

# 物理モデルと言葉による観測を統合した状態推定

三好優輝<sup>1</sup> 井上正樹<sup>1</sup> 藤本悠介<sup>2</sup><sup>1</sup> 慶應義塾大学 <sup>2</sup> 大阪大学

{yukiki0423,minoue.z6}@keio.jp fujimoto.yusuke.es@osaka-u.ac.jp

## 概要

本稿では、人由来のテキスト形式の観測情報を用いる状態推定法 Language-Aided Particle Filter (LAPF) を整理し、複数の人センサを想定した動的配置を見据えた拡張課題について述べる。さらに、観測精度の異なる3人の人間をセンサとして用いた河川の水位推定実験を行ない、人の配置により推定誤差が変化することを確認した。

## 1 背景・目的

SNS やチャットサービスの普及により、自然言語形式のデータを容易に取得することが可能になっている。このようなテキストデータには、人間による要望や評価などの情報が多く含まれている。このような多様な情報を含むテキストデータを活用することで、物理システムの状態に関する情報を得ることができるのではないだろうか。

筆者らの先行研究 [1] において、テキストを活用した状態推定手法として Language-Aided Particle Filter (以降, LAPF) を提案した。状態推定は、推定対象の状態が従うシステムのモデルと、センサによる観測情報を元に行なわれ、一般的にはセンサとして電氣的・機械的な素子を用いて数値化する物理センサが使用される。これに対して、LAPF では、人間を自然言語形式の観測情報を返す人センサとして扱い、観測情報が非構造化データのテキストである場合にパーティクルフィルタ [2] を実行する方法を考案した。本稿では、LAPF を改めてまとめなおし、さらなる展開をはかることを目的とする。

文献 [1] では、人センサの観測位置は時不変という単純な状況を考えていた。しかし、人センサの特徴は、物理センサより観測精度が低いものの、観測位置を柔軟に変えることができるという点であり、文献 [1] で考えた状況はその特徴をまだ充分には活用できていない。そこで、本研究では複数の人センサを用いて彼らの観測位置も動的に変化させながら

状態推定を行なう問題に取り組む。

本研究では、人センサの動的な配置を含む状態推定法を開発するため、二つの問題に取り組む。一つは各人センサの観測精度などの特徴をもとに彼らの最適な配置をおこなう方法を開発することである。もう一つは LAPF をセンサ位置の変更を許容できるように拡張することである。前者の最適センサ配置の準備として、人センサの個人差を観測精度という形で見積もる方法にも取り組むことになる。

本稿の構成は、2 節で LAPF について述べ、3 節で人センサ配置問題のための準備実験について述べ、4 節で本稿をまとめる。

**表記:** 実数の集合は  $\mathbb{R}$ 、テキストの集合は  $\mathbb{T}$ 、空集合は  $\emptyset$  と表す。ここで、本稿におけるテキストは、単一および複数のトークンの列のことであり、テキストの集合  $\mathbb{T}$  とは「NLP2026 は宇都宮で開催される」のようなテキストを要素にもつ集合である。確率分布  $p(x)$  からサンプリングした値  $x$  は、 $x \sim p(x)$  と表す。時刻  $T$  までの時系列データ  $\{x_1, x_2, \dots, x_T\}$  は  $x_{1:T}$  と表す。

## 2 Language-Aided Particle Filter

本節では、文献 [1] で提案した LAPF について簡単にまとめる。

### 2.1 問題設定

本小節では、LAPF で考えている問題設定について述べる。LAPF は、図 1 に示すような、人センサ  $S_H$  を用いた状態推定を行なうための手法である。物理センサを用いた一般的な状態推定との違いは、センサによって得られる観測情報がテキストデータであり、構造化されていないデータである点である。LAPF では、人センサ  $S_H$  から得られる観測テキスト  $s_k$  を元に、状態  $x_k$  がどの値である確率が高いかという状態推定分布  $\pi(x_k)$  を計算することで状態推定を行なう。最後に、LAPF で考える問題設定を問題 1 にまとめる。

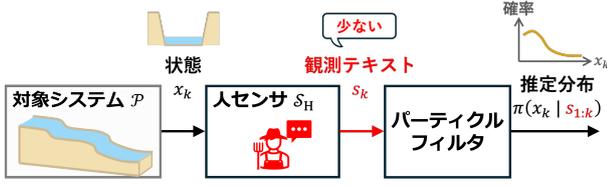


図1 人センサを用いた状態推定

**問題 1.** 対象システム  $\mathcal{P}$  と人センサ  $S_H$  のモデルが既知とする。そして、初期時刻の状態推定分布が  $\pi(x_0)$ 、時刻  $k$  までの観測テキストの時系列データ  $s_{1:k}$  が与えられた元で、時刻  $k$  における状態推定分布  $\pi(x_k | s_{1:k})$  を計算せよ。

## 2.2 対象システムと人センサのモデル

まず、対象システム  $\mathcal{P}$  のモデルについて述べる。対象システム  $\mathcal{P}$  として、次式の離散時間の状態方程式を考える。

$$\mathcal{P}: x_k = f(x_{k-1}, w_k), \quad \forall k \geq 1 \quad (1)$$

ここで、 $x_k \in \mathbb{R}^n$  は時刻  $k$  における状態、 $w_k \in \mathbb{R}^\ell$  は時刻  $k$  におけるシステム雑音、関数  $f: \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^n$  は、状態  $x_{k-1}$  とシステム雑音  $w_k$  から次時刻  $k$  の状態  $x_k$  を出力する状態遷移関数である。また、システム雑音  $w_k$  は特定の確率分布  $\mathcal{W}$  に従う乱数とする。

次に、人センサ  $S_H$  のモデルについて述べる。ここでは説明のため、観測者は一人とし、その観測者が報告するテキスト  $s_k$  も一つ、つまり  $s_k \in \mathbb{T}$  という単純な状況で考える。そして、LAPF では、人センサ  $S_H$  のモデルとして、認知器  $C$  と表現器  $E$  のカスケード構造を考える。これは、観測者はまず状態  $x_k$  を認知することによって脳内で認知値  $y_{H,k} \in \mathbb{R}$  を想定し、その認知値  $y_{H,k}$  を表現することで観測テキスト  $s_k$  が出力されていると仮定したモデルである。以下では、認知器  $C$  と表現器  $E$  のモデルについて述べる。

認知器  $C$  は、次式の観測方程式で記述する。

$$C: y_{H,k} = h_H(x_k, v_{H,k}), \quad \forall k \geq 1 \quad (2)$$

ここで、 $v_{H,k} \in \mathbb{R}^r$  は観測者の認知上の観測雑音、関数  $h_H: \mathbb{R}^n \times \mathbb{R}^r \rightarrow \mathbb{R}$  は、状態  $x_k$  と認知雑音  $v_{H,k}$  から認知値  $y_{H,k}$  を出力する観測関数である。また、認知雑音  $v_{H,k}$  は特定の確率分布  $\mathcal{V}_H$  に従う乱数とする。

表現器  $E$  は、量子化部と言語化部で構成されると考える。このように考えられる理由は、観測者によ

る言語表現能力には限界があり、認知値  $y_{H,k}$  に対して一対一で対応するほど細かく言語化することはできないためである [3, 4]。そのため、LAPF では、観測者が認知値  $y_{H,k}$  を表現する際、 $y_{H,k}$  から直接言語化を行なうのではなく、 $y_{H,k}$  を  $m$  段階の値に量子化した量子ラベル  $q_k \in \{1, \dots, m\}$  から言語化を行なっていると考える。したがって、表現器  $E$  は次式で記述する。

$$E: \begin{cases} q_k = Q_m(y_{H,k}), & \forall k \geq 1 & (3a) \\ s_k \sim \text{Label2Prob}(q_k), & \forall k \geq 1 & (3b) \end{cases}$$

ここで、関数  $Q_m: \mathbb{R} \rightarrow \{1, \dots, m\}$  は、認知値  $y_{H,k}$  を  $m$  段階の量子ラベル  $q_k$  に量子化する関数である。また、関数  $\text{Label2Prob}$  は、量子ラベル  $q_k$  に対応した観測テキスト  $s_k$  の確率分布を生成する関数である。ただし、この関数  $\text{Label2Prob}$  は、人間の言語化を数式で記述するために導入した抽象的な関数である。(3b) 式では、観測テキスト  $s_k$  の生成を、関数  $\text{Label2Prob}$  で得られる確率分布からサンプリングすることで表現している。

最後に、量子化関数  $Q_m$  の定義を述べる。認知値  $y_{H,k}$  の取り得る範囲を  $\Lambda = [y_{H,\min}, y_{H,\max}]$  とする。また、 $\Lambda$  上の  $m$  個の量子化区間  $\{\Lambda_i\}_{i=1}^m$  は、次式を満たすとする。

$$\bigcup_{i=1}^m \Lambda_i = \Lambda \quad (4a)$$

$$\Lambda_i \cap \Lambda_j = \emptyset, \quad \forall i \neq j \quad (4b)$$

このとき、量子化関数  $Q_m$  を次式で定義する。

$$Q_m(y_{H,k}) := i, \quad \text{if } y_{H,k} \in \Lambda_i \quad (5)$$

## 2.3 LAPF のアルゴリズム

本小節では、問題 1 を解くためのアルゴリズムである LAPF について簡単に説明する。LAPF とは、状態推定分布  $\pi(x_k | s_{1:k})$  の計算をパーティクルフィルタ [2] の理論に基づいて行なう手法であり、その計算において肝になるのは尤度  $p(s_k | x_k)$  の計算方法である。尤度  $p(s_k | x_k)$  とは、状態が  $x_k$  のときに観測情報として  $s_k$  が得られる確率であり、状態  $x_k$  に対する観測情報  $s_k$  の尤もらしさである。一般的なパーティクルフィルタでは、観測情報は数値やベクトルなどの構造化されたデータであるのに対し、 $s_k$  はテキストという構造化されていないデータであるため、尤度  $p(s_k | x_k)$  を計算するためには工夫が必要である。以下では、LAPF における尤度  $p(s_k | x_k)$  の計算方法を簡単にまとめる。

---

**Algorithm 1** Language-Aided Particle Filter
 

---

- 1: **Initialization:**
  - 2: Sample particles  $x_{0|0}^i \sim \pi(x_0)$  for  $i = 1, \dots, N_p$
  - 3: Set initial weights  $\alpha_0^i = 1/N_p$
  - 4: **for**  $k = 1$  to  $T$  **do**
  - 5:  $x_{k|k-1}^i \sim p(x_k | x_{k-1|k-1}^i)$
  - 6:  $p(s_k | x_{k|k-1}^i) \leftarrow \sum_{j=1}^m p(q_k^j | s_k) p(q_k^j | x_{k|k-1}^i)$
  - 7:  $\alpha_k^i \leftarrow p(s_k | x_{k|k-1}^i)$
  - 8:  $\alpha_k^i \leftarrow \alpha_k^i / \sum_{j=1}^{N_p} \alpha_k^j$
  - 9:  $\tilde{\pi}(x_k | s_{1:k}) = \sum_{i=1}^{N_p} \alpha_k^i \delta(x_k - x_{k|k-1}^i)$
  - 10: Resample particles  $x_{k|k}^i \sim \tilde{\pi}(x_k | s_{1:k})$
  - 11: **end for**
- 

準備として、量子ラベル  $q_k$  の確率分布に対して次の仮定をおく。

**仮定 1.** 量子ラベル  $q_k$  の確率分布  $p(q_k)$  は次式で記述できると仮定する。

$$p(q_k) = \frac{1}{m}, \quad \forall q_k \in \{1, \dots, m\} \quad (6)$$

この仮定は、量子ラベル  $q_k$  の事前分布を一様分布とするものであり、事前情報がない状況における  $q_k$  の確率分布を一様分布にするという妥当な仮定である。ここで、尤度  $p(s_k|x_k)$  について次の命題が成り立つ。

**命題 1.** 仮定 1 の元で、次式が成り立つ。

$$p(s_k|x_k) \propto \sum_{q_k=1}^m p(q_k|s_k) p(q_k|x_k) \quad (7)$$

**証明** 証明は省略する。詳細は文献 [1] へ。 □

**命題 1** より、尤度  $p(s_k|x_k)$  は、 $p(q_k|s_k)$  と  $p(q_k|x_k)$  の計算ができれば求めることができる。 $p(q_k|s_k)$  は言語モデルを利用することで計算でき、 $p(q_k|x_k)$  は次の命題に示す式で計算できる。

**命題 2.** 状態が  $x_k$  である条件のもとで量子ラベルが  $q_k$  である確率  $p(q_k | x_k)$  は次式で記述できる。

$$p(q_k | x_k) = \int_{\Lambda_{q_k}} p(y_{H,k} | x_k) dy_{H,k} \quad (8)$$

ただし、 $\Lambda_{q_k}$  は  $i = q_k$  の場合の  $\Lambda_i$  であり、量子ラベル  $q_k$  に対応する量子化区間である。

**証明** 証明は省略する。詳細は文献 [1] へ。 □

最後に、LAPF のアルゴリズムを **Algorithm 1** に示す。ここで、**Algorithm 1** に現れる  $p(x_k | x_{k-1|k-1})$  は状態遷移モデルであり、(1) 式の状態方程式に対応している。

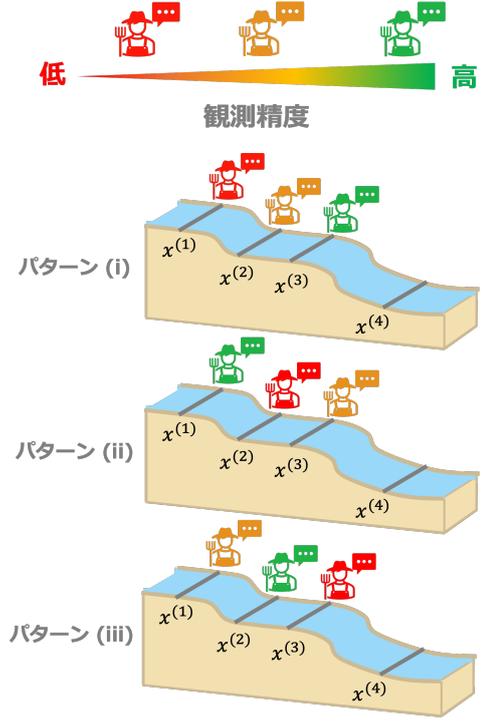


図 2 本実験の概要

### 3 数値実験

本節では、人センサ配置問題に取り組むための準備の実験として、複数の個体差のある人センサを用いて状態推定を行ない、人センサの配置を変化させた場合に状態推定の精度が変化することを確認する。具体的には、図 2 に示したように、観測精度の異なる 3 つの人センサを用いて河川の水位の推定を行ない、人センサの配置パターンを変えて推定誤差の変化を確認する。

#### 3.1 実験設定

対象システム  $\mathcal{P}$  として、次式の河川の水量伝播モデルを考えた。

$$\mathcal{P} : x_k = \text{proj}_{[0,5]}(Ax_{k-1} + w_k), \quad \forall k \geq 1 \quad (9)$$

ここで、状態  $x_k = [x_k^{(1)} \dots x_k^{(4)}]^\top$  は 4 つの地点の水位、関数  $\text{proj}_{[a,b]} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  は、ベクトルの各成分に対して区間  $[a, b]$  への直交射影を行なう関数である。したがって、(9) 式は各地点の最高水位が 5 である河川の水位の時間発展を記述したものである。

表1 各パターンにおける認知行列の数値設定

パターン	$C_H^{(1)}$	$C_H^{(2)}$	$C_H^{(3)}$
(i)	[1 0 0 0]	[0 1 0 0]	[0 0 1 0]
(ii)	[0 1 0 0]	[0 0 1 0]	[1 0 0 0]
(iii)	[0 0 1 0]	[1 0 0 0]	[0 1 0 0]

システム行列  $A$  の数値を次式に示す.

$$A = \begin{bmatrix} 0.4 & 0 & 0 & 0 \\ 0.6 & 0.3 & 0 & 0 \\ 0 & 0.7 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.4 \end{bmatrix} \quad (10)$$

システム雑音  $w_k$  は, 平均が  $u_k = [1 \ 0 \ 0 \ 0]^\top$ , 共分散行列が  $0.1I_4$  とした正規分布  $\mathcal{N}(u_k, 0.1I_4)$  に従うとした. また, 各地点の初期水位  $x_0^{(i)}$  は 2.5 とした.

本実験では 3 つの人センサ  $s_H$  により状態推定を行なった.  $j$  人目の人センサ  $s_H^{(j)}$  のモデルとして, 次式を考えた.

$$s_H^{(j)} : \begin{cases} y_{H,k}^{(j)} = \text{proj}_{[0,5]} \left( C_H^{(j)} x_k + v_{H,k}^{(j)} \right), & \forall k \geq 1 \quad (11a) \\ q_k^{(j)} = Q_5 \left( y_{H,k}^{(j)} \right), & \forall k \geq 1 \quad (11b) \\ s_k^{(j)} \sim \text{Label2Prob} \left( q_k^{(j)} \right), & \forall k \geq 1 \quad (11c) \end{cases}$$

(11a) 式において直交射影関数  $\text{proj}_{[0,5]}$  を使っている理由は, 河川の様子を観測した人間がその水位の下限・上限を越えて認知することはないと考えられるためである. 認知行列  $C_H^{(j)} \in \mathbb{R}^{1 \times 4}$  は one-hot な横ベクトルであり,  $j$  人目の人センサがどの状態  $x_k^{(i)}$  を観測するかを表現している. 本実験では 3 つの人センサの配置パターンとして, 表 1 に示す 3 つのパターンを考えた. 認知雑音  $v_{H,k}^{(j)} \in \mathbb{R}$  は, 正規分布  $\mathcal{N}(0, r_H^{(j)})$  に従うとし, 分散は  $r_H^{(1)} = 10, r_H^{(2)} = 5, r_H^{(3)} = 1$  とした. つまり, 各人センサの観測精度は, 1 人目が最も低く, 3 人目が最も高くなるようにした. 量子化関数  $Q_5$  は, 区間  $\Lambda = [0, 5]$  を 5 等分した量子化区間  $\{\Lambda_i\}_{i=1}^5$  による量子化関数である. (11c) 式における関数  $\text{Label2Prob}$  は, 量子ラベル  $q_k$  に対応する観測テキスト  $s_k$  の確率分布を出力する関数である. (11c) 式の具体的な実装は, 収集した認知値と観測テキストの  $N_d$  組のデータセット  $\{y_H^i, s^i\}_{i=1}^{N_d}$  から, 量子ラベルと観測テキストの  $N_d$  組のデータセット  $\{q^i, s^i\}_{i=1}^{N_d}$  を構築し, 入力量子ラベル  $q_k$  に対応する観測テキストデータ  $s^i$  の中からランダムに一つ選択することにより行なった.

本実験で使用したデータセット  $\{y_H^i, s^i\}_{i=1}^{N_d}$  や, 確率  $p(q_k^i | s_k^i)$  を計算するための言語モデルは, 文献 [1] と同様にした.

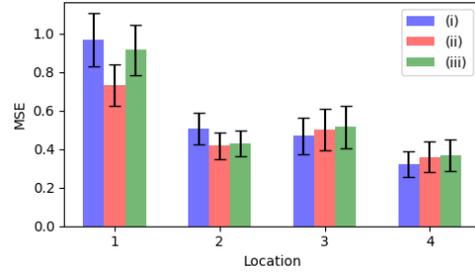


図3 人センサの各配置パターンにおける推定誤差の比較

## 3.2 実験結果

初期時刻  $k = 0$  における近似事後分布  $\tilde{\pi}(x_0)$  は, 平均が初期状態の真値  $x_0$ , 共分散行列が単位行列  $I_4$  の正規分布  $\mathcal{N}(x_0, I_4)$  とした. また, 総時間ステップ数は 100, パーティクルの数  $N_p$  は 1000 とした. 以上の設定における状態推定を 1000 回試行した.

得られた近似事後推定分布  $\tilde{\pi}(x_k | s_{1:k}^{(1)}, s_{1:k}^{(2)}, s_{1:k}^{(3)})$  の平均値を推定値  $\hat{x}_k$  とし, 各試行・各時刻において推定値  $\hat{x}_k$  と真値  $x_k$  の二乗誤差を計算した. そして, その二乗誤差の全試行・全時刻にわたって平均を取り, 各状態  $x_k^{(i)}$  における平均二乗誤差 (Mean Squared Errors; 以下, MSE) は図 3 のようになった. 人センサの配置パターンによって, 各状態の MSE の大きさが変化することが確認できた. 特に, 最も観測精度が高い 3 人目の人センサが観測した状態は, パターン (i) では  $x_k^{(3)}$ , パターン (ii) では  $x_k^{(1)}$ , パターン (iii) では  $x_k^{(2)}$  であり, 各パターンにおいてその状態の MSE が小さくなった. そして, 全状態にわたる MSE の平均値は, パターン (i) では  $0.57 \pm 0.08$ , パターン (ii) では  $0.50 \pm 0.07$ , パターン (iii) では  $0.56 \pm 0.08$  となり, パターン (ii) が最小となった. また, 人センサを配置しなかった状態  $x_k^{(4)}$  の MSE に注目すると, パターン (i) が最小となった. したがって, 3 つの配置パターンの中で最適なセンサ配置は, 全状態の MSE を平均的に低くしたい場合はパターン (ii), 人センサを配置できていない状態  $x_k^{(4)}$  の MSE を最小にしたい場合はパターン (i) となり, 目的に応じて最適なセンサ配置が変わることが確認できた.

## 4 まとめ

本稿では LAPF の枠組みを再整理し, 人センサの動的配置を含む状態推定へ向けた課題を示した. 今後は, 各人センサの観測精度の推定と, それを考慮した配置最適化および位置可変 LAPF の設計に取り組む.

## 謝辞

本研究は JSPS 科研費基盤研究 (B) 25K01254 の助成を受けたものです。

## 参考文献

- [1] Yuki Miyoshi, Masaki Inoue, and Yusuke Fujimoto. Language-aided state estimation, 2025.
- [2] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In **IEE Proceedings F (Radar and Signal Processing)**, Vol. 140, pp. 107–113. IET, 1993.
- [3] RS Ogden, FR Simmons, and JH Wearden. Verbal estimation of the magnitude of time, number, and length. **Psychological Research**, Vol. 85, No. 8, pp. 3048–3060, 2021.
- [4] W Todd Maddox and F Gregory Ashby. Comparing decision bound and exemplar models of categorization. **Perception & psychophysics**, Vol. 53, No. 1, pp. 49–70, 1993.