Learning Action Models from Disordered and Noisy Plan Traces

Hankz Hankui Zhuo\textsuperscript{1}, Jing Peng\textsuperscript{1}, Subbarao Kambhampati\textsuperscript{2}
\textsuperscript{1}School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
\textsuperscript{2}Department of Computer Science and Engineering, Arizona State University, Tempe, Arizona, US
zhuohank@mail.sysu.edu.cn, pengj69@mail2.sysu.edu.cn, rao@asu.edu

Abstract

There is increasing awareness in the planning community that the burden of specifying complete domain models is too high, which impedes the applicability of planning technology in many real-world domains. Although there are many learning systems that help automatically learning domain models, most existing work assumes that the input traces are completely correct. A more realistic situation is that the plan traces are disordered and noisy, such as plan traces described by natural language. In this paper we propose and evaluate an approach for doing this. Our approach takes as input a set of plan traces with disordered actions and noise and outputs action models that can best explain the plan traces. We use a MAX-SAT framework for learning, where the constraints are derived from the given plan traces. Unlike traditional action models learners, the states in plan traces can be partially observable and noisy as well as the actions in plan traces can be disordered and parallel. We demonstrate the effectiveness of our approach through a systematic empirical evaluation with both IPC domains and the real-world dataset extracted from natural language documents.

Introduction

Most work in planning assumes that complete domain models are given as input in order to synthesize plans. However, there is increasing awareness that building domain models at any level of completeness presents steep challenges for domain creators. As planning issues become more and more realistic, the action model will become more and more complex. So it is necessary for us to consider how to automatically attain the domain action model. Indeed, recent work in web-service composition (c.f. [Bertoli et al., 2010; Hoffmann et al., 2007]) and workflow management (c.f. [Blythe et al., 2004]) suggest that dependence on complete models can well be the real bottlenecks inhibiting applications of current planning technology.

Attempts have been made to design systems to automatically learn domain models from pre-specified (or precollected) plan traces (a plan trace is composed of an action sequence with zero or more partial states). For example, Amir [Amir, 2005] presented a tractable and exact technique for learning action models known as Simultaneous Learning and Filtering (SLAF). Yang et al. [Yang et al., 2007; Zhuo et al., 2010; 2014] proposed to learn STRIPS action models [Fikes and Nilsson, 1971] from plan traces with partially observed states. Bryce et al. propose an approach called Marshal to issue queries, and learns models by observing query answers, plan solutions, and direct changes to the model [Bryce et al., 2016]. [Aineto et al., 2018] propose to learn STRIPS models via classical planning. These systems, however, are all based on the assumption that actions in plan traces are correct and totally ordered.

In many real-world applications, however, plan traces are often extracted or built from raw data, such as sensing signals or texts, by off-the-shelf systems, due to the high cost of collecting plan traces by hand. Those plan traces are often disordered and noisy. For example, the text “before you go into the bus station, you should buy a ticket. Meanwhile please remember to bring the luggage with you” addresses the procedure of going traveling. The action sequences that can be extracted by off-the-shelf approaches such as [Feng et al., 2018; Branavan et al., 2012] is “goto(station), buy(ticket), remember(luggage), bring(luggage)”, where “buy(ticket)” and “bring(luggage)” should be executed before “goto(station)”, and action “remember(luggage)” is noisy. Moreover, more than one action may be executed at the same time, i.e., parallel actions are pervasive [Aghighi and Bäckström, 2017].

Although Mourao et al. propose to learn action models from noisy observations of intermediate states, they assume actions are totally ordered. Zhuo and Kambhampati consider actions are noisy in plan traces [Zhuo and Kambhampati, 2013]. They, however, do not allow states to be noisy, and employ strong assumption that noisy actions are similar to their corresponding correct actions when building the conditional probability of correct actions given the input noisy actions.

In this paper, we aim to learn action models from plan traces with disordered actions, parallel actions, and noisy states. This is challenging in the sense that each of the three uncertain cases can harm the learning quality of action models. To address the challenge, we build three types of constraints, namely disorder constraints, parallel constraints, and noise constraints, to capture information from disordered actions, parallel actions, and noisy states, respectively. We then
solve the constraints with an off-the-shelf weighted MAX-SAT [Bacchus et al., 2018; Li et al., 2007] solver and convert the solution to action models. We denote our approach by AMDN, which stands for Learning Action Models from Disordered and Noisy plan traces. We will evaluate AMDN on both IPC1 domains and the real-world dataset extracted from natural language documents.

Although our AMDN approach uses the same MAX-SAT framework with previous work such as ARMS [Yang et al., 2007] and ML-CBP [Zhuo and Kambhampati, 2017], the constraints we build based on disordered and noisy plan traces are totally different from the ones built by previous work. Compared to building constraints based on perfectly correct plan traces as done by previous work, building constraints based on disordered and noisy traces is challenging.

In the remainder of the paper, we first review previous work related to our approach, and then present the formal definition of our problem. After that, we provide the detailed description of our AMDN algorithm and evaluate AMDN in three planning domains. Finally we conclude our paper with future work.

Related Work

There have been many approaches on learning action models from plan traces. Previous research efforts differ mainly on whether plan traces consist of actions and intermediate states (c.f. [Gil, 1994]) or only actions (c.f. [Yang et al., 2007; Zhuo et al., 2014]). While the latter assumption makes the problem harder than the former, in both cases, whatever observed is assumed to be observed perfectly. Both of them assume non-noisy plan observations. Recently, Gregory et al. present an algorithm called LOP to induce static predicates to the learning system, which finds a set of minimal static predicates for each operator that preserves the length of the optimal plan [Gregory and Cresswell, 2016]. Instead of an action model, a set of successfully executed plans are given and the task is to generate a plan to achieve the goal without failing [Roni Stern, 2017]. Lindsay et al. propose to learn action models action descriptions in the form of restricted template [Lindsay et al., 2017]. Despite the success of those approaches, they all assume actions are correctly ordered and states are not noisy.

Preliminaries and Problem Definition

A complete STRIPS domain is defined as a tuple $D = \langle R, M \rangle$, where $R$ is a set of predicates with typed objects and $M$ is a set of action models. Each action model is a quadruple $(a, \text{PRE}(a), \text{ADD}(a), \text{DEL}(a))$, where $a$ is an action name with zero or more parameters, $\text{PRE}(a)$ is a precondition list specifying the conditions under which $a$ can be applied, $\text{ADD}(a)$ is an adding list and $\text{DEL}(a)$ is a deleting list indicating the effects of $a$. We denote $R_O$ as the set of propositions instantiated from $R$ with respect to a set of typed objects $O$. Given $D$ and $O$, we define a planning problem as $P = (O, s_0, g)$, where $s_0 \subseteq R_O$ is an initial state, $g \subseteq R_O$ are goal propositions.

We denote a set of parallel actions in $A$ in a plan, where actions in $A$ can be executed in any order or simultaneously. For example, let $A = \{a_1, a_2\}$. The sequence $\langle s_0, A, s_1 \rangle$ is equivalent to $\langle s_0, a_1, a_2, s_1 \rangle$ or $\langle s_0, a_2, a_1, s_1 \rangle$. A solution plan to $P$ with respect to model $D$ is a sequence of parallel actions $p = \{A_1, A_2, \ldots, A_n\}$ that achieves goal $g$ starting from $s_0$. If actions $a_x \in A_i$ and $a_y \in A_j$, we define the distance of $a_x$ and $a_y$ by $|a_x - a_y| = |i - j|$. If the positions of $a_x$ and $a_y$ are exchanged, we say they are disordered with respect to their correct order in the plan.

A plan trace $t$ is defined by $t = \langle s_0, A_1, s_1, \ldots, A_n, g \rangle$, where actions can be disordered. We assume that the probability of two actions being disordered decreases as their distance increases. This is reasonable in the sense that actions in a short-term horizon are more likely to be disordered than a long-term horizon. State $s_i$ is both partial and noisy, indicating some propositions are missing in $s_i$ and some propositions in $s_i$ are incorrect. We denote $T$ as a set of plan traces. Our problem can be defined by: given as input a set of observed plan traces $T$, our approach outputs a domain model $M$ that best explains the observed plan traces.

An example of our learning problem for the depots domain can be found in Figure 1, which is composed of two parts: plan traces as input (Figure 1(a)) and action models as output (Figure 1(b)). In Figure 1(a), $t^1$ and $t^2$ are two plan traces, where initial states and goals are drawn over. The dark parts indicate the incorrect propositions or disorder actions. For example, in $t^1$, “drop(h1 c0 p1 dp1)” and “load(h0 c0 t0 dp0)” are disordered, “(at t0 dp0)” is a noisy proposition which should be “(at t0 dp1)”. In Figure 1(b), the action model “drive” is one of the output action models of our algorithm.

The AMDN Approach

In this section we present AMDN approach in detail. An overview of our approach is shown in Algorithm 1. We first build sets of disorder constraints, parallel constraints and noise constraints, and solve these constraints with an off-the-shelf MAX-SAT solver. We then directly convert the solution to action models $M$.

#### Algorithm 1

An overview of our AMDN approach

**Input:** a set of plan traces $T$.

**Output:** a set of action models $M$.

1: build disorder constraints;
2: build parallel constraints;
3: build noise constraints;
4: solve all constraints with a MAX-SAT solver
5: convert the solution to action models $M$;
6: return $M$;

---

1 http://www.icaps-conference.org/index.php/Main/Competitions
2 http://planning.cis.strath.ac.uk/competition/
Building disorder constraints

In Step 1 of Algorithm 1, we build a set of disorder constraints. The disorder constraints are soft constraints, which are designed to exploit the information carried by the execution order of actions. We denote these constraints by DC.

**Constraint DC** Consider two adjacent sets of parallel actions \( \langle A_i, A_{i+1} \rangle \). We assume that all actions in \( A_i \) are not parallel with some action in \( A_{i+1} \), vice versa. In other words, for each action \( a_y \) in \( A_{i+1} \), there exists \( a_x \) in \( A_i \), such that there are conflicts between \( a_y \) and \( a_x \). The conflicts may be as shown below:

1. 1. A proposition \( r \) in the precondition list of \( a_x \) but not in the delete list of \( a_x \) will be deleted by \( a_y \), which can be encoded by: \( \exists r \in \text{PRE}(a_x) \land r \notin \text{DEL}(a_x) \land r \in \text{DEL}(a_y) \).

2. 2. A proposition \( r \) in the precondition list of \( a_y \) is added by \( a_x \), which can be encoded by: \( \exists r \in \text{ADD}(a_x) \land r \in \text{PRE}(a_y) \).

3. 3. A proposition \( r \) is in the add list of \( a_x \) and it is in the delete list of \( a_y \), which can be encoded by: \( \exists r \in \text{ADD}(a_x) \land r \in \text{DEL}(a_y) \).

4. 4. A proposition \( r \) is in the delete list of \( a_x \) and it is in the add list of \( a_y \), which can be encoded by: \( \exists r \in \text{DEL}(a_x) \land r \in \text{ADD}(a_y) \).
We formulate the constraints as follows and denote them by DC:

\[ \langle A_i, A_{i+1} \rangle \Rightarrow \forall a_y \in A_{i+1} \exists a_x \in A_i \exists r \]
\[ (r \in \text{PRE}(a_x) \land r \notin \text{DEL}(a_x) \land r \in \text{DEL}(a_y)) \]
\[ \lor (r \in \text{ADD}(a_x) \land r \in \text{PRE}(a_y)) \]
\[ \lor (r \in \text{ADD}(a_x) \land r \in \text{DEL}(a_y)) \]
\[ \lor (r \in \text{DEL}(a_x) \land r \in \text{ADD}(a_y)). \]

**Weight of DC** If the actions are ordered, DC can be hard constraints. When the actions are disordered, we set the weight as follows: \( w_\text{dc} = w_{do} p_d \), where \( w_{do} \) is the original weight (we empirically set \( w_{do} \) to be 5) and \( p_d \) is the belief of this constraint. In this paper, we calculate \( p_d \) by

\[ p_d = \begin{cases} \frac{\theta_{\text{disco}}}{K_1} & K_1 \neq 0 \\ 1 & K_1 = 0 \end{cases} \]

where \( \theta_{\text{disco}} \) is the probability of disorder. \( K_1(a_i, a_j) \) is the distance between \( A_i \) and \( A_{i+1} \), which can be defined by \( K_1(a_i, a_j) = |i + 1 - j| \). For instance, if \( a_{i1}, a_{i2} \in A_i, a_j \in A_{i+1} \) and \( a_k \in A_{i+2} \), where \( A_i, A_{i+1} \) and \( A_{i+2} \) are adjacent sets of parallel actions in a plan trace \( t \), we have \( K_1(a_{i1}, a_{i2}) = 1, K_1(a_{i1}, a_j) = 0 \) and \( K_1(a_{i1}, a_k) = 1 \).

**Building parallel constraints**

In Step 2 of Algorithm 1, we build a set of parallel constraints. In this subsection, to make sure that the learned action models are succinct and consistent with the STRIPS language, we first enforce a set of hard constraints that must be satisfied by action models, which were also built by [Yang et al., 2007]. We then build soft constraints between actions in the same set of parallel actions. We formulate the constraints as follows and denote them by PC.

**Constraint P.1** An action may not add a proposition which already exists before the action is applied. We formulate the constraints as follows and denote them by P1: \( r \in \text{ADD}(a) \Rightarrow r \notin \text{PRE}(a) \).

**Constraint P.2** An action may not delete a proposition which does not exist before the action is applied. We formulate the constraints as follows and denote them by P2: \( r \in \text{DEL}(a) \Rightarrow r \in \text{PRE}(a) \).

**Constraint P.3** For the case that two action \( a_i \) and \( a_j \) are in the same parallel actions \( A_i \), if \( a_i \) add or delete a proposition \( r \), \( r \) can not be in the add list or delete list of \( a_j \). In other words, \( r \) can not belong to all two of \( \text{ADD}(a_i), \text{DEL}(a_j), \text{ADD}(a_j) \) and \( \text{DEL}(a_j) \) at the same time. We formulate the constraints as follows and denote them by P3:

\[ a_i, a_j \in A \Rightarrow \text{ADD}(a_i) \cap \text{DEL}(a_i) \cap \text{ADD}(a_j) \cap \text{DEL}(a_j) = \emptyset. \]

**Weight of PC** In P1 and P2, to make sure these constraints are hard, we assign these constraints with a large enough weight, denoted by \( w_{\text{max}} \). Similar to DC, if the actions are ordered, P3 can be hard constraints. When the actions are disordered, we set the weight as follows: \( w_p = w_{pp} p_p \), where \( w_{pp} \) is the original weight (similar to \( w_{do} \), we empirically set \( w_{pp} \) to be 5) and \( p_p \) is the belief of this constraint. In this paper, we calculate \( p_p \) by

\[ p_p = \begin{cases} \frac{\theta_{\text{disco}}}{K_2} & K_2 \neq 0 \\ 1 & K_2 = 0 \end{cases} \]

where \( K_2(a_i, a_j) \) is the distance between \( A_j \) and \( A_i \), which can be defined by \( K_2(a_i, a_j) = |i - j| \). For instance, if \( a_{i1}, a_{i2} \in A_i, a_j \in A_{i+1} \), and \( a_k \in A_{i+2} \), where \( A_i, A_{i+1} \) and \( A_{i+2} \) are adjacent sets of parallel actions in a plan trace \( t \), we have \( K_2(a_{i1}, a_{i2}) = 0, K_2(a_{i1}, a_j) = 1 \) and \( K_2(a_{i1}, a_k) = 2 \).

**Building noise constraints**

In Step 3 of Algorithm 1, we build a set of noise constraints. The noise constraints are designed to be soft, directing the MAX-SAT algorithm towards learning the most probable complete description of actions. These constraints are used to explain why the observed states exist in a plan. We denote these constraints by NC.

**Constraint N.1** Consider a pair \((a, r)\), where \( a \) is an action and \( r \) is a proposition, \( r \) should not be deleted by \( a \) since \( r \) exists after \( a \). We formulate the constraints as follows and denote them by N1: \((a, r) \Rightarrow r \notin \text{DEL}(a)\).

**Constraint N.2** Consider a sequence \( \langle s_0, a_0, a_1, \ldots, a_n, r \rangle \), where \( a_i (0 \leq i \leq n) \) is an action and \( r \notin s_0 \) is a proposition. \( r \) must be in the add list of some \( a_i (0 \leq i \leq n) \). We formulate the constraints as follows and denote them by N2:

\[ \langle s_0, a_0, a_1, \ldots, a_n, r \rangle \land r \notin s_0 \Rightarrow r \in \text{ADD}(a_0) \cup \text{ADD}(a_1) \cup \ldots \cup \text{ADD}(a_n). \]

**Constraint N.3** Consider a pair \((r, a)\), where \( r \) is a proposition appearing before action \( a \). If the probability of its occurrence over all of the plan traces is higher than the probability threshold \( \theta_s \), we set the constraints as follows and denote them by N3:

\[ \langle r, a \rangle \land \theta(\langle r, a \rangle) > \theta_s \Rightarrow r \in \text{PRE}(a), \]

where \( \theta(\langle r, a \rangle) = \frac{\text{number of } \langle r, a \rangle}{\text{number of all } \langle r, a \rangle} \).

**Weight of NC** These constraints are designed to be soft, since the observed states are noisy and the actions are disordered. Even without noise, we can not be certain that N3 is correct. So the weight of these constraints is dynamic, which are designed to be \( w_n = w_{no} p_n P_f \), where \( w_{no} \) is the original weight (similar to \( w_{do} \), we empirically set \( w_{no} \) to be 5), \( P_n \) is the belief of the constraint and \( P_f \) is the frequency of observation. In N1, we calculate \( P_n \) by using the function \( P_n = \frac{\theta_{\text{disco}}}{K_3} \) if \( K_3 \neq 0 \), and \( P_n = 1 \) otherwise. \( K_3(a, r) \) is the distance between \( A_i \) and the first parallel actions before
r, where \( a_i \in \mathcal{A}_i \). So \( K_3(a_i, r) = |i - j| \), where \( a_i \in \mathcal{A}_i \), \( r \in s \) and \( \mathcal{A}_j \) is the first set of parallel actions before \( s \). For instance, if \( a_i \in \mathcal{A}_i, a_j \in \mathcal{A}_{i+1}, r \in s_i \) and \( (\mathcal{A}_i, s_i, \mathcal{A}_{i+1}) \), we have \( K_3(a_i, r) = 0 \) and \( K_3(a_j, r) = 1 \). In N.2, \( P_n \) is defined by

\[
P_n = \frac{\text{the number of actions executed before } r}{\text{the total number of actions in this trace}}.
\]

In N.3, we calculate \( P_n \) using the function \( P_n = \frac{\theta_{S_{\text{occ}}}}{\mathcal{A}_i} \) if \( K_4 \neq 0 \), \( P_n = 1 \) otherwise. \( K_4(r, a) \) is the distance between \( \mathcal{A}_i \) and the first set of parallel actions after \( r \), where \( a_i \in \mathcal{A}_i \). So \( K_4(r, a) = |i - j| \), where \( a_i \in \mathcal{A}_i \), \( r \in s \) and \( s_i, \mathcal{A}_i, \mathcal{A}_{i+1} \). For instance, if \( a_i \in \mathcal{A}_i, a_j \in \mathcal{A}_{i+1}, r \in s_i \) and \( (s_i, \mathcal{A}_i, \mathcal{A}_{i+1}) \), we have \( K_4(r, a_i) = 0 \) and \( K_4(r, a_j) = 1 \). \( P_f \) is derived from the statistics. For instance, in N.1, if we observed \( (a, r_1) \) one time and observed \( (a, r_2) \) nine times, \( P_f((a, r_1)) = 0.1 \) and \( P_f((a, r_2)) = 0.9 \). This is similar in N.2 and N.3.

### Solving Constraints

In Steps 4 and 5 of Algorithm 1, we aim to solve all the constraints and convert the solution to action models. We put all constraints together and solve them with a weighted MAXSAT solver [LI et al., 2007]. The greater the weight of the constraint, the higher his priority in the MAX-SAT solver. We exploit MaxSat [LI et al., 2007] to solve all the constraints, and attain a true or false assignment to maximally satisfy the weighted constraints. Given the solution assignment, the construction of action models \( \mathcal{M} \) is straightforward; e.g., if “\( r \in \text{ADD}(a) \)” is assigned true in the result of the solver, \( r \) will be converted into an effect of \( a \).

### Experiments

In the section, we first evaluate our AMDN algorithm on planning domains to carefully investigate the properties of our approach in handling disordered and noisy scenarios. We then evaluate our approach on a real-world domain which is described in natural language to further look into the effectiveness of our approach in practice.

### Criterion

We define the accuracy \( \text{Acc} \) of our AMDN algorithm by comparing its learned action models with the artificial action models which are viewed as the ground truth. We define the error rate of the learning result by calculating the missing and extra predicates of the learned action models. Specifically, for each learned action model \( a \), if a precondition of \( a \) does not exist in the ground-truth action model, the number of errors increases by one; if a precondition of the ground-truth action model does not exist in \( a \’ \)s precondition list, the number of errors also increases by one. As a result, we have the total number of errors of preconditions with respect to \( a \). We define the error rate of the total number of errors among all the possible preconditions of \( a \), that is,

\[
\text{Err}_{\text{pre}}(a) = \frac{\text{the total number of errors of preconditions}}{\text{all the possible precondition of } a}.
\]

Likewise, we can calculate the error rates of adding effects and deleting effects of \( a \), and denote them as \( \text{Err}_{\text{add}}(a) \) and \( \text{Err}_{\text{del}}(a) \), respectively. Furthermore, we define the error rate of all the action models \( \mathcal{M} \) (denoted as \( \text{Err}(\mathcal{M}) \)) as an average of \( \text{Err}_{\text{pre}}(a) \), \( \text{Err}_{\text{add}}(a) \) and \( \text{Err}_{\text{del}}(a) \) for all the actions \( a \) in \( \mathcal{M} \), that is,

\[
\text{Err}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{a \in \mathcal{M}} \left( \text{Err}_{\text{pre}}(a) + \text{Err}_{\text{add}}(a) + \text{Err}_{\text{del}}(a) \right),
\]

and define the accuracy as \( \text{Acc} = 1 - \text{Err}(\mathcal{M}) \). Note that domain model \( \mathcal{M} \) is composed of a set of action models.

### Evaluation on planning domains

We first evaluate AMDN algorithm in three planning domains: \( \text{blocks}^3 \), \( \text{driverlog}^2 \), and \( \text{depots}^4 \). In each domain, we generate 20 to 120 plan traces for learning domain models. In our datasets, the rate of observations \( \theta_1 \) means that only \( \theta_1 \) of the propositions can be observed. If the probability of noise is \( \theta_2 \), for each one of our observed propositions, the probability of being replaced by another different and random proposition is \( \theta_2 \). If the probability of disorder is \( \theta_3 \), for all \( a_i \in \mathcal{A}_i \), \( a_g \in \mathcal{A}_g \) and \( a_j \in \mathcal{A}_j \), the probability of exchanging the order of \( a_x \) and \( a_y \) is \( \frac{\theta_3}{K} \), where \( K \) is the distance between \( a_x \) and \( a_y \). In the experiments, we generate \( n \) plan traces (training data) by the following steps:

1. We first generate a set of correct plan traces \( \{t\} \) by using a planner (such as FF\(^3\)) to solve randomly generated planning problems, assuming correct action models (ground truth action models) are available.

2. For each plan trace \( t \), we generate parallel actions according to the definition in the problem definition section. Initialize an empty trace \( t’ = \{\} \) and a parallel actions \( A = \{a_0\} \). We denote by \( s_{\text{first}}, s_{\text{second}} \) and \( s_{\text{third}} \) the first, the second and the third state in the trace \( t \), respectively. While there are actions in \( t \), we pop the first action \( a \) and test if \( (s_{\text{first}}, A, a, s_{\text{third}}) \) is equivalent to \( (s_{\text{first}}, a, A, s_{\text{third}}) \). If true, we push \( a \) into \( A \) and delete \( s_{\text{second}} \). If false, we push \( s_{\text{first}} \) from \( t \) and replace \( A \) by \( \{a\} \). For instance, we have \( t = \{s_0, a_0, s_1, a_1, ..., s_g\} \). We first initialize \( t’ = \{\} \) and \( A = \{a_0\} \). Then we test if \( (s_0, A, a_1, s_2) \) is equivalent to \( (s_0, a_1, A, s_2) \). If true, \( A \) will be \( \{a_0, a_1\} \) and \( t \) will be \( \{s_0, A, s_2, a_2, ..., g\} \). If false, \( t’ \) will be \( (s_0, A) \), \( A \) will be \( \{a_1\} \) and \( t \) will be \( \{s_1, A, s_2, a_2, ..., g\} \).

3. For each proposition in \( \{t’\} \), we ignore it with a \( (1 - \theta_1) \) probability. If we do not ignore it, we replace it by another different and random proposition according the probability of \( \theta_2 \).

4. For each \( t’ \) in \( \{t’\} \) and for all \( a_i \) and \( a_j \) in \( t \), we exchange their order with the probability \( \frac{\theta_3}{d(a_i, a_j)} \), where \( d(a_i, a_j) \) is the distance between \( a_i \) and \( a_j \).

We evaluate AMDN in the following aspects. We first compare the accuracies of the action models learned by AMDN, AMAN [Zhuo and Kambhampati, 2013] and ARMS [Yang et al., 2013].
by increasing number of plan traces or increasing rate of observation. We then test the sensitivity to noise and disorder by comparing to the baselines AMAN and ARMS. Finally, we show the running time to see the efficiency of AMDN. In all experiments we set \( \theta_1 = 0.1 \), \( \theta_{diso} = \theta_3 \), \( w_{do} = w_{po} = w_{no} = 5 \) and \( w_{\max} = 9999 \).

Comparison between AMDN, AMAN and ARMS

To see the benefit of our algorithm when solving the disorder and noisy plan traces, first we generate from 20 to 120 plan traces with \( \theta_1 = 0.2, \theta_2 = 0.05 \) and \( \theta_3 = 0.05 \). And we compare the accuracies of action models learned by AMDN, AMAN and ARMS increasing number of plan traces. Then we generate 60 traces with \( \theta_2 = 0.05 \) and \( \theta_3 = 0.05 \). And we compare the accuracies of action models learned by AMDN, AMAN and ARMS increasing rate of observations from 0.2 to 1. The accuracy is an average over 20 runs.

Sensitivity to disorder

We also would like to see the sensitivity of AMDN with respect to disorder of actions. We generate 60 traces with \( \theta_1 = 0.6 \) and \( \theta_2 = 0 \) and take AMAN and ARMS as the baselines. It is easy to see whether AMDN is robust by comparing the accuracies with AMAN and ARMS with \( \theta_3 \) from 0 to 0.2. The accuracy is an average over 20 runs. The results are shown in Figure 4.

Sensitivity to noise

We also would like to see the sensitivity of AMDN to noise. We generate 60 traces with \( \theta_1 = 0.6 \) and \( \theta_3 = 0 \) and take AMAN and ARMS as the baselines. It is easy to see whether AMDN is robust with respect to noise by comparing the accuracies with AMAN and ARMS with \( \theta_2 \) from 0 to 0.2. The accuracy is an average over 20 runs.
The result is shown in Figure 5. We find that in all three domains, as the noise increases, the accuracies of all three algorithms decline, but AMDN drops slower than AMAN and ARMS significantly, which implies AMDN is more robust to noise than the other two. Without considering the noise, AMAN and ARMS will get a lot of wrong information. In contrast, AMDN adjusts the weight of each constraint based on the probability of noise. Frequently observed constraints will be given greater weight. So the wrong constraints will get small weights, which minimizes the effect of noise on the MAXSAT solver.

**Running time**

To study the efficiency of AMDN, we ran AMDN over 50 problems and calculated an average of the running time with respect to different number of plan traces in the blocksworld domain. The result is shown in Figure 6. As can be seen from the figure, the running time increases polynomially with the number of input plan traces. This can be verified by fitting the relationship between the number of plan traces and the running time to a performance curve with a polynomial of order 2 or 3. For example, the fit polynomial in Figure 6 is $0.3845x^2 + 4.837x - 15.32$. The results for the other two domains are similar to blocksworld, i.e., the running time also polynomially increases as plan traces increase.

**Evaluation on real-world dataset**

We conducted experiments on the dataset “CookingTutorial”\(^5\), which is about how to cook food in the form of natural language, e.g., “Cook the rice the day before, or use leftover rice in the refrigerator. The important thing to remember is not to heat up the rice, but keep it cold.”, which addresses the procedure of making egg fired rice. We exploited an off-the-shelf approach EASDRL [Feng et al., 2018] to extract action sequences from the dataset, e.g., an action sequence of “cook(rice), keep(rice, cold)” or “use(leftover rice), keep(rice, cold)” is expected to be extracted based on the above-mentioned example. There are 116 texts with 24284 words in total, describing about 2500 actions. Since different action sequences can be extracted from a single text (describing optional ways of cooking in the text), we generated about 400 action sequences, which are generally disordered and noisy (we viewed actions that occur less than 1% times as noisy actions). In order to learn action models from the action sequences, we built rules based on syntactical parsing of texts to generate initial states and goals for each action sequence. We also manually built action models as ground-truth models to evaluate the model learnt by our approaches.

![Figure 6: The running time of AMDN for domain blocksworld.](image)

The experimental results are shown in Table 1, where the first column is the number of the plan traces. From Table 1, we can see that AMDN performs much better than both ARMS and AMAN in all cases, which indicates our constraints built based on noisy and disordered actions can indeed help improve the learning accuracy. We can also find that the accuracy of our approach generally increases with respect to the increase of plan traces. This is consistent with our intuition since the more the plan traces are, the more the knowledge we have for handling disordered and noisy scenarios.

![Figure 6: The running time of AMDN for domain blocksworld.](image)

| number of traces | AMDN   | ARMS  | AMAN   |
|------------------|--------|-------|--------|
| 100              | 0.656  | 0.622 | 0.634  |
| 200              | 0.788  | 0.651 | 0.682  |
| 300              | 0.854  | 0.714 | 0.746  |
| 400              | 0.882  | 0.732 | 0.756  |

We compared our AMDN approach to ARMS and AMAN. The experimental results are shown in Table 1, where the first column is the number of the plan traces. From Table 1, we can see that AMDN performs much better than both ARMS and AMAN in all cases, which indicates our constraints built based on noisy and disordered actions can indeed help improve the learning accuracy. We can also find that the accuracy of our approach generally increases with respect to the increase of plan traces. This is consistent with our intuition since the more the plan traces are, the more the knowledge we have for handling disordered and noisy scenarios.

**Conclusion**

In this paper, we presented a system called AMDN for learning domain models for disordered and noisy plan traces. AMDN is able to integrate knowledge from parallel actions and a set of disordered and noisy plan traces to produce action models. With the plan traces, we first build disorder constraints, parallel constraints and noise constraints, and then using the weighted MAX-SAT solver to solve them. Our approach is well suited for scenarios where high-quality plan traces is hard to attain. Our experiments exhibit that our approach is effective in both planning domains and real-world domain. In the future, we would like to extend AMDN to investigating the executability of the learnt action models from real-world domains.

\[^5\]http://cookingtutorials.com/
References

[Aghighi and Bäckström, 2017] Meysam Aghighi and Christer Bäckström. Plan reordering and parallel execution - A parameterized complexity view. In AAAI, pages 3540–3546, 2017.

[Aineto et al., 2018] Diego Aineto, Sergio Jiménez, and Eva Onaindia. Learning STRIPS action models with classical planning. In ICAPS, pages 399–407, 2018.

[Amir, 2005] Eyal Amir. Learning partially observable deterministic action models. In IJCAI, pages 1433–1439, 2005.

[Bacchus et al., 2018] Fahiem Bacchus, Antti Hyttinen, Matti Järvisalo, and Paul Saikko. Reduced cost fixing for maximum satisfiability. In IJCAI, pages 5209–5213, 2018.

[Bertoli et al., 2010] Piergiorgio Bertoli, Marco Pistore, and Paolo Traverso. Automated composition of web services via planning in asynchronous domains. Artificial Intelligence Journal, 174(3-4):316–361, 2010.

[Blythe et al., 2004] J. Blythe, E. Deelman, and Y. Gil. Automated composed workflows for grid environments. IEEE Intelligent Systems, 19(4):16–23, 2004.

[Branavan et al., 2012] S. R. K. Branavan, Nate Kushman, Tao Lei, and Regina Barzilay. Learning high-level planning from text. In ACL, pages 126–135, 2012.

[Bryce et al., 2016] Daniel Bryce, J. Benton, and Michael W. Boldt. Maintaining evolving domain models. In IJCAI, pages 3053–3059, 2016.

[Feng et al., 2018] Wenfeng Feng, Hankz Hankui Zhuo, and Subbarao Kambhampati. Extracting action sequences from texts based on deep reinforcement learning. In IJCAI, 2018.

[Fikes and Nilsson, 1971] Richard Fikes and Nils J. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. Artif. Intell., 2(3/4):189–208, 1971.

[Gil, 1994] Yolanda Gil. Learning by experimentation: Incremental refinement of incomplete planning domains. In ICML, pages 87–95, 1994.

[Gregory and Cresswell, 2016] Peter Gregory and Stephen Cresswell. Domain model acquisition in the presence of static relations in the LOP system. In IJCAI, pages 4160–4164, 2016.

[Hoffmann et al., 2007] Joerg Hoffmann, Piergiorgio Bertoli, and Marco Pistore. Web service composition as planning, revisited: In between background theories and initial state uncertainty. In Proceedings of AAAI, 2007.

[LI et al., 2007] Chu Min Li, Felip Manya, and Jordi Planes. New inference rules for Max-SAT. Journal of Artificial Intelligence Research, 30:321–359, October 2007.

[Lindsay et al., 2017] Alan Lindsay, Jonathon Read, João F. Ferreira, Thomas Hayton, Julie Porteous, and Peter Gregory. Framer: Planning models from natural language action descriptions. In ICAPS, pages 434–442, 2017.

[Pasula et al., 2007] Hanna M. Pasula, Luke S. Zettlemoyer, and Leslie Pack Kaelbling. Learning symbolic models of stochastic domains. J. Artif. Intell. Res., 29:309–352, 2007.

[Roni Stern, 2017] Brendan Juba Roni Stern. Efficient, safe, and probably approximately complete learning of action models. In IJCAI-17, pages 4405–4411, 2017.

[Yang et al., 2007] Qiang Yang, Kangheng Wu, and Yunfei Jiang. Learning action models from plan examples using weighted MAX-SAT. Artif. Intell., 171(2-3):107–143, 2007.

[Zettlemoyer et al., 2005] Luke S. Zettlemoyer, Hanna Pasula, and Leslie Pack Kaelbling. Learning planning rules in noisy stochastic worlds. In AAAI, pages 911–918, 2005.

[Zhuo and Kambhampati, 2013] Hankz Hankui Zhuo and Subbarao Kambhampati. Action-model acquisition from noisy plan traces. In IJCAI, pages 2444–2450, 2013.

[Zhuo and Kambhampati, 2017] Hankz Hankui Zhuo and Subbarao Kambhampati. Model-lite planning: Case-based vs. model-based approaches. Artif. Intell., 246:1–21, 2017.

[Zhuo et al., 2010] Hankz Hankui Zhuo, Qiang Yang, Derek Hao Hu, and Lei Li. Learning complex action models with quantifiers and logical implications. Artif. Intell., 174(18):1540–1569, 2010.

[Zhuo et al., 2014] Hankz Hankui Zhuo, Héctor Muñoz-Avila, and Qiang Yang. Learning hierarchical task network domains from partially observed plan traces. Artif. Intell., 212:134–157, 2014.