Revised Proof Of AAMAS Paper

1 Analysis with the realizability assumption

Theorem 1. *Given an MDP with generalized eluder coefficient $d_{\mathrm{GEC}}(\cdot)$ and a hypothesis set $\mathcal{H}$ that the realizability assumption holds, by setting*

$$\eta = \sqrt{\frac{d_{\mathrm{GEC}}(1/\sqrt{HK})}{\log(Hn^{\psi}(\mathcal{H})/\delta)\cdot HK}},$$

the regret of the MEX algorithm, given $\mathcal{H}$ and oracle ψ, after K episodes is upper bounded by, with probability at least $1-\delta$,

$$\mathrm{Reg}(K) \lesssim \sqrt{d_{\mathrm{GEC}}(1/\sqrt{HK})\cdot\log(Hn^{\psi}(\mathcal{H})/\delta)\cdot HK}.$$

The statement and the proof of Theorem 1 is correct. However, the meaning of Theorem 1 should be revised. In our AAMAS paper, we said:

The sole term related to the size of the hypothesis set is $n^{\psi}(\mathcal{H})$. Consequently, the magnitude of regret is solely influenced by the type number associated with a hypothesis set, as opposed to the size of the hypothesis set. This phenomenon occurs because policies that are categorized under the same type yield identical rewards when selected by the MEX algorithm. Note that the analysis does not rely on merging or eliminating the same type of policies from the hypothesis set, the MEX algorithm can ignore the same type of policies automatically.

This should be revised into:

We reduce the regret upper bound from $\sqrt{d_{\mathrm{GEC}}(1/\sqrt{HK})\cdot\log(H\,|\mathcal{H}|/\delta)\cdot HK}$ (proved in the original MEX paper) into $\sqrt{d_{\mathrm{GEC}}(1/\sqrt{HK})\cdot\log(Hn^{\psi}(\mathcal{H})/\delta)\cdot HK}$. The size of the hypothesis set $|\mathcal{H}|$ is reduced to the type number $n^{\psi}(\mathcal{H})$. However, the term $d_{\mathrm{GEC}}(1/\sqrt{HK})$ may depend on $|\mathcal{H}|$ for some hypothesis sets. Eliminating these hypothesis sets requires future research on the structure of MDPs, which won't be considered in this paper. The reduction of sample complexity occurs because policies that are categorized under the same type yield identical rewards when selected by the MEX algorithm. Note that the analysis does not rely on merging or eliminating the same type of policies from the hypothesis set, the reduction is done automatically.

2 Analysis without the realizability assumption: revised proof

Theorem 2. *Given an MDP with generalized eluder coefficient $d_{\mathrm{GEC}}(\cdot)$ and a hypothesis set $\mathcal{H}$, by setting*

$$\eta = \sqrt{\frac{d_{\mathrm{GEC}}(1/\sqrt{HK})}{\log(Hn^{\psi}(\mathcal{H})/\delta)\cdot HK}},$$

the regret of the MEX algorithm, given $\mathcal{H}$ and oracle ψ, after K episodes is upper bounded by, with probability at least $1-\delta$,

$$\mathrm{Reg}(K) \lesssim \sqrt{d_{\mathrm{GEC}}(1/\sqrt{HK})\cdot\log(Hn^{\psi}(\mathcal{H})/\delta)\cdot HK} + \varepsilon_{\mathrm{fin}}K + \sqrt{\frac{d_{\mathrm{GEC}}(1/\sqrt{HK})\cdot HK}{\log(Hn^{\psi}(\mathcal{H})/\delta)}}\varepsilon_{\mathbb{P}}K,$$

where $\varepsilon_{\text{fin}} = \min_{\pi \in \mathcal{H}} |V^*(\pi) - V^*(\pi^*)|$, $\pi^*_{\text{fin}} = \text{argmin}_{\pi \in \mathcal{H}} |V^*(\pi) - V^*(\pi^*)|$, *and* $\varepsilon_{\mathbb{P}} = \max_{(s,a,s')} \log \frac{\mathbb{P}(s' | s, a, \pi^*_{\text{fin}}(s))}{\mathbb{P}(s' | s, a, \pi^*(a))}$

Proof. Follow the same proof in our AAMAS paper gives us

$$V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k) \leq \eta \sum_{s=1}^{k} \sum_{h=1}^{H-1} (L_h^{s-1}(\pi^*_{\text{fin}}) - L_h^{s-1}(\pi^k)) + \varepsilon_{\text{fin}}.$$

Note that $L_h^{k-1}(\pi^*_{\text{fin}}) - L_h^{k-1}(\pi^k)$ is the sum of

$$L_h^{k-1}(\pi^*_{\text{fin}}) - L_h^{k-1}(\pi^*)$$

and $L_h^{k-1}(\pi^*) - L_h^{k-1}(\pi^k)$. In the wrong proof of our AAMAS paper, we ignored the term $L_h^{k-1}(\pi^*_{\text{fin}}) - L_h^{k-1}(\pi^*)$. Thus, to revise the proof, the only thing we need to do is to add an upper bound of

$$\eta \sum_{k=1}^{K-1} \sum_{s=1}^{k} \sum_{h=1}^{H-1} L_h^{k-1}(\pi^*_{\text{fin}}) - L_h^{k-1}(\pi^*). \tag{1}$$

By the definition of L_h^{k-1}, we have

$$L_h^{k-1}(\pi^*_{\text{fin}}) - L_h^{k-1}(\pi^*) = \log \frac{\mathrm{P}(s_{h+1}^k | s_h^k, a_h^k, \pi^*_{\text{fin}}(s_h^k))}{\mathrm{P}(s_{h+1}^k | s_h^k, a_h^k, \pi^*(s_h^k))} \leq \varepsilon_{\mathbb{P}}$$

Substituting the above equation into (1) gives

$$\begin{aligned} \eta \sum_{k=1}^{K-1} \sum_{s=1}^{k} \sum_{h=1}^{H-1} L_h^{k-1}(\pi^*_{\text{fin}}) - L_h^{k-1}(\pi^*) &\leq \frac{\eta K^2 H \varepsilon_{\mathbb{P}}}{2} \\ &= \sqrt{\frac{d_{\text{GEC}}(1/\sqrt{HK}) \cdot HK}{\log(H n^{\psi}(\mathcal{H})/\delta)}} \varepsilon_{\mathbb{P}} K. \end{aligned}$$

□