確率的レベルセット推定

日野 英逸　先端データサイエンス研究系 教授

1 TL;DR

- レベルセット推定は得られた測定結果を用いて次の測定点を決定する適応的実験設計の一つ
- 各測定点を入力とし，対応する測定結果を出力とするブラックボックス関数を考慮し，データセットから推定された代理関数を用いて測定されていない点が閾値を超えるかどうかを予測する
- 代理関数が閾値を超える確率に基づいてレベルセット推定のための獲得関数を提案する

2 レベルセット推定

レベルセット推定 (Level set estimation: LSE) はブラックボックス関数の出力値がある閾値より大きい（または小さい）入力空間上の領域を特定することを目的とした問題であり，特にできるだけ少ない実験反復でLSEを実行するための能動学習アプローチが提案されている

例 材料の物理的特性が所望の品質を満たしていない不良領域を特定することが重要な問題．精製された材料の様々な部分でX線回折のような技術を用いて物理的特性を測定し，それらが許容下限値を超えているかどうかを判定することによって決定する

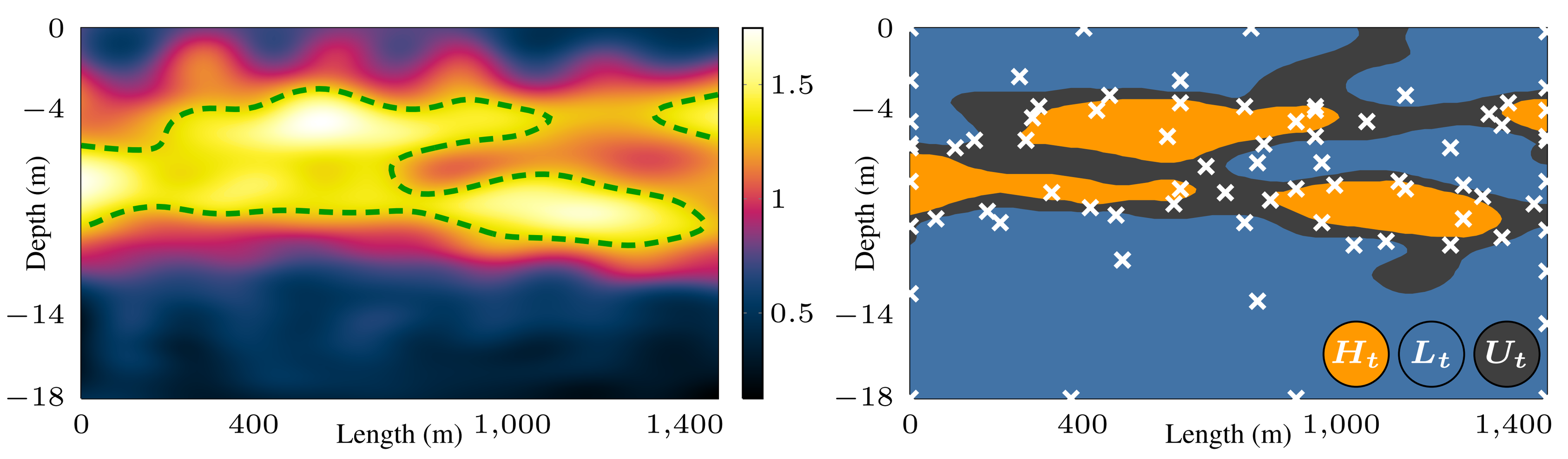

図 1: LSEの概念図

3 定式化

$\mathcal{X}$ を入力 $\boldsymbol{x}$ のドメインとし，$f : \mathcal{X} \to \mathbb{R}$ を $\boldsymbol{x}$ を入力とする未知の関数とする．LSEの目的は閾値 θ が与えられた時に可能な限り少数のデータセットから候補点集合 $X = \{\boldsymbol{x}_m\}_{m=1}^M$ に対応するそれぞれの出力 $\{f(\boldsymbol{x}_m)\}_{m=1}^M$ が θ を超えるかどうかを分類することである．

LSEでは以下の手続きを繰り返すことでこの目的を達成する．

1. 得られたデータセットからサロゲート関数 $\hat{f}$ を推定
2. サロゲート関数を用いてそれぞれの候補点を上位レベル集合 (出力が閾値を超える集合) か下位レベル集合（出力が閾値を下回る集合）のどちらかに分類
3. サロゲート関数に基づいて次の探索点を選択
4. 選択した点をオラクルに問い合わせ対応する出力を取得
5. 得られた点をデータセットに追加

サロゲート関数としてはガウス過程回帰 (GP) が用いられることが多い．今，入出力のペアの集合 $S = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$ が与えられたとする．このときGPでは関数 $\hat{f}$ は平均関数 $m(\boldsymbol{x})$，共分散関数 $k(\boldsymbol{x}, \boldsymbol{x}')$ のガウス過程に従って生成されると仮定する．GPを用いたLSEではGP事後分布に基づいて次の探索点が決められる．具体的には $p(\hat{f}|\boldsymbol{y})$ をパラメータとする獲得関数を $\alpha : \mathcal{X} \to \mathbb{R}$ とすると次の探索点は以下のように決められる．

$$\boldsymbol{x}_t = \operatorname*{argmax}_{\boldsymbol{x} \in \mathcal{X}} \alpha(\boldsymbol{x}; p(\hat{f} \mid \boldsymbol{y}))$$

代表的な獲得関数としてはStraddle, Maximum Improvement in Level-set Estimation (MILE) などがある．これらはいずれも，GP事後分布の平均 ± 標準偏差がしきい値を含むような点の中から，上位・下位集合に分類可能となる点が最も大きくなるような新規観測を決定論的に提案する．

4 確率的LSE

真の関数 f が事後分布 $p(\hat{f} \mid \boldsymbol{y})$ から生成されると仮定し，生成されたサンプルパス $\hat{f}$ に基づいて m 番目の候補点 $\boldsymbol{x}_m$ が上位レベル集合か下位レベル集合のどちらに所属するかを考える．$\boldsymbol{x}_m$ が上位レベル集合に所属するなら1，下位レベル集合に所属するなら0としたとき $\boldsymbol{x}_m$ がどちらに所属するかは以下のように表現できる．

$$z_m := z(\boldsymbol{x}_m) = \mathbb{1}_{\mathbb{R}^+}(\hat{f}(\boldsymbol{x}_m) - \theta).$$

このとき $\Phi(\cdot)$ を標準正規分布の累積分布関数とすると $z_m = 0$ である確率は以下のように表すことができる．

$$p(z_m = 0) = \int_{-\infty}^{\theta} p(\hat{f}(\boldsymbol{x}_m) \mid \boldsymbol{y}) d\hat{f}(\boldsymbol{x}_m) = \Phi\left(\frac{\theta - \mu(\boldsymbol{x}_m)}{\sigma(\boldsymbol{x}_m, \boldsymbol{x}_m)}\right) =: p_m^0.$$

同様に $p(z_m = 1) = 1 - p_m^0 =: p_m^1$ である．すなわち，$p(\hat{f} \mid \boldsymbol{y})$ に従ってサンプルパス $\hat{f}$ を生成したときに，候補点 $\boldsymbol{x}_m$ が上位レベル集合か下位レベル集合に分類される確率はベルヌーイ分布に従う．このとき $p_m^{\min} = \min\{p_m^0, p_m^1\}$ とすると $p_m^{\min}$ は候補点 $\boldsymbol{x}_m$ の分類の困難さを表す．実用上は，しきい値 θ と真の関数値 $f(\boldsymbol{x})$ 関数値が一致するような点は分類が困難であるため，マージン $\gamma > 0$ を設定して，候補点 $\boldsymbol{x}_m$ が $f(\boldsymbol{x}_m) \in \Gamma = [\theta - \gamma, \theta + \gamma)$ ならばこの点は候補点集合から取り除く．$f(\boldsymbol{x}_m) \in \Gamma = [\theta - \gamma, \theta + \gamma)$ になる事象 w_m の確率は

$$\begin{aligned} p(w_m = 1) &= \int_{\theta-\gamma}^{\theta+\gamma} p(\hat{f}(\boldsymbol{x}_m) \mid \boldsymbol{y}) d\hat{f}(\boldsymbol{x}_m) \\ &= \Phi\left(\frac{\theta + \gamma - \mu(\boldsymbol{x}_m)}{\sigma(\boldsymbol{x}_m, \boldsymbol{x}_m)}\right) - \Phi\left(\frac{\theta - \gamma - \mu(\boldsymbol{x}_m)}{\sigma(\boldsymbol{x}_m, \boldsymbol{x}_m)}\right) =: q_m^1. \end{aligned}$$

同様に $p(w_m = 0) = q_m^0 = 1 - q_m^1$．ここで $r_m^{\min} := \min\{p_m^0, p_m^1, q_m^0\}$ とし，以下のように獲得関数を定義する．

$$\boldsymbol{x}_m = \operatorname*{argmax}_{\boldsymbol{x}_m \in X} r_m^{\min}$$

これは，ベルヌーイ分布のパラメタが0.5に近い点が次の観測の候補点として選ばれるという直観的にも自然な獲得関数であり，さらに真の関数が Γ に含まれない確率 q_m^0 も評価している．

4.1 停止基準

$1 - \sum_{m=1}^M r_m^{\min} \geq \tilde{\delta}$ を満たした時にLSEを停止する．つまり全ての候補点の獲得関数の値の総和が十分に小さくなるとLSEを停止する．このとき，LSEを停止したタイミングにおいて以下の確率不等式が成り立つ．

Theorem 4.1. *Let the candidate point set be $X = \{\boldsymbol{x}_m\}_{m=1}^M$, then, the following holds for $\epsilon_m = \max\{p_m^{\min}, q_m^1\}$ and $\eta_m = \mathbb{1}_{\mathbb{R}^+}(q_m^1 - p_m^{\max})$:*

$$p\left(\forall \boldsymbol{x}_m \in X, |z_m - E[z_m]| \leq \epsilon_m, w_m \geq \eta_m\right) \geq 1 - \sum_{m=1}^M r_m^{\min}.$$

つまり $1 - \sum_{m=1}^M r_m^{\min} \geq \tilde{\delta}$ を満たす時にLSEを停止することでこの確率不等式が成り立つ確率が少なくとも $\tilde{\delta}$ 以上であることを保証できる．ここで $\tilde{\delta}$ はユーザが許容できる「停止が早すぎて候補点を誤分類してしまうか真の関数がマージンに含まれていない確率」をコントロールする，ユーザ定義の値である．

謝辞

本研究は石橋英朗氏（九工大），沓掛健太郎氏（理研AIP），松井孝太氏（名大）との共同研究です．