Bases and Objective Novelty Measures for Plain Association Rules

Confidence Width, Blocking, and Confidence Boost: Theory and Applications

José L. Balcázar, Cristina Tîrnăucă,
Marta E. Zorrilla, Diego García

Departamento de Matemáticas, Estadística y Computación
Universidad de Cantabria

WorSe, LSI, UC, May 2010

Three Desired Properties

For useful data mining processes

Intuitive aims of a Data Mining conceptual tool

Several standard references propose the following desiderata:

- Actionability
- Correctness
- Novelty

This series of talks describe our UC-LSI path to them, starting with Logic, and reaching to Educational Data Mining.

The Logic of Implications, I

2350 years of occidental culture

Formal Logic has provided us with a beautiful study of human inference processes.

(Aristotle's great advance over Plato on dialectics!)

The Logic of Implications, I

2350 years of occidental culture

Formal Logic has provided us with a beautiful study of human inference processes.

(Aristotle's great advance over Plato on dialectics!)

Among many others:

- Deduction: from $p \to q$, if we observe p, we infer q.
  (Beautiful properties of *soundness* and *completeness*.)

The Logic of Implications, I

2350 years of occidental culture

Formal Logic has provided us with a beautiful study of human inference processes.

(Aristotle's great advance over Plato on dialectics!)

Among many others:

- Deduction: from $p \rightarrow q$, if we observe p, we infer q.
  (Beautiful properties of *soundness* and *completeness*.)
- Abduction: from $p \rightarrow q$, if we observe q, we infer p.
  (Unsound, but models frequent human inference steps.
  Statistics may help to quantifying some "level of soundness".)

The Logic of Implications, I

2350 years of occidental culture

Formal Logic has provided us with a beautiful study of human inference processes.

(Aristotle's great advance over Plato on dialectics!)

Among many others:

- Deduction: from $p \rightarrow q$, if we observe p, we infer q.
  (Beautiful properties of *soundness* and *completeness*.)
- Abduction: from $p \rightarrow q$, if we observe q, we infer p.
  (Unsound, but models frequent human inference steps.
  Statistics may help to quantifying some "level of soundness".)
- But, how are we to "observe $p \rightarrow q$"?

The Logic of Implications, I

2350 years of occidental culture

Formal Logic has provided us with a beautiful study of human inference processes.

(Aristotle's great advance over Plato on dialectics!)

Among many others:

- Deduction: from $p \rightarrow q$, if we observe p, we infer q.
  (Beautiful properties of *soundness* and *completeness*.)
- Abduction: from $p \rightarrow q$, if we observe q, we infer p.
  (Unsound, but models frequent human inference steps.
  Statistics may help to quantifying some "level of soundness".)
- But, how are we to "observe $p \rightarrow q$"?
- Even... what do we "mean" by $p \rightarrow q$?

(Definite) Horn Formulas, I

Definiteness issues glossedover

A world of propositional variables: Boolean-valued.

- Models (binary strings): a Boolean value per variable
- (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) Horn Formulas, I

Definiteness issues glossedover

A world of propositional variables: Boolean-valued.

- Models (binary strings): a Boolean value per variable
- (Definite) Horn Clause: one single positive disjunct, like $\neg a \lor \neg b \lor c$.
- Equivalent form as implication, like $a \land b \Rightarrow c$.
- Horn Formula: conjunction of Horn Clauses.
- Implications: $(a \land b \Rightarrow c) \land (a \land b \Rightarrow c) \equiv (a \land b \Rightarrow c \land d)$.

(Definite) Horn Formulas, I

Definiteness issues glossedover

A world of propositional variables: Boolean-valued.

- Models (binary strings): a Boolean value per variable
- (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 bitwise conjunction.

(Note the Syntax/Semantics two-sided view.)

(Definite) Horn Formulas, II

A real-life example

Logs from virtual learning platform:

Propositional variables:

one for each "area" of the course.

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

(Definite) Horn Formulas, II

A real-life example

Logs from 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;

(Definite) Horn Formulas, II

A real-life example

Logs from 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.

(Definite) Horn Formulas, II

A real-life example

Logs from 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:

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

It is again the conjunction of two Horn clauses.

The Logic of Implications, II

A deductive calculus

The Armstrong inference schemes, originally from functional dependency analysis in Databases:

- 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 $X \implies Z$.

Key property: using these rules, 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.

Logical Terminology:

Soundness and consistency properties of deductive calculi.

The Logic of Implications, III
Formal Concept Analysis

Properties of the implications:

- competing algorithms exist to extract them from data;
- they characterize the smallest Horn theory that contains the data;
- they will provide accurate predictions if employed in a context where the source of data is well-approximated by a Horn theory;
- they look great! their syntax suggests the existence of an actionable causality relation!
- sets closed under the implications are exactly those that have as many attributes as possible without loss of supporting data transactions (a closure space).

The Logic of Implications, IV
Optimal axiomatizations

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

- some of them may be redundant, due to the Armstrong rules;
- 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 axiomatizations: the Guigues-Duquenne basis

- a canonical, minimum-size basis for implications;
- no information loss;
- there is no such canonical, minimum-size basis for mere Horn clauses.

The Logic of Implications, V

How useful they actually are?

Some examples:

From a “machine learning abstracts” dataset (and more).

- descent $\Longrightarrow$ gradient
- hilbert $\Longrightarrow$ space
- carlo $\Longrightarrow$ monte
- monte $\Longrightarrow$ carlo
- margin support $\Longrightarrow$ vector

The Logic of Implications, V

How useful they actually are?

Some examples:

From a “machine learning abstracts” dataset (and more).

- descent $\implies$ gradient
- hilbert $\implies$ space
- carlo $\implies$ monte
- monte $\implies$ carlo
- margin support $\implies$ vector
- In a “census” dataset: Husband $\implies$ Male. . .

The Logic of Implications, V

How useful they actually are?

Some examples:

From a "machine learning abstracts" dataset (and more).

- descent $\Longrightarrow$ gradient
- hilbert $\Longrightarrow$ space
- carlo $\Longrightarrow$ monte
- monte $\Longrightarrow$ carlo
- margin support $\Longrightarrow$ vector
- In a "census" dataset: Husband $\Longrightarrow$ Male. . . does not hold! (see tuple 7110)

The Logic of Implications, V

How useful they actually are?

Some examples:

From a "machine learning abstracts" dataset (and more).

- descent $\implies$ gradient
- hilbert $\implies$ space
- carlo $\implies$ monte
- monte $\implies$ carlo
- margin support $\implies$ vector
- In a "census" dataset: Husband $\implies$ Male… does not hold! (see tuple 7110)
- Similarly, Wife $\implies$ Female does not hold either: there are two tuples declaring Male and Wife.

The Logic of Implications, V

How useful they actually are?

Some examples:

From a “machine learning abstracts” dataset (and more).

- descent $\implies$ gradient
- hilbert $\implies$ space
- carlo $\implies$ monte
- monte $\implies$ carlo
- margin support $\implies$ vector
- In a “census” dataset: Husband $\Longrightarrow$ Male... does not hold! (see tuple 7110)
- Similarly, Wife $\implies$ Female does not hold either: there are two tuples declaring Male and Wife.
- Black $\Longrightarrow$ makes less than 50000\$/year

  (does not hold, but... “in many cases” it holds...)

From Implications to Associations
Motivation

There are reasons to be satisfied with an implication even in the presence of counterexamples.

- Transmission or keying errors;
- mistakes in filling up forms;
- mixed populations;
- ...

Partial "approximate" implications that allow for exceptions: we will need some implication intensity criterion.

Standard Association Rules, I

A lovely syntax for the policy makers

Working assumptions today

1. Stay not too far from propositional logic.

Standard Association Rules, I

A lovely syntax for the policy makers

Working assumptions today

1. Stay not too far from propositional logic.
2. Focus on entailment and axiomatizations (or: bases).

Standard Association Rules, I

A lovely syntax for the policy makers

Working assumptions today

1. Stay not too far from propositional logic.
2. Focus on entailment and axiomatizations (or: bases).
3. Keep deduction calculi as research target.

Standard Association Rules, I

A lovely syntax for the policy makers

Working assumptions today

1. Stay not too far from propositional logic.
2. Focus on entailment and axiomatizations (or: bases).
3. Keep deduction calculi as research target.
4. We accept an implication $p \rightarrow q$ if we "see it happen".

Standard Association Rules, I

A lovely syntax for the policy makers

Working assumptions today

1. Stay not too far from propositional logic.
2. Focus on entailment and axiomatizations (or: bases).
3. Keep deduction calculi as research target.
4. We accept an implication $p \to q$ if we "see it happen".
5. Key question: must "see it happen always"? or does it suffice to "see it happen often"?

Standard Association Rules, I

A lovely syntax for the policy makers

Working assumptions today

1. Stay not too far from propositional logic.
2. Focus on entailment and axiomatizations (or: bases).
3. Keep deduction calculi as research target.
4. We accept an implication $p \rightarrow q$ if we "see it happen".
5. Key question: must "see it happen always"? or does it suffice to "see it happen often"?
6. Two major questions behind:
   - How to settle on "how often".
   - How do we measure "how often".

Standard Association Rules, II

The meaning of actionability

The very syntax in the form of implications provides a causal relationship, making them great from the point of view of actionability.

Hopes of profit from actions may require that the association rules apply to a large enough set of “users” or “clients”.

Standard Association Rules, II

The meaning of actionability

The very syntax in the form of implications provides a causal relationship, making them great from the point of view of actionability.

Hopes of profit from actions may require that the association rules apply to a large enough set of "users" or "clients".

Politically incorrect variants:

The very syntax in the form of implications suggests an illusion of a causal relationship, making them look great from the point of view of actionability.

Exponentially many subsets: we impose a support constraint to make life a bit better.

Intensity of Implication, I

A more precise definition of association rules

For implications (or "deterministic association rules"), all reasonable criteria of "intensity" of the implication are equivalent.

Intensity of Implication, I
A more precise definition of association rules

For implications (or "deterministic association rules"), all reasonable criteria of "intensity" of the implication are equivalent.

Are these association rules equivalent?

high-wife-education → high-husband-education
high-wife-education → high-wife-education high-husband-education

Not all the measures of degree of implication proposed so far would rate them equal!

We only really define association rules when we have selected a measure of "degree of implication".

Intensity of Implication, II

The meaning of correctness

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

$$\text{conf}(A \to C) = \frac{\text{support of } AC}{\text{support of } A}$$

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

- The notion that we will use as definition of association rule.
- A lower confidence threshold gives more rules.
- 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 $X \rightarrow Y$, almost all on the basis of the supports of X, Y, XY, and $X \cap Y$.
- Complex landscape, leading to an “axiomatic” study of all these alternatives.

Intensity of Implication, III

Alternatives to confidence

Dozens of criteria for intensity of implication.

- Lift (interest, strength, deviation from independence),
- Conviction,
- Gini index,
- Prevalence,
- Leverage,
- Jaccard,
- Relative risk,
- J-Measure,
- Influence,
- ...

Intensity of Implication, III

Alternatives to confidence

Dozens of criteria for intensity of implication.

- Lift (interest, strength, deviation from independence),
- Conviction,
- Gini index,
- Prevalence,
- Leverage,
- Jaccard,
- Relative risk,
- J-Measure,
- Influence,
- ...

I believe now that none of them will ever beat confidence in user acceptation.

Novelty

An inherently relative notion?

There is no absolute notion of novelty: a piece of information is novel relative to “something else already known”.

Novelty

An inherently relative notion?

There is no absolute notion of novelty: a piece of information is novel relative to "something else already known".

Or is it?

Novelty

An inherently relative notion?

There is no absolute notion of novelty: a piece of information is novel relative to “something else already known”.

Or is it?

Our proposal:

There are absolute notions of novelty: if we mine a set of rules, is each rule “novel” with respect to the others?

Novelty

An inherently relative notion?

There is no absolute notion of novelty: a piece of information is novel relative to "something else already known".

Or is it?

Our proposal:

There are absolute notions of novelty: if we mine a set of rules, is each rule "novel" with respect to the others?

- "Logical" notion of redundancy, $X \to Y \models X' \to Y'$: in every dataset where $X \to Y$ holds, $X' \to Y'$ also does.
- Irredundant rules can be identified.
- From irredundancy, we can fine-tune the intuition into a first notion of absolute novelty (confidence width).
- Yet, this notion is not robust and general enough, and we will propose a better alternative (confidence boost).

Redundancy in Association Rules, I

A Logic-based view

We may wish to present the user a smallish set of mined rules.

- For a fixed γ, user-chosen, we focus on statements of the form $\mathrm{conf}(X \rightarrow Y) \geq \gamma$.
- Our (rather obvious) proposal of *plain* redundancy: $X \rightarrow Y$ is redundant with respect to $X' \rightarrow Y'$ if $\mathrm{conf}(X \rightarrow Y) \geq \mathrm{conf}(X' \rightarrow Y')$ in every dataset.
- Natural generalization: redundancy with a set of "premise" rules instead of a single one.
- Implications "sneak in" anyway, hence we are better off treating them explicitly since they allow better summarization (GD basis).

Redundancy in Association Rules, II

The natural notion for the logician

We should consider several premise rules for redundancy.

However: reality is stubborn, and imposes itself.

- Only well-understood when just one of the premises is a (partial) association rule, and the others are all implications ("rules" of confidence 1).
- Equivalently: same rendering as *plain* redundancy, but under a condition that datasets share the same closure space.

Redundancy in Association Rules, II

The natural notion for the logician

We should consider several premise rules for redundancy.

However: reality is stubborn, and imposes itself.

- Only well-understood when just one of the premises is a (partial) association rule, and the others are all implications ("rules" of confidence 1).
- Equivalently: same rendering as *plain* redundancy, but under a condition that datasets share the same closure space.
- Slight understanding with two partial rules: complex characterization, slow algorithms of little productivity.
- State of the art unable to work out beyond two partial rules.

Redundancy in Association Rules, III

About the Armstrong Calculus

A confidence threshold γ is assumed to be in place.

(Here $X \rightarrow Y$ means that the rule reaches the confidence threshold, whereas $X \Rightarrow Y$ denotes standard implication.)

Among the Armstrong deduction schemes, only Reflexivity is still valid. It gives that $X \rightarrow Y$ is equivalent to $X \rightarrow XY$.

Re Augmentation:

From $A \rightarrow B$ and $A \rightarrow C$ it does not follow $A \rightarrow BC$;
and from $A \rightarrow C$ it does not follow $AB \rightarrow BC$.

Re Transitivity:

From $A \rightarrow B$ and $B \rightarrow C$ it does not follow $A \rightarrow C$;
and from $A \rightarrow AB$ and $AB \rightarrow C$ it does not follow $A \rightarrow C$ either.

Redundancy in Association Rules, IV
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 $Z \to Y$.

Simpler variant when implications are not considered separately.

Key property:

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

Logical Terminology:

Soundness-like and consistency-like properties of deductive calculi.

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.

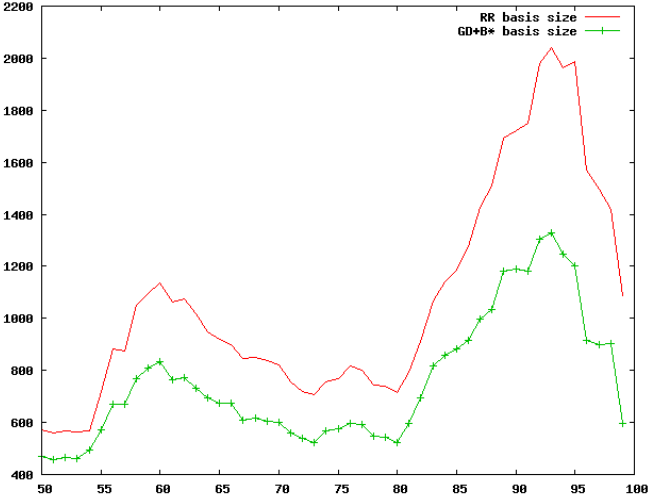
Basis $\mathcal{B}^*$: $X \to Y - X$ for all closed Y and all basic antecedents X of Y, provided $Y - X \neq \emptyset$.

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



Confidence Width

Quantifying absolute novelty

How robust is irredundancy?

Assume we have run an association miner for a given support and confidence thresholds (say γ for confidence).

- First, we discard all redundant rules, and just the basis is left.
- Each rule discarded is entailed by the basis.
- Each rule kept, say R, is not entailed by the others.
- This means that the other rules would not suggest that R passes the confidence threshold γ.

Confidence Width

Quantifying absolute novelty

How robust is irredundancy?

Assume we have run an association miner for a given support and confidence thresholds (say γ for confidence).

- First, we discard all redundant rules, and just the basis is left.
- Each rule discarded is entailed by the basis.
- Each rule kept, say R, is not entailed by the others.
- This means that the other rules would not suggest that R passes the confidence threshold γ.
- But R becomes redundant at a lower confidence, say γ'.

Confidence Width

Quantifying absolute novelty

How robust is irredundancy?

Assume we have run an association miner for a given support and confidence thresholds (say γ for confidence).

- First, we discard all redundant rules, and just the basis is left.
- Each rule discarded is entailed by the basis.
- Each rule kept, say R, is not entailed by the others.
- This means that the other rules would not suggest that R passes the confidence threshold γ.
- But R becomes redundant at a lower confidence, say γ'.
- The confidence width of R "is" the quotient γ/γ'.

Low Confidence Width

Novel, but barely

Suppose:

- Confidence of R is γ.
- Other rules of confidence γ do not entail it.
- Thus, it is a representative rule, irredundant, and novel with respect to the rest of the rules mined at confidence γ.
- But, if we had mined at a confidence slightly lower, say $\gamma' < \gamma$, maybe some R' would have been found that entails R.

R only belongs to the representative basis during the short interval of confidences $(\gamma', \gamma]$.

At its own confidence, it is novel, but really not too much.

High Confidence Width

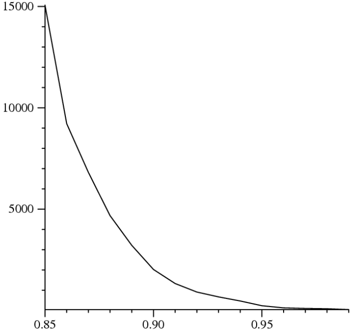
Novel, and robustly so

Suppose:

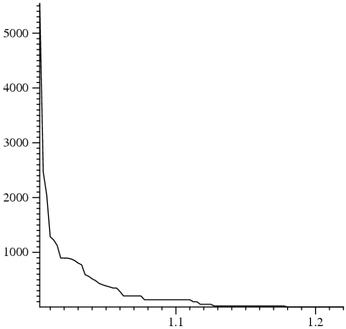
- Confidence of R is γ.
- Other rules of confidence γ do not entail it.
- Thus, it is a representative rule, irredundant, and novel with respect to the rest of the rules mined at confidence γ.
- We mine at lower confidence thresholds and R is still in the representative basis.
- We keep going lower and lower, and still R is in the basis.

Any other rule R' that we could consider to entail R has confidence $\gamma' << \gamma$: the confidence suggested for R by the rest of the rules in the dataset is γ', but R turned out to have much larger confidence than suggested.
R is actually novel.

Results for `FIMI Chess`
Promising...



chess rule confidences, s:80%, c:0.85



chess rule widths, s:80%, c:0.85

Further Case Studies

The idea looks nice, but...

Often, one finds large amounts of similar implications.

Let's have a look again at the Adult dataset.

Recall the odd tuple 7110.

- Counterexample to Husband $\Longrightarrow$ Male.
- Consequence: over sixty full-confidence rules of the form Husband,SomethingElse $\Longrightarrow$ Male.
- Problematic because these implications actually are the closure operator! Shall we stop trusting closures?
- Need yet another threshold for discarding this sort of uninteresting rules.

More Generally

This happens in many other cases.

We observe often this particularly peculiar case:

One finds many rules of the form $XA \to B$ for fixed A, B and many different X.

It can be due to rule $A \to B$ which would be the one that is really informative. But it could even disappear due to confidence just below the threshold.

Confidence width does not seem strong enough to help here: to make $XA \to B$ really redundant we need $A \to XB$; $A \to B$ does not suffice.

Blocking

Block one rule with another

$Z \subset X$ blocks $X \implies Y$ when the confidence of $X \implies Y$ is not larger enough than that of $Z \implies Y$.

For instance, Husband$\rightarrow$Male, of confidence very close to 1, would block Husband,SomethingElse $\implies$ Male of confidence 1.

But note that either could have higher confidence: we are comparing

- the density of Y's among the support of X with
- the density of Y's among the support of Z, which is larger.

Blocking Factor

Level at which we block one rule with another

Blocking factor: measures the extent to which the distribution of Y among the X is really denser than among the Z's, proper subsets of X.

Similar to width but different, neither bounds the other.

Good tricks for computing width fast; no such thing (as of now) for blocking.

Empirical results comparing confidence width and (an approximation to) the blocking factor suggest that best is to bound both.

Confidence Boost

Relates plain redundancy and representative rules

Confidence Boost measures "both at once", simultaneously adapting the related thresholds.

Properties:

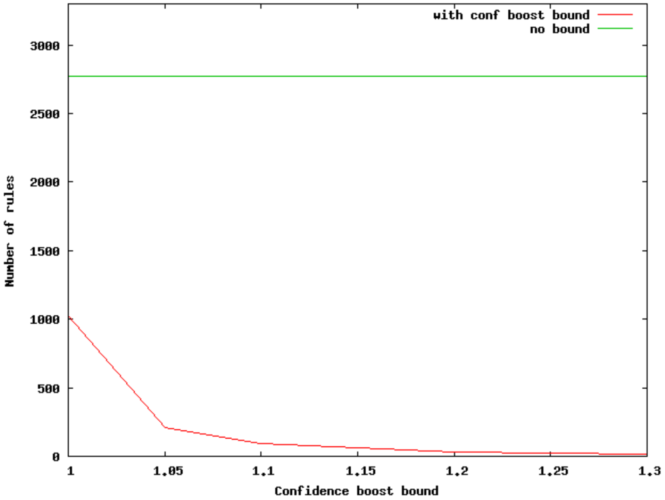
A formally tiny change in one of the formulations of the confidence width leads to our new notion:

- Low confidence boost if and only if low confidence width or rule blocked.
- A negative correlation like those that "fool confidence" will result in a low confidence boost (below 1).
- In particular, low lift implies low confidence boost.
- It can be computed more efficiently than the best algorithm we knew for blocking (but not as fast as width).
  - Involves a search on subsets of antecedents that risks exploring an exponential search space.

Confidence Boost in Action

An example

Irredundant association rules from the Adult dataset at 2% support and 75% confidence, with or without a confidence boost bound.



Closure-Based Confidence Boost

For closure-based redundancy and the $\mathcal{B}^*$ basis

Towards additional advantages

Confidence boost is good for representative rules, but needs adjustment for the $\mathcal{B}^*$ basis.

- Appropriate definition,
- formal properties,
- tuning of the algorithmics.

Excellent results, except that we lost the implications along the way.

How about considering negations?

> In some cases, only the presence of items is not really sufficiently informative.

Comparing Objective Novelty Measures

Making up for the shortcomings of confidence via novelty

A basic, objective novelty of an association rule can be measured by comparing its confidence with the confidence of related rules.

- Confidence width is a logical connection, formally excellent, lacks sufficient "statistical robustness" to be really useful in practice.

Comparing Objective Novelty Measures

Making up for the shortcomings of confidence via novelty

A basic, objective novelty of an association rule can be measured by comparing its confidence with the confidence of related rules.

- Confidence width is a logical connection, formally excellent, lacks sufficient “statistical robustness” to be really useful in practice.
- Rule blocking is relatively similar, has better statistical behavior, but is computationally very expensive, and fails to capture low width, so we still need both.

Comparing Objective Novelty Measures

Making up for the shortcomings of confidence via novelty

A basic, objective novelty of an association rule can be measured by comparing its confidence with the confidence of related rules.

- Confidence width is a logical connection, formally excellent, lacks sufficient "statistical robustness" to be really useful in practice.
- Rule blocking is relatively similar, has better statistical behavior, but is computationally very expensive, and fails to capture low width, so we still need both.
- Confidence boost is equivalent to bounding both at once, and implies bounding the lift and avoiding the negative correlations that make confidence too fallible:
  - Faster computation than blocking,
  - Slower computation than width,
  - Still affordable running times for many cases.

Perspectives

Looks too good to be true?. . . maybe. . .

Ongoing work:

If an absolute confidence threshold is treacherous, maybe we could make do with a relative confidence threshold.

- Further empirical studies of the behavior of confidence boost:
  - does this notion prune off rules heavily enough?
  - how satisfactory are the chosen rules for the domain expert?
  - how real in practice is the risk of exponential search space?
- any intuitions on how to set the thresholds?
- how (or *whether*?) to apply the closure-aware version to the GD basis?
- mine association rules by setting a single, robust parameter?
- other studies and evaluations?
- Python-based fully open implementation on top of Borgelt's apriori available in slatt.googlecode.com (download v.0.2.2 and edit file main.py – but usual no-guarantees disclaimer).