

## 序章 連環データ分析とは

### 1. 始めに

#### ■連環データ分析とは

“連環データ分析” (DCB Analysis: Dual ComBine Analysis) は、クロス表の形式で表現されたデータを集約展開し、MAP化するソフトウェアである。複雑なデータを理解し易いように図示することで、データから有意な情報や、有用な知識を掘り出すことを支援するパーソナルBI (ビジネス・インテリジェンス) ツールである。

#### ■連環データ分析におけるクロス表とは

通常、“クロス表は、表の行側(表側)に付けられたアイテム名の集合と、列頭(表頭)につけられたアイテム名の集合の2組の集合が取り上げられ、それらのアイテムの関係を表現したデータ” であると考えることができる。

連環データ分析では、表側のアイテムの集合は分析の主たる関心の対象として“サブジェクト”と呼ぶ。表頭のアイテムの集合はサブジェクトを説明するもので、アトリビュートと呼ぶ。サブジェクトとアトリビュートの各アイテムを結び付けているデータ(表体)は、“連環度”というデータ値として連環データ分析ソフトに入力する。

例えば、ある商品の購入者300人のサンプルをとり、その性別と年齢を調べたとき、サブジェクトのアイテムは300人の個々の顧客であり、性別のアトリビュートは男女の2アイテムとなり、年齢のアトリビュートは、年齢を10代毎に区切れば、10代、20代、、60代の6個のアイテムとなり、全体は300行×8列のクロス表として表わされる。

この300人を性別と年齢の組み合わせで集計して、性別の2行と年齢の6列に集計し集約したものは、クロス集計表と呼ばれる。連環データ分析は、クロス表でもクロス集計表でも扱うことができる。また、連環データ分析では、あるサブジェクトまたはアトリビュートを共有して、複数のクロス表を、統合して扱うこともできる。

#### ■連環度について

連環度は、サブジェクトとアトリビュートのそれぞれのアイテム同士を結びつける強度を示す非負の値である。結びつき

の強度が高いほど大きな値となり、低い場合は小さくなる。

従って、マイナスのデータを含む場合や、結びつきの強度が逆に示されている場合は、データの変換が必要となる。

上の300人の顧客の例では、初めのクロス表のあるサンプルが男性ならその列に“1”を女性の列に“0”を立てる。年齢も同様に該当する列に“1”をたてます。クロス集計表では、男性で10代のセルにその人数が入ることになる。

連環度の“環”は“玉の輪”を示す言葉である。連環データ分析は、大切な輪がアイテム間にある強度を持って繋ぐネットワーク構造を分析する方法。

## 2. 連環データ分析の特徴

従来のデータ解析法との機能の違いについて

■サブジェクトとアトリビュートの同時布置ができる。

連環データ分析は、クロス表の情報を圧縮して、少数次元空間上に各アイテムの座標を決め、サブジェクトとアトリビュートの各アイテムを同じ空間に同時布置ができる。また、サブジェクトのアイテムは、アトリビュートのアイテムの中で連環度が高いものと近くなるように布置ができる。逆に、アトリビュートのアイテムとサブジェクトのアイテムとの関係も同じように布置ができる。

その結果、サブジェクトとアトリビュートのアイテムの似たもの同志が、それぞれ近くなるように空間座標として布置されてMAPが造られることになる。こうした同時布置は、従来の対応分析、双対尺度法、林の数量化3類などではできなかった。

例えば、就職先人気企業のアンケートの集計結果を連環データ分析にかけることを考えてみよう。企業をサブジェクト、選社理由をアトリビュートとし、選んだ企業の選択理由をクロス集計した表を入力する。結果のMAPでは選社理由の似ている企業が近くに配置される。また同時に、似た企業を選ぶ要因となる選社理由同士も近くに布置されることになる。

■アトリビュートの三つ以上のアイテムの相互関係を扱うことができる。

従来相関分析や正準相関分析では、複数のサブジェクトに対し、一対のアイテム（2つのアイテムまたは変数）しか扱え

なかった。3つ以上のアイテムがある場合でも、一対のアイテム(正準変数)に変換し、その相対的な大小関係の整合性の強さしか扱えなかった。つまりサブジェクトは多数でも、アトリビュートのアイテムは一対または一対のアイテムに変換して扱うことが前提だった。またその大小関係の整合性ということで、データは、計量値に限られていた。

例えば、日本人の5歳別に切り分けた世代をサブジェクトとし、その趣味の種類をアトリビュートとし、その連環度を関心のある割合とするクロス集計表を扱うことを考えよう。趣味のアイテムが3つ以上の集合で、その連環度をその世代の関心の%という計量値の場合、相関分析や正準相関分析では、幾つかのペアーを造らないと分析できなかった。

■三つ以上のクロス表の相互関係を扱うことができる。

連環データ分析では、2つ以上のクロス表を扱える。ただし、サブジェクトが共通という条件で、アトリビュートという表頭の集合自体が複数でも扱うことができる。アトリビュートが共通の場合は、サブジェクトの集合が複数でも可能となった。

例えば、上の世代別の趣味の動向の例で、さらにもし、それらの世代とビールのブランドの好のような二組目のアトリビュートとの関係を分析することが可能となった。

■複数のアイテムでも平等に扱うことができる。

従来、回帰分析等や林の数量化1,2類などでは、目的のアイテムを決め、それが他のアイテムでどの程度説明できるかという、1つの従属アイテム(変数)に対し、独立な複数のアイテム(変数)の重み付きの組み合わせモデルを造っていた。そこでは、目的アイテムの見かけの説明精度を上げるために、説明アイテム間の関係を歪めてしまい、得られたモデルを脆弱にする弊害が指摘されてきた。

また、たとえ簡単な2つのアイテムの場合であっても、得られたモデルは、逆の説明には使えなかった。

連環データ分析では、全てのアイテムを平等に扱い、頑健性があり、可用性(バーサビリティ)の高いモデルを得ることが可能になった。

■サブジェクトとアトリビュートの同時クラスタリングが可能。

サブジェクトに対して似たアトリビュートのアイテム同士

が近くに布置され、またアトリビュートに対して似たサブジェクトのアイテム同士が近くに布置される。さらに共通するサブジェクトの対し、それらを特徴付けるための別のアトリビュートのアイテム同士も近くに同時布置できる。

そして、それらの各アイテムの似たもの同士を同時クラスタリングすることができる。このようなソフトは、従来なかった。

具体的には、世代別趣味の種類と差代別の好みのビールブランドを、同じMAP空間で、同時クラスタが可能になった。これによって、分類された世代の特徴を理解したり、ビールのブランドイメージのポジショニング戦略を検討する新しい方策を検討することができる。

### 3. 連環データ分析の優位性

連環データ分析を応用する場合の、多様なデータタイプを扱える優位性から来るアドバンテージは、以下の通り。

■連環度が多様なデータのタイプのクロス表の分析ができる。

従来のデータ解析の手法は、質的データや量的データ等の扱うデータタイプによって分析法が分かれていた。そのため、ソフトも別々なものを用意して、使い分けが必要でした。連環データ分析では、連環度が、質的データ、量的データでも扱えるようになった。

例えば、日本の男性をランダムに300人サンプリングしてそれらをサブジェクトとし、幾つかのビールのブランドを“好き”、“嫌い”、“どちらでもない”で答えてもらったとき、アトリビュートは3つのアイテムとなり、答えが各アイテムに該当したら“1”、そうでなかったら“0”として質的なデータを、連環度データとして扱うことができる。

また、単に好みの程度ではなく、各ブランドをアトリビュートのアイテムとしその消費量がデータとして得られた場合は、消費量をそのまま連環度として、分析することができる。

■アトリビュートのアイテムが量的なデータタイプのクロス表の柔軟な分析ができる。

サブジェクトに対し、アトリビュートのアイテムが量的な属性を持つ場合、それらの量の範囲を区分し、それらをアイテムとしてラベルを附加し、サブジェクトのアイテムがその区

間に属するか否かを、“1”、“0”を着けて連環度とすることで、量的なデータを扱うことができる。

従来は、量的なデータは、その“量に比例した性質を持つこと”を前提とした分析が多く、そのため量の多寡に比例しない問題では、結果が歪むことがあったが、連環データ分析では、そうした前提に捉われない、柔軟な分析が可能になった。

例えば、年齢によって趣味がどのように変わるかという問題は、必ずしも年齢に比例して、ある種の性質が強くなるとか弱くなるとかいった傾向が現れるわけではない。また、投与する薬も量に比例して効果が高くなるとは限らない。

■アトリビュートのアイテムがランク等の順序データを持つクロス表の分析ができる

あるサブジェクトに対し、アトリビュートのアイテムがランクのような順序を持ったデータを扱う場合、ランク自体をアイテムとして、サブジェクトのアイテムが該当するか否かで、そのランクに該当するか否かで“1”、“0”を連環度とすることで、扱うことができる。

従来は、順序データは、それが説明アイテムとした場合に、より高いランクがそれに比例した効果が現れるか、また、逆に順序データを目的アイテムとした場合に、説明要因アイテムの増減と上手く比例関係や加算関係を使って説明ができるか等の扱いが難しく、そのため有効な分析手法がなかった。

連環データ分析では、そうしたランクのような順序を属性として持つアイテムや、ランクそのものをアイテムとすることで、柔軟な分析が可能となった。また、ランクの効果の連続性を考慮して分析することができる。

例えば、知識創造の場としてオフィスの職場満足度を上げたい場合、どのような要因がどの位効いているかのモデルを得て改善に結びつけることは、ちょっと、難しい問題だった。

#### 4. 連環データ分析の利点

ユーザにとって、多種の情報源を統合できるベネフィットは以下の通り：

■テキストデータから意味のある情報を獲得することができる

知的ワーカーにとって、ビジネスで重要なデータは、議事録、講演、論文、メモなどの記録に80%があると言われる。

従来、いわゆるテキストマイニングでは、これらからキーワードを切り出し、それらをタグ付けして構造化し、情報を取り出す言葉のネットワーク図や、ツリー図やキーグラフとして作成して、言葉の関係性の概念図の作成で終わっていた。

また、これらのネットワーク・ダイアグラムは、解釈が難しく、検索の重要度付けやレコメンデーション等重要な要因との関係付けが困難で、特に、他の複数のデータとの組合せに対応できないという制約があった。

連環データ分析では、こうした制約を、各アイテムに最適な空間座標を定めるため、こうした難点のブレークスルーを可能にした。

■テキストや数値データが混在する多種のデータを統合して意味のある情報を獲得できる

ICTの進展によって、知的ワーカが日常的に処理しているビジネスで重要なデータは、個人のPC、ノートやメールの中に、孤立し、無秩序のいわばゴミ箱状態で、格納されており活用することができなかった。それは、扱えるデータタイプが異なることにより統合した処理が難しかったこと、また、複数のクロス表を扱うことが難しかったこと、およびそれらを平等に扱う方法が無かったからである。

連環データ分析では、多様なタイプのデータが扱え、また多くのデータのタイプが混在したクロス表を組み合わせで分析できることから、今までは関連すると判っていても、それらを統合して解析することが困難だったが、そうしたデータを構造化し、統合して有用な役に立つ知識として回収することが出来るようになってきた。

例えば連環データ分析では、ある顧客の各種の属性やパフォーマンスが、どのような理由で決まってくるかということ、その顧客の言動から分析することも可能となった。つまりある目標指標（KPI）を改善するのに、言葉や動作を表現するような質的な変数の関数で表現するモデルを使うことも可能となった。

例えば、知的ワーカの職場満足度とその要因関係のモデル化が可能となった。満足度の高い場合の言動や低い場合の言動をアイテム化し、連環データ分析で、モデル化することが可能となった。

■成功の要因理解と新しい成功のコンセプトの開発を支援

従来は、現在成功しているサービスや商品の成功の理由、背

景や消費者が抱くイメージの解明が困難だった。そのため、それらを継続させ発展させる新しいサービスや商品のコンセプトの開発には、適切な方法が見つからなかったと言える。また似た商品やサービスがヒットしない理由の解明も困難があった。

連環データ分析は、コンセプトの分布を可視化し、現在のイメージポジションを明確にし、そこにおける強み・弱味を調べ、脅威やチャンスに関する仮説を得ること、また、チャンスと思われるイメージポジションに関するプロフィールを検討し、可能性がある仮説の立案を強力に支援する。

例えば、ある化粧品の新製品がヒットしても、その効用のどのような特徴が受け入れられ、どのようにリ・イノベーションの伝播が起きているのかの理解は難しい問題である。先進ユーザ層の集まるSNSの解析では、効用の変化の追従には限界があり、プロダクト・ライフサイクル・マネジメントには、困難があった。

#### ■ アイテムやその関係する要因の推移を読むこと

サブジェクトを説明するアトリビュートの値が時系列的に変化していく場合などに、他のアイテムの配置場所との比較により、アイテムのトレンドの変化を読み取ることが出来る。

複数の自動車メーカーの年代の組み合わせをサブジェクトとし、販売台数をアトリビュートとし、その年のそのメーカーの出荷台数を連環度とし、連環データ分析にかけることを考えてみます。結果では販売台数の近い、年毎のメーカー名が近くに配置できる。ひとつのメーカーに注目すると、メーカーの販売台数の変化がMAP上の位置の移動により示すことが可能。つまり、アイテムの時系列的な量的な変化やトレンドを他のメーカーとの比較で読み取ることが出来るようになる。

## 5. イノベーション支援

連環データ分析には、大きく2つの応用分野が考えられるが、誰が、何に使うと便利か、ここでは、イノベーションの支援に関して説明する。

#### ■ 新製品・新サービス・新概念開発

製品・サービス・ビジネスの企画担当者や、開発・設計担当者が、新製品、新サービス等の開発にあたり、新しいコンセプトや新しいニーズ発掘に利用すると便利である。

## ■マーケティング企画・宣伝プロモーション企画

マーケティングやセールスプロモーションの企画戦略担当、広告宣伝の企画立案担当者が、新しいコンセプトや、新しいニーズ定義、新しいキャッチフレーズの発見や新しいチャンネル発掘のヒントを得るのに便利である。

◎従来何となく気がついていて、データを示して説明ができなかった提案を説明し易く支援する。また、従来なぜ良かったのか、なぜ飽和してきたのか等の、新しいチャンスの可能性や、新しい脅威の可能性を、従来は手に負えなかったデータに基づいて仮説として提言する可能性を広げる。これらの問題は、あるべき姿 (To Be Position) が明確でないいわゆる構造化されていない難しい問題ですが、意外と現実の姿 (As Is Position) 自体が明確に認識できていないことが多く、連環データ分析は、そうした問題にも有効性がある。

### 6. K P I 改善支援

連環データ分析は、大きく2つの応用分野が考えられるが、誰が、何に使うと便利か、ここでは、K P I : キー・パフォーマンス・インジケータの改善支援に関して説明する。

■製造担当、資材担当者などが、日々の活動の歩留やコスト改善などのK P I に関する種々の要因や条件の差異点を見つけ、改善提案に結びつけることを支援する。

さらにK P I への要因別の影響度を調べることも可能である。

■営業担当や流通担当者などが、売上や納期達成、またサービス活動当りコスト等を改善するため、関連するサービスや担当者やそれらの活動に関するいろいろな属性との差異点を明確にし、することで、K P I の改善に結びつけることが容易になる。

◎評価インデックスの改善のためには、何がパフォーマンスを決める要因であり、何がビジネスの価値を上げるドライバーであり、そのビジネス資源の配分や活動のベストモードを探ることが大切となる。イノベーションは、そうした外的規準が無い問題であるが、改善活動では、ビジネスパフォーマンスという外的規準があるので、構造化された問題の中でも、取り組みが比較的明確な問題といえる。

### 7. 連環データ分析の原理

■多重対応分析法の拡張に成功



従来のデータ解析では、相関度や回帰度等が説明モデルの主流だった。単一のクロス表から情報を抽出する対応度を指標とする対応分析(Correspondence Analysis)がフランスで生まれ、現在は、サブジェクトを共通とする複数のクロス表から情報を抽出する同質性分析(Homogeneity)がオランダで成長中です。これをさらにデータ・ケーキベーカ(株)が、拡張したのが連環データ分析である。

拡張された方向としては、同質性分析がまだ、連環度として計数型のデータに限られていたが、連環データ分析では計量型や順序型も扱えるようにした。また、サブジェクトだけが共通な複数のクロス表だけでなく、かなり自由なクロス表の配列と組合せを扱えるようにした。さらに多くのサブジェクトやアトリビュートの各アイテムを同時布置できるようにしたこと等がある。

#### ■データを圧縮し展開して意味のある情報を読み取りやすくするための可視化

一般に、クロス表は、アトリビュートのアイテムの数だけの次元の空間にサブジェクトのアイテムを布置すれば、完全にデータの持つ全ての情報量を表現することができる。逆にサブジェクトのアイテム数の次元空間にアトリビュートのアイテムを布置しても情報は失われない。連環データ分析では、これらのデータを情報圧縮し2～3次元の空間にマップ化し可視化する。つまり、情報を一端圧縮した後に、データの意味を解釈して味わえるように、展開し可視化する。データを圧縮して固め、焼いて膨らませて美味しくし味わえるようにしている。

#### ■計算の原理

サブジェクトのあるアイテムが、対応するアトリビュートの中でより強い連環度で連環されたアイテムとより近くなるように空間座標を計算し、また逆にアトリビュートのあるアイテムが、対応するサブジェクトの中でより強い連環度で連環されたアイテムとより近くなるように空間座標を計算している。

複数のクロス表が有る場合は、関連するクロス表の各サブジェクトやアトリビュートの空間座標は、それらが重なるように、1) 原点を合わせ、2) スケールを合わせ、3) 方向を合わせる必要がある。つまり、似たアイテム同士を近くに布置するというだけでは無数にあり得る空間座標を限定することが出来ないという問題もあるからである。

## ■ 自己連環性と相互連環性の両立する空間

サブジェクトにおけるアイテムの特徴はアトリビュートで説明され、アトリビュートのアイテムの特徴もまたサブジェクトで説明されます。また、同時にそれぞれの同じアイテム同士でもその特長が説明できる必要がある。

クロス表を判り易く可視化するためには、サブジェクトの特徴がアトリビュートとの連関度のプロファイルで決まり、逆にアトリビュートの特徴もサブジェクトとの連環度のプロファイルで決まるということから、相互連環の対称性が必要となります。

サブジェクトとアトリビュートの関係である相互連環性は、それらのアイテムがなるべく重複が少ない形で同時クラスタリングされ類型化されたとき、特徴がより強くより明確であると言える。また、サブジェクトのまたはアトリビュートの中のアイテム同士の関係である自己連環性は、なるべく重複が少なくなる形でクラスタリングされ類型化されたとき、特徴がより強くより明確になると言える。

サブジェクトのあるアイテムのイメージを説明する問題では、相互連環性で説明するか、自己連環性で説明するかのどちらかでしかできない。つまり、そのサブジェクトを説明するアトリビュートの集合との関係のプロファイルで説明するか、サブジェクトの集合の中で似たプロファイルを持つアイテムで説明するか以外には、方法は無いと言えるからである。

以上