

最先端嗅覚IoTセンサに基づくニオイデータマイニング

Smell Data Mining on the Cutting-edge IoT Olfactory Sensor

江藤 力^{*1} 吉川 元起^{*2} 今村 岳^{*2}

Riki Eto Genki Yoshikawa Gaku Imamura

^{*1}NEC データサイエンス研究所 ^{*2}国立研究開発法人 物質・材料研究機構
NEC Data Science Research Laboratories National Institute of Material Science

As key components at the edge of IoT systems, there is a growing need for sensors that can accurately mimic the five human senses. Among them, olfactory sensors for IoT have not yet been established because of several difficulties in accurately sensing various smells using small devices. Recently, a nanomechanical sensor named "MSS", which can resolve such difficulties, has been developed and is currently under verification. This paper reports a result of a data mining experiment that employs MSS output, which shows the potential use of the MSS.

1. はじめに

IoT システムのエッジ端末として、ヒトの五感を模倣するセンサの重要性が盛んに叫ばれているが、嗅覚については、多種多様なニオイを小型端末で高精度に分析するという難しさゆえ、最もセンサ開発が遅れていた。しかし最近になって、上記の課題を克服する超小型・高感度・化学的多様性を同時に実現する新たなセンサ「膜型表面応力センサ」(Membrane-type Surface stress Sensor, MSS) [Yoshikawa 11, Yoshikawa 12] が開発され、実用化に向けて様々な検証が行われ始めた [吉川 16]。

MSS は、ナノメカニカルセンサと呼ばれるタイプのセンサであり、標的分子の吸着に伴う機械的な変形を電気抵抗の変化として高精度に読み取ることでセンシングを行う。この機械的な変形を生じさせるものとして、化学的選択性を有する「感応膜」をセンサ素子表面に被覆して使用する。この感応膜には有機・無機・バイオ系など様々な材料を利用可能であり、異なる感応膜を用いて得られる多様なセンサ出力を組み合わせることで、混合気体から成る複雑なニオイの分析も可能となる。

しかし、このセンサ出力の多様性により、分析時の特徴量が増えて多重共線性や過学習の問題が生じる恐れがあるため、慎重に特徴量を選択する必要がある。一方、この特徴選択によって、上記問題を回避することに加え、食品管理、環境測定、安全確認、医療やヘルスケアといった各種アプリケーションに最適な感応膜の組み合わせの提示など、より高精度な分析に向けたデバイス設計へのフィードバックも期待できる。特に、近年「異種混合学習技術」として提案された区分疎線形判別分析 [藤巻 16] は、決定木形式のルールによって特徴空間を複数の領域に分割し、各領域におけるローカルな特徴量の寄与を学習できることから、「温度が 20 度以上だとこの感応膜を使えばよい」など、物理的に解釈し易い知見を得ることができる。

本稿では、4つの異なる感応膜をもつ MSS のセンサ出力に基づくデータマイニングによって、ニオイの分析を行った一例を報告する。分析対象として市販のシャンプーを用い、ニオイからシャンプーと水を判別できるかを検証する。まず主成分分析による可視化を行い、センサ出力にシャンプーと水を判別できる情報が含まれていることを確かめる。そして、上述の区分疎線形判別分析を用いて、どの程度の判別性能が達成できるかを確かめ、さらに、疎な線形判別モデルの解釈を試みる。

2. データ諸元と特徴量

本稿で用いるデータは、温度を 20 度で固定し、湿度を 10% および 30% とした環境において、市販のシャンプーと水それぞれのニオイを MSS で計測したものである。今回用いた MSS デバイスには気体をセンサに導入するポンプが 2 つ搭載されており、一方から分析対象の気体 (試料ガス) を、他方から大気 (パージガス) を交互に吸引することで、感応膜への試料ガス分子の吸着/脱離を繰り返し、図 1 に示されるような波形 (1 山 1 谷) が連なるデータを得ることができる。データのサンプリングレートを 10Hz とし、ポンプの駆動を約 30 秒ずつ 20 回交互に切り替える操作を行うことで、感応膜ごとに 585 点の電圧値で構成される波形 20 個を含むニオイデータを得た。その諸元をまとめたものを以下に示す。

- 試料: シャンプー、水
- 環境条件 (温度, 湿度): (20 度, 10%) (20 度, 30%)
- 感応膜数: 4 種類
- 波形個数: 上記 3 項の組合せごとに 20 個
- 1 波形あたりのデータ点数: 585 点

今回はシャンプーと水の判別に加え、疎学習による特徴選択の妥当性を検証するために、分析時の特徴量として、同条件下で同時刻に取得された感応膜ごとの 4 波形を単純にマージした $585 \times 4 = 2340$ 次元のベクトルを用いる。すなわち、分析するデータとして、試料ガスと環境条件の組合せ 4 通りごとに、2340 次元ベクトルが 20 個含まれるものを用いる。

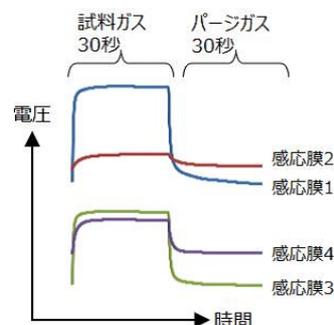


図 1: センサ出力例: シャンプー (温度 20 度, 湿度 10%)

連絡先: 江藤 力, NEC データサイエンス研究所, 神奈川県川崎市中原区下沼部 1753, e-mail: r-eto@bc.jp.nec.com

3. 可視化

まずは主成分分析による可視化を行い、センサ出力にシャンプーと水を判別できる情報が含まれていることを確かめる。第2,3主成分を用いた可視化結果（散布図）を図2に示す。

図2より、条件（試料、温度湿度）ごとにクラスタ化されていることや、左上部から右下部に向かう方向に湿度の影響が出ていることが分かる。左上部は湿度10%、右下部は湿度30%に対応しているが、それぞれにおいてシャンプーの方が第2主成分が大きいという傾向があり、このことから、図3のように湿度で特徴空間を分割し、それぞれの分割された特徴空間に線形判別式を配した2分木構造をもつモデル^{*1}でシャンプーと水の判別が可能であることが想定される。

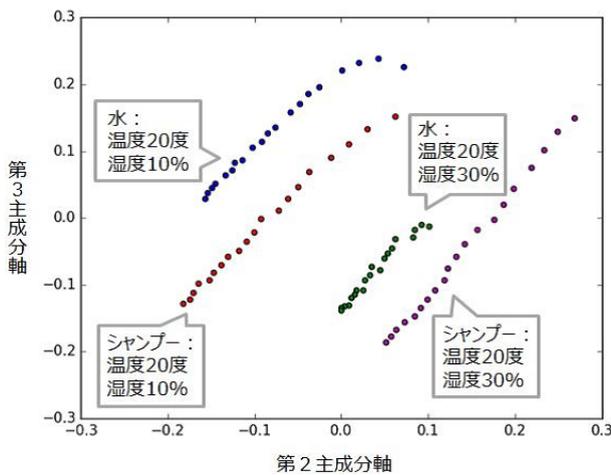


図2: 主成分分析による可視化

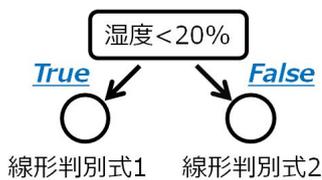


図3: 可視化から想定される2分木構造をもつ判別モデル

4. 判別分析

つぎに区分線形判別分析 [藤巻 16] を用いて、シャンプーと水の判別を実行する。ここでは合計80サンプルを9対1の割合で学習データ（72サンプル）とテストデータ（8サンプル）へとランダムに分割した。なお、それぞれのデータにおけるシャンプーと水のサンプル数は同じとしている。区分線形判別分析ではEMアルゴリズムに似た反復計算を実行している。特徴空間の分割数は初期値として想定より大きい値に設定しておき、反復計算中に割り当てられるサンプル数（期待サンプル数）が小さい特徴空間を除いていくのだが、今回は [藤巻 16] にない分割数32を初期値とし、また、除く条件としての期待サンプル数の閾値は全体の1%と設定している。

可視化結果から、図3のようなモデルが得られることが期待されるが、学習の結果、特徴空間の分割は生じず、図4に示

*1 湿度での分岐条件は10%と30%の中間値（20%）を閾値とした。

された3特徴のみを使用する疎な線形判別式が1つ得られた。ただし、得られたモデルは学習データ、テストデータともに完全に判別できており、想定よりも非常にコンパクトなモデルで判別可能であることが分かった。

この特徴選択結果は、波形での立ち上がり部分と立ち下がり部分が有効であると示しており、これは感応膜に試料ガス分子が吸着/脱離するときにその分子の情報が載るという物理化学的事実と整合している。また、感応膜3と4は使用されておらず、今回の用途ではこれらを除いてもよいといえる。

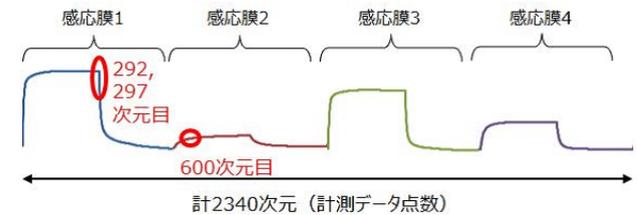


図4: 疎学習による特徴選択結果

5. おわりに

本稿では、最先端の嗅覚IoTセンサであるMSSのセンサ出力を用いた二オイデータマイニングの一例として、シャンプーと水の判別を行った。そして、そのセンサ出力に試料ガス判別に資する情報が含まれること、試料ガスの感応膜への吸着/脱離に対応するデータが特徴選択で選ばれること、さらには感応膜の組合せ最適化への知見も得られることが確認できた。様々な用途が見込まれる嗅覚IoTセンサの社会実装を効率的に進めていくためには、ソフト側のみならず、感応膜の組合せも含めたデバイス側の最適化も同時に進める必要がある。我々は区分線形判別やその回帰版 [Eto 14] という解釈性が高い予測モデルを利用し、物理化学的事実との整合を意識した二オイデータマイニングの研究開発に取り組んでいく所存である。

参考文献

[Yoshikawa 11] Yoshikawa, G., Akiyama, T., Gautsch, S., Vettiger, P., Rohrer, H.: Nanomechanical Membrane-type Surface Stress Sensor, *Nano Lett.* 11, 1044-1048 (2011).

[Yoshikawa 12] Yoshikawa, G., Akiyama, T., Loizeau, F., Shiba, K., Gautsch, S., Nakayama, T., Vettiger, P., de Rooij N. F., Aono, M.: Two Dimensional Array of Piezoresistive Nanomechanical Membrane-type Surface Stress Sensor (MSS) with Improved Sensitivity, *Sensors*, 12, 15873-15887 (2012).

[吉川 16] 吉川 元起, 今村 岳, 柴 弘太: 最先端センサシステムとビッグデータ解析技術で実現する二オイの標準化と呼気診断, 第30回人工知能学会全国大会, 2016.

[藤巻 16] 藤巻 遼平, 山口 勇太郎, 江藤 力: 因子化漸近ベイズ推論による区分線形判別, *人工知能学会論文誌* 31 巻6号 AI30-I, 2016.

[Eto 14] Eto, R., Fujimaki, R., Morinaga, S., and Tamano, H.: Fully-Automatic Bayesian Piecewise Sparse Linear Models, in *AISTATS*, pp. 238-246 (2014)