



Deep Learning搭載ソフトウェアの 品質評価/テストを前進させる取り組み

2019/08/30

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

Contents

1. 背景
2. 帳票認識ソフトウェアの開発について
3. 課題
4. 施策
5. 施策の効果
6. まとめと今後の課題



1. 背景

AI (Artificial Intelligence : 人工知能)

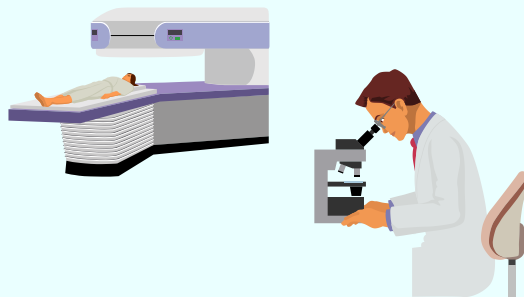
推論・認識・判断など、人間と同じ知的な処理能力をもつコンピュータシステム

製造業



故障検知 人員配置
生産計画最適化

医療



画像診断
癌の存在予測

交通



渋滞予測 配車計画
自動運転

AI搭載ソフトウェアの活用分野が広がっている

一方、AI搭載ソフトウェアの品質保証技術の進歩は大幅に遅れている



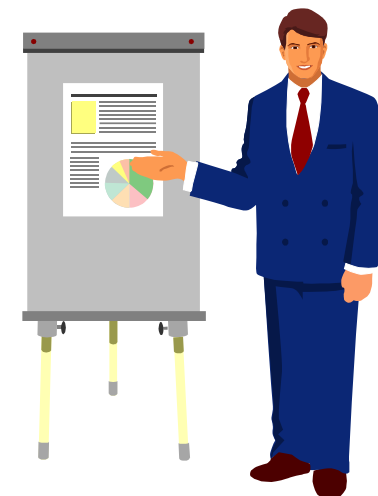
品質確保の考え方やテスト技法が提案され始めているが
実案件への適用例は少ない

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は重要テーマ

AI搭載

帳票認識ソフトウェアの開発を担当

⇒品質確保に向けて取り組んだ施策を報告



AI搭載ソフトウェアの品質保証を前進させる取り組み



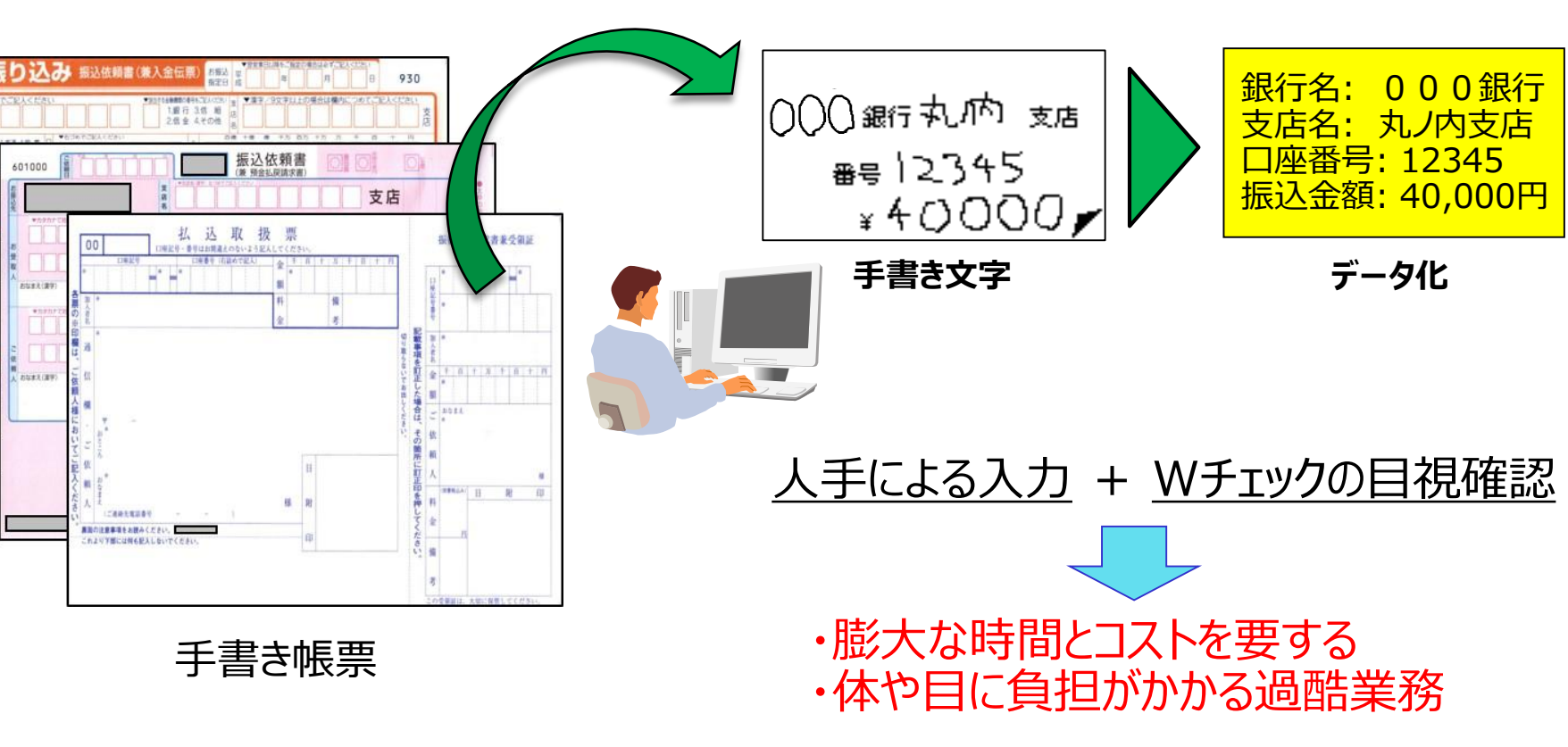
2. 帳票認識ソフトウェアの開発について

2-1 生活に根付く帳票



我々の暮らしの中には様々な手書き帳票が存在

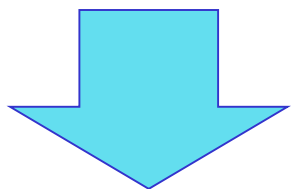
ビジネスの効率化/コスト削減には、手書き文字のデータ化が必須



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

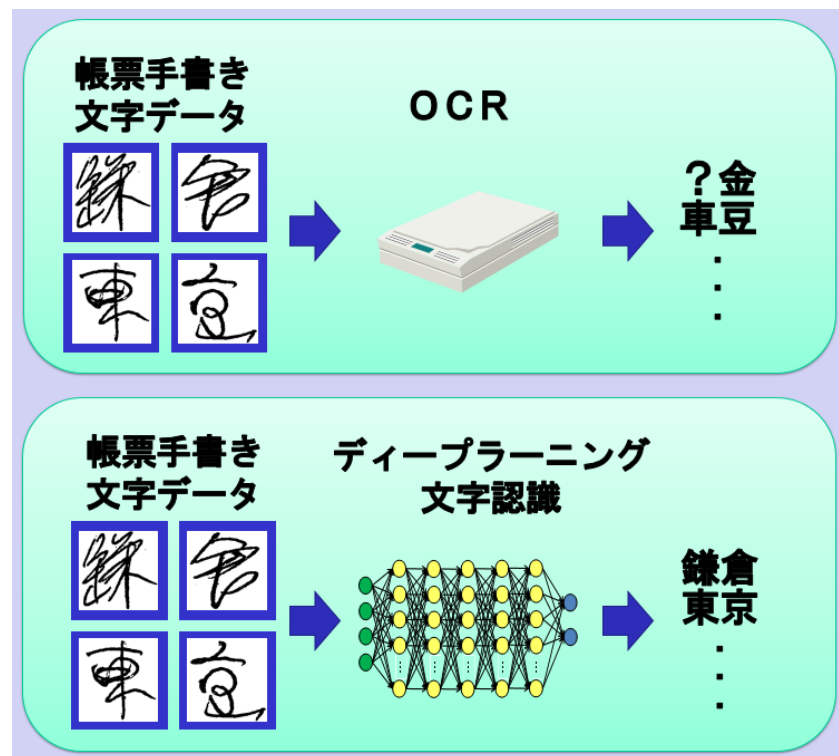
【OCRの課題】

従来のOCRは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技術との組み合わせ

⇒読み取りにくい手書き文字の認識精度向上



ディープラーニングとは

⇒機械学習手法の一つで、コンピュータが大量のデータから自動的に特徴を抽出する手法

2-4 本開発におけるQAの役割

開発部署



設計/実装/テスト実施

開発製品



品質保証部署(QA部署)



開発部署とは独立して
第三者視点で製品検査を実施

ディープラーニングの導入により、OCRの課題を克服できているか
QA部署独自の検査データを作成して確認する



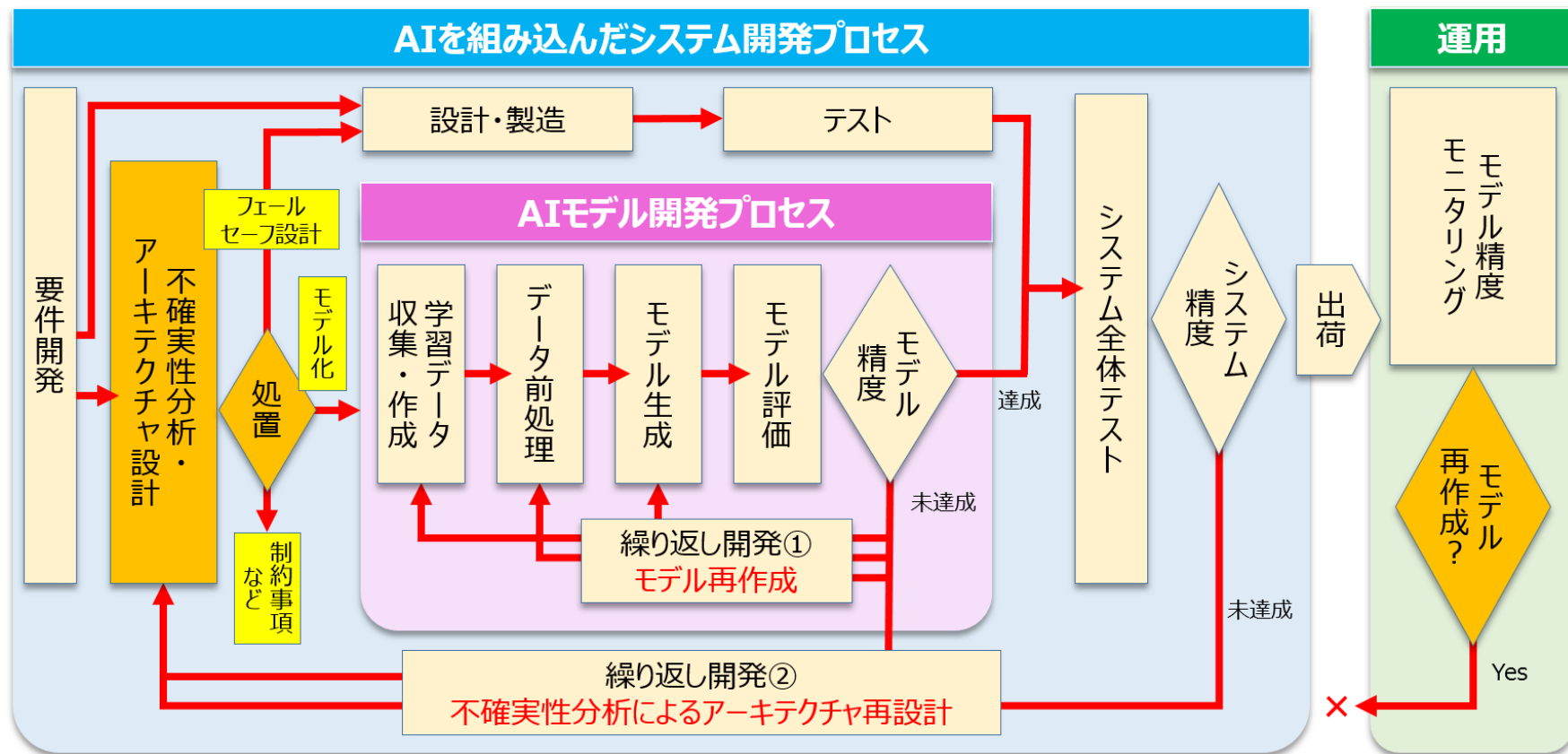
3. 課題

3-1 帳票認識ソフトウェアの品質確保に向けた課題

課題

AIモデルのテストにおいて、従来のテスト設計が通用しない

3-2 AI搭載システムの開発プロセスと品質観点



品質観点

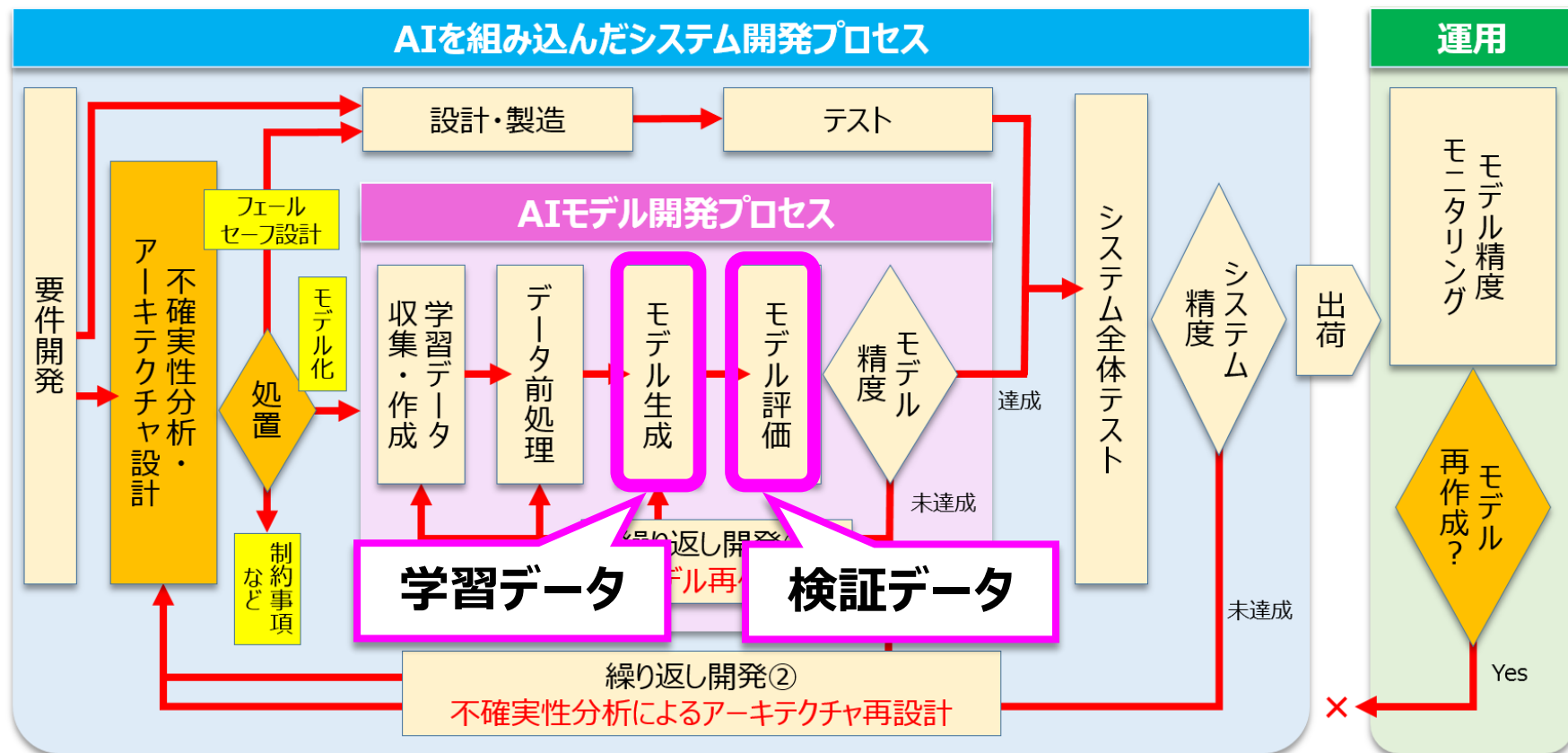
(A)データの品質

(B)AIモデルの品質

(C) システムの品質

(D)運用の品質

3-2 AI搭載システムの開発プロセスと品質観点



品質観点

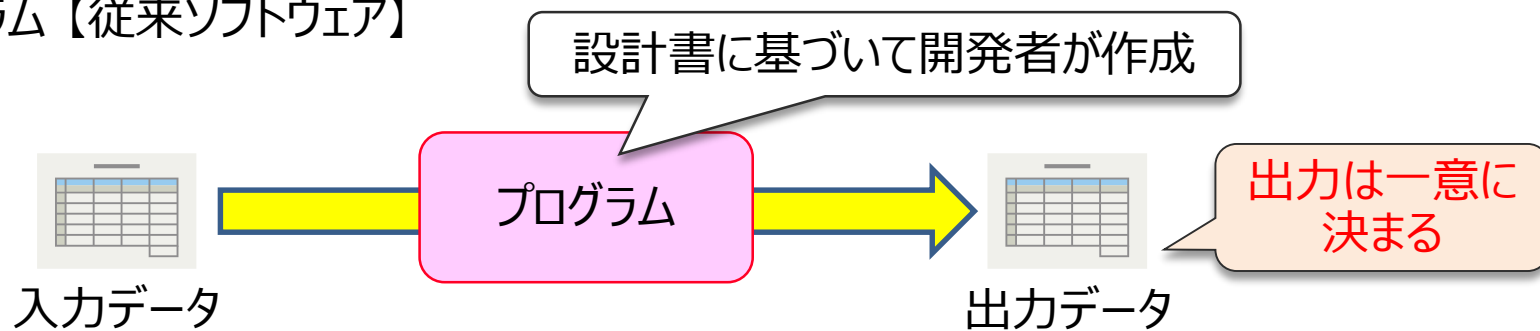
(A) データの品質

(B) AIモデルの品質

(C) システムの品質

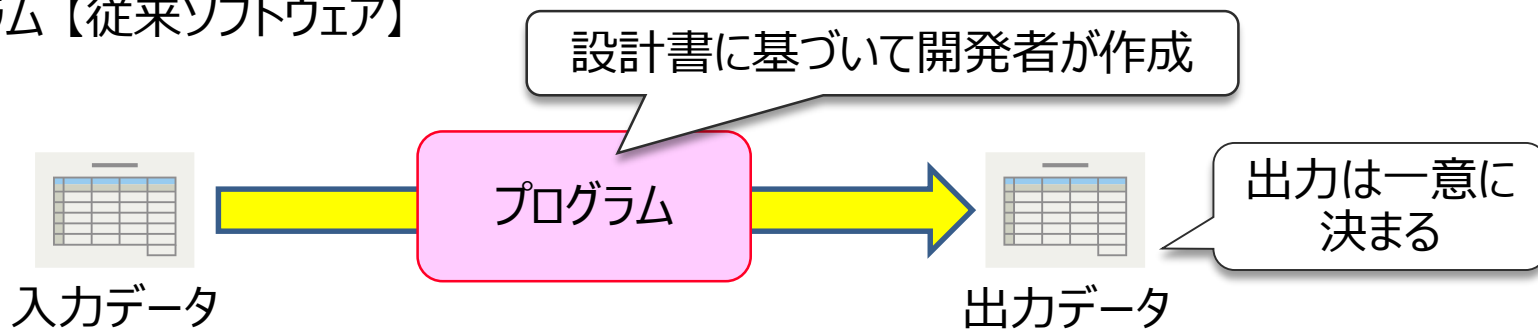
(D) 運用の品質

◆演繹的プログラム【従来ソフトウェア】

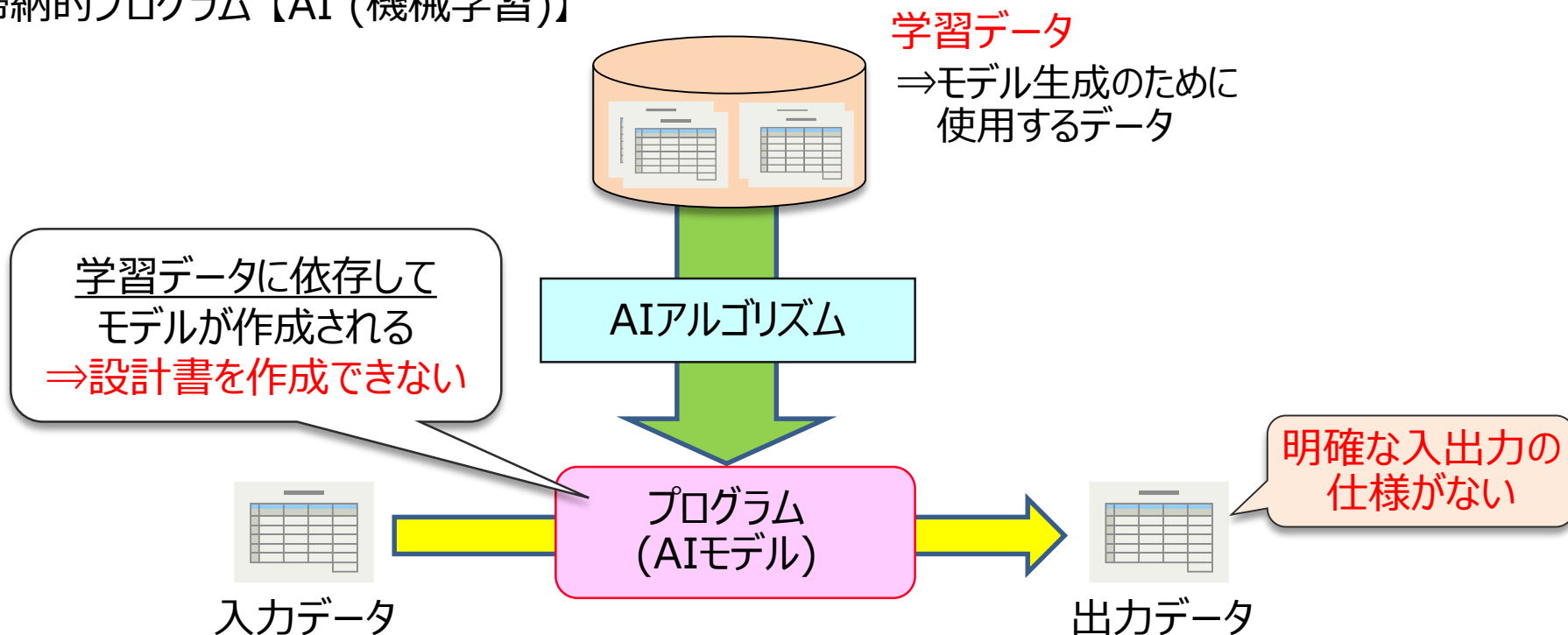


3-3 AIモデルの特徴

◆演繹的プログラム【従来ソフトウェア】

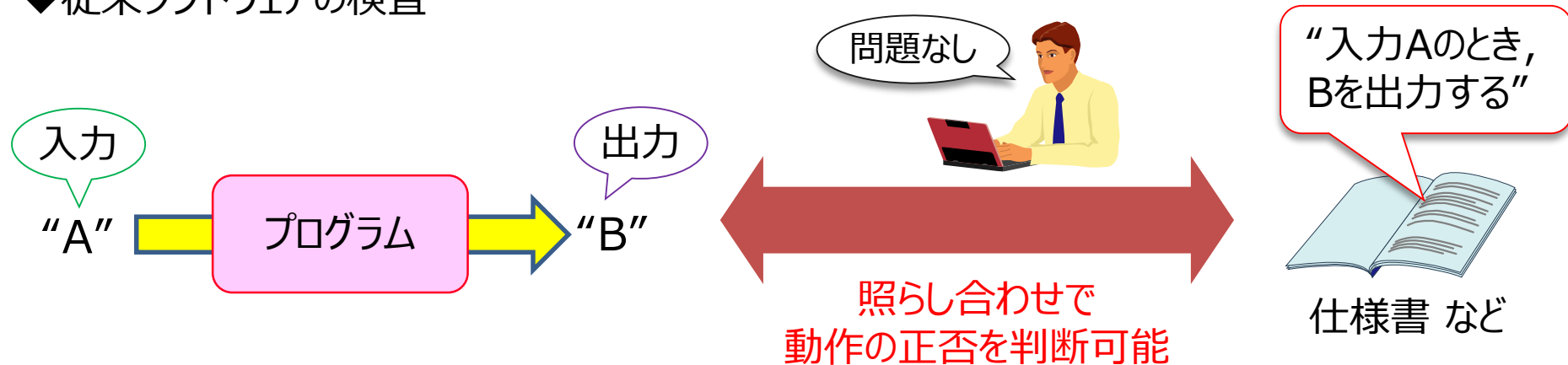


◆帰納的プログラム【AI (機械学習)】



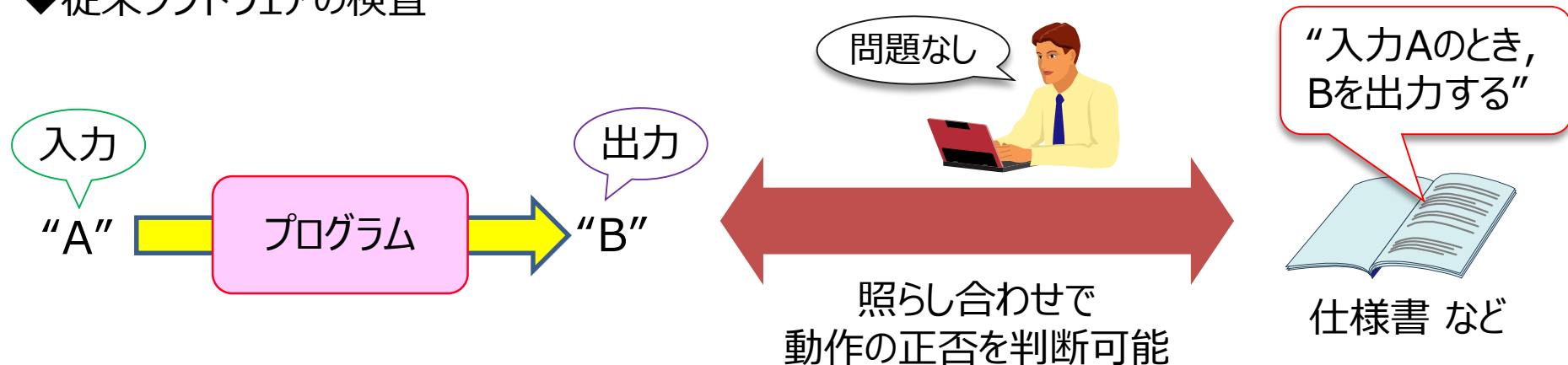
3-4 AIモデルの検査における問題点①

◆従来ソフトウェアの検査

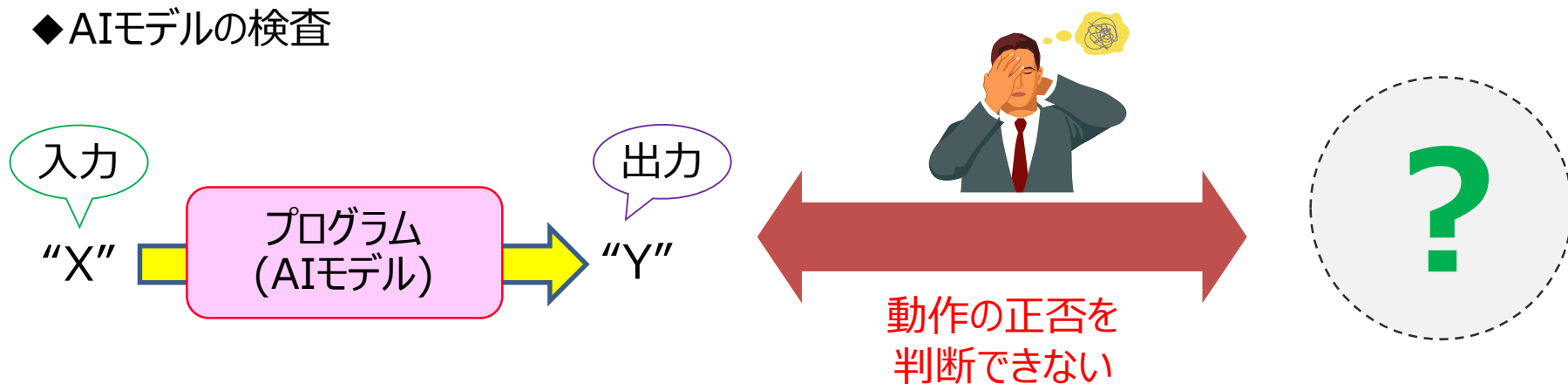


3-4 AIモデルの検査における問題点①

◆従来ソフトウェアの検査



◆AIモデルの検査



AIモデルでは入力に対する絶対的な正解が不明

従来のソフトウェアテストと同様の入出力を意識した検証(Verification)が不可能

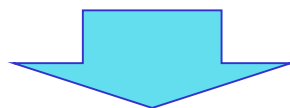


妥当性確認(Validation)でAIモデルの品質を評価する

妥当性確認のための方法は？




AI案件は初担当であり、ノウハウがない



妥当性確認のためのテスト設計を検討する必要がある





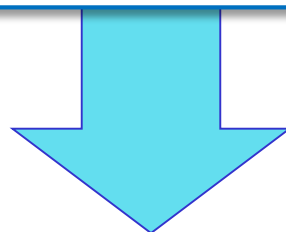
4. 課題に対する施策

課題

AIモデルのテストにおいて、従来のテスト設計が通用しない

課題

AIモデルのテストにおいて、従来のテスト設計が通用しない



施策

メタモルフィック関係を利用したQA独自の
検証データを作成し、テストをする

QA検査でAIモデルの妥当性確認をしたい

ブラックボックス
テスト

DNNカバレッジテスト ❌

制約条件なし

形式手法 ❌

検証データの作成

メタモルフィックテスト

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

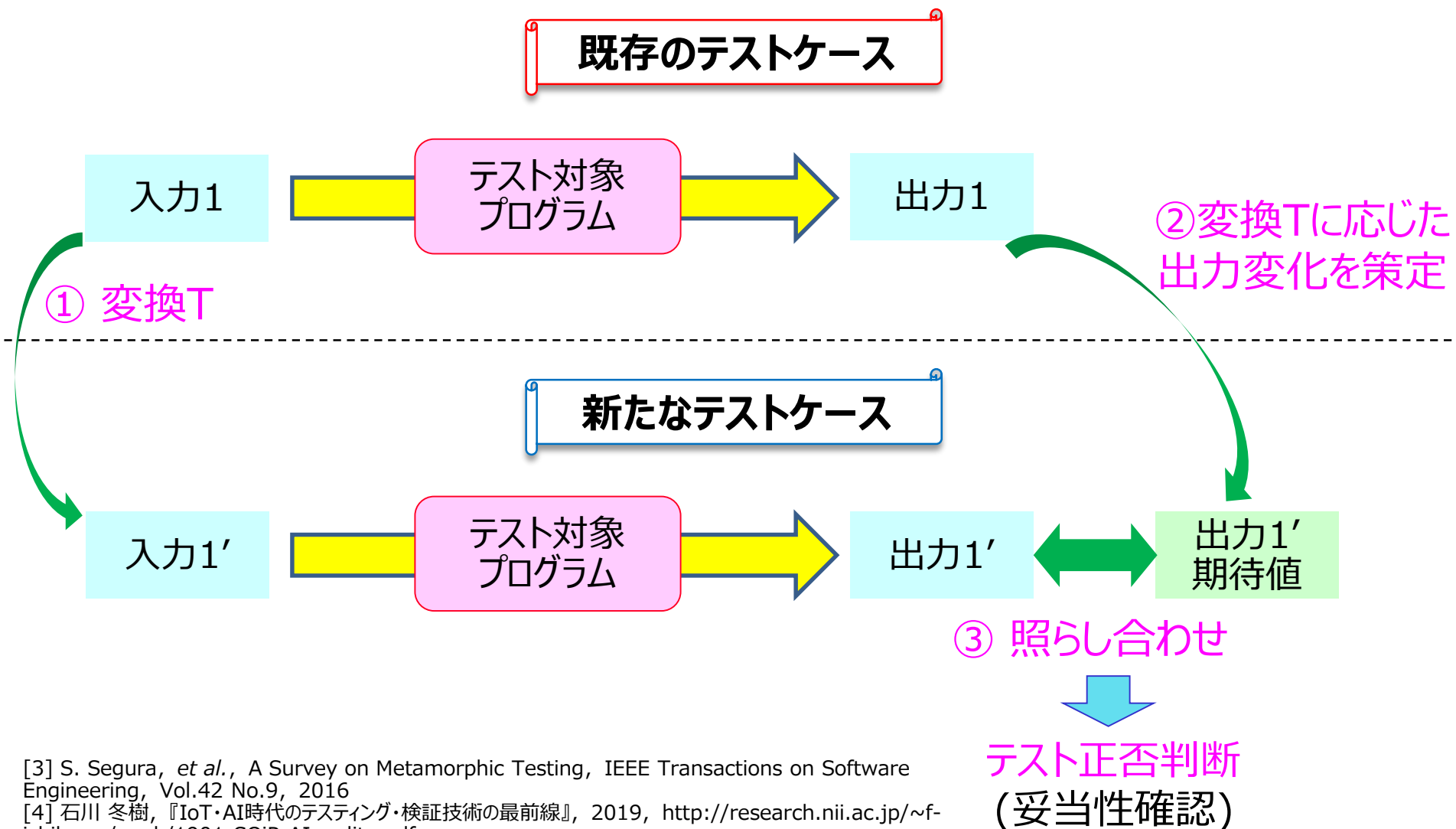
テストケース間の出力結果の比較を繰り返し、
プログラムの妥当性を確認



[1] 『AIプロダクト品質保証ガイドライン』, AIプロダクト品質保証コンソーシアム(QA4AIコンソーシアム)編, 2019.05版 2019年5月17日公開, <http://www.qa4ai.jp/>
[2] Tsong Yueh Chen, et al., Metamorphic Testing: A Review of Challenges and Opportunities, ACM Computing Surveys, Vol.51 No.1, 2018

4-4 メタモルフィックテストについて②

◆メタモルフィックテストの概念図

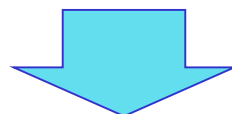


[3] S. Segura, *et al.*, A Survey on Metamorphic Testing, IEEE Transactions on Software Engineering, Vol.42 No.9, 2016

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

4-5 メタモルフィックテストング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>

4-6 メタモルフィックテストの適用例①

<例1> 入力数列の最小値を求めるプログラム

数列の入れ替え
⇒ **Permutative**

変換T

{24, 27, 22, 20, 29} ⇒ 20



{29, 24, 27, 22, 20} ⇒ 20

結果は
同じになるはず
⇒ テストで確認

<例2> ランキングを出力するプログラム

	好きな動物ランキング
1	犬
2	猫
3	うさぎ
4	ライオン
5	象

変換T

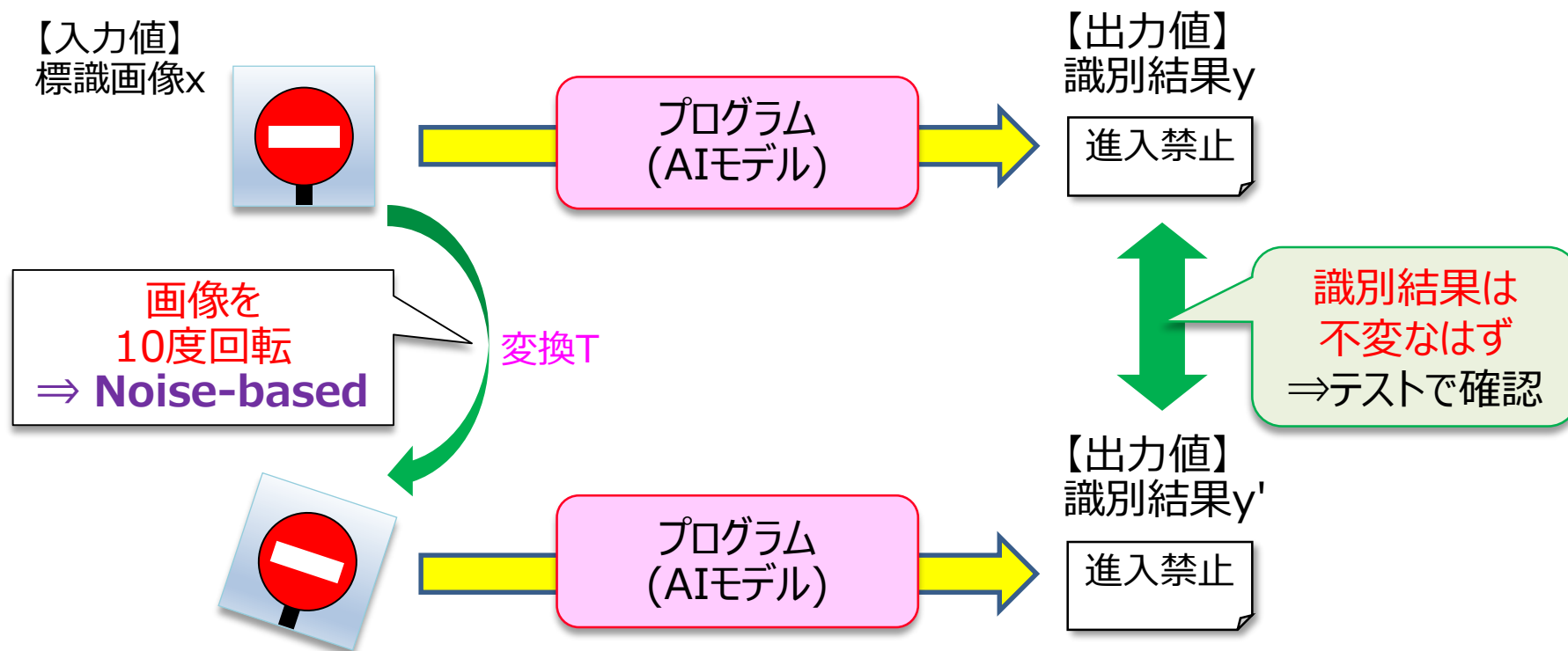
「猫」データを削除
⇒ **Exclusive**

	好きな動物ランキング
1	犬
2	うさぎ
3	ライオン
4	象

結果はこのように
変化するはず
⇒ テストで確認

4-7 メタモルフィックテストの適用例②

<例3> 道路標識を識別するプログラム(AIモデル)



① 入力帳票に対する変換を考えやすく、多様なテストケースを容易に作成できる

- ✓ 太文字にする
- ✓ 文字の色を変える
- ✓ 文字の間隔を広くする
- ✓ 帳票の角度を変える … etc

メタモρφック関係に
基づく変換例

“日立”

変換

日立

日立

日立

② 他のテストケースとの相対比較から認識の弱点を予測しやすい



北海道

北海道

照らし合わせ



北海道

北海“進”

下線付与
⇒ Noise-based

下線の変化に
弱いかも…

メタモρφックテストによるAIモデルの妥当性確認が
帳票認識ソフトウェアでも有効と予測



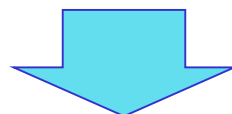
テストケース作成時の課題

メタモルフィック関係は無数のパターンが考えられ、
テストケースが膨大となる

きりがない…



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>

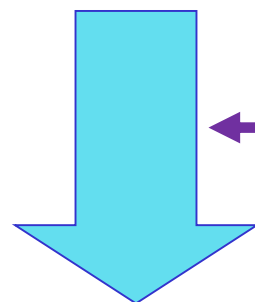
テストケース作成時の課題

メタモルフィック関係は無数のパターンが考えられ、
テストケースが膨大となる

きりがない…



現実的なテストケース数で最大限の効果を得るため、
帳票に対する変換を工夫する必要がある

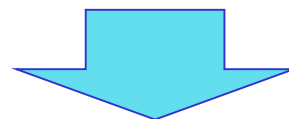


本検査の目的
⇒従来のOCRの課題を
克服できているか確認する

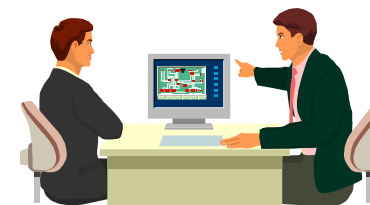
施策

従来のOCRが不得意としていた認識パターンを
メタモルフィック関係に基づく変換に利用する

OCRが不得意とする認識パターンの洗い出し



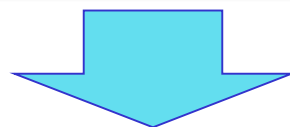
社内の有識者にヒアリング



**従来OCRの
不得意パターン**

- ① 帳票の背景色が濃いケース
- ② 項目値が複数行になるケース
- ③ 文字と罫線が接触するケース
- ④ 文字の形が似ているケース
- ⑤ 手書き文字

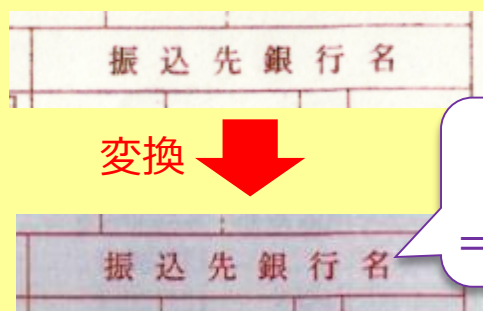
…etc



**得た知見をメタモルフィック関係に
基づく変換に利用**

意図的に変換を施してテストケースを派生

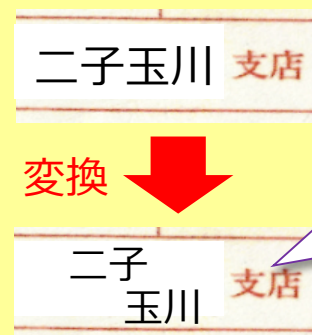
①背景色が濃いケース



背景色を
濃くする

⇒ Noise-based

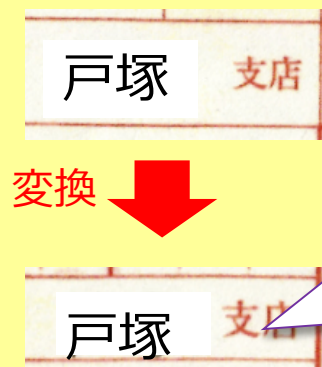
②項目値が複数行のケース



複数行で
記入する

⇒ Noise-based

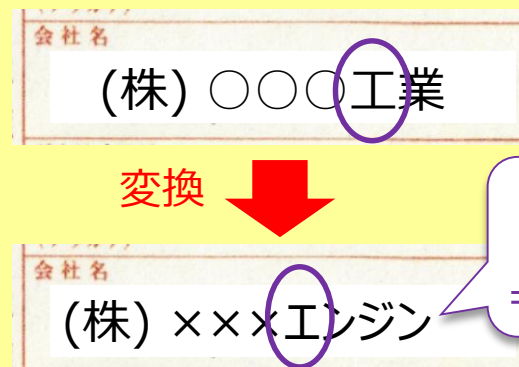
③文字と罫線が接触するケース



罫線と接触
させる

⇒ Noise-based

④文字の形が似ているケース



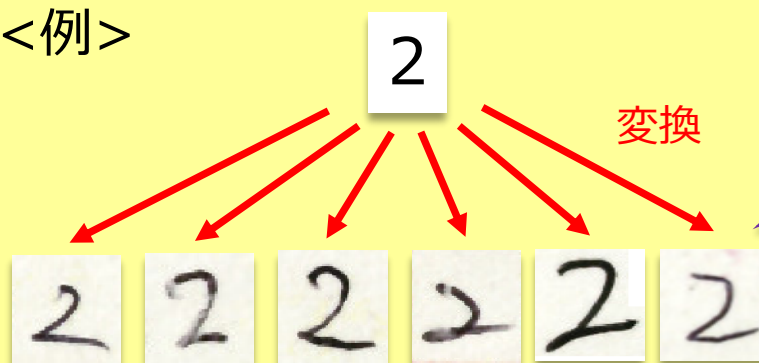
片仮名の
"工"に変える

⇒ Heuristic

⑤手書き文字

同じ字でも様々な書き方や字体があり、
書き手によって字が変化する点をメタモρφック関係として利用

<例>



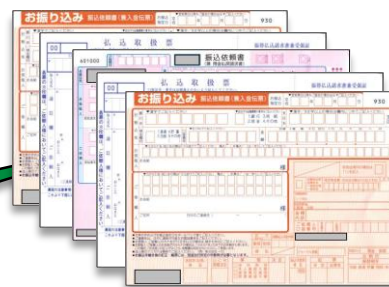
手書き文字の変化
⇒ Noise-based
⇒ Semantically
Equivalent

手書き文字の特徴は様々
(くせ字, 崩し字, 流れ字など)

QA部署メンバー



帳票記入依頼



多様な手書き文字
データを収集

担当QA



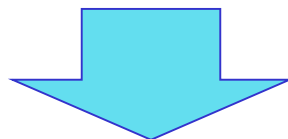
認識精度検証



5. 施策の実施と効果

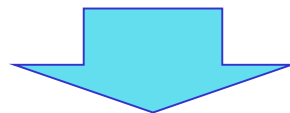
メタモルフィックテストを用いたQA検査

正しく認識できなかったケースを抽出し、学習データの不足や
内部ロジックの不良といった問題を検出する



施策の効果を確認

QA検証データの認識率を測定



認識結果に対する見解を開発部署にフィードバック

OCRで認識不可能
だった文字が
認識できています

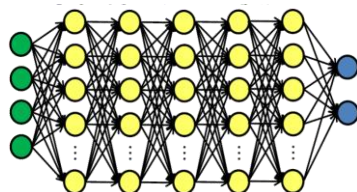
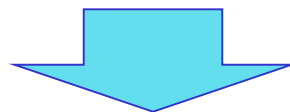
特定の字種において
認識率が低下しています

崩し字の認識が
まだ少し弱いです

文字が罫線と重なる
ケースの認識結果が
間違っています

メタモルフィックテストによる検査を通して、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

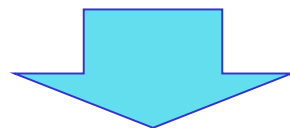
認識に失敗したテストケースの原因調査



- ・内部ロジックの不良
- ・学習データの不足



OCRの苦手な認識パターンを克服できたケース、
克服できていないケースの明確化



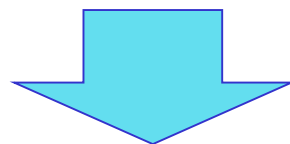
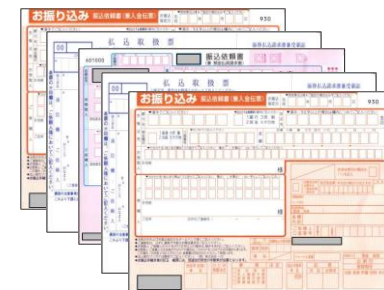
メタモルフィックテストングが帳票認識ソフトウェアにおける
AIモデルの妥当性確認に有効



6. まとめと今後の課題

帳票認識ソフトウェアの品質確保に向けたQA施策

- ✓ メタモルフィックテストによる検証データの作成
およびAIモデルの妥当性確認
- ✓ OCRの苦手パターンをメタモルフィック関係に基づく変換に
利用した効果的な検査



多くの分野での活用が予想される
帳票認識ソフトウェアの品質確保に繋がる知見

AIテスト技法を実案件に適用し、
AI搭載ソフトウェアの品質評価/テストを前進させた

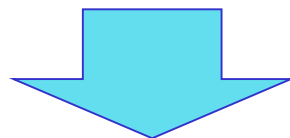


現状の問題点

テスト十分性を評価できていない

➡ 以下を確認できればテスト十分性を評価可能と考える

テストケースが量的/質的基準を満たしていること



今後の課題

下記事項の検討が必須

- 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



END



Deep Learning搭載ソフトウェアの品質評価/テストを前進させる取り組み

2019/08/30

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

HITACHI
Inspire the Next 