An Intelligent Broker Agent for Energy Trading: An MDP Approach

Rodrigue T. Kuate, Minghua He, Maria Chli and Hai H. Wang
School of Engineering and Applied Sciences, Aston University
Birmingham, United Kingdom
{tallakur, m.he1, m.chli, H.WANG10}@aston.ac.uk

Abstract

This paper details the development and evaluation of AstonTAC, an energy broker that successfully participated in the 2012 Power Trading Agent Competition (Power TAC). AstonTAC buys electrical energy from the wholesale market and sells it in the retail market. The main focus of the paper is on the broker's bidding strategy in the wholesale market. In particular, it employs Markov Decision Processes (MDP) to purchase energy at low prices in a day-ahead power wholesale market, and keeps energy supply and demand balanced. Moreover, we explain how the agent uses Non-Homogeneous Hidden Markov Model (NHHMM) to forecast energy demand and price. An evaluation and analysis of the 2012 Power TAC finals show that AstonTAC is the only agent that can buy energy at low price in the wholesale market and keep energy imbalance low.

1 Introduction

Due to the privatisation and decentralisation of the electricity provision system in many countries, electricity markets have undergone several restructuring processes in order to improve the market efficiency. Given the fact that the storage of electricity is very expensive, one of the key indicators of the electricity market efficiency is the imbalance between the energy demand and energy supply. As a market participant, an energy broker that buys energy from energy generators and sells to consumers in the retail market, plays a prominent role in ensuring the market efficiency.

To this end, the Power Trading Agent Competition (Power TAC) [Ketter *et al.*, 2012] is provided by the Trading Agent Competition (TAC) community [TAC Community, 2013], which is an international research forum that promotes the development of intelligent agents. The Power TAC simulates an open and competitive electricity market where broker agents compete against each other [Power TAC Community, 2013b]. In more detail, it models a wholesale market and a retail market. Simultaneously, while the wholesale market simulates energy markets such as the European or North American wholesale energy markets, the retail market simulates energy consumers. A distribution utility (DU) owns the distribution network and ensures real time energy balancing between supply and demand. The simulation environment provides two types of customers: (1) elemental customers, such as households, small businesses, small energy producers and electric vehicles; and (2) factored customers, such as greenhouse complexes and manufacturing facilities. A Power TAC game generally runs for about 60 virtual days or 1440 simulated hours. Every simulated hour lasts 5 seconds in real world. Starting the game without money in its bank account, the broker earns money by getting payment for the energy sold and loses money by paying for the energy required or market fees (e.g., the energy distribution fee, energy imbalance fee, tariff publication fee, tariff revocation fee and the bank interest on the debt). During the game, the bank always loans the broker money to purchase energy and charges interest. At the end of the simulation, the broker with the highest bank balance wins the game.

During each simulated hour (time slot), an energy broker can perform the following activities:

- place bids or/and asks on the wholesale market. The wholesale market operates as a periodic double auction market where it is cleared once every simulated hour. Broker agents can buy and sell energy for future delivery up to 24 hours ahead (thus called day-ahead).
- determine, publish and modify energy tariffs for the retail market. The broker can publish two types of tariff: production and consumption tariffs. The power TAC environment enables the design of tariffs with real world tariff features (e.g., periodic payments, tiered rates, sign-up bonuses, dynamic pricing). These features enable brokers to manage their customer portfolio.

Furthermore, the Power TAC server provides the following information to energy brokers during the game:

- Public information that is available to all the brokers. During each time slot, a broker receives market clearing prices and the cleared volume, current weather situation (temperature, cloud cover, wind speed and wind direction) and the weather forecast for the next 24 simulated hours. For every six time slots, all brokers receive information about all the newly published tariffs. Each Power TAC game is initialised with a setup game, which provides initial data for 360 time slots. The initial data includes the hourly power consumption and production

volume for each customer, the clearing price and cleared volume in the wholesale market, and the hourly weather report.

- Private information that is available only to the broker in question. This includes successful bids and asks in the wholesale market and the information about the tariff that are already published on behalf of the broker. Only the broker that owns a published tariff receives information about the tariff. This tariff information may be customer subscription, withdrawals or payments. General information such as cash position, total energy distribution or energy imbalance is also considered as private.

Against this background, we have developed and evaluated an intelligent broker agent called AstonTAC. The focus of the paper is on the Markov Decision Process (MDP) model that the agent uses for the energy bidding in the wholesale market. In order to apply the MDP model, AstonTAC forecasts the energy demand and the energy price using Non-Homogeneous Hidden Markov Models (NHHMM) [Bengio and Frasconi, 1995; Bishop, 2006]. During the Power TAC competition in December 2012, AstonTAC performed stably and successfully. Moreover, it was the only agent able to buy energy at a low price in the wholesale market and keep the energy imbalance low. The key contribution of our work is: (1) the energy bidding strategy for the energy broker and (2) the forecast models for energy demand and price.

The remainder of the paper is organised as follows. Section 2 presents related work on energy trading. Section 3 describes our agent. Section 4 evaluates the agent and the trading strategy. Finally, Section 5 concludes.

2 Related Work

A vast number of techniques have been proposed to deal with energy trading, energy demand forecast and price forecast.

In [Song *et al.*, 2000], an optimal bidding strategy for energy suppliers was presented. The MDP presented in [Song *et al.*, 2000] assumes to have complete information about the competing agents. However, the Power TAC environment considers any successful bids and asks as private information. Using the HMM to define the state of the environment, our MDP indirectly considers the behaviour of the other market participants. Most of the publications about the bidding strategy in an energy market consider only the supplier viewpoint [Tellidou and Bakirtzis, 2006; Bach *et al.*, 2012]. In contrast to the energy broker, the energy supplier is only concerned with selling all the available energy but not with the appropriate volume that would enable the supply and demand to be balanced.

The techniques used in energy demand and price forecasting can be classified in two trends: times series models and machine learning. The commonly used time series models include Autoregressive Integrated Moving Average (ARIMA) [Cancelo *et al.*, 2008; Contreras *et al.*, 2003; Conejo *et al.*, 2005], Generalized Autoregressive Conditional Heteroskedasticity (GARCH) [Garcia *et al.*, 2005; Zheng *et al.*, 2005], structural time series models [Harvey and Koopman, 1993] and multiple regression models [Ramanathan *et al.*, 1997]. The machine learning techniques include Artificial Neural Networks (ANN) [Mandal *et al.*, 2005; Zhang and Luh, 2005; Gao *et al.*, 2000; Yao *et al.*, 2000], Wavelet transform and Kalman filter [Nguyen and Nabney, 2010]. HMM provides us a more robust and faster way for generating prediction models at run time. Our HMMs automatically consider the intra-weekly and intra-daily behaviour of the energy demand and price. Moreover, by updating our HMMs transition matrices during the game, HMMs adapt to environment changes, which is a desirable feature of AstonTAC.

3 AstonTAC MDP

Given the complexity, and the dynamic and uncertain nature of the Power TAC environment, AstonTAC is built to adapt to environmental changes and to act autonomously during the simulation. For every time slot, AstonTAC employs MDP models to determine the bids (price, volume and time slot) submitted to the wholesale market. In order to calculate the volume to buy, AstonTAC needs to know the demand consumption of the contracted customers and the energy production of some of the contracted customers. In the Power TAC environment, some customers are able to produce electrical energy using renewable energy sources and to sell their energy production to the broker agent. Thus, the net customer demand consumption is the difference between the energy consumed and the energy produced. In order to set a bidding price to buy energy and the time slot for delivery, it is vital to predict the clearing price, so that the agent can buy the energy when the clearing price is low.

Figure 1 illustrates the architecture of AstonTAC MDP. The *Clearing Price Predictor* (see Section 3.4) predicts the clearing prices of the wholesale market. The *Customer Energy Production Predictor* (see Section 3.2) predicts the energy consumption of the contracted customers and the *Customer Energy Consumption Predictor* (see Section 3.1) predicts the energy production of the contracted customers. The predictors use the Power TAC Server information (both public and private information described in Section 1). The customer energy consumption and production predictors are used by the *Customer Energy Demand Manager* (see Section 3.3) to determine the volume of the net energy demand to buy each hour. Based on the predicted clearing price and the net energy demand, AstonTAC MDP decides the bids to place in the wholesale market. AstonTAC MDP aims to enable the agent to buy energy at a low price and keep the supply and

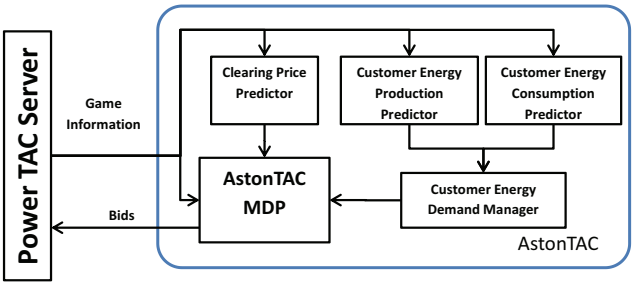

Figure 1: AstonTAC MDP Architecture

demand volume as equal as possible.

3.1 Customer Energy Consumption Predictor

AstonTAC uses Non-Homogeneous Hidden Markov Models (NHHMM) to predict the customer energy consumption. Our HMM is considered to be non-homogeneous, because the transition probability between two states is time dependent. The following explains the hidden states, the training and the update of the NHHMM model at run time.

Hidden States

The historical data from the setup game (see Section 1) are used to determine the hidden states of the customer energy consumption. The hidden states are computed at the beginning of the game for all the customers that are able to consume energy. The net energy usage of each customer represents the observed variables. We use Matlab's implementation of the k-means clustering algorithm to determine the hidden states from the observed variables.

Time slot	0	1	2	3	4	5	6
Net usage	7882.4	7445.3	7585.2	9091.6	8316.3	10088.5	10696.2
Hidden states	2	1	2	3	2	4	4
Cluster centroids	8000	7000	8000	9000	8000	10000	10000

Table 1: An Example of K-Means Clustering of Customer Consumption.

Table 1 shows an example of k-means (k=4) clustering with seven observed random variables (net usage). Matlab k-means provides the clusters 1, 2, 3 and 4 which are also the hidden states. Moreover, k-means provides the centroid of each cluster, which we use for the prediction.

After defining the hidden states, a graphical representation of the HMM can be drawn. Figure 2 shows an example of a graphical representation of the latent states and observed variables where $x_0, ..., x_4$ represent the net energy consumption of the customer. In order to predict the values of the observations $x_0, ..., x_4$, one needs to predict the state of the latent variables $z_0, ..., z_4$.

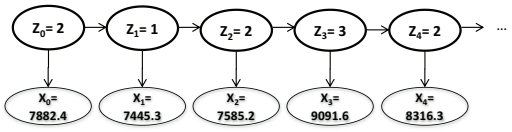

Figure 2: Graphical Representation of the HMM model for Energy Consumption. The nodes represent the random variables and the arrows the dependencies.

Training of the Consumption NHHMM

The training of an HMM model is essentially the calculation of the transition matrix and of the emission matrix. The transition matrix is the matrix of transition probabilities, which is denoted by: given i, j, z_{t-1} and z_t the hidden states; k and x_t the observed variables. $T_{ij} = p(z_t = j|z_{t-1} = i)$ where the time $t\epsilon\mathbb{N}$, $0 \leq T_{ij} \leq 1$ with $\sum_j T_{ij} = 1$. The emission matrix is the matrix of emission probabilities, which is denoted by: $E_{jk} = p(x_t = k|z_t = j)$ where the time $t\epsilon\mathbb{N}$, $0 \leq E_{jk} \leq 1$ with $\sum_k E_{jk} = 1$. Our HMM is considered to be non-homogeneous, because the transition probability between two states is time dependent. This means that it is possible to have two energy consumption states i and j so that: $p(z_t = j|z_{t-1} = i) \neq p(z_{t+2} = j|z_{t+1} = i)$.

Prediction and Update of the Consumption NHHMM

The aim of the prediction is to obtain an approximation of the consumption value of each customer. Instead of predicting the observed values $x_0, ..., x_4$, we only need to predict the hidden states $z_0, ..., z_4$. The predicted values of the observed variables are the corresponding k-mean centroids. The resulting transition matrix is a $24 \times 20 \times 20$ matrix where there is a 20×20 transition matrix for each hour of the day. Thus, at each hour of the day, considering the current state of the consumption, the predicted values for the future hours can be calculated using the resulting transition matrix.

After each prediction of the energy consumption, the agent receives the actual values of the customer consumption from the Power TAC server. AstonTAC records these values for each customer, computes the corresponding hidden states and updates the transition matrix with the new information. The HMM transition tables are updated hourly (each simulated hour). The update does not require running k-mean clustering.

3.2 Customer Energy Production Predictor

AstonTAC uses Input-Output HMM (IOHMM) [Bengio and Frasconi, 1995] to predict the customer energy production. The IOHMM is a form of NHHMM where inputs variables condition the states of the hidden variables or/and the value of the observed variables.

There are two types of energy production in the simulation environment: solar energy production and wind energy production. The solar energy production is influenced by temperature and cloud cover; and the wind energy production is influenced by wind speed and wind direction. The weather states are the inputs of our IOHMM model. Hidden states of the energy production are determined using the same approach employed for the energy consumption forecast (see Section 3.1). The states of the weather are also determined using the same technique: Power TAC server provides initial weather data that are sent to Matlab for a k-means clustering.

For each of the energy production types, the weather states are defined as a combination of the corresponding parameter states. For example, considering the wind production, if four states are defined for the wind direction and four states for the wind speed, then the resulting number of weather states is sixteen. Since Power TAC provides during the game a rudimentary weather forecast for the next 24 simulated hours, we use the predicted weather values to predict the most probable hidden states of the energy production. The predicted values of the observed random variables are the corresponding k-means centroids. The IOHMM conditioned transition matrix is also updated during the game using the information

(weather report and energy production values) provided by the game server.

3.3 Customer Energy Demand Manager

The customer energy demand manager calculates the total remaining energy to buy, which is the net customer demand less the energy bought for that time slot so far in the game. From the energy broker viewpoint, for each time slot, the net predicted customer energy demand that has to be satisfied is the difference between the predicted energy volume consumption and the predicted energy volume production. The energy broker needs to buy energy from the wholesale market only if there is a positive difference. Furthermore, the energy broker has 24 trading hours to buy the energy volume in order to cover the predicted demand. This means that for every hour (up to 24 hours) ahead the broker can buy a portion of the energy demand. The actual volume to buy for a specified time slot is determined by the MDP model.

3.4 Clearing Price Predictor

AstonTAC uses an NHHMM to predict the wholesale market clearing price. In the wholesale market, the market is cleared every hour for the 24 hours ahead. We design a $24\times20\times20$ conditioned transition matrix to predict the hidden states of the clearing prices. For each hour ahead, there is a 20×20 transition matrix. The calculation of the hidden states, the training of NHHMM, prediction of the observed values and the update of the NHHMM are similar to the methods used for the energy consumption NHHMM (see Section 3.1).

3.5 AstonTAC MDP Model

Upon receiving the predicted energy price for each time slot and the remaining energy to buy in the wholesale market, the AstonTAC MDP decides the bids to place for each hour ahead. By using the MDP model, AstonTAC aims to buy energy at a low price for the time slots desired and to keep the imbalance as low as possible. This subsection explains the four parameters of our MDP and the computation of the optimal value and policy.

Parameters of the AstonTAC MDP

The parameters of the AstonTAC MDP are defined by the states, actions, state transition probabilities and reward function.

Set of states (S): The state is represented by the *number of hours ahead*, the *remaining quantity of energy needed* and the *predicted clearing price*. The remaining needed quantity is denoted as a ratio as $R_r = q_{r_t}/q_{n_k}$ where $0 \leq R_r \leq 1$; q_{r_t} is the remaining quantity of energy needed at time t. q_{n_k} is the predicted quantity of energy needed that satisfies the customer demand at a particular time slot k in the game. t is the time ahead or the time remaining and $0 < t \leq T$; T is the finite horizon time $T = 24$. q_{r_t} is provided by the customer energy demand manager. The predicted clearing price is mapped in 24 price states of the MDP. The predicted clearing price is classified in one of 24 states where the best (lowest) prices are in state 1 and the highest price in state 24.

Set of actions (A): The MDP action is a bid that is placed in the wholesale market. The bid is a combination of the bidding price and the bidding volume for a time slot in the game. For each time slot, the MDP decides the bid to place for the future 24 time slots that are available for trading in the wholesale market. Bid prices are the predicted clearing prices provided by the clearing price predictor. The energy volume of the bids are decided by the MDP. The bid energy volume is represented as the ratio of the remaining energy needed that should be ordered. For instance, AstonTAC MDP can choose to place bids with 10, 20 or 80% of the remaining energy needed as bid energy volume.

Transition probability (P): Probability ($P(s'|s,a)$) is that the AstonTAC MDP transitions from one state $s \in S$ to another $s' \in S$ after taking the action $a \in A$.

Reward function (R): The agent receives two types of reward: an immediate reward for buying at low price and a delayed reward for the energy balancing after 24 decision steps. The immediate reward is given to the agent according to bid setting: bidding price and bidding volume. For example, if the predicted clearing price is high, the agent will receive a positive reward for buying a low energy volume and negative reward for buying high energy volume. The value of the delayed reward is based on the imbalance ratio at the end of the 24 steps. The simulation environment informs the agent each hour about the value of the energy imbalance. In order to guarantee a low energy supply and demand imbalance, the highest positive reward is received when the imbalance is the lowest (i.e., between to 0 and 5%).

Optimal Value and Policy

AstonTAC MDP is a combination of a state-action rewards MDP [Littman, 1996; Singh, 1993] and a finite horizon MDP [Watkins, 1989; Li and Littman, 2005]. The overall strategy is the sequence of the decisions that can be made in order to maximise the total rewards. This overall strategy is denoted by the policy function π where $\pi : S \rightarrow A$. Following the policy π, at each state s of a run, the agent takes the decision and transitions the system with a probability of $P(s'|s,a)$ to the next state s' as follows. Let n_d be the total number of states to visit in the run.

- At state s_0 and time $t = 0$, take action a_0
- Go to s_1, with a transition probability of $P(s_1|s_0,a_0)$
- At state s_1 and time $t = 1$, take action a_1
- Go to s_i, at time $t = i$ with a transition probability of $P(s_i|s_{i-1},a_{i-1})$, where $2 \leqslant i \leq n_d$ and $i \in \mathbb{N}$

MDP enables the agent to find the optimal policy $\pi^*(s)$ that maximises the expected cumulative rewards at state s. Let $V_\pi(s)$ denote the cumulative expected reward function that starts from state $s = s_0$ at time $t = 0$ and uses a policy $\pi(s)$. The best policy $\pi(s)$ at state s is therefore the one that has the maximal value of $V_\pi(s)$. The maximal value of $V_\pi(s)$ is denoted by $V^*(s)$. We define the AstonTAC MDP as follows given $\pi(s)$ and $0 \leq t \leq T-1$ with $T = 24$.

$$V_\pi(s) = E\left[R(s,a_0) + R(s_1,a_1) + R(s_2,a_2) + ... + R(s_{23},a_{23})\right]$$

at time $t = 0$.

$$V_t^*(s) = \max_a \left[R_t(s,a) + \sum_{s'} P_t(s'|s,a) V_{t+1}^*(s') \right]$$

where $a \in A,\ s \in S$ and $s' \in S$.

$$\pi_t^*(s) = \arg\max_a \left[R_t(s,a) + \sum_{s'} P_t(s'|s,a) V_{t+1}^*(s') \right]$$

where $a \epsilon A,\ s \epsilon S$ and $s' \epsilon S$.

The computation of π^* and V^* requires the model parameters (reward function, expected total pay-off and transition probabilities) to be available. AstonTAC MDP learns the model parameters online using a technique inspired by the Monte Carlo Methods presented in [Sutton and Barton, 2000]. The agent learns from on-line experience and decides by comparing the average of experienced returns at each state. During the simulation, AstonTAC uses the pay-off average to decide about the optimal action to take instead of computing the transition probabilities and the expected reward values. The learning occurs at the end of each episode (24 simulated hours) after the delayed reward for the energy balancing is defined. At the beginning of the training, the initial policy that will be evaluated follows the immediate reward setting: buy the maximum of the needed energy if the predicted clearing price of the time slot is the lowest and lowest energy volume if the price is predicted to be the highest.

4 Evaluation

Our evaluation is composed of two parts: the results from the 2012 Power TAC and our analysis of the tournament games.

4.1 Competition Results

The 2012 Power TAC finals [Power TAC Community, 2013a] consisted of 184 games with three individual competitions: 63 games with two players, 105 games with four players and 16 games with seven players. The winner of the competition is the broker agent with the highest normalised total profit of all games. Table 2 shows the result of the competition. This table presents the accumulated profit for each broker in seven-player (annotated Size 1), four-player (annotated Size 2) and two-player (annotated Size 3) games. Considering the non-normalised total, AstonTAC is second after CrocodileAgent. According to the normalised total, AstonTAC is third with a score of -1.217 after CrocodileAgent (with a score of 7.348) and SotonPower (with a score of -1.215). Although CrocodileAgent achieved the highest score, we fear that its success took advantage of the weaknesses of the Power TAC server. Its strategy is not realistic. For instance, when the CrocodileAgent offered tariffs with very high rates, it still attracted a lot of customers for a long period in the game.

In fact, among all the games played with SotonPower, AstonTAC outperforms SotonPower in 63% of the games (27 out of 43) and outperforms MinerTA in 60% of the games (28 out of 47). Moreover, AstonTAC won all two-player games against SotonPower or MinerTA. Table 3 compares the performance of AstonTAC with SotonPower (3(a)) and MinerTA (3(b)) in two-player games."Energy Demand" represents the average energy demand (in MWh) of the broker agent in all the two-player games. This is the sum of the net energy usage for all customer consumption. "Energy Bought" represents

Agent Name	Size 1	Size 2	Size 3	Total (Non normalised)	Total (normalised)
Crocodile Agent	1.26E+10	6.31E+10	2.74E+10	1.03E+11	7.348
AstonTAC	**3.11E+06**	**9.58E+07**	**5.00E+07**	**1.49E+08**	**-1.217**
Soton Power	1.03E+07	6.68E+07	6.22E+07	1.39E+08	-1.215
MinerTA	2.33E+07	6.48E+07	1.83E+07	1.06E+08	-1.217
Mertacor	-5.11E+05	-1.13E+07	4.98E+06	-6.79E+06	-1.227
LARGE power	-4.16E+07	-5.25E+07	1.01E+07	-8.40E+07	-1.238
UTest	1.55E+06	-6.67E+07	-4.68E+07	-1.12E+08	-1.235

Table 2: Result of the Power TAC Tournament

the average energy volume (in MWh) the agent bought from the wholesale market. "Imbalance" shows the average hourly supply demand imbalance ratio (in %). The hourly imbalance ratio is computed by dividing the absolute hourly imbalance for each broker by the hourly customer energy consumption. "Buy Price" is the average price (in EURO/MWh) of the successful bids of the broker agent. "Average Cash" represents the average profit (in EURO) the broker has at the end of the game. According to Table 3, AstonTAC has a bigger market share than its opponents: an average of 92.01% in the games against SotonPower and 98.81% in the games against MinerTA. These two-player games demonstrate that although AstonTAC has more than 92% of the market share, it is able to keep the energy imbalance lower than 10.5%. SotonPower and MinerTA are able to buy energy at lower price in the wholesale market but are not able to control the market or to keep the energy imbalance lower than 80%.

	Agent Name	Energy Demand	Energy Bought	Imbalance (%)	Buy Price	Average Cash
(a)	**AstonTAC**	**78478.84**	**79821.74**	**10.40**	**34.25**	**3.63E+06**
	SotonPower	6810.82	14212.93	128.86	27.17	1.06E+06
(b)	**AstonTAC**	**85492.29**	**84661.96**	**10.43**	**33.98**	**3.66E+06**
	MinerTA	1025.30	1662.22	80.10	26.99	3.43E+05

Table 3: Two-Player Games Result with (a) SotonPower and (b) MinerTA

4.2 Competition Game Analysis

Agent Name	Energy Demand	Energy Bought	Imbalance (%)	Buy Price	No. of Games
AstonTAC	**29009.72**	**34731.37**	**21.52**	**29.22**	**67**
Mertacor	40530.53	40350.92	18.16	51.84	69
LARGE power	22094.77	23434.50	35.21	37.99	71
MinerTA	3477.17	4356.54	75.08	17.44	74
SotonPower	9934.13	17683.05	146.29	19.26	68
Crocodile Agent	25099.90	63609.87	425.70	32.46	68

Table 4: Brokers' Performance in Wholesale Market

To further investigate the performance of AstonTAC in the wholesale market, we analysed the games in Power TAC fi-

nals. We mainly focused on the performance of AstonTAC in terms of the bidding in wholesale market and balancing supply and demand. Table 4 shows the brokers' performance in all successful games.[1] Mertacor did well in keeping the energy imbalance low. MinerTA and SotonPower managed to buy at low prices. AstonTAC performs well: both in energy balance and in buying at low price. As shown in Table 4, AstonTAC is the only agent that can buy energy at low price in the wholesale market and keep energy imbalance low. The fact that using the MDP, AstonTAC has an energy imbalance of 21.52% indirectly shows that the NHHMMs used by AstonTAC MDP for energy production and consumption provide acceptable prediction values.

In order to evaluate AstonTAC MDP and the clearing price prediction, we analyse the wholesale market performance of several agents in a randomly chosen game 418 and the result is shown in Figure 3. Figure 3(a) shows the average clearing price in each hour ahead, Figures 3(b) and 3(c) illustrate the average volume of energy bought by each broker.

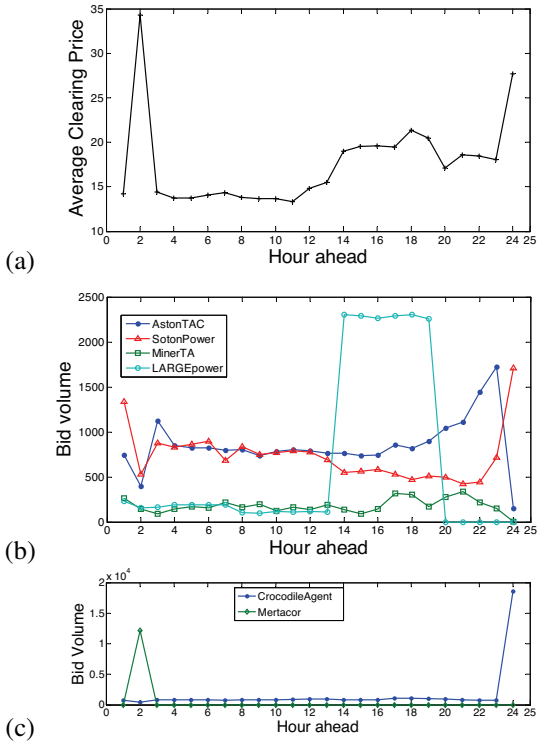

Figure 3: Wholesale Market Performance in Game 418

According to Figure 3(a), the average energy clearing prices are the highest at time slots two and twenty-four hours ahead. The lowest clearing prices are observed at time slots one, and between three and eleven hours ahead. From Figures 3(b) and 3(c), we can see Mertacor, CrocodileAgent and LargePower do not adapt their bidding behaviour to the market clearing price. In contrast, AstonTAC, SotonPower and MinerTA try to adapt their bidding behaviour to the clearing price. Thus, agents that are not adapting to the wholesale market may end up buying energy at high price and adaptive agents can manage to buy less energy at high price. This is the case for AstonTAC and MinerTA which buy less energy volume at time two and twenty-four hours ahead. Although SotonPower adapts its bidding behaviour to the market, it ends up buying more energy twenty-four hours ahead when the energy is high. Using the MDP to balance the retail market, AstonTAC is able to buy less energy at high price and buy a constant volume of energy otherwise. The fact that AstonTAC buys less energy at time two and twenty-four hours ahead, demonstrates that the NHHMM price prediction provides a reliable clearing price prediction.

[1]A successful game is a game that lasts over 1320 time slots which is the minimum duration of a game.

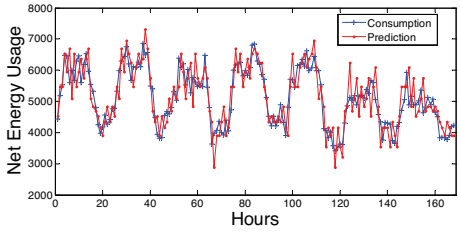

Figure 4: Prediction of the Energy Consumption in Game 562

To provide some details about the performance of our prediction techniques in the PowerTAC environment, we decide to analyse a randomly chosen game with two brokers. We observe the performance of our prediction one hour ahead over several days. Figure 4 shows the performance of the NHHMM for the forecast of the energy consumption in KWh for a customer called "MedicalCenter" (Customer ID: 513) during the game 562 between time slots 1131 and 1298. The standard deviation of the prediction is 501.24 with a mean value of 5027.46 of the observations. In general, the standard deviation of our NHHMM predictions is 10% of the mean value of the observations.

5 Conclusion

This paper details the design, implementation and evaluation of the AstonTAC MDP. The focus of the paper is on the bidding in the energy wholesale market. Through analysing the actual tournament, we found out that AstonTAC is the only agent that can buy energy at low price in the wholesale market and keep energy imbalance low. AstonTAC MDP is designed independently from the Power TAC simulation environment. Moreover, AstonTAC MDP provides a concrete bidding approach for the future energy market where the energy demand will be more satisfied by renewable energy sources. Thus, the approach used in this paper could be generalised to other energy wholesale markets in real world. For the future, we plan to improve the prediction techniques for energy demand and price in order to improve the performance of our MDP.

Acknowledgments

The authors would like to thank the Royal Academy of Engineering for the research exchanges with China award.

References

[Bach *et al.*, 2012] T. Bach, J. Yao, J. Wang, and S. Yang. Research and application of the Q-learning for wholesale power markets. In *2nd International Conference on Consumer Electronics, Communications and Networks*, pages 1192–1197. IEEE, 2012.

[Bengio and Frasconi, 1995] Y. Bengio and P. Frasconi. An input output HMM architecture. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in neural information processing systems*, volume 7, pages 427–434. MIT Press, Cambridge, MA, 1995.

[Bishop, 2006] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science and Business Media, LLC, 2006.

[Cancelo *et al.*, 2008] J.R. Cancelo, A. Espasa, and R. Grafe. Forecasting the electricity load from one day to one week ahead for the spanish system operator. *International Journal of Forecasting*, 24(4):588–602, 2008.

[Conejo *et al.*, 2005] A.J. Conejo, M.A. Plazas, R. Espinola, and A.B. Molina. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, 20(2):1035–1042, 2005.

[Contreras *et al.*, 2003] J. Contreras, R. Espinola, F.J. Nogales, and A.J. Conejo. ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18(3):1014–1020, 2003.

[Gao *et al.*, 2000] F. Gao, X. Guan, X.R. Cao, and A. Papalexopoulos. Forecasting power market clearing price and quantity using a neural network method. In *Power Engineering Society Summer Meeting*, volume 4, pages 2183–2188. IEEE, 2000.

[Garcia *et al.*, 2005] R.C. Garcia, J. Contreras, M. Van Akkeren, and J.B.C. Garcia. A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, 20(2):867–874, 2005.

[Harvey and Koopman, 1993] A. Harvey and S.J. Koopman. Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88(424):1228–1236, 1993.

[Ketter *et al.*, 2012] W. Ketter, J. Collins, P. Reddy, C. Flath, and M. Weerdt. *The Power Trading Agent Competition*. ERIM Report Series, 2012.

[Li and Littman, 2005] L. Li and M.L. Littman. Lazy approximation for solving continuous finite-horizon MDPs. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, page 1175. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press, 2005.

[Littman, 1996] M.L. Littman. *Algorithms for sequential decision making*. PhD thesis, Brown University, 1996.

[Mandal *et al.*, 2005] P. Mandal, T. Senjyu, K. Uezato, and T. Funabashi. Several-hours-ahead electricity price and load forecasting using neural networks. In *Power Engineering Society General Meeting*, volume 3, pages 2146–2153. IEEE, 2005.

[Nguyen and Nabney, 2010] H.T. Nguyen and I.T. Nabney. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy*, 35(9):3674–3685, 2010.

[Power TAC Community, 2013a] Power TAC Community. Overview for tournament: finals-2012-12. http://wolf-08.fbk.eur.nl:8080/TournamentScheduler/faces/tournament.xhtml?tournamentId=9, 2013. [Online; accessed 28-January-2013].

[Power TAC Community, 2013b] Power TAC Community. Power Trading Agent Competition. http://www.powertac.org, 2013. [Online; accessed 28-January-2013].

[Ramanathan *et al.*, 1997] R. Ramanathan, R. Engle, C.W.J. Granger, F. Vahid-Araghi, and C. Brace. Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, 13(2):161–174, 1997.

[Singh, 1993] S.P. Singh. *Learning to solve Markovian decision processes*. PhD thesis, University of Massachusetts, 1993.

[Song *et al.*, 2000] H. Song, C.-C. Liu, J. Lawarrée, and R.W. Dahlgren. Optimal electricity supply bidding by Markov Decision Process. *IEEE Trans. Power Systems*, 15(2):618–624, 2000.

[Sutton and Barton, 2000] R.S. Sutton and A.G. Barton. *Reinforcement Learning, an Introduction*. The MIT press, 2000.

[TAC Community, 2013] TAC Community. Trading Agent Competition. http://www.sics.se/tac, 2013. [Online; accessed 28-January-2013].

[Tellidou and Bakirtzis, 2006] A.C. Tellidou and A.G. Bakirtzis. Multi-agent reinforcement learning for strategic bidding in power markets. In *3rd International Conference on Intelligent Systems*, pages 408–413. IEEE, 2006.

[Watkins, 1989] C.J. Watkins. *Learning from Delayed Rewards*. PhD thesis, Cambridge University, 1989.

[Yao *et al.*, 2000] S.J. Yao, Y.H. Song, L.Z. Zhang, and X.Y. Cheng. Wavelet transform and neural networks for short-term electrical load forecasting. *Energy Conversion and Management*, 41(18):1975–1988, 2000.

[Zhang and Luh, 2005] L. Zhang and P.B. Luh. Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method. *IEEE Transactions on Power Systems*, 20(1):59–66, 2005.

[Zheng *et al.*, 2005] H. Zheng, L. Xie, and L.-Z. Zhang. Electricity price forecasting based on GARCH model in deregulated market. In *The 7th International Power Engineering Conference*, pages 1–410. IEEE, 2005.