

Instance Based Methods — A Brief Overview

Peter Baumgartner and Evgenij Thorstensen

Instance-based methods are a specific class of methods for automated proof search in first-order logic. This article provides an overview of the major methods in the area and discusses their properties and relations to the more established resolution methods. It also discusses some recent trends on refinements and applications.

This overview is rather brief and informal, but we provide a comprehensive literature list to follow-up on the details.

1 Introduction

Automated Reasoning (AR) is an application-relevant subfield of both Logic in Computer Science and Artificial Intelligence. The main tool is mathematical logic, which provides both a language for modeling a domain of interest and inference mechanisms for logical reasoning. AR involves the design of efficient reasoning algorithms based on logical calculi, their software implementation and application. AR techniques have been proven useful in areas such as Computer-Aided Verification, Database Systems, Programming Languages, Computer Security, Artificial Intelligence, Logic Programming, and Software Engineering.

In this article we focus on AR in first-order logic (FOL), which has a long-standing tradition in AI. The perhaps best-known methods are based on the resolution calculus, which dates back to 1965 [Rob65], or on analytic tableaux methods. (For example, Prolog’s SLD-resolution can be appropriately described as a tableau method.) With the development of *instance based methods (IMs)* since the 1990s, a comparably new family of AR methods for FOL is available now. While also being sound and complete, IMs typically explore a different search space and exhibit different termination behaviour, which makes them attractive from a practical point of view as a complementary method. For instance, IMs naturally provide decision procedures for a certain fragment of first-order logic, the Bernays-Schönfinkel class, which currently attracts a lot of attention as a compact knowledge representation language. See also [Pla94] for a comparison of various calculi and strategies, including an instance based method.

2 Ideas and History

Like many other AR methods, IMs assume the input formula given as a set of clauses. A *clause* is an implicitly universally quantified formula of the form $L_1 \vee \dots \vee L_n$, where each L_i is a *literal*, i.e. an atomic formula or its negation (also known as a *positive* or *negative* literal, respectively). The members of a given clause set are implicitly connected by “and”.

As there are efficient translations from a more general first-order logic syntax into clause logic [NW01, e.g.], the assumption above does not lead to a loss of generality. Determining the validity of a given formula—the most common reasoning problem—translates into a question of unsatisfiability of the clause set obtained from the negation of the given formula

(“refutational theorem proving”).

The general idea behind IMs is to prove the unsatisfiability of a clause set by generating sets of ground instances of its clauses, then checking the unsatisfiability of these sets. This is a sufficient test for the unsatisfiability of the given clause set, because, if a set of instances is not satisfiable then neither is the given clause set. Consider an example that does this in a very inefficient way. We generate ground instances (i.e. variable-free instances) of every clause in the clause set below by looping over possible constants to try, then check to see, e.g. with a propositional logic SAT solver, whether the generated instances are unsatisfiable (x, y, z are variables, a, b are constants).

$$P(x, y) \quad \neg P(a, z) \vee Q(a, z) \quad \neg P(b, z) \vee R(b, z) \quad \neg R(b, c)$$

We could start by generating a set of instances where every variable is replaced by a constant, say a :

$$P(a, a) \quad \neg P(a, a) \vee Q(a, a) \quad \neg P(b, a) \vee R(b, a) \quad \neg R(b, c) .$$

This set is satisfiable, so we continue the loop. If the enumeration of instances is done in a fair way, we will eventually end up with an unsatisfiable set, such as

$$P(b, c) \quad \neg P(a, c) \vee Q(a, c) \quad \neg P(b, c) \vee R(b, c) \quad \neg R(b, c)$$

and we can stop.

There are two important points of inefficiency here. The first one is that the clause $\neg P(a, z) \vee Q(a, z)$ is superfluous, as the clause set is unsatisfiable without it. We should thus avoid generating instances of this clause altogether. The second point of inefficiency is the lack of direction in the instantiation process, leading to many dead-ends as we generate instances that do not lead to an unsatisfiable set. Usually, the dead-ends can be avoided by keeping the generated sets, but one still has to deal with all the superfluous instances generated. Contemporary IMs try to avoid these and other problems in order to obtain practically useful methods.

The idea explained above is already present in the work by Davis, Putnam, Logemann and Loveland, and others, in the early sixties of the last century [DP60, DLL62]. One of their algorithms to check the satisfiability of the sets of instances is still popular as the basis of modern propositional SAT solvers, commonly referred to as the “DPLL procedure”. However, because of its primitive treatment of quantifiers by uninformed guessing,

their first-order logic procedures have quickly been overshadowed by a reasoning procedure developed soon afterwards, the *resolution calculus* [Rob65].

One of the key insights in [Rob65] concerns the use of MGUs (most general unifiers) for reasoning on clauses, as in the *resolution inference rule*:¹

$$\frac{C \vee K \quad L \vee D}{(C \vee D)\sigma} \quad \text{if } \sigma \text{ is a MGU of } K \text{ and } \bar{L}$$

That σ is a *most general* unifier means that, given any unifier τ (a substitution that makes the involved literals equal), there is a substitution γ such that $\sigma\gamma = \tau$. The notation \bar{L} refers to the complement of L .

In contrast to resolution, IMs work with instances of clauses without combining them into new clauses as the resolution inference rule does. With that view, bottom-up model generation procedures like SATCHMO [MB88] and hyper tableau [BFN96] qualify as IMs. They employ unification at the core of their inference rules, but still require blind guessing of ground instances in certain circumstances. A stream of research that avoids this was initiated with Lee and Plaisted's first IM, the Hyper-Linking calculus [LP92]. Hyper-Linking is akin to the instantiation loop described above, but uses unification to guide the instantiation of clauses and capitalizes on the latest SAT solver technology for the unsatisfiability check of the instance sets. Since then, other IMs have been developed by Plaisted and his coworkers [CP94, PZ00]. Another influential approach is Billon's disconnection calculus [Bil96], which was picked up by Letz and Stenz and has been significantly developed further since then into a tableaux-like IM [LS01, SL04]. To some extent, the tableau structure of Disconnection Tableaux enables one to avoid the problem of superfluous clauses.

One of the authors of this paper introduced a first-order version of the propositional DPPL procedure mentioned above, First-Order DPPL (FDPLL) [Bau00], which is now subsumed by the Model Evolution (ME) calculus [BT03]. These two calculi are less concerned with generating sets of instances of the clauses themselves; instead they focus on finding a potential model for the clause set by a semantic-tree construction. The model representation formalism in the ME calculus has been studied in its own right by [FP05].

Other IMs have also been described in [Bau98, BEF99], by Hooker [HRCS02], and by Ganzinger and Korovin [GK03]. See [JW07] for a thorough comparison of some of these IMs. A rather recent development is [dMB08], which employs tuple-at-a-time reasoning for sets of ground instances of clauses, which are represented by BDD.

In the following we describe the idea behind some of these methods in more detail. A key notion to several of them is that of a *link* between two clauses. Given two clauses $C \vee K$ and $L \vee D$, the literals L and K constitute a *link* if there is an MGU σ for L and \bar{K} .

¹It took another 25 years until the development of the "modern" theory of resolution had begun in the 1990s [BG90]. This led to a breakthrough in resolution theory by unifying more or less all resolution variants and improvements until then in a single theoretical framework, yet more elegant, general and powerful [BG01].

3 Inst-Gen

Inst-Gen [GK03] is perhaps the conceptually simplest IM. Unlike the unguided IM in Section 2, Inst-Gen uses unification to generate clause instances, which is realized with the following inference rule:

$$\frac{C \vee K \quad L \vee D}{(C \vee K)\sigma \quad (L \vee D)\sigma} \quad \text{if } \sigma \text{ is a MGU of } K \text{ and } \bar{L}$$

Notice that the side condition in the rule can be replaced by " L and K constitute a link with unifier σ ". The rule can be strengthened by requiring that at least one conclusion clause must be a *proper* instance of its premise, that is, the conclusion must be an instance of its premise but not the other way round.

The Inst-Gen inference rule differs from the resolution rule by keeping the instantiated premises separate instead of combining them into one resolvent. While resolution derives the empty clause \perp to indicate unsatisfiability, Inst-Gen uses a SAT solver to check propositional unsatisfiability of the clause set after instances have been added. For that, every variable in every clause is uniformly replaced by the same constant. Intuitively, by this process an unsatisfiable clause set gets closer and closer to being propositionally unsatisfiable.

The beauty of Inst-Gen is that it is "trivially" sound and not so difficult to prove complete. It is sound because it always adds instances of clauses already present, hence consequences thereof. To get a deeper understanding of how the calculus works it is instructive to sketch its completeness proof. As a prerequisite for that, let \perp denote the substitution already mentioned above that uniformly replaces every variable by some fixed constant. Also, assume as given a clause set M that is closed under the application of the Inst-Gen inference rule, modulo alphabetic variants of clauses. Completeness then amounts to showing that if M is unsatisfiable then $M\perp$ is likewise unsatisfiable. It is advantageous, however, to work in the contrapositive direction. Thus assume that $M\perp$ is satisfiable; it suffices to construct a model for M .

Because $M\perp$ is satisfiable, there is a satisfiable *path* P through $M\perp$, that is, a set of literals that is obtained by picking exactly one literal from each clause in $M\perp$ and such that P does not contain a link, i.e., two complementary literals.

We sketch how P guides the construction of a model I for M , better said, a partial model for M that can be extended to a total one (a partial model is a consistent set of ground literals): initially let I be the empty set. For every clause $C \in M$ and every ground instance $C\gamma$ do the following: if

1. C is the *most specific representation* of $C\gamma$, that is, there is no clause $D \in M$ that is a proper instance of C such that $C\gamma = D\gamma'$ for some γ' , and
2. $C\gamma$ is false in I , and
3. K is a literal in C such that $K\perp \in P$ and $K\gamma$ is undefined in I (i.e., neither $K\gamma$ nor its complement $\bar{K}\gamma$ are in I)

then add $K\gamma$ to I . In this case we say that $C\gamma$ *generates* $K\gamma$ (in I).

Now, assume to the contrary that I falsifies some ground instance $D\gamma$ of a clause $D \in M$. We show that "enough" inferences must have been applied so that I satisfied $D\gamma$, this way contradicting the assumption. Without loss of generality assume that D is a most specific representation of $D\gamma$ (among all representations there is always a most specific one). Let

$L \in D$ be a literal such that $L \perp \in P$ (recall that P contains a literal from every clause in $M \perp$, thus we can find such an L). As $D\gamma$ is false in I , $D\gamma$ cannot be generating and we must have $\bar{L}\gamma \in I$. That is, some clause $C\gamma$ generates $\bar{L}\gamma$ in I . Let $K \in C$ be a literal such that $K\gamma = \bar{L}\gamma$. Now, the Inst-Gen inference rule is applicable to C and D by using K and L as the link literals. Let σ be the MGU of K and \bar{L} . By condition 3 above it holds that $K \perp \in P$. As $L \perp \in P$ the MGU σ must replace at least one variable in K or in L by a non-variable term, otherwise P would contain complementary literals. Therefore, $C\sigma$ is a proper instance of C , or $D\sigma$ is a proper instance of D (or both). Both cases give a contradiction: in the first case, $C\gamma$ would not be generating as $C\sigma$ is a more specific representation of $C\gamma$ (recall that M is closed under Inst-Gen, and hence $C\sigma \in M$). Similarly, in the second case D would not be a most specific representation of $D\gamma$ by virtue of $D\sigma \in M$. Thus, the assumption that I falsifies a $D\gamma$ cannot hold, and the proof is complete.

The Inst-Gen calculus is extensible by many refinements, including hyper-style inference, semantic selection (add instances that are not yet satisfied by the propositional model $P \perp$), and certain forms of redundancy elimination, which can all be justified by the model-generation completeness proof above.

4 Hyper-Linking

The Hyper-Linking calculus [LP92] works, similarly to Inst-Gen, by looking for links between clauses. The process is divided into “rounds”; each round computes for every clause C the set of hyper-links for this clause. A hyperlink for $C = L_1 \vee \dots \vee L_n$ is a set of links between C and other clauses such that for every i , there is one and only one link featuring L_i .

For every such hyper-link, one can compose the unifiers of every link in it (possibly renaming variables) to get a single substitution for the hyper-link. Call this substitution θ ; the instance added to the current clause set is $C\theta$. After all the clauses have been processed (this process terminates), the calculus temporarily grounds every clause by substituting some distinct constant for every variable to obtain a set of ground clauses, as for Inst-Gen. This set can be checked for satisfiability by e.g. resolution, and if it is unsatisfiable the procedure stops. Otherwise, a new round is initiated with the set of instances that the last round failed to refute as input.

In this calculus, the instances computed are an attempt to create complementary clauses. This helps to avoid superfluous clauses, as they would have few or no links to other clauses. However, if there is a “cluster” of superfluous clauses, i.e. clauses that have links between them but no links to other clauses, and this cluster is satisfiable, Hyper-Linking would still generate instances of clauses in the cluster.

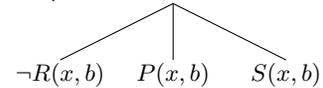
The Hyper-Linking calculus has been developed further, and has lead to IMs that take advantage of semantic guidance and of certain refinements based on ordering restrictions [CP94, PZ00].

5 Disconnection Tableaux

The Disconnection Tableaux calculus [LS01] works on links, which are also called connections in this context. In contrast to Hyper-Linking, where all the links of a clause are used together, Disconnection looks at one link at a time. The general idea is

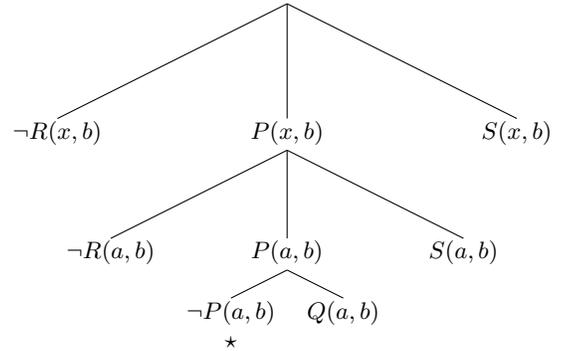
as follows: One starts by setting up an *initial path* through the input clause set by arbitrarily picking one literal from each input clause. This initial path is fixed throughout the entire subsequent tableau construction. Starting with the initial path, the calculus’ single inference rule takes a branch constructed so far and looks for a link between two literals on the branch. If such a link exists and, say, C and D are two clauses containing the link literals, the calculus expands the branch by the clause instances $C\sigma$ and $D\sigma$ using the unifier σ for this link. This way, the link is “disconnected”. Every relevant instance that can be found this way eventually ends up on the tableau, and if a branch contains literals that become complementary when grounded to a special constant, it is closed. A proof is a tableau where every branch is closed.

Consider an example expansion with the clause $\neg P(a, z) \vee Q(a, z)$, with $\neg P(a, z)$ being the literal on the initial path to the following tableau (the initial path is not drawn):



To use the link between $P(x, b)$ and $\neg P(a, z)$, the variable x has to be instantiated to a , and z to b . Before doing the expansion with $\neg P(a, z) \vee Q(a, z)$, the instantiated tableau clause has to be put on the branch. The branch marked by \star is closed, as it contains complementary literals $P(a, b)$ and $\neg P(a, b)$.

After the expansion, the tableau looks like this:



In general, the test for complementary literals is carried out after the tableau has been temporarily grounded by replacing every variable by a constant, as for Inst-Gen and Hyper-Linking. Any branch satisfying this condition without looking at the initial path can be closed, and there is always at least one such branch in every expansion.

This way of looking for single links that are “needed”, i.e. can be used together with the clauses that are already there, removes the problem of superfluous clauses to a large extent, as they have few or no links to other clauses. Hyper-Linking, for example, does not have a way to avoid this possibility. On the negative side, Disconnection generates many similar branches, something that can lead to near-copies of derivations.

6 FDPLL and Model Evolution

The First-Order DPLL (FDPLL) [Bau00] calculus and its successor, Model Evolution (ME) [BT03] share some features with the calculi described above. The main object of a derivation is a tree, as in Disconnection Tableaux, but branching is on complementary literals instead of clauses. As in Hyper-Linking, every

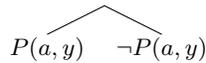
literal of the current clause must be simultaneously linked with literals on the current branch to drive the inference rules.

FDPLL and ME have been introduced as a lifting of the propositional core of the DPLL procedure to the first-order level. To describe how they work, it is instructive to recapitulate propositional DPLL first: Given a propositional clause set S , one picks a propositional variable, say A , from a clause in S , and creates two new clause sets $S[A/\top]$ and $S[A/\perp]$ (by $S[A/\perp]$ we mean the set S with every instance of A replaced by \perp), to be analyzed separately. The two clause sets created are simpler than S , and can be further simplified by propositional rules, e.g. $A \vee \top \equiv \top$. If we find an elementary contradiction during this simplification, that clause set is unsatisfiable. If not, a new propositional variable to split on is picked, until all propositional variables have been exhausted (with the conclusion that S is satisfiable), or all the sets generated are shown to be unsatisfiable, which means that S is not satisfiable either.

FDPLL lifts this splitting rule to first-order clauses by case analysis on possibly non-ground literals. FDPLL can be described as a semantic tree construction method akin to Disconnection tableaux, but branching on complementary literals instead of clauses. The intention of the calculus is then to construct a model for the input clause set. As an example, consider the two clauses

$$P(a, y) \quad P(x, b) \vee \neg P(z, y) \vee Q(x, y, z)$$

and suppose that the following tree has already been derived:



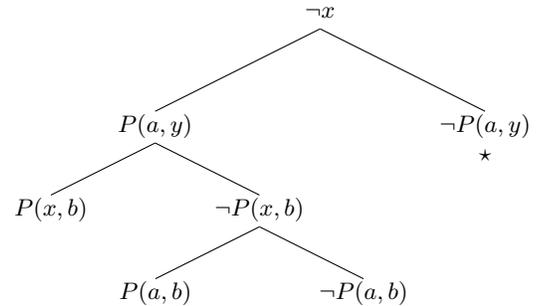
The left branch specifies the interpretation that assigns true to all ground instances of $P(a, y)$, and false to all other atoms. This interpretation assigns false to the second clause — in fact, it assigns false to every instance of $P(x, b) \vee \neg P(a, y) \vee Q(x, y, a)$ where x is different from a . This situation is detected by unifying the branch literal $P(a, y)$ with the complement of the clause literal $\neg P(x, y)$. Such links with branch literals have to be found for every clause literal. Positive clause literals can also be linked with the pseudo-literal “ $\neg x$ ” in the root node, where x unifies with any positive literal. In general, if a clause instance C computed this way contains a literal L such that neither L nor the complement of L is on the branch, the branch is extended by splitting on L and its complement. In the example, the literal $P(x, b)$ can be used for this, but $\neg P(a, y)$ cannot. On the left branch the interpretation will at least partially be “repaired” towards a model for C , and on the right branch the clause C is implicitly shrunk by the presence of the complement of L . Both cases mark some progress in either finding a model or a refutation. The test for closing branches is similar to the one in Disconnection Tableaux, and is based on simultaneously unifying all the literals of a clause with complementary branch literals after temporarily grounding them.

FDPLL is sound and complete. Regarding soundness recall that branch closure is based on temporarily replacing every variable in every branch literal by some fixed constant. A closed tree can therefore be seen as one that branches on complementary propositional literals. Soundness of FDPLL then follows easily.

The completeness proof is of the model-generating kind, akin to the one for Inst-Gen (cf. Section 3). In FDPLL, the central concept is that of a *candidate model* induced by a (current)

branch. Similarly to Inst-Gen, one can argue that any open branch in a fair derivation has had “enough” inferences applied to it, so no falsified clause can exist. We just note one key property here: Whenever the candidate model falsifies a clause C , there is a clause instance $C\sigma$, where σ simultaneously unifies the links between all literals in C and branch literals. The branch is then either closed or the branch can be split with some literal in $C\sigma$. Conversely, if no such unifier σ exists then C holds true in the candidate model. This fact can be seen as a semantically justified redundancy criterion – inferences with “true” clauses need not be carried out.

The following semantic tree can be derived from the example tree above in two steps:



The ME calculus subsumes FDPLL by also lifting several simplification rules from DPLL to the first-order level. ME operates on sequents of the form $\Lambda \vdash \Phi$. The set of literals Λ corresponds to a branch in an FDPLL semantic tree and is called a context, while Φ is the current clause set to be refuted. As said earlier, the chief advantage of ME over FDPLL is that it manages to lift the simplification rules of DPLL to first-order logic. An example of this is the Subsume rule, which in the propositional case allows one to simplify a clause set $\{L, L \vee C, \dots\}$ to $\{L, \dots\}$ — as L has to be true, the satisfiability of this clause set does not depend on $L \vee C$. A dual rule, Unit Resolution, allows to remove a literal from a clause in the presence of a unit clause with a complementary literal. Such rules are mandatory in practice, and ME contains first-order logic variants of both of them. They are used to simplify the current clause set Φ based on the current context Λ .

In addition to the simplification rules there are two rules that add literals to the context, Assert and Split. The Split rule is similar to the splitting rule in FDPLL, while the Assert rule is a lifting of the propositional rule that assigns \top to a clause containing a single literal, called a *unit clause*. Such a clause can only be made true in one way, namely by assigning true to the literal (in the propositional case), or assigning true to every instance of the literal (in the first-order case).

7 EPR

In the previous sections we have reviewed several IMs by summarizing their main underlying ideas and differences. IMs are conceptually different to more established methods based on resolution or unification-based free-variable tableaux. This way, they contribute as alternative methods to the repository of available first-order AR methods. In particular, IMs are strong on a certain fragment of first-order logic that proves difficult for many other methods. More precisely, all instance-based methods

can be used as decision procedures for the Bernays-Schönfinkel (BS) class of first-order logic. This class consists of formulas that, when written in prenex normal form, have the form $\exists \vec{x} \forall \vec{y} \phi(\vec{x}, \vec{y})$, where ϕ is a quantifier-free formula without function symbols. Such formulas are sometimes also referred to as effectively propositional (EPR), since they can be effectively translated into propositional logic by a finite process of grounding. However, the cost of that is an exponential blow-up in formula size. The translation into clause logic always yields a clause set that may contain variables and constants, but no terms built from proper function symbols (“Datalog”).

This property helps to explain why IMs decide the satisfiability problem of the EPR class. It is simply because the set of instances of a finite clause set without function symbols is finite modulo renaming of clause variables, something that is easy to control in IMs. By contrast, resolution (see the resolution inference rule in Section 2) might derive clauses of unbounded length, which makes it less suitable for EPR. This circumstance may partially explain why the winners in the EPR category in the annual CADE ATP Systems competition (CASC [SS06]), a major competition for automated theorem provers, have for the last six years been instance-based provers instead of resolution-based provers.

The decidability problem of EPR is complete for NEXPTIME. In practical terms this means that much more succinct problem specifications are possible with EPR than with propositional logic. This suggests to capitalize on IMs as decision procedures for EPR and to investigate practically feasible reductions of application problems into EPR. For instance, it is already known that bounded model checking problems can be encoded in BS logic much more succinctly than in propositional logic [NV07].

Another example is the optimized functional translation of modal logics [OS97] to BS logic [Sch99]. Many benchmark problems obtained this way are contained in the TPTP problem library [SS98], and implementations of instance based methods consistently score very well on them. In the description logic context, [MSS04] show how to translate the expressive description logic $SHIQ(D)$ to BS logic, but with a different motivation.

Other potentially useful applications of IMs as decision procedures for EPR lie within the *constraint programming* area. IMs are possibly not the preferred choice as solvers for search problems, typically in NP, as this is the domain of the traditional constraint programming paradigm. More appropriate seems the application to e.g. model expansion problems [MTHM06] (with NEXPTIME combined query/data complexity), which can be reduced to EPR in a way similar to finite model computation mentioned above. Another application is to analyze constraint models for certain “interesting” properties, like proving of functional dependencies and symmetries between decision variables [CM05, CM04]. Quite often, the resulting proof obligations lie within BS logic.

A “generic” application area is *finite model computation*. Finite model computation is the problem of computing a model with a finite domain for the given formula or clause set, if one exists. One application of this is in computing counterexamples of “false” theorems, which arise frequently in software verification or modelling in early stages. See [BT98, BS06, dNM06] for IM-related methods. Other methods for finite model computation essentially work by stepwise reduction to formulas in proposi-

tional logic and use a SAT solver on the result. [Sl94, McC94, ZZ95, CS03, Pel03b, e.g.]. In [BFdNT09] it was shown how this model computation paradigm can be rooted in the Model Evolution calculus instead of a SAT solver, something that can lead to space advantages. Actually, the results in [BFdNT09] are more general, and any method that decides the EPR class can be used.

8 Conclusions and Outlook

In this paper we surveyed methods for instantiation-based theorem proving and indicated their strength for the EPR class of first-order logic formulas. We concentrated on the basic versions of four typical IMs, Inst-Gen, Hyper-Linking, Disconnection Tableaux and FDPLL/Model Evolution. The theoretical power of each of them is the same, they all are sound and complete methods for first-order clausal theorem proving. So what are the differences? First, there are conceptual differences in the way they lift propositional reasoning to the first-order level. IMs can broadly be classified as *one-level vs. two-level methods*: two-level methods like Inst-Gen and Hyper-Linking directly employ a propositional SAT-solver as a subroutine for periodic unsatisfiability tests of the grounded version of the current clause set. This way, these methods can always capitalize on the latest advances in SAT-solving technology. In the extreme case, when the given clause set is already ground, their performance is the same as if their SAT-solver had been directly called. In contrast, one-level methods like Disconnection Tableaux and FDPLL/Model Evolution work directly with “lifted” first-order logic data structures and inference rules of their propositional base calculi — in these cases propositional tableaux and propositional DPLL. These first-order data structures allow some optimizations that are difficult to replicate on the propositional level; see the discussion of “candidate model” and redundancy criterion in Section 6. On the other hand, two-level methods can take advantage of SAT-solving technologies only by adapting them individually, both with respect to theory and implementation. A good example for this is *lemma learning*, a key factor in modern SAT-solving which helps to avoid repeating identical parts of a refutation. It required some efforts to integrate lemma learning into Model Evolution [BFT06] but, on the upside, led to a more powerful lemma learning mechanism.

Other differences between the IMs considered here would require a deeper technical treatment, something that is beyond the scope of this paper. See [JW07] for a comparison of IMs with respect to simulations of refutations.

In most applications, e.g., software verification, an efficient treatment of equality by specialized inference rules is mandatory. Fortunately, research on efficient equality reasoning in IMs can capitalize on the results developed for the resolution calculus over the last 20 years. Indeed, some promising approaches along these lines have been developed for Inst-Gen [GK04], for Disconnection tableaux [LS02], and for Model Evolution [BT05]. These approaches all employ ordering restrictions as pioneered for the resolution calculus (see [NR01] for an overview). How this is concretely realized and the discussion of the differences is beyond the scope of this paper. A related but less developed topic is the integration of reasoning modulo more general background theories, such as integer arithmetics or theories of certain

data structures (lists, arrays, sets, etc). This is currently a hot topic, and only some initial results are available [GK06, BFT08]. One motivation for this stream of research to address a major weakness of the prevailing approach to theory reasoning, *Satisfiability Modulo Theories (SMT)* [RT06]. To explain, current SMT systems are practically very successful for quantifier-free (i.e. ground) input formulas. However, they do not natively support quantifiers and resort to incomplete instantiation heuristics for quantified formulas. In contrast, IMs are devised as first-order logic calculi at the outset and provide a systematic treatment of quantifiers. Equipping IMs with theory reasoning could thus provide alternatives to SMT. Under certain restrictions it is even possible to design refutational complete calculi over, say, integer arithmetics.

A different line of research has only just begun, the combination of instance based methods and resolution calculi. The motivation for that is to combine their individual strengths in a single framework. See [BW09] for an integration of Model Evolution and Superposition, and [LM09] for an integration of Inst-Gen and Resolution.

Many of the calculi we discussed have been implemented, yielding insight into their practical applicability. For Hyper-Linking, there is a prover CLIN [Lee90], with improved versions CLIN-S (semantic restriction) [CP97] and CLIN-E (equality handling) [Ale97]. For Disconnection Tableaux, there is a prover DCTP [Ste02] featuring both equality handling and various refinements. The same is true for Inst-Gen, with the iProver [K08], and the Model Evolution calculus, with prover Darwin [BFT04]. DCTP, iProver and Darwin regularly participate in the CASC competition.

Other basic research questions concern, for example, search space improvements, implementation techniques, variants for deciding more fragments of first-order logic than are currently known, better understanding of theoretical properties, and clarifying the relationships between IMs and other methods.

Acknowledgements. We thank the reviewers for their helpful comments and suggestions.

References

- [Ale97] G.D. Alexander. CLIN-E — Smallest Instance First Hyper-Linking. *Journal of Automated Reasoning*, 18(2):177–182, 1997.
- [Bau98] P. Baumgartner. Hyper Tableaux — The Next Generation. In *Proc. TABLEAUX'98*, LNCS 1397, pp. 60–76, 1998.
- [Bau00] P. Baumgartner. FDPLL – A First-Order Davis-Putnam-Logeman-Loveland Procedure. In *Proc. CADE-17*, LNAI 1831, pp. 200–219, 2000.
- [BEF99] P. Baumgartner, N. Eisinger, U. Furbach. A confluent connection calculus. In *Proc. CADE-16*, LNAI 1632, pp. 329–343, 1999.
- [BFdNT09] P. Baumgartner, A. Fuchs, H. de Nivelle, C. Tinelli. Computing finite models by reduction to function-free clause logic. *Journal of Applied Logic*, 7(1):58–74, 2009.
- [BFN96] P. Baumgartner, U. Furbach, I. Niemelä. Hyper Tableaux. In *Proc. JELIA'96*, LNAI 1126, 1996.
- [BFT04] P. Baumgartner, A. Fuchs, C. Tinelli. Implementing the Model Evolution Calculus. *International Journal of Artificial Intelligence Tools*, 15(1):21–52, 2006.
- [BFT06] P. Baumgartner, A. Fuchs, C. Tinelli. Lemma learning in the model evolution calculus. In *Proc. LPAR*, LNAI 4246, pp. 572–586, 2006.
- [BFT08] P. Baumgartner, A. Fuchs, C. Tinelli. ME(LIA) – Model Evolution With Linear Integer Arithmetic Constraints. In *Proc. LPAR'08*, LNAI 5330, pp. 258–273, 2008.
- [BW09] P. Baumgartner, U. Waldmann. Superposition and Model Evolution Combined. In *Proc. CADE-22*, LNAI 5663, pp. 17–34, 2009.
- [BG90] L. Bachmair and H. Ganzinger. On Restrictions of Ordered Paramodulation with Simplification. In *Proc. CADE-10*, LNAI 449, pp. 427–441, 1990.
- [BG01] L. Bachmair and H. Ganzinger. Resolution Theorem Proving. In A. Robinson and A. Voronkov, eds., *Handbook of Automated Reasoning*. North Holland, 2001.
- [Bil96] J. Billon. The Disconnection Method. In *Proc. TABLEAUX'96*, LNAI 1071, pp. 110–126, 1996.
- [BS06] P. Baumgartner, R. Schmidt. Blocking and other enhancements for bottom-up model generation methods. In *Proc. IJCAR*, LNAI 4130, pp. 125–139, 2006.
- [BT98] F. Bry and S. Torge. A Deduction Method Complete for Refutation and Finite Satisfiability. In *Proc. JELIA*, LNAI, pp. 122–136, 1998.
- [BT03] P. Baumgartner, C. Tinelli. The Model Evolution Calculus. In *Proc. CADE-19*, LNAI 2741, pp. 350–364, 2003.
- [BT05] P. Baumgartner, C. Tinelli. The model evolution calculus with equality. In *Proc. CADE-20*, LNAI 3632, pp. 392–408, 2005.
- [Bun94] A. Bundy, ed. *Proc. CADE-12*, LNAI 814, 1994.
- [CM04] M. Cadoli, T. Mancini. Exploiting functional dependencies in declarative problem specifications. In *Proc. JELIA'04*, LNCS 3229, pp. 628–640, 2004.
- [CM05] M. Cadoli, T. Mancini. Using a theorem prover for reasoning on constraint problems. In *AI*IA*, pp. 38–49, 2005.
- [CP94] H. Chu, D. Plaisted. Semantically Guided First-Order Theorem Proving using Hyper-Linking. In Bundy [Bun94], pp. 192–206.
- [CP97] H. Chu, D.A. Plaisted. CLIN-S — A Semantically Guided First-Order Theorem Prover. *Journal of Automated Reasoning*, 18(2):183–188, 1997.
- [CS03] K. Claessen, N. Sörensson. New techniques that improve mace-style finite model building. In *CADE-19 Workshop: Model Computation – Principles, Algorithms, Applications*, 2003.
- [DLL62] M. Davis, G. Logemann, D. Loveland. A machine program for theorem proving. *Communications of the ACM*, 5(7), 1962.
- [dMB08] L. de Moura, N. Bjørner. Deciding effectively propositional logic using DPLL and substitution sets. In *Proc. IJCAR 2008*, LNAI 5195, pp. 410–425, 2008.
- [dNM06] H. de Nivelle, J. Meng. Geometric resolution: A proof procedure based on finite model search. In *Proc. IJCAR 2006*, LNAI 4130, 2006.
- [DP60] M. Davis, H. Putnam. A computing procedure for quantification theory. *Journal of the ACM*, 7(3):201–215, 1960.
- [FP05] C. Fermüller, R. Pichler. Model representation via contexts and implicit generalizations. In *Proc. LPAR'01*, LNCS 2250, pp. 409–423, 2001.
- [GK03] H. Ganzinger, K. Korovin. New directions in instance-based theorem proving. In *Proc. LICS*, 2003.

- [GK04] H. Ganzinger and K. Korovin. Integrating equational reasoning into instantiation-based theorem proving. In *Proc. CSL'04*, LNCS 3210, pp. 71–84, 2004.
- [GK06] H. Ganzinger and K. Korovin. Theory Instantiation. In *Proc. LPAR'06*, LNCS 4246, pp. 497–511, 2006.
- [HRCS02] J.N. Hooker, G. Rago, V. Chandru, A. Shrivastava. Partial Instantiation Methods for Inference in First Order Logic. *Journal of Automated Reasoning*, 28(4):371–396, 2002.
- [JW07] S. Jacobs, U. Waldmann. Comparing instance generation methods for automated reasoning. *Journal of Automated Reasoning*, 38(1-3):57–78, 2007.
- [K08] K. Korovin. iProver - An Instantiation-Based Theorem Prover for First-Order Logic (System Description). In *Proc. IJCAR 2008*, LNAI 5195, pp. 292–298, 2008.
- [Kor09] K. Korovin. Instantiation-Based Automated Reasoning: From Theory to Practice. In *Proc. CADE-22*, LNCS 5663, pp 163–166, 2009.
- [Lee90] S. Lee. *CLIN: An automated reasoning system using clause linking*. PhD thesis, The University of North Carolina at Chapel Hill, 1990.
- [LP92] S.-J. Lee and D. Plaisted. Eliminating Duplicates with the Hyper-Linking Strategy. *Journal of Automated Reasoning*, 9:25–42, 1992.
- [LS01] R. Letz, G. Stenz. Proof and Model Generation with Disconnection Tableaux. In *Proc. LPAR'01*, LNCS 2250, 2001.
- [LS02] R. Letz, G. Stenz. Integration of Equality Reasoning into the Disconnection Calculus. In *Proc. TABLEAUX 2002*, LNCS 2381, pp. 176–190, 2002.
- [LM09] C. Lynch, R. McGregor. Combining Instance Generation and Resolution. In *Proc. FroCoS 2009*, LNCS 5749, 2009.
- [MB88] R. Manthey, F. Bry. SATCHMO: a theorem prover implemented in Prolog. In *Proc. CADE-9*, LNCS 310, pp. 415–434, 1988.
- [McC94] W. McCune. A davis-putnam program and its application to finite first-order model search: Quasigroup existence problems. Technical report, Argonne National Laboratory, 1994.
- [MSS04] B. Motik, U. Sattler, R. Studer. Query answering for owl-dl with rules. In *International Semantic Web Conference*, pp. 549–563. AAAI Press, 2004.
- [MTHM06] D. Mitchell, E. Ternovska, F. Hach, R. Mohebbi. Model Expansion as a Framework For Modelling and Solving Search Problems. Technical Report TR 2006-24, Simon Fraser University, December 2006.
- [NR01] R. Nieuwenhuis, A. Rubio. Paramodulation-Based Theorem Proving. In *Handbook of Automated Reasoning*, pp 371–443. Elsevier and MIT Press, 2001.
- [NV07] J. A. Navarro Pérez, A. Voronkov. Encodings of bounded LTL model checking in effectively propositional logic. In *Proc. CADE-21*, LNCS 4603, pp. 346–361, 2007.
- [RT06] S. Ranise, C. Tinelli. Satisfiability Modulo Theories. *Trends and Controversies - IEEE Intelligent Systems Magazine*, 21(6):71–81, 2006.
- [NW01] A. Nonnengart, C. Weidenbach. Computing small clause normal forms. In J. A. Robinson, A. Voronkov, eds., *Handbook of Automated Reasoning*, pp. 335–367. Elsevier and MIT Press, 2001.
- [OS97] H. J. Ohlbach, R. A. Schmidt. Functional translation and second-order frame properties of modal logics. *Journal of Logic and Computation*, 7(5):581–603, 1997.
- [Pel03b] N. Peltier. A more efficient tableaux procedure for simultaneous search for refutations and finite models. In *Proc. TABLEAUX 2003*, LNCS 2796, pp. 181–195, 2003.
- [Pla94] D. Plaisted. The Search Efficiency of Theorem Proving Strategies. In Bundy [Bun94].
- [PZ00] D. A. Plaisted, Y. Zhu. Ordered Semantic Hyper Linking. *Journal of Automated Reasoning*, 25(3):167–217, 2000.
- [Rob65] J.A. Robinson. A machine-oriented logic based on the resolution principle. *JACM*, 12(1):23–41, 1965.
- [Sch99] R. A. Schmidt. Decidability by resolution for propositional modal logics. *Journal of Automated Reasoning*, 22(4):379–396, 1999.
- [SL04] G. Stenz, R. Letz. Generalized handling of variables in disconnection tableaux. In *Proc. IJCAR 2004*, LNCS 3097, pp. 289–306, 2004.
- [Sla94] J. Slaney. FINDER: Finite Domain Enumerator (System Description) In Bundy [Bun94].
- [Ste02] G. Stenz. DCTP 1.2 — System Abstract. In *Proc. TABLEAUX'02*, LNAI 2381, pages 335–340, 2002.
- [SS98] G. Sutcliffe, C. Suttner. The TPTP Problem Library: CNF Release v1.2.1. *Journal of Automated Reasoning*, 21(2):177–203, 1998.
- [SS06] G. Sutcliffe, C. Suttner. The State of CASC. *AI Communications*, 19(1):35–48, 2006.
- [ZZ95] J. Zhang, H. Zhang. Sem: a system for enumerating models. In *Proc. IJCAI-95*, pp. 298–303. Morgan Kaufmann, 1995.

Contact

Dr. habil. Peter Baumgartner
 NICTA Canberra and Australian National University
 7 London Circuit
 Canberra ACT 2601
 Australia
 Email: Peter.Baumgartner@nicta.com.au

Bild

Peter Baumgartner is a Principal Researcher and Research Group Manager with NICTA, Australia's center of excellence in information and communication technology. (NICTA is funded by the Australian Government's *Backing Australia's Ability* initiative.) Before joining NICTA, he mainly worked at the University of Koblenz from 1990 until 2003, and at the Max-Planck-Institut for Computer Science in Saarbrücken from 2003 to 2005.

Bild

Evgenij Thorstensen is a graduate student at the University of Oxford. He holds a master's degree in computer science from the University of Oslo.