Bayesian Optimization over High-Dimensional Continuous and Categorical Mixture Variables

Kazuaki TAKEHARA SOKENDAI, Statistical Science, 2nd year of 3 year doctoral program

Background

Searching for items like real estate, movies, or music is a laborious process. Because of the enormous number of things and their high-dimensional properties, the users can only understand their preference gradually and spend a very long time finding the preferable one. Therefore interactive recommender system is getting a lot of attention to solve such problems [Takehara 2017]. The system has to capture a user's preference and recommend items that the user might like.

Problem Settings

We consider user's preference as a block-box function and maximize it as little function evaluation as possible.

$$\mathbf{x}^* = \mathrm{argmax}_{\mathbf{x}}\ f(\mathbf{x})$$

To optimize the black-box function, we use the framework of Bayesian Optimization .

The input $\mathbf{x}$ means properties of an item. It consists of categorical variables $\mathbf{h}$ and continuous variables $\mathbf{z}$, $\mathbf{x} = [\mathbf{h}, \mathbf{z}]$. $\mathbf{h} = [h_1, ..., h_c]$, each variable $h_j \in \{1, ..., N_j\}$ is one of N_j different values. $\mathbf{z}$ is a d-dimensional hypercube Z. We can observe user's reaction y which is one of ordered set $\{0, ..., N_y\text{-}1\}$, for example, 0: no reaction, 1: viewed, and 2: bought.

Bayesian Optimization

Input: A black-box function f, observation data D_0,
maximum number of iteration T.
Output : The best recommendation $\mathbf{x}_T = [\mathbf{h}, \mathbf{z}]$.

1. **for** t = 0, 1, ..., T **do**
2. $\mathbf{x}_{t+1} = \mathrm{argmax}_{\mathbf{x}}\ u(\mathbf{x}|D_t)$ # u(.) : acquisition function
3. $D_{t+1} = D_t \cup \{(\mathbf{x}_{t+1}, f(\mathbf{x}_{t+1}))\}$ # Acquire next data point
4. update $u(\mathbf{x}|D_{t+1})$ # update acquisition function
5. **end for**

Issues & Challenges

One of the most crucial challenges is scaling Bayesian Optimization (BO) to high dimensions. Despite recent progress, BO is limited to moderate dimensions, works well in around 10 ~ 15 parameters. But, for example, in the real estate domain, the item's dimension will be more than 100. There are two main parts to make the BO scalable in the above algorithm. Identifying low-dimensional structure for Step2 and fast algorithm to update acquisition function for Step4 are keys.

Related Work

[Ru 2020] is similar to the above problem settings, but low-dimensional structure identification is not discussed.

There are some approaches to identify low-dimensional structures. [Wang 2013] and [Djolonga 2013] use a projection from high-dimensional space to low-dimensional subspace, model the original function over the low-dimensional space using GP, and optimize the surrogate function. (**Fig** 1)

[Rolland 2018] decomposes a high-dimensional function as a sum of lower-dimensional functions on subsets of the variables. They defined a learning algorithm of dependency graph structure among variables based on Gibbs sampling. (**Fig** 2)

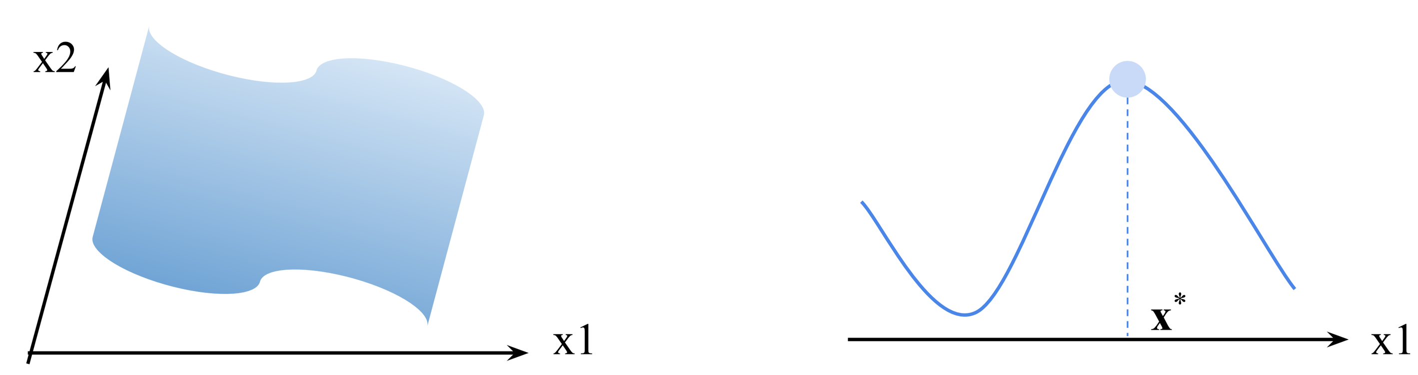

Fig 1: This function has an active dimension x1. Left is the original function over a high dimension, right is the surrogate function over a lower dimension.

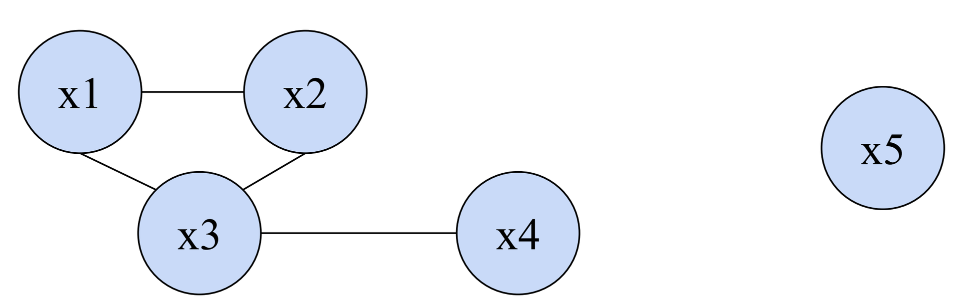

Fig 2: If variables $\mathbf{x}$ consists of x1, x2, x3, x4, x5 and its dependency graph structure is above then f($\mathbf{x}$) is decomposed to f(x1, x2, x3) + f(x3, x4) + f(x5) .

References

- [Takehara 2017] Takehara, Kazuaki, "深層強化学習による対話メディアのモデリング", オペレーションズ・リサーチ(特集 自然言語処理と数理モデル), 725-730, 2017.
- [Ru 2020] Ru, Binxin, et al. "Bayesian optimisation over multiple continuous and categorical inputs." International Conference on Machine Learning. PMLR, 2020.
- [Rolland 2018] Rolland, Paul, et al. "High-dimensional Bayesian optimization via additive models with overlapping groups." International conference on artificial intelligence and statistics. PMLR, 2018.
- [Wang 2013] Wang, Ziyu, et al. "Bayesian Optimization in High Dimensions via Random Embeddings." IJCAI. 2013.
- [Djolonga 2013] Djolonga, Josip, Andreas Krause, and Volkan Cevher. "High-dimensional gaussian process bandits." Neural Information Processing Systems. No. CONF. 2013.