

多次元データ分析システムによる保守業務分析支援の検討

藤野 友也[†] 今村 誠[†] 菅野 幹人[†]

[†]三菱電機株式会社 情報技術総合研究所 〒247-8501 神奈川県鎌倉市大船 5-1-1

E-mail: [†] {Fujino.Tomoya@cj, Imamura.Makoto@bx, Kanno.Mikihito@bc}.MitsubishiElectric.co.jp

あらまし 今日、保全業務に関する多次元データの分析は、設備・機器を効率的に維持するために重要となってきた。例えば OLAP(On-Line Analytical Processing) のような典型的な多次元データ分析システムでは、膨大な分析視点を全て確認できないこと、また試行錯誤に伴い分析時間が長くなることなどの課題がある。本論文では、分析推薦機能と対話的集計機能を備えた多次元データ分析システム dCOMPASS による保守業務分析の効率化の検討を行う。故障率の経年変化などの典型的な分析業務において、分析の過程における GUI 操作量(クリック数)、分析担当者の判断回数、および所要時間の改善効果について報告する。

キーワード 多次元データ, 保守業務, 分析支援, OLAP, 層別分析

1. はじめに

近年の公共施設、製造プラントおよびビルファシリティにおいては、設備・機器の電子化による高機能化・多機能化が進んでいる。それに伴い、設備・機器が故障などで停止した時の影響範囲は年々拡大している。これらの設備・機器の平均故障間隔 (MTBF, Mean Time Before Failure), 平均修理時間 (MTTR, Mean Time To Repair) および稼働率の改善のため、通年的な点検・保守・故障検知および修理作業などの保守業務の重要性が増している。保守業務を提供する企業もしくは部門にとって、保守業務活動の効率化と保守品質の向上は、経営上の重要な課題である。

設備・機器における自動診断機能の標準化や、通信やストレージの高速化・大容量化・低コスト化が進み、観測データを選別せずに蓄積することが容易になっているが、蓄積されたデータを活用する方式については発展途上である。特に、各地に点在する機器や設備に保守作業者が向向き修理する形態の保守業務では、蓄積されたデータを確認する個人の経験が現場依存となり、集約されにくい。そのため、それらの情報を集約し効率化につなげるための取り組みとして、保守カルテ[1]や不具合事例知識[2]などを、企業知[3]として集約・共有するシステムが提案されている。

設備・機器の高機能化により、観測可能なデータ項目は多様化しているが、どのデータ項目が分析に有意義であるかについては、分析担当者の経験からの判断に委ねられている。項目が膨大であるため、試行錯誤が必要であるが、全ての試行錯誤に対して人が判断を下すことは量的に難しいという問題がある。

本論文では、保守業務データの分析において、分析の際に注目すべき項目の候補が多く、全てを人が判断することが量的に難しいという課題を解決し、保守業務活動の効率化に繋げるシステムを提案する。

2. 保守業務向けデータ分析

2.1. 保守業務とは

本論文で述べる保守業務の定義を以下に示す。

- ・ 保守作業は、目視点検、校正、部品交換、故障修理の4種類に分類される。目視点検では、設備・装置に異常がないことを外観から確認する。校正では、清掃、校正、高負荷動作確認などを行う。部品交換では、経年消耗品を交換する。故障修理では、故障した部品に対する清掃、校正、部品交換を行う。
- ・ 保守作業者が、顧客との契約に従い、定期・不定期に設備・機器が設置された現地を訪問して目視点検・保守・部品交換を行う。
- ・ 機器に不定期に発生する故障に対し、保守作業者が故障修理を行う。
- ・ 機器は部品で構成され、故障は部品で発生する。

各々の保守作業者が機器に対して行う保守・修理作業の品質は、機種毎の習熟度の差や、個別の機器・現場に対する慣れなど、様々な要因により左右される。保守業務を継続する中で、現場の保守作業者はさまざまな仮説を経験的に認識するが、仮説が成立することを、実際のデータに基づいて裏付けることが重要となる。

2.2. 保守業務に関するデータ

本節では、保守業務にて管理されるデータの内容について述べる。保守業務に関連するデータとして、例えば以下に示す9種類のデータが管理され蓄積される。

- ・ 顧客データは、保守業務を保守業務部門に委託している顧客を管理する。(数百～数万件)
- ・ 保守作業データは、保守業務部署に属して各種の保守作業を担当する保守作業者に関する情報を格納する。(数十～数千人)
- ・ 機種データは、設備・装置をカテゴリ分けする種類

の情報を格納する。(数十～数百種類)

- ・ 機器データは、個別の設備・装置に関する情報を格納する。(数千～数万件)
- ・ 部品データは、機器を構成する部品に関する情報を格納する。(数百～数千種類)
- ・ 故障データは、機器・部品に発生した故障に関する情報を格納する。(数百～数千件)
- ・ 故障内容データは、故障の詳細内容に関する情報を格納する。(数十種類)
- ・ 保守作業データは、保守作業が実施した保守作業の情報を格納する。(数万～数十万件)
- ・ 作業内容データは、保守作業の詳細内容に関する情報を格納する。(数十種類)

2.3. 保守業務向けデータ分析ニーズ

保守業務向け分析ニーズの例として、以下の3種類が挙げられる。

[分析ニーズ 1] 経年劣化

稼働年数の増加に対して故障件数が増加する傾向を強く示す条件を特定する。

[分析ニーズ 2] 保守後早期故障

保守実施後の経過時間が短い段階で、故障が発生する件数が多い傾向を強く示す条件を特定する。

[分析ニーズ 3] 保守回数別故障

保守の頻度に対応し、保守頻度の低い場合に特に故障件数の多い傾向を強く示す条件を特定する。

上記の分析ニーズはいずれも、設備・機器の故障が増大することが想定される条件を明らかにし、保守を適切なタイミング・頻度で実施するために把握しておくべき内容である。それぞれ確認したい傾向は明確であるが、条件として考慮すべきデータ項目や、それぞれの項目が取り得る値の種類が多く、確認すべき組み合わせが膨大となる。

分析ニーズ 1～3 に対応する分析タスクは、以下の通りである。

[分析タスク 1] 経年劣化分析

稼働年数に対して故障発生数が増加する傾向を発見することを目的として分析する。稼働年数と故障件数を比較する。稼働年数の値が大きいほど故障件数が多い右肩上がり(上昇傾向)を示す条件を発見する。

[分析タスク 2] 保守後早期故障分析

保守作業を実施した後、短時間での故障発生が多い傾向を発見することを目的として分析する。保守後経過時間と故障件数を比較する。保守後経過時間の値が小さい範囲で故障件数が多い右肩下がり(下降傾向)を示す条件を発見する。

[分析タスク 3] 保守回数別故障分析

保守回数が少ない設備で、故障発生が多い傾向を発見することを目的として分析する。累計保守回数と故障件数を比較する。累積保守回数の値が小さい範囲で故障件数が多い右肩下がり(下降傾向)を示す条件を発見する。

2.4. 保守業務向けデータ分析の応用課題

保守業務向けデータ分析は、分析する際に候補となる項目が多く、全てに対して人が判断を下すことが量的に難しいという応用面の課題がある。多数の部品からなる機器では、個々の部品から取得可能な動態情報の種類が多く、累積して膨大な件数となる。百数十～数百、多い場合には数千単位で存在するデータ項目を全て人手で確認することはコストを要する。コストを抑制するために、分析の視点を一部のみに限定する場合に、適切な分析の視点を適切に選択することは、十分な経験が必要である。

[応用課題 1] 分析時間の増大

分析における複数の選択肢をすべて人が確認することは分析時間を多大に要する。

[応用課題 2] 有用結果の見逃しの発生

一部の選択肢に限定して分析する場合、対象外とした選択肢に有用な結果があっても認知できない。

3. 従来方式の技術課題

3.1. 従来方式

多次元データベース[4]などに格納された、多数のデータ項目からなるデータを分析対象とする分析を、多次元データ分析と呼ぶ。多次元データ分析向けに広く知られている手法に、OLAP (On-Line Analytical Processing) [5]がある。OLAPは、複数の項目からなるデータを、視点を自由に切り替えながら分析する手法の概念として、IBM サノゼ研究所の Edger F. Codd 博士が 1993 年に提唱した[6]。OLAPによる分析では、主に対話的な操作で分析を進める。基本的な操作として、データの分類をより細かい分類に変更するドリルダウン、データを絞り込むスライシング、データの視点を変更するダイシングの3つの操作がある。

OLAPは、大きく分類して MOLAP (Multidimensional OLAP) と ROLAP (Relational OLAP) の2種類の形式のいずれかで実装される[7]。MOLAP と ROLAP の特徴を Table 1 にまとめて示す。MOLAPは、多項目からなる値が変化しないデータ(以降、大福帳型データと呼ぶ)を、事前にキューブと呼ばれる多次元テーブル[8]の形式に集計して管理する。MOLAPは、キューブから集計値を算出することで、元データからの処理に比べて高速に集計ができる。集計処理に使用されるクエ

りは、ベンダの独自言語で実装される。キューブは、一般に元データよりもサイズが大きくなるため、分析可能なデータ量は ROLAP より小さい。大福帳型データを格納するデータベースを大福帳型データベースとも呼ぶ。ROLAP は、リレーショナルデータベース (RDB) と連携してテーブル形式でデータを保持し、その都度集計値を算出する。ROLAP は、RDB が対応する柔軟な集計を可能とするが、集計時間が遅い。集計処理に使用されるクエリには、SQL クエリが用いられる。分析可能なデータ量は、RDB で格納可能なデータ量であり、大容量である。

Table 1. MOLAP と ROLAP

特徴	MOLAP	ROLAP
データ形式	キューブ形式	RDB 形式
クエリ	ベンダ独自	SQL クエリ
集計速度	速い	遅い
データ量	中規模	大規模

OLAP 分析作業のフローチャートを Fig. 1 に示す。入力に対してシステムが集計処理を行う。分析担当者は、集計結果を判断し、十分と判断しなければ分析軸変更を行い、再度システムが集計処理を行う。分析担当者が十分と判断した場合は、分析結果を出力する。入力から出力までの所要時間が分析時間である。

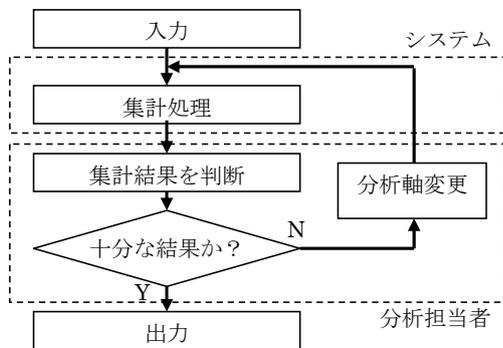


Fig. 1. 分析フローチャート

OLAP による分析を効率化する技術について、複数の取り組みが報告されている。磯部らは、視覚化方式の汎用化により表現の多様性を確保したシステム INFOVISER[9]を提案している。Hasaan らは、MOLAP において一度作成したキューブを分析中に拡張可能なキューブの実装方式を提案している[10]が、50 万件のレコードに対する挿入処理に、データの分割度合いに応じて 20 秒～500 秒を要する[11]など、対話的処理に使用する場合には課題がある。

3.2. 従来方式の技術課題

2.4 節で述べた応用課題に対応する、OLAP による分

析を保守業務に適用する場合の技術課題は、MOLAP、ROLAP のいずれも、分析の視点の選択を分析担当者に委ねており、選択肢を評価する機能がないことである。

応用課題 1,2 に対応する技術課題として、以下の 2 つがある。

[技術課題 1] 分析選択肢の限定

分析担当者の目的に合わせて、分析選択肢を限定する必要がある。

[技術課題 2] 全選択肢の有用性評価の提示

分析の見逃しが発生しないよう、すべての選択肢について、分析の有用性を可視化し分析者に提示する必要がある。

技術課題 1 は、Fig. 1 の分析軸変更のループ回数を減少させることに対応する。技術課題 2 は、分析担当者が見逃しなく Fig. 1 の集計結果を判断するのに必要な操作量を減少させることに対応する。

3.3. 解決のアプローチ

本論文では、3.2 節で述べた技術課題に対して、以下に示すアプローチで分析軸推薦技術にて解決する。

前述の技術課題を解決するために、分析軸推薦技術は、分析担当者にとって選択が可能な膨大な分析の視点に対して、重要度を算出して可視化する。分析軸推薦技術により、分析の視点の変更・選択におけるスライシング、ダイシングなどの OLAP 操作の各段階で、次に見るべき分析の視点に複数種類の指標から優先度をつけて表示することで、分析担当者の目的に応じた選択を支援し、試行錯誤を効率化する。提案する技術およびシステムでは常に対話的に分析を進め、その対話的分析での選択肢に優先度を付与し効率化する方式であるため、人の分析途中での判断基準の変更にも、容易に対応することが可能である。

4. 分析軸推薦技術

4.1. 技術概要

分析型推薦技術は、複数の保守業務データを統合して得られる大福帳型データに対し、分析担当者が選択した分析軸に応じて、次に選択すべき分析軸を推薦する技術である。

多次元データ分析は、分析対象の膨大なデータからデータの一部を切り出し、複数の視点から確認する技術である。通常、有効な知見を得るまでに数多くの試行を伴い手間を要する分析作業が行われる。分析作業の負荷を軽減するためによく使われる OLAP は、試行ごとの手間を軽減するが、試行回数は軽減されない。

3.2 節で述べた OLAP に存在する技術課題を解決するために、分析軸推薦技術によって、保守業務向けデ

ータ分析において重要な視点である分析軸を、分析中に少ない操作数で対話的に更新するために、分析軸の推薦機能を提供する。ここで分析軸とは、評価すべき指標（集計方式）、比較すべきデータ項目、絞込みすべき条件（絞込み条件）の3種類の要素をまとめて表現したものである。

分析軸推薦技術のデータフローを Fig. 2 に示す。

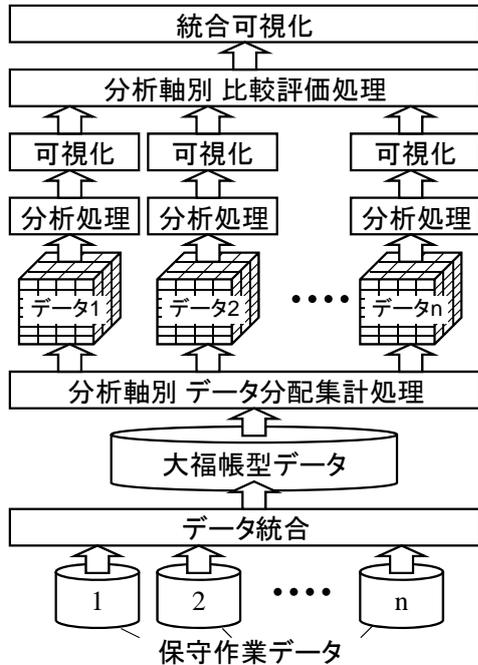


Fig. 2. 分析軸推薦技術のデータフロー

推薦結果を可視化するインターフェースとして、AVI2004にて報告されたインターフェースを使用する[12].

分析軸推薦技術は、保守業務データ分析作業に対し、次の効果をもたらす。分析担当者は、従来は個別に集計し比較確認していた分析軸の層別分析について、分析軸ごとに算出される重要性を基に優先度を付け、その優先度の可視化により効率的に探索することが可能となる。このことから、分析担当者が分析過程において行う分析操作数が削減される。

4.2. 技術要素

分析軸推薦技術は、分析軸重要性評価技術と分析軸評価結果可視化技術の2つからなる。分析軸重要性評価技術によって3.2節の技術課題1を解決し、分析軸評価結果可視化技術によって3.2節の技術課題2を解決する。

4.2.1. 分析軸重要性評価技術

分析軸重要性評価技術は、同一の分析軸の要素ごとにデータや集計値を分類する層別評価を行い、それぞれの分析軸の要素の重要度を定量評価する。分析選択

肢を重要度が上位の選択肢に限定することを可能として、3.2節の技術課題1を解決する。

重要度を定量算出するための関数（評価関数）を複数用意し、分析担当者が目的に応じて適切な評価関数を選択する。評価関数が算出した値を、推薦評価指数と呼ぶ。それぞれの指標に基づく評価関数を分析担当者側で設定可能とするため、評価関数の定義インターフェースを提供する。評価関数は、次の4種類からなる。

- (a) 上昇傾向など、グラフ単体で評価可能な指標
- (b) 複数の条件毎グラフに対して主成分分析を実施し、得られる主成分スコアから、個別のグラフの特異性を評価するなど、複数のグラフを元に評価する指標
- (c) アソシエーション分析など、生データに対しデータマイニングの手法により評価する指標
- (d) その他、部署固有のノウハウに基づく指標

分析軸重要性評価技術が、推薦評価指数を出力する方式について述べる。分析対象データが D_0 、条件セットが p_0 、x軸が c_x 、y軸が c_y であるとき、分析担当者は集計関数 agg 、評価関数 eva を指定して、データ絞込み関数 Sel を用いた Fig. 3 に示すデータフローにて推薦評価指数 $\{e_p\}$ を算出する。

for all $p \in p_0$ の層別分割, $e_p = eva(agg(Sel(D_0, p_0), c_x, c_y))$

ここで、条件セット p_0 とは、データ D_0 が持つ項目に対して、分析対象としたい絞込み条件の集合を管理するものである。絞込み条件は、数値項目ならば数値範囲、離散項目ならば限定すべき値のリストを指定するものである。x軸を指す c_x 、y軸を指す c_y は、それぞれデータ D_0 が持つ項目の1つを指す。集計関数 agg は、第1引数にデータ、第2引数にx軸項目、第3引数にy軸項目を指定する関数である。集計関数 agg は、複数種類の値を持つx軸項目に対し、第1引数のデータのレコードを、x軸項目が同一の値を持つグループ（x軸同値グループと呼ぶ）毎に分類し、同一のx軸同値グループに属するレコードのy軸項目の値の集合に対して、集計値を算出する関数である。例えば、平均値や最大値を算出する関数を集計関数とすることができる。ここで、事前に規定されたx軸同値グループの順序付けに従い、集計関数 agg により算出された集計値を並べた系列を、y軸項目集計値系列と呼ぶ。評価関数 eva は、y軸項目集計値系列を引数とし、規定された基準による推薦評価指数を算出する関数である。例えば、単調増加性を評価する目的で、y軸項目集計値系列で隣接する集計値を比較し、順序が後の集計値の方が大きい隣接点の割合を算出する関数を、評価関数とすることができる。データ絞込み関数 Sel は、第1引数に指定されたデータに対し、第2引数に指定された条件セットの条件に従い、レコードの絞込みを行った結果を出力する関数である。

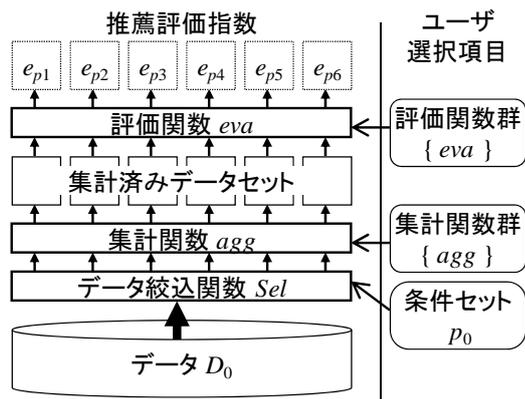


Fig. 3. 分析軸重要性評価技術のデータフロー

4.2.2. 分析軸評価結果可視化技術

分析軸評価結果可視化技術は、分析軸ごとの重要度の強さに応じて、GUI上で分析軸を選択するボタンを強調するよう表示する。また同時に、選択可能な分析軸における集計結果について、重要性の高い順番に配置した画面を分析軸集計結果可視化パネルに表示する。全ての分析選択肢の重要度の一覧による可視化を可能として、3.2節の技術課題2を解決する。

異なる分析軸を推薦した結果を表示した結果を可視化した画面例を、Fig. 4に示す。Fig. 4には、支社、設置環境、機種種の項目名と、それぞれの項目名が持つ値が表示されたボタンが、タイル状に並んでいる。各々のボタンが、それぞれの項目の値を絞込み条件に追加した、新たな分析軸に対応する。分析軸評価結果可視化技術は、分析軸重要性推薦技術が算出した推薦評価指数を元に、ボタンの色の濃淡を用いて、各々の分析軸の重要性を可視化する。Fig. 2のインタフェースを、分析軸評価結果可視化パネルと呼ぶ。Fig. 2の例では、分析軸評価結果可視化技術は、支社項目においては支社07、支社08、設置環境項目においては設置環境05、機種項目においては機種07、機種20、機種21のボタンに対応する分析軸の重要性が高いと判定した結果を受けて、色を濃く可視化している。

表示	データ項目名	種別	絞り込み条件
支社	支社01	支社02	支社03 支社04 支社05 支社06
	支社07	支社08	支社09 支社10 支社11 支社12
	支社13	支社14	支社15 支社16 支社17 支社18
	支社19	支社20	
設置環境	設置環境01	設置環境02	設置環境03 設置環境04
	設置環境05	設置環境06	設置環境07 設置環境08
	設置環境09	設置環境10	
機種	機種01	機種02	機種03 機種04 機種05 機種06
	機種07	機種08	機種09 機種10 機種11 機種12
	機種13	機種14	機種15 機種16 機種17 機種18
	機種19	機種20	機種21 機種22 機種23 機種24
	機種25	機種26	機種27 機種28 機種29 機種30

Fig. 4. 分析軸評価結果可視化パネル

分析軸評価結果可視化技術が、推薦評価指数の高い順に分析軸層別集計結果のグラフを並べて一覧表示した例を Fig. 5に示す。

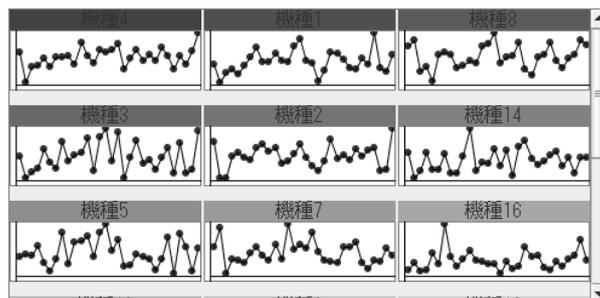


Fig. 5. 分析軸推薦結果に基づくグラフ表示例

Fig. 5の例では、横軸を月、縦軸を保守作業件数として、機種毎の月別の保守作業件数の推移を示している。月別の件数の上昇傾向を推薦評価指数としている。個々のグラフの上部に表示される機種名の背景色が濃いほど、保守出動件数が月を経る毎に増大する傾向が強いことを示す。推薦評価指数が高い順に左上から行方向へ配置している。

5. 多次元データ分析システム dCOMPASS

本章では、分析軸推薦技術を用いて MOLAP を拡張した多次元データ分析システム dCOMPASS を提案する。分析軸重要性評価技術の実装により 2.4 節の応用課題 1 を解決し、4.2.2. 分析軸評価結果可視化技術の実装により 2.4 節の応用課題 2 を解決する。

dCOMPASS システムは、Fig. 6に示す構成によって運用される。

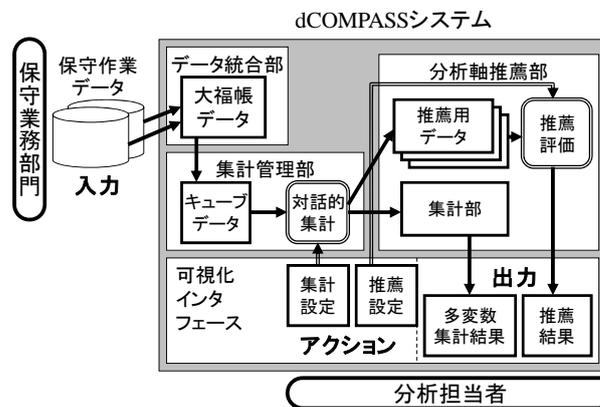


Fig. 6. dCOMPASS システム構成

dCOMPASS システムは大きく、データ統合部、集計管理部、分析軸推薦部および可視化インタフェースにて構成される。保守業務部門にて保有する複数の保守業務データは、大福帳型データとして一つのテーブルとして統合される。集計管理部において、大福帳型データはキューブデータとして管理される。可視化インタフェースにて、分析担当者から集計設定、評価設定が対話的に指定される。キューブデータおよび集計設定を元に、分析軸推薦部が、直接分析担当者が確認す

る集計結果データと、分析軸の推薦のために内部で確保する分析軸推薦データを算出する。分析軸推薦処理では、分析軸推薦データと評価条件を元に、分析軸推薦結果を出力する。分析担当者が出力として確認する情報は、多次元集計結果および分析軸推薦結果である。

6. 評価

本章では、MOLAP と dCOMPASS の分析作業の効率の差異について評価する。本章では、分析軸推薦技術による分析作業の効率化について、最大操作回数の効率化評価および人の判断による平均操作回数の効率化評価の結果を述べる。

6.1. 最大操作回数評価

本節では、大量データに対する必要な分析捜査数の理論値から、MOLAP と比較した dCOMPASS の効率の差異を評価した結果を示す。

本節で評価する保守業務データの内容は、以下の通り構成されるものとする。保守作業 800 人、機種 200 種類、部品 2,000 種類、機器 50,000 個、故障原因 20 種類が存在する。10 年分の保守業務データとして、故障 2,500 件、保守作業 2,300,000 件が蓄積されている。

故障発生数が年毎に増加する傾向の強い機種および故障原因を評価し、故障発生が顕著な機器を特定する分析を評価する。この分析では、増加傾向にある条件の上位 10% を重点的に確認するものとする。年毎に増加傾向にある機種の特定、故障原因の特定、機器の特定の順で深掘り分析を行う場合の、分析担当者による最大操作回数を評価する。dCOMPASS では、年毎の増加傾向が強い条件を推薦する評価関数により、試行錯誤を支援する。MOLAP を用いた場合の操作回数と、dCOMPASS による分析の操作回数と、MOLAP と比較した dCOMPASS による分析操作数の削減率について、Table 2 に示す。

Table 2. 分析操作の最大操作回数

	スライ シング	ダイシ ング	推薦 回数	クリック 数
MOLAP	120	26,200	0	26,320
dCOMPASS	120	2,620	121	2,861
削減率	0.0%	90.0%	---	89.1%

Table 2 にて太枠で示している分析の操作回数 (クリック数) を分析操作作業量とする。MOLAP と比較し、dCOMPASS における分析軸推薦機能を用いた分析操作作業量は 89.1% 削減される。

6.2. 平均操作回数評価

本節では、実務上の平均分析操作回数の評価のため、

実際に試験協力者が MOLAP ツールと dCOMPASS を用いて同一の分析作業を行った実評価実験の結果を示す。

MOLAP の代表ツールとして Microsoft® Excel® 2003 ピボットテーブルを使用した。実評価実験は、試験協力者 7 人が 2.3 節にて説明した 3 つの分析タスクを分析評価する形式で実施した。評価対象のデータとして、2 つの項目がそれぞれ特定の値を持つ場合に、2.3 節の分析ニーズ 1~3 で目的とする傾向を強く示すよう作成したランダム生成データを用いた。このデータは、Microsoft® Excel® 2003 にて使用可能なデータ量の上限の制約から、6.1 節で使用したデータではなく、データサイズを 61,796 件 14 項目に限定した別のデータである。dCOMPASS で使用する評価関数として、各グラフの上昇傾向を推薦評価指数とする評価関数および下降傾向を推薦評価指数とする評価関数を使用した。実評価実験では分析開始から、分析ニーズ 1~3 の傾向が強く確認される絞込み条件を特定するまでの操作を評価対象とした。分析タスク毎に、分析操作の指標として、クリック数、分析判断のため注視した画面数、分析時間を計測した。実評価実験の結果として、7 人の計測結果の平均値を Table 3 に示す。

Table 3. 人の判断による操作回数 実験結果

分析 タスク	方式	クリッ ク数	注視 画面数	分析時間 [分]
1	MOLAP	92.1	20.6	9.15
	dCOMPASS	7.57	3.00	1.87
	削減率	91.8%	85.4%	79.6%
2	MOLAP	26.9	7.29	2.26
	dCOMPASS	8.29	2.71	1.51
	削減率	69.2%	62.8%	33.2%
3	MOLAP	106	18.7	7.04
	dCOMPASS	7.86	2.71	1.30
	削減率	92.6%	85.5%	81.5%
平均削減率		84.5%	77.9%	64.8%

MOLAP と比較して、dCOMPASS では、3 つの分析タスクに対して、クリック数が平均 84.5% 削減され、判断のため確認した画面数が平均 77.9% 削減され、分析時間が平均 64.8% 削減される実験結果を得た。

7. 考察

本章では、本論文の主要な新規性である分析軸推薦技術を構成する要素である「(1) 分析軸重要性評価技術」と「(2) 分析軸評価結果可視化技術」による分析作業改善への寄与について考察する。

(1) 分析軸重要性評価技術の効果

分析軸重要性評価技術の効果は、分析担当者の判断の回数の減少として観測される。Fig. 1 のフローチャートにおけるループ回数の減少効果を意味する。Table 3 にて、分析担当者が目的とする結果を得る

ために確認すべき画面数を平均 77.9%削減したことで、効果を確認することができる。

(2) 分析軸評価結果可視化技術の効果

分析軸評価結果可視化技術の効果は、分析担当者の判断に対する分析軸変更における分析操作量の減少として観察される。Fig. 1 のループ 1 回における分析操作量の減少効果を意味する。Table 3 にて、分析に必要なクリック数を平均 84.5%削減したことで、効果を確認することができる。分析時間に注目すると、分析時間の削減率は、クリック数および確認すべき画面数の削減率に比べて相対的に低い。

Table 3 の結果から算出した分析操作毎の所要時間を Table 4 に示す。クリック毎の分析時間は平均 2.39 倍、確認すべき画面数毎の分析時間は平均で 1.49 倍に増大していることが確認できる。分析担当者による操作・判断に時間を要していることが推定される。分析軸評価結果可視化技術において、分析軸集計結果可視化パネルにて複数のグラフを同時表示し、推薦する分析軸を強調表示して情報量が増大している影響から、MOLAP に比べて分析担当者が 1 画面を確認する時間が増加している。

Table 4. 1 操作当たりの所要時間(秒数)の削減率

分析タスク	方式	分析時間 [秒]	1クリック当たり秒数	1判断当たり秒数
1	MOLAP	549	5.96	26.7
	dCOMPASS	112	14.9	37.5
	削減率	79.6%	-150%	-40.4%
2	MOLAP	136	5.05	18.6
	dCOMPASS	90.9	11.0	33.5
	削減率	33.2%	-118%	-80.1%
3	MOLAP	422	3.97	22.6
	dCOMPASS	78.0	9.93	28.7
	削減率	81.5%	-150%	-27.0%
平均削減率		64.8%	-139%	-49.2%

dCOMPASS による効果について、Fig. 1 のループ回数を減少させる効率化において、分析軸重要性評価技術と分析軸評価結果可視化技術は同等の削減率を示している。Fig. 1 のループ毎の分析時間では、分析軸評価結果可視化技術による画面内情報の増加により、所要時間が平均 1.5 倍程度に増加していることが確認された。分析全体としては、ループ回数の削減効果により、dCOMPASS は分析時間を平均 64.8%削減する。

8. おわりに

本論文では、分析軸推薦技術を特長とした、多次元データ分析システム dCOMPASS による保守業務向け分析支援環境を提案した。実評価実験により、経年劣

化分析、保守後早期故障分析、保守回数別故障分析という分析タスクに対して、dCOMPASS による分析作業の所要時間を平均 64.8%削減した。dCOMPASS により、多数の分析軸を持つ大規模な保守業務データ向け分析作業の効率化の有効性を確認した。

今後の課題として、以下の 2 つが挙げられる。

- ・ 保守業務データ特有の分析事例のテンプレートを蓄積し、効果的な分析が得られる分析テンプレートの推薦など推薦機能の適用範囲を拡大すること。
- ・ 既存のデータ分析技術を簡便に取り込むことのできる分析処理プラットフォームとして、拡張開発を進めること。

参考文献

- [1] 岸原正樹・山本聖哲・佐藤淳: XML データベースへの統合と蓄積による保守情報の一元化と付加価値の創出, 東芝ソリューション テクニカルニュース, Vol. 14, pp.6-7 (2008)
- [2] 坂上聡子・岩田雅史: 昇降機現場知識共有システム, 三菱電機技報, Vol. 82, No. 9, pp.31-34 (2008)
- [3] 松本匡孝・瀬戸川教彦: ソーシャルテクノロジーによる企業知の戦略的活用, 日立評論, Vol. 92, No. 7, pp.536-541 (2010)
- [4] T. B. Pedersen, C. S. Jensen: "Multidimensional database technology", IEEE Computer, 34(12), pp.40-46 (2001)
- [5] E. Thomsen: "OLAP Solutions: Building Multidimensional Information Systems", John Wiley & Sons, New York (1997)
- [6] E. F. Codd, S. B. Codd and C. T. Salley: "Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate", Codd & Date, Inc. Technical Report (1993)
- [7] T. J. ティオリ: 「データベースの設計」, 勁草書房, 第一版, pp. 234-247 (2000)
- [8] Y. Zhao, P. M. Deshpande, J. F. Naughton: "An array based algorithm for simultaneous multidimensional aggregates", Proc. of ACM SIGMOD, pp.159-170 (1997)
- [9] 磯部成二・黒川清・塩原寿子・飯塚哲也: 視覚化による多次元データ分析システム: INFOVISER, 情報処理学会論文誌, Vol. 40, No. 5, pp. 2417-2428 (1999)
- [10] K. M. Azharul Hasan, Tatsuo Tsuji and Ken Higuchi: "An efficient implementation for MOLAP basic data structure and its evaluation", Proceedings of the 12th international conference on Database systems for advanced applications (DASFAA'07), pp. 288-299 (2007)
- [11] 土田隼之・都司達夫・樋口健: MOLAP用多次元配列構築のためのバッファリング方式, 日本データベース学会論文誌, Vol. 7, No. 1, pp. 19-24 (2008)
- [12] N. Lesh and M. Mitzenmacher: "Interactive data summarization: an example application", AVI 2004: pp. 183-187 (2003)