

# Generating Diverse Plans Using Quantitative and Qualitative Plan Distance Metrics

Alexandra Coman, Héctor Muñoz-Avila

Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015  
alc308@lehigh.edu | hem4@lehigh.edu

## Abstract

Diversity-aware planning consists of generating multiple plans which, while solving the same problem, are dissimilar from one another. Quantitative plan diversity is domain-independent and does not require extensive knowledge-engineering effort, but can fail to reflect plan differences that are relevant to users. Qualitative plan diversity is based on domain-specific characteristics, thus being of greater practical value, but may require substantial knowledge engineering. We demonstrate a domain-independent diverse plan generation method that is based on customizable plan distance metrics and amenable to both quantitative and qualitative diversity. Qualitative plan diversity is obtained with minimal knowledge-engineering effort, using distance metrics which incorporate domain-specific content.

## Introduction

Diversity-aware problem solving involves generating two or more distinct solutions to the same problem. In planning, solution diversity has practical value for multiple domains (e.g. military planning, route planning, intrusion detection), particularly in mixed-initiative planning environments (Myers, 2002), where awareness of available alternatives is particularly useful. Sets of diverse plans cover a large portion of the solution space, providing a good indication of the range of available possibilities, and potentially highlighting solutions that may otherwise not be considered (Myers and Lee, 1999).

Previous approaches to diverse plan generation can be seen as belonging to one of two categories: *qualitative* and *quantitative*.<sup>1</sup> Quantitative plan diversity is domain-independent and has the advantage of not requiring extensive domain knowledge. However, the plan differences it reflects may be irrelevant to users. Qualitative plan diversity is based on domain-specific knowledge, thus reflecting significant differences that a human expert might take into account when comparing two plans (Myers and Lee 1999).

---

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup> We adopt the terminology used by Myers and Lee (1999).

Our objective is to generate qualitatively diverse plans with minimal additional knowledge requirements aside from the usual domain transition model.

One possible approach to obtaining plan diversity is by modifying heuristic search planners to add diversity criteria to the plan generation process, thus balancing the competing needs of plan generation efficiency and plan diversity (Srivastava et al. 2007). We show how such approaches can be used to obtain not only quantitative, but also qualitative plan diversity, through the use of customizable plan distance metrics which, when the aim is to obtain qualitative diversity, incorporate the minimal domain-specific content required for differentiating between plans, thus eliminating the need for a comprehensive domain model and its associated knowledge-engineering effort.

We demonstrate this approach on 4 planning domains: 3 preexisting, synthetic domains and a real-time strategy game domain. In the latter case, we test the diversity of the generated plans by running them in the game environment and assessing the variation of game scores thus obtained.

## The Plan Diversity Problem

Let  $D: \Pi \times \Pi \rightarrow [0, \infty)$  be a plan distance metric (a measure of the dissimilarity between two plans). For a non-empty set of plans  $\Pi$ , let the plan-set diversity  $Div(\Pi)$ , and the relative diversity  $RelDiv(\pi, \Pi)$  of a plan  $\pi$  relative to  $\Pi$ , be defined as:

$$Div(\Pi) = \frac{\sum_{\pi, \pi' \in \Pi} D(\pi, \pi')}{|\Pi| \times (|\Pi| - 1)} \quad (1)$$

$$RelDiv(\pi, \Pi) = \frac{\sum_{\pi' \in \Pi} D(\pi, \pi')}{|\Pi|} \quad (2)$$

These definitions are similar to those used in case-based reasoning by Smyth and McClave (2001). Myers and Lee (1999) use a definition similar to Equation 1 for the “dispersion” of sets of plans.

The maximal plan diversity problem can be defined as follows:

**Definition (maximal plan diversity problem).** Given a plan distance metric  $D$ , a parameter  $k$ , and a new problem  $P$ , obtain a collection of  $k$  solution plans  $\Pi$  solving  $P$ , such that no other set of  $k$  solution plans  $\Pi'$  solving  $P$  exists such that  $Div(\Pi) < Div(\Pi')$ .

Finding a set  $\Pi$  of  $k$  solution plans with maximum  $Div(\Pi)$  can be impossible, depending on how  $D$  is defined. Let us assume that we are attempting to find two maximally distant plans (i.e.,  $k = 2$ ). If we define  $D = D_{\text{stability}}(\pi, \pi')$  as the number of actions in  $\pi$  not in  $\pi'$  plus the number of actions in  $\pi'$  not in  $\pi$  (Fox et al. 2006)<sup>2</sup>, then the generation of maximally diverse sets of plans can diverge (e.g., in the logistics transportation domain, one could repeatedly move a truck back and forth between two locations, thereby, with every repetition, making the plan more distant to another one not containing these movement actions).

Therefore, we propose to solve a relaxation of this problem, where a distance metric  $D$  is taken into account during plan set generation, but maximal plan diversity is not guaranteed. We refer to this as the **plan diversity problem**.

## Quantitative and Qualitative Plan Distance

The distance metric  $D$ , on which the diversity  $Div$  of a set of plans is based, can be either quantitative or qualitative.

*Quantitative plan distance* is based on plan elements (e.g. actions) derivable from the domain transition model, and not interpreted in any domain-specific way. It follows that any two distinct plan elements are considered equally distant from one another (e.g. in an ancient warfare reenactment domain, the action of attacking using a battle ram is equally distant from the action of attacking using a catapult and the action of building fortifications). For example,  $D_{\text{stability}}$ , as described above, is a quantitative distance metric. Two plans which, due to minimal action overlap, are identified as distant using  $D_{\text{stability}}$  could appear very similar to a human expert.

*Qualitative plan distance* is based on interpretation, using domain knowledge, of the components of plans (e.g. battle rams and catapults are both siege engines; using siege engines is an offensive measure; building fortifications is a defensive measure). It follows that the qualitative distance between two distinct plan elements can be zero (e.g. the actions of attacking with a battle ram and attacking with a catapult might be deemed similar enough to be considered identical for the purposes of diverse plan generation) or any non-zero value, based on any number of domain-specific criteria. As multiple bases for qualitative distance can be defined for the same domain, it is possible to vary the set of features along which one would like to see diversity (e.g. in a travel domain, variation of ticket cost, but not means of transportation). Obtaining qualitative plan diversity requires domain-specific knowledge to be en-

coded and utilized. Previously, this was achieved by Myers and Lee by requiring that the domain transition model be supplemented by a “metatheory”: an extended description of the planning domain in terms of high-level attributes, based on which plans can be compared qualitatively.

We propose obtaining both quantitative and qualitative plan diversity based solely on the domain transition model and quantitative/qualitative plan distance metrics. Qualitative distance metrics should incorporate only the minimal domain-specific content that is required for the purposes of differentiating plans discerningly, rendering a comprehensive qualitative model of the domain unnecessary for the purposes of obtaining plan diversity.

## Generating Diverse Plans

Our diverse plan generation method enhances heuristic search planning with diversity techniques emulating those used by Smyth and McClave (2001) for obtaining solution diversity in case-based reasoning. While the case-based-reasoning approach to problem solving generally involves retrieving library cases on the basis of maximal similarity to a problem, the additional retrieval criterion of case diversity helps ensure that the retrieved results provide valuable alternatives (Smyth and McClave 2001, McSherry 2002, McGinty and Smyth 2003). We address the challenge of balancing the planner’s own heuristic function (which ensures planning speed) with diversity considerations in a manner similar to that used in case-based reasoning to handle the trade-off between similarity to the problem and diversity within the retrieved case set.

## Diversity-aware Retrieval in Case-Based Reasoning

*Case Diversity Greedy Selection* (Smyth and McClave, 2001) retrieves a set of  $k$  cases as follows: first, it automatically adds to the retrieved set the case that is maximally similar to the new problem. Then, for  $k-1$  steps, it retrieves the case that maximizes a metric taking into account both the similarity to the new problem and the relative diversity to the set of solutions selected so far (where the relative diversity is based on a distance metric that is the inverse of the similarity metric).

---

### CaseDiversityGreedySelection( $P, CL, k$ )

```

1:  $S \leftarrow \{\}$ ;  $R \leftarrow CL$ 
2: Sort  $R$  by  $Sim(c, P)$ 
3:  $S \leftarrow S + \text{first}(R)$ ;  $R \leftarrow R - \text{first}(R)$ 
4: Repeat
5:   Sort  $R$  by  $BalancedCase(c, P, S)$ 
6:    $S \leftarrow S + \text{first}(R)$ 
7:    $R \leftarrow R - \text{first}(R)$ 
8: Until  $|S| = k$  library cases have been retrieved
9: Return  $S$ 

```

---

In the pseudocode above,  $P$  is the problem,  $CL$  the case library,  $c$  an individual case in  $CL$ ,  $k$  the number of cases to

<sup>2</sup> This metric is used by Fox et al. (2006) in the context of plan stability (obtaining repaired plans which are similar to a source plan).

be retrieved,  $S$  the set of retrieved cases, and  $R$  the current set of library cases that are candidates to be added to the solution set.  $Sim(c, P)$  is the similarity of case  $c$  to problem  $P$ , while  $BalancedCase(c, P, S)$  is a metric that takes into account both the similarity of  $c$  and  $P$  and the relative diversity between  $c$  and  $S$ .

## Heuristic Search Planning

In this type of planning, heuristic functions are used to assess candidate refinements of partial solutions, as in the highly abstracted pseudocode below.  $PL$ ,  $h_{PL}$ , and  $P$  are the heuristic search planner, its heuristic function, and the problem, respectively. The call  $getCandidates(P, \pi)$  returns the set of possible one-step refinements for the partial plan  $\pi$ . The call  $selectCandidate(C, h_{PL})$  chooses the candidate refinement that is ranked highest based on  $h_{PL}$ .  $h_{PL}$  is usually computed based on the solution plan of a relaxed version of the planning problem.

---

### GeneratePlan( $PL, P$ )

```

1:  $\pi \leftarrow$  empty-plan
2: Repeat
3:    $C \leftarrow$  getCandidates( $P, \pi$ )
4:    $\pi \leftarrow$  selectCandidate( $C, h_{PL}$ )
5: Until  $\pi$  is a solution for  $P$ 
6: Return  $\pi$ 

```

---

The FF (Hoffmann and Nebel 2001) forward state-space planner uses a goal-distance planning heuristic: the heuristic value of a candidate state is the length of the solution plan, obtained by running the Graphplan planner (Blum and Furst 1997) with the candidate state as initial state, of a relaxation of the planning problem. The relaxed problem is obtained by eliminating all negative effects of the actions (Bonet, Loerincs, and Geffner 1997), thereby guaranteeing that a solution can be found in polynomial time. FF’s primary search type, *Enforced Hill-Climbing* (using a filter called *Helpful Actions* for preliminary action pruning) is efficient, but not complete. When it fails to produce a result (which occurs if dead-end states are reached), FF switches to complete search, with no action pruning.

## Diversity-aware Heuristic Search Planning

Below, we describe the algorithm *GreedyDiversePlanSet*, a combination of *GeneratePlan* and *CaseDiversityGreedySelection*, which generates a set of  $k$  diverse plans. Initially, it uses the heuristic search planner to generate the first plan in the set. Then, the modified, diversity-aware version of the planner is run  $k-1$  times, each time generating a plan by using two criteria for candidate selection: a plan-diversity metric (estimating the relative diversity between a candidate plan and the plan set  $\Pi$  generated so far), and the planner’s regular heuristic.  $PL$  is the heuristic search planner,  $h_{mixed}$  a heuristic function that evaluates a candidate partial plan (Equation 3),  $P$  the problem and  $k$  the number of diverse plans to be generated. The call  $generatePlan(PL, P)$  runs an unmodified version of the planner on problem

$P$ , while  $getCandidates(P, \pi)$  returns the set of possible one-step refinements for the partial plan  $\pi$ .  $selectCandidate(C, h_{mixed}, \Pi)$  chooses a candidate partial plan that maximizes  $h_{mixed}$ , instead of one that minimizes the regular planning heuristic,  $h_{PL}$ .

---

### GreedyDiversePlanSet( $PL, h_{mixed}, P, k$ )

```

1:  $\Pi \leftarrow \{\}$ ;  $\pi \leftarrow$  generatePlan( $PL, P$ ); Add  $\pi$  to  $\Pi$ 
2: Repeat
3:    $\pi \leftarrow$  BalancedPlan( $PL, h_{mixed}, P, \Pi$ )
4:   Add  $\pi$  to  $\Pi$ 
5: Until  $|\Pi| = k$  plans have been generated
6: Return  $\Pi$ 

```

---

### BalancedPlan( $PL, h_{mixed}, P, \Pi$ )

```

1:  $\pi \leftarrow$  empty-plan
2: Repeat
3:    $C \leftarrow$  getCandidates( $P, \pi$ )
4:    $\pi \leftarrow$  selectCandidate( $C, h_{mixed}, \Pi$ )
5: Until  $\pi$  is a solution for  $P$ 
6: Return  $\pi$ 

```

---

$$h_{mixed}(\pi, \Pi) = -\alpha h_{PL}(\pi) + (1 - \alpha) h_{diversity}(\pi, \Pi) \quad (3)$$

where  $h_{diversity}$  is defined as follows:

$$h_{diversity}(\pi, \Pi) = RelDiv(\pi_{relax}, \Pi) \quad (4)$$

The plan  $\pi_{relax}$  is a relaxed solution of  $P$ , that augments  $\pi$ . It is constructed based on the relaxed plan  $\pi_{relaxPL}$ , produced by  $PL$ . With FF, for example,  $\pi_{relax}$  can be obtained by appending  $\pi_{relaxPL}$  (i.e. the sequence of actions in the relaxed domain leading from the current state to the goal state) to the current partial plan  $\pi$  (i.e., the sequence of actions chosen so far, leading from the initial state to the current state). The regular planning heuristic ( $h_{PL}$ ) is subtracted from the diversity metric ( $h_{diversity}$ ) because one must seek to minimize  $h_{PL}$  (an estimate of the effort required for finding a solution), and to maximize  $h_{diversity}$  (an estimate of the diversity of the generated plan set).  $\alpha$  is a parameter used for varying the complementary weights assigned to the two criteria.

## Experimental Evaluation

We implement FF<sup>GrDiv</sup>, a diversity-aware version of the JavaFF (Coles et al. 2008) implementation of FF. FF<sup>GrDiv</sup> was created by modifying JavaFF as indicated by *GreedyDiversePlanSet*, where  $h_{PL}$  is the standard FF heuristic.

We test FF<sup>GrDiv</sup> on 4 domains: the first 3 are synthetic, while the 4th is a real-time strategy game domain. In our experiments, we use two types of distance metrics to compute *RelDiv* (Equation 2). For the synthetic domains, we use a quantitative distance metric. For the Wargus domain, we use both a quantitative and a qualitative distance metric (the abbreviations *Quant* and *Qual* are used to distinguish between FF<sup>GrDiv</sup> variants using these types of metrics).

As a baseline, we use  $\varepsilon$ -Greedy FF, a slightly modified version of JavaFF, which generates sets of  $k$  plans by occasionally injecting random diversity: whenever choosing between candidate states,  $\varepsilon$ -Greedy FF will, with probability  $(1-\varepsilon)$ , pick a candidate state randomly. In all other cases, it will behave as JavaFF would (picking the state with the best heuristic value).

## Planning Domains

The 3 synthetic domains we use are DriverLog, Depots, and Rovers, which are available with the JavaFF distribution. For obtaining diversity in these domains, we use the quantitative plan distance metric  $D_{\text{stability}}$ , discussed above.

The fourth domain is Wargus, a real-time, non-deterministic action outcome, partial-state observable, adversarial strategy game. Wargus exhibits many of the characteristics of real domains of practical interest, and is used herein to highlight the value of qualitative plan diversity for such domains. It should be stressed that our aim is not to produce plans demonstrating expert-gamer-like behavior, but to create game sessions that are diverse, providing a varied sample of possible tactical approaches to the game. This can, for example, be of practical value in the modeling of AI enemies, which, to make the game environment realistic and engaging, should vary in intelligence and ability. The games take place on a small-sized map (32x32 tiles). By restricting the size of the map and the types of units used, we purposely do not allow for considerable intrinsic game variation. Plans indicate what units in a team should do when competing against the built-in Wargus enemy AI. Units can “move” from a current location to an indicated one, “attack” any enemies at an indicated location on the map or “guard” a location (attacking any enemy that comes within a certain range). A restriction for these actions is that no two units can be occupying the same map location at the same time. Because of the size of the map and the number of units involved, experiments are restricted to five problems, corresponding to game scenarios that are significantly different from one another. Problems indicate the available friendly unit armies and a number of map locations (waypoints which the units can visit), a subset of which are “attackable” (they can be the target of an “attack” action).

$$D_{\text{wargus}}(\pi_1, \pi_2) = \begin{cases} 1, \text{attackUnitsType}(\pi_1) \neq \\ \text{attackUnitsType}(\pi_2) \\ 0, \text{attackUnitsType}(\pi_1) = \\ \text{attackUnitsType}(\pi_2) \end{cases} \quad (5)$$

For generating diverse Wargus plans, we define a qualitative distance metric,  $D_{\text{wargus}}$ , which reflects domain knowledge: a relevant characteristic setting plans apart is the type of units used for attacking, as different units have their specific strengths and weaknesses (e.g. an archer is adept at long range attacks, but weak in close combat), and their losses lead to different score penalties. Given this, the qualitative plan distance metric  $D_{\text{wargus}}(\pi_1, \pi_2)$  is defined as

in Equation 5, where  $\text{attackUnitsType}(\pi)$  is the type of units in the attacking army of plan  $\pi$ .

For the Wargus domain, we suppress the Helpful Actions filter employed by JavaFF (replacing it with the Null Filter, which does not perform preliminary action pruning), for both  $\text{FF}^{\text{GrDiv}}$  and  $\varepsilon$ -Greedy FF. The Helpful Actions filter only considers a limited subset of the applicable actions, potentially making it impossible to obtain qualitatively diverse plans, if the actions required for doing so are not in the subset in question. In Equation 3, we assign  $\alpha = 0.8$  (thus giving more weight to the FF heuristic) for the synthetic domains, in order to increase the chances of generating a solution, as we found empirically that, with lower values of  $\alpha$ ,  $\text{FF}^{\text{GrDiv}}$  generates solutions for fewer problems. For Wargus, we assign a weight of  $\alpha = 0.55$ , which is sufficient to generate solutions for all problems, and still ensure that these solutions are diverse.

## Evaluation Methods

For the synthetic domains, we analyze the diversity of the plan sets (generated with  $\text{FF}^{\text{GrDiv}}$  and  $\varepsilon$ -Greedy FF,  $\varepsilon = 0.99; 0.8; 0.7$ ) by computing the values of the diversity metric  $Div$  (Equation 1) for the plans generated (using  $D = D_{\text{stability}}$ ). The results are not redundant because, during plan generation,  $\text{FF}^{\text{GrDiv}}$  uses the relaxed plans as a heuristic estimate, whereas, in the evaluation, we compute the diversity of the actual set of generated plans.

In the Wargus domain, we test the diversity of the generated plans by running them in the game and observing the variation of the built-in Wargus score (which reflects damage inflicted and incurred). We compare  $\text{FF}^{\text{GrDiv}}$  Qual ( $D = D_{\text{wargus}}$ ) with  $\varepsilon$ -Greedy FF, and with  $\text{FF}^{\text{GrDiv}}$  Quant ( $D = D_{\text{stability}}$ ).

In order to assess  $\varepsilon$ -Greedy FF and  $\text{FF}^{\text{GrDiv}}$  on equal terms, out of the problems in each domain, we only report the results obtained on those which, using both planner variants, were able to produce a complete set of plans (i.e. did not run out of memory) and did so using only Enforced Hill Climbing (not complete search), as the use of a different algorithm may, by itself, influence plan diversity (e.g. if two plans in a set are generated using different algorithms, they are more likely to differ), potentially creating bias in favor of one or the other of the compared algorithms.

## Experimental Results

For all three synthetic domains, the diversity of the plan sets generated using  $\text{FF}^{\text{GrDiv}}$  is, in most instances, greater than that of plan sets obtained using the three  $\varepsilon$ -Greedy FF variants, while plan generation time is comparable, as can be seen in Figure 1: each point indicates the average of the diversity or planning time (as indicated by the y-axis label) over 4 planning sessions (with  $k=4$ ) on one domain problem. It should be noted that, as  $\varepsilon$  decreases,  $\varepsilon$ -Greedy FF produces increasingly greater diversity (because more random choices are being made), but the number of failed planning attempts also increases. This causes 0.7-Greedy

FF to repeatedly fail on the last two problems in the DriverLog domain, never producing results (hence, the two missing data points in the DriverLog section of Figure 1). Increasing diversity by greatly increasing  $\epsilon$  is, therefore, not a feasible approach.

In-game results for the Wargus domain (Figure 2) indicate that  $\text{FF}^{\text{GrDiv}}$  Qual generates sets of plans containing, on average, more distinct values than both  $\text{FF}^{\text{GrDiv}}$  Quant and 0.7-Greedy FF. In Figure 2, each point indicates the average game score for 4 game runs of the same plan (the score's fluctuation is shown in the error bars). For Problem 1,  $\text{FF}^{\text{GrDiv}}$  Qual produces 3 out of 4 distinct scores on the first and third plan sets and 4 distinct scores on the second one, while  $\text{FF}^{\text{GrDiv}}$  Quant produces 2 out of 4 distinct scores on the first two sets and 3 out of 4 distinct scores on the third set. For Problem 2,  $\text{FF}^{\text{GrDiv}}$  Qual produces 3 out of 4 distinct scores on the first two plan sets and 2 out of 4 distinct scores on the last set, while  $\text{FF}^{\text{GrDiv}}$  Quant produces 2 out of 4 distinct scores on all three sets. For Problem 3,  $\text{FF}^{\text{GrDiv}}$  Qual produces three maximally diverse sets of scores (4 distinct scores per plan set), while  $\text{FF}^{\text{GrDiv}}$  Quant produces 2 out of 4 distinct scores on the first set, 4 out of 4 distinct scores on the second set and 3 out of 4 distinct scores on the third set. For Problems 4 and 5,  $\text{FF}^{\text{GrDiv}}$  Qual produces 3 out of 4 distinct scores on all plan sets, while  $\text{FF}^{\text{GrDiv}}$  Quant produces 2 out of 4 distinct scores on all plan sets. 0.7-Greedy FF only produces 2 out of 4 distinct scores on the third plan set in Problem 3 and the third plan set in Problem 4, achieving no score diversity at all in any other plan set.

In addition to showing that both variants of  $\text{FF}^{\text{GrDiv}}$  outperform a randomized diverse plan generation method, the results suggest that the qualitative distance metric can, indeed, help generate plan sets of greater genuine diversity, as attested by their behavior in Wargus.

It should also be pointed out that, for each problem, a subset of  $\text{FF}^{\text{GrDiv}}$  Qual plans perform at least as well as any plans generated with  $\text{FF}^{\text{GrDiv}}$  Quant or 0.7-Greedy FF (as reflected in game scores). Furthermore, we have observed that  $\text{FF}^{\text{GrDiv}}$  Quant tends to produce plans of questionable quality, by inflating them with actions not necessary for reaching the goal state, added solely for the purpose of increasing the distance to the previously-generated set of plans. In contrast,  $\text{FF}^{\text{GrDiv}}$  Qual generally restricts itself to adding actions which are necessary for reaching the goal state. This is reflected in the lower average length of  $\text{FF}^{\text{GrDiv}}$  Qual plans, and suggests that a well-chosen qualitative plan distance metric may help ensure that the generated diverse plans are also of good quality. We plan to investigate this matter further in future work.

## Related Work

Myers and Lee (1999) have proposed a method for qualitative-diversity-aware plan generation. They use a domain metatheory to partition the plan space into  $k$  regions, then generate a plan for each such region. Quantitative plan diversity has been explored by Srivastava et al. (2007) in ge-

nerative planning, and by Coman and Muñoz-Avila (2010) in case-based planning. We explore both quantitative and qualitative plan diversity in generative planning, without requiring extensive knowledge engineering. To our knowledge, aside from Coman and Muñoz-Avila (2010) in case-based planning, no other work assesses plan diversity by running plans in their environment and observing the results thus obtained. Instead, assessment is conducted solely by analyzing the sets of plans themselves.

Outside planning, solution diversity has been explored extensively in case-based reasoning (Smyth and McClave 2001, McSherry 2002, McGinty and Smyth 2003), and by Hebrard et al. (2005) in constraint programming.

## Conclusions and Future Work

We investigate domain-independent diverse plan generation with heuristic search planners, using a method that is amenable to both quantitative and qualitative distance metrics.

Our work brings two main contributions to diverse generative planning. First, we obtain qualitative diversity solely through the use of a qualitative distance metric, without requiring a comprehensive domain model. Second, we evaluate the diversity of generated plans by running them in their intended environment and observing their performance.

In future work, we intend to explore qualitative plan diversity in various real domains of practical interest, as well as address the trade-off between plan set diversity and plan quality. We anticipate that, as the distance metrics become more complex, the required planning effort will increase, requiring optimization techniques for balancing planning efficiency and plan set diversity.

## Acknowledgements

This work was supported in part by NSF grant 0642882. We thank the anonymous reviewers for their comments.

## References

- Blum, A., and Furst, M. 1997. Fast Planning Through Planning Graph Analysis. *Artificial Intelligence* 90:281-300.
- Bonet, B.; Loerincs, G.; and Geffner, H. 1997. A Fast and Robust Action Selection Mechanism for Planning. In *Proceedings of the National Conference on Artificial Intelligence*, 714-719. AAAI Press/MIT Press.
- Coles, A. I.; Fox, M.; Long, D.; and Smith, A. J. 2008. Teaching Forward-Chaining Planning with JavaFF. *Colloquium on AI Education, Twenty-Third AAAI Conference on Artificial Intelligence*.
- Coman, A., and Muñoz-Avila, H. 2010. Case-based Plan Diversity. In *Proceedings of the Eighteenth International Conference on Case-Based Reasoning*, 66-80. Springer.

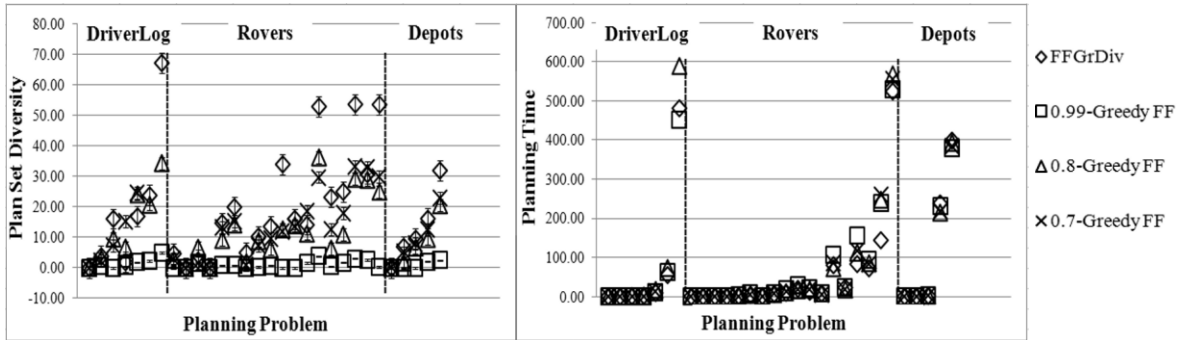


Figure 1. Synthetic domains: diversity (left) and planning time (right).

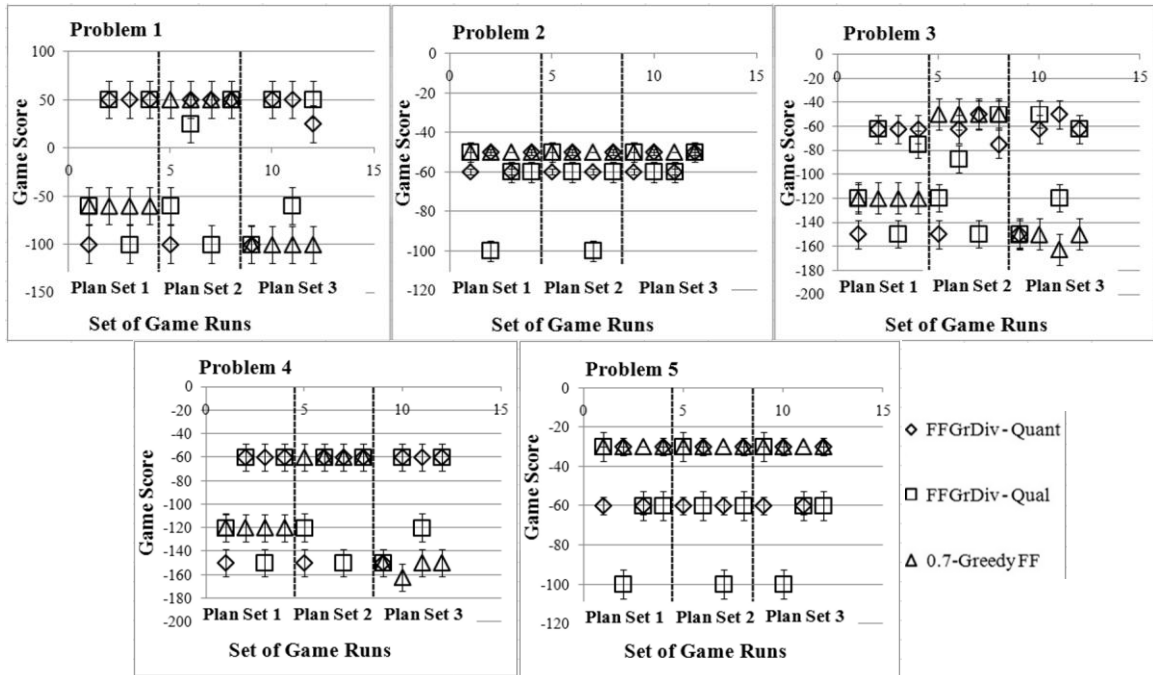


Figure 2. Wargus domain: game scores of sets of plans for the five test problems.

Fox, M.; Gerevini, A.; Long, D.; and Serina, I. 2006. Plan Stability: Replanning versus Plan Repair, In *Proceedings of the Sixteenth International Conference on Automated Planning and Scheduling*, 212–221.

Hebrard, E.; Hnich, B.; O’Sullivan, B.; and Walsh, T. 2005. Finding Diverse and Similar Solutions in Constraint Programming. In *Proceedings of The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference*, 372–377. AAAI Press/MIT Press.

Hoffmann, J., and Nebel, B. 2001. The FF Planning System: Fast Plan Generation Through Heuristic Search, *Journal of Artificial Intelligence Research*, 14:253–302.

McGinty, L., and Smyth, B. 2003: On the Role of Diversity in Conversational Recommender Systems. In *Proceedings of the Fifth International Conference on Case-Based Reasoning*, 276–290. Springer.

McSherry, D. 2002. Diversity-Conscious Retrieval. In *Proceedings of the Sixth European Conference on Case-Based Reasoning*, 219–233. Springer.

Myers, K. L., and Lee, T. J. 1999. Generating Qualitatively Different Plans through Metatheoretic Biases. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and Eleventh Conference on Innovative Applications of Artificial Intelligence*, 570–576. AAAI Press/MIT Press.

Myers, K. L.; Tyson, W. M.; Wolverson, M. J.; Jarvis, P. A.; Lee, T. J.; and desJardins, M. 2002. PASSAT: A User-centric Planning Framework. In *Proceedings of the 3rd Intl. NASA Workshop on Planning and Scheduling for Space*.

Smyth B., and McClave, P. 2001: Similarity vs. Diversity. In *Proceedings of the Fourth International Conference on Case-Based Reasoning*, 347–361. Springer.

Srivastava, B.; Kambhampati, S.; Nguyen, T.; Do, M.; Gerevini, A.; and Serina, I. 2007. Domain Independent Approaches for Finding Diverse Plans. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2016–2022.