Time-Critical Decision Making in Interactive Dynamic Influence Diagram

Yifeng Zeng
Dept. of Computer Science
Aalborg University
Denmark
yfzeng@cs.aau.dk

Yanping Xiang
School of Computer Science and Engineering
University of Electronic Science and Technology of China
P. R. China
yanping_xiang@yahoo.com.cn

***Abstract*—Time-critical dynamic decision making is a quite challenging task in many real-world applications. It requires to play a trade-off between solution optimality and computational tractability. It is especially true for multiagent settings under uncertainty. In this paper, we model time-critical dynamic decision problem using the representation of interactive dynamic influence diagram (I-DID). We formalize I-DID by providing time-index to nodes within the model. This results in a model that has the ability to represent space-temporal abstraction. In addition, we propose a new method for selecting the abstract model without arbitrarily compromising solution optimality. We evaluate the performance of our method in two benchmark settings and provide results in support.**

***Keywords*-Interactive Dynamic Influence Diagram; Time-Critical Decision Making; Model Abstraction**

I. Introduction

Many real-world applications demand actions to be taken in a time pressured situations. A decision maker is expected to take a proper amount of time in the modeling and solution so that he/she is allowed to have sufficient time for really executing the actions. This is a trade-off between model quality and computational tractability. A complete model may provide exact solutions while it consumes a large amount of time for compilation and execution.

Much research has been seen, mainly on a single-agent setting, about addressing time-critical dynamic decision problems [1], [2], [3]. In particular, Xiang and Poh [2], [4] proposed a formal representation of time-critical dynamic influence diagrams that provide explicit support for modeling temporal processes and dealing with time-critical situations. Their work is applicable in a single-agent decision domain. The issue becomes more complicated when a multiagent setting is considered since multiple agents may interact with each other over time. It involves the complicated modeling process and solutions. For example, a team of rescue agents expect to take a fast collaboration in a natural disaster while their decisions shall be made with the consideration of all other agents in the team.

In this paper, we utilize the language of *interactive dynamic influence diagrams* (I-DIDs) [5] to study time-critical dynamic decision making in multiagent settings. I-DID provides an efficient representation for modeling sequential multiagent decision making in an uncertain environment. It stands on the viewpoint of an individual agent and explicitly models other agents into the subject agent's state space. In this way, I-DID generalizes dynamic influence diagrams to multiagent settings and resorts to many standard solutions of probabilistic graphical models [6].

One important aspect of time-critical decision analysis is the framing and formulation of the decision problem, which requires the modeling of temporal process in an explicit way. Following the same vein as time-critical dynamic influence diagrams [2], we further formalize I-DIDs into time-critical I-DIDs by providing time index to each node in the model. This offers a possibility to represent temporal relations of the underlying random variables. We propose two forms of time-critical I-DIDs: the *condensed* form and the *deployed* form. The condensed form provides a static model of time-critical I-DIDs and is transformed to the deployed form in a dynamic process. The deployed form is the final time-critical I-DID model that will be compiled and solved for seeking optimal time-critical policies.

As expansion in I-DID, the transformation from the condensed form to the deployed form is a complicated process and may result in a computationally intractable time-critical I-DID model. We need to find a proper way to specify the condensed form so that the deployed form becomes more abstract; meanwhile, the reduction shall not have a serious impact on the model solution. In doing so, we may try all possible condensed forms and select the one by transforming that would approximate the model solution within a certain accuracy. This greedy selection has to consume much computation,which may not be allowed in a time pressured situation. We propose an entropy-based method to select a condensed form. The new method may further reduce the solution error while it is computationally cheap. We formalize the selection strategy and experimentally evaluate the performance of our method. We show the approach may elicit the condensed form efficiently and really strengthen the utilization of time-critical I-DIDs in dynamic decision making.

The rest of this paper is organized as follows. In Section II, we review necessary background knowledge on I-DIDs. In Section III, we propose both the condensed and deployed forms of time-critical I-DIDs. More importantly, in Section IV, we implement the entropy-based method for selecting a proper condensed form. We conduct the experiment and show positive results on two well-studied domains in Section V. Finally, we discuss relevant works and conclude the paper with remarks.

II. Background: Interactive DIDs

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 [5] for more details.

A. Syntax

In addition to the usual chance, decision, and utility nodes, I-IDs include a new type of node called the *model node* (hexagonal node, $M_{j,l-1}$, in Fig. $1(a)$). 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 other agent's actions, and a dashed link, called a *policy link*.

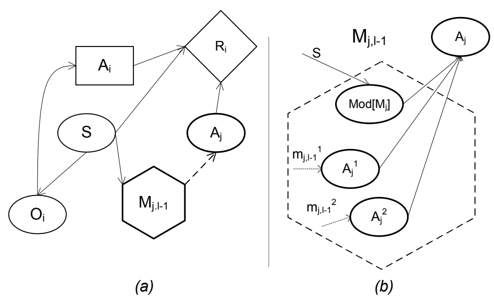


Figure 1. (a) A generic level $l > 0$ I-ID for agent i situated with one other agent j. The hexagon is the model node ($M_{j,l-1}$) and the dashed arrow is the policy link. (b) Representing the model node and policy link using chance nodes and dependencies between them. The decision nodes of the lower-level I-IDs or IDs ($m^1_{j,l-1}$, $m^2_{j,l-1}$) are mapped to the corresponding chance nodes (A^1_j, A^2_j), which is indicated by the dotted arrows.

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 $Pr(a_j \in A^1_j) = \frac{1}{|OPT|}$ 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^1_j, A^2_j) depending on the value of $Mod[M_j]$. In other words, when $Mod[M_j]$ has the value $m^1_{j,l-1}$, the chance node A_j assumes the distribution of the node A^1_j, and A_j assumes the distribution of A^2_j when $Mod[M_j]$ has the value $m^2_{j,l-1}$. 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.

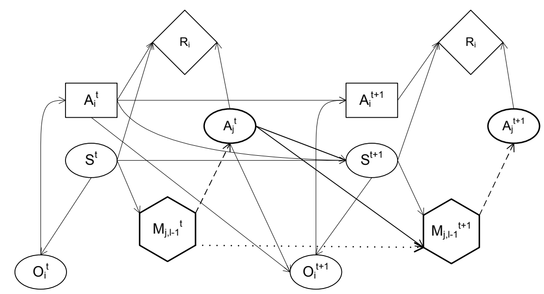


Figure 2. A generic two time-slice level l I-DID for agent i. Notice the dotted model update link that denotes the update of the models of j and of the distribution over the models, over time.

I-DIDs extend I-IDs to allow sequential decision making over several time steps. We depict a general two time-slice I-DID in 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.

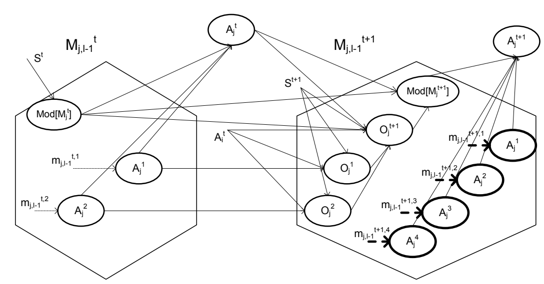


Figure 3. The semantics of the model update link. Notice the growth in the number of models in the model node at $t+1$ in bold.

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 $|\mathcal{M}^t_{j,l-1}||A_j||\Omega_j|$ models. Here, $|\mathcal{M}^t_{j,l-1}|$ 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 $Mod[M^{t+1}_{j,l-1}]$ encodes the function, $\tau(b^t_{j,l-1}, a^t_j, o^{t+1}_j, b^{t+1}_{j,l-1})$ which is 1 if the belief $b^t_{j,l-1}$ in the model $m^t_{j,l-1}$ using the action a^t_j and observation o^{t+1}_j updates to $b^{t+1}_{j,l-1}$ in a model $m^{t+1}_{j,l-1}$; 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 in Fig. 3, transforming it into a flat DID.

B. Solution

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. $SE(b^t_j, a_j, o_j)$ is an abbreviation for the belief update. 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). We point out that 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)
<u>Expansion Phase</u>
1. **For** t **from** 0 **to** $T-1$ **do**
2. **If** $l \geq 1$ **then**
   Populate $M^{t+1}_{j,l-1}$
3. **For each** m^t_j **in** $\mathcal{M}^t_{j,l-1}$ **do**
4. Recursively call algorithm with the $l-1$ I-DID (or DID) that represents m^t_j and the horizon, $T-t$
5. Map the decision node of the solved I-DID (or DID) $OPT(m^t_j)$, to the chance node A^t_j
6. **For each** a_j **in** $OPT(m^t_j)$ **do**
7. **For each** o_j **in** O_j (part of m^t_j) **do**
8. Update j's belief, $b^{t+1}_j \leftarrow SE(b^t_j, a_j, o_j)$
9. $m^{t+1}_j \leftarrow$ New I-DID (or DID) with b^{t+1}_j as init. belief
10. $\mathcal{M}^{t+1}_{j,l-1} \overset{\cup}{\leftarrow} \{m^{t+1}_j\}$
11. Add the model node, $M^{t+1}_{j,l-1}$, and the model update link between $M^t_{j,l-1}$ and $M^{t+1}_{j,l-1}$
12. Add the chance, decision, and utility nodes for $t+1$ time slice and the dependency links between them
13. Establish the CPTs for each chance node and utility node

<u>Solution Phase</u>
14. **If** $l \geq 1$ **then**
15. Represent the model nodes and the model update link as in Fig. 3 to obtain the DID
16. Apply the standard look-ahead and backup method (or other approaches)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.

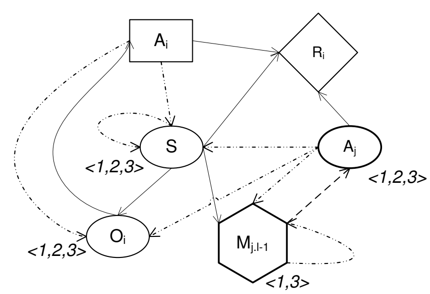


Figure 5. Time-critical I-IDs (the condensed form of time-critical I-DIDs). Nodes are indexed with a time sequence while long dashed arcs denote temporal relations.

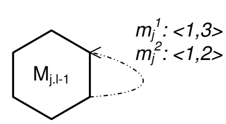


Figure 6. Time-critical model node in which two models, m^1_j and m^2_j, have different sub-time sequences $< 1, 3 >$ and $< 1, 2 >$ respectively.

III. TIME-CRITICAL I-DIDS

Time-critical I-DIDs extend I-DIDs to time-critical decision making analogously to how time-critical dynamic influence diagrams extend DIDs. We start with the proposal of the condensed form of time-critical I-DIDs.

A. The Condensed Form

The condensed form of time-critical I-DIDs serves as a static model of time-critical multiagent sequential decision making. We may formalize I-IDs to be the condensed form of time-critical I-DIDs. We show an example of time-critical I-IDs in Fig. 5.

Additional arcs (*long dashed* ones) are called temporal arcs in time-critical I-IDs and represent both probabilistic and temporal (time-lag) relations between nodes. We also index nodes with a time sequence (numbers in a angle bracket) and the set of time indices may be different from

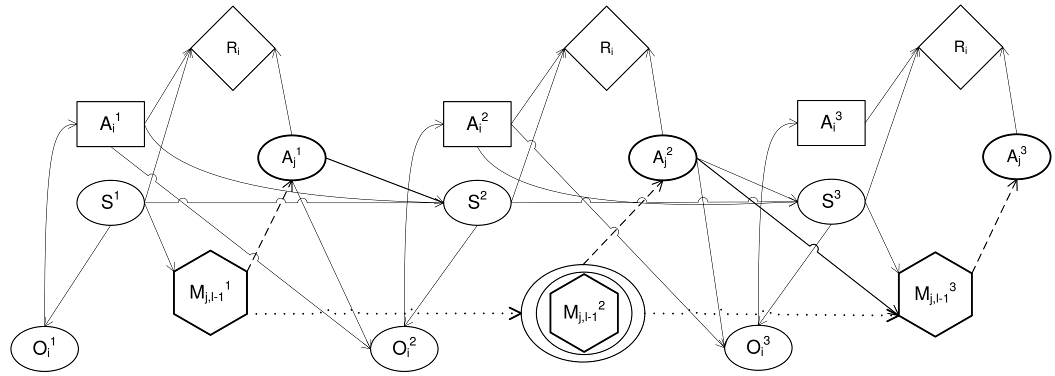


Figure 7. Time-critical I-DIDs (the deployed form) resulted from the transformation of the condensed form in Fig. 5. Nodes with a double circle are deterministic nodes.

one node to another. The longest sequence is called a master time sequence (mts) while others shall be a sub-time sequence (sts) that is a subset of the master time sequence. To avoid cumbersome denotations, we may not index nodes whose time sequence is equivalent to the master time sequence. For example, the master sequence of time-critical I-ID in Fig. 5 is $< 1, 2, 3 >$ indexed to chance nodes S, A_j and O_i. Decision node A_i and utility node R_i share the master time sequence and the model node $M_{j,l-1}$ has the sub-time sequence $< 1, 3 >$. The time sequence will be later utilized to expand time-critical I-IDs into time-critical I-DIDs.

Recall that the model node contains all candidate models of other agents. Those models may be abstracted in a different way - they may have different condensed forms. This requires to index each model with a unique time sequence in the model node. Consequently, we must attach a list of time sequences to the model node and each model is associated with a time sequence. Formally, let $m_j^{k,sts}$ be the condensed form of model m_j^k where m_j^k is indexed by a sub-time sequence sts. Assume that agent j has two candidate models, we show one example of indexed model node in Fig. 6. In this case, sts for model m_j^1 is equal to the sub-time sequence $< 1, 3 >$ while it is $< 1, 2 >$ for m_j^2. We may also index the model node using a single time sequence if all models share the same time sequence. This is exactly the case for the model node in Fig. 5 where all models have the sub-time sequence $< 1, 3 >$.

We notice that the candidate model indexed by a sub-time sequence has a shorter time horizon than the models with the master time sequence. In other words, the subject agent i is envisioned to suppose the bounded rationality of other agent j. Agent j may take actions of fewer time steps and play an intervention only at the indexed times. For instance, in Fig. 6, m_j^1 is the model of 2 time horizons, and the solutions are j's actions only at times $t = 1$ and $t = 3$.

B. The Deployed Form

The deployed form of time-critical I-DID is the expansion of time-critical I-ID (the condensed form of time-critical I-DID). We transform the time-critical I-ID (in Fig. 5) to the deployed form in Fig. 7. The transformation expands the time-critical I-IDs similarly to the expansion phase for I-DIDs in Fig. 4. However,it must rely on various time sequences indexed to relevant nodes. We repeat a normal node (except the model node) only if its time sequence is equivalent to the master time sequence; otherwise, the node will be casted into a deterministic node (which is deterministically dependent on its parent nodes) for the time step where the time is not indexed in the time sequence.

For the model node, we update the model only at the time step if the time is indexed in the time sequence to the model inside the model node. Otherwise, we retain all models from the previous time step and do not perform any model update - we also mark the model node using the type of deterministic nodes. Recall that the model with a sub-time sequence has a shorter time horizon. There is no solutions (actions performed by other agents) from the model at a particular time step which is not indexed in the time sequence. For facilitating the CPT setting of action node A_j (in Fig. 1), we assume a uniform distribution of actions from the model, e.g. assigning the probability $\frac{1}{|A_j|}$ to the columns corresponding to the model.

Consequently, we need to modify lines 4-10 in Fig. 4 and may expand time-critical I-IDs in a different way. Lines 4-10 are only followed if the current time step t is equivalent to the time index in the time sequence attached to model m_j^t. Otherwise, we do not need to update m_j^t, but assign a uniform distribution $\frac{1}{|A_j|}$ to chance node A_j^t for columns of model m_j^t.

We expand the time-critical I-ID of Fig. 5 and get the deployed form of time-critical I-DID in Fig. 7. We observe the model node at time $t = 2$, $M_{j,l-1}^2$, is deterministically dependent on $M_{j,l-1}^1$. In other words, the set of lower level's models do not change when the model node is expanded at $t = 2$. In addition, CPTs of node A_j^2 have a uniform distribution.

C. Discussion

Time-critical I-IDs serve as the condensed form of time-critical I-DIDs while the deployed form is the expansion of the time-critical I-IDs. Converting the condensed form to the deployed form is similar to the expansion from I-IDs to I-DIDs. The complexity still relies on the model update where the number of candidate models ascribed to other agents grows exponentially over time. On the other hand, the conversion depends on the condensed form in which the time sequence indexed to the mode node has determined the complexity of lower level's models and manipulates the model growth [1]. For example, in Fig. 7, we do not need to expand the models from time $t = 1$ to $t = 2$ and the model space does not increase at $t = 2$. This results in a time-critical I-DID computationally cheap compared to the I-DID with all expanded models in every time step.

It is evident that the condensed form for the model node may reduce the complexity of model expansion. This makes time-critical I-DIDs tractable on the solutions. Meanwhile, the conversion may compromise the solution quality since we use the abstract models, instead of complete models, at the lower level. Hence it is quite critical to prepare the condensed form of time-critical I-IDs in a proper way. We shall handle this issue in the next section.

IV. Selection of Time-critical I-IDs

We proceed to build the appropriate condensed form by converting which would not incur much damage to the solutions of time-critical I-DIDs. Essentially, we need to index the model node with a suitable sub-time sequence in time-critical I-IDs.

A. Greedy Selection

We may build the condensed form by indexing a random sub-time sequence to the model node. By doing this it is unknown to the solution quality of the converted time-critical I-DIDs comparing to the I-DIDs that results from the expansion of the model node indexed by the master time sequence. We need a way to measure the impact of the condensed form on time-critical I-DID solutions when it is converted.

We follow the concept of behavioral equivalence [7] that quantifies the influence of other agent j's models (at a low level) on solutions of agent i's I-DIDs. Muthu *et. al.* [7] state: Two models of agent j, m_j and $\hat{m}_j$, are ϵ-behavioral equivalence if the probability distributions of agent i's action-observation paths induced by j's models, $Pr_{m_j}(A_i^T, O_i^T, \cdots, A_i^1)$ and $Pr_{\hat{m}_j}(A_i^T, O_i^T, \cdots, A_i^1)$, have the distance of ϵ measured by the Kullback-Leibler divergence (KL) [8]. The error of agent i's expected values (I-DID solutions) is bounded by $(R_i^{max} - R_i^{min})T \times 2\epsilon$, where T is the time horizon of I-DIDs and R_i^{max} (R_i^{min}) is the agent i's maximum (minimum) immediate reward, if an ϵ-behavioral equivalence model is pruned from the model node. More details refer to [7].

Following the same vein, we may define ϵ-behavioral equivalence between two condensed forms of agent j's models, m_j^{mts} and m_j^{sts}, that are indexed by a master time sequence and a sub-time sequence respectively. Recall that m_j^{mts} is the complete model of time-critical I-IDs that expands the model node at every time step. Subsequently we may limit the solution error by $(R_i^{max} - R_i^{min})T \times 2\epsilon$ if we select m_j^{sts} that has at most ϵ distance from m_j^{mts}. Here the distance is the KL divergence between the probability distributions of i's action-observation paths induced by m_j^{mts} and m_j^{sts} respectively. Appropriate usage will be self-evident in the rest of this paper.

Having the quantitative measurement, we are able to perform a greedy selection of the best condensed form for model m_j. The optimal way is to select the one, m_j^{sts}, that has the *least* distance from the condensed form, m_j^{mts}, indexed by a master sequence . To achieve this, we need to list all of the condensed forms m_j^{sts} for model m_j, and compare each of them to m_j^{mts}. For a T-time horizon model, the number of all condensed forms is $2^T - 1$. Recall that the computation of $Pr_{m_j}(A_i^T, O_i^T, \cdots, A_i^1)$ is very time-consuming in the comparison [7]. Meanwhile, the selection will be applied to all models $\mathcal{M}_j$ in the model node. Consequently, the greedy selection would become intractable - choosing the optimal m_j^{sts} for all models.

B. Entropy-based Selection

The condensed form is an abstract model of I-ID since its conversion may not expand/update other agents' models in every time step. Recall that other agents are supposed to take a random action for the particular time that is not indexed in the condensed form. An arbitrary selection may cause all models to be indexed with the same sub-time sequence. This results in completely unknown actions of other agents at some particular steps, which may incur much loss to the solution quality. Hence we need a strategic selection of the condensed form for all models in the model node.

To limit the solution error, we may consider only a subset of condensed forms that have at most ϵ distance from m_j^{mts}. We expect to avoid the case that all agent j's models select the same condensed form from the subset. Towards this insight, we propose an entropy-based method to measure the diversity of the condensed forms of all models in the selection. We take a Shannon based definition of information entropy that quantifies the information uncertainty [9].

[1] We agree that the complexity is also relevant to the time sequence indexed to other nodes. Currently, we focus on the condensed form of the model node since model expansion has more impact on time-critical I-DIDs as it does for normal I-DIDs.

Let $\{sts_1, sts_2, \cdots, sts_N\}$ be sub-time sequences for N agent j's models of the condensed form. We may cluster the same sub-time sequence into a class, c_i, and get K classes of sub-time sequences. Formally, the diversity of the condensed forms for N models is calculated in Eq. 1.

$$\begin{aligned} EP &= -\{\frac{|c_1|}{N} ln \frac{|c_1|}{N} + \cdots + \frac{|c_K|}{N} ln \frac{|c_K|}{N}\} \\ &= -\sum_{i=1}^{K} \frac{|c_i|}{N} ln \frac{|c_i|}{N} \end{aligned} \tag{1}$$

where $|c_i|$ is the cardinality of the i^{th}'s class c_i.

Eq. 1 offers us a quantitative selection when we are facing multiple condensed forms of model m_j and all of them are within at most ϵ distance from m_j^{mts}. We may choose the one that would increase the current diversity of the condensed forms. This may prevent the condensed forms from running into a single type. We formalize the entropy-based selection in Fig. 8.

ENTROPY-BASED SELECTION (Model set $\mathcal{M}_j$, I-ID m_i, ϵ)
returns Condensed form $CF(\mathcal{M}_j)$
1. **While** $\mathcal{M}_j$ not empty
2. Prepare the condensed form with the master time sequence, $m_j^{k,mts}$, for a model, $m_j^k \in \mathcal{M}_j$
3. Compute the probability distributions $Pr_{m_j^{k,mts}}(\cdot)$ [a] by calling **Get-Probability(**$m_j^{k,mts}$**,**m_i**)**
4. **Repeat**
5. Select the condensed form m_j^{k,sts_q} for the model m_j^k
6. Compute the probability distributions $Pr_{m_j^{k,sts_q}}(\cdot)$ [b] by calling **Get-Probability(**m_j^{k,sts_q}**,**m_i**)**
7. **If** $D_{KL}(Pr_{m_j^{k,mts}}(\cdot)||Pr_{m_j^{k,sts_q}}(\cdot)) \leq \epsilon$
8. Compute the diversity of the condensed forms $EP_{k,q}$ in Eq. 1
9. **If** $EP_{k,q-1} \leq EP_{k,q}$
10. $CF(\mathcal{M}_j) \stackrel{+}{\leftarrow} m_j^{k,sts_q}$, $\mathcal{M}_j \stackrel{-}{\leftarrow} m_j^k$
11. **While** $EP_{k,q-1} \leq EP_{k,q}$
12. Return $CF(\mathcal{M}_j)$

GET-PROBABILITY (m_j,I-ID m_i)
returns $Pr(A_i^T, O_i^T, \cdots, A_i^1)$
1. Convert the condensed form of I-ID, m_i, to the deployed form using the model m_j
2. Transform the deployed form (DID) into dynamic Bayesian networks by replacing i's decision nodes with chance nodes having uniform distribution
3. Compute the distribution $Pr(A_i^T, O_i^T, \cdots, A_i^1)$ from the dynamic Bayesian networks inference

[a] $Pr_{m_j^{k,mts}}(\cdot)$ refers to $Pr_{m_j^{k,mts}}(A_i^T, O_i^T, \cdots, A_i^1)$
[b] $Pr_{m_j^{k,sts}}(\cdot)$ refers to $Pr_{m_j^{k,sts}}(A_i^T, O_i^T, \cdots, A_i^1)$

Figure 8. Algorithm for selecting the condensed form for each j's model m_j^k using the entropy-based method. The function **Get-Probability(**m_j**,**m_i**)** computes the probability distributions on i's action-observation paths [7] when m_j is expanded in m_i.

The algorithm aims to prepare the model set, $CF(\mathcal{M}_j)$, in which each model has a suitable condensed form $m_j^{k,sts}$ using the entropy-based selection. We firstly compute the probability distributions of agent i's action-observation paths by expanding only the complete model $m_j^{k,mts}$ in I-ID (lines 2-3). The probability distributions are exact values since the model, $m_j^{k,mts}$ indexed by the master time sequence, exactly expands all model node over T time steps. The values provide a benchmark for measuring the condensed forms with sub-time sequences in the selection. Similarly, we obtain the probability distributions when one condensed form, m_j^{k,sts_q}, is selected to be included in m_i (lines 5-6). After that, we compute the KL distance, $D_{KL}(\cdot)$ [2], between the distributions induced by the two models (line 7). If the distance is within ϵ range we further check whether by including the sub-time sequence sts_q (indexed to m_j^{k,sts_q}) would increase the diversity of the condensed forms at the current time (lines 7-9). We select the condensed form m_j^{k,sts_q} if the diversity is increased (lines 9-10); otherwise, we take a new selection of the condensed form for the model m_j^k. In the worst case, we have to iterate all condensed forms of m_j^k. We select the condensed form that has the least distance, $D_{KL}(\cdot)$, from $m_j^{k,mts}$ if none of the condensed forms increases the diversity.

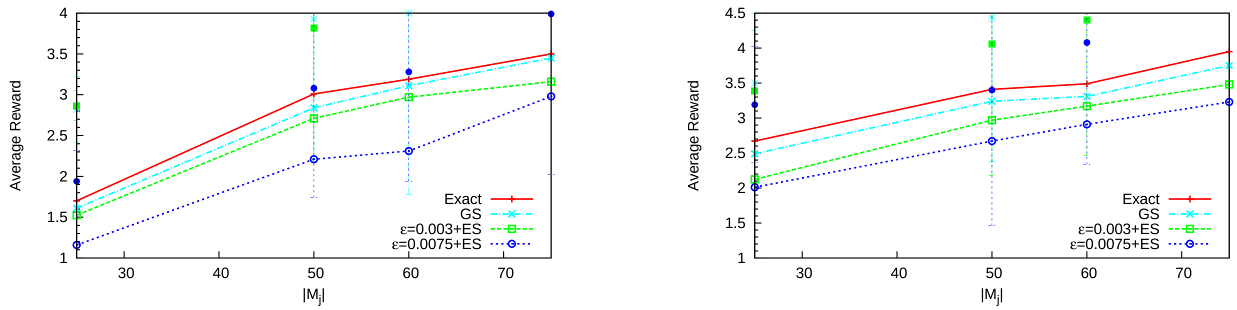
In summary, the entropy-based selection still bounds the solution error by $(R_i^{max} - R_i^{min})T \times 2\epsilon$, and may further reduce the actual error by providing a large diversity of the condensed forms in $\mathcal{M}_j$. It may gain much efficiency in comparison to the greedy selection.

V. EXPERIMENTAL RESULTS

We implemented both the greedy selection (GS) and the entropy-based algorithm (ES), and empirically demonstrate the time-critical I-DIDs (singly nested) on two well-studied domains: the multi-agent tiger [10] and a multi-agent version of the machine maintenance problem [11]. We show that the performance of ES algorithm (with different ϵ values) approaches GS performance regarding the solution quality of the time-critical I-DIDs. We compare to an exact method that actually indexes the master sequence to all j models and expand them exactly at each time step. The exact one serves as an optimal solution of I-DIDs. More importantly, we demonstrate that ES achieves much efficiency indicated by the low run times.

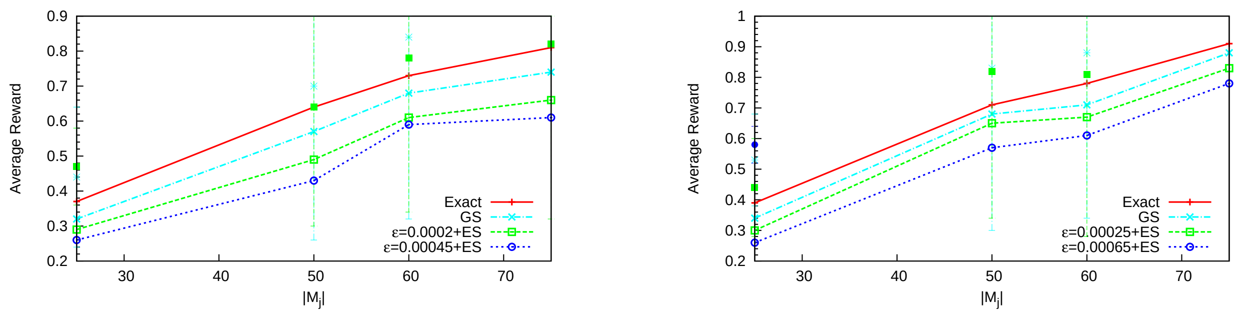
We select the condensed form of j's models through either GS or ES method. By including j's models, we convert the condensed form of i's I-IDs to the deployed form of time-critical I-DIDs. We compute i's optimal policies and play with agent j. In Figs. 9 and 10, we show the average rewards

[2] $D_{KL}(Pr_{m_j^{k,mts}}(\cdot)||Pr_{m_j^{k,sts_q}}(\cdot)) = \frac{1}{2}\left(Pr_{m_j^{k,mts}}(\cdot) log \frac{Pr_{m_j^{k,mts}}(\cdot)}{Pr_{m_j^{k,sts_q}}(\cdot)} + Pr_{m_j^{k,sts_q}}(\cdot) log \frac{Pr_{m_j^{k,sts_q}}(\cdot)}{Pr_{m_j^{k,mts}}(\cdot)}\right)$



$(a) Time Horizon = 4$ $(b) Time Horizon = 5$

Figure 9. Performance profile obtained by solving a level 1 time-critical I-DID for the multiagent tiger problem using both the greedy and entropy-based selections for (a) 4 horizons and (b) 5 horizons. The performance of ES selection approaches the GS while solution errors are bounded for both methods.



$(a) Time Horizon = 3$ $(b) Time Horizon = 4$

Figure 10. Performance profile obtained by solving a level 1 time-critical I-DID for the multiagent machine maintenance problem using both the greedy and entropy-based selections for (a) 3 horizons and (b) 4 horizons.

gathered by executing the policies obtained from solving the I-DIDs. We test all methods across different numbers of j's models ($|M_j|$). Each data point here is the average of 30 runs where the true model of the other agent, j, is randomly picked according to i's belief distribution over j's models.

We observe that the performance of both GS and ES selections are quite close to the performance of the exact method on the solution quality. Recall that the GS selection results in the optimal time-critical I-DIDs that has the least distance from the exact I-DIDs. Hence it is evident that time-critical I-DIDs really upset the solution quality, but do not incur much loss if the condensed form is appropriately selected. The performance of ES selection approaches ES and becomes better when ϵ is reduced. This is true for different sets of j's models on both problem domains.

Finally, we show the efficiency of ES selection measured by run times (seconds) in Table I. We ran 25 initial models of agent j in time-critical I-DIDs, and set 0 .005 and 0.0005 to ϵ values for the multiagent tiger and machine maintenance (MM) domains respectively. The ES selection achieves significantly lower run times in comparison to the GS. This may make time-critical I-DIDs more applicable in the real problem.

Level 1	**T**	**Time (s)**	
		GS	ES
Tiger	4	87.176	26.287
	5	188.973	67.245
MM	3	41.213	14.86
	4	122.897	33.158

Table I

COMPARISON OF RUN TIMES. ALL EXPERIMENTS ARE RUN ON A WINXP PLATFORM WITH A DUAL PROCESSOR XEON 2.0GHZ WITH 2GB MEMORY.

VI. RELATED WORKS

Timely action is often critical on facing rapid changes in the real world. The time-critical dynamic decision problem is to decide or select a course of actions that satisfies some objective in an environment under time constraints. Normative methods have been used to formalize time-critical decision problems in dynamic decision making systems e.g. Bayesian networks and influence diagrams [12]. Xiang and Poh [2], [4] proposed time-critical dynamic influence diagrams for the modelling of time and dealing with time-pressured situations to develop time-critical dynamic decision-support systems. Their work has been successfully applied to medical decision systems e.g. dynamic decision making for the treatment of cardiac arrest. Tseng and Gmytrasiewicz [13] studied real-time actions in recommendation systems through the development of informative influence diagrams. Meanwhile, Noh and Gmytrasiewicz [14] investigated time-critical multiagent decision making in the recursive modeling framework. They use performance profile to determine the appropriate scope of modeling, mainly on nested levels, and provide experimental results on an anti-air defense domain.

Interactive dynamic influence diagram contributes to a growing line of work on sequential multiagent decision making that extends the static single play games to multiagent interactive domain [5]. It explicitly models how other agents behave in a real time and provides online solutions for the subject agent. Due to its interactive state space that contains both the physical states and other agents' models, the solutions become extremely complicated. Several effective methods have been proposed mainly to limit an exponential growth of models over time [5], [15]. Extending interactive dynamic influence diagram for time-critical decision making would face more challenge. In principle, solutions follow two types of schemes: One is to explicitly encode time as one factor into a utility function of model and then maximize the utilities. The other is to maintain the utility form and then propose fast solutions of limited quality. This paper provides a solution of the second type.

VII. Conclusion

We propose a representation of time-critical I-DIDs. The representation contains the condensed form - time-critical I-IDs - and the deployed form that is time-critical I-DIDs and provides online solutions to the subject agent. The deployed form is converted from the condensed form similarly to the expansion of I-IDs to I-DIDs. The transformation faces the trade-off between model tractability and solution quality. We propose an entropy-based method for selecting the condensed form by including it would limit the solution error of time-critical I-DIDs and meanwhile increases the diversity of the condensed forms. Consequently, the method avoids a greedy selection of the condensed form for time-critical I-IDs and achieves significant computational savings on the solutions. Currently, we are exploring a more efficient implementation of time-critical I-DIDs using meta-reasoning.

Acknowledgement

The first author acknowledges partial support from both the Obel Family Foundation in Denmark and National Natural Science Foundation of China (No. 60974089 and No. 60975052). Yanping Xiang thank the support from Natural Science Foundation of China (No. 60974089).

References

[1] E. Horvitz and A. Seiver, "Time-critical action: Representations and application," in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 1997, pp. 250–257.

[2] Y. Xiang and K. leng Poh, "Time-critical dynamic decision making," in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 1999, pp. 688–695.

[3] I. Cohen, "Improving time-critical decision making in life-threatening situations: Observations and insights," *Decision Analysis*, vol. 5, no. 2, pp. 100–110, 2008.

[4] Y. Xiang and K. leng Poh, "Knowledge-based time-critical dynamic decision modeling," *Journal of the Operational Research Society*, vol. 53, no. 1, pp. 79–87, 2002.

[5] P. Doshi, Y. Zeng, and Q. Chen, "Graphical models for interactive pomdps: Representations and solutions," *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, vol. 18, no. 3, pp. 376–416, 2009.

[6] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (Second Edition)*. Prentice Hall, 2003.

[7] M. Kumaran, P. Doshi, and Y. Zeng, "Approximate solutions of interactive dynamic influence diagrams using ϵ-behavioral equivalence," in *International Symposium on Artificial Intelligence and Mathematics (ISAIM)*, 2010.

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

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Second Edition)*. Wiley, 2006.

[10] P. Doshi and P. J. Gmytrasiewicz, "A particle filtering based approach to approximating interactive pomdps," in *Proceedings of the Twentieth Conference on Association for the Advancement of Artificial Intelligence (AAAI)*, 2005, pp. 969–974.

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

[12] E. Horvitz, "Computation and action under bounded resources," *Ph.D. Thesis, Stanford University*, 1990.

[13] C. che Tseng and P. Gmytrasiewicz, "Real time decision support system for portfolio management," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 2002, p. 79.

[14] S. Noh and P. Gmytrasiewicz, "Identifying the scope of modeling for time-critical multiagent decision-making," in *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI)*, 2001, pp. 1043–1048.

[15] Y. Zeng and P. Doshi, "Speeding up exact solutions of interactive dynamic influence diagrams using action equivalence," in *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, 2009, pp. 1996–2001.