Validity of FOL is undecidable

Jochen Hoenicke

May 15, 2012

Theorem 1 (FOL is undecidable (Turing & Church))**.** *There is no algorithm for deciding if a FOL formula F is valid, i.e. an algorithm that always halts and says "yes" if F is valid or says "no" if F is invalid.*

Proof. We reduce the halting problem for deterministic Turing machines on the empty tape to the validity problem for first order-logic. For a TM τ we build a first-order-logic formula F_τ such that τ terminates when started on the empty tape if and only if F_τ is valid.

Let $\tau = (Q, \Sigma, \Gamma, \delta, q_0, q_n)$ be a deterministic Turing Machine with states $Q = \{q_0, \ldots, q_n\}$, input alphabet $\Sigma = \{\}$ (we consider the halting problem on an empty tape), tape alphabet $\Gamma = \{a_0, \ldots, a_m\}$ where a_0 is the blank symbol, start state q_0, final state q_n, and a total transition function $\delta : Q \times \Gamma \to Q \times \Gamma \times \{L, R\}$. We build a formula that encodes the run of τ. There is one constant *zero* and two one-argument functions *succ*, *pred*. Furthermore we have $n + m + 2$ predicates of arity 2, $q_0, \ldots, q_n, a_0, \ldots, a_m$. The intended meaning of the predicate $q_i(s, p)$ is that in the sth step, the Turing Machine is at position p in state q_i. The intended meaning of the predicate $a_i(s, p)$ is that at the sth step the symbol at position p is a_i.

The formula F_τ consists of several components:

- Functions *succ* and *pred* are inverse to each other:

$$F_1 = \forall s\ (\mathit{pred}(\mathit{succ}(s)) = s \land \mathit{succ}(\mathit{pred}(s)) = s)$$

- Always at every position there is at most one symbol on the tape:

$$F_2 = \forall s\ \forall p \bigwedge_{\substack{i,j \in \{0,\ldots,m\} \\ i \neq j}} (\neg a_i(s,p) \lor \neg a_j(s,p))$$

  Note that this can be written as a valid first-order formula once the number of symbols m is known. In particular there is an algorithm that computes formula F_2 from a given Turing Machine τ.

- Always the TM is only in one state

$$F_3 = \forall s\ \forall p_1\ \forall p_2 \bigwedge_{\substack{i,j \in \{0,\ldots,n\} \\ i \neq j}} (\neg q_i(s,p_1) \lor \neg q_j(s,p_2))$$

- Always the TM is only at one position

$$F_4 = \forall s\ \forall p_1\ \forall p_2 \bigwedge_{i \in \{0,\ldots,n\}} (p_1 \neq p_2 \to \neg q_i(s,p_1) \vee \neg q_i(s,p_2))$$

- Only the symbol at the position of the TM may change.

$$F_5 = \forall s\ \forall p \bigwedge_{i \in \{0,\ldots,m\}} (a_i(s,p) \wedge \neg a_i(succ(s),p) \to \bigvee_{j \in \{0,\ldots,n\}} q_j(s,p))$$

- The TM writes the correct symbol: For each $q \in Q, a \in \Gamma$ with $\delta(q,a) = (q', a', R)$, we define

  $$F_{q,a} = \forall s\ \forall p\ (a(s,p) \wedge q(s,p) \to a'(succ(s),p) \wedge q'(succ(s), succ(p)))$$

  For each $q \in Q, a \in \Gamma$ with $\delta(q,a) = (q', a', L)$, we define

  $$F_{q,a} = \forall s\ \forall p\ (a(s,p) \wedge q(s,p) \to a'(succ(s),p) \wedge q'(succ(s), pred(p)))$$

  then F_6 is the conjunction of these formulas.

- The TM starts at step zero on the empty tape:

$$F_7 = q_0(zero, zero) \wedge \forall p\ a_0(zero, p)$$

The formula F_τ specifies that every run of τ is terminating:

$$F_\tau = F_1 \wedge \ldots \wedge F_7 \to \exists s\ \exists p\ q_n(s,p)$$

We show that F_τ is valid if and only if τ terminates when starting on the empty tape.

only if We show that there is a falsifying model I for F_τ if τ does not terminate on the empty tape. Let $D_I = \mathbb{Z}$, $\alpha_I(zero) = 0$, $\alpha_I(succ)(x) = x + 1$, $\alpha_I(pred)(x) = x - 1$.

We set $\alpha_I[q_i](s,p) = \top$ if and only if $s \geq 0$ and the TM τ is in step s at position p in state q_i. Note that for $s < 0$ the predicate $q_i(s,p)$ is always false. This is consistent with $F_1, \ldots, F_7$.

We set $\alpha_I[a_i](s,p)$ if and only if $s < 0$ and $i = 0$ or $s \geq 0$ and the tape contains symbol a_i at position p in step s.

One can see that $F_1, \ldots, F_7$ are true and $\exists s\ \exists p\ q_n(s,p)$ is false. Hence I is a falsifying interpretation for F_τ.

if Let $succ^i(zero)$ denote the term $succ(\ldots(succ(zero)\ldots)$ with i applications of $succ$. If $i < 0$ we denote by $succ^i(zero)$ the term $pred(\ldots(pred(zero)\ldots)$ with $-i$ applications of $pred$.

One can show by induction over i that for every interpretation satisfying $F_1, \ldots, F_7$ that if at step i the TM is in state q_j and at position p the predicate $q_j(succ^i(zero), succ^p(zero))$ holds and that if at step i the tape contains symbol a_j at position p the predicate $a_j(succ^i(zero), succ^p(zero))$ holds. Since τ terminates, there is a step i and a position p at which the τ reaches the final state, hence $q_n(succ^i(zero), succ^p(zero))$ holds. Hence F_τ is true for every interpretation. $\square$