

テキスト・マイニング研究会 活用セミナー

2006年9月7日

於: 日本電子計算株式会社

テキスト・マイニング研究会
http://wordminer.comquest.co.jp/
統計数理研究所
大隅 昇
ohsumi@ss.ij4u.or.jp

All rights reserved. Copyright by Noboru Ohsumi, ISM Professor Emeritus.

質的研究への新たな関心, そして回帰現象

- 大量のまとまった量的データの確保, 取得が次第に困難となってきた. 例: **調査環境の悪化**
- とくに社会調査の標準的な方法である**定量型調査**(選択肢形式の質問文による調査など)が難しくなってきた.
- 従来型の調査方法論, とくに統計的な標本調査理論の適用の困難性が増した. 例: **新たな調査方式への移行現象**
- 一方, 大量の, 見かけ上は情報が豊富に見える**電子化データ**が増えてきた.
- 定性調査・質的調査における**取得データの電子化**も機器類の改善で容易となってきた.
- 結果としてテキスト型データの(形式的な)取得・確保が容易になった. 例: **Web調査, Web検索情報, 新聞記事コーパス**など
- 質的情報からの現象解明への期待感の高まり.
- しかし質的研究, 定性調査などの研究は**昔からある**.
- **対象分野の拡がり**が見られる(看護・福祉, 社会心理, MR, ...).

テキスト・マイニングへの過剰な期待?

- テキスト・マイニング(TM)は「**魔法のランプ**」「**打ち出の小槌**」**ではない**.
- **何となく集めたデータ**から, うまい話し, 画期的な知見, 期待する結果は得られない.
- データマイニング(DM), TMは宝の山を発掘する方法というが, **そもそも宝の山はどこにあるのか**, それを知りたい(分からない), つまり論理矛盾がある.
- 要するにある現象解明, 現象理解に役立つというが, **元となるデータ資源(⇒情報源)は何処にあるのか**.
- DM, TMで画期的な大発見があったか, あるいは多大な利益を生むことがあったか? 冷静かつ謙虚に考えるべき.
- **データ解析の手間が減った訳ではない**, むしろ**煩雑化**した.

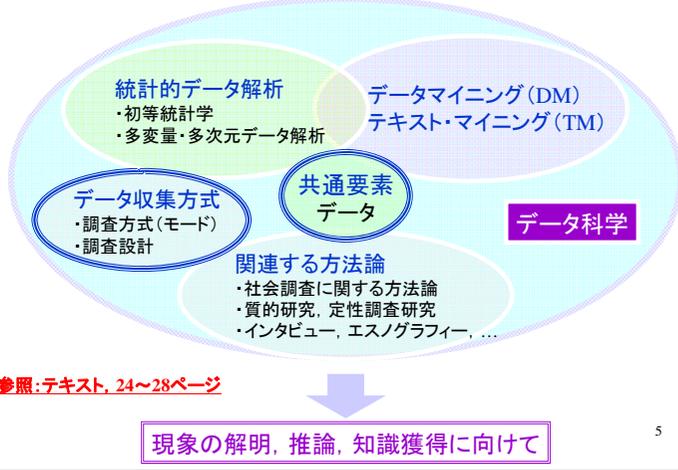
3

統計的データ解析とマイニングの違いは?

- 統計的データ解析・統計分析の弱点は「データの収集から始めねばならないこと」, しかし「大量に集積したデータから探す方法論=DM, TM」を使えばよい, という主張は正しいか?
- つまり構造化された(大規模)データベースのような資源があればよい, データウェアハウス, ドキュメント・ウェアハウスなどを使えばよいのだ, という主張は?
- 大規模なデータ(≡宝の山)を対象にコンピュータの高速処理性を活かして宝探しを行えばよいとの発想.
- これらは本当だろうか?
- 基本的には「**ほとんどNo**」といえるだろう. 新しい言葉・表現に惑わされてはいけけない. **そもそも対立の構図で考えるべきではない**. 相互補完的である.
- そもそも, **データ解析=マイニング=鉱脈探し=探索**そのものである.

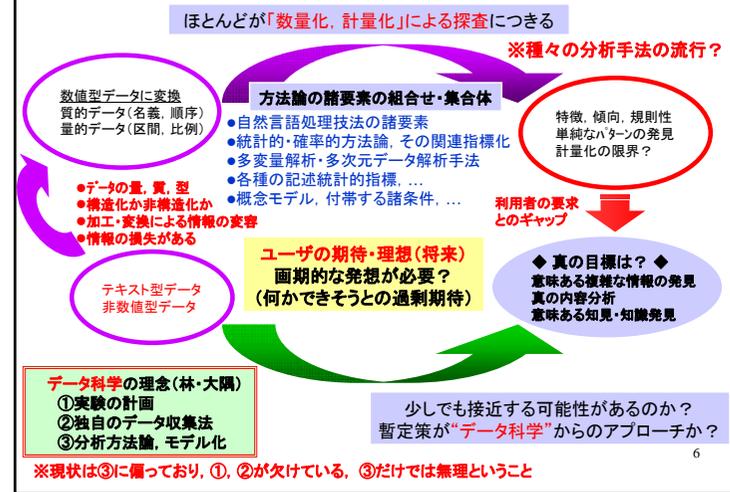
4

統計的データ解析, TM・DM, 他の関係



5

現状のテキスト・マイニングは？



6

質的研究関連

- Norman, K. Denzin, Yvonna, S. Lincoln (eds.) (1994): *Handbook of Qualitative Research*, Sage Publications, Inc. の翻訳版.
- 平山満義 (監訳)として, 以下の3分冊で発刊されている.
 - ①Norman, K. Denzin, Yvonna, S. Lincoln (編), (2006): 質的研究ハンドブック(第1巻), 質的研究のパラダイムと眺望, 北大路書房.
 - ②Norman, K. Denzin, Yvonna, S. Lincoln (編), (2006): 質的研究ハンドブック(第2巻), 質的研究の設計と戦略, 北大路書房.
 - ③Norman, K. Denzin, Yvonna, S. Lincoln (編), (2006): 質的研究ハンドブック(第3巻), 質的研究資料の収集と解釈, 北大路書房.
- Uwe Flick (1995): *An Introduction to Qualitative Research*, second edition, Sage Publications, Inc.
(訳書)ウヴェ・フリック著, 小田博志他訳(2002,2004): 質的研究入門, 春秋社.
- Sage Publications, Inc.から関連書が多数発刊されている.

7

インタビュー調査技法関連

- 桜井厚(2002,2006): インタビューの社会学, せりか書房.
- 河西宏祐(2005,2006): インタビュー調査への招待, 世界思想社.
- 吉田寿夫編著(2006): 心理学研究法の新しいかたち, 誠信書房.
- 木下康仁(2003,2006): グラウンデッド・セオリー・アプローチの実践—質的研究への誘い, 弘文堂.
- 桜井厚, 小林多寿子(編著)(2005): ライフストーリー・インタビュー—質的研究入門—, せりか書房.
- ジェイムズ・ホルスタイン, ジェイバー・グブリアム著, 山田富秋, 兼子一, 倉石一郎, 矢原隆行訳(2004): アクティヴ・インタビュー—相互行為としての社会調査—, せりか書房.
- 山崎敬一編(2004): 実践エスノメソドロジー入門, 有斐閣.

8