目的関数値の悪化を抑制するベイズ最適化に基づくオンライン学習に関する研究

1X18C066-6　中村友香
指導教員　後藤正幸

1. 研究背景・目的

近年多くの EC サイトでは，蓄積されたログデータを活用して，各ユーザの好みのアイテムを提示する推薦システムが利用されており，様々な手法が提案されている．これらの推薦システムでは，過去のログデータからユーザが好むであろうアイテムを推定し，リスト化して提示するものが多く，その性能は推薦リストの精度によって測られることが多かった．しかし，提示したアイテムに対してユーザの反応が逐次的に得られ，次の推薦に活用できるログデータには，すでに過去の推薦が影響を与えてしまっている．すなわち，推薦は一時点のみで単発的に実施されるものではなく，各ユーザに対して継続的に実施されるものである．そのため，推薦系列全体に対する累積損失によって性能を議論することが重要だと考えられる．

従来，このような逐次的な推薦と評価を取り扱うことができる枠組みとしてオンライン学習があるが，その目的は学習の効率化にあり，各時点における推薦の累積損失を考慮しない手法がほとんどである．一方，近年，目的関数の悪化を抑制しながら探索を行う手法として Safe Exploration for Optimization(以下，SafeOpt)[1] が提案された．この手法では，リスクの少ない探索を繰り返しながら，評価値が閾値以上となる連続的な入力の領域 (以下，安全領域) を効率よく求めることができる．しかし SafeOpt は，安全領域に属する入力が少なくとも 1 つ既知であることが前提であり，かつ初期入力に依存した安全領域しか求めることができない．各ユーザにとって未知のジャンルにも高い評価をつけるアイテム群は存在すると考えられるため，SafeOpt では，潜在的な高評価アイテムを見逃す可能性が高い．

そこで本研究ではガウス過程 (Gaussian Process)[2] を基とした全体探索を行う手法である Gaussian Process Upper Confidence Bound(以下，GP-UCB)[2] を導入することで SafeOpt を拡張し，広い探索領域に対して安全領域を推定できる手法を提案する．これにより，ユーザが好まないアイテムを極力提示することなく，真の好みを効率よく捉えた推薦に導くことが可能となる．最後に，探索されたアイテムに対して逐次的にユーザの評価が得られることを想定した人工データを生成し実験を行い，提案手法の有効性を示す．

2. 準備

2.1. 問題設定

入力を $\boldsymbol{x} \in \mathbb{R}^M$，出力を $y \in \mathbb{R}$ としたとき，現時点で得られている n 個のデータ集合を $(\boldsymbol{X}^n, \boldsymbol{Y}^n) = \{(\boldsymbol{x}_t, y_t)\}_{t=1}^n$，$t$ 番目のデータの特徴量を $\boldsymbol{x}_t = (x_{t1}, \cdots, x_{tM})^\top$ とする．$\boldsymbol{x}$ と y に関数関係 $y = f(\boldsymbol{x})$ が成り立つと仮定したとき，効率よく関数の推定値 $y = \hat{f}(\boldsymbol{x})$ を得ることが一般的な機械学習である．本研究では，n 個のデータ集合を $(\boldsymbol{X}^n, \boldsymbol{Y}^n)$ と関数の推定値 $y = \hat{f}(\boldsymbol{x})$ から得られる統計量を用いて，探索対象集合 $\mathcal{D}$ から，推定値 $\hat{y}$ に何らかの目標値を設定し，それに合致した次の探索点となる入力 $\boldsymbol{x}$ を決定する問題を扱う．

2.2. ガウス過程回帰

ガウス過程回帰 [2](Gaussian Process Regression, 以下 GPR) は，入力 $\boldsymbol{x}$ と出力 y の関係性を表す関数 $y = f(\boldsymbol{x})$ を推定する手法の 1 つである．GPR の最も重要な特徴は，入力の集合 $\boldsymbol{X}^n = \{\boldsymbol{x}_t\}_{t=1}^n$ に対応する出力 $\boldsymbol{Y}^n = \{y_t\}_{t=1}^n$ が多変量ガウス分布に従うと仮定することで，関数の不確実性を表現できることである．GPR を用いることで，入力 $\boldsymbol{x}$ に対応する出力 y の推定幅を導出することができる．推定幅の大きさは，対応する入力周辺の不確実性を表しており，推定幅が大きい入力 $\boldsymbol{x}$ を探索することで，効率よく関数の不確実性を小さくすることができる．SafeOpt および GP-UCB は，このような考え方を用いた手法である．

2.3. SafeOpt[1]

SafeOpt[1] は安全領域を効率よく推定することができる手法である．GPR を用いて求めた安全領域 $S_t \subseteq \mathcal{D}$ と，安全領域を増やすための集合 $G_t \subseteq S_t$ を各ステップで定義し，その中で探索を行う．具体的なアルゴリズムを以下に示す．

Step.A1 $t = 0$ とする．

Step.A2 $t = 0$ であれば初期入力 S_0，$t \neq 0$ であれば S_t を導出する．

$$S_t = \bigcup_{\boldsymbol{x} \in S_{t-1}} \{\boldsymbol{x}' \in \mathcal{D} \mid l_t(\boldsymbol{x}) - Ld(\boldsymbol{x}, \boldsymbol{x}') \geq h\} \quad (1)$$

Step.A3 G_t を導出する．G_t が空集合であればアルゴリズム終了．

$$g_t(\boldsymbol{x}) := \|\boldsymbol{x}' \in \mathcal{D} \backslash S_t \mid u_t(\boldsymbol{x}) - Ld(\boldsymbol{x}, \boldsymbol{x}') \geq h\| \quad (2)$$

$$G_t = \{\boldsymbol{x} \in S_t \mid g_t(\boldsymbol{x}) > 0\} \quad (3)$$

Step.A4 下記の基準で $\boldsymbol{x}_t$ を探索し，対応する y_t を得て GPR を更新する．

$$\boldsymbol{x}_t = \underset{\boldsymbol{x} \in G_t}{\operatorname{argmax}} \{u_t(\boldsymbol{x}) - l_t(\boldsymbol{x})\} \quad (4)$$

Step.A5 $t = t + 1$ とし，Step.A2 に戻る．

ただし，安全領域 S_t を定義する閾値を h とし，閾値を上回る任意の初期入力を S_0 とする．また，L はリプシッツ連続に用いられる定数，$l_t(\boldsymbol{x})$，$u_t(\boldsymbol{x})$ はそれぞれ y_t の推定幅の下限値と上限値であり，これらは GPR から算出される．また，$\boldsymbol{x}'$ は任意の探索対象の点を表し，$d(\boldsymbol{x}, \boldsymbol{x}')$ は $\boldsymbol{x}$ と $\boldsymbol{x}'$ のユークリッド距離である．

2.4. GP-UCB[2]

GP-UCB は，未知の関数における大域的最適解を少ない試行回数で求める手法の一つである．具体的には，式 (5) の基準で探索対象 $\boldsymbol{x}$ を定める．

$$\boldsymbol{x}_t = \underset{\boldsymbol{x}}{\operatorname{argmax}} \left\{\mu_{t-1}(\boldsymbol{x}) + \sqrt{\beta}\sigma_{t-1}(\boldsymbol{x})\right\} \quad (5)$$

ここで，$\mu_{t-1}(\boldsymbol{x}), \sigma_{t-1}(\boldsymbol{x})$ は GPR から算出される $\boldsymbol{x}$ の期待値と標準偏差であり，β は探索の度合いを決めるハイパーパラメータである．β の値が小さいと $\mu_{t-1}(\boldsymbol{x})$ が支配的となり，現状の最適解に最も近い入力が，大きいと $\sigma_{t-1}(\boldsymbol{x})$ が支配的となり，不確実性が大きい入力が選択される．β の値を調節することで効率よく最適解を探索することが可能となる．

3. 提案手法

3.1. 着想

SafeOpt は，安全領域に属する初期入力から連続的に安全領域を広げていく手法である．しかし，存在する全ての安全領域は必ずしも連続的ではないため，初期値に依存した一部の安全領域しか求めることができない．一方，GP-UCB は関数を全体的に探索するものの，あくまで最適解を見つけることに主眼を置いた手法であり，安全領域を求めることはできない．そこで本研究では，初期入力に依存せずに全体の安全領域を求めることが可能な手法を提案する．

3.2. 提案アルゴリズム

本研究では，GP-UCB を用いて SafeOpt のアルゴリズムを拡張し，関数全体の安全領域を算出するアルゴリズムを提案する．具体的なアルゴリズムを以下に示す．

Step.B1 $i = 0$ とし，任意の S_0 を定義する．

Step.B2 SafeOpt をアルゴリズムが停止するまで用いて安全領域 S_{T_i} を得る．

Step.B3 $\mathcal{D} = \mathcal{D} \backslash S_{T_i}$，$t = t + 1$ として，GP-UCB の探索基準を用いて $\boldsymbol{x}_t$ の探索を行う．

Step.B4 $t = T' + T_i$ であればアルゴリズム終了．

Step.B5 $y_t \leq h$ であれば Step.B3 に戻る．さもなくば，$i = i + 1$，$S_0 = \{\boldsymbol{x}_t\}$ として Step.B2 に戻る．

ここで，T' は安全領域外の探索回数を制限するための規定回数，T_i は i 個目の安全領域を推定する際に要した探索回数である．

4. 評価実験

4.1. 実験条件

本実験では，学習が行われるオンラインシステムを想定し，テストユーザが好むアイテム群を捉えられるか否かの評価を行う．対象のログデータには，探索したアイテムに対して逐次的にユーザの評価が得られると仮定した人工データを生成し使用する．人工データ生成においては，アイテムが 5 つのジャンルに分類でき，各ユーザはそのうち 2 つのジャンルを好むものとした．評価値は 5 段階で，好みのジャンルのアイテムには評価値 3〜5，それ以外のジャンルのアイテムには 1〜3 をつけると仮定し，評価のランダム性を加味している．ここで，評価値 4 以上をユーザが好むアイテムと定義し，好みのジャンルではなく，個人に依存する好みを得ることを目的とできるようにした．全アイテム数を 500 個とし，各アイテムの特徴量は各アイテムに対する全ユーザ (1000 人) の評価値の人工データに NMF[3] を適用して得た．本研究では，安全領域を決定する閾値 h を 3.5，初期入力を評価値 5 のアイテムとした．また提案手法で用いる GP-UCB では $\beta = 9$ とし，$T' = 10$ とした．比較手法は，従来手法の SafeOpt と GP-UCB とし，GP-UCB は $\beta = 3$ と $\beta = 9$ を用いて実験を行った．また，本実験では，総探索回数を 50 と設定した．本研究では，安全領域に属する点のうち探索過程において安全であると判断した点の割合 (以降，安全領域のカバー率)，安全であると判断した点 (以下，安全点) の数に対する閾値を下回る点を探索した回数の比率，安全領域のカバー効率の 3 つの評価指標を用い，有用性を示す．安全領域のカバー効率は，図 1 のように縦軸を安全領域のカバー率，横軸を探索回数とし，全体面積を 1 としたときの関数以下の面積を表している．この指標値が高いと効率よく安全領域を探索できていることを表す．比較手法の GP-UCB においては，安全領域を求める手法ではないため，安全点は SafeOpt の定義式を利用して求めた．

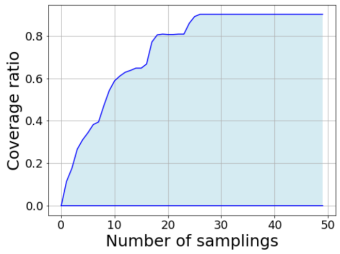

図 1: 安全領域のカバー効率 (網掛け部)

4.2. 結果と考察

表 1 に，各手法による評価指標値を示す．

表 1: 各手法による探索過程の評価

	カバー効率	カバー率	安全でない探索回数/安全点の数
SafeOpt	0.440	0.510	**0/78**
GP-UCB($\beta = 3$)	0.217	0.229	**0/35**
GP-UCB($\beta = 9$)	0.252	0.261	**0/40**
提案手法	**0.623**	**1.000**	4/153

表 1 より，提案手法が最も効率的かつ，精度よく全体の安全領域を求めていることが分かる．また，安全でない点の探索比率は増加するものの，得られた安全点の数を考慮するとリスクは最小限に抑えられたと言える．GP-UCB のカバー率は低い値となったが，これは評価値が最大である 5 をとるアイテムが複数あることから，収束が早かったためであると考えられる．以上よりログデータの性質上，提案手法がより有効であることが示された．

5. まとめと今後の課題

本研究では，最小限のリスクで全体的な安全領域を効率よく探索するアルゴリズムを提案し，人工データを用いた実験により手法の有効性を示した．今後の課題として，性別や年齢などのユーザ情報の考慮等が挙げられる．

参考文献

[1] Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *International Conference on Machine Learning*, pp. 997–1005. PMLR, 2015.

[2] 持橋大地，大羽成征． ガウス過程と機械学習． 講談社，2019.

[3] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, pp. 788–791, 1999.