

ディープラーニング手法の探索

研究分担者 大塚玲 情報セキュリティ大学院大学・教授

一般家庭の1秒単位の水使用量データと水使用行動を収集し、ディープラーニングの手法を適用することにより、水使用量データのみから水使用行動を推定する手法を開発する。ディープラーニングの手法としては、水使用量データをスペクトログラムによる時間一周波数空間の2次元画像に変換し、畳み込みニューラルネットワークの画像-ラベル識別問題として水使用行動ラベルを推定する手法を検討した。当初、水使用量データのスペクトログラムによる解析を試みたが、実験を重ねた結果、バイアス成分に重点をおいた流量ヒストグラムによる解析性能が高いことが見いだされ、当該流量ヒストグラムを入力とした回帰問題およびラベル識別問題の両方を実践し、ラベル識別問題に関しては95%以上の識別性能を達成した。

機械学習を用いた1秒毎の水使用の流量パターン解析により、家庭全体の使用量を用途別に分類(ディスアグリゲーション)することを目的とする。ディスアグリゲーションの問題は、音声データにおける話者分離 (speech separation), 音源分離(audio source separation)と関係が深い。現時点で最高性能を持つ Autoflow は、オーストラリアで収集されたデータを使って学習し、74~94%の精度を達成したと報告されている。水の消費量を7つの主要な水使用カテゴリ(シャワー、トイレ、水道、衣類洗浄機、食器洗浄機、水冷空調器、散水)に分解している。Autoflow では決定木、動的時間伸縮法(DTW: Dynamic Time Warping)、自己組織化マップ(SOM: Self Organize Map)、隠れマルコフモデル(HMM: Hidden Markov Model)を用いている。一般に流量データの識別問題は、1) サンプリング速度が遅いため流量の特徴が流量データの波形に現れにくいこと、2) 水道の元栓に流量センサを設置するため、同時に複数箇所で水をした場合に重畳した波形が観測される。このため、識別の際には重畳した波形を分解して、それぞれの水使用用途を出力する必要があること、3) センサ設置コストが高額なため、多数の世帯データを得ることが困難であるなどの課題がある。また、Autoflow では、14mlの分解能で5秒から10秒毎に積算した高精度流量計を用いて従来にない高い識別精度を上げている。しかし、従来の研究では、被験者宅の元栓に流量センサを設置し、被験者の水使用日記データから元栓の流量センサ波形にラベル付けしたものを Ground Truth として識別器を学習する研究が多い。しかし、記載ミス等による誤差を取り除けないことから、90%以上の識別率を要求するモデルの学習には問題が多いと考えられる。

本研究では、これまでの研究により、1秒毎のバイアス成分を1分程度集積したヒストグラムを特徴量とした独自の深層学習により、重畳使用にも対応し、95%以上の精度で用途別の水使用量を推定する手法の開発に成功した。さらに2mlの分解能での毎秒毎に積算した超高精度流量データを用い、深層学習の一種である Multi-task Convolutional Autoencoder を用いて既に Autoflow よりも高い識別精度を達成することを示唆する実験データが得られた。具体的には、水道の秒間総流量データから用途別流量予測器および用途識別器を作成した。用途別流量予測については予測データと実データのヒストグラム形状が酷似していることから、ほぼ正確に予測できていることを確認した。また、用途識別器については、用途毎に利用の有無を出力する独立した二値識別器を複数構成し、全ての識別器で95%を超える精度で用途別の利用状況を推定できることを確認した。

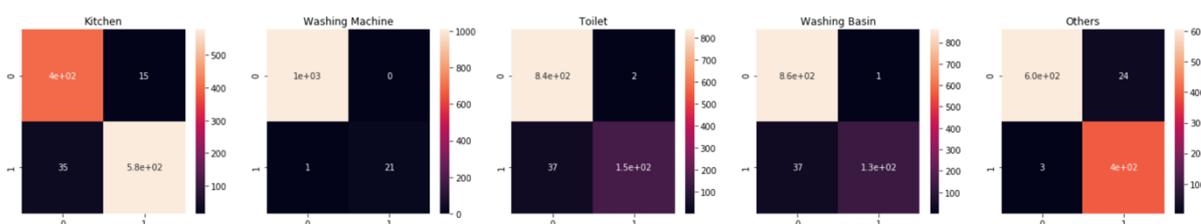


図 1-1: 5 種類の用途別識別器の Confusion Matrix. 図は Kitchen, Washing Machine, Toilet, Washing Basin, Others(その他の用途)の5つの用途について、縦軸は True Label, 横軸は Predicted Label を表している。それぞれ、0 は利用なし, 1 は利用ありを示す。Confusion Matrix から 95% を超える精度で用途別の利用状況を推定できることを確認できる。

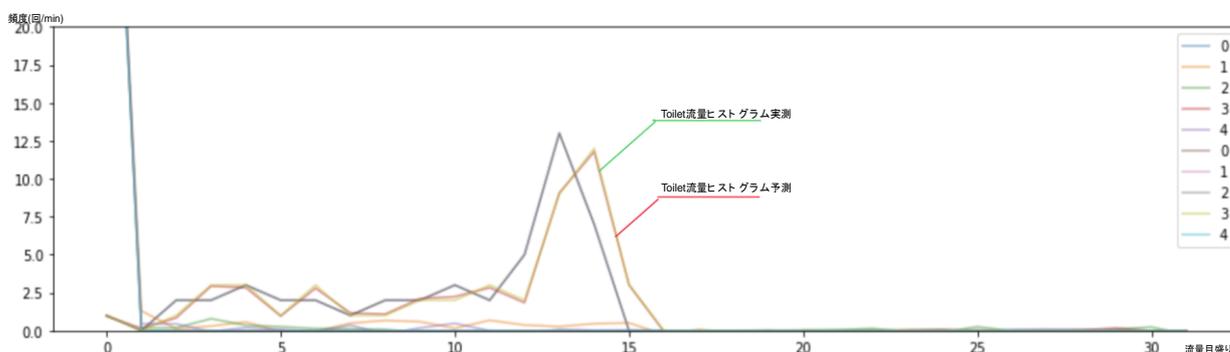


図 1-2: 5 種類の用途別流量予測器の出力結果. 図は 1 分間の総流量の分単位ヒストグラム(秒毎の水量計の目盛りをヒストグラム化したもの)を入力とし、

各用途の分単位ヒストグラムを予測するモデルの出力例を示したものである。図では3番(Toilet)の予測ヒストグラムが実測ヒストグラムに一致していることが分かる。他方, 2番(Washing Machine)のヒストグラムについては、実測値と予測値が異なっていることが判別できる。この時間の2番については使用量が少なく、総量ヒストグラムからの推定が困難であったことが原因として考えられる。