

<https://doi.org/10.21122/2227-1031-7448-2019-18-6-519-524>

UDC 349

Signal Pre-Selection for Monitoring and Prediction of Vehicle Powertrain Component Aging

A. Udo Sass¹⁾, E. Esatbeyoglu¹⁾, T. Iwwerks¹⁾

¹⁾Volkswagen AG (Wolfsburg, Federal Republic of Germany)

© Белорусский национальный технический университет, 2019
Belarusian National Technical University, 2019

Abstract. Predictive maintenance has become important for avoiding unplanned downtime of modern vehicles. With increasing functionality the exchanged data between Electronic Control Units (ECU) grows simultaneously rapidly. A large number of in-vehicle signals are provided for monitoring an aging process. Various components of a vehicle age due to their usage. This component aging is only visible in a certain number of in-vehicle signals. In this work, we present a signal selection method for in-vehicle signals in order to determine relevant signals to monitor and predict powertrain component aging of vehicles. Our application considers the aging of powertrain components with respect to clogging of structural components. We measure the component aging process in certain time intervals. Owing to this, unevenly spaced time series data is pre-processed to generate comparable in-vehicle data. First, we aggregate the data in certain intervals. Thus, the dynamic in-vehicle database is reduced which enables us to analyze the signals more efficiently. Secondly, we implement machine learning algorithms to generate a digital model of the measured aging process. With the help of Local Interpretable Model-Agnostic Explanations (LIME) the model gets interpretable. This allows us to extract the most relevant signals and to reduce the amount of processed data. Our results show that a certain number of in-vehicle signals are sufficient for predicting the aging process of the considered structural component. Consequently, our approach allows to reduce data transmission of in-vehicle signals with the goal of predictive maintenance.

Keywords: predictive maintenance, feature extraction, signal selection, time series, machine learning, model explanation

For citation: Udo Sass A., Esatbeyoglu E., Iwwerks T. (2019) Signal Pre-Selection for Monitoring and Prediction of Vehicle Powertrain Component Aging. *Science and Technique*. 18 (6), 519–524. <https://doi.org/10.21122/2227-1031-7448-2019-18-6-519-524>

Предварительный выбор сигнала для мониторинга и прогнозирования старения компонентов силовой передачи автомобиля

А. Удо Сасс¹⁾, Э. Эсатбейоглу¹⁾, Т. Ивверкс¹⁾

¹⁾АО «Фольксваген» (Вольфсбург, Федеративная Республика Германия)

Реферат. Прогнозное техническое обслуживание является важным для предотвращения незапланированных простоев современных транспортных средств. С расширением функциональности одновременно происходит быстрый рост обмена данными между электронными блоками управления. Большое количество бортовых сигналов позволяет осуществлять мониторинг процесса старения. Старение компонентов автомобиля зависит от того, как они используются. Элементы старения выявляются благодаря наличию ряда бортовых сигналов. В данной статье предложен метод выбора бортовых сигналов с целью определения соответствующих для проведения мониторинга и прогнозирования старения компонентов силовой передачи транспортных средств. Процесс старения рассматривается на основе степени засорения конструктивных элементов. Измерение процесса старения компонентов осуществляется в определенные промежутки времени. Благодаря такому подходу данные, полученные в неравномерно распределенные промежутки времени, предварительно обрабатываются для формирования сравнимых бортовых показателей. На первом этапе агрегируем данные в определенные интервалы. Тем самым динамическая база бортовых данных уменьшается, что позволяет более эффективно анализировать сигналы. Также используем алгоритм машинного обучения с целью создания цифровой модели для измерения процесса старения. С помощью методологии локальных интерпретируемых модельно-агностических объяснений модель становится интерпретируемой. Это позволяет извлекать наиболее релевантные сигналы и тем самым сокращать объем обрабатываемых данных. Полученные результаты показывают,

Адрес для переписки

Удо Сасс Андреас
АО «Фольксваген»
Бриффх 17772,
38436, г. Вольфсбург, Федеративная Республика Германия
Тел.: +7 916 940-00-06
andreas.udo.sass@volkswagen.de

Address for correspondence

Udo Sass Andreas
Volkswagen AG
Brieffach 17772,
38436, Wolfsburg, Federal Republic of Germany
Tel.: +7 916 940-00-06
andreas.udo.sass@volkswagen.de

что для прогнозирования процесса старения рассматриваемого структурного компонента достаточно определенного количества бортовых сигналов. Таким образом, предлагаемый подход позволяет сократить передачу данных бортовых сигналов для проведения профилактического обслуживания.

Ключевые слова: прогнозное техническое обслуживание, выделение признаков, выбор сигнала, временные ряды, машинное обучение, объяснение модели

Для цитирования: Удо Сасс, А. Предварительный выбор сигнала для мониторинга и прогнозирования старения компонентов силовой передачи автомобиля / А. Удо Сасс, Э. Эсатбейоглу, Т. Ивверкс // *Наука и техника*. 2019. Т. 18, № 6. 519–524. <https://doi.org/10.21122/2227-1031-7448-2019-18-6-519-524>

Introduction

A massive amount of information is transmitted in a modern vehicle. This information is used for communication between various Electronic Control Units (ECU). These information are transmitted via the Control Area Network (CAN) bus in form of signals, which can be triggered by different ECUs (e. g. vehicle speed, outside temperature, turn signal status). Due to the utilization and prioritization of the CAN bus, the signals cannot be transmitted in real-time. Because of more safety functions, more driver assistant systems and a higher in-vehicle entertainment the complexity of such vehicles grows rapidly.

Our goal is to provide component aging indicators for the use in a health management or in sense of predictive maintenance. Therefore, we identify relevant groups of signals concerning an observed physical aging process. With the help of this approach, the transmitted in-vehicle signals can be reduced to a small group of relevant signals. Due to the massive amount of data and complex aging process (the selected physical aging process cannot be identified by using only one signal), a manually identification of aging-relevant signals is not suitable. The analyzed prototypes transmit hundreds of signals in a day. After preprocessing, the transmitted data of a prototype's signal contains up to 65 mio. value samples. Therefore statistical features are extracted out of the raw information to reduce this massive amount of data.

This paper is structured as follows. In the section 2 we briefly present some of the related work and background information. Afterwards, we describe the analyzed data and the preprocessing step for generating suitable datasets in section 3. In section 4 our approach for preselecting relevant in-vehicle signals is presented. In section 5 our results are given and evaluated. To the end, section 6 concludes this paper and we give an outlook of future work.

Background

In this section, we provide background information and related works. Different commercial vehicles are equipped with CAN loggers to save the in-vehicle signal streams, including sensor readings, actuator readings and internal parameters of control models. With the help of these loggers, hundreds of in-vehicles signals can be recorded.

In our work we estimate a degree of aging of an Exhaust Gas Recirculation (EGR) cooling system. This aging value can be used for further predictive maintenance approaches. Some authors used special sensors to detect faulty state of the observed component [1–3]. As described in section 1, the amount of information is massive and has to be aggregated to filter the important information. H. Guo et al. shows an approach, to reduce the transmitted information with the help of a cloud [4]. Common feature extraction is a widely used method to keep the amount of samples in a suitable way [5, 6].

The examined prototypes with diesel engines have in common, that the EGR cooling system is observed in various time periods in workshops. With the help of this information, a health status of this component aging is given. The recirculated exhaust gas of the engine back into the intake tract is set from the EGR valve. Due the combustion of sulfur-containing diesel fuel, different types of emissions are released. The EGR rate determines the proportion of exhaust gas mass flow based on the total mass flow filling the engine cylinders. By changing the EGR rate the emissions can be influenced positively [7, 8]. Though, a too high EGR rate causes a fouling in EGR coolers [9, 10].

In addition to the given explanation of the physical aging process, the quality of Machine Learning models for predicting fault diagnosis or health states are related to the quality of input data. In order to increase the quality of the models and to reduce the computational complexity a feature selection is necessary. K. H. Hui et al. present an

approach for selecting a subset of features for predicting machinery faults by using vibration signals [11]. In the first iteration step the best feature is selected. Furthermore, in the next iteration step all other combinations of this feature are tested. The best combination of this iteration is selected. This is repeated until all features are selected. In the end, a subset of feature combinations with the best accuracy and the lowest number of features is selected. The approach shows a accuracy improvement from 74 % to 81 % selected by the features [11].

R. Prytz et al. implement an approach to identify dependencies between on-board signals in a truck [12]. First, the external influences are separated from the internals. In the next step, important relations are found by using Least Absolute Shrinkage and Selection approach (LASSO) and Recursive Least Squares approach (RLS). Zhang et al. select different features for predicting the Remaining Useful Life (RUL) of rolling element bearings [13]. Different statistical features are calculated from the origin monitored signals by processing time domain, frequency domain and time-frequency domain. The features are evaluated by different goodness metrics, such as correlation, monotonicity and robustness. The signals are selected by calculating the weighted linear combinations of the several goodness metrics.

A. Mrowca et al. identify groups of signals in in-vehicle network traces [14]. In this approach, redundant signals are discovered to reduce potentially identical information on the bus load. The authors of this work distinguish between categorical and numerical signals. In order to represent the signal behavior, various features are extracted with respect to the overlapping windows. Furthermore, the feature subset with the best prediction quality is found. With the help of the most important feature subset, the signals can be clustered in several groups [14].

J. A. Crossman et al. analyze signal fault analysis of vehicle engine data in order to find relevant signal features [15]. In a first step the input data is segmented in several dynamic windows. These windows are used to generate signal features. The algorithm ranks the features according to the linear separability of these features. For the current features set the error rate is calculated. In order to reduce the feature set, the backward

selection algorithm eliminates the lowest ranked feature until the best error rate. The authors show, that classification accuracy rises from 61.92 % to 83.84 % by reducing the selected features [15].

Besides the signal reduction, machine learning methods are also implemented to predict an observed target value. M. J. Kane et al. shows in the work, that the Mean Squared Error can be improved with a Random Forest approach for prediction of avian influenza outbreaks [16]. The Support Vector Regression (SVR) is used for predicting the urban air quality of Beijing and cities next to Beijing.

In our paper, we implement a signal pre-selection to reduce the whole signal variance by using Local Interpretable Model-agnostic Explanations (LIME). M. T. Ribeiro et al. present an algorithm to interpret complex classifiers or regressors in a faithful way [17]. In order to achieve this, the algorithm has to be interpretable for humans and the results should be similar to the origin prediction locally. The explanation is defined as following

$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g),$$

where x – origin representation of an instance and x' the binary vector of its interpretable instance; G – class of potentially interpretable models; $\Omega(g)$ – measured complexity [17].

For the representation randomized samples around π are used and with the distance weighted. The fidelity function Λ approximates the global function locally. Furthermore, LIME tries to minimize the complexity and to maximize the quality.

Besides the relevance of a signal, different samples of a signal can cover various amount of the whole database. In order to identify the samples with the highest coverage, LIME has implemented a submodule pick algorithm [17].

Data description

For the further analytics the vehicle internal network traces from the CAN-bus are applied. This network traces consists of a various signal variety of sensors/actors readings and internal parameters of control models. The given prototypes are analyzed regarding their EGR cooler aging. This aging-value is observed for every prototype in certain intervals in workshops. We use these values as training set target value (ground truth).

The in-vehicle signals from the internal network traces include high frequent and unevenly spaced time-series. It is necessary to transform this data in a analyzable form. First, the time-series are cleaned from invalid values. Afterwards, the time-series are synchronized to a 100 ms raster. Instead of interpolating the values, only the associated timestamps of each value-time-pairs are changed to a unified and equidistant time raster. With the help of this method it is ensured, that no measured value is changed and the signals keep its origin behavior. For each recording a trigger signal defines the start and end time. All signals for that recording are cut regarding the length of the trigger signal. Therefore, all signals have the same length within the recordings. Afterwards, for every signals the recordings are merged to a coherently dataset.

The data is segmented regarding several time periods. The associated target value is averaged for that given time period. As mentioned in our previous work [18], a too high dynamic in the dataset will predict an aging-value with a less quality. In order to decrease this dynamic and to reduce the amount of samples, we calculate statistical features directly from the original equidistant signal for a given time period of 9 h. We use similar statistical features as mentioned in [15] as basic features to aggregate the signals: arithmetical mean, 25th and 75th percentile and the standard deviation of the values in each time period. The Fig. 1 shows the signal segmentation and aggregation. This approach is done for the whole vehicle lifetime and for all signals.

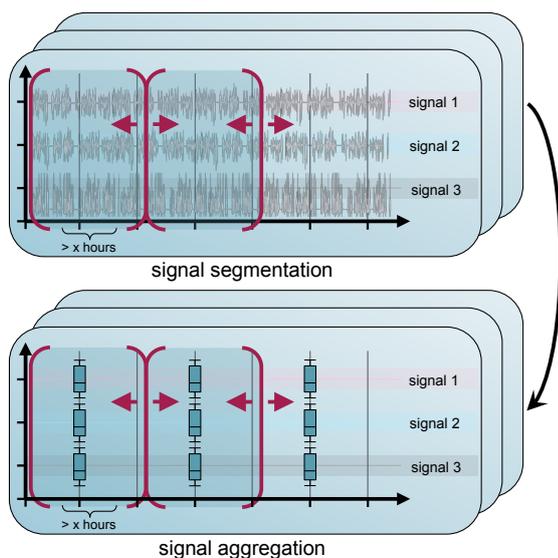


Fig. 1. Example for the signal segmentation and signal aggregation for the synchronized time-series of the vehicle lifetime

Signal preselection

In this section we present our approach to preselect a small amount relevant signals from the whole dataset. First, a machine learning model is used to predict the observed aging-value of the component with the help of the calculated signal aggregations. The focus of our work is the preselection of the signals, for that purpose we use the Random Forest Regression as default model.

As described in the previous section, the final datasets have the same length for all signals. Thus, the dataset is split into a training (2/3) and test dataset (1/3).

Different representative samples of the trained model are picked to explain the model locally. For that purpose we apply the LIME algorithm [17]. As a result we get the weights for explaining the signal relevance locally. To show the local relevance heatmaps are created for every sample. The heatmaps include local scores for the used features (arithmetical mean, 25th and 75th percentile and the standard deviation) of the signals.

Finally, the global score φ is created from the local heatmaps of every signal regarding all statistical features. In order to calculate the score φ_s , h is defined as the amount of heatmaps, the amount of feature for every signal is n and the processed signal s of all signals S . s_{ij} is defined as local weight of a feature j from the signal s within the heatmap i . The score calculation is defined as follows

$$\varphi_s = \sum_{i=1}^h \frac{\sum_{j=1}^n \frac{s_{ij}}{n} \cdot 100}{h} \quad \forall s \in S.$$

After determination we sort the results regarding the global score (relevance).

Results

In this section, we present our results for preselecting in-vehicle signals of a powertrain aging component purpose. The determined in-vehicle signals are evaluated with the help of the root mean square error (RMSE) regarding the predicted aging-value.

The Fig. 2 shows the score map of the top ten in-vehicle signals regarding the relevance (global score) for predicting the aging-value for a selected prototype.

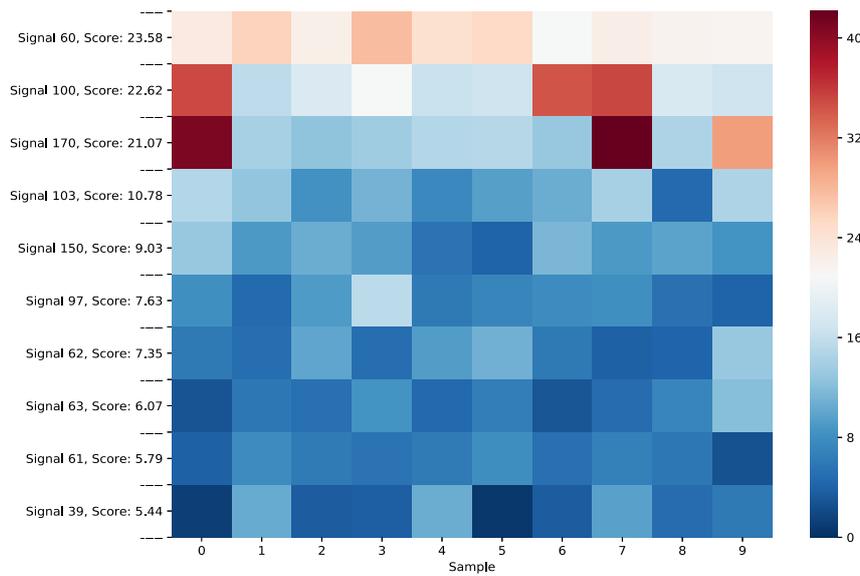


Fig. 2. Score map of the top ten in-vehicle signals regarding the relevance (global score) for predicting the aging purpose for a selected prototype, the x-axis shows different samples with the highest coverage regarding LIME

Different samples cover a various amount of the analyzed database. The submodule pick algorithm identifies the samples with the highest coverage regarding the aging purpose. The colored boxes in a row represent the local score of each signal in the given sample. Each local score is determined by using the four statistical features. The following Tab. 1 shows the description of the determined signals in Fig. 2.

Table 1
Description of the relevant signals for predicting the aging purpose for a selected prototype

Signal number	Description
39	Information about activity on CAN bus
60–63	Different log information
97	Mass of ash in particle filter
100	Mass of soot in particle filter
103	EGR mass flow
150	Oil level information
170	Temperature in EGR valve

In order to evaluate the calculated signals, we apply the RMSE to calculate the goodness of the model. A tuple of ten signals is used to predict the aging-value, for that prediction the RMSE is calculated. The Fig. 3 shows the RMSE, which result from the tuples of signals. The error of the trained models increases with the selection of the worse rated signals. Some outliers do not fit into the global trend. These outliers can be caused dif-

ferently. On the one hand, the score calculation based on the four statistical features of the signal. If a signal has only a single highly relevant feature and another signal has four semi-relevant features, the resulting score can be the same. For this reason, a lower rated signal could return better results than a higher rated. On the other hand, LIME uses different models as explanations. In order to keep the origin signal behavior, the analyzed statistical feature from the signals are not normalized.

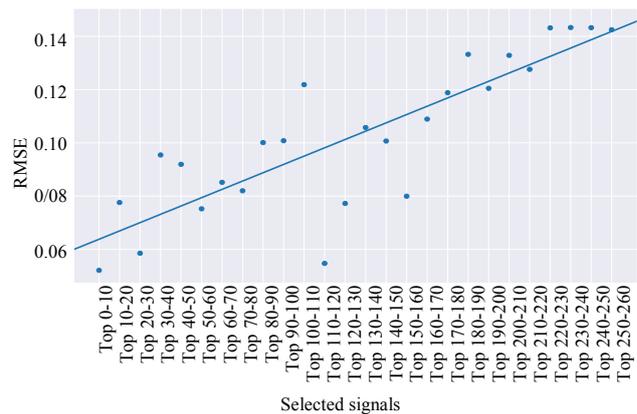


Fig. 3. Overview of the RMSE regarding the in-vehicle signals to predict the aging-value, ten signals (sorted by global score) are combined to calculate the error

When using a linear model, a feature with a very high absolute value could be represented from a low factor in the linear model, although the influence

of that feature is very relevant. Because of using features in a similar range, this behavior appears only in exceptional cases. Despite the outliers, the figure shows well weighted scores in order to explain their relevance regarding the physical aging process.

CONCLUSIONS

1. In this work we analyzed dynamic in-vehicle signals and predicted the aging value. The recorded network traces are cleansed and synchronized. After that, the time series of the signals are segmented and aggregated to equidistant datasets. With the help of LIME a small group of relevant signals are preselected for further analytics. The whole amount of analyzed data is compressed multiply by comparison the origin network traces to the preselected aggregated datasets. Our goal is to deliver component aging indicators for the use in predictive maintenance. With the help of our approach, a selected physical aging process can be assigned to a unique group of relevant signals.

2. In the future, unknown aging processes can be identified by using the assignment of preselected signal groups and the aging processes. A cloud can save various signal groups and aging type configurations and transmit it to all the vehicles within analyze cluster. In this case, an expensive aging observation must be only for one vehicle done and the resulting signal groups can be used for all the other vehicles. Furthermore, an aging model of all known aging processes can be used to implement in a health management by using the relevant signal groups and can inform the customers in sense of predictive maintenance.

REFERENCES

- Goyal D, Pabla B. S. (2015) Condition Based Maintenance of Machine Tools – a Review. *CIRP Journal of Manufacturing Science and Technology*, 10, 24–35. <https://doi.org/10.1016/j.cirpj.2015.05.004>.
- Teti R., Jemielniak K., O'Donnell G., Dornfeld D. (2010) Advanced Monitoring of Machining Operations. *CIRP Annals*, 59 (2), 717–739. <https://doi.org/10.1016/j.cirp.2010.05.010>.
- Bediaga I., Mendizabal X., Arnaiz A., Munoa J. (2013) Ball Bearing Damage Detection Using Traditional Signal Processing Algorithms. *IEEE Instrumentation & Measurement Magazine*, 16 (2), 20–25. <https://doi.org/10.1109/mim.2013.6495676>.
- Guo H., Crossman J. A., Murphey Y. L., Coleman M. (2000) Automotive Signal Diagnostics Using Wavelets and Machine learning. *IEEE transactions on Vehicular Technology*, 49 (5), 1650–1662. <https://doi.org/10.1109/25.892549>.
- Carino J. A., Delgado-Prieto M., Iglesias J. A., Sanchis A., Zurita D., Millan M., Ortega Redondo J. A., Romero-Troncoso R. (2018) Fault Detection and Identification Methodology under an Incremental Learning Framework Applied to Industrial Machinery. *IEEE Access*, 6, 49755–49766. <https://doi.org/10.1109/access.2018.2868430>.
- Carino J. A., Delgado-Prieto M., Zurita D., Millan M., Ortega Redondo J. A., Romero-Troncoso R. (2016) Enhanced Industrial Machinery Condition Monitoring Methodology Based on Novelty Detection and Multi-Modal Analysis. *IEEE Access*, 4, 7594–7604. <https://doi.org/10.1109/access.2016.2619382>.
- Ladommatos N., Balian R., Horrocks R., Cooper L. (1996) The Effect of Exhaust Gas Recirculation on Combustion and NO_x Emissions in a High-Speed Direct-Injection Diesel Engine. *SAE Technical Paper Series*. <https://doi.org/10.4271/960840>.
- Zelenka P., Aufinger H., Reczek W., Cartellieri W. (1998) Cooled EGR – a Key Technology for Future Efficient HD Diesels. *SAE Technical Paper Series*. <https://doi.org/10.4271/980190>.
- Hoard J., Abarham M., Styles D., Giuliano J. M., Sluder C. S., Storey J. M. E. (2008) Diesel EGR Cooler Fouling. *SAE International Journal of Engines*, 1 (1), 1234–1250. <https://doi.org/10.4271/2008-01-2475>.
- Bravo Y., Moreno F., Longo O. (2007) Improved Characterization of Fouling in Cooled EGR Systems. *SAE Technical Paper Series*. <https://doi.org/10.4271/2007-01-1257>.
- Hui K. H., Ooi C. S., Lim M. H., Leong M. S., Al-Obaidi S. M. (2017) An Improved Wrapper-Based Feature Selection Method for Machinery Fault Diagnosis. *PLOS ONE*, 12 (12), e0189143. <https://doi.org/10.1371/journal.pone.0189143>.
- Prytz R., Nowaczyk S., Byttner S. (2011) Towards Relation Discovery for Diagnostics. *Proceedings of the First International Workshop on Data Mining for Service and Maintenance – KDD4Service '11*. ACM Press, San Diego, California, 23–27. <https://doi.org/10.1145/2018673.2018678>.
- Zhang B., Zhang L., Xu J. (2016) Degradation Feature Selection for Remaining Useful Life Prediction of Rolling Element Bearings. *Quality and Reliability Engineering International*, 32 (2), 547–554. <https://doi.org/10.1002/qre.1771>.
- Mrowca A., Moser B., Gunnemann S. (2018) Discovering Groups of Signals in In-Vehicle Network Traces for Redundancy Detection and Functional Grouping. *Machine Learning and Knowledge Discovery in Databases*, Springer, Cham, 86–102. https://doi.org/10.1007/978-3-030-10997-4_6.
- Crossman J. A., Hong Guo, Murphey Y. L., Cardillo J. (2003) Automotive Signal Fault Diagnostics. I. Signal Fault Analysis, Signal Segmentation, Feature Extraction and Quasi-Optimal Feature Selection. *IEEE Transactions on Vehicular Technology*, 52, 1063–1075. <https://doi.org/10.1109/tvt.2002.807635>.
- Kane M. J., Price N., Scotch M., Rabinowitz P. (2014) Comparison of ARIMA and Random Forest Time Series Models for Prediction of Avian Influenza H5N1 Outbreaks. *BMC Bioinformatics*, 15 (1). <https://doi.org/10.1186/1471-2105-15-276>.
- Ribeiro M. T., Singh S., Guestrin C. (2016) “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '16*. ACM Press, San Francisco, California, USA, 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
- Sass A. U., Esatbeyoglu E., Fischer T. (2019) Monitoring of Powertrain Component Aging Using In-Vehicle Signals. *Diagnose in Mechatronischen Fahrzeugsystemen XIII: Neue Verfahren für Test, Prüfung und Diagnose von E/E-Systemen im Kfz*. Books on Demand, 15–28.

Received: 08.10.2019

Accepted: 29.11.2019

Published online: 06.12.2019