Are exploration-based strategies of interest for repeated stochastic coalitional games?

Josselin Guéneron, Grégory Bonnet

Are exploration-based strategies of interest for repeated stochastic coalitional games?

Josselin Guéneron[1] and Grégory Bonnet[1]

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC
{firstname.name}@unicaen.fr
https://www.greyc.fr/

Abstract. Coalitional games are models of cooperation where selfish agents must form groups (coalitions) to maximize their utility. In these models, it is generally assumed that the utility of a coalition is fixed and known. As these assumptions are not realistic in many applications, some works addressed this problem by considering repeated stochastic coalitional games. In such games, agents repeatedly form coalitions and observe their utility a posteriori in order to update their knowledge. However, it is generally assumed that agents have a greedy behavior: they always form the best coalitions they estimate at a given time step. In this article, we study if other strategies (behaviors) that allow agents to explore under-evaluated coalitions may be of interest. To this end, we propose a model of repeated stochastic coalitional game where agents use a neural network to estimate the utility of the coalitions. We compare different exploration strategies, and we show that, due to the structure of the coalitional games, the greedy strategy is the best despite the fact exploration-based strategies better estimate the utilities.

Keywords: Sequential decision · Coalition formation · Cooperative game theory · Multi-agent systems.

1 Introduction

In multi-agent systems, individual agents are not always able to realize some tasks on their own. In such case, they can decide to cooperate with each other in forming coalitions, i.e. forming groups of agents able to realize a given task, and sharing the gains generated afterwards. As agents are selfish and rational, they will try to earn as much as possible, and can refuse to form certain coalitions deemed uninteresting for themselves. In the literature, the majority of works about coalition formation makes two strong hypotheses. The first one is that agents have perfect *a priori* knowledge of their payoff when forming a given coalition. The second hypothesis is that this payoff is deterministic. These two hypotheses do not seem to fit with real situations where the exact payoff brought by a coalition is known only a posteriori. Moreover, if this coalition is formed again, this payoff may not always be the same, if the agents are more or less efficient in their tasks. For example, we can consider scientists having to repeatedly form consortia in order to temporarily work on projects. These consortium

formations are repeated with the same pool of scientists but the quality of the results produced by each consortium may vary due to internal factors. For example, a internal factor can be the individual skills of the scientists and their ability to interact better with some rather than others. Moreover, externalities independent of the consortia formed may also stochastically impact the quality of the result. In the literature, some works proposed to relax these hypotheses by considering repeated stochastic coalitional games. The agents play the same game – and thus form coalitions – repeatedly. They observe the payoff they obtain and use this information to estimate the value of each coalition at the next time step. However, in those works agents use greedy strategies: they form the coalitions they estimate the best. We can thus wonder if exploration-based strategies, which are successful in other contexts, may be of interest in the coalition formation domain. We then propose in this article a high-level repeated stochastic coalition formation model, and we experimentally assess the performance of several strategies compared to a greedy strategy. We finally highlight that, due to the structure of the coalitional games, the greedy strategy remains the best. This article is structured as follows. In Section 2, we present the basic notions related to coalitional games, then we review some works both on stochastic characteristic functions and repeated coalitional games. In Section 3, we describe our repeated stochastic coalitional game model, and detail how agents learn the characteristic function, and the different strategies they can use. Finally, Section 4 is devoted to evaluate these strategies.

2 State of the art

We present here the basic notions about coalitional games [15], repeated coalitional games [1, 2, 4] as well as stochasticity in coalitional games [4, 7, 11].

2.1 Coalition formation

In a coalition game, a set of agents is partitioned into separate *coalitions* which produce an amount of *utility*. Such partition is called a *coalition structure*.

Definition 1 (Coalitional game). *A* game *is a tuple* $\mathcal{G} = \langle N, v \rangle$ *where:*

- $N = \{a_1, \ldots, a_n\}$ *is a set of agents,*
- $v : 2^N \to \mathbb{R}$ *is a characteristic function that assigns a real value to each coalition, called the coalition utility and denoted* $v(C_k)$ *where* $C_k \subseteq N$*.*

We consider in this article coalitional games with *transferable utility*, i.e. where agents must decide how to distribute the coalition utility among its members [9]. A *solution* to such a game is defined as follows.

Definition 2 (Solution). *A* solution *to* $\mathcal{G}$ *is a tuple* $S_\mathcal{G} = \langle \mathcal{C}, \boldsymbol{x} \rangle$ *where:*

- $\mathcal{C}$ *is a coalition structure of* N*,*
- $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ *is a payoff vector where* $x_i \geq 0$ *is the payoff of agent* a_i*.*

As agents are selfish, when a solution is proposed, all of them must accept it, i.e. that they must not wish to form or join another coalition where they could earn more. This is why we are interested in solutions which belong to a *solution concept*. A solution concept is the set of solutions that respect a certain notion of stability. While many concepts have been proposed in the literature such as the nucleolus or the kernel [5], we focus in this article on the concept of core and its generalization, the ϵ-*core* [14, 19]. The core is the set of solutions $\langle \mathcal{C}, \boldsymbol{x} \rangle$ for which no other coalition that could be formed produces a sum of gains greater than that which agents obtain with $\boldsymbol{x}$. The ϵ-core allows agents to make a concession, i.e. agree to reduce their gain by ϵ, in order to form a stable coalition structure.

Definition 3 (ϵ-core). *A solution $(\mathcal{C}, \boldsymbol{x})$ belongs to the ϵ-core if and only if:*

$$\forall C \subseteq N, x(C) \geq v(C) - \epsilon \text{ with } x(C) = \sum_{i \in C} x_i$$

The ϵ-core allows to define the *least core*, which contains all ϵ-core solutions for the smallest value of ϵ that make the solution concept non-empty.

2.2 Stochastic characteristic functions

In the literature, some works have proposed stochastic coalitional games [4, 7, 8, 11]. The nature of uncertainty in these models differs. For instance, Ieong and Shoham proposed a probability distribution on worlds representing coalitional games, each of them having a deterministic characteristic function [11]. Chalkiadakis and Boutilier considered a deterministic characteristic function modeled in a stochastic environment with partially observable Markovian decision processes [4]. Agents have beliefs about capabilities of other agents and the same coalition structure can lead to different world states. Charnes and Granot simply considered that the value of a coalition is a random variable [7, 8]. The characteristic function is then rewritten as $v : 2^N \rightarrow \mathcal{X}_{2^N}$. Thus, when a coalition is formed, the utility produced is determined by the random variable, that follows a normal distribution. In this model, they compute their payoff vectors by associating to each agent in a coalition an equal part of the expectation of the random variable associated to the coalition. In this article, we position ourselves in the continuity of the Charnes and Granot's work. Indeed, their model allows us to deal with the heterogeneity of stochasticity (both internal and external factors as cited in Section 1) through the use of a single random variable.

2.3 Repeated coalitional games

If we also relax the hypothesis of perfect knowledge of the characteristic function, whether it is stochastic or not, it becomes interesting to move on to a repeated game [1, 2, 4, 12]. For instance, Konishi and Debraj [12] have shown that repeated coalition formation processes converge towards equilibria if agents sequentially form Pareto-efficient coalition structures. Moreover, repeated coalitional games

allow to observe the utilities when coalitions are formed in order to learn an estimation of the characteristic function. Agents can then use this estimation, and can be able to find an optimal stable solution over time. Models in the literature essentially differ on the nature of what the agents learn and how they estimate the coalitions. For instance, Blankenburg *et al.* [2] learn a reliability value for each agent, which impact the utility of coalitions. In Chalkiadakis and Boutilier [4], agents learn both the others' skills and a stochastic transition between a given coalition structure and the states it may reach (and therefore a payoff). In all those works, agents use a greedy strategy: they form the coalition structure which is estimated the best at each time step. As in other sequential decision-making problems it has been demonstrated that there is an interest to explore, i.e. making *a priori* sub-optimal decision in order to acquire knowledge [10, 13], we study in this article if exploration-based strategies are efficient in the context of coalition formation.

3 A general model of repeated stochastic coalitional game

First of all, let us define a model of repeated stochastic coalitional game.

3.1 Game and solution

Definition 4 (Repeated stochastic coalitional game). *Let* $\mathcal{G} = \langle N, \mathbb{T}, v \rangle$ *be a* repeated stochastic coalitional game *(RSCG) where:*

- $N = \{a_1 \dots a_n\}$ *is a set of agents,*
- $\mathbb{T} \subset \mathbb{N}^+$ *is a set of distinct time steps,*
- $v : 2^N \to \mathcal{X}^{2^N}$ *is a characteristic function that associates a random variable to each coalition. For a given coalition* $C \subseteq 2^N$*, we note* $v(C) = \mathcal{X}^C$*. This characteristic function is unknown to the agents.*

At each time step, agents in N have to decide on a solution to the game, despite the fact that they do not know the characteristic function v *a priori*. A solution is, like in a deterministic context, a tuple made of a coalition structure and a payoff vector. Here, the payoff is an *ex ante* payoff, i.e. the estimated payoff based on what the agents know about v.

Definition 5 (Solution to a RSCG). *A* solution S^t *at the time step* $t \in \mathbb{T}$ *to a RSCG* $\mathcal{G}$ *is a tuple* $S^t = (\mathcal{C}^t, \boldsymbol{x}^t)$ *such as:*

- $\mathcal{C}^t$ *is a coalition structure (disjointed partition) of* N*,*
- $\boldsymbol{x}^t = \{x_1^t, \dots, x_n^t\}$ *is a payoff vector such as* $x_i^t \geq 0$ *is the gain of the agent* a_i *calculated according to the estimated value of the coalition of which he is a part in the structure* $\mathcal{C}^t$*.*

3.2 Coalition formation process

We consider the following process:

1. Agents are initialized with an *a priori* knowledge about the game, i.e. an estimation of the characteristic function, which may reflect either ignorance, or an expert-knowledge (e.g. larger coalitions produce a higher value);
2. Agents form a coalition structure according to a given strategy based on their current knowledge of the characteristic function (see Section 3.4);
3. Agents observe the payoff they obtain by forming the structure, and they update their knowledge of the characteristic function (see Section 3.3). We assumed that all agents observe the payoff produced by each coalition: hence they have the same knowledge, and consequently the same estimation;
4. The process is repeated from step 2.

3.3 Estimating the characteristic function

As we assume the *ex-post* payoffs are observed by all the agents, we denote by X_t^C the observation of the payoff of coalition C at the time step t.

Definition 6 (Observations). *Let* $\mathcal{O}_t = \{(C, t', X_{t'}^C)) : C \subseteq 2^N, t' \in \mathbb{T}, t' < t\}$ *be a set of observations at time step* t *corresponding to the set of the coalitions formed at each time step before* t *and their ex-post payoffs. Knowing a solution* S^t *of a RSCG at time step* t, $\mathcal{O}_{t+1} = \mathcal{O}_t \cup \{(C, t, X_t^C)) : \forall C \in \mathcal{C}^t\}$.

Thereafter, let us note $\mathcal{O}_t(C)$ the set of observations at time step t associated with the coalition $C \subseteq 2^N$. Then the agents can use different methods to estimate the future payoff, e.g. tabular representation, bayesian network, or neural network. In order to remain general, we simply consider that the agents know a function that produces an estimation according to the observations. Such function has to be instanciated (see experiments in Section 4).

Definition 7 (Payoff estimation). *Let* $\mathbb{E}(C, t)$ *be the payoff estimation of a coalition* $C \subseteq 2^N$ *at time step* $t \in \mathbb{T}$.

3.4 Decision strategies

Once the agents have estimated the characteristic function, they need to decide which coalitions to form, according to a given strategy that take exploration into account. We consider two kinds of strategies: ϵ-greedy strategies (also known as semi-uniform strategies) and contextual strategies. Adapted in the context of coalition formation, ϵ-greedy strategies are strategies where agents choose the best coalition structure according to the least core solution concept with a given probability, or choose a random coalition structure otherwise.

Definition 8 (ϵ-greedy strategy). *The* ϵ-greedy strategy *selects a solution from the least core solution concept with a probability of* ϵ*, or a solution drawn uniformly at random among all solutions otherwise.*

Obviously, when ϵ is set to 1, the ϵ-greedy strategy becomes a simple greedy strategy as considered in the literature [1, 2, 4, 12]. When ϵ is set to 0, the agents choose their coalition structure uniformly at random among all coalition structures. Contextual strategies are strategies where agents value the information they can gain as they value a payoff. We consider firstly a strategy, we called UCB-core strategy, inspired from UCB strategies in multi-armed bandits problem [10, 13]. Information value is a bias defined as follows.

Definition 9 (Exploration bias). *Let $\gamma(C,t) : 2^N \mapsto \mathbb{R}$ a bias such as:*

$$\gamma(C,t) = \sqrt{\frac{2.\log(|\mathcal{O}_t|+1)}{|\mathcal{O}_t(C)|+1}}$$

We now adapt the UCB-strategy in the context of coalition formation. To this end, we consider a variant of the ϵ-core solution concept, called the UCB-core.

Definition 10 (UCB-core). *A solution $S^t = (\mathcal{C}^t, \boldsymbol{x}^t)$ belongs to the* UCB-core solution concept *if, and only if:*

$$\forall C \in N, x^t(C) + \Gamma(C,t) \geq \mathbb{E}(C,t) - \epsilon + \gamma(C,t),$$

with:

$$x^t(C) = \sum_{a_i \in C} x_i^t \quad \textit{and} \quad \Gamma(C,t) = \sum_{a_i \in C} \frac{\gamma(C_{a_i},t)}{|C_{a_i}|},$$

where C_{a_i} is the coalition of a_i in $\mathcal{C}^t$

The previous definition means a coalition structure is stable if, and only if, there is no coalition such that its payoff plus its exploration bias is higher than what its members earn currently in the coalition structure, knowing the value given by the exploration bias is equally shared between agents. Hence,

Definition 11 (UCB-core strategy). *The* UCB-core strategy *selects a solution uniformly at random from the non-empty UCB-core solution concept with the smallest ϵ.*

The UCB-core strategy may allow solutions that are irrational for some agents, i.e. solutions where the payoff of at least one agent is lesser than the payoff he would have received alone. As rationality is an important concept in coalition formation, we propose another contextual strategy that preserves the rationality, called the *δ-core strategy*. The idea is to allow agents to sacrifice a part of their surplus, i.e. the part of the payoff they received in excess of their singleton coalition, proportional to the exploration biais.

Definition 12 (Surplus). *Let $\Omega^t(a_i, S^t)$ be the surplus of the agent a_i for a given solution S^t at time step $t \in \mathbb{T}$. This surplus is computed as*

$$\Omega^t(a_i, S^t) = x_i^t - \mathbb{E}(C,t)$$

where $C = \{a_i\}$, i.e. the singleton coalition of the agent a_i. If the surplus is negative, it means that the given solution is irrational *for agent i, so it will never be stable.*

Secondly, let us consider a normalized exploration bias.

Definition 13 (Normalized exploration bias).

$$\zeta(C,t) = \frac{\gamma(C,t)}{\max\limits_{\forall C' \subseteq 2^N}(\gamma(C',t))}$$

Once the normalization of exploration factors done, agents can compute how much of their surplus they accept to not gain. Thus, the payoff an agent can sacrifice is given by:

Definition 14 (Sacrificable payoff). *The* sacrificable payoff *for a agent a_i and a given solution S^t at time step t is given by $\delta^t(a_i, S^t) = \Omega^t(a_i, S^t) \times \zeta(C,t)$ where C is the coalition of a_i in S^t.*

We can now define the δ-core solution concept. In this solution concept, a coalition structure is stable if there are no coalitions – that are not part of the structure – whose estimated payoff, minus the payoff that agents accept not to earn to form the structure, is greater than the estimated payoff the agents obtain with the structure.

Definition 15 (δ-core). *The solution $S^t = (\mathcal{C}^t, \boldsymbol{x}^t)$ belongs to the* δ-core *solution concept if, and only if:*

$$\forall C \in N, x^t(C) \geq \mathbb{E}(C,t) - \epsilon - \Delta^t(C),$$

where:

$$x^t(C) = \sum_{a_i \in C} x_i^t \quad \textit{and} \quad \Delta^t(C) = \sum_{a_i \in C} \delta^t(a_i, S^t)$$

Hence obviously,

Definition 16 (δ-core strategy). *The* δ-core strategy *selects a solution uniformly at random from the δ-core solution concept.*

4 Experimentations

To compare the different strategies given above, we generate random games where agents play repeatedly and observe the evolution of the chosen solutions.

4.1 Experimental protocol

We generate games with 5, 6 and 7 agents, thus for 52, 203, 877 possible coalition structures respectively. The stochastic characteristic functions of those games are generated with normal distributions whose their μ parameter is drawn from the normal, uniform and NDCS model proposed in [16, 17]. Hence for each coalition $C \subseteq N$, the μ parameter is $|C| \times \mathcal{N}(1, 0.1)$ for normal models, $|C| \times \mathcal{U}(0,1)$ for

uniform models and $\mathcal{N}(|C|, \sqrt{|C|})$ for NDCS models. The variance σ associated to each coalition is given by $\sigma = \mathcal{U}(0, \frac{\mu}{2})$. As the maximal variance is related to μ and as μ is higher with large coalitions, the larger a coalition, the higher its variance may be.

We now need to instantiate the payoff estimation function (Definition 7). In order to be general and abstract, we use in the experiments a neural network with two hidden layers. Each layer is a dense layer with a ELU activation function. The input layer represents coalitions with one neuron dedicated to each agents: input of 1 for his presence in the coalition, 0 otherwise. The output layer consists in a single neuron that produces a real value. Such neural network is able to learn non-linear functions. Here, we use stochastic gradient descent with adaptive moment estimation to train the network. The loss function is the mean square error. It is important to notice that this network is not trained offline before playing the repeated game, but trained at the runtime. Each time the agents observe a payoff (see the process described in Section 3.2), the network is trained with a set of examples (coalition, payoff).

For each kind of model (normal, uniform or NDCS) we perform 1000 runs with different characteristic functions each time (but their type does not change) where the agents play over 100 time steps. We made two experiments: the first one compares the performances of the ϵ-greedy strategy when ϵ varies; the second one highlights the performances of the UCB-core and δ-core strategies compared to the greedy and the random strategy.

4.2 Performance measure

In order to evaluate our model, we measure both the efficiency of the decisions taken (seen as the optimality of the stable solutions found) over time, and the accuracy of the estimated characteristic function. The efficiency is measured from the *instant regret* at step t, which is the sum of the differences between the maximum social welfare (the maximum sum of the real expected utilities of the coalition structures) and the sum of the actual expected utilities of the coalitions of the structure formed at time step t. Formally, instant regret is defined as:

Definition 17 (Instant regret). *At one time step t, let $S^* = (\mathcal{C}^*, \boldsymbol{x}^*)$ be the optimal solution, the instant regret at this time step, noted R^t, is defined by:*

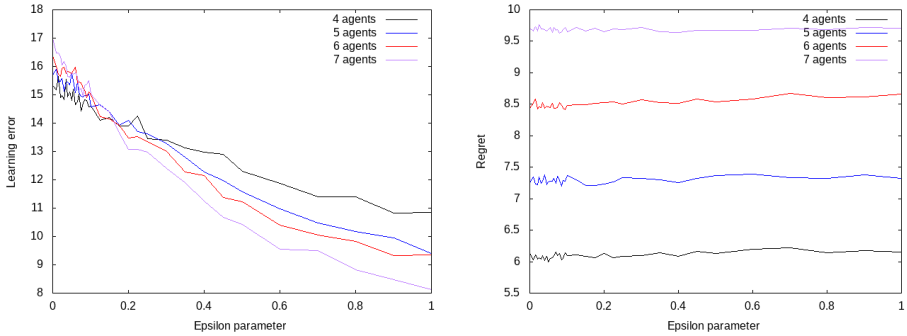
$$R^t = \sum_{C^* \in \mathcal{C}^*} \mu_{C^*} - \sum_{C \in \mathcal{C}^t} \mu_C$$

As instant regret can oscillate due to stochasticity, we consider in the sequel the *cumulative regret*, i.e. the sum of instant regrets from the beginning of an experiment to a given time step. The accuracy of the estimated characteristic functions is given by the *mean absolute error* (MAE) measure over the coalitions.

Definition 18 (Mean absolute error). *The distance D^t_{MAE} between two characteristic functions at time step t is defined by:*

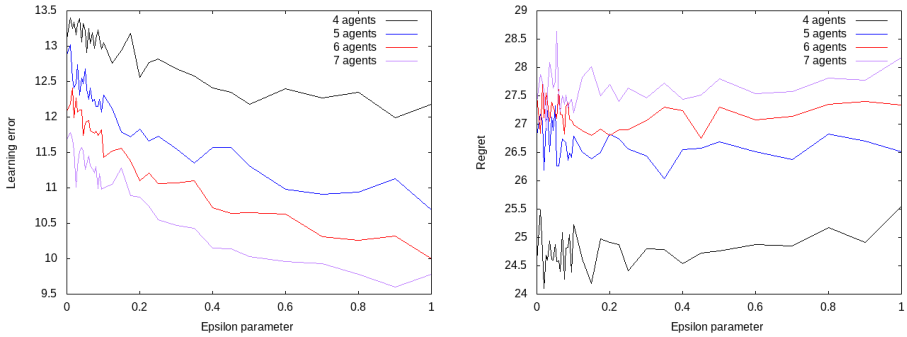
$$D^t_{MAE} = \frac{\sum_{C \in 2^N} |\hat{v}(C) - v(C)|}{|2^N|}$$

4.3 Results



(a) Final mean absolute error (b) Cumulative regret

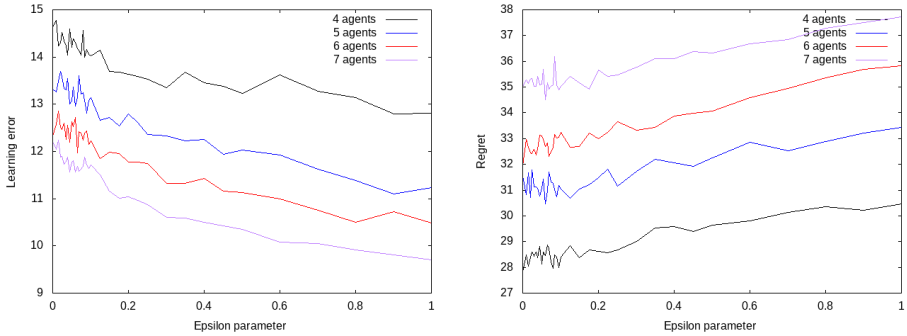
Fig. 1: Results for the ϵ-greedy strategy on normal characteristic functions



(a) Final mean absolute error (b) Cumulative regret

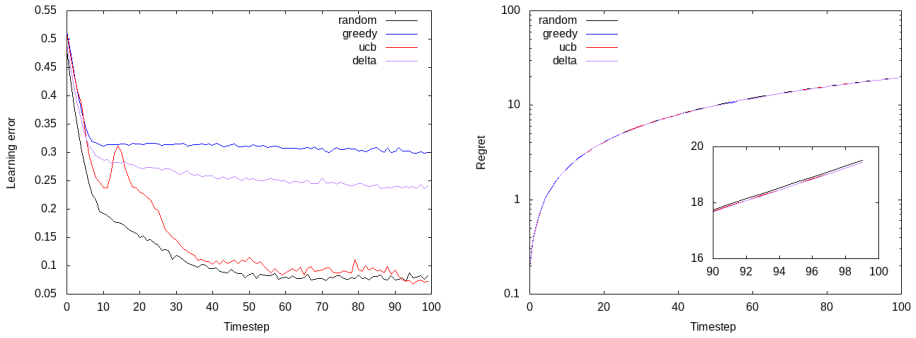
Fig. 2: Results for the ϵ-greedy strategy on uniform characteristic functions

Figures 1, 2 and 3 show respectively for normal, uniform and NDCS characteristic functions the MAE and regret at the end of the experiment for the ϵ-greedy strategy, when ϵ vary between 0 and 1, and when the number of agents increase. The oscillations when ϵ is low are due to the higher number of data points. Independently of the number of agents, the learning error decreases when ϵ increases, i.e. when going from a greedy strategy to a random exploration. However, the regret remains the same or increases when ϵ increases. Thus, while semi-uniform exploration is interesting to better estimate the characteristic function as expected, it is helpless to decrease the regret in the context of coalition formation. Pure greedy strategies are still the best. Figures 4, 5 and 6 show



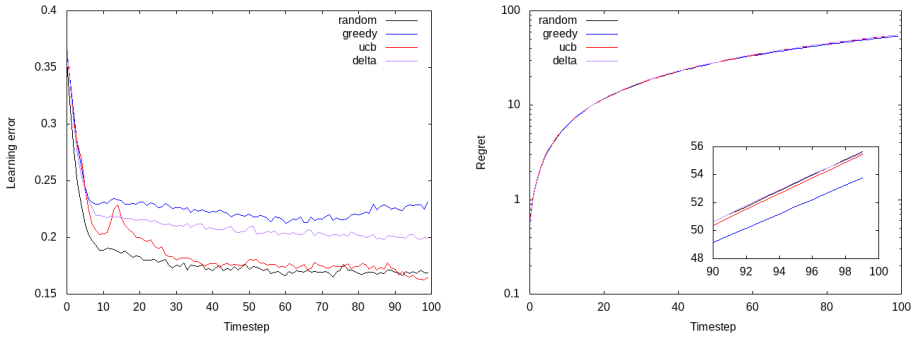
(a) Final mean absolute error (b) Cumulative regret

Fig. 3: Results for the ϵ-greedy strategy on NDCS characteristic functions



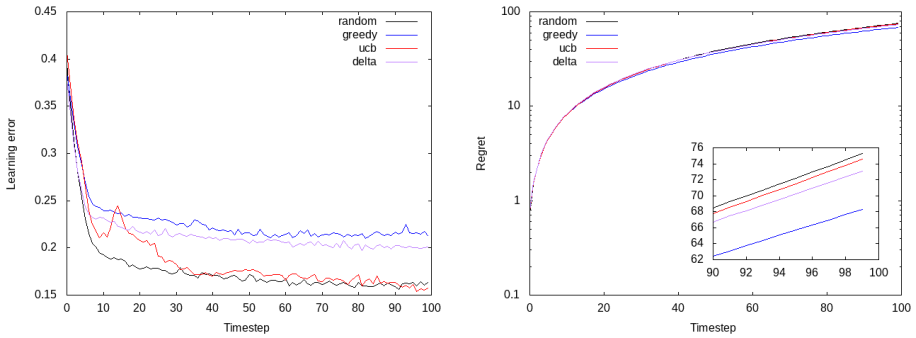
(a) Mean absolute error (b) Cumulative regret

Fig. 4: Results for 7 agents with contextual strategies on normal functions



(a) Mean absolute error (b) Cumulative regret

Fig. 5: Results for 7 agents with contextual strategies on uniform functions



(a) Mean absolute error (b) Cumulative regret

Fig. 6: Results for 7 agents with contextual strategies on NDCS functions

respectively for normal, uniform and NDCS characteristic functions the MAE and the cumulative regret for 7 agents. Due to space and readability constraints, we do not give the figures for 5 and 6 agents, but in all experiments, they present the same shapes. Concerning the MAE, as expected, the random strategy learns the best, while the greedy strategy learns the worse. UCB-core strategy is very efficient and close to the random strategy. δ-core strategy is worse but better than the greedy strategy. Concerning the cumulative regret, we can see its convergence. As the curve as close to each other, we provide a zoom on the final steps. For the NDCS model, the greedy strategy remains the best, followed by UCB-core strategy, δ-core strategy (and finally random strategy). Greedy strategy remains the best for uniform models. In the particular case of normal models, all strategies tend to be confounded.

5 Conclusion

We studied in this article repeated stochastic coalitional games, which relax hypothesis which may be too strong for real work application. In such models, agents use greedy strategies, i.e. they form at each time step the best coalition structure they estimate, form the coalitions, and update their knowledge accordingly. However, are exploration-based strategies, known to be efficient in other contexts, interesting for such games? To answer this question, we proposed a high-level model of repeated coalitional games, and experiment several strategies: ϵ-greedy strategy, UCB-core strategy and δ-core strategy. The results show that, as expected exploration-based strategies allows to better estimate the characteristic function. However, the greedy strategy remains the best for repeated coalition formation. Indeed, learning the precise value of each coalition independently is not useful in coalition formation as long as the agents correctly rank the coalitions. In terms of perspectives, these results must be consolidated on other models of characteristic function, and with a higher number of agents.

References

1. Jean-Pierre Benoit and Vijay Krishna. Finitely repeated games. *Econometrica*, Vol. 53, Issue 4, pages 905-922, 1985.
2. Bastian Blankenburg, Rajdeep K. Dash, Sarvapali D. Ramchurn, Matthias Klusch and Nicholas R. Jennings. Trusted kernel-based coalition formation. *Proc. of 4th AAMAS*, pages 989-996, 2005.
3. Georgios Chalkiadakis and Craig Boutilier. Bayesian reinforcement learning for coalition formation under uncertainty. *Proc. of 3rd AAMAS*, pages 1090-1097, 2004.
4. Georgios Chalkiadakis and Craig Boutilier. Sequential decision making in repeated coalition formation under uncertainty. *Proc. of 7th AAMAS*, pages 347-354, 2008.
5. Georgios Chalkiadakis, Edith Elkind and Michael Wooldridge. Computational aspects of cooperative game theory. *Synth. Lect. Artif. Intell. Mach. Learn.*, Vol. 5, No. 6, pages 1-168, 2011.
6. Georgios Chalkiadakis, Evangelos Markakis and Craig Boutilier. Coalition formation under uncertainty: Bargaining equilibria and the Bayesian core stability concept. *Proc. of 6th AAMAS*, pages 400-407, 2007.
7. Abraham Charnes and Daniel Granot. Prior solutions: Extensions of convex nucleus solutions to chance-constrained cames. *Texas Univ. Austin Center for Cybernetic Studies*, No. CS-118, 1973.
8. Abraham Charnes and Daniel Granot. Coalitional and chance-constrained solutions to n-person games. I: The prior satisficing nucleolus. *SIAM J. Appl. Math.*, Vol. 31, No. 2, pages 358-367, 1976.
9. Ray Debraj. A game-theoretic perspective on coalition formation. *Oxford University Press*, 2007.
10. John C. Gittins. Bandit processes and dynamic allocation indices. *J. R. Stat. Soc. Series B Stat. Methodol.*, Vol. 41, No. 2, pages 148-164, 1979.
11. Samuel Ieong and Yoav Shoham. Bayesian coalitional games. *Proc. of 23rd AAAI*, pages 95-100, 2008.
12. Hideo Konishi and Ray Debraj. Coalition formation as a dynamic process. *J. Econ. Theory*, Volume 110, Issue 1, Pages 1-41, 2003.
13. Aditya Mahajan and Demosthenis Teneketzis. Multi-armed bandit problems. *Foundations and applications of sensor management*, Springer, pages 121-151, 2008.
14. Rami Mochaourab and Eduard Jorswieck. Coalitional games in MISO interference channels : Epsilon-core and coalition structure stable set. *IEEE Trans. Signal Process.*, Vol. 62, No. 24, pages 6507-6520, 2014.
15. Oskar Morgenstern and John von Neumann. Theory of games and economic behavior. *Princeton University Press*, 1953.
16. Talal Rahwan, Tomasz Michalak, Michael Wooldridge and Nicholas R. Jennings. Anytime coalition structure generation in multi-agent systems with positive or negative externalities. *AI*, Vol. 186, pages 95-122, 2012.
17. Talal Rahwan, Sarvapali D. Ramchurn, Andrea Giovannucci, Nicholas R. Jennings. An anytime algorithm for optimal coalition structure generation. *JAIR*, Vol. 34, pages 521-556, 2009.
18. Lloyd S. Shapley. A value for n-person games. *Contributions to the Theory of Games*, Vol. 2, No. 28, pages 307-317, 1953.
19. Lloyd S. Shapley and Martin Shubik. Quasi-cores in a monetary economy with non-convex preferences. *Econometrica*, Vol. 34, No. 4, pages 805-827, 1966.