Using Support Vector Machine model for fault detection along a water canal

José Duarte[1]
d10401@alunos.uevora.pt
Luís Rato[1]
lmr@uevora.pt
Manuel Rijo[2]
rijo@uevora.pt

[1] Departamento de Informática, Escola de Ciências e Tecnologia, Centro de Inovação em Tecnologias da Informação, Universidade de Évora, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal

[2] Departamento de Engenharia Rural, Escola de Ciências e Tecnologia, Núcleo de Hidráulica e Controlo de Canais, Universidade de Évora, Núcleo da Mitra, Apartado 94, 7002-554 Évora, Portugal

Abstract

This paper reports a work in progress, the training of a Support Vector Machine model to detect faults in an experimental water supply canal. The work took place at the experimental canal of Núcleo de Hidráulica e Controlo de Canais at the Universidade de Évora. The main objective is to identify faults in the water depth sensors and to detect unauthorized water withdrawals using pattern recognition. The preliminary accuracy tests, in and out of sample, have shown an accuracy over 90% to identify 28 different patterns.

1 Introduction

Fault Detection and Isolation (FDI) is a research field where knowlegde--based models have been used with some success. Artificial Neural Networks and Fuzzy Systems are often used but its dependency of large amounts of training data and the slow convergence speed leads them to lose ground to models mathematical-based [13].

Support Vector Machines (SVMs) are been used as classification tool with a huge success in research areas like computer vision [7], health [9] and entertainment [4]. Lately, many studies propose the use of SVM in FDI problems [2, 12, 13].

This paper reports the work, still in progress, of the creation of FDI model for an experimental water supply canal using a SVM.

1.1 Presentation of the water canal

The work took place at the experimental canal of Núcleo de Hidráulica e Controlo de Canais (NuHCC) [10] at the Universidade de Évora.It is a canal with 145m of length and it is divided in four pools. Each one separated by an undershot gate with an overshot gate at the canal end, as can be seen in Figure 1. Sensors installed at the upstream, center and downstream allows to monitor the water depth in each pool. The maximum depth is 900mm, is the equivalent to the height of the pools. There is an offtake valve at upstream of every gates that allows to implement water withdrawals. An electric MONOVAR valve controls the canal inlet with a maximum design flow of $0.09m^3s^{-1}$.

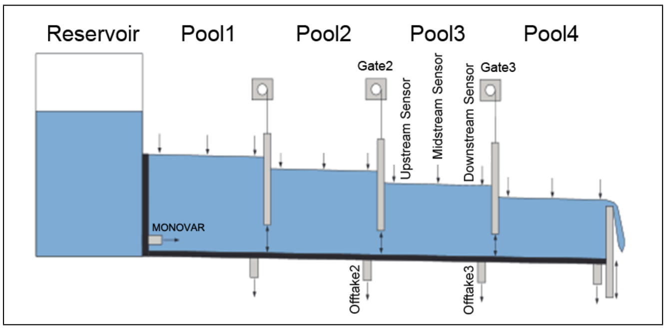

Figure 1: Diagram of the experimental water canal (adapted from[8])

The NuHCC facility is monitored and controlled by a MODBUS/Serial network of six Programmable Logic Controllers (PLCs) and a Supervisory Control and Data Acquisition (SCADA) system. A multi-platform controller interface [3] was used to interact with SCADA and collect data. Further details about this canal can be read in [8].

1.2 Objectives

At this moment, the focus is only in the third pool of the NuHCC canal. In Figure 2, it is possible to identify, from upstream to downstream, all the elements of interest to the study. Two offtakes in the left, two gates and water stream in the center and in the right the three water depth sensors.

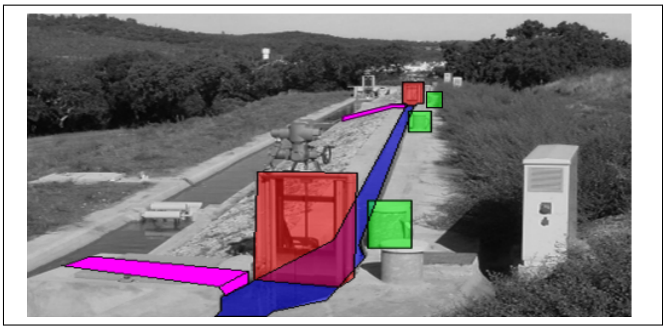

Figure 2: The third pool of the canal (Red - Gates 2 and 3; Pink - Offtakes 2 and 3; Green - Sensores Up, Center and Downstream).

The water depth sensors measure values between 0 and 900mm with an reading error of 0.005mm. The opening values of the offtake are measured in percentage, ie. from 0% to 100%. The height of the three first gates varies from 0 to 800mm, the last one from 0 to 700mm.

The main objective is to identify faults in the water depth sensors and to detect unauthorized water withdrawals using pattern recognition.

2 Preparing the data

The SVM is a popular and powerful classification technique. Considered easier to understand than Neural Networks, but users less familiar with it often have problems to get satisfactory results [6]. Classify the existing data is often the first task to create a model using a SVM. In this case. due to the nature of the problem, the first step was to collect the data.

Tests were made in order to collect the more significant data as possible. Tests like filling and emptying the pool were made and all readings of sensors and actuator recorded. The two offtake valves were used to simulate water uptake upstream or downstream of the pool. The sample time used was 1s.

2.1 Supervising the samples

The model should be able to classify correctly any given sample. To allow that to happen, every instance in the training set should be correctly classified. The model it is able to identify 28 different canal states. This means that there are at least 28 recognizable patterns (Table 1) provided by the training set.

2.2 Selecting the features

One problem detected in these kind of water depth sensors is inconsistent readings. It is an oscillatory behavior between the real value and a shift of it. Though temporary, this kind of error can lead of misclassifications. A threshold of 10mm has been defined in order to identify this error state. An instance of the training set that violates this condition it is classified as faulty sample. The resulting 26 labels are described in Table 1.

# of classes	Description
One	Nominal state
One	Loss of water
Three	Positive shift in one sensor
Three	Negative shift in one sensor
Three	Positive shift in two sensors
Three	Negative shift in two sensors
Six	Alternated shifts in two sensors
One	Positive shifts in three sensors
One	Negative shifts in three sensors
Six	Alternated shifts in three sensors

Table 1: Description of all classes

Selecting the right features for a SVM it is not a easy task. In fact, many studies were made about that matter [1, 5, 11]. The used features were kept to a minimum without compromising the performance. Just six features are used, the reading of the three depth sensors present in the pool and the difference between the actual and the previous reading. A further look into this difference (1) it is possible to understand it as the slope of the reading values along the time , for a sample time of $1s$.

$$\Delta sensor_i = \frac{sensor_i - sensor_{(i-1)}}{t_i - t_{(i-1)}} \quad (1)$$

An instance of the training set it is described in (2).

$$\overbrace{class}^{\text{Label}} \ \overbrace{sensor_1 \ \ sensor_2 \ \ sensor_3 \ \ \Delta sensor_1 \ \ \Delta sensor_2 \ \ \Delta sensor_3}^{\text{Six features}} \quad (2)$$

3 Training and testing the model

Experimental tests were made in the canal in order to cover several nominal and faulty states. Seven data independent sets were created. One was used to train the model, the six others to test it. The nominal values of the test sets can been seen in Table 2. A linear kernel with the penalty parameter of the error term equal to 1 (ie. $C = 1$) was used.

Test #	MONOVAR (l/s)	Gates (mm)	Water level (mm)	Offtakes $(\%)$
1	$25 \rightarrow 25$	$0 \rightarrow 0$ (G3)	$0 \nearrow 800$	none
2	$25 \rightarrow 25$	$0 \rightarrow 0$ (G3)	$0 \nearrow 800$	none
3	$25 \searrow 0$	$400 \rightarrow 400$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT2)
4	$25 \searrow 0$	$400 \rightarrow 400$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT3)
5	$25 \searrow 0$	$400 \searrow 0$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT2)
6	$25 \searrow 0$	$400 \searrow 0$ (G4)	$400 \searrow 0$	$0 \nearrow 50$ (OT3)

Table 2: Nominal values for the six test sets

3.1 Faults detected and results

Three kinds of tests were made to measure the model accuracy. *In sample* test where the same set is used to train and test the model. The *10-Fold Cross-validation* technique has been used to estimate accuracy of the model. And then, six test sets were used to measure the *out of sample* accuracy. All tests have shown an accuracy over 90%. The results are presented in Table 3.

4 Conclusions and future perspectives

Besides the good results of the model, much work has to be done in order to create an online FDI system. Scaling the values of the features it is always important to avoid numerical problems during the calculation [6]. In this case it revealed to be crucial. The values of the features resulting from (1) are ten or a hundred times smaller than the others ones. The values with a greater range were dominating the smaller ones, leading to worse results.

The model still need more training in order to detect these and more faults in real time. Other kernels and its parameters must be studied. Techniques like leave-one-out cross-validation should be used in order to understand if more features are needed in the sets.

In sample accuracy
Accuracy = 100% (1038/1038) (classification)
Out of sample accuracy
Train set 1 - Accuracy = 99.3506% (918/924) (classification)
Train set 2 - Accuracy = 99.3506% (918/924) (classification)
Train set 3 - Accuracy = 98.1722% (1665/1696) (classification)
Train set 4 - Accuracy = 99.0538% (2303/2325) (classification)
Train set 5 - Accuracy = 90.8019% (385/424) (classification)
Train set 6 - Accuracy = 90.0463% (389/432) (classification)
All sets - Accuracy = 97.8141% (6578/6725) (classification)
10-fold cross-validation
Cross Validation Accuracy = 99.3256%

Table 3: Results of the performance tests

5 Acknowledgements

This work was financially supported by the Project FCT - ORCHESTRA (PTDC/EMS-CRO/2042/2012) of the Fundação para a Ciência e Tecnologia, Portugal.

References

[1] Y.-W. Chen and C.-J. Lin. Combining svms with various feature selection strategies. In *Feature extraction*, pages 315–324. Springer, 2006.

[2] D. Dehestani, F. Eftekhari, Y. Guo, S. Ling, S. Su, and H. Nguyen. Online support vector machine application for model based fault detection and isolation of hvac system. In *International Journal of Machine Learning and Computing*, volume 1, pages 1–7, 2011.

[3] J. Duarte, L. Rato, P. Shirley, and M. Rijo. Multi-platform controller interface for scada application. In *Proceedings of 18th IFAC World Congress (IFAC2011)*, volume 18, pages 7885–7890. IFAC, Milan, Italy, 2011.

[4] E. Eyjolfsdottir, G. Tilak, and N. Li. Moviegen: A movie recommendation system. Technical report, Computer Science Department, University of California Santa Barbara, 2010.

[5] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. In *The Journal of Machine Learning Research*, volume 3, pages 1157–1182. JMLR. org, 2003.

[6] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003.

[7] T. Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.

[8] J. Lemos, F. Machado, N. Nogueira, L. Rato, and M. Rijo. Adaptive and non-adaptive model predictive control of an irrigation channel. In *Networks and Heterogeneous Media*, volume 4, pages 303–324. American Institute of Mathematical Sciences, 2009.

[9] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. Mesirov, et al. Multiclass cancer diagnosis using tumor gene expression signatures. In *Proceedings of the National Academy of Sciences*, volume 98(26), pages 15149–15154. National Acad Sciences, 2001.

[10] M. Rijo. Local automatic control modes in an experimental canal. In *Irrigation and Drainage System*, volume 17, pages 179–193, 2003.

[11] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik. Feature selection for svms. In *NIPS*, volume 12, pages 668–674, 2000.

[12] W. Yin, W. Zhang, and X. Sun. A svm-based multiple faults classification scheme design in flight control fdi system. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, pages 187–187. IEEE, 2007.

[13] J. Zeng, D. Lu, Y. Zhao, Z. Zhang, W. Qiao, and X. Gong. Wind turbine fault detection and isolation using support vector machine and a residual-based method. In *American Control Conference (ACC), 2013*, pages 3661–3666. IEEE, 2013.