

マイクロデータの利用技術と EBPM（統計数理研究所共同利用重点テーマ）

セッションテーマ解説：マイクロデータに関する構造化技術と意思決定

統計数理研究所 山下智志

1. はじめに

平成 29 年 2 月より首相官邸が中心となって統計改革推進会議が開催された。そこでは国民により信頼される行政を展開するためには、統計を積極的に利用して、証拠に基づく政策立案 (EBPM) が必要であるとされた。そのためには、ユーザーニーズを反映した統計データと提供システムが不可欠とされ、正確さや利便性の改善が提案されている。この流れとして、オンサイト拠点の設営や政府マイクロデータの研究利用が推進されている。一方で、個人情報保護法に代表されるプライバシー保護も重要な課題として認識されるようになった。このような現況は学術的研究としては以下のような課題を生み出している。

- ・政府マイクロデータ特有の性質に見合った分析方法の開発
- ・政府マイクロデータと他のデータの併用による分析
- ・統計分析の結果を社会実装や意思決定に活用する方法
- ・匿名性の定義、匿名化の方法論
- ・名寄せやデータリンケージの技術
- ・マイクロデータの欠損値・異常値処理などのクレンジング技術
- ・EBPM に関する体制や実効性の評価方法
- ・オンサイト拠点の活用

2. データ構造化技術とデータ結合

マイクロデータは集計されたマクロデータとは違い「未加工」のまま供給される。そのためデータベースには欠損値、異常値が存在し、またデータ間の結合に用いられる ID キーが欠落しているなどの問題点がある。そのため、実際にマイクロデータを統計分析に用いるにはデータ構造化処理が不可欠である。欠損値補間についてはこれまで多くの実績があるが、異常値処理、データ結合については発展途上である。また、匿名化やプロファイリング技術も重要な課題となっている。

本発表および本セッションにおいて、マイクロデータの構造化の方法論とそれを利用した実証分析について議論する。

本発表および本セッションにおいて、マイクロデータの構造化の方法論とそれを利用した実証分析について議論する。

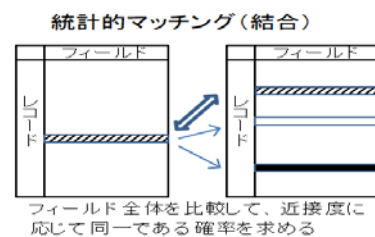
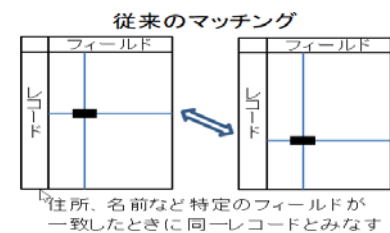


図1 従来のマッチングと統計的マッチングの違い

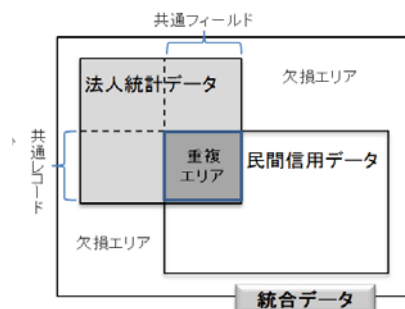


図2 統合データベース作成のイメージ

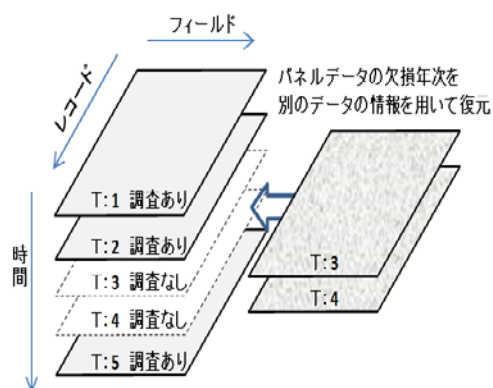


図3 データフュージョン(パネル)のイメージ

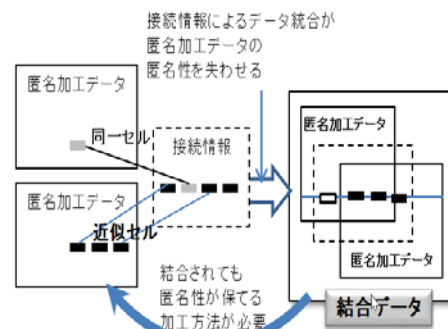


図4 プロファイリングによる非匿名化