

大気化学的知見に基づく学習データの最適化によるオキシダント機械学習予測精度の向上

Improving performance of oxidant prediction using machine learning by optimizing input data based on atmospheric science knowledge

*佐藤 知紘¹、趙 培江¹、是津 耕司¹

*Tomohiro Sato¹, Peijiang Zhao¹, Koji Zettsu¹

1. 情報通信研究機構

1. National Institute of Information and Communications Technology

【背景】近年、大気汚染による早期死亡者数は世界で約370万人であり、これは交通事故による死亡者数の約3倍に相当する（WHO報告書, 2014）。日本では、NO₂やPM_{2.5}等のほとんどの大気汚染物質が減少傾向を示す中、オキシダント（Ox）は環境基準達成率が0.1%と極めて低く、依然として横ばいまたは増加傾向を示している（環境省報告書, 2018）。Oxは光化学スモッグの主成分であり、高濃度になると喘息や肺炎などの発症リスクを高める他、農作物にも被害を与える（Zheng et al., 2018等）。Oxの生成メカニズムはNO₂や揮発性有機化合物（VOC）等の前駆体の濃度や太陽光、風量などの様々な要因が非線形に結びついているため、その物理化学過程の完全なモデル化は難しく、また計算コストが莫大になるという課題がある。そこで、従来の物理化学モデルに代わる予測手法として、機械学習を用いた研究が進められている。Keller et al., (2017)はGEOS-Chemモデルの計算結果を学習データとしてOxの主成分であるオゾン（O₃）の濃度を予測するRandom Forestモデルを構築し、計算コストを100分の1に小さくした上で10日後までの予測を誤差10%以内で実現できることを示した。他にも、Multiple Linear RegressionやArtificial Neural Network（Pires and Martins, 2011）、Gene Expression Programming（Samadianfard et al., 2013）等の方法によるオゾン濃度予測研究が報告されている。

【目的】これまでの機械学習による大気汚染短期予測研究は、モデルのアルゴリズムについての発展は行われているものの、学習データそのものについては、あまり注目されていなかった。一方で、Data Augmentation等の学習データの改良により予測精度が向上できることが報告されており（Guenec et al., 2016）、本研究では、大気化学的知見に基づくOx学習データを最適化することにより、機械学習によるオキシダント短期予測精度の向上を試みる。

【データセット・手法】 Ox（特に主成分であるO₃）は光化学反応によって生成されるため、日の出とともに増加（NO₂ + hn → NO + O, O + O₂ + M → O₃ + M等）、日中に最大となり、夜は他の物質との反応や地表面との接触により徐々に減少する。この規則的な日変化は、天気や季節による太陽光量や人間活動による前駆体発生量に依存して絶対量は変化するものの、その傾向は一年を通してほとんど変わらない。そこで本研究では、学習データにOx測定値（V_{obs}）をそのまま用いてOx短期予測を行った場合（①）と、1日の中の時間ごとの日変化（V_{diurnal}）からの差分（DeltaV = V_{obs} - V_{diurnal}）を入力した場合（②）を比較した。学習データセットには、福岡市内16地点の1時間毎の測定値（国立環境研究所 環境数値データベース 時間値データファイル）を使用し、Oxの他にNO₂やSPM、PM_{2.5}、SO₂、CO等の主要大気汚染物質、温度や湿度、風向、風速の気象条件を対象とした。期間は2014～2016年度の3年分とし、基準となるOxのV_{diurnal}は、年度毎の全データについての時間ごとの平均値とした。機械学習モデルには、Convolutional Neural Network（CNN）とLong Short-Term Memory（LSTM）を合わせたConvolutional Recurrent Neural Network（CRNN）を使用した（Zhao and Zettsu, 2018）。このモデルでは、測定点の緯度・経度座標を基に各測定値を2次元データ化してCNNにインプットし、空間特徴量を抽出、その後LSTMにより時間発展後の予測値を出力する。入力データは過去24時間分の全測定点毎のデータ、出力データは6時間後の全測定地点を平均したOx濃度とした。

【結果・考察】①と②の場合におけるCRNNによる予測値と実測値を比較し、平方根平均二乗誤差（Root Mean Squared Error: RMSE）と正確率（Accuracy）により予測精度を評価した。各年度のRMSEとAccuracyは、①10.3 ppb/75.4%、②6.7ppb/84.7%（2014年度）、①11.0 ppb/72.7%、②8.3ppb/79.6%（2015年度）、①9.4 ppb/76.7%、②8.6ppb/79.4%（2016年度）であり、入力データに差分DeltaVを用いることで正確率が最大約10%向上した。また、予測値と実測値を1日の中の時間ごとに比較すると、①では夜間は過大評価、昼間は過小評価する日内バイアスが確認された。一方、②ではこのバイアスが大きく解消され、時間による偏りが少ない予測を実現した。なお、このバイアスの解消は、全ての年度について確認された。

【まとめ・提言】本研究では、日本国内で近年横ばいまたは増加傾向を示す有害物質であるオキシダント（Ox）の機械学習による高精度短期予測を実現するために、従来研究に多く見られるアルゴリズムの改良ではなく、大気化学的知見に基づくOx学習データの改良による予測精度向上を試みた。その結果、入力データとしてOx日変化の特徴を除いた差分データを用いることで、予測正確率を最大10%向上させることに成功した。本研究成果は、機械学習を用いた大気環境予測において、大気化学の知見に基づく学習データの拡張や調整が重要であることを示唆している。

キーワード：大気汚染、オキシダント、機械学習、最適化、短期予測

Keywords: Air pollution, Oxidant, Machine learning, Optimization, Short-term prediction