Abductive Reasoning

CMPT 411/721

Topics

- Reasoning with Prime Implicates
- Abduction

Bottom-Up Reasoning Via Prime Implicates

Here we give a different flavour for reasoning with ground (variable-free) sets of clauses.

Idea:

- Begin with formulas expressed in *clause normal form*.

Bottom-Up Reasoning Via Prime Implicates

Here we give a different flavour for reasoning with ground (variable-free) sets of clauses.

Idea:

- Begin with formulas expressed in *clause normal form*.
- Then find the set of *prime implicates* of the set of clauses.

Bottom-Up Reasoning Via Prime Implicates

Here we give a different flavour for reasoning with ground (variable-free) sets of clauses.

Idea:

- Begin with formulas expressed in *clause normal form*.
- Then find the set of *prime implicates* of the set of clauses.
  - ☞ Notably, prime implicates allow for very efficient reasoning.

Bottom-Up Reasoning Via Prime Implicates

Here we give a different flavour for reasoning with ground (variable-free) sets of clauses.

Idea:

- Begin with formulas expressed in *clause normal form*.
- Then find the set of *prime implicates* of the set of clauses.
  - ☞ Notably, prime implicates allow for very efficient reasoning.

- This is an example of *knowledge compilation*
  - I.e. transform a KB so that inference is efficient.
  - Reasoning will be done using the prime implicates.
  - This is also used in a form of diagnosis, called *abduction*.
    - (Abduction can be though of reasoning backwards from symptoms to causes)

Knowledge Compilation

Idea: Given a general knowledge base KB, transform it to KB' so that

- $KB \equiv KB'$ but
- Determining whether $KB' \models \phi$ can be carried out more efficiently than $KB \models \phi$.

☞ Here KB' is the set of prime implicates (next slide) of KB

Prime Implicates

- An *implicate* of a theory is a clause that logically follows from that theory.
  - I.e. an implicate of KB is a clause c such that $KB \models c$.
  - E.g. $\neg a \vee b$ and $\neg b \vee c$ have implicates $\neg a \vee c \vee d$ and $\neg a \vee a \vee d$.

Prime Implicates

- An *implicate* of a theory is a clause that logically follows from that theory.
  - I.e. an implicate of KB is a clause c such that $KB \models c$.
  - E.g. $\neg a \vee b$ and $\neg b \vee c$ have implicates $\neg a \vee c \vee d$ and $\neg a \vee a \vee d$.
- A *minimal implicate* is a clause that has no strict subset as an implicate.
  - I.e. a minimal implicate of KB is an implicate c such that for every implicate c' of KB, $c' \not\subset c$
    - Recall: treating clauses as sets of literals.
  - E.g. $\neg a \vee c$ and $\neg a \vee a$ are minimal implicates of $\neg a \vee b, \neg b \vee c$.

Prime Implicates

- An *implicate* of a theory is a clause that logically follows from that theory.
  - I.e. an implicate of KB is a clause c such that $KB \models c$.
  - E.g. $\neg a \vee b$ and $\neg b \vee c$ have implicates $\neg a \vee c \vee d$ and $\neg a \vee a \vee d$.
- A *minimal implicate* is a clause that has no strict subset as an implicate.
  - I.e. a minimal implicate of KB is an implicate c such that for every implicate c' of KB, $c' \not\subset c$
    - Recall: treating clauses as sets of literals.
  - E.g. $\neg a \vee c$ and $\neg a \vee a$ are minimal implicates of $\neg a \vee b, \neg b \vee c$.
- A *prime implicate* is a minimal implicate that is not trivial, i.e., does not contain complementary literals (of the form a, $\neg a$).
  - E.g. $\neg a \vee c$ is a prime implicate of $\neg a \vee b, \neg b \vee c$.
    The other prime implicates are $\neg a \vee b, \neg b \vee c$.

Bottom-Up Procedure

Motivation:

- We have the result that:

  A clause

  $$L_1 \vee \cdots \vee L_k$$

  is a logical consequence of a theory iff either

  - *there are some* L_i *and* L_j *such that* $L_i = \neg L_j$*, or*
  - *some subset of* $\{L_1, \ldots, L_k\}$ *is a prime implicate of the theory.*

Bottom-Up Procedure

Motivation:

- We have the result that:

  A clause

  $$L_1 \vee \cdots \vee L_k$$

  is a logical consequence of a theory iff either

  - *there are some* L_i *and* L_j *such that* $L_i = \neg L_j$, *or*
  - *some subset of* $\{L_1, \ldots, L_k\}$ *is a prime implicate of the theory.*

- So, for a query ϕ and a KB made up of prime implicates,
  1. Convert ϕ to clause form, and
  2. for each such clause, see if it is a superset of a prime implicate.

Bottom-Up Procedure

Motivation:

- We have the result that:

  A clause

  $$L_1 \vee \cdots \vee L_k$$

  is a logical consequence of a theory iff either

  - *there are some* L_i *and* L_j *such that* $L_i = \neg L_j$, *or*
  - *some subset of* $\{L_1, \ldots, L_k\}$ *is a prime implicate of the theory.*

- So, for a query ϕ and a KB made up of prime implicates,
  1. Convert ϕ to clause form, and
  2. for each such clause, see if it is a superset of a prime implicate.
- Thus if we can compute the prime implicates of a theory, we can perform deduction by table lookup, which is very fast.

Bottom-Up Procedure: Notes

- The main operation for computing prime implicates is *binary resolution*.
- Recall:
  Rule of resolution: from $R \vee L$ and $S \vee \neg L$ we can infer $R \vee S$.
  - Since we implicitely use sets, disjuncts of the form $A \vee A$ can be collapsed to A.

Bottom-Up Procedure: Notes

- The main operation for computing prime implicates is *binary resolution*.
- Recall:
  Rule of resolution: from $R \vee L$ and $S \vee \neg L$ we can infer $R \vee S$.
  - Since we implicitely use sets, disjuncts of the form $A \vee A$ can be collapsed to A.
- Earlier, we used resolution in a top-down procedure.

Bottom-Up Procedure: Notes

- The main operation for computing prime implicates is *binary resolution*.
- Recall:
  Rule of resolution: from $R \vee L$ and $S \vee \neg L$ we can infer $R \vee S$.
  - Since we implicitely use sets, disjuncts of the form $A \vee A$ can be collapsed to A.
- Earlier, we used resolution in a top-down procedure.
- Here, resolution is used as a bottom-up procedure to *compile out* all resolution steps.

Bottom-Up Procedure for Computing Prime Implicates

Input: Theory T in clause form

repeat

choose $\{L\} \cup R \in T$ and $\{\neg L\} \cup S \in T$ such that

$\nexists$ atom A such that $\{A, \neg A\} \subseteq R \cup S$ and

$\nexists C \in T$ such that $C \subseteq R \cup S$;

remove all C from T for which $R \cup S \subset C$;

$T := T \cup \{R \cup S\}$

until no more choices

Example

- Consider:

$$c \Rightarrow a \vee \neg b$$
$$\neg c \Rightarrow \neg e$$
$$b \vee d$$
$$d \Rightarrow a \vee b$$
$$\neg a \Rightarrow e$$

- In clause form:

$$\{\{a, \neg b, \neg c\}, \{c, \neg e\}, \{b, d\}, \{a, b, \neg d\}, \{a, e\}\}$$

Example (continued)

- These clauses have prime implicates
  $a, \quad b \vee d, \quad \neg e \vee c$

Example (continued)

- These clauses have prime implicates
  $a, \quad b \vee d, \quad \neg e \vee c$
- We can now quickly answer queries. For example:

  $?a \vee \neg b$ is

Example (continued)

- These clauses have prime implicates
  $a, \quad b \vee d, \quad \neg e \vee c$
- We can now quickly answer queries. For example:

$?a \vee \neg b$	is	*yes*
$?\neg e \vee \neg b$	is	

Example (continued)

- These clauses have prime implicates
  $a, \quad b \vee d, \quad \neg e \vee c$

- We can now quickly answer queries. For example:

$?a \vee \neg b$	is	*yes*
$?\neg e \vee \neg b$	is	*no*
$?b \vee \neg e \vee \neg b$	is	

Example (continued)

- These clauses have prime implicates
  $a, \quad b \vee d, \quad \neg e \vee c$

- We can now quickly answer queries. For example:

$?a \vee \neg b$	is	*yes*
$?\neg e \vee \neg b$	is	*no*
$?b \vee \neg e \vee \neg b$	is	*yes*

Inference Procedure using Prime Implicates

Input: Knowledge base KB expressed as prime implicates
Query Q, a formula of propositional logic.

$Q' := CNF(Q)$;
for each $C \in Q'$
    If $\nexists P \in KB$ s.t. $P \subseteq C$ return "fail"
return "yes"

More Notes

- This is an example of *knowledge compilation*, i.e. translating knowledge (usually offline) into a form for faster reasoning.

More Notes

- This is an example of *knowledge compilation*, i.e. translating knowledge (usually offline) into a form for faster reasoning.
- Computing prime implicates can be expensive, since resolution is exponential in the worse case.
  - I.e. determining whether a formula follows by resolution may take exponential time.
  - Also there may be an exponential number of prime implicates.

More Notes

- This is an example of *knowledge compilation*, i.e. translating knowledge (usually offline) into a form for faster reasoning.
- Computing prime implicates can be expensive, since resolution is exponential in the worse case.
  - I.e. determining whether a formula follows by resolution may take exponential time.
  - Also there may be an exponential number of prime implicates.
- Nonetheless prime implicates have played an important role in several areas of KR.

Applications

Prime implicates have found extensive use in KR.

- One major area is *abduction* or *diagnosis*.
  - This derives from earlier work on the *assumption-based truth-maintenance system (ATMS)*.
- As well, there has been work on using prime implicates in belief revision.

Application: Abduction

(See Chapter 13 of the Brachman and Levesque text)

So far, reasoning has been primarily deductive:

- Main question: Given KB, is α a logical consequence?

Application: Abduction

(See Chapter 13 of the Brachman and Levesque text)

So far, reasoning has been primarily deductive:

- Main question: Given *KB*, is α a logical consequence?

Now consider a new type of question:

Given:

- A KB, and a fact α,

Application: Abduction

(See Chapter 13 of the Brachman and Levesque text)

So far, reasoning has been primarily deductive:

- Main question: Given KB, is α a logical consequence?

Now consider a new type of question:

Given:

- A KB, and a fact α,

Ask:

- what would be a sufficient *reason* for α to be true?

Application: Abduction

(See Chapter 13 of the Brachman and Levesque text)

So far, reasoning has been primarily deductive:

- Main question: Given KB, is α a logical consequence?

Now consider a new type of question:

Given:

- A KB, and a fact α,

Ask:

- what would be a sufficient *reason* for α to be true?
- or, if I didn't believe α, what else would I have to believe for α to become an implicit belief?

Application: Abduction

(See Chapter 13 of the Brachman and Levesque text)

So far, reasoning has been primarily deductive:

- Main question: Given KB, is α a logical consequence?

Now consider a new type of question:

Given:

- A KB, and a fact α,

Ask:

- what would be a sufficient *reason* for α to be true?
- or, if I didn't believe α, what else would I have to believe for α to become an implicit belief?
- or, what would *explain* α being true?

Aside: Forms of Reasoning

Deduction: Given $p \Rightarrow q$, from p, deduce q

Aside: Forms of Reasoning

Deduction: Given $p \Rightarrow q$, from p, deduce q
Abduction: Given $p \Rightarrow q$, from q, abduce p

- I.e. p is *sufficient* for q, or
  one way for q to be true is for p to be true.
- Can be used for causal reasoning: ($cause \Rightarrow effect$)

Aside: Forms of Reasoning

Deduction: Given $p \Rightarrow q$, from p, deduce q

Abduction: Given $p \Rightarrow q$, from q, abduce p

- I.e. p is *sufficient* for q, or
  one way for q to be true is for p to be true.
- Can be used for causal reasoning: ($cause \Rightarrow effect$)

Induction: Given $p(t_1)$, $q(t_1)$, ..., $p(t_n)$, $q(t_n)$,
induce $\forall x(p(x) \Rightarrow q(x))$.

Using Abduction for Diagnosis

- One simple version of diagnosis uses abductive reasoning
- KB has facts about symptoms and diseases including:

$$Disease \wedge Hedges \Rightarrow Symptoms$$

- Goal: Find disease(s) that best explain observed symptoms

Abduction Example

Example:

$$TennisElbow \Rightarrow SoreElbow$$
$$TennisElbow \Rightarrow TennisPlayer$$
$$Arthritis \land \neg Treated \Rightarrow SoreJoints$$
$$SoreJoints \Rightarrow SoreElbow$$
$$SoreJoints \Rightarrow SoreHips$$

Abduction Example

Example:

$$TennisElbow \Rightarrow SoreElbow$$
$$TennisElbow \Rightarrow TennisPlayer$$
$$Arthritis \land \neg Treated \Rightarrow SoreJoints$$
$$SoreJoints \Rightarrow SoreElbow$$
$$SoreJoints \Rightarrow SoreHips$$

Explain: $SoreElbow$

Abduction Example

Example:

$$TennisElbow \Rightarrow SoreElbow$$
$$TennisElbow \Rightarrow TennisPlayer$$
$$Arthritis \wedge \neg Treated \Rightarrow SoreJoints$$
$$SoreJoints \Rightarrow SoreElbow$$
$$SoreJoints \Rightarrow SoreHips$$

Explain: $SoreElbow$

Want: $TennisElbow$, $Arthritis \wedge \neg Treated$

☞ Obtain multiple equally-good explanations

Criteria for an Explanation

Given KB, and β to be explained, we want an α such that:

Criteria for an Explanation

Given *KB*, and β to be explained, we want an α such that:

1. α is sufficient to account for β
   - $KB \cup \{\alpha\} \models \beta$ or $KB \models \alpha \Rightarrow \beta$

Criteria for an Explanation

Given *KB*, and β to be explained, we want an α such that:

1. α is sufficient to account for β
   - $KB \cup \{\alpha\} \models \beta$ or $KB \models \alpha \Rightarrow \beta$
2. α is not ruled out by *KB*
   - $KB \cup \{\alpha\}$ is consistent or $KB \not\models \neg\alpha$

Criteria for an Explanation

Given KB, and β to be explained, we want an α such that:

1. α is sufficient to account for β
   - $KB \cup \{\alpha\} \models \beta$ or $KB \models \alpha \Rightarrow \beta$
2. α is not ruled out by KB
   - $KB \cup \{\alpha\}$ is consistent or $KB \not\models \neg\alpha$
3. α is as simple as possible
   - Parsimonious: as few terms as possible
   - Explanations should not unnecessarily strong or weak

Criteria for an Explanation

Given *KB*, and β to be explained, we want an α such that:

1. α is sufficient to account for β
   - $KB \cup \{\alpha\} \models \beta$ or $KB \models \alpha \Rightarrow \beta$
2. α is not ruled out by *KB*
   - $KB \cup \{\alpha\}$ is consistent or $KB \not\models \neg\alpha$
3. α is as simple as possible
   - Parsimonious: as few terms as possible
   - Explanations should not unnecessarily strong or weak
4. α is in the appropriate vocabulary
   - Atomic sentences of α should be drawn from a set *H* of possible *hypotheses*.
   - E.g. diseases, original causes

Criteria for an Explanation

Given *KB*, and β to be explained, we want an α such that:

1. α is sufficient to account for β
   - $KB \cup \{\alpha\} \models \beta$ or $KB \models \alpha \Rightarrow \beta$
2. α is not ruled out by *KB*
   - $KB \cup \{\alpha\}$ is consistent or $KB \not\models \neg\alpha$
3. α is as simple as possible
   - Parsimonious: as few terms as possible
   - Explanations should not unnecessarily strong or weak
4. α is in the appropriate vocabulary
   - Atomic sentences of α should be drawn from a set H of possible *hypotheses*.
   - E.g. diseases, original causes

Call such α an *explanation* of β wrt *KB*.

Simplifying the Problem

We can simplify explanations in the propositional case, as follows:

Simplifying the Problem

We can simplify explanations in the propositional case, as follows:

- Assume that we are only going to explain an atom p, rather than an arbitrary formula.

Simplifying the Problem

We can simplify explanations in the propositional case, as follows:

- Assume that we are only going to explain an atom p, rather than an arbitrary formula.
- An explanation will be (equivalent to) a conjunction of literals (that is, the negation of a clause)
  - Why?
    - ☞ If α is a purported explanation, and
      $$DNF[\alpha] = (d_1 \vee d_2 \vee \cdots \vee d_n)$$
      then each d_i is also an explanation that is simpler than α

Simplifying the Problem

A simplest explanation is then the negation of a clause with a *minimal* set of literals

Simplifying the Problem

A simplest explanation is then the negation of a clause with a *minimal* set of literals

☞ To explain a literal l, it will be sufficient to find a minimal clause $C = \neg c_1 \vee \cdots \vee \neg c_n$ such that

1. $KB \not\models C$ (consistent)
2. $KB \cup \neg C \models l$ or
   $KB \models \neg C \Rightarrow l$ or
   $KB \models (C \cup \{l\})$ (sufficient)

Recall that the clause $C \cup \{l\}$ represents $\neg c_1 \vee \cdots \vee \neg c_n \vee l$.

Using Prime Implicates

Recall: Clause C is a prime implicate of KB iff

1. $KB \models C$
2. For no $C^* \subset C$ do we have $KB \models C^*$

Using Prime Implicates

Recall: Clause C is a prime implicate of KB iff

1. $KB \models C$
2. For no $C^* \subset C$ do we have $KB \models C^*$

For explanations:

- Want minimal C such that $KB \models (C \cup \{l\})$ and $KB \not\models C$

Using Prime Implicates

Recall: Clause C is a prime implicate of KB iff

1. $KB \models C$
2. For no $C^* \subset C$ do we have $KB \models C^*$

For explanations:

- Want minimal C such that $KB \models (C \cup \{l\})$ and $KB \not\models C$

Hence:

- Find prime implicates C such that $l \in C$.
- Then $\neg(C \setminus l)$ must be an explanation for l.

Example

$$KB = \{\ p \wedge q \wedge r \Rightarrow g,\ \neg p \wedge q \Rightarrow g,\ \neg q \wedge r \Rightarrow g\ \}$$

Example

$$KB = \{\ p \wedge q \wedge r \Rightarrow g,\ \neg p \wedge q \Rightarrow g,\ \neg q \wedge r \Rightarrow g\ \}$$

$$PI(KB) = \{p \vee \neg q \vee g,\ \neg r \vee g\} + \text{tautologies}$$

Example

$$KB = \{ \; p \wedge q \wedge r \Rightarrow g, \; \neg p \wedge q \Rightarrow g, \; \neg q \wedge r \Rightarrow g \; \}$$

$$PI(KB) = \{p \vee \neg q \vee g, \; \neg r \vee g\} + \text{tautologies}$$

Explanations for g:

- 2 non-trivial prime implicates contain g, so get 2 explanations: $\neg p \wedge q$ and r.