「様式３」

プロジェクト番号 P-53

複数のセンシングデータ解析による位置推定モデルの開発とノイズデータの活用法の検討

［１］組織

代表者：松田　健
（静岡理工科大学総合情報学部）
対応者：青木　徹
（静岡大学電子工学研究所）
分担者：
木谷　友哉（静岡大学大学院）
青木　徹（静岡大学電子工学研究所）

［２］研究経過

（本プロジェクトの目的・概要，及び，研究会，研究打ち合わせ・研究討論会，研究発表会，研究集会等の開催状況を記載して下さい。）

（２−１）目的と概要

本研究の目的は，人や物に装着された複数のセンシング装置から得られるセンシングデータを地図情報と照合することで人や物の位置情報を推定する数学的手法を開発することである．

近年のスマートフォンの普及によってジオメディアと呼ばれるイメージングデバイスを応用したネットワーク上のサービスが拡大している．しかしながら，GPS（Global Positioning System）を用いた位置推定の精度はデータを取得する場所の環境に依存して変化するため，GPS 以外のセンシングデータを活用することで位置推定の精度を向上させるための研究が進められている．これは本研究の目的と一致するところでもあるが，本研究では⑴センシングデータとインターネット上のテキストデータを統合して位置推定に重要なデータの特徴を抽出・加工すること，そして，⑵加工したデータと対応する地図情報の組から学習データを作成し，新たに取得したセンシングデータから精度の高い位置推定を可能とする数学モデルを開発することを目的として研究を遂行するための準備を行った．

（２−２）　研究活動状況の概要

初年度の本年は前述した目的を達成すべく，以下の活動を行った．

[A: 数学モデル開発の準備]

本研究では，センシングデータと非構造化データであるテキストデータを統合解析して精度の高い位置推定を行うことを目的としている．センシングデータとテキストデータを統合させることを検討している理由は，センシングデータはラベル無しデータであり，これらのデータの特徴をラベルデータとして自動的に付与するためである．ラベル無しデータであるセンシングデータを解析する手法として，教師無し学習と呼ばれる機械学習のアルゴリズムが存在する．本研究では，センシングデータへのラベル付けを行うことから，教師無し学習を用いてデータ分類を行い，関連するテキストデータを適切に付与する方法について検討している．そのために，センシングデバイスに精通している木谷友哉先生（静岡大学大学院）と共に，既存の教師なし機械学習のアルゴリズムを整理し，それぞれのアルゴリズムの数学的な特徴について考察してきた．その議論の中で本研究では，実際に観測されるデータと，実際には観測されない潜在的なデータを関連づける状態空間モデルに基づいた数学モデルを検討し，センシングデータに適用する方針を定めた．なお，本プロジェクトは，木谷友哉先生，青木徹先生（静岡大学電子工学研究所）からの知識補填により遂行することができた．

[B: データ取得の準備]

本研究は初年度であり，実際に解析対象となるデータを収集することが必要となる．人や物に装着されたセンシング装置からのデータを取得するため，EPSON のスマートグラス（ウェアラブルグラス）MOVERIO を購入し，人が装着した場合のデータ収集を行っている．MOVERIO には GPS，地磁気センサー，加速度センサー，ジャイロなどのセンサーが搭載されており，この装置を装着したときのデータ収集作業を進めている．取得したデータ解析は今後の課題である．

［３］成果

（３－１）研究成果

分担者の木谷友哉先生と協力し，パーティクルフ

ィルタ研究会（代表：九州工業大学・統計数理研究所　生駒哲一先生）のローカルアレンジメント担当として2015年1月9日に静岡大学浜松キャンパスで開催した．パーティクルフィルタとは，ベイズの定理に基づいて多数のサンプルにより確率密度関数を近似する手法であり，オンラインアルゴリズムとして利用することができる．以下，簡単にパーティクルフィルタについて紹介する．時刻tでの状態を表現するデータをx_t，そのときに観測されるデータをy_tとする．時刻tの状態x_tはその一つ前の状態x_{t-1}で決まり，$p\left(x_t \middle| x_{t-1}\right)$に従って発生するものとする．また，観測データは$q\left(y_t \middle| x_t\right)$に従って発生するものとする．パーティクルフィルタでは，時刻tまでの観測データ$y_{1:t}=\left\{y_1, y_2, \cdots, y_t\right\}$を用いて，目的分布$f\left(x_{1:t} \middle| y_{1:t}\right)$に従うサンプルと重みと呼ばれる量の組$\left\{\left(x_{1:t}^{(i)}, w_t^{(i)}\right)\right\}_{i=1}^{I}$を計算する．ただし，$x_{1:t}=\left\{x_1, x_2, \cdots, x_t\right\}$である．$\left\{\left(x_{1:t}^{(i)}, w_t^{(i)}\right)\right\}_{i=1}^{I}$を導出する前提として，$\left\{\left(x_{1:t-1}^{(i)}, w_{t-1}^{(i)}\right)\right\}_{i=1}^{I}$が与えられているとする．時刻$t$の粒子$x_t^{(i)}$はプロポーザル分布と呼ばれる目的分布$f\left(x_{1:t} \middle| y_{1:t}\right)$に近い計算が容易な確率分布$g\left(x_t^{(i)} \middle| x_{1:t-1}^{(i)}, y_{1:t}\right)$から生成する．重み$w_t^{(i)}$はプロポーザル分布と目的分布の比を用いて計算される．具体的には

$$w_{t-1}^{(i)} \frac{p\left(x_t^{(i)} \middle| x_{t-1}^{(i)}\right) q\left(y_t \middle| x_t^{(i)}\right)}{h\left(x_t^{(i)} \middle| x_{1:t-1}^{(i)}, y_{1:t}\right)}$$

から計算する．分母の$h\left(x_t^{(i)} \middle| x_{1:t-1}^{(i)}, y_{1:t}\right)$はプロポーザル分布から得られる．重みは正規化して使用する．本研究では，この計算の枠組みを超幾何分布に従うデータに適用し，統計的検定を行うアルゴリズムに関する検討を行った．パーティクルフィルタでは状態x_tを推定することを目的としているが，本研究はパーティクルフィルタのアルゴリズムの一部を用いて観測y_tを推定した．周辺頻度を固定したときの分割表のセル頻度は超幾何分布に従う．分割表のセル頻度xが与えられれば，Fisherの正確検定のp値計算に必要なカイ二乗値の大小関係をカウントする0,1の2値をとる関数$g(x)$の値が観測される．

Fisherの正確検定では，周辺頻度を固定した際の分割表を全て数え上げる必要があり，この実現が計算量的に困難であることが知られているが，本研究の手法を用いて分割表のセル頻度xを多数のサンプルとして発生することでp値の近似計算が可能となることを実験により示した[1]．具体的には，$p\left(x_t^{(i)} \middle| x_{t-1}^{(i)}\right)$をベイズ予測分布とし，$\Pr\left(y_t=1 \middle| y_{1:t-1}^{(i)}\right)$を$\sum_{i=1}^{I} s\left(y_t=1 \middle| y_{1:t-1}^{(i)}\right) w_{t-1}^{(i)}$として近似計算する．ここで，$\Pr\left(y_t=1 \middle| y_{1:t-1}^{(i)}\right)$は時刻$t-1$までの$g(x)$の値がすべて与えられているときに$y_t=g\left(x_t\right)$が１をとる（$x_t$に対応する分割表のカイ二乗値が分割表を構成する変数が独立であると仮定したときに計算される期待値より大きくなる）確率を表し，$s\left(y_t=1 \middle| y_{1:t-1}^{(i)}\right)$は$t-1$個のサンプルのうち$l$個のデータが$y=g(x)=1$となるとき，$\frac{l}{t-1}$と計算するものと定義した．

（３－２）波及効果と発展性など

本プロジェクトを通じて，分担者の木谷先生と共に多くの機械学習のアルゴリズムの性質をまとめることができた．また，パーティクルフィルタ研究会の開催により，センシングデータへのパーティクルフィルタ理論の応用についても，統計的検定の手法と関連させることで検討を進めていくための準備も完了しただけでなく，本プロジェクトにより他分野の研究者との交流の機会を作ることができた．今後は本プロジェクトを通して得られた知見を基に，センシングデータをより高精度に解析するための数学的基礎を確立させていく所存である．

［４］成果資料（以下１０．５ポイント）

（１）松田健，パーティクルフィルタを用いた分割表の検定への応用とその検討，平成27年１月９日，パーティクルフィルタ研究会（口頭発表）

出張報告

当該経費による出張はなし。