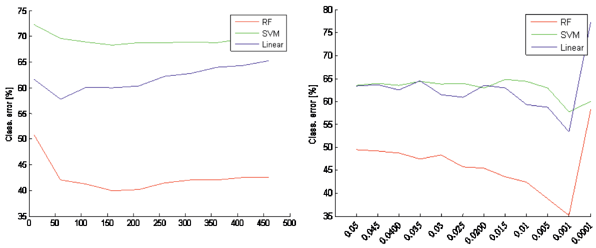

Figure 6: Classification error, Lasso and RLS[1]

We have decided to use supervised learning to evaluate whether parameters of our models (both from LASSO and RLS estimators) can be useful for detecting faults. We tried three different classifiers: linear regression, support vector machine (SVM) and random forest. Both the forgetting factor (for RLS) and the number of data slices (for LASSO) are parameters for tuning. We have found that less slices and larger forgetting factor gives better signal to noise ratio and a more robust solution. However, they are a lot less sensitive to faults that are only apparent under certain conditions. This is due to the smoothing larger slices and forgetting factor results in. As an example, a partially clogged air filter will only have a visible effect if the engine is running at high power, since this is the only situation when a large air flow is required.

We have run the classification task a number of times, varying the time slice size the and forgetting factor. It is easy to see from Figure 6 that choosing too small forgetting factor for RLS is detrimental. On the other hand, the effect of choosing too many data slices is hardly visible.

In general, the random forest classifier outperforms both SVM and linear classifier by a pretty large margin. We do not know why this is the case, since we have not investigated the classification itself in great depth. More interestingly, RLS estimator appears to give slightly better results than the LASSO estimator, but it probably is not worth the increased computational complexity.

As a final comment, the resulting classification error appears to be rather high, but it is important to take into account that this data set is a very difficult one. There is a lot of different external influences that disturb the "normal" operation of a truck, and the low quality of available sensors result in high levels of noise in the data. The lack of dedicated sensors is also a problem: neither of the four faults we have analysed is being monitored in any way for current in-production vehicles.

7 Conclusions

In this paper we present a project that we are involved in, developing an unsupervised algorithm for discovering interesting relations between time series of vehicle signal data, to be used for fault detection and predictive maintenance. We present our approach and show initial evaluation, using supervised learning, on the data collected from a Volvo truck during a fault injection experiment.

This is a step towards a system that would be able to analyse on-board data on real vehicles and detect anomalies in an autonomous way. Ideas presented here are very much work in progress and there are numerous directions to extend those results. Primarily, we have not really answered the question of how to distinguish "interesting" relations from "uninteresting" ones, especially taking into account that we are looking for those that hold *most*, but definitely not *all*, of the time.

It is also not quite clear if the supervised classification is the best way of evaluating usefulness of discovered relations. We intend to explore other possibilities, especially those connected to the service records database we have access to.

Acknowledgements. This work has been partially funded by grants from VINNOVA and from the Knowledge Foundation.

References

[1] S. Byttner, T. Rögnvaldsson, and M. Svensson. Consensus self-organized models for fault detection (COSMO). *Engineering Applications of Artificial Intelligence*, 24(5):833–839, 2011.

[2] S.H. D'Silva. Diagnostics based on the statistical correlation of sensors. Technical Report 2008-01-0129, Society of Automotive Engineers (SAE), 2008.

[3] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.

[4] H. Kargupta et al. VEDAS: A mobile and distributed data stream mining system for real-time vehicle monitoring. In *Int. SIAM Data Mining Conference*, 2003.

[5] J. Lacaille and E. Come. Visual mining and statistics for turbofan engine fleet. In *IEEE Aerospace Conf.*, 2011.

[6] Rune Prytz, Sławomir Nowaczyk, and Stefan Byttner. Towards relation discovery for diagnostics. In *Proceedings of the First International Workshop on Data Mining for Service and Maintenance*, pages 23–27, 2011.

[7] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58, 1996.

[8] G. Vachkov. Intelligent data analysis for performance evaluation and fault diagnosis in complex systems. In *IEEE International Conference on Fuzzy Systems*, pages 6322–6329, July 2006.

[9] Y. Zhang, G.W. Gantt, et al. Connected vehicle diagnostics and prognostics, concept, and initial practice. *IEEE Transactions on Reliability*, 58(2), 2009.