

"Model and Methods in Social Network Analysis" 6章
Network Models and Methods for Studying
the Diffusion of Innovations

(ネットワークモデルと
イノベーション普及の解法)

東京理科大学大学院
理工学研究科 情報科学専攻
M2 古川 忠延

第8回 社会ネットワーク分析 輪読会 2005 @ 05/07/27

概要

- ネットワーク上におけるイノベーション (技術革新:新しい物事)の普及が,
 - ・どのような要因に基づいているのかそして
 - ・どのようなモデルとして考えればよいかを分析していく.

構成

- 1. 導入:いろいろなモデルの紹介
 - (A)マクロモデル
 - (B)空間的自己相関
 - (C)ネットワークモデル
- 2. 実証研究
 - (A)イベント履歴分析
- 3. ネットワークに基づく干渉
- 4. まとめ

6.1 イノベーション普及学

- 新しいアイデアや慣習がコミュニティ内もしくはコミュニティ間でどのように広まっていくか
- もともとは人類学・経済学・地理・社会学・マーケティングの分野の話

(A)マクロモデル

- 累積的な普及パターンは1変数のロジスティック関数で近似できる

$$y_t = b_0 + \frac{1}{1 + e^{-b_1 t}}$$

- y: 採用比率
- b₀: y切片
- t: 時間
- b₁: 比率変数

- 但し, 適用範囲が限られる Bassモデル

Bassモデル(1)

$$y_t = b_0 + (b_1 - b_0)Y_{t-1} - b_1 Y_{t-1}^2$$

- y: 採用比率
- b₀: イノベーションの比率変数
- b₁: 模倣の比率変数(今回だけの採用度合い)
- 各時間での採用者の割合を合計

- (1)普及の予想水準の予測
- (2)別の視点からの普及
 - b₀: 外的要因
 - b₁: 内的要因

Bassモデル(2)

- ある中心となる発信源から広まっていく場合に適用
 - 汚染された食物・個人の接触による感染などによる病気の伝染
 - メディア広告による普及

7

マクロモデル - まとめ

- マクロレベルでの研究
 - 例えば国を一つのネットワークとするなら、得られる分析結果は、国によって要因が異なっているということを示す？
 - 但し、扱っているネットワーク内全体で、ある一つの同じ性質をもっている、と仮定してしまっている。(ネットワーク内でも、人によって異なる行動を起こすかもしれないことを考えていない)

8

(B) 空間的自己相関

- イノベーション(新しい製品・風習など)が、隣接するエリアに広まっていくかどうか?を測る
- 近接データを利用し、距離を考慮したネットワークを考える

9

Moran's I

- 採用の地理的なクラスタリング

$$I = \frac{N \sum_i \sum_j D_{ij} (y_i - \bar{y})(y_j - \bar{y})}{S \sum_i (y_i - \bar{y})^2}$$

- N: サンプル数
- D: 距離行列
- y: 採用
- S: D中の距離の合計値
- 平均した行動から逸れた行動をしているノードの割合
 - 平均スコアより大きくズレていると高い値

10

空間自己相関 - まとめ

- ネットワーク自己相関がどの程度、イノベーション普及に適用できるかは不明確
空間自己相関は普及をマクロレベルで測っている
- ネットワーク構造が普及に影響を与えているかは分からない
ネットワークモデルを改めて考える

11

(C) ネットワークモデル

- 図6.1.1
カメルーンのボランティア組織における調査: 組織内の友人関係は、避妊薬の選択と関係があるか

12

Exposure (1)

- ネットワークの影響は, exposureによって測ることができる.
- 個人が採用する見込みは, 近接するノードが増えることによって上昇する.

$$E_i = \frac{\sum w_{ij} y_j}{\sum w_i}$$

- w: 重み行列
- y: 採用ベクトル

13

Exposure (2)

- exposureは社会ネットワークの特徴を表すWを用いて計算できる.
Wは隣接ノードに対してだけでなく, さまざまな形で表すことができる
 - ex. 人間関係の構造同値をあらわしたもの
exposure: 同値の他人との競争による伝播
- exposureはネットワーク性質 (centrality: オピニオンリーダーからの影響力, など) によっても重み付けできる

14

図6.1.3について

15

6.2 実証研究

- 多くの研究は断面的なもので, 回顧的にある一時のデータを使用する.
- カメルーンの避妊薬の例
 - 初期採用者の選択は明らかに後期採用者の選択に影響している
 - 友人の行動を知ること, 特に友人が使用を勧めた薬は, 非常に行動に影響している

16

モデル

$$\log \frac{\Pr(y_i = 1)}{(1 - \Pr(y_i = 1))} = \alpha + \sum \beta_k X_k + \beta_{k+1} \omega y_i$$

- y: 採用・不採用を表す(0,1)のベクトル
- : 切片
- k: K個の社会人口学的特性(年齢・財産など)に対する係数(重み)
- : 社会ネットワーク行列
- y_t: 同時発生のネットワークexposure
- k+1が大きい: 伝染力が強い(exposureが採用と関係あることを示す)

17

クラスタリングの修正

- 推定値の分散はバイアスがかかってしまう
 - 観測者が独立でないから
- 対策として, クラスタリングを修正する
 - 例えば, 別々の組織から抽出した2人より, 同じ組織内から抽出した2人のほうが行動が似ている可能性が高い
 - 結果: 表6.2.3
 - 修正前はexposureは採用と強い関係があったが, 修正後はわずかにしかない
- クラスタリング制御は重要である

18

社会的影響

- クラスタリング制御を行っても、社会的影響は行動に対して影響力が強い。
 - 自分の周りの大多数がタバコを吸っている場合、自分自身も喫煙者である可能性は2倍
 - 親友が喫煙者であればさらに2倍

19

信頼性の向上(1)

- 繋がり信頼性を上げたい
 - アンケート形式では、ミスや偽造ができてしまう
- 社会的影響を調査するのであれば、最低でも時間上の2点を取った長期のデータを使うべきである。

20

信頼性の向上(2)

$$\log \frac{\Pr(y_i = 1)}{(1 - \Pr(y_i = 1))} = \alpha + \sum \beta_k X_k + \beta_{(k+1)} \omega y_i + \beta_{k+2} \omega_{t-1} y_{t-1}$$

- y: 採用・不採用のバイナリ
- : 切片
- k: 特性変数のための係数
- : 社会ネットワーク重み行列
- k+2: 高い正の値のとき、高い exposure の応答者は時点 2 のときに採用する可能性が高い
- k+1: 高い正の値のとき、exposure が変化すると行動も変化しやすい

21

イベント履歴分析(1)

- とにかく多くの時点でのデータを使って分析
- データは一点からの観測ではなく、場合と時間に応じた(各期間・各個人における状況別の)データとして扱う。
 - 100人について、7つの時点でのデータがある場合、700の状況について、exposureのための変数と採用・非採用のデータセットがある
- 最尤推定法を用いて、互いに独立の変数が、独立でない変数(=採用 / 非採用)と関係があるかどうかを決める。

22

イベント履歴分析(2)

$$\log \frac{\Pr(y_i = 1)}{(1 - \Pr(y_i = 1))} = \alpha + \sum \beta_j X_j + \sum \beta_{kt} X_{kt} + \sum \beta_{k+1} \omega y_i$$

- y: 採用・不採用のバイナリ
- : 切片
- j: 特性変数(X_j)のための係数
- kt: 時間変化する特性変数(X_{kt})のための係数
- : 社会ネットワーク重み行列
- t: 時間

23

イベント履歴分析 - まとめ

- exposureは必ずしも採用と関係がないかもしれない
 - 全ての人において当てはまらないかもしれない
 - 各個人における採用のための閾値を変化させているかもしれない

24