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　４ 年　 ６ 月 ２３ 日現在

機関番号：６２６１５
研究種目：基盤研究(B)（一般）
研究期間：2018～2020
課題番号：１８Ｈ０３２２４

研究課題名（和文）機械学習ソフトウェアの高信頼化に関わるデータセット多様性の研究

研究課題名（英文）On Dataset Diversity for Achiving High Reliability of Machine Learning Software

研究代表者
中島　震（Nakajima, Shin）

国立情報学研究所・大学共同利用機関等の部局等・名誉教授

研究者番号：６０３５０２１１

交付決定額（研究期間全体）：（直接経費）　　9,400,000円

研究成果の概要（和文）：深層ニューラルネットワーク訓練学習基盤を検査するソフトウェア・テスティングの新しい技術としてメタモルフィック・テスティングを用いる方法の研究を行った。その成果として、セマンティックノイズ合成による入力テストデータ自動生成方法、ニューロン内部活性状態をもとにしたメタモルフィック関係定義法、統計的仮説検定を応用いしたテスティングフレームワークを考案した。手書き数字画像分類の問題に適用し、検査法の有効性を確認した。

研究成果の学術的意義や社会的意義
深層ニューラルネットワークの技術は高度な信頼性を求められるシステムに応用され、不具合が生じると社会的な影響が大きいことから、品質評価方法の確立が求められている。学術的には、セマンティックノイズによるデータセット多様性というアイデアから、メタモルフィック・テスティングを深層ニューラルネットワーク訓練学習基盤の検査に応用する方法を示したことである。

研究成果の概要（英文）：We proposed a new metamorphic testing method applicable to checking the correctness of machine learning programs constituting the core mechanism of machine learning frameworks. The contributions include a test generation method employing the notion of semantic noises, metamorphic relations referring to active neuron states, and a testing framework to combine the statistical hypothesis testing method and the metamorphic testing.

研究分野：ソフトウェア

キーワード：ソフトウェア工学　ソフトウェア・テスティング　ニューラル・ネットワーク　ディペンダビリティ

１．研究開始当初の背景

深層ニューラルネットワーク（Deep Neural Networks, DNN）に代表される機械学習技術が発展し、自動運転・医療診断などの社会基盤システムで活用される段階に達した。このような高度な信頼性を求められるシステムに不具合が生じると社会的な影響が大きい。DNN ソフトウェアの品質を評価する方法の確立が期待されている。

DNN ソフトウェア構築の標準的な方法では、訓練学習基盤・学習モデル・訓練データセットの３つの技術要素が関係する。品質評価では、不具合の検知と欠陥の有無を調べる作業を実施する。通常の DNN ソフトウェア開発では、訓練学習基盤として、ベンダー提供のブラックボックス化された機械学習フレームワーク[1]を利用することが多い。DNN ソフトウェア構築者は、機械学習フレームワークの品質を確認することが難しい。自身が作成した DNN ソフトウェアの機能・振舞いを検査することで、間接的に調べざるを得ない。そこで、実績のある、つまり品質が確認されている学習モデル・訓練データセットを用いて構築した DNN ソフトウェアを用いて、機械学習フレームワークの欠陥の有無を調べる。

訓練学習基盤の標準的な方法は、数値最適化の問題を勾配法によって解くプログラム[文献 1]である。基本的には、収束解が得られるまで状態空間を探索する繰り返し処理からなる。ところが、収束解は未知であり、予め正解がわかっていることはない。仮に正解となる収束解が既知であれば、この数値最適化問題を解く必要がない。このようなプログラムの検査では、正解か否かを確認するテストオラクルが存在しない数値計算プログラムのテスト法を確立が課題となる。メタモルフィック・テスティング（Metamorphic Testing, MT）の方法[文献 2]が有力なアプローチとする議論と考えられていた。実際、分類学習問題の SVM などに MT の方法を適用した成功事例[文献 3][文献 4]が知られていた。しかし、DNN 訓練学習基盤を実現する最適化問題は SVM を解く問題よりも複雑であり、新たな技術確立が必要とされていた。

２．研究の目的

DNN ソフトウェアの訓練学習基盤（機械学習フレームワーク）の欠陥の有無を調べるソフトウェア・テスティングの新しい技術として MT を用いる方法を確立する。

３．研究の方法

MT の方法では、入力テストデータ生成方法と、プログラム実行結果を比較するメタモルフィック関係の定義法を確立する必要がある。第１項で述べたように、実績のある学習モデル・訓練データセットを用いて、方式検討・実験による確認を行う。具体的には、手書き数字画像（MNIST データセット）の分類問題を対象とし、研究を進める。

４．研究成果

入力テストデータ生成方法・メタモルフィック関係定義法・テスティングフレームワーク・関連する知見の順に報告する。

（１）入力テストデータ生成方法

学習モデルを決めた時、訓練学習基盤の働きは、訓練データセットを入力し、訓練済み学習モデルを出力することなので、MT の方法で用いる入力テストデータは訓練データセットである。そこで、訓練学習基盤のプログラムを効率良く検査するには、どのような訓練データセットを入力すれば良いかという問題である。つまり，検査に有用な多様なデータセットを自動生成する方法の確立が課題である。

この検査対象プログラムは状態空間の数値探索処理を行うことから、特に、繰り返しの収束判定処理（計算誤差）に影響を与える可能性が高い入力を準備すれば良い。一方、今、対象としている学習問題の訓練データセットの特徴は保存する必要がある。そこで、セマンティックノイズ付加によるデータセット多様性の方法を考案した[文献 5]。

対象問題（実験では MNIST）の訓練データセットから選んだデータ x（クリーンデータ）に、ノイズ δx を付加し、x'=x+ δx を得る。選んだデータ x の正解タグを t とする時、x'の正解タグ同じ t となるようにノイズを付加する。今回の実験では、条件付き最適化問題によって、x'を求める方法を考案した[文献 6]。与える条件によって、付加するノイズの大きさ・形状を制御可能であり、元の学習問題の特徴が保存されていることを確認できるという長所がある。一方、ひとつひとつについて条件付き最適化問題を解く必要があることから、規模の大きい訓練データセットを対象とする時、セマンティックノイズを付したデータセットの生成に時間がかかるという短所がある。実験に用いた MNIST では、画像が小さいことから、通常の PC 上でも問題なく作動することを確認した。

---

[1] Google 社の Tensorflow や Facebook 社の PyTorch など。

（2）メタモルフィック関係定義法

MT の方法を DNN 訓練学習基盤のテストに応用する方法では、クリーンデータを入力する場合とセマンティックノイズを付加した入力の場合の実行結果を比較する。ところが、訓練学習基盤の出力は訓練済み学習モデルであり、2 つの訓練済み学習モデルを比較し、その関係を調べることが困難である。そこで、別途、準備した評価用データを両者に入力して予測分類を実行させた時の振舞いを調べるという間接的な検査方法を用いる。実験では、評価用データとして、MNIST で提供される試験データセットを用いた。

基本となる訓練学習基盤と意図的に欠陥を混入させた訓練学習基盤から得られた各々の訓練済み学習モデルに対して実験を行った。ところが、評価用データ入力に対する正解率に大きな差がなく、正解率という指標から、欠陥の有無を判断することができない。そこで、訓練済み学習モデルの内部状態を調べる方法を考案した。評価用データを入力した時に内部で活性化されるニューロンの状態に着目すると、欠陥の有無と高い相関がある[文献 7]。

この内部活性状態の方法は、用いた学習モデルの構造に大きく影響される。MNIST 問題は取扱いが容易なことから古典的な全結合ネットワークを用いて検査できる。より複雑な画像の場合、CNN を用いることが多い。そこで、CNN に対する内部活性状態の定義法を検討した。一般に、全結合では、ニューロンの入換えに対して出力が変化しないという性質（スワップ不変性）が成り立つ。一方、CNN は層とニューロンに機能的に異なる役割が与えられていることから、スワップ不変性が成り立たず、内部活性状態を一意に定義できない。本研究が解決しようとする課題は、訓練学習基盤の欠陥を調べることから、全結合ネットワークを用いた検査を行えば良い。

（3）テスティングフレームワーク

評価用データは単独ではなくデータの集まり（実際は試験データセット）で、得られる結果も指標値（内部活性状態から得られる指標値）の集まりで、統計的な取扱いを要する。そこで、MT の方法と統計的仮説検定を組み合わせたテスティングフレームワークを考案し、実験によって、欠陥を意図的に混入するかしないかで、統計的な有意水準下で、異なる結果が生じることを確認した[文献 8]。

（4）関連する知見

DNN ソフトウェアの品質評価は、実用的な重要性が高い一方で、従来のソフトウェアに対する品質評価とは異なる面が多い。本研究は、多数ある課題のうち、訓練学習基盤の欠陥検査を行うソフトウェア・テスティングの方法を取り扱った本研究課題の実験では、品質が確認されているデータセットを用いた。学習データの品質を確認する方法が別途必要になる。そこで、従来ソフトウェアのデータ品質に関する国際標準 SQuaRE を学習データに適用する方法を考察した[文献 9]。

＜参考文献＞

[文献 1] S. Haykin, Neural Networks and Learning Machines (3rd), Prentice Hall 2008.
[文献 2] T.Y. Chen, S.C. Chung, and S.M. Yiu, Metamorphic Testing - A New Approach for Generating Next Test Cases, HKUST-CS98-01, The Hong Kong University of Science and Technology, 1998.
[文献 3] X. Xie, J.W.K. Ho, C. Murphy, G. Kaiser, B. Xu, and T.Y. Chen, Testing and Validating Machine Learning Classifiers by Metamorphic Testing, J. Syst. Softw., 84(4), pp.544-558, 2011.
[文献 4] S. Nakajima and H.N. Bui, Dataset Coverage for Testing Machine Learning Computer Programs, In Proc. 23rd APSEC, pp.297-304, 2016.
[文献 5] S. Nakajima, Dataset Diversity for Metamorphic Testing of Machine Learning Software, Proc. 8th SOFT+MSVL, pp.21-38, 2018.
[文献 6]S. Nakajima and T.Y. Chen, Generating Biased Dataset for Metamorphic Testing of Machine Learning Programs, Proc. IFIP-ICTSS 2019, pp.56-64, 2019.
[文献 7] S. Nakajima, Distortion and Faults in Machine Learning Software, In Proc. 9th SOFL+MSVL, pp.29-41, 2019.
[文献 8] S. Nakajima, Software Testing with Statistical Partial Oracles - Application to Neural Networks Software -, In Proc. 10th SOFL+MSVL, pp.175-192, 2020.
[文献 9] S. Nakajima and T. Nakatani, AI Extension of SQuaRE Data Quality Model, In Proc. 21st QRS-C, pp.306-313, 2020.

５．主な発表論文等

〔雑誌論文〕　計6件（うち査読付論文　6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<br>Shin Nakajima | ４．巻<br>- |
| ２．論文標題<br>Distortion and Faults in Machine Learning Software | ５．発行年<br>2020年 |
| ３．雑誌名<br>Proc. The 9th International Workshop on SOFL + MSVL for Reliability and Security (SOFL+MSVL 2019) | ６．最初と最後の頁<br>29-41 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1007/978-3-030-41418-4_3 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| １．著者名<br>Shin Nakajima, T.Y. Chen | ４．巻<br>- |
| ２．論文標題<br>Generating Biased Dataset for Metamorphic Testing of Machine Learning Programs | ５．発行年<br>2019年 |
| ３．雑誌名<br>Proc. The 31st IFIP International Conference on Testing Software and Systems (IFIP-ICTSS 2019) | ６．最初と最後の頁<br>56-64 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1007/978-3-030-31280-0_4 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>該当する |

| | |
|---|---|
| １．著者名<br>Shin Nakajima | ４．巻<br>- |
| ２．論文標題<br>Quality Evaluation Assurance Levels for Deep Neural Networks Software | ５．発行年<br>2019年 |
| ３．雑誌名<br>Proc. The 24th International Conference on Technologies and Applications of Artificial Intelligence (TAAI 2019) | ６．最初と最後の頁<br>1-6 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1109/TAAI48200.2019.8959916 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| １．著者名<br>Shin Nakajima | ４．巻<br>- |
| ２．論文標題<br>Dataset Diversity for Metamorphic Testing of Machine Learning Software | ５．発行年<br>2019年 |
| ３．雑誌名<br>Post-Proc. 8th SOFL+MSVL (LNCS) | ６．最初と最後の頁<br>21-38 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1007/978-3-030-13651-2_2 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| 1．著者名<br>Shin Nakajima | 4．巻<br>- |
| 2．論文標題<br>Quality Assurance of Machine Learning Software | 5．発行年<br>2018年 |
| 3．雑誌名<br>Proc. IEEE 7th Global Conference on Consumer Electronics | 6．最初と最後の頁<br>601-604 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.1109/GCCE.2018.8574766 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

| | |
|---|---|
| 1．著者名<br>中島震 | 4．巻<br>35 |
| 2．論文標題<br>データセット多様性のソフトウェア・テスティング | 5．発行年<br>2018年 |
| 3．雑誌名<br>コンピュータ・ソフトウェア | 6．最初と最後の頁<br>26-32 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>10.11309/jssst.35.2_26 | 査読の有無<br>有 |
| オープンアクセス<br>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<br>－ |

〔学会発表〕　計10件（うち招待講演　2件／うち国際学会　2件）

| |
|---|
| 1．発表者名<br>中島震 |
| 2．発表標題<br>訓練済み機械学習モデル歪みの定量指標 |
| 3．学会等名<br>電子情報通信学会 ソフトウェア・サイエンス研究会 |
| 4．発表年<br>2020年 |

| |
|---|
| 1．発表者名<br>中島震 |
| 2．発表標題<br>機械学習ソフトウェア・テスティングの技術動向 |
| 3．学会等名<br>電子情報通信学会 システム数理と応用研究会 |
| 4．発表年<br>2020年 |

１．発表者名
中島震

２．発表標題
AIビジネスリスク軽減への価値共創アプローチ

３．学会等名
日本ソフトウェア科学会ソフトウェア工学の基礎ワークショップ

４．発表年
2020年

１．発表者名
中島震

２．発表標題
ファズ・データセットを用いたメタモルフィック・テスティング ~ 機械学習ソフトウェアの検査 ~

３．学会等名
日本ソフトウェア科学会第36回大会

４．発表年
2019年

１．発表者名
中島震

２．発表標題
モデルの歪みと機械学習プログラムの欠陥

３．学会等名
情報処理学会第202回ソフトウェア工学研究発表会

４．発表年
2019年

１．発表者名
中島震

２．発表標題
機械学習ソフトウェアの品質評価保証レベル

３．学会等名
電子情報通信学会ソフトウェア・サイエンス研究会

４．発表年
2019年

１．発表者名
中島震

２．発表標題
機械学習ソフトウェアの品質：製品，サービス，プラットフォーム

３．学会等名
電子情報通信学会知能ソフトウェア工学研究会

４．発表年
2018年

１．発表者名
Shin Nakajima

２．発表標題
Dataset Diversity for Metamorphic Testing of Machine Learning Software

３．学会等名
8th International Workshop SOFL+MSVL（国際学会）

４．発表年
2018年

１．発表者名
Shin Nakajima

２．発表標題
Quality Assurance of Machine Learning Software

３．学会等名
IEEE 7th Global Conference on Consumer Electronics（招待講演）（国際学会）

４．発表年
2018年

１．発表者名
中島震

２．発表標題
機械学習ソフトウェアの品質保証とは

３．学会等名
機械学習工学研究会キックオフシンポジウム（招待講演）

４．発表年
2018年

〔図書〕　計1件

| | |
|---|---|
| １．著者名<br>中島　震 | ４．発行年<br>2020年 |
| ２．出版社<br>丸善出版 | ５．総ページ数<br>192 |
| ３．書名<br>ソフトウェア工学から学ぶ　機械学習の品質問題 | |

〔産業財産権〕

〔その他〕

Researchmap
https://researchmap.jp/nkjm/

６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（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 |
|---|---|