

Rule mining2

Oleg Shpynov

JetBrains Biolabs

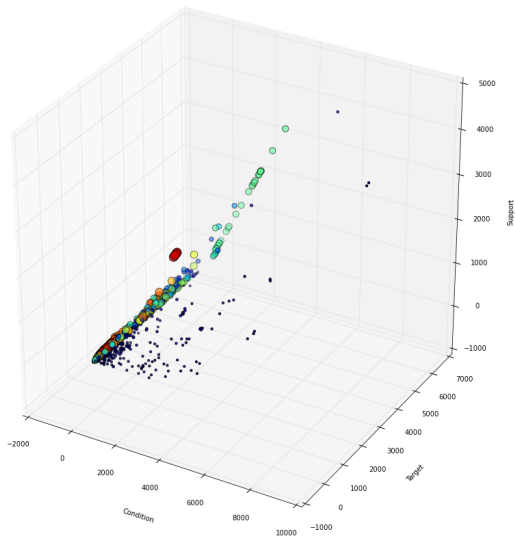
September 30, 2015

Rule mining VS ML

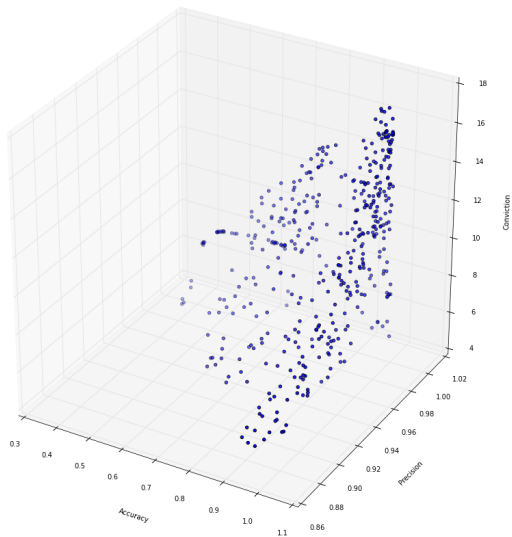
- *Conviction* still looks fine
- We can forbid predicates negations
- Rules can still have unobservable variables
- Human readable interpretation instead of ML

Rules are classifier?

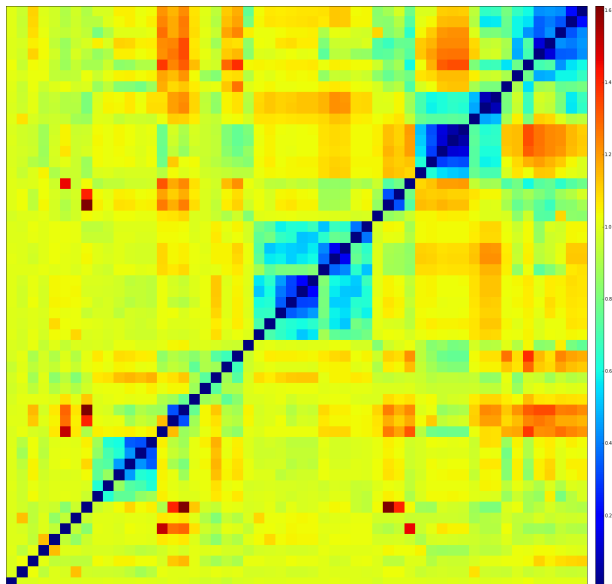
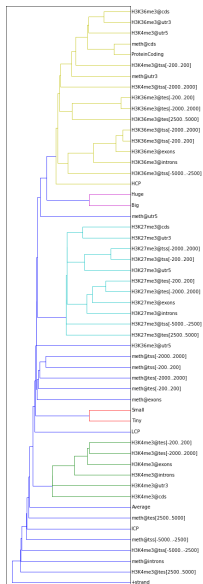
Conviction



Accuracy Precision



GSE47819



Collocations vs correlation

Colocations of ≥ 0.5 show correlation ≈ 0.7 .

Let's use threshold correlation 0.5.

Example:

Filtered by correlation with *H3K4me2@exons* on GSE26320 GM12878:
H3K4me3@introns, *H3K4me3@exons*, *H3K4me2@introns*,
H3K4me2@tes2000, *H3K27ac@exons*, *H3K9Ac@introns*,
H3K4me3@tes2000, *H3K9Ac@tes200*, *H3K9Ac@exons*,
H3K4me3@tes200, *H3K4me2@tes200*

Example for rules within GSE26320 with *conviction* ≥ 15

cell	condition	target	conviction	accuracy	precision	AdaBoost	AdaBoost precision	Decision Tree	Decision Tree precision	Random Forest	Random Forest precision
GM12878	H3K4me2@utr3 AND H3K9Ac@introns	H3K4me2@cds	16.220816	0.9553	1	0.968846	0.867526	0.963647	0.973384	0.959737	0.978520
GM12878	H3K27ac@tss200 AND H3K4me3@tss200...	H3K4me2@exons	15.399389	0.9591	1	0.986660	0.895085	0.970740	0.710200	0.958609	0.888889
NHEK	NOT HCP AND NOT ICP AND NOT LCP	Small	15.825031	0.9908	1	0.991214	0.996517	0.810052	0.519168	0.795865	0.000000
NHEK	H3K4me2@tss-200 AND H3K4me2@tss2000...	H3K4me2@exons	15.822808	0.9714	1	0.986660	0.864322	0.973037	0.713781	0.971788	0.916996
HUVEC	H3K27ac@introns AND H3K4me2@utr3	H3K4me2@cds	15.537027	0.9732	1	0.982911	0.882653	0.976181	0.891967	0.970256	0.985714

ML

Example RULE:

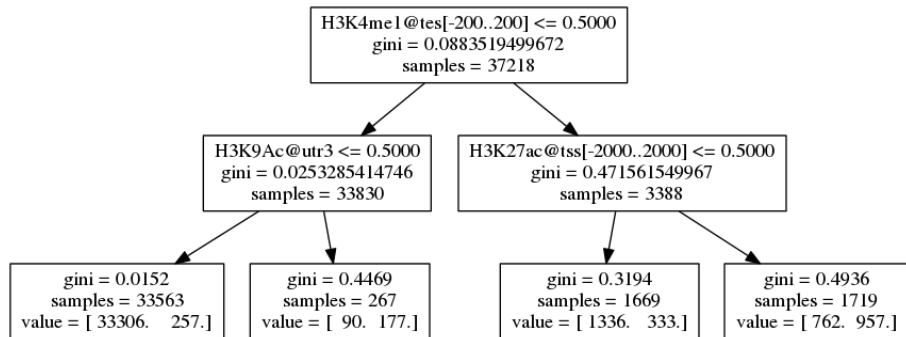
$H3K27ac@tss2000 \wedge Small \rightarrow H3K27ac@tes2000$

algorithm	accuracy	precision
rule mining	0.9673	0.9652174
AdaBoost	0.9722714815411897	0.75026910656620016
Decision Tree	0.95921328389488958	0.56051502145922749
Random Forest	0.95731903917459293	0.81333333333333335

Example RULE:

$H3K27ac@tss2000 \wedge Small \rightarrow H3K27ac@tes2000$

Decision tree



References

- Rules vs ML IPython notebook
- RulesVisualizer - class for rules visualization