転移学習

～様々なデータに適応するための機械学習の方法論～

松井　孝太† Zhi Li†† 米川　慧††† 黒川　茂莉†††

† 名古屋大学大学院医学系研究科　〒 466-8550 名古屋市昭和区鶴舞町 65
†† 大阪大学大学院情報科学研究科　〒 565-0871 大阪府吹田市山田丘 1-5
††† KDDI 総合研究所　〒 356-8502 埼玉県ふじみ野市大原 2-1-15
E-mail: †matsui.k@med.nagoya-u.ac.jp, ††li.zhi@ist.osaka-u.ac.jp,
†††{ke-yonekawa,mo-kurokawa}@kddi-research.jp

あらまし　学習済みモデルを学習時とは異なるデータ・タスクに適応させるための機械学習の方法論である転移学習は，学習済みモデルを有効利用し，様々なデータに適応するための技術として注目を集めている．本稿では，転移学習の問題設定や手法の分類を説明した後,「何を転移するか」に着目して事例転移，特徴転移，パラメータ転移に関する具体的な手法を説明する．さらに発展として，データの標本空間も異なる異種/異質的な設定での転移学習の手法についても説明する．

キーワード　転移学習，同種/同質的転移学習，異種/異質的転移学習

1 はじめに

従来の機械学習の重要な性質である汎化能力は，モデルを学習するデータ（訓練データ）と学習したモデルを適用するデータ（テストデータ）が独立同一分布に従うという仮定に大きく依存している．しかし，現実には訓練データとテストデータの生成分布が異なっている問題や，あるデータで学習したモデルや特徴量を大きく質の異なるデータのモデリングに利用する問題が多数存在する．例えば，ある疾患に罹患しているかどうかを予測する診断法開発では，モデルを訓練するデータを取得した病院と，学習したモデルを運用する病院で来院する患者の特性が大きく異なっている可能性がある．また，近年注目を集めている VisionLanguage のようなマルチモーダルな問題設定では自然言語から得られる特徴を画像の生成に利用するなど，大きく質の異なるデータの間でモデルや特徴量などの「知識」をやり取りする必要がある．そのような問題に対して通常の機械学習の方法を適用すると，期待される性能から大きく悪化した結果が得られる可能性が高い．転移学習（transfer learning）とは，そのような問題においてテストデータに対して高い性能を示すようなモデルを適切に学習するための拡張された機械学習のフレームワークである．

転移学習では，一般に「診断法を開発したい」や「ある種の画像を生成したい」といった解決しようとしているタスクが定義されている領域（正確には標本空間とデータ生成分布の組）を目標ドメインと呼ぶ．また，目標ドメインのタスクを効率的に解決するために利用する別のタスク領域を元ドメインと呼ぶ．転移学習の目的は，元ドメインにおける補助タスクで得られる「知識」，例えばデータそのものやデータから抽出される特徴量，あるいは訓練された何らかの機械学習モデル，を利用して目標ドメインにおけるタスクを効率的に解決することである．

転移学習を考える上で,「いつ転移するか（When to transfer）」「何を転移するか（What to transfer）」「どう転移するか（How to transfer）」は常に意識しなければならない基本的な問題である.「いつ転移するか」では，転移学習がうまく機能する条件を考察する．一般に，転移学習は無条件で成功するということはなく，元ドメインと目標ドメインが何らかの意味で類似していることが必要であるという事実が経験的にも理論的にも報告されている．もしそのような類似性のないドメイン間で転移学習を実行してしまうと，単に目標ドメインのみで機械学習を行う場合に比べてタスク性能をより悪化させてしまう負転移（negative transfer）が起こり得る.「何を転移するか」では，元ドメインから目標ドメインへ転移させる知識を具体化する．これまで考察されている主なアプローチとしては，元ドメインのデータそのものを目標ドメインのデータとして扱う事例転移（instance transfer），元ドメインと目標ドメインのデータ間で不変な特徴表現を学習する特徴転移（feature transfer），元ドメインで訓練した機械学習モデルのパラメータを目標ドメインで利用するパラメータ転移（parameter transfer）などが挙げられる．最後の「どう転移するか」では，元ドメインから目標ドメインへの具体的な転移学習アルゴリズムを考える．転移学習は非常に多くの問題設定を含んでおり，例えば元ドメインと目標ドメインでデータ生成分布のみが異なるのか（これを異種/同質的な設定と呼ぶ），分布のみでなく標本空間もドメイン間で異なるのか（これを同種/異質的な設定と呼ぶ）によって適切な転移学習アルゴリズムは異なる．詳細は次節で説明する．

2 転移学習の手法

事例転移のアプローチでは，元ドメインのデータ（事例）そ

のものを目標ドメインのデータ（事例）として扱う．元ドメインと目標ドメインとでデータの分布が異なる状況を想定し，元ドメインのデータを重要度に従って重み付けする方法が知られている．重要度としては，元ドメインと目標ドメインにおける確率密度の比である確率密度比が用いられることが多く，密度比の推定方法が多数提案されている．KLIEP [1] は，目標ドメインにおける真の確率密度とのカルバックライブラーダイバージェンスを最小化することにより確率密度比を推定する，Kernel Mean Matching [2] は，確率密度比で重み付けした元ドメインの確率分布の積率と目標ドメインの確率分布が適合するように最適化する．一方，HEGS [3] のように，元ドメインのデータを重み付けするのではなく，両ドメインのデータを特徴空間上でクラスタリングし各ドメイン由来のデータができるだけ均等なクラスタを選択することにより，元ドメインのデータを選択する方法もある．

特徴転移のアプローチでは，異なるドメインでも有効となるように特徴空間を変化させる．変化させる方法としては，特徴次元の拡張，共通の特徴空間への変換（対称変換），一方のドメインの特徴空間から他方のドメインの特徴空間への変換（非対称変換），などがある．目標ドメインのラベルを用いるか否かで，教師ありドメイン適応と，教師なしドメイン適応に大別される．教師ありドメイン適応の手法としては，特徴次元を複製してドメイン共通の傾向とドメイン固有の傾向のそれぞれをモデルに学習させる FAM [4] がある．教師なしドメイン適応の手法としては，非対称変換により共分散を整合させる CORAL [5] とその DNN 拡張である DeepCORAL [6]，共有エンコーダの対称変換により特徴表現のドメイン由来を識別不能にする DANN [7]，DANN を逐次方式とした ADDA [8]，ドメイン固有のエンコーダを追加した DSN [9] がある．

パラメータ転移のアプローチでは，元ドメインで学習したモデルのパラメータを目標ドメインの学習に転用する．パラメータ転移の技術としては事前学習とファインチューニング，知識蒸留，メタ学習などが一般的に用いられる．

事前学習とファインチューニングでは，元ドメインのデータで学習された事前学習済みモデルのパラメータの一部または全部をファインチューニングし，目標ドメインの学習タスクに適応する．この方法では，元ドメインのモデルパラメータという事前知識を用いることにより過学習の抑制が期待でき，特に目標ドメインの学習データが不足している場合に有効性が期待できる．

元ドメインで学習したモデルから知識を蒸留する方法として知識蒸留法（Knowledge Distillation）が知られている．知識蒸留の初期の研究 [10] においては，元ドメインで学習したモデルを教師モデルとして目標ドメインにおける生徒モデルを学習する際，教師モデルと生徒モデルの予測結果の乖離度を最小化するように学習していた．FitNets や Mutual Learning などの近年の知識蒸留法 [11], [12], [13] は，両ドメインで学習したモデルの中間層の出力間の乖離度を最小化することにより，目標ドメインのパラメータ探索空間を制約し過学習を抑えることができる．さらに知識蒸留法は，大規模パラメータを持つ教師モデルから学習された知識を小規模な生徒モデルに転移するため，推論時の計算コストの減らすというメリットもある．

メタ学習は，特定の目標ドメインに限定せず，モデルの汎化性能，すなわち未知のタスクに対してモデルがどれほどうまく機能するかに着目する．メタ学習の手法は，メトリックベース，モデルベース，最適化ベースの 3 つのカテゴリに分類される．一般に，メトリックベースのメタ学習 [14] は分類問題に用いられる．メトリックベースのメタ学習の学習段階では，各タスク中の各クラスに対応する特徴ベクトルを学習し，サンプルベクトルと個々のクラスベクトルとの類似度を比較することにより，サンプルを分類している．そして，推論段階では，未知のタスクにおける未知のクラスの特徴ベクトルを抽出し，サンプルベクトルとの類似度を計算する．モデルベースのメタ学習 [15] は，以前のタスクで学習した知識を保存できるメモリメカニズムが追加されている．そして，メモリ中の知識を転移することにより，新しいタスクでの予測性能を向上させることができる．モデルベースのメタ学習 [16] は，教師あり学習だけでなく，強化学習にもよく応用されている．最適化ベースのメタ学習の目的は，汎化性能の良いモデルの初期パラメータを学習することである．そして，このパラメータをファインチューニングするだけで，未知のタスクに最適なモデルを学習することができる．このため，モデルベースのメタ学習は，タスクごとの学習データが少ないという Few-shot learning の問題によく応用されている．

異種転移では入力空間が異なるケースを扱う．入力空間の違いに対処するため，必然的に特徴転移のアプローチをとる．変換の学習には，ドメイン間で共通のラベルを用いる場合，ドメイン横断で対応関係のある特徴ペア（ブリッジインスタンス）を用いる場合，いずれの情報も用いない場合がある．ラベルを用いる手法としては，FAM においてドメイン共通の次元を対称変換により構成する HFA [17]，各ドメインからの事例ペアのラベルの一致・不一致を予測するような非対称変換を求める Arc-t，DANN のエンコーダをドメイン固有にし目標ドメインのラベルを前提とした HeDANN [19] がある．ブリッジインスタンスを用いる手法としては，カーネル正準相関分析による共通潜在空間で SVM を学習するときに相関係数で正則化する CT-SVM [20] がある．いずれの情報も用いない手法としては，再構成損失と共通潜在空間の乖離度を最小化するよう行列因子分解する HeMap [21] がある．

3　まとめと今後の展望

本稿では，転移学習の問題設定や手法の分類を説明した後，「何を転移するか」に着目して事例転移，特徴転移，パラメータ転移に関する具体的な手法を説明した．さらに発展として，データの標本空間も異なる異種/異質的な設定での転移学習の手法についても説明した．本稿では一対一のドメイン間での転移学習を中心に解説したが，複数の元ドメインがある状況（マルチソース転移 [22]），第三のドメインが媒介する状況（遷移的転移 [23]），逐次的に来訪するタスクに対し繰り返し転移が

行われる状況（Lifelong Learning [24]， 継続学習 [25]）など，実際的な各種状況に対応した手法についても研究が進んでいる．転移学習のライブラリとして ADAPT [26] などのライブラリも出ており，実装や比較評価がしやすい環境が整いつつある．また，転移学習の手法選択 [27] や負の転移への対処も課題であり，今後さらなる発展が期待される．

4 謝　　辞

本研究・調査は，JST CREST JPMJCR21F2 の一部支援を受けたものである．

文　　献

[1] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bnau, and M. Kawanabe, “Direct importance estimation for covariate shift adaptation,” Annals of the Institute of Statistical Mathematics, vol. 60, no. 4, pp. 699–746, 2008.

[2] A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, and B. Schlkopf, “Covariate shift by kernel mean matching,” Dataset shift in machine learning, MIT Press, pp. 131-160, 2009.

[3] X. Shi, Q. Liu, W. Fan, Q. Yang, and P.S. Yu, “Predictive modeling with heterogeneous sources,” SDM, pp. 814–825, 2010.

[4] H. Daum´ e III, “Frustratingly Easy Domain Adaptation,” in Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007, pp. 256-263.

[5] B. Sun, J. Feng, and K. Saenko, “Return of Frustratingly Easy Domain Adaptation,” Proc. AAAI Conf. Artif. Intell., vol. 30, no. 1, 2016.

[6] B. Sun and K. Saenko, “Deep CORAL: Correlation alignment for deep domain adaptation,” ECCV 2016 Work. ECCV 2016. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2016.

[7] Y. Ganin et al., “Domain-Adversarial Training of Neural Networks,” J. Mach. Learn. Res., vol. 17, no. 1, pp. 2096-2030, May 2016.

[8] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial Discriminative Domain Adaptation,” Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2962-2971, Feb. 2017.

[9] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, “Domain Separation Networks,” in Advances in Neural Information Processing Systems 29, 2016, pp. 343-351.

[10] G. E. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” CoRR, vol. abs/1503.02531, 2015.

[11] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, “Fitnets: Hints for thin deep nets,” in 3rd International Conference on Learning Representations, 2015.

[12] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu, “Deep mutual learning,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 4320–4328, 2018.

[13] J. Yim, D. Joo, J. Bae, and J. Kim, “A gift from knowledge distillation: Fast optimization, network minimization and transfer learning,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 7130–7138, 2017.

[14] J. Snell, K. Swersky, and R. S. Zemel, “Prototypical networks for few-shot learning,” in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, pp. 4077–4087, 2017.

[15] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. P. Lillicrap, “Meta-learning with memory-augmented neural networks,” in Proceedings of the 33nd International Conference on Machine Learning, JMLR Workshop and Conference Proceedings, vol. 48, pp. 1842–1850, 2016.

[16] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in Proceedings of the 34th International Conference on Machine Learning, PMLR, vol. 70, pp. 1126–1135, 2017.

[17] W. Li, L. Duan, D. Xu, and I. W. Tsang, “Learning With Augmented Features for Heterogeneous Domain Adaptation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 6, pp. 1134-1148, Jun. 2014.

[18] B. Kulis, K. Saenko, and T. Darrell, “What You Saw is Not What You Get: Domain Adaptation Using Asymmetric Kernel Transforms,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1785-1792, 2011.

[19] K. Yonekawa et al., A Heterogeneous Domain Adversarial Neural Network for Trans-Domain Behavioral Targeting, vol. 11607 LNAI, 2019.

[20] Yi-Ren Yeh, Chun-Hao Huang, and Y.-C. F. Wang, “Heterogeneous Domain Adaptation and Classification by Exploiting the Correlation Subspace,” IEEE Trans. Image Process., vol. 23, no. 5, pp. 2009-2018, 2014.

[21] X. Shi, Q. Liu, W. Fan, P. S. Yu, and R. Zhu, “Transfer Learning on Heterogenous Feature Spaces via Spectral Transformation,” in Proceedings of the IEEE International Conference on Data Mining, pp. 1049-1054, 2010.

[22] Z. Xu, and S. Sun, “Multi-source transfer learning with multi-view adaboost,” In International conference on neural information processing, pp. 332–339, 2012.

[23] B. Tan, Y. Song, E. Zhong, and Q. Yang, “Transitive Transfer Learning,” in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1155-1164, 2015.

[24] Z. Chen, and B. Liu, “Lifelong Machine Learning,” Morgan Claypool, 2018.

[25] M. Delange et al., “A continual learning survey: Defying forgetting in classification tasks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[26] A. de Mathelin, F. Deheeger, G. Richard, M. Mougeot, and N. Vayatis, “ADAPT: Awesome Domain Adaptation Python Toolbox,” arXiv preprint arXiv:2107.03049, 2021.

[27] W. Ying, Y. Zhang, J. Huang, and Q. Yang, “Transfer learning via learning to transfer.” In International Conference on Machine Learning, PMLR, pp. 5085–5094, 2018.