

Abstract

The integration of renewable energy sources yields multifaceted requirements for controlling building energy systems (BES). Simultaneously, the increasing availability of sensors and monitoring data supports a deeper integration of data driven algorithms. Previous researches have proven that Model Predictive Control (MPC) can handle the multidimensionality and volatility of system components, the complex constraints from BES, and the inherent uncertainties of the renewable energy supply. To create the model for MPC, hybrid automata can be employed and implemented by rule mining from sensor data. Automata are characterized by being more human-understandable in comparison to the other black-box machine learning methods. For the achievement of a data-driven modeling, there is a strong demand to extract meaningful knowledge automatically from the collected data in the preliminary stage, which avoids time-consuming manual evaluation and deployment of expert knowledge.

In this thesis we present a tool for the data mining of time series data, which offers knowledge on operating state segmentation and favors the creation of hybrid automata as control models for MPC. This tool works firstly on filtering the sensor data while preserving the information density. Afterward, meaningful features are extracted from filtered data by using the time series feature extraction library (TSFEL) package. After that, features are selected by low-variance filtering, transformed by principal components analysis (PCA), scored by a wrapper method and then evaluated individually by calculating how they improve the key performance indicators (KPI) of cluster results. Additionally, the best number of clusters is estimated by a prior calculation of KPIs. Whereafter, hidden patterns are discovered through cluster analysis by using k-Means, k-Medoids and Motif-Aware State Assignment (MASA). Finally, state segmentation results are analyzed by expert knowledge and assessed through its performance in Sparse Identification of Nonlinear Dynamics (SINDy), which indicates how accurate the controlled variables are predicted within each operation state.

Zusammenfassung

Durch die Integration erneuerbarer Energie und die zunehmende Verfügbarkeit von Sensoren und Überwachungsdaten ergeben sich vielseitigere Anforderungen an die Regelung von Gebäudeenergiesystem (BES). Frühere Forschungen haben die Fähigkeit der modellprädiktiven Regelung (MPC) bewiesen, dass sie in der Lage ist, die Mehrdimensionalität und Volatilität der Systemkomponenten, die komplexen Nebenbedingungen des BES und die sich aus erneuerbarer Energie ableitenden inhärenten Unsicherheiten in der Vorhersagbarkeit der Energieversorgung, zu handhaben. MPC führt die Optimierung der Regelgrößen des aktuellen Zeitfensters gemäß der Vorhersage des nächsten Zeitfensters. Die Generierung des Prozessmodells einer MPC kann viele Schwierigkeiten mit sich bringen, z. B. bei der automatischen Identifikation dominanter Modellparameter, der Beherrschbarkeit von multivariablen Systemen und der Simulation der Nichtlinearitäten und hybriden Eigenschaften. Um das Modell für MPC zu erstellen, können hybride Automaten eingesetzt werden, welche für den Menschen verständlicher als Black-Box maschinelle Lernverfahren sind und durch Regelextraktion aus Sensordaten implementiert werden können. Zur Implementierung einer datengesteuerten Modellierung besteht ein starker Bedarf, aus den gesammelten Daten in der Vorstufe automatisch aussagekräftiges Wissen zu extrahieren, wodurch eine zeitaufwändige manuelle Auswertung und der Einsatz von Expertenwissen vermieden werden.

In dieser Arbeit wird ein Tool für das Data Mining von Zeitreihendaten vorgestellt, welche die Segmentierung zu den Betriebszuständen anbieten kann und die Erstellung von hybriden Automaten als Steuerungsmodelle für MPC begünstigt. Zur Verifizierung des Tools wird ein aus HiL-Experiment einer Wärmepumpe gesammelter Datensatz verwendet. Das Tool arbeitet zunächst mit dem Filtern der Sensordaten unter Beibehaltung der Informationsdichte. Danach werden aussagekräftige Merkmale aus den gefilterten Daten mit Hilfe des Time Series Feature Extraction Library (TSFEL) Pakets extrahiert. Nachfolgend werden die Merkmale durch Low-Variance-Filterung ausgewählt, durch Principle Component Analysis (PCA) transformiert, durch eine Wrapper-Methode bewertet und dann einzeln ausgewertet, wie sie die Key Performance Indicators (KPI) der Cluster-Ergebnisse verbessern. Zusätzlich wird die beste Anzahl von Clustern durch eine vorherige Berechnung der KPIs eingeschätzt. Danach werden verborgene Muster durch die Cluster-Analyse mit Motif-Aware State Assignment (MASA) aufgedeckt, deren Ergebnisse dann mit denen von Baseline-Algorithmus d. h. k-Means und k-Medoids verglichen werden. Schließlich werden die Ergebnisse der Zustandssegmentierung aus Cluster-Analyse mit Hilfe von Expertenwissen analysiert und anhand ihrer Leistung in der Sparse Identification of Nonlinear Dynamics (SINDy) bewertet, die angibt, wie genau die Regelgrößen innerhalb jedes Betriebszustands vorhergesagt werden können.

Table of Contents

Nomenclature	VII
List of Figures	IX
List of Tables	XIII
1 Introduction	1
1.1 Motivation	1
1.2 Aim and structure of this thesis	3
2 Background	5
2.1 Introduction to building automation system and model predictive control	5
2.1.1 Building automation systems	5
2.1.2 Model predictive control	6
2.1.3 Model creation	9
2.2 Change point detection	10
2.2.1 Change points	10
2.2.2 Different cluster methods	11
2.2.3 KPIs	14
2.3 Filtering	16
3 Methodology	19
3.1 Workflow overview	19
3.2 Data preprocessing	22
3.3 Parameter estimation for the Butterworth filter	22
3.4 Feature engineering and feature extraction	25
3.5 Feature selection	27
3.6 Ground truth labeling	30
3.7 Evaluation of cluster results	31
3.8 System identification	32
4 Case study	33
4.1 Dataset	33

4.2	Result	36
4.2.1	Filtering	36
4.2.2	Feature extraction and prefiltering	42
4.2.3	Feature selection	46
4.2.4	Best number of clusters	51
4.2.5	Evaluation	52
4.3	Comparison with the selected dataset	61
5	Discussion	77
6	Conclusion and future work	81
	Bibliography	83