

F A Q

連環データ分析の全体に関する事項

Q1. 連環分析とは？

A1. 従来、相関分析がありましたが、2種類の属性の大小関係に絞って、その対応の度合いを調べるものでした。連環データ分析では、多くの属性の繋がり方の度合いの関係を可視化するためのものです。

従って、相関関係をもっと広げた考え方とも言えます。また属性単独の大小関係ではなく、属性と属性の繋がりや近さ関係だけに注目しておりますので、相関関係のように一对の属性自身が計量値や、順序値などの数値だけではなく、複数の区別できるアイテム(定性的、質的なデータによる)属性群であっても扱うことができます。

連環分析には、今のところ3種類がありそうです。

「連環データ分析」では、サブジェクト(表側)のアイテム(要素)と、アトリビュート(表頭)のアイテム(要素)との環連性の強さを、連環度といい、そのデータを処理するものです。

「連環情報分析」では、関心のある基準を持った尺度、あるいはその基準となるクロス表を持ち、それを説明することによって、データから意味のある情報を抽出する分析をさします。

また、「連環情報分析」では、意味を持つ言葉を属性に持ち、文章を構造化し、外部に通用する言葉と繋ぎ、誰でも使える知識を抽出する分析を扱うものとしましょう。

Q2. 連環度とは？

A2. 連環度は、ある想定したサブジェクトとそれに関わるアトリビュートのそれぞれの要素(アイテム)の繋がり方の関係を表現するための尺度です。各アイテム間の繋がりが強いほど大きな値とし、弱いほど小さな値とします。ただし連環度は、全て0以上の正の値でなくてはなりません。

一般に、ネットワークモデルでは、有向性を持つ性質の連環度を扱うこともあるが、連環データ分析では、イメージや連想等の相関度の拡張として、双方向性を尊重するため、それを扱わない。ただし、クロス表の一般配置型の連環度分析では、クロス表の間のネットワーク分析の中では、それを扱う必要があるかもしれない。

Q3. 連環とは？

A3. 連環とは、“環を連ねること”、“連環馬：中国では、騎兵が乗馬を鎖で繋ぎ合せ、一斉に敵陣に突撃すること”(広辞苑)。

“環 “とは、”玉の輪”、”Rung: 集合の2要素の間に、2種類の算法が定義されており、この算法について通常の数の中の加法・減法と同様に、結合法則・分配法則があり、場合により乗法は別として、加法に対する交換法則が成り立つ集合をいう(広辞苑)”。

因みに、英語では、Dual ComBine としておりますが、これは、日本語では双対結合とする方が、合うかもしれません。これは、サブジェクトはアトリビュートによって説明され、アトリビュートはサブジェクトとによって説明される、相互関係によっ

て、お互いが存在しているからです。

Q4. クロス表の連環性とは？

A4. 相関性の場合、2つの属性について、Aの属性が大きければ、Bの属性は大きい（あるいは小さい）ということのなり、逆に、Bの属性が大きければ、Aの属性は大きくなる（あるいは小さくなる）、というように可逆性が成り立っております。しかし、連環データ分析では、“Aの中で a1 ならば、Bのなかでは b1.である”と言っても、“Bの中で b1.ならば、Aの中では a1.である”と両方のメッセージが成り立つとは限りません。

これが両方が強く成り立つとき、サブジェクトとアトリビュートの連環性が高いと言うことにします。

Q5. アイテム間多次元連環度とは？

A5. クロス表を情報圧縮して、元の次元より少ない次元空間で、連環マップを作り、サブジェクトとアトリビュートのアイテムを、そのマップ上に同時布置しますが、[0,1]度の大きなアイテム同士が近くなるように計算されております。しかし、アイテムのネットワークが、力の引き合いのバランスで、結果が決まりますので、最終的なアイテム毎の近さ関係は、マップの空間における原点から見込む角度が狭いほど“近く”なります。この距離を多次元空間における原点からの見込む角度の余弦（コサイン θ ）で相関係数として使うことも有りますが、連環データ分析では、これに(+1)を加えて、[0~2]までの数値に変換し、マップ上で“近い”アイテムほど大きな値として、アイテム間多次元空間度としております。

Q6. 連環マップとは？

A6. あるサブジェクトとアトリビュートと、それらのアイテム間の連環度で構成されたクロス表データを、情報圧縮技術で少数次元に縮小して近似した空間を構成し、可視化した図です。

通常、n個のサブジェクトとm個のアトリビュートからなるクロス表は、全体ではnかmの少ない次元空間で、全ての情報量を表現できますが、2~5次元空間に縮小したとき、nやmが元々大きい場合などでは、近似度が落ちることになります。しかし、人間が見易く、理解しやすくなります。

連環データ分析のアドバンテージは、同じ図面に、サブジェクトとアトリビュートの各アイテムを、同時布置します。また、同時にクラスタリングも可能となりました。このことによって、連環性を読み取ることが、比較的容易になりました。

Q7. 対応分析との違いとアドバンテージは？

A7. 対応分析 (Correspondence Analysis) や 多重対応分析 (Multiple Correspondence Analysis) も、連環度マップと同じアプローチで、可視化を実現しております。ただし、連環度（彼らはそう呼んでいませんが）を、[0,1]か計数値(数えあげられる整数値)に限っており、まだ、順序値や計量値を扱って居りません。

また、アイテムの同時布置も、明確な基準で、実現できておりません。そのため、同時クラスタも、実現できて居りません。

Q8. 連環度マップにおけるアイテム間の近さとは？

A8. 基本的に連環度の強いアイテム同士は、連環度マップ上で、より近くに布置さ

れ、弱いアイテム同士は、より遠くに布置されます。

とはいえ、クロス表で見て、あるアイテム同士で、大きな連環度データがあってもそれらがマップ上で必ずしも近くに布置されるとは限りません。それは、他のアイテムとの引っ張り合いで、ポジションが決まるからです。あるアイテム同士の連環度が小さくても、他のアイテムとの連環度が全て小さければ、そのアイテムとの全ての連環度ネットワーク作用で、結果として、近くなることもあり得るのです。

連環マップの見方

Q1. 軸の意味は？

A1. 連環マップに現れる軸には、特定の意味はありません。それは、使う方々の解釈に委ねられます。しかし、コンピュータが計算する過程で、最初に最も情報量を多く反映できる（データ全体のバラツキを最も反映できる）軸を第一軸としてありますので、多くの場合、連環度の大きなアイテムG p. とそうでないG p. を分離する軸が現れることが多くなります。（DCB Analysis では、デフォルトでは、第一軸を横軸、第2軸を縦軸としております）

第2軸は、第一軸と直交して、残りの情報量をできるだけ反映させた軸として、設定されています。第3軸以降も同様に計算されております。

Q2. アイテムの方向の意味は？

A2. アイテムの位置は、原点からの方向に意味があります。

サブジェクトあるアイテムと、アトリビュートのあるアイテムが同じ方向にあれば、それらの連環性は、他の組み合わせに比べ、強いと判断できます。

同様に、サブジェクトのアイテム同士が同じ方向にあれば、それらのアイテムの連環性は、他のアイテムより強いといえます。アトリビュートのアイテムについても同じです。

Q3. アイテムの原点からの距離の意味は？

A3. アイテムの意味は、原点からの方向で決まりますので、同じ方向にあるサブジェクトやアトリビュートのアイテムを見ることによって、意味を掴むことができます。

アイテムの原点からの距離が、その意味の強さを示します。従って、方向の意味を理解するには、原点から遠い距離にあるアイテムに注目することで、意味のある情報の読み取りが、容易になります。

Q4. 連環マップ上に描かれている三本のサークルの意味は？

A4. 連環マップには、三本の破線のサークルが描かれております(デフォルトで)。これは、アイテムの分布の標準偏差を示しております。各軸のプラス、マイナス1シグマ、2シグマ、3シグマの区切りを通る円を表しております。

従って、通常、3本目の破線を飛び出すアイテムは、あまりありませんが、著しく飛び出した場合は、異常値である可能性があります。

また、一番内側の破線の内部のアイテムは、他のアイテムから遠いか、あるいは、全てのアイテムと強い関係があつて、中心に近くなつており、意味が不明確であることを示しております。従って、中心に近いアイテムの特徴を読むことは危険性があります。

Q5. 基本的なクラスタの標準的な使い方は？

A5. 社会現象では、正規分布が現れることは極めてまれで、むしろベキ乗分布がよく現われます。事前の何らかの知識がなく、全体を理解したい場合は、クラスタの数を3つに設定して分析することを、お奨めします。

軸の意味の解説で述べましたように、第一軸が、数値の大小を分離しますと、非正規分布では、小さな値の連環度を持った多数のアイテムが座標の中心近くに分布し、大きな連環度を持った少数のアイテムが座標の中心から遠くに飛ばされます。

そして、第2軸は、これらの少数を分離してくれます。従って、3個のクラスタリングを行いますと、多くの場合、1つのマジョリティと、対抗する2個のマイノリティのクラスタが得られます。これは、マーケティングでは、しばしば、かなり有効な知見が得られることがあります。

データの処理の仕方

Q1. 平準化の役割は？

A1. 社会現象では、正規分布よりもベキ乗分布が多く現われます。連環データ分析では、連環度として、計量値と同時に離散数やそのアイテム（例えば、男性か、20歳代か等）に属するなら[1]、そうでないなら[0]等の数値を同時に扱うこととなります。計量値は、4桁や9桁などになることもあります。こうした桁数の異なったデータを別なアイランドに入れるにしても、バランスが失われたり、収束するまでの計算時間が掛かることがあります。

また、個別の表体（セル）の値に拘るのではなく、サブジェクトの特性と、アトリビュートの特性を見ることが大切な場合には、平準化オプションを使います。

従来の統計学における分散分析では、行の効果と列の効果とを、加算分解モデルで分離しておりましたが、連環データ分析では、それぞれの積分解モデルを使っています。

Q2. 計量値の扱いは？

A2. 計量値は、そのまま、直接連環度として、利用することができます。しかしアトリビュートに計量値のアイテムを選んだ場合、そのアイテムは、3個以上のアイテムで、そのアトリビュートの対象集団を代表するように選ぶ必要があります。(Cf. 連環データ分析への招待 連環データ分析ベシックの応用では、アルコール飲料企業をサブジェクトとし、アトリビュートはアルコールの種類をバランスよく数種類取り上げております。この場合は、アルコールの種類によって、数量の単位が異なっておりますが、マップは破綻せず、業界の特徴が良くでております)

しかし、計量値を、適当な区間で区切り、サブジェクトが、区間で示されたアイテムの何処に種属するかで、連環度を[0,1]として入れて、クロス表を構成することもできます。これは、アトリビュートが線形の効果をもたらすとは限らないとき、有効な情報をもたらしてくれる可能性があります。

Q3. サブジェクトとアトリビュートの入れ替えは？

A3. サブジェクトとアトリビュートは、対称的に計算されて居りますので、入れ替えても構いません。

ただ、入力データとして、エクセルを使う関係で、サブジェクトとのアイテム数が多く取れるように、表側側を使っております。エクセル 2007 では、表頭側も大きく使えるようになりましたので、どちらを使っても結構です。

Q4. 個票データとクロス集計表データのどちらがよいか？

A4. 大勢のパネルをサンプルしたアンケート調査や、POSなどのトランザクションデータでは、アイテム数がかなり多数になります。もしこれらがすべて[0,1]のようなデータタイプであれば、これらを全てのアトリビュートの“クロス集計表データ”（パート表）でも本来の情報量は、変わりません。しかし、連環データ分析に掛けると、両者は、若干微妙に変化します。それは、連環データ分析が、元のマトリクスの次元を圧縮して近似していることに起因します。

どちらが良いかは、分かりません。ただし、[0,1]以外のアトリビュートがあり、それが、KPI等の計量値の場合は、個票をサブジェクトにすることをお奨めします。それは、スタンダード版以上の連環データ分析では、KPIの分解表示機能を全てに搭載する予定でその応用が可能になるからです。また、クロス集計表データは、サンプルや個性を無視しているからです。

もちろん、クロス集計した切り口が、共通のサブジェクトとして、それまで統合解析ができなかったデータを統合できるなら、それは別なメリットが発揮されることになります。

言葉の定義

Q1. クロス表とは？

A1. 通常、エクセルなどで使われている、行と列とで区切られ、そのクロスしたセルに数字が入れた表のことです。

行は、表側（または表足）、列は表頭とも呼ばれ、クロスしたセルの部分を表体と呼ぶこともあります。

データは、何か他のものと比較して始めて意味が判りやすくなるので、表形式のデータは、ビジネスでは、非常に便利に使われております。

Q2. サブジェクトとは？

A2. サブジェクトとは、クロス表の表側に置かれるアイテムの集合のことです。解析、関心の対象としてのアイテムの集合を、サブジェクトと呼ぶことにしました。連環データ分析では、関心の対象を、通常法則側に置き、それを修飾する属性や、それを喩えたり関連するアイテムとの関係性を分析します。

サブジェクトは、関心のテーマの集合を代表するアイテムから構成されている必要があります。

Q3. アトリビュートとは

A3. アトリビュートは、通常表頭に置かれるアイテムの集合のことです。

解析、関心の対象としてのサブジェクトを飾り、または喩え、または関係のあるアイテムの集合として、アトリビュートと呼ばれることから、踏襲しております。

アトリビュートも、サブジェクトに対応した集合を代表するアイテムから構成されている必要があります。

Q4. アイテムとは？

A4. アイテムは、サブジェクトやアトリビュートの構成要素です。通常は、クロス表の表側や表頭に付けられた言葉で表現されており、連環マップ上に、その言葉が布置されます。

どのサブジェクトのどのアイテムと、どのアトリビュートのどのアイテムとの関係が大きいか、近いかが主要な課題になります。

連環データ分析は、最低でもサブジェクトとアトリビュートの2種類のアイテムの集合の関係を検討しつつことで、意味のある情報を抽出することです。従って、アイテムは、どちらの集合の出身母体（母集団）を正しく代表している不必要があります。

Q5. クロスピークサーチとは？

A5. サブジェクトのあるアイテムとアトリビュートのあるアイテムの組合せ効果が大きな時、つまり連環度が高いとき、人は、“サブジェクトでそのアイテムと言えば、アトリビュートでは、そのアイテムを想起し、または連想”します。逆にアトリビュートのあるアイテムがとサブジェクトのそのアイテムが連想されたとき、連環性が最も強くなります。こうした連環度のピークは、サブジェクトとアトリビュートのクロス値のピークを探す問題でもあります。これは、サブジェクトとの集団とアトリビュートの集団との組合せデータ全体の特徴を捉え、理解するための重要な手順です。

具体的には、あるサブジェクトのアイテムをみて、その中で最も大きな連環度となるアトリビュートのアイテムを探し、そのアイテムとの組合せで最も大きな連環度となるサブジェクトのアイテムを探し、それが落ち着くところまで繰り返します。最初に見つかったピークが最高峰で、いわばK1です。同様な手順でK2., K3と次々に探して行きます。そして、サブジェクトかアトリビュートの全てで、ピークが1つづつ決まったとき、そのクロス表は、連環性が非常に高いということができます。

連環データ分析では、連環マップ上で、こうしたクロスピークを見つけ易く、原点から離れて、同じ方向に、同時布置をします。

Q6. ベキ乗分布とは？

A6. ハード技術や工業の領域では、目標に向かって、多くの要素を標準化し、コントロールを行うので、多くの誤差が累積して、特性値が正規分布をする場合が、良く見かけられます。しかし、社会現象では、ある主体が、限られた原料資源空間を占有してそこから得た資源を使って原料資源を経ち取るパワー資源を獲得することで、強いものがますます強くなり、非対称な特性分布を持つことがむしろ自然です。例えば、お金持ちはますます富みを蓄積し、大きな都市がますます大きくなってきたといような例です。

また、細かい存在物の寿命が短いところに多く指数分布に従っており、それらの特性の積の形で決まる特性でも指数分布が現れることが知られております。道路を走る自動車の間隔の分布などもそうです。また、自然現象では、小さな変動が多く発生し、大きな変動はまれに発生することが知られて居りますが、人工物でも、全体の仕組の中で、最も弱いリンクが破損して現れる分布は、“極値分布”と呼ばれます。

こうした分布では、小さな値と大きな値が、混在しますが、こうした環境情報に対し、人は自動的に尺度調整をして、認知し、区別することが知られております。

連環データ分析では、社会現象を扱い、また、アトリビュートが男女や年齢や、その問いにイエスかノーかなど、離散的で0か1等の小さな値を、統合して扱うため、

前処理として、“平準化”機能をオプションとして持って居ります。

商品形態・契約条件

Q1. なぜレンタルか？

A1. 連環データ分析は、まだ生まれ立てで、理論的根拠もまだ十分明確になって居りません。また応用領域や応用法も開発途上です。従って、多くの実際の問題で、その有効性を探索しつつ、成長させて頂きたい、いわばオープンイノベーション向きのテーマといえます。お陰様で、我々は、良き指導者に恵まれ、理論よりも、実証的研究を優先させて来ております。また良きパートナーに恵まれ、日々刻々と進化し続けており、常に最新のソフトのバージョンを提供する仕組として、レンタルモデルを採用させて頂いております。

丁度、SaaSという考え方や、クラウドコンピューティングの時代になり、ご理解を頂けると幸いです。

Q2. なぜ、エクセルを入出力として使うのですか？

A2. パソコンの普及発展には、文章を処理するワードプロセッサと、クロス表集計計算をするスプレッドシートが両輪で、支えてきました。現在は、その後者の主流が、マイクロソフトのエクセルであり、多くの方が、使っておられますので、入力と、結果の座標計算のアウトプットに、エクセルを使っております。

企業では、非常に多くのクロス表が使われて居りますが、いわば、“パーソナル・ビジネス・インテリジェンス”が、組織のパフォーマンスを決める時代ではないでしょうか？ 連環データ分析が、そうしたデータリテラシーを上げるのにお役に立てればと思っております。

Q3. なぜ、ダウンロードして使うのか？

A3. クラウドコンピューティングが注目されておりますが、まだ日本では、ブロードバンドの回線が安く自由に使えるわけではありません（ただ2000年以降、日本でも香港や韓国に近くなったことは特筆されますが）。また、パソコンのパワーの向上が著しく、かなり大きなデータでも扱えるようになって、料金を気にせず使えることもあります。

Q4. 期限切れのお知らせは何時？

A4. 期限切れを5週間前にお知らせさせて頂きます。それによって、引き続きご利用いただけますよう、早めの手続きなどのご準備をお願いします。引き続きご利用の場合は、期限が切れる2ヶ月以前から受付ております。有償版の有効期間は、お申込頂き、お振込日から、その翌年の前日までの通常年は、365日分とさせて頂いて居ります。継続前の有効期間が終了してから、継続される新しいご契約期間に引き継がれます。

Q5. 仮キーの有効期間は？

A5. 仮キーの有効期間は、約4週間です。

ソフトをダウンロードして頂く関係で、お持ちのPCに無事ダウンロードでき、稼

働できることをご確認いただきましてから、料金のお振込をお願いしております。それまでは、仮キーでのご利用をお願いしておりますが、その期間のご利用に関しては、料金の発生はありません。

この間、弊社では、受付手続きや、お振込みの確認、本キーの発行などの手続きや、お客様でも、稼働のご確認、社内手続きなどのご準備に時間が取られるかと思われまます。早めのご準備やお手続きをお願いします。

Q6. 期間限定の無償版の期間は？

A6. 現在、連環データ分析ベイシックトライアルは、2011年3月31日までとさせて頂いております。上位版に比べて、機能はかなり絞りはありますが、それでも非常に豊富な機能を持っております。ぜひ、実際の問題で、使い方を工夫して頂き、有効性を確かめて頂きたいと思っております。

その先の、こうした試みの延長につきましては、未定です。

Q7. インターネットに接続できない環境では使えませんか？

A7. 常にソフトの最新バージョンをお使い頂く形をとらせて頂いておりますので、ソフトの立上げ時に、自動的に最新版をダウンロードして頂くことになっております。また、同時に、ご本人確認をさせて頂いております。

プレゼンテーションなどをされる場合は、ご不便をお掛けしますが、ご了解を頂ければ幸いです。

推奨システム環境

Q1. PC本体は？

A1. 現在は、PC/AT 互換機のみです。

Q2. オペレーティングシステム：OSは？

A2. Microsoft Windows XP Home Edition 以上

Q3. CPU速度は？

A3. 処理するデータ量が多い場合は、速いほど良いのですが、2.0GHz以上を推奨しております。

Q4. 主記憶装置（RAM）の容量は？

A4. 処理するデータ量によりますが、2.0GB以上を推奨します。

Q5. ビデオメモリー：VRAMの容量は？

A5. 表示するアイテム数によりますが、3D表示をさくさくと実行するために、128MBとやや大きめの容量をお願いしております。

Q6. 補助記憶装置の容量は？

A6. データ量にも寄りますが、ハードディスクで、空き容量として 2.0GB以上をお願いいたします。

Q7. Microsoft Excel のバージョンは？

A7. Excel は、2007 もサポートしております。エクセルが使える状態であることが必要です。

Q8. Java のバージョンは？

A8. Java Runtime Environment 6(JRE 6)以上がインストール済みであることが必要です。

プログラムのダウンロード環境条件

・・・スムーズにダウンロードするために

Q1. ダウンロードできないが？

A1.1. 送りしましたユーザID、およびライセンスキーの前後にスペース等が入り込んでいないかを調べてください。

A1.2. Java 6 以降のバージョンがインストールされているかを調べてください。

A1.3. ブラウザーの設定で、「ポップアップの禁止」を解除していなるか確認してください。解除しないとポップアップアイコンがデスクトップ上に布置されません。

Google 等のブラウザーをブラウザーのアドインとして、ツールバーとして導入している場合も同様です。

以上