

無限関係モデルを用いた道路ネットワークの接続性評価

金沢大学 学生会員 ○奥村 颯太

金沢大学 正会員 中山 晶一郎

金沢大学 正会員 山口 裕通

1. 研究の背景と目的

道路は生活に欠かせない都市基盤の一つである。地震や大雪などの災害が発生した場合、道路ネットワークが寸断することで、人命救助や災害復旧のための物流に大きな影響を与える可能性がある。非常時における災害による影響を少なくするためには、効率的に道路ネットワークの各拠点間の連結性を高めることが必要となる。道路ネットワークの連結性が低い、つまり脆弱性のあるネットワークを効率的に見つけ出し、優先的に対策を行うべきである。

本研究では、クラスタリングの手法の1つである無限関係モデル (IRM) を用いて、道路ネットワーク内で連結性や関連性の高いノード同士をグループ分けするとともに、ネットワーク内での脆弱な部分などを見つけ出すことを目的とする。IRMにより道路の連結性評価を行うことで効率的かつ理論的に評価することができると考えている。

2. クラスタリングについて

類似したデータをグルーピングすることをクラスタリングという。また、行・列を適切に並び替えることで2つのデータを同時にクラスタリングすることを共クラスタリングという。ノードの接続を表す隣接行列の行の成分と列の成分の2つのデータに対して共クラスタリングを行うことで、各ノードのグループ分けを行う。

グループ数であるクラスタ数が未知の状態で共クラスタリングを実現する手法として、無限関係モデル

(IRM) がある。IRM の特徴としては、潜在的クラスタ数を自動的に決定できる点である。確率的ブロックモデル (SBM) あるいはK-meansクラスタリングのような一般的なクラスタリングの手法の多くでは事前にクラスタ数を決定する必要がある。一方、IRM ではクラスタ数を未知パラメータとして自動的に決定するため、最適なクラスタ数を試行錯誤で探索せずにすむ。IRM のクラスタ生成モデルにおける初期値は中華料理店過程 (CRP) という確率分布によって求める。

3. 中華料理店過程 (CRP) について

中華料理店過程 (CRP) とは無数個の要素集合の分割に対して確率を与える確率過程である。CRP から生成されるクラスタ数は固定されず、サンプリングのたびに確率的に変動する。第1ドメイン (行) と第2ドメイン (列) を複数のクラスタにそれぞれ分解して、関係データ行列をクラスタ同士のパラメータ θ_{ij} に従って確率的に関係データを生成する。CRP の説明は中華料理店を比喩的に用いることが通例である。

1. 各テーブルに着席する客数に上限はない
2. 中華料理店に配置するテーブル数に上限はない
という中華料理店を仮定する。そのうえで、入店した客をどのテーブルに配置するか、つまり与えられたデータ要素 (客) をどのようにクラスタリング (テーブル配置) して分割を与えるかが CRP である。いま客 k が i 番目のテーブル ω_i に着席していることを $s_k = \omega_i$ と表現する。このテーブルの着席確率は以下のようになる。

$$P(s_k = \omega_i) = \begin{cases} \frac{n_i}{n-1+\alpha} & (\text{既存クラスタの場合}) \\ \frac{\alpha}{n-1+\alpha} & (\text{新規クラスタの場合}) \end{cases}$$

n :すでに来店した客数

n_i : i 番目のテーブルにすでに着席している客数

α :新規テーブルに与えられるパラメータ

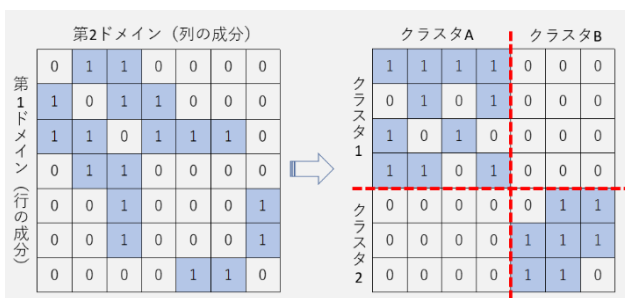


図-1 共クラスタリングのイメージ

もし、 N 人の客が入店した場合、テーブル数は最大で N 個になることがわかる。また、CRP ではなく多くの人が着席しているテーブルに新たな客が集まりやすい設計となっているため、データ要素に対するクラスタリング効果がある。これを道路ネットワークに当てはめると、多くの需要があるクラスターにノードが集まることになると考えられる。

4. 無限関係モデル (IRM) について

IRMの学習とは、観測データの背後にある共クラスタリング構造($\mathbf{s}^1, \mathbf{s}^2$)を、観測データとIRMの生成モデルから推定することである。ここで($\mathbf{s}^1, \mathbf{s}^2$)はクラスター所属のリストを表す。具体的には、ベイズの公式を用いて、観測データ $\mathbf{R}$ が得られた下で、パラメータを積分消去した $\mathbf{s}^1, \mathbf{s}^2$ の事後確率

$$P(\mathbf{s}^1, \mathbf{s}^2 | \mathbf{R}) = \frac{P(\mathbf{R} | \mathbf{s}^1, \mathbf{s}^2) P(\mathbf{s}^1) P(\mathbf{s}^2)}{P(\mathbf{R})}$$

を最大化する $\mathbf{s}^1, \mathbf{s}^2$ を求める問題に帰着される。ここで、 $\mathbf{s}^1, \mathbf{s}^2$ の事前分布 $P(\mathbf{s}^1), P(\mathbf{s}^2)$ は共に上述のCRP(α)である。この最大化問題を解くには、ディリクレ混合過程モデルに対するギブスサンプリング(パラメータを消去する場合)を適用する。すなわち、($\mathbf{s}^1, \mathbf{s}^2$)を逐次サンプリングし、 $P(\mathbf{R} | \mathbf{s}^1, \mathbf{s}^2)$ を最大化する($\mathbf{s}^1, \mathbf{s}^2$)を解とすればよい。

5. 分析方法

IRMの計算はpythonで行う。入力データは関係データ行列 $\mathbf{R}$ 、出力データは行オブジェクト、列オブジェクトの所属クラスターを表す潜在変数の値 $\mathbf{s}^1 = \{s_1^1, \dots, s_K^1\}$, $\mathbf{s}^2 = \{s_1^2, \dots, s_L^2\}$ 、行クラスター数 c^1 、列クラスター数 c^2 である。また、CRPのパラメータ $\alpha = 1.0$ とする。一般に N が大きいときは $\alpha = 0.1$ 、 N が小さいときは $\alpha = 1.0$ とすることが多い。

6. 道路ネットワークへの適用

IRMを仮想ネットワーク(ノード数13)と実ネットワークである石川県の緊急輸送道路(ノード数411)の隣接行列において適用する。仮想ネットワークで、ノード1~4はクラスター1、ノード5~8はクラスター2、ノード9~13はクラスター3に分かれ、このとき境目はノード4と9、ノード7と9であり、これらのリンクが脆弱であることがわかる。

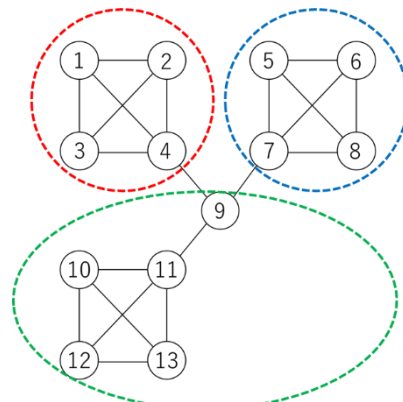


図-2 仮想ネットワーク(ノード数13)

また石川県の緊急輸送道路においてIRMを適用すると図-3に示すような結果となった。クラスターの境目は奥能登・七尾・中能登・金沢・加賀の大きく5つとなることが確認できた。クラスターの境目の1つは犀川流域と重なっており、川がクラスターの分割に影響している可能性がある。

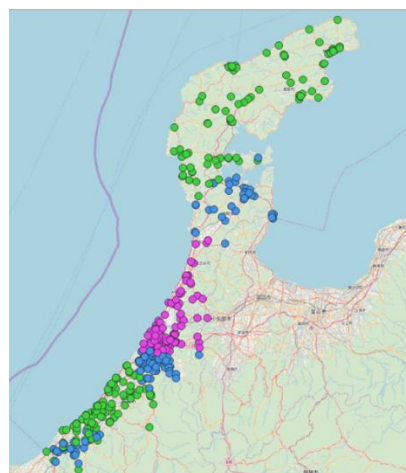


図-3 石川県緊急輸送道路への適用(ノード数411)

7. 今後の展望

今回用いた関係データは道路の接続を表す隣接行列のみであったため道路長や交通量が考慮できていない。実際の道路ネットワークにおいてこれらの要素を関係データに組み込み、より現実的な分析が必要である。また、最終的なクラスター数とIRMが収束する計算回数との関係性も今後考える必要がある。

参考文献

- 1) 石黒勝彦, 林浩平: 関係データ学習, pp. 45-86
- 2) 石井健一郎, 上田修功: 続・わかりやすいパターン認識, pp. 277-289