

# A Survey on Deep Transfer Learning

## 備考

### 著者

Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu

### 掲載

International Conference on Artificial Neural Networks, Springer, pp. 270-279, 2018.

## Deep Transfer Learning

転移学習は、機械学習の基本的な問題である学習データの不足を解決するための重要なツールです。これは、学習データとテストデータが等値でなければならないという前提を緩和することで、学習元の領域から学習対象の領域へ知識を転移しようとするもので、学習データが不足しているために改善が困難な多くの領域に大きな効果をもたらします。図 1 に示す転送学習の学習過程を説明する。

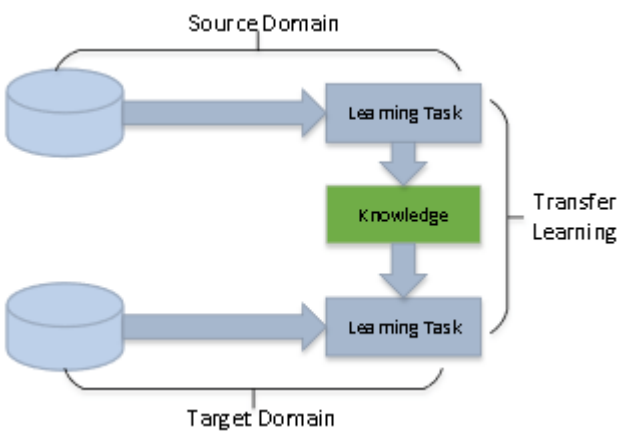


Fig. 1. 転送学習の学習過程

本調査では、いくつかの表記法を明確に定義する必要がある。まず、ドメインとタスクをそれぞれ定義する。ドメインは  $D = \{\chi, P(X)\}$  で表現され、特徴空間  $\chi$  とエッジ確率分布  $P(X)$  の2つの部分を含み、 $X = \{x_1, \dots, x_n\} \in \chi$  で表される。タスクは  $T = \{y, f(x)\}$  で表されます。 $f(x)$  は条件付き確率関数  $P(y|x)$  とみなすこともできる。そうすると、転移学習は次のように正式に定義することができる。

定義 1. (転移学習)

$D_{t\$}$  に基づいて学習タスク  $T_{t\$}$  が与えられ、その学習タスク  $T_{s\$}$  に対して  $D_{s\$}$  からの助けを得ることができる。転送学習は、 $D_{s\$} \neq D_{t\$}$  および/または  $T_{s\$} \neq T_{t\$}$  の  $D_{s\$}$  と  $T_{s\$}$  から潜在知識を発見して転移することで、学習タスク  $T_{t\$}$  に対する予測関数  $f_{T_{t\$}}(.)$  の性能を向上させることを目的としている。また、多くの場合、 $D_{s\$}$  の大きさは  $D_{t\$}$  の大きさよりもはるかに大きい。 $N_{s\$} \gg N_{t\$}$ 。

調査[19]や[25]では、転移学習の手法を大きく3つに分類しており、その中でも特にソースドメインとターゲットドメインの関係については、広く受け入れられている。これらの調査結果は、これまでの転移学習に関

する研究をまとめたものであり、多くの古典的な転移学習手法が紹介されている。さらに、最近では、より新しく、より優れた方法が数多く提案されています。近年の転移学習の研究コミュニティでは、主に領域適応と多元領域転移の2つの側面に焦点が当てられています。

現在、深層学習は近年多くの研究分野で一世を風靡しています。そのためには、以下のように定義された深層学習と呼ばれる深層ニューラルネットワークを用いて、効率的に知識を転移する方法を見つけることが重要である。

定義2, (深層転移学習)

転移学習タスクが  $\langle D_s, T_s, D_t, T_t, f, T(\cdot) \rangle$  で定義されているとする。  $f, T(\cdot)$  が深層ニューラルネットワークを反映した非線形関数である深層転移学習課題である。

### 3. Categojrries

深層転移学習は、深層ニューラルネットワークを用いて他分野の知識をどのように活用するかを研究するものである。深層ニューラルネットワークは様々な分野で普及しているため、かなりの量の深層転移学習手法が提案されており、それらを分類してまとめることは非常に重要である。本論文では、深層転移学習に用いられている手法をもとに、深層転移学習をインスタンスベースの深層転移学習、マッピングベースの深層転移学習、ネットワークベースの深層転移学習、敵対者ベースの深層転移学習の4つに分類し、表1に示す。

表1：深層転移学習のカテゴライズ。

Approach category	Brief description	Some related works
Instances-based	ソースドメイン内のインスタンスを適切な重みで利用する。	
Mapping-based	2つのドメインのインスタンスを、より類似性の高い新しいデータ空間にマッピングする。	
Network-based	ソースドメインで事前に学習したネットワークの一部を再利用する。	
Adversarial-based	2つのドメインに適した転移可能な機能を見つけるために敵対技術を使用する。	

#### 3.1. Instances-based deep transfer learning

インスタンスベースの深層転移学習とは、特定の重み調整戦略を用いて、ソースドメインから部分的なインスタンスを選択し、その選択したインスタンスに適切な重みを割り当てることで、ターゲットドメインの学習セットを補完することをいう。これは、「2つのドメイン間では異なるが、ソースドメインの部分的なインスタンスは、適切な重みをもってターゲットドメインで利用できる」という前提に基づいている。インスタンスベースの深層転移学習のスケッチマップを図2に示す。

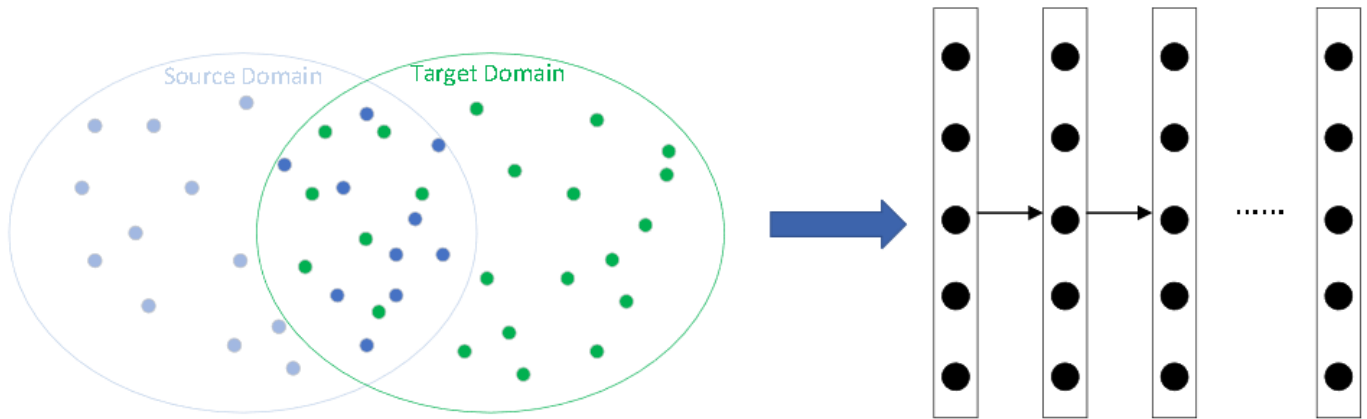


図2：インスタンスベースの深層転移学習のスケッチマップ。ソース領域の意味がターゲット領域と異なる水色のインスタンスは学習データセットから除外し、ソース領域の意味がターゲット領域と類似している濃紺色のインスタンスは適切な重みをつけて学習データセットに含める。

[4]で提案されたTrAdaBoostは、\*\*AdaBoostをベースにした技術を用いて、ソースドメイン内のターゲットドメインとは異なるインスタンスをフィルタリングする。ソースドメインのインスタンスを再重み付けして、ターゲットドメインに近い分布を構成する。最後に、ソースドメインの再重み付けされたインスタンスとターゲットドメインのオリジンインスタンスを用いてモデルを学習する。\*これにより、AdaBoostの特性を維持したまま、異なる分布領域での重み付け学習誤差を低減することができる。[27]で提案されたTaskTrAdaBoostは、新しいターゲットに対して迅速な再学習を促進する高速なアルゴリズムである。TrAdaBoostが分類問題用に設計されているのとは異なり、回帰問題をカバーするために[20]によってExpBoost.R2とTrAdaBoost.R2が提案された。Bi-weighting domain adaptation (BIW)は、[24]で提案されているように、2つのドメインの特徴空間を共通の座標系に合わせて、元のドメインのインスタンスに適切な重みを割り当てることができる。[10]は、砂岩顕微鏡画像の領域間分類の問題を処理するための拡張TrAdaBoostを提案している。[26]は、2つの異なるドメインのインスタンス重みと距離を並列に学習するメトリック転移学習フレームワークを提案し、ドメイン間の知識転移をより効果的に行う。[11]は、ソースドメインのインスタンスを利用できるアンサンブル転送学習をディープニューラルネットワークに導入している。

## 3.2. Mapping-based deep transfer learning

マッピングベースの深層転移学習とは、ソースドメインとターゲットドメインのインスタンスを新しいデータ空間にマッピングすることである。この新しいデータ空間では、2つのドメインからのインスタンスが類似しており、ユニオンディープニューラルネットワークに適している。これは、\*\*「2つの原点ドメインの間には異なるものがあるが、精巧な新しいデータ空間では、より類似したものに行うことができる」という前提に基づいています。\*インスタンスベースの深層転移学習のスケッチマップを[図3]に示す。

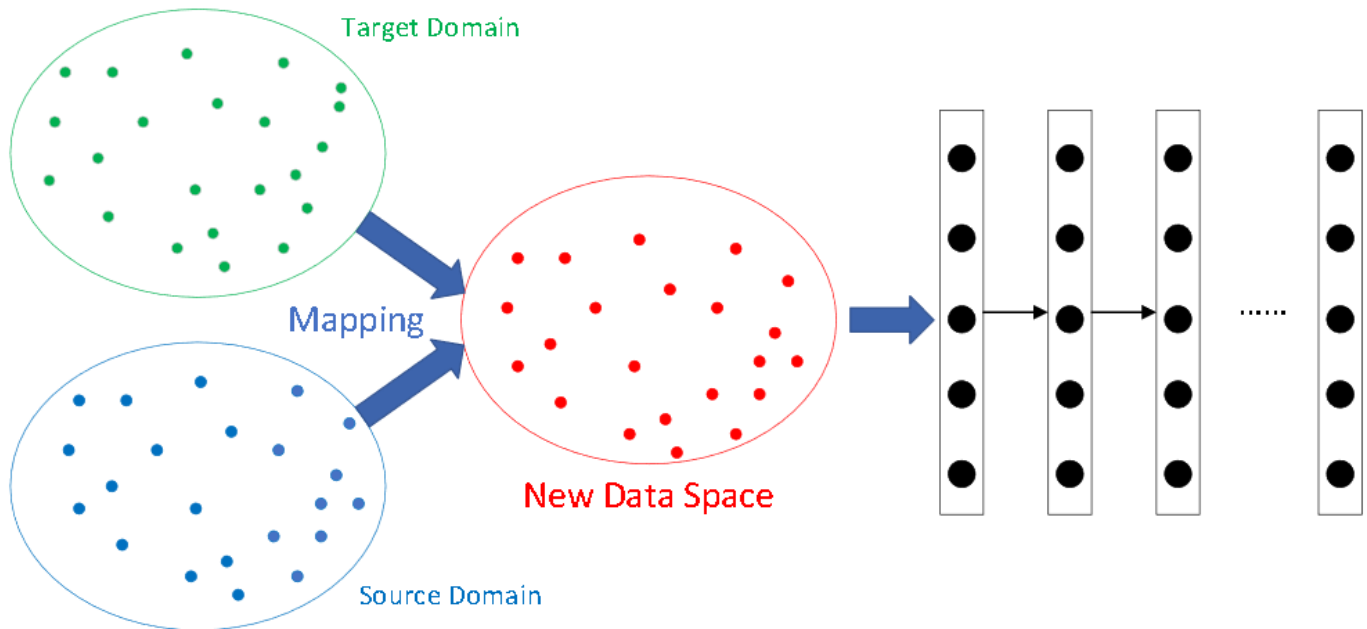


図3：マッピングベースの深層転移学習のスケッチマップ。同時に、ソースドメインとターゲットドメインのインスタンスが、より類似性の高い新しいデータ空間にマッピングされる。新しいデータ空間のすべてのインスタンスをニューラルネットワークの学習セットとみなす。

[18]で導入された転移成分分析(TCA)やTCAベースの手法[29]は、従来の転移学習の多くの応用分野で広く利用されてきた。TCA法をディープニューラルネットワークに拡張するのは自然な発想である。[23]はMMDをディープニューラルネットワークでの分布比較に拡張し、適応層と追加のドメイン混同損失を導入して、意味的に意味があり、ドメイン不変な表現を学習している。この研究で使用されているMMD距離は次のように定義されている。

$$D_{\text{MMD}}(X_s, X_T) = \sqrt{\frac{1}{|X_s|} \sum_{x_s \in X_s} \phi(x_s) \frac{1}{|X_T|} \sum_{x_t \in X_T} \phi(x_t) - \frac{1}{|X_s| |X_T|} \sum_{x_s \in X_s} \sum_{x_t \in X_T} \phi(x_s) \phi(x_t)}}$$

であり、損失関数は次のように定義されます。

$$\mathcal{L} = \mathcal{L}_C(X_L, y) + \lambda D_{\text{MMD}}^2$$

[12]は、MMD距離を[8]で提案された多重カーネルバリエーションMMD(MK-MMD)距離に置き換えることで、これまでの研究を改良したものである。畳み込みニューラルネットワーク(CNN)の学習タスクに関連する隠れ層を再生カーネルヒルベルト空間(RKHS)にマッピングし、マルチコア最適化法により異なる領域間の距離を最小化する。[14]は共同分布の関係性を測定するために共同最大平均不一致(JMMD)を提案している。JMMDはディープニューラルネットワーク(DNN)の転移学習能力を一般化し、異なる領域のデータ分布に適応させるために用いられており、従来の研究を改良したものである。また、[2]で提案されたワッサーシュタイン距離は、より良いマッピングを見つけるための新たなドメインの距離計測として利用できる。

### 3.3. Network-based deep transfer learning

ネットワークベースの深層転移学習とは、ソース領域で事前に学習した部分的なネットワークを、ネットワーク構造や接続パラメータなどを含めて再利用し、ターゲット領域で使用するディープニューラルネットワークの一部として転送することを指します。これは、「ニューラルネットワークは人間の脳の処理機構に似ており、反復的で連続的な抽象化プロセスである」という前提に基づいています。ネットワークのフロント

レイヤーは特徴抽出器として扱うことができ、抽出された特徴は汎用性がある」という前提に基づいています。図4にネットワークを用いた深層転移学習のスケッチマップを示す。

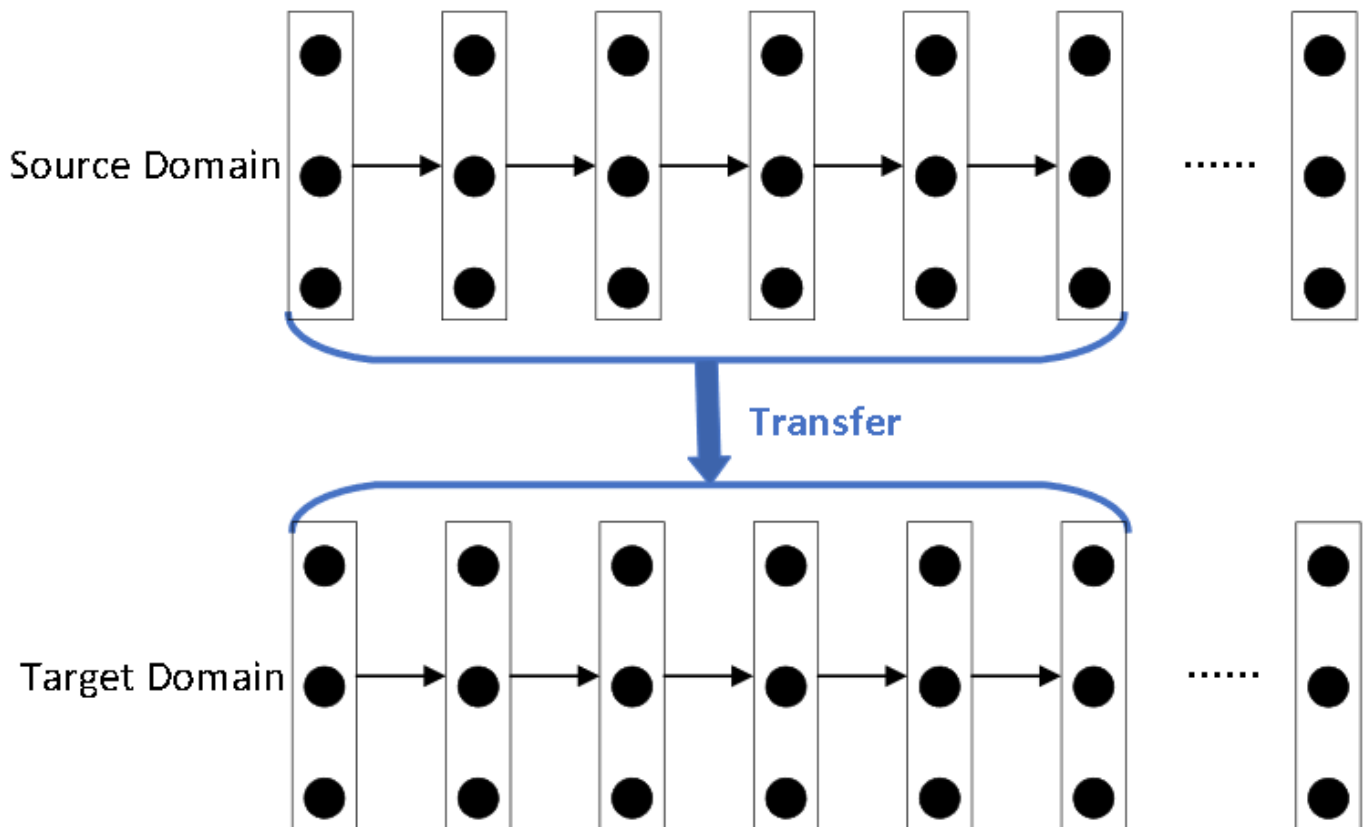


図4：ネットワークベースの深層転移学習のスケッチマップ。まず、大規模な学習データセットを用いてソース領域でネットワークを学習する。次に、学習元領域で学習したネットワークの一部を、学習対象領域用に設計された新しいネットワークの一部として移行する。最後に、移行した部分ネットワークを更新することで、ネットワークの微調整を行うことができる。

[9]はネットワークを2つの部分に分け、前者は言語に依存しない特徴変換、後者は言語関連分類器である。言語に依存しない特徴変換は多言語間での移行が可能である。[17]は、CNNによって訓練されたフロントレイヤーをImageNetデータセット上で再利用し、他のデータセットの画像の中間画像表現を計算している。[15]は、ソース領域のラベル付きデータとターゲット領域のラベルなしデータから適応分類器と転送可能な特徴量を共同で学習するアプローチを提案しており、ディープネットワークに複数の層を差し込んでターゲット分類器を参照して残差関数を明示的に学習している。[30]はDNNでドメイン適応とディープハッシュ特徴を同時に学習している。[3]は新規なマルチスケール畳み込みスパースコーディング法を提案している。この手法は、学習したパターンのスケール特異性を強制した上で、異なるスケールのフィルタバンクを共同で自動的に学習することができ、転移可能なベース知識を学習し、目標タスクに向けて微調整するための教師なしのソリューションを提供する。[6]は、実世界の物体認識タスクの知識を、複数の重力波信号検出器のグリッチ分類器に転移するために、深層転移学習を適用しています。その結果、DNNが教師なしクラスタリング法のための優れた特徴抽出器として利用できることを示している。

もう一つの注目すべき結果は、[28]がネットワーク構造と転移性の関係を指摘していることです。この研究では、領域内精度には影響しないが、転送性には影響するモジュールがあることを示している。また、ディープネットワークではどのような特徴が転移可能で、どのようなタイプのネットワークが転移に適しているかを指摘しています。その結果、LeNet, AlexNet, VGG, Inception, ResNetは、ネットワークベースの深い転移学習に適しているという結論が得られた。



## 3.4. Adversarial-based deep transfer learning

敵対ベースの深層転移学習とは、ソースドメインとターゲットドメインの両方に適用可能なトランスファー可能な表現を見つけるために、敵対生成ネット（GAN）[7]に触発された敵対技術を導入することを意味します。これは、「効果的な転移のためには、良い表現が主な学習タスクを識別し、ソースドメインとターゲットドメインを区別できなければならない」という前提に基づいています。[図5]に、逆境に基づく深層転移学習のスケッチマップを示す。

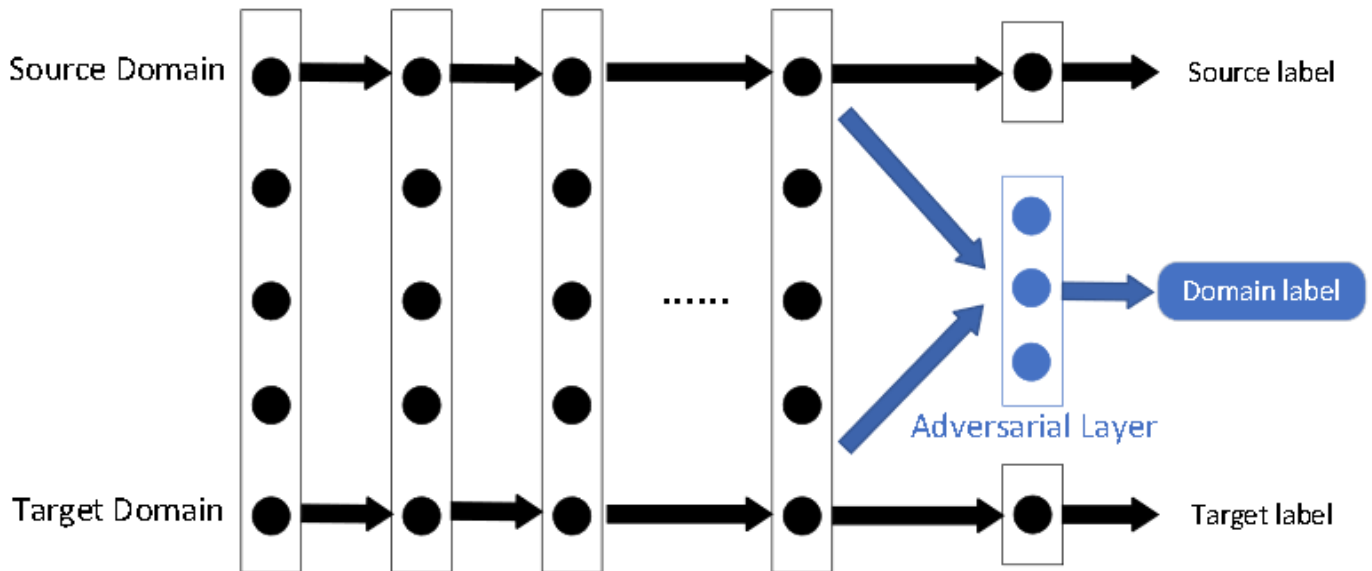


図5：逆境に基づく深層転移学習のスケッチマップ。ソースドメインの大規模データセットの学習過程では、ネットワークのフロントレイヤーを特徴抽出器とみなす。2つの領域から特徴量を抽出し、それを逆参照層に送る。アドバーサリアル層は、特徴量の起源を識別しようとする。もし、逆説的ネットワークの性能が悪い場合は、2種類の特徴量の差が小さく、転送性が良いことを意味し、逆に性能が悪い場合は、2種類の特徴量の差が小さく、転送性が良いことを意味する。次の学習プロセスでは、逆問題層の性能を考慮して、転送ネットワークがより転送性の高い一般的な特徴を発見するように強制することを検討する。

敵対的方式による深層転移学習は、その効果の高さと実用性の高さから、近年盛んに発展している。[1]では、損失関数に領域適応の正則化項を用いることで、領域適応のための転移学習に逆問題技術を導入している。[5]は、少数の標準層と単純な勾配反転層を用いることで、ほとんどのフィードフォワードニューラルネットワークモデルに適した学習法を提案している。また、[21]は、ラベル付けがまばらな対象領域データに対して、クロスドメインとクロスタスクの同時性を考慮した知識転移アプローチを提案している。この研究では、 $\mathcal{L}_D = \mathcal{L}_C + \lambda \mathcal{L}_{adver}$ と定義されたドメイン間距離の最適化をCNNに強制的に行わせるために、特別なジョイントロス関数を用いた。ここで、 $\mathcal{L}_C$ は分類損失、 $\mathcal{L}_{adver}$ はドメインの敵対的損失です。2つの損失は互いに正反対の関係にあるため、一方の損失が固定されたときに他方の損失を更新する反復最適化アルゴリズムが導入されている。[22]は、新しいGAN損失を提案し、識別的モデリングと組み合わせる新しいドメイン適応法を提案した。[13]は、複数の特徴層と分類器層を利用したランダム化された多線型敵対者ネットワークを提案し、ディープな敵対適応と識別的な敵対適応の両方を可能にしている。[16]は、領域逆説的損失を利用し、深層転移学習におけるより扱いやすい特徴を見つけるために、メトリック学習に基づいたアプローチを用いて、新しいタスクへの埋め込みを一般化している。

# Conclusiton

---

本調査論文では、深層転移学習の現在の研究を概観し、分類した。深層転移学習は、初めてインスタンスベースの深層転移学習、マッピングベースの深層転移学習、ネットワークベースの深層転移学習、敵対者ベースの深層転移学習の4つに分類される。ほとんどの実用的なアプリケーションでは、より良い結果を得るために、上記の複数の技術を組み合わせて使用することが多い。現在の研究の多くは教師付き学習に焦点を当てているが、教師なし学習や半教師付き学習において、深層ニューラルネットワークを用いてどのように知識を転移するかは、今後ますます注目されていくと思われる。従来の転移学習では、負の転移と転移可能性の尺度が重要な問題となっている。また、この2つの問題が深層転移学習に与える影響については、さらなる研究が必要である。また、非常に魅力的な研究分野として、物理学者、神経科学者、コンピュータ科学者の協力が必要とされる深層ニューラルネットワークにおける転移知識の物理的なサポートをより強力に見出すことが挙げられます。深層ニューラルネットワークの発展に伴い、多くの困難な問題を解決するために、深層転移学習が広く応用されることが予想されます。