様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月　 ６ 日現在

機関番号：１７１０２
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：１９Ｋ０４４５５

研究課題名（和文）逆最適化を用いた人の意思決定のモデル構築

研究課題名（英文）Modelling of human decision making using inverse optimization

研究代表者
村田　純一（Murata, Junichi）

九州大学・システム情報科学研究院・教授

研究者番号：６０１９０９１４

交付決定額（研究期間全体）：（直接経費）　　2,400,000円

研究成果の概要（和文）：人は自らの判断基準に基づいて行動を決定したり複数選択肢の中から良いものを選択したりする．この判断基準は熟練者のマネをする装置など幅広い応用に活用できる．しかし，判断基準を観測することはできない．観測可能なのは実際の行動や選択された選択肢である．本研究では，この観測可能なデータから判断基準を推定する方法を開発した．この時の最大の問題点はデータから判断基準を一意に決定できないことである．これを解決するために，データの量に応じて判断基準の適切な表現方法を自動的に定める方法，人の行動の揺れ，すなわち毎回最適な行動をとるとは限らないことを活用する方法などを構築し，例題を用いてその有効性を確認した．

研究成果の学術的意義や社会的意義
人が持つ判断基準を推定することができると，その人が好む意匠の発見支援，熟練者と同等の機能を持つ装置，人に不快感を感じさせない行動誘導，運転者個人の嗜好にあった車の自動運転などに活用することができ，大きな社会的意義を持つ．本研究で取り扱っている観測可能なデータから判断基準を推定する過程は，逆最適化問題として捉えることができる．逆最適化問題は唯一解が存在しない不良設定問題であるが，本研究では，表現方法の複雑さとデータの量のバランスをとる方法や，最適解以外の解も活用して利用する情報を増やす方法により，この問題点の解決を図っている．ここに大きな学術的意義がある．

研究成果の概要（英文）：Humans decide their actions and select the best one among available alternatives based on their own judgment criteria. The criteria can be utilized, for example, to design machines that compete with skilled craftsmen. We, however, cannot directly observe them. We only can observe the actions taken and the alternatives selected. The research proposes methods that estimate judgment criteria using those observable data. The biggest issue here is that the data only cannot determine the criteria uniquely. To solve this, the research proposed a method that determines a representation of judgment criterion with a complexity suitable for the amount of given data, and another method was developed that utilizes fluctuations in human actions, i.e., the fact that humans do not always take the best actions, as additional and useful information. Applications of these methods to example problems verified their validity.

研究分野：システム工学・制御工学

キーワード：人のモデル　人の判断　社会サービス　個人化　逆強化学習　多目的最適化　対話型進化計算

１．研究開始当初の背景

情報通信技術の発達や持続可能な福祉・幸福への関心の高まりを受け，社会の機能・サービスは個人への適応や個人の満足を一層指向すると想定される．その実現には人の満足や行動の把握が不可欠である．人は自らの判断基準に従って判断し行動する．しかし従来は，人の行動を把握し表現するモデルとして，状況に対する人の反応を表す外形的表現が多用され，また，人が内面に持つ判断基準が用いられるとしても万人共通の標準的なものが利用されていた．しかし，個々人が納得する車の自動運転，無理がない行動の誘導，個人に応じた学習指導などの「個人化」を行うには，個々人の判断基準の把握が必要である（図１）．

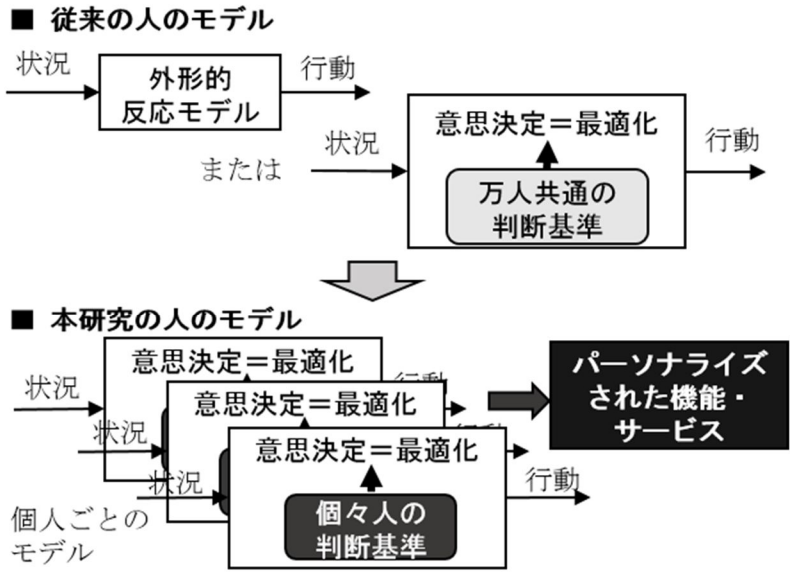

図1 従来の人のモデルと本研究で構築する人のモデル

しかし，人が持つ判断基準は外部からは観察できない．アンケート等による把握は手間がかかる．意思決定の結果である行動のデータから判断基準が推定できると実用的である．人は自らの判断基準に照らしてベストな行動をとると想定される．すなわち判断基準の最適化によって意思決定すなわち行動決定を行う．このことから，人の行動から判断基準を推定する過程は，最適化問題の解から逆に最適化問題を見出す逆最適化となる．しかし，$x=0$ で最大値をとる関数 $f(x)$ が無数にあるように，ある行動を導く判断基準も無数に存在する．また，人は常に最適行動をしているとは限らず，データも不完全な場合がある．このような困難さがある中で行動データから判断基準をどのように推定すればよいかを解決する必要がある．

逆最適化の研究は 2000 年頃から行なわれている[1]．また，報酬和の最大化問題として定式化される強化学習についての逆強化学習も積極的に研究され[2]．[3]，ロボットカー制御のための目的関数推定にも応用されている[4]．しかし，推定のために利用するデータの質・量と目的関数の表現の複雑さ・自由度とのバランスについては，十分な検討がなされていない．

一方，人を対象としたサービスの設計の分野では多数の研究がなされているが，人は，購入量の価格弾性モデル，電力需要と気温等の関係を表す回帰モデルのように，価格や気温などの状況に反応するだけの存在として表現されたり，人の車運転時の評価関数モデル[5]のように熟練度に応じた標準的・平均的な評価関数を使って表現されたりしており，個々人が持つ判断基準を推定して活用しているものは少ない．

２．研究の目的

研究の目的は，人の行動結果のデータを基に逆最適化を利用して人が内面に持つ判断基準を推定する方法を，さまざまな状況に適用できるよう，以下の３項目に対処しながら構築することである．

1. さまざまなタイプの意思決定が存在すること
2. 逆最適化問題の解は無数に存在すること
3. 人の行動やデータは常に理想的とは限らないこと

この 3 項目の課題に対処するために以下を明らかにする．

- 対象とする意思決定問題のタイプに応じた判断基準の表現方法
- 行動データ以外に活用できる付加的情報も含む情報の活用法
- 判断基準の表現と推定に活用する情報の間の適切なバランスの実現方法

これらを踏まえて，さまざまな現実の状況に適用できる判断基準推定手法を構築し，その検証を行う．

３．研究の方法

人が持つ判断基準は行動（例：車の加減速操作）の結果生ずる状態 x（例：車の速度，車間距離）を評価する．状態 x を評価する判断基準を，パラメータ p で規定される関数 f(x;p)を用いて表すと，判断基準の推定は p の推定に帰着される．このとき，観測データ x*が最適行動の結果であれば, x*はどのような x よりも優れている（f(x*;p)≧f(x;p)）という条件が活用できる．この条件は逆最適化問題を解く上での基本条件である．研究すべき内容は以下の４項目である（図２）．

(a)判断基準を表現する関数形 f(x;p)の決定
(b)逆最適化問題の定式化
(c)求める表現と利用する情報の適切なバランスを図りながら解を得る具体的な解法の開発
(d)検証

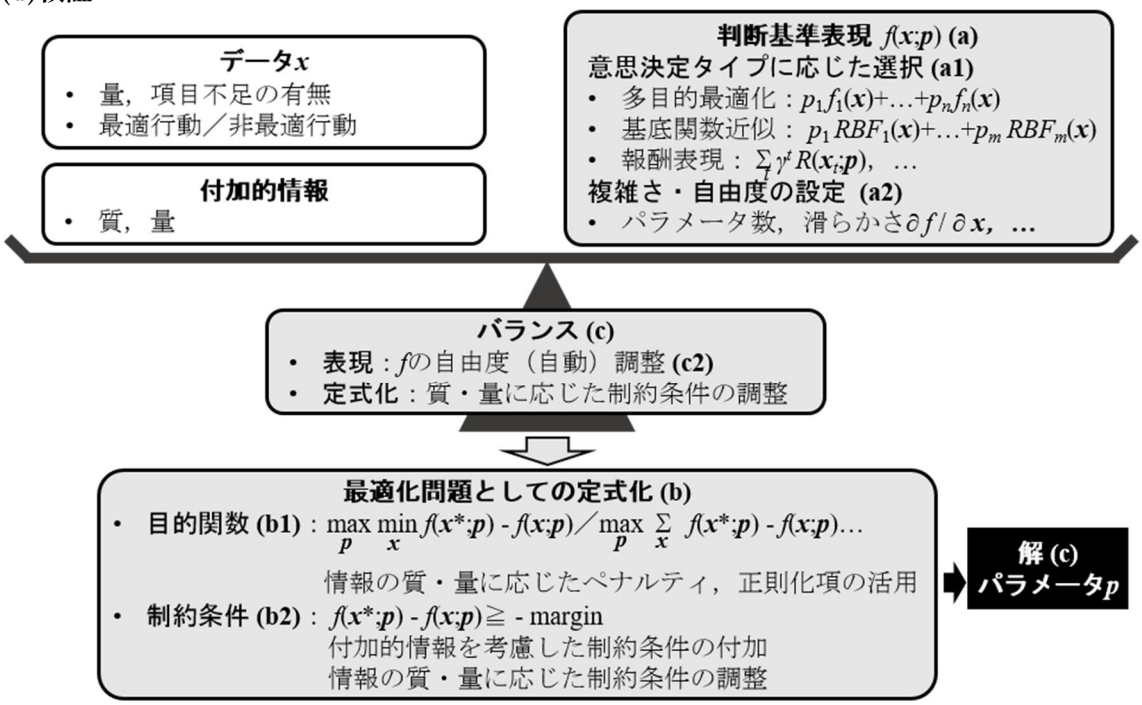

図２　全体像

(a)では判断基準 f(x;p)の関数形を，(a1)対象とする意思決定問題のタイプ，(a2)入手した情報の質・量，に応じて決定する．(a1)に関しては，判断基準が一つか複数か（単目的か多目的か）など問題タイプに応じた表現形式を考案する．(a2)では，入手データの量，対象者の行動が最適と想定されるか，付加的情報の利用可能性，に依存して定まる入手情報の質・量に応じて，関数形 f の複雑さを適切に抑制する（図２の「判断基準表現」部分参照）．

(b)では，不等式条件 f(x*;p)≧f(x;p)を満たす p は無数にあるため，解を一意に定めるために最適化問題として定式化する．その適切な(b1)目的関数と(b2)制約条件の決定を行う．(b1)では，意思決定問題のタイプに応じた目的関数とする他，入手情報の質・量に応じた不等式条件の「強さ」の調整を活用する．(b2)では，基本の制約条件 f(x*;p)≧f(x;p)に加え，データ以外の付加的情報も考慮して適切な f(x;p)の表現自由度が得られる制約条件を定める（図２の「最適化問題としての定式化」部分参照）．

(c)では(a)と(b)の結果として得られる最適化問題の解法を開発する．特に，情報の質・量に合わせた自由度バランスの維持を可能とする．この実現のために，f(x;p)の関数形の複雑さを自動決定する（図２の「バランス」部分参照）．

(d)では条件設定が容易なシミュレーションデータと現実的なデータを用いて検証を行う．後者として車の運転状況のデータを利用する．

４．研究成果

(1)研究内容(a)「判断基準の表現」に関する成果

人の行動あるいは判断が１回で完結する静的な場合と，変化する状況に応じて複数回繰り返して行われる動的な場合とに分けて適切な表現方法を考案した．

複数の選択肢の中から最良のものを選ぶ場合，行動あるいは判断は１回で完結する．選択肢 x の良さが複数の観点 f1，...，fn から評価されるとすると，判断基準は f(x;p) = p1 f1(x)+ … + pn fn(x) と（近似的に）表すことができる［表現１］．ここで，f1(x)から fn(x)は選択肢 x を異なる観点それぞれで評価した結果である．これは，購入する電気製品を価格，性能，大きさなどに基づいて決める場合などが該当する．

また，判断基準が複数個あり，人による選択肢比較結果が得られる場合の典型例である対話型

進化計算は，人と計算機が対話を行いながら最適化を行う方法である．この対話型進化計算の過程で得られる情報を基に，複数個の判断基準のうちのどれを重視して判断を行っているのかを，f(x;p)の関数形を明示的に規定せずに推定する方法を開発した［表現２］．

一方，複数回繰り返し行動あるいは判断が行われる動的な場合，x はこれら複数回をまとめて表現するベクトル **x** になる．また，**x** を評価する判断基準は毎回の行動 x1，...，xT の判断結果の総和になる．すなわち，F(**x**;p) = f(x1;p) + … + f(xT;p) と表される．ここで T は行動の総数である．人は F(**x**;p)が最良の値になるように行動の系列 **x** を決めている．f(xt;p),t=1,…,T は強化学習では報酬と呼ばれ，F(**x**;p)は価値と呼ばれる．判断基準の推定は報酬の推定となる．これは逆強化学習問題と呼ばれる．逆強化学習によって推定する報酬の表現として次の２つを考案した．

対象とする報酬 f(xt;p)に関する情報が得られていない場合は，報酬を表現する関数の形 f(xt;p)を特定のものとすることはできず，一般性の高いものとする必要がある．そこで，複数のガウス関数型の基底関数 ri(xt)に係数を乗じたものの和 p1 r1(xt) + … + pn rn(xt)による表現を採用することとした．これは，小さな山型のグラフで表される関数 ri(x)を n 個用意し，各々の高さ pi を適切に設定して，その総和で任意の曲面を表現することに相当する［表現３］．

一方，判断基準に関する情報が得られている場合は，それを活用した関数形を採用し，それに含まれるパラメータを推定する．たとえば，車を運転する際に障害物との距離が判断基準に影響を与え，かつ，距離が長いほど良いと判断されることがわかっているが，距離に反比例するのか距離の 2 乗に反比例するのかまではわかっていない場合，時刻 t での障害物との距離を g(xt)とすると，判断基準を p1 1/g(xt) + p2 1/(g(xt)^2) + p3 1/(g(xt)^3)などと表すことができる．ここで，a^2 は a の 2 乗を表す．係数 p1,p2,p3 の値は逆強化学習によって求めることができる［表現４］．

(2)研究内容(b)逆最適化問題の定式化に関する成果

上記の［表現３］に含まれるパラメータ p1,…,pn を求める逆最適化問題を，パラメータの絶対値の和を，不等式条件 f(x*;p)≧f(x;p)+ε を制約条件として最小化する最適化問題として定式化した．各パラメータ pi の絶対値の和の最小化により，値がゼロに近くなる pi の個数をできるだけ増やすことによって，表現の自由度が過大になることを抑制し，また，判断基準がシンプルで解釈がしやすいものとなるようにした．制約条件中の ε の値は，この不等式が成り立つ強さをコントロールし，データの量と表現の自由度のバランスを調整する働きを持つ．

一方，人は常に最適な行動をするとは限らない．最適から少し外れた行動も行う可能性がある．同様の状況下で人間の行動を複数回観察すると，最適な行動は高い頻度で，最適から外れた行動はそれよりも低い頻度で観察される．したがって，この行動の頻度はその行動の良さの指標となる．上記の［表現４］を対象として，この行動頻度によって推測できる行動の良さを付加的な情報として活用した問題の定式化を考案した．

(3)研究内容(c)具体的な解法の開発についての成果

［表現１］の形式で表現される判断基準の推定は，異なる観点 fi と fj のどちらが重要な観点であるかについての人の判断という主観的情報と，fi(x)の値の x によるばらつきという客観的情報の両方を用いて行うことができる．主観的情報からは Analytical Hierarchical Network [6]などの方法を用いてパラメータ pi の値を求めることができる．客観的情報である fi(x)のばらつきが小さいと，これは，x が異なっても fi(x)の値はあまり変化しないことを表し，fi(x)は x の良さの識別に役立たないことを意味するので，対応する係数 pi は小さくすべきである．ばらつきの大きさをエントロピーで表すことによって，この定量的情報に基づく pi の値を定量的に定めることができる．これら主観的情報から得た pi と客観的情報から定めた pi を適切に合成して総合的な pi を求める方法を提案した．この方法では，主観的情報を複数人から入手し，その各々から pi の値を求める．得られた複数の pi の値の標準偏差が小さければ，複数人の意見は一致しており，主観的情報の信頼度は高いと判断されるため，主観的情報から得た pi の寄与率を高くして客観的情報から得た pi の値と合成する．この標準偏差は観点の総数 n に依存して変化するため，この依存性を排除して寄与率を算出する方法を開発した．

［表現２］は対話型進化計算を対象とする．対話型進化計算では選択肢の評価を人が行う．その際の人の疲労を抑制するために，選択肢の絶対評価ではなく，二つの選択肢のいずれが良いかを人が回答する対比較による相対評価が多く用いられる．多数の選択肢について対比較を繰り返し，良いと判断された選択肢から新しい選択肢を生成して，再度評価を行う．これを繰り返すことによって，多数の選択肢が最適なものに近づいていく．各選択肢が複数の値を集めたベクトルで構成されている場合，たとえば，好みのシャツを見つける際に，あるシャツがその形，色，柄などで規定されている場合を考える．このとき，ある座標軸（たとえば色の軸）に沿って，選択肢が類似した値を持つとき，その座標軸（色）が好みに大きく影響しており，値が散らばっているとき，その座標軸（色）に対応する値は好みに影響しないと判断される．この集中・分散の程度から，どの値が重要視されているかを判定する方法を開発した．

［表現３］は，複数のガウス関数型の基底関数 ri(x)に係数を乗じたものの和 p1 r1(x) + … + pn rn(x)による表現であり，ガウス関数型の基底関数の個数，位置，拡がり，高さを，判断基

準推定に用いるデータに応じて適切に決める必要がある．これを行う方法を提案した．このうち，高さは逆強化学習によって決定されるが，それに先立って，他のパラメータを決定する必要がある．逆強化学習で推定した報酬を用いて，人の行動を模擬的に予測した際の誤差が最小となるように，これらを決定する方法を開発した．これは多くの計算量を要するので，個数の決定と位置および拡がりの決定の２段階に分けて決定することとし，さらに，適切な位置と拡がりの関係を導入してベイズ最適化手法を活用することにより，計算量の削減を行った．

［表現４］による判断基準を求めるために，同様の状況下で人間の行動を複数回観察した際に得られた結果を活用する方法を提案した．複数回行動を観察すると，最適な行動は高い頻度で，最適から外れた行動はそれよりも低い頻度で観察される．したがって，この行動の頻度はその行動の良さの指標となる．ここでは，統計力学的な考えに基づき，人は行動の良さの指数関数に比例した確率で行動を選び，実行すると想定した．この想定が正しいとすると，２種類の異なる行動の良さの差は，観測された各行動の実行頻度の対数の差に等しくなることから，これを制約条件とし，表現の自由度抑制に貢献するパラメータの絶対値和を最小にする最適化問題を解くことによって，パラメータを推定する方法を提案した．

(4)研究内容(d)検証についての成果

上記の方法を，対象とする問題の特徴を持つ人工的な例題，電力の消費量の変更を誘導するデマンドレスポンスの評価およびデマンドレスポンスに参加している消費者の判断基準の推定，さらに，実際の自動車運転データからの運転者の評価基準の推定に適用し，各方法の有効性を確認した．

＜引用文献＞

[1] R.K.Ahuja, J.B.Orlin, “Inverse Optimization”, Operations Research, Vo.49, No.5, 771-783 (2001).
[2] A.Y.Ng, S.Russel, “Algorithms for Inverse Reinforcement Learning”, Proc. The 17th Int. Conf. on Machine Learning, 663-670 (2000).
[3] R.Syed and R.E.Schapire, “A Game-Theoretic Approach to Apprenticeship Learning”, Advances in Neural Information Processing Systems 20, 1449-1456 (2008).
[4] P.Abbeel, D.Dolgov, A.Y.Ng, S.Thrun, “Apprenticeship Learning for Motion Planning with Application to Parking Lot Navigation”, Proc. 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2008).
[5] 古賀，奥田，田崎，鈴木，原口，康，「運転個性を反映したモデル予測型自動運転システム-評価関数推定手法の提案-」，自動車技術会論文集， 47 巻，6 号，1431-1437（2016）
[6] T.L.Saaty, L.G.Vargas, “Decision making with the analytic network process. Economic, political, social and technological applications with benefits, opportunities, costs and risks”, Springer US (2006).

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　0件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 村田 純一，船木 亮平	４．巻 59
２．論文標題 目的関数の推定：循環・螺旋型システムズアプローチに資するモデリングと意思決定過程把握	５．発行年 2020年
３．雑誌名 計測と制御	６．最初と最後の頁 918-921
掲載論文のDOI（デジタルオブジェクト識別子） 10.11499/sicejl.59.918	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計7件（うち招待講演　1件／うち国際学会　0件）

１．発表者名 内山 海渡，船木 亮平，村田 純一
２．発表標題 逆強化学習における報酬関数を表す基底関数群のベイズ最適化による決定
３．学会等名 計測自動制御学会 システム・情報部門 学術講演会 2021
４．発表年 2021年

１．発表者名 徐　士豪，船木 亮平，村田 純一
２．発表標題 逆強化学習によるコスト関数推定に基づく自動運転の軌道最適化
３．学会等名 計測自動制御学会 システム・情報部門 学術講演会 2021
４．発表年 2021年

１．発表者名 村田 純一，船木 亮平
２．発表標題 人間の判断基準の推定
３．学会等名 日本オペレーションズリサーチ学会「エネルギーシステムの進化とOR」研究部会 第5回研究会（招待講演）
４．発表年 2021年

１．発表者名
甲斐田 幸希，船木 亮平，村田 純一

２．発表標題
行動頻度から推定した状態価値の相対的関係を活用した逆強化学習

３．学会等名
計測自動制御学会 システム・情報部門 学術講演会 2020

４．発表年
2020年

１．発表者名
Chen Zhang, Ryohei Funaki, Junichi Murata

２．発表標題
Evaluation Model for Demand Response Based on Integrated ANP-Entropy Method

３．学会等名
計測自動制御学会 システム・情報部門 学術講演会 2020

４．発表年
2020年

１．発表者名
杉本 顕武郎，船木 亮平，村田 純一

２．発表標題
対話型進化計算における決定変数が評価に与える影響の個体群エントロピーを用いた推定

３．学会等名
進化計算シンポジウム2019

４．発表年
2019年

１．発表者名
酒井 優也，船木 亮平，村田 純一

２．発表標題
対比較ベース対話型差分進化における優劣拮抗個体の探索への活用

３．学会等名
進化計算シンポジウム2019

４．発表年
2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究協力者	船木　亮平 (Funaki Ryohei) (60775287)	九州大学・大学院システム情報科学研究院・助教 (17102)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関