Approximately Optimal Search Algorithms from Data

Evangelia Gergatsouli
University of Wisconsin - Madison

Joint work w/
Shuchi Chawla, Yifeng Teng, Christos Tzamos, Ruimin Zhang

Pandora's box problem

- **n** boxes each containing a value $\mathbf{v_i}$
- values are drawn from a known distribution
- pay $\mathbf{c_i}$ to find exact value of box **i** (*probing cost*)
- How to search to find a single good value?

Pandora's box problem

Pay:	1	4	2	20	10
Receive:	??	10	1	??	25

- Objective → 2 Versions:
  - *Maximization*: max value - probing cost: 25-(4+2+10)
  - *Minimization*: min value + probing cost: 1+(4+2+10)

Our focus today: mainly the minimization objective

Optimal Strategy

[Weitzman'1979] When values are drawn independently, a simple greedy strategy is optimal:

- Assign an index to every box
- Search boxes in order of index until current value is better than the index of the next box

Boxes are independent, what about **correlation**?

Can we approximate the optimal?

- Really hard problem:
  - location of best box may be encoded in values of other boxes

Example: values 4 and 2 means go to box 42 to find best value
Even worse: Information of best box might be arbitrarily encrypted

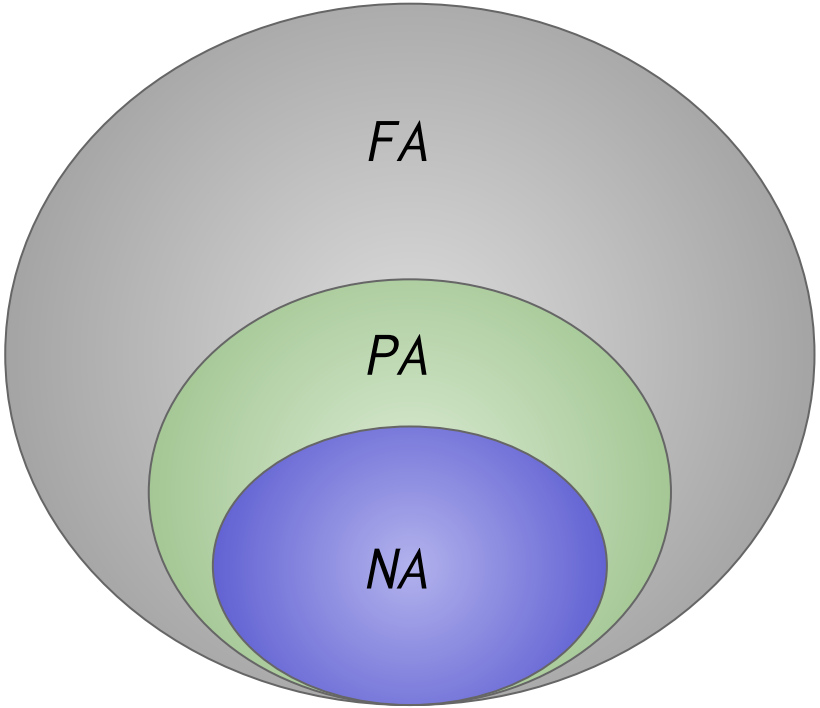
Restricted Strategy types

Fully Adaptive (FA): decision based on any previous step

Partially Adaptive (PA): fix order of boxes + stopping rule

Note: PA strategies are optimal for independent values

Non Adaptive (NA): decide on set of boxes from the start



What can we approximate?

- Approximate FA? **Hard!**
- Approximate NA? **Hard!** **Similar to Set Cover**

Main Positive Result:
Can approximate the optimal partially adaptive (PA) strategy within a constant factor!

A weaker goal:

Approximate NA by any FA

Lower Bounds:

- Minimization
  - NP-hard to approximate NA better than a constant
  - Our PA algorithm is asymptotically optimal
- Maximization
  - NP-hard to approximate NA within any constant

Extensions

- Selecting **k** boxes
  - constant factor approximation to optimal PA
- Selecting a basis of a matroid of rank **k**
  - **O(log k)** approximation to optimal PA
  - No **o(log k)** algorithm even for NA vs FA
- Learning optimal algorithm from samples
  - No need to know distribution of values

Extensions

Thank you!

- Selecting **k** boxes
  - constant factor approximation to optimal PA
- Selecting a basis of a matroid of rank **k**
  - **O(log k)** approximation to optimal PA
  - No **o(log k)** algorithm even for NA vs FA
- Learning optimal algorithm from samples
  - No need to know distribution of values