

平成 30 年 6 月 20 日現在

機関番号：82627

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26350446

研究課題名(和文) 国際輸送に関する異種統計データの融合と高精度化

研究課題名(英文) statistical data fusion and accuracy improvement for international transport data

研究代表者

小坂 浩之 (KOSAKA, Hiroyuki)

国立研究開発法人海上・港湾・航空技術研究所・その他部局等・研究員

研究者番号：90349225

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究は、国際輸送を対象とした同種・異種の統計データに関するデータフュージョンを検討し、データの高精度化や新しい情報の作成を行うことを目的とする。国際輸送の同種統計データとして貿易統計を取り上げ、国際輸送の異種統計データとして、貿易統計と船舶動静データを取り上げた。貿易統計に関しては、数量データの詳細な特性分析による国際貨物流動量の高精度化や国際輸送コストの推計を行った。貿易統計と船舶動静データに関しては、欠損値補完法の枠組みにより船舶の積載貨物量の推計を実施した。

研究成果の概要(英文)：This study examines data fusion method for same type data and different types data for international transport statistics. The data fusion is aimed at increasing the accuracy of statistical data and creating new information. We focus on import/export data of trade statistics as the same type data, and trade statistics/ship movement data as the different type data. In the analysis of trade statistics, we developed a method for improving the accuracy of international freight flow data and estimating international transportation costs. In the analysis of trade statistics and ship movement data, we created a method to estimate ship laden cargos according to the framework of missing value imputation method.

研究分野：国際輸送

キーワード：データフュージョン 国際輸送 貿易統計

1. 研究開始当初の背景

社会・経済学の分野では、欠損値補完の問題としてデータフュージョンを定義し、多くの適用事例が存在する。この際のデータフュージョンは、「複数のサンプルから得られたデータを、模擬的な単一のサンプルから得られたデータに統合する手法」と定義されている。これは、サンプルA群とサンプルB群では、異なる変数が観測され、A群とB群の変数を統合したサンプルC群(完全データ)を作成することを目的とした手法である。変数の統合にはマッチング手法等が存在し、A群とB群の共通変数の特性が最も近い場合、A群のサンプルとB群のサンプルを統合することが行われている。よって、A群(もしくはB群)のみに存在する変数は、B群(もしくはA群)の欠損値と考えられるため、欠損値を補完することがデータフュージョンと定義できる。従来のデータフュージョンは、異なる対象物に対して、異なる変数がデータとして得られ、それを統合して、ある1つの対象物に関する完全データを得ることを前提としている。この前提条件に依存して、従来のデータフュージョンの手法は、欠損値の問題に対処する方法論と位置付けられる。しかし、運輸・交通分野では、欠損値の問題として扱うだけでは、利用可能なデータを有効に活用することはできない。その原因として、複数の異種統計データ群(貿易統計、港湾統計、船舶移動データ等)が、異なる状態(金額、数量、キャパシティ等)で共通変数を捉えているため、マッチング手法等を単純には利用できないことが考えられる。

2. 研究の目的

本研究は、国際輸送を対象とした同種・異種の統計データに関するデータフュージョンを検討し、さらにはデータの高精度化を行うことを目的とする。国際輸送の同種統計データとして貿易統計を取り上げ、国際輸送の異種統計データとして、貿易統計と船舶動静データを取り上げる。国際輸送の利用可能なデータは、多様性と複雑性を持つことから、適切なデータフュージョンの実施により、高度な知見を提供することが可能である。

3. 研究の方法

多くの研究は、複数の統計を同一の精度と考慮して組み合わせ使用しているが、統計間で精度が異なり、高い精度の統計で一方の統計を修正するような試みも存在する。本研究は、そのような統計データの精度や修正等に着眼する上で、多くの研究実績があり、国際輸送の分野で利用価値の高い貿易統計を基礎的な対象とする。約200カ国・地域は、貿易統計データを国連に報告し、世界の90%以上の貿易量が国連によって把握されている。貿易統計では、ある国の輸入と、それに対応する相手国の輸出が、それぞれの国のデータとして記録される。同一の貿易フローが異なる

輸入データと輸出データとして捉えられることから、輸入と輸出のデータ群を同種の異なるデータ群と捉えることができる。そのため、貿易統計の輸入データと輸出データを同種統計と考え、データフュージョンの検討を行う。貿易統計における国際貨物と輸送機関の移動データは、米国等の一部の地域でのみ同時にデータが収集されている。そのため、貿易統計と輸送機関の移動データに関するデータフュージョンは、国際輸送を検討する際に有用となる。国際海運では、海運コンサルタント等が、世界の国際船舶を対象とした港湾間移動のデータベースを作成している。この船舶動静データでは、バルク貨物・原油・石油製品等の品目を単一的に積載するタンカーやバルクキャリアについて、積載貨物量の特定が進められている。本研究は、貿易統計の貨物流動量と船舶動静データの船舶移動量を異種統計データと考え、データフュージョンの検討を行う。

4. 研究成果

(1) 貿易統計の数量データの活用

貿易統計は、上述の輸入データと輸出データが一致しない不整合問題と呼ばれる特性が存在する。不整合問題は、価額について多くの検討事例がある。その際、整合性は、輸入価額と輸出価額の比率である整合率(輸入価額/輸出価額)によって評価されている。不整合の要因は、不可避な要因、エラー、構造的な要因の3つに分けられる。1つ目の不可避な要因とは、輸出時点と輸入時点のタイムラグ、輸入と輸出の記録方式の相違である。記録方式は、輸入価額において運賃・保険料を含むCIF(Cost, Insurance and Freight)形式、輸出価額において運賃・保険料を含まないFOB(Free On Board)形式で記録することが一般的である。2つ目の要因であるエラーとは、密輸、意図的な過大・過小申告、品目名の偽申告等である。3つ目の構造的な要因とは、貿易統計の作成方法の相違等が挙げられ、経由貨物に関しては、相手国の定義(輸入の原産地国、輸出の仕向け国等)の相違が大きく影響する。本研究は、アジア27カ国・地域の2000年から2006年の貿易統計を使用して、価額の整合性を確認した。アジア諸国では、香港とシンガポールの経由貨物(再輸出額)が、不整合の主要な要因となる。

既存研究は、貿易統計の価額に着目した検討が中心であることから、本研究は、数量に着目した検討を進めた。貿易統計の輸入データと輸出データの相互比較を実施し、数量データの特性から整合的と考えられる調整データを作成した。数量整合率は、輸入と輸出で記録方式の相違が存在しないため、1.0が整合的と明確に判断できる。図1は、品目が特定できる一般的なデータ(品目特定データ)と調整データに関して、品目別価額整合率を算出し、その常用対数の頻度分布を示している。価額整合率は、CIF価額とFOB価額

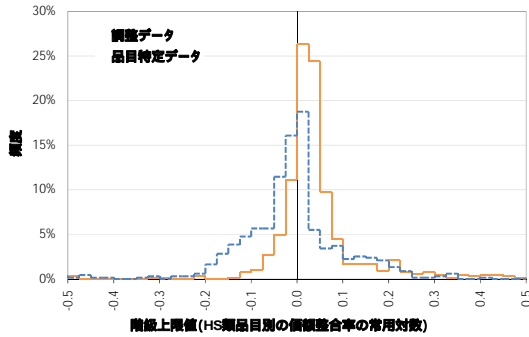


図1 品目特定データと調整データの品目別価額整合率の分布

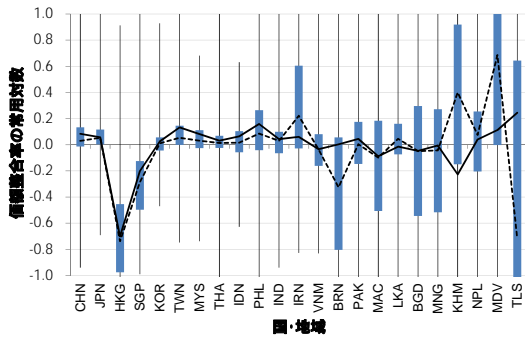


図2 国・地域別輸入価額の品目別整合率の統計値 (品目特定データ)

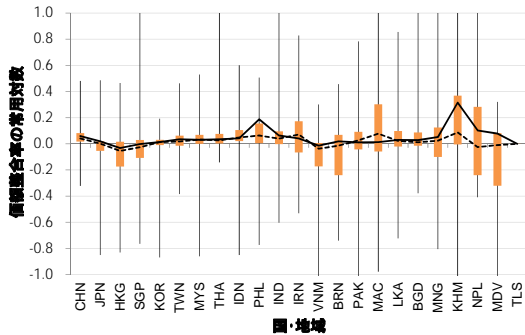


図3 国・地域別輸入価額の品目別整合率の統計値 (調整データ)

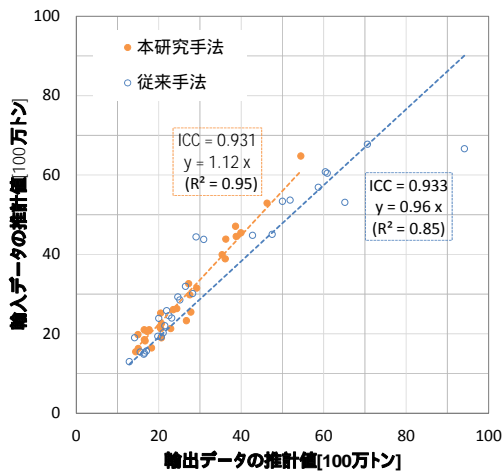


図4 調整データの換算係数による国際貨物流動量推計結果の改善

の相違から1.0よりも大きくなる必要があり、調整データは価額整合率を改善させることを明らかにした。図2と図3は、品目特定データと調整データに関して、国・地域別の輸出額整合率の箱ひげ図を作成したものである。調整データにおいて、価額整合率のばらつきが小さく、調整データの有効性が示されている。また、香港やシンガポールの輸出額の整合性も改善する。

整合的な貿易統計を作成することは、様々な国際輸送の検討に有益である。本研究は、国際貨物流動量を推計する視点で、貿易統計の整合性が改善する利点を示した。国際貨物の検討の際、貿易統計を使用して国間の貨物量を推計することが広く行われている。一般的な貿易統計は、価額データと複数の数量単位を持つ数量データで構成されているため、数量単位を統一する作業が伴う。例えば、重量単位に統一する場合、価額データと重量データが存在する品目に関して、両データを品目グループで集計し、品目グループでの価額と重量間の換算係数を作成する。品目グループ内の重量データが存在しない品目に関して、その価額データに品目グループの換算係数を掛け合わせて重量データに換算できる。本研究は、調整データに基づき換算係数を算出し、重量単位の国際貨物流動量の推計を行い、簡便な方法で精度良く重量単位の貨物量が推計可能であることを示した。図4は、調整データを利用することによって、輸入データと輸出データに基づく国際貨物流動量が、大きな乖離がなく推計できることを示している。

(2) 数量データの詳細分析と欠損値補完

貿易統計の価額データと数量データを相互に利用することの有用性から、両データの整合性に関して詳細な分析を行った。その結果、価額整合率と数量整合率に比例関係が存在することを確認した。この比例関係は、品目別に異なる強さを持つ。具体的には、ASEAN+6 (ラオスを除く)、台湾、マカオ、香港の18ヶ国・地域における2000年から2014年の貿易統計を使用し、詳細な品目別の価額整合率と数量整合率を算出した。その後、HS類品目グループ(96品目)別に、価額整合率と数量整合率の相関係数(r)を算出した所、平均で0.87となった(最小0.57、最大0.95)。図5は、HS26類品目(鉱石、スラグ及び灰)における価額と数量の整合性の関係を示している。

数量データの利用において、欠損値の処理が課題となる。国連が各国・地域の貿易統計に基づいて作成するデータベースは、1つのレコードの基礎的な項目として、品目バージョン、貿易統計の報告国、貿易フロー(輸入、輸出、再輸出、再輸入)、相手国、品目コード、価額、数量単位、数量、重量、推計フラッグを持つ。国連貿易統計は、数量と重量において、欠損値(0もしくは空欄)が存在し

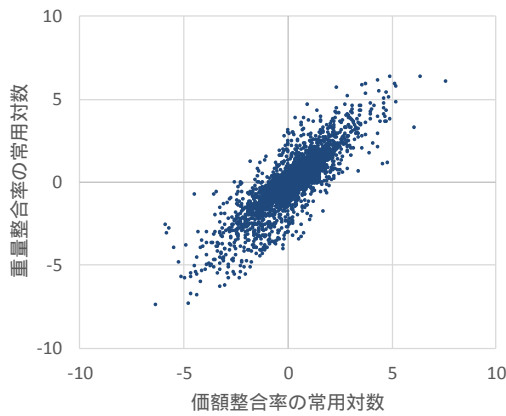


図5 価額整合率と重量整合率の関係
<HS26 (鉱石、スラグ及び灰) $r=0.85$ >

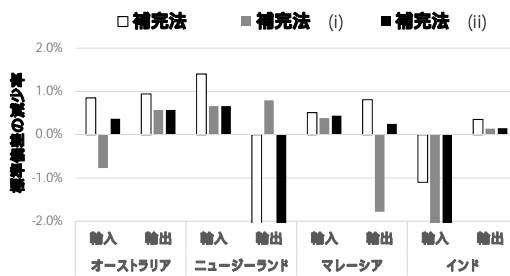


図6 補完法 における数量整合率の変化

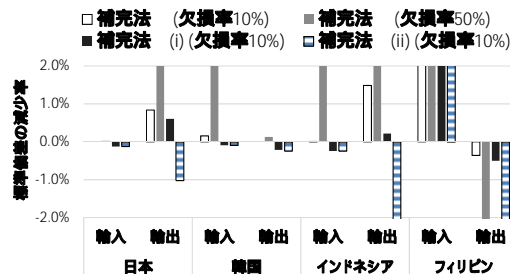


図7 補完法 における数量整合率の変化

ている。本研究は、上述の 18 ヶ国・地域の 1996 年から 2006 年の貿易統計を使用し、数量と重量のデータの欠損値を計数した。その結果、1999 年までは 10%程度、2000 年以降は状況が改善し 4%程度の欠損値が存在することを確認した。なお、貿易統計の価額では欠損が生じていない。本研究は、数量データの欠損値補完法を検討するため、以下の ~ を実施した。

実在する欠損値を同一年のデータで補完：報告国 A の k 年における数量欠損データに対し、対応する相手国 B で k 年に数量データが存在する場合、相手国 B の単価を用いて報告国 A の数量データを算出。
 実在する欠損値を過去年のデータで補完：と同様の作業で過去年の単価を使用。
 仮想的に欠損値を作成し、同一年のデータで補完：と同様の方法を仮想で作成した欠損データに対して実施。欠損データの割合は 10%と 50%でランダムに抽出。
 仮想的に欠損値を作成し、過去年のデータで補完：と同様の方法を仮想的に作成し

た欠損データに対して行う。欠損データの割合は 10%と 50%でランダムに抽出。
 補完法 は、欠損データ数が多いオーストラリア等の 4 カ国の貿易統計に対して実施した。品目は、HS84 類「原子炉、ボイラー及び機械類並びにこれらの部分品」に含まれる詳細品目を対象にする。HS84 類では、重量単位に加え個数単位が多く用いられている。補完法 は、欠損データ数の少ない日本等の 4 カ国の貿易統計データに対して実施した。品目は、HS72 類「鉄鋼」に含まれる詳細品目を対象にした。HS72 類では、基本的に重量単位が使用されている。各補完法は、補完実施前後の数量整合率の標準偏差の変化から評価した。その結果、 の同一年のデータを用いた補完法が有力であり、図 6 と図 7 はその結果を示している。

(3) 貿易統計による国際輸送コストの推計
 貿易統計の輸入データと輸出データにおいて、不整合問題の要因である CIF 価額と FOB 価額の相違を明らかにすることは、国際輸送における運賃・保険料を算定することにも貢献する。本研究は、データフュージョンによって、通常の貿易統計のみでは得られない新たな情報を得ることを目的として、貿易統計に基づく国際輸送コストの推計を実施した。対象は、1990 年から 2013 年の主要国における鉄鉱石、石炭等のバルク貨物の国際輸送とした。対象の貿易統計に関して、輸入データと輸出データの相互比較を行い、数量データの精度に基づき国際輸送コストを推計するためのデータを抽出した。海運コンサルタントが作成する運賃データと国際輸送コスト

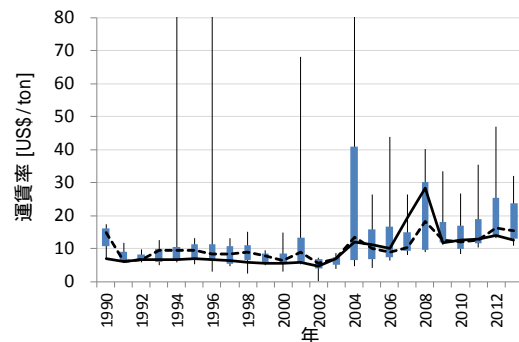


図8 豪州の鉄鉱石輸出の運賃率推計値

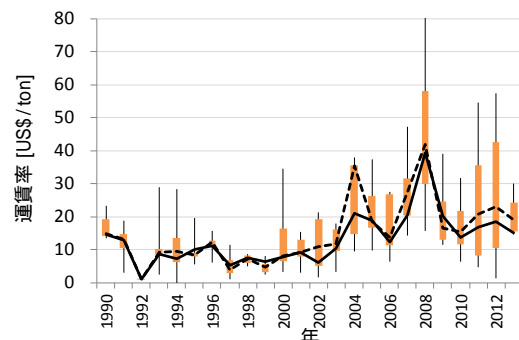


図9 インドネシアの石炭輸出の運賃率推計値

の推計値を比較し、一部の国間では、鉄鉱石と石炭の推計値に関する妥当性を示した。図8と図9は、推計結果を箱ひげ図で示している。また、貿易統計の月間値を利用した国際輸送コストの推計に関して特性を検討した結果、輸送時間に対応する貿易統計の輸入時点と輸出時点のタイムラグを考慮し、国際輸送コストを推計することが有効であることを示した。

(4) 船舶と国際貨物の移動データの統合

貿易統計の国際貨物流動量と船舶動静データの船舶移動量のデータフュージョンに関しては、世界的に把握可能な船舶動静データに基づき、船舶の積載貨物量を推計することを目的に検討を行った。米国の詳細貿易統計では、船舶の積載貨物量が記録されている。また、船舶動静データに存在する船舶の積載能力や積載時の喫水量等は、積載貨物量と関係する。そのため、船舶動静データに基づいて、船舶の積載貨物量を推計することを、欠損値補完法の枠組みで検討を行った。対象は、アジア・米国間における2010年8月のコンテナ船の運航とした。米国の詳細貿易統計から、船舶の積載貨物量を特定し、船舶動静データ等から対応する船舶の積載能力、運航時喫水量、港湾停泊時間を特定し、すべての変数を備えた155隻のデータセットを作成した。その内、ランダムに抽出した100隻のデータセットを欠損値の無い完全データとし、残りの55隻のデータセットについて、積載貨物量のデータを削除し、欠損値データとした。完全データに基づいて積載貨物量を推計する欠損値補完法として、統計的手法である回帰分析と多重代入法、機械学習に位置付けられる人工ニューラルネットワーク(Artificial Neural Network, ANN)とK近傍法を適用し、仮定の欠損値の積載貨物量を推計した。機械学習では、完全データが訓練データに対応し、欠損値データがテストデータに対応する。ANNについては、隠れ層ノードを1とした場合(ANN1)と隠れ層ノードを5とした場合(ANN2)を検討した。各推計方法の推計値と実績値から、平均二乗誤差平方根(RMSE)を算出した結果が図10であり、本研究の実証では機械学習の欠損値補完法が有効となった。

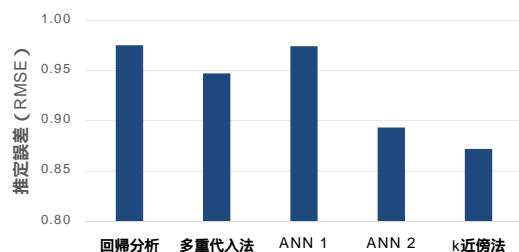


図10 各推計方法における実績値と推計値の推定誤差(RMSE)

5. 主な発表論文等

〔雑誌論文〕(計4件)

小坂 浩之、貿易統計の国際標準化と課題、査読無、環境情報科学、Vol.44、No.4、2016、pp.13-18

小坂 浩之、鹿島 茂、坂本 将吾、布施 正暁、貿易統計不整合問題の調整による国際貨物流動量の推計精度の検討、土木学会論文集 D3(土木計画学)、査読有、Vol.71、No.5、2015、pp.673-680

小坂 浩之、鹿島 茂、貿易統計に基づく三大バルク貨物の国際海上輸送コストの推計、日本物流学会誌、査読有、No.23、2015、pp.79-86

〔学会発表〕(計5件)

小坂 浩之、荒谷 太郎、AISデータに基づくコンテナ船の積載貨物量推計、日本航海学会、2018

小坂 浩之、鹿島 茂、貿易統計データの欠損値補完法の検討、日本ロジスティクスシステム学会、2017

小坂 浩之、鹿島 茂、貿易統計による国際海上輸送コストの推計と特性分析、日本物流学会、2016

6. 研究組織

(1) 研究代表者

小坂 浩之 (KOSAKA, Hiroyuki)
国立研究開発法人海上・港湾・航空技術研究所・その他部局等・研究員
研究者番号：90349225

(2) 研究分担者

鹿島 茂 (KASHIMA, Shigeru)
中央大学・理工学部・教授
研究者番号：70108207

坂本 将吾 (SAKAMOTO, Shogo)
一般財団法人電力中央研究所・環境科学研究所・主任研究員
研究者番号：50580057

(3) 連携研究者

布施 正暁 (FUSE, Masaaki)
広島大学・工学研究院・准教授
研究者番号：70415743