Online Learning Algorithms for Network Optimization with Unknown Variables

Yi Gai
Advisor: Bhaskar Krishnamachari

Ming Hsieh Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089

ygai@usc.edu
http://www-scf.usc.edu/∼ygai/

April 20, 2012

Summary of My PhD Research (1)

Bio: BS'05, Tsinghua Univ. -> MS'07, Tsinghua Univ. -> Join ANRG, USC in Fall 2007

A. Online Learning Algorithms:

1. MAB with Linear Rewards
   - i.i.d. and Markovian formulation
   - DySPAN'10, IEEE/ACM Trans. Networking, Globecom'11, Machine Learning (under submission), Infocom'12 (mini-conf), SECON'12
   - joint work with Bhaskar Krishnamachari, Rahul Jain, Mingyan Liu.
2. Learning in Decentralized Settings
   - Globecom'11
   - joint work with Bhaskar Krishnamachari
3. Learning with Non-Linear Rewards
   - ITA'12 (under submission)
   - joint work with Bhaskar Krishnamachari
4. Non-Bayesian Restless Multi-Armed Bandits
   - ICASSP'11, IEEE Trans. Information Theory (under submission), Allerton'11
   - joint work with Bhaskar Krishnamachari, Qing Zhao, Wenhan Dai, Naumaan Nayyar.

Summary of My PhD Research (2)

B. Network Game Theory, Algorithmic Game Theory and Economics

1. Incentive Mechanisms for M/M/1 Queueing Game
   - Infocom'11, IEEE Trans. Automatic Control (under submission)
   - joint work with Bhaskar Krishnamachari, Hua Liu.
2. Funding Games
   - EC'12 (under submission)
   - joint work with Bhaskar Krishnamachari, Amotz Bar-Noy, Matthew Johnson, George Rabanca

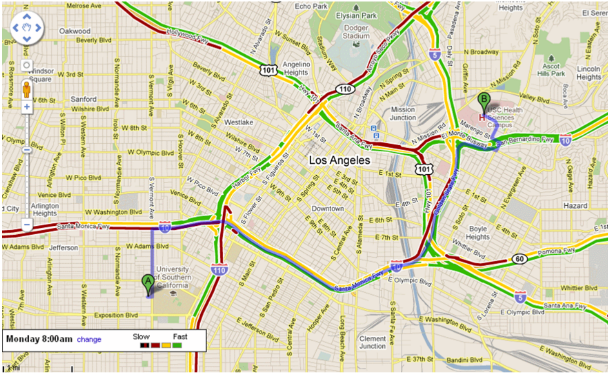
C. Wireless Networks and Communications

1. The Saturation Throughput Region of p-Persistent CSMA
   - ITA'11
   - joint work with Bhaskar Krishnamachari, Shankar Ganesan.
2. Subcarrier Allocation in Multiuser OFDM Systems
   - WCNC'10
   - joint work with Bhaskar Krishnamachari, Pai-Han Huang, Ashwin Sridharan.

Today's focus: Online Learning Algorithms

Online Learning Algorithms: Motivating Example 1

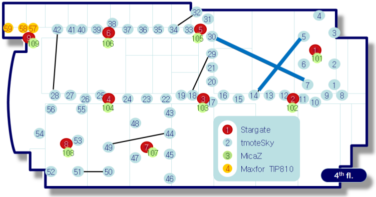
- Finding the lowest expected delay path through traffic using prior observations.



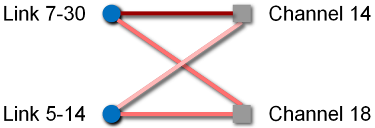
A sample path from Google Maps.

Motivating Example 2

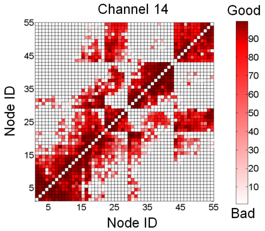
- Channel allocation for wireless links.



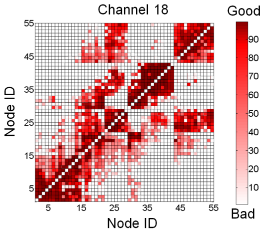
The TutorNet testbed at USC.



Bipartite link channel allocation graph.



Link qualities on channel 14.



Link qualities on channel 18.

Online Learning for Stochastic Network Optimization

- Common theme: find an optimal network structure (best path / matching), assuming the underlying edge weights are unknown random variables.
- Problem formulation:

$$\begin{aligned} \max \quad & \mathbb{E}[\sum_{\tau=1}^{t} f(\mathbf{a}(\tau), \mathbf{X}(\tau))] \\ s.t. \quad & \mathbf{a}(\tau) \in \mathcal{F} \end{aligned} \tag{1}$$

where $\mathbf{X}$ are unknown random variables; $\mathbf{a}(\tau)$ is action at time τ; $\mathcal{F}$ is a finite set.

Our focus on this topic

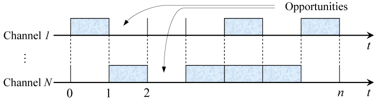
- Develop online learning algorithms for stochastic network optimization.

Multi-Armed Bandits (MAB)

- Multi-armed bandit (MAB) problems provide a fundamental approach to learning under stochastic rewards.
- It has rich applications in networking contexts

Trade-off

- Exploration vs Exploitation



Cognitive Radio Networks [e.g. Anandkumar *et al.*'10]

Internet advertising [e.g. Pandey *et al.*'07]

Summary of Proposed Algorithms

Problems	Random Process	Proposed Algorithms	Regret Bound *
MAB with Linear Rewards	i.i.d.	LLR	$O(N^4 \ln t)$
		LLR-K	$O(N^4 \ln t)$
		LLR with a β-approximation algorithm	$O(N^4 \ln t)^\natural$
MAB with Linear Rewards	Rested Markovian	MLMR	$O(N^4 \ln t)^\sharp$
	Rested Markovian		$O(L(t)N^4 \ln t)^\dagger$
MAB with Linear Rewards	Restless Markovian	CLRMR	$O(N^4 \ln t)^\sharp$
	Restless Markovian		$O(L(t)N^4 \ln t)^\dagger$
Distributed Learning with Prioritization	i.i.d.	DLP	$O(M(N+M) \ln t)$
Distributed Learning with Fairness	i.i.d.	DLF	$O(M(N-M) \ln t)$
Selective learning of the K-th largest arm	i.i.d.	SL(K)	$O(N \ln t)$
MAB with Non-Linear Rewards	i.i.d.	CWF1	$O(N^4 \ln t)$
		CWF2	$O(\frac{N^2}{B(N)^2} \ln t)$
Non-Bayesian Restless MAB with identical transition matrices	Restless Markovian	SPUDC	$O(L(t) \ln t)^\dagger$
Non-Bayesian Restless MAB with non-identical transition matrices	Restless Markovian	R2PC	$O(L(t) \ln t)^\dagger$

Notes:
$*$. Upper bounds on regret are achieved uniformly.
$\natural$. β-approximation regret.
$\sharp$. weak regret; an upper bound on L is known.
$\dagger$. $L(t)$ is any arbitrarily slowly diverging non-decreasing sequence.

Thanks!

ygai@usc.edu