

A_Survey_of_Transfer_Learning

備考

著者

Karl Weiss, Taghi M. Khoshgoftaar, DingDing Wang

掲載

Journal of Big Data, Vol. 3, Article No. 9, 40 pages, 2016.

Abstract

機械学習とデータマイニングの技術は、数多くの実世界での応用で使用されてきた。従来の機械学習手法の前提として、訓練データとテストデータが同じ領域から取得され、入力特徴空間とデータ分布の特性が同じであることが挙げられます。しかし、現実の機械学習では、この仮定が成り立たない場合もある。訓練データが高価であったり、収集が困難であったりする場合がある。そのため、異なる領域からより容易に得られるデータで訓練された高性能な学習者を作成する必要がある。この方法論をトランスファー学習と呼ぶ。本サーベイペーパーでは、転移学習を簡単に定義し、現在の解決策を紹介し、転移学習に適用されているアプリケーションをレビューする。最後に、様々な移転学習ソリューションのためのソフトウェアのダウンロードに関する情報と、今後の研究の可能性についての議論を掲載しています。調査した転写学習ソリューションは、データサイズに依存せず、ビッグデータ環境にも適用可能である。

Background

データマイニングや機械学習の分野は、過去の情報（訓練データ）からパターンを抽出して将来の結果を予測することができる多くのアプリケーションで広く利用されており、成功している[129]。従来の機械学習は、訓練データとテストデータが同じ入力特徴空間と同じデータ分布を持つことを特徴としている。訓練データとテストデータの間にデータ分布の違いがある場合、予測学習者の結果は劣化

する可能性がある[107]。特定のシナリオでは、テストデータの特徴空間と予測データの分布特性が一致する学習データを得ることは困難であり、コストがかかる場合がある。そのため、関連するソースドメインから訓練されたターゲットドメインのための高性能な学習器を作成する必要がある。これが転移学習の動機である。

転移学習とは、ある領域の学習者を、関連する領域の情報を転送することで、ある領域の学習者を向上させるために用いられる。なぜ転送学習が可能なのかを理解するために、実世界の非技術的な経験から引き出すことができます。ピアノを習いたいと思っている二人の人の例を考えてみましょう。一人は音楽の経験がなく、もう一人はギターを弾いて音楽の知識を豊富に持っています。音楽の知識が豊富な人は、以前に学んだ音楽の知識をピアノを弾くことを学ぶ課題に移すことで、より効率的な方法でピアノを学ぶことができる[84]。ある人は、以前に学習したタスクから情報を取り、それを有益な方法で利用して、関連するタスクを学習することができる。

機械学習の具体的な例として、デジタルカメラのレビューからラベル付けされたデータが豊富に存在する商品レビューのテキストセンチメントを予測する課題を考えてみましょう。学習データと対象データの両方がデジタルカメラのレビューから得られたものであれば、従来の機械学習技術を用いて良好な予測結果を得ることができます。しかし、学習データがデジタルカメラレビューからのもので、対象データが食品レビューからのものである場合には、領域データの違いにより予測結果が劣化する可能性がある。デジタルカメラレビューと食品レビューには、完全に同じではないにしても、共通して多くの特徴があります。どちらも同じ言語を使用してテキスト形式で書かれており、購入した製品に関する意見を表現しています。これら2つのドメインは関連しているので、転移学習は、潜在的に対象学習者の結果を改善するために使用することができます[84]。転移学習環境におけるデータドメインを見る別の方法は、トレーニングデータとターゲットデータが、高レベルの共通ドメインによってリンクされた異なるサブドメインに存在するということである。例えば、ピアノ奏者とギター奏者は、音楽家ドメインのサブドメインである。さらに、デジタルカメラのレビューと食品のレビューは、レビュードメインのサブドメインである。高レベル共通ドメインは、サブドメインがどのように関連しているかを決定します。

前述したように、転送学習の必要性は、対象となる学習データの供給が限られている場合に発生します。これは、データが希少であったり、データを収集してラベルを付けるのに高価であったり、データにアクセスできなかったりすることが原因かもしれません。ビッグデータリポジトリの普及に伴い、興味のある対象領域に関連しているが全く同じではない既存のデータセットを使用することで、転移学習ソリューションは魅力的なアプローチとなります。テキストセンチメント分類[121]、画像分類[30, 58, 146]、人間活動分類[46]、ソフトウェア欠陥分類[77]、多言語テキスト分類[145, 91, 144]など、転移学習が成功裏に適用されている機械学習アプリケーションは数多くあります。

本調査論文は、学習に関心のある研究者に、関連する著作の概要、転移学習によって取り組まれている応用例、転移学習の分野に関連する問題点と解決策を提供することを目的としている。本調査論文では、分類、回帰、クラスタリング問題のデータマイニングに関連して、転移学習の分野で現在用い

られている手法の概要を紹介するが、強化学習のための転移学習には焦点を当てていない（強化学習についてはTaylor[112]を参照のこと）。転移学習の歴史や分類法に関する情報は、本調査論文では提供されていないが、Pan [84]の論文を参照されたい。2010 年に Pan [84]による転移学習の調査論文が発表されて以来、転移学習の進歩と革新を取り上げた学術論文は 700 本を超えている。これらの論文は、新しいアルゴリズムの開発、既存の転移学習アルゴリズムの改良、新しいアプリケーション領域へのアルゴリズム展開の分野を幅広くカバーしています。本論文では、過去5年間に調査された転移学習ソリューションの多様性と代表的なものを取り上げています。ほとんどの論文は汎用的な転移学習ソリューションを提供しているが、中には個々のアプリケーションに特化したソリューションを提供している論文もある。この論文は、読者が機械学習の実務知識を持っていることを前提に書かれています。機械学習の詳細については、Witten [129]を参照されたい。この論文で調査した論文は、提案された解決策の高度な説明を、ユニークで顕著な点を強調して提示することを目的としています。調査対象論文の実験は、応用例、テストされた他の競合するソリューション、および経験の全体的な相対的な結果に関して記述されています。この調査論文では、私たちの知る限りではユニークな異種転移学習のセクションを提供しています。さらに、様々な調査対象論文のソフトウェアダウンロードのリストも提供しています。

本稿の残りの部分は以下のように構成されている。"転移学習の定義" では、転移学習の定義と表記法を提供する。"ホモジニアス転移学習(同種転移学習)"と"ヘテロジニアス転移学習(異種転移学習)"のセクションでは、同種転移学習と異種転移学習に関する解答を提供し、"ネガティブ転移"のセクションでは、転移学習に関連するネガティブ転移に関する情報を提供する。"transfer learning-ing application"では、転移学習の応用例を提供しています。"Conclusion and discussion"では、今後の研究の可能性についてまとめ、議論しています。「付録」では、転移学習に関連したソフトウェアのダウンロード情報を提供している。

Define itions of transfer learning (転移学習の定義)

以下のセクションでは、本論文の残りの部分で使用した表記法と定義を列挙する。このセクションの表記法と定義は、両方の調査の一貫性を維持するために、Pan [84]による調査論文の表記法と一致しています。以下に挙げた定義の例を説明するために、ソフトウェアモジュールの欠陥分類の機械学習アプリケーションを使用して、学習者がソフトウェアモジュールが欠陥が発生しやすいかどうかを予測するように訓練します。

ドメイン D は、特徴空間 χ と限界確率分布 $P(X)$ の2つの部分で定義され、 $X = \{x_1, \dots, x_n\} \in \chi$ である。例えば、機械学習アプリケーションがソフトウェアモジュールの欠陥分類であり、各ソフトウェアメトリックを特徴量とすると、 x_i は i 番目のソフトウェアモジュールに対応する i 番目の特徴ベクトル（インスタンス）であり、 n は X の特徴ベクトルの数であり、 χ はす

すべての可能な欠陥ベクトルの空間であり、 X は特定の学習サンプルである。与えられたドメイン D について、タスク T は、ラベル空間 Y と、特徴ベクトルとラベル対 $\{x_i \in X, y_i \in Y\}$ から学習される予測関数 $f(\cdot)$ の2つの部分によって定義される。ソフトウェアモジュールの欠陥分類アプリケーションでは、 Y はラベルの集合であり、ここでは *true* と *false* を含み、 y_i は *true* または *false* の値をとり、 $f(x)$ はソフトウェアモジュール x のラベル値を予測する学習者である。ここで、 D_S は、 $D_S = \{(x_{S1}, y_{S1}), \dots, (x_{Sn}, y_{Sn})\}$ であるソースドメインデータとして定義され、 $x_{Si} \in X_S$ は D_S の i 番目のデータインスタンスであり、 $y_{Si} \in Y_S$ は x_{Si} に対応するクラスラベルである。同様に、 D_T は、 $D_T = \{(x_{T1}, y_{T1}), \dots, (x_{Tn}, y_{Tn})\}$ のターゲットドメインデータとして定義され、 $x_{Ti} \in X_T$ が D_T の i 番目のデータインスタンスであり、 $y_{Ti} \in Y_T$ が x_{Ti} に対応するクラスラベルである。さらに、ソースタスクを T_S 、ターゲットタスクを T_T 、ソース予測関数を $f_S(\cdot)$ 、ターゲット予測関数を $f_T(\cdot)$ と表記する。

転移学習を正式に定義します。

相関関係のあるソースタスク T_S を持つソースドメイン D_S と、対応するタスク T_T を持つターゲットドメイン D_T があるとすると、 $D_S \neq D_T$ または $T_S \neq T_T$ のように、 D_S と T_S からの関連情報を用いてターゲット予測関数 $f_T(\cdot)$ を改善するプロセスを**転移学習**と呼びます。

ここで定義されている単一ソースドメインは、複数のソースドメインに拡張することができます。転移学習の定義を考えると、 $D_S = \{X_S, P(X_S)\}$ 、 $D_T = \{X_T, P(X_T)\}$ であるから、 $D_S \neq D_T$ の条件は、 $X_S \neq X_T$ および/または $P(X_S) \neq P(X_T)$ であることを意味します。転移学習に関して

$X_S \neq X_T$ となる場合を**ヘテロジニアス(異種)転移学習**と定義します。

また、 $X_S = X_T$ の場合を**ホモジニアス(同種)転移学習**と定義します。

ソフトウェアモジュールの欠陥分類の例に戻りますが、異種転移学習とは、ソースソフトウェアプロジェクトとターゲットソフトウェアプロジェクトのメトリクス (特徴) が異なる場合のことです。また、同種転移学習とは、ソフトウェアのメトリクスがソースとターゲットの両方のソフトウェアプロジェクトで同じである場合のことです。

転移学習の定義に引き続き、

$P(X_S) \neq P(X_T)$ の場合は、入力空間の限界確率分布がソースドメインとターゲットドメインで異なることを意味します。

下平[107]は、入力領域の限界確率分布が異なる場合、与えられた入力領域で訓練された学習者は、目標領域に対して最適なパフォーマンスを発揮しないことを実証しました。ソフトウェアモジュールの欠陥分類アプリケーションに関連して、限界確率分布の違いの例として、ソースソフトウェアプログラムがユーザインタフェースシステム用に使われており、ターゲットソフトウェアプログラムがDSPシグナリングデコーダアルゴリズム用に使われている場合がある。もう一つ考えられる転移学習の条

件 (上記の定義から) は $T_S \neq T_T$ であり, $T = \{Y, f(.)\}$, あるいはこれを書き換えると $T = \{Y, P(Y|X)\}$ と記載されていました. したがって, 転移学習環境では, $Y_S \neq Y_T$, $P(Y_S|X_S) \neq P(Y_T|X_T)$ となる可能性があります.

$P(Y_S|X_S) \neq P(Y_T|X_T)$ の場合は, ソースドメインとターゲットドメインの間の条件付き確率分布が異なることを意味します.

条件付き確率分布の不一致の例としては, 特定のソフトウェアモジュールがソースドメインとターゲットドメインで異なる障害が発生しやすい結果をもたらす場合があります.

$Y_S \neq Y_T$ の場合は, クラス空間の不一致を意味します.

このケースの例としては, ソースソフトウェアプロジェクトでは, 欠陥が発生しやすいものは *true*, 欠陥が発生しないものは *false* のバイナリラベル空間を持ち, ターゲットドメインでは, 欠陥が発生しやすいモジュールの5つのレベルを定義するラベル空間を持っている場合が挙げられます.

他にも, 欠陥のある分類器の劣化を引き起こすケースとして, $P(Y_S) \neq P(Y_T)$ の場合がありますが, これはソースドメインとターゲットドメインの間でラベル化されたデータセットが不均衡であることが原因となっています.

従来の機械学習の場合は, $D_S = D_T$, $T_S = T_T$ です.

本論文で使用する共通の表記法を表 1 にまとめました.

Table 1 よく使われる表記のまとめ

表記	説明	文章中での例
\mathcal{X}	入力特徴量空間	全ての可能な欠陥ベクトルの空間
\mathcal{Y}	ラベル空間	<i>true</i> と <i>false</i>
T	予測学習のタスク	ラベル空間 \mathcal{Y} と, 特徴ベクトルとラベル対 $\{x_i \in \mathcal{X}, y_i \in \mathcal{Y}\}$ から学習される予測関数 $f(.)$ の2つの部分によって定義される.
下付き S	ソース	D_S は, (省略) ソースドメインデータとして定義され
下付き T	ターゲット	D_T は, (省略) ターゲットドメインデータとして定義され
$P(X)$	限界確率分布	
$P(Y X)$	条件付き確率分布	

表記	説明	文章中での例
$P(Y)$	ラベルの確率分布	
D_S	ソースドメインデータ	
D_T	ターゲットドメインデータ	