Artificial Intelligence Research and Development
A. Cortés et al. (Eds.)

doi:10.3233/FAIA220330

Feature Engineering and Machine Learning Predictive Quality Models for Friction Stir Welding Defect Prediction in Aerospace Applications

Marta CAMPS[a,1], Maddi ETXEGARAI[a], Francesc BONADA[a], William LACHENY[b], Sylvain PAULEAU[b] and Xavier DOMINGO[a]
[a] *Eurecat, Centre Tecnològic de Catalunya, Unit of Applied Artificial Intelligence, Av. Carrer de Bilbao, 72, 08005 Barcelona, Spain*
[b] *Ariane Group, 51/61 route de Verneuil, Bâtiment 71 - Bureau 142, 78131 – Les Mureaux – France*

Abstract. Data-Driven Predictive Quality solutions are of utmost importance for Industry 4.0 in general and for high added value and complex manufacturing systems in particular. A unique Friction Stir Welding process is performed for the manufacturing of the new Ariane 6 aerospace launchers. This work presents a novel feature engineering approach that correlates Friction Stir Welding process data and quality inspection data to build a Machine Learning-based predictive quality solution. This solution predicts the presence of welding defects, empowering end-user's quality assurance and reducing quality inspection time and associated costs.

Keywords. Machine Learning, Predictive Quality, Industry 4.0, Aerospace Industry

1. Introduction

For product quality assurance, solutions that build upon data analytics, Machine Learning (ML) and Artificial intelligence (AI) can provide major benefits, empowering end-users' decision making that can lead to better products and optimized production lines. H2020 SESAME [1] project aims to bring innovative Industry 4.0 and Data-Driven solutions to the aerospace industry, focusing on reducing the costs of the new Ariane launcher production through data science models for quality prediction, predictive maintenance, and supply chain agility. The aerospace manufacturing industry its highly challenging due to their strict quality requirements and the low production volume.

The use case introduced here studies the Friction Stir Welding (FSW) [2] process on the Ariane 6 Pre-Final Assembly Line. The proposed model, consisting in a set of data analytics and AI/ML-based tools, digests a large amount of welding process data to provide the expert with valuable quality insights that can lead to better and more cost and time-efficient quality control procedures. Following the preliminary study introduced in Camps, et al. [3], this work presents a feature engineering layer as well as the modelling of a binary classifier for the prediction of the presence of the welding defect is presented.

[1] Eurecat, Centre Tecnològic de Catalunya, 08005 Barcelona, Spain; E-mail: marta.camps@eurecat.org

2. Data Fusion & Feature Engineering Layer

For each welding test, the evolution along the 360º (cylindrical tank) of 53 process parameters, including temperature, force, position, and current variables of different FSW subsystems and sensors are acquired together with the FSW station configuration set-up. The sampling frequency of the process parameters varies from one to another (ranging from 0.7 Hz to 330 Hz), resulting in an inconstant timestamping. To obtain a uniform and constant sampling frequency, a resampling methodology based on linear interpolation is implemented.

Once the cylindrical tanks are welded, strategic perpendicular cuts are done to assess the quality of the welding, which is directly related to the penetration of the welding pin into the interface between two welded parts. Even though the process data is acquired for several tests, due to the destructive and expensive nature of the quality inspection, it is performed only in 6 welded tanks and in a few positions of each of those welded tanks. Hence the importance of data-driven solutions that can estimate their values without destructing the welded part. To increment the quality label granularity, linear interpolation is applied.

To get uniform inputs to feed the predictive quality algorithm, a correspondence between process data (timestamped) and quality inspection (space stamped) needs to be implemented. The data is segmented in the same spatial length windows and a set of 9 features are computed for each window of each parameter: the minimum value, the maximum value, the mean value, the standard deviation, the maximum value of the first-order derivative and the first two Power Spectral Density (PSD) computed with Welch's method [4] and their amplitudes. Thus, each window is characterized by 477 new features. To accelerate the computation of the model and to reduce the high correlation observed among some of the features, the Principal Component Analysis (PCA) [5] dimensionality reduction algorithm is applied. Additionally, the two configuration parameters (pin diameter and length) and the Crown variable are added to the input set without the PCA transformation. The input data (process features & configuration parameters) is normalized with the Standard Scaler, which transforms the data to a standard normal distribution with a mean of 0 and a standard deviation of 1 [6].

The process data initially acquired based on the timestamp, and the quality data, based on the position of the quality inspection, are now both referenced to window position, implementing a time-spatial correlation.

3. Model and Hyperparameter Tuning

The proposed predictive quality model is based on a binary classifier algorithm which estimates the quality of the segment, defined as the presence of a defect or not. To explore different strategies and select the best one, different model training strategies determined by the size of the windows or segments, the algorithm and the hyperparameters associated are considered. The exploration focuses in four different windows sizes: 1, 2, 4 and 6 degrees; and six different ML classifiers: Support Vector Machine (SVM) [7], K-Nearest Neighbours (KNN) [8], Decision Trees (DT) [9], Logistic Regression (LR) [10], Naïve Bayes (NB) [11], Bag decision trees (BDT) [9]. For the hyperparameter tuning of each model, an exhaustive grid-search method is implemented, generating candidates from a grid of hyperparameter values specified to evaluate their impact on the model performance.

As there are only 6 tests or files available to train, which are quite different due to the FSW configuration, the training is done by saving one file for the test and using the remaining 5 files to train the model, following a Leave-One-Out strategy [12]. Furthermore, 10 subsets are generated from each training dataset choosing randomly half of the population, thus implementing a dual cross-validation strategy. The best model is chosen by computing the mean of the accuracy metric for each combination of window, algorithm and hyperparameters across all iterations.

4. Results and validation

In this section, the results obtained in the dual cross-validation strategy are presented. Figure 1 shows the accuracy of some of the models generated, for the training in dark blue and test in light blue. Each of the plots in Figure 1 exhibits the results for a given window size and the best hyperparameters for each algorithm. A significant difference is observed between the train and test datasets, especially for 1-degree windows, implying models susceptible to overfitting in this configuration. The best performance is found in the 6-degrees moving window (0.72 for the test), with the smallest difference between train and test accuracy (around 0.28). Therefore, we can ensure that the models with 6-degree windows are more robust and reliable.

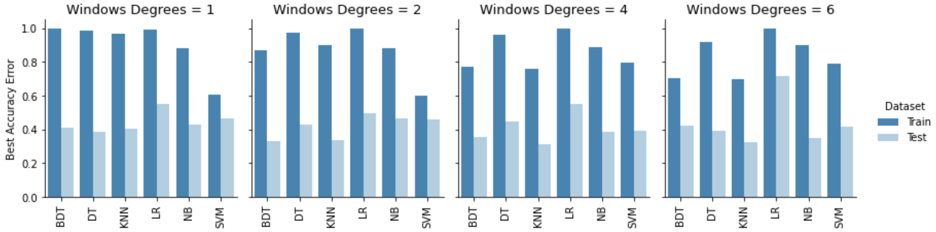

Figure 1. Classifier comparison of accuracy error by the window size.

It can be seen that LR outperforms the rest of the models, tuned with the hyperparameters, *C=1*, *penalty=none*, proving an accuracy error of 0.72. It is important to note also that LR shows a constant performance throughout all the cross-validation folds and hyperparameter combinations, and consequently presents a more reliable generalization. On the other hand, the KNN classifier shows a strong dependency on the train and test files selected, indicating that it is prone to overfitting.

Finally, the cross-validation confusion matrix for the LR model is presented in Table 1. The model performs better for the absence of defect due to the class imbalance.

Table 1. Cumulative cross-validation confusion matrix for quality prediction. Counts and percentage

Predicted	**Good**	4646	1649
	Bad	1110	2595
		Good	**Bad**
		Real	

Predicted	**Good**	81%	39%
	Bad	19%	61%
		Good	**Bad**
		Real	

5. Conclusions

The model introduced focuses on predictive quality solutions for the FSW stations of the Ariane Pre-Final Assembly Line. The design proposed, developed, and benchmarked builds upon a data processing and feature engineering layer. It establishes the spatial-time correlation, based on a moving window that determines the relation between process parameters timestamps and quality inspection position in degrees.

To predict the presence of the defect, a strategy based on model competition and benchmarking for different ML classifiers is presented. Employing two-step cross-validation, the performance of the different algorithms, hyperparameters and moving window sizes is evaluated. Larger moving windows obtain more balanced results between train and test data, therefore a compromise solution of 6-degree windows is selected. The presence of defects can be predicted with over 70% of accuracy with a LR model. There is an important decrease in the performance of test data compared to train data. Large differences may indicate that the model is still not completely able to generalize and perform under new welding conditions, indicating that models suffer from overfitting (what is learned in the training phase does not match new unseen data). This is due to two main reasons: limited available data together with different configurations of the FSW station in each dataset.

Overall, these results, although preliminary given the small dataset, show the potential of the predictive quality module carried out within the SESAME project for complex manufacturing systems in the aerospace industry which is particularly challenging due to the low production cadence and extreme quality requirements. Furthermore, it can provide very valuable insights to process experts to focus their attention on specific welding sections that need to be inspected for quality assurance.

6. Acknowledgment

This work has been carried out in the framework of the SESAME project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821875. The content of this paper reflects only the author's view; the EU Commission/Agency is not responsible for any use that may be made of the information it contains.

7. References

[1] https://sesame-space.eu
[2] Thomas WM, et al., Friction stir welding for the transportation industries, Mater. Des. 1997;18, 4–6.
[3] Camps M., et al., Data-Driven Analysis of Friction Stir Welding for Aerospace Applications. In Artificial Intelligence Research and Development 2021 (pp. 181-184). IOS Press.
[4] Welch P. The use of the fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. 1967; 15: 70-73.
[5] Pearson K. On lines and planes of closest fit to systems of points in space. Philos. Mag. Lett. 1901. 2.11:559
[6] Iglewicz B. Robust scale estimators and confidence intervals for location. Understanding robust and exploratory data analysis, 1983, 404: 431.
[7] Boser BE, et al., A training algorithm for optimal margin classifiers. Proc. COLT. 1992: 144-152.
[8] Fix E, et al., Discriminatory analysis. Nonparametric discrimination: Consistency properties. ISR, 1989, 57.3: 238-247.
[9] Sutton CD. Classification and regression trees, bagging, and boosting. Handb. Stat, 2005, 24: 303-329.
[10] Hosmer Jr. DW. Lemeshow, S.; Sturdivant, R. X. Applied logistic regression. John Wiley & Sons, 2013.
[11] Murphy KP, et al. Naive bayes classifiers. University of British Columbia, 2006, 18.60: 1-8.
[12] Hastie T, et al. The elements of statistical learning: data mining, inference, and prediction. Springer, 2009.