Bilateral Contract Prices Estimation using a Q-Learning based approach

Jaime Rodríguez-Fernández[1], Tiago Pinto[2], Francisco Silva[1], Isabel Praça[1], Zita Vale[1], Juan Manuel Corchado[2]

[1]GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development
Polytechnic of Porto (ISEP/IPP), Porto, Portugal
{fspsa, icp, zav}@isep.ipp.pt
[2]BISITE Research Centre, University of Salamanca (USAL), Calle Espejo, 12, 37007 Salamanca, Spain
{tpinto, corchado}@usal.es

Abstract—**The electricity markets restructuring process encouraged the use of computational tools in order to allow the study of different market mechanisms and the relationships between the participating entities. Automated negotiation plays a crucial role in the decision support for energy transactions due to the constant need for players to engage in bilateral negotiations. This paper proposes a methodology to estimate bilateral contract prices, which is essential to support market players in their decisions, enabling adequate risk management of the negotiation process. The proposed approach uses an adaptation of the Q-Learning reinforcement learning algorithm to choose the best from a set of possible contract prices forecasts that are determined using several methods, such as artificial neural networks (ANN), support vector machines (SVM), among others. The learning process assesses the probability of success of each forecasting method, by comparing the expected negotiation price with the historic data contracts of competitor players. The negotiation scenario identified as the most probable scenario that the player will face during the negotiation process is the one that presents the higher expected utility value. This approach allows the supported player to be prepared for the negotiation scenario that is the most likely to represent a reliable approximation of the actual negotiation environment.**

Keywords—bilateral contracts, decision support, electricity market, learning algorithm, negotiation process.

I. INTRODUCTION

All over the world electricity industry restructuring have placed several challenges to governments and to the companies that are involved in generation, transmission and distribution of electrical energy. The result of the electricity markets restructuring, which was performed in order to the competitiveness could be increased and consequently encourage a decrease in electricity prices, has led to a constant process of changes in electricity markets operation rules. However, it also had exponential implications in the increase of the complexity and unpredictability in these markets [1, 2].

Nowadays several market models exist, with a set of complex rules and particular regulations creating the need to foresee market behavior. Some implemented market type are such as day-ahead spot markets, balancing markets, intraday markets, bilateral contracts, forward and future contracts' negotiations, among others [2]. Each electricity market has its own characteristics and clearing price mechanisms, taking into account the power systems reality and the available energy mix.

Some markets have the clearing mechanism based on the optimization of offers, such as most electricity markets in the U.S. [3]; and other based on symmetric or asymmetric bids, as is the case of most European countries [4]. However, electricity trade worldwide is also supported by means of bilateral contracts [5].

Thus, the continuous change of the existing market mechanisms requires the need to understand market's behavior and how the interaction between the players affect the market has contributed to the increased use of simulation and decision support tools [2]. Multi-agent based software is particularly directed to analysis of dynamic and adaptive systems with complex interactions among its constituents, such as the energy sector [6]. Several modeling tools for the study of electricity markets have emerged. Some relevant examples are Electricity Market Complex Adaptive System (EMCAS) [7], Agent-based Modeling of Electricity Systems (AMES) [8], Genoa Artificial Power Exchange (GAPEX) [9], and Multi-Agent System for Competitive Electricity Markets (MASCEM) [10, 11]. Although several modelling tools have emerged to the study of restructured electricity markets, particularly using multi-agent systems, these tools present a common limitation: the lack of adaptive learning capabilities that enable them to provide effective support to the decisions of market entities. Current tools are directed to the study of market mechanisms and interactions among participants, but are not suitable for supporting the decision of the players' negotiators in obtaining higher profits in energy transactions.

These restrictions encourage the need for the development of adaptive tools, able to provide effective support decision to market negotiating players. Such tools should provide the means for a real improvement in players' markets results analyzing the data generated by the simulations and the real electricity markets operation. A new multi-agent adaptive learning system - AiD-EM (Adaptive Decision Support for Electricity Markets Negotiations) – has been integrated with MASCEM with the purpose of providing decision support to electricity markets' negotiating players. This multi-agent system provides decision support to market players for different types of negotiations, such as DECON (Decision Support for Energy Contracts Negotiation), which is addressed in this paper. DECON provides specific decision support for players when negotiating by means of bilateral contracts. DECON implements several methodologies to analyze competitor

players' negotiation profiles enabling the adjustment of the adopted negotiation strategies and tactics [12].

The market players' behavior is mainly based on strategic definition of prices and quantities in energy transactions, in which each competitor player may have a prediction about possible contract prices before reaching an agreement for the negotiation process. Hence, relevant information concerning competitors' history of previous negotiations can be used to improve the decision making process, providing support in order to the definition of computational models to choose the most appropriate parties to negotiate [13]. Techniques such as adaptive learning and game theory [14], which explores the study of algorithms that can learn from and make predictions or decisions on data, allow the assessment of different negotiation contexts and to dynamically learn over time. [15, 16].

This paper presents a learning method which has the aim of supporting the decisions of players in bilateral contracts' pre-negotiating stage. The aim is achieving an advantageous position that allows identifying the ideal negotiators to trade, enhancing the outcomes of the negotiation process. This method is integrated in DECON and is based on the application of the Q-Learning reinforcement learning algorithm to learn the potential forecasting contract price that is the most likely to occur. These forecast scenarios are determined using several methodologies, such as data mining techniques [17], artificial neural networks (ANN) [18], support vector machines (SVM) [19], fuzzy logic [20], among other methods [6]. However, no method presents a better performance than all others in every situation, only in particular cases and contexts [6]. Thus these contract prices forecasting are submitted to some degree error. As a consequence of that is essential for supporting of decision process to localize the best forecast method.

After this introductory section, Section II presents an overview the decision support capabilities provided by MASCEM and AiD-EM, highlighting the DECON system. Section III provides the proposed learning method to estimate bilateral contract prices using a Q-Learning based approach, and section IV presents a case study that shows the experimental results of the proposed methodology using the alternative negotiation scenarios furnished by AiD-EM and historic bilateral contracts data. Finally, Section V presents the most relevant conclusions of this work.

II. AiD-EM DECISION SUPPORT SYSTEM

AiD-EM is a multi-agent system that provides adaptive decision support for electricity market negotiations, resulting from the integration of several decision support methodologies developed in GECAD (Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development) [12]. This system consists of several modeling tools, where each one is independent of the other and oriented to decision support for different types of trading [12].

Figure 1 shows an overview of the AiD-EM system with a representation of this main components. The goal is to provide the most realistic and complete decision support for market negotiating players as possible. Thus, AiD-EM uses important inputs such as real market data, data extracted from simulations and external sources data (e.g. weather conditions such as wind speed, solar intensity and temperature; raw materials prices, among many others). In addition, AiD-EM includes an interface with the MASCEM electricity market simulator, which allows to simulate the main types of negotiations normally considered in electricity markets, such as day-ahead and intraday pool (symmetric and asymmetric) markets, bilateral contracts, forward markets and mixed markets [10, 11]. This connection to MASCEM plays an essential role that allows to recreate realistic simulation scenarios of several European electricity markets to test and validate the AiD-EM's methodologies for decision support [12].
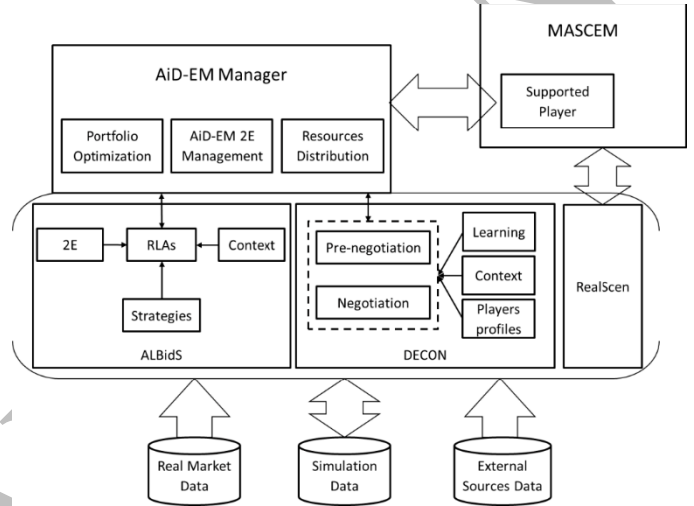

Fig. 1. Overview of AiD-EM's main components [12]

As can be observed from Fig. 1, AiD-EM is includes a manager agent, which is responsible for managing the decision support process, communicating with the market player, and also executing the portfolio optimization model [21], which decides in which markets should the player negotiate at each time, depending on the expected market prices. The manager agent uses two distinct multi-agent decision support systems, depending on the negotiation type. ALBidS [12] is used for decision support in auction based markets, and DECON [14] is applied as decision support system for bilateral contracts negotiations. The bilateral contracts negotiation process in DECON considers two distinct phases. The first is the pre-negotiation step, in which the decision support system evaluates the players available for negotiations, and decides with whom it would be more advantageous to negotiate with, according to the expected negotiation prices. This is the phase addressed by this paper. The second phase corresponds to the actual negotiation process, in which different strategies and tactics are used depending on the negotiation profile of the opponents.

III. PROPOSED METHODOLOGY

The behavior of electricity markets players is normally based on strategies whose purpose is to define energy price and transacted amounts. As a consequence it is essential that the negotiators can be able to predict the expected prices resulting from potential negotiations. Using historical data established in previous agreements, several forecasting methodologies are applied in DECON system to prognosticate the expected established contract price for each player, for different

transacted amounts. As previously mentioned, no method presents a better performance than all others in every circumstances [6]. Therefore, these forecasting method are submitted to some degree error. Hence it is crucial for supporting of decision process to localize the best forecast method. These issues motivated the development of the present work that proposes to undertake a learning process to recognize the potential forecasting contract price that present the higher probability of occurrence in each current context.

The learning process allows an agent to acquire a skill or knowledge that is not available. In fact, an analysis and appropriate learning can improve the results of the participation of stakeholders. The proposed method uses a learning process based on the assessment of likelihood of occurrence of each alternative scenario of trading. Thus, this approach allows the supported player to be prepared for the negotiation scenario that is the most likely to occur and perform the action that generates better results.

The bilateral contract price estimation approach based on the application of the Q-Learning reinforcement learning algorithm that is proposed in [14], due to dynamic environment such as bilateral negotiations where an agent learns through attempt and error, *i.e.* the agent operates in an environment formed by a set of possible states where the agent can choose actions within a set of possible actions, so each time that the player performs an action, this receives a reinforcement value indicating the immediate value of the resulting state transition. Thus, the only learning source is the agents' own experience, whose goal is to acquire an actions policy that maximizes its overall performance [22].

The proposed methodology proposes an adaptation of the Q-Learning algorithm [23] to undertake the learning process. Q-Learning is a very popular reinforcement learning method. It is an algorithm that allows the autonomous establishment of an interactive action policy. It is demonstrated that the Q-Learning algorithm converges to the optimal proceeding when the learning state-action pairs Q is represented in a table containing the full information of each pair value [24]. The basic concept behind the proposed Q-Learning adaption is that the learning algorithm is able to learn a function of optimal evaluation over the whole space of context-scenario pairs c x s. This evaluation thus defines the confidence value Q that each scenario is able to represent the actual encountered negotiation scenario s in context c. The Q function performs the mapping as in (1):

$$Q: c \: x \: s \: \to \: U \qquad (1)$$

where U is the expected utility value when selecting a scenario s in context c; Q function (c, s), the expected future reward when choosing the scenario s in context c, it is learned through try and error according to equation (2):

$$Q_{t+1}(c_t,s_t) = Q_t(c_t,s_t) + \alpha(c_t,s_t)[r_{s,c,t} + \gamma \cdot U_t(c_{t+1}) - Q_t(c_t,s_t)] \quad (2)$$

where c_t is the kind of context when performing under scenario s_t at time t:

- $Q_t(c_t,s_t) \to$ represents the value of the previous iteration (each iteration represents each new contract with the target agent). Generally, the Q value is initialized to 0.

- $\alpha(c_t,s_t)$ $(0 < \alpha \leq 1) \to$ is the learning rate which determines the extent to which the newly acquired information will replace the old information, e.g. assuming a value of 0 learns nothing; on the other hand, to a value of 1, it is fully deterministic environment.

- $r_{s,c,t} \to$ is the reward, which represents the quality of the pair context-scenario *(c x s)*. It appreciates the positive actions with high values and negative with low values, all of them are normalized on a scale from 0 to 1. The reward r is defined in the following equation (3):

$$r_{s,c,t} \: = \: 1 - |RP_{c,t,a,p} - EP_{s,c,t,a,p}| \qquad (3)$$

where $RP_{c,t,a,p}$ represents the real price that has been established in a contract with an opponent p, in context c, in time t, referring to an amount of power a; and $EP_{s,c,t,a,p}$ is the estimation price of scenario that corresponds to the same player, amount of power and context in time t. All r values are normalized in a scale from 0 to 1.

- γ $(0 \leq \gamma \leq 1) \to$ is the discount factor which determines the importance of future rewards, e.g. assumes a value of 0 only evaluates current rewards, and higher values than 0 takes into account future rewards.

- $U_t(c_{t+1}) \to$ is the estimation of the optimal future value which determines the utility of scenario s, resultant in context c. U_t is obtained using the Q function that has been learned to date, and is calculated as in (4):

$$U_t(c_{t+1}) \: = \: \max_s Q(c_{t+1},s) \qquad (4)$$

The Q learning algorithm is executes as follows:

- For each c and s, initialize $Q(c,s) = 0$;

- Observe new event (new established contract);

- Repeat until the stopping criterion is satisfied:
    - Select new scenario for current context;
    - Receive immediate reward $r_{s,c,t}$;
    - Update $Q(c,s)$ according to the equation (2);
    - Observe new context c';
    - $c \to c'$.

After each update all Q values are normalized according to the equation (5), in order to facilitate the interpretation of values of each scenario in a range from 0 to 1.

$$Q'(c,s) \: = \: \frac{Q(c,s)}{max[Q(c,s)]} \qquad (5)$$

The proposed learning model assumes the confidence of Q values as the probability of a scenario in a given context. $Q(c,s)$ learns by treating a forecasting error, updating each time a new observation (new established contract) is available again. Once all pairs context-scenario have been visited, the scenario that presents the highest Q value in the last update is chosen by the learning algorithm to identify the most likely scenario to occur under actual negotiation.

IV. CASE STUDY

A. Specifications

This section presents a case study with the goal of demonstrating the performance of the proposed methodology for the bilateral contract price estimation using a Q-Learning based approach. For this case study, a historic database concerning the past log of established contracts of different electricity market players is used to apply the proposed methodology and assess its performance. The used data has been based on real data extracted from the Iberian Electricity Market -MIBEL [25].

The simulations involve several competitor players and their historic logs of negotiations with the supported player, in order to compare and validate the learning mechanism of algorithm. For each competitor player are considered:

- 5 forecasting contract price methods, where there is an expected price for each amount of energy (from 1 until 10 MW). These forecast scenarios are undertaken using DECON as explained in section II;

- Bilateral contracts' established prices and power amounts, concerning deals established in MIBEL [25];

Hence, the overall goal is to update the Q value of each forecast method (scenario) for each competitor player whenever they establish new contracts. In addition, different combinations of input parameters are tested, such as the discount factor and learning rate, in order to analyze their impact on the evolution of Q values and to have a suitable learning mechanism which chooses the most likely forecast method to occur, i.e. the scenario with a lower forecast error.

B. Results

This sub-section presents the results of the implemented learning model. Since the reinforcement learning algorithm is going to learn throughout established contracts in time, the results show the evolution and convergence of Q values for each parameter combination over each iteration (newly established contracts). In order to define the Q-Learning algorithm, a discount factor (DF) value of 0.8 is applied, which has been considered as the most suitable value by means of a previous sensitivity analysis. Figures 2, 3, 4 and 5 show a learning rate (LR) variation analysis whose assessment allows to determine how LR influences in the results, i.e. in the evolution of Q values. For this analysis different values of LR (0.1:0.4:0.7:1.0) have been uses, using the fixed value of 0.8 for the DF.
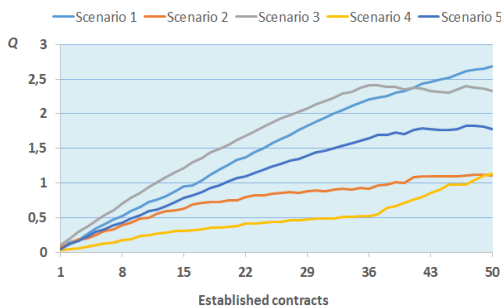

Fig. 2. Q-Learning algortihm evolution throughtout established contract the subject player ID. Parameters: LR=0.1; DF=0.8.
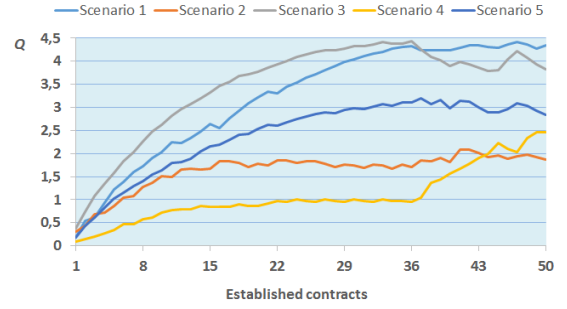

Fig. 3. Q-Learning algortihm evolution throughtout established contract the subject player ID. Parameters: LR=0.4; DF=0.8.
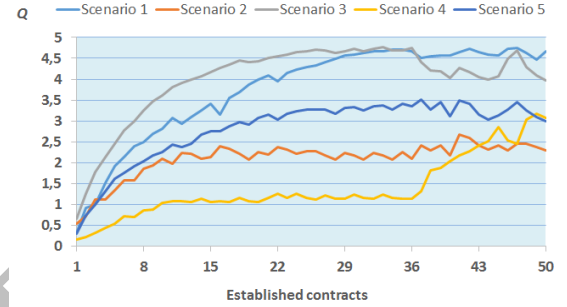

Fig. 4. Q-Learning algortihm evolution throughtout established contract the subject player ID. Parameters: LR=0.7; DF=0.8.
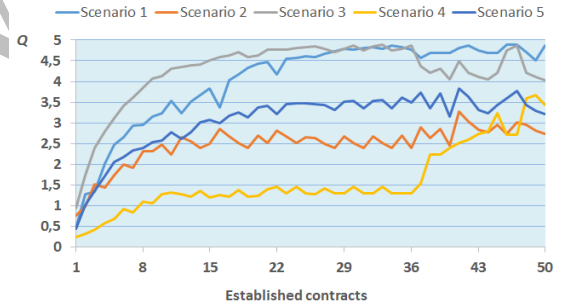

Fig. 5. Q-Learning algortihm evolution throughtout established contract the subject player ID. Parameters: LR=1.0; DF=0.8.

By analyzing Figures 2, 3, 4 and 5, it is clear that the variation of LR throughout the iterations changes the shape of the Q-Learning algorithm convergence process, since it determines the extent to which the acquired information will complement and replace the old information.

Using high values of LR such as Figure 5, where the learning process is fast; the value of the Q function will only reflect the latest iterations (not the goal of our learning process); the main goal is not to forget the previous values, but to use new and old information as complementary. Thus, the ideal situation to choose the best trading scenario for a player would be a high number of iterations (many established contracts) and a LR relatively low, so that it will reflect scenarios near reality. However, a learning process that is too slow is also not advisable, such as the one shown in Figure 2, since the observed new information becomes almost irrelevant for the learning process; therefore a suitable balance is needed. On one hand, by analyzing the results for a number of 50 established contracts,

the most appropriate learning rate is 0.4 as shown in Figure 3. These are the values that apparently reflect a better balance between the consideration of new events and the previous values already learned by the system. On the other hand, it can be observed that for low LR values such as Figure 2, the variation of the Q values is very low (almost linear) and there is a very slow convergence, while for high LR values such as Figure 4 and Figure 5, the variation is very high at every iteration (waves recorded on the graphs).

With the purpose of determining the most probable negotiation scenario, the Q-Learning based algorithm is applied, using DF = 0.8 and LR = 0.4, in order to provide the learning algorithm with a rather quick, yet balanced learning rate, with the aim at facilitating the fast adaptation to the most recent perceived events. The confidence value Q in each of the scenarios created using the forecasts resulting from each of the 5 algorithms throughout 50 observations (newly established contracts) is presented in Table I.

TABLE I. Q values of each scenario throughout 50 iterations.

Contract	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
5	0.76	0.55	1	0.22	0.64
10	0.77	0.57	1	0.27	0.62
15	0.79	0.50	1	0.25	0.64
20	0.84	0.47	1	0.23	0.67
25	0.86	0.44	1	0.23	0.67
30	0.92	0.40	1	0.23	0.68
35	0.96	0.39	1	0.22	0.70
40	1	0.42	0.94	0.34	0.71
45	1	0.44	0.89	0.48	0.68
50	1	0.41	0.88	0.52	0.65

These normalized Q values of Table 1 allow a better identification of the recommended scenario as the most reliable over the iterations (the value of 1 indicates the recommended forecast method by the algorithm until the last observation).

C. Results analysis

In order to validate the Q-learning results, the obtained outcomes are compared to the input data in order to assess if the results are in accordance with the best expected forecast method throughout the iterations. Figure 6 shows the forecasting contract price methods resulting from 5 different algorithms generated by DECON, where there is an expected price for each amount of energy (from 1 until 10 MW). From Figure 8 it is visible that Scenario 1 has high contract prices for low amounts of power and low prices for high amounts too. The opposite occurs in Scenario 2, low contract price values for low power amounts and large prices for large amounts of energy. In Scenario 3 a large contract price is always predicted to any amount of energy. The latter scenario, the Scenario 5 shows intermediate values but for large power amounts can be verified some high prices as well.

Figure 7 presents the historic log of bilateral negotiation agreements between the subject player and the supported player. Figure 7 thus represents the previous bilateral contracts' established prices for some power amounts. It can be observed that the first contracts always have large prices for almost all power amounts, but at the end (contract 37 onwards) the trend starts to change to low contract prices for high power amounts.
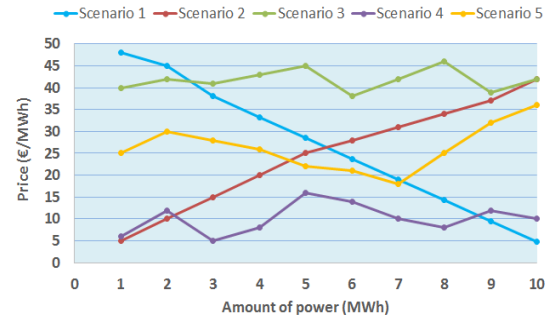

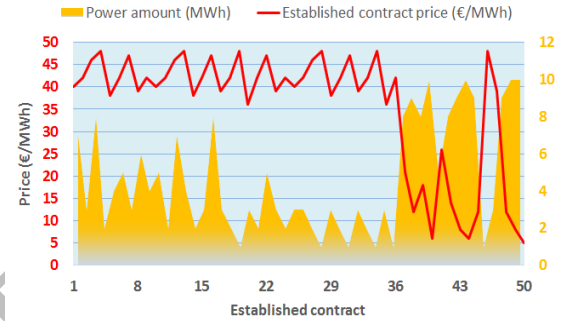Fig. 6. Forecast contract price methods of the subject player ID.

Fig. 7. Historic log of established bilateral contracts of the subject player ID with the supported player.

As previously mentioned, these analyses are essential to evaluate the expected results of the proposed Q-Learning approach. Table II presents a summary of the expected scenario (the best expected forecast method) throughout the established contracts with the subject player.

TABLE II. Expected forecast method throughout iterations.

Established Contracts	Power amount (MWh)	Prices (€/MWh)	Expected forecast
1 → 37	All	High	3
37 → 50	High	Low	1 or 4

In summary, by matching these analyses of data inputs with the obtained results, it is visible by Table I and the represented graphs of Q values (in Fig. 2 to 5) that the best scenarios (with higher Q value) are the Scenario 3 (grey line in Fig. 2 to 5) for the first contracts (until contract 36), and from then forward is Scenario 1 increases to become the scenario with best Q value. This rise is faster in the cases where the LR is higher. Additionally, as verified in Table II, the last contracts also impose a larger confidence value for Scenario 4. However since Scenario 4 was having bad results from the start, its Q value was low, as despite the quick rise of Scenario 4's Q value after contract 36, it was still not enough to become the best scenario in the end of the 50 contracts. Hence, from the analyzed data inputs and their comparison to the achieved results, it can be concluded that the confidence value Q is certainly precise to learn which of the potential scenarios are most likely to represent a reliable approximation of the negotiation environment that the player will face.

V. Conclusions

Currently, automated negotiations in electricity market are an active area of research within the field of computing, particularly with the development of artificial intelligence. This paper proposes a machine learning model, integrated in the DECON system, to provide decision support for the pre-negotiation step of bilateral contracts in electricity markets. The pre-negotiation is a stage that has great importance in the negotiation process because it is responsible for all the preparation and planning of the actual negotiation, i.e., this process aims to identify the ideal negotiators and target prices to face the potential negotiations in order to obtain the greatest possible benefit for the supported player.

The proposed methodology addresses the bilateral contract price estimation based on the application of the Q-Learning reinforcement learning algorithm. Using the proposed approach the implemented model learns which of the potential negotiation scenarios are most likely to represent a reliable approximation of the possible actual negotiation environment for supported player.

Results demonstrate that a balance of the learning parameters is very important for the quality of the Q-Learning algorithm results. It has been shown that the Q values have a huge variation with higher learning rates, and thus it is not so reliable in the long run, although being capable to adapt quickly to new situations, such as fast variations in the contract prices. Moreover, the learning algorithm with high LR is quickly adapted providing a suitable solution for situations where the expected number of contracts is small (e.g. contexts where contracts are not frequently established). Nevertheless, the opposite can be verified with smaller LR. The algorithm is not able to adapt so fast, but it is more consistent because it uses with greatest influence the previously learned values. Therefore, the learning rate should be selected depending on the number of expected observations and according to the rate of variation in the learning conditions (contract prices in this case). In what regards the discount factor (DF) parameter, it can be concluded that higher values are more desirable for this study. Finally, it is noteworthy that it was also demonstrated that the simulated process was in accordance with the a priori analysis of the input data; i.e. the scenario which should be chosen as the most reliable was effectively the scenario that has obtained the largest Q values at the end of learning process, thus confirming the adequacy of the proposed methodology.

As future work, regarding to the learning process of the scenarios that are most likely to occur in the actual negotiation, it is proposed to experiment with alternative learning techniques, such as adapting Roth-Erev algorithm or models based on Bayes theory of conditional probability, in order to enable a comparison with the proposed model in this paper and thus facilitate the choice of the most appropriate learning method for each problem.

References

[1] Sioshansi, F.P., "Evolution of Global Electricity Markets – New paradigms, new challenges, new approaches", *Academic Press*, 2013.

[2] L. Meeus, K. Purchala and R. Belmans, Development of the internal electricity market in Europe, *The Electricity Journal* 18(6) (2005), 25–35.

[3] MISO Energy, homepage, http://www.misoenergy.org (accessed on August 2016).

[4] NordPool, homepage, http://www.nordpoolspot.com (accessed on August 2016).

[5] Algarvio, H.; Lopes, F.; Santana, J. Bilateral Contracting in Multi-agent Energy Markets: Forward Contracts and Risk Management. *Highlights of Practical Applications of Agents, Multi-Agent Systems, and Sustainability-The PAAMS Collection*, Springer International Publishing 2015, 260- 269.

[6] Pinto T., Vale Z., Sousa T., Praça I., Santos G. and Morais H. Adaptive Learning in Agents Behaviour: A Framework for Electricity Markets Simulation. *Integrated Computer-Aided Engineering*, IOS Press, vol. 21, no. 4, pp. 399-415, September 2014.

[7] V. Koritarov, Real-world market representation with agents: modeling the electricity market as a complex adaptive system with an agent-based approach, IEEE Power Energy Mag. (2004) 39–46.

[8] H. Li, L. Tesfatsion, Development of open source software for power market research: the AMES test bed, J. Energy Mark. 2 (2) (2009) 111–128.

[9] S. Cincotti, G. Gallo, Genoa artificial power-exchange, Agents Artif. Intell. (2013).

[10] I. Praça, et al., MASCEM: a multi-agent system that simulates competitive electricity markets, IEEE Intell. Syst. 18 (6) (2003) 54–60.

[11] Z. Vale, et al., MASCEM – electricity markets simulation with strategically acting players, IEEE Intell. Syst. 26 (2) (2011) (Special Issue on AI in Power Systems and Energy Markets).

[12] T. Pinto. 2016. "Decision Support for the Strategic Behaviour of Electricity Market Players". *Doctorate Degree. University of Trás-os-Montes e Alto Douro.*

[13] F. Lopes, M. Wooldridge, A.Q. Novais. 2008. "Negotiation among autonomous computational agents: Principles, analysis and challenges". *Artificial Intelligence Review*.

[14] T. Pinto, Z. Vale, I. Praça, E.J. Solteiro, F. Lopes. 2015. "Decision Support for Energy Contracts Negotiation with Game Theory and Adaptive Learning". *Energies*.

[15] D.A. Goldberg, Genetic algorithms in search, optimization, and machine learning, Addison-Wesley, 1989.

[16] D.A. Gaines, R. Pakath, An examination of evolved behavior in two, reinforcement learning systems, Decision Support Systems 55 (1) (April 2013) 194–205.

[17] Jain, A. K. Data Clustering: 50 years beyond K-Means. Pattern Recognition Letters, Elsevier, 2010, 31, 8, 651-666.

[18] Pinto, T.; Sousa, T. M.; Vale, Z. Dynamic artificial neural network for electricity market prices forecast. *2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES)*, 2012, 311–316.

[19] Pinto, T.; Sousa, T. M.; Praça, I.; Vale, Z.; Morais, H. Support Vector Machines for Decision Support in Electricity Markets' Strategic Bidding. *Neurocomputing, Elsevier*, in press, August 2015.

[20] G.G. Rigatos, Adaptive fuzzy control for differentially flat MIMO nonlinear dynamical systems, *Integrated Computer- Aided Engineering*, IOS Press 20(2) (2013), 111–126.

[21] Pinto, T., Morais, H., Sousa, T.M., Sousa, T., Vale, Z., Praça, I., Faia, R., Solteiro Pires, E.J., "Adaptive Portfolio Optimization for Multiple Electricity Markets Participation" IEEE Transactions on Neural Networks and Learning *Systems*, in press, August, 2015.

[22] R.S. Sutton, A.G. Barto. 1998. "Reinforcement learning: An introduction". *Reinforcement Learning*.

[23] Rahimi-Kian, A.; Sadeghi, B.; Thomas, R.J. Q learning based supplier-agents for electricity markets. *IEEE Power Engineering Society General Meeting*, 2005, 1, 420-427

[24] C.J.C.H. Watkins, P. Dayan. 1992. "Q-learning. Machine Learning". *Machine Learning*.

[25] MIBEL - Iberian Electricity Market Operator, homepage, http://www.omie.es, (accessed on September 2016).