Fast and Diverse Narrative Planning through Novelty Pruning

Rachelyn Farrell and **Stephen G. Ware**

rfarrell@uno.edu, sgware@uno.edu
Narrative Intelligence Lab
University of New Orleans
New Orleans, LA, USA

Abstract

Novelty pruning is a simple enhancement that can be added to most planners. A node is removed unless it is possible to find a set of n literals which are true in the current state and have never all been true in any of that plan's previous states. Expanding on the success of the *Iterated Width* algorithm in classical planning and general game playing, we apply this technique to narrative planning. Using a suite of 8 benchmark narrative planning problems, we demonstrate that novelty pruning can be used with breadth-first search to solve smaller problems optimally and combined with heuristic search to solve larger problems faster. We also demonstrate that when many solutions to the same problem are generated, novelty pruning can produce a wider variety of solutions in some domains.

Introduction

Narrative planning (Riedl and Young 2010) is a variant of classical planning which searches for a sequence of actions to achieve the author's goal such that all actions are clearly motivated and goal-oriented for the agents who take them. Plan-based models of narrative have proven a popular paradigm for representing, generating, and adapting stories (Young et al. 2013), and planning algorithms have been used to control a variety of interactive narrative experiences (Cavazza, Charles, and Mead 2002; Pizzi and Cavazza 2007; Porteous, Cavazza, and Charles 2010; Ware and Young 2015; Robertson and Young 2015).

We previously introduced the Glaive Narrative Planner (Ware and Young 2014), which reasons about the intentional structure of narrative plans and leverages advances in state-space heuristic search to increase the size of problems that can be solved in a practical amount of time. This paper introduces a technique called novelty pruning which can be used independently of or in conjunction with other narrative planning techniques.

Novelty pruning (Geffner and Lipovetzky 2012) removes a plan from the search space if it fails to find a set of literals of size n in the current state which are true but have never been true at any time in the past. This saves the planner considerable effort by avoiding redundant plans and focusing on those which introduce the most change into the current state.

In our experience, novelty pruning:

- is simple and easy to implement.
- can be paired with uninformed search techniques like breadth-first search to solve smaller problems optimally.
- can be paired with heuristic search techniques to improve their efficiency on large problems.
- generates a wider variety of solutions in some domains when used to produce more than one for a problem.

This paper is organized into two sections which explore these claims. The first section demonstrates how novelty pruning can improve breadth-first search and heuristic-driven search on a set of benchmark narrative problems. The second section demonstrates that novelty pruning improves the diversity of solutions generated by a narrative planner in some domains.

Planning Speed

Related Work

Novelty Pruning The concept of novelty pruning is taken directly from the *Iterated Width* (or *IW*) algorithm (Geffner and Lipovetzky 2012). In its simplest form, *IW* is just breadth-first search through the space of states that prunes any node which is not sufficiently novel.

The *state space* of a planner is a directed tree whose nodes are states and whose edges are actions. An edge $s_1 \xrightarrow{a} s_2$ exists when the preconditions for action a are true in state s_1 and the effects of a have been applied to state s_2. A node represents a *current state*, and its position in the tree (i.e. the path of edges from the root to that node) is a plan. A node's *previous states* are the root node (i.e. the initial state) and all states on the path from the root to that node.

Novelty is defined as an integer n. A plan is said to have novelty n if, in its current state, it is possible to find a set of n literals which have never all been true at the same time in any of that plan's previous states. States with lower values of n are more novel. Novelty pruning works by setting a threshold for n and pruning any nodes in the search space whose minimum novelty exceeds the threshold.

With threshold $n = 1$, *IW* prunes any plan whose current state fails to make some new literal true (one which has never been true in any of that plan's previous states). With threshold $n = 2$, *IW* tries to find a pair of literals $\{A, B\}$

which have never both been true before, and if it fails, the plan is pruned. Note that A may have been true in some previous state, and B may have been true in some previous state, but if the conjunction $A \land B$ has never been true in a previous state, the plan has novelty 2. So on for $n > 2$.

It may seem counterintuitive that lower values of n mean higher novelty. Another way one might define n is the size of the set of literals needed to prove a state is novel. To prove novelty, one first considers all sets of literals of size 1, and if none can prove the state novel, one tries all the sets of size 2, etc. The smaller n is, the easier it is to prove novelty.

Geffner and Lipovetzky (2012) demonstrated that many classical planning benchmarks have a bounded and low width (meaning that *IW* can solve them for some low threshold of n, usually $n = 1$ or $n = 2$). The threshold value can be fixed if the width of the problem is known in advance, or *IW* can start with the threshold $n = 1$ and, if it fails to find a solution, try higher and higher thresholds for n until it solves the problem or runs out of resources. If the threshold is fixed and the width of the problem is unknown, the planner may be incomplete; hence *IW* increases the threshold each time the planner fails and tries again. Variants of *IW* have achieved competitive results on classical planning benchmarks (Geffner and Lipovetzky 2012). Recent results (Geffner and Geffner 2015) demonstrated that *IW* achieves state-of-the-art results in the General Video Game AI competition with thresholds of $n = 1$ and $n = 2$.

One advantage of novelty pruning is that it is agnostic to the search technique being used. It can be combined with other techniques besides breadth-first search, such as heuristic search. Intelligent pruning can make a dramatic difference in the size of problems that planner can solve, as demonstrated by Teutenberg and Porteous's IMPRACTical planner (2013; 2015), which can be used for online narrative generation in some domains thanks to its intelligent pruning of the search space.

Narrative Planning A full description of narrative planning and the Glaive heuristic is outside the scope of this paper (see Ware and Young, 2014). In short, a planning problem takes as input the initial state of the world, a set of actions which change the state, and a goal. A plan is a sequence of actions such that the preconditions of every action are satisfied immediately before it is taken and, after all actions have been taken, the problem goal has been met.

Narrative planning (Riedl and Young 2010) is an extension to planning which reformulates the problem goal as the author's goal and also considers the goals of individual characters. Each action is annotated with a list of characters who must consent to take that action. Characters only consent to take an action if it contributes to one of their goals or satisfies the precondition of a later action which contributes to one of their goals. The solution to a narrative planning problem is a sequence of actions that achieves the author's goal using only actions which can be explained in terms of the individual goals of the characters who take them.

The challenge faced by forward-chaining narrative planners like Glaive is that steps taken at the beginning of the plan may not get explained until later or may never be explained at all. The search must consider not only distance to the author's goal but also how difficult it will be to explain any currently unexplained steps. The Glaive heuristic accounts for these challenges by considering two different metrics. The first is based on the Fast Forward heuristic (Hoffmann and Nebel 2001), which estimates how many steps need to be taken to achieve the author's goal. The second estimates how many steps need to be taken to explain those steps which are currently unexplained. Glaive uses a maximum of these two estimates when deciding in what order to visit nodes during search.

In addition to its heuristic, the Glaive planner performs motivation pruning. This technique provides an efficient and necessary but not sufficient means of recognizing when a step cannot possibly be explained by any future sequence of steps. In other words, when motivation pruning removes a node it is certain that such a node can never lead to a solution, but some such nodes may be missed. In the results below, all planners use motivation pruning.

Results

We compared the performance of six narrative planners:

- Breadth-first search (as a baseline)

- Breadth-first search + novelty pruning ($n = 1$ and 2)

- A* + Glaive heuristic

- A* + Glaive heuristic with novelty pruning ($n = 1$ and 2)

These techniques were tested on a suite of 8 benchmark narrative planning problems described in Table 1. That table gives the problem name, its authors, the total number of ground actions and axioms after pruning (as described by Ware and Young, 2014), the width of the problem (minimum n for which a solution can be found), and the number of actions in the shortest known solution. We did not consider any thresholds above $n = 2$ because none of these benchmark problems have a width greater than 2, which is consistent with the findings of Geffner and Lipovetzky (2012).

Each planner was tested on each problem. Tests were performed on a Dell Precision 5810 desktop with 3.5 GHz Intel Xeon processor. Each planner was given 100 GB of RAM and 1 hour to solve each problem. The results are given in Table 2. Each problem shows the number of nodes visited during search, the total number of nodes generated, the number of nodes pruned (by motivation pruning and novelty pruning, if applicable), and the time spent in milliseconds (average of 10 runs). When a solutions is listed as $\neg\exists$, such as breadth-first search with threshold $n = 1$ on the Ark problem, it means that the planner visited all the nodes in the search space without finding any solutions (which is expected, because the width of the Ark problem is 2). The anomalous time of 3.2 milliseconds for breadth-first search with threshold $n = 2$ on the $Space$ problem is likely due to a poorly timed run of the garbage collector.

The table below gives the improvement factors for each planner in terms of time (milliseconds spent) and space (nodes visited). It assumes the lowest n for which a solution can be found. For example, on the *Fantasy* problem,

Table 1: Benchmark Narrative Planning Problems

Problem	Authors	Actions	Axioms	Width	Solution
Heist	Niehaus (2009)	1830	0	2	30
Space	Ware, Young, Harrison, and Roberts (2014)	40	0	1	2
Fantasy	Ware, Young, Harrison, and Roberts (2014)	53	0	1	6
Western	Ware, Young, Harrison, and Roberts (2014)	632	0	1	5
Ark	Ware (2014)	76	6	2	7
BLP-Die	Ware and Young (2015)	845	625	2	7
BLP-Win	Ware and Young (2015)	845	625	2	10
Life	Farrell and Ware (forthcoming)	67	0	1	11

Table 2: Narrative Planning Results on Benchmark Problems (best results for each problem highlighted)

Problem		Planner					
		BFS	BFS $n = 1$	BFS $n = 2$	Glaive	Glaive $n = 1$	Glaive $n = 2$
Heist	solution	out of time	out of time	out of time	out of time	out of time	out of time
	visited	19,266,167	16,595,975	18,941,422	1,046,098	1,333,410	1,115,670
	generated	365,567,309	287,355,532	352,547,797	22,723,488	23,267,401	20,563,751
	pruned	117,503,623	162,865,744	153,051,720	6,267,374	12,656,829	8,914,355
	time (ms)	3,633,640	3,600,001	3,600,002	3,600,002	3,600,006	3,600,002
Space	solution	2 steps	2 steps	2 steps	2 steps	2 steps	2 steps
	visited	6	6	6	3	3	3
	generated	56	56	56	24	24	24
	pruned	14	17	19	6	6	7
	time (ms)	< 1	< 1	3.2	< 1	< 1	< 1
Fantasy	solution	6 steps	6 steps	6 steps	6 steps	6 steps	6 steps
	visited	55,394	10,835	12,447	20,221	3,602	3,608
	generated	818,894	163,230	185,589	338,248	58,196	58,299
	pruned	135,487	89,787	90,903	46,595	35,139	34,180
	time (ms)	3,233	705	818	151,724	1,414	1,494
Western	solution	5 steps	5 steps	5 steps	6 steps	6 steps	6 steps
	visited	14,879	6,780	7,132	176,028	29,639	32,574
	generated	414,165	186,536	196,398	4,341,287	580,324	642,283
	pruned	132,605	96,393	95,316	799,078	242,693	239,601
	time (ms)	2,366	1,111	1,334	695,648	22,752	29,460
Ark	solution	7 steps	¬∃	7 steps	7 steps	¬∃	7 steps
	visited	4,132	14,101	1,882	186	14,101	174
	generated	26,601	68,843	12,202	950	68,843	876
	pruned	5,039	54,743	4,123	178	54,743	261
	time (ms)	127	464	63	22	927	16
BLP-Die	solution	7 steps	out of time	7 steps	7 steps	out of time	7 steps
	visited	1,271,055	1,066,334	278,103	1,250,936	879,805	39,294
	generated	23,472,945	26,116,416	5,519,664	21,924,886	17,500,100	703,891
	pruned	16,460,269	18,530,959	4,072,342	15,479,744	13,913,513	558,090
	time (ms)	1,662,262	3,600,004	441,093	2,886,313	3,600,002	78,200
BLP-Win	solution	10 steps	out of time	10 steps	11 steps	out of time	11 steps
	visited	1,271,055	1,062,338	278,103	1,057,967	880,716	18,819
	generated	23,472,945	26,012,432	5,519,664	18,416,493	17,518,071	338,157
	pruned	16,460,269	18,458,943	4,072,342	13,040,600	13,926,999	267,211
	time (ms)	1,663,905	3,600,004	440,635	2,440,509	3,600,003	36,544
Life	solution	11 steps	11 steps	11 steps	12 steps	12 steps	12 steps
	visited	37,144	37,144	37,144	97,191	97,191	97,191
	generated	225,850	225,850	225,850	613,110	613,110	613,110
	pruned	117,101	117,101	117,101	407,505	407,505	407,505
	time (ms)	5,066	5,249	5,280	39,999	40,542	40,581

Glaive with novelty pruning (threshold $n = 1$) runs 107.3 times faster than Glaive without novelty pruning.

Domain	Time Improve		Space Improve	
	BFS	Glaive	BFS	Glaive
Space	1.00	1.00	1.00	1.00
Fantasy	4.59	107.30	5.11	5.61
Western	2.13	30.58	2.19	5.94
Ark	2.02	1.38	2.20	1.07
BLP-Die	3.77	36.90	4.57	31.84
BLP-Win	3.78	66.78	4.57	56.22
Life	0.97	0.99	1.00	1.00

Novelty pruning improves performance on all problems with the exception of *Space* (whose solution is so trivial that all planners find it in under 1 millisecond) and *Life*, because novelty pruning fails to prune any nodes and incurs a slight overhead penalty. In many cases the improvements are dramatic. The *Best Laid Plans: Win* domain takes Glaive over 40 minutes to solve, but this is reduced to only 37 seconds with novelty pruning. It also visits about 56 times fewer nodes in the process.

It is interesting to note that for smaller problems breadth-first search with motivation and novelty pruning is sufficient and possibly even better in terms of time spent. Breadth first search always returns the optimal (fewest steps) solution, which is desirable in many cases. Most modern state-space planning heuristics, including Glaive, are satisficing heuristics that may overestimate. This is evident in the *Western* and *Life* domains where Glaive's solution is 1 step longer than the optimal plan.

Solution Diversity

Related Work

Narrative planners may be used to generate multiple solutions to the same problem. This can be used by interactive narrative systems to build branching stories that provide unique content in response to user actions but still accomplish the author's goals. In many cases it is desirable for the solutions generated to contain meaningfully different story content; for example, to increase the user's sense of agency in interactive stories, or to allow a human author to choose from a diverse set of possible solutions to use. To measure the diversity of a set of story plans, we require a domain-independent plan distance metric that captures semantically relevant differences between stories.

Kypridemou and Michael (2013) demonstrated that, from a human standpoint, evaluating story similarity can be reduced to the comparison of "common summaries" of each story. To that end, Amos-Binks et al. (2016) proposed a new distance metric that compares two plans by means of an *Important-Step Intention-Frame* (ISIF) story plan summary, which takes into account story-centric information; namely plot progression and character goals. The former is summarized by a set of *important steps*—the set of executed steps with the highest causal degree, where the causal degree of a step is its number of satisfied preconditions plus its number

of used effects. The latter is summarized by a set of *intention frame summaries*.

In narrative planning, an *intention frame* is a 5-tuple $< c, g, m, f, S >$ where c is a character, g is a goal of that character, m is the step that motivated that goal, f is the final satisfying step which has g as an effect, and S is the full set of steps taken by c in order to achieve g. In other words, an intention frame describes how a character acquired a goal and the steps they took to achieve it. An intention frame exists for each character goal that is satisfied in some branch of the story. The *summary* of an intention frame simply ignores the set of steps taken in between the motivation and the final step. Thus, an intention frame summary is the quadruple $< c, g, m, f >$.

The ISIF distance between two plans is defined as the Jaccard (1912) distance between the two plan summaries:

$$\delta_{ISIF}(\varphi_1, \varphi_2) = 1 - \frac{1}{2}\left(\frac{|E(\varphi_1) \bigcap E(\varphi_2)|}{|E(\varphi_1) \bigcup E(\varphi_2)|} + \frac{|J(\varphi_1) \bigcap J(\varphi_2)|}{|J(\varphi_1) \bigcup J(\varphi_2)|}\right),$$

where φ_1 and φ_2 are ISIF plan summaries, E is the set of important steps of a plan summary, and J is the set of intention frame summaries of a plan summary. In other words, for each of the two sets, it divides the intersection of the sets for each plan by their union. This means that plans with a high number of character goals that were motivated by the same step and achieved by the same step will have an ISIF distance closer to 0, as will plans with a high number of important steps in common.

Amos-Binks et al. compared the ISIF distance metric to two others: *action distance* and *causal link distance* (Srivastava et al. 2007), which only take into account the set of actions in each plan and the set of causal links, respectively. They demonstrated that the ISIF metric is able to capture semantically meaningful differences between story plans that are syntactically very similar. They used two example plans from the *Space* domain which had the same number of actions and causal links but with a single difference in how the author goals were achieved vs. how a character goal was achieved. The ISIF metric recognized that this change had a significant impact on the story; it reflected an additional character goal and a change in the number of important steps. While the three metrics tended to agree with each other, the authors concluded that ISIF distance does offer additional insight into story plan similarity.

Results

Using the same domains from the previous section (except *Heist*, which no planner could solve), we generated 100 unique solutions first with breadth-first search and then with A* using the Glaive heuristic, both with and without the novelty pruning enhancement (with n set to the lowest threshold for which the problem could be solved). Unique solutions were generated by allowing the planner to continue exploring the search space rather than stopping after the first solution was found. In other words, instead of returning the first solution found, we return the first 100 solutions found. After each new solution was generated, we calculated the average ISIF distance between the current set of solutions. The final results are presented in Figures 1 and 2.
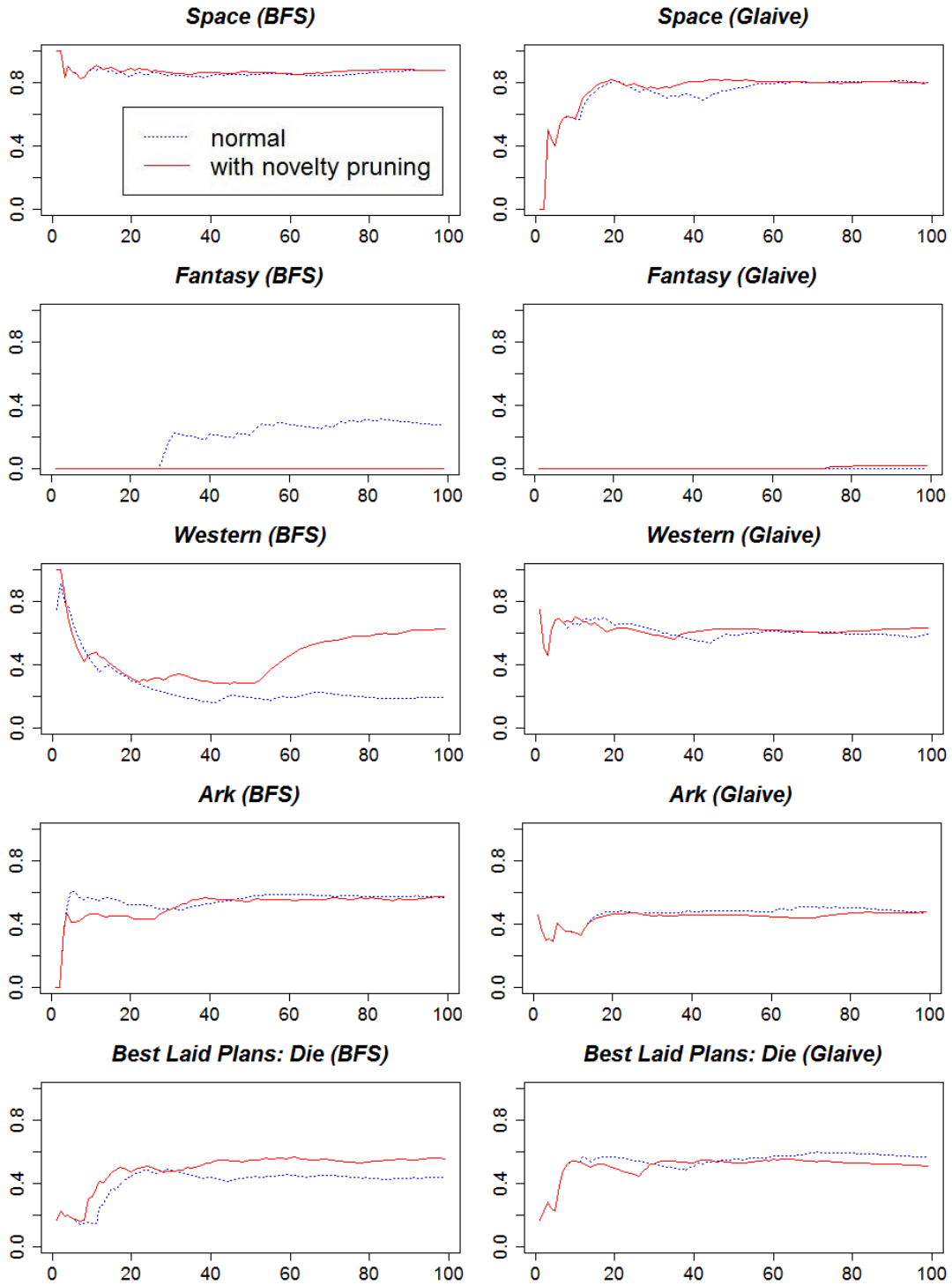
Figure 1: Average diversity among solutions generated. The x axis gives the number of plan in the set. The y axis gives the average diversity (expressed as average distance from 0 to 1 between all plans in the set).
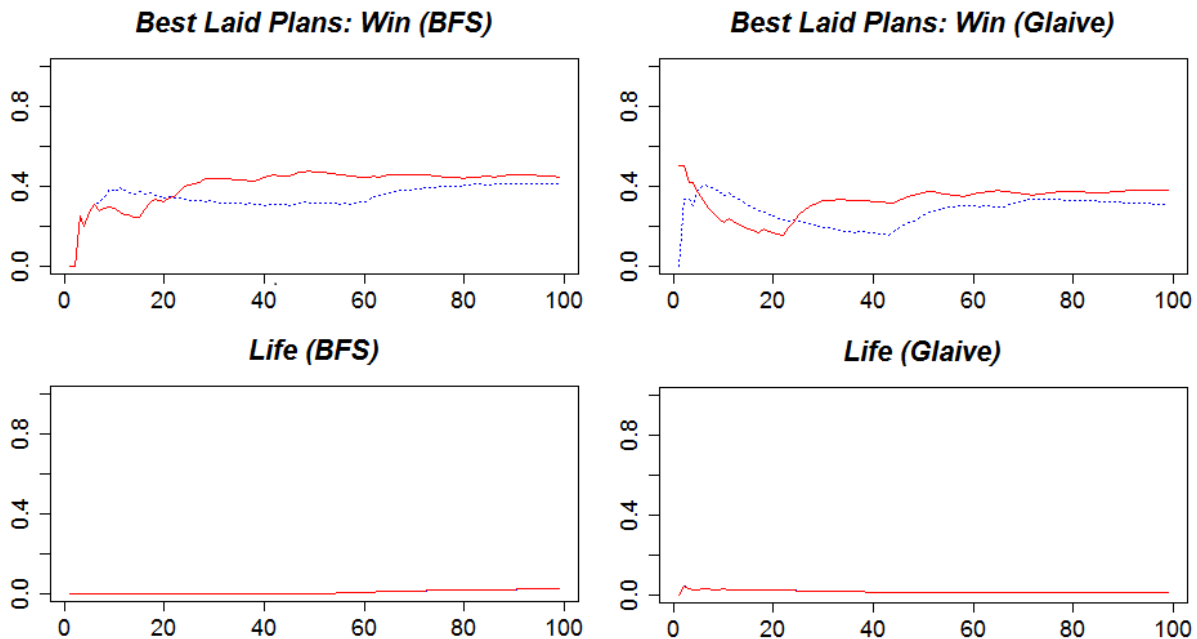
Figure 2: Average diversity among solutions generated. The x axis gives the number of plan in the set. The y axis gives the average diversity (expressed as average distance from 0 to 1 between all plans in the set).

We hypothesized that after we had generated 100 solutions, the solutions generated using novelty pruning would have a higher average diversity than those without it. Our hypothesis was supported in six out of the fourteen tests. In two out of the fourteen tests, the set of solutions generated without novelty pruning was more diverse. In most cases there was not a large difference between the two configurations, and some domains such as *Life* showed no difference between them at all.

These results are only somewhat encouraging, but it should be noted that the domains themselves are limiting factors in how diverse any set of solutions within them can be. *Life*, for example, was engineered to tell a specific story about how a character achieves two goals, where each goal can be accomplished in exactly two possible ways. The Author goals constrain the rest of the story such that all solutions must contain the exact same set of steps, although the order may change. (The ISIF distance metric does not capture differences in orderings between the same steps.) *Western*, by contrast, is a more open-ended domain where the Author's goals can be achieved in many different ways, intuitively allowing for more diverse solutions to be found.

We expected novelty pruning to produce a wider variety of plans because it prunes redundant plans from the search space. Consider a story where a character needs to travel from location A to location B to location C. If the solution contains other steps after this, search techniques like A* may consider redundant plans such as going from A to B to A to B to C. Novelty pruning prevents these kind of redundant paths from being considered, which may explain

its excellent performance in game playing domains (Geffner and Geffner 2015).

While it does provide greater diversity in some domains, it does not do so reliably in all domains. We believe this is because Glaive already precludes some redundancy such as, "I got to location A by going to location A, then location B, then back to location A again." Many of the redundant plans that novelty pruning excels at removing are already precluded by Glaive. However, it is important to note that the planner still spends effort considering these plans and ruling them out as non-solutions, so novelty pruning can save significant time by preventing it from even considering these redundant plans in the first place.

Conclusion

Novelty pruning is a simple technique for removing redundant plans from the search space of a narrative planning problem. It can be used to speed up uninformed and informed search techniques alike. In some domains it can also increase the variety of solutions generated.

Acknowledgements

References

Amos-Binks, A.; Roberts, D. L.; and Young, R. M. 2016. Summarizing and comparing story plans. In *Proceedings of the 2016 Workshop on Computational Models of Narrative*. (forthcoming).

Cavazza, M.; Charles, F.; and Mead, S. J. 2002. Character-based interactive storytelling. *IEEE Intelligent Systems special issue on AI in Interactive Entertainment* 17(4):17–24.

Geffner, T., and Geffner, H. 2015. Width-based planning for general video-game playing. In *Proceedings of the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 23–29.

Geffner, H., and Lipovetzky, N. 2012. Width and serialization of classical planning problems. In *Proceedings of the 20th European Conference on Artificial Intelligence*, 540–545. IOS 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.

Jaccard, P. 1912. The distribution of the flora in the alpine zone. *New phytologist* 11(2):37–50.

Kypridemou, E., and Michael, L. 2013. Narrative similarity as common summary. In *Proceedings of the 2013 Workshop on Computational Models of Narrative*, 129–146.

Niehaus, J. M. 2009. *Cognitive models of discourse comprehension for narrative generation*. Ph.D. Dissertation, North Carolina State University.

Pizzi, D., and Cavazza, M. 2007. Affective storytelling based on characters' feelings. In *AAAI Fall Symposium on Intelligent Narrative Technologies*, 111–118.

Porteous, J.; Cavazza, M.; and Charles, F. 2010. Applying planning to interactive storytelling: Narrative control using state constraints. *ACM Transactions on Intelligent Systems and Technology* 1(2):1–21.

Riedl, M. O., and Young, R. M. 2010. Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research* 39(1):217–268.

Robertson, J., and Young, R. M. 2015. Automated gameplay generation from declarative world representations. In *Proceedings of the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*.

Srivastava, B.; Nguyen, T.; Gerevini, A.; Kambhampati, S.; Do, M.; and Serina, I. 2007. *Domain independent approaches for finding diverse plans*. 2016–2022.

Teutenberg, J., and Porteous, J. 2013. Efficient intent-based narrative generation using multiple planning agents. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, 603–610.

Teutenberg, J., and Porteous, J. 2015. Incorporating global and local knowledge in intentional narrative planning. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 1539–1546.

Ware, S. G., and Young, R. M. 2014. Glaive: a state-space narrative planner supporting intentionality and conflict. In *Proceedings of the 10th AAAI International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 80–86. (awarded Best Student Paper).

Ware, S. G., and Young, R. M. 2015. Intentionality and conflict in The Best Laid Plans interactive narrative virtual environment. (forthcoming).

Ware, S. G.; Young, R. M.; Harrison, B.; and Roberts, D. L. 2014. A computational model of narrative conflict at the fabula level. *IEEE Transactions on Computational Intelligence and Artificial Intelligence in Games* 6(3):271–288.

Ware, S. G. 2014. *A plan-based model of conflict for narrative reasoning and generation*. Ph.D. Dissertation, North Carolina State University.

Young, R. M.; Ware, S. G.; Cassell, B. A.; and Robertson, J. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37(1-2):41–64.