

Kombination von Messdaten und wissensbasierter Modellierung zur Fehlerdiagnose bei Weichen

Connecting measurement data and knowledge-based engineering for heavy rail switch fault diagnosis

Susanne Reetz | Thorsten Neumann | Gerrit Schrijver | Arnout van den Berg

Die Anwendung Künstlicher Intelligenz (KI) im Bereich Prognostics and Health Management (PHM) der Eisenbahninfrastruktur, insbesondere in der Fehlerdiagnose, wird durch hinsichtlich Umfang und / oder Labelling unzureichende Datenbestände und die Notwendigkeit der Rückverfolgbarkeit aufgrund strenger Sicherheitsvorschriften erschwert. Vielversprechende Ansätze sind Feature Engineering, unüberwachtes Lernen und wissensbasierte Systeme. Vor diesem Hintergrund wird nachfolgend erörtert, wie Stromumlaufkurven von Weichenantrieben ausgewertet und mit einem für den Menschen interpretierbaren Bayes'schen Netzmodell für Diagnosezwecke verbunden werden können.

1 Einleitung

Mit zunehmender Digitalisierung im Eisenbahnsektor gewinnen Methoden der KI immer mehr an Bedeutung [1]. Datengetriebene Methoden können oft Probleme lösen, die für eine physikalische Modellierung zu komplex sind, und eröffnen damit zahlreiche neue Möglichkeiten. Viele der etablierten datengetriebenen Methoden wie neuronale Netze sind jedoch weder transparent noch für einen Menschen unmittelbar interpretierbar. Im Falle einer großen gelabelten Datengrundlage ist die Performanz dieser Modelle so gut, dass dies oft vertretbar erscheint, anders verhält es sich jedoch bei einem Mangel an gelabelten Daten (guter Qualität) oder einer unvollständigen Datenbank. Zudem machen im Eisenbahnsektor strenge Sicherheitsvorschriften [16] die Rückverfolgbarkeit von Modellen besonders wichtig. Diese Rahmenbedingungen stellen das PHM insbesondere bei der Fehlerdiagnose vor wesentliche Herausforderungen.

Transparente, auf Domänenwissen basierende Methoden sind ein vielversprechender Ansatz zur Überwindung des Problems unzureichender Trainingsdaten, sofern es gelingt, die Modelle in geeigneter Weise mit Messdaten aus der automatisierten Anlagenüberwachung zu kombinieren und ihre Performanz zu validieren.

Vor diesem Hintergrund erörtert die vorliegende Arbeit die Analyse von Stromumlaufkurvenmessungen von Weichenantrieben und deren Kombination mit einem interpretierbaren Bayes'schen Netzmodell zur Fehlerdiagnose. In Abschnitt 2 werden die verwendeten theoretischen Konzepte vorgestellt, gefolgt von ihrer Anwendung auf Stromumlaufkurven und Weichen zu Diagnosezwecken in Abschnitt 3. Der Beitrag schließt mit einer Zusammenfassung des aktuellen Stands der Arbeiten und einem Ausblick auf zukünftige Möglichkeiten.

The application of Artificial Intelligence (AI) in prognostics and health management, especially fault diagnosis, for railway infrastructure is complicated by the largely unlabelled databases and the necessity for traceability due to strict safety regulations. Promising approaches include feature engineering, unsupervised learning and knowledge-based systems. This article discusses how to treat the current curve measurements of railway point machines against this background and connect them with a human-interpretable Bayesian network model for diagnostic purposes.

1 Introduction

Increasing digitalisation in the railway sector means that AI methods are gaining more and more attention [1]. Data-driven methods can often solve problems that are too complex for physical modelling and thus open up numerous new possibilities. However, many established data-driven methods, such as neural networks, are neither transparent nor human-interpretable. In the case of a large labelled database, data-driven models are often so accurate that this seems justifiable. However, the situation is different in the case of a lack of (good quality) labelled data or an incomplete database. Furthermore, the strict regulations for safety-critical applications in the railroad sector [16] make model traceability especially important.

The rail infrastructure sector, and especially the fault diagnostics in this area, exhibits all these challenges at the same time. Therefore, transparent modelling approaches based on domain knowledge (instead of black-box models based on pure machine learning) are expected to overcome the problem of insufficient training data, providing it proves possible to suitably combine and validate the models with measurement data from automatic asset monitoring.

In this regard, this article discusses how to treat current curve measurements from railway point machines and combine them with a human-interpretable Bayesian network model for fault diagnostics. Section 2 introduces the applied theoretical concepts followed by their application to current curves and railway switches for diagnostic purposes in Section 3. The article concludes with a recap of the current state of the work and the outlook for future possibilities.

2 Theoretical concepts

2. Dynamic time warping

Dynamic time warping [2] is a well-known technique used to find the optimal alignment between two given time-dependent sequences under certain restrictions and rules. Any warping be-

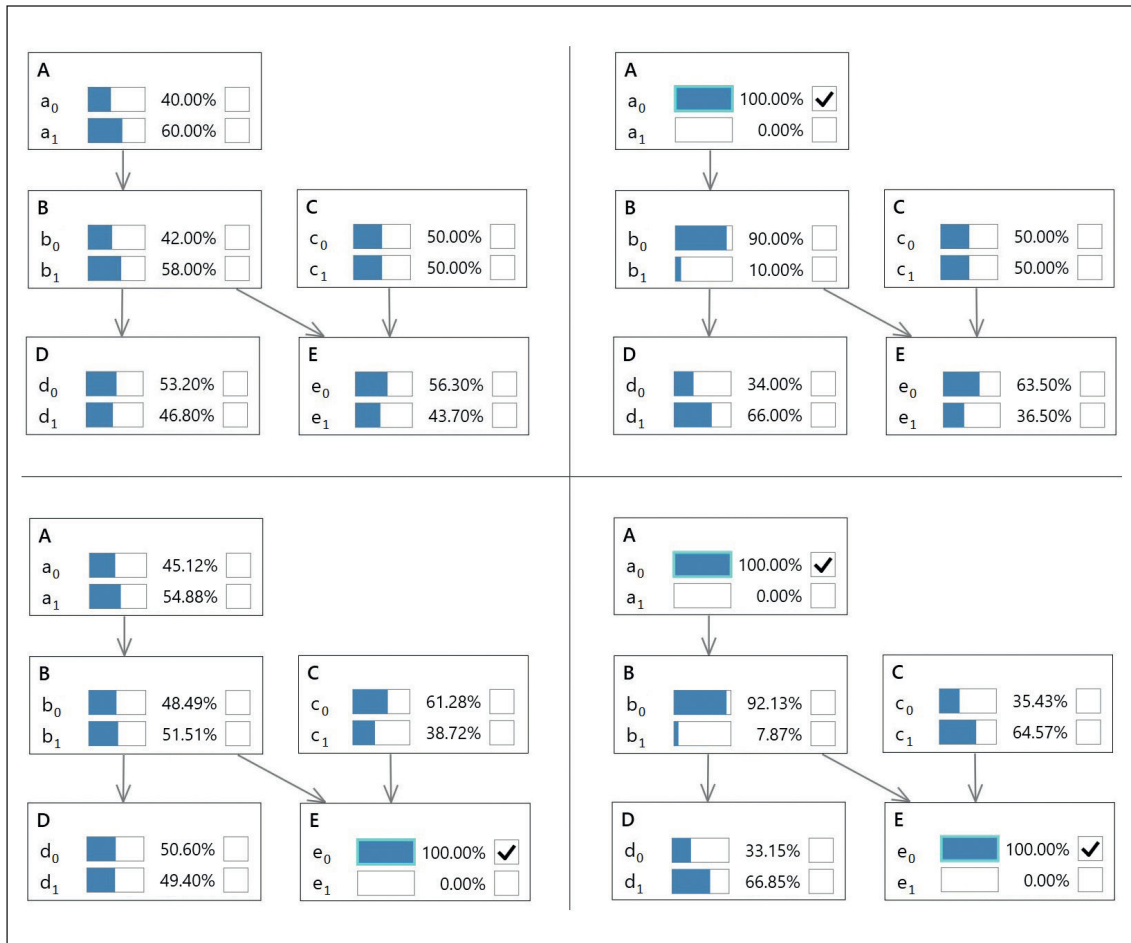


Bild 1: Arten von Inferenz in Bayes'schen Netzen: Keine Evidenzen (oben links), kausale Inferenz (oben rechts), diagnostische Inferenz (unten links) und interkausale Inferenz (unten rechts)

Fig. 1: Reasoning in Bayesian networks: no evidence (top left), causal reasoning (top right), diagnostic reasoning (bottom left) and intercausal reasoning (bottom right)

Quelle (alle Bilder) / Source (all fig.): DLR

2 Theoretische Konzepte

2.1 Dynamic Time Warping

Dynamic Time Warping [2] ist eine etablierte Technik, um eine optimale mathematische Abbildung zwischen zwei gegebenen (zeitabhängigen) Wertefolgen unter bestimmten Einschränkungen und Regeln zu finden. Ein Warping zwischen zwei (nicht notwendig gleich langen) Wertefolgen muss die folgenden Bedingungen erfüllen:

- Randbedingung: Das Warping bildet das jeweils erste und letzte Element der beiden Folgen aufeinander ab.
- Monotonie: Die Zuordnung der Indizes der ersten Folge zu den Indizes der zweiten Folge ist monoton steigend und umgekehrt.
- Schrittweite: Die Schrittweite des Warpings liegt zwischen null und eins, d. h. es werden keine Indizes ausgelassen.

Dynamic Time Warping berechnet den günstigsten Pfad zwischen den beiden betrachteten Folgen anhand eines lokalen Abstandsmaßes für deren Punkte. Ursprünglich wurde es in der Spracherkennung eingesetzt [3] und ist seitdem erfolgreich in verschiedenen Anwendungen für das Mining und Clustering von zeitabhängigen Daten verwendet worden [4].

2.2 Bayes'sche Netze

Bayes'sche Netze [5] ermöglichen es, probabilistische Zusammenhänge zwischen verschiedenen diskreten und / oder kontinuierlichen Zufallsvariablen transparent und interpretierbar zu modellieren. Die Zufallsvariablen (= Knoten mit Zuständen) und deren Verbindungen (= Kanten) werden durch einen gerichteten, azyklischen Graphen abgebildet. Eine Kante steht typischerweise für einen kausalen Zusammenhang zwischen dem Knoten und sei-

tween two time series (of not necessarily the same length) has to fulfil the following conditions:

- the boundary condition: the warping starts with a match between the first element of both series and ends with a match between the last elements of each series.
- the monotony condition: the mapping of the indices of the first sequence to the indices of the other sequence increases monotonically and vice versa.
- the step size condition: the step size of the matching lies between zero and one, i.e. no indices are left out.

In summary, this means that the head and tail need to be matched. There is no cross-matching and no part of any sequence is left out. Dynamic time warping calculates the optimal warping path given a local distance measure for the points of the time series. It was originally applied to speech recognition [3] and has since been used successfully in various applications for time series data mining and clustering [4].

2.2 Bayesian networks

Bayesian networks [5] allow the transparent, humanly interpretable modelling of complex probabilistic relations between different random variables with continuous or discrete states. The random variables (= nodes with states) and their relations (= edges) are arranged in a directed, acyclic graph. An edge represents a typically causal relationship between a node and its successor. The parameterisation of a Bayesian network comprises the base probabilities of the states of the root nodes and the (conditional) probabilities of the states of all the other nodes given the individual constellations of the states of their parent nodes.

nem Nachfolger. Die Parametrisierung eines Bayes'schen Netzes umfasst die Basiswahrscheinlichkeiten für die Zustände der sogenannten Wurzelknoten und die (bedingten) Wahrscheinlichkeiten für die Zustände aller anderen Knoten gegeben die jeweilige Konstellation der Zustände ihrer individuellen Elternknoten.

Die Entwicklung eines Bayes'schen Netzes umfasst zunächst die Konstruktion dessen Struktur (Knoten, Zustände, Kanten) und dann die Parametrisierung. In der Anwendung kann die Information über den aktuellen Zustand einer oder mehrerer der modellierten Zufallsvariablen, sofern bekannt, als sogenannte Evidenz in das Netz eingegeben werden. Die Zustandswahrscheinlichkeiten im übrigen Netz werden daraufhin entsprechend aktualisiert [6]. Die elementaren Muster logischen Schlussfolgerns in Bayes'schen Netzen sind in Bild 1 für ein Beispielnetz dargestellt. Im Vergleich zum Ausgangszustand ohne Evidenz (oben links) gibt die Annahme, dass der Knoten A sich im Zustand a_0 befindet, Aufschluss über die Zustände der nachfolgenden Knoten B, D und E durch kausale Inferenz (oben rechts). Diagnostische (bzw. antikausale) Inferenz hingegen propagiert Informationen im Netzwerk entgegen den Pfeilrichtungen, z. B. ändert die Kenntnis des Zustands von Knoten E die Zustandswahrscheinlichkeiten seiner Vorgänger C, B, und A (unten links) und damit auch wiederum die des Knotens D. Durch Kombination beider Inferenzmuster kann je nach Netzstruktur ferner interkausale Inferenz entstehen, die insbesondere zum Effekt des sogenannten „Explaining Away“ führen kann: Evidenz für a_0 in A führt z. B. zu einer erhöhten Wahrscheinlichkeit für b_0 in B und zu keiner Veränderung bei C, während der Nachweis von e_0 für E sowohl die Wahrscheinlichkeiten für b_0 als auch für c_0 erhöht. Sind beide Evidenzen (d. h. a_0 und e_0) gesetzt, hat b_0 sogar eine noch höhere Wahrscheinlichkeit, während die Wahrscheinlichkeit für c_0 sinkt. Im probabilistischen Sinne eliminiert das Wissen um a_0 also (zu einem gewissen Anteil) c_0 als mögliche andere Ursache für e_0 .

3 Anwendung in der Weichendiagnose

3.1 Weichenüberwachung mit Stromumlaufkurven

Weichen und ihre (typischerweise elektrischen) Antriebe sind kritische Elemente der Eisenbahninfrastruktur. Die Motoren werden daher häufig durch indirekte Messung des elektrischen Stroms während des Umstellvorgangs überwacht, indem die Induktion um das zum Weichenantrieb führende Stromkabel über eine Spule gemessen wird. Viele Aspekte des Weichenverhaltens spiegeln sich in diesen sogenannten Stromumlaufkurven wider, sodass diese Daten eine gute Grundlage für die Bewertung des aktuellen Zustands (hauptsächlich, aber nicht ausschließlich, mechanischer Art) der Weiche darstellen. Im Bereich der Anomaliedetektion, d. h. der automatisierten Erkennung von atypischem Verhalten von Zielobjekten durch regelmäßige Auswertung von Messdaten, wurde die Überwachung von Weichen auf der Basis von Stromumlaufkurven z. B. bereits durch [7] erfolgreich umgesetzt. Sobald eine Warnung ausgegeben wird, besteht der nächste Schritt in der (automatischen) Identifizierung der Ursache(n), die das anormale Verhalten auslösen (d. h. Fehlerdiagnose). Da umfangreich und korrekt gelabelte Fehlerdaten in der Praxis oft nicht verfügbar sind, basiert der im Folgenden vorgeschlagene Ansatz auf ungelabelten Daten und Expertenwissen, wobei Dynamic Time Warping, manuelles Feature Engineering und Bayes'sche Netze zum Einsatz kommen. Besonderes Augenmerk liegt dabei auf der Transparenz und Interpretierbarkeit der Algorithmen.

[8] listet eine Reihe von Ansätzen zur Fehlerdiagnose auf Basis von Stromumlaufkurven auf. [9] detektiert beispielsweise verschiedene Fehlertypen mittels Dynamic Time Warping, während die Publikation [10] einen stärkeren Fokus auf unüberwachtes Lernen

The development of a Bayesian network first entails the construction of a model structure (nodes, states and edges) and its subsequent parameterisation. If the current state of a given random variable is known (exactly) in the application, this information can then be entered into the network as so-called evidence to update the state probabilities in the remaining network accordingly [6].

In order to demonstrate the elementary reasoning patterns in Bayesian networks, consider the network in fig. 1, with its initial state without any evidence displayed in the top left corner. Evidence that node A is in state a_0 provides insight into the states of successive nodes B, D and E through causal reasoning (top right). Diagnostic reasoning in turn moves the information upstream in the network, e.g. knowing the state of node E changes the state probabilities of its predecessors C, B, and A (bottom left) and thus of node D as well. The combination of both (i.e. intercausal reasoning) can lead to so-called “explaining away”. Evidence of a_0 for node A, for instance, results in an increased probability of b_0 for B and no changes in C, whereas evidence of e_0 for E increases both the probabilities for b_0 and c_0 . Given both pieces of evidence, b_0 has an even higher probability, while the probability for c_0 drops. In this case, knowing a_0 eliminates c_0 as a reason for e_0 , at least to some extent.

3 Application in railway switch diagnosis

3.1 Switch monitoring using current curves

Switches and their point machines, commonly powered by electric motors, are critical elements of the railway infrastructure. Therefore, the motors are often monitored by measuring the electric current indirectly during the transition process, whereby the induction around the cable is measured via a coil. Many aspects of the switch behaviour are reflected in these current curves, so that this data constitutes a good basis for assessing the existing (mainly, but not exclusively, mechanical) condition of the switches.

In the area of anomaly detection, i.e. the automated detection of any atypical behaviour in target assets via the regular assessment of measurement data, the monitoring of railway switches based on current curves has already been successfully implemented by [7], for instance. The next step, once an alert is issued, is to (automatically) identify the root cause(s) behind the abnormal behaviour (i.e., diagnosis). As labelled fault data is often hard to come by in practice, the following approach is based on unlabelled data and expert knowledge using dynamic time warping, manual feature engineering and Bayesian networks. Special focus has been given to the transparency and interpretability of the algorithms.

[8] lists a number of approaches to fault diagnosis based on current curves. For example, [9] detects different fault types based on dynamic time warping, while [10] has a stronger focus on unsupervised learning. The idea of using Bayesian networks has previously been promoted for fault diagnosis in general [11]. Moreover, their applicability to railway switches, in particular, has been discussed from an academic perspective in [12]. Since then, several efforts have been made to build up a valid Bayesian network model for a common switch based on expert assessment, further literature review and practical experiments in the field (cf. fig. 2a/2b). The modelling process is still ongoing, but detailed discussions about which features in the current curves are likely to be related to which functional aspects or components of the switch have helped to raise the for-

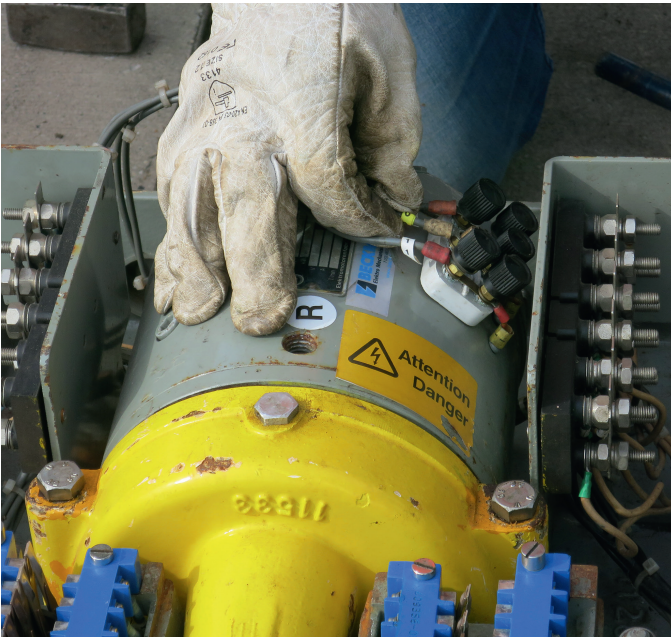


Bild 2: a) Simulation eines Wackelkontaktes; b) fehlerhaft eingestellte Zungenroller
 Fig. 2: a) The simulation of a loose contact; b) misadjusted rollers

legt. Die Idee, Bayes'sche Netze zu verwenden, wurde bereits früher für die Fehlerdiagnose im Allgemeinen propagiert [11]. Darüber hinaus wurde in [12] aus akademischer Sicht ihre Anwendbarkeit für Eisenbahnweichen diskutiert.

Seitdem wurden mehrere Anstrengungen unternommen, ein valides Bayes'sches Netzwerkmodell für einfache Weichen auf der Grundlage von Experteneinschätzungen, weiterer Literaturrecherche und praktischen Experimenten im Feld (vgl. Bild 2a/2b) zu erstellen. Der Modellierungsprozess ist zwar noch nicht abgeschlossen, aber die detaillierte Diskussion darüber, welche Merkmale in den Stromumlaufkurven potenziell mit welchen funktionalen Aspekten oder Komponenten der Weiche zusammenhängen, hat dazu beigetragen, die früheren (rein akademischen) Ideen von [12] auf ein realistischeres Niveau zu heben.

Bild 3 zeigt hierzu einige typische Stromumlaufkurven, die auf mögliche Fehler in der Weiche hinweisen. Die Kurve oben links etwa könnte auf eine schwergängige (aber noch funktionierende) Verriegelung

mer (purely academic) ideas espoused by [12] to a much more realistic level.

Here, fig. 3 shows some typical current curves indicating possible faults in the switch system. The curves in the top left diagram, for instance, could be indicative of a heavy (but still working) locking while the camel-shaped curve in the second plot has likely resulted from misadjusted rollers. Furthermore, the left and middle bottom plots, where the current has not dropped to zero, are likely to have been caused by mechanical blockages during movement that have prevented the blades from reaching their end positions.

The examples show that it is very important to consider and define segment-specific features (i.e., features related to the phases of the switch movement). As such, valid tools are needed for the automatic segmentation of not only regular, but also abnormal current curves. Current curve segmentation based on dynamic time warping (see Section 3.2) is expected to considerably improve the validity of the final Bayesian network model within this context.

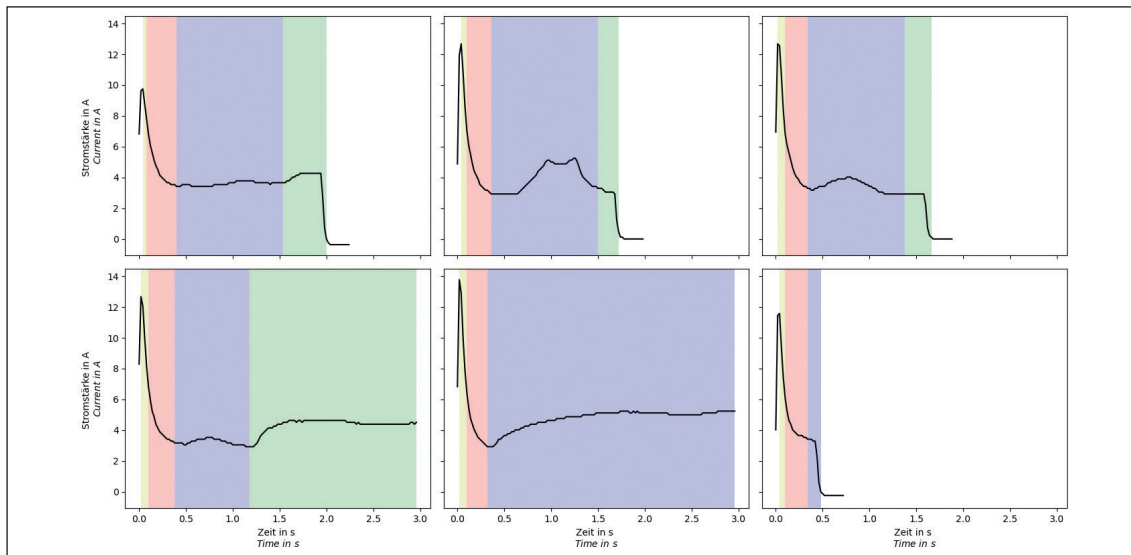
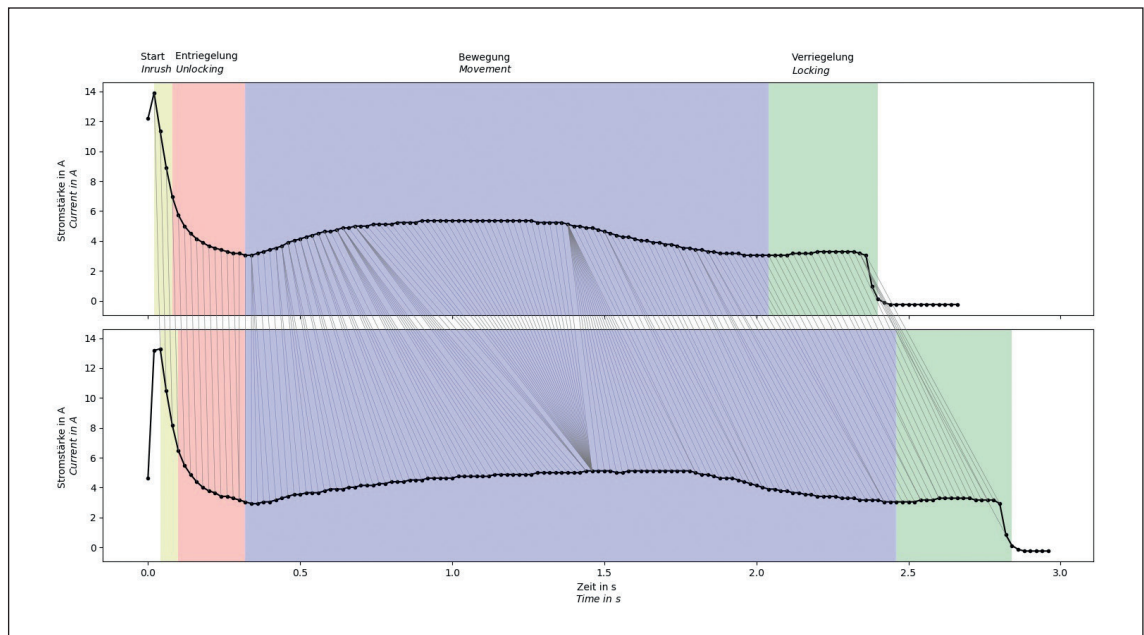


Bild 3: Beispiele abnormaler Stromumlaufkurven und deren Segmentierung in Start (gelb), Entriegelung (rot), Bewegung (blau) und Verriegelung (grün)
 Fig. 3: Examples of abnormal current curves with their segmentation into the inrush (yellow), unlocking (red), movement (blue) and locking phases (green)

Bild 4: Dynamische Segmentierung einer neuen Stromumlaufkurve (unten) auf Basis einer Musterkurve (oben)

Fig. 4: Dynamic segmentation of a new current curve (bottom) based on a template curve (top)



hindeuten, während die kamelhöckerförmige Kurve im zweiten Diagramm wahrscheinlich von falsch eingestellten Zungenrollern herrührt. Die Diagramme links unten in der Mitte, bei denen die Stromstärke am Ende nicht auf null abfällt, sind wahrscheinlich auf mechanische Blockaden während der Bewegung zurückzuführen, welche die Zungen daran hindern, ihre Endposition zu erreichen.

Die Beispiele zeigen unter anderem, wie wichtig es ist, Merkmale der Stromumlaufkurven segmentspezifisch (d.h. differenziert nach den Phasen des Weichenumschaltens) zu berücksichtigen und zu definieren. Somit werden zwingend valide Werkzeuge für die automatische Segmentierung von nicht nur regulären, sondern auch anomalen Stromumlaufkurven benötigt. Die Segmentierung von Stromumlaufkurven mittels Dynamic Time Warping (siehe Abschnitt 3.2) hat in diesem Zusammenhang das Potenzial, die Validität des finalen Diagnosemodells erheblich zu verbessern.

3.2 Segmentierung von Stromumlaufkurven

Unabhängig von der genauen Bauart können die Stromumlaufkurven von Weichenantrieben mit Elektromotoren in verschiedene Abschnitte unterteilt werden. So können Kurven wie in Bild 4 in Start-, Entriegelungs-, Bewegungs- und Verriegelungsphase unterteilt werden. Die automatisierte Segmentierung der Kurven ist vor allem für die Fehlerdiagnose relevant, da sich so aus den Daten genauere und vor allem interpretierbare Informationen ableiten lassen.

Gerade im Fehlerfall nehmen Stromumlaufkurven jedoch sehr unterschiedliche Formen an, sodass ein einfacher regelbasierter Ansatz zur automatischen Segmentierung nicht sinnvoll ist. Gängige überwachte Lernansätze hingegen benötigen eine große Menge an gelabelten Trainingsdaten, d.h. manuell segmentierte Stromumlaufkurven, die mit großem Aufwand händisch erstellt werden müssten. Daher wird im Folgenden ein alternatives Verfahren für die Segmentierung von Stromumlaufkurven vorgeschlagen, welches auf Dynamic Time Warping basiert. Der Algorithmus erfordert keinen Trainingsschritt und ist flexibel genug, um gewisse Unterschiede zwischen den zeitabhängigen Sequenzen kompensieren zu können.

Angenommen, es gibt eine kleine Datenbank mit händisch segmentierten Stromumlaufkurven (einige wenige oder sogar nur eine pro typischer Fehlerform sollten für den Anfang ausreichen), die im Folgenden als Musterkurven bezeichnet werden. Die Seg-

3.2 Current curve segmentation

The current curves of point machines with electric motors can be divided into different segments regardless of the exact construction type. As such, the curves can be divided into the inrush, unlocking, movement and locking phases as in fig. 4. Automated curve segmentation is particularly relevant for fault diagnosis, as it allows more precise and, above all, interpretable information to be derived from the data.

However, current curves can take on very different forms, especially in the case of errors, meaning that a simple rules-based approach to automatic segmentation is not expedient. Common supervised learning approaches on the other hand require a large amount of labelled training data, i.e. manually segmented current curves. Therefore, an alternative pipeline for current curve segmentation based on dynamic time warping is proposed below. The algorithm does not require a training step and is flexible enough to cope with any differences between the time-dependent sequences. Let's assume there is a small database of labelled current curves (just a few or even one for every general fault shape should be enough at the beginning) called template curves. The proposed segmenting pipeline for a new current curve would then require the following steps:

1. up-sampling and interpolating the new current curve to prevent any artefacts based on measurement errors.
2. detecting the inrush peak and, if it exists, the end point of the new curve where the current drops to zero and remains there.
3. finding the template current curve that is (most) similar to the new current curve.
4. calculating the transition between both curves using dynamic time warping.
5. transferring the segmentation from the template onto the new current curve.

Fig. 4 exemplarily illustrates the transition between two curves using dynamic time warping and the transfer of the segmentation labels.

3.3 The diagnostic model

The basic idea of the proposed diagnostic model for railway switches is to represent the point machine, the switch's mechani-

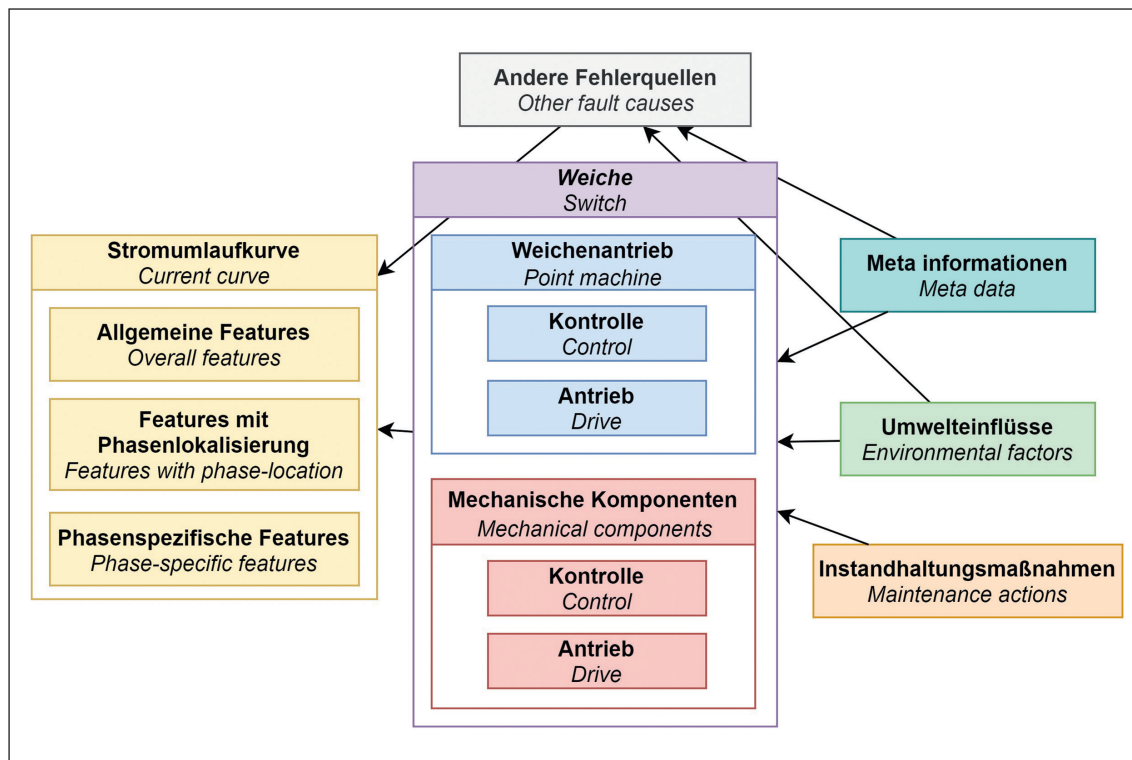


Bild 5: Vereinfachte Darstellung des Diagnosemodells

Fig. 5: A greatly simplified overview of the proposed diagnostic model

mentierung einer neuen Stromlaufkurve erfolgt sodann mittels folgender Schritte:

1. Up-Sampling und Interpolation der neuen Stromlaufkurve, um Artefakte aufgrund von Messfehlern zu vermeiden.
2. Lokalisierung des Beginns der Startphase und, falls vorhanden, des Endpunkts der neuen Kurve, an dem die gemessene Stromstärke auf null abfällt und dort verbleibt.
3. Auswahl einer Musterkurve, die der neuen Stromlaufkurve (möglichst) ähnlich ist.
4. Berechnung der Abbildungsfunktion zwischen beiden Kurven durch Dynamic Time Warping.
5. Übertragung der Segmentierung der Musterkurve auf die neue Kurve. Bild 4 veranschaulicht exemplarisch den Übergang zwischen zwei Kurven durch Dynamic Time Warping und die Übertragung der Segmentierungsetiketten.

3.3 Diagnosemodell

Die Grundidee des vorgeschlagenen Diagnosemodells für Eisenbahnweichen besteht darin, den Weichenantrieb, die mechanischen Komponenten der Weiche, externe Faktoren und Messdaten als miteinander verbundene probabilistische Knoten darzustellen. Die zentralen Teile sind der Weichenantrieb selbst und die mechanischen Komponenten, d. h. das violett dargestellte Weichenmodul in Bild 5, wobei die aktuelle Funktionsfähigkeit der Weichenkomponenten (d. h. normales Verhalten, Fehlerart x, Fehlerart y, ...) als Zustände der Knoten modelliert werden. Im Folgenden werden der Konstruktionsprozess sowie die in der aktuellen Modellstruktur (Bild 5) getroffenen Design-Entscheidungen näher erläutert. Die Entwicklung eines Bayes'schen Netzes für ein komplexes Zielobjekt ist stets mit Kompromissen verbunden. Insbesondere muss ein Gleichgewicht zwischen Detailgrad und Genauigkeit auf der einen versus Vollständigkeit und Klarheit auf der anderen Seite gefunden werden. D. h., die Komponenten und ihre Fehlertypen könnten sehr detailliert modelliert werden, was aus ingenieurwissenschaftlicher Sicht vorteilhaft ist und gut mit anderen gängigen Ansätzen wie der

cal components, any external factors and the measurement data as interconnected probabilistic nodes. The central parts are the point machine itself and the mechanical components, i. e. the purple switch module in fig. 5, where the health condition or current functionality of the switch components (i. e., normal behaviour, fault type x, fault type y, ...) are modelled as states of the nodes. The following paragraph provides a closer look at the construction process and the choices made in the current model structure (fig. 5).

The development of a Bayesian network model for a complex asset is subject to trade-offs. First, it is necessary to balance detail and accuracy versus completeness and clarity. That is, the components and their fault types can be modelled in great detail, which is favourable from an engineering point of view and aligns well with other common traditional tools such as FMECA [13]. On the other hand, the greater the level of detail the more likely it is that a certain fault type or subcomponent will be forgotten in the modelling, resulting in an incomplete model. Furthermore, the connections between components are harder to model correctly and the increasing size makes the model less manageable and the parametrisation more complex.

Secondly, modelling can be performed from either a physical or a functional point of view. This particularly concerns the state of the nodes, i. e. the definition of fault types in the model. A physical view promotes fault types closely related to the actually occurring component faults, while functional fault types describe the effects that a fault can have on the working of an asset or component and the measurement data.

With regard to the model from fig. 5, a decision was made to model all the central components (e. g., the motor or the clutch) as a subset of the elements listed in the detailed FMECA tables (as provided by Strukton) that have been analysed during this process, but not the smaller subcomponents such as the motor's carbon brushes. The state of

FMECA [13] korrespondiert. Andererseits gilt: Je höher der Detailgrad, desto wahrscheinlicher ist es, dass ein bestimmter Fehlertyp oder eine Unterkomponente bei der Modellierung vergessen wird, was zu einem unvollständigen Modell führt. Außerdem sind die Verbindungen zwischen den Komponenten schwieriger korrekt zu modellieren, und mit zunehmender Größe wird das Modell schwerer zu handhaben und die Parametrisierung komplexer.

Darüber hinaus kann die Modellierung von einem physikalischen oder einem funktionalen Standpunkt aus erfolgen. Dies betrifft insbesondere die Definition der möglichen Zustände der Netzknoten, d. h. die konkret abgebildeten Fehlertypen im Modell. Eine physikalische Sichtweise favorisiert dabei Fehlertypen, die eng mit den tatsächlich auftretenden Komponentenfehlern zusammenhängen, während funktionale Fehlertypen die Auswirkungen beschreiben, die entsprechende Fehler auf die Funktionsweise der Anlage oder Komponente und die Messdaten haben können.

Das in Bild 5 skizzierte Modell beruht unter anderem auf detaillierten (von Strukton zur Verfügung gestellten) FMECA-Tabellen. Es wurde allerdings entschieden, zwar alle zentralen Komponenten (z. B. den Motor oder die Kupplung), nicht jedoch kleinere Unterkomponenten wie beispielsweise die Kohlebürsten des Elektromotors explizit zu modellieren. Die Zustände der Komponenten werden größtenteils funktional verstanden, wenngleich versucht wurde, nahe an den tatsächlichen physikalischen Fehlertypen (d. h. den komponentenspezifischen Fehlern) zu bleiben. Darüber hinaus ist zu beachten, dass einige primäre Fehler im Weichensystem, wie z. B. die Einstellung des Weichengestänges, nicht einer einzelnen Komponente zugeordnet werden können, da sie vielmehr aus dem Zusammenspiel verschiedener Komponenten resultieren und auch durch weitere Faktoren wie Gleisgeometrie oder Zustand des Schotterbetts beeinflusst sein können. Das finale Bayes'sche Netz wird also eine Mischung aus einer funktionalen Perspektive und der komponentenzentrierten Sichtweise sein. Insgesamt erleichtert dies die Verknüpfung des Diagnosemodells mit relevanten Merkmalen der Stromlaufkurven, ohne die spätere Modellierung des Einflusses externer Faktoren auf die Weichenkomponenten (d. h. deren Fehlerwahrscheinlichkeiten) unnötig zu verkomplizieren.

Ziel des Entwurfs ist eine robuste Modellstruktur in dem Sinne, dass kleine Änderungen in der Weichenkonstruktion möglichst keine Anpassungen der Modellstruktur erfordern, sondern durch eine Modifikation der Parametrisierung abgefangen werden können. Geringfügige Änderungen des Weichenaufbaus (z. B. verschiedene Arten von Zungenrollern) dürften damit ohne einen größeren Modellumbau abbildbar sein.

Ausgehend von einem vollständigen Modell des Weichenantriebs und der mechanischen Komponenten, werden die aus einer gegebenen Stromlaufkurve extrahierten Informationen über das aktuelle Verhalten der Anlage in Form sogenannter Feature-Knoten in das Netz eingefügt (vgl. gelbes Modul in Bild 5). Hierbei ermöglicht die dynamische Segmentierung der Stromlaufkurve die Berechnung möglichst präziser Features, die eine (wahrscheinlichkeitsbasierte) Unterscheidung zwischen verschiedenen Fehlerfällen wie den Beispielen aus Bild 3 ermöglichen. Dabei können grob drei Kategorien von Features differenziert werden:

- Allgemeine Features, die sich auf die Stromkurve als Ganzes beziehen (z. B. Länge oder Integral der Kurve)
- Features mit Phasenlokalisierung, d. h. solche, die höchstens einmal in einer Stromkurve auftreten können, wobei es für die Interpretation entscheidend ist, ob und in welcher Phase das Ereignis auftritt (z. B. vorzeitiger Abbruch der Kurve)
- Phasenspezifische Features, die nur für bestimmte Phasen definiert sind (z. B. Höhe des Peaks in der Startphase).

each component is mostly understood in a functional way, though the idea was to stay as close as possible to the actual physical fault types (i.e. the component-specific defects). Moreover, note the fact that some primary faults in the switch system, such as misadjusted rods for instance, cannot even be assigned to a single component, as they result from the interaction of many different components and could also be affected by further factors like track geometry or ballast conditions. Thus, the final Bayesian network will most certainly be a mixture of a functional perspective (which is closer to the observations from the current curves) and a component-centred view of the given switch. All in all, this facilitates the connection of the diagnostic model with the relevant current curve features derived from the measurement data in the following, while avoiding unnecessary complications when modelling the influence of any external factors on the switch components (i.e. their fault probabilities) later on.

The goal of the design is to ensure that, while the parameterisation might need to be adjusted for each type or generation of point machines, the model structure mostly remains the same and is thus "robust" with regard to any small changes in the point machine construction. Minor changes in the switch design (e.g., different types of blade rollers) are expected to be implemented without the need for a major model rebuild.

Given a complete model of the point machine and the mechanical components, the information on the health status of the asset as extracted from a current curve is added to the network in the form of so-called feature nodes (cf. the yellow module in fig. 5). Within this context, the dynamic segmentation of the current curve allows the calculation of more precise features that help to differentiate between various fault cases such as the examples in fig. 3. There are three categories of features:

- general features that refer to the current curve as a whole.
- singular event features that can occur a maximum of one time in a current curve, whereby if and which phase the event occurs in is critical for the interpretation.
- phase-specific features defined for certain phases only.

In general, an increase (decrease) in mechanical resistance is met by an increase (decrease) in the motor current limited by the clutch. When taking all the possible faults (including multiple simultaneous faults) into account, the connection of features and fault types is non-trivial and still a work in progress. In practice, calculated current curve features can be entered into the model as evidence and provide insights on possible switch fault types through diagnostic reasoning. Furthermore, the assessment of the fault states is further influenced by external factors, including meta data such as the age or load (turquoise), environmental information such as precipitation (green) and maintenance information (brown), all shown on the right-hand side of fig. 5. As evidence, this information changes the switch fault probabilities in the network through causal reasoning.

4 Conclusion and outlook

4.1 The state of development

The presented approach to segmenting current curves based on dynamic time warping (cf. Section 3.2) has been verified based on several examples. The open challenge is now to develop a sta-

Im Allgemeinen bedingt eine Zunahme (Abnahme) des mechanischen Widerstands eine Zunahme (Abnahme) des Motorstroms, die durch die Kupplung mechanisch nach oben begrenzt wird. Dennoch ist die Zuordnung von Features und Fehlertypen unter Berücksichtigung aller möglichen (ggf. auch gleichzeitig auftretenden) Fehler nicht trivial und aktuell nicht abgeschlossen.

In der praktischen Anwendung können die berechneten Features der Stromumlaufkurven als Evidenz in das Modell eingegeben werden und durch vorrangig diagnostische Inferenz Abschluss über die daraus resultierende Wahrscheinlichkeit möglicher Weichen- bzw. Komponentenfehler geben. Dabei ist zu bedenken, dass die Bewertung der Fehlerzustände durch verschiedene externe Faktoren (vgl. rechte Seite in Bild 5) beeinflusst sein kann, u. a. das Bauteilalter oder die Weichenbeanspruchung (türkis), Umweltbedingungen wie Temperatur oder Niederschlag (grün) und / oder Qualität und Häufigkeit von Wartungseinsätzen (braun). Diese Informationen verändern die Fehlerwahrscheinlichkeiten im Bayes'schen Netz durch kausale Inferenz.

4 Fazit und Ausblick

4.1 Aktueller Stand der Entwicklung

Der vorgestellte Ansatz zur Segmentierung von Stromumlaufkurven auf der Basis von Dynamic Time Warping (vgl. Abschnitt 3.2) wurde bisher anhand mehrerer Beispiele überprüft. Die Herausforderung besteht nun darin, eine stabile Implementierung zu entwickeln, die konstant präzise Ergebnisse einschließlich einer sorgfältigen Qualitätsvalidierung liefert.

Bayes'sche Netze sind ferner ein etablierter Ansatz für die diagnostische Modellierung in verschiedenen Wissenschafts- und Anwendungsdisziplinen, unter anderem in der Medizin [14] und der Bioinformatik [15]. Zum Zeitpunkt der Erstellung des vorliegenden Beitrags ist die Entwicklung der Struktur des in Abschnitt 3.3 beschriebenen Diagnosemodells für Weichenantriebe in seiner Grundversion weitgehend abgeschlossen. In einem nächsten Schritt wird eine Parametrisierung auf der Grundlage von Literaturwerten, Experteneinschätzungen und verfügbaren Wartungsinformationen durchgeführt. Anschließend erfolgt die qualitative Verifizierung des Modells durch Expertenbewertungen unter Verwendung gängiger Fehler Szenarien aus der Praxis.

4.2 Fazit

Die vorgeschlagene Methode zur dynamischen Segmentierung der Stromumlaufkurven ermöglicht die Gewinnung interpretierbarer Features, ohne dass eine größere Menge an gelabelten Daten benötigt wird. Im günstigsten Fall können mehrere Anomalien, die gleichzeitig in einer einzelnen Kurve auftreten, ohne weiteren Aufwand für das Feature Engineering analysiert werden. Auf der Grundlage von Expertenwissen können diese Features sodann zusammen mit anderen Informationen auf transparente, interpretierbare Weise in das Diagnosemodell eingegeben werden. Die Aussagekraft des Diagnosemodells steigt dabei mit der Menge der gesetzten Evidenzen. Es müssen jedoch nicht alle Informationen sofort verfügbar sein, damit das Modell funktioniert. Insgesamt kann der Ansatz also die Entscheidungsfindung bei der Fehlerursachenanalyse von abnormalen Weichenumläufen unterstützen.

Es ist jedoch zu beachten, dass Systeme, die auf ungelabelten Daten und manueller Modellierung von Expertenwissen basieren, anfällig für strukturelle Fehler sind und aufgrund menschlicher Unzulänglichkeiten bei der Modellierung anfangs oft nur eine eingeschränkte Genauigkeit aufweisen. Insbesondere die Parametrisierung ist eine schwierige Aufgabe, wenn keine umfassenden Fehler- und

ble implementation which provides highly accurate results, including profound validation of the quality.

Bayesian networks are a well-established approach for diagnostic modelling in various disciplines, including medicine [14] and bioinformatics [15]. At the moment of writing, the development of the structure of the diagnostic model for heavy rail switches as described in Section 3.3 has almost been completed in its basic version and parametrisation based on any available literature values, expert assessments and data will be carried out in the next step. A qualitative verification will then be achieved through expert evaluations within the context of common fault scenarios taken from reality.

4.2 Conclusions

The proposed method for the dynamic segmentation of current curves allows for interpretable feature engineering without requiring a larger amount of labelled data. Multiple anomalies occurring in a single curve can be analysed without any further effort on the part of feature engineering. These features could then be entered into the diagnostic model alongside other information in a transparent, interpretable way based on expert knowledge. The power of the diagnostic model increases with the amount of entered evidence, but not all the information has to be immediately available for the model to be functional. Altogether, this approach can support the decision-making for the fail cause analysis of any abnormal point machine measurements.

However, systems based on unlabelled data and the manual modelling of expert knowledge are often prone to structural errors and lower accuracy due to human errors (of judgment). Parametrisation is an especially challenging task, if no comprehensive statistics are available. Thus, such models can usually support experts in their work and standardise the processes, but must not completely assume the responsibility for any decision-making without further extensive validation and testing.

4.3 Outlook

To date, the diagnostic model discussed above has mainly been built on a conceptual design and with qualitative considerations. However, performance is a key factor from the perspective of the practitioners. This can only be evaluated by implementing the entire diagnostic pipeline in a real-world environment in combination with a feedback loop on any true faults as identified during the execution of the relevant maintenance actions.

With regard to the scope of the model, its structure does not yet incorporate any information on the switch positioning at the relay house or on the influences of any fluctuations in the power supply and other factors. Needless to say, the more information sources that are utilised, the stronger the diagnostic model is. As such, additional factors such as environmental conditions or rail operations data, as well as measurement data other than current curves (e.g., track geometry, ballast conditions,...), should prospectively be integrated step by step. Platforms used to collect and manage this kind of data, including standardised maintenance feedback, will become more and more important in the future within this context.

Research conducted within the Shift2Rail project IN2SMART2; EU Horizon 2020 research and innovation program, grant agreement 881574. ■

Wartungsstatistiken verfügbar sind. Daher können solche Modelle Experten und Wartungsteams bei ihrer Arbeit zwar mit wertvollen Hinweisen unterstützen und Prozesse standardisieren, dürfen aber ohne weitere umfangreiche Validierung und Tests nicht vollständig die Verantwortung für die Entscheidungsfindung übernehmen.

4.3 Ausblick

Bislang basiert das oben beschriebene Diagnosemodell hauptsächlich auf konzeptionellen und qualitativen Überlegungen. Aus Sicht der Praktiker ist jedoch nicht zuletzt auch die Performanz ein Schlüsselfaktor. Diese kann nur durch die Implementierung der gesamten Verfahrenskette (inkl. Segmentierung und Inferenz im Bayes'schen Netz) im Realbetrieb abschließend bewertet werden. Dabei ist vor allem auch der systematische Abgleich mit den tatsächlich eingetretenen Fehlern, die im Rahmen der Durchführung der entsprechenden Wartungsmaßnahmen zu ermitteln und sorgfältig zu dokumentieren sind, entscheidend.

Was den Umfang des Modells anbelangt, so enthält dessen Struktur aktuell noch keine Informationen aus dem Stellwerk, zu Einflüssen von Schwankungen in der Stromversorgung und zu etwaigen weiterführenden Faktoren. Perspektivisch sollten folglich nach und nach weitere Aspekte wie Bahnbetriebsdaten und andere Messdaten (z. B. Gleisgeometrie, Zustand des Schotterbetts, ...) in das Diagnosemodell integriert werden. In diesem Zusammenhang dürften Plattformen zur Bündelung und Verwaltung von Zustands- und Wartungsdaten einschließlich standardisierter Rückmeldungen zu durchgeführten Instandhaltungsmaßnahmen in Zukunft immer wichtiger werden.

Die dargestellten Forschungsergebnisse basieren auf Arbeiten, welche als Teil des Shift2Rail-Projekts IN2SMART2 (Grant Agreement 881574) im Rahmen des Forschungs- und Innovationsprogramms Horizon 2020 der EU durchgeführt wurden. ■

AUTOREN | AUTHORS

Susanne Reetz

Research Assistant

Institut für Verkehrssystemtechnik / *Institute of Transportation Systems*

Deutsches Zentrum für Luft- und Raumfahrt

Anschrift/Address: Lilienthalplatz 7, D-38108 Braunschweig

E-Mail: susanne.reetz@dlr.de

Dr.-Ing. Thorsten Neumann

Research Assistant

Institut für Verkehrssystemtechnik / *Institute of Transportation Systems*

Deutsches Zentrum für Luft- und Raumfahrt

Anschrift/Address: Rutherfordstraße 2, D-12489 Berlin

E-Mail: thorsten.neumann@dlr.de

Gerrit Schrijver

Maintenance Analyst

Strukton Rail bv Control Center (SCC)

Anschrift/Address: Westkanaaldijk 2, NL-3542 DA Utrecht

E-Mail: gerrit.schrijver@strukton.com

Arnout van den Berg

Maintenance Analyst

Strukton Rail bv Control Center (SCC)

Anschrift/Address: Westkanaaldijk 2, NL-3542 DA Utrecht

E-Mail: arnout.vandenberg@strukton.com

LITERATUR | LITERATURE

- [1] Xie, J.; Huang, J.; Zeng, C.; Jiang, S.-H.; Podlich, N. (2020): Systematic Literature Review on Data-Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering (M. D. Institute, Hrsg.), *Geosciences*, 10, S. 425
- [2] Müller (2007): Dynamic Time Warping. In M. Müller: *Information Retrieval for Music and Motion* (S. 69-84), Berlin, Heidelberg: Springer Berlin Heidelberg
- [3] Sakoe, H.; Chiba, S. (1978): Dynamic programming algorithm optimization for spoken word recognition. (IEEE, Hrsg.) *IEEE transactions on acoustics, speech, and signal processing*, 26, S. 43-49
- [4] Li, H. (2015): On-line and dynamic time warping for time series data mining. (Springer, Hrsg.) *International Journal of Machine Learning and Cybernetics*, 6, S. 145-153
- [5] Koller, D.; Friedman, N. (2009): *Probabilistic graphical models*. Cambridge, Mass.: MIT Press
- [6] Charniak, E. (1991): *Bayesian Networks without Tears*. *AI Magazine* (Winter 1991), S. 50-63
- [7] Narezo Guzmán, D.; Hadzic, E.; Baasch, B.; Heusel, J.; Neumann, T.; Schrijver, G.; ... Groos, J. C. (2020): Anomaly Detection and Forecasting Methods Applied to Point Machine Monitoring Data for Prevention of Railway Switch Failures. In Ball, A.; Gelman, L.; Rao, B. K.: *Advances in Asset Management and Condition Monitoring* (Bd. 166, S. 307-318). Cham: Springer International Publishing
- [8] Hamadache, M.; Dutta, S.; Olaby, O.; Ambur, R.; Stewart, E.; Dixon, R. (2019): On the Fault Detection and Diagnosis of Railway Switch and Crossing Systems: An Overview. *Applied Sciences*, 9(23), S. 5129
- [9] Huang, S.; Zhang, F.; Yu, R.; Chen, W.; Hu, F.; Dong, D. (2017): Turnout Fault Diagnosis through Dynamic Time Warping and Signal Normalization. *Journal of Advanced Transportation* (2017)
- [10] Guo, Z.; Wan, Y.; Ye, H. (2020): An Unsupervised Fault-Detection Method for Railway Turnouts. *IEEE Transactions on instrumentation and measurement*, 69(11), S. 8881-8901
- [11] Cai, B.; Huang, L.; Xie, M. (2017): Bayesian networks in fault diagnosis. (IEEE, Hrsg.) *IEEE Transactions on industrial informatics*, 13, S. 2227-2240
- [12] Neumann, Th.; Narezo Guzmán, D.; Groos, J. C. (2019): Transparente Fehlerdiagnose bei Weichenstörungen mittels Bayes'scher Netze – Transparent failure diagnostics for railway switches using Bayesian networks, *SIGNAL+DRAHT* (111), 12/2019, S. 23-31
- [13] IEC 60812, (2018): Analysis techniques for system reliability - Procedure for failure mode and effects analysis (FMEA). International Electrotechnical Commission (IEC)
- [14] Sierra, B.; Inza, I.; Larrañaga, P. (2000): Medical bayes networks. *International Symposium on Medical Data Analysis* (S. 4-4), Springer
- [15] Friedman, N.; Linial, M.; Nachman, I.; Pe'er, D. (2000): Using Bayesian networks to analyze expression data. (I. Mary Ann Liebert, Hrsg.) *Journal of computational biology*, 7, S. 601-620
- [16] European Union (2004): *Railway Safety Directive*. Official Journal of the European Union (L164), S. 44-113