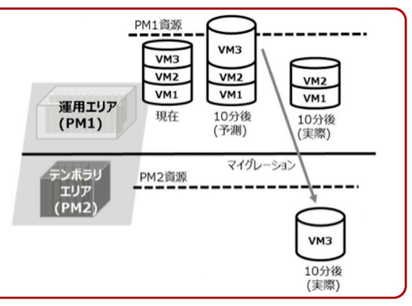
# 小口研究室　研究紹介　(2019年度)
## (お茶の水女子大学理学部情報科学科)

## クラスタリングを用いた時系列データ回帰モデル化手法の提案
## (研究担当:高橋　佑里子)

### 研究概要

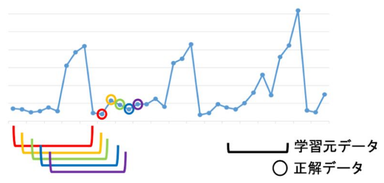
◆ 研究背景
・クラウドサービス事業者では、サーバの仮想化が行われている
・仮想サーバ(VM)を制御する上で、あらゆるVMのCPU使用率を高精度で予測する必要がある
・しかし、現状の予測モデルは汎用性が低い

◆ 研究目的
**VMのCPU使用率の汎用的な深層学習予測モデルの生成に向けた時系列データの回帰モデル化手法の提案を行う**




### アプローチ

◆ 時系列データの学習の特徴
・学習元データ数と正解データ数を設定
・設定された値に基づいて学習元データとその直後の正解データのペアを作成し学習




**→長い波形をそのまま使わず細かい波形を多数使う**

◆ アプローチ
1. 学習に必要な長さでデータセットを細かく抽出
2. 細分化後のデータセットをクラスタリング
3. 2.の結果を再学習データの選定時に活用

### 実験準備

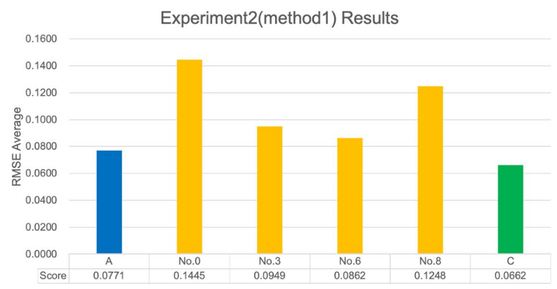
◆ 全体の流れ
1. 前処理
・平滑化：ノイズを減らす
・正規化：最小値を0、最大値を1に揃える
2. 使用するデータ(103個*3種類)の選択
・A：学習元データセット
・B：Aと似ているデータセット
・C：Aと似ていないターゲットデータセット
3. データ粒度最適化
・最適な学習元データ数(=history)を調査
→比較実験により、24であることが判明
4. データセット細分化
・データセットを学習に必要な長さごとに抽出
・24(学習元データ)+1(正解データ)=25点ごと
5. クラスタリング
・k-means法で細分化後のCをクラスタリング
・クラスタ数：10, 20, 30

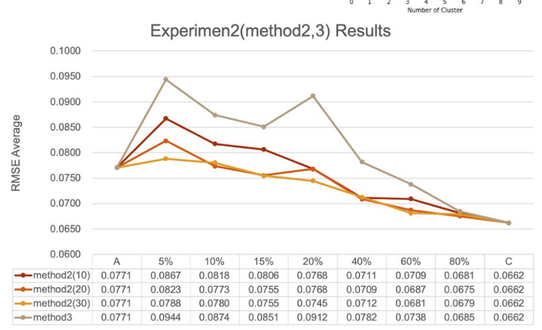
### 実験・考察

◆ 実験1
・既存予測モデルが、学習元と似ているターゲットに対してどの程度の回帰精度を発揮するかを調査
**→ターゲットで学習したモデルと同等の精度を発揮することを確認**

| 学習元 | 予測対象 | RMSE の平均値 |
|---|---|---|
| A | B | 0.1399 |
| B | B | 0.1381 |

◆ 実験2
・既存予測モデルを別のターゲットに応用するための、適切なfine tuningの方法を検討
・方法1：クラスタ数10としたときに多く分類された4つのクラスタ(No.0,3,6,8)のデータのみを使用
**→偏ったデータでfine tuningを行うと精度は低下することを確認**
・方法2：各クラスタのデータを一定の割合で使用
・方法3：細分化前の大きな波形の状態でデータを選択し使用
**→細分化を行うことで、より適切なデータ選定が可能になることを確認**
**→細分化後のクラスタリングが有効とまでは言えないことを確認**


| | A | 5% | 10% | 15% | 20% | 40% | 60% | 80% | C |
|---|---|---|---|---|---|---|---|---|---|
| method2(10) | 0.0771 | 0.0867 | 0.0818 | 0.0806 | 0.0768 | 0.0711 | 0.0709 | 0.0681 | 0.0662 |
| method2(20) | 0.0771 | 0.0823 | 0.0773 | 0.0755 | 0.0768 | 0.0709 | 0.0687 | 0.0675 | 0.0662 |
| method2(30) | 0.0771 | 0.0788 | 0.0780 | 0.0755 | 0.0745 | 0.0712 | 0.0681 | 0.0679 | 0.0662 |
| method3 | 0.0771 | 0.0944 | 0.0874 | 0.0851 | 0.0912 | 0.0782 | 0.0738 | 0.0685 | 0.0662 |

### 今後の課題

・新たな方法を考えたり、fine tuningの規模を変えたりしながら引き続き実験を行う
・回帰の次の段階として、予測に向けた取り組みを行う

## IoTデバイスにおける共通鍵暗号と完全準同型暗号を組み合わせた暗号化の高速化
## (研究担当:松本　茉倫)

### 研究背景

❑ **クラウドと情報漏洩の危険**
ユーザから取得したセンサデータを
クラウドサービスで統計分析したい
→ クラウドサービスは安全とは限らない


❑ **個人情報保護とデータ分析の需要**
企業側　:データ分析結果が欲しい
ユーザ側:個人情報を知られたくない
→ 完全準同型暗号 (FHE) が有用

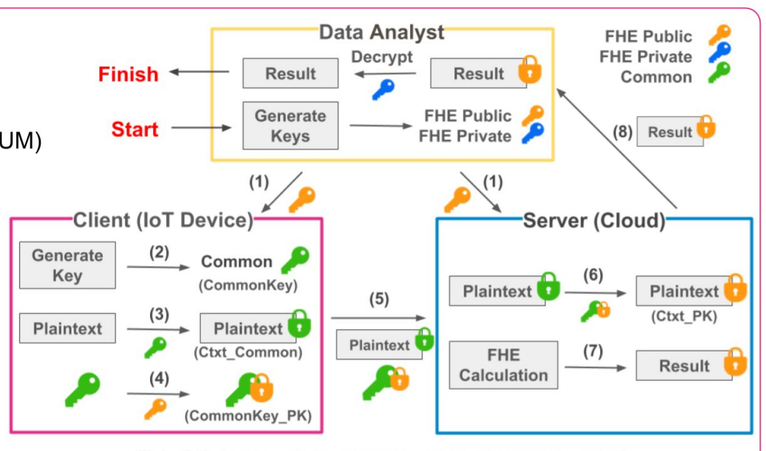
❑ **完全準同型暗号 (FHE)**
暗号文同士の加算・乗算が可能な暗号方式
処理に時間がかかり、暗号文サイズが大きい
→ 低性能マシン (IoTデバイス) において
暗号化時間・通信量を減らしたい

### 実験概要

❑ **従来手法 (FHE only)**
平文をFHEで暗号化する
❑ **提案手法**
共通鍵暗号 (AES, KATAN, TRIVIUM)
とFHEを組み合わせる (右図)
❑ **実験環境**
Client:Google Pixel3
Server:MacBookPro

従来手法・提案手法で3つを比較
1. **Clientへの負荷 (実行時間)**
2. **通信量 (ファイルサイズ)**
3. **Serverへの負荷 (実行時間)**

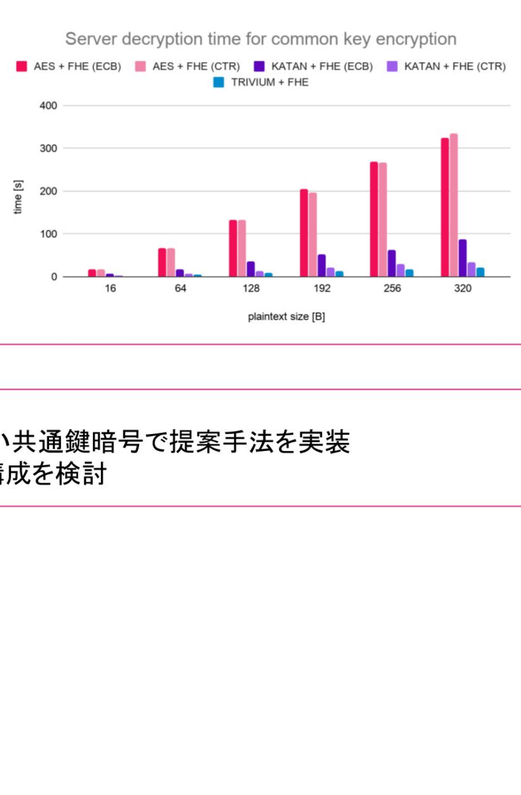
※ブロック暗号のAES・KATANでは
ECBモードとCTRモードを用いる




提案手法 (AES + FHE, KATAN + FHE, TRIVIUM + FHE)

### 実験結果

❑ **平文が短い場合**
提案手法よりも従来手法がClientへの負荷が少なく
通信量も少ない
❑ **平文が長い場合**
従来手法よりも提案手法がClientへの負荷が少なく
通信量も少ない
特にKATAN + FHEのCTRモードが最も効率的
❑ **Serverへの負荷**
AES + FHEは高負荷
TRIVIUM + FHEが最も負荷が少ない










### 今後の課題

❑ AES・KATAN・TRIVIUM以外のFHEと相性の良い共通鍵暗号で提案手法を実装
❑ 提案手法よりもクライアントへの負荷を減らせる構成を検討


http://ogl.is.ocha.ac.jp/
oguchi@is.ocha.ac.jp