Perovskite Model Description

December 7, 2016

Please add a short outline what this document provides, i.e. a description of the BO approach and the results on a first problem.

1 Single-halide Model

1.1 Model setup

Let V_x to denote the solubility of solution x. Since there are 3 cations, 3 halides and 15 solvents, $x \in \{1, ..., 135\}$. We use vector Z^x to describe the components of solution x. In particular, $Z_i^x, i \in \{1, 2, 3\}$ is a binary variable that indicate whether a cation is present in solution x. We require that $\sum_{i=1}^{3} Z_i^x = 1, \forall x$, since there is one cation present in each solution. Likewise, $Z_i^x, i \in \{4, 5, 6\}$ is a binary variable that indicate whether a halide is present in solution x. $Z_7^x \in \{s_1, ..., s_15\}$ indicates which solvent is used in solution x. For example, $Z^x = (1, 0, 0, 0, 1, 0, s_{12})$ means that solution x contains the first type of cation, second type of halide, and the 12^{th} type of solvent. In addition, for future computation, we let s_i be a 2-d vector with the first entry being the Mayer b and order and the second entry being the polarity of the i^{th} solvent, $i \in \{1, ..., 15\}$.

We assume the presence of cations and halides contributes linearly to the solubility, and we use α_i to quantify the amount contributed by cation/halide i. To correct for the possible non-linear effect of cations and halides, we introduce β_x for each of the solution x. For solvents, we assume they affect the solubility through an unknown function $f : \mathbb{R}^2 \mapsto \mathbb{R}$

which takes into consideration the Mayer bond order and polarity of a solvent. Moreover, since a perovskite solution contains more than just a cation, a halide and a solvent, we use ζ to denote the invariant part of the solution. Hence solubility of solution x can be written in the following way:

$$V_x = \sum_{i=1}^{6} \alpha_i Z_i^x + \beta_x + \zeta + f(Z_7^x). \tag{1}$$

We place the same prior distribution $N(\mu_\alpha, \sigma_\alpha^2)$ on each of α_i, and $N(0, \sigma_\beta^2)$ on β_x. ζ is assumed to follow $N(\mu_\zeta, \sigma_\zeta^2)$ We also view $f(\cdot)$ as a Gaussian process with prior mean function $\mu_0(\cdot)$ and covariance $\Sigma_0(\cdot, \cdot)$. Since ζ captures ~~the invariant part of the solution, and somewhat a fixed amount to the solubility value,~~ it is reasonable for us to assume $\mu_0(\cdot) = \mathbf{0}$. For $\Sigma_0(\cdot, \cdot)$, we use 5/2 Matern kernel. Let $r = \sqrt{\sum_{i=1}^{d} \ell_i (x_{1,i} - x_{2,i})^2}$ denote the distance between point x_1 and x_2 weighed by each dimension. In our case $d = 2$. The covariance between x_1 and x_2 under a 5/2 Matern kernel is calculated as:

$$\Sigma_0(x_1, x_2) = \sigma_m^2 \left(1 + \sqrt{5}r + \frac{1}{3}5r^2\right) exp\left(-\sqrt{5}r\right), \tag{2}$$

where σ_m^2, ℓ_i are hyper-parameters. Hence we have:

$$\begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_{135} \end{pmatrix} \sim N(\mu^0, \Sigma^0), \tag{3}$$

where

$$\begin{aligned}\mu_x^0 &= \mathbb{E}\left[\sum_{i=1}^{6}\alpha_i Z_i^x + \beta_x + \zeta + f(Z_7^x)\right] \\ &= \sum_{i=1}^{6}\mu_\alpha Z_i^x + \mu_\zeta + \sum_{i=1}^{15}\mathbb{1}(Z_7^x = s_i)\mu_{0,i} \\ &= 2\mu_\alpha + \mu_\zeta,\end{aligned}$$

and

$$\begin{aligned}\Sigma^0_{x,x'} &= COV(V_x, V_{x'}) \\ &= \sum_{i=1}^{6}\sigma_\alpha^2\mathbb{1}(Z_i^x = 1)\mathbb{1}(Z_i^{x'} = 1) + \Sigma_0(Z_7^x, Z_7^{x'}) + \mathbb{1}(x = x')\sigma_\beta^2 + \sigma_\zeta^2\end{aligned}$$

1.2 Estimation of Hyper-parameters

We estimate the hyper-parameters by maximum likelihood estimation method. Let $\theta = \{\mu_\alpha, \sigma_\alpha, \sigma_\beta, \mu_\zeta, \sigma_\zeta, \ell_1, \ell_2\}$ ~~to~~ be the vector of all the hyper-parameter we want to estimate. Let $V = \{v_{i_1}, ..., v_{i_m}\}$ ~~to~~ be the ~~solubilities we have observed~~ for solution $i_1, ..., i_m$. Then the likelihood of V given θ is

$$L(\theta) = P(V|\theta) = (2\pi)^{-m/2} * |\Sigma^0(i_1, ..., i_m)| * \exp\left(-\frac{1}{2}u^0(i_1, ..., i_m)^T\Sigma^0(i_1, ..., i_m)u^0(i_1, ..., i_m)\right), \tag{4}$$

where $u^0(i_1, ..., i_m) = (u^0_{i_1}, ..., u^0_{i_m})$ is a vertical vector, and $\Sigma^0(i_1, ..., i_m)u^0(i_1, ..., i_m)$ is a $m \times m$ matrix with $\{k, j\}$ entry $= \Sigma^0_{i_k, i_j}$.

The MLE of θ is set to be

$$\hat{\theta} = \underset{\theta}{\text{argmax}}\ log(L(\theta)). \tag{5}$$

A maximizer of Eg. (5) can be found via gradient-based optimization algorithms such as BFGS.

~~(5) can be solved numerically using algorithms such as BFGS.~~

1.3 Posterior update

We can update the mean and covariance in (3) after we observe the solubility of a solution. Let μ^n and Σ^n denote the mean and covariance ~~parameter~~ in (3) after n observations are made. We can update the posterior distribution in the following way: given μ^n and Σ^n, and a new observation $\hat{y}^{n+1}$ which is made about solution x, the new posterior is:

It is a well-known fact that we can...

$$\mu^{n+1} = \mu^n + \frac{\hat{y}^{n+1} - \mu_x^n}{\Sigma_{xx}^n}\Sigma^n \mathbf{e}_x, \tag{6}$$

$$\Sigma^{n+1} = \Sigma^n - \frac{\Sigma^n \mathbf{e}_x \mathbf{e}_x^T \Sigma^n}{\Sigma_{xx}^n} \tag{7}$$

1.4 Selection of next observation

Two methods can be used to determine on which solution x the next observation is to be made. The first method, which is named expected improvement, involves simpler computations, whereas the second method, which is called knowledge gradient, has slightly better performance in general. For now we opt for the first method.

1.4.1 expected improvement

supposing a maximization problem

Let $\hat{y}^* = \max_{1,\ldots,n}\{\hat{y_1}, ..., \hat{y_n}\}$ be the largest observed value so far. The expected improvement (EI) of a solution x is:

$$\begin{aligned} EI(x) &= \mathbb{E}\Big[\max\{V_x - \hat{y}^*, 0\}\Big|\mu^n, \Sigma^n\Big] \\ &= (\mu_x^n - \hat{y}^*)\Phi\left(\frac{\mu_x^n - \hat{y}^*}{(\Sigma_{xx}^n)^{1/2}}\right) + (\Sigma_{xx}^n)^{1/2}\phi\left(\frac{\mu_x^n - \hat{y}^*}{(\Sigma_{xx}^n)^{1/2}}\right), \end{aligned}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are standard normal cdf and pdf respectively.

We pick the solution with the largest EI:

$$x^* = \underset{x \in \{1,...,135\}}{\text{argmax}} \; EI(x) \tag{8}$$

~~**1.4.2 knowledge gradient**~~

Please write a summary of the experimental findings, in particular adding the plot for BO.

2 Mixed-Halide Model

2.1 Model setup

Let V_x to denote the solubility of solution x. For each solution, there are three halides, one cation and a solvent. Since ~~order~~ the positioning of the halides matters, there are in total 27*3*8 = 648 ~~solutions.~~ possible combinations (i.e., $x \in \{1, ..., 648\}$). We use vector Z^x to describe the components of solution x. In particular, $Z^x_i, i \in \{1, 2, 3\}$ is a binary variable that indicate which halide is in the 1st position. And ~~We~~ require that $\sum_{i=1}^{3} Z^x_i = 1, \forall x$. For example, $Z^x_{1,2,3} = (0, 0, 1)$ means that iodine (iodide) is present in the 1st position. Likewise, $Z^x_i, i \in \{4, 5, 6\}$ is a binary variable that indicates which halide is present in the second position and $Z^x_i, i \in \{7, 8, 9\}$ in indicates which halide is present in the third position. Z^x_i, $i \in \{10, 11, 12\}$, indicates which cation is present in the solution. Again it is required that $\sum_{i=10}^{12} Z^x_i = 1, \forall x$. $Z^x_{13} \in \{s_1, ..., s_8\}$ indicates which solvent is used in solution x. In addition, for future computation, we let s_i be a 2-d vector with the first entry being the Mayer bond order and the second entry being the polarity of the i^{th} solvent, $i \in \{1, ..., 8\}$.

We still assume the presence of cations and halides contributes linearly to the solubility, and we use α_i to quantify the amount contributed by cation/halide i. To correct for the possible non-linear effect of cations and halides, we introduce β_x for each of the solution x. For solvents, we assume they affect the solubility through an unknown function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ which takes into consideration the Mayer bond order and polarity of a solvent. Moreover,

since a perovskite solution contains more than just a cation, a halide and a solvent, we use ζ to denote the invariant part of the solution. Hence solubility of solution x can be written in the following way:

$$V_x = \sum_{i=1}^{12} \alpha_i Z_i^x + \beta_x + \zeta + f(Z_{13}^x). \tag{9}$$

We place the same prior distribution $N(\mu_\alpha, \sigma_\alpha^2)$ on each of α_i, and $N(0, \sigma_\beta^2)$ on β_x. ζ is assume to follow $N(\mu_\zeta, \sigma_\zeta^2)$ We also view $f(\cdot)$ as a Gaussian process with prior mean function $\mu_0(\cdot)$ and covariance $\Sigma_0(\cdot, \cdot)$. Since ζ captures the invariant part of the solution, and somewhat a fixed amount to the solubility value, it is reasonable for us to assume $\mu_0(\cdot) = \mathbf{0}$. For $\Sigma_0(\cdot, \cdot)$, we use 5/2 Matern kernel. Let $r = \sqrt{\sum_{i=1}^d \ell_i (x_{1,i} - x_{2,i})^2}$ denote the distance between point x_1 and x_2 weighed by each dimension. In our case $d = 2$. The covariance between x_1 and x_2 under a 5/2 Matern kernel which is the same as the one described in section 1.

We use the same procedure described in section 1 to place prior, to update posterior and to select the next sample.