Prospect Theory-inspired Automated P2P Energy Trading with Q-learning-based Dynamic Pricing

Ashutosh Timilsina
Department of Computer Science
University of Kentucky
Lexington, USA
ashutosh.timilsina@uky.edu

Simone Silvestri
Department of Computer Science
University of Kentucky
Lexington, USA
simone.silvestri@uky.edu

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001173

***Abstract*—The widespread adoption of distributed energy resources, and the advent of smart grid technologies, have allowed traditionally passive power system users to become actively involved in energy trading. Recognizing the fact that the traditional centralized grid-driven energy markets offer minimal profitability to these users, recent research has shifted focus towards decentralized *peer-to-peer (P2P) energy* markets. In these markets, users trade energy with each other, with higher benefits than buying or selling to the grid. However, most researches in P2P energy trading largely overlook the user perception in the trading process, assuming constant availability, participation, and full compliance. As a result, these approaches may result in negative attitudes and reduced engagement over time. In this paper, we design an *automated* P2P energy market that takes user perception into account. We employ *prospect theory* to model the user perception and formulate an optimization framework to maximize the buyer's perception while matching demand and production. Given the non-linear and non-convex nature of the optimization problem, we propose Differential Evolution-based Algorithm for Trading Energy called $DEbATE$. Additionally, we introduce a risk-sensitive Q-learning algorithm, named Pricing mechanism with Q-learning and Risk-sensitivity (PQR), which learns the optimal price for sellers considering their perceived utility. Results based on real traces of energy consumption and production, as well as realistic prospect theory functions, show that our approach achieves a 26% higher perceived value for buyers and generates 7% more reward for sellers, compared to a recent state of the art approach.**

***Index Terms*—Peer-to-peer energy trading, differential evolution, dynamic pricing, prosumer, prospect theory, Q-learning.**

I. Introduction

Distributed Energy Resources (DER), such as rooftop solar and wind turbine, have seen widespread proliferation among consumers in recent years [1]. In addition, the advent of Smart Grid (SG) technologies, Advanced Metering Infrastructures (AMI), and home energy management systems, have added flexibility in energy generation/consumption for consumers. This, in turn, has allowed traditionally passive consumers to become actively involved in energy trading by sharing the excess energy generated at their premise to either grid or other buyers [2], [3]. These active consumers with energy production capabilities have been referred to as *prosumers* [3], as a portmanteau of "producers" and "consumers". The role of prosumers in energy market has been recognized to some extent with the adoption of incentive schemes like *Feed-in-Tariff* (FiT) mechanism [4], [5]. FiT allows prosumers to sell excess energy to the grid and buy from grid when required [5]. However, existing energy trading modalities offer limited benefits to participating prosumers. This is due to the minimal prices at which energy is purchased by grid, as well as the low limits on the amount of energy that can be purchased [3]–[5].

A. Literature Review and Motivation

Peer-to-peer (P2P) energy trading is a recently proposed decentralized modality for energy sharing aiming at solving limitations of centralized techniques. This modality has been gaining significant traction recently [4], [5]. Specifically, P2P energy trading allows prosumers to trade energy among each other at a negotiated price with or without the involvement of the grid [4]. It generates better monetary incentives for prosumers compared to existing mechanisms while also reducing their grid dependency [5]. Additionally, increased local energy generation/consumption resulting from P2P trading leads to the minimization of overall system energy loss while providing an effective way to achieve demand side management [6]. Benefits extend also to the grid operator, by providing savings in investments that would have been otherwise required to develop/maintain transmission infrastructure in a centralized power distribution architecture [3], [4].

P2P energy trading has received attention from the research community in recent years. The works in [7], [8] present game theoretic approaches in a P2P setting, while a greedy rule-based P2P mechanism to assign energy among prosumers is proposed in [9] that includes mid-market pricing. Similarly, the physical aspects of P2P energy trading, such as power loss minimization and voltage regulation, have been explored in [10], [11]. These works, however, largely overlook the user behavior in designing their solutions. As established in [2], [3], [5], accommodating the user behavioral modeling in P2P energy trading ensures sustained participation from prosumers while incentivizing their contribution. In fact, the papers [7], [8] consider prosumers to be actively involved and fully compliant with the system as rational decision-makers. First concern with this assumption is that the continuous online presence of participating prosumers with the system might not always be possible in real-world application. Secondly,

research on user behavioral models and decision making [12], [13] have found users to have *bounded rationality*. Therefore, requiring constant active participation overwhelms the users and incentivizes non-rational decisions [14]. In the worst case, it might even result in users opting to terminate their participation altogether [2], [13]. In that light, the works in [2], [15] incorporates bounded rationality and user preferences into P2P energy trading. However, it requires continuous human participation and assumes a simplistic linear model for user perception. Conversely, the authors of [5] limit their focus on coalition formation in game theoretic setting and do not explicitly consider user behavioral modeling.

As a result, a prosumer-centric P2P energy trading model, that effectively incorporates the prosumers' decision-making behavior and their perceived loss/gain value from trading, is still lacking in the existing literature. Such a trading modality is expected to require minimal active participation from users while also ensuring their sustained involvement through the adoption of user behavioral modeling. To this end, the framework of *Prospect Theory* (PT) [16] can be used to model the non-rational user behavior in the face of uncertain decision-making. It is often regarded as fairly accurate mathematical representation of human behavior [16]–[18].

Recently, there has been few efforts in integrating PT in energy related applications as well to capture the irrationality of users [18]–[21]. In relation to P2P energy trading, the authors in [21] have proposed a PT-based distributed energy trading model to optimize trading decisions for prosumers in a competitive market. Although these papers model the user behavior in some ways, they require active participation from users and also assume that such behavior (e.g., the parameters of PT) is homogeneous for all the users. Social science studies, such as the one conducted in Italy [22] to investigate the social acceptance of nuclear energy using an online survey, show that users exhibit significant heterogeneity in their preferences for the sources of energy. Neuroscience studies have also stressed the heterogeneity of humans in reference to PT parameters [23]. Not capturing such heterogeneity provides little benefits in terms of user behavioral modeling.

B. Paper Contributions

In this paper, we design a PT-based optimization framework for prosumer-centric P2P energy trading as shown in Fig. 1. The framework aims at matching energy production and consumption (step 1 in Fig. 1) to maximize the perceived utility of individual buyers while taking into account the intrinsic heterogeneity of human perception. Given that the optimization problem is non-linear and non-convex, we further devise a *Differential Evolution*-based [24] metaheuristic algorithm called $DEbATE$ to solve the problem (*energy allocation*, step 2). In order to ensure minimal active participation of prosumers, we employ a Reinforcement Learning (RL) framework, called PQR, in tandem with $DEbATE$ to automate the pricing mechanism for sellers (*pricing mechanism*, step 3). In doing so, PQR learns the selling price for each sellers using a PT-based risk-sensitive Q-learning algorithm [25]. The output of the algorithms is then returned to the prosumers for executing the physical energy transactions (step 4). Using real datasets for energy production and consumption, paired with recent survey data for PT perception modeling, results show that $DEbATE$ performs 25% higher in buyer's perception and 7% higher in seller's reward compared to state-of-the-art approach.

The major contributions of the paper are the following:

- We develop a PT-inspired optimization framework for P2P energy trading;
- We design a metaheuristic algorithm $DEbATE$ to solve the non-linear energy allocation problem;
- We design dynamic pricing mechanism with PQR algorithm using risk-sensitive Q-learning approach;
- Experiments using real data show the superiority of proposed approach compared to the state-of-the-art;

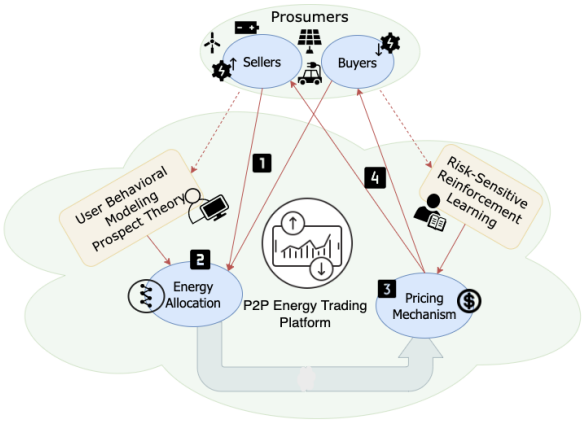

Fig. 1. P2P Energy Trading System Overview.

II. System Model and Problem Formulation

We consider a P2P energy trading system as shown in Fig. 1. The system consists of prosumers that can exchange energy among each other through an existing distribution network. The grid serves as backup for prosumers to either buy or sell energy, if the local energy trading is insufficient or not possible. Let P be the set of all prosumers participating in the P2P energy market. We refer to $B_t \subset P$ as the set of *Buyers*, i.e. the set of prosumers that have higher self-consumption than generation at a timeslot t, and consumers without energy generation capabilities. Similarly, $S_t \subset P$ is the set of *Sellers*, i.e., prosumers that have excess generation at a timeslot t. For simplicity of notation, we drop the subscript t in the following.

We model the perceived loss and gain of prosumers using the *prospect theory* (PT) value function to capture user perception on gains and losses. Specifically, consider the excess energy generation of seller $i \in S$ be r_i and demand of buyer $j \in B$ be w_j. Then, let $x_{ij} \in [0, 1]$ represent the fraction of w_j that a buyer j is willing to buy from seller i at ρ_i price per kWh amount of energy. There is an *energy loss* during the physical energy transfer through wires [6], which depends on the wire-length between i and j and directly proportional

to the amount of energy exchanged. The loss is modeled as a fraction $l_{ij} \in [0,1]$ of the energy exchanged. Assume ρ_{gs}, ρ_{gb} be the energy selling and purchasing prices from the grid. We adopt a modified PT value function to model realistic user perception in an energy market [16]. The function quantifies perceived utility of humans towards gain and loss based on degree of deviation from a reference point. Particularly, in our problem, it captures the difference of total actual buying cost y_j from the buyer's desired total reference cost $\rho_j w_j$ where ρ_j is the *reference price* of buyer j for purchasing energy. This utility function is formulated as

$$v(y_j) = \begin{cases} k_{+,j}(\rho_j w_j - y_j)^{\zeta_{+,j}}, & y_j < \rho_j w_j \\ -k_{-,j}(y_j - \rho_j w_j)^{\zeta_{-,j}}, & y_j \geq \rho_j w_j \end{cases} \quad (1)$$

where $k_{+,.}, k_{-.}, \zeta_{+,.}, \zeta_{-,.}$ are the parameters that control the degree of loss-aversion and risk-sensitivity. These parameters are found to be highly heterogeneous and vary from person to person based on factors like gender and age group [26], [27]. y_j is the total actual cost of buying energy for j^{th} buyer s.t.

$$y_j = \sum_{i \in S} \rho_i x_{ij} w_j + \rho_{gs}(1 - \sum_i x_{ij}) w_j$$

Note that, similar to the PT value function in [16], the utility function in Eq. (1) is concave in the gain domain (i.e. case $y_j < \rho_j w_j$) while convex in loss domain (i.e. case $y_j \geq \rho_j w_j$).

The problem of matching demand and production of heterogeneous prosumers is formalized as follows.

$$\text{maximize} \quad f(y) : \sum_{j \in B} v(y_j) \quad (2)$$

$$\text{s.t.} \quad \sum_{j \in B} (1 + l_{ij}) x_{ij} w_j \leq r_i, \quad \forall i \quad (2a)$$

$$\sum_{i \in S} x_{ij} \leq 1, \quad \forall j \quad (2b)$$

$$x_{ij} = 0, \text{ if } l_{ij} \geq l_{max}, \quad \forall i \quad (2c)$$

$$\rho_{gb} \leq \rho_i, \rho_j \leq \rho_{gs}, \quad \forall i \quad (2d)$$

$$x_{ij} \in [0,1], \quad \forall i, j \quad (2e)$$

The problem maximizes the sum of perceived utility for buyers in Eq. (2). Constraint in Eq. (2a) prevents the problem from exceeding the amount of energy being sold by each sellers while incorporating the losses in electric lines. The constraint in Eq. (2b) ensures that the energy demand for each buyers is not exceeded, while constraint (2c) limits the loss between sellers and buyers to be within the loss threshold l_{max}. Finally, the constraint (2d) limits upper and lower bound for energy price to the selling and buying price of the grid.

It is to be noted that the problem in Eq. (2) is non-linear, non-convex optimization problem. Hence, we propose a heuristic based on Differential Evolution Algorithm (DEA) [24] described in the following section. Additionally, in the above problem, the selling price is considered as a fixed amount for a trading period. However, the reference price ρ_j of buyer j is a personal value which may under- or over-estimate the competitiveness of market. In order to maximize the sellers' perceived objectives through prospect theory, we resort to the risk-sensitive Q-learning algorithm [25].

Algorithm 1: DEbATE

Input : set of buyers B, sellers S, fitness function $f(.)$, max iterations G_{max}, population size NP, crossover probability CR, differential weight F
Output: best identified feasible solution $\mathbf{x}^*$

1 Update set of buyers B and sellers S, $count = 0$;
2 Generate initial population $\mathcal{X} = \{\mathbf{x_k} | \ k = 1, \ldots, NP\}$;
3 **while** $count < G_{max}$ **do**
4 **for** *each* $\mathbf{x_k} \in \mathcal{X}$ **do**
5 Choose 3 different vectors $\{\mathbf{x_a}, \mathbf{x_b}, \mathbf{x_c}\} \in \mathcal{X}$ at random and $R \sim U(1, |S| \times |B|)$;
6 Create mutated solution $\bar{\mathbf{x}}_\mathbf{k} = \mathbf{x_k}$;
  /* **Mutation and Crossover** */
7 **for** *each* $i \in |S|$, $j \in |B|$ **do**
8 Select $u \sim U(0,1)$;
9 **if** $u < CR || (i \times j) == R$ **then**
10 $\bar{x}_{ij}^{(k)} = x_{ij}^{(a)} + F \times (x_{ij}^{(b)} - x_{ij}^{(c)})$;
11 $\bar{x}_{ij}^{(k)} = \min(1, \max(0, \bar{x}_{ij}^{(k)}))$
12 **end**
  /* **Check Constraints** */
13 $\forall i, j$, **if** $l_{ij} \geq l_{max}$ **then** $\bar{x}_{ij} = 0$;
14 $\forall i$, **if** $\sum_j (1 + l_{ij}) \bar{x}_{ij} w_j > r_i$ **then** $\bar{x}_{ij} = \frac{\bar{x}_{ij} r_i}{\sum_j (1 + l_{ij}) \bar{x}_{ij} w_j}$;
15 $\forall j$, **if** $\sum_i \bar{x}_{ij} > 1$ **then** $\bar{x}_{ij} = \frac{\bar{x}_{ij}}{\sum_i \bar{x}_{ij}}$;
  /* **Compare fitness** */
16 **if** $f(\bar{\mathbf{x}}_\mathbf{k}) > f(\mathbf{x_k})$ **then** $\mathcal{X} = (\mathcal{X} \setminus \{\mathbf{x_k}\}) \cup \{\bar{\mathbf{x}}_\mathbf{k}\}$;
17 **end**
18 count = count++;
19 **end**
  /* **Find the best solution to execute trading** */
20 Let $\mathbf{x}^* = \arg\max_{\mathbf{x_k} \in \mathcal{X}} f(\mathbf{x_k})$;
21 Execute transactions for each prosumers to $\mathbf{x}^*$;

III. The DEbATE and PQR Heuristics

In this section, we describe the *Differential Evolution-based Algorithm for Trading Energy (DEbATE)* (Alg. 1), designed for the problem presented in Section II, and the *Pricing mechanism with Q-learning and Risk-sensitivity (PQR)*, designed to dynamically adjust the sellers' prices.

A. DEbATE

$DEbATE$ is executed at each trading period (e.g., 12 hours) to solve the non-linear optimization problem in Eq. (2). It uses differential evolution to determine an optimal amount of energy to be traded between prosumers that maximizes the perceived utility of buyers. *DEbATE* initially updates the list of buyers (B) and sellers (S) based on the expected production and consumption for current trading period. These can be predicted accurately with recent approaches [28], [29]. The differential evolution-based optimization begins on line 2 where an *initial population* $\mathcal{X}$ is generated with population size of NP. An element $\mathbf{x_k} \in \mathcal{X}$, with $k = 1, 2, \ldots, NP$ is a *candidate solution* vector of variables x_{ij} representing the amount of energy to be traded between each seller i and buyer j . These variables correspond to the decision variables of our optimization problem.

The $while-$loop (line $3-19$) is the differential evolution loop that aims at finding solution to the non-linear optimization problem with Eq. (2) as the fitness function. The loop is executed for G_{max} iterations. At each iteration, for each candidate solution $\mathbf{x}_k \in \mathcal{X}$, the algorithm creates a *mutated solution* $\bar{\mathbf{x}}_\mathbf{k}$. Initially, $\bar{\mathbf{x}}_\mathbf{k} = \mathbf{x}_\mathbf{k}$. The mutated solution is subsequently updated through mutation and crossover with 3 random candidates $\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c \in \mathcal{X}$ (line 5). A value $R \in [1, |S| \times |B|]$ is selected at random. R will be used in the following $for-$loop to ensure a minimum mutation. The for loop in line 7 iterates over the components (dimensions in evolutionary terms) of $\bar{\mathbf{x}}_\mathbf{k}$. During each iteration, a value $u \in [0, 1]$ is sampled at random as mutation probability (line 8). Subsequently, a mutation occurs for the component ij of $\bar{\mathbf{x}}_\mathbf{k}$ with crossover probability CR (line 9). The mutation occurs irrespective of the probability if $(i \times j) = R$ (to ensure at least one minimum mutation). A mutation is executed by combining the corresponding component of $\mathbf{x_a}$, $\mathbf{x_b}$, and $\mathbf{x_c}$ with the differential weight parameter $F \in [0, 2]$ as in line 10. The mutated component $\bar{\mathbf{x}}_\mathbf{ij}^{(\mathbf{k})}$ is clipped to ensure that it falls within $[0, 1]$ as minimum and maximum threshold to satisfy constraint Eq. (2e) in line 11 of the algorithm.

After the mutated solution is finalized, it is checked, and adjusted if needed, to meet the constraints in Eqs. (2a)-(2c) of the optimization problem. Specifically, line 13 ensures that no exchange occurs (i.e., $\bar{\mathbf{x}}_\mathbf{ij}^{(\mathbf{k})} = 0$) between users having a loss higher than l_{max}. Lines $14-15$ ensure that the production of a seller and the demand of each buyer are not exceeded, respectively. Finally, in line 16, the fitness function $f(.)$ of the mutated solution $\bar{\mathbf{x}}_\mathbf{k}$ is compared against the original candidate solution $\mathbf{x}_\mathbf{k}$. If $f(\bar{\mathbf{x}}_\mathbf{k}) > f(\mathbf{x}_\mathbf{k})$, then $\bar{\mathbf{x}}_\mathbf{k}$ replaces $\mathbf{x}_\mathbf{k}$ in the set of candidate solutions $\mathcal{X}$. At the end of the while loop, $DEbATE$ selects the best solution $\mathbf{x}^*$ in $\mathcal{X}$ (line 20) and executes the transactions accordingly (line 21). In the following theorem 1, we show that the $DEbATE$ has polynomial complexity and hence, computationally efficient.

Theorem 1. *The complexity of the $DEbATE$ algorithm is $O(G_{max} \times NP \times |S||B|)$.*

Proof. The complexity is dominated by the $while$ loop (lines $3-19$), which is executed G_{max} times. Within this loop, the $for-$loop (lines $4-17$) does $|\mathcal{X}| = NP$ total iterations. In each iteration, the inner $for-$loop (lines $7-12$) iterates over the sets S and B, and only contains constant operations. Similarly, checking the constraints (lines $13-15$) requires to iterate over the same sets. Finally, calculating the function $f(.)$ (line 16) has cost $|B|$. Overall, the complexity is $O(G_{max} \times NP \times (|S||B| + 3|S||B| + |B|)) = O(G_{max} \times NP \times |S||B|)$ □

B. PQR

After determining the solution to the energy allocation problem in $DEbATE$, the selling price for sellers is then updated through the PQR algorithm. In order to learn the optimal selling price dynamically over time, we model the sellers as independent learning agents. Note that, to preserve the privacy and avoid the conflict between prosumers, these agents do not have access to information about other sellers or buyers. The state space in the Q-learning formulation consists of the prices between the grid buying (ρ_{gb}) and selling (ρ_{gs}), discretized by a step size, δ, i.e., $\rho_i \in \{\rho_{gb},\ \rho_{gb}+\delta,\ \rho_{gb}+2\delta,\ ...,\ \rho_{gb}+\left(\frac{\rho_{gs}-\rho_{gb}}{\delta}-1\right)\delta,\ \rho_{gs}\}$.

The action space consists of a price increasing action, price decreasing action, and no change action, i.e. $a \in \{+\delta, -\delta, 0\}$, where δ is the amount by which price is increased or decreased. Seller i reward function is the total revenue generated at the current trading period i.e. $R_i = (\rho_i + a) \sum_{j \in B} x_{ij} w_j$. For updating Q-values, we modify the approach proposed in [25] by considering the following Q-learning update rule that includes the PT-based perceived utility of sellers.

$$Q^{(new)}(s, a) = Q^{(old)}(s, a) + \alpha v(y_i) \tag{3}$$

$$v(y_i) = \begin{cases} k_{+,i}(y_i)^{\zeta_{+,i}}, & y_i > 0 \\ -k_{-,i}(-y_i)^{\zeta_{-,i}}, & y_i \leq 0 \end{cases} \tag{4}$$

where, $y_i = R_i + \gamma \max_a Q(s_{new}, a) - Q(s, a)$ is the Temporal Difference (TD) error of i^{th} seller for current iteration, and $v(y_i)$ is transformation of TD error to capture each seller's personalized perceived utility on loss and gain. α refers to the learning rate for updating Q-values in Eq. (3). The action is selected based on an ϵ-*greedy* exploration-exploitation strategy [30]. Specifically, ϵ refers to the probability of exploration and it is initially set to 1. It is then decreased over time using an $\epsilon-$*decay* value, as the system learns the optimal policy. Based on the selected action, the new selling price, reward, and Q-value are updated as per Eqs. (3) and (4). Updated selling price is then sent to the respective seller i for next trading period.

The system runs both $DEbATE$ and PQR sequentially at every trading period. Input of $DEbATE$ is updated based on the prices computed by PQR. PQR then takes as input the reward from executing energy transactions by $DEbATE$.

IV. Experimental Results

A. Experimental Setup

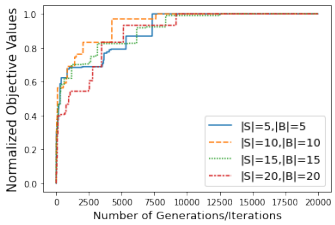

Fig. 2. Normalized objective value vs. number of iterations.

In this section, we evaluate the performance of *DEbATE* and *PQR*, hereafter jointly referred as $DEbATE - PQR$, against a recent state-of-the-art approach referred to as *Rule* [9]. $Rule$ allocates energy using a greedy heuristic that assigns cheapest sellers to buyers based on their registration order in the system, while final price of each transaction follows mid-market pricing, i.e., mid value of seller's and buyer's asking

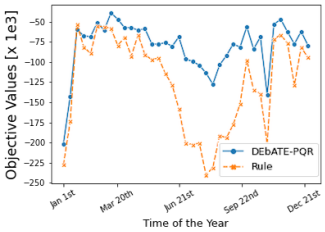

Fig. 3. Buyers' perceived values.

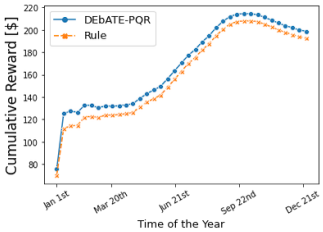

Fig. 4. Sellers' cumulative reward.

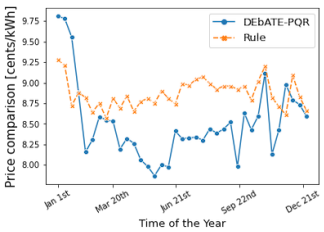

Fig. 5. Average price.

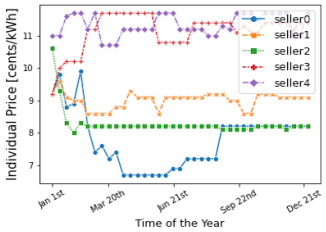

Fig. 6. Individual prices.

price. We consider a system with 40 prosumers, split evenly as buyers and sellers. This is considered a representative number of prosumers in a microgrid or set of houses supplied by a single distribution transformer. We use a realistic dataset for buyers' energy consumption obtained from [31]. Similarly, we consider sellers equipped with $4kW$ rooftop solar located in Lexington, Kentucky, USA. The energy generated is estimated using NREL's PVWatts Calculator [32] given the solar irradiance in Lexington and size of solar panels. Losses are assigned uniformly at random from set $\{1\%, 2\%, 3\%, 4\%\}$ and maximum loss threshold $L_{max} = 2.5\%$.

We assume that prosumers complete a survey before joining the system to estimate their individual prospect theory parameters, similar to [23], [26], [27], and use realistic prospect theory parameters determined by them. Specifically, we sample the risk-averting parameter for gains $(\zeta_+) \in [0.60, 0.88]$, the risk-seeking parameter for losses $(\zeta_-) \in [0.52, 1.0]$, the loss-aversion parameters for gain and loss $(k_+), (k_-) \in [2.10, 2.61]$ for each individual prosumers. The grid energy buying price is set to $\rho_{gb} = \$0.06$ and the selling price to $\rho_{gs} = \$0.12$. The reference price for each sellers is initially randomly sampled from range $[0.09, 0.12]$. It is then updated using PQR at each iteration. The reference price for each buyer is selected in the range $[0.06, 0.10]$ and considered static for the duration of experiments, which is 365 days. The parameters for PQR algorithm are set as follows: learning rate $\alpha = 10^{-4}$, step size for discretizing state space $\delta = \$0.001$, and $\epsilon-$decay $= 0.965$.

B. Results

We consider several experimental scenarios and performance metrics, as discussed in the following.

Experimental Scenario 1: We first run experiments to study the convergence of *DEbATE*. We considered different system size by scaling the number of sellers and buyers. Fig. 2 shows the normalized objective value as a function of the number of iterations using a population size $NP = 20$. The plot averaged over 10 runs shows that $10,000$ iterations are sufficient for the algorithm to converge in the considered settings. As a result, in the following scenarios we set $G_{max} = 10,000$ and the population size $NP = 20$.

Experimental Scenario 2: In the second experimental scenario we study the performance of the considered approaches over time. Two performance metrics are considered, namely the buyers' objective value and the sellers' cumulative reward. These are represented in Figs. 3 and 4, respectively, with a moving average of 10 days. In this experiments we consider 15 buyers and 15 sellers. The benefits of $DEbATE - PQR$ over $Rule$ are more prominent from April through October, when the energy demand and production is higher. The greedy nature of $Rule$ penalizes the quality of the resulting matching, significantly reducing the buyers' perceived value. Note that, the buyers' objective values are negative because they are paying higher prices than their reference purchase price. Therefore, transactions are seen as loss from a prospect theory perspective. Nevertheless, our approach optimizes the energy assignment to maximize the buyers perceived value. Additionally, our approach is able to generate higher rewards than $Rule$ by dynamically learning the prices for sellers through the PQR algorithm. The the sellers' reward decreases after mid-september for both the approaches due to the reduced energy production during winter.

We further study the performance over time by considering the evolution of average and individual sellers' prices. We consider a smaller system of 5 sellers and 5 buyers for ease of representation of the results. Fig. 5 shows the average price while Fig. 6 shows the individual prices. $DEbATE - PQR$ is able to learn and adjust the price over time to improve the buyers' perceived value considering their competitiveness. The competitiveness is a function of a buyer's reference price, their production, and their location in the system (e.g., loss w.r.t. sellers). As a result, our approach is able to improve the perception of both buyers and sellers while ensuring the competitiveness of the market.

Experimental Scenario 3: In this scenario we test the scalability with respect to the system size. Specifically, we increase the system proportionately from 5 sellers and 5 buyers to 20 sellers and 20 buyers. Figs. 7-8 show the buyers' total perceived value and the sellers' reward, respectively, over a year. By considering the loss-averse and risk-seeking PT-value functions, $DEbATE-PQR$ achieves an increasing advantage as the system size increases compared to $Rule$, for both sellers and buyers. As a numerical example, $DEbATE-PQR$ achieves as much as 26% increase in buyers' perceived value while ensuring 7% profit improvement for sellers.

Experimental Scenario 4: In the final scenario we study the learning trends of prices for $DEbATE - PQR$. Specifically, we observe the average price at 90, 300, and 365 days, for different system sizes. Fig. 9 shows the results. $DEbATE - PQR$ learns more profitable prices over time under all system sizes. It is to be noted that prices are lower for larger systems,

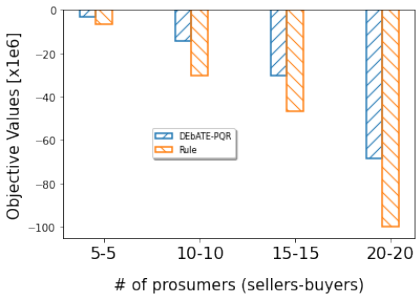

Fig. 7. Obj. values for buyer vs. network size.

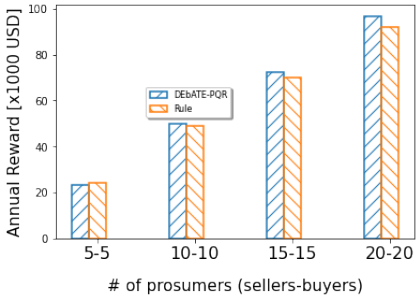

Fig. 8. Total rewards for sellers vs. network size.

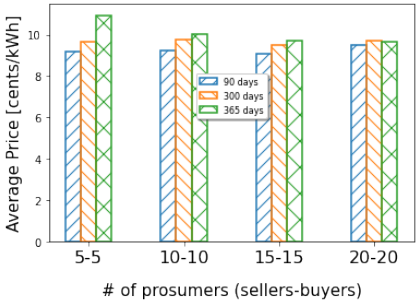

Fig. 9. Avgerage price vs. network size.

due to the increased competitiveness.

V. Concluding Remarks

In this paper, we bring together the concept of perceived utility from behavioral economics and reinforcement learning into the P2P energy trading scene. Unlike existing literature, we propose an automated and dynamic P2P energy trading problem that maximizes the perceived value for buyers while simultaneously learning the optimal selling price. Given the non-linear and non-convex nature of the problem, we propose a novel differential evolution-based metaheuristic algorithm, called $DEbATE$. $DEbATE$ is paired with a prospect theory enhanced Q-learning algorithm, called PQR, to adjust the selling price over time. Results show the advantages of the proposed approaches with respect to a state of the art solution using real energy consumption and production data.

Acknowledgment

This work is supported by the NSF grant EPCN-1936131 and NSF CAREER grant CPS-1943035.

References

[1] "Iea org," https://www.iea.org/reports/electricity-information-overview/.
[2] A. Timilsina, A. R. Khamesi, V. Agate, and S. Silvestri, "A reinforcement learning approach for user preference-aware energy sharing systems," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1138–1153, 2021.
[3] Y. Parag and B. Sovacool, "Electricity market design for the prosumer era," *Nature Energy*, vol. 1, p. 16032, March 2016.
[4] W. Tushar, T. K. Saha, C. Yuen *et al.*, "Peer-to-peer trading in electricity networks: an overview," *IEEE Transactions on Smart Grid*, 2020.
[5] W. Tushar, T. K. Saha, C. Yuen, P. Liddell, R. Bean, and H. V. Poor, "Peer-to-peer energy trading with sustainable user participation: A game theoretic approach," *IEEE Access*, vol. 6, pp. 62 932–62 943, 2018.
[6] T. Zhu, Z. Huang, A. Sharma, J. Su, D. Irwin, A. Mishra, D. Menasche, and P. Shenoy, "Sharing renewable energy in smart microgrids," in *ACM/IEEE ICCPS*, 2013.
[7] W. Tushar, C. Yuen, H. Mohsenian-Rad, T. Saha, H. V. Poor, and K. L. Wood, "Transforming energy networks via peer-to-peer energy trading: The potential of game-theoretic approaches," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 90–111, 2018.
[8] W. Tushar, T. K. Saha, C. Yuen, T. Morstyn, H. V. Poor, R. Bean *et al.*, "Grid influenced peer-to-peer energy trading," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1407–1418, 2019.
[9] M. I. Azim, S. Pourmousavi, W. Tushar *et al.*, "Feasibility study of financial p2p energy trading in a grid-tied power network," in *IEEE PESGM*, 2019.
[10] M. Nasimifar, V. Vahidinasab, and M. S. Ghazizadeh, "A peer-to-peer electricity marketplace for simultaneous congestion management and power loss reduction," in *IEEE Smart Grid Conference*, 2019.
[11] A. Paudel, L. Sampath, J. Yang, and H. B. Gooi, "Peer-to-peer energy trading in smart grid considering power losses and network fees," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4727–4737, 2020.
[12] G. Gigerenzer and R. Selten, *Bounded rationality: The adaptive toolbox*. MIT press, 2002.
[13] D. E. Agosto, "Bounded rationality and satisficing in young people's web-based decision making," *Journal of the American society for Information Science and Technology*, vol. 53, no. 1, pp. 16–27, 2002.
[14] P. E. Earl, "Bounded rationality in the digital age," in *Minds, Models and Milieux*. Springer, 2016, pp. 253–271.
[15] V. Agate, A. R. Khamesi, S. Silvestri, and S. Gaglio, "Enabling peer-to-peer user-preference-aware energy sharing through reinforcement learning," in *IEEE ICC*, 2020.
[16] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 2013, pp. 99–127.
[17] G. El Rahi, W. Saad, A. Glass *et al.*, "Prospect theory for prosumer-centric energy trading in the smart grid," in *IEEE PES ISGT*, 2016.
[18] G. El Rahi, S. R. Etesami, W. Saad, N. B. Mandayam, and H. V. Poor, "Managing price uncertainty in prosumer-centric energy trading: A prospect-theoretic stackelberg game approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 702–713, 2017.
[19] W. Saad, A. L. Glass, N. B. Mandayam, and H. V. Poor, "Toward a consumer-centric grid: A behavioral perspective," *Proceedings of the IEEE*, vol. 104, no. 4, pp. 865–882, 2016.
[20] Y. Wang, L. Zhang, Q. Ding, and K. Zhang, "Prospect theory-based optimal bidding model of a prosumer in the power market," *IEEE Access*, vol. 8, pp. 137 063–137 073, 2020.
[21] Y. Yao, C. Gao, T. Chen *et al.*, "Distributed electric energy trading model and strategy analysis based on prospect theory," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 106865, 2021.
[22] D. Contu, E. Strazzera, and S. Mourato, "Modeling individual preferences for energy sources: The case of iv generation nuclear energy in italy," *Ecological Economics*, vol. 127, pp. 37–58, 2016.
[23] C. R. Fox and R. A. Poldrack, "Prospect theory and the brain," in *Neuroeconomics*. Elsevier, 2009, pp. 145–173.
[24] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.
[25] Y. Shen, M. J. Tobia, T. Sommer, and K. Obermayer, "Risk-sensitive reinforcement learning," *Neural computation*, vol. 26, no. 7, 2014.
[26] V. Baláž, V. Bačová, E. Drobná *et al.*, "Testing prospect theory parameters," *Ekonomicky časopis*, vol. 61, 2013.
[27] M. O. Rieger, M. Wang, and T. Hens, "Estimating cumulative prospect theory parameters from an international survey," *Theory and Decision*, vol. 82, no. 4, pp. 567–596, 2017.
[28] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on lstm recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2017.
[29] E. Casella, E. Sudduth, and S. Silvestri, "Dissecting the problem of individual home power consumption prediction using machine learning," in *IEEE SMARTCOMP*, 2022.
[30] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
[31] Pecan street inc. [Online]. Available: www.pecanstreet.org
[32] Solar Resource Data. [Online]. Available: pvwatts.nrel.gov/pvwatts.php