類似タスクにおける経験に基づいた探索アルゴリズム

—人工データを用いた探索性能評価—

唐澤 宏之*・福井 類*・割澤 伸一*

Search Algorithm Using Prior Knowledge of Similar Task

—Evaluation of Search Ability Using Toy Problem—

Hiroyuki KARASAWA*, Rui FUKUI* and Shin'ichi WARISAWA*

This study introduces a concept of transfer learning to search tasks such as function approximation and optimization, and aims to achieve effective search using fewer number of samples. We propose a search procedure transfer (SPT) algorithm which extracts knowledge of efficient search procedure from well-known task and uses it to search similar unknown tasks. Experiments reveal that the closer the target task becomes to the source task, the higher the performance of the SPT algorithm becomes. In addition, experiments also demonstrate the SPT algorithm shows higher performance than the existing method using the Bayesian optimization algorithm (BOA) when the difference between a source task and a target task is small. Applying the idea of domain randomization, the SPT algorithm can utilize even human ambiguous heuristic knowledge.

Key Words: design of experiments, transfer learning, Bayesian optimization, search strategy

1. 緒言

近年、製品の設計¹⁾から顧客への販促戦略²⁾まで幅広い分野において、計算機による人間の判断の補助が欠かせない存在となっている。特に注目されているのが機械学習によるデータ駆動的アプローチであり、そこでは過去に取得されたデータから知見を抽出し予測に用いることによって、さまざまな因果関係が交錯する複雑な対象に対する高い精度での未来予測が可能となった。すでに、一部のタスクにおいては、計算機による判断が人間の能力を超えることが示されている。計算機上のアルゴリズムはタスク特化型ではあるものの、一人の人間には考慮しきれない量の蓄積されたデータを考慮したり³⁾、数多くのシミュレーションを行ない知識を抽出する⁴⁾ことにより、人間を超える能力を獲得している。

データ駆動的アプローチが華々しい成功を挙げているタスクがある一方で、多くのタスクではデータ取得コストが膨大になるという問題に直面している。たとえば多自由度マニピュレータによる物品の把持タスクに関する研究⁵⁾では、アルゴリズムの能力が成熟しきるまでに14台のマニピュレータによる2ヵ月間の訓練データ収集を必要とした。このような実機械を用いたデータ収集は時間や動作コスト、故障リスクな

どを考えると、広く社会に行なわれるようになるにはまだ手法面での改善が必要である。

データ取得コストを低減するために、転移学習⁶⁾に注目が集まっている。転移学習とは、あるタスク（ソースタスク）で獲得した知識を転用し、新しいタスク（ターゲットタスク）をこなすために用いる手法である。タスク間で共通な知識を抽出することができれば、新しいタスクにおけるデータ取得の手間を削減することが可能である。先述のようなマニピュレータの制御タスクでは、データ取得コストが小さい物理シミュレータ内のソースタスクにおいて訓練を行ない、知識を転用することにより現実世界におけるマニピュレータの訓練回数を低減することに成功している⁷⁾。

われわれが直面する多くの問題において、現在の状態と未来に起こる現象との間には因果関係が存在する。本論文では現在の観測できる状態を x 、未来に起こる現象を y と定義し、その間の因果関係を関数 $f(\cdot)$ で表現する。データ駆動的予測アルゴリズム目標の主たるものとして、関数近似と最適化がある。関数近似は状態 x と現象 y の間の関数 $f(\cdot)$ を近似するもので、最適化は x が変更可能な場合において、関数の出力 $y = f(x)$ を最大化するような x を求める問題といい表わせる。

これらの背景から、あるタスクに関して過去の取得したデータがない状態から、少ない手数、つまり少ない実験コストで高精度な関数近似や高精度な最適値の提示を行なうことが求められているといえる。その要求を受け、本研究では関数近似・最適化の問題に対して転移学習の考え方を導入し、少な

* 東京大学大学院新領域創成科学研究科 柏市柏の葉 5-1-5

* Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa

(Received February 14, 2019)

(Revised June 28, 2019)

い実験回数で目標を達成することを目指す。また本論文ではこれらの課題を総称して探索課題と呼称する。

本論文の貢献は以下のとおりである。

- 探索課題において、探索手順の転移という形で知識転移手法を提案し、実験により手法の有効性や特性を明らかにする。
- 経験則的な法則があいまいに得られている探索課題において、経験則的な法則を取り入れることで探索効率を上げられるような探索手法を提案する。

本論文の構成は以下のとおりである。まず、2章において、従来から存在する探索手法および転移学習手法の導入を行なう。続く3章では本研究の提案する探索手順の転移アルゴリズムを述べ、4章においては本論文内での検証時の具体的実装について述べる。5章で複数の探索課題に対して提案アルゴリズムの実証実験の結果を述べる。最後の6章は結論である。

2. 関連研究

2.1 人間による探索課題へのアプローチ

少ない手順で多くの情報を得る手法群は実験計画法と呼ばれており、さまざまな場面で用いられている。以下に示す2手法⁸⁾は人間が行なう実験計画法の代表的なものである。

- タグチメソッド：直交表を用いて感度を解析する。
- 一時一事法：多くの変数により現象が決まる系において、1度に1つのみ変数を変更して各変数の感度解析や、各変数と現象との間の詳細なメカニズムを考察する。

自然界の現象はエネルギー保存則や遺伝の法則などいくつかの基本的な法則に従って発生する。人間はさまざまな計測から背後の法則を理解したり、法則と法則をつなぎ合わせて新たな法則を生み出したりすることができるため、上記のような実験計画法により、要素還元的に法則を理解し一定の成功を挙げてきた。現状、人間が計算機上のアルゴリズムに対してもつ優位な特性として、メタな思考がある。メタな思考とは、法則を探し出す手順に対しても何らかの法則を見つけ出すといったような、与えられた問題の枠にとらわれない思考である。人間はこの思考法によって、つぎつぎと新しい発明を行なうことができた。

2.2 機械による探索課題へのアプローチ

実験の手続きが複雑ではなく、計算機の性能を活かしたい探索課題に対しては、自動的に探索を行なうアルゴリズムが開発されてきた。ただ、人間の手を介さず自動的に実験を行なうといえども、グリッドサーチのような愚直な探索方法ではコストに見合わないような探索課題が多く存在している。たとえば、時間や金銭のコストがかかる化学物質の合成パラメータ探索や、顧客に対して有効な広告提示方法の探索などがそれにあたる。

少ない回数の実験で最適値を見つけ出す手法の代表的なものとして、ベイズ最適化 (Bayesian Optimization Algorithm: BOA) がある。ベイズ最適化とは多くの解候補の中から適切な解候補を選択する操作と、解候補を探索して得られた情報

を用いて解候補の選択に用いる知識を得る操作とを交互に繰り返すことにより、少ない実験回数で最適値を見つけ出すことを目指した手法である。

BOAの代表的な実装として、ガウス過程回帰 (Gaussian Process Regression: GPR)⁹⁾を用いたものがある (GPR-BOA)¹⁰⁾。ガウス過程回帰とはカーネル法を用いることで、複雑なモデルの近似が可能な回帰手法であり、近似に用いる関数群を選ぶステップを内部にもつことによって近似式の滑らかさを担保可能な手法である。また、通常の回帰手法と異なり、推定値のみならず、推定値の分散も同時に求まる。この性質を利用して、BOAの解候補の選択時に分散の大きな点を優先的に訪れることにより効果的な関数近似を行なうことができる。一方で、推定値が高い点を優先的に訪れることにより獲得報酬を高くすることもできる。

GPR-BOAのようなBOAの手法は一定の成果を上げているものの、ある探索課題に取り組んだ際に得られた知見を別の探索に用いることができていない。Fig. 1に示すように、ある1つの探索課題によって得られた知識を別の探索課題に用いることで効果的な探索が行なえる可能性がある。身近な例でたとえると、人間は機械の故障箇所を同定するときにまず点検すべき箇所の知識をもっており、機械のマニュアルにはそのような知識が掲載されている。従来のBOAの手法では、このような探索の定石に関する知識を利用することができない。

2.3 知識の転移手法

データ駆動的アプローチにおいては、なるべく抽象的な知識を抽出し、複数の課題に共通な知識を取り出す転移学習やメタ学習といった分野の研究が盛んである⁶⁾。抽象的な知識は多くの課題において成立するため、各課題を解く際の事前知識として用いることができ、学習コストの低減につながる。先述のような、シミュレーションにおいてマニピュレータの把持方法を学び現実世界に転用する手法⁷⁾や、画像認識に共通して有用な特徴量抽出方法をローカルな画像認識課題に転用する手法¹¹⁾などが存在する。

これらの手法の目標は、与えられた入力 (観測値, 画像) に対して適した出力 (つぎの動作, クラスラベル) を返すような関数を近似することである。一方で、このような関数近似のためのデータ収集手法に関して、メタ的に知識を抽出してタスクの枠を超えて利用するという研究は少ない。

Fig. 1 Two types of time axes

以上のような背景から、本論文では探索課題において効率的な探索手法に関する知識を抽出し、類似の探索課題に用いる手法の開発を目指す。実験においてはデータの用意のしやすさを考慮し仮想のタスクを設定し、提案手法のふるまいの考察を行なう。

3. Search Procedure Transfer Algorithm の提案

3.1 問題設定

本論文では、探索課題において、関数近似誤差を小さく、探索時の平均獲得報酬を大きくすることを目標とする。一般的な最適化問題においては後者の獲得報酬最大化を主要な目標とし、関数近似のための探索というのは獲得報酬最大化の一手段に過ぎない。本論文では以下の2つの理由から関数近似誤差最小化も獲得報酬最大化と並ぶ目標と設定している。

1つ目の理由は、関数全体の形状を知っておくことによる評価関数変動時の対応の柔軟性にある。たとえば、機械加工においては、表面精度や1ワークあたりの加工速度、工具への負荷など複数の評価基準のもと評価基準の合計値が最も高くなる点が最適加工条件とされる。しかし、加工の需要が下がりワークの加工速度に対する優先度が下がるようなことが起こると、評価基準の合計が最大値となる最適加工条件は変動する。この際に、全体の関数形状を知っておけば最適加工条件更新の対応が容易となる。

2つ目は、人間によるアルゴリズムの動作チェックが容易になるというものである。データ駆動的アプローチが社会で用いられるようになるにつれて、人間とは異なる基準で判断を行なうこともあるアルゴリズムの判断¹²⁾を信用してよいのかという問題が盛んに議論されるようになってきた。関数近似を行なっていれば、計算機上の近似関数を用いることで、人間が関数への入力を変えてみてシミュレーションを行なうことができる。このシミュレーションにより、過学習が発生していないか、人間の経験則に反した予測をしていないかということを人間が直感的に確認することができる。

3.2 タスクの定義

本論文では、アルゴリズムが解くタスクとして関数の探索課題を設定している。Fig. 2のようにソースタスクを既知の関数、ターゲットタスクを形状が未知の関数とする。本手法の利用想定として、人間がタスクの類似性を判断しやすい問題を想定しており、具体的な適用先としては、回路設計¹³⁾におけるパラメータチューニング、液体流路の設計最適化など、人間がある程度背後の物理現象を理解しているが、高速なシミュレーションが難しいような問題を想定している。また、本手法はターゲットタスクのデータ取得コストが高く、ソースタスクのデータ取得コストが低いときに効果を発揮することを期待した手法である。たとえば、目標が調整パラメータが10個ある機械のパラメータ最適化であり、人間が各パラメータの大まかなふるまいがわかっているときに用いることを想定している。そのような場合、ターゲットタスクは目標

Fig. 2 Concept of knowledge transfer

Fig. 3 Domains in single variable task

としている機械の評価関数の探索であり、ソースタスクには人間がもっている各パラメータの感度や影響度合いの知識を関数化したものを設定する。本手法はFig. 2のように関数形が既知で高速な計算が可能なソースタスクの探索を行ない、探索に関する知識を抽出し、ターゲットタスクの探索を効率化することを目指している。Fig. 3は本論文の中で用いるソースタスク、ターゲットタスクの具体的な例を表わしている。Fig. 3では、どのタスク（関数）も1入力1出力の正弦波であるが、それぞれの振幅や位相が異なる。

3.3 探索知識の転移手法

データ取得コストが大きいタスクにおいて効率的な探索を行なうために、類似する既知のタスクから探索手順を転移し、解きたいタスクに用いる手法、Search Procedure Transfer Algorithmを提案する。以降、当手法をSPTと呼称する。

Fig. 4にSPTアルゴリズムの流れを示す。アルゴリズムは3つの大きなステップにわけられる。1つ目は探索手順生成ステップであり、何らかの指針に基づき訪れる探索点の手順を生成し、生成した探索手順と既知であるソースタスクを用いてシミュレーションを行なう。本論文の実装では関数の定義域の範囲で乱数を用いて生成する。つぎのステップ2では、探索を行なった結果と評価関数を元に、各手順の性能を評価する。評価関数には関数近似誤差 e と平均獲得報酬 r の2つの指標を用いる。どちらの指標の優先度を高くするかはユーザが設定する。どちらかの指標の重みを0とすることも可能であるが、手法の性質を考えると e に対する重みを残しておくことにより、転移した手順に沿って探索した後の追加のサンプリングで用いることのできる情報が豊かになることが期待される。以上のステップでは2種類のハイパーパラメータ N_{sample} , N_{cand} が存在する。 N_{sample} は探訪する点の数で

Fig. 4 Search procedure transfer algorithm

あり、ソースタスク、ターゲットタスクの両方で共通である。 N_{cand} はソースタスクで評価される探索手順候補の数であり、 N_{cand} が大きくなるにつれて手順の中の最高性能が高くなることが期待される。最後のステップ3では性能が高いと判定された探索手順を転移させ、ターゲットタスクを探索する。本論文の実装においては N_{cand} 個の手順の中から最も性能が高い手順を転移している。また、 N_{domain} はターゲットタスクの個数を表す。本論文の実験においては5を用いた。

今回のように探索に関するメタな知識を抽出することなく、ソースタスクの知識を利用してターゲットタスクの探索を効率化することも考えられる。たとえば、ソースタスクの関数の値に応じてターゲットタスクの探索優先度を決める方法があるが、ハイパーパラメータを増やさず関数近似と最適化を両立した探索を行なうことが難しい。

3.4 ソースタスクの設定

SPT アルゴリズムの実利用では、ソースタスクはターゲットタスクとなるべく似るように設定するのが良い。しかし、SPT アルゴリズムが用いられるような多くの問題ではターゲットタスクの関数形ははっきりとはわかっていないため、似るように設定することには困難が伴う。その困難への対策として、Domain randomization⁷⁾の考え方を導入する。Domain randomization は多自由度マニピュレータによる物体の把持をシミュレーション内で訓練させる際の工夫である。その背景として、カメラにより観測される現実世界の物体の色合いや影は変動しうが、その変動によって把持行為の本質はほぼ変化しないという事実がある。Domain randomization では色合いや影の変動は本質的に意味がないという事実をアルゴリズムに対し暗に教え込むために、シミュレータ内の物体の色や影の落とし方をランダムに変えて学習を進める。

本論文において SPT アルゴリズムの応用例として提案する、Domain randomization の概念を Fig. 5 に示す。ターゲットタスクの具体的な関数形はわからないが、大まかにあ

Fig. 5 Application of domain randomization

る特定の範囲 (Fig. 5 の破線内の部分) に入っているという知識は得られていると仮定する。機械の例で示すと、あるパラメータを増加させたときに性能が向上することが知られているが、線形な比例関係にあるのか、指数関数のような非線形な関係があるのかわかっていない場合などである。提案する手法では、その特定の範囲を取り囲むように複数のソースタスクを設定し、複数のソースタスクにおいて有効な手順を抽出する。上記の例では、あるパラメータ x_i が性能 a に与える影響を $a = bx_1 + c$, $a = bx_1^2 + c$, $a = be^{x_1} + c$ などとモデル化し、複数のソースタスクを設定することを想定している。そのようなソースタスクの設定により、ソースタスクによって囲まれるターゲットタスクの存在領域全体を通して有効な手順が抽出されることが期待される。つまり、人間が試行錯誤により経験則的に獲得した物理モデルが多少のあいまいさをもっていても、そのあいまいさに従ってソースタスクを設定することにより、一定の探索性能を示す手順の算出が可能になると期待できる。一方、先述のソースタスクの関数の値に応じてターゲットタスクの探索優先度を決める方法では、Domain randomization のような複数のタスクの情報を使うことが難しい。

4. 実装

この章では実験に用いる各タスクの生成方法や SPT アルゴリズムをはじめとした実験に用いるアルゴリズムの実装、探索性能の評価基準について述べる。

4.1 タスクの生成方法

多変数入力を受け付け、なめらかで多峰性のある関数を生成するために係数 a と関数セット $S_f = \{(ax)^2, 2^{ax}, \sin(ax), \cos(ax)\}$ の中から基本関数を変数の数だけ選択し、それらの総乗をとったものをソースタスクとした。ターゲットタスクはソースタスクの係数ベクトルを変形して生成した。なお、ここでの変形とはベクトルの要素に対してある実数 (本論文では 1.1) をべき乗した値を掛け合わせるものであり、べき指数は平均 0、分散 1 の正規乱数発生器の出力を用いる。本論文では、実験のために 1 つのソースタスクに対して 5 つのターゲットタスクを生成した。また、論文内においてはソースタスクとターゲットタスクを合わせてタスクセットと呼ぶ。関数生成にあたり、関数同士の類似性を示す値である距離 d という値を定義した。関数同士の距離 d は 2 つの関数値の差を

$\mathbf{x}$ の定義域全体において足し合わせたものであり, (1) 式のように定義される. 5 つ目のターゲットタスク (以降 #5 と表記) とソースタスクとの距離が既定の値になるように, 係数ベクトルの変形量であるべき指数を調整し #5 を生成し, #1 ~ #4 の各タスクに関しては # の数が 1 小さくなるにつれて d が半分になるように係数ベクトルを変形して生成する. 2 変数関数, 既定の $d = 0.16$ において生成された関数の例を **Table 1** に示す. また, Fig. 3 は本節の生成方法とは係数の与え方が異なるものの, 1 変数関数のタスクセットを図示したものである.

$$d = \int_D |g(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} \quad (1)$$

D : domain of a function, f, g : functions

4.2 アルゴリズム実装

SPT アルゴリズムの手順生成器には乱数発生器を用いた. SPT アルゴリズムでは #0 をソースタスク, #1 ~ #5 をターゲットタスクと設定した. 比較するアルゴリズムとして, SPT-best, SPT-worst, GPR-BOA の報酬優先, GPR-BOA 近似優先, ランダム探索の 5 つを設定した. SPT-best と SPT-worst はそれぞれソースタスクで最も性能が高かった手順と低かった手順を転移するものである. 報酬優先 (BOA-exploit) はベイズ的に探索を行なう際に, つぎの点で得られる報酬の期待値が最も高くなるように探索点を選ぶ方法, 近似優先 (BOA-explore) は近似関数の分散が大きい点を探索点として選ぶ方法である. BOA においては報酬優先と近似優先をバランスさせることも考えられるが, 今回の実験設定においては付録 A に示すように BOA の性能に大きな影響は与えなかったため, 比較においては単純な報酬優先, 近似優先のみを取り上げる. どのアルゴリズムにおいても関数近似には GPR を用いており, GPR のカーネルには放射基底関数 (RBF) を用いている. GPR には RBF の場合のみならずカーネルの設定値がハイパーパラメータとして存在する. ある問題において高い性能を示したアルゴリズムはほかの問題を解かせると低い性能を示し, 万能な最適化アルゴリズムは存在しない (No Free Lunch Theorem¹⁴⁾) といわれており, 最適化アルゴリズムの開発時は, パラメータが手法の性能に及ぼす感度の考察が必要とされる. 一方で, SPT アルゴリズムは, アルゴリズムの利用想定を限定しており, ある単一のタスクを解くためにほかのタスクで得た知識を転用するという戦略をとっているため, 一般的な最適化手法とは用いる情報が異なる. SPT アルゴリズムにおいて, GPR のパラメータはタスク間で転用

される情報に対して質的な影響を及ぼさないため, 本研究においては GPR のパラメータの変動が最終的な性能に及ぼす影響の考察は割愛する.

4.3 性能評価方法

探索手順の性能評価および, 探索の性能評価には Fig. 4 中段および, (2) 式に示すような近似誤差 e , 平均報酬 r の線形和を用いた. この性能評価指標を Search score と呼ぶ. 本論文では (2) 式中の w_1, w_2 にはそれぞれ $-1, 0.5$ を用いた.

$$\text{Search score } s = w_1 e + w_2 r \quad (2)$$

誤差や報酬に関してはタスクごとの比較を行ないやすくするため, 算出の際に関数の値域が $[0, 1]$ になるように値を正規化している. 近似誤差の算出では, 本章で述べた 2 つの関数同士の誤差算出関数式 (1) を用いて, 実際の関数と近似された関数の距離を求め, 近似誤差とした.

5. 実験

本章では実施した 3 つの実験の設定および結果について述べる. 実験の最も基本的な目標は探索手順の転移により探索が効率化されることの確認である. 1 つ目の実験は Fig. 3 に示す 1 変数関数のタスクセットを用いた実験であり, 探索評価基準の妥当性や手順転移の有効性を示す. 2 つ目の実験は 4.1 節で述べた生成方法によって生成した多変数関数を用いた実験であり, 提案手法が用いられるような探索課題における手法の有効性を示す. 3 つ目の実験は複数のソースタスクを用いた実験であり, Domain randomization の応用のための予備実験である.

5.1 1 変数関数を用いた実験

1 つ目の実験では, Fig. 3 における #0 をソースタスクとして有効な探索手順の抽出を行ない, 探索手順を #1 ~ #5 のターゲットタスクの探索へ転用した. なお, ハイパーパラメータ 2 種, 実験試行回数はつぎのように設定した. 探索点数 $N_{\text{sample}} = 3$, 手順候補数 $N_{\text{cand}} = 100$, 試行数 $N_{\text{trial}} = 10$.

Fig. 6 はソースタスクにおいて最も性能が高いと判定された探索手順 (Fig. 6 (a)), 最も性能が低いと判定された探索手順 (Fig. 6 (b)) を示す. (a) は関数値の最高値に近い 3 点を探訪しているのに対し, (b) は最低値に近い 3 点を探訪している. また, (b) に比べ (a) は関数近似 (Fig. 6 中破線) が上手くいっていることがわかる.

Fig. 7 のグラフはターゲットタスクにて探索を行なった際

Table 1 Generated functions in two variable task

Domain #0 (source)	: $\sin(4.37x_0) \times (1.74x_1)^2$
Domain #1 (target)	: $\sin(4.31x_0) \times (1.76x_1)^2$
Domain #2 (target)	: $\sin(4.24x_0) \times (1.77x_1)^2$
Domain #3 (target)	: $\sin(4.11x_0) \times (1.81x_1)^2$
Domain #4 (target)	: $\sin(3.86x_0) \times (1.87x_1)^2$
Domain #5 (target)	: $\sin(3.41x_0) \times (2.01x_1)^2$

Fig. 6 Procedure examples

Fig. 7 Search score of each algorithm (single variable)

Fig. 8 Search score of each algorithm (multivariable)

の各アルゴリズムの性能を示す。横軸がソースタスクとの距離 d を表わし、縦軸が近似誤差と平均報酬の線形和の探索性能を表わす。図中の線は試行の平均値、パターンが描かれた領域は分散の大きさを表わす。

ソースタスクからの距離が近い4つのタスクではSPT-bestが最高性能を示した。一番遠い $d = 0.30$ のタスクではBOA-exploitが最高性能を上げている。一方、SPT-worstはどの条件においても最も性能が悪く、想定どおりの動作をしている。また、ソースタスクから近いタスク #1, #2 ではSPT-bestの性能の分散が小さく、手順の転移の性能が安定しているといえる。以上のことから1次元の正弦波の探索課題において、関数同士の距離が $d = 0.16$ 以下の条件では探索手順の転移の有効性があるといえる。

5.2 多変数関数を用いた実験

2つ目の実験では4.1節で述べた生成方法によって生成された2変数関数、5変数関数のタスクセットを用いて前節のような実験を行なった。5変数関数の場合もTable 1と同様に4.1節にて示した関数セットの中から変数の数と同じ5個の関数を選び5変数関数を構成した。ソースタスクと #5 のターゲットタスクとの距離は2変数では0.16、5変数では0.12と設定した。そのほかのハイパーパラメータ2種、実験試行回数2種はつぎのように設定した。探索点数は2変数と5変数で、 $N_{\text{sample}} = 3, 5$, 手順候補数 $N_{\text{cand}} = 100$, タスクセットの数 $N_{\text{set}} = 10$, セットごとの試行数 $N_{\text{trial}} = 10$ 。

Fig. 8のグラフにターゲットタスクにて探索を行なった結果を示す。Fig. 7と同じく、横軸はソースタスクからの距離、縦軸は探索性能を表わす。1変数関数の場合と同様にSPT-best

Fig. 9 Effect on number of candidates

が最も高い探索性能を示し、探索性能はソースタスクとの距離が遠いタスクであればあるほど低下した。2次元、5次元の探索課題において、関数同士の距離が $d = 0.16$, $d = 0.12$ 以下の条件では探索手順の転移の有効性があるといえる。

Fig. 9は生成する手順候補数 N_{cand} と転移後の探索性能の関係を示す。2変数関数を利用し、先ほどの実験条件の N_{cand} のみを変動させた。Fig. 9に示す性能は、前述の実験においてSPT-bestに次いで性能が高くなったBOA-exploitの性能を用いてタスクごとに正規化を施している。Fig. 9の点線右側が各 N_{cand} における性能を示しており、生成する手順の候補数 N_{cand} を増やすことで、探索性能を向上させることができるといえる。

一方、関数の変数の数や探索点数 N_{sample} が増えるに従い、探索手順に関する探索空間の大きさが指数的に増加することから、多変数における手順の生成には、生成した手順と探索性能の対応から手順の良し悪しを決める要素を抽出するなどの対策が必要になると考えられる。

5.3 複数のソースタスクを利用する実験

Domain randomization (Fig. 5)の応用では探索したいターゲットタスクの大体の予測をつけ、ターゲットタスクの周囲に複数のソースタスクを配置することにより、ターゲットタスクにおいて有効な探索手順の抽出を狙う。この実験では複数のソースタスクを利用した際に、ソースタスク間の領域全体に対して有効な手順を抽出できるかを確認する。

Fig. 10は比較を行なう2手法を表わしたものである。Fig. 10上SPT-singleが1つのソースタスクを用いるもの、Fig. 10下SPT-multiが2つのソースタスクを用いるものである。Fig. 10中の横軸は係数ベクトルの変形量を表わしており、四角と三角はそれぞれソースタスクとターゲットタスクを表わす。また、4.1節にあるように、ターゲットタスク #1 から番号が1増えるにつれて距離 d が倍増していく。実験には2変数関数を用い、各種パラメータはつぎのように設定した。 $d = 0.16$, $N_{\text{sample}} = 3$, $N_{\text{cand}} = 100$, $N_{\text{set}} = 10$, $N_{\text{trial}} = 5$ 。

Fig. 11に結果を示す。複数ソースタスクを利用した場合のSPT-multi (図中実線)は、全体を通してBOA-exploitやSPT-singleの最低スコアよりも高いスコアになっており、複

Fig. 10 Setting of multi-domain experiment

Fig. 11 Result of domain randomization

数のターゲットタスク間で安定した成績を示している。SPT-multiはソースタスクとして#0, #5を用いているため、その間で安定してスコアが高いが、#0に近い#1~#3では#0のみをソースタスクとしたSPT-singleの性能が優っている。複数のソースタスクを設定し、ターゲットタスクを挟みこむことで、間のターゲットタスクに対して安定して高い性能を示す手順を抽出できているといえる。

6. 結 言

本研究は、関数近似・最適化といった探索課題に対して転移学習の考え方を導入し、少ない手数で効果的な探索を達成することを目標とした。目標達成のため、探索課題における効率的な探索手順に関する知識を既知の課題から抽出し、類似の探索課題に用いるという手法であるSPTアルゴリズムを提案した。

実験から、1, 2, 5変数関数において転移する先の関数であるターゲットタスクが転移元のソースタスクから近いほど探索手順の転移(SPT)の性能が高くなること、および、1変数関数、2変数関数のターゲットタスクが関数形状の乖離度を表わす d が $d = 0.16$ より小さい場合において探索手順の転移(SPT)がベイズ最適化(BOA)を用いた既存手法に比べて高い性能を示すことが明らかになった。また、SPTアルゴリズムに、Domain randomizationの考え方を適用し、人間のもつ経験則的な物理モデルのあいまいさを許容する探索手順抽出手法を提案した。Domain randomizationにより複数ソースタスクに囲まれた領域内のターゲットタスクに対して有効な探索手順の抽出が可能であることが実験により示唆された。

本論文は探索手順の転移の基本的なアルゴリズムの提案と人工データを用いた実証実験を行なった。複数タスクから抽出される探索手順はタスク間の乖離度が小さいほどそれらのタスクに特化したものとなる。一方でタスク間の乖離度が大きいときは探索空間全域の均一幅の探索など一般的な探索手順が抽出されると考えられる。今後はタスクの乖離度が抽出される探索手順に与える影響の確認や、手順の候補の効率的な生成手法の開発を行なう。また、データ取得コストが課題となっている現実世界の課題に対して提案手法の効果検証を行なう。

謝辞 本研究は三菱重工工作機械株式会社の助成を受けたものである。ここに謝意を表わす。

参 考 文 献

- G.G. Wang and S. Shan: Review of metamodeling techniques in support of engineering design optimization, *Journal of Mechanical Design*, **129**-4, 370/380 (2007)
- 本多, 中村: バンディット問題の理論とアルゴリズム, 講談社 (2016)
- H. Hassan, A. Aue, C. Chen, V. Chowdhary, J. Clark, C. Federmann, X. Huang, M. Junczys-Dowmunt, W. Lewis, M. Li, S. Liu, T. Yan Liu, R. Luo, A. Menezes, T. Qin, F. Seide, X. Tan, F. Tian, L. Wu, S. Wu, Y. Xia, D. Zhang, Z. Zhang and M. Zhou: Achieving human parity on automatic Chinese to English news translation, arXiv preprint arXiv:1803.05567 (2018)
- D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel and D. Hassabis: Mastering the game of Go without human knowledge, *Nature*, **550**-7676, 354/359 (2017)
- S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz and D. Quillen: Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection, *The International Journal of Robotics Research*, **37**-4-5, 421/436 (2018)
- S.J. Pan and Q. Yang: A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering*, **22**-10, 1345/1359 (2009)
- J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba and P. Abbeel: Domain randomization for transferring deep neural networks from simulation to the real world, *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, 23/30 (2017)
- E. Alpaydin: *Introduction to Machine Learning*, MIT Press (2009)
- C.E. Rasmussen and C.K.I. Williams: *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press (2005)
- E. Contal, V. Perchet and N. Vayatis: Gaussian process optimization with mutual information, *Proceedings of the 31st International Conference on Machine Learning (ICML)*, 253/261 (2014)
- G.E. Hinton and R.R. Salakhutdinov: Reducing the dimensionality of data with neural networks, *Science*, **313**-5786, 504/507 (2006)
- C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow and R. Fergus: Intriguing properties of neural networks, arXiv preprint, arXiv:1312.6199 (2013)
- H. Wang, J. Yang, H. Lee and S. Han: Learning to design

circuits, arXiv preprint, arXiv:1812.02734 (2018)

- 14) D.H. Wolpert and W.G. Macready: No free lunch theorems for optimization, *IEEE Transactions on Evolutionary Computation*, 1-1, 67/82 (1997)

《付 録》

A. BOA における評価値と探索性能の関係

BOA には 4 章の実験において扱った, 報酬優先, 近似優先の 2 つ以外にも両者をバランスしたような探索点の評価値を設定するという実装がある. 本付録では, BOA における探索点決定時の評価値と最終的な探索性能の関係について述べる. **Fig. A. 1** と **Fig. A. 2** は 5.2 節で述べた実験と同じ条件のもと, 複数の探索点評価値における BOA の探索スコアを示したものである. 図中の k は報酬優先と近似優先のバランスを決定するパラメータであり, (A. 1) 式のように分散の大きさにかける重み係数である.

$$\text{Priority metrics } p = \frac{\mu(x)}{\mu_{\text{mean}}} + k \frac{\sqrt{\sigma^2(x)}}{\sigma_{\text{mean}}} \quad (\text{A. 1})$$

また, タスクによって推定値の大きさと分散値の大きさのオーダーが大きく異なる可能性があるため, 推定値と分散値の定義域全体の値を用いて正規化している. Fig. A. 1 と Fig. A. 2 から, k が 0.2 ~ 5.0 の間では BOA-exploit, BOA-explore と同程度の性能を示す. BOA は過去の情報を使ってつぎの探索点を決定するという設計になっているが, 今回の実験設定では N_{sample} が 3, 5 と小さく, 過去の情報の精度が悪かったため, BOA 全体を通して探索性能が上がりなかったのだと考えられる.

Fig. A.1 Search score in several parameters of BOA (2 variables)

Fig. A.2 Search score in several parameters of BOA (5 variables)

[著 者 紹 介]

唐 澤 宏 之

2017 年東京大学大学院工学系研究科機械工学専攻修士課程修了. 同年大学院新領域創成科学研究科人間環境学専攻博士課程進学, 現在に至る. 機械システムの制御パラメータ最適化, 少数データによる機械学習の研究に従事.

福 井 類 (正会員)

2009 年東京大学大学院情報理工学系研究科知能機械情報学専攻博士課程修了. 三菱重工業 (株), 東京大学機械工学専攻特任講師, 新領域創成科学研究科准教授. ロボット工学, 機械システムデザイン, 知能センシングシステムの研究に従事. 博士 (情報理工学).

割 澤 伸 一

1994 年東京大学大学院工学系研究科産業機械工学専攻博士課程修了. 東京工業大学精密工学研究所助手, 東京大学工学系研究科講師, 准教授, 新領域創成科学研究科教授. ナノメカニクス, ナノ・マイクロ加工, 生産システム, 医療支援システム, 生産文化の研究に従事. 博士 (工学).