# An Overlapping Coalition Game for Individual Utility Maximization in Federated Learning

## ABSTRACT

To tackle the challenge of data heterogeneity in federated learning (FL), personalized FL has been proposed to maximize individual utility (model performance) by customizing personalized models for clients. Considering the significance of *individual rationality*, existing works have formulated clients' participation decisions problem as *hedonic* games. However, they assume that clients can participate in only one collaborative coalition, constraining players' attempts to join multiple coalitions. Different from prior works, we approach personalized FL from the perspective of hedonic *overlapping coalition formation* (OCF) games where rational clients can join multiple coalitions and generate their personalized model by weighting the local and coalition models. Nevertheless, the key challenge in analyzing the game is how to achieve a stable coalition structure where no clients would deviate from the current structure. This leads to our main question: *what does a stable OCF structure look like?* To address this problem, we first investigate the linear FL models for theoretical insights. Then, we design a heuristic algorithm for achieving an *individually stable* OCF structure. Experimental results demonstrate the feasibility of our algorithm, and show that our mechanism can improve the personalized model performance by up to 19% over existing methods.

## KEYWORDS

Personalized Federated Learning, Individual Rationality, Overlapping Coalition Formation Game

## 1 INTRODUCTION

Recently, data isolation and concerns of privacy preservation have aroused the wide study of federated learning (FL), where models are trained across multiple decentralized clients holding local data without exchanging data directly. In the vanilla FL setting, federating participants form a grand coalition and collaborate to train a globally shared model. In practice, however, clients' individual data distributions are generally non-independent identically distributed (non-IID), and thus it is difficult for a global model to perform well on clients' local distributions.

To this end, personalized FL is proposed to maximize individual utility (i.e., model performance) by customizing personalized models for each client. Among the mainstream personalization strategies, group collaboration techniques [11] encourage cluster-level collaboration among similar clients to customize personalized models for each client. However, in such group collaboration approaches, *the collaboration arrangement for personalized FL designed by the central server may end up being an unstable coalition structure, as the rational clients may deviate from the designed coalition structure to join their favorite coalition for higher utility (i.e., personal model performance).* Taking into account the significance of individual rationality [12], some recent study has dissected the clients' participation decisions problem in FL as a *hedonic* coalition formation game [8], where the utility-maximizing clients (referred to as players interchangeably) arrange themselves into federating coalitions.

However, a key limitation of the above game design is that they require players *join no more than one coalition*, which constrains players' attempts or possibilities to participate in multiple coalitions. Intuitively, allowing players to join multiple coalitions could potentially lead to an enhanced utility, as players can leverage the knowledge embedded in different coalition-wide global models to further optimize their personalized model.

We advocate a novel mechanism that allows players to engage in multiple coalitions, and formulate their participation decisions problem into an *overlapping coalition formation* (OCF) game. In the proposed mechanism, each player weights their local model and the coalition models to generate a personalized model. However, an open problem in the proposed mechanism is to identify a stable OCF structure where no player deviates from the current structure, which leads to our key question: *how does a stable OCF structure look like?*

In this paper, we characterize the model performance of each potential coalition metric in a machine learning task by delving into the linear regression model, which has proven to be insightful for nonlinear models, as in [9]. Through the linear model setting, we can quantify players' utility (i.e., model error) in different coalitions. These error values provide fundamental insights into understanding which coalition a player would prefer to take part in during the OCF game. We propose an effective and low-complexity heuristic algorithm to obtain a stable OCF structure. The algorithm follows a greedy strategy, where we rank players' participation preferences based on the error of coalition models and prioritize satisfying the highest-ranked preference of each player in each round of the game. This strategy accelerates the convergence process of coalition formation, ensuring computational practicality and efficiency. Moreover, among a set of players, there may exist multiple *individually stable* OCF structures where no player can deviate from the current structure for higher utility. Our algorithm guarantees to converge to one of these stable structures.

## 2 SYSTEM MODEL

In this section, we start by presenting our personalized FL model and outlining the technical assumptions. Then, we provide essential definitions and notions related to the OCF game. Finally, we formally formulate the problem that we aim to solve.

### 2.1 FL Model and Technical Assumptions

Consider a scenario with a fixed set of clients $\mathcal{N} = \{1, 2, \cdots, N\}$ seeking to participate in FL, and each client has their true model parameters $\theta_i$ which they aim to estimate. We focus on the linear regression task and $\theta_i$ is a $D$-dimensional linear vector representing the coefficients for the classification function. The private dataset consists of feature vectors $\mathbf{X}_i$ and the corresponding labels $\mathbf{Y}_i$. Each client utilizes the known data $(\mathbf{X}_i, \mathbf{Y}_i)$ to estimate their unknown model parameter $\theta_i$. We assume that each client corresponds to an input distribution $\mathcal{X}_i$ and $n_i$ input data points are drawn from the distribution $\mathbf{X}_i \sim \mathcal{X}_i$ such that $\mathbb{E}_{x \sim \mathbf{X}_i}[\mathbf{X}_i^T \mathbf{X}_i] = \Sigma_i$. They then observe the noisy outputs which are drawn with a variance $\epsilon_i^2$ around the true mean of the distribution, i.e., $\mathbf{Y}_i \sim \mathcal{D}_i(\mathbf{X}_i \theta_i, \epsilon_i^2)$. The expectation of the error parameters is $\mu_e = \mathbb{E}_{\epsilon_i^2 \sim \mathcal{P}}[\epsilon^2]$. A client can learn its model parameter through stochastic gradient descent (SGD) [2, 5] or ordinary least squares (OLS) [6]. For theoretical game reasoning to be feasible, we assume that $\mathbf{X}_i^T \mathbf{X}_i$ is invertible such that each client can use OLS to compute their local model parameters. The local estimation is given by

$$\hat{\theta}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{Y}_i \tag{1}$$

We consider that clients organize themselves into multiple collaborative groups, and each collaborative group is called a coalition, denoted by $C_j$. Let $N_j$ denote the total number of samples in a coalition, given by $N_j = \sum_{i \in C_j} n_i$. A coalition performs FL and generates a coalition-wide global model:

$$\hat{\Theta}_j = \sum_{i \in C_j} \frac{n_i}{N_j} \cdot \hat{\theta}_i \tag{2}$$

We use $\Pi_i$ to indicate the collection of coalitions that player $i$ participates in. The aggregation function of the personalized model is formalized as follows

$$\hat{\theta}_i^{\Pi} = w_i \cdot \hat{\theta}_i + \sum_{C_j \in \Pi_i} w_{ij} \cdot \hat{\Theta}_j \tag{3}$$

for $w \in [0, 1]$ and $w_i + \sum_{C_j \in \Pi_i} w_{ij} = 1$. We use $w_i, w_{ij}$ to respectively refer to the aggregation weight of local model and the corresponding coalition model. Note that $(w_i, w_{ij})$ are player-specific parameters that can be tuned.

## 2.2 An OCF Game Formulation

In our mechanism, each client enjoys the autonomy to make independent decisions. They prioritize their own interests when choosing which coalition to join. To capture clients' behaviors in making participation decisions, we formulate their strategic interactions into an OCF game.

Definition 1 (OCF game). *In an OCF game with player set $\mathcal{N}$, $C \subseteq \mathcal{N} \neq \emptyset$ is called a coalition on $\mathcal{N}$. An OCF structure of $\mathcal{N}$ is a collection $\Pi = \{C_1, C_2, \cdots, C_m\}$ such that $\bigcup_{i=1}^m C_i = \mathcal{N}$. In addition, the condition $C_i \cap C_j = \emptyset$ for any $C_i, C_j \in \Pi$ and $C_i \neq C_j$ is not a prerequisite.*

*1)Player Payoff:* In an OCF game, each player participating in a coalition would expect to receive a positive payoff. The payoff of player $i$ in coalition $C_j$ is the reduction in model error between their local model and the coalition model, which is formalized as

$$\varphi_i^+(C_j) = err_i(\hat{\theta}_i) - err_i(\hat{\Theta}_j) \tag{4}$$

*2)Collaboration Cost:* In some scenarios, the computation and communication costs incurred from FL are non-negligible, especially for players who joins multiple coalitions. Let $c_i$ denote the unit cost of player $i$'s data training when it participates in a coalition, therefore, the cost of player $i$ in coalition $C_j$ is defined as

$$\varphi_i^-(C_j) = c_i n_i \tag{5}$$

*3)Player Utility:* In our formulated game, the utility of a player relates to the performance of its personalized model and the computation/communication cost incurred from joining multiple coalitions. Formally, the utility of player $i$ in an OCF structure $\Pi$ is defined as the difference between the error reduction and total collaboration cost

$$\mathbf{U}_i(\Pi) = err_i(\hat{\theta}_i) - err_i(\hat{\theta}_i^{\Pi}) - \sum_{C_j \in \Pi_i} \varphi_i^-(C_j) \tag{6}$$

Note that the limit of the utility function is $err_i(\hat{\theta}_i) - \sum_{C_j \in \Pi_i} \varphi_i^-(C_j)$ and players always pursue a positive utility, hence, the maximum number of coalitions that a player would join can be estimated by the condition $\lim \mathbf{U}_i(\Pi) > 0$, i.e., $max(|\Pi_i|) = err_i(\hat{\theta}_i)/c_i n_i$. Players would stop attempting to join new coalitions when reaching the maximum number of coalitions that they could participate in.

*4) Individual Rationality:* The players make independent decisions on which coalitions to participate in, based on the potential payoff they may get from the coalition. Individual rationality states that an individual always chooses the action that maximizes its utility. We assume that each player always tends to join the coalition with the highest payoff in each round, i.e., the coalition $C_j$ chosen by player $i$ satisfies

$$\varphi_i^+(C_j) \geq \varphi_i^+(C_k), \forall C_k \in \Pi, j \neq k \tag{7}$$

*5) Joining Operation and Rule:* Considering the stability of the OCF structure, players cannot join their preferred coalition at will. The execution of joining operations requires adherence to specific rules. We denote the operation of player $i$ joining coalition $C_j$ as $J_i C_j$, and give related definitions as follows

Definition 2 (Joining Operation). *Given an OCF structure $\Pi$, if the operation of player i joining coalition $C_j$ is performed, then the OCF structure is modified to $\Pi^* = \{\Pi \setminus C_j \cup \{C_j \cup i\}\}$.*

Definition 3 (Joining Rule). *A joining operation can be executed only if player i can get a positive payoff from coalition $C_j$ and the utility of existing players in that coalition should not be decreased.*

$$\begin{cases} \varphi_i^+(C_j \cup i) > \varphi_i^+(\{i\}) \\ \varphi_k^+(C_j \cup i) \geq \varphi_k^+(C_j), \forall k \in C_j \end{cases} \tag{8}$$

*6) Stability of the OCF structure:* There are various stability notions in *hedonic* games [1, 13]. Stability concepts based on coalitional deviations (*core stable* and *strictly core stable*) are too strong to assure the existence of stable coalition structures in the general setting. Therefore, we adopt the stability notion based on individual deviations [3] to guarantee the existence of different solutions. In the following, we give the notion of individually stable.

Definition 4 (Individually Stable). *An OCF structure $\Pi$ is individually stable if there does not exist a pair $(i, C_j)$ of $i \in \mathcal{N}$, $C_j \in \Pi$ and $i \notin C_j$ such that*

$$\begin{cases} \varphi_i^+(C_j \cup i) > \varphi_i^+(\{i\}) \\ \varphi_k^+(C_j \cup i) \geq \varphi_k^+(C_j), \forall k \in C_j \end{cases} \tag{9}$$

## 2.3 Problem Formulation

In this work, we aim to solve the client utility maximization problem in personalized FL, while taking into account individual rationality and ensuring the stability of the OCF structure.

**Algorithm 1** The OCF algorithm

```
Input: Player set $\mathcal{N}$;
Output: OCF structure $\Pi$;
1: // OCF structure initialization, each play organizes themselves as a single-member coalition
2: Let OCF structure $\Pi \leftarrow \{\{i\}\}_{i=1}^{N}$
3: // overlapping coalition formation
4: while $\Pi^t \neq \Pi^{t-1}$ do
5:     Broadcast OCF structure information $\Pi^{t-1}$.
6:     // active player set
7:     $\mathcal{N}^* = \{i | i \in \mathcal{N}, |\Pi_i| < max(|\Pi_i|)\}$
8:     for client $i \in \mathcal{N}^*$ in parallel do
9:         $\Pi_{\setminus i} = \{C_j \mid i \notin C_j\}$
10:        $C_k \leftarrow \min_{C_k \in \Pi_{\setminus i}} \mathrm{err_i}(\hat{\Theta}_k)$
11:        if Rule (8) is satisfied then
12:            $J_i C_k$
13:    Update OCF structure information $\Pi^t$.
14: Return $\Pi$
```

The essence of personalized FL is to customize models for each client who aims to minimize their average local loss at private data distribution. For each personalized model $\hat{\theta}_i^{\Pi}$, personal distribution $\mathcal{D}_i$ and local loss function $f$, the objective of our utility maximization problem in personalized FL can be formulated as follows

$$\min_{\{\hat{\theta}_i^{\Pi}\}_{i=1}^{N}} \sum_{i=1}^{N} f(\hat{\theta}_i^{\Pi}, \mathcal{D}_i)$$
$$\text{s.t. } \Pi \text{ is } \textit{individually stable}$$

The aggregation function of each personalized model in Eq. (3) indicates that coalition-wide global models have a significant impact on personalized model performance. This means that the problem of minimizing local error can be essentially translated into the identification of coalition structure. Under a stable OCF structure, each client is motivated to engage in FL and benefits from the coalition-wide global model.

# 3 PERSONALIZED FL WITH OVERLAPPING COALITION

In this section, we start with a theoretical analysis of players' participation decisions. Then, based on the fundamental insights obtained from the theoretical analysis, we design a heuristic algorithm for identifying a stable OCF structure. At last, we present the workflow of personalized FL under a stable OCF structure.

## 3.1 The OCF algorithm

The pseudo-code presented in Algorithm 1 outlines the process of our OCF game.

At the initialization phase, each player organizes themselves as a single-member coalition, as shown in Line 2. Only players who haven't joined the maximum number of coalitions would attempt to join new coalitions, and we call these players *active* players. At the beginning of each iteration, the latest coalition structure information will be broadcast to players. Each active player then sorts their unjoined coalitions based on the estimated expected error obtained from the coalition models, as illustrated in Line 9 - Line 10. Subsequently, they assess whether joining the coalition with the least error would provide greater benefits than local training. A joining operation can be executed only if a player can get a positive payoff and the utility of existing members in that coalition should not be decreased, and the corresponding code is Line 11 -Line 12. The coalition structure will be updated after all players have made their decisions.

**Algorithm 2** personalized FL with stable OCF structure

```
Input: OCF structure $\Pi$;
Output: Personalized models $\{\hat{\theta}_i^{\Pi}\}_{i=1}^{N}$;
1: // coalition-wide global training
2: for coalition $C_j \in \Pi$ do
3:     for client $i \in C_j$ do
4:         Local estimation based on Eq. (1)
5:     Coalition model $\hat{\Theta}_j$ aggregation based on Eq. (2)
6: //Personalized model generation
7: for client $i \in \mathcal{N}$ do
8:     Personalized model $\hat{\theta}_i^{\Pi}$ generation based on Eq. (3)
9: Return $\{\hat{\theta}_i^{\Pi}\}_{i=1}^{N}$
```

The above procedure may take several iterations until no player can successfully join any coalition.

Theorem 1. *Starting from the initial OCF structure and after several operations, our proposed OCF algorithm will converge to an individually stable OCF structure in finite iterations.*

Proof. See details in appendix. □

## 3.2 Personalized FL with stable OCF structure

The Algorithm 1 outputs a stable OCF structure where players collaborate to train the coalition-wide global model. Personalized FL begins after the stable OCF structure comes into being. Here, we outline the workflow of personalized FL in Algorithm 2.

During the federated training phase (Line 2 - Line 6), players would perform FL in the coalitions that they have successfully joined. Correspondingly, each coalition aggregates the uploaded parameters to generate a coalition-wide global model. The coalition model is accessible to the players in that coalition. After the coalition models converge, each player adjusts the aggregation weight between their local model and the accessible coalition models to produce their personalized model, as shown in Line 8 - Line 9.

# 4 EXPERIMENT

In this section, we provide a comprehensive evaluation of our mechanism, which comprises three parts. First, we validate the feasibility and convergence of the OCF algorithm. Second, we evaluate the performance of our mechanism under the OCF structure output by the OCF algorithm and investigate the impact of aggregation weight on the personalized model.

## 4.1 Experimental Setup

**Platforms.** We conduct experiments both in a simulated environment and on a networked hardware prototype system. In the

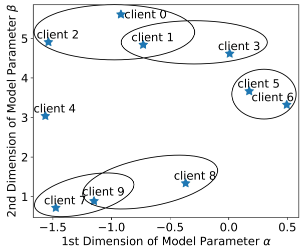

**Figure 1: True model parameters distribution ($\theta_i = [\alpha_i, \beta_i]$) and the OCF algorithm output.**

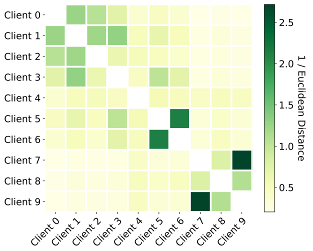

**Figure 2: Local model similarity (The darker color of the block indicates a higher similarity).**

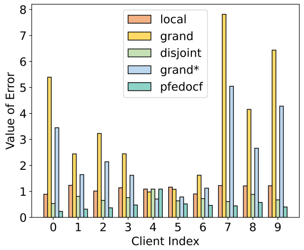

**Figure 3: Model error of the 10 clients under different collaboration mechanisms.**

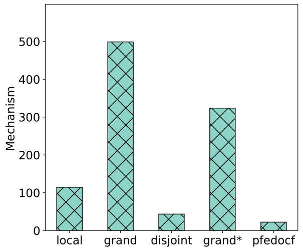

**Figure 4: Overall error of the 100 clients under different collaboration mechanisms.**

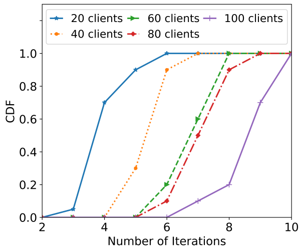

**Figure 5: The CDF of the total number of iterations.**

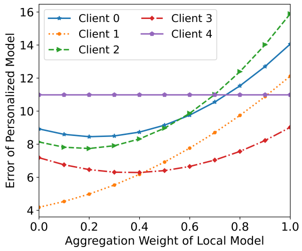

**Figure 6: The personalized model error aggregation weights.**

**Table 1: The personalized model performance improvement of the 100 clients.**

| | performance improvement compared to *local* | | | | |
|---|---|---|---|---|---|
| | <0% | [0,50)% | [50,70)% | [70,90)% | [90,100)% |
| *grand* | 72 | 14 | 4 | 6 | 4 |
| *grand** | 1 | 17 | 42 | 40 | 0 |
| *disjoint* | 69 | 9 | 10 | 8 | 4 |
| *pfedocf* | 0 | 1 | 11 | 75 | 13 |

simulated system, we simulate 100 virtual devices and a virtual central server to experiment with linear regression tasks. On the prototype system, we conducted a 10-client scale experiment. Raspberry Pis serves as clients, and a laptop computer acts as the central server.

**Model & Dataset.**We adopt a 2-dimensional linear regression model for demonstration. Furthermore, the distribution of input values $\mathcal{X}$ follows the multivariate normal distribution with 0 mean. For a given data point $x = [x_0, x_1]$, the true output is $y = \alpha x_0 + \beta x_1$. However, each client observes noisy outputs $\hat{y} = y + \epsilon$ and endeavors to estimate the authentic parameter $\theta = [\alpha, \beta]$. For each coalition, the experiments are conducted on the joint dataset.

**Compared Methods & Evaluation Metric.** To evaluate the performance of our proposed method *pfedocf*, we compare it with several benchmarks. The details of each method are introduced as follows:

- *local*: Each client estimates their model parameter locally without collaboration.
- *grand(FedAvg):* Vanilla FL setting. All clients form a grand coalition and cooperatively train a uniform model. Without fine-tuning, clients adopt this global model as their personalized model.
- *disjoint:* In the context of hedonic games, each client is constrained to join no more than one coalition and prioritize choosing the one that offers the highest potential payoff. Within each coalition, clients collectively train a coalition model. This coalition-wide global model servers as their personalized model.
- *gand*(FedAvg with fine-tuning):* This approach is a variation of the previous *grand* method, the only difference is that each client can weight the global model with their local model to produce a personalized model.
- *pfedocf:* In the context of hedonic games, clients have the flexibility to join multiple coalitions. Each client weights the coalition models with their local model to produce a personalized model.

## 4.2 Performance Results

We present the experiment results and analyze the observation to demonstrate our mechanism.

*1) Feasibility of the OCF algorithm:* Fig. 1 pictures the true model parameters distribution of 10 clients and the derived OCF structure from our OCF algorithm. Each circle in the diagram represents a coalition, and the points inside the circle represent the members of that coalition. Overlapping circles indicate that the clients join multiple coalitions. Fig. 2 illustrates the local model similarity between the 10 clients. We observe that clients with high pairwise similarity in Fig. 2 have formed partnerships in the stable OCF structure output by Fig. 1. This observation demonstrates the feasibility of our algorithm in identifying the formation of stable collaboration, as in general, clients with similar distribution are more likely to see a reduction in model error and thus more willing to collaborate with each other.

*2) Performance under the derived OCF structure:* For a comprehensive evaluation, we carried out experiments involving two different client scale scenarios. Fig. 3 illustrates the model error of 10 clients under various mechanisms, while Fig. 4 demonstrates the overall model error of 100 clients under different collaboration approaches. It is evident that our approach outperforms other methods in both scenarios. Worth noting that client 4 in the 10-client scenario opts for local estimation as it was unable to identify any viable coalition to join during the game.

*3) Superiority of the proposed mechanism:* We summarize the percentage distribution of 100 clients across different performance improvement intervals in Table 1. By comparing the results of *grand* and *grand**, we can conclude that weighting the global model with the local model enables the global model to better accommodate clients' local distributions, thereby reducing errors. The experimental results of *disjoint* and *pfedocf* demonstrate that enabling clients to engage in multiple coalitions helps them further improve personal model performance.

*4) Convergence analysis:* Fig. 5 shows the convergence result of the OCF algorithm. We can observe the cumulative distribution function (CDF) of the total number of iterations, versus the number of iterations, with a different number of clients. Higher CDF in fewer rounds, indicating a faster convergence rate. It is obvious that our OCF algorithm can converge within 10 iterations in all the scenarios, revealing that the computational complexity of our proposed algorithm is rather smaller.

*5) Impact of model aggregation weight:* Fig. 6 illustrates the fluctuation of personalized model errors as the aggregation weight of the local model ranges from 0 to 1. It's noteworthy that the personalized model error of client 4, who does not participate in any coalition, remains unaffected by the model aggregation weights and equals its local model error. We can also see that the personalized model performance of client 4 who participates in multiple coalitions is getting worse when the local model weights more. Moreover, it can be observed from the results that the optimal weight of model aggregation differs for each client.

## 5 RELATED WORK

To address the statistical diversity challenge of clients with non-IID data, personalized FL has recently emerged as a promising solution that is attracting more attention.

FedAMP[10] and FedFomo[14] encourage pairwise collaboration between clients with relevant local target distribution. Although pairwise collaboration methods have achieved good results, they only rely on one-to-one model similarity, and the communication efficiency is also adversely affected owing to iterative pair comparison. Superior to pairwise collaboration, group collaboration schemes [4, 11] encourage cluster-level collaboration and have achieved a promising performance. Yet, the traditional group collaboration approaches assume that all clients voluntarily take part in FL, which neglects the significance of individual rationality.

There are existing works that take individual rationality into consideration where game theory is employed as a powerful tool to study this issue. For instance, one recent work [7] considers the clustering of clients in the form of hedonic games and investigates how clients make decisions to participate in the FL setting. Much of this paper analyzes the stability of coalition structures instead of searching for a stable coalition structure. However, they assume that players join no more than one coalition, which constrains clients' participation and attempts to take part in multiple coalitions.

Paper [15] explores the conditions under which stable coalition structures can be formed and provides insights into their computational aspects, which has inspired us to introduce the concept of overlapping games into the research field of personalized FL. Their research applies to more general classes of games, while we focus on a specific problem domain, resulting in a distinct game model. In contrast to all existing works, we propose a mechanism where utility-maximizing clients can strategically join multiple coalitions, and then we formulate clients' participation decisions problem into an OCF game. In our heuristic algorithm, the game will finally converge to an individually stable OCF structure.

## 6 CONCLUSION

In this work, we have studied personalized FL from a cooperative game theoretical perspective. We formulate the strategic interaction among the clients into an OCF game where each client can participate in multiple coalitions and generate their personalized model by weighting the local model and coalition models. Then, we design a heuristic algorithm to derive an individually stable OCF structure. Extensive experimentation results demonstrate the effectiveness of our proposed algorithm and validate the superiority of our mechanism. Although we focus on a linear regression task for game theoretical reasoning to be feasible, the OCF mechanism can also be applied to other popular non-linear models.

## ACKNOWLEDGMENTS

The research results of this work are sponsored by the Kunshan Municipal Government research funding.

## REFERENCES

[1] Krzysztof R Apt and Tadeusz Radzik. 2006. Stable partitions in coalitional games. *arXiv preprint cs/0605132* (2006).

[2] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*. Springer, 177–186.

[3] Felix Brandt, Martin Bullinger, and Anaëlle Wilczynski. 2021. Reaching individually stable coalition structures in hedonic games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5211–5218.

[4] Christopher Briggs, Zhong Fan, and Peter Andras. 2020. Federated learning with hierarchical clustering of local updates to improve training on non-IID data. In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–9.

[5] Jeffrey Dean, Greg Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Mark Mao, Marc'aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, et al. 2012. Large scale distributed deep networks. *Advances in neural information processing systems* 25 (2012).

[6] Clara Dismuke and Richard Lindrooth. 2006. Ordinary least squares. *Methods and designs for outcomes research* 93, 1 (2006), 93–104.

[7] Kate Donahue and Jon Kleinberg. 2021. Model-sharing games: Analyzing federated learning under voluntary participation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5303–5311.

[8] Kate Donahue and Jon Kleinberg. 2023. Fairness in model-sharing games. In *Proceedings of the ACM Web Conference 2023*. 3775–3783.

[9] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani. 2022. Surprises in high-dimensional ridgeless least squares interpolation. *Annals of statistics* 50, 2 (2022), 949.

[10] Yutao Huang, Lingyang Chu, Zirui Zhou, Lanjun Wang, Jiangchuan Liu, Jian Pei, and Yong Zhang. 2021. Personalized Cross-Silo Federated Learning on Non-IID Data. In *AAAI*. 7865–7873.

[11] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. 2020. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems* 32, 8 (2020), 3710–3722.

[12] Arkady Sobolev. 1995. The nucleolus for cooperative games with arbitrary bounds of individual rationality. *International Journal of Game Theory* 24 (1995), 13–22.

[13] Shao Chin Sung and Dinko Dimitrov. 2007. On myopic stability concepts for hedonic games. *Theory and Decision* 62, 1 (2007), 31–45.

[14] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M Alvarez. 2020. Personalized federated learning with first order model optimization. *arXiv preprint arXiv:2012.08565* (2020).

[15] Yair Zick, Georgios Chalkiadakis, Edith Elkind, and Evangelos Markakis. 2019. Cooperative games with overlapping coalitions: Charting the tractability frontier. *Artificial Intelligence* 271 (2019), 74–97.

# A ANALYSIS OF PARTICIPATION DECISION

Given the nature of utility-maximizing problems in OCF games, which involve players' participation decisions about which coalitions to join, it is crucial to analyze players' utility. To facilitate this analysis, we provide the exact *mean square error* (MSE) that a player would experience when joining different coalitions.

Lemma 1 (Lemma 4.1, from Donahue and Kleinberg [7]). *For linear regression, the expected MSE that player i with $n_i$ samples derives from local estimation is:*

$$\mu_e \cdot tr[\Sigma_i \mathbb{E}_{\mathbf{X}_i \sim \mathcal{X}_i}[(\mathbf{X}_i^T \mathbf{X}_i)^{-1}]] \tag{10}$$

*if the distribution of input values $\mathcal{X}_i$ is a D-dimensional multivariate normal distribution with 0 mean, the expected MSE of local estimation can be simplified to:*

$$\mu_e \cdot \frac{D}{n_i - D - 1} \tag{11}$$

Lemma 2. *For linear regression, the expected MSE that player i with $n_i$ samples derives from the personalized model is:*

$$\begin{aligned} &\mu_e \sum_{k \in P_i, k \neq i} ( \sum_{C_j \in \Pi_i, k \in C_j} w_{ij} \frac{n_k}{N_j})^2 \cdot tr[\Sigma_i \mathbb{E}_{\mathbf{X}_k \sim \mathcal{X}_k}[(\mathbf{X}_k^T \mathbf{X}_k)^{-1}]] \\ &+ \mu_e (w_i + \sum_{C_j \in \Pi_i} w_{ij} \frac{n_i}{N_j})^2 \cdot tr[\Sigma_i \mathbb{E}_{\mathbf{X}_i \sim \mathcal{X}_i}[(\mathbf{X}_i^T \mathbf{X}_i)^{-1}]] \\ &+ \Sigma_i \left( \sum_{k \in P_i, k \neq i} \sum_{C_j \in \Pi_i, k \in C_j} w_{ij} \frac{n_k}{N_j} (\theta_k - \theta_i) \right)^2 \end{aligned} \tag{12}$$

*if the distribution of input values $\mathcal{X}$ is a D-dimensional multivariate normal distribution with 0 mean, it can be simplified to:*

$$\begin{aligned} &\mu_e \sum_{k \in P_i, k \neq i} ( \sum_{C_j \in \Pi_i, k \in C_j} w_{ij} \frac{n_k}{N_j})^2 \cdot \frac{D}{n_k - D - 1} \\ &+ \mu_e (w_i + \sum_{C_j \in \Pi_i} w_{ij} \frac{n_i}{N_j})^2 \cdot \frac{D}{n_i - D - 1} \\ &+ \Sigma_i \left( \sum_{k \in P_i, k \neq i} \sum_{C_j \in \Pi_i, k \in C_j} w_{ij} \frac{n_k}{N_j} (\theta_k - \theta_i) \right)^2 \end{aligned} \tag{13}$$

Proof. Recall that each player's local model $\hat{\theta}_i$ can be estimated as follows:

$$\hat{\theta}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{Y}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} (\mathbf{X}_i \theta_i + \eta_i)$$

The aggregation function of each coalition model is given by

$$\Theta_j = \sum_{i \in C_j} \frac{n_i}{N_j} \cdot \theta_i \tag{14}$$

where $N_j = \sum_{i \in C_j} n_i$ denotes the total number of samples in coalition $C_j$.

The personalized model generation is a combination of local model and coalition models, which is formalized as

$$\theta_i^{\Pi} = w_i \cdot \theta_i + \sum_{C_j \in \Pi_i} w_{ij} \cdot \Theta_j \tag{15}$$

for $w \in [0, 1]$ and $w_i + \sum_{C_j \in \Pi_i} w_{ij} = 1$.

With Eq. (14), we can equivalently transform Eq. (15) as follows:

$$\theta_i^{\Pi} = \sum_{j \in P_i} v_{ij} \theta_j \tag{16}$$

for $v_{ij}$ denoting the weight of player $j$'s local model in player $i$'s personalized model and we have $v_{ij} = \sum_{C_k \in \Pi_i, j \in C_k} \frac{n_j}{N_k} w_{ik}$ and $v_{ii} = w_i + \sum_{C_k \in \Pi_i} \frac{n_i}{N_k} w_{ik}$.

First, note that the expected error of a set of parameters at a particular point $x$ is determined by the $\mathbb{E}_{x \sim \mathbf{X}_i}[(x^T \hat{\theta}_i - x^T \theta_i)^2]$. Let's expand the inner expression

$$\begin{aligned} (x^T \theta_i - x^T \hat{\theta}_i^{\Pi})^2 &= (x^T \theta_i + x^T \theta_i^{\Pi} - x^T \theta_i^{\Pi} - x^T \hat{\theta}_i^{\Pi})^2 \\ &= (x^T \theta_i - x^T \theta_i^{\Pi})^2 + (x^T \theta_i^{\Pi} - x^T \hat{\theta}_i^{\Pi})^2 \\ &+ 2(x^T \theta_i - x^T \theta_i^{\Pi})(x^T \theta_i^{\Pi} - x^T \hat{\theta}_i^{\Pi}) \end{aligned} \tag{17}$$

As we have $\mathbb{E}_{\mathbf{Y} \sim \mathcal{D}(\theta_i, \epsilon_i^2)}[x^T \theta_i^{\Pi} - x^T \hat{\theta}_i^{\Pi}] = 0$, the last term in Eq. (17) equals to 0. Next, we expand the second term in Eq. (17) as follows

$$\begin{aligned} (x^T \theta_i^{\Pi} - x^T \hat{\theta}_i^{\Pi})^2 &= \left( x^T \sum_{j \in P_i} v_{ij} \theta_j - x^T \sum_{j \in P_i} v_{ij} \hat{\theta}_j \right)^2 \\ &= \left( \sum_{j \in P_i} v_{ij} x^T (\theta_j - \hat{\theta}_j) \right)^2 \\ &= \sum_{j \in P_i} \left( v_{ij} x^T (\theta_j - \hat{\theta}_j) \right)^2 \\ &+ \sum_{j \in P_i} \sum_{j \neq k} \left( v_{ij} x^T (\theta_j - \hat{\theta}_j) v_{ik} x^T (\theta_k - \hat{\theta}_k) \right) \end{aligned} \tag{18}$$

Note that we assume the true model parameter of each client is irrelevant, i.e., the expectation of $\theta_j - \hat{\theta}_j$ is 0. Therefore, the second term in Eq. (18) ends up being 0. Remaining the first term and rewrite it as follows:

$$\sum_{j \in P_i} v_{ij}^2 (x^T \theta_j - x^T \hat{\theta}_j)^2$$

The term $(x^T \theta_j - x^T \hat{\theta}_j)^2$ is the value of local estimation in Lemma 1, which has been calculated in prior work. Hence, we can rewrite it as

$$\mu_e \sum_{j \in P_i} v_{ij}^2 \cdot tr[\Sigma_i \mathbb{E}_{\mathbf{X}_j \sim \mathcal{X}_j}[(\mathbf{X}_j^T \mathbf{X}_j)^{-1}]]$$

if $\mathbf{X}_j$ follows the multivariate normal distribution with 0 mean, it can be simplified to

$$\mu_e \sum_{j \in P_i} v_{ij}^2 \cdot \frac{D}{n_j - D - 1}$$

As we have

$$\begin{aligned} \sum_{j \in P_i} v_{ij}^2 &= \sum_{j \in P_i j \neq i} v_{ij}^2 + v_{ii}^2 \\ &= \sum_{j \in P_i, j \neq i} ( \sum_{C_k \in \Pi_i, j \in C_k} w_{ik} \frac{n_j}{N_k})^2 + (w_i + \sum_{C_k \in \Pi_i} w_{ik} \frac{n_i}{N_k})^2 \end{aligned}$$

Eq. (18) isequivalent to

$$
\begin{aligned}
&(x^T\theta_i^{\Pi} - x^T\hat{\theta}_i^{\Pi})^2\\
=&\mu_e \sum_{j\in P_i, j\neq i} \Big( \sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}\Big)^2 \cdot tr[\Sigma_i \mathbb{E}_{\mathbf{X}\sim\mathcal{X}_j}[(\mathbf{X}_j^T\mathbf{X}_j)^{-1}]]\\
+&\mu_e\Big(w_i + \sum_{C_k\in\Pi_i} w_{ik}\frac{n_i}{N_k}\Big)^2 \cdot tr[\Sigma_i \mathbb{E}_{\mathbf{X}\sim\mathcal{X}_i}[(\mathbf{X}_i^T\mathbf{X}_i)^{-1}]]
\end{aligned}
$$

At last, we consider the first term in Eq. (17)

$$
\begin{aligned}
(x^T\theta_i - x^T\theta_i^{\Pi})^2 &= (x^T(\theta_i - \theta_i^{\Pi}))^T x^T(\theta_i - \theta_i^{\Pi})\\
&= (\theta_i - \theta_i^{\Pi})^T x x^T (\theta_i - \theta_i^{\Pi})
\end{aligned}
$$

We note that the above quantity is a scalar. For a scale, $tr(a) = a$, and for any matrix $tr(AB) = tr(BA)$. Taking the expectation through the cyclic property of trace, we can translate the above equation:

$$
\begin{aligned}
&(x^T\theta_i - x^T\theta_i^{\Pi})^2\\
=&tr[(\theta_i - \theta_i^{\Pi})^T \mathbb{E}_{x\sim\mathcal{X}}[xx^T](\theta_i - \theta_i^{\Pi})]\\
=&tr[\Sigma_i(\theta_i - \theta_i^{\Pi})^T(\theta_i - \theta_i^{\Pi})]
\end{aligned}
\tag{19}
$$

Using Eq. (16), we can simplify the inner term of Eq. (19) involving the $\theta$ values:

$$
\begin{aligned}
&(\theta_i - \theta_i^{\Pi})^T(\theta_i - \theta_i^{\Pi})\\
=&(\theta_i - \sum_{j\in P_i} v_{ij}\theta_j)^T(\theta_i - \sum_{j\in P_i} v_{ij}\theta_j)\\
=&\left((1-v_{ii})\theta_i - \sum_{j\in P_i, j\neq i} v_{ij}\theta_j\right)^T\left((1-v_{ii})\theta_i - \sum_{j\in P_i, j\neq i} v_{ij}\theta_j\right)
\end{aligned}
$$

Note that $v_{ii} + \sum_{j\in P_i, j\neq i} v_{ij} = 1$ and $v_{ij} = \sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}$, so we can rewrite the above equation as

$$
\begin{aligned}
&(\theta_i - \theta_i^{\Pi})^T(\theta_i - \theta_i^{\Pi})\\
=&\Big(\sum_{j\in P_i, j\neq i} v_{ij}\theta_i - \sum_{j\in P_i, j\neq i} v_{ij}\theta_j\Big)^T\Big(\sum_{j\in P_i, j\neq i} v_{ij}\theta_i - \sum_{j\in P_i, j\neq i} v_{ij}\theta_j\Big)\\
=&\left(\sum_{j\in P_i, j\neq i} v_{ij}(\theta_i - \theta_j)\right)^T\left(\sum_{j\in P_i, j\neq i} v_{ij}(\theta_i - \theta_j)\right)\\
=&\left(\sum_{j\in P_i, j\neq i}\sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}(\theta_j - \theta_i)\right)^T\\
&\cdot\left(\sum_{j\in P_i, j\neq i}\sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}(\theta_j - \theta_i)\right)\\
=&\left(\sum_{j\in P_i, j\neq i}\sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}(\theta_j - \theta_i)\right)^2
\end{aligned}
$$

Finally, we can recombine our simplification into Eq. (17) to rewrite it:

$$
\begin{aligned}
&\mu_e \sum_{j\in P_i, j\neq i}\Big(\sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}\Big)^2 \cdot tr[\Sigma_i\mathbb{E}_{\mathbf{X}_j\sim\mathcal{X}_j}[(\mathbf{X}_j^T\mathbf{X}_j)^{-1}]]\\
&+\mu_e\Big(w_i + \sum_{C_k\in\Pi_i} w_{ik}\frac{n_i}{N_k}\Big)^2 \cdot tr[\Sigma_i\mathbb{E}_{\mathbf{X}_i\sim\mathcal{X}_i}[(\mathbf{X}_i^T\mathbf{X}_i)^{-1}]]\\
&+\Sigma_i\left(\sum_{j\in P_i, j\neq i}\sum_{C_k\in\Pi_i, j\in C_k} w_{ik}\frac{n_j}{N_k}(\theta_j - \theta_i)\right)^2
\end{aligned}
\tag{20}
$$

At this point, the proof of Lemma 2 is complete. □

Lemma 3. *For linear regression, the expected MSE that player i with $n_i$ samples derives from the uniform model of coalition $C_j$ is*

$$
\begin{aligned}
&\mu_e\sum_{k\in C_j}\frac{n_k^2}{N_j^2}\cdot tr[\Sigma_i\mathbb{E}_{\mathbf{X}_k\sim\mathcal{X}_k}[(\mathbf{X}_k^T\mathbf{X}_k)^{-1}]]\\
&+\Sigma_i\left(\sum_{k\in C_j, k\neq i}\frac{n_k}{N_j}\cdot(\theta_k - \theta_i)\right)^2
\end{aligned}
\tag{21}
$$

*if the distribution of input values $\mathcal{X}$ is a D-dimensional multivariate normal distribution with 0 mean, the expected MSE can be simplified to:*

$$
\mu_e\sum_{k\in C_j}\frac{n_k^2}{N_j^2}\frac{D}{n_k - D - 1} + \left(\sum_{k\in C_j, k\neq i}\frac{n_k}{N_j}(\theta_k - \theta_i)\right)^2
\tag{22}
$$

Proof. Lemma 3 can be derived from Lemma 2 by setting $w_i$ to 0. Remarkably, when considering only one coalition, $P_i$ is equivalent to $C_j$ and $\Pi_i$ equals to $\{C_j\}$. □

Theorem 2. *In our model, a player i with $n_i$ samples would receive positive payoff from a new coalition $C_j$ iff*

$$
n_i < D + 1 + \frac{\mu_e D(1 - (1 - \sum_{k\in C_j}\frac{n_k}{N_j})^2)}{\mu_e\sum_{k\in C_j}\frac{n_k^2}{N_j^2}\frac{D}{n_k-D-1} + \left(\sum_{k\in C_j}\frac{n_k}{N_j}(\theta_k - \theta_i)\right)^2}
\tag{23}
$$

Proof. Expr. (11) in Lemma 1 and Expr. (22) in Lemma 3 respectively give the exact errors that a player would expect from local training and collaborative training in a coalition. Let Expr. (11) > Expr. (22), we can determine the possible range of sample sizes when a player would consider joining a new coalition.

Firstly, we emphasize some key points in derivation. To estimate a player's benefit in a coalition, we would initially pretend player $i$ joins the coalition $C_j$ to help the coalition model generation, i.e., $C_j^* = \{i\}\cup C_j$. Thus, we have

$$
\begin{aligned}
&\mu_e\sum_{k\in C_j^*}\frac{n_k^2}{N_j^2}\frac{D}{n_k - D - 1}\\
=&\mu_e\sum_{k\in C_j^*, k\neq i}\frac{n_k^2}{N_j^2}\frac{D}{n_k - D - 1} + \mu_e\frac{n_i^2}{N_j^2}\frac{D}{n_i - D - 1}
\end{aligned}
$$

Referring to the model aggregation equation within a coalition defined in Eq. (2) of Section 2.1, we have $\frac{n_i}{N_j} = 1 - \sum\limits_{k\in C_j^*, k\neq i}\frac{n_k}{N_j}$. Review that Expr. (11) is $\mu_e\frac{D}{n_i-D-1}$, Expr. (22) is $\mu_e\sum_{k\in C_j^*}\frac{n_k^2}{N_j^2}\frac{D}{n_k-D-1} + \left(\sum_{k\in C_j^*, k\neq i}\frac{n_k}{N_j}(\theta_k - \theta_i)\right)^2$. Let Expr. (11) > Expr. (22), i.e.,

$$
\begin{aligned}
\mu_e\frac{D}{n_i - D - 1} > \mu_e\frac{n_i^2}{N_j^2}\frac{D}{n_i - D - 1} +\\
\mu_e\sum_{k\in C_j^*, k\neq i}\frac{n_k^2}{N_j^2}\frac{D}{n_k - D - 1} + \left(\sum_{k\in C_j^*, k\neq i}\frac{n_k}{N_j}(\theta_k - \theta_i)\right)^2
\end{aligned}
$$

As we have $\sum \frac{n_i}{N_j} = 1 - \sum_{k \in C_j^*, k \neq i} \frac{n_k}{N_j}$, the above inequation becomes

$$\mu_e \frac{D}{n_i - D - 1} > \mu_e (1 - \sum_{k \in C_j^*, k \neq i} \frac{n_k}{N_j})^2 \frac{D}{n_i - D - 1} +$$

$$\mu_e \sum_{k \in C_j^*, k \neq i} \frac{n_k^2}{N_j^2} \frac{D}{n_k - D - 1} + \left( \sum_{k \in C_j^*, k \neq i} \frac{n_k}{N_j} (\theta_k - \theta_i) \right)^2$$

Solving the above inequation, we derive

$$n_i < D + 1 + \frac{\mu_e D (1 - (1 - \sum_{k \in C_j^*, k \neq i} \frac{n_k}{N_j})^2)}{\mu_e \sum_{k \in C_j^*, k \neq i} \frac{n_k^2}{N_j^2} \frac{D}{n_k - D - 1} + \left( \sum_{k \in C_j^*, k \neq i} \frac{n_k}{N_j} (\theta_k - \theta_i) \right)^2}$$

For $C_j^* = \{i\} \cup C_j$,

$$n_i < D + 1 + \frac{\mu_e D (1 - (1 - \sum_{k \in C_j} \frac{n_k}{N_j})^2)}{\mu_e \sum_{k \in C_j} \frac{n_k^2}{N_j^2} \frac{D}{n_k - D - 1} + \left( \sum_{k \in C_j} \frac{n_k}{N_j} (\theta_k - \theta_i) \right)^2}$$

proof is completed. □

**Remark.** Theorem 2 plays a significant role in the OCF algorithm design. Even though the true model parameter $\theta_i$ remains unknown, it's logical for players to utilize their local model $\hat{\theta}_i$ for gauging the potential gains from joining a new coalition.

Theorem 3. *In our game, the error of a player's personalized model would not increase after joining a new coalition.*

Proof. Let $err_i(\Pi_{\setminus C_j})$, $err_i(\Pi)$ respectively denote the personalized model error of player $i$ before and after joining coalition $C_j$. As we can observe from Lemma 2, $err_i(\Pi_{\setminus C_j})$ is a special case of $err_i(\Pi)$ when the coefficient $w_{ij}$ is set to 0. As long as there exists a coalition $C_k$ that player $i$ joined previously and $err_i(\hat{\Theta}_k)$ is greater than $err_i(\hat{\Theta}_j)$, we can find out a set of aggregation weights to make $err_i(\Pi)$ less than $err_i(\Pi_{\setminus C_j})$. Even in the worst-case scenario where such a coalition does not exist, we can guarantee that $err_i(\Pi)$ equals to $err_i(\Pi_{\setminus C_j})$. □

## B PROOF OF THEOREM 1

Proof. In iteration $t$, the OCF structure changes from $\Pi^{t-1}$ to $\Pi^t$. During this process, player $i$ checks the current structure structure $\Pi^{t-1}$ and tend to join its preferred coalition, which satisfies $\mathbf{U}_i(\Pi^t) \geq \mathbf{U}_i(\Pi^{t-1})$. Rule (8) ensures that player $i$'s strategy doesn't negatively impact the utilities of other players, i.e., $\mathbf{U}_j(\Pi^t) \geq \mathbf{U}_j(\Pi^{t-1}), \forall j \neq i$. Consequently, the utility of any player will not decrease from $\Pi_{t-1}$ to $\Pi_t$. The pseudo-code reveals that the number of players per coalition gradually increases over iterations. Given that the collaboration cost limits the maximum number of coalitions that each player could join, the possible OCF structures are finite. More importantly, the number of players per coalition will not continuously grow due to data heterogeneity. This implies that, as the game processes on, few joining operations will be performed. The algorithm terminates when the OCF structure no longer updates, i.e., no joining operations are successfully performed during the last iteration. According to the Definition 4, an OCF structure is called individually stable when players cannot deviate from their current structure by joining any new coalition to achieve higher utility. Our algorithm satisfies this definition and ultimately converges to an individually stable OCF structure. □