DNF Hypotheses in Bottom-directed ILP

Katsumi Inoue

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

1 Introduction

To explain a given observation O, *explanatory induction* seeks a hypothesis H accommodated to the existing background knowledge B in such a way that

$$B \wedge H \models O, \quad \text{and} \tag{1}$$

$$B \wedge H \text{ is consistent.} \tag{2}$$

Often, H is constructed with some restricted vocabulary Γ called a *bias*. A formula H is called a *hypothesis* for the inductive problem (B, O) or (B, O, Γ).

Traditionally, H has been computed as a set of clauses, which is also interpreted as a formula in *conjunctive normal form* (CNF). CNF is useful in knowledge representation in AI because a CNF formula represents a set of rules and constraints, and each clause is regarded as a declarative statement of knowledge that holds in a domain. On the other hand, a formula in *disjunctive normal form* (DNF) is a conjunction of disjunctions of literals. A DNF formula can be regarded as a set of (partial) interpretations, and DNF has been used in the domains of logic circuit design and machine learning. However, most previous works of DNF hypotheses in machine learning have been done in the context of computational learning theories, and no previous results can be directly applied to explanatory induction in full clausal theories.

In this paper, we investigate an inductive framework which outputs hypotheses in DNF instead of CNF. We will show logical foundations for this framework and procedures to compute DNF hypotheses in explanatory induction. *Abduction* also infers hypotheses satisfying (1) and (2) in the form of sets (or conjunctions) of literals, which can be regarded as DNF hypotheses with single disjuncts. Our setting to compute DNF hypotheses is related to *model-based* inductive reasoning, in which propositional reasoning methods such as SAT techniques and prime implicant computation can be utilized. Then, enumeration of all DNF hypotheses can be more systematically performed than that of all CNF hypotheses.

2 Preliminaries

In this extended abstract, we mainly consider a propositional language, but will extend the work to the first-order case in Section 5. Given the set V of propositional variables, an *assignment* (or *interpretation*) I to V is a vector

in $\{0,1\}^V$, which can be identified with a subset I_t of V by interpreting the elements of I_t as true and those of $V \setminus I_t$ as false.

A (propositional) *formula* is constructed from V using the connectives $\vee$, $\wedge$, $\neg$, $\rightarrow$ and $\leftrightarrow$. An assignment I *satisfies* a formula φ if φ evaluates to true under I, and is called a *model* of φ. The set of all models of φ is denoted as $M(\varphi)$. A formula φ is *valid* if φ is true under any assignment, and is *unsatisfiable* if φ is false under any assignment, i.e., $M(\varphi) = \emptyset$. For formulas φ and ψ, we write $\varphi \models \psi$ if $M(\varphi) \subseteq M(\psi)$. In this case, φ is said to be a *generalization* of ψ, and ψ is *weaker than* (or equal to) φ. Note that $\varphi \models \psi$ iff $\varphi \rightarrow \psi$ is valid.

A *literal* is either a propositional variable $v_i \in V$ or its negation $\neg v_i$. A *term* is a conjunction or a disjunction of literals; a conjunction is called a *monomial*, and a disjunction is called a *clause*. Terms are also considered as sets of literals. Term T_1 *covers* (or *subsumes*) term T_2 if $T_1 \subseteq T_2$. A formula is in *disjunctive normal form* (DNF) if it is a disjunction of monomials. A formula is in *conjunctive normal form* (CNF) if it is a conjunction of clauses. A DNF (resp. CNF) formula φ (*theory-*)*subsumes* a DNF (resp. CNF) formula ψ if, for any term $T_2 \in \psi$, there exists a term $T_1 \in \varphi$ such that T_1 covers T_2.

An *implicant* of a formula φ is a monomial C such that $C \rightarrow \varphi$ is valid, i.e., $C \models \varphi$. An implicant C of φ is *prime* iff for every proper subset $S \subseteq C$ it holds that S is not an implicant of φ. Then, a prime implicant of φ is a weakest implicant of φ. Similarly an *implicate* of φ is a clause D such that $\varphi \rightarrow D$ is valid, i.e., $\varphi \models D$. *Prime implicates* are defined in the same way as prime implicants.

The next property is important: If $C_1, \ldots, C_n$ are implicants of a formula φ, then $C_i \models C_1 \vee \cdots \vee C_n \models \varphi$ for any $i = 1, \ldots, n$. That is, a DNF formula $C_1 \vee \cdots \vee C_n$ is a weaker generalization of φ than each implicant C_i of φ.

Proposition 2.1. *The disjunction δ^* of prime implicants of a formula φ is equivalent to φ, i.e., $\delta^* \leftrightarrow \varphi$ is valid. Hence, δ^* is a weakest generalization of φ.*

3 DNF Hypotheses

We here consider an inductive problem (B, O) in Section 1, where B is a background theory and O is an observation. Here, the inductive bias Γ is set to the representation language itself, but we will allow any bias Γ in Section 4. It is also assumed that B and O are CNF formulas such that $B \wedge O$ is consistent; otherwise O cannot be explained. In this case, the next property holds.

Proposition 3.1. *$(B \rightarrow O)$ is a weakest hypothesis for (B, O).*

The weakest hypothesis $H^* = (B \rightarrow O)$ is called the *(bottommost) bottom theory*. Any ILP method based on *inverse entailment* (IE) [6, 9, 3] also constructs some bottom theory $\bot(B, O)$ that satisfies $\bot(B, O) \models H^*$. For example, Progol [6] searches a hypothesis that subsumes the *bottom clause*, and CF-induction [3] starts from the *characteristic clauses* instead of the bottom clause. In IE, the relation (1) is converted to $B \wedge \neg O \models \neg H$, and consequences CC of $B \wedge \neg O$ are computed in CNF. To get a CNF hypothesis H from the CNF formula CC,

it is necessary to convert a DNF formula $\neg CC$ into CNF (*distribution*), which is expensive in general [9]. Moreover, the *generalization* task to obtain a CNF formula H from $\bot(B, O)$ such that $H \models \bot(B, O)$ is achieved in various ways [9, 3]. To avoid those computational problems in explanatory induction, we take another approach that constructs DNF hypotheses instead of computing CNF hypotheses, and will show two methods for their computation.

3.1 Model-based Approach

A set $\mathcal{M}$ of (partial) models can be identified with a DNF formula $\delta_{\mathcal{M}}$ by $\delta_{\mathcal{M}} = \bigvee_{I \in \mathcal{M}} (\bigwedge_{l:I(M)=true} l \wedge \bigwedge_{l:I(l)=false} \neg l)$. For $H^* = (B \rightarrow O) = (\neg B \vee O)$, the disjunction δ_{H^*} of all models in $M(\neg B \vee O)$ is thus in DNF, and is equivalent to H^*. Hence, any subdisjunction of δ_{H^*} is a generalization of H^*. Then, a possible way to compute a DNF hypothesis H is to select a set S of models of $\neg B \vee O$, i.e., $S \subseteq M(\neg B \vee O)$. The disjunction $D_S = \bigvee_{I \in S} I$ is in DNF, and is a hypothesis provided that $B \wedge D_S$ is consistent, i.e., $M(B \wedge D_S) \neq \emptyset$. By $M(D_S) = S$, $M(B \wedge D_S) = M(B) \cap M(D_S) = M(B) \cap S \neq \emptyset$. By assumption, $M(B) \neq \emptyset$. Then, we must have $M(S) \neq \emptyset$. Hence, a naïve (non-deterministic) *model-based procedure* to compute DNF hypotheses, $\mathsf{MB\text{-}DNF}(B, O)$, is obtained as follows:

1. Let $\mathcal{M} = M(\neg B \vee O)$;
2. Select a non-empty subset S of $\mathcal{M}$ such that $M(B) \cap S \neq \emptyset$ as follows;
   (a) Identify $\mathcal{D} = \mathcal{M} \cap M(B) = M(B \wedge O)$;
   (b) If $\mathcal{D} = \emptyset$, then output "No solution";
   (c) Else, select any set S such that $S \subseteq \mathcal{M}$ and $S \cap \mathcal{D} \neq \emptyset$;
3. Output the disjunction of the elements of S as a hypothesis.

Model computation in Steps 1 and 2(a) can be done by a model enumeration procedure based on efficient modern SAT techniques, e.g., [4].

Proposition 3.2 (Soundness and Completeness of MB-DNF).

(1) *If* $\mathsf{MB\text{-}DNF}(B, O)$ *returns a formula* φ*, then* φ *is a hypothesis for* (B, O).
(2) *For any hypothesis* H *for* (B, O)*, there is a DNF formula* φ *obtainable from* $\mathsf{MB\text{-}DNF}(B, O)$ *such that* $M(\varphi) = M(H)$.

Example 3.1. Suppose $V = \{p, q\}$ and consider $B = \neg p$ and $O = q$. Then $\mathcal{M} = M(\neg B \vee O) = M(p \vee q) = \{\{p\}, \{q\}, \{p, q\}\}$. Also $\mathcal{D} = M(B \wedge O) = M(\neg p \wedge q) = \{\{q\}\}$. We can choose any set from: $S_1 = \{\{q\}\}$, $S_2 = \{\{q\}, \{p\}\}$, $S_3 = \{\{q\}, \{p, q\}\}$, and $S_4 = \mathcal{M}$. These sets form DNF hypotheses: $\varphi_1 = (\neg p \wedge q)$, $\varphi_2 = (\neg p \wedge q) \vee (p \wedge \neg q)$, $\varphi_3 = (\neg p \wedge q) \vee (p \wedge q)$, and $\varphi_4 = (\neg p \wedge q) \vee (p \wedge \neg q) \vee (p \wedge q)$. Note that the last three DNF formulas are respectively equivalent to the formulas: $H_2 = (\neg p \leftrightarrow q)$, $H_3 = q$, and $H_4 = (p \vee q)$.

3.2 Prime Implicant-based Approach

Although $\mathsf{MB\text{-}DNF}(B, O)$ is correct, we must compute $M(\neg B \vee O)$ and $M(B \wedge O)$ in the *complete* form, i.e., assigning true/false to all variables in V, which are not feasible for a large B with many propositional variables. It is thus desirable to have a set of partial models of $\neg B \vee O$ which can replace $M(\neg B \vee O)$ at Step 1.

For any formula φ, we denote the set of its prime implicants as $PI(\varphi)$. Instead of computing $M(\neg B \vee O)$, the next procedure computes $PI(\neg B \vee O)$. Since each prime implicant is a weakest implicant of $\neg B \vee O$, any disjunction π of prime implicants can be a hypothesis provided that $B \wedge \pi$ is consistent. Since the consistency condition can be transformed to $\pi \not\models \neg B$, we need to check whether π is not an implicant of $\neg B$. Hence, a (non-deterministic) *PI-based procedure* to compute DNF hypotheses, $\mathsf{PI\text{-}DNF}(B, O)$, is obtained as follows:

1. Let $\mathcal{P} = PI(\neg B \vee O)$;
2. Select a non-empty subset S of $\mathcal{P}$ such that $S \not\subseteq PI(\neg B)$ as follows;
   (a) Identify $\mathcal{N} = \mathcal{P} \setminus PI(\neg B)$;
   (b) If $\mathcal{N} = \emptyset$, then output "No solution";
   (c) Else, select any set S such that $S \subseteq \mathcal{P}$ and $S \cap \mathcal{N} \neq \emptyset$;
3. Output the disjunction of the elements of S as a hypothesis.

Prime implicants in Steps 1 and 2(a) can be computed by some procedures in the literature, e.g., Tison's consensus method and its variants [5]. Weak completeness of the procedure holds by Proposition 2.1: For any hypothesis H for (B, O), there is a DNF formula π obtainable from $\mathsf{PI\text{-}DNF}(B, O)$ such that π is weaker than or equal to H. In fact, we have a stronger completeness as follows.

Proposition 3.3 (Soundness and Completeness of PI-DNF).

(1) *If* $\mathsf{PI\text{-}DNF}(B, O)$ *returns a formula* π*, then* π *is a hypothesis for* (B, O)*.*
(2) *For any DNF hypothesis* H *for* (B, O)*, there is a DNF formula* π *obtainable from* $\mathsf{PI\text{-}DNF}(B, O)$ *such that* π *theory-subsumes* H*.*

Example 3.2. Suppose the same inductive problem (B, O) as Example 3.1, i.e., $B = \neg p$ and $O = q$. Then $\mathcal{P} = PI(\neg B \vee O) = \{p, q\}$, and $PI(\neg B) = \{p\}$. Hence $\mathcal{N} = \mathcal{P} \setminus PI(\neg B) = \{q\}$. We can choose any set from: $T_1 = \{q\}$ and $T_2 = \mathcal{P}$. These sets form DNF hypotheses: $\pi_1 = q$ and $\pi_2 = (p \vee q)$. Note that $\pi_1 = H_3$ and $\pi_2 = H_4$ in Example 3.1, and that π_1 covers φ_1 and π_2 covers H_2.

As seen in the last example, $\mathsf{PI\text{-}DNF}(B, O)$ generally outputs DNF hypotheses in much simpler forms than $\mathsf{MB\text{-}DNF}(B, O)$. Actually, each monomial in π output by $\mathsf{PI\text{-}DNF}(B, O)$ is *subsumption-minimal*, so we have a compact representation of the DNF hypotheses. The difference of average sizes of hypotheses between these two procedures becomes much larger as the number of variables increases.

Example 3.3. The following theory in [3] is originally by Wray Buntine (1988):

$$B = (\mathit{cat} \rightarrow \mathit{pet}) \ \wedge \ (\mathit{small} \wedge \mathit{fluffy} \wedge \mathit{pet} \rightarrow \mathit{cuddly\_pet}),$$
$$O = (\mathit{fluffy} \wedge \mathit{cat} \rightarrow \mathit{cuddly\_pet}).$$

Then $\neg B \vee O$ is the DNF formula:

$$(\mathit{cat} \wedge \neg \mathit{pet}) \vee (\mathit{small} \wedge \mathit{cat} \wedge \neg \mathit{cuddly\_pet}) \vee (\neg \mathit{fluffy} \vee \neg \mathit{cat} \vee \mathit{cuddly\_pet}).$$

Each of the following prime implicants is not an implicant of $\neg B$:

$$PI(\neg B \vee O) = \{\neg \mathit{fluffy}, \neg \mathit{cat}, \neg \mathit{pet}, \mathit{cuddly\_pet}, \mathit{small}\}. \tag{3}$$

Taking the disjunction of any non-empty subset of (3) provides a hypothesis, and hence $2^5 - 1 = 31$ hypotheses are obtainable. The average number of literals

in each hypothesis is approximately 2.5. For example, if all monomials are taken into account, we have a weakest hypothesis that is equivalent to $B \to O$:

$$H = (\mathit{fluffy} \wedge \mathit{cat} \wedge \mathit{pet} \to \mathit{cuddly\_pet} \vee \mathit{small}).$$

On the other hand, if $\mathsf{MB\text{-}DNF}(B, O)$ is used, the number of models in $\mathcal{M} = M(\neg B \vee O)$ is $2^5 - 1 = 31$ and the number of models in $\mathcal{D} = M(B \wedge O)$ is $2^5 - (2^3 + 2^1 + 2^2) = 18$, then the number of possible combinations of a non-empty subset of $\mathcal{D}$ and a subset of $\mathcal{M} \setminus \mathcal{D}$ becomes $(2^{18} - 1) \times (2^{31-18}) \approx 2.147 \times 10^9$, and the average size of each hypothesis before simplification is $15.5 \times 5 = 77.5$.

4 Abduction and Signature Restriction

In the previous section, we do not assume any restriction on hypotheses. Here we consider a bias Γ given as a set of literals allowed to appear in hypotheses. In the case of abduction, each literal in Γ is called an *abducible*. We now characterize abductive hypotheses in terms of prime implicants. Let $\mathcal{A}$ be a set of literals. An implicant C of a formula φ is *signature-restricted with respect to* $\mathcal{A}$ (or an $\mathcal{A}$*-restricted implicant* of φ) if all literals in C belongs to $\mathcal{A}$. Then, the set of signature-restricted implicants is closed under subsumption: If C is an $\mathcal{A}$-restricted implicant of φ and C' subsumes C, then C' is also an $\mathcal{A}$-restricted implicant. The set of all $\mathcal{A}$-restricted prime implicants of φ is denoted as $PI_{\mathcal{A}}(\varphi)$.

Proposition 4.1. *Let (B, O, Γ) be an abductive problem. The set of subsumption-minimal abductive hypotheses for (B, O, Γ) is exactly $PI_\Gamma(\neg B \vee O) \setminus PI(\neg B)$.*

Proposition 4.1 is contrasted with formalizations of abductive hypotheses [8, 5, 2], which are based on prime implicates instead of prime implicants. In fact, there is a *duality* between these formalizations, and hence procedures to compute $\mathcal{A}$-restricted prime implicates [2] can be utilized for our purpose. Moreover, when B is a set of propositional Horn clauses, there are some efficient procedures to enumerate abductive hypotheses [1]. The procedure $\mathsf{PI\text{-}DNF}(B, O)$ can then be adapted to compute abductive hypotheses by computing Γ-restricted prime implicants and selecting those elements in $\mathcal{N}_\Gamma = PI_\Gamma(\neg B \vee O) \setminus PI(\neg B)$ at Step 2. Similarly, inductive hypotheses with a bias Γ can also be devised by changing Step 2(c) to select $S \subseteq PI_\Gamma(\neg B \vee O)$ such that $S \cap \mathcal{N}_\Gamma \neq \emptyset$.

5 Extension to the First-order Case

We here transfer the concept of *characteristic clauses* [2] to its dual concept in first-order logic. A monomial C_1 *subsumes* a monomial C_2 if there is a substitution θ such that $C_1\theta \subseteq C_2$. Given a first-order open formula φ whose variables are assumed to be existentially quantified at the front, a monomial C is called a *generalization* of φ if $\exists C \models \exists\varphi$, where $\exists\varphi$ is the existential closure of φ, and is further said $\mathcal{A}$*-restricted* if all literals in C belongs to $\mathcal{A}$. The set of $\mathcal{A}$-restricted

generalizations of φ is denoted as $Gen_{\mathcal{A}}(\varphi)$. Then the *characteristic monomials of φ with respect to $\mathcal{A}$* is defined as $CM(\varphi, \mathcal{A}) = \mu\, Gen_{\mathcal{A}}(\varphi)$, where μS denotes the set of monomials in S that are minimal with respect to subsumption.

We can now define the dual concept of CF-induction [3], called *GF-induction* (generalization-finding induction), by lifting the procedure $\mathsf{PI\text{-}DNF}(B, O)$ to the first-order case, in which the prime implicants PI are simply replaced with the characteristic monomials CM. Computing characteristic monomials can be done in the dual form using the first-order consequence-finding procedure SOLAR [7]. For a formula α, the *dual* of α, denoted as α^d, is the formula obtained from α by swapping $\wedge$ and $\vee$ (and $\forall$ and $\exists$). For a CNF formula Σ, $Carc(\Sigma, \mathcal{A})$ is the *characteristic clauses* of Σ with respect to $\mathcal{A}$, which is the dual concept of characteristic monomials, and $\mathtt{Skolem}(\Sigma)$ is a Skolemization of Σ. Then, the resulting procedure of GF-induction, $\mathsf{GF\text{-}ind}(B, O, \mathcal{A})$, is defined as follows:

1. Let F be the CNF formula $(\neg B)^d \wedge \mathtt{Skolem}(O^d)$, and $\mathcal{P} = Carc(F, \mathcal{A})$;
2. Select a non-empty subset S of $\mathcal{P}$ as follows:
   (a) Identify $\mathcal{N} = \mathcal{P} \setminus Carc((\neg B)^d, \mathcal{A})$;
   (b) If $\mathcal{N} = \emptyset$, then output "No solution";
   (c) Else, select any set $S \subseteq \mathcal{P}$ such that $S \cap \mathcal{N} \neq \emptyset$;
3. Let S^d be the DNF formula $\bigvee_{D \in S}(\bigwedge_{l \in D} l)$;
4. Generalize S^d to a hypothesis H such that $B \wedge H$ is consistent.

In $\mathsf{GF\text{-}ind}(B, O, \mathcal{A})$, Steps 1, 2 and 4 are similar to the corresponding steps in CF-induction [3], but Step 3 converts the dual of a hypothesis to the original form and distribution (from CNF to DNF) is not necessary unlike CF-induction.

As in the case of CF-induction, GF-induction can be proved to be sound and complete in the class of induction problems with first-order full clausal theories.

References

1. Eiter, T. and Makino, K., On computing all abductive explanations from a propositional Horn theory, *J. ACM*, 54(5), Article 24 (2007)
2. Inoue, K., Linear resolution for consequence finding, *Artificial Intelligence*, 56:301–353 (1992)
3. Inoue, K., Induction as consequence finding, *Machine Learning*, 55:109–135 (2004)
4. Jin, H.S., Han, H.J. and Somenzi, F., Efficient conflict analysis for finding all satisfying assignments of a Boolean circuit, *Proceedings of TACAS'05*, LNCS 3440, pp.287–300, Springer (2005)
5. Kean, A. and Tsiknis, G., An incremental method for generating prime implicants/implicates, *J. Symbolic Computation*, 9:185–206 (1990)
6. Muggleton, S., Inverse entailment and Progol, *New Generation Computing*, 13:245–286 (1995)
7. Nabeshima, H., Iwanuma, K., Inoue, K. and Ray, O., SOLAR: An automated deduction system for consequence finding, *AI Communications*, 23:183–203 (2010)
8. Reiter, R. and de Kleer, J., Foundations of assumption-based truth maintenance systems: preliminary report, *Proceedings of AAAI-87*, pp.183–187 (1987)
9. Yamamoto, A. and Fronhöfer, B., Hypotheses finding via residue hypotheses with the resolution principle, in: *Proceedings of ALT 2000*, LNAI 1968, pp.156–165, Springer (2000)