

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 22 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330270

研究課題名(和文) SNSクラスタリングのための伝統的理論vs.発見的技法

研究課題名(英文) Traditional method versus heuristic techniques for Social Network clustering

研究代表者

宮本 定明 (MIYAMOTO, Sadaaki)

筑波大学・システム情報系(名誉教授)・名誉教授

研究者番号：60143179

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：近年、ソーシャルネットワークにどのようなグループが存在しているのかが問題とされ、解析ツールとしてクラスタリングという互いに緊密につながったグループの抽出技法が注目されている。よって、ネットワーククラスタリング技法が開発されてきた。一方、伝統的なクラスタリング技法はソーシャルネットワークには適していないとみなされている。ここではこの考えが正しいかどうかを検証すべく、伝統的方法の延長上にある新規技法でネットワーククラスタリングを行い、既知技法との比較を行った。メドイド法と階層的技法を組み合わせた方法が有効であること、Ward法として知られた方法が直接ネットワークに適用できること等を示した。

研究成果の概要(英文)：Analysis of social networking services (SNS) is recently focused on by many researchers and techniques of network clustering have been developed. In contrast, traditional methods of clustering are considered to be inappropriate for network clustering. In this study new methods based on traditional ideas of clustering are developed and performances are compared with known methods of network clustering. Main results of this study include the development of two-stage clustering methods using the medoids and agglomerative hierarchical clustering which is effective in SNS clustering. Moreover the Ward method of agglomerative hierarchical clustering is shown to be applicable directly to networks without the use of a positive-definite kernel.

研究分野：ソフトコンピューティング

キーワード：ソーシャル・ネットワーク クラスタリング 階層的クラスタリング メドイド法 2段階階層的クラスタリング Ward法 ファジィクラスタリング

1. 研究開始当初の背景

SNS (Social Networking Service) におけるコミュニティ発見の重要性はここ数年認識が高まり、多くの研究がなされるようになった。どのようなコミュニティが存在するかは不明な場合が多く、教師付き分類のみならず、教師のないクラスタリング技法が注目され、いくつかの方法が開発された。その際、既存の方法は用いられなかった。その理由として、非対称類似度が問題となること、クラスターの代表点を効果的に抽出したいこと、大量のデータを扱う必要があること、などが挙げられる。実際は既存の方法でもこれらの問題点にはある程度対応できるが、新たに開発された技法は既存の方法をしのぐ、と考えられている。新たに開発された技法の代表的なものとして、Newman らの階層的アルゴリズム、とそれを発展させた Louvain 法、DBSCAN に代表される density seeking などがある。これらは、発見的 (ヒューリスティック、思いつきの) アルゴリズムであり、理論的性質がわかりにくいという難点がある。一方、伝統的な方法 (階層的技法、K-means など) は、非対称類似度をもち、かつ大量データの場合に対応しにくいという欠点がある。また理論的に良く整理された混合分布モデルは、この応用分野では分類対象がグラフの頂点とみなされるため、適用できないことが多い。

2. 研究の目的

新たな発見的技法と伝統的技法を組み合わせるといのは自然な発想であるが、単に組み合わせるといっただけでは、学術的観点からみて、意義は乏しい。しかしながら、クラスタリング技法を俯瞰的に観察すると、この組合せの意義を明らかにする理論を構築することができる。

本研究では、これまでの研究成果に則って、SNS におけるクラスタリング諸技法を新たなものと既存のもの両面から対比させつつ考察し、上記の例に述べたように、それらの関連性を明らかにすることによって、理論体系を確立するとともに、新たな技法を生み出し、SNS におけるデータに適用してその効果と効率を調べることを目的とする。

3. 研究の方法

(1) ネットワーククラスタリングに関する従来の諸技法を俯瞰的観点から考察することにより、技法間の関連性を考察し理論的性質を解明する。

(2) 新規性のあるアルゴリズムを開発する。たとえば、伝統的な階層的アルゴリズムを大量のデータに適用できるようにする 2 段階クラスタリング技法など。また、非対称類似度をもちネットワーククラスタリングについて考察する。

(3) Newman 法や Louvain 法と伝統的アルゴリズム (およびそれから導かれる新規アルゴリ

ズム) との比較検討を行う。また、開発したアルゴリズムの SNS データへの適用と評価を行う。

(4) ネットワーククラスタリングにおけるファジィクラスタリング技法の導入可能性について検討する。

4. 研究成果

以下では、研究の方法(1)~(4)に即して研究成果を述べ、あわせて当初予想していなかった研究成果と要約について述べる。

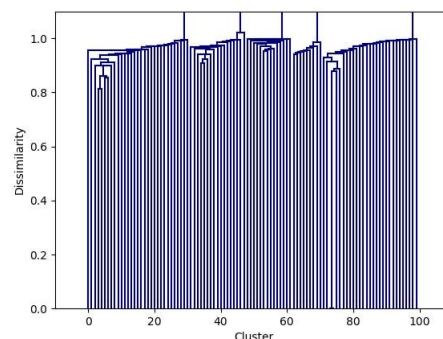
(1) 理論的考察に関して、DBSCAN が single linkage 技法の変形であることを示した。また、DBSCAN と spectral clustering を組み合わせることによって、アルゴリズムの効率化ができることを示した (論文,)。

また、従来 Ward 法は一般にネットワーククラスタリングに直接的には利用できないと考えられてきたが、そうではなく、緩い条件のもとでネットワーククラスタリングに利用可能であることを示した (論文, 発表)。

この結果は「厳格でないユーザ」というモデルを導入することで得られたもので、従来の常識を書き換える成果といえる。

(2) 伝統的技法の延長として得られる新規アルゴリズムとして、2 段階階層的クラスタリングを主に考察した。ネットワークにおいてクラスター中心の働きをするのはメドイドが自然と考えられることから、メドイド (medoid) を用いたアルゴリズムを中心に考察した。2 段階の 1 段階目では、繰り返し回数を 1 ~ 2 回に制限したメドイド法などを考察し、2 段階目では、Medoid-Ward 法 (Ward 法における重心をメドイドで置き換えた方法) などを提案した。さらに 1 段階目では、K-means++ と同様の初期中心設定法を用いた (論文, , 発表)。

また、メドイドの概念を非対称類似度に適用した。出方向 medoid と入方向 medoid を定義し、これら 2 つのメドイドをもつクラスターを考察し、さらに 2 段階アルゴリズムを開発した (論文, , 発表,)。なお、下記の図は、提案技法を以下に述べる party data に適用して得られた樹形図の一例である。



(3) Newman 法・Louvain 法と本研究で開発したアルゴリズム(提案技法と呼ぶ)を比較するため、既存の repository data の他に、研究室で収集したデータ(Party data と呼ぶ)を用いた。収集したデータは、既存のものに比べて、解釈が簡単で、クラスターへの分離の良さが適度である点、方法の評価に適していると考えられる。

比較の結果を要約すると、千~数万のオーダーのデータについては、Newman 法・Louvain 法と提案技法は適用可能な計算時間をもち、出力である樹形図からの結果の解釈の容易さでは提案技法が優り、計算時間の点では Newman 法・Louvain 法が優る。これよりも個体数が増えてくると、Louvain 法が計算時間の点ではるかに優れている。Newman 法・Louvain 法ではクラスター数が自動推定できるが、極めてクラスター分離の良いデータ以外では、推定されたクラスター数は適切とはいえない。全体的には、提案技法と従来技法とは一長一短であり、両者を合わせ用いるのが良いと考えられる(論文 , , 発表 , ,)

(4) ファジィクラスタリング技法は主にメドイド技法に関連して導入すべきであることを示し、その際生じるファジィ識別関数の性質について考察した(論文 , , 発表 , ,)。また、ネットワーククラスタリングに関連するファジィクラスタリング技法としてファジィ近傍法(論文) やネットワーク上のクラスター回帰技法(論文 , , 発表)についても考察した。

(5) これらの成果の中で代表的なものは、理論では(1)に挙げた Ward 法に関する性質、新規の提案法では(2),(3)に挙げた 2 段階メドイド階層技法である。Ward 法に関する理論的性質は過去数十年解かれていなかった問題を新たな観点から解決したものである。2 段階メドイド階層技法は樹形図が解釈しやすい点で従来法より優れている。さらに、ネットワーク上のクラスター回帰(c-回帰)技法は従来考察されてこなかった問題であり、これら 3 つはすべて著しい結果と思われる。これらを要約した自己評価としては、所期の目的を達したという意味で一定の成果を挙げたと同時に、研究開始時には想定していなかった 2 つの成果(Ward 法に関する成果とネットワーク上の c-回帰技法)を挙げたという点で、優れた成果が得られたとして良いものと思われる。

今後の研究課題として、メドイド法の効率化と伝統的な階層的クラスタリングではクラスター数の推定はできないかどうかという問題が新たに生じてきた。これらについてさらに考察する予定である。

<引用文献>

- M. E. J. Newman, Physical Review E, 69, 066133(2004).
M. Girvan, M.E.J. Newman, PNAS, 99(12)

pp.7821-7826(2002).

D.Blondel, J.L.Guillaume, R.Lambiotte, E.Lefebvre, Journal of Statistical Mechanics : Theory and Experiment, p10008(2008).

M.Ester, H.P.Kriegel, J.Sander, X.Xu, Proc. of KDD-96, pp.226-231(1996).
宮本定明、クラスター分析入門、森北出版、1999.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

(雑誌論文)(計 11 件)

Sadaaki Miyamoto, Yoshiyuki Komazaki, Yasunori Endo, Generalizations of Fuzzy c-Means and Fuzzy Classifiers, Lecture Notes in Artificial Intelligence (LNAI 9978), pp.151-162, 2016 (査読有). DOI: 10.1007/978-3-319-49046-5_13

Sadaaki Miyamoto, Yousuke Kaizu, Yasunori Endo, Hierarchical and Non-hierarchical Medoid Clustering Using Asymmetric Similarity Measures, Proc. of SCIS&ISIS2016, Aug.25-28, 2016, Hokkai-Gakuen University, Sapporo, Japan, pp.400-403, 2016 (査読有). DOI: 10.1109/SCIS&ISIS.2016.129

Tatsuya Higuchi, Sadaaki Miyamoto, Yasunori Endo, Fuzzy c-Regression Models for Fuzzy Numbers on a Graph, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.20, No.4, pp. 521-534, 2016 (査読有). DOI: 10.20965/jaciii.2016.p0521

Sadaaki Miyamoto, Ryosuke Abe, Yasunori Endo, Jun-ichi Takeshita, Ward Method of Hierarchical Clustering for Non-Euclidean Similarity Measures, Proc. of SoCPaR 2015, Nov. 13-15, 2015, Kyushu University, pp.60-63, 2015 (査読有). DOI: 10.1109/SOCPAR.2015.7492784

Sadaaki Miyamoto, So Miyahara, Yusuke Tamura, Reduction of Points for Network Data, CD-ROM Proc. of MDAI 2015, pp.137-145, 2015 (査読有).

Yousuke Kaizu, Sadaaki Miyamoto, Yasunori Endo, Hard and Fuzzy c-Medoids for Asymmetric Networks, Proc. of 16th World Congress of the International Fuzzy Systems Association (IFSA 2015), June 30 - July 3, Gijon, Spain, pp.435-440, 2015 (査読有). DOI: 10.2991/ifsa-eusflat-15.2015.63

Yusuke Tamura, Sadaaki Miyamoto, Two-Stage Clustering Using One-Pass K-Medoids and Medoid-Based Agglomerative Hierarchical Algorithms, Proceedings of the Joint 7th International Conference on Soft Computing and Intelligent Systems and 15th International Symposium on Advanced Intelligent Systems (SCIS&ISIS 2014), pp. 484-488, Kitakyushu, Japan, 2014 (査読有). DOI:10.1109/SCIS-ISIS.2014.7044641

Yusuke Tamura, Sadaaki Miyamoto, A Method of Two Stage Clustering Using Agglomerative Hierarchical Algorithms with One-Pass k-Means++ or k-Median++, Proc. of 2014 IEEE International Conference on Granular Computing (GrC2014), Noboribetsu Grand Hotel, Noboribetsu, Hokkaido, Japan, Oct. 22-24, 2014, pp.281-285 (査読有). DOI: 10.1109/GRC.2014.6982834

So Miyahara, Sadaaki Miyamoto, A Family of Algorithms Using Spectral Clustering and DBSCAN, Proc. of 2014 IEEE International Conference on Granular Computing (GrC2014), Noboribetsu Grand Hotel, Noboribetsu, Hokkaido, Japan, Oct. 22-24, 2014, pp.196-200 (査読有). DOI:10.1109/GRC.2014.6982850

Zhang Canlun, Sadaaki Miyamoto, Text Clustering Using Fuzzy Neighborhood and Evaluation of Clusters, Proc. of 2014 IEEE International Conference on Granular Computing (GrC2014), Noboribetsu Grand Hotel, Noboribetsu, Hokkaido, Japan, Oct. 22-24, 2014, pp.19-24 (査読有). DOI: 10.1109/GRC.2014.6982800

Tatsuya Higuchi, Sadaaki Miyamoto, Fuzzy c-Regression Models Combined with Support Vector Regression, Proc. of 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2014) July 6-11, 2014, Beijing, China, pp.2489-2493, 2014 (査読有). DOI:10.1109/FUZZ-IEEE.2014.6891624

[学会発表](計12件)

Sadaaki Miyamoto, Hierarchical Cluster Analysis for Large-Scale Data Sets with Application to Finance and Social Networks, Annual International Conference of Thailand Econometric Society 2017, 招待講演, 2017.1.11-13, Chiang Mai University, Chiang Mai, Thailand

Sadaaki Miyamoto, Ward Method Applied to Non-Positive Definite Matrices with

a Model of Unstrict Users, 5th German-Japanese Workshop on Classification, 招待講演, 2016.9.11-13, University of Ulm, Guenzburg, Germany

Sadaaki Miyamoto, Methodological Aspects of Fuzzy Clustering, Annual International Conference of Thailand Econometric Society 2016, 招待講演, 2016.1.6-8, Chiang Mai University, Chiang Mai, Thailand

宮本定明, 阿部亮介, 豊田章倫, 遠藤靖典, 竹下潤一, 非ユークリッド的類似度に対する階層的クラスタリング: Ward法の利用について, 第6回横幹連合コンファレンス, 2015.12.5-6, 名古屋工業大学

藤原秀平, 宮本定明, 遠藤靖典, Core-pointを利用した有向ネットワークに対する階層的クラスタリング, 第31回ファジィシステムシンポジウム, 2015.9.2-4, 電気通信大学, 調布市

宮本定明, クラスタ分析の理論と研究動向, 大阪大学数理・データ科学セミナー, 招待講演, 2015.9.29, 大阪大学基礎工学研究科, 大阪府豊中市

Sadaaki Miyamoto, Two-stage agglomerative hierarchical clustering using medoids for network clustering, IFCS 2015, 2015.7.6-7.8, University of Bologna, Bologna, Italy

Keisuke Minakawa, So Miyahara, Yusuke Tamura, Sadaaki Miyamoto, Comparing Different Methods on Clustering Follow-Follower Relations on Twitter, 11th International Conference on Modelling for Artificial Intelligence (MDAI2014), 2014.10.30, 筑波大学文京キャンパス, 東京

Yousuke Kaizu, Sadaaki Miyamoto, Asymmetric K-Medoids, 11th International Conference on Modelling for Artificial Intelligence (MDAI2014), 2014.10.30, 筑波大学文京キャンパス, 東京

Tatsuya Higuchi, Sadaaki Miyamoto, c-Regression Models on Vertices for Graphs, 11th International Conference on Modelling for Artificial Intelligence (MDAI2014), 2014.10.30, 筑波大学文京キャンパス, 東京

Sadaaki Miyamoto, Variations of K-Means Using the Concepts of Rough Sets, 2014 IEEE International Conference on Granular Computing (GrC2014) 招待講演, 2014.10.24, 登別グランドホテル, 北海道登別市

Sadaaki Miyamoto, Theories of Clustering, Web Intelligence Congress (WIC2014), 招待講演, 2014.8.12 University of Warsaw, Warsaw, Poland

〔その他〕

ホームページ:

<http://soft.risk.tsukuba.ac.jp/miyamoto>

6. 研究組織

(1) 研究代表者

宮本 定明 (MIYAMOTO, Sadaaki)

筑波大学・名誉教授

研究者番号: 6 0 1 4 3 1 7 9

(2) 研究分担者

遠藤 靖典 (ENDO, Yasunori)

筑波大学・システム情報系・教授

研究者番号: 1 0 2 6 7 3 9 6

(3) 研究協力者

Vicenc Torra

Professor, University of Skovde, Sweden