A note on the analysis of theorem-proving strategies

Maria Paola Bonacina
(bonacina@cs.uiowa.edu)

In recent work [3] David Plaisted proposed an approach to analyze *the complexity of theorem-proving strategies*, and applied it to *Horn propositional logic*. Most of the research in complexity of theorem proving studies the complexity of propositional proofs, while most of the work on search concentrates on the design of heuristics[1]. To my knowledge, Plaisted's approach has been the first one to treat the complexity of *searching for a proof*. This note is a comment on this ground-breaking work. The framework of [3] is a starting point to discuss open problems that need to be addressed in order to analyze theorem-proving strategies in first-order logic. These include the formalization of the search plan, the representation of contraction, the duality of forward and backward reasoning, and the definition of complexity measures for infinite search spaces.

A theorem-proving strategy is made of an inference system and a search plan. The inference system defines the possible inferences, and the search plan selects at each step of a derivation the inference rule and the premises for the next step. *Contraction-based strategies* include the strategies that integrate resolution, Knuth-Bendix completion and term rewriting. These strategies work primarily by forward reasoning. They use *expansion rules* such as resolution to generate consequences from existing clauses. Forward-reasoning methods do not generally distinguish goal clauses from the others, seeking to obtain a contradiction from the whole set. Generated clauses are kept and used for further inferences, so that the strategy works on a *database of clauses*. A major source of redundancy for these methods is that they generate and retain clauses that do not contribute to proving the theorem. Contraction-based strategies counter this problem by using *contraction rules*, such as simplification and subsumption, to delete redundant clauses. Since the inference system features multiple inference rules, and generated clauses are kept, creating many possible choices of premises, the search plan plays an important role. Forward-reasoning strategies admit various search plans, which sort the inferences by different criteria. Contraction-based strategies employ *eager-contraction* search plans, that give priority to contraction over expansion, in order to maintain the database minimal with respect to the contraction rules.

Subgoal-reduction strategies include model elimination, tableau-based methods, Prolog technology theorem proving, the MESON strategy, and the problem-reduction-format strategies. These strategies work by reducing the goal to subgoals. Each inference step involves the current goal and at most another premise. The current goal is initially a selected input clause, and then the most recently generated goal. If no rule applies to the current goal, the strategy backtracks to the previous one. Since generated clauses are goals retained for backtracking, the strategy works on a *stack of goals*. Since there is no database, the application of contraction is limited, and the search plan is fixed, usually *depth-first search with iterative deepening*. A source of redundancy for these strategies is that by focusing on the most recently generated goal, they may reduce repeatedly the same subgoals. Subgoal-reduction methods counter this problem by using techniques for *lemmaizing* or *caching*, that enable them to keep track of already solved or failed subgoals.

[1] See [3] and [1] for most references relevant to this note.

Plaisted's approach to the analysis of strategies

The model of the search space

A theorem-proving strategy is formalized in [3] as a 5-tuple $< S, V, i, E, u >$, where S is a set of states, and E is a set of pairs of states, in such a way that $< S, E >$ forms a directed graph. The component V is a set of labels for the states in S. The component i is a mapping from the set of input clauses to S, which identifies the initial state(s). The component u is a function from S to $\{true, false\}$, where $u(s) = true$ if and only if s is a state where unsatisfiability is detected. It is assumed that u is an almost trivial test, such as recognizing that a set of clauses contains the empty clause. A more precise definition of V and u depends on the specific strategy. A function *label*, such that $label(s)$ denotes the label of state s, is also used. It is required that no two distinct edges (s_1, t_1) and (s_2, t_2) in E have $t_1 = t_2$. This implies that the directed graph is a set of *trees*. A strategy is said to be *linear* if for all states s, there is a unique state t such that $(s, t) \in E$, that is, every state has unique successor.

Given a strategy $G =< S, V, i, E, u >$ and an input set of clauses R, a state $s \in S$ is *reachable* from R if there is a path from a state in $i(R)$ to s. $S(R)$ denotes the subset of S containing all states that are reachable from R, and $E(R)$ denotes the restriction of E to $S(R)$. Then, $G(R) =< S(R), E(R) >$ is the search space for the theorem-proving problem R according to the strategy G.

For *resolution*, the labels of the states are finite sets of clauses. If R is the input set of clauses, and $s_0 \in S$ is the state with label R, the input function $i\colon R \rightarrow S$ is the function such that $i(\psi) = s_0$ for all $\psi \in R$. In other words, all input clause are mapped to a single initial state, whose label is the set of input clauses. It follows that $G(R)$ is *a tree* with R as the label of the root. An arc connects state s to state t if the label of t is the union of the label of s and all the resolvents in s according to the inference system. This means that each state has a unique successor, that is, the tree $G(R)$ for resolution degenerates to *a list*. Accordingly, all resolution-based strategies are *linear* in [3]. The function u is defined by $u(s) = true$ if and only if the label of s contains the empty clause.

For *model elimination*, each state is labelled by a single chain[2], which is the current goal in that state. If $R = \{\psi_1, \ldots, \psi_n\}$ is the input set of clauses, there are states $s_1, \ldots, s_n$ with labels $\{\psi_1\}, \ldots, \{\psi_n\}$, respectively, and $i(\psi_j) = s_j$, for all $1 \leq j \leq n$. In other words, there is an initial state for each input clause. This captures the fact that any input clause can be chosen as the initial chain. It follows that $G(R)$ is *a set of trees*, one for each possible initial chain. An arc connects state s to state t if the chain labelling t is generated from the chain labelling s in one ME-step (ME-extension or ME-reduction or ME-contraction). The function u is defined in the same way as for resolution. The strategies based on model elimination are *not linear*, because an arc represents a single step, and more than one step may be applicable to a chain (e.g., ME-extension steps with different input clauses).

[2] A *chain* is a sorted clause with plain literals, called B-literals, and framed literals, called A-literals for ancestors.

The complexity measures to evaluate and compare the strategies

The measures of complexity of search proposed in [3] aim at measuring *the total size of the search space*. The total size of $G(R)$ is defined as $||G(R)|| = \sum_{s \in S(R)} |label(s)|$. For resolution, the labels are finite sets of clauses, and therefore $||G(R)||$ is the sum of the cardinalities of the sets of clauses labelling the nodes of $G(R)$. For model elimination, the labels are singleton sets, so that $||G(R)||$ is equal to the number of nodes in $G(R)$.

The measure $||G(R)||$ is refined into three measures, called *duplication by iteration*, *duplication by case analysis* and *duplication by combination*. Duplication by iteration is the maximum length of a path in $G(R)$. Duplication by case analysis is the maximum size of a subset of $S(R)$ no two elements of which are on the same path. Duplication by combination is the maximum cardinality of a label of a state in $S(R)$, i.e., $max_{s \in S(R)}|label(s)|$. Since $G(R)$ is a tree or a set of trees, duplication by case analysis reduces to the number of paths, and $||G(R)||$ is bounded by the product of duplication by iteration, duplication by case analysis and duplication by combination. The analysis proceeds by establishing whether these measures are exponential or linear or constant in the length of the input set R read as a single string.

The analysis of the strategies in Horn propositional logic

According to the analysis in [3], basic resolution strategies have linear duplication by iteration and exponential duplication by combination. The duplication by case analysis is trivially constant (more precisely, equal to 1), because each state has unique successor. The exponential duplication by combination captures the complexity of generating and keeping clauses, since resolvents are generated by combining in different ways the input literals. Since one of the measures is exponential, resolution strategies are regarded as inefficient. For positive hyperresolution, however, the duplication by combination is linear, because positive hyperresolvents in Horn logic are unit clauses, so that all generated clauses are unit clauses, and literals are not combined. Also, the duplication by combination of positive resolution reduces from exponential to linear if the strategy is equipped with an ordering on predicate symbols, and only the negative literals with smallest predicate are resolved upon. In summary, resolution strategies are either efficient, but not goal-sensitive (e.g., positive hyperresolution and positive resolution), or goal-sensitive, but not efficient (e.g., negative resolution), or neither efficient nor goal-sensitive (e.g., ordered resolution). These results are worst-case results. For instance, ordered resolution is efficient for some sets of clauses and orderings (e.g., all well-ordered sets[3]), but there are sets for which an ordering that makes the strategy efficient cannot be found. The same is true for goal-sensitivity.

Model elimination has exponential duplication by iteration and exponential duplication by case analysis. The duplication by combination is trivially constant (equal to 1), because each state is a singleton. The exponential duplication by iteration captures the redundancy of solving subgoals repeatedly. Duplication by case analysis is also exponential because a state may have multiple successors. Thus, resolution and model elimination are sort of dual: resolution has low

[3] A set of Horn clauses is *well-ordered* if there is a partial ordering $<$ such that if $P :- P_1, \ldots, P_n$ is a clause in the set then $P_i < P$ for all i, $1 \leq i \leq n$.

duplication by case analysis and iteration, but high duplication by combination, and vice versa for model elimination.

If model elimination is enriched with unit lemmaizing[4], duplication by iteration becomes linear, because lemmaizing prevents solving repeatedly the same successful subgoals. Lemmatization adds a forward-reasoning character to the strategy, because lemmas are generated and retained. In the framework of [3] this means exponential duplication by combination and constant duplication by case analysis, so that model elimination with lemmaizing has the same duplication as the resolution strategies. In Horn logic model elimination can be enhanced with caching[5]. The use of caching reduces the duplication further, because not only successful subgoals (success caching), but also failed subgoals (failure caching) are not repeated. Assuming depth-first search with iterative deepening, duplication by combination becomes linear (more precisely, quadratic). Similar results apply to the other subgoal-reduction strategies. In summary, subgoal-reduction strategies are inefficient, but goal-sensitive. Subgoal-reduction strategies with caching are efficient and goal-sensitive.

A caveat Working by worst-case analysis means exhibiting a set of clauses where a strategy performs poorly. Examples are the parametric sets S_n^2 and A_n that give the exponential upper bounds on duplication by combination for negative resolution and ordered resolution, respectively. Worst-case sets may display regularities of structure, such as symmetries, or recurrence relations. The worst-case sets used in [3] incorporate recurrence relations on the indices of the literals in the clauses. The recurrence relations have exponential solution, and this is used in the proofs of the exponential behaviour of the strategies. It is not known whether these patterns occur in real derivations, and how often. Also, worst-case sets are hard from a combinatorial point of view for the targeted strategy, but may be easy in the practice of theorem proving. S_n^2 is satisfiable and contains no positive clauses, so that positive hyperresolution or positive resolution would not apply a single inference and establish that the set is satisfiable in the time needed to read it. Similarly, A_n is satisfiable and contains neither positive nor negative clauses, so that it is trivial also for negative resolution.

Discussion

The following comment is written having first-order logic in mind, which is also the final purpose of the approach of [3].

Infinite search spaces

The complexity of search is measured in [3] by measuring the total size of the search space in the worst-case. The total size of the search space can be measured only if the search space is *finite*, as in the propositional problems considered in [3]. Thus, the first problem is:

[4]In Horn logic all lemmas generated by lemmaizing are unit lemmas.

[5]Caching is not consistent with ME-reduction, which is necessary for the completeness of model elimination in first-order logic. In first-order logic one may use lemmaizing or more complicated caching schemes.

1. Explore complexity measures and modes of analysis that are suitable for *infinite* search spaces, such as those of first-order problems, whose total size cannot be measured.

The search plan

Since the intent of [3] is to measure the total size of the search space, and the latter depends on the inference system, not on the search plan, the model of search in [3] does not represent the search plan and its application. However, the search plan is important in the practice of first-order theorem proving. A prover may find a proof with a certain search plan and run out of memory with another, or find a proof hours later. Therefore, the following questions remain open:

2. Give the mathematical definition of a search plan.
3. Represent the application of a search plan to a search space, that is *the search process*.
4. Compare strategies with the same inference system but different search plans.

Contraction

The analysis in [3] considers first all resolution-based methods as expansion-only strategies. Results for strategies with subsumption and clausal simplification are obtained by modifying the results for the expansion-only versions of the strategies. For most strategies, the analysis of contraction in [3] consists in showing that subsumption and clausal simplification do not apply in the sets of clauses used to establish the worst-case results, e.g., S_n^2 and A_n for negative resolution and ordered resolution, respectively. For positive resolution, since in Horn logic positive clauses are unit clauses, each positive-resolution step generates a resolvent that subsumes its non-unit parent. Thus, if a positive resolution strategy applies subsumption after each resolution step, the duplication by combination reduces from exponential to linear.

Because the analysis is conducted in this way, the formalism of [3] does not include contraction inferences as a basic element. The arcs in the directed graph represent only expansion inferences, such as the addition of resolvents, or the generation of a successor chain from a chain in model elimination. The choice of representing resolution inferences by arcs that add all resolvents is an obstacle to including contraction in the model. For instance, the search space of a strategy that applies subsumption after every resolution step cannot be represented in this model, because subsumption by a resolvent may delete the parent clauses of other resolution steps that were originally enabled.

Contraction is most useful in practice in equational logic and first-order logic. For these problems the search space is *infinite*. Therefore, the approach of exhibiting a worst-case set and showing that contraction does not affect it may not apply, so that the following problem remains:

5. Provide a model of the search space and search process which includes contraction as a basic element.

Contraction-based and subgoal-reduction strategies

The search space of subgoal-reduction strategies has been traditionally represented as a tree (e.g., AND/OR-tree). The nodes are labelled by the goals (e.g. the chains of model elimination), and the input clauses are viewed as "operators" that may be applied to reduce the goals. The treatment of subgoal-reduction strategies in [3] follows this pattern. On the other hand, there is no standard for the representation of the search space of contraction-based strategies (e.g., the model of [2] is for expansion-only strategies). The search space formalism of [3] does not change this state of affairs.

The approach of [3] stipulates that an arc represents the addition of all the resolvents, and the search space of resolution is a list. (If an arc represents a single resolution inference, the search space is a general graph as in [2].) This move of [3] may appear surprising. First, it makes the semantics of the formalism ambiguous, because an arc represents an inference for model elimination and all the possible inferences in the given state for resolution. Second, it has the counterintuitive effect that resolution is linear, whereas model elimination is not. A possible interpretation of this move is that the approach of [3] implicitly envisions resolution as a subgoal-reduction strategy, where the whole database of clauses is the goal (to be reduced to the empty clause) and the only operator is adding all resolvents. This interpretation is supported by the following observations:

- The tree structure, which is natural for subgoal-reduction strategies, but not for contraction-based strategies, is assumed already in the general characterization of a strategy as a tuple $< S, V, i, E, u >$. Then, the search space of subgoal-reduction strategies is represented properly as a set of trees, while the search space of resolution is reduced to a list.

- Resolution has a very high degree of non-determinism, because of all the possible selections of clauses in the database. However, it is represented as a completely deterministic strategy. On the other hand, the "degrees of freedom" of model elimination (e.g., choice of input clause for ME-extension steps and choice of initial chain) are represented (e.g., a state may have multiple successors and there is a tree for each choice of initial chain).

- The importance of the search plan is directly proportional to the degree of non-determinism of the inference system. Thus, the choices of making resolution deterministic and excluding the search plan from the representation are related. This does not affect subgoal-reduction strategies, for which depth-first search with iterative deepening is assumed. On the other hand, the search plan is fundamental for the modelling of forward-reasoning strategies.

- The positive effect of lemmaizing and caching in reducing the duplication of subgoal-reduction strategies is successfully captured. On the other hand, contraction, which counters the duplication of forward reasoning, is not included in the model.

Thus, we have the following problem:

6. Provide a common framework which is sufficiently rich to analyze subgoal-reduction, expansion-only and contraction-based strategies.

A different approach to the modelling of search and the analysis of strategies appeared in [1]. It presents the beginning of an approach to Problems 1, 2, 3, and 5. More problems remain open, as the field of strategy analysis is only at its infancy.

Acknowledgements

Thanks to David Plaisted, for answering my questions on his paper, and to Jieh Hsiang, for our discussions on Plaisted's work.

References

[1] M. P. Bonacina and J. Hsiang. On the representation of dynamic search spaces in theorem proving. In C.-S. Yang (ed.), *Proc. of the Int. Computer Symposium*, 85–94, 1996, and Technical Report, Dept. of Computer Science, Univ. of Iowa, 95-05.

[2] R. Kowalski. Search strategies for theorem proving. In B. Meltzer and D. Michie (eds.), *Machine Intelligence 5*, 181–201, Edinburgh University Press, 1969.

[3] D. A. Plaisted. The search efficiency of theorem proving strategies. In A. Bundy (ed.), *Proc. of CADE-12*, Springer Verlag, LNAI 814, 57–71, 1994, and Technical Report MPI I-94-233.