Learning, Games, and Networks

Abhishek Sinha
Laboratory for Information and Decision Systems

MIT

ML Talk Series @CNRG

December 12, 2016

Outline

1. Prediction With Experts' Advice
2. Application to Game Theory
3. Online Shortest Path (OSP)
4. Open Problems

Outline

1. Prediction With Experts' Advice
2. Application to Game Theory
3. Online Shortest Path (OSP)
4. Open Problems

Prediction with Experts' Advice - Setup [1]

- Setting: We are playing a game with an opponent for T rounds. At each round t we first output a real number $\hat{x}_t$ and then the opponent reveals another real number y_t. Our (convex) loss at round t is $\ell(x_t, y_t)$.

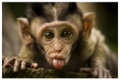
Prediction with Experts' Advice - Setup [1]

- Setting: We are playing a game with an opponent for T rounds. At each round t we first output a real number $\hat{x}_t$ and then the opponent reveals another real number y_t. Our (convex) loss at round t is $\ell(x_t, y_t)$.
- The Experts: There are N experts to assist us. At round t, the expert i forecasts a number $f_i(t)$.
- $\hat{x}_t$ could be taken to be any function of the past and present forecasts.

Prediction with Experts' Advice - Setup [1]

- Setting: We are playing a game with an opponent for T rounds. At each round t we first output a real number $\hat{x}_t$ and then the opponent reveals another real number y_t. Our (convex) loss at round t is $\ell(x_t, y_t)$.
- The Experts: There are N experts to assist us. At round t, the expert i forecasts a number $f_i(t)$.
- $\hat{x}_t$ could be taken to be any function of the past and present forecasts.

Expert 1  forecasts$=\{10, -4, 2, 30, \ldots\}$

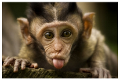
Expert 2  forecasts$=\{20, 6, -50, 10, \ldots\}$

Expert 3  forecasts$=\{30, -5, -10, 42, \ldots\}$

Prediction with Experts' Advice - Setup [1]

- Setting: We are playing a game with an opponent for T rounds. At each round t we first output a real number $\hat{x}_t$ and then the opponent reveals another real number y_t. Our (convex) loss at round t is $\ell(x_t, y_t)$.
- The Experts: There are N experts to assist us. At round t, the expert i forecasts a number $f_i(t)$.
- $\hat{x}_t$ could be taken to be any function of the past and present forecasts.

Expert 1  forecasts=$\{10, -4, 2, 30, \ldots\}$

Expert 2  forecasts=$\{20, 6, -50, 10, \ldots\}$

Expert 3  forecasts=$\{30, -5, -10, 42, \ldots\}$

- Observed Info: After each round we get to observe the loss incurred by each experts $\ell(f_{it}, y_t)$.

Prediction Strategy

Problem Design an online strategy for selecting $\{\hat{x}_t\}$ so that our cumulative loss remains close to the best fixed expert in the hindsight.

Prediction Strategy

Problem Design an online strategy for selecting $\{\hat{x}_t\}$ so that our cumulative loss remains close to the best fixed expert in the hindsight.
Let $L(i, t)$ denote the total loss of expert i up to time $t-1$, i.e.,

$$L(i, t) = \sum_{\tau=1}^{t-1} \ell(f_i(\tau), y_\tau) \quad (1)$$

Prediction Strategy

Problem Design an online strategy for selecting $\{\hat{x}_t\}$ so that our cumulative loss remains close to the best fixed expert in the hindsight.
Let $L(i, t)$ denote the total loss of expert i up to time $t-1$, i.e.,

$$L(i,t) = \sum_{\tau=1}^{t-1} \ell(f_i(\tau), y_\tau) \tag{1}$$

Performance Metric: A natural metric to measure the performance of an *online* algorithm π is to bound its regret $R^\pi(T)$ up to time T defined as follows:

$$R^\pi(T) = \max_{\boldsymbol{y}} \left(L^\pi(T) - \min_i L(i, T)\right)$$

- A natural strategy is to take a weighted combination $\boldsymbol{w}(t)$ of experts' prediction at time t, where the weights are *decreasing* with the expert's cumulative losses up to time $t-1$.
- This ensures that we give more weights to the experts who are doing better in terms of their total losses

Prediction Strategy

Problem Design an online strategy for selecting $\{\hat{x}_t\}$ so that our cumulative loss remains close to the best fixed expert in the hindsight.
Let $L(i,t)$ denote the total loss of expert i up to time $t-1$, i.e.,

$$L(i,t) = \sum_{\tau=1}^{t-1} \ell(f_i(\tau), y_\tau) \tag{1}$$

Performance Metric: A natural metric to measure the performance of an *online* algorithm π is to bound its regret $R^\pi(T)$ up to time T defined as follows:

$$R^\pi(T) = \max_{\boldsymbol{y}} \left(L^\pi(T) - \min_i L(i,T) \right)$$

- A natural strategy is to take a weighted combination $\boldsymbol{w}(t)$ of experts' prediction at time t, where the weights are *decreasing* with the expert's cumulative losses up to time $t-1$.
- This ensures that we give more weights to the experts who are doing better in terms of their total losses

In other words, for a suitable weight vector $\boldsymbol{w}(t)$, we predict

$$\hat{x}_t = \frac{\sum_i w_{i,t} f_i(t)}{\sum_i w_{i,t}} \tag{2}$$

Performance of EXP

Question: How to select the weights?

Performance of EXP

Question: How to select the weights?

Strategy EXP

$$w_{i,t} = \exp(-\eta L(i, t-1)), \quad (3)$$

for some $\eta > 0$.

[Theorem: Regret of EXP]

$$R^{\mathrm{EXP}}(T) \leq \frac{\ln N}{\eta} + \frac{\eta}{T}$$

Performance of EXP

Question: How to select the weights?

Strategy EXP

$$w_{i,t} = \exp(-\eta L(i, t-1)), \quad (3)$$

for some $\eta > 0$.

[Theorem: Regret of EXP]

$$R^{\mathrm{EXP}}(T) \leq \frac{\ln N}{\eta} + \frac{\eta}{T}$$

By choosing $\eta \equiv \sqrt{T \ln N}$, we obtain a $\mathcal{O}(\sqrt{T})$ regret bound for the exponential algorithm. Thus the average regret per round diminishes to zero as $T \to \infty$. Proof involves a Lyapunov argument with the logarithm of the total weights $L(t) = \ln(\sum_i w_{i,t})$ as the Lyapunov function.

Lower Bounds : Can we do better?

Here is my existential proof that we cannot do better than $\mathcal{O}(\sqrt{T})$ in terms of regret, asymptotically.

Proof: Assume that there are two experts and an environment, all of which output iid binary sequences $\{\boldsymbol{f}(t), y(t)\}$, independent of each other. Choose the loss function to be $\ell(x, y) = |x - y|$.

Now observe:

- The expected loss of each expert at any slot t is simply $\frac{1}{2}$, independent of $\boldsymbol{y}$.
- The expected loss of the player at each slot is $\frac{1}{2}$, *irrespective* of his output $\hat{x}(t) \in \{0, 1\}$.

Lower Bounds : Can we do better?

Here is my existential proof that we cannot do better than $\mathcal{O}(\sqrt{T})$ in terms of regret, asymptotically.

Proof: Assume that there are two experts and an environment, all of which output iid binary sequences $\{\boldsymbol{f}(t), y(t)\}$, independent of each other. Choose the loss function to be $\ell(x, y) = |x - y|$.

Now observe:

- The expected loss of each expert at any slot t is simply $\frac{1}{2}$, independent of $\boldsymbol{y}$.
- The expected loss of the player at each slot is $\frac{1}{2}$, *irrespective* of his output $\hat{x}(t) \in \{0, 1\}$.

Suffices to show that regret with respect to *any* expert (say, the first expert) is $\mathcal{O}(\sqrt{T})$.

Proof Contd.

Analysis:

$$\begin{aligned}\mathbb{P}(R^\pi(T) \geq \sqrt{T}) &\geq \mathbb{P}(L^\pi(T) - L(1,T) \geq \sqrt{T}) \\ &= \mathbb{P}(\sum_{t=1}^{T} (|\boldsymbol{Y}(t) - \hat{\boldsymbol{X}}(t)| - |\boldsymbol{Y}(t) - \boldsymbol{F}_1(t)|) \geq \sqrt{T})\end{aligned}$$

Define the random variable $\boldsymbol{Z}(t) = (|\boldsymbol{Y}(t) - \hat{\boldsymbol{X}}(t)| - |\boldsymbol{Y}(t) - \boldsymbol{F}_1(t)|)$.
Clearly, the sequence of random variables $\{\boldsymbol{Z}(t)\}$ are i.i.d. with zero mean and variance $= \frac{1}{2}$.
Thus, we have

$$\lim_{T\to\infty} \mathbb{P}(R^\pi(T) \geq \sqrt{T}) \geq \lim_{T\to\infty} \mathbb{P}(\sum_{t=1}^{T} \boldsymbol{Z}(t) \geq \sqrt{T}) \overset{\mathrm{CLT}}{=} 1 - \Phi(\sqrt{2}) \geq 0.07$$

Thus, for large enough T, there is a strictly positive probability that regret is greater than $\sqrt{T}$. This shows that there exists a sequence of forecasts $\boldsymbol{f}(t)$ and an adversarial sequence y_t such that *no* online strategy achieves regret smaller that $\mathcal{O}(\sqrt{T})$. ■

Outline

1. Prediction With Experts' Advice
2. Application to Game Theory
3. Online Shortest Path (OSP)
4. Open Problems

Set Up: Two Player zero-sum game [2]

- Consider a finite, two player (row and column), zero-sum game, with the payoff matrix given by $\boldsymbol{M}$.
- The row player plays a randomized strategy $\boldsymbol{P}$ and the column player plays a randomized strategy $\boldsymbol{Q}$.
- Loss of the row and column player is $\boldsymbol{P}^T\boldsymbol{M}\boldsymbol{Q}$ and -$\boldsymbol{P}^T\boldsymbol{M}\boldsymbol{Q}$.

Set Up: Two Player zero-sum game [2]

- Consider a finite, two player (row and column), zero-sum game, with the payoff matrix given by $\boldsymbol{M}$.
- The row player plays a randomized strategy $\boldsymbol{P}$ and the column player plays a randomized strategy $\boldsymbol{Q}$.
- Loss of the row and column player is $\boldsymbol{P}^T\boldsymbol{M}\boldsymbol{Q}$ and -$\boldsymbol{P}^T\boldsymbol{M}\boldsymbol{Q}$.

Assume that the game is played sequentially in the following orders:

- Case 1: Row player makes the first move
  - Optimal strategy is given by the solution of the following problem

$$V_1 = \min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{P}^T\boldsymbol{M}\boldsymbol{Q} \tag{4}$$

- Case 2: Column player makes the first move
  - Optimal strategy is given by the solution of the following problem

$$V_2 = \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{P}^T\boldsymbol{M}\boldsymbol{Q}, \tag{5}$$

where V_1 and V_2 are the corresponding pay off of the row player.

Minimax Theorem

Intutively, for any player making move after the other player is advantageous as he knows his opponent's move. Hence, it is no surprise that

$$V_2 = \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} \leq \min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} = V_1 \tag{6}$$

Minimax Theorem

Intutively, for any player making move after the other player is advantageous as he knows his opponent's move. Hence, it is no surprise that

$$V_2 = \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} \leq \min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} = V_1 \tag{6}$$

The startling result by Von Neumann is that the second player does not have any advantage *in expectation*, if the first player plays optimally:

Theorem (Minimax Theorem)

For any payoff matrix $\boldsymbol{M}$*:*

$$\min_{\boldsymbol{Q}} \max_{\boldsymbol{P}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} = \max_{\boldsymbol{P}} \min_{\boldsymbol{Q}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} \tag{7}$$

Minimax Theorem

Intutively, for any player making move after the other player is advantageous as he knows his opponent's move. Hence, it is no surprise that

$$V_2 = \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} \leq \min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} = V_1 \tag{6}$$

The startling result by Von Neumann is that the second player does not have any advantage *in expectation*, if the first player plays optimally:

Theorem (Minimax Theorem)

For any payoff matrix $\boldsymbol{M}$*:*

$$\min_{\boldsymbol{Q}} \max_{\boldsymbol{P}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} = \max_{\boldsymbol{P}} \min_{\boldsymbol{Q}} \boldsymbol{P}^T \boldsymbol{M} \boldsymbol{Q} \tag{7}$$

EXP and Game Theory

Questions:

- How to find the minimax strategy ?
  - Classically the minimax strategy is derived by solving an LP, which requires knowledge of the full pay-off matrix.
- What has it to do with the theory of experts' advice?

EXP and Game Theory

Questions:

- How to find the minimax strategy ?
  - Classically the minimax strategy is derived by solving an LP, which requires knowledge of the full pay-off matrix.
- What has it to do with the theory of experts' advice?

Answers:

- We will see that the EXP algorithm gives an alternative *online* strategy for finding the MiniMax strategy, *without* requiring knowledge of the full payoff matrix M.
- It also yields an alternative concise proof of the MiniMax Theorem (which is classically proved via LP duality).

A Minimax Strategy for the Learner

Setting: Repeated Games

- Imagine the learner to be the Row player and the Adversary to be the column player.
- Experts are the actions of the row players.
- At each round the learner plays a randomized strategy $\boldsymbol{P}_t$ and the adversary plays a randomized strategy $\boldsymbol{Q}_t$ (which may depend on the current choice and the past history).
- The learner observes the losses of each action $M(i, \boldsymbol{Q}_t)$ at time t.
- At time t, the learner incurs a loss $\boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t) = \boldsymbol{P}_t^T \boldsymbol{M} \boldsymbol{Q}_t$.

A Minimax Strategy for the Learner

Setting: Repeated Games

- Imagine the learner to be the Row player and the Adversary to be the column player.
- Experts are the actions of the row players.
- At each round the learner plays a randomized strategy $\boldsymbol{P}_t$ and the adversary plays a randomized strategy $\boldsymbol{Q}_t$ (which may depend on the current choice and the past history).
- The learner observes the losses of each action $M(i, \boldsymbol{Q}_t)$ at time t.
- At time t, the learner incurs a loss $\boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t) = \boldsymbol{P}_t^T \boldsymbol{M}\boldsymbol{Q}_t$.

Learner's strategy: EXP

$\boldsymbol{P}_t$ is chosen accoding to the EXP rule: probability of playing action i is proportional to Exponential of the minus of cumulative loss incurred by the action i, i.e.,

$$\begin{aligned} w_{t+1}(i) &= w_t(i)\exp(-\eta M(i, \boldsymbol{Q}_t)) \\ \boldsymbol{P}_t &\propto \boldsymbol{w}_t \end{aligned}$$

Performance Analysis

Invoking our general theory of EXP strategy, we readily obtain the following regret bound. With $\eta \leftarrow \sqrt{T \ln(N)}$, we have

$$\text{Avg. Loss} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t) \leq \min_{\boldsymbol{P}} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}_t) + \Delta_T, \tag{8}$$

where $\Delta_T = \mathcal{O}\sqrt{\frac{\ln N}{T}} \to 0$ as $T \to \infty$.

Performance Analysis

Invoking our general theory of EXP strategy, we readily obtain the following regret bound. With $\eta \leftarrow \sqrt{T \ln(N)}$, we have

$$\text{Avg. Loss} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t) \leq \min_{\boldsymbol{P}} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}_t) + \Delta_T, \tag{8}$$

where $\Delta_T = \mathcal{O}\sqrt{\frac{\ln N}{T}} \rightarrow 0$ as $T \rightarrow \infty$.

Corollary: Note that the above bound holds for any choice of the opponent's strategy $\boldsymbol{Q}_t$.
First, let $\boldsymbol{Q}_t = \boldsymbol{Q}_t^*$ be the MinMax strategy against $\boldsymbol{P}_t$. We have from the LHS,

$$\text{Avg. Loss} \geq \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t^*) \geq \min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) \tag{9}$$

Performance Analysis

Invoking our general theory of EXP strategy, we readily obtain the following regret bound. With $\eta \leftarrow \sqrt{T \ln(N)}$, we have

$$\text{Avg. Loss} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t) \leq \min_{\boldsymbol{P}} \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}_t) + \Delta_T, \tag{8}$$

where $\Delta_T = \mathcal{O}\sqrt{\frac{\ln N}{T}} \to 0$ as $T \to \infty$.

Corollary: Note that the above bound holds for any choice of the opponent's strategy $\boldsymbol{Q}_t$.
First, let $\boldsymbol{Q}_t = \boldsymbol{Q}_t^*$ be the MinMax strategy against $\boldsymbol{P}_t$. We have from the LHS,

$$\text{Avg. Loss} \geq \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q}_t^*) \geq \min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) \tag{9}$$

Bound the RHS by the MaxMin strategy $\boldsymbol{P} = \boldsymbol{P}^*$

$$\text{Avg. Loss} \leq \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{M}(\boldsymbol{P}^*, \boldsymbol{Q}_t^*) + \Delta_T \leq \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) + \Delta_T, \tag{10}$$

Proof Contd.

Combining the above two Eqns., we get

$$\min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) \leq \text{Avg. Loss} \leq \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) + \Delta_T \tag{11}$$

Letting $T \to 0$, and combining it with the previous bound, we recover the famous minmax theorem.

Proof Contd.

Combining the above two Eqns., we get

$$\min_{\boldsymbol{P}} \max_{\boldsymbol{Q}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) \leq \text{Avg. Loss} \leq \max_{\boldsymbol{Q}} \min_{\boldsymbol{P}} \boldsymbol{M}(\boldsymbol{P}, \boldsymbol{Q}) + \Delta_T \quad (11)$$

Letting $T \to 0$, and combining it with the previous bound, we recover the famous minmax theorem.
Moreover, the above proof shows that by choosing $\boldsymbol{Q}_t = \arg\max_{\boldsymbol{Q}} \boldsymbol{M}(\boldsymbol{P}_t, \boldsymbol{Q})$ and updating $\boldsymbol{P}_t$ according to EXP and letting $T \to \infty$, we may obtain the Nash Equilibrium of both players as follows

$$\bar{\boldsymbol{P}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{P}_t$$

$$\bar{\boldsymbol{Q}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{Q}_t$$

Outline

1. Prediction With Experts' Advice
2. Application to Game Theory
3. Online Shortest Path (OSP)
4. Open Problems

Online Shortest Path

Consider the classic problem of online shortest path. Here we are given a graph $\mathcal{G}(V, E)$ with a source-destination pair s, t. At every slot t, an adversary assigns cost $c(e, t)$ to each edge e.
The goal is to pick an $s - t$ path at each slot, so that the cumulative regret upto time T remains small w.r.t. the best path.

Online Shortest Path

Consider the classic problem of online shortest path. Here we are given a graph $\mathcal{G}(V, E)$ with a source-destination pair s, t. At every slot t, an adversary assigns cost $c(e, t)$ to each edge e.
The goal is to pick an $s - t$ path at each slot, so that the cumulative regret upto time T remains small w.r.t. the best path.

- Can frame the problem as an expert's prediction problem, with each path acting as an expert.
- The problem with this naive approach is that there are exponentially many $s - t$ paths and hence experts. The algorithm will suffer an undesirable exponential slow down.

Online Shortest Path

Consider the classic problem of online shortest path. Here we are given a graph $\mathcal{G}(V, E)$ with a source-destination pair s, t. At every slot t, an adversary assigns cost $c(e, t)$ to each edge e.
The goal is to pick an $s - t$ path at each slot, so that the cumulative regret upto time T remains small w.r.t. the best path.

- Can frame the problem as an expert's prediction problem, with each path acting as an expert.
- The problem with this naive approach is that there are exponentially many $s - t$ paths and hence experts. The algorithm will suffer an undesirable exponential slow down.

The solution to get around this issue is the following randomized strategy : Follow The Perturbed Leading Path (FPL).

The Algorithm FPL [3]

At each time t:

1. For each edge e, pick a number $p_t(e)$ randomly from a two-sided exponential distribution (with parameter ϵ).
2. Use the shortest path π_t on the graph $\mathcal{G}$ with edge e's weight $C_t(e) + p_t(e)$, where $C_t(e)$ is total cost incurred by traversing on edge e, i.e.,

$$C_t(e) = \sum_{\tau=1}^{t-1} c(e, \tau)$$

The Algorithm FPL [3]

At each time t:

1. For each edge e, pick a number $p_t(e)$ randomly from a two-sided exponential distribution (with parameter ϵ).
2. Use the shortest path π_t on the graph $\mathcal{G}$ with edge e's weight $C_t(e) + p_t(e)$, where $C_t(e)$ is total cost incurred by traversing on edge e, i.e.,

$$C_t(e) = \sum_{\tau=1}^{t-1} c(e, \tau)$$

Theorem (Performance of FPL)

$$\mathbb{E}(Cost) \leq (1+\epsilon)C^* + \frac{\mathcal{O}(mn \log n)}{\epsilon} \tag{12}$$

The Algorithm FPL [3]

At each time t:

1. For each edge e, pick a number $p_t(e)$ randomly from a two-sided exponential distribution (with parameter ϵ).
2. Use the shortest path π_t on the graph $\mathcal{G}$ with edge e's weight $C_t(e) + p_t(e)$, where $C_t(e)$ is total cost incurred by traversing on edge e, i.e.,

$$C_t(e) = \sum_{\tau=1}^{t-1} c(e, \tau)$$

Theorem (Performance of FPL)

$$\mathbb{E}(Cost) \leq (1+\epsilon)C^* + \frac{\mathcal{O}(mn \log n)}{\epsilon} \tag{12}$$

For completeness, we also mention that, there exists a clever way for efficiently implementing the original EXP algorithm with exponentially many experts, if the underlying graph is a DAG [1].

Outline

1. Prediction With Experts' Advice
2. Application to Game Theory
3. Online Shortest Path (OSP)
4. Open Problems

Online Sabotaged Shortest Path [4]

The setting is similar to OSP, however at each slot some edges are *sabotaged*, and hence, some paths are blocked.

The regret with respect to a constant path P is defined as follows

$$R_T(P) = \sum_{t:\text{P is open}} \boldsymbol{P}_t^{\pi} \cdot \boldsymbol{c}_t - \boldsymbol{P} \cdot \boldsymbol{c}_t \tag{13}$$

Again, the objective is to keep the regret small with respect to *all* paths P simultaneously.

Online Sabotaged Shortest Path [4]

The setting is similar to OSP, however at each slot some edges are *sabotaged*, and hence, some paths are blocked.
The regret with respect to a constant path P is defined as follows

$$R_T(P) = \sum_{t:\text{P is open}} \boldsymbol{P}_t^{\pi} \cdot \boldsymbol{c}_t - \boldsymbol{P} \cdot \boldsymbol{c}_t \tag{13}$$

Again, the objective is to keep the regret small with respect to *all* paths P simultaneously.

Baseline Algorithm : Sleeping Experts

1. Maintain an expert for each path.
2. At each slot, if an expert is awake (path open), update its weight according to the EXP algorithm.
3. If an expert is asleep (path closed), keep its weight unchanged.
4. Select an open path with probability proportional to its weight.

Performance

Regret bound is satisfactory:

$$\mathbb{E}R_T(P) \leq K\sqrt{T \ln D}, \tag{14}$$

where K is the length of the longest $s-t$ path and D is the total number of $s-t$ paths (which could be exponential but that does not matter as we are taking log of it).

Issues:

1. The regret bound is suboptimal (by a factor $\sqrt{K}$)
2. The run time could be exponential

Open Problem [4] : Can we design an efficient prediction strategy similar to FPL with good regret guarantee ?

References

Nicolo Cesa-Bianchi and Gábor Lugosi.
Prediction, learning, and games.
Cambridge university press, 2006.

Yoav Freund and Robert E Schapire.
Game theory, on-line prediction and boosting.
In *Proceedings of the ninth annual conference on Computational learning theory*, pages 325–332. ACM, 1996.

Adam Kalai and Santosh Vempala.
Efficient algorithms for online decision problems.
Journal of Computer and System Sciences, 71(3):291–307, 2005.

Wouter M Koolen, Manfred K Warmuth, and Dmitry Adamskiy.
Open problem: Online sabotaged shortest path.
In *Proceedings of The 28th Conference on Learning Theory*, pages 1764–1766, 2015.