University of Strathclyde

Department of Electronic and Electrical Engineering

Learning to Trade Power

by

Richard W. Lincoln

A thesis presented in fulfilment of the
requirements for the degree of

Doctor of Philosophy

2010

Signed: Date: November 15, 2010

Acknowledgements

I wish to thank Professor Jim McDonald for giving me the opportunity to study at The Institute for Energy and Environment and for permitting me the freedom to pursue my own research interests. I also wish to thank my supervisors, Professor Graeme Burt and Dr Stuart Galloway, for their guidance and scholarship. Most of all, I wish to thank my parents, my big brother and my little sister for all of their support throughout my PhD.

This thesis leverages several open source software projects developed by researchers from other institutions. I wish to thank the researchers from Cornell University, especially Dr Ray Zimmerman, for their work on optimal power flow, the researchers from the Dalle Molle Institute for Artificial Intelligence (IDSIA) and the Technical University of Munich for their work on reinforcement learning algorithm and artificial neural network implementations and Charles Gieseler from Iowa State University for his implementation of the Roth-Erev method.

This research was funded by the United Kingdom Engineering and Physical Sciences Research Council through the Supergen Highly Distributed Power Systems consortium under grant GR/T28836/01.

Abstract

In electrical power engineering, learning algorithms can be used to model the strategies of electricity market participants. The objective of this thesis is to establish if *policy gradient* reinforcement learning algorithms can be used to create participant models superior to those involving previously applied *value function* based methods.

Supply of electricity involves technology, money, people, natural resources and the environment. All of these aspects are changing and electricity market designs must be suitably researched to ensure that they are fit for purpose. In this thesis electricity markets are modelled as non-linear constrained optimisation problems, which are solved using a primal-dual interior point method. Policy gradient reinforcement learning algorithms are used to adjust the parameters of multi-layer feed-forward artificial neural networks that approximate each market participant's policy for selecting power quantities and prices that are offered in the simulated marketplace.

Traditional reinforcement learning methods, that learn a value function, have been previously applied in simulated electricity trade, but they are mostly restricted to use with discrete representations of a market environment. Policy gradient methods have been proven to offer convergence guarantees in continuous environments and avoid many of the problems that mar value function based methods.

Five types of learning algorithm are compared in a series of Nash equilibrium and constraint exploitation simulations. Policy gradient methods are found to be a valid option for modelling the strategies of electricity market participants, but they are outperformed by a traditional action-value function algorithm in all of the tests. Further development of this research could provide opportunities for advanced learning algorithms to be used in descision support and automated energy trade applications.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis examines reinforcement learning algorithms in the domain of electric power trade. In this chapter the motivation for research into electricity trade is explained, the problem under consideration is defined and the principle research contributions are stated.

1.1 Research Motivation

Quality of life for a person is directly proportional to his or her electricity usage (Alam, Bala, Huo, & Matin, 1991). The world population is currently 6.7 billion and forecast to exceed 9 billion by 2050 (United Nations, 2003). Electricity production currently demands over one third of the annual primary energy extracted (The International Energy Agency, 2010) and as more people endeavour to improve their quality of life, finite fuel resources will become increasingly scarce. Market mechanisms, such as auctions, where the final allocation is based upon the claimants' willingness to pay for the goods, provide a device for efficient allocation of resources in short supply. Two decades ago the UK became the first large industrialised country to introduce competitive markets for electricity generation.

The inability to store electricity, once generated, in a commercially viable quantity prevents it from being traded as a conventional commodity. Trading mechanisms must allow shortfalls in electric energy to be purchased at short notice from quickly dispatchable generators. Designed correctly, a competitive electricity market can promote efficiency and drive down costs to the consumer, while design errors can quickly allow market power to be abused and market prices to be elevated. It is essential to research electricity market architectures and ensure that their unique designs are fit for purpose.

The value of electricity to society makes it impractical to experiment with radical changes to trading arrangements on real systems. The average total demand for electricity in the United Kingdom (UK) is approximately 45GW and the cost of buying 1MW for one hour is around £40 (Department of Energy and Climate Change, 2009). This equates to yearly transaction values of £16 billion. The value of electricity becomes particularly apparent when supply fails. The New York black-out in August 2003 involved a loss of 61.8GW of power supply to approximately 50 million consumers. The majority of supplies were restored within two days, but the event is estimated to have cost more than $6 billion (Minkel, 2008; ICF Consulting, 2003).

An alternative approach is to study abstract mathematical models of markets with sets of appropriate simplifying approximations and assumptions applied. Market characteristics can established by simulating the models using digital computer programs. Competition between participants is a fundamental feature of all markets, but the strategies of humans can be difficult to model mathematically. One option is to use reinforcement learning algorithms from the field of artificial intelligence. These methods can be used to represent adaptive behaviour in competing players and, when correctly configured, have been shown to be capable of learning highly complex strategies (Tesauro, 1994). This thesis makes advances in electricity market participant modelling through the application of a relatively new genre of reinforcement learning methods called policy gradient algorithms.

1.2 Problem Statement

Individuals participating in an electricity market (be they representing generating companies, load serving entities, firms of traders etc.) must utilise multi-dimensional data to their advantage. This data may be noisy, sparse, corrupt, have a degree of uncertainty (e.g. demand forecasts) or be hidden from the participant (e.g. competitor bids). Reinforcement learning algorithms must be capable of operating with this kind of data if they are to successfully model participant strategies.

Traditional reinforcement learning methods, such as Q-learning, attempt to find the *value* of each available action in a given state. When discrete state and action spaces are defined, these methods become restricted by Bellman's Curse of Dimensionality (Bellman, 1961) and can not be readily applied to complex problems. Function approximation techniques (e.g. artificial neural networks) can be used to apply these methods to continuous environment representations, but have been shown to cause convergence or even diverge problems when approximating a value function (Tsitsiklis & Roy, 1994; Peters & Schaal, 2008; Gordon, 1995; Baird, 1995).

Policy gradient reinforcement learning methods do not attempt to approximate a value function, but to approximate a *policy-function* that, given the current perceived state of the environment, returns an action. They do not suffer from many of the problems that mar value-function based methods in high-dimensional problems. They have strong convergence properties, do not require that all states be continuously visited and work with state and action spaces that are continuous, discrete or mixed. Policy performance may be degraded by uncertainty in state data, but the learning methods do not need to be altered. They have been successfully applied in many operational settings, including: robotic control (Peters & Schaal, 2006), financial trading (Moody & Saffell, 2001) and network routing (Peshkin & Savova, 2002) applications.

It is proposed in this thesis that agents which learn using policy gradient methods may outperform those using value function based methods in simulated competitive electricity trade. It is further proposed that policy gradient methods may operate better under dynamic electric power system conditions, achieving greater profit by exploiting constraints to their financial benefit. This thesis will use electricity market simulation techniques to compare value function based and policy gradient learning methods to explore these proposals.

1.3 Research Contributions

The research presented in this thesis pertains to the academic fields of Electric Power Engineering, Artificial Intelligence and Economics. The principle contributions made by this thesis are:

- The first application of policy gradient reinforcement learning methods in simulated electricity trade.
- The first application of a non-linear optimal power flow formulation in agent based electricity market simulation.
- A new Stateful Roth-Erev reinforcement learning method.
- Simulation results comparing the convergence to a Nash equilibrium of policy gradient and value function based reinforcement learning methods.

- Simulation results that examine the exploitation of electric power system constraints by policy gradient reinforcement learning methods.
- An implementation of a power exchange auctions market model and multi-learning-agent system for simulating electricity trade.
- The idea of applying Neuro-Fitted Q-Iteration and GQ(λ) in simulations of competitive energy trade.
- A model of the UK transmission system derived from data in the National Grid Seven Year Statement.

The publications that have resulted from this thesis are: Lincoln, Galloway, and Burt (2009, 2007); Lincoln, Galloway, Burt, and McDonald (2006).

1.4 Thesis Outline

The presentation of this thesis is organised into nine chapters. Chapter **??** provides background information on electricity supply, wholesale electricity markets and reinforcement learning. It describes how optimal power flow formulations can be used to model electricity markets and defines the reinforcement learning algorithms that are later compared.

In Chapter **??** the research in this thesis is described in the context of previous work that is related in terms of application field and methodology. Publications on agent based electricity market simulation are reviewed with emphasis on the reinforcement learning methods used. Previous applications of policy gradient learning methods in other types of market setting are reviewed also.

Chapter **??** describes the power exchange auction market model and the multi-agent system used to simulate electricity trade. It defines the association of learning agents with portfolios of generators, the process of offer submission and the reward process.

Simulations that examine the convergence to a Nash equilibrium of systems of multiple electric power trading agents is reported in Chapter **??**. A six bus test case is used and results for four learning algorithms under two cost configurations are presented and analysed.

Chapter **??** examines the ability of agents to learn policies for exploiting constraints in simulated power systems. The 24 bus model from the IEEE Reliability Test System provides a complex environment with dynamic loading conditions.

The primary conclusions drawn from the results in this thesis are summarised in Chapter **??**. Shortcomings of the approach are noted and the broader implications are addressed. Some ideas for further work are also outlined, including alternative reinforcement learning methods and uses for a model of the UK transmission system.

Bibliography

Alam, M. S., Bala, B. K., Huo, A. M. Z., & Matin, M. A. (1991). A model for the quality of life as a function of electrical energy consumption. Energy, 16(4), 739-745.

Baird, L. (1995). Residual algorithms: Reinforcement learning with function approximation. In Proceedings of the Twelfth International Conference on Machine Learning (p. 30-37). Morgan Kaufmann.

Bellman, R. E. (1961). Adaptive control processes – A guided tour. Princeton, New Jersey, U.S.A.: Princeton University Press.

Department of Energy and Climate Change. (2009). Digest of United Kingdom Energy Statistics 2009. In (chap. 5). National Statistics – Crown.

Gordon, G. (1995). Stable function approximation in dynamic programming. In Proceedings of the Twelfth International Conference on Machine Learning (p. 261-268). Morgan Kaufmann.

ICF Consulting. (2003, August). The economic cost of the blackout: An issue paper on the northeastern blackout. (Unpublished)

Lincoln, R., Galloway, S., & Burt, G. (2007, May 23-25). Unit commitment and system stability under increased penetration of distributed generation. In Proceedings of the 4th International Conference on the European Energy Market, 2007. EEM 2007. Cracow, Poland.

Lincoln, R., Galloway, S., & Burt, G. (2009, May). Open source, agent-based energy market simulation with Python. In Proceedings of the 6th International Conference on the European Energy Market, 2009. EEM 2009. (p. 1-5).

Lincoln, R., Galloway, S., Burt, G., & McDonald, J. (2006, 6-8). Agent-based simulation of short-term energy markets for highly distributed power systems. In Proceedings of the 41st International Universities Power Engineering Conference, 2006. UPEC '06. (Vol. 1, p. 198-202).

Minkel, J. R. (2008, August 13). The 2003 northeast blackout–five years later. Scientific American.

Moody, J., & Saffell, M. (2001, July). Learning to trade via direct reinforcement.

IEEE Transactions on Neural Networks, 12(4), 875-889.

Peshkin, L., & Savova, V. (2002). Reinforcement learning for adaptive routing. In Neural Networks, 2002. IJCNN 2002. Proceedings of the 2002 International Joint Conference on (Vol. 2, p. 1825-1830).

Peters, J., & Schaal, S. (2006, October). Policy gradient methods for robotics. In Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on (p. 2219-2225).

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.

The International Energy Agency. (2010, Septemeber). Key world energy statistics 2010. Paris.

Tsitsiklis, J. N., & Roy, B. V. (1994). Feature-based methods for large scale dynamic programming. In Machine learning (p. 59-94).

United Nations. (2003, December 9). World population in 2300. In Proceedings of the United Nations, Expert Meeting on World Population in 2300.