

機械学習を用いた匂いの印象予測・分類

人間の五感の一つである嗅覚の感覚は本能的であり、専門家以外の評価者による匂いの評価は非常に直感的で簡潔な表現となります。また、工業製品の分野において望み通りの匂いをプロダクトが有するように調香師と呼ばれる専門家によって日々検討されています。

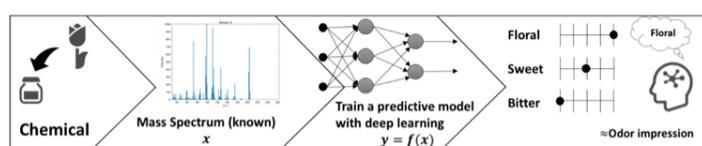
しかし、このプロセスは未だに人の手を介しており、自動化による匂いの印象の予測・測定への恩恵は非常に大きいものと考えられています。

上記の背景から、当研究室では機械学習を用いて匂い物質の特性から物質の匂いの印象を予測するモデルの検討を行っています。

マススペクトルを用いた匂い物質の印象予測

本研究ではある香気物質から対応する官能評価データを予測するモデルの研究を行っています。入力データにはマススペクトルを使用し、テストデータに官能評価データを使用しています。マススペクトルとは、質量分析計を用いて質量電荷数比に応じて分離・検出されたイオンをもとに作成した棒グラフのことです。また、本研究での官能評価データとは、人間の感覚を用いて物質の特性(今回は匂いの印象)を定量的に表したデータを意味します。概略図は図1に示します。訓練時に入力データにはマススペクトルを、テストデータには官能評価データを用いて学習を行うことで、任意のマススペクトルに対してその官能評価データを生成することを目的としています。

(i) Modeling



(ii) Prediction

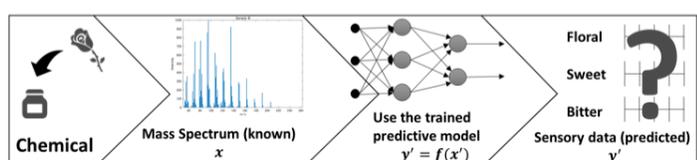


図1. 匂い印象予測モデルの概略図

より詳細な機械学習の手法を図2に示します。本研究では2つのオートエンコーダを用いて両データの特徴抽出及び、次元圧縮を行い、それらを多層パーセプトロンを用いて写像を行うことで、9層の予測モデルを構築しています。

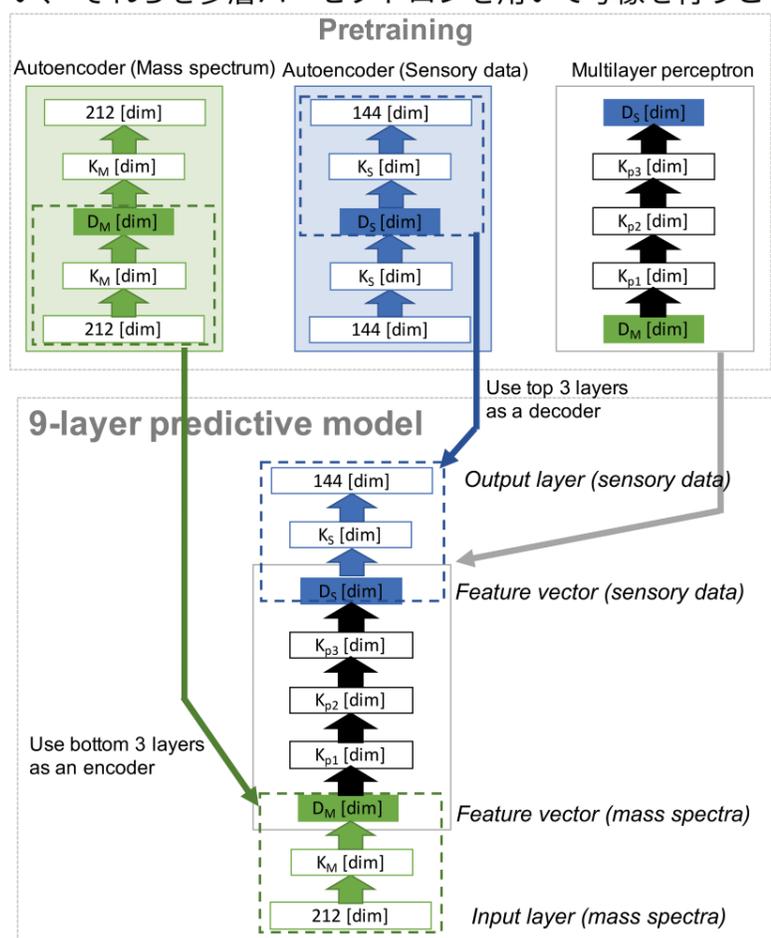


図2. 9層の予測モデル

以上より構築したモデルの予測精度は相関係数が $R=0.76$ であり、従来法は $R=0.61$ から制度の向上を確認しています。

匂い記述子の分類

本研究ではある分子のマススペクトルが与えられた時、その分子の官能評価データ内の記述子への当てはまり(所属)の有無を分類予測するニューラルネットワークモデルの構築を行っています。官能評価データの中には本来高い相関を持つべき記述子間の相関が失われていることがあります。そのため、ニューラルネットワークの学習を困難にしている可能性があると考えられています。そこで、この問題に対処するために本研究では類似した記述子を内包するより大きな粒度のクラスタを用いて匂いの分類を行うモデルの



構築を行いました。実験において、Word2vecを用いて単語(記述子)間の類似性を用いて記述子のクラスタリングを行いました(図3)。

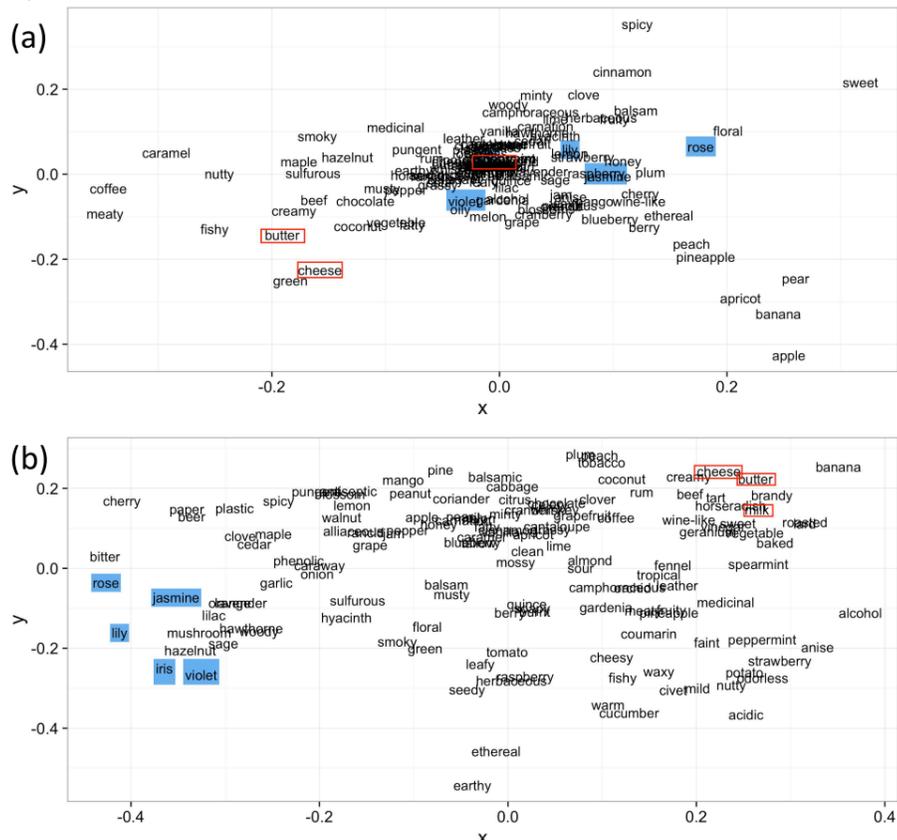


図3. カタログデータにおける記述子間の相関係数より作成したMDS

図3の結果から、図4のように類似した単語同士をクラスターに分類し、最終的に深層学習を用いて匂いの印象予測モデルの構築へと繋げた研究を行っています。

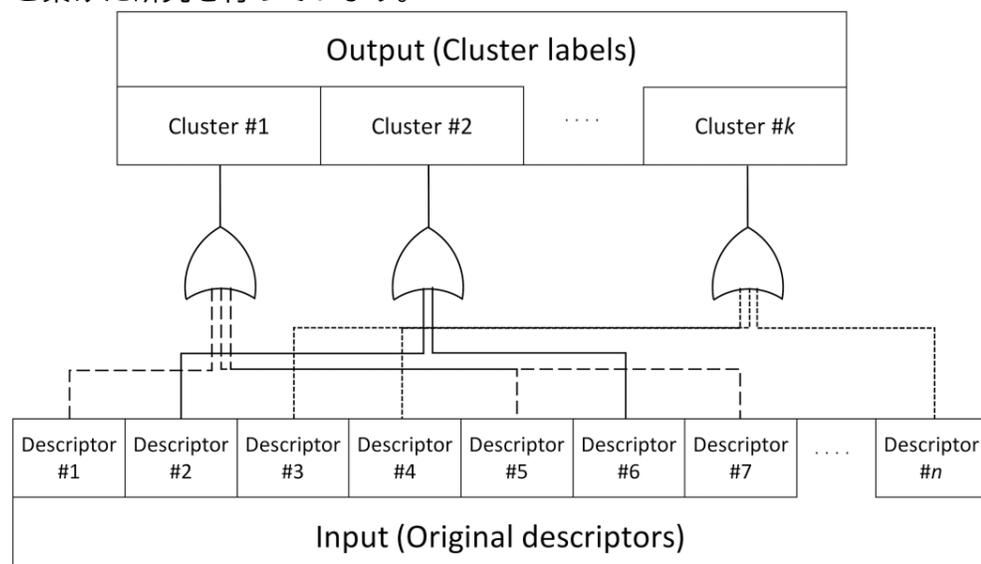


図4. Word2Vecモデルから得られた単語ベクトル間のコサイン距離より作成したMDS

本研究では、匂いクラスターの数を20に設定した場合、モデルの予測精度は、真の陽性で53%、真の陰性で85%であり、粒度が大きい未知の化学物質の臭気特性の予測が実現しています。

Reference

1. Nozaki Y, Nakamoto T (2016) Odor Impression Prediction from Mass Spectra. PLoS ONE 11(6): e0157030. <https://doi.org/10.1371/journal.pone.0157030>
2. Nozaki Y, Nakamoto T (2018) Correction: Predictive modeling for odor character of a chemical using machine learning combined with natural language processing. PLOS ONE 13(12): e0208962. <https://doi.org/10.1371/journal.pone.0208962>