

# プリンター使用状況の計測・分析における 機械学習を用いたデータ・クラスタリングの検討

キヤノン株式会社

菅原和平

sugawara.wahei @ mail.canon

## 開発における問題点

プリンターの使われ方を計測し、躓き箇所を分析することで使い勝手を改善したい。  
しかし、UXテストでの計測では多様化するユーザー環境の模擬が難しい。プリンター使用状況の分析では項目の組み合わせが多く、明確な基準での着目が難しい。また、組込み機器では保存や送信できるデータ量に制約がある。

## 手法・ツールの適用による解決

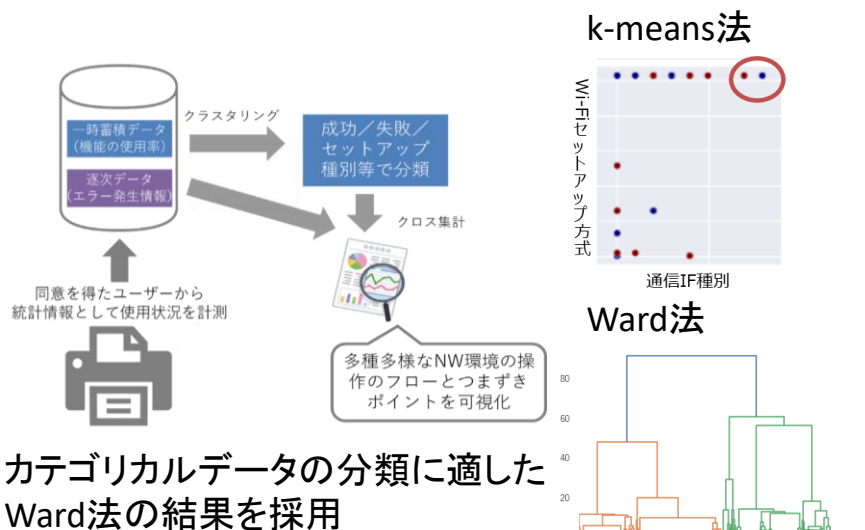
プリンターの初期セットアップの操作フローとネットワーク接続エラーを題材に、プリンター使用状況から躓き箇所を分析する。  
躓き箇所の分析で着目すべき操作や属性を抽出するために、機械学習によるデータ・クラスタリングを適用する。カテゴリカルデータの分類に適したWard法の結果を採用し分析に用いた。

## 着目するクラスタの抽出

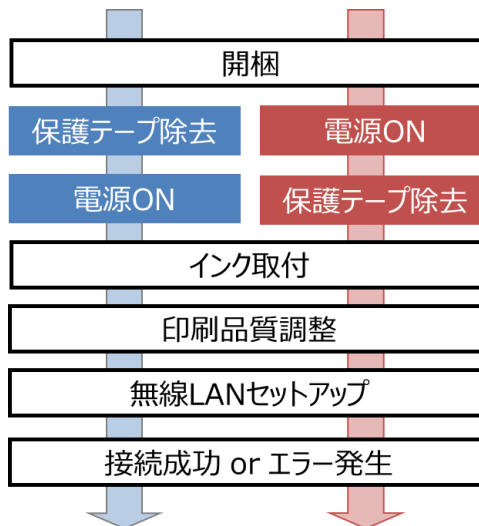
1. 前処理: 外れ値の除去、カテゴリカルデータを数値化
2. 分類 分析対象として着目するクラスタを、k-means法 (非階層的クラスタリング) と Ward法 (階層的クラスタリング) で抽出した

## 典型的な操作フローの抽出

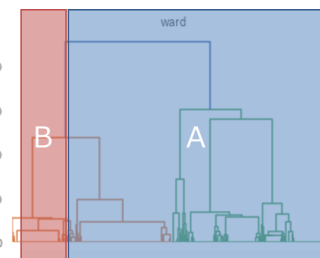
3. 抽出したクラスタごとに典型的な操作フローを抽出  
操作フロー毎の無線LANエラー率の差を集計した



### 初期セットアップ手順の例



クラスタごとに典型的な操作フローを抽出  
「保護テープ未除去」有のクラスタは、無線LANエラー率が高い傾向があった



## 効果と注意点

### <効果>

- ・人手による分類と比較して、より多くの項目間の関係を可視化出来た
- ・項目間の関係に基づいて、着目すべき項目を自動で選択することが出来た

### <注意点>

- ・前処理では欠損値や外れ値の除去に加えて、圧縮して記録された値の数値化が必要
- ・データ間の因果関係の解釈はドメイン知識が必要
- ・ログ設計時はデータ量削減と機械可読性を考慮

## まとめと今後の取り組み

### <まとめ>

プリンター使用状況の分析に機械学習を用いたデータ・クラスタリングを適用し、効果と注意点を示した

### <今後の取り組み>

- ・データの前処理を自動化し分析のコストを下げ、適用範囲を拡大し、初期セットアップ全体につまづき箇所の可視化を可能にする
- ・項目間の因果関係の説明に適した手法(重回帰分析や決定木)による分析を試す