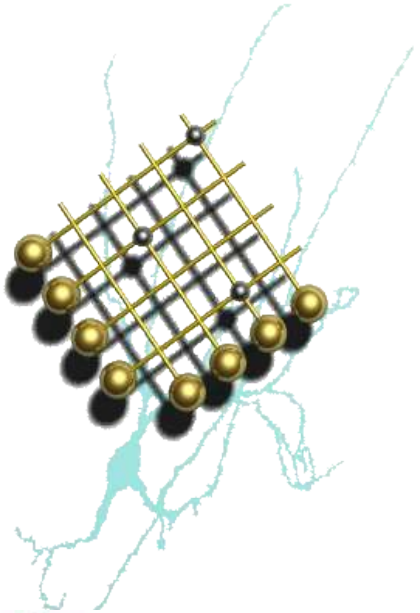
MEDPUB - Maximum Entropy Decisions about PUBlications

David Sterratt



Institute for
Adaptive & Neural Computation

School of
informatics

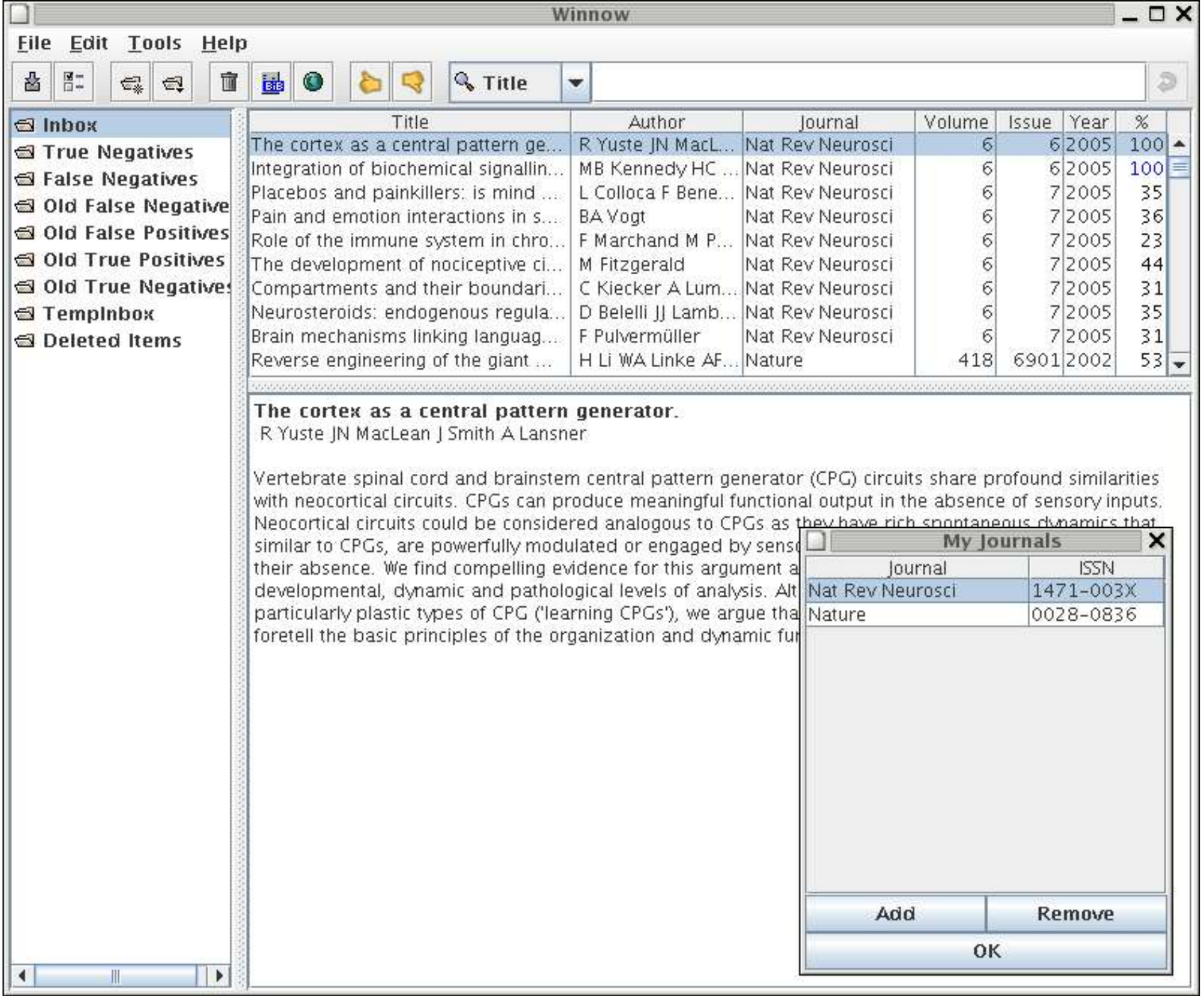
THE UNIVERSITY OF EDINBURGH

The problems with email tables of contents alerts

- Small fraction of articles relevant
- Hence alerts go unread
- To read abstracts and articles, researcher has to deal with a variety of journal-specific interfaces
- Setting up tables of contents alerts requires dealing with multiple websites too

A solution

- Collect the latest journal tables of contents from a database (PUBMED)
- Use machine learning algorithm trained by user to rank articles in order of interest
- Helpful GUI to displays abstracts, give easy access to PDFs, save references in BibTeX
- Which machine learning algorithm?


Winnow
File Edit Tools Help
Title
Inbox
True Negatives
False Negatives
Old False Negative
Old False Positives
Old True Positives
Old True Negatives
TempInbox
Deleted Items
Title Author Journal Volume Issue Year %
The cortex as a central pattern ge... R Yuste JN MacL... Nat Rev Neurosci 6 6 2005 100
Integration of biochemical signallin... MB Kennedy HC ... Nat Rev Neurosci 6 6 2005 100
Placebos and painkillers: is mind ... L Colloca F Bene... Nat Rev Neurosci 6 7 2005 35
Pain and emotion interactions in s... BA Vogt Nat Rev Neurosci 6 7 2005 36
Role of the immune system in chro... F Marchand M P... Nat Rev Neurosci 6 7 2005 23
The development of nociceptive ci... M Fitzgerald Nat Rev Neurosci 6 7 2005 44
Compartments and their boundari... C Kiecker A Lum... Nat Rev Neurosci 6 7 2005 31
Neurosteroids: endogenous regula... D Belelli JJ Lamb... Nat Rev Neurosci 6 7 2005 35
Brain mechanisms linking languag... F Pulvermüller Nat Rev Neurosci 6 7 2005 31
Reverse engineering of the giant ... H Li WA Linke AF... Nature 418 6901 2002 53
The cortex as a central pattern generator.
R Yuste JN MacLean J Smith A Lansner
Vertebrate spinal cord and brainstem central pattern generator (CPG) circuits share profound similarities with neocortical circuits. CPGs can produce meaningful functional output in the absence of sensory inputs. Neocortical circuits could be considered analogous to CPGs as they have rich spontaneous dynamics that
similar to CPGs, are powerfully modulated or engaged by sens
their absence. We find compelling evidence for this argument a
developmental, dynamic and pathological levels of analysis. Alt
particularly plastic types of CPG ('learning CPGs'), we argue tha
foretell the basic principles of the organization and dynamic fu
My Journals
Journal ISSN
Nat Rev Neurosci 1471-003X
Nature 0028-0836
Add
Remove
OK

Naïve Bayes Algorithm

Class $c \in \{\text{interesting}, \text{boring}\}$

Document d comprises N_w instances of word w

Likelihood of generating document d given class c:

$$p(d|c) \propto \prod_w \frac{p(w|c)^{N_w}}{N_w!}$$

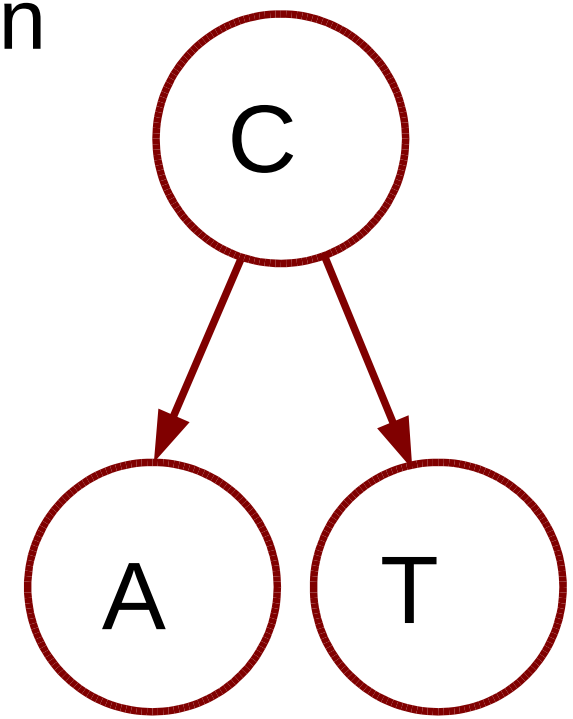
(Multinomial formula; independence; bag of words)

Bayes: $p(c|d) = \frac{p(d|c)\,p(c)}{\sum_{c'} p(d|c')\,p(c')}$

Estimate $p(w|c)$ and $p(c)$ from document set D

Different Fields

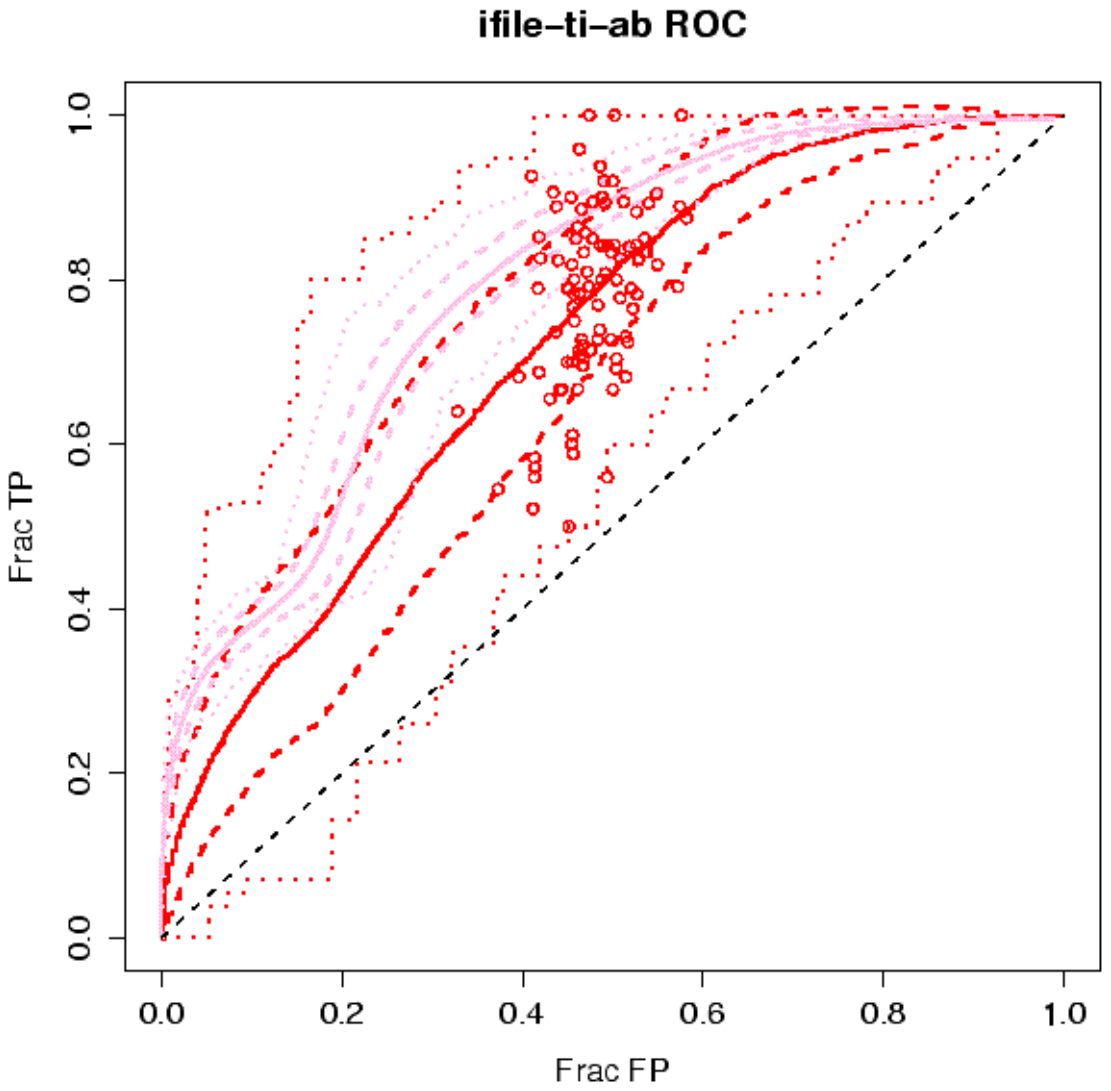
- Titles, abstract and authors contain different types of information
- Some articles contain only title & author information
  - e.g. Nature N&V, ZETOC alerts
- Hence have combine separate conditional probability tables
  - Bayesian chain rule



Testing performance (I)

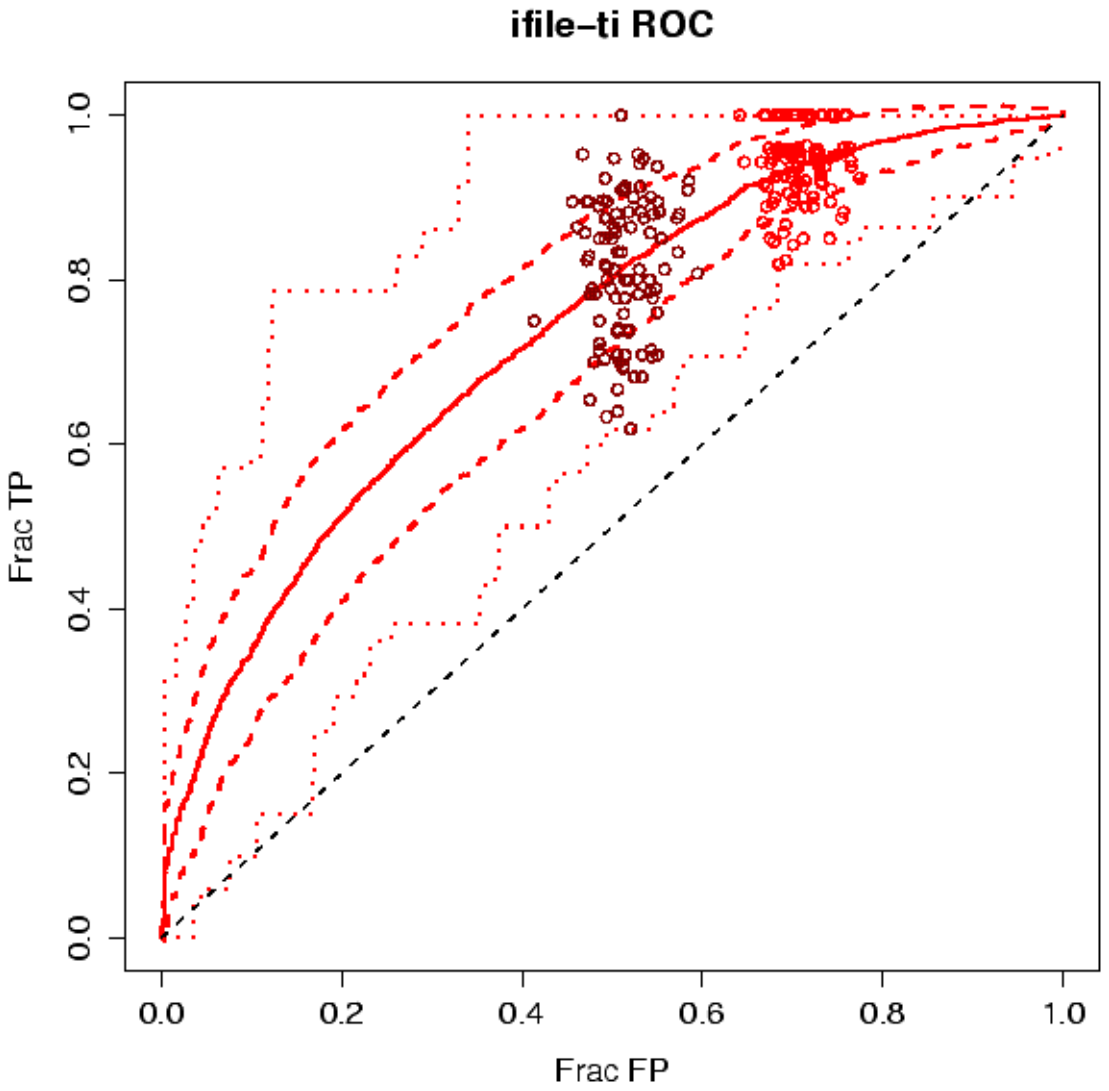
- Corpus of 2662 articles, 1047 with empty abstracts collected over 10-week period of testing software
- 218 interesting articles, 2444 boring
- Ten by tenfold crossvalidation procedure
- Naïve Bayes (ifile implementation)

Naïve Bayes: title and abstract (lumped together)



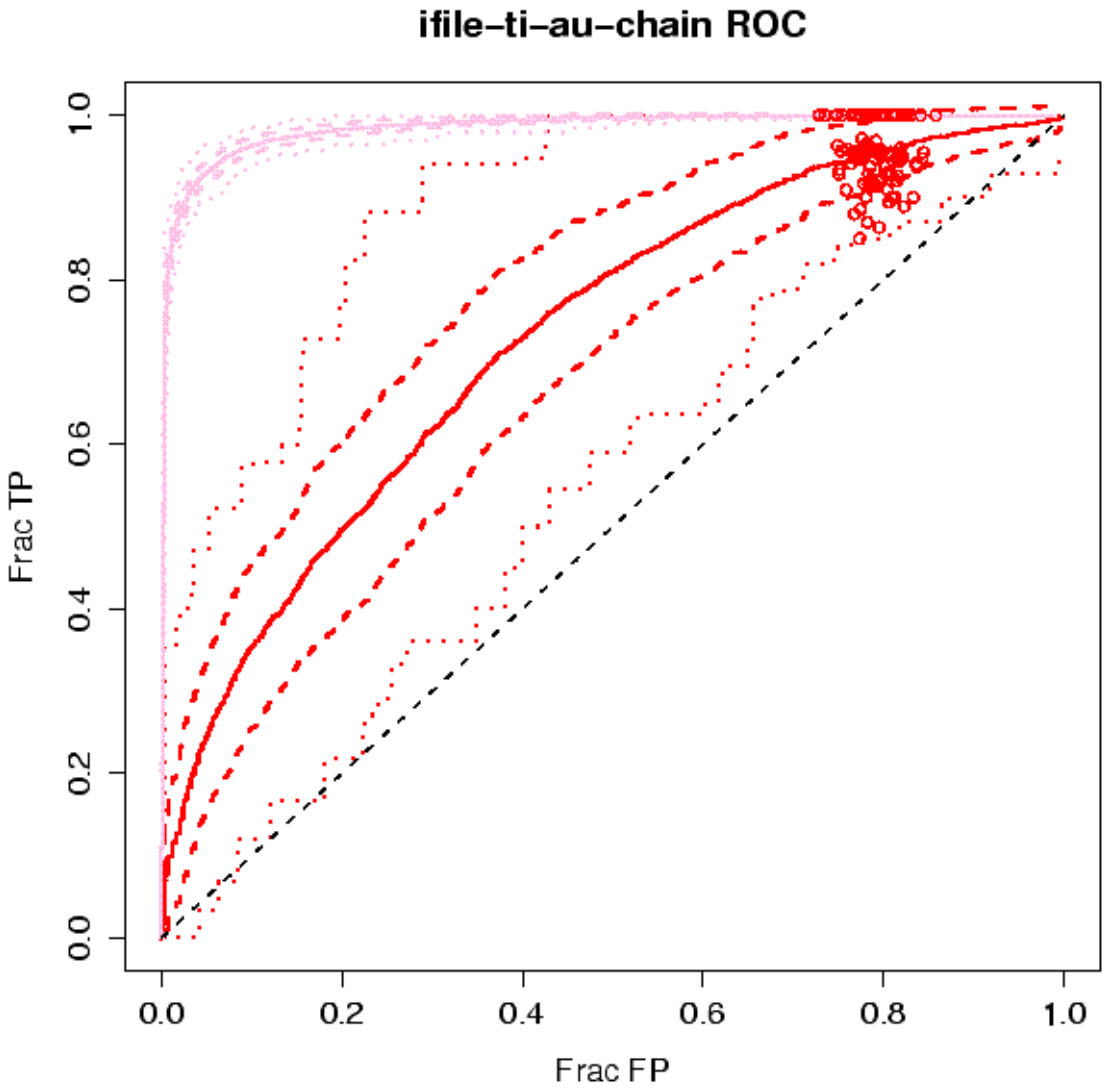
48±4% FP
78±11% TP

Naïve Bayes – titles only



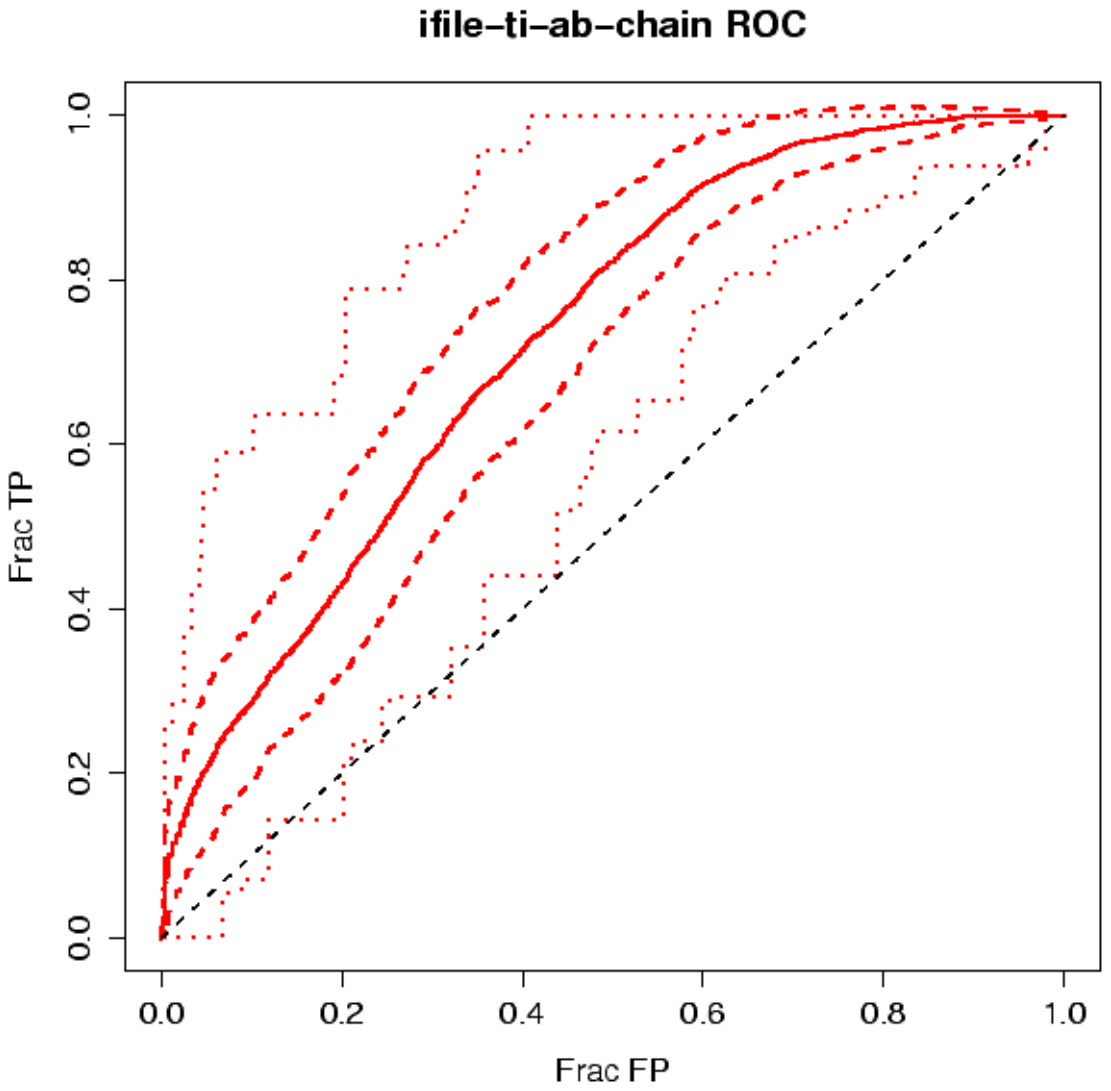
71±3% FP
94±5% TP

Naïve Bayes – titles and authors



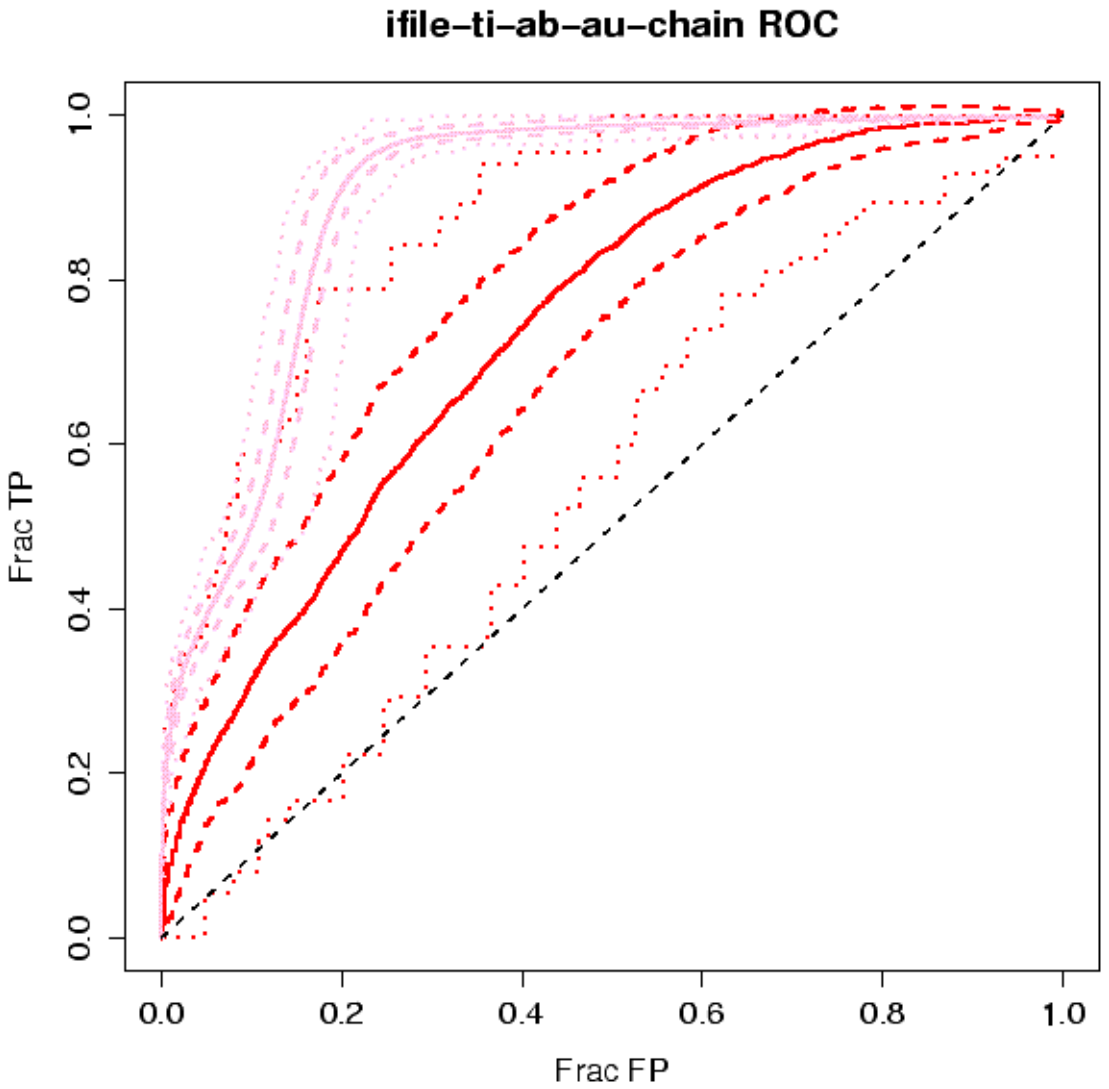
79±2% FP
96±4% TP

Naïve Bayes – Abstracts and Titles



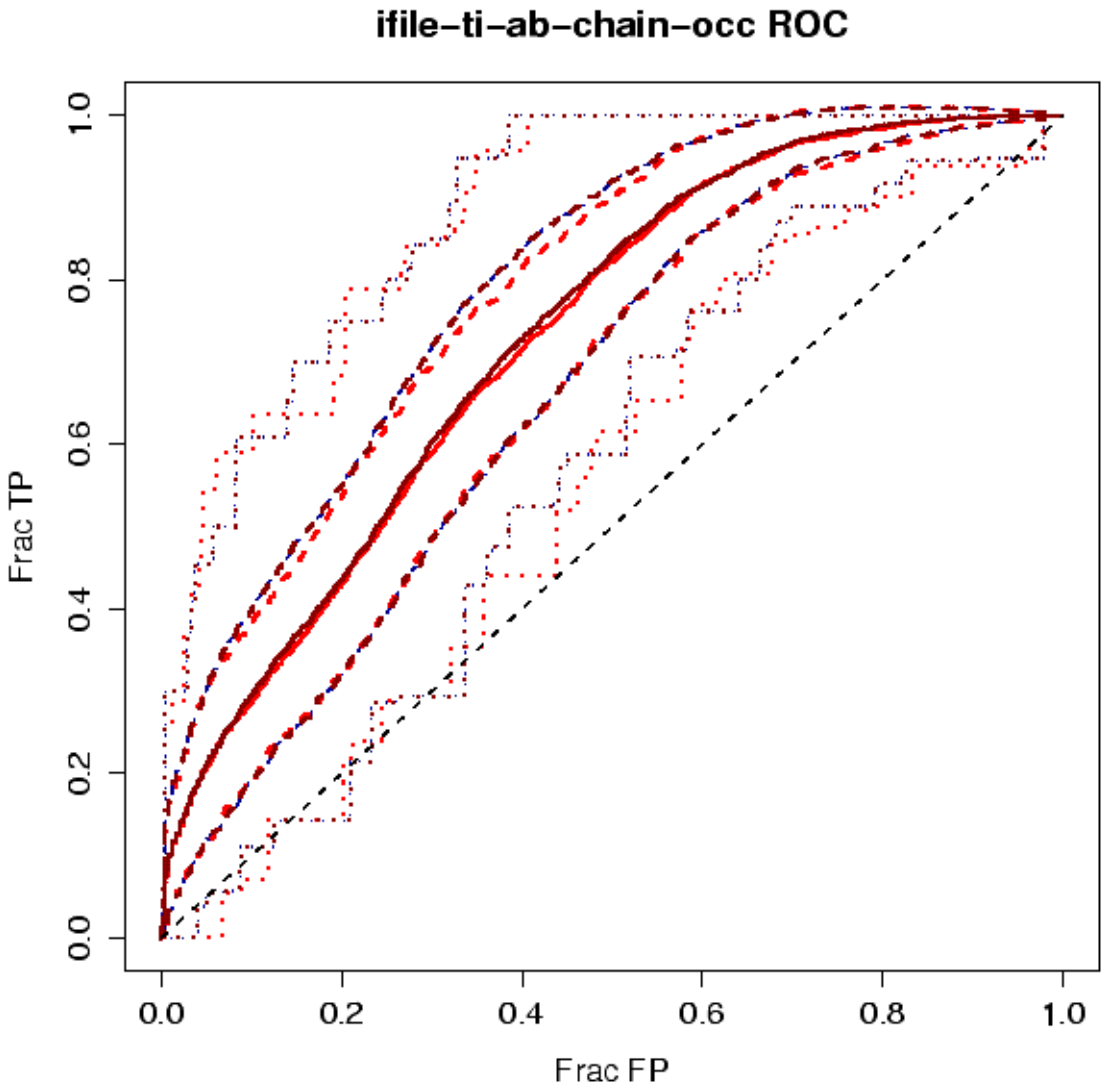
59±4% FP
90±7% TP

NB Titles + Abstracts + Authors



64±4% FP
93±6% TP

Naïve Bayes – occurrence vs counts



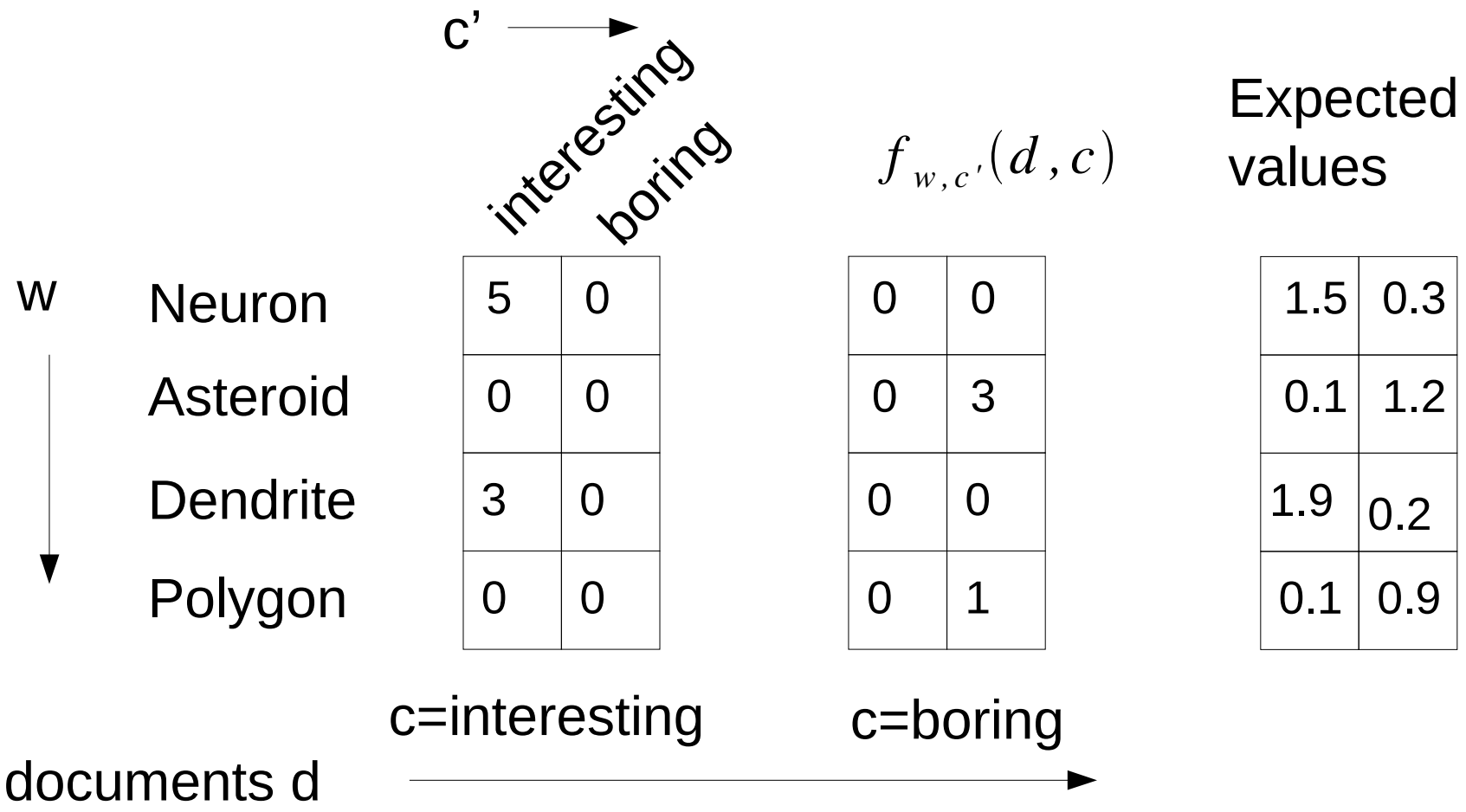
Can we do better?

- Work by Nigram et al. suggests that the Maximum Entropy algorithm outperforms Naïve Bayes

Maximum Entropy

- Choose a model that is consistent with the facts, but otherwise as uniform as possible
  - Modern formulation due to Jaynes, see also “Principle of Insufficient Reason” (Laplace), Occam’s Razor...
- e.g. if we have no information about document, assign 50% chance of it being interesting or boring

Features


c'
interesting
boring
$f_{w,c'}(d,c)$
Expected values
w
Neuron
Asteroid
Dendrite
Polygon
5 0
0 0
3 0
0 0
0 0
0 3
0 0
0 1
1.5 0.3
0.1 1.2
1.9 0.2
0.1 0.9
c=interesting
c=boring
documents d

Constraints

- Our distribution of class conditional on document should match expected value of each feature in data:

$$\frac{1}{|D|}\sum_{d\in D} f_{w,c}(d, c(d)) = \frac{1}{|D|}\sum_{d\in D} p(c|d)\sum_{c} f_{w,c}(d, c(d))$$

Maximum entropy

- Conditional entropy of distribution should be maximised subject to constraints

$$H(p) = -\sum_{c,d} p(d)\, p(c|d) \log p(c|d)$$

Solution

- The distribution has the exponential form

$$p(c|d)=\frac{1}{Z(d)}\prod_{w,c'} e^{\lambda_{w,c'} f_{w,c'}(d,c)}$$

- The parameters are adjusted to maximise the (log) likelihood of the data. This also maximises the entropy.

$$l(D)=\log \prod_{d\in D} p(c(d)|d)$$

Naïve Bayes versus MaxEnt

Naïve Bayes

$$p(c|d) \propto \tilde{p}(c) \prod_w \frac{\tilde{p}(w|c)^{N_w}}{N_w!}$$

Parameters derived directly from training data

Number of times w appears in d

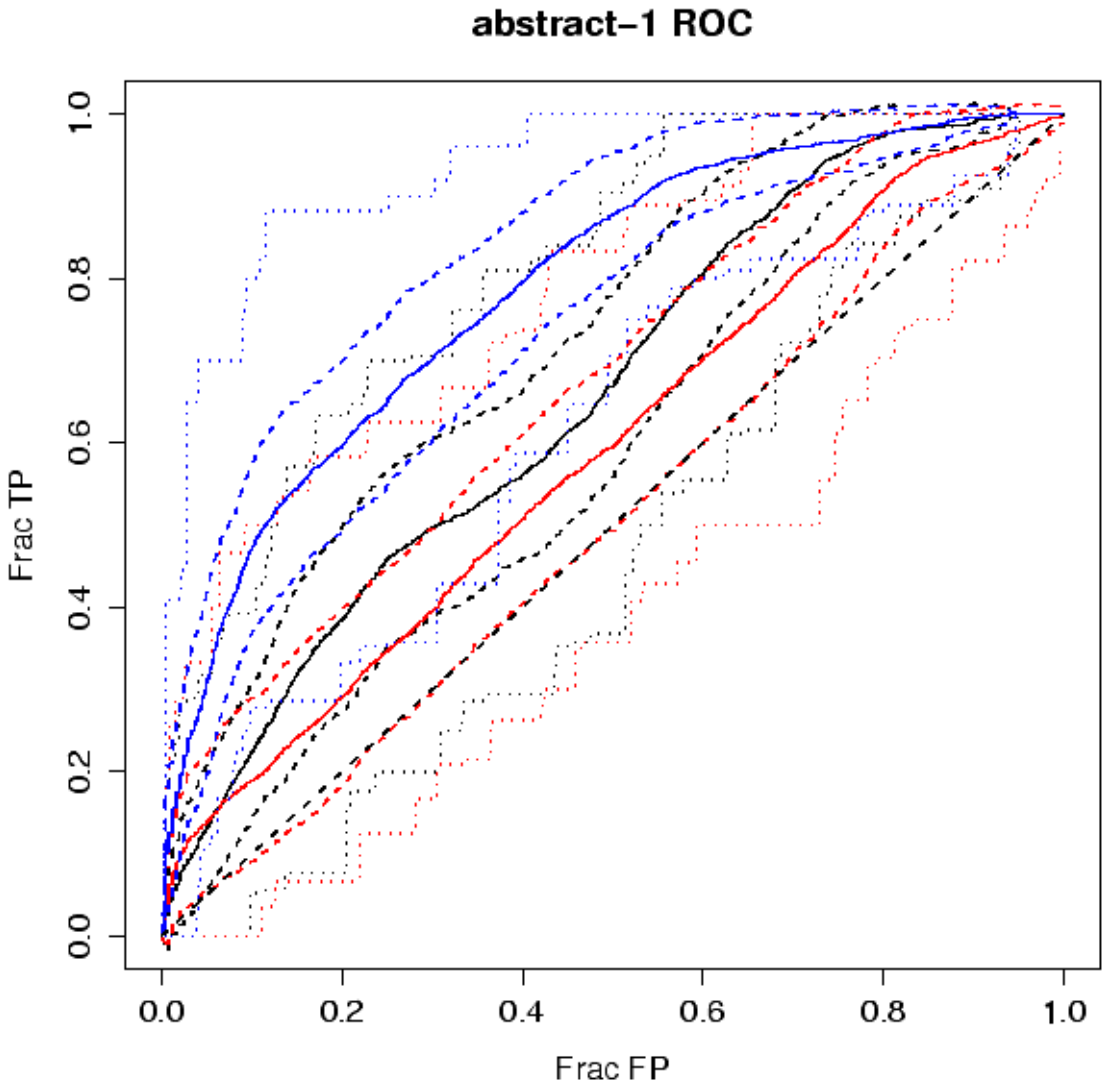
MaxEnt

$$p(c|d) \propto \prod_w e^{\lambda_{w,c} N_w}$$

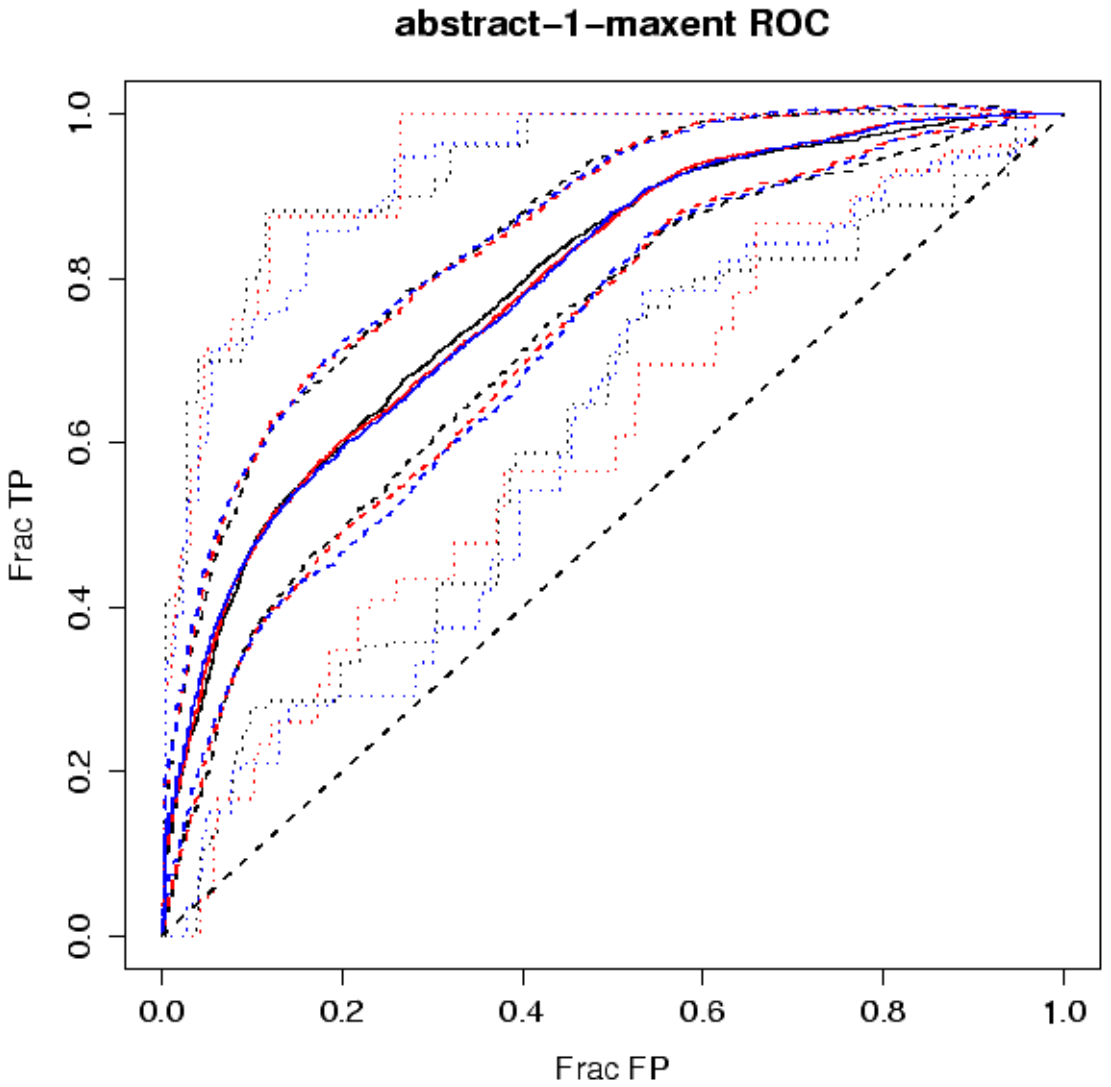
Testing performance (II)

- Similar to before, but only consider articles with abstracts, hence 1615 articles in corpus
- Test using open source, Java-based MALLET library implementations of various algorithms

Naïve Bayes, Decision Trees, Maxent



Maxent with bigrams and trigrams



Why does MaxEnt perform better?

- No independence assumptions (Nigram et al)
  - e.g. Naïve Bayes would count both *Boltzman* and *machine* in *Boltzman machine*
  - MaxEnt will discount the weights for these features so that their weight towards classification is approximately half
- Other reasons?

Conclusions

- MaxEnt and Naïve Bayes cut down on the number of uninteresting articles to be skimmed for finding a given fraction of interesting articles
  - Is the improvement worthwhile?
- MaxEnt > Naïve Bayes

The future?

- Already open source project on sourceforge
  - Put MaxEnt version there
- Algorithm improvements anyone?
  - Star rating system?
  - Performance improvements
- More data sources?