CS344: Introduction to Artificial Intelligence

Lecture: 22-23

Herbrand’s Theorem

Proving satisfiability of logic formulae using semantic trees

(from *Symbolic logic and mechanical theorem proving*)

By

Raunak Pilani

Under the guidance of Prof. P. Bhattacharyya

Basic Definitions

- <u>Interpretation</u>: Assignment of meaning to the symbols of a language
- Interpretations of Predicate logic requires defining:
  - Domain Of Discourse (D), which is a set of individuals that the quantifiers will range over
  - Mappings for every constant, n-ary function and n-ary predicate to elements, n-ary functions ($D^n \rightarrow D$) and n-ary relations on D, respectively

Basic Definitions (contd.)

- Satisfiability (Consistency)
  - A formula G is satisfiable iff there exists an interpretation I such that G is evaluated to "T" (True) in I
  - I is then called a model of G and is said to satisfy G
- Unsatisfiability (Inconsistency)
  - G is inconsistent iff there exists no interpretation that satisfies G

Need for the theorem

- Proving satisfiability of a formula is better achieved by proving the unsatisfiability of its negation
- Proving unsatisfiability over a large set of interpretations is resource intensive
- Herbrands Theorem reduces the number of interpretations that need to be checked
- Plays a fundamental role in Automated Theorem Proving

Skolem Standard Form

- Logic formulae need to first be converted to the Skolem Standard Form, which leaves the formula in the form of a set of clauses
- This is done in three steps
  - Convert to Prenex Form
  - Convert to CNF (Conjunctive Normal Form)
  - Eliminate existential Quanitifiers using Skolem functions

Step 1: Converting to Prenex Form

- Involves bringing all quantifiers to the beginning of the formula
  - $(Q_i\ x_i)\ (M)$, i=1, 2..., n

    Where,

    - Q_i is either ∀ (Universal Quantifier) or ∃ (Existential Quanitifier) and is called the *prefix*

    - M contains no Quantifiers and is called the *matrix*

Example

$$\neg((\forall x)\ P(x)\ \rightarrow\ (\exists y)(\forall z)\ Q(y,\ z))$$

$$\Rightarrow\ \neg(\ \neg((\forall x)\ P(x))\ \vee\ ((\exists y)(\forall z)\ Q(y,\ z)\ )\ )$$

$$\Rightarrow\ ((\forall x)\ P(x))\ \wedge\ \neg((\exists y)(\forall z)\ Q(y,\ z)\ )\ )$$

$$\Rightarrow\ ((\forall x)\ P(x))\ \wedge\ ((\forall y)\neg((\forall z)\ Q(y,\ z)\ )))$$

$$\Rightarrow\ (\forall x)\ P(x)\ \wedge\ (\forall y)(\exists z)\ \neg Q(y,\ z)$$

$$\Rightarrow\ (\forall x)(\ \forall y)(\exists z)\ P(x)\ \wedge\ \neg Q(y,\ z)$$

Step 2: Converting to CNF

- Remove $\longleftrightarrow$ and $\longrightarrow$
- Apply De Morgan's laws
- Apply Distributive laws
- Apply Commutative as well as Associative laws

Example

$$(\forall x)(\exists y)(\exists z)((P(x,y) \vee \neg Q(x,z)) \rightarrow R(x,y,z))$$

$$\Rightarrow (\forall x)(\exists y)(\exists z)(\neg(P(x,y) \vee \neg Q(x,z)) \vee R(x,y,z))$$

$$\Rightarrow (\forall x)(\exists y)(\exists z)((\neg P(x,y) \wedge Q(x,z)) \vee R(x,y,z))$$

$$\Rightarrow (\forall x)(\exists y)(\exists z)((\neg P(x,y) \vee R(x,y,z)) \wedge (Q(x,z) \vee R(x,y,z)))$$

Step 3: Skolemization

- Consider the formula, $(Q_1\ x_1)\ldots(Q_n\ x_n)M$
- If an existential quantifier, Q_r is not preceded by any universal quantifier, then
  - x_r in M can be replaced by any constant c and Q_r can be removed
- Otherwise, if there are 'm' universal quantifiers before Q_r, then
  - An m-place function $f(p_1, p_2, \ldots, p_m)$ can replace x_r where $p_1, p_2, \ldots, p_m$ are the variables that have been universally quantified
- Here, c is a skolem variable while f is a skolem function

Example

$$(\forall x)(\exists y)(\exists z)((\neg P(x, y) \vee R(x, y, z)) \wedge (Q(x, z) \vee R(x, y, z)))$$

We eliminate y and z using skolem functions

$\therefore$ y becomes f(x) and z becomes g(x)

as x is the only preceding universal quantifier

$$\Rightarrow (\forall x)((\neg P(x, f(x)) \vee R(x, f(x), g(x))) \wedge$$

$$(Q(x, g(x)) \vee R(x, f(x), g(x))))$$

Herbrand Universe

- It is infeasible to consider all interpretations over all domains in order to prove unsatisfiability

- Instead, we try to fix a special domain (called a Herbrand universe) such that the formula, S, is unsatisfiable iff it is false under all the interpretations over this domain

Herbrand Universe (contd.)

- H_0 is the set of all constants in a set of clauses, S
- If there are no constants in S, then H_0 will have a single constant, say $H_0 = \{a\}$
- For i=1,2,3,…, let H_{i+1} be the union of H_i and set of all terms of the form $f^n(t_1,\ldots, t_n)$ for all n-place functions f in S, where t_j where j=1,…,n are members of the set H
- H_∞ is called the Herbrand universe of S

Herbrand Universe (contd.)

- <u>Atom Set</u>: Set of the ground atoms of the form $P^n(t_1,\ldots, t_n)$ for all n-place predicates P^n occuring in S, where $t_1,\ldots, t_n$ are elements of the Herbrand Universe of S
  - Also called the *Herbrand Base*
- A <u>ground instance</u> of a clause C of a set of clauses is a clause obtained by replacing variables in C by members of the Herbrand Universe of S

Example

$$S = \{P(a), \neg P(x) \vee P(f(x)), Q(x)\}$$

$$H_0 = \{a\}$$

$$H_1 = \{a, f(a)\}$$

$$H_2 = \{a, f(a), f(f(a))\}$$

$$\vdots$$

$$H_\infty = \{a, f(a), f(f(a)), f(f(f(a))), \ldots\}$$

Let, $C = Q(x)$

Here, $Q(a)$ and $Q(f(f(a)))$ are both ground instances of C

Atom Set : $A = \{P(a), Q(a), P(f(a)), Q(f(a)), \ldots\}$

H-Interpretations

- For a set of clauses S with its Herbrand universe H, we define I as an H-Interpretation if:
  - I maps all constants in S to themselves
  - An n-place function f is assigned a function that maps $(h_1, \ldots, h_n)$ (an element in H^n) to f $(h_1, \ldots, h_n)$ (an element in H) where $h_1, \ldots, h_n$ are elements in H
- Or simply stated as $I = \{m_1, m_2, \ldots, m_n, \ldots\}$

where $m_i = A_i$ or $\sim A_i$ (*i.e.* A_i *is* set to true or false) and $A = \{A_1, A_2, \ldots, A_n, \ldots\}$

H-Interpretations (contd.)

- Not all interpretations are H-Interpretations
- Given an interpretation I over a domain D, an H-Interpretation I* corresponding to I is an H-Interpretation that:
  - Has each element from the Herbrand Universe mapped to some element of D
  - Truth value of $P(h_1, \ldots, h_n)$ in I* must be same as that of $P(d_1, \ldots, d_n)$ in I

Example

$$Let,\, S = \{P(x) \vee Q(x), R(f(y))\}$$

$\Rightarrow$ Herbrand Universe $H = H_\infty = \{a, f(a), f(f(a)), \ldots\}$

$\Rightarrow$ Atom Set A is given by

$$A = \{P(a), P(f(a)), P(f(f(a))), \ldots, Q(a), \ldots, R(a), \ldots\}$$

$\Rightarrow$ Some Herbrand Interpretations are

$$I_1 = \{P(a), P(f(a)), P(f(f(a))), \ldots, Q(a), \ldots, R(a), \ldots\}$$

$$I_2 = \{\neg P(a), \neg P(f(a)), \neg P(f(f(a))), \ldots, \neg Q(a), \ldots, \neg R(a), \ldots\}$$

$$I_3 = \{\neg P(a), \neg P(f(a)), \neg P(f(f(a))), \ldots, Q(a), \ldots, R(a), \ldots\}$$

Use of H-Interpretations

- If an interpretation I satisfies a set of clauses S, over some domain D, then any one of the H-Interpretations I* corresponding to I will also satisfy H
- <u>A set of clauses S is unsatisfiable iff S is false under all H-Interpretations of S</u>

Semantic Trees

- Finding a proof for a set of clauses is equivalent to generating a semantic tree
- A semantic tree is a tree where each link is attached with a finite set of atoms or their negations, such that:
  - Each node has only a finite set of immediate links
  - For each node N, the union of sets connected to links of the branch down to N does not contain a complementary pair
- If N is an inner node, then its outgoing links are marked with complementary literals

Semantic Trees (Contd.)

- Every path to a node N does not contain complementary literals in $I(N)$, where $I(N)$ is the set of literals along the edges of the path
- A <u>Complete Semantic Tree</u> is one in which every path contains every literal in Herbrand base either +ve or –ve, but not both
- A <u>failure node</u> N is one which falsifies I_N but not $I_{N'}$, where N' is predecessor of N
- A semantic tree is <u>closed</u> if every path contains a failure node

Example

$$S' = \{C_2 : q \vee r,\ C_3 : \neg p \vee \neg q,\ C_4 : p \vee \neg r,\ C_5 : p \vee \neg q \vee r\}$$

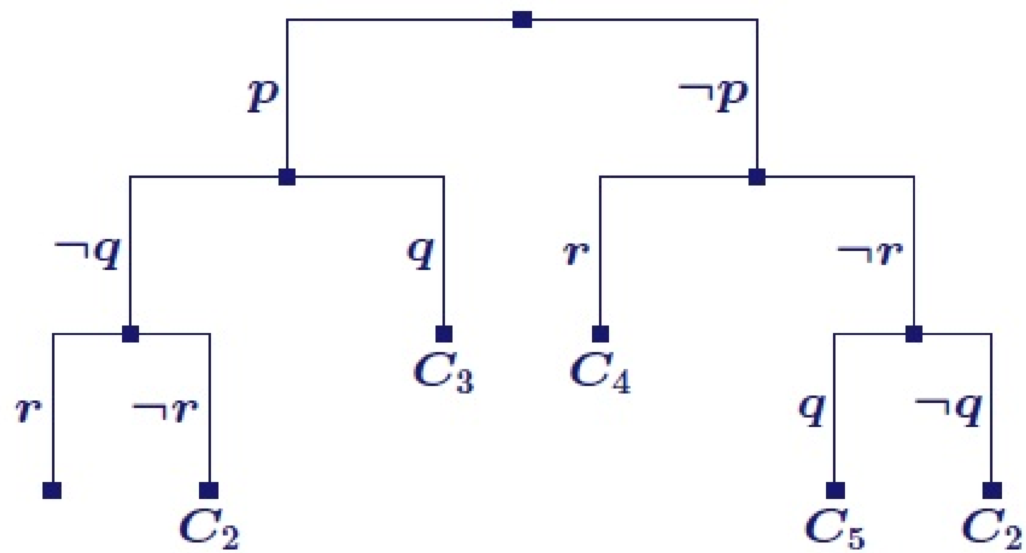

Image courtesy: http://www.computational-logic.org/iccl/master/lectures/summer07/sat/slides/semantictrees.pdf

S' is satisfiable because it has at least one branch without a failure node

Example

$$S = \{C_1 : q \vee \neg r,\ C_2 : q \vee r,\ C_3 : \neg p \vee \neg q,\ C_4 : p \vee \neg r,\ C_5 : p \vee \neg q \vee r\}$$

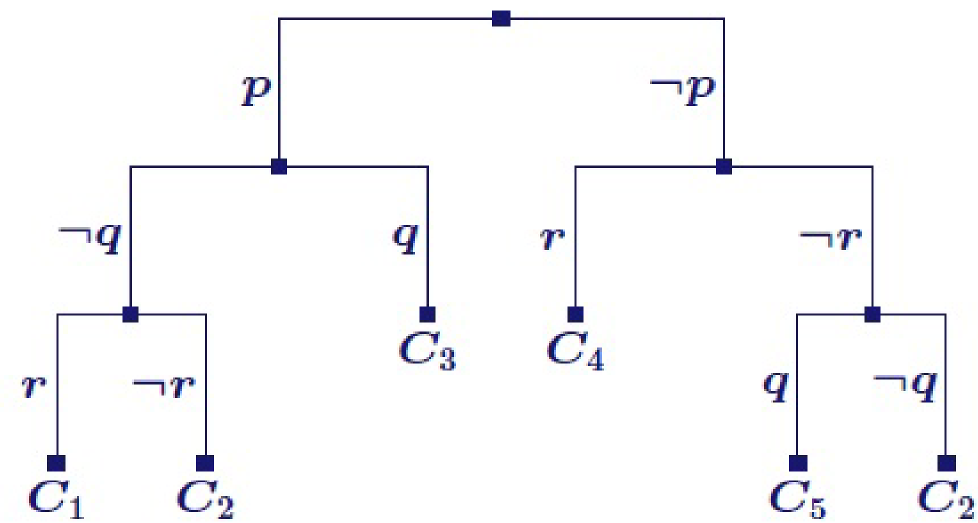

Image courtesy: http://www.computational-logic.org/iccl/master/lectures/summer07/sat/slides/semantictrees.pdf

S is unsatisfiable as the tree is closed

Herbrand's Theorem (Ver. 1)

<u>Theorem:</u>

A set S of clauses is unsatisfiable iff corresponding to every complete semantic tree of S, there is a finite closed semantic tree

<u>Proof:</u>

Part 1: Assume S is unsatisfiable

- Let T be the complete semantic tree for S

- For every branch B of T, we let I_B be the set of all literals attached to the links in B

Version 1 Proof (contd.)

- I_B is an interpretation of S (by definition)
- As S is unsatisfiable, I_B must falsify a ground instance of a clause C in S, let's call it C'
- T is complete, so, C' must be finite and there must exist a failure node N_B (a finite distance from root) on branch B
- Every branch of T has a failure node, so we find a closed semantic tree T' for S
- T' has a finite no. of nodes (Konig's Lemma)

<u>Hence, first half of thm. is proved</u>

Version 1 Proof (contd.)

Part 2: If there is a finite closed semantic tree for every complete semantic tree of S

- Then every branch contains a failure node
- i.e. every interpretation falsifies S
- Hence, S is unsatisfiable

Thus, both halves of the theorem are proved

Herbrand's Theorem (Ver. 2)

<u>Theorem:</u>

A set S of clauses is unsatisfiable iff there is a finite unsatisfiable set S' of ground instances of clauses of S

<u>Proof:</u>

Part 1: Assume S is unsatisfiable

- Let T be a complete semantic tree of S
- By ver. 1 of Herbrand Thm., there is a finite closed semantic tree T' corresponding to T

Version 2 Proof (contd.)

- Let S’ be a set of all the ground instances of clauses that are falsified at all failure nodes of T’
- S’ is finite since T’ contains a finite no. of failure nodes
- Since S’ is false in every interpretation of S’, S’ is also unsatisfiable

<u>Hence first half of thm. is proved</u>

Version 2 Proof (contd.)

Part 2: Suppose S' is a finite unsatisfiable set of gr. instances of clauses in S

- Every interpretation I of S contains an interpretation I' of S'
- So, if I' falsifies S', then I must also falsify S'
- Since S' is falsified by every interpretation I', it must also be falsified by every interpretation I of S
- i.e. S is falsified by every interpretation of S
- Hence S is unsatisfiable

<u>Thus, both halves of the thm. are proved</u>

Example

Let $S = \{P(x), \neg P(f(a))\}$

This set is unsatisfiable

Hence, by Herbrand's Theorem

there is a finite unsatisfiable set S' of ground

instances of clauses of S

One of these sets can be $S' = \{P(f(a)), \neg P(f(a))\}$

References

- Chang, Chin-Liang and Lee, Richard Char-Tung
  Symbolic Logic and Mechanical Theorem Proving
  Academic Press, New York, NY, 1973