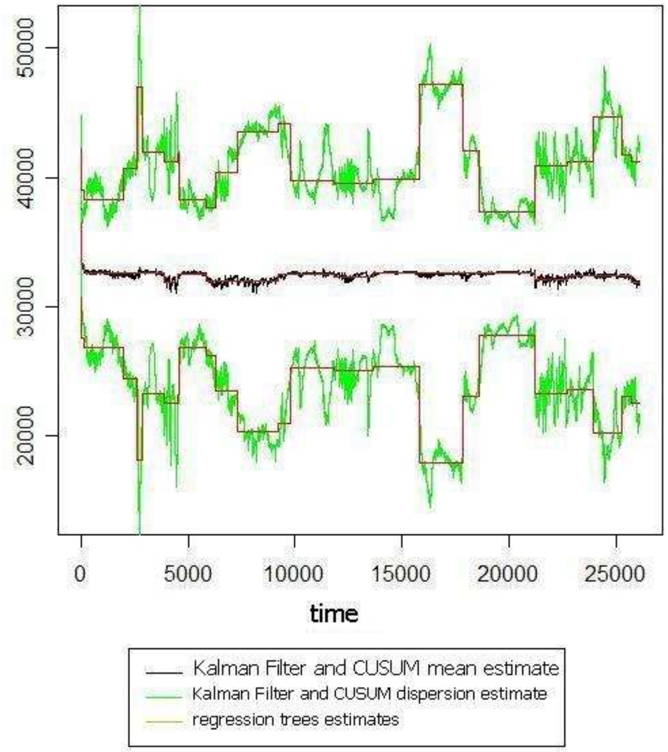

Fig. 3. The regression tree segmentation of the airflow and airflow dispersion filter by Kalman Filter and CUSUM.

significant changes in the output of the Kalman filter. The CUSUM works as a wrapper over the learning algorithm (the regression model plus the Kalman filter), monitoring the residuals of the actual regression model. If CUSUM detects an increase of the error, a new regression model is learned using only the most recent examples. As shown in the experimental section the Kalman filter application to the residuals gives a good on-line estimation of the learning algorithm state. The results of the method for change detection in regression problems show that it can be applied with efficiency when the information is available sequentially over time. An advantage of the proposed method is that it is independent of the learning algorithm. The results of the change detection algorithm mainly depend on the efficiency of the learning algorithm. They also show, that the Kalman filter has a good performance in detecting real changes from noisy data.

Acknowledgments:

This work was developed in the context of project ALES II (POSI/EIA/55340/2004). Thanks to Andre Carvalho for several useful comments.

References

1. D. Andre and P. Stone. Physiological data modeling contest. A ICML'04 Workshop, 2004.
2. Michele Basseville and Igor Nikiforov. *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, 1993.
3. G. Bhattacharyya and R. Johnson. *Statistical Concepts and Methods*. New York, John Willey & Sons, 1977.
4. Giuseppe Bianchi and Ilenia Tinnirello. Kalman filter estimation of the number of competing terminals in ieee. The 22nd Annual Joint Conference of IEEE Computer and Communications, 2003.
5. Cauwenberghs, Gert, Poggio, and Tomaso. Incremental and decremental support vector machine learning. *In Advances in Neural Information Processing Systems*, 13, 2001.
6. P. Domingos and G. Hulten. Mining high-speed data streams. In *Knowledge Discovery and Data Mining*, pages 71–80, 2000.
7. W. Ward Flemons, Michael R. Littner, James A. Rowley, W. McDowell Anderson Peter Gay, David W. Hudgel, R. Douglas McEvoy, and Daniel I. Loube. Home diagnosis of sleep apnoeas: A systematic review of the literature. *Chest*, 1543-1579:211–237, 2003.
8. J. Friedman. Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1–141, 1991.
9. J. Gama, P. Medas, and G. Castillo. Learning with drift detection. pages 286–295. Brazilian AI Conference, Springer Verlag, 2004.
10. Eugene Grant and Richard Leavenworth. *Statistical Quality Control*. McGraw-Hill, 1996.
11. G. Guimarães, J.H. Peter, T. Penzel, and A Ultsch. A method for automated temporal knowledge acquisition applied to sleep-related breathing disorders. *Artificial Intelligence in Medicine*, 23:211–237, 2001.
12. C. M. Higgins and R. M. Goodman. Incremental learning using rule-based neural networks. In *International Joint Conference on Neural Networks*, pages 875–880. Seattle, WA, 1991.
13. G. Hulten, L. Spencer, and P. Domingos. Mining time-changing data streams. Proc. of KDD2001, ACM Press, 2001.
14. R. E. Kalman. A new approach to linear filtering and prediction problems. *Transaction of ASME - Journal of Basic Engineering*, pages 35–45, mar 1960.
15. Kwok Pan Pang and Kai Ming Ting. Improving the centered cusums statistic for structural break detection in time series. In *Proc. 17th Australian Join Conference on Artificial Intelligence*. LNCS, Springer Verlag, 2004.
16. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2005. ISBN 3-900051-07-0.
17. Raul Rojas. The Kalman filter. Technical report, Freie University of Berlin, 2003.
18. Greg Welch and Gary Bishop. An introduction to the Kalman filter. Technical report, 95-041, Department of Computer Science, Department of Computer Science, University of North Caroline at Chapel Hill, Apr 2004.

	reg. trees	**linear mod.**	**tree off-line**	**type of change**
1	2017	1965	1962.5	disp,+
2	2600	2682	2631.5	disp,+
3	—	—	2851.5	disp,-
4			3864.5	mean,-
5	4172	4135	—	—
6	—	—	4551.5	mean,+
7	—	—	4595.5	disp,-
8	5835	—	—	—
9	—	—	5875.5	mean,-
10	—	6165	6324.5	disp,+
11	—	7415	7322.5	disp,+
12	—	9287	9202.5	mean,+
13	—	—	9764.5	disp,-
14	10207	10211	—	—
15	11106	11112	—	—
16	11531	—	—	—
17	—	—	11793.5	mean,-
18	12318	12452	—	—
19	13404	13396	13632.5	mean,+
20	14686	14927	—	—
21	15848	15802	15808.5	disp,+
22	—	17762	—	—
23	—	—	17833.5	disp,-
24	18046	—	—	—
25	—	—	18609.5	disp,-
26	20456	20463	—	—
27	—	—	21207.5	mean,-
28	21216	21280	21222.5	disp,+
29	22505	22253	—	—
30	—	—	22743.5	mean,+
31	23139	—	—	—
32	24018	—	23961.5	disp,+
33	24581	24400	—	—
34	—	—	25290.5	disp,-
35	—	—	25733.5	mean,-

Table 7. Alarms detected by DSKC with regression trees and linear models in sleep Apnoea dataset and the leafs cut point of regression tree applied off-line