

## 道路構造の類似性に着目した機能クラスタ抽出の高速化

伏見 卓恭† 齊藤 和巳‡ 風間 一洋§

†筑波大学図書館情報メディア系 ‡静岡県立大学経営情報学部 §和歌山大学システム工学部

## 1 はじめに

近年、ハイパーリンクネットワークや SNS 内でのユーザ関係ネットワークだけでなく、電力網や道路網の分析に対しても複雑ネットワーク分析を適用することで、興味深い研究成果が報告されている [1]. 著者らは、ネットワークにおいてノードが果たす機能に着目し、類似機能を有するノード群である機能クラスタを抽出する FCE 法を提案した [4]. ネットワークにおける機能の例として、社内ネットワークにおける「部長」、「一般社員」や、道路ネットワークにおける「市街地」、「郊外」などが挙げられる. この手法は、ネットワーク上でのランダムウォークに着目して、各ノードの確率が収束する過程を各ノードの機能ベクトルとし、そのコサイン類似度でノード間の機能類似度を定義する. そして、機能類似度に基づき  $K$ -medoids 法によりノードをクラスタリングすることで、機能的に類似するノード群を同定する.

大規模なネットワークにおいては、全ノードペアの機能類似度を計算するために膨大な時間が掛かることに加えて、メモリ上に保持することも困難になる. この問題に対して、サンプリングが最も簡単な解決策であるが、サンプリングで選ばれたノード群による解の精度は保証されない. 著者らは、FCE 法における  $K$ -medoids 法の貪欲解法を高速化するアルゴリズムを提案した [3]. このアルゴリズムでは、幾つかの機能ベクトルをピボットとして抽出し、ピボットとの距離下界値を利用し、距離計算不要なノードペアの計算を省略することで高速化を図っている. 機能ベクトルは高次元で表現されるため、計算回数削減の効果は大きい. 本稿では、既存研究 [2] における、「各都市ネットワークの機能クラスタは共通した特性を有する」という知見を利用し、転移学習に基づく高速なクラスタリング手法を提案する.

## 2 提案手法

転移学習の考え方にに基づき、FCE 法における  $K$ -medoids クラスタリングのフェーズを高速化する. 形式的には、

Acceleration of Functional Cluster Extraction Focused on Similarity of Road Network Structure  
†Takayasu FUSHIMI ‡Kazumi SAITO §Kazuhiro KAZAMA  
†University of Tsukuba

転移学習における元ドメインと目標ドメインの機能ベクトル群を  $\mathbf{X}^{(S)}$  と  $\mathbf{X}^{(T)}$  と表記する. 元ドメインにおいて、 $K$  個のメドイドノードの機能ベクトル  $\{\mathbf{x}_1^{(S)}, \dots, \mathbf{x}_K^{(S)}\} \subset \mathbf{X}^{(S)}$  は、あらかじめ FCE 法により抽出されているとする. これらの  $K$  個のベクトルを利用し、目標ドメインのメドイドベクトル  $\{\mathbf{x}_1^{(T)}, \dots, \mathbf{x}_K^{(T)}\} \subset \mathbf{X}^{(T)}$  を  $\mathbf{x}_k^{(T)} = \arg \max_{\mathbf{x}^{(T)} \in \mathbf{X}^{(T)}} \langle \mathbf{x}_k^{(S)}, \mathbf{x}^{(T)} \rangle$  として抽出する. ここで、両ドメインの機能ベクトルは正規化されているとする.

## 3 評価実験

評価実験では、OpenStreetMap (OSM) データから抽出した以下の 6 都市を用いた: 静岡県, 神奈川県, 京都府, サンフランシスコ, ニューヨーク, バルセロナ. 各都市の OSM データから交差点と道を全て抽出し、交差点をノード, 交差点間の道をリンクとした空間ネットワークを構築し、機能ベクトルの次元を  $S = 10,000$ , クラスタ数を  $K = 5, 10$  とし、計算時間とクラスタリングの正解率を以下の既存手法と比較し定量的に評価する. 紙面の都合上、目標ドメインを静岡県道路網とした結果のみを示すが、他の都市データにおいても同様の結果が得られることを確認した. 比較する既存手法として、遅延評価技術 (LE) [5], 貪欲法において抽出済みのメドイドをピボットとし距離計算回数をプルーニングする技術 (MP) [3], 外れ値をピボットとして距離計算回数をプルーニングする技術 (OP) [3] を組み合わせた手法を用いる. OP 技術において、外れ値ピボット数を  $H = 10, 20$  とした.

図 1 は、横軸をクラスタ数 (メドイド数), 縦軸をクラスタリングの計算時間を対数でプロットしたものである. LE (黒点線), LE+MP (青実線), LE+MP+OP (赤実線), 提案手法 (緑実線) を示す. 図 1 から、提案手法が桁外れに高速であることが確認できる.

提案手法による目標ドメインのメドイド選定は、比較している既存手法と異なり、近似的な解を出力する. したがって、クラスタリング精度について評価する必要がある. 本稿では、多クラス分類タスクで用いられる混同行列の対角成分  $M_{kk}$  の和と分類の総数  $M = \sum_{k=1}^K \sum_{h=1}^K M_{kh}$  により、 $ACC = \frac{1}{M} \sum_{k=1}^K M_{kk}$  として計算される指標 (正解率) により評価する. 目標ドメインのネットワークにおける全機能ベクトルを、FCE 法に

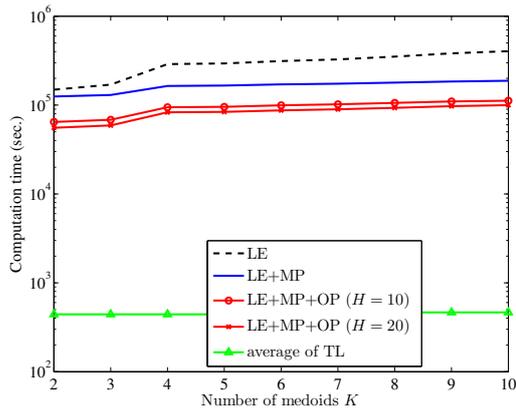


図 1: 計算時間比較

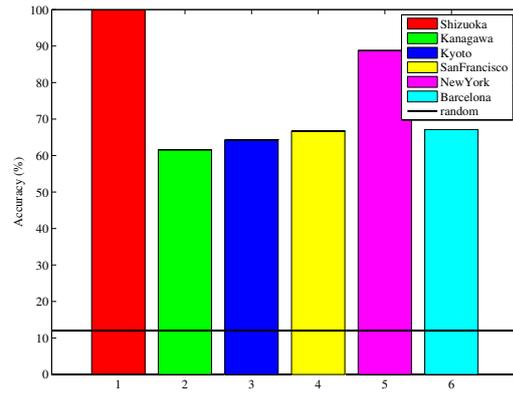


図 2: クラスタ正解率 ( $K = 5$ )

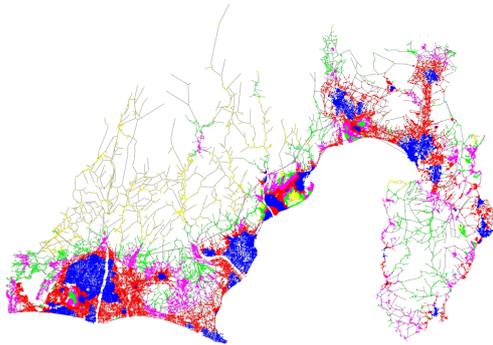


図 3: FCE 法による正解機能クラスタ ( $K = 5$ )

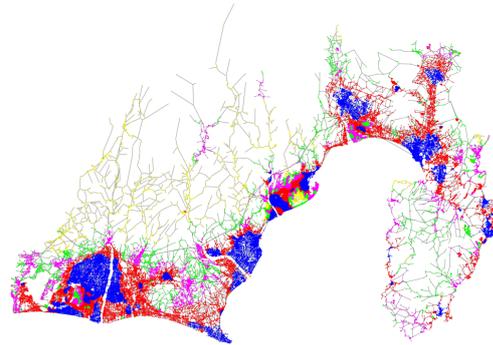


図 4: 提案手法による機能クラスタ ( $K = 5$ )

より分類した  $K$  個のクラスタを正解とする．さらに、ランダムに  $K$  個選んだオブジェクトを用いてクラスタリングした random と比較する．図 2 は、各元ドメインから目標ドメイン（静岡県）のネットワークをクラスタリングした際の正解率 ACC をプロットした．図 2 から、random に比べて提案手法によるクラスタリング結果の正解率が有意に高いことが確認できる．さらに、静岡県はニューヨーク（マゼンタ）のメドイドを元ドメインとした際、正解率が 9 割程度と最も高い正解率を示している．

図 3, 4 に、直接 FCE 法を用いて目標ドメインの機能クラスタを抽出した場合（正解）と提案手法により最も正解率の高い元ドメインを用いて機能クラスタを抽出した場合の可視化結果を示す．図 3, 4 から、静岡県の主要部分は、ほぼ忠実に再現できていることが確認できる．さらに、誤りが生じる部分は、次数が高い都市部ではなく、次数が低い都市部分の辺縁部や山間部の一部であることがわかった．

#### 4 おわりに

本研究では、ある地域においてあらかじめ抽出されたメドイドを利用し、他の地域のメドイドを抽出する

ことで高速に機能クラスタを抽出する手法を提案した．評価実験の結果から、最先端技術と比べて著しく高速に、かつ、ランダムと比べて高い正解率でクラスタリングできることを確認した．今後は、対象地域の構造的な特性との関連について分析を深めていきたい．

謝辞 本研究は、JSPS 科研費 16K16154 および JSPS 特別研究員奨励費 15J00735 の助成を受けたものです．

#### 参考文献

- [1] Crucitti, P. et al.: Centrality Measures in Spatial Networks of Urban Streets, *Physical Review E*, Vol. 73, No. 3, pp. 036125+ (2006).
- [2] Fushimi, T. et al.: Extracting and Characterizing Functional Communities in Spatial Networks, *Proc. of the Workshop on Artificial Intelligence for Tourism (AI4Tourism2016)*, pp. 182–193 (2016).
- [3] Fushimi, T. et al.: Functional Cluster Extraction from Large Spatial Networks, *Proc. of the 2016 IEEE/ACM Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM2016)*, pp. 57–62 (2016).
- [4] Fushimi, T. et al.: Extracting Communities in Networks based on Functional Properties of Nodes, *Proc. of the 12th Pacific Rim Knowledge Acquisition Workshop (PKAW2012)*, Berlin, Heidelberg, Springer-Verlag, pp. 328–334 (2012).
- [5] Leskovec, J. et al.: Cost-effective Outbreak Detection in Networks, *Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Mining*, New York, NY, USA, ACM, pp. 420–429 (2007).