Learning Note Aug 2020

Contextual MDP Note

Baizhi (Daniel) Song

- Theoretical Note for CMDP's formation

1 Contextual Setting and Function Approximation

Here we summarize and discuss several contextual settings in different problems. To recapitulate, the form of context and reduction of parameters is related to whether the context information is independent with state space $\mathcal{S}$ and action space $\mathcal{A}$.

1.1 Example for contextual setting

1.1.1 Contextual MNL

Multinomial logit (MNL) model offer a sequence of assortments of at most K items from a set of N possible items that minimize regret. The utility or the degree of attractive for a specific item may depend on both user and item's features (context $X(i) \subseteq \mathbb{R}^d$) and the probability of being chose is dynamically depends on that context.

- In MNL's setting (Agrawal et al., 2017), the probability of an item being chose is depend on all items in the selected subset, we intuitive use "utility" v_i for item i in subset $\mathcal{S}$ as parameters of its probability. The number of parameter to sample is S.

$$P_i(S) = \frac{v_i}{\sum_{j \in \mathcal{S}} v_j + 1} \qquad i \in \mathcal{S}$$

- By introducing contextual setting (Min-hwan et al., 2019), we represent the utility of each item as a function of feature x(i): $v_i = exp(\theta x(i))$. The number of parameter to sample is d.

$$P_i^x(S) = \frac{exp(\theta^{(i)} x(i))}{\sum_{j \in \mathcal{S}} exp(\theta^{(j)} x(j)) + 1} \qquad \theta \subseteq \mathbb{R}^d$$

1.1.2 Clinical trial CMDP

Patients with different health condition (context $X \subseteq \mathbb{R}^d$) may have different clinical settings (MDP with different $P(s'|s,a)$ and $r(s,a)$)

- Given n patients, we have n different $M_i(A, S, P_i(s'|s,a), R_i(s,a))$ with $n \times S \times S \times A$ parameters for transition matrix. If we consider the MDP with all different contexts, the state and parameter become infinite.

- By introducing contextual setting (Modi et al., 2020), we reduce the number of parameters from $n \times S \times S \times A$ to $d \times S \times S \times A$:

$$P_x(s_i|s,a) = \frac{exp(\theta_{sa}^{(i)} x)}{\sum_{j=1}^{S} exp(\theta_{sa}^{(j)} x)} \qquad \theta_{sa} \subseteq \mathbb{R}^{S \times d}$$

1.2 Function approximation

Use selected feature to approximate large scale MDP, the approximate function can be linear or more complex version (deep neural network)

1.2.1 Q-learning with linear approximation

For a large state set S, we approximate the $Q(s,a)$ by a linear combination of feature set $\overrightarrow{x}(s,a)$:

$$Q(s,a) \approx Q_\theta(s,a) = \theta_0 + \theta_1 x_1(s,a) + ... + \theta_d x_d(s,a) = \theta \overrightarrow{x}(s,a)$$

In this case we reduce the parameters to estimate from $S \times A$ to d.

1.2.2 Transit matrix with function approximation(proposed)

Here we propose four forms of approximation:

- Case-1; feature set $x(s,a,s')$

$$P(s'|s,a) = \frac{exp(\theta x(s,a))}{\sum_{j=1}^{S} exp(\theta x(s,a))} \qquad \theta \subseteq \mathbb{R}^d$$

  $\Rightarrow d$ parameters

- Case-2: feature set $x(s,a)$

$$P(s'|s,a) = \frac{exp(\theta^{(i)} x(s,a))}{\sum_{j=1}^{S} exp(\theta^{(j)} x(s,a))} \qquad \theta \subseteq \mathbb{R}^{S \times d}$$

  $\Rightarrow d \times S$ parameters

- Case-3: feature set $x(s)$

$$P(s'|s,a) = \frac{exp(\theta_a^{(i)} x(s))}{\sum_{j=1}^{S} exp(\theta_a^{(j)} x(s))} \qquad for\ a \in \mathcal{A},\ \theta_a \subseteq \mathbb{R}^{S \times d}$$

  $\Rightarrow d \times S \times A$ parameters

- Case-4: feature set x, similar form to case 2 and 3, and this is the case in the common CMDP setting (clinical trail)

$$P(s'|s,a) = \frac{exp(\theta_{sa}^{(i)} x)}{\sum_{j=1}^{S} exp(\theta_{sa}^{(j)} x)} \qquad for\ s,a \in \mathcal{S} \times \mathcal{A},\ \theta_{sa} \subseteq \mathbb{R}^{S \times d}$$

  $\Rightarrow d \times S \times S \times A$ parameters

In the case of recommendation problem, if the state is a list of user's previous purchase like (x_1, x_3, x_4) then the state space it self could be infinite, in this case we could not take case 2,3 and 4's setting due to the infinite S. Case 1's setting is not covered a lot in the current CMDP's research.

1.2.3 Normalization and sparse structure

In previous part we discussed four different cases of the contextual approximation of transition matrix in MDP. It can be seen that the main difference between different cases is the dependency of contextual feature x with state space $\mathcal{S}$ and action space $\mathcal{A}$. Apart from the feature x, the rest of structure is quite similar: exponential like feature aggregation and normalization. Here we discuss more about the idea of normalization and potential way to solve the infinite state space problem for the case 2 to 4.

$$u(s'|s,a) = exp(\theta_{sa}x(s,a))$$

$$P(s'|s,a) = \frac{u(s'|s,a)}{\sum_{s_i \in S} u(s_i|s,a)}$$

The normalization method used in above section is based on the fact the the sum of the probability within all events is 1 ie. $\sum_{s' \in \mathcal{S}} P(s'|s,a) = 1$.Currently the problem in case 2 to 4 is that the state space $\mathcal{S}$ is extremely large, which make it highly costly to normalize along all states. However, under certain situations (many open AI games where agent moves step by step), it is reasonable to make such an assumption that the next state could only be in a certain 'near subset' ie. $P(s'|s,a) = 0$ for $s' \notin K_{|s,a}$, which also means that we assume the transition matrix is sparse. In this case the normalized probability can be represented as:

$$P(s'|s,a) = \frac{u(s'|s,a)}{\sum_{s_i \in S} u(s_i|s,a)} = \frac{u(s'|s,a)}{\sum_{s_i \in K_{|s,a}} u(s_i|s,a)}$$

So far it is not difficult to realize the above sparse form of transition matrix is similar to the problem setting in contextual MNL. In contextual MNL problem, the probability of an item being chose is normalized along the assortment subset S_t with a finite and tractable cardinality. Here in CMDP we achieve the same from by adding the sparse assumption.

2 Posterior Sampling for CMDP

To sample the parameter in contextual setting, we may not get a conjugate prior so we will do it in by the way of MCMC or variational inference.

3 Contextual PSRL Algorithm

Here we proposed the algorithm of PSRL in Contextual setting.