

Covering Landmark Interactions for Semantically Diverse Plans

Daniel Bryce
SIFT, LLC.
dbryce@sift.net

Renée C. Bryce
University of North Texas
rennee.bryce@unt.edu

Abstract

Prior approaches to generating diverse plans in domain-independent planning seek out variations on plan structure such as actions or causal links used, or states entered. As a result, these approaches can achieve great plan set diversity by synthesizing unnecessarily long plans. This type of misleading diversity may be useful if, for example, the plans are used as training data for a learner; however, they have arguably low value to a human decision maker.

We present an approach to domain-independent diverse planning that systematically varies the semantic attributes of plans so that we do not arbitrarily inflate plan length to achieve diversity. Our contribution is based upon the fact that landmarks, which represent the minimally necessary subgoals (semantics) of a planning domain, can be disjunctive and hence satisfied in a number of ways. Varying the disjuncts that must be satisfied by alternative plans leads to a form of diversity that does not encourage irrelevant plan structure. We present an extension of the LAMA planner called DLAMA that generates multiple plans, each required to systematically satisfy alternative landmark disjuncts. We show that, in comparison with prior diverse planners, DLAMA reduces average plan length while achieving plan set diversity.

Introduction

Diverse planning is an important tool for decision support scenarios where a human analyst wants to understand the space of possible plans or has difficulty specifying the planning domain model that exactly matches their application (a similar motivation underlies information retrieval systems that return multiple results (Carbonell and Goldstein 1998)). Diversity is measured by the average distance between plans in a plan set, and many domain-independent metrics have been explored. We find a common deficiency among the distance measures that is best illustrated by considering, not how plans should be different, but how they are the same. Nominally, every plan must be applicable in the initial state and satisfy the goal. Furthermore and by definition, each plan must satisfy every landmark. We note that landmarks are often disjunctive, meaning that alternative plans can satisfy a landmark in different ways (i.e., by satisfying a different disjunct). The deficiency noted above is that all prior

Copyright © 2013, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

diverse planners may fail to diversify the most essential aspects of a planning domain – landmarks.

We argue that the tuples of landmark disjuncts satisfied by a plan semantically distinguish it from other plans. By considering solely landmarks, our distance measure will not be sensitive syntactic properties of plans that introduce deceptive diversity. For example, applying and reversing an action might increase the number of unique actions in a plan from those in the the plan set but only *cosmetically* changes the nature of the plan. In contrast, varying which disjunct of a landmark is satisfied represents a difference in the *essential* nature of a plan. As such we develop a new measure of plan distance that is based upon landmark disjuncts it satisfies.

We evaluate a modified version of the LAMA planner (Richter and Westphal 2010), called DLAMA, on a set of planning tasks and compare it with FF^{GrDiv} (Coman and Muñoz-Avila 2011) and LPG-d (Nguyen et al. 2012). Our evaluation shows that plan set diversity can be attained without creating overly long plans that include different non-essential plan structure.

Plan Distance

We explore two distances between a pair of plans: action set distance, and landmark interaction distance. Action set distance, as defined in prior work (Coman and Muñoz-Avila 2011), is the cardinality of the symmetric difference between two plans' action sets:

$$d_{act}(\pi_1, \pi_2) = |(A(\pi_1) \cup A(\pi_2)) \setminus (A(\pi_1) \cap A(\pi_2))|$$

where $A(\pi)$ is the set of actions in plan π . The landmark interaction distance is defined similarly as the cardinality of the symmetric difference of the sets of t -strength landmark interactions satisfied by each plan:

$$d_{L_t}(\pi_1, \pi_2) = |(L_t(\pi_1) \cup L_t(\pi_2)) \setminus (L_t(\pi_1) \cap L_t(\pi_2))|$$

where $L_t(\pi)$ is a set of t -strength landmark interactions.

A t -strength landmark interaction is a set of t landmark disjuncts, each a literal appearing in a different disjunctive landmark. Thus, the possible 1-strength landmark interactions correspond to the literals appearing in each disjunctive landmark. The possible 2-strength landmark interactions are each pair of literals occurring in different landmarks. Measuring interactions, above and beyond counting individual landmark disjuncts, signifies semantic differences between

plans. Increasing the interaction strength considered magnifies the differences between plans.

Example of Landmark Interaction Distance

Consider an example from the first depot domain instance (Long and Fox 2003), where we have the following disjunctive landmarks, as found by the LAMA planner (Richter and Westphal 2010):

$$\begin{aligned} & \text{in}(c1, tr0) \vee \text{in}(c1, tr1) \\ & \text{in}(c0, tr0) \vee \text{in}(c0, tr1) \\ & \text{at}(tr1, dist0) \vee \text{at}(tr0, dist0) \end{aligned}$$

Assume that plans π_1 , π_2 , and π_3 satisfy the disjuncts:

$$\begin{aligned} \pi_1 : & \text{in}(c1, tr0), \text{in}(c0, tr0), \text{at}(tr0, dist0) \\ \pi_2 : & \text{in}(c1, tr1), \text{in}(c0, tr1), \text{at}(tr0, dist0) \\ \pi_3 : & \text{in}(c1, tr0), \text{in}(c0, tr1), \text{at}(tr0, dist0) \end{aligned}$$

The plan π_1 satisfies the 1- and 2-strength interactions

$$L_1(\pi_1) = \{\text{(in}(c1, tr0)\text{)}, \text{(in}(c0, tr0)\text{)}, \text{(at}(tr0, dist0)\text{)}\}$$

$$\begin{aligned} L_2(\pi_1) = & \{\text{(in}(c1, tr0)\text{), in}(c0, tr0)\text{,} \\ & (\text{in}(c1, tr0), \text{at}(tr0, dist0)), \\ & (\text{in}(c0, tr0), \text{at}(tr0, dist0))\} \end{aligned}$$

The size of the symmetric difference of 1- and 2-strength landmark interactions among the plans are the distances

d_{L_1}	π_1	π_2	π_3	d_{L_2}	π_1	π_2	π_3
	0	4	2		0	6	4
π_1	0	4	2		6	0	4
π_2	4	0	2		4	4	0
π_3	2	2	0				

where there is a total of six possible 1-strength landmark interactions and twelve 2-strength interactions.

DLAMA

DLAMA extends LAMA to find a diverse set of plans by constructing a t -strength covering array (Fisher 1992; Bryce, Colbourn, and Cohen 2005) for the set of disjunctive landmarks. In software testing, t -strength covering arrays represent alternative configurations of program inputs that collectively guarantee all t -way settings of program inputs are tested. In this setting, the covering array provides a similar guarantee by selecting landmark disjuncts that must be satisfied by each plan. For each configuration of the landmark disjuncts, we modify (mask) the LAMA landmark graph and find a plan that must satisfy the given disjuncts.

Similar to the set cover problem, finding a minimal covering array implies that the alternative configurations of the parameters (landmarks) satisfy highly different sets of t -way interactions. We use a greedy algorithm (Bryce, Colbourn, and Cohen 2005) to construct the covering array that selects each new configuration to cover the most new t -way interactions in each step. Thus, if we invoke LAMA on the configurations in the order found by the greedy algorithm (and find a plan) we are likely to find distant plans. We can also parallelize DLAMA quite easily because the configurations are independent. We note that it is possible to find plans with high landmark distance by extending the plans to satisfy additional landmark interactions that are not required, but doing so increases plan length.

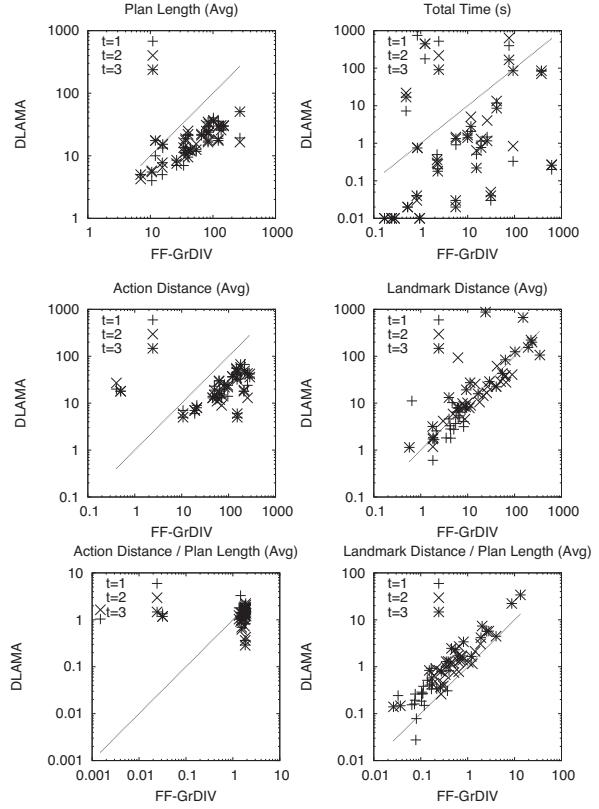


Figure 1: Plan set statistics.

Empirical Results

We compare DLAMA with FF^{GrDIV} and LPG-d on instances from the driverlog, storage, and satellite domains where each planner could generate four plans in 30 minutes and 2 GB of memory. Each instance has several disjunctive landmarks (as found by LAMA), rendering DLAMA applicable. The scatterplots in Figure 1 compare DLAMA with FF^{GrDIV} (LPG-d results are not shown because the trends are identical to that of FF^{GrDIV} , aside from a better overall run time) with respect to three landmark interaction strengths $t \in \{1, 2, 3\}$ used to compute landmark distance and parameterize DLAMA. We see that FF^{GrDIV} has a considerably higher average plan length (among the plans in its plan set) on most instances and generally takes longer as compared with DLAMA. The average action set distance between plans generated by FF^{GrDIV} is generally higher and the landmark interaction distance is more comparable, but often higher with FF^{GrDIV} – generating longer plans allows FF^{GrDIV} to satisfy more landmarks in each plan. FF^{GrDIV} is superior in terms of diversity, but at the expense of plan length. If we discount the diversity by dividing it by plan length (as shown by the lower plots), then we see that both planners are comparable in terms of action sets, but DLAMA is superior in terms of landmarks.

The trend is that diversity can be increased by inflating plan length. By normalizing with respect to plan length, landmark interaction distance is much greater when systematically varying the landmarks satisfied in DLAMA.

References

- Bryce, R. C.; Colbourn, C. J.; and Cohen, M. B. 2005. A framework of greedy methods for constructing interaction test suites. In Roman, G.-C.; Griswold, W. G.; and Nuseibeh, B., eds., *International Conference on Software Engineering*, 146–155. ACM.
- Carbonell, J., and Goldstein, J. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 335–336. ACM.
- Coman, A., and Muñoz-Avila, H. 2011. Generating diverse plans using quantitative and qualitative plan distance metrics. In Burgard, W., and Roth, D., eds., *National Conference on Artificial Intelligence (AAAI-11)*. AAAI Press.
- Fisher, R. 1992. The arrangement of field experiments. In *Breakthroughs in Statistics, Vol II*. J. Wiley and Sons, Inc. Originally published in 1926.
- Long, D., and Fox, M. 2003. The 3rd international planning competition: Results and analysis. *J. Artif. Intell. Res. (JAIR)* 20:1–59.
- Nguyen, T. A.; Do, M. B.; Gerevini, A.; Serina, I.; Srivastava, B.; and Kambhampati, S. 2012. Generating diverse plans to handle unknown and partially known user preferences. *Artif. Intell.* 190:1–31.
- Richter, S., and Westphal, M. 2010. The LAMA planner: Guiding cost-based anytime planning with landmarks. *J. Artif. Intell. Res. (JAIR)* 39:127–177.