

# Titel der Abschlussarbeit

Bachelorarbeit  
von  
Marius Wodtke

an der Fakultät für Informatik

In dem Studiengang  
Informatik (B.Sc.)

eingereicht am 01.01.01 beim  
Institut für Angewandte Informatik  
und Formale Beschreibungsverfahren  
des Karlsruher Instituts für Technologie

Referent: Prof. Dr. Hartmut Schneck  
Betreuer: Kaibin Bao

KIT – Universität des Landes Baden-Württemberg und  
nationales Forschungszentrum in der Helmholtz-Gemeinschaft





## Eidesstattliche Erklärung

Ich versichere hiermit wahrheitsgemäß, die Arbeit und alle Teile daraus selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderung entnommen wurde.

*Karlsruhe, den DATUM*

*NAME*



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Fragestellung . . . . .	1
1.3	Weiterer Aufbau . . . . .	2
<b>2</b>	<b>Datensatz</b>	<b>3</b>
2.1	Gemessene Daten . . . . .	3
2.2	Berechnete Daten . . . . .	3
2.3	Die Waschmaschine . . . . .	4
<b>3</b>	<b>Vorverarbeitung</b>	<b>5</b>
3.1	Stand der Technik . . . . .	5
<b>4</b>	<b>Fazit</b>	<b>6</b>
<b>A</b>	<b>Anhang</b>	<b>7</b>

## Abkürzungsverzeichnis

NILM	Non-intrusive load monitoring
KNN	Künstliches Neuronale Netz

# Abbildungsverzeichnis



# Tabellenverzeichnis

# 1 Einleitung

## 1.1 Motivation

Energie zu sparen ist seit Langem ein Ziel der Umweltpolitik, dennoch steigt der Energieverbrauch in Industrieländern kontinuierlich und es werden nach wie vor Wege gesucht, diesen zu senken. In den USA verursachen private Haushalte 40% des CO<sub>2</sub>-Ausstoßes [VBG08] und stehen deshalb im Fokus vieler Programme zum Energiesparen.

Je nach Studie besteht für die Haushalte ein Einsparungspotential von bis zu 20% [AGSA13], Fischer [Fis08] untersucht mehrere dieser Studien und beschreibt ein durchschnittliches Einsparungspotential von 5-12%. Um dieses Einsparungspotential auszunutzen, muss der Nutzer regelmäßig und möglichst genau über seinen Verbrauch aufgeklärt werden. Er kann sich oft nicht mit seinem Verbrauch identifizieren, weil dieser intransparent und durch die langen Rechnungsintervalle nicht präsent ist.

Damit der Nutzer anfängt sich selbst zu kontrollieren, muss er begreifen, dass sich sein Verhalten auf seinen Verbrauch auswirkt und er ihn auch durch die gezielte Veränderungen seines Handelns senken kann [Fis08].

Non-intrusive load monitoring (NILM) bietet nun die Chance dem Nutzer eine regelmäßige, detaillierte Rückmeldung zu seinem Energiekonsum zu geben, Kolter und Matthew [KJ11] beschreibt NILM als die Aufgabe aus einem, für den gesamten Haushalt messenden Stromzähler, Rückschlüsse über die elektrische Last einzelner Geräte zu ziehen. Ein solches System wird dem Konsumenten helfen, verschwenderische Verhaltensweisen und Geräte zu identifizieren ohne den Haushalt mit vielen digitalen Zählern für die individuellen Geräte ausrüsten zu müssen, wie es derzeit z.B. mit sogenannten *Energiekostenmessgeräten* auf Steckerbasis praktiziert wird.

## 1.2 Fragestellung

Viele Systeme zur Disaggregation oder zu verschiedenen Vorhersage-Aufgaben benutzen Ansätze des Maschinellen Lernens. Sie benötigen annotierte (gelabelte) Trainingsdaten, um robuste Modelle zu erzeugen. Annotieren bedeutet in diesem Zusammenhang, die Daten mit Metadaten wie etwa dem Gerätezustand zu versehen. Mehr Trainingsdaten führen in der Regel zu besseren Modellen, die Annotation ist allerdings sehr aufwändig, weil sie manuell erfolgen muss. Der Ansatzpunkt dieser Arbeit ist, das Problem der Disaggregation auf ein Gerät zu reduzieren und so zu vereinfachen. Nun kann man ein robustes Modell mit nur wenigen Trainingsdaten erstellen und mit diesem dann eine große Menge an Daten schnell annotieren. Diese gelabelten Daten sollen dann wiederum als Trainingsdaten für schwierigere Aufgaben verwendet werden. Aufgabe dieser Bachelorarbeit ist es ein System zu entwickeln, welches in der Lage ist die Zustände von ausgewählten Geräten in einem disaggregierten Lastgang zu klassifizieren und so eine Segmentierung für diese Geräte zu erstellen. Dies beinhaltet sowohl die nötige Vorverarbeitung der Daten sowie eventuelle nachträgliche Formatierungen. Die eigentliche Klassifikation soll mit Künstlichen

Neuronalen Netzen (KNN) stattfinden, dabei sollte eine möglichst hohe Akkurarität erreicht werden, da Systeme, die mit den resultierenden Daten trainiert werden keine Möglichkeit haben Fehler, die bereits in den Trainingsdaten sind, zu korrigieren.

### 1.3 Weiterer Aufbau

Im folgenden Kapitel 2 werden die verwendeten, disaggregierten Energiedatensätze beschrieben. Insbesondere werden die gemessenen Werte und die daraus berechenbaren Werte untersucht. Kapitel 3 gibt zunächst einen Überblick über die aktuelle Forschung im Bereich Vorverarbeitung der Daten und Feature-Auswahl, anschließend werden die für diese Arbeit verwendeten Vorverarbeitungsschritte und Features erläutert. In Kapitel 4 soll schließlich die eigentliche Klassifizierung beschrieben werden, hier werden insbesondere das verwendete Netz und die Trainingsmethoden besprochen. Auch hier wird es einen kurzen Überblick über die aktuelle Forschung geben. In Kapitel 5 findet die Klassifikation statt, hier werden verschiedene Klassifizierungsaufgaben ausgeführt und ausgewertet. Am Schluss steht das Fazit, in dem die Ergebnisse zusammengefasst und mit ursprüngliche Fragestellung verglichen werden.

## 2 Datensatz

Die verwendeten Daten stammen aus dem *KIT Smart Home Lab*. Sie wurden über fast 3 Jahre zwischen dem 22.08.2011 und dem 31.07.2014 aufgezeichnet, weisen aber einige Lücken auf. Für die Messungen wurden mehrere Smart Meter verwendet, bei Großverbrauchern wie etwa dem Boiler wurden dabei alle drei Phasen gemessen, sie sind durch die Nummer des Smart Meters in den Daten identifizierbar. Bei den restlichen Geräten wurde nur je eine Phase gemessen, das heißt, dass je 3 Geräte an einem Smart Meter angeschlossen sind, sie sind durch die Nummer des Smart Meters in Kombination mit Nummer des verwendeten Ports identifizierbar. Wie viele Phasen zur Messung eines Geräts verwendet wurden lässt sich dem Controller entnehmen. (Tabelle mit Beispiel Zuordnung)

### 2.1 Gemessene Daten

Die Daten des *KIT Smart Home Lab* besitzen folgendes Format. (Tabelle mit Beispielen aus dem Datensatz)

Controller, Meter und Port werden, wie oben beschrieben, verwendet, um einen Datenpunkt einen eindeutigen Gerät zuzuordnen. Hier ist zu beachten, sich diese Werte immer nur gemeinsam verwendet werden dürfen. Die Spalte mit der UUID-Zuordnung wird ignoriert, denn sie enthält ebenfalls die Information um welches Gerät es sich handelt und ist somit Redundant zu Controller, Meter und Port.

Ein Datenpunkt enthält die gemessene Spannung, die Stromstärke und die Wirkleistung. Es wird zusätzlich die aggregierte Wirkleistung gemessen, welche aber zunächst ignoriert wird, da sie sich aus der Wirkleistung berechnen lässt.

Eine Besonderheit stellt das Aufzeichnungsintervall dar. Die Smart Meter erzeugen jede Sekunde einen neuen Datenpunkt für jeden Port, dieser wird aber nur gespeichert, wenn sich die Wirkleistung im Vergleich zum letzten gespeicherten Punkt um 5 Watt Wirkleistung unterscheidet. So werden insbesondere lange *Off* Phasen auf einen Datenpunkt komprimiert, es gehen aber auch kleinere Schwankungen in der Wirkleistung von sehr verbrauchsarmen Geräten verloren.

### 2.2 Berechnete Daten

Die Messung von Spannung und Stromstärke erlaubt es eine Vielzahl von Energiewerten zu berechnen, in diesem Abschnitt werden die Schein- und Blindleistung vorgestellt, das Kapitel zur Vorverarbeitung wird zusätzlich eine Form der Normalisierung vorstellen. Die Scheinleistung ist das Produkt aus Spannung und Stromstärke und ist die gesamte, aus dem Netz gezogene Leistung:

$$\text{Scheinleistung} = \text{Spannung} * \text{Stromstärke}$$

Zur Berechnung der Blindleistung werden Schein- und Wirkleistung genutzt, die Blindleistung ist dabei Leistung, die aus dem Netz gezogen wird ohne tatsächlich genutzt zu werden:

$$\text{Blindleistung} = \sqrt{\text{Scheinleistung}^2 - \text{Wirkleistung}^2}$$

## 2.3 Die Waschmaschine

Die Waschmaschine ist ein besonders interessantes Gerät für die Klassifikation, denn sie hat nicht nur einen *On*- und einen *Off-Zustand*, sondern einen Motor, der mit verschiedenen Drehzahlen operieren kann, ein Heizelement und eine Pumpe. Für die Waschmaschine existieren außerdem einige annotierte Waschgänge die als Trainingsdaten verwendet werden können. (Tabelle mit Angaben zu Profil 0,1,4,5,8)

Die Waschgänge sind mit sechs verschiedenen Klassen Sigma 0 bis Sigma 5 annotiert, Sigma 0 ist der *Off-Zustand*. Sigma 1 ist ein normaler Schleudervorgang, hier sind Phasen mit 200-350 Watt und dazwischen kurze Pausen mit ca. 4 Watt Verbrauch typisch. Sigma 2 ist ein Heizvorgang, dieser kennzeichnet sich durch einen Verbrauch über 1700 Watt. Sigma 3 und Sigma 4 sind Pumpvorgänge und werden durch Verbrauchspitzen von wenigen Sekunden charakterisiert. Unterschieden werden sie anhand der Höhe der Spitze, ein Vorgang Sigma 4 folgt immer auf einen Vorgang Sigma 3 mit einem Abstand von 50 bis 80 Sekunden. Sigma 5 ist das Ausschleudern, es ist ebenfalls ein Schleudervorgang, der aber im Gegensatz zu Sigma 1 ohne Pausen ausgeführt wird. (Bild Profil 0 mit Klassen angezeichnet)

## 3 Vorverarbeitung

### 3.1 Stand der Technik

Die Wirkleistung ist das ursprünglichste Merkmal um die Zustände von Geräten zu unterscheiden. Sie gibt die elektrische Leistung an, die von einem Gerät in andere Leistungen (z.B. Wärme oder Bewegung) umgewandelt werden kann. Oft wird die Wirkleistung durch gröbere Quantisierung oder Normalisierung geglättet, um die Erkennung der Klassen zu vereinfachen. [Har92] verwendet hierzu folgende Formel:

$$P_{norm}(t) = \left(\frac{Nennspannung}{V(t)}\right)^2 * P(t)$$

mit  $P :=$  Wirkleistung in Watt und  $V :=$  Spannung in Volt, die Nennspannung beträgt in Europa 230 Volt.

Die Normalisierung bietet den Vorteil, dass sie nicht verlustbehaftet ist und wird deshalb häufig verwendet, wenn die Spannung gemessen wurde.

Mit der Wirkleistung allein ist man jedoch nicht in der Lage Geräte mit sehr ähnlichem Verbrauch zu unterscheiden, ein Beispiel wären hier ein 2kW Motor und ein 2kW Heizelement. Diese lassen sich mit Hilfe der Blindleistung, welche nicht in andere Leistungen umgewandelt wird, unterscheiden, weil der Motor als induktiver Verbraucher mehr Blindleistung als das Heizelement aus dem Stromnetz zieht. Nachteil hier ist jedoch, dass die Blindleistung (und auch die Spannung) zusätzlich gemessen werden muss, die Berücksichtigung dieses Features kostet also gegebenenfalls Geld [ZR11].

Vergleicht man jedoch eine 60W Glühbirne und einen Laptop mit einem 60W Netzteil, dann stellt man fest, dass diese sich auch unter Hinzunahme der Blindleistung kaum unterscheiden lassen [LLC<sup>+</sup>03]. Hier hilft die Betrachtung von micro level Features. Dabei werden die Wellenformen und die harmonischen Komponenten des Frequenzspektrums verwendet, um Verbraucher zu unterscheiden. Laughman et al. [LLC<sup>+</sup>03] zeigen z.B., dass sich Glühbirne und Netzteil in der 3. harmonischen Komponente deutlich unterscheiden. Zusätzlich verändern viele Geräte die Wellenform des Stroms, so dass in diesem Bereich weiteres Potential für die Unterscheidung verschiedener Geräte vorliegt [LNKC10].

Micro level Features bieten sehr gute Unterscheidungsmöglichkeiten, benötigen aber auch ein sehr hochfrequent aufgelöstes Signal, üblich sind mehrere kHz. Anderson et al. [AOB<sup>+</sup>12] bietet einen Datensatz mit 12kHz Auflösung.

Messgeräte für eine solche Auflösung sind wesentlich teurer als die häufig für Wirk- und Blindleistung verwendeten Geräte mit einer Auflösung von 1Hz und sind im Gegensatz zu Letzteren oft nicht in normalen Haushalten mit Smartmeter vorhanden [ZR11].

Da sie in der Praxis häufiger vorhanden sind wird in dieser Thesis mit den macro level Features mit einer Auflösung von 1Hz gearbeitet.

## 4 **Fazit**

# **A Anhang**



## Literatur

- [AGSA13] ARMEL, K CARRIE, ABHAY GUPTA, GIREESH SHRIMALI und ADRIAN ALBERT: *Is disaggregation the holy grail of energy efficiency? The case of electricity*. Energy Policy, 52:213–234, 2013.
- [AOB<sup>+</sup>12] ANDERSON, KYLE, ADRIAN OCNEANU, DIEGO BENITEZ, DERRICK CARLSON, ANTHONY ROWE und MARIO BERGES: *BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research*. In: *Proceedings of the 2nd KDD workshop on data mining applications in sustainability (SustKDD)*, Seiten 1–5, 2012.
- [Fis08] FISCHER, CORINNA: *Feedback on household electricity consumption: a tool for saving energy?* Energy efficiency, 1(1):79–104, 2008.
- [Har92] HART, GEORGE WILLIAM: *Nonintrusive appliance load monitoring*. Proceedings of the IEEE, 80(12):1870–1891, 1992.
- [KJ11] KOLTER, J ZICO und MATTHEW J JOHNSON: *REDD: A public data set for energy disaggregation research*. In: *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, Band 25, Seiten 59–62. Citeseer, 2011.
- [LLC<sup>+</sup>03] LAUGHMAN, CHRISTOPHER, KWANGDUK LEE, ROBERT COX, STEVEN SHAW, STEVEN LEEB, LES NORFORD und PETER ARMSTRONG: *Power signature analysis*. Power and Energy Magazine, IEEE, 1(2):56–63, 2003.
- [LNKC10] LIANG, JIAN, SIMON KK NG, GAIL KENDALL und JOHN WM CHENG: *Load signature study - Part I: Basic concept, structure, and methodology*. Power Delivery, IEEE Transactions on, 25(2):551–560, 2010.
- [VBG08] VANDENBERGH, MICHAEL P, JACK BARKENBUS und JONATHAN M GILLIGAN: *Individual carbon emissions: The low-hanging fruit*. UCLA Law Review, 55:08–36, 2008.
- [ZR11] ZEIFMAN, MICHAEL und KURT ROTH: *Nonintrusive appliance load monitoring: Review and outlook*. IEEE Transactions on Consumer Electronics, Seiten 76–84, 2011.