View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repositório Científico do Instituto Politécnico do Porto

Identifying Most Probable Negotiation Scenario in Bilateral Contracts with Reinforcement Learning

Francisco Silva[1], Tiago Pinto[1,2], Isabel Praça[1], and Zita Vale[1]

[1] GECAD - Knowledge Engineering and Decision Support Research Center, Institute of Engineering – Politechnic of Porto (ISEP/IPP), Porto, Portugal *

[2] BISITE – Research Centre, University of Salamanca, Salamanca, Spain

{fspsa,tmcfp,icp,zav}@isep.ipp.pt

[2]tpinto@usal.es

AUTHOR'S VERSION

Abstract. This paper proposes an adaptation of the Q-Learning reinforcement learning algorithm, for the identification of the most probable scenario that a player may face, under different contexts, when negotiating bilateral contracts. For that purpose, the proposed methodology is integrated in a Decision Support System that is capable to generate several different scenarios for each negotiation context. With this complement, the tool can also identify the most probable scenario for the identified negotiation context. A realistic case study is conducted, based on real contracts data, which confirms the learning capabilities of the proposed methodology. It is possible to identify the most probable scenario for each context over the learned period. Nonetheless, the identified scenario might not always be the real negotiation scenario, given the variable nature of such negotiations. However, this work greatly reduces the frequency of such unexpected scenarios, contributing to a greater success of the supported player over time.

Keywords: Automated Negotiation, Bilateral Contracts, Decision Support System, Electricity Markets, Reinforcement Learning Algorithm

1 Introduction

The world is constantly changing. Nowadays, these changes happens a lot faster than before and the tendency is for this rate to keep increasing. This is mainly due to the easier access to information and increased interaction and change of knowledge between people. As the world always seeks equilibrium, when something changes, all other things will change too in order to re-establish the equilibrium.

The Electricity Markets (EMs) are not any different. They have been constantly changing to keep up with the society needs. However, sometimes the

* This work has received funding from National Funds through FCT (Fundaçao da Ciencia e Tecnologia) under the project SPET – 29165, call SAICT 2017.

change takes too long to take place and when it happens, it requires more profound changes. That is what happened to the EMs back in 2000 [1]. Those changes allowed the sector liberalization, introducing free competition in its various segments such as production, transportation and energy distribution. Nowadays there is another key change in the EMs paradigm that needs to be addressed, which is the increased use of energy from renewable sources. The use of this source is being highly encouraged with the aim of adopting a more sustainable growth (by reducing CO_2 emissions), as well as archiving energy independence [2]. The "20-20-20" program [3], which has been introduced by the European Union, is a good example of such encouragements. The mentioned program sets ambitious energy goals to be met by 2020, contributing to the large scale implementation of distributed generation.

While the evolution of EMs contributed to keep their stability and address society needs, it also brought new challenges for the participating entities. The restructuring of the sector introduced new entities with complex interactions that increased its unpredictability. That unpredictability only got bigger with the introduction of renewable energy sources, due to their intermittent nature. Consequently, the participating entities face higher risks and a lot more variables, increasing the importance and impact of decision-making. This way, the participating entities in EMs need proper tools to keep up with the changes, increasing their knowledge and improving their participation. The literature presents several simulators with focus in modelling EMs [4]. However, they are mostly focused on auction-based market models such as Day-ahead spot and Intra-day, overlooking Bilateral Contracts model (negotiation between players). The process of negotiation itself is a subject, common to several different domains, that has been widely explored. A relevant review in automated negotiation identifies the main phases of negotiation and exposes the features that are partially or completely missing in current models [5]. One of the identified gaps is the poor opponents analysis in the pre-negotiation phase, where the negotiator has to define its objectives and the opponent(s) to trade with. In EMs domain, players can establish bilateral contracts with several different players and the selection of the right opponent(s), according to its objectives, can have a great impact in the negotiation outcome. For this purpose, the negotiator needs to be able to identify the most probable negotiation scenario that it can face, among all the different scenarios that can occur in its current negotiation context. Each scenario is composed by the expected prices that each opponent may offer for each power amount.

To address the identified gap, this work presents a methodology, based on a reinforcement learning algorithm (Q-Learning [6]), to determine the most probable scenario that the supported player can face, in a future negotiation. The methodology is integrated in a Decision Support System (DSS) for the pre-negotiation of bilateral contracts, which is able to analyse the opponents and generate multiple negotiation scenarios.

AUTHOR'S VERSION

2 Bilateral Contracts in Electricity Markets

The EMs are usually composed of several market types [7, 8], based on several different models such as: day-ahead spot; intra-day, both usually auction based; and bilateral contracts.

In the scope of EMs, bilateral contracts are long-term contracts established between two entities, buyer and seller, for energy transaction, without the involvement of a third entity. The transaction is usually carried out several weeks or months after the contract is made [9] and usually has the following specifications: start and end dates and times; Price per hour (/MWh) and amount of energy (MW), variable throughout the contract and, finally, a range of hours relative to the delivery of the contract. Players can use customized long-term contracts, trading "over the counter" and electronic trading to conduct bilateral transactions [10]. In MIBEL, there are four types of bilateral contracts: the first type are Forward Contracts, that consist in energy exchange between a buyer and a seller for a future date, for the price negotiated at that moment; the second type are Future Contracts, which are similar to Forward Contracts except that they are managed by a third party responsible for ensuring compliance with the agreement; the third type are Option Contracts, that are similar to the Forward and Future contracts with the difference that the two entities only guarantee a buy/sell option; the last one are Contracts for Difference, that allows concerned entities to protect themselves from the energy price change between the agreement establishment date and the agreed exchange date.

With the exception of Contracts for Difference, this type of negotiation allows players to control the price at which they will transact energy, in contrast to what happens in spot markets, due to the proposals' instability. In establishing a Forward or Future contract, players are committing themselves to transact energy for a given price at a future time, with the risk of making a transaction at a lower price than the expected and lose competitive power. Option Contracts or Contracts for Difference can avoid this risk. The first allows the player to choose not to go through with the exchange while the second ensures that the transaction is carried out at the market price. However, the first option also has the risk of not guaranteeing whether or not the other party will exercise their option to exchange and the second option does not allow better prices than the market. This way, it is possible to understand the risk associated with the negotiation of bilateral contracts and the need that players have of tools that help them reduce this risk and even optimize their profits.

AUTHOR'S VERSION

Automated Negotiation

The process of negotiation itself has been widely explored in the literature of several different domains such as social psychology [11], economics and management science [12], international relations [13] and artificial intelligence [14, 15]. The last one, artificial intelligence, is the most related area with the present work. As this area itself is also very rich in research about the process of negotiation, it has motivated the conduction of a very thorough review [5]. In the review, the authors present the state-of-art of the existing negotiation models and, as result of their study, they are able to present the most common phases of automated negotiation for computational agents: (I) Preliminaries, (II) Pre-Negotiation, (III) Actual Negotiation and (IV) Renegotiation. However, it is important to note that the Preliminaries and Pre-Negotiation phases are often joined together as the Preliminaries can be considered as part of the Pre-Negotiation phase. Despite being a common phase, the Pre-Negotiation is often very simple, not exploring its full potential. The last phase, Renegotiation, is also not present in all models, as some do not allow the final agreement modification. This way, it is possible to verify that the main focus of the existing models is the Actual Negotiation phase. However, the other phases are also important and can have a great impact in the negotiation process.

Decision Support Systems for the Negotiation of Bilateral Contracts

Some DSS for the negotiation of bilateral contracts can be found in the literature, such as Electric Market Complex Adaptive Systems (EMCAS) [16], General Environment for Negotiation with Intelligent Multi-purpose Usage Simulation (GENIUS) [17], and Multi-Agent Negotiation and Risk Management in Electricity Markets (MAN-REM) [18].

EMCAS [16] is a multi-agent simulator that aims to simulate various EMs market models, including Bilateral Contracts. The tool considers the objectives of all the participating players. The players can be either demand or generation company agents. The demand agents formulate their proposals and then each generation agents decide the price for the amount of power they want to sell. At last, the demand agent decides to accept or reject the generation agent conditions.

GENIUS [17] is a multi-agent simulator with the main focus of facilitating and evaluating automated negotiators strategies. The tool main features are: bilateral and multilateral negotiations; agent-to-agent and human-to-agent negotiations; domain independent; negotiators performance analysis, including comparison between results and optimal solution. The negotiation process follows three phases: Preparation, when the agents, protocol and domain of the negotiation are defined; Negotiation, when the actual negotiation occurs; and Post-negotiation, when the negotiation is analysed in detail.

MAN-REM [18] is a framework that combines small multi-agent EMs simulators for the simulation of bilateral contracts. In the simulations, the framework models two agents, besides the expected Seller and Buyer agents, which are: the Trader agent, which distributes the energy; and the Market Operator agent, which validates the contracts. The negotiation process follows three phases: Pre-Negotiation phase, when the proposer agent defines its contract preferences and its response to counter-offers; Actual Negotiation phase, when the Buyer and Seller agents trade offers; and Post-Negotiation phase, when the two entities reach an agreement.

The presented tools are mainly focused in the actual negotiation, being in accordance with the analysis of the previous subsection. Regarding Pre-Negotiation

AUTHOR'S VERSION

phase, EMCAS has a basic approach while the others explore it further, specially GENIUS. However, these tools are not capable to address the identified gap of poor opponents analysis in the Pre-Negotiation phase which does not allow the identification of the most probable negotiation scenario.

3 Proposed Methodology

Reinforcement Learning Approach

This paper proposes the adaptation of Q-Learning [6], a reinforcement learning algorithm, for the identification of the most probable negotiation scenario that the supported player can face, under a certain context. A negotiation scenario is composed by the expected prices that each opponent may offer for each power amount.

The Q-Learning is a very popular reinforcement learning algorithm. The concept of this algorithm is that an agent can take an action, from a set of possible actions, in each state that it can be. The aim of the algorithm is to help the agent identify the best action to take in each state. For that purpose, every state-action pair have an Q value that represents the utility of taking that action in that state. The agent will always choose the action with the highest Q value. Being a reinforcement learning method, Q-Learning is able to update the Q value of each state-action pair. Every time the agent repeats an action in a given state, the Q value of that state-action pair gets a reward r. The algorithm contains two variables related to the future learning: Learning Rate (α), which defines the contribution of the reward to the previous Q value; and Discount Factor (γ), which defines if the algorithm should only consider the current reward or look forward for highest rewards in the future.

The proposed methodology is an adaptation of Q-Learning that, instead of evaluating the best action for each state, evaluates the closest negotiation scenario to the one that the player will face in reality, under a certain context. In this adaptation, the agent is the supported player, the states are each different context c, in which the player may trade, and the actions are each scenario s that the player may face. The equation 1 presents the mapping performed by the Q function.

$$Q : c \times s \rightarrow U \tag{1}$$

The U is the expected utility value, which represents how close the scenario s is to the real negotiation scenario, that the supporter player may face under context c. The utility value (Q value) of each context-scenario gets rewarded once the information about the real scenario is available. This way, it is possible to evaluate how close each scenario was to the real negotiation scenario. The reward r is defined in Equation 2.

$$r_{s,c,t} = 1 - norm|RP_{c,t,a,p} - EP_{s,c,t,a,p}| \tag{2}$$

AUTHOR'S VERSION

The $RP_{c,t,a,p}$ is the real price negotiated by opponent p, in contract c, at the time t, for the amount of power a. The $EP_{s,c,t,a,p}$ is the price expected in scenario s under context c for the same opponent, amount of power and time. To simplify the analysis of each Q value, the r values are normalized, in a scale between 0 and 1, to keep the Q value inside that interval. This way, the Q value resulting from the $Q(c, s)$ function, can be interpreted as the probability of occurrence of the scenario s under context c. After the calculation of the reward of a context-scenario pair, its current Q value gets updated by following Equation 3.

$$\begin{aligned} Q_{t+1}(c_t, s_t) = Q_t(c_t, s_t) + \alpha(c_t, s_t) \\ \times [r_{s,c,t} + \gamma U_t(c_{t+1}) - Q_t(c_t, s_t)] \end{aligned} \tag{3}$$

The α is the learning rate, γ is the discount factor and $U_t(c_{t+1})$ is the maximum utility expected in all of the scenarios of c_{t+1}, as represented by Equation 4.

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

Every time the Q value of a context-scenario pair is updated, its value is normalized, in a scale between 0 and 1, to represent the proximity of the given scenario to the actual negotiation scenario, under the given context. This normalization is represented in Equation 5.

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

This way, the scenario of each context with the highest Q value will have the value 1, after normalization, as it is the closest scenario to the actual negotiation scenario under the same context.

The adapted Q Learning execution is presented in Algorithm 1.

Algorithm 1 Adapted Q-Learning Execution

```
initialize Q(c, s) ← 0
repeat
    wait for new event                          ▷ (new established contract)
    for all scenarios of current contract do
        calculate reward                        ▷ (equation 2)
        update Q(c, s)                          ▷ (equation 3)
    end for
    normalize Q(c, s)                           ▷ (equation 5)
until stopping criteria
```

Decision Support System

The proposed methodology presents itself as a good solution for the determination of the most probable scenario that the supported player may face in a

AUTHOR'S VERSION

future negotiation. However, this methodology requires a tool that is capable to detect different negotiation contexts and generate alternative negotiation scenarios per context. The proposed methodology has been included in a DSS, whose architecture is presented in Figure 1, which meets the identified prerequisites.

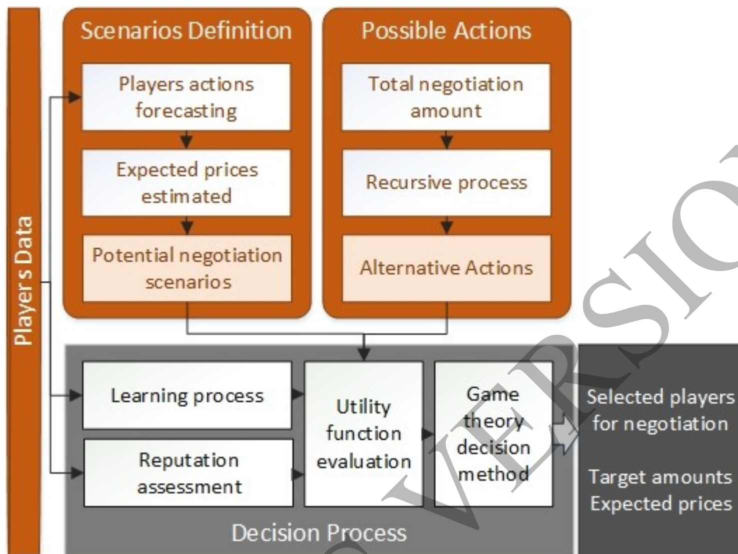

Fig. 1. Architecture of the DSS

AUTHOR'S VERSION

As it can be seen in Figure 1, the DSS is composed by three main components: Scenarios Definition, Possible Actions and Decision Process.

The Scenarios Definition is responsible for the generation of several alternative scenarios for each different context. A scenario is composed by the expected price that each one of the possible opponents find acceptable for the negotiation of different amounts of power (from the minimum negotiable amount to the amount that the supported player wants to trade). The expected price is obtained through forecasting, considering the historical data of the possible opponents. However, it is not always possible to forecast all the power amounts as the historical data might not have enough information about those quantities. In this cases, an estimation is performed instead, based on the prices that were possible to forecast. Each scenario uses a different forecast algorithm to make it possible to generate different scenarios.

After the Scenarios Definition, the generation of all Possible Actions, that the supported player can take, is performed. For this purpose, it is generated all the possible distributions, of the desired power amount to trade, among all the possible opponents. Each possible distribution is one action that the supported player can take. The actions can range from trading all the desired power amount, with only one of the possible opponents, to trade with all of the possible opponents, an equally split amount. Then, to evaluate each action individually and make it possible to select the best action of each scenario, the utility of each action is calculated. The utility of each action takes into account how profitable the action can be (taking into account the expected prices) and the reputation of the involved opponents. The impact of each component depends on the risk that the supported players is willing to face. The lower the impact of the reputation component, the higher the risk. Each action has an utility value per each possible scenario that the supported player can face under the current context.

The generation of all possible actions per each possible scenario creates the need of a decision method that enables the selection of the best action that supported player can take. The third component, Decision Process, is responsible for that final decision. For that matter, the system contains three different decision methods: Optimistic, Pessimistic and Most Probable (made possible by the proposed methodology). The optimistic method chooses the action with the highest utility value among all the scenarios. On the other hand, the pessimistic method, follows the mini-max game theory approach, where the selected action is the one with the highest utility value of the scenario with the lowest global utility (sum of the utility of all the actions of a scenario). At last, the contribution of the proposed methodology to this DSS: the Most Probable decision method, which selects the action with the highest utility, of the scenario that is most probable to occur in reality.

As result of the simulation, the supported player obtains the opponent(s) to trade with and how much power to trade with each one, considering the amount of power that it wants to trade, the list of possible opponents, the risk that it is willing to take and the preferred decision method.

4 Experimental Findings

This section presents a case study that has been conducted to test the proposed methodology as well as the impact of its integration in the presented DSS. For that purpose, the DSS will be run to aid the supported player in the scenario presented in Table 1.

Table 1. Case Study Scenario

Power Amount	40
Transaction Type	Purchase
Context	Weekday
Possible Opponents	5
Reputation Calculation	50% Personal Opinion and 50% Social Opinion
Risk	50% *(The economical and reputation components have the same weight)*
Decision Method	Most Probable *(Proposed Methodology)*
Possible Actions	135 751

AUTHOR'S VERSION

The DSS provides, in this scenario, two different contexts: Weekday and Weekend. For each context, the DSS will generate five different scenarios, provided by three different methods: Artificial Neural Network (ANN), Support Vector Machine (SVM) and Average. The last two consider the last 1000 contracts in their training while ANN has three different methods: ANN1 (500 contracts), ANN2 (1000 contracts) and ANN3 (1500 contracts).

This is a simple amount of scenarios that are capable to test the proposed methodology without being too complex to analyse. However, the DSS is capable to execute a much higher number of scenarios, which makes it possible to use other forecasting techniques and test different configurations (as exemplified with the ANN). Three different versions of the ANN algorithm will be used, where the only difference is the amount of contracts used in its training.

AUTHOR'S VERSION

Besides the presented DSS internal definitions for this case study (contexts and scenario generation methods), there is also another very important definition: the historical data of the possible opponents. The data of the five possible opponents could be generated and still be able to test which scenario is capable to better represent the generated negotiation scenarios. However, the optimal test can only be performed with real data. Therefore, a real dataset is used instead, which contains executed physical bilateral contracts declared in the Spanish System Operator (SO)[19]. The dataset data ranges from 1 July 2007 to 31 October 2008 (16 months / 488 days) and each day contains 24 negotiation periods (one per hour), in a total of 11 712 periods. The negotiations were performed by 132 different players (88 Buyers and 44 Sellers) which established 1 797 996 contracts. The Table 2 presents a detailed overview of the dataset.

Table 2. Dataset Overview

	MIN	AVG	STDEV	MAX
Contracts / Period	128	157	17,78	180
Contracts / Day	147	3 753	485,78	4 287
Contracts / Player	2	27 244	58 653,22	288 160
Contracts / Player / Period	1	5	6,83	29
Power / Period / Contract	1	69,04	6,25	3 575
Power / Player / Contract	1	89,05	223,17	3 575
Power / Period	7 718	10 813	1 346,38	14 128
Power / Day	8 210	258 405,89	34 317,46	316 801
Power / Player	30	1 875 400,33	4 503 101,94	26 081 833

AUTHOR'S VERSION

However, as it can be seen in Table 2, there is not information about contract prices, because the dataset only contains the traded power amount. The established price of a contract is a key information and the involved entities avoid sharing it. The share of such information can negatively affect their future negotiations. This way, there is the need to generate a price for each contract present in the dataset.

To guarantee an increased realism, the contracts price are generated taking into account the market price of the same negotiation period of the Spanish Day-ahead Market [20]. Nevertheless, the contracts established in the same negotiation period, between different players, can not have the same price. It would not make sense. Therefore, each player can have one of five different negotiation profiles: Profile 1, in which the player defines a minimum price, based on the market price, and keeps increasing it according to the power amount increase, until reaching a maximum price; Profile 2, which is similar to Profile 1 but with an higher minimum and lower maximum prices; Profile 3, which follows market price; and Profiles 4 and 5 which are the reverse of Profiles 2 and 1, respectively, in which the price keeps decreasing according to the decrease in the traded power amount.

Besides the contracts price definition, the dataset also needs another complement, to make it possible to use the full capabilities of the DSS: the reputation assessment. The DSS is capable of calculating the reputation of each opponent, based on various components, but it requires the personal opinion of each player about the other players. Therefore, the personal opinions of each player have been defined taking into account three components: personal experience with the evaluated player; number of opponents that traded with that player; and number of contracts established by the player. Then, there is also the need to define groups of players, as the DSS uses that information in the social component of its reputation assessment process. The players are divided into four groups according to their average traded power amount.

After the analysis of the dataset, there is the need to select: (I) the period of time that the proposed methodology will start learning; (II) the negotiation period when the supported player will attempt to trade; (III) the five possible opponents, which the supported player may trade with; and (IV) the supported player. First, as the EM will consider the contexts Weekday and Weekend, a good starting point is the most recent Sunday in the dataset. This way, the proposed methodology can learn both contexts. It also allows the supported player to attempt to negotiate on the next Tuesday, in the Weekday context, as specified in this case study. The last day of the dataset (31 October 2008) is a Friday, therefore the chosen date is 26 October 2008, the previous Sunday. This way, the proposed methodology will be able to learn during 48 negotiation periods, evaluating more than 2000 contracts per context. Second, the negotiation period of the supported player is the first period of 28 October 2008, the following Tuesday. Third, the possible opponents selection is made by identifying the players with more contracts in the learning period and with power amounts ranging from 1 to 40 (amount of power that supported player wants to trade). Therefore, the following players are selected: Player 1 (Profile 2, Group 3); Player 2 (Profile 3, Group 4); Player 3 (Profile 1, Group 2); Player 4 (Profile 4, Group 4); and Player 5 (Profile 3, Group 3).

At last, the selected player to support is a buyer that established contracts with each one of the possible opponents in the learning period.

After the preparation of the case study, the DSS used the proposed methodology to learn what is the most probable scenario, during the selected period (26 and 27 October 2008). For that purpose, the proposed methodology is run with a Learning Rate of 0.3, allowing a slow learning, and Discount Factor of 0.8, favouring future rewards, considering the available amount of data. The final Q values for each scenario under each context are presented in Table 3.

Table 3. Final results (Q) of the learning process

Context	**ANN1**	**ANN2**	**ANN3**	**SVM**	**Average**	**Contracts**
Weekday	4,862	4,858	4,845	**4,883**	4,762	2 714
Weekend	**4,902**	4,876	4,713	4,801	4,585	2 671

As it can be observed in Table 3, the proposed methodology learned over 2 500 contracts for each context. At the end of the learning process, the most

AUTHOR'S VERSION

probable scenario in a Weekday is SVM and, in a Weekend, is ANN1. The results shows that, in both contexts, the ANN results improve with the reduction of the number of contracts considered in its training. The ANN1, which has the lower number of contracts (500), presents better results than ANN2 (1000) and even better than ANN3 (1500). However, ANN1 is slightly surpassed by SVM (1000) in the Weekday. On the other hand, the ANN1 and ANN2 are better than SVM in the Week, with an higher distance. The ANN proves to be a better algorithm overall, when considering the same amount of contracts, and even better with a lower number. The Average method presents the worst results in both contexts, representing the great uncertainty present in the real negotiation scenarios, where the players keep changing their behaviours. The Q values for each scenario under each context are very close, ranging from 4.585 to 4.902, which is caused by the converging nature of Q-Learning, over its iterations.

The Figure 2 and Figure 3 presents the normalized Q value of each scenario under Weekday and Weekend contexts respectively, over all the analysed contracts.

AUTHOR'S VERSION

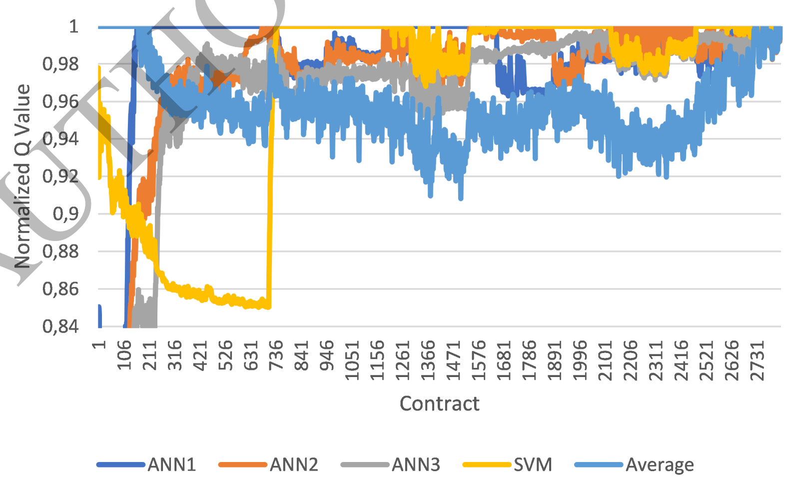

Fig. 2. The learning process for the Weekday context

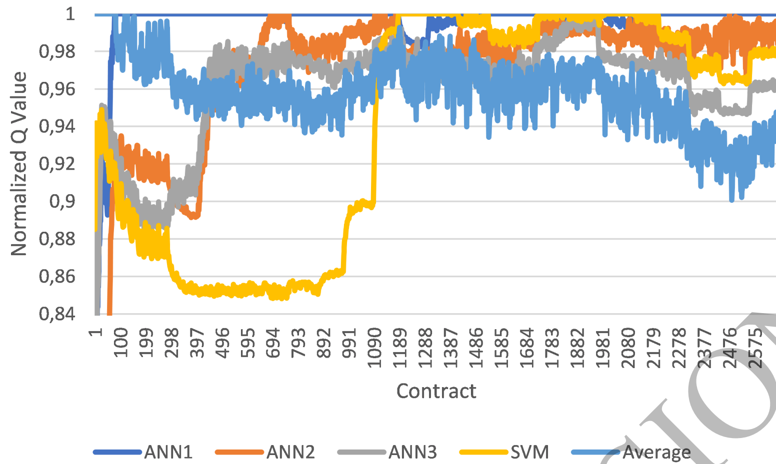

Fig. 3. The learning process for the Weekend context

AUTHORS' VERSION

The Figure 2 proves that SVM scenario is really dominant, being the scenario with the maximum Q value during more contracts. However, it is not always the most probable scenario. In fact, the SVM scenario were very far from reality in the first 708 contracts, the period in which ANN1 dominated $[102, 707]$, after the initial success of the Average scenario $[1, 101]$. Then, the SVM scenario is only surpassed by ANN1 $[1314, 1516]$, and ANN2 $[2122, 2425]$. As seen in Table 3, the success of each ANN method is measured by the amount of contracts considered. The fewer the number of contracts, the better the results. The Average scenario only had success in the beginning of the learning process, as it is a simple average, which does not requires much learning to know its potential, contrary to the other scenarios. Having seen the learning process of the Weekday context, it may be interesting to see how it compares to a different context, which in this case is the Weekend (Figure 3).

It is visible in Figure 3 that SVM scenario does not have as much success as the one presented in the Weekday context. Nonetheless it still presents good results, being the second most successful. The Table 3 shows that ANN1 finished the learning process as the most probable scenario. However, it was not just a momentary achievement since it was the most dominant scenario. The Average scenario only had success in the first 200 contracts, like has been verified in the Weekday context. The ANN2 only had a few successful periods but it is the most regular scenario, after the ANN1.

After the learning process, the DSS is ready to provide the supported player with the best action to take under the most probable scenario under Weekday context (SVM). The Table 4 presents the best action of SVM scenario as well as a comparison with the best action of the other scenarios.

Table 4. Best Action per Scenario

Scenario	**Player 1**	**Player 2**	**Player 3**	**Player 4**	**Player 5**	**Total Price**
ANN1	**1** 36,87	**37** 45,88	**1** 32,61		**1** 42,07	1809,08
ANN2	**38** 65,08		**1** 39,45		**1** 44,21	2556,80
ANN3	**1** 40,56		**39** 62,87			2492,46
SVM	**1** 40,60	**37** 49,03	**1** 35,34		**1** 44,88	**1935,06**
Average	**40** 29,98					1199,01

It is possible to observe in Table 4 that the most probable scenario (SVM) and the second most successful scenario (ANN1) identified the same best action, with a slight price variance. The ANN1 presented a less expensive action (about 9% less). However, the less expensive action was identified by the Average scenario while ANN2 and ANN3 present the most expensive ones. If the supported player selected the Optimistic decision method, instead of the Most Probable, the selected action would be the Average scenario one as it present the highest utility (which in this case is the highest profit, as all the possible opponents have similar reputations - about 0.62). On the other hand, if the Pessimistic decision method were selected, the best Action would be the one of ANN3 scenario, as it is the scenario with the lowest global utility (ANN3: 39 905.94; ANN2: 41 521.69; ANN1: 52 674.94; SVM: 60 689.93; Average: 73 904.44).

The supported player could take three different actions depending on the decision method but which one will be the closest to reality? Will SVM confirm itself as the most probable? After learning the real results of the negotiation period, for which supported player required decision support, an unexpected result is observed. The Average scenario presented the lowest average error per contract (Average: 9.31%; ANN1: 12.38%; ANN2: 16.55%; SVM: 18.30%; ANN3: 18.57%). Consequently, after learning from the real contracts, the ANN1 and Average scenarios surpassed the Q value of SVM scenario (ANN1: 4.920; Average: 4.784; SVM: 4.780; ANN2: 4.777; ANN3: 4.767). The improvement of the Average scenario shows that the participating players have been more regular, being closer to their average behaviour. The ANN1 is once again the leading scenario resultant of a slow, but constant, growing verified in the most recent negotiation periods.

This case study scenario could not detect the most probable scenario for the negotiation period in matter. However it could do it for most of the preceding negotiation periods. The reason is that the chosen negotiation period coincided with a turning point in the learning process. The ANN1 and Average scenarios were slowly improving while SVM was slowly decaying. As seen in this case

AUTHOR'S VERSION

study, there is always room for the unexpected to happen but that is exactly why the proposed methodology arose: to reduce the frequency of such situations as much as possible.

5 Conclusion

Nowadays, the EMs are constantly facing new challenges which result in constant changes and, consequently, an increased complexity for the involved entities. There is a growing need of tools that are capable to ease the experience of those entities, providing them with a better insight of what is going on as well as supporting them in their decisions. Various tools have emerged in the literature but they do not cover all of the current needs.

This paper identifies the lack of decision support for the pre-negotiation phase of bilateral contracts negotiation in the EMs. Although there are some tools, as analysed in this work, they do not address one of the key aspects of the pre-negotiation phase: the possible opponents analysis. For that purpose, a DSS has been developed which allows a good analysis of the possible opponents that the supported player may trade with. However, with the use of such tool, another problem arises: it is capable to generate various alternative negotiation scenarios under different contexts but it does not know which one is the most probable to occur in reality, under each context.

An adaptation of the Q-Learning algorithm is proposed to address that problem. Through its use, it is possible to learn through time, what is the most probable scenario to occur under a given context. The methodology confronts the generated scenarios with the real negotiation scenarios, being able to update their probability of occurrence.

By executing the proposed methodology in the presented case study, it can be concluded that it fulfils its purpose. By analysing real contracts under two different contexts (Weekday and Weekend), it was capable to determine which of the five scenarios had the highest probability of occurrence for each context, over the simulated period. The most probable scenario for each context kept changing over time and the scenarios had different results according to the context in which they were inserted. This way, it is possible to verify the importance of the presented methodology. There is not a scenario that will be always the most probable and it will also always depend on the context. The only way to guarantee a good selection is to keep learning through time, taking into account previous information without underestimating the new information.

AUTHOR'S VERSION

References

1. M. Shahidehpour, H. Yamin, Z. Li, and John Wiley & Sons., *Market operations in electric power systems : forecasting, scheduling, and risk management.* Institute of Electrical and Electronics Engineers, Wiley-Interscience, 2002.
2. R. Wüstenhagen and E. Menichetti, "Strategic choices for renewable energy investment: Conceptual framework and opportunities for further research," *Energy Policy*, vol. 40, pp. 1–10, 2012.

3. European Commission, "The 2020 climate and energy package," 2009.
4. P. Ringler, D. Keles, and W. Fichtner, "Agent-based modelling and simulation of smart electricity grids and markets - A literature review," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 205–215, 2016.
5. F. Lopes, M. Wooldridge, and A. Q. Novais, "Negotiation among autonomous computational agents: principles, analysis and challenges," *Artificial Intelligence Review*, vol. 29, no. 1, pp. 1–44, 2008.
6. A. Rahimi-Kian, B. Sadeghi, and R. Thomas, "Q-learning based supplier-agents for electricity markets," in *IEEE Power Engineering Society General Meeting, 2005*, pp. 2116–2123, IEEE.
7. F. Silva, B. Teixeira, T. Pinto, G. Santos, Z. Vale, and I. Praça, "Generation of realistic scenarios for multi-agent simulation of electricity markets," *Energy*, vol. 116, pp. 128–139, dec 2016.
8. G. B. Sheblé, *Computational Auction Mechanisms for Restructured Power Industry Operation*. Boston, MA: Springer US, 1999.
9. H. Algarvio and F. Lopes, *Risk Management and Bilateral Contracts in Multi-agent Electricity Markets*, pp. 297–308. Cham: Springer International Publishing, 2014.
10. D. S. Kirschen and G. Strbac, *Fundamentals of power system economics*. John Wiley & Sons, 2004.
11. L. Thompson, *Mind and Heart of the Negotiator, Second Edition, the*. Upper Saddle River, NJ, USA: Prentice Hall Press, second ed., 2000.
12. G. H. Snyder and P. Diesing, *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton University Press, 1977.
13. N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, M. Wooldridge, and C. Sierra, "Automated negotiation : prospects, methods and challenges," *Group Decision and Negotiation*, vol. 10, no. 2, pp. 199–215, 2001.
14. I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. Mcburney, S. Parsons, and L. Sonenberg, "Argumentation-based Negotiation," *Knowl. Eng. Rev.*, vol. 18, no. 4, pp. 343–375, 2003.
15. T. Veselka, G. Boyd, G. Conzelmann, V. Koritarov, C. Macal, M. North, B. Schoepfle, and P. Thimmapuram, "Simulating the Behavior of Electricity Markets with an Agent-based Methodology: the Electric Market Complex Adaptive Systems (emcas) Model," 2002.
16. V. Koritarov, "Real-world market representation with agents," *IEEE Power and Energy Magazine*, vol. 2, pp. 39–46, jul 2004.
17. R. Lin, S. Kraus, T. Baarslag, D. Tykhonov, K. Hindriks, and C. M. Jonker, "GENIUS: An Integrated Environment for Supporting the Design of Generic Automated Negotiators," *Computational Intelligence*, vol. 30, pp. 48–70, feb 2014.
18. F. Lopes, T. Rodrigues, and J. Sousa, "Negotiating Bilateral Contracts in a Multi-agent Electricity Market: A Case Study," in *2012 23rd International Workshop on Database and Expert Systems Applications*, pp. 326–330, 2012.
19. OMIE, "ejecucioncbfom." `http://www.omie.es/aplicaciones/datosftp/datosftp.jsp?path=/ejecucioncbfom/`, 2017. [Online; accessed 18-October-2017].
20. OMIE, "marginalpdbc." `http://www.omie.es/aplicaciones/datosftp/datosftp.jsp?path=/marginalpdbc/`, 2017. [Online; accessed 18-October-2017].

AUTHOR'S VERSION