

機械学習による広域大気汚染予報システムの開発

岡山大学異分野基礎科学研究所

松本 正和

序章

本研究課題では、公的にアクセス可能な過去の大気質と気象の大規模データセットを学習し、光化学オキシダント(Ox)濃度を予測するソフトウェア「あんでるさん」を開発した。

光化学オキシダントは、工場や自動車から排出される窒素酸化物(NOx)や揮発性有機化合物(VOC)が太陽光を受けて反応することで生成される二次汚染物質である。この大気汚染物質は、人間の健康に深刻な影響を及ぼす。特に、呼吸器系への影響が顕著であり、ぜんそくの悪化、気管支炎、肺機能の低下などを引き起こす可能性がある。さらに、目や喉への刺激も強く、広範囲で健康被害が発生する恐れがある。小児や高齢者、呼吸器疾患を持つ人々は特に脆弱であり、高濃度に達した場合には、外出自粛などの対策が呼びかけられることもある。市民の健康と生活の質を守る上で、高濃度予測は不可欠な情報となる。

光化学オキシダントは、人間の健康だけでなく、植物や農作物、森林といった環境生態系にも大きなダメージを与える。植物の光合成能力を低下させ、成長を阻害し、葉の壊死や収穫量の減少を引き起こすことがある。長期的な影響は、生態系のバランスを崩し、生物多様性の損失にも繋がりがかねない。持続可能な社会を築くためには、自然環境への負荷を軽減し、生態系の健全性を保つための対策が喫緊の課題であり、そのためにも光化学オキシダントの正確な予測が求められる。

光化学オキシダントによる被害は、健康や生態系に留まらず、社会経済活動にも間接的な影響を及ぼす。例えば、農作物の収量減少は食料供給に影響を与え、経済的損失を生じさせる。また、高濃度予測による外出自粛や屋外活動の制限は、観光業やイベント産業などにも影響を及ぼし、経済活動の停滞を招く可能性がある。さらに、医療費の増加や労働生産性の低下といった形で、社会全体にコストを強いることにもなる。光化学オキシダント注意報が発令された時点で、工場等の稼働に一定の制限が加えられることがあり、事前の予測は円滑な経済活動を続ける上で有用である。

現在も神奈川県などで光化学オキシダントの予報は行われているが、その精度や広域的な対応には課題が残されている。大気汚染物質の生成と拡散は、気象条件(日射量、気温、風向・風速など)や複雑な化学反応、地形の影響を強く受けるため、非常に予測が難しい現象である。既存の物理化学モデルだけでは捉えきれない複雑な非線形性を考慮し、より高精度な予測を行うことが求められている。

このような背景から、本研究では機械学習技術を導入することで、広域的な光化学オキシダントの発生メカニズムをより詳細に解明し、高精度な予報システムを開発することを目指す。これにより、健康被害や環境負荷の低減、そして持続可能な社会の実現に貢献できると考える。

えだまめ

えだまめは、あんでるさんの前身で、1つの大気観測局の過去データを学習し、その局の8時間先までの光化学オキシダント濃度を予測するシステムである。えだまめは、主に予測アルゴリズムの改良と検証を目的として作られた。我々は2022年10月から2023年3月ごろまでこれを開発していた。

えだまめは、環境省の環境展望台(<https://tenbou.nies.go.jp/download/>)で提供される全国の大気測定局の測定値アーカイブのみを利用するように設計されたので、現在の実測値を読みこんで未来予測をする機能

はない。その代わり、過去のどの時刻を起点にしても予測を試すことができ、その時点での実測値と比較が可能なので、予測アルゴリズムの性能評価に適している。えだまめには、開発者向けのウェブユーザーインターフェースがあり、数十のアルゴリズムを同時比較できるようになっている。

あんでるさんへの発展

えだまめを開発するうちに、これを多少改良しても、我々が期待しているような予報システムに到達できず、抜本的な改良が必要であることが明らかになってきた。

えだまめは、個々の測定局の場所での予測しかできない。大気測定局は都市部に多い一方で、広範囲にわたってまったく測定局の無い領域がある。また、測定項目が局ごとに違う。予測に必要な測定項目を測定していない測定局では、データが完備していないため、予測ができない。さらに、それぞれの局で、ときどき欠測がある。欠測の原因は不明だが、おそらく装置不具合や装置の整備交換などによると思われる。欠測が起こるのは、それ以外にも、そもそも自排局などのように、一部の測定項目を測定する能力がない場合や、測定局の移転・統合など、さまざまな要因が関わる。欠測があるたびに、近隣の測定局の測定値に代替させることは不可能ではないが、自動化に適さず、長期的にコストをかけずに運用する妨げとなる。

このような問題を解決するために、あんでるさんの開発に着手した。あんでるさんでは、地域を地理院タイルと呼ばれる長方形のタイルに切りわけ、それぞれの格子点での測定値を、Delaunay 三角形分割¹⁾によって割当てられる近隣3測定局の測定値を内挿することで推定する。欠測がある局を排除して動的に三角形分割を行うことで、欠測の影響を非常に受けにくい。

また、隣接格子点との相関もとり入れ、より正確な予測を行う、いわゆる Convolutional neural network (CNN)に発展することも可能である。(なお、現在開発を進めているバージョンでは、隣接格子点の寄与はとりいれていない。その理由については後述する)

システム構成

あんでるさんのシステムはいくつかの独立な要素で構成されている。

Airpollutionwatch²⁾

現在の大気測定値の速報値が、そらまめくんに反映されるには 61 分かかる。できるだけ速い予測のために、各県の提供する大気測定値の速報値をスクレイピングにより直接入手するための Python モジュールを開発した。各県は異なるソフトウェアを用い、異なるフォーマットで情報提供しているが、このモジュールを使うことで、一貫した方法でデータを入手できる。我々はこの汎用ライブラリを andersan とは別個にオープンソースソフトウェアとして提供する。

Andersan³⁾

アーカイブデータの管理、気象データの入手、データの単位の共通化、測定局ごとのデータを格子点(タイル)ごとのデータに変換する処理、データキャッシュなどを担う Python モジュールである。

Andersan-train

ある時刻を原点とし、それより過去の大気測定値の履歴と、原点よりも未来の気象データをインプットとし、光化学オキシダントを教師データとして、Convolutional recurrent neural network (CRNN)を用いて学習するソフトウェア基盤。この部分は事前の訓練にのみ用いる。非公開。

Andersan-core⁴⁾

あんでるさんの中核を担う Python モジュールで、andersan-train で訓練した予測モデルを用い、予測を行う。andersan-api から呼びだされる。

Andersan-API⁵⁾

andersan-train で学習したモデルを用い、現在からすこし前の大気測定値と、現在からすこし未来の気象データを airpollutionwatch から読みこんで、現在の予測を行い提供する Application programming interface (API)。Python FastAPI で書かれている。

指定時刻以降の、指定領域内での光化学オキシダント予測値、指定時刻の、指定領域内での大気環境情報の速報値、緯度経度から住所への変換、予測値から注意報発報確率への変換テーブルなどを提供する。

Andersan-UI⁶⁾

API が提供する情報を、一般ユーザーが利用しやすい形で表示する Web ユーザーインターフェース (UI)。Svelte で書かれている。表示されている確率予報の原理については後述する。

仮設サイトは <https://andersan.riis.okayama-u.ac.jp:8080> で試用できる。(現在の大気質状況ではなく、過去の、注意報レベルの濃度が観測された日の朝の予報をサンプルとして表示している)

Andersan-Pro⁷⁾

API が提供する情報を、地方公共団体の予報担当者が利用しやすい形で表示する Web UI。Svelte で書かれている。

仮設サイトは <http://andersan.riis.okayama-u.ac.jp:8081> で試用できる。時刻の始点には現在もしくは過去に光化学オキシダント濃度が高くなり注意報が発令された日を選べる。県内のすべてのタイル展での、24 時間先までの光化学オキシダントの予測値が一覧できる。

なお、UI の開発では、AI による自動コーディングが非常に有用であった。UI 開発では、思ったような見た目を実現するための手順がこみいっている一方、AI に出すべき指示(プロンプト)ははっきりしているので、人力でマニュアルを引き引き開発するよりも AI にまかせたほうが圧倒的にはやく完成する。一方、API の設計や、その背後で動くアルゴリズムの設計は、まだ AI に任せてよいのかはわからない。

アルゴリズム

えだまめ開発で得た経験をもとに、あんでるさんでも非常に多数のアルゴリズムを検討したが、現在のモデルでは Gated recurrent unit(GRU)タイプの Recurrent neural network (RNN)⁸⁾を用いて 24 時間先までの予測を行う。検討したアルゴリズム(RNN の種類、階層、パラメータ数などのバリエーション)には複数あるが、ここでは v0a と呼ばれているアルゴリズムについて述べる。

あんでるさんでは、上述の通り、地域を格子に分け、予測は格子点ごとに行う。各格子点での測定値は、その格子点に最も近い 3 測定点での測定値の内挿値を用いる。

RNN への入力と出力

RNN に与えるデータは、3 種類ある。

- 23 時間前～現在の 1 時間ごとの 6 測定値(計 144 個)
- 1 時間後～24 時間後の 1 時間ごとの 5 予報値(計 120 個)
- 23 時間前～現在の 1 時間ごとの予報気象コード(24 個)

これらはいずれも時系列データではあるが、入手経路が違い、データの種類によって時間長さも違ってくる可能性もあるので、まずそれぞれ独立な GRU に読みこませる。その上で、3 入力を統合し、さらに数段の密結合ニューラルネットワーク(ドロップアウトあり)を通して出力に至る。

出力されるのは、24 時間先までのオキシダント濃度(24 個)である。これを実測値と照合することで RNN の訓練を行う。

隣接タイル点の情報も含めた、CNN の構成にすれば、近隣からの汚染原因物質の移送の効果もとおりいれ、より正確な予測ができると考えられる。しかしながら、実際には隣接格子点をとりに入れることによる精度の向上はごく僅かであったため、現在の最終的なアルゴリズムでは隣接格子点の情報をとりにいていない。

隣接格子点をとりにいても予測精度が向上しないのは、おそらく光化学オキシダントの生成が主に無風状態で亢進するという、光化学反応の特性によるものと思われる。無風状態であれば、隣接格子点からの移送を考慮する必要はない。本システムは、光化学オキシダントが高濃度となる可能性を的確に捉えることを目的としており、風が強いなどの理由で高濃度とならない時期に関して、正確に予測できることを目指していない。学習時にも、高濃度時のデータの重みを大きくするよう傾斜学習を行っている。今後、光化学オキシダント以外の大気質予測に転用する場合には、移送の効果が必要になる場合もあるだろう。

異なるタイルでの予測に、同じモデルを使用する(タイルごとに別々に学習しない)ことで、教師データの量を増やし、本番運用時の計算量を減らす。

学習データのソース

過去の大気質測定値は環境展望台(<https://tenbou.nies.go.jp/download/>)から入手した。また、過去の気象情報は Open-Meteo (<https://open-meteo.com>)から入手した。学習に用いたのは、2009 年 4 月 1 日～2021 年 3 月 31 日年の 12 年分のデータである。これらをさらに神奈川県内 77 地点のタイルに展開したものを SSD 上に配置して、学習に利用した。

訓練

v0a モデルの場合、神奈川県内の 77 タイルの 12 年分のデータを用いた訓練には、NVIDIA RTX4000 Ada GPU を用いて丸一日程度かかる。日々新しい測定値が増えるが、統計的な変化はわずかなので、再訓練は年に一回程度行えば十分である。一方、訓練結果を用いた予測には GPU を要しない。

評価

v0a アルゴリズムの予測精度を評価するためのいくつかのデータを示す。地方自治による光化学オキシダント予報の発報の形式として、同日予報(その日のうちに OX が既定値よりも大きくなるかどうかを同日朝、始業時前後に予報する)と、前日予報(翌日に OX が既定値よりも大きくなるかどうかを前日夕方)に予報する)の 2 通りを想定する。

同日予報

光化学オキシダントが最大値をとるのは、午後 2 時～4 時であることが多いことから、午前 8 時頃に、8 時間先までの正確な予測ができることが望ましい。そこで、図 4 では、神奈川県内の過去 12 年間の大気測定値と気象情報を用いて、8 時間先の光化学オキシダント濃度をどの程度予測できるかどうかを検証した。図 4 から、予測は実測に比べてやや低めの値になることがわかる。

前日予報

前日予報を行っている自治体は神奈川県のみであるため、神奈川県の評価手法に沿った検証を行った。神奈川県の前日予報では、翌日に県内のどこかの地点で光化学オキシダントの濃度が基準値(120 ppb)を越えるかどうかを午後 5 時時点で予報する。そこで、図 5 では、神奈川県内のすべてのタイル点の、未来 24 時間以内での最大値を、予測値と実測値で比較した。

等高線が対数尺度になっていることを考えると、かなり正確に翌日の光化学オキシダントの最大値を予測できていると言えよう。

運用体制

システムはフロントエンドとバックエンドに大分される。フロントエンドの andersan-ui と andersan-pro はインターネット上のウェブサーバ andersan.riis.okayama-u.ac.jp (岡山大学が管理)を利用し、バックエンドの andersan-api は岡山大学内に設置された高性能 GPU 計算サーバが受けもつ。後者から前者に向けて SSH port forwarding で接続することで、逆に前者から後者に対してファイアウォールを越えてアクセスし、API の提供する情報が UI に反映される。フロントエンドは JavaScript 系のフレームワークで構成され、軽量でサーバを選ばないが、後者は予測のための大量のデータと訓練環境が必要であるため、設置場所を分けてある。

運用には人力が介入する余地はないが、県ごとの最新大気データの提供方法や URL が変わったりすると、その都度迅速にソフトウェアの修正が必要である。なお、データの取得に失敗すると、その日の予測には影響が生じるが、最新データを随時学習に利用しないので、訓練済みニューラルネットワークには影響しない。訓練に関しては、アーカイブデータを用いて年に一度程度行って新しいデータを学ばせることで、環境の経時変化による精度劣化を避けられる。

リアルタイム運用

API からデータを取得する際に、現在時刻を指定すると、リアルタイムでの予測が入手できる。各都道府県の大気環境速報サイトにデータが残っている期間内であれば、現在よりもさかのぼった時刻を起点とした予測もできる。また、2022 年以前に関しては、環境展望台と Open-Meteo のアーカイブデータを用いた予測を行う。

訓練には、未来値(予報値)も含めすべてアーカイブされた実測値を利用できる。一方、本番運用時には、気象予報サービスの予報値(実測ではない)を利用するので、データの品質が異なる可能性があり、それは精度の低下に結びつくことに留意すべきである。また、リアルタイムデータで得られない情報を、訓練に用いることはできないので、モデル設計時に、本番運用時のデータ取得フローまで考慮しておく必要がある。言いかえるなら、本番運用でいくつかの予報値が満足のいく精度で入手できないとなると、ニューラルネットワークの訓練までさかのぼってやりなおすことになる。

確率予報の原理

図 4、5 を見るとわかるように、相関プロットは対角線よりもやや下側に傾いており、光化学オキシダントが高濃度の時ほど、やや低めの予測になっていることが読みとれる。それでも以前の重回帰モデル等による予測に比べるとかなり精度は上がっているのだが、予測値をそのまま予報の判断基準にすると、常に低めの数字を示すことになり、適切でない。

そこで、予測モデルの予測値をそのまま予報に使うのではなく、注意報の発報確率という形で注意を促すことにする。図 4 の縦軸の予測値を 5 ppb 単位の区間に分け、各区間での過去の実測値の分布を作成する。そして、分布の全面積に対する、120 ppb (神奈川県における光化学スモッグ注意報の発令基準値)以上の領域の面積を、注意報の発報確率として、ユーザーインターフェースに表示する。図 2 で表示されている確率はこのように計算したものである。過去 12 年間のすべてのタイル点でのデータを用いて学習しているので、統計量は十分あり、信頼性の高い確率予報が可能だと考えている。

まとめ

本研究課題では、公的にアクセス可能な過去の大気質と気象の大規模データセットを学習し、光化学オキシダント(Ox)濃度を予測するソフトウェア「あんでるさん」を開発した。光化学オキシダントは、人間の健康や環境生態系、社会経済活動に深刻な影響を及ぼすため、その高精度な予測は不可欠である。

既存の予報が抱える精度や広域対応の課題に対し、本研究では機械学習技術、特に Gated recurrent unit(GRU)タイプの Recurrent neural network (RNN)を導入し、この課題解決に取り組んだ。

開発された「あんでるさん」は、地域を地理院タイルと呼ばれる長方形のタイルに細分化し、各格子点での測定値を近隣 3 測定局の測定値を Delaunay 三角形分割により内挿することで推定する手法を採用している。これにより、従来の個別測定局での予測しかできない「えだまめ」が抱えていた、測定局の偏在や欠測によるデータ不備の問題を克服した。

予測アルゴリズムの性能評価として、神奈川県内の過去 12 年間のデータを用いた検証では、8 時間先の光化学オキシダント濃度の同日予報において、予測は実測に比べてやや低めの値となる傾向が認められたものの、高濃度時の予測も実測と高い相関を示した。また、翌日の最大値を予測する前日予報においては、実測値と予測値の相関図から、特に高濃度域で予測精度が高いことが確認され、目標の水準を達成したと考えている。予測値をそのまま使用するのではなく、注意報発令基準値（神奈川県では 120 ppb）以上の「注意報発報確率」として提示する原理により、ユーザーはより実用的な情報を得られる。

現在、開発を進めているウェブポータルは、一般ユーザー向け UI（図 2）がほぼ完成しており、プロフェッショナル向け UI（図 3）も開発中である。2025 年 6 月時点では API が神奈川県内に限定されているが、2025 年度早期の全国展開を目指している。また、一般ユーザー向け実運用のためには、API サーバーのインターネットへの移設と、その常時運用コストの確保が今後の重要な課題である。プロ向けインターフェースの完成に向けては、地方公共団体の予報担当者との連携を深めていく計画である。

謝辞

本プロジェクトは、神奈川県環境科学センターの北岡勇樹氏との共同研究として進められました。プロジェクトの根幹をなす、過去の大気測定ビッグデータを AI に学習させ、高濃度 OX を予測するという着想は、静岡県環境衛生科学研究所の小田祐一主査によるものです。小田氏は、クラウドファンディングによる資金調達にも挑戦されながら、私たちにソフトウェアの設計を詳細に解説してくださいました。その熱意に触発され、私たちは神奈川県版の予測システム(えだまめ)の開発に着手した経緯があります。えだまめおよびあんでるさんで我々が開発したコードには、小田氏のコードは直接的には含まれていません。しかし、大気測定ビッグデータからオキシダントを予測するという基本設計は共通しています。さまざまな政治的な企てによって小田氏との共同開発は実現しませんでした。小田氏の構想が形を変えて実現したことを大変感慨深く思います。この場を借りて、小田氏に改めて深謝いたします。公益財団法人天野工業技術研究所からは、研究助成金をいただき、多岐にわたる挑戦ができました。大気汚染観測精度向上のためのミニ測定局開発や、大気汚染状況の VR 表示システムなど、次期研究に繋がる試みもその一つです。また、岡山と神奈川の二拠点で進める本プロジェクトにおいて、助成金による共同研究基盤の強化は、情報交換、試作画面の共有、開発環境の共有を円滑に進める上で不可欠でした。心より感謝申し上げます。

神奈川県環境科学センターの加藤陽一所長、三島聡子調査研究部長には、私たちの萌芽的な研究を温かく見守り、支援していただきました。この場を借りて、厚く御礼申し上げます。

大気観測データのクレジット表示: 国立環境研究所 環境展望台 大気汚染常時監視データファイル。

<https://tenbou.nies.go.jp/download/>, (参照 2025-04-04)

参考文献

- 1) Delaunay, B. (1934). Sur la sphère vide. A la mémoire de Georges Voronoï. *Известия Российской академии наук Серия математическая*, (6), 793–800.
- 2) Matsumoto, M. (n.d.). *Airpollutionwatch*. Github. Retrieved from <https://github.com/vitroid/airpollutionwatch>
- 3) *Andersan: Andersan developers' python library*. (n.d.). Github. Retrieved from <https://github.com/photochemistry/andersan>
- 4) *Andersan-core: Core of andersan*. (n.d.). Github. Retrieved from <https://github.com/photochemistry/andersan-core>
- 5) *Andersan-api: API for andersan*. (n.d.). Github. Retrieved from <https://github.com/photochemistry/andersan-api>
- 6) *Andersan-ui*. (n.d.). Github. Retrieved from <https://github.com/photochemistry/andersan-ui>
- 7) *Andersan-pro*. (n.d.). Github. Retrieved from <https://github.com/photochemistry/andersan-pro>
- 8) Salem, F. M. (2022). Gated RNN: The Gated Recurrent Unit (GRU) RNN. In *Recurrent Neural Networks* (pp. 85–100). Cham: Springer International Publishing.

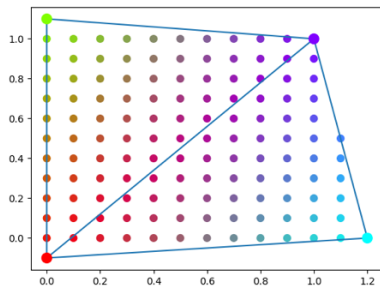


図 1 Delaunay 三角形分割による内挿の例。三角形の頂点が測定局、色が測定値を表す。三角形内部の格子点の色は、それを囲む三角形の三頂点の内挿で一意に定まる。

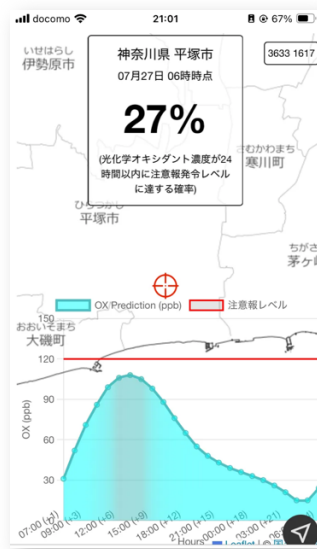


図 2 iOS での Web UI の表示例。地図をスクロールして地点を動かすと、その場所での予報が表示される。

<

図 3 MacOS Safari での Web UI の表示例。灰色の部分はすこし確率が高くなっている部分。さらに確率が高くなると赤い表示に変わる

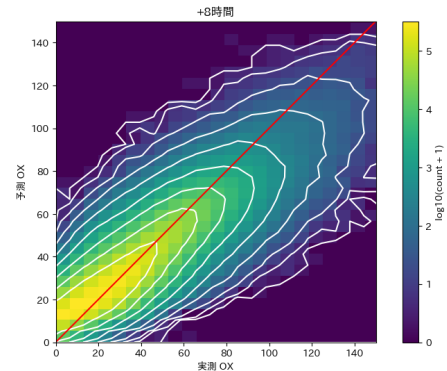


図 4 光化学オキシダントの予測と実測の相関図。縦軸が 8 時間後予測値、横軸が実測値。等高線は対数尺度であり、等間隔であれば指数減衰をあらわしている。

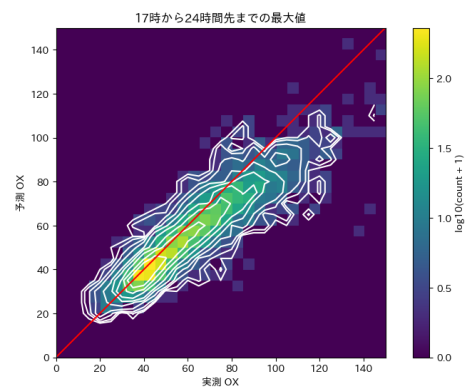


図 5 17 時時点での、24 時間先までの神奈川県内 77 地点での OX の最大値の、実測値と予測値の相関。等高線は対数尺度となっている。