T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	深層学習による画像認識における推論根拠と未知不均衡ドメイン学習に関する研究
Title(English)	
著者(和文)	桑島洋
Author(English)	Hiroshi Kuwajima
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12361号, 授与年月日:2023年3月26日, 学位の種別:課程博士, 審査員:田中 正行,奥富 正敏,蜂屋 弘之,中臺 一博,川上 玲
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12361号, Conferred date:2023/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

Powered by T2R2 (Science Tokyo Research Repository)

（論文博士）
(Dissertation Doctorate)

論　文　要　旨　（和文2000字程度）

Dissertation Summary (approx. 2000 characters in Japanese)

報告番号 For administrative use only	乙 第　　号	氏　名 Name	桑島　洋

（ 要　旨 ）
(Summary)

大量のデータからデータ駆動で複雑なモデル（膨大なパラメータを持つ深いニューラルネットワーク）を学習する深層学習は、画像認識などで成果をあげ、自動車などのセーフティクリティカルな用途でも使われ始めている。深層学習モデルの開発では、第一に、学習データとテストデータを収集する。第二に、学習データから深層学習モデルをデータ駆動で学習する。第三に、深層学習モデルが推論した結果をテストデータで検証する。深層学習は、学習データとテストデータの分布が一致することを仮定しており、この仮定が成り立てば、高い性能を発揮する。

深層学習には「複雑なモデルの限界」と「データ駆動の限界」がある。深層学習モデルは、パラメータ数が膨大で構造も複雑であり、特徴量の活性値は観測できるが、実際の現象は把握できない。複雑なモデルの限界は、推論の過程や根拠を人間が理解できないため、問題を特定、修正、改善するというエンジニアリングが難しいことである。一方、深層学習を利用した高度運転支援システムや自動運転システムで事故が起きるなど、リスクが顕在化している。これらの事故は、道路を横断する大型トレーラーの側面や深夜に道路を横断する歩行者など、事故につながる希少な環境条件（ドメイン）に遭遇したときに起きている。学習データとテストデータの分布が一致するという深層学習の仮定に対して、実際の開発では、学習データとテストデータの分布の一致は不明である。現実的には、学習データには異なるドメインの標本が不均衡に混在し、標本の元ドメインはわからない。この状況を「未知不均衡ドメイン」と呼ぶ。未知不均衡ドメインの状況で、学習データに十分に含まれていない少数派ドメインにおいて深層学習モデルの性能が低下し、事故につながったと考えられる。データ駆動の限界は、未知不均衡ドメインに対して十分な性能を発揮できないことである。

ところで、深層学習では、学習データを所与のものとしてモデルを学習するだけではなく、現実の世界（母集団）から標本を抽出し学習データを作成することが重要である。データ作成に関連する技術として、アノテーションのコスト削減を主な目的とする能動学習がある。そのため、未知不均衡ドメインへのアプローチとして、「機械学習（モデル学習）」と「能動学習」がある。

そこで本研究では、深層学習による画像認識において、推論根拠と未知不均衡ドメイン学習（機械学習と能動学習）を基礎検討する。推論根拠に関する研究の目的は、特徴量に着目して深層学習の推論過程を分析し、人間に理解できる形式で提示する手法を構築することである。未知不均衡ドメイン学習に関する研究の目的は、未知不均衡ドメインの標本を含む学習データに対して、多数派ドメインの性能を保持しつつ、少数派ドメインの性能を向上させる機械学習（モデル学習）と能動学習の手法

を構築することである。

本論文は全5章で構成され、各章の概要は以下のとおりである。

第1章「緒言」では、本研究の背景や目的、本論文の構成などを述べる。

第2章「深層学習における推論根拠」では、深層学習モデルの特徴量に着目した推論根拠の分析方法と、クラウドソーシングを用いて分析結果の推論根拠の妥当性を評価する方法を提案する。提案手法で分析した深層学習モデルの推論根拠は人間の推論過程とおおむね一致し、人間に理解できる分析結果であることを実験で示す。また、推論根拠を理解することで、モデルの拡張や学習データの追加など、改善のための手掛かりが得られる可能性も示す。

第3章「深層学習における未知不均衡ドメイン機械学習」では、未知不均衡ドメインが混在する学習データを用いて深層学習モデルを学習し、ドメイン別のテストデータを用いて深層学習モデルをドメインごとにテストする、「未知不均衡ドメイン機械学習」の問題設定を提案する。未知不均衡ドメイン機械学習の問題設定に対して、center loss（損失関数）と特徴量の空間における標本間の距離に基づいたミニバッチ抽出を組み合わせた手法を提案する。提案手法により、多数派ドメインの性能を維持しながら少数派ドメインの性能が向上することを実験で示す。

第4章「深層学習における未知不均衡ドメイン能動学習」では、第3章の内容を能動学習に拡張する。深層学習モデルの学習において能動学習が用いられる実際の状況を想定し、現実的な巨大データプールの生成と、現実的なモデル学習を考慮した実験設定を提案する。未知不均衡ドメイン機械学習を拡張し、能動学習とモデル学習を組み合わせることで未知不均衡ドメインに対処する「未知不均衡ドメイン能動学習」の問題設定を提案する。能動学習の現実的な実験設定のもとで、モデル学習と能動学習の手法の組み合わせを比較し、softmax marginによる能動学習と第3章の提案手法（モデル学習）を組み合わせることで、多数派ドメインの性能を維持しながら少数派ドメインの性能が向上することを実験で示す。

最後に、第5章「結言」で本研究のまとめと今後の課題を述べる。

備考：論文要旨は、和文2000字と英文300語を1部ずつ提出するか、もしくは英文800語を1部提出してください。
Note: Dissertation summaries must be written in either of the following formats: (A) both in Japanese (approx. 2000 characters) and in English (approx. 300 words), or (B) in English (approx. 800 words).
注意：論文要旨は、東工大リサーチリポジトリ（T2R2）にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Important: Dissertation summaries will be published online on the Tokyo Tech Research Repository (T2R2). Do not include information treated as confidential under certain circumstances.

（論文博士）
(Dissertation Doctorate)

論　文　要　旨（英文）　（300語程度）

Dissertation Summary (approx. 300 words in English)

報告番号 For administrative use only	乙 第　　号	氏　名 Name	桑島　洋

（要　旨）
(Summary)

Deep learning, which involves data-driven training of complex models (deep neural networks with a massive number of parameters) from large amounts of data, has achieved success in image recognition. Deep learning uses complex black box models where we can observe feature activation; however, we cannot understand the actual phenomenon. Therefore, a limitation of complex models is that the engineering of identifying, correcting, and improving problems is difficult because humans cannot understand the computational process and inference arguments. Furthermore, for real-world applications, deep learning is with blind imbalanced domains where training data contains multiple (source) domains; however, the domain assignment of samples is unknown and imbalanced. Therefore, a limitation of data-driven training is that it cannot perform adequately for blind imbalance domains. Thus, this dissertation will study inference arguments and blind imbalanced domain learning for image recognition using deep learning.

In Chapter 1, Introduction, we describe the dissertation's background, purpose, and document structure. In Chapter 2, Inference Arguments in Deep Learning, we propose a method to generate inference arguments based on deep features and evaluate them using crowdsourcing. Then, we discuss that the inference arguments may provide clues for improving deep learning models, such as extending the model or adding training data. In Chapter 3, Unknown Imbalanced Domain Machine Learning in Deep Learning, we propose a problem setting of blind imbalanced domain learning and a corresponding method that combines center loss (loss function) with mini-batch weighted sampling based on the distance between samples in a deep feature space. In Chapter 4, Imbalanced Domain Active Learning in Deep Learning, we extend Chapter 3 to active learning and identify an appropriate method that combines simple active learning with softmax margin and model training with the proposed method in Chapter 3. In Chapter 5, Conclusion, we summarize this dissertation and discuss future works.

備考：論文要旨は、和文2000字と英文300語を1部ずつ提出するか、もしくは英文800語を1部提出してください。
Note: Dissertation summaries must be written in either of the following formats: (A) both in Japanese (approx. 2000 characters) and in English (approx. 300 words), or (B) in English (approx. 800 words).
注意：論文要旨は、東工大リサーチリポジトリ（T2R2）にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Important: Dissertation summaries will be published online on the Tokyo Tech Research Repository (T2R2). Do not include information treated as confidential under certain circumstances.