Journal of Machine Engineering, Vol. 12, No. 1, 2012

cognition, classification
self-optimization
gear testing, contact pattern

Robert SCHMITT[1]
Christian NIGGEMANN[1]
Matthis LAASS[1]

COGNITIVE FAILURE CLUSTER ENHANCING THE EFFICIENCY AND THE PRECISION OF THE SELF-OPTIMIZING PROCESS MODEL FOR BEVEL GEAR CONTACT PATTERNS

The contact patterns of bevel gear sets are an important indicator for the acoustic quality of rear axle drives. The contact patterns are the result of complex interactions in the production process. This is due to many process steps, numerous influencing factors and interdependencies. In general, their effect on product variations is not fully comprehended. This impedes the design and control of the production process based on a holistic analytical model for new variants fulfilling the acoustic requirements. The approach with self-optimization is possible but can take a long time for the training of the artificial neural networks and the necessary iterations until a satisfying precision for the predicted process parameters is achieved. Also it can occur that the algorithm is not converging and therefore no satisfactory result is turned out at all. In this paper an approach is presented combining the flexibility of self-optimizing systems with the higher precision of delimited solution finders called the Cognitive Failure Cluster (CFC). The improvements provided by the clustering of the optimization program are evaluated regarding the training time and the precision of the result for a production lot of bevel gear sets.

1. INTRODUCTION

The manufacturing industry in high-wage countries is in a challenging situation due to increasing pressure from competitors offering lower labour costs. This is especially evident for standard products. So the high-wage countries have to get more into the production of more sophisticated or custom-made products [1]. This requires a continuous improvement of production technologies and capabilities. However, this can mean more complex processes, which are more difficult to control due to numerous influencing factors or that the production process has to be adapted more frequently due to smaller production lots. The outlined situation poses a polylemma between scale and scope and value orientation and planning orientation [2]. So the challenge is the economically feasible manufacturing of innovative and technologically demanding products of high precision and quality.

[1] Laboratory for Machine Tools and Production Engineering, RWTH Aachen University, Germany

A good example depicting this polylemma is the manufacturing of rear axle drives, because for these many variants exist, which are produced merely at medium lot sizes (~ 10,000 units). Sometimes even different manufacturing processes are used for another variant. The frequent change in variants leads to a high planning effort for the necessary adaptation of the manufacturing processes. However, the customer does not accept that this planning effort is added to the product price. Thus, to solve the conflict the manufacturing of rear axle drives in high-wage countries has to be reorganised to keep it competitive. As outlined by Schmitt et al. [3-5] self-optimization can be an answer.

2. CONCEPT OF SELF-OPTIMIZATION

A self-optimized system is designed in that way that it can pursue different goals and adapts its behaviour on the actual conditions of the production system. While a change in the production system is triggered externally by humans, i.e. the decision which variant will be manufactured next, the decision's effect on the production system could be predicted by the self-optimized system. This means that it defines how single parameters of a manufacturing process have to be adapted based on the part history up to that point to fulfil the production function. This concept is called Cognitive Tolerance Matching (CTM). By this the step from the optimization of single production steps to the function-oriented optimization of the final product is done. This benefits the value-orientation and reduced the planning effort, solving the aforementioned polylemma of production. According to Frank et al. self-optimizing systems have to continuously repeat the following three actions [6]:

- Analysis of the actual conditions → sensors and metrology → valid process data.
- Determination of new system targets → cognition → derive well-founded decision.
- Adaptation of the system behaviour → actuators → altered processes.

Sensors are necessary to acquire the actual process conditions. The sensors have to return a quantitative value for the measurand with sufficiently small uncertainty. For the production predominately the sensing of geometric and functional features is important. So the main task is that sensors provide valid process data of the actual state. Otherwise there could be a high risk taking a disadvantageous decision, i.e. altering the process unnecessarily or even worse doing nothing while the process is out of tolerance. Based on the valid process data by means of cognitive methods a founded decision can be derived. This decision comprises new or adapted target values for actuating variables, for instance machining or assembly parameters. The decision taking is then based on the function-oriented optimization of the production process considering several command variables, e.g. adaptive functional tolerances. Subsequently, the decision has to be implemented in the production line by actuators such as manipulators, handling systems, drives and also humans, e.g. a worker who is changing the process parameters at the machine tool.

For the application scenario rear-axle drive the self-optimising behaviour is achieved by using cognitive methods based on Artificial Neural Networks (ANN) [3-5]. These partly model the human capacity of solving complex problems by formalized rules, which are implemented using the Soar programming platform [7]. Soar provides different learning techniques such as chunking and reinforcement learning so that one can benefit from its experiences. So it is possible to use already acquired knowledge about the work pieces, the manufacturing processes, similar cases and process interdependencies. This knowledge has to be stored in the self-optimizing system in a way that it can be automatically processed.

However, beside the beneficial qualities of the cognitive methods with increasing complexity of the production process the number of rules rises significantly. This could increase the processing time for the optimization beyond the economically feasible margin. Also it could reach the limit of the Random Access Memory (RAM) of standard computers used for process planning. Another issue is the precision of the suggested solution. For some processes the solution is unfeasible, i.e. a too high deviation between simulation and reality.

3. COMBINATION OF COGNITIVE TOLERANCE MATCHING AND COGNITIVE FAILURE CLUSTER

The drawback of the current self-optimizing system is attended to in this paper. To achieve a more efficient and robust system the flexibility of the current self-optimizing systems is combined with the higher precision of delimited solution finders.

For this the Cognitive Failure Cluster (CFC) is integrated in the program structure of the existing self-optimizing system for the rear axle drive production process. The adapted program structure is displayed in Fig. 1. This structure describes the information flow for the testing of the gear sets with single-flank testing machines, which is an important preliminary process step to the final assembly of the gear box. By this testing the face clearance and the block dimension are determined. Based on these parameters the mounting of the single gears in the housing can be adapted to improve the running and acoustic behaviour of the rear axle drive. A certain face clearance is necessary to compensate pitch and tooth thickness deviations and avoid meshing interferences. On the other hand a too high face clearance displaces the contact pattern and may induce disturbing noise in operation.

The testing machine measures the working variation of the gear set. For the test the crown gear is mounted at fixed axis-centre distance to the gauge gear, a master gear with well known deviation. This gear is fixed on the driven axis perpendicularly to the crown gear axis. The gauge gear is then moved in that way that the gear flanks become engaged (single flank contact with right flank left flank sequence). While the engaged gears are rotated, the difference in the angle of revolution between the driving and the driven axis is measured. Before the gears are rotated the flanks are sprayed with paint. It was assured that the flanks are fully wetted but not dripping. As the paint film is not thicker than a remaining film of lubricants used for the gear cutting no additional effect on the measurement is expected. The paint is squeezed out in those areas where the opposite flanks are in contact. Having achieved a stable state, i.e. after several revolutions, a distinct contact pattern is left.

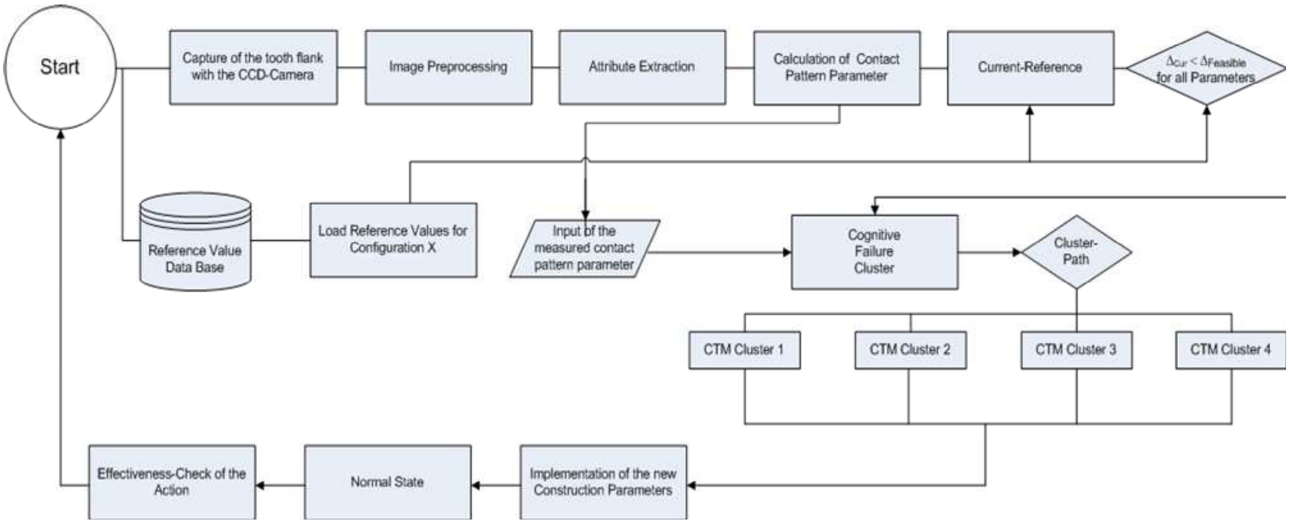


Fig. 1. Operation sequence for Soar program enhanced by the Cognitive Failure Cluster

The contact pattern is a 2D feature and results from the cumulative deviations of the gear set. So it is a more powerful function-oriented quality indicator than standard deviation parameters as it considers the whole contact area. The contact patterns are acquired by 2 CCD cameras, one for the traction flanks and one for the thrust flanks. So in this case the cameras are the sensing components to assess the actual process state. The CCD cameras are arranged within the single flank gear testing machine so that the contact patterns can be acquired in-line. Due to the harsh environmental conditions (paint spray, dust and light intensity) additional measures have to be taken to assure the high-quality imaging, e.g. spot light illumination of region of interest, lens protection, counter air current device.

The outputs of the CCD cameras are colour images of the flanks (Fig. 2 left). In these images the contact patterns are coherent dark regions. The images have to be processed further to obtain the pattern features, e.g. filtering, edge enhancement, edge detection, morphological operations to render a closed contour. Algorithms for these operations can be found in standard works for image processing, e.g. in [8], [9].

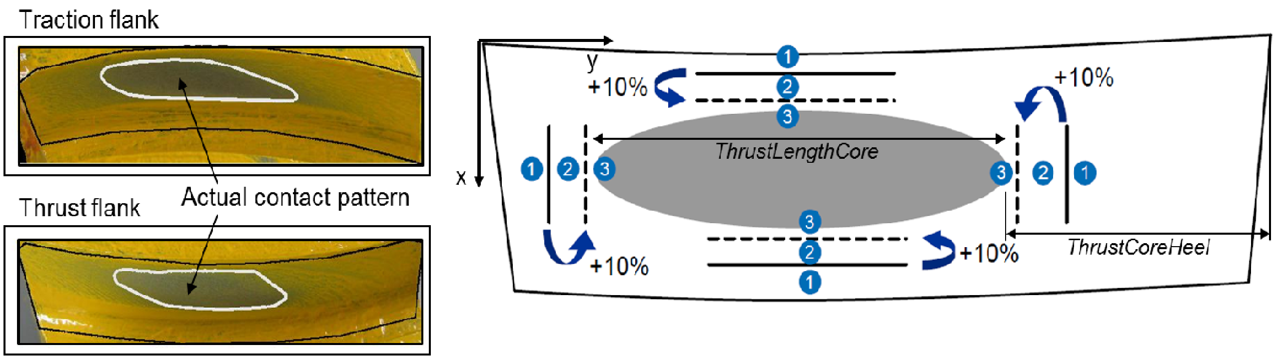


Fig. 2. Contact pattern images and tolerances for contact pattern

The pattern features are for instance the pattern length (“ThrustLengthCore”) and width, the distances of the pattern towards each edge of the gear flank (e.g. “ThrustCoreHeel”), the centroid position of the pattern or the inclination of its main axis (Fig. 2 right). Displayed are the tolerance regions for the pattern on the flat projection of the flank, whereas region 1 beyond the continuous line lies out of tolerance (not in order), region 2 has an A rating (in order) and region 3 within the dashed lines an AA rating.

For the latter the edge distances are increased by 10 % regarding the lower feature limit. This also means that at the same time the permissible length and width for the contact pattern are reduced. The 10 %-rule applies for all parameters with the intention to be always within the AA region and to reduce the still quite high permissible variation of the contact pattern alignment. The A tolerance is derived from established boundary values, which are based on the experience with the regular manufacturing and assembly processes. For the 10 %-margin a significant improvement of the gear box acoustic compared to the standard set-up has been observed while still having a sufficient contact surface necessary for the power transmittance. However, in case the tighter tolerance would be taken instead of the actual tolerance, the production would become more costly as more accurate machines would be required for each manufacturing step to keep the process variation of the resulting contact pattern low. So at this stage a flexible optimization system is needed. Taking into account the tolerances the measurement process has to be sufficiently repeatable so that a founded decision could be taken based on the measurement. It has been observed that for 50 repetitions the process variation for several contact pattern features could be below 0.1 mm.

Having determined the contact pattern features, the CFC then classifies them depending on their position deviation towards the nominal position (distance and direction). For each class a separate Soar-based optimization routine is implemented to obtain adequate assembly parameters, which lead to an AA rating. So this is the step for which the new target values, in this case the assembly parameters, are deduced by cognitive methods.

By applying the CFC one big optimization program for the current system is replaced by a number of smaller optimization programs. The notion is to reduce the complexity of the optimizer by eliminating unreasonable parameter combinations. For instance, in case of a shift of the contact pattern towards the upper edge of the flank a combination of the assembly parameters face clearance and block dimension decreasing the distance towards this edge further would not make sense. So with the CFC the optimization is related to a smaller set of rules thus having a more efficient program. Hence, a classification oriented at the shifting direction is sensible. How the classification is done is explained in more detail in the next section.

4. FUNCTION-ORIENTED CLUSTERING OF GEOMETRICAL CONTACT PATTERN DEVIATIONS

The clustering, i.e. the generation of feature classes, is done to aggregate observed pattern shifts for a lot of gear sets, e.g. a shift to the upper left edge of the thrust flank, to a defined number of classes, whereas the coherence of each cluster is characterised by a distance measure, e.g. the Euclidean distance in the multi-dimensional parameter space (i.e. Edge distances ThrustLengthCore, ThrustWidthCore, ThrustCoreHead… altogether 10 parameters) of the single observation towards the cluster centroid. A certain class of pattern deviations correlates with a set of countermeasures implemented as rules in the corresponding optimization program. So for a new inquiry the acquired set of features is then assigned to one of the generated classes revealing the adequate contact pattern shift.

Several established clustering methods have been analysed to assess their robustness, the plausibility of the clustering and the calculation performance for the present case. Typical clustering methods are density-based, hierarchical or perform a partitioning of the data set. The clustering ends when either a predefined number of classes have been reached or the classes have exceeded the margin of the distinguishing feature. Among the tested methods were k-means, DBSCAN, k-medoid and Expectation Maximization Clustering. For the k-means method also different kernels (radial, polynomial, sigmoid, Epanechnikov) have been evaluated. For the performance assessment 80 gear sets randomly taken out of the serial production have been measured and evaluated using RapidMiner 5, an open- source data mining development environment.

The best clustering results have been achieved with the k-means method [11] using the Epanechnikov kernel [10]. This clustering method partitions n observations into k clusters, whereas each observation is assigned to the class with the nearest mean. This method requires that the number of clusters k is defined a priori to the run of the algorithm. For this expert knowledge about the assembly process and the cause-effect-relationship between a certain pattern shift and the acoustic quality is utilised. Also a systematic refinement of k has been done to find out how sensitive the clustering reacts upon a variation of this parameter.

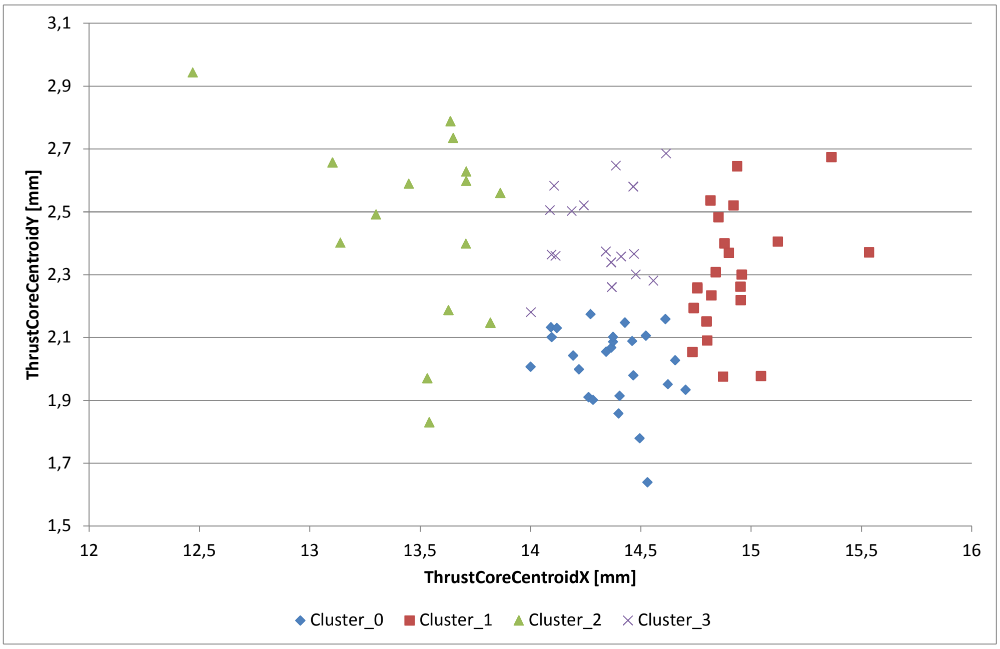


Fig. 3. K-means clustering of centroid coordinates of contact pattern for the thrust flank

Exemplary, the result of the clustering is visualized in Fig. 3 for the centroid coordinates of the contact pattern for the thrust flank. The centroid coordinates in the unit mm are defined in the coordinate system of the flank (view Fig. 2), whereas the origin is located at the tooth tip of the toe, meaning the flank edge facing to the gear centre. Each spot in the figure marks a different gear set. The results show that with the chosen clustering algorithm four distinct regions can be detected indicating a shift of the centroid to the left, to the right, upward or downward. The outcome was that these four classes are sufficient to perform the function-oriented optimization of the contact pattern.

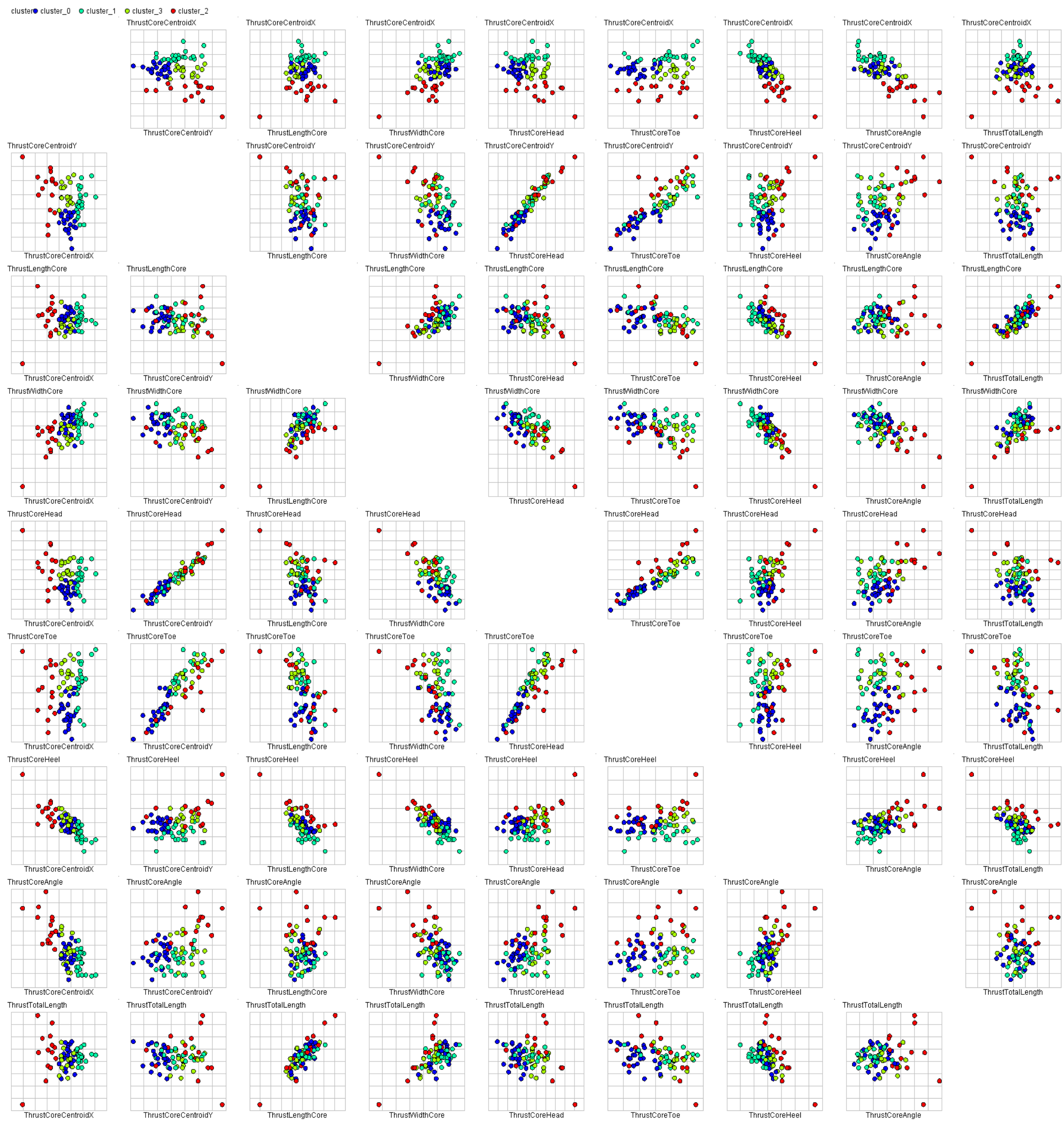


Fig. 4. Overview of clustering results for all contact pattern features

The clustering has to be applied simultaneously to the other contact pattern features so that a certain gear set is properly assigned to the right class of rules. So basically the result shown in Fig. 3 is only a small excerpt of the multi-dimensional feature space. The cross-feature results are visualized in Fig. 4 by plotting the cluster marks in the feature 1 (horizontal axis) vs. feature 2 (vertical axis) coordinate system. It is evident that the diagonal is left out and that the matrix is symmetric as only the axes are exchanged. It appears that some features are quite well correlated while others have no linear dependence. Their correlation can be characterized by the correlation coefficient. For instance, the feature “ThrustCoreHead” is highly correlated with the feature “ThrustCoreCentroidY”. This is evident as with a shift of the centroid in the positive y-direction (see coordinate system in Fig. 2) also the distance to the edge at the tooth head increases. The oppositional effect occurs for the feature “ThrustCoreBase”. Uncorrelated are the length of the contact pattern and the angle of the main axis of the contact pattern.

In case of a new measurement of the gear set the derived features will be assigned to the appropriate class. For consecutive runs of the k-means clustering algorithm it can occur that a data set is assigned once to the class A and once to the class B. This variation is related to the iterative procedure for k-means, whereas the initial k class centroids are arbitrarily distributed in the feature space and then each data set is assigned to that class to which the distance measure to the centroid becomes smallest. This does not generally affect the precision of the optimization program. But it can occur that due to an inappropriate assignation the necessary number of runs for the optimizer can increase.

After the clustering is done the corresponding optimization routine will be initiated providing recommended assembly parameters, which lead to an AA contact pattern after the final assembly. The scenario is that the optimization is done after the single-flank testing, which is a 100%-testing. At this stage the contact patterns and their features are available. Subsequently, the optimization with the cognitive system must be done in the time after having finished the single-flank testing and before the final assembly. This limits the available calculation time as it has to abide by the production tact by all means to prevent bottlenecks at the assembly line or disturbances in the flow of the components. So it is important that the assembly parameters are known well before the gear set is mounted. The improvement in performance accomplished by embedding the CFC in the cognitive system is discussed in the next section.

5. PERFORMANCE EVALUATION OF COMBINED SYSTEM

The performance of the cognitive system with embedded CFC is compared to the performance of the original CTM system. Criterions for this comparison are the average number of loop cycles required until a satisfactory solution (i.e. having a solution with AA rating) is rendered by the optimizing program and the number of processed Soar rules per computer core. Both criterions affect the calculation time which is crucial if the cognitive system should be utilised in the production line. For the performance evaluation a Pentium4 with 2,8GHz, 2 cores, 2GB RAM and Windows XP Pro operating system has been used.

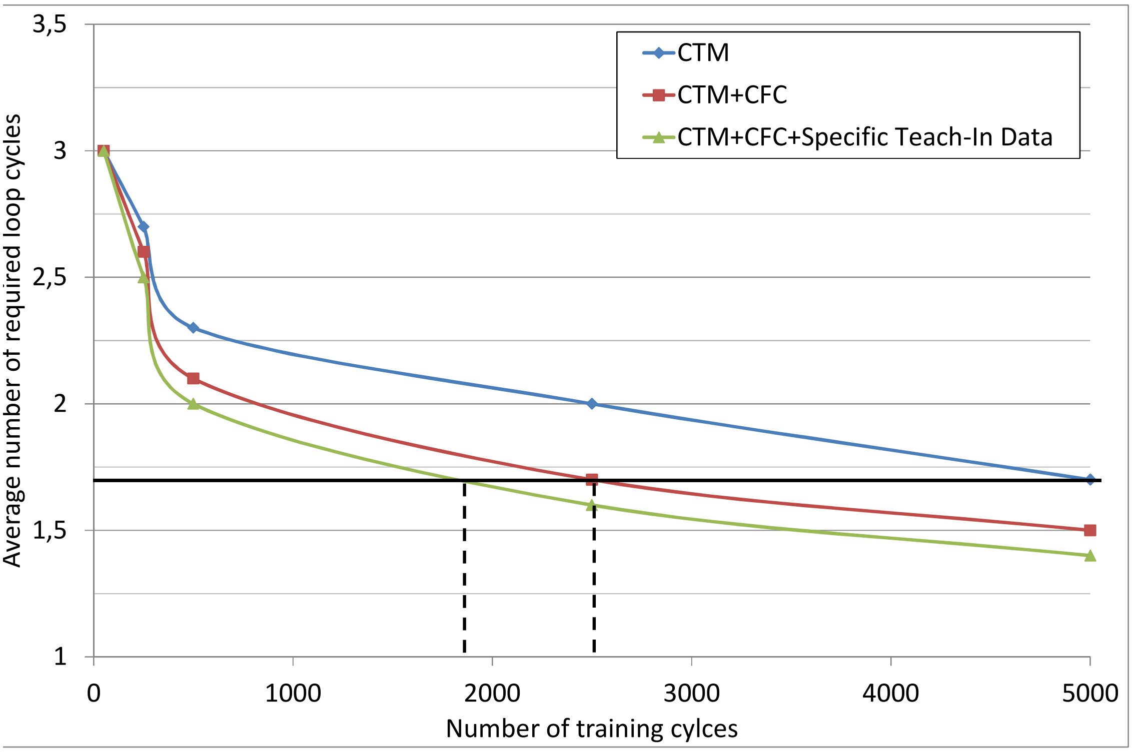
In the left diagram in Fig. 5 the average number of loop cycles is plotted vs. the number of training cycles to illustrate the relationship between the training effort and the number of runs required until the AA rating is achieved. The loop cycles characterise the analysis with real test data. For the cognitive system based merely on CTM random training data (various deviations, e.g. data, which are out of tolerance for the tooth tip distance, the tooth base distance, etc., one Soar program) have been used. For the cognitive system enhanced with the CFC also random training data (various deviations, 4 Soar programs) were utilised while for the cognitive system with CFC specific Teach-In data (4 Soar programs, specific training data for each program, e.g. out of tolerance data for the head distance for Soar program corresponding to cluster_3) have been taken into account.

It can be observed that the latter set-up needs in average 0.3 less loop cycles to achieve the AA rating than the CTM-based system. The maximum number of loop cycles was 60 even though this was an exemption, while obviously the minimum was 1. Generally, for the first 500 training cycles the average number decreased from 3 to 2 loop cycles for the CTM+CFC+Specific Teach-In data set-up. Then for a higher number of training cycles the decline of the necessary number of loop cycles became weaker. For 5.000 training cycles the CTM-based system still needed in average 1.7 loop cycles. For the same performance the enhanced system needed about 2.300 cycles, so 55 % less. This corresponds to a similar cut in training time (15 minutes for 5.000 cycles compared to 8 minutes for 2.300 cycles). An explanation for this is that through the a priori clustering the probability that the rule with the highest function value for the optimization task is taken increases (Reinforcement learning according to the SARSA-(State-Action-Reward-State-Action method)). So the desired results can be achieved faster. A higher effort in training is beneficial later on in the application of the system as the number of necessary loop cycles will be reduced.

An important issue for the training time is also the number of tolerance regions (in this case 3). For a higher number of these tolerance regions a further enhancement in precision is possible. However, the calculation effort would increase over proportionally reaching the processing limit of standard shop floor computers. Thinkable would be a higher precision for one or two features only, e.g. ThrustCoreHead, not for the whole set.

Taking a look at the right diagram in Fig. 5 it can be observed that for the CTM-based system with one Soar program almost 768.000 rules have to be processed per core for each run. For the system with CTM+CFC (4 Soar programs) the number diminishes to 230.000 rules, which is almost 70 % less. For the first case this amount of rules could reach the limit that could be processed with standard PCs typically used on the shop floor, e.g. also the one used for this evaluation. Here the RAM could pose the bottleneck. So the enhanced system is a good option also for computers with fewer cores and less calculation power.

So with the cognitive system enhanced by the CFC it seems feasible that it can be used in the production line for rear axle drives. By this the step from the current CTM-based off-line system is made towards the inline application. Also due to the higher efficiency of the enhanced system even more complex processes could be modelled in adequate time. Furthermore it is thinkable to combine several specific optimization modules modelling single manufacturing processes to an optimization chain.



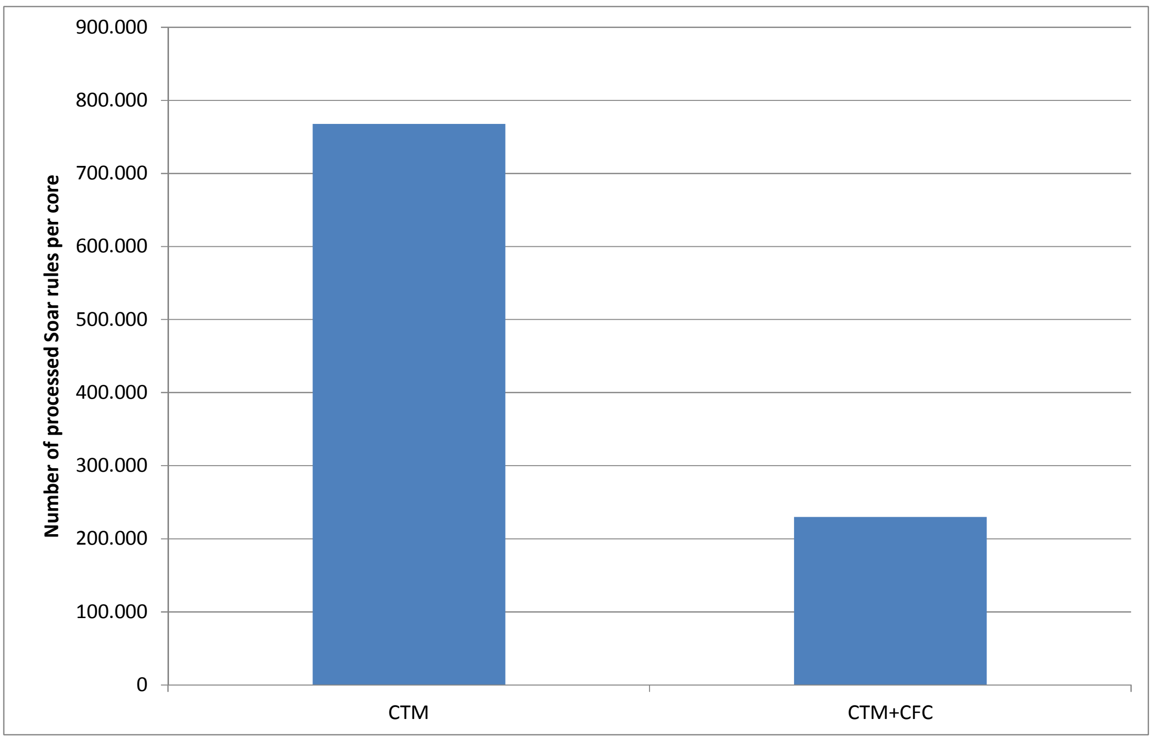


Fig. 5. Average number of required loop cycles vs. Number of training cycles (left) and Number of processed Soar rules per core for old and enhanced set-up of the self-optimizing system (right)

6. CONCLUSIONS

The work presented how a cognitive system with self-optimizing behaviour could be enhanced by integrating a Cognitive Failure Cluster. This does a classification on which a smaller and less complex optimization program based on Cognitive Tolerance Matching can operate. This method offers great autonomy to find an adequate solution but can result in long processing times due to an enormous number of possible solutions. By the classification implausible parameter combinations are eliminated so that the set of processed rules diminishes while keeping up the desired autonomy. This leads to a more efficient method. The enhanced system is applied to the manufacturing of rear axle drives, in special to the optimization of the acoustic behaviour, which is essentially influenced at the final assembly. Subject to the optimization are the assembly parameters face clearance and block dimension. The optimization is based on the evaluation of the contact patterns as an important indicator for the running behaviour of the engaged gears. The contact patterns are acquired by an inline Image processing system. Several geometric features and tolerances for each pattern are defined. Based on the feature deviations the data sets are assigned to certain cluster using the k-means method. Then the optimization run with the cluster-specific program revealed a significant improvement of more than 50 % regarding the number of loop cycles to reach the required precision per number of training cycles. Also by rendering smaller optimization programs the hardware requirements could be reduced.

ACKNOWLEDGEMENTS

The depicted research has been funded by the German Research Foundation DFG as part of the Cluster of Excellence "Integrative Production Technology for High-Wage Countries".

REFERENCES

[1] SCHUH G., KLOCKE F., BRECHER C., SCHMITT R., 2007, *Excellence in Production*. 1st edition, Apprimus Aachen
[2] ORILSKI S., SCHUH G., 2007, *Roadmapping for Competitiveness of High Wage Countries*. Proc. of the XVIII ISPIM Conference: On Innovation for Growth – the Challenges for East and West, Warsaw
[3] SCHMITT R., LAASS M., ISERMANN M, WAGELS C., *Cognitive learning in self-optimization production systems*, in: Proc. of MITIP 2011, Norwegian University of Science and Technology, Trondheim, Norway
[4] SCHMITT R., ISERMANN M.; WAGELS C., MATUSCHEK N., 2010, *Cognitive optimization of an automotive rear-axle drive production process*. Jour. of Machine Engineering, 9/4/71-80
[5] SCHMITT R., NIGGEMANN C., ISERMANN M., LAASS M., MATUSCHEK N., 2011, *Cognition-based self-optimisation of an automotive rear axle drive production process*. Jour. of Machine Engineering, 10/3/68-77
[6] FRANK U., GIESE H., KLEIN F., OBERSCHELP O., SCHMIDT A., SCHULZ B., VÖCKING H., WITTING K. 2004, *Selbstoptimierende Systeme im Maschinenbau. Definitionen und Konzepte*. Paderborn: HNI Verlag
[7] SOAR TECHNOLOGY. http://www.soartech.com, 17.11.2011.
[8] JÄHNE B., 2005, *Digital image Processing*. 6th revised and extended edition, Springer Berlin
[9] GONZALEZ R.C., WOODS R.E., 2001, *Digital Image Processing*, 2nd int. edition, Prentice Hall Int.
[10] EPANECHNIKOV V.A., 1969, *Non-parametric estimation of a multivariate probability density*, Theory of Probability and its Applications, 14/153-158.
[11] MACQUEEN J.B., (1967, *Some Methods for classification and Analysis of Multivariate Observations*, Proc. of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, 281-297.