Utilizing Partial Policies for Identifying Equivalence of Behavioral Models

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

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

Hua Mao  
Dept. of Computer Science  
Aalborg University  
DK-9220, Aalborg, Denmark  
huamao@cs.aau.dk

Muthukumaran Chandrasekaran  
Dept. of Computer Science  
University of Georgia  
Athens, GA 30602, USA  
mkran@uga.edu

Yinghui Pan  
Dept. of Automation  
Xiamen University  
Xiamen, China  
panyinghui@xmu.edu.cn

Jian Luo  
Dept. of Automation  
Xiamen University  
Xiamen, China  
jianluo@xmu.edu.cn

Abstract

We present a novel approach for identifying exact and approximate behavioral equivalence between models of agents. This is significant because both decision making and game play in multiagent settings must contend with behavioral models of other agents in order to predict their actions. One approach that reduces the complexity of the model space is to group models that are behaviorally equivalent. Identifying equivalence between models requires solving them and comparing entire policy trees. Because the trees grow exponentially with the horizon, our approach is to focus on partial policy trees for comparison and determining the distance between updated beliefs at the leaves of the trees. We propose a principled way to determine how much of the policy trees to consider, which trades off solution quality for efficiency. We investigate this approach in the context of the interactive dynamic influence diagram and evaluate its performance.

Introduction

Several areas of multiagent systems such as decision making and game playing benefit from modeling other agents sharing the environment, in order to predict their actions (Schadd, Bakkes, & Spronck 2007; Del Giudice, Gmytrasiewicz & Bryan 2009). If we do not constrain the possible behaviors of others, the general space of these models is very large. In this context, a promising approach is to group together behaviorally equivalent (BE) models (Dekel, Fudenberg, & Morris 2006; Pynadath & Marsella 2007) in order to reduce the number of candidate models. Models that are BE prescribe identical behavior, and these may be grouped because it is the prescriptive aspects of the models and not the descriptive that matter to the decision maker. Essentially, we cluster BE models of other agents and select a representative model for each cluster.

One particular decision-making framework for which BE has received much attention is the interactive dynamic influence diagram (I-DID) (Doshi, Zeng, & Chen 2009). I-DIDs 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 of unknown types. They generalize DIDs to multiagent settings. Expectedly, solving I-DIDs tends to be computationally very complex. This is because the state space in I-DIDs includes the models of other agents in addition to the traditional physical states. As the agents act, observe, and update beliefs, I-DIDs must track the evolution of the models over time. 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.

Previously, I-DID solutions mainly exploit BE to reduce the dimensionality of the state space (Doshi, Zeng, & Chen 2009; Doshi & Zeng 2009). For example, Doshi and Zeng (2009) minimize the model space by updating only those models that lead to behaviorally distinct models at the next time step. While this approach speeds up solutions of I-DID considerably, it does not scale desirably to large horizons. This is because: (a) models are compared for BE using their solutions which are typically policy trees. As the horizon increases, the size of the policy tree increases exponentially; (b) the condition for BE is strict: entire policy trees of two models must match exactly.

Progress could be made by efficiently determining if two models are BE and by grouping models that are approximately BE. We expect the latter to result in lesser numbers of classes each containing more models, thereby producing less representatives at the cost of prediction error. In this paper, we seek to address both these issues. We determine BE between two models by comparing their partial policy trees and the updated beliefs at the leaves of the policy trees. This leads to significant savings in memory as we do not store entire policy trees. Furthermore, we may group models whose partial policy trees are identical but the updated beliefs diverge by small amounts. This defines an approximate measure of BE that could group more models together.

We use the insight that the divergence between the updated beliefs at the leaves of the two policy trees will not be greater than the divergence between the initial beliefs. Boyen and Koller (1998) show that the change in the divergence is a contraction controlled by a rate parameter, $\gamma$. We show how we may calculate $\gamma$ in our context and use...
it to obtain the depth of the partial policy tree to use for a
given approximate measure of BE. We bound the prediction
error due to grouping models that could be approximately
BE. Finally, we evaluate the empirical performance of this
approach in the context of multiple problem domains, and
demonstrate that it allows us to scale the solution of I-DIDs
significantly more than previous techniques.

**Background: Interactive DID and BE**

We briefly describe interactive influence diagrams (I-IDs)
for two-agent interactions followed by their extensions to
dynamic settings, I-DIDs, and refer the reader to (Doshi,
Zeng, & Chen 2009) for more details.

**Syntax**

I-IDs include a new type of node called the model node
(hexagonal shaded node, $M_{j,t-1}$, in Fig. 1(a)). The prob-
ability 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 have a
chance node, $A_j$, that represents the distribution over the
other agent’s actions, and a dashed link, called a policy link.

The model node contains as its values the candidate computa-
tional models ascribed by $i$ to the other agent. We de-
note the set of these models by $\mathcal{M}_{j,t-1}$. A model in the
model node may itself be an I-ID or ID, and the recur-
sion terminates when a model is an ID or a simple proba-
bility distribution over the actions. Formally, we denote a
model of $j$ as, $m_{j,t-1} = (b_{j,t-1}, \theta_j)$, where $b_{j,t-1}$ is the
level $l-1$ belief, and $\theta_j$ 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 trans-
formed 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 $\text{Mod}[M_j]$.
The distribution over $\text{Mod}[M_j]$ 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 linked together using a policy link,
for each other agent.

![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 ($M_{j,t-1}$) and the dashed arrow is the policy link. (b) Representing the model node and policy link using chance nodes and dependencies between them. The decision nodes of the lower-level I-IDs or IDs ($m_{j,t-1}^1, m_{j,t-1}^2$) are mapped to the corresponding chance nodes ($A_j^1, A_j^2$).](image1)

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 step time $t + 1$ will have up
to $|\mathcal{M}_{j,t-1}^1| |\Omega_j|$ models. Here, $|\mathcal{M}_{j,t-1}^1|$ is the number
of models at time step $t$, $|A_j^1|$ and $|\Omega_j|$ are the largest
spaces of actions and observations respectively, among all
the models. The CPT of $\text{Mod}[M_{j,t-1}^1]$ encodes the func-
tion, $\tau(b_{j,t-1}^l, a_{j,t-1}^l, O_{j,t-l-1}^l)$ which is 1 if the belief $b_{j,t-1}^l$
in the model $m_{j,t-l-1}^1$ using the action $a_{j,t-1}^l$ and observation $O_{j,t-l-1}^l$
updates to $b_{j,t-l-1}^{l+1}$ in a model $m_{j,t-l-1}^{l+1}$; otherwise it is 0. Sec-
ond, we compute the new distribution over the updated mod-
els, 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

![Figure 2: A generic two time-slice level $l$ I-DID for agent $i$. The dotted model update link denotes the update of $j$’s models and of the distribution over the models, over time.](image2)

![Figure 3: Semantics of the model update link. Notice the growth in the number of models in the model node at $t + 1$ shown in bold.](image3)
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.

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 BE (Pynadath & Marsella 2007) – 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 behavior of the other agent only; thus we need not distinguish between BE models. Let \textbf{PruneBehavioralEq} (\(M_{j,l-1}\)) be the procedure that prunes BE models from \(M_{j,l-1}\) returning the representative models.

Solving an I-DID (and I-ID) proceeds in a bottom-up manner, and is implemented recursively (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, and 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 BE 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 such as Hugin API.

Approximating Behavioral Equivalence

Although BE represents an effective exact criteria to group models, identifying BE models requires us to compare the entire solutions of models – all paths in the policy trees which grow exponentially over time. This is further complicated by the number of candidate models of the other agents in the model node growing exponentially over time. In order to scale BE to large horizons, we seek to (a) reduce the complexity of identifying BE by comparing partial policy trees; and (b) group together more models that could be approximately BE. We do this by grouping models that have identical partial policy trees of depth \(d\) and whose updated beliefs at the leaves of the policy trees do not diverge much.

Revisiting BE

For the sake of clarity, we assume that the models of the other agent \(j\) have identical frames (possibly different from \(i\’s\)) and differ only in their beliefs. We focus on the general setting where a model, \(M_{j,l-1}\), is itself a DID or an I-DID, in which case its solution could be represented as a policy

\begin{algorithm}
\caption{I-DID Exact (level \(l \geq 1\) I-DID or level \(0\) DID, horizon \(T\))}
\begin{algorithmic}[1]
\Procedure{Expansion Phase}{(level \(l \geq 1\) I-DID or level \(0\) DID, horizon \(T\))}
\end{algorithmic}
\end{algorithm}
Two models of agent $\gamma \in \mathcal{T}$ and the divergence of pairs $\min_{\epsilon} d_{\epsilon}$ may not be derived. If two distributions $b_{\gamma,\epsilon} \in B_{\gamma,\epsilon} \cap B_{\epsilon,\gamma}$ are the same, then beliefs at the leaves of the two ordered policy trees do not diverge: $D_{KL}(b_{\gamma,\epsilon}^d || b_{\epsilon,\gamma}^d) = 0 \quad \forall k = 1 \ldots |\Omega_j|^d$.

Proposition 1 holds because of the well-known fact that beliefs updated using an action-observation sequence in a partially observable stochastic process is a sufficient statistic for the history. Consequently, future behavior is predicated only on the beliefs. Therefore, pairs of models that satisfy the two conditions in Prop. 1 for some $d$ will necessarily conform to Def. 1. Furthermore, Prop. 1 is not particularly sensitive to the measure of divergence between distributions that we utilize. While it holds because $D_{KL}(b_{\gamma,\epsilon}^d || b_{\epsilon,\gamma}^d) = 0$ if and only if the two distributions are equal, the same is also true for, say, the $L_1$ distance. However, KL divergence has some desirable properties lacked by other norms, which we will exploit later.

Notice that the redefinition produces the same grouping of BE models as previously for the case $d = T - 1$ because it collapses into Def. 1. For the case of $d < T - 1$, it may group less models in a BE class because belief sets that do diverge could still result in the same set of policy trees. Hence, it may lead to more BE classes than needed.

The advantage of Prop. 1 is that we may elegantly generalize it to the notion of approximate BE:

**Definition 2 ($\epsilon, d$-BE)** Two models of agent $j$, $m_{j,1}$ and $\hat{m}_{j,1}$, are $(\epsilon,d)$-BE, $\epsilon \geq 0$, $d \leq T - 1$, if their depth-$d$ policy trees are identical, $\pi_{m_{j,1}}^d = \pi_{\hat{m}_{j,1}}^d$, and if $d < T - 1$ then beliefs at the leaves of the two ordered policy trees diverge by at most $\epsilon$: $\max_{k = 1 \ldots |\Omega_j|^d} D_{KL}(b_{\gamma,\epsilon}^d || b_{\epsilon,\gamma}^d) \leq \epsilon$.

Intuitively, two models are $(\epsilon, d)$-BE if their solutions share an identical depth-$d$ tree and the divergence of pairs of the ordered beliefs at the leaves of the depth-$d$ tree is not larger than $\epsilon$. As $\epsilon$ approaches zero, $(\epsilon, d)$-BE converges to Prop. 1. While the definition above is parameterized by the depth $d$ of the policy trees as well, we show in the next section that $d$ may be determined given some $\epsilon$.

**Depth of the Partial Policy**

Definition 2 introduces a measure of approximate BE between two models. It is parameterized by both the amount of approximation, $\epsilon$, and the partialness of the comparison, $d$. However, we show that the depth $d$ may be uniquely determined by the amount of approximation that is allowed between the equivalence of two models. We begin by reviewing an important result for a Markov stochastic process.

While it is well known that a stochastic transition never increases the KL divergence between two distributions over the same state space in a Markov stochastic process (Cover & Thomas 1991), Boyen and Koller (1998) show that the KL divergence between the distributions contracts at a geometric rate given a stochastic transition, and the rate of contraction is based on a mixing rate, $\gamma$.

In our context, we may apply this result to bound the divergence between the beliefs of two models updated using an action-observation sequence:

$$D_{KL}(b_{\gamma,\epsilon}^1 || b_{\epsilon,\gamma}^1) \leq (1 - \gamma_{F_{a,o}}) D_{KL}(b_{\gamma,\epsilon}^0 || b_{\epsilon,\gamma}^0)$$

where $F_{a,o}(s'|s)$ is the “stochastic transition” from state $s$ to $s'$ obtained by multiplying the state transition probability due to action, $a$, and the likelihood of observation, $o$, for $j$. $\gamma_{F_{a,o}}$ is the minimum probability mass on some state due to the transition, and is called the minimal mixing rate:

$$\gamma_{F_{a,o}} = \min_{s_{m_{j,1}}} \sum_{s' \in S} \min \{ F_{a,o}(s'|s_{m_{j,1}}), F_{a,o}(s'|s_{\hat{m}_{j,1}}) \}$$

Next, we may extend Eq. 1 over an action-observation sequence of length $d$ that corresponds to a path in a depth-$d$ policy tree:

$$D_{KL}(b_{\gamma,\epsilon}^d || b_{\epsilon,\gamma}^d) \leq (1 - \gamma_F)^d D_{KL}(b_{\gamma,\epsilon}^0 || b_{\epsilon,\gamma}^0)$$

Here, because a path may involve different action and observation sequences, $\gamma_F = \min \{ \gamma_{F_{a,o}} | a \in A_j, o \in \Omega_j \}$.

The definition of approximate BE in the previous section (Def. 2) limits the maximum divergence between any pair of beliefs at the leaves of the partial policy trees to at most $\epsilon$. Because Eq. 2 bounds this divergence as well, we may equate the bound to $\epsilon$ and obtain the following:

$$(1 - \gamma_F)^d D_{KL}(b_{\gamma,\epsilon}^0 || b_{\epsilon,\gamma}^0) = \epsilon$$

In the above equation, the only unknown is $d$ because $\gamma_F$ may be obtained as shown previously and $b_{\gamma,\epsilon}^0$, $b_{\epsilon,\gamma}^0$ are the given initial beliefs in the models. Therefore, we may derive $d$ for a given value of $\epsilon$ as:

$$d = \min \left\{ T - 1, \max \left\{ 0, \left[ \frac{\ln \gamma_F}{\ln (1 - \gamma_F)} \right] \right\} \right\}$$

where $\gamma_F \in (0, 1)$ and $\epsilon > 0$. Eq. 3 gives the smallest depth that we could use for comparing the policy trees. In general, as $\epsilon$ increases, $d$ reduces for a model pair until it becomes zero when we compare just the initial beliefs in the models.

We note that the minimal mixing rate depending on the function, $F_{a,o}$, may also assume two extreme values: $\gamma_F = 1$ and $\gamma_F = 0$. The former case implies that the updated beliefs have all probability mass in the same state, and the KL divergence of these distributions is zero after a transition. Hence, we set $d = 1$. For the latter case, there is at least one pair of states for which the updated beliefs do not agree at all (one assigns zero mass). For this null mixing rate, the KL divergence may not contract and $d$ may not be derived. Thus, we may arbitrarily select $d \leq T - 1$.

**Computational Savings and Error Bound**

Given that we may determine $d$ using Eq. 3, the complexity of identifying whether a pair of models are approximately BE is dominated by the complexity of comparing two depth-$d$ trees. This is proportional to the number of comparisons made as we traverse the policy trees. As there are a maximum of $|\Omega_j|^d$ leaf nodes in a depth-$d$ tree, the following proposition gives the complexity of identifying BE classes in the model node of agent $i$’s I-DID at some time step.
Proposition 2 (Complexity of BE) The asymptotic complexity of the procedure for identifying all models that are $\epsilon$-BE is $O(|M_j|^{d-1}|\Omega_j|^d)$ where $|M_j|_{j-1}$ is the number of models in the model node.

While the time complexity of comparing two partial policy trees is given by Prop. 2 (set $|M_j|_{j-1}=2$), we maintain at most $2(|\Omega_j|^d)$ paths ($d \leq T - 1$) at each time step for each pair of models that are being compared, with each path occupying space proportional to $d$. This precludes storing entire policy trees containing $(|\Omega_j|^T)^{-1}$ possible paths, leading to significant savings in memory when $d \ll T$.

We analyze the error in the value of $j$’s predicted behavior. If $\epsilon = 0$, grouped models are exactly BE and there is no error. With increasing values of $\epsilon$ (resulting in small $d$ values), a behaviorally distinct model, $m_{j,l-1}$, may be erroneously grouped with the model, $m_{j,l-1}$. Let $m_{j,l-1}$ be the model associated with $m_{j,l-1}$, resulting in the worst error. Let $\alpha^T$ and $\alpha^T$ be the exact entire policy trees obtained by solving the two models, respectively. Then, the error is: 

$$\rho = |T - T| - |T - T|.$$  

Because the depth-$d$ policy trees of the two models are identical (Def. 2), the error becomes:

$$\rho = \alpha^T - \alpha^T.$$ 

Here, $R_j^{max}$ and $R_j^{min}$ are the maximum and minimum rewards of $j$, respectively. Of course, this error is tempered by the probability that agent $i$ assigns to the model, $m_{j,l-1}$, in the model node at time step, $d$.

**Experimental Results**

We implemented our approach of determining $\epsilon$-BE between models and use it to group models into a class. This is followed by retaining the representative for each class while pruning others, analogously to using exact BE. This procedure now implements PruneBehaviorEq (line 6) in Fig. 4.

Because our approach is the first to formalize an approximation of BE (to the best of our knowledge), we compare it with the previous most efficient algorithm that exploits exact BE while solving I-DIDs. This technique (Doshi & Zeng 2009) groups BE models using their entire policy trees and updates only those models that will be behaviorally distinct from existing ones; we label it as DMU.

We evaluate both using two standard problem domains and a scalable multiagent testbed with practical implications: the two-agent tiger problem ($|S|=2$, $|A|=|A|=3$, $|\Omega|=6$, $|\Omega|=3$) (Gmytrasiewicz & Doshi 2005), the multiagent version of the concert problem ($|S|=2$, $|A|=|A|=3$, $|\Omega|=4$, $|\Omega|=2$)\(^1\), and a much larger domain: the two-agent unmanned aerial vehicle (UAV) problem ($|S|=25$, $|A|=|A|=5$, $|\Omega|=|\Omega|=5$) (Doshi & Sonu 2010).

We report on the performance of both techniques ($\epsilon$-BE and DMU) when used for solving level 1-I-DIDs of increasing horizon in the context of the above three domains. We show that the quality of the solution generated by $\epsilon$-BE converges to that of the exact DMU as $\epsilon$ decreases (with the corresponding increase in $d$). However, the multiaagent tiger problem exhibits a minimal mixing rate of zero, due to which the partial depth, $d$, is selected arbitrarily: we select increasing $d$ as $\epsilon$ reduces. In Fig. 5(a), we show the average rewards gathered by simulating the solutions obtained for decreasing $\epsilon$ for each of the three problem domains. We used a horizon of 10 for the small domains, and 6 for the UAV. Each data point is the average of 500 runs where the true model of $j$ is sampled according to $i$’s initial belief. For a given number of initial models, $M_i$, the solutions improve and converge toward the exact (DMU) as $\epsilon$ reduces. While the derived partial depths varied from 0 up to the horizon minus 1 for extremely small $\epsilon$, we point out that the solutions converge to the exact for $d < T - 1$, including the tiger problem (at $d=3$) despite the zero mixing rate. Fig. 5(b) shows the best solution possible on average for a given time allocation. Notice that $\epsilon$-BE consistently produces better quality solution than DMU. This is because it solves for a longer horizon than DMU in the same time. Finally, Fig. 5(c) confirms our intuition that $\epsilon$-BE leads to significantly less model classes for large $\epsilon$ (small $d$), although more than DMU for $\epsilon = 0$. Importantly, comparing partial policy trees is sufficient to obtain the same model space as in the exact case, which is responsible for the early convergence to the exact reward we observed in Fig. 5(a).

<table>
<thead>
<tr>
<th>Level</th>
<th>T</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU</td>
<td>10</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>14.5</td>
</tr>
</tbody>
</table>

Table 1: $\epsilon$-BE shows scalability to a large horizon.

In Table 1, we compare different techniques based on the time each takes to solve problems of increasing horizon. We additionally include a heuristic approach (Zeng, Chen, & Doshi 2011), labeled TopK, that samples $K$ paths from a policy tree that are approximately most likely to occur, and uses just these paths to compute for equivalence. $\epsilon$-BE demonstrates significant scalability over DMU, solving for much longer horizons than exactly possible. It shows significant run time speed up over TopK as well, which needs to maintain complete paths that grow long. $\epsilon$ and $K$ were varied to get the same reward as DMU if appropriate, otherwise until the model space stabilized.

---

\(^1\)We adapt the single-agent concert problem from the POMDP repository: http://www.cs.brown.edu/research/ai/pomdpl/.
In the face of an unconstrained model space, BE provides a way to compact it. We showed how we may utilize partial solutions of models to determine approximate BE and applied it to significantly scale solutions of I-DIDs. Our insight is that comparing partial solutions of models is likely sufficient for grouping models similarly to using exact BE, as our experiments indicate. While we use a principled technique to determine the partialness given the approximation measure, not all problem domains may allow this.

**Conclusion**

In the face of an unconstrained model space, BE provides a way to compact it. We showed how we may utilize partial solutions of models to determine approximate BE and applied it to significantly scale solutions of I-DIDs. Our insight is that comparing partial solutions of models is likely sufficient for grouping models similarly to using exact BE, as our experiments indicate. While we use a principled technique to determine the partialness given the approximation measure, not all problem domains may allow this.

**Acknowledgments**

Yifeng Zeng acknowledges support from the Obel Family Foundation (Denmark), NSFC (#60974089 and #60975052). Prashant Doshi acknowledges support from an NSF CAREER grant (#IIS-0845036).

**References**


Zeng, Y.; Chen, Y; and Doshi, P. 2011. Approximating behavioral equivalence of models using Top-K policy paths (extended abstract). In *AAMAS*.