

# Transfer Learning for Visual Categorization: A Survey

---

## 備考

---

## Abstract

---

通常の機械学習やデータマイニング技術では、将来のデータが同じ特徴空間内にあるか、あるいは訓練データと同じ分布を持っているという大前提の下で、将来の推論のための訓練データを学習する。しかし、人間がラベルを付けた訓練データの利用可能性は限られているため、将来のデータが同じ特徴空間内に留まる、あるいは将来のデータと同じ分布を持つ訓練データは、オーバーフィットの問題を回避するのに十分な量を保証することはできない。実際のアプリケーションでは、ターゲット領域のデータとは別に、異なる領域の関連データを含めることで、ターゲットとなる未来のデータに関する事前知識の利用可能性を拡大することができる。転移学習は、このような領域横断的な学習問題を解決するために、関連する領域のデータから有用な情報を抽出し、それを転移して対象タスクで利用することを目的としています。近年では、視覚分類への応用が進んでおり、行動認識タスクでのビュー発散や画像分類タスクでの概念ドリフトなどの典型的な問題を効率的に解決することができるようになってきている。本論文では、物体認識、画像分類、人間の行動認識などの視覚的な分類に適用されている最新の転移学習アルゴリズムを調査する。

## Introduction

---

過去数年の間に、コンピュータビジョンのコミュニティでは、ビデオ検索、監視、ロボット工学などの分野でかなりの量の応用が見られるようになってきました。通常の機械学習アプローチ[1]-[7]は、学習データとテストデータが同じ特徴空間内にあるか、同じ分布を共有しているという大前提の下で、有望な結果を得てきた。しかし、実社会での応用では、人間の手作業によるラベリングの高価さや環境の制約から、テストデータと同じ特徴空間や同じ分布に属する十分な学習データが必ずしも得られない場合がある。代表的な例としては、訓練のために各アクションクラスに対して1つのアクションテンプレートのみが提供されている[8]-[11]や、訓練サンプルが異なる視点からキャプチャされている[12]などがある。このような状況では、通常の機械学習技術は失敗する可能性が高い。このことは、人間の視覚システムの能力を思い起こさせる。物体の巨大な幾何学的変動やクラス内変動を考えると、人間は一生のうちに何万もの視覚カテゴリを学習することができ、これは、人間が情報と知識の蓄積によってこのような能力を達成しているという仮説につながる[13]。世界には約1~3万個の物体クラスがあると推定されており[14]、子供は1日に4~5個の物体クラスを学ぶことができるとされている[13]。子供が一日のうちに見ることができるオブジェクトの数には限りがあるため、対応する大量のオブジェクトデータから新しいオブジェクトクラスを学習することはできない。したがって、以前に知られていたオブジェクトから得られた既存の知識が、新しいオブジェクトクラスとの関連性を介して新しい学習プロセスを支援すると考えられている。例えば、スイカが何であるかを知らなかったとすると、新しい物体カテゴリであるスイカを記憶するためには、スイカの丸い形や緑の色などの以前の知識と一緒に、スイカの学習サンプルを1つだけ用意すればよい。転写学習は、与えられた領域で新たなタスクを実行する際に、他の関連領域の十分な量の予備知識を利用することで、人間の視覚システムを模倣したものである。転移学習では、学習データとテストデータの両方が2種類の領域に貢献することができます。

1. 目標ドメイン
2. ソースドメイン

対象領域には、カテゴリ化システムのタスクであるテストインスタンスが含まれ、ソース領域には、対象領域データとは異なる分布下にあるトレーニングインスタンスが含まれます。多くの場合、転移学習タスクのターゲットドメインは1つだけであるが、ソースドメインは単一でも複数でもよい。例えば、[15]では、異なるドメインのデータセットをまたいで行動認識を行い、背景がきれいで視点や規模の変化が限定的なKTHデータセット[16]をソースデータセットとし、現実的なシナリオからキャプチャしたMicrosoftの研究行動データセットとTRECVIDの監視データ[17]をターゲットデータセットとしている。[18]では、映像概念検出の課題に対して、異なるテレビ番組のチャンネルからソースデータセットとターゲットデータセットを選択している。

転移学習は、使用される学習データの一部/全部がテストデータとは異なる分布下にある特殊な学習パラダイムと考えることができる。視覚学習問題の観点から知識移転の意義を理解するために、文献（[19]-[21]を参照）では、移転プロセスに関する3つの一般的な問題を結論づけている。

1. いつ移転するか
2. 何を移転するか
3. どのように移転するか

まず、「いつ移転するか」には、特定の学習課題に対して移転学習が必要かどうか、移転元のドメインデータと移転先のドメインデータが関連しているかどうかという問題が含まれている。[22]-[24]のシナリオでは、学習サンプルが十分であり、目標領域の制約を受けながらも、他の領域を含めて原領域が余計なものになってしまうような場合に、優れた性能を発揮することが可能である。ソースドメインとターゲットドメインのデータの異なるペア間には様々な発散レベルが存在し、発散に関係なくソースドメインからターゲットドメインに知識を強要すると、ある種の性能低下を引き起こしたり、さらに悪い場合にはターゲットドメインでの元のデータの整合性を壊したりすることになります。第二に、何を移転するかという答えは、3つの側面から結論づけられる。

1. 帰納的移転学習、すなわち、すべてのソースドメインのインスタンスとそれに対応するラベルを知識移転のために使用
2. インスタンス移転学習、すなわち、ソースドメインのインスタンスのみを使用
3. パラメータ移転学習、すなわち、ソースドメインのインスタンスとラベルに加えて、ソースドメインから事前に学習されたモデルのいくつかのパラメータを利用して、ターゲットドメインでの性能向上を支援 最後に、どのように移行するかは、具体的な移行学習手法のすべてを含み、移行学習文献の中で最も研究されてきた部分でもある。例えば、[25]-[27]では、知識の移転を非負行列トリファクタリゼーションのフレームワークに基づいて行っており、[28]では、次元削減を介した移転学習を行っている。従来の機械学習アプローチと知識移転アプローチの基本的なフレームワークを図1に示す。従来の機械学習アプローチでは、テストインスタンスの車を予測するための訓練セットの理想的な選択は、車を含むことである。しかし、知識移転の場合には、学習セットは車ではなく、車の車輪に似ている車輪、車の車輪と車輪の知識を共有する自転車、あるいは、車とは無関係に見えるが、実際には車の画像の局所的な部分と特定のエッジや幾何学的なレイアウトを共有するラップトップや鳥など、いくつかの関連するカテゴリを含むことができる。

ビッグデータの時代が到来した今、転移学習は、より関連性の高いデータを用いて目的の問題を解決するために、より多くのメリットを提供することができる。このように、今後の研究では、より多くの応用が期待されている。本調査では、視覚的な分類課題に対する転移学習の手法を包括的に概観することで、読者の皆様が、視覚的な分類課題に転移学習がどのように適用できるのか、あるいは適切な転移学習手法を用いて課題を解決するために、本調査での分析や議論を活用できる可能性があることを示したいと思います。視覚的な分類タスクは、学習プロセスで利用できる視覚的特性、例えば、物体の部分の外観や形状、物体の局所的な対称性、構造などに起因するいくつかのユニークな特性を持っている。これらのユニークな特性はすべて転

移学習アルゴリズムを設計する際に採用することができ、我々の研究は前者がデータマイニングタスクに関連した分類、回帰、クラスタリング問題に焦点を当て、後者が強化学習に焦点を当て、正しくラベル付けされた例ではなく、限られた環境フィードバックのみで問題を解決する[19]や[29]の研究とは異なるものとなっている。

本調査の残りの部分は以下のように構成されている。第II節では概要を説明する。第III節と第IV節では、特徴表現と分類器を用いて知識の伝達を実行する2つの伝達学習カテゴリについて詳細に議論し、「何をどのように伝達するか」「どのように伝達するか」という問題を解決する。第V節では、複数のソースドメインからのモデル選択方法、すなわち、いつ移転するかについて議論する。第VI節では、述べられた移転学習法の評価、分析、考察を行う。最後に、第VII節で結論を述べる。

## 2. OverView

### A. Developing Interests on Transfer Learning

転移学習（クロスドメイン学習、ドメイントランスファー、ドメインアダプテーション）は、前世紀にその概念が提唱されて以来、特定の機械学習手法として研究されてきた歴史があります。近年、インターネット上での情報（音声、画像、動画など）の爆発的な普及に伴い、精度、データ規模、計算効率の面に対象タスクへの要求が高まってきており、パターン認識や機械学習のあらゆる研究分野から転移学習アプローチへの関心が高まってきている。通常の機械学習技術が限界に達したとき、転移学習は新しい流れを開き、これまでの学習方法や分類や回帰タスクの扱い方を根本的に変えてしまう可能性があります。この流れに沿って、いくつかのワークショップやチュートリアルが開催されてきた（例えば、機械学習やデータマイニングの分野では、NIPS 1995 postconference workshopが開催され、強化学習の分野では、[29]で移転学習の調査が行われている）。本調査では、行動認識、物体認識、画像分類を含む視覚分類への転移学習技術の応用に焦点を当てている。

$D^T = D_l^T \cup D_u^T$  を対象領域データとし、部分的にラベル付けされている部分を  $D_l^T$ 、ラベル付けされていない部分を  $D_u^T$  とする。対象領域データの他に、半ラベル化または完全ラベル化された補助データの集合をソース領域データと見なし、単一ソースの場合は  $D^s = (x_i, y_i)_{i=1}^a$ 、複数ソースの場合は  $D_k^s = (x_i^k, y_i^k)_{i=1}^{N_k^a}$  の表現  $D_1^s, D_2^s, \dots, D_M^s$  とする。ここで、 $x_i \in \mathbb{R}^d$  は、 $i$  番目の特徴ベクトルであり、 $d$  はデータ次元、 $y_i$  は  $i$  番目のサンプルのクラスラベルを表す。

先行提案によれば、知識伝達に関する共通の問題点は2つある。第一に、補助サンプルは一般的に適応時に相互依存性を考慮せずに扱われるため、適応されたデータが任意に分散し、補助データの単一データサンプルを超えた構造情報が損なわれる可能性がある。第二に、適応の間、ノイズ、特に補助領域からの可能性のある外れ値は、盲目的にターゲット領域に強制される[30]。

補助領域から対象領域へ知識を伝達する際には、対象領域データと各ソース領域データとの間の分布類似度を知ることが重要である。これまでのところ、2つのドメインの分布類似度を測定するための最も一般的な基準は、最大平均不一致（MMD）と呼ばれるノンパラメトリック距離測定法である。MMDは[31]で提案されており、再生カーネルヒルベルト空間におけるデータ分布を比較するものである。