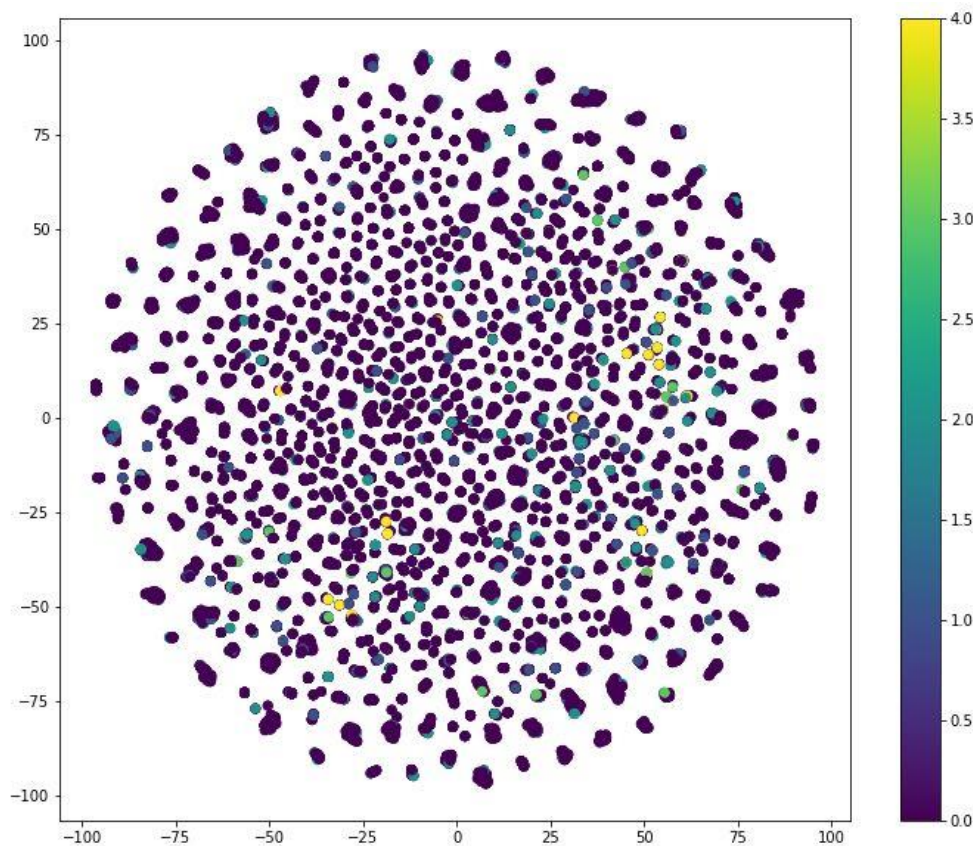


AI 霧予測チャレンジ



フルタイムクラス

加藤芳樹

私のバックグラウンド

- 民間気象会社で気象予報士として約14年勤務
- LCCで運航管理の仕事を約1年経験

データサイエンスを学ぶモチベーション

- 気象解析・予測そのものに応用
- 気象ビジネスに役立てたい

飛行機のフライトに影響する気象現象の予測をしてみたい

乱気流・雷・雪も影響が大きいですが、今回は霧をチョイス

【霧がフライトに与える影響】

- ◆ 視界不良で離陸できず、スケジュールの遅延が発生（次の便にも影響）
- ◆ 視界不良で着陸できない
 - 上空で霧が薄まるまで待機すると、到着遅延・燃料消費量が増える
 - 他の空港へダイバートすると、スケジュールの大幅遅延、飛行計画の再提出、燃料補給、乗客のケア、乗務員の勤務時間制限などのリスクとコストが発生

AI霧予測にチャレンジ

そこで...

機械学習を用いた霧の発生予測にチャレンジ！

- できるだけ多くのアルゴリズムを試す（スキルアップを兼ねて）
- 気象庁はじめ気象分野の人がやってないアルゴリズムを使う
- 予測地点は日本の主要な空港で、霧の発生頻度が比較的多く、データも入手できる**成田空港**とする



濃霧の成田空港に着陸する飛行機(YouTubeより引用 - oyazi49様)

★目標★

- 気象の知識とデータサイエンスの技術を駆使して、気象庁と同等以上の精度を出す
- 民間気象会社時代に自ら開発した視程予測式の精度を上回る

霧について

- ✓ 霧とは、大気が飽和して水蒸気が凝結し、無数の小さな水滴となって地表近くに浮かんでいる状態
- ✓ 視程が5000m以下の場合に**もや**、1000m未満の状態を**霧**と呼ぶ
- ◆ 霧も成因によっていくつか種類があるが、メジャーなのは放射霧
- ◆ 夜間に晴れていて放射冷却により地表付近の気温が下がって飽和し、朝にかけて霧が発生する

放射霧のポイント

1. Moisture（十分な水蒸気）
2. Cooling（夜間の放射冷却）
3. Calm Wind（風が弱い）

1. and 2.は湿度の上昇を、
3.は大気が安定で穏やかであることを意味する



市街地の霧(兵庫県三田市) - Wikipediaより引用

実験方法と使用するデータ

【分析方法】

- [1] 霧の発生の有無を予測する（二値問題とする）
- [2] 霧の深さの予測をする（視程カテゴリ or 連続量）

- ・ 2014～2016年のデータで学習する
（ Train : Validation = 7 : 3 ）
- ・ 学習したモデルで2017年の予測テストをする

【データ】

■ 空港観測データ(METAR) : ターゲット変数

- ・ 2014～2016年のデータが手元にある
- ・ 2017年のデータはWEB入手

■ 気象庁の数値予報GPVデータ(MSM) : 特徴量

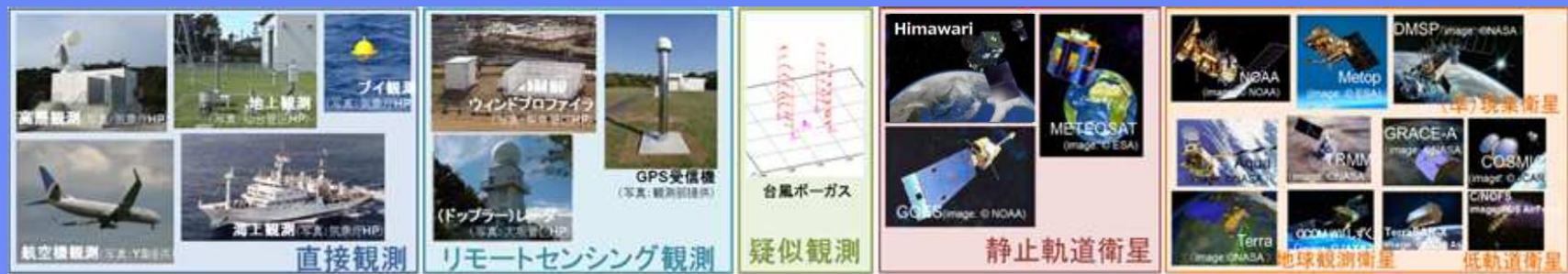
- ・ 京都大学の生存圏データベースからダウンロード
- ・ 成田空港の地点データを切り出す（経験あり）
- ・ 6時間ごとに更新する数値予報モデルのFT=6～11（6コマ）
を集めることで、1日24時間連続のデータを作る（4年分）

視程カテゴリ	
Category	Range
0	5000m以上
1	3200～5000m
2	1600～3200m
3	800～1600m
4	800m未満

※FTとは数値予報の初期時刻から何時間後の未来の予測か、
ということを示す数字（FT=6なら6時間後の予測）

天気予報ができるまで

気象観測

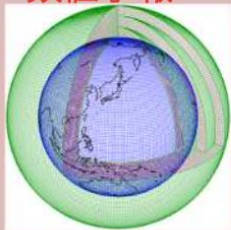


数値予報

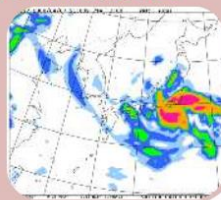


スーパーコンピュータ

数値予報



数値予報を使った応用処理



天気予報/防災情報



予報官

予報
警報
情報

作成

民間気象会社

関係機関
報道機関



国民

この部分に相当する霧予測モデルを作る

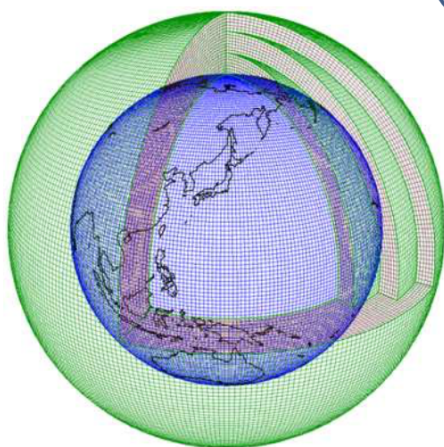


気象庁 Japan Meteorological Agency

※「気象庁における機械学習の利用」より抜粋

数値予報を使った応用処理とは？

数値予報



計算結果は未来の大気状態に対応する様々な数値の羅列（1億以上の格子点：ビックデータ）

機械学習
ニューラルネット
カルマンフィルタ
ロジスティック回帰
...

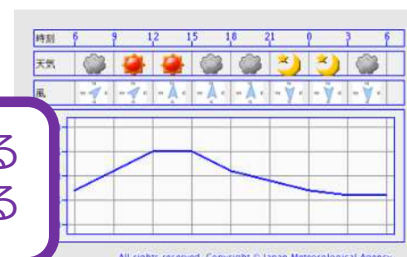
翻訳

この部分に相当する
霧予測モデルを作る

予報をガイドすることから（予報）ガイダンスと呼ばれる

天気予報・防災情報に必要な情報

晴・曇り・雨 ☀️
最高気温27℃ 🌡️
発雷確率30% ... ⚡️



初期仮説

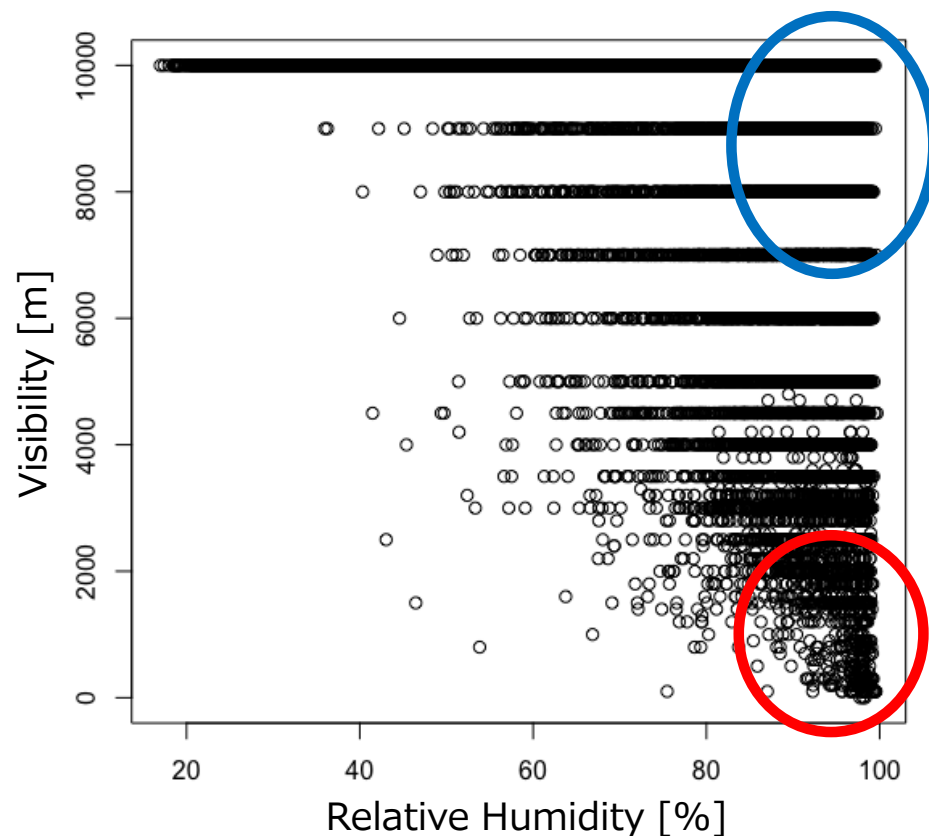
- 基本的には湿度が100%近くなると霧は発生する
- しかし実際には100%近くとも霧が発生していない場合が多く、湿度だけで予想すると空振りが多くなってしまう
- 他の気象要素を含めることで、予測精度アップを図る

[思い出してください]

放射霧のポイント

1. Moisture (十分な水蒸気)
2. Cooling (夜間の放射冷却)
3. Calm Wind (風が弱い)

1. and 2.は湿度の上昇を、
3.は大気安定で穏やかであることを意味する



特徴量の選択と作成

- 基本的な気象要素

- [地上の要素]

- ・ 湿度・気圧・風向・風速・気温・露点温度・降水量・雲量

- [上空の要素]

- ・ 1000, 975, 950hPa高度：風速・鉛直速度
 - ・ 1000, 975, 950, 850, 700, 500, 300hPa高度：湿度

- METAR観測から作った特徴量

- ・ 前24時間に雨が降ったかどうか (0 or 1)

- ドメイン知識をもとに導入した特徴量

- ・ 風の鉛直シア（風ベクトルの差の大きさ）
 - ・ 大気の安定度（気温の鉛直勾配）
 - ・ 前線霧パラメータ（気温の鉛直勾配×降水量）
 - ・ 時間帯（1日を8つの時間帯に分割）
 - ・ 季節（月単位）
 - ・ 前1時間の変化量（気温、露点温度、地上気圧）

評価方法について

【気象の分野でよく使われる評価指標】

→ 霧は極度の不均衡データになっている

■ **スレツスコア(TS) = $D / (B + C + D)$**

このうち適当に予測しても当たってしまう

割合を考慮したものを**ETS**と呼ぶ

● **バイアススコア(BI) = $(C + D) / (B + D)$**

空振り・見逃しのバランスを見る

		予報	
		0 (なし)	1 (あり)
実況	0 (なし)	A	C
	1 (あり)	B	D

【ベンチマーク】

気象庁の視程ガイダンスの精度：**成田空港はETS = 0.08 (± 0.02)**

ここで霧かどうかの閾値は **1600m** となっている

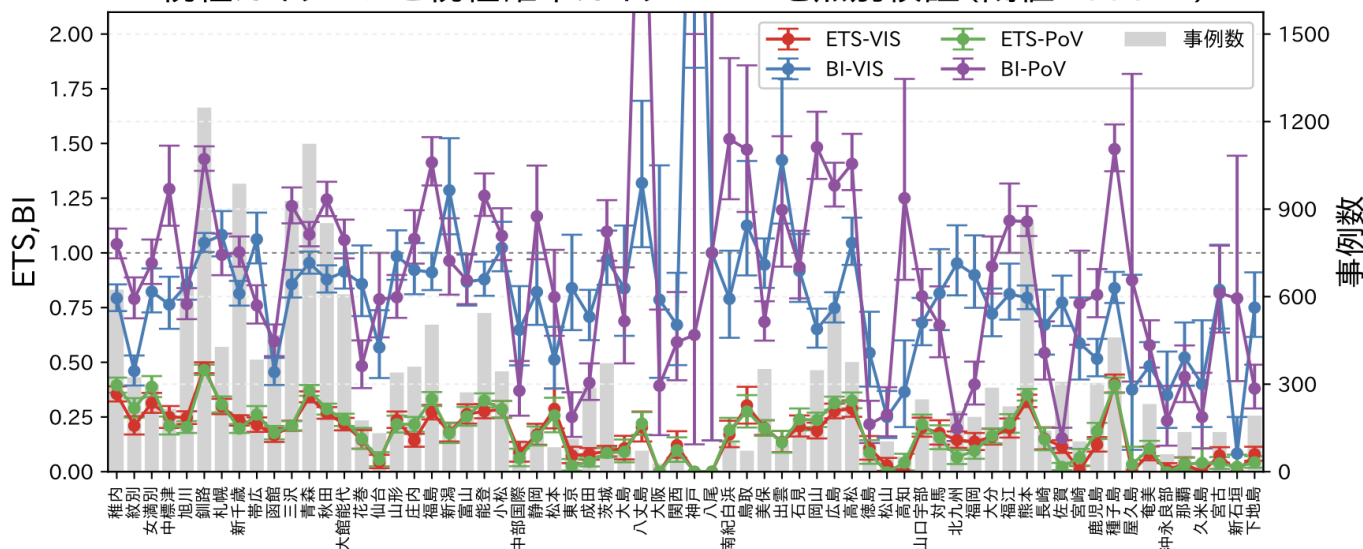
→ 分析[1]の閾値も 1600m とする

※気象庁の手法はカルマンフィルター＋頻度バイアス補正

通常公開されていないデータも使える

→ 基本、不利な戦いではある

視程ガイダンスと視程確率ガイダンスの地点別検証(閾値1600 m)



【参考 1】

- 上のグラフは、数値予報課報告・別冊第64号より引用した、空港ごとの視程ガイダンスの精度評価を表す
- グラフからの読み取りで、成田空港はだいたい $ETS=0.08$ / $BI=0.7$

【参考 2】

- 民間気象会社時代に自ら開発した視程予測式を再現計算した
- 2017年の予測テストでは、 $ETS=0.08$ / $BI=4.3$
- 精度は視程ガイダンスと同程度だが、空振りが非常に多かった

1. ロジスティック回帰 （霧の有無の予測）
2. Elastic Net （霧の有無の予測）
3. ランダムフォレスト （霧の有無の予測）
4. xgboost （霧の有無の予測）
5. サポートベクターマシン （霧の有無の予測）
6. ディープラーニング （視程カテゴリー予測）
7. LSTM （視程の時系列予測）

ロジスティック回帰

まずはロジスティック回帰から…

学習時のスコア

		予報	
		0	1
実況	0	7577	151
	1	66	61
AUC		0.94	
ETS		0.21	
BI		1.67	

予測テストのスコア

		予報	
		0	1
実況	0	7879	159
	1	68	31
AUC		0.91	
ETS		0.11	
BI		1.92	

- いきなり気象庁と同等以上！
- BIはちょっと大きめ（空振りが多い）
- 学習時と予測テスト時の精度の差が大きすぎないか？
（over-fitting?）

ロジスティック回帰

● 回帰係数とオッズ比を確認

Step関数、VIFでマルチコ排除、
P値で*2個以上の変数のみ表示

- ✓ ほとんどの変数が気象ドメイン知識と合致する効果
- ✓ 950hPa風速は謎（でも解釈可）

・ただ気になるのが…

気温・露点温度の変化量、回帰係数がおかしい（乾燥するの？）

→ はずしてみよう

地上湿度の係数がデカ過ぎない？

→ over-fittingの可能性あり
正則化項をつけてみる

→ Elastic Net

※特徴量はスケーリングしている

特徴量	回帰係数	オッズ比	検定結果
(Intercept)	-18.1	0.00	***
前24時間に雨	0.7	2.04	***
地上湿度	10.3	30570.78	***
地上気圧	2.9	18.32	**
下層の雲量	0.8	2.26	***
中層の雲量	-1.2	0.30	**
3月	-1.0	0.37	**
4月	-1.2	0.30	**
5月	-0.9	0.41	**
7月	-0.8	0.43	***
気温・変化量	5.2	179.70	**
露点温度・変化量	-12.6	0.00	***
風の鉛直シア	-2.7	0.07	***
大気安定度1	4.8	115.90	***
大気安定度2	1.6	5.12	**
1000hPa湿度	3.9	48.84	**
1000hPa鉛直流	6.2	472.53	**
1000hPa風速	-5.6	0.00	***
950hPa風速	6.5	654.78	***

ロジスティック回帰

気温変化量と露点温度変化量をはずしたら、
精度は少し下がるが、空振りが大きく減った！

【なぜ？考察してみた】

実況で霧が発生すると、日の出後も気温が上がりにくく、空気が入れ替わりにくい。このとき数値予報は霧の発生を予測できていないと、日の出とともに日射による気温上昇を予測し、空気の入れ替わりで露点温度を下げる予測をする（特に冬に起こり得る）。こうした状況を学習していたためと考えられる。

		Before	
		予報	
		0	1
実況	0	7879	159
	1	68	31
AUC		0.91	
ETS		0.11	
BI		1.92	

		After	
		予報	
		0	1
実況	0	7929	109
	1	78	21
AUC		0.9	
ETS		0.09	
BI		1.31	

Elastic Net

- ✓ さらに空振りが減り、かつ精度は $ETS=0.1$ を維持できた！
- ✓ 湿度の大きすぎた回帰係数は程良くなり、時間変化量は係数ゼロに！
- ただし予測テストで学習時より精度が大幅に下がる点は変わらず…

【考察】

地上湿度による過剰な予測が抑えられた結果、バランスの良い予測になったといえる

Elastic Net

		予報	
		0	1
実況	0	7938	100
	1	78	21
AUC		0.91	
ETS		0.1	
BI		1.22	

特徴量	回帰係数	オッズ比
前24時間の雨	0.7	1.99
地上湿度	1.9	6.51
地上風速	-0.3	0.72
地上風向・西	0.2	1.19
午前3～5時	0.5	1.69
午後6～8時	-0.3	0.78
2月	0.4	1.43
11月	0.5	1.69
気温・変化量	.	
露点温度・変化量	.	
風の鉛直シア	0.1	1.15
大気安定度1	2.8	16.88
大気安定度2	1.2	3.18
950hPa速度	.	

※一部の特徴量のみ掲載

なぜ学習時の精度が予測時には出ないのか？

【数値予報モデルの要因】

- 気象庁の数値予報システムは毎年改修されており、特に2017年2月のVer.UPは大きな変化だった。特徴量の特性が変化した可能性がある。
- そもそも論、数値予報モデルの未来の予想がはずれることがある。本音では、数値予報がはずれることも含めて機械学習がうまくやってくれることを期待していた。

【気候の年々変動】

- 気候は毎年異なっており、気候特性が変化した可能性がある。特に2015～2016年は大規模なエルニーニョが発生し、全国的にかなり高温だったが、2017年は平年並み・特に夏は不順だった。
- 地球大気はとても複雑なふるまいをしており、数年程度の学習では母集団のすべての特徴をカバーしきれないかもしれない。感覚的には10年かそれ以上が必要。

ランダムフォレスト

すでに目標は達成してしまいましたが、やっぱり本命の機械学習！
まずはランダムフォレスト

学習時のスコア

		予報	
		0	1
実況	0	7516	223
	1	48	68
AUC		0.95	
ETS		0.19	
BI		2.51	

予測テストのスコア

		予報	
		0	1
実況	0	7923	115
	1	79	20
AUC		0.9	
ETS		0.09	
BI		1.36	

- ロジスティック回帰ほどの精度が出ない
- やっぱり学習時ほどの精度が予測テストでは出ない

※不均衡データなのでSMOTEでサンプルを増やし、GridSearchでチューニングすると学習時は ETS=0.8以上！ にもなるが、予測テスト時はかえって精度が悪化する（霧をあまり予測しなくなる）

【仮説1】主成分分析でいくつかの主成分に絞ったら、年々変動成分は削られ、汎用性の高い学習ができるのではないかな？

1. SMOTEでOver Sampling
2. PCAで次元圧縮
3. ランダムフォレストのclass_weightを調整

学習時のスコア

		予報	
		0	1
実況	0	12914	66
	1	481	424
AUC		0.96	
ETS		0.42	
BI		0.54	

予測テストのスコア

		予報	
		0	1
実況	0	8000	38
	1	85	14
AUC		0.87	
ETS		0.1	
BI		0.53	

- ETS = 0.1にあがったが、学習時ほどの精度はでないかつ予報が過小になっている
- 仮説に反して、霧の有り無しで特徴量の差が薄れてしまったそのためかなり無理矢理の調整をしてこの精度になった

【仮説2】定期的に学習を更新すれば、過学習を防げるのではないか？

1. 1か月ごとに過去1年のデータで学習を繰り返す
2. 2015年からスタートし、2017年まで学習&予測を続ける

3年間の総スコア

		予報	
		0	1
実況	0	24947	267
	1	286	58
ETS		0.09	
BI		0.94	

年ごとの精度比較

Year	ETS
2015	0.09
2016	0.14
2017	0.03

- 総スコアではETS = 0.09で、改善は見られない…
- 年ごとの精度の違いが大きく、気候の違いが原因と考えられる

高性能ともっぱら評判のxgboostですが...

学習時のスコア

		予報	
		0	1
実況	0	7701	38
	1	62	54
AUC		0.96	
ETS		0.34	
BI		0.79	

予測テストのスコア

		予報	
		0	1
実況	0	8009	29
	1	94	5
AUC		0.9	
ETS		0.04	
BI		0.34	

- 学習性能は高いが、予測テストでは精度が大幅に落ちる（霧の予測が少ない）
- SMOTEでOver Samplingするとほとんど完璧に学習するが、予測テストではぜんぜん霧を予想しなくなる

→ 高性能ゆえに、学習した年に過剰にフィットしてしまう（過学習の呪縛）

サポートベクターマシン

一応やってみた

● Kernel SVM : たいした精度は出ない

学習時のスコア

		予報	
		0	1
実況	0	7648	91
	1	56	60
ETS		0.28	
BI		1.3	

予測テストのスコア

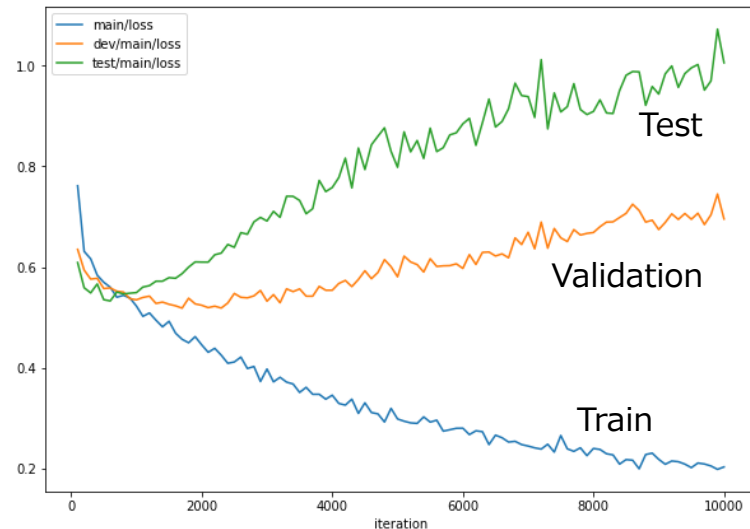
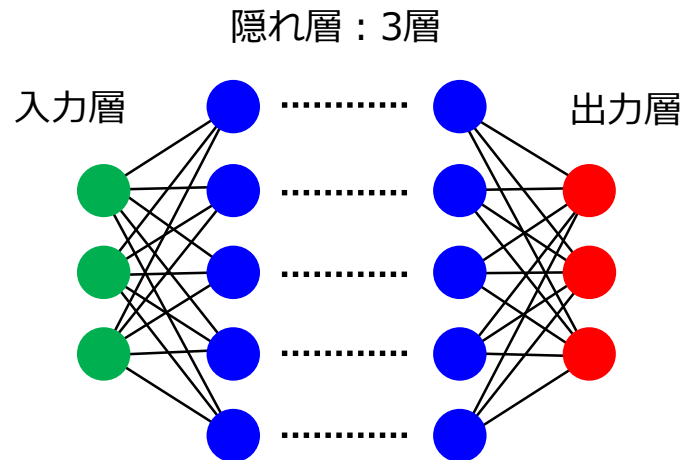
		予報	
		0	1
実況	0	7957	81
	1	89	10
ETS		0.05	
BI		0.92	

※なおスケーリングしないほうが精度が良かった

● Linear SVM : 全然ダメ、霧の予測ゼロ

ディープラーニング

ランダムフォレストやxgboostがだめなら、ディープラーニングだ！
視程カテゴリーの予測にトライしてみた



学習時のLOSSの推移

Batch Size = 128, Iteration = 10000回

- ✓ やっぱり過学習してしまう！
- ✓ TestだけでなくValidationでもLOSSが増加
- ✓ 予測精度もほどほど…

ETS 800m : 0.05

ETS 1600m : 0.08

		予測カテゴリー				
		0	1	2	3	4
実況カテゴリー	0	6850	312	122	7	11
	1	198	108	66	2	2
	2	156	100	99	4	1
	3	27	13	10	4	1
	4	20	5	14	2	3

- 2014～2015年でしっかり学習する
Optimizer = Adam, Batch Size = 1024, Epoch = Early Stopping
- 2016～2017年は1か月毎に新データでオンライン学習
Optimizer = SGD, Sample = 720, Batch Size = 24, Epoch = max 2
- LOSS関数のclass_weightを調整

【ポイント】

事前にしっかり学習させ、オンライン学習では過学習を防ぐために、OptimizerはSGDを使い、バッチサイズやエポック数を小さくした

		予測カテゴリー				
		0	1	2	3	4
実況カテゴリー	0	7540	77	171	20	42
	1	263	37	101	17	27
	2	135	52	131	7	18
	3	30	6	15	1	13
	4	20	5	7	2	14

- ✓ 精度は**最大ETS = 0.12**まで上昇！
- ✓ ただし何回か計算すると精度にブレがあり、だいたいETS = 0.07～0.12
- ✓ ネットワークは深くしても精度は上がらない（おそらくOver Fittingするだけ）

ETS 800m : 0.09

ETS 1600m : 0.12

(注：一例です)

ここまでの予測は、すべて気象庁の数値予報モデルの精度に依存していた。
なので、観測値だけでLSTMを使って予測をすることを試す。
長時間先の予測は無理だが、目先1～3時間先ならやる価値がある。

【観測値から得られる要素】

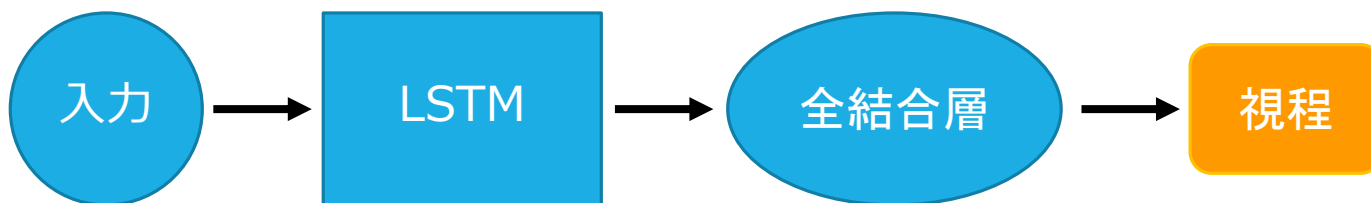
視程・風向・風速・瞬間風速・雨の有無・霧雨の有無・雪の有無・雲量
気温・露点温度・湿数(≒湿度)・地上気圧

【モデルの作り方】

時刻Tの視程以外の要素を特徴量とする

時刻T+1, T+3, T+6の視程をターゲット変数とする

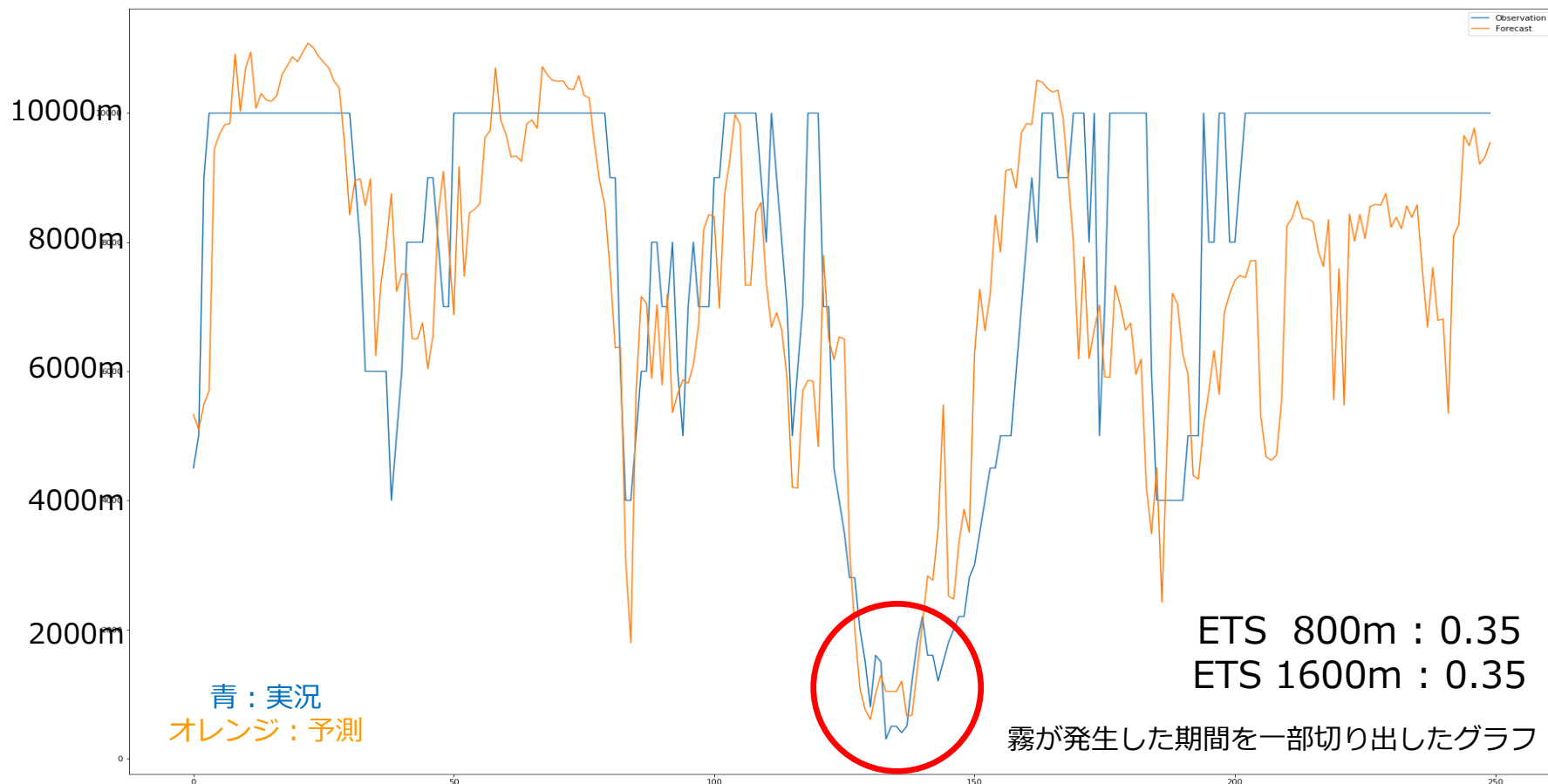
(このとき、視程は対数をとっておく)



2014年～2015年梅雨くらいまでをTrain、2015年残りをValidationとし、
2016年の1年間を予測Test期間とする

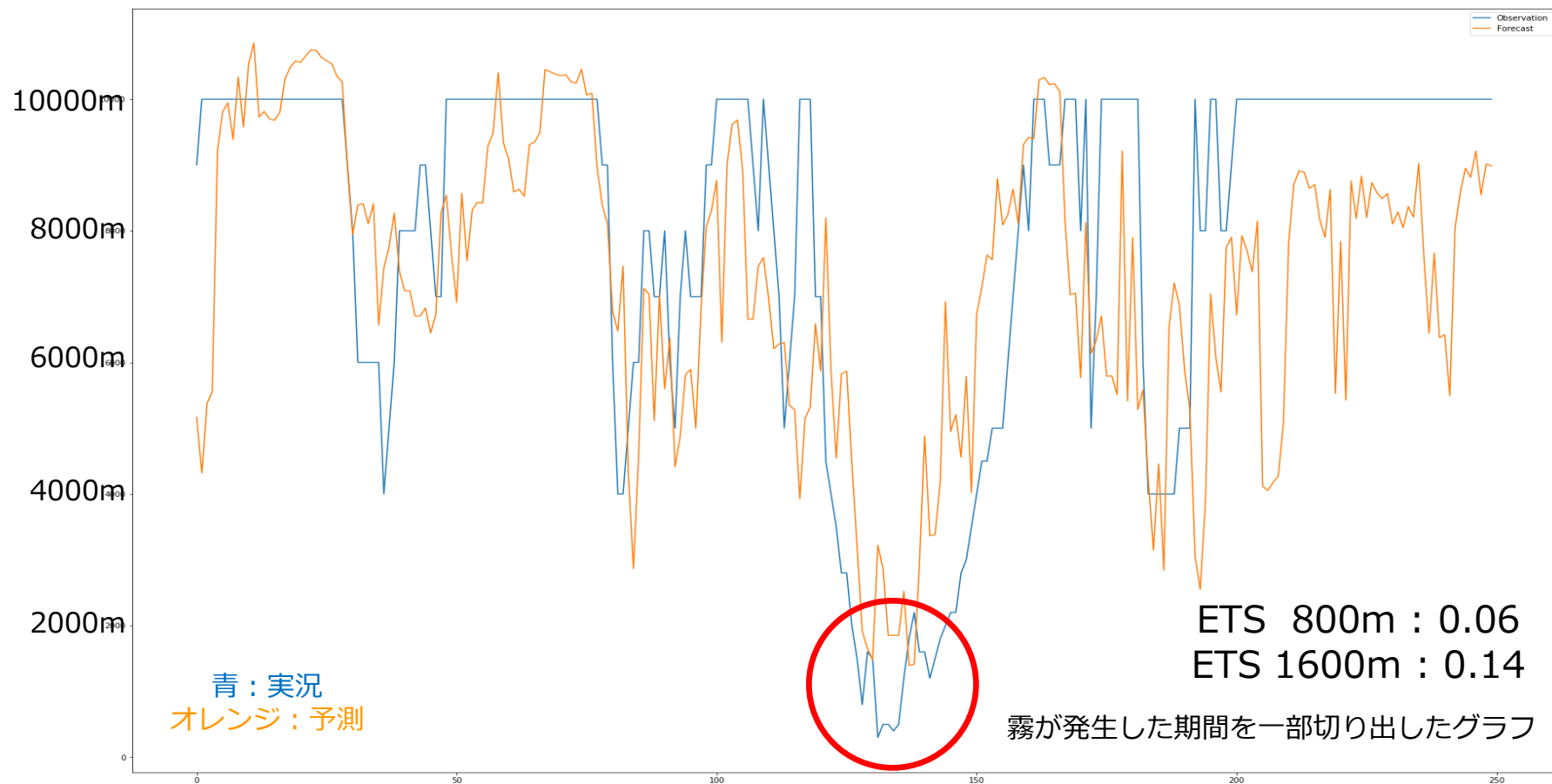
LSTM (1時間先の予測)

- 1時間後はなかなか良い予測結果が出た
- 実況で視程が急激に下がる様子を事前に予想できているところがある



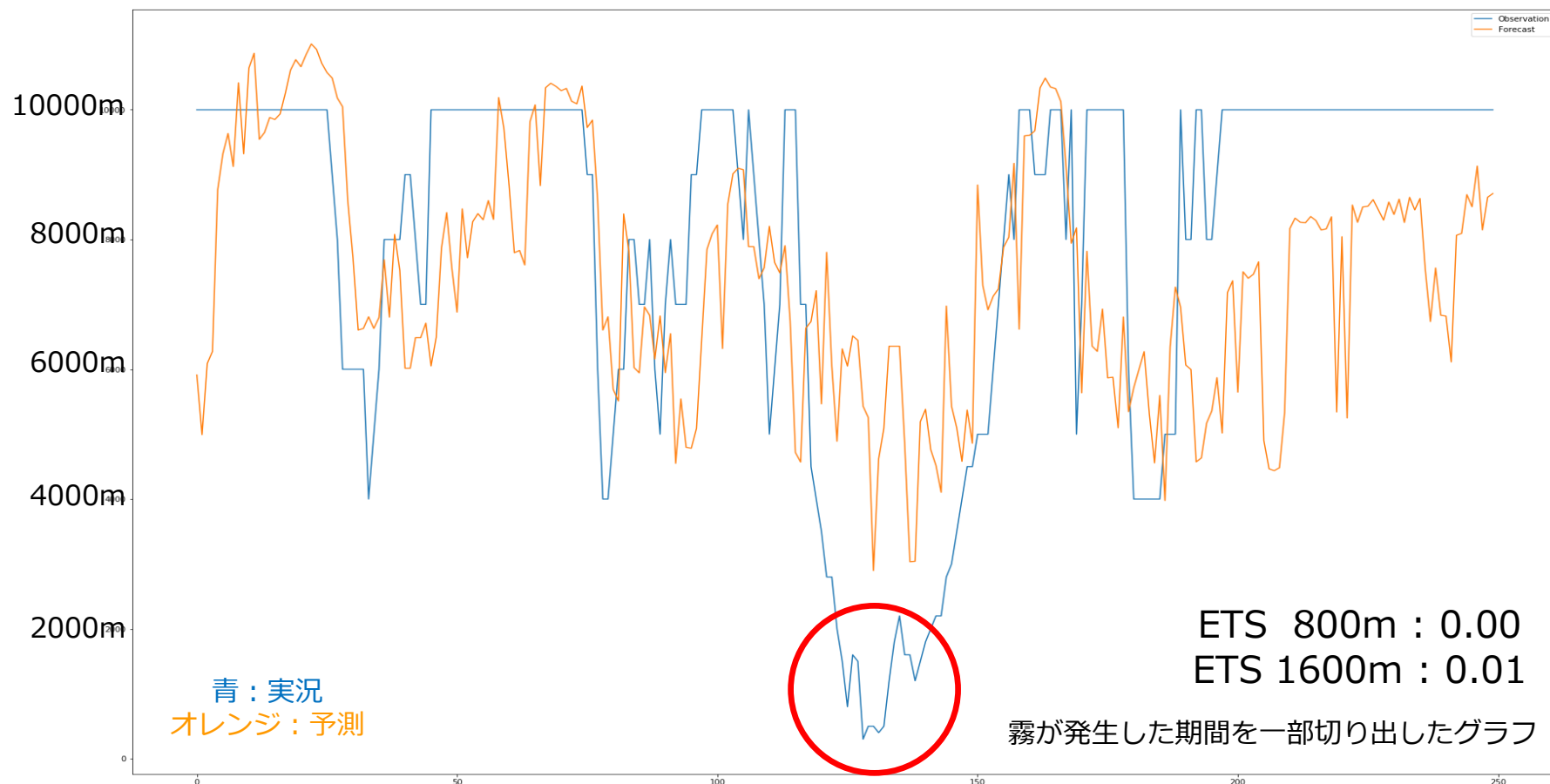
LSTM (3時間先の予測)

- 3時間後にもなるとだいぶ苦戦している
- 健闘しているものの、実況にワンテンポ遅れる傾向が現れる
- 霧の予測数が少なくなっている



LSTM (6時間先の予測)

- 6時間後にもなると霧の予測はほぼなくなる…
- ただなんとなくの変動の傾向は出ており、実況だけを使って6時間後の変化傾向がここまで出るとは！

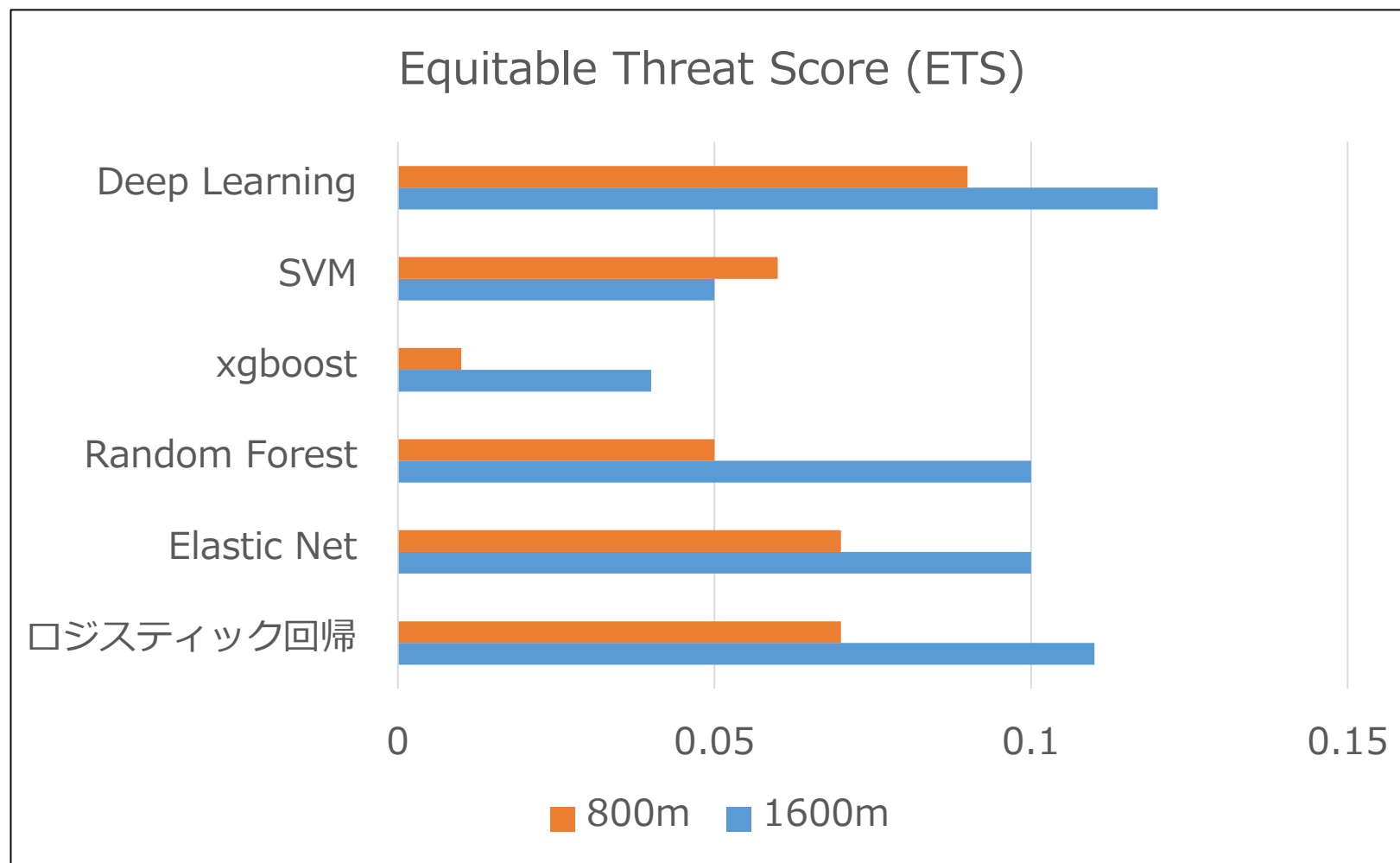


結果・考察・今後の課題

- いくつかのアルゴリズムで、気象庁と同等以上の精度が出せた
- 気象のドメイン知識を用いた特徴量の増減が有効であり、気象予報士とAIの合わせ技の効果といえる
- あえて1つ選ぶなら **Elastic Net** (扱いやすさ、解釈しやすさ)
- ❑ ランダムフォレストやディープラーニングでは、条件によって気象庁と同等以上の精度が出るが、期待したほどの精度は出せなかった
- ❑ **LSTM**は1~3時間先のナウキャスト予測として利用価値ありそう
- ✓ どのアルゴリズムを用いても、学習時に比べてテスト時の精度が大幅に落ちていた
- ✓ 特徴量としている数値予報データのブレや不確実性、年による気候の違いが原因として考えられる
- ✓ 高性能と言われる機械学習でも、この問題には対処できない（むしろ過学習の呪縛にはまる傾向がある）
- ◆ DNNやLSTMなどは、まだチューニングの余地あり
- ◆ 季節ごとや時間帯ごとにモデルを分けると精度向上の余地あり

ちなみに視程800m未満を閾値にすると…

飛行機の運航では800m未満の霧を予報できるかどうかで大きな違いになる…

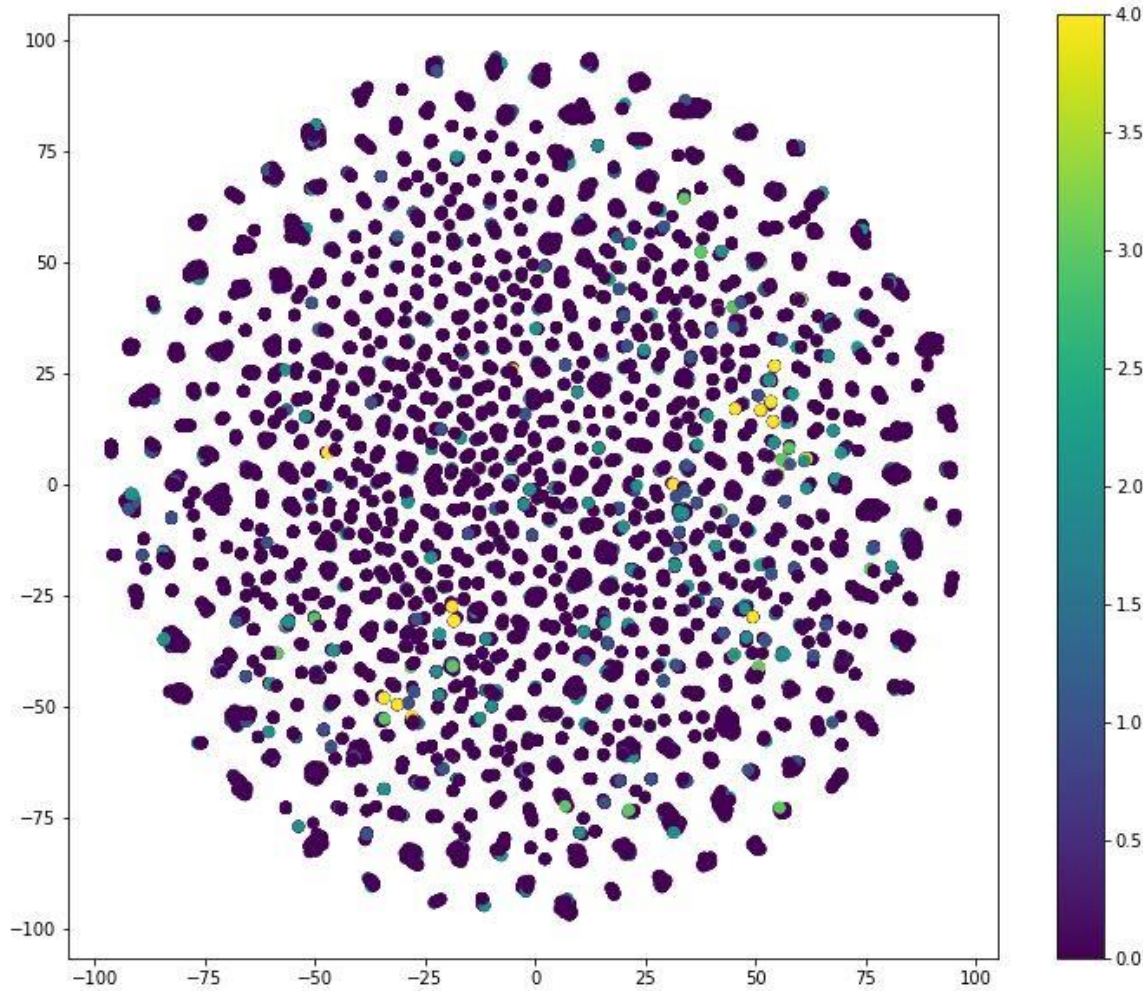


※各アルゴリズムの、試行錯誤を通じたMAXスコアです

※アルゴリズムによってパワーのかけ具合に違いがあり、一概に優劣は言えません

最後に

表紙に使ったこの図、TSNEで特徴量を2次元まで圧縮してみた
色は視程カテゴリーで、黄色ほど視程が低い霧の状態



キレイに分類されたら大発見！と思ったが、世の中そんなに甘くなかった…
気象の複雑さを物語る絵ができたと思っています

近年のAI技術の発達が目覚ましく、
データロボットなんてものまで現れましたが…

人としての気象予報士と
技術としてのAIが
どちらも必要です

完