

ニューロンカバレッジ技法を用いた AI モデル特性分析によるテスト十分性向上施策
Measures to improve test sufficiency by analyzing AI model characteristics using neuron coverage method
中川 純貴 junki.nakagawa.gs@hitachi.com 株式会社日立製作所 システム&サービスビジネス統括本部 品質保証統括本部
発表要旨： AI モデルの品質要件の一つに、さまざまな外乱要素に耐え得る「頑健性」がある。弊社では手書き文字を含む帳票から必要な情報を読み取りデータ化する帳票認識ソフトウェアを開発しており、本ソフトウェアが搭載する AI モデルにも高い頑健性が要求されている。 頑健性の評価手法として、効果的なメタモルフィック関係を利用してテストケースを作成するメタモルフィックテストが有効であることが報告されている。しかし、本テストにおけるテスト十分性向上およびテスト十分性評価が大きな課題となっている。 そこで本報告では、メタモルフィックテストのテスト十分性向上施策として、ニューロンカバレッジ技法を用いた AI モデルの特性分析を提案する。ニューロンカバレッジ技法とは、ニューラルネットワーク内のニューロン活性化を基準としたテストケース評価技術およびカバレッジ率を上昇させるテストケース生成技術である。弊社で開発したニューロンカバレッジツールを使用し、帳票認識ソフトウェアが搭載する AI モデルのカバレッジ率が外乱によるデータ変化に対してどのような遷移を示すか分析した(AI モデルの特性分析)。そして分析結果をもとに、メタモルフィックテストのテスト十分性向上に寄与する効果的なテストケース範囲を明確にした。 また今後検討を重ねる必要があるが、カバレッジ率を品質指標とすることでテストケースの量的/質的基準を策定でき、テスト十分性評価に繋がると考えている。
キーワード： 帳票認識ソフトウェア, メタモルフィックテスト, ニューロンカバレッジ技法
想定している聴衆 AI 搭載ソフトウェアの開発者および品質保証担当者
発表者の紹介 (全角 100 文字)： 2016 年～ (株) 日立製作所にて、日立ソフトウェア・ミドルウェアの品質保証に従事。 2018 年～ 同社内にて AI 搭載ソフトウェア開発における品質確保プロセスおよびテスト技術の整備に従事。

ニューロンカバレッジ技法を用いたAIモデル特性分析による テスト十分性向上施策

2021/9/9

株式会社 日立製作所
システム&サービスビジネス統括本部
品質保証統括本部

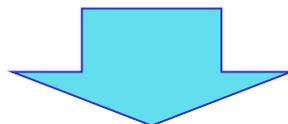
中川 純貴

AI搭載ソフトウェアの品質保証技術の進歩が遅れている

さまざまな悩みが存在

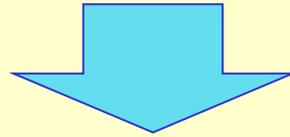


従来のソフトウェアテストや品質保証の考え方が通用しない



信頼されるAI搭載ソフトウェアの実現に向けた品質保証技術の確立は重要テーマ

さまざまなAIテスト技法が提案されているが、実案件への適用例は少ない

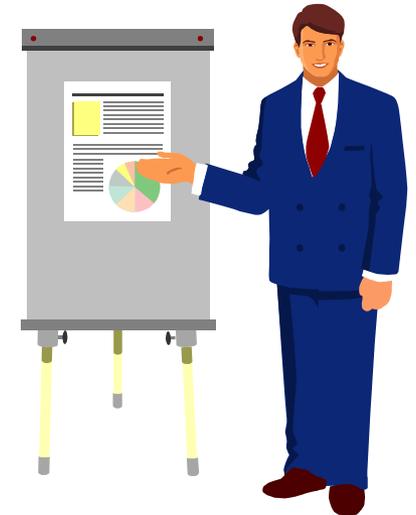


実案件へのAIテスト技法適用を推進したい

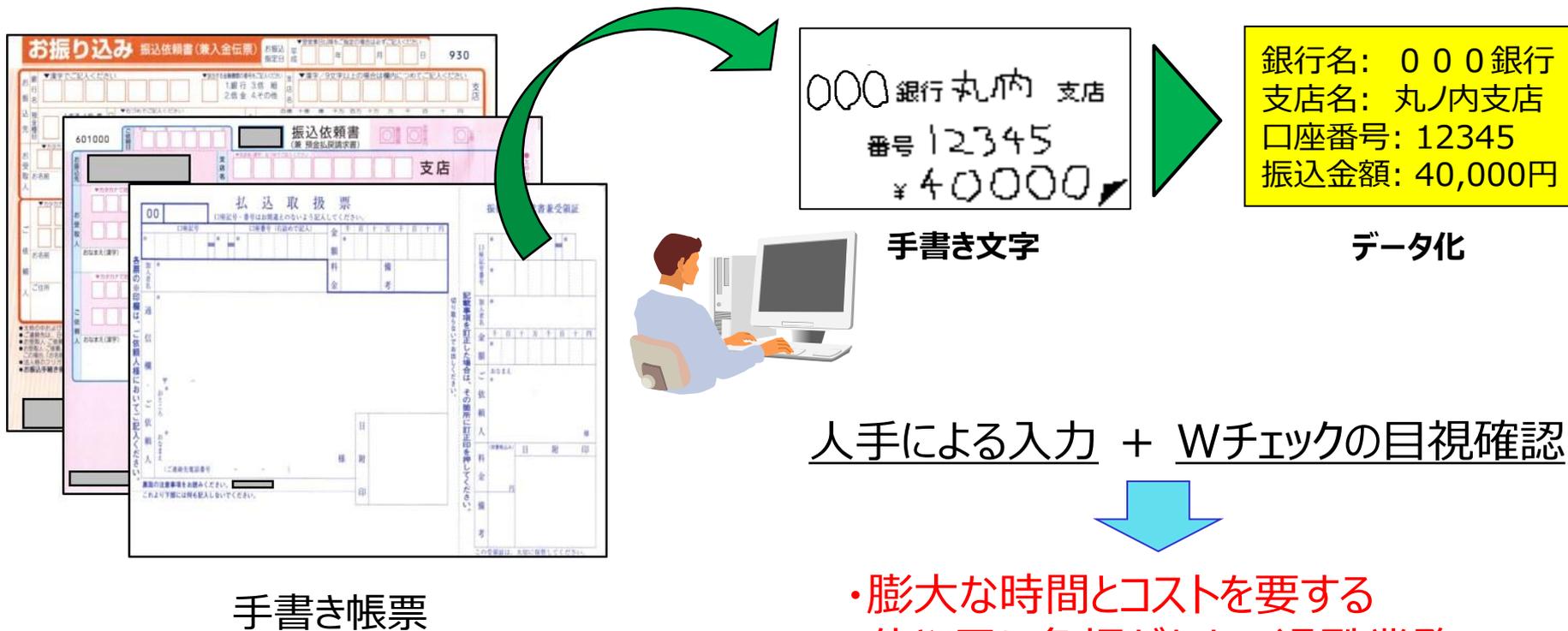
品質確保のために
テストは必要

本日は

AIテスト技法のひとつであるメタモルフィックテストの
課題に対する施策内容を報告



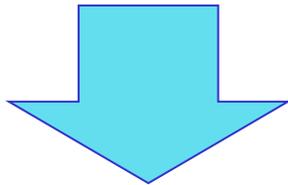
ビジネスの効率化/コスト削減には、手書き文字のデータ化が必須



日立はOCR(Optical Character Recognition)技術の研究開発を重ね、手書き帳票のデータ化業務の効率化に貢献

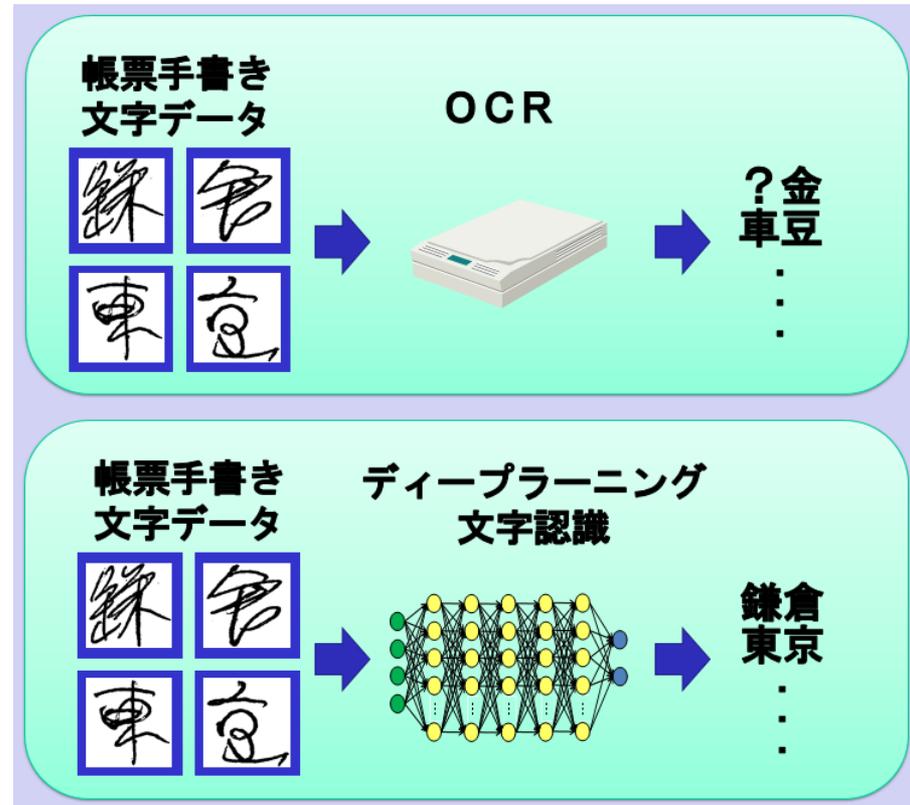
【従来のOCRの課題】

パターンマッチングを用いており、
認識できない手書き文字が多い
(くせ字, 崩し字など)



ディープラーニングとの組み合わせ

⇒読み取りにくい手書き文字等の
認識精度向上

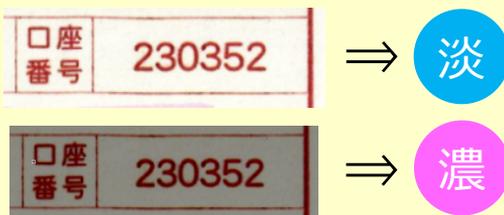


ディープラーニングとは

⇒生物の脳にある神経細胞(ニューロン)の繋がりを模倣したニューラルネットワークを
ベースとした機械学習手法。コンピューターが大量のデータから自動的に特徴を抽出する。

業務で発生する帳票はさまざまな外乱要素を含む

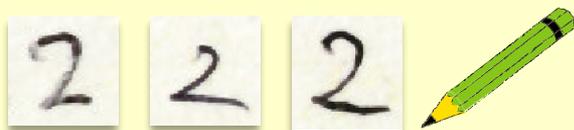
背景色の变化



罫線と文字の接触



手書きによる文字の乱れ



外乱に耐え、正確に認識できる
AIモデルが必要だな



帳票認識ソフトウェアのAIモデルには高い「**ロバストネス(頑健性)**」が求められる



ロバストネス評価のテスト技法として、**メタモルフィックテスト**が有効

※帳票認識ソフトウェアでのロバストネス要件例：入力データに対して画像の回転処理、ぼかし処理、ノイズ等の加工をしても正しく認識できること

メタモルフィックテスト

メタモルフィック関係を利用して既存のテストケースから新たなテストケースを作成し、プログラムの妥当性を確認するテスト技法



メタモルフィック関係とは

⇒「入力に対して一定の変化を与えると、出力の変化を理論上予測できる」という関係

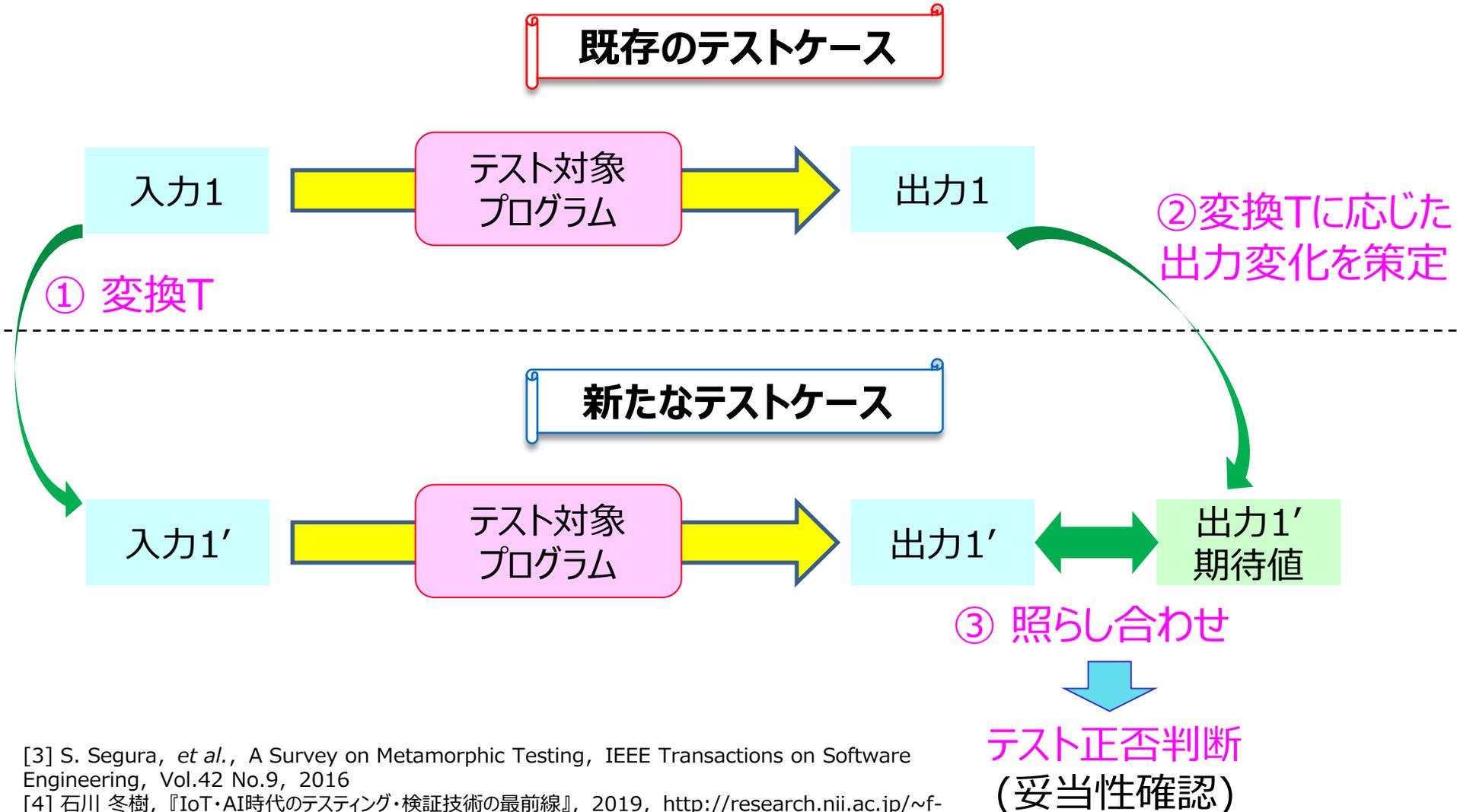
メタモルフィック関係を利用して
多数のテストケースを作成し、テスト

テストケース間の出力結果の比較を繰り返し、
プログラムの妥当性を確認



3-2. メタモルフィックテストの概要②

◆メタモルフィックテストの概念図

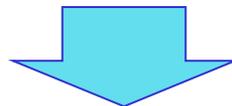


[3] S. Segura, et al., A Survey on Metamorphic Testing, IEEE Transactions on Software Engineering, Vol.42 No.9, 2016

[4] 石川 冬樹, 『IoT・AI時代のテストング・検証技術の最前線』, 2019, <http://research.nii.ac.jp/~f-ishikawa/work/1901-SQiP-AIquality.pdf>

3-3. メタモルフィックテストにおける代表的な変換

Additive	データ点の属性に加算(減算)
Multiplicative	データ点の属性に乗算
Permutative	データ点の入れ替え
Inversive	正解ラベルの反転
Inclusive	新しいデータ点の追加
Exclusive	既存データ点の削除
Compositional	データの合成
Noise-based	出力に影響を及ぼさない入力に変化
Semantically Equivalent	元データと同様の意味をもつ入力に変化
Heuristic	元データに近い入力に変化
Statistical	元データと統計的に同じ性質を示す入力に変化



入力の変換に伴う出力の期待値を予測する

[5] C. Murphy, *et al.*, Properties of Machine Learning Applications for Use in Metamorphic Testing, SEKE2008, 2008

[6] C. Murphy, Applications of Metamorphic Testing, 2011, <http://www.cis.upenn.edu/~cdmurphy/pubs/MetamorphicTesting-Columbia-17Nov2011.ppt>

3-4. メタモルフィック関係を利用したデータの作成例

ロバストネス要件をもとに意図的に変換を施し、外乱を含んだテストケースを派生

①背景色が濃いケース



背景色を濃くする

⇒ Noise-based

②項目値が複数行のケース



複数行で記入する

⇒ Noise-based

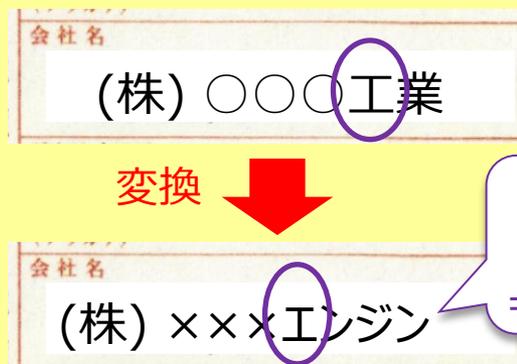
③文字と罫線が接触するケース



罫線と接触させる

⇒ Noise-based

④文字の形が似ているケース



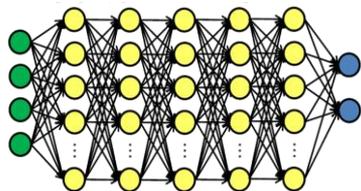
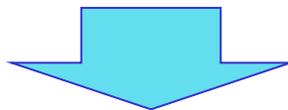
片仮名の“工”に変える

⇒ Heuristic



メタモルフィックテストを用いた検査により、開発部署では気づかなかったAIモデルの弱点が明確になった

認識に失敗したテストケースの原因調査



- ・内部ロジックの不良
- ・学習データの不足

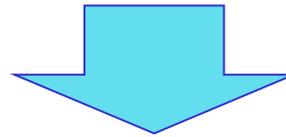


メタモルフィックテストが帳票認識ソフトウェアにおけるAIモデルの頑健性評価に有効

問題点

テスト十分性を評価できていない

➡ テスト十分性とは、ロバストネス要件を評価するメタモルフィックテストのテストケースの量的/質的十分性を指す

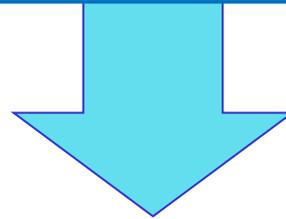


課題

- ① 品質を担保するテストケースの量的/質的基準
- ② 基準を満たしているか否かの評価方法
- ③ テスト十分性を向上させるための技法

メタモルフィックテストのテスト十分性評価の実現に向けて必要な検討課題

- ① 品質を担保するテストケースの量的/質的基準
- ② 基準を満たしているか否かの評価方法
- ③ テスト十分性を向上させるための技法

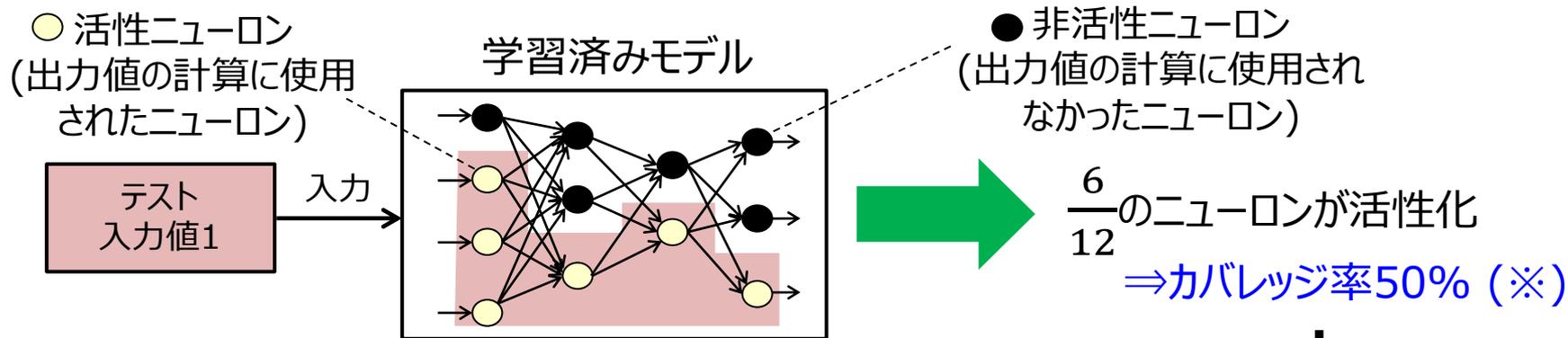


提案施策

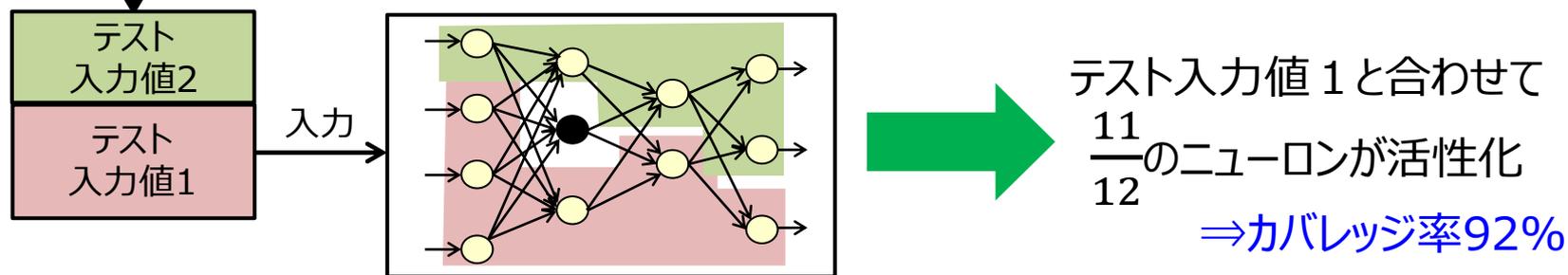
ニューロンカバレッジ技法を用いたAIモデルの特性分析



ニューラルネットワーク内のニューロン活性化を基準としたテストケース評価技術 およびカバレッジ率を上昇させるテストケース生成技術



カバレッジ率が増加するようなテストケースを生成



※カバレッジ率：全体のニューロンに対する活性ニューロンの存在割合

カバレッジ率の上昇がテストケースの充足に繋がると考えられ、
ニューラルネットワークに対するテストケースの定量的な測定が可能になる

ニューラルネットワークのカバレッジ率が増加するように入力値を逆算して生成

<ロバストネス要件例>

“太い文字でも認識できること”

メタモルフィックテストのテストケース作成時に
考慮するロバストネス要件をデータの加工処理に利用

要件に基づき、
ユーザが加工処理を
ツールへ実装

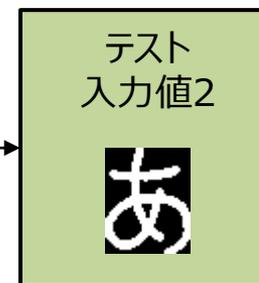


テスト
入力値1

- 加工処理例
- 1pt太くする
 - 2pt太くする
 - 3pt太くする

ニューロンカバレッジ
ツール

2pt太くする



テスト
入力値2

カバレッジ率を増加させる加工処理を
ツールが選択し適用

カバレッジ率を増加させる
データを生成

ロバストネス要件をもとにカバレッジ率を向上させるテストデータを作成できる

本施策のポイント

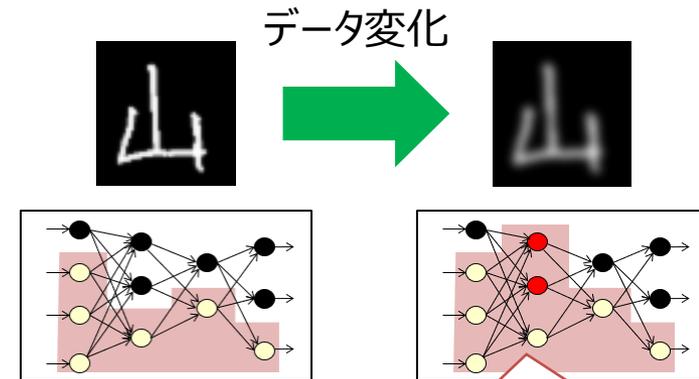
メタモルフィックテストのテスト十分性向上施策として
ニューロンカバレッジ技法を適用



AIモデルの特性分析

AIモデルのカバレッジ率がロバストネス要件による
データ変化に対してどのような遷移を示すか分析

カバレッジ率上昇が大きいデータ加工範囲が
テスト十分性向上に寄与すると考える

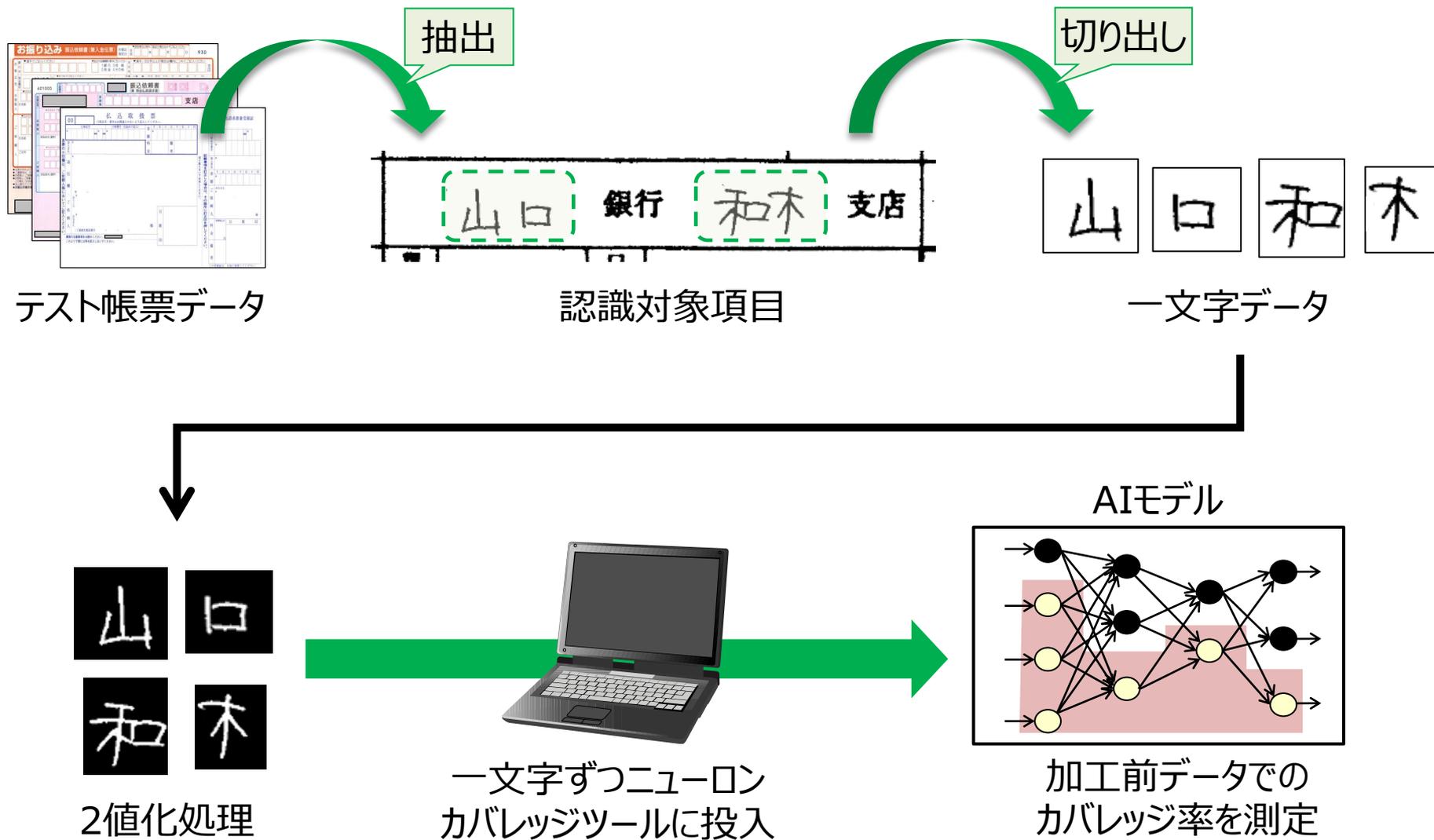


活性化ニューロンの増加
⇒ カバレッジ率上昇

メタモルフィックテストの効果的なテストケース範囲を推測できると仮定

4-5. 施策の手順①

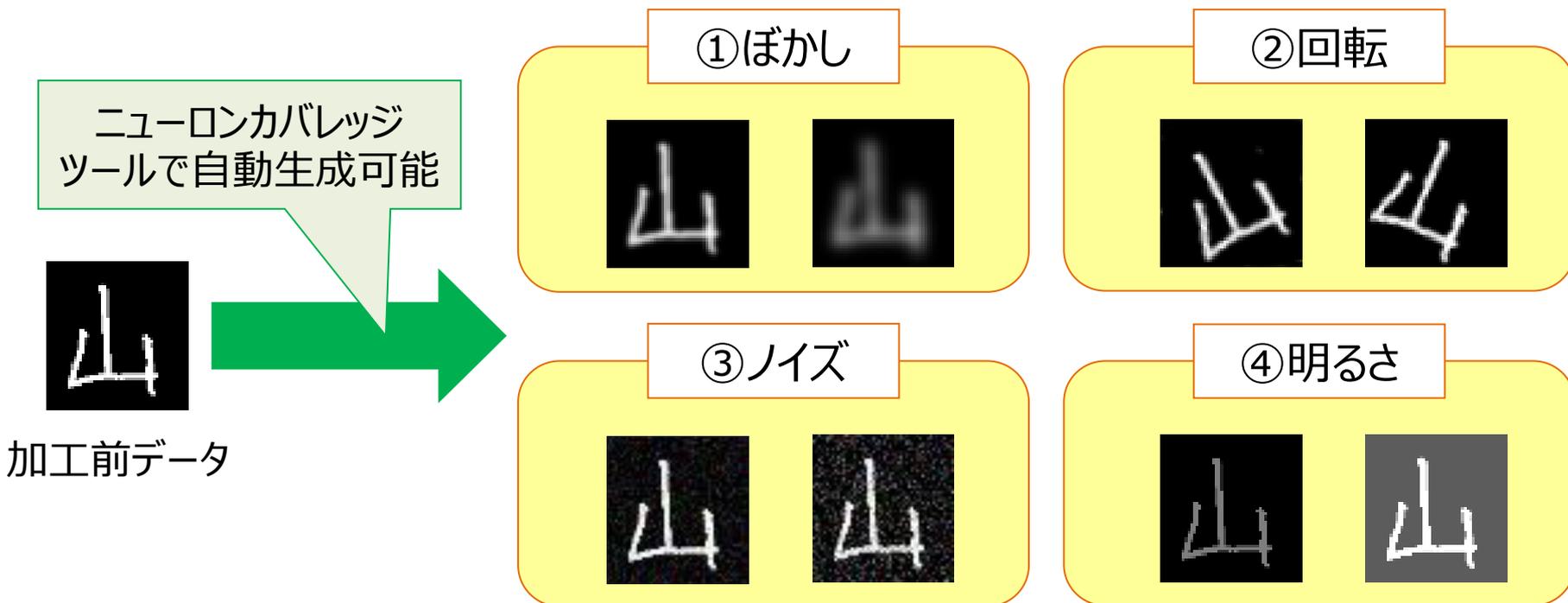
手順①：帳票データから文字データを切り出し、加工前データでのカバレッジ率を測定



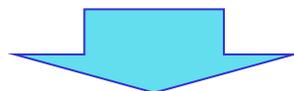
4-6. 施策の手順②

手順②：4種類のロバストネス要件(※)をもとに文字データに外乱を施す

※ 実際の発生頻度が高い要件を選定



指定した刻み幅で徐々に加工を施す

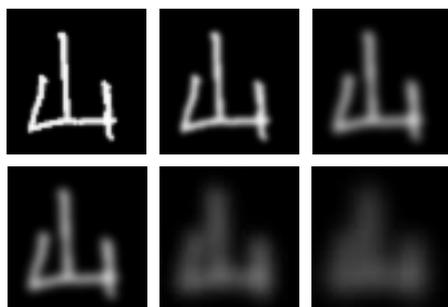


カバレッジ率の遷移測定に最適なテストデータセットを用意できる



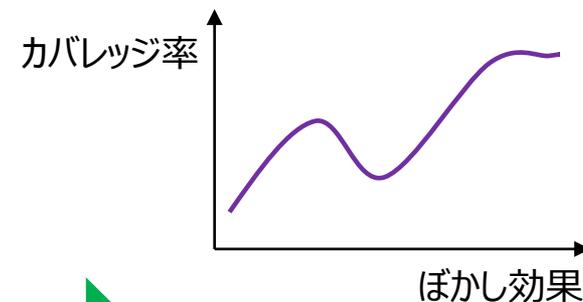
手順③：作成したテストデータセットをツールに入力し，カバレッジ率遷移を測定

(例)ぼかし画像



テストデータセット

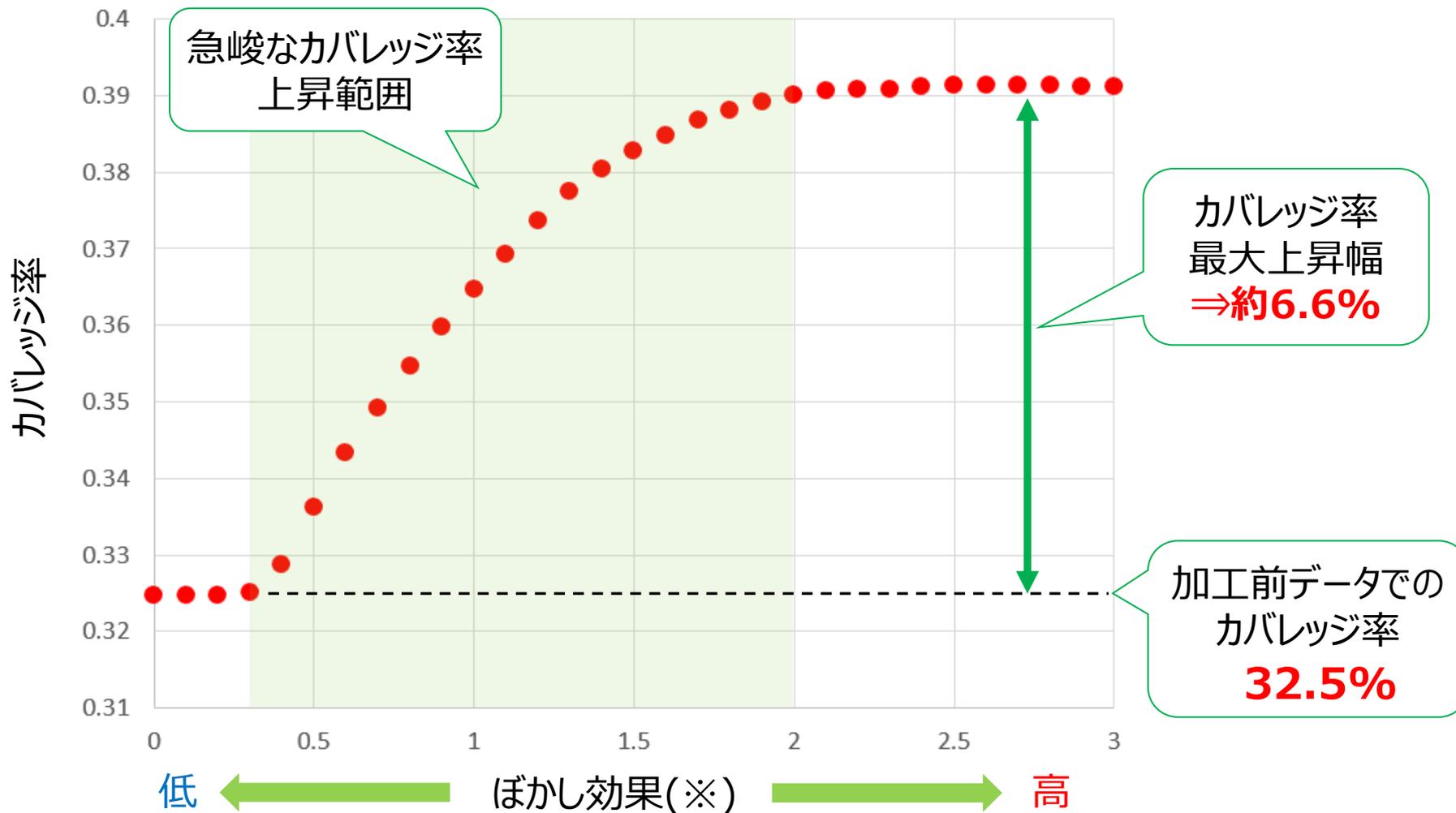
データを順々に
ツールに入力



カバレッジ率がどのような
遷移を示すか分析

分析結果からメタモルフィックテストのテスト十分性向上に
効果的なテストケース範囲を明確にする

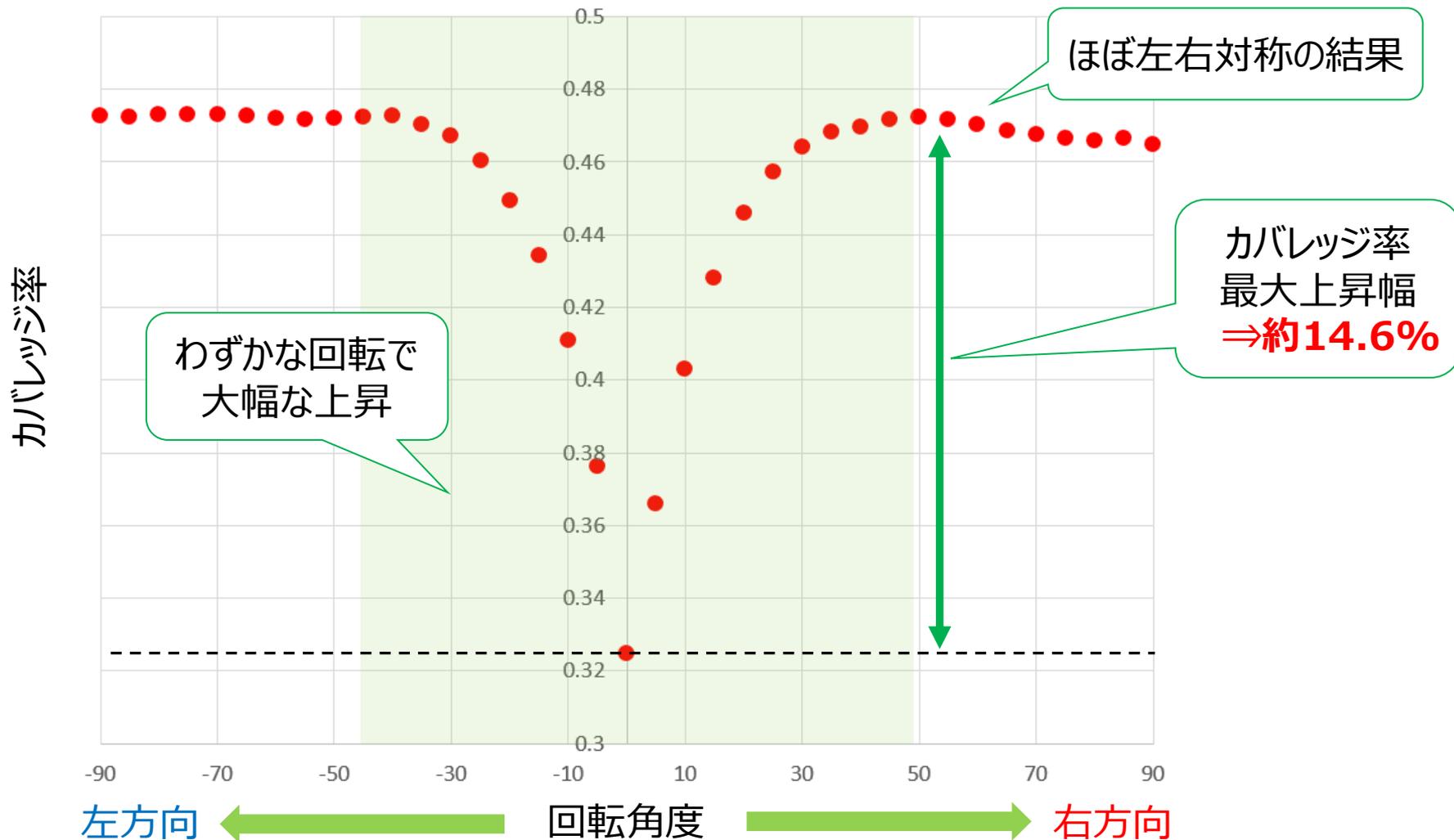
5-1. カバレッジ率遷移の分析結果①ぼかし



※ 横軸：ガウシアンフィルタで用いる正規分布の標準偏差

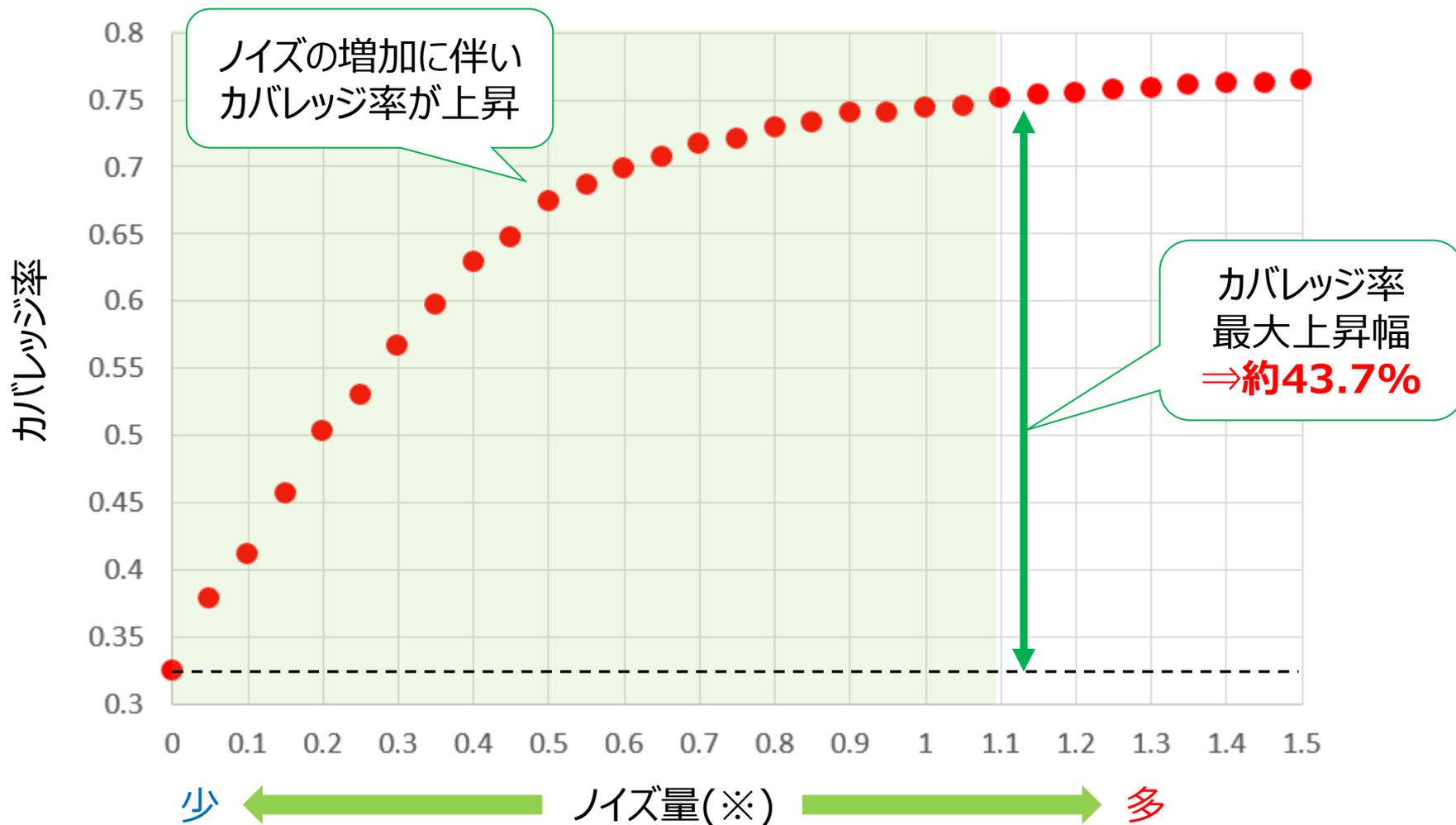
0.3～2.0の範囲がテスト十分性向上に大きく寄与する

5-2. カバレッジ率遷移の分析結果②回転



どちらか一方方向に0°～50°回転させたテストケース作成が効果的

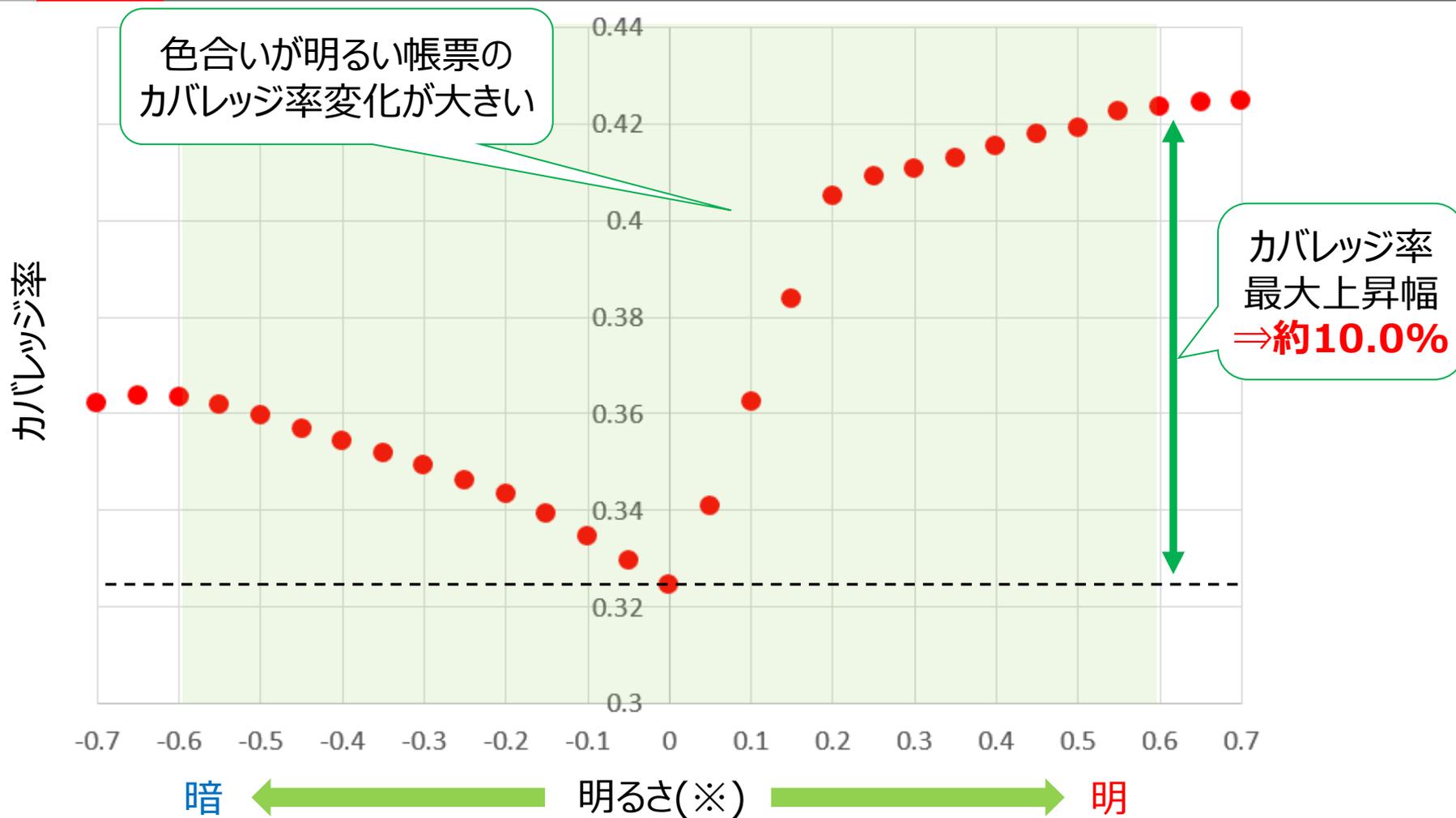
5-3. カバレッジ率遷移の分析結果③ノイズ



※ 横軸：ノイズ生成で用いる正規分布の標準偏差

0.0～1.1の範囲のテストケースがテスト十分性向上に大きく寄与する

5-4. カバレッジ率遷移の分析結果④明るさ



※ 横軸：輝度値

0.0~0.6の明るい色合いの帳票データ作成が効果的

5-5. カバレッジ率遷移分析結果のまとめ

	ロバストネス要件	カバレッジ率上昇に寄与する範囲	最大カバレッジ率 (%)	カバレッジ率の最大上昇幅 (%)
①	ぼかし	0.5 ~ 2.5	39.1	6.6
②	回転	0° ~ 50°	47.2	14.7
③	ノイズ	0.0 ~ 0.8	76.2	43.7
④	明るさ	0.0 ~ 0.6	42.5	10.0

カバレッジ率遷移を測定

テスト十分性向上に寄与するテスト範囲の明確化

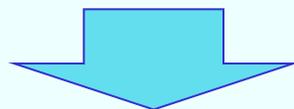
メタモルフィックテストのテストケース作成の効率化



本施策を通して得た知見を生かし、
今後のメタモルフィックテストのテストケース作成を推進する

本報告のポイント

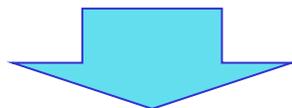
メタモルフィックテストのテスト十分性向上施策として、
ニューロンカバレッジ技法を用いたAIモデルの特性分析を提案



カバレッジ率上昇に大きく寄与する加工範囲を分析し、
メタモルフィックテストの効果的なテストケース作成に繋がる知見を得た



外乱変化が大きくなるに伴いカバレッジ率は上昇するが
同時に実際の帳票では明らかに存在しないケースも増加傾向に



テストケースを検査時にどこまで採用すべきか
検討する必要がある



最終的な目標はメタモルフィックテストのテスト十分性評価

Step 1 : 検査用帳票データの作成

本施策で得た知見をもとに帳票データを作成

➡ 再度ニューロンカバレッジツールに投入し、
本施策と同様のカバレッジ率が得られるか検証する



Step 2 : 残存する課題事項の検討

① 品質を担保するテストケースの量的/質的基準

➡ カバレッジ率を品質指標として基準策定をめざす

② 基準を満たしているか否かの評価方法

➡ ニューロンカバレッジ技法により評価可能になると考える

