2016

Changing States of Multistage Process Chains

Thorsten Wuest
West Virginia University, thwuest@mail.wvu.edu

Christopher Irgens
University of Strathclyde

Klaus-Dieter Thoben
University of Bremen

Research Article

Changing States of Multistage Process Chains

Thorsten Wuest,[1] Christopher Irgens,[2] and Klaus-Dieter Thoben[3]

[1]*West Virginia University, Morgantown, WV, USA*
[2]*University of Strathclyde, Glasgow G1 1XQ, UK*
[3]*University of Bremen, 28359 Bremen, Germany*

Correspondence should be addressed to Thorsten Wuest; thwuest@mail.wvu.edu

Generally, a process describes a change of state of some kind (state transformation). This state change occurs from an initial state to a concluding state. Here, the authors take a step back and take a holistic look at generic processes and process sequences from a state perspective. The novel perspective this concept introduces is that the processes and their parameters are not the priority; they are rather included in the analysis by implication. A supervised machine learning based feature ranking method is used to identify and rank relevant state characteristics and thereby the processes' inter- and intrarelationships. This is elaborated with simplified examples of possible applications from different domains to make the theoretical concept and results more feasible for readers from varying domains. The presented concept allows for a holistic description and analysis of complex, multistage processes sequences. This stands especially true for process chains where interrelations between processes and states, processes and processes, or states and states are not fully understood, thus where there is a lack of knowledge regarding causations, in dynamic, complex, and high-dimensional environments.

1. Introduction

In most areas of modern life, especially in business and manufacturing settings, there often is a process perspective. In the majority of cases, unsurprisingly, there is a clear influence of process quality on the process outcome, for example, product or service quality [1, 2]. In this paper, the application domain of the presented approach can be any domain. However, as the approach was successfully applied in the manufacturing and financial services industry as of today, the case studies are depicted from those domains to make the theoretical construct more feasible.

A process defines combinations of actions or events which are elated by causal mechanisms and/or timed actions or events, either internal to the process itself or externally [3] is rather broad. Simply put, the goal of a process is to add value to the product, service, and so forth through a transformation activity (see Figure 1). Often this goes hand in hand with a change of state; in manufacturing this may be mostly a change of physical properties of the product [4, 5]. In other domains like business and healthcare, this may be more diverse with

the input and output being information or knowledge and the state transformation is more virtual.

Processes today are rarely independent from previous and following activities. Those activities can more often than not also be described as being processes. On the other hand, a process can be anything from a single operation to a whole operations sequence. A possible generic view on the used terminology in this paper is depicted in Figure 2.

The aforementioned change of state typically happens in a progression of successive states until the final state of the focal object, for example, product or service. The final state is generally defined by the requirements towards the object from the customers' side and specific to the individual programme. Once the final state is reached, the object is considered ready for delivery to customers. The progression of states from start to finish can thus be observed; each state presents an individual "picture" of a process' precondition and postcondition.

A major challenge of today's optimization regarding process sequences is the increasing complexity and dynamic nature of processes. In a succession of processes and thus

FIGURE 1: Transformation model in manufacturing and value creation.

states, it has been shown (e.g., [6]) that different processes are connected across the whole programme through a variety of relations influencing the outcome significantly. Often these relations are not considered even though their influence can cause major discrepancies with regard to, for example, scrap and rework. Reasons for this are manifold, ranging from ignorance to lack of knowledge to technical limitations and budget considerations.

In this paper, a concept is proposed that targets this issue of, currently ignored, implicit inter- and intrarelations across multistage process sequences. It defines complex, dynamic, and high-dimensional process environments through a series of progressing states and the subsequent analysis of this enriched process picture. This analysis is done by state of the art supervised machine learning based analysis which allows taking, at times hidden, implicit relations into consideration in a simple and economic way. This ease of application is a prerequisite for successful employment in the targeted domains, like, for example, manufacturing.

The paper is structured as follows: first, the theoretical state concept is developed step by step before the data model, analytical methodology, and results are presented in Section 3. These two sections are purposely held rather generic and neutral to highlight the broad applicability in different domains defined by similar characterizations, like, for example, complexity, high-dimensionality, and dynamics. Following that in Section 4, the simplified, theoretical results from Section 3 are discussed by mapping them on real applications in three different scenarios to allow the reader to put them into perspective. Following that, a critical discussion of the limitations and benefits of the proposed method is illustrated before Section 5 which concludes the paper and gives an outlook on further research destinations.

2. Development of the Generic State Concept

Any process, physical or otherwise, creates a state change (state transformation) on an object upon which the process works. It is possible to view a process as having an initial state (input) and a concluding state (output). The process may work on concrete and physical objects, as well as on abstract objects. Processes can be found both within and outside of human activities. An example of a process is the following: a manufacturing process produces certain artifacts from material(s), thus transforming the material object's state from an initial state to a concluding state. Other process examples function differently, such as those in human services. These processes include education, medical health services,

banking, and tourism. Within these processes, it is once again possible to identify a process' transformation of an object's initial state to a concluding state. Naturally, it is common to sequence processes. This is so that successive processes transform an object's initial state to a conclusion state through a number of successive process steps.

In Figure 3, there are two processes implied affecting two state changes from initial state, State 1, to concluding state, State 3. Exactly what the processes do and what the objects are are irrelevant here. What is relevant is that the effective transformation from the initial state to the concluding state has been completed. The final result of the implied processes may be in some way ascertained by analyzing the concluding state, State 3. It should be observed that the intermediate state, State 2, is expressing the progression from the object's initial state towards the concluding state. Therefore, in relation to that, it should be noted that it may be possible to ascertain the success of the progress from State 1 to State 3. It should also be noted that the concluding state in this case may be the initial state for a subsequent set of state transformations. Similarly, the initial state, State 1, may be the concluding state of a previous state transformation. In this way we may use state space to observe the actions and progress of the underlying processes.

2.1. States. An object's state may be characterized in a number of ways. A physical object's states may be characterized in terms of dimensional measurements and other physical parameters. An abstract object's state may be characterized by measurements/assessments of abstract parameters such as profit and loss, ability/level of skill, and profitability. Thus each state will be characterized by its state characteristics (SCs).

$$\text{State} = \{\text{SCs}\} \text{ such that}$$
$$\text{SCs} = \{\text{SC}_1, \text{SC}_2, \ldots, \text{SC}_n\}. \tag{1}$$

This allows for accumulating states within the state space, for example, depending on availability or relevancy:

$$\text{Accumulated\_State\_Space}_i = \{\text{SC}_{11}, \text{SC}_{12}, \ldots, \text{SC}_{1k},$$
$$\text{SC}_{21}, \text{SC}_{22}, \ldots, \text{SC}_{2n}, \text{SC}_{i1}, \text{SC}_{i2}, \ldots, \text{SC}_{im}\}, \tag{2}$$

where k, n, and m are the number of SCs for each included state within the accumulated state space. In this manner, as required for analysis, it is possible to select the initial state and the concluding state at any state point, thus grouping states and therefore processes flexibly. This is reflecting the need to include the intermediate states (e.g., "State 2") as component states between the initial state "State 1" and the concluding state "State$_i$" (see Figure 4).

2.2. Complex Process Sequences. There are many situations in which process sequences must be analyzed. These analyses serve as tests to see whether their actions and progression of the process sequence are continuing as planned (or required), in order to produce the required result. This applies to all types of processes, whether they work with physical objects and/or abstract objects.
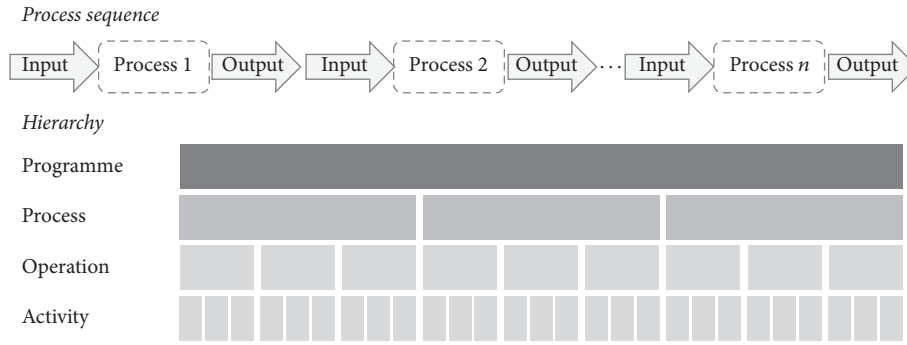
Figure 2: Process sequence and hierarchy (adapted from [7]).
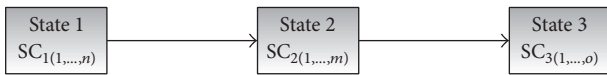

Figure 3: Basic three-state model.

Figure 5: Basic three-process model with disruptive state transformations implied processes.

Figure 4: Basic three-state model with descriptive state characteristics (SCs).

Figure 6: State indicates relationship between successive processes by implication.

It is now possible to treat a number of successive processes as one state space entity, that is, a state transformation from initial state to concluding state with intermediate states.

In this case {*State 1*, *State 2*} can be one state space for analysis, as can {*State 2*, *State 3*} and {*State 1*, *State 2*, *State 3*}. This now imposes the assumption of time, as only successive states can be grouped in this way. Thus successive transformations imply successive processes and therefore continuous time progression.

Conventionally, using a process output state enables the monitoring and control of the processes. This is established through a manufacturing process monitoring and control method; thus using successive states should enable the monitoring and control of successive processes. Introducing time, it is thus easy to see how {*State 1*, *State 2*, *State 3*} defines the progression of the successive processes 1 and 2 (see Figure 5). Compounding the successive subsets of the states will therefore define partial progress such that (*State 1* + *State 2*) gives the progress of process 1 from *precondition* to *postcondition*. (*State 2* + *State 3*) gives the progress of process 2 from precondition to postcondition, while (*State 1* + *State 3*) gives the progress from precondition of process 1 to postcondition of process 2, where "+" indicates the set operation of union. One can then view process 1 and process 2 as a single process defined by the states "State$_i$," *where i* = 1, 2 *or i* = 2, 3 (see Figure 6). The final state is defined by the needs and requirements of the stakeholders, for example, customers (see Figure 7).
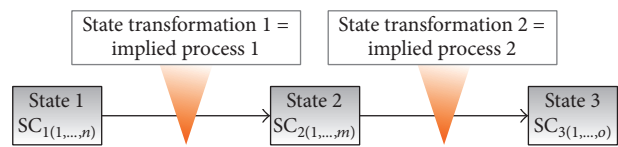
2.3. Accumulation of States along a Process Sequence.
In Figure 8, the different perspectives of the simplified three-process model are summarized. Figure 8(a) just shows the individual states along a process chain. However, they are not connected and, thus, neither accumulated nor implicit information (e.g., from (inter-)relations) is considered at that point.

Looking at Figure 8(b), the accumulation of successive states along the process chain describes the main argument for a state perspective as described in this paper. The different colours (red, yellow, and blue) indicate the growing and maturing of the state vector along the process chain (time). As the vector grows from "red" to "yellow" and "blue" it includes more and more of the hidden causal chains which influence the "final state." This final state is in most process chains the one on which the quality considerations are based upon and thus it is highly important to assure the requirements are met.

The increasingly rich picture that is being generated about the progressing state as we accumulate the vectors includes not only the state information but also implicit information
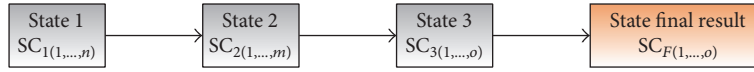
FIGURE 7: State model with final state (final result).

(a) States

(b) Accumulated states

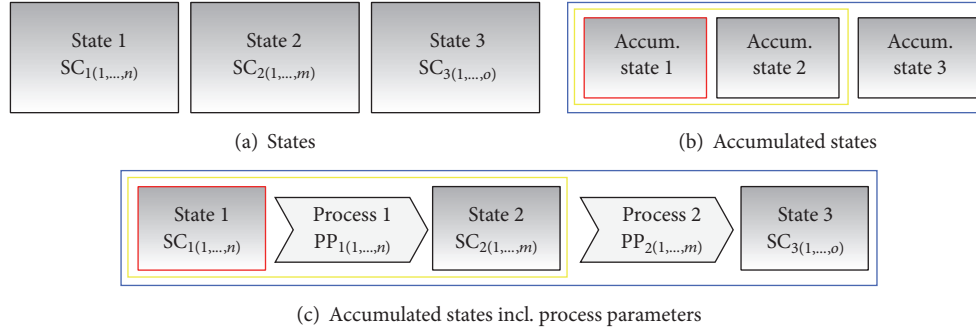(c) Accumulated states incl. process parameters

FIGURE 8: (a) Three-state model. (b) Three-state model with accumulated states. (c) Three-state model with accumulated parameters including process parameters.

about the relation between states (causal chains). By basing the analysis on accumulated states rather than individual ones, state drivers can be identified and used for further optimization/knowledge generation, and so forth, within the process chain. This does not necessarily mean the explicit capture/identification of those drivers, but rather implicit consideration through KPIs/KPVs/drivers/correlating factor, and so forth. Of course, where the vector elements/parameters are independent variables, they might represent a direct causal driver; otherwise they may represent "indirect" ones, that is, the one which is itself being driven. However, this can be considered a valuable insight into what can be very complex causal chains.

The final Figure 8(c) illustrates that although it is not always necessary, it is possible to include process information in form of, for example, process parameters (PP), in the state vector too. Adding additional process (or environmental, etc.) parameters will allow the creation of even a richer picture of the developing state along the process chain and thus capturing even more implicit relations.

3. Data Model, Methodology, and Application Results

In this section the previously discussed theoretical state concept is applied in a generic example. In order to allow for a broad applicability, the data set used is synthesized and designed as simple as possible to show the principles behind the state concept. In Section 4, different case studies referring to the application results in this section highlight the opportunities of the concept in a wide variety of domains.

3.1. Data Model. The synthesized data model is designed to represent a process sequence consisting of three processes, thus representing three states. The following parameters were used to generate the data set:

(i) There are 3 sets/processes (set size: 1000; dimensionality: 5).

(ii) Each process data set is generated as a random uniformly distributed set, with mean = 0.5 and incorporating two distinct clusters.

(iii) The data sets are noise-free and the values are normalised in the range (0-1).

The synthetic data sets were created using KNIME. The objective was to create a very simple data set for illustrative purposes, yet a data set that would be realistic enough to show the use of a suitable supervised learning method. Thus, the focus of this paper at this point is to show the applicability with relative complexity and data sets as they may be commonly found in "real world" processes, in order to highlight the principles and potential of the proposed state concept. Later in the discussion section, the application of the method and the results of an application in a more complex, dynamic, and high-dimensional environment are presented to extend this simplified theoretical explanation.

For each of the processes/states the dimensionality was chosen to be five. It would have been possible to vary the number of features/parameters per state. However, this would have added unnecessarily to the complexity and would contradict the goal of simplicity in illustrating the application.

Process 1 (TOM) is as follows:

$$f_{10} \quad f_{11} \quad f_{12} \quad f_{13} \quad f_{14}. \tag{3}$$

Process 2 (DICK) is as follows:

$$f_{20} \quad f_{21} \quad f_{22} \quad f_{23} \quad f_{24}. \tag{4}$$

Process 3 (HARRY) is as follows:

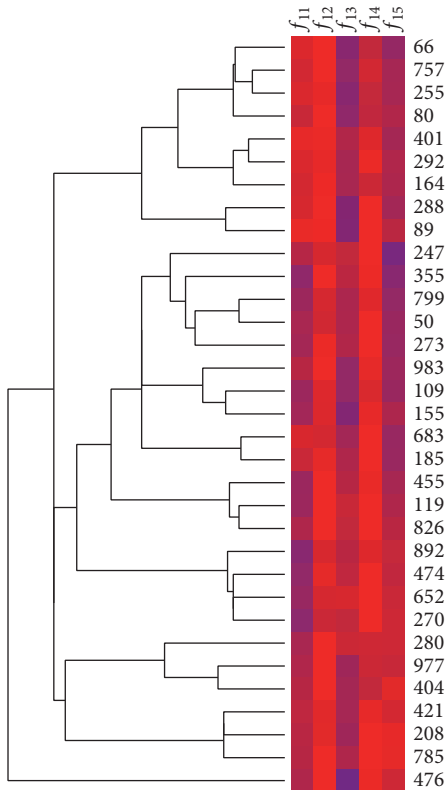$$f_{30} \quad f_{31} \quad f_{32} \quad f_{33} \quad f_{34}. \tag{5}$$

FIGURE 9: Exemplary cluster analysis performed on vectors TOM.

f_{ij} indicates feature i of process j. The total set of $\{f_{ij}\}$, $i = 1, \ldots, 3$, $j = 0, \ldots, 4$ is 1000 vectors. Each vector is identified as vector 1 to 1000, to establish a time framework such that the sequence $1, \ldots, 1000$ denotes a sampling sequence.

From this it should be clear that the accumulated vectors TOM_DICK (TD) (State 1 and State 2) and TOM_DICK_HARRY (TDH) (State 1, State 2, and State 3) are a combination of the two, respectively, three, individual process vectors TOM (State 1), DICK (State 2), and HARRY (State 3). As for those combined vectors (accumulated states) no labels are available. In order to simulate postprocess inspection and associated vector labelling into {*good, bad*}, the following approach was taken.

A hierarchical cluster (agglomerative) analysis was performed and the labels were awarded based on the cluster population (see Figure 9). This should emulate postprocess inspection based upon unrecorded parameters, for example, inspection of weight, colour, smell, sound, or other unrecorded product parameters. In this case clearly outlying clusters were labelled "*bad*."

This allows classified vectors, which are needed for the following analysis using supervised machine learning on all process states TOM, TOM_DICK, and TOM_DICK_HARRY (T, TD, TDH), represented by the various accumulations of the feature vectors.

3.2. Methodology. In this section, a methodology to analyze a data set based on a process sequence as described above is introduced. The objective of the presented methodology is to

derive information from the data set to increase transparency and advance the knowledge about the overall system. By using advanced machine learning techniques, it is possible to utilize "hidden" information that would normally not be taken into consideration.

A key factor which differentiates this methodology from many other is the combination of advanced machine learning techniques with strategic preparation of the data set according to the state view described earlier. This has to be done prior to the application of supervised machine learning based feature ranking as described below, ideally as part of the data preprocessing.

Machine learning promises the ability to handle complicated optimization problems. It is able to handle high-dimensional data in a complex and dynamic, even chaotic environment rather well depending on the algorithm used [8, 9]. ML algorithms provide the opportunity to learn from the dynamic system and adapt to changing environment automatically to a certain extent [10, 11].

A common metaclassification of machine learning algorithms is "unsupervised," "supervised," and "reinforcement" machine learning [12, 13]. Whereas for unsupervised machine learning the data is not labelled, thus no feedback if provided from an external source, supervised machine learning, and Reinforcement Learning (RL) rely on external feedback. RL is based on an evaluation of a chosen action whereas for supervised machine learning, the correct label is provided by a teacher [12].

Supervised machine learning was found to be a good fit for problems and application with access to labelled data and available expert feedback. A domain with such properties is, for example, manufacturing [11]. In the application cases of this methodology, the availability of labels and expert feedback is assumed. Thus, supervised machine learning is focused on from here on.

There are several supervised machine learning algorithms and variations available today. Each of them has distinct advantages and challenges. Kotsiantis [13] compared several algorithms according to different dimensions. A rather promising algorithm for the illustrated application of this methodology is *Support Vector Machines* (SVM), introduced by Cortes and Vapnik [14] as a new machine learning technique for two-group classification problems. Burbidge et al. [15] found SVM to be a "robust and highly accurate intelligent classification technique well suited for structure-activity relationship analysis." SVM can be understood as a practical methodology of the theoretical framework of statistical learning theory (STL) [16]. The idea behind it is that input vectors are nonlinearly mapped to a very high-dimensional feature space [14]. SVM can be combined with different kernels (e.g., neural networks; Anova) [17]. Hence, SVM is a very adaptable algorithm, suitable for a broad range of applications, requirements, and problem characteristics (Table 1).

SVM as a classification technique has its roots in statistical learning theory [18, 19] and has shown promising empirical results in a number of practical manufacturing applications [20, 21] and works very well with high-dimensional data [19, 22–26]. Another aspect of this approach is that

Table 1: Characteristics and corresponding requirements of Whisky production.

Characteristic	Requirement
Appearance	Clear and colourless
Wholesomeness	Freedom from taint
Mineral salt and metallic contents	Contents that meet all process requirements
Microbiological standard	Reducing water must be of drinkable standard
Reliability of supply	Water must be available at all times

it represents the decision boundary using a subset of the training examples, known as the support vectors.

In order to identify the relevant state drivers of the accumulated state vector, a ranking of the descriptive state characteristics and process parameters (summarized as features in machine learning terms) is necessary. SVM has the possibility to take the different, often just implicitly known process, intra- and interrelations into account. This can be done as a simple matter using SVM. The general idea behind this *feature ranking*, also known as feature selection, is to identify or rank "relevant" characteristics which are either able to represent a system through generalization or are important to monitor as they may allow prediction of a certain (future) outcome/behaviour. This enables the identification of relevant state drivers in multistage process sequences.

There are several feature ranking methods available in literature, for instance, the method of SVM based feature ranking using recursive feature elimination. This technique was first introduced by Guyon et al. [27] and has been successfully applied in various other scenarios [28]. The findings indicate that the performance of this feature ranking method, with its relatively simple application, is very high and even outperforms several more complicated casual discovery methods. This method was especially created for situations in which the number of features (dimensionality) is very much higher than the number of vectors. Especially in an environment characterized by complexity, dynamical behaviour, and high-dimensional data SVM based feature ranking algorithm provides good results. It can be concluded, that SVM, as a classification method based on maximizing the margin between two groups of data points is suitable for the task of identifying state drivers within a process sequence of a system. As stated before, the soft, maximum margin SVM algorithm, as a population separator and state classifier, is the key advantage of the SVM algorithm.

In application, the challenge lies in transferring the relationships of state characteristics and process parameters in the algorithm and being able to interpret the results accordingly. The classification hyperplane, being constructed in the multidimensional space, is able to reflect these relationships implicitly. Thus, SVM utilizing the hyperplane allows for classification in multidimensional space and furthermore deriving relevant state drivers. These state drivers are (partly) responsible for or have a strong impact on, a change in

Feature weight T		Feature weight TD		Feature weight TDH	
f_{13}	2.360497	f_{13}	1.812961	f_{10}	0.71884
f_{12}	2.037506	f_{12}	1.348599	f_{11}	0.470044
f_{14}	0.122294	f_{14}	0.198171	f_{34}	0.42577
f_{11}	−0.54468	f_{23}	0.123906	f_{33}	0.255275
f_{10}	−2.00167	f_{24}	0.073222	f_{21}	0.249973

Figure 10: Feature ranking of the accumulated T, TD, and TDH state vectors sorted according to ranking. The features highlighted by blue and red show interesting shifts in importance from state to state.

class, which in this case would translate to a change between desirable or undesirable state (e.g., "good"/"bad"). Following that, the results of the methodology application on the data set are presented before three scenarios from different domains make the concept more feasible.

3.3. Application Results. In this section, the results of applying the method and concept described above on the data sets are depicted. This serves as a theoretical foundation for the subsequent sections presenting domain specific application scenarios in the discussion section.

3.3.1. Feature Ranking. Figure 10 depicts the resultant ranking of the features for process TOM (T) as well as the accumulated processes TOM_DICK (TD) and TOM_DICK_HARRY (TDH). The ranking was derived by ranking the SVM calculated feature weights. The following should be noted:

(1) The SVM learning set was the total vector population of 1000, labelled as indicated above

(2) The learning phase was carried out using a 10-fold cross-validation approach in order to minimise overfitting; overfitting occurs when the number of features (dimensions) is large relative to the number of vectors. In this case overfitting is not considered a serious threat, but nevertheless the cross-validation approach was taken in order to minimise any risk. The subsequent feature ranking was done by the recursive feature elimination method proposed by Guyon et al. [27].

(3) The labelled data sets were *oversampled* with respect to the minority class in order to achieve balanced data sets. As a result the learning data sets varied in size from approximately 1200 vectors to 1500.

The performed cross-validation results in form of confusion matrices show good separation performance using a linear kernel as used in the feature ranking method following Guyon et al. [27]. Furthermore, no indication of overfitting was observed, partly due to the balanced nature of the data set. Therefore, the application of the outlined feature ranking method is feasible in this case. The rankings for the individual processes TOM (T), TOM_DICK (TD), and TOM_DICK_HARRY (TDH) are provided in Figure 10. Given the large ratio of number of vectors to features (dimension) these rankings are relevant for determining the individual feature's importance at each state.
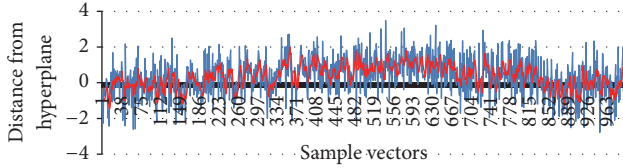
FIGURE 11: State drift of TOM_DICK_HARRY (TDH) over time based on "state strength." The slower variant shows a 5-sample moving average.
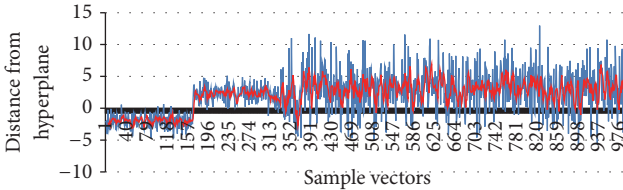

FIGURE 12: State drift of TOM (T). This shows a clear shift from *"bad"* process performance to *"good"* at approx. sample 170.

3.3.2. State Drift. In Figures 11 and 12 the state drift is based upon the "state strength" which is derived from the individual vector's distance from the relevant hyperplane. For this purpose the calculated functional value of the SVM kernel model is used. This is because $d(\mathbf{x}_i) = g(\mathbf{x}_i)/\|\mathbf{w}\|$, where $d(\mathbf{x}_i)$ is vector $\mathbf{x}_i$'s distance from the hyperplane, $g(\mathbf{x}_i)$ is the calculated classification function, and $\mathbf{w}$ is the hyperplane vector. $\|\mathbf{w}\|$ is a constant and therefore $g(\mathbf{x}_i)$ can be used as a measure of the distance $d(\mathbf{x}_i)$. Thus, the individual vectors' distance from the hyperplane was plotted and is shown in Figures 11 and 12. The plots show the distances in the vector sequence, emulating the process sampling sequence; the larger the distance, the "stronger" the process vector's membership in class *"good,"* that is, the more likely the process is to produce *"good"* output.

As mentioned above, it was decided to build into the process data sets two distinct clusters. This is reflected in the results (see Figure 12) which shows an abrupt state change. However, it is also possible to have a more slowly progressing state change (see Figure 11). The approximating state drift shows a slow drift from "acceptable" state to "unacceptable" state. Both can be found in "real world" process sequences and allow for different conclusions.

3.3.3. Parameter Shift. It can be observed that the "relative importance" of individual process parameters (features) may change with successive states, leading to a so-called parameter shift. "Relative importance" means in this context that the ranking position of a certain parameter (feature) of a previous state is not necessarily reflected in a successive state's ranking.

In this case, this can be illustrated by looking at parameter f_{10} (see Figure 10). This feature belongs to the first process TOM (T). In the first observed state, (T), f_{10} is ranked at the 5th position of all the process features. In the subsequent accumulated state TOM_DICK (TD), the same parameter is ranked below 5th. However, the resultant importance of f_{10} shifts to the 1st position in the final state
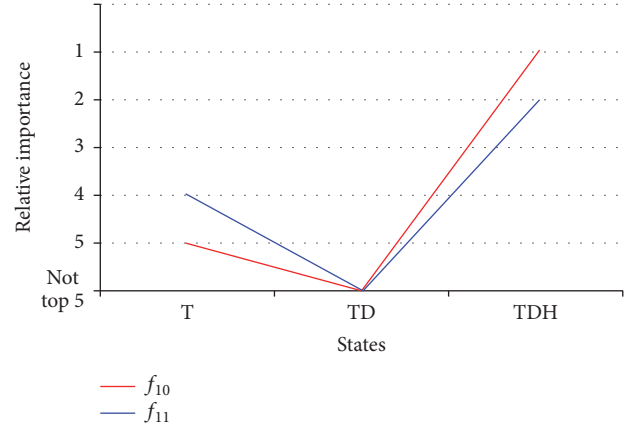

FIGURE 13: Parameter shift along the process sequence; the vertical axis indicates the relative importance by ranks 1–5.

TOM_DICK_HARRY (TDH), having originally been the least influential parameter of the original TOM (T) process set. Similar shift can be observed for parameter f_{11}, also the 3 most important parameters for T and TD do not appear among the 5 most important parameters for TDH.

It should be noted that in this situation there are 3 different hyperplanes being used, a different one for each T, TD, and TDH state of the process chain. This comparison of parameters, in this example, f_{10} (red) and f_{11} (blue), within their original setting and the relative shift of importance among each other with progressing state, is regarded as "relative importance." This sample parameter shift is depicted in graphical form in Figure 13.

This "relative importance" provides interesting insights in the interprocess relations influencing the process outcome (final state). For stakeholders of these process sequences, such additional information may prove beneficial depending on the domain and the domain specific mechanics. In the later presented application examples, this paradigm will be depicted from a more practical perspective.

In the following, the so-far rather theoretical results are discussed with a focus on application. After a more general discussion, three domain specific application scenarios are introduced, utilizing the previous results within the domain specific context.

4. Discussion

In this section, the theoretical results of the methodology's application on the showcase data set are discussed. Previously, different obtainable results were presented in a generic way. In the following, the three different application scenarios are projecting the challenges and requirements of varying domains on the theoretical results of the methodology application, in order to indicate its broad applicability. Matching the selected theoretical outcomes on real-life problems in the different domains illustrates the possible benefit of the developed methodology. After the brief description of the three scenarios, the results are critically discussed and the limitations of the approach are presented.

4.1. Domain Specific Application Scenarios. The selected theoretical analysis results and their counterparts in "real-life" application are discussed below. The three scenarios compromise industrial manufacturing, whisky production, and financial market analysis.

4.1.1. Manufacturing. In complex, multistage manufacturing processes, such as is common in semiconductor [29] and chemical [30] and mechanical manufacturing, the accumulating stream of data and information presents a challenge to process analysis and control. However, there is a growing need for process monitoring and control as well as continuing optimization. In the following, the previously presented results are put into perspective within the manufacturing domain.

The benefit of the analysis of relevant state drivers as depicted in Figure 10 is rather easy to comprehend in a manufacturing environment, at least when considering the individual processes and states. The most relevant features have the highest relevance on the chosen outcome parameter (e.g., quality). So this can be directly utilized in optimization activities.

The benefit of the results of the feature ranking of the accumulated state vectors is not as clearly derivable by common sense compared to the one for individual states. However, the results include the important implicit and explicit process inter- and intrarelations. This is reflected in the ranking and thus these influences can be taken into consideration with relative simplicity during optimization activities. This is most likely the most important aspect in industrial application; the presented method offers additional potential learning outcomes.

Looking at the state drift presented before (see Figures 11 and 12), the repercussions on an industrial manufacturing application can be observed. The drift of manufactured products from a desirable to an undesirable state can happen in different variations. The variations approximating and distinctly depicted in Figure 11 may allow for a prediction of future problematic situations in the manufacturing process. Thus, maintenance and analysis activities may be triggered before a potential problem arises. However, in some cases the behaviour can be considered chaotic, in which case a useful prediction is rather difficult. These situations may be caused by, for example, tool wear (approximating), tool tear (distinct), and environmental impact like vibrations (chaotic).

The parameter shift, depicted in Figure 13, presents an interesting analysis tool and insight in the manufacturing process sequence and the causal mechanisms between processes. The importance of individual parameters' shift in comparison to their peers throughout a process sequence can be observed. A parameter being of high importance during the early stages of the manufacturing programme may even be irrelevant at a later stage, again regarding a previously decided outcome like quality. However, the importance of parameters can also switch between "important, less important, and important" along a three-stage process. This may reflect the complex interrelations of modern manufacturing processes. A practical example of such a setting may be the

internal stress induced during clamping with a three-chuck-jaw during the process of machining. Whereas this internal stress allocation has no effect on several processes, it becomes very important with regard to the quality outcome during heat treatment (incl. quenching). In that later process, the previously induced stress can lead to a significant deformation [6, 31].

The simplicity of applying the presented model and the timely generation of results corresponds very well with the requirements of modern manufacturing process chains with their complexity and dynamic environment. The results not only benefit practitioners directly by presenting focal points for optimization activities but also may act as a starting point for in-depth analysis that will lead to a better understanding of the implicit relations which are currently unknown to a large extent. With products, materials and processes are becoming ever more optimized; this contribution to knowledge and transparency can be considered highly relevant.

The practical application of process parameter (feature) selection has been used in the manufacture of aeroengine parts. Critical process parameters have been identified in a component forming process. The focus of process analysis was the component defects with the view to identify important features which can be seen as "drivers" for the various defect types as well as their locations on the component. Again, the approach defined in the section on feature ranking and Figure 10 applies. In this case there is only a single process whose features are ranked in order to determine the features which can be seen as defect drivers. The results look promising, at the time of writing this paper.

4.1.2. Financial Portfolio Management. The financial market is a complex, evolutionary, and nonlinear, chaotic dynamic system. The field of financial forecasting is characterized by noisy data, unstructured nature, high degree of uncertainty, and hidden relationships [32, 33]. A large number of market factors interact, including political events, micro- and macroeconomic conditions, and traders' expectations and "herd instincts." This creates in effect a large number of process parameters (features) which operate on, and are effected by, different parts of the market independently

Each day, a portfolio will show an accumulated result of the market process, exhibiting an overall profit or loss from the day/period. The continuously accumulating portfolio effect can be viewed as a "process regime" valid for the portfolio as a whole and the "drivers" of the regime can be determined using the supervised machine learning method such as SVM. Thus, the consequential changes in the importance of the individual portfolio holdings with respect to the resulting profit/loss are reflected in the changing ranking of these parameters (features). Looking at the state drift presented before (see Figures 11 and 12), the repercussions on a financial portfolio management process can be observed. The drift of profitable portfolio to a loss making portfolio can be determined in the same way as shown in Figure 12.

The method has successfully handled from 150 to 500 dimensions in a financial assets management application. It has been used daily for portfolio sizes of a few hundred

positions, to analyze financial portfolio positions. Processing is done on a 2.6 GHz CORE i7 CPU, 8 GB, laptop machine for portfolios containing up to a couple of hundred positions.

4.1.3. Whisky Production.

The whisky production chain contains the main processes of malting, mashing, distilling and storage. These processes are dependent on specific process parameters which affect the final product in what may be seen a directly local process outcome and thus an indirect accumulated or delayed effect during later processes. Simple observations from this industry could illustrate this.

Firstly, the characteristics of the water affects the characteristics and thus the quality of the whisky produced, and this fact has resulted in some areas being famous for their whisky. The quality of the water can also affect the efficiency of all the processes. Thus it may be seen that certain water parameters will remain important "drivers" throughout the whole process sequence.

Secondly, the grain must be stored so that it does not start *uncontrolled germination*. Storage happens when the barley has been dried to approximately 12% moisture. It is transferred to storage, without cooling, to facilitate dormancy breaking. The temperature of the grain should be between 18°C and 25°C at this stage. In the case of severe dormancy, it has been suggested that the temperature should be increased to around 30°C for a short period as a way of lessening the duration of the dormancy. During this period of storage samples are taken for determination of germination energy, and once dormancy has fully broken the grain bulk is gradually cooled to an ambient temperature of about 5°C for storage. Volumes of air of around $0.15 \, \text{m}^3$/min per ton are normally used to cool barley. During this part of the process, temperature and moisture control are important and can have indirect effects during the later processes.

Finally, the "malting" of the barley is the first main process. The malting stage is crucial in determining production costs as the sugar levels created dictate the alcohol yield further down the line. Oversteeping, especially with inadequate aeration, can lead to uneven germination, thus effecting this dramatically. The required levels of enzymes can be produced by using multiple wettings to achieve a moisture content of at least 46 per cent. Abrasion as well as related processes, in which the end of the barleycorn is damaged, allowing faster ingress of water is one way of managing the process. It can thus be seen that the malting parameters of controlling the water uptake and thus enzyme production becomes a new final product quality "driver" which may in turn overshadow water quality and critical storage parameters.

As the whisky processes proceed, the individual importance of each process' parameters can be reflected by the accumulated process knowledge. Mapping this onto Figure 8 it corresponds with *process 1* being the *controlled storage* process and *process 2* being the *malting* process. The water characteristics and the grain characteristics form the initial conditions as represented by State 1. The individual rankings of each process' parameters at the end of the *malting* process are then expressed by the accumulated states in the yellow box in Figure 8(b).

4.2. Discussion.

The presented methodology addresses challenges which have been, and are, emerging in many applications where complex and dynamic systems require high-dimensional data. The developed concept brings together the object and process perspective in a multistage system and, most importantly, allows taking implicit relations along the process sequence into account by means of an accumulating state vector. A major asset of the concept is the possibility to identify currently unknown relations which may provide a basis for further, in-depth research and experimentation.

The strength of the approach is its relevance, simplicity, and efficacy and its applicability in actual and realistic applications. The selected approach and methodology is such that it is not constraint to specific objects, processes, sequences or systems nor is it constraint to specific outcomes. The outcomes may be of a quality nature but can also be defined as seen fit by the owner.

4.3. Limitations.

Features which are not available are not included in the analysis. Thus, if something is not measured, accessible. or communicated it will not be identified as relevant and may dilute the results to some extent. Furthermore, as in most data applications, the preprocessing has a large impact on the outcome. Depending on the original data quality, for example, the amount of missing data or noise, the application of the method and finally the results may be affected.

The practical limitations of the method are the commonly observed problem of reliable and readily available data. Industrial data collection is typically beset with noise and incomplete data due to operator and sensor problems. This leads to considerable effort needed to preprocess the data. Also, process knowledge is required in order to effect a feasible data selection for the training data. In the abovementioned engineering/manufacturing cases, considerable effort was expended on "cleaning" and completing the process data, as well as on creating a sufficient level of process understanding to generate process features and set up the data selection rules and use these. It is important to note that establishing classified vector sets for each state (T, TD, and TDH) is important as is the maintenance of these. Given the routine procedures of quality inspections, especially of high cost products, this is not considered problematic. Once the training data is established the limitations become a practical computing limitation which is a function of data dimensionality and data volume. This limitation has not been a problem with cases of dimensionality to 500 and corresponding data vector volume of 10000, other words considerably beyond the present example's bounds.

5. Conclusion and Outlook

The work so far has indicated the practical application and evaluation of the methodology to a limited process spectrum, manufacturing, and financial services. When looking at the thus evaluated cases, one observes that the processes and their characteristics, for example, complexity, dynamics, and high-dimensionality as well as inter- and intrarelations along the process sequence, seem rather universal. This indicates that

the proposed methodology could be successfully investigated for a variation in process types and domains. There are practical issues associated with the further development of this.

The first is addressing the problems of interpreting raw process data. Such data is most often based on very low level information such as simple temperature readings exchange rate values and similar instant low level information. This can be generally described as challenges in feature generation. It is considered necessary that higher level process features are generated based on such data prior to analysis. For instance, instantaneous process parameter readings are transformed into inflexion points, first differentials with respect to time, integrated impulse values, and so forth. Investigation into the automatic versus manual approach to this is important.

The second issue is the investigation into methods for selecting suitable machine learning parameters with regard to the data feature characteristics. This is highly relevant as real world data almost always presents distinct challenges like unbalanced and/or nonseparable data. This leads to further issues, like how and when to use Synthetic Minority Oversampling Techniques (SMOTE) [34] to handle such issues. The presently reported work has been largely driven by manual selection and set-up of parameters and features.

Additional Points

Highlights. Processes describe a change of state of any kind. By accumulating subsequent processes, exploitable information can be enhanced. Analyzing accumulated states may lead to variety of possible results. With this generic concept, process sequences of various domains may be analyzed.
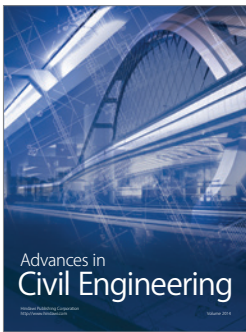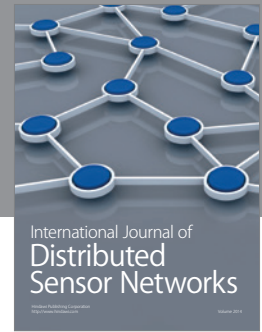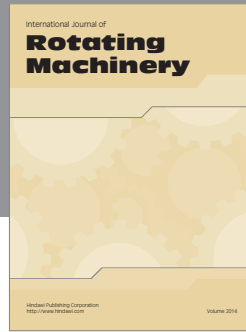
Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] E. Brinksmeier, *Prozeß- und Werkstückqualität in der Feinbearbeitung*, Fortschritt-Berichte VDI, Reihe 2: Fertigungstechnik, VDI-Verlag, Düsseldorf, Germany, 1991.

[2] J. Jacob and K. Petrick, "Qualitätsmanagement und Normung," in *Masing Handbuch Qualitätsmanagement*, R. Schmitt and T. Pfeifer, Eds., pp. 101–121, Carl Hanser Verlag, München, Germany, 2007.

[3] M. Hoffmann, T. Goesmann, and A. Kienle, "Analyse und Unterstützung von Wissensprozessen als Voraussetzung für erfolgreiches Wissensmanagement," in *Geschäftsprozessorientiertes Wissensmanagement*, A. Abecker, K. Hinkelmann, H. Maus, and H. J. Müller, Eds., pp. 159–181, Springer, Berlin, Germany, 2002.

[4] S. Kumar, *Intelligent Manufacturing Systems*, B.I.T. Mesra, Ranchi, India, 2002, http://pchats.tripod.com/int_manu.pdf.

[5] S. Kalpakjian and S. R. Schmid, *Manufacturing Engineering and Technology*, Prentice Hall, New Jersey, NJ, USA, 2009.

[6] J. Sölter, "Relationship between strain distributions and shape deviations of rings caused in clamping," *Materialwissenschaft und Werkstofftechnik*, vol. 43, no. 1-2, pp. 23–28, 2012.

[7] T. Becker, *Prozesse in Produktion und Supply Chain Optimieren*, Springer, Berlin, Germany, 2nd edition, 2008.

[8] K. Yang and J. Trewn, *Multivariate Statistical Methods in Quality Management*, McGraw-Hill, New York, NY, USA, 2004.

[9] G. Köksal, İ. Batmaz, and M. C. Testik, "A review of data mining applications for quality improvement in manufacturing industry," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13448–13467, 2011.

[10] H. A. Simon, "Why should machines learn?" in *Machine Learning: An Artificial Intelligence Approach*, R. Michalski, J. Carbonell, and T. Mitchell, Eds., pp. 25–38, Tioga Press, Charlotte, NC, USA, 1983.

[11] S. C.-Y. Lu, "Machine learning approaches to knowledge synthesis and integration tasks for advanced engineering automation," *Computers in Industry*, vol. 15, no. 1-2, pp. 105–120, 1990.

[12] L. Monostori, "AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 4, pp. 277–291, 2003.

[13] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.

[14] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[15] R. Burbidge, M. Trotter, B. Buxton, and S. Holden, "Drug design by machine learning: support vector machines for pharmaceutical data analysis," *Computers and Chemistry*, vol. 26, no. 1, pp. 5–14, 2001.

[16] V. Cherkassky and Y. Ma, "Another look at statistical learning theory and regularization," *Neural Networks*, vol. 22, no. 7, pp. 958–969, 2009.

[17] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural Computation*, vol. 15, no. 7, pp. 1667–1689, 2003.

[18] R. Khemchandani, Jayadeva, and S. Chandra, "Knowledge based proximal support vector machines," *European Journal of Operational Research*, vol. 195, no. 3, pp. 914–923, 2009.

[19] K. Salahshoor, M. Kordestani, and M. S. Khoshro, "Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers," *Energy*, vol. 35, no. 12, pp. 5472–5482, 2010.

[20] R. B. Chinnam, "Support vector machines for recognizing shifts in correlated and other manufacturing processes," *International Journal of Production Research*, vol. 40, no. 17, pp. 4449–4466, 2002.

[21] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[22] J. Sun, M. Rahman, Y. S. Wong, and G. S. Hong, "Multiclassification of tool wear with support vector machine by manufacturing loss consideration," *International Journal of Machine Tools and Manufacture*, vol. 44, no. 11, pp. 1179–1187, 2004.

[23] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," in *Data Mining Techniques for the Life Sciences*, O. Carugo and F. Eisenhaber, Eds., vol. 609 of *Methods in Molecular Biology*, pp. 223–239, Humana Press, Totowa, NJ, USA, 2010.

[24] Q. Wu, "Product demand forecasts using wavelet kernel support vector machine and particle swarm optimization in manufacture system," *Journal of Computational and Applied Mathematics*, vol. 233, no. 10, pp. 2481–2491, 2010.

[25] A. Azadeh, M. Saberi, A. Kazem, V. Ebrahimipour, A. Nour-mohammadzadeh, and Z. Saberi, "A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization," *Applied Soft Computing*, vol. 13, no. 3, pp. 1478–1485, 2013.

[26] T. Wuest, C. Irgens, and K.-D. Thoben, "An approach to monitoring quality in manufacturing using supervised machine learning on product state data," *Journal of Intelligent Manufacturing*, vol. 25, no. 5, pp. 1167–1180, 2014.

[27] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.

[28] Y. Chang and C. Lin, "Feature ranking using linear SVM," *JMLR: Workshop and Conference Proceedings*, vol. 3, pp. 53–64, 2008.

[29] M. Mccann, Y. Li, L. Maquire, and A. Johnston, "Causality Challenge: benchmarking relevant signal components for effective monitoring and process control," *Journal of Machine Learning Research: Workshop and Conference Proceedings*, vol. 6, pp. 277–288, 2010.

[30] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, New York, NY, USA, 2013.

[31] J. Sölter, "Modeling and simulation of ring deformation due to clamping," *Materialwissenschaft und Werkstofftechnik*, vol. 40, no. 5-6, pp. 380–384, 2009.

[32] Y. S. Abu-Mostafa and A. F. Atiya, "Introduction to financial forecasting," *Applied Intelligence*, vol. 6, no. 3, pp. 205–213, 1996.

[33] J. W. Hall, "Adaptive selection of US stocks with neural nets," in *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, G. J. Deboeck, Ed., pp. 45–65, Wiley, New York, NY, USA, 1994.

[34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.