

機械学習を用いたイオン水和状態のテラヘルツ分光計測データの解析

Analysis of Terahertz Spectroscopic Measurement Data of Ion Hydration State Using Machine Learning

東京大学大学院工学系研究科¹ ○(M1)川上 大貴¹, 田畑 仁¹

The Univ. of Tokyo Graduate School of Engineering¹, ○Daiki Kawakami¹, Hitoshi Tabata¹

E-mail: kawakami@bioxide.t.u-tokyo.ac.jp

1. 概要

テラヘルツ波は、0.1~10 THz の周波数を持つ電磁波で、この周波数領域は、抗体抗原反応の分子間結合や水分子と溶質との間の水和による分子間結合などの分子間結合エネルギーに対応している。そのため近年、バイオ分野において、上記の特徴を持つテラヘルツ波を用いた生体関連物質の分光センシング応用が研究されている^{[1][2]}。しかし一般に、分子間結合力の分散は大きいため、テラヘルツ帯に現れる分子間結合エネルギーに由来するピークはブロードである。また、テラヘルツ波のエネルギー(0.4~40 meV)は低く、熱エネルギー(~25 meV)に埋もれやすい。そのため、赤外分光や光電子分光に比べて、テラヘルツ分光ではピークを抽出することが難しい。そこで本報では、機械学習を用いて、テラヘルツ分光データのスペクトル形状から学習モデルを作製し、分光データの解析を試みた。本報では、生体にとって重要な役割を持つイオン性分子(K^+ , Na^+ , Ca^{2+} など)の水和状態を測定したテラヘルツ分光データを機械学習させ、その学習モデルを評価した結果について述べる。

2. 実験方法とその結果

テラヘルツ全反射分光法により、0 M (=mol/L), 0.1 M, 0.3 M, 0.5 M, 1.0 M, 1.5 M, 2.0 M, 3.0 M の濃度の塩化カリウム水溶液の分光測定を行った。各濃度での測定は 10 回ずつ行い、合計で 80 個の反射率の分光スペクトルデータを得た。その後、80 個のデータのうち 64 個を学習用のデータとして用い、機械学習を行った。学習用データは最小二乗法、Ridge 法、Lasso 法^[3]によりそれぞれ学習させ、分光データから塩化カリウム濃度を予測するモデルを作成した。そして、16 個のテスト用データを用いて各学習モデルの評価を行った。Table1 に、学習用データおよびテスト用データから塩化カリウム濃度を予測した結果の決定係数(真値と予測値の一致具合)を示す。これを見ると、今回、Ridge 法による学習モデルが最も精度よくテストデータの塩化カリウム濃度を予測できている。Ridge 法による学習モデルを用いたテストデータの予測結果を Fig. 1 に示す。Fig. 1 より、 $R^2 = 0.880$ の精度でテストデータの塩化カリウム濃度の予測ができていることがわかる。

謝辞

本研究の特に機械学習において、阪大産研の小口教授、兼平氏、神田氏にご指導いただいた。ここに感謝を記す。また、本研究は、独立行政法人日本学術振興会の「研究拠点形成事業 (A. 先端拠点形成型)」の助成を得て行われた。

Reference

- [1] K. Kawase, *et al.*, *Opt. Express* 11.20 (2003), pp. 2549-2554.
 [2] S. Kawabe, *et al.*, *IRMMW-THz*, Mainz, Germany, 2013.
 [3] W. Richert, *et al.*, “実践 機械学習システム,” 株式会社オライリー・ジャパン, 2014.

Table1 The coefficient of determination at each learning method.

Learning Method	R^2	R^2
	(@Training data)	(@Test data)
Least-square Method	1.000	0.320
Ridge Method	0.922	0.880
Lasso Method	0.967	0.806

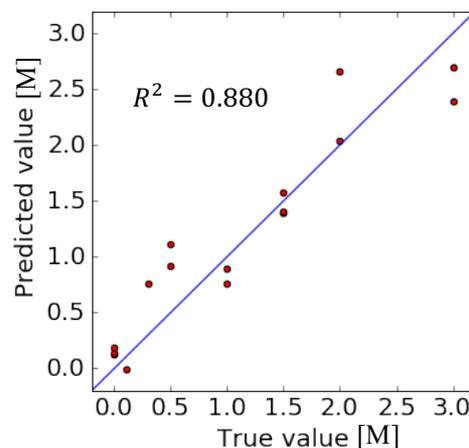


Fig. 1 Prediction result of test data using learning model of Ridge method.