Text Categorization for Evidence Based Medicine

Automatic Identification of Purpose and Quality of Articles In Journals Of Internal Medicine

Yin Aphinyanaphongs M.S. ,
Constantin Aliferis M.D., Ph.D.
Department of Biomedical Informatics
Vanderbilt University
AMIA symposium, November 8-12, 2003

Research Focus

- Evidence Based Medicine
  - Main tenet: Clinical decisions based on solid clinical evidence.
  - 3 steps:
    - Obtaining the best evidence
    - Appraising it
    - Applying it

A Practical Example

- Let’s suppose a physician has a patient with Congestive Heart Failure and is exploring the latest treatment options.
- As a start, the physician tries the following MeSH PubMed query

- "Heart Failure, Congestive/diet therapy"[MeSH] OR "Heart Failure, Congestive/drug therapy"[MeSH] OR "Heart Failure, Congestive/therapy"[MeSH])

Using PubMed Only (10/29/2003)

1. Gadolinium cardiovascular magnetic resonance predicts reversible myocardial dysfunction and remodeling in patients with heart failure undergoing beta-blocker therapy.
2. Hospitalization rates and length of stay for cardiovascular conditions in Canada, 1994 to 1999.
3. Hormone replacement therapy is associated with improved survival in women with advanced heart failure.
4. Noncardiac comorbidity increases preventable hospitalizations and mortality among Medicare beneficiaries with chronic heart failure.
5. **Effect of trapidil on cardiovascular events in patients with coronary artery disease (results from the Japan Multicenter Investigation for Cardiovascular Diseases-Mochida [JMIC-M]).**

•
•
•

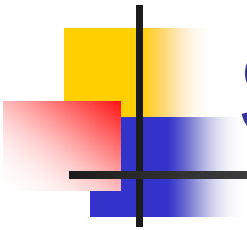
Getting to this.

1. **Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction.**
2. **Hawthorn extract for treating chronic heart failure: meta-analysis of randomized trials.**
3. **Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study).**
4. **Vasopressin V2-receptor blockade with tolvaptan in patients with chronic heart failure: results from a double-blind, randomized trial.**
5. **Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial.**

-
-
-

Defining High Quality Evidence

- Evidence Based Medicine emphasizes use of studies with excellent methodology.
- In treatment studies, good design includes:
  - Double blind, controlled, randomized.
  - Sufficient sample, good follow-up, pertinent clinical outcomes.



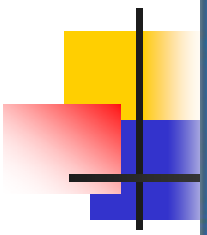
Solution

- From an information retrieval/ computer science perspective, this class of problems can be solved by:
  - Having a pre-classified document collection
  - Suitable classifier that labels new articles.
- This is the methodology used by Dr. Brian Haynes and colleagues.
  - Document collection built by reviewers.
  - Brute force classifier used to build Boolean queries.

PubMed's Clinical Query Filters

- The discovered sensitivity optimized treatment filter used is:
  - "Randomized Controlled Trial" [Publication Type] OR "Drug Therapy" [MeSH Subheading] OR "Therapeutic Use" [MeSH Subheading] OR "Random" [Textword]

Haynes, B., et al., *Developing Optimal Search Strategies for Detecting Sound Clinical Studies in MEDLINE.* JAMIA, 1994. **1**(6): p. 447-458.



NCBI

PubMed Clinical Queries

PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | OMIM | Books

About Entrez

Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Related Resources
Order Documents
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

Privacy Policy

Select from two filters to limit your retrieval. Choose either Clinical Queries or Systematic Reviews. Enter your search topic in the box below and click Go.

◉ Clinical Queries using Research Methodology Filters

This specialized search is intended for clinicians and has built-in search "filters" based largely on Haynes RB et al.. Four study categories are provided, and the emphasis may be more sensitive (i.e., most relevant articles but probably some less relevant ones) or more specific (i.e., mostly relevant articles but probably omiting a few). See filter table for details.

Indicate the category and emphasis below:

Category: ◉ therapy ○ diagnosis ○ etiology ○ prognosis
Emphasis: ○ sensitivity ◉ specificity

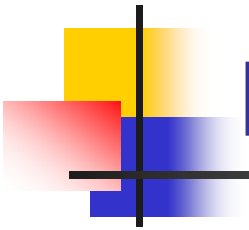
○ Systematic Reviews

This feature retrieves systematic reviews and meta-analysis studies for your search topic(s). For more information, see Help. Related sources are also provided.

Enter subject search:

[ ] Go Clear

Note: If you want to retrieve everything on a subject area, you should not use this screen. The objective of filtering is to reduce the retrieval to articles that report research conducted with specific methodologies.



Hypothesis

- Possible to improve both aspects of CQF construction by improving...
  - The operational gold standard.
  - The learning algorithms.
- Advances in machine learning provide many opportunities for the latter.
- However, selecting an improved operational gold standard is more challenging.

Operational gold standard for high quality articles

- The ACP journal club.
- They bring together world class editorial members whose job is to apply criteria to articles in the top journals in internal medicine.

Gold standard selection criteria

- The treatment criteria -ACP journal club
  - “Random allocation of participants to comparison groups.”
  - “80% follow up of those entering study.”
  - “Outcome of known or probable clinical importance.”

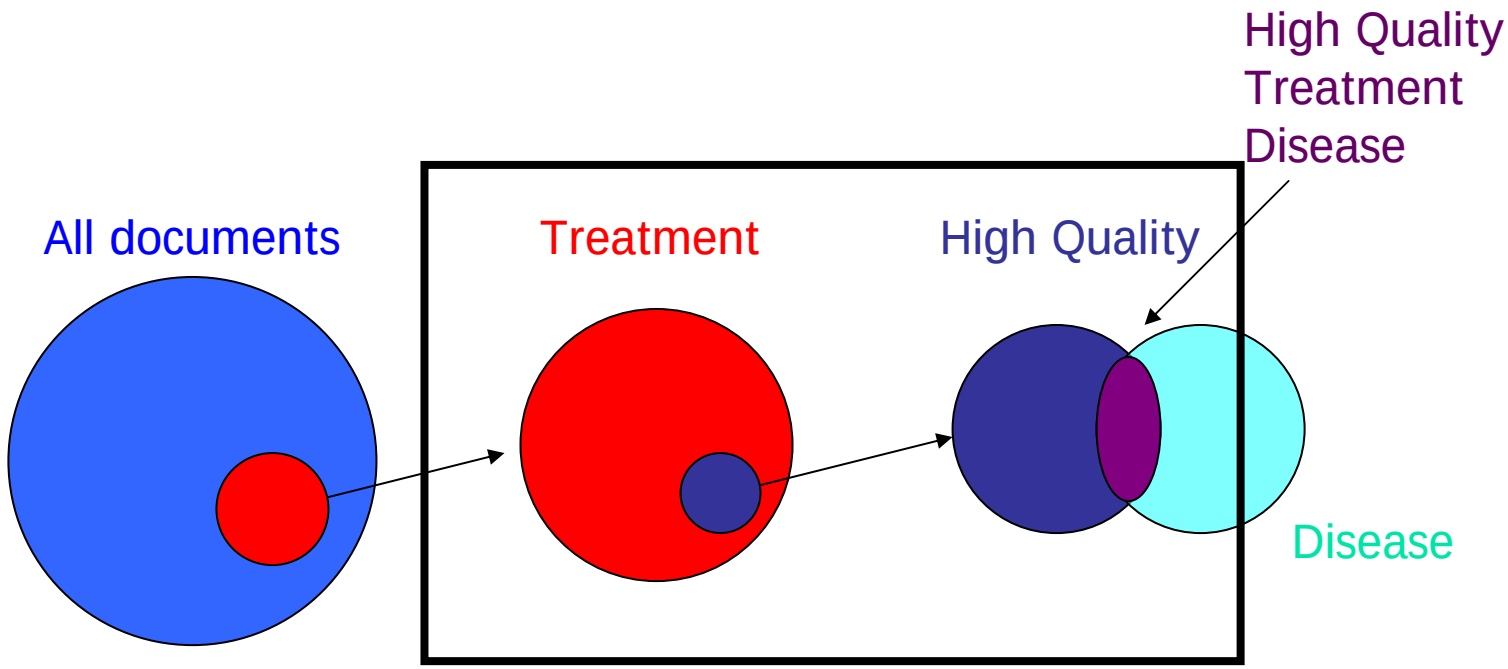
Purpose and Procedure. ACP Journal, 1999. **131**(1): p. A-15 - A-16.

Content Category

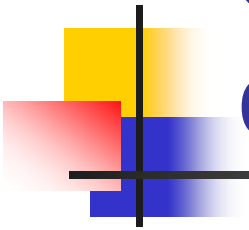
- As a proof of concept, for this study, we identify high quality articles pertaining to treatment.
  - Good candidate as it is the most often asked clinical question.

Jerome, R.N., et al., Information Needs of clinical teams: analysis of questions received by the Clinical Informatics Consult Service. Bull Med Libr Assoc, 2001. 89(2): p. 177-184.

Overall process



Steps for a text categorization engine.

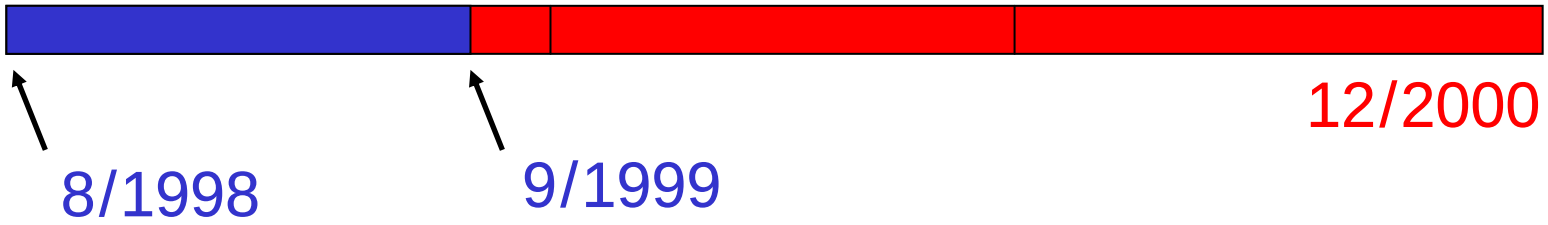


- Corpus Construction
- Document Representation
- Study Design
- Train classifiers.
- Evaluate the classifiers.

Select the Journals

- Age and ageing
- AMERICAN JOURNAL OF CARDIOLOGY
- AMERICAN JOURNAL OF EPIDEMIOLOGY
- AMERICAN JOURNAL OF GASTROENTEROLOGY
- AMERICAN JOURNAL OF MEDICINE
- AMERICAN JOURNAL OF PUBLIC HEALTH
- AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE
- ANNALS OF EMERGENCY MEDICINE
- ANNALS OF INTERNAL MEDICINE
- ANNALS OF MEDICINE
- ARCHIVES OF FAMILY MEDICINE
- ARCHIVES OF INTERNAL MEDICINE
- ARCHIVES OF NEUROLOGY
- ARTHRITIS AND RHEUMATISM
- BRITISH MEDICAL JOURNAL
- BRITISH JOURNAL OF GENERAL PRACTICE
- CANADIAN MEDICAL ASSOCIATION JOURNAL
- CANADIAN JOURNAL OF CARDIOLOGY
- CANADIAN JOURNAL OF GASTROENTEROLOGY
- Chest
- Circulation
- CLINICAL AND INVESTIGATIVE MEDICINE
- CRITICAL CARE MEDICINE
- Diabetes Care
- Gastroenterology
- Gut
- Heart
- Hypertension
- J Am Board Fam Pract
- JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
- JOURNAL OF THE AMERICAN GERIATRICS SOCIETY
- JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION
- JOURNAL OF CLINICAL EPIDEMIOLOGY
- JOURNAL OF FAMILY PRACTICE
- JOURNAL OF GENERAL INTERNAL MEDICINE
- JOURNAL OF INFECTIOUS DISEASES
- JOURNAL OF INTERNAL MEDICINE
- JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY
- JOURNAL OF VASCULAR SURGERY
- JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
- Lancet
- MEDICAL CARE
- MEDICAL JOURNAL OF AUSTRALIA
- NEW ENGLAND JOURNAL OF MEDICINE
- Neurology
- Pain
- Spine
- Stroke
- Thorax

Downloading and marking articles in the study period.

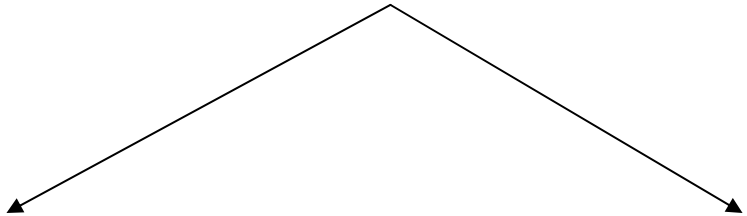


Downloaded all articles from the journals in the study period.

Review ACP Journal from 8/1998 to 12/2000 for articles that are cited by the ACP.

Corpus Composition

For the study period from 8/98 – 9/99
15803 original articles



396 high quality Treatment articles

15407 non-treatment articles

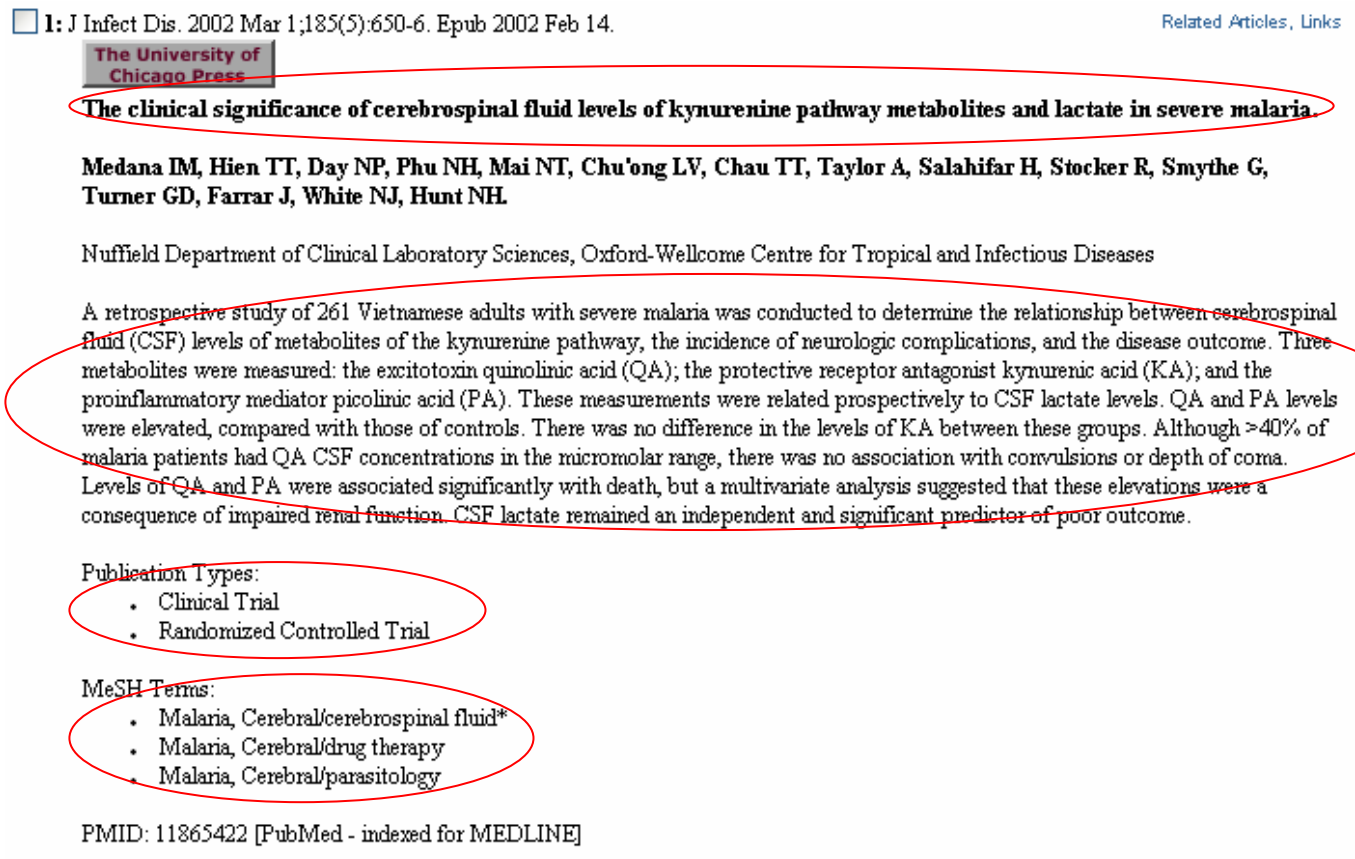
What words to use?

1

2

3

4



1: J Infect Dis. 2002 Mar 1;185(5):650-6. Epub 2002 Feb 14.

Related Articles, Links

The University of Chicago Press

The clinical significance of cerebrospinal fluid levels of kynurenine pathway metabolites and lactate in severe malaria.

Medana IM, Hien TT, Day NP, Phu NH, Mai NT, Chu'ong LV, Chau TT, Taylor A, Salahifar H, Stocker R, Smythe G, Turner GD, Farrar J, White NJ, Hunt NH.

Nuffield Department of Clinical Laboratory Sciences, Oxford-Wellcome Centre for Tropical and Infectious Diseases

A retrospective study of 261 Vietnamese adults with severe malaria was conducted to determine the relationship between cerebrospinal fluid (CSF) levels of metabolites of the kynurenine pathway, the incidence of neurologic complications, and the disease outcome. Three metabolites were measured: the excitotoxin quinolinic acid (QA); the protective receptor antagonist kynurenic acid (KA); and the proinflammatory mediator picolinic acid (PA). These measurements were related prospectively to CSF lactate levels. QA and PA levels were elevated, compared with those of controls. There was no difference in the levels of KA between these groups. Although >40% of malaria patients had QA CSF concentrations in the micromolar range, there was no association with convulsions or depth of coma. Levels of QA and PA were associated significantly with death, but a multivariate analysis suggested that these elevations were a consequence of impaired renal function. CSF lactate remained an independent and significant predictor of poor outcome.

Publication Types:

- Clinical Trial
- Randomized Controlled Trial

MeSH Terms:

- Malaria, Cerebral/cerebrospinal fluid*
- Malaria, Cerebral/drug therapy
- Malaria, Cerebral/parasitology

PMID: 11865422 [PubMed - indexed for MEDLINE]

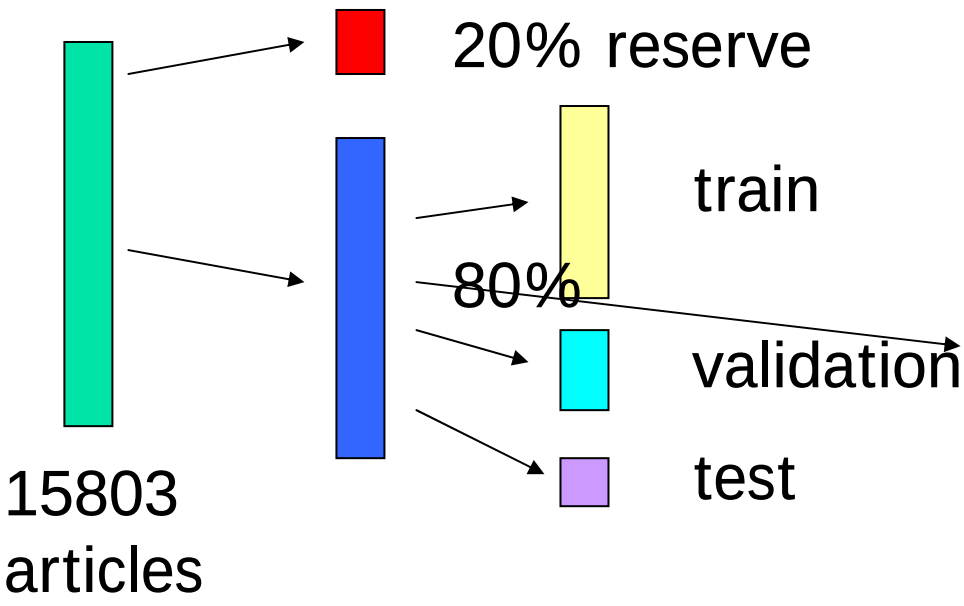
Document Preparation

- "The clinical significance of cerebrospinal."

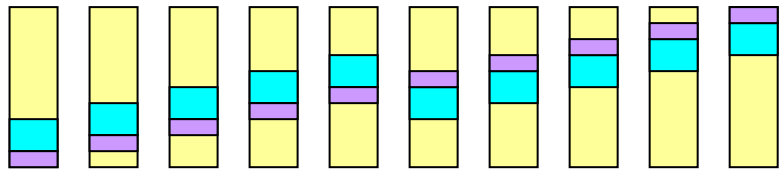
1. Representation
   - "The", "clinical", "significance", "of", "cerebrospinal"
2. Stop word removal
   - "Clinical", "Significance", "Cerebrospinal"
3. Porter Stemming (i.e. getting the roots of words)
   - "Clinic*", "Signific*", "Cerebrospin*"
4. Word weighting
   - This number is a statistical measure of how well the word separates the data into the 2 categories.
   - The measure is called log frequency with redundancy.

Study Design

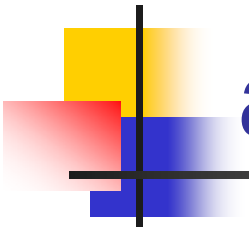
- Build a model.
- Estimate the performance of the methodology.



10 fold cross validation to estimate performance and error



Train the classifier with learning algorithms



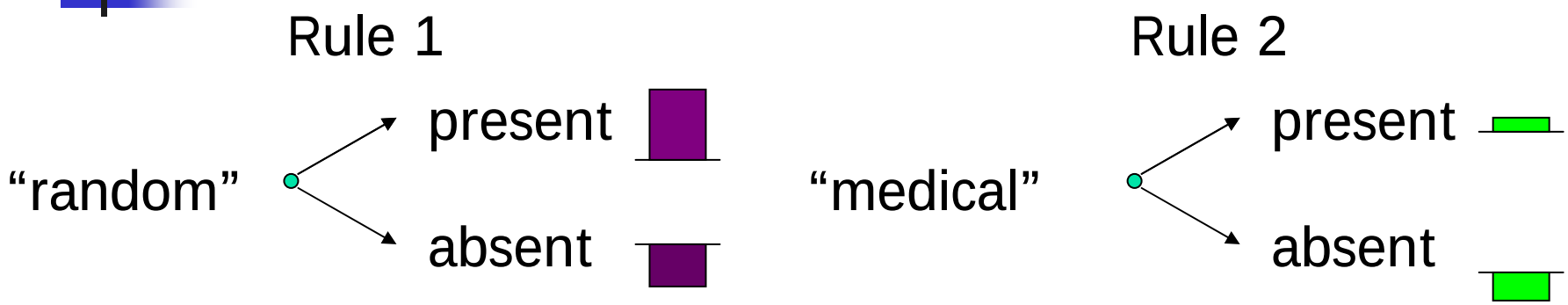
- 4 classifiers
  - Naïve Bayes
  - Decision Trees with Boosting
  - Linear Support Vector Machines
  - Polynomial Support Vector Machines

Naïve Bayes

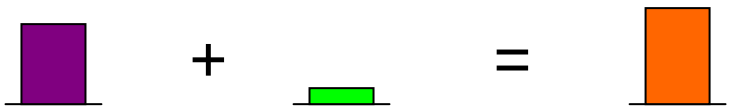
- A baseline learning algorithm used in text categorization.
- Application of Simple Bayes

$$C_{learned} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_i P(w_i \mid c_j)$$

Decision Trees with Boosting

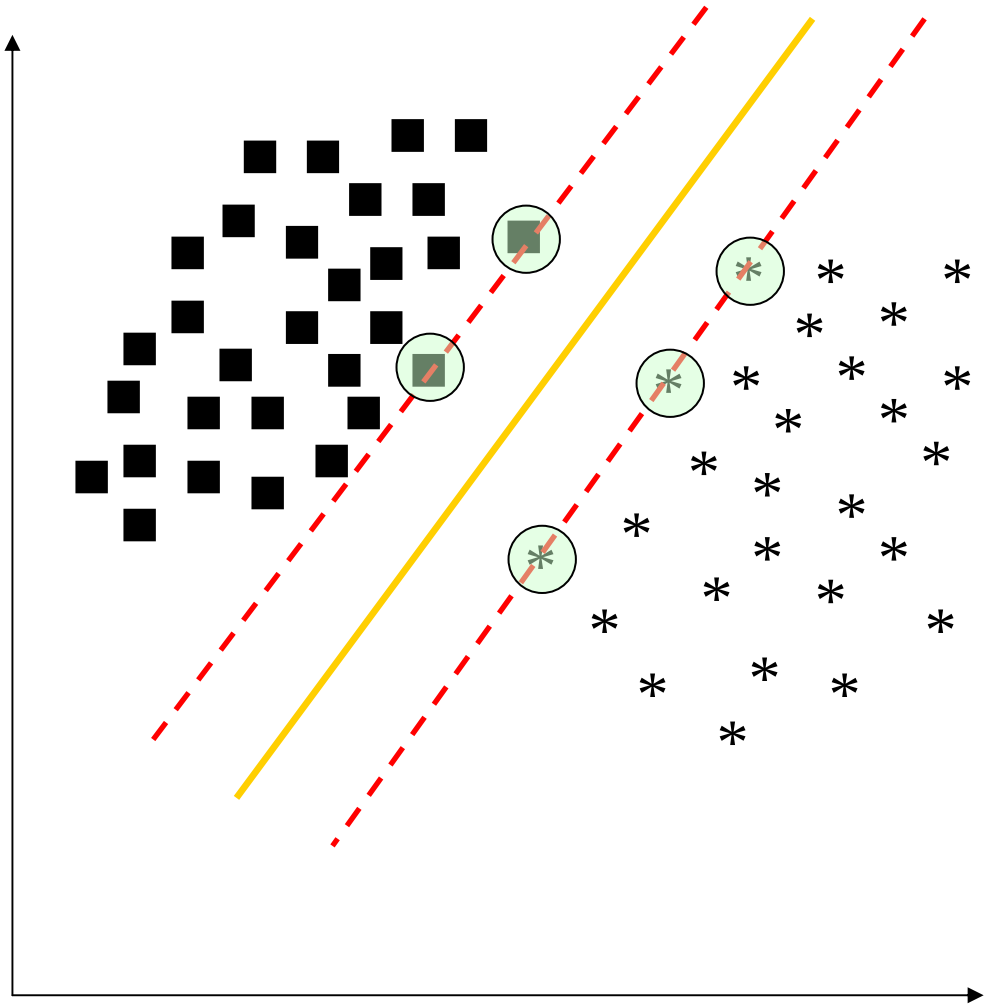


Document 1: <random> <medical>

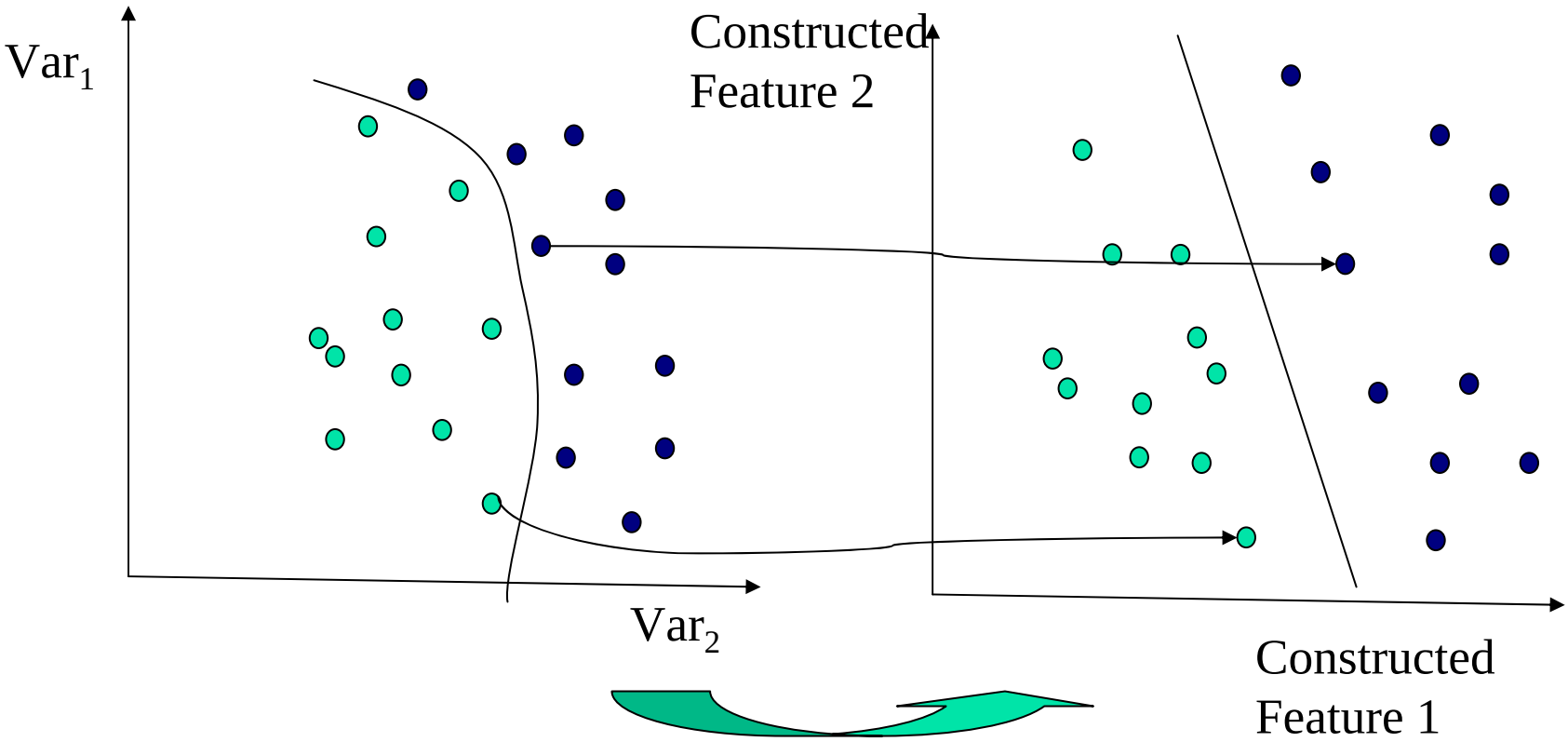


In class

Linear Support Vector Machine



Non-linear Support Vector Machine



Find function Φ(x) to map to a different space

Results – Area under the ROC curve

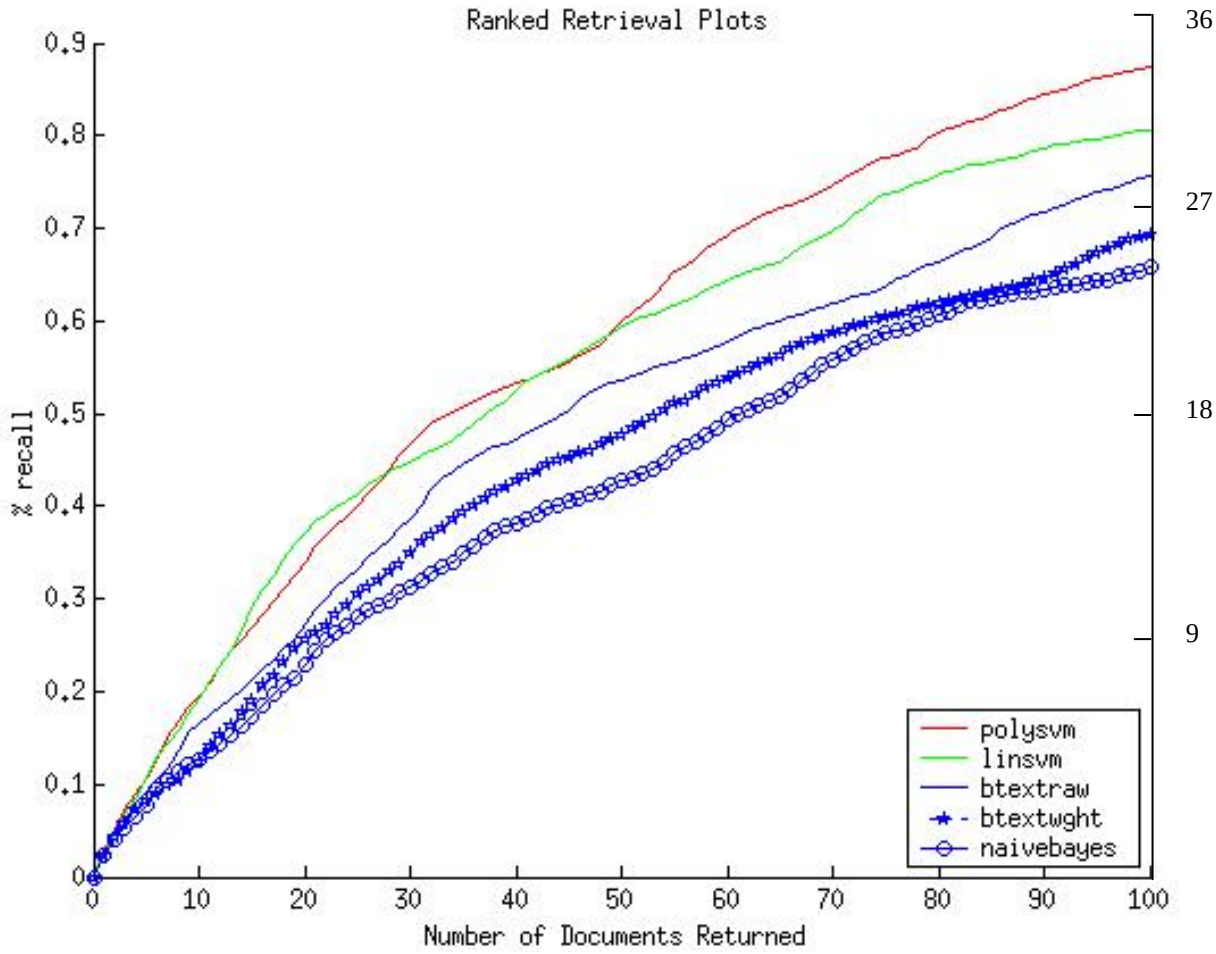
Classifiers	Average AUC over 10 folds	Range over 10 folds	p-value compared to largest
LinSVM	0.965	0.948 – 0.978	0.01
PolySVM	**0.976**	**0.970 – 0.983**	**N/A**
Naïve Bayes	0.948	0.932 – 0.963	0.001
Boost Raw	0.957	0.928 – 0.969	0.001
Boost Wght	0.941	0.900 – 0.958	0.001

Results – PubMed CQF vs. Our Methods (with 95% CI)

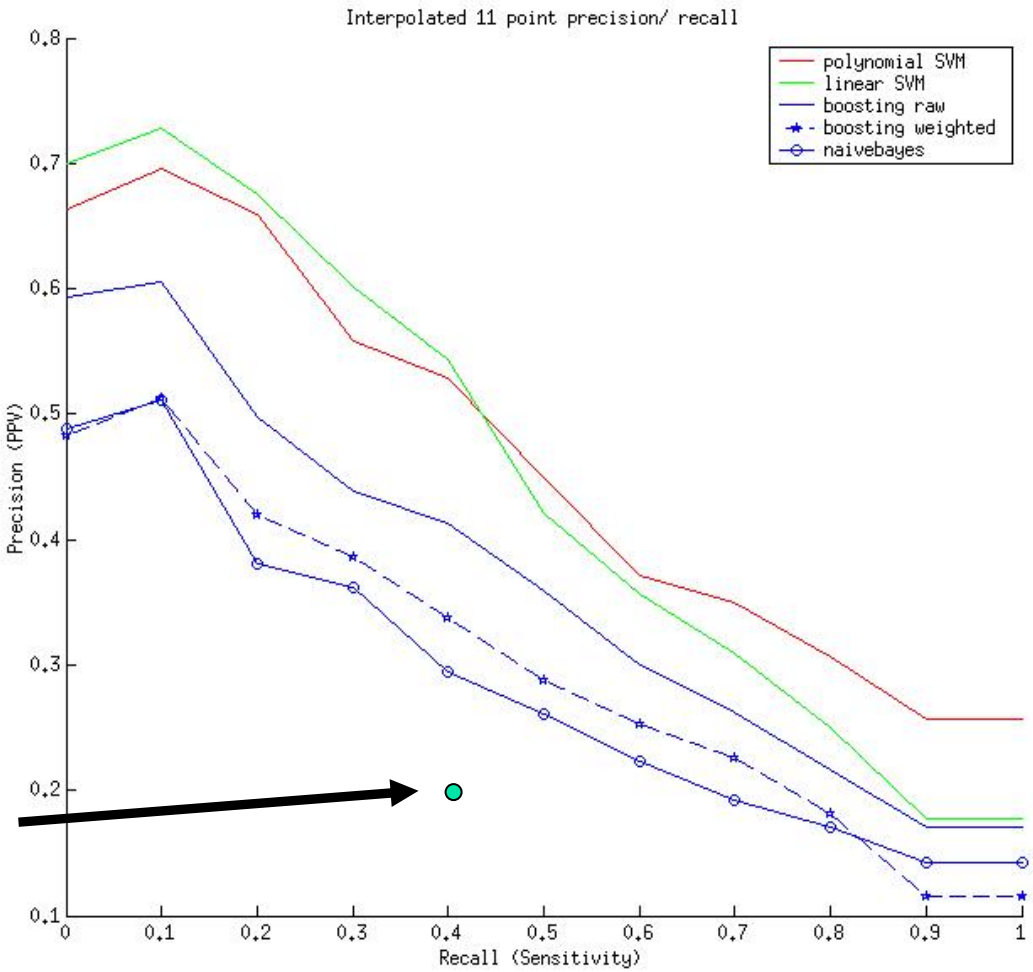
	Sensitivity	Specificity	Precision	Number Needed to Read (Average)
CQF	0.96	0.75	0.071	**14**
Poly SVM	0.9673 (0.830-0.99)	0.8995 (0.884-0.914)	0.1744 (0.120-0.240)	**6**

	Sensitivity	**Specificity**	Precision	Number Needed to Read (Average)
CQF	0.367	0.959	0.149	**6.7**
Poly SVM	0.8181 (0.641-0.93)	0.959 (0.948-0.97)	0.2816 (0.191-0.388)	**3.55**

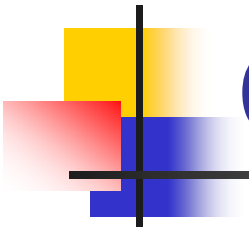
Retrieving High Quality Articles



Precision Recall Comparisons



Clinical Query Filter Performance



Conclusions

- These experiments provide evidence in support of the following two conclusions:
  - Text categorization methods can learn models of quality and content using the operational gold standard of ACP inclusion or citation. Polynomial SVMs perform the best in the above task.
  - Using a variety of evaluation metrics , models built with these learning methods and gold standard outperform the clinical query filters of PubMed.

Acknowledgements

- MSTP Program and NLM for funding.
- My advisor Dr. Constantin Aliferis and my other committee members