Topics in Supervised Ranking

Cynthia Rudin
Assistant Professor, OR/Stats group, MIT Sloan
My work

1) Unconventional strategies for convergence of ML algorithms
2) Design of algorithms for ML
3) Probabilistic Generalization Bounds
4) Applications of ML / Knowledge discovery
My work

1) Unconventional strategies for convergence of ML algorithms
   - boosting and margins (RDS JMLR 04, RSD AnnStat 07)
   - AdaBoost and RankBoost equivalence (RS JMLR 09)

2) Design of algorithms for ML

3) Probabilistic Generalization Bounds

4) Applications of ML / Knowledge discovery
My work

1) Unconventional strategies for convergence of ML algorithms
2) Design of algorithms for ML
3) Probabilistic Generalization Bounds
4) Applications of ML / Knowledge discovery
My work

1) Unconventional strategies for convergence of ML algorithms

2) Design of algorithms for ML
   - “smooth margin” algorithms for classification and ranking (RS JMLR 09)
   - P-Norm Push (R JMLR 09)

3) Probabilistic Generalization Bounds
   - Supervised Ranking (RS JMLR 09, R JMLR 09)

4) Applications of ML / Knowledge discovery
My work

1) Unconventional strategies for convergence of ML algorithms
2) Design of algorithms for ML
3) Probabilistic Generalization Bounds
4) Applications of ML / Knowledge discovery
My work

1) Unconventional strategies for convergence of ML algorithms
2) Design of algorithms for ML
3) Probabilistic Generalization Bounds
4) Applications of ML / Knowledge discovery
   - Ranking Methods for Smart Grid Maintenance
     (RPRII Mach Learn 10)
May 2, 2008
Harvard Square, MA
Sample ECS Ticket

ticket: ME00010305
borough: M
house no: 135-6
cp: W
street: 4
arty: ST
act: EDSACB
time: 2000-01-12 10:31:00
cross st: MAC DOUGAL ST

ticket: ME00010305
lines: 31
Remarks: C.I.B POWERS REPORTS CONTRACTOR WORKING IN SB#158622 F/O 135 W.4 ST HAS 7W COPPERED AT W/DUCT ------------------------JM
01/12/00 13:20 MDEO'HARA DISPATCHED BY 44729
01/12/00 14:00 MDEO'HARAARRIVED BY 58101
01/12/00 14:25 OHARA REPORTS IN SB-158622 F/O 135 W. 4TH ST HAVE 7W MAIN IN TROUBLE GOING OUT THE WEST WALL.....UNABLE TO TELL IF IT GOES WEST OR SOUTH IN THE CROSSING AS PER THE M&S PLATE... FLUSH REQUIRED... ORDERED... PM
01/12/00 17:16 OHARA REPORTS FLUSH TRUCK STILL WORKING IN THE TROUBLE HOLE.............................. PM
01/12/00 20:10 OHARA REPORTS C/F/R 3-4/0DC,4-4/0AC,4"43' FROM SB-3622 F/O 135 W.4TH ST TO SB-3627 F/O 136 W.4TH ST. PARKING IS NORTH SIDE NO STANDING MON TO FRI 8AM TO 6PM SOUTH SIDE METER PARKING 8AM TO 9AM MON TO FRI
---------------------------------------------------------------------
IN SB-3627 F/O 136 W.4TH ST CLEARED B/O'S
IN SB-3622 F/O 135 W.4TH ST TIEING NUTRAL'S IN.........MB
01/12/00 20:40 OHARA REPORTS IN SB-3622 F/O 135 W.4TH ST HE CLAERED ALL B/O'S AND RETIED ALL NUTRALS............MB
01/12/00 20:40 MDEO'HARA COMPLETE BY 40075
******************************* E-00-12017-M *******************************
01/13/00 00:00 REFERRED TO: M.EC.UGBT ME1353 BY 07525
DOCS JOB: 019552984 TASK: 0195520988 AWU: 019552989
01/13/00 02:45 ORIGINAL LAYOUT # E00-12017-M WAS ENTERED IN ERROR THE CORRECT LAYOUT SHOULD READ AS FOLLOWED E00-12018-M & HAS BEEN CORRECTED ON THE ORIGINAL REFERAL TICKET........................................WNB
01/13/00 01:51 REFERRED TO: M.EC.CABBT EDSACB FYI BY 07525
01/13/00 01:51 REFERRED TO: M.EC.UGBT EDSACB FYI BY 07525
My work

1) Unconventional strategies for convergence of ML algorithms
2) Design of algorithms for ML
3) Probabilistic Generalization Bounds
4) Applications of ML / Knowledge discovery
   - Ranking Methods for Smart Grid Maintenance
     (RPRII Mach Learn 10)

The P-Norm Push was used to create ranked lists currently used for prioritization of repair work / inspections on the NYC power grid.
My work

1) Unconventional strategies for convergence of ML algorithms
   - boosting and margins *(RDS JMLR 04, RSD AnnStat 07)*
   - AdaBoost and RankBoost equivalence *(RS JMLR 09)*

2) Design of algorithms for ML

3) Probabilistic Generalization Bounds

4) Applications of ML / Knowledge discovery
Let’s say we’re ranking movies for Netflix:

\[ \{(x_i, y_i)\}_{i=1}^m \]  \hspace{1cm} (x_i, y_i) \in X \times \{-1, +1\} \quad \text{“Examples”}

movies

whether I liked it

\[ \{h_j\}_{j=1}^n \]  \hspace{1cm} h_j : X \rightarrow \{-1,1\} \quad \text{“Features”}

Is it an action movie?  
Is it a horror movie?  
Was it pre-1980?

Score for movie \( (x) \) = \[ f(x) = \sum_j \lambda_j h_j(x) \]
Two machine learning subfields:

Classification

Bipartite Ranking

Misclassification Error ($f$)

$$= \sum_{i} 1_{[y_i f(x_i) \leq 0]}$$
Two machine learning subfields:

Classification

Misclassification Error \((f)\)

\[
\frac{1}{n} \sum_{i} 1[y_i f(x_i) \leq 0]
\]

Bipartite Ranking

Misranking Error \((f)\)

\[
\frac{1}{n} \sum_{i} \sum_{k} 1[f(x_i) \leq f(x_k)]
\]

\{i: y_i = 1\} \{k: y_k = -1\}
Two machine learning subfields:

Classification

Misclassification Error ($f$)

$$= \sum_{i} 1[y_i f(x_i) \leq 0]$$

Bipartite Ranking

Misranking Error ($f$)

$$= \sum \sum 1[f(x_i) \leq f(x_k)]$$

\{i:y_i=1\} \{k:y_k=-1\}

Are the minimizers of the misclassification error the same as the minimizers of the misranking error?

In general, no.

In the case of well-known algorithms AdaBoost and RankBoost, yes.
Classification and Bipartite Ranking are *not* the same

\[ f > 0 \]
\[ f = 0 \]

Misclassification error = \(|Y_+| \) (huge) \[ Y_+ = \{i: y_i = +1\} \]
Misranking error = 0
Classification and Bipartite Ranking are *not* the same

Misclassification error = 1
Misranking error = 0
Classification and Bipartite Ranking are *not* the same.

Misclassification error = 1
Misranking error = $|Y_+|$

$Y_+ = \{i: y_i = +1\}$
AdaBoost

- Very powerful and popular supervised classification algorithm

Top 10 algorithms in data mining

Xindong Wu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang · Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu · Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg

Received: 9 July 2007 / Revised: 28 September 2007 / Accepted: 8 October 2007
Published online: 4 December 2007
© Springer-Verlag London Limited 2007

Abstract This paper presents the top 10 data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) in December 2006: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, KNN, Naive Bayes, and CART. These top 10 algorithms are among the most influential data mining algorithms in the research community. With each algorithm, we provide a description of the algorithm, discuss the impact of the algorithm, and review current and further research on the algorithm. These 10 algorithms cover classification,

(Data mining: the process of extracting patterns from data)
Statistical View of Boosting

(Breiman, Duffy, Helmbold, Friedman, Hastie, Tibshirani, Mason, Baxter, Bartlett, Frean, Ratsch Onoda, Muller)

Misclassification Error ($f$)  
\[ = \sum_{i} 1_{[y_i f(x_i) \leq 0]} \leq \sum_{i} e^{-y_i f(x_i)} \]

\[ = \sum_{i} e^{-y_i \sum_{j} \lambda_j h_j(x_i)} \]

\[ =: F^{Ada}(\lambda) \]
Statistical View of Boosting

(Breiman, Duffy, Helmbold, Friedman, Hastie, Tibshirani, Mason, Baxter, Bartlett, Frean, Ratsch Onoda, Muller)

Misclassification Error \( (f) \)

\[
\sum_i 1[y_i f(x_i) \leq 0] \leq \sum_i e^{-y_i f(x_i)}
\]

\[
= \sum_i e^{-y_i \sum_j \lambda_j h_j(x_i)} =: F^{\text{Ada}}(\lambda)
\]

AdaBoost iteratively computes \( \{\lambda_t\}_t \) such that:

\[
\lim_{t \to \infty} F^{\text{Ada}}(\lambda_t) = \inf_{\lambda} F^{\text{Ada}}(\lambda)
\]

AdaBoost objective minimized

Predicted class for new example \( x \) is \( \text{sign}(f(x)) \), where \( f(x) = \sum_j \lambda_j h_j(x) \)
RankBoost

- Designed by Freund, Iyer, Schapire, Singer in ’03 as a ranking version of AdaBoost.
- One of the most popular algorithms for information retrieval.
Statistical View of Boosting

Misranking Error \( f \) 
\[
= \sum_{i \in Y_+} \sum_{k \in Y_-} 1[f(x_i) \leq f(x_k)] \leq \sum_{i \in Y_+} \sum_{k \in Y_-} e^{-[f(x_i) - f(x_k)]}
\]
\[
= \sum_{i \in Y_+} \sum_{k \in Y_-} e^{-\sum_j \lambda_j [h_j(x_i) - h_j(x_k)]} =: F^{RB}(\lambda)
\]

RankBoost iteratively computes \( \{\lambda_t\}_t \) such that:
\[
\lim_{t \to \infty} F^{RB}(\lambda_t) = \inf_{\lambda} F^{RB}(\lambda) \quad \text{RankBoost objective minimized}
\]

Predicted rank for \( x \) determined by \( f(x) \), where \( f(x) = \sum_j \lambda_j h_j(x) \)
AdaBoost

- Very powerful and popular supervised classification algorithm

---

**An Empirical Comparison of Supervised Learning Algorithms**

Rich Caruana  
Alexandru Niculescu-Mizil  
Department of Computer Science, Cornell University, Ithaca, NY 14853 USA

---

With excellent performance on all eight metrics, calibrated boosted trees were the best learning algorithm overall. Random forests are close second, followed by uncalibrated bagged trees, calibrated SVMs, and uncalibrated neural nets. The models that performed poorest were naive bayes, logistic regression, decision trees, and boosted stumps. Although some methods clearly perform better or worse than other methods, there is significant variability across the problems and metrics. Even the best models sometimes perform poorly, and models with poor average

Metrics include misclassification error and misranking error
An Equivalence Between AdaBoost and RankBoost
Intuition
AdaBoost $\rightarrow$ RankBoost

$F^{Ada}(\lambda)$

$F^{RB}(\lambda)$

Misranking Error

$\lambda$
Intuition
AdaBoost $\Rightarrow$ RankBoost

$F^{Ada}(\lambda)$

$F^{RB}(\lambda)$

Main Theorem*

Misranking Error

* Under conditions
Intuition

AdaBoost ➔ RankBoost

$F^{Ada}(\lambda)$

$F^{RB}(\lambda)$

↓ Main Theorem*

Misranking Error

* Corollary 1 says the conditions are trivial.
Intuition
AdaBoost ➔ RankBoost

$F^{Ada}(\lambda)$

$F^{RB}(\lambda)$

Main Theorem

Theorem 2

Misranking Error
Intuition
RankBoost ➔ AdaBoost

$F^{Ada}(\lambda)$

$F^{RB}(\lambda)$

Misclassification Error
Intuition
RankBoost $\rightarrow$ AdaBoost

Misclassification Error

$F_{Ada}^{\lambda}$

$F_{RB}^{\lambda}$

Main Theorem*

* Under conditions
Intuition

RankBoost → AdaBoost

* Corollary 3 transforms RankBoost’s solutions into solutions for AdaBoost.
• AdaBoost produces a solution to the ranking problem that is just as good as RankBoost’s.

• RankBoost can be trivially altered to produce a solution to the classification problem that is just as good as AdaBoost’s.
Implications

• Coolness

• AdaBoost performs well with respect to eval metrics for classification and ranking.

• If you already programmed AdaBoost, you don’t need to program RankBoost.
Thanks

For details: