Over the Horizon: ML+SBSE = What?

tim@menzies.us
wvu

30 - 31 January 2013
The 24th CREST Open Workshop, UCL, London, UK
Machine Learning and Search Based Software Engineering (ML & SBSE)

Data miners can find signals in SE artifacts

- apps store data
- recommender systems
- emails ➔human networks
- process data ➔ project effort
- process models ➔ project changes
- bug databases ➔ defect prediction
- execution traces ➔ normal usage patterns
- operating system logs ➔ software power consumption
- natural language requirements ➔ connections between program components
- Etc

So what's next?

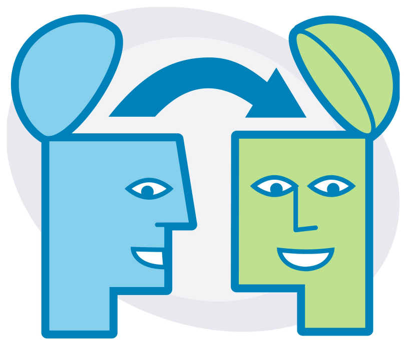
Better algorithms ≠ better mining (yet ...)

- Dejaeger, K.; Verbeke, W.; Martens, D.; Baesens, B.; , "Data Mining Techniques for Software **Effort Estimation**: A Comparative Study," *Software Engineering, IEEE Transactions, doi: 10.1109/TSE. 2011*
  - Simple, understandable techniques like Ordinary least squares regressions with log transformation of attributes and target perform as well as (or better than) nonlinear techniques.
- Hall, T.; Beecham, S.; Bowes, D.; Gray, D.; Counsell, S.; , "A Systematic Review of **Fault Prediction** Performance in Software Engineering," *Software Engineering, IEEE Transactions,* doi: 10.1109/TSE.2011.103
  - Support Vector Machine (SVM) perform less well.
  - Models based on C4.5 seem to under-perform if they use imbalanced data.
  - Models performing comparatively well are relatively simple techniques that are easy to use and well understood.. E.g. Naïve Bayes and Logistic regression

What matters: sharing

Tutorials : ICSE'13	Workshops : ICSE'13

- Data Science for SE
  - How to share data and models



- SE in the Age of Data Privacy
  - If you do want to share data ...
  - … how to privatize it

- DAPSE'13:
  - If you can't share data, or models….
  - At least, share our analysis methods
  - Data analysis patterns in SE

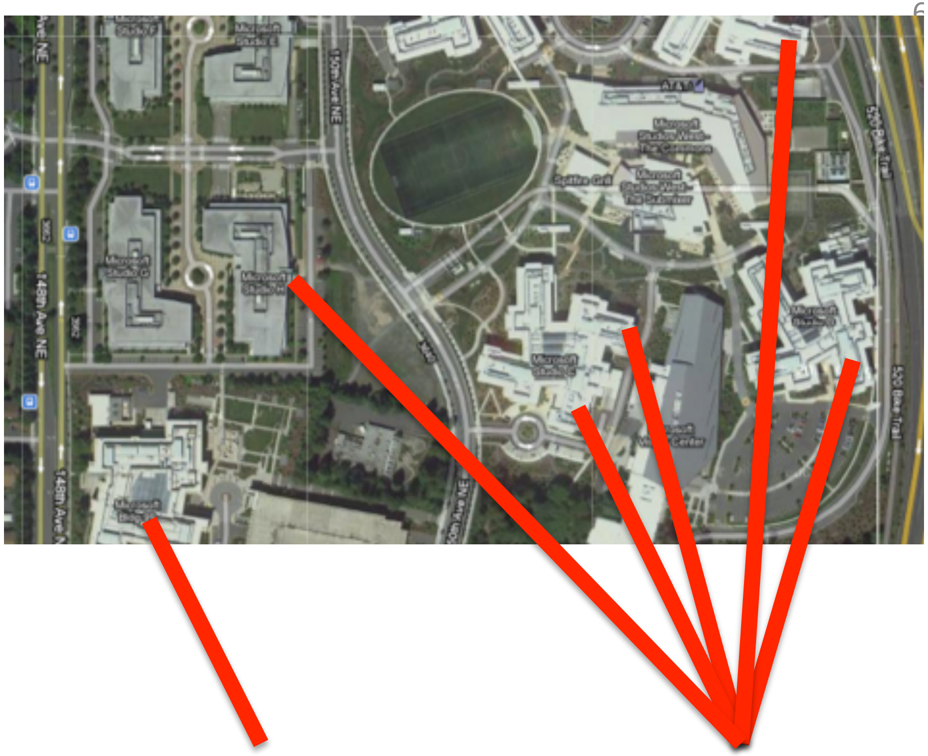
- RAISE'13:
  - realizing AI synergies in SE
  - State of the art (archival)
  - Over the horizon (short, non-archival)

What else matters

- Tools, availability
  - Simpler, the better
- Not algorithms, but users:
  - CS discounts "user effects"
  - The notion of 'user' cannot be precisely defined and therefore has no place in CS and SE -- Edsger Dijkstra, ICSE'4, 1979
- Not predictive power
  - Need "insight"and "engagement"

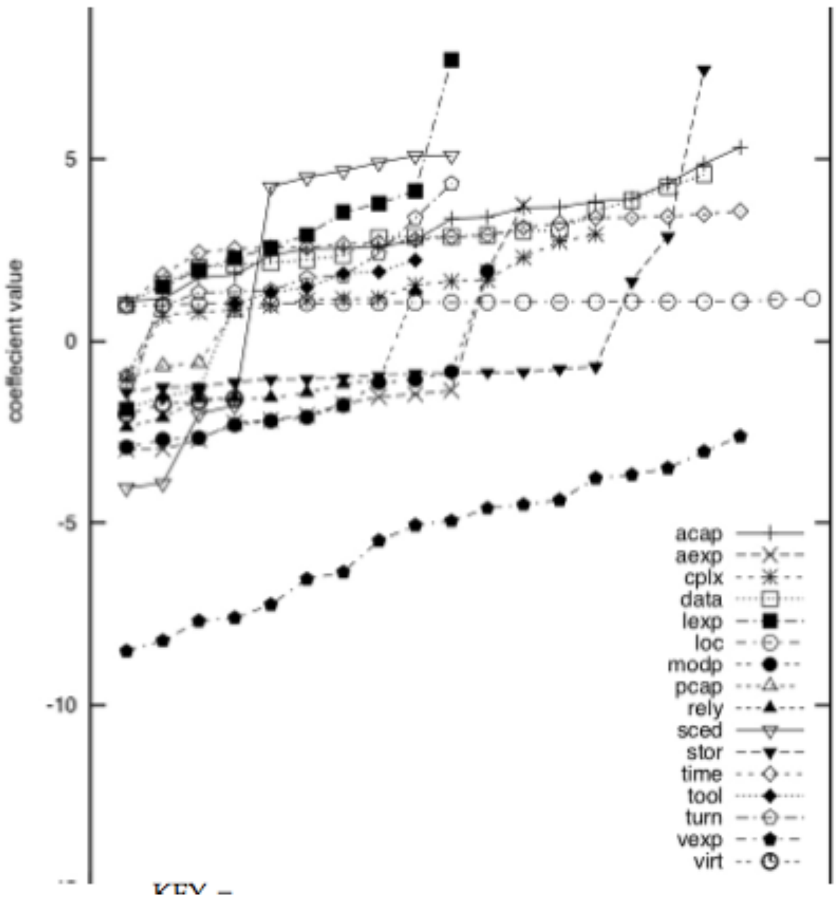
No one thing matters

SE = locality effects



Microsoft research, Redmond, Building 99

Other studios, many other projects



Localism:
Not general models

ref	cbo	rfc	lcom	dit	noc	wmc	#proj	size	type
[13]	+	+	+	-	-	+	6	95-201 classes	6 versions of rhino (java)
[14]	+	+	+	-	-	+	12	86 classess (3-12kloc)	student
[15]	+	+	-				1	1700 classes (110kloc)	commercial telecom
[16]	+	+	-	+	+	+	8	113 classes	student
[17]	+	+	-	+	+	+	8	114 classes	student
[18]	+	+	+	+	-		1	83 classes	commercial: lalo (c++)
[19]				+	+		1	32 classes	commercial: telecom c++
[20]				+	-		1	42-69 classes	commercial java word proc.
[21]	+	-	-	-	-	-	1	85 classes	telecom c++
[22]	-	+		-	-	+	3	92 classes	3 c++ subsystems,commercial
[23]	+	+	+	-	+	+	1	123 classes (34kloc)	java commercial
[24]	+			+		+	1	706 classes	commercial c++ and java
[25]	+	+	+	-	+	+	1	145 classes	kc1-nasa
[26]	+	+	+	+	-	+	1	3677 classes	open source:mozilla
[27]	+	+	+			+	1	?	java (sap) commercial
[28]	+	+	+	+	+	+	3	?	eclipse 2.0, 2.1, 3.0
[29]	-	+	+	-	-	+	8	113 classes	student
[30]		+	+	+	+		2	64 classes	sales and cd-selection system
[31]		-		-	-	-	1	3344 modules (2mloc)	commercial telecom c++
[32]	+	+	+	-	-	+	5	395 classes	commercial telecom c++
[33]	+	+		-	-	+	1	1412 classes	open source:jdt
[34]	+	+		-	-	+	2	9713 classes	eclipse 2.0, 2.1
[35]	+	+	-	-	-	+	1	145 classes	kc1-nasa
[36]				+	-		1	145 classes	commercial java xml editor
[37]	-	-	-	-	-	-	1	174 classes	commercial telecom c++
[38]	-					-	0	50 classes	student
[39]	+	+	-	-	-	+	1	145 classes	kc1-nasa
[40]		+		+	+		2	294 classes	commercial c++
total +	18	20	11	11	8	17			
total -	4	3	7	14	16	4			
Total percents:	"*" denotes majority conclusion in each column								
+	* 64%	* 71%	* 39%	39%	29%	* 61%			

KEY:

- Strong consensus (over 2/3rds)
- Some consensus (less than 2/3rds)
- Weak consensus (about half)

Goal: general methods for building local models

	effort		defect						
cluster	NasaCoc	china	lucene2.4	xalan2.6	jedit4.0	velocity1.6	synapse1.2	tomcat	xerces1.4
global	kloc=1	afp=1	rfc=2	loc=1	rfc=2	cam=7	amc=1	loc=2	cbo=1
C0									
C1	rely=n	added=4	amc=7	amc=1	ic=7	noc=1	dit=4	cbm=1	dit=1
C2	prec=h	deleted=1	ca=1	cam=2	noc=1	dam=1 or 5		dam=1	dam=1
C3		deleted=1	dam=5	cam=3	amc=6	avg_cc=4		noc=1	ca=1 or 7
C4			mfa=1	dit=2 or 4	noc=1	moa=1		rfc=5	**cbo=1**
C5			moa=1	**loc=1**				lcom3=5	
C6				**loc =1 or 2**				max_cc=1	
C7				moa=1				cbm=1	

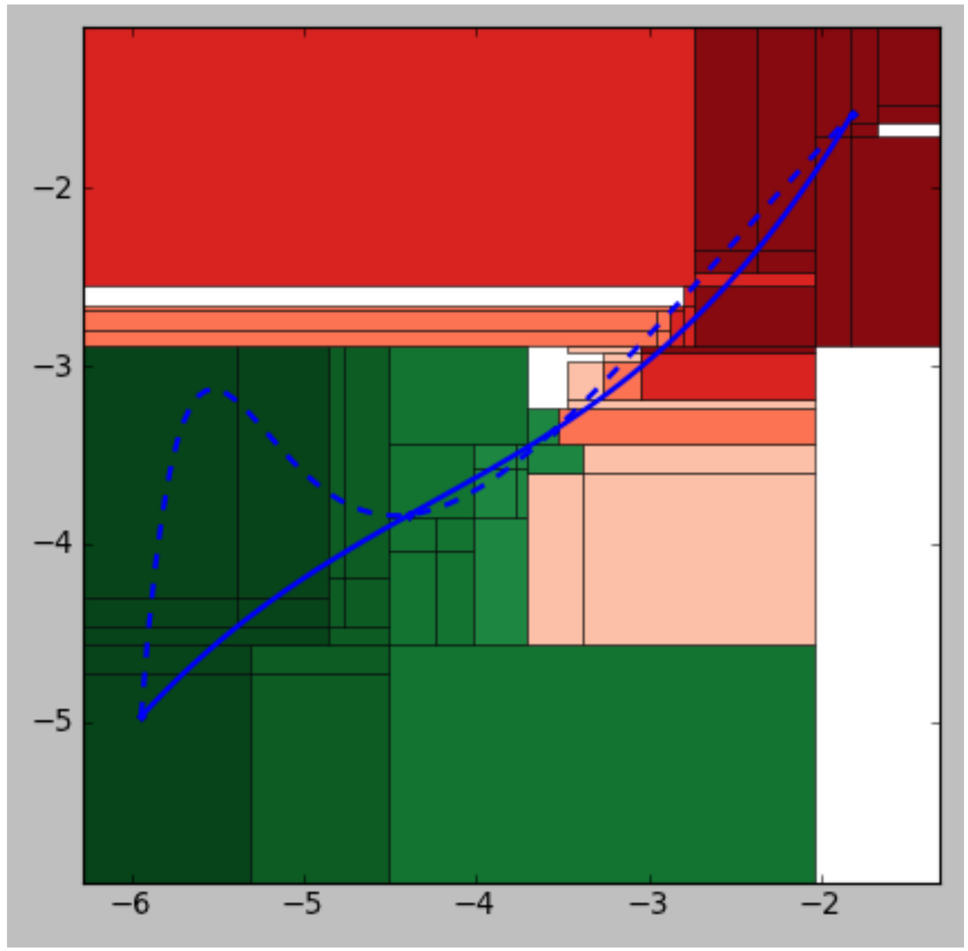
Local models:

- very simple,
- very different to each other

"Discussion Mining" : guiding the walk across the space of local models

- Assumption #1:
  - Principle of rationality [Popper, Newell] ; "If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action."

- Assumption #2:
  - Agents walk clusters of data to make decisions
  - Typology of the data = space of possible decisions



1. <u>Landscape mining:</u> **find local regions &score them**

2. <u>Decision mining:</u> **find intra-region deltas**

3. <u>Discussion mining:</u> **help a team walk the regions**

A formal model for "engagement"

- Christian Bird, knowledge engineering,
  - Microsoft Research, Redmond
  - Assesses learners not by correlation, accuracy, recall, etc
  - But by "engagement"

A successful "Bird" session:

- Knowledge engineers enter with sample data
- Users take over the spreadsheet
- Run many ad hoc queries
- In such meetings, users often…
  - demolish the model
  - offer more data
  - demand you come back next week with something better

Over the Horizon: ML+SBSE = Discussion mining

yesterday		*today*		*tomorrow*		*future*
0. algorithm mining	→	1. landscape mining	→	2. decision mining	→	3. discussion mining

Q: why call it mining?

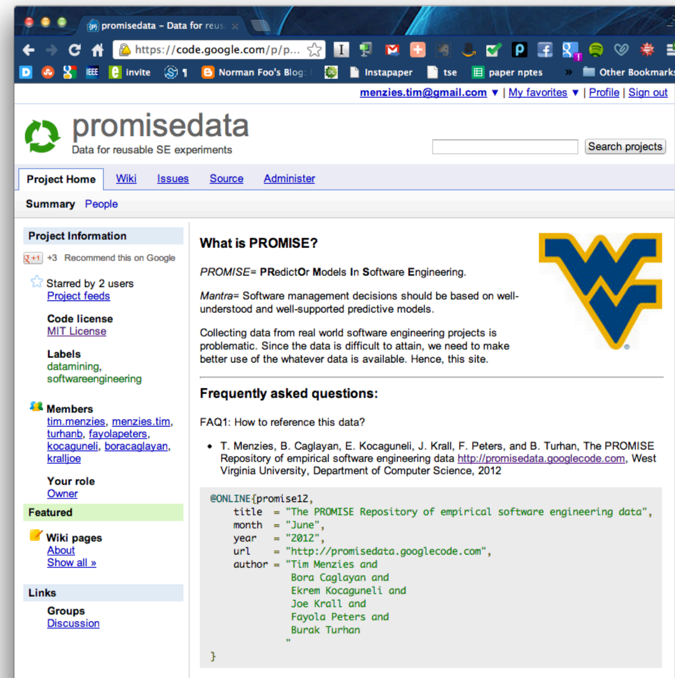
- A1: because all the primitives for the above are in the data mining literature
  - So we know how to get from here to there
- A2: because data mining scales

Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
     • E.g. Rutherford scattering
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Towards a science of localism

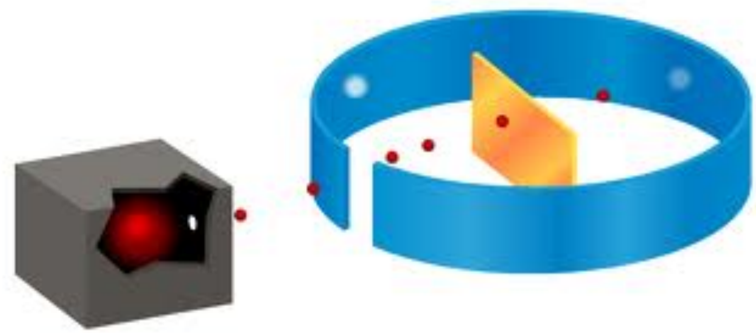
1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
     • E.g. Rutherford scattering
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

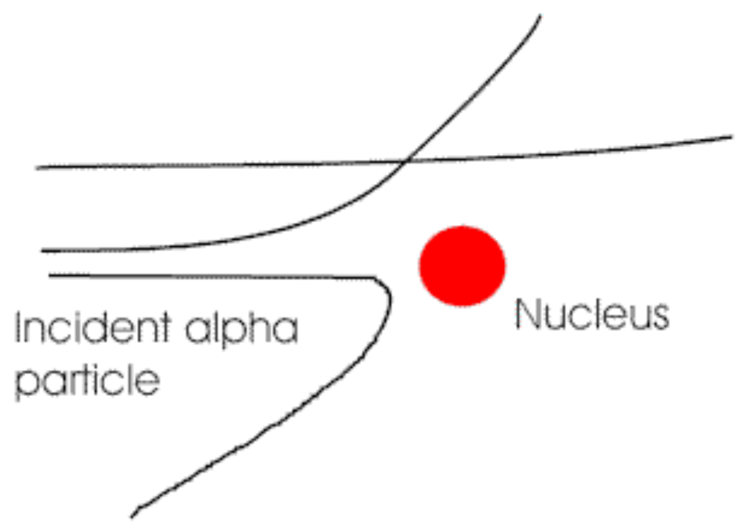


STOP PRESS: ISBSG samplers now in PROMSE

Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
      • E.g. Rutherford scattering
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students





Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
     • E.g. Rutherford scattering
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Christian Bird:
- The engagement pattern.
- users don't want correct models
- They want to correct their own models

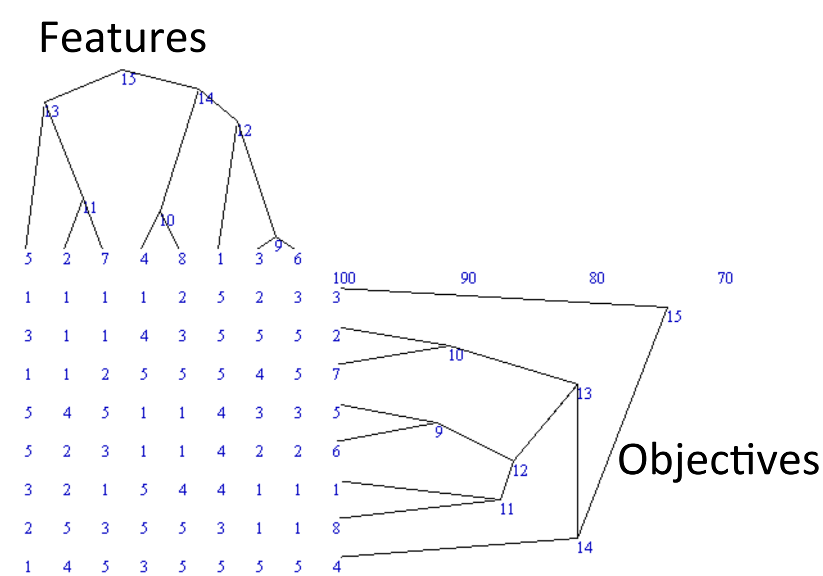
Towards a science of localism

1. Data
   - that a community can share
2. A spark
   - a (possibly) repeatable effect that questions accepted thinking
     - E.g. the photo-electric effect
3. Maths
   - E.g. a data structure
4. Synthesis:
   - N olde things are really 1 thing
5. Baselines
   - Tools a community can access
   - Results that scream for extension, improvement
6. Big data:
   - scalability, stochastics
7. Funding, industrial partners, grad students

F = features = Observable+ | Controllable+
O = objectives = O_1, O_2, O_3,.. = score(Features)

Eg = example = <F,0>

C = bicluster of examples, clustered by F and O
- Each cell has N examples

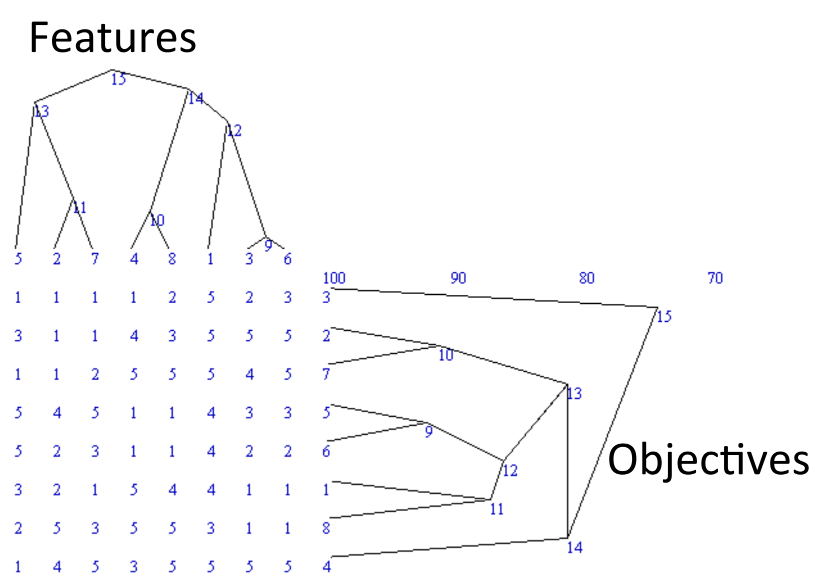


Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
      • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Menzies & Shepperd, EMSE 2012

- Special issue on conclusion instability
- Of course solutions are not stable- they specific to each cell in the bicluster.



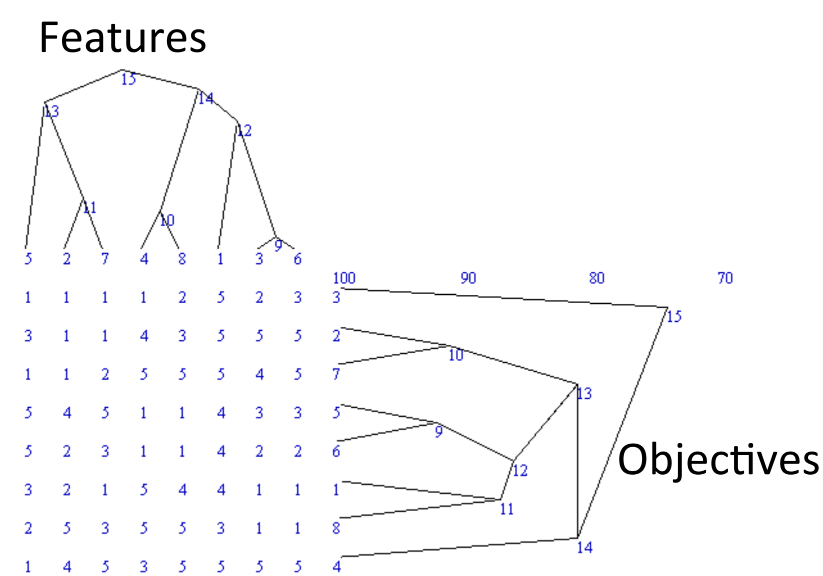
Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
     • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Applications to MOEA

Krall & Menzies (in progress) :

- Cluster on objectives using recursive Fastmap
- At each level, check dominance on N items from left& right branch
  - Don't recurse on dominated branches
- Selects examples in clusters on Pareto frontier



20 to 50 times fewer objective evaluations

- Find a large variance dimension in O(2N) comparisons
  - W= any point;
  - West= furthest of W;
  - East= furthest of West
  - c = dist(X,Y)
- Each example X:
  - a = dist(X ,West);
  - b = dist(X, East)
  - Falls $x=(a^2+c^2-b^2)/2c$
  - And at $y = \mathrm{sqrt}(x^2 – a^2)$
- If X or Y dominates, don't recurse on the other
  - Finds clusters on the Pareto frontier in linear time
- Split on median points, & recurse
  - Stop when leaf has less than, say, 30 items
  - These are parents of next generation

```
class Moo(BinaryTree):
  def __init__(i,t):
    BinaryTree.__init__(i)
    i.table = t; i.abort=False
    i.east, i.west, i.c, i.x = None,None,None,None

  def project(i,rows,mid):
    "Uses the O(2N) Fastmap heuristic."
    w    = one(rows) # any row, selected at random
    west = w.furthest()
    east = west.furthest()
    c    = west.distance(east)
    for row in rows:
       a    = row.distance(west)
       b    = row.distance(east)
       row.x= (a**2 + c**2 - b**2) / (2*c+0.00001)
    rows = sorted(rows, key= lambda row: row.x)
    i.west, i.east, i.c, i.x = \
      west,   east,   c, rows[mid].x
    return rows[:mid],rows[mid:]

  def divide(i,abort=False,min=30):
    def aFew(rows) :
      all = map(lambda r: r.cells,rows)
      return i.table.clone(some(all,The.alpha))
    n = len(i.table.rows);
    m = n/2
    i.abort = The.allowDomination and abort
    if not i.abort and n >= min :
      wests,easts  = i.project(i.table.rows,m)
      if i.west != i.east:
        i.lhs = Moo(aFew(wests))
        i.rhs = Moo(aFew(easts))
        i.lhs.divide(abort=i.east.dominates(i.west),
                     min  =min)
        i.rhs.divide(abort=i.west.dominates(i.east),
                     min  =min)
    return i
```

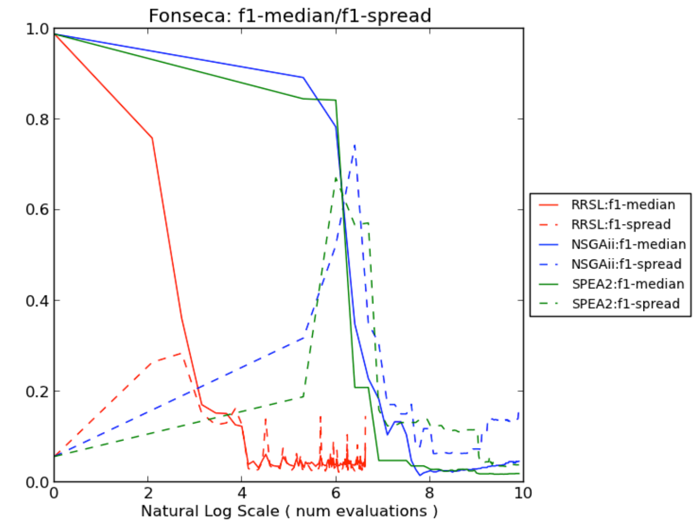
20 to 50 times fewer objective evaluations

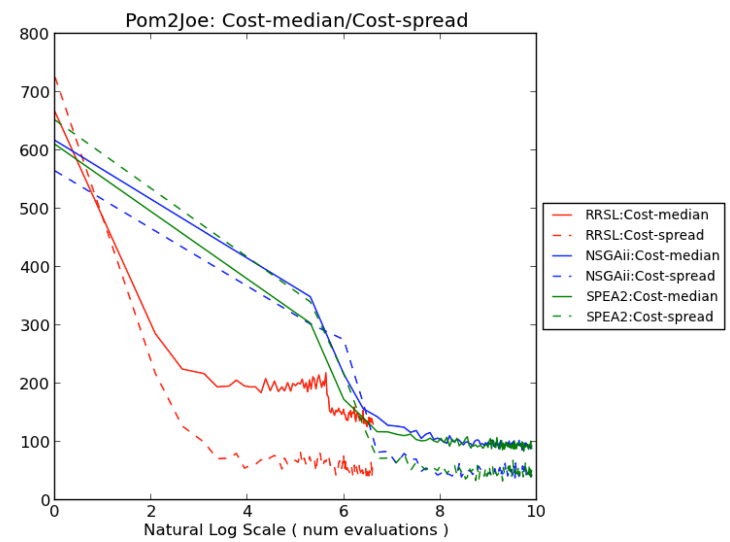
- Find a large variance dimension in O(2N) comparisons
  - W= any point;
  - West= furthest of W;
  - East= furthest of West
  - c = dist(X,Y)

- Each example X:
  - a = dist(X ,West);
  - b = dist(X, East)
  - Falls $x=(a^2+c^2-b^2)/2c$
  - And at $y = \mathrm{sqrt}(x^2 - a^2)$

- If X or Y dominates, don't recurse on the other
  - Finds clusters on the Pareto frontier in linear time

- Split on median points, & recurse
  - Stop when leaf has less than, say, 30 items
  - These are parents of next generation



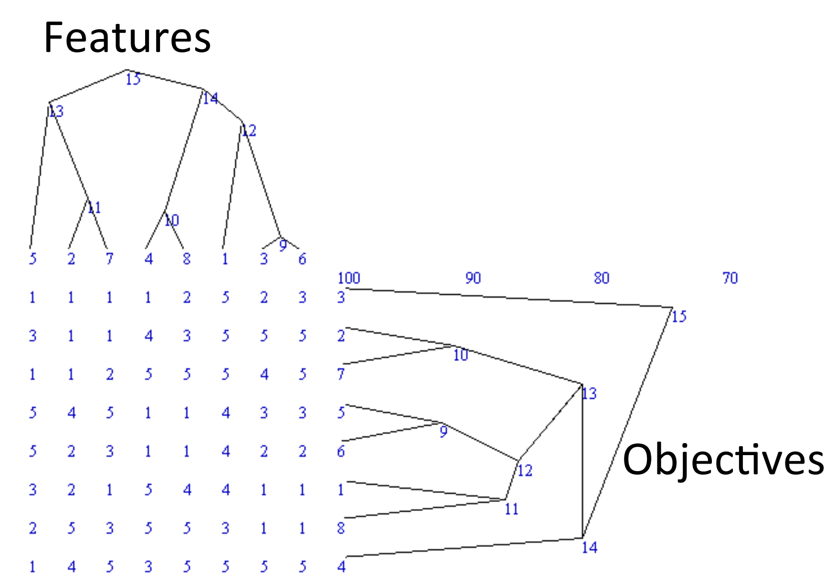


Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
      • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Transfer learning

- Domain = <examples, distribution>
- One data set may have many domains
  - i.e. . multiple clusters of features & objectives
- Kocaguneli, & Menzies EMSE'11
- TEAK select best cluster for cross-company learning
  - Cross company== within company learning



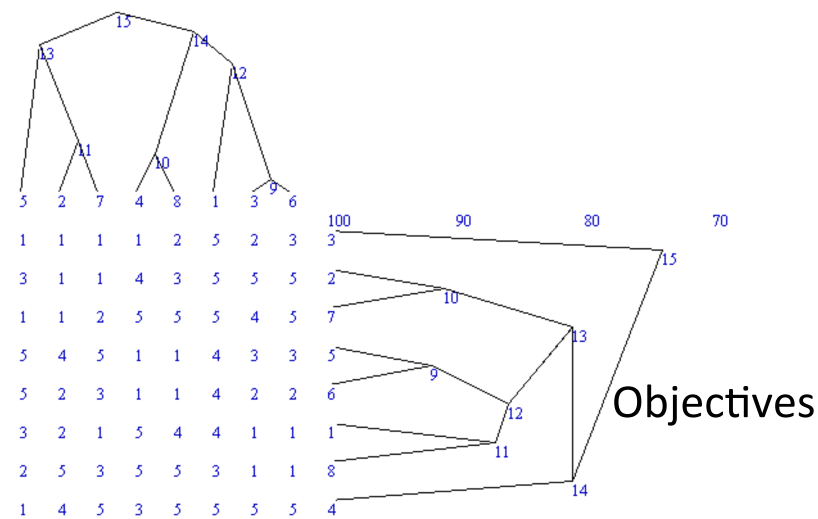
Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
     • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Kocaguneli & Menzies et al, TSE'12 (March)

- TEAK : recursively cluster on features
- Kill clusters with large objective variance
- Recursively cluster the survivors
- kNN, select k via sub-tree inspection
- Generated only a few clusters per data set

Features



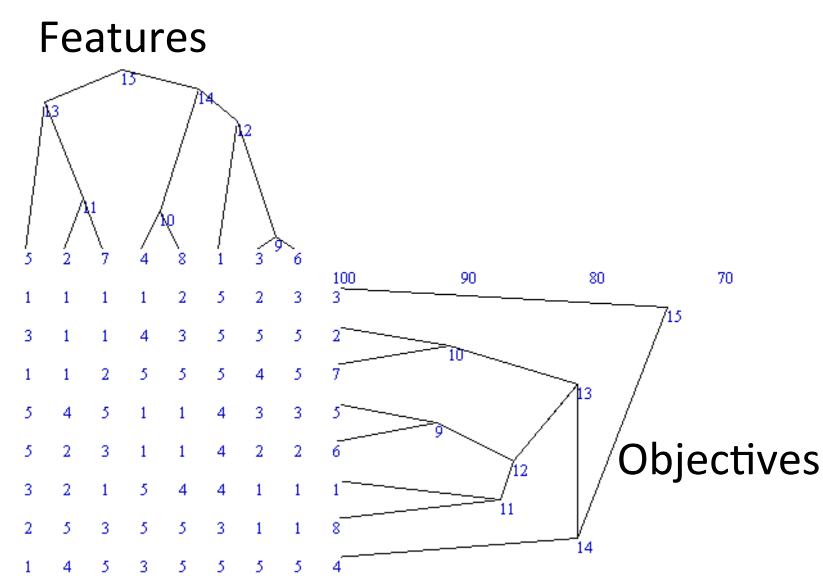
Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
     • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Active learning

Kocaguneli & Menzies et al, TSE'13 (pre-print)

- Grow clusters via reverse nearest neighbor counts
- Stop at N+3 if no improvement over N examples
    - Evaluated via k=1 NN on a holdout set
- Finds the most informative next question

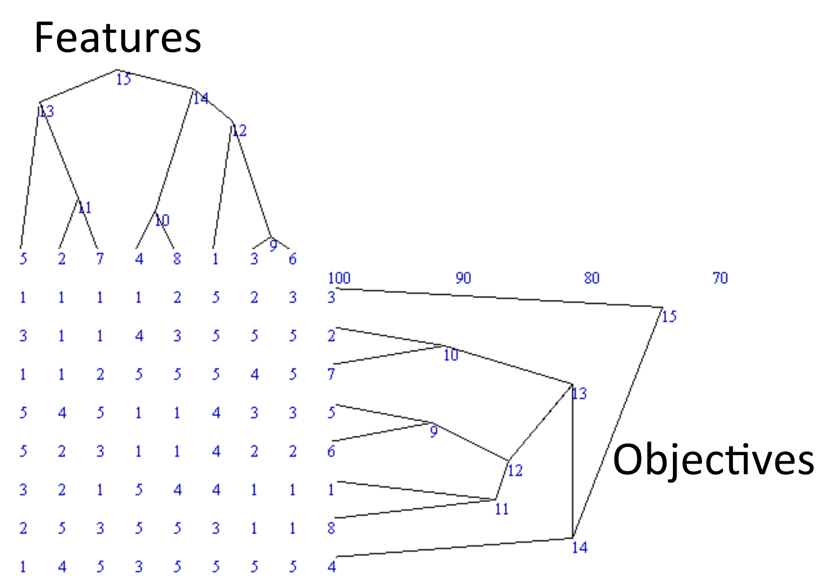


Towards a science of localism

1. Data
   - that a community can share
2. A spark
   - a (possibly) repeatable effect that questions accepted thinking
     - E.g. the photo-electric effect
3. Maths
   - E.g. a data structure
4. Synthesis:
   - N olde things are really 1 thing
5. Baselines
   - Tools a community can access
   - Results that scream for extension, improvement
6. Big data:
   - scalability, stochastics
7. Funding, industrial partners, grad students

Peters & Menzies & Zhang et al., TSE'13 (pre-print)

- Find divisions of data that separate classes
- To effectively privatize data…
  - Find ranges that drive you to different classes
  - Remove examples without those ranges
  - Mutate survivors, do not cross boundaries

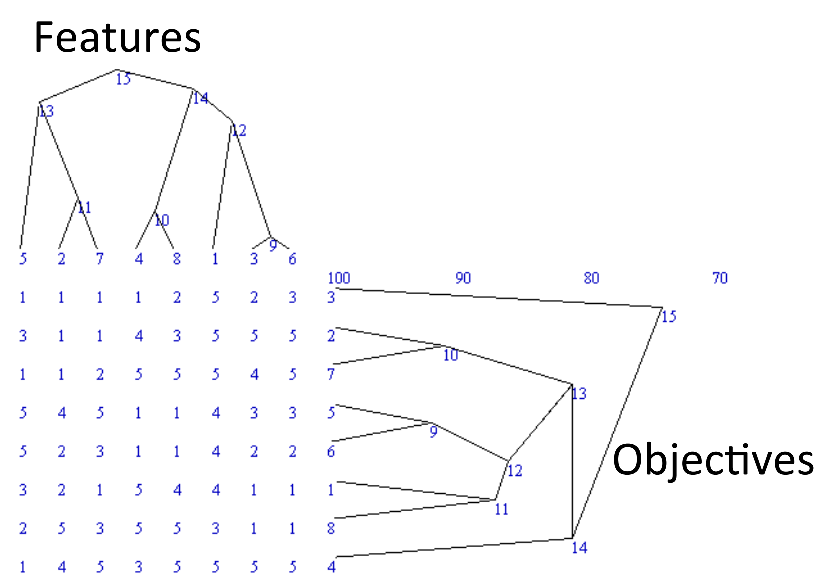


Towards a science of localism

1. Data
   - that a community can share
2. A spark
   - a (possibly) repeatable effect that questions accepted thinking
     - E.g. the photo-electric effect
3. Maths
   - E.g. a data structure
4. Synthesis:
   - N olde things are really 1 thing
5. Baselines
   - Tools a community can access
   - Results that scream for extension, improvement
6. Big data:
   - scalability, stochastics
7. Funding, industrial partners, grad students

Menzies & Butcher & Marcus et al., TSE'13 (pre-print)

- Cluster on features using recursive Fastmap
- Combine sibling leaves if their density not too low
- Train from cluster you most "envy"; test on you
- Envy-based models out-perform global models



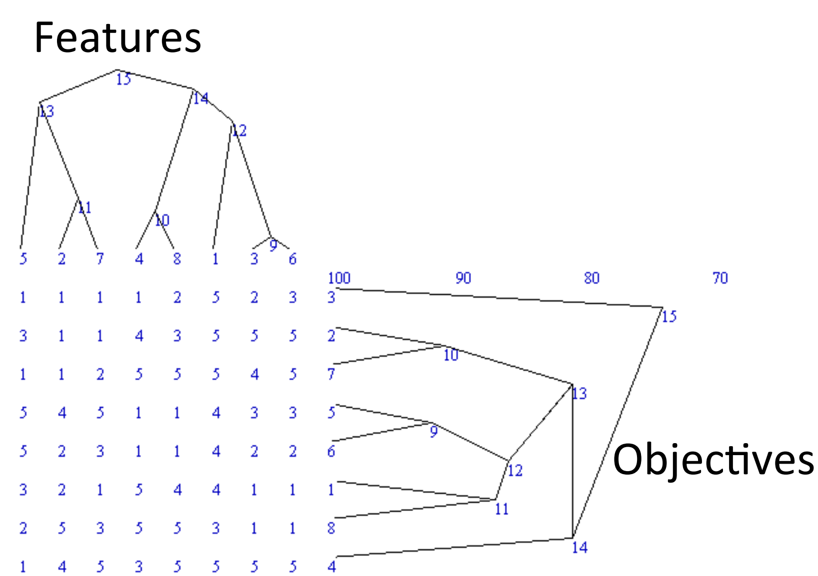
Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
      • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

Baseline results: See above. Got a better clusterer?
Source: http://unbox.org/things/var/timm/13/sbs/rrsl.py
Tutorial: http://unbox.org/things/var/timm/13/sbs/rrsl.pdf

Open issues:

- Can we track engagement.
- The acid test for structured reviews.
- Is data mining and MOEA really different.



Towards a science of localism

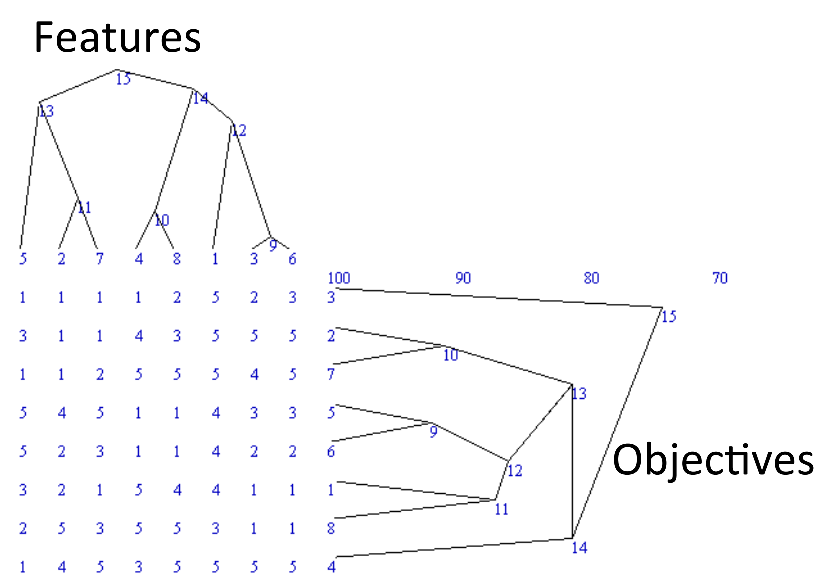
1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
      • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. **Big data:**
   – **scalability, stochastics**
7. Funding, industrial partners, grad students

Stochastic recursive Fastmap: O(N.logN)

- Can't explore all data?
- Just use a random sample

On-line learning:

- Anomalies = examples that fall outside leaf clusters
- Re-learn, but just on sub-trees with many anomalies



Towards a science of localism

1. Data
   – that a community can share
2. A spark
   – a (possibly) repeatable effect that questions accepted thinking
      • E.g. the photo-electric effect
3. Maths
   – E.g. a data structure
4. Synthesis:
   – N olde things are really 1 thing
5. Baselines
   – Tools a community can access
   – Results that scream for extension, improvement
6. Big data:
   – scalability, stochastics
7. Funding, industrial partners, grad students

New 4 year project:

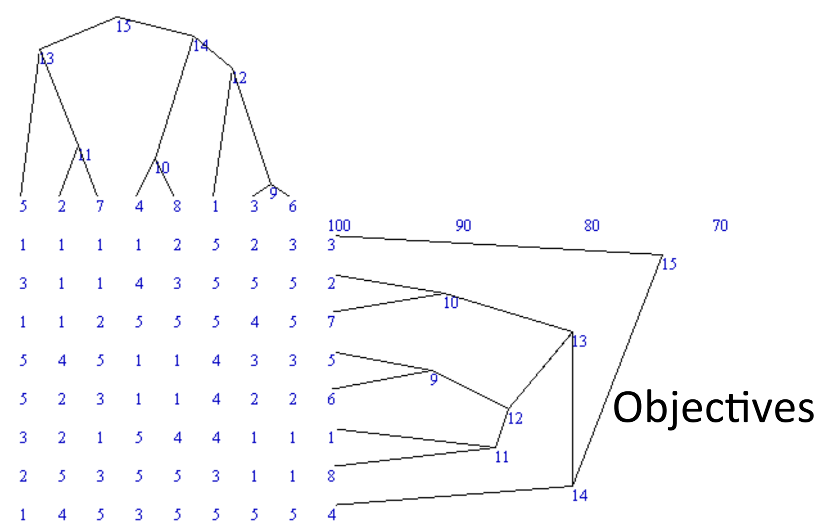
- WVU + Fraunhofer , Maryland (director = Forrest Shull)

Transfer learning on data from

- PROMISE (open source)
- Fraunhofer (proprietary data)

Core data structure? See below…

Features



Over the Horizon: ML+SBSE = Discussion mining

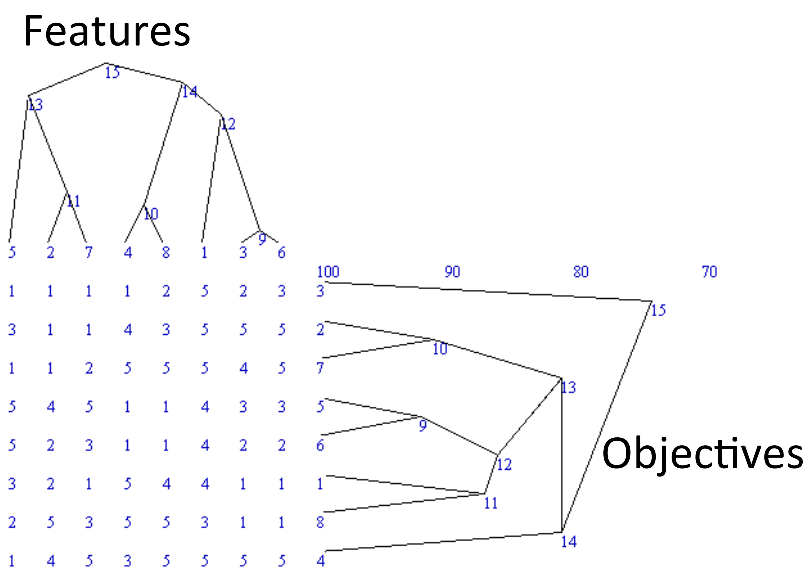
yesterday: 0. algorithm mining → *today*: 1. landscape mining → *tomorrow*: 2. decision mining → *future*: 3. discussion mining

Q: why call it mining?

- A1: because all the primitives for the above are in the data mining literature
  - So we know how to get from here to there
- A2: because data mining scales

Beyond Data Mining, T. Menzies, IEEE Software, 2013, to appear

Care to join a new cult?



ML+SBSE = exploring clusters of features and objectives

Questions? Comments?

Backup slides

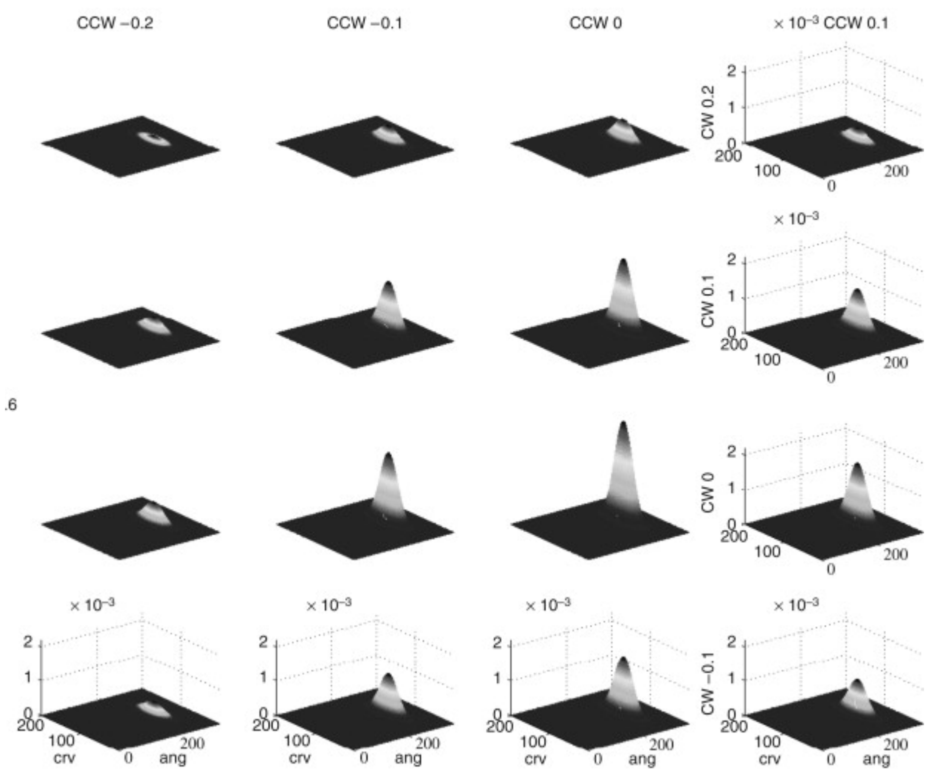
Local Regions

Bettenburg, Hassan et al. MSR'12
Menzies et al. TSE'12 (pre-print)

- "Domains" in transfer learning
  - <Examples, Distribution>
  - One data set has N domains
  - Each with different "best" policies
    - Posnett et al. ASE'11
- If you localize learning
  - Do better than learning over all domains
- Domains != manual stratifications
  - Find domains automatically
    - Cluster, then learn

application domain	avionics	fixed ground	missile	mobile ground	shipboard	unmanned airbor	unmanned space	total
business systems		6		4	2			12
command & control	1	41		16	35			93
communications	4	77			17		2	100
controls & display	8	6		2	5			21
executive		4			3			7
information assurance		1						1
infrastructure		11			23			34
maintain & diagnose	1				5			6
mission management	42	2	3	2		1		50
mission planning	1	17						18
modeling & simulation		1						1
process control		3		6	1			10
scientific systems					3			3
sensor processing	12	15			18			45
simulation & modeling		19			17			36
spacecraft BUS							9	9
spacecraft payload							16	16
test & evaluation		2			2			4
tool & tool systems		6	1					7
training				2	6			8
weaps delivery & control	11		19		9			39
totals	80	211	23	32	146	1	27	520

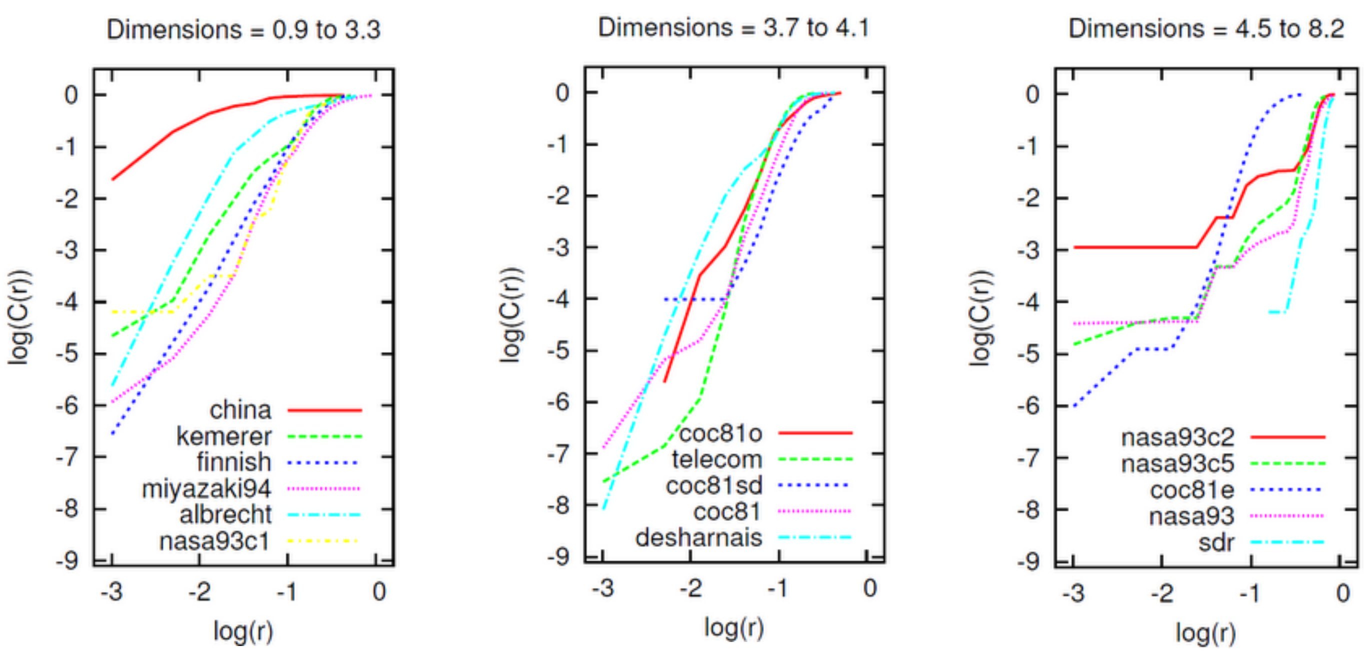
Low Intrinsic Dimensionality



- Ball1.radius = r
- Ball2.radius = 2r
  - Holds more points than ball1
- Data1 :
  - Spreads in all dimensions
- Data2 : data is “squashed”
  - Does not spread in all directions
- Ball2 in Data2 has fewer points
  - than Ball2 in Data1

To measure intrinsic dimensionality, use "correlation dimension" [1]

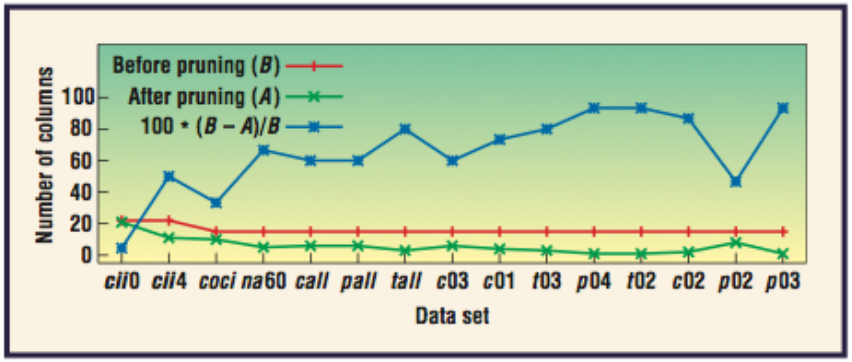
- n = # points
- c = #points separated by no more than r
- C(r) = 2/(n*(n-1)) * c
- Cd = max slope of log(C(r)) vs log(r)
- CD for effort estimation data:

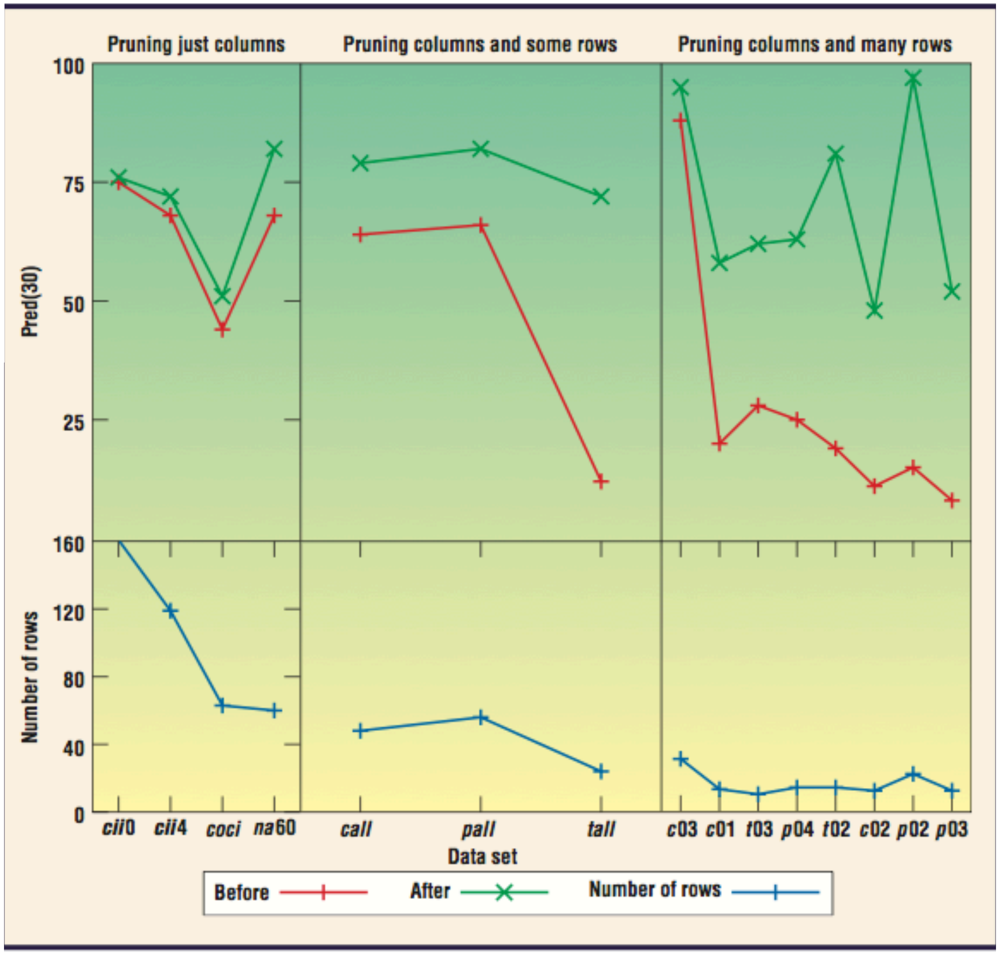


1. Elizaveta Levina, Peter J. Bickel, Maximum Likelihood Estimation of Intrinsic Dimension In Advances
2. *NIPS,* vol. 17. MIT Press, 2005

In practice

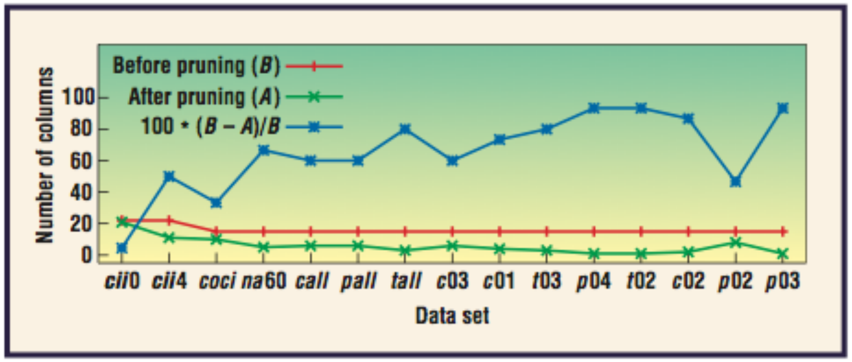
[Chen, Menzies, Boehm, IEEE Software 2005]

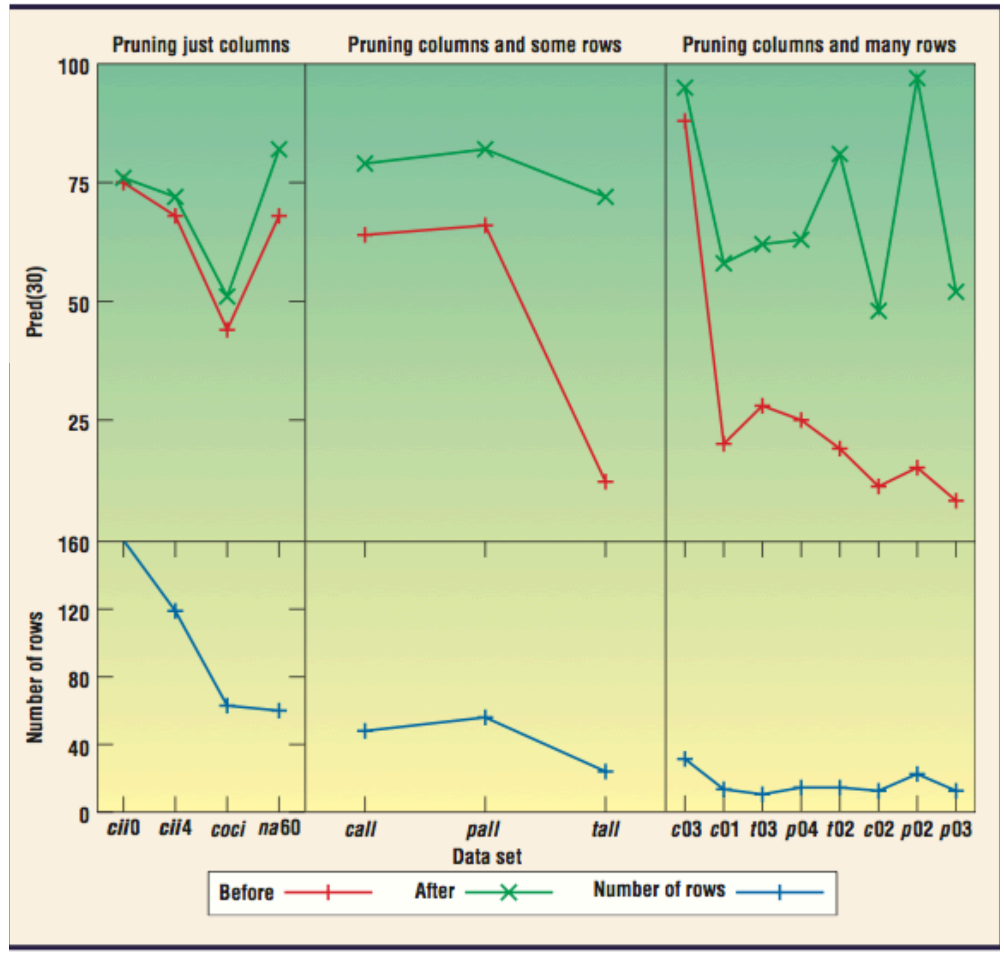




In practice (2)

[Kocaguneli, Menzies, Keung, Cok, 2013, TSE (pre-print)]





Wanted : a new idea

Old ones getting a little dull

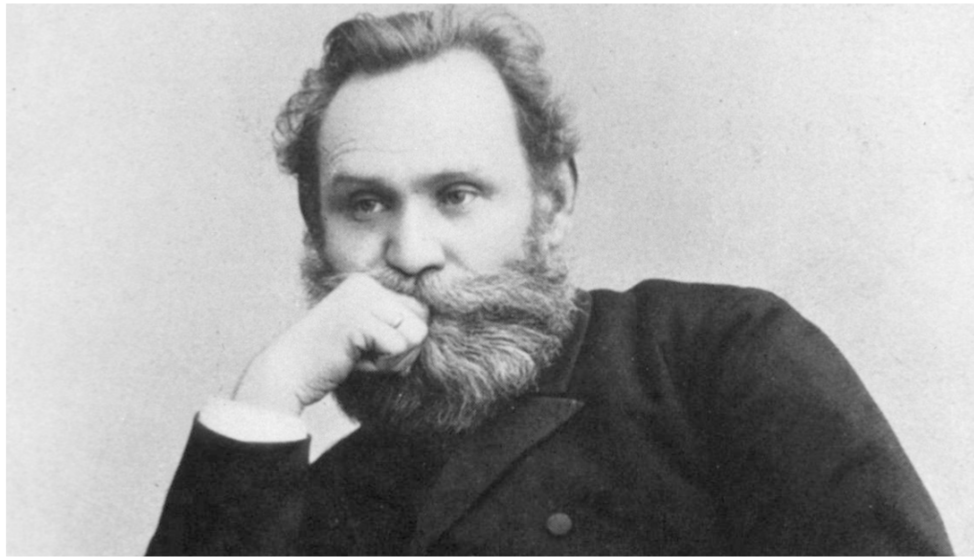
Algorithm mining = DULL

- Our results should be insights about data
  - Not trivia about (say) decision tree learning
- The thing that most predicts for performance is the data, not the algorithm

Table 1. Classification accuracies and sample standard deviations, averaged over 20 random training/test splits. "Bayes" is the Bayesian classifier with discretization and "Gauss" is the Bayesian classifier with Gaussian distributions. Superscripts denote confidence levels for the difference in accuracy between the Bayesian classifier and the corresponding algorithm, using a one-tailed paired t test: 1 is 99.5%, 2 is 99%, 3 is 97.5%, 4 is 95%, 5 is 90%, and 6 is below 90%.

Data Set	Bayes	Gauss	C4.5	PEBLS	CN2	Def.
Audiology	73.0 ± 6.1	73.0 ± 6.1^{6}	72.5 ± 5.8^{6}	75.8 ± 5.4^{3}	71.0 ± 5.1^{5}	21.3
Annealing	95.3 ± 1.2	84.3 ± 3.8^{1}	90.5 ± 2.2^{1}	98.8 ± 0.8^{1}	81.2 ± 5.4^{1}	76.4
Breast cancer	71.6 ± 4.7	71.3 ± 4.3^{6}	70.1 ± 6.8^{5}	65.6 ± 4.7^{1}	67.9 ± 7.1^{1}	67.6
Credit	84.5 ± 1.8	78.9 ± 2.5^{1}	85.9 ± 2.1^{3}	82.2 ± 1.9^{1}	82.0 ± 2.2^{1}	57.4
Chess endgames	88.0 ± 1.4	88.0 ± 1.4^{6}	99.2 ± 0.1^{1}	96.9 ± 0.7^{1}	98.1 ± 1.0^{1}	52.0
Diabetes	74.5 ± 2.4	75.2 ± 2.1^{6}	73.5 ± 3.4^{5}	71.1 ± 2.4^{1}	73.8 ± 2.7^{6}	66.0
Echocardiogram	69.1 ± 5.4	73.4 ± 4.9^{1}	64.7 ± 6.3^{1}	61.7 ± 6.4^{1}	68.2 ± 7.2^{6}	67.8
Glass	61.9 ± 6.2	50.6 ± 8.2^{1}	63.9 ± 8.7^{6}	62.0 ± 7.4^{6}	63.8 ± 5.5^{6}	31.7
Heart disease	81.9 ± 3.4	84.1 ± 2.8^{1}	77.5 ± 4.3^{1}	78.9 ± 4.0^{1}	79.7 ± 2.9^{3}	55.0
Hepatitis	85.3 ± 3.7	85.2 ± 4.0^{6}	79.2 ± 4.3^{1}	79.0 ± 5.1^{1}	80.3 ± 4.2^{1}	78.1
Horse colic	80.7 ± 3.7	79.3 ± 3.7^{1}	85.1 ± 3.8^{1}	75.7 ± 5.0^{1}	82.5 ± 4.2^{2}	63.6
Hypothyroid	97.5 ± 0.3	97.9 ± 0.4^{1}	99.1 ± 0.2^{1}	95.9 ± 0.7^{1}	98.8 ± 0.4^{1}	95.3
Iris	93.2 ± 3.5	93.9 ± 1.9^{6}	92.6 ± 2.7^{6}	93.5 ± 3.0^{6}	93.3 ± 3.6^{6}	26.5
Labor	91.3 ± 4.9	88.7 ± 10.6^{6}	78.1 ± 7.9^{1}	89.7 ± 5.0^{6}	82.1 ± 6.9^{1}	65.0
Lung cancer	46.8 ± 13.3	46.8 ± 13.3^{6}	40.9 ± 16.3^{5}	42.3 ± 17.3^{6}	38.6 ± 13.5^{3}	26.8
Liver disease	63.0 ± 3.3	54.8 ± 5.5^{1}	65.9 ± 4.4^{1}	61.3 ± 4.3^{6}	65.0 ± 3.8^{3}	58.1
LED	62.9 ± 6.5	62.9 ± 6.5^{6}	61.2 ± 8.4^{6}	55.3 ± 6.1^{1}	58.6 ± 8.1^{2}	8.0
Lymphography	81.6 ± 5.9	81.1 ± 4.8^{6}	75.0 ± 4.2^{1}	82.9 ± 5.6^{6}	78.8 ± 4.9^{3}	57.3
Post-operative	64.7 ± 6.8	67.2 ± 5.0^{3}	70.0 ± 5.2^{1}	59.2 ± 8.0^{2}	60.8 ± 8.2^{4}	71.2
Promoters	87.9 ± 7.0	87.9 ± 7.0^{6}	74.3 ± 7.8^{1}	91.7 ± 5.9^{3}	75.9 ± 8.8^{1}	43.1
Primary tumor	44.2 ± 5.5	44.2 ± 5.5^{6}	35.9 ± 5.8^{1}	30.9 ± 4.7^{1}	39.8 ± 5.2^{1}	24.6
Solar flare	68.5 ± 3.0	68.2 ± 3.7^{6}	70.6 ± 2.9^{1}	67.6 ± 3.5^{6}	70.4 ± 3.0^{2}	25.2
Sonar	69.4 ± 7.6	63.0 ± 8.3^{1}	69.1 ± 7.4^{6}	73.8 ± 7.4^{1}	66.2 ± 7.5^{5}	50.8
Soybean	100.0 ± 0.0	100.0 ± 0.0^{6}	95.0 ± 9.0^{3}	100.0 ± 0.0^{6}	96.9 ± 5.9^{3}	30.0
Splice junctions	95.4 ± 0.6	95.4 ± 0.6^{6}	93.4 ± 0.8^{1}	94.3 ± 0.5^{1}	81.5 ± 5.5^{1}	52.4
Voting records	91.2 ± 1.7	91.2 ± 1.7^{6}	96.3 ± 1.3^{1}	94.9 ± 1.2^{1}	95.8 ± 1.6^{1}	60.5

Domingos & Pazzani: Optimality of Simple Bayesian Classifier under Zero-One Loss, *Machine Learning*, Vol. 29, 103-130, 1997



If you want old ideas, read new books.
If you want new ideas, read old books.
-- Ivan Pavlov
1849-1939

Alan Turing, 1939

"The well-known theorem of Gödel (1931) shows that every system of logic is in a some sense incomplete, but at the same time it indicates means whereby from a system L of logic a more complete system L′ may be obtained. By repeating the process we get a sequence

L, L1 = L′, L2 = L1, ...

each more complete than the proceeding. A logic Lω may then be constructed in which the provable theorems are the totality of theorems provable with the help of logics L, L1, L2..." .

Communities of agents

- Some silicon, some carbon
- Helping each other out