

【研究報告】

長瀬 永遠

富山県立大学 情報基盤工学講座

December 15, 2023

背景

近年、コンピューターの普及によって社会にはあらゆる分野において大量の情報が溢れるようになった。そのため、大量のデータから有用な情報を取り出す研究が数多くなされている。その中の一つに複数の入力データと単一の出力データからそれらの間に成り立つ数法則を発見する手法が挙げられる。

目的

社会事象について、データ間に成り立つ数法則を発見することによってそれらを表現する手法を提案する。また、それを行う上で課題となるデータの不完全さに考慮した手法を作成する。

新しい政策の効果検証がしたい…

理想

- 予算 ∞
- すぐに適用できる
- すぐに効果が出る

現実

- 予算簡単につかない
- 適用に数年
- 効果が出るのに数年

疑似的に効果検証できるようにすればよいのでは？

➡ そのためには社会を表現したモデルが必要

RF5.0

複数の入力変数（量的）と一つの出力変数（量的）との間に成り立つ数法則を求めることができる 3 層パーセプトロンを用いた多変量多項式回帰法

数法則発見法を社会現象に適用する上での課題

行政のオープンデータは数年に一度しか更新されない
＝データを時間的な観点のみで収集するのは難しい
＝空間的な観点でのデータ収集が必要
→ 全ての自治体を同じ土俵で扱っていいのかという疑問が生まれる

疑問

日本全国を見てみると人口や主要産業、地価など大きなばらつきがあるが、全てを同じ土俵で扱っていいのか？

仮説

- 全国の市区町村は潜在的にいくつかのタイプに分かれるのではないか
- タイプ分けすることで数法則発見の精度が向上するのではないか

アイデア

多数の入力変数（量的）を用いてクラスタリングを行い、発生したクラスを質的変数として入力変数に組み込むことで RF6.3 を適用する.

計算量削減とインタラクティブな分析への工夫

潜在プロファイル分析（LPA）における計算量削減のために LPA には全データではなくサンプリングしたデータを用いる．サンプリングによって除外されたデータにおける各クラスへの所属確率は以下のファジィニューラルネットワークでモデルを作成することで求める．

また，所属確率を求めるためのモデルを作成しておくことで，ユーザが一部の入力を調整した場合の結果を見る際に LPA を行う必要がなくなる．

ファジィニューラルネットワーク

提案手法の概要

はじめに
提案手法
今後の課題

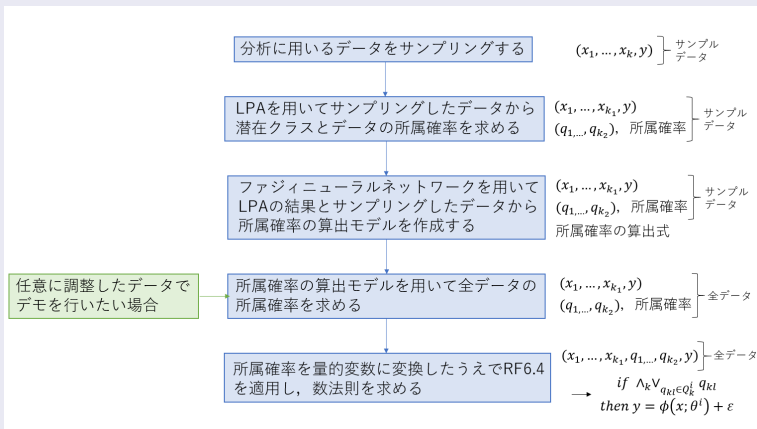


図 1: 提案手法のフロー

所属確率から量的変数への変換

LPA によって求められた潜在クラスの数に基づいてカテゴリ数と閾値を設定し、所属確率を量的変数に変換する。

(潜在クラス数が 3 の場合)

各データの所属確率の合計は必ず 1 になることから以下のパターンが想定される。

- 1 3つの潜在クラスに属する $\rightarrow q_1 = 0.33, q_2 = 0.33, q_3 = 0.33$
- 2 2つの潜在クラスに属する $\rightarrow q_1 = 0.50, q_2 = 0.50, q_3 = 0.00$
- 3 1つの潜在クラスに属する $\rightarrow q_1 = 1.00, q_2 = 0.00, q_3 = 0.00$

よって、以下のカテゴリによって量的変数を求める。

カテゴリ 1 = 1.00 ~ 0.50

カテゴリ 2 = 0.50 ~ 0.33

カテゴリ 3 = 0.33 ~ 0.00

- 提案手法のコーディング
- 数値実験の実施
- 修士論文の執筆