

令和 4 年 5 月 29 日現在

機関番号：34419

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K12146

研究課題名(和文) 構造的ゆらぎを伴うネットワークデータに対するクラスタリング手法の拡張と高度化

研究課題名(英文) Extending and Advancement of Clustering Methods for Network Data with Structural Fluctuations

研究代表者

濱砂 幸裕 (Hamasuna, Yukihiro)

近畿大学・情報学部・准教授

研究者番号：70610559

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：本研究課題では、構造的ゆらぎを伴うネットワークデータマイニングの実現に向けた新たなクラスタリングの方法論の構築に取り組んだ。はじめに、先行研究で得られた知見を発展させることで、ネットワークデータに伴う構造的ゆらぎを扱う数理モデルを構築した。次に、構築したモデルに基づくクラスタリング手法の新規開発を行った。その後、構築したモデルを時系列データの解析に援用することで、本研究において構築した数理モデルの汎用性を示した。研究全体を通じて、様々な数値実験を行うことで、開発手法が持つ分類性能・実行時間・扱えるデータの規模などの項目について、既存手法との比較評価を行い、開発手法の実用化に向けて取り組んだ。

研究成果の学術的意義や社会的意義

本研究課題では、大規模データのマイニングを目的に、ネットワークデータに伴う構造的ゆらぎに対する知識のモデル化とクラスタリング手法の開発に取り組んだ。さらに、数理モデルおよび開発手法について、理論的検討および数値実験を通じて得られた知見により、アルゴリズムのみならずクラスタリングの方法論について、包括的發展に取り組んだ。これらの成果により、大規模ネットワークデータに隠された因果関係や相互作用を明らかにするデータマイニングの実現に向けた方法論の基盤を築いた。

研究成果の概要(英文)：This research project aimed to establish a novel cluster analysis framework to achieve data mining for network data with structural fluctuations. First, mathematical models to handle network data with structural fluctuations are investigated based on the previous research. Second, clustering methods based on proposed mathematical models are constructed. Next, the versatility of investigated mathematical models is verified to apply the models for time-series data clustering. The proposed clustering methods are organized throughout the research project through comparative numerical experiments with conventional methods regarding cluster partition, execution time, and data scale.

研究分野：機械学習

キーワード：クラスタリング ネットワークデータ 構造的ゆらぎ 機械学習 ソフトコンピューティング ガウス過程 サイズコントロール

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

情報通信技術および計測技術の飛躍的発展により、蓄積されたビッグデータを活用した知識発見の重要性は極めて強く認識されている。ビッグデータ解析の一例として、ノードとエッジで構成されるネットワークデータのクラスタリング(ネットワーククラスタリング)が注目を集めており、SNS 上のコミュニティ抽出によるトレンド予測、交通ネットワークの解析による感染症の流行予測など、大規模ネットワークデータに潜む因果関係や相互作用を解析するネットワーククラスタリングは大きな期待を寄せられている。しかし、情報通信技術および計測技術の発達した現代においても、あらゆることについて正確なデータが観測されているわけではない。例えば、ネットワーククラスタリングにおいては、エッジの欠損や誤差、情報の遅れや動的変化などの構造的ゆらぎを避けることは事実上不可能である。また、大規模ネットワークデータの利活用が不可欠であるリスク管理やマーケティングなどの分野では、データに伴う不確実性の解析こそが現象の本質を明らかにする、という見方がされており、通常のデータマイニングを上回る価値が存在すると認識されている。

そのような背景から、研究代表者はネットワークデータのゆらぎに対する知識ベース構築とモデル開発を通じた新たなデータ解析の方法論構築を進めてきた。これまでの研究では、分類性能、欠損データの解析、ネットワーク構造の数理モデル開発などにおいて、既存研究を上回る成果を得ている。しかし、実データの解析を進める過程で、これまでに想定されていなかったデータの規模・解析の目的・ユーザの要求といった新たな課題が明らかとなった。そこで本研究課題では、これらの課題を解決するために、実世界の大規模・不確実・不確定なネットワークデータに対する新たなデータ解析の方法論構築に取り組んだ。簡潔に述べると、本研究課題の概要は、人間が持つ知識とデータ解析の融合による大規模ネットワークデータマイニングの実現である。

2. 研究の目的

本研究課題では、構造的ゆらぎを伴うネットワークデータのマイニングをターゲットとして進めてきた先行研究の継続的発展を通じて、実世界の大規模・不確実・不確定なネットワークデータの解析を可能とする知識融合型クラスタリング技法の高度化に取り組んだ。特に、ネットワークデータの構造的ゆらぎに対する数理モデルと知識ベースの深化を通じて、構造的ゆらぎを活用する新たなデータ解析の方法論を構築するという観点から、以下3つの課題を設定する。

- (1) 構造的ゆらぎを扱う数理モデルの拡張に基づく知識融合型クラスタリングの高度化
- (2) 新たなデータ化政期の方法論を構築するための知識ベースの深化と体系化
- (3) 実世界の大規模・不確実・不確定なネットワークデータに対するマイニングの実用化

以上を踏まえ、本研究では、構造的ゆらぎに対する人間の知識を抽出し、位相データ解析、ソフトコンピューティング、機械学習の技法を用いて、知識融合型クラスタリングの開発を進めた。また、開発したモデルを類型に分類し、汎用性を高めることで、知識ベースに基づくアプローチが実世界の大規模データを柔軟に処理する糸口となる可能性について検討した。さらに、ベンチマークデータおよび実データの解析事例を蓄積し、大規模データマイニングの実現に必要な項目の向上を図った。

本研究課題で想定しているマルチエージェントシミュレーション、IoT センサーネットワーク、SNS などから得られるソーシャルデータは、個体差、通信の途絶、突発的なアクシデントにより、その構造が著しく変容することが想定される。このような特性を持つデータをマイニングし、新たな知見を獲得するには、解析対象に関する知識、すなわち構造的ゆらぎに対する知識ベース構築およびモデル化が不可欠である。そこで、これまで蓄積した成果の継続的発展から、同じ条件下でも全く異なるデータが得られる広範な実問題の解析対象を柔軟に処理する新たなクラスタリング手法の開発を進めた。次に、実データ解析を実現するモデルを他の機械学習手法へ援用することで、大規模データに対する新たな方法論の構築に取り組んだ。さらに、様々な特性を持つ大規模ネットワークデータの解析事例を蓄積し、実用化に必要な項目の向上を図ることで、大規模データマイニングを促進する基盤の構築を進めた。本研究課題は、大規模データマイニングの実現が緊要である現状において、直面している喫緊の課題に対する解決策を示す研究である。

3. 研究の方法

本研究では、構造的ゆらぎを伴うネットワークデータに対する知識ベースと知識融合型クラスタリングの発展を通じて、大規模データマイニングの実用化を進めた。本研究の骨子は次の6項目からなる。概ね、課題(1)に項目・、課題(2)に項目・、課題3に項目・が該当する。

項目：構造的ゆらぎを伴うネットワークデータに対する数理モデルの拡張

先行研究で構築した知識ベースを拡張し、データの規模・解析の目的・ユーザの要求などをクラスタリング手法へ組み込むためのモデル化を行った。知識のモデル化には、位相データ解析、カーネル法、サイズコントロールを用いた。

項目 : ネットワークデータに対する知識融合型クラスタリングの新規開発

項目 の検討を基に、知識融合型クラスタリング技法の開発を進めた。ここで開発した技法とは、独立した手法だけではなく、ネットワークデータに対する知識融合型クラスタリングのフレームワークを含んでいる。具体的には、ネットワークデータを分割するためのアルゴリズム、分割結果を評価するための評価指標、実データ解析・応用における性能などがクラスタリングのフレームワークに構成される。

項目 : ベンチマークテスト、実データ解析による開発手法の評価および特徴把握

Stanford Large Network Dataset Collection で公開されているベンチマークデータ、RoboCup Rescue における各エージェントの位置情報および地図情報などから取得した比較的規模の大きい実データを用いて開発手法の評価を行った。特に、分類規則の特徴、適切に処理できるデータの規模、計算時間について、Louvain 法やスペクトラルクラスタリングなどとの比較を行った。ここで得られた結果を基に、項目 ・ について再検討を行った。

項目 : 知識ベースの効果検証および類型分類を通じた汎用性の向上

項目 の検討を基に、有効なモデルをネットワークデータ以外にも適用できるように拡張した。特に、時間変化する情報を表現する時系列データに対する有効性について、開発したモデルの有効性を検証した。また、開発したモデルのクラスタリング以外のデータ解析手法への応用の可能性を検討するべく、ガウス過程回帰に基づく c-回帰モデルへの適用について検討した。また、数値実験から得られた結果を定量的・定性的に評価することで、項目 ・ の再検討を行った。

項目 : 知識融合型クラスタリング技法の再開発

項目 までを通じて構築した知識ベースを基に、検討が不十分な領域を明らかにし、該当する手法の再開発および評価を行った。特に、位相データ解析、サイズコントロール、カーネル法をはじめとして、本研究課題の知識ベースに適切な諸技法についても再検討した。

項目 : 実用化に向けた知識ベースの高度化および実問題への応用

実データの検討を通じて、分類性能・データの規模・複雑さなどの観点から、本研究課題で開発した知識融合型クラスタリングの特徴を明らかにした。解析対象として、公開されているネットワークデータおよび時系列データに関するベンチマークデータおよび RoboCup Rescue から得られる比較的規模の大きいデータを用いて検証を行った。本項目における数値実験については、研究協力者である大学院生らとともに解析事例を蓄積し、実用化に必要な項目の包括的發展を図った。

4. 研究成果

研究計画の初年度から最終年度にかけて、上記に記載した研究項目 から を概ね番号順に実施した。課題(1)として設定した「構造的ゆらぎを扱う数理モデルの拡張に基づく知識融合型クラスタリングの高度化」に関連して、項目 ・ に取り組み、複数のクラスタリング手法の新規開発に取り組んだ。具体的には、データ構造の局所的構造を考慮したクラスタ分割を行う手法であるカーネル密度推定と JS ダイバージェンスを用いたクラスタリング手法、クラスタ内に含まれる個体数を制約条件として扱うサイズコントロール付きネットワーククラスタリング、クラスタ内に含まれるエッジの総量を制約条件として扱うエッジコントロール付きネットワーククラスタリングを手法の開発を進めた。これらの手法は、データの局所的な構造や得られるクラスタ分割に関する情報を確率モデルや制約条件として扱うことで、より複雑なクラスタリングアルゴリズムを構成する手法であり、従来手法よりも良好なクラスタ分割を得ることを実験的に示した。また、開発手法で使用するパラメータが与える影響について、複数のベンチマークデータを用いて検証した。これらの研究成果については、国内研究会 2 件、査読付き国際会議 2 件、査読付き学術論文雑誌 1 件で発表している。

次に、課題(2)として設定した「新たなデータ化政期の方法論を構築するための知識ベースの深化と体系化」に関連して、項目 ・ に取り組み、ベンチマークデータ・実データ解析による開発手法および知識ベースの効果について検証した。具体的には、ネットワーククラスタリング手法を用いた RoboCup Rescue から得られるエージェントの位置情報および地図情報の解析、クラスタ分割の結果として得られるボロノイ図に基づくクラスタ妥当性基準の開発を進めた。これらの研究では、ネットワークデータへの変換やボロノイ図を考慮することで既存手法よりも良好な結果を得ることを示した。これらの研究成果については、国内研究会 4 件で発表してい

る。

上記に加えて、課題(3)として設定した「実世界の大規模・不確実・不確定なネットワークデータに対するマイニングの実用化」に関連して、項目・ に取り組み、知識融合型クラスタリングの再開発と実問題への応用について検討した。具体的には、これまでの研究により検証した中で有用な数理モデルと考えられたカーネル法に基づく手法として、ガウス過程に基づく c -回帰法を提案し、その有用性と実データを扱う際の問題点について明らかにした。また、開発した数理モデルを時系列データへ援用することで、従来手法よりも良好なクラスタ分割を得る手法を構築した。さらに、実問題への応用として、時系列データに対する異常検知にも着手した。これらの研究成果については、国内研究会4件、査読付き国際会議2件、査読付き学術論文雑誌1件で発表している。

これまでの検討により、本研究課題における成果は以下に総括される。

課題(1) 構造的ゆらぎを扱う数理モデルの拡張に基づく知識融合型クラスタリングの高度

データ構造の局所的構造を考慮した数理モデルの開発

ネットワークデータの構造を考慮した数理モデルの開発

開発した数理モデルに基づくクラスタリング手法の構築

課題(2) 新たなデータ化政期の方法論を構築するための知識ベースの深化と体系化

ベンチマークデータ・実データ解析による開発手法の評価

クラスタ分割の構造に着目した評価指標の構築

課題(3) 実世界の大規模・不確実・不確定なネットワークデータに対するマイニングの実用化

開発した数理モデルに基づく機械学習の高度化の検討

開発した数理モデルに基づく時系列データの解析と異常検知への応用

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Yuto Kingetsu, Yukihiro Hamasuna	4. 巻 25
2. 論文標題 Jensen-Shannon Divergence-Based k-Medoids Clustering	5. 発行年 2020年
3. 雑誌名 Journal of Advanced Computational Intelligence and Intelligent Informatics	6. 最初と最後の頁 226-233
掲載論文のDOI（デジタルオブジェクト識別子） 10.20965/jaciii.2021.p0226	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Yukihiro Hamasuna, Yuto Kingetsu, Shusuke Nakano	4. 巻 11676
2. 論文標題 k-medoids Clustering based on Kernel Density Estimation and Jensen-Shannon Divergence	5. 発行年 2019年
3. 雑誌名 Modeling Decisions for Artificial Intelligence. MDAI 2019. Lecture Notes in Computer Science	6. 最初と最後の頁 272,282
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-26773-5_24	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 津田 暢彦, 濱砂 幸裕, 遠藤 靖典	4. 巻 33
2. 論文標題 時系列データに対する3種類のサイズコントロールクラスタリング	5. 発行年 2021年
3. 雑誌名 知能と情報	6. 最初と最後の頁 608 ~ 616
掲載論文のDOI（デジタルオブジェクト識別子） 10.3156/jssoft.33.2_608	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Hamasuna Yukihiro, Nakano Shusuke, Endo Yasunori	4. 巻 12898
2. 論文標題 Network Clustering with Controlled Node Size	5. 発行年 2021年
3. 雑誌名 Modeling Decisions for Artificial Intelligence. MDAI 2021. Lecture Notes in Computer Science	6. 最初と最後の頁 243 ~ 256
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-85529-1_20	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計12件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 金月優斗, 瀧砂幸裕
2. 発表標題 ガウス過程を用いたファジィc-回帰の検討
3. 学会等名 第36回ファジィシステムシンポジウム (FSS2020)
4. 発表年 2020年

1. 発表者名 津田暢彦, 瀧砂幸裕
2. 発表標題 時系列データに対するサイズ調整クラスタリングに関する検討
3. 学会等名 第36回ファジィシステムシンポジウム (FSS2020)
4. 発表年 2020年

1. 発表者名 北村壮馬, 西村佳洋, 瀧砂幸裕
2. 発表標題 Louvain法を用いたRoboCupRescueにおける地図の分割とエージェント制御への応用に関する検討
3. 学会等名 第36回ファジィシステムシンポジウム (FSS2020)
4. 発表年 2020年

1. 発表者名 Nobuhiko Tsuda, Yukihiro Hamasuna
2. 発表標題 Controlled-Sized Clustering for Time-Series Data
3. 学会等名 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS) (国際学会)
4. 発表年 2020年

1. 発表者名 Yukihiro Hamasuna, Daiki Kobayashi, Yasunori Endo
2. 発表標題 On Fuzzy c-Regression Models Based on Gaussian Process Regression Models
3. 学会等名 The 17th International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 金月優斗, 瀧砂幸裕
2. 発表標題 JS ダイバージェンスを用いた k-medoids
3. 学会等名 第35回ファジィシステムシンポジウム (FSS2019)
4. 発表年 2019年

1. 発表者名 津田暢彦, 瀧砂幸裕
2. 発表標題 ボロノイ図に基づくクラスタ分割の妥当性評価
3. 学会等名 第35回ファジィシステムシンポジウム (FSS2019)
4. 発表年 2019年

1. 発表者名 大津拓登, 北村壮馬, 瀧砂幸裕
2. 発表標題 RoboCup 2D リーグにおけるパッキングレートを用いた評価
3. 学会等名 第35回ファジィシステムシンポジウム (FSS2019)
4. 発表年 2019年

1. 発表者名 北村壮馬, 大津拓登, 瀧砂幸裕
2. 発表標題 RoboCup 2D リーグに対する 5 レーン理論の実装と評価
3. 学会等名 第35回ファジィシステムシンポジウム (FSS2019)
4. 発表年 2019年

1. 発表者名 西村佳洋, 瀧砂幸裕
2. 発表標題 クラスタリングを用いた時系列データに対する異常検知におけるLocal Outlier Factorを用いた評価
3. 学会等名 2021年度SICE 関西支部・ISCIEシンポジウム
4. 発表年 2022年

1. 発表者名 越川遥太, 瀧砂幸裕
2. 発表標題 エッジサイズの制約に基づくネットワーククラスタリングの検討
3. 学会等名 2021年度SICE 関西支部・ISCIEシンポジウム
4. 発表年 2022年

1. 発表者名 西村佳洋, 瀧砂幸裕
2. 発表標題 クラスタリングを用いた時系列データに対する異常検知
3. 学会等名 第23回日本知能情報ファジィ学会九州支部学術講演会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------