

能動推論を用いた IoT センサ情報収集と実世界の状態推定を同時に解決する偽情報判定手法の提案と評価

大阪大学 基礎工学部 情報科学科

村田研究室 前野仁愛

特別研究報告会 2025/2/13

● 偽情報問題の深刻化

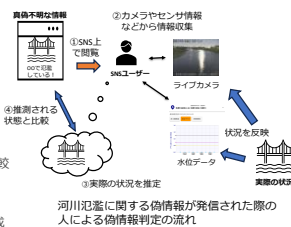
- 災害やパンデミックでの偽情報の拡散
 - 例: 2024年8月の台風10号の際の偽情報
 - 実際には発生していないにもかかわらず、「多摩川氾濫」と書かれた偽情報が拡散

● 人間による偽情報判定

- 周辺のカメラやセンサ等の情報を収集
- 収集した情報をもとに推測される状況と比較

● 人間のみによる偽情報判定の限界

- 近年では生成AIによって大量の偽情報が生成
- 人間による判断では偏見や思い込みが混入



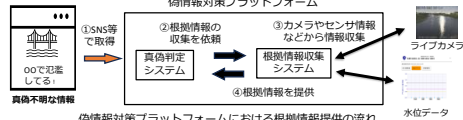
偽情報判定システム

● 偽情報対策プラットフォーム^[1]

- 真偽不明な情報に対し、真偽の判定結果とその理由説明を提供する
- 偽情報の検知、根拠収集・統合管理、総合的な分析、社会的な影響度評価の4つによって構成

● 根拠情報収集システム

- センサやカメラ等の情報源にアクセスし、真偽判定のための根拠情報を提供
偽情報対策プラットフォーム



[1]富士通株式会社, "富士通と産学組織が9 府で共創し、世界初の偽情報対策プラットフォームの構築を開始," <https://jp.fujitsu.com/jp/news/2024/10/16.html>, 2024. (Accessed: 2025-01-21)

研究目的・アプローチ

● 研究目的

- 偽情報対策プラットフォームのための根拠情報収集手法の考案

● アプローチ

- 真偽判定の対象として河川氾濫に関する情報をシナリオとして想定
 - 特定の地点が浸水しているとする真偽不明な情報が発信された場合
 - 特定の地点の浸水の有無に関する確率を根拠情報として提供する
- 能動推論を用いることで真偽判定のための根拠情報収集における各要素をモデル化する
 - カメラやセンサ等からの情報による環境の推測
 - 状況を確認するために重要なカメラやセンサ類の選択

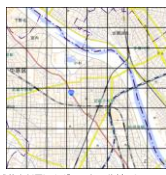
想定するシナリオの設定-1

● 武蔵小杉周辺の地域で真偽不明な情報が発信された状況を想定

- 2024年の台風10号における偽情報を題材にする

● センサ等の情報源や浸水状況は単純化して表現する

- 浸水状況の設定
 - 地域内をグリッドで分割し、各マス内の浸水の有無として表現
 - 時間経過によって変化しないと仮定
- 情報源の設定
 - 真偽判定の対象の地点を除くすべてのマスに存在
 - アクセスによって、そのマスでの浸水の有無を提供
 - 提供する情報には誤情報は含まれていない



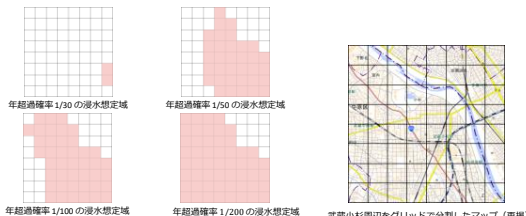
武蔵小杉周辺をグリッドで分割したマップ
国土地理院の公開する電子国土基本図※を加工して作成

[2]国土地理院, "地理院タイル一覧," <https://maps.gsi.go.jp/development/ichiran.html>, (Accessed: 2025-02-03).

想定するシナリオの設定-2

● システムは事前に氾濫の起こりうる確率を与えられる

- 国土交通省の公開している多段階の浸水想定図^[3]をもとに設定



年超過確率 1/30 の浸水想定域
年超過確率 1/50 の浸水想定域
年超過確率 1/100 の浸水想定域
年超過確率 1/200 の浸水想定域
武蔵小杉周辺をグリッドで分割したマップ (再掲)

[3]国土交通省 関東地方整備局, "流域治水プロジェクトー 京浜河川事務所," <https://www.ktr.mlit.go.jp/keihin/keihin/index136.html>, (Accessed: 2025-01-26)

能動推論

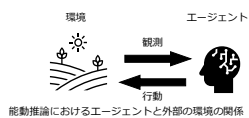
7

● 人間や生物の情報取得行動をモデル化したフレームワーク^[4]

- 人間や生物など(エージェント)と周囲の環境が相互に影響を及ぼし合う
 - 周囲の環境はエージェントが受ける感覚入力(観測)を生成
 - エージェントは行動 a を起こすことによって周囲の環境を変更

● エージェントは推定と行動を繰り返す

- 状態の推定
 - 観測 o をもとに周囲の環境の状態 s を推定
 - 変分自由エネルギーの最小化によってモデル化
- 行動の実行
 - エージェントにとって好ましく周囲の状況の不確実性を減らすような観測 o が得られる行動 a を実行
 - 期待自由エネルギーを最小化する行動を選択するとモデル化



[4] R. Smith, K. J. Friston, and C. J. Whyte, "A step-by-step tutorial on active inference and its application to empirical data," *Journal of Mathematical Psychology*, vol. 107, pp. 1–60, Feb. 2022.

7

能動推論の適用

8

● 能動推論における各要素を根拠情報収集における要素と対応付ける

- 観測 o : システムが情報源から得る情報
- 行動 a : システムがアクセスする情報源の切り替え
- 状態 s : 河川周辺の状況とシステムが現在アクセスしている情報源の位置

● 自由エネルギーの最小化によって情報収集と状況推定を実行

1. ハザードマップ等の情報から初期状態の信念を設定
2. 期待自由エネルギーの最小化によってアクセスする情報源を選択しアクセス
 - 期待自由エネルギー : $G(\pi) = -E_{Q(o|\pi)}[D_{KL}[Q(s|o, \pi)||Q(s|\pi)] - E_{Q(o|\pi)}[\text{log}P(o)]]$
3. 情報源から得られた情報をもとに、変分自由エネルギーの最小化によって真の状態を推論
 - 変分自由エネルギー : $F = E_{Q(s)}[\ln Q(s) - \ln P(o, s)]$
4. 浸水の有無に対する不確実性が減少するまで2と3を繰り返す

8

評価方法

9

● 浸水の有無に関して有用な情報を提供できるかを検証

- マップ内の特定の地点に関する情報が発信されたと仮定
 - 右図の色をついた四角で示す地点でそれぞれ実施
 - 情報が発信された地点の浸水の有無に関する信念を測定
 - 浸水の状況を変更しそれぞれの場合で検証



● アクセスする情報源を適切に選択できているかの検証

- 比較手法として情報源を無作為に選択した場合の平均を測定
- 初期状態や状態の更新方法は提案手法と同じ設定

9

評価結果：小規模マップの青色のマスでの測定結果

10

● マップ内のどのマスでも浸水が発生していない場合

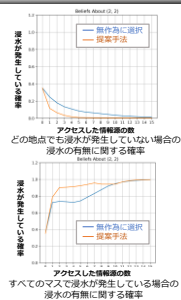
- 対象の地点で浸水が発生している確率が 0 に収束
- 浸水が発生している可能性が低いとする根拠情報を出力

● すべてのマスで浸水が発生している場合

- 対象の地点で浸水が発生している確率が 1 に収束
- 浸水が発生している可能性が高いとする根拠情報を出力

● 両方の場合で無作為に選択した場合よりも早く収束

- 提案手法によって、より少ない回数のアクセスで判定に必要な情報を収集



10

まとめと今後の課題

11

● まとめ

- 偽情報判定プラットフォームのための根拠情報収集手法として、能動推論を用いて情報収集と状態推定を行い、根拠情報を提供する手法を考案した
- シミュレーションによって河川氾濫シナリオにおいて浸水の確率を出力できることを確認した
- 広域マップにおいて計算量を削減する手法を考案し、広域マップを用いて動作を確認した

● 今後の課題

- 実際の事例や観測結果をもとにした環境での検証
 - 過去の河川水位や雨量などのデータを使用したシミュレーションを実施する
- 不正確な情報を提供する可能性がある情報源が存在する場合を想定した手法の実装
 - 提案手法では、情報源の信頼性に応じた推論が可能だと考えられる

11

各変数の詳細

12

● 状態 s

- 氾濫の状況と現在システムがアクセスしている情報源の位置を表す
- 氾濫の状況の表し方は以下の通り
 - マップをグリッドに分割する
 - 各マスは「マス内に浸水している地点がある」「浸水している地点がない」の二値を持つ
 - 氾濫の状況は各マスの状態の組み合わせとして表現する

● 観測 o

- システムが情報源のアクセスによって得る情報を表す
- その情報源が存在するマスが「浸水している地点がある」「浸水している地点がない」のどちらの状態なのかを表す

● 行動 a

- システムがアクセスする情報源の変更を表す

12

生成モデル

13

● 周囲の環境によって観測がどのように生成されるかを示すモデル

- $P(o_{0:T}, s_{0:T} | \pi) = P(s_0) \prod_{t=0}^{T-1} P(s_{t+1} | s_t, \pi) \prod_{t=0}^{T-1} P(o_t | s_t)$
- 各要素の意味は以下の通り
 - $P(s_0)$: 初期状態の信念
 - $P(s_{t+1} | s_t, \pi)$: ポリシー π において、1 ステップ進んだ際の状態の遷移
 - $P(o_t | s_t)$: 状態 s_t において観測 o_t が得られる確率
- 河川氾濫に関する偽情報判定における各要素の意味
 - $P(s_0)$: ハザードマップ等から事前に分られる各氾濫状況の起こりうる確率
 - $P(s_{t+1} | s_t, \pi)$: ポリシー π によるアクセスする情報源の変化
 - $P(o_t | s_t)$: 現在アクセスしている情報源のある地点である観測を生成する確率

13

自由エネルギー

14

● 変分自由エネルギー

- エージェントは変分自由エネルギーの最小化によって状況を推論
- $F = \mathbb{E}_{Q(s)} [\ln Q(s) - \ln P(o, s)]$
- F が小さくなるほど $Q(s)$ が $P(s|o)$ に近づく
 - ➔ 変分自由エネルギーの最小化によって事後分布の近似が可能

● 期待自由エネルギー

- エージェントは期待自由エネルギーを最小化する行動を選択
 - 情報利得と事前の嗜好の両方を可能な限り大きくすることを旨とする
 - $G(\pi) = -\mathbb{E}_{Q(o|\pi)} [D_{KL}[Q(s|o, \pi) || Q(s|\pi)] - \mathbb{E}_{Q(o|\pi)} [\log P(o)]]$
- $P(o)$: 事前の嗜好
- 情報利得: 観測が得られたときの信念の変化の大きさを表す。
 - 事前の嗜好の項: 観測が事前の嗜好と一致する度合いを表す

14

評価結果：小規模マップの赤色のマスでの測定結果

15

● 浸水が発生していない場合は判定可能

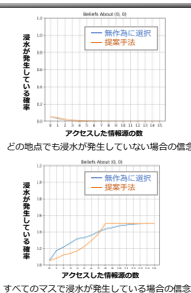
- 対象の地点で浸水が発生している確率が 0 に収束
- 浸水が発生している可能性が低いと判明

● 断定が不可能な状況では 0.5 に収束

- ここよりも浸水する確率の低い地点が存在しないため周辺の情報収集による判定が不可能



小規模マップ (再掲)



どの地点でも浸水が発生していない場合の信念

すべてのマスで浸水が発生している場合の信念

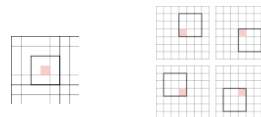
15

広域マップにおける推論の効率化

16

● 提案手法を広域なマップに用いると莫大な計算量が必要になる

- 提案手法では状態をマップ内の各マスの組み合わせとして表現
- マップのマスの数 n に対して 2^n で計算時間が増加
- 小規模なマップで推論範囲をスライドしていくことによって対応する
- 最初は真偽判定の対象の地点(図の赤色)に隣接した領域で推論を実施
- 最初の推論で対象の地点の浸水の有無が判定できない場合は推論範囲をスライド



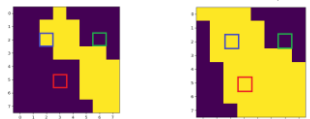
16

広域マップにおける提案手法の評価

17

● 広域マップにおける推論の効率化の手法を評価

- 右図の青、赤、緑で示す3つのマスそれぞれの場合で検証
 - それぞれマスC、マスD、マスEと呼ぶ
- 真の状態として以下の状態を想定
 - マップ内のどの地点でも浸水が発生していない場合
 - すべてのマスの地点で浸水が発生している場合
 - 浸水しているマスとしていないマスが混在する場合(2パターン)



浸水しているマスとしていないマスが混在する場合 (パターン1)

浸水しているマスとしていないマスが混在する場合 (パターン2)

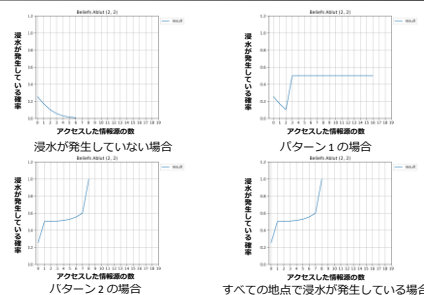


広域マップ

17

広域マップにおける評価の結果：マスCの場合

18



浸水が発生していない場合

浸水が発生している場合

パターン2の場合

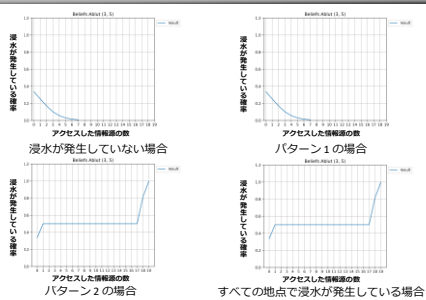
パターン1の場合

すべての地点で浸水が発生している場合

18

広域マップにおける評価の結果：マス D の場合

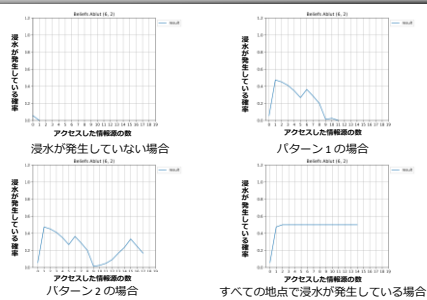
19



19

広域マップにおける評価の結果：マス E の場合

20



20