Distributed Monitoring with Collaborative Prediction

Dawei Feng*, Cécile Germain-Renaud*, Tristan Glatard**

* Laboratoire de Recherche en Informatique, CNRS – Université Paris-Sud 11
** CREATIS, CNRS - INSERM - Université Lyon 1 - INSA Lyon

Motivating application

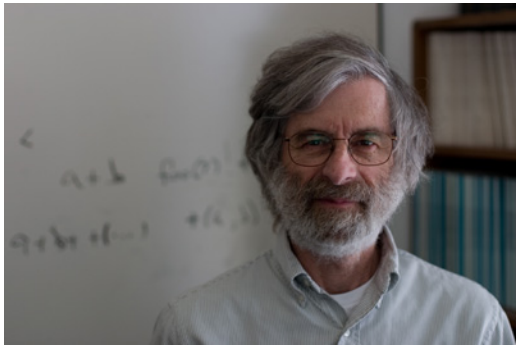
- File access failures for Biomed VO in EGI
  - European Grid Infrastructure
    - Large-scale, production grid: 350 sites, 250K cores
    - Significant failure challenge: typical resource availability ~90%
    - EGI Vocabulary:
      - a Computing Elements (CE) a computing system
      - a Storage Element (SE) is the grid disk system
  - Biomed Virtual Organization (VO)
    - Biomedical computational research
    - File access is essential and the most problematic
    - Limited manpower for resource monitoring / incident reporting

Cloud thunder

- Amazon cloud service crash
  - 2011, in April and in August, …
  - Wiped out many customers' data permanently
  - Netflix, reddit, instagram, quora, foursquare…

Failure: it's not a bug, it's a feature

« *A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable* »



Realistic goal: limit the impact of failures

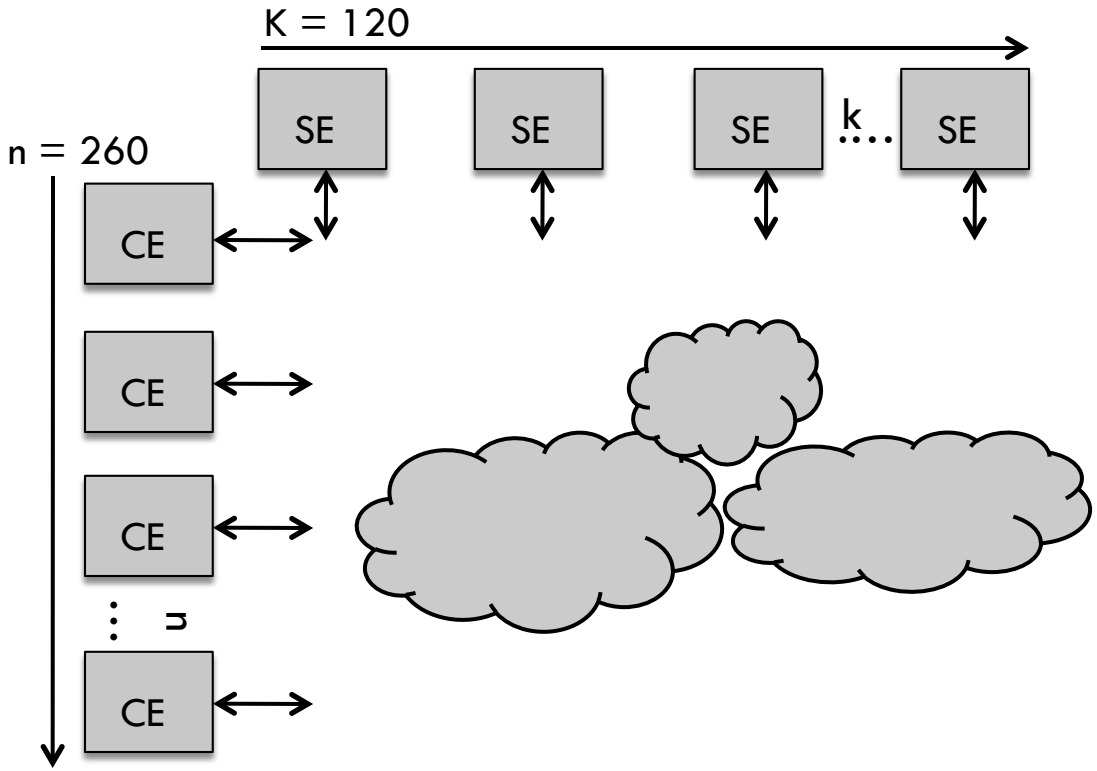
Challenges

- Distributed systems are complex black box
  - Thousands of systems/services/applications is typical
  - Multiple levels of abstractions and interactions between components
- Failures caused by uncountable reasons
  - Systems/Applications (S/A) change rapidly
  - S/A's environments evolve dynamically
- Monitoring info has gone beyond humans' capabilities
  - Drowning in data, but starving for knowledge
  - Multiple types of data sources
  - Repetitive tasks for the same problem

* Ira Cohen, Machine Learning for Automated Diagnosis of Distributed Systems Performance , 2006

Available tools: Probes

- A Probe is
  - A test transaction whose outcome depends on some of the system's components;
  - general examples: ping, trace route, web-access, database query, etc.

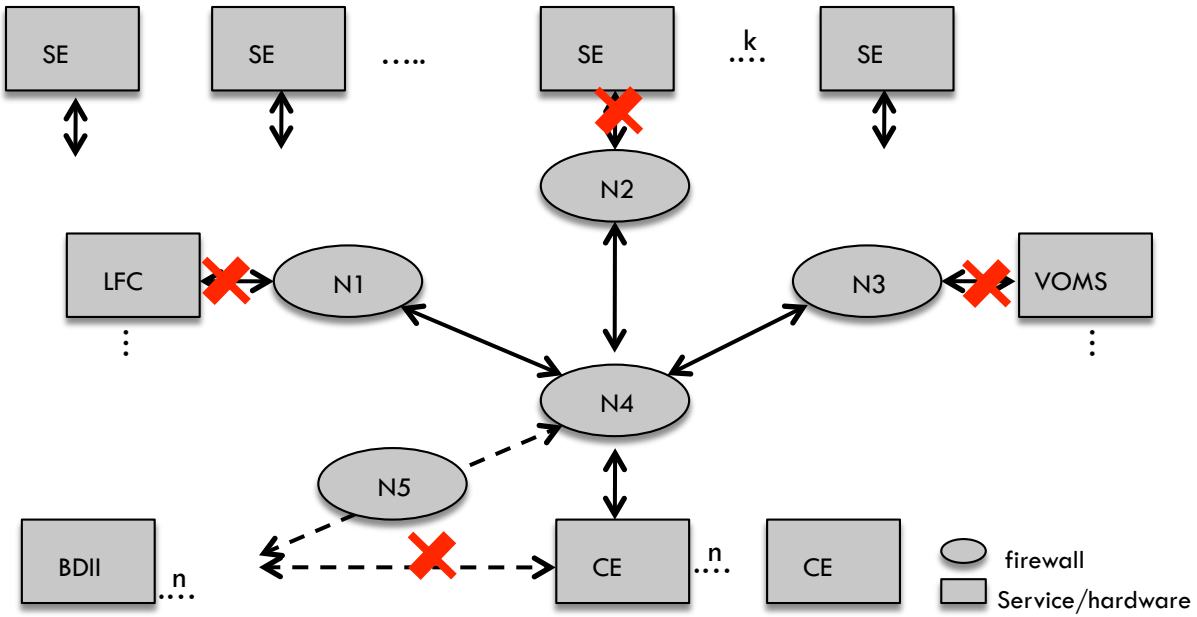


In our case:
Lcg-cr:
Read ability from CE to SE
Lcg-cp:
Write ability from CE to SE
Srm-ls:
File list ability from CE to SE
…
ALL2ALL is unrealistic

Reduce the number of tests with similar evaluation performance?

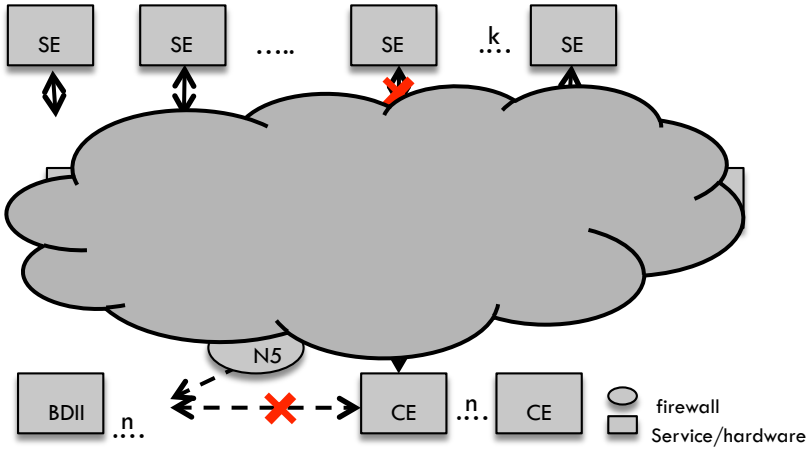
Mining faults: detection and diagnosis

- Hypothesis
  - info of the internal hardware and middleware components are known
- Goal: discovering the culprit(s) aka mining faulty components
  - Detection: say if any of the components is DOWN
  - Diagnosis: exhibit all DOWN components



Mining faults, detection & diagnosis

- Provides an explanation of the failure
- Strongly relies on a priori knowledge, dependency matrix
  - which components are required for a probe to succeed.
- A priori knowledge is useful, but not available
  - ”A distributed system is one in which the failure of a computer you didn’t even know existed can render your own computer unusable” (L. Lamport)

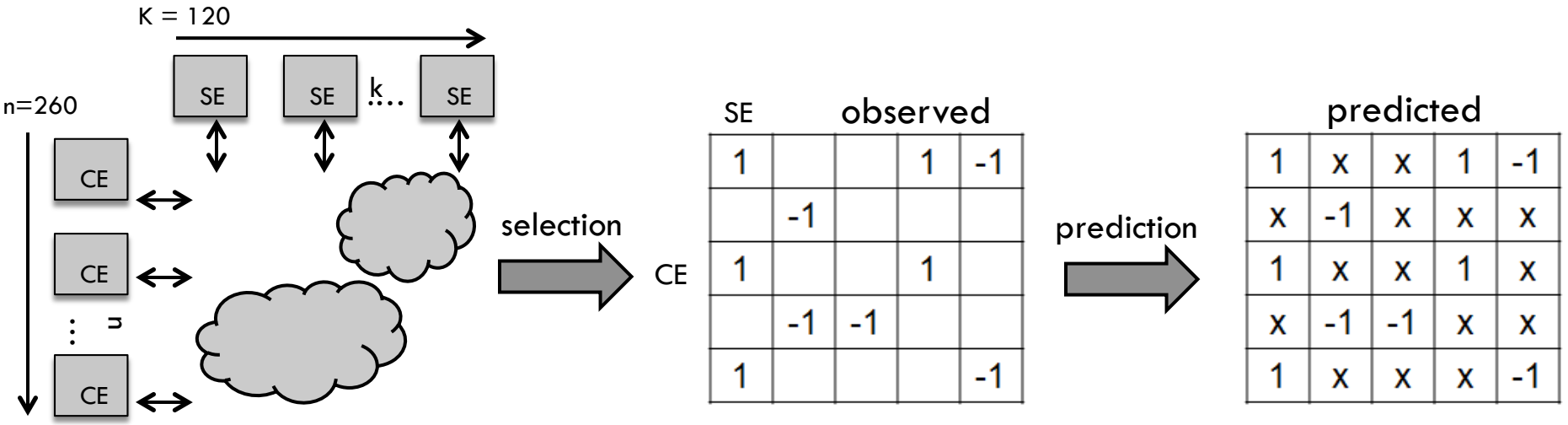


Dependency matrix

	ce-hd	bdii	lfc	voms	….
lcg-cr	1	1	1	1	
nmap	1			0	
srm-ls	1	0	1	1	
….					

Our goal: fault prediction

- Given a small number of probe results, infer the capacities for other (CE, SE) pairs
  - The overall infrastructure is a black box, with no priori knowledge of its structure.
- Two sub-goals
  - Probe selection: which subset of the (CE,SE) pairs should actually be tested?
  - Prediction: predict the availability of all (CE,SE) pairs from a small number of probes.



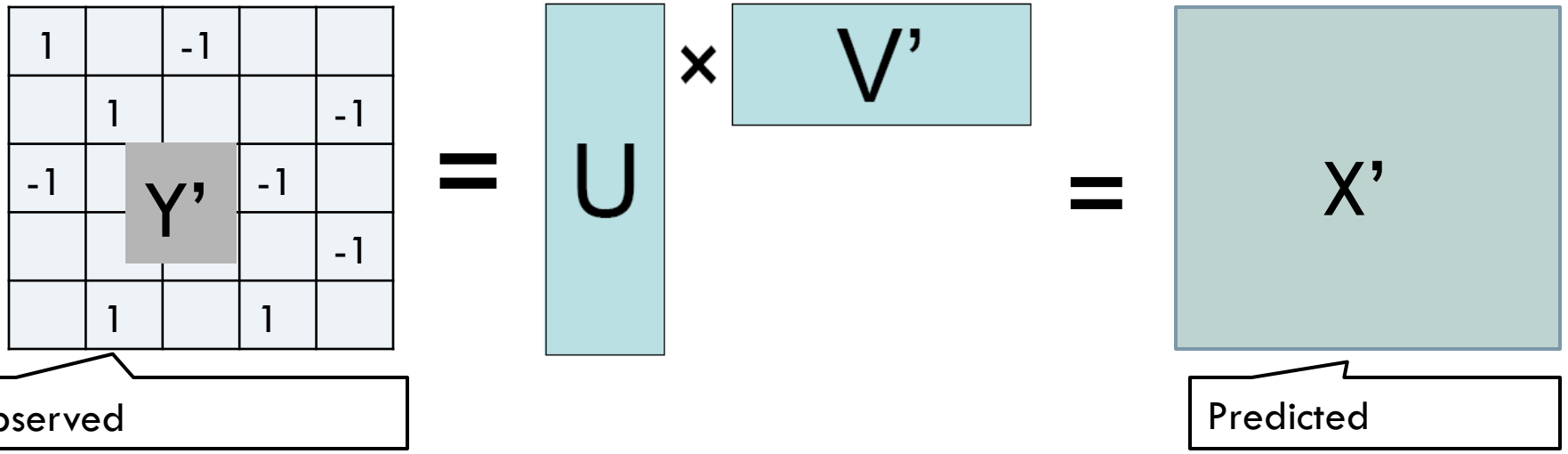
Sub-goal 1: probe selection

- Static uniform
  - Uniform random selection amongst all <CE, SE> pairs
- Active probing
  - An initial set of probes for basic info, then adaptively and incrementally select further probes
- Differentiated cost
  - Different costs for false negative & false positive
  - Real failures are more harmful than fake failures

Sub-goal 2: Prediction

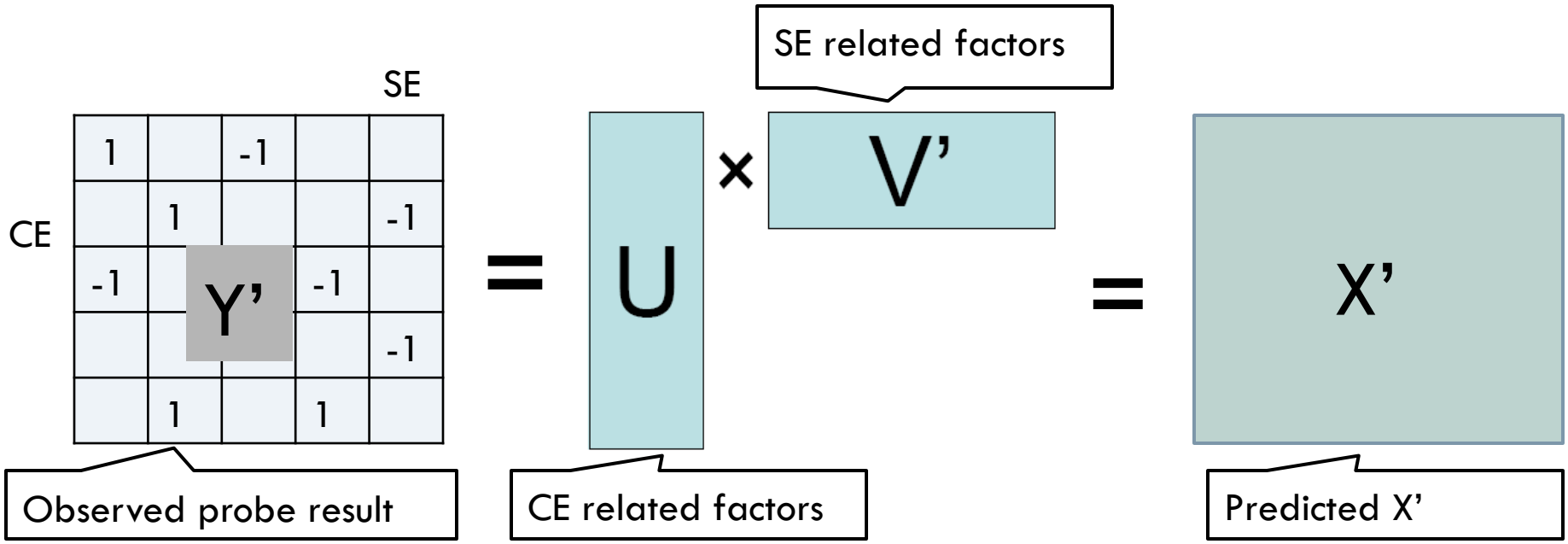
- Based on Collaborative Prediction
  - Predicting unknown ratings based on observed data from other users and products
- Hypothesis: Factor model
  - Hidden and partially shared factors affect the matrix entries
  - The techniques search for Low Rank factorization: ~ number of hidden factors
  - Technically, Maximum Margin Matrix Factorization [Srebo et al. 2005]

$$\|X\|_\Sigma + C \sum_{ij \in S} \max(0, 1 - X_{ij} Y_{ij})$$



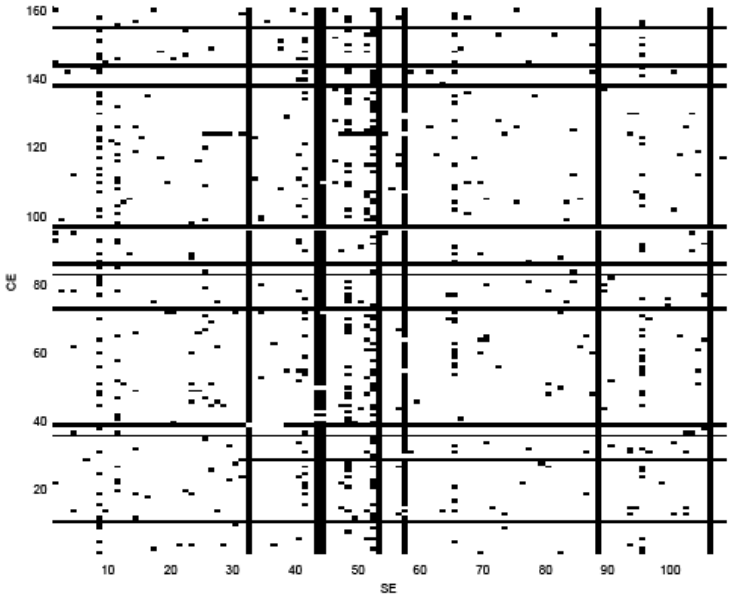
Collaborative monitoring

- Underlying hypothesis for monitoring
  - <SE, CE> pairs share some hidden factors, network connectivity, brokering, authentication, etc.
  - Low rank, # of hidden factors is small

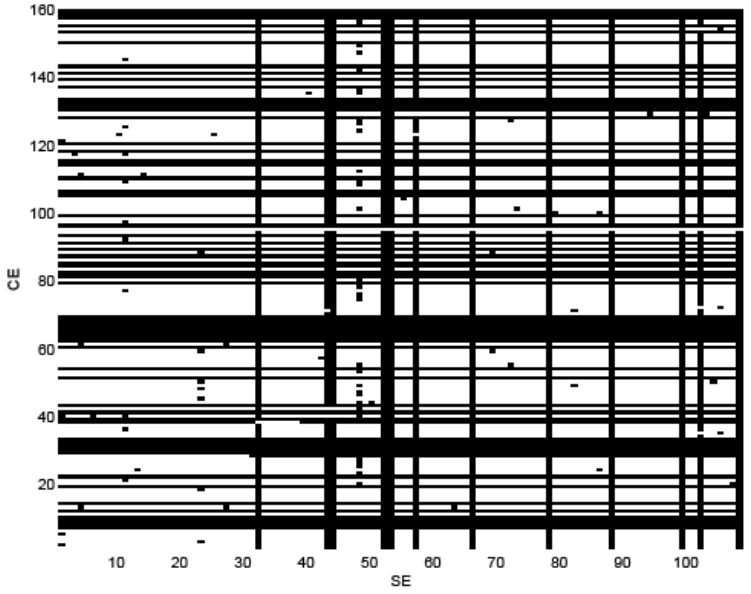


Data sets

- Fully observed all-to-all probe results of 51days on biomed
  - lcg-cr, lcg-cp, srm-ls
  - matrix model ->One measure per (CE,SE) pair per day



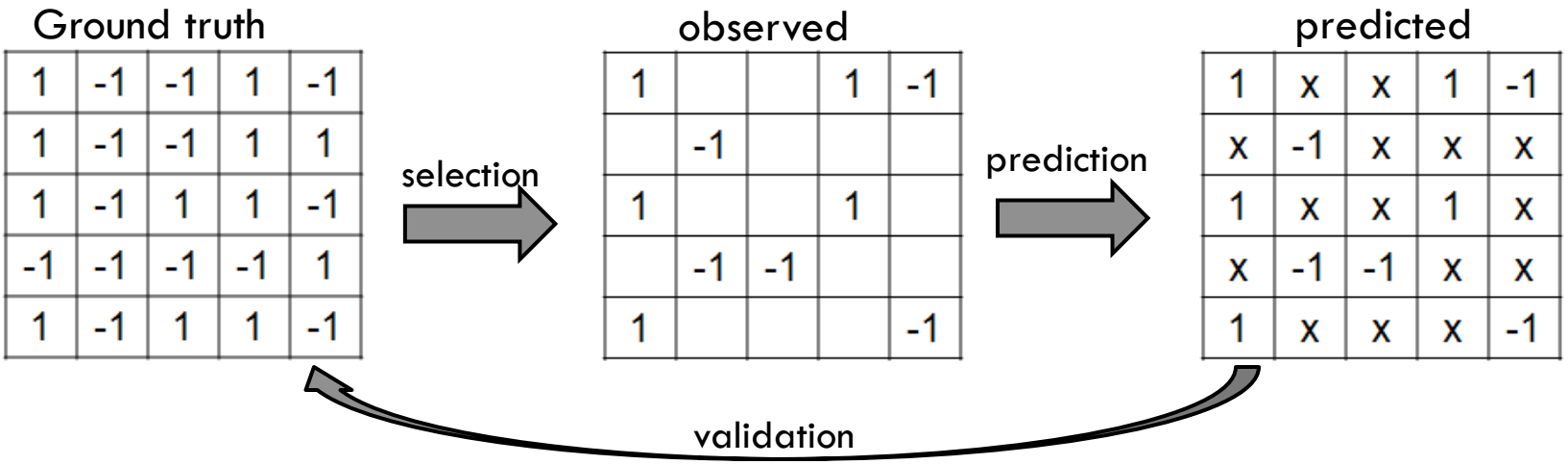
lcg-cp



srm-ls

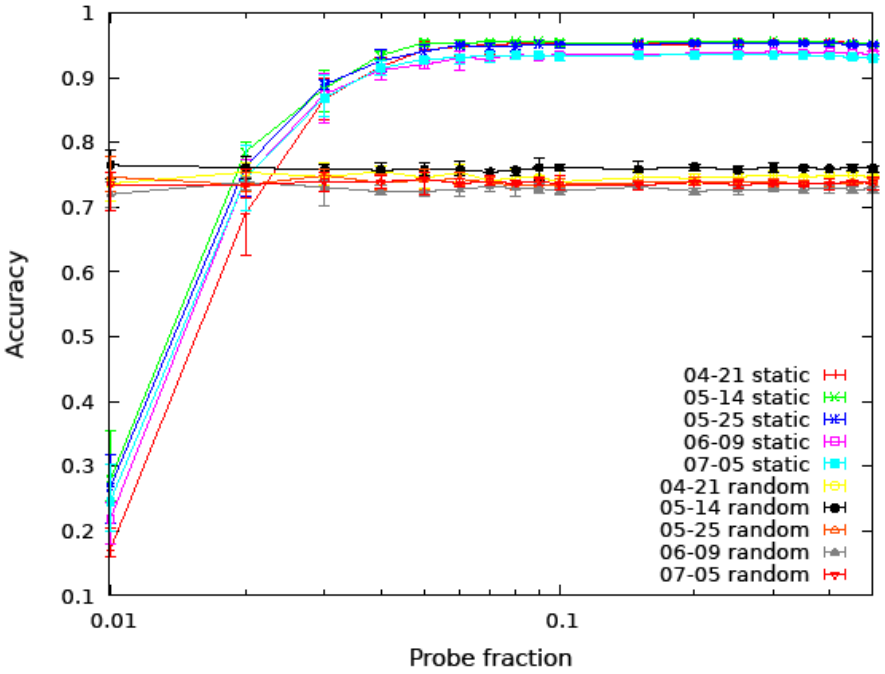
Validation methodology

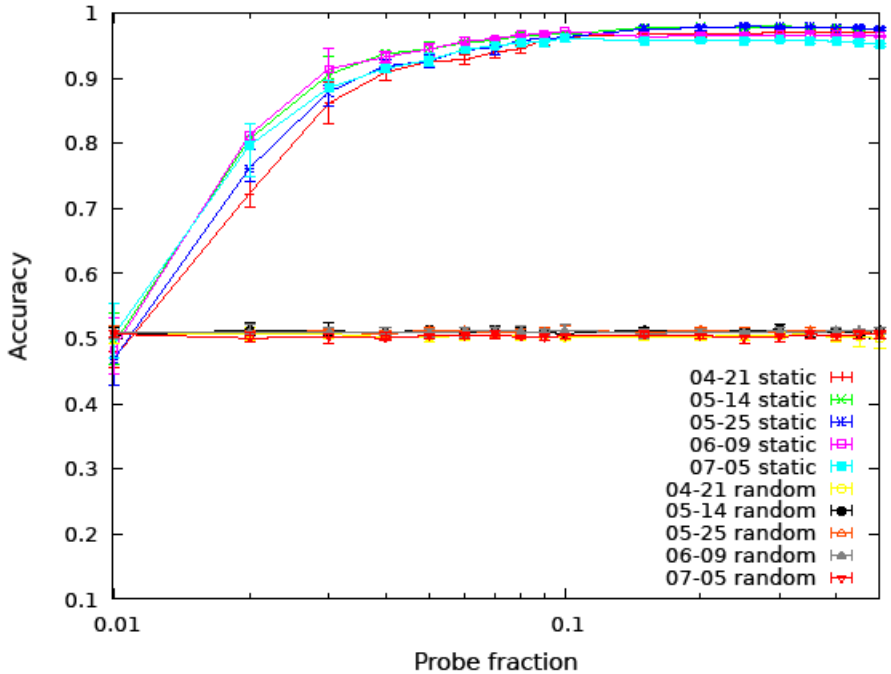
- Classical evaluation criteria
  - Accuracy,
    - Ratio of correct predictions over total predictions
  - Indicators associated with risks
    - Sensitivity, specificity, precision, MCC
  - AUC, area under curve
    - Intrinsic quality of a binary classifier independent of decision threshold



Static-uniform

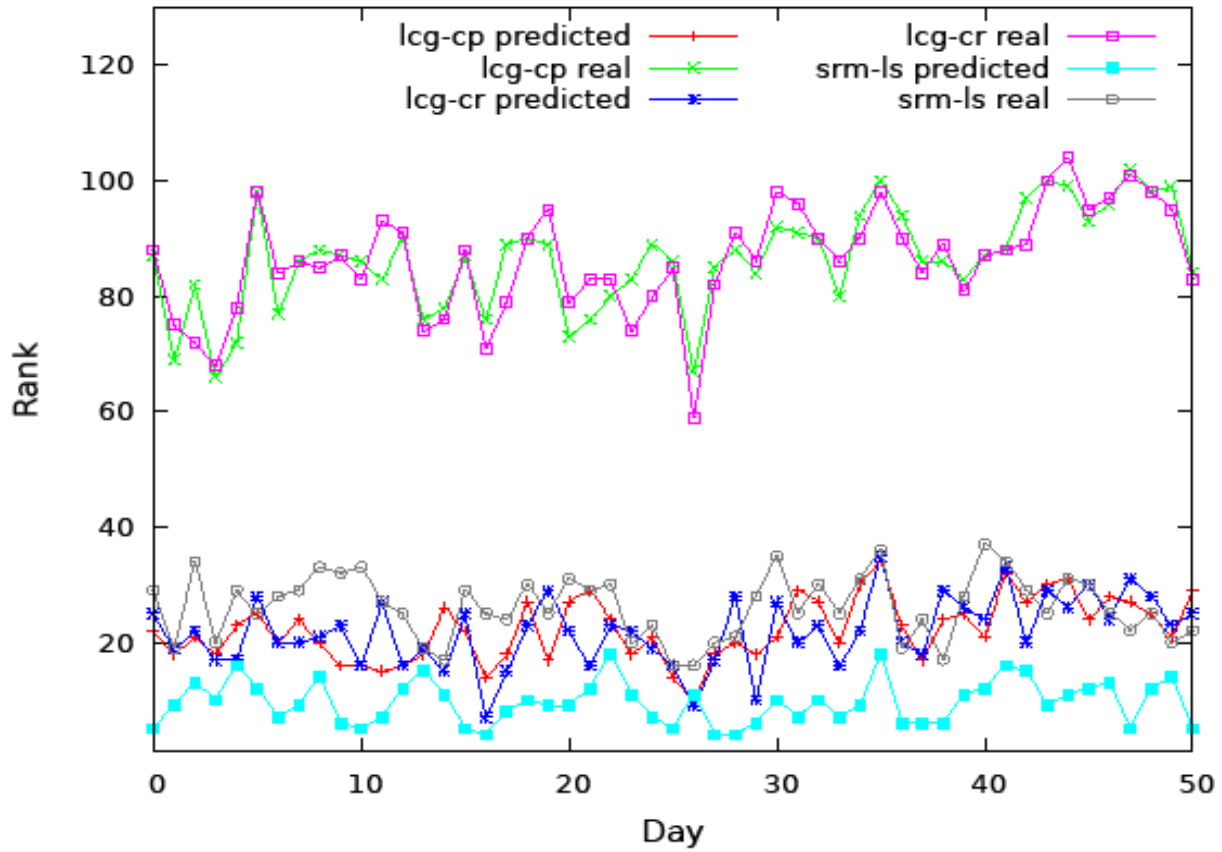
- Select probes uniformly at random among all <CE, SE> pairs.
- 95% accuracy can be reached by 5% of the probes





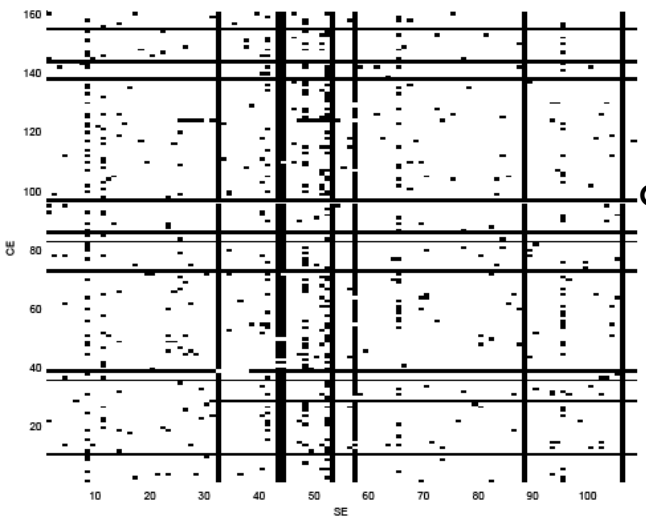
Static-uniform

- Rank, predicted vs. real
  - a small number of causes dominates the overall behavior

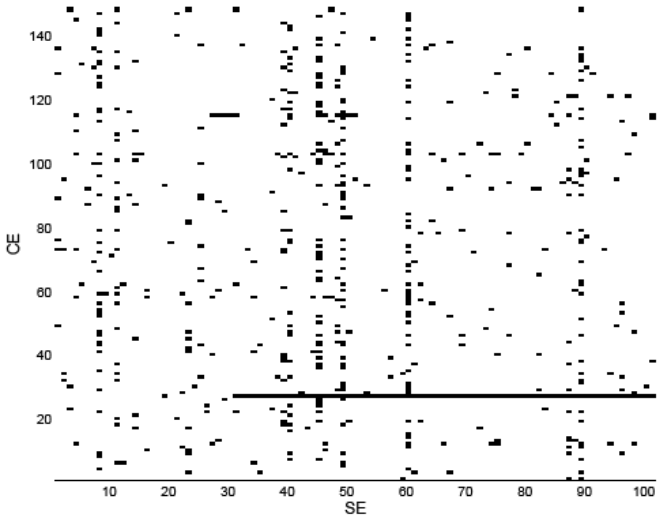


Without systematic faults

- Systematic faults:
  - decommissioned hardware, no authorization,...
- When systematic faults are eliminated (curated matrices)
  - Accuracy is easy, but uninteresting: predicting always OK gets ~97% accuracy
  - Precision is not that good with static uniform, eg
    - sensitivity is 0.32, meaning that 68% of the failures are not predicted
    - precision is 0.49, meaning that amongst the predicted failures, 51% are spurious.

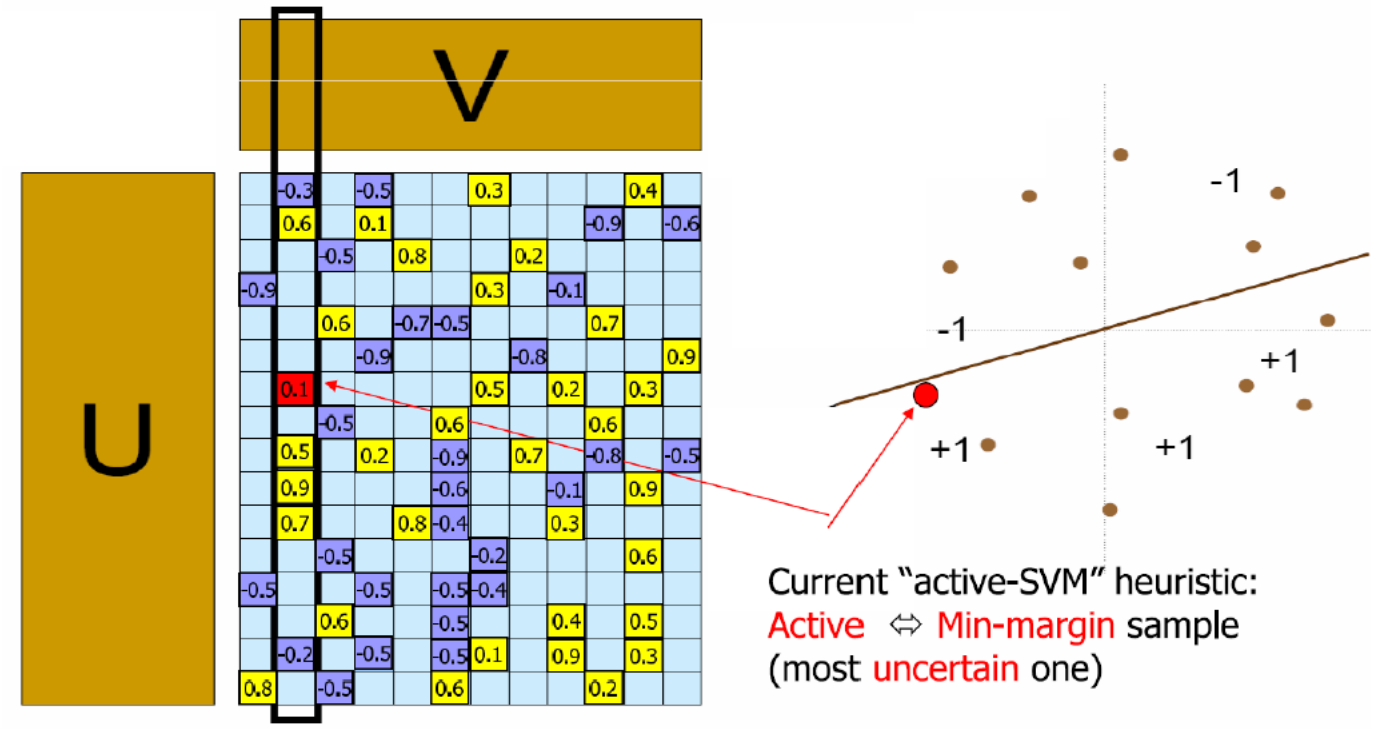


delete systematic faults



Active probing

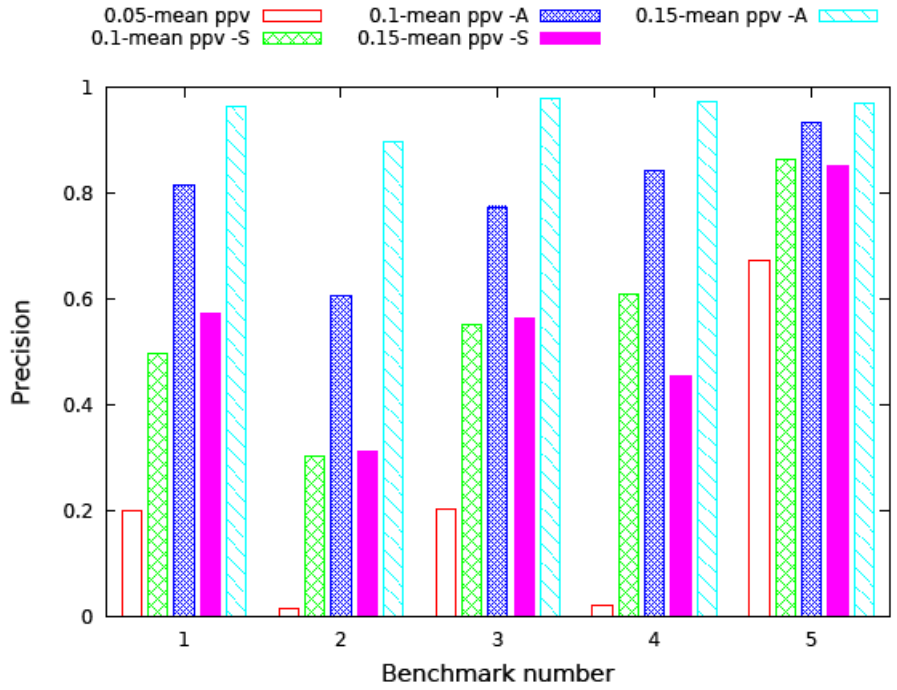
- Construct probes set dynamically
- Here, *min margin* (or *most uncertain*) heuristic

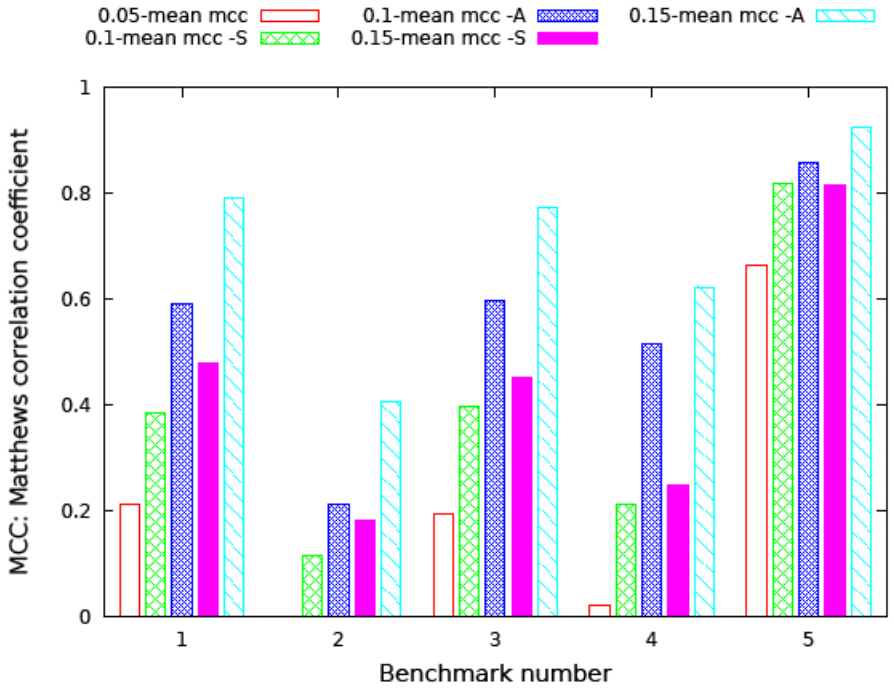


* Slide from 'Active collaborative prediction with maximum margin matrix factorization. Rish 2007'

Active probing

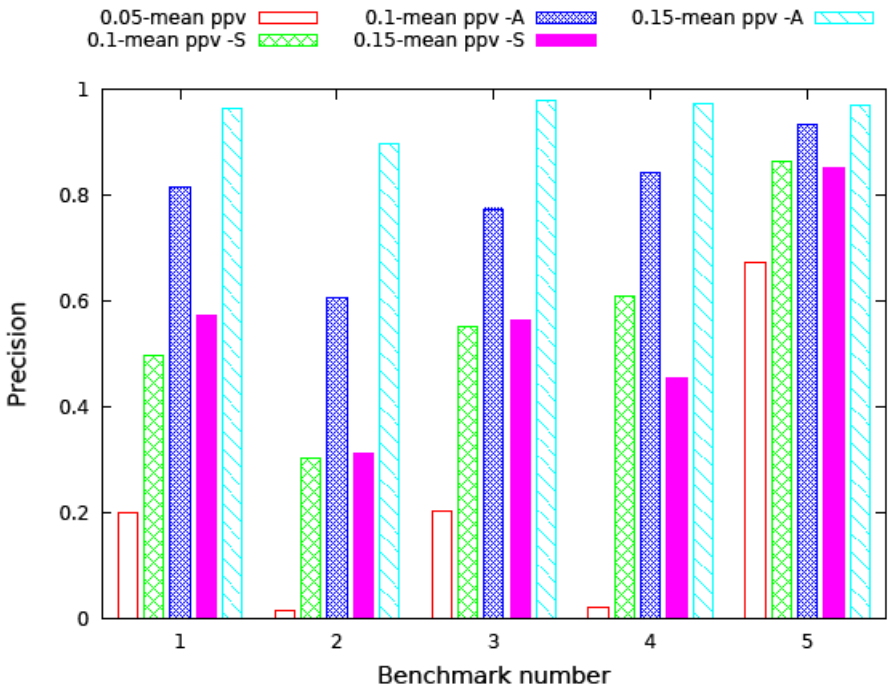
- Experimental setting: initial select 5% of probes at random, the add probes by x increment
- Compare at constant probe cost
- Active probing increases performance significantly for all criteria

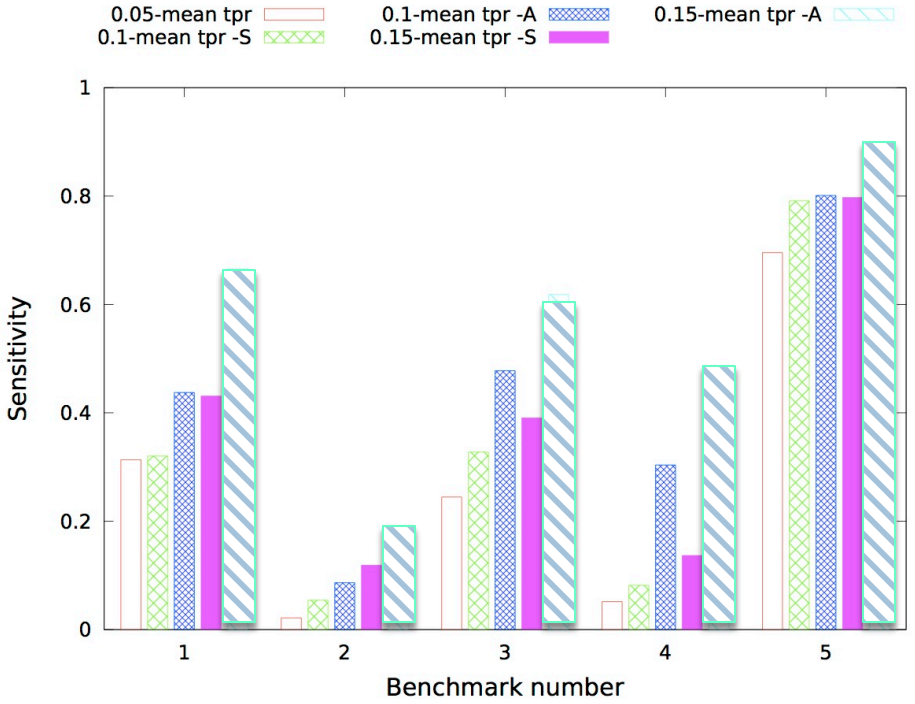




Active probing

- Experimental setting: initial select 5% of probes at random, the add probes by x increment
- Compare at constant probe cost
- Active probing increases performance significantly for all criteria





Cost sensitive learning

- FN (missing a failure) is more dangerous than FP (predicting a failure incorrectly) from system management view
- Cost sensitive optimization

$$min \ \|X\|_{tr} + c \sum_{ij \in S} C(Y_{ij}) \max(0, 1 - X_{ij} Y_{ij})$$

$$C(Y_{ij}) = C_+, \text{ if } Y_{ij} == 1; \ C(Y_{ij}) = C_- \text{ if } Y_{ij} == -1$$

- The C+/C- ratio parameterize the tradeoff between sensitivity and precision

Conclusion

- Conclusion
  - Real world large scale data indicates that large proportion of faults are caused by a small number of hidden variables
  - Collaborative prediction is a promising strategy for distributed monitoring in large scale system
- Future work
  - Extract hidden causes uncovered by CP
    - Bi-LDA instead of MMMF
  - Improving prediction with temporal models
    - Very different from Recommendation Systems: users don't rate twice the same product

More details

- CCGRID paper *Distributed Monitoring with Collaborative Prediction*

  http://hal.inria.fr/hal-00673148