

令和元年度（平成 31 年度）学内研究助成金 研究報告書

研究種目	<input checked="" type="checkbox"/> 奨励研究助成金	<input type="checkbox"/> 研究成果刊行助成金
	<input type="checkbox"/> 21 世紀研究開発奨励金 (共同研究助成金)	<input type="checkbox"/> 21 世紀教育開発奨励金 (教育推進研究助成金)
研究課題名	時系列データに対するクラスタリングの高度化	
研究者所属・氏名	研究代表者：理工学部情報学科 講師 濱砂幸裕 共同研究者：	

1. 研究目的・内容

本研究は、**大規模時系列データ解析をターゲットとして、新たな知識融合型クラスタリング技法を開発することを目的**とする。特に、不変性と呼ばれる時系列データ特有の性質に対して人間が持つ知識の抽出を通じたモデル化を進め、新たなデータ解析の方法論を確立する。本研究では、時系列データが持つクラスタ構造のモデル化に基づくアルゴリズムの構築に取り組んだ。

2. 研究経過及び成果

具体的な研究課題と本研究で明らかにすること

本研究では、**時系列データのクラスタ構造に対して人間が持つ知識の抽出を通じた知識ベース構築とモデル開発を通じて、新たなデータ解析の方法論を確立する**という観点から、以下 3 つの課題を設定した。

- 課題 1：時系列データに対する知識ベース構築と数理モデル開発
- 課題 2：開発した数理モデルに基づくクラスタリング手法の新規開発
- 課題 3：大規模時系列データに対するマイニングの実用化のための知識ベースの高度化

以降、上記の研究項目に従い、これまでの研究成果を報告する。

課題 1：時系列データに対する知識ベース構築と数理モデルの開発

時系列データが持つ固有の性質である各種の不変性、クラスタ構造などの観点から時系列データの特徴に対する知識ベースと数理モデルの開発を目的とし、以下の項目について実施した。

- (1) **時系列データに対する知識ベースのプロトタイプ開発**：先行研究および実データの事例から、時系列データに固有の特徴やクラスタ構造を調査した。さらに、時系列データに対する代表的なクラスタリング手法が生成するクラスタ構造の特性について調査した。
- (2) **クラスタリングの枠組みで時系列データに対する知識を扱う数理モデルの開発**：(1)の調査内容を基に、時系列データに対する知識抽出を行い、知識ベースのプロトタイプを開発した。その際、計算コスト、生成されるクラスタ分割、アルゴリズムの観点から、大規模データに適用可能かどうか注目して整理した。

課題 2：開発した数理モデルに基づくクラスタリング手法の新規開発

課題 1 の検討を基に、知識融合型クラスタリングの開発を目的とし、以下について実施した。

- (3) **時系列データに対する知識融合型クラスタリングの開発**：知識ベースに基づいた知識のモデル化を行い、時系列データに対する知識融合型クラスタリングを開発した。本研究に着手する以前に開発したネットワークデータに対するサイズ均等 (S. Nakano et al., SCIS&ISIS2018) を援用し、従来手法とは異なるクラスタ分割を生成する新たなクラスタリング手法のプロトタイプを開発した。開発手法に対して、小規模な人工データを用いて、既存手法との比較実験を行い、計算コストおよび生成されるクラスタ分割の特徴について検討した。

- (4) **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**：UCR Time Series Classification Archive (https://www.cs.ucr.edu/%7Eeamonn/time_series_data_2018/)などで公開されているベンチマークデータを用いて開発手法の評価を行った。分類精度、扱えるデータ規模、生成されるクラスタ構造について、既存手法と比較検討し、開発手法の性能評価を行った。分類精度については既存手法と同等以上の性能を持つことが確認できたが、データの規模が増加した際の計算コスト、データの特徴と生成されるクラスタ構造の特徴把握については現在も継続して比較実験を含めた検討を進めている。
- (5) **数理モデルの効果検証および類型分類を通じた汎用性の向上**：上記に引き続き、数値実験を行い、開発手法の基盤となった知識ベースおよび数理モデルの効果について検討した。得られた結果を基に、対象とするデータの特徴、知識ベース、数理モデル、アルゴリズムの関連性について検討し、実用化に必要な項目について検討した。
- (6) **実用化に向けた知識ベースの高度化および開発手法の包括的発展**：ベンチマークおよび実データの検討を通じて、扱えるデータの規模、生成されるクラスタ分割の特徴、アルゴリズムの関連性などの観点から、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検討した。これまでに得られた結果では、小規模のデータについては従来手法を上回る性能を示すと考えられるが、比較的規模の大きいデータに対しては、計算コストの観点からより一層の検討を進める必要があると考えられる。また、本研究では検討することのできなかった時系列データの特徴を扱うことも必要となる。それらの観点を踏まえ、本研究で開発した手法を実用化に近づけるための継続的な課題について検討した。

最終的な総括・自己評価

本研究課題の最終的な自己評価は次の3点から行うこととなる。

1. 時系列データに対する新たなクラスタリング手法を開発し、その優位性を示せたか
2. 知識融合型クラスタリングが大規模データマイニング実現の糸口となることを示せたか
3. 大規模時系列データに対するクラスタリング手法の新たな方法論を構築できたか

これまでのところ、1については十分な成果が得られたと考えており、今後も発展継続的に取り組むことで、さらなる成果を得られると考えている。2、3については、時系列データの特徴を扱う数理モデルの構築が実用化の糸口となることが考えられるため、本研究課題により得られた成果の発展を目標に設定し、継続的に検討を進める。さらに、開発した数理モデルを確率モデルの観点からも検討することで、ファジィクラスタリングなどの伝統的手法との関連性について検討を進める必要がある。

3. 本研究と関連した今後の研究計画

今後の研究計画

本研究課題により得られた成果を足掛かりとし、今後の研究計画として以下を予定している。

- (1) 実問題の解決に向けた知識ベースの再構築とクラスタリング手法の新規開発
- (2) 様々な時系列データに適用可能な新たな数理モデルの構築
- (3) 開発手法の特性に対する理論的考察

特に、加速度センサーなどから取得できる大規模実データの分析、株価や検索数などのデータマイニングへの応用、開発手法の特性に対する理論的考察に基づくアルゴリズムの高度化に取り組む。また、研究成果をまとめ、令和2年度中に査読付き学術論文雑誌へ投稿し、本研究課題を総括する。

4. 成果の発表等

発表機関名	種類(著書・雑誌・口頭)	発表年月日(予定を含む)
第36回ファジィシステムシンポジウム	口頭(国内、査読なし)	2020年9月7~9日
SCIS&ISIS2020	口頭(国際、査読あり)	2020年12月5~8日