論文審査の要旨及び担当者

報告番号	甲　乙　第　　号	氏　名	光廣　正基

論文審査担当者	主　査	： 長倉　大輔	（慶應義塾大学経済学部教授、Ph.D.）
	副　査	： 中妻　照雄	（慶應義塾大学経済学部教授、Ph.D.）
		： 星野　崇宏	（慶應義塾大学経済学部教授、博士（経済学））
	面接担当	： 新井　拓児	（慶應義塾大学経済学部教授、博士（理学））
		： 片山　翔太	（慶應義塾大学経済学部准教授、博士（理学））

（論文審査の要旨）

論文題目

Latent Variable Modeling for Statistical Data Combination Problem

1. 論文の概要

本論文の主たる目的は，ソースが異なる複数の多変量データを“（擬似的な）シングルソースデータ”として統合させる新たなデータ融合法の提案である．マーケティングや経済学の分野では，一つのソースから関心のある変数が全て観測されたシングルソースデータを取得することは難しいが，代わりに個人やユニットが異なる複数のソースからデータを取得することは比較的容易であり，これらのマルチソースデータ活用のニーズは高い．各データセットでのみ観測される変数間の関係を調べるため，マルチソースデータをシングルソースデータとして統合する、統計的データ融合やデータコンビネーションという方法が提案されている．これらの手法は，マルチソースデータをシングルソースデータとみなしたとき，ユニットごとに発生する欠測値を推測、補完する方法でもあり，欠測データの問題とも関連がある．

他にも，擬似パネルという方法が提案されており，クロスセクションデータからユニットごとに推論を行う方法は，政策や意思決定などに活かす実証研究で有効とされている．さらに，政府統計の一つである家計調査では，企業が保有する購買履歴データを使い，共通属性を共変量としてデータ融合を実施することで，国民全体の総購買額などの集計値に対して，回答者の偏りを補正する研究が行われている．

近年のデータ活用の拡大に伴い，データ分析の場面ではマルチソースデータを融合する機会が増え，欠測データの分析も必要とされている．このような背景を踏まえ本論文では，既存のデータ融合法の問題点を整理し，それらを状況設定に応じて解決するための新たな 4 つの手法を提案している．欠測データのメカニズムのモデリングを行う，潜在変数を活用するといった工夫を行うことで，マルチソースデータを統合する際に発生する欠測値の推定精度の改善を図るこれらの手法により，既存のデータ融合法の適用が不向きなマルチソースデータに対しても適用可能とすることを目的としている．

本論文は 7 つの章で構成されており，第 1 章では，導入として各分野で共通しているデータ融合問題について述べ，第 2 章では既存のデータ融合法の特徴を比較し，欠測メカニズムとの関係を述べ，第 3 章から第

論 文 審 査 の 要 旨

6 章において、新たなデータ融合の手法を提案し、第 7 章は結論として本論文の内容をまとめる。以下では本研究の中心である第3 章から第6 章の内容を要約する.

第 3 章では，カーネル正準相関分析と統計的マッチングを組み合わせたデータ融合法を提案している．経済学での分析でも広く利用されている統計的マッチングでは，ペアとなった相手の観測値を欠測値に代入することで欠測値が推定される．しかし，マルチソースデータの共変量が高次元な場合は，共変量の値の組み合わせが複雑になったり，注目している変数間の非線形な関係を扱ったりすることがあるため、この方法を使うのは難しい．本章では，多変量解析で使用されるカーネル正準相関分析に着目し，共変量を低次元に縮約した後に統計的マッチングを適用する 2 段階の手法を提案している．この提案手法は，2 つのデータの共変量とアウトカム変数間の関係性を保持した共通の正準変量を推定し，その正準変量の距離を利用して統計的マッチングを行う．シミュレーションデータと実データの適用結果から，提案手法とマハラノビスマッチングの予測値と真値の二乗誤差を比較し，提案手法の有用性を示している.

第 4 章では，誤差最小化で補正ウエイトを推定する制約付きマッチングによるデータ融合法を提案している．統計的マッチングでは，共変量の分布がマッチング前後で異なる場合があり，特にランダムでない欠測のときは影響を受けやすい．マッチングの前後でモーメント情報が大きく変化しないように，平均値や分散といった代表値に制約を追加した多対多マッチングが提案されているが，加えた制約に合わせて最適化するマッチングアルゴリズムが必要となる．本章では，多変量解析の因子分析などで用いられる行列分解で推定されるウエイトを応用し，共変量の平均値が近づくような制約を追加した多対 1 マッチングを提案している．この提案手法は，制限したいモーメント情報の数が増えても行列演算により同じ更新式で推定可能となる．適用例では，制約の強さの違いにより，マッチングの予測精度とマッチング前後の共変量の平均値の差における変化の様子が確認できた.

第 5 章では，ガウス過程による潜在変数を用いたデータ融合法を提案している．データ融合法の多変量解析的アプローチとして提案されている既存の潜在変数モデルではパラメトリックな推定が必要であり，モデルの仮定から逸脱すると欠測値の予測精度が悪化することが知られている．本章では，近年機械学習の分野で応用されているガウス過程に着目し，これに従う潜在変数を用いたノンパラメトリックモデルを提案している．この提案手法は，ガウス過程による統一的な枠組みで離散変数と連続変数を持つような複雑なデータに適用できる確率生成モデルであり，ベイズ推定を用いて各種パラメータを推定し，推定された潜在変数から欠測値を推測する．シミュレーションでは，ランダムな欠測データへの適用を前提とした上で，既存手法のマハラノビスマッチングや潜在変数モデルと比較し，提案手法による改善が示された.

第 6 章では，ランダムでない欠測データに対するガウス過程を用いたデータ融合法を提案している．第5 章で提案したランダムな欠測に対するガウス過程を用いたデータ融合法を，ランダムでない欠測でも扱

論 文 審 査 の 要 旨

えるように拡張した．回答者や調査方法の違いにより，選択バイアスを持ったマルチソースデータに対して，既存のデータ融合法を適用した場合の予測精度が悪く，欠測するかどうかの確率もモデリングする必要がある．本章では，欠測するかどうかの確率に離散選択モデル，今回はプロビットモデルを仮定し，ガウス過程による潜在変数を用いたデータ融合法に組み込んだ．この離散選択モデルに対してもガウス過程による潜在変数を用いるため，観測値におけるモデリングと統一的なフレームワークでのモデル構築となっている．シミュレーションでは，第 5 章で比較した手法に加え，ランダムでない欠測に対応した潜在変数モデルと本章での提案手法の予測結果を比較し，提案手法の有用性が示された．適用例では，企業財務データから得られる広告宣伝費と企業ブランド調査データから得られる購入意向率との関係を調べるため，これらのデータを統合し，提案手法と既存手法の欠測値の予測結果を比較している．企業財務データは企業戦略によって情報の開示内容が異なることで，シングルソースデータの入手が難しく，さらに企業ブランド調査データは調査対象企業しか得られないため，すべての企業からすべての項目を得ることが不可能な例である．適用例の結果では，既存手法やランダムな欠測を前提とした手法と比較し，提案手法による予測値の改善が見られ，ガウス過程を用いてランダムでない欠測を考慮したデータ融合法の有用性が示された．

2. 論文の評価

近年，企業が所有するビッグデータや大規模な調査データなどあらゆる目的で取得されたデータが活用可能となってきており，経済学においても，個人の経済行動の実態把握など蓄積されたデータの活用シーンは増えつつある．一方で今後はデータ活用の法整備が進み，一定の制限が加わることも考えられる。本研究で提案されたデータ融合法は応用研究の重要なツールとしての利用が期待され，欠測データの検証結果は，データ分析とその周辺分野において重要な示唆を与えていると評価できる．特にガウス過程を用いた第 5 章の内容、これをランダムでない欠測の状況でも利用可能なように拡張した第 6 章の成果は独創的であり，機械学習系の学会においても複数の研究者からの注目を集め、このテーマについての講演を依頼されるなどしている。なお，本論文の元となった研究成果は，第 3 章の内容は査読付きジャーナルである Japanese Journal of Statistics and Data Science に採択されており，第 5 章や第 6 章は統計学の国際誌に現在投稿準備中である．

しかし，潜在変数による条件付き独立性の仮定を置くデータ融合法を実データに適用する際，その仮定の妥当性や、モデルの精度の評価方法など課題が残るという声も審査委員の中にはあった．他にも，2 つのデータだけでなく，複数データの統合を行う際の方法論についてまだ十分研究が行われていないなどといった課題があるという指摘が出された．これらは，本論文の学術的な価値や貢献を損なうものではなく，

論 文 審 査 の 要 旨

本論文で提案された手法が幅広く利用されるための今後の研究課題といえる.

以上の所見から，本論文はデータ分析を行う多くの分野での応用研究の発展に貢献する新規性のある内容であり，審査委員会は全員一致で本論文が博士（経済学）の学位を授与するにふさわしいと判断する.