Faculté
des sciences économiques
et de gestion
Université de Strasbourg

El Farol Bar Problem

Moritz MULLER

DOBROSSY Samuel
PETIT-DANGEON Jeanne
M2 DS2E

Table des matières

I. Introduction 3

II. Basic model 4

III. Extension 6

IV. Implementation 8

V. Analysis 10

VI. Limitations and Conclusion 13

VII. References 15

I. Introduction

The **El Farol Bar Problem**, that we chose to modelize, was introduced by the economist **W. Brian Arthur** in **1994**. It is a theoretical concept that explores the challenges individuals face in making decisions when confronted with uncertainty in a social context. The problem is named after a fictional bar in Santa Fe, New Mexico, called El Farol, which has a limited capacity.

In this scenario, individuals must decide whether to go to the bar on a given night, knowing that the enjoyment of the evening depends on the crowd size. If too many people attend, the bar becomes overcrowded and unpleasant; if too few attend, it's boring. Each individual decision is influenced by their expectations of how many others will show up.

The El Farol Bar Problem highlights the complexities of decision-making in situations where individuals must consider the choices and expectations of others. It has implications for various fields, including economics, game theory, and behavioral science, as it challenges traditional models that assume individuals make decisions in isolation without considering the actions of others. The problem has been used to study topics such as coordination, cooperation, and the emergence of patterns in social systems.

If this problem cannot be solved by homogenous rational agents - either they all go, or all stay at home, real people coordinate on such problems every day. How will reinforcement learning agents adapt ?

We are answering this question by creating a model of this situation and an agent that has to choose to go or not to go considering a reward curve depending on the number of people in the bar. Although the model is really simple, we implement an extension to add realistic settings. Here, the extension we chose is an event every seven days that gives an extra reward to the agent if he goes, no matter how many people are there.

II. Basic model

We establish a neural network-based solution for tackling the El Farol Bar Problem within the realm of reinforcement learning.
The model comprises two primary components: **the Critic** and **the Actor**.

The Critic class defines a neural network that takes both the state and action as inputs, producing a Q-value as its output. This Q-value represents the anticipated cumulative reward associated with a specific action in a given state. The architecture of the critic network consists of three fully connected linear layers. The first two layers employ the **ReLU (Rectified Linear Unit)** activation function, introducing **non-linearity** to the model. The third layer generates the Q-value without applying any activation function. By concatenating the state and action vectors, the model captures the interaction between the state and the chosen action.
The Actor class serves as the policy function, determining the probability distribution over actions given a certain state. Similar to the Critic, the actor network features three fully connected layers. The initial two layers use ReLU activation, while the third layer employs the sigmoid activation function. The sigmoid function outputs probabilities within the [0, 1] range, making it suitable for representing action probabilities. Notably, the actor network takes only the state as input and outputs a probability distribution over actions.
Both the Critic and the Actor networks employ a three-layer feedforward neural network architecture. The inclusion of activation functions (ReLU and sigmoid) introduces non-linearity to the model, enabling it to capture intricate relationships between inputs and outputs. These networks are designed to operate within an actor-critic reinforcement learning framework, where the actor suggests actions based on the learned policy, and the critic evaluates these actions in the given state.

Finally, the model parameters, such as input size, hidden size, and output size, can be fine-tuned based on the specific requirements of the El Farol Bar Problem or other analogous reinforcement learning tasks.

In our code, we defined a reward function for the agent in the context of the El Farol Bar Problem. This function evaluates the individual actions of agents based on the overall crowd attendance in the bar. The objective is to encourage the agent to make decisions that align with the optimal crowd size of 60, first without the extension and then when we consider that an event night occurs every 7 days.
The reward function starts by calculating the cumulative crowd attendance based on the individual actions. It then assigns rewards to each agent according to predefined criteria. The reward structure encourages a balanced attendance, with penalties for overcrowding or underutilization of the bar (**Figure 1**)

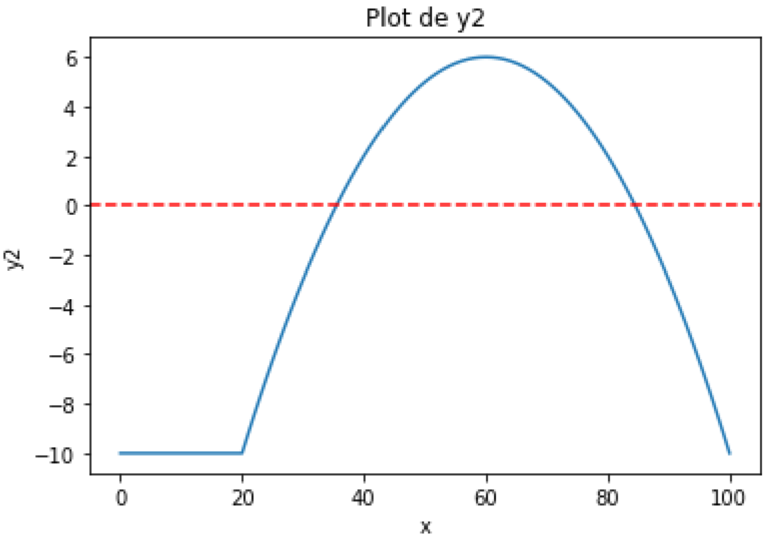

Figure 1 : *reward curve for the basic model*

Notably, the rewards incorporate randomness by adding noise generated from a normal distribution for the event night. This introduces variability in the rewards, simulating the unpredictable nature of real-world scenarios.

Upon closer inspection, it appears that the reward values are predominantly negative, especially in scenarios of extreme overcrowding or underutilization. This suggests a tendency for the agent to receive penalties in these situations.

Despite the nuanced reward design, it's mentioned that during the training of the model, the agent consistently **converges towards a 0.6 probability of attending the bar**. Additionally, it's specified that the "event night" occurs every 7 days, implying a recurring pattern in the environment.

The observed convergence at a 0.6 attendance probability could be indicative of a stable policy learned by the agent, striking a balance between attending and avoiding the bar. The recurring event every 7 days might influence the agent's decision-making process, leading to a consistent behavior.

III. Extension

Then, we defined an extension to the El Farol Bar Problem, introducing **an event occurring every 7 days**, and comparing it with the simpler model that lacks this periodic event. The extension encompasses the reward function and the agent's decision-making process.

In the extended reward function, the impact of the event night is considered. When an event night occurs, additional rewards are provided to the agent. The rewards are primarily influenced by the crowd size, with penalties for extreme overcrowding or underutilization of the bar (basic model). Random noise, simulated by a normal distribution, introduces variability to the rewards, reflecting the unpredictability of real-world scenarios.

The step function simulates the agent's decision-making process for each of the 100 individuals. The action probability determines whether an individual chooses to go to the bar or not. The resulting crowd attendance contributes to the reward calculation. Additionally, if it's an event night, an extra reward is added. The agent's state is updated, and the episode is marked as done.
The extension is implemented within the Memory class, where experiences are stored in a buffer for training the agent. Experiences consist of the current state, action, reward, next state, and a flag indicating the episode's completion.

Comparing the extended model with the simpler one (without the 7-day event), it is noted that both models lead the agent to converge toward a 0.6 probability of attending the bar. However, with the extension, there is a slightly higher variance and more pronounced extremes during the initial exploration phase. This suggests that the periodic event introduces additional complexity and variability to the agent's learning process, impacting its exploration and decision-making. The agent may need to adapt to the periodicity of the events, leading to a more nuanced policy compared to the simpler model.

In principle, the original solution concept of the basic model can be applied to the extended model with the introduced periodic event. The core idea of the El Farol Bar Problem remains consistent: agents aim to make decisions on attending the bar based on the historical attendance patterns and the current situation.
However, the introduction of the periodic event every 7 days in the extended model adds an additional layer of complexity. The agents now need to adapt to the recurring pattern of events, potentially influencing their decision-making strategies. The original solution concept would need to account for this periodicity and its impact on the agent's learning and exploration process.
The extended model, with its additional dynamics, may require adjustments to the original solution concept to effectively capture and exploit the patterns introduced by the periodic event. This could involve refining the reward structure, adjusting the learning parameters, or employing specific strategies to handle the recurring nature of the events.
In summary, while the basic solution concept can serve as a foundation for the extended model, the introduction of the 7 day periodic event necessitates an extension or modification

of the solution approach to address the new dynamics and enhance the agent's ability to make optimal decisions.

IV. Implementation

In this fourth part, let’s describe more in detail our code and our implementation of the model.
Our code presents an implementation of the **Deep Deterministic Policy Gradient (DDPG)** algorithm to tackle the El Farol Bar Problem. This problem is a multi-agent system where individuals decide whether to attend a bar or not based on past attendance patterns. The DDPG algorithm is a model-free, off-policy reinforcement learning approach suitable for continuous action spaces.

The model structure consists of two neural networks: the Actor and the Critic. The Critic evaluates the quality of the action taken by the Actor by estimating the Q-value, a measure of the expected cumulative reward. Both networks employ three fully connected layers, with the Critic taking the concatenation of state and action as input, while the Actor takes only the state. The hidden layers utilize Rectified Linear Unit (ReLU) activation, and the Actor outputs a sigmoid-activated probability representing the action.

The simulation process revolves around a modified El Farol Bar environment encapsulated within the *NormalizedEnv* class. This class handles the reward computation, event night occurrence, and the agent's decision-making process. The event night, occurring probabilistically every 7 days, influences the reward structure and adds a temporal dimension to the problem.

The DDPG algorithm follows a standard reinforcement learning pipeline. During exploration, the agent may take random actions with a probability defined by an exploration rate. In exploitation, the Actor suggests actions based on the current state. The experiences (state, action, reward, next state, done) are stored in a replay memory buffer.

The training process involves sampling batches from the replay memory and updating the Critic and Actor networks iteratively. The Critic's objective is to minimize the **Mean Squared Error (MSE)** loss between predicted Q-values and target Q-values, while the Actor aims to maximize the Q-value for the current state-action pair. Target networks, used to stabilize training, are updated through a soft update mechanism.

The neural networks architectures and hyperparameters are defined in the DDPGagent class. The *hidden_size* parameter determines the size of the hidden layers, while *actor_learning_rate* and *critic_learning_rate* control the learning rates for the Actor and Critic networks, respectively. The discount factor for future rewards (*gamma*), the soft update parameter (*tau*), and the maximum size of the replay memory buffer (*max_memory_size*) are other critical parameters influencing the training process.

The training loop spans a specified number of episodes, during which the agent iteratively interacts with the environment, updates the replay memory, and refines its policy. The exploration rate decreases over episodes, balancing exploration and exploitation. The final plot displays the actions taken by the agent throughout training, providing insights into the learned behavior over time.

This DDPG-based solution addresses the El Farol Bar Problem's challenges, and the periodic events occurring every 7 days introduce additional complexity to the learning process, potentially influencing the agent's decisions and exploration strategy.

V. Analysis

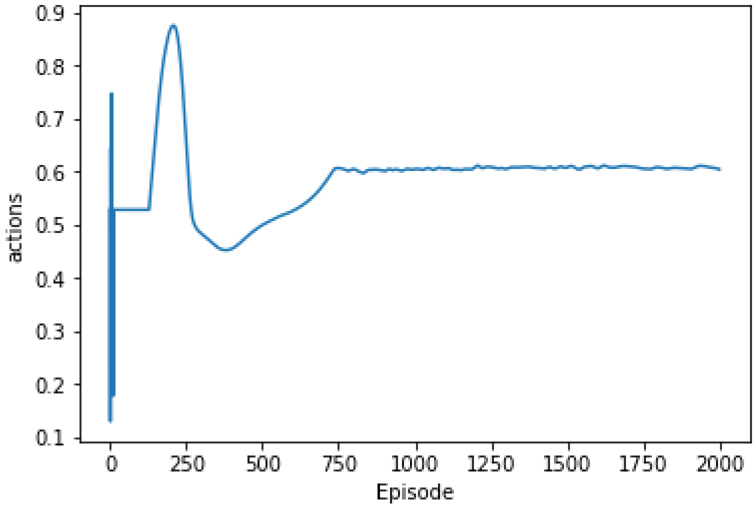

Figure 2 : *Results of the basic model, 2000 episodes*

In **Figure 2**, we observe strong variations from episode 0 to 750. From 750 and onwards the model stabilizes around 0.6, which is what we're looking for if we want an optimal crowd of 60.

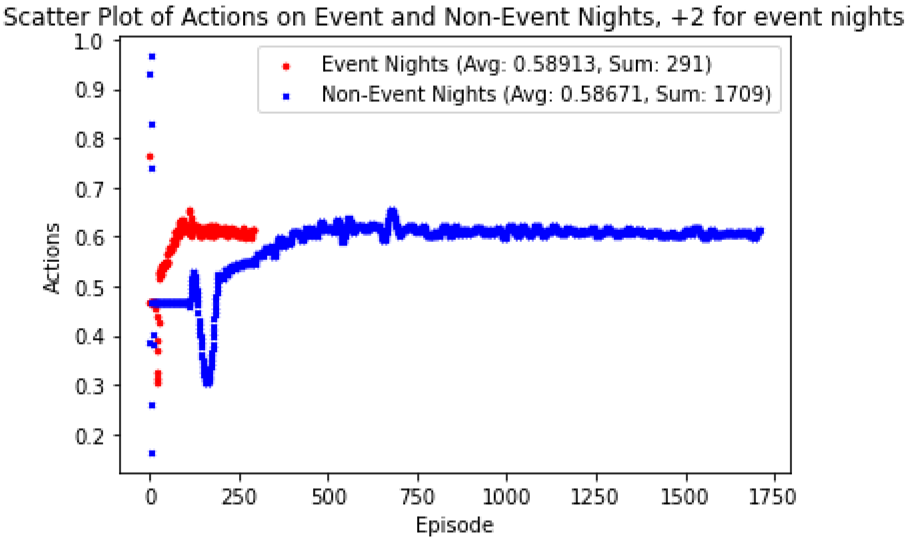

Figure 3: Results for +2 reward for Event nights, 2000 episodes

From **Figure 3** and forth we added the extension of a 'Event Night'. We added +2 to the reward for event nights to encourage and push for a higher attendance. In red we can observe the results for event nights and in blue regular nights. The odds of going to an event night is slightly higher than for a regular night as we can see, 0.589 > 0.586. Nothing too dramatic though.

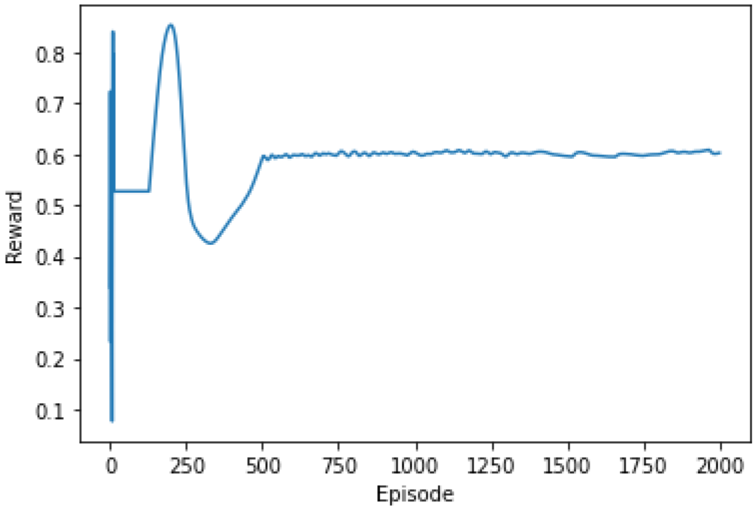

Figure 4: Results for +2 reward + noise (normal distribution) for Event nights, 2000 episodes

In **Figure 4**, we added 2 and some noise (normal distribution) to the reward for attending event nights. Besides stronger variations in the learning curve, there doesn't seem to be a strong impact on the attendance for event nights. Adding the noise to the reward is supposed to create some uncertainty.

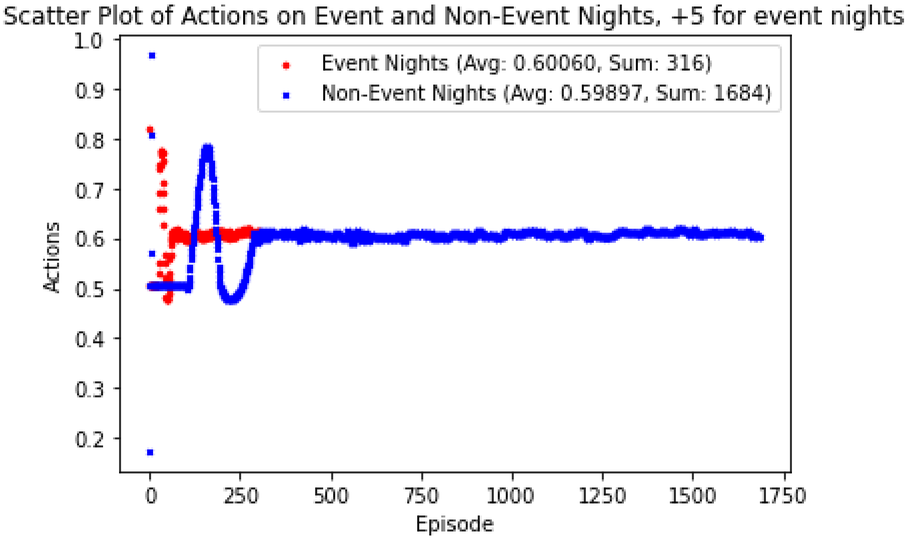

Figure 5: Results for +5 reward for Event nights, 2000 episodes

In **Figure 5**, we added +5 to the reward for event nights to encourage and push for a higher attendance. In red we can observe the results for event nights and in blue regular nights. The odds of going to an event night is slightly higher than for a regular night as we can see, 0.600 > 0.598. The difference remains marginal.

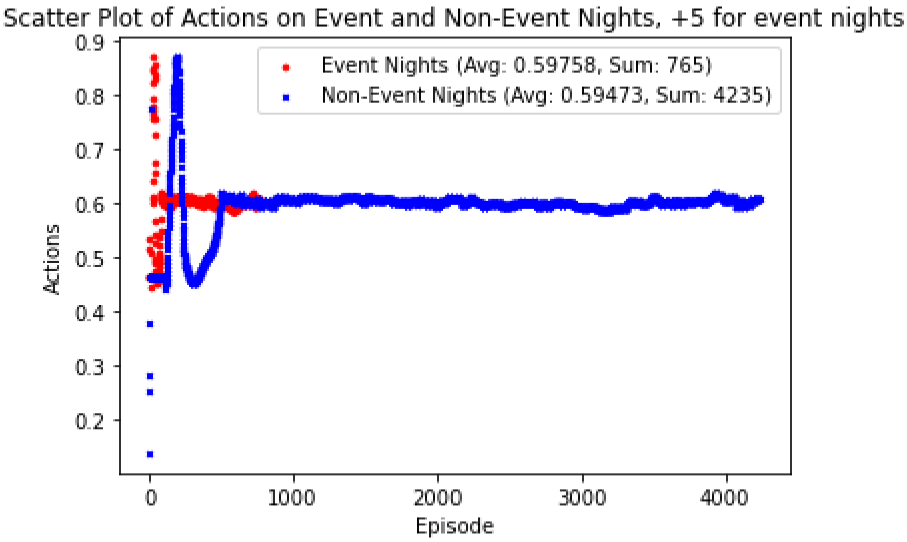

Figure 6: Results for +5 reward for Event nights, 5000 episodes

In **Figure 6**, we kept the same reward system as in **Figure 5** and just increased the number of episodes. Same story as in the odds of going to an event night remain slightly higher than for a regular night as we can see, 0.597 > 0.594. The difference remains negligible.

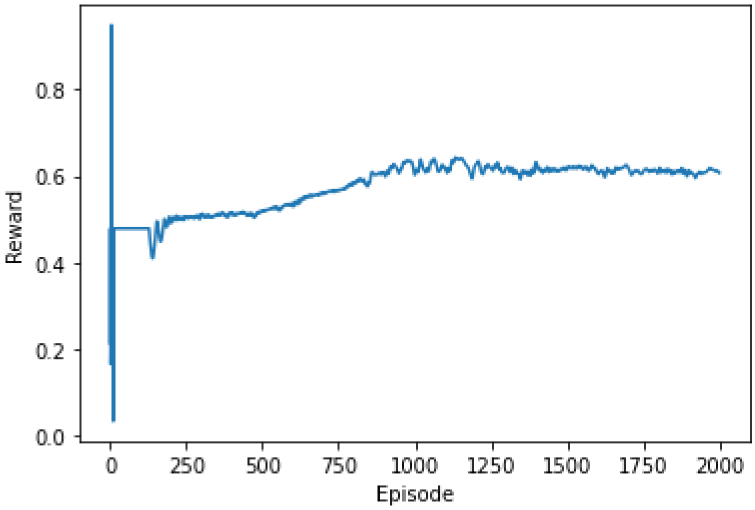

Figure 7: Results for +5 reward + noise (normal distribution) for Event nights, 2000 episodes

In **Figure 7**, we added 5 and some noise (normal distribution, to create some uncertainty) to the reward for attending event nights. We observe that it takes a longer amount of episodes to stabilize at around 0.6. We also observe stronger variations in the odds of going out to the bar between episodes 800 and 1500.

VI. Limitations and Conclusion

The DDPG-based solution for the El Farol Bar Problem demonstrates promising outcomes, yet several aspects need consideration for improvement. The convergence of the algorithm and adaptability of the policy throughout training episodes represent notable strengths. The selected neural network architectures, featuring appropriately structured hidden layers and activation functions for the Actor and Critic networks, effectively capture the decision-making dynamics inherent to the problem.

The incorporation of a replay memory buffer contributes to the stability and efficiency of learning, addressing potential challenges related to correlated experiences !

Additionally, the implementation's soft update mechanism for target networks aids in stabilizing the training process by gradually transferring knowledge from the main networks. The extension of the model to accommodate periodic events occurring every 7 days showcases the adaptability of the DDPG approach to variations in the environment.

However, certain limitations and areas for potential improvement exist. The exploration strategy relies on a relatively simple epsilon-greedy approach, which may not be exhaustive for exploring the action space. Incorporating more advanced exploration techniques, such as parameter noise or sophisticated exploration policies, could enhance the model's exploration capabilities, kind of like what we did with the reward of the event night.

The reward function, while capturing the essence of the El Farol Bar Problem, might benefit from additional complexity. A more nuanced reward function considering individual preferences or contextual information could potentially enhance the model's decision-making abilities (Why would everybody want to go to the bar ? Maybe some agents would prefer to stay home on a Friday or would only be able to go on the weekends). The sensitivity of the model to hyperparameters, including learning rates, hidden layer sizes, and exploration rates, emphasizes the importance of systematic fine-tuning for optimal performance.

The model's assumption of periodic events occurring probabilistically every 7 days may oversimplify the dynamic nature of real-world scenarios. Exploring more realistic and dynamic event occurrence patterns could add further complexity and realism to the problem. Moreover, the performance of the model is inherently tied to hyperparameter configurations, necessitating careful and systematic tuning.

In terms of potential enhancements without resource constraints, advanced exploration strategies, such as intrinsic motivation or entropy regularization, could be incorporated for more effective exploration. Additionally, exploring ensemble methods where multiple actors and critics operate in parallel could introduce diversity and enhance the model's robustness.

Reflecting on the implementation, it underscores the complexity of real-world scenarios where individual decisions collectively influence outcomes. The El Farol Bar Problem serves as a reminder of the challenges associated with modeling such scenarios and the importance of adapting to dynamic changes in the environment. The delicate balance between exploration and exploitation becomes apparent, especially in the face of uncertain events.

In conclusion, while the DDPG-based solution demonstrates commendable performance, ongoing refinement and experimentation, informed by a deeper understanding of the problem domain, are essential for achieving optimal results. Continuous efforts toward incorporating advanced strategies and refining model components will contribute to the model's overall effectiveness in addressing the complexities of the El Farol Bar Problem.

VII. References

Arthur, B. 1994. Inductive reasoning and bounded rationality. The American Economic Review, 84(2):406-411.

Lillicrap et al. 2016. Continuous Control With Deep Reinforcement Learning.

Silver et al. 2014. Deterministic Policy Gradient Algorithms.