ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-1756

Efficient Semantic Tableau Generation for Abduction in Propositional Logic

Yifan Yang [1] and **Ricardo De Aldama** [1,2] and **Jamal Atif** [2] and **Isabelle Bloch** [1]

1 Introduction

Abduction is a backward chaining inference, finding the best explanations of an observation with regard to a knowledge base in a two-steps process: i) hypotheses generation, and ii) explanations selection according to a minimality criterion. A propositional abduction problem $\mathcal{P}$ is a tuple $\langle \mathcal{V}, \mathcal{O}, \mathcal{T} \rangle$ [3], where $\mathcal{V}$ is a set of variables, $\mathcal{T}$ denotes the background knowledge of an application domain, $\mathcal{O}$ represents the observation, which is not directly entailed by the knowledge base $\mathcal{T} \not\models \mathcal{O}$, and $\mathcal{H}$ is a hypothesis of the problem $\mathcal{P}$ if $\mathcal{T} \cup \{\mathcal{H}\} \models \mathcal{O}$.

Example 1
Let us consider $\mathcal{V} : \{r, c, s, w, n, d\}$, $\mathcal{O} : \{w \wedge d\}$, $\mathcal{T} = \{\phi_1, \cdots, \phi_4\} : \phi_1 = r \rightarrow w$, if it rains, the ground will be wet; $\phi_2 = s \rightarrow w$, when the sprinkler is on, the ground will be wet; $\phi_3 = c \rightarrow r \wedge d$, if there are some heavy clouds, it will rain and get dark; $\phi_4 = n \rightarrow d$, during the night, it is dark. Potential explanations $\mathcal{H}$ can be c, $r \wedge n$, $s \wedge n$, $r \wedge c$, $s \wedge c$, and $c \wedge n$.

In our paper, we restrict $\mathcal{O}$ and $\mathcal{H}$ to be a conjunction of a set of literals. A hypothesis is an explanation if it satisfies the consistency ($\mathcal{T} \cup \{\mathcal{H}\}$ is consistent) and explanatory ($\mathcal{H} \not\models \mathcal{O}$) conditions. Semantic minimality is one of the important criteria in propositional logic [5, 6] and Description Logics [2] to select the most general explanations. $\mathcal{H}$ is a semantic-minimal explanation if there does not exist an explanation $\mathcal{H}'$ of $\mathcal{P}$ such that $\mathcal{T} \cup \{\mathcal{H}\} \models \mathcal{H}'$. One concrete algorithm to solve the propositional abduction problem, based on semantic tableaux, is proposed in [1, 5] and [7]. In this paper, we focus on improving the performance of the computation process of abductive reasoning in propositional logic by developing an optimized semantic tableau method to find semantic minimal explanations for a propositional abduction problem.

2 An optimized semantic tableau for propositional logic

2.1 Optimized semantic tableau

Definition 2.1 (*Relevant formula*)
Given a literal l and a formula ϕ, ϕ is a relevant formula for l if l appears in the normal form of ϕ. The set of relevant formulas for l is denoted by $J(l)$.

In Example 1, we have for instance $J(r) = \{c \rightarrow r \wedge d\}$.

[1] LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France, email: {yifan.yang, isabelle.bloch}@telecom-paristech.fr, ricardo.de.aldama@gmail.com
[2] PSL, Université Paris-Dauphine, LAMSADE, UMR 7243, Paris, France, email: jamal.atif@dauphine.fr

Definition 2.2 (*And-Or semantic tableau*)
Given a propositional abduction problem $\mathcal{P} = \langle \mathcal{V}, \mathcal{O}, \mathcal{T} \rangle$, the And-Or semantic tableau, denoted by $T_O(\mathcal{V}, \mathcal{O}, \mathcal{T})$ (T_O for short), is an And-Or tree structure where an And-node (called rule node) consists of a formula $\phi \in \mathcal{T}$ or $\neg O$ and an Or-node (literal node) consists of a set of literals $\{l_i\}$ derived from its parent rule node. The level of a node is the distance from the root to this node.

Remark 1. The semantic tableau proposed in [1], denoted by T_A, can be represented as a special structure using And-Or semantic tableau, denoted by T'_A: i) every literal node has a unique rule node as a child and this rule node is not necessarily a relevant formula of its parent literal node; ii) all the rule nodes appearing at the same level are identical formula; iii) all the rule nodes in the same branch are different formulas.

2.2 Tableau construction and hypotheses generation

We propose to construct a connected And-Or semantic tableau (denoted by T_G) exhibiting hierarchical connection structures such that, in contrast to the traditional tableau method, the construction process involves only relevant formulas in the knowledge base iteratively and intermediate hypotheses are generated in the node along with the development of the tableau. T_G is expanded step by step and terminated when a sub-tableau is closed or no more relevant formulas can be applied: i) level 0 is composed by the negation of the observation; ii) level 1 is composed by the union of negation of observation literals ($\neg o_i$) by applying expansion rules; iii) an edge derived from a rule node (ϕ) at level $2k$ to a literal node (a) at level $2k+1$ by applying expansion rules is denoted by $\phi \longrightarrow a$; iv) an edge derived from a literal node (a) at level $2k+1$ to a rule node (ϕ) at level $2(k+1)$ is denoted by $a \dashrightarrow \phi$, where ϕ is a relevant formula of $\neg a$ ($\phi \in J(\neg a)$). The connected And-Or semantic tableau of Example 1 is shown in Figure 1. An intermediate hypothesis $\mathcal{H}_{inter}$ of an elementary sub-tableau[3] is the complement of the literals in the open branches.

Definition 2.3 (*Hypothesis model*)
A hypothesis $\mathcal{H}$ of $\mathcal{P}$ is the conjunction of $\{\mathcal{H}^k_{inter} | k \in \{1, \cdots, n\}\}$ where $\{\mathcal{H}^k_{inter} | k \in \{1, \cdots, n\}\}$ represents n terminal elementary sub-tableaux in a simple sub-tableau[4] of T_O. The intermediate hypothesis model $\mathcal{M}(\mathcal{H}_{inter})$ of a sub-tableau S consists of truth assignments of a set of literals $\{l_i\}$, where l_i is the complement of the literal of an open branch or the literal of a closed branch in the leaf

[3] An elementary sub-tableau is a two-level sub-tableau rooted in a rule node.
[4] An And-Or semantic tableau is called *simple* if every literal node has at most a unique child node.

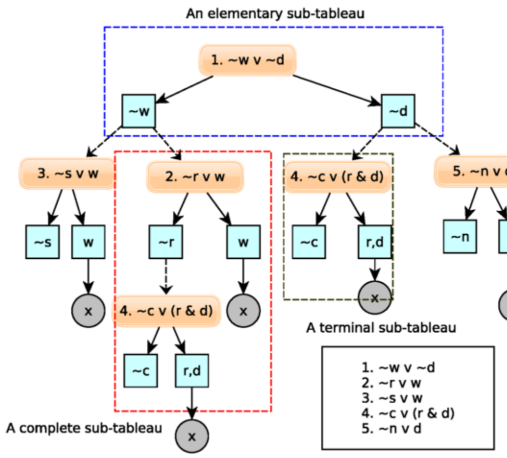

Figure 1. The tree-structure of the connected And-Or semantic tableau method for Example 1.

nodes including S and its ancestors. This set represents the semantic model of the intermediate hypothesis $\mathcal{H}_{inter}$ in the current elementary sub-tableau with respect to the ancestor elementary sub-tableau $\mathcal{T}_l = \bigwedge_{i \in I_l} \phi_i$. A hypothesis model is the union of the intermediate hypotheses models ($\mathcal{M}(\mathcal{H}) = \bigcup_{k \in \{1,\cdots,n\}} \mathcal{M}(\mathcal{H}^k_{inter})$).

A resolved knowledge base is a knowledge base with supplementary formulas derived by applying resolution on the original one $\mathcal{T}$, denoted by $\mathcal{T}_c$. $\mathcal{T}_c$ is represented by formulas rewritten in disjunctive normal form. For each literal l_i in $\mathcal{M}(\mathcal{H})$, we remove the conjunction containing its complement l_i in each formula. If one formula becomes an empty formula, the intermediate hypothesis is considered as inconsistent. We choose one rule node among alternative children nodes under each literal node, and the conjunction of the complements of all leaf literal nodes is one potential hypothesis.

In the exploration part, we update the initial explanation set once a new elementary sub-tableau is added in the tableau. The intermediate hypothesis model of this elementary sub-tableau is compared with the one in the hypotheses set. When $\mathcal{M}(\mathcal{H}_{inter}) \subseteq \mathcal{M}(\mathcal{H})$ is satisfied, the new elementary sub-tableau will be considered as a candidate to construct a hypothesis.

In Example 1, the initial hypotheses are $s \wedge n$, $s \wedge c$, $r \wedge n$ and $r \wedge c$. When the elementary sub-tableau associated with $\phi_3 = \neg c \vee (r \wedge d)$ is added below $\phi_1 = \neg r \vee w$, c is a new intermediate hypothesis. Therefore, a new explanation is a conjunction of the literal set obtained by replacing r by c in the initial hypothesis literal set $\{r, c\}$. The new explanation is c.

2.3 Soundness and completeness

Proposition 2. *The connected And-Or semantic tableau (T_G) and the semantic tableau proposed in [1] (T_A) provide the same sets of minimal hypotheses.*

Sketch of the proof: The proof relies on the following steps. We first show that T_G can be obtained from T'_A by a sequence Φ of transformations, combining deletion of branches that would lead to inconsistencies (closed branches), and "cut and paste", which consists in moving a semi-closed sub-tableau closer to the root of the tableau. We then prove that $\Phi(T'_A)$ and T'_A provide the same sets of hypotheses. Finally we prove that $\Phi(T'_A)$ and T_G provide the same sets of hypotheses.

3 Experimental evaluation

In this section, we present an empirical study to compare the computational cost of the proposed optimized semantic tableau with the algorithms in [1] and [4]. The theories used in our experiments to compare with semantic tableau method in [1] are generated in 3-CNF using a random generator[5]. The experiments are carried out by varying the number of formulas in the theory (namely for 10, 15 and 20 formulas). As evidenced in Table 1, our optimized semantic tableau based abduction has a significantly lower computation time for all the experiments compared to the semantic tableau method in [1].

Number of formulas	10	15	20
mean value (T_A)	999.64	3147.23	21435.59
std. deviation (T_A)	1083.80	2916.39	90244.14
mean value (T_G)	2.61	4.58	8.50
std. deviation (T_G)	4.29	7.24	13.63

Table 1. Computation cost of the two semantic tableau methods by varying the number of formulas in the theory.

The comparison between the semantic tableau method with the other abduction approaches using the diagnosis benchmark in [4] and Example 1 is shown in Table 2. The proposed method is competitive for the simple example presented in Example 1. However, the computation time increases along with the number of formulas in the theory. In the second example, the methods evaluated in [4] are more efficient. However, some explanations (c) are ignored for this example and the employed subset minimality is less restricted than semantic minimality.

Example (in ms)	T_A	T_G	ATMS	HS-DAG	MUS
Example 1	713	7	8	9	13
Diagnosis Horn theory	out of memory	85	17	17	30

Table 2. Computation time of all the methods.

Acknowledgements: This work has been supported by the French ANR project LOGIMA.

REFERENCES

[1] A. Aliseda-Llera, *Seeking explanations: abduction in logic, philosophy of science and artificial intelligence*, Ph.D. dissertation, University of Amsterdam, 1997.

[2] S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, and M. Mongiello, 'A uniform tableaux-based approach to concept abduction and contraction in $\mathcal{ALN}$', in *17th International Workshop on Description Logics (DL)*, volume 104, pp. 158–167, (2004).

[3] Nadia Creignou and Bruno Zanuttini, 'A complete classification of the complexity of propositional abduction', *SIAM Journal on Computing*, **36**(1), 207–229, (2006).

[4] R. Koitz and F. Wotawa, 'Finding explanations: an empirical evaluation of abductive diagnosis algorithms', in *International Workshop on Defeasible and Ampliative Reasoning (DARe-15)*, pp. 36–42, (2015).

[5] M. C. Mayer and F. Pirri, 'First order abduction via tableau and sequent calculi', *Logic Journal of IGPL*, **1**(1), 99–117, (1993).

[6] Gabriele Paul, 'Approaches to abductive reasoning: an overview', *Artificial Intelligence Review*, **7**(2), 109–152, (1993).

[7] F. Soler-Toscano, A. Nepomuceno-Fernández, and A. Aliseda-Llera, 'Abduction via C-tableaux and δ-resolution', *Journal of Applied Non-Classical Logics*, **19**(2), 211–225, (2009).

[5] Note that the generated inconsistent theories are not considered in our experiments.