**Online learning**
Real-world tasks:

Many models: A, B, ... Best model typically changes with time:

A D G A G D G A

Need adaptive algorithms that can exploit repeats

**Mixing Past Posteriors [BW02] does the job**

But while we do understand the simpler algorithms

Bayes

Fixed Share

Fixed Share can be regarded as a Bayesian mixture of all the possible partition models. (Number is exponential in time, but FS collapses)

Fixed Share = Bayes update + mix in small amount of initial prior

the superior MPP remains a baffling mystery

\[ \hat{d}_{t+1}(m) = \frac{P(y_t|m)w_t(m)}{\sum_{m} P(y_t|m)w_t(m)} \]

where

\[ w_t(m) = \sum_{s=0}^{t-1} \hat{d}_s(m) \gamma_t(s) \]

MPP = Bayes update + mix in small amount of all past posteriors

Bayes for specialists crash course

A specialist may or may not issue a prediction [FSSW97]. Prediction \( P(y|m) \) only available for awake \( m \in W \).

Key insight: complete specialists to full models [CV09]:

\[ P(y|m) := P(y) \quad \text{for all asleep } m \notin W. \]

With prior \( P(m) \) on specialists, the Bayesian predictive distribution

\[ P(y) = \sum_{m \in W} P(y|m)P(m) + \sum_{m \notin W} P(y)P(m) \]

has solution

\[ P(y) = \frac{\sum_{m \in W} P(y|m)P(m)}{\sum_{m \in W} P(m)}. \]

The posterior distribution is incrementally updated by

\[ P(m|y) = \frac{P(y|m)P(m)}{P(y)} \]

if \( m \in W \),

\[ P(m|y) = P(m) \]

if \( m \notin W \).

Bayes is fast: predict in \( O(M) \) time per round.

Bayes is good: regret w.r.t. specialist \( m \) on data \( y_{\leq T} \) bounded by

\[ \sum_{t \leq T} \left( - \ln P(y_t|y_{<t}) + \ln P(y_t|y_{<t}, m) \right) \leq - \ln P(m). \]

**Our breakthrough**

Bayesian interpretation for MPP

We interpret Mixing Past Posteriors as a Bayesian mixture of partition specialists which can be asleep:

MPP

We craft a prior on all partition specialists for which Bayes is both

- fast: collapses to \( O(M) \) time per trial, \( O(M) \) space
- good: regret close to information-theoretic lower bound

Conclusion

Proper Bayesian interpretation of Mixing Past Posteriors using "prediction with specialists"

- Simplified tuning
- Fastest algorithm
- Sharpest bounds
- Mysterious factor 2 in bound explained

Application of specialists technology to multitask learning

- significantly improved bounds
- intriguing collapsed algorithm

all grandiose details are in the paper

Thank you!