ϵ-Subjective Equivalence of Models for Interactive Dynamic Influence Diagrams

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

Muthukumaran Chandrasekaran
Institute for AI (IAI)
University of Georgia
Athens, GA 30602
mkran@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 behaviorally equivalent models is one way toward minimizing the model set. We seek to further reduce the complexity by additionally pruning models that are approximately subjectively equivalent. Toward this, we define subjective equivalence in terms of the distribution over the subject agent's future action-observation paths, and introduce the notion of ϵ-subjective equivalence. We present a new approximation technique that reduces the candidate model space by removing models that are ϵ-subjectively equivalent with representative ones.**

I. Introduction

Interactive dynamic influence diagrams (I-DID) [1] are recognized 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. They generalize DIDs [2] to multiagent settings, and provide a way to model and exploit the embedded structure often present in real-world decision-making situations. For comparisons with related graphical models, MAIDs [3] and NIDs [4], see [1].

I-DIDs acutely suffer from both the curses of dimensionality and history [5]. This is because the state space in I-DIDs includes the models of other agents as well. 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. Thus, 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 further contributes to the state space.

Previous approaches for solving I-DIDs [1], [6] focus on limiting the number of candidate models of other agents. Using the insight that beliefs that are spatially close are likely to be *behaviorally equivalent* [7], [8], Doshi, Zeng and Chen [1] 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 significant loss in the optimality of the decision maker. However, this approach often retains more models than needed. Doshi and Zeng [6] further minimize the model set. At each time step, only those models are updated which will result in predictive behaviors that are distinct from others in the updated model space. The initial set of models are solved and merged to obtain a policy graph, which assists in discriminating between model updates. Pynadath and Marsella [7] proposed utility equivalence to additionally cluster models; its applicability in the context of I-DIDs is not straight forward.

In this paper, we aim to reduce the model space by pruning models that are approximately subjectively equivalent. Toward this objective, we introduce the concept of ϵ-*subjective equivalence* among candidate models. We define subjective 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. We relate subjective equivalence to the previous concept of behavioral equivalence. Subsequently, models that induce distributions over the paths, which are no more than $\epsilon \geq 0$ apart are termed as being ϵ-subjectively equivalent. Intuitively, this results in a lesser number of equivalence classes in the partition than behavioral equivalence. If we pick a single representative model from each class, we typically end up with no more models than the number of subjectively distinct ones, which need be solved. This improves on approaches that utilize exact behavioral equivalence.

We begin by selecting a model at random and grouping together ϵ-subjectively 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 subjective equivalence and the model set consists of all the subjectively unique models. Our novel approach provides a unique opportunity to bound the error in optimality of the subject agent. Furthermore, we experimentally evaluate our approach on I-DIDs formulated for benchmark problem domains and show significant qualitative improvement. However, this improvement is tempered by increased time complexity of ascertaining ϵ-subjective equivalence of models.

II. Background: Interactive DID

We outline interactive influence diagrams (I-IDs) for two-agent interactions followed by their extensions to dynamic settings, I-DIDs [1].

A. Syntax

In addition to the usual 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*.

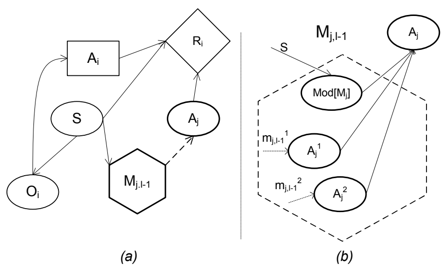


Figure 1. (a) A generic level $l > 0$ I-ID for agent i situated with one other agent j. (b) Representing the model node and policy link using chance nodes and dependencies.

The model node contains as its values the alternative computational 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 $Mod[M_j]$. The distribution over $Mod[M_j]$ is i's belief over j's models given the state.

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, I-DIDs include the *model update link* shown as a dotted arrow in Fig. 2. We briefly explain the semantics of the model update. The update of

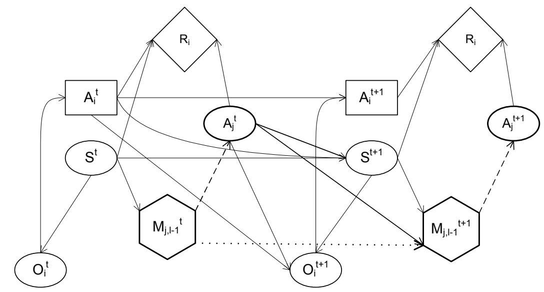


Figure 2. A generic two time-slice level l I-DID for agent i.

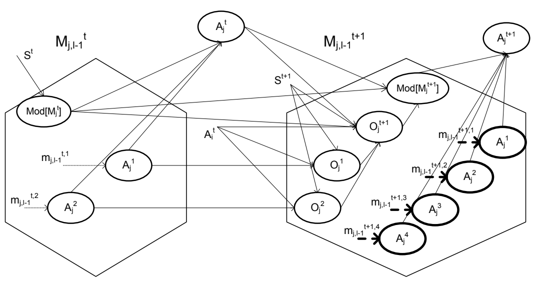


Figure 3. The semantics of the model update link. Notice the growth in the number of models at $t+1$ shown in bold.

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. The CPT of $Mod[M_{j,l-1}^{t+1}]$ 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 (Fig. 3) transforming it into a flat DID.

B. 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* [7], [8] – whose behavioral predictions for the 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.

The solution of an I-DID (and I-ID) proceeds in a bottom-up manner, and is implemented recursively. 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. This is done by combining j's actions obtained by solving its models with its possible observations. The updated set of j's models 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.

I-DID EXACT(level $l \geq 1$ I-DID or level 0 DID, T)

Expansion Phase

1. **For** t **from** 1 **to** $T-1$ **do**
2. **If** $l \geq 1$ **then**
   Minimize $M_{j,l-1}^t$
3. **For each** m_j^t **in** $\mathcal{M}_{j,l-1}^t$ **do**
4. Recursively call algorithm with the $l-1$ I-DID (or DID) that represents m_j^t and the horizon, $T-t$
5. Map the decision node of the solved I-DID (or DID), $OPT(m_j^t)$, to the chance node A_j^t
6. $\mathcal{M}_{j,l-1}^t \leftarrow$ **BehavioralEq**($\mathcal{M}_{j,l-1}^t$)
   Populate $M_{j,l-1}^{t+1}$
7. **For each** a_j **in** $OPT(m_j^t)$ **do**
8. **For each** o_j **in** O_j (part of m_j^t) **do**
9. Update j's belief, $b_j^{t+1} \leftarrow SE(b_j^t, a_j, o_j)$
10. $m_j^{t+1} \leftarrow$ New I-DID (or DID) with b_j^{t+1} as belief
11. $\mathcal{M}_{j,l-1}^{t+1} \xleftarrow{\cup} \{m_j^{t+1}\}$
12. Add the model node, $M_{j,l-1}^{t+1}$, and the model update link between $M_{j,l-1}^t$ and $M_{j,l-1}^{t+1}$
13. Add the chance, decision and utility nodes for t+1 time slice and the dependency links between them
14. Establish the CPTs for each chance node and utility node

Solution Phase

15. **If** $l \geq 1$ **then**
16. Represent the model nodes and the model update link as in Fig. 3 to obtain the DID
17. Apply the standard look-ahead and backup method to solve the expanded DID

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

III. Subjective Equivalence

We assume that the models of j have identical frames and differ only in their beliefs. Recall that models $m_{j,l-1}, \hat{m}_{j,l-1} \in \mathcal{M}_{j,l-1}$ are behaviorally equivalent 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 [8]. If the model is a DID or an I-DID, its solution is a policy tree. 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 behavioral equivalence, motivating further investigations.

We note that subsets of models may impact the decision making of the modeling agent similarly, thereby motivating interest in grouping such models together. We utilize this insight toward introducing the new concept of *subjective equivalence* (SE)[1]. 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 SE below:

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

In other words, SE models induce an identical distribution over agent i's future action-observation paths. This reflects the fact that such models impact i's behavior similarly and could be grouped.

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, $Pr(h_{T-t}|m_{i,l}^t, m_{j,l-1}^t)$. Of course, the probabilities over all possible paths sum to 1.

Proposition 1: $Pr(h_{T-t}|m_{i,l}^t, m_{j,l-1}^t) = Pr(a_i^t, o_i^{t+1}|m_{i,l}^t, m_{j,l-1}^t) \sum_{a_j^t, o_j^{t+1}} Pr(h_{T-t-1}|m_{i,l}^{t+1}, m_{j,l-1}^{t+1}) Pr(a_j^t, o_j^{t+1}|a_i^t, m_{i,l}^t, m_{j,l-1}^t)$

[1]We will use SE as an acronym for both, subjectively equivalent (adjective form) and subjective equivalence (noun form). Appropriate usage will be self-evident.

where

$$Pr(a_i^t, o_i^{t+1}|m_{i,l}^t, m_{j,l-1}^t) = Pr(a_i^t|OPT(m_{i,l}^t))\ \sum_{a_j^t} Pr(a_j^t| OPT(m_{j,l-1}^t)) \sum_{s^{t+1}} O_i(s^{t+1}, a_i^t, a_j^t, o_i^{t+1}) \times \sum_{s,m_j} T_i(s, a_i^t, a_j^t, s^{t+1})\ b_{i,l}^t(s, m_j) \tag{1}$$

$$Pr(a_j^t, o_j^{t+1}|a_i^t, m_{i,l}^t, m_{j,l-1}^t) = Pr(a_j^t|OPT(m_{j,l-1}^t))\ \sum_{s^{t+1}} O_j(s^{t+1}, a_j^t, a_i^t, o_j^{t+1}) \sum_{s,m_j} T_i(s, a_i^t, a_j^t, s^{t+1}) b_{i,l}^t(s, m_j) \tag{2}$$

In Eq. 1, $O_i(s^{t+1}, a_i^t, a_j^t, o_i^{t+1})$ is i's observation function contained in the CPT of the node, O_i^{t+1}, in the I-DID, $T_i(s, a_i^t,\ a_j^t, s^{t+1})$ is i's transition function contained in the CPT of the node, S^{t+1}, $Pr(a_i^t|OPT(m_{i,l}^t))$ is obtained by solving agent i's I-DID, $Pr(a_j^t|OPT(m_{j,l-1}^t))$ is obtained by solving j's model and appears in the CPT of A_j^t. In Eq. 2, $O_j(s^{t+1}, a_j^t,\ a_i^t, o_j^{t+1})$ is j's observation function contained in the CPT of the chance node, O_j^{t+1}, given j's model is $m_{j,l-1}^t$. Proposition 1 may be derived recursively over future paths and by noting that j's level $l-1$ actions and observations are independent of i's observations. We provide a concise proof in the Appendix.

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

Proposition 2: If $OPT(m_{j,l-1}^t) = OPT(\hat{m}_{j,l-1}^t)$, then $Pr(H_{T-t}|m_{i,l}^t, m_{j,l-1}^t) = Pr(H_{T-t}|m_{i,l}^t,\ \hat{m}_{j,l-1}^t)$, where $m_{j,l-1}^t$ and $\hat{m}_{j,l-1}^t$ are j's models.

Proof sketch: The proof is reducible to showing the above for some individual path, $h_{T-t} \in H_{T-t}$.

Given $OPT(m_{j,l-1}^t) = OPT(\hat{m}_{j,l-1}^t)$, we may write, $Pr(a_j^t|OPT(m_{j,l-1}^t)) = Pr(a_j^t|OPT(\hat{m}_{j,l-1}^t))$ for all a_j^t. Because all other terms in Eqs. 1 and 2 are identical, it follows that $Pr(h_{T-t}|m_{i,l}^t, m_{j,l-1}^t)$ must be same as $Pr(h_{T-t} \mid m_{i,l}^t,\ \hat{m}_{j,l-1}^t)$. ■

Consequently, the set of SE models includes those that are behaviorally equivalent. It further includes models that induce identical distributions over agent i's action-observation paths, but these models could be behaviorally distinct over those paths that have zero probability. Thus, these latter models may not be behaviorally equivalent. Doshi and Gmytrasiewicz [10] call these models as (strictly) observationally equivalent. Therefore, the converse of Prop. 2 is not true.

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 $Pr(a_i \in A_i^t) = \frac{1}{|OPT(m_{i,l}^t)|}$ 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.

IV. ϵ-SUBJECTIVE EQUIVALENCE

Our definition of SE formalizes the intuition that SE models impact the subject agent identically. While rigorous, it has the advantage that it permits us to measure the degree to which models are SE, allowing the introduction of *approximate SE*.

A. Definition

We introduce the notion of ϵ-subjective equivalence (ϵ-SE) and define it as follows:

Definition 2 (ϵ-SE): Given $\epsilon \geq 0$, two models, $m_{j,l-1}^t$ and $\hat{m}_{j,l-1}^t$, are ϵ-SE if the divergence between the distributions $Pr(H_{T-t}|m_{i,l}^t, m_{j,l-1}^t)$ and $Pr(H_{T-t}|m_{i,l}^t, \hat{m}_{j,l-1}^t)$ 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 [11] in its symmetric form, in part because its mathematical properties are well studied. Consequently, the models are ϵ-SE if,

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

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-SE collapses into exact SE. Sets of models exhibiting ϵ-SE for some non-zero but small ϵ do not differ significantly in how they impact agent i's decision making.

B. Approach

We proceed by picking a model of j at random, $m_{j,l-1}^{t=1}$, from the model node in the first time step, which we call the *representative*. All other models in the model node that are ϵ-SE with $m_{j,l-1}^{t=1}$ are grouped together. 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 SE. This results 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-SE 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 emerge only from the models that did not get previously grouped. We observe that the representatives of different classes are ϵ-subjectively distinct, otherwise they would have been grouped together. However, this set is not unique and the partition could change with different representatives.

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-SE with it. Unlike exact SE, ϵ-SE relation is not necessarily transitive. Consequently, we may not arbitrarily select a model from each class as the representative since others may not be ϵ-SE with it. Let $\hat{\mathcal{M}}_j$ be the largest set of behaviorally distinct models. Then, the following holds:

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

Intuitively, the Prop. follows from the fact that in the worst case, ϵ = 0, resulting in subjectively distinct models. This set is no larger than the set of behaviorally distinct models.

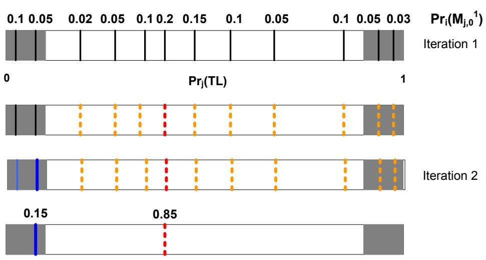


Figure 5. Illustration of iterative ϵ-SE model grouping using the multiagent tiger problem. Black vertical lines denote beliefs contained in different models of j included in the initial model node, $M^1_{j,0}$. Decimals on top indicate i's probability distribution over j's models. We pick a representative model (red line) and group models ϵ-SE with it. Unlike exact SE, models in a different behavioral (shaded) region also get grouped. Of the remaining models, another is selected as representative. i's distribution over the representatives is obtained by summing probabilities assigned to individual models in each class.

Transfer of probability mass 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-SE 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, $Pr(H_{T-t}|h^t,\ m_{i,l}, m^t_{j,l-1})$, is conditioned on the history of i's observations, h^t, as well. For a time-extended I-DID, because the model grouping is performed at every subsequent time step at which 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 I-DID.

Beginning with the first time step, we pick an action, a^t_i, at random assuming that each action is equally likely. An observation is then sampled from the distribution given i's sampled action and belief, $o^{t+1}_i \sim Pr(\Omega_i|a^t_i, b^t_{i,l})$. We utilize this sampled action and observation pair as the history, $h^t \xleftarrow{\cup} \langle a^t_i, o^{t+1}_i\rangle$. We may implement this procedure by entering as evidence i's action in the chance node, A^t_i, of the DBN mentioned previously and sampling from the inferred distribution over the chance node, O^{t+1}_i.

Finally, we note that in computing the distribution over the paths, solution to agent i's I-DID is needed as well ($Pr(a^t_i|OPT(m^t_{i,l}))$ term in Eq. 1). As we wish to avoid this, we assume a uniform distribution over i's actions. However, this may change the set of SE models. Specifically, this does not affect the set of behaviorally equivalent models, but a different set of models of j may now be observationally equivalent. Nevertheless, a uniform distribution minimizes any change as models that are now observationally equivalent would continue to remain so for any other distribution over i's actions. This is because given a model of j, a uniform distribution for i induces a distribution that includes the largest set of paths in its support.

V. Algorithm

We present the algorithm for partitioning the models in the model node of the I-DID at each time step according to ϵ-SE, in Fig. 6. The procedure, **ϵ-SubjectiveEquivalence** replaces the procedure, **BehaviorEq**, in the algorithm in Fig. 4. The procedure takes as input, the set of j's models, $\mathcal{M}_j$, 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 III and entering the model of j as evidence – this implements Eqs. 1 and 2.

We then pick a representative model at random, and using the cached distributions group models whose distributions exhibit a divergence less than ϵ from the distribution of the representative model. We iterate over 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-subjectively distinct, are returned.

VI. 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 time step t, there could be $|\mathcal{M}^0_j|(|A_j||\Omega_j|)^t$ models of the other agent j, where $|\mathcal{M}^0_j|$ is the number of models considered initially. Nested modeling further contributes to the complexity since solution of each model at level $l-1$ requires solving the lower level $l-2$ models, and so on recursively up to level 0. 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 $\mathcal{O}((N|\mathcal{M}|)^l)$

ϵ-**SUBJECTIVEEQUIVALENCE**(Model set $\mathcal{M}_j$, DID m_i, current time step tt, horizon T, ϵ) **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 (Proposition 1)

Clustering Phase

While $\mathcal{M}_j$ not empty
    Select a model, $m_j^{\hat{k}} \in \mathcal{M}_j$, at random as representative
    Initialize, $\mathcal{M}_j^{\hat{k}} \leftarrow \{m_j^{\hat{k}}\}$
    For each m_j^k **in** $\mathcal{M}_j$ **do**
        If $D_{KL}(P[\hat{k}]||P[k]) \leq \epsilon$
            $\mathcal{M}_j^{\hat{k}} \xleftarrow{\cup} m_j^k, \quad \mathcal{M}_j \xleftarrow{-} m_j^k$

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 model space using ϵ-SE. This function replaces **BehaviorEq()** in Fig. 4.

models. As mentioned in Proposition 3, ϵ-SE approximation reduces the number of models at each level to at most the size of the minimal set, $|\hat{\mathcal{M}}^t|$. In doing so, it solves $|\mathcal{M}_j^0|$ 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 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. Generally, $|\hat{\mathcal{M}}| \ll |\mathcal{M}|$, resulting in a substantial reduction in computation. Reducing the number of models in the model node also reduces the size of the state space, making the upper-level I-DID more memory efficient.

Given that lower-level models of other agent are solved exactly, we analyze the conditional error bound of this approach. [2] Trivially, if $\epsilon = 0$ there is no optimality error in the solution. If we limit the pruning of ϵ-SE 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. Our definition of SE provides us with a unique opportunity to bound the error for i. Observe that the expected value of the I-DID could be obtained as the expected reward of following each path weighted by the probability of that path. Let $\rho_{b_{i,l}}(H_T)$ be the vector of expected rewards for agent i given its belief when each path in H_T is followed. Here, T is the I-DID's horizon. The expected value for i is:

$$EV_i = Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T)$$

where $m_{j,l-1}$ is the model of j.

If the above model of j is pruned in the Mod node, let model $\hat{m}_{j,l-1}$ be the representative that replaces it. Then $\hat{b}_{i,l}$ is i's belief in which model $m_{j,l-1}$ is replaced with the representative. Expected value for i, $\hat{EV}_i$, is:

$$\hat{EV}_i = Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{\hat{b}_{i,l}}(H_T)$$

Then, the effective error bound is:

$$\begin{aligned}
\Delta \quad &= ||\hat{EV}_i - EV_i||_\infty = ||Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{\hat{b}_{i,l}}(H_T) \\
&\quad - Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T)||_\infty \\
&= ||Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{\hat{b}_{i,l}}(H_T) \\
&\quad - Pr(H_T|m_{i,l}, \hat{m}_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T) \\
&\quad + Pr(H_T|m_{i,l}, \hat{m}_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T) \\
&\quad - Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T)||_\infty \quad \text{(add zero)} \\
&\leq ||Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{\hat{b}_{i,l}}(H_T) \\
&\quad - Pr(H_T|m_{i,l}, \hat{m}_{j,l-1}) \cdot \rho_{\hat{b}_{i,l}}(H_T) \\
&\quad + Pr(H_T|m_{i,l}, \hat{m}_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T) \\
&\quad - Pr(H_T|m_{i,l}, m_{j,l-1}) \cdot \rho_{b_{i,l}}(H_T)||_\infty \; (|\rho_{\hat{b}_{i,l}}| \leq |\rho_{b_{i,l}}|) \\
&\leq ||\rho_{\hat{b}_{i,l}}(H_T) - \rho_{b_{i,l}}(H_T)||_\infty \cdot ||Pr(H_T|m_{i,l}, m_{j,l-1}) \\
&\qquad - Pr(H_T|m_{i,l}, \hat{m}_{j,l-1})||_1 \qquad \text{(Hölder's inequality)} \\
&\leq (R_i^{max} - R_i^{min})T \times 2\epsilon \qquad \text{(Pinsker's inequality)}
\end{aligned}$$

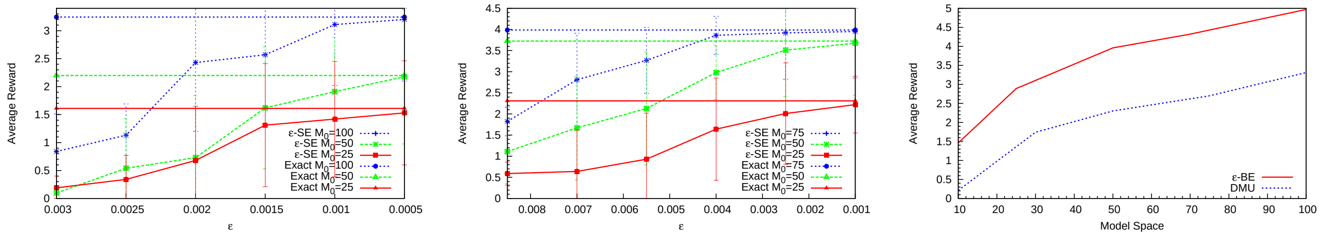
Matters become more complex when we additionally prune models in the subsequent model nodes as well. This is because rather than comparing over distributions given each history of i, we sample i's action-observation history. Hence, additional error incurs due to the sampling.

VII. EXPERIMENTAL EVALUATION

We implemented the approach in Figs. 4 and 6 utilizing Hugin API for DIDs and show results for the well-known two-agent *tiger problem* ($|S|$=2, $|A_i|$=$|A_j|$=3, $|\Omega_i|$=6, $|\Omega_j|$=3) [1], [9] and the multiagent version of the machine maintenance (MM) problem ($|S|$=3, $|A_i|$=$|A_j|$=4, $|\Omega_i|$=2, $|\Omega_j|$=2) [12]. We formulate level 1 I-DIDs of increasing time horizons for the problems and solve it approximately for varying ϵ. We show that, (i) the quality of the solution generated using our approach (ϵ-SE) improves as we reduce ϵ for given numbers of initial models of the other agent, M_0, and approaches that of the exact solution. This is indicative of the flexibility of the approach; (ii) in comparison to the previous approach of updating models discriminatively (DMU) [6], which is the current efficient technique, ϵ-SE is able to obtain larger rewards for an identical number of initial models. This indicates a more informed clustering and pruning using ϵ-SE in comparison to DMU, although it is less efficient in doing so.

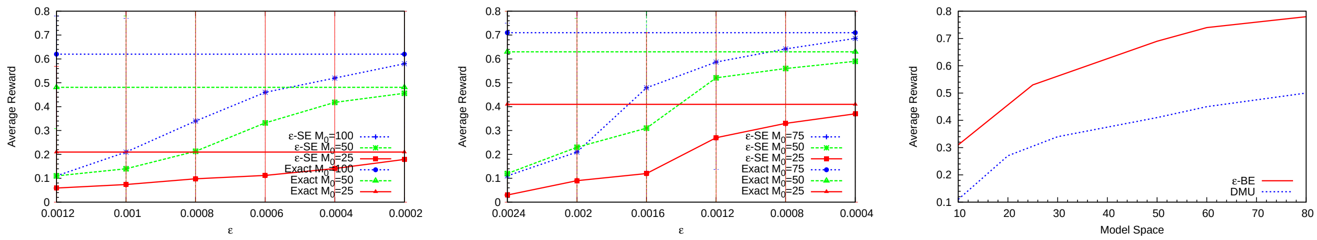
In Figs. 7 and $8(a, b)$, we show the average rewards gathered by executing the policies obtained from solving

[2] Doshi and Zeng [6] show that, in general, it is difficult to usefully bound the error if lower-level models are themselves solved approximately.



(a) (b) (c)

Figure 7. Performance profile obtained by solving a level 1 I-DID for the **multiagent tiger problem** using the ϵ-SE 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-SE and DMU in terms of the rewards obtained given identical numbers of models in the initial model node after clustering and pruning.



(a) (b) (c)

Figure 8. Performance profile for **multiagent MM problem** by solving level 1 I-DIDs approximately using ϵ-SE for (a) 3, and (b) 4 horizon. Reducing ϵ results in better quality solutions. (c) Significant increase in rewards obtained for ϵ-SE given identical numbers of retained models in the initial model node.

level 1 I-DIDs approximately within a simulation of each of the two problem domains. 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, across horizons and problem domains. Values of these policies increase as i considers greater numbers of models thereby improving it's chances of modeling j correctly. [3]

Next, we compare the performance of this approach with that of DMU. While both approaches cluster and prune models, DMU does so in the initial model node only, thereafter updating only those models which on update will be behaviorally distinct. Thus, we compare the average rewards obtained by the two approaches when an identical number of models remain in the initial model node after clustering and selection. This is done by varying ϵ in both approaches until the desired number of models are retained. In DMU, models whose beliefs are within ϵ of a representative are pruned. This allows comparison between clustering and selection techniques of the two methods. From Figs. 7 and 8(c), we observe that ϵ-SE results in better quality policies that obtain significantly higher average reward. This indicates that models pruned by DMU were more valuable than those pruned by ϵ-SE, thereby testifying to the more *informed* way in which we compare between models by directly gauging the impact on i's history. DMU's method of measuring simply the closeness of beliefs in models for clustering results in significant models being pruned. However, the trade off is the increased computational cost in calculating the distributions over future paths. To illustrate, ϵ-SE consumed an average of 23.7 secs in solving a 4 horizon I-DID with 25–100 initial models for the tiger problem and differing ϵ, on a Xeon 2GHz, 2GB RAM machine. This represents approximately a two-fold increase compared to DMU. For the MM problem, the approach incurred on average 38.1 secs exhibiting a three-fold increase in time taken compared to DMU to solve a horizon 4 I-DID with 25–100 initial models. On the other hand, while ϵ-SE continues to solve I-DIDs of 5 horizons, the exact approach runs out of memory.

VIII. Discussion

Our results demonstrate flexible solutions of I-DIDs by pruning models that are approximately SE. Defining SE by explicitly focusing on the impact that other agents' models have on the subject agent allows us to better identify model similarity. This translates into solutions of better quality given a limit on the number of models that could be held in memory. Consequently, other approaches would need more models to achieve comparable quality, which could translate into better efficiencies for our approach. However, we face the challenge of computing distributions over a number of paths that grow exponentially with horizon, which translates into increased time complexity. Although the approach is not yet viable as a scalable approximation technique, we are

[3]Note that the error bound of Section VI does not apply here because we prune models in subsequent time steps as well.

optimistic that the technique may be combined synergistically with DMU, and this will facilitate application to larger multiagent problem domains. Given the informed clustering and selection, this approach also serves as a benchmark for other techniques that seek to prune models.

Acknowledgment

This work was supported in part by grant #FA9550-08-1-0429 from AFOSR and in part by CAREER grant #IIS-0845036 from NSF. The authors would like to acknowledge the helpful comments of anonymous reviewers.

References

[1] P. Doshi, Y. Zeng, and Q. Chen, "Graphical models for interactive pomdps: Representations and solutions," *JAAMAS*, vol. 18, no. 3, pp. 376–416, 2009.

[2] J. A. Tatman and R. D. Shachter, "Dynamic programming and influence diagrams," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 2, pp. 365–379, 1990.

[3] D. Koller and B. Milch, "Multi-agent influence diagrams for representing and solving games," in *IJCAI*, 2001, pp. 1027–1034.

[4] K. Gal and A. Pfeffer, "Networks of influence diagrams: A formalism for representing agents' beliefs and decision-making processes," *JAIR*, vol. 33, pp. 109–147, 2008.

[5] J. Pineau, G. Gordon, and S. Thrun, "Anytime point-based value iteration for large pomdps," *JAIR*, vol. 27, pp. 335–380, 2006.

[6] P. Doshi and Y. Zeng, "Improved approximation of interactive dynamic influence diagrams using discriminative model updates," in *AAMAS*, 2009, pp. 907–914.

[7] D. Pynadath and S. Marsella, "Minimal mental models," in *AAAI*, Vancouver, Canada, 2007, pp. 1038–1044.

[8] B. Rathnas., P. Doshi, and P. J. Gmytrasiewicz, "Exact solutions to interactive pomdps using behavioral equivalence," in *AAMAS*, 2006, pp. 1025–1032.

[9] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella, "Taming decentralized pomdps : Towards efficient policy computation for multiagent settings," in *IJCAI*, 2003, pp. 705–711.

[10] P. Doshi and P. J. Gmytrasiewicz, "On the difficulty of achieving equilibrium in interactive pomdps," in *AI and Math Symposium*, Ft. Lauderdale, FL, 2006.

[11] S. Kullback and R. Leibler, "On information and sufficiency," *Annals of Math. Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[12] R. Smallwood and E. Sondik, "The optimal control of partially observable markov decision processes over a finite horizon," *OR* vol. 21, pp. 1071–1088, 1973.

Appendix

Proof of Proposition 1: $Pr(h_{T-t}|m^t_{i,l}, m^t_{j,l-1}) = Pr(h_{T-t-1}, a^t_i, o^{t+1}_i|m^t_{i,l}, m^t_{j,l-1}) = Pr(h_{T-t-1}|a^t_i, o^{t+1}_i, m^t_{i,l}, m^t_{j,l-1})Pr(a^t_i, o^{t+1}_i|m^t_{i,l}, m^t_{j,l-1})$ (using Bayes rule)

We focus on the first term next:
$Pr(h_{T-t-1}|a^t_i, o^{t+1}_i, m^t_{i,l}, m^t_{j,l-1}) = \sum_{a^t_j, o^{t+1}_j} Pr(h_{T-t-1}| a^t_i, o^{t+1}_i, m^t_{i,l}, a^t_j, o^{t+1}_j, m^t_{j,l-1})Pr(a^t_j, o^{t+1}_j|a^t_i, m^t_{i,l}, m^t_{j,l-1})$
$= Pr(h_{T-t-1}|m^{t+1}_{i,l}, m^{t+1}_{j,l-1})\ Pr(a^t_j, o^{t+1}_j|a^t_i, m^t_{i,l}, m^t_{j,l-1})$
In the above equation, the first term results due to an update of the models at time step t with actions and observations. This term is computed recursively. For the second term, j's level $l-1$ actions and observations are independent of i's observations.

We now focus on the term, $Pr(a^t_i, o^{t+1}_i|m^t_{i,l}, m^t_{j,l-1})$:
$Pr(a^t_i, o^{t+1}_i|m^t_{i,l}, m^t_{j,l-1}) = Pr(o^{t+1}_i|a^t_i, m^t_{i,l}, m^t_{j,l-1})$
$\times\ Pr(a^t_i|OPT(m^t_{i,l}))$ (i's action is conditionally independent of j given its model)
$= Pr(a^t_i|OPT(m^t_{i,l}))\sum_{a^t_j} Pr(o^{t+1}_i|a^t_i, a^t_j, m^t_{i,l}, m^t_{j,l-1})$
$\times\ Pr(a^t_j|OPT(m^t_{j,l-1}))$
$= Pr(a^t_i|OPT(m^t_{i,l}))\sum_{a^t_j} Pr(o^{t+1}_i|a^t_i, a^t_j, m^t_{i,l})$
$\times\ Pr(a^t_j|OPT(m^t_{j,l-1}))$ (i's observation is conditionally independent of j's model)
$= Pr(a^t_i|OPT(m^t_{i,l}))\ \sum_{a^t_j}\ Pr(a^t_j|OPT(m^t_{j,l-1}))\ Pr(o^{t+1}_i|a^t_i, a^t_j,\ b^t_{i,l})$ ($b^t_{i,l}$ is i's belief in model, $m^t_{i,l}$)
$= Pr(a^t_i|OPT(m^t_{i,l}))\sum_{a^t_j} Pr(a^t_j|OPT(m^t_{j,l-1}))$
$\times \sum_{s^{t+1}} Pr(o^{t+1}_i|s^{t+1}, a^t_i, a^t_j)\ Pr(s^{t+1}|a^t_i, a^t_j, b^t_{i,l})$
$= Pr(a^t_i|OPT(m^t_{i,l}))\sum_{a^t_j} Pr(a^t_j|OPT(m^t_{j,l-1}))\sum_{s^{t+1}} O_i(s^{t+1}, a^t_i, a^t_j, o^{t+1}_i)\sum_{s,m_j} T_i(s, a^t_i, a^t_j, s^{t+1})\ b^t_{i,l}(s, m_j)$
where O_i and T_i are i's observation and transition functions respectively, in the I-DID denoted by model, $m^t_{i,l}$. This proves Eq. 1 in Proposition 1.

Finally, we move to the term, $Pr(a^t_j, o^{t+1}_j|a^t_i, m^t_{i,l}, m^t_{j,l-1})$, to obtain Eq. 2:
$Pr(a^t_j, o^{t+1}_j|a^t_i, m^t_{i,l}, m^t_{j,l-1}) = Pr(o^{t+1}_j|a^t_j, a^t_i, m^t_{i,l}, m^t_{j,l-1})\ Pr(a^t_j|a^t_i, m^t_{i,l}, m^t_{j,l-1})$
$= Pr(o^{t+1}_j|a^t_j, a^t_i, m^t_{i,l}, m^t_{j,l-1})\ Pr(a^t_j|OPT(m^t_{j,l-1}))$ (j's action is conditionally independent of i given model)
$= Pr(a^t_j|OPT(m^t_{j,l-1}))\sum_{s^{t+1}} Pr(o^{t+1}_j|a^t_j, a^t_i, s^{t+1})$
$\times\ Pr(s^{t+1}|a^t_j, a^t_i, m^t_{i,l}, m^t_{j,l-1})$
$= Pr(a^t_j|OPT(m^t_{j,l-1}))\ \sum_{s^{t+1}}\ O_j(s^{t+1}, a^t_j, a^t_i, o^{t+1}_j)\sum_{s,m_j} Pr(s^{t+1}|a^t_j, a^t_i, s)b^t_{i,l}(s, m_j)$ ($b^t_{i,l}$ is i's bel. in $m^t_{i,l}$)
$= Pr(a^t_j|OPT(m^t_{j,l-1}))\ \sum_{s^{t+1}}\ O_j(s^{t+1}, a^t_j, a^t_i, o^{t+1}_j)\sum_{s,m_j} T_i(s, a^t_i, a^t_j, s^{t+1})b^t_{i,l}(s, m_j)$ (i's I-DID is used)
where O_j is j's observation function in model $m^t_{j,l-1}$, which is a part of i's I-DID. ■