

ファジィ集合質的比較分析 (fsQCA)

東京大学法学部助教
森 大輔

<http://park18.wakwak.com/~mdai/qca/>

2010/5/9

1

質的比較分析(QCA)

- Qualitative Comparative Method
- csQCA(クリスプ集合QCA)
 - 1987年、Charles Ragin(チャールズ・レイガン)が提唱
- 特徴
 - ① 小規模なデータから共通法則を見つけ出せる
 - ② 必要条件・十分条件の関係を分析できる
 - ③ 複数の条件の組み合わせを分析できる

2

QCAのキーワード

- 必要条件・十分条件
- 結合因果・多元因果
- ブール代数
- 真理表
 - 矛盾
 - 論理残余
- 論理式
 - 簡単化
 - 複雑解・簡略解・中間解

3

質的分析と量的分析

- 質的分析
 - ケーススタディー(事例研究)
 - それぞれの事例の特殊事情がよく分かる
 - 一般的な状況に適用できる法則が見えにくい
 - 事例を増やすことは困難
- 量的分析
 - 統計
 - 多くのデータから共通する法則を見つけ出す
 - それぞれの事例の特殊事情を捨象
 - 小規模なデータは分析が困難

4

相関関係

- 線形の関係
- 多い方がよいもの
 - 原因あり・結果あり
 - 原因なし・結果なし
- 少ない方がよいもの
 - 原因あり・結果なし
 - 原因なし・結果あり

	結果あり	結果なし
原因あり	22	8
結果なし	5	15

	結果あり	結果なし
原因あり	○	×
結果なし	×	○

5

必要条件

- 結果が起こるにはその原因が必要
- 多い方がよいもの
 - 原因あり・結果あり
- 少ない方がよいもの
 - 原因あり・結果なし
- 無関係
 - その他

	結果あり	結果なし
原因あり	16	14
結果なし	0	20

	結果あり	結果なし
原因あり	○	-
結果なし	×	-

6

十分条件

- 結果が起こるにはその原因があれば十分
- 多い方がよいもの
 - 原因あり・結果あり
- 少ない方がよいもの
 - 原因あり・結果なし
- 無関係
 - その他

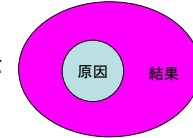
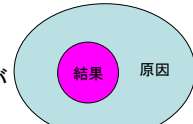
	結果あり	結果なし
原因あり	30	0
結果なし	12	18

	結果あり	結果なし
原因あり	○	×
結果なし	-	-

7

ベン図

- 必要条件
 - 結果は原因の部分集合
 - 当該原因があっても結果が生じないことがある
- 十分条件
 - 原因は結果の部分集合
 - 他にも結果を生じる原因がある



8

必要条件・十分条件

- 社会科学の命題は、必要条件か十分条件関係を含むことが実は多い
- 必要条件の例
 - 民主主義国であるためには選挙が行われなければならない
- 十分条件の例
 - 民主主義国同士は戦争をしない (民主的平和論)

9

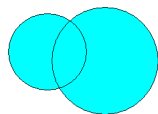
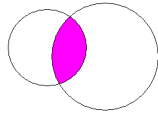
条件の組み合わせ

- 原因条件が複数ある場合
 - 原因 A, B, C, D, E 結果 Y
 - 多重クロス表
 - 4重クロス表以上は理解が困難
 - 対数線形(ログリニア)モデル
 - 4変数以上は分析が困難
- 交互作用
 - 条件付きの関係
 - e.g. 条件ZがあるときXとYの増加は関係するが、条件Zがないときは関係しない
 - 多数の変数の交互作用を統計分析で扱うことは困難

10

結合因果・多元因果

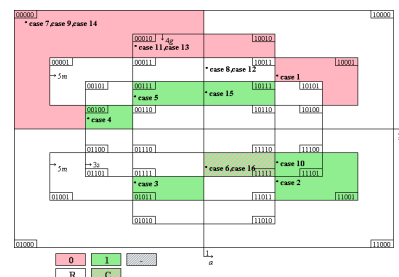
- 結合因果
 - 条件が結びついて初めて結果をもたらす
 - AかつBのとき結果Y
 - 積集合
- 多元因果
 - 異なるいくつかの原因が同じ結果をもたらす
 - AまたはBのとき結果Y
 - 和集合



11

原因条件の数を多くした場合

- ベン図:とても複雑になり理解が困難



12

ブール代数

- 集合を機械的に計算するための方法
- 論理式
 - 例えば、原因条件をAとB、結果をYとする
 - $A \leq Y$ のような式
- \leq は「ならば」を表す
 - $A \leq Y$ は「AならばY」
 - 同時に集合の大きさの関係も表している
- 2値変数を使用
 - Aが存在している場合: Aの値は1
 - Aが欠如している場合: Aの値は0
- 否定
 - 否定は小文字で表すことにする
 - aは「Aが存在していない場合」

13

論理式の計算

- 論理積
 - 「かつ」は*で表す 例: $A*B$ は「AかつB」
- 論理和
 - 「または」は+で表す 例: $A+B$ は「AまたはB」
- 簡単化定理
 - 条件が存在していても欠如していても同じ結果をもたらすならば、その条件は不要
 - $A*B + A*b \leq Y$
 - $A*(B + b) \leq Y$ (分配法則)
 - $A \leq Y$ (簡単化定理)
 - 論理式を簡単にしていくのが主要な目標

14

例: 戦間期のヨーロッパの民主主義

- 付表の表1
- Rihoux & Ragin (2008)の例
- 戦間期のヨーロッパ諸国
 - 民主主義が継続した国 例: イギリス、フランス
 - 民主主義が崩壊した国 例: ドイツ、スペイン
 - 比較政治学者Lipsetの主張「裕福な国ほど民主主義は生き延びやすい」
 - ドイツなどは裕福だが、他の事情で崩壊した

15

真理表

- まず表1の各種条件を2値化(表2)
 - 閾値
 - 各事例の知識やデータの分布をもとに設定
- 条件の組み合わせごとにまとめる(表3)
 - 真理表
- 矛盾する組み合わせ
 - 0と1の両方の結果がある条件の組み合わせ

16

矛盾の解決方法

1. 条件を増やす(表4)
 - 分類しきれないのは情報が足りないから
 2. 閾値の値を変える(表4)
 - 機会主義的に行わないように
 3. 結果の定義を見直す
 4. 結果の値を0とみなす
- 表5: 条件の追加・閾値の変更後の真理表

17

論理残余(logical remainder)

- 観察されなかった条件の組み合わせ
- 条件の組み合わせは、論理的には 2^5 通り
表5には9通りしかない
残りの23通りはデータにない
- これらの論理残余をどう扱うか?
 1. 論理式の計算に、論理残余を含めない(複雑解)
 2. 簡単化に役立つ論理残余は全て含める(簡略解)
 - 結果を予測して、結果についての仮定を置く
 - 例: GNP*CAP*URBANIZA*literacy*indlab*STABというデータはないが、結果SURVIVALが1になると仮定する
 3. 妥当な論理残余のみ含める(中間解)

18

論理式の導出

- 複雑解
GNPCAP*LITERACY*STAB*(INDLAB+urbaniza) ≤ SURVIVAL
(BEL,CZE,NET,UK+FRA,SWE)+(FIN,IRE+FRA,SWE)
- 簡略解
GNPCAP*STAB ≤ SURVIVAL
(BEL,CZE,NET,UK+FIN,IRE+FRA,SWE)
- 中間解
GNPCAP*STAB*LITERACY ≤ SURVIVAL
(BEL,CZE,NET,UK+FIN,IRE+FRA,SWE)

19

単純化の仮定

- 使用した論理残余のこと
- 簡略解
 - GNPCAP*urbaniza*literacy*indlab*STAB
 - GNPCAP*urbaniza*literacy*INDLAB*STAB
 - GNPCAP*URBANIZA*literacy*indlab*STAB
 - GNPCAP*URBANIZA*literacy*INDLAB*STAB
 - GNPCAP*URBANIZA*LITERACY*indlab*STAB
- 中間解
 - 5つの条件は全て、値が1のときに結果1に影響
 - 解の複雑さが、複雑解と簡略解の中間
 - 詳しい導出方法はRagin(2008)を参照

20

論理式について補足

- 3つの解を比較し、最も理に適ったものを選ぶ
 - ただし、3つとも提示はすること
- ここでは、結果が1になる条件のみ求めた
 - 実際は結果が0になる条件も求めるべき
- 計算にはソフトウェアを使う
 - Tosmana, fs/QCA, R, Stataなど
 - 詳しくは、
<http://park18.wakwak.com/~mdai/qca/software.html>

21

ファジィ集合質的比較分析(fsQCA)

- 2000年、Charles Ragin(チャールズ・レイガン)が提唱
- csQCA(クリस्प集合質的比較分析)の弱点
 - データの2値化が必要
- fsQCAはこの弱点を解消
 - ファジィ集合を使用

22

fsQCAのキーワード

- ファジィ集合・クリस्प集合
- キャリブレーション
- 必要条件・十分条件
- 整合度
- 真理表
 - 度数の閾値・整合度の閾値
- 被覆度
 - 解被覆度・粗被覆度・固有被覆度

23

ファジィ集合

- 集合に属する度合いを0から1の間の値で表す
 - ⇔クリस्प集合(通常の2値の集合のこと)
- 1965年、Lotfi Zadeh(ロトフィ・ザデー)提唱
- メンバーシップ関数
 - 集合に属する度合い(メンバーシップ度)を表す関数

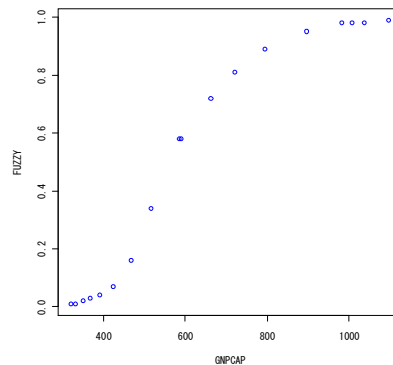
24

キャリブレーション(calibration)

- データをメンバーシップ度に変換すること(表6)
- 1つの方法(直接法)
 1. 以下の3つの点と元のデータの対応を決める
 - ①完全帰属(メンバーシップ度0.95)
 - ②中間点(メンバーシップ度0.5)
 - ③非帰属(メンバーシップ度0.05)
 2. 3つの点を手がかりにロジット変換を行う
S字型のロジスティック曲線
完全帰属や非帰属の中でのデータの差はそれほど問題ではないから
例: 先進国・途上国の分類で、米国と日本の差はそれほど問題ではない
※ fs/QCAのcalibrate関数で容易に変換できる

25

キャリブレーション



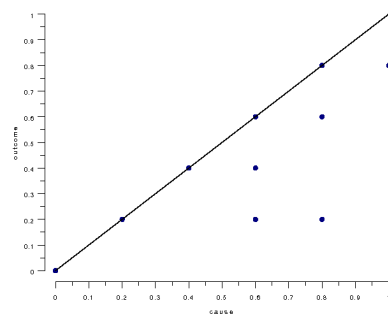
26

必要条件・十分条件

- ファジィ集合での部分集合
 - Xを原因、Yを結果とする
 - 集合Xの各要素を X_i 、Yの各要素を Y_i とする(iは各事例)
 - $\forall i, X_i \leq Y_i$ のとき、XはYの部分集合
- YがXの部分集合のとき、XはYの必要条件
 - $Y_i \leq X_i$
 - 散布図の右下三角形
- XがYの部分集合のとき、XはYの十分条件
 - $X_i \leq Y_i$
 - 散布図の左上三角形

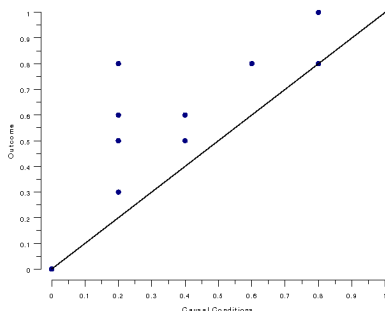
27

必要条件



28

十分条件



29

論理積・論理和・否定

- 集合Aの各要素を A_i 、Bの各要素を B_i とする(A、Bは原因条件、iは各事例)
- 論理積
 - AかつBの集合 $A*B$
 - $(A*B)_i = \min(A_i, B_i)$
 - 表7を参照
- 論理和
 - AまたはBの集合 $A+B$
 - $(A+B)_i = \max(A_i, B_i)$
- 否定
 - $a=1-A$

30

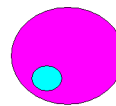
fsQCAの真理表

- 元のデータをキャリブレーションする(表6)
- 条件の組み合わせごとに整合度を求める(表8)
 - 全ての事例の全ての条件の組み合わせのメンバーシップ度が、整合度の計算に使われる
- 度数の閾値で論理残余を決める(表9)
 - 度数とは、0.5より大きいメンバーシップ度を持つ事例の数
 - 一定以上の度数のない組み合わせは論理残余にする
 - 通常は閾値は1でよい(度数0のものが論理残余)
- 整合度の閾値で結果を決める(表9)
 - 一定以上の整合度があれば結果1
 - 閾値は高めに設定する(表9では0.8)
 - csQCAと同じ結果にするには、もっと低い閾値である必要

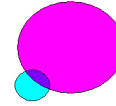
31

整合度(consistency)

- 一方の集合が他方の集合の部分集合である度合い
- 当該関係と整合する事例がどれぐらいあり、矛盾する事例がどれぐらいあるかを示す



整合度1



整合度小

32

整合度の計算方法

- 最も素朴な計算方法
 - 散布図の三角形に入る事例の割合を計算
 - 十分条件の場合は左上の三角形に入る割合
- 精密化した方法(推奨方法)
 - 45度線から遠く離れた不整合事例に大きなウェイトを置く
 - 十分条件の場合

$$\text{Consistency}(X_i \leq Y_i) = \frac{\sum(\min(X_i, Y_i))}{\sum(X_i)}$$
 - 計算例: 表8
- さらに精密化した方法
 - PRI整合度(詳細は省略)

33

論理式の導出

- 真理表まで作れば後はcsQCAとほぼ同じ
- ここでは、結果が1の場合の中間解を求める
- 中間解

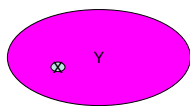
$$\text{GNPCAP} * \text{LITERACY} * (\text{URBANIZA} + \text{indlab}) \leq \text{SURVIVAL}$$

	粗被覆度	固有被覆度	整合度
GNPCAP*LITERACY*indlab	0.284038	0.191315	0.814815
GNPCAP*LITERACY*URBANIZA	0.470657	0.377934	0.813387
解被覆度	0.661972		
解整合度	0.806867		

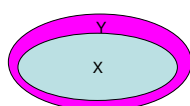
34

被覆度(coverage)

- 一方の集合が他方の集合の中に占める割合
- 原因条件の組み合わせXが結果Yを説明する度合い



被覆度小



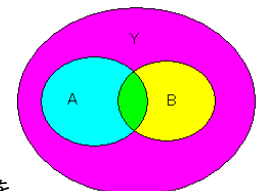
被覆度大

- $$\text{Coverage}(X_i \leq Y_i) = \frac{\sum(\min(X_i, Y_i))}{\sum(Y_i)}$$

35

解被覆度

- 解被覆度
 - 論理式が全体として結果を説明する割合
 - 例: $(A+B)/Y$
- 粗被覆度
 - 論理式の各項が結果を説明する割合
 - 例: A/Y 及び B/Y
- 固有被覆度
 - 論理式の各項が単独で結果を説明する割合
 - 例: 水色/Y 及び 黄色/Y



36

計量法学(ジュリメトリックス)

- 1960年代、アメリカ
 - Baade(1963)を参照
- 判決予測
 - 例: 弁護人依頼権を奪われたか
 - 11. 上告人は起訴認否手続きの時期に弁護人の援助を受けられなかった
 - 12. 上告人は起訴認否手続きから後半までの期間または起訴認否手続きから有罪の答弁までの期間に弁護人の援助を得られなかった
 - 19. 上告人が弁護人依頼剣を明示的に放棄したことはなかった
- この後、あまり用いられた形跡はない
 - 多変量解析に取って代わられた
 - 統計的手法との差異の無自覚
 - 原因条件を多く取り込みすぎ分析が複雑化
- しかし、QCAの手法は判例分析となじみやすいと思われる
 - 真理表の矛盾の解決と判例分析におけるdistinguishの類似性

37

QCAの欠点

- 全ての事例を同じ重みで扱う
 - 決定論 vs 確率論
 - 推測統計では少数の逸脱は誤差として扱う
 - QCAでは矛盾はなるべく解消を図る
- 各変数の影響の大きさが分からない
 - 回帰分析では各変数に係数がある
- 分析者の選択により結果が大きく左右される
 - 各種の閾値の選択
 - キャリブレーションの3つの点の選択
- * 旧来からの批判の幾つかはfsQCAで解消されている(2値化、影響の数値化etc)

38

QCAをいませんか？

- 各事例の詳細な知識のあるケーススタディーにこそQCAが生かせる
 - 法社会学と親和性
- 判例分析に使われていた
 - 法解釈学と親和性
- QCAの用途
 - データの要約
 - 探索的分析
 - 仮説の検証
 - 仮説の構築
 - 思考実験
 - 分類

39

参考文献

- 鹿又伸夫ほか(2001)『質的比較分析』
- Baade H.W.(1963) Jurimetrics (早川武夫・碧海純一編訳『ジュリメトリックス』, 1969)
- Cronqvist L. (2007) Tosmana User Manual (森大輔訳「Tosmanaユーザーマニュアル」<http://park18.wakwak.com/~mdai/qca/software/Tosmana/Japanese.pdf>)
- Lipset S.M.(1959) Political Man (内山秀夫訳『政治のなかの人間』, 1963)
- Ragin, C.C. (2008) Redesigning Social Inquiry: Fuzzy Sets and Beyond
- Rihoux, B. & C.C. Ragin (2008) Configurational Comparative Methods

40