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

Industrialised societies have become increasingly reliant on the supply of electric energy since the connection of large power stations began in 1938. The extent to which this is true can be seen in the financial impact that loss of supply has on society.

In June 2004 the United Kingdom (UK) became a net importer of natural gas for the first time in 8 years[EIA DOE]. Since energy industry privatisation by the Conservatives in the 1980s, use of domestic gas reserves for electricity generation has been encouraged and exploited.

As UK natural gas production has now peaked and consumption continues to grow, concern over reliance on imports from less stable regions has increased.

The UK is hugely reliant on fossil fuels. More than three quarters of UK electricity is generated from a relatively small number of large coal and gas fired power stations[Energy Digest]. Of the 298 stations with capacity over 1MW in the UK, 63 gas fuelled (including CCGT) and 13 coal fired stations supply over 290 TWh of the UK's 393TWh annual electric energy production[DUKES]. Much of the remainder is generated through nuclear fission.

Concerns over climate change and security of supply have caused the UK government to pursue self-sustainable sources of electric energy. This is illustrated by the government?s recent decision to make a legally binding commitment to an 80% reduction in carbon dioxide emissions by 2050, relative to 1990 levels[].

Relative to most other commodities, trade of electric energy is still in its infancy. Liberalisation and unbundling of electricity supply industries costs many millions of pounds to implement[]. Countries, having made this investment, continue to restructure and adjust their energy markets in the hope of further reducing costs to the consumer and promoting innovation and efficiency through competition.

1.1 Conventional national power systems

Economies of scale prompted construction of the first large-scale power stations in the UK at the beginning of the twentieth century. Following the introduction of the Electricity (Supply) Act 1926 the largest and most efficient of these were connected by a series of regional high-voltage three-phase AC grids synchronised at 50 Hertz. Integration was completed and a national transmission system made operational in the UK for the first time in 1938. This approach to electricity supply is largely the same as that still employed throughout the UK to the present day. Alternating current, mainly from rotating synchronous machines, is transformed to high voltages for bulk transmission over long distances with high efficiency. Power from the transmission system is fed through distribution networks in a uni-directional fashion at lower voltages before final usage.

While maintenance and extension of the transmission system and of distribution networks is an everyday activity for energy utilities, many power system components have extremely long operational lifetimes[]. This and the magnitude of the capital investment made post-war in construction of the electricity networks suggests that there is likely to be little change in the topology of the system of wires in the foreseeable future. This is further compounded by the fact that distribution networks are often radial in their structure. Reliability and protection are major issues in the operation of power systems and the task of detecting and isolating a fault is often more difficult in systems that are meshed.

Large-scale thermal power stations operate steam and gas turbines around 13 metres in length, approximately 400 tons in weight and rotate at up to 3600 revolutions per minute[SIEMENS]. The kinetic energy stored in these turbines and the connected synchronous machines is vast and the associated flywheel effect plays an important role in smoothing short-term imbalances in supply and demand[].

Synchronous machines used in large thermal power stations invariably use a rotor winding that is excited and a magnetic flux created by a DC current. Controlling the magnitude of the rotor field current allows the reactive power provided or absorbed by the machine to be adjusted. Reactive power is essential in maintaining system voltage within permissable limits[].

1.2 Highly distributed national power systems

The highly distributed power system is a conceptual future for the UK's national energy network. It is capable of meeting current targets for reduced greenhouse gas emissions[] and reduces reliance on foreign fuel imports. It is a system in which all but the least polluting fossil fuelled power stations have been decommissioned. Their output supplanted by distributed generation, supported with demand-side management measures and advances in energy storage technology.

Distributed (or embedded) generation is most easily defined as electricity production plant connected to the national electricity network at the distribution level[DG Book]. The transmission system being defined as that which operates at of above 275kV in England and Wales or at or above 132kV in Scotland. This encompasses most small-scale plant, but there exist exceptions such as large-scale wind farms.

Hydro-electric dams, biomass fuelled power stations and wind farms are the three sources of renewable energy currently available in the UK with capacities equivalent to that of fossil fuelled plant. Hydro-electric power stations are often large-scale and transmission connected for bulk transport of power to load centres. The low energy density of biomass fuels, relative to that of fossil fuels, often dictates that related generating plant is located close the fuel source origin and connected to a distribution network. The time varying nature of wind energy necessitates the use of induction machines for generation. These are typically sinks of reactive power and require support for operation and maintenance of voltage.

1.2.1 Increased source granularity

In terms of demand the UK power system is already highly distributed. The number of generators, controllable loads and storage systems is expected to be much greater in a HDPS. However, there are at present already around 260 generators supplying energy to consumers and the mechanisms currently in place to facilitate the trade of their power may be adaptable.

Consumers and owners of small-scale generation will continue to desire protection from the risks associated with the wholesale marketplace. Today?s typical domestic and commercial supply contracts between consumers and energy retailers offer such protection. As the penetration of DER grows the role of energy retailers offering aggregation services will likely grow as more plant owners come to desire representation in the marketplace. While there are a great many ways in which this aggregation may be arranged it remains essentially a contractual arrangement. A much larger challenge lies in the operation and control of a system in which the state of plant is principally determined by a decentralised free market. The network may be divided into ?cells? for the provision of a single point of control, but each cell may contain individual items of plant being aggregated by different companies. Controlling the DER, so as to adhere to system constraints, must be done in the most economically efficient, in the least environmentally damaging manner and according to contracted access arrangements. Also, the details of any control measures undertaken must be fed back to the marketplace such that they may be taken into consideration during the settlement process. This interface between the technical domain and the commercial mechanisms remains an open research topic.

1.2.2 Inversion of control

Consumers have grown accustomed to using power from the grid at will. System loads are by and large passive and, with the exception of dual rate white metered loads, only the largest make adjustments to their consumption according to price signals or system conditions. It is likely that the fluctuating nature of the power output from generators exploiting renewable energy sources will necessitate an increase in the adoption of active loads. As control shifts from being almost purely supply-side, an opportunity opens for the energy marketplace to offer appropriate consumer choice, not only in terms of cost, but Quality of Service (QoS) also. Advances in smart metering promise to support such a migration.

1.2.3 Reduced network reticulation

The connection of controllable energy resources to lower voltage and less reticulated areas of the network may also offer new possibilities for structuring the relationship between the market and the system for management of network constraints. There are generally three options for power system constraint management.

- Only permit the formation of energy trades if delivery is physically feasible.
- Impose delivery charges which increase as network constraints are approached.
- Request extended bids and offers which include costs associated with the adjustment of participant?s desired position.

The third option most closely describes the method currently used in the United Kingdom. In a highly reticulated network it is difficult to determine the direction of each participant?s energy flows. In turn, this poses difficulties in determining if a particular delivery is feasible and which participants are responsible for congesting the network. Distribution networks are typically less reticulated than the transmission network to which the majority of generation is connected at present. Therefore, there may be opportunities to utilise options one or two in an energy marketplace for an HDPS.

Furthermore, the lower voltage of distribution networks may open up opportunities for more widespread use of power electronics technologies such as FACTS and phase shifting devices. These would provide a limited ability to direct the flow of electrical energy. How this might be managed on a wide scale and how efficient interaction with the marketplace might be achieved is open to investigation

1.2.4 Dual objective optimisation

Along with concern over the UK's dependence on natural gas imports, concern over the environmental impact of electricity generation is a primary motivator for a move to HDPS. The energy market is expected to simultaneously minimise costs to the consumer and encourage reduced emission of greenhouse gasses.

The European Union Greenhouse Gas Emission Trading Scheme (EU ETS) is an example of how a second marketplace running parallel to electricity trade, in this case trading allowances, may give weight to this new objective. There are other ways in which the greenhouse gas output of certain technologies may be taken into consideration and this remains an open and important research topic.

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.

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.