

敵対的訓練によるドメイン適応を用いた転移学習手法の

HR 分野への適用

行武 俊秀 櫻井 義尚

明治大学 総合数理学部

1 はじめに

近年、様々なデータに対して機械学習を適用することが多くなっている。しかしながら、機械学習の精度を上げるためには十分な量のデータが必要であり、データを大量に収集することが困難なドメインに対しては、機械学習によるアプローチをとることが難しい、という課題がある。

データが十分でない場合に機械学習の精度を向上させる手段として、転移学習という手法が取られることがある。本研究では、データの十分な入手が難しいドメインである従業員の離職予測タスクにおける、転移学習を用いた機械学習モデルの提案を行う。

2 関連研究

2019年のNeil Brockettらの論文[3]では、従業員の離職しやすさを離職スコアとし、特徴量のクラスラリングやパターンマイニングといった手法で得られた結果を結合することでスコアを求めた。提案されたフレームワークでは、離職スコアの高い職員について、特徴量の類似した他職員との比較を用いて、そのスコアを改善させるようなアプローチの提案まで行うことを新規性としている。

従業員の離職というデータのドメインの性質上、実際のアプリケーション運用に重きを置いた研究に用いられることが多い。一方、このデータセットにも見受けられるような、HRデータが持つ機械学習への適用時の課題である、データ数の少なさに焦点を当てるような論文は非常に少ない。

3 提案手法

本論文では、少数データでの機械学習に用いる転移学習を用いて精度の向上を図る。

転移学習とは、あるタスクを解く際に、異なるタスクの知識を活用することで、効率的かつ

効果的に、目標タスクに対して有用な知識を得るような機械学習手法の総称である。

転移学習の課題として、解きたい問題と近いドメインのデータが必要であること、解きたい問題のデータと同じ特徴量のデータが必要であることの2つが挙げられる。本論文では、こうした課題に対して、国勢調査から得られた収入予測モデルが、HRタスクに対して元タスクとなりうることを示すとともに、特徴量の定義域が異なる問題について、共通する小数の一部分のみでも学習の効率化が可能であることを示す。

本論文では、元タスクを収入予測、目標タスクを離職予測としている。それぞれのタスクには、年齢や学歴、学歴といった一部共有可能な特徴量がある。これらの共有可能な特徴量部分を個別で抜きだし、その部分について転移学習を行う。また、元タスクと目標タスクで扱う問題が異なるため、ラベルも異なるものを取り扱う。そのため、ニューラルネットワークの浅い層を特徴抽出器として取り出し、その部分を転移学習に用いる。これは、ネットワーク内の浅い層が特徴量の抽象的な要素の取得をするという仮定に基づくものである。また、元ドメインが学習した層が、本来異なるタスクを解くためのものであることから、目標タスクに適化するために、後述する敵対的ドメイン適応を用いることで、より目標タスクにより適化された特徴抽出器の作成を目指す。

ドメイン適応：機械学習タスクにおいて、訓練データとテストデータで、それぞれのデータがサンプルされた母集団が異なる場合に、学習モデルをテストデータに適用する際に精度が低下することがある。そのとき、テストデータに適したモデルになるような変換や、損失関数の導入を行うことで、よりテストデータに対して高精度なモデルの構築を目指す手法をドメイン適応という。本実装のドメイン適応には、DANN[4]のアーキテクチャの損失部分に敵対的学習[6]を導入したモデルを用いる。本モデルは敵対的ドメイン適応を行いながら、同じく敵対的手法を用いた ADDA[5]と比較して、学習のステップ数を一つ減らし、簡略化されたものとなって

Transfer learning with adversarial domain adaptation to human resources

Yukitake Toshihide Sakurai Yoshitaka
Meiji University

いる。事前学習した元ドメインの特徴抽出器を、元ドメインと目標ドメイン間で共有し、同一の層からの出力を、それがどちらドメインのものかドメイン識別器が識別する。

本学習部分では、転移学習可能な特徴量部分を中間層で分離させ、その部分に事前学習した特徴抽出器を挿入する。

実装に際して、転移学習可能な特徴量部分のみを分離させるモデル A (図 2-(a)) と、特徴量の種類ごとに中間層を分割するモデル B (図 2-(b)) の 2 種を提案する。

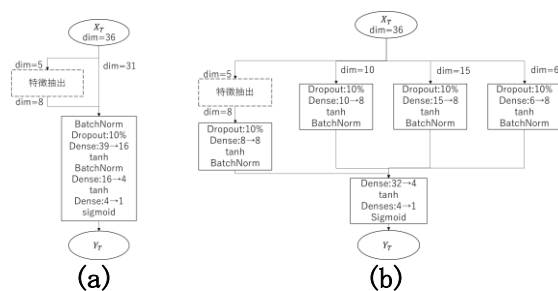


図 2 モデル図

4 実験

実装には、kaggle にて IBM が提供している、従業員の離職予測データセット [1]を用いる。このデータセットから離職予測モデルを構築し、このモデルの精度向上を目標に、転移学習用のサブデータセットとして、UCI が提供する Adult データセット [2]を用いる。

モデル A, モデル B について、転移学習をしない場合、する場合、ドメイン適応を行う場合、転移学習部分の特徴量を削除した場合の 4 パターンで学習データ数を減らしながら比較を行う。また、既存手法との比較として、ランダムフォレストと LightGBM による予測との比較を行う。

評価には Accuracy を用い、学習回数は最大 200epoch まで回し、100epoch ごとに評価値を保存し、最も高評価だったものを評価値とする。

表 1

学習データ数	300	150	100	75
ドロップA	0.714778	0.702333	0.692852	0.685444
転移なしA	0.709667	0.700778	0.687926	0.685111
転移ありA	0.707556	0.700444	0.69463	0.689333
ドメイン適応A	0.711333	0.703556	0.69563	0.695222
ドロップB	0.697889	0.696667	0.689407	0.684889
転移なしB	0.712	0.703111	0.699556	0.687778
転移ありB	0.715556	0.702667	0.70263	0.694333
ドメイン適応B	0.724889	0.712778	0.702926	0.700889
RandomForest	0.697556	0.694667	0.685852	0.679333
LGBM	0.710889	0.702889	0.699926	0.693333

実験結果: モデル B にドメイン適応を用いた転移学習モデルが最も高精度となった。また、モデル A では少数データで転移学習が高精度となり、転移学習のデータの減少への堅牢性が見受けられた。モデル問わず、ドメイン適応は行うことで、ドメイン適応を用いない転移学習モデルの精度を上回ったため、敵対的ドメイン適応が転移学習による特徴抽出をより予測に適した形にしたと考えられる。モデル同士を比較すると、モデル B がより、部分的な転移学習およびドメイン適応の恩恵を受けやすいモデルであると推測できる。

5 まとめ

本論文では転移学習を用いた、少数データを対象としたニューラルネットワークを実装し、HR データに対して適用した。特徴量の一部のみでも転移学習が精度の向上に効果的であることを明らかにし、敵対的ドメイン適応によってより効果的な特徴抽出を可能にした。また、より部分的な転移学習に適したモデル構造を示した。今後は、現実の問題に適用された場合の実用性や、元ドメインや目標ドメインが変わることが手法の有効性にどの程度影響を与えるかが明らかにし、これから有効性を検証していく。

参考文献

- [1] IBM HR Analytics Employee Attrition & Performance. 2017. 2021 年 1 月 1 日. <<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>>.
- [2] Adult Data Set. 1996. 2021 年 1 月 1 日. <<http://archive.ics.uci.edu/ml/datasets/Adult>>
- [3] Brockett, Neil et al. "A System for Analysis and Remediation of Attrition." 2019 IEEE International Conference on Big Data (Big Data) (2019): 2016-2019.
- [4] Ganin, Yaroslav et al. "Domain-Adversarial Training of Neural Networks." ArXiv abs/1505.07818 (2016): n. pag.
- [5] Tzeng, E. et al. "Adversarial Discriminative Domain Adaptation." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 2962-2971.
- [6] Goodfellow, Ian J. et al. "Generative Adversarial Nets." NIPS (2014).