

機関番号：24403

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500209

研究課題名（和文）時系列データの言葉による表現とその検索および機械学習への応用

研究課題名（英文）Linguistic Expression of Time Series and Its Applications to Retrieval and Machine Learning

研究代表者

馬野 元秀 (UMANO MOTOHIDE)

大阪府立大学・理学系研究科・教授

研究者番号：10131616

研究成果の概要（和文）：時系列データとは、時間の経過に従って計測されたデータで、毎日の気温や株価の変化など、我々の身近に数多く存在している。我々は、時系列を言葉で表現された全体的な傾向と局所的な特徴により理解していると思われる。そこで、与えられた時系列データの全体的な傾向と局所的な特徴と振動を言葉により表現する方法を定式化し、その方法に基づいて時系列の全体的な傾向と局所的な特徴と振動を言葉により表現するシステムを作成し、その応用として、類似時系列検索システムと時系列知識学習システムのプロトタイプを作成した。検索は全体的傾向の類似度と局所的特徴の類似度と振動の類似度の重み付き平均の値を計算し、その順に必要なものを検索する。また、知識学習システムでは、時系列の全体的傾向と局所的特徴と振動の属性を追加し、ファジィ C4.5 を用いて知識を獲得する。

研究成果の概要（英文）：We have many kinds of data of time series such as daily temperatures and stock prices. It is likely that we understand them via linguistic expressions for the global trend and local features. In this research, we formulate a method to extract linguistic expressions for the global trend, local features and oscillation using fuzzy sets and implement a system for extracting them from a time series data. Then we implement two prototype systems for retrieving similar time series in a database of time series and learning knowledge from a set of data including time series. The retrieval of similar time series is based on a similarity degrees of linguistic expressions for global trend, local features and oscillation. The learning of knowledge is an extension of fuzzy C4.5 to including attributes of global trend, local features and oscillation.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	1,100,000	330,000	1,430,000
2012年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：複合領域

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング

キーワード：ファジィ理論

1. 研究開始当初の背景

時系列データとは、時間の経過に従って計測されたデータで、毎日の気温や株価の変化など、我々の身近に数多く存在している。

我々は、以前からファジィ集合を含む問合せによりデータベースから必要なデータを検索するシステム[1]や数値と記号のデータ集合からファジィ知識を学習するシステム[2][3]に関する研究を行ってきた。データには、気温や株価のように基本的に時系列であるものが多く存在しているが、検索において時刻属性や時間属性を用いることはできても、時系列としての性質をうまく用いることができなかった。また、知識を学習する場合でも、時系列の性質をうまく用いれば、診断や予測の性能を大幅に改善することが可能のように思われる(例えば、故障診断では、通常と異なる時系列の変化を見つければよいし、予測では時系列としの変化の仕方が重要になる)が、うまく利用することができなかった。

時系列データを検索や知識学習で利用する場合には、結果の良し悪しが人間の判断に密接に関係しているため、人間がどのようにして時系列データを表現・理解しているかが重要となる。例えば、図1のような時系列データ(ある時期のレギュラーガソリンの全国平均店頭価格のデータ)が与えられると、人間は、例えば「全体的にはやや下に凸である」というように、全体的な傾向を言葉で表現することによって大雑把に理解していると思われる。そして、全体的な傾向だけで表現できない場合には、「後期において少し減少している部分がある」(17ヶ月目あたりで少し減少している)のような局所的な特徴にも注目すると思われる。すなわち、言葉で表現された全体的な傾向と局所的な特徴により、時系列データを理解している。

時系列データを解析する方法では、自己回帰モデルなどの確率過程モデルを用いてデータの変化のしかたを詳細に調べるのが普通で[4]、時系列データの特徴を言葉で表現し

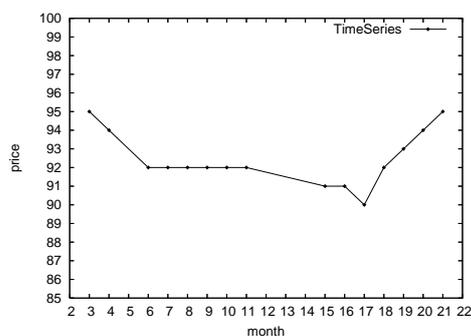


図1. 時系列データの例

ようという研究はほとんどない。そのなかで、J. Kacprzykらは時系列データの中の細かい特徴を言語的に表現する研究[5]を行なっているが、この方法では時系列データ全体を理解できない。

- [1] Umamo and S. Fukami: "Fuzzy Relational Algebra for Possibility-Distribution-Fuzzy-Relational Model of Fuzzy Data," *Journal of Intelligent Information Systems*, Vol.3, No.1, pp.7-27, 1994
- [2] K. Hori, M. Umamo, H. Satoh and Y. Uno: "Fuzzy C4.5 for Generating Fuzzy Decision Trees and Its Improvement," *Fourth Asian Fuzzy Systems Symposium*, pp.881-884, 2000
- [3] 松本, 馬野, 籾丸, 瀬田: 「知識表現を変化させる学習について」, 知能と情報(日本知能情報ファジィ学会誌), Vol.19, No.3, pp.276-286, 2007
- [4] 田中: 現代時系列分析, 岩波書店, 2006
- [5] J. Kacprzyk and A. Wilbik: "Linguistic Summarization of Time Series Using Linguistic Quantifiers: Augmenting the Analysis by a Degree of Fuzziness," *2008 IEEE International Conference on Fuzzy System*, pp.1146-1153, 2008

2. 研究の目的

本研究では、与えられた時系列データの全体的な傾向と局所的な特徴を言葉により表現するシステムを作成し、その応用として、時系列データ検索システムと時系列知識学習システム(のプロトタイプ)を作成することを目的とする。

まず、与えられた時系列データの全体的な傾向と局所的な特徴を抽出し、それをファジィ集合で定義された言葉により表現する方法を定式化し、時系列データを言葉で表現することができるシステムを作成する。

そして、その応用として、言葉による表現を付加した時系列のデータベースから、言葉や時系列データを用いて時系列データを検索するシステム(のプロトタイプ)を作成する。これにより、例えば「全体的に穏やかに値上がりしてきて、最近急に値上がりした株は」という問合せによりデータベースを検索できるようになると期待できる。さらに、時系列データの集合から時系列に関する規則性を抽出するシステム(のプロトタイプ)を作成する。これにより、例えば「全体的に穏やかに値下がりしてきて、最近少し値上がりした株は「その直後に大きく値上がりしている」というような規則を抽出できるようになると期待できる。

3. 研究の方法

本研究を遂行するために、時系列データの全体的な傾向と局所的な特徴を取り出す方法を定式化し、与えられた時系列データをファジィ集合により定義された言葉で表現可能なシステムを作成する。そして、応用として、時系列のデータベースから必要な時系列データを検索するシステム(のプロトタイプ)と時系列データ集合から時系列に関する知識を学習するシステム(のプロトタイプ)を作成する。

(1) 全体的傾向と局所の特徴を取り出す方法の定式化: 次の項目の検討を行なう。

(a) 期間のファジィ集合の決め方について、等分割のものとデータに基づくものとの比較を行なう。

(b) 各期間の代表値を用いて全体的な傾向の表現を決めているが、どのような代表値を用いるのが適切かについて検討する。

(c) 全体的傾向と局所の特徴に加えて、振動のしかたを考慮に入れる。

定式化の際に、どちらがよいかの判断ができない場合には、シミュレーションの結果を用いて判断する。

(2) シミュレーション・プログラムの作成: 定式化に基づいて、シミュレーション・プログラムを作成する。

(3) 時系列データへの適用: 作成したシミュレーション・プログラムをいくつかの時系列データに適用し、定式化の妥当性を調べる。

(4) 検索手法の定式化: 時系列に関する属性を用いて時系列データを検索するための方法を定式化する。可能ならば、データベースの検索用のファジィ関係演算子として定式化する。

(5) 検索システムの作成: 定式化に基づいて、検索システム(のプロトタイプ)を作成する。いまのところ、主記憶上で稼動するものを考えている。

(6) 時系列データへの適用と結果の検討: 適当な時系列データを用いて、実際に検索プログラムを動かし、検索結果を検討する。

(7) 知識学習手法の定式化と知識学習システムの検討: 時系列に関する属性を用いて、知識を学習する手法について検討する。ファジィ C4.5 を基にして定式化を行なう。

(8) 知識学習システムの作成: 定式化に基づいて、知識学習システム(のプロトタイプ)を

作成する。

(9) 時系列データ集合への適用と結果の検討: 適当な時系列データを用いて、実際に知識学習プログラムを動かし、結果を検討する。

(10) 全体的な検討と大規模化の検討: 本研究で作成したシステムをより大規模なものへの展開が可能か検討する。

4. 研究成果

研究方法にしたがって、時系列データの全体的な傾向と局所的な特徴を取り出す方法を定式化し、与えられた時系列データをファジィ集合により定義された言葉で表現可能なシステムを作成し、応用として、時系列のデータベースから必要な時系列データを検索するシステムのプロトタイプと時系列データ集合から時系列に関する知識を学習するシステムのプロトタイプを作成した。

(1) 全体的傾向と局所の特徴を取り出す方法の定式化:

(a) 等分割の期間のファジィ集合とデータに基づくものとの比較を行なった結果、データに基づいて決める方法がよいことが分かった。具体的な方法については、全体的な傾向を表すデータと実際のデータとの差を小さくする方法と局所の特徴を少なくする方法を検討した。

(b) 各期間の代表値としては、一般的には、期間の重み付き平均を用いるのが妥当である。しかし、特定の状況では、最大値や最小値に近い値を使う方がよいと思われる場合があることが分かった。

(c) 全体的傾向と局所の特徴に加えて、振動を定式化した。

(2) シミュレーション・プログラムの作成: 定式化に基づいて、シミュレーション・プログラムを作成した。

(3) 時系列データへの適用: 作成したシミュレーション・プログラムをいくつかの時系列データに適用し、定式化の妥当性を確認した。

(4) 検索手法の定式化: 全体的傾向と局所の特徴と振動の言葉による表現を用いて、時系列データを検索するための方法を定式化した。全体的傾向の類似度と局所の特徴の類似度と振動の類似度の重み付き平均を計算し、その値の順に検索する方法である。データベースの検索用のファジィ関係演算子として定式化することはできなかった。

(5) 検索システムの作成: 定式化に基づいて、検索システムのプロトタイプを作成した。シ

システムは C 言語と Java 言語により作成した。このうち、Java 言語によるものはタブレット PC 上で時系列をタブレット上で直接入力すると、全体的傾向と局所の特徴と振動の言葉による表現を表示し、指示に従って、これと類似の時系列を検索するようになっている。

(6) 時系列データへの適用と結果の検討: いくつかの時系列データを用いて、実際に検索プログラムを動かす、ある程度妥当な検索結果が得られた。

(7) 知識学習手法の定式化と知識学習システムの検討: 時系列に関する属性を用いて、知識を学習する手法について検討し、ファジィ C4.5 を基にして定式化を行なった。新たに全体的傾向と局所の特徴という属性を追加するという方法である。

(8) 知識学習システムの作成: 定式化に基づいて、知識学習システムのプロトタイプを C# 言語を用いて作成した。

(9) 時系列データ集合への適用と結果の検討: 学習にはある程度の量の時系列データが必要であるが、今回使用した時系列データ集合では明確な知識が得られなかった。今後、データ集合の吟味と手法の検討が必要である。

(10) 全体的な検討と大規模化の検討: 手法的には、大規模なものへの展開が可能であることは確認したが、そのためにはシステムの大規模な拡張が必要であることが分かった。

5. 主な発表論文等

[雑誌論文] (計 5 件)

- (1) K. Seta, L. Cui, M. Ikeda, N. Matsuda and M. Okamoto: "Meta-Cognitive Skill Training Program for First-Year Bachelor Students Using Thinking Process Externalization Environment," *International Journal of Knowledge and Web Intelligence*, 査読有, Vol.4, 2013 (to appear)
- (2) 瀬田, 崔, 池田, 松田, 岡本: 「思考外化と知識共創によるメタ認知スキル育成プログラム — 大学初年次生を対象として」, 教育システム情報学会誌, 査読有, Vol.30, pp.77-91, 2013
- (3) 砂山, 高間, D. Bollegala, 西原, 徳永, 串間, 松下: 「Total Environment for Text Data Mining: テキストデータマイニングのための統合環境」, 人工知能学会論文誌, 査読有, Vol.26, pp.483-493, 2011
- (4) K. Seta and M. Ikeda: "An Intelligent Support System for Presentation-based Meta-Learning," *Artificial Life and Robotics*, Vol.16, 査読有, pp.352-355, 2011
- (5) K. Seta, C. Hara, M. Fujiwara, M. Umamo and M. Ikeda: "Presentation Based Meta-Learning Support System Tightens Learners' Evaluation Criteria," *Lecture Notes in Artificial Intelligence*, 査読有, Vol.6884, pp.288-297, 2011

[学会発表] (計 23 件)

- (1) K. Seta, L. Cui, M. Ikeda, M. Okamoto and M. Umamo: "Prompting Internal Self-conversation Enhances Learning by Discussion to Train Meta-cognitive Skills," *1st Asian Conference on Information Systems*, Sem Reap, Cambodia, December 6 – 8, 2012
- (2) H. Ooe, T. Kojiri and K. Seta: "Design Pattern Learning Support System Based on Redesigning," *20th International Conference on Computers in Education*, Singapore, November 26 – 30, 2012
- (3) 高橋, 馬野: 「全体的傾向と局所の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の特徴を用いて —」, 第 28 回ファジィ システム シンポジウム, 名古屋工業大学, 2012 年 9 月 12~14 日
- (4) 馬野, 立野, 岸本, 伊瀬: 「カーレースゲームへのファジィ Q 学習の適用 — 効率的に目標を通過するための学習方法の提案 —」, 第 28 回ファジィ システム シンポジウム, 名古屋工業大学, 2012 年 9 月 12~14 日
- (5) 馬野, 森下, 岸本, 瀬田: 「カーレースゲームにおけるファジィ Q 学習を用いた段階的学習の改善」, 第 28 回ファジィ システム シンポジウム, 名古屋工業大学, 2012 年 9 月 12~14 日
- (6) 馬野, 藤井, 小路, 細谷: 「実数値環境の追跡問題におけるエージェントへの段階的視覚の適用」, 第 28 回ファジィ システム シンポジウム, 名古屋工業大学, 2012 年 9 月 12~14 日
- (7) K. Takahashi and M. Umamo: "Retrieval of Similar Time Series with Similarity Degree of Linguistic Expressions for Global Trend and Local Features," *21st IEEE International Conference on Fuzzy Systems*, Brisbane, Australia, June 10 – 15, 2012
- (8) Y. Hosoya and M. Umamo: "Dynamic Fuzzy Q-Learning with Facility of Tuning and Removing Fuzzy Rules," *21st IEEE International Conference on Fuzzy Systems*, Brisbane, Australia, June 10 – 15, 2012
- (9) 末吉, 松下, 白水: 「複数の時系列データの比較に基づくグラフの言語表現生成手法」, 人工知能学会 第 1 回情報アクセスと可視化マイニング研究会, 東京大学, 2012 年 4 月 14 日
- (10) 末吉, 田中, 白水, 松下: 「比較対象に着目したグラフの言語表現の生成」, 電子情報通信学会 第 21 回 Web インテリジェンスとインタラ

- クシオン研究会, 大阪大学, 2011 年 11 月 7~8 日
- (11) 高橋, 馬野: 「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」, 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, 福井大学, 2011 年 9 月 12~14 日
- (12) 馬野, 森下, 岸本, 瀬田: 「ファジィ Q 学習を用いた段階的学習のカーレースゲームへの適用」, 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, 福井大学, 2011 年 9 月 12~14 日
- (13) 馬野, 藤井, 細谷, 瀬田: 「実数値環境の追跡問題におけるファジィ強化学習の比較」, 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, 福井大学, 2011 年 9 月 12~14 日
- (14) 馬野, 林田, 藤丸, 瀬田: 「知識表現を変化させる学習について —保存データ集合の更新方法—」, 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, 福井大学, 2011 年 9 月 12~14 日
- (15) 馬野, 西村, 藤丸, 瀬田: 「知識表現を変化させる学習について —類似データの数に応じて異なる抽象度の知識表現を用いる方法—」, 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, 福井大学, 2011 年 9 月 12~14 日
- (16) K. Seta, H. Maeno, M. Umano and M. Ikeda: “An Integrated Framework as a Foundation to Develop Meta-Learning Support Systems,” *15th International Conference on Artificial Intelligence in Education*, Auckland, New Zealand, June 26 – July 1, 2011
- (17) 藤原, 原, 野口, 瀬田: 「内発的動機付けの高まりを捉えた知識創造型学習プロセスの獲得支援」, 電子情報通信学会 2011 年度第 1 回教育工学研究会, 大阪府立大学, 2011 年 5 月 14 日
- (18) 松下, 末吉: 「言語表現による時系列データ検索のための基礎検討」, 電子情報通信学会 第 19 回 Web インテリジェンスとインタラクション研究会, 学術総合センター, 2011 年 3 月 7~8 日
- (19) 馬野, 高橋, 瀬田: 「全体的傾向と局所の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—」, 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, 広島大学, 2010 年 9 月 13~15 日
- (20) 馬野, 小谷, 瀬田: 「2 つの色表現による画像のファジィ検索」, 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, 広島大学, 2010 年 9 月 13~15 日
- (21) 馬野, 岸本, 瀬田: 「カーレースゲームへのファジィ Q 学習の適用」, 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, 広島大学, 2010 年 9 月 13~15 日
- (22) 馬野, 西村, 藤丸, 瀬田: 「知識表現を変化さ

せる学習について —知識表現を段階的に変化させる方法—」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, 広島大学, 2010 年 9 月 13~15 日

- (23) 馬野, 丸山, 藤丸, 瀬田: 「知識表現を切り換える学習用パラメータの変更 —学習用パラメータのしきい値の制御—」, 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, 広島大学, 2010 年 9 月 13~15 日

6. 研究組織

(1) 研究代表者

馬野 元秀 (UMANO MOTOHIDE)
大阪府立大学・理学系研究科・教授
研究者番号: 10131616

(2) 研究分担者

瀬田 和久 (SETA KAZUHISA)
大阪府立大学・理学系研究科・教授
研究者番号: 50304051

(3) 研究分担者

松下 光範 (MATSUSHITA MITSUNORI)
関西大学・総合情報学部・教授
研究者番号: 50396123