様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ６ 月 １８ 日現在

機関番号：１７１０２
研究種目：研究活動スタート支援
研究期間：2019～2020
課題番号：１９Ｋ２４３４８

研究課題名（和文）深層学習システムの大規模展開に向けた差動テスト技術に関する研究

研究課題名（英文）Differential Testing Techniques Towards Large-scale Deployment of Deep Learning Systems

研究代表者
馬　雷（MA, LEI）

九州大学・システム情報科学研究院・学術研究員

研究者番号：７０８４２０６１

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：本研究では、当初の提案内容は全て順調に完成された。まず、深層学習(DL)の開発・展開における具体的な課題を理解するために調査を行った。その結果、深層学習システムの展開品質には多くの課題があることが分かった。これに踏まえて、不確実性の観点から差分テスト基準を提案し、DiffChaser差分テストフレームワークを提案した。多様なDL展開場面に対して系統的な評価を行い、有効性を示した。さらに、不確実性やデータ分散の観点から、DLの挙動解析手法について詳細な検討を行い、有望な結果を獲得した。本研究の成果は、今後の研究や産業界への応用に向けて、DL展開の品質保証に関する重要な基盤を構築することができた。

研究成果の学術的意義や社会的意義
近年、小型デバイスを目標としたAIチップの急速な進歩は、DLのパワーを小型デバイスにもたらす新たな機会をもたらした。しかし、DLシステムの品質保証技術に関する研究はまだ初期段階である。本研究では、DLの展開段階から多様な小型デバイスへの関連テスト基準、差分テストフレームワーク、および品質向上技術を構築し、重要だが欠けている部分を埋める。高品質のDL手法で小型デバイスを強化することで、応用の範囲がさらに拡大され、知能システムの恩恵が世界中のあらゆる社会にもたらされる。本研究の成果がDLシステム展開のための品質保証を提供し、将来の知能社会の発展を加速するための基盤と応用を築くことが期待される。

研究成果の概要（英文）：In this project, we successfully completed the planned research tasks in original research proposal.(1)At an early stage, we made a comprehensive survey to better understand the concrete challenges of deep learning(DL) development and deployment. We found that there are indeed lots of issues causing the deployment quality issues of DL systems.(2)Based on this, we propose multiple differential testing criteria from the uncertainty perspective. We further proposed a differential testing framework named DiffChaser to systematically detect the buggy behavior of DL deployment. We performed systematic evaluations on diverse DL deployment scenarios and found our proposed methods are effective.(3)Furthermore, we conducted in-depth studies on the behavior analysis methods of DL from uncertainty and data distribution perspective with promising results. The results of this project set important foundations on quality assurance of DL deployment for further research and industry applications.

研究分野：ソフトウェア工学、機械学習工学

キーワード：差動テスト　信頼性と安全性　深層学習システム　深層学習システムの展開　テスト基準

１．研究開始当初の背景

近年、自動運転の交通事故、音声アシスタントによる誤った注文などのような事件が発生しているため、社会における実用的な用途に採用されるにつれて、DL システムの品質および信頼性保証を改善する余裕がある。また、DL をより広い範囲に展開 (Deployment)で きるために、DL システムは小型デバイス(例えば、モバイルデバイス、IoT エッジデバイ ス)への適用の要求が高まっている。このため、強力な GPU 搭載サーバー/クラウドから小 型デバイスへの DL システムの導入と適用の要請が急増している。ただし、サーバー/クラウ ドと小型デバイスの間の DL のフレームワーク、プラットフォーム、およびターゲットデバイスのハードウェアの多様性と差異により、DL システムを小型デバイスに効果的に展開することは非常に困難である。また、展開段階では、DL 変換において意図しない動作を伴うソフトウェアのバグを引き起こす問題がある。以上の問題に対して、申請者は DL システム展開階段において品質保証に関する予備的な実証調査を行い、その結論としては、既存の品質保証手法は DL システム開発段階のみに対応でき、DL システム展開段階 の品質保証が対応できない。このため、展開された DL システムが高いリスクに直面し、誤った決定をしてしまう可能性がある。従って、従来のソフトウェアシステムの展開と同様に、 信頼性及び安全性が重要である DL システムの展開プロセスにおいて、潜在的な不具合や望ましくない動作を検出するために系統的にテストすることが極めで重要である。しかし、実際に、実世界 DL システムの展開において、完全自動化による潜在的な不具合を系統的に究明することは非常に困難である。このため、本研究課題の核心をなす学術的「問い」は DL システム展開段階においてその品質はどう保証するかということである。

２．研究の目的

本研究では、以上の問題を解決するために、DL システム展開の品質保証のための自動差動テストフレームワーク (Differential Testing Framework) を構築することを目的としている。具体的には、() DL システム展開中の品質を評価するための差動テスト基準の設計、(2) DL システム展開中に導入されたソフトウェアの欠陥を検出するための自動的な差動テストフレームワークの開発、 (3) 実用的な小型モバイル機器や組込みシステムにおける提案手法の系統的評価となっている。本研究の進展により、DL システム展開中の品質保証環境が整い、より信頼性の高い DL システム展開プロセスを支援することが期待できる。

３．研究の方法

- **DL システム展開のための差動テスト基準の設計**

最初は DL システムの展開段階で使用できる品質保証測定のための差動テスト基準を提案する。特に、提案された基準は、展開された DL システムとその原始バージョンの動作の変化および相違について体系的かつ定量的な測定を提供することを目的としている。以下に示すように、3 つの観点から差動テスト基準を設計する: (1) 入力空間および特徴空間における決定境界の不確実性、(2) 統計的不確実性推定に基づくカバレッジ、(3) 情報理論 に基づく多様性測定のテス ト。この一連の基準は、バグ領域に関する定量的測定 することが可能にし、DL システム展開の品質保証に 関する理解とさらなる分析 のため基盤を構築する。

- **DL システム展開の自動差動テストフレームワークの開発**

次はDL展開中のバグ検出と品質評価のために、自動化された差動テストフレー ムワークを開発する。具体的には、まず、オリジナルのDNNorg、そ の小型デバイスに展開されたバージョンDNNdeploy、そしてユーザーがシードデータセットの 入力を指定したことを初めとしている。初期化のためにシードデータがテストプールに配置 された後、差分テストの反復が開始されます。 各反復において、テストはテストプールか ら無作為に選択され、メタモルフィック変換による新しい候補テスト生成のためのシードと して使用される。メタモルフィックミューテーター（Metamorphic Mutator）は、プール内の試験 データを無作為にサンプリングし、変換を実行して新しい候補試験データを生成する。関連 研究 DeepTestは、実世界のケースをシミュレートするためのイメージテストデータを生 成するためにメタモルフィック変換を実行するが、その設計はまだ初期段階であり、入力ド メインの外にある無効イメージの生成において高い誤検出がある。ここでは、(1) 物理的世界のシナリオをシミュレートしながら、低い偽陽性率で新しい有効テストデータを自動的に 生成する一連のより正確なメタモルフィックミューテーターを設計する。(2) 画像処理用の メタモルフィックミューテータの提案を含み、音声認識および自然言語処理用のメタモルフィックミューテータも設計し、全体として広範囲のDL応用を対応する。さらに、(3)ランタイム分析と組み合わせたKDE（Kernel Density Estimation-based）テストフィルタを使用し､生成された新しいテストをチェックし､その中の無効なテストを排除する。取得されたテストは目 標デバイスへの展開の前後に DL システムで分析される。それらの失敗したテストは、さら なる品質評価のために出力される。その後、差動カバレッジ分析が新しく生成された有効な データに対して実行され、それらの潜在的に有効なテストに優先順位を付けてテストプールに入れるように導く。テストの反復は、一定レベルの品質信頼度が得られるまで続けられる。

- **実用的な小型携帯機器と組込みシステムにおける提案手法の系統的評価**

このタスクでは、提案された技術を2つの実用的なアプリケーションに適用し、その有効性を評価する。DL システムをモバイルデバイスや組み込みシステムに移行するときの欠陥の検出方法の精度を評価し、精度、堅牢性、エネルギー効率などの観点からその品質を向上させる。これらの2つの小型デバイスはDLシステム展開の最も広く採用されているシナリオを表しているため、提案された技術は展開中に系統的にバグを検出し、包括的なレポートを提供する。それに基づいて、提案されDLシステム展開の自動差動テストフレームワークの有効性を評価することができる。

４．研究成果

本研究では、当初の提案内容は全て順調に完成された。(1)まず、深層学習(DL)の開発・展開における具体的な課題を理解するために、包括的な調査を行った。その結果、深層学習システムの展開品質には多くの課題があることが分かった[3][5][6] (TSE 20, ISSRE 19, ASE 19)。(2) (1)で得られた結果に基づいて、不確実性の観点から複数の差分テスト基準を提案した。さらに、DiffChaserと呼ばれる差分テストフレームワークを提案し、DLシステムの展開におけるバグを体系的に検出することに成功した。多様なDL展開シナリオに対して系統的な評価を行い、提案した手法が有効であることを示した[4](IJCAI 2019)。(3)さらに、不確実性やデータ分散の観点から、DLの挙動解析手法について詳細な検討を行い、有望な結果を獲得した[2][1] (ICSE 20, ASE 20)。本研究の成果は、今後の研究や産業界への応用に向けて、DL展開の品質保証に関する重要な基盤を構築することができた。

参考文献：

[1] David Berend, Xiaofei Xie, Lei Ma, Lingjun Zhou, Yang Liu, Chi Xu, Jianjun Zhao
Cats Are Not Fish: Deep Learning Testing Calls for Out-Of-Distribution Awareness
The 35th IEEE/ACM International Conference on Automated Software Engineering. (ASE 2020, CORE Rank A*)
[2] Xiyue Zhang, Xiaofei Xie, Lei Ma, Xiaoning Du, Qiang Hu, Yang Liu, Jianjun Zhao, Meng Sun
Towards Characterizing Adversarial Defects of Deep Learning Software from the Lens of Uncertainty
The 42nd International Conference on Software Engineering, 12 pages, 23-29 May 2020, Seoul, South Korea (ICSE'20, CORE Rank A*)
[3]Jie M. Zhang, Mark Harman, Lei Ma, Yang Liu
Machine Learning Testing: Survey, Landscapes and Horizons
IEEE Transactions on Software Engineering, 2020. (TSE, SCI-indexed, Impact Factor:4.778)
[4]Xiaofei Xie, Lei Ma, Haijun Wang, Yuekang Li, Yang Liu, Xiaohong Li
DiffChaser: Detecting Disagreements for Deep Neural Networks
The 29th International Joint Conference on Artificial Intelligence, Macau, China, August 2019 (IJCAI'19, CORE Rank A*)
[5]Qianyu Guo, Sen Chen, Xiaofei Xie, Lei Ma, Qiang Hu, Hongtao Liu, Yang Liu, Jianjun Zhao, Xiaohong Li
An Empirical Study towards Characterizing Deep Learning Development and Deployment across Different Frameworks and Platforms.
In Proc. 34th IEEE/ACM Conference on Automated Software Engineering (ASE 2019), San Diego, California, USA, November 11-15, 2019. (ASE'19, Core Rank A*)
[6]Tianyi Zhang, Cuiyun Gao, Lei Ma, Michael R. Lyu and Miryung Kim
An Empirical Study of Common Challenges in Developing Deep Learning Applications
The 30th International Symposium on Software Reliability Engineering (ISSRE'19, CORE Rank A)

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　0件）

１．著者名 Zhang Jie M.、Harman Mark、Ma Lei、Liu Yang	４．巻 1
２．論文標題 Machine Learning Testing: Survey, Landscapes and Horizons	５．発行年 2020年
３．雑誌名 IEEE Transactions on Software Engineering	６．最初と最後の頁 1～37
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TSE.2019.2962027	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計26件（うち招待講演　1件／うち国際学会　26件）

１．発表者名 Xiaofei Xie, Lei Ma, Haijun Wang, Yuekang Li, Yang Liu, Xiaohong Li
２．発表標題 DiffChaser: Detecting Disagreements for Deep Neural Networks
３．学会等名 The 29th International Joint Conference on Artificial Intelligence, Macau, China, August 2019 (IJCAI'19, CORE Rank A*)（国際学会）
４．発表年 2019年

１．発表者名 Zhenya Zhang, Deyun Lyu, Paolo Arcaini, Lei Ma, Ichiro Hasuo and Jianjun Zhao
２．発表標題 Effective Hybrid System Falsification Using Monte Carlo Tree Search Guided by QB-Robustness
３．学会等名 The 33rd International Conference on Computer-Aided Verification, 2021 (CAV 2021, CORE Rank A*)（国際学会）
４．発表年 2021年

１．発表者名 Maryam V. Pour, Li Zhuo, Lei Ma and Hadi Hemmati
２．発表標題 A Search-Based Testing Framework for Deep Neural Networks of Source Code Embedding
３．学会等名 IEEE International Conference on Software Testing, Verification and Validation (ICST 2021, CORE Rank A)（国際学会）
４．発表年 2021年

１．発表者名
Xiyue Zhang, Xiaoning Du, Xiaofei Xie, Lei Ma, Yang Liu, Meng Sun

２．発表標題
Decision-Guided Weighted Automata Extraction from Recurrent Neural Networks

３．学会等名
Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2021, CORE Rank A*)（国際学会）

４．発表年
2021年

１．発表者名
Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, Yang Liu, and Jianjun

２．発表標題
EfficientDeRain: Learning Pixel-wise Dilation Filtering for High-Efficiency Single-Image

３．学会等名
Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2021, CORE Rank A*)（国際学会）

４．発表年
2021年

１．発表者名
Qing Guo, Felix Juefei-Xu, Xiaofei Xie, Lei Ma, Jian Wang, Bing Yu, Wei Feng, Yang Liu

２．発表標題
Watch out! Motion is Blurring the Vision of Your Deep Neural Networks

３．学会等名
Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS 2020, CORE Rank A*)（国際学会）

４．発表年
2020年

１．発表者名
Zi Peng, Jinqiu Yang, Tse-Hsun Chen, Lei Ma

２．発表標題
A first look at the integration of machine learning models in complex autonomous driving systems: a case study on Apollo

３．学会等名
The 28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2020, CORE Rank A*)（国際学会）

４．発表年
2020年

１．発表者名
Xiaoning Du, Yi Li, Xiaofei Xie, Lei Ma, Yang Liu, Jianjun Zhao

２．発表標題
MARBLE: Model-Based Robustness Analysis of Stateful Deep Learning Systems

３．学会等名
The 35th IEEE/ACM International Conference on Automated Software Engineering. (ASE 2020, Rank A*) (国際学会)

４．発表年
2020年

１．発表者名
David Berend, Xiaofei Xie, Lei Ma, Lingjun Zhou, Yang Liu, Chi Xu, Jianjun Zhao

２．発表標題
Cats Are Not Fish: Deep Learning Testing Calls for Out-Of-Distribution Awareness

３．学会等名
The 35th IEEE/ACM International Conference on Automated Software Engineering. (ASE 2020, Rank A*) (国際学会)

４．発表年
2020年

１．発表者名
Xuhong Ren, Bing Yu, Hua Qi, Felix Juefei-Xu, Zhuo Li, Wanli Xue, Lei Ma and Jianjun Zhao

２．発表標題
Few-Shot Guided Mix for DNN Repairing

３．学会等名
The 36th IEEE International Conference on Software Maintenance and Evolution (ICSME, CORE Rank A) (国際学会)

４．発表年
2020年

１．発表者名
Run Wang,Felix Juefei-Xu, Yihao Huang, Qing Guo, Xiaofei Xie, Lei Ma, Yang Liu

２．発表標題
DeepSonar: Towards Effective and Robust Detection of AI-Synthesized Fake Voices

３．学会等名
Proceedings of the 28th ACM International Conference on Multimedia (ACM MM, CORE Rank A*) (国際学会)

４．発表年
2020年

１．発表者名
Run Wang,Felix Juefei-Xu, Yihao Huang, Qing Guo, Xiaofei Xie, Lei Ma, Yang Liu

２．発表標題
Amora: Black-box Adversarial Morphing Attack

３．学会等名
Proceedings of the 28th ACM International Conference on Multimedia (ACM MM, CORE Rank A*)（国際学会）

４．発表年
2020年

１．発表者名
Huangzhao Zhang, Zhuo Li, Ge Li, Lei Ma, Yang Liu, Zhi Jin

２．発表標題
Generating Adversarial Examples for Holding Robustness of Source Code Processing Models

３．学会等名
The 34th AAAI Conference on Artificial Intelligence, New York, USA, Feb 7-12, 2020. (AAAI'20, CORE Rank A*)（国際学会）

４．発表年
2020年

１．発表者名
Jianwen Sun, Tianwei Zhang, Xiaofei Xie, Lei Ma, Yan Zheng, Kangjie Chen, Yang Liu

２．発表標題
Stealthy and Efficient Adversarial Attacks against Deep Reinforcement Learning

３．学会等名
The 34th AAAI Conference on Artificial Intelligence, New York, USA, Feb 7-12, 2020. (AAAI'20, CORE Rank A*)（国際学会）

４．発表年
2020年

１．発表者名
Huangzhao Zhang, Zhuo Li, Ge Li, Lei Ma, Yang Liu, Zhi Jin

２．発表標題
Generating Adversarial Examples for Holding Robustness of Source Code Processing Models

３．学会等名
The 34th AAAI Conference on Artificial Intelligence, 8 pages, New York, USA, Feb 7-12, 2020. (AAAI ' 20, CORE Rank A*)（国際学会）

４．発表年
2020年

１．発表者名
Jianwen Sun, Tianwei Zhang, Xiaofei Xie, Lei Ma, Yan Zheng, Kangjie Chen, Yang Liu

２．発表標題
Stealthy and Efficient Adversarial Attacks against Deep Reinforcement Learning

３．学会等名
The 34th AAAI Conference on Artificial Intelligence, 9 pages, New York, USA, Feb 7-12, 2020. (AAAI ' 20, CORE Rank A*) (国際学会)

４．発表年
2020年

１．発表者名
Xiyue Zhang, Xiaofei Xie, Lei Ma, Xiaoning Du, Qiang Hu, Yang Liu, Jianjun Zhao, Meng Sun

２．発表標題
Towards Characterizing Adversarial Defects of Deep Learning Software from the Lens of Uncertainty

３．学会等名
The 42nd International Conference on Software Engineering, 12 pages, 23-29 May 2020, Seoul, South Korea (ICSE ' 20, CORE Rank A*) (国際学会)

４．発表年
2020年

１．発表者名
Wuji: Automatic Online Combat Game Testing Using Evolutionary Deep Reinforcement Learning

２．発表標題
Yan Zheng, Xiaofei Xie, Ting Su, Lei Ma, Jianye Hao, Zhaopeng Meng, Yang Liu, Ruimin Shen, Yinfeng Chen, Changjie Fan

３．学会等名
The 34rd IEEE/ACM International Conference on Automated Software Engineering, pp.772-784, San Diego, California, USA, November 11-15, 2019 (ASE ' 19, Rank A*) (国際学会)

４．発表年
2019年

１．発表者名
Qianyu Guo, Sen Chen, Xiaofei Xie, Lei Ma, Qiang Hu, Hongtao Liu, Yang Liu, Jianjun Zhao, Xiaohong Li

２．発表標題
An Empirical Study towards Characterizing Deep Learning Development and Deployment across Different Frameworks and Platforms

３．学会等名
In Proc. 34th IEEE/ACM Conference on Automated Software Engineering, pp.810-822, San Diego, California, USA, November 11-15, 2019. (ASE ' 19, Rank A*) (国際学会)

４．発表年
2019年

１．発表者名
Xiaoning Du, Xiaofei Xie, Yi Li, Lei Ma, Yang Liu, and Jianjun Zhao

２．発表標題
A Quantitative Analysis Framework for Recurrent Neural Network.

３．学会等名
In Proc. 34th IEEE/ACM Conference on Automated Software Engineering, pp.1062-1065, San Diego, California, USA, November 11-15, 2019. (ASE ' 19, Rank A*) (国際学会)

４．発表年
2019年

１．発表者名
Xiaofei Xie, Hongxu Chen, Yi Li, Lei Ma, Yang Liu, and Jianjun Zhao

２．発表標題
Coverage-guided Fuzzing for Feedforward Neural Networks

３．学会等名
In Proc. 34th IEEE/ACM Conference on Automated Software Engineering, pp.1162-1165, San Diego, California, USA, November 11-15, 2019. (ASE ' 19, Rank A*) (国際学会)

４．発表年
2019年

１．発表者名
Qiang Hu, Lei Ma, Xiaofei Xie, Bing Yu, Yang Liu, and Jianjun Zhao

２．発表標題
DeepMutation++: a Mutation Testing Framework for Deep Learning Systems

３．学会等名
In Proc. 34th IEEE/ACM Conference on Automated Software Engineering, pp.1158-1161, San Diego, California, USA, November 11-15, 2019. (ASE ' 19, Rank A*) (国際学会)

４．発表年
2019年

１．発表者名
Xiaoning Du, Xiaofei Xie, Yi Li, Lei Ma, Yang Liu and Jianjun Zhao

２．発表標題
DeepStellar: Model-Based Quantitative Analysis of Stateful Deep Learning Systems

３．学会等名
The 27th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 477-487, Tallinn, Estonia, August 2019 (FSE ' 19, CORE A*) (国際学会)

４．発表年
2019年

１．発表者名
Yang Liu, Lei Ma, Jianjun Zhao

２．発表標題
Secure Deep Learning Engineering: a Road towards Quality Assurance of Intelligent Systems

３．学会等名
The 21st International Conference on Formal Engineering Methods, pp.3-15, ShenZhen, China, Nov. 2019 (ICFEM 2019, CORE Rank B)（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Tianyi Zhang, Cuiyun Gao, Lei Ma, Michael R. Lyu and Miryung Kim

２．発表標題
An Empirical Study of Common Challenges in Developing Deep Learning Applications

３．学会等名
The 30th International Symposium on Software Reliability Engineering, 12 pages, Oct. 2019, Berlin, Germany (ISSRE'19, CORE Rank A)（国際学会）

４．発表年
2019年

１．発表者名
Xiaofei Xie, Lei Ma, Felix Juefei-Xu, Minhui Xue, Hongxu Chen, Yang Liu, Jianjun Zhao, Bo Li, Jianxiong Yin, Simon See

２．発表標題
DeepHunter: A Coverage-Guided Fuzz Testing Framework for Deep Neural Networks

３．学会等名
The 28th International Symposium on Software Testing and Analysis, pp.146-157, Beijing, China, July 2019 (ISSTA'19, CORE Rank A)（国際学会）

４．発表年
2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
カナダ	University of Calgary			
カナダ	Concordia University			
シンガポール	Nanyang Technological University			
中国	Peking University			
米国	UIUC			
米国	UCLA			
英国	University College London			