the hedge algorithm and applications

Virgil Pavlu

will it rain tomorrow ?

- ☒ I say it will rain
- ☑ Nick says it will not rain
- ☒ CNN says it will rain
- ☑ FOX says it will not rain

- TOMORROW : sunny all day…
- what about Saturday?

weighted majority algorithm

- Give the “weather-men” weights
  - Initial uniform or according to some prior belief
- Run the forecast for several days. Every day :
  - Make our prediction by weighted-majority-vote
  - Get the real outcome
  - Update weights
    - “penalize” wrong predictors
    - “reward” good predictors

more problems

- more problems
  - trading stocks
  - IR metasearch
  - disease classification
    - where we need to “train” doctors
- common underlying idea
  - generalization of weighted-majority
- why *not* exactly WM
  - losses are real numbers instead of discrete 0-1
  - our loss may be a weighted sum of expert losses

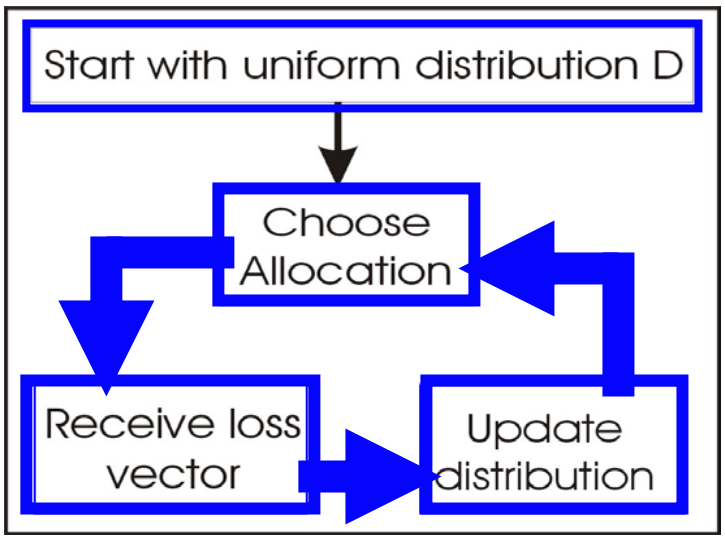
this talk

- Hedge algorithm for online allocation [schapire, freund ’96]
- Applications
  - On classification : Adaboost
  - On IR : metasearch algorithm

"how to use expert advice" ?

- N experts (or strategies)
- maintain a set of weights over experts
- loop for T episodes t=1,2,…..,T
  - allocate Resources (believe) - based on weights
  - receive loses
  - reweight the experts

online allocation - hedge algorithm



$\beta \in [0,1]$; N strategies (systems)

initial weights $\mathrm{w}^1 \in [0,1]^N ; \sum_{i=1}^{N} w_i^1 = 1$

$$p_i^t = \frac{w_i^t}{\sum_{i=1}^{N} w_i^t}$$

$$l^t \in [0,1]^N$$

$$L_H^t = p^t \cdot l^t$$

$$w_i^{t+1} = w_i^t \cdot \beta^{l_i^t}$$

$$L_{HEDGE} = \sum_{t=1}^{T} p^t \cdot l^t$$

why hedge

- goal : Hedge loss close to the Loss of the best expert (bound)
- proof idea :
  - relate episodic Hedge loss with the sum of weights

$$\sum_{i=1}^{N} w_i^{t+1} \le \left(\sum w_i^t\right)\left(1-(1-\beta)L^t_{Hedge\ (\beta)}\right)$$

  - relate cumulate Hedge loss with sum of final weights

$$\ln\left(\sum_{i=1}^{N} w_i^{T+1}\right) \le -(1-\beta)L_{\mathbf{Hedge}(\beta)}$$

  - relate sum of final weights with loss of the best expert

$$L_{Hedge(\beta)} \le \frac{\ln(1/\beta)}{1-\beta} L_{best} + \frac{1}{1-\beta}\ln(N)$$

episodic hedge loss

$$\sum_{i=1}^{N} w_i^{t+1} \le \left(\sum w_i^t\right)\left(1-(1-\beta)L_{Hedge}^t\right)$$

$$\sum_{i=1}^{N} w_i^{t+1} = \sum_{i=1}^{N} w_i^t \beta^{l_i^t}$$

$$\le \sum_{i=1}^{N} w_i^t (1-(1-\beta) l_i^t)$$

$$= \sum_{i=1}^{N} p_i^t \left(\sum w_i^t\right)(1-(1-\beta) l_i^t)$$

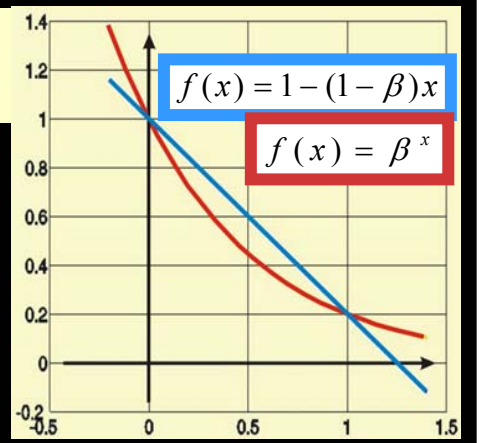
$$= \left(\sum w_i^t\right)\sum_{i=1}^{N} \left(p_i^t - (1-\beta) p_i^t l_i^t\right)$$

$$= \left(\sum w_i^t\right)\left(1-(1-\beta)\sum_{i=1}^{N} p_i^t l_i^t\right)$$

$$= \left(\sum w_i^t\right)\left(1-(1-\beta)\mathrm{p}^t \cdot 1^t\right)$$

$\beta, l \in [0,1]$

$\beta^l \le 1-(1-\beta) l$



$$p_i^t = \frac{w_i^t}{\sum_{i=1}^{N} w_i^t}$$

$$L_{Hedge}^t = \mathrm{p}^t \cdot 1^t$$

cumulative hedge loss

$$\ln\left(\sum_{i=1}^{N} w_i^{T+1}\right) \le -(1-\beta)L_{\mathbf{Hedge}(\beta)}$$

$$\sum_{i=1}^{N} w_i^{t+1} \le \left(\sum w_i^t\right)\left(1-(1-\beta)L_{Hedge}^t\right)$$

Applying repeatedly for $t = 1, \ldots, T$ yields

$$\sum_{i=1}^{N} w_i^{t+1} \le \left(\sum w_i^1\right)\prod_{t=1}^{T}\left(1-(1-\beta)\mathbf{p}^t \cdot \mathbf{l}^t\right)$$

$$\le \left(\sum w_i^1\right)\prod_{t=1}^{T}\exp\left(-(1-\beta)\mathbf{p}^t \cdot \mathbf{l}^t\right)$$

$$\le \exp\left(-(1-\beta)\sum_{t=1}^{T}\mathbf{p}^t \cdot \boldsymbol{\ell}^t\right)$$

$$1-x \le \exp(x)$$

$$L_{Hedge} = \sum_{t=1}^{T} p^t \cdot l^t$$

[almost] as good as the best expert

$$\ln\left(\sum_{i=1}^{N} w_i^{T+1}\right) \leq -(1-\beta)L_{\mathbf{Hedge}(\beta)}$$

$$\sum_{i=1}^{N} w_i^{T+1} \geq w_{any}^{T+1} = w_{any}^{1}\,\beta^{L_{any}}$$

$$w_{any}^{1} = \frac{1}{N}$$

$$L_{Hedge(\beta)} \leq \frac{-\ln\left(\frac{1}{N}\right) - L_{any}\ln\beta}{1-\beta}$$

optimality

$$L_{Hedge(\beta)} \le \frac{-\ln\left(\frac{1}{N}\right) - L_{best}\ln\beta}{1-\beta} = \frac{\ln(1/\beta)}{1-\beta}L_{best} + \frac{1}{1-\beta}\ln(N)$$

Theorem 3 *Let B be an algorithm for the on-line allocation problem with an arbitrary number of strategies. Suppose that there exist positive real numbers a and c such that for any number of strategies N and for any sequence of loss vectors $\ell^1, \ldots, \ell^T$*

$$L_B \le cL_{best} + a\ln(N)$$

Then for all $\beta \in (0,1)$, either

$$c \ge \frac{\ln(1/\beta)}{1-\beta} \quad or \quad a \ge \frac{1}{(1-\beta)}$$

how to choose β

$$L_{Hedge(\beta)} \leq \frac{\ln(1/\beta)}{1-\beta} L_{best} + \frac{1}{1-\beta} \ln(N)$$

- β = confidence parameter
- think of it as a trade-off
- try to make the bound tight
- binary search : perfect expert + (β=0)

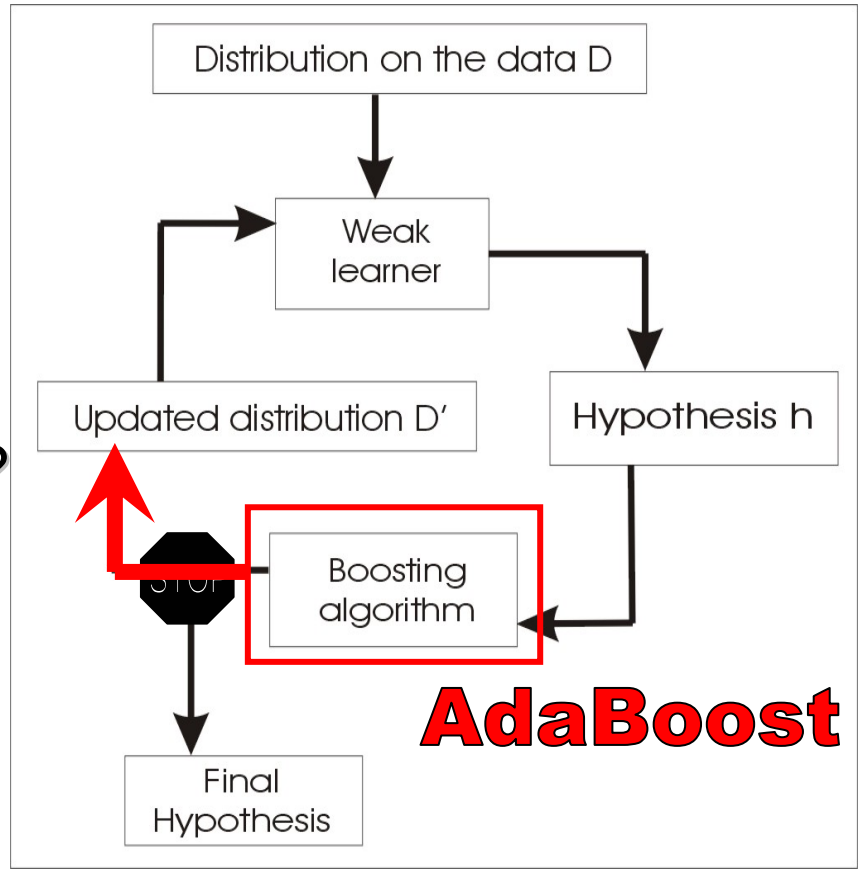
that wasn't so bad

☑Hedge algorithm for online allocation

- Applications
  - On classification : Adaboost
  - On IR : metasearch algorithm

boosting

- disease classification….
  - get past data
  - “train” a disease-predictor “doctor” (“hypothesis”,”weak learner”)
- how good is it (on training data) ?
  - where is it wrong ?
  - train a new predictor to correct mistakes for the first one



hedge application : AdaBoost

[schapire, freund '96]

HEDGE

- given : experts
- incoming : loses
- reweight: experts

BOOSTING

- given : datapoints
  - experts in weak-learners performance
- incoming: weak learners
  - compute error (loss)
  - compute "believe" (β)
- reweight : datapoints

ADABOOST

$$D_1(i) = \frac{1}{m}$$

$$h_t : X \rightarrow \{-1,1\}$$

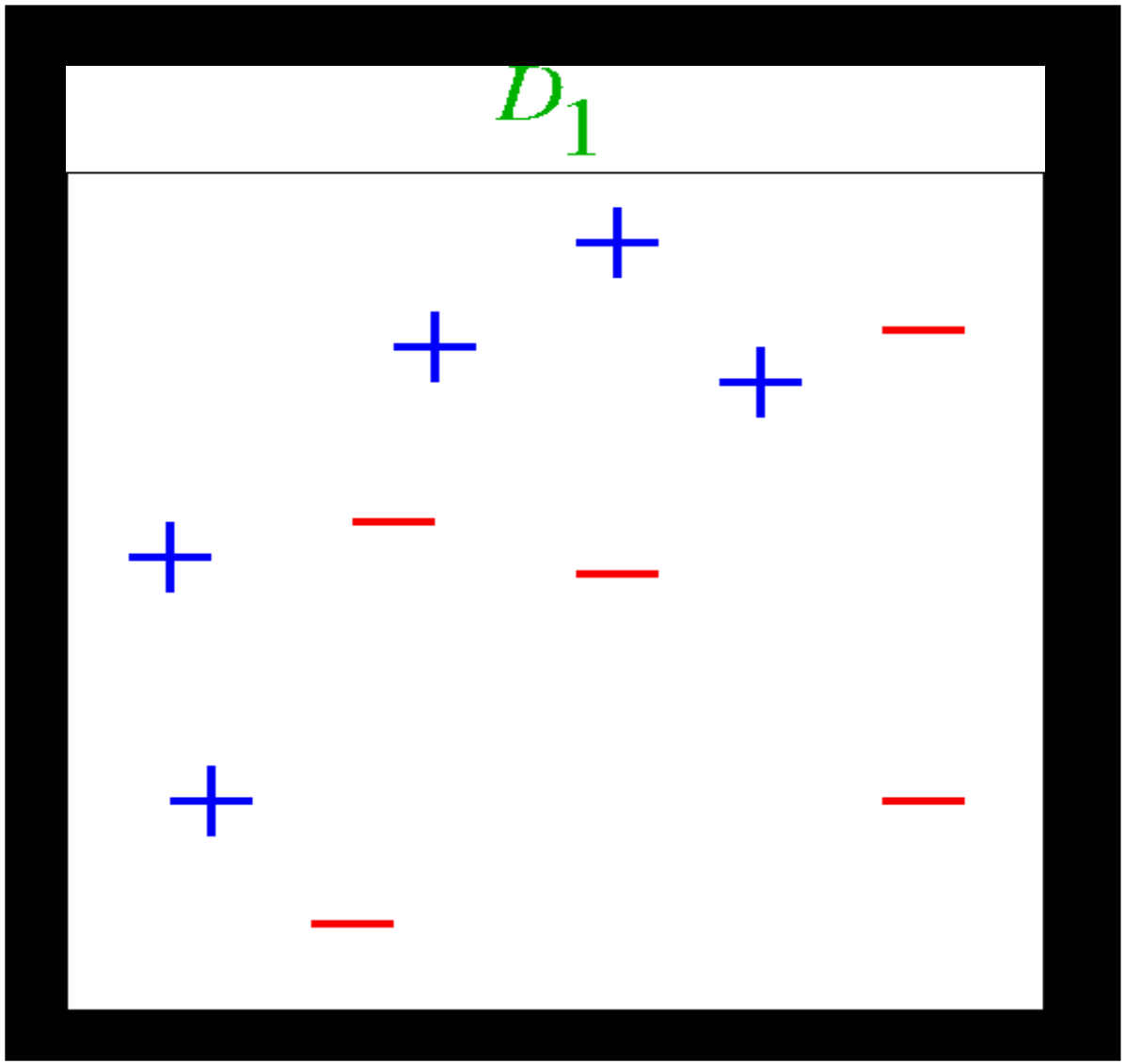
$$\varepsilon_t = \Pr_{D_t}[h_t(x_i) \neq y_i]$$

$$\beta_t = \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} > 1$$

$$D_{t+1} = \frac{D_t}{Z_t} \cdot \begin{cases} 1/\beta_t & \text{if } y_i = h_t(x_i) \\ \beta_t & \text{if } y_i \neq h_t(x_i) \end{cases}$$

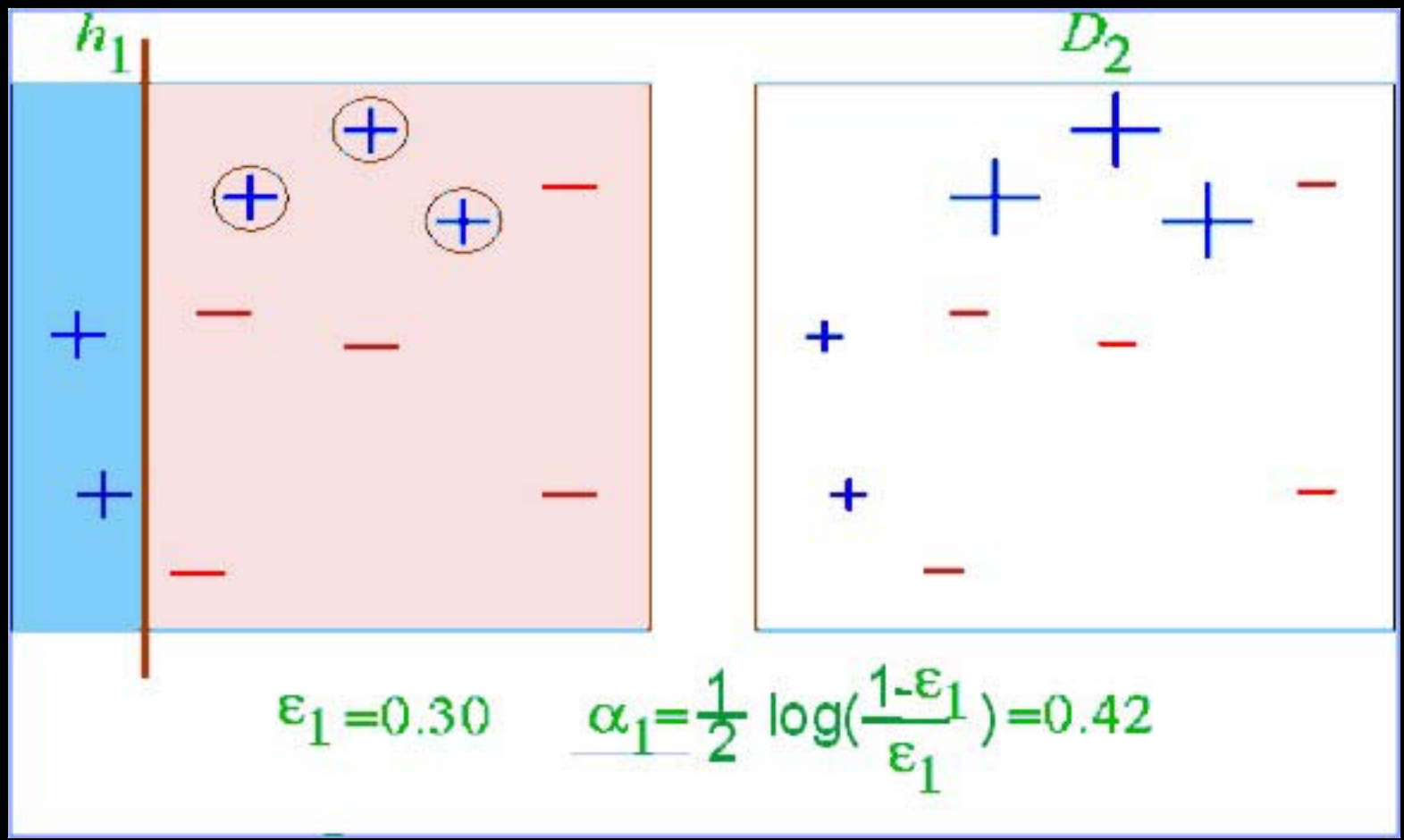
$$H_{\text{final}}(x) = \text{sgn}\left(\sum_t \ln(\beta_t) h_t(x)\right)$$

AdaBoost - example

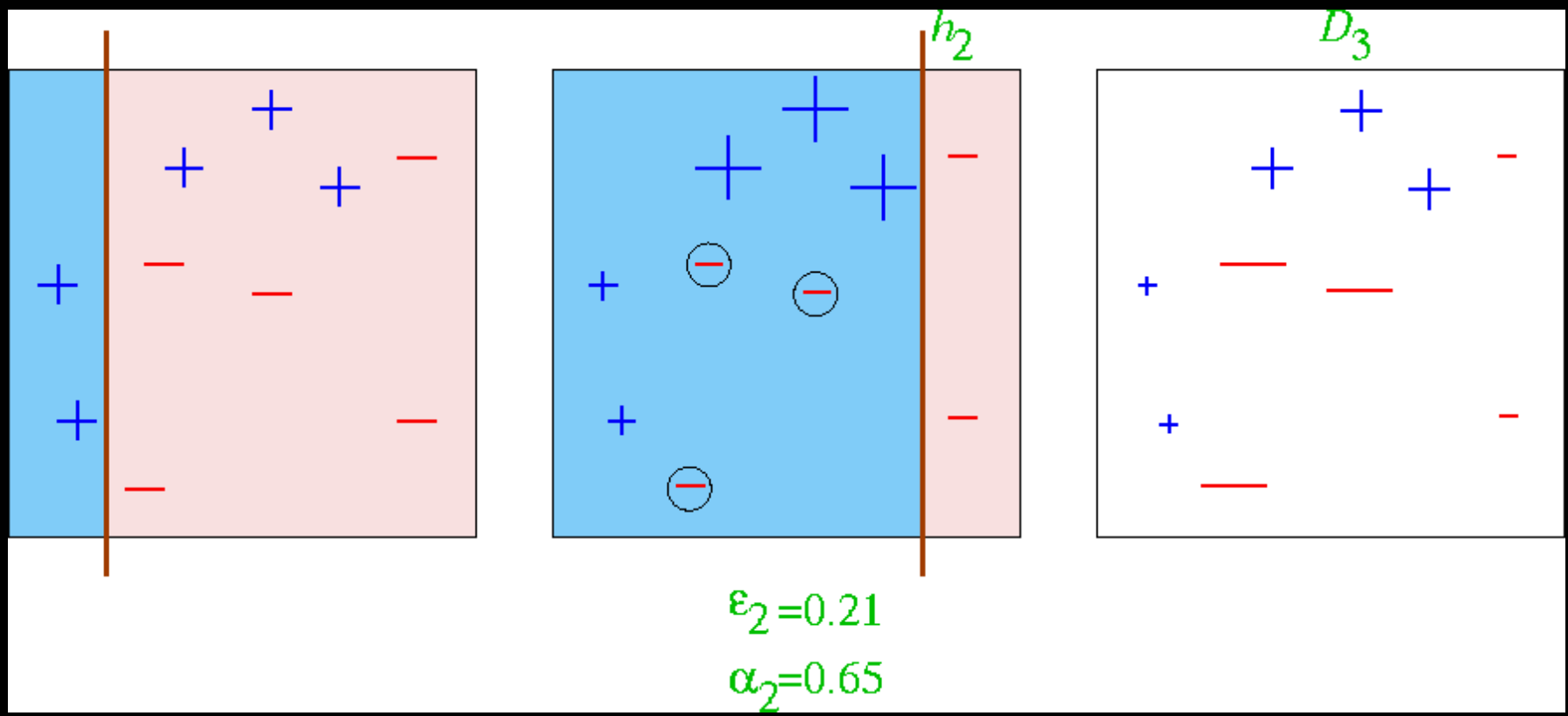


- Start with uniform distribution on data
- Weak learners = halfplanes

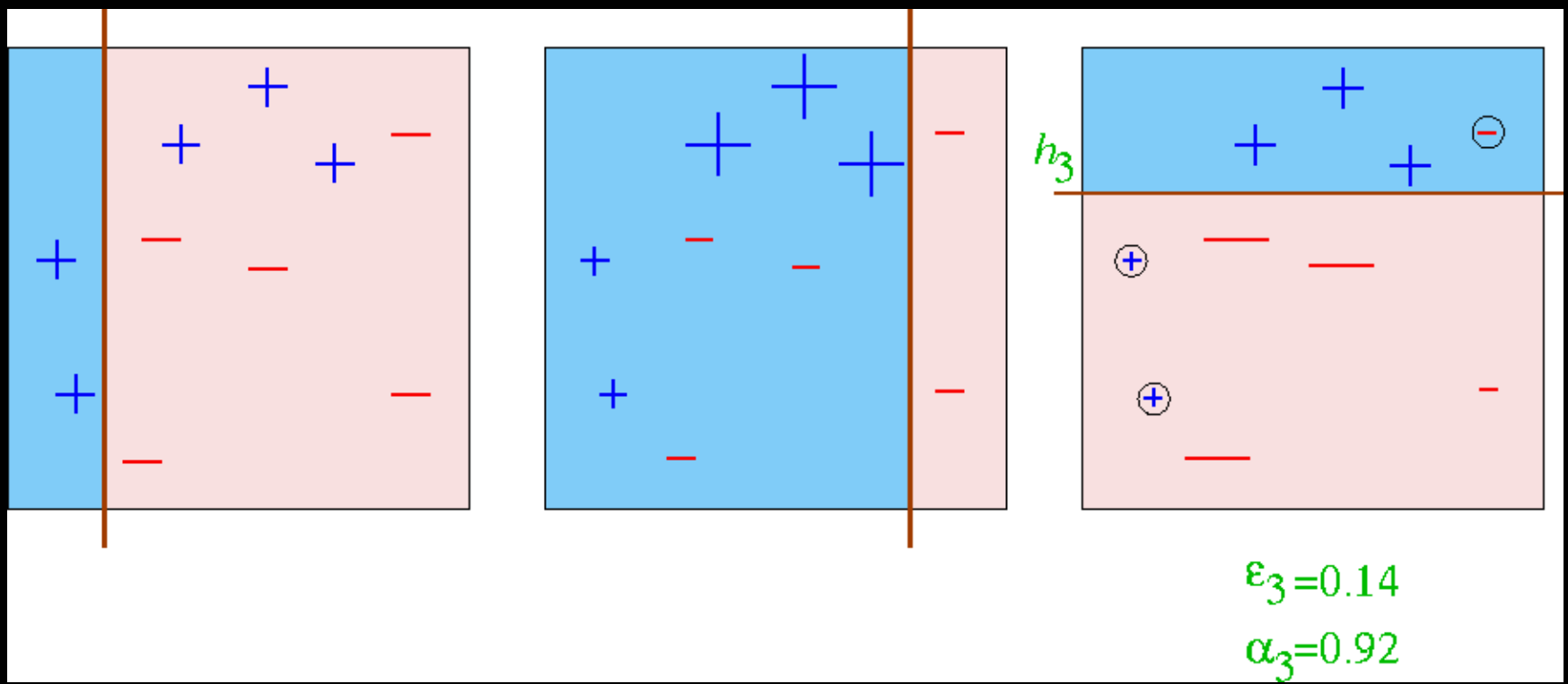
round 1


h_1
D_2
$\varepsilon_1 = 0.30$
$\alpha_1 = \frac{1}{2} \log(\frac{1-\varepsilon_1}{\varepsilon_1}) = 0.42$

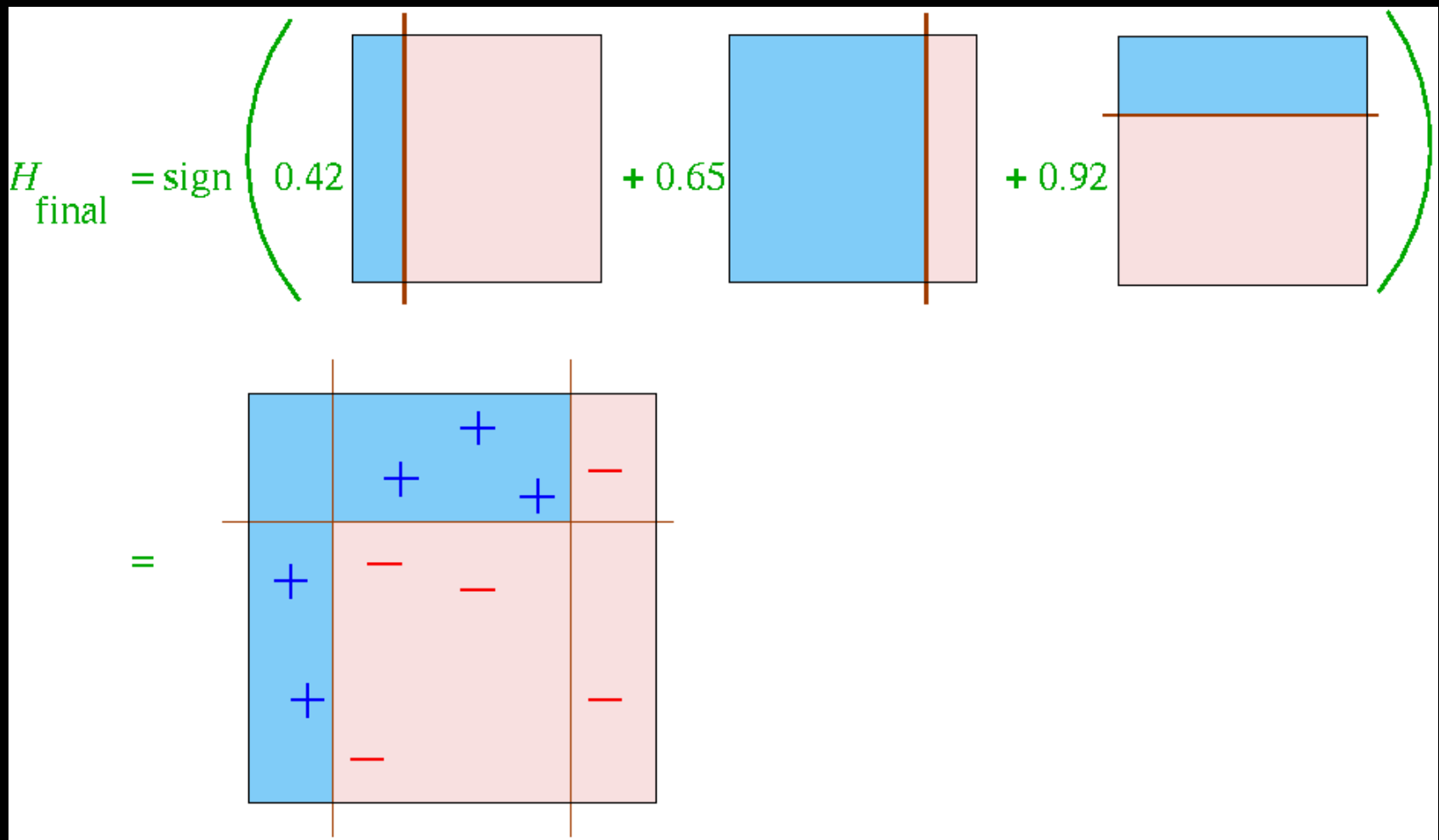
round 2



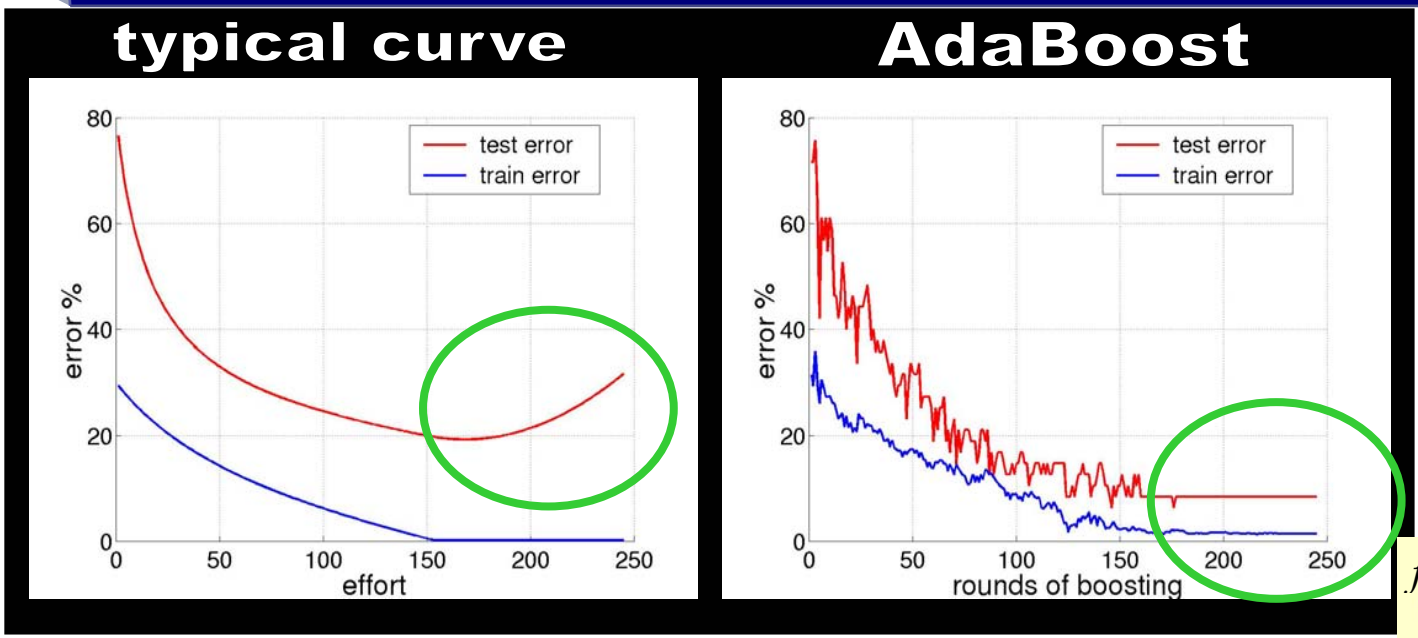
round 3



final hypothesis

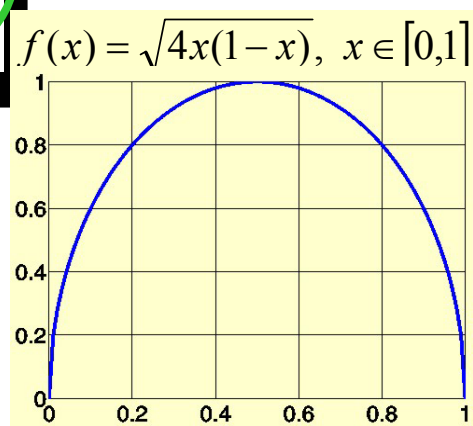


AdaBoost - analysis



- training error

$$\text{training error}(H_{\text{final}}) \le \prod_t \sqrt{4\varepsilon_t(1-\varepsilon_t)}$$



- generalization error

$$\mathbf{P}_{\mathcal{D}}\left[yf(x) \le 0\right] \le \mathbf{P}_{S}\left[yf(x) \le \theta\right] + O\left(\frac{1}{\sqrt{m}}\left(\frac{\log m \log|\mathcal{H}|}{\theta^2} + \log(1/\delta)\right)^{1/2}\right)$$

2003 Gödel Prize
Yoav Freund and Robert Schapire

- "The prize was awarded to Yoav Freund and Robert Schapire for their paper "A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences 55 (1997), pp. 119-139. This paper introduced AdaBoost, an adaptive algorithm to improve the accuracy of hypotheses in machine learning. The algorithm demonstrated novel possibilities in analysing data and is a permanent contribution to science even beyond computer science. Because of a combination of features, including its elegance, the simplicity of its implementation, its wide applicability, and its striking success in reducing errors in benchmark applications even while its theoretical assumptions are not known to hold, ***the algorithm set off an explosion of research in the fields of statistics, artificial intelligence, experimental machine learning, and data mining***. The algorithm is now widely used in practice. The paper highlights the fact that theoretical computer science continues to be a fount of powerful and entirely novel ideas with significant and direct impact even in areas, such as data analysis, that have been studied extensively by other communities."

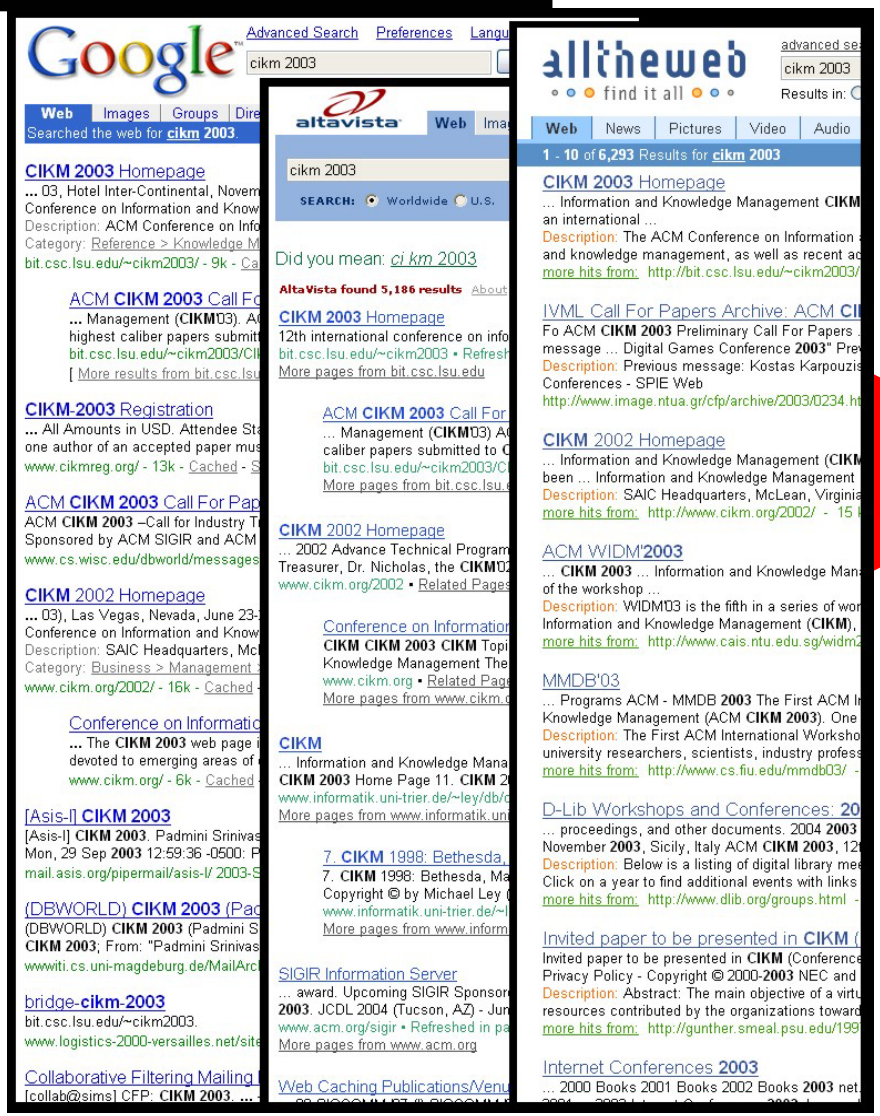
what we did last year [Aslam, Pavlu, Savell]

☑Hedge algorithm for online allocation

- Applications
  - ☑On classification : Adaboost
  - On IR : metasearch algorithm

metasearch problem

Search for: CIKM 2003



metasearch

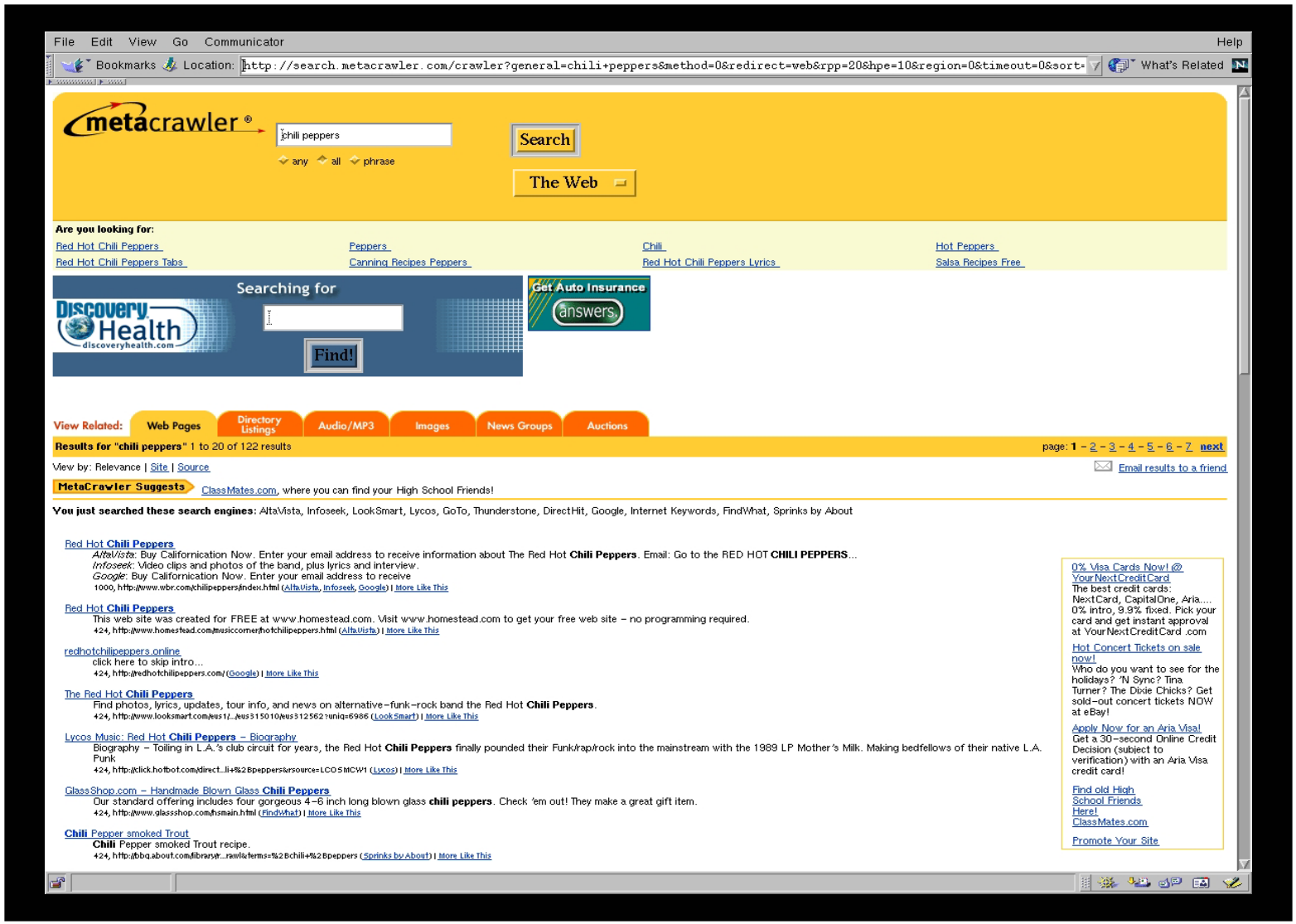
CIKM 2003 Homepage
ACM CIKM 2003 Call For Pape
CIKM–2003 Registration
ACM CIKM 2003 Call For Pape
CIKM 2002 Homepage
Conference on Information an
[Asis–l] CIKM 2003
(DBWORLD) CIKM 2003 (Padm
bridge–cikm–2003
Collaborative Filtering Mailing
CIKM Home Page ACM DL: CIK
Yahoo! Groups : webir Messag
dBforums – Cfp: Cikm 03
Mailing List ARL–ERESERVE@a
Received: from cni.org by b.cr
ACM WIDM 2003
ACM – MMDB 2003
ACM CIKM 2003 PRELIMINARY
Selected Publications
Web Caching Publications/Ver
cikm '03
Call for papers
Mario A. Nascimento – Person
Iceved
Conferences and Journals on T
Conferences On Information V
Calendrier des manifestations
(DBWORLD) Final Call for ACM
Collaborative Filtering Mailing

search and metasearch

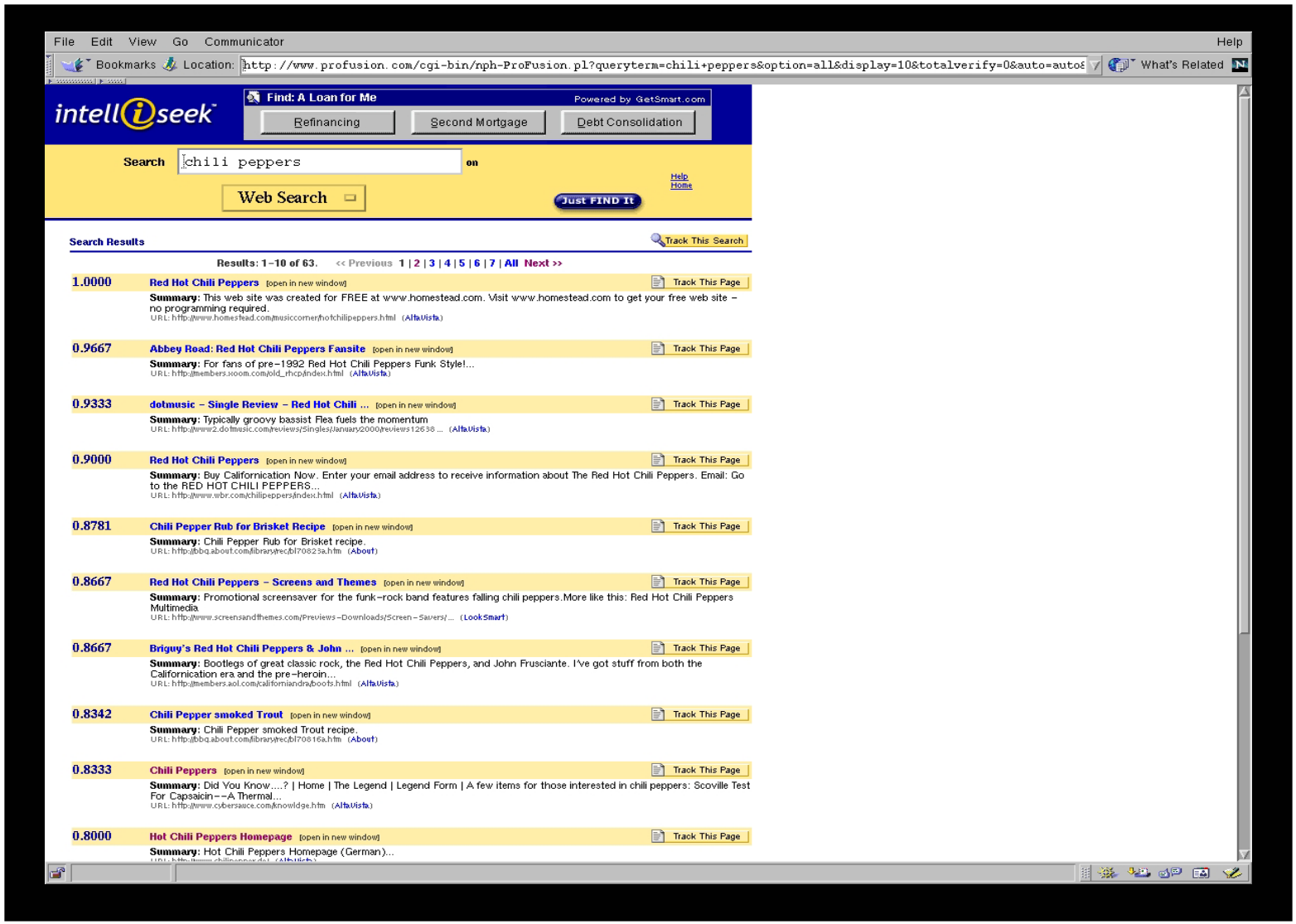
- Search engines:
  - Provide a ranked list of documents.
  - May provide relevance scores.
- Metasearch engines:
  - Query multiple search engines.
  - May or may not combine results.

metasearch: Dogpile

metasearch: Metacrawler



metasearch: Profusion

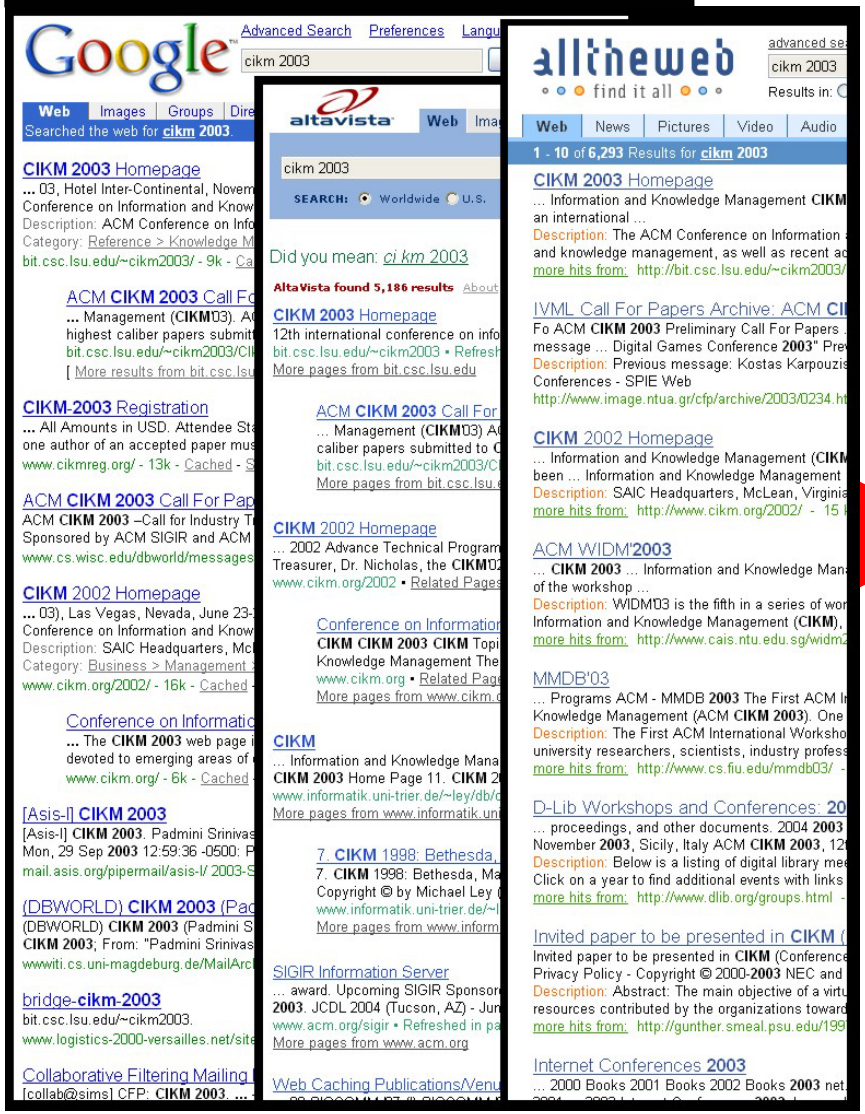


metasearch algorithms

- Heuristics and hacks:
  - Interleave, average rank, sum scores, etc.
- Principled models:
  - Bayesian inference, election theory, etc.
  - *On-line combination of expert advice.*

online metasearch

Search for: CIKM 2003



metasearch

CIKM 2003 Homepage
ACM CIKM 2003 Call For Pape
CIKM-2003 Registration
ACM CIKM 2003 Call For Pape
CIKM 2002 Homepage
Conference on Information an
[Asis-l] CIKM 2003
(DBWORLD) CIKM 2003 (Padm
bridge-cikm-2003
Collaborative Filtering Mailing
CIKM Home Page ACM DL: CIK
Yahoo! Groups : webir Messag
dBforums - Cfp: Cikm 03
Mailing List ARL-ERESERVE@a
Received: from cni.org by b.cr
ACM WIDM 2003
ACM - MMDB 2003
ACM CIKM 2003 PRELIMINARY
Selected Publications
Web Caching Publications/Ver
cikm '03
Call for papers
Mario A. Nascimento - Person
Iceved
Conferences and Journals on T
Conferences On Information V
Calendrier des manifestations
(DBWORLD) Final Call for ACM
Collaborative Filtering Mailing

metasearch

CIKM 2003 Homepage
CIKM-2003 Registration
CIKM 2002 Homepage
ACM CIKM 2003 Call For Pape
ACM CIKM 2003 Call For Pape
CIKM Home Page ACM DL: CIK
Conference on Information an
ACM CIKM 2003 PRELIMINARY
CIKM Home Page ACM DL: CIK
ACM CIKM 2003 PRELIMINARY
[Asis-l] CIKM 2003
(DBWORLD) CIKM 2003 (Padm
bridge-cikm-2003
Collaborative Filtering Mailing
Yahoo! Groups : webir Messag
dBforums - Cfp: Cikm 03
Mailing List ARL-ERESERVE@a
Received: from cni.org by b.cr
ACM WIDM 2003
ACM - MMDB 2003
Selected Publications
Web Caching Publications/Ver
cikm '03
Call for papers
Mario A. Nascimento - Person
Iceved
Conferences and Journals on T
Conferences On Information V
Calendrier des manifestations

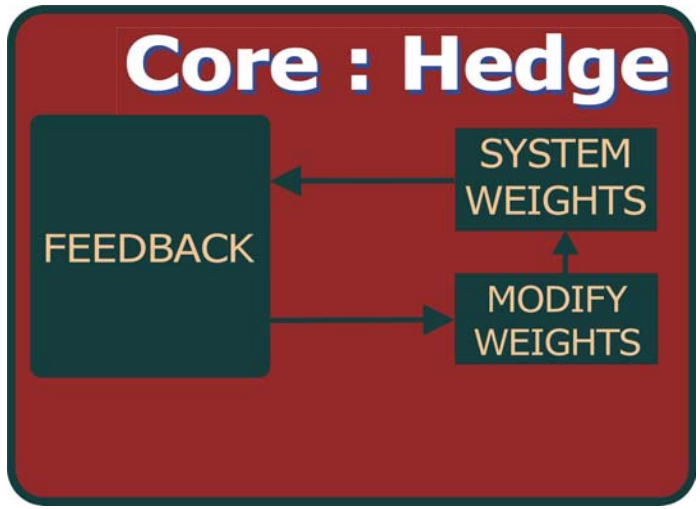
hedge application : metasearch

HEDGE

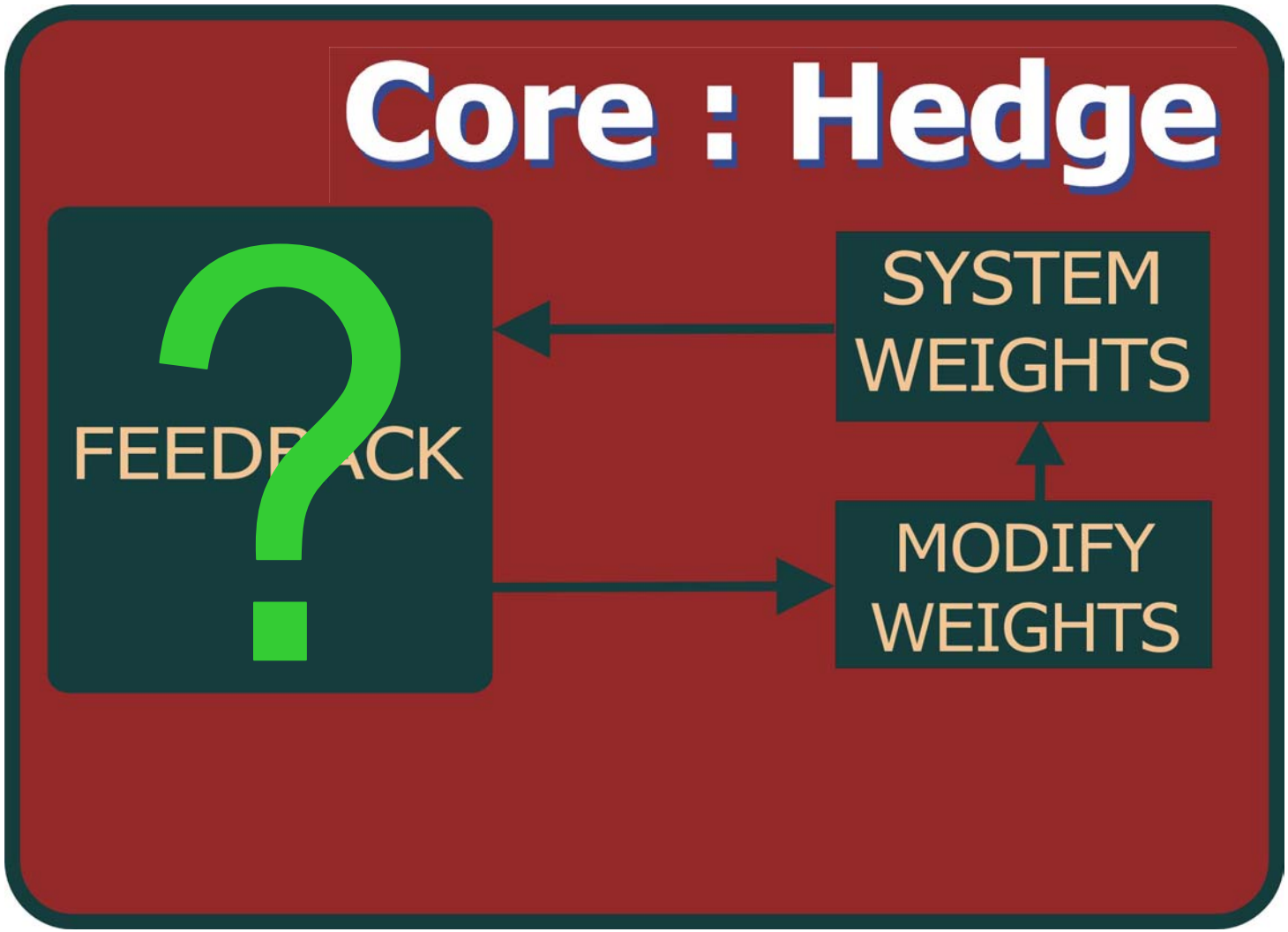
- given : experts
- incoming : loses
- reweight: experts

METASEARCH

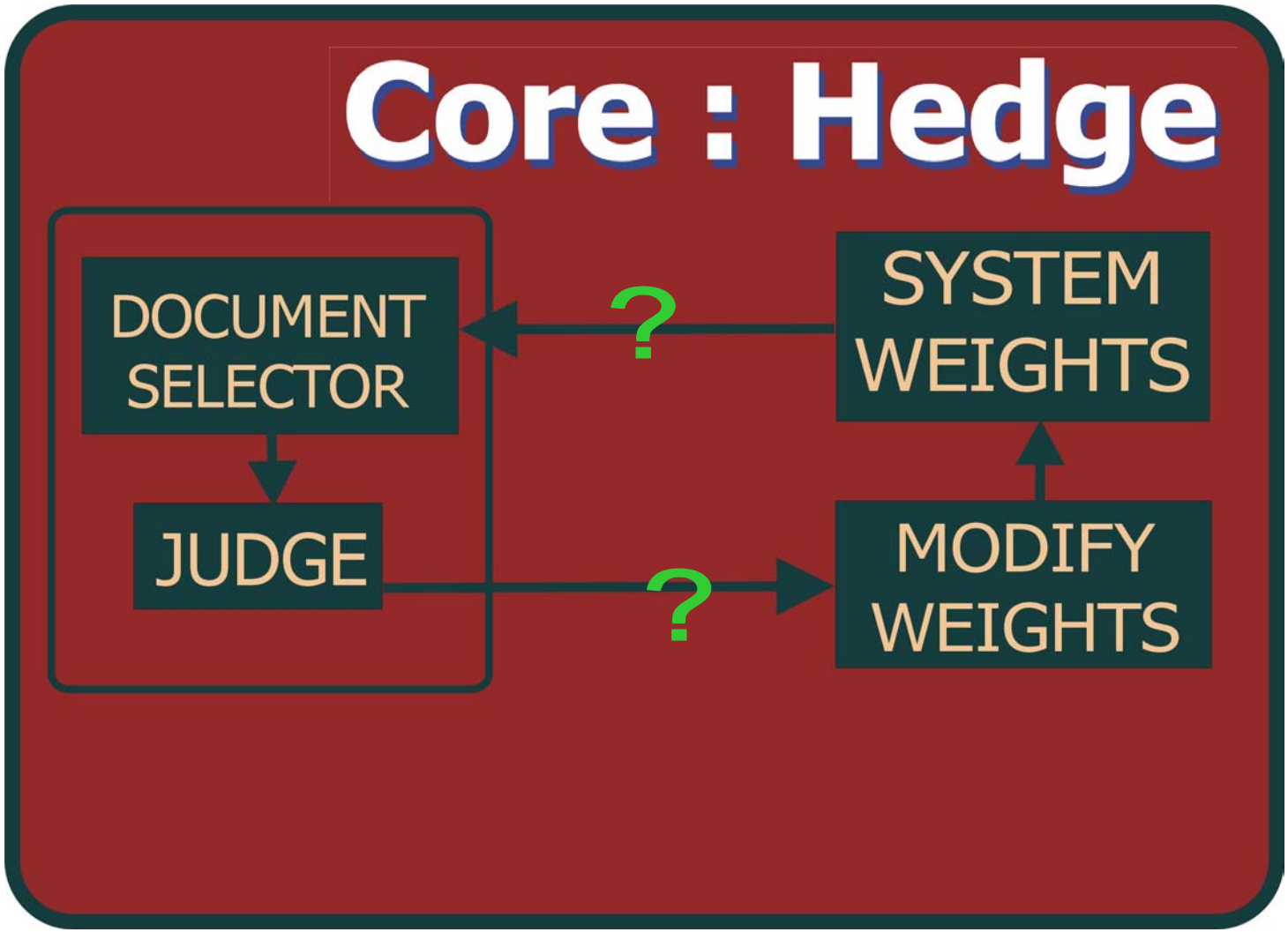
- given : search engines
- incoming: documents
  - compute losses
- reweight : search engines



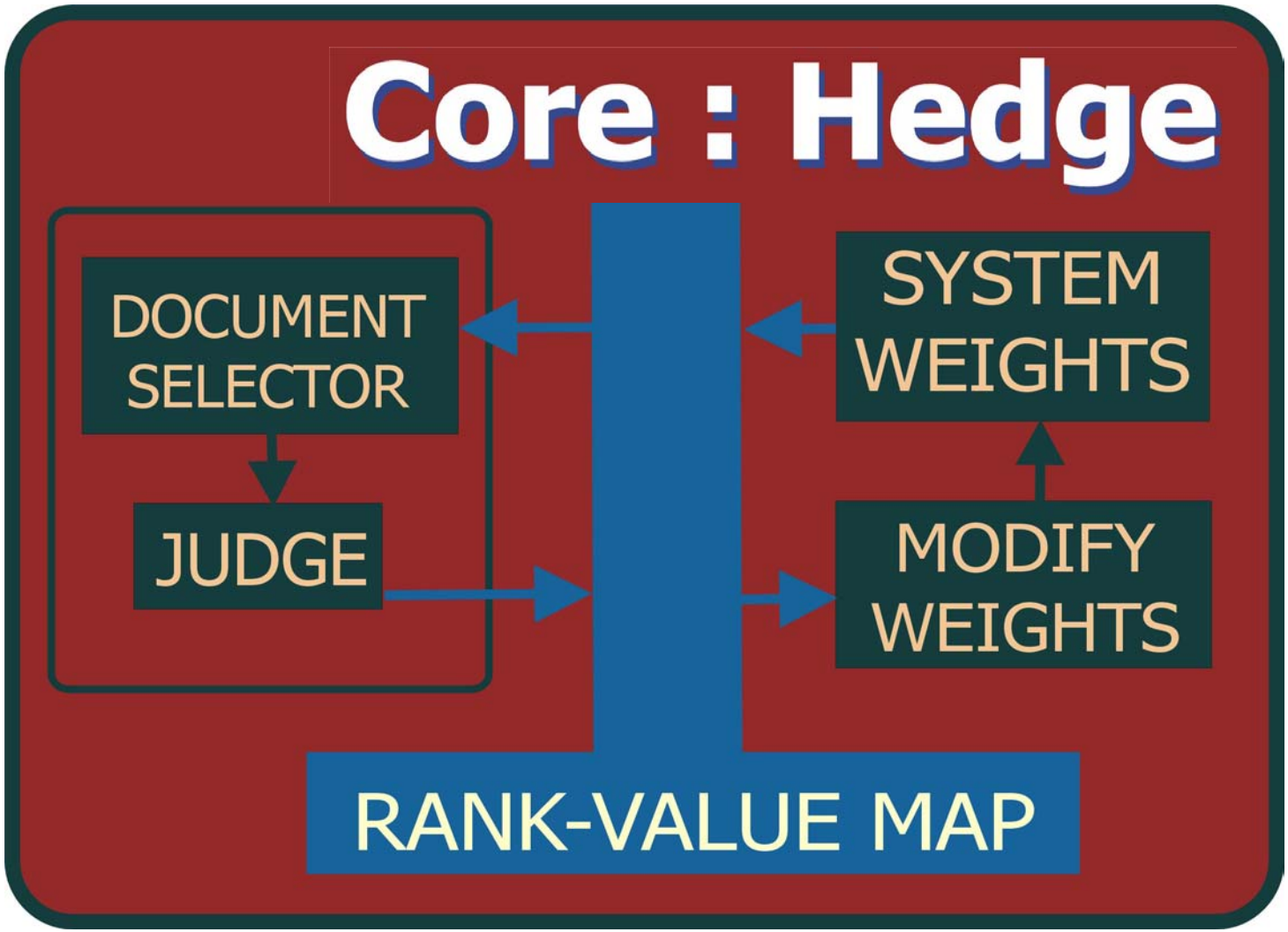
a unified model



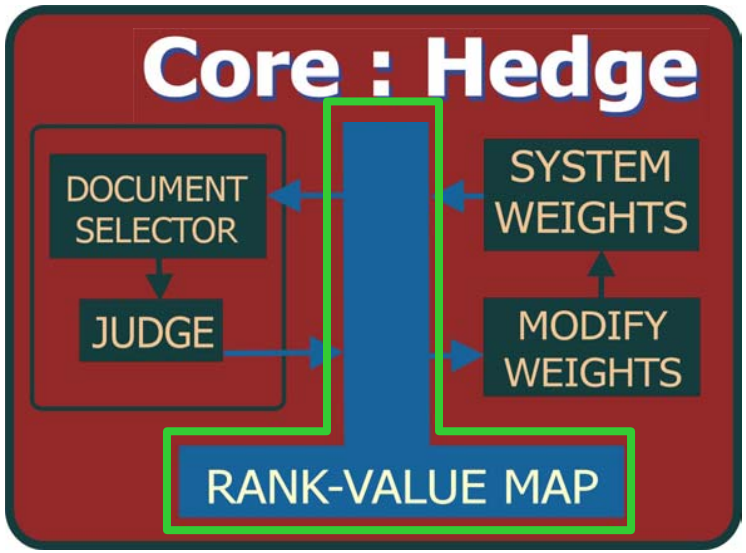
a unified model



a unified model



ranks importance



- map ranks to values

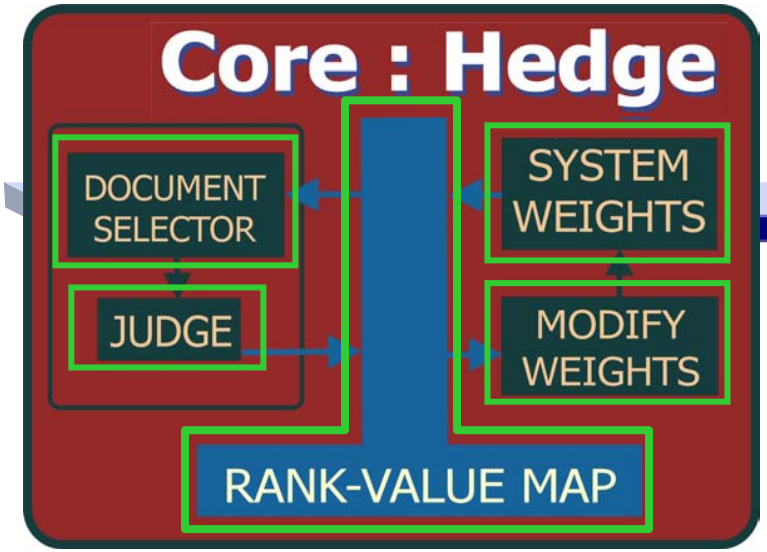
$$value(r) = \frac{1}{r} + \frac{1}{r+1} + \ldots + \frac{1}{Z} \approx \ln \frac{Z}{r}$$

$$RELEVANT = -1$$
$$NONRELEVANT = +1$$

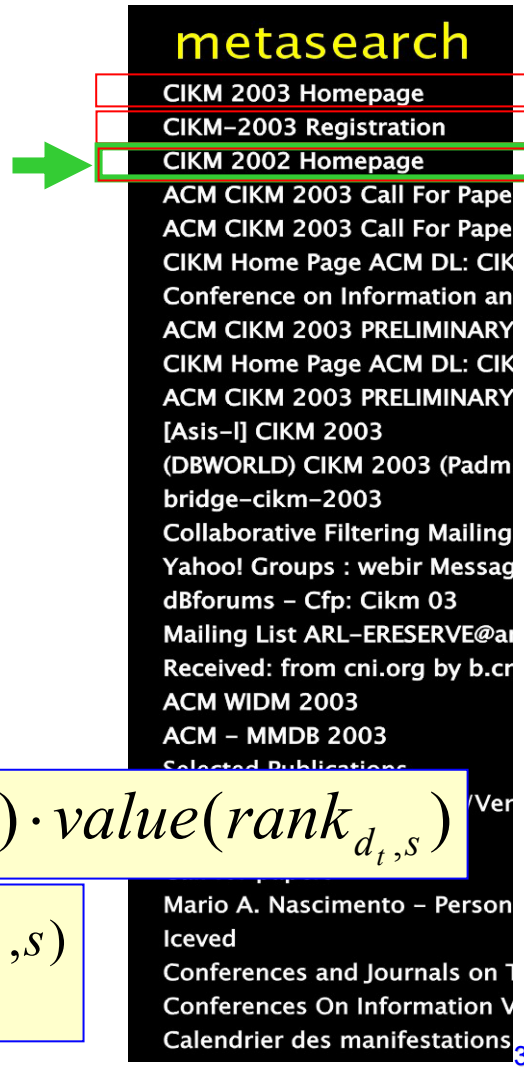
$$LOSS(d,s) = label(d) \cdot value(rank_{d,s}) \approx label(d) \cdot \ln \frac{Z}{r}$$

total loss vs. total precision vs. average precision

feedback loop



$$average\_value_t(d) = \sum_{s=1}^{N} w_s^{t-1} \cdot value(rank_{d,s})$$



- average_value at episode t
  - "trust" in systems change with t
- metasearch list : order docs by

$$average\_value_t(d)$$

- get feedback

$$label(d_t)$$

- compute losses

$$LOSS\ (d_t, s) = label(d_t) \cdot value(rank_{d_t,s})$$

- modify weights

$$w_s^{t+1} = w_s^t \cdot \beta^{LOSS_t\ (d_t,s)}$$

experiments

- TREC 3,5,6,7,8
  - 41-129 systems
  - 50 queries per TREC
  - Metasearch combines *all* systems
- Use TREC judgments as user feedback

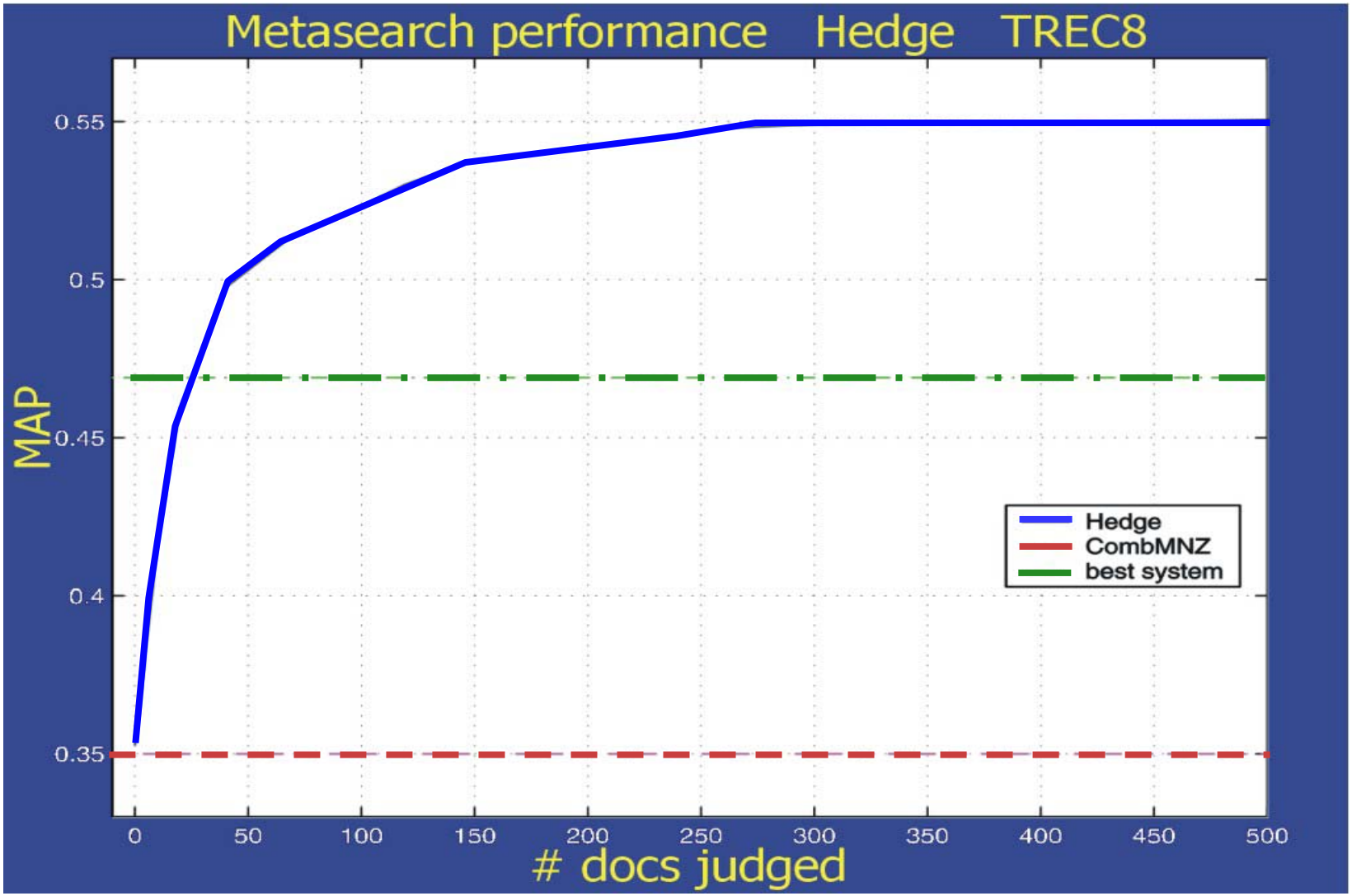
metasearch - no feedback (yet)

no relevant judgements → uniform weights

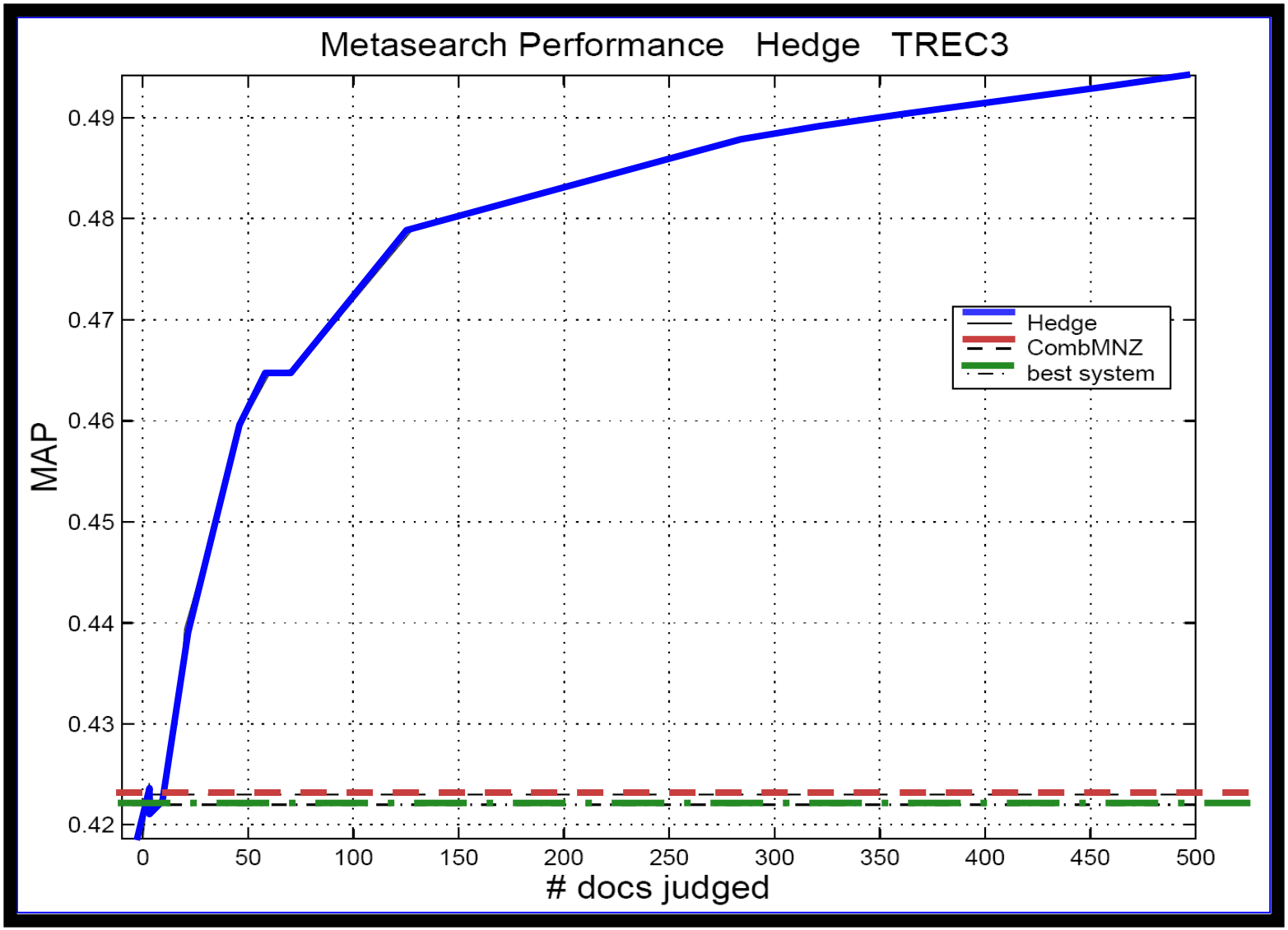
TREC	MNZ	COND	Hedge-0	%MNZ	%COND
3	0.423	0.403	0.418	−1.2	+3.7
5	0.294	0.307	0.309	+5.1	+0.6
6	0.341	0.315	0.345	+1.2	+9.5
7	0.320	0.308	0.323	+0.9	+4.9
8	0.350	0.343	0.352	+1.4	+2.6

MNZ=CombMNZ(Fox,Shaw,Lee et al)
COND=Condorcet(Aslam,Montague)

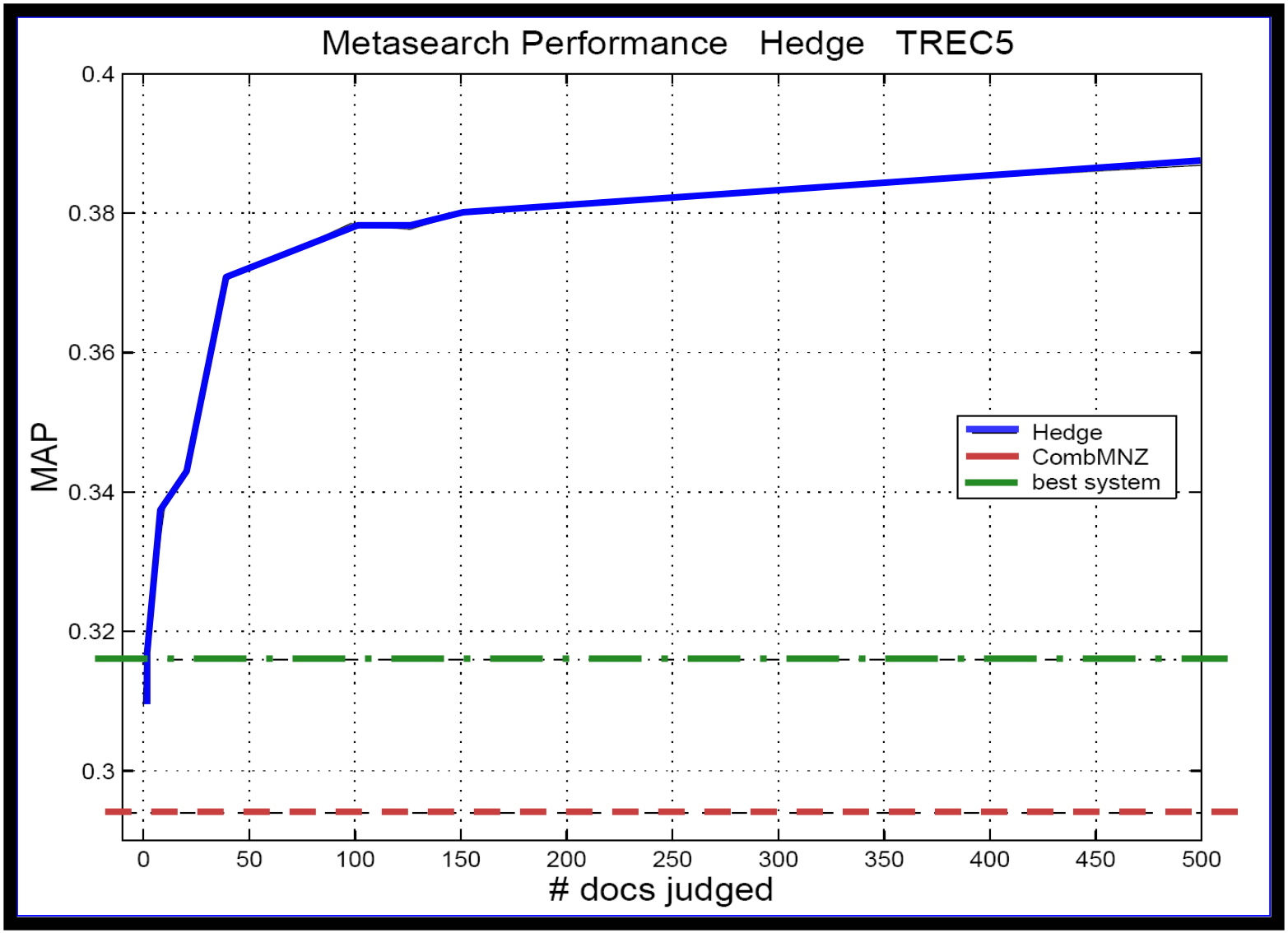
metasearch – TREC8



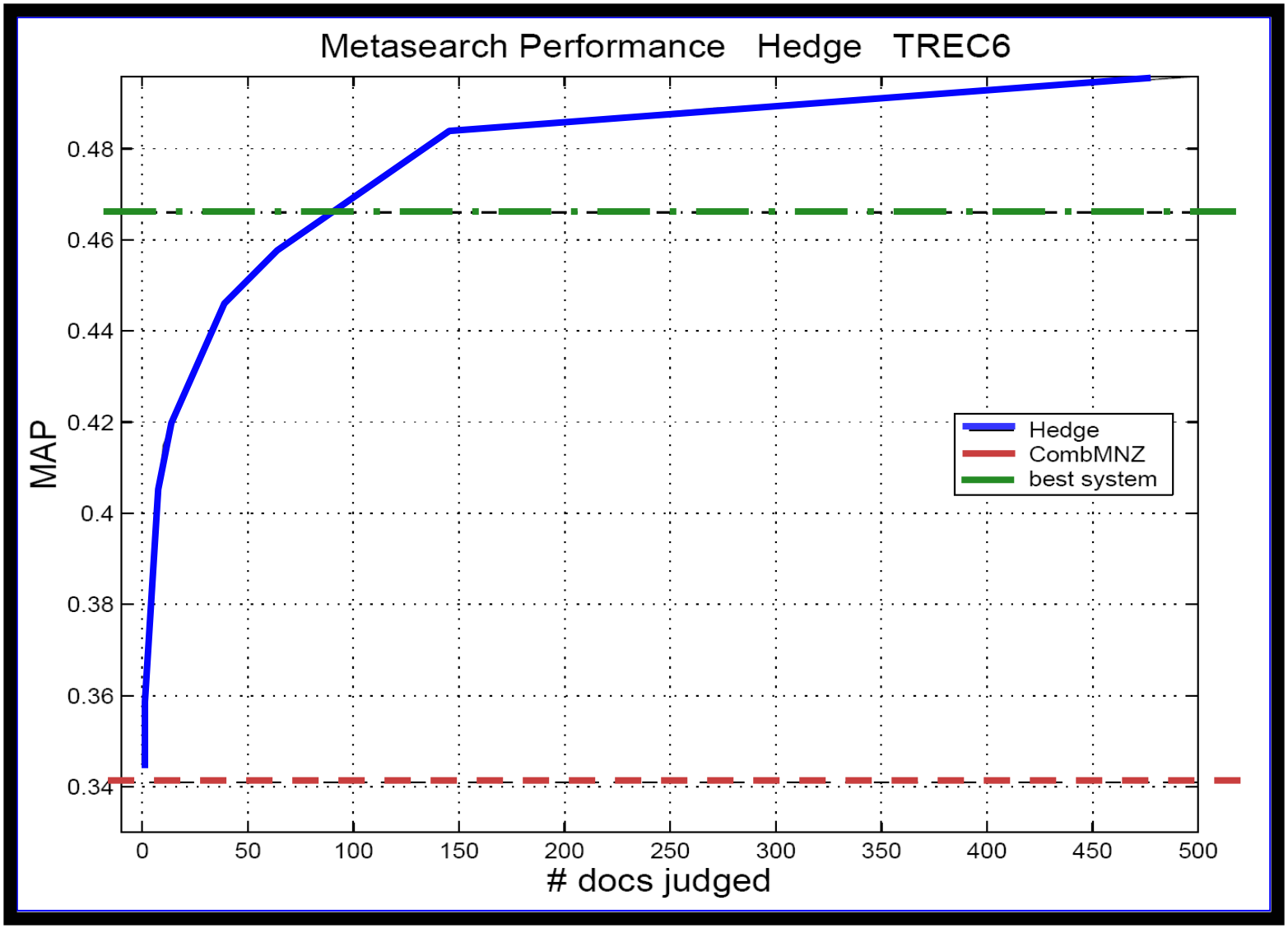
metasearch – TREC 3



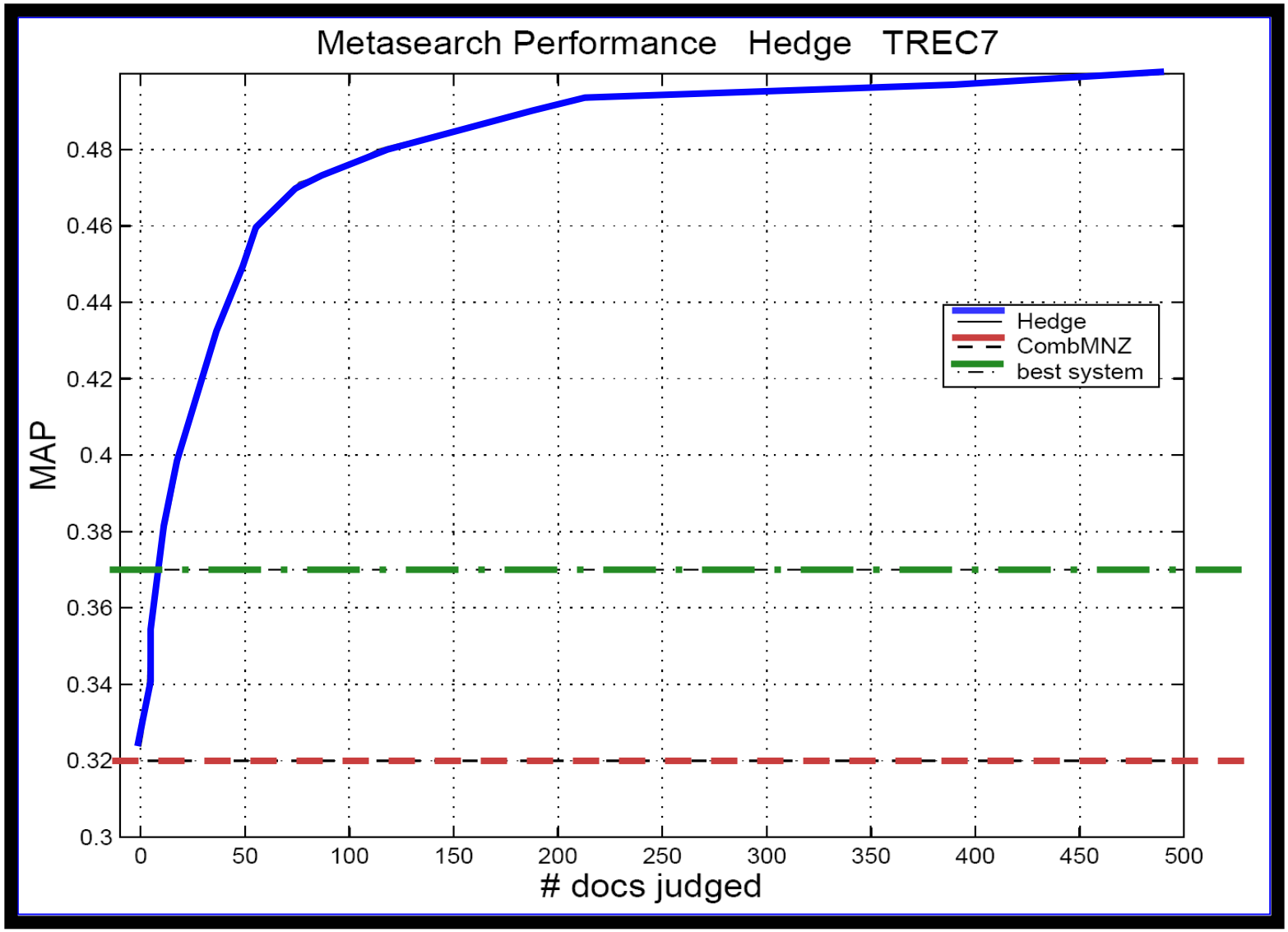
metasearch – TREC 5



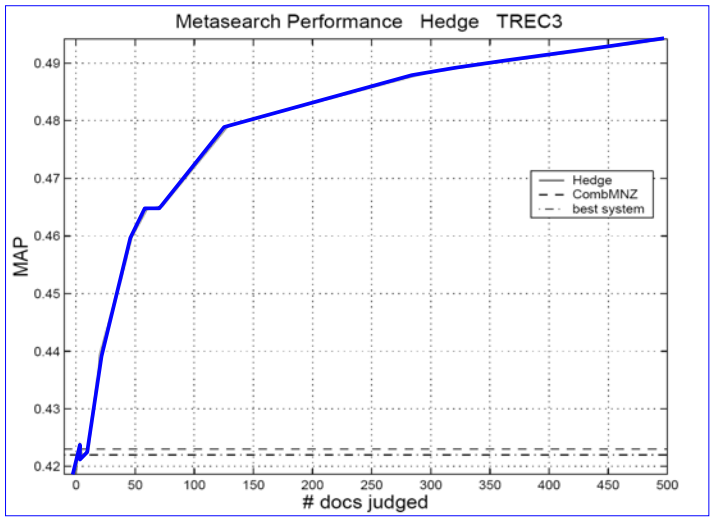
metasearch – TREC 6

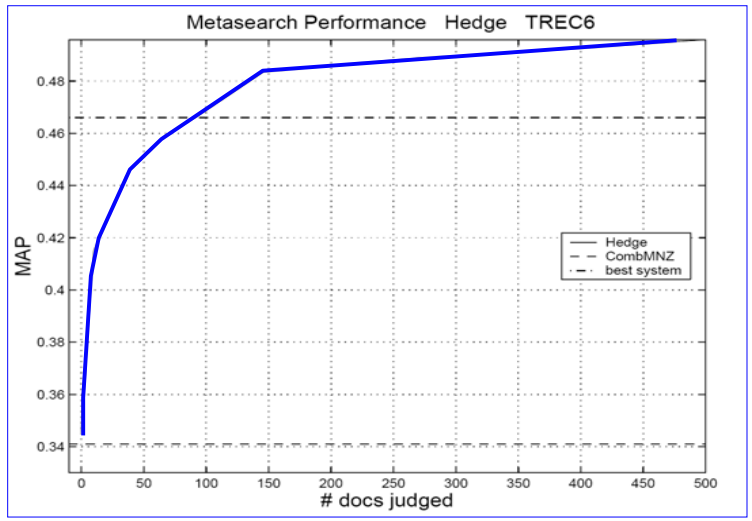


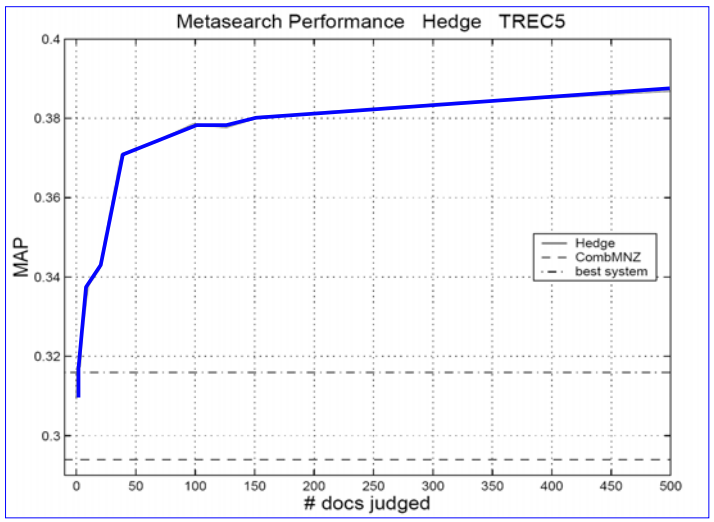
metasearch – TREC 7

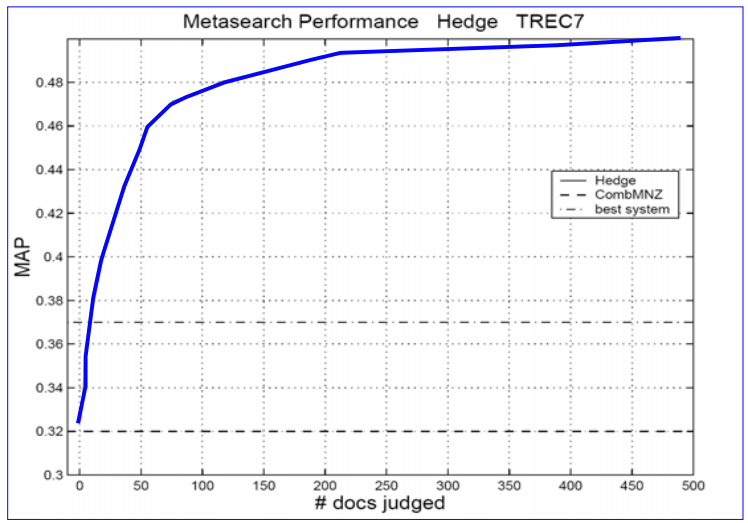


metasearch – TREC 3,5,6,7

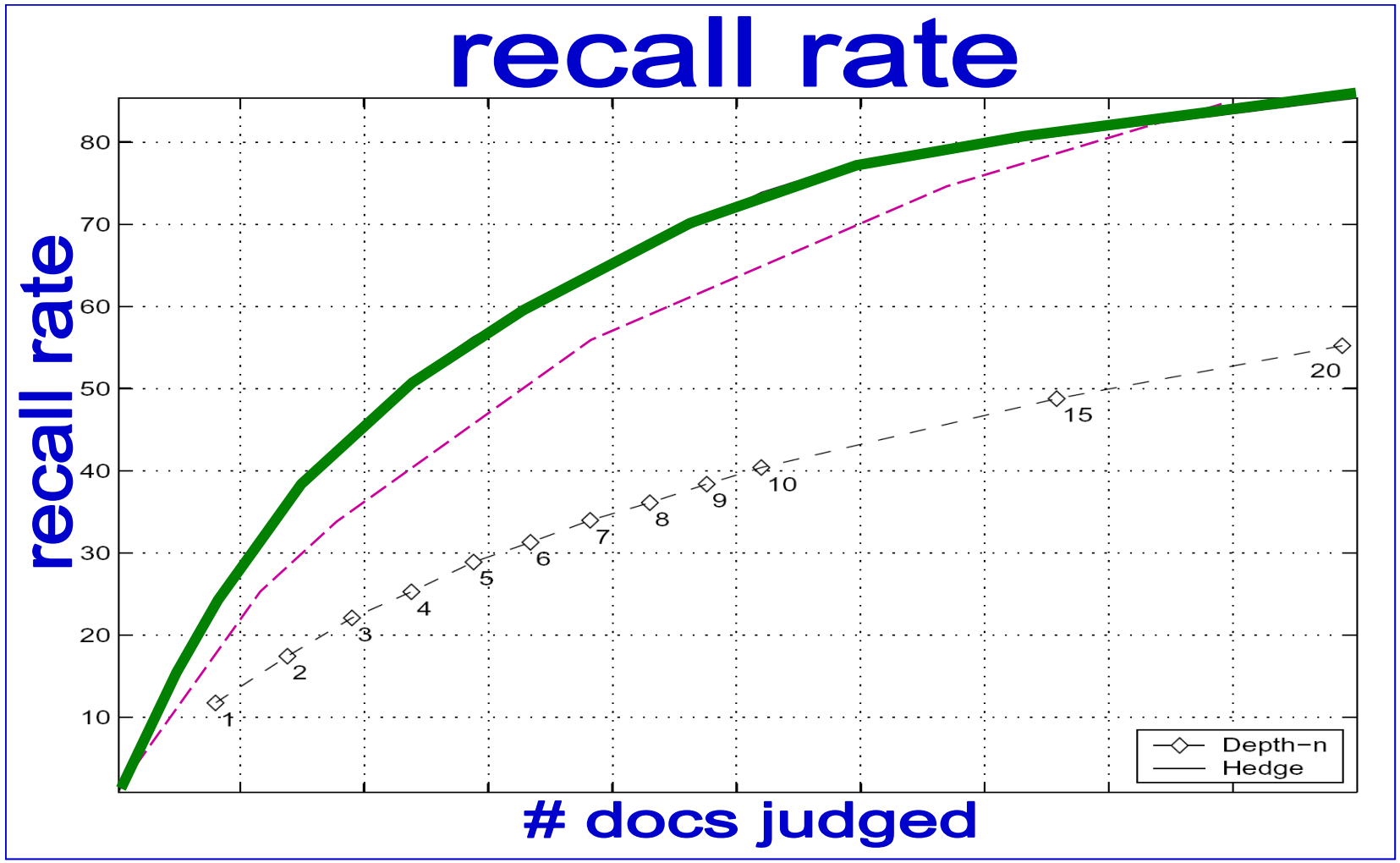




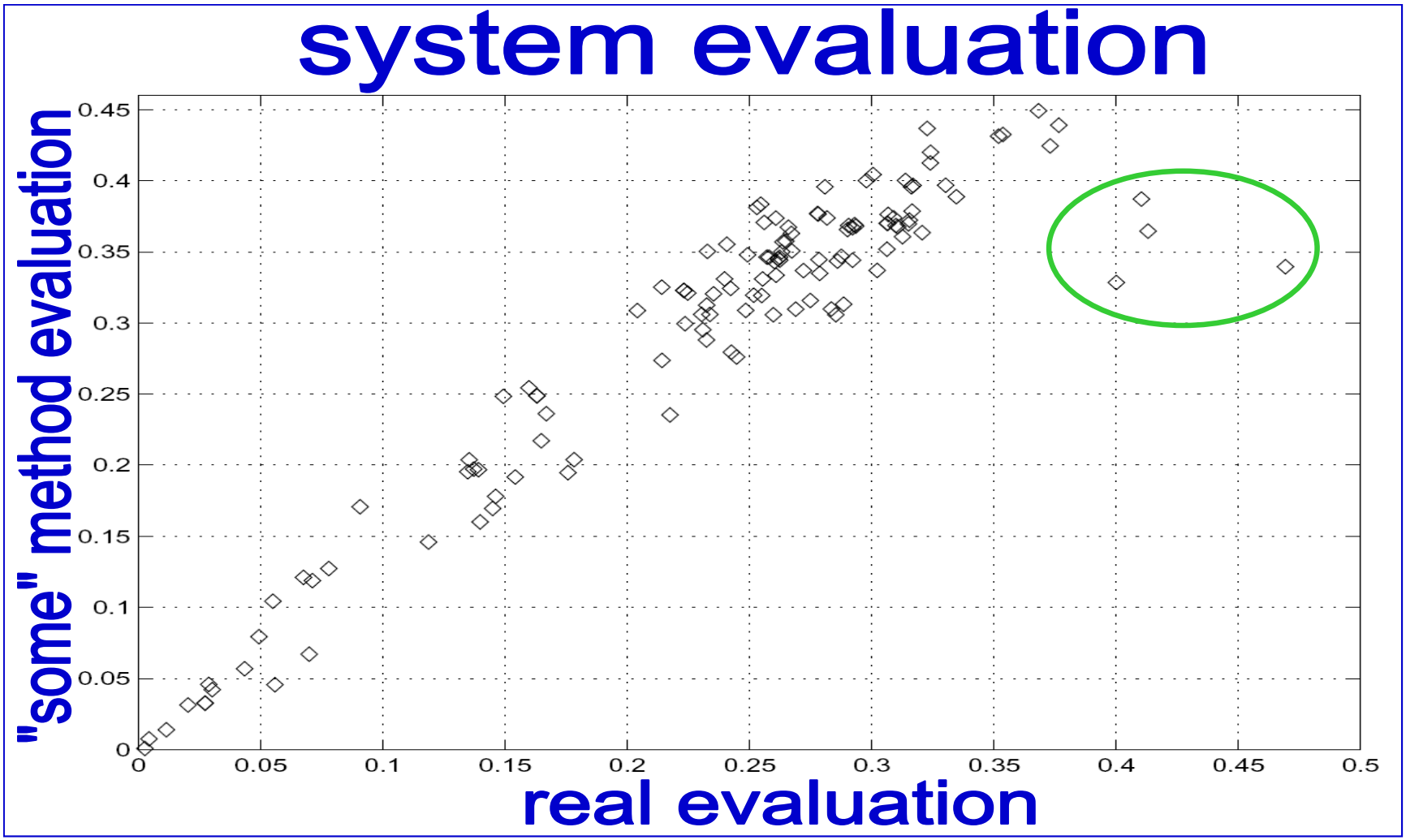




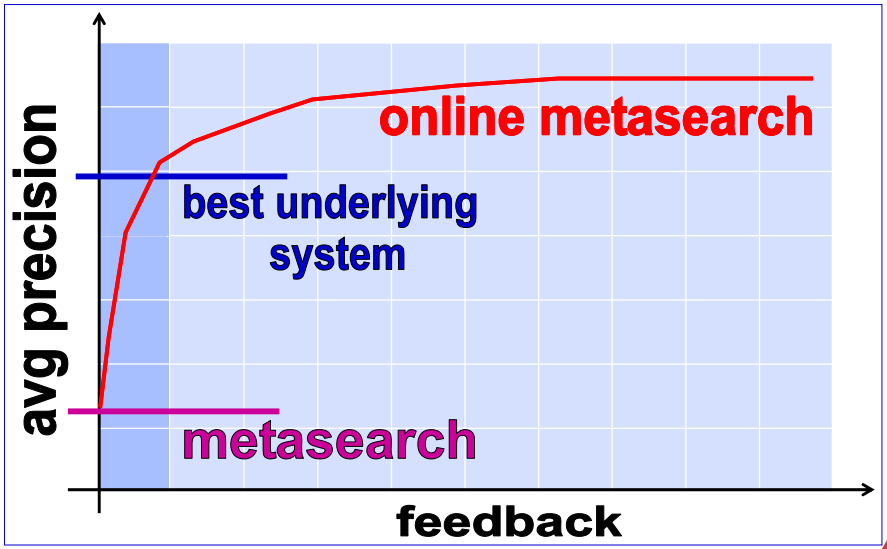
actually...we do more than metasearch

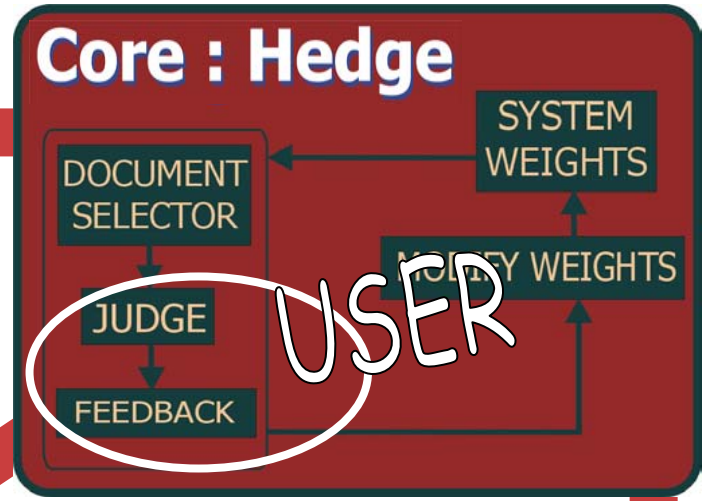


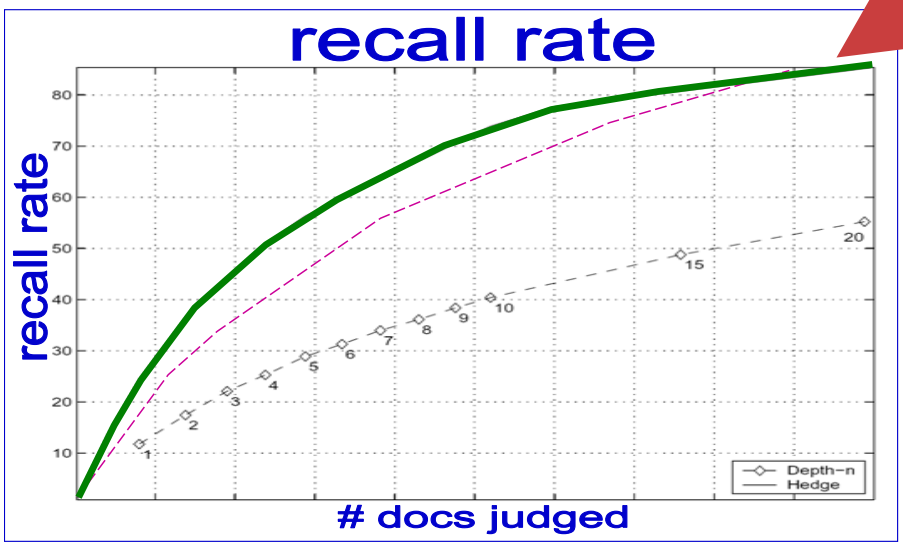
actually...we do more than metasearch

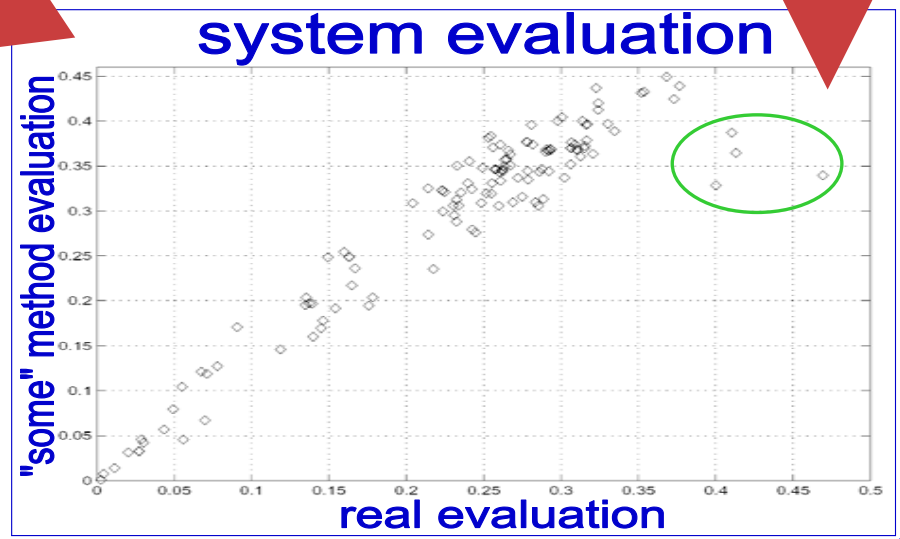


"a unified model for metasearch, pooling and system evaluation"









conclusion

- theoretic explanation of “being adaptive”
- simple, elegant, intuitive
- usually performs much better than the bound

END

AdaBoost - technical

$$D_1(i) = \frac{1}{m}$$

- Start with uniform distrib D_1:
- At every round t=1 to T

  given D_t

  find weak hypothesis $h_t : X \rightarrow \{-1,1\}$

  with error $\varepsilon_t = \Pr_{D_t}[h_t(x_i) \neq y_i]$

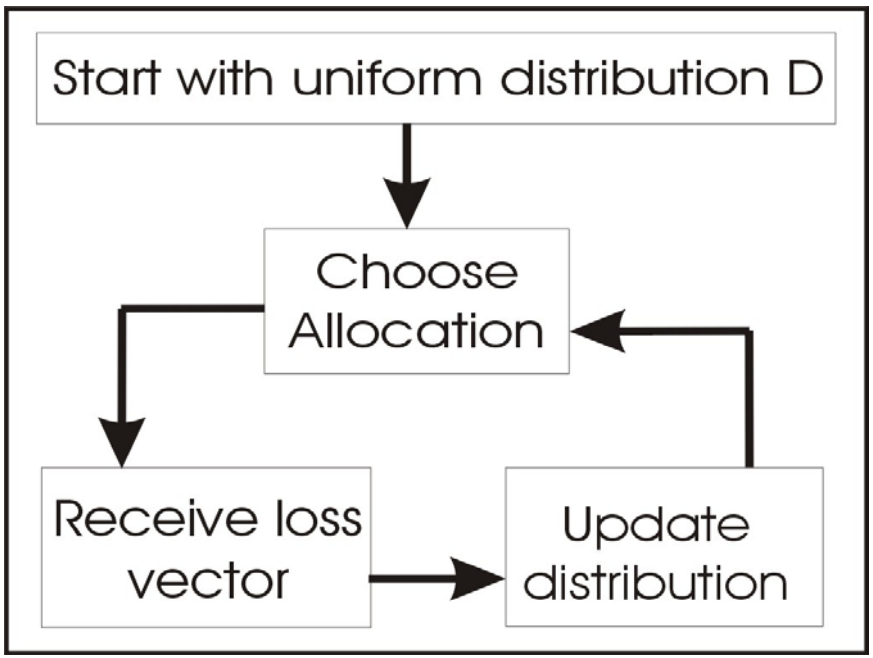
  $$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) > 0$$

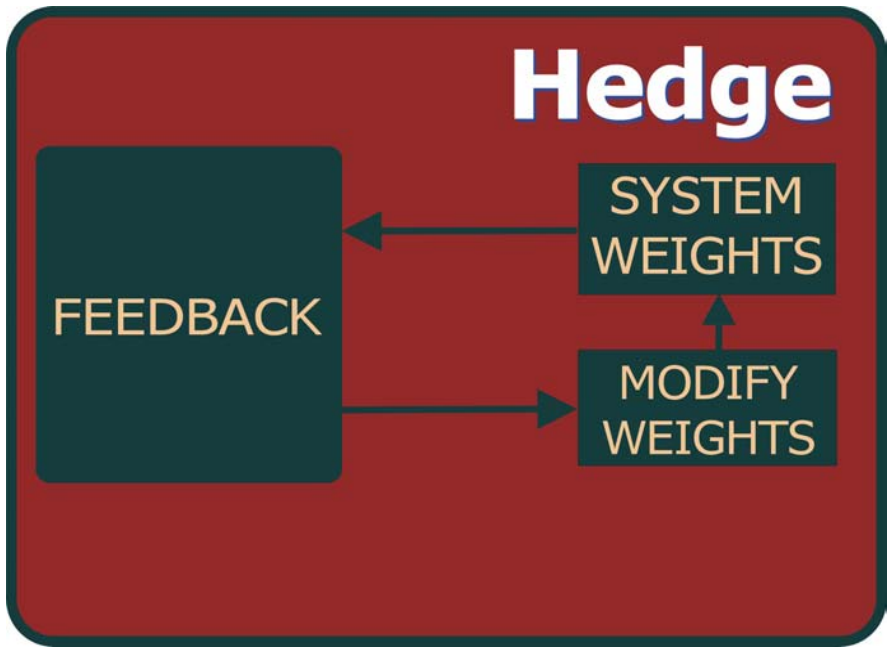
  compute "belief" in h_t

  update distribution $D_{t+1} = \frac{D_t}{Z_t} \cdot \begin{cases} e^{-\alpha_t} & \text{if } y_i = h_t(x_i) \\ e^{\alpha_t} & \text{if } y_i \neq h_t(x_i) \end{cases}$

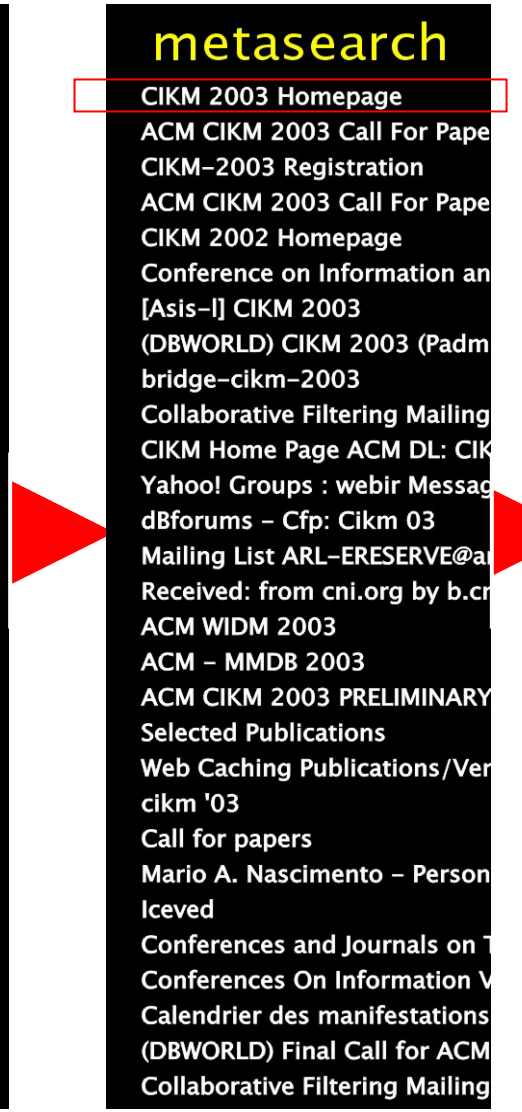
- final hyp $H_{\text{final}}(x) = \text{sgn}\left(\sum_t \alpha_t h_t(x)\right)$

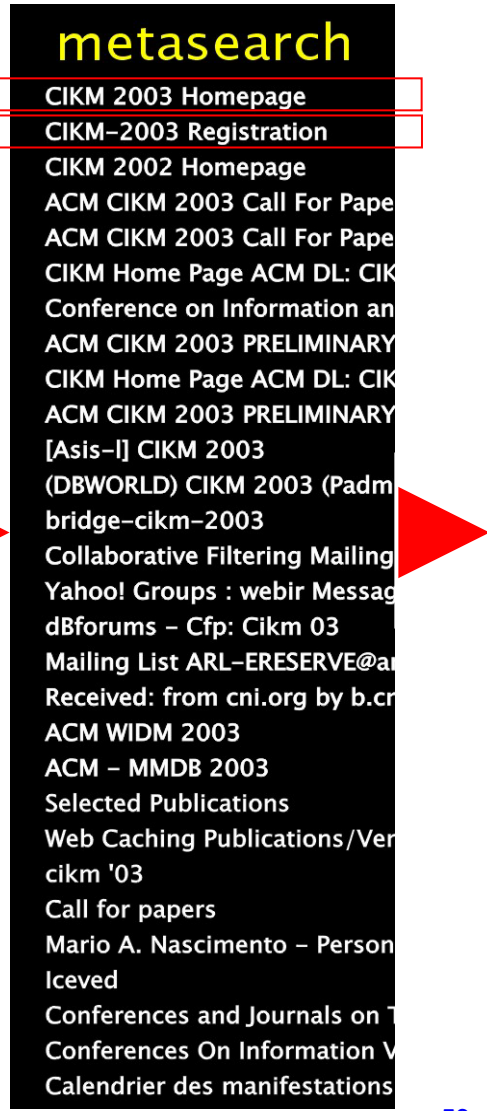
On-line Setup





On-line Metasearch [Aslam, Pavlu, Savell]





hedge application : metasearch

- problem : metasearch
  - search engines
  - document judgment
  - metaserach
- [hedge] solution
  - search engines are “experts”
  - judgments on documents are “loses”

an IR setup

generalization error- based on margins

- training error [Schapire,Freund 1996]

$$\text{training error}(H_{\text{final}}) \le \prod_t \sqrt{4\varepsilon_t(1-\varepsilon_t)}$$

- generalization error [Schapire,Freund,Barlett,Lee 1998]

Theorem 1 *Let $\mathcal{D}$ be a distribution over $X \times \{-1, 1\}$, and let S be a sample of m examples chosen independently at random according to $\mathcal{D}$. Assume that the base-classifier space $\mathcal{H}$ is finite, and let $\delta > 0$. Then with probability at least $1 - \delta$ over the random choice of the training set S, every weighted average function $f \in \mathcal{C}$ satisfies the following bound for all $\theta > 0$:*

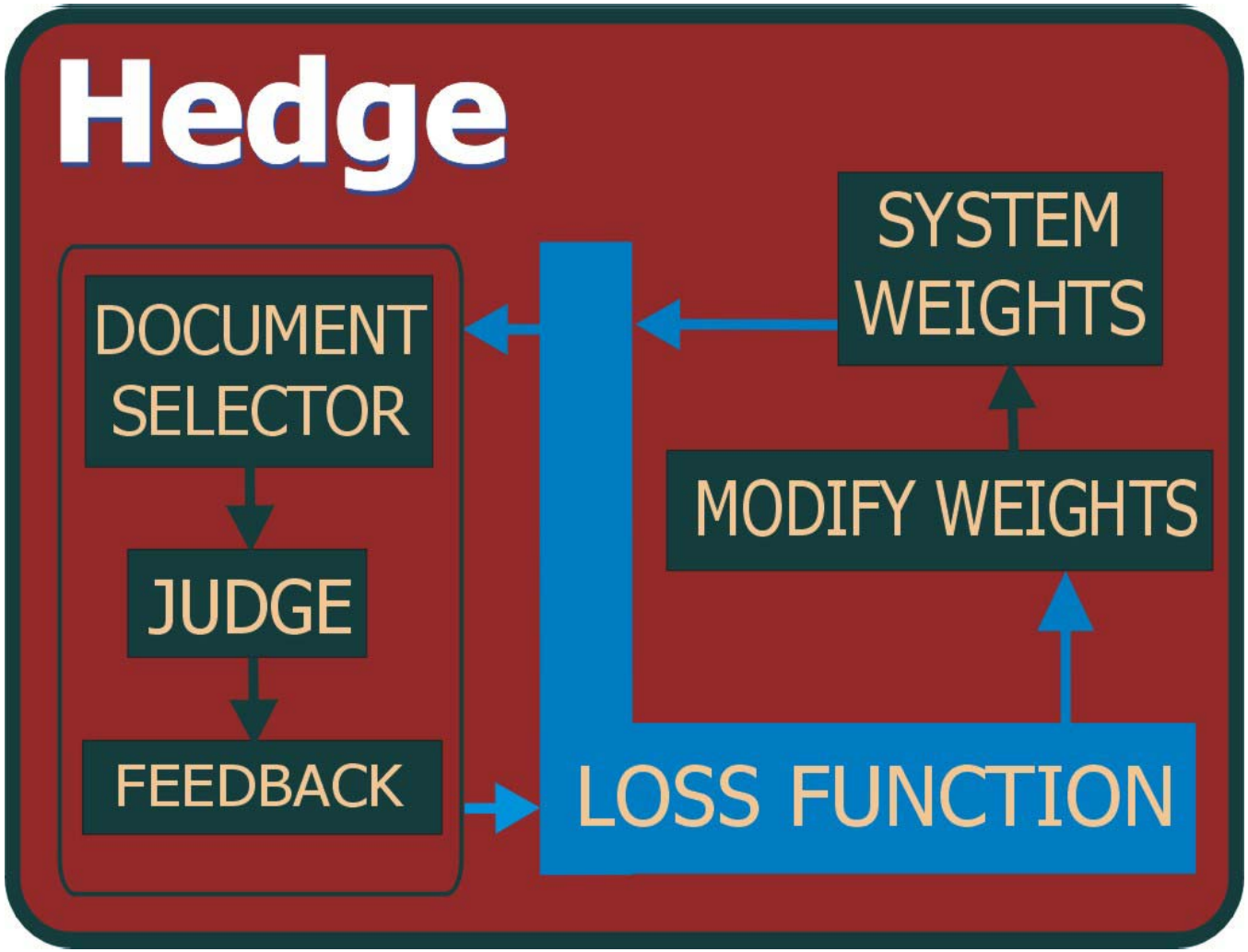
$$\mathbf{P}_{\mathcal{D}}[yf(x) \le 0] \le \mathbf{P}_S[yf(x) \le \theta] + O\left(\frac{1}{\sqrt{m}}\left(\frac{\log m \log|\mathcal{H}|}{\theta^2} + \log(1/\delta)\right)^{1/2}\right).$$

testing error

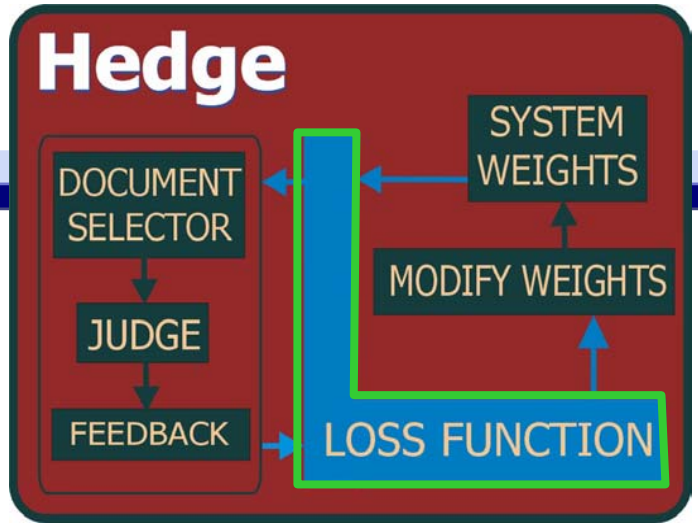
training miss-confidence

does not depend on T

hedge approach



loss function



- map ranks to values

$$value(r) = \frac{1}{r} + \frac{1}{r+1} + \ldots + \frac{1}{Z} \approx \ln \frac{Z}{r}$$

$RELEVANT = -1$

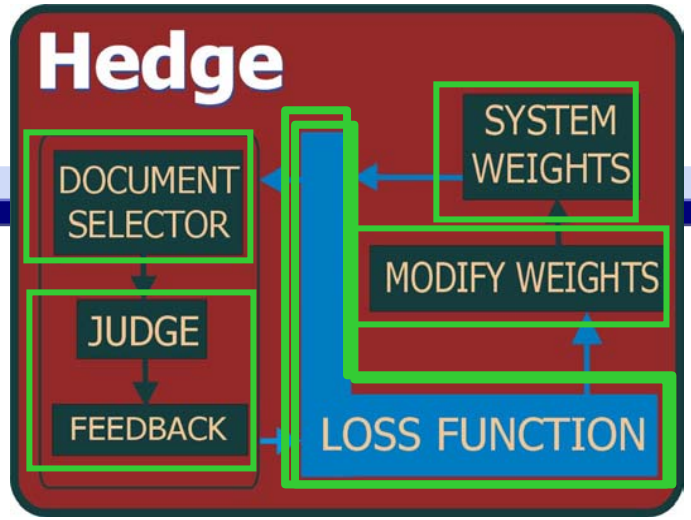
$NONRELEVANT = +1$

$$LOSS(d,s) = label(d) \cdot value(rank_{d,s}) \approx label(d) \cdot \ln \frac{Z}{r}$$

$TP(s) =$ total precision of $s =$ average of precision values at all ranks

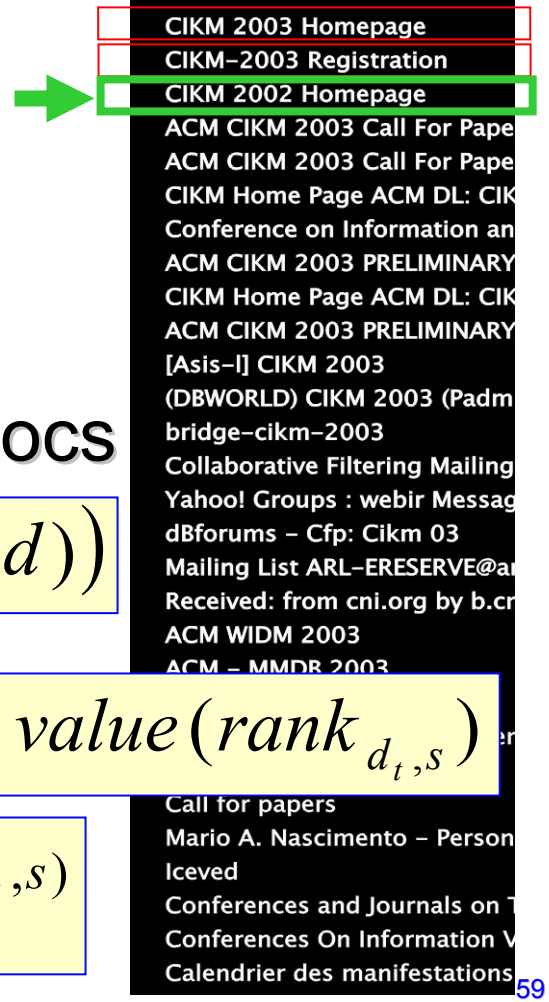
FACT : $\sum_{\text{all docs}} LOSS(d,s) = C * (Z - 2 * TP(s))$

- Average the precision at ALL ranks
- Normalize so ideal system gets TP=1
- math is simpler



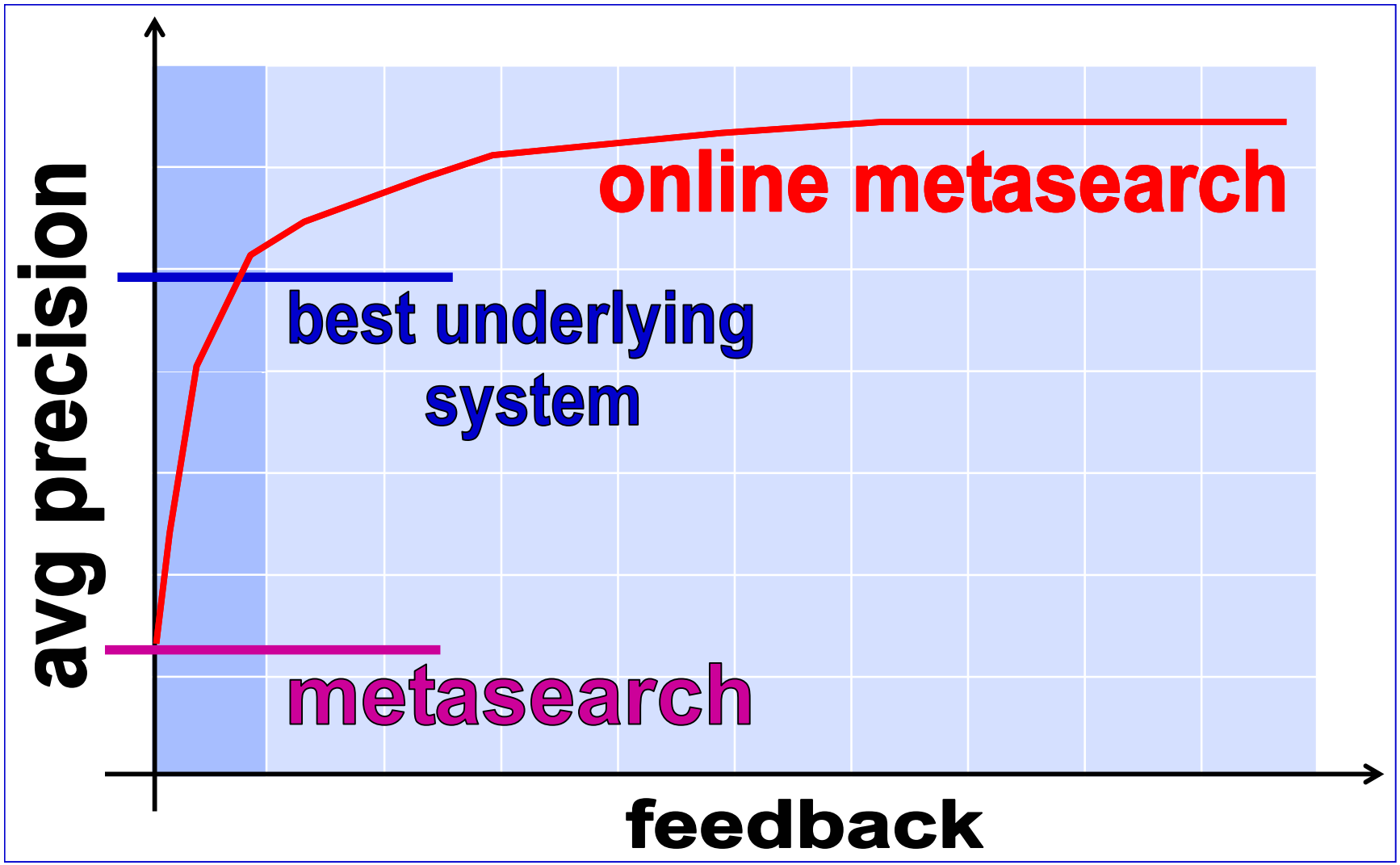
$$average\_value_t(d) = \sum_{s=1}^{N} w_s^{t-1} \cdot value(rank_{d,s})$$

ork ?



- average_value at episode t
  - "trust" in systems change with t
- metasearch : include already judged docs
- pooling $d_t = \arg\max_d \left(average\_value(d)\right)$
  - system evaluation
- get feedback $LOSS\ (d_t, s) = label(d_t) \cdot value(rank_{d_t,s})$
- modify weights $w_s^{t+1} = w_s^t \cdot \beta^{LOSS_t(d_t,s)}$

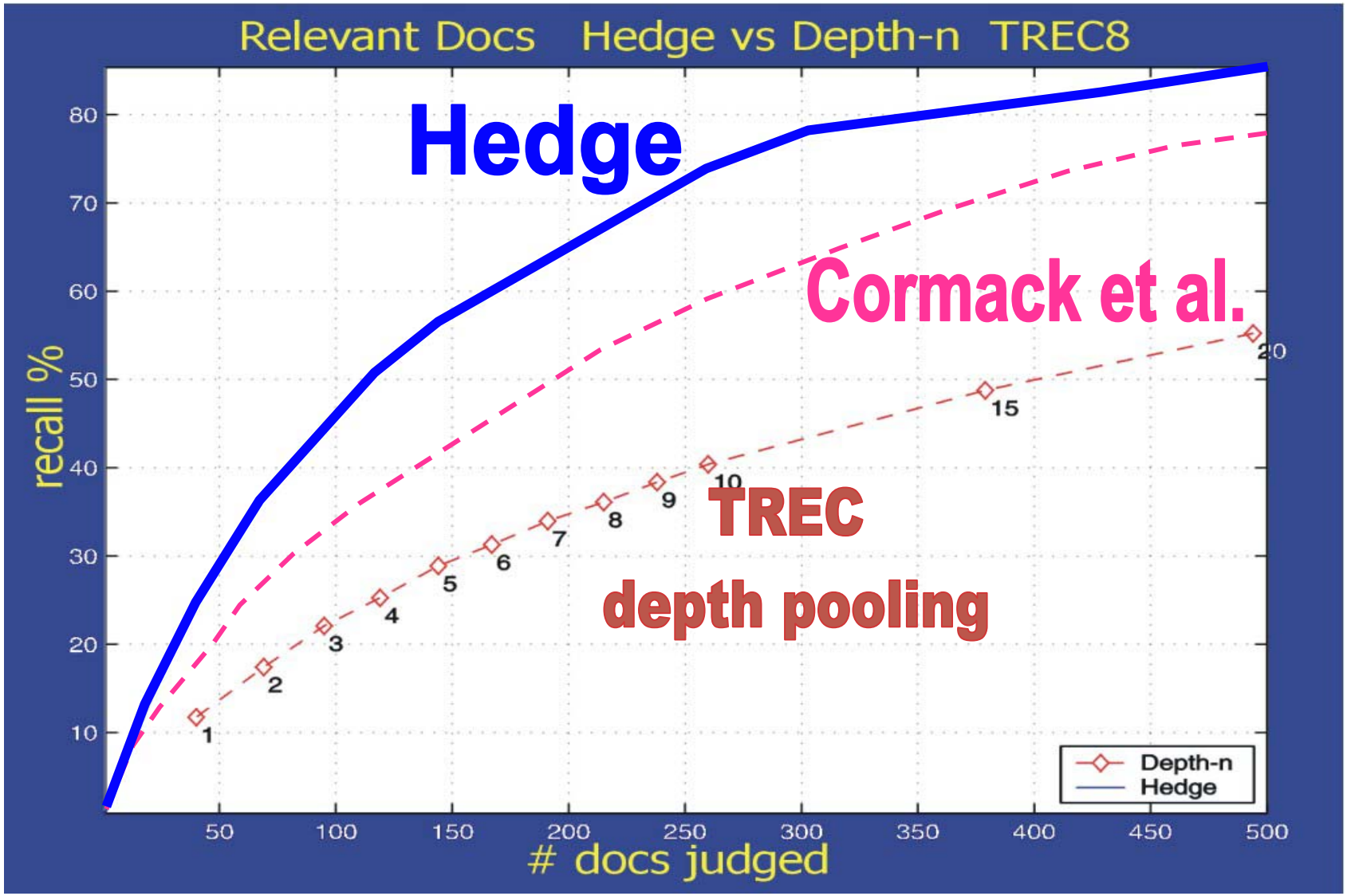
actually…we do more than metasearch



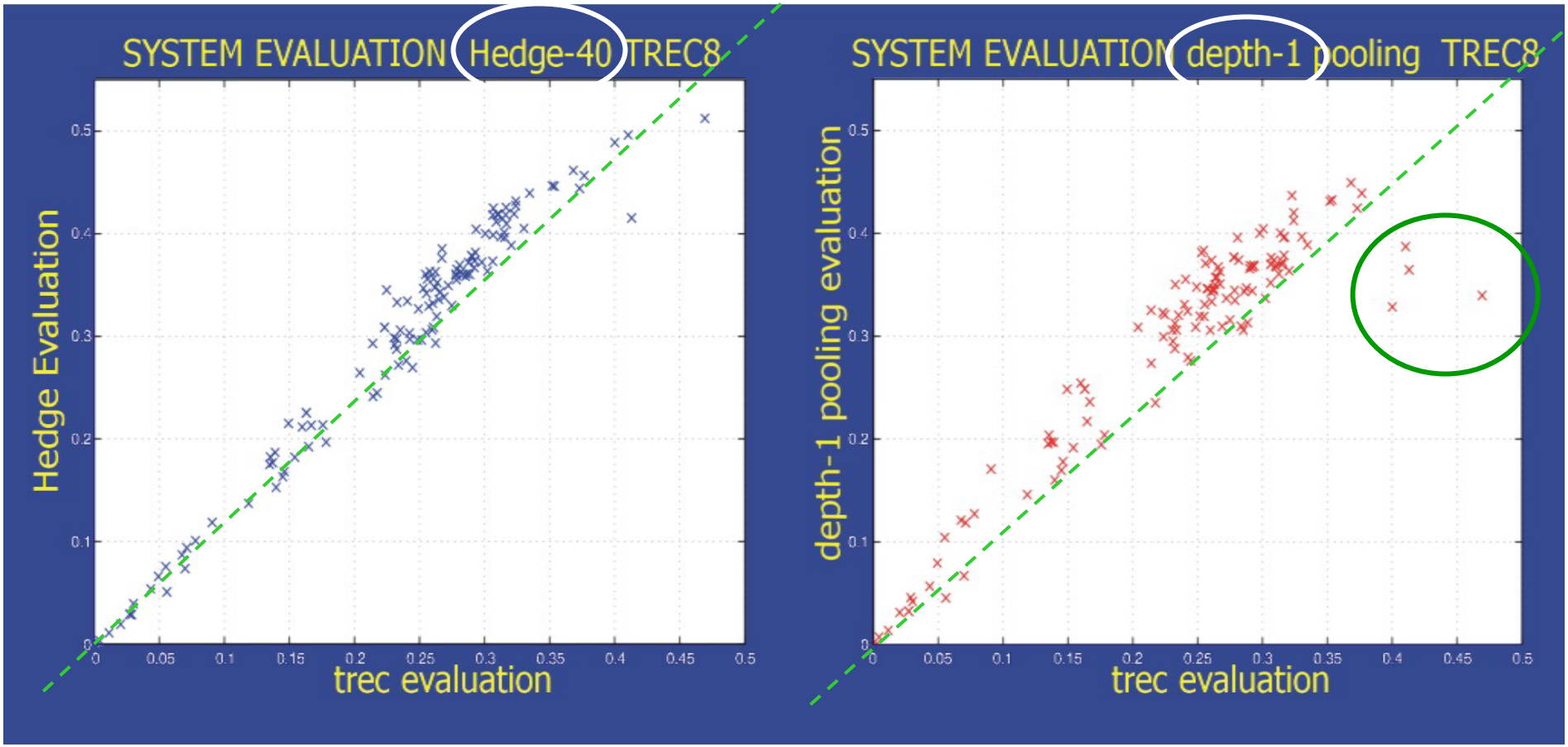
experiments

- TREC 3,5,6,7,8
  - 41-129 systems
  - 50 queries per TREC
  - metasearch uses *all* systems
- use TREC judgments as user feedback
- system evaluation : incomplete judgments

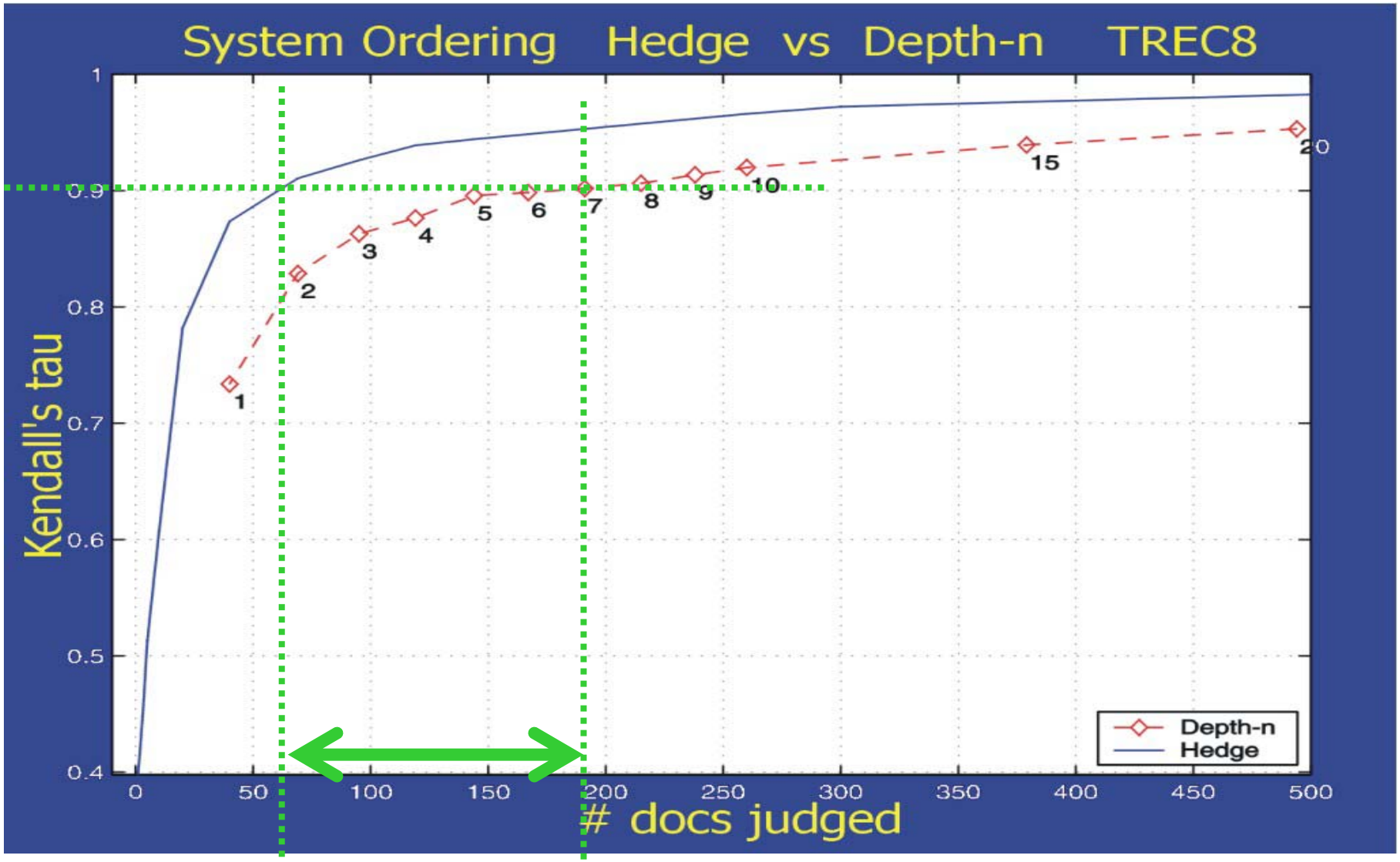
experiments - relevant docs found



experiments - system evaluation



system evaluation – kendall's τ



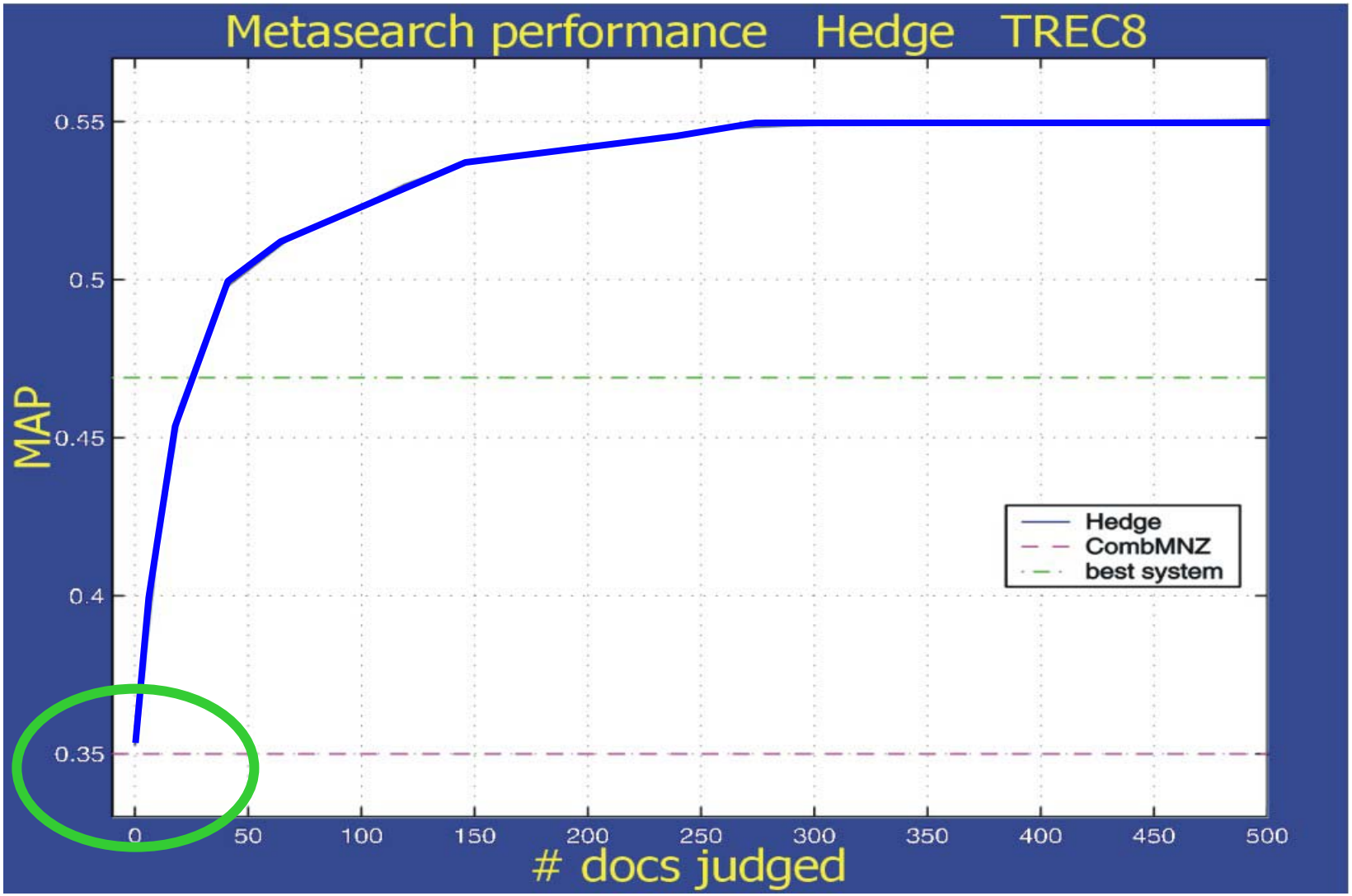
metasearch - no feedback (yet)

no relevant judgements → uniform weights

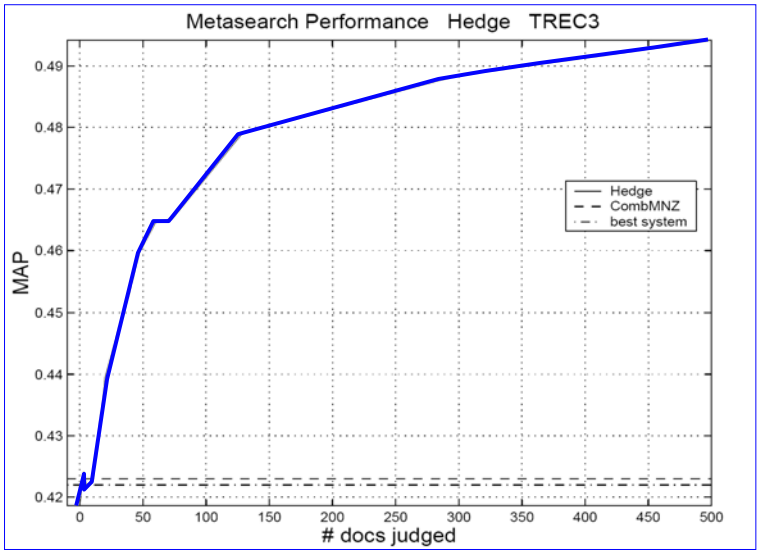
TREC	MNZ	COND	Hedge-0	%MNZ	%COND
3	0.423	0.403	0.418	−1.2	+3.7
5	0.294	0.307	0.309	+5.1	+0.6
6	0.341	0.315	0.345	+1.2	+9.5
7	0.320	0.308	0.323	+0.9	+4.9
8	0.350	0.343	0.352	+1.4	+2.6

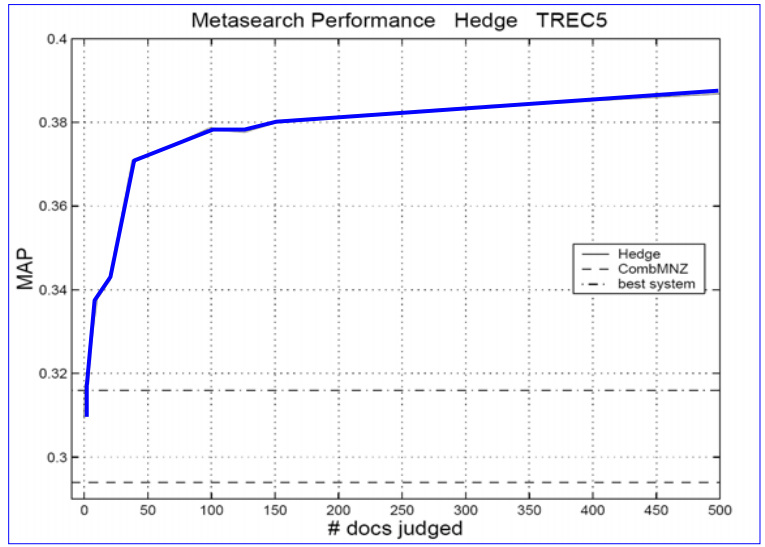
MNZ=CombMNZ(Fox,Shaw,Lee et al)
COND=Condorcet(Aslam,Montague)

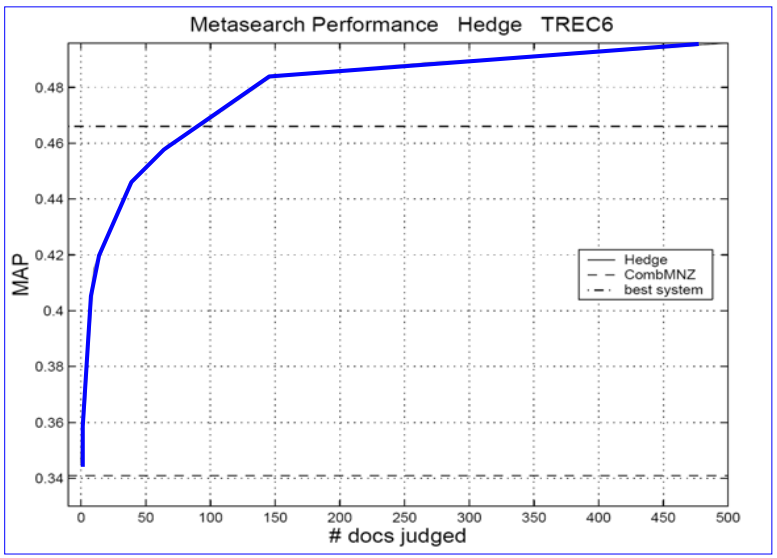
experiments – metasearch – TREC8

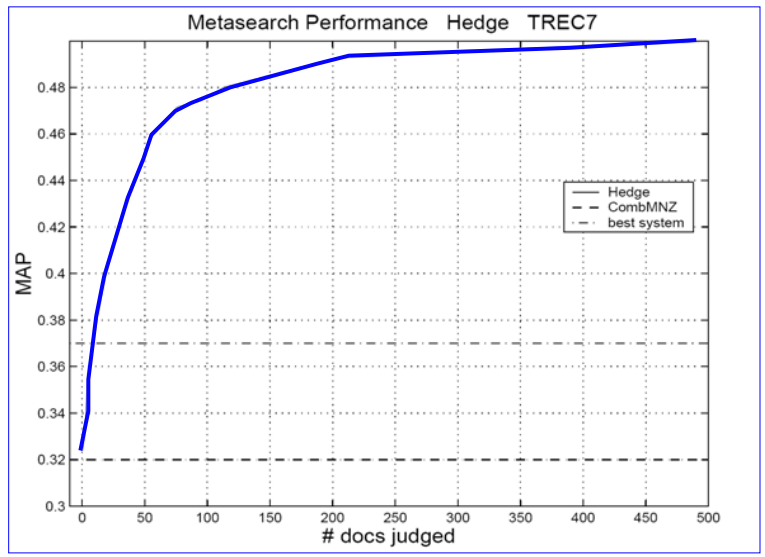


metasearch – TREC 3,5,6,7









conclusion

- a powerful machine learning approach
  - Hedge = AdaBoost core
- works [usually] better than anything else we've seen
- true, it uses feedback
  - but without feedback there are provable limitations

Why hedge [schapire, freund '96]

$$\sum_{i=1}^{N} w_i^{T+1} \le \left(\sum_{i=1}^{N} w_i^{T}\right)\left(1-(1-\beta)\mathrm{p}^{T}\cdot l^{T}\right) \le \ldots.$$

$$\ldots \le \exp\left(-(1-\beta)\sum_{t=1}^{T}\mathrm{p}^{t}\cdot l^{t}\right)$$

hedge loss at episode T $= p^{T}\cdot l^{T}$

cummulat loss L_{HEDGE}

$$L_{HEDGE} \le \frac{\ln\left(\frac{1}{\beta}\right)L_{SYSTEM} + \ln N}{1-\beta}$$

$L_{\mathrm{Hedge}(\beta)}$

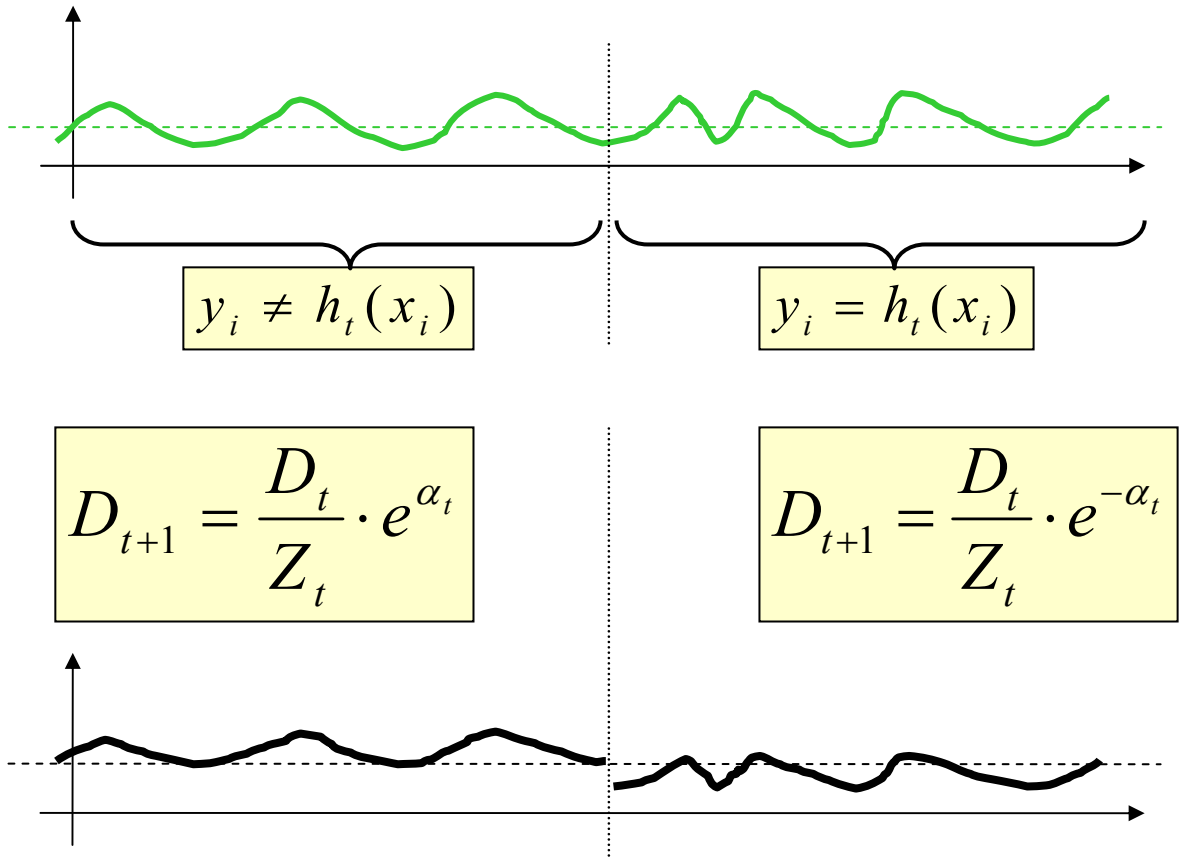
$$\sum_{i=1}^{N} w_i^{T+1}$$

AdaBoost – distribution update

$$h_t : X \rightarrow \{-1,1\}$$

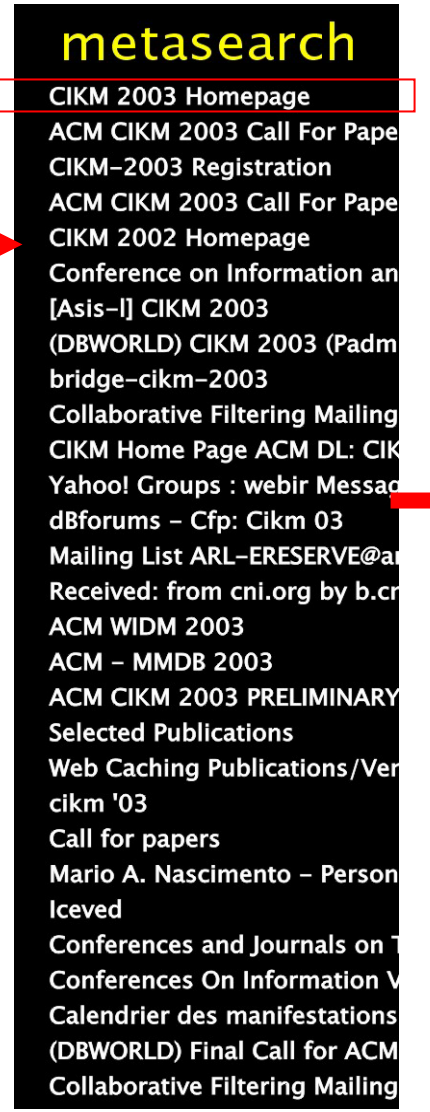
$$\varepsilon_t = \Pr_{D_t}[h_t(x_i) \neq y_i]$$

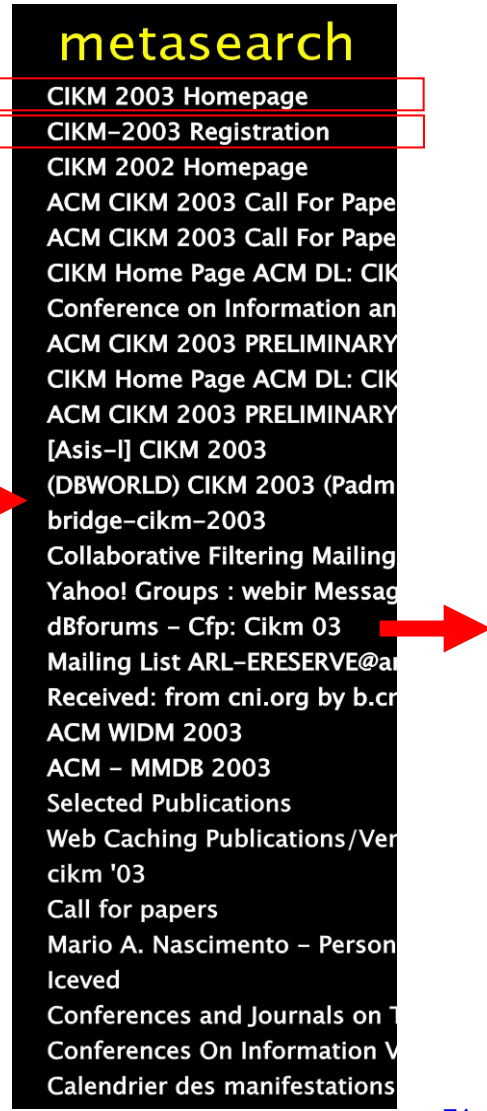
$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) > 0$$



- "neutralize" the last weak hypothesis

Applicaltion - metasearch



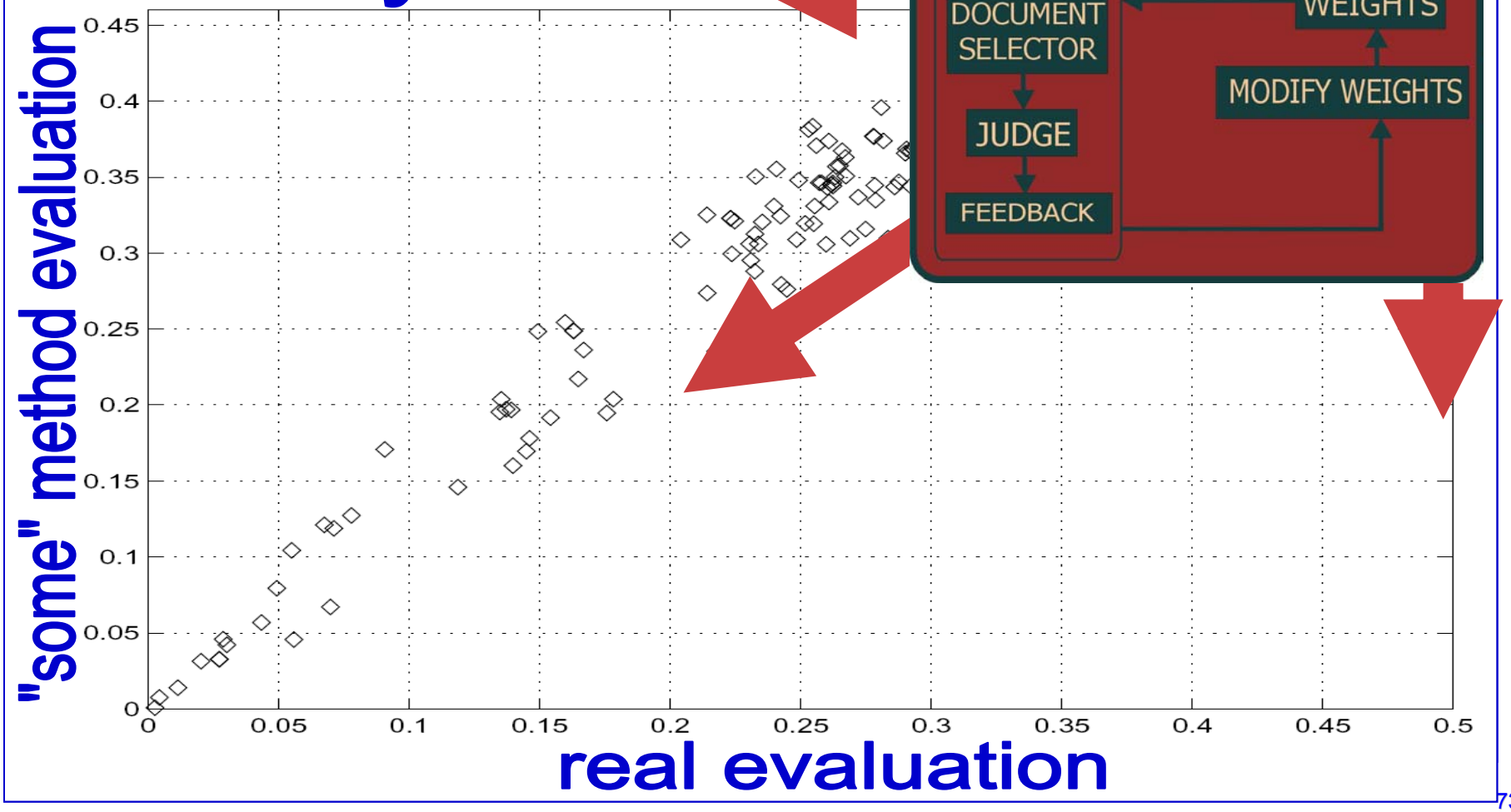


problem setup

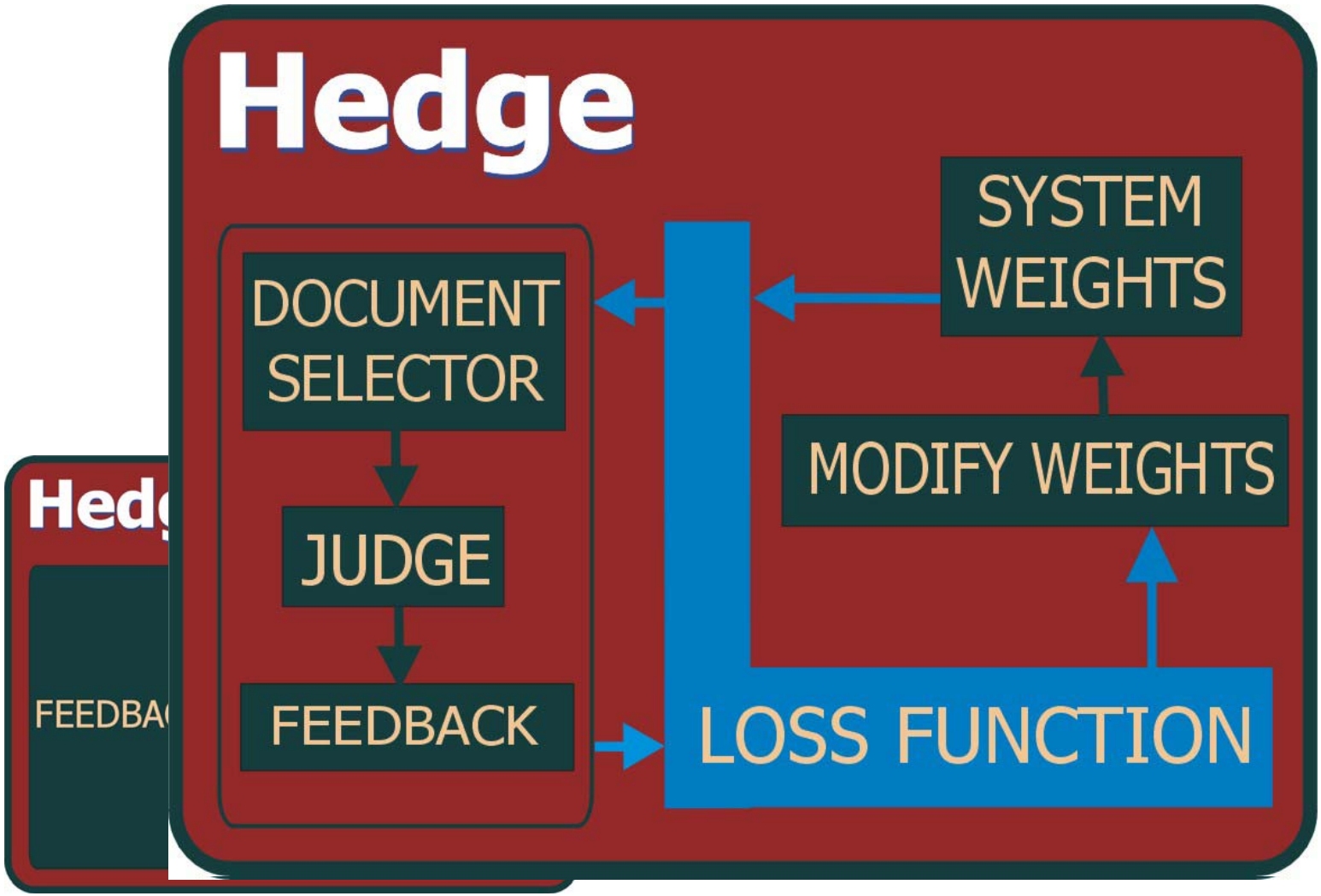
- On the same query
- Set of underlying systems
- User feedback
- Goal
  - Find relevant documents
  - Produce online metasearch lists
  - Perform online system evaluation
- We are looking for an adaptive approach

motivation

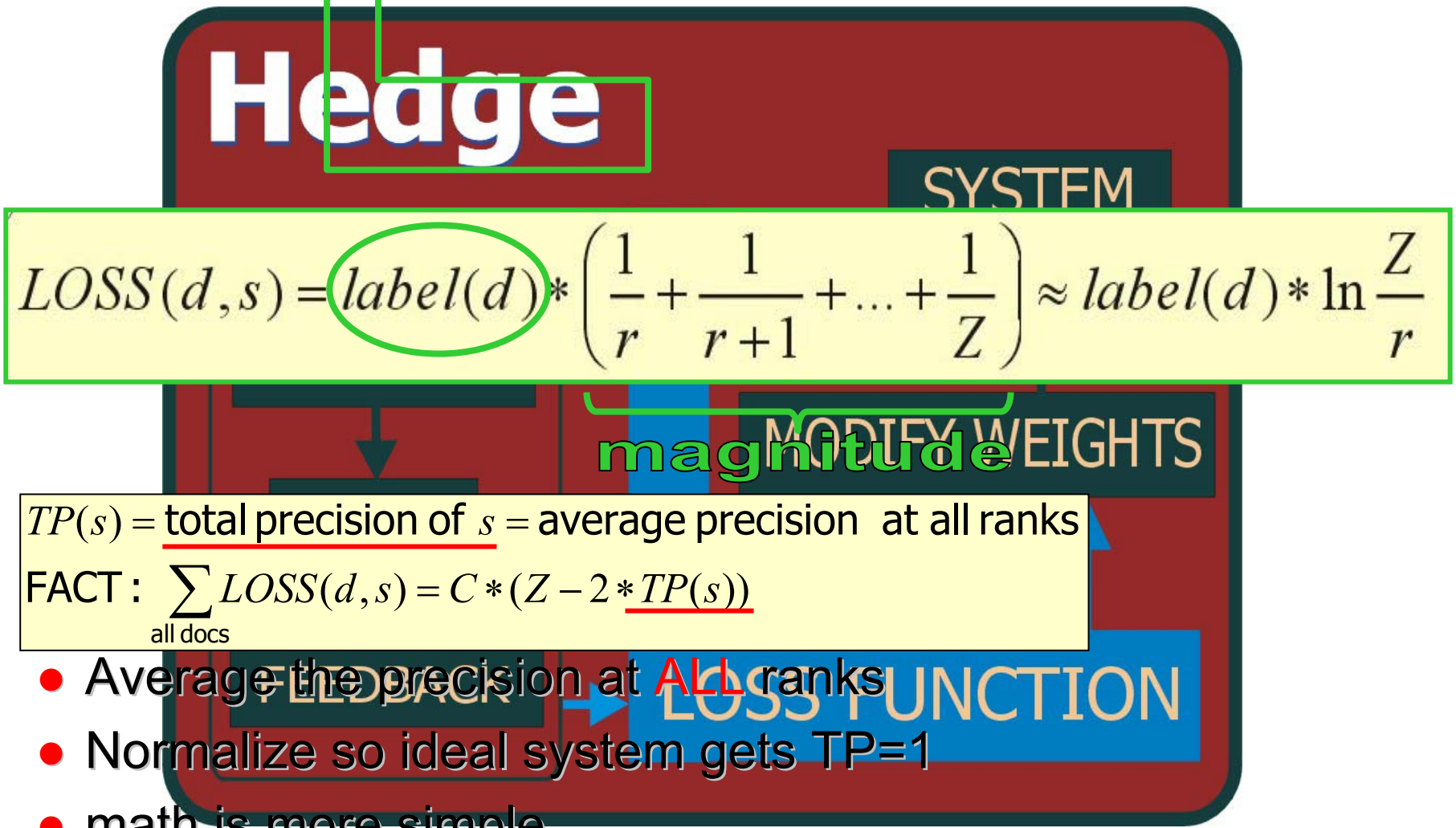
system e



our unified model



loss function



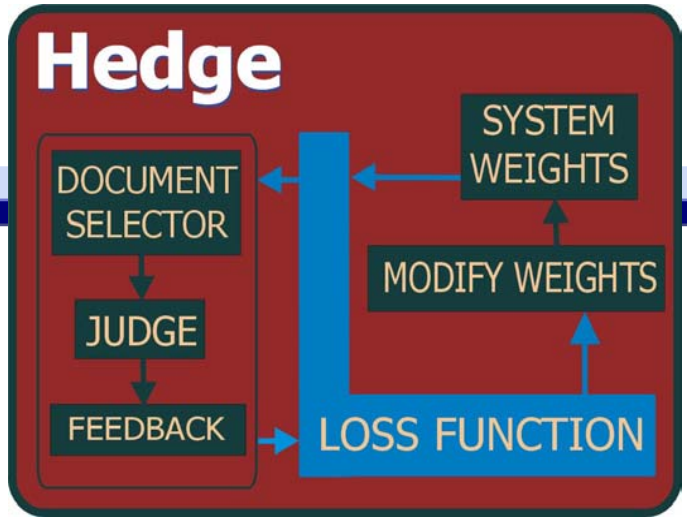
$$LOSS(d,s) = label(d) * \left(\frac{1}{r} + \frac{1}{r+1} + \ldots + \frac{1}{Z}\right) \approx label(d) * \ln\frac{Z}{r}$$

magnitude

$TP(s) =$ total precision of $s =$ average precision at all ranks

FACT : $\sum_{\text{all docs}} LOSS(d,s) = C * (Z - 2 * TP(s))$

- Average the precision at ALL ranks
- Normalize so ideal system gets TP=1
- math is more simple



pooling – howto

pooling value(d)

$$\text{- select } d_t = \arg\max_{d \text{ not labeled}} \left[\sum_{s=1}^{N} w_s^{t-1} * LOSS(d, s \mid d = NR) \right]$$

- Naturally “want” top ranks
- If NON RELEVANT, then a NR in top ranks of the system lists
- If RELEVANT, bingo.

system evaluation – howto

- assume all docs not judged (so many ?) to be NON RELEVANT
- compute AvegPrecision for every system
- one (or few) very good systems – use small β

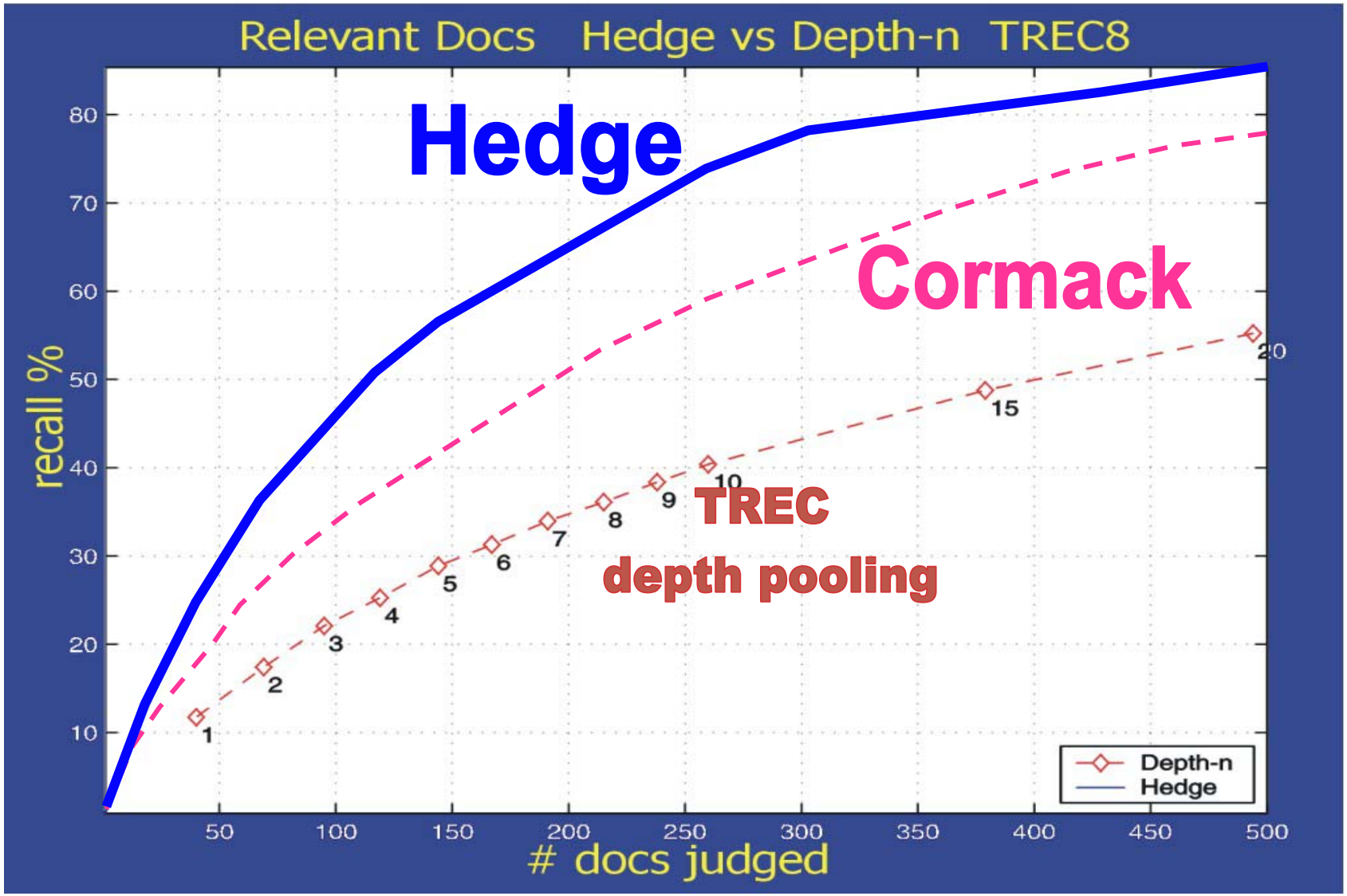
metasearch - howto

- Compute “pooling value” for each doc
  - Instead of “select the top doc” for pooling
  - do “select the top 1000 doc” for metasearch
- almost 1000 – docs already pooled are automatically in top of metasearch list

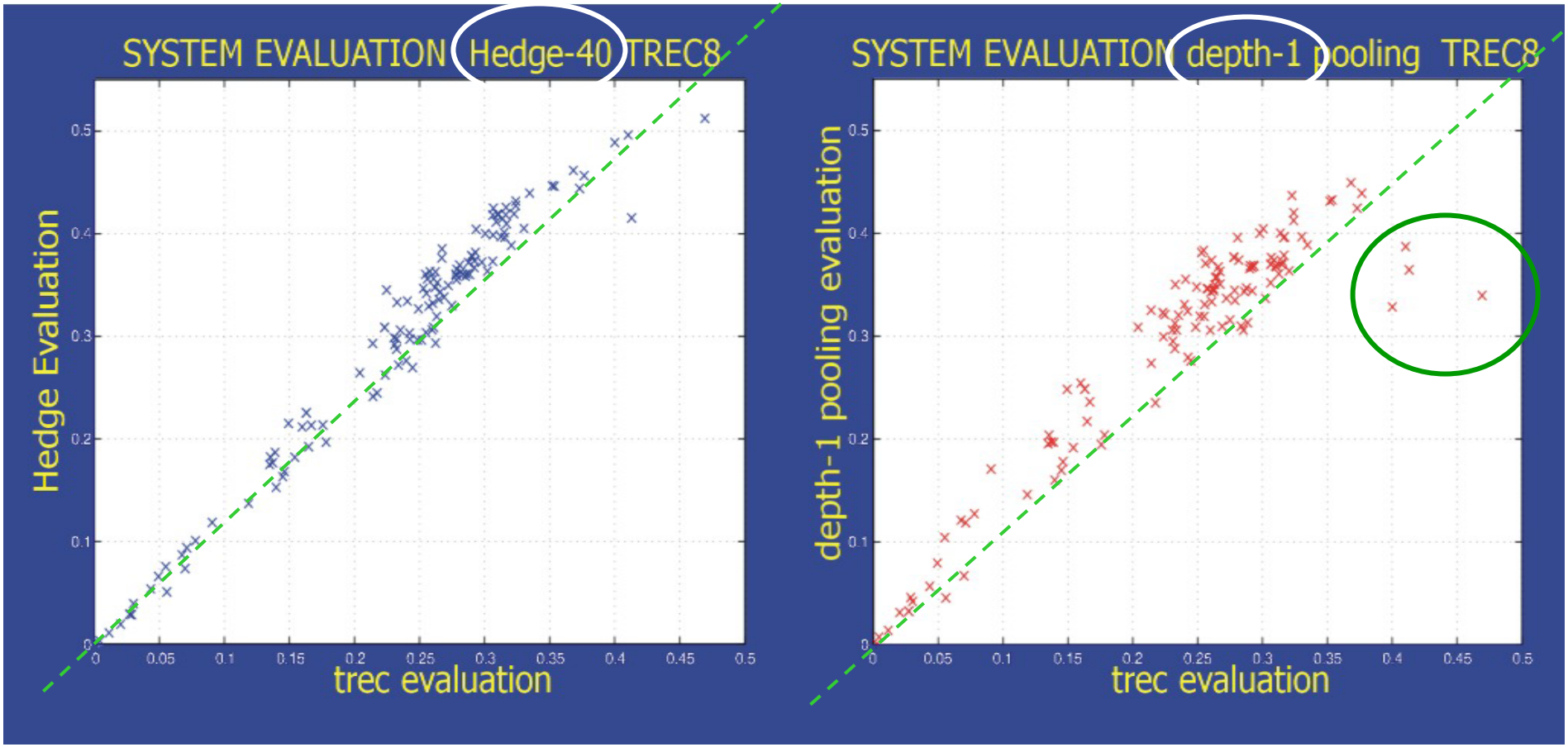
experiments

- TREC
  - ~100 systems
  - 50 queries each competition
- Use TREC qrels as user feedback
  - incomplete feedback
- For comparison with depth-pooling we use average number of pools (over queries)

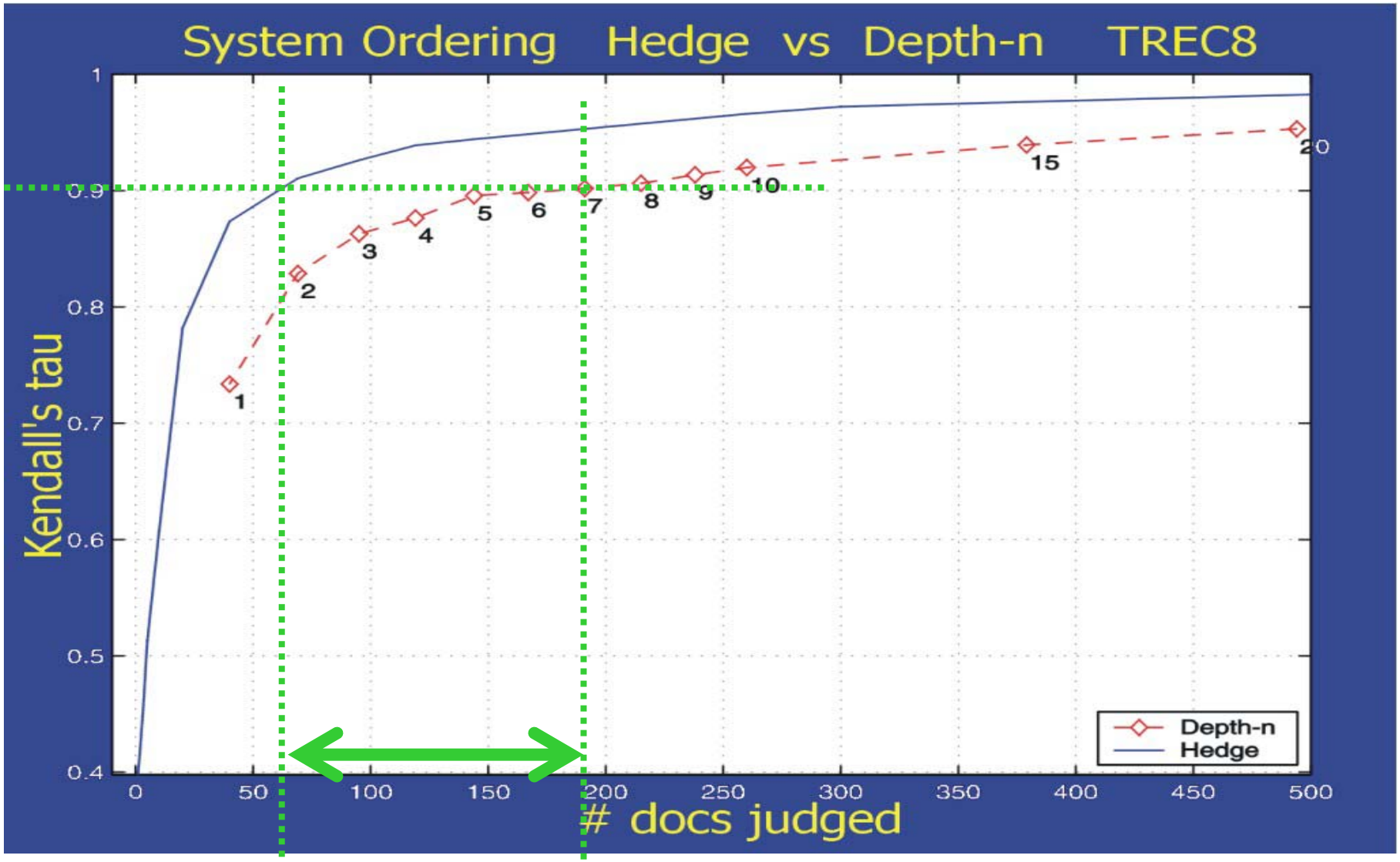
experiments - relevant docs found



experiments - system evaluation



system evaluation – kendall's tau

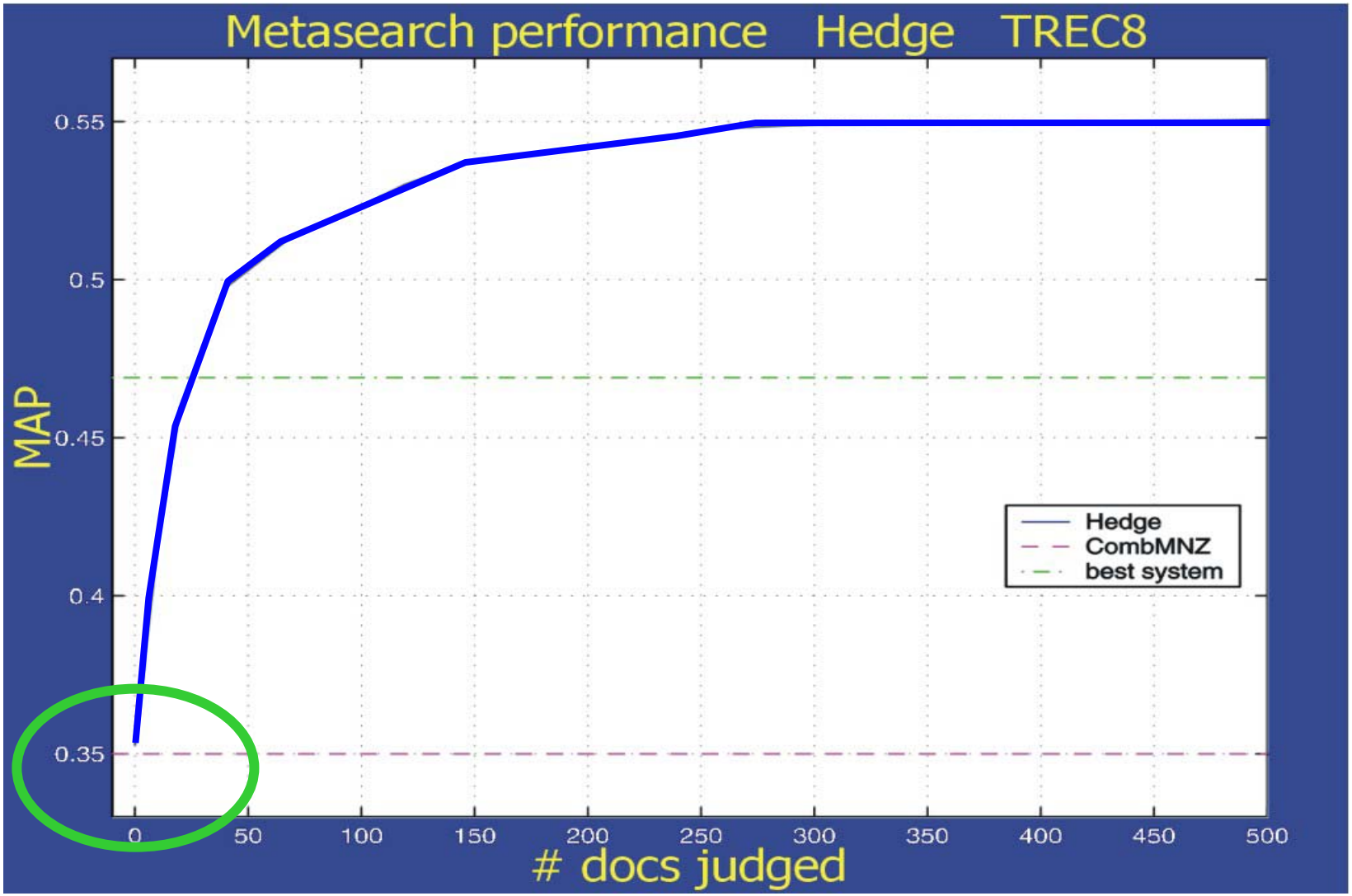


metasearch - no feedback (yet)

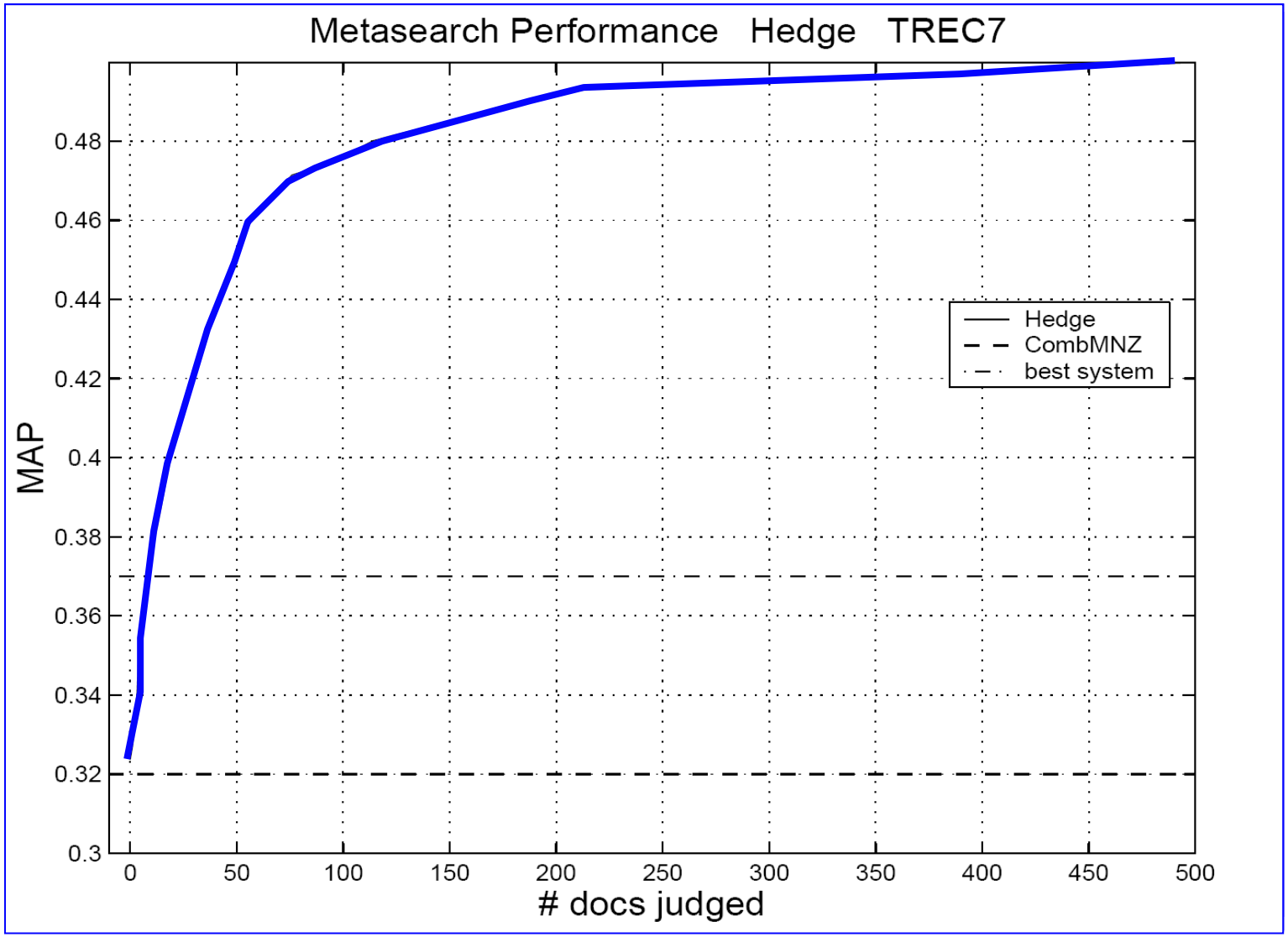
no relevant judgements

TREC	MNZ	COND	Hedge-0	%MNZ	%COND
3	0.423	0.403	0.418	−1.2	+3.7
5	0.294	0.307	0.309	+5.1	+0.6
6	0.341	0.315	0.345	+1.2	+9.5
7	0.320	0.308	0.323	+0.9	+4.9
8	0.350	0.343	0.352	+1.4	+2.6

experiments – metasearch – TREC8



metasearch – TREC 3,5,6,7



conclusion

- A powerful machine learning approach
  - Hedge = AdaBoost core
- Works [usually] better than anything else we've seen
- True, it uses feedback
  - But without feedback there are provable limitations
- It is missing a rigorous analysis
  - We are not very far away with that
  - Need a model assumption

After conclusion – don't read

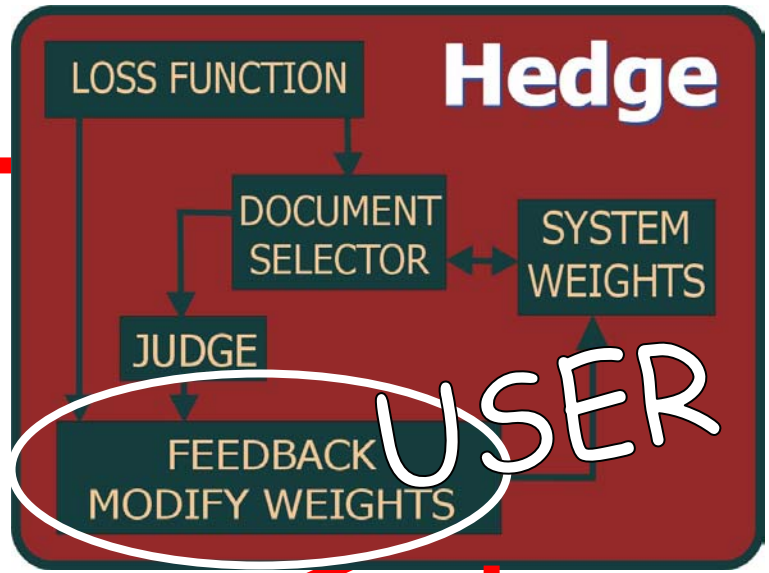
differentiate (clasify) the search engines

RANK	SYS1	SYS2	SYS3	SYS129
1	8	111	2	1013
2	2	2	821	1
3	1	4	13	8
4	134	7284	9 5	5
5	14215	769	106	2038
6	24	84	439	4
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1000	8521	11003	84	3967

84=RELEVANT 821=NON RELEVANT

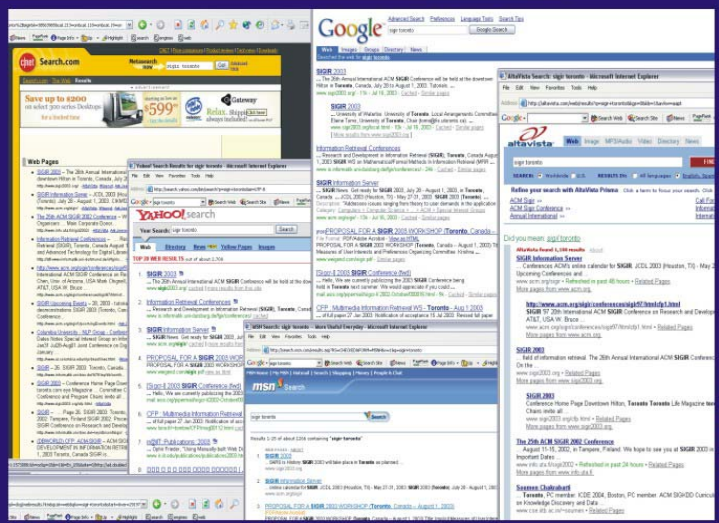
SYSTEM	ORDERING(MAP)
READWARE2	0.469245750504929
orcl99man	0.413044919590457
iit99ma1	0.4103611 615777
[illegible]	40 658
CL9 [illegible]Top	0 6527 2 016
CL9 [illegible]T	0 0432 2 631
[illegible]	0 6 1
CL99SD	0.353710984082113
CL99SDopt2	0.351955930155801
8manexT3D1N0	0.334646677365296
pir9Aatd	0.330293582307575
ok8alx	0.324046294809153
Flab8atdn	0.323972048518037
MITSLStdn	0.32268558626124
pir9Attd	0.320651128642633
ok8amxc	0.31687789001205
att99atde	0.316549112790725
Flab8ax	0.316270864248346
apl8p	0.315425192468865
apl8c221	0.314991172213836
UniNET8Lg	0.313842898958697
apl8c621	0.312624929467966
fub99tt	0.310608000938879
fub99tf	0.309923242027166
att99atdc	0.308935872304471
fub99a	0.306768289435422
GE8MTD2	0.306560441783452
fub99td	0.306376965574526
pir9At0	0.306288595615414
pir9Atd0	0.302238406288713

TREC8 competition
129 search engines (systems)
~1000 docs returned /query
TREC Evaluation DEPTH 100
-judge top 100 docs (each sys); use MAP to rank systems



fuse the lists (metasearch)

toronto sigir Search

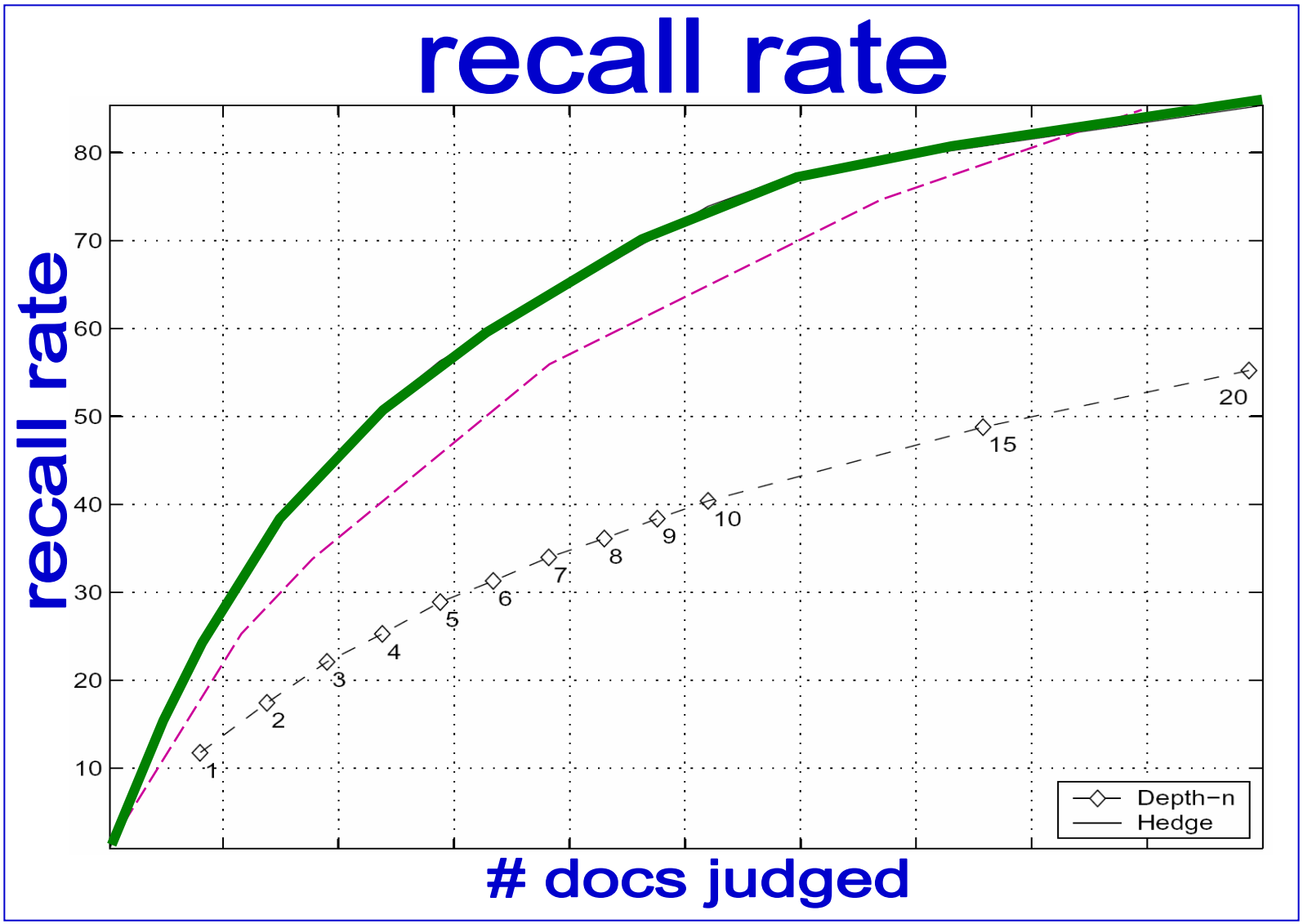


metasearch list

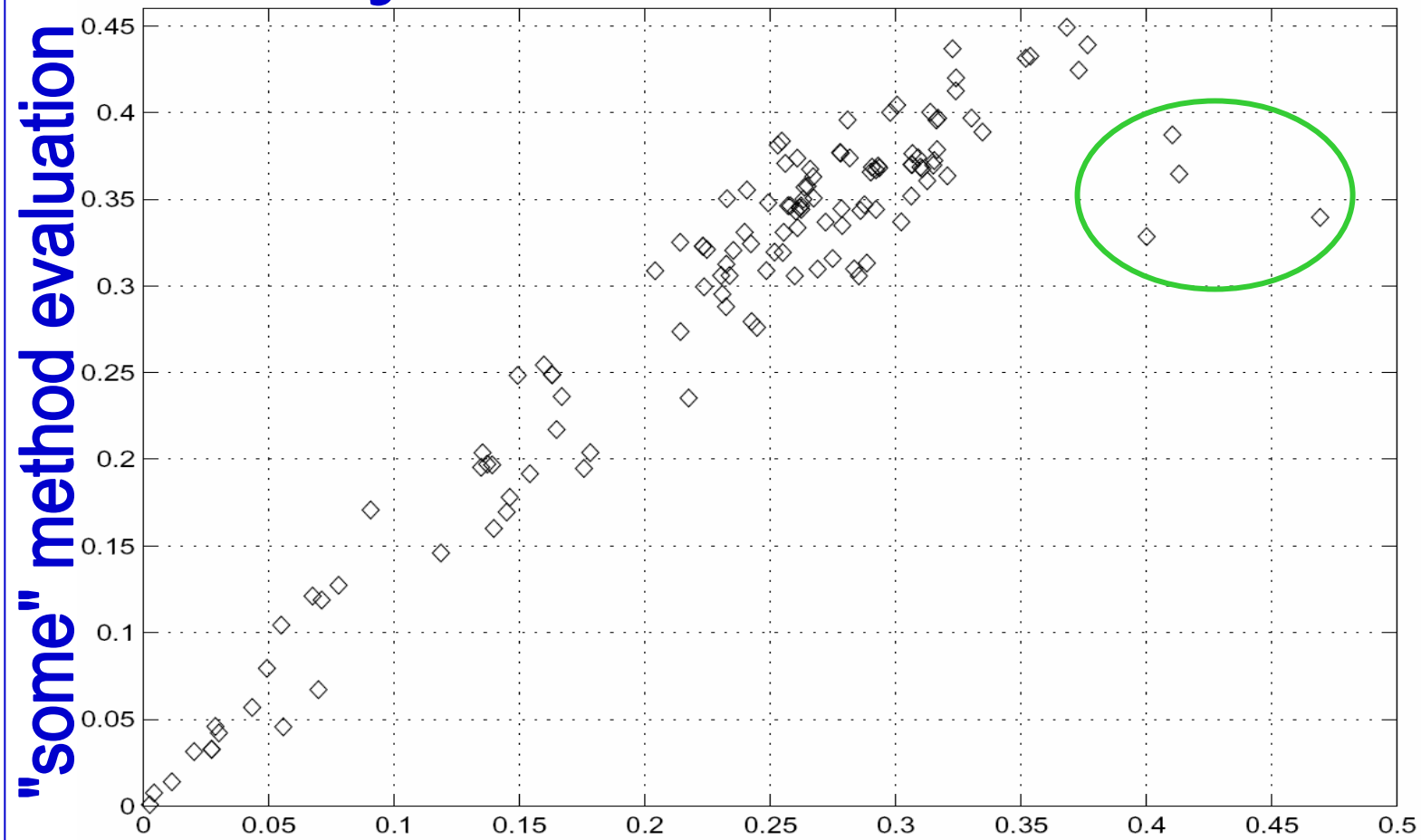
1 SIGIR 2003
2 Information Retrieval Conferences
3 SIGIR Information Server
4 SIGIR Upcomming events
5 PROPOSAL FOR A SIGIR 2003 WORKSHOP
6 Andreas S. WEIGEND, PhD
7 [Sigir-l] 2003 SIGIR Conference -
8 CFP : Multimedia Information Retrieval
9 The 25th ACM SIGIR 2002 Conference
10 Mark Girolami's home page -
11 Conferences on Information Retrieval
12 OntoWeb - SIGIR 2003 -2003-01-
13 Yahoo! Groups : webir Messages
14 DBWorld Message /2002-
15 Publications and Presentations of Thomas
16 SIGIR 2003 Workshop on the Evaluation
17 David Crow: Upcoming Conferences
18 Upcoming conferences for the WUME Lab
19 SIGIR 99 Preliminary Program
20 SIGIR 99 CFP/PC
21 Mounia Lalmas's list of activities
22 A Partial List of Publications of Hongyuan
23 ACM SIGIR 2002 Tutorial Bibliography
24 SIGIR <http://www.informatik.uni-
25 Information Retrieval Links
26 iSchool Events

KEY:
FIND RELEVANT DOCS

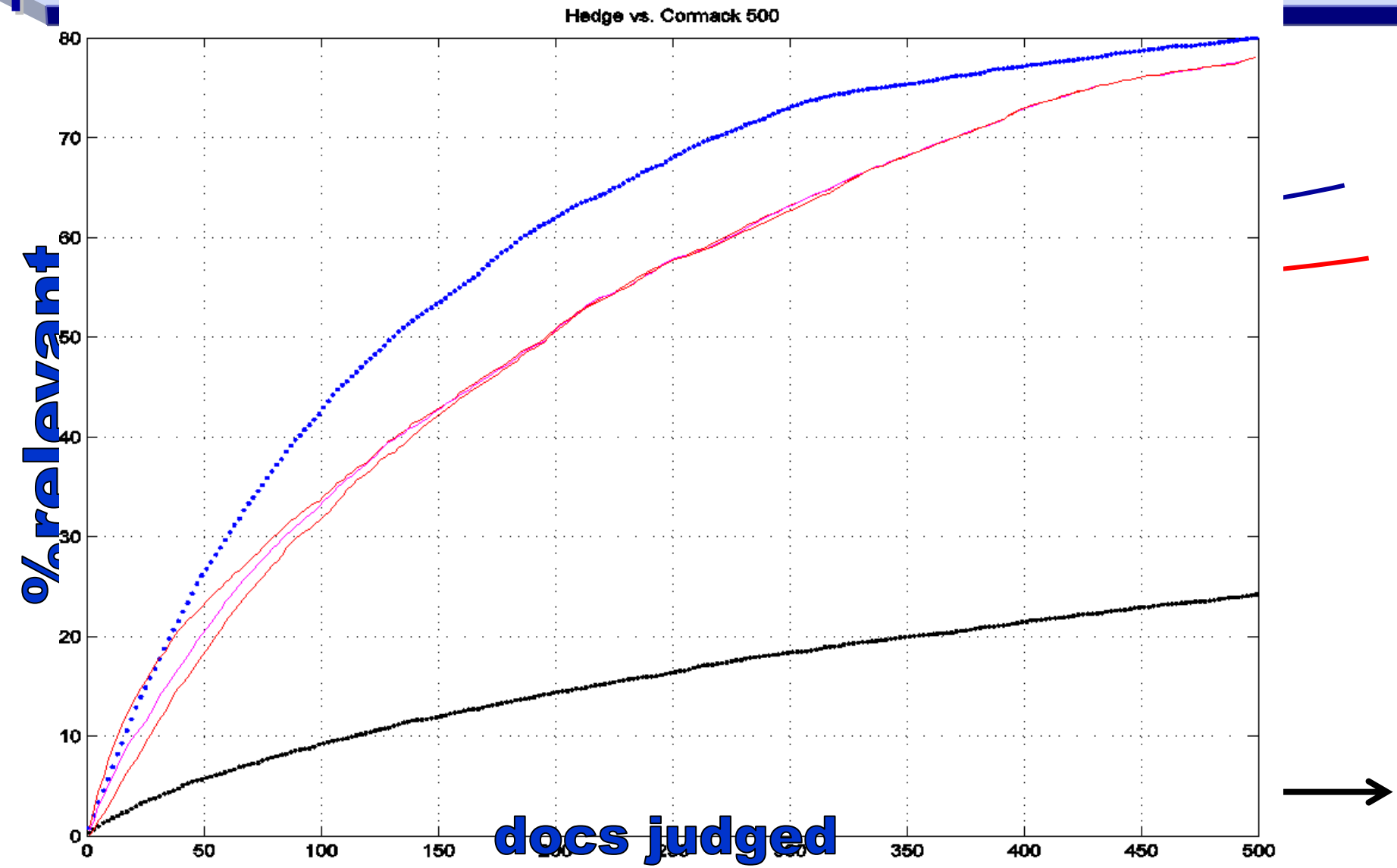
test


recall rate
recall rate
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10
15
20
Depth−n
Hedge
docs judged

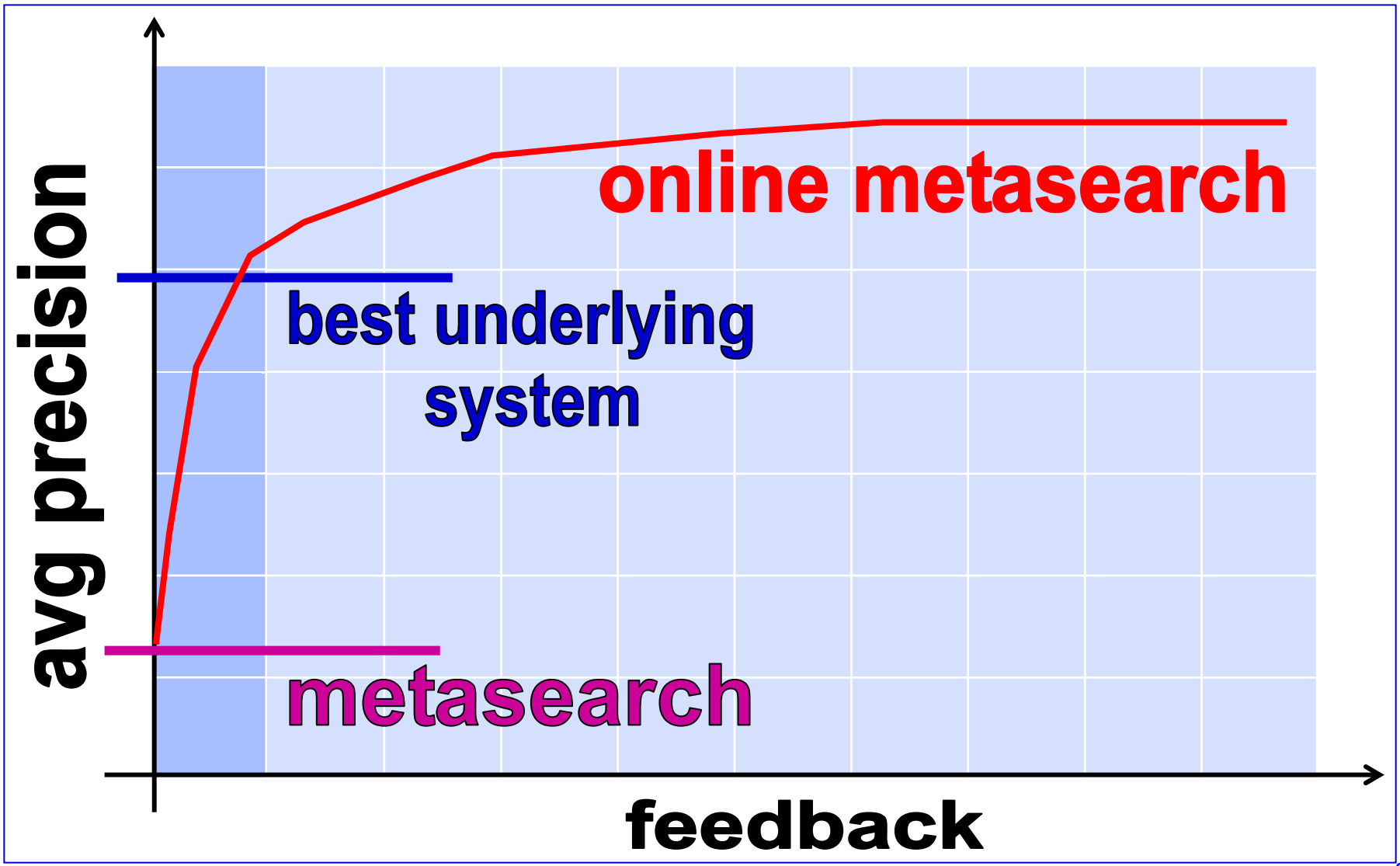
system evaluation

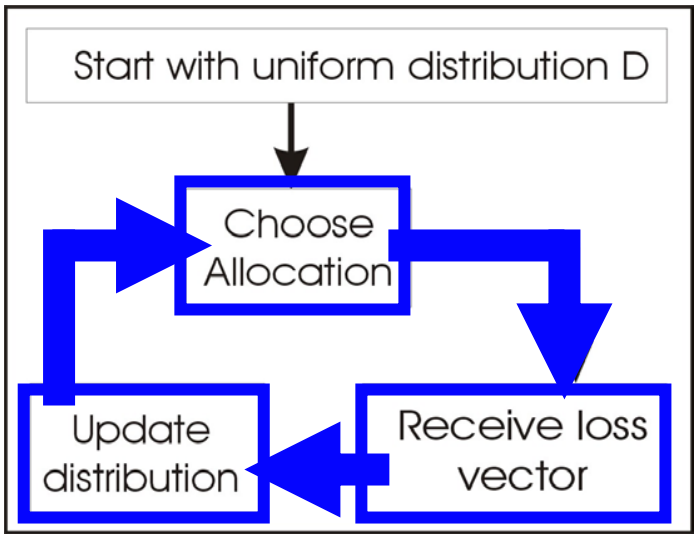


pooling - comparison with Cormack



motivation


avg precision
online metasearch
best underlying system
metasearch
feedback


Start with uniform distribution D
Choose Allocation
Receive loss vector
Update distribution

pooling - howto

$$\text{- select } d_t = \underset{d \text{ not labeled}}{\arg\max} \left[\sum_{s=1}^{N} w_s^{t-1} * LOSS(d, s \mid d = NR) \right]$$

pooling value(d)

$$LOSS(d, s \mid d = NR) = \left(\frac{1}{r} + \frac{1}{r+1} + \ldots + \frac{1}{Z} \right) \approx \ln \frac{Z}{r}$$

- Naturally “want” top ranks
- If NON RELEVANT, then a NR in top ranks of the system lists
- If RELEVANT, bingo.

metasearch – howto

- Before the next episode
  - Compute "pooling value"

for each doc

$$\left[\sum_{s=1}^{N} w_s^{t-1} * LOSS(d, s \mid d = NR)\right]$$

  - Instead of "select the top doc" for pooling

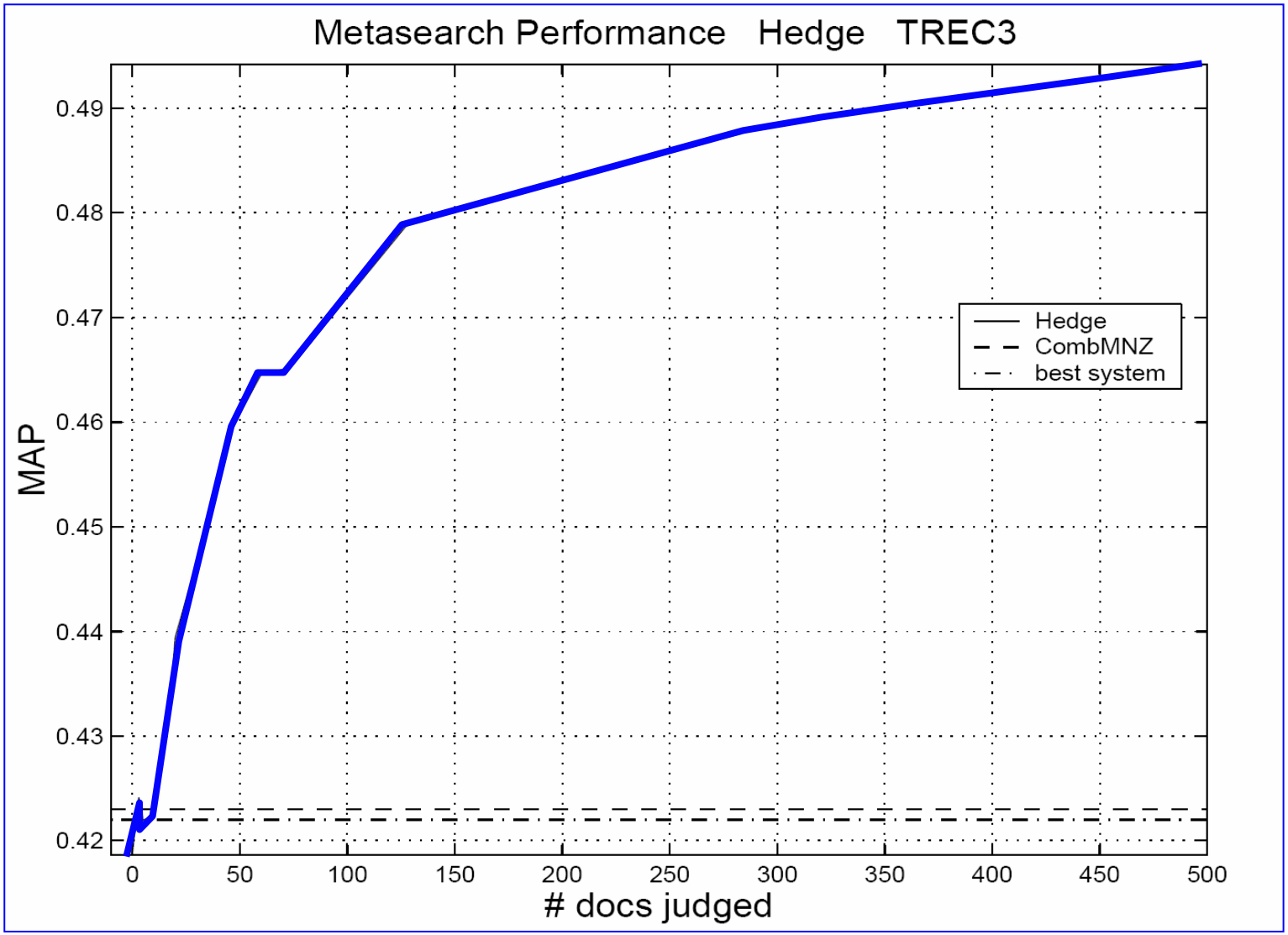
do "select the top 1000 doc" for metasearch

  - almost 1000 – docs already pooled are automatically in top of metasearch list

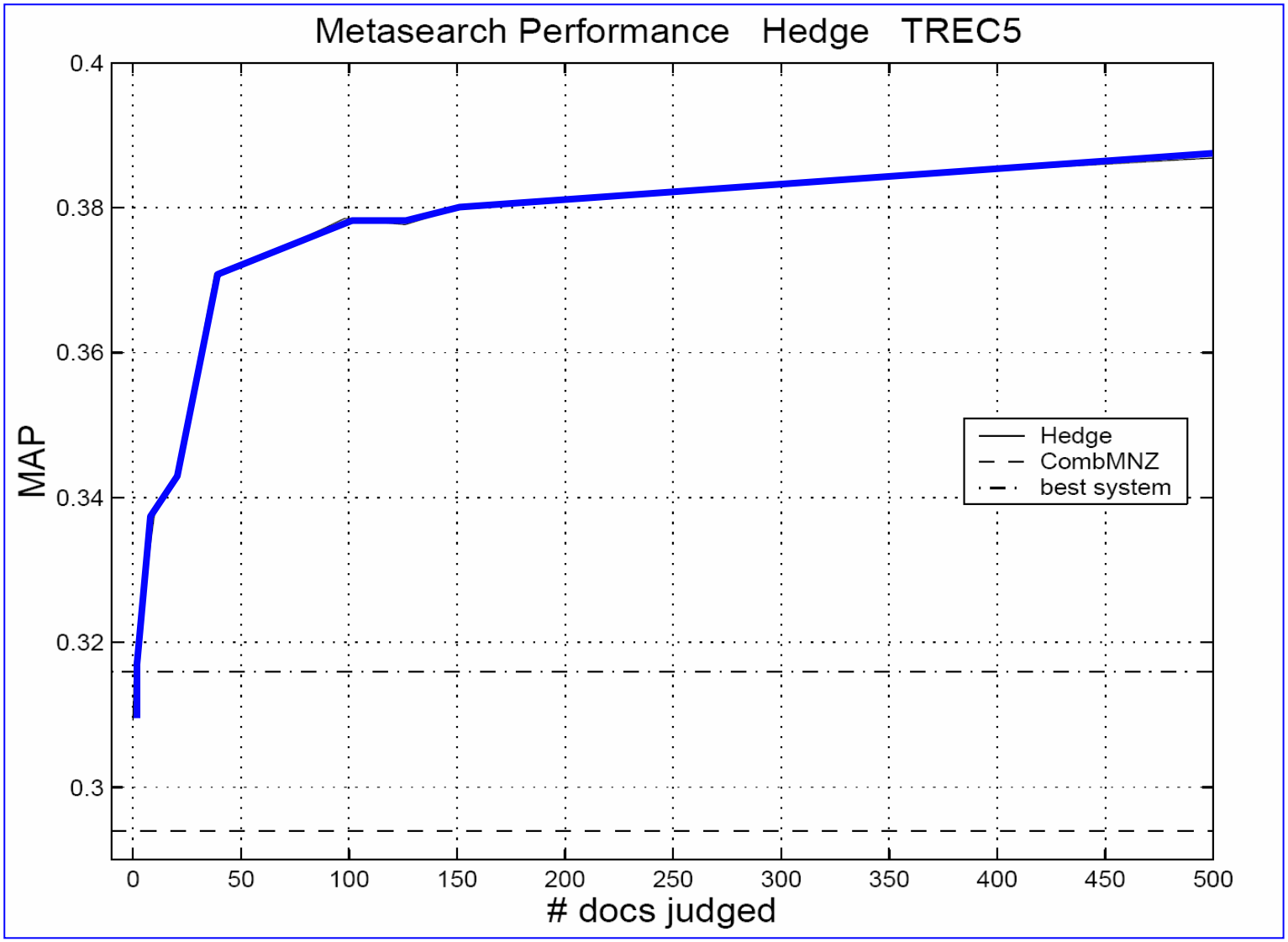
“total” precizion

- Average the precision at ALL ranks
  - Normalize so ideal system gets TP=1
- math is more simple

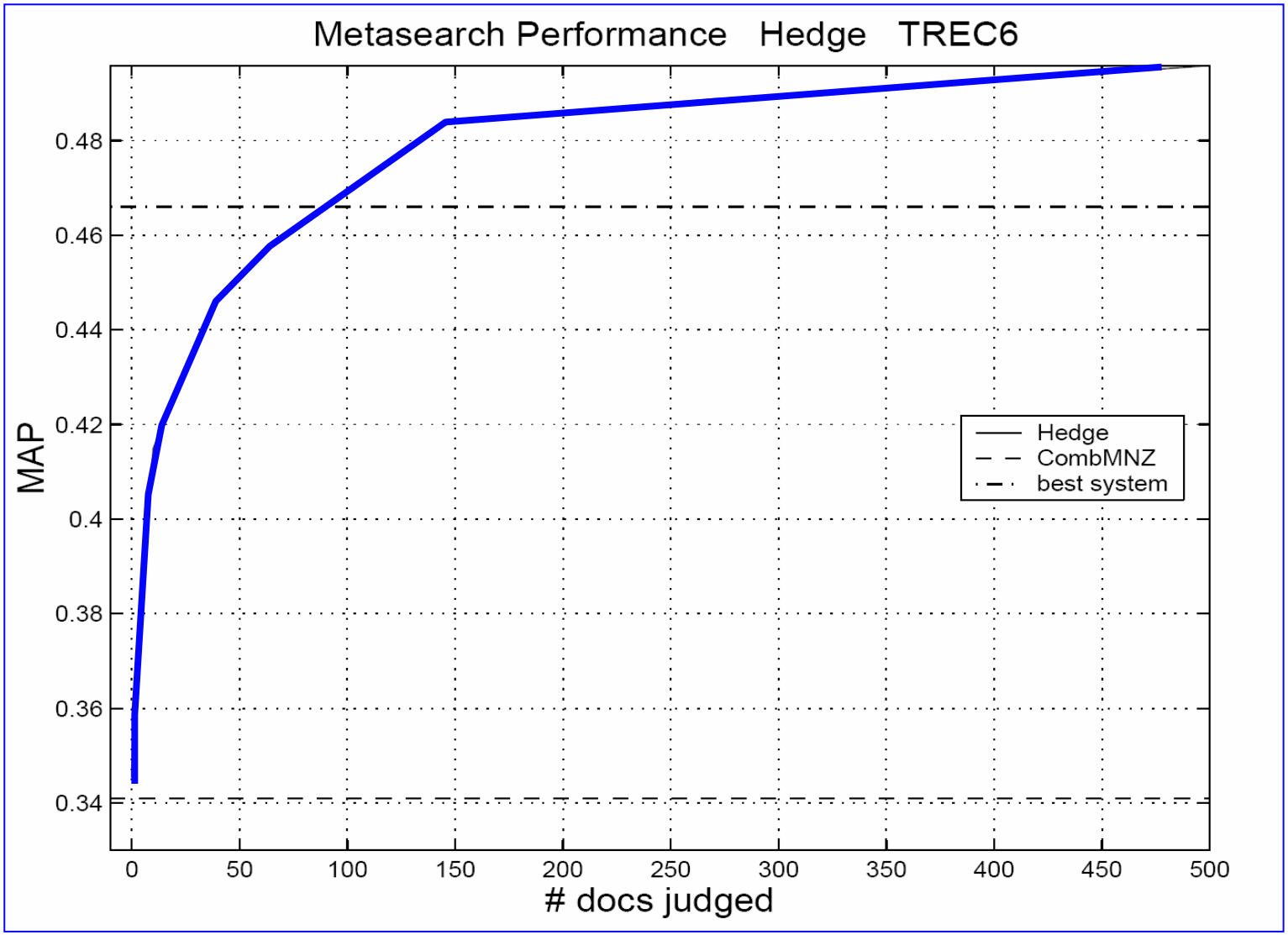
Metas – trec3


Metasearch Performance Hedge TREC3
MAP
0.49
0.48
0.47
0.46
0.45
0.44
0.43
0.42
0
50
100
150
200
250
300
350
400
450
500
docs judged
Hedge
CombMNZ
best system

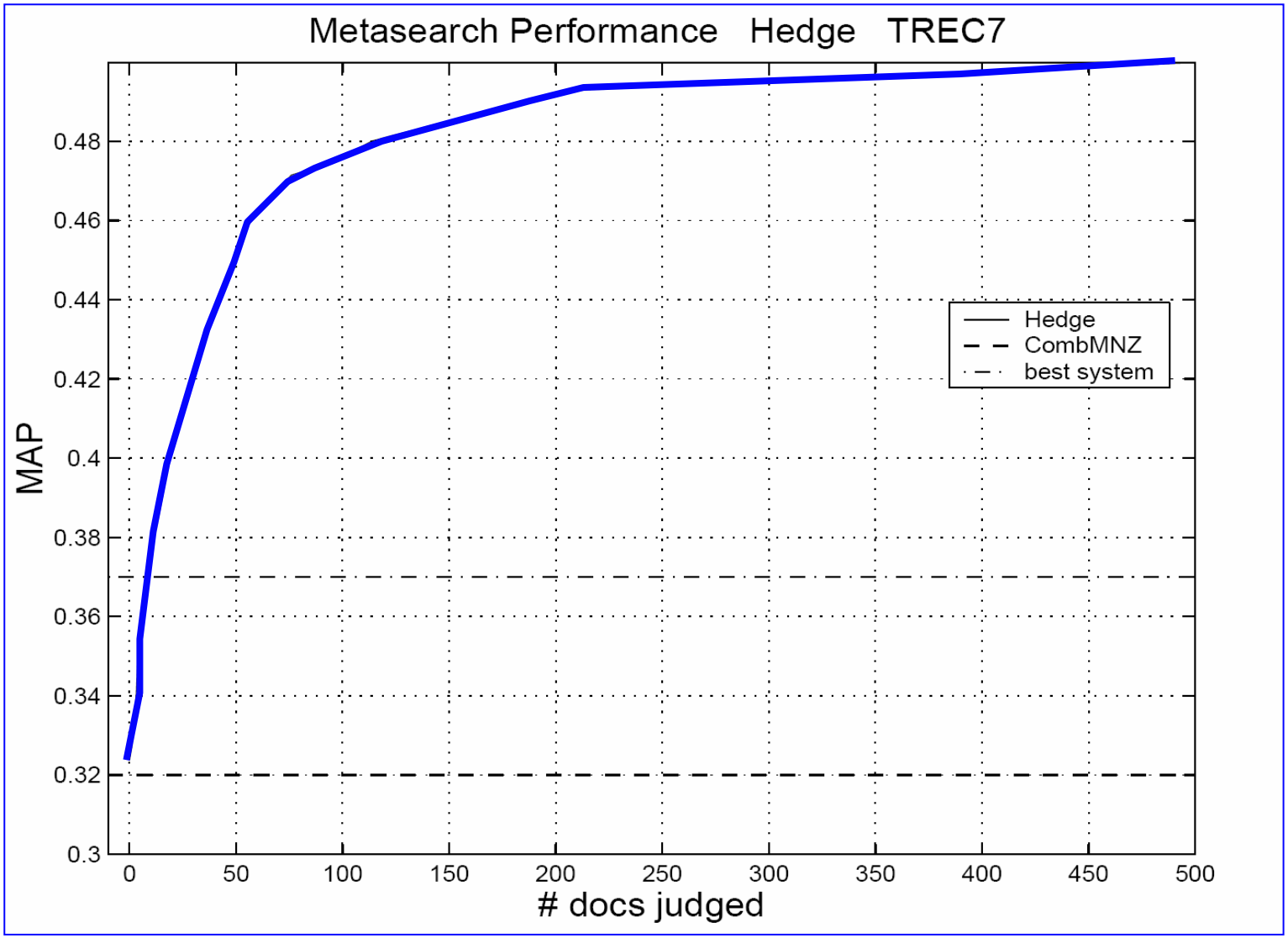
Metas – trec5


Metasearch Performance Hedge TREC5
MAP
0.4
0.38
0.36
0.34
0.32
0.3
0
50
100
150
200
250
300
350
400
450
500
docs judged
Hedge
CombMNZ
best system

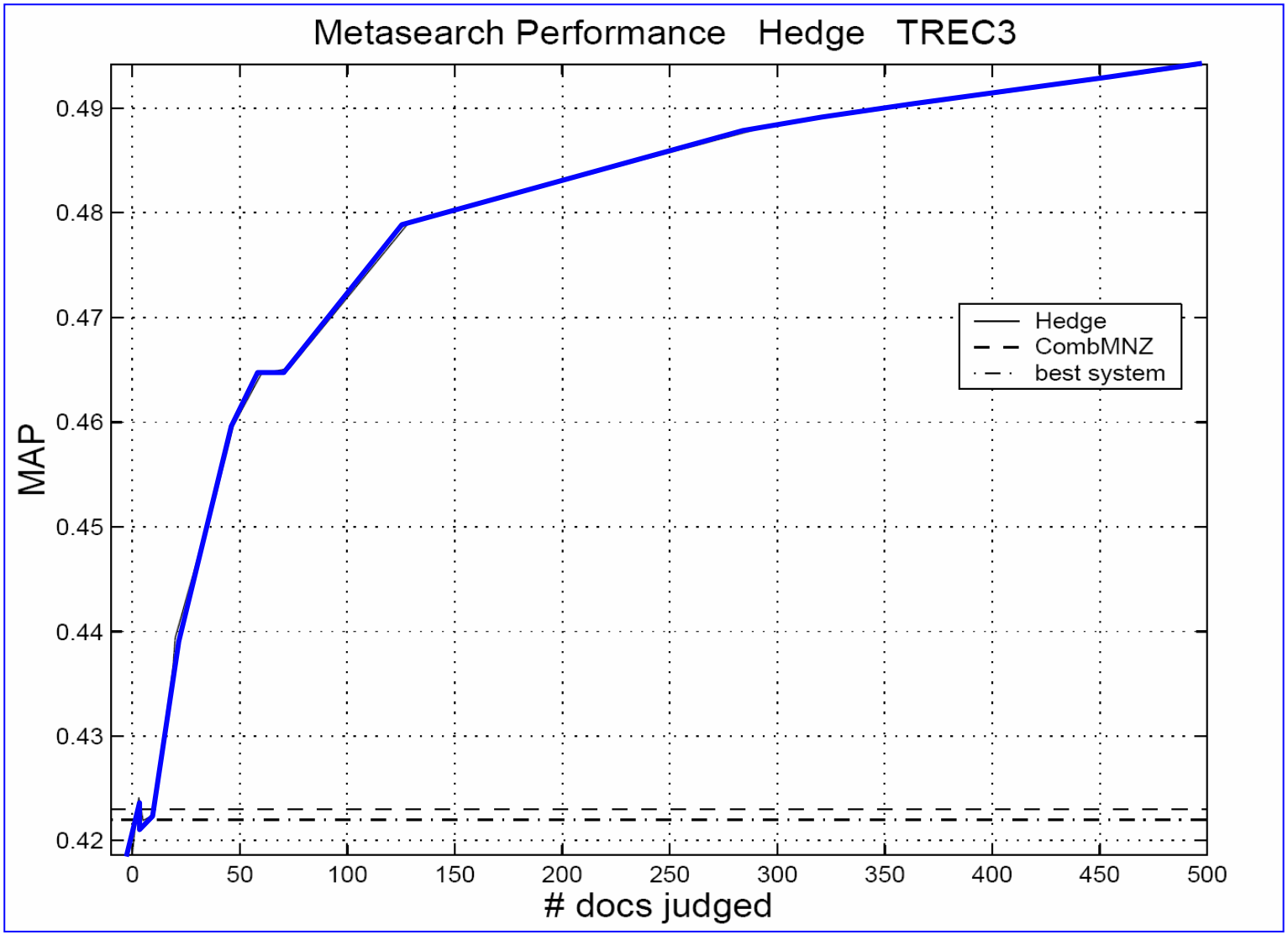
Metas – trec6


Metasearch Performance Hedge TREC6
MAP
0.48
0.46
0.44
0.42
0.4
0.38
0.36
0.34
0
50
100
150
200
250
300
350
400
450
500
docs judged
Hedge
CombMNZ
best system

Metas – trec7


Metasearch Performance Hedge TREC7
MAP
docs judged
0.48
0.46
0.44
0.42
0.4
0.38
0.36
0.34
0.32
0.3
0
50
100
150
200
250
300
350
400
450
500
Hedge
CombMNZ
best system

Metas – trec3


Metasearch Performance Hedge TREC3
MAP
docs judged
0.49
0.48
0.47
0.46
0.45
0.44
0.43
0.42
0
50
100
150
200
250
300
350
400
450
500
Hedge
CombMNZ
best system