

明星大学 理工学部 総合理工学科 環境科学系

令和 6 年度卒業論文

単一地点における DNN を用いた O_x 短期予測の
最適パラメータ探索

学籍番号 21T7-008

氏名 今給黎 樹

研究室名 大気科学研究室

指導教官 櫻井 達也

要旨

大気汚染物質の一つである光化学オキシダント(Ox)は大気汚染物質の一つであり、人体に悪影響を与えることが分かっている。そこで、環境省では大気汚染防止法に基づき光化学オキシダント注意報を定めている。これは大気汚染防止法に基づき Ox 濃度の 1 時間値が 0.12ppm 以上になり、気象条件からみてその状態が継続すると認められる場合に都道府県知事等が発令を行うものである。発令の判断は難しいため、AI を用いた短期予測によって発令の補助をおこなおうとする研究がなされている。この研究において AI モデルを作成する際の特徴量の探索が不十分であったため本研究にて高濃度(80ppb)を高精度に予測できるモデルを作成できる特徴量の探索を行うことにした。

仮説として特徴量需要度が高い物が最適になると考えた。そこで、9 地点を対象に各地点で全データの 24 時間分で学習を行いそこで特徴量需要度の算出を行った。需要度の算出は高濃度以上の時、高濃度以下(低濃度)の時に分けて計算した。こうして算出された特徴量需要度を降順に並び替え高濃度以上の時の上位 10 個、上位 20 個、上位 30 個、高濃度以上と低濃度の時のそれぞれの上位 20 個を合わせた高低上位 20 個のものを学習データとした。また、比較のためのベンチマークとして Ox の生成に関与し、先行研究で用いられていた Ox 、NMHC、 NOx 、TEMP の 24 時間分のデータを選定した。これらのデータで学習を行い時間値別、日別での評価をもとに比較を行ったところ、1 時間後予測に関しては上位 20 個を用いたものが多かった。2~3 時間後予測に関しては特段規則性がなかった。このことから地点ごと n 時間後予測ごとに特徴量を選定する必要があることが分かった。そのため、東京と埼玉の二つの地域に分けて個々の特徴量を見たところ東京、埼玉どちらともすべて重複していたのは 1 時間前の Ox 、1 時間前の湿度、1 時間前の気温であった。

今回出た結果から 1 時間前の Ox 、1 時間前の湿度、1 時間前の気温が高濃度域の Ox を予測するうえで欠かせない特徴量であることが示唆された。また、「探索を行うにあたって、高濃度域で特徴量需要度が高いものが最適な特徴量になる」という仮説を完全に支持するものではないが、本研究の結果はその可能性を示している。

目次

1. 研究の背景と目的	5
2. 機械学習について	6
2.1 機械学習	6
2.2 ニューラルネットワーク	6
2.3 DNN (Deep Neural Network)	7
3. 研究手法	7
3.1 対象地点	7
3.2 使用データ	8
3.3 モデル詳細	9
3.4 評価方法	9
3.4.1 モデル評価	9
3.4.2 特徴量評価	10
3.5 特徴量の探索	11
4. 学習結果	11
4.1 東秩父	11

4.2	鴻巣.....	15
4.3	幸手.....	18
4.4	東青梅.....	21
4.5	所沢.....	24
4.6	草加.....	27
4.7	多摩市愛宕.....	30
4.8	世田谷.....	33
4.9	南葛西.....	36
5.	考察.....	39
5.1	各地点の最適な特徴量.....	39
6.	まとめ.....	41
7.	謝辞.....	41
8.	参考文献.....	42

1. 研究の背景と目的

光化学オキシダント(Ox)は大気汚染物質の一つである。これが大気中に滞留すると上空がもやがかかったようになり、視界が悪くなる状態である「光化学スモッグ」を引き起こし、呼吸器系や循環器系、代謝系、神経系などといった器官に悪影響を及ぼす。そのため環境省では光化学スモッグによる健康被害を抑えるために、大気汚染防止法に基づき光化学オキシダント注意報を定めている。これは大気汚染防止法に基づき Ox 濃度の 1 時間値が 0.12ppm 以上になり、気象条件からみてその状態が継続すると認められる場合に都道府県知事等が発令を行うものである¹⁾。令和 5 年の Ox 注意報等の発令状況は、発令都道府県数が 17 都府県、発令延日数が 45 日であり、令和 4 年(12 都府県、41 日)と比べ増加している¹⁾。また、被害届出状況として令和 5 年の光化学大気汚染によると思われる被害の届出人数は 2 人(1 県)であり、令和 4 年(0 人)と比較して増加している¹⁾。

注意報の発令は各自治体が常時監視局の実測値などをもとにこれから先の Ox 濃度を推測し、注意報発令の判断を行う。これには専門的知識や経験が必要になるため、的確な判断を行うことは容易ではない。このような背景のもと、AI(人工知能)技術の一種である機械学習(ディープラーニング)を用いて、短期的な高濃度予測を行い判断の補助を行おうという研究がなされている。細越(2022)²⁾は、Ox 高濃度が継続する可能性を事前に予測できれば健康被害の抑制、そして自治体による早期対策の検討を可能にすることに繋がるという考えのもと、Deep Neural Network(DNN)を用いて、常時観測値からリアルタイムで大気汚染(Ox 濃度)の動向を予測することの有用性を示した。また現在、細越(2024)では、複数地点のデータを取り込むことによる予測精度向上を目的とした研究が行われている。これは、関東地域特有の海陸風循環を表現する気象条件のデータ、更には化学物質の輸送を表現するための他地点のデータなどを学習させ、主に高濃度域の予測精度を上げるといった試みである。具体的には、選定された代表地点における限定データの学習から、それらの周辺地域まで含めた広範囲における濃度予測を目指している

機械学習において、目的変数(予測対象)に対してどのような説明変数(特徴量)を学習させればよいかを検討することは予測精度に影響するため重要である。細越(2022)の研究において選定された特徴量は、オキシダントの生成に関与する NOx、Ox、NMHC、TEMP であった。対象時間から何時間前までのデータを取り込むのかの検討は行われていたが、これらが Ox を予測するのに最適な特徴量であるかの検討は行われていない。また、Ox を予測するために最適な特徴量に関する知見は限られている。このことから、Ox を予測するために最適な特徴量を探索することによって、予測精度の向上が期待できると考えた。自治体等の現場において重要視されるのは、注意報を発令すべきか否かという点である。そのため、高濃度に関して高精度な予測できることが求められる。そこで、発令の基準である 120ppb からバッファを持たせ、本研究では 80ppb 以上を高濃度と定義した。作成した AI モデルが高濃度を高精度に予測できるかを Ox 生成に関与する NOx、Ox、NMHC、TEMP を基準とし、単一の常時監視局を対象としたベンチマーク試験を実施することで、Ox 予測における

最適な特徴量の選定を試みることにした。

2. 機械学習について

2.1 機械学習

近年、情報化が進み多岐の分野にわたり大量のデータを蓄積している。機械学習はそのデータをコンピューターに学習させることによって解析や予測を行わせる。これは、様々な分野で活用されており、物品販売量と顧客の購買情報に関するデータを収集・分析するマーケティングや去の販売データ・購入履歴を基に、自動でその顧客に合ったおすすめ商品やサービスが表示されるレコメンデーション、不良品を検知するための画像認識などと実用化が進んでいる。機械学習を行うにあたっては、目的変数と説明変数が必要になる。目的変数は、解析もしくは予測対象であり、説明変数はこれを用いて解析もしくは予測対象を表現するために用いられるもので、特徴量や独立変数などと呼ばれるものである。これらを与えることによって学習を行うことでモデルを作成する。このようにできたモデルで誰でも予測が簡単に行え、説明が難しい事象に対して特徴を見出して解析もしくは予測を行うことができるのが機械学習の優れている点である。ただ欠点として、コンピューターによって見出される特徴はブラックボックス化されており、そのモデル内での特徴量がどのように影響を及ぼしているかはわかりづらい。そのため、目的変数を適切に表現するための特徴量は目的変数毎に適切なものを選択しないと行けず、そこに定跡はなく基本は手作業で探索することが求められる。

2.2 ニューラルネットワーク

機械学習に用いられるニューラルネットワークは、人間の脳のニューロンの構造や働きをもとに考案されたモデルである。ニューラルネットワークとは、入力された特徴量の値に対して重みを掛け、バイアスを加え、活性化関数に通すことで出力の値を表現する非線形の数式であり、全結合層とも称される。活性化関数は、各層のニューロンから得た値を通すことで出力の値に整える働きをする様々な式の総称である。この活性化関数の働きにより、ニューラルネットワークを用いた様々な非線形的特徴量と出力の関係に対しても近似を行うことが可能となった。ここでの層とは重みとバイアス、活性化関数を用いた処理を1回行う単位のことであり、通常のニューラルネットワークではこの層を重ねて構成している。

ニューラルネットワークにおいて、外部から入力を受け取る層を入力層、外部に値を出力する層を「出力層」といい、前の層の出力を入力として処理を行い次の層の入力に対して値を出力する層を「隠れ層」という。図1における y_pred が「出力層」、 x_input と y_pred の間のものが「隠れ層」のイメージとなる。また、入力層から隠れ層、出力層へ値を前に進め出力を求める処理を伝播という。ニューラルネットワークを用いた学習では、入力に対して得られた出力をより正確にするために重みとバイアスを更新する誤差逆伝播法を用いる。

誤差逆伝播法では伝播の処理とは逆向きに、出力の値と正解の値の誤差から重みとバイアスを調整する⁴⁾。

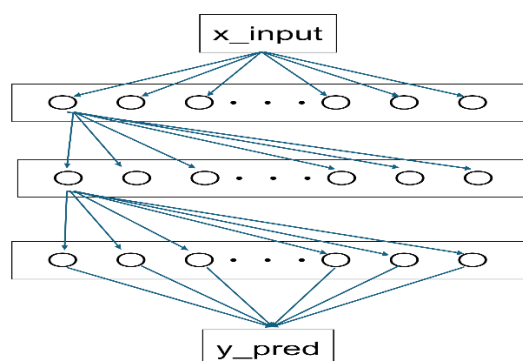


図1 ニューラルネットワークのイメージ

2.3 DNN (Deep Neural Network)

本研究で用いる DNN は機械学習の一種であり、ニューラルネットワークの中層が多層になった、多層ニューラルネットワークを用いた機械学習手法の総称である。従来は人が設計した特徴量を用いて機械学習を行っていたが、ディープラーニングは機械が自ら特徴を設計、抽出し、学習を行う。その結果ディープラーニングを用いた手法は従来手法よりも高い識別率を示している。DNN は機械学習の技術として幅広い分野での利用が進められており、画像認識、音声認識、翻訳等といった場面で使用されている。

3. 研究手法

本研究において、高濃度を高精度に予測できるモデルを作成できることが目的である。単一点において、この目的を満たす特徴量の探索を行った。

3.1 対象地点

本実験において先行研究³⁾との比較を行うため先行研究でテスト地点として選定されていた地点を対象とした。なお、その地点は図2に示した東秩父、鴻巣、幸手、東青梅、所沢、草加、多摩市愛宕、世田谷区、南葛西の計9地点である。



図2 対象地点

3.2 使用データ

使用データには、国立環境研究所の測定物質全ての時間値データを用いた。データの前処理として、欠損値に関しては前後の値から線形補完を行い、正規化や標準化を行わずそのままのデータを投入した。また、データテーブルは下記の表1のようにになっている。ここでtはある日時のことを指し、目的変数として1~3時間後のOx濃度があり、それに対する特徴量としてある日時tから24時間前までのデータを取り込んだ。

対象期間は以下の通りとし、この期間の予測を行わせた。

学習データ : 2018年 4月1日~3月31日 (データ数 n:8505)

テストデータ : 2019年 4月1日~3月31日 (データ数 n:8550)

表1 Ox濃度予測モデルに使用するデータセット構造

	目的変数			特徴量				
	測定日時	Ox_t	Ox_t+1	Ox_t+2	測定物質_t-1	測定物質_t-2	測定物質_t-3	.. 測定物質_t-24
t	時間値データ							

また、特徴量においてその地点で測定しているすべてのデータを候補にすること、地点ごとに測定している物質が異なるため対象地点において測定している物質を表2にまとめた。表2で○がついているものはその地点で測定している物質であり、×がついている物質は測定していない物質のことを指す。

表 2 対象地点における測定物質

	CH ₄	CO	HUM	NO	NO ₂	NO _x	O ₃	PM _{2.5}	SO ₂	SPM	TEMP	THC	WD	WS
東秩父	○	○	○	○	○	○	○	○	○	○	○	○	○	○
鴻巣	○	×	○	○	○	○	○	○	○	○	○	○	○	○
幸手	○	×	○	○	○	○	○	○	○	○	○	○	○	○
東青梅	○	○	○	○	○	○	○	○	○	○	○	○	○	○
所沢	○	×	○	○	○	○	○	○	○	○	○	○	○	○
草加	○	○	○	○	○	○	○	×	○	○	○	○	○	○
多摩市愛宕	○	○	○	○	○	○	○	○	○	○	○	○	○	○
世田谷	○	○	○	○	○	○	○	○	○	○	○	○	○	○
南葛西	○	×	○	○	○	○	○	○	×	○	○	○	○	○

3.3 モデル詳細

学習時のユーザーが設定するハイパラメータを表 3 にまとめた。

表 3 モデルのハイパラメータ

使用言語	Python
ライブラリ	PyTorch
損失関数	RMSE(二乗平均平方根誤差)
活性化関数	ReLU
学習係数	0.001
中間層	4層
エポック数	70

Python のバージョンは 3.12 であり、学習には Facebook 社の提供するオープンソースフレームワークである PyTorch を用いた。計算実験に使用したハードウェアは以下の通りである。

OS Windows 11 HOME バージョン 23H2
CPU AMD Ryzen 7 5700X @4.6GHz 8 コア
GPU NVIDIA GeForce RTX 3070

3.4 評価方法

3.4.1 モデル評価

本実験において優れたモデルは注意報等を発令する補助となるものであり、高濃度を高

精度に予測することが求められる。そのため、評価方法として時間値データの (i) 予測値と実測との RMSE(平均二乗平方根誤差)値が低いこと、(ii) 高濃度時の追跡率が高いこと、(iii) 高濃度時の RMSE が低いこと、といった 3 つの観点をもとに優れたモデルを作成できる最適な特徴量を探索することとした。(i) はモデル全体の精度を確認するものであり、一般的に RMSE 値が低いと実測と予測との誤差がすくないと解釈できる。ただ、これだけであると低濃度帯での予測精度はよいが高濃度帯での予測精度は悪いといった際にも RMSE 値が低いことがありうる。これだけでは本実験での目的を満たすことはできない。そこで、(ii)～(iii)のような高濃度域に対する評価が必要である。(ii) では、実測値が高濃度を越えた時に対する予測が高濃度を越えている割合であり、そのモデルで高濃度を予測できているかの可否を評価している。そして、(iii) の実測値が高濃度を越えた時のみに絞り RMSE を算出することによって、どれほどの精度で高濃度を予測しているかを評価した。

また、再現率、適合率、調和平均を算出し評価を行った。実測で高濃度を観測した日に予測で高濃度を越えた日の割合である再現率は式に表される。予測で高濃度を観測した日に実測で高濃度を観測した日の割合である適合率は式で表される。これら二つの割合の平均の調和平均(F 値)は式(1)で表される。ここにおける調和平均は高濃度域に関してどれほど高精度に予測できるかを示した値となる。そのためこの値が一番高いものが目的を満たせるモデル、特徴量だと判断できる。これらの評価をもとに各モデルを評価した。

$$\text{調和平均 (\%)} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \cdots \text{式 (1)}$$

最終的な評価は調和平均、高濃度 RMSE、特徴量の個数の順で優先度を設定し評価を最も優れていたものを抽出する。

モデルを作成する際、同じデータを与えても違う精度のモデルができる。そのため、評価する際はモデルを 10 回作成しそれぞれの値の平均をとり評価を行った。

3.4.2 特徴量評価

特徴量の評価を行うためオープンソースである SHAP (SHapley Additive exPlanations) を用いて特徴量の需要度を計算させ、モデルの評価同様 10 回算出した平均を用いて特徴量需要度として特徴量の評価を行った。SHAP ではシャープレイ値と呼ばれる目的変数に対して、ある特徴量がどれほど貢献しているか(目的変数を説明するために重要か)を表す。これは、平均的な予測値をベースに、ある特徴量に加わったときの予測値の変動量を、その特徴量の限界貢献度として扱い、全ての順序の組み合わせで限界貢献度を求め、その平均をとることで求まる⁵⁾。すなわち、値が大きい時予測を行うにあたって需要度が高い。このようにして求めた値を用いて各特徴量が Ox の予測に対してどれほどの需要なのかを評価した。この時、高濃度域と低濃度域(高濃度域以外)に対する特徴量需要度を算出させた。

3.5 特徴量の探索

探索を行うにあたって、高濃度域で特徴量需要度が高いものが最適な特徴量になると仮説を立てた。それぞれの地点において測定している物質の時間値データを全て学習させる。これにより、どの物質の何時間前のデータが需要度として高いかを確認する。特徴量需要度は高濃度以上の時と、高濃度以外の時で特徴量需要度を分けて算出した。

ここで算出した特徴量需要度を降順に並び替えた。高濃度以上の上位 10 個、20 個、30 個、高濃度以外の上位 20 個、高濃度以上の上位 20 個を合わせたものを学習させた。この時の特徴量を評価させた。

4. 学習結果

以下に地点ごとの結果を示した。また、下記における「lag=n」は n 時間後予測のことを指す。特徴量に関して「物質名_n」という命名規則になっており、この時の n は n 時間前のデータを表している。また、再現率、適合率の表中にあるかっこ内の数字は n 数のことを指す。

4.1 東秩父

表 4 に今回用いた東秩父における特徴量の詳細を示した。後述する特徴量に関しては表 4 の特徴量と対応している。

表 4 東秩父における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	HUM_01	TEMP_01	SPM_01	Ox_03	TEMP_02	HUM_22	PM _{2.5} _05	TEMP_04	PM _{2.5} _22
上位10個_lag=2	Ox_01	HUM_01	TEMP_01	SPM_01	PM _{2.5} _04	PM _{2.5} _05	HUM_21	PM _{2.5} _21	Ox_06	TEMP_11
上位10個_lag=3	Ox_01	HUM_01	TEMP_01	SPM_01	Ox_06	PM _{2.5} _04	TEMP_02	HUM_20	PM _{2.5} _14	TEMP_11
上位20個_lag=1	Ox_01	HUM_01	TEMP_01	SPM_01	Ox_03	TEMP_02	HUM_22	PM _{2.5} _05	TEMP_04	PM _{2.5} _22
	TEMP_12	Ox_08	Ox_05	PM _{2.5} _15	TEMP_11	Ox_06	WD_01	TEMP_19	PM _{2.5} _02	SPM_24
上位20個_lag=2	Ox_01	HUM_01	TEMP_01	SPM_01	PM _{2.5} _04	PM _{2.5} _05	HUM_21	PM _{2.5} _21	Ox_06	TEMP_11
	Ox_07	PM _{2.5} _15	PM _{2.5} _14	TEMP_12	Ox_05	PM _{2.5} _01	HUM_20	WD_01	TEMP_13	WD_02
上位20個_lag=3	Ox_01	HUM_01	TEMP_01	SPM_01	Ox_06	PM _{2.5} _04	TEMP_02	HUM_20	PM _{2.5} _14	TEMP_11
	WD_01	Ox_05	PM _{2.5} _03	TEMP_12	PM _{2.5} _13	PM _{2.5} _22	PM _{2.5} _05	TEMP_13	PM _{2.5} _21	TEMP_14
上位30個_lag=1	Ox_01	HUM_01	TEMP_01	SPM_01	Ox_03	TEMP_02	HUM_22	PM _{2.5} _05	TEMP_04	PM _{2.5} _22
	TEMP_12	Ox_08	Ox_05	PM _{2.5} _15	TEMP_11	Ox_06	WD_01	TEMP_19	PM _{2.5} _02	SPM_24
	TEMP_03	PM _{2.5} _04	PM _{2.5} _09	PM _{2.5} _20	PM _{2.5} _14	SPM_14	TEMP_16	Ox_10	TEMP_24	Ox_22
上位30個_lag=2	Ox_01	HUM_01	TEMP_01	SPM_01	PM _{2.5} _04	PM _{2.5} _05	HUM_21	PM _{2.5} _21	Ox_06	TEMP_11
	Ox_07	PM _{2.5} _15	PM _{2.5} _14	TEMP_12	Ox_05	PM _{2.5} _01	HUM_20	WD_01	TEMP_13	WD_02
	PM _{2.5} _22	PM _{2.5} _13	PM _{2.5} _03	TEMP_14	Ox_03	TEMP_03	TEMP_04	HUM_15	HUM_02	TEMP_15
上位30個_lag=3	Ox_01	HUM_01	TEMP_01	SPM_01	Ox_06	PM _{2.5} _04	TEMP_02	HUM_20	PM _{2.5} _14	TEMP_11
	WD_01	Ox_05	PM _{2.5} _03	TEMP_12	PM _{2.5} _13	PM _{2.5} _22	PM _{2.5} _05	TEMP_13	PM _{2.5} _21	TEMP_14
	HUM_19	HUM_21	HUM_02	Ox_04	TEMP_10	PM _{2.5} _15	TEMP_03	PM _{2.5} _20	Ox_07	PM _{2.5} _01
高低上位20個_lag=1	TEMP_12	TEMP_24	Ox_05	PM _{2.5} _05	TEMP_08	PM _{2.5} _22	PM _{2.5} _15	HUM_04	Ox_06	TEMP_01
	SPM_01	TEMP_23	Ox_03	TEMP_02	Ox_01	HUM_22	TEMP_03	WD_01	PM _{2.5} _02	TEMP_20
	TEMP_06	HUM_03	TEMP_19	TEMP_04	Ox_02	HUM_01	Ox_08	SPM_24	TEMP_11	
高低上位20個_lag=2	TEMP_24	TEMP_12	Ox_05	PM _{2.5} _21	PM _{2.5} _05	PM _{2.5} _04	PM _{2.5} _15	Ox_07	Ox_06	TEMP_13
	PM _{2.5} _14	TEMP_01	SPM_01	PM _{2.5} _01	TEMP_23	TEMP_02	HUM_21	Ox_01	Ox_04	WD_02
	TEMP_20	TEMP_05	TEMP_06	HUM_20	Ox_02	HUM_01	WD_01	TEMP_10	TEMP_11	
高低上位20個_lag=3	TEMP_24	TEMP_12	Ox_05	PM _{2.5} _21	TEMP_09	SPM_02	PM _{2.5} _05	PM _{2.5} _22	PM _{2.5} _04	Ox_06
	TEMP_13	PM _{2.5} _14	TEMP_01	SPM_01	TEMP_14	TEMP_23	Ox_03	TEMP_02	Ox_01	Ox_04
	TEMP_05	PM _{2.5} _13	TEMP_06	HUM_20	TEMP_19	TEMP_04	Ox_02	PM _{2.5} _03	HUM_01	WD_01
	TEMP_10	TEMP_11								

表 5、6 にそれぞれ東秩父の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関して全データが最も高い 85.156%であり、次点で上位 20 個の 81.538%であった。高濃度 RMSE、RMSE では、ベンチマークの精度よりも、全データや上位 20 個の方が精度は悪く、全データよりも上位 20 個の方が精度は悪い。ただ、上位 20 個の方が特徴量の個数が明らかに少なく、調和平均、高濃度 RMSE に関して全データと上位 20 個に明確な差はほぼない。また、日別評価を見ると上位 10 個が最も高く 86.364%と、次点で上位 20 個の 83.333%であった。このことから 1 時間後予測に関しては時間別、日別での評価が高い上位 20 個が高精度であると考えた。

2 時間後予測の時間別評価では、調和平均に関して上位 20 個が最も高い 65.487%で、次点で上位 30 個の 64.681%であった。高濃度 RMSE では、上位 30 個に関して 15.009、上位 20 個で 15.687 と調和平均とは逆で上位 30 個の方が精度はよかった。また、日別評価を見ると上位 20 個で最も高い 71.429%であり、上位 30 個で 76.697%であった。このことから、2 時間後予測に関して、時間別、日別で評価が最も高い上位 20 個が高精度であると考えた。

3 時間後予測の時間別評価では、調和平均に関してベンチマークが最も高く 58.621%で、次点で高低上位 20 個が 57.851%であった。高濃度 RMSE では、高低上位 20 個が最も精度

が良く 17.536 であった。また、日別でみると調和平均に関して、全データで最も高い 65.116%、次点で上位 20 個の 60.465%であり、高低上位 20 個は日別でみると 53.333%と他と比べると低い値となっている。しかし、ベンチマークよりも特徴量の個数が少なく、時間値で見たときに調和平均、高濃度 RMSE に関して精度良いため 3 時間後予測に関して高低上位 20 個が高精度であると考え

表 5 東秩父における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	74.615	84.348	79.184	4.055	8.240
全データ_lag = 1	83.846	86.508	85.156	4.096	8.403
上位 10 個_lag = 1	73.077	84.821	78.512	3.924	8.631
上位 20 個_lag = 1	81.538	81.538	81.538	4.200	8.455
上位 30 個_lag = 1	67.692	83.019	74.576	4.007	8.814
高低上位 20 個_lag = 1	63.077	88.172	73.543	3.991	9.753
ベンチマーク_lag = 2	42.308	80.882	55.556	6.122	16.664
全データ_lag = 2	44.615	79.452	57.143	6.320	16.222
上位 10 個_lag = 2	39.231	72.857	51.000	6.251	17.470
上位 20 個_lag = 2	56.923	77.083	65.487	6.524	15.687
上位 30 個_lag = 2	58.462	72.381	64.681	6.102	15.009
高低上位 20 個_lag = 2	39.231	80.952	52.850	6.511	19.143
ベンチマーク_lag = 3	52.308	66.667	58.621	7.782	18.342
全データ_lag = 3	57.692	57.692	57.692	7.626	18.418
上位 10 個_lag = 3	37.692	79.032	51.042	7.517	20.406
上位 20 個_lag = 3	43.846	71.250	54.286	7.876	18.494
上位 30 個_lag = 3	52.308	61.818	56.667	8.468	17.798
高低上位 20 個_lag = 3	53.846	62.500	57.851	7.680	17.536

表 6 東秩父における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	70.833	89.474	79.070
全データ_lag = 1	75.000	81.818	78.261
上位 10 個_lag = 1	79.167	95.000	86.364
上位 20 個_lag = 1	83.333	83.333	83.333
上位 30 個_lag = 1	70.833	94.444	80.952
高低上位 20 個_lag = 1	66.667	94.118	78.049
ベンチマーク_lag = 2	41.667	90.909	57.143
全データ_lag = 2	41.667	83.333	55.556
上位 10 個_lag = 2	41.667	66.667	51.282
上位 20 個_lag = 2	62.500	83.333	71.429
上位 30 個_lag = 2	62.500	78.947	69.767
高低上位 20 個_lag = 2	37.500	81.818	51.429
ベンチマーク_lag = 3	45.833	84.615	59.459
全データ_lag = 3	58.333	73.684	65.116
上位 10 個_lag = 3	33.333	80.000	47.059
上位 20 個_lag = 3	54.167	68.421	60.465
上位 30 個_lag = 3	58.333	58.333	58.333
高低上位 20 個_lag = 3	50.000	57.143	53.333

4.2 鴻巣

表 7 に今回用いた鴻巣における特徴量の詳細を示した。後述する特徴量に関しては表 7 の特徴量と対応している。

表 7 鴻巣における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	TEMP_03	Ox_04	Ox_03	Ox_21	Ox_02	Ox_15
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_02	Ox_03	Ox_21	Ox_05	WS_01	TEMP_03
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	TEMP_24	Ox_03	WS_01	Ox_22	HUM_02	Ox_21
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	TEMP_03	Ox_04	Ox_03	Ox_21	Ox_02	Ox_15
	Ox_22	WS_02	Ox_19	TEMP_04	Ox_17	SPM_01	Ox_20	WS_01	HUM_20	TEMP_21
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_02	Ox_03	Ox_21	Ox_05	WS_01	TEMP_03
	Ox_23	Ox_20	Ox_22	WS_02	Ox_15	Ox_16	TEMP_24	TEMP_07	Ox_14	TEMP_04
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	TEMP_24	Ox_03	WS_01	Ox_22	HUM_02	Ox_21
	Ox_02	TEMP_07	Ox_20	TEMP_06	TEMP_05	TEMP_03	Ox_15	Ox_05	TEMP_19	TEMP_20
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	TEMP_03	Ox_04	Ox_03	Ox_21	Ox_02	Ox_15
	Ox_22	WS_02	Ox_19	TEMP_04	Ox_17	SPM_01	Ox_20	WS_01	HUM_20	TEMP_21
	TEMP_24	Ox_23	NOx_01	Ox_12	TEMP_14	NMHC_01	SPM_06	TEMP_16	HUM_02	Ox_06
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_02	Ox_03	Ox_21	Ox_05	WS_01	TEMP_03
	Ox_23	Ox_20	Ox_22	WS_02	Ox_15	Ox_16	TEMP_24	TEMP_07	Ox_14	TEMP_04
	TEMP_08	Ox_17	SPM_01	TEMP_20	TEMP_16	HUM_20	HUM_19	NO_07	HUM_02	TEMP_17
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	TEMP_24	Ox_03	WS_01	Ox_22	HUM_02	Ox_21
	Ox_02	TEMP_07	Ox_20	TEMP_06	TEMP_05	TEMP_03	Ox_15	Ox_05	TEMP_19	TEMP_20
	TEMP_04	Ox_14	WS_02	TEMP_18	TEMP_08	TEMP_17	Ox_19	TEMP_22	Ox_16	TEMP_21
高低上位20個_lag=1	NOx_01	TEMP_24	TEMP_12	Ox_15	Ox_22	Ox_19	TEMP_01	SPM_01	TEMP_14	TEMP_21
	TEMP_22	Ox_03	Ox_17	NOx_01	TEMP_02	Ox_21	Ox_01	Ox_04	WS_02	NOx_02
	TEMP_03	NOx_02	HUM_20	TEMP_04	Ox_20	Ox_02	HUM_02	TEMP_07	HUM_01	WS_01
	TEMP_11									
高低上位20個_lag=2	NOx_01	TEMP_24	Ox_05	Ox_15	Ox_22	Ox_14	TEMP_08	TEMP_01	TEMP_14	TEMP_21
	TEMP_22	Ox_03	TEMP_02	Ox_21	Ox_01	Ox_04	WS_02	TEMP_03	Ox_23	NOx_02
	TEMP_20	TEMP_06	Ox_16	TEMP_04	Ox_20	Ox_02	TEMP_07	HUM_01	TEMP_18	WS_01
	TEMP_11									
高低上位20個_lag=3	NOx_01	TEMP_24	Ox_05	Ox_15	Ox_22	TEMP_01	TEMP_21	TEMP_22	Ox_03	TEMP_02
	Ox_21	Ox_01	Ox_04	TEMP_17	TEMP_03	TEMP_20	TEMP_05	TEMP_06	TEMP_19	TEMP_04
	Ox_20	Ox_02	HUM_02	TEMP_07	HUM_01	TEMP_18	WS_01			

表 8、9 に鴻巣の時間値評価、日別評価を示した。

1 時間後予測の時間値データでは、調和平均が上位 20 個において最も高い 76.852%であった。モデル全体の評価である RMSE を見るとベンチマークよりも精度が悪い。ただ、高濃度 RMSE に関して、最も精度がよく 8.995 であった。日別評価で見た時、調和平均は上位 20 個が最も高い 85.714%であった。このことから 1 時間後予測に高精度なのは上位 20 個だと考えられる。

2 時間後予測の時間値データでは、調和平均が上位 10 個において抜きんで高く 66.055%であった。高濃度 RMSE、RMSE どちらにおいても上位 10 個が最も精度がよかった。日別評価で見ても、上位 10 個の調和平均が抜きんで高い。このことから 2 時間後予測に高精度なのは上位 10 個だと考えられる。

3 時間後予測では、調和平均が上位 10 個において最も高い 56.652%であり、次点で上位 20 個の 56.502%であった。高濃度 RMSE だと上位 20 個が最も精度が良い 19.062 であり、

次点で上位 10 個は 20.845 であった。また、日別の調和平均で上位 10 個の 65.625%で最も精度がよい。時間値と日別で精度がよい上位 10 個が 3 時間後予測に高精度であると考えた。

表 8 鴻巣における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	70.635	83.178	76.395	4.125	9.955
全データ_lag = 1	58.730	88.095	70.476	4.429	11.559
上位 10 個_lag = 1	50.000	90.000	64.286	4.880	12.916
上位 20 個_lag = 1	87.302	74.324	80.292	4.289	8.995
上位 30 個_lag = 1	76.984	80.833	78.862	4.164	9.050
高低上位 20 個_lag = 1	69.048	89.691	78.027	3.817	9.941
ベンチマーク_lag = 2	42.857	75.000	54.545	6.925	15.994
全データ_lag = 2	48.413	68.539	56.744	6.694	15.686
上位 10 個_lag = 2	57.143	78.261	66.055	6.350	14.576
上位 20 個_lag = 2	43.651	85.938	57.895	6.696	16.245
上位 30 個_lag = 2	26.984	85.000	40.964	6.999	19.214
高低上位 20 個_lag = 2	30.952	88.636	45.882	6.841	19.894
ベンチマーク_lag = 3	40.476	69.863	51.256	8.885	19.222
全データ_lag = 3	32.540	68.333	44.086	8.576	21.492
上位 10 個_lag = 3	52.381	61.682	56.652	8.808	20.845
上位 20 個_lag = 3	50.000	64.948	56.502	8.637	19.062
上位 30 個_lag = 3	26.190	89.189	40.491	9.071	23.733
高低上位 20 個_lag = 3	50.794	60.377	55.172	8.893	17.052

表 9 鴻巣における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_鴻巣_lag = 1	75.758	83.333	79.365
全データ_鴻巣_lag = 1	66.667	88.000	75.862
上位 10 個_鴻巣_lag = 1	57.576	86.364	69.091
上位 20 個_鴻巣_lag = 1	90.909	81.081	85.714
上位 30 個_鴻巣_lag = 1	78.788	83.871	81.250
高低上位 20 個_鴻巣_lag = 1	69.697	92.000	79.310
ベンチマーク_鴻巣_lag = 2	54.545	78.261	64.286
全データ_鴻巣_lag = 2	45.455	62.500	52.632
上位 10 個_鴻巣_lag = 2	69.697	82.143	75.410
上位 20 個_鴻巣_lag = 2	48.485	88.889	62.745
上位 30 個_鴻巣_lag = 2	30.303	71.429	42.553
高低上位 20 個_鴻巣_lag = 2	30.303	83.333	44.444
ベンチマーク_鴻巣_lag = 3	42.424	66.667	51.852
全データ_鴻巣_lag = 3	33.333	68.750	44.898
上位 10 個_鴻巣_lag = 3	63.636	67.742	65.625
上位 20 個_鴻巣_lag = 3	48.485	57.143	52.459
上位 30 個_鴻巣_lag = 3	21.212	77.778	33.333
高低上位 20 個_鴻巣_lag = 3	54.545	56.250	55.385

4.3 幸手

表 10 に今回用いた幸手における特徴量の詳細を示した。後述する特徴量に関しては表 10 の特徴量と対応している。

表 10 幸手における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	WS_02	TEMP_05	TEMP_06	WS_01	TEMP_04	TEMP_11
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	TEMP_05	WS_01	TEMP_04	TEMP_06	WS_02	Ox_03	Ox_15
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	TEMP_05	TEMP_04	TEMP_06	WS_01	SPM_01	Ox_02	Ox_15
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	WS_02	TEMP_05	TEMP_06	WS_01	TEMP_04	TEMP_11
	TEMP_12	TEMP_03	Ox_17	Ox_15	SPM_01	TEMP_02	Ox_24	TEMP_13	TEMP_17	SPM_03
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	TEMP_05	WS_01	TEMP_04	TEMP_06	WS_02	Ox_03	Ox_15
	Ox_16	TEMP_03	TEMP_11	Ox_14	TEMP_12	WD_11	SPM_02	TEMP_07	SPM_01	TEMP_19
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	TEMP_05	TEMP_04	TEMP_06	WS_01	SPM_01	Ox_02	Ox_15
	TEMP_03	Ox_14	Ox_19	SPM_02	Ox_18	TEMP_24	WS_02	TEMP_07	HUM_02	TEMP_22
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	WS_02	TEMP_05	TEMP_06	WS_01	TEMP_04	TEMP_11
	TEMP_12	TEMP_03	Ox_17	Ox_15	SPM_01	TEMP_02	Ox_24	TEMP_13	TEMP_17	SPM_03
	Ox_16	NOx_08	HUM_21	CH4_07	THC_07	NOx_01	PM2.5_19	NMHC_09	NO_11	SPM_06
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	TEMP_05	WS_01	TEMP_04	TEMP_06	WS_02	Ox_03	Ox_15
	Ox_16	TEMP_03	TEMP_11	Ox_14	TEMP_12	WD_11	SPM_02	TEMP_07	SPM_01	TEMP_19
	THC_06	Ox_20	NOx_17	TEMP_17	Ox_04	WS_17	HUM_21	CH4_06	PM2.5_19	Ox_19
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	TEMP_05	TEMP_04	TEMP_06	WS_01	SPM_01	Ox_02	Ox_15
	TEMP_03	Ox_14	Ox_19	SPM_02	Ox_18	TEMP_24	WS_02	TEMP_07	HUM_02	TEMP_22
	TEMP_18	Ox_13	TEMP_19	THC_06	WD_11	Ox_16	Ox_17	TEMP_17	NO_08	WD_10
高低上位20個_lag=1	NOx_01	TEMP_12	TEMP_24	Ox_15	TEMP_13	TEMP_01	SPM_01	Ox_03	Ox_17	NOx_01
	TEMP_02	Ox_01	WS_02	TEMP_17	NOx_02	TEMP_03	NOx_02	TEMP_05	TEMP_06	TEMP_04
	SPM_03	Ox_02	HUM_02	Ox_24	HUM_01	TEMP_10	WS_01	TEMP_11		
高低上位20個_lag=2	NOx_01	TEMP_16	TEMP_12	Ox_05	Ox_15	Ox_14	SPM_02	TEMP_01	WD_11	SPM_01
	TEMP_23	Ox_03	NOx_01	TEMP_02	Ox_01	WS_02	TEMP_17	TEMP_03	Ox_23	TEMP_05
	TEMP_06	Ox_16	TEMP_19	TEMP_04	Ox_02	TEMP_07	HUM_01	WS_01	TEMP_11	
高低上位20個_lag=3	TEMP_16	TEMP_24	Ox_15	Ox_14	SPM_02	Ox_19	Ox_18	TEMP_01	SPM_01	TEMP_21
	TEMP_22	TEMP_23	TEMP_02	Ox_01	WS_02	TEMP_17	TEMP_03	TEMP_20	TEMP_05	TEMP_06
	TEMP_19	TEMP_04	Ox_02	HUM_02	TEMP_07	HUM_01	TEMP_18	WS_01		

表 11、12 に幸手の時間値評価、日別評価を示した。

1 時間後予測の時間値データでは、調和平均が上位 10 個において最も高い 76.852%であった。また、高濃度 RMSE に関してはベンチマークが最も精度がよく 8.654 であり、上位 10 個の高濃度 RMSE で 8.701 であった。モデル全体の評価である RMSE を見ると、ベンチマークよりも上位 10 個の精度がよく特徴量の個数も上位 10 個の方が明らかに少ない。日別評価では上位 30 個の調和平均が 80.000%と一番高く次点で上位 10 個の 78.125%であった。日別評価で見た時上位 30 個がよく見えるが時間値で見た時の高濃度 RMSE が 11.159 と上位 10 個と比べた時に精度が悪い。そのため、1 時間後予測に高精度なのは上位 10 個だと考えられる。

2 時間後予測の時間値データでは、調和平均が上位 10 個において抜きんでて高い 62.745%であった。日別評価に関しても同様に上位 10 個において抜きんでて高い 66.667%であった。このことから、2 時間後予測に高精度なのは上位 10 個であると考ええる。

3 時間後予測では、調和平均が上位 20 個において最も高い 47.619%であり、次点で全デ

ータの 43.210%であった。高濃度 RMSE は、全データが 19.077 で、上位 20 個が 19.369 であり、全データの方が精度はよい。ただ、特徴量の個数は、上位 20 個の方が明らかにデータ数が少ない。また、日別評価に関して上位 20 個が抜きんでて高い 46.154%であった。このことから、3 時間後予測に高精度なのは上位 20 個であると考ええる。

表 11 幸手における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	83.495	69.919	76.106	4.818	8.654
全データ_lag = 1	34.951	94.737	51.064	4.527	14.481
上位 10 個_lag = 1	80.583	73.451	76.852	4.100	8.701
上位 20 個_lag = 1	67.961	81.395	74.074	4.381	9.331
上位 30 個_lag = 1	60.194	89.855	72.093	4.085	11.159
高低上位 20 個_lag = 1	60.194	84.932	70.455	3.849	10.536
ベンチマーク_lag = 2	47.573	63.636	54.444	6.966	14.153
全データ_lag = 2	28.155	70.732	40.278	6.912	18.899
上位 10 個_lag = 2	62.136	63.366	62.745	6.715	13.273
上位 20 個_lag = 2	37.864	75.000	50.323	6.661	17.040
上位 30 個_lag = 2	37.864	76.471	50.649	6.695	16.631
高低上位 20 個_lag = 2	20.388	80.769	32.558	7.586	21.279
ベンチマーク_lag = 3	25.243	63.415	36.111	9.098	21.029
全データ_lag = 3	33.981	59.322	43.210	8.708	19.077
上位 10 個_lag = 3	24.272	65.789	35.461	8.716	23.174
上位 20 個_lag = 3	38.835	61.538	47.619	9.093	19.369
上位 30 個_lag = 3	31.068	69.565	42.953	8.707	20.338
高低上位 20 個_lag = 3	27.184	70.000	39.161	8.528	21.118

表 12 幸手における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	80.645	73.529	76.923
全データ_lag = 1	41.935	86.667	56.522
上位 10 個_lag = 1	80.645	75.758	78.125
上位 20 個_lag = 1	70.968	84.615	77.193
上位 30 個_lag = 1	70.968	91.667	80.000
高低上位 20 個_lag = 1	61.290	79.167	69.091
ベンチマーク_lag = 2	48.387	60.000	53.571
全データ_lag = 2	22.581	63.636	33.333
上位 10 個_lag = 2	67.742	65.625	66.667
上位 20 個_lag = 2	41.935	65.000	50.980
上位 30 個_lag = 2	38.710	80.000	52.174
高低上位 20 個_lag = 2	12.903	80.000	22.222
ベンチマーク_lag = 3	19.355	66.667	30.000
全データ_lag = 3	29.032	56.250	38.298
上位 10 個_lag = 3	19.355	60.000	29.268
上位 20 個_lag = 3	38.710	57.143	46.154
上位 30 個_lag = 3	29.032	64.286	40.000
高低上位 20 個_lag = 3	22.581	63.636	33.333

4.4 東青梅

表 13 に今回用いた東青梅における特徴量の詳細を示した。後述する特徴量に関しては表 13 の特徴量と対応している。

表 13 東青梅における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	HUM_02	Ox_03	TEMP_08	Ox_23	Ox_19	WS_02
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	TEMP_08	Ox_23	Ox_03	Ox_05	Ox_04	Ox_18	TEMP_24
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_05	TEMP_24	WS_01	Ox_17	TEMP_08	Ox_03
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	HUM_02	Ox_03	TEMP_08	Ox_23	Ox_19	WS_02
	TEMP_03	TEMP_09	HUM_24	HUM_09	TEMP_21	Ox_11	Ox_24	TEMP_13	SPM_02	TEMP_15
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	TEMP_08	Ox_23	Ox_03	Ox_05	Ox_04	Ox_18	TEMP_24
	TEMP_09	WS_01	WS_02	Ox_22	HUM_24	PM _{2.5} _01	Ox_19	HUM_05	Ox_06	TEMP_10
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_05	TEMP_24	WS_01	Ox_17	TEMP_08	Ox_03
	Ox_22	TEMP_09	Ox_18	Ox_23	HUM_24	TEMP_07	TEMP_20	TEMP_18	TEMP_10	TEMP_22
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	HUM_02	Ox_03	TEMP_08	Ox_23	Ox_19	WS_02
	TEMP_03	TEMP_09	HUM_24	HUM_09	TEMP_21	Ox_11	Ox_24	TEMP_13	SPM_02	TEMP_15
	Ox_04	HUM_07	Ox_02	Ox_09	PM _{2.5} _01	Ox_21	HUM_06	HUM_05	SPM_01	TEMP_24
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	TEMP_08	Ox_23	Ox_03	Ox_05	Ox_04	Ox_18	TEMP_24
	TEMP_09	WS_01	WS_02	Ox_22	HUM_24	PM _{2.5} _01	Ox_19	HUM_05	Ox_06	TEMP_10
	HUM_06	TEMP_03	Ox_17	TEMP_19	TEMP_20	Ox_10	TEMP_21	Ox_11	TEMP_04	TEMP_18
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_05	TEMP_24	WS_01	Ox_17	TEMP_08	Ox_03
	Ox_22	TEMP_09	Ox_18	Ox_23	HUM_24	TEMP_07	TEMP_20	TEMP_18	TEMP_10	TEMP_22
	TEMP_19	PM _{2.5} _01	TEMP_21	Ox_06	HUM_05	WS_02	Ox_21	Ox_10	TEMP_02	HUM_02
高低上位20個_lag=1	TEMP_16	TEMP_09	SPM_02	TEMP_08	Ox_11	Ox_19	TEMP_13	TEMP_01	TEMP_14	TEMP_21
	TEMP_23	Ox_03	TEMP_02	Ox_01	WS_02	TEMP_03	Ox_23	HUM_24	TEMP_15	TEMP_05
	HUM_03	TEMP_04	HUM_02	Ox_24	HUM_09	HUM_01	WD_01	WS_01		
高低上位20個_lag=2	TEMP_24	Ox_05	Ox_22	TEMP_09	TEMP_08	Ox_19	Ox_06	Ox_18	TEMP_01	PM _{2.5} _01
	TEMP_21	TEMP_22	TEMP_23	Ox_03	TEMP_02	Ox_01	Ox_04	WS_02	TEMP_03	Ox_23
	HUM_24	TEMP_05	HUM_05	TEMP_04	Ox_02	HUM_02	TEMP_07	HUM_01	WD_01	TEMP_10
	WS_01									
高低上位20個_lag=3	TEMP_24	Ox_05	Ox_22	TEMP_09	TEMP_08	Ox_18	TEMP_01	TEMP_21	TEMP_22	TEMP_23
	Ox_17	Ox_03	TEMP_02	Ox_01	Ox_04	TEMP_03	Ox_23	HUM_24	TEMP_20	TEMP_05
	TEMP_06	TEMP_04	Ox_02	TEMP_07	HUM_01	TEMP_18	TEMP_10	WS_01		

表 14、15 にそれぞれ東青梅の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関してベンチマークが最も高い 80.714%であり、次点で上位 20 個の 80.443%であった。高濃度 RMSE に関してベンチマークで最も精度が良い 8.869 であり、上位 20 個は 9.505 であった。また、日別評価の調和平均に関して、ベンチマークで最も高い 82.667%で上位 20 個は 3 番目に高い 80.556%であった。時間値評価でベンチマークが最も優れているが、特徴量の数が少なく、精度がそれなりに担保されている上位 20 個が 1 時間後予測に高精度であると考えた。

2 時間後予測の時間値評価では、調和平均に関して上位 30 個に関して 70.758%で、次点でベンチマークの 70.313%であった。高濃度 RMSE はベンチマークの 11.850 で最も精度が良く、次点で上位 30 個の 12.232 であった。また、日別評価の調和平均に関して、ベンチマークが最も高い 73.846%であり、次点で上位 30 個の 70.270%であった。ベンチマークより上位 30 個の方が特徴量の数が少なく、ベンチマークと変わらない精度である。このことから

ら、2 時間後予測では上位 30 個が高精度であると考えた。

3 時間後予測の時間値評価では、調和平均に関して上位 30 個が最も高い 58.921%であり、次点でベンチマークの 58.559%であった。高濃度 RMSE では上位 30 個の 14.179 で最も精度が良く、次点で上位 20 個の 14.220 であった。また、日別評価の調和平均では全データが 61.538%であり、上位 30 個が 55.385%で 3 番目に高かった。上位 30 個に関して時間値、日別で精度が高いため 3 時間後予測に関して上位 30 個が高精度であると考え

表 14 東青梅における時間値評価

	再現率 (%)	適合率(%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	81.884	79.577	80.714	4.252	8.869
全データ_lag = 1	73.188	85.593	78.906	4.202	8.760
上位 10 個_lag = 1	73.913	83.607	78.462	4.207	9.562
上位 20 個_lag = 1	78.986	81.955	80.443	4.385	9.505
上位 30 個_lag = 1	67.391	85.321	75.304	4.177	9.278
高低上位 20 個_lag = 1	71.739	81.148	76.154	4.285	9.382
ベンチマーク_lag = 2	65.217	76.271	70.313	7.001	11.850
全データ_lag = 2	42.754	86.765	57.282	6.891	15.130
上位 10 個_lag = 2	44.928	86.111	59.048	7.718	15.398
上位 20 個_lag = 2	58.696	81.818	68.354	7.260	13.199
上位 30 個_lag = 2	71.014	70.504	70.758	7.405	12.232
高低上位 20 個_lag = 2	55.797	84.615	67.249	6.948	12.430
ベンチマーク_lag = 3	47.101	77.381	58.559	8.787	15.862
全データ_lag = 3	43.478	84.507	57.416	8.548	17.715
上位 10 個_lag = 3	39.855	82.090	53.659	10.034	19.346
上位 20 個_lag = 3	65.217	49.724	56.426	10.594	14.220
上位 30 個_lag = 3	51.449	68.932	58.921	9.075	14.179
高低上位 20 個_lag = 3	22.464	88.571	35.838	9.556	22.517

表 15 東青梅における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	86.111	79.487	82.667
全データ_lag = 1	77.778	87.500	82.353
上位 10 個_lag = 1	75.000	81.818	78.261
上位 20 個_lag = 1	80.556	80.556	80.556
上位 30 個_lag = 1	66.667	80.000	72.727
高低上位 20 個_lag = 1	72.222	78.788	75.362
ベンチマーク_lag = 2	66.667	82.759	73.846
全データ_lag = 2	38.889	77.778	51.852
上位 10 個_lag = 2	52.778	95.000	67.857
上位 20 個_lag = 2	58.333	80.769	67.742
上位 30 個_lag = 2	72.222	68.421	70.270
高低上位 20 個_lag = 2	55.556	83.333	66.667
ベンチマーク_lag = 3	47.222	77.273	58.621
全データ_lag = 3	44.444	100.000	61.538
上位 10 個_lag = 3	47.222	73.913	57.627
上位 20 個_lag = 3	58.333	44.681	50.602
上位 30 個_lag = 3	50.000	62.069	55.385
高低上位 20 個_lag = 3	27.778	90.909	42.553

4.5 所沢

表 16 に今回用いた所沢における特徴量の詳細を示した。後述する特徴量に関しては表 16 の特徴量と対応している。

表 16 所沢における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	Ox_04	Ox_07	Ox_02	TEMP_03	WS_02	TEMP_23
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_06	TEMP_03	Ox_04	TEMP_04	TEMP_21	TEMP_24
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_02	TEMP_03	Ox_05	TEMP_20	Ox_06	Ox_19	Ox_04
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	Ox_04	Ox_07	Ox_02	TEMP_03	WS_02	TEMP_23
	NMHC_08	Ox_24	Ox_21	HUM_02	TEMP_13	Ox_03	HUM_06	SPM_20	SPM_02	PM2.5_07
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_06	TEMP_03	Ox_04	TEMP_04	TEMP_21	TEMP_24
	WS_02	Ox_07	WS_01	TEMP_20	SPM_19	NMHC_07	Ox_18	HUM_24	Ox_05	Ox_14
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_02	TEMP_03	Ox_05	TEMP_20	Ox_06	Ox_19	Ox_04
	TEMP_04	Ox_18	WS_02	WS_01	TEMP_24	HUM_24	TEMP_19	TEMP_05	TEMP_06	TEMP_07
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	TEMP_02	Ox_04	Ox_07	Ox_02	TEMP_03	WS_02	TEMP_23
	NMHC_08	Ox_24	Ox_21	HUM_02	TEMP_13	Ox_03	HUM_06	SPM_20	SPM_02	PM2.5_07
	Ox_14	Ox_23	PM2.5_12	NO2_02	TEMP_22	NMHC_07	Ox_18	SPM_08	TEMP_05	Ox_19
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_06	TEMP_03	Ox_04	TEMP_04	TEMP_21	TEMP_24
	WS_02	Ox_07	WS_01	TEMP_20	SPM_19	NMHC_07	Ox_18	HUM_24	Ox_05	Ox_14
	Ox_03	Ox_20	Ox_19	HUM_06	Ox_24	SPM_07	PM2.5_01	SPM_18	SPM_08	HUM_05
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_02	TEMP_03	Ox_05	TEMP_20	Ox_06	Ox_19	Ox_04
	TEMP_04	Ox_18	WS_02	WS_01	TEMP_24	HUM_24	TEMP_19	TEMP_05	TEMP_06	TEMP_07
	Ox_14	SPM_18	SPM_07	Ox_17	HUM_05	TEMP_18	Ox_13	NMHC_06	PM2.5_01	TEMP_21
高低上位20個_lag=1	NO2_01	SPM_20	NMHC_08	SPM_02	TEMP_08	Ox_07	TEMP_13	TEMP_01	TEMP_22	TEMP_23
	Ox_03	NOx_01	TEMP_02	Ox_21	PM2.5_07	Ox_01	Ox_04	WS_02	NO2_02	TEMP_03
	NOx_02	TEMP_05	HUM_03	TEMP_19	TEMP_04	HUM_06	Ox_02	HUM_02	Ox_24	HUM_01
	WS_01									
高低上位20個_lag=2	NO2_01	TEMP_24	Ox_05	Ox_14	TEMP_08	Ox_07	Ox_06	Ox_18	TEMP_01	TEMP_21
	TEMP_02	SPM_19	Ox_01	Ox_04	WS_02	NO2_02	TEMP_03	HUM_24	NOx_02	TEMP_20
	TEMP_05	TEMP_06	TEMP_19	TEMP_04	Ox_02	NMHC_07	TEMP_07	HUM_01	WS_01	
高低上位20個_lag=3	TEMP_24	Ox_05	Ox_14	Ox_19	Ox_06	Ox_18	TEMP_01	TEMP_21	TEMP_22	Ox_03
	TEMP_02	Ox_01	Ox_04	WS_02	TEMP_03	HUM_24	TEMP_20	TEMP_05	TEMP_06	TEMP_19
	TEMP_04	Ox_02	TEMP_07	HUM_01	TEMP_18	WS_01				

表 17、18 にそれぞれ東青梅の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関して全データ 78.049%で最も高く、次点で上位 10 個の 77.419%であった。高濃度 RMSE では、高低上位 20 個の 10.061 で最も高く、次点で上位 10 個の 10.117 であった。また、日別評価の調和平均に関して全データの 85.000%であり、上位 10 個は 78.378%で 3 番目に高い値であった。時間別、日別ともに精度が高い上位 10 個が 1 時間後予測に高精度であると考えた。

2 時間後予測の時間値評価では、調和平均に関して高低上位 20 個の 66.418%で最も高く、次点で上位 10 個の 65.979%であった。高濃度 RMSE に関して上位 10 個に関して 13.625 で最も精度がよい。また、日別評価の調和平均に関しても上位 10 個の 70.130 で最も精度が良い。時間別、日別において精度がよい上位 10 個が 2 時間後予測に高精度であると考えた。

3 時間後予測の時間値評価では、調和平均に関して全データの 54.694%で最も高く、次点で高低上位 20 個の 51.883%であった。高濃度 RMSE に関して高低上位 20 個の 18.964 で最も精度がよい。また、日別評価の調和平均では上位 20 個の 55.882%で最も高く、高低上位

20 個は 48.387%と 3 番目に高い。高低上位 20 個に関して時間値、日別評価で精度がある程度担保されて特徴量が少ない。そのため、高低上位 20 個が 3 時間後予測に高精度であると考えた。

表 17 所沢における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	56.000	89.362	68.852	4.897	12.294
全データ_lag = 1	85.333	71.910	78.049	4.830	10.169
上位 10 個_lag = 1	72.000	83.721	77.419	4.521	10.117
上位 20 個_lag = 1	89.333	68.020	77.233	5.848	10.155
上位 30 個_lag = 1	71.333	81.061	75.887	4.764	10.160
高低上位 20 個_lag = 1	71.333	80.451	75.618	4.253	10.061
ベンチマーク_lag = 2	62.667	69.630	65.965	8.418	14.074
全データ_lag = 2	58.000	72.500	64.444	7.531	14.727
上位 10 個_lag = 2	64.000	68.085	65.979	7.561	13.625
上位 20 個_lag = 2	60.000	69.231	64.286	7.556	14.291
上位 30 個_lag = 2	51.333	65.812	57.678	7.764	16.510
高低上位 20 個_lag = 2	59.333	75.424	66.418	7.327	15.017
ベンチマーク_lag = 3	39.333	64.835	48.963	10.223	20.164
全データ_lag = 3	44.667	70.526	54.694	9.446	19.062
上位 10 個_lag = 3	40.667	69.318	51.261	9.851	20.067
上位 20 個_lag = 3	41.333	62.000	49.600	9.525	19.133
上位 30 個_lag = 3	31.333	64.384	42.152	9.974	22.340
高低上位 20 個_lag = 3	41.333	69.663	51.883	9.456	18.964

表 18 所沢における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	60.526	92.000	73.016
全データ_lag = 1	89.474	80.952	85.000
上位 10 個_lag = 1	76.316	80.556	78.378
上位 20 個_lag = 1	92.105	72.917	81.395
上位 30 個_lag = 1	76.316	80.556	78.378
高低上位 20 個_lag = 1	71.053	75.000	72.973
ベンチマーク_lag = 2	60.526	69.697	64.789
全データ_lag = 2	55.263	72.414	62.687
上位 10 個_lag = 2	71.053	69.231	70.130
上位 20 個_lag = 2	65.789	67.568	66.667
上位 30 個_lag = 2	65.789	71.429	68.493
高低上位 20 個_lag = 2	63.158	75.000	68.571
ベンチマーク_lag = 3	34.211	56.522	42.623
全データ_lag = 3	39.474	75.000	51.724
上位 10 個_lag = 3	36.842	56.000	44.444
上位 20 個_lag = 3	50.000	63.333	55.882
上位 30 個_lag = 3	39.474	55.556	46.154
高低上位 20 個_lag = 3	39.474	62.500	48.387

4.6 草加

表 19 に今回用いた草加における特徴量の詳細を示した。後述する特徴量に関しては表 19 の特徴量と対応している。

表 19 草加における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	WS_02	TEMP_06	TEMP_22	TEMP_02	WS_01	NO _z _02
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	WS_01	TEMP_22	TEMP_06	TEMP_21	WS_02	TEMP_05	SPM_04
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	TEMP_21	HUM_02	Ox_02	WS_01	TEMP_04	TEMP_20	TEMP_05
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	WS_02	TEMP_06	TEMP_22	TEMP_02	WS_01	NO _z _02
	Ox_02	TEMP_04	TEMP_03	TEMP_23	Ox_07	Ox_04	WD_04	TEMP_21	Ox_18	Ox_20
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	WS_01	TEMP_22	TEMP_06	TEMP_21	WS_02	TEMP_05	SPM_04
	Ox_02	WD_03	WD_04	Ox_06	TEMP_07	TEMP_20	HUM_02	WS_04	NO _z _15	Ox_17
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	TEMP_21	HUM_02	Ox_02	WS_01	TEMP_04	TEMP_20	TEMP_05
	TEMP_06	WS_02	TEMP_22	TEMP_24	WD_03	WS_03	WD_02	SPM_03	SPM_04	HUM_03
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	WS_02	TEMP_06	TEMP_22	TEMP_02	WS_01	NO _z _02
	Ox_02	TEMP_04	TEMP_03	TEMP_23	Ox_07	Ox_04	WD_04	TEMP_21	Ox_18	Ox_20
	SPM_05	NO _x _02	TEMP_24	WS_05	SPM_21	NO _x _15	NO _z _11	NO _z _17	NO _z _01	TEMP_07
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	WS_01	TEMP_22	TEMP_06	TEMP_21	WS_02	TEMP_05	SPM_04
	Ox_02	WD_03	WD_04	Ox_06	TEMP_07	TEMP_20	HUM_02	WS_04	NO _z _15	Ox_17
	SPM_20	NO _x _15	TEMP_24	NO _z _16	WS_05	Ox_07	NO _z _14	NO _x _16	HUM_03	NO _z _10
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	TEMP_21	HUM_02	Ox_02	WS_01	TEMP_04	TEMP_20	TEMP_05
	TEMP_06	WS_02	TEMP_22	TEMP_24	WD_03	WS_03	WD_02	SPM_03	SPM_04	HUM_03
	WD_04	WS_04	TEMP_19	Ox_16	Ox_05	NO _z _15	TEMP_07	Ox_18	Ox_17	NO _x _15
高低上位20個_lag=1	NO _z _01	WD_04	Ox_07	Ox_18	TEMP_01	TEMP_14	TEMP_21	TEMP_22	TEMP_23	Ox_03
	NO _x _01	TEMP_02	Ox_01	WS_02	Ox_04	NO _z _02	TEMP_03	NO _x _02	TEMP_06	TEMP_19
	TEMP_04	Ox_20	HUM_02	Ox_02	HUM_01	WS_01	TEMP_11			
高低上位20個_lag=2	NO _z _01	TEMP_24	WD_04	SPM_04	WD_03	Ox_06	TEMP_01	TEMP_21	TEMP_22	TEMP_23
	Ox_17	NO _x _01	TEMP_02	Ox_01	WS_02	TEMP_03	NO _z _15	WS_04	TEMP_20	TEMP_05
	TEMP_06	TEMP_04	Ox_02	HUM_02	TEMP_07	HUM_01	TEMP_10	WS_01	TEMP_11	
高低上位20個_lag=3	NO _z _01	TEMP_24	WD_03	SPM_04	TEMP_01	TEMP_21	TEMP_22	TEMP_23	TEMP_02	Ox_01
	WS_02	TEMP_17	WD_02	TEMP_03	TEMP_20	TEMP_05	TEMP_06	HUM_03	TEMP_04	SPM_03
	Ox_02	HUM_02	TEMP_07	WS_03	HUM_01	TEMP_10	WS_01	TEMP_11		

表 20、21 にそれぞれ東青梅の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関して全データの 74.126%で最も高く、次点で上位 20 個が 72.727%で高い。高濃度 RMSE に関して、上位 30 個の 11.399 で最も精度が良く、次点で上位 20 個の 11.564 であった。また、日別評価の調和平均に関して全データの 71.429%で最も高く、次点で上位 20 個の 69.565%であった。上位 20 個において精度が担保されており、特徴量の数が少ない。そのため、上位 20 個が 1 時間後予測に高精度であると考えた。

2 時間後予測の時間値評価では、調和平均に関して上位 20 個の 67.133%で最も高い。高濃度 RMSE に関して、全データの 15.383 で最も精度がよく、上位 20 個は 17.075 で 3 番目に精度がよい。また、日別評価の調和平均に関して上位 20 個の 58.537%で最も高かった。上位 20 個において精度が良く、特徴量の数が少ない。そのため、上位 20 個が 2 時間後予測に高精度であると考えた。

3 時間後予測の時間値評価では、調和平均に関して高低上位 20 個の 58.824%で最も高い。高濃度 RMSE に関して、高低上位 20 個が 18.636 で最も精度が良い。また、日別評価の調和平均に関して上位 30 個の 50.000%で最も高く、高低上位 20 個が 45.714%と 3 番目に高い。高低上位 20 個において時間別、日別において精度がよい。そのため、高低上位 20 個が 3 時間後予測に高精度であると考ええる。

表 20 草加における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	60.563	78.182	68.254	4.823	13.504
全データ_lag = 1	74.648	73.611	74.126	5.415	12.091
上位 10 個_lag = 1	57.746	83.673	68.333	4.471	14.123
上位 20 個_lag = 1	78.873	67.470	72.727	5.119	11.564
上位 30 個_lag = 1	76.056	67.500	71.523	4.547	11.399
高低上位 20 個_lag = 1	60.563	78.182	68.254	4.465	13.956
ベンチマーク_lag = 2	50.704	67.925	58.065	8.056	18.718
全データ_lag = 2	63.380	55.556	59.211	7.525	15.383
上位 10 個_lag = 2	61.972	65.672	63.768	7.518	17.748
上位 20 個_lag = 2	67.606	66.667	67.133	7.826	17.075
上位 30 個_lag = 2	64.789	58.228	61.333	8.072	17.547
高低上位 20 個_lag = 2	70.423	59.524	64.516	7.260	16.919
ベンチマーク_lag = 3	25.352	66.667	36.735	10.542	25.099
全データ_lag = 3	36.620	68.421	47.706	9.078	20.983
上位 10 個_lag = 3	45.070	62.745	52.459	9.087	20.443
上位 20 個_lag = 3	49.296	67.308	56.911	9.464	19.510
上位 30 個_lag = 3	33.803	82.759	48.000	10.145	25.499
高低上位 20 個_lag = 3	49.296	72.917	58.824	9.084	18.636

表 21 草加における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	59.091	76.471	66.667
全データ_lag = 1	68.182	75.000	71.429
上位 10 個_lag = 1	54.545	80.000	64.865
上位 20 個_lag = 1	72.727	66.667	69.565
上位 30 個_lag = 1	72.727	64.000	68.085
高低上位 20 個_lag = 1	54.545	66.667	60.000
ベンチマーク_lag = 2	36.364	47.059	41.026
全データ_lag = 2	50.000	45.833	47.826
上位 10 個_lag = 2	50.000	57.895	53.659
上位 20 個_lag = 2	54.545	63.158	58.537
上位 30 個_lag = 2	54.545	57.143	55.814
高低上位 20 個_lag = 2	54.545	54.545	54.545
ベンチマーク_lag = 3	31.818	63.636	42.424
全データ_lag = 3	31.818	58.333	41.176
上位 10 個_lag = 3	40.909	56.250	47.368
上位 20 個_lag = 3	36.364	57.143	44.444
上位 30 個_lag = 3	36.364	80.000	50.000
高低上位 20 個_lag = 3	36.364	61.538	45.714

4.7 多摩市愛宕

表 22 に今回用いた多摩市愛宕における特徴量の詳細を示した。後述する特徴量に関しては表 22 の特徴量と対応している。

表 22 多摩市愛宕における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	Ox_04	TEMP_03	SPM_01	TEMP_02	Ox_06	TEMP_04
上位10個_lag=2	Ox_01	TEMP_01	Ox_03	HUM_01	TEMP_04	TEMP_05	TEMP_03	Ox_04	SPM_01	Ox_17
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_03	TEMP_04	TEMP_03	Ox_17	TEMP_05	SPM_01	WS_01
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	Ox_04	TEMP_03	SPM_01	TEMP_02	Ox_06	TEMP_04
	Ox_23	TEMP_05	Ox_17	SPM_14	SPM_18	Ox_20	HUM_02	Ox_19	SPM_17	Ox_21
上位20個_lag=2	Ox_01	TEMP_01	Ox_03	HUM_01	TEMP_04	TEMP_05	TEMP_03	Ox_04	SPM_01	Ox_17
	Ox_18	SPM_14	Ox_16	TEMP_18	WS_01	Ox_20	Ox_22	SPM_17	Ox_19	TEMP_24
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_03	TEMP_04	TEMP_03	Ox_17	TEMP_05	SPM_01	WS_01
	TEMP_24	Ox_18	Ox_16	TEMP_17	SPM_14	TEMP_06	TEMP_18	Ox_21	HUM_06	Ox_19
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	Ox_04	TEMP_03	SPM_01	TEMP_02	Ox_06	TEMP_04
	Ox_23	TEMP_05	Ox_17	SPM_14	SPM_18	Ox_20	HUM_02	Ox_19	SPM_17	Ox_21
	Ox_09	TEMP_18	HUM_08	Ox_18	WS_02	TEMP_09	HUM_07	Ox_08	SPM_12	TEMP_24
上位30個_lag=2	Ox_01	TEMP_01	Ox_03	HUM_01	TEMP_04	TEMP_05	TEMP_03	Ox_04	SPM_01	Ox_17
	Ox_18	SPM_14	Ox_16	TEMP_18	WS_01	Ox_20	Ox_22	SPM_17	Ox_19	TEMP_24
	HUM_07	HUM_06	WS_02	Ox_21	TEMP_17	TEMP_08	SPM_11	SPM_15	Ox_08	Ox_05
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_03	TEMP_04	TEMP_03	Ox_17	TEMP_05	SPM_01	WS_01
	TEMP_24	Ox_18	Ox_16	TEMP_17	SPM_14	TEMP_06	TEMP_18	Ox_21	HUM_06	Ox_19
	Ox_04	Ox_20	TEMP_07	SPM_13	TEMP_20	Ox_15	Ox_09	TEMP_19	Ox_02	HUM_02
高低上位20個_lag=1	NO_01	TEMP_09	Ox_19	Ox_06	TEMP_01	SPM_01	TEMP_22	TEMP_23	Ox_03	Ox_17
	TEMP_02	Ox_21	Ox_01	Ox_04	SPM_18	NO_02	TEMP_03	Ox_23	NOx_02	TEMP_05
	TEMP_06	HUM_03	SPM_14	TEMP_04	Ox_20	HUM_02	SPM_17	HUM_01	TEMP_10	WS_01
高低上位20個_lag=2	TEMP_24	Ox_22	Ox_19	Ox_18	TEMP_01	SPM_01	TEMP_21	TEMP_22	TEMP_23	Ox_03
	Ox_17	TEMP_02	Ox_01	Ox_04	TEMP_03	TEMP_05	Ox_16	TEMP_06	HUM_03	SPM_14
	TEMP_04	Ox_20	Ox_02	HUM_02	SPM_17	HUM_01	TEMP_18	WS_01		
高低上位20個_lag=3	TEMP_24	Ox_15	Ox_19	Ox_18	TEMP_01	SPM_01	TEMP_21	TEMP_22	TEMP_23	Ox_03
	Ox_17	Ox_21	TEMP_02	Ox_01	TEMP_17	TEMP_03	TEMP_20	TEMP_05	TEMP_06	Ox_16
	TEMP_19	SPM_14	TEMP_04	HUM_06	Ox_02	TEMP_07	HUM_01	TEMP_18	WS_01	

表 23、24 にそれぞれ多摩市愛宕の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関して全体的に差が目立ちにくく、ほぼ同じ値を推移している。高濃度 RMSE に関して、上位 10 個で 8.089 と最も精度がよい。また、日別評価の調和平均に関して、全データの 65.455%と最も高く、次点でベンチマークの 65.306%、3 番目に上位 20 個の 64.151%であった。この時、調和平均に差は少なく高濃度 RMSE に関して上位 10 個が精度は良いが日別の調和平均で一番低い 59.016%である。そのため、高濃度 RMSE は他と比べ高いが日別の調和平均で 3 番目に高く、特徴量の数が少ない上位 20 個が 1 時間後予測に高精度であると考えた。

2 時間後予測の時間値評価では、調和平均に関してベンチマークが 71.111%で最も高く、次点で上位 20 個の 69.231%であった。高濃度 RMSE に関して上位 10 個が 10.976 で最も精度がよく、次点で上位 20 個の 12.437 であった。また、日別評価の調和平均に関して全データで最も高い 59.091%であった。日別評価で見た時に上位 20 個は 4 番目であるが、3 番目

である高低上位 20 個との差はほぼなく、時間値評価での精度がよい。さらに特徴量に関して高低上位 20 個よりも少ない。このことから、2 時間後予測に関して上位 20 個が高精度であると考えた。

3 時間後予測の時間値評価では、調和平均に関して上位 10 個が最も高い 66.986%であった。高濃度 RMSE に関しても、上位 10 個の 14.178 と最も精度がよい。日別評価の調和平均に関して、上位 10 個が最も高い 54.902%であった。時間値、日別評価で精度がよい上位 10 個が 3 時間後予測に高精度だと考えた。

表 23 多摩市愛宕における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	63.636	93.902	75.862	4.683	10.509
全データ_lag = 1	72.727	87.129	79.279	4.457	8.545
上位 10 個_lag = 1	76.033	79.310	77.637	4.325	8.089
上位 20 個_lag = 1	68.595	88.298	77.209	4.525	9.613
上位 30 個_lag = 1	84.298	70.833	76.981	4.823	9.168
高低上位 20 個_lag = 1	74.380	84.112	78.947	4.293	8.381
ベンチマーク_lag = 2	66.116	76.923	71.111	7.454	13.081
全データ_lag = 2	49.587	93.750	64.865	7.155	14.499
上位 10 個_lag = 2	72.727	65.672	69.020	7.847	10.976
上位 20 個_lag = 2	66.942	71.681	69.231	7.324	12.437
上位 30 個_lag = 2	58.678	78.889	67.299	7.521	14.000
高低上位 20 個_lag = 2	61.983	77.320	68.807	6.897	13.360
ベンチマーク_lag = 3	64.463	59.091	61.660	10.410	15.399
全データ_lag = 3	30.579	94.872	46.250	9.350	20.900
上位 10 個_lag = 3	57.851	79.545	66.986	9.414	14.178
上位 20 個_lag = 3	47.934	66.667	55.769	9.320	16.305
上位 30 個_lag = 3	46.281	87.500	60.541	9.206	17.976
高低上位 20 個_lag = 3	42.975	80.000	55.914	9.059	16.567

表 24 多摩市愛宕における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	53.333	84.211	65.306
全データ_lag = 1	60.000	72.000	65.455
上位 10 個_lag = 1	60.000	58.065	59.016
上位 20 個_lag = 1	56.667	73.913	64.151
上位 30 個_lag = 1	66.667	55.556	60.606
高低上位 20 個_lag = 1	56.667	62.963	59.649
ベンチマーク_lag = 2	53.333	61.538	57.143
全データ_lag = 2	43.333	92.857	59.091
上位 10 個_lag = 2	53.333	44.444	48.485
上位 20 個_lag = 2	53.333	59.259	56.140
上位 30 個_lag = 2	43.333	56.522	49.057
高低上位 20 個_lag = 2	50.000	65.217	56.604
ベンチマーク_lag = 3	43.333	39.394	41.270
全データ_lag = 3	30.000	100.000	46.154
上位 10 個_lag = 3	46.667	66.667	54.902
上位 20 個_lag = 3	40.000	60.000	48.000
上位 30 個_lag = 3	33.333	83.333	47.619
高低上位 20 個_lag = 3	36.667	64.706	46.809

4.8 世田谷

表 25 に今回用いた世田谷における特徴量の詳細を示した。後述する特徴量に関しては表 25 の特徴量と対応している。

表 25 世田谷における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	Ox_05	Ox_06	Ox_23	TEMP_03	TEMP_04	NO ₂ _01
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	Ox_05	Ox_06	Ox_23	TEMP_03	TEMP_04	NO ₂ _01
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_03	Ox_05	Ox_06	Ox_23	TEMP_03	TEMP_04	NO ₂ _01
高低上位20個_lag=1	NO ₂ _01	TEMP_16	TEMP_12	Ox_05	Ox_15	SPM_22	TEMP_08	Ox_06	TEMP_01	SPM_01
	PM _{2.5} _01	TEMP_22	TEMP_23	Ox_03	NO _x _01	TEMP_02	Ox_01	NO ₂ _02	TEMP_03	Ox_23
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_05	TEMP_04	Ox_03	TEMP_03	Ox_19	Ox_20
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_05	TEMP_04	Ox_03	TEMP_03	Ox_19	Ox_20
	Ox_23	SPM_21	TEMP_22	Ox_22	WS_02	Ox_06	SPM_22	TEMP_21	SPM_01	Ox_21
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_05	TEMP_04	Ox_03	TEMP_03	Ox_19	Ox_20
	Ox_23	SPM_21	TEMP_22	Ox_22	WS_02	Ox_06	SPM_22	TEMP_21	SPM_01	Ox_21
	TEMP_19	WD_03	PM _{2.5} _01	Ox_14	NO ₂ _01	Ox_24	WS_01	Ox_18	TEMP_08	TEMP_18
高低上位20個_lag=2	NO ₂ _01	TEMP_16	TEMP_24	Ox_05	Ox_22	SPM_22	Ox_19	Ox_06	TEMP_01	SPM_01
	SPM_21	TEMP_21	TEMP_22	TEMP_23	Ox_03	NO _x _01	TEMP_02	Ox_21	Ox_01	Ox_04
	WS_02	TEMP_03	Ox_23	TEMP_15	NO _x _02	TEMP_05	TEMP_04	Ox_20	Ox_02	HUM_01
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_03	Ox_05	Ox_18	TEMP_03	TEMP_21	TEMP_22
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_03	Ox_05	Ox_18	TEMP_03	TEMP_21	TEMP_22
	Ox_19	TEMP_20	Ox_21	Ox_23	SPM_21	TEMP_04	Ox_22	WS_01	Ox_20	SPM_20
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_04	Ox_03	Ox_05	Ox_18	TEMP_03	TEMP_21	TEMP_22
	Ox_19	TEMP_20	Ox_21	Ox_23	SPM_21	TEMP_04	Ox_22	WS_01	Ox_20	SPM_20
	SPM_01	TEMP_19	TEMP_18	TEMP_24	Ox_17	WD_02	WS_02	TEMP_07	WD_03	PM _{2.5} _01
高低上位20個_lag=3	NO ₂ _01	TEMP_24	Ox_05	Ox_22	SPM_20	Ox_19	Ox_18	TEMP_01	SPM_21	TEMP_21
	TEMP_22	Ox_03	NO _x _01	TEMP_02	Ox_21	Ox_01	Ox_04	TEMP_03	Ox_23	TEMP_15
	TEMP_20	TEMP_05	TEMP_06	TEMP_19	TEMP_04	Ox_20	Ox_02	HUM_01	TEMP_10	WS_01

表 26、27 にそれぞれ世田谷の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関して高低上位 20 個で最も高い 79.646% であり、次点で上位 20 個の 79.612% であった。高濃度 RMSE に関して高低上位 20 個で最も精度がよい 13.103 であり、上位 20 個は 14.605 であった。また、日別評価の調和平均では全データで 81.818% であり、3 番目に上位 20 個の 76.596% であった。時間値評価で高低上位 20 個と上位 20 個では精度にあまり精度がなく、上位 20 個の方が特徴量の個数が少ない。このことから 1 時間後予測に高精度なのは上位 20 個であると考えた。

2 時間後予測の時間値評価では、調和平均に関して高低上位 20 個で最も高い 61.616% であり、次点で上位 30 個の 60.793% であった。高濃度 RMSE に関して調和平均と同じで高低上位 20 個が最も精度が高く、次点で上位 30 個が精度はよかった。また、日別評価の調和平均では上位 20 個が 58.824% であり、高低上位 20 個は上から 4 番目ではあるが 48.980% であった。日別評価の調和平均に関して、高低上位 20 個が他と比べ低い制度ではあるが時間値

で見たときに一番精度がよい。そのため、2 時間後予測に関して高低上位 20 個が高精度であると考えた。

3 時間後予測の時間値評価では、調和平均に関して上位 30 個が最も精度がよい 53.714% であった。高濃度 RMSE に関して全データの 24.626 で最も精度がよく、上位 30 個は 3 番目に高い 25.235 であった。また日別評価の調和平均では、上位 30 個が 60.000% で最も高い。このことから、上位 30 個が 3 時間後予測に高精度であると考えた。

表 26 世田谷における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	60.550	90.411	72.527	5.268	15.275
全データ_lag = 1	62.385	89.474	73.514	4.765	14.979
上位 10 個_lag = 1	86.239	72.308	78.661	5.188	13.980
上位 20 個_lag = 1	75.229	84.536	79.612	4.876	14.605
上位 30 個_lag = 1	71.560	78.000	74.641	5.014	14.208
高低上位 20 個_lag = 1	82.569	76.923	79.646	4.830	13.103
ベンチマーク_lag = 2	35.780	79.592	49.367	8.091	24.658
全データ_lag = 2	46.789	73.913	57.303	7.444	21.507
上位 10 個_lag = 2	42.202	76.667	54.438	7.897	24.301
上位 20 個_lag = 2	33.945	82.222	48.052	8.119	26.301
上位 30 個_lag = 2	63.303	58.475	60.793	8.265	20.726
高低上位 20 個_lag = 2	55.963	68.539	61.616	7.637	20.532
ベンチマーク_lag = 3	43.119	62.667	51.087	9.857	26.071
全データ_lag = 3	41.284	67.164	51.136	9.353	24.626
上位 10 個_lag = 3	39.450	66.154	49.425	9.664	27.100
上位 20 個_lag = 3	44.954	62.025	52.128	9.968	24.858
上位 30 個_lag = 3	43.119	71.212	53.714	9.463	25.235
高低上位 20 個_lag = 3	30.275	73.333	42.857	9.785	28.223

表 27 世田谷における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	75.000	85.714	80.000
全データ_lag = 1	75.000	90.000	81.818
上位 10 個_lag = 1	83.333	55.556	66.667
上位 20 個_lag = 1	75.000	78.261	76.596
上位 30 個_lag = 1	75.000	69.231	72.000
高低上位 20 個_lag = 1	83.333	62.500	71.429
ベンチマーク_lag = 2	41.667	76.923	54.054
全データ_lag = 2	41.667	58.824	48.780
上位 10 個_lag = 2	41.667	71.429	52.632
上位 20 個_lag = 2	41.667	100.000	58.824
上位 30 個_lag = 2	50.000	38.710	43.636
高低上位 20 個_lag = 2	50.000	48.000	48.980
ベンチマーク_lag = 3	45.833	57.895	51.163
全データ_lag = 3	54.167	61.905	57.778
上位 10 個_lag = 3	41.667	58.824	48.780
上位 20 個_lag = 3	45.833	50.000	47.826
上位 30 個_lag = 3	50.000	75.000	60.000
高低上位 20 個_lag = 3	41.667	66.667	51.282

4.9 南葛西

表 28 に今回用いた南葛西における特徴量の詳細を示した。後述する特徴量に関しては表 28 の特徴量と対応している。

表 28 南葛西における特徴量詳細

	特徴量									
上位10個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_23	Ox_24	TEMP_03	Ox_04	Ox_22	NOx_02
上位10個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_23	Ox_22	Ox_24	TEMP_06	TEMP_07	Ox_06
上位10個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_23	Ox_22	Ox_21	TEMP_05	TEMP_06	Ox_04	Ox_05
上位20個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_23	Ox_24	TEMP_03	Ox_04	Ox_22	NOx_02
	PM _{2.5} _01	NO ₂ _02	SPM_02	TEMP_23	TEMP_07	PM _{2.5} _09	TEMP_02	Ox_07	Ox_13	TEMP_08
上位20個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_23	Ox_22	Ox_24	TEMP_06	TEMP_07	Ox_06
	Ox_05	Ox_04	WS_02	Ox_13	NOx_02	Ox_21	TEMP_05	WD_05	NO ₂ _15	NO ₂ _14
上位20個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_23	Ox_22	Ox_21	TEMP_05	TEMP_06	Ox_04	Ox_05
	TEMP_04	NOx_01	TEMP_07	Ox_20	WS_02	TEMP_21	TEMP_22	Ox_02	Ox_24	NO ₂ _14
上位30個_lag=1	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_23	Ox_24	TEMP_03	Ox_04	Ox_22	NOx_02
	PM _{2.5} _01	NO ₂ _02	SPM_02	TEMP_23	TEMP_07	PM _{2.5} _09	TEMP_02	Ox_07	Ox_13	TEMP_08
	NMHC_21	WS_02	SPM_16	NO ₂ _15	Ox_19	PM _{2.5} _11	TEMP_11	NO ₂ _16	Ox_06	HUM_02
上位30個_lag=2	Ox_01	TEMP_01	HUM_01	Ox_02	Ox_23	Ox_22	Ox_24	TEMP_06	TEMP_07	Ox_06
	Ox_05	Ox_04	WS_02	Ox_13	NOx_02	Ox_21	TEMP_05	WD_05	NO ₂ _15	NO ₂ _14
	TEMP_22	NO ₂ _02	NOx_01	NMHC_20	Ox_03	TEMP_03	Ox_19	TEMP_04	TEMP_23	WD_06
上位30個_lag=3	Ox_01	TEMP_01	HUM_01	Ox_23	Ox_22	Ox_21	TEMP_05	TEMP_06	Ox_04	Ox_05
	TEMP_04	NOx_01	TEMP_07	Ox_20	WS_02	TEMP_21	TEMP_22	Ox_02	Ox_24	NO ₂ _14
	Ox_12	NO ₂ _13	WD_04	WD_05	NO_01	NO ₂ _15	WD_06	Ox_09	WD_07	TEMP_02
高低上位20個_lag=1	NO ₂ _01	Ox_22	Ox_13	SPM_02	TEMP_08	Ox_07	TEMP_01	PM _{2.5} _01	TEMP_22	TEMP_23
	NOx_01	TEMP_02	PM _{2.5} _09	Ox_01	Ox_04	NO ₂ _02	TEMP_03	Ox_23	NOx_02	TEMP_05
	TEMP_06	TEMP_19	TEMP_04	Ox_02	Ox_24	TEMP_07	HUM_01	TEMP_18	WS_01	TEMP_11
高低上位20個_lag=2	NO ₂ _01	NO ₂ _14	Ox_05	Ox_22	Ox_13	Ox_06	TEMP_01	TEMP_21	TEMP_22	TEMP_23
	TEMP_02	Ox_21	Ox_01	Ox_04	WS_02	TEMP_17	WD_05	TEMP_03	Ox_23	NO ₂ _15
	NOx_02	TEMP_05	TEMP_06	TEMP_04	Ox_02	Ox_24	TEMP_07	HUM_01	NO_01	TEMP_18
	TEMP_10									
高低上位20個_lag=3	NO ₂ _14	TEMP_16	Ox_05	Ox_22	TEMP_01	TEMP_21	TEMP_22	TEMP_23	NOx_01	Ox_21
	Ox_01	Ox_04	WS_02	TEMP_17	TEMP_03	Ox_23	TEMP_15	TEMP_20	TEMP_05	TEMP_06
	TEMP_04	Ox_20	Ox_02	Ox_24	TEMP_07	HUM_01	NO_01	TEMP_18	TEMP_10	WS_01

表 29、30 にそれぞれ世田谷の時間別評価、日別評価を示した。

1 時間後予測の時間値評価では、調和平均に関して上位 20 個の 73.381%で最も高い値であった。高濃度 RMSE に関して 12.297 で最も精度がよい。また、日別評価の調和平均では 79.070%であった。このように時間値別、日別ともに精度がよい上位 20 個が 1 時間後予測に高精度であると考えた。

2 時間後予測の時間値評価では、調和平均に関して上位 10 個の 55.556%で最も高い値であった。高濃度 RMSE に関して 23.716 で最も精度がよかった。また、日別評価の調和平均では全データの 60.606%で最も高く、次点でベンチマーク、上位 10 個の 58.824%であった。時間値別で最も精度がよく、日別でも精度がよい中で最も特徴量が少ない。そのため、上位 10 個が 2 時間後予測に高精度であると考えた。

3 時間後予測の時間値評価では、調和平均に関して上位 30 個の 45.361%で最も高い値であった。高濃度 RMSE に関して、上位 10 個の 30.298 で最も精度がよく、次点で上位 30 個

の 30.852 で精度がよい。日別評価での調和平均に関して上位 30 個の 53.333%であった。このように時間値別、日別での精度がよい上位 30 個が 3 時間後予測に高精度であると考えた。

表 29 南葛西における時間値評価

	再現率 (%)	適合率 (%)	調和平均 (%)	RMSE	高濃度 RMSE
ベンチマーク_lag = 1	66.667	77.193	71.545	4.944	14.099
全データ_lag = 1	51.515	82.927	63.551	5.933	16.329
上位 10 個_lag = 1	53.030	89.744	66.667	5.410	17.301
上位 20 個_lag = 1	77.273	69.863	73.381	5.215	12.297
上位 30 個_lag = 1	57.576	80.851	67.257	4.851	15.754
下位上位 20 個_lag = 1	62.121	89.130	73.214	4.679	15.505
ベンチマーク_lag = 2	37.879	71.429	49.505	7.902	24.939
全データ_lag = 2	36.364	75.000	48.980	7.459	23.024
上位 10 個_lag = 2	45.455	71.429	55.556	7.747	23.716
上位 20 個_lag = 2	39.394	63.415	48.598	8.189	23.433
上位 30 個_lag = 2	40.909	72.973	52.427	7.902	23.893
下位上位 20 個_lag = 2	31.818	80.769	45.652	7.992	26.184
ベンチマーク_lag = 3	27.273	62.069	37.895	9.965	31.154
全データ_lag = 3	21.212	77.778	33.333	9.734	32.953
上位 10 個_lag = 3	27.273	46.154	34.286	10.045	30.298
上位 20 個_lag = 3	24.242	72.727	36.364	9.864	31.175
上位 30 個_lag = 3	33.333	70.968	45.361	10.291	30.852
下位上位 20 個_lag = 3	19.697	86.667	32.099	10.047	34.571

表 30 南葛西における日別評価

	再現率(%)	適合率(%)	調和平均(%)
ベンチマーク_lag = 1	71.429	78.947	75.000
全データ_lag = 1	61.905	100.000	76.471
上位 10 個_lag = 1	61.905	100.000	76.471
上位 20 個_lag = 1	80.952	77.273	79.070
上位 30 個_lag = 1	71.429	83.333	76.923
高低上位 20 個_lag = 1	71.429	100.000	83.333
ベンチマーク_lag = 2	47.619	76.923	58.824
全データ_lag = 2	47.619	83.333	60.606
上位 10 個_lag = 2	47.619	76.923	58.824
上位 20 個_lag = 2	42.857	69.231	52.941
上位 30 個_lag = 2	42.857	75.000	54.545
高低上位 20 個_lag = 2	38.095	100.000	55.172
ベンチマーク_lag = 3	33.333	70.000	45.161
全データ_lag = 3	19.048	100.000	32.000
上位 10 個_lag = 3	38.095	50.000	43.243
上位 20 個_lag = 3	28.571	85.714	42.857
上位 30 個_lag = 3	38.095	88.889	53.333
高低上位 20 個_lag = 3	23.810	100.000	38.462

5. 考察

5.1 各地点の最適な特徴量

各地点で n 時間後予測に高精度であった特徴量を表 31 に示した。

表 31 n 時間後予測に高精度な特徴量

	lag_1	lag_2	lag_3
東秩父	上位 20 個	上位 20 個	高低上位 20 個
鴻巣	上位 20 個	上位 10 個	上位 10 個
幸手	上位 10 個	上位 10 個	上位 20 個
東青梅	上位 20 個	上位 30 個	上位 30 個
所沢	上位 10 個	上位 10 個	高低上位 20 個
草加	上位 20 個	上位 20 個	高低上位 20 個
多摩市愛宕	上位 20 個	上位 20 個	上位 10 個
世田谷	上位 20 個	高低上位 20 個	上位 30 個
南葛西	上位 20 個	上位 10 個	上位 30 個

1 時間後予測に関して幸手と所沢以外では、上位 20 個が高精度な特徴量あった。2～3 時間後予測以降に関しては地点ごとに様々であり、規則性は見受けられない。また、各地点で 1～3 時間後予測に高精度なものは時間ごとによって変わっておりこちらも規則性などは見受けられなかった。特徴量を時間ごとに変えるのであれば、現場で用いる際に 3 時間分のモデルを作成しなくてはならなくなる。これでは非効率的であり、使用するモデルを間違える可能性があり、実用的とは言えないと考える。そのため、本実験の方法では特徴量を一般化することはできないであろう。

東京として東青梅、多摩市愛宕、世田谷、南葛西とし、埼玉として草加、所沢、東秩父、鴻巣、幸手でグループ分けを行い重複する特徴量をそれぞれ表 32、33 に示した。

表 32 東京における重複した特徴量

	特徴量					
lag = 1	Ox_01	TEMP_01	HUM_01	TEMP_03	TEMP_02	Ox_23
lag = 2	Ox_01	TEMP_01	HUM_01	Ox_22		
lag = 3	Ox_01	TEMP_01	HUM_01			
all	Ox_01	TEMP_01	HUM_01			

表 33 埼玉における重複した特徴量

	特徴量			
lag = 1	Ox_01	TEMP_01	HUM_01	
lag = 2	Ox_01	TEMP_01	HUM_01	
lag = 3	Ox_01	TEMP_01	HUM_01	TEMP_24
all	Ox_01	TEMP_01	HUM_01	

東京では時間がたつにつれて重複している特徴量の個数は減少していることが確認できた。2 時間後予測では、1 時間後予測にて重複していた特徴量である 2、3 時間前の気温、23 時間前の Ox が消え、22 時間前の Ox が重複した特徴量として出てきた。これらの特徴量はその他の時間帯で重複して出てくることはなかった。ただ、1 時間前の Ox、1 時間前の湿度、1 時間前の気温はすべての時間で見られた。そのため東京では 1 時間前の Ox、1 時間前の湿度、1 時間前の気温が必須の特徴量である可能性が考えられる。

埼玉では 3 時間後予測にのみ 24 時間前の気温が重複していた特徴量として出ていた。これは 1 時間前の気温と 24 時間前の気温の挙動が似ているために偶発的に出てきたものではないかと考える。また、1 時間前の Ox、1 時間前の湿度、1 時間前の気温はすべての時間で見られた。そのため埼玉では 1 時間前の Ox、1 時間前の湿度、1 時間前の気温が必須の特徴量である可能性が考えられる。

東京、埼玉の重複している特徴量をまとめたものを表 33 に示した。

表 34 重複していた特徴量

	特徴量		
lag = 1	Ox_01	HUM_01	TEMP_01
lag = 2	Ox_01	HUM_01	TEMP_01
lag = 3	Ox_01	HUM_01	TEMP_01
all	Ox_01	HUM_01	TEMP_01

表 34 を見ると全てにおいて 1 時間前の Ox、1 時間前の湿度、1 時間前の気温が使われていた。これらの事から、高濃度の Ox を予測するために必要になる特徴量は 1 時間前の Ox、

1 時間前の湿度、1 時間前の気温ではないかと考える。また、高濃度の Ox を予測するにあたって n 時間後及び地点によって最適な特徴量は違うと考えられる。そのため、地点と n 時間予測に合わせた特徴量の剪定を行うことが重要だと考える。

今回のように全データを学習させた後に重要度を算出、その後需要度が高い特徴量を抽出することは、ベンチマークよりも高精度なモデルが作成できた。そのため、手当たり次第に特徴量を抜き差しするよりも有効な方法であったといえるだろう。すなわち 3.5 で立てた「探索を行うにあたって、高濃度域で特徴量需要度が高いものが最適な特徴量になる」という仮説を完全に支持するものではないが、本研究の結果はその可能性を示している。

6. まとめ

本研究では高濃度を高精度に予測できる特徴量の探索を行った。その結果、 n 時間後予測や地点によって最適な特徴量は違っていた。その中で、重複していた特徴量は 1 時間前の Ox 、1 時間前の湿度、1 時間前の気温であり、これらが高濃度域の Ox を予測するうえで欠かせない特徴量であることが示唆された。また、「探索を行うにあたって、高濃度域で特徴量需要度が高いものが最適な特徴量になる」という仮説を完全に支持するものではないが、本研究の結果はその可能性を示している。

7. 謝辞

本研究を行うにあたって指導教官である櫻井先生には、研究の方針を考える際や行き詰った際などにたくさんのご助言をいただきました。また、本研究室の院生である細越さん、入澤さん、高井さんにご助言やデータの抽出など様々な場面でご協力をしていただき、同期には違った視点からの意見をいただき新たな知見を得ることができました。ここで上げさせていただいた皆様のおかげで本研究を行うことが出来ましたことこの場をお借りして感謝申し上げます。

8. 参考文献

- 1) . . . 環境省（令和 5 年）：“令和 5 年光化学大気汚染の概要－注意報等発令状況、被害届出状況－”，環境省
- 2) . . . 細越英彰(2022)：“ディープラーニングを用いた光化学オキシダント濃度の短期予測”，明星大学 理工学部 総合理工学科 環境科学系，令和 4 年度卒業論文
- 3) . . . 細越英彰(2024)：“ディープラーニングを用いた光化学オキシダント濃度の短期予測の有効性の検討 明星大学大学院 理工学研究科 環境システム学専攻 博士前期課程，令和 6 年度修士論文
- 4) . . . 星野 智恵子,吉野 裕紀：計測データの欠損を考慮した重機の経路最適化
- 5) . . . 川越雄介(2021/4/14)：SHAP を用いて機械学習モデルを説明する Data Robot