12-4-2009

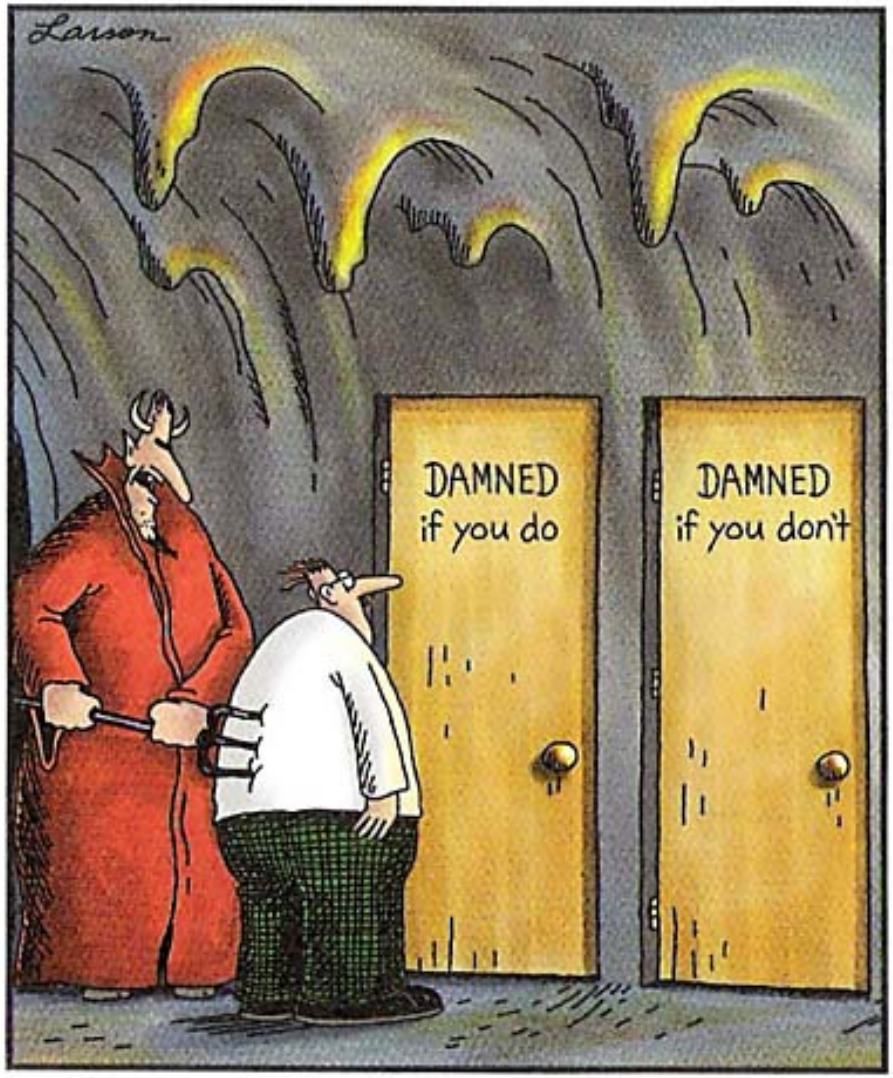
Creating Insanity in Learning Systems: Addressing Ambiguity Effects of Predicting Non-linear Continuous Valued Functions with Reconstructabilty Analysis from Large Categorically Valued Input Data Sets

William D. Eisenhauer
Portland State University

Creating insanity in learning systems

Addressing ambiguity effects of predicting non-linear continuous valued functions with reconstructability analysis from large categorically valued input data sets

William “Ike” Eisenhauer
Systems Science PhD Candidate - Portland State University
December 4, 2009

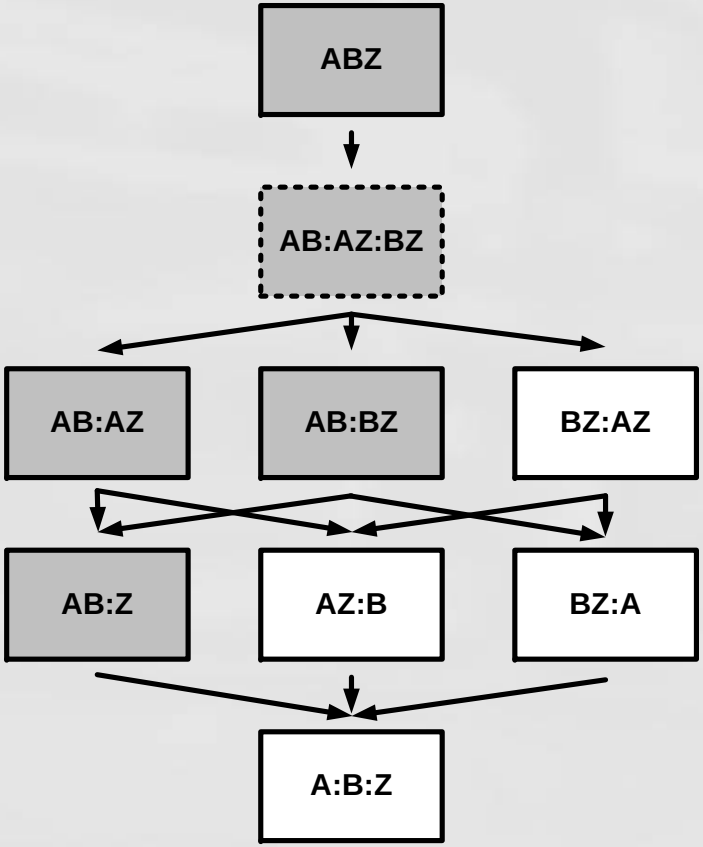


"C'mon, c'mon—it's either one or the other."

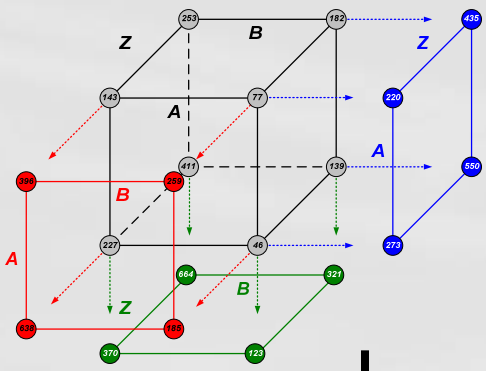
Overview

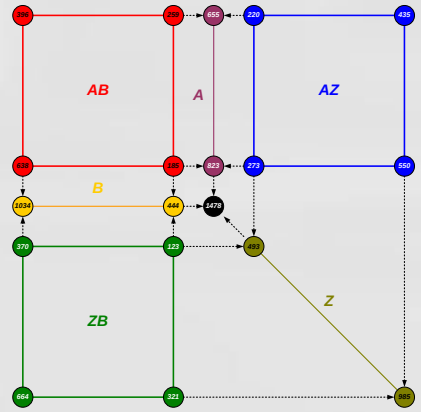
- Reconstructability Analysis
- Text Mining with RA
- New York Times – US Unemployment
- K-Systems
- Double Bind
- Expected Value Problem
- Questions for Discussion

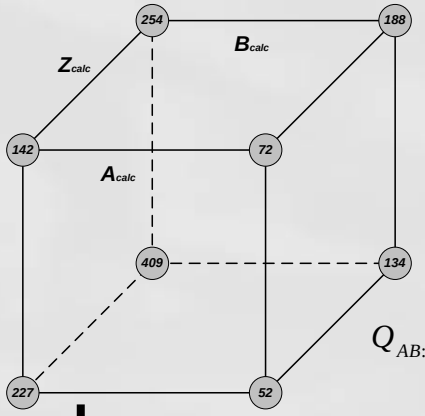
Reconstructability Analysis



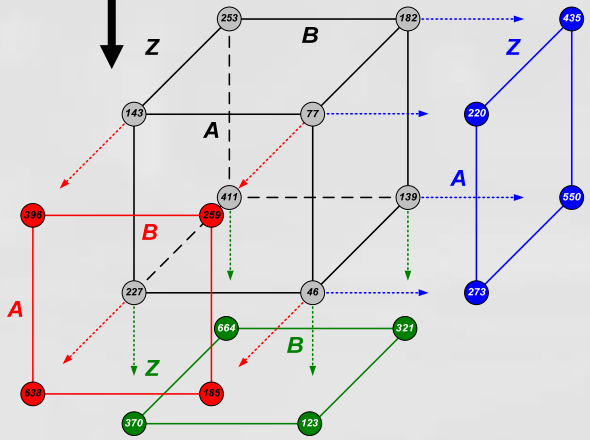
Reconstructability Analysis



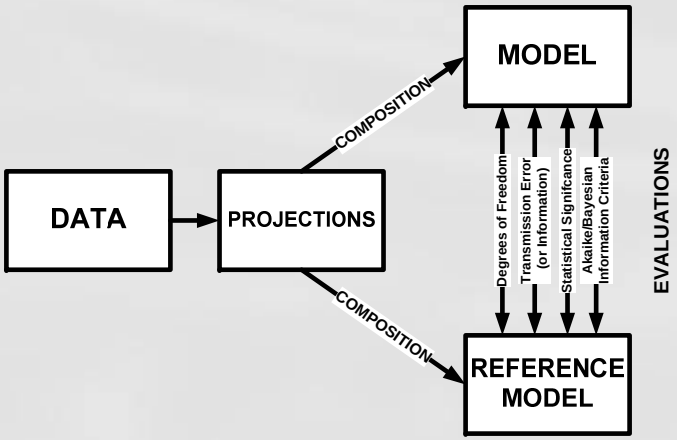




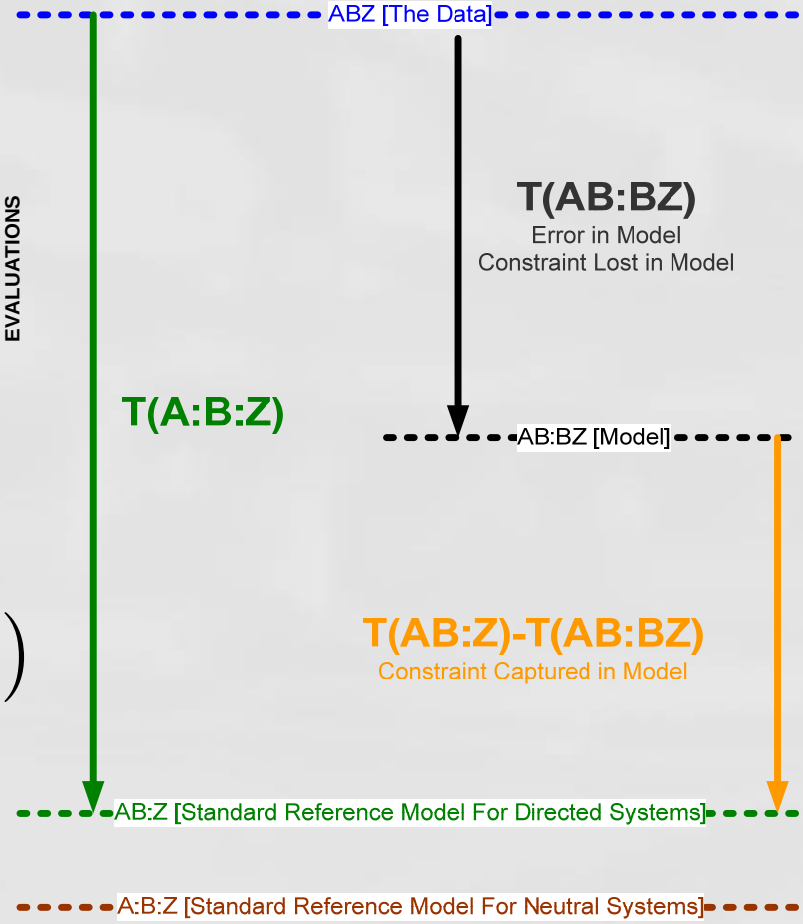
$$Q_{AB:BZ} = \frac{P\left(A,B\right)P\left(B,Z\right)}{P\left(B\right)}$$



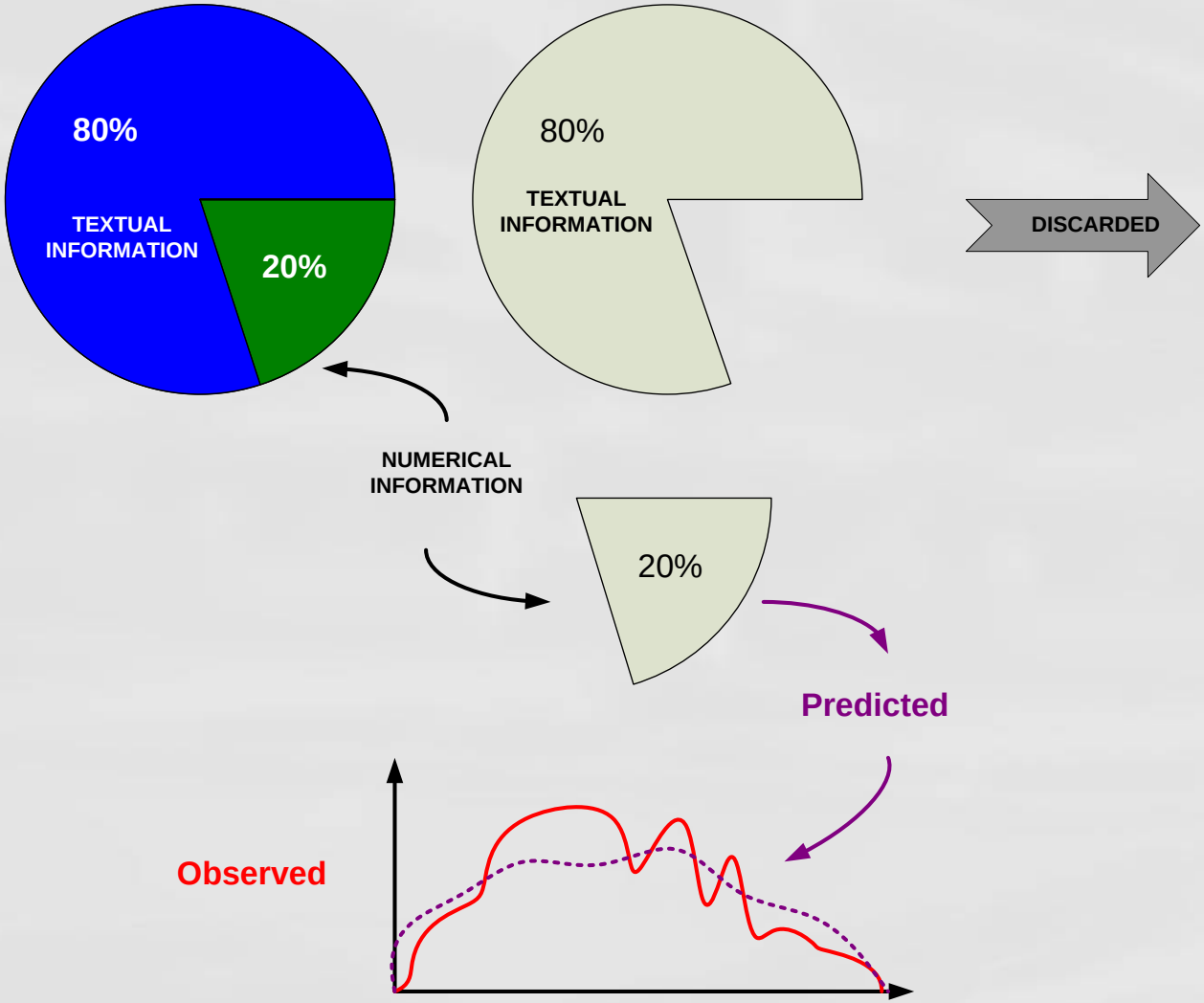
Reconstructability Analysis



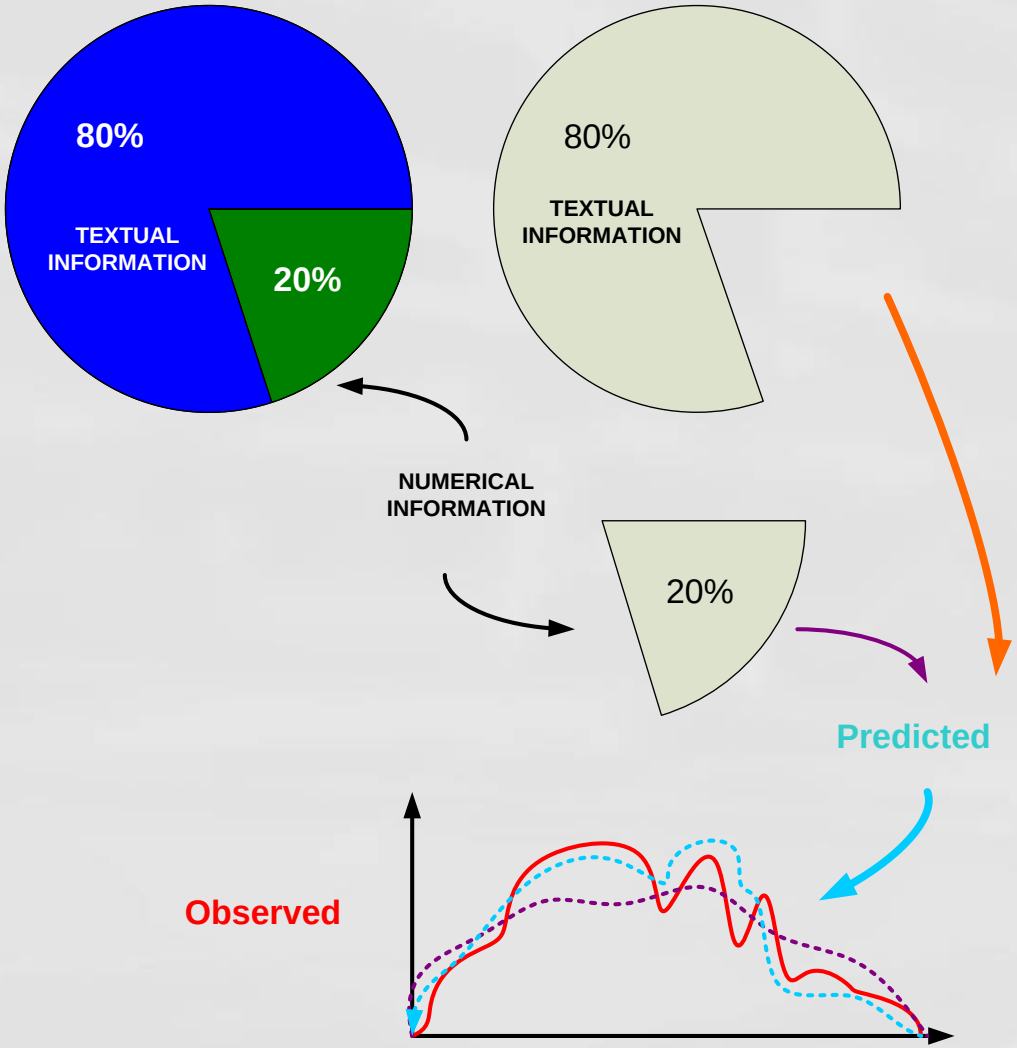
$$H = -\sum p \cdot \log(p)$$



Text Mining


80%
TEXTUAL INFORMATION
20%
80%
TEXTUAL INFORMATION
DISCARDED
NUMERICAL INFORMATION
20%
Predicted
Observed

Text Mining


80%
TEXTUAL INFORMATION
20%
80%
TEXTUAL INFORMATION
NUMERICAL INFORMATION
20%
Predicted
Observed

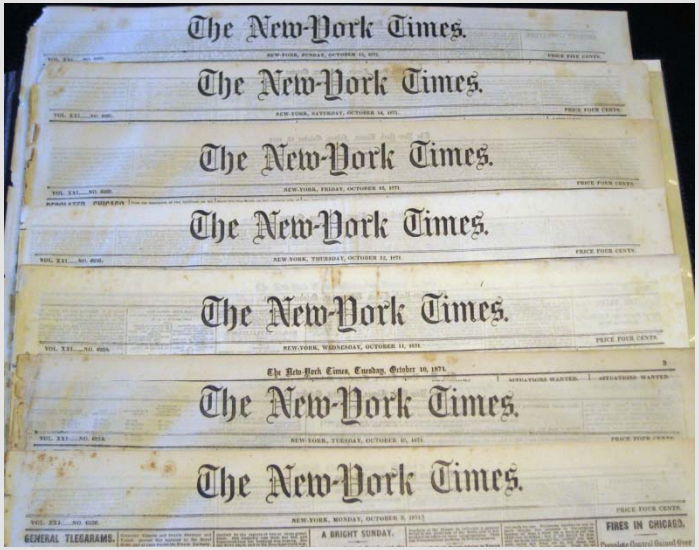
Research Questions

- Can Reconstructability Analysis (RA) methods enable the direct prediction of a continuous observed variable from textual data?
- What characteristics of continuous observed variables allow the effective use of RA?

Predictions

- Various short term, wide market variables that are implied to be affected by textual communications
  - Example
    - United States Non-Farm Non-seasonally adjusted unemployment numbers
    - Dow Jones Industrial Average
    - Consumer Price Index
- These variables are expected to lag behind the communication that potentially influences them.

NYT/Unemployment Example Data



NYT/Unemployment Example Data

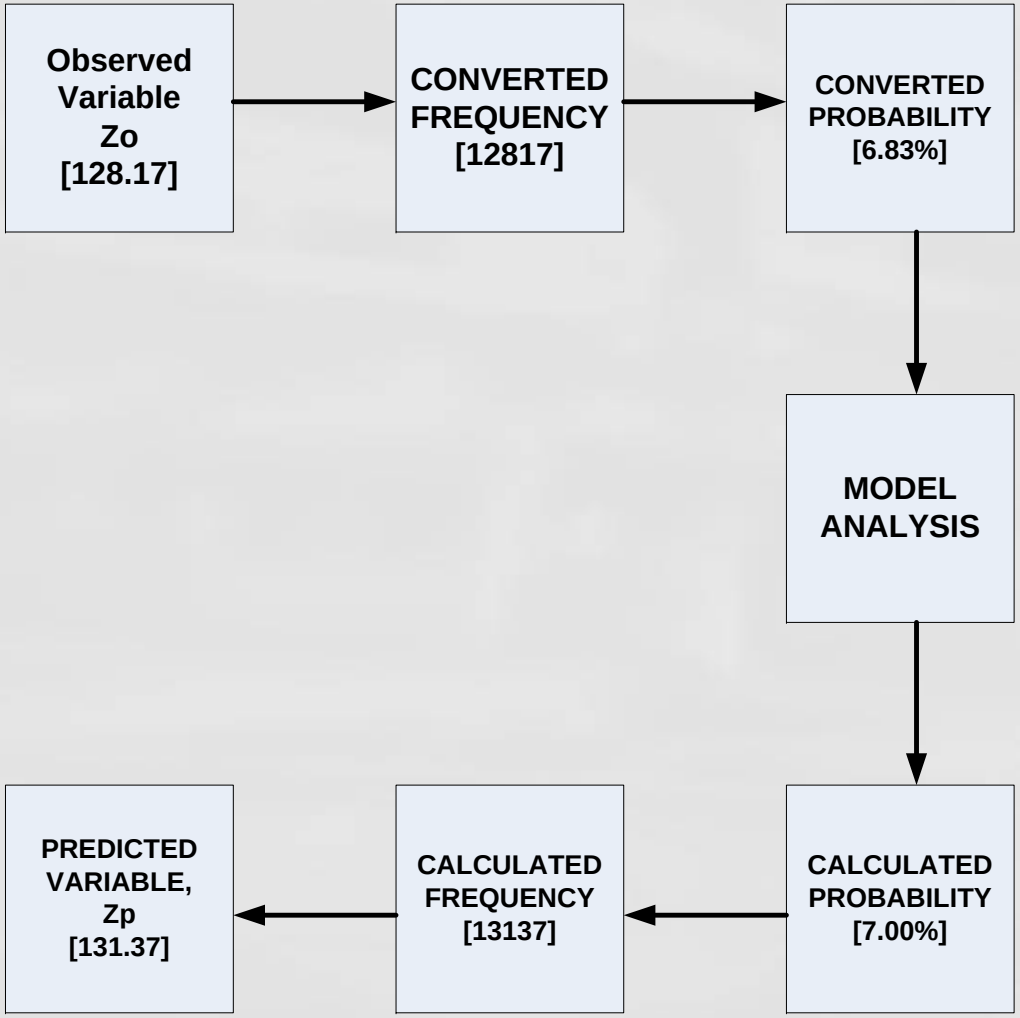
Start Date	1/1/1966
End Date	10/29/2008
Unique Dates	15,555
Missing Dates	90 day
Total Number of Words	30,212,348
Final Articles Used	3,616,957
Total Number of Words	20,371,039
Total Number of Unique Words	235,368
Total Number of Unique Words Freq >= 2	123,368
Average Words Per Headline	5.63

1	REPORTS
2	EARNINGS
3	QTR
4	NEWS
5	BUSINESS
6	BRIEFING
7	WORLD
8	COMPANY
9	CITY
10	SPORTS
11	YORK
12	CORP
13	PEOPLE
14	PLAN
15	REVIEW
16	DAY
17	STATE
18	JUNE
19	TIMES
20	MARCH

K-Systems Analysis

- Jones 1985
- Continuous dependent variable viewed as a non-negative frequency count.
- Normalized over the count of all observations
- Transform of the calculated frequency provides the “predicted” continuous value.
- Advantages are in the non-requirement of a maximum range parameter to be established or an a priori binning process to occur.
- Key disadvantage is the assumption, which is not valid in the text mining problem, of non-repeatability of a set of independent variables augmented by a differing continuous value.

K-Systems Analysis



K-Systems Analysis

Obs #	A	B	C	D	E	Z	FREQ	OBS	CALC	MODEL_Z
1	0	1	1	1	1	128.17	12817	0.0683	0.070	131.37
2	0	1	1	0	0	139.35	13935	0.0743	0.087	163.27
3	0	0	1	1	1	79.89	7989	0.0426	0.036	67.56
4	1	1	1	1	1	178.00	17800	0.0948	0.081	152.01
5	1	0	1	0	0	165.65	16565	0.0883	0.080	150.13
6	0	0	1	0	0	69.04	6904	0.0368	0.033	61.93
7	1	0	1	1	1	163.51	16351	0.0871	0.106	198.93
8	1	1	0	1	1	167.59	16759	0.0893	0.064	120.11
9	0	1	0	0	0	90.60	9060	0.0483	0.037	69.44
10	0	1	0	1	0	110.40	11040	0.0588	0.058	108.85
11	1	1	0	0	1	168.27	16827	0.0897	0.129	242.09
12	0	0	0	1	0	8.42	842	0.0045	0.005	9.38
13	0	0	0	1	1	1.00	100	0.0005	0.007	13.14
14	0	1	0	1	1	108.59	10859	0.0579	0.041	76.94
15	1	0	0	1	0	147.36	14736	0.0785	0.056	105.09
16	1	0	0	0	1	150.82	15082	0.0804	0.079	148.26

Table 6: K-Systems Example [Calculated Z]

Challenges in K-Systems Analysis

- Key disadvantage is the assumption, which is not valid in the text mining problem, of non-repeatability of a set of independent variables augmented by a differing continuous value.
- This becomes even more of an issue in text mining with RA due to the significant reduction in the variable sets, resulting in potential for repeatability in the independent variables

Double Bind in Human Learners

- Gregory Bateson
- “Communication in the context of an emotionally important relationship in which there is unacknowledged contradiction between messages of different logical levels.”
- Mary Poppins Scene

Double Bind in Human Learners

- Two individuals - one of them feels under pressure that it is vitally important to discriminate what sort of message is being communicated so she/he can respond appropriately.
- The “dependent” individual cannot survive without the others’ co-operation. They need to “please” the other to survive.
- The other person, is expressing two messages which conflict, and the second is likely to be expressed non-verbally.

Adapted from Bateson ‘ Steps to an Ecology of Mind’

Double Bind in Human Learners

- a primary negative injunction: “do not do ‘x’ or I will punish you”
- a secondary injunction conflicting with the first at a more abstract level, enforced by punishments or consequences which threaten survival.
- a third message is key to the structure - implication the dependent person is wrong, that he/she is the cause of the problem situation.

Adapted from Bateson ‘ Steps to an Ecology of Mind’

Double Bind in Human Learners

- There are three other important elements of a double bind including;
  - The “dependent” individual cannot comment on the messages being expressed or not allowed to meta-communicate.
  - Negative injunction prohibiting escape from the situation
  - Repeated Experience, not a single traumatic event, so the double bind becomes an expectation.

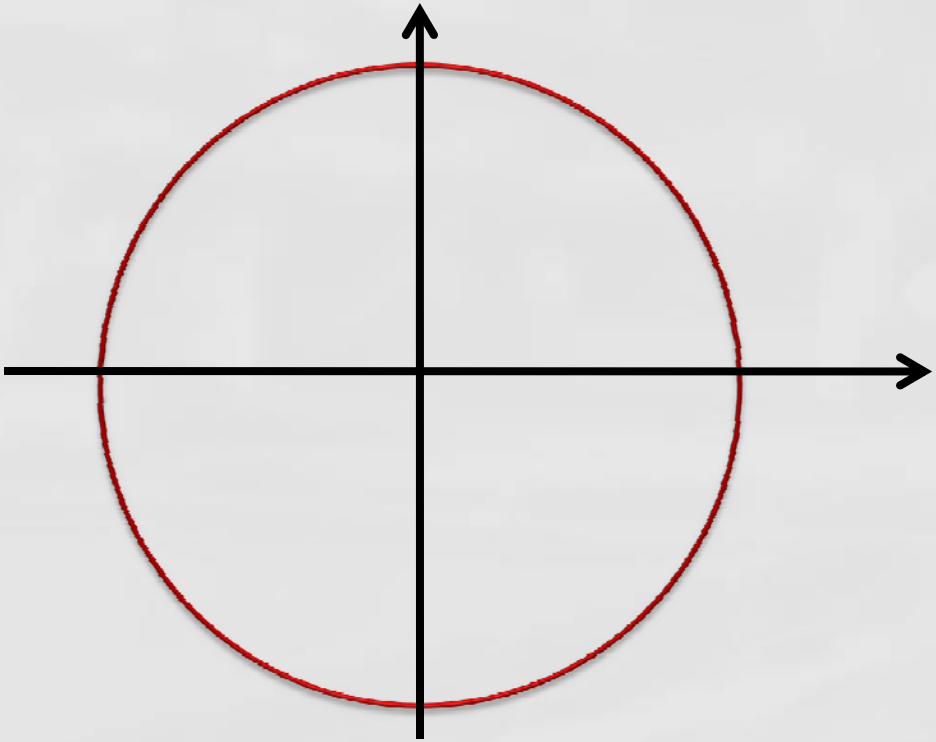
Adapted from Bateson ‘ Steps to an Ecology of Mind’

Machine Learning Double Bind

- Think of the situation where a model selection or learning system is locked in a relationship with its trainer
- However it has been trained that there is more than one “correct” response
- Through metrics of performance or survival the model is “punished” for giving the “other” response

Issues in Attempting to Minimize Error by Expected Value Response

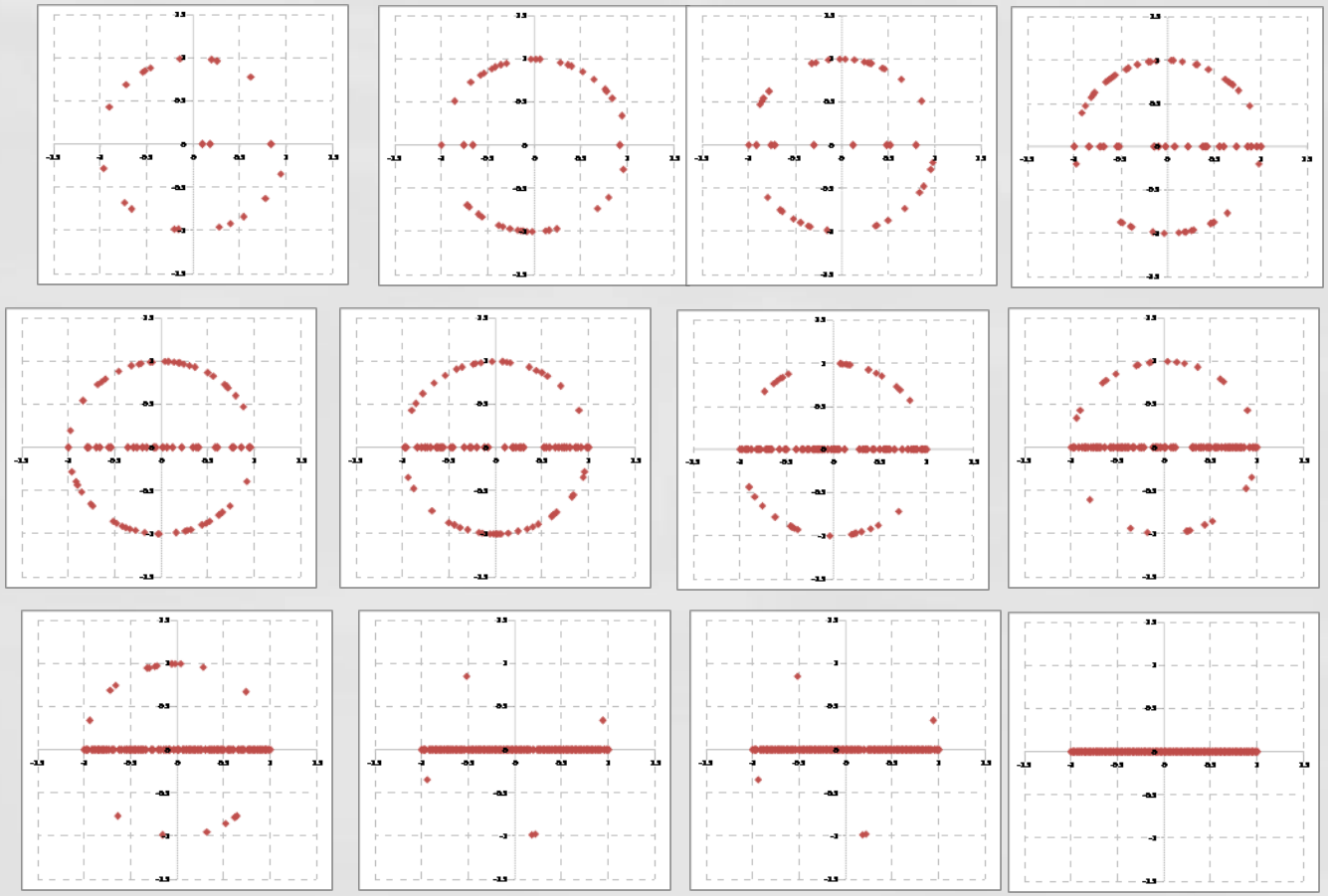
- Reality: $X^2 + Y^2 = 1$
  - Unit Circle Conic relationship



Issues in Attempting to Minimize Error by Expected Value Response

- Trained in 10% of data available increments
- If it gets two inputs the same it averages the responses in order to minimize error
- The more it trains the more potential for double binding ends up resulting in maximizing the error instead of minimizing it

Issues in Attempting to Minimize Error by Expected Value Response



Questions for Discussion

- Given the potential for double binding and the seemingly inapplicability of EV, how should we be measure predictive performance for models reflecting non-linear systems?
- Should we be cautious in purposefully subjecting learning systems to potential double blinding training sets?