

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 10 日現在

機関番号：32601

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24710173

研究課題名(和文)学習ベクトル量子化による多次元データの可視化管理図

研究課題名(英文)Visualization of Multi-dimensional Chart Based on Vector Quantization Model

研究代表者

齊藤 史哲(SAITOH, FUMIAKI)

青山学院大学・理工学部・助教

研究者番号：30625132

交付決定額(研究期間全体)：(直接経費) 1,300,000円

研究成果の概要(和文)：複雑化した製品やサービスに関するデータの理解支援を目的としてベクトル量子化モデルを用いた多次元管理図の構築を行った。管理者は自己組織化マップに代表されるベクトル量子化モデルによる低次元空間において視覚的な異常検知を可能にした。さらに、外れ値の検出において利用される指標の一つであるPeculiarity Factorの計算量を、ベクトル量子化モデルを用いて近似的に計算することによって、抑制する方法を提案した。これらの手法は複雑かつ高度化した製品サービスの管理技術として期待できる。

研究成果の概要(英文)：The works presented here are of constructing robust multi-dimensional control chart using vector quantization model for the purpose of understanding support of the complex data. We have realized the visualization of anomaly data by the vector quantization model represented by self-organizing map. In addition, we focused on the peculiarity factor as an indicator of anomaly detection, and we have proposed a method of suppressing the amount of its calculation. These techniques can be expected as a management technology of complexed and sophisticated products and services.

研究分野：知的情報処理

キーワード：ベクトル量子化 クラスタリング データ可視化 異常検知

1. 研究開始当初の背景

近年における製品・サービスの複雑化や高度化に伴い、各製品の品質に影響を及ぼす要因は増加している。このような要因の管理において、良・不良や正常・異常の識別は重要な課題である。一般に、この判定においては管理図や散布図を用いたデータの可視化に基づいて、異常検知やデータの特徴把握が行われる。このようなデータの可視化は管理者の知識に基づく問題解決を行う上で重要なプロセスの一つであるといえる。

単変数による管理図や各変数対の散布図を用いたデータの可視化のみでは変数間の関連性の把握は難しく、高次元データへの対処においては限界がある。また、一般的に用いられる多次元管理図においても設定や仮定などの制約が多く、必ずしもこの条件を満足するデータばかりではないことが指摘されている。

特に、これらの手法は多重共線性や非線形性を持つデータおよび大規模疎なデータに対して脆弱であると考えられており、これらの課題に対処する必要があると考えた。

2. 研究の目的

本研究では、上記の課題に対処するべく、データに対して制約を設けず頑健性の高いベクトル量子化モデルをこれらの分析へ拡張することにより、新たな管理手法の構築を目指している。すなわち、本研究の目的は多次元データを統合的な視覚化に基づく、管理者による良・不良の識別・把握を可能にする方法の構築である。

これにより、従来の管理図や散布図の利用における異常検知をサポートできる。ここで提案された方法により従来の方法で対処しきれない状況を補える可能性があり、複雑かつ高度化した製品・サービスに対する柔軟な対応が期待できる。

3. 研究の方法

データマイニング・知的情報処理の分野において広く利用されているベクトル量子化の技術を本研究における中心的な手法として用いる。ベクトル量子化モデルは多次元データを代表点ベクトルの集合により近似する手法である。代表的な手法には、データの可視化に主眼を置いた自己組織化マップ(SOM)、データの識別・クラスタリングに主眼を置いた学習ベクトル量子化(LVQ)・K-means クラスタリングなどがある。

本研究では、これらの手法をデータの特徴把握において適用することによって、異常検知や異常データの可視化を実現する。さらに、このようなタスクにおける問題点を解消する際に、ベクトル量子化が有効に機能することを示す。

本研究は以下に示すような大枠(1)~(3)によって構成されている。(1)可視化マップに対するデータの評価値の付与、(2) Peculiarity Factor とベクトル量子化モデルの併用による異常データの検知、(3)自己組織化マップによる異常データの可視化である。それぞれに対する説明は以下に示す通りである。

4. 研究成果

(1) データ可視化のための低次元マッピング

低次元マッピングによるデータ可視化を行う際に、そのデータがどのような性質を持つかを利用者に理解しやすくする必要がある。各データが持つ評価値などをマップと対応付ければよいわけだが、自己組織化マップには学習に関与しないデッドノードが発生するという問題やデータ自体の評価が不明であるといった問題からこの対応付けは難しい。

そこで、これら課題に対処するために、評価値が未定なノードに対する評価値を付与

する方法を提案した。ここでは、自己組織化マップ上の各ノードの近傍関係とその類似度に着目し、既に与えられている評価値を用いて未知の評価値を補間する方法を提案した。

(2) Peculiarity Factor とベクトル量子化モデルの併用による異常データの検知

ここでは、このようなタスク(異常検知・特異性の分析)において用いられている指標の一つとして、一定の成果を収めている Peculiarity Factor に着目した。この指標は各データに対して与えられる指標で、多次元データ空間中において相対的な外れ度を評価することができる。

この指標は外れ値検出において有効である一方で、各データに対して相対的な外れ度が計算されるため、データの増加に伴い計算量が急激に増加するという問題がある。そこで、ベクトル量子化モデルを用いて近似的に Peculiarity Factor を計算する方法を提案し、データの増加への対処をしつつ安定的な外れ値検出を実現した。

データ数と提案法による計算量の削減量の関連性は図2に示す通りである。

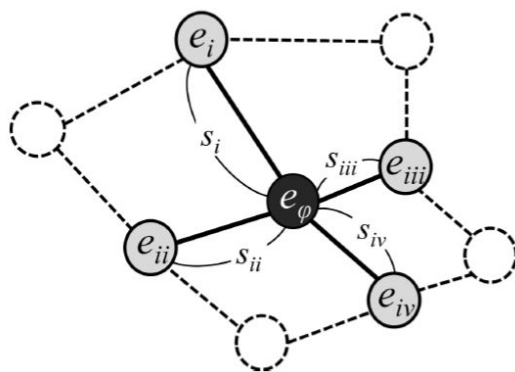


図1. 評価値が未知であるノードの補間法の概略図

(3) の自己組織化マップによる異常データの可視化

ここでは、低次元にマッピングされたデータが、その中でどの程度外れているか(異常

であるか)をマップ上で可視化する方法を提案した。ここでは、自己組織化マップの各ノードが超球状の正規分布と考えると、その最尤法に基づいてデータの分布を推定できる。そこから、データがどの程度外れているかが算出でき、マップ上にその程度が可視化できる。

以上を要するに、(1)~(3)の研究により、本研究で目指していたベクトル量子化モデルによる多次元管理図の構築は概ね達成できたと考えている。

また、高次元データの可視化を目的として行った(1)の研究に対する副次的な成果として、Web カスタマーレビューの可視化に拡張することができた。これにより、Web レビューの文書データ(顧客の声)のような高次元かつ疎なデータを低次元マッピングし、それと関連する評価情報を付与することでデータの可視化を実現した。異常データの視覚的把握のみならず、製品サービスに対する評価とテキストデータとの関連性への拡張可能性を示す事ができたと考えている。

なお、将来的な展望として、(3)の研究を今後も継続していく予定である。研究期間内で煮詰まらなかった議論を十分に行うことで、更なる発展を見込んでいる。

全研究期間を通じて得られた研究成果は以下に示す通り、査読付き学术论文2編、学会発表7件、受賞2件である。

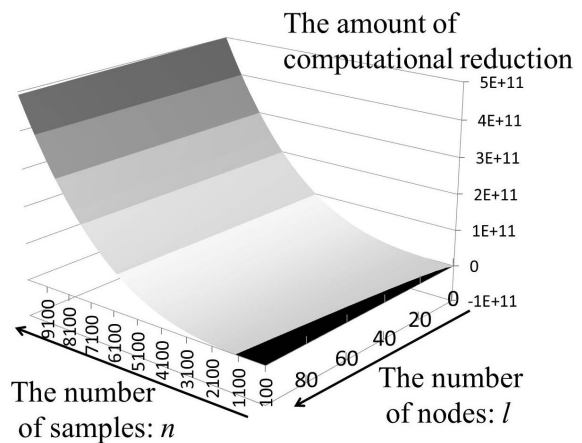


図2. データ数・ノード数と削減された計算量の関係

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 2 件)

(1) 齊藤 史哲, 石津 昌平, “ベクトル量子化モデルによる Peculiarity Factor の近似計算,” 電気学会論文誌 C, Vol. 135, No. 3, pp. 304-311, (2014) (査読有)

(2) 齊藤 史哲, “Web カスタマーレビュー文の理解支援を目的とした自己組織化マップによる評価分布の可視化法,” 日本経営工学会論文誌, Vol.65, No.3, pp.180-190 ,(2014) (査読有)

[学会発表](計 7 件)

(1) 齊藤 史哲 “自己組織化マップの確率密度関数に基づいた多次元管理図,” 日本経営工学会平成 26 年度秋季大会予稿集, pp.200-201, 広島大学 (広島県・東広島市), 2014 年 11 月 8 日

(2) Fumiaki SAITOH, “Visualization of Online Customer Reviews and Evaluations Based on Self-organizing Map,” The 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC2014), pp. 182-187, San Diego (CA, USA), 5.10.2014

(3) 齊藤 史哲 “自己組織化マップを用いたカスタマーレビューにおける評価分布図の可視化,” 第 24 回インテリジェントシステムシンポジウム(FAN2014), 北見工業大学 (北海道・北見市), 2014 年 9 月 18 日

(4) Fumiaki SAITOH and Syohei ISHIZU, “Visualization of Anomaly Data Using Peculiarity Detection on Learning Vector Quantization,” the 15th International Conference on Human-Computer Interaction International (HCI2013) Human Interface and the Management of Information Part.2, pp.181-188, Las Vegas (NV, USA), 22.7.2013

(5) 齊藤 史哲, 石津昌平 “学習ベクトル量子

化による Peculiarity Factor の計算量の抑制,” 第 23 回インテリジェントシステムシンポジウム(FAN2013)講演論文集, pp.25-30, 九州大学 (福岡県・福岡市), 2013 年 9 月 25 日

(6) Fumiaki SAITOH and Syohei ISHIZU, “Visualization of Nonlinear Multidimensional Quality Data Using Self-organizing Map and Bootstrap,” 10th Asian Network for Quality Congress (ANQ2012), pp. 699-704, Hong Kong (China), 31.7.2012

(7) 齊藤 史哲, 石津昌平 “自己組織化マップとリサンプリングを用いた品質データの可視化法,” 日本品質管理学会第 42 回年次大会研究発表要旨集, pp.137-140, コマツウェイ研修センタ (石川県・小松市), 2012 年 10 月 28 日

[その他]

受賞:

(1) 平成 27 年 日本経営工学会論文奨励賞: 齊藤史哲, “Web カスタマーレビュー文の理解支援を目的とした自己組織化マップによる評価分布の可視化法,” 日本経営工学会論文誌, Vol.65, No.3, pp.180-190,(2014)

(2) 平成 26 年 電気学会 電子・情報・システム部門 技術委員会奨励賞: 齊藤史哲 “学習ベクトル量子化による Peculiarity Factor の計算量抑制”

6. 研究組織

(1)研究代表者

齊藤 史哲 (SAITOH, FUMIAKI)

青山学院大学・理工学部・助教

研究者番号: 30625132