

戦略的創造研究推進事業 CREST
研究領域「ビッグデータ統合利活用のための次世
代基盤技術の創出・体系化」
研究課題「膨大なマルチメディアデータの理解・要
約・検索基盤の構築」

研究終了報告書

研究期間 2014年 10月～2020年 3月

研究代表者：原田 達也
(東京大学先端科学技術研究センター、
教授)

§ 1 研究実施の概要

(1) 実施概要

本研究課題の目的は、膨大なマルチメディアデータを統合的に理解・要約する基盤技術の構築にある。膨大なマルチメディアデータを理解・要約する基盤技術構築には、(a) データのコンテンツ認識アルゴリズム、(b) 認識を支えるための知識獲得、(c) 認識技術の有益さの検証、の3つの要素が必要不可欠である。さらに、これらを容易に実行するための(d) 認識実行フレームワークの構築も大切である。これらの項目の成果を以下に述べる。

(a) データのコンテンツ認識アルゴリズム

膨大なマルチメディアデータを効率よく要約するためには、重要部分の切り出しが重要である。しかし、変化前と後のデータの生成確率分布をそれぞれ推定し、それらを比較することによって重要部分を見つけるという標準的なアプローチでは、問題の次元が上がるにつれて検知性能が著しく低下するという問題があった。そこで、生成確率分布を明示的に同定することなく、統計的な依存関係の変化を直接検出するという全く新しいアプローチを考案した。これは密度比推定と呼ばれる技術に基づいており、高次元データに対しても効率よく計算できる実用的なアルゴリズムを開発するとともに、変化が低次元である限りは元のデータの次元数に依らずに精度良く変化が検知できることを理論的に示した。

切り出したデータの要素理解・認識においては、正負両方のクラスからデータを集めるのが困難、あるいは、多大なコストがかかることがある。そこで、正のクラスとラベルなしのデータだけから正しく分類境界を学習できる画期的な学習アルゴリズムを開発し、正負両方のクラスのデータを用いた場合と同程度か、条件によってはそれを上回る漸近性能を達成できるという非自明な理論的保証を得ることに成功した。また、この手法を一般化し、間違っただけのラベルがついたデータのみからの学習、信頼度のついた正例のみからの学習、類似データのみからの学習、2セットのラベルなしデータのみからの学習など、様々な問題設定に対する弱教師付き学習手法を提案し、漸近的には教師付き学習と同程度の分類性能が得られることを理論的に示した。

同様に、限られた教師データしか存在しない状況で、認識に頻繁に利用される深層ニューラルネットワーク(DNN)の分類性能を最大限に引き出す学習手法の開発は重要課題である。そこで、異なる二つのクラスからパターンをランダムな比率で混合し入力パターンを作成し、混合した二つのカテゴリと比率を教師情報として与えることで DNN を学習させるクラス間学習を提案した。この手法により、画像認識や環境音認識において大幅な性能向上を得られることが分かった。また、クラス間学習は、データ拡張の一種とも考えられるが、学習効率が悪い場合がある。そこで、幾何学的不変特徴を理論的に設計し、この不変性を事前知識としてモデルに組み込むことにより効率的に学習可能な手法を考案した。

さらに、時間も考慮した要素間関係性を把握することも認識には大切である。つまり、実世界情報から原因と結果の関係である因果関係を推定することは認識には重要となる。そこで、従来の因果指標の多くはある条件下で偏正準相関分析に帰着されることを活用し、実世界では頻出するデータ欠損にロバストな因果指標の推定手法を構築した。

(b) 認識を支えるための知識獲得

予測モデルを学習した状況と、学習した予測モデルを適用する状況が微妙に異なる場合でも、予測性能が大幅に低下する問題は、実世界では頻出する。そこで、ソースドメインには教師ありデータが存在し、ターゲットドメインには教師ありデータが存在しない教師なしドメイン適合の研究を進めてきた。これは、ソースドメインとしては大量の教師付きデータが容易に生成可能なシミュレーション環境、ターゲットドメインとして人手でのアノテーションが困難な実環境などが想定される。本プロジェクトでは、二つの分類器の予測不一致度を利用した高い適合性能をもつ手法などを構築した。さらに、ドメイン適合を物体検出問題や画像生成モデルに応用した新規手法も開発した。

また、世の中に存在するあらゆる物体カテゴリをあらかじめ学習させることは不可能であるため、カテゴリが学習されていない物体に関する情報をシステムが自動的に得る方法が必要である。画

像認識システムが新しいカテゴリを学習する方法の一つとして、未知物体について人間に尋ねることが考えられる。そこで本プロジェクトでは、学習していないカテゴリに関する知識獲得の手段として、画像中の未知物体に関する自然言語での質問を自動的に生成する手法を構築した。

さらに、深層学習の発展により画像認識の認識精度は向上したが、その動作原理がブラックボックスであることから、なぜそのような認識結果になったのか説明可能なシステムが求められている。そこで、入力パターンが与えられると、特定のカテゴリに分類するだけでなく、言語と視覚情報の両方による説明を出力するシステムを構築した。

(c) 認識技術の有益さの検証

認識技術の有益さの検証は、人間行動データ、パーソナルファブ리케이션データ、ロボット等の自律能動型機械から獲得されるデータ等、への適応を通じて行った。

大量の人間行動データの取り扱いに必須である、データ間の関係性に注目した学習モデルを開発した。スマートフォン端末から得られる大規模位置履歴情報に基づく活動人口の時空間パターンのモデル化を題材に、生成的・識別的双方の観点から多タスク学習モデルを開発した。また、実世界情報処理において問題となるデータ収集のコストに注目し、データ収集後の学習手法の改善のみならず学習に用いるデータの品質を向上する、効率的なデータ収集手法を開発した。スマートフォン端末で得られる大規模位置履歴情報のモデル化について研究を進めるとともに、産学連携を進め、商用ウェブサービスへの技術移転を達成した。

パーソナルファブ리케이션における実世界のモノづくり過程をマルチモーダルに蓄積・共有・再現する基盤システムの構築を進めた。具体的には、大規模データに対応したサーバ、手軽にデータ収集が可能なスマートフォンクライアント、及び一般的なブラウザからデータ閲覧と編集が可能な Web クライアントをクラウド上に実装した。さらに、モノづくりデータ収集を行う支援システムを構築した。一例として、両腕に付けたスマートウォッチから動作データを収集し、基盤システムとして連携可能なデータ収集とタグ付けシステムを実装した。さらに、FabLab等のコミュニティと連携してデータ収集と、認識検索や手順書作成の支援等を進めた。

自律型の探査システムを用いて、災害に関係する情報を能動的に獲得、実時間で状況を認識し、地図として情報共有する基盤システムの構築を進めた。具体的には、クローラロボットや、小型自動運転車両、サイバー救助犬、球殻ドローンなどの探査システムに搭載した LiDAR やカメラや IMU や GNSS などのセンサを利用して、能動的に実世界の情報を収集・蓄積することに取り組んだ。また、収集したヘテロなデータを確率ロボティクスや機械学習を利用して、3次元の地図に統合し可視化した。異なる日のデータを比較することで意味のある情報を抽出する機能を開発した。成果の一部として、他車両の観察に基づく駐車可能な場所の特定、イヌの探査軌跡から土砂の堆積高を推定する方法の成果が得られた。また、本プロジェクトで提案したアルゴリズムを応用した3次元点群の密なセマンティクスの認識と知識の抽出にも取り組んだ。

(d) 認識実行フレームワークの構築

実世界情報を高い精度で認識するには、深層ニューラルネットワーク(DNN)を利用することが一般的になっているが、その計算負荷の高さや環境構築の煩雑さが応用の際の問題となっている。この問題を解決するために、WebDNN というウェブブラウザを利用したインストール不要な DNN 実行環境を開発した。DNN の学習フェーズでは学習に必要な中間データを保持しておく必要があるが、推論フェーズではこれらのデータを保持しておく必要はない。このように、推論フェーズに特化することで、より積極的な計算グラフ最適化が可能となる。WebDNN はユーザエンドデバイスでの推論フェーズ実行に特化しており、学習済みモデルの積極的な最適化をサポートしている。

(2) 顕著な成果

< 優れた基礎研究としての成果 >

1.

概要:

マルチメディアデータの認識システム構築には、高品質な教師データの作成が必要不可欠である。しかしながら、画像のように対象物体以外の物体が不可避に映り込むようなデータでは、正例は集められても負例を正確に収集することは事実上不可能である。このように膨大なマルチメディアデータの分類問題では、正負両方のクラスからデータを集めるのが困難、あるいは、多大なコストがかかることがある。そこで、正のクラスとラベルなしのデータだけから正しく分類境界を学習できる画期的な理論的枠組を考案し、正負両方のクラスのデータを用いた場合と同程度の漸近性能を達成できることを理論的に証明した。更に、凸最適化に基づく高速かつ実装が容易である実用的なアルゴリズムを開発した。

Marthinus Christoffel du Plessis, Gang Niu, Masashi Sugiyama. Convex formulation for learning from positive and unlabeled data. In F. Bach and D. Blei (Eds.), Proceedings of 32nd International Conference on Machine Learning (ICML2015), JMLR Workshop and Conference Proceedings, vol.37, pp.1386-1394, Lille, France, Jul. 6-11, 2015.

2.

概要:

従来の正例とラベルなしデータからの学習技術は、深層ニューラルネットワークに対しては、訓練誤差が負の値を取ってしまう傾向があり、過適合によりテスト誤差が大きくなっていく傾向があった。それに対して、修正を施した新手法では、過適合が抑えられ、通常の教師付き学習(PN)よりもテスト誤差が小さくなり得ることを示した。

Ryuichi Kiryo, Gang Niu, Marthinus C. du Plessis, Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. NIPS2017, Oral Presentation (採択率: 40/3240=1.2%)

3.

概要:

ラベル付きデータが大量に入手可能なソースドメインで獲得した知識を教師データが得られないターゲットドメインに転移する教師なしドメイン適合の新規手法を実現した。従来はソースとターゲットの特徴分布をカテゴリ情報を無視して一致させる手法が主流である。一方、提案手法は二つの分類器の予測不一致度を利用することで、ソースにおけるカテゴリ毎の分布にターゲットの分布を一致させるように特徴抽出器を学習可能であり、カテゴリ情報を反映したより適切なドメイン適合が実現できる。

Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, Tatsuya Harada. Maximum Classifier Discrepancy for Unsupervised Domain Adaptation. The 31st IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp.3723-3732, 2018. (oral, 採択率: 70/3309=2.1%)

< 科学技術イノベーションに大きく寄与する成果 >

1.

概要:

深層ニューラルネットワークの新しい学習方法としてクラス間学習を提案した。クラス間学習では、異なるクラスに属する二つのデータをランダムな比率で混合して、クラス間データを生成する。次に、混合データをモデルに入力し、混合比を出力するようにモデルをトレーニングする。クラス間学習により、教師データのバリエーションが増加されるだけでなく、特徴空間におけるフィッシャー基準の拡大と、クラスの特徴分布間の位置関係の正則化が実現される。実験結果、クラス間学習が画像認識と音声認識のさまざまなネットワーク、データセットにおいて認識精度が大幅に改善されることが示された。

Yuji Tokozume, Yoshitaka Ushiku, Tatsuya Harada. Between-class Learning for Image Classification. The 31st IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp.5486-5494, 2018.

Yuji Tokozume, Yoshitaka Ushiku, Tatsuya Harada. Learning from Between-class Examples for Deep Sound Recognition. 6th International Conference on Learning Representations (ICLR), 2018.

2.

概要:

構築した大規模人行動データに基づく人口パターン予測法は、本研究課題でこれまでに開発した技術を展開したものである。一方、この従来技術は、民間企業提供の検索サービスにおいて、街中で発生する様々なイベントを予測する技術として活用され、人口パターン予測の社会実装が進みつつある。本手法は従来法に比べ空間分解能の飛躍的な向上をもたらすものであり、都市・社会の関連分野でさらなる活用が期待できる。

Masamichi Shimosaka, Yuta Hayakawa, Kota Tsubouchi. Spatiality preservable factored Poisson regression for large scale fine-grained GPS-based population analysis. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), pp.1142-1149, February 2019.

3.

概要:

近年、深層ニューラルネットワーク(DNN)が様々なタスクで著しい成果をあげ注目されているが、計算負荷の高さがアプリケーション応用の際の問題となっている。ハードウェアアクセラレーションによる解決方法では、煩雑な計算環境セットアップ手順やハードウェアの価格等が問題となり、ユーザエンド端末に同様の環境を構築することは非常に困難である。そこで本研究では、ウェブブラウザを利用したインストールフリーな DNN 実行環境である。WebDNN を開発した。WebDNN により、ウェブブラウザ上での実行を前提とした積極的な最適化が学習済みモデルに行われるため、パラメータデータの配信サイズを小さく抑え、高速な実行が可能となる。既存のライブラリと比較して大幅な高速化を達成した。

Masatoshi Hidaka, Yuichiro Kikura, Yoshitaka Ushiku, Tatsuya Harada. WebDNN: Fastest DNN Execution Framework on Web Browser. ACM Multimedia Open Source Software Competition, pp.1213-1216, 2017.

<代表的な論文>

1. Marthinus Christoffel du Plessis, Gang Niu, Masashi Sugiyama. Convex formulation for learning from positive and unlabeled data. In F. Bach and D. Blei (Eds.), Proceedings of 32nd International Conference on Machine Learning (ICML2015), JMLR Workshop and Conference Proceedings, vol.37, pp.1386-1394, Lille, France, Jul. 6-11, 2015.
2. Ryuichi Kiryo, Gang Niu, Marthinus C. du Plessis, Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. NIPS2017, Oral Presentation (採択率:40/3240=1.2%)
3. Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, Tatsuya Harada. Maximum Classifier Discrepancy for Unsupervised Domain Adaptation. The 31st IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp.3723-3732, 2018. (oral, 採択率:70/3309=2.1%)

§ 2 研究実施体制

(1) 研究チームの体制について

① 原田グループ

研究代表者: 原田 達也 (東京大学先端科学技術研究センター 教授)

研究項目

- ・研究全体の統合・統括
- ・膨大なマルチメディアデータを認識する手法の開発
- ・獲得したデータ間の関係性を発見し記述・要約する手法の開発

② 杉山グループ

主たる共同研究者: 杉山 将 (東京大学大学院新領域創成科学研究科 教授)

研究項目

- ・マルチメディアデータの認識技術を支える数理的基盤の構築

③ 大野グループ

主たる共同研究者: 大野 和則 (東北大学未来科学技術共同研究センター 准教授)

研究項目

- ・能動的センシングにより獲得される実世界データの理解

④ 塚田グループ

主たる共同研究者: 塚田 浩二 (公立はこだて未来大学システム情報科学部 准教授)

研究項目

- ・パーソナルファブリケーションにより獲得される実世界データの理解

⑤ 下坂グループ

主たる共同研究者: 下坂 正倫 (東京工業大学情報理工学院情報工学系 准教授)

研究項目

- ・人の行動データにより獲得される実世界データの理解と能動的知識獲得

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

ソルボンヌ大学・IRCAM の Gérard Assayag 教授, Philippe Esling 准教授の音楽情報処理チームとそれぞれの指導学生を数名, 約半年間交換留学してマルチメディア情報処理に関する共同研究を実施している. 具体的には Erasmus Mundus プログラムを利用して, ソルボンヌ大学の博士課程の学生 1 名が原田チームに約半年滞りし, また, 原田チームの修士学生 2 名が Esling 准教授のチームの研究室に約半年間滞在した. 本年度の Esling 准教授の来日時には理研 AIP にて講演を行った. さらに, Assayag 教授と Esling 准教授は, JST-ERC 間研究者交流推進を利用して, 今年度数か月間原田チームの研究室に滞在予定である. また, 日仏の共同研究プロジェクトの提案準備を行っている.

インド工科大学ハイデラバード校 (IITH) の V. N. Balasubramanian 准教授と原田グループは, 過去3年ほど, 相互訪問, 短期学生インターンシップ, 最近の合同ワークショップなどのさまざまなチャネルを通じて少数データからの認識モデルの学習方法に関して, 定期的に交流を行ってきた. 具体的には以下のものがあげられる.

- Balasubramanian 准教授は, 2018 年 2 月に原田グループの研究室を訪問して講演を行った.
- 続いて Balasubramanian 准教授は, 2018 年 10 月にコラボレーションの議論のために原田グループの研究室に二度目の訪問を行った.

- 原田は、2019年3月に IITH-RIKEN AI ワークショップの一環として講演を行うためにために IIT-Hyderabad を訪問した。このワークショップは、IIT-Hyderabad と理研 AIP (原田は理研 AIP のチームリーダーの立場でもある) の研究者間のコラボレーションを模索するための特別な取り組みである。
- JST さくらサイエンスプランとホンダ YES フェローシップを通じて、Balasubramanian 准教授の大学院生と学部生の両方の複数の学生が原田グループの研究室で短期間インターンシップを実施した。

中国の電子科技大学から、Yanli Jie 准教授を原田チームの研究室に1年間招いて、人間の動作認識に関して共同研究を行っている。共同研究の成果は、マルチメディアのトップ会議である ACM Multimedia2019 に採択されている。また、原田は2019年5月に電子科技大学を訪問して講演を行っている。

大阪大学の柳澤教授が PI を務める、JST さきがけのプロジェクト「思考・行動を予想する脳ビックデータ」と共同研究を行った。原田グループが開発した環境音認識のアルゴリズムを柳澤教授が進める脳波情報 (MEG) から の てんかん 推定 に 応用 すること で、従来手法と比較して高精度の推定結果を得ることができた。この成果は Scientific Reports にアクセプトされた。

塚田グループが進めている Fab データ収集にあたって、COI「ファブ地球社会創造拠点」のメンバーである、明治大学 渡邊恵太准教授や、慶應義塾大学 田中浩也教授と連携している。さらに、国内／アジア地域の FabLab とも連携を進めており、特に FabLab Sendai を運営する小野寺 志乃氏、大網 拓真氏とは密にデータ収集／意見交換を行っている。