BRAVE INDUCTION

CHIAKI SAKAMA
Wakayama University
KATSUMI INOUE
National Institute of Informatics

ILP 2008, Prague

Background

- Typical ILP tasks construct hypotheses to explain observations using background knowledge.
- Given the background knowledge B and an observation O, a hypothesis H covers O under B if
  - $B \wedge H \models O$
  - $B \wedge H$ is consistent.
- The condition is often too strong for building possible hypotheses.

Motivating Example

- There are 30 students in a class, of which 20 are European, 7 are Asian, and 3 are American.

  B: student(1) $\land \cdots \land$ student(30),

  O: euro(1) $\land \cdots \land$ euro(20) $\land$ asia(21) $\land \cdots \land$ asia(27) $\land$ usa(28) $\land \cdots \land$ usa(30)

- In this situation, the clause

  H: euro(x) $\lor$ asia(x) $\lor$ usa(x) $\leftarrow$ student(x)

  appears a good candidate of hypothesis.

- However, H does not satisfy the relation $B \land H \models O$. In fact, $B \land H$ has many models in which O is not true, e.g., {student(1),...,student(30),euro(1),...,euro(30)}.

What is the Problem?

- When B and H are Horn theories, the relation $B \wedge H \models O$ (*) represents that O is true in the unique minimal model of $B \wedge H$.
- When $B \wedge H$ contains indefinite information, $B \wedge H$ becomes a non-Horn theory which has multiple minimal models in general.
- In this case, the relation (*) excludes a possible hypothesis H due to the existence of a single minimal model of $B \wedge H$ in which O is not true.

Contribution

- To cope with the problem, we introduce a new form of induction called brave induction.
- We investigate formal properties and develop an algorithm of brave induction.
- The framework is extended to induction in answer set programming.
  (This part is not included in this talk.)

Brave Reasoning vs. Cautious Reasoning

- Brave (or credulous) reasoning and cautious (or skeptical) reasoning are used in nonmonotonic logics and disjunctive logic programs.
- A formula F is a consequence of brave/cautious inference in a theory T (under the minimal model semantics) if F is true in some/every minimal model of T.
- Brave/cautious reasoning is used in hypothetical reasoning in AI such as abduction.

Brave/Cautious Abduction

- B: light_off ← power_failure,
  light_off ← broken_bulb,
  broken_bulb v melted_fuse ← high_current,
  Abducible: power_failure, high_current.
- O: light_off
- E1=power_failure is the unique (minimal) explanation in cautious abduction, since O is true in every minimal model of B ∧ E1.
- In addition to E1, E2=high_current is also a (minimal) explanation in brave abduction, since O is true in some minimal model of B ∧ E2.

Brave Induction in Clausal Logic

- B, O, and H are all consistent clausal theories.
- A hypothesis H covers O under B in brave induction if a consistent theory $B \land H$ has a minimal model satisfying O. In this case, H is called a solution of brave induction.
- In this sense, explanatory induction in ILP is considered cautious induction.

Properties (1)

- Existence of solutions: Brave induction has a solution iff $B \wedge O$ is consistent.
- Relation to cautious induction: If H covers O under B in cautious induction, H is a solution of brave induction. The converse holds when B is a Horn theory.
- Necessary condition of solutions: If H is a solution of brave induction, $B \wedge H \wedge O$ is consistent.

Properties (2)

- Generalization of solutions: For any theory $H' \models H$ such that $B \wedge H'$ is consistent, if H is a solution of brave induction, so does H'.
- Nonmonotonicity: The fact that H1 and H2 are solutions of brave induction does not imply that $H1 \wedge H2$ is a solution.
- Disjunctive combination of solutions: If H1 and H2 are solutions of brave induction, so is $H1 \vee H2$.

Properties (3)

- Conjunctive combination of observations:
  The fact that H covers both O1 and O2 under B does not imply H covers O1∧O2 under B.
- Ex) Let B={ p(x) v q(x) ← r(x), s(a)←}, O1={p(a)}, and O2={q(a)}.
  Then, H={ r(x) ← s(x) } covers both O1 and O2 under B in brave induction, but H does not cover O1∧O2 under B.
  Note that B∧H has two minimal models {p(a),r(a),s(a)} and {q(a),r(a),s(a)}.

Comparison of Properties

	Brave ind.	Cautious ind.
Generalization of solutions	●	●
Nonmonotonicity	●	●
Disjunctive combination of solutions	●	●
Conjunctive combination of observations	×	●

Computation

- Proposition:
  B: background knowledge,
  H: a hypothesis,
  O: an observation.

  $B \wedge H$ has a minimal model satisfying O
  iff there is a disjunction F of ground atoms s.t.

  $B \wedge H \models O \vee F$ and $B \wedge H \not\models F$.

- Assumption:
  - O is a conjunction of ground atoms.
  - H is a finite clausal theory s.t. the head of each clause has the predicate appearing in O.

Terms and Notions

- A ground clause C is prime wrt T if $T \models C$ but $T \not\models C'$ for any $C' \subset C$.
- A DNF formula $F = C_1 \vee \cdots \vee C_k$ is irredundant if $F \not\equiv F'$ for any $F = C_1 \vee \cdots \vee C_{i-1} \vee C_{i+1} \vee \cdots \vee C_k$.
- Given an atom A,
  pred(A) is the predicate of A,
  term(A) is the set of terms in A,
  const(A) is the set of constants in A.
- Given a clause C, head(C) is the head of C, and body(C) is the body of C.

Step 1: Computing ground hypotheses

- $B \wedge H \models O \vee F$ implies $B \wedge \neg O \models \neg H \vee F$.
- H is a clausal theory and F is a disjunction of ground atoms, then $\neg H \vee F$ is a DNF formula.
- To get $\neg H \vee F$, first compute prime CNF formulas from $B \wedge \neg O$. Then, construct an irredundant DNF formula from it.
- From $\neg H \vee F$, extract $\neg H$.
- By $\neg H$, we can obtain H.

Step 2: Generalization

- O is partitioned into disjoint subsets

  $O = O1 \wedge \cdots \wedge On$

  where Oi is a conjunction of ground atoms having the same predicate.
- Correspondingly, H is partitioned as

  $H = H1 \wedge \cdots \wedge Hn$

  where Hi is a conjunction of clauses whose heads contain the predicate in Oi.
- Compute the least generalization under subsumption (LGS) of each Hi and collect it as

  $lgs(H) = lgs(H1) \wedge \cdots \wedge lgs(Hn).$

Step 3: Construct a weak form of hypotheses

- Given a set S of atoms, suppose two atoms A1 and A2 in S s.t. pred(A1)≠pred(A2). Then, pred(A1) and pred(A2) are synchronous in S if const(A1)∩const(A2)≠φ.
  Otherwise, they are asynchronous in S.
- A set S is asynchronous if any pair of different predicates is asynchronous in S.
- When an observation O is an asynchronous set, take the greatest specialization under implication (GSI) of lgs(H1),...,lgs(Hn) as
  gsi(lgs(H1),...,lgs(Hn)) = lgs(H1)∨・・・∨ lgs(Hn).

Step 4: Optimization

- Two atoms A1 and A2 are linked if $term(A1) \cap term(A2) \neq \varphi$.
- Given a clause C, an atom $A \in body(C)$ is isolated in C if there is no atom $A'(\neq A)$ in C s.t. A' and A are linked.
- For any clause C in lgs(Hi),

  1. remove any atom A from head(C) s.t. pred(A) is not included in O,
  2. remove any atom A from body(C) s.t. A is isolated in C.

The result of such reduction is denoted by lgs*(Hi).

An Algorithm for Brave Induction

Procedure: BRAIN

Input: B and O;

Output: hypotheses $H^{\wedge}$ and $H^{\vee}$.

Step 1: Compute ground and irredundant DNF formulas $\neg H \vee F$ from $B \wedge \neg O$, and extract $\neg H$ from $\neg H \vee F$.

Step 2: Compute lgs(H).

Step 3: If O is asynchronous and is partitioned into $O = O1 \wedge \cdots \wedge On$, compute gsi(lgs(H1),...,lgs(Hn)).

Step 4: If $B \wedge lgs^*(Hi)$ is consistent, put

$H^{\wedge} = lgs^*(H1) \wedge \cdots \wedge lgs^*(Hn)$ and

$H^{\vee} = lgs^*(H1) \vee \cdots \vee lgs^*(Hn)$.

Main Theorem

- Any hypothesis computed by BRAIN becomes a solution of brave induction.

Example

B: $teacher(0) \wedge student(1) \wedge \cdots \wedge student(30)$,

O: $euro(1) \wedge \cdots \wedge euro(20) \wedge asia(21) \wedge \cdots \wedge asia(27)$
$\wedge usa(28) \wedge \cdots \wedge usa(30)$

First, the ground and irredundant DNF formula
$\neg H1 \vee \neg H2 \vee \neg H3$ is derived from $B \wedge \neg O$ where

$H1 = (\neg B \vee euro(1)) \wedge \cdots \wedge (\neg B \vee euro(20))$

$H2 = (\neg B \vee asia(21)) \wedge \cdots \wedge (\neg B \vee asia(27))$

$H3 = (\neg B \vee usa(28)) \wedge \cdots \wedge (\neg B \vee usa(30))$.

Example

Next, O is partitioned into $O = O1 \wedge O2 \wedge O3$ where
$O1=\{euro(1),...,euro(20)\}$, $O2=\{asia(21),...,asia(27)\}$,
$O3=\{usa(28),...,usa(30)\}$.

Then, $lgs(H) = lgs(H1) \wedge lgs(H2) \wedge lgs(H3)$ where
$lgs(H1) = \neg teacher(0) \vee \neg student(x) \vee euro(x)$,
$lgs(H2) = \neg teacher(0) \vee \neg student(x) \vee asia(x)$,
$lgs(H3) = \neg teacher(0) \vee \neg student(x) \vee usa(x)$.

As O is asynchronous, $gsi(lgs(H1),...,lgs(Hn))$ becomes
$lgs(H1) \vee lgs(H2) \vee lgs(H3)$.

Finally, teacher(0) is isolated in each lgs(Hi),
so removing it from each lgs(Hi), and get lgs*(Hi).

Example

As a result, $H^{\wedge} = lgs^*(H1) \wedge \cdots \wedge lgs^*(Hn)$ becomes

$euro(x) \leftarrow student(x)$

$\wedge\ asia(x) \leftarrow student(x)$

$\wedge\ usa(x) \leftarrow student(x)$,

and $H^{\vee} = lgs^*(H1) \vee \cdots \vee lgs^*(Hn)$ becomes

$euro(x) \vee asia(x) \vee usa(x) \leftarrow student(x)$.

Thus, $H^{\wedge}$ and $H^{\vee}$ becomes two solutions of brave induction.

Discussion
Relation to Learning from Satisfiability (LFS)

- A hypothesis H covers O under B in LFS if B∧H has a model satisfying O (De Raedt and Dehaspe, 1997).
- If H covers O under B in brave induction, H covers O under B in LFS. The converse implication does not hold in general.
- LFS does not require the minimality of models, and any H which is consistent with B∧O becomes a solution. (e.g., B={p(a)}, O={q(a)}, H={r(b)}.)
- LFS generally produces many useless hypotheses, and brave induction reduces hypotheses space.

Discussion

Relation to Confirmatory Induction

- A hypothesis H covers O under B in confirmatory induction (or descriptive induction) if $Comp(B \wedge O) \models H$ where Comp is predicate completion.
- There is no stronger/weaker relation between confirmatory induction and brave induction.
- Confirmatory induction does not explain why particular individuals are observed under B, and the aim is to learn relations between concepts.

Discussion

Relation to CF-Induction

- CF-induction (Inoue, 2004) applies Muggleton's inverse entailment to full clausal theories.
- CF-induction is cautious induction and is stronger than brave induction.

Discussion
Handling Negative Observations

- Given a negative observation N, it is requested that $B \wedge H \not\models N$.
- Given a positive observation P and a negative one N, H is a solution of brave induction if $B \wedge H$ has a minimal model M such that $M \models P$ and $M \not\models N$.
- By putting $O = P \wedge \neg N$, negative observations are handled within the framework of this paper.

Discussion
Computational Complexity

- Given a ground theory B and a ground observation O, deciding the existence of solutions in brave induction is NP-complete.
(This is also the case for cautious induction.)
- Identifying whether a propositional theory H is a solution of brave induction is Σ_2^P-complete.
In cautious induction, the task is coNP-complete.
- Brave induction appears more expensive than cautious induction for identifying solutions.

Conclusion

- Brave induction is weaker than explanatory (or cautious) induction, and stronger than learning from satisfiability. It is useful for learning indefinite or incomplete theories.
- Brave induction is used for automated negotiation in multiagent systems for building proposals [Sakama, DALT-08].
- A candidate of practical application is system biology which would have indefinite or incomplete information in the background knowledge and observations.