

カーブフィッティングと SOM アルゴリズムによる可視化分析の研究 On A Study of Data Visualization Analysis Using Curve Fitting and SOM Algorithm

馮鑫[†] 能上 慎也[‡]
Xin Feng Shinya Nogami

1. はじめに

本研究では、データの可視化を対象としており、今回はカーブフィッティングと SOM アルゴリズムを中心に、ビジネスデータを用いて可視化分析の改良モデルについて報告する。

2. 研究背景

これまでの研究において、既に様々な可視化分析手法を利用して、より効率的な可視化分析モデルが提案されてきており、筆者らもまた提案してきた^[1]。今回は、ビジネスデータを利用して、カーブフィッティングと SOM アルゴリズムに基づいた可視化分析方法を中心に報告する。

3. 先行研究

3.1 カーブフィッティング

カーブフィッティングは、平面上のそれぞれの離散点で表される座標間の関数の関係を連続曲線で近似的に描く処理方法である。このようなデータ処理の方法は、データ可視化の過程において特に重要である^[2]。

現在よく使用されている方法には、セグメント線形法がある^[3]。しかし、この方法を用いて多項式関数の係数計算を行う場合、行列の演算において、計算速度が遅く、計算時間が長くなってしまふ。これは、特に大量の時系列データの場合には、計算性能は大きな影響を与えることになる^[4]。今回は、時系列データの特徴に対して、多項式フィッティングの計算時間を減少する方法を提案し、サンプル行列の中間計算結果を再利用することにより、多項式フィッティングの計算時間を大幅に減少させることを試みる。

3.2 SOM アルゴリズム

現在の次元削減可視化方法は主に主成分分析法、投影追跡法、自己組織化マッピング法 (SOM) や多次元尺度法などがある^[5]。しかし、このような可視化方法の集積モデルは、原始データの可視化表現に集中しており、相互関連性が強くなく、演算量が大きく、現在の時空間データの全面的、高効率、動的な深い分析とマイニングの要求をうまく満たすことができない^[6]。単一の可視化方法では時空間データの多角度表示と分析の要求を満たすことができないため、統合された可視化方法を用いることになるが、これは時空間データ可視化ツールの組み合わせになり、本質を考慮し独立して時空間データを表現して説明する (図 1)。

今回の研究では、SOM がデータの次元削減を実現すると同時に、そのトポロジー構造を保持し、「次元災害」の

発生を避けることができるというところを目指す。

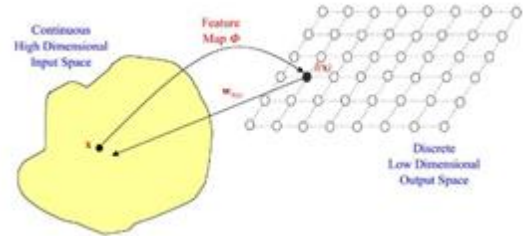


図 1 SOM アルゴリズムの表示

4. 研究目的と手法

4.1 研究目的

本研究では、カーブフィッティングモデルを改良し、予測速度の上がり、そして、SOMにより可視化分析を行う。

4.2 データソース

本研究では、ある会社 (ピンドウドウ) における売れ筋商品 TOP3 の代表的な商品を対象とする。それらの商品の 2015~2019 年の四半期別販売量データを参照し、分析を行う。

4.3 研究手法

4.3.1 カーブフィッティング

複数の時系列データをフィッティング際に、中間計算結果を再利用することができれば、計算時間現象の役割を果たすことができる。

4.3.2 SOM に基づく可視化

SOM に基づくマルチビュー協調可視化分析方法を提案する (図 2)。

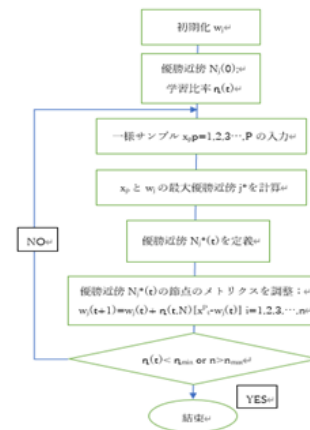


図 2 SOM による計算過程

[†] 東京理科大学大学院 Graduate school of Tokyo University of Science

[‡] 東京理科大学 Tokyo University of Science

この方法は、次元削減データを可視化するという観点から複数の可視化分析ツールを統合し、次元削減後のデータに対して可視化表現を行い、従来の可視化分析ツールでは高次元多属性時空間データを可視化できないという問題を解決するだけでなく、各表現ツール間の連動をも実現する。

5. 研究過程

5.1 カーブフィッティング

最小二乗法を用いて多項式フィッティングを行う場合、データが持つタイムスタンプから最小二乗法に必要な x 座標の値が得られるが、これらの x 座標の値を定義する系列を時刻 $T\{x_1, x_2, \dots, x_n\}$ とする。縦座標の値を時系列 Y 、 Y を T に対応する時系列データ $\{y_1, y_2, \dots, y_n\}$ とする。サンプリングの時系列データを平均セグメント化してフィッティングを行う場合には、変換後の時系列セグメントのいずれも時刻番号 T は同じであり、その後の時系列データセグメントのフィッティング計算にも、1 回目に算出した行列値をそのまま用いる。この場合、従来の 3 回の行列乗算と 1 つの行列逆演算を 1 回の行列乗算に減少させることができ、計算ステップを減少させ、フィッティングの加速を実現することができた (図 3)。

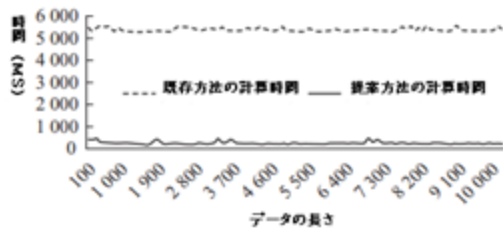


図 3 販売データを用いた計算時間の比較

5.2 SOM に基づく可視化の実現

ここでは、SOM アルゴリズムを利用して高次元データをクラスター化した後、U-Matrix 法、並列座標、時空間キューブ、およびその他の可視化ツールを使用して、時空間データを共同で可視化を実現させる。これにより、従来の時空間関係の欠陥を補い、時間次元での多次元時空間データの地理的および空間的变化を明確に示し、データの集約と分布を説明して、データの潜在的な関係の分析を容易にすることができる (図 4)。

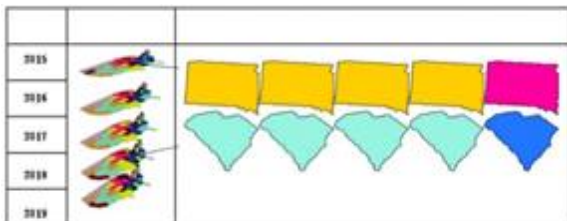


図 4 販売データの時間変化のイメージ

6. 結論と展望

時系列販売データの重要な特性は、膨大な量であるということであり、それをモデル的に表現することは後続のデータマイニングと可視化に対して重要な意義を持っている。実験結果により、平均セグメントのフィッティングに対しても、非平均セグメントのフィッティングに対しても、改良後の計算方法は有効にフィッティング時間を減少でき、フィッティング速度を向上でき、良好な計算速度を実現することができた。

多次元情報の可視化において、SOM の役割は主に二つの方面に現れる：(1) クラスター分析や多変量可視化では、SOM のオリジナルベクトルはサンプルデータと見なすことができ、元のサンプルデータは次元数の低いデータセットに置き換えられているため、SOM のオリジナルベクトルから得られる可視化情報は元のデータと一致していると考えられる。(2) 他の技術では、SOM をデータのモデルと見なして、データをこのモデルと比較したり、モデルに基づいて様々なサンプルデータセットを反転したりすることができる。

SOM では、データに対して自律的な学習を行い、この学習方法では、多次元のベクトルデータに対してのみ次元削減処理を行い、多次元情報は時空間情報を大量に含み、SOM は多次元情報の時空間特性に対する反映は少ないと考えられる。どのようにしてこの方法において時空間情報を持つ多次元データをはっきり反映させることができるのかについて、今後さらに深い研究が必要である。

7. おわりに

本研究では、可視化分析方法の改良、特に可視化分析中に、カーブフィッティング、SOM、補間法など様々な分析方法をどのように改良して、どのようなスパンで、どのように組み合わせる分析を行うことが最も効率的なのかを整理して示すことを目指しているが、今回はそのうちのカーブフィッティングと SOM アルゴリズムに基づいた研究手法や計算過程について報告した。

今後は、様々な利用環境においてカーブフィッティングと SOM アルゴリズムによる可視化モデルを評価し、そしてこれまでの研究手法と組み合わせ、より効率的、正確的な可視化モデルを提案することを目指す。

謝辞

本研究を進めるにあたり、ご議論いただきコメントをいただきました皆様に感謝いたします。

参考文献

- [1] 馮鑫, 能上慎也, “ビジネス向けのデータ可視化分析とその応用” 信学技報, IN2020-18(2020-08), pp.55-58, 2020.
- [2] 馮鑫, 能上慎也, “データ可視化の手法について—時系列補間とニューラルネットワークアルゴリズムを中心に” 信学技報, vol. 120, no. 324, LOIS2020-40, pp. 108-112, 2021 年 1 月
- [3] Bruce H. McCormick, Thomas A. DeFanti and Maxine D. Brown (eds.), Visualization in Scientific Computing. ACM Press (1987).
- [4] Chen, W., L. Ren, M. Zwicker, and H. Pfister, Hardware-accelerated adaptive EWA volume splatting, IEEE Visualization '04, pp.67-74 (2004).
- [5] Kohonen, Teuvo. Intro to SOM. SOM Toolbox. 2005 [2006-06-18].
- [6] Haykin, Simon. 9. Self-organizing maps. Neural networks - A comprehensive foundation 2nd. Prentice-Hall. 1999. ISBN 0-13-908385-5.