Girdle: Four Steps to Automatic Theorem Proving

David Anthony Bau

October 7, 2017

1 The Model

The first step to automatic theorem proving is to set up a representation for sentences ("sentences" in the model-theoretic sense). We require that every language in which Girdle works contain at least the following symbols:

1. $\forall x.\psi[x]$, the universal quantifier

2. $\exists x.\psi[x]$, the existential quantifier

3. $x \wedge y$, logical and

4. $x \vee y$, logical or

5. $x \implies y$, logical implies

6. $x \iff y$, logical equivalence

7. $= [x, y]$, the equality relation, more frequently written as $x = y$

These operations all have the associated normal axioms (for instance, $=$ is reflexive, transitive, and commutative, and De Morgan's laws apply).

Working with arbitrary trees of sentences is dramatically inconvenient, so we move everything to conjunctive normal form (CNF). This is a "flat" representation of sentences where every sentence is a conjunction of disjunctions of (optionally) negations of terms. This can be reached simply by applying De Morgan's laws and the distributive property several times.

$$(a \wedge b) \vee c = (a \vee c) \wedge (b \vee c)$$
$$\neg(a \wedge b) = (\neg a) \vee (\neg b)$$
$$\neg(a \vee b) = (\neg a) \wedge (\neg b)$$
$$\neg \exists x.\psi[x] = \forall x.\ \neg\psi[x]$$
$$\neg \forall x.\psi[x] = \exists x.\ \neg\psi[x]$$

We then move further into Skolem conjunctive normal form by Skolemizing out all of the existential quantifiers. Skolemization is the motion:

$$\forall x.\exists y.\psi[x, y] = \exists Y.\forall x.\psi[x, Y(x)]$$

In other words, saying "for every x, I can find a y" is equivalent to saying "there is some finder-function Y that can find such a value for every x".

Now that everything is in Skolem normal form, variables exist in exactly one of two context: either they were declared as $\exists$ or declared as $\forall$ (order no longer matters). We then make every $\exists$ variable a constant, which is a refutationally complete operation (equivalent to saying "one exists, so choose one"). If we show inconsistency, i.e. if we show that **no** model can fill in the constants correctly, then we will have shown that no model can satisfy the existence conditions either, since if one did we could just pick the correct interpreation and it would fill in the constants correctly too. Conversely, if we can't show inconsistency, i.e. if there exists some model that fills in these constants correctly, then some model satisfies the existence conditions too trivially. Thus this operation is refutationally complete. Now we can treat all variables the same, and all constants the same.