ノート

機械学習を用いたFT-IRスペクトルデータ解析

天城　裕子*

FT-IR Spectrum Analysis Using Machine Learning

AMAKI Yuko*

1. 緒　　言

当所の依頼試験などにおいては，未知サンプルの赤外分光分析（以下FT-IR）スペクトルデータは混合物のデータとして取得される場合が多い。FT-IR装置に付属のスペクトルデータ検索ソフトでは単一成分でないと該当成分が一致しないことが多く，比較品の入手が不可の場合，測定者の知見に基づく逆引き検索や，手作業で差分をとり差分データで改めて解析を行うなどの方法で対応する必要がある。

また，酸化・劣化やモルフォロジの変化でもスペクトルデータの形状は変化するが，そのようなデータはデータベースに登録がない場合がほとんどで，解析には測定者の経験を要する。

そこで，これらの作業の効率化のため機械学習を用いた解析方法について検討した。

2. 各手法による検討

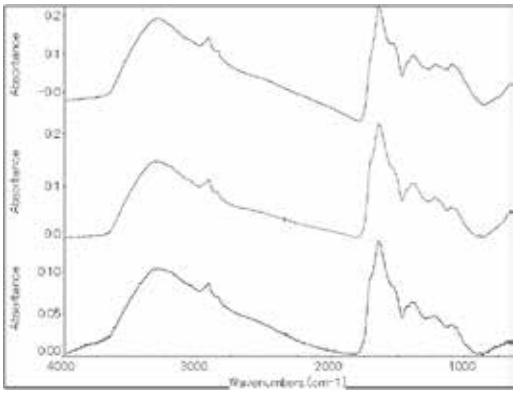
2.1　教師なし学習：独立成分分析法（ICA）

同一サンプルで複数データを取得し，Pythonのscikit-learnライブラリのICA（Independent Component Analysis）で分離を行った結果を図1に示す。サンプルは鍋の付着物で，食材由来成分の混合データであることが予想される。図1(a)が任意の3か所を採取し測定した分離前データ，図1(b)が分離後である。分離後スペクトルは上から焦げなどでみられるカルボン酸塩，糖類，タンパク質由来のスペクトルに近い形状を示した。いずれも食材成分でよくみられる成分である。

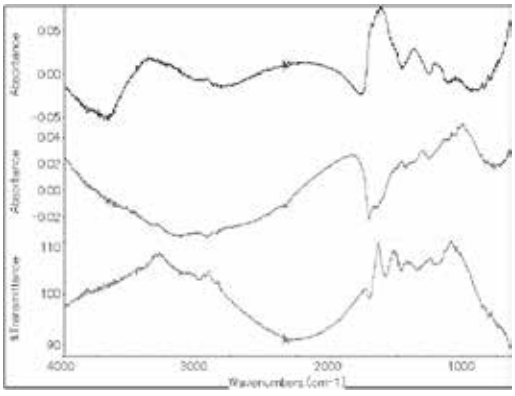
* 下越技術支援センター

2.2　教師あり学習：スパースモデリング

スパースモデリングは，入力データを（なるべく少ない）辞書用データの組み合わせで表現する手法である。今回は製品や標準品などのラベル付けをした実測定サンプル約700点を辞書用データとし，scikit-learnライブラリのSparse Coderを用いて分離を試みた。SparseCoderの計算用のアルゴリズム[1)]で，デフォルトのOMPは小さくても鋭いピークをよく拾う傾向，



(a) 分離前



(b) 3成分分離後

図1　ICAテスト結果

Lasso-LARS（Least Angle Regression）はブロードなピークも含めてスペクトルの全体的な形状に合わせこむ傾向がみられた。FT-IR のスペクトルデータの処理に関しては Lasso-LARS のほうが適していると思われる。図 2 はシリコーンと炭酸カルシウム，カオリン，テフロンを混合したテストデータで評価を行った結果である。混合元に近い物質を返すことが確認できた。

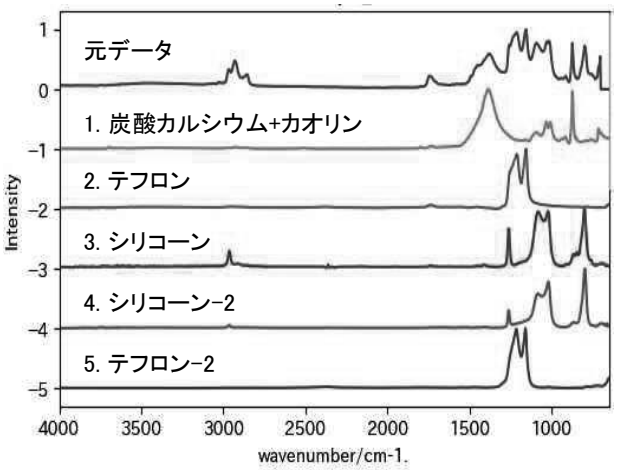


図 2　スパースモデリングのテスト結果

2.3　教師あり学習：ディープラーニング

前項でのデータセットからランダム 4 種を強度比 1：0.5：0.25：0.1 で混合した 20 万データを新しいデータセットとし，Google Colaboratory[2] 上で Keras[3] の畳み込みニューラルネットワークでテストを行った。混合データの 85%（17 万）を学習用データ，残りを評価用とし，混合元の 4 種データを判定した。モデルは畳み込み層とドロップアウト層を適宜組み合わせて作成した。

なお，FT-IR のスペクトルデータでは波数位置が重要なため，画像抽出などでよく用いられる Max プーリング層は設けなかった。

表 1 に 5 回学習後の損失関数と精度を示す。表内で label1 から label4 の順に混合強度比の高いデータラベルである，混合比の高い第 1，第 2 成分までは比較的良好に学習できていることが確認できる。一方で第 3 成分以降は，うまく評価できていないことがわかる。原因はデータによるもの，学習不足，もしくはモデルが適当でないことなどが考えられる。今後の検討課題である。

表 1　損失関数と精度

label	loss	accuracy
label1	0.206	0.93
label2	0.913	0.75
label3	2.669	0.45
label4	5.447	0.17
total loss	9.235	-

3. 結　　言

(1) ICA は，辞書データに一致するものがないと想定される場合に有効である。ただし，分離後データにラベルがないため，その後の解析にやや慣れが必要である。

(2) スパースコーディングは分離目的以外にも類似データとの照合が可能であり，一般的な検索ソフトの使用感に近い。辞書用データの拡充が課題である。

(3) ディープラーニングについては，主成分は学習可能であることを確認した。主成分以外の成分を認識するためには，データセットや学習モデルの再検証が必要である。

参考文献

1) https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.SparseCoder.html

2) https://colab.research.google.com/notebooks/intro.ipynb?hl=ja#scrollTo=5fCEDCU_qrC0

3) https://keras.io/ja/