

オペレーショナルリスク管理に活用できる データ分析技術

Data Analysis Technology Focusing on Operational Risk Management

早川 ルミ

江川 誠二

小高 聡

■ HAYAKAWA Rumi

■ EGAWA Seiji

■ ODAKA Satoshi

東芝ソリューション（株）は、企業内に蓄積された多種多様な大量データの中から企業活動に役立つ知識を抽出するための、データ分析技術の研究開発に取り組んでいる。

今回当社は、金融機関のオペレーショナルリスク管理に活用できるデータ分析技術を開発した。業務作業量である事務量と、業務中に発生する事務ミス件数の実データを分析することで、事務ミスの発生条件を定量的に示すことができる。また、複数組織のデータを横断して分析することで、個々の組織の特徴を明らかにできる、などの効果を得ている。2011年5月に発足した金融機関ユーザー向けコンソーシアムにおいて、データの横断分析技術を適用し、ユーザー間での情報共有や事務業務改善の検討に活用していく。

Toshiba Solutions Corporation has been engaged in research and development of data analysis technologies to extract useful information and knowledge for business activities from large volumes of different data in companies.

We have now developed a data analysis technology for application to the management of operational risk in financial institutions. This technology can quantitatively show the conditions of occurrence of mistakes in office work by analyzing the correlation between the volume of tasks and number of mistakes from actual data, and can identify the characteristics of each organization in the company by cross-checking and evaluating the data of multiple organizations. We have been participating in a data analysis consortium for financial institutions established in May 2011, and will offer the results obtained by this data analysis technology to the consortium aiming at both the promotion of information sharing among users and the improvement of their office work.

1 まえがき

企業活動においては、効率化のために、業務の自動化やシステムによる支援が進められてきた。更にPC（パソコン）の処理能力向上により、短時間に複雑な処理を実現することが容易になっている。また、データベースの大容量化や多様化が進み、様々な企業活動で発生する、数値、テキスト、音声、画像など、多様なデータを収集し蓄積する仕組みも整ってきた。

このような環境の下、企業には、日々増加するデータを効率よくかつ安全に管理し、企業活動に活用することが求められている。このため、企業が保有する大量のデータに対して、“データの状態を把握するために可視化する”、“大量のデータから迅速に目的のデータを探し出す”、“複数のデータ間の関係を明らかにする”など、企業活動に役立つ情報の可視化や知識抽出に関する技術の研究開発が盛んに進められている。

東芝ソリューション（株）は、ソフトウェア開発での進捗や品質検査の数値データ、開発ドキュメントのテキストデータなどを総合的に分析することで、ソフトウェアの品質低下を防ぐ要因を発見したり、業務での作業量と事務ミス件数といった数値データを分析し、業務上のリスク管理などに有効なデータ分析手法の研究開発を進めている。これらの取組みは、大量デー

タの分析が可能となった環境を生かして新たな知見を得るために、ある特定の一組織のデータに着目して実施するものから、複数の組織を横断して実施する分析に広がりつつある。

ここでは、大量かつ多様なデータ分析を実施するうえでの課題と、金融機関のオペレーショナルリスクについてデータ分析を行う場合の実例と効果について述べる。

2 データ分析における課題

大量かつ多様なデータの分析結果を活用するユーザー側の視点に立つと、次の課題がある。

- (1) 大量のデータを保有しているが、どのデータをどのような視点で分析すればよいのかわからない。
- (2) データ分析から得られた結果を、どのように判定し活用することが適切なかが判断できない。
- (3) 他の組織ではデータ分析結果を活用してどのような成果を上げているのかを知りたい。

また、データ分析結果を提供する分析者側の視点に立つと、課題を次のように整理できる。

- (1) 組織によって、データ項目名称や採取しているデータの粒度、採取間隔などにばらつきがあり、そのままでは複

数組織のデータをまとめて分析することが困難である。

(2) 組織によって採取しているデータ項目が異なり、一部の組織ではあるデータ項目がまったく採取されていないという場合があるが、一般的な前処理方法では十分な対処ができない。

大量かつ多様なデータを十分に活用するためには、これらの課題を解決してユーザーニーズを満たすことができるようなデータ分析を行う必要がある。

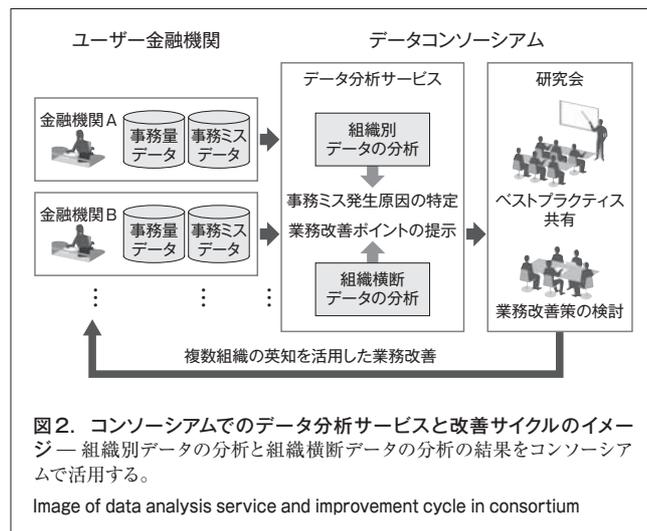
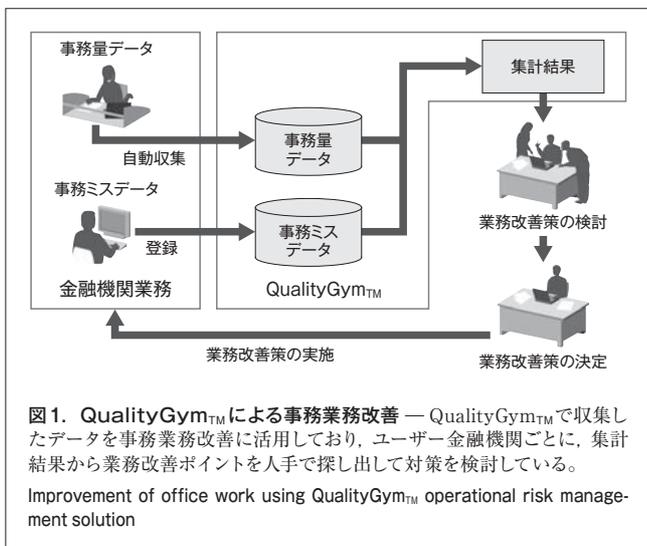
3 コンソーシアムにおけるデータ分析

当社は、金融機関のユーザー向けに、事務品質を向上させ業務改善をサポートするオペレーショナルリスク管理ソリューションQualityGym™を提供しており、既に20社以上への導入実績がある。QualityGym™は、時々刻々発生する事務業務の作業量を表す大量の事務量データと、事務作業中に発生したミスを表す事務ミスデータの収集、蓄積、集計が可能である。現在、ユーザー金融機関ごとに、集計結果から業務改善ポイントを人手で探し出し対策を検討している(図1)。

事務業務を改善し、事務品質を高めるために、当社は次のアプローチを取ることとした。

まず、単純集計にとどまらずに、統計解析やデータマイニングを用いて、事務ミスの発生条件や原因を追究できるような分析結果を提供する。分析方法としては、ある一つのユーザー金融機関における事務量データや事務ミスデータから事務ミスが発生しやすい条件を探る“組織別データの分析”と、複数のユーザー金融機関のデータを一括して分析する“組織横断データの分析”の2種類がある。

更に、ユーザー金融機関が参加できるコンソーシアムという“場”により、事務品質改善の活動について他社のベストプラクティスを知り、ディスカッションや情報共有できるようにする。



このため、前述の組織別データの分析結果や組織横断データの分析結果をコンソーシアム内で活用できるようにする(図2)。

3.1 分析対象となる事務品質データ

ユーザー金融機関においてQualityGym™で収集されている事務品質データは、次の2種類の内容を含んでいる。

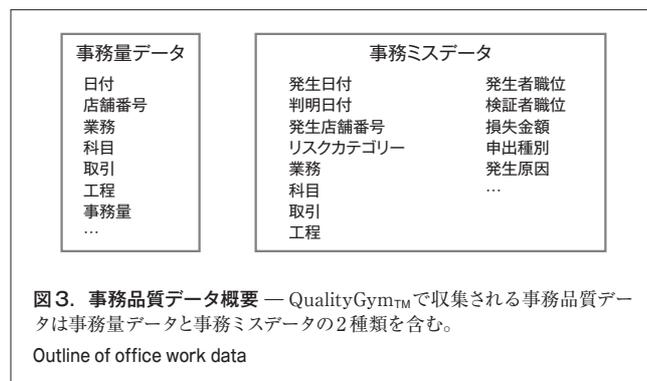
- (1) 事務量データ いつ、誰が、どの業務においてどのくらいの事務作業を実施したのかをシステムが自動的に収集したデータ
- (2) 事務ミスデータ “振込先まちがい”、“融資書類の不備”など、いつ、誰が、どの業務でどのようなミスを発生させたか、などの情報を人手によりシステム登録したデータ

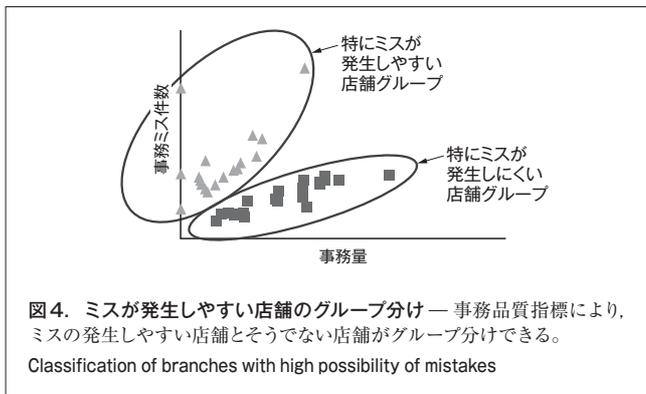
事務量データと事務ミスデータの概要を図3に示す。

3.2 組織別データの分析

組織別データの分析では、ある一つの組織だけに注目し、ミスの発生条件を探るためのデータ分析を行う。ある金融機関の実データを分析した事例を用いて、組織別に分析した二つのステップについて述べる。

- (1) 事務品質指標によるミスが発生しやすい店舗の特定 事務ミスの発生のしやすさを定量的に把握するため





に、事務量の大きさと事務ミス件数の大きさから算出する事務品質指標を新たに定義した。この指標を用いて、事務量の大きさに比べて事務ミス件数が少ない、つまり事務ミスが発生しにくい店舗と、事務量の大きさに比べて事務ミス件数が多い、つまり事務ミスが発生しやすい店舗を判別する。ある金融機関の全店舗にこの指標を適用した結果、事務ミスが発生しやすい店舗グループと発生しにくい店舗グループを分けることができた(図4)。

(2) アソシエーション分析^(注1)による事務ミスが発生しやすい条件の抽出 事務ミスが発生しやすいグループについて、アソシエーション分析を用いて、事務ミスが発生した“業務”とその他の状況の組合せを明らかにした。アソシエーション分析で出力される数千個の条件の組合せから、重要と判定されたものを抜粋した結果を図5に示す。重要度の判定には、次に示す三つの指標を用いた。

- (a) 信頼度 特定の条件との結び付きやすさを評価
- (b) 支持度 条件の組合せの出現しやすさを評価
- (c) リフト値 条件間の関係の強さを評価

アソシエーション分析で得られる条件の中には、図5の1、2行目にあるような、金融商品とその部分集合である生命保険の組合せや、金融商品とその処理がなされる預かり資産窓口という場所の組合せという、金融機関業務では当然の組合せも抽出される。

信頼度	支持度	リフト値	条件1	条件2
0.617	0.091	6.780	業務：金融商品	科目：生命保険
0.416	0.061	5.791	業務：金融商品	場所：預り資産窓口
...
0.651	0.075	5.163	業務：内国為替	原因：人員不足
...

図5. ミスが発生しやすい条件 — アソシエーション分析の結果からミスが発生しやすい条件の組合せが得られる。
Conditions of mistakes

(注1) 頻繁に同時に起こる条件を、相関の強い関係として抽出する手法。

当社は、この分析結果から当然の組合せを排除した後に、“内国為替業務”と“人員不足”という条件がそろると特に事務ミスが発生しやすいという組合せに注目し、内国為替業務について、なんらかの理由で一時的な人員不足が発生していると考えた。業務について詳しく調査した結果、内国為替は窓口が閉まる間際に取引が飛び込んでくることが多いため、このタイミングで一時的な人員不足が起こることが原因となり、事務ミスが発生しやすくなっている可能性があることがわかってきた。この結果から、閉店間際の時間帯だけ内国為替の対応シフトを変更する、などの改善策を検討できる。

3.3 組織横断データの分析

組織横断データの分析では、コンソーシアムに参加する複数の金融機関を対象とし、ミスの発生の仕方などに関して、共通点や相違点を明らかにするためのデータ分析を行う。このためには、複数金融機関のデータを一つの入力データとして扱う必要がある。QualityGymTMで収集されるデータは、基本的なデータ項目については統一されているが、個別の金融機関の事情に応じてカスタマイズしている部分があり、その点でデータが異なる箇所があるため、2章で述べたような課題がある。当社は、次の手段で対処している。

- (1) データ項目名称とそれに対応するデータについて標準形を定義し、各金融機関のデータ項目名称と標準形を対応付ける。
- (2) データの粒度や収集タイミングの問題に対しては、分析内容に合わせて方法を変えて対応する。
 - (a) 事務ミス“件数”などの実数ではなく、なんらかの比率に換算した値を用いる。
 - (b) どのような金融機関でも同程度の粒度になる“店舗”レベルのデータに分解して分析する。
- (3) あるデータ項目全体が欠損している場合は、他の観測されているデータを十分に活用して分析結果が得られるような新たなアルゴリズムを提案する。

クラスタリング^(注2)を行う場合について、今回提案したアルゴリズムと従来手法の違いを図6に示す。従来手法は、クラスタリング分析の前処理として、欠損値を一つ以上含む項目を削除するが、今回提案した手法は、項目の削除をせずに観測されているデータを十分に活用してクラスタリングを行う。

今回提案したアルゴリズムの効果を調べるために、図6の状況を想定して、試験データで従来手法との比較を行った。提案手法の精度は、従来手法より10%高いという結果が得られた(表1)。

複数の金融機関を対象に、全金融機関に共通する性質や、ある金融機関にだけ見られる特徴を明らかにするため、五つの金融機関の事務品質データを用いて分析を進めている。現

(注2) 与えられたデータを自動的に分類する手法。共通する特徴を持つデータが一つのクラスタにまとめられる。

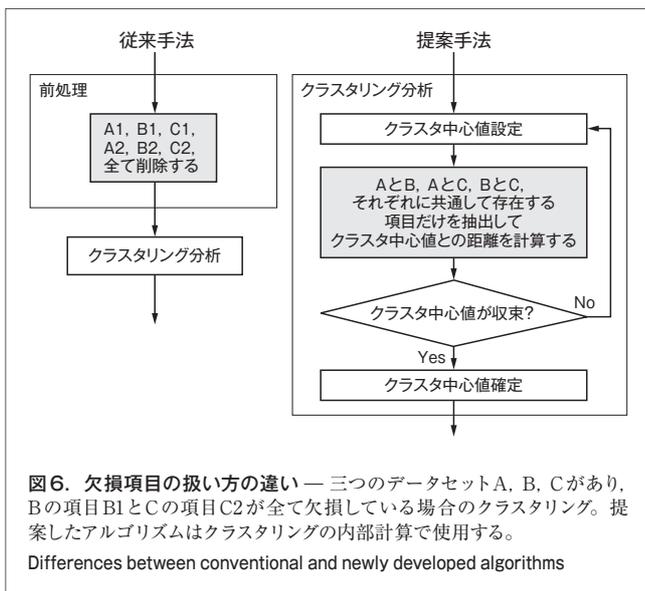


表1. 欠損項目があるデータをクラスタリングするアルゴリズムの精度比較
Comparison of accuracies obtained by conventional and newly developed algorithms

データとクラスタリング手法の組合せ	精度 (%)
欠損項目がない完全なデータ×従来手法	73
欠損項目があるデータ×従来手法	53
欠損項目があるデータ×提案手法	63

在, 一部のクラスタリングから次の結果を得ている。

- (1) ある特徴を持つクラスタ (例えばクラスタKとする) に含まれる店舗の90%は, ある一つの金融機関 (例えば金融機関Lとする) の店舗である。
- (2) 金融機関Lの店舗の60%は前記クラスタKに含まれる。これは, クラスタKが金融機関Lの特徴をそのまま表しているものと考えられる。複数の金融機関データを合わせて分析することで, 個々の金融機関ごとの特徴を浮き彫りにできる効果があると言える。

4 データ分析結果の活用

組織別データの分析により, 組織ごとにミスが発生しやすい条件を明らかにできる。また, 組織横断データの分析により, 複数組織に共通したミスの出方の特徴, 逆に, ある組織にだけ見られる特徴を明らかにできる。コンソーシアムでは, これらの分析結果をユーザー金融機関に提供し, 業務改善に向けて様々なディスカッションと情報共有を進める。更に, 業務改善策を適用後の各金融機関の状況を分析し比較検討すれば, 対策の評価と更なる改善策を検討できる。複数のユーザー金融機関が参加する場でデータ分析結果を活用すること

で, オペレーショナルリスク管理のPDCA (Plan-Do-Check-Act) サイクルをより強固なものとして実現できるようになる。

5 あとがき

コンソーシアムが発足して以降, データ項目名称の標準形の定義など, これまでに基本的なディスカッションを進めてきた。また, 参加ユーザー金融機関の事務品質データを分析し, 2012年春に開催予定のコンソーシアムの研究会では, 今回紹介した組織横断データの分析結果を紹介する予定である。コンソーシアムでは, ユーザー金融機関が自組織の業務改善の手掛かりを得られるようサポートするほか, ユーザー金融機関とのディスカッションにより得られるニーズや課題に応えるために, 今後もデータ分析方法の改良と提案を進める。

更に, 次に挙げるような, 別の観点からの分析や他分野への横展開の可能性も視野に入れている。

- (1) 文書分析技術との連携 数値データに関連する自由記述文書も対象とすれば, より正確で詳細な分析ができる可能性が高い。
- (2) 金融機関以外の業種への展開 今回紹介した分析手法や成果は, 様々な業種や業務に活用できる。
- (3) 組織外部のデータを用いた分析 気象データ, 経済データ, 新聞記事データなどもいっしょに分析すれば, まったく新しい知見が得られる可能性がある。

今後も, ユーザーニーズを捉え, 種々の課題を解決するために, 新たな技術の創出や既存技術の強化に取り組んでいく。

文 献

- (1) 東芝ソリューション. “新しいステージを迎えたオペレーショナル・リスク管理の高度化”. 第65回 金融機関リスクマネジメントセミナー. 東京, 2010-09.
- (2) 江川誠二 他. “企業の業務改善に向けた作業履歴データ分析”. 情報処理学会 第73回全国大会. 東京, 2011-03, 情報処理学会. 2011, p.4-407-4-408. (DVD-ROM).



早川 ルミ HAYAKAWA Rumi

東芝ソリューション(株) IT技術研究所 研究開発部研究主務。情報知識利活用技術の研究・開発に従事。日本統計学会, 人工知能学会, 言語処理学会会員。
Toshiba Solutions Corp.



江川 誠二 EGAWA Seiji

東芝ソリューション(株) IT技術研究所 研究開発部。情報知識利活用技術の研究・開発に従事。情報処理学会会員。
Toshiba Solutions Corp.



小高 聡 ODAKA Satoshi

東芝ソリューション(株) 流通・金融ソリューション事業部 金融ソリューション技術第一部主任。金融機関向けソリューション構築の技術支援に従事。
Toshiba Solutions Corp.