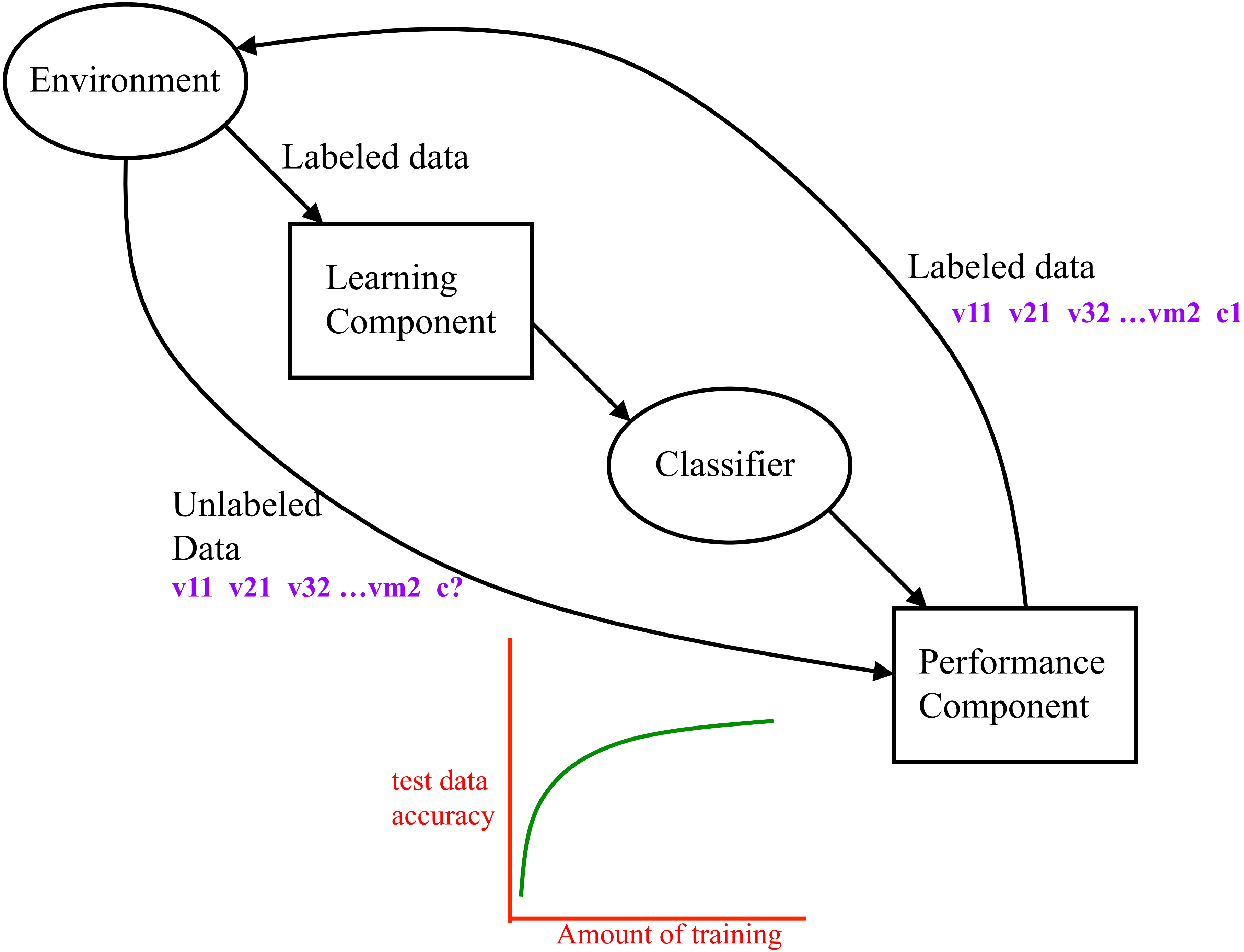

Environment
Labeled data
Learning Component
Labeled data
v11 v21 v32 …vm2 c1
Classifier
Unlabeled Data
v11 v21 v32 …vm2 c?
Performance Component
test data accuracy
Amount of training

Early Days of AI (and still)

AI (or KB) analysts would interview experts (e.g., medicine, manufacturing)

To elicit the "rules" the experts used for diagnosis and other processing

From these rules, build an "expert system"

Often a painful process

Supervised machine learning (early 1980s)

Let the human expert do what the expert does best: exercise "rules by labeling data

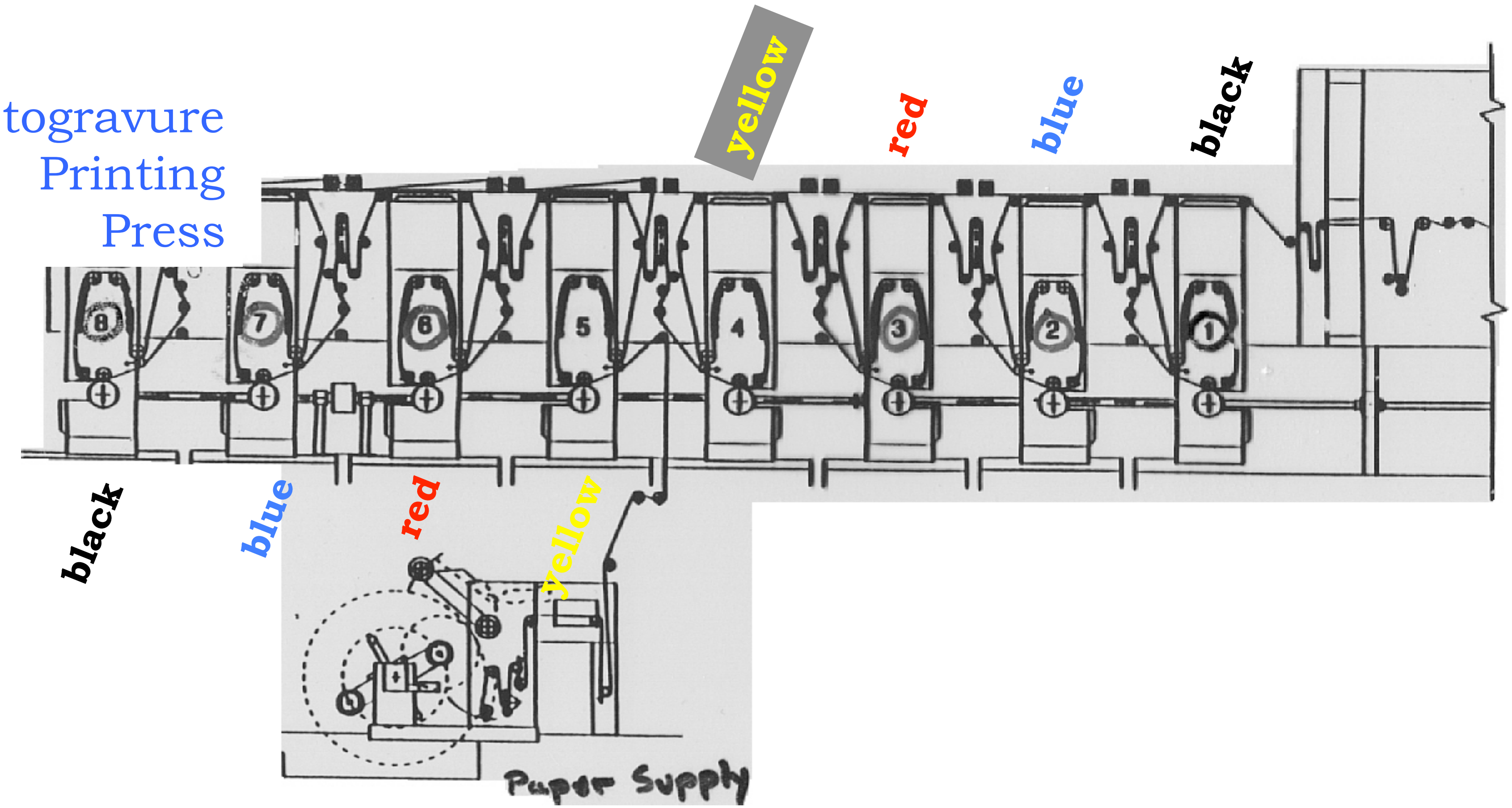
Let the machine do what the machine does best: find patterns that are predictive of labels

Issues, variations, optimizations, etc:

- continuous attributes
  - hard versus soft splits
- other node types (e.g., perceptron trees)
- continuous classes (regression trees)
- termination conditions (pruning)
- selection measures (see problem DT1)
- missing values
  - during training
  - during classification (see expansion)
- noise in data
- irrelevant attributes
- less greedy variants (e.g., lookahead)
- incremental construction
- applications (e.g., Banding)
- cognitive modeling (e.g., Hunt)
- DT based approaches to nearest neighbor search, object recognition
- background **knowledge** to augment feature space

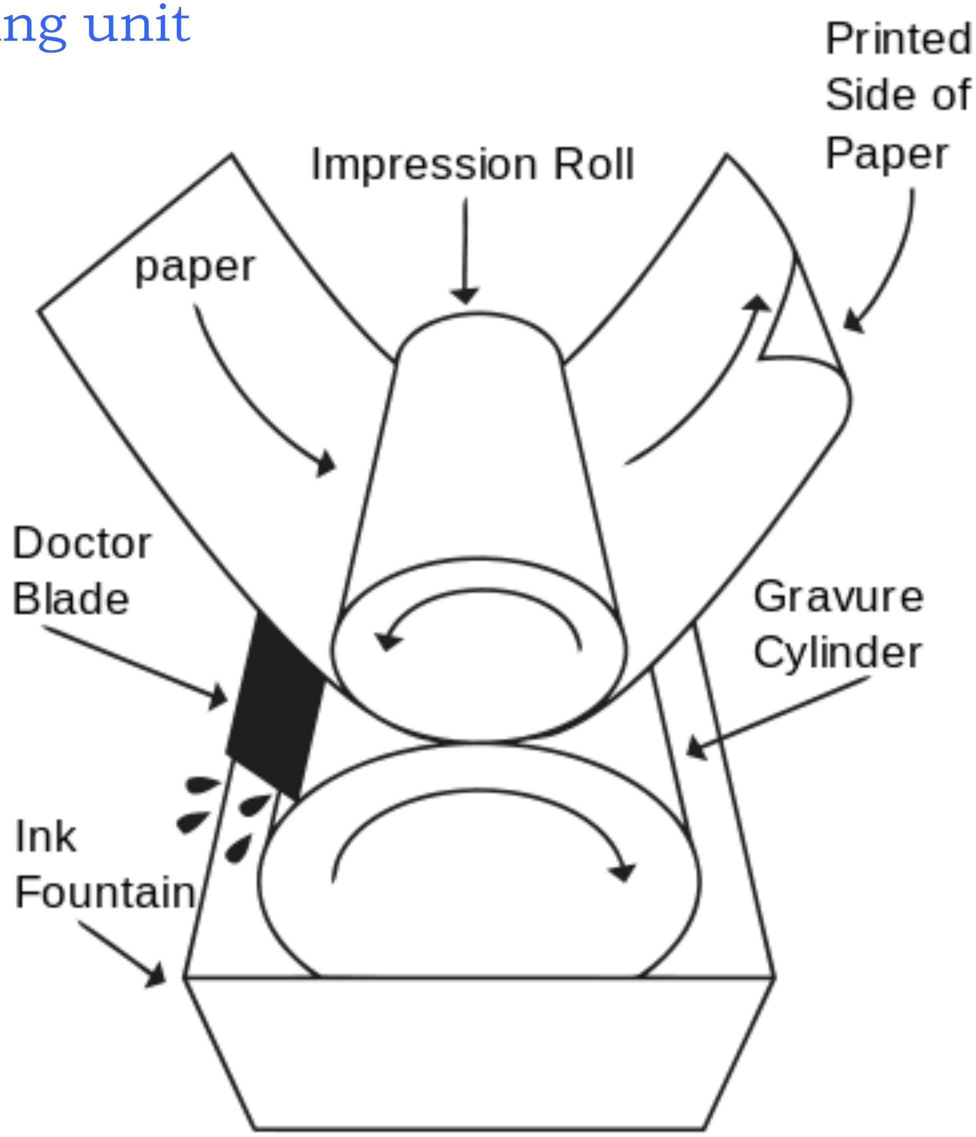
Features and Constraints
(and machine learning and rotogravure printing)

Rotogravure
Printing
Press

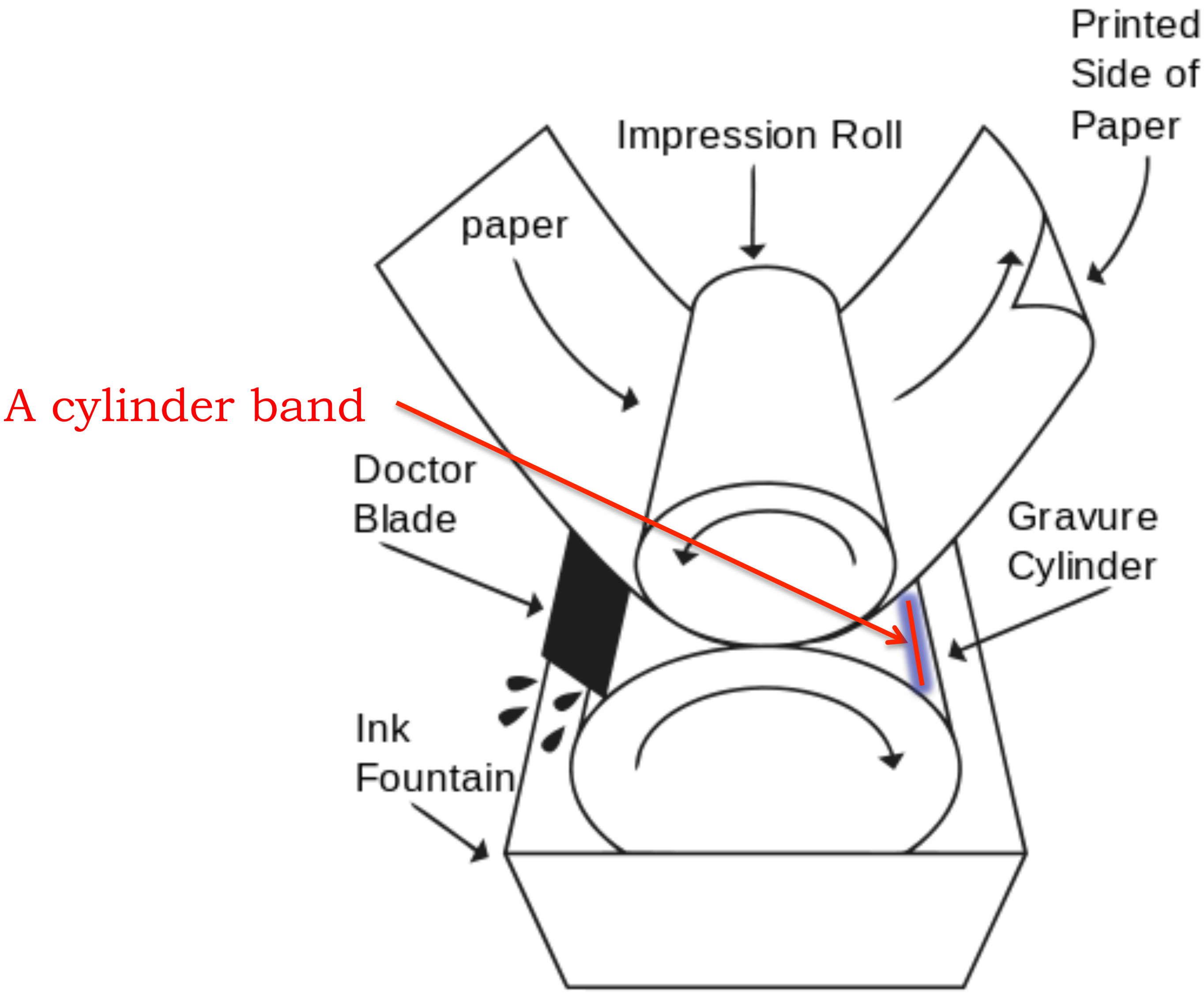


http://en.wikipedia.org/wiki/CMYK_color_model
http://en.wikipedia.org/wiki/Rotogravure

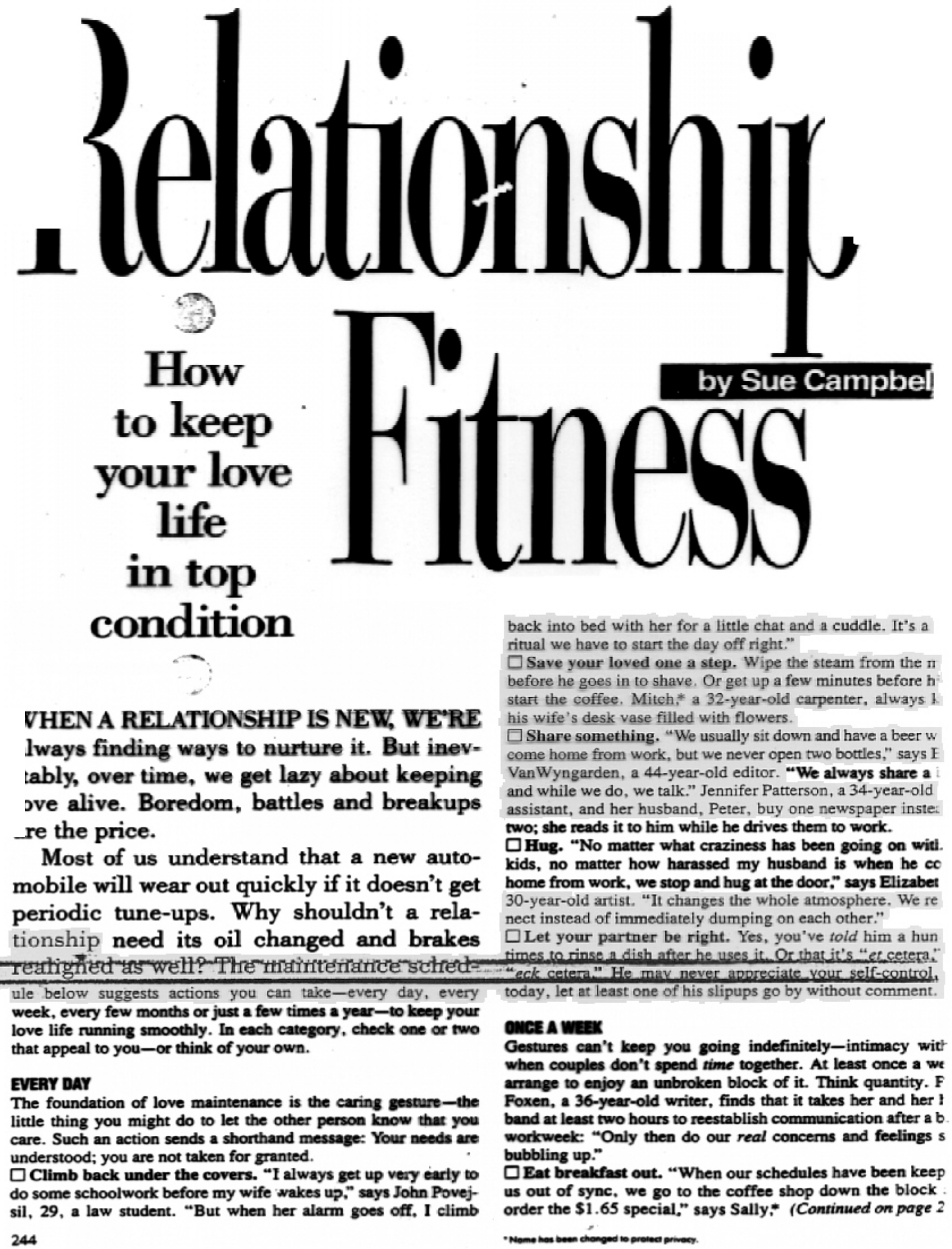
A single printing unit



http://en.wikipedia.org/wiki/File:Rotogravure_PrintUnit.svg



http://en.wikipedia.org/wiki/
File:Rotogravure_PrintUnit.svg


Relationship Fitness

How to keep your love life in top condition

by Sue Campbel

WHEN A RELATIONSHIP IS NEW, WE'RE always finding ways to nurture it. But inevitably, over time, we get lazy about keeping love alive. Boredom, battles and breakups are the price.

Most of us understand that a new automobile will wear out quickly if it doesn't get periodic tune-ups. Why shouldn't a relationship need its oil changed and brakes realigned as well? The maintenance schedule below suggests actions you can take—every day, every week, every few months or just a few times a year—to keep your love life running smoothly. In each category, check one or two that appeal to you—or think of your own.

EVERY DAY

The foundation of love maintenance is the caring gesture—the little thing you might do to let the other person know that you care. Such an action sends a shorthand message: Your needs are understood; you are not taken for granted.

☐ **Climb back under the covers.** "I always get up very early to do some schoolwork before my wife wakes up," says John Povejsil, 29, a law student. "But when her alarm goes off, I climb back into bed with her for a little chat and a cuddle. It's a ritual we have to start the day off right."

☐ **Save your loved one a step.** Wipe the steam from the m before he goes in to shave. Or get up a few minutes before h start the coffee. Mitch,* a 32-year-old carpenter, always k his wife's desk vase filled with flowers.

☐ **Share something.** "We usually sit down and have a beer w come home from work, but we never open two bottles," says E VanWyngarden, a 44-year-old editor. "We always share a and while we do, we talk." Jennifer Patterson, a 34-year-old assistant, and her husband, Peter, buy one newspaper inste two; she reads it to him while he drives them to work.

☐ **Hug.** "No matter what craziness has been going on with kids, no matter how harassed my husband is when he co home from work, we stop and hug at the door," says Elizabet 30-year-old artist. "It changes the whole atmosphere. We re nect instead of immediately dumping on each other."

☐ **Let your partner be right.** Yes, you've *told* him a hun times to rinse a dish after he uses it. Or that it's "*et cetera*," "*eck cetera*." He may never appreciate your self-control, today, let at least one of his slipups go by without comment.

ONCE A WEEK

Gestures can't keep you going indefinitely—intimacy with when couples don't spend *time* together. At least once a we arrange to enjoy an unbroken block of it. Think quantity. F Foxen, a 36-year-old writer, finds that it takes her and her band at least two hours to reestablish communication after a b workweek: "Only then do our *real* concerns and feelings s bubbling up."

☐ **Eat breakfast out.** "When our schedules have been keep us out of sync, we go to the coffee shop down the block order the $1.65 special," says Sally.* *(Continued on page 2*

*Name has been changed to protect privacy.

result of a cylinder band

Douglas H. F...er

Context at one large printing plant of R R Donnelley & Sons

- Plant runs 24/7 with hard deadlines
- 538 banding incidents in 1989
- Each band required 1.5 to 6 hours to remedy
- Time of 3 to 10 crew members
- What to do?

Machine Learning (ML) to the Rescue …
but ML systems require data

“attribute” is synonymous with **“variable”**

Attributes of *printing unit* (e.g., ink viscosity and temp); *printing press* (paper type, rotation speed), *ambient conditions* (e.g., temperature, humidity), etc

30+ variables

Table 3. Banding attributes, domain values, and units. The algorithm treats both continuous and ordinal attributes as “numeric” attributes.

Attribute	Domain values/range	Domain type/units
Roughness	0–0.5	Continuous/microns
Anode distance	0–100	Continuous/millimeters
Chrome solution ratio	0–200	Continuous/ratio of chemical mix
Current density	0–200	Continuous/amperes per square decimeter
Plating tank	1910, 1911, Other	Nominal/plating tank
Viscosity	0–30	Continuous/seconds
Proof press cut	0–100	Continuous/percentage
Proof on coated ink	Yes, No	Boolean
Humidity	0–100	Continuous/percentage
Ink temperature	0–120	Continuous/degrees Fahrenheit
Blade oscillation	0–2	Continuous/inches
Blade pressure	0–100	Continuous/pounds
Type on cylinder	Yes, No	Boolean
Blade manufacturer	Benton, Daetwyler, Udeholm	Nominal/blade manufacturer
Varnish percentage	0–100	Continuous/percentage
Ink percentage	0–100	Continuous/percentage
Solvent percentage	0–100	Continuous/percentage
Wax	0–100	Continuous/gallons
Hardener	0–100	Continuous/gallons
Press speed	0–4,000	Continuous/feet per minute
Paper type	Uncoated, Super, Coated	Nominal/paper class
Ink type	Uncoated, Coated	Nominal/ink grade
Steam bar	On, Off	Boolean
Solvent type	Line, Lactol, Xylol, Naphtha, Other	Nominal/commercial solvent
Grain screened	Yes, No	Boolean
Press	TR802, TR813, TR815, TR816, TR821, TR824, TR827, TR828	Nominal/presses
Unit	1 ... 10	Ordinal/printing unit
ESA current	0–5	Continuous/milliamps
ESA voltage	0–5	Continuous/Kilovolts
Cylinder size	Catalog, Spiegel, Tabloid	Nominal/cylinder circumference
Basis weight	0–120	Continuous/pounds per ream

Table from
Bob Evans and Doug Fisher, *“Overcoming Process Delays with Decision Tree Induction”* IEEE Expert, Vol. 9, No. 1, Feb 1994, pp. 60-66.

Machine Learning approach to Banding

- Take system "snaphots" along 30+ variables
  - when job banded, and
  - when it did not! *(evidence-based reasoning!)*

- Over 500+ such snapshots (of 30+ features each),
  - About half of banded and half of not,
  - Learn patterns of variable values (aka features) of when banding would likely occur and when not

- From discovered patterns from the data,
  - Make ***constraints*** on operating conditions that press crews (or automated agents) should adhere to

Machine Learning approach to Banding

- Take system “snaphots” along 30+ variables
  - when job banded, and
  - when it did not! *(evidence-based reasoning!)*

- Over 500+ such snapshots (of 30+ features each),
  - About half of banded and half of not,
  - Learn patterns of variable values (aka features) of when banding would likely occur and when not

- From discovered patterns from the data,
  - Make ***constraints*** on operating conditions that press crews (or automated agents) should adhere to

Machine Learning approach to Banding

- Take system "snaphots" along 30+ variables
  - when job banded, and
  - when it did not! *(evidence-based reasoning!)*

- Over 500+ such snapshots (of 30+ features each),
  - About half of banded and half of not,
  - Learn patterns of variable values (aka features) of when banding would likely occur and when not

- From discovered patterns from the data,
  - Make ***constraints*** on operating conditions that press crews (or automated agents) should adhere to

Chrome Solution Ratio > t1 (high) and
Ink Temperature <= t2 (low) and
Ink Viscosity > t3 (high) ➔ No Banding

Machine Learning approach to Banding

- Take system “snaphots” along 30+ variables
  - when job banded, and
  - when it did not! *(evidence-based reasoning!)*

- Over 500+ such snapshots (of 30+ features each),
  - About half of banded and half of not,
  - Learn patterns of variable values (aka features) of when banding would likely occur and when not

- From discovered patterns from the data,
  - Make ***constraints*** on operating conditions that press crews (or automated agents) should adhere to

Years	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
Bands	**538**	**384**	**138**	**66**	**42**	**109**	**21**	**26**	**37**	**26**

Decrease in bands due to increase in acceptance of learned rules

Conta-
Minated
ink

Machine Learning Relationship(s)
to
Constraint Satisfaction

Variables (30+)	Humidity	ink visc	ink temp	CSR	...	Banded?
Snapshots	85	25	75	180		NO
	92	27	81	179		NO
	. . .					
	70	18	100	105		YES
	86	16	105	120		YES

GIVEN
- **a set of variables**
- **a domain for each variable**
- **a set of constraints**

FIND

an assignment of variables
to particular values
that satisfy constraints

Constraint: Banded? = NO

Trivial Solution: Each assignment fou
in NOT Banded snapshots

Machine Learning Relationship(s)
to
Constraint Satisfaction

Variables (30+)	Humidity	ink visc	ink temp	CSR	…	
Snapshots	85	25	75	180		NOT Banded
	92	27	81	179		NOT Banded
	. . .					
	70	18	100	105		Banded
	86	16	105	120		Banded

FIND

- ~~**an assignment of variables to particular values that satisfy constraints**~~

➔ **discovery of 'maximal' subdomains for each variable and**
relationships between variables

Chrome Solution Ratio > t1 (high) and
Ink Temperature <= t2 (low) and
Ink Viscosity > t3 (high) and
don't care on remainder

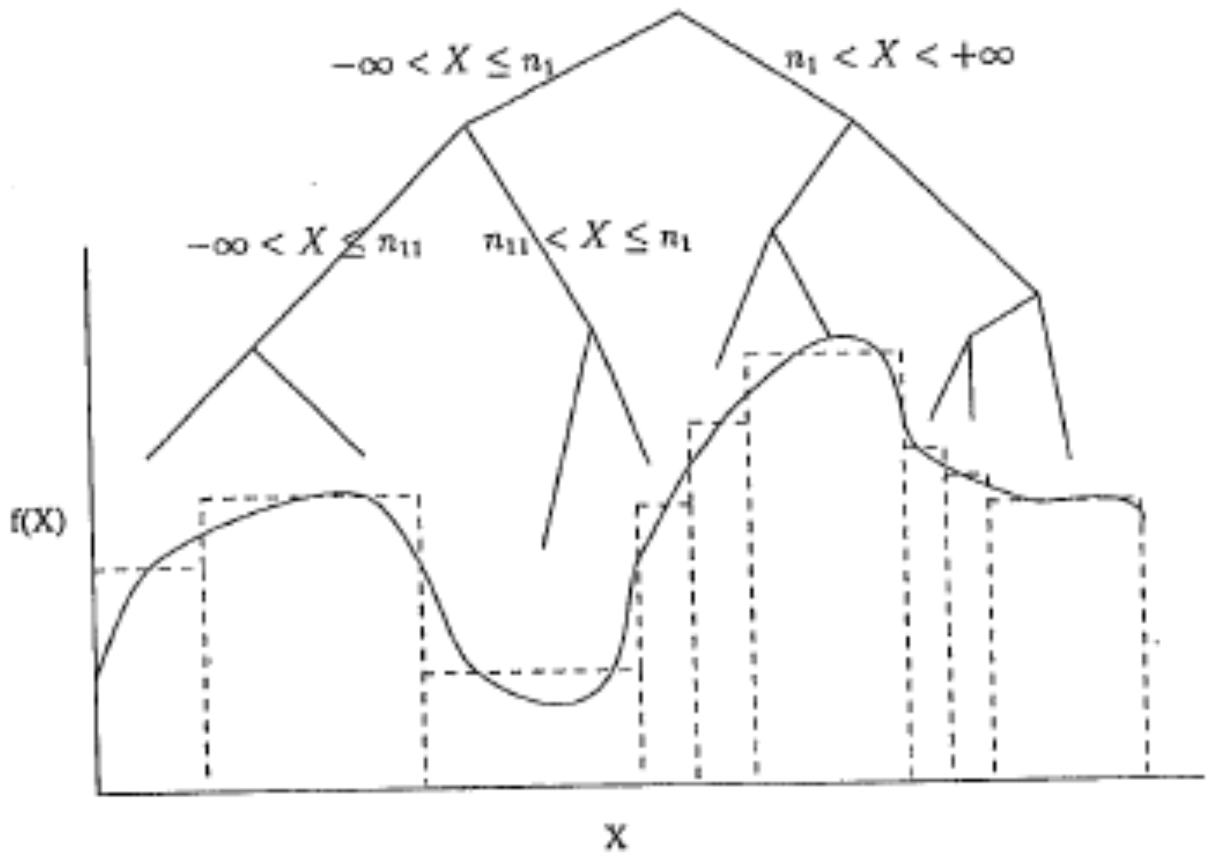
If you want to learn more about the machine learning application to rotogravure printing

Evans, B. and Fisher, D. (2002) "*Using Decision tree Induction to Minimize Process Delays in the Printing Industry*." In Handbook of Data Mining and Knowledge Discovery, W. Klosgen and J. Zytkow (Eds), Oxford University press, Retrieved from http://www.vuse.vanderbilt.edu/~dfisher/KDD-Handbook/printing.pdf

Bob Evans and Doug Fisher, "*Overcoming Process Delays with Decision Tree Induction*" IEEE Expert, Vol. 9, No. 1, Feb 1994, pp. 60-66.

Variations of DT Induction

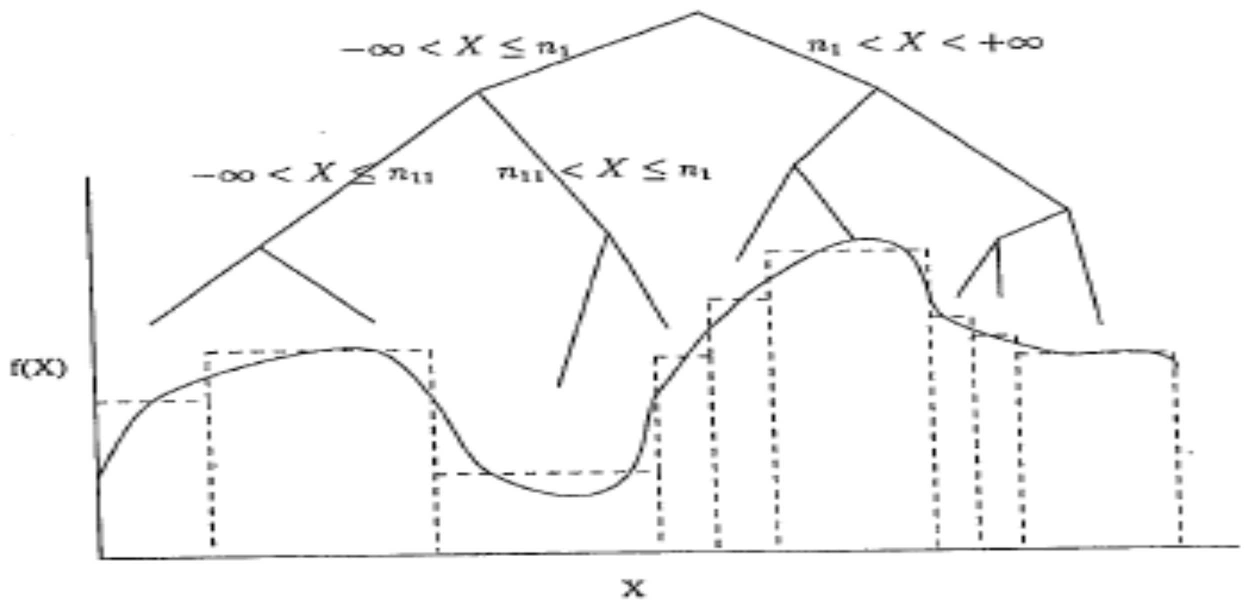
Regression trees predict values along a continuously-valued dependent variable



Regression tree over one variable, with an illustration from Srinivasan and Fisher (1995) *IEEE Software Engineering* paper on estimating software development time (http://dl.acm.org/citation.cfm?id=205309)

Variations of DT Induction

R regression tree over one variable, with an illustration from the IEEE Software Engineering paper on estimating software development time.



We also discussed using linear regression at each regression tree leaf instead of using zero-order models (i.e., $h(x) = \theta_0$) at each leaf. For example, using a linear regression model over x, we might have the following at two leaves of the regression tree.

To make a prediction of y for a given x, we would classify the x to a leaf and then use the linear model over that leaf to estimate y by h(x).

Lots of different search algorithms possible !!

Ensembles of classifiers

Other supervised approaches: ANNs, SVMs, …

Relational (e.g., first-order) representations, such as:

$$\textbf{IF } R(?c1, ?r1) \wedge R(?c2, ?r1) \wedge R(?c3, ?r2) \wedge R(?c4, ?r2) \wedge R(?c5, ?r2) \wedge \neq(?c1, ?c2) \wedge \neq(?c3, ?c4) \wedge \neq(?c3, ?c5) \wedge \neq(?c4, ?c5)$$
$$\textbf{THEN } FullHouse(?c1, ?c2, ?c3, ?c4, ?c5)$$

K H	K C	5 S	5 C	5 D	2 D	9 H	2 C	2 H	9 D
6 S	6 H	7 C	7 D	7 H	A C	3 C	3 D	A H	3 H

The matching problem (on sets of feature vectors)

Empirical, Supervised Learning

Example: Naïve Bayesian Classifiers

Subclass: Supervised Rule Induction

Example: Decision tree induction

Example: Brute-force induction of decision rules

Empirical, Unsupervised Learning

Unsupervised Rule Induction

Association Rule Learning

Bayesian Network Learning

Clustering

Analytical Learning

Explanation-Based Learning

Empirical/Analytic Hybrids

Unsupervised Performance Task: Pattern Completion

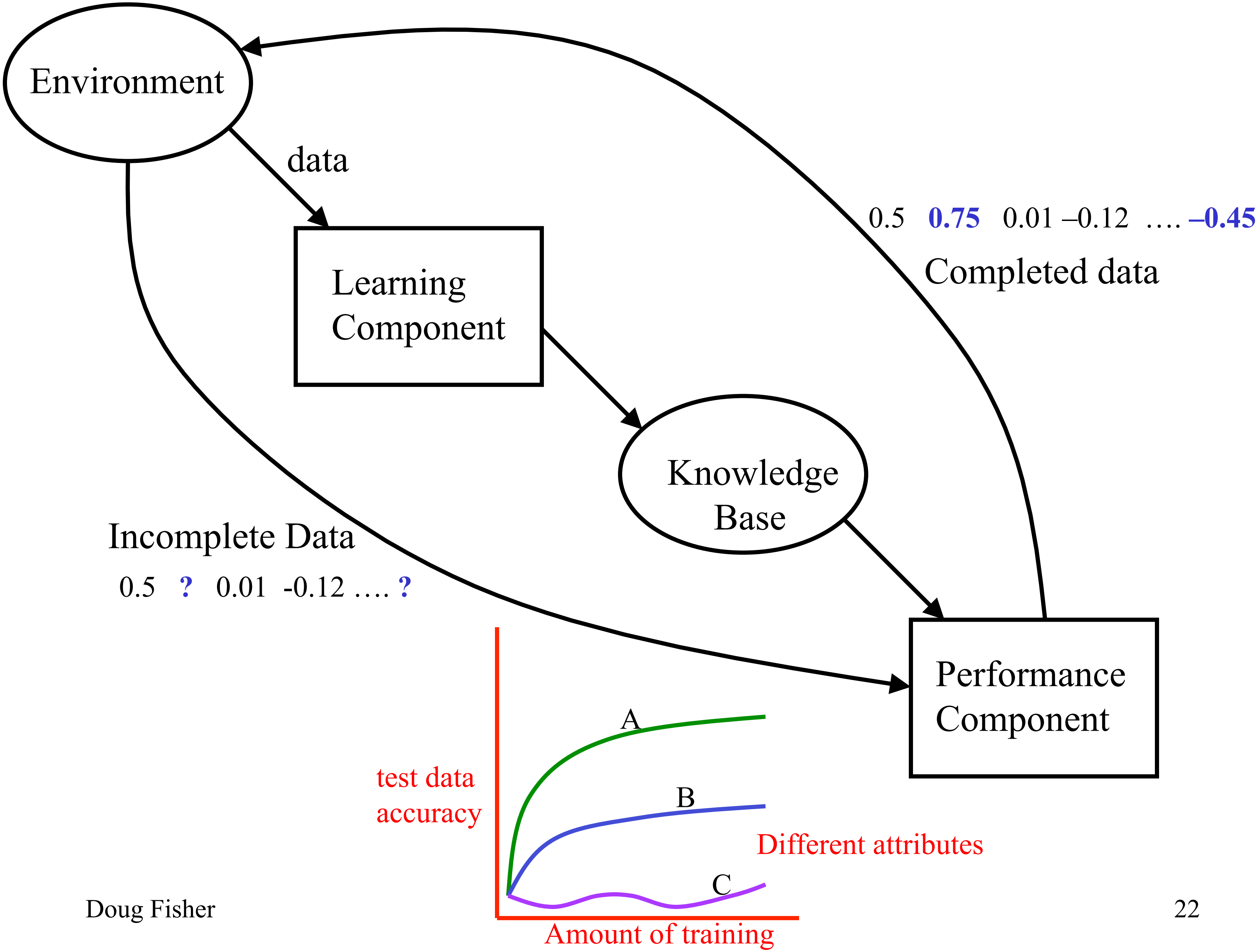
0.5   ?   0.01   -0.12 …. ?   ~~C'   C''~~

↓

Knowledge
Base

↓

0.5   **0.75**   0.01 –0.12   …. **–0.45**   ~~c'1   c''2~~


Environment
data
Learning Component
Knowledge Base
Performance Component
0.5 0.75 0.01 –0.12 …. –0.45
Completed data
Incomplete Data
0.5 ? 0.01 -0.12 …. ?
test data accuracy
A
B
C
Different attributes
Amount of training

Example: Unsupervised rule induction of Association Rules (market-basket analysis)

In a nutshell: run brute force rule discovery for all possible consequents, not simply single variable values (e.g., V1=v12), but consequents that are conjunctions of variable values (e.g., V1=v12 & V4=v42 & V5=v51).

Retain rules A → C such that P(A & C) >= T1 and P(C|A) >= T2. These thresholds enable pruning of the search space (A and C are themselves conjunctions).

Problem: a plethora of rules, most uninteresting, are produced.

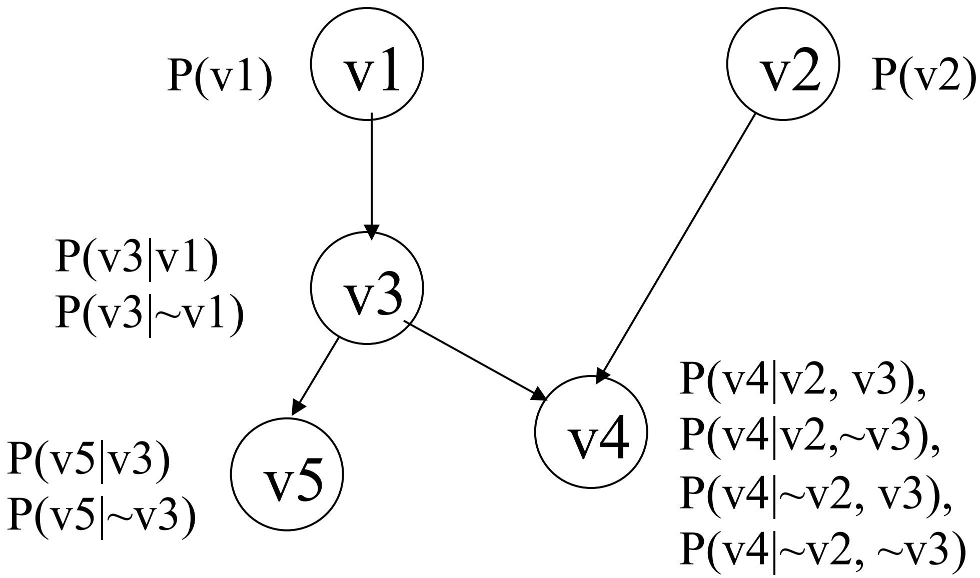
Solutions: Organize/prune rules by

a) Interestingness (e.g., A→C interesting if P(A, C) >> P(A)P(C) or << P(A)P(C)

b) confidence (a confidence interval around coverage and/or accuracy)

c) support for top-level goal

Example (Empirical, Unsupervised): Learning Bayesian Networks



Components of a Bayesian Network: a topology (graph) that qualitatively indicates displays the conditional independencies, and probability tables at each node

Semantics of graphical component: for each variable, v, v is independent of all of its non-descendents conditioned on its parents

A Bayesian Network is a graphical representation of a joint probability distribution with (conditional) independence relationships made explicit

Example (Empirical, Unsupervised): Clustering

Given data (vectors of variable values)

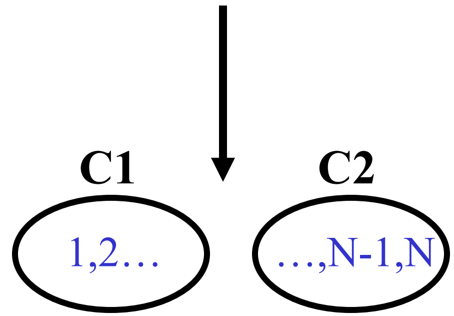
Compute a partition (clusters) of the vectors, such that vectors within a cluster tend to be similar, and vectors across clusters tend to be dissimilar

For example,

	V1	V2	V3	V4	……………	VM
1	0.3	0.7	0.1	-0.2	………….	-0.5
2	0.4	0.8	0.01	0.1	………….	-0.4
		…………………….				
N-1	-0.3	0.1	1.01	0.8	………….	1.3
N	-0.5	0.03	1.1	0.9	………….	0.9

Cluster summary representations (e.g., the centroid)

	V1	V2	V3	V4	…………..	VM
1	0.3	0.7	0.1	-0.2	………….	-0.5
2	0.4	0.8	0.01	0.1	………….	-0.4
		……………………				
N-1	-0.3	0.1	1.01	0.8	………….	1.3
N	-0.5	0.03	1.1	0.9	………….	0.9



0.35 0.75 0.05 -0.05 …. –0.45
(centroid for C1)

-0.4 0.05 1.05 0.85 …. 1.1
(centroid for C2)

Using summary representations for inference

0.5 **?** 0.01 -0.12 …. **?**

C1 **C2**

1,2… …,N-1,N

0.35 0.75 0.05 -0.05 …. –0.45
(centroid for C1)

-0.4 0.05 1.05 0.85 …. 1.1
(centroid for C2)

0.5 **0.75** 0.01 –0.12 …. **–0.45**

K-means

```
Clustering K-Means (Data, K) {
    ClusterCentroids = K randomly selected vectors from Data
    for each d in Data
        assign d to cluster with closest centroid
    do {
        compute new cluster centroids
        for each d in Data
            assign d to cluster with closest centroid
    } while NOT termination condition
}
```

"closest": Euclidean distance

Empirical, Supervised Learning

- Example: Naïve Bayesian Classifiers
- Subclass: Supervised Rule Induction
  - Example: Decision tree induction
  - Example: Brute-force induction of decision rules

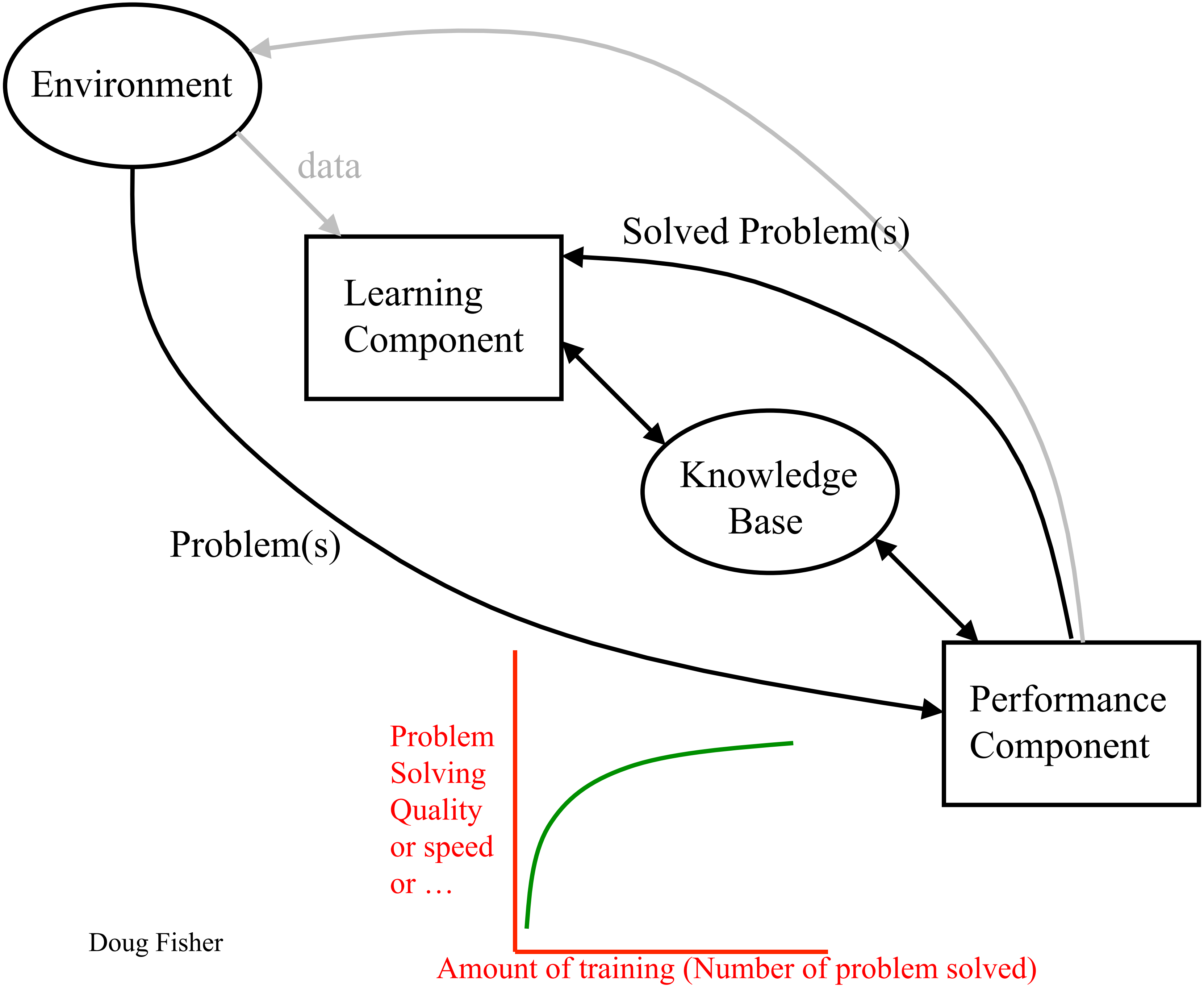
Empirical, Unsupervised Learning

- Unsupervised Rule Induction
  - Association Rule Learning
- Bayesian Network Learning
- Clustering

Analytical Learning

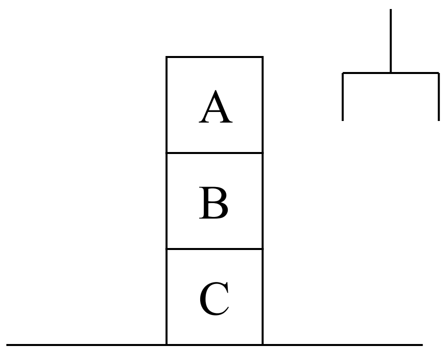
- Explanation-Based Learning

Empirical/Analytic Hybrids

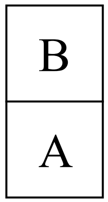

Environment
data
Solved Problem(s)
Learning Component
Knowledge Base
Problem(s)
Performance Component
Problem Solving Quality or speed or …
Amount of training (Number of problem solved)

Learning macros: Given a plan, generalize the plan so that the generalized plan can be applied in a greater number of situations

Objective: reusing previously-developed generalized plans (aka macro-operators) will reduce the cost (improve the “speed”) of subsequent planning



Start State



GoalSpec

Unstack(A,B) → Putdown(A) → Unstack(B,C) → Stack(B,A)

(Generalize) →

Unstack(?x1, ?y1) → Putdown(?x1) → Unstack(?y1, ?z1) → Stack(?y1, ?x1)