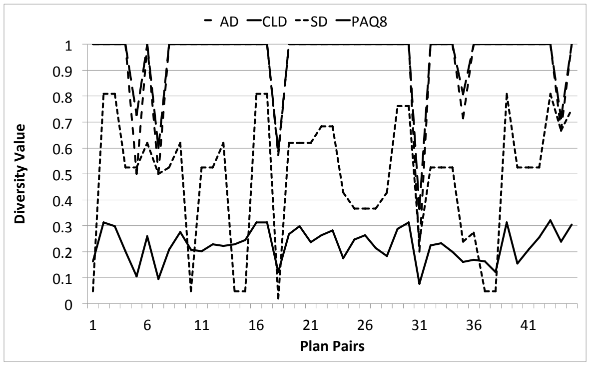

Figure 1: Comparisons of AD, CLD, SD, and NCD with pac8 compressor in problems with correlated fluents.

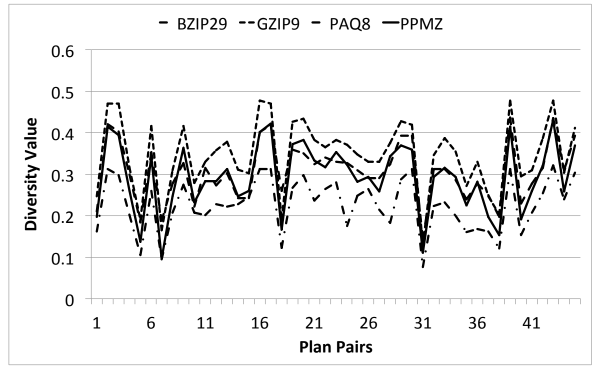

Figure 2: Comparisons of different compressors used by NCD in problems with correlated fluents. Above, NCD was run with gzip, bzip2, ppmz, and pac8.

We have also conducted an experiment in which we show that variations in naming and plan representations could pose challenges to NCD. These can cause NCD to behave poorly, but can be addressed by pre-processing. More problematic are cases involving plans with causally independent actions, and parameter shuffling. We illustrate the challenges, and suggest directions for future work.

Finally, we conducted an experiment to illustrate the utility of our diversity measure. We used diversity measures to choose training sets for the HTN-Maker learning system (Hogg, Muñoz-Avila, and Kuter 2008). We found that HTN Maker learns better with a training set that NCD labels as diverse versus one that it labels less diverse, while the value of AD cannot similarly predict learning performance.

In our experiments, we used the typed STRIPS dialect of PDDL (Planning Domain Description Language) (McDermott 1998) to represent planning domains, planning problems and plans. We have implemented the AD, CLD, and SD diversity measures, as described in (Nguyen et al. 2012a), and NCD as described in (Li et al. 2004) in Common Lisp. For NCD, we have used four different publicly-available compressors: gzip (Gailly and Adler 2014), bzip2 (Seward 2014), ppmz (Bloom 2014), and paq8 (Mahoney 2014). Since paq8 gave us the best compression, whenever we report only a single result for NCD, we report for paq8. We use all of the compressors at the highest compression effort setting, with the exception of paq8, for which we use -8 instead of -9 because -9 often crashed for us. The following sections describe our experimental scenarios and results. Experimental data (including plans and domain and problem definitions) will be available at www.sift.net/aaai14-diversity/.

Correlated Fluents. We have written an abstract planning domain for this experiment, where there are two agents one of which mirrors the other's actions. The problems involve an agent reaching a particular goal location in a grid, and we compare plans for problems with different initial positions of the two ninjas with different goals.

Figure 1 shows experimental results where we evaluated AD, CLD, SD, and NCD on a suite of randomly-generated problems in this planning domain. Figure 1 shows the results. As expected, AD, CLD, and SD all give inflated estimates of diversity. For some pairs (see the downward spiking entries), these measures can detect some commonality, but often they cannot make finer distinctions below the level of "these plans are completely different." NCD, on the other hand, is sensitive to the qualitative features of the plans and therefore, its distance measures are much lower.

Comparing Compressors. The different compressors use different kinds of internal models while compressing, and make different tradeoffs between run-time and compression. Standard compressors such as gzip and bzip are more concerned about compression and decompression times, than compressors such as ppmz and pac8, that try for maximal compression. In order to investigate how NCD's behavior changes with different compressors, we compare NCD on the "Correlated fluents" problem. We compared NCD's performance with compressors gzip, bzip2, ppmz, and pac8. Figure 2 shows the results. As one would expect, ppmz and pac8 are better at extracting commonalities, since they spend more effort in compression. An interesting feature, though, is that the shapes of the different curves are broadly similar, despite the difference in techniques.

We conjectured that the more in-depth compressors would do better when plans were jumbled. To test this, we took a single, 20-step plan for the driverlog domain, and split it into subsequences of length 2, 4, 5, and 10. Then we permuted these subsequences (we did a maximum of 100 permutations, choosing randomly when we could not exhaustively explore the permutations), and averaged the pairwise comparisons. As one would expect, ppmz and paq8 are better at identifying the underlying similarities. Note that this is an artificial test, since the "plans" here are not well-formed. But this simulates plans with causally-independent subplans.

Action Orderings. We hypothesized that because most file compressors exploit adjacency relations in their input, NCD would find it difficult to recognize certain similarities

References

Bloom, C. 2014. Ppmz website. www.cbloom.com/src/ppmz.html. Date is retrieval date.

Boddy, M. S.; Gohde, J.; Haigh, T.; and Harp, S. A. 2005. Course of action generation for cyber security using classical planning. In Biundo, S.; Myers, K. L.; and Rajan, K., eds., *International Conference on Automated Planning and Scheduling*, 12–21. AAAI.

Bryce, D. 2014. Landmark-based plan distance measures for diverse planning. In *Proceedings of the International Conference on Automated Planning and Scheduling*.

Coman, A., and Muñoz-Avila, H. 2011. Generating diverse plans using quantitative and qualitative plan distance metrics. In *Proceedings AAAI*.

Fox, M.; Gerevini, A.; Long, D.; and Serina, I. 2006. Plan stability: Replanning versus plan repair. In Long, D.; Smith, S. F.; Borrajo, D.; and McCluskey, L., eds., *ICAPS*, 212–221. AAAI.

Gailly, J., and Adler, M. 2014. Gzip website. www.gzip.org. Date is retrieval date.

Goldman, R. P.; Kuter, U.; and Schneider, A. 2012. Using classical planners for plan verification and counterexample generation. In *Proceedings of AAAI Workshop on Problem Solving Using Classical Planning*.

Hogg, C.; Kuter, U.; and Muñoz-Avila, H. 2009. Learning hierarchical task networks for nondeterministic planning domains. In *IJCAI-09*.

Hogg, C.; Kuter, U.; and Muñoz-Avila, H. 2010. Learning methods to generate good plans: Integrating htn learning and reinforcement learning. In *AAAI-10*.

Hogg, C.; Muñoz-Avila, H.; and Kuter, U. 2008. HTN-MAKER: Learning HTNs with minimal additional knowledge engineering required. In *AAAI-08*.

Kuter, U.; Burstein, M.; Benton, J.; Bryce, D.; Thayer, J.; and McCoy, S. 2014. HACKAR: helpful advice for code knowledge and attack resilience. Under review.

Lempel, A., and Ziv, J. 1976. On the complexity of finite sequences. *IEEE Transactions on Information Theory* 22(1):75–81.

Li, M., and Vitányi, P. 2008. *An introduction to Kolmogorov complexity and its applications*. Springer, third edition.

Li, M.; Chen, X.; Li, X.; Ma, B.; and Vitányi, P. M. B. 2004. The similarity metric. *IEEE Transactions on Information Theory* 50(12):3250–3264.

Mahoney, M. 2014. The PAQ data compression programs. cs.fit.edu/~mmahoney/compression/paq.html. Date is retrieval date.

McDermott, D. 1998. PDDL, the planning domain definition language. Technical report, Yale Center for Computational Vision and Control.

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*, 570–576. Menlo Park, Cal.: AAAI/MIT Press.

Nau, D.; Au, T.-C.; Ilghami, O.; Kuter, U.; Murdock, W.; Wu, D.; and Yaman, F. 2003. SHOP2: An HTN planning system. *JAIR* 20:379–404.

Nguyen, T. A.; Do, M. B.; Gerevini, A.; Serina, I.; Srivastava, B.; and Kambhampati, S. 2012a. Generating diverse plans to handle unknown and partially known user preferences. *Artificial Intelligence* 190:1–31.

Nguyen, T. A.; Do, M. B.; Gerevini, A.; Serina, I.; Srivastava, B.; and Kambhampati, S. 2012b. Generating diverse plans to handle unknown and partially known user preferences. In *Artificial Intelligence* (2012a) 1–31.

Roberts, M.; Howe, A.; and Ray, I. 2014. Evaluating diversity in classical planning. In *Proceedings ICAPS*.

Sayood, K. 2012. *Introduction to data compression*. Morgan Kaufmann, fourth edition.

Searcóid, M. O. 2007. *Metric Spaces*. SUMS. Springer.

Seward, J. 2014. Bzip2 website. www.bzip.org. Date is copyright.

Srivastava, B.; Nguyen, T. A.; Gerevini, A.; Kambhampati, S.; Do, M. B.; and Serina, I. 2007. Domain independent approaches for finding diverse plans. In Veloso, M. M., ed., *Proceedings of the International Joint Conference on Artificial Intelligence*.

Thayer, J.; Burstein, M.; Goldman, R. P.; Kuter, U.; Robertson, P.; and Laddaga, R. 2013. Comparing strategic and tactical responses to cyber threats. In *SASO Workshop on Adaptive Host and Network Security AHANS*.

Veloso, M. M. 1992. Learning by analogical reasoning in general problem solving. PhD thesis CMU-CS-92-174, School of Computer Science, Carnegie Mellon University.