Optimal Index Rules for Single Resource Allocation to Stochastic Dynamic Competitors

Peter Jacko*

Valuetools
May 19, 2011

*Basque Center for Applied Mathematics (BCAM), Bilbao, Spain

Known Optimal Index Rules

- Job sequencing problem
  - ▷ $c\mu$-rule: Cox & Smith (1961)
- Multi-armed bandit problem
  - ▷ Gittins index rule: Gittins & Jones (1974)
- Klimov network: Klimov (1974)
- Tax problem: Varaiya, Walrand & Buyukkoc (1985)
- Object search and detection: Bertsekas (2001)

Motivation

- Can we cover the five problems by a unified model?
  - ▷ frozen-if-not-allocated assumption
- Can we optimally solve by index rules problems with restless competitors using the same approach?
  - ▷ much more general model
  - ▷ Whittle (1988): It may be too much to expect the index policy to be optimal in the restless case
  - ▷ asymptotic optimality with ∞ resources
  - ▷ in applications we often have single resource
    - – no (asymptotic) optimality results

Outline

- Unified MDP formulation
- Relaxations and decomposition into subproblems
- Optimal solution to subproblems
  - ▷ frozen-if-not-allocated assumption
  - ▷ reward normalization and the Gittins index
  - ▷ optimal solution to relaxations
  - ▷ optimal policy to original problem
    - – as a sequence of solutions to relaxed problems
- New sufficient conditions for restless case

Markov Decision Process Framework

- Decision time epochs $t = 0, 1, 2, \ldots$
- K independent competitors
- Two possible control actions for each $k \in \mathcal{K}$:
  - ▷ allocate the resource (action 1)
  - ▷ do not allocate the resource (action 0)
- Competing for a resource that is allocated to a single competitor at a time
- Decisions captured by action processes $a_k(t)$ for each k

Model of Competitor

- Each competitor k is defined by
  - ▷ finite state space $\mathcal{N}_k$
  - ▷ capacity consumption (work) $W_{k,n}^a$
  - ▷ reward $R_{k,n}^a$
  - ▷ transition probability matrix $\boldsymbol{P}_k^a$
- Dynamics captured by state process $X_k(t) \in \mathcal{N}_k$
- Key assumptions (to be used)
  - ▷ [binary work] $W_{k,n}^a = a$
  - ▷ [frozen if not allocated] $\boldsymbol{P}_k^0 = \boldsymbol{I}$ (identity matrix)

Unified Optimization Criterion

- β-average quantity: for $0 \le \beta \le 1$,

$$\mathbb{B}_\tau^\pi \left[Q_{X(\cdot)}^{a(\cdot)}, \beta \right] := \lim_{T \to \infty} \frac{\displaystyle\sum_{t=\tau}^{T-1} \beta^{t-\tau} \, \mathbb{E}_\tau^\pi \left[Q_{X(t)}^{a(t)} \right]}{\displaystyle\sum_{t=\tau}^{T-1} \beta^{t-\tau}}$$

- Recovers traditional measures
  - ▷ time-average criterion, when $\beta = 1$
  - ▷ myopic criterion, when $\beta = 0$
  - ▷ (scaled) total β-discounted criterion, otherwise

Optimization Problem

- Maximizing β-average reward

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} R_{k,X_k(\cdot)}^{a_k(\cdot)}\right] \tag{P}$$

$$\text{subject to } \mathbb{E}_t^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} a_k(t)\right] = 1, \text{ for all } t\in\mathcal{T}$$

- Where
  - ▷ $\mathcal{I}_{\boldsymbol{X},\boldsymbol{a}}(t) := \{\boldsymbol{X}(0), \boldsymbol{a}(0), \ldots, \boldsymbol{X}(t-1), \boldsymbol{a}(t-1), \boldsymbol{X}(t)\}$
  - ▷ $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ is the space of policies adapted to $(\mathcal{I}_{\boldsymbol{X},\boldsymbol{a}}(t))_{t\in\mathcal{T}}$
  - ▷ $\mathbb{E}_t^{\boldsymbol{\pi}}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\boldsymbol{\pi}, \mathcal{I}_{\boldsymbol{X},\boldsymbol{a}}(t)]$

Two-Step Relaxation

- 1: Whittle's Relaxation:
  - ▷ allocate 1 competitor on β-average

$$\mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} W_{k,X_k(\cdot)}^{a_k(\cdot)}\right] = \mathbb{B}_0^{\boldsymbol{\pi}}\left[1\right]$$

  - ▷ instead of infinite number of sample-path constraints

- 2: Dualize this constraint using Lagrangian multiplier ν

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}}\left(R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)}\right)\right] + \nu$$

  - ▷ solvable unconstrained problem!

Properties of Relaxations

- Whittle relaxation gives an upper bound on (P)
- Lagrangian relaxation for every ν gives an upper bound on the Whittle relaxation and on (P)
- Key property: If an optimal solution to a relaxation is feasible for (P), then it is optimal for (P)

Decomposition

- Joint policy $\boldsymbol{\pi} \in \Pi_{\boldsymbol{X},\boldsymbol{a}}$ defines single-competitor policies $\widetilde{\pi}_k$ for all $k \in \mathcal{K}$

- $\widetilde{\pi}_k$ depends on $\boldsymbol{X}(\cdot)$ but decides only $a_k(\cdot)$, $\widetilde{\pi}_k \in \Pi_{\boldsymbol{X},a_k}$

- Let us study single-competitor parametric subproblems

$$\max_{\widetilde{\pi}_k \in \Pi_{\boldsymbol{X},a_k}} \mathbb{B}_0^{\widetilde{\pi}_k} \left[R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)} \right]$$

- This is an MDP
  - ▷ with net reward $Q_{k,X_k(\cdot)}^{a_k(\cdot)} := R_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu W_{k,X_k(\cdot)}^{a_k(\cdot)}$
  - ▷ there is an optimal policy only depending on $X_k(\cdot)$

Normalization

- From now on, frozen-if-not-allocated assumption
- It is enough to solve the normalized problem:
  - ▷ normalizing reward vectors (under $\beta \neq 1$)

$$\widehat{\boldsymbol{R}}_k^1 := \boldsymbol{R}_k^1 - \frac{(\boldsymbol{I} - \beta \boldsymbol{P}_k^1)\boldsymbol{R}_k^0}{1-\beta}, \qquad \widehat{\boldsymbol{R}}_k^0 := \boldsymbol{0}$$

  - ▷ $\boldsymbol{W}_k^0 = \boldsymbol{0}$ already due to the binary-work assumption
  - ▷ proof by analyzing the Bellman equation
- Then it is the one-armed bandit problem
  - ▷ solved by the Gittins index

Gittins Index

- There exist break-even values $\nu_{k,n}$ of ν, called Gittins index values, such that at state n
  - ▷ it is optimal to use resource if $\nu_{k,n} \geq \nu$
  - ▷ it is optimal not to use resource if $\nu_{k,n} \leq \nu$
- $\nu_{k,n}$ measures the maximal achievable average reward if using the resource
- it is the value of an optimal stopping problem
  - ▷ once competitor does not use the resource, it continues not using it forever

Optimal Solutions to Relaxations

- For Lagrangian relaxation: For each competitor,
  - ▷ it is optimal to use the resource if $\nu_{k,n} \geq \nu$
  - ▷ it is optimal not to use the resource if $\nu_{k,n} \leq \nu$
- Policy "use the resource iff $\nu_{k,n} > \nu$" results in using a non-increasing number of resource units
- For Lagrangian relaxation with $\nu_0 := \max\{\nu_{k,X_k(0)}\}$, it is optimal
  - ▷ to allocate a single competitor for a positive number of periods (while its index value is $\geq \nu_0$)
  - ▷ and not to use the resource afterwards

Sequence of Parametric Problems

- Consider a sequence of parametric problems
  - ▷ each problem is of the form of the Lagrangian relaxation defined earlier
  - ▷ the problems only differ by the initial state
  - ▷ each problem's parameter is defined by the highest index value (given by the initial state)
- They create a finite graph with problems as nodes
  - ▷ when it becomes optimal at the current node not to use the resource, move to the node with such initial state

Optimal Solution to Original Problem

- Frozen-if-not-allocated assump. implies: graph is a tree!
- Bandit evolution realizations give rise to a sequence of realized problems
  - ▷ with nonincreasing parameter
  - ▷ therefore, the sequence is finite $\left(\leq \sum_{k \in \mathcal{K}} |\mathcal{N}_k| \right)$
- We apply an optimal policy in each realized problem
- Merging the realized problems gives the original problem
- This merged solution is optimal and allocates exactly one competitor at every time epoch

Properties of Optimal Solution

- We recover the Gittins index rule:
  - ▷ allocate the resource at every period to the competitor of highest current Gittins index value
- It is optimal to stay on a winner
- Optimal learning by doing model:
  - ▷ after a finite number of periods (exploration phase), the resource is allocated to the same competitor forever (exploitation phase)
  - ▷ if all competitors are irreducible, then that competitor is the one whose smallest Gittins index value is largest

General Competitors

- Frozen-if-not-allocated assumption dropped: restless
- Reward normalization also possible (more general)
- Competitor is indexable, iff there exist unique values $-\infty \leq \nu_{k,n} \leq +\infty$, called (Whittle) index values, such that at state n
  - ▷ it is optimal to use resource if $\nu_{k,n} \geq \nu$
  - ▷ it is optimal not to use resource if $\nu_{k,n} \leq \nu$
- Existence not guaranteed!
- Not an optimal stopping problem (unlike Gittins index)

General Competitors

- If all the competitors are indexable, the same relaxation and decomposition approach holds
- But monotonicity in resource usage is not guaranteed
- Sufficient conditions for optimality of index rule
  - ▷ existence of dominant competitor: if smallest index value of one competitor is greater or equal to index values of the others
  - ▷ reinitializing if not allocated and if index value of the initial state is greatest

Reinitializing Competitors

- Buffer management: TCP Tahoe drops sending rate to 1 packet
- Optimal search: forgetting of relevant information and replacing it with a prior state
- Job scheduling: new job arrival

Conclusion

- Brief account of powerful Lagrangian approach to study of optimality of index rules
- New sufficient conditions for optimality
- Proofs of optimality are scarce; wanted:
  - ▷ if competitors are symmetric
  - ▷ if Whittle index is myopic
  - ▷ if several competitors can be allocated resource
  - ▷ if several resource units can be allocated to competitor

Thank you for your attention!