

Qualitative vs. Quantitative Plan Diversity in Case-Based Planning

Alexandra Coman and Héctor Muñoz-Avila

Department of Computer Science and Engineering, 19 Memorial Drive West,
Lehigh University, Bethlehem, PA 18015
{alc308,hem4}@lehigh.edu

Abstract. Plan diversity has practical value in multiple planning domains, including travel planning, military planning and game planning. Existing methods for obtaining plan diversity fall under two categories: quantitative and qualitative. Quantitative plan diversity is domain-independent and does not require extensive knowledge-engineering effort, but can fail to reflect plan differences that are truly meaningful to users. Qualitative plan diversity is based on domain-specific characteristics which human experts might use to differentiate between plans, thus being able to produce results of greater practical value. However, the previous approach to qualitative plan diversity assumes the availability of a domain metatheory. We propose a case-based planning method for obtaining qualitative plan diversity through the use of distance metrics which incorporate domain-specific content, without requiring a domain metatheory. To our knowledge, this is the first time qualitative plan diversity is being explored in a case-based planning context.

Keywords: diversity, case-based planning, qualitative diversity, quantitative diversity, diversity metrics.

1 Introduction

Diversity-aware planning consists of generating two or more plans which, while solving the same problem, are dissimilar from one another, thus covering a large portion of the solution space, and providing a good indication of the range of available possibilities.

Plan diversity has practical value in multiple planning domains, including military planning [14] (e.g., offensive versus defensive plans, or defensive plan variants), travel planning [12] and route planning (e.g. using local roads versus using highways). In mixed-initiative planning environments [13], diverse plans can provide the user with genuine alternatives, potentially highlighting useful solutions that may otherwise not be considered. In plan-based intrusion-detection [1], they raise awareness of manifold threats. In game environments, plan diversity could be used to assist the player in exploring multiple different game-play strategies, as well as for modeling non-player characters exhibiting varied behavior, adding to the realistic atmosphere and enjoyment factor of the gaming experience [16].

Depending on the type of plan distance (the measure of the dissimilarity between two plans) on which they are based, previous approaches to diversity-aware planning can be seen as belonging to one of two categories: *qualitative* and *quantitative*.

Quantitative plan distance is domain-independent and has the advantage of not requiring domain knowledge aside from the usual domain transition model (e.g., a collection of STRIPS operators). The most common example of a quantitative distance metric is an action-set metric that counts the number of actions two plans do not have in common. This approach is, however, inflexible, as well as likely to produce misleading results: two plans identified as distant using a quantitative, action-set metric could, in essence, be similar (e.g. in combat-based games, two plans may have very little overlap in terms of the actions they execute, while being both implementations of a defensive strategy).

Qualitative plan distance, on the other hand, is based on domain-specific knowledge, thus having the potential to reflect subtler semantic differences that a human expert might take into account when comparing two plans (e.g. even if consisting of otherwise identical actions, a plan involving first-class air-travel will, from the point of view of a budget-conscious customer, be radically different from its economy-class counterpart). In contrast to the more mechanical quantitative approach, qualitative plan comparison should “see” plans much like human users would: as endeavors characterized by cost, risk, degrees of preference, etc. On the downside, the method for achieving qualitative plan diversity proposed previously in generative planning [12] is knowledge-intensive, requiring, in addition to the domain transition model, an extended domain theory (“metatheory”).

In our previous work [3], we took the first steps in exploring plan diversity in case-based planning [5,17,20], and did so using a quantitative approach. We now propose a case-based planning method for obtaining qualitative plan diversity without the need for a domain metatheory. This is achieved through the use of distance metrics which, themselves, incorporate the minimal domain-specific content that is required for the purposes of obtaining diversity.

We aim to show that, when run in a real environment, qualitatively-diverse plan sets can produce more varied results than quantitatively-diverse plan sets. To our knowledge, this is the first time qualitative plan diversity is explored in a case-based planning context.

In Section 2, we describe qualitative and quantitative plan distance in more detail. Then, in Section 3, we exemplify and compare quantitative and qualitative plan diversity in a planning domain. In Section 4, we present a case-based retrieval algorithm that is amenable to both quantitative and qualitative distance metrics. Section 5 is dedicated to the comparative experimental evaluation of the diversity of plans obtained using quantitative and qualitative distance metrics. Section 6 provides an overview of related work, followed by final remarks in Section 7.

2 Qualitative and Quantitative Plan Distance

Adapting the case diversity definition formulated by Smyth and McClave [19], we can define the diversity $Div(\Pi)$ of a set of plans Π as the average dissimilarity between pairs of plans in the set:

$$Div(\Pi) = \sum_{\pi, \pi' \in \Pi} \frac{Dist(\pi, \pi')}{\frac{|\Pi| \times (|\Pi| - 1)}{2}} \quad (1)$$

where the *plan distance* $Dist: \Pi \times \Pi \rightarrow [0, 1]$ is a measure of the dissimilarity between two plans. This formula is a direct adaptation, for the purposes of plan comparison, of the case diversity formula proposed by Smyth and McClave [19],¹ as $Dist$ can be considered to be the complement of a similarity measure Sim :²

$$Dist(\pi, \pi') = 1 - Sim(\pi, \pi') \quad (2)$$

It should immediately be pointed out that the problem of comparing plans is nontrivial: each plan may have an arbitrary number of actions, each with any number of parameters. Furthermore, the plan space for a given problem is potentially infinite (e.g., in a travel domain, a plan can be arbitrarily lengthened by repeatedly going back and forth between two locations). It follows that the notion of completeness [11], as defined for analysis tasks, such as recommender systems, cannot be applied to case-based planning (the set of possible solutions to planning problems is not limited to the contents of the case base, but includes all adapted plans which could possibly be obtained from those cases; and there are infinitely many such plans).

The types of plan distance used in generative planning fall under two categories, which we will be referring to as *quantitative* and *qualitative*.

Quantitative plan distance is based on plan elements (such as actions) derivable from the domain transition model,³ which are 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 a cooking domain, the action of adding lemon juice to a dish is considered equally distant from the action of adding vinegar and the action of adding sugar). Quantitative plan comparison, therefore, generally consists of counting the plan elements which plans have, or do not have, in common. An example of a quantitative distance metric (a normalized version of the metric used in [6]) is:

$$Dist_{Quant}(\pi_1, \pi_2) = 1 - \frac{common(\pi, \pi')}{\max(|\pi|, |\pi'|)} \quad (3)$$

where $common(\pi, \pi')$ is the number of actions that plans π and π' have in common and $|\pi|$ is the number of actions in plan π .

Qualitative plan distance is based on interpretation, using domain knowledge, of the components of plans (e.g. in a cooking domain: lemon juice and vinegar are both sour, but sugar is sweet; in a travel domain: a first-class plane ticket is expensive, while an economy one is affordable). 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

¹ The same formula is used in a generative planning context by Myers and Lee [12], under the name of “dispersion”.

² We are expressing diversity in terms of distance metrics. However, all formulas for distance metrics can be rewritten in terms of the complementary similarity metric. We maintain this interchangeability by always using normalized versions of the metrics, so that their values fall in the $[0, 1]$ interval.

³ The minimal domain theory required in planning.

means of transportation). This has a practical advantage over quantitative diversity. However, with the greater potential benefits of qualitatively diverse plan generation, comes the greater complexity of obtaining it. Unlike quantitative metrics, which are domain-independent, qualitative metrics require domain-specific knowledge to be encoded and utilized. Previously, this was achieved by Myers and Lee [12] in an HTN planning context. Their approach, however, involves considerable knowledge engineering effort: for the purposes of diverse plan generation, Myers and Lee require a “metatheory” providing additional domain information, thus allowing plans to be compared in terms of high-level features, such as the objects which fulfill various “roles” in plans and the domain-specific characteristics of various types of actions (e.g. the speed of travel by a given means of transportation).

We propose a method of obtaining qualitative diversity which requires neither an HTN planning context, nor a domain metatheory. Instead, it is based solely on the domain transition model, a case base of plans, and *qualitative distance metrics* which incorporate all the domain-specific, qualitative content that is required for the purposes of creating diversity.

Our approach is motivated by the observation that obtaining plan diversity does not require a comprehensive qualitative model of the domain. It is sufficient to “equip” the diversity metric with minimal knowledge regarding the selected features it should base its differentiation between plans on. It immediately follows that multiple qualitative metrics can be defined for any domain, each metric reflecting the minimal useful information necessary for a particular diverse-plan retrieval task. These metrics can then be used separately or compounded as needed, offering much greater power and flexibility in generating diverse plan sets that are truly useful in practical situations.

3 Qualitative and Quantitative Plan Diversity in a Real-Time Strategy Game Context

To exemplify possible uses of quantitative and qualitative plan diversity, we assume a real-time strategy game context, which is characterized by many of the complexities of real domains of practical interest: it is dynamic (the world state evolves while the agent deliberates), non-deterministic (no specific action outcome can be guaranteed), partially observable, and adversarial (agents in each team seek to maximize their performance metric by minimizing the opponents’ performance) [15].

Assume the following game configuration: the types of available units are peasants, soldiers, archers and mages. Units vary in terms of attack capabilities (e.g. soldiers are close-combat units, archers and mages long-range attack units) and robustness (e.g. peasants are very weak). The game score is computed by adding points for enemy kills and subtracting points for loss of friendly units. The amount added/subtracted on the destruction of a unit depends on the type of unit in question. The actions that can be taken by units are: *move* (the unit attempts to move to a specified location on the map), *patrol* (the unit moves back and forth between its current location and a specified location on the map) and *attack* (the unit attacks any enemies at a given location). The action signature is <action name (parameter1, parameter2)>, where *parameter1* specifies the unit which will undertake the action and *parameter2* specifies the target location of the action (e.g. action *Move(soldier1, loc1)* instructs

unit *soldier1* to move to the map location *loc1*). There are two teams, one controlled using our plans, the other controlled by the built-in enemy AI.

Consider the set of 3 plans in Fig. 1, and assume that we have already retrieved Plan 1 and are now trying to find a second plan, out of the two remaining ones, that is maximally distant from Plan 1, making the resulting pair of retrieved plans maximally diverse (using the diversity-aware retrieval algorithm described in Section 4).

Plan1: Move (soldier1, loc1), Move (soldier2, loc2), Move (mage1, loc3),
Attack (soldier3, loc4)
Plan 2: Attack (soldier2, loc4)
Plan 3: Move (soldier1, loc1), Move (soldier2, loc2), Move (mage1, loc3),
Attack (archer1, loc4)

Fig. 1. Sample plans for a real-time strategy domain. The action parameters specify the unit which will be undertaking the action and the map location at which the action will take place.

First, let us consider quantitative diversity (in our previous work [3], we demonstrated quantitative plan diversity in a real-time strategy game domain). To do so, we use the quantitative metric $Dist_{Quant}$ (Formula 3). As a result, the plan that is chosen is Plan 2: it shares no actions with Plan 1 (the *attack* actions in the two plans use distinct soldier units), therefore the distance between them is 1 (the maximum possible distance).

However, an informed analysis, using domain-specific information, of the individual actions yields significant information: an *attack* action indicates an offensive approach to the game; a more neutral *move* action could be interpreted in various ways: moving to a location on one's own side of the map may be considered a defensive action, while attempting to move towards the enemy side is likely offensive, indicating the intention to engage in battle. Therefore, Plans 1 and 2 may not be meaningfully different at all. They both culminate in an *attack* action at the same map location, using units, which, while distinct, are of the same type (soldiers). The three other actions that differentiate Plan 1 from Plan 2 may not be of great consequence at all, if the locations the units are moving to are on the friendly side of the map and not very far from their initial locations.

Let us now consider, instead, a qualitative distance metric which considers two plans maximally diverse if they attack using a different *type* of unit, and identical if they use units of the same type to attack, even if the units are distinct (we will be using a more elaborated variant of this metric in our experiments).

This method assesses Plan 2 as being maximally similar to Plan 1: they use units of the same type to attack, and the other actions in Plan 1 are ignored for the purposes of comparison, as they were not specified in the metric definition (this is an example of a qualitative metric including only the minimal amount of domain information that is relevant to the task at hand, thus reducing the knowledge engineering effort, and improving retrieval performance). As a result, the qualitative method picks the maximally distant Plan 3, which attacks using an archer, a unit very different from a

soldier: it is long-range, weaker in close combat, and its loss incurs a different score penalty than the loss of a soldier. This makes the selected plans significantly different relatively to the rules of the game.

4 Plan-Diversity-Aware Retrieval Algorithm

To demonstrate plan-diversity-aware case retrieval, we use a variant of the Greedy Selection⁴ algorithm proposed by Smyth and McClave [19] (Fig. 2). The algorithm retrieves a set of k diverse cases. 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 an evaluation metric taking into account both the similarity to the new problem and the relative diversity to the set of solutions selected so far. The key difference between the original Greedy Selection method (used for analysis tasks) and our variant (used for planning, which is a synthesis task) stems from the fact that plan-diversity-aware retrieval needs to take the solution plan into account, in addition to the problem. During retrieval, the problem is considered for similarity purposes, while the solution is considered for diversity purposes.

```

1. define PlanDiversityGreedySelection(n,C,k)
2. begin
3.   R := {}
4.   For i := 1 to k
5.     Sort C by SimPIDiv(n,c,R) for each c in C
6.     R := R + First(C)
7.     C := C - First(C)
8.   EndFor
9. return R
10. end

```

Fig. 2. The Plan Diversity Greedy Selection algorithm: a case-based planning variant of Greedy Selection [19]

We assume a transformational-analogy adaptation method, in which the contents of a case are a problem (consisting of an initial and/or final state) and a solution, consisting of a plan. The new problem is defined in terms of initial and/or final state.

In Fig. 2, n is the new problem, C the case-base, and k the number of cases we aim to retrieve. In our variant of the algorithm, the quality based on which retrieval occurs is:

⁴ We chose to use general Greedy Selection, rather than its variant Bounded Greedy Selection [19], which improves performance for large case bases, as retrieval from our particular case base is manageable with the general algorithm. Alternatively, we can assume our case base to consist of only the top bk most similar cases, making our algorithm a variant of Bounded Greedy Selection.

$$SimPlDiv(n, c, R) = \alpha Sim(n, c) + (1 - \alpha) RelPlDiv(c, \pi, R, \Pi) \quad (4)$$

where *Sim* is the case similarity measure used for traditional similarity-based retrieval (most generally, similarity of initial and/or final states), α a parameter used for varying the complementary weights assigned to the similarity and diversity retrieval criteria, c, π the solution plan of case c , R, Π the set of solution plans in the set of cases R , and $RelPlDiv(\pi, \Pi)$, the diversity of a plan π relative to a set of plans Π (adapted from the *RelDiversity* formula proposed by Smyth and McClave [19]):

$$RelPlDiv(\pi, \Pi) = \frac{\sum_{\pi' \in \Pi} Dist(\pi, \pi')}{|\Pi|} \quad (5)$$

Dist can be any distance metric, either quantitative or qualitative.

5 Experimental Evaluation

Our experimental environment is real-time strategy game Wargus, which has previously been used in case-based planning work [3,15].

5.1 Experimental Setup

Game Configuration. We run two-player Wargus games on two 32x32 tile maps (Fig. 3), with our team’s plans executed against the built-in Wargus enemy AI. The types of units and available actions are as described in Section 3. Each plan represents an individual battle (in which one of our armies engages the enemy), rather than a complete, prolonged game. This restriction was necessary so as not to allow excessive implicit game-play diversity, which might render meaningless the difference in variance between results produced using different metrics. The two maps on which we test our plans are topologically different: the first has one gap in the forest separating the two armies, while the second has two gaps, located at different coordinates than the gap in the first map. This difference is meaningful for the following reason: on the second map, units will sometimes make different choices as to which gap to use to pass to the other side: sometimes, all units will use the same gap, at other times, they will split up, sometimes they will even “hesitate”, marching towards one gap, then returning to the other one. This ensures considerably different game behavior between the two maps.

Case-based Planning System. In our case-based planning system, we use the following convention: the cases are interpreted as battle-plan blueprints, so that every unit in a case is an abstracted representation of an entire army of units of that type (e.g. a soldier stands for an army of soldiers). New problems consist of actual game configurations, specifying number of armies of each type, as well as number of units in each army.

The *case base* consists of 100 distinct cases, each composed of an initial state (the problem) and a plan (the solution). The *initial state* is represented in terms of numbers of armies of each type. Each of these armies is represented by one unit in the plan.

The *plans* were generated using the FastForward [7] generative planner, modified so as to generate multiple plans for the same problem. All plans contain an attack action by one unit (which represents the entire attacking army in the adapted plan). No goal state is specified: the general goal is to obtain the highest possible score, and there is never one single final state through which this is achieved.



Fig. 3. The two topologically-different game maps, with archer armies highlighted. Note how the archer army in the second map has split up into two divisions, each using a different gap to pass.

The *new problems* consist of initial game states, indicating the number of armies of each type (soldier, archer, mage, peasant), as well as the number of units in each of the armies. All units in an army are of the same type. There are 5 new problems, with varying numbers of armies of each type, as well as number of units per army.

The *adaptation* algorithm is consistent with the idea of a retrieved plan serving as blueprint. As each unit in the retrieved plan represents an army, each army A in the new problem will be matched to a unit U (of the same type as the units in A) in the retrieved plan. All units in A will then perform all actions performed by U in the retrieved plan. The matching will occur in order of the numbering of units in the retrieved plan, with one exception: if unit U is the attacking unit in the retrieved plan, U will be the first to be assigned to an army of its type in the new problem, assuming such an army exists. This will always be the case with our problems: they all contain at least one army of each type, in order to be able to take at least partial advantage of any retrieved plan.

For *case retrieval*, we use the PlanDiversityGreedySelection retrieval algorithm (Fig. 2), where $k=4$, $\alpha=0.5$, and Sim is a similarity metric Sim_{InitSt} , based on the initial states of the compared cases:

$$Sim_{InitSt}(c_1 \cdot IS, c_2 \cdot IS) = \frac{\sum_{i=1}^n \frac{\min(numArmiesType_i(c_1 \cdot IS), numArmiesType_i(c_2 \cdot IS))}{\max(numArmiesType_i(c_1 \cdot IS), numArmiesType_i(c_2 \cdot IS))}}{n} \quad (6)$$

In Formula 6, n is the number of types of units (in our experimental setup, $n=4$) and $numArmiesType_i(c \cdot IS)$ is the number of armies of units of type i in the initial state of case c .

As the distance metric $Dist$, we use the quantitative metric $Dist_{Quant}$ (Formula 3), as well as a game-specific qualitative metric, which we call $Dist_{Wargus}$:

$$\begin{cases}
 Dist_{Wargus}(\pi_1, \pi_2) = & \\
 0, & \text{if } attackUnitsType(\pi_1) = attackUnitsType(\pi_2) \\
 d, & 0 < d \leq 1, \text{ if } attackUnitsType(\pi_1) \neq attackUnitsType(\pi_2)
 \end{cases} \tag{7}$$

<u>NEW PROBLEM</u>	<u>ADAPTED PLAN</u>
<p><i>Initial State</i> 1 soldier army (4 units) 1 peasant army (4 units) 2 mage armies (4 units each) 2 archer armies (4 units each)</p>	<p><i>move (archer1, 05_05)</i> <i>move (archer2, 05_05)</i> <i>move (archer3, 05_05)</i> <i>move (archer4, 05_05)</i> <i>move (peasant1, 03_02)</i> <i>move (peasant2, 03_02)</i> <i>move (peasant3, 03_02)</i> <i>move (peasant4, 03_02)</i> <i>move (mage1, 04_07)</i> <i>move (mage2, 04_07)</i> <i>move (mage3, 04_07)</i> <i>move (mage4, 04_07)</i></p>
<p><u>RETRIEVED CASE</u></p> <p><i>Initial State</i> 1 soldier army 1 peasant army 1 mage army 1 archer army</p> <p><i>Plan</i> <i>move (archer1, 05_05)</i> <i>move (peasant1, 03_02)</i> <i>move (mage1, 04_07)</i> <i>move (archer1, 24_07)</i> <i>patrol (soldier1, 01_04)</i> <i>move (soldier1, 05_04)</i> <i>attack (archer1, 24_07)</i></p>	<p><i>move (archer1, 24_07)</i> <i>move (archer2, 24_07)</i> <i>move (archer3, 24_07)</i> <i>move (archer4, 24_07)</i> <i>patrol (soldier1, 01_04)</i> <i>patrol (soldier2, 01_04)</i> <i>patrol (soldier3, 01_04)</i> <i>patrol (soldier4, 01_04)</i> <i>move (soldier1, 05_04)</i> <i>move (soldier2, 05_04)</i> <i>move (soldier3, 05_04)</i> <i>move (soldier4, 05_04)</i> <i>attack (archer1, 24_07)</i> <i>attack (archer2, 24_07)</i> <i>attack (archer3, 24_07)</i> <i>attack (archer4, 24_07)</i></p>

Fig. 4. Sample case, new problem, and the corresponding adapted plan (units in the adapted plan are not annotated with the army they belong to because, in this example, there is only one army of each type). The second action parameter indicates the coordinates of the location at which the action should take place.

In Formula 7, $attackUnitsType(\pi)$ is the type of units in the attacking army of plan π , and d is the degree of difference between two types of units, as defined based on game-specific knowledge, and indicated in Table 1 below.

Table 1. Domain-specific degrees of distance between types of Wargus units

Unit type 1	Unit type 2	d
Peasant	Soldier	0.50
Peasant	Archer	0.75
Peasant	Mage	0.90
Soldier	Peasant	0.50
Soldier	Archer	0.50
Soldier	Mage	0.75
Archer	Peasant	0.75
Archer	Soldier	0.50
Archer	Mage	0.50
Mage	Peasant	0.90
Mage	Soldier	0.75
Mage	Archer	0.50

5.2 Experimental Evaluation

Evaluation Method. To evaluate the diversity of game-play sessions which are based on the sets of generated plans, we observe the variation of two game-specific *evaluation metrics*. The primary metric is *Wargus score* (computed as in Section 3); the secondary metric is *time* (the duration, in game cycles, of game-play sessions).

Our *hypothesis* is that plans obtained using retrieval based on the qualitative plan-diversity metric $Dist_{Wargus}$ will produce greater game-play variation (reflected in the evaluation metrics), than plans obtained using the action-set quantitative distance metric $Dist_{Quant}$. We expect that, when run in the game, adaptations of plans retrieved using the qualitative distance metric will produce significantly more variation (as measured using standard deviation and assessed using the F-test) of Wargus scores⁵ than adaptations of quantitatively-diverse sets of plans. We expect to see a similar behavior with regard to time, but with less confidence, as we have observed that game duration tends to vary more between runs of the same plan, on the same map.⁶

⁵ Note how we have chosen one of the countless possible domain-specific, qualitative distance metrics in accordance with our purpose: that of obtaining easily quantifiable diversity. Had our objective been different, we might have opted for a distance metric producing some form of diverse game behavior which is not so clearly reflected in score variation.

⁶ Had we chosen time as the primary metric, we might have retrieved plans which use diverse route waypoints, encouraging the variation of game duration more clearly than that of score.

Results. In Fig. 5, each point in each chart represents the standard deviation of score or time (as indicated) for one plan set of 4 plans, where each plan is run in the game 5 times. The two data sets in each chart correspond to results obtained using the quantitative distance metric $Dist_{Quant}$ and the qualitative metric $Dist_{Wargus}$ in retrieval. There are 5 plan sets for each of the 5 new problems, on each of the 2 maps (50 plan sets in all).

As can be seen in the charts, for *score*, the standard deviation of $Dist_{Wargus}$ results per plan set is consistently higher than that of $Dist_{Quant}$ results.⁷ Being highly diverse, the $Dist_{Wargus}$ score sets always include the highest recorded score per problem/map combination (while $Dist_{Quant}$ sets do not).

The F-test score results indicate that the difference between the variances of the $Dist_{Wargus}$ and $Dist_{Quant}$ score data sets is statistically significant, at the 95% confidence level, for all problems, on both maps, with the $Dist_{Wargus}$ data set displaying the greater variance.

For the secondary metric of *time*, the standard deviation of $Dist_{Wargus}$ results is greater than that of $Dist_{Quant}$ results on all but 2 of the 25 plan sets on the first map, and all but 3 out of the 25 plan sets on the second map.

The F-test indicates that the $Dist_{Wargus}$ data sets display greater variance, and the variance difference is statistically significant, at the 95% confidence level, on 4 of the 5 problems on each map. On the second map, the difference is statistically significant (with greater variance for the $Dist_{Wargus}$ data set), at the 90% confidence level, on the remaining problem. For the remaining problem on the first map, the variance is slightly greater for the $Dist_{Quant}$ data set, but the difference is not statistically significant.

To sum up, $Dist_{Wargus}$ results are significantly more diverse than $Dist_{Quant}$ results on all problems for the score metric, and on the majority of problems for the time metric. This is consistent with our expectations.

In terms of plan quality, we have noticed that plans retrieved using $Dist_{Wargus}$ (and, consequently, the adapted plans based on them) tend to be shorter, on average, than plans retrieved using $Dist_{Quant}$ (the reason for this should be obvious from the way the two metrics are computed, with $Dist_{Quant}$ easily increasable by lengthening any of the compared plans, as long as the added actions are not encountered in the other plan). Plan length relates to the time it takes to execute the strategy outlined in the plan. It follows that shorter plans may, in this context, be preferable to longer ones. This suggests that well-chosen qualitative distance metrics can also help ensure that retrieved plans are of good quality.

⁷ The question might be raised whether plan sets producing highly diverse scores, from high to low (rather than all of the plans playing the game expertly) are ever of practical value: a simple example is the modeling of AI enemies, which, to make the game environment realistic (as well as not discouragingly difficult) should vary in intelligence and ability. Also, in partially unknown environments (e.g. the map remains the same, but the enemy force may vary over consecutive plans), we may benefit from experimenting with multiple diverse plans, even if some of them behaved poorly in a slightly different game configuration.

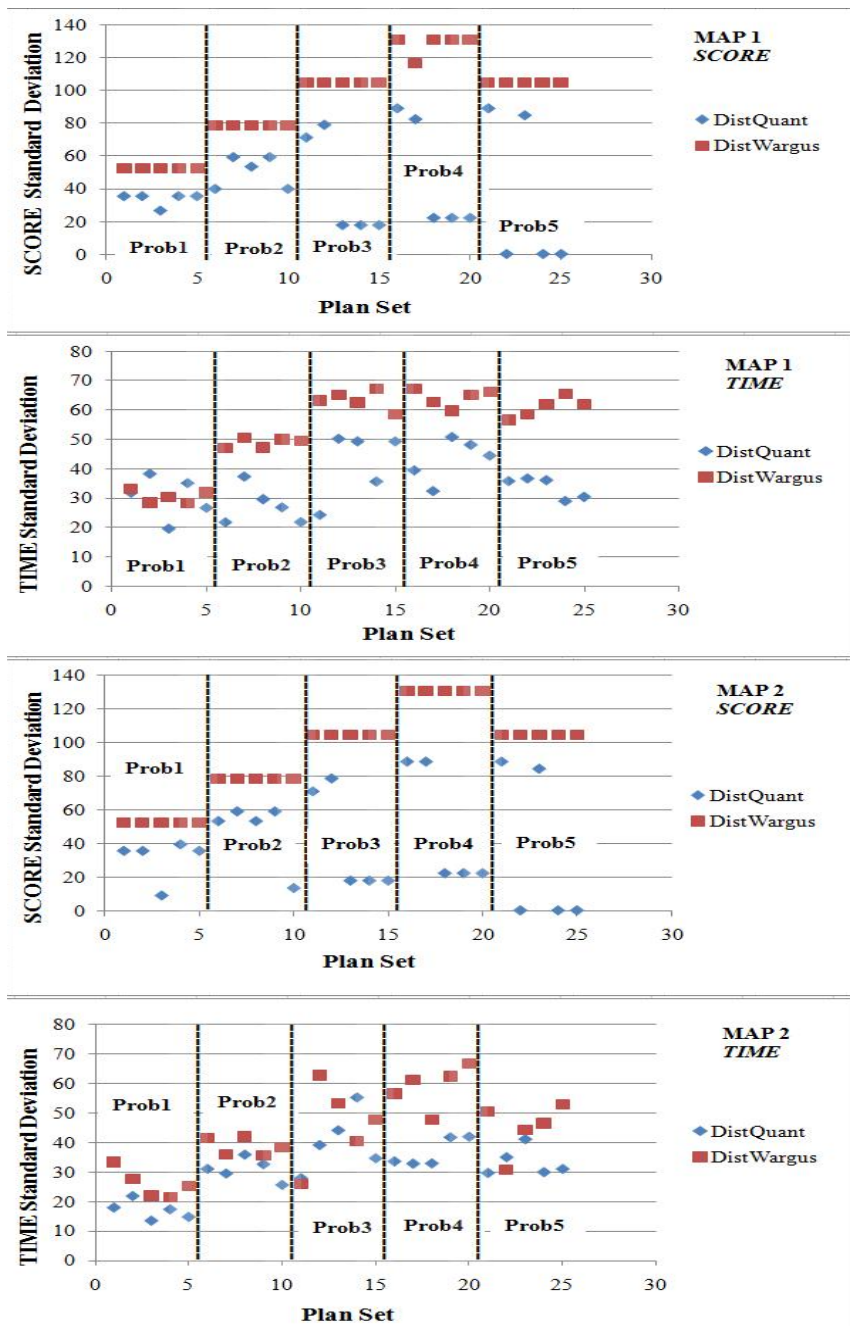


Fig. 5. Standard deviation of game scores and time (game duration)

6 Related Work

Case diversity was explored extensively in case-based recommender systems [2,8,9,10,18,19].

In case-based planning, we began to explore diversity in our previous work [3]. However, the focus there was on comparing plan diversity with state diversity, and we only demonstrated quantitative plan distance. While we also tested the diversity of plans by running them in the Wargus game, the game configurations were less sophisticated (with fewer unit types and simpler plans), as was our case-based planning system.

In generative planning (which involves generating plans from scratch, rather than through case retrieval and adaptation), quantitative plan diversity has been explored by Srivastava *et al.* [21]. A method for qualitative-diversity-aware plan generation has been proposed by Myers and Lee [12], in HTN planning. Their knowledge-intensive approach does not use distance metrics at plan generation time. Instead, it directs the generative planner towards regions of the search space which are identified as representing qualitatively different plan attributes, using a domain metatheory (an extended description of the planning domain in terms of high-level attributes, supplementing the standard domain model). In [4], we explore quantitative and qualitative plan diversity in generative planning.

To our knowledge, apart from our previous work [3,4], no other work on plan diversity (generative or case-based) assesses plan diversity by running plans in their environment, and observing behavior and results thus obtained. Instead, this is achieved by analyzing the plans themselves [12,21].

Myers [14] explores qualitative plan comparison (identifying similarities and differences between plans) through the use of a domain metatheory. The approach assumes an HTN planning paradigm, and defines plan distance purely on the basis of high-level characteristics specified in the metatheory. It does not deal with diverse plan generation, but with the computation of distance between already available plans. The related problem of plan stability is explored by Fox *et al.* [6]. Plan stability aims at reducing the difference between an original plan and a repaired plan. The similarity metric they use for this purpose is quantitative. Storyline diversity in a gaming environment, for the purpose of enhancing the player's experience, is explored by Paul *et al.* [16].

7 Conclusions and Future Work

Our work brings two main contributions to case-based plan diversity research. First, to our knowledge, we approach qualitative diversity in case-base planning for the first time. Second, we obtain qualitative plan diversity through the use of a qualitative plan distance metric at case retrieval time.

In a game domain, we show how qualitative plan diversity can, by reflecting characteristics specific to the domain in question, produce more meaningful plan variation than quantitative diversity. In addition, we evaluate the diversity of generated plans by running them in the environment and observing their behavior, as opposed to examining the structure of the plans themselves.

In future work, we aim to explore qualitative plan diversity in various real domains of practical interest, once again testing the diversity of plans by running them in the environments. We are also interested in exploring diversity in online planning, which should be particularly interesting in game domains, such as the one used herein.

We also plan to analyze the trade-off between plan diversity and plan quality; and to explore whether qualitative distance metrics can be used to help ensure that the sets of retrieved plans are not only diverse, but also composed of individual plans of good quality.

Acknowledgements. We thank the anonymous reviewers for their comments.

References

1. Boddy, M., Gohde, J., Haigh, T., Harp, S.: Course of Action Generation for Cyber Security Using Classical Planning. In: Biundo, S., et al. (eds.) Proc. of ICAPS 2005, pp. 12–21. AAAI Press, Menlo Park (2005)
2. Bridge, D., Kelly, J.P.: Ways of Computing Diverse Collaborative Recommendations. In: Wade, V.P., Ashman, H., Smyth, B. (eds.) AH 2006. LNCS, vol. 4018, pp. 41–50. Springer, Heidelberg (2006)
3. Coman, A., Muñoz-Avila, H.: Case-based Plan Diversity. In: Bichindaritz, I., Montani, S. (eds.) ICCBR 2010. LNCS, vol. 6176, pp. 66–80. Springer, Heidelberg (2010)
4. Coman, A., Muñoz-Avila, H.: Generating Diverse Plans Using Quantitative and Qualitative Plan Distance Metrics. In: Proc. of AAAI 2011 (to appear, 2011)
5. Cox, M.T., Muñoz-Avila, H., Bergmann, R.: Case-based Planning. *The Knowledge Engineering Review*, 1–4 (2005)
6. Fox, M., Gerevini, A., Long, D., Serina, I.: Plan Stability: Replanning Versus Plan Repair. In: Long, D., et al. (eds.) Proc. of ICAPS 2006, pp. 212–221. AAAI Press, Menlo Park (2006)
7. Hoffmann, J., Nebel, B.: The FF Planning System: Fast Plan Generation Through Heuristic Search. *Journal of Artificial Intelligence Research* 14, 253–302 (2001)
8. McGinty, L., Smyth, B.: On the Role of Diversity in Conversational Recommender Systems. In: Ashley, K.D., Bridge, D.G. (eds.) ICCBR 2003. LNCS (LNAI), vol. 2689, pp. 276–290. Springer, Heidelberg (2003)
9. McSherry, D.: Increasing Recommendation Diversity Without Loss of Similarity. In: Proc. of the Sixth UK Workshop on Case-based Reasoning, Cambridge, UK, pp. 23–31 (2001)
10. McSherry, D.: Diversity-Conscious Retrieval. In: Craw, S., Preece, A.D. (eds.) ECCBR 2002. LNCS (LNAI), vol. 2416, pp. 219–233. Springer, Heidelberg (2002)
11. McSherry, D.: Completeness Criteria for Retrieval in Recommender Systems. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) ECCBR 2006. LNCS (LNAI), vol. 4106, pp. 9–29. Springer, Heidelberg (2006)
12. Myers, K.L., Lee, T.J.: Generating Qualitatively Different Plans through Metatheoretic Biases. In: Proc. of AAAI 1999, pp. 570–576. AAAI Press, Menlo Park (1999)
13. Myers, K.L., Tyson, W.M., Wolverton, M.J., Jarvis, P.A., Lee, T.J., des-Jardins, M.: PASSAT: A User-centric Planning Framework. In: Proc. of the 3rd Intl. NASA Workshop on Planning and Scheduling for Space, Houston, TX (2002)
14. Myers, K.L.: Metatheoretic Plan Summarization and Comparison. In: Long, D., et al. (eds.) ICAPS 2006, pp. 182–192. AAAI Press, Menlo Park (2006)

15. Ontañón, S., Mishra, K., Sugandh, N., Ram, A.: On-line Case-Based Planning. *Computational Intelligence Journal* 26(1), 84–119 (2010)
16. Paul, R., Charles, D., McNeill, M., McSherry, D.: MIST: An Interactive Storytelling System with Variable Character Behavior. In: Aylett, R., Lim, M.Y., Louchart, S., Petta, P., Riedl, M. (eds.) *ICIDS 2010. LNCS*, vol. 6432, pp. 4–15. Springer, Heidelberg (2010)
17. Serina, I.: Kernel Functions for Case-based Planning. *Artificial Intelligence* 174(16-17), 1369–1406 (2010)
18. Shimazu, H.: ExpertClerk: Navigating Shoppers' Buying Process with the Combination of Asking and Proposing. In: Nebel, B. (ed.) *Proc. of IJCAI 2001*, pp. 1443–1448. Morgan Kaufmann, Seattle (2001)
19. Smyth, B., McClave, P.: Similarity vs. Diversity. In: Aha, D.W., Watson, I. (eds.) *ICCBR 2001. LNCS (LNAI)*, vol. 2080, pp. 347–361. Springer, Heidelberg (2001)
20. Spalazzi, L.: A Survey on Case-Based Planning. *Artificial Intelligence Review* 16, 3–36 (2001)
21. Srivastava, B., Kambhampati, S.T., Nguyen, M., Do, B., Gerevini, A.: Domain Independent Approaches for Finding Diverse Plans. In: Veloso, M.M. (ed.) *Proc. of IJCAI 2007*, pp. 2016–2022. AAAI Press, Menlo Park (2007)