



Universität Stuttgart

Institut für Parallele und Verteilte Systeme
Anwendersoftware

Feature-Extraktion für Sensordaten zur Maschinenüberwachung

Seminararbeit

Advanced Topics in Data Management (WS 2018/2019)

Betreuer: Mathias Mormul

Niklas Kleinhans

Stuttgart, 05.11.2018

Feature-Extraktion für Sensordaten zur Maschinenüberwachung

Niklas Kleinhans

Zusammenfassung. Maschinen werden bis zum kleinsten Bauteil immer intelligenter und ermöglichen es Daten unterschiedlicher Struktur und Komplexität in großen Mengen aufzunehmen. Im industriellen Umfeld wird dabei von „Industrie 4.0“ gesprochen. Der Begriff ummantelt die Beschreibung einer neuen Industriegeneration, in der System durch technologische Weiterentwicklungen intelligent werden und miteinander vernetzt sind. Teil der Intelligenz ist die Maschinenüberwachung. Die Maschinen sollen mithilfe von Livedatenanalyse überwacht und möglichst die preventive Instandhaltung durch präventive ersetzt werden. Um diese Intelligenz zu erreichen werden unter anderem Sensordaten aufgenommen und anschließend zur Analyse weiterverarbeitet. Eine breite Masse komplexer Daten mit der Anforderung an eine schneller Verarbeitung stellt eine große Herausforderung an Analysealgorithmen. Herkömmliche statistische Methoden reichen für den hohen Anspruch an Informationsgewinn nicht mehr aus. In dieser Arbeit werden Verfahren vorgestellt die mithilfe von Feature Extraktion Sensordaten analysieren. Dabei steht der Begriff „Feature“ für ein Merkmal, wodurch sich Daten unterscheiden lassen. Diese werden extrahiert und bilden die Grundlage zur Datendifferenzierung. Es werden die Herausforderungen in der Sensordatenanalyse, speziell in der Livedatenanalyse, beschrieben, Algorithmen vorgestellt und bezüglich ihrer Korrektheit und Laufzeit diskutiert.

Schlüsselwörter: Maschinenüberwachung, Feature Extraktion, Maschinelles Lernen, Sensordaten, Livedaten, Livedatenanalyse

1 Einleitung

-TODO- Umlaute

Ein großer Bereich der vierten Industriegeneration ist das „Internet der Dinge“ oder im Englischen „Internet of Things (IoT)“. Darunter versteht man ein erweitertes Konzept des aktuell bestehenden Internets. Es findet nicht mehr nur ein reiner Daten-/Informationsaustausch mittels unterschiedlicher Medientypen statt, sondern große Systeme bis hin zu kleinsten Komponenten werden Vernetzt und bilden eine Kommunikationsschicht [4]. Durch diese Kommunikation entstehen Datenflüsse die einerseits in großen Rechenzentren gespeichert werden können und andererseits bieten sie eine Grundlage zur Livedatenanalyse.

Maschinen generieren Daten durch Controllersysteme oder Sensordaten. Diese werden zur Steuerung und Kontrolle des Gesamtsystems verwendet. Von

„Condition-Monitoring“ wird gesprochen, wenn diese Daten mithilfe von Diagrammen für den Menschen lesbar dargestellt werden. Dabei können diese unverarbeitet angezeigt oder durch analytische Verfahren vorverarbeitet werden, um beispielsweise Anomalien anzuzeigen. In der „Anomaliedetektion“ wird versucht anhand der generierten Maschinendaten untypische Verhaltensmuster zu erkennen, um auf diese anschließen zu reagieren. Damit diese Aktionen möglichst rechtzeitig ausgeführt werden können, muss die Analyse der Maschinendaten möglichst ohne große Berechnungszeiten geschehen. Diese Livedatenanalyse stellt harte Randbedingungen an Algorithmen, da es sich hier auch meist um große Datenmengen handelt.

Um zusammenhängende Informationen aus den Daten zu gewinnen werden nicht nur einzelne Daten betrachtet, sondern auch ganze Datenpakete, beispielsweise über einen gewissen Zeitraum hinweg. Man spricht von Zeitreihendaten. Im industriellen Umfeld wird diese Analyse meist angewendet um „preventive Instanthaltung“ durch „präventive Instanthaltung“ zu ersetzen. Die preventive Instanthaltung reduziert Maschineneausfälle durch vorbeugende meist durch Zyklen gesteuerte Wartungsarbeiten. Intelligente Systeme sollen dabei unterstützen diese Zyklen durch gezielte Vorhersagen zu ersetzen. Dann spricht man von „prädektive Instanthaltung“. Um Livedaten auf diese Weise zu analysieren, muss meist die Komplexität der Algorithmen so reduziert werden, dass die anschließende Analyse möglichst durch lineare Verfahren durchgeführt werden kann [2]. Man berechnet eine „Kernfunktion“ wendet diese auf die Daten an und analysiert anschließend die Daten auf der neuen Dimensionsebene.

All diesen Methoden und Verfahren liegt das Trainieren von Verfahren auf gewissen Datenmerkmale zu Grunde. In dieser Arbeit wird zu Beginn in Kapitel 3 die grundlegende Struktur von Sensordaten beschrieben, der Zusammenhang mit Livedaten diskutiert und anschließend die damit verbundene Verwendung von „Feature-Extraktions Verfahren“ besprochen. In Kapitel 4 werden die Randbedingungen und Herausforderungen an Algorithmen zur kontinuierlichen Livedatenanalyse besprochen und Ansätze sowie Algorithmen vorgestellt, die sich mit diesem Problem auseinandersetzen.

2 Verwandte Arbeiten

Viele Arbeiten beschäftigen sich damit Sensordaten mit Hilfe von analytischen Methoden zu verarbeiten. Ein Beispiel ist die Arbeit „Using Machine Learning on Sensor Data“ von Alexandra Moraru et al. [6]. In dieser wird die Anzahl von Mitarbeitern im Büro anhand von Sensordaten vorhergesagt. Dabei werden klassische Klassifikations- und Regressionsverfahren angewendet und validiert. Die Verfahren werden auf einen Trainingsdatensatz trainiert und anschließend auf weitere Datenpakete angewendet. Dabei wurden die Merkmale, an welchen die Anzahl der Mitarbeiter vorhergesagt werden sollen vordefiniert.

Bei komplexeren Problemen, wie bei der Maschinenüberwachung, müssen die Merkmale oft erst gefunden werden. Diese Merkmale können neben einfachen Schwellwertüberschreitungen auch beispielsweise gewisse Datenmuster sein. Diese

Andre Gensler, Thiemo Gruber und Bernhard Sick beschreiben Verfahren in ihrer Arbeit „Fast Feature Extraction For Time Series Analysis Using Least-Squares Approximations with Orthogonals Basis Functions“, mit welchen sie diese Merkmale, unter harter Laufzeit und Speicheranforderungen, erkennen.

Die Analyseverfahren benötigen oft sehr viel Rechenleistung da die Probleme, welche einer nicht linearen Komplexität entsprechen, sehr aufwändig zu Berechnen sind. Man spricht dabei von Höherdimensionalen Problemen. Fabian Mörchen beschreibt in seiner Arbeit „Time series feature extraction for data mining using DWT and DFT“ eine Methode um die Dimension von Zeitreihendaten zu reduzieren. Er Optimiert dabei die Auswahl der Koeffizienten und reduziert damit den Berechnungsaufwand für Analyseverfahren.

-TODO- Die weiteren Referenzen zusammenfassen

3 Grundlagen

In diesem Kapitel werden einige Grundlagen für die in Kapitel 4 diskutierten Algorithmen besprochen.

3.1 Was sind Sensordaten

Um Maschinen zu steuern werden oft bidirektionale Datenflüsse verwendet. Das Kontrollsystem gibt Anweisungen und reagiert anschließend auf das Feedback der Maschine. Teilweise werden auch einfache Zeitschranken als Feedback verwendet. Sensoren optimieren das Feedback für das Kontrollsystem. Um Maschinen zu überwachen, können diese Feedback- und Sensordaten verwendet werden um beispielsweise Datenabweichungen festzustellen. Weitere Umgebungssensoren, wie auch Kamerasysteme, Temperatur- und Kontaktsensoren, bieten meist einen Informationshaltigen Datenbestand. Dabei können Temperaturunterschiede, Verschlusszeiten, Druckaufbauzeiten u.ä. betrachtet werden. Die spezielle Herausforderung bei Sensordaten ist der kontinuierliche Datenfluss. Bei laufender Maschine werden stetig Daten generiert sowohl bei Gut- als auch Schlechtdurchläufen. Die Anforderung an das Überwachungssystem ist somit die kontinuierliche Datenverarbeitung. Es wird von einer Livedatenanalyse gesprochen. In der Livedatenanalyse müssen die Daten genauso schnell verarbeitet werden wie sie als Eingabe zur Verfügung stehen, damit sich kein „Datenstau“ ergibt.

-TODO- Beispieldatensatz

Abbildung 1 zeigt die Sensordatenabweichung im Büro innerhalb eines Arbeitstages. Zu jedem Zeitpunkt kann der Datensatz ausgewertet und differenziert werden. Beispielsweise unterscheiden sich die Datenwerte von 8:00 Uhr und 9:99 Uhr maßgeblich an der Helligkeit. Mathematisch würde dieser Zeitpunkt als Vektor dargestellt werden.

$$x_1 = \begin{pmatrix} 38 \\ 20 \\ 25 \end{pmatrix} \quad (1)$$

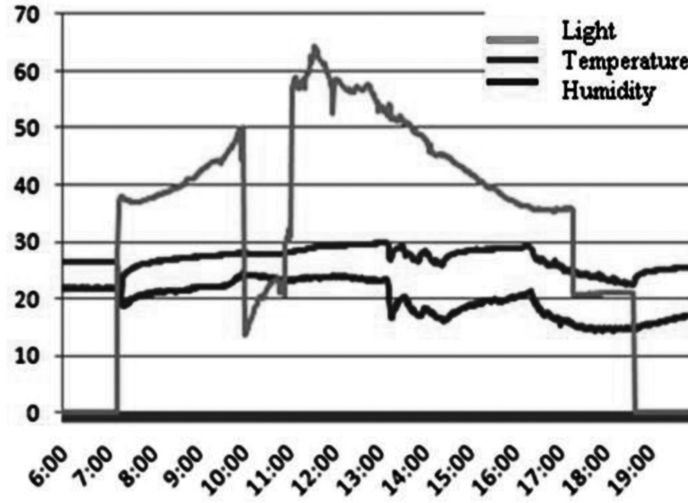


Abb. 1. Sensordatenabweichung anhand von Licht, Temperatur und Drucksensoren im Büro innerhalb von einem Arbeitstag [6]

Durch die Vektordarstellung können die Datenpunkte anhand ihrer Parameter unterschieden werden. Durch betrachten von Zeitreihen, lassen sich an mehreren Tagen Muster erkennen. Solche Zeitreihendaten können Mathematisch als Matrix dargestellt werden.

$$x_{1,4} \begin{pmatrix} 38 & 40 & 45 & 20 \\ 20 & 22 & 23 & 23 \\ 25 & 28 & 29 & 29 \end{pmatrix} \quad (2)$$

Um Mathematisch mehrere Tage in einen Datensatz darzustellen erhalten wir schon Tensoren. Schon anhand dieses simplen Beispiels wird die Datenmenge und Komplexität der Daten ersichtlich. Algorithmen zur Livedatenanalyse solcher komplexer Datenbestände werden in Kapitel 4 besprochen.

-TODO- Es fehlt noch ein Kapitel für kurze ML einleitung. Regressionen, Klassifikationen. Und den Bezug zur Maschinenüberwachung.

Speziell in der Maschinenüberwachung ist das Ziel solche Daten nicht nur im aktuellen Zeitpunkt zu bewerten, sondern eine Vorhersage zu treffen. Angenommen wir betrachten einen Datensatz zum Zeitpunkt t_1 . Dann besteht der zu bewertender Datensatz aus einem Tupel $\tau_1 = (t_1, x_1)$. Diesem Tupel wird abhängig von den verwendeten Verfahren eine Klasse y_1 zugewiesen [2]. Für den Zeitraum (t_1, \dots, t_n) mit $n \in \mathbb{N}$ erhalten wir den Datensatz

$$D = \{(\tau_1, y_1), \dots, (\tau_n, y_n)\} \quad (3)$$

Das Ziel ist das Tupel (t_{n+1}, y_{n+1}) vorherzusagen.

3.2 Was versteht man unter Feature Extraktion

In Kapitel 3.1 wurde ein Datensatz durch seine Muster und Merkmale beschrieben. Merkmale können einfache Parameter, wie Schwellwertüberschreitungen sein. In Komplexeren Datenstrukturen, können die Daten anhand von Verlaufsmustern unterschieden werden.

Um Algorithmen zu trainieren die Daten durch diese Merkmale und Muster zu unterscheiden, gibt es zwei wesentliche herangehensweisen. Es können die Merkmale dem Algorithmus vorgegeben und auf diese Merkmale trainiert werden, oder es werden Verfahren angewendet um diese Merkmale zu extrahieren. Es ist die Rede von „Feature Extraktion“. Ein „Feature“ ist eines dieser Merkmale, wodurch Daten in einem Datensatz voneinander unterschieden werden können.

Dabei ist das Ziel die Features so zu wählen und zu parametrisieren, dass das ermittelte Wert dem tatsächlichen Wert möglichst ähnelt. Betrachten wir die Gleichung 3 und die daraus ermittelte Funktion $p(x_i)$, welche das Ergebnis des gewählten Algorithmus und des Trainingsdatensatzes ist. Es soll versucht werden den Fehlerunterschied

$$\sigma = p(x_i) - y_i \quad (4)$$

möglichst zu reduzieren [3]. Konkret wird meist die Summe der Quadratischefehler

$$\sigma^2 = \frac{1}{N} \sum_{i=0}^K (p(x_i) - y_i)^2 \quad (5)$$

versucht zu reduzieren.

- TODO - Ein konkretes Beispiel für Feature Extraktion raussuchen und daran erklären. Es muss auch noch die Notwendigkeit und die herangehensweise von Dimensionsreduzierung erläutert werden.

- Kurze ML einleitung mit erklärung zur Feature-Extraktion
- Feature extraction vs Feature selection
- „extrahieren von Merkmalen, wodurch die Daten in einem Datensatz voneinander unterschieden werden können“

4 Feature Extraktion bei kontinuierlichen Livedaten

- Grundsätzliche Ansätze beschreiben [3]
- Time Series Decomposition [5]

Beispiel Algorithmen:

- Benutzung von Least-squares approximations um Features zu extrahieren
- DWT DFT [2]
- Noch aufstocken mit Algorithmen für WSN [1]

5 Zusammenfassung

Literatur

1. Alsheikh, M.A., Lin, S., Niyato, D., Tan, H.: Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys Tutorials* **16**(4), 1996–2018 (Fourthquarter 2014). <https://doi.org/10.1109/COMST.2014.2320099>
2. Gay, D., Guigourès, R., Boullé, M., Clérot, F.: Feature extraction over multiple representations for time series classification. In: *International Workshop on New Frontiers in Mining Complex Patterns*. pp. 18–34. Springer (2013)
3. Gensler, A., Gruber, T., Sick, B.: Fast feature extraction for time series analysis using least-squares approximations with orthogonal basis functions. In: *Temporal Representation and Reasoning (TIME)*, 2015 22nd International Symposium on. pp. 29–37. IEEE (2015)
4. Jasperneite, J.: Was hinter begriffen wie industrie 4.0 steckt. *Computer & Automation* (19 Dezember 2012), <https://www.computer-automation.de/steuerungsebene/steuern-regeln/artikel/93559/0/>
5. Michale A. Tovero, M.J.L.: Time series feature extraction. SAS Institute Inc. (2018), <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2020-2018.pdf>
6. Moraru, A., Pesko, M., Porcius, M., Fortuna, C., Mladenic, D.: Using machine learning on sensor data. *Journal of computing and information technology* **18**(4), 341–347 (2010)

Alle Links wurden zuletzt am 10.12.2018 geprüft.