Reinforcement Learning for Power Trade

Richard W. Lincoln
Department of Electronic and Electrical Engineering
University of Strathclyde

A thesis submitted for the degree of
Doctor of Philosophy

2009

Acknowledgements

I would like to acknowledge ...

Abstract

Abstract goes here...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis describes the application of value-function and policy gradient reinforcement learning algorithms to the electric energy trade problem. This chapter introduces the problem and explains the motivation for the research. The goals of the research are stated along with the principle contributions. Finally, a reading guide is provided that provides an overview of the remaining chapters.

1.1 Motivation/Setting the scene

Free market democracy has underpinned the transformation of large western economies in the post war era and continues to be relied upon. Towards the end of the nineteenth century the principals of competitive trade were successfully applied to electric power industries, beginning in the UK in March 1990 with the creation of The Electricity Pool.

1.2 Problem statement/Aims & Objectives/Problem Description

Engineers must strive for complexity in their work. Rarely will a simple solution will perform a function to a higher degree than a more complex one. Certainly, where a function is either performed or not performed, prefer the simpler one, but most often problems can be solved to varying degrees.

The broad aim of the research presented in this thesis is to prove that the above conjecture applies to reinforcement learning algorithms for power trade. Previous research in this field (See Chapter 3 below) has used very simple algorithms in relation to those from the latest advances in artificial intelligence (See Sections 2.3.4 and 2.3.3 below). The goal is to prove that policy gradient methods, using artificial neural networks for policy function approximation, are better suited to learning the complex dynamics of a power system.

1.3 Research contributions

The research presented in this thesis pertains to the academic fields of power engineering, artificial intelligence and economics. The principle contributions in these areas are

- The proof that policy gradient reinforcement learning algorithms outperform value-function algorithms when applied to the power trade problem,
- A novel coupling of power system models and optimal power flow algorithm results with agents capable of handling discrete and continuous sensor and action spaces,
- Implementations of Roth-Erev reinforcement learning algorithms and continuous versions of Q-learning and Q(λ) for the open source PyBrain library,
- Open source implementations of power flow and optimal power flow algorithms in the Python programming language, preserving sparsity throughout the optimisation using the open source CVXOPT library.

1.4 Thesis structure/Overview/Reading guide

This thesis is focussed on the application of standard and advanced reinforcement learning algorithms to a particular problem domain. The reader will require a certain degree of prior knowledge, or must be willing to read much of the referenced material, to fully understand the methodology taken. The intended audience is engineering and economics researchers interested in the application of reinforcement learning algorithms to the problem of trading energy in electric power systems.

Chapter 2

Background

2.1 Power Flow

2.2 Optimal Power Flow

$$\begin{array}{ll} \text{minimize} & \sum_i c_i(P_i) \\ \text{subject to} & B_{bus}\theta = P_g - P_d - P_{bus,shift} - G_{sh} \end{array}$$

2.3 Reinforcement Learning

This section provides an introduction to the reinforcement learning problem and some of the associated terminology. Definitions for the value-function and policy gradient algorithms, that are later applied to power trade implementations of the problem, are given.

For a comprehensive introduction to reinforcement learning with evaluations of algorithm designs through mathematical analysis and computational experiments the intersted reader is directed to the seminal work by Barto and Sutton (?, ?).

2.3.1 Introduction

The problem of learning how best to interact with an environment so as to maximise some long-term reward is one that arises in many aspect of life. Reinforcement learning is a term that is typically applied to understanding, automating and solving this problem through computational approaches. Unlike with the majority of Machine Learning techinques, the algorithms are not instructed as to which actions to take, but must learn to maximise the long-term reward through trial-and-error.

Reinforcement learning starts with an interactive, goal-seeking individual and an associated environment. The individuals require the ability to sense aspects of their environment, perform actions that influence the state of their environment and be assigned rewards as a response to their chosen action. An agent is said to follow a particular *policy* when mapping the perceived state of its environment to an action choice.

Value-based methods attempt to find the optimal policy by approximating a *value-function* which returns the total reward an agent can expect to accumulate, given an initial state and following the current policy thereafter.

Policy-gradient methods are an alternative to this which represent a policy using a learned function approximator with its own parameters (?, ?). The function approximator is updated according to the gradient of expected reward with respect to these parameters.

2.3.2 Sarsa

Sarsa is an on-policy Temporal Difference control method. The policy is represented by a $M \times N$ table, where M and N are arbitrary positive numbers equal to the total number of feasible states and actions. The action-value update for agent j is defined by

$$Q_j(s_{jt}, a_{jt}) + \alpha[r_{jt+1} + \gamma Q_j(s_{jt+1}, a_{jt+1}) - Q_j(s_{jt}, a_{jt})]. \tag{2.1}$$

While the Q-learning algorithm updates action-values using a greedy policy, which is different to that being followed, Sarsa uses the discounted future reward of the next state-action observation following the original policy.

2.3.3 REINFORCE

2.3.4 ENAC

Chapter 3

Related Work

Game theoretic models are commonly associated with economics and attempt to capture behaviour in strategic situations mathematically. They have been applied to electric energy problems of many forms, including but not limited to analysis of market structure, market liquidity, pricing methodologies, regulatory structure, plant positioning and network congestion. More recently, agent-based simulation has received a certain degree of attention from researchers and has been applied in some of these fields also.

While popular and seemingly promising, agent-based simulation is still centred around abstracted models. The assumptions made is this abstraction must be subjected to the same verification and validation as with equation-based models. Verification of assumptions and model validation are often overlooked in agent-based simulations of energy markets, yet they are possibly the most important steps in the model building process. Techniques used to develop, debug and maintain large computer programs can often be used to verify that a model does what it is intended to do.

Validation of an energy market model is more difficult. It can be accomplished using the intuition of experts or through comparison of simulation results with either historical market data or theoretical results from more abstract representations of the model. Finding verifyable trends in existing markets is a very large challenge. To then prove that a computational model replicates these characteristics with suitable fidelity is yet more challenging still. Only when a model is suitably verified and validated can any conclusions be drawn from results obtained through implementation and simulation of suitable scenarios.

Chapter 4

Methodology

Societies reliance on secure energy supplies and the high volumes of electricity typically consumed render it impractical to experiment with radically new approaches to energy trade on real systems. This section explains the approach taken modelling real systems in software such that they may be simulated computationally. The method by which the physical power systems, that deliver electricity to consumers, were modeled is given, as well as for the mechanisms that facilitate trade and participants that utilise these mechanisms.

4.1 Electricity network model

High voltage transmission and distribution networks are the mechanisms by which traded electric energy is delivered to consumers. Limits to line/cable power flows, outages and reactive power availability can impose constraints on particular trades. As such, certain technical characteristics of the networks are fundamental to energy market operation and must be duly modeled.

4.1.1 Power Flow

The problem to be solved is finding the steady-state operating point of the network when given levels of generation and load are present. The primary constraints in a power system are the branch flow limits and the voltage limits at each bus. The system must be operated such that these constraints are not violated.

4.1.2 Common Information Model

Many tools exist for steady-state analysis of balanced three-phase AC networks and most are centred around bespoke models that describe the power system data. Several attempts have been made in the past to standardise the format in which power system data is stored [CDF, UKGDS, ODF] and latest and most popular is the Common Information Model.

The Common Information Model (CIM) is an abstract ontological model that describes the elements of national electric power systems and the associations between them. CIM is an evolving international standard approved by the International Electrotechnical Commission (IEC).

Unlike many tool specific models the CIM does not simplify the power system into a graph of buses connected by branches. Instead it describes each of the components in the system and the electrical connectivity between them. Conventional numerical techniques for steady-state analysis of AC power systems require a simplified bus-branch model such that when the voltage angle and magnitude at each bus is determined the power flows on each branch may be calculated.

4.1.3 Energy market model

Mechanisms for facilitating competitive trade between electricity producers and consumers differ greatly in the specifics of their implementations in coutries throughout the world. However, fundamentally they either provide a centralised pool through which all electricity is bought and sold or they permit producers and suppliers to trade directly.

The UK transmission network is frequently congested[]. The thermal limits of transmission lines between particular areas are often reached. The balancing mechanism is the financial instrument used by the system operator to resolve constraint issues and energy imbalances. Should the market not be suitably effective in this function the system operator may choose to contract outwith the balancing mechanism. By way of incentive to match demand and avoid congestion, imbalance charges are imposed on responsible participants. There is some evidence to suggest that centralised resolution by a system operator and socialisation of the incurred costs leads to inefficient despatch of generators[Neuhoff].

There are a number of alternative approaches to congestion resolution(?, ?).

4.1.4 Transmission capacity rights

One approach is to issue contracts for transmission capacity rights or equivalent financial rights. The maximum available transmission capacity being auctioned for certain periods of time and firm contracts made entitling owners to full compensation upon curtailment or withdrawl(?, ?).

The states of Pensylvania, New Jersey and Maryland (PJM) operate a non-compulsory power pool with nodal market-clearing prices based on competitive bids. This is complemented by daily and monthly capacity markets plus the monthly auction of Financial Transmission Rights to provide a hedging mechanism against future congestion charges.

4.1.5 Transmission charging

Impose delivery charges which increase as network constraints are approached.

4.1.6 Extended bids/offers

Request extended bids and offers which include costs associated with the adjustment of participant's desired position.

4.2 Market participant model

Without competition between market participants there is no driver for individuals to improve efficiency and reduce costs paid by the consumers. Traders are typically resposible for this, but it is not feasible to use humans for this project. In a highly distributed power system, a very large number of items of plant may be supplying the demand and, depending on the levels of aggregation, this could require many traders to be used. Also, this project requires that experiments be repeated numerous times under a variety of scenarios.

4.2.1 Software agents

Participants are modeled in software also. The nature of a highly distributed power system dictates that a very large number of entities may be interacting in the marketplace. Economic studies regularly integrate participant logic into the same optimisation problem as the market. However, this does not scale to large numbers of individual participants. Separating participant logic into individual software agents allows their action selection procedures to be processed in simultaneously. The definition of an agent in this context emerges from the machine learning technique employed to implement the competitive decision making process.

4.2.2 Reinforcement learning

While there is a wealth of data available on past energy market activity involving conventional transmission connected plant, there exists no such resource for trade performed in highly distributed power systems. Consequently, reactive machine learning techniques that use new data to influence the decision making policy are used.

Reinforcement learning is a sub-area of machine learning and can be applied to a wide variety of problems(?, ?). To allow the same learning algorithms developed for traditional, academic reinforcement learning problems (chess, backgammon, lift scheduling etc.) to be applied to models of energy markets (and vice versa) a modular machine learning library is used.

Chapter 5

Results

Chapter 6

Discussion

Chapter 7

Critical analysis

Chapter 8

Future work

Chapter 9

Summary conclusions

References

Peters, J., Schaal, S. (2008). Natural actor-critic. *Neurocomputing*, *71*(7-9), 1180–1190.

Tesauro, G. (1994). Td-gammon, a self-teaching backgammon program, achieves master-level play. *Neural Computation*, *6*(2), 215-219.