Partial Implications in Data Mining and Logic

Deduction, Optimum Axiomatizations, and Data Analysis

José L. Balcázar

LSI, UPC

Lisboa, April 2013

The Dual Vision of Logic

Syntax meets Semantics, or vice versa

Syntactic deduction, semantic entailment

Derive logical formulas... from other logical formulas.

- Syntactic view: formula manipulation rules.
- Semantic view: state formally what each formula "means" in a given "context" (model):
  - models assign meaning to the "atoms" of the formula, and then the "connectives" are applied;
  - leads to the notion of a formula being true or false in a model;
  - entailment refers to a consequent formula being surely true whenever antecedents are.

Beautiful properties of soundness and completeness: in many important cases, deduction corresponds exactly to entailment (for First-Order Logic: Gödel 1930).

Computing Entailments

Bad News

We run into computational difficulties

Entailment turns out to be:

- Algorithmically undecidable for reasonably expressive logics,
- Algorithmically infeasible for even quite modest ones,
- Still not really understood regarding its efficient neural implementations in living organisms.

What can we do?

There exists a relatively natural sublanguage that is (almost) about as rich as you can hope for, while being at the same time algorithmically feasible.

(Definite) Propositional Horn Formulas

Definiteness issues glossed over

Very simple case: a fragment of the propositional world

Boolean-valued variables.

- Models (binary strings): a Boolean value per variable; equivalently: the set of variables true in it.
- (Definite) Horn Clause: one single positive disjunct, like $\neg a \vee \neg b \vee c$.
- Equivalent form as implication, like $a \wedge b \Rightarrow c$.
- Horn Formula: conjunction of Horn Clauses.

(Definite) Propositional Horn Formulas

Definiteness issues glossed over

Very simple case: a fragment of the propositional world

Boolean-valued variables.

- Models (binary strings): a Boolean value per variable; equivalently: the set of variables true in it.
- (Definite) Horn Clause: one single positive disjunct, like $\neg a \vee \neg b \vee c$.
- Equivalent form as implication, like $a \wedge b \Rightarrow c$.
- Horn Formula: conjunction of Horn Clauses.
- Implications: $(a \wedge b \Rightarrow c) \wedge (a \wedge b \Rightarrow c) \equiv (a \wedge b \Rightarrow c \wedge d)$.

(Definite) Propositional Horn Formulas

Definiteness issues glossed over

Very simple case: a fragment of the propositional world

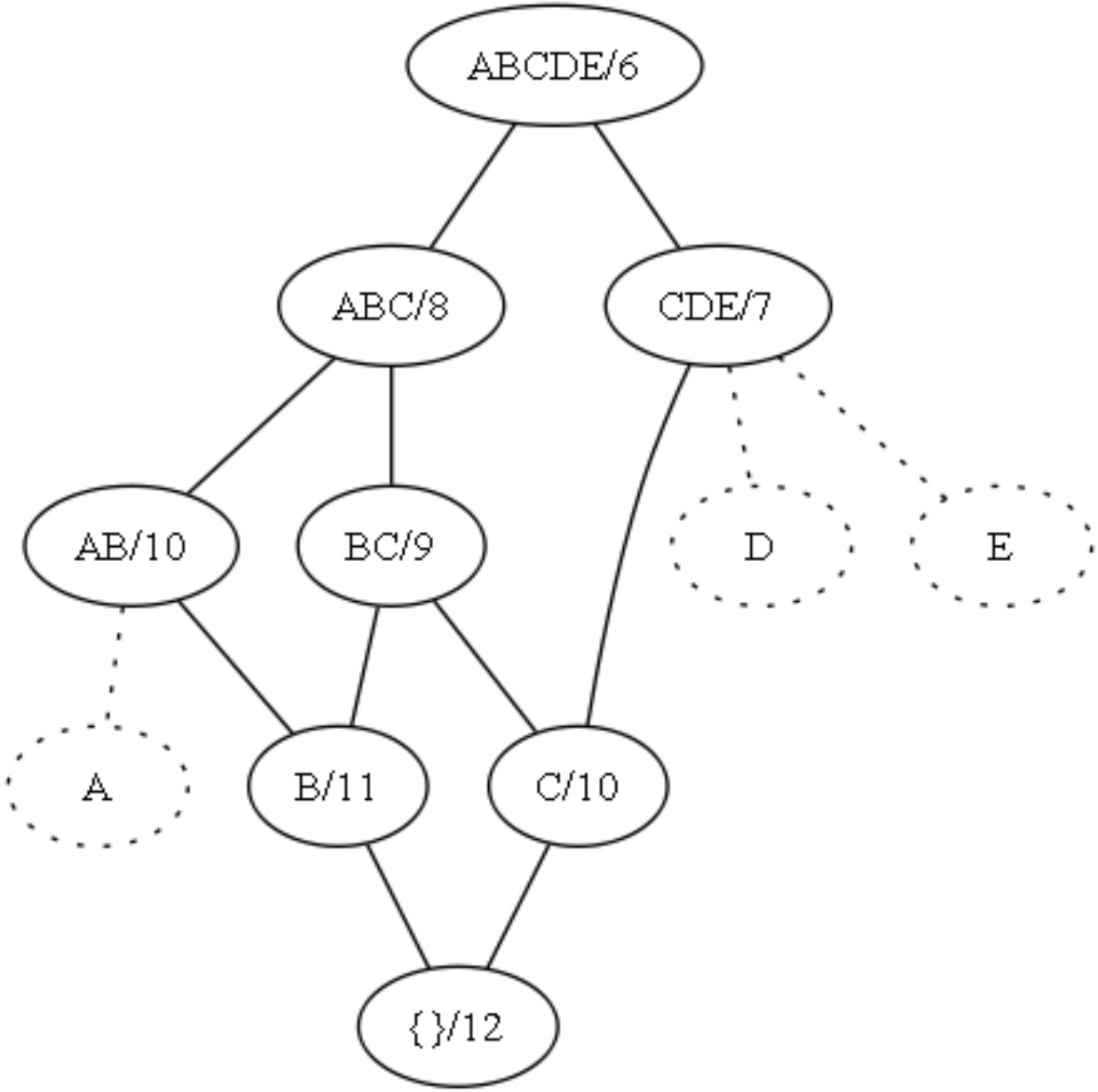
Boolean-valued variables.

- Models (binary strings): a Boolean value per variable; equivalently: the set of variables true in it.
- (Definite) Horn Clause: one single positive disjunct, like $\neg a \vee \neg b \vee c$.
- Equivalent form as implication, like $a \wedge b \Rightarrow c$.
- Horn Formula: conjunction of Horn Clauses.
- Implications: $(a \wedge b \Rightarrow c) \wedge (a \wedge b \Rightarrow c) \equiv (a \wedge b \Rightarrow c \wedge d)$.

Main Property

A set of models can be axiomatized by a Horn Formula if and only if it is closed under intersection: they form a closure space lattice.

A Closure Lattice



ABCDE (x6),
ABC (x2),
AB (x2),
CDE (x1),
BC (x1)

$A \Rightarrow B$,
$D \Rightarrow CE$,
$E \Rightarrow CD$,
$BD \Rightarrow A$,
$BE \Rightarrow A$

Implications, I

A real-life example

Logs from a virtual learning platform

Propositional variables:

- one for each "area" of the course.
  - `announcements`, `assessments`, `assignments`, `contents`, `forum`, `organizer`, ...

Implications, I

A real-life example

Logs from a virtual learning platform

Propositional variables:

one for each "area" of the course.

`announcements`, `assessments`, `assignments`, `contents`, `forum`, `organizer`, ...

- Student's sessions are logged;
- for each session, we know whether each "area" was visited in that session;

Implications, I

A real-life example

Logs from a virtual learning platform

Propositional variables:

one for each "area" of the course.

`announcements`, `assessments`, `assignments`, `contents`, `forum`, `organizer`, ...

- Student's sessions are logged;
- for each session, we know whether each "area" was visited in that session;
- therefore each session is a propositional model.

Implications, I

A real-life example

Logs from a virtual learning platform

Propositional variables:

one for each "area" of the course.

`announcements`, `assessments`, `assignments`, `contents`, `forum`, `organizer`, ...

- Student's sessions are logged;
- for each session, we know whether each "area" was visited in that session;
- therefore each session is a propositional model.

Example of an implication:

$\texttt{announcements} \wedge \texttt{assignments} \Rightarrow \texttt{assessments} \wedge \texttt{organizer}$

It is again the conjunction of two Horn clauses.

Implications, II

As a data analysis tool

Implications are a classic in the Data Mining field.

Examples from a "Machine Learning abstracts" dataset

descent $\Longrightarrow$ gradient

machines svms $\Longrightarrow$ support vector

Implications, II

As a data analysis tool

Implications are a classic in the Data Mining field.

Examples from a "Machine Learning abstracts" dataset

descent $\Longrightarrow$ gradient

machines svms $\Longrightarrow$ support vector

hilbert $\Longrightarrow$ space

Implications, II

As a data analysis tool

Implications are a classic in the Data Mining field.

Examples from a "Machine Learning abstracts" dataset

descent $\Longrightarrow$ gradient

machines svms $\Longrightarrow$ support vector

hilbert $\Longrightarrow$ space

carlo $\Longrightarrow$ monte

monte $\Longrightarrow$ carlo

Implications, II

As a data analysis tool

Implications are a classic in the Data Mining field.

Examples from a "Machine Learning abstracts" dataset

descent $\Longrightarrow$ gradient

machines svms $\Longrightarrow$ support vector

hilbert $\Longrightarrow$ space

carlo $\Longrightarrow$ monte

monte $\Longrightarrow$ carlo

Example from a "census" dataset

Exec-managerial Husband $\Longrightarrow$ Married-civ-spouse

Learning Via Queries

Identifying Horn Formulas

A major property of Horn clauses: polynomial time learnability.

Active Learning

We (the learning algorithm) "interact" with a Horn theory by querying whether a given model satisfies it and by "showing up for an exam" when we believe we know a correct set of axioms.

More precisely

Identify a target Horn formula through

- membership queries: whether a specific model, chosen by us, satisfies the target formula;
- equivalence queries: whether our current hypothesis is finally correct; otherwise, we are given a counterexample model.

AFP algorithm (Angluin, Frazier, Pitt 1992): constructs a Horn axiomatization of the target.

The Logic of Implications, I

A Deductive Calculus

Implications obey the Armstrong inference schemes, originally from functional dependency analysis in Databases (Armstrong 1974):

- Reflexivity: if $Y \subseteq X$, infer $X \implies Y$;
- Augmentation: from $X \implies X'$ and $Y \implies Y'$, infer $XY \implies X'Y'$;
- Transitivity: from $X \implies Y$ and $Y \implies Z$, infer $X \implies Z$.

Soundness and completeness

Using these schemes, one can infer from a set of implications exactly those implications that become logically entailed by them: any dataset in which the premises are satisfied must satisfy as well the conclusions.

The Logic of Implications, II

Optimal Axiomatizations

Given all the implications that hold for a set of models,

- some of them may be redundant (logically entailed);
- taking these out would give an irredundant basis;
- but there may be various ways to reach irredundant bases,
- and they may be of very different sizes.

The Logic of Implications, II

Optimal Axiomatizations

Given all the implications that hold for a set of models,

- some of them may be redundant (logically entailed);
- taking these out would give an irredundant basis;
- but there may be various ways to reach irredundant bases,
- and they may be of very different sizes.

Minimum-size axiomatization (Guigues, Duquenne 1986)

- a canonical, minimum-size basis for implications;
- equivalent notion in functional dependencies;
- the Horn Query Learning algorithm AFP constructs it.
- There is no such canonical, minimum-size basis for Horn clause syntax.

Towards Standard Association Rules

A relaxed notion of "correctness"

Many other logics exist

Alternative views of various notions, even the most basic ones:

- Infinitary connectives,
- truth "degrees" or "probabilities",
- unsound but useful deductive processes (like abduction)...
- Here we will consider standard conjunction, and our variables have crisp truth values, but the implication connective will be quantitative.

Towards Standard Association Rules

A relaxed notion of "correctness"

Many other logics exist

Alternative views of various notions, even the most basic ones:

- Infinitary connectives,
- truth "degrees" or "probabilities",
- unsound but useful deductive processes (like abduction)...
- Here we will consider standard conjunction, and our variables have crisp truth values, but the implication connective will be quantitative.

Consider some facts found in a "census" dataset:

- Husband $\Longrightarrow$ Male

Towards Standard Association Rules

A relaxed notion of "correctness"

Many other logics exist

Alternative views of various notions, even the most basic ones:

- Infinitary connectives,
- truth "degrees" or "probabilities",
- unsound but useful deductive processes (like abduction)...
- Here we will consider standard conjunction, and our variables have crisp truth values, but the implication connective will be quantitative.

Consider some facts found in a "census" dataset:

- Husband $\Longrightarrow$ Male... does not hold!

Towards Standard Association Rules

A relaxed notion of "correctness"

Many other logics exist

Alternative views of various notions, even the most basic ones:

- Infinitary connectives,
- truth "degrees" or "probabilities",
- unsound but useful deductive processes (like abduction)...
- Here we will consider standard conjunction, and our variables have crisp truth values, but the implication connective will be quantitative.

Consider some facts found in a "census" dataset:

- Husband $\Longrightarrow$ Male... does not hold!
- Similarly, Wife $\Longrightarrow$ Female does not hold either: there are two tuples declaring Male and Wife.

Intensity of Implication

The meaning of correctness

If implication does not hold universally, a natural measure is confidence:

$$\mathrm{conf}(X \to Y) = \frac{\text{support of } XY}{\text{support of } X}$$

that is, the frequentist approximation to the conditional probability of the consequent with respect to the antecedent.

- A lower confidence threshold gives more rules.
- Lower bounds on the joint support of XY are usually enforced as well.
- Educated people from other disciplines (like many candidates to become data mining users) would expect us to use it!

Intensity of Implication

The meaning of correctness

If implication does not hold universally, a natural measure is confidence:

$$\mathrm{conf}(X \to Y) = \frac{\text{support of } XY}{\text{support of } X}$$

that is, the frequentist approximation to the conditional probability of the consequent with respect to the antecedent.

- A lower confidence threshold gives more rules.
- Lower bounds on the joint support of XY are usually enforced as well.
- Educated people from other disciplines (like many candidates to become data mining users) would expect us to use it!
- But, unfortunately, it is easily misled by negative correlations.

The Danger Of Absolute Confidence Thresholds

But, how to convince everyone else?

Dataset CMC (Contraceptive Method Choice)

A rule of over 10% support and 90% confidence:

`near-low-wife-education` `no-contraception-method`

$\longrightarrow$

`good-media-exposure`

The Danger Of Absolute Confidence Thresholds

But, how to convince everyone else?

Dataset CMC (Contraceptive Method Choice)

A rule of over 10% support and 90% confidence:

`near-low-wife-education` `no-contraception-method`

$\longrightarrow$

`good-media-exposure`

But the support of "`good-media-exposure`" is over 92%.

The Danger Of Absolute Confidence Thresholds

But, how to convince everyone else?

Dataset CMC (Contraceptive Method Choice)

A rule of over 10% support and 90% confidence:

`near-low-wife-education` `no-contraception-method`

$\longrightarrow$

`good-media-exposure`

But the support of "`good-media-exposure`" is over 92%.

- The most natural normalization to avoid this problem (deviation from independence, also called lift) is symmetric.
- Many alternative definitions of $X \to Y$, almost all on the basis of the supports of X, Y, XY, and $X \cap Y$.
- Rich and complex landscape, leading to an "axiomatic" study of all these alternatives.

Redundancy in Association Rules, I

A Logic-based view

Standard Association Mining Process

User provides dataset and thresholds for support and confidence, and gets all rules that hold in the dataset at those levels or higher.

Huge set of rules, growing further for lower thresholds. How to offer the user a smallish set of output rules?

- Our (rather obvious) proposal of "plain" redundancy: $X \to Y$ is redundant with respect to $X' \to Y'$ if $\mathrm{conf}(X \to Y) \geq \mathrm{conf}(X' \to Y')$ in every dataset.
- A natural variant, closure-based redundancy, reads the same, but under a condition to share the same "closure space" lattice.

Redundancy in Association Rules, II

A Calculus

Schemes of Augmentation or of composition with an Implication, each applied at the ℓeft hand side or at the right hand side.

- (rA) from $X \to Y$ and $X \Rightarrow Z$ infer $X \to YZ$;
- (rI) from $X \to Y$ and $Y \Rightarrow Z$ infer $X \to YZ$;
- (ℓA) from $X \to YZ$ infer $XY \to Z$;
- (ℓI) if $Z \subseteq X$, from $X \to Y$ and $Z \Rightarrow X$ infer $Z \to Y$.

Soundness and completeness

Using these rules, one can infer from a partial rule $X \to Y$, plus a set of implications, exactly those implications that are redundant with them.

Redundancy in Association Rules, IV

The natural notion for the logician

We should consider several premise rules for redundancy

Something interesting happens!

- Quite clear intuition: from two partial rules of confidence $\gamma < 1$, any combination will lead to rules of confidence less than γ;

Redundancy in Association Rules, IV

The natural notion for the logician

We should consider several premise rules for redundancy

Something interesting happens!

- Quite clear intuition: from two partial rules of confidence $\gamma < 1$, any combination will lead to rules of confidence less than γ; this intuition is wrong.

Redundancy in Association Rules, IV

The natural notion for the logician

We should consider several premise rules for redundancy

Something interesting happens!

- Quite clear intuition: from two partial rules of confidence $\gamma < 1$, any combination will lead to rules of confidence less than γ; this intuition is wrong.
- $A \to BC,\ A \to BD \models ACD \to B$

Redundancy in Association Rules, IV

The natural notion for the logician

We should consider several premise rules for redundancy

Something interesting happens!

- Quite clear intuition: from two partial rules of confidence $\gamma < 1$, any combination will lead to rules of confidence less than γ; this intuition is wrong.
- $A \to BC,\ A \to BD \models ACD \to B$
- Redundancy, formalized as logical entailment, is only well-understood when just one of the premises is a (partial) association rule, and the others are all implications ("rules" of confidence 1).
- Slight understanding with two partial rules: complex characterization, slow algorithms of little productivity.

Redundancy in Association Rules, V

Minimum-Size Bases

Basic antecedent X of Y (with $X \subseteq Y$):

- work only among closures: both X and Y must be closed;
- "representative rules" variant: X not necessarily closed;
- confidence of $X \to Y$ must be at least γ;
- but falls below γ if either we enlarge Y, or we reduce X.

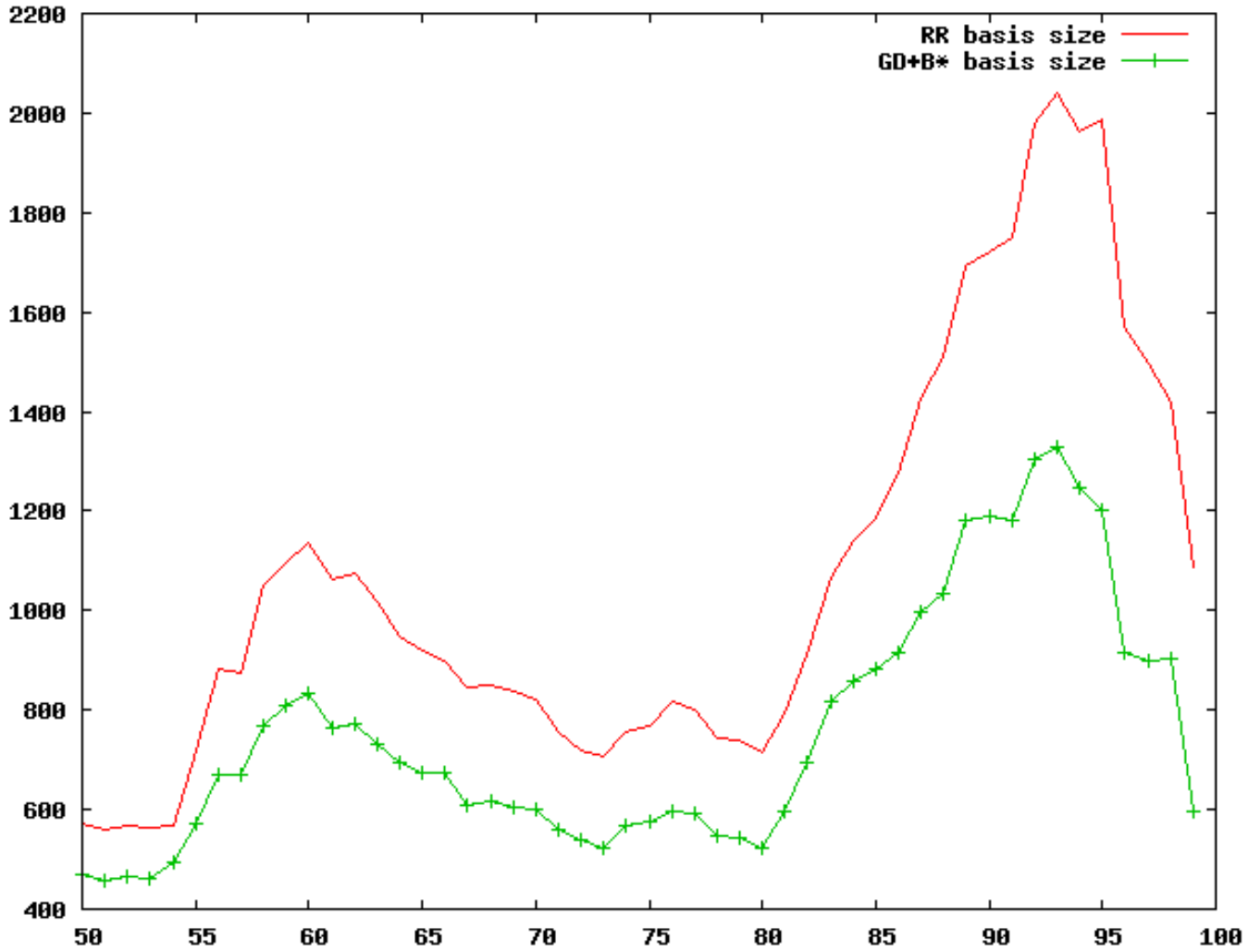
Basis $\mathcal{B}^*$: $X \to Y - X$ for all closed Y and all basic antecedents X of Y, provided $Y - X \neq \emptyset$.

Facts

1. These rules hold with confidence γ,
2. All the rules that hold with confidence γ can be inferred from these rules plus the implications, and
3. Any alternative set of rules with the same properties has at least as many rules as this one.

Irredundant Rules for Dataset FIMI `pumsb-star`

In a couple of alternative formulations



Inspires a notion of "novelty".

Further Case Studies

The logical notion of redundancy is still far from intuition

Often, one finds large amounts of similar implications.

Back to the "census" dataset

Let's have a look again at the odd tuple:

- Counterexample to Husband $\Longrightarrow$ Male.
- Consequence: over sixty full-confidence implications of the form Husband,SomethingElse $\Longrightarrow$ Male.
- Are they "redundant"?

Either of $A \to C$ and $AB \to C$ can have arbitrarily higher confidence than the other. But discarding $AB \to C$ in such cases "works".

Confidence boost:

Bound the relative confidence instead of its absolute value: a partial implication is novel if it has substantially more confidence than "similar" rules.

Conclusions

Progress so far on the logic of partial implications

Propositional Horn logic

A key notion to understand certain sorts of data in specific ways.

- Association Rules are an interesting variation of Horn clauses.
- We begin to understand the Logic of Association Rules:
  - A solid preliminary notion of redundancy as entailment;
  - A deductive calculus sound and complete for it;
  - A way of selecting bases of optimum size.

Conclusions

Progress so far on the logic of partial implications

Propositional Horn logic

A key notion to understand certain sorts of data in specific ways.

- Association Rules are an interesting variation of Horn clauses.
- We begin to understand the Logic of Association Rules:
  - A solid preliminary notion of redundancy as entailment;
  - A deductive calculus sound and complete for it;
  - A way of selecting bases of optimum size.
- A family of proposals for formalizing "novelty":
  - The notion that "works" intuitively speaking... does not really correspond to the "logician's" view of redundancy; but logic-inspired notions do not really capture human intuition.
- `sourceforge.net/projects/yacaree`

Conclusions

Progress so far on the logic of partial implications

Propositional Horn logic

A key notion to understand certain sorts of data in specific ways.

- Association Rules are an interesting variation of Horn clauses.
- We begin to understand the Logic of Association Rules:
  - A solid preliminary notion of redundancy as entailment;
  - A deductive calculus sound and complete for it;
  - A way of selecting bases of optimum size.
- A family of proposals for formalizing "novelty":
  - The notion that "works" intuitively speaking... does not really correspond to the "logician's" view of redundancy; but logic-inspired notions do not really capture human intuition.
- `sourceforge.net/projects/yacaree`
- Full entailment with several partial rules as premises: rough road ahead as the case of two premises is already near the current limit of human understanding.