Leibniz Universität Hannover
Wirtschaftswissenschaftliche Fakultät
Institut für Wirtschaftsinformatik

Erkennung von Störungen in periodischen Vorgängen durch intelligente Systeme

Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M. Sc.)" im Studiengang Wirtschaftsingenieur der Fakultät für Elektrotechnik und Informatik, Fakultät für Maschinenbau und der Wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover

vorgelegt von

Alexander Wladimirowitsch Axjonow
[redacted]

Prüfer: Prof. Dr. Michael H. Breitner
Betreuer: Jens Passlick

Hannover, den 29. März 2018

Inhaltsverzeichnis

Abbildungsverzeichnis I

Tabellenverzeichnis IV

Formel- und Abkürzungsverzeichnis V

Quellcodeverzeichnis VII

1 Einleitung 1

2 Grundlagen 3

2.1 Periodischer Vorgang 3

2.2 Intelligentes System 6

2.3 Knowledge Discovery in Databases 8

2.3.1 Domänenverständnis und Zieldefinition 11

2.3.2 Datenselektion 12

2.3.3 Datenvorverarbeitung und -bereinigung 12

2.3.4 Datentransformation 13

2.3.5 Data Mining 14

2.3.6 Interpretation und Evaluation 15

2.4 Data Mining 16

2.4.1 Data Mining Ansätze 18

2.4.2 Datenbanken 23

2.4.3 Modellierungskulturen 24

2.4.3.1 Statistik 25

2.4.3.2 Machine Learning 28

3 Literaturübersicht 31

4 Knowledge Discovery in Databases 36

4.1 Domänenverständnis und Zieldefinition 36

4.2 Versuchsaufbau 37

4.3 Analyse und Auswahl des Data Mining Verfahrens 41

4.3.1 Bewertung und Auswahl der Modellierungskulturen 41

4.3.2 Bestimmung des Data Mining Ansatzes 46

4.3.3 Bestimmung der Lernart 47

4.3.4 Bewertung und Auswahl der Methoden der ersten Ebene 48

4.3.4.1 Künstlich neuronale Netze 48

4.3.4.2 Entscheidungsbäume 51

4.3.4.3 Gegenüberstellung der Verfahren der ersten Ebene 55

4.3.5 Bewertung und Auswahl der Verfahren der zweiten Ebene 56

4.3.5.1 Vorwärtsbetrieben neuronale Netze 56

4.3.5.2 Rekurrent neuronale Netze ... 60
4.3.5.3 Gegenüberstellung der Verfahren der zweiten Ebene ... 62
4.3.6 Bewertung und Auswahl der Verfahren der dritten Ebene ... 63
4.3.6.1 Jordan-Netze ... 64
4.3.6.2 Elman-Netze ... 69
4.3.6.3 Long Short-Term Memory ... 72
4.3.6.4 Gegenüberstellung der Verfahren der dritten Ebene ... 76
4.4 Datenselektion ... 78
4.5 Datenvorverarbeitung und -bereinigung ... 79
4.6 Datentransformation ... 86
4.7 Data Mining ... 98
4.8 Interpretation und Evaluation ... 106
5 Diskussion und Handlungsempfehlung ... 124
6 Limitation ... 128
7 Fazit und Ausblick ... 131
8 Literaturverzeichnis ... 135
Anhang ... i
Ehrenwörtliche Erklärung ... v

1 Einleitung

> *"Maintenance costs are a major part of the total operating costs of all manufacturing or production plants. Depending on the specific industry, maintenance costs can be represent between 15 and 60 percent of the cost of goods produced."* [1]

Kosten die im Lebenszyklus eines Produkts anfallen, bestehen zu einem bedeutenden Anteil aus Instandhaltungskosten. Wie das angeführte Zitat verdeutlicht, wird aufgrund der Kostenverteilung diesem Unternehmensbereich eine große Bedeutung zugewiesen. Das produzierende Gewerbe wird durch eine Vielzahl von Anlagen mit hohem Automatisierungsgrad geprägt. In Zeiten immer komplexer werdender Fertigungsprozesse, stellen Faktoren wie die Termintreue und Produktqualität nach wie vor ein unabdingbares Kriterium für den Erhalt der Wettbewerbsfähigkeit eines Unternehmens dar. Für die Einhaltung dieser Kriterien kann die Instandhaltung maßgeblich verantwortlich sein.[2] Im Zuge der Entwicklung neuer Technologien, unterliegen eine Vielzahl an Unternehmensbereichen einem stetigen Wandel. Diese kontinuierliche Weiterentwicklung gilt insbesondere, aufgrund des Stellhebels in der Wertschöpfungskette, für den Bereich der Instandhaltung. Schon heute lautet der Tenor vieler Technologie-CEOs *AI-First*. Ein Grund für den derzeitigen Trend des Einsatzes besonders rechenintensiver lernender Algorithmen und Verfahren, liegt in der Verfügbarkeit großer Datenmengen und günstigen Rechenleistung. Lernende Maschinen stellen jedoch keine Neuheit dar, da die Grundlagen hierfür schon in den 60er Jahren geschaffen wurden. Dennoch wurde erst durch die Verfügbarkeit preiswerter Rechenleistung eine vielfältige praktische und besonders wirtschaftliche Anwendung dieser Technologien möglich.[3]

Gründe für notwenige Instandhaltungen sind unterschiedlichsten Ursprungs. In diesem Kontext stellen Störungen aufgrund ihrer Auswirkungen einen maßgeblichen Kostenfaktor dar. Zudem sind sie durch ihre große Vielfältigkeit schwer zu detektieren, da sie verschiedenste unbekannte Merkmale in den periodischen Sensorverläufen besitzen. Jedoch haben alle Ursachen die Gemeinsamkeit, dass sie früher oder später bei Nichtbehebung einen Stillstand der Anlage bedingen. Um diesem Vorgang entgegenzuwirken, kann eine intelligente Überwachung ausfallkritischer Komponenten, zu einer Effizienzsteigerung führen und sich damit zu einem deutlichen Wettbewerbsmerkmal für

---

[1] Mobley (2002): An Introduction to Predictive Maintenance, S. 1
[2] Vgl. Weißmann (2002): Total Quality Management für Industriebetriebe, S. 114
[3] Vgl. Böttcher et al (2017): Machine Learning im Unternehmenseinsatz, S. 3 ff & vgl. Yadav et al. (2015): An Introduction to Neural Network Methods for Differential, S. 14

ein Unternehmen entwickeln. Oftmals werden neben den modellbasierten Methoden auch besonders rechenintensive Algorithmen unter dem Einbezug von Wissensextraktionsmethoden zur Detektion von Störungen eingesetzt. So kann das frühzeitige Aufzeigen von Störungen durch die Erfassung der Sensordaten eines mechanischen Elements größeren Schäden vorbeugen. Für den Einsatz rechenintensiver Algorithmen ergeben sich in Bezug auf die Wirtschaftlichkeit und Genauigkeit der Systeme jedoch viele Fragen, da sich die Vorgehensweisen von den bisher verwendeten ingenieurswissenschaftlichen modellbasierten Ansätzen unterscheiden. Je nach Anwendungsfall bestehen unterschiedliche Anforderungen, die durch eine Vielzahl an Algorithmen erfüllt werden können, weshalb eine fallspezifische Bewertung der Algorithmen vorgenommen werden muss. Weiterhin ergeben sich im Bezug der zur Verfügung stehenden Rechenleistung und des Genauigkeitsbedarfs der Algorithmen unterschiedliche Anforderungen an die Sensorik der mechanischen Systeme. Je größer die Reduzierung der Abtastrate ist, desto mehr wird die Komplexität der erfassten periodischen Sensordatenverläufe reduziert, wodurch potentiell wichtige charakteristische Merkmale nicht mehr erfasst werden. Zudem kann durch den frühzeitigen Einbezug von Expertenwissen im Prozess der Datenvorverarbeitung, ein großer Einfluss auf den weiteren Verlauf der Analyse ausgeübt werden.

Daher wird in dieser Arbeit die folgende Fragestellung untersucht:

> *Wie können Anomalien in periodischen Vorgängen mithilfe von intelligenten Systemen erkannt werden und welche Faktoren besitzen einen Einfluss auf die Störungserkennungsgenauigkeit?*

Zu Beginn der Arbeit werden in Kapitel 2 die Grundlagen erarbeitet und in Kapitel 3 die Literaturanalyse durchgeführt. Mithilfe des Knowledge Discovery in Databases Prozesses wird anschließend die Anomalieerkennung angegangen. Hierfür werden zunächst an einer realen Förderbandanlage Sensordaten erfasst. Bei den zu untersuchenden Faktoren stellt die Abtastrate für die Praxis eine hohe Relevanz dar. Weiterhin wird aufgrund der Beschaffenheit von Perioden eine horizontale Zerlegung dieser als vielversprechend angesehen, weshalb die Faktoren auf diese beiden Größen beschränkt werden. Auf Basis der erarbeiteten Grundlagen werden anschließend im Kapitel 4.3 unterschiedliche, dem Data Mining zugeordneten Verfahren und Algorithmen beschrieben und bewertet. Mithilfe der Bewertungsergebnisse wird in Kapitel 4.3.6.4 ein Algorithmus für die Anomalieerkennung ausgewählt. In Kapitel 4.5 und 4.6 erfolgt die Entwicklung und Programmierung von Datenvorverarbeitungs- und Transformationsalgorithmen. Anschließen wird in Kapitel 4.7 der Algorithmus programmiert und auf den Datensatz angewendet. Abschließend erfolgt in Kapitel 4.8 eine Gegenüberstellung der Szenarien und die Diskussion der erzielten Ergebnisse.

7 Fazit und Ausblick

Im Folgenden werden die Ergebnisse der Arbeit zusammengefasst und ein Ausblick für weiterführende Forschungen gegeben.

Im Rahmen dieser Arbeit wurde die Fragestellung beantwortet, inwiefern Anomalien in periodischen Vorgängen mithilfe von intelligenten Systemen erkannt werden können und welche Faktoren einen Einfluss auf die Störungserkennungsgenauigkeit besitzen. Aufgrund der hohen Relevanz für die Praxis und der in der Literatur bestehenden Forschungslücke wurden die zu untersuchenden Faktoren auf die Abtastrate und die horizontale Zerlegung der Perioden beschränkt. Hierfür wurden nach Erarbeitung der Grundlagen und der Literaturanalyse zunächst die Forschungslücke identifiziert und darauf aufbauend die Forschungsfrage hergeleitet. Mithilfe des Knowledge Discovery in Databases Prozesses erfolgte anschließend die Durchführung der Anomalieerkennung. Hierfür wurden zunächst die in Frage kommenden Modelle und Algorithmen detailliert analysiert und gegenübergestellt. Anschließend wurden die am Versuchsstand erfassten Sensordaten mittels der entwickelten und programmierten Algorithmen automatisiert vorverarbeitet, bereinigt und transformiert. Darauf aufbauend erfolgte das Trainieren und Anwenden der programmierten LSTM-Netze der unterschiedlichen Szenarien. Abschließend wurden die erzielten Genauigkeiten evaluiert und mit den Szenarien gegenübergestellt.

Zunächst lässt sich festhalten, dass bei der Analyse der Algorithmen, durch die Aufteilung in Bewertungsebenen eine schnelle und begründbare Auswahl des geeignetsten Algorithmus vorgenommen werden konnte. Damit kann eine reproduzierbare Vorgehensweise vorgestellt werden, die die Grundlage für die Erzielung eine hohe Störungsklassifikationsgenauigkeit lieferte.

Durch die Anwendung der LSTM-Netze sowohl auf störungsfreie als auch störungsbehaftete Datensätze ergaben sich je nach Szenario unterschiedliche Genauigkeiten. In der ersten Analysephase wurde der Einfluss der Abtastrate auf die Anomalieerkennungsgenauigkeit untersucht. Dabei konnte nachgewiesen werden, dass eine Reduzierung der Abtastrate einen Genauigkeitsverlust bedingt. In dem zweiten Analyseabschnitt erfolgte eine horizontale Zerlegung des Datensatz. Dabei wurde jede Periode in deren drei Bestandteile des Geschwindigkeitsprofils zerlegt und drei LSTM-Modelle auf die Phasen antrainiert. In diesem Analyseabschnitt konnte nachgewiesen werden, dass die abschnittweise Spezialisierung der LSTM-Modelle zu keinem Genauigkeitsgewinn führt. Zwar erzielte dieses System eine deutlich höhere Genauigkeit als das

System in dem die Abtastrate reduziert wurde, im Vergleich zum Grundmodell ohne Phasenzerlegung fiel die Genauigkeit jedoch deutlich geringer aus.

Im Hinblick auf weiterführende Forschungen, wäre der Vergleich der Genauigkeiten zwischen den LSTM-Modellen und bereits bestehenden ingenieurswissenschaftlichen Ansätzen interessant. Zwar konnte mit den LSTM-Modellen, trotz der relativ geringen Datenmengen, eine hohe Genauigkeit erzielt werden, jedoch stellen ingenieurswissenschaftliche modellbasierte Ansätze, wie bspw. die Fast Fourier-Transformation für die Frequenzanalyse, in der Praxis eine bewährte Methode dar. Daher könnte ein Genauigkeits- und Robustheitsvergleich beider Ansätze spannende Ergebnisse liefern.

Die Klassifizierung der störungsbehafteten und störungsfreien Perioden, wurde mithilfe von ermittelten Grenzwerten für Fehlerflächen durchgeführt. Hierbei zeigte sich im Rahmen der Analyse, dass die Fehlerfläche eine Zeitreihe darstellt, welche in einer weiterführenden Forschung ebenfalls durch ein rekurrent neuronales Netz approximiert werden könnte. Damit könnte untersucht werden, ob durch eine weitere Zeitreihenanalyse eine bessere Unterscheidung zwischen störungsfreien- und behafteten Perioden möglich ist.

Zu weiteren wichtigen Erkenntnissen könnte die Ermittlung von Fehlerursachen führen. Allerdings ergab sich im Rahmen der Arbeit die Schwierigkeit, dass keine Aussage über die exakte zeitliche Position möglich war, die für das Auftreten einer Störung verantwortlich ist. Daher könnte eine vorgelagerte Kennzeichnung der für die Anomalie verantwortlichen Phasen vielversprechend sein. Anschließend könnte durch den Einsatz eines weiteren Algorithmus die Ursache der Fehler ermittelt werden.

Zu Beginn des Knowledge Discovery in Databases-Prozesses, wurde aufgrund der Beschaffenheit der Zeitreihe eine Genauigkeitsverbesserung durch eine horizontale Aufteilung der Messreihen vermutet, welche jedoch nicht nachgewiesen werden konnte. Jedoch wurde deutlich, dass LSTM-Netze Schwierigkeiten bei der Approximation von hohen Ausschlägen in periodischen Zeitreihen besitzen. Daher könnte in einer weiterführenden Arbeit die Untersuchung der Genauigkeitsentwicklung in Abhängigkeit einer vertikalen Datensatzaufteilung und Spezialisierung der LSTM-Modelle auf unterschiedliche Ausschläge zu weiteren Erkenntnissen führen.

In der Industrie weisen eine Vielzahl an mechanischen Komponenten periodische Vorgänge auf. Aufgrund der Vielfalt und Komplexität der Anwendungen ist diese Gemeinsamkeit nicht immer auf den ersten Blick, sondern erst bei Zerlegung der Anwendung

auf ihre elementaren Bestandteile, erkennbar. Für eine zukünftige Analyse der Robustheit eines Systems, könnte die Verwendung von Nenngrößen neben einer besseren Vergleichbarkeit auch zu einem Genauigkeitsgewinn führen.