

# モデルの性能劣化を検出するドリフト検知手法の検討

キヤノンイメージングシステムズ株式会社  
 東芝デジタルソリューションズ株式会社

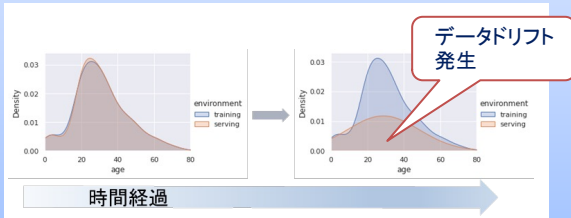
丸山 優  
 矢島 旭

株式会社デンソー

千葉 裕介

## ドリフトとは

- データの予期しない変化によって、予測モデルの精度が劣化する現象
- 「データドリフト」とは、入力データの統計的分布が変化すること



## 課題と目標

### 課題

データセットやユースケースに応じて、複数のドリフト検知手法から選択しなければならない

### 目標

手法を調査・比較し特徴をまとめ、選択のための指針を作成する

## 調査のアプローチ

データ  
入力

扱えるデータの種類を調査

ドリフト  
検知処理

検知アルゴリズムや閾値設定等の調整の自由度を調査

検知結果  
出力

Yes/Noのみかレポート形式か等の出力方法を調査

## 調査結果：手法の比較表

対象	未知のデータの通知	得意なデータ形式	扱えるデータ		ドリフト検知結果の出力方法			調整の自由度	
			数値	文字列	レポート出力	ドリフト検知の判断	ドリフト検知の理由説明	検出アルゴリズムの数	閾値設定
Tensorflow Data Validation (TFDV)	○	テーブルデータ TFRecord	○	○	○	○	○	2	○
scikit-multiflow. drift_detection	×	一次元の 時系列データ	○	○	×	○	×	7	アルゴリズムによる
TorchDrift	×	画像データ	○	×	×	×	×	2	-
Evidently AI	×	テーブルデータ	○	○	○	○	×	18	○

- 「TFDV」はレポートがリッチで比較元データに存在しない未知のデータが発生したことを通知する
- 「TorchDrift」はドリフト検知判断材料となるデータのみ提供するので、別途ドリフトを検知したかどうかの判断処理が必要

## 評価：2つのユースケースへの適用

### テーブルデータ(TFDV、Evidently AI)： 運賃予測AIのケース

- 未知の値を検知したい
- 検知結果の詳細なレポートが欲しい

Anomaly short description	Feature name	Anomaly long description
Unexpected string values	'payment_type'	Examples contain values missing from the schema: Prcard (<1%).

未知の支払方法  
(Prcard)を検知

Anomaly short description	Feature name	Anomaly long description
High Linty distance between training and serving	'payment_type'	The Linty distance between training and serving is 0.0225 (up to six significant digits), above the threshold 0.01. The feature value with maximum difference is: Credit Card

#### TFDVレポート

Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
payment_type	cat			Detected	Jensen-Shannon distance	0.018164

データセットの異常を検知

#### Evidently AIレポート

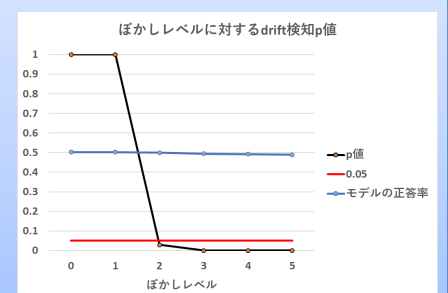
統計情報と結果がまとまっている

### 画像データ(TorchDrift)： カメラ画像識別AIのケース

- カメラ故障・異常による入力データの変化を検知したい



性能劣化が微小でも  
ドリフト検知



## 今後の課題

- 指針(比較表)の改善・拡充
- 4ライブラリの深掘り調査
- 自然言語・音声等の追加調査
- コンセプトドリフトに対応
- チートシートの拡充

