

---

# Hedging Structured Concepts

Wouter M. Koolen  
CWI Amsterdam

Manfred K. Warmuth  
UC Santa Cruz

Jyrki Kivinen  
University of Helsinki

Sunday 27 June, 2010

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Conclusion

- Prediction With Expert Advice

Hedge algorithm

- Structured Concepts

SC + Hedge  $\Rightarrow$  range factor problem

- Component Hedge

SC + CH  $\Rightarrow$  range factor problem solved

- Conclusion

# Prediction with Expert Advice

Prediction With  
Expert Advice

▷ Prediction with  
Expert Advice

The Hedge  
Algorithm

Structured Concepts

Component Hedge

Conclusion

## □ Setting

- Several sources of predictions (experts)
- Choose an expert each trial (randomised)
- Incur loss of the selected expert (0/1)
- Observe loss of all experts (full information)

## □ Goal

- Cumulative loss close to the best expert
- Efficient algorithm

# The Hedge Algorithm (Freund & Schapire 1997)

Prediction With  
Expert Advice

Prediction with  
Expert Advice

▷ The Hedge  
Algorithm

Structured Concepts

Component Hedge

Conclusion

- Maintains uncertainty as a distribution  $w_t$  on  $n$  experts  
 $w_1$  is uniform
- For each trial  $t = 1, 2, \dots$ 
  - Select expert  $i$  with probability  $w_{t,i}$
  - Receive loss vector  $\ell_t \in [0, 1]^n$ , incur loss  $\ell_{t,i}$
  - Expected loss  $w_t \cdot \ell_t$
  - Update  $w_{t+1,i} \propto w_{t,i} \beta^{\ell_{t,i}}$
- With  $\ell^H = \sum_{t=1}^T w_t \cdot \ell_t$  and  $\ell^* = \min_i \sum_{t=1}^T \ell_{t,i}$ ,

$$\ell^H - \ell^* \leq \sqrt{2\ell^* \ln n} + \ln n$$

# Structured Concepts

Prediction With  
Expert Advice

Structured Concepts

▷ Structured  
Concepts

Prediction with  
Structured Concepts  
Expanded Hedge

Component Hedge

Conclusion

- Concepts composed of components

concept	component
set	element
permutation	assignment
bipartite matchings	edges
spanning trees	edges
paths	edges

# Prediction with Structured Concepts

Prediction With  
Expert Advice

Structured Concepts

Structured Concepts  
Prediction with  
Structured

▷ Concepts

Expanded Hedge

Component Hedge

Conclusion

- Goal: on-line prediction with “combinatorial experts”
  - Route planning: shortest path
  - Media multicasting: directed spanning trees
- Loss of concept is sum of losses of its components
- Helps: losses of concepts highly related
- Hurts: combinatorial explosion (many concepts)

# Expanded Hedge (EH)

Prediction With  
Expert Advice

Structured Concepts

Structured Concepts  
Prediction with  
Structured Concepts

▷ Expanded Hedge

Component Hedge

Conclusion

- Treat each structured concept as an expert
- Run Hedge algorithm
- Consider size  $k$  subsets of  $n$  elements
  - Component loss in  $[0, 1]$ , so concept loss in  $[0, k]$ .
  - Number of concepts  $\binom{n}{k} \approx n^k$ .
  - Regret bound

$$\ell^{\text{EH}} - \ell^* \leq \sqrt{2\ell^* k k \ln n} + k k \ln n$$

- But lower bound has  $k \ln n$ . **Range factor problem**

# Usages

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

▷ Usages

Usages Example

Expanded Hedge

Component Hedge

Component Hedge II

Implementation

Lower Bounds

Conclusion

- Identify concepts with incidence vectors
- Loss of  $C$  is  $C \cdot \ell$  (with  $\ell$  component losses)
- Randomly select a concept  $C$  with probability  $W_C$
- Expected loss is

$$\sum_C W_C (C \cdot \ell) = \underbrace{\left( \sum_C W_C C \right)}_{\text{usage of } W} \cdot \ell$$

- Only the *usage* (i.e. mean concept) matters
- Set of usages is the convex hull of concepts



# Usages Example

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Usages

▷ Usages Example

Expanded Hedge

Component Hedge

Component Hedge II

Implementation

Lower Bounds

Conclusion

- Sets of 2 out of 4 elements

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

- The usage of the distribution  $(.3, .3, .2, .1, .1, 0)$  on sets

$$.3 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + .3 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + .2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + .1 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + .1 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} .8 \\ .5 \\ .4 \\ .3 \end{pmatrix}$$

# Expanded Hedge (EH)

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Usages

Usages Example

▷ Expanded Hedge

Component Hedge

Component Hedge II

Implementation

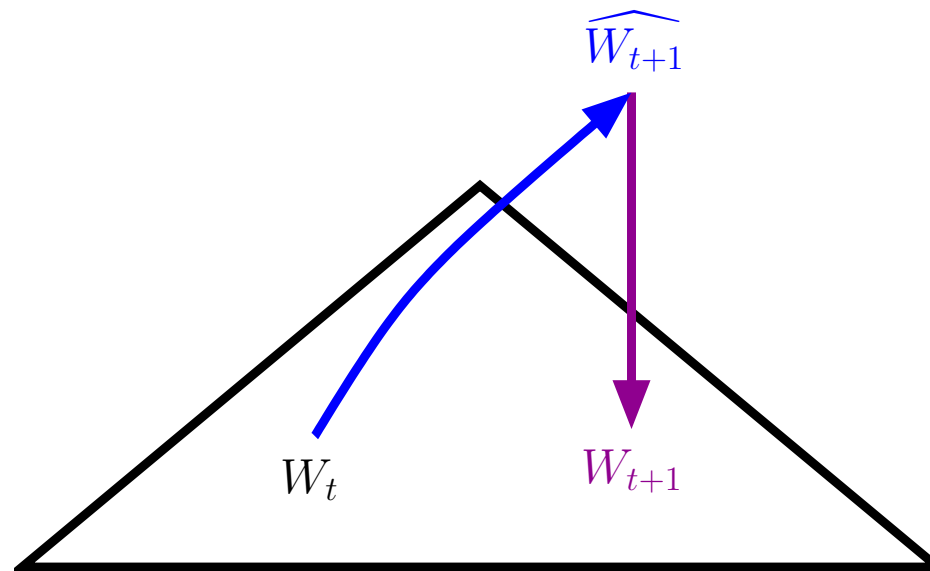
Lower Bounds

Conclusion

## Two-step EH update

$$\widehat{W}_{t+1} = \operatorname{argmin}_W \Delta(W \| W_t) + \sum_C W_C (C \cdot \ell_t)$$

$$W_{t+1} = \operatorname{argmin}_{W \text{ a p.d.}} \Delta(W \| \widehat{W}_{t+1})$$



$$\Delta(x \| y) = \sum_i x_i \ln \frac{x_i}{y_i} - x_i + y_i$$

# Component Hedge(CH)

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Usages

Usages Example

Expanded Hedge

Component  
▷ Hedge

Component Hedge II

Implementation

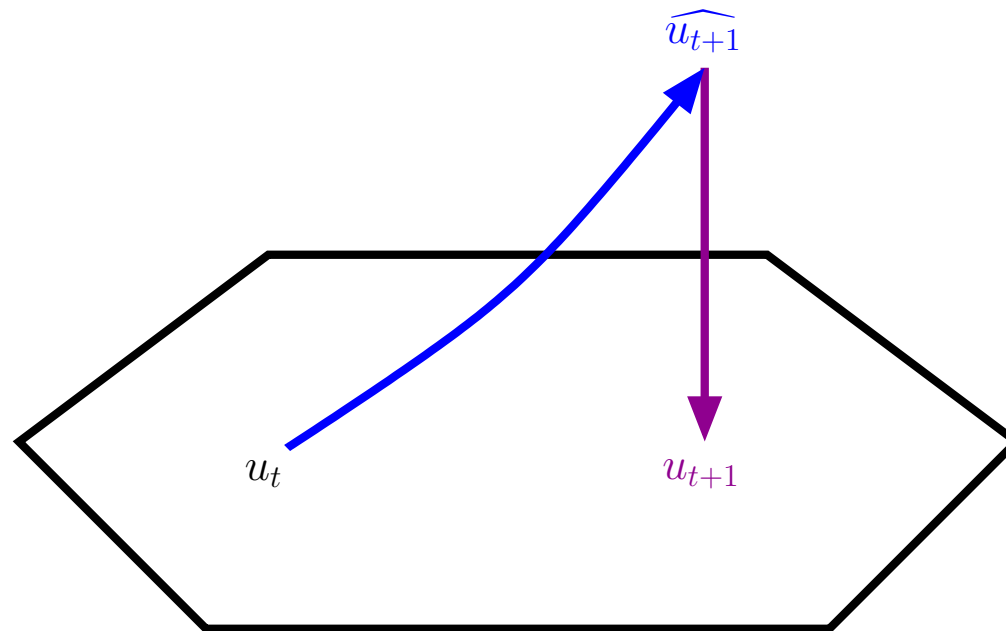
Lower Bounds

Conclusion

Idea: do the same trick on the level of usages

$$\widehat{u}_{t+1} = \operatorname{argmin}_u \Delta(u \| u_t) + u \cdot \ell_t$$

$$u_{t+1} = \operatorname{argmin}_{u \text{ a usage}} \Delta(u \| \widehat{u}_{t+1})$$



# Component Hedge II

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Usages

Usages Example

Expanded Hedge

Component Hedge

▷ Component  
Hedge II

Implementation

Lower Bounds

Conclusion

□ Let  $u_1$  be the usage of the uniform distribution

□ For trial  $t = 1, 2, \dots$

– Decompose  $u_t = \sum_i \alpha_i C_i$

– Sample  $C_i$  with probability  $\alpha_i$

– Expected loss  $u_t \cdot \ell_t$

– Update and relative entropy projection

□ Regret has no range factor. E.g. for  $k$ -of- $n$  sets

$$\ell^{\text{CH}} - \ell^* \leq \sqrt{2\ell^* k \ln n} + k \ln n$$

# Implementation

Prediction With  
Expert Advice

Structured Concepts

Component Hedge  
Usages

Usages Example

Expanded Hedge

Component Hedge

Component Hedge II

▷ Implementation

Lower Bounds

Conclusion

- Usage vectors  $u_t$  are small
- No closed form for relative entropy projection

$$u_{t+1} = \underset{u \text{ a usage}}{\operatorname{argmin}} \Delta(u \| \widehat{u}_{t+1})$$

- The usage polytope is the convex hull of exponentially many concepts. Fortunately, it can often be represented by polynomially many linear inequalities. E.g. Birkhoff and flow polytope.
- Idea: iteratively reestablish most violated constraint
- Known as Sinkhorn balancing for permutations

# Lower Bounds

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Usages

Usages Example

Expanded Hedge

Component Hedge

Component Hedge II

Implementation

▷ Lower Bounds

Conclusion

- CH is optimal: we have matching lower bounds for sets, permutations, bipartite matchings, spanning trees and paths.
- In each case, reduction from the basic expert case.

Prediction With  
Expert Advice

Structured Concepts

Component Hedge

Conclusion

▷ Philosophy

## □ Uncertainty

- EH: Probability distribution on concepts
- CH: Convex combination of concepts

## □ Relative entropy regularisation seems universal

- Possible to incorporate constraints into divergence
- But RE works in all cases