Fast Downward Merge-and-Shrink

Silvan Sievers
University of Basel
Basel, Switzerland
silvan.sievers@unibas.ch

Abstract

Fast Downward Merge-and-Shrink uses the optimized, efficient implementation of the merge-and-shrink framework available in the Fast Downward planning system. We describe the techniques used in this implementation. To further push the performance of single-heuristic merge-and-shrink planners, we additionally discuss and evaluate partial merge-and-shrink abstractions, which we obtain through imposing a simple time limit to the merge-and-shrink computation.

Classical Planning

In this planner abstract, we discuss most of the concepts informally. We consider planning tasks in the SAS$^+$ representation (Bäckström and Nebel 1995), which are defined over a finite set of finite-domain variables. States are assignments over these variables. The planning task comes with a set of operators that have preconditions, effects, and a cost, and which allow to transform a state which satisfies the precondition into another state that satisfies the effect and remains unchanged otherwise. The task also specifies an initial state and a goal condition. The semantics of a planning task can naturally be described in terms of the *labeled transition system* it induces.

A labeled transition system, or transition system for short, has a set of states, a set of labels with associated costs, a transition relation that specifies the transitions which are triples of predecessor state, label, and successor state, an initial state from the set of states, and a set of goal states which is a subset of the set of states. Paths are sequences of labels that lead from a given state to some goal state. Their cost is the sum of the label costs of the sequence.

The transition system induced by a planning task consists of the states of the planning task and has transitions induced by the operators of the task, respecting the applicability of operators. Planning is the task of finding a path from the initial state to some goal state, called a plan. Optimal planning, which we are concerned with, deals with finding plans of minimal cost or proving that no plan exists.

Merge-and-Shrink

Merge-and-shrink (Dräger, Finkbeiner, and Podelski 2009; Helmert et al. 2014) is an algorithm framework to compute abstractions of transition systems. While it has very succesfully been used to compute heuristics for planning tasks (e.g., Sievers, Wehrle, and Helmert 2014; Fan, Müller, and Holte 2014; Sievers et al. 2015; Sievers, Wehrle, and Helmert 2016; Fan, Müller, and Holte 2017; Fan, Holte, and Müller 2018), it can in principle be used for any problem that can be represented as a state space which exhibits a *factored represententation*. Using such compact factored representations of both transition systems and abstraction mappings is a key aspect of merge-and-shrink that allows computing arbitrary abstractions of transition systems of interest which are generally too large to be explicitly represented.

Factored transition systems are tuples of labeled transition systems, also called factors, with the same label set that serve as a compact representation of their *synchronized product*. The synchronized product is the transition system consisting of the Cartesian product of states, where labels are used to synchronize the factors of the factored transition system via the labeled transitions: there is a transition between two states in the product system iff all factors have a transition between the corresponding component states labeled with the same label. A state is an initial/goal state in the product if all its components are initial/goal states in the respective factors.

To represent state mappings, merge-and-shrink uses *factored mappings* (Sievers 2017), which have previously also been called cascading tables (Helmert et al. 2014; Torralba 2015) and merge-and-shrink representations (Helmert, Röger, and Sievers 2015). Factored mappings are tree-like data structures where each leaf node is associated with a variable and a table that maps values of the variable to some values, and each inner node has two children factored mappings and a table that maps pairs of values computed by the children to some values. Factored mappings represent a function defined on assignments over the associated variables of all leaf nodes to some value set. To represent state mappings between factored transition systems, merge-and-shrink uses a tuple of factored mappings, called F2F mapping, that each correspond to one factor of the target factored transition system, i.e., each factored mapping computes the state mapping from states of the source factored transition system to the corresponding factor of the target factored transition system.

With the addition of generalized label reduction (Sievers, Wehrle, and Helmert 2014), the merge-and-shrink algorithm

can be understood as a framework that repeatedly applies *transformations* of a factored transition system, which essentially need to specify the transformed factored transition system and the F2F mapping that maps from the given factored transition system to the transformed one. In the context of planning, the algorithm first computes the induced factored transition system of the given task that consists of *atomic factors* which each represent a single variable of the task. It further initializes the F2F mapping to the identity mapping of the factored transition system.

In the main loop, the algorithm then repeatedly selects a transformation of the current factored transition system, choosing from the four available types of merge-and-shrink transformations: *merge* transformations replace two factors by their synchronized product, *shrink* transformations apply an abstraction to a single factor, *prune* transformations discard unreachable or irrelevant states, i.e., states from which no goal state can be reached, of a single factor, and *label reductions* map the common label set of the factored transition system to a smaller one. Applying the selected transformation means to replace the previous factored transition system by the transfomred one, and to compose the previous F2F mapping with the one of the transformation. The main loop terminates if the maintained factored transition system only contains a single factor. Together with the factored mapping, this factor induces the merge-and-shrink heuristic.

Concrete instantiations of the algorithm framework need to decide on a general strategy that decides on which type of transformation to apply in each iteration of the main loop, and it needs to provide *transformation strategies* that specify how to compute the individual transformations. For example, *shrink strategies* compute a state equivalence relation for a given transition system, reducing the size of the transition system below a given limit, and *merge strategies* decide which two factors to replace by their synchronized product.

Since our efficient implementation of the merge-and-shrink relies on label equivalence relations, we briefly discuss this concept in the context of label reductions. Sievers, Wehrle, and Helmert (2014) showed that label reductions are exact, i.e., preserve the perfect heuristic, if they only combine labels of the same cost that are Θ-*combinable* for some factor Θ of a given factored transition system F. Labels are Θ-combinable if they are *locally equivalent* in all factors $\Theta' \neq \Theta$ of F, i.e., if they label exactly the same transitions in all other factors than Θ.

For more details and a formal presentation of the transformation framework and the merge-and-shrink transformations, we refer to the work by Sievers (2017).

Implementation

In this section, we briefly mention some of the techniques used in the efficient implementation of the merge-and-shrink framework in Fast Downward (Helmert 2006). More details can be found in the work by Sievers (2017).

To represent transition systems, we do not store transitions as an adjacency list as it is commonly done to represent graphs, but rather store all transitions grouped by labels. This allows an efficient application of all merge-and-shrink transformations, as we will see below. Furthermore, we store *label groups* of locally equivalent labels for each factor, disregarding their cost (the cost of a label group is the minimum cost of any participating label). This allows storing the transitions of locally equivalent labels once rather than separately for each label.

Depending on the chosen transformation strategies, we need to compute g- and h-values of individual factors already during the merge-and-shrink computation. (Of course, we need to compute h-values in the end to compute the heuristic.) These are computed using Dijkstra's algorithm (Dijkstra 1959). This is the only place where we need an explicit adjacency list representation of transition systems.

We now turn our attention to the different merge-and-shrink transformations. When applying a shrink transformation, the shrink strategy computes a state equivalence relation for the given factor. We first compute the explicit state mapping from this equivalence relation, assigning a consecutive number to each equivalence class to allow a compact representation. Then we use this state mapping for an in-place modification of the factor by going over all transitions and updating their source and target states (compared to an adjacency list, this avoids the need to move transitions of different states), and for an in-place modification of the corresponding factored mapping by applying the state mapping to its table. From the equivalence relation on states, we get the set of new goal states.

When applying a merge transformation to the factored transition system, merging two transition systems Θ_1 and Θ_2, we do not compute the full product of states and their transitions because this would require to compute the local equivalence relation on labels from scratch after computing the product. Instead, we use a more efficient, bucket-based approach to directly compute the refinement of the local equivalence relations on labels of Θ_1 and Θ_2, collecting their transitions accordingly. Computing the factored mapping that maps states to the product factor is straightforward and merely a composition of the two component factored mappings.

When applying a prune transformation, we first determine the set of to-be-pruned states using g- and/or h-values. We prune them by entirely removing them and their transitions from the factor. The table of the corresponding factored mapping is updated to map removed states to a special symbol which is evaluated to ∞ by the heuristic.

For an efficient computation of exact label reductions based on Θ-combinability, we need to be able to efficiently refine the local equivalence relations of all (but one) factors of a factored transition system. This is possible using linked lists, which we therefore use to store label equivalence classes, i.e., label groups, for each factor. Applying the label reduction, i.e., the label mapping, is simple for all factors $\Theta' \neq \Theta$ for which we know that the reduced labels are locally equivalent: all we need to do is to relabel the set of transitions of the reduced labels, remove the labels from their group and add the new label to it. For the factor Θ, we need to collect all transitions of all reduced labels and combine them to form the transitions of the new label. We update the local equivalence on labels by removing reduced labels from their (different) groups and the groups themselves if

they become empty, and by adding a new singleton group for the new label.

Partial Merge-and-Shrink Abstractions

The literature on merge-and-shrink so far always considered computing merge-and-shrink abstractions over all variables of a given planning task to the best of our knowledge. That is, the main loop of the algorithm is stopped only if the factored transition systems contains a single factor. However, there is no conceptual or technical reason to not stop the algorithm early, ending up with several factors and factored mappings that can be used to compute several induced *factor heuristics*. Additionally, we observed that state-of-the-art merge-and-shrink planners fail to finish computing the abstraction in the given time and memory limits in a non-negligible number of cases (152–272 out of 1667 tasks for state-of-the-art-configurations[1]). As a simple stop-gap measure, we added a time limit to the merge-and-shrink algorithm, allowing to terminate the computation even before having computed all atomic factors.

When the computation terminates early, we face the decision of computing a heuristic from the set of factor heuristics induced by the remaining factors. A straight-forward way is to compute the *max-factor heuristic* (h^{mf}_F) that maximizes over all factor heuristics. The second, presumably less expensive alternative is to choose a *single* factor heuristic (h^{sg}_F) and use it as the merge-and-shrink heuristic. We use the following simple rule of thumb for this choice: we prefer the heuristic with the largest estimate for the initial state (rationale: better informed heuristic), breaking ties in favor of larger factors (rationale: more fine-grained abstraction), and choose a random heuristic among all remaining candidates of equal preference.

In the following, we briefly evaluate using partial merge-and-shrink abstractions to compute merge-and-shrink heuristics. To do so, we ran our planner on all (optimal) benchmarks of all all IPCs up to 2014, a set comprised of 1667 planning tasks distributed across 57 domains.[2], using A* search in conjunction with different merge-and-shrink heuristics. We limit time to 30 minutes and memory to 3.5 GiB per task, using Downward-Lab (Seipp et al. 2017) for conducting the experiments on a cluster of machines with Intel Xeon Silver 4114 CPUs running at 2.2 GHz.

Both variants of our planner, FDMS1 and FDMS2, use the state-of-the-art shrink strategy based on bisimulation (Nissim, Hoffmann, and Helmert 2011) with a size limit of 50000, always allowing (perfect) shrinking. We use full pruning of dead states and exact label reductions based on Θ-combinability with a fixed point algorithm using a random order on factors. FDMS1 uses the merge strategy based on *strongly connected components* of the causal graph (Sievers, Wehrle, and Helmert 2016), which uses DFP (Sievers, Wehrle, and Helmert 2014) for internal merging (SCCdfp). FDMS2 uses the merge strategy *score-based MIASM* (sbMIASM, previously also called DYN-MIASM), which is a simple variant of the entirely precomputed merge strategy *maximum intermediate abstraction size minimizing* (Fan, Müller, and Holte 2014).

		h^{sg}			h^{mf}			
	base	450s	900s	1350s	450s	900s	1350s	
Coverage	802	835	**836**	**836**	**836**	**836**	835	FDMS2
# constr	1395	**1637**	1629	1615	1636	1629	1614	FDMS2
Constr time	241.90	135.99	197.57	230.70	**135.73**	196.58	229.45	FDMS2
Constr oom	**21**	**21**	**21**	**21**	**21**	**21**	**21**	FDMS2
Constr oot	251	**9**	17	31	10	17	32	FDMS2
E 75th perc	**1342k**	1368k	**1342k**	**1342k**	1368k	**1342k**	**1342k**	FDMS2
Coverage	814	**844**	**844**	842	**844**	**844**	841	FDMS1
# constr	1505	**1622**	1620	1611	**1622**	1621	1611	FDMS1
Constr time	97.93	61.62	80.59	91.17	**61.29**	79.82	89.84	FDMS1
Constr oom	**21**	**21**	**21**	**21**	**21**	**21**	**21**	FDMS1
Constr oot	141	**24**	26	35	**24**	25	35	FDMS1
E 75th perc	**1860k**	**1860k**	**1860k**	**1860k**	**1860k**	**1860k**	**1860k**	FDMS1

Table 1: Compassion of the baseline against two versions of partial merge-and-shrink, using different time limits.

Table 1 shows the number of solved tasks (coverage), the number of tasks for which the heuristic construction completed (# constr), the runtime of the heuristic construction (constr time), the number of failures of the heuristic construction due to running out of time (constr oot) or memory (constr oom), and the number of expansions until the last f-layer. The table compares the baseline (base) with the two variants of computing a single merge-and-shrink heuristic (h^{sg} and h^{mf}) using time limits of 450s, 900s, and 1350s.

As expected, adding a time limit is a very effective measure for greatly increasing the number of successful heuristic constructions, which also directly transfers to a significant increase in coverage of all configurations, with 900s being a sweet spot for both planners. Stopping the computation early does *not* affect the heuristic quality as one might have expected. The likely reason is that with limiting the time, we catch precisely those tasks for which the construction otherwise does not terminate or terminate too late for a successful search. Tasks which we can already solve without imposing a time limit (base) usually require a rather short construction time, and therefore limiting the time to 900s or more does not stop the heuristic computation early and hence does not reduce heuristic quality in these cases.

We also observe that there is no significant difference between h^{mf} and h^{sg}. While h^{mf} theoretically dominates any factor heuristic by definition, evaluating the former can be slightly more expensive. Furthermore, in scenarios where there is one large factor and many small (e.g., atomic) factors in the end, the large one likely dominates the others, and thus h^{sg} is equally informed as h^{mf}.

Our submissions to the competition both use a time limit of 900s. In addition to pure A* search with the described merge-and-shrink heuristics, they use pruning based on partial order reduction by using simple stubborn sets (Wehrle and Helmert 2014). We extended the implementation in Fast

[1] See rows "# constr" of column "base" of Table 1.

[2] From the collection at `https://bitbucket.org/aibasel/downward-benchmarks`, we use the "optimal strips" benchmark suite.

Downward with a mechanism that disables pruning if, after the first 1000 expansions, only 10% or fewer states have been pruned, and with support for conditional effects. We further use pruning based on structural symmetries (Shleyfman et al. 2015) by using the DKS algorithm (Domshlak, Katz, and Shleyfman 2012). Finally, after translating PDDL with the translator of Fast Downward (Helmert 2009), we also post-process the resulting SAS^+ representation using the implementation of h^2 mutexes by Torralba and Alcazar.

References

Bäckström, C., and Nebel, B. 1995. Complexity results for SAS^+ planning. *Computational Intelligence* 11(4):625–655.

Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. *Numerische Mathematik* 1:269–271.

Domshlak, C.; Katz, M.; and Shleyfman, A. 2012. Enhanced symmetry breaking in cost-optimal planning as forward search. In McCluskey, L.; Williams, B.; Silva, J. R.; and Bonet, B., eds., *Proceedings of the Twenty-Second International Conference on Automated Planning and Scheduling (ICAPS 2012)*, 343–347. AAAI Press.

Dräger, K.; Finkbeiner, B.; and Podelski, A. 2009. Directed model checking with distance-preserving abstractions. *International Journal on Software Tools for Technology Transfer* 11(1):27–37.

Fan, G.; Holte, R.; and Müller, M. 2018. Ms-lite: A lightweight, complementary merge-and-shrink method. In *Proceedings of the Twenty-Eighth International Conference on Automated Planning and Scheduling (ICAPS 2018)*. AAAI Press.

Fan, G.; Müller, M.; and Holte, R. 2014. Non-linear merging strategies for merge-and-shrink based on variable interactions. In Edelkamp, S., and Barták, R., eds., *Proceedings of the Seventh Annual Symposium on Combinatorial Search (SoCS 2014)*, 53–61. AAAI Press.

Fan, G.; Müller, M.; and Holte, R. 2017. Additive merge-and-shrink heuristics for diverse action costs. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 4287–4293. AAAI Press.

Helmert, M.; Haslum, P.; Hoffmann, J.; and Nissim, R. 2014. Merge-and-shrink abstraction: A method for generating lower bounds in factored state spaces. *Journal of the ACM* 61(3):16:1–63.

Helmert, M.; Röger, G.; and Sievers, S. 2015. On the expressive power of non-linear merge-and-shrink representations. In Brafman, R.; Domshlak, C.; Haslum, P.; and Zilberstein, S., eds., *Proceedings of the Twenty-Fifth International Conference on Automated Planning and Scheduling (ICAPS 2015)*, 106–114. AAAI Press.

Helmert, M. 2006. The Fast Downward planning system. *Journal of Artificial Intelligence Research* 26:191–246.

Helmert, M. 2009. Concise finite-domain representations for PDDL planning tasks. *Artificial Intelligence* 173:503–535.

Nissim, R.; Hoffmann, J.; and Helmert, M. 2011. Computing perfect heuristics in polynomial time: On bisimulation and merge-and-shrink abstraction in optimal planning. In Walsh, T., ed., *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, 1983–1990. AAAI Press.

Seipp, J.; Pommerening, F.; Sievers, S.; and Helmert, M. 2017. Downward Lab. `https://doi.org/10.5281/zenodo.790461`.

Shleyfman, A.; Katz, M.; Helmert, M.; Sievers, S.; and Wehrle, M. 2015. Heuristics and symmetries in classical planning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*, 3371–3377. AAAI Press.

Sievers, S.; Wehrle, M.; Helmert, M.; Shleyfman, A.; and Katz, M. 2015. Factored symmetries for merge-and-shrink abstractions. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*, 3378–3385. AAAI Press.

Sievers, S.; Wehrle, M.; and Helmert, M. 2014. Generalized label reduction for merge-and-shrink heuristics. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI 2014)*, 2358–2366. AAAI Press.

Sievers, S.; Wehrle, M.; and Helmert, M. 2016. An analysis of merge strategies for merge-and-shrink heuristics. In *Proceedings of the Twenty-Sixth International Conference on Automated Planning and Scheduling (ICAPS 2016)*, 294–298. AAAI Press.

Sievers, S. 2017. *Merge-and-shrink Abstractions for Classical Planning: Theory, Strategies, and Implementation*. Ph.D. Dissertation, University of Basel.

Torralba, Á. 2015. *Symbolic Search and Abstraction Heuristics for Cost-Optimal Planning*. Ph.D. Dissertation, Universidad Carlos III de Madrid.

Wehrle, M., and Helmert, M. 2014. Efficient stubborn sets: Generalized algorithms and selection strategies. In *Proceedings of the Twenty-Fourth International Conference on Automated Planning and Scheduling (ICAPS 2014)*, 323–331. AAAI Press.