Opponent Modeling in Automated Negotiation Applied to P2P Energy Trading

by

Lichen Xia

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday July 21, 2022 at 2:00 PM.

Student number:	5395704	
Project duration:	November 15, 2021 – July 21, 2022	
Thesis committee:	Prof. dr. C. M. Jonker,	TU Delft, thesis advisor
	Dr. L. C. Siebert,	TU Delft, supervisor
	R. Isufaj,	UAB, co-supervisor
	Dr. P. V. Barrios,	TU Delft
	Dr. T. Koca,	Independent Researcher

Contents

Abstract

Automated negotiation is a key form of interaction in systems composed of multiple autonomous agents with different preferences. Such interactions aim to reach agreements through an iterative process of making offers. With the growth of Peer-to-Peer (P2P) energy markets due to the development and deployment of a variety of small-scale electricity generation and storage devices (DERs), automated negotiation is seen as one of the advanced techniques that can improve the efficiency of energy distribution with the consideration of preferences of different entities. Opponent modeling is one of the essential abilities of automated negotiation agents that can further benefit automated negotiation. This project introduces a new opponent modeling technique considering the specific characteristics of P2P energy markets. These particular characteristics are *a*) Two automated negotiation agents can negotiate with each other many times, and *b*) The preferences of the users of agents are decided mainly by their energy consumption patterns, which usually do not have massive fluctuation across the year. The proposed opponent modeling method is developed from the idea of modeling the policy of a Reinforcement Learning agent. It uses a neural network to approximate the bidding strategy of the opposite automated negotiation agent. The network is learned based on the observations of offers exchanged in negotiations. With the learned network, the negotiation agent can predict the future actions of its opponent and make better decisions. We evaluated our opponent modeling with an existing automated negotiation system designed for off-grid energy trading. In experiments, the introduced opponent modeling always performs better than a random-guess model while modeling basic bidding strategies. Its performance is stable in dynamic environments where its opponent's preference and bidding strategy may change randomly. It is also proved that the introduced method has potential for further improvement with the help of advanced opponent modeling techniques, which model the preference profile of the opponent. With our new opponent modeling method, the automated negotiators who take part in the P2P energy markets should be able to find better joint agreements that are preferred by both itself and its opponents. And in this case, a better joint agreement means a more efficient way of distributing energy.

Preface

It has been a wonderful experience to conduct my master project thesis in the past nine months. I am grateful to all the people who helped and accompanied me during this thesis journey. Firstly, I would like to say thank you to my daily supervisor Dr.Luciano Cavalcante Siebert, who guided me into this exciting topic and constantly encouraged me throughout the whole thesis process. I also want to thank my thesis advisor, Prof. Catholijn Jonker, for providing me with valuable feedback at crucial points. And I also thank Ralvi Isufaj, who is my co-supervisor and gives me concrete advice. Second, I want to thank all my family and friends, who accompanied and encouraged me until this thesis journey's end.

Lichen Xia
Delft, July 2022

1

Introduction

With the integration of a variety of small-scale electricity generation and storage devices (DERs), such as photovoltaics panels and micro-wind turbines installations for commercial buildings and residential dwellings (Nair and Garimella, 2010), individual households can consume as well as produce energy. This kind of individual has the name which is prosumer (Chen, 2012). For an individual prosumer, the DERs may continue to generate energy when the energy demand has already been fulfilled. They may also stop generating energy when there is still a huge energy requirement, which inevitably causes waste and inefficiency. With the growth of prosumers who can participate in the local energy market (The number of prosumers have increased in the Netherlands by 200.000 in 2019 ([1])), cooperative sharing of the produced energy among neighborhoods (local energy market) is seen as an effective way of efficiently utilizing energy. However, there are two main challenges to effectively sharing energy. Firstly, people may have different preferences regarding the trading of energy, but they may not be willing to engage in competition and local trading themselves. Secondly, how to optimize the sharing of energy is still a big challenge. In centralized solutions to these challenges, mediators are at the center of local prosumers collecting information from participating prosumers and trying to optimize resource distribution. However, centralized solutions suffer from scaling problems as the number of participating prosumers increases rapidly. Besides, collecting personal data and preferences also raises privacy concerns. As a result, more and more research is starting to focus on the peer-to-peer (P2P) markets where prosumers can trade and transfer energy directly with each other, with the development of DER and smart metering technologies along with communications systems (Andoni et al., 2019, Andoni et al., 2017, Jogunola et al., 2018). The P2P market with decentralized management and collaborative principles allows for a bottom-up approach that would empower prosumers (Sousa et al., 2019). Among the advanced technologies applied on the P2P energy market (Paudel et al., 2020, M. R. Alam et al., 2017, Moret and Pinson, 2019), automated negotiation is one of the key technologies. For instance, Chakraborty et al., 2019 and Etukudor et al., 2019 propose their own bilateral automated negotiations applied in the field of P2P energy market.

1.1. Automated negotiation

Automated negotiation is a key form of interaction in systems composed of multiple autonomous agents with different preferences. The aim of such interactions is to reach agreements through an iterative process of making offers. The content of such proposals is a function of the strategy of the agents (Faratin et al., 2002).

1.1.1. Automated negotiation for P2P energy market

Negotiation technologies are seen as a key coordination mechanism for the interaction of providers and consumers in future electronic markets (Chakraborty et al., 2019). In the case of the local P2P energy market, each agent represents a prosumer or a group of prosumers and only knows its owner's preference such as how much energy is required. The final agreement is about how much energy will be transferred among prosumers. Since each agent only needs to care about the participants of the negotiation instead considering the whole local P2P energy market, the agent can have less pressure on computing and have the potential to apply intelligent algorithms. In addition, because an automated negotiation agent's goal is always to maximize its owner's achievement, entities' incentive of attending local energy market can also be improved by applying an automated negotiation system. In concern of the privacy problem, the data such as preference profiles and bidding strategies are not shared between negotiators. Therefore, to further improve the benefit brought by automated negotiation, it is essential for an agent to build the model of opposite negotiators.

1.1.2. Opponent modeling

Baarslag et al., 2014 divided a negotiation agent into three components: bidding strategy, acceptance strategy and opponent modeling (BOA). Different parts are responsible for different functions. Bidding strategy defines how an agent proposes offers at each round. Acceptance strategy decides what offer to accept and what offer to reject. And opponent modeling of a negotiation agent is defined as the ability to make a model of the opponent. It is an important component of a negotiation agent because efficient and effective negotiation requires the bidding agent to take the other's wishes and future behavior into account when deciding on a proposal (Baarslag et al., 2016). In general, an automated negotiation agent has two important components that can be modeled, namely strategy and preference profile. Strategy refers to the agent's bidding strategy and acceptance strategy, and preference decides the value of each offer for the agent. There have already been some strategy estimation (Mudgal and Vassileva, 2000, Masvoula et al., 2011) and preference estimation techniques (Tunali et al., 2017, Baarslag et al., 2013) with good performance. However, no opponent mod-

[1] https://www.uu.nl/en/news/nearly-200000-new-pv-prosumers-in-the-netherlands

eling method so far is typically designed for the automated negotiation systems applied to the P2P energy market. Consequently, their functions are limited and may even not fit in domain of P2P energy market. We will talk about the limitations of existing opponent modeling methods in detail in related works in chapter 2.

1.1.3. Modeling Reinforcement Learning agent's policy

While training Reinforcement Learning (RL) agents in a multi-agent environment, Lowe et al., 2020 find that RL agents can perform better if they know the policy of other RL agents interacting with them. Therefore, they propose a method that one RL agent can model the policies of other RL agents. We see the opportunity of transferring the idea of modeling RL agent's policy to the design of a new opponent modeling method considering the particular characteristics of the P2P energy market. We will make explicit assumptions based on the P2P energy market's specific characteristics in the section Problem statement. In chapter 2, we will introduce the concept of RL agents and the method of modeling RL agents. In chapter 3, we will explain how we convert the idea of modeling RL agent to opponent modeling in automated negotiation.

1.2. Problem statement

Opponent modeling have been added and evaluated into automated negotiation systems applied in different domains such as strategic video game (Afiouni and Ovrelid, 2013), cloud computing (Alsrheed et al., 2014) and supply chain management (Fang et al., 2008). However, to the best of our knowledge, no research has focused on opponent modeling methods of automated negotiation systems in the field of the P2P energy market. Besides, some special characteristics of the field of the P2P energy market can be utilized. Based on these features, we make assumptions needed for developing novel opponent modeling techniques.

- $A1$: In the case of agent-agent negotiation applied in the P2P energy market, two prosumers may trade energy over and over again, which means two automated negotiation agents may negotiate with each other more than once.

- $A2$: We assume in this project that the agent's bidding strategy is stable, which means an automated agent always uses similar bidding strategies in different negotiations. Furthermore, the preferences of the users of agents in the field of the P2P energy market are decided mainly by their energy consumption patterns, which usually do not change rapidly or dramatically during the year.

Therefore, there is a potential for an agent to build a model of an opponent's bidding strategy based on the bidding history of previous negotiations between the agent and the opponent and use and update the built model in future negotiations. So far, no bidding strategy modeling technique utilizes data from prior negotiations between the agent and the opponent instead of one negotiation. Additionally, an agent-agent negotiation can take over a hundred or a thousand rounds. For example, in their automated negotiation system, Chakraborty et al., 2019 set the negotiation deadline as 5000 rounds. The availability of a vast amount of data opens the door to deep learning methods.

1.3. Research question

To explore possible methods that can utilize the specific features of the domain of the P2P energy market discussed in the previous section, the research question of this work is:

- How can an automated negotiation agent model its opponent in the field of the P2P energy market.?

To answer this research question, we propose an opponent modeling technique based on the new assumptions of the field of the P2P energy market. To evaluate our proposed opponent modeling technique, we formulated three sub-questions:

1. Can our technique model different bidding strategies with good accuracy?

2. How stable is our opponent modeling method to opponent's preference profile changes?

3. Can our opponent modeling react to the changing of opponent's bidding strategy?

1.4. Contributions

This project introduces a new opponent modeling method, designed considering assumptions made in the case of agent-agent negotiation applied in the field of the P2P energy market:

1. The concept of the new opponent modeling method is developed from the idea of modeling the policy of Reinforcement Learning agents. With the new opponent modeling method, the agent can model the bidding strategy of the opponent by repeatedly negotiating with the opponent. With the opponent's bidding strategy model, the agent can predict the opponent's future actions, which gives the agent ability of what-if analysis, thus making better decisions.

2. To answer sub-questions, we applied our opponent modeling method to an automated negotiation system designed for off-grid P2P energy trading and evaluated the method with load profiles of typical off-grid energy consumers. Our new method has stable and sufficient performance in experiments while modeling basic time-dependent and behavior-dependent bidding strategies.

3. Our opponent modeling method's performance relies on the used opponent's preference model, and there is still a potential to improve the performance of our new opponent modeling method.

1.5. Outline

Chapter 2 will discuss related works about Automated negotiations applied in the domain of the P2P energy market and existing opponent modeling methods. Besides, the theoretical background needed to understand our approach and experiments will also be included. In chapter 3, the general setting of automated negotiation in our project will be introduced, and the design of the new opponent modeling will be presented in detail. Chapter 4 includes the design of the experiment and the results. Conclusions and future works will be given in chapter 5.

2

Related works and Theoretical background

In this chapter, we will study the related works from automated negotiation systems applied to the P2P energy market and opponent modeling techniques. For the opponent modeling, we will focus on existing preference estimation and bidding strategy estimation techniques and how these techniques compare and contribute to our new opponent modeling method. Then we will talk about the theoretical background of our project, including different bidding strategies of automated negotiation agents used in our experiments, the concept of deep reinforcement learning (DRL) agent, and the method of modeling the policy of DRL agents.

2.1. Related works

2.1.1. Automated negotiation applied in the domain of P2P energy market

In general, two kinds of negotiation protocols have been analyzed in the field of electronic markets, which are multilateral negotiation protocol and bilateral negotiation protocol. In multilateral protocol, agents negotiate with multiple agents at the same time. For example, M. Alam et al., 2015 presents a negotiation protocol for decentralized, concurrent negotiation over energy exchange between off-grid houses. This protocol has been proven to reduce battery charging and can be scaled to 100 houses. However, the multilateral negotiation protocol still suffers from the problem of complexity in designing agents since each agent needs to make offers to all potential negotiating partners simultaneously. This is why bilateral negotiation protocol has become a focus of many studies.

Chakraborty et al., 2018 present a bilateral negotiation protocol. This protocol focuses on settling energy contracts among prosumers considering heterogeneous prosumer preferences. Each offer has two issues in this protocol: the volume of energy to be transferred and the time to pay back the transferred energy. This protocol has been evaluated over real residential demand, generation, and storage data and proved that it can increase system efficiency and fairness. The same protocol has been improved by adding a reinforcement learning method to help select negotiation partners (Chakraborty et al., 2019).

Etukudor et al., 2019 introduce another bilateral negotiation framework that has a different structure of offers. Besides, the agents in their framework are implemented with three different negotiation strategies (Zero Intelligence Strategy, Linear Heuristic Strategy, and Expert Agent Strategy). By case study with Community-scale Energy Demand Reduction in India [1]. It is demonstrated that this framework allows prosumers to increase their revenue while providing electricity access to the community at a low cost. Also, the Linear Heuristic Strategy and Expert Agent Strategy perform better than the Zero Intelligence Strategy. In another similar study, bilateral negotiation heuristics applied to a low-income, off-grid, community P2P energy market is proposed (Etukudor et al., 2020). Five negotiation strategies are implemented and compared. The result shows that local P2P markets using negotiation strategies such as the Boulware strategy are solutions to bridge the electricity-deficit gap effectively. Although the frameworks mentioned above have already been implemented with some negotiation strategies and achieved good results in the field of the P2P energy market, no agent in these frameworks can model opponents. However, efficient and effective negotiation requires the bidding agent to consider the other's wishes and future behavior when deciding on a proposal (Baarslag et al., 2016). In the P2P energy market field, a negotiator may encounter opponents with different preferences and strategies, and no single bidding strategy performs the best against all opponents. Therefore, models of the opponent are needed for negotiators to make better decisions at each round of the negotiation. We believe that the agents in these frameworks can be further improved by proper opponent modeling methods, thus making more efficient energy distribution.

2.1.2. Opponent modeling

Opponent modeling of a negotiation agent is the ability to make a model of the opponent. A good opponent modeling can improve the benefit of automated negotiation, including but not limited to achieving a win-win outcome and avoiding failure of the negotiation. Baarslag et al., 2012 prove that proper opponent modeling techniques can result in significant gains in both the time-based and round-based negotiation protocols. Besides, opponent modeling can help agents find fairer agreements without sacrificing any agent. For example, Sanchez-Anguix et al., 2021 introduce a social agent relying the opponent modeling. Experiments prove that this social agent can achieve better performance in terms of individual utility and social fairness. Automated negotiation agents capable of modeling opponents have already been applied in different fields. Afiouni and Ovrelid, 2013 implements a multi-issue negotiation system for the strategic video game Civilization IV and focuses on improving negotiation results using opponent modeling. Alsrheed et al., 2014 proposes an automated negotiation system for providers and consumers in the field of cloud computing. They evaluate the

[1] (www.cedri.hw.ac.uk/)

system with agents that can model the opponent and achieves a good performance. Fang et al., 2008 improves the automated negotiation agent applied in supply chain management by equipping the agent with the ability to retrieve the opponent's knowledge. However, no opponent modeling method so far has been designed and applied to the domain of the P2P energy market. Therefore, new opponent methods need to be introduced to automated negotiations in the P2P energy market.
There has already been a method to model the policy of other RL agents in a multi-agent scenario (Lowe et al., 2020). However, such methods need continuous interactions between agents, which is usually not the case for automated negotiations. But in the P2P energy market field, we see opportunities to use such methods to model the bidding strategy of the opponent thanks to assumptions $A1$ and $A2$. Besides, preference estimation techniques are needed to compensate for the incomplete information about the opponent's preference profile to clearly describe the interactions between the negotiation agent and its opponent. Therefore, this project focuses on developing a new bidding strategy estimation technique by utilizing existing preference estimation technique.
Some representative bidding strategy estimation techniques have been introduced before our work. Mudgal and Vassileva, 2000, Hou, 2004 and Brzostowski and Kowalczyk, 2006 apply regression analysis to model the bidding strategy of opponent negotiator. However, their methods need knowledge of the modeled bidding strategy type in advance. Masvoula et al., 2011 trains Multi-layer Perceptrons (MLPs), which predict the next offer from the opponent, with counter offers proposed by the opponent, and no assumption of the modeled bidding strategy is needed. In this method, only the offers received from the opponent are considered, but a bidding strategy can depend on other factors such as offers received by the opponent and the number of negotiation rounds that have passed. Besides, this method does not assume that a negotiator can negotiate with one opponent many times. Therefore, this method only utilizes the history of the current negotiation instead of all previous negotiations. Our opponent modeling method models the opponent based on data from past negotiations between the agent and its opponent. Additionally, since the MLPs directly use offers as input and output in the design from Masvoula et al., 2011, their method also does not consider the case that the opponent's preference can be different in different negotiations. Although there are many limitations, this method is one of the inspirations for our new opponent modeling method. Our opponent modeling method also models the opponent by training an MLP. However, instead of using offers as input and output, our method takes utilities of offers and other related variables as inputs and predicts how the utility of the opponent's offer will change. In this way, our method can handle changing preference profiles with the help of existing preference estimation techniques. In the P2P energy market, the preference of each prosumer largely depends on their energy consumption pattern, which can slightly change through the days or months. And our method's ability to handle changing preference reduce the possible negative influence on modeling accuracy due to the changing consumption pattern. To efficiently handle changing preferences of opponents, the choice of preference estimation method is important.
Baarslag et al., 2013 has evaluated and compared three preference estimation methods. These three methods are Bayesian models, Frequency models, and Value models. Among them, Frequency models and Value models have a better and more robust performance in general, and CUHK value mode and Smith Frequency model are the best Value model and Frequency model, respectively (Baarslag et al., 2013). Since Value models assume equal issue weights, which is not usually the case in the field of the P2P energy market, we choose the Smith Frequency model as the help function in our new opponent modeling method.

2.2. Theoretical background

2.2.1. Bidding strategies of automated negotiation agents

There are two categories of bidding strategies: time-dependent and behavior-dependent (Faratin et al., 1998). Generally, an automated negotiation agent changes its target utility U_{target} at each round based on either the negotiation round or the behavior of the opponent negotiator, or both. To propose an offer, the agent first finds all offers with a utility higher than the target utility U_{target} and randomly chooses one offer from those found offers as a counteroffer for the next negotiation round.

1. $time-dependent\ strategy$. The negotiator with a time-dependent strategy concedes with time. At each round r, the target utility U_{target} is defined as:

$$U_{target} = U_{rev} + F(r)(U_{max} - U_{rev}) \tag{2.1}$$

where U_{rev} is the reservation utility of the negotiator and U_{max} is the maximum utility the negotiator

can obtain. The $F(r)$ is defined as:

$$F(r) = 1 - r^{\frac{1}{e}} \tag{2.2}$$

where e is the concession rate.

2. $behaviour - dependent\ strategy$. The negotiator with a behavior-dependent strategy bases its action on its opponent's action. The naive Tit-For-Tat(TFT) strategy (Mohammad et al., 2020) is one representative of the behavior-dependent strategy family, At each round r, the target utility U_{target} of a naive-TFT negotiator is defined as:

$$U_{target} = min(U_{max} - max(U_{max} - U_{r-1}(w_{oppo}), 0), U_{rev}) \tag{2.3}$$

where $U_{r-1}(w_{oppo})$ is the utility of the offer received from the negotiator's opponent at the previous round $r-1$. A naive-TFT negotiator always calculates the utility of offers with its own preference profile.

2.2.2. Reinforcement Learning

Reinforcement Learning (RL) is learning what to do—how to map situations to actions—so as to maximize a numerical reward signal. It is about training a learning agent that is able to sense the state of its environment to some extent, take actions that affect the state and have a goal or goals relating to the state of the environment (Sutton and Barto, 2018). Generally, a RL agent connects to an environment via perception and action. At each step of interaction between the RL agent and the environment, the agent observes the state of the environment and chooses an action to conduct. The conducted action changes the state of the environment, and the environment sends a reward to the agent. The RL agent learns a policy $\pi(A|O)$ to maximize the long-run sum of received rewards by systematic trial and error. The policy $\pi(A|O)$ takes an observation O as input and outputs an action A. The policy of an RL agent can be represented by a tabular or a complex neural network. The RL agents whose policy is approximated by neural networks are called Deep Reinforcement Learning (DRL) agents.

2.2.3. DRL agents and opponent's strategy modeling

Lowe et al., 2020 trains approximate policy networks to model the policy of DRL agents in a multi-agent environment. The approximate policy network is trained online with the collection of previous observations and actions of the DRL agent to be modeled. With the help of approximate policy networks, a DRL agent can predict the actions of other DRL agents in the environment and learns to make better decisions. This work forms the base of our opponent modeling method. Our method views an intelligent automated negotiation agent as a DRL agent with a good and stable policy and trains an approximate policy network to model the policy of the automated negotiation agent. A considerable amount(more than 100 epochs if each epoch consists of around 100 steps for one particular agent) of collections of observations and actions are needed to train the approximate policy network. According to the assumption $A1$, acquiring the required amount of data becomes much softer in the field of the P2P energy market since two agents can negotiate with each other many times, and one negotiation may consist of hundreds or thousands of rounds.

3

Approach

This chapter firstly discusses the general setting of automated negotiation used throughout this project. Then, the detailed design of our new opponent modeling method is introduced, and the main challenge of realizing our method is discussed. We propose two solutions to solve the main challenge of our work. The advantages and disadvantages of each solution will be discussed in detail.

3.1. Setting of automated negotiation

The negotiation we are concerned about within this project is the automated bilateral negotiation, where two automated agents negotiate with each other during a negotiation. The setting of an automated negotiation consists of negotiation protocol, negotiators and the negotiation scenario (Baarslag, 2014). The negotiation protocol defines how two negotiators interact with each other. The protocol used in this project is the stacked alternating offers protocol (Aydoğan et al., 2017), where one negotiation session consists of rounds of consecutive turns. At every turn, each negotiators can choose to propose the next offer, accept the offer from the opponent, or end the negotiation. The negotiation ends if a joint agreement is found by two negotiators, the deadline is reached, or one negotiator decides to end the negotiation. The deadline is defined as the maximum number of rounds a negotiation session can last. The negotiation scenario contains the negotiation domain and the preference profile of each agent. The negotiation domain contains one or more issues. To propose an offer, the value of each issue should be set. The outcome space of the negotiation domain is defined as $\Omega = \{w_1, w_2, ...w_n\}$ where w_n is a possible offer and n is the number of possible offers in this domain. In this project, the negotiation domain is limited to the P2P energy market. It is assumed that the negotiations can happen between each pair of negotiators repeatedly, in addition to what the stacked alternating offers protocol usually assumes. The preference profile of each negotiator is defined as the utility function $U(w)$. Each negotiator only knows its own utility function. Besides, each negotiator has its reservation utility U_{rev}, which is the utility they will get if the negotiation ends with no joint agreement.

3.1.1. The structure of the negotiator

In Fig. 3.1, the structure of the automated negotiator used in this project is presented. The negotiator is doing the ith negotiation (assumption $A1$) with the same opponent who has a stable bidding strategy and preference profile (assumption $A2$). The negotiator is based on a BOA agent (Baarslag et al., 2014), which consists of three components: bidding strategy (B), opponent model (O) and acceptance strategy (A). In this project, we mainly focus on developing a new method for building a bidding strategy model (highlighted in red in figure 3.1) of the opponent with the help of a preference profile model built by existing preference estimation methods. The bidding strategy estimation method, which is the core part of our opponent modeling method, learns a MLP that approximates the bidding strategy of the opponent based on the bidding history of previous negotiations between the negotiator and the opponent, such as the offers received from and by the opponent. For each opponent, one model is learned. For better decisions, information such as predicting the opponent's next actions can be extracted from the learned model and utilized by the negotiator.

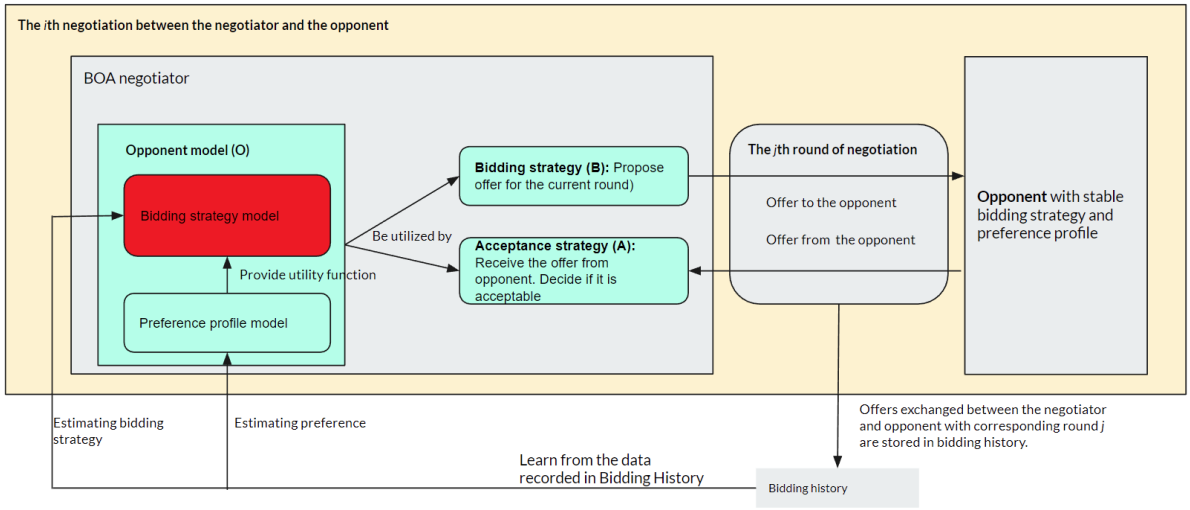


Figure 3.1: Structure of the negotiator in this project

3.2. Design of the new opponent modeling technique

Our opponent modeling method consists of bidding strategy estimation and preference estimation, and building a bidding strategy model of the opponent is the core part of our method.
Lowe et al., 2020 propose a method that can infer the policy network of other RL agents in a multi-agent cooperative and competitive environment. If the data of its previous observations and corresponding actions are available for a specific agent, its policy network can be approximated. This method can also be transferred as an opponent modeling method in automated negotiation in the domain of the P2P energy market.
Firstly, an intelligent automated negotiator makes decisions at every round of negotiation based on the previous interactions such as proposed and received offers between it and its opponent. Previous interactions are the observation of the negotiator, and its decisions can be viewed as actions it chooses to take based on its observation. The bidding strategy of the negotiator defines how the observations are transferred to specific actions, which is similar to the policy of an RL agent. An RL agent's policy decides which actions to take based on its observations. Therefore, we can assume that each negotiator has a policy network that represents its bidding strategy and then uses a MLP to approximate it. However, in a multi-agent environment, the observations and actions of each RL agent are usually available to other RL agents, which is not the case in negotiation since the negotiator do not know the preference of their opponents. Therefore, we use preference estimations to help the negotiator to describe their opponent's observations and actions. We will talk about this in detail in the following sub-sections. Secondly, in the domain of the P2P energy market, one negotiator can negotiate with the same opponent more than hundreds of times, and each negotiation can consist of more than hundreds of rounds (assumption $A1$). With more and more negotiation happening between the negotiator and the opponent, the increased collection of the opponent's observations and actions should improve the accuracy of inferring the opponent's policy.

3.2.1. Structure

Fig. 3.2 presents the process of our opponent modeling method modeling an opponent's bidding strategy. It is assumed that the opponent to be modeled has a Target Policy network P_{target} which has observations as input and next actions as output (demonstrated at the top of the figure). This policy network represents the opponent's bidding strategy. During negotiation, the opponent's observations and corresponding actions are collected, and our method trains an Approximate Policy network (P_{appro}) with collected data by minimizing the loss function (3.1). P_{appro} takes the opponent's observations as input and predicts the opponent's subsequent actions (demonstrated at the bottom of the figure). For each opponent, there is a separate P_{appro} to model their bidding strategy.
The loss function to be minimized while training P_{appro} is defined as:

$$L(P_{appro}) = L_{CE}(P_{appro}, P_{target}) + (-\lambda H(P_{appro})) \tag{3.1}$$

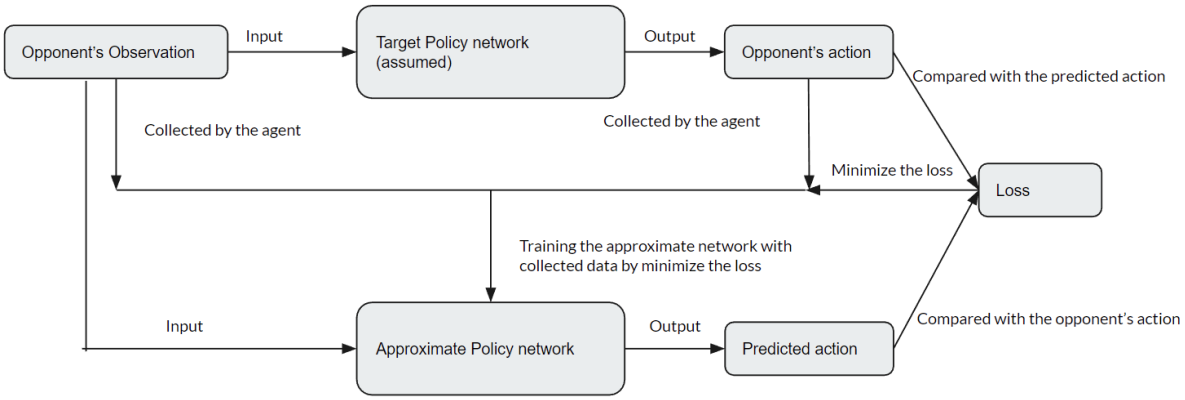


Figure 3.2: Opponent Modeling Structure

where $L_{CE}(P_{appro}, P_{target})$ is the cross-entropy loss between P_{target} and P_{appro}, and $H(P_{appro})$ is the entropy regularizer. Since the P_{appro} is trained with the collection of the opponent's observations and actions, clearly and accurately describing the opponent's observations and actions is essential to building a good model of the opponent's bidding strategy.

3.2.2. Observation and action space

The core of modeling the opponent policy is the design of the observation and action space of the opponent. In each round r of negotiation, the observation o_r is represented as:

$$o_r = \{U_o(w_{prop}^{r-1}), U_o(w_{rec}^{r-1}), r\} \tag{3.2}$$

where U_o is the utility function of the opponent, and w_{prop}^{r-1} and w_{rec}^{r-1} are the offers proposed and received by the opponent in the last round $r-1$ respectively. The policy network (bidding strategy) P_{target} of the opponent takes O_r as input returns a discrete action a_r (3.3). There are three discrete actions: the opponent can choose to propose an offer with a utility higher/lower/equal to the utility of the most recent proposed offer.

$$P_{target}(O_r) = a_r \tag{3.3}$$

To clearly explain our design. The negotiator equipped with our opponent modeling method is referred to as neg_o in the remaining part of the section.
Since it is assumed that the neg_o only knows its own utility function (U_s) and has no knowledge of the opponent's utility function (U_o), in practice, our method can only use the estimated opponent's utility function $U_{o'}$ as the replacement of U_o. In this case, the observation o_r is redefined as:

$$o_r = \{U_{o'}(w_{prop}^{r-1}), U_{o'}(w_{rec}^{r-1}), r\} \tag{3.4}$$

The actions observed by the neg_o are also different from the actions done by the opponent due to using of the estimated opponent's utility function $U_{o'}$. To clearly divide these two kinds of actions in this report, we use two terminologies:

1. **Relative action** a_r^{rel}: action observed by the neg_o and used in online training. It is calculated by estimated opponent's utility function $U_{o'}$.

2. **Absolute action** a_r^{abs}: action done by the opponent. It is calculated by U_o which is not available to the neg_o.

To estimate the U_o, we came up with two options:

1. The first option is using U_s as $U_{o'}$ all the time, which means that the neg_o looks at the world only in the view of its own utility function. On the one hand, since U_s is always known to the neg_o, the learned P_{appro} should be stable if U_s is stable. On the other hand, the neg_o directly ignores the opponent's preference, and the changes in preferences may affect how the neg_o estimates the observations and

actions of the opponent. Therefore, the learned P_{appro} should react to changes in the opponent's preference. In other words, the P_{appro} may need to be relearned once the preference of one side of the negotiation changes. The relation between relative actions a_r^{rel} and absolute actions a_r^{abs} is complex and unpredictable for the first option. This option is concise, and its performance only relates to how well the P_{appro} is trained. Therefore, we use this option as the first and base option.

2. In automated negotiation, one efficient way of estimating the opponent's utility function during negotiation is using existing preference estimation techniques, and frequency opponent modeling is one of the most successful preference estimation techniques. Therefore, the second option uses the modeled utility function from frequency opponent modeling as $U_{o'}$. However, the frequency of opponent modeling itself is not always stable or accurate, which brings extra complexity and problems of accuracy and stability of learned P_{appro}. With the estimated opponent's utility function $U_{o'}$, the neg_o can calculate the relative actions a_r^{rel} done by the opponent, but the opponent's absolute actions a_r^{abs} are not available during the negotiation. Suppose the frequency opponent modeling is stable and accurate enough, in that case, the relative actions a_r^{rel} will be close to the absolute actions a_r^{abs}, and our model can predict the future absolute actions a_r^{abs} of the opponent with high accuracy. It is worth mentioning that the frequency of opponent modeling itself initializes after every negotiation because we want the second option to work in a general setting.

In the following part of this report, we refer to the first option as $option_{noF}$, which means the first option model the opponent's bidding strategy without the help of frequency opponent modeling. And we use $option_F$ to represent the second option which is the option using the frequency opponent modeling.

4

Experiments and Discussions

4.1. Automated negotiation system for off-grid P2P energy trading

To evaluate the performance of our opponent modeling method, we applied our method to a bilateral automated negotiation system proposed by Etukudor et al., 2020. The system is designed for off-grid P2P energy markets. In the system, a negotiator, on behalf of a seller, negotiates with a buyer who another negotiator represents. The seller and buyer negotiate with each other for the quantity and price of the energy to be traded the next day. Fig. 4.1 gives an overview of the system. In experiments, the seller negotiator equips with our method to model the buyer negotiator's bidding strategy. We include essential formulas from the original paper in this report to illustrate the seller and buyer agents in the system and the whole set-up of experiments. Some formulas are modified to make the agents compatible with different bidding strategies. For that modified formulas, we will explain what changes have been made to them.

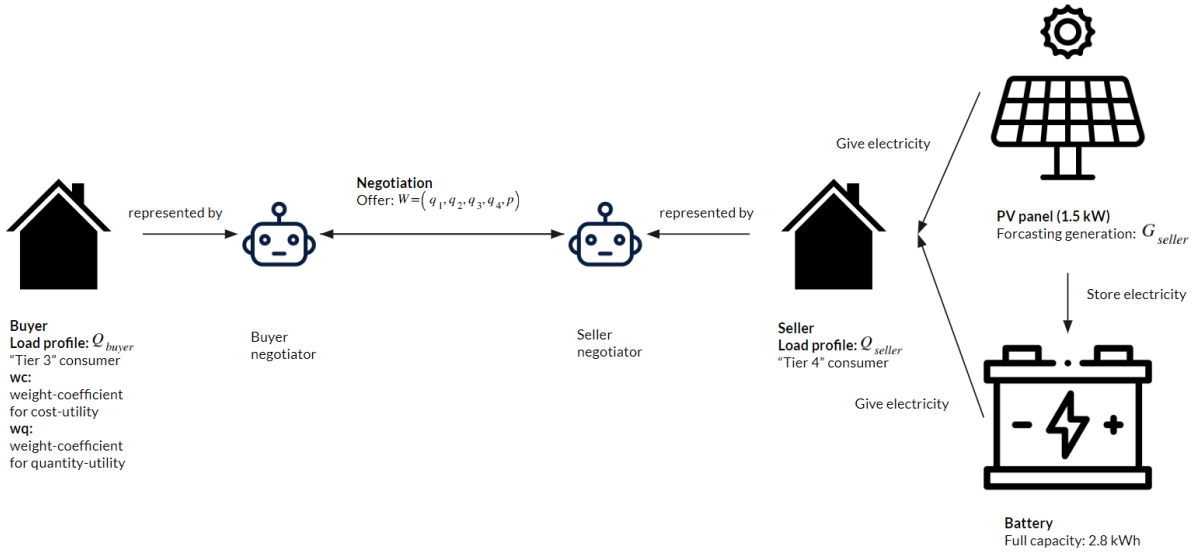


Figure 4.1: Automated negotiation system for off-grid P2P energy markets based on Etukudor et al., 2020

4.1.1. Domain

A day is divided into four sectors, which are night (0:00-6:00), morning (6:00-12:00), afternoon (12:00-18:00) and evening (18:00-24:00). Therefore, an offer W is defined as:

$$W = (q_1, q_2, q_3, q_4, p) \tag{4.1}$$

where q_1, q_2, q_3 and q_4 are the quantities of energy to be transferred at night, morning, afternoon and evening, respectively. p is the price of each unit of energy.

4.1.2. Seller negotiator

For a seller, if the offer $W = (q_1, q_2, q_3, q_4, p)$ is feasible, the utility $U(W)$ of the offer will be as same as the revenue utility $R(W)$ after conducting the offer. Otherwise, the utility $U(W)$ will be set to -1, and the offer will be directly ignored during the negotiation. An offer W is feasible if the seller has enough energy to conduct the offer in all four sectors. The algorithm for checking infeasible offers is presented in appendix A. The revenue utility $R(W)$ is defined as:

$$R(W) = \frac{\sum_{i=1}^{4} q_i p_i - \sum_{i=1}^{4} q_i MC_i}{maxR} \tag{4.2}$$

where MC_i is the marginal cost of generating energy in each sector, and $maxR$ is the maximum revenue the seller can expect from all possible offers. To limit the $R(W)$ between 0 and 1 (the utility of the offer most preferred by the seller exactly equals 1) our definition of $maxR$ is different from the original paper:

$$maxR = \max_{W} \sum_{i=1}^{4} q_i p_i - \sum_{i=1}^{4} q_i MC_i \tag{4.3}$$

The reservation utility of the seller U_{rev}^{seller} is defined as:

$$U_{rev}^{seller} = U((0,0,0,0,p^T)) \tag{4.4}$$

where $(0,0,0,0,p^T)$ is an offer which means no energy will be traded at all, and p^T is the minimally acceptable price for the seller.

4.1.3. Seller profile

The load profile of a seller is defined as $Q_{seller} = (q_1^{required}, q_2^{required}, q_3^{required}, q_4^{required})$, which represents the quantities of energy the seller needs to consume itself during different sectors of one day. The World Bank has categorized consumers into different tiers based on their electricity needs (Bhatia and Angelou, 2015). In our experiment, we assume that the seller is a "Tier 4" consumer and use load profiles constructed for a representative "Tier 4" off-grid household. The data of load profiles are from the work of Narayan et al., 2020. Following the case study in Etukudor et al., 2020, we also assume that a seller owns a small solar PV system of 1.5 kW with a battery of 2.8 kWh of available capacity. The battery is assumed to be full at the start of the experiment, and $G_{seller} = (0,2,2.5,0)kWh$ is the forecasting of the energy generated by the solar PV system per day.

4.1.4. Buyer negotiator

For a buyer, the utility $U(W)$ of an offer $W = (q_1, q_2, q_3, q_4, p)$ is defined as:

$$U(W) = w_c * C(W) + w_q * Q(W) \tag{4.5}$$

where w_c and w_q are weight coefficients and add up to 1. w_c is the importance of the cost of an offer for the buyer, and w_q is the importance of quantities of traded energy in an offer. $C(W)$ represents the utility of the overall cost of the offer w. It is defined as:

$$C(W) = \frac{\sum_{i=1}^{4} p^{max} q_i^{max} + \sum_{i=1}^{4} q_i^{required} p^{max} - \sum_{i=1}^{4} q_i p_i}{\sum_{i=1}^{4} p^{max} q_i^{max} + \sum_{i=1}^{4} q_i^{required} p^{max} - \sum_{i=1}^{4} p^{min} q_i^{min}} \tag{4.6}$$

where $q_i^{required}$ is the quantity of the energy the buyer requires in each sector of the day, p^{max} and q_i^{max} are the highest possible price and maximum energy quantity that can be set in an offer, while p^{min} and q_i^{min} are the lowest possible price and minimum quantity of energy that can be traded. With this definition, an offer's cost-utility is always higher than another offer with the same amount of energy to trade but a higher price. To limit the $C(W)$ between 0 and 1 (the highest $C(W)$ exactly equals 1), we add $\sum_{i=1}^{4} p^{max} q_i^{max}$ to both numerator and denominator in our definition of $C(W)$.
Additionally, $Q(W)$ is the buyer's utility for the quantities of energy according to the offer W, and it is defined as:

$$Q(W) = \sum_{i=1}^{4} m_i w_i \tag{4.7}$$

where m_i represents the matching between the buyer's required energy quantity and traded energy quantity for sector i according to the offer, it is defined as:

$$m_i = \begin{cases} \frac{min(q_i, q_i^{required})}{q_i^{required}} & q_i <= q_i^{required} + \phi_{buyer} \\ 0 & q_i > q_i^{required} + \phi_{buyer} \end{cases} \tag{4.8}$$

where ϕ_{buyer} is the flexibility the buyer has for overconsumption.
w_i represents the importance of each sector i for a buyer, and it is defined as:

$$w_i = \frac{\frac{q_i^{required}}{\max_i q_i^{required}}}{\sum_{i=1}^{4} w_i} \tag{4.9}$$

where $\max_i q_i^{required}$ is the maximum energy quantity required by the buyer among four sectors of the day, and $\sum_{i=1}^{4} w_i = 1$. The reservation utility of the buyer U_{rev}^{buyer} is defined as:

$$U_{rev}^{buyer} = min(U((0,0,0,0,p^T)), U((q_1^{rq}, q_2^{rq}, q_3^{rq}, q_4^{rq}, p^T))) \tag{4.10}$$

where $(q_1^{rq}, q_2^{rq}, q_3^{rq}, q_4^{rq}, p^T)$ is the offer that the energy requirement of the buyer is perfectly satisfied, and p^T here is the maximum acceptable price for the buyer. Therefore, the value of U_{rev}^{buyer} depends on the buyer's profile.

4.1.5. Buyer profile

The load profile of a buyer is defined as $Q_{buyer} = (q_1^{required}, q_2^{required}, q_3^{required}, q_4^{required})$. In our experiment, we assume the buyer is on behalf of a "Tier 3" consumer and use load profiles constructed for a representative "Tier 3" off-grid household.

4.2. Pool of opponents

There are two categories of bidding strategies: time-dependent and behavior-dependent (Faratin et al., 1998). To evaluate the generality of our modeling methods, we test our methods on a diverse pool of opponents, including negotiators with bidding strategies from both categories. In our experiments, there are three time-dependent strategies, which are Boulware, Linear and Conceder, with concession rates e equal to 0.3, 1, 3 respectively. For behavior-dependent strategy, we use a naive Tit-For-Tat(TFT) strategy (Mohammad et al., 2020) as the representative.

4.3. Metrics

A good opponent model should predict the opponent's following action accurately after being trained with data from bidding histories. Therefore, we use accuracy

$$acc = \frac{num\_correct\_predictions}{num\_predictions} \tag{4.11}$$

to evaluate the performance of our opponent modeling method.
There are two kinds of accuracy used as metrics in our experiments:

1. The first accuracy acc_{rel} is computed with action predicted by our model a_{pred} and relative actions a_r^{rel}. The opponent model of our agents is trained online with relative actions a_r^{rel}. So the first accuracy acc_{rel} aims at checking whether the opponent model can learn from past bidding histories.

2. The second accuracy acc_{abs} is computed with predicted actions a_{pred} and absolute actions a_r^{abs} because, for $option_F$, we are also interested in the accuracy of predicting the opponent's absolute action a_r^{abs} while training the approximate network with the help of estimated preference profile from frequency opponent modeling.

4.4. Experimental set-up

We evaluate our opponent modeling method in four different settings to assess the opponent modeling method and answer the research questions. In each experiment, a seller negotiator with our opponent modeling method models the strategy of its opponent, which is a buyer negotiator. Table 4.1 presents the parameters of experiments and bilateral automated negotiation. One experiment consists of 600 negotiations, which means the seller negotiator with our opponent modeling method negotiates with one single opponent 600 times during one experiment. We set the number of negotiations to 600 per experiment because we want to make sure our method has sufficient amount data to model the opponent and find out what will happen after a model is made by our method. Besides, 600 also gives us enough room to manipulate the buyer negotiator's bidding strategy and preference profile during experiments. The $deadline$ of one negotiation is 200 rounds. The seller and buyer negotiator's bidding strategy and preference profile are fixed during one negotiation. However, the bidding strategy and preference of the buyer negotiator may change during one experiment based on different settings. The load profile of the seller is presented in appendix A. The hyper-parameters of training P_{appro} are presented in Table 4.2. It is worth noting that the batch size is equal to the $deadline$ of

one negotiation, which means our method updates the learned model at the end of each negotiation during experiments. The parameters presented in the tables are constant in all four settings unless mentioned explicitly.
Table 4.3 compares four different settings of experiments, and we will discuss the details in the following subsections.

Experiment settings	value
Negotiations per experiment	600
Deadline per negotiation	200 rounds
Bidding strategy of the seller negotiator	Linear
Q_{seller}	(0.18, 1.07, 1.10, 0.95)kWh
G_{seller}	(0, 2, 2.5, 0)kWh
U_{rev}^{seller}	0
ϕ_{buyer}	0.05kWh

Table 4.1: Parameters for experiments.

Hyper-parameters	value
Learning rate $option_{noF}$	0.003
Learning rate $option_F$	0.006
λ	0.003
Layers of P_{appro}	4
Size of hidden layers	64
Batch size	200

Table 4.2: hyper-parameters for training P_{appro}.

Settings	Bidding strategy our negotiator	Bidding strategy opponent negotiator	Preference our negotiator	Preference opponent negotiator
Setting 1	fixed	fixed	fixed	fixed
Setting 2	fixed	fixed	fixed	randomly changes every negotiation
Setting 3	fixed	randomly changes every 100 negotiations	fixed	fixed
Setting 4	fixed	randomly changes every 100 negotiations	fixed	randomly changes every negotiation

Table 4.3: Comparison between four settings .

4.4.1. Setting 1: Fixed strategy and preference profiles

The goal of setting 1 is to know whether our method can model the bidding strategy of time-dependent negotiators and behavior-dependent agents. Besides, it is also essential to evaluate and compare the overall performance of $option_{noF}$ and $option_F$. For setting 1, we conduct four experiments. In each experiment, the seller negotiator models the buyer negotiator which has one of Boulware, Linear, Conceder and naive-TFT as its bidding strategy. During the experiment, the preference and bidding strategy of the seller negotiator

and the buyer negotiator are fixed. The load profile of the buyer negotiator is set to $(0.15, 0.3, 0.17, 0.32)kWh$, which is a profile of a representative "Tier 3" consumer, and the weight coefficients w_c and w_q are both set to 0.5. The detailed load profile of the buyer is presented in appendix A.

4.4.2. Setting 2: Fixed strategy and varying preference profile

As discussed in section 3.2.3, we expected that the performance of the $option_{noF}$ should be influenced negatively by the changing preference of the opponent. Therefore, for setting 2, it is assumed that the preference of the buyer negotiator changes randomly after every negotiation. In contrast, the bidding strategy of both the seller and buyer negotiators is fixed during each experiment. To randomly change the preference of the buyer negotiator, we randomly draw a load profile from a set of load profiles. The set contains the load profiles of a "Tier 3" consumer across one year, so there are in total 365 load profiles in this set. Besides, the buyer negotiator's weight coefficient w_c also changes randomly between 0.1 and 0.9 per negotiation, and $w_q = 1 - w_c$

4.4.3. Setting 3: Varying strategy and fixed preference profile

To find out if our opponent modeling method can react to the changing of the opponent's bidding strategy quickly, we conduct experiments in setting 3. It is assumed that during each experiment, the bidding strategy of the buyer negotiator changes randomly among Boulware, Linear, Conceder and naive-TFT every 100 negotiations. The preference of both the seller and buyer negotiators is fixed during each experiment. In this setting, the load profile of the buyer negotiator is set to $(0.15, 0.3, 0.17, 0.32)kWh$, and the weight coefficients w_c and w_q are both set to 0.5.

4.4.4. Setting 4: Varying strategy and preference profile

Setting 4 is a combination of setting 2 and 3, which is closer to a real scenario. During each experiment, the bidding strategy of the buyer negotiator changes randomly among Boulware, Linear, Conceder and naive-TFT every 100 negotiations. The preference of the buyer negotiator changes randomly after every negotiation during experiments.

4.4.5. Baseline

Our opponent modeling method is unique in three main aspects. Firstly, our method models the bidding strategy of the opponent based on the bidding history of a series of previous negotiations. Secondly, our method views the opponent negotiator as a DRL agent and tries to model the agent's policy. Thirdly, our method can collaborate with the existing preference estimating methods. To the best of our knowledge, there is no similar work in the field of opponent modeling in automated negotiation. Therefore, to verify our method can indeed learn the pattern of the bidding strategy of the opponent to some extent, we use a random-guess model as our baseline.

4.5. Results

With experiments in the above four settings, we evaluate the performance of our method while modeling the opponent's bidding strategies. We will demonstrate and explain the experiment results in four settings in detail in the following subsections.

4.5.1. Setting 1: Fixed strategy and preference profiles

Plots in Fig. 4.2 present the performance of $option_{noF}$ and $option_F$ while modeling the buyer negotiator which has time-dependent (Boulware, Linear and Conceder) and behavior-dependent (naive-TFT) as its bidding strategy respectively in different instances of the experiment. Plots 4.2a, 4.2c, 4.2e and 4.2g compare the performance of $option_{noF}$ and $option_F$ to the baseline in terms of relative accuracy acc_{rel}. Plots 4.2b, 4.2d, 4.2f and 4.2h show the performance of $option_{noF}$ and $option_F$ while using absolute accuracy acc_{abs} as the metric since we are interested in our method's ability of predicting the absolute actions a_r^{abs} with the help of existing preference estimation methods as well ($option_F$).

We can find that the acc_{rel} of $option_{noF}$ is stable while modeling different time-dependent and behavior-dependent bidding strategies. The acc_{rel} of $option_F$ is between 0.6 and 0.7 most of the time while modeling opponents with varying bidding strategies. From modeling the Boulware strategy to modeling the naive-TFT strategy, the performance gap between $option_{noF}$ and $option_F$ increases in terms of acc_{rel}. Although $option_F$ does not perform as well as $option_{noF}$, both $option_{noF}$ and $option_F$ outperform the baseline

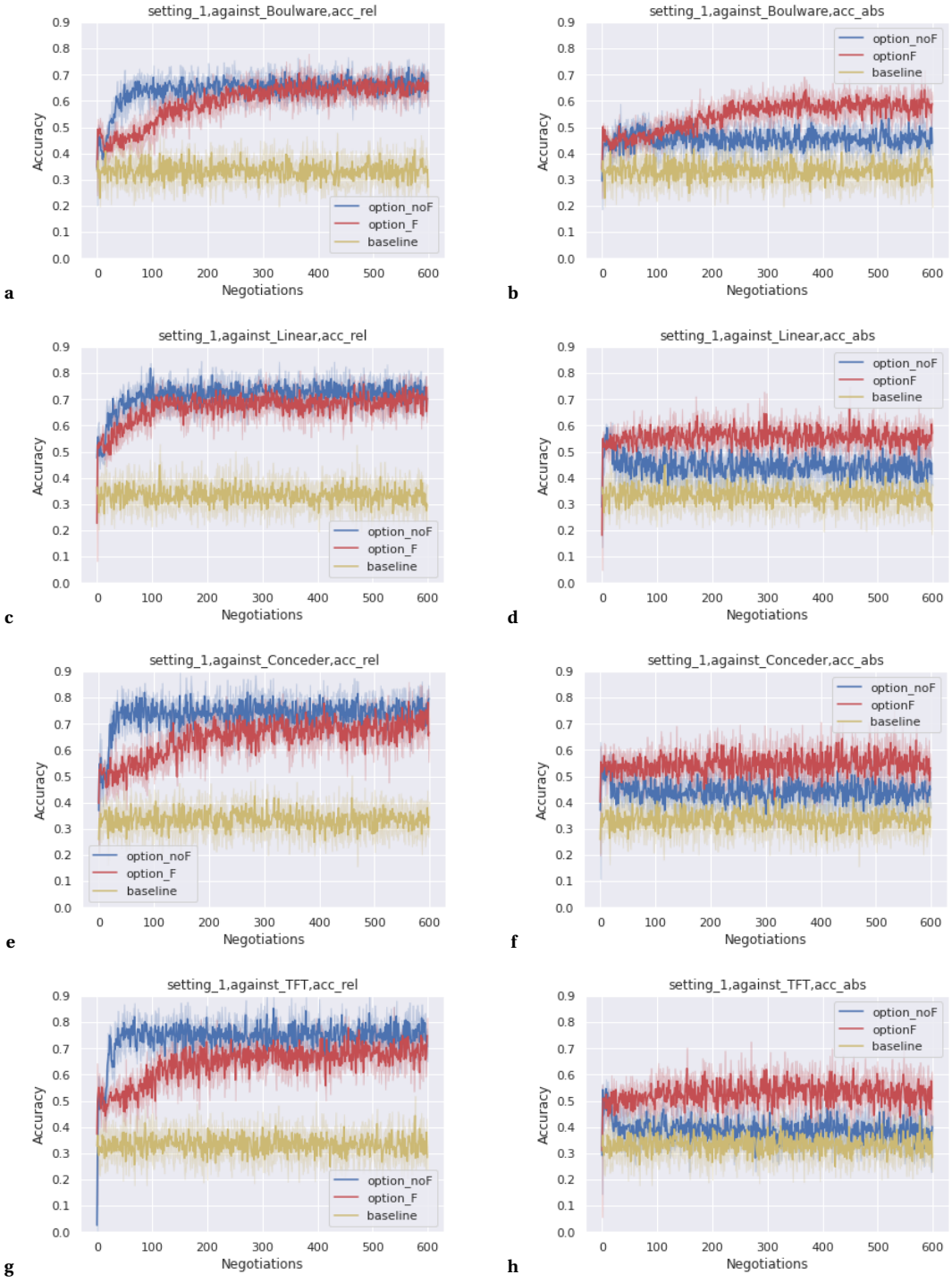


Figure 4.2: Results of experiments in setting 1 where the preference and bidding strategy of the opponent (buyer negotiator) are both fixed. Plots a, c, e and g show the performance of two different options in acc_{rel}. Plots b, d, f and h compare the performance of two options in terms of acc_{abs}

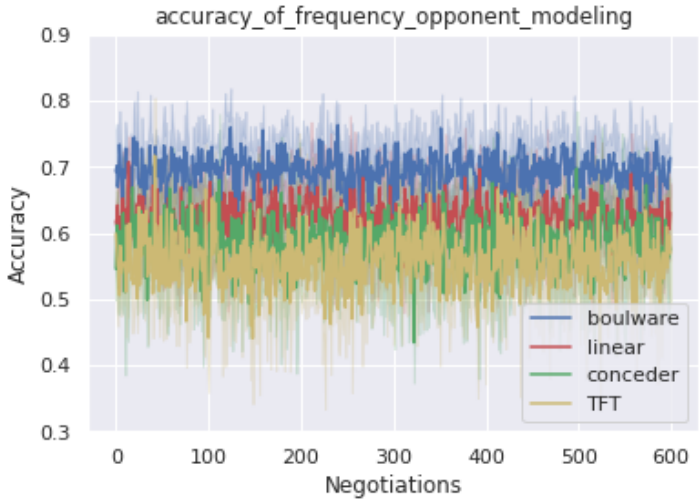


Figure 4.3: The accuracy of preference estimation while using frequency opponent modeling. The model has the highest accuracy while estimating the preference of the Boulware negotiator and the lowest accuracy while estimating the preference of the naive-TFT negotiator

model in terms of acc_{rel}.
The acc_{abs} of $option_F$ are even lower compared to the acc_{rel}. Meantime, the difference between acc_{abs} and acc_{rel} of $option_F$ slightly grows from the plots on the first row to the plots on the last row. However, our method's $option_F$ can still build a model of the buyer negotiator's bidding strategies to some extent before the 200th negotiation and outperforms the baseline model in terms of acc_{abs} as well. As we expected, $option_{noF}$ cannot predict the absolute actions a_r^{abs} without preference estimation methods. The acc_{abs} of $option_{noF}$ is always lower than that of $option_F$ and close to the baseline.
The reason for such a performance gap between $option_{noF}$ and $option_F$ and the difference between two used metrics of $option_F$ is that, for $option_F$, the accuracy of modeling an opponent's bidding strategy is highly influenced by the accuracy of modeling the opponent's preference (we use frequency opponent modeling as the preference estimation method). The frequency opponent modeling performs best while modeling the preference of the Boulware negotiator, and performs worst while modeling the preference of the Conceder negotiator and naive-TFT negotiator. As presented in Fig. 4.3, the accuracy of frequency opponent modeling changes when the buyer negotiator has a different bidding strategy.

4.5.2. Setting 2: Fixed strategy and varying preference profile

Plots in Fig. 4.4 present the performance of $option_{noF}$ and $option_F$ in experiments of setting 2, where the preference of the buyer negotiator changes randomly. Since acc_{abs} is only influenced by the frequency opponent modeling as we explained in the previous sub-section, the plots only show acc_{rel} of $option_{noF}$ and $option_F$. Plots 4.4a, 4.4c, 4.4e and 4.4g present how the performance of $option_{noF}$ in setting 2 compares to the performance in setting 1 while modeling three different time-dependent bidding strategies (Boulware, Linear, and Conceder) and one behaviour-dependent bidding strategy (naive-TFT). Plots 4.4b, 4.4d, 4.4f and 4.4h present how the performance of $option_F$ in setting 2 compares to the performance in setting 1.
As we expected, the performance of our method is influenced by changing preferences. On the one hand, the adverse effects of randomly changing preferences are more evident for $option_{noF}$ than $option_F$ while modeling time-dependent strategies. The learned opponent's bidding strategy model of $option_{noF}$ in setting 2 is more unstable and less accurate than the model of $option_{noF}$ in setting 1 since $option_{noF}$'s estimation of the actions of the opponent will change once the preference of one side of the negotiation changes. Although struggling with the instability due to the evolving preferences, $option_{noF}$ can still build a model of the opponent with an accuracy of around 0.6, which is still higher than a naive model. One reason could be that changing preferences is not significant enough to break the bidding strategy model learned with previous preferences, which is usually the case in the real world since we used load profiles constructed from data on the energy consumption of representative households. Fig. 4.5 shows the averaged daily load profiles of a "Tier 3" consumer across one year. The energy consumption pattern is relatively stable with low variance.

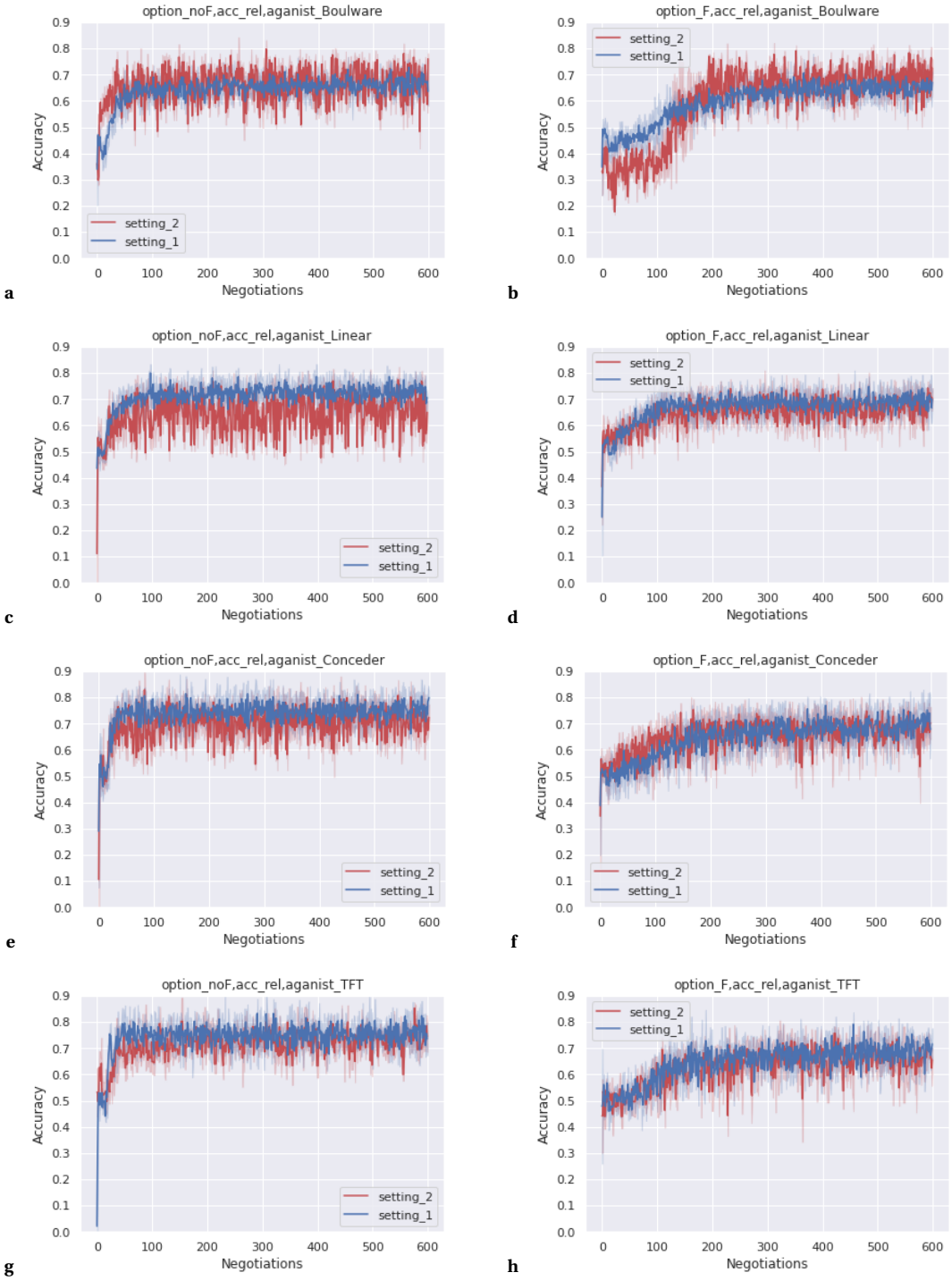


Figure 4.4: Results of experiments in setting 2 where the preference of the opponent changes every negotiation and the opponent's bidding strategy is fixed during an experiment. Plots a, c, e and g compare the performances of $option_{noF}$ in setting 1 and setting 2. Plots b, d, f and h compare the performances of $option_F$ in setting 1 and setting 2

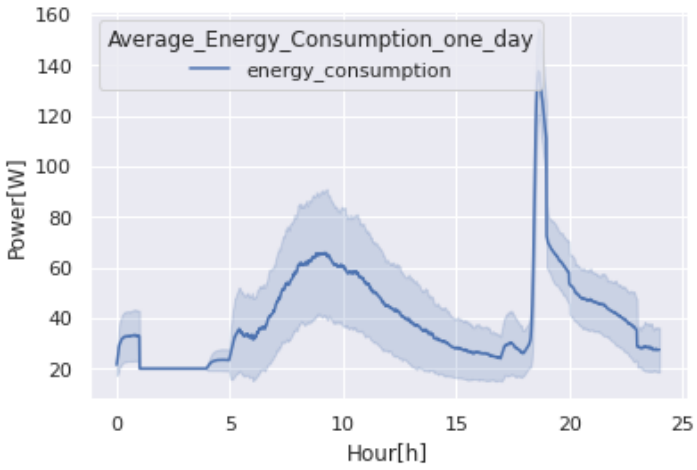
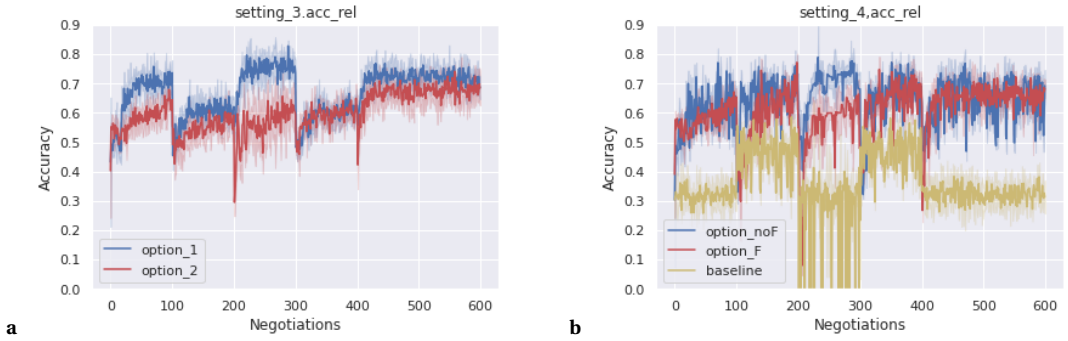


Figure 4.5: Averaged daily energy consumption of "Tier 3" consumer across the year

On the other hand, such influence is not significant for $option_{noF}$ while modeling the naive-TFT strategy. The reason can be that the behavior of a naive-TFT negotiator will also change if the preference of one side of the negotiation changes because the naive-TFT negotiator always tries to imitate its opponent's behavior. The changing of behavior mediates the effect of changing preference to some extent.
Opposite to $option_{noF}$, $option_F$ has decent and stable performances while modeling time-dependent strategies because $option_F$ uses the utility function estimated by the frequency opponent model rather than its own utility function. However, $option_F$'s performance is compromised a bit while modeling the behavior-dependent strategies because the behavior of the naive-TFT strategy will change if the preference of one side of the negotiation changes.
Since the energy consumption pattern of a particular consumer is relatively stable across one year, we also tried to improve the performance of $option_F$ by not initializing the frequency opponent modeling at the start of each negotiation. The design and results of this extra setting are presented in appendix B.



a b

Figure 4.6: Results of experiments in the third and fourth settings. In the third setting, the bidding strategy of the opponent randomly changes every 100 negotiations where the opponent's preference is fixed during an experiment. The fourth setting is a combination of the second and third setting.

4.5.3. Setting 3: Varying strategy and fixed preference profile

Plot 4.6a shows the performance of $option_{noF}$ and $option_F$ in setting 3 with acc_{rel} as metric. Both $option_{noF}$ and $option_F$ can sense the change in bidding strategy and model the new strategies. However, as explained while analyzing the results of setting 1, different bidding strategies can influence the accuracy and stability of frequency opponent modeling used in $option_F$. Therefore, $option_F$ is more struggling with reacting to the changes in bidding strategy and has a lower performance than $option_{noF}$ in most of time.

4.5.4. Setting 4: Varying strategy and preference profile

Setting 4 is the combination of setting 2 and setting 3. Therefore, based on the results of setting 2 and 3, it is not hard to predict the performance of our new opponent modeling method. Plot 4.6b presents how $option_{noF}$ and $option_F$ of our method performs in setting 3. . As we expected, the performance of both $option_{noF}$ and $option_F$ is always better than that of the naive model.

4.6. Discussion

In the experiments, we evaluated our opponent modeling method with an existing automated negotiation system for the off-grid P2P energy market and real-life load profiles of representative off-grid households. Our method generally has stable performance: our method can build a model of all bidding strategies in experiments. Once the model is made, the model's accuracy does not decrease as more negotiations happen between seller and buyer agents. Besides, although prediction accuracy may drop while modeling negotiators with changing preferences and inconsistent bidding strategies, both options of our method outperform the random-guess model (baseline) in all scenarios. However, since, in some cases, our method can only have a prediction accuracy of around 0.6, which cannot be counted as a remarkable prediction, it is critical to decide the timing of using the current version of our method. On the one hand, in some automated negotiation systems where negotiation failure can lead to severe harmful consequences, wrong predictions may increase the probability of failure. People should be cautious about relying on our method. On the other hand, in some systems where collapses of negotiations do not influence the daily life of consumers a lot, and the correct predictions can help negotiators to find better joint agreements, which can benefit the whole market, then it is worth giving our method a try.

Furthermore, during experiments, we notice that $option_{noF}$ and $option_F$ are good at different scenarios. Therefore, deciding which option to use during negotiations is also essential.

Last but not least, our system has the potential to be further improved and adjusted to other P2P systems where agents need to interact with each other repeatedly and knowing the strategy of other agents is vital for decision-making.

5

Conclusion and Future works

5.1. Conclusion

In this project, we designed a new opponent modeling method to model the bidding strategy of opponent negotiators. Our new method is dedicated to the automated negotiation applied to the P2P energy market because *a*) Two negotiators can negotiate with each other many times (assumption $A1$). Therefore, our method can collect sufficient data to build a good model *b*) Agents' preferences in the P2P energy market usually depend on their energy consumption patterns, which are generally stable, as demonstrated during experiments. Besides, agents always utilize similar bidding strategies in different negotiations (Assumption $A2$). Therefore, the learned model can be used in future negotiations without the requirement of significant modification. To the best of our knowledge, there is no similar opponent modeling technique that can utilize the characteristics of the P2P energy market in the domain of automated negotiation. With the ability to model the opponent's bidding strategy and predict the opponent's future actions, an automated negotiator attending in a P2P energy market can conduct what-if analyses and finally make a better decision, which can improve the efficiency of the energy distribution in the P2P energy market. With the improved ability of automated negotiators, the prosumers should have a larger incentive to take part in the P2P energy market where automated negotiation is applied.

To evaluate the overall ability of our new opponent modeling method, we applied it to a bilateral automated negotiation system designed for the P2P off-grid energy market. To make our experiments close to reality, we also used load profiles from Narayan et al., 2020. The used load profiles are all constructed from data on the energy consumption of representative consumers. With the experiments' results, we can now answer the research questions we proposed.

1. Can our opponent modeling technique models different bidding strategies with good accuracy?
   Answer: Based on the results of experiments in setting 1, our method can model representative time-dependent and behavior-dependent strategies.

2. How stable our opponent modeling method is while the opponent's preference profile changes?
   Answer: Both the $option_{noF}$ and $option_F$ of our method can be negatively influenced by changing preference profiles. However, since the preferences of the household in the P2P energy market mainly depend on their consumption patterns which are usually stable across the year, our method is still stable in the case of changing preferences. Besides, in the experiment, we found that $option_{noF}$ and $option_F$ are good at modeling different bidding strategies. Therefore, the two options of our method can sometimes be complementary to each other.

3. Can our opponent modeling reacts to the changing of opponent's bidding strategy?
   Answer: Based on the experiments in setting 3, both $option_{noF}$ and $option_F$ of our method can quickly react to the change of bidding strategies.

5.2. Future works

There are three directions for future works. The first direction is about the further evaluation of our opponent modeling method. So far, only straightforward time-dependent and behavior-dependent negotiators are used in experiments. We are interested to see the performance of our method while modeling negotiators with more complex bidding strategies and the ability of opponent modeling. Besides, the preference and consumption pattern of agents in the P2P-energy market can also be influenced by season and weather. For example, households usually spend more power on air conditions and less energy on lights during hotter seasons. As a result, they are more sensitive to the price in the summer (Filippini and Pachauri, 2004). It would be nice to include more such factors in experiments. The second direction is about further improvement of our opponent modeling method. The accuracy of $option_F$ of our method is largely affected by the accuracy of the used preference estimation method. Therefore, our method can be further improved with better preference estimation methods. We used frequency opponent modeling as the preference estimation method in this project, which can probably be improved if more detailed data on the consumption patterns of different consumers are available. Additional, we found in experiments that $option_{noF}$ and $option_F$ are good at modeling different bidding strategies. It would be good if there were a mechanism to decide which option to use in different scenarios. The third direction is about utilizing the model built by our method. A model is useless without proper utilization of it. An automated negotiator can probably conduct a search to find good decisions by using the learned model's predicted actions. Furthermore, if a RL (reinforcement learning) negotiator includes predicted actions in their observation, it may be able to learn a better policy with proper training.

Bibliography

Afiouni, E. N., & Ovrelid, L. J. (2013). Negotiation for strategic video games [Accepted: 2014-12-19T13:40:36Z Publisher: Institutt for datateknikk og informasjonsvitenskap]. 140. Retrieved June 3, 2022, from https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/253482

Alam, M., Gerding, E. H., Rogers, A., & Ramchurn, S. D. (2015). A scalable, decentralised multi-issue negotiation protocol for energy exchange [Num Pages: 7]. Retrieved December 13, 2021, from https://eprints.soton.ac.uk/376618/

Alam, M. R., St-Hilaire, M., & Kunz, T. (2017). An optimal p2p energy trading model for smart homes in the smart grid. Energy Efficiency, 10(6), 1475–1493. https://doi.org/10.1007/s12053-017-9532-5

Alsrheed, F., El Rhalibi, A., Randles, M., & Merabti, M. (2014). Intelligent agents for automated cloud computing negotiation. 2014 International Conference on Multimedia Computing and Systems (ICMCS), 1169–1174. https://doi.org/10.1109/ICMCS.2014.6911305

Andoni, M., Robu, V., & Flynn, D. (2017). Crypto-control your own energy supply [Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7666 Primary_atype: Correspondence Publisher: Nature Publishing Group Subject_term: Energy;Mathematics and computing Subject_term_id: energy;mathematics-and-computing]. Nature, 548(7666), 158–158. https://doi.org/10.1038/548158b

Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., McCallum, P., & Peacock, A. (2019). Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renewable and Sustainable Energy Reviews, 100, 143–174. https://doi.org/10.1016/j.rser.2018.10.014

Aydoğan, R., Festen, D., Hindriks, K. V., & Jonker, C. M. (2017). Alternating offers protocols for multilateral negotiation [Series Title: Studies in Computational Intelligence]. In K. Fujita, Q. Bai, T. Ito, M. Zhang, F. Ren, R. Aydoğan, & R. Hadfi (Eds.), Modern approaches to agent-based complex automated negotiation (pp. 153–167). Springer International Publishing. https://doi.org/10.1007/978-3-319-51563-2_10

Baarslag, T. (2014). What to bid and when to stop. (Doctoral dissertation) [ISBN: 9789461863058 OCLC: 905871136]. [s.n.] S.l.

Baarslag, T., Hendrikx, M., Hindriks, K., & Jonker, C. (2012). Measuring the performance of online opponent models in automated bilateral negotiation [Series Title: Lecture Notes in Computer Science]. In M. Thielscher & D. Zhang (Eds.). D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, & G. Weikum (**typeredactors**), AI 2012: Advances in artificial intelligence (pp. 1–14). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35101-3_1

Baarslag, T., Hendrikx, M., Hindriks, K., & Jonker, C. (2013). Predicting the performance of opponent models in automated negotiation. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 59–66. https://doi.org/10.1109/WI-IAT.2013.91

Baarslag, T., Hendrikx, M. J. C., Hindriks, K. V., & Jonker, C. M. (2016). Learning about the opponent in automated bilateral negotiation: A comprehensive survey of opponent modeling techniques. Autonomous Agents and Multi-Agent Systems, 30(5), 849–898. https://doi.org/10.1007/s10458-015-9309-1

Baarslag, T., Hindriks, K., Hendrikx, M., Dirkzwager, A., & Jonker, C. (2014). Decoupling negotiating agents to explore the space of negotiation strategies. https://doi.org/10.1007/978-4-431-54758-7_4

Bhatia, M., & Angelou, N. (2015, July). Beyond connections: Energy access redefined (Working Paper) [Accepted: 2016-05-31T19:15:42Z ISSN: 2628-5649]. World Bank. Washington, DC. Retrieved July 1, 2022, from https://openknowledge.worldbank.org/handle/10986/24368

Brzostowski, J., & Kowalczyk, R. (2006). Adaptive negotiation with on-line prediction of opponent behaviour in agent-based negotiations. 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, 263–269. https://doi.org/10.1109/IAT.2006.26

Chakraborty, S., Baarslag, T., & Kaisers, M. (2018). Energy contract settlements through automated negotiation in residential cooperatives. 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 1–6. https://doi.org/10.1109/SmartGridComm.2018.8587537

Chakraborty, S., Baarslag, T., & Kaisers, M. (2019). Automated peer-to-peer negotiation for energy contract settlements in residential cooperatives. Applied Energy, 259. https://doi.org/10.1016/j.apenergy.2019.114173

Chen, K. K. (2012). Artistic prosumption: Cocreative destruction at burning man [Publisher: SAGE Publications Inc]. American Behavioral Scientist, 56(4), 570–595. https://doi.org/10.1177/0002764211429362

Etukudor, C., Couraud, B., Robu, V., Früh, W.-G., Flynn, D., & Okereke, C. (2020). Automated negotiation for peer-to-peer electricity trading in local energy markets [Number: 4 Publisher: Multidisciplinary Digital Publishing Institute]. Energies, 13(4), 920. https://doi.org/10.3390/en13040920

Etukudor, C., Robu, V., Couraud, B., Kocher, G., Früh, W.-G., Flynn, D., & Okereke, C. (2019). Automated negotiation for peer-to-peer trading of renewable energy in off-grid communities. 2019 IEEE PES/IAS PowerAfrica, 1–6. https://doi.org/10.1109/PowerAfrica.2019.8928640

Fang, F., Xin, Y., Yun, X., & Haitao, X. (2008). An opponent's negotiation behavior model to facilitate buyer-seller negotiations in supply chain management. 2008 International Symposium on Electronic Commerce and Security, 582–587. https://doi.org/10.1109/ISECS.2008.93

Faratin, P., Sierra, C., & Jennings, N. R. (2002). Using similarity criteria to make issue trade-offs in automated negotiations. Artificial Intelligence, 142(2), 205–237. https://doi.org/10.1016/S0004-3702(02)00290-4

Faratin, P., Sierra, C., & Jennings, N. R. (1998). Negotiation decision functions for autonomous agents [Multi-Agent Rationality]. Robotics and Autonomous Systems, 24(3), 159–182. https://doi.org/https://doi.org/10.1016/S0921-8890(98)00029-3

Filippini, M., & Pachauri, S. (2004). Elasticities of electricity demand in urban indian households. Energy Policy, 32(3), 429–436. https://doi.org/https://doi.org/10.1016/S0301-4215(02)00314-2

Hou, C. (2004). Predicting agents tactics in automated negotiation. Proceedings. IEEE/WIC/ACM International Conference on Intelligent Agent Technology, 2004. (IAT 2004)., 127–133. https://doi.org/10.1109/IAT.2004.1342934

Jogunola, O., Adebisi, B., Anoh, K., Ikpehai, A., Hammoudeh, M., Harris, G., & Gacanin, H. (2018). Distributed adaptive primal algorithm for p2p-ETS over unreliable communication links. Energies, 11. https://doi.org/10.3390/en11092331

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2020). Multi-agent actor-critic for mixed cooperative-competitive environments. arXiv:1706.02275 [cs]. Retrieved April 14, 2022, from http://arxiv.org/abs/1706.02275

Masvoula, M., Halatsis, C., & Martakos, D. (2011). Predictive automated negotiators employing risk-seeking and risk-averse strategies. 363, 325–334. https://doi.org/10.1007/978-3-642-23957-1_37

Mohammad, Y., Nakadai, S., & Greenwald, A. (2020). Negmas: A platform for automated negotiations. PRIMA 2020: Principles and Practice of Multi-Agent Systems: 23rd International Conference, Nagoya, Japan, November 18–20, 2020, Proceedings, 343–351. https://doi.org/10.1007/978-3-030-69322-0_23

Moret, F., & Pinson, P. (2019). Energy collectives: A community and fairness based approach to future electricity markets [Conference Name: IEEE Transactions on Power Systems]. IEEE Transactions on Power Systems, 34(5), 3994–4004. https://doi.org/10.1109/TPWRS.2018.2808961

Mudgal, C., & Vassileva, J. (2000). Bilateral negotiation with incomplete and uncertain information: A decision-theoretic approach using a model of the opponent. In M. Klusch & L. Kerschberg (Eds.), Cooperative information agents IV - the future of information agents in cyberspace (pp. 107–118). Springer. https://doi.org/10.1007/978-3-540-45012-2_11

Nair, N.-K. C., & Garimella, N. (2010). Battery energy storage systems: Assessment for small-scale renewable energy integration. Energy and Buildings, 42(11), 2124–2130. https://doi.org/10.1016/j.enbuild.2010.07.002

Narayan, N., Qin, Z., Popovic, J., Diehl, J. C., Bauer, P., & Zeman, M. (2020). Stochastic load profile construction for the multi-tier framework for household electricity access using off-grid dc appliances. Energy Efficiency, 13. https://doi.org/10.1007/s12053-018-9725-6

Paudel, A., Sampath, L. P. M. I., Yang, J., & Gooi, H. B. (2020). Peer-to-peer energy trading in smart grid considering power losses and network fees [Conference Name: IEEE Transactions on Smart Grid]. IEEE Transactions on Smart Grid, 11(6), 4727–4737. https://doi.org/10.1109/TSG.2020.2997956

Sanchez-Anguix, V., Tunalı, O., Aydoğan, R., & Julian, V. (2021). Can social agents efficiently perform in automated negotiation? [Number: 13 Publisher: Multidisciplinary Digital Publishing Institute]. Applied Sciences, 11(13), 6022. https://doi.org/10.3390/app11136022

Sousa, T., Soares, T., Pinson, P., Moret, F., Baroche, T., & Sorin, E. (2019). Peer-to-peer and community-based markets: A comprehensive review [Publisher: Elsevier]. Renewable and Sustainable Energy Reviews, 104, 367–378. Retrieved November 29, 2021, from https://econpapers.repec.org/article/eeerensus/v_3a104_3ay_3a2019_3ai_3ac_3ap_3a367-378.htm

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (Second). The MIT Press. http://incompleteideas.net/book/the-book-2nd.html

Tunali, O., Aydogan, R., & Sanchez-Anguix, V. (2017). Rethinking frequency opponent modeling in automated negotiation [PRIMA 2017 : 20th International Conference on Principles and Practice of Multi-Agent Systems ; Conference date: 30-10-2017 Through 03-11-2017]. In B. An, A. Bazzan, J. Leite, S. Villata, & L. van der Torre (Eds.), Prima 2017 (pp. 263–279). Springer. https://doi.org/10.1007/978-3-319-69131-2_16

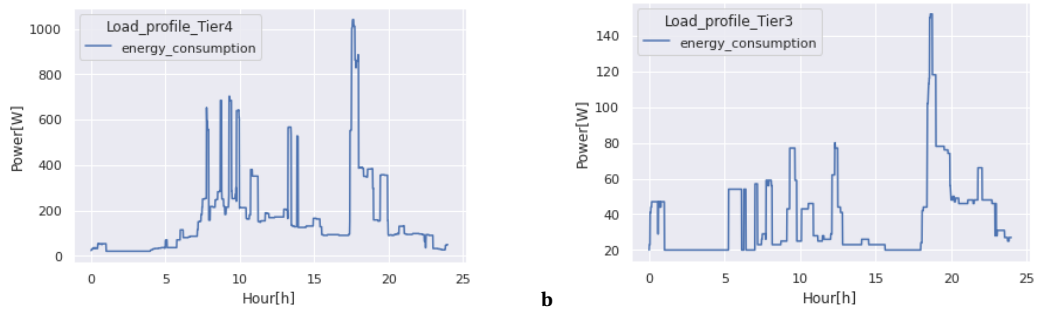
A

appendix-a

$Algorithm$ 1 demonstrates the procedure of checking whether an offer $W = (q_1, q_2, q_3, q_4, p)$ is feasible to the seller.

Algorithm 1 Check whether an offer is feasible

SoC : Amount of energy in the battery
SoC_{init} : initial energy in the batter
SoC_{max} : battery capacity
$Q_{seller} = (q_i^{required})$: energy requirement of the seller in each sector
$G_{seller} = (g_i)$: forecasting energy generation of the seller in each sector
procedure Feasible_Offer($W = (q_1, q_2, q_3, q_4, p)$)
    $SoC \leftarrow SoC_{init}$
    for $i \leftarrow 1,4$ **do**
        $SoC \leftarrow min(SoC + g_i - q_i^{required}, SoC_{max})$
        if $SoC \geq q_i$ **then**
            $SoC \leftarrow SoC - q_i$
        else
            return $not\ feasible$
        end if
    end for
    return $is\ feasible$
end procedure



a **b**

Figure A.1: One day load profiles of Tier 4 and Tier 3 consumers in experiments

Daily load profiles of Tier 4 and Tier 3 consumers used in our experiments are presented in figure A.1.

B

appendix-b

During the experiments of setting 2 where the opponent's preference randomly changes every negotiation, we found that the magnitude of changing load profiles is even lower than we expected. Therefore, to explore the probabilities of improving our method's $option_F$, we conducted experiments with one extra setting $setting_{ex}$. During $setting_{ex}$, the load profiles of the buyer negotiator (opponent) are drawn from a list of load profiles. The list consists of one particular "Tier 3" consumer's load profiles during one year and is ordered by date. During the experiment, the buyer negotiator draws one load profile from the list in order after each negotiation. Meantime, the preference estimation in our method does not initialize after each negotiation anymore. Instead, it keeps what it has learned in previous negotiations and updates itself in subsequent negotiations. Plots B.1a and B.1b compare the performances of $option_F$ in the case that the frequency opponent modeling is initialized every negotiation and the case that the frequency opponent modeling is not initialized every negotiation. Unluckily, we found that turning off the initialization of the frequency opponent modeling cannot improve our method further. A more dedicated way of updating and using the frequency opponent modeling is needed in order to improve our method.

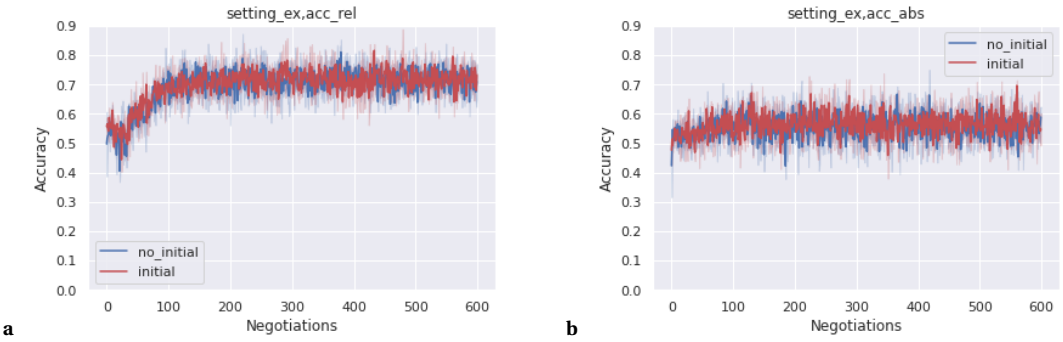


Figure B.1: Results of experiments of $setting_{ex}$. In this setting, the preference of the opponent changes each negotiation regularly. Plots a compares the performances of $option_F$ with and without the initialization of frequency opponent modeling after each negotiation in terms of acc_{rel}. Plot b shows the performance in terms of acc_{abs}.