Towards Cooperation in Adversarial Search with Confidentiality

Branislav Bošanský, Viliam Lisý, Michal Pěchouček

Agent Technology Center, Dept. of Cybernetics, FEE, Czech Technical University
Technická 2, 16627 Prague 6, Czech Republic
{bosansky, lisy, pechoucek}@agents.felk.cvut.cz

Abstract. We investigate the problem of cooperation of self-interested agents with respect to the confidentiality of their plans and without a presence of any third-party mediator. We base our approach on non-zero-sum n-player games in the extensive form focusing on two main goals: (1) the analysis of the maximal improvement of the utility value that an agent can gain by negotiation, and (2) the dependence of this improvement on the basic characteristics of the game – the number of agents, size of the branching factor, and correlation of the utility functions. We provide theoretical definitions together with experimental results evaluated on games abstracting real-world scenarios.

Introduction

Complex multi-agent environments modeling real-world scenarios (e.g. models of societies, economies, or war-games) have been objects of research interest for some time. Such environments often describe competitive situations where involved agents are self-interested and pursue their private goals. On the other hand, in the complex systems the agents frequently interact with each other in terms of either cooperation or conflicts. An example can be a military operation with several allied forces or a market shared by multiple competing companies.

The interactions among the agents make finding an optimal course of actions for an agent in such environments difficult. Single-agent planning approach is directly unusable for not taking actions of other agents into consideration, which can lead to frequent failures of the plan. On the other hand, approaches based on adversarial-search are explicitly taking into consideration actions of all other agents which exponentially increases the search space for an agent. The computational complexity of the adversarial search can be addressed by utilizing further knowledge about the domain (using heuristic, or utilizing the sparsity of the interactions between the agents). Another key disadvantage of adversarial-search methods is unnecessary cautiousness in cases modeled scenarios are not strictly adversarial – agents selfishly maximize their utility values, however, it can be beneficial for them to partially cooperate. In this paper we therefore focus on incorporating the concept of cooperation into an adversarial search.

One way to address the concept of cooperation of agents in multi-agent environments is the multi-agent planning (e.g. described in [1]). The multi-agent planning assumes that agents are either fully cooperative (i.e. pursue a common goal) or there is a presence of a fully-trusted mediator. However, in competitive scenarios none of these assumptions hold – the involved agents do not want to completely reveal their future plans or their utility functions as this information can be exploited by other agents.

We aim to create a method of cooperation that works with non-cooperative (even adversarial) rational agents. In the method each agent uses an adversarial search to create a set of possible plans and then uses negotiation about selected future situations in order to improve its plans. Then they use negotiation about parts of the possible courses of actions and agree on commitments binding an agent to adapt a certain course of action in case the game reaches a specified state. However, the state does not have to be reached during the game. This way the agents can control their confidential plans and limit the amount of information given to the other agents. This concept is in the game theory studied as a pre-play communication or more specifically – cheap talk.

In this paper, we provide a feasibility study of the proposed negotiation-based approach. We tackle two main goals: (1) If such a pre-play communication (i.e. a negotiation prior to actual playing) is allowed among the adversarial-search-based agents, *how often* the agents can improve their utility value and what is the *amount of maximal possible improvement* of the utility value that an agent can gain by negotiation of strategies in sub-games with other agents? (2) In what way these values depend on parameters of the game, such as number of involved agents, number of possible actions, or correlation of utility functions of the involved agents? Answers to these questions are the primary contributions of this paper – we investigate a negotiation-space of utility values for complex competitive scenarios and provide a basic description of its dependence on general characteristics of games.

The rest of the paper is organized as follows. The following section contains brief review of related work. The next section describes the used game-theoretical model, and gives an example of a situation where an improvement of the utility can be reached via pre-play communication. We follow by describing a framework that allows measuring the possibilities of utility improvement in the described setting. Then we discuss the relevant properties of the real world domain and describe a class of games we use for the feasibility experiments presented afterwards.

Related Work

The problem of cooperation of self-interested agents is being widely studied, however, there is not much approaches that would incorporate the concerns about the plan confidentiality. The work closest to our approach is presented in [2], where authors model the multi-agent planning problem as a general-sum stochastic game and use negotiation to reach the subgame-perfect equilibrium. The crucial difference is that the authors do not consider the issue of plan confi-

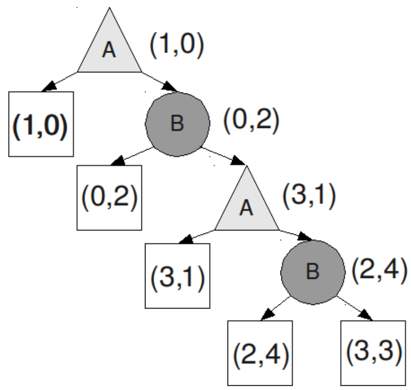

Fig. 1. Centipede Game – an example of a game where max^n does not find the optimal solution. Utility values in leafs are tuples of two elements, where the first agent (A) maximizes the first element and the second agent (B) the second one. There is a calculated max^n value next to the internal nodes.

dentiality hence agents negotiate the strategies of the whole game, while in our approach only strategies in sub-games are being considered.

In [3] authors work with the concept of Correlated Equilibria (which can be reached using a pre-play communication) in the context of games in extensive form. Their focus is however on two-player games, for which they present a polynomial algorithm, while our approach is not limited in the number of players.

In the field of multi-agent planning the cooperation is being tackled by either merging the plans that agents create separately (discussed e.g. in [1]), or by defining specific rules for the game by methods of mechanism design (e.g. shown in [4]). Solution for introducing the confidentiality of agents' plans applied for both cases is based on a mediator, to whom the agents reveal their plans. Our approach however, does not need the mediator and agents can limit the amount of revealing information about their plans. Distributed approach without a mediator is proposed in [5] using logical framework based on resources constrains, but only collaborative agents are considered, while our approach is applicable for more general self-interested agents with varying correlation of utility functions.

Problem Description

As the game-theoretical model for our approach we use n-player non-zero-sum games with perfect information in the extensive form. These models are represented as game-trees. The basic algorithm for finding a solution of the game (a *Nash equilibrium*[1]) is a known modification of the classical minimax algorithm – max^n [6]. The max^n algorithm uses backward induction and in each node chooses the branch that maximizes the utility value of the player associated

[1] An equilibrium state where no player has the incentive to deviate from it.

with the node. However, an equilibrium found by backward induction (termed *subgame-perfect equilibrium*[2]) is not optimal in terms of maximal utility values for each player. Example can be found in so called Centipede game (Figure 1), where the subgame-perfect equilibrium has lower utility values than the another outcome in the game – $(3, 3)$ – that improves utility values of both agents. We are interested in definition and analysis of such possible improvements of agents' utility values.

We have one important additional assumption. It is that agents trust each other in terms of committing to the strategy they agreed upon in the negotiation phase. This assumption might not be fully satisfied, but is a reasonable in case of longer or repeated interaction of the agents. The proposed mechanism is beneficial for all the agents involved so the agents are motivated to fulfill their commitments often enough so that the others do not refuse to negotiate any longer.

Game Definition

Based on the definition of extensive form games presented in [7], we define the turn-taking n-player non-zero-sum game in extensive form as a tuple $G = (\mathcal{I}, \mathcal{A}, \mathcal{H}, \mathcal{L}, \rho, \sigma, \pi, u)$, where:

$\mathcal{I}$ is a set of n players indexed $i = 1 \dots n$
$\mathcal{A}$ is set a of actions, $A = \bigcup_{i \in \mathcal{I}} \mathcal{A}_i$, where $\mathcal{A}_i$ is a set of actions an agent $i \in \mathcal{I}$ can perform
$\mathcal{H}$ is a set of non-terminal nodes in the game
$\mathcal{L}$ is a set of terminal nodes (leafs)
$\rho : \mathcal{H} \mapsto \mathcal{I}$ is the player function which assigns to each non-terminal node a player $i \in \mathcal{I}$
$\sigma : \mathcal{H} \times A_i \mapsto \mathcal{H} \cup \mathcal{L}$ is the transition function realizing one move of agent i in the game
π is a permutation on $\mathcal{I}$ representing the order of agents assigned to nodes on the path from the root of the tree towards the leafs. This order is repeated until leafs are reached.
$\overrightarrow{u} = (u_1, \dots, u_n)$ is global utility, where $u_i : \mathcal{L} \mapsto \mathbb{R}$ is a real-valued utility function for player i on leafs $\mathcal{L}$

□

Furthermore, we use a notation, where:

$r \in \mathcal{H}$ is the root of the game tree;
$\tau(h) = \{j \in \mathcal{H} \cup \mathcal{L}; (\exists a \in A_{\rho(h)} : j = \sigma(h, a))\}$ is a set of child nodes of a non-terminal node $h \in \mathcal{H}$
τ can be expanded for a set $\tau(H) = \bigcup_{h \in H} \tau(h)$ where $H \subseteq \mathcal{H}$.

[2] A refinement of a more general Nash equilibrium – the combination of players' strategies is in an equilibrium state if they are in Nash equilibrium for each smaller sub-game of the original game.

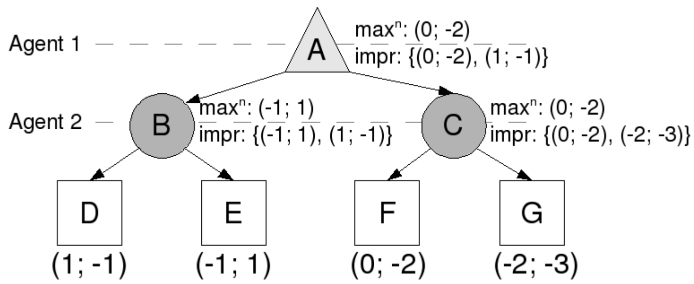

Fig. 2. Example of the max^n algorithm. Utility values in leafs are tuples of two elements, where Agent 1 maximizes the first element and Agent 2 the second one. The max^n values next to the nodes *A, B,* and *C* show calculated max^n value. The set *impr* corresponds to the set of improvements calculated by the algorithm in Figure 3

$d \in \mathbb{N}$ is the depth of the tree calculated as number of non-terminal nodes on the path from the root to any leaf divided by n: $\underbrace{\tau(\tau(\ldots\tau}_{d\cdot n}(r))\ldots) = \mathcal{L}$

Utility Improvement Analysis

Based on the definition of the game we present the algorithm for computing the maximal utility improvement for an agent. The example of the improvement is visualized in Figure 2. The max^n value of the game (the value in the root A) is inferior for both agents to the strategy leading to the leaf D. More precisely, for each node $h \in \mathcal{H}$ a max^n value $\overrightarrow{v}^{maxn} \in \mathbb{R}^n$ is calculated, and it corresponds to max^n value of the child node, for which the $v^{maxn}_{\rho(h)}$is maximal. For leaf $l \in \mathcal{L}$, $\overrightarrow{v}^{maxn} = \overrightarrow{u}(l)$. The max^n value $\overrightarrow{v}^{maxn}$ of the root of the tree represents the max^n value of the game, together with the max^n solution of the game, which is the path on which was the value propagated from the leaf with utility value equal to $\overrightarrow{v}^{maxn}$.

The algorithm for the calculation of the maximal utility improvement is inspired by the *soft-max*^{n} algorithm described in [8] – a modification of the max^n algorithm. The main idea is that, in contrary to max^n algorithm, the *soft-max*^{n} algorithm remembers not only a single best value in each node, but it remembers all best values in case the maximal value is not strict. Propagation of multiple values reflects uncertainty about the choice in a node in case more child nodes have the same utility value. Hence, instead of finding a single solution, which can be seen as a path from the root to a leaf in the game-tree, the *soft-max*^{n} algorithm considers a sub-tree of possible solutions. More precisely, the *soft-max*^{n} value of the node $h \in \mathcal{H}$ is a set of utility values $\overrightarrow{v}^{sfmaxn} \in \mathbb{R}^n$, for which the $\rho(h)$-th index is equal to the associated max^n value.

In our algorithm that computes maximal improvement of the utility value in the game we adopt the idea of multiple values propagation. The pseudocode

Input: $h \in \mathcal{H}$ current node; $i \in \mathcal{I}$ player performing the search
Output: $(\overrightarrow{v}^{maxn}, V^{impr})$ is a tuple, where $\overrightarrow{v}^{maxn}$ is a max^n utility value of the game, and V^{impr} is a set of utility values containing possible improvements.

1 $\overrightarrow{v}^{\text{maxn}} \leftarrow (-\infty, \ldots, -\infty)$
2 $V^{impr} \leftarrow \emptyset$
3 **if** $h \in \mathcal{L}$ **then**
4   **return** $(u(h), \emptyset)$
5 **end**
6 **foreach** $c \in \tau(h)$ **do**
7     $(v', V') \leftarrow$imprMaxN(c, i)
8   **if** $v^{maxn}_{\rho(h)} < v'_{\rho(h)}$ **then**
9         $\overrightarrow{v}^{\text{maxn}} \leftarrow v'$
10   **end**
11     $V^{impr} \leftarrow V^{impr} \cup V' \cup \{v'\}$
12 **end**
13 **if** $\rho(h) = i$ **then**
14   **foreach** $v' \in V^{impr}$ **do**
15   **if** $v^{maxn}_i > v'_i$ **then**
16             $V^{impr} \leftarrow V^{imrp} \setminus v'$
17   **end**
18   **end**
19 **end**
20 **return** $(\overrightarrow{v}^{\text{maxn}}, V^{impr})$

Fig. 3. imprMaxN(c, i) – algorithm for calculation of the *impr-max*^{n} value of the game.

of the algorithm is depicted in Figure 3. Let us assume, that we run the search algorithm for an agent $i \in \mathcal{I}$. Now, for each node $h \in \mathcal{H}$ we calculate a so called *impr-max*^{n} value that is a tuple $(\overrightarrow{v}^{maxn}, V^{impr})$, where $\overrightarrow{v}^{maxn}$ is the max^n value of the node (set in line 9), and V^{impr} is a set of utility values of other child nodes (line 11). For the nodes that are assigned to agent i we remove from the set V^{impr} those utility values where the searching agent i has lower utility value than its max^n value (lines 13-19).

After agent i calculates the *impr-max*^{n} $= (\overrightarrow{v}^{\text{maxn}}, V^{impr})$ value of the game tree using algorithm in Figure 3, we say that there is an improvement possible in the game iff

$$\exists \overrightarrow{v}' \in V^{impr}, \forall j \in \mathcal{I} : v^{maxn}_j \leq v'_j \tag{1}$$

Let $\mathcal{M}$ be a set of all such utility values v' for which formula 1 holds, and

$$v^{impr\_max}_i = \arg\max_{v' \in \mathcal{M}} v'_i$$

be the utility value with maximal value for agent i. Using this maximal value from the set $\mathcal{M}$ we can now define the amount of the utility improvement. We define this amount as a rate:

$$Impr = \frac{v_i^{maxn} - v_i^{impr\_max}}{2 \cdot 1.96 \cdot w} \quad (2)$$

where w is a standard deviation of the probability distribution of all utility values in leafs. We normalize the improvement of utility difference to the width of the interval that contains 95% of all utility values in leafs in case of normal distribution – we divide the difference of the utility values with the term $(2 \cdot 1.96 \cdot w)$. The normalization is essential in order to put the difference of the utility values to the context of all utility values in the game (i.e. if the $v_i^{impr\_max}$ is 120% of the max^n value it can either represent small improvement of bad max^n value, or good improvement of good man^n solution).

Reaching the Utility Improvement via Negotiation

In order to reach the utility improvement in a multi-agent environment a negotiation-based algorithm can be used. The solution with the $v_i^{impr\_max}$ value (or any other solution with the utility value in the set $\mathcal{M}$) can be reached via negotiation about specific parts of the game tree (termed *sub-tree*) that represent a single turn of the game (a sub-game where each agent moves once) – it is a sub-tree with a root node $h \in \mathcal{H}$ where the searching player makes a decision $\rho(h) = i$.

As this paper acts as a proof of concept whether it is reasonable to consider such a negotiation between self-interested agents, the $v_i^{impr\_max}$ value represents the maximal amount of the utility improvement. We investigate the ideal case from the view of searching agent i, where the other agents always agree with her proposition in case they do not lower their own utility value. The $v_i^{impr\_max}$ value can therefore be reached by searching agent i by proposing the strategies that lead to its best solution in each subgame in the order from the bottom of the game-tree towards the root.

In general, this basic negotiation-based algorithm can be modified to be more fair – each of the agents can propose their solution to the other agents and reach a different value from the set $\mathcal{M}$.

Moreover, using a concept of negotiations about strategies in specific subtrees, the agents can control the confidentiality of their plans by limiting the amount of information they provide to the other agents. An example can be seen in limiting the minimal depth for the root of the sub-tree that agent can negotiate about – the agent can commit to a different strategy in the future, however, it has not revealed the action in the current turn.

Experiments

We perform the experimental evaluation on a class of abstract games with properties often found in real-world situations. We use an extensive-form game created with respect to the definition in the previous section for 2 or 3 players. The searching agent is the first one ($i = 1$), and we use a fixed order of agents where

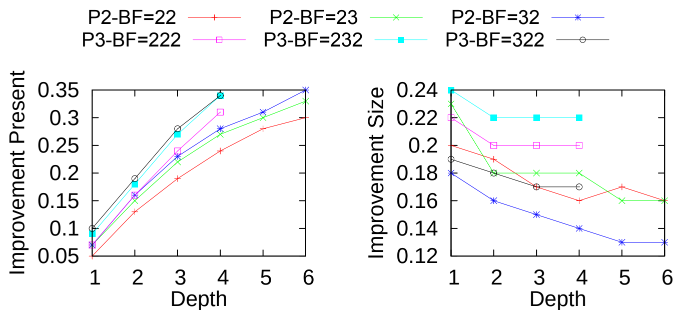

Fig. 4. Results of the dependence of the improvement presence (the left figure) and the amount of the improvement (the right figure) on depth (the x-axis). The legend: P2 represents games with 2 agents, P3 represents games with 3 agents; value of BF represents the branching factors for the agents in fixed order (each digit for one agent).

the π is equal to identity. We use varying branching factors (number of actions of agents) from the interval $2 \dots 3$, and we varied depth of the game $d = 1, \dots, 6$. Finally, we put an assumption on the utility values in leafs of the game tree. Often in complex competitive environments the agents have many possibilities of their future courses of actions that can slightly improve or slightly worsen their current situation. However, there is significantly less courses of actions that lead to a radical change of the utility value. Therefore, considering a part of the game where each agent moves only once, the utility values on leafs would form a normal distribution with center in the estimation of the utility value in the root of the game-tree. For a game with arbitrary number of moves this principle holds for each turn of the game (a sub-game where each of the agents moves once). Therefore, the utility values in leafs were generated with respect to this observation.

For each configuration of experimental parameters we created a set of 2000 variants (with different values on leafs of the tree). We are interested in two aspects: firstly, we want to know how often there is a utility improvement possible in games (i.e. the number of games where $\mathcal{M} \neq \emptyset$ divided by number of all games; we refer to this aspect as the *improvement presence*). Secondly, we are interested in the amount of the utility improvement $Impr$ as defined in formula 2 (we also use a term *improvement size*).

In the first experiment setting, we analyzed the dependence of these two aspects on the depth of the game tree, number of agents involved in the game and the branching factors (BFs) of agents. The results depicted in Figure 4 show, that the improvement presence increases with the depth of the tree, while the maximal value of the utility improvement is slightly decreasing. Both of these results are expected: in the first case, there is an increasing opportunity for occurrence of situations similar to the one visualized in Figure 2 with increasing size of the game tree. For the second case we argue that the game tree has

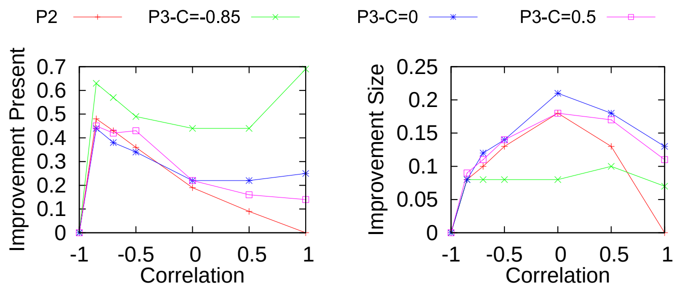

Fig. 5. Results of the dependence of the improvement presence (the left figure) and the amount of the improvement (the right figure) on correlation of agents' utility functions (the x-axis). The legend: P2 represents games with 2 agents, P3 represents games with 3 agents; in P3, the value C represents the utility correlation between the first and the third agent; the correlation between the first and the second agent is varying at the x-axis.

wider interval of all utility values in leafs, hence the improvement $Impr$ is slowly decreasing in spite of fact that there is a higher presence of improvements.

Utility Correlation

In the next experiment setting, we analyzed the impact of the correlation of the agent's utilities on the monitored aspects. We use pairwise correlation with the searching agent $i = 1$ – i.e. in 3-agents scenario we analyzed the impact by correlating the utility functions of the first and the second agent, and of the first and the third agent.

The experiment results are visualized in Figure 5. The results show an interesting phenomenon – the negative correlation between agents highly increases the improvement presence in games. This phenomenon is present due to the fact that in positively correlated examples the max^n algorithm is able to find already good solution, however, it is unnecessarily cautious in the case of the negative correlation.

For the value of the maximal improvement, the results are as expected – negative correlation directly negatively affects the amount of utility improvement, while in case of positive correlation, there is not much space left for utility improvement as the agents are driven by the utility to the "implicit cooperation".

Conclusions

In this paper we investigated the problem of cooperation of self-interested agents with respect to the confidentiality of their plans. We used extensive-form games and the concept of pre-play communication to analyze the negotiation-space of maximal improvement of utility values for an agent that is reachable using a negotiation-based algorithm. We proposed an algorithm and experimental evaluation to answer two main questions: (1) how often and how much can a self-interested agent improve its utility via sub-game negotiation, and (2) how these measures depend on basic properties of the game such as number of agents involved, size of the branching factor, and correlation of the utility functions of involved agents.

The results showed, that there is quite large potential in using the pre-play communication even for domains with self-interested agents based on adversarial search. Particularly interesting is the high improvement presence when the agents' utility functions are correlated negatively.

We point out the aspect of trust in the problem description. Currently, the agents always keep their promises and use the strategy they agreed upon. However, if the adversarial search in a later stage of the game explores consequences of an agreed strategy to higher depths, the agreement may become unprofitable for both the agents. This problem of horizon can be overcome by allowing cancellation of the agreements under certain conditions.

We would address this issue in our future work, together with a negotiation-based algorithm that gives us better perspective of applicability this concept in real multi-agent environments.

Acknowledgments. This research was partially funded by Air Force Office of Scientific Research USAF grant number FA8655-09-1-3060, the Ministry of Education of the Czech Republic grants ME09053 and MSM6840770038, and the Grant Agency of the Czech Technical University in Prague, grant No. SGS10/188/OHK3/2T/13.

References

1. Weerdt, M.D., Mors, A.T., Witteveen, C.: Multi-agent planning: An introduction to planning and coordination. Technical report, In: Handouts of the European Agent Summer (2005)
2. Murray, C., Gordon, G.J.: Multi-robot negotiation: Approximating the set of subgame perfect equilibria in general-sum stochastic games. In: NIPS. (2006) 1001–1008
3. von Stengel, B., Forges, F.: Extensive-form correlated equilibrium: Definition and computational complexity. Math. Oper. Res. **33**(4) (2008) 1002–1022
4. van der Krogt, R., de Weerdt, M., Zhang, Y.: Of mechanism design and multiagent planning. In: Proceedings of ECAI'08. (2008) 423–427
5. Ma, J., Russo, A., Broda, K., Lupu, E.: Multi-agent planning with confidentiality. In: Proceedings of AAMAS'09. (2009) 1275–1276
6. Luckhardt, C., K.B.Irani: An algorithmic solution of n-person games. In: Proc. of the National Conference on Artificial Intelligence (AAAI-86). (1986) 158–162
7. Shoham, Y., Leyton-Brown, K.: Multiagent systems: Algorithmic, game-theoretic, and logical foundations. (February 2009) 1–532
8. Sturtevant, N., Bowling, M.: Robust game play against unknown opponents. In: Proceedings of AAMAS '06. (2006) 713–719