

機械学習ソフトウェアの品質評価

中島 震 (国立情報学研究所)

どんな研究？

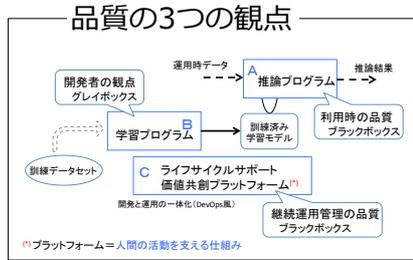
機械学習の技術を、極めて高度な信頼性が必要とされるアプリケーションに適用することが期待されています。不具合が社会問題になることから、品質評価の技術確立が必須です。

何がわかる？

一般に、品質評価の実用的な方法は、ソフトウェア・テストです。機械学習プログラムの検査では、データセット多様性とメタモルフィック・テストを組み合わせてみます。

状況設定

機械学習ソフトウェアの「気が利く」という特徴は、多少のいい加減さ・理由がわからないブラックボックスさと関連します。厳密さ・確実さという、高い品質達成の基本的な前提と正反対です。



機械学習ソフトウェアの品質には3つの観点があります。本年は、開発者の視点からみた「品質」として、高い信頼性を達成する技術を中心に考察します。

研究内容

ソフトウェア・テスト

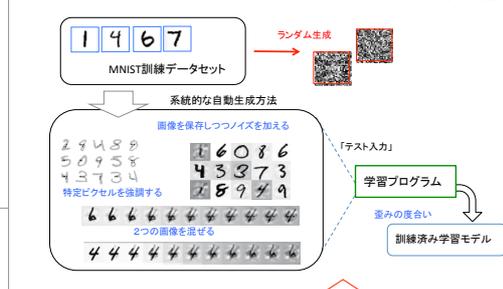
正常系テスト
→ 想定内の入力データ
→ 期待する機能をはたす

例外系テスト
→ 想定外の入力データ
→ 不具合を生じない

→ さまざまな入力の生成方法！

多様性 ⇔ 異なる経験分布
(「意図的な」標本選択バイアス)

データセット多様性



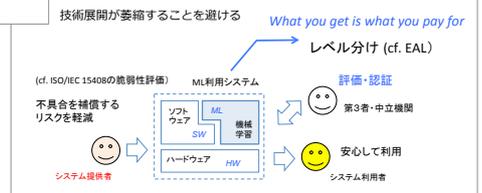
ファズ・データセット

メタモルフィック・テストの方法で歪みの度合いの「現れ」を検査

MSEA ℓ : 品質評価保証レベル

| 品質保証 | 評価項目 | MSEA-1 | MSEA-2 | MSEA-3 | MSEA-4 | MSEA-5 |
|------------|------------------|--------|--------|--------|--------|--------|
| サービス品質 | S-BEL 信頼性保証 | 1 | 1 | 1 | 2 | 3 |
| | S-DM 脆弱性対策 | 1 | 1 | 2 | 2 | 2 |
| | S-DOC 脆弱性対策 | 1 | 2 | 2 | 2 | 4 |
| | S-ADV 脆弱性対策(脆弱性) | 1 | 1 | 1 | 2 | 2 |
| | S-DPT 脆弱性 | 1 | 1 | 1 | 2 | 2 |
| | S-DM 脆弱性 | 1 | 1 | 1 | 2 | 2 |
| | S-DM 脆弱性 | 1 | 1 | 1 | 2 | 2 |
| | S-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | S-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | S-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| 学習ソフトウェア品質 | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | T-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| プラットフォーム品質 | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |
| | P-DM 脆弱性 | 1 | 2 | 2 | 2 | 2 |

第三者評価・認証の枠組み



一部、NEDOの委託業務 (NIIへは再委託) によって実施

最近の関連発表

- [1] Quality Assurance of Machine Learning Software, Invited talk, Proc. 7th GCCE 2018, pp.601-604, October 2018.
- [2] Dataset Diversity for Metamorphic Testing of Machine Learning Software, Proc. 8th SOFL+MSVL, pp.21-38, February 2019.
- [3] 機械学習ソフトウェアの品質評価保証レベル, SIGSS, March 2019.
w/ 妹尾義樹・大岩寛・磯部祥尚 (産業技術総合研究所)
- [4] モデルの歪みと機械学習プログラムの欠陥, SIGSE, July 2019.

