**単一地点におけるDNNを用いたOx短期予測の最適パラメータ探索**

21T7-008　今給黎樹

指導教員　櫻井達也

**1研究背景と目的**

　大気汚染物質の一つである光化学オキシダント(Ox)ハ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)を有意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。

**2研究方法**

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。この時の特徴量の有意性を評価させた。

**2.1入力データの選定**

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）

**2.2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。

グラフ, マップ

自動的に生成された説明

図2　対象地点

**2.3評価方法**

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、高濃度を有意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予測値と実測とのRMSE(平均二乗平方根誤差)値が低いこと、高濃度時のRMSEが低いこと、といった2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。

また、再現率、適合率、調和平均を算出し評価を行った。再現率は実測で高濃度を観測した日に予測で高濃度を超えた日の割合、適合率は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。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有意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ければ優れた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。

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結果と考察**

各地点でn時間後予測に有意であった特徴量を表1に示した。

表31　n時間後予測に有意な特徴量

テーブル

自動的に生成された説明

　表1でもちいた特徴量においてn時間後の予測に使われた特徴量で重複していた特徴量を表2に示した。

表2　重複していた特徴量

テーブル

自動的に生成された説明

表2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。これらの特徴量はOxを予測するうえで必ず入れなければならない特徴量であったといえるだろう。

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時間前の気温だといえる。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であろう。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中手することは、ベンチマークよりもよい結果が出ていた。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。すなわち最初に立てた「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。

**4今後の展望**

今後、特徴量の選択パターンを増やすことで最適な特徴量の規則性が見つけられと考えた。そのため多くの試行を試す必要があると考える。

**5参考文献**

[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