

was developed, one based on a least-squares solution, the other based on nonlinear optimization.

6 Conclusions

In this paper we proposed a method to predict component failures based on system logs. The results show, that such a type of prediction indeed is possible. However, the effort to reach this goal is high. For every type of failure a separate classifier has to be trained, although the results of experiment 4 (cf. Section 4.4) show that a combined classification is possible. Nevertheless, such a combination still is prone to result in a reduced classification performance compared to treating each problem separately. Also, and perhaps most importantly, the requirement for methods such as proposed in this paper, namely data quality is not yet met. To assure a consistent data environment for the training, the data needs to be complete for the whole fleet. The data used for this research had huge gaps in the recorded timespan. This is a problem *DB Schenker Rail* has to solve, before methods like this can be used in a real-world environment. We also assume that the prediction quality will significantly increase given the data quality is improved.

Also, the problem of imbalanced classes is certainly present in domains such as failure prediction as usually cases of failure are quite rare compared to the regular cases where everything is fine. The results show that the *RandomForest* classifier seems to work well on imbalanced data, but, if tackled appropriately the performance will even increase.

For future work it is planned to carefully tune the parameters of the machine learning algorithms for each single problem. Also, given the data quality is enhanced, the approach has to be re-run on the new data. Another interesting topic is to inspect the effects of the values v and w also for the other experiments and figure out whether or not similar trends are present.

References

1. Fulp, E.W., Fink, G.A., & Haack, J.N., Predicting Computer System Failures Using Support Vector Machines. *WASL'08 Proceedings of the First USENIX conference on Analysis of system logs*, 2008.
2. Guo, P., & Bai, N., Wind Turbine Gearbox Condition Monitoring With AAKR And Moving Window Statistic Methods. *Energies* 2011, 4, 2077-2093., 2011.
3. Kauschke, S., Nutzung Bahnbezogener Sensordaten Zur Vorhersage Von Wartungszyklen, Diploma Thesis, TU Darmstadt, Knowledge Engineering Group, http://www.ke.tu-darmstadt.de/lehre/arbeiten/diplom/2014/Kauschke_Sebastian.pdf, 5 2014.
4. Létourneau, S., Famili, F. & Matwin, S., Data Mining For Prediction of Aircraft Component Replacement. *IEEE Intelligent Systems Jr., Special Issue on Data Mining*, p. 59-66, 1999.
5. Lipowsky, H., *Entwicklung Und Demonstration Eines Integrierten Systems Zur Zustandsüberwachung Von Gasturbinen*, Phd. Thesis, Stuttgart, Institut für Luftfahrtantriebe, 2010.
6. Swets, J., *Signal Detection And Recognition By Human Observers*, New York, Wiley, 1964.
7. Witten, I.H., Frank, E., & Hall, M., *Data Mining - Practical Machine Learning Tools And Techniques*. Burlington: Morgan Kaufmann Publishers, 3rd edition, 2011.