Constraints on Pattern Sets

A. Zimmermann

13th December 2005

Outline

Setting

- Mining process finished
- Result: patterns and/or rules
- Question: How to select subsets?
- Answer: Constrain the sets!
- Question: How?

Setting

- Mining process finished
- Result: patterns and/or rules
- Question: How to select subsets?
- Answer: Constrain the sets!
- Question: How?

Types of constraints (I can think of)

- Syntactic constraints
  - Size
  - Which patterns allowed (possibly dependent on already included ones)
- Apriori constraints
  - Coverage
  - Overlap
- Ranking constraints
  - Significance in distribution changes
  - Accuracy
  - ROC

Types of constraints (I can think of)

- Syntactic constraints
  - Size
  - Which patterns allowed (possibly dependent on already included ones)
- Apriori constraints
  - Coverage
  - Overlap
- Ranking constraints
  - Significance in distribution changes
  - Accuracy
  - ROC

Types of constraints (I can think of)

- Syntactic constraints
  - Size
  - Which patterns allowed (possibly dependent on already included ones)
- Apriori constraints
  - Coverage
  - Overlap
- Ranking constraints
  - Significance in distribution changes
  - Accuracy
  - ROC

Setting revisited

- Mining process finished
- Result: patterns and/or rules
- 1st Question: How to select subsets?
- Answer: Constrain the sets!
- 2nd Question: How?
- 3rd Question: How to efficiently use constraints?
  Note: disjunctions!

Setting revisited

- Mining process finished
- Result: patterns and/or rules
- 1st Question: How to select subsets?
- Answer: Constrain the sets!
- 2nd Question: How?
- 3rd Question: How to efficiently use constraints?
  Note: disjunctions!

Syntactic constraints

- Size
  - (Anti-)monotone
  - Techniques exist
- Patterns allowed
  - Ad hoc: filter beforehand
  - Dynamic: enumeration tree

Syntactic constraints

- Size
  - (Anti-)monotone
  - Techniques exist
- Patterns allowed
  - Ad hoc: filter beforehand
  - Dynamic: enumeration tree

Apriori constraints

- Coverage (possibly only on subset)
  - (anti-)monotone
- Overlap
  - Pairwise: decided at 2-level, Apriori-like checks at deeper levels
  - Full set: (anti-)monotone
  - Average (on size): bounds, possibly based on shallower levels

Apriori constraints

- Coverage (possibly only on subset)
  - (anti-)monotone
- Overlap
  - Pairwise: decided at 2-level, Apriori-like checks at deeper levels
  - Full set: (anti-)monotone
  - Average (on size): bounds, possibly based on shallower levels

Ranking constraints

- Significance in distributions
  - Non-overlap: upper bounds
  - Overlap: “trivial” ubs, based on all P,N added, possibly lb, based on overlap
- Accuracy
  - Non-overlap: upper bounds
  - Overlap: trivial ub 1
  - Somewhat remedied by ordered lists, new problems arise

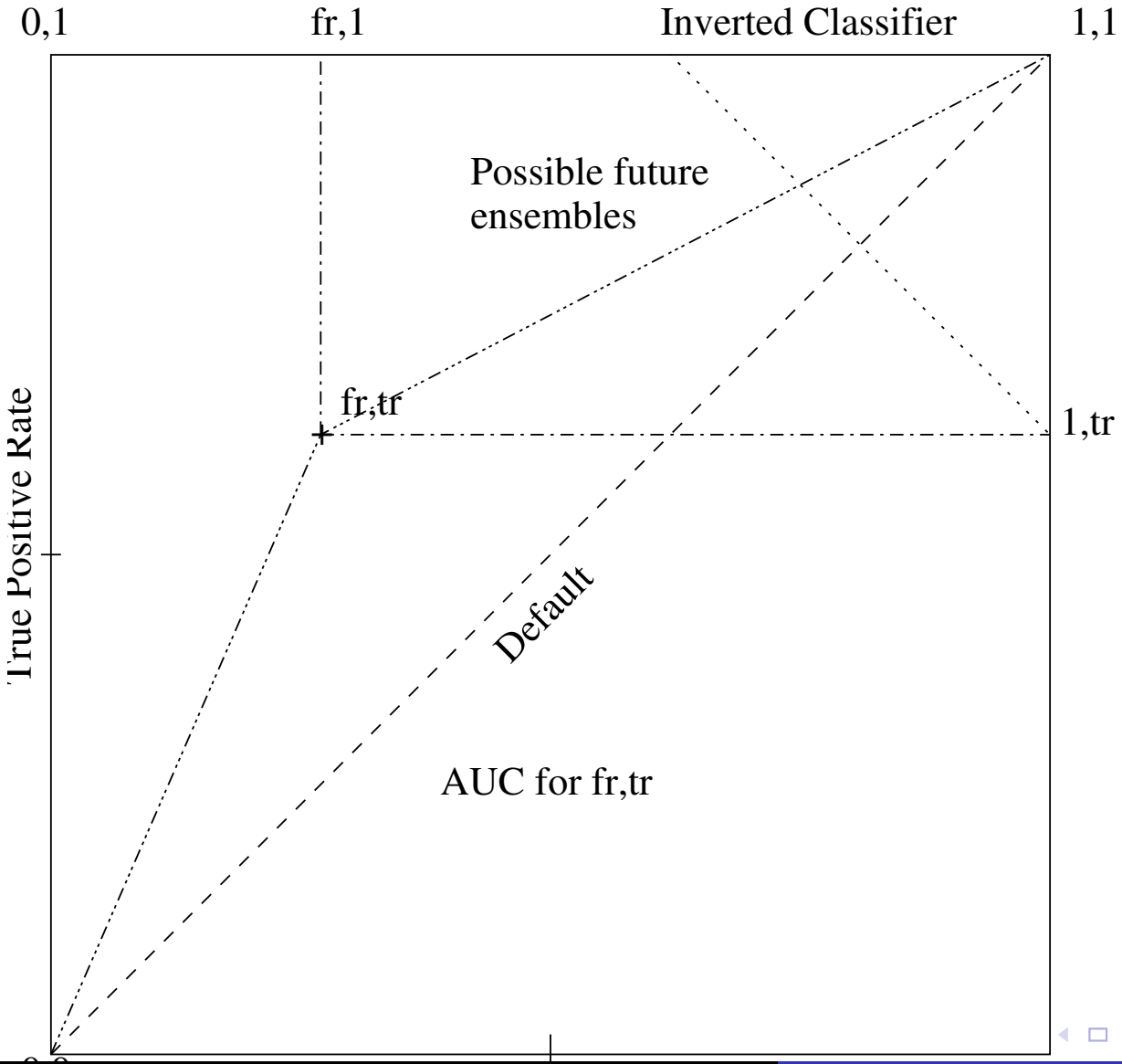
Ranking constraints

- Significance in distributions
  - Non-overlap: upper bounds
  - Overlap: “trivial” ubs, based on all P,N added, possibly lb, based on overlap
- Accuracy
  - Non-overlap: upper bounds
  - Overlap: trivial ub 1
  - Somewhat remedied by ordered lists, new problems arise

Ranking cont. - ROC

- ROC
  - AUC above threshold, non-overlap: upper bounds
  - AUC above threshold, overlap: very trivial ubs
  - Convex hull
    - Consider each pattern as single classifier, build hull: deterministic
    - Non-overlap, build one best AUC value classifier, extend hull: upper bound by slope patterns
    - Non-overlap, build classifiers parallel: upper bound based on slope, checked against competing patterns

ROC visualized - trivial upper bounds



Overlap effects

- On significance: possible overlap with either positives or negatives $\Rightarrow$ trivial bounds
- Accuracy: possible re-classification of all current errors $\Rightarrow$ trivial bound 1
- ROC: possible re-classification of all current instances $\Rightarrow$ ubs $(1, 0), (0, 1)$

Alternatives to Non-Overlap

- Bounded pairwise overlap: adjustment bounds, probably the looser the deeper
- Full set overlap: makes adjustment harder, bounds looser
- For both: how to decide on value?
- Ordered lists: how to order, skipping one pattern effects following, possibly explicit relation between patterns needed

First Hands-On Idea

- Molfea-style mining
- Non-overlap
- ROC: AUC maximization, possibly convex hull

Your wisdom Here!