Approximate Solutions of Interactive Dynamic Influence Diagrams Using ϵ-Behavioral Equivalence

Muthukumaran C.
Institute for AI (IAI)
University of Georgia
Athens, GA 30602
mkran@uga.edu

Prashant Doshi
Dept. of Computer Science and IAI
University of Georgia
Athens, GA 30602
pdoshi@cs.uga.edu

Yifeng Zeng
Dept. of Computer Science
Aalborg University
DK-9220 Aalborg, Denmark
yfzeng@cs.aau.dk

Abstract

Interactive dynamic influence diagrams (I-DID) are graphical models for sequential decision making in uncertain settings shared by other agents. Algorithms for solving I-DIDs face the challenge of an exponentially growing space of candidate models ascribed to other agents, over time. Pruning the behaviorally equivalent models is one way toward identifying a *minimal* model set. We further reduce the complexity by pruning models that are approximately behaviorally equivalent. Toward this, we redefine behavioral equivalence in terms of the distribution over the subject agent's future action-observation paths, and introduce the notion of ϵ-behavioral equivalence. We present a new approximation method that reduces the candidate models by pruning models that are ϵ-behaviorally equivalent with representative ones.

1 Introduction

Interactive dynamic influence diagrams (I-DID) (Doshi, Zeng, & Chen 2009) are graphical models for sequential decision making in uncertain multiagent settings. I-DIDs concisely represent the problem of how an agent should act in an uncertain environment shared with others who may act in sophisticated ways. I-DIDs may be viewed as graphical counterparts of interactive POMDPs (I-POMDPs) (Gmytrasiewicz & Doshi 2005), providing a way to model and exploit the embedded structure often present in real-world decision-making situations. They generalize DIDs (Tatman & Shachter 1990), which are graphical representations of POMDPs, to multiagent settings analogously to how I-POMDPs generalize POMDPs.

As we may expect, I-DIDs acutely suffer from both the curses of dimensionality and history. This is because the state space in I-DIDs includes the models of other agents in addition to the traditional physical states. These models encompass the agents' beliefs, action and sensory capabilities, and preferences, and may themselves be formalized as I-DIDs. The nesting is terminated at the 0^{th} level where the other agents are modeled using DIDs. As the agents act, observe, and update beliefs, I-DIDs must track the evolution of the models over time. Consequently, I-DIDs not only suffer from the curse of history that afflicts the modeling agent, but more so from that exhibited by the modeled agents. The exponential growth in the number of models over time also further contributes to the dimensionality of the state space. This is complicated by the nested nature of the space.

Previous approaches for approximating I-DIDs focus on reducing the dimensionality of the state space by limiting the number of candidate models of other agents. Using the insight that beliefs that are spatially close are likely to be behaviorally equivalent (Rathnas., Doshi, & Gmytrasiewicz 2006), Zeng et al. (2007) cluster the models of other agents and select representative models from each cluster. Intuitively, a cluster contains models that are likely to be behaviorally equivalent and hence may be replaced by a subset of representatives without a significant loss in the optimality of the decision maker. However, this approach often retains more models than needed. Doshi and Zeng (2009) formalize the concept of a *minimal set* of models using behavioral equivalence. At each step, only those models are updated which will result in predictive behaviors that are distinct from others in the updated model space. Minimal sets of models were previously discussed by Pynadath and Marsella (2007) which, in addition to discussing behavior equivalence proposed to further cluster models using utility equivalence. Notice that models that are behaviorally equivalent are also utility equivalent for the subject agent. We are currently investigating the applicability of utility equivalence in the context of I-DIDs.

In this paper, we aim to reduce the model space by additionally pruning models that are approximately behaviorally equivalent. Toward this objective, we introduce the concept of *ϵ-behavioral equivalence* among candidate models. In doing so, we redefine behavioral equivalence as the class of models of the other agents that induce an identical distribution over the subject agent's future action-observation paths in the interaction. Subsequently, models that induce distributions over the paths, which are no more than $\epsilon \geq 0$ apart are termed as being ϵ-behaviorally equivalent. Intuitively, this results in a lesser number of equivalence classes in the partition. If we pick a single representative model from each class, we typically end up with no more models than in the minimal set which need be solved thereby improving on approaches that utilize exact behavioral equivalence.

We begin by selecting a model at random and grouping together ϵ-behaviorally equivalent models with it. We repeat this procedure for the remaining models until all models have been grouped. The retained model set consists of the representative model from each equivalence class. In the worst case ($\epsilon = 0$), our approach identifies exact behavioral equivalence and the model set consists of all the behaviorally unique models. We discuss the error introduced by this approach in the optimality of the solution. More importantly, we experimentally evaluate our approach on I-DIDs formulated for a benchmark problem, and mention its limitations.

2 Background: Interactive DID

2.1 Syntax

In addition to the usual chance (oval), decision (rectangular), and utility (diamond shaped) nodes, I-IDs include a new type of node called the *model node* (hexagonal node, $M_{j,l-1}$, in Fig. $1(a)$). We note that the probability distribution over the chance node, S, and the model node together represents agent i's belief over its *interactive state space*. In addition to the model node, I-IDs differ from IDs by having a chance node, A_j, that represents the distribution over the other agent's actions, and a dashed link, called a *policy link*.

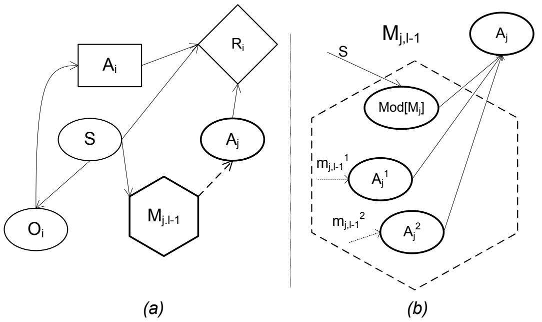


Figure 1: (a) A generic level $l > 0$ I-ID for agent i situated with one other agent j. The hexagon is the model node ($M_{j,l-1}$) and the dashed arrow is the policy link. (b) Representing the model node and policy link using chance nodes and dependencies.

The model node contains as its values the alternative computable models ascribed by i to the other agent. We denote the set of these models by $\mathcal{M}_{j,l-1}$. A model in the model node may itself be an I-ID or ID, and the recursion terminates when a model is an ID or a simple probability distribution over the actions. Formally, we denote a model of j as, $m_{j,l-1} = \langle b_{j,l-1}, \hat{\theta}_j \rangle$, where $b_{j,l-1}$ is the level $l-1$ belief, and $\hat{\theta}_j$ is the agent's *frame* encompassing the action, observation, and utility nodes. We observe that the model node and the dashed policy link that connects it to the chance node, A_j, could be represented as shown in Fig. $1(b)$. The decision node of each level $l-1$ I-ID is transformed into a chance node. Specifically, if OPT is the set of optimal actions obtained by solving the I-ID (or ID), then $Pr(a_j \in A_j^1) = \frac{1}{|OPT|}$ if $a_j \in OPT$, 0 otherwise. The conditional probability table (CPT) of the chance node, A_j, is a *multiplexer*, that assumes the distribution of each of the action nodes (A_j^1, A_j^2) depending on the value of $Mod[M_j]$. In other words, when $Mod[M_j]$ has the value $m_{j,l-1}^1$, the chance node A_j assumes the distribution of the node A_j^1, and A_j assumes the distribution of A_j^2 when $Mod[M_j]$ has the value $m_{j,l-1}^2$. The distribution over $Mod[M_j]$, is i's belief over j's models given the state. For more than two agents, we add a model node and a chance node representing the distribution over an agent's action linked together using a policy link, for each other agent.

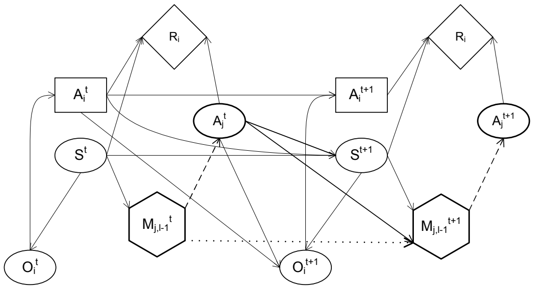


Figure 2: A generic two time-slice level l I-DID for agent i.

I-DIDs extend I-IDs to allow sequential decision making over several time steps (see Fig. 2). In addition to the model nodes and the dashed policy link, what differentiates an I-DID from a DID is the *model update link* shown as a dotted arrow in Fig. 2. We briefly explain the semantics of the model update next.

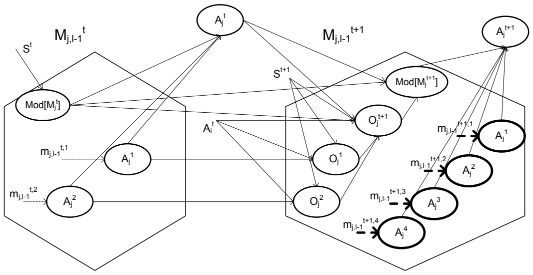


Figure 3: The semantics of the model update link. Notice the growth in the number of models at $t + 1$ shown in bold.

The update of the model node over time involves two steps: First, given the models at time t, we identify the updated set of models that reside in the model node at time $t+1$. Because the agents act and receive observations, their models are updated to reflect their changed beliefs. Since the set of optimal actions for a model could include all the actions, and the agent may receive any one of $|\Omega_j|$ possible observations, the updated set at time step $t + 1$ will have up to $|\mathcal{M}_{j,l-1}^t||A_j||\Omega_j|$ models. Here, $|\mathcal{M}_{j,l-1}^t|$ is the number of models at time step t, $|A_j|$ and $|\Omega_j|$ are the largest spaces of actions and observations respectively, among all the models. The CPT of $Mod[M_{j,l-1}^{t+1}]$ encodes the function, $\tau(b_{j,l-1}^t, a_j^t, o_j^{t+1}, b_{j,l-1}^{t+1})$ which is 1 if the belief $b_{j,l-1}^t$ in the model $m_{j,l-1}^t$ using the action a_j^t and observation o_j^{t+1} updates to $b_{j,l-1}^{t+1}$ in a model $m_{j,l-1}^{t+1}$; otherwise it is 0. Second, we compute the new distribution over the updated models, given the original distribution and the probability of the agent performing the action and receiving the observation that led to the updated model. The dotted model update link in the I-DID may be implemented using standard dependency links and chance nodes, as shown in Fig. 3 transforming it into a flat DID.

2.2 Behavioral Equivalence and Solution

Although the space of possible models is very large, not all models need to be considered in the model node. Models that are *behaviorally equivalent* (Pynadath & Marsella 2007; Rathnas., Doshi, & Gmytrasiewicz 2006) – whose behavioral predictions for the other agent are identical – could be pruned and a single representative model considered. This is because the solution of the subject agent's I-DID is affected by the predicted behavior of the other agent only; thus we need not distinguish between behaviorally equivalent models. Let **BehavioralEq**($\mathcal{M}_{j,l-1}$) be the procedure that prunes the behaviorally equivalent models from $\mathcal{M}_{j,l-1}$ returning the set of representative models.

The solution of an I-DID (and I-ID) proceeds in a bottom-up manner, and is implemented recursively as shown in Fig. 4. We start by solving the level 0 models, which may be traditional DIDs. Their solutions provide probability distributions which are entered in the corresponding action nodes found in the model node of the level 1 I-DID. The solution method uses the standard look-ahead technique, projecting the agent's action and observation sequences forward from the current belief state, and finding the possible beliefs that i could have in the next time step. Because agent i has a belief over j's models as well, the look-ahead includes finding out the possible models that j could have in the future. Consequently, each of j's level 0 models represented using a standard DID in the first time step must be solved to obtain its optimal set of actions. These actions are combined with the set of possible observations that j could make in that model, resulting in an updated set of candidate models (that include the updated beliefs) that could describe the behavior of j. $SE(b_j^t, a_j, o_j)$ is an abbreviation for the belief update. The updated set is minimized by excluding the behaviorally equivalent models. Beliefs over these updated set of candidate models are calculated using the standard inference methods through the dependency links between the model nodes (Fig. 3). The algorithm in Fig. 4 may be realized using the standard implementations of DIDs.

3 Redefining Behavioral Equivalence

We assume that the models of j have identical frames and differ only in their beliefs. As mentioned previously, two models of the other agent are behaviorally equivalent (BE) if they produce identical behaviors for the other agent. More formally, models $m_{j,l-1}, \hat{m}_{j,l-1} \in \mathcal{M}_{j,l-1}$ are BE if and only if $OPT(m_{j,l-1}) = OPT(\hat{m}_{j,l-1})$, where $OPT(\cdot)$ denotes the solution of the model that forms the argument. If the model is a DID or an I-DID, its solution is a policy tree.

Our aim is to identify models that are *approximately* BE. While a pair of policy trees may be checked for equality, disparate policy trees do not directly permit intuitive behavioral comparisons. This makes it difficult to define a measure of approximate BE, motivating investigations into a more rigorous formalization of BE.

```
I-DID EXACT(level $l \geq 1$ I-DID or level 0 DID, $T$)
Expansion Phase
For $t$ from $1$ to $T-1$ do
    If $l \geq 1$ then
        Populate $M_{j,l-1}^{t+1}$
        For each $m_j^t$ in $\mathcal{M}_{j,l-1}^t$ do
            Recursively call algorithm with the $l-1$ I-DID(or DID)
            that represents $m_j^t$ and the horizon, $T-t$
            Map the decision node of the solved I-DID (or DID),
            $OPT(m_j^t)$, to the chance node $A_j^t$
            For each $a_j$ in $OPT(m_j^t)$ do
                For each $o_j$ in $O_j$ (part of $m_j^t$) do
                    Update $j$'s belief, $b_j^{t+1} \leftarrow SE(b_j^t, a_j, o_j)$
                    $m_j^{t+1} \leftarrow$ New I-DID (or DID) with $b_j^{t+1}$ as belief
                    $\mathcal{M}_{j,l-1}^{t+1} \overset{\cup}{\leftarrow} \{m_j^{t+1}\}$
        Add the model node, $M_{j,l-1}^{t+1}$, and the model update link
        between $M_{j,l-1}^t$ and $M_{j,l-1}^{t+1}$
    Add the chance, decision and utility nodes for $t$+1 time slice
    and the dependency links between them
    Establish the CPTs for each chance node and utility node
Solution Phase
If $l \geq 1$ then
    Represent the model nodes and the model update link
    as in Fig. 3 to obtain the DID
    Minimize model spaces
    For $t$ from $1$ to $T$ do
        $\mathcal{M}_{j,l-1}^t \leftarrow$ BehavioralEq($\mathcal{M}_{j,l-1}^t$)
Apply the standard look-ahead and backup method to solve the
expanded DID (other solution approaches may also be used)
```

Figure 4: Algorithm for exactly solving a level $l \geq 1$ I-DID or level 0 DID expanded over T time steps.

Recall that BE models impact the decision-making of the modeling agent similarly, thereby motivating interest in grouping such models together. We utilize this insight toward introducing a new definition of BE. Let $h = \{a_i^t, o_i^{t+1}\}_{t=1}^T$ be the action-observation path for the modeling agent i, where o_i^{T+1} is null for a T horizon problem. If $a_i^t \in A_i$ and $o_i^{t+1} \in \Omega_i$, where A_i and Ω_i are i's action and observation sets respectively, then the set of all paths is, $H = \Pi_1^T(A_i \times \Omega_i)$, and the set of action-observation histories up to time t is $H^t = \Pi_1^{t-1}(A_i \times \Omega_i)$. The set of future action-observation paths is, $H_{T-t} = \Pi_t^T\ (A_i \times \Omega_i)$, where t is the current time step.

We observe that agent j's model together with agent i's perfect knowledge of its own model and its action-observation history induces a predictive distribution over i's future action-observation paths. This distribution plays a critical role in our approach and we denote it as, $Pr(H_{T-t}|h^t, m_{i,l}, m_{j,l-1}^t)$, where $h^t \in H^t$, $m_{i,l}$ is i's level l I-DID and $m_{j,l-1}^t$ is the level $l-1$ model of j in the model node at time t. For the sake of brevity, we rewrite the distribution term as, $Pr(H_{T-t}|m_{i,l}^t, m_{j,l-1}^t)$, where $m_{i,l}^t$ is i's horizon $T-t$ I-DID with its initial belief updated given the actions and observations in h^t. We define BE below:

Definition 1 (Behavioral Equivalence)**.** *Two models of agent j, $m^t_{j,l-1}$ and $\hat{m}^t_{j,l-1}$, are behaviorally equivalent if and only if $Pr(H_{T-t}|m^t_{i,l}, m^t_{j,l-1}) = Pr(H_{T-t}|\ m^t_{i,l}, \hat{m}^t_{j,l-1})$, where H_{T-t} and $m^t_{i,l}$ are as defined previously.*

In other words, BE models are those that induce an identical distribution over agent i's future action-observation history. This reflects the fact that such models impact agent i's behavior similarly.

Let h_{T-t} be some future action-observation path of agent i, $h_{T-t} \in H_{T-t}$. In Proposition 1, we provide a recursive way to arrive at the probability, $Pr(h_{T-t}|m^t_{i,l}, m^t_{j,l-1})$. Of course, the probabilities over all possible paths sum to 1.

Proposition 1. $Pr(h_{T-t}|m^t_{i,l}, m^t_{j,l-1}) = Pr(a^t_i, o^t_i|m^t_{i,l}, m^t_{j,l-1}) \sum_{a^t_j, o^{t+1}_j} Pr(h_{T-t-1}|m^{t+1}_{i,l}, m^{t+1}_{j,l-1}) \times Pr(a^t_j, o^{t+1}_j|a^t_i, m^t_{i,l}, m^t_{j,l-1})$
where

$$Pr(a^t_i, o^{t+1}_i|m^t_{i,l}, m^t_{j,l-1}) = Pr(a^t_i|OPT(m^t_{i,l}))\ \sum_{a^t_j} Pr(a^t_j| OPT(m^t_{j,l-1})) \sum_{s^{t+1}} O_i(s^{t+1}, a^t_i, a^t_j, o^{t+1}_i) \times \sum_{s,m_j} T_i(s, a^t_i, a^t_j, s^{t+1})\ b^t_{i,l}(s, m_j) \tag{1}$$

and

$$Pr(a^t_j, o^{t+1}_j|a^t_i, m^t_{i,l}, m^t_{j,l-1}) = Pr(a^t_j|OPT(m^t_{j,l-1}))\ \sum_{s^{t+1}} O_j(s^{t+1}, a^t_j, a^t_i, o^{t+1}_j) \sum_{s,m_j} T_i(s, a^t_i, a^t_j, s^{t+1}) b^t_{i,l}(s, m_j) \tag{2}$$

In Eq. 1, $O_i(s^{t+1}, a^t_i, a^t_j, o^{t+1}_i)$ is i's observation function contained in the CPT of the chance node, O^{t+1}_i, in the I-DID, $T_i(s, a^t_i,\ a^t_j, s^{t+1})$ is i's transition function contained in the CPT of the chance node, S^{t+1}, $Pr(a^t_i|OPT(m^t_{i,l}))$ is obtained by solving agent i's I-DID, $Pr(a^t_j|OPT(m^t_{j,l-1}))$ is obtained by solving j's model and appears in the CPT of node, A^t_j. In Eq. 2, $O_j(s^{t+1}, a^t_j,\ a^t_i, o^{t+1}_j)$ is j's observation function contained in the CPT of the chance node, O^{t+1}_j, given j's model is $m^t_{j,l-1}$.

Now that we have a way of computing the distribution over the future paths, we may relate Definition 1 to our previous understanding of BE models:

Proposition 2 (Correctness)**.** *$Pr(H_{T-t}|m^t_{i,l}, m^t_{j,l-1}) = Pr(H_{T-t}|m^t_{i,l},\ \hat{m}^t_{j,l-1})$ if and only if $OPT(m^t_{j,l-1}) = OPT(\hat{m}^t_{j,l-1})$, where $m^t_{j,l-1}$ and $\hat{m}^t_{j,l-1}$ are j's models.*

A simple method for computing the distribution over the paths given models of i and j is to replace agent i's decision nodes in the I-DID with chance nodes so that $Pr(a_i \in A^t_i) = \frac{1}{|OPT(m^t_{i,l})|}$ and remove the utility nodes, thereby transforming the I-DID into a dynamic Bayesian network (DBN). The desired distribution is then the marginal over the chance nodes that represent i's actions and observations with j's model entered as evidence in the Mod node at t.

4 ϵ-Behavioral Equivalence

4.1 Definition

We introduce the notion of ϵ-behavioral equivalence (ϵ-BE):

Definition 2 (ϵ-BE)**.** *Given $\epsilon \geq 0$, two models, $m^t_{j,l-1}$ and $\hat{m}^t_{j,l-1}$, are ϵ-BE if the divergence between the distributions $Pr(H_{T-t}|m^t_{i,l}, m^t_{j,l-1})$ and $Pr(H_{T-t}|m^t_{i,l}, \hat{m}^t_{j,l-1})$ is no more than ϵ.*

Here, the distributions over i's future paths are computed as shown in Proposition 1. While multiple ways to measure the divergence between distributions exist, we utilize the well-known Kullback-Leibler (KL) divergence (Kullback & Leibler 1951) in its symmetric form, in this paper. Consequently, the models are ϵ-BE if,

$$D_{KL}(Pr(H_{T-t}|m^t_{i,l}, m^t_{j,l-1})||Pr(H_{T-t}|m^t_{i,l}, \hat{m}^t_{j,l-1})) \leq \epsilon$$

where $D_{KL}(p||p')$ denotes the symmetric KL divergence between distributions, p and p', and is calculated as:

$$D_{KL}(p||p') = \frac{1}{2}\sum_k \left(p(k) log \frac{p(k)}{p'(k)} + p'(k) log \frac{p'(k)}{p(k)} \right)$$

If $\epsilon = 0$, ϵ-BE collapses into exact BE. Sets of models exhibiting ϵ-BE for some non-zero but small ϵ do not differ significantly in how they impact agent i's decision making. These models could be candidates for pruning.

4.2 Approach

We proceed by picking a model of j at random, $m^{t=1}_{j,l-1}$, from the model node in the first time step, which we call the representative. All other models in the model node that are ϵ-BE with $m^{t=1}_{j,l-1}$ are grouped together with it. Of the remaining models, another representative is picked at random and the previous procedure is repeated. The procedure terminates when no more models remain to be grouped. We illustrate the process in Fig. 5. We point out that for $\epsilon > 0$, in general, more models will likely be grouped together than if we considered exact BE. This will result in a fewer number of classes in the partition.

We first observe that the outcome is indeed a partition of the model set into ϵ-BE classes. This is because we continue to pick representative models and build classes until no model remains ungrouped. There is no overlap between classes since new ones are built only from the models that did not get previously grouped. We observe that the representatives of different classes are ϵ-behaviorally distinct, otherwise they would have been grouped together. However, this set is not unique and the partition could change with different representatives. Furthermore, let $\hat{\mathcal{M}}_j$ be the largest set of behaviorally distinct models, also called the minimal set (Doshi & Zeng 2009). Then, the following holds:

Proposition 3 (Cardinality)**.** *The ϵ-BE approach results in at most $|\hat{\mathcal{M}}_j|$ models after pruning.*

Intuitively, the Proposition follows from the fact that in the worst case, ϵ = 0, resulting in behaviorally distinct models.

Transfer of probability mass From each class in the partition, the previously picked representative is retained and all other models are pruned. The representatives are distinguished in that all models in its group are ϵ-BE with it. Unlike exact BE, ϵ-BE relation is not necessarily transitive.

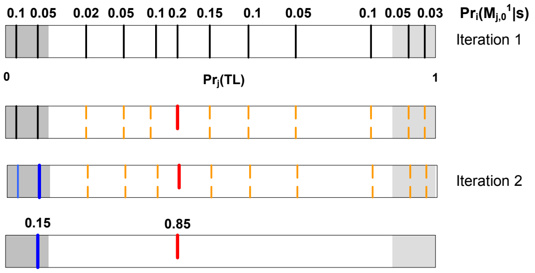


Figure 5: Illustration of the iterative ϵ-BE model grouping using the tiger problem. Black vertical lines denote the beliefs contained in different models of agent j included in the initial model node, $M_{j,0}^1$. Decimals on top indicate i's distribution over j's models. We begin by picking a representative model (red line) and grouping models that are ϵ-BE with it. Unlike exact BE, models in a different behavioral (shaded) region get grouped as well. Of the remaining models, another is selected as representative. Agent i's distribution over the representative models is obtained by summing the probability mass assigned to the individual models in each class.

Consequently, we may not select any model from each class as the representative since others may not be ϵ-BE with it.

Recall that agent i's belief assigns some probability mass to each model in the model node. A consequence of pruning some of the models is that the mass assigned to the models would be lost. Disregarding this probability mass may introduce further error in the optimality of the solution. We avoid this error by transferring the probability mass over the pruned models in each class to the ϵ-BE representative that is retained in the model node (see Fig. 5).

Sampling actions and observations Recall that the predictive distribution over i's future action-observation paths, $Pr(H_{T-t}|h^t, m_{i,l}, m_{j,l-1}^t)$, is conditioned on the history of i's observations, h^t, as well. Because the model grouping is performed while solving the I-DID when we do not know the actual history, we obtain a likely h^t by sampling i's actions and observations for subsequent time steps in the I-DID.

Beginning with the first time step, we pick an action, a_i^t, at random assuming that each action is equally likely. An observation is then sampled from the distribution given i's sampled action and belief, $o_i^{t+1} \sim Pr(\Omega_i|a_i^t, b_{i,l}^t)$, where $b_{i,l}^t$ is the prior belief. We utilize this sampled action and observation pair as the history, $h^t \xleftarrow{\cup} \langle a_i^t, o_i^{t+1}\rangle$. We may implement this procedure by entering as evidence i's action in the node, A_i^t, of the DBN (mentioned in Section 3) and sampling from the inferred distribution over the node, O_i^{t+1}.

Finally, we note that in computing the distribution over the paths, solution to agent i's I-DID is needed as well ($Pr(a_i^t|OPT(m_{i,l}^t))$ term in Eq. 1). As we wish to avoid this, we observe that ϵ-BE is based on the *comparative impact* that j's models have on i, which is independent of i's decisions. Therefore, we assume a uniform distribution over i's actions, $Pr(a_i^t|OPT(m_{i,l}^t)) = \frac{1}{|A_i|}$, which does not change the ϵ-BE of models.

5 Algorithm

We present the algorithm for partitioning the models in the model node of the I-DID at each time step according to ϵ-BE, in Fig. 6. The procedure, **ϵ-BehaviorEquivalence** replaces the procedure, **BehaviorEq**, in the algorithm in Fig. 4. The procedure takes as input, the set of j's models, $\mathcal{M}_j$, the agent i's DID, m_i, current time step and horizon, and the approximation parameter, ϵ. The algorithm begins by computing the distribution over the future paths of i for each model of j. If the time step is not the initial one, the prior action-observation history is first sampled. We may compute the distribution by transforming the I-DID into a DBN as mentioned in Section 3 and entering the model of j as evidence – this implements Eqs. 1 and 2.

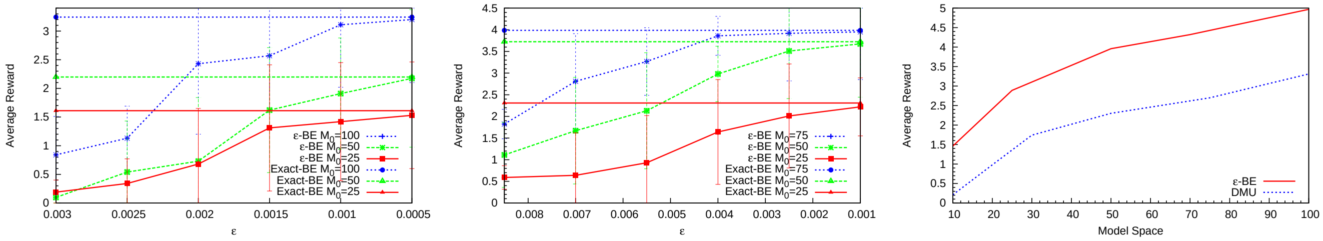
```
ε-BEHAVIOREQUIVALENCE(Model set $\mathcal{M}_j$, DID $m_i$,
current time step $tt$, horizon $T$, $\epsilon$) returns $\mathcal{M}'_j$
  Transform DID $m_i$ into DBN by replacing $i$'s decision nodes
  with chance nodes having uniform distribution
  For $t$ from 1 to $tt$ do
      Sample, $a_i^t \sim Pr(A_i^t)$
      Enter $a_i^t$ as evidence into chance node, $A_i^t$, of DBN
      Sample, $o_i^{t+1} \sim Pr(O_i^{t+1})$
      $h^t \xleftarrow{\cup} \langle a_i^t, o_i^{t+1}\rangle$
  For each $m_j^k$ in $\mathcal{M}_j$ do
      Compute the distribution, $P[k] \leftarrow Pr(H_{T-t}|h^t, m_i, m_j^k)$,
      obtained from the DBN by entering $m_j^k$ as evidence (Prop. 1)
Clustering Phase
  While $\mathcal{M}_j$ not empty
      Select a model, $m_j^{\hat{k}} \in \mathcal{M}_j$, at random
      Initialize, $\mathcal{M}_j^{\hat{k}} \leftarrow \{m_j^{\hat{k}}\}$
      For each $m_j^k$ in $\mathcal{M}_j$ do
          If$D_{KL}(P[\hat{k}]||P[k]) \leq \epsilon$
              $\mathcal{M}_j^{\hat{k}} \xleftarrow{\cup} m_j^k$,  $\mathcal{M}_j \xleftarrow{-} m_j^k$
Selection Phase
  For each $\mathcal{M}_j^{\hat{k}}$ do
      Retain the representative model, $\mathcal{M}'_j \xleftarrow{\cup} m_j^{\hat{k}}$
  Return $\mathcal{M}'_j$
```

Figure 6: Algorithm for partitioning j's models using ϵ-BE. This function replaces **BehaviorEq()** in Fig. 4.

We then pick a representative model at random, and using the cached distributions group together models whose distributions exhibit a divergence less than ϵ from the distribution of the representative model. We iterate over the models left ungrouped until none remain. Each iteration results in a new class of models including a representative. In the final selection phase, all models except the representative are pruned from each class in the partition. The set of representative models, which are ϵ-behaviorally distinct, are returned.

6 Computational Savings and Error Bound

As with previous approaches, the primary complexity of solving I-DIDs is due to the large number of models that must be solved over T time steps. At some time step t, there could be $|\mathcal{M}_j^0|(|A_j||\Omega_j|)^t$ many models of the other agent j, where $|\mathcal{M}_j^0|$ is the number of models considered initially.



(a) (b) (c)

Figure 7: Performance profile obtained by solving a level 1 I-DID for the multiagent tiger problem using the ϵ-BE approach for (a) 3 horizons and (b) 4 horizons. As ϵ reduces, quality of the solution improves and approaches that of the exact. (c) Comparison of ϵ-BE and DMU in terms of the rewards obtained given identical numbers of models in the initial model node after clustering and pruning.

The nested modeling further contributes to the complexity. In an N+1 agent setting, if the number of models considered at each level for an agent is bound by $|\mathcal{M}|$, then solving an I-DID at level l requires the solutions of $\mathcal{O}((N|\mathcal{M}|)^l)$ many models. As we mentioned in Proposition 3, the ϵ-BE approximation reduces the number of agent models at each level to at most the size of the minimal set, $|\hat{\mathcal{M}}^t|$. In doing so, it solves $|\mathcal{M}_j^0|$ many models initially and incurs the complexity of performing inference in a DBN for computing the distributions. This complexity while significant is less than that of solving DIDs. Consequently, we need to solve at most $\mathcal{O}((N|\hat{\mathcal{M}}^*|)^l)$ number of models at each non-initial time step, typically less, where $\hat{\mathcal{M}}^*$ is the largest of the minimal sets, in comparison to $\mathcal{O}((N|\mathcal{M}|)^l)$. Here $\mathcal{M}$ grows exponentially over time. In general, $|\hat{\mathcal{M}}| \ll |\mathcal{M}|$, resulting in a substantial reduction in the computation. Additionally, a reduction in the number of models in the model node also reduces the size of the state space, which makes solving the upper-level I-DID more efficient.

We assume that lower-level models of the other agent are solved exactly, and analyze the conditional error bound of this approach. In the trivial case, ϵ=0, and there is no optimality error in the solution. If we limit the pruning of ϵ-BE models to the initial model node, the error is due to transferring the probability mass of the pruned model to the representative, effectively replacing the pruned model with the representative. The maximum error in the solution of i's I-DID due to this transfer could be $(R_i^{max} - R_i^{min})T$, where T is the horizon of the I-DID. However, the divergence in the impact of the pruned model and the representative on i's action-observation path is no more than ϵ. Hence, the effective error bound is: $(R_i^{max} - R_i^{min})T \times \epsilon$.

Matters become more complex when we additionally prune models in the subsequent model nodes as well. This is because rather than comparing over distributions given each history of i, we sample i's action-observation history. Consequently, additional error incurs due to the sampling, which is difficult to bound. Finally, Doshi and Zeng (2009) show that it is difficult to usefully bound the error if lower-level models are themselves solved approximately. This limitation is significant because approximately solving lower-level models could bring considerable computational savings.

In summary, error in i's behavior due to pruning ϵ-BE models in the initial model node may be bounded, but we continue to investigate how to usefully bound the error due to multiple additional approximations.

7 Experimental Evaluation

We implemented the algorithms in Figs. 4 and 6 and show preliminary results for the well-known two-agent *tiger problem* ($|S|$=2, $|A_i|$=$|A_j|$=3, $|\Omega_i|$=6, $|\Omega_j|$=3) (Gmytrasiewicz & Doshi 2005). We formulate a level 1 I-DIDs for the problem, and solve them approximately for varying ϵ. We show that, (i) the quality of the solution generated using our approach (ϵ-BE) improves as we reduce ϵ for given numbers of initial models of the other agent, M_0, and approaches that of the exact solution; and (ii) in comparison to the approach of updating models discriminatively (DMU) (Doshi & Zeng 2009), which is the current efficient technique, ϵ-BE is able to obtain larger rewards for an identical number of initial models. This indicates a more informed clustering and pruning using ϵ-BE although it is less efficient in doing so.

In Fig. 7(a, b), we show the average rewards gathered by executing the policies obtained from solving the level 1 I-DIDs approximately. Each data point is the average of 300 runs where the true model of j is picked randomly according to i's belief. Notice that as we reduce ϵ the policies tend to converge to the exact (denoted by flat lines) and this remains true for different numbers of initial models. Values of these policies increase as i considers greater numbers of models thereby improving it's chances of modeling j correctly. Next, we compare the performance of this approach with that of DMU (Fig. 7(c)). While both approaches cluster and prune models, DMU does so only in the initial model node, thereafter updating only those models which on update will be behaviorally distinct. Thus, we compare the average rewards obtained by the approaches when an identical number of models remain in the initial model node after clustering and selection. This allows us to compare between the clustering and selection techniques of the two approaches. From Fig. 7(c), we observe that ϵ-BE results in better quality policies that obtain significantly higher average reward. This indicates that the models pruned by DMU were more valuable than those pruned by ϵ-BE, thereby testifying to the more informed way in which we compare between models by gauging the impact on i's history. DMU's approach of measuring simply the closeness of beliefs in models for clustering results in significant models being pruned. However, the tradeoff is the increased computational cost in calculating the distributions over future paths. To illustrate, ϵ-BE consumed an average of 9.1 secs in solving a 4 horizon I-DID with 25 initial models and differing ϵ, which represents approximately a three-fold increase compared to DMU.

8 Conclusion

Our initial results demonstrate the potential for obtaining flexible approximations of I-DIDs by pruning models that are approximately BE, and motivates further investigations. However, we face the challenge of computing distributions over a number of paths that grow exponentially with horizon. Nevertheless, we expect to be able to solve I-DIDs of longer time horizons in reasonable time and with larger numbers of models, as we optimize our implementation and seek ways to mitigate the curse of history.

Acknowledgement This research is supported in part by AFOSR grant #FA9550-08-1-0429 and in part by NSF CAREER grant IIS-0845036.

References

Doshi, P., and Zeng, Y. 2009. Improved approximation of interactive dynamic influence diagrams using discriminative model updates. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 907–914.

Doshi, P.; Zeng, Y.; and Chen, Q. 2009. Graphical models for interactive pomdps: Representations and solutions. *Journal of Autonomous Agents and Multiagent Systems (JAAMAS)* 18(3):376–416.

Gmytrasiewicz, P., and Doshi, P. 2005. A framework for sequential planning in multiagent settings. *Journal of Artificial Intelligence Research* 24:49–79.

Kullback, S., and Leibler, R. 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22(1):79–86.

Pynadath, D., and Marsella, S. 2007. Minimal mental models. In *Twenty-Second Conference on Artificial Intelligence (AAAI)*, 1038–1044.

Rathnas., B.; Doshi, P.; and Gmytrasiewicz, P. J. 2006. Exact solutions to interactive pomdps using behavioral equivalence. In *Autonomous Agents and Multi-Agent Systems Conference (AAMAS)*, 1025–1032.

Tatman, J. A., and Shachter, R. D. 1990. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics* 20(2):365–379.

Zeng, Y.; Doshi, P.; and Chen, Q. 2007. Approximate solutions of interactive dynamic influence diagrams using model clustering. In *Twenty Second Conference on Artificial Intelligence (AAAI)*, 782–787.