

称号及び氏名 博士(理学) 高橋 勝稔

学位授与の日付 平成 31 年 3 月 31 日

論文名 時系列データの階層的クラスタリングを用いた分割と言葉による表現への適用に関する研究

論文審査委員 主査 藤本 典幸
副査 宇野 裕之
副査 瀬田 久和
副査 馬野 元秀

論文要旨

高橋 勝稔

[研究概要]

時系列データとは、時間順に並んだ数値データのことで、身近な例としては毎日の気温の変化や株価の変化、学生の試験の点数の推移などがある。人間は時系列データや、それを視覚的に表したグラフなどの情報に対して優れた処理能力を持っている。その中で代表的なものとして、データの言葉による表現、データの期間への分割、類似時系列の検索などがある。これらの処理を計算機上で行うことは人間と機械のコミュニケーションという観点で重要である。

時系列データを解析するには、確率モデルなどが用いられるが、我々人間が時系列データを理解するときは言葉で理解していると思われる。時系列データを言葉で表現する方法としては、J. Kacprzyk らが提案した方法や馬野、小泉らが提案した方法がある。前者の方法は、“Among all segments, most are slowly increasing” や “Among all short segments, most are increasing” などのように、時系列データをゆるやかな線形性を持つ期間の集まりと捉え、線形でなくなる時刻で分割し、各期間の長さやデータの傾きにより表現する。この方法は、外れ値や振動の部分が細かく分割され、うまくその様子を捉えることができない。

後者の方法は、「全体的には少し減少しており、後期に少し増加している部分がある」などのように、全体的傾向と局所的特徴に基づいて表現する。この方法は、まず時系列データを期間のファジィ集合で分割し、各期間での代表値を求めて、それらの変化量から全体的傾向の表現を決定する。また、元の時系列データと全体的傾向の時系列データとの差の絶対値が大きい所の時刻と値を局所的特徴として表現する。表現で用いる言葉は、時系列データのグラフを表す語彙（少し増加、中ぐらい減少、下に凸など）の中からグラフと一致度が大きいものを一致度とともに選り出す。さらに、振動している時系列データを表現する方法として、馬野、岡村らが提案した方法があり、元の時系列データと全体的傾向を表す時系列データの差の標準偏差を用いて振動の大きさを表現し拡張された増減回数により振動の細かさを表現している。

本論文では、馬野、小泉らの方法の改良を行う。従来は「前期」、「中期」、「後期」に対して図1の緑色の実線のような全期間を均等に分割したファジィ集合を用いていたが、より適切な言葉による表現を得るために、データに基づいて期間のファジィ集合を調整する方法を提案する。例えば、図1では時刻4で減少から増加

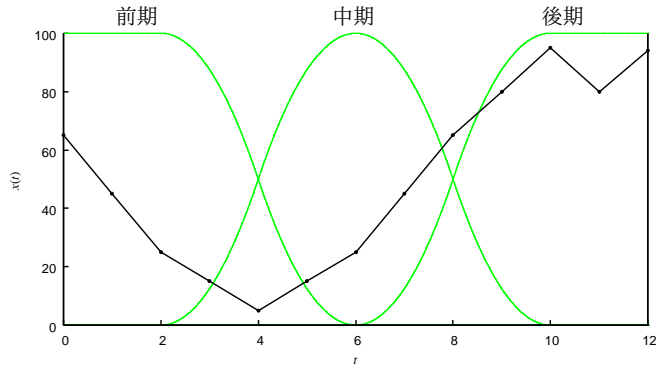


図 1: 期間のファジィ集合

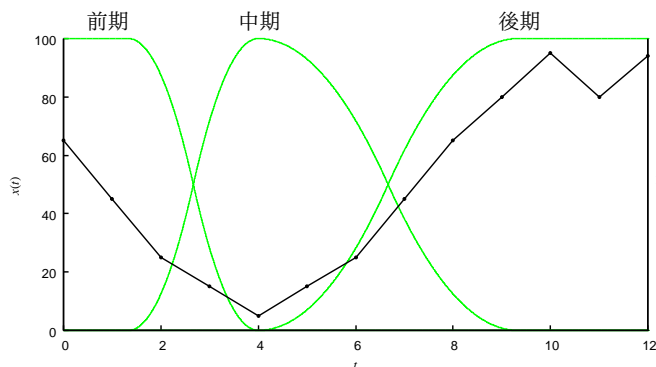


図 2: 調整された期間のファジィ集合

に傾向が変わるので、図 2 のように時刻 4 が中期の真ん中となるように調整する。そのために、データが変化している時刻で時系列を期間に分割する方法を提案する。

時系列データの分割については、J. Kacprzyk らの言葉による表現でデータを分割するとき用いられた J. Sklansky らの線形性を持つ期間の集まりに分割する方法や、P. Wang らの方法や松原、櫻井らの方法のように隠れマルコフモデル (HMM: Hidden Markov Model) を利用して分割する方法などが提案されている。これらの方法は線形性やマルコフ性があることを仮定しているために、適用できない時系列データも少なくない。また、図 1 のように 3 つの期間に分割することを考えると、期間の数を指定できないなどの問題もある。本論文では、まず一般的に階層的クラスタリングを用いて時系列データを適切な数の期間に分割する方法を提案する。この方法では、分割された各期間をクラスタとみなし、隣り合うクラスタ間の類似性が大きいものをまとめていく。類似性としてクラスタの値の大きさ、変化の大きさ、振動の大きさの変化量 (の絶対値) を考え、これらの変化量の重み付き平均の小さいものが類似していると考え。このとき、重みを固定するとうまくクラスタリングができないので、クラスタの大きさや各変化量の関係によって重みを動的に変化させる。また、振動している部分を 1 つのクラスタにまとめるために、小さいクラスタをまとまりやすくするパラメータを導入する。さらに、はずれ値の部分を超えて前後のクラスタがまとまるように、1 つ越えたクラスタの類似性も考慮する。こ

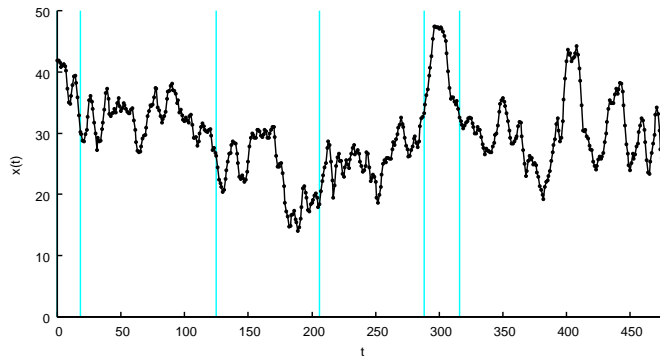


図 3: 長期間の時系列データの例

の方法は、直前からの値の変化や、確率的な切り替わりで分割するのではなく、入力された時系列データ全体の傾向や特徴で期間に分割している。

この方法を期間のファジィ集合の調整に適用する。クラスタ数が2の境界が前期と後期の境界（すなわち、中期の真ん中）になると考え、期間のファジィ集合を決める。例えば、図1の時系列データでは時刻4で2分割され、図2のように時刻4が中期の真ん中となる期間のファジィ集合が得られる。これにより、従来の方法と比べて、より適切な言葉による表現が得られる。

また、この分割の方法を用いて、図3のような長期間の時系列データを複数の期間に分割し、各期間毎に言葉で表現することで傾向の推移を表現できる。図3はある終了条件では6つに分けられ（各期間の境界は青色の実線で表されている）、左から順に、「大きく減少」、「中ぐらゐの値でほぼ一定」、「低い値で少し上に凸」、「少し増加」、「上に凸」、「中ぐらゐ振動しながらほぼ一定」となる。

時系列データの言葉による表現の応用として、時系列データを用いた異常の検知や予測などが挙げられる。本論文では、その第一歩として、得られた言葉による表現が似ている時系列データを検索する方法を提案する。この方法は、各時刻の値の差の二乗和など値そのものを用いた非類似度よりも人間の感覚に近い検索結果が得られる。また、言葉による表現を用いるのでデータ数の違いにも対応できる。まず、似た言葉間を考慮するために言葉間の類似度を用いて表現の拡張を行い、複数の言葉間の類似度を求める。そして、これを基に似た時系列データを検索する。検索結果は、期間のファジィ集合の調整をした場合の方が、妥当な結果が得られる。

以下、第2章では、本論文で必要とするファジィ理論、時系列データの言葉による表現、類似時系列の検索について述べる。第3章では階層的クラスタリングを用いて時系列データを分割する方法を提案する。第4章では階層的クラスタリングを用いて期間のファジィ集合を調整し、より適切な言葉による表現を得る方法を提案する。また、長期間の時系列データの言葉による表現について述べる。さらに、期間の調整を行った言葉による表現を用いた類似時系列の検索について述べる。第5章では、本研究で得られた結果をまとめ、今後の課題を示す。

論文 (査読あり)

1. Retrieval of Similar Time Series with Similarity Degree of Linguistic Expressions for Global Trend and Local Features, K. Takahashi and M. Umamo, *Proc. of 2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2012)*, pp.817–824, Brisbane, Australia, 2012.
2. Partition of Time Series Using Hierarchical Clustering K. Takahashi, M. Umamo and N. Fujimoto, *Proc. of Joint 11th International Conference on Soft Computing and Intelligent Systems and 19th International Symposium on Advanced Intelligent Systems*, pp.843–848, Toyama, Japan, 2018.
3. 階層的クラスタリングを用いた時系列データの分割, 高橋 勝稔, 馬野 元秀, 藤本 典幸, 知能と情報 (日本知能情報ファジィ学会誌), Vol.31, No.2 (2019) に掲載予定.

口頭発表等

1. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —期間のファジィ集合の決定—, 馬野 元秀, 高橋 勝稔, 瀬田 和久, 日本知能情報ファジィ学会 第 25 回ファジィシステムシンポジウム, 茨木, 5 ページ (ページ番号なし) (2009).
2. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —言葉による表現の類似度—, 馬野 元秀, 高橋 勝稔, 瀬田 和久, 日本知能情報ファジィ学会 第 26 回ファジィシステムシンポジウム, 広島, pp.102–107 (2010).
3. 言葉による表現の類似度を用いた類似時系列データの検索, 高橋 勝稔, 馬野 元秀, 日本知能情報ファジィ学会 第 27 回ファジィシステムシンポジウム, 福井, pp.496–501 (2011).
4. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —値の変化の全体的傾向や局所的特徴をもちいて—, 高橋 勝稔, 馬野 元秀, 日本知能情報ファジィ学会 第 28 回ファジィシステムシンポジウム, 愛知, pp.1044–1049 (2012).

5. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —OWA 演算子による代表値の算出—, 高橋 勝稔, 馬野 元秀, 日本知能情報ファジィ学会 第 29 回ファジィシステムシンポジウム, 大阪, pp.810–815 (2013).
6. 値にあいまいさを含む時系列データの言葉による表現, 馬野 元秀, 岩浅 貴大, 高橋 勝稔, 日本知能情報ファジィ学会 第 29 回ファジィシステムシンポジウム, 大阪, pp.804–809 (2013).
7. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —データの傾向に基づく期間のファジィ集合の決定—, 馬野 元秀, 岩浅 貴大, 高橋 勝稔, 日本知能情報ファジィ学会 第 30 回ファジィシステムシンポジウム, 高知, pp.808–813 (2014).
8. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —振動回数が多の場合の表現—, 馬野 元秀, 島野 佳之, 高橋 勝稔, 日本知能情報ファジィ学会 第 30 回ファジィシステムシンポジウム, 高知, pp.8014–819 (2014).
9. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —振動を考慮した期間のファジィ集合の決定—, 岩浅 貴大, 高橋 勝稔, 馬野 元秀, 計測自動制御学会 第 6 回コンピューターショナル・インテリジェンス研究会, 大阪, pp.40–45 (2014).
10. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —振動を取り出すための三角移動平均のパラメータの調整—, 島野 佳之, 高橋 勝稔, 馬野 元秀, 計測自動制御学会 第 6 回コンピューターショナル・インテリジェンス研究会, 大阪, pp.46–51 (2014).
11. 全体的傾向と局所的特徴に基づく時系列データの言葉による表現 —上側と下側の概形を用いた振動の表現—, 馬野 元秀, 島野 佳之, 高橋 勝稔, 日本知能情報ファジィ学会 第 31 回ファジィシステムシンポジウム, 東京, pp.300–305 (2015).
12. 言葉による表現を用いた類似時系列検索の診断への応用, 高橋 勝稔, 濱田 賢吾, 馬野 元秀, 日本知能情報ファジィ学会 第 31 回ファジィシステムシンポジウム, 東京, pp.580–585 (2015).
13. 階層的クラスタリングを用いた時系列データの分割, 高橋 勝稔, 馬野 元秀, 田邊 朋加, 日本知能情報ファジィ学会 第 32 回ファジィシステムシンポジウム, 佐賀, pp.597–602 (2016).
14. 時系列における異常状態に対する前兆傾向の抽出, 高橋 勝稔, 塩澤 大輝, 馬野 元秀, 日本知能情報ファジィ学会 第 32 回ファジィシステムシンポジウム, 佐賀, pp.607–612 (2016).

15. 言葉で表現された前兆データにより学習したニューラルネットワークを用いた時系列データの異常検知, 高橋 勝稔, 濱田 賢吾, 馬野 元秀, 日本知能情報ファジィ学会 第 32 回ファジィシステムシンポジウム, 佐賀, pp.613–618 (2016).
16. 階層的クラスタリングを用いた時系列データの分割 —ヒープを用いた一括クラスタリング—, 高橋 勝稔, 馬野 元秀, 藤吉 誠, 藤本 典幸, 日本知能情報ファジィ学会 第 33 回ファジィシステムシンポジウム, 山形, pp.19–24 (2017).
17. 忘却型ファジィ・ニューラルネットワークを用いた時系列データの異常検知, 高橋 勝稔, 塩澤 大輝, 馬野 元秀, 藤本 典幸, 藤吉 誠, 日本知能情報ファジィ学会 第 33 回ファジィシステムシンポジウム, 山形, pp.25–30 (2017).
18. 階層的クラスタリングを用いた時系列データの分割手法のオンライン環境への適用, 伊瀬 顕史, 高橋 勝稔, 馬野 元秀, 藤本 典幸, 日本知能情報ファジィ学会 第 34 回ファジィシステムシンポジウム, 愛知, pp.565–570 (2018).

学位論文審査結果の要旨

時系列データとは、時間順に並んだデータのこと、例えば、毎日の気温や株価の変化などがその例である。人間は時系列データに対して優れた処理能力を持っており、例えば、時系列データを言葉により簡潔に表現したり、長期の時系列データを適切な期間に分割したり、2つの時系列データが似ているかどうかを判定することが容易にできる。

最近では、与えられた時系列データを「全体的には少し減少しており、後期に少し増加している部分がある」などのようにファジィ的な言葉を用いて表現する方法が提案されている。この学位論文では、まずその方法ではうまく表現できない場合があることを指摘し、それを改善する方法を提案している。そのために、まず階層的クラスタリングを用いて時系列データを必要な数の期間に分割する方法を提案し、それにより言葉の表現を改良している。そして、言葉による表現を用いて、似ている時系列を検索する方法を提案している。

本論文の主要な成果は次の通りである。

- (1) 時系列データの言葉による表現に関して、前期、中期、後期のファジィ集合をデータに基づいて決定する必要があることを指摘し、時系列データを必要な数の期間に分割する方法(項目(2))を用いて、それらのファジィ集合を決定する方法を提案している。
- (2) 時系列データを期間に分割する方法に関して、分割された各期間をクラスタとみなし、階層的クラスタリングを用いて隣り合うクラスタ間の類似性が大きいものをまとめていく方法を提案している。このとき、類似性としてクラスタの値の大きさ、変化の大きさ、振動の大きさに関する変化量(の絶対値)を考え、これらの重み付き平均の小さいものが類似していると考えている。そして、重みを固定するとうまくクラスタリングできないことを例により示し、各変化量の関係やクラスタの大きさによって重みを動的に変化させる方法を提案している。さらに、小さいクラスタをまとまりやすくしたり、はずれ値の前後のクラスタがまとまるようにする工夫も行っている。この方法をごみ焼却発電施設の実際のセンサデータに適用し、その有効性を示している。
- (3) 似ている時系列の検索に関して、表現された言葉が類似しているものを検索する方法を提案している。言葉の類似性を用いることで、より人間の感覚に近い検索結果が得られることをシミュレーションにより示している。

本研究の成果を用いると、時系列データのある程度適切な期間に分割することが可能になる。これにより、言葉を用いて時系列データをうまく表現できるようになり、類似検索の結果も人間の感覚に近いものにすることができる。これは、計算機利用の発展に大きく貢献するものと考えられる。また、時系列の期間への分割の手法は新たに提案されたもので、さまざまな分野に適用でき、時系列データの利用に大きく貢献するものである。

以上のことから、本委員会は本論文の審査、最終試験の結果に基づき、高橋勝稔氏に博士(理学)の学位を授与することを適当と認める。

学位論文審査委員会

主査 藤本 典幸
宇野 裕之
瀬田 和久
馬野 元秀