The "yacaree" approach to association rules

José L Balcázar
Departamento de Matemáticas, Estadística y Computación
Universidad de Cantabria, Santander, Spain
joseluis.balcazar@unican.es

WAPA 2011

Association Rules and Implications

Brief review

Association rules (slight generalization):

Syntactic form of implications among sets of items: $AB \longrightarrow CD$.

- Items can be seen as propositional values.
- Antecedent and consequent are positive minterms.
- Sets of implications are fully equivalent to Horn formulae.
- A canonical minimum-size basis exists for implications, with respect to a natural notion of entailment.
- The AFP query learning algorithm for Horn formulae constructs exactly this canonical basis.
- The notion of entailment has a sound and complete syntactical counterpart: a derivation calculus.
- Association rules generalize implications by allowing a limited amount of exceptions.

Implications, I

And a real-life example

Main Property

A set of models can be axiomatized by a Horn Formula if and only if it is closed under intersection.

Immediate connection to closure spaces.

Examples

From log files of a virtual learning platform.

- 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

hilbert $\Longrightarrow$ space

margin support $\Longrightarrow$ vector

carlo $\Longrightarrow$ monte

monte $\Longrightarrow$ carlo

Example from a "census" dataset:

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

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

Similarly, Wife $\Longrightarrow$ Female does not hold either:

there are two tuples declaring Male and Wife.

The Confidence-and-Support Framework, I

The support bound

Algorithms extract all association rules not disproved in a dataset.

The "not-disproved" condition can be made somewhat strong through the notion of support.

Why do we impose a support lower bound?

Two reasons:

- "Una golondrina no hace verano":
  reduce potential spurious statistical artifacts.
- Exponential powerset size may blow up core memory:
  - Slow-down due to virtual memory leads to stalling.
  - Even the hard drive availability can be exhausted.
  - Huge closure lattices take looooooong to explore.
- The price: How to set the support threshold? There are examples of datasets leading to both extreme behaviors:
  - algorithms choke if you ask them to go below 98%,
  - algorithms find nothing until reaching down to 0.1%.

The Confidence-and-Support Framework, II

The confidence bound

The very notion of association rules is not fully defined until the measure of "intensity of the implication" is agreed to.

Most popular measures:

Confidence (that is, frequentist conditional probability); lift, leverage, weighted relative accuracy...; but:

- How to set the threshold?
- Many measures take values in unpredictable intervals: what does the threshold "mean"?
- The basic properties of association rules depend of the implication intensity notion.

  (Is $A \longrightarrow B$ equivalent to $A \longrightarrow AB$?)

Comparing Basic Data Mining Tools

Associations are not the most successful technology so far

End-user point of view

According to our very limited experience so far (mainly, in the Educational Data Mining area):

- Clustering may end up being rather successful,
- classification also, maybe to a lesser degree,
- but associations do not really work straight away.
  - Most association miners yield very redundant rules.
  - Hardly any sensible rule is found:
    "Most sessions start at the front page"
    "Most sessions visiting assignments start at the front page"
    "Most sessions visiting grades start at the front page"
    "Most sessions visiting contents start at the front page"
    ...

Example

Dataset on Contraceptive Method Choice

Standard miner output

At 90% confidence and 10% support, among others,

1. wife-education=4 contraception=2 207

⇒

media-exposure=0 207 conf 1

2. husband-education=4 no-working-now=1 standard-of-living=3 202

⇒

media-exposure=0 202 conf 1

...

64. husband-occupation=1 contraception=2 156

⇒

media-exposure=0 153 conf 0.98

Example

Dataset on Abstracts of PASCAL Reports

Standard miner output

At 70% confidence and 5% support, among others,

vector $\leftarrow$ support (12.6, 81.3)

support $\leftarrow$ vector (13.3, 77.1)

vector $\leftarrow$ machines support (6.4, 100.0)

support $\leftarrow$ machines vector (6.5, 97.9)

vector $\leftarrow$ support using (6.0, 88.4)

support $\leftarrow$ vector using (6.2, 84.4)

vector $\leftarrow$ support data (5.4, 82.1)

support $\leftarrow$ vector data (5.8, 76.2)

vector $\leftarrow$ support paper (5.4, 82.1)

support $\Leftarrow$ vector paper (5.3, 84.2)

...

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: $X \rightarrow Y$ is redundant with respect to $X' \rightarrow Y'$ if $\mathrm{conf}(X \rightarrow Y) \geq \mathrm{conf}(X' \rightarrow Y')$ in every dataset.
- A natural variant, closure-based redundancy, reads the same, but under a condition to share the same closure space.
- That variant offers a way to treat separately implications from partial rules; implications "sneak in" anyway, and they allow better summarization through the canonical basis.

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, III

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.
- Beyond two partial rules?

Redundancy in Association Rules, IV

Minimum-Size Bases

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

- work only among closures: both X and Y must be closed;
- confidence of $X \longrightarrow Y$ must be at least γ;
- but falls below γ if either we enlarge Y, or we reduce X.

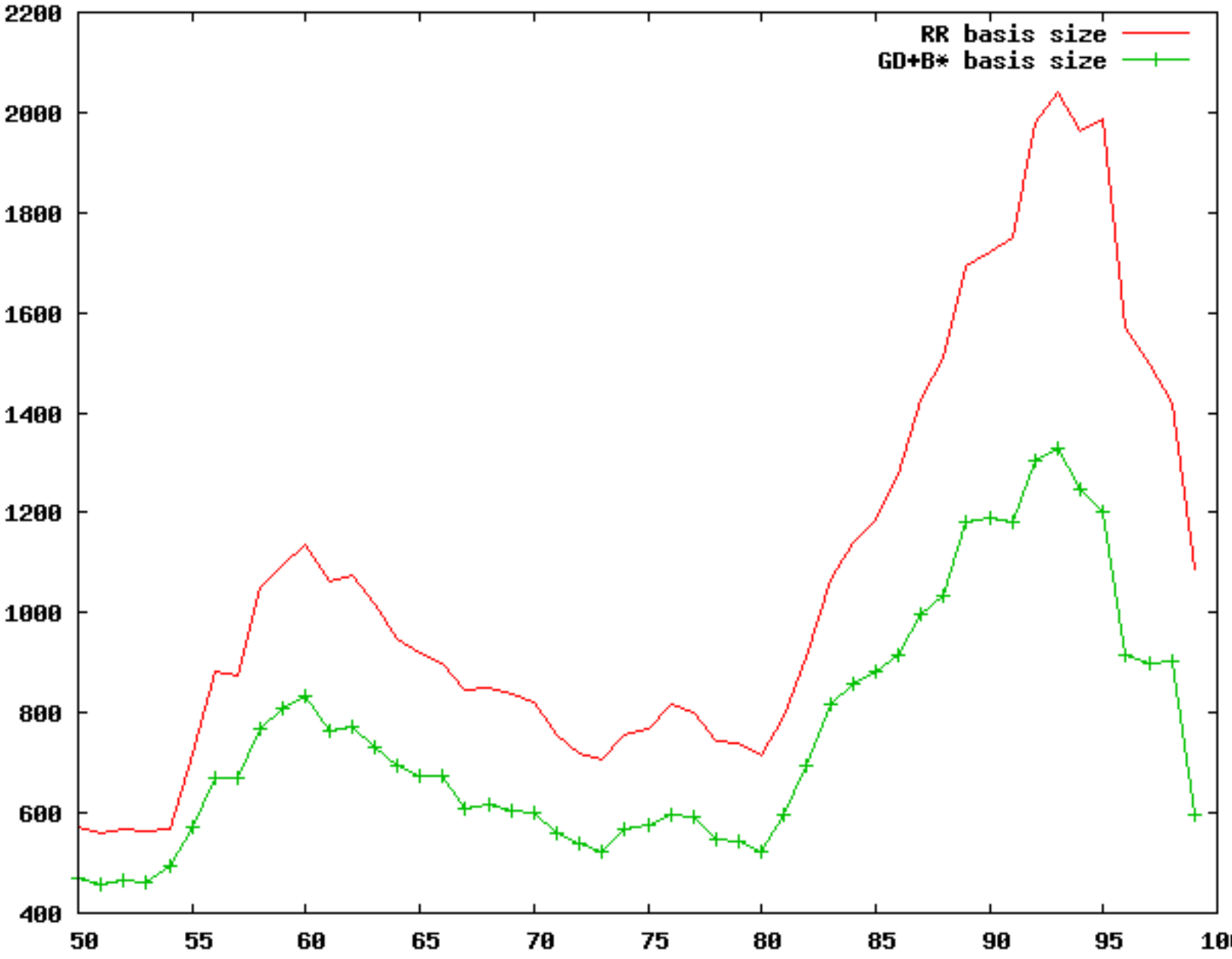
Basis $\mathcal{B}^*$: $X \longrightarrow Y - X$ for all closed Y and all basic antecedents X of Y, provided $Y - X \neq \emptyset$.

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`

Inspires a notion of "novelty"



Closure-Based Confidence Boost

One way through

For the usual confidence-and-support scheme:

High thresholds give nothing of interest, but lowering them (specially confidence) leads to too many rules to browse manually. How to discard rules?

- "Logical" redundancy approach: $\mathcal{B}^\star$ irredundant basis;
- "logical" novelty: confidence width (promising, but still somewhat unsatisfactory);
- (closure-based) confidence boost:
  - an "intuitive" variant of redundancy,
  - a measure of novelty,
  - advantageous not only in that it sets apart few rules, but also in the intuitive quality of the rules.
- Ongoing end user validation of the scheme for educational datasets.

Relative Confidence
Key intuition

Picture yourself reading association rules:

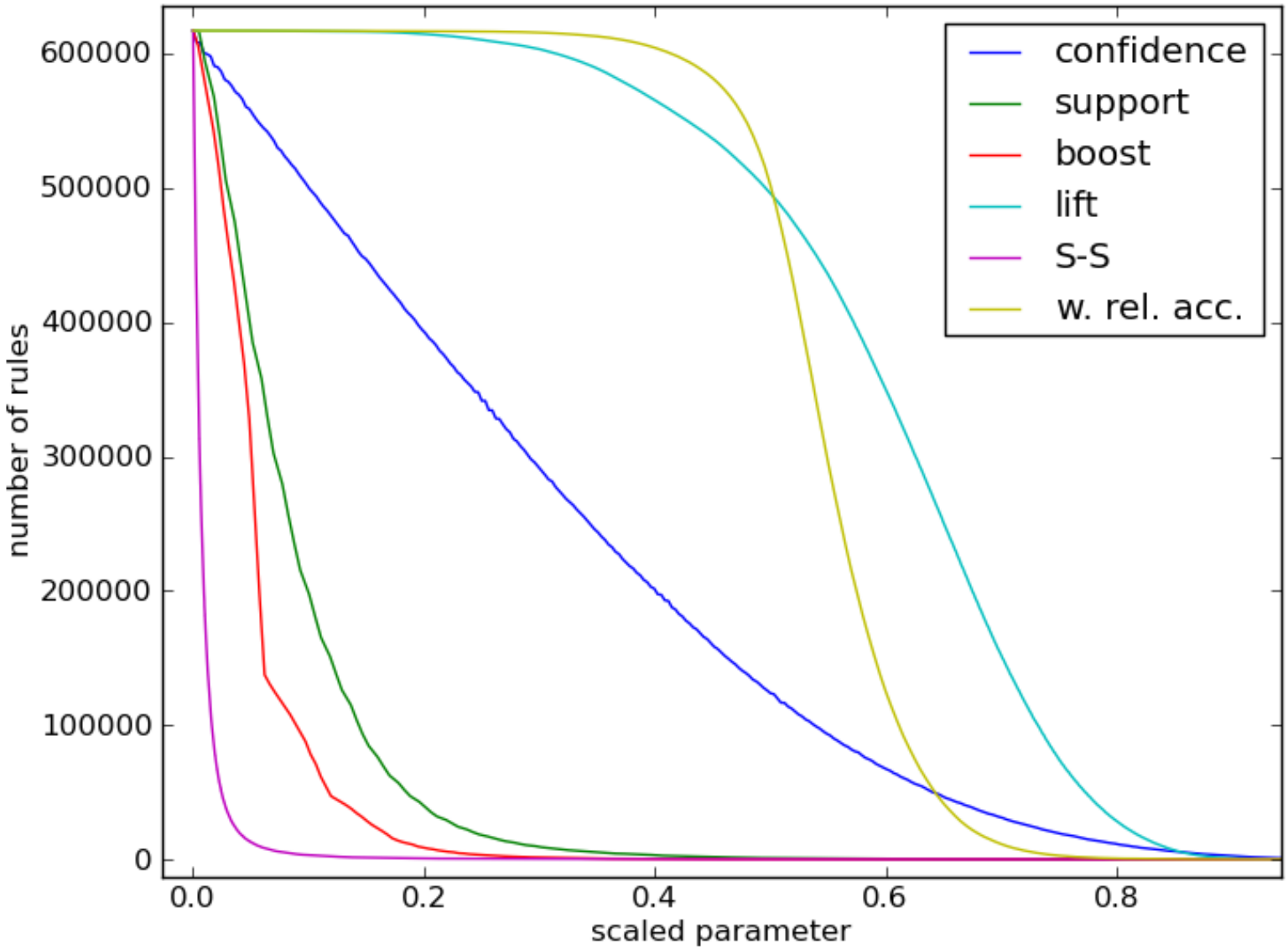
Do you really feel compelled to work out and understand a rule that you know only gives you, say, 3% more confidence than another, simpler one that would make it redundant?

- An irredundant rule is so because its confidence is higher than what the rest of the rules would suggest;
- then, one can ask: "how much higher?"
- $\overline{\beta}(X \to Y) =$

$$= \frac{c(X \to Y)}{\max\{c(X' \to Y') \mid (\overline{X} \neq \overline{X'} \vee \overline{XY} \neq \overline{X'Y'}),\ X' \subseteq \overline{X},\ Y \subseteq \overline{X'Y'}\}}$$

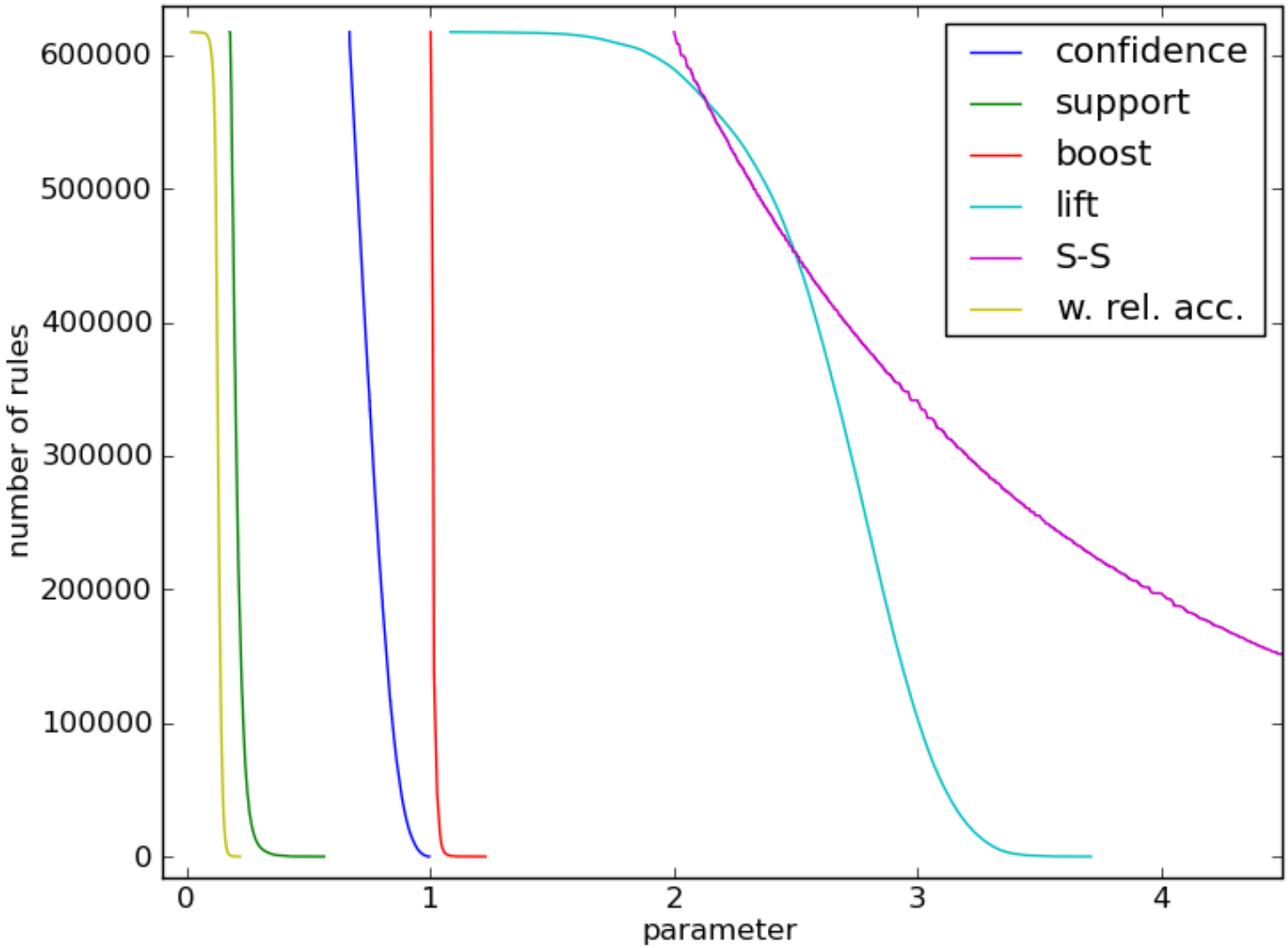
Behavior of Some Measures, I

Comparative evaluation, scaled (dataset "house-votes-84")



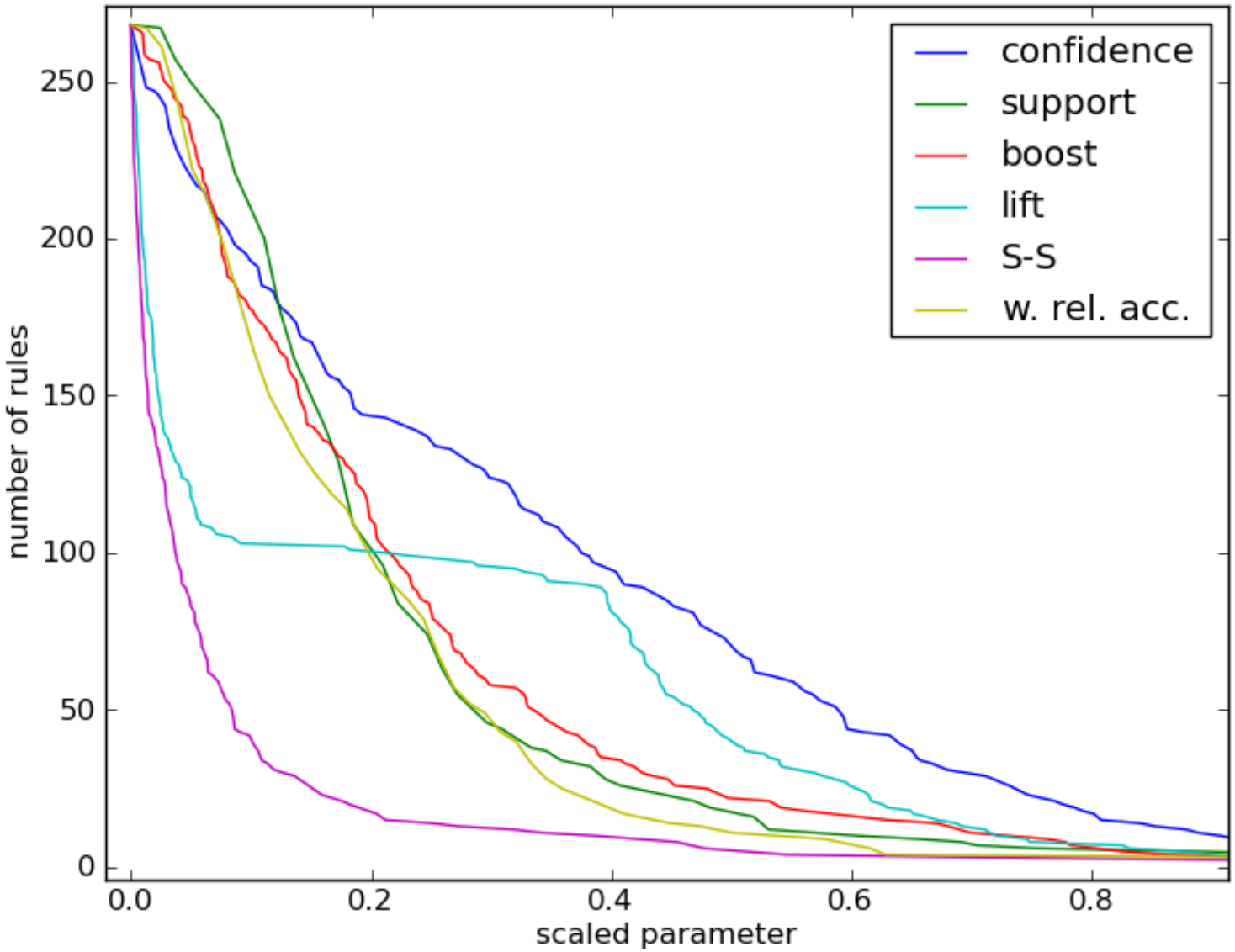
Behavior of Some Measures, II

Comparative evaluation, not scaled (dataset "house-votes-84")



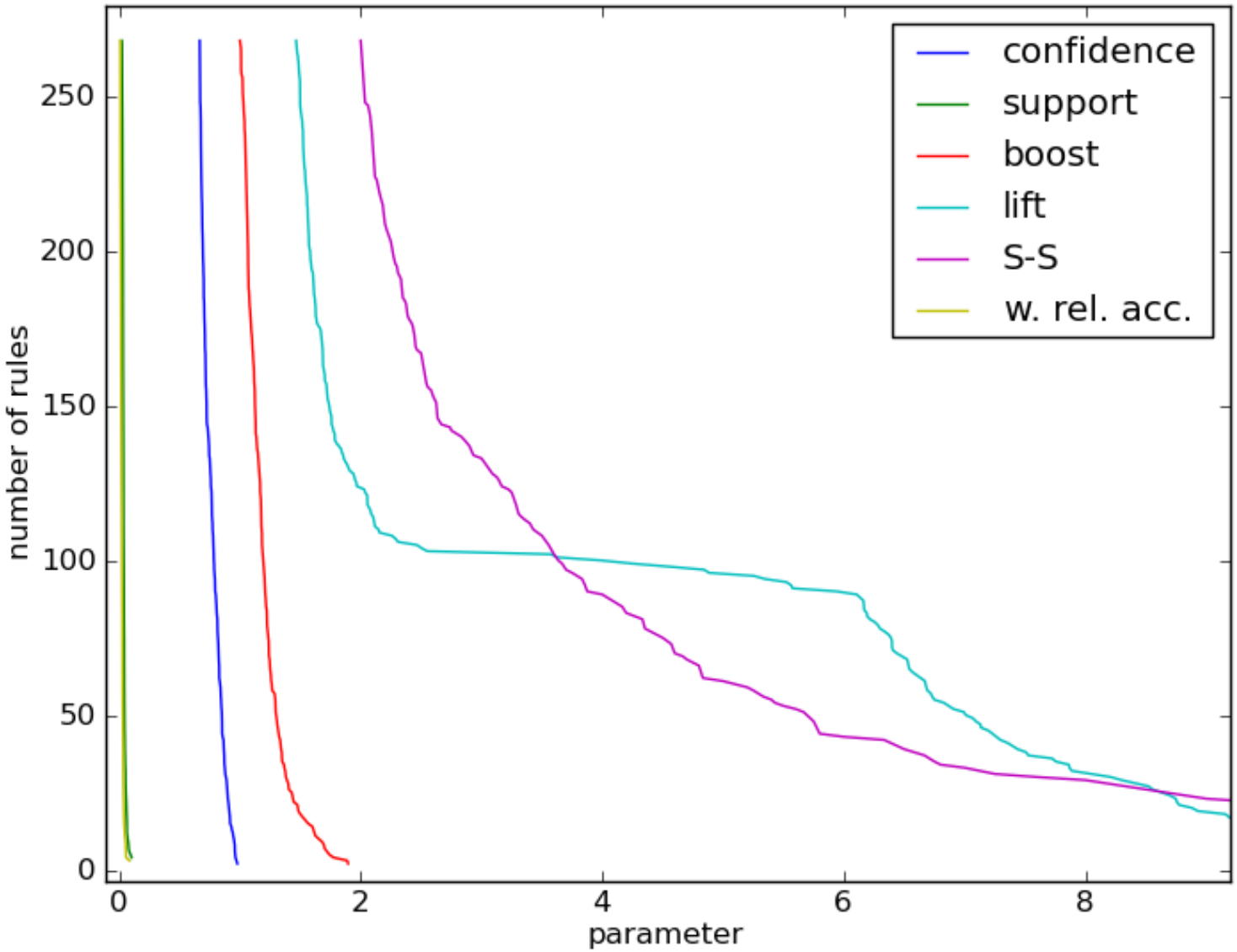
Behavior of Some Measures, III

Comparative evaluation, scaled (dataset "Pascal Reports")



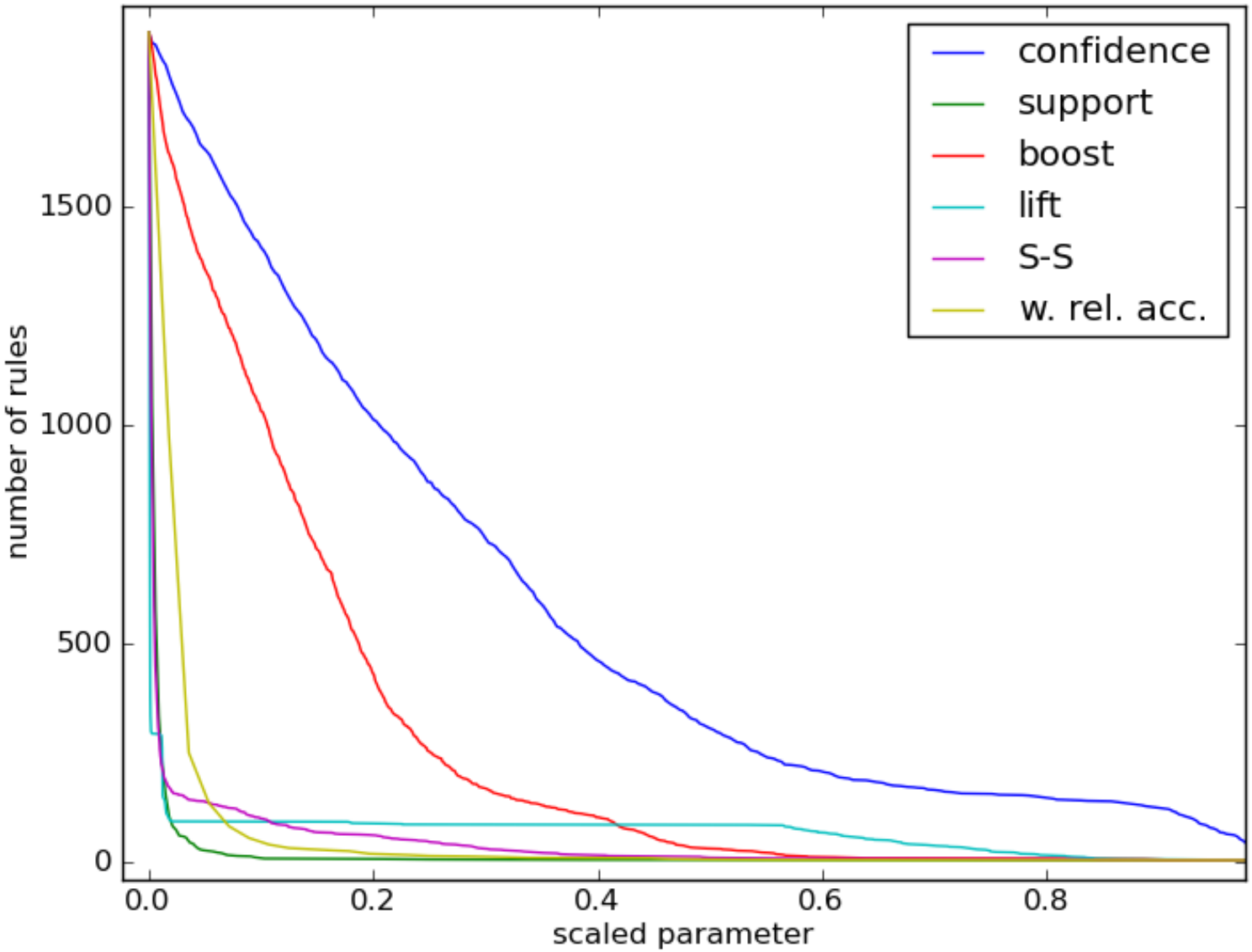
Behavior of Some Measures, IV

Comparative evaluation, not scaled (dataset "Pascal Reports")



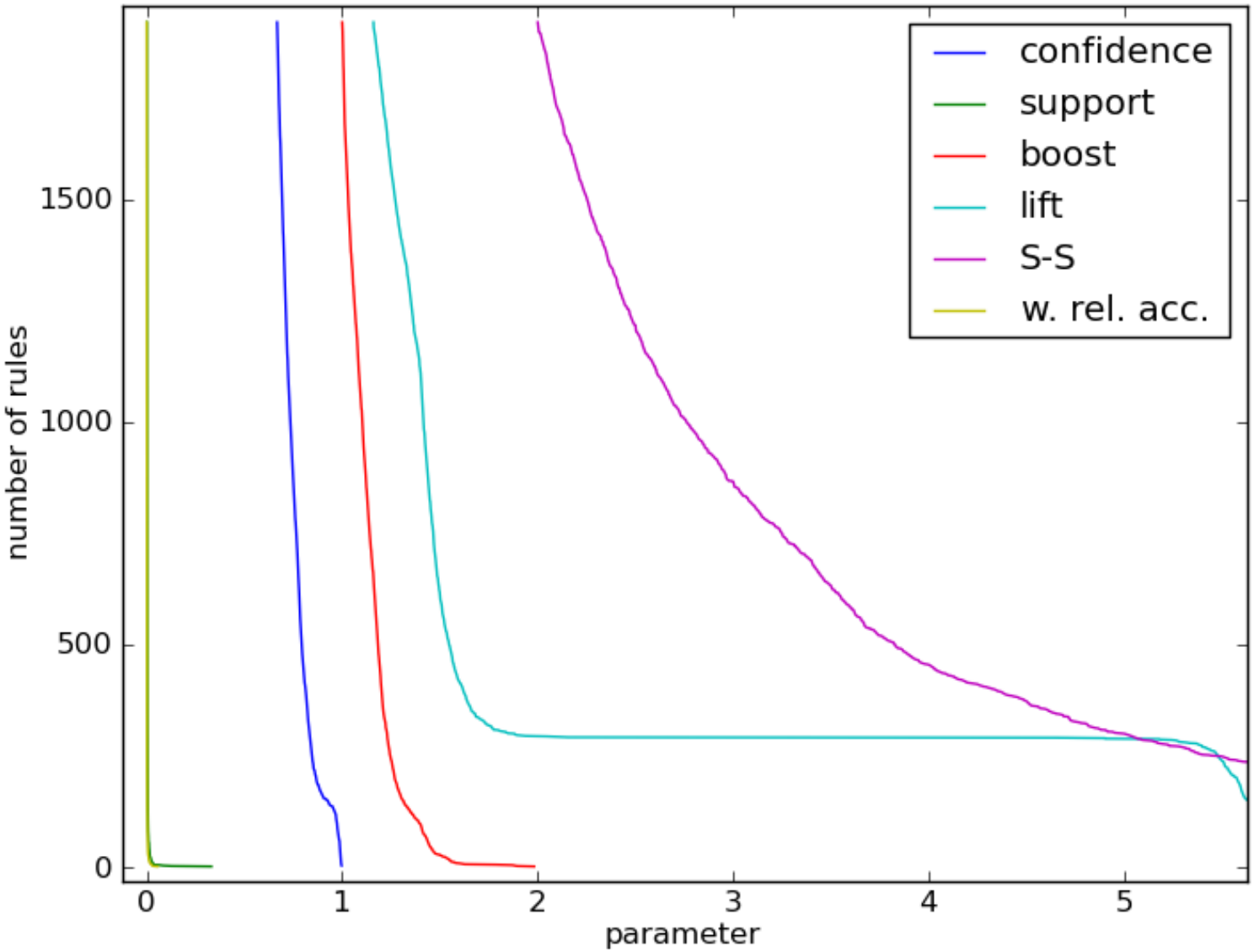
Behavior of Some Measures, V

Comparative evaluation, scaled (dataset "retail")



Behavior of Some Measures, VI

Comparative evaluation, not scaled (dataset "retail")



Some Empirical Observations

That suggest that this notion may work

Compared to other measures

No one really wins on all tests.

- Closure-based confidence boost may reach quite large values,
- but it does so for really few rules (about zero to ten);
- in most datasets, few or no rules reach boost above 1.5;
- in many datasets, few rules reach boost above 1.2;
- in many datasets, not too many rules reach boost above 1.05.

Parameter-Less, "Self-Tuning" Option

A somewhat bold attempt

Ask nothing from the user

(except of course the file name of the dataset).

- Confidence? Set it somewhat low, keep it constant!
- Support? Set it low and keep it constant as far as you can afford it, then increase it if needed.
- Confidence boost? Set it high and keep it constant as long as you keep finding rules, then decrease it if needed.
  - Connection to support ratio allows one to partially push the confidence boost constraint into the mining process.
  - Connection to lift for certain syntactic form of (usually quite abundant) rules allows one to monitor the rules and trigger the weakening of the confidence boost threshold.

Implementation
It sort of works!

`yacaree.sf.net`

(yet another closure-based association rule exploration environment)

Self-adjusting support and confidence boost,

Plus hardwired constants (easy to modify for a programmer) for

- the confidence threshold,
- the "learning rate" of the confidence boost, and
- start and limit values of both support and confidence boost.

Further Questions and Perspectives, I

Improvements necessary (sort of immediately)

To do:

Improve quite a few suboptimal algorithmics; also, some technological "dirty tricks" to be removed (and others added).

- Improve the ridiculously slow test to find out when to increase the support threshold; trying out one option for version 1.1.
- Better algorithm to construct the Hasse edges that define the closure space on which we work; theorems proved, implemented in version 1.1 — 10% running time improvement.
- Being a bit smarter in pushing the support ratio constraint into the miner; to be implemented in version 1.1.
- Finish up at once version 1.1 please!

  (Have been telling this to myself for about six months.)
- Make the system more widely available?

Further Questions and Perspectives, II

Beyond!

Next issues to study:

- Handling implications?
  - The rule basis we work with is only for partial associations.
  - Needs a separate processing of implications. But...
  - Closure-based confidence boost is inappropriate for them!
  - Recent results and algorithms: plain confidence boost, "ignoring" the closure operator, seems to work.
  - For version 1.2?
- A better closure miner?
  - We think we have the algorithm.
  - We have half of its correctness proof.
  - For version 1.3?

Further Questions and Perspectives, III

Even further beyond!

- Implement approximations to handling negated items
  - Inmensely helpful in EDM if implemented "the right way".
  - Ongoing theoretical development hitting some difficulties.
  - For version 1.4?
- Apply recent advances regarding statistical validation tests.
  - Is the data "everything there is"? Or, is it a sample under a probability distribution, allowing us to study things like p-values?
  - How to reconcile this with the "closure-based" approach?
  - Preliminary experiments on combining sampling with closure spaces give really funny results!

Further Questions and Perspectives, IV

Even further!

Alternative proposals for formalizing novelty

Some of them quite close to the logical entailment notion.

However:

- The notion that "works" intuitively speaking does not correspond to the "logician's" view of redundancy: logic-inspired notions do not capture human intuition.
- Can we give a more refined meaning of implication within our logic that would reconcile it with the practically working notion of redundancy?

Also: 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.