

データ特性の対話的可視化技術

藤野友也*
高山茂伸*
菅野幹人*

An Interactive Visualization of Features on Versatile Data

Tomoya Fujino, Shigenobu Takayama, Mikihiro Kanno

要 旨

近年、企業ではPOS(Point Of Sale)データや設備の稼働ログデータなど、膨大かつ多項目からなるデータの収集・蓄積環境が整備されてきており、データの発生から活用までのデータライフサイクルを効率的に管理するデータマネジメントの重要性が増している。一方で、売上げ分析や稼働状態監視などへのデータの活用には、大量のデータから目的に合致する部分を発見する必要がある、データの抽出条件を繰り返し確認する煩雑な試行錯誤が課題となる。

一般に分析手法は、次の2種類のいずれかに二極化される。一つは、分析処理がブラックボックスであり、結果は得られるが、分析者の意図の反映や結果の導出理由の推察が困難である手法であり、もう一つは分析者の意図を反映し段階的に分析を進めることができ、全体像を把握しやすいが、専門知識やスキルを必要とする手法である。

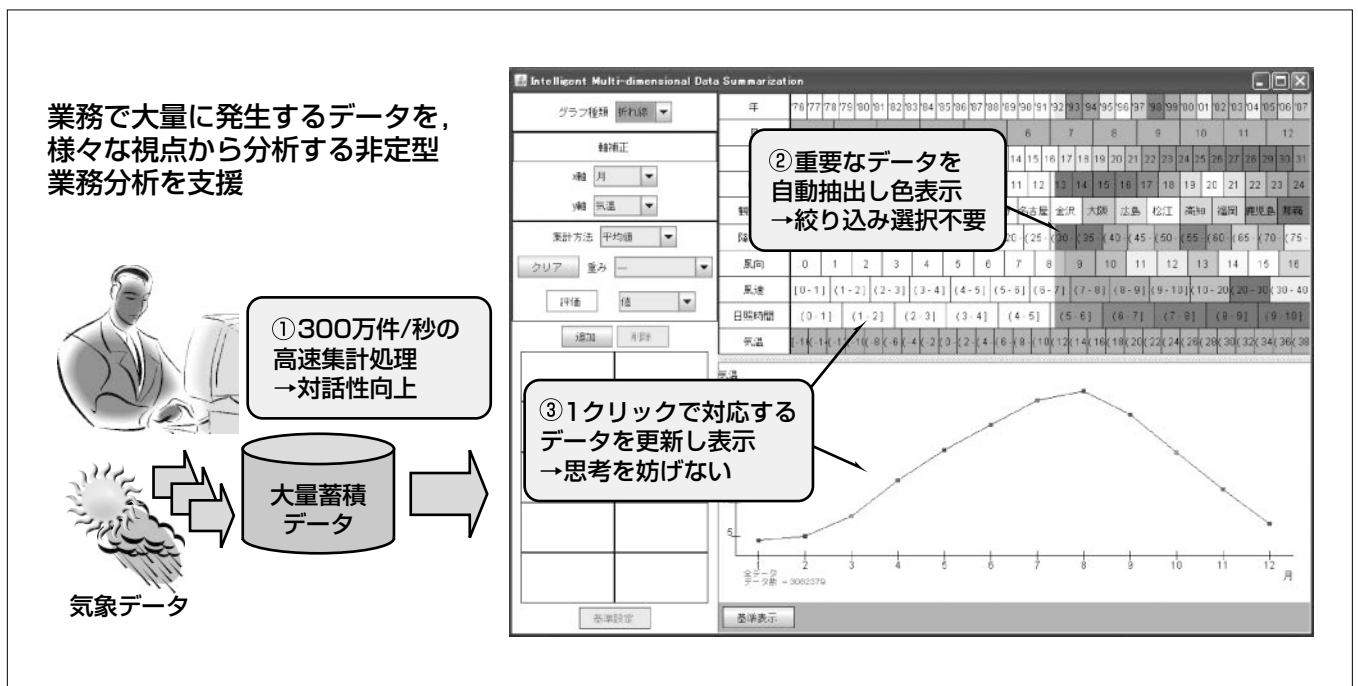
三菱電機では、双方の手法の長所を取り入れ、専門知識を必要とせずに対話的な分析操作を効率化し、少ない作業量で効率的に目的の知見を得るためのデータマイニングについて研究開発を行っている。その成果の一つに、データ

特性可視化技術(Intelligent Multi-dimensional Data Summarization:IMDS)がある。IMDSは、分析の各過程で発生する複数の選択肢に対して、どの選択肢を選択すべきかを定量的に評価し、選択肢を一覧配置した画面上で視覚的に表示するソフトウェアである。IMDSは、優先すべき選択肢を分析者に容易に把握させることが可能であり、次の特長を持つ。

- (1) 高速集計技術によって、視点の変更に伴う表示を瞬時に更新し、対話的な分析操作を実現(300万件/秒)
- (2) 着目すべき条件を強調表示して、分析を誘導
- (3) 選択可能な条件が一画面で表示され、探索が容易

IMDSが提供する視覚的な選択の誘導によって、専門知識不要でかつ分析者の意図を反映可能な分析を実現する。

今後は、データの蓄積が想定される各事業・各企業の販売部門や品質管理部門などへの適用や、現場で容易に即時分析を行う環境整備への応用を想定し、この技術を展開していく。



データ特性可視化技術IMDS

このソフトウェアの特長は、①大量のデータに対する対話的分析操作を可能とする省メモリ高速集計技術による即時性、②分析において通常の特長と異なるデータの条件をユーザーに提示する分析誘導技術による視認性、③データの特性をわかりやすく表現する、簡単に直観的なユーザーインターフェースによる操作性である。

1. ま え が き

現在、各企業はPOSデータや機器動作ログデータなど、億単位にものぼる膨大な件数のデータを保有しているが、企業がデータ資産をいかに活用していくかは大きな課題である。データ資産の活用にはデータ分析が有効であるが、一般の分析手法は次の2つのパターンに二極化している。

- (1) 分析過程がブラックボックス化され、結果は容易に得られるが、その分析結果が導出された理由の推察が困難な分析手法(例：ニューラルネットワーク分析)
- (2) 自らの意図を反映させながら段階的に分析を行い、各段階での分析内容からも知見が得られる反面、高い専門知識やスキルを求める分析手法(例：統計解析)

多くの種類の項目を持つデータを様々な視点から非定型に分析する場合、(1)の手法では分析の視点ごとに、どのような方式でデータを入力し結果をどのように解釈すればよいかを分析者が考案する必要がある、分析手法に関する知識が必要となる。したがって、多項目データの非定型な分析では、分析の過程で意図を反映させることを前提とし、分析手法に関する知識を要しない(2)の分析手法を用いることが多い。この場合、大量のデータの中から注目すべき部分を発見するために、繰り返しデータの抽出条件を変更し確認するという煩雑な試行錯誤が必要となる。分析作業の負荷を軽減するため、例えばOLAP(OnLine Analytical Processing)ツールを使用することで、試行ごとの手間を軽減することができる。

当社では、先に述べた(1)、(2)の手法の長所を取り入れ、操作が容易で、かつ詳細な知見が得られるデータマイニング手法として、分析の各過程で発生する複数の選択肢に対して、いずれの選択肢を選択すべきかを視覚的に表示することで、優先すべき選択肢を分析者に容易に把握させるデータ特性可視化技術(IMDS)⁽¹⁾⁽²⁾を開発している。

IMDSが提供する視覚的な選択の誘導によって、専門知識不要でかつ分析者の意図を反映可能な分析が実現できる。IMDSは、分析のしやすさを追求するとともに、“特異な特性の検知”を目的として開発している。IMDSに独自に導入されている“分析誘導機能”を利用することによって、多数の分析における選択肢の中で、分析者がどの選択肢を選択すべきかを視覚的に強調して誘導するため、無駄な試行錯誤を回避し、効率的に分析を進めることができる。

なおこの技術は、目的が明確でない分析者を支援する技術ではなく、目的が明確な分析者に対し、その分析作業の効率化を支援することを主眼に置く技術である。

この技術は、主に次の点を目的としている。

- (1) 通常時には同じパターンのデータが観測され、稀(まれ)に特異な挙動を示す場合に、特異事例が発生する条

件を特定する目的(流通・製造の品質管理、建物・施設のエネルギー消費量の分析など)

- (2) 品質や利益などの指標に対して、カテゴリーを絞り込んだ場合の傾向の違いを分析(層別分析)する目的(商品ごとの売上げ分析、年齢層ごとの健康管理など)

2. IMDSの特長

2.1 概 要

IMDSは、操作の簡略化に主眼を置いたレイアウトによって、分析手法に関する知識を必要とせずに目的に沿った分析を可能とするソフトウェアである。平均的・一般的な傾向を逸脱する特性を持つ部分データは、分析する価値が高いという方針で、選択肢の中で最も知見が得られると判断される選択肢を強調表示し、目的に合致した分析が得られるように誘導して、分析者をサポートする。IMDSは次のような特長を備えている。

- (1) 膨大なデータを遅延なく集計する即時性
- (2) 分析の要所を評価し視覚的に強調する視認性
- (3) 試行錯誤にかかる作業負荷を軽減する操作性

2.2 技 術 内 容

IMDSは、分析業務における対話処理に必要な視認性や操作性を確保するユーザーインタフェース技術⁽¹⁾を特長としたソフトウェアである。特長は次の3点である。

(1) 即時性

データ抽出条件の変更による再集計を効率的に行う省メモリ高速集計技術によって、集計・評価処理にかかる負荷を軽減し、操作ごとの即応性を実現した。

省メモリ高速集計技術では、起動時に中間集計値を算出し、起動後に指定される集計的に条件に応じて、差分的に新たな集計値を算出することで、既存ツールとの実測比数十倍の高速集計を行い、対話的高速性を確保する。

(2) 視認性

基準となる特性と、部分データの特性との違いを定量的に評価する分析誘導技術によって、特性の違いが大きいほど、絞り込み条件に対応するボタンの色を濃く表示し、各条件の重要度を分析者に視覚的に提示する。

(3) 操作性

抽出条件に対応するボタンや制御アイコンを一画面に配置することで、最小限の操作でデータ分析を可能とするユーザーインタフェースを提供する。

図1に示すインタフェースの右上部に、タイルのようにボタンが並んでいるGUI(Graphical User Interface)部分が、集計対象のデータを絞り込む条件を選択する部分である。例えば“月”の行の“8”のボタンを押すと、データは8月に観測されたデータのみ絞り込みがされ、瞬時に右下のグラフに反映される。

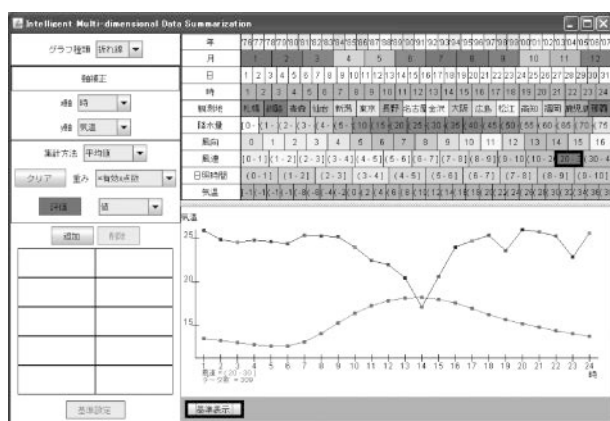


図 1. IMDS インタフェース例

表 1. 分析対象アメダスデータ諸元

データ内容	アメダス観測気象データ
観測期間	1976年1月1日～2007年12月31日
観測地点	16地点(札幌, 釧路, 青森, 仙台, 新潟, 東京, 長野, 名古屋, 金沢, 大阪, 広島, 松江, 高知, 福岡, 鹿児島, 那覇)
測定間隔	1時間ごと
測定項目	5項目(降水量, 平均風速, 風向, (平均)気温, 日照時間)
データ数	3,062,379レコード
ファイル形式	CSV形式(98,143,773バイト)

年	月	日	時	観測地	降水量	風向	風速	日照時間	気温
1976	12	1	2	東京	0	11	0.2	0	6.0
1976	12	1	3	東京	0	12	0.3	0	5.2
1976	3	17	5	名古屋	2	1	0.2	0	10.0
1976	3	17	6	名古屋	3	16	0.2	0	10.0
1977	5	15	8	福岡	0	1	0.6	0	13.2
1977	5	15	9	福岡	0	1	0.7	0	13.5
1977	6	8	12	高知	0	7	0.4	1.0	25.1

図 2. CSVデータ例

3. 分析例

3.1 気象データによる分析例

気象データは、観測結果そのものの広報や、過去に蓄積されたデータの解析による予報など、最も古くから取り扱われ、一般生活で触れる機会の多いデータの一つである。近年でも、気象データを小売店のPOSシステムと連携させ客足予測に活用したり、異常気象に起因する損害を補填(ほてん)する天候デリバティブが金融商品として市場を拡大したりするなど、気象データの重要性は増す傾向にある。

ここでは、気象データを用いて分析の流れについて述べる。30年分16観測地点からなる、時別のアメダス気象データ⁽³⁾(表1)を用いたIMDSによる異常気象分析の例を示す。

図2に示すような汎用(はんよう)的な形式のCSV(Comma Separated Values)ファイルからデータを読み込み、300万レコードのデータに対し、1GHz程度のCPU(Central Processing Unit)のノートパソコン上で、各集計操作が1秒以内に完了する高速性を実現し、対話的な処理を可能とする。

ここでは分析の一例として、“気温に関する異常気象”

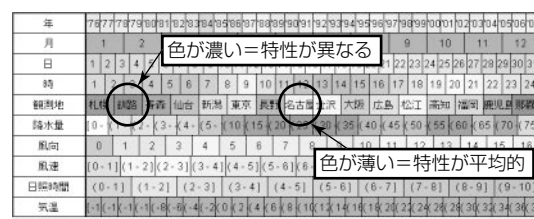


図 3. 月別気温変化に関する異常候補評価



図 4. 年ごとの特異性可視化結果

を抽出する。30年間の全国のデータから求めた、月ごとの平均気温の推移を通常の気温推移ととらえ、そこから逸脱する気温変化が発生した年の絞り込みを行う例を示す。

月別の気温推移を観測中に、“評価”機能を実行すると、条件を絞り込むボタンの色が、それぞれ図3のように、様々な濃淡で彩色される。この濃淡が、各条件への絞り込みによって、どの程度特性が変化するかを表現している。

図3では、“観測地”の“釧路”や“那覇”の色が濃いことが確認できる。濃い色に対応するボタンに対応する条件に絞り込むことで、平均と異なる特性を確認できる。

一方、釧路や那覇が全国平均と異なる気温推移をすることは既知の事実であるため、当初の目的どおり、異常気象であった年の分析に注目する。年による気温推移の特性の差は、観測地による差より小さいため、年を示す項目の色は薄い、図4のとおり注目したい“年”項目で相対的に色が濃い(特性が異なる)年が1984年であることがわかる。

図5に、年を1984年に限定した集計結果を示す。図を確認すると、1～3月の気温が30年間の平均と比べて3度近く低いことが確認でき、この年が例年に比べて特に冬の気温が低い異常気象であったことがわかる。実際、当時の報道でも平均最高気温が昭和になって最低を記録するような“寒冬”であったことが報じられる異常気象であった⁽⁴⁾。

本来ならば、単に年に注目して異常気象を探索する場合でも、30年間の各年を一つ一つ確認していかなければ、このような事例を発見できない。事前処理で優先度を可視化することで、通常と異なる箇所を優先的に分析することが

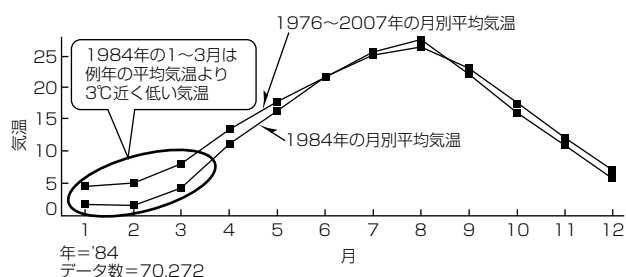


図5. 平均的月別気温推移と1984年の月別気温推移

可能になる。一定の作業量の範囲内での分析業務で、従来の広く浅い分析から狭く深い分析へと移行することで、より質の高い分析作業を実現することができる。

ここで述べた分析例では、最も特性が異なると判定された“観測地”の“釧路”や“那覇”よりも、“年”の中で相対的に特性の異なる点に注目した。分析の目的を定め目的に応じて重視する項目を適宜選択することが、分析の手法の選択と同等に重要である。

3.2 その他応用例

IMDSは汎用的なデータ形式からの可視化が可能であるため、次のような種々の応用が考えられる。

(1) 省エネルギー施策の検討

建物の電力消費情報から、不適切なエネルギー消費をしている場所・部署を洗い出し、効率的に環境整備や指導などを行い、一層の省エネルギーを実現することを目的とした分析に活用できる。例えば、フロアごと、消費機器ごと、日時ごとに区別し収集されたエネルギー消費データを基に、IMDSを使用して全体の平均的消費から逸脱する条件を洗い出す。その結果、対象の事業所の消費エネルギーから、例えば冬季に月曜日朝の空調消費電力が高い傾向が観測された場合に、サーバ類の時差起動による消費電力ピークの解消施策を実施するなどといった対策の立案が可能となる。

(2) 品質管理

製造業やインフラ系公共機関における生産物管理や設備管理では、個々の事象の特性は現場で経験的に把握しているものの、多数の要因が影響することによって、単純に予想された効果が現れないことがあるという課題がある。こ

の課題に対応するために、定期的に生産物や設備の状態を測定し、その結果を蓄積して、過去の実績に基づく管理を行うことで、生産物や設備の評価基準を確立するアプローチをとることが望まれる。IMDSを用いることで、分析時の条件と同じ条件の過去実績を参照することが容易であるため、適切な処置を短時間で実施することが可能である。これによって、生産の歩留りや改善処置が必要最小限に抑えられ、生産性の向上が見込まれる。

4. む す び

データマネジメントでは、データの発生から取得、収集、管理、統合を経て、分析・可視化を通じて活用に至るデータライフサイクルが存在し、分析・可視化技術は活用現場に最も近いフェーズに位置する。

IMDSは、そのフェーズの中でもデータ全体を見通して特異な特性が発生している時期・箇所を特定することを可能とし、そこを足掛かりに詳細な分析を行うことで、目的の知識までの試行錯誤を最小に抑えて大きな成果・発見をもたらすことに寄与するものである。

IMDSは、現状では基本的に汎用ソフトウェアとして開発しているが、分野ごとに固有の課題に対して、より効果的な分析支援が可能となるよう、引き続き技術開発を進めていく。

参 考 文 献

- (1) Lesh, N. B., et al.: Interactive Data Summarization, An Example Application, Advanced Visual Interfaces (AVI), 183~187 (2004)
- (2) 藤野友也, ほか: データ特性可視化による分析システム, 電子情報通信学会総合大会講演論文集, 2005年情報・システム(1), 27 (2005)
- (3) 財気象業務支援センター: アメダス年報
<http://www.jmbc.or.jp/>
- (4) 異例ずくめ“寒冬”裏付け, 読売新聞夕刊, 14 (1984-2-29)