

周辺・境界的知識のイノベーション予測モデルの構築

森基金 成果報告書

2018/02/21

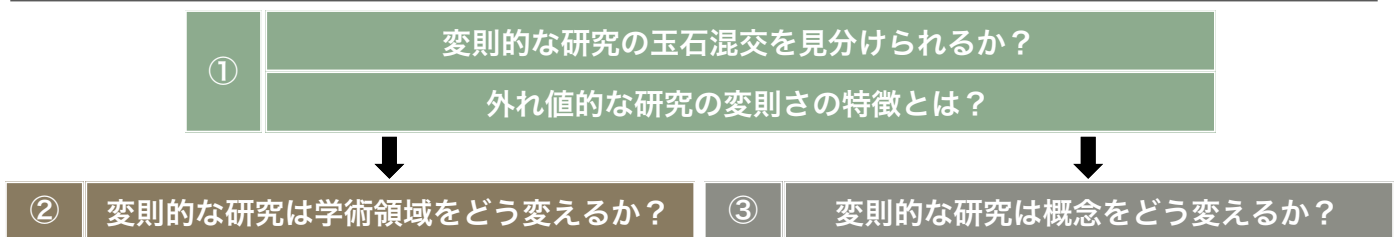
小野塚 亮

慶應義塾大学大学院政策・メディア研究科後期博士課程

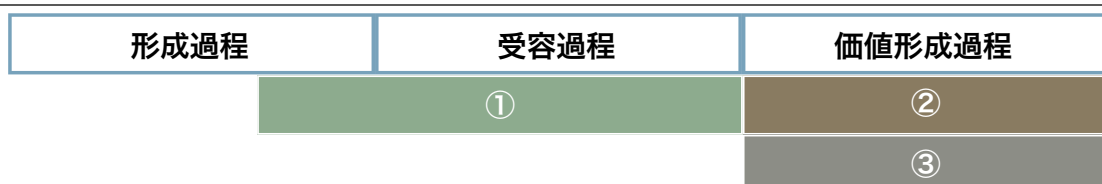
変則的な研究はどのように新しい知識を形成するのか？

変則的な研究の主流化について、経営戦略論の主流理論の変遷を事例に論じる。
機械学習による外れ値検知とネットワーク分析によって変則さを計測。

■ 現段階の構成



■ 知識のイノベーションのプロセス



研究①

変則的な研究の玉石混交は見分けられるのか？

リソース・ベースト・ビューを提唱するも長い間無視された過去を持つWernerfelt (1984)の潜在的な価値を1984年時点で予測する。論文の変則さを4側面から計測することで、10年後の被引用回数
の多寡を65.1%の精度で予測するモデルを構築した。さらに、外れ値的な論文に特徴的な変則さを抽出することで、Wernerfelt (1984)の潜在的な価値の見積もりに成功した。

INTRODUCTION(1) 社会的背景と学術的背景

社会的背景 社会の不確実性に対応した知識社会の発展のためには、将来の主流になりうるイノベ
ーティブな知識を玉石混交の中からいち早く見分ける方法が必要である。

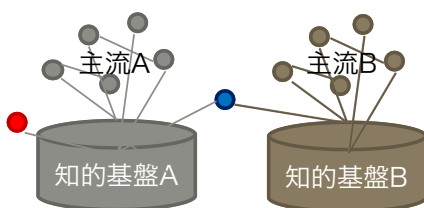
学術的背景 経営学は、組織の周辺・境界においてイノベーションが生まれやすいこと、
また、外れ値的な組織は他とは異なる生成メカニズムを持つことを主張。
しかし、実証的でなく外的妥当性が乏しいという問題がある。

計量書誌学は、論文の参考文献のつながりから知識のイノベーションを論じてきた。
しかし、論文の内容の分析を十分に行えていないという問題がある。

研究①

変則的な研究の玉石混交は見分けられるのか？

INTRODUCTION(2) 主要な概念の定義



■ 知的基盤 (intellectual base)

- 研究コミュニティの方向性を決める共有された参考文献群

■ 主流 (mainstream)

- 同じ知的基盤を共有する研究コミュニティが生産する知識

■ 変則性 (anomaly)

- 現在のパラダイムには当てはまらない科学の発展の起点となる知識 (Kuhn, 1970)

■ 周辺性 (peripherality) 上図の赤丸

- 同じ知的基盤上の議論であっても、主流の一般的な議論とは外れる議論を行っている知識

■ 境界性 (borderlineness) 上図の青丸

- 異なる知的基盤を組み合わせることで生産された知識の所属領域が曖昧になること

■ 分析対象

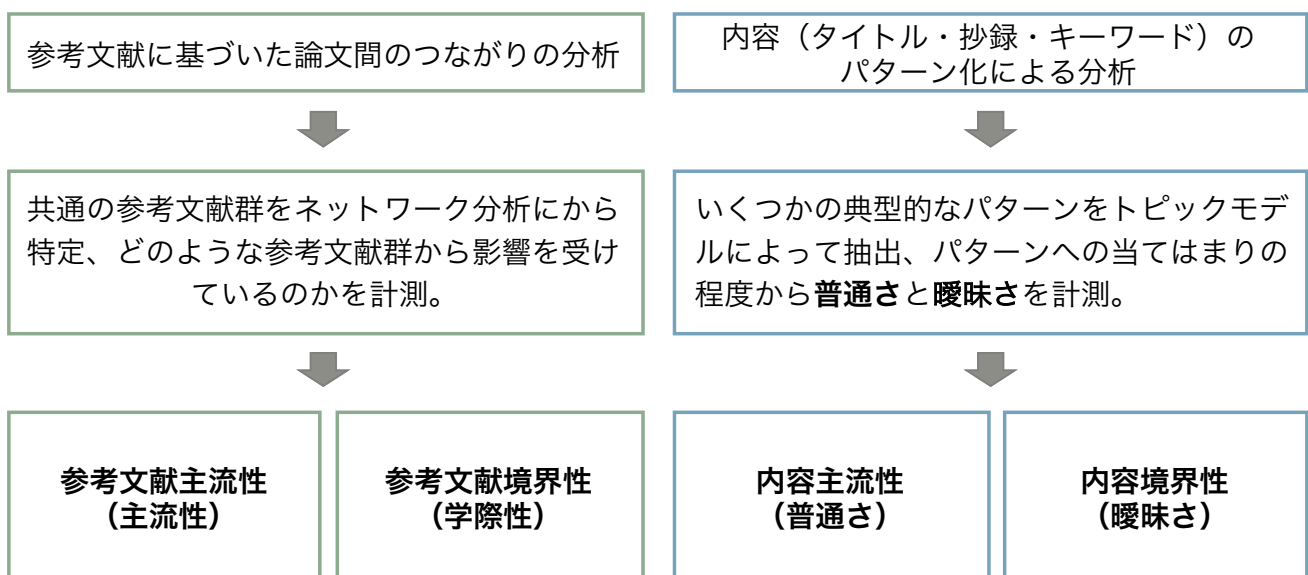
- 経営学の主流理論のひとつであるリソース・ベースド・ビュー(RBV)の主流化過程
RBVが提唱された1984年に経営学の上位ジャーナルに掲載された論文が対象

■ リサーチクエスチョン

1. 主流に受容されるかどうかに対して、変則性はどのように影響を与えるか？
2. 外れ値的な論文の変則性は他の知識とどのように異なるのか？
3. 見過ごされていたWernerfelt (1984)の価値を正しく見積もることはできるのか？

■ 仮説

1. 論文が主流的であり、境界的でないほど、主流に受容されやすい
2. 外れ値的な論文は周辺性・境界性が高い



■ 使用したデータ

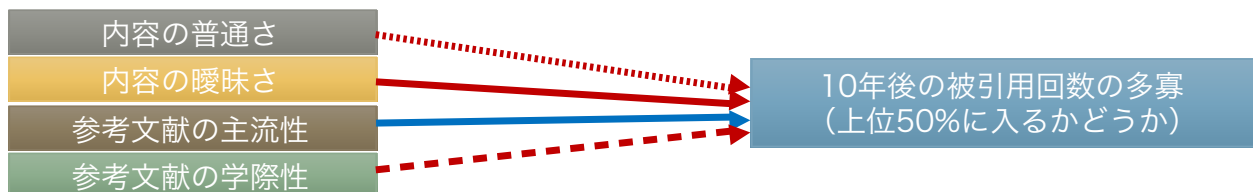
- Scopus®から取得した、1970年から1984年までにSNIP1.0以上の経営学上位ジャーナルに掲載された7,499本の論文からネットワークとモデルを作成した
- 1984年の論文962本に対して4つの指標を計算し、参考文献主流性・境界性が計算できなかった論文を除いた628本を分析に使用した
- この628本の論文の10年後（1994年）の被引用回数を目的変数とした

■ 3つの分析の方法

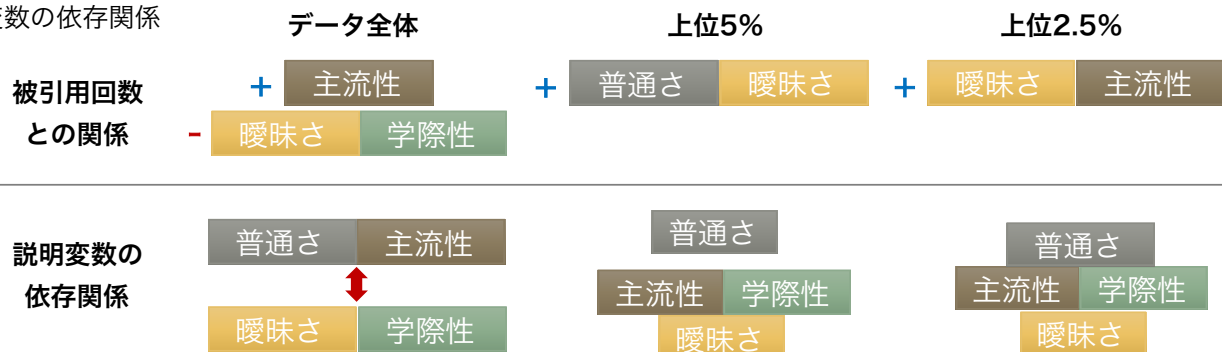
1. 被引用回数が少なくとも上位50%に入るかどうかを予測するモデルをベイズロジスティック回帰分析で作成後、各指標の影響度とモデルの評価を考察した
2. 被引用回数の高さによって変数間の依存関係がどのように変化するのかを考察した
3. 以上を踏まえてWernerfelt (1984)の予測確率を算出して考察した

7

- ベイズロジスティック回帰分析結果（点線：符号が反対になる可能性がある）。分類精度は65.1%。



■ 変数の依存関係

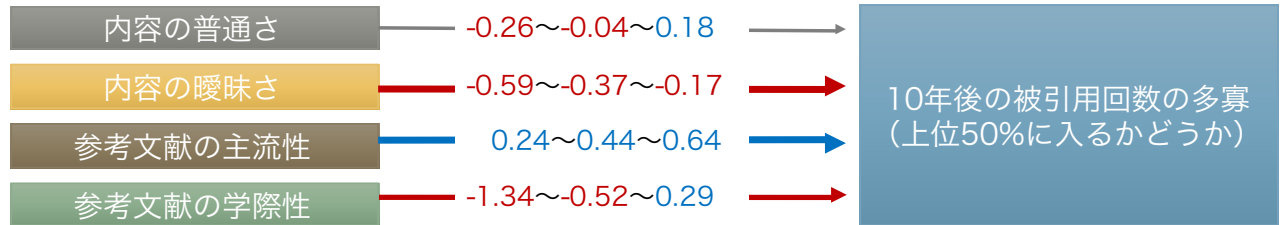


8

RESULT : ベイズロジスティック回帰分析結果

■ ベイズロジスティック回帰分析結果

ベイズ確信区間 (2.5%~50%~97.5%)



■ モデルの評価：データの8割から学習させ、残り2割を使ってモデルの評価をした

	実際に上位50%	実際に50%未満
上位50%と予測	20 ○	16 ×
50%未満だと予測	28 ×	62 ○

精度は65.1%
汎化性能を考えると妥当

9

Wernerfelt (1984)の潜在的な価値の評価

■ Wernerfelt(1984)の10年後の被引用回数が上位50%に入る確率は0.514

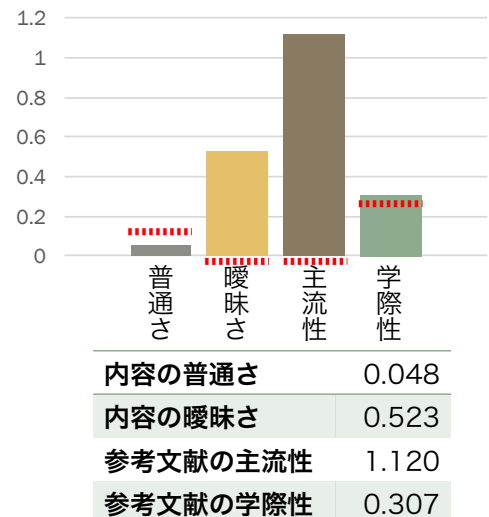
- 実際には80年代には全く引用されず、90年から94年までに44回引用された。

■ 予測確率の大きさに影響を与えた変数

- 正の影響：参考文献の主流性
- 負の影響：内容の曖昧さと参考文献の学際性

■ 変数の依存関係

- 参考文献の主流性・学際性と内容の曖昧さが共に高くなるのは、被引用回数の多い論文に見られるパターンである。
- データ全体の変数の依存関係とは似ていないことから、異なる生成メカニズムを持つと言える。



10

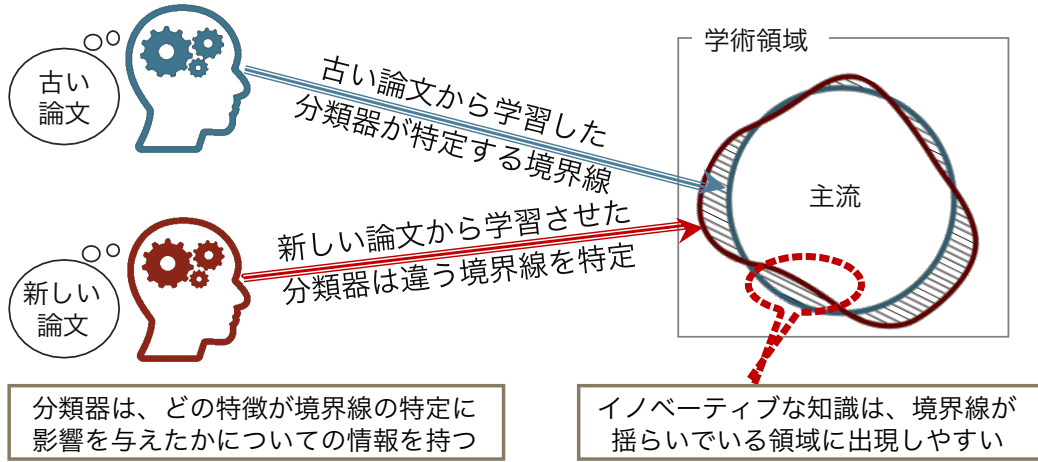
DISCUSSION & CONCLUSION

- **内容と参考文献の主流性と学際性から将来の被引用回数を予測することができる**
 - 全体としては、参考文献の主流性が正の影響を持つ一方、参考文献の学際性と内容の曖昧さは負の影響を持つ
- **外れ値的な論文は普通とは異なる生成メカニズムを持つ**
 - 参考文献の主流性と学際性、内容の曖昧さの正の連動が特徴
- **Wernerfelt(1984)は普通とは異なる生成メカニズムを持っている**
 - 10年後の被引用回数が上位10%になるポテンシャルを1984年の時点で持っていた
- **本研究の限界**
 - 他の分野では、これらの指標は別の動きをする可能性が高いため、分野ごとにモデルを構築する必要がある
 - 外れ値の分析はサンプル数が少なくならざるをえないため一般化可能性に乏しいことから、今回対象とした標本内のみで妥当な議論にこの議論は限定される

研究②と研究③について

研究② 変則事例は学術領域をどう変えるのか？

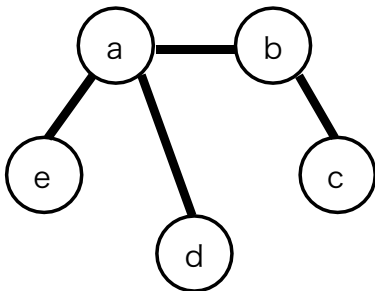
以前は主流ではなかった論文が、後に主流と考えられるようになるという変化を機械学習を利用してモデル化した。また、この変化の原因の探索を可能にした。
International Conference on Information Systems 2016の発表が原案。



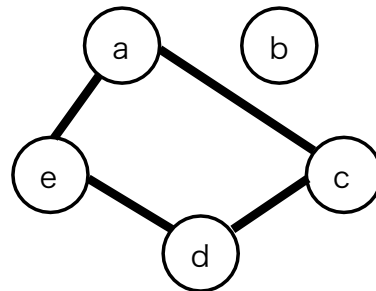
研究③ 変則事例は概念の意味をどう変えるのか？

清木ら（1996）の「意味の数学モデル」では文脈が単語の意味を決めると考える。論文群を文脈と捉えると、主流の論文群と変則的な論文群では同じ単語の意味が異なることが予想できる。この意味の異なりは科学の発展の本質である（Kuhn, 1970）ことから、新しい概念と研究領域の出現を検知することが期待できる。

主流の論文群の単語の共起性



変則的な論文群の単語の共起性



単語の共起性から意味が創発すると考えると、この2つは意味が異なる

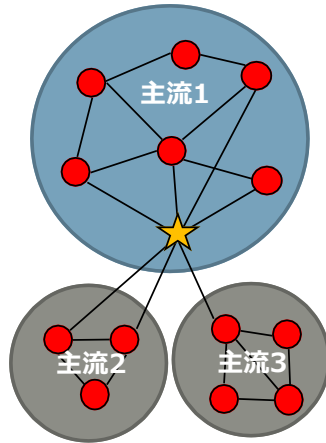
参考文献

1. Guimera, R., & Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature*, 433(February), 895–900. <https://doi.org/10.1038/nature03286.1>.
2. Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
3. Hofmann, T. (1999). Probabilistic Latent Semantic Analysis. *Proc of Uncertainty in Artificial Intelligence UAI'99*, 50(4), 21. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>.
4. Varadarajan, J., & Odobez, J. (2009). Topic Models for Scene Analysis and Abnormality Detection. In *The ICCV Visual Surveillance (ICCV-VS) workshop*. Japan: Kyoto.
5. Wernerfelt, B. (1984). The Resource-Based View of the Firm. *Strategic Management Journal*. 5 (2): 171–180.
6. 清木康・金子昌史・北川高嗣. (1996). 「意味の数学モデルによる画像データベース探索方式とその学習機構」. 『電子情報通信学会論文誌』, 79(4), 509–519.

APPENDIX

APPENDIX(1) 参考文献の類似度から主流度・境界度を計測

- Guimera & Amaral (2005) を基に指標を作成
- 経営学の学術領域の中には複数の主流があると想定



■ 参考文献の主流性（主流性）z

主流の中でどれだけ論文は中心的か？

主流の中のつながりの数に比例

★のzは次数4に比例（主流内で標準化）

■ 参考文献の境界性（学際性）P

他の主流の影響をどれだけ論文が受けているか？

論文のつながりのうち、他の主流とのつながりの割合

★のPは(2/7 + 1/7)に比例

APPENDIX(2) 外れ値検知による内容主流性（普通さ）の計測

- Varadarajan & Odobez, 2009を基に指標を作成
- 内容の主流性（普通さ）の計測
 - トピックモデル（pLSA）は、論文の内容に共通するいくつかの**典型的なパターン**を学習する（Hoffman, 1999）。そこで、これらのパターンへの当てはまりの悪い論文を外れ値と考える。
 - パターンへの適合度合いである**尤度**によって、論文の内容の普通さを計測する。
- 内容の境界性（曖昧さ）の計測
 - ある論文が複数のパターンに所属しうる場合に、論文が境界的（曖昧）であると考えられる。
 - パターンへの所属確率の**平均情報量（エントロピー）**によって曖昧さを計測する。

研究① 異端な研究の価値の有無は見分けられるのか？

RESULT : 変数の依存関係の分析結果

■ 変数間の距離を相関係数で計測。外れ値群はデータ全体とは異なる依存関係を持つことが明らかになった。

	データ全体	上位5%	上位2.5%
被引用回数との関係	主流性と正の相関 曖昧さ・学際性と負の相関	普通さ・曖昧さ・主流性と 正の相関	曖昧さ・主流性と 正の相関
説明変数の依存関係	<p>$r = 1.0$</p>	<p>$r = 0.54$ n.s.</p>	<p>$r = 0.48$ n.s.</p>

研究① 変則的な研究の玉石混交は見分けられるのか？

APPENDIX(3) ベイズロジスティック回帰分析結果

■ モデルの概要：Scopus®から取得した、1970年から1984年までにSNIP1.0以上の経営学上位ジャーナルに掲載された7,499本の論文からネットワークとモデルを作成した。1984年の論文962本に対して4つの指標を計算し、参考文献主流性・境界性が計算できなかった論文を除いた628本を分析に使用した。この628本の論文の10年後（1994年）の被引用回数を目的変数とした。

	オッズ比	オッズ比が1未満になる確率	オッズ比が1以上になる確率	有効標本数	収束判定 Rhat
定数	0.869	0.899	0.101	4275	1
内容の普通さ	0.961	0.645	0.355	5020	1
内容の曖昧さ	0.684	1	0	5188	1
参考文献の主流性	1.553	0	1	5600	1
参考文献の学際性	0.589	0.894	0.106	4209	1

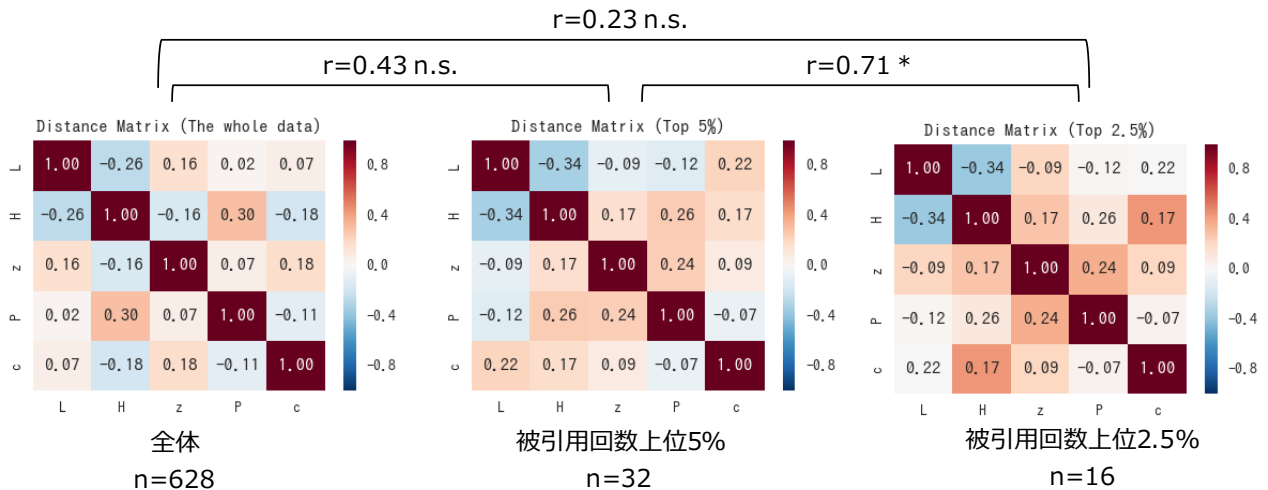
■ モデルの評価：データの8割から学習させ、残り2割を使ってモデルの評価をした

	実際に中央値より多い	実際に中央値未満
中央値より多いと予測	20 ○	16 ×
中央値未満だと予測	28 ×	62 ○

精度は65.1%
汎化可能性を考えると妥当

APPENDIX(4) 4つの変則さと被引用回数の依存関係

それぞれのグループで変数間の距離を相関係数によって計測することで変数間の依存関係を計測した。距離行列の類似性（構造同値性）を相関係数で計測、外れ値群は特殊な依存関係を持つことが明らかになった。



L: 内容の普通さ H: 内容の曖昧さ z: 参考文献の主流性 P: 参考文献の学際性 c: 10年後の被引用回数