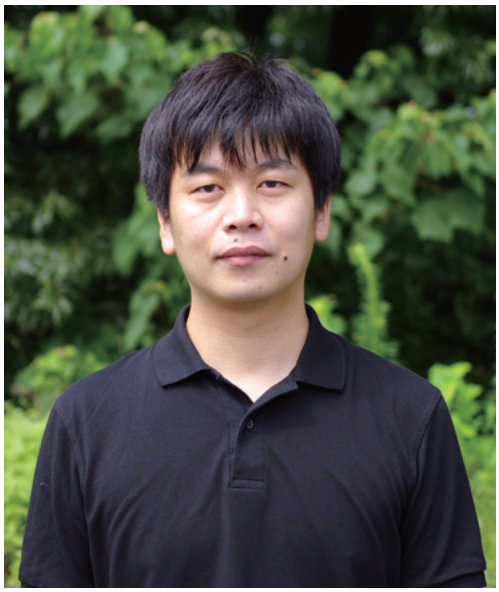
明日のトップランナー

NTTコンピュータ＆データサイエンス研究所・NTT社会情報研究所

熊谷　充敏 特別研究員

AIにも人間同様の汎用性・器用さを与える「転移学習」の研究



AIを構築するために必要な機械学習．機械学習を実施するには大量の学習データが必要です．しかし，質の高い学習データが常に必要なだけ得られるとは限りません．さらにAIの経年劣化を防ぐためには，刻々と変化するデータに合わせて定期的に再学習を行う必要もあります．今回は，理想的なデータが得られない場合においてもAIの性能向上を可能とする「転移学習」の技術について，熊谷充敏特別研究員に伺いました．

◆PROFILE：2012年日本電信電話株式会社入社，機械学習とサイバーセキュリティの研究に従事．NTTセキュアプラットフォーム研究所（2012年～2021年6月），NTTソフトウェアイノベーションセンタ（2018年～2021年6月），NTTコンピュータ＆データサイエンス研究所（2021年7月～），NTT社会情報研究所（2021年7月～）に所属．

データが足りなくてもAIの性能を向上できる「転移学習」

◆「転移学習」とはどのような技術なのでしょうか．

学習データが不足しているタスクや未知のタスクが与えられたとき，関連するタスクの学習データを活用しデータ不足を補い，性能を向上させるのが「転移学習」技術です．転移学習技術を用いることで，AIに人間同様の汎用性・器用さを与えられるのではないかと考えて研究しています．どういうことかというと，例えば，人間は料理をする，簡単な計算をする，読み書きをする，話す，走る，ボールを投げるなど，さまざまなタスクをこなすことができます．また，今まで見たことのないような新しいタスクを与えられたときにも，過去の知識や経験などをうまく活かすことで比較的少ない試行により身につけたり，適応したりすることができます．

一方，最近のAIは画像認識など大量のデータを用意できる特定のタスクでは人間と同等，あるいはある意味人間の能力を超えるような性能を出せるようになってきました．囲碁の分野で「AlphaGo（アルファ碁）」というAIが人間のトップ棋士を打ち負かしたというニュースがありましたが，それはその一例でしょう．しかし，人間が持っているような汎用性や器用さはありません．これは私だけでなく多くの研究者の共通認識だと思います．

そこで，汎用性や器用さを備えたAIを構築しようとしたときに，重要な要素となり得るのが「転移学習」技術です．汎用性を持ったAIを構築するためのもっとも単純な方法は，すべてのタスクを列挙し，それぞれのタスクに対して大量の学習データを用意することでしょう．しかし，それを実行する場合，「すべてのタスクを列挙する」のがまず難しく，仮にできたとしてもそれぞれのタスクに対して大量の学習データを集めるのは，コストとリソースを考えると現実的ではありません．そこで，「転移学習」技術を用いることで，学習データが不足しているタスクや未知のタスクでも，関連する学習データを活用しデータ不足を補い，性能を向上させられますので，汎用性・器用さを備えたAIの構築に役立つのではと考えています．

◆具体的にはどのような研究をされているのでしょうか．

1番目は「正常データしか得られない状況で適切な異常検知器を高速に生成する」技術の研究です．「異常検知」は，正常データとは異なる性質を持つ「異常データ」を見つけ出すタスクです．通常，高精度な異常検知器を作成するためには，学習データとして正常データ群，異常データ群の両方が必要です．ところが，異常データはもともと非常に希少なもので実際にはなかなか手に入らないため，異常検知器の学習に利用できないことが実用では多いです．

そこで，正常データしかない状況でも，正常データ・異常データの両方を含む「似たような」データセットを活用し，異常検知

器を生成しようというのがこのアプローチです．**図1**は，正常データ・異常データを含む複数の関連データセット（1,2,…）を活用し，正常データ群から異常検知器へのマッピングをNNモデル（ニューラルネットワークモデル）であらかじめ学習しておき，異常データのない目標データセットを学習済のNNモデルに入力することにより，適切な異常検知器を生成するイメージ図です．

通常のAI学習では，新たな問題が提示されるたびに再学習という非常に計算コストの高い処理を行う必要がありますが，この方法では新しいデータセットをNNモデルに入力するだけで異常検知器を生成できるため，リアルタイム性を求められるようなケースや計算リソースが限られているなかで異常検知を行いたいケースなどにフィットするのではないかと思います．

2番目は「時間変化するタスクのための転移学習」技術の研究です．機械学習ではデータを分類する「分類器」というものを学習します．ところがデータは刻々と変化しますから，それに伴って分類器も変化していきます．こういった変化は現実社会の問題においても非常に多く起きていて，例えばマルウェア検知器の場合，攻撃者は検知器をかいくぐろうと次々に新たな攻撃手法を編み出します．そのためいったん学習したマルウェア検知器をそのまま使い続けていると，分類精度がどんどん劣化してしまいます．

分類器の精度の劣化を防ぐ有効な対策として，最新の追加データを利用して分類器を適時アップデートしていく方法があります．この方法はシンプルで非常に効果的ですが，追加データを収集するコストがかかります．例えば画像データには「この画像は猫」「この画像は犬」などラベル付けが必要となりますが，これらの作業は基本的に人間の手で行うためコストがかかります．また，それ以外の要因，例えば個人情報保護の観点などから，ラベル付きのデータが得にくいような場合もあるでしょう．

そこで，ラベルありデータを使用して学習した決定境界から，追加のラベルありデータを使用することなく未来の決定境界を予測しようというのがこのアプローチです．**図2**でいうと，ラベル付きのクラスA，クラスBのサンプルがある決定境界が時間変化に伴い3段階に変化したとして，その先の決定境界の変化を推測しようというものです．転移学習では通常，適用先にデータを仮定するものですが，この研究では追加のラベル付きデータを用いないという点が面白いところです．

◆現在の課題は何でしょうか．

一番の課題は，学習したモデルが正しいことをいかに担保するかです．一般の機械学習では，学習したモデルの妥当性を確認するために検証用のデータを別に用意します．検証用のデータをどれだけ正しく判別できるかで学習の正確さを評価できるわけです．

ところが今回紹介したような研究では，そもそも検証用のデータを作成しにくい，もしくは検証用データが存在しないというこ

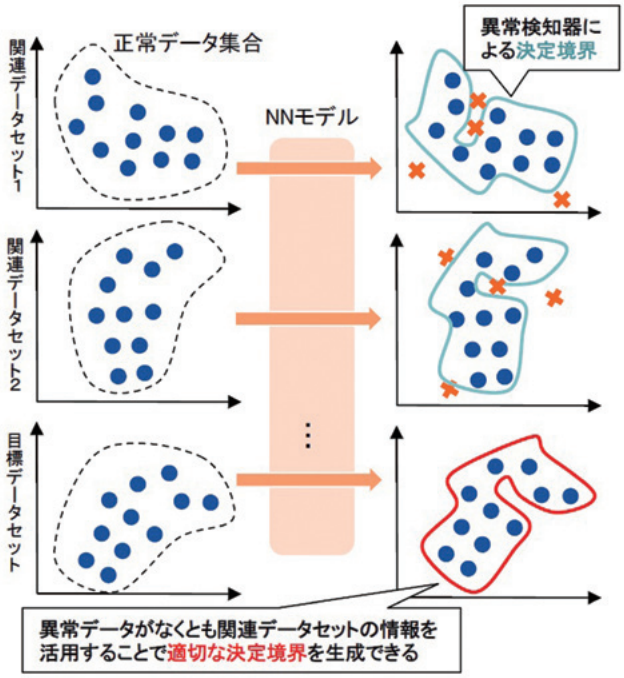


図1　正常データのみから異常検知器を高速に生成

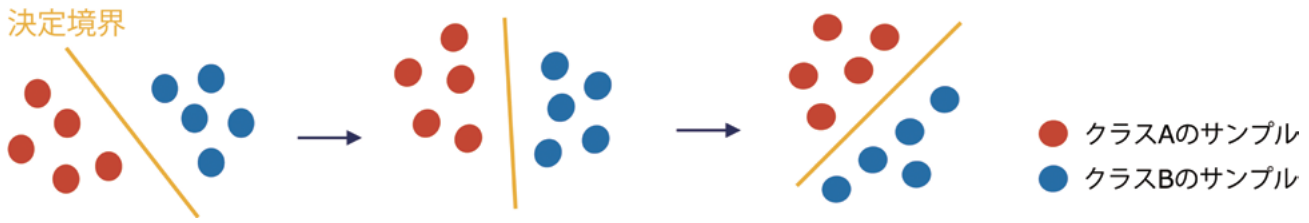


図 2　時間変化による分類器の決定境界の変化

とも起こり得ます．特に未来予測については検証用データがない状態となるため，予測したものがどれだけ正しいのかを判断できないことが難しいところです．

今後の実用化，特にミッションクリティカルな分野への応用を考えたときには，安全性も担保しなければならないため，この点は大きな課題となります．

将来的には汎用性・器用さを備えたAIの構築をめざす

◆本研究はどのような分野に応用が可能なのでしょうか．

「異常検知」の分野でいうと，例えば複数の工場を所有するメーカがあったとして，新たな工場を建設し，そこでの機器監視業務を自動化するようなAIをつくりたいとします．これは機器の正常データとは異なる性質を持つ「異常データ」を見つけ出す異常検知器をつくることに相当します．当然，新工場には稼働実績がないため，希少な機器の異常データが十分集まるまで性能の良い異常検知器をつくれないかもしれません．しかし，ほかに長期間稼働している工場があれば，そこでの正常データ・異常データのデータセットを活用し，新工場の正常データを入力すれば即座に高精度な異常検知器を作成することができます．また，複数顧客のネットワークのセキュリティを送信サービスにより一元管理するような場合では，正常データを入力するだけで顧客に応じた条件を自動的に作成することもできます．

（今回はリモートにてインタビューを実施しました）

「時間変化するタスクのための転移学習」技術でいうと，先ほども例に挙げたセキュリティの面では自動更新するアンチウイルスソフトなどへの応用も考えられます．また，企業によってはマルウェアなどから自社を守るために接続してはいけないサイトの「ブラックリスト」のようなものを使って運用することがありますが，そのブラックリスト作成に私たちのつくった技術を使うことも考えられます．

そのほかにも，eコマースへの応用なども考えられますね．年齢とともに変化する顧客の趣味・嗜好を的確に予測し適切な商品を推薦するとか，新規ユーザや利用頻度の少ないユーザのニーズを転移学習により推測するなどです．いきなり長期にわたって予測し続けることは難しいかもしれませんが，更新が必要となるまでの時間を少しだけ延ばしコストを削減できる可能性はあると思います．

複数のデータセットがある状況で精度を高めるという研究は汎用的なものですので，応用範囲は広いのではないかと思います．

◆今後の研究の方向性について教えてください．

現時点でも，画像認識や言語処理など，転移学習をうまく活用して実用レベルの精度を確保できている分野はあります．しかし，可能性が見逃され，転移学習をうまく使えていないブルーオーシャン的な分野も存在するのではないか，と考えています．中期的な目標として，そのような分野を重点的に探っていき，転移学習の適用範囲を広げるような研究を行っていきたいと思っています．学習データの不足は結構いろいろな分野で問題になり得ますから，転移学習がそういった実用的な問題を解決できる手段になるのではないかと期待しています．

さらに長期的には，最初に申し上げたような汎用性・器用さを備えたAIを構築したいと考えています．新しい問題やタスクが生じるたびに人間がAIに学習やチューニングを行うのではなく，AIが勝手に学習し1つのモデルで何でもできるようになれば面白いですね．