Learning with Abduction

Antonis Kakas
Department of Computer Science
University of Cyprus

Fabrizio Riguzzi
DEIS - University of Bologna

Overview of the Talk

- Preliminaries on Abduction
- Abduction in Logic Programming
- Integration of induction and abduction
- An example of an abductive learning problem
  - learning from incomplete background data
- Proposed algorithm
- Experiments
- Conclusions and future works

Preliminaries on Abduction

- Abductive problem: given an observation that can not be derived in the theory, make assumptions that explain it

  shoes_are_wet $\leftarrow$ *grass_is_wet*

  grass_is_wet $\leftarrow$ *rained_last_night*

  grass_is_wet $\leftarrow$ *sprinkler_was_on*

  – the observation *shoes_are_wet* can be explained by the two sets of assumptions {*rained_last_night*} or {*sprinkler_was_on*}
- If we have as well the fact *electrical_black_out* and the integrity constraint

  $\neg$ (*electrical_black_out* and *sprinkler_was_on*)

  then the only allowed abductive explanations is

  {*rained_last_night*}

Abductive Logic Programming

- Abductive Logic Program: $T=\langle P,A,I\rangle$ where
  - P: normal logic program, A: set of abducible predicates, I: integrity constraints
- Abductive Problem (formally):
  - Given an observation O and a theory $T=\langle P,A,I\rangle$
  - Find an *abductive explanation* Δ such that
    - $P \cup \Delta \models O$
    - $P \cup \Delta \models I$ (Δ is consistent with IC)
  - We say that T *abductively entails* O ($T \models_A O$)
- Abducibles $\Leftrightarrow$ Incompleteness in T
- Abductive Proof Procedure [Kakas, Mancarella 90].
- Example:

```
?- shoes_are_wet.
yes, Δ ={rained_last_night}.
```

Integration of Abduction and Induction

- Both background and target theories are abductive theories.
- Coverage through deductive entailment replaced by coverage through abductive entailment
- Motivation:
  - increased expressive power of the background and learned theory
    - learning theories for diagnosis, default reasoning, planning, natural language understanding, high level vision
  - learning in presence of incomplete knowledge
    - abduction used to make assumptions about missing facts in the background knowledge

Abductive Concept Learning

- Given

  an **abductive** logic program $T=\langle P,A,I\rangle$ (the background knowledge)

  a set E^+ of positive examples

  a set E^- of negative examples

- Find an **abductive** logic program $T'=\langle P',A,I'\rangle \in \mathcal{P}$ with $P' \subseteq P$ and $I' \subseteq I$ such that

  $T' \models_A E^+$

  $\forall\ e^- \in E^-,\ T' \not\models_A e^-$

[Dimopoulos and Kakas 96].

•Abductive entailment for coverage

Example: learning from incomplete data

- Background knowledge
  - P={*parent(john,mary), male(john),*
    parent(david,steve),
    parent(kathy,ellen), female(kathy)}
  - A={*male, female*}
  - I={ }
- Training sets
  - E^+={*father(john,mary), father(david,steve),* }
  - E^-={*father(kathy,ellen), father(john,steve)*}
- Solution:

  P'=P ∪ {*father(X,Y)* ← *parent(X,Y), male(X).*}

  I'=I ∪ {¬(*male(X), female(X)*).}

  Δ={*male(david),not_male(kathy)*}

Two Step Algorithm:

1) Learn rules (Intermediate-ACL, I-ACL): find a set of rules R and a set of abductive explanations Δ^+_i and Δ^-_{jk} such that
   – all positive examples have an abductive explanation from $\langle P \cup R, A, I \rangle$
   – all the abductive explanations for each negative example are non empty
2) Learn constraints: find a set of integrity constraints C such that
   – all the abductive explanations for positive examples are allowed by the constraints
   – none of the abductive explanations for negative examples is allowed

First step (I-ACL)

- Discriminant learning: top-down algorithm with beam search.
- Prolog proof procedure substituted by abductive proof procedure.
- Each derivation starts from the set of assumptions abduced in the derivations of previous positive examples (1st example has $\Delta = \emptyset$).
  - to avoid incompatible assumptions for different positive examples
- For each negative example, compute all possible sets of assumptions. None of them must be empty.

Heuristic Function

- Expected Classification Accuracy using m-estimate.
- Different importance to example covered with or without abduction.

$$H = \frac{n^+ + k^+ \times n_A^+ + m \times p}{n^+ + n^- + k^+ \times n_A^+ + k^- \times n_A^- + m}$$

n^+_A pos. ex. cov. with abduction

n^-_A neg. ex. not cov. with all non empty Δs

k^+ estimate of the probability that the assumption made for pos. ex. are true

k- estimate of the probability that the assumption made for neg. ex. are true

Second Step (ACL)

- Characteristic learning of ICs using ICL
- Problem definition for ICL:
  - Given
    - B, a definite clause background
    - P, a set of true interpretations (models) for the target theory.
    - N, a set of false interpretations for the target theory
  - Find a clausal theory C such that
    - $\forall\ p \in P$, $M(B \cup p)$ is a model for C, i.e.
    - $\forall\ n \in N$, $M(B \cup n)$ is a false interpretation of C, i.e.
- Input to ICL:
  - $B = P \cup R$
  - one positive interpretation equal to the union of all explanations for positive examples
  - one negative interpretation for each explanation for negative examples

Example Solution: Step 1

- Output

  R= $father(X,Y) \leftarrow parent(X,Y), male(X).$

  – Positive examples:

    $\leftarrow father(john,mary).$ $\Delta^+_1 = \emptyset$

    $\leftarrow father(david,steve).$ $\Delta^+_2 = \{male(david)\}$.

  – Negative examples:

    $\leftarrow father(kathy,ellen).$ $\Delta^-_1 = \{male(kathy)\}$

    $\leftarrow father(john,steve).$ fails (i.e. no set of assumptions)

Example Solution: Step 2

- Input to ICL
  - B ={*parent(john,mary), male(john),*
    parent(david,steve),
    parent(kathy,ellen), female(kathy),
    $father(X,Y) \leftarrow parent(X,Y), male(X).$}
  - P = {p}, p = $\Delta^+_1 \cup \Delta^+_2$ = {*male(david)*}.
  - N = {Δ^-_1} , Δ^-_1= {*male(kathy)*}.
- Output:

$\leftarrow male(X),female(X).$

Experiments: 6 Bit Multiplexer

- Examples are 6 bits tuples,
- Positive if the bit addressed by the first 2 is at 1: e.g. 11 0110
- 64 examples: 32 positive, 32 negative
- Representation of 11 0110:
  - in E^+: mul(e1)
  - in B:
    bit1at1(e1). bit2at0(e1). bit3at0(e1). bit4at1(e1). bit5at1(e1). bit6at0(e1).
- 3 experiments (I-ACL):
  - complete dataset
  - incomplete dataset (3 bits taken out randomly from 12 examples)
  - incomplete dataset plus constraints

6 Bit Multiplexer Results

- Accuracy: number of examples correctly classified over total number of examples.

Experiments	I-ACL	ICL-Sat
Complete Background	100%	100%
Incomplete Background	98.4%	82.8%
Incomplete Background plus constraints	96.9%	92.2%

- Loss of accuracy due to positive examples not covered: 1 for the second experiment, 2 for the third.
- Overspecific rules due to wrong assumptions for some covered examples.

Classifying Incomplete Examples

- Learned theory tested against the incomplete training set

Experiments	n^+	n^-_{ACL}	Accuracy
Incomplete data	32	5	92.2 %
Incomplete data plus constraints	32	2	96.9 %

n^-_{ACL}: negative examples e^- for which $\leftarrow e^-$ succeeds (there exist an abductive explanation)

Learning Integrity Constraints

- Some negative examples taken out from the training set.
- I-ACL now learns the incorrect rule
  $mul(X) \leftarrow bit1at1(X), bit2at1(X).$
- Run Claudien on background plus positive examples.
- Among the generated constraints:
  $\leftarrow mul(X), bit1at1(X), bit2at1(X), bit6at0(X).$
- that corrects the rule above.
- Integration of characteristic and discriminant induction.

Learning Family Relations

- Learning father from a database containing 740 facts about parent, male and female. 36 positive, 34 negative examples.
- Two series of experiments with increasing degree of incompleteness:
  - 100%, 90%,..., 50% of facts
  - with or without the constraint $\leftarrow male(X), female(X).$
- Results (I-ACL):
  - without constraint: correct result up to 80%
  - with constraint: correct result up to 60%

Conclusions

- ACL: Abduction for coverage relation
- Two step algorithm for learning abductive theories:
  - First step: discriminant learning (top-down algorithm)
  - Second step: characteristic learning (ICL)
- Applications
  - learning from incomplete data
  - learning theories for abductive reasoning
  - multiple predicate learning
  - learning logic programs with negation

Future work

- Experiments on real world data from a marketing company
- Full ACL on real world data
- Investigate the impact of the integration on
  - multiple predicate learning
  - learning logic programs with negation
- Interleave the generation of rules and constraints. Two levels:
  - generate constraints after the generation of each rule,
  - new refininement operator on the new space of rules and constraints