Model Selection in Markovian Processes

Shie Mannor

Department of Electrical Engineering
Technion
Joint work with Dotan Di-Castro (Technion) and Duncan Simester (MIT)

NIPS December 2011

What matters in policies?

Planning under uncertainty: We typically want to maximize the expected average/discounted reward

- Model is known.
- Assumes we can tradeoff different elements and monitize.
- Expectation can be tricky.
- Rare events can be meaningful (black swans).

What matters in policies?

Planning under uncertainty: We typically want to maximize the expected average/discounted reward

- Model is known.
- Assumes we can tradeoff different elements and monitize.
- Expectation can be tricky.
- Rare events can be meaningful (black swans).

What matters in policies?

Planning under uncertainty: We typically want to maximize the expected average/discounted reward

- Model is known.
- Assumes we can tradeoff different elements and monitize.
- Expectation can be tricky.
- Rare events can be meaningful (black swans).

What matters in policies?

Planning under uncertainty: We typically want to maximize the expected average/discounted reward

- Model is known.
- Assumes we can tradeoff different elements and monitize.
- Expectation can be tricky.
- Rare events can be meaningful (black swans).

What matters in policies?

Planning under uncertainty: We typically want to maximize the expected average/discounted reward

- Model is known.
- Assumes we can tradeoff different elements and monitize.
- Expectation can be tricky.
- Rare events can be meaningful (black swans).

Types of uncertainty

- Deterministic uncertainty in the parameters $\rightarrow$ Robust MDPs. (Known model)
- Probabilistic uncertainty in the parameters $\rightarrow$ Bayesian RL. (Known model)
- Uncertainty due to random transitions/rewards $\rightarrow$ Risk sensitive optimization (mean-variance, percentile, coherent risk measures).
- Model uncertainty $\rightarrow$ This talk

Types of uncertainty

- Deterministic uncertainty in the parameters $\rightarrow$ Robust MDPs. (Known model)
- Probabilistic uncertainty in the parameters $\rightarrow$ Bayesian RL. (Known model)
- Uncertainty due to random transitions/rewards $\rightarrow$ Risk sensitive optimization (mean-variance, percentile, coherent risk measures).
- Model uncertainty $\rightarrow$ This talk

Types of uncertainty

- Deterministic uncertainty in the parameters $\rightarrow$ Robust MDPs. (Known model)
- Probabilistic uncertainty in the parameters $\rightarrow$ Bayesian RL. (Known model)
- Uncertainty due to random transitions/rewards $\rightarrow$ Risk sensitive optimization (mean-variance, percentile, coherent risk measures).
- Model uncertainty $\rightarrow$ This talk

Types of uncertainty

- Deterministic uncertainty in the parameters $\rightarrow$ Robust MDPs. (Known model)
- Probabilistic uncertainty in the parameters $\rightarrow$ Bayesian RL. (Known model)
- Uncertainty due to random transitions/rewards $\rightarrow$ Risk sensitive optimization (mean-variance, percentile, coherent risk measures).
- Model uncertainty $\rightarrow$ This talk

Motivation I

- Open pit mining (with BHP Biliton)
- Objective: dig out gold
- Objective: Keep throughput reasonable, but under severe variance constraints
- Model is not known for mining but "known" for the rest of the supply chain
  Model is terribly complicated

Motivation I

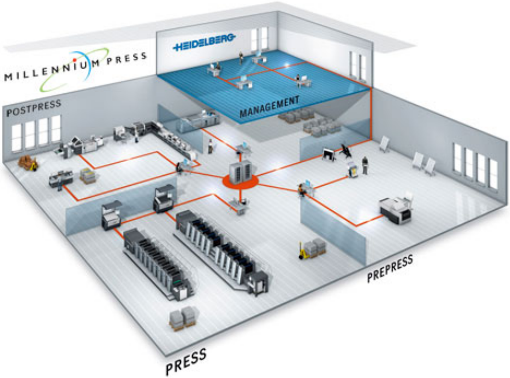
- Open pit mining (with BHP Biliton)
- Objective: dig out gold
- Objective: Keep throughput reasonable, but under severe variance constraints
- Model is not known for mining but "known" for the rest of the supply chain
  Model is terribly complicated

Motivation I

- Open pit mining (with BHP Biliton)
- Objective: dig out gold
- Objective: Keep throughput reasonable, but under severe variance constraints
- Model is not known for mining but “known” for the rest of the supply chain
  Model is terribly complicated

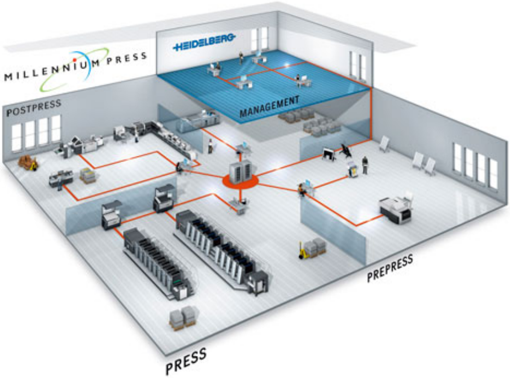
Motivation II

- Print Service Providers (with HP-Research Labs)
- A scheduling problem with many machines
- Objective: Maximize reward, but under operational constraints
- Where does the model come from?
- A fairly stochastic problem.



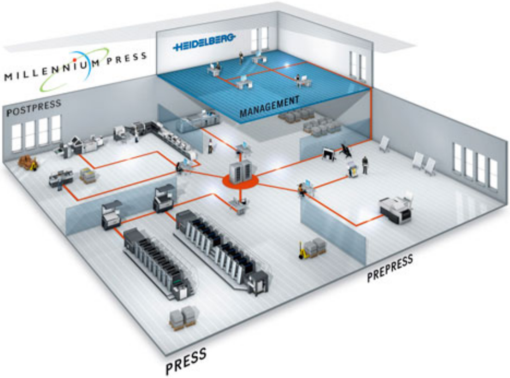
Motivation II

- Print Service Providers (with HP-Research Labs)
- A scheduling problem with many machines
- Objective: Maximize reward, but under operational constraints
- Where does the model come from?
- A fairly stochastic problem.



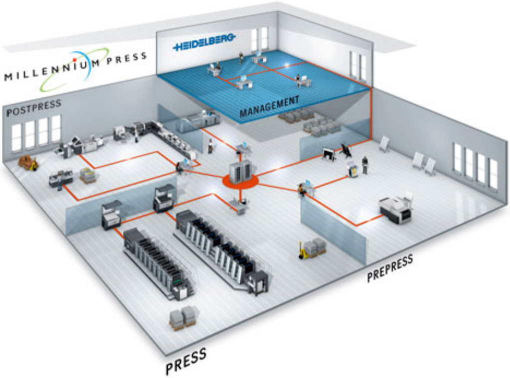
Motivation II

- Print Service Providers (with HP-Research Labs)
- A scheduling problem with many machines
- Objective: Maximize reward, but under operational constraints
- Where does the model come from?
- A fairly stochastic problem.



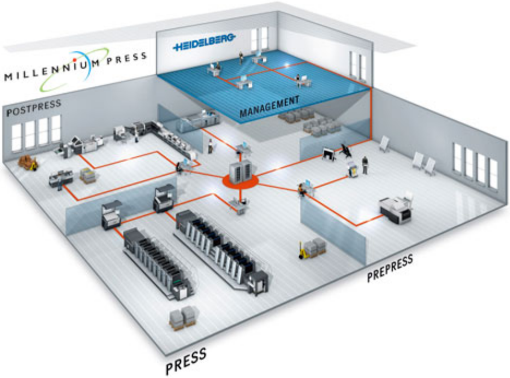
Motivation II

- Print Service Providers (with HP-Research Labs)
- A scheduling problem with many machines
- Objective: Maximize reward, but under operational constraints
- Where does the model come from?
- A fairly stochastic problem.



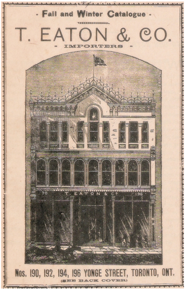
Motivation II

- Print Service Providers (with HP-Research Labs)
- A scheduling problem with many machines
- Objective: Maximize reward, but under operational constraints
- Where does the model come from?
- A fairly stochastic problem.



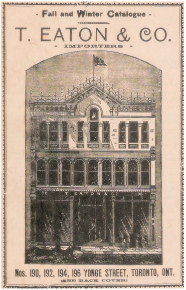
Motivation III

- Large US retailer (Fortune 500 company)
- Marketing problem: send or not send coupon/invitation/mail order catalogue
- Common wisdom: per customer look at RFM
- Recency, Frequency, Monetary value
- Dynamics matter
- How to discretize?



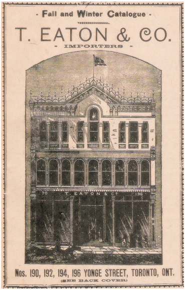
Motivation III

- Large US retailer (Fortune 500 company)
- Marketing problem: send or not send coupon/invitation/mail order catalogue
- Common wisdom: per customer look at RFM
- Recency, Frequency, Monetary value
- Dynamics matter
- How to discretize?



Motivation III

- Large US retailer (Fortune 500 company)
- Marketing problem: send or not send coupon/invitation/mail order catalogue
- Common wisdom: per customer look at RFM
- Recency, Frequency, Monetary value
- Dynamics matter
- How to discretize?

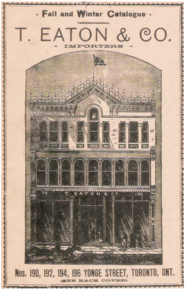


Motivation III

- Large US retailer (Fortune 500 company)
- Marketing problem: send or not send coupon/invitation/mail order catalogue

- Common wisdom: per customer look at RFM
- Recency, Frequency, Monetary value

- Dynamics matter
- How to discretize?



Common to the problems

Much $$$ on the line

- Real state space is huge with lots of uncertainty and parameters. Problem may not even be Markov.
- Batch data are available with no opportunity for exploration
- Operative solution: build a small MDP ($<$ 300 states!), solve, apply.
- Function approximation does not seems to buy much here $\rightarrow$ isolated problems are solved with special solvers.
- Risk and uncertainty are of the essence

Common to the problems

Much $$$ on the line

- Real state space is huge with lots of uncertainty and parameters. Problem may not even be Markov.
- Batch data are available with no opportunity for exploration
- Operative solution: build a small MDP (< 300 states!), solve, apply.
- Function approximation does not seems to buy much here $\rightarrow$ isolated problems are solved with special solvers.
- Risk and uncertainty are of the essence

Common to the problems

Much $$$ on the line

- Real state space is huge with lots of uncertainty and parameters. Problem may not even be Markov.
- Batch data are available with no opportunity for exploration
- Operative solution: build a small MDP (< 300 states!), solve, apply.
- Function approximation does not seems to buy much here $\rightarrow$ isolated problems are solved with special solvers.
- Risk and uncertainty are of the essence

Common to the problems

Much $$$ on the line

- Real state space is huge with lots of uncertainty and parameters. Problem may not even be Markov.
- Batch data are available with no opportunity for exploration
- Operative solution: build a small MDP ($<$ 300 states!), solve, apply.
- Function approximation does not seems to buy much here $\rightarrow$ isolated problems are solved with special solvers.
- Risk and uncertainty are of the essence

Common to the problems

Much $$$ on the line

- Real state space is huge with lots of uncertainty and parameters. Problem may not even be Markov.
- Batch data are available with no opportunity for exploration
- Operative solution: build a small MDP ($<$ 300 states!), solve, apply.
- Function approximation does not seems to buy much here $\rightarrow$ isolated problems are solved with special solvers.
- Risk and uncertainty are of the essence

Common to the problems

Much $$$ on the line

- Real state space is huge with lots of uncertainty and parameters. Problem may not even be Markov.
- Batch data are available with no opportunity for exploration
- Operative solution: build a small MDP ($<$ 300 states!), solve, apply.
- Function approximation does not seems to buy much here $\rightarrow$ isolated problems are solved with special solvers.

- Risk and uncertainty are of the essence

Two important questions

1. What model to use?

2. If I choose a model - how to optimize? (Not today.)

The Model Selection Problem

We will focus on a simpler problem:

1. Ignore action completely (MRP). We have: State $\rightarrow$ reward $\rightarrow$ next state.
2. We observe a sequence of T observations and rewards that occur in some space $\mathcal{O} \times \mathbb{R}$ ($\mathcal{O}$ is complicated)

$$\mathcal{D}(T) = (o_1, r_1, o_2, r_2, \ldots, o_T, r_T).$$

3. We are given K mappings from $\mathcal{O}$ to states spaces $\mathcal{S}_1, \ldots, \mathcal{S}_K$, belong to MRPs $M_1, \ldots, M_K$, respectively.
   Each mapping $H_i : \mathcal{O} \rightarrow \mathcal{S}_i$ describes a model where $\mathcal{S}_i = \{x_1^{(i)}, \ldots, x_{|\mathcal{S}_i|}^{(i)}\}$ is the state space of the MRP M_i.
4. We do not describe how the mappings $\{H_i\}_{i=1}^K$ are constructed.

The Identification Problem

A model selection criterion takes as input D_T and the models $M_1, \ldots, M_k$, and it returns one of the k models as the proposed true model.

Definition: A model selection criterion is weakly consistent if

$$\mathbb{P}^i\left(\hat{M}(D_T) \neq i\right) \to 0 \quad \text{as } n \to \infty,$$

where $\mathbb{P}^i$ is the probability induced when model i is the correct model.

Penalized Likelihood Criteria

In abstraction: data samples $y_1, y_2, \ldots, y_T$.

$$L_i(T) = \max_{\theta}\{\log P(y_1, \ldots, y_T | M_i(\theta))\}.$$

We denote the dimension of θ by $|M_i|$. Then, an MDL model estimator has the following structure

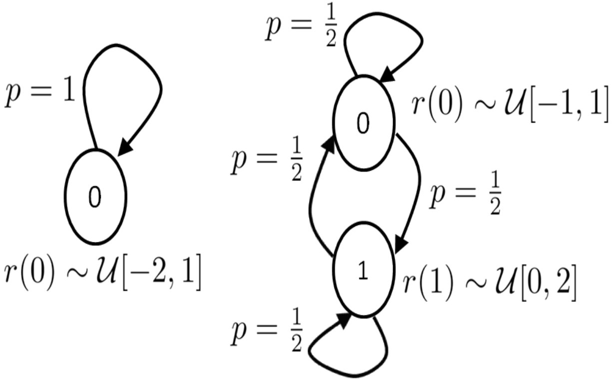
$$\mathrm{MDL}(i) \triangleq |M_i| f(T) - L_i(T),$$

where $f(T)$ is some sub-linear function.

Many related criteria: AIC, BIC, and many others.

Impossibility result

Theorem: There does not exist a consistent MDL-like criterion.



Identifying Markov Reward Processes

We look at the aggregate prediction error.

Two types of aggregations:

1. Reward aggregation
2. Transition probability aggregation

We will focus on refined models: $M_1 \preceq M_2$ is M_2 is a refinement of M_1.

Reward Aggregation

Define *reward mean square error* (RMSE) operator to be

$$\mathcal{L}^i_{RMSE}(D_T) = \frac{1}{T} \sum_{j \in \mathcal{S}_i} \epsilon(x^i_j),$$

where $\epsilon(x^i_j)$ is the error in state j in model i of the reward estimate.

Observation: $\lim_{T\to\infty} \mathcal{L}^i_{RMSE}(D_T) = \sum_{x \in \mathcal{S}_i} \pi(x)\mathrm{Var}(x)$.

Lemma: Suppose M_i contains M_k. Then, for a single trajectory D_T we have $\mathcal{L}^i_{RMSE}(D_T) \leq \mathcal{L}^k_{RMSE}(D_T)$. Moreover, if the states aggregated in M_i are with different mean rewards, then the inequality is sharp.

Corollary: Consider a series of refined models $M_1 \preceq \ldots \preceq M_k$. Then,

$$\mathcal{L}^1_{RMSE}(D_T) \geq \mathcal{L}^2_{RMSE}(D_T) \geq \ldots \geq \mathcal{L}^k_{RMSE}(D_T).$$

Reward Aggregation Score

The *(reward) score* for the j-th model to be

$$\hat{M}(j) = |M_j| \frac{f(T)}{T} + \mathcal{L}^j_{RMSE}(D_T), \tag{1}$$

where $f(T)$ is a sub-linear increasing function with $\lim_{T\to\infty} f(T)/\sqrt{T} \to \infty$.
Based on the RMSE, we consider the following model selector

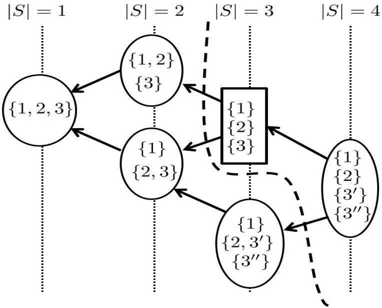
$$\hat{M}_{RMSE} = \arg\min_j \left\{ \hat{M}(j) \right\}.$$

Theorem: The model selector $\hat{M}_{MSE}$ is weakly consistent.

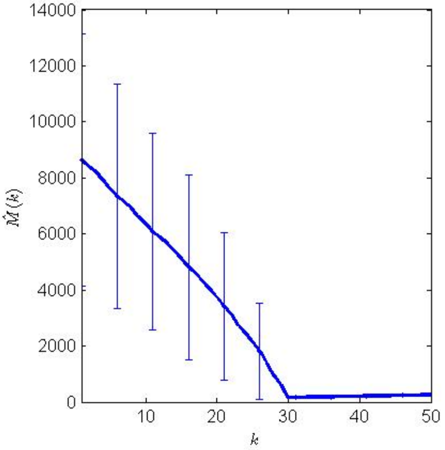
Comment: Not hard to get finite time analysis.

Hierarchical model selection

Have a comparative test: $\rightarrow$ select the best model in a hierarchy

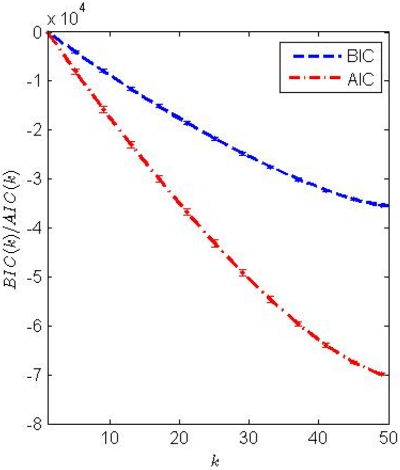


Experiments with artificial data



The figure reports the test statistic $\hat{M}(k)$ for different model dimensions k. The error bars are one standard deviation from the mean.

Experiments with artificial data



The test statistic $\hat{A}IC(k)/\hat{B}IC(k)$ for different model dimensions k.

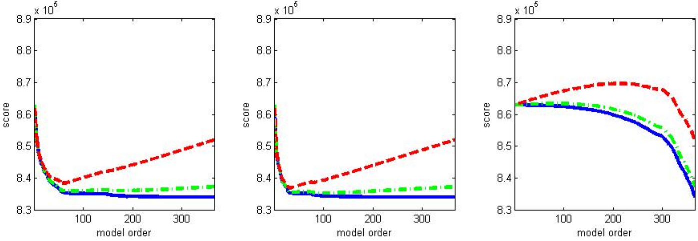
Experiments with real data

Large US apparel retailer.
RFM measures: Recency, Frequency and Monetary value

Problem: How to aggregate? Focus on recency

1. Randomly
2. Most recent
3. Least recent

Real data



(random) (lowest) (highest)

Each graph is for a different value of $f(T)/T$: blue =1, green = 10, red = 50.

Conclusion

- A very special model selection problem.
- Standard approaches fail - but not all is lost
- Mismatched models?
- No word on optimization
- How to aggregate?

Outlook

- Learning from batch: What is the objective?
- Finding the model is “easy"
- Learning the model actively (cf. Maillard, Munos, and Ryabko, this NIPS).
- Todo: Handling model, parametric and inherent uncertainty
- Todo: Large state space (but - who cares?)

Outlook

- Learning from batch: What is the objective?
- Finding the model is "easy"
- Learning the model actively (cf. Maillard, Munos, and Ryabko, this NIPS).
- Todo: Handling model, parametric and inherent uncertainty
- Todo: Large state space (but - who cares?)

Outlook

- Learning from batch: What is the objective?
- Finding the model is “easy"
- Learning the model actively (cf. Maillard, Munos, and Ryabko, this NIPS).
- Todo: Handling model, parametric and inherent uncertainty
- Todo: Large state space (but - who cares?)

Outlook

- Learning from batch: What is the objective?
- Finding the model is “easy"
- Learning the model actively (cf. Maillard, Munos, and Ryabko, this NIPS).
- Todo: Handling model, parametric and inherent uncertainty
- Todo: Large state space (but - who cares?)

Outlook

- Learning from batch: What is the objective?
- Finding the model is “easy"
- Learning the model actively (cf. Maillard, Munos, and Ryabko, this NIPS).
- Todo: Handling model, parametric and inherent uncertainty
- Todo: Large state space (but - who cares?)