Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Master Thesis Computer Science

Title of thesis

Ludwig Bald

Datum

Reviewers

Dr. Nicole Ludwig
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Jun. Prof. Dr.-Ing. Setareh Maghsudi
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Bald, Ludwig:
Title of thesis
Master Thesis Computer Science
Eberhard Karls Universität Tübingen
Thesis period: von-bis

Abstract

Write here your abstract.

Zusammenfassung

Bei einer englischen Masterarbeit muss zusätzlich eine deutsche Zusammenfassung verfasst werden.

Saake et al. (1997) Age Canteli and Ramón (2020)

Acknowledgements

Write here your acknowledgements.

Contents

List of Figures

List of Tables

List of Abbreviations

BLAST	Basic Local Alignment Search Tool
...	...

Chapter 1

Introduction

Start with a comprehensive introduction about the questions of your thesis. 1-2 pages:

When proofreading: Check that all terms and abbreviations are introduced here

Climate Change is the global challenge of our lifetime. Carbon introduced in the atmosphere when burning fossil fuels for human needs causes global warming, destabilizing the climate, ecosystems and societies around the world. In the Paris Agreement of ... governments have committed to an ambitious goal of drastically reducing carbon emissions in order to keep global warming from increasing beyond 2°C, compared to ... The latest IPCCC report urges governments to take stronger actions, or their previous commitment will not be reached. A key strategy for reducing carbon emissions from a range of sources is the combination of two measures: The first step is to electrify current processes that use fossil fuels, like replacing gas-fired furnaces with heat pumps. The second step is to replace carbon-intensive electricity generation with renewable options like solar and wind power.

cite

cite

While much better for the natural environment, renewable sources of energy pose a challenge for a grid built for fossil fuels: Unlike fossil fuelled power plants, renewable power production depends on the weather, and it can not react flexibly to changes in demand.

As the share of installed renewable sources of electricity continues to grow from today's...%, the reliability of electricity supply goes down.

cite

talk about reacting to less reliability in renewable production, not only reacting to demand

In order to keep the grid stable, supply and demand must always be in balance. Before the green transition, this was achieved by flexible power generation: When demand was high, electricity producers were able to react and increase production. This was incentivized by a complex and tightly regulated market constructed on top of the physical layer. As the share of renewable power increases, fossil-fuelled plants remain the only market participants that can flexibly react to changes in demand. In order to phase out fossil-fuelled power generation, this flexibility needs to be provided by different parts of the system.

There are dedicated electricity storage facilities that can react very quickly to stabilize the grid by storing and releasing energy as needed. However, consumer electricity demand can flexibly react to changes in supply, a scheme called Demand Response. Grid-scale storage in Europe is mainly comprised of hydropower, which has been installed where mountainous geography allows, and capacity has reached its natural limit. More expensive battery-powered storage facilities are slowly being built, but are largely not cost-efficient in the current economic setting.

cite

In this thesis, I focus on a possibility to make consumer electricity demand more flexible.

Buildings require energy mainly for heating and cooling the air and the water supply. Today, they are responsible for ...% of total energy demand. On the other hand, buildings often contribute to electricity production through photovoltaic panels. New buildings often come with a battery, which enables them to more efficiently use their solar electricity.

cite

Building's electricity consumption is already largely automated and is therefore a prime candidate for automated demand response.

mention incentive-based human DR

The Reinforcement Learning family of control algorithms has successfully been applied for control of the battery of simulated buildings. . A popular simulation framework for this purpose is CityLearn. In this thesis, my goal is to test Uncertainty-Aware Deep Q-Networks on CityLearn. UA-DQN is an uncertainty-aware adaptation of Deep Q-Learning.

list other automated control algorithms and their goals

cite algorithms

cite CityLearn

cite algorithm

cite DQN

The algorithm's better treatmeant of uncertainty should lead to overall better performance with less need for data, better robustness for novel data, and can even be leveraged for differently risk-aware charging and discharging strategies.

talk about results!

These aspects are important for a demand response algorithm, which provides flexibility to the grid while reliably meeting building demands for energy.

This thesis is structured as follows: In the following chapter, I motivate in more detail the need for Automated Demand Response. I introduce the theory of Reinforcement Learning and lay the foundations for the uncertainty-aware algorithm. In chapter 3, I present the uncertainty-aware algorithm, as well as a detailed description of the experimental setup. I present the results in chapter 4. A discussion and short outlook conclude the thesis.

rephrase

—— - Results (abstrakt success vs failure) (technical, numbers)

Terms: - Demand Response vs. Demand Side Management - Reinforcement

Learning

Chapter 2

Background

Give all the necessary background that is needed to justify and understand the problem and the approach. - comment on employed hardware and software - describe methods and techniques that build the basis of your work - review related work(!) - roughly 1/3 of thesis

- flesh out beginning from introduction, support general argument
1. supply side rigidity (plot of different energieträger over time) 2. status quo demand side rigidity 3. ansatzpunkte für Flexibilität in Demand Side mit RL oder so, konkrete Beispiele
4. RL ...

Missing figure

Show reliability of renewables over a day

——- Erster Entwurf below ——

2.1 The Electricity Grid: A Technical and Structural Overview

In places with a well-managed electricity grid, consumers know electricity as a reliable resource that is always available, in whatever quantity needed. This is a remarkable feat of engineering and cooperation between a number of different actors, who I will briefly introduce in this section. Each actor acts according to their own interests.

diagram that explains the grid. Physical, technical, market abstractions. Include the environment and politics.

move electricity grid to introuction, and make it much shorter

2.1.1 Consumers

- private: expect electricity for convenience and safety - industrial: need electricity for economic activity - now also Producers, "Prosumers" - are electrifying more and more processes: Transportation, heating, appliances, etc. - risks: Price risks; running out of electricity; producing more than they are able to sell,.

2.1.2 Utility Companies

- buy electricity on the markets

2.1.3 Electricity Markets

- electricity lake - buy/sell contracts - different timescales and area scales.

- decoupled from the actual power transmission grid.

2.1.4 Electricity Producers

- define: renewable (solar, wind, maybe others?) vs conventional (nuclear, coal, gas)

2.1.5 Transmission System Operators

- manage transmission and electricity flow - are responsible for keeping the grid synchronized - can request electricity production if there is not enough power

2.1.6 Local Grid Operators / Utility companies

- manage local grid, responsible for getting power to consumers

2.1.7 Politics / Society?

- guarantees comfort for consumers - international relations - pressure to save CO_2 - designs and oversees the system's mechanisms

2.1.8 Environment

- provides energy as fossil fuels, renewables, other. - Has Limits to the CO2 it can absorb - can be responsibly exploited

2.1.9 Challenges for the Grid/ Sustainable transition

- Political tensions, global dependence - Ecological dependence on the environment - Green transition: Renewables are much less reliable - Prosumers: power at new places in the grid - Electrification of everything (heat pumps, transport, industry) –¿ Need much more capacity, quickly

- Programs to cope with this (alternatives/complements to Demand Response): - increasing grid and production capacity - increasing storage capacity to smooth out renewable production - increase grid efficiency (better grid management) - - ... TODO Research

2.1.10 Flexibility (/Demand Response)

This section introduces the notion of flexibility and therefore motivates demand response. It should also maybe talk about conflicts of interest, and incentives.
need to be defined before this section: - conventional power - renewable power - supply/demand vs production/use

- flexibility is needed both in the market layer and in the physical layer

Renewable electricity is less flexible than conventional electricity production, therefore there is a need for more flexibility elsewhere in the system. Electricity production and electricity use must always match. If there is more or less electricity used than is available, the entire grid is no longer stable. Partly, this coordination problem between producers and consumers is solved by the electricity market: An electricity producer will produce and feed to the grid only as much electricity as they can sell. This system works fine if the electricity producer can freely regulate how much power they produce based on the market.

This is not true for renewable energy, where production capacity is both determined by external factors and harder to predict. Renewable sources of electricity are therefore less flexible. In order to keep production and consumption in balance, there needs to be a system component with enough flexibility to respond to sudden changes in electricity supply and demand. This can be a hydroelectric or grid-scale battery storage system which can both produce and consume electricity. It can also be a conventional power plant that's kept in spinning reserve, to take over when there's less renewable electricity than

define

needed. This causes carbon emissions. Often, renewable power plants can't feed all of their production into the grid, because they can't find a flexible buyer for unexpected production. This motivates the need for flexibility on the demand side.

2.1.11 Demand Response

- a scheme to operate the grid more efficiently as demand rises quickly. - breaks with one of the central guarantees: electricity is freely available, accepting that renewable energy is much less reliable. - Mechanism: Incentivize consumers to use electricity when there's capacity. - implementations: TODO examples (Large-scale AC cuts, incentive programs, etc.) - already in place for large industrial consumers who buy electricity on the markets. (TODO: Is that true?) - some industrial consumers are paid for providing flexibility, they can stop their processes if there's not enough electricity being produced. - more difficult for private consumers, who don't want to think about efficiency all the time. Solution: Automate where possible.

2.2 Reinforcement Learning

2.2.1 Fundamentals: Markov Decision Process

Many animals are able to learn complex behaviors by performing an action, observing the results and - if the action got them closer to their goal - repeating the action when facing a similar situation. For example, consider a food-dispensing lever in a hamster's cage. At first, the hamster might not notice the lever. But sooner or later, by accident or out of curiosity, the hamster will push the lever, dispensing an item of food. After a few repetitions, the hamster will know to walk towards the lever, and push the lever when it wants food. Our hamster uses a biological implementation of *Reinforcement Learning (RL)*. In an RL environment, an *agent* is able to repeatedly perform an *action*, observe the consequences and be rewarded or punished. RL therefore is an adaptive approach to solve feedback-based control problems.

Formally, RL environments are *Markov Decision Processes (MDP)*:

$$\text{MDP} = (S, A, p)$$

where S is the set of possible states the environment can be in, A is the set of actions available to the agent, and the transition distribution $p(s', r \mid s, a)$ specifies the environment dynamics, giving the joint probability of transitioning from state $s \in S$ to state $s' \in S$ after performing action $a \in A$, and getting reward $r \in \mathbb{R}$.

Interacting with the MDP, an agent first observes the environment's state s, then takes an action a, and then the next state s' of the environment is sampled from p. There are no further restrictions placed on the structure of state and action spaces. They can be merely labelled or ordered, finite or infinite, single- or multidimensional. Usually the agent observes a collection of variables and chooses an action along one or several dimensions.

Crucially, the states have the *Markov Property*: The probability of transitioning to state s' only depends on the current state s and action a. The history of past states does not matter, and there are no hidden facts that could change the probabilities. In practice, this is a simplifying modelling assumption. Returning to our hamster: In reality, the food-dispensing mechanism runs out of food after being activated a number of times. However, modelling the situation as an MDP, we do not keep track of history, so we can't tell in advance whether pushing the lever will actually produce food. Instead, we assume there's a certain probability of the action being unsuccessful, that is $p(s', \text{food} \mid s, \text{lever pushed}) < 1$. One generalization of the MDP that does enable us to model such hidden variables is the Partially Observable MDP, which I will cover in more detail later on .

cite section

2.2.2 Reinforcement Learning

introduce RL terms and algorithms

Let's now turn to the question how to solve a Markov Decision Process. Solving an MDP usually means finding the sequence of actions that maximize the expected total reward over all future time steps.

todo:

- episode - policy - value function - exploration vs exploitation - Bellmann Update

Baseline algorithm: - One RL algorithm, which I use as a baseline, is ??? - Explain algorithm

which one? depends on my treatment of uncertainty in the other one, so on the approach

2.3 Uncertainty in Reinforcement Learning

Introduce Uncertainty terms and formalisms from different perspectives. Then apply to RL.

There is a rich body of work on uncertainty. Mathematical and statistical notions of uncertainty, perspectives from economics for decision making under uncertainty.

Notes:

- Motivation: Most information is uncertain to some extent. Making good decisions under uncertainty requires an awareness of the uncertainty. - Uncertainty vs Risk: Uncertainty is a measure of the information content of a random variable or an observation???, Risk is the cost associated with different situations. - formal framework (maybe borrow from Econ: When to buy or sell a given asset?) - Decision making under uncertainty - which objective (Expected value vs risk metrics) - different types of uncertainty (e.g. aleatoric vs epistemic) - there are different types of uncertainty: Some uncertainty can be reduced by learning more about the problem, other uncertainty can not. - this stems from the formulation of RL as a stochastic MDP - for example, a biased coin. You will be able to learn something about it, but not actually predict the outcome ???

Uncertainty in RL: (maybe this should already be in approach?) - How is Uncertainty commonly modelled? - epistemic uncertainty in the observations: not explicitly modelled, somewhat represented in state-value function - stochasticity in the environment dynamics (+consequences of actions): accepted in the MDP. learned as transition probabilities in model-based RL, subsumed in e.g. Q-function in model-free RL. - epistemic uncertainty in the environment dynamics: modelled as transition probabilities - stochasticity in the reward function: not usually explicitly modelled - epistemic uncertainty in the reward function: modelled implicitly in the state-value function - uncertainty about causality? - probably not really relevant? should I discuss it somewhere else? - Adaptations for explicit treatment of uncertainty: - Formalism for non-perfect observations: POMDP (usually there are hidden variables) -¿ Usually solved by estimating an MDP, solving that. - POMDP does not assume the Markov property on observations, but does assume a hidden MDP - ??? - RL from human preferences? (for learning a reward function) - Benefits of explicit treatment of uncertainty/Motivation: - Risk-aware strategies - better performance (maybe? TODO: Test this) - more rubustness (possibly? TODO: support this or not) - better interpretability (possibly?) - TODO: other - Drawbacks: - more complex models require more training data - less efficient algorithms - not as well understood theoretically - might perform worse than just learning everything implicitly!

Risk-aware strategies: - can either specify a risk tolerance at time of inference or during training - during training: change reward function - at time of inference: requires model of the environment (I think) or a Q-function - more robust: can hand over control to e.g. humans when uncertain

Uncertainty in Multi-Agent Learning: (maybe exclude this completely) - Multi-Agent Environments are characterized by simultaneous actions by multiple agents, who each learn and act according to their own rewards. - More realistic and resilient than centralized control - absent trust, might be stuck in a suboptimal equilibrium

Chapter 3

Approach

Roughly 1/3 of thesis

Done: - restate the research question, and how I set out to answer it. Plan for this chapter: - introduce CityLearn and motivate the choice - mention citylearn challenge. - Introduce and explain the UADQN algorithm and motivate the choice (other algorithms have been tried) - Explain Experiment Setup, High-level overview to implementation details each. 1. Hand-Engineered Benchmark agent and Discretization Experiment 2. Hyperparameter Tuning Setup: - High-Level Overview of Experiments - Design decisions of Tuning Experiment - Implementation details (all the way down to hardware) 3. Test performance of Tuned Agents - High-Level Overview of Experiment, motivated by story - Investigate performance of Tuned Agents - High-Level Overview of Investigation

In order to efficiently control electricity demand, one needs to act with foresight on uncertain information. I aim to find out whether existing approaches can benefit from explicitly modelling various uncertainties. In order to do so, I introduce CityLearn, a sample control environment and model its uncertainties. I model CityLearn from different perspectives. Drawing from the theoretical background, I then construct an uncertainty-aware RL model to solve CityLearn. I compare this uncertainty-aware model empirically against existing baselines. An uncertainty-aware model can have additional benefits, like risk-awareness. I explore these.

- Divide it into subquestions, and treat each subquestion separately.

3.1 RL for Demand Response

- Let's focus in on Automated Demand Response in buildings: - Mention incentive-based behavior-change models, but ML models can do better! - Define the setting: - Focus on buildings, they need lots of energy and the physical processes are comparably simple, standardized and predictable. - Ignore industrial DR programs. - decentralized (why?) - What is the state of the art for Automated Demand Response? - How else can we implement Automated DR?

- Introduce CityLearn and alternatives

3.1.1 CityLearn: Applying RL for Demand Response

This section of the approach/background introduces the CityLearn Framework, so I can later mention it. It motivates the choice while highlighting alternative options. It introduces the data, modelling limitations and setting, but not the exact experimental setup.

CityLearn as a model for building-based demand response: - Model description (data, available observations and actions, rationale) - Success measures, cost function design - CityLearn challenge - limitations: modelling errors/simplifications and how much they matter - CityLearn was built to assess the general usefulness of RL for DR, not for my specific question. - Battery efficiency - Weather Forecasts are perfect oracles - Only available action is battery charging/discharging. - Other processes are assumed to be fixed (at least in 2022 challenge). Automatic Washing Machine starting. - Human behavior can not be influenced (models humans as inflexible) - Hourly control instead of continuous -¿ Makes perfect control more difficult - Energy can not be sold to the grid, not even to the microgrid. -¿ unrealistic - 2022 challenge: initially did not encourage cooperation. - Strengths: For what tasks is this environment adequate? - Makes it possible to apply RL to DR without expensive computational overload* - Test Multi-Agent behavior and cooperation in the DR context, as opposed to only single-building frameworks - Enables comparison of different approaches as a benchmark - Formal treatment of the CityLearn environment - Why? What kind of formal treatment?

Alternatives to CityLearn

3.2 Theoretical approach

3.2.1 Setting the scene

- start with a theoretical approach: - formal model of the environment,

In order to develop our uncertainty-aware approach to demand side management, we need to have a good understanding of the information environment. Let's formalize all relevant goals, available information and uncertainties.

The idea of Demand Side Management was introduced to enable a more efficient usage of the grid. Depending on who implements Demand Side Management, different incentives and information infrastructure might evolve. In our model of demand response, consumers themselves are empowered to make decisions about their electricity demand.

- !!! where does uncertainty come from in CityLearn !!!,

- forecasting problems: - uncertainty about future occupant behavior (electricity demand) - uncertainty about future solar power production (weather forecasts) - uncertainty about future costs (price forecasts) - measurement uncertainty: - observations might be imprecise - coordination uncertainty: - uncertainty about other actors' strategy and current actions - reward uncertainty: - uncertainty about the consequences of actions - the observed reward might be imprecise - the observed reward might not be the actual desired reward ???

Which uncertainties need to be specially treated?

What advantages would an uncertainty-aware strategy have?

- better risk management (possibly) - better performance (possibly) - better interpretability and robustness (possibly)

3.2.2 Developing the approach

- developing the algorithm (this is important. Maybe it includes some actual theoretical work and insight!),

- build off approaches that are cited in background.

- explain and motivate decisions.

3.3 Practical Approach

3.3.1 Implementation Details

- high-level overview of the practical approach and setup

- describe hand-engineered strategy

- Train one single agent that works for all buildings independently

Adaptations from the Risk and Uncertainty Paper algorithm: - make it work with continuous action space (discretize) - make it work for the single-agent case.

- describe implementation details - (of my algorithm and of CityLearn) - I found/fixed CityLearn Bugs - I engineered the Reward Function

3.3.2 Experimental Setup

- Design/select an algorithm that is fit for the task. Run the experiment and investigate results (in separate chapter)

- Data? - CityLearn.

- I want to test the performance of different strategies: - doing nothing as baseline - basic rule-based controller as improved baseline - various risk-aware strategies

Chapter 4

Results

In this chapter which also could be more than one chapter, depending on the nature of the thesis, the results of the thesis are presented. Make sure you illustrate your results with appropriate figures and tables, but do not discuss the results here. This should be done in a separate discussion chapter. Or maybe do combine results and discussion and split by research questions.

4.1 Exploratory Analysis of the Environment

4.1.1 Patterns in the Training Data

Describe the Training Data and patterns observed

4.1.2 Discretization

Describe the Discretization considerations

4.2 Hyperparameter Tuning

Describe the results of hyperparameter tuning. Which hyperparameters are important for which algorithm? How much computational power did tuning take? How robust were the algorithms w.r.t. hyperparameter choice?

4.3 Comparison of different algorithms

The main event. Results are different.

- Results of comparison / experiment

Chapter 5

Discussion and Outlook

Of course very important! You need to discuss the informatics as well as econ part of your thesis topic.

Take your time for writing the discussion, besides the introduction chapter it is the most important chapter of your thesis. Also do not subsection the discussion too heavily.
At least 5 pages,
Outlook can become an extra chapter.

Chapter 6

Conclusion and Outlook

- 1 page - summarize again what your paper did, but now emphasize more the results, and comparisons - write conclusions that can be drawn from the results found and the discussion presented in the paper - future work (be very brief, explain what, but not much how)

Appendix A

Further Tables and Figures

Viele Arbeiten haben einen Appendix. Besondere Sorgfalt muss beim Nummerieren der Tabellen und Abbildungen gewährleistet sein.

Nummer	Datum
1	1.1.80
2	1.1.90

Table A.1: Erste Appendix-Tabelle

Nummer	Datum
1	1.1.80
2	1.1.90

Table A.2: Zweite Appendix-Tabelle

Bibliography

Agent-Based Models of the Economy. (document)

Vázquez Canteli and José Ramón. *Multi-Agent Reinforcement Learning for Demand Response and Load Shaping of Grid-Interactive Connected Buildings*. Thesis, September 2020. (document)

G. Saake, I. Schmitt, and C. Türker. *Objektdatenbanken — Konzepte, Sprachen, Architekturen*. International Thomson Publishing, Bonn, 1997. ISBN 3-8266-0258-7. (document)

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Masterarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Ort, Datum　　　　　　　　　　　　　　　　　　　　Unterschrift