# Recurrent boosting for classification of natural and synthetic time-series data

Robert D. Vincent[1], Joelle Pineau[1],
Philip de Guzman[2] and Massimo Avoli[2]

[1]School of Computer Science, McGill University
[2]Montreal Neurological Institute

30 May 2007

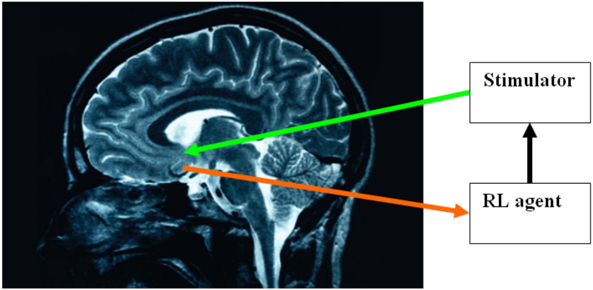
McGill

# Overview

- We describe Recurrent AdaBoost, a new method for time-series classification
- We present an empirical validation of the algorithm
- We provide a dataset for an interesting problem

# Goals of this research

- Electrical stimulation devices for the treatment of epilepsy
- An adaptive stimulation algorithm using reinforcement learning




- An automatic seizure detection algorithm will be part of the RL agent's state representation.

# Prior work in seizure detection

Automatic seizure detection is a longstanding problem in biomedical literature, relatively unstudied in the machine learning community:

- RBF networks (Schuyler et al. 2007)
- Wavelet analysis (Khan & Gotman 2003)
- Neural networks (Gabor at al. 1996)
- Energy methods (Zaveri et al. 1993)

# Why boosting?

We considered existing time series methods such as CRF (Lafferty et al. 2001), HMM (Rabiner, 1989)

Boosting offers two advantages:

- Relatively interpretable with simple learners
- Requires no prior model of the distribution of features

However, boosting is a general ensemble method that typically assumes i.i.d. training examples.

# Real AdaBoost.MH (Schapire & Singer 1999)

Multiclass form of AdaBoost (Freund & Schapire, 1997) giving real-valued predictions for $k$ classes, $\ell \in \mathcal{Y}$:

Given $m$ training examples $(x_i, Y_i), x_i \in \mathcal{X}, Y_i \subseteq \mathcal{Y}$
Initialize $D_1(i, \ell) = \frac{1}{mk}$
**for** $t = 1$ to $T$ **do**
  Train weak learner on distribution $D_t$
  Get new weak rule $h_t(x_i, \ell) \to \mathbb{R}$
  Increase $D_{t+1}(i, \ell)$ for all $i : \mathrm{sign}(h_t(x_i, \ell)) \neq Y_i(\ell)$
Return strong rule:

$$H(x, \ell) = \sum_{t=1}^{T} h_t(x, \ell)$$

# Recurrent AdaBoost.MH

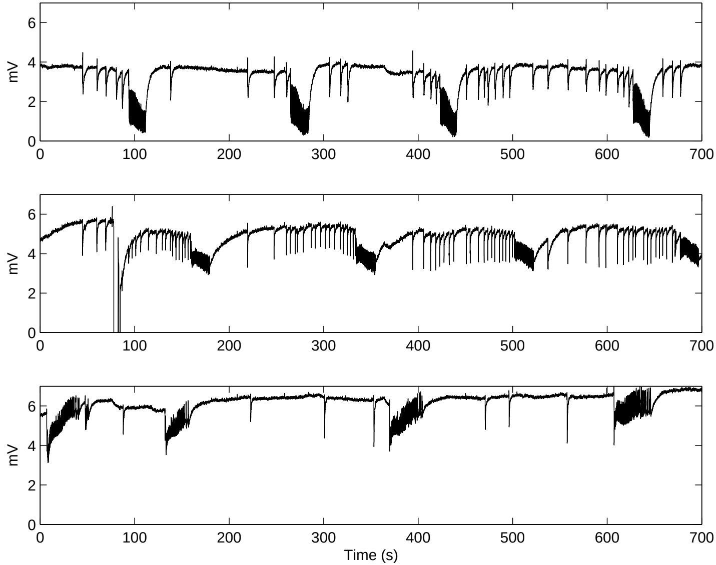
Adds both a set of class predictions $C_i$ and a vector of features $w_i$ containing the $C_i$ values for $n$ prior frames.

Given $m$ training examples $(\{x_i, w_i\}, Y_i), x_i \in \mathcal{X}, Y_i \subseteq \mathcal{Y}$
Initialize $D_1(i, \ell) = \frac{1}{mk}$, $C_i(\ell) = 0$, $w_i = 0$
**for** $t = 1$ to $T$ **do**
  Train weak learner on distribution $D_t$
  Get new weak rule $h_t(\{x_i, w_i\}, \ell) \to \mathbb{R}$
  Increase $D_{t+1}(i, \ell)$ for all $i : \mathrm{sign}(h_t(\{x_i, w_i\}, \ell)) \neq Y_i(\ell)$
  $C_i(\ell) = C_i(\ell) + h_t(\{x_i, w_i\}, \ell)$ for all $i = 1, ..., m$, $\ell \in \mathcal{Y}$
  $w_i = \{C_{i-n}(\ell) ... C_{i-1}(\ell)\}$ for all $i = 1, ..., m$, $\ell \in \mathcal{Y}$
Return strong rule:

$$H(\{x, w\}, \ell) = \sum_{t=1}^{T} h_t(\{x, w\}, \ell)$$

# Real data

Recordings of electrical activity from animal models of epilepsy



# Real data

- 4096-sample frames hand-labeled as either *normal* (82%), *spike* (3%), and *seizure* (14%).
- Windowed, Normalized, FFT
- Final vector of 83 real-valued features:
  - Mean: $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Range: $\max x_i - \min x_i$
  - Energy: $E = \sum_{i=1}^{n}(x_i - \mu)^2$
  - 80 FFT magnitudes

# Algorithms evaluated

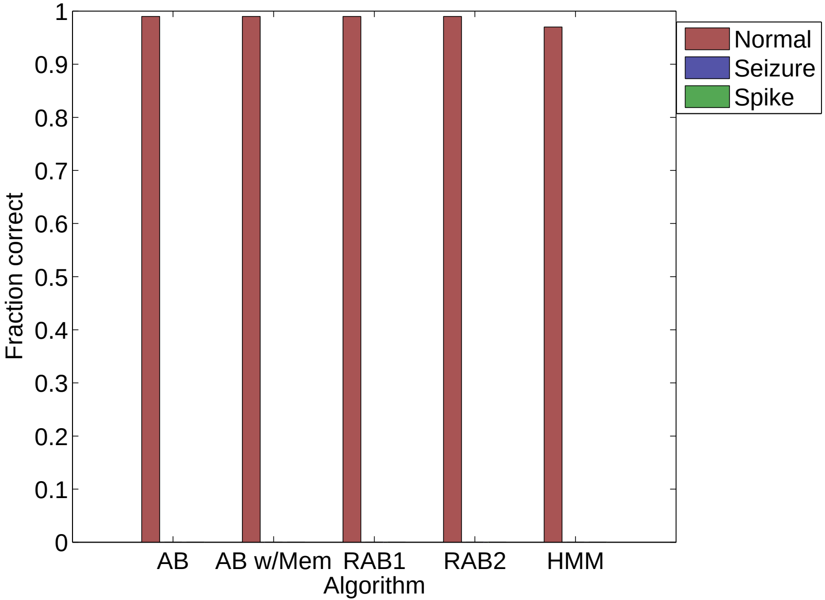
In *Recurrent AdaBoost*, the real-valued predictions for prior time frames are used as features of the current time frame.

In *AdaBoost with Memory*, the entire feature vectors of prior time frames are concatenated with the current time frame.

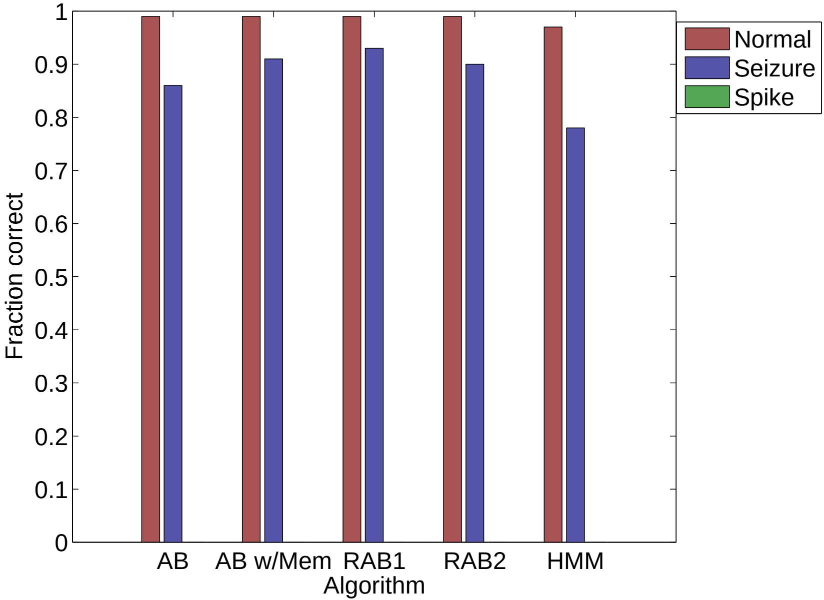
| | |
|---|---|
| $y_i = f(x_i)$ | Standard AdaBoost.MH |
| $y_i = f(x_i, y_{i-1} ... y_{i-n})$ | Recurrent AdaBoost.MH |
| $y_i = f(x_i, x_{i-1} ... x_{i-n})$ | AdaBoost.MH with Memory |

We also present results for a standard Hidden Markov Model with a single Gaussian per class.

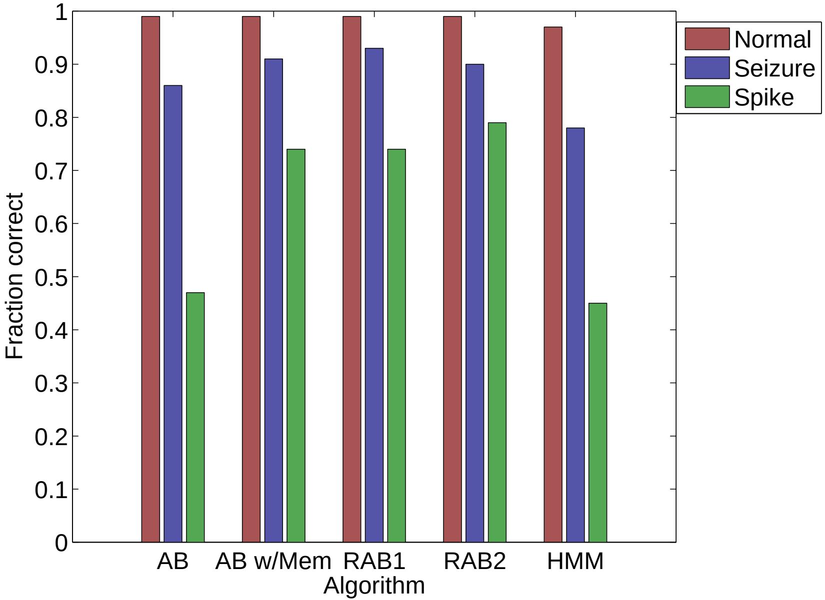
# Results with real data



# Results with real data

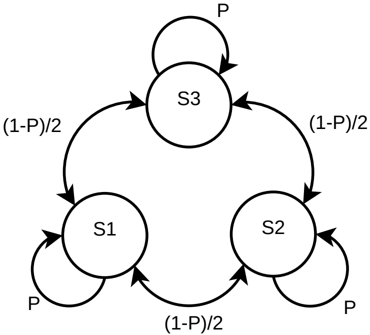


# Results with real data

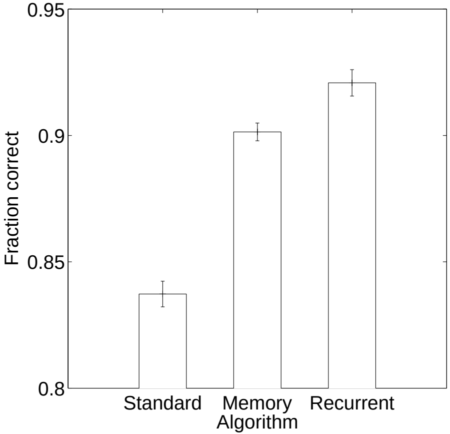


# Synthetic data

- Based on the "waveform" problem from the UCI ML repository (Breiman et al. 1984, Aha et al. 1991)
- 3 classes of 21 noisy continuous features with 86% Bayes-optimal accuracy
- Added Markov sequence with probability $P$ of *not* changing class

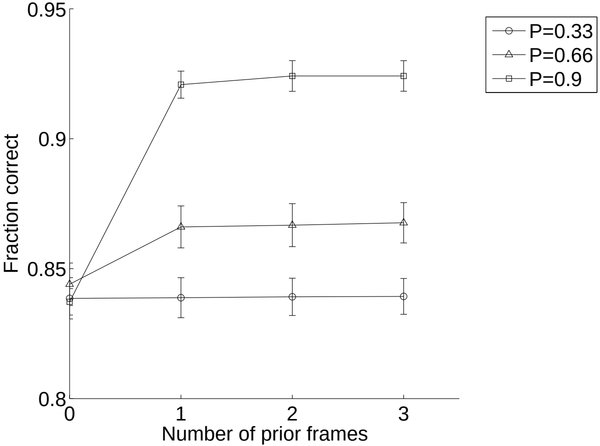


# Results with synthetic data




Results with the three boosting algorithms for $P = 0.9$

# Results with synthetic data




Recurrent AdaBoost with various values of $P$ and $n$

# Complexity

- AdaBoost.MH is $O(mk)$ per round (Schapire and Singer, 1999)
- Recurrent AdaBoost.MH is $O(mnk)$ per round
- The complexity of AdaBoost is dominated by the weak learner, typically $O(md)$ where $d = ||x_i||$.
- Both Recurrent AdaBoost and AdaBoost with Memory increase $d$, however, given $k \ll d$,

$$d_{RA} = d + nk < d_{AM} = d(n+1)$$

| | |
|---|---|
| $m$ | Number of training examples |
| $d$ | Dimensionality of training data |
| $n$ | Number of prior time frames incorporated |
| $k$ | Number of classes |

# Conclusion

- We have introduced Recurrent AdaBoost a new, potentially general method for time-series classification.
- Boosting methods can avoid strong prior assumptions about feature distributions
- Boosted classifiers can provide useful insights about the application

# Acknowledgments

For further information:

rl.cs.mcgill.ca

Labeled dataset:

http://www.cs.mcgill.ca/~jpineau/datasets/epilepsy.tar.gz

NSERC
CRSNG

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

Fonds de recherche
sur la nature
et les technologies
Québec