# Выборка

Прогнозируемое значение - минимальная температура воздуха на высоте 2 метров, измеренные на метеостанции Чумикан (номер 31286) в 21 ВСВ.

Данные для отбора признаков были взяты из расчетов модели WRF (Weather Research and Forecasting) за период 2013-2018 г с заблаговременностью 9, 21, 33, 45, 57, 69. Каждый признак из Таблицы 1 представлен в виде минимума, максимума и среднего значения за полусутки (с 9 до 21 и с 21 до 9 часов ВСB).

Согласно [1], расчет WRF производился по области с шагом сетки 5 км на 27-ми вертикальных уровнях. Параметры применяемой конфигурации приведены в приложении 2. Начальные и граничные условия для WRF были заданны по данным оперативных прогнозов сроком на 72 ч. Национального центра прогнозирования атмосферы NCEP (США). Использовались следующие данные о подстилающей поверхности: высота поверхности над уровнем моря, категория подстилающей поверхности в соответствии с классификацией Геологической службы США USGS (например, лес, водная поверхность, и т.д.), тип растительности, тип почвы, средняя сезонная температура, среднемесячное альбедо. Соответствующие данные брались с сайта USGS.

В качестве признаков взяты среднее значение из Таблицы 1 за полусутки. Так как зависимость лучше искать в группах с одно заблаговременностью, была оставлена группа с заблаговременностью 69, как самая многочисленная. Время актуальности было выбрано 21 ВСВ, так как совпадало с временем измерения прогнозируемого значения.

Из выборки прогнозируемого значения были исключены значения, которые в окне 10 дней выше 95-го процентиля или ниже 5-го процентиля. Этот процесс называется усечением, и нужна для устранения выбросов.

# Составление и оценка прогноза

Для построения тренировочного и тестового набора использовалась 5-кратная последовательная кросс-валидация. В отличие от стандартного метода перекрестной проверки, тренировочные наборы - это надмножества тех, которые предшествуют им.

С помощью кросс-валидации можно объективно сравнить два метода относительно их ошибок на наборе данных. Так же эта техника позволяет проверить насколько успешно применяемый метод способен работать на независимом наборе данных.

Для каждого тренировочного набора рассчитывалось несколько показателей. Коэффициент детерминации

*,*

где .

Он представляет собой доля дисперсии, которая была объяснена независимыми переменными в модели и показывает насколько хорошо модель будет предсказывать на других данных.

Средняя абсолютная ошибка

.

Средняя квадратичная ошибка

*.*

# Линейная регрессия

Для построения линейной регрессии используется Stepwise регрессия, основанная на p-значения из api.OLS библиотекеstatsmodels.

На первом шаге из списка всех возможных входных переменных выбирается признак, наиболее коррелирующий с прогнозируемым значением. После включения в схему производится построение модели statsmodels api.OLS для отобранных предикторов и предиктанта. Переменная добавляется в модель если её p-значения меньше заданного порога 0.01. Это шаг называется прямым выбором. Затем происходит исключение в обратном направлении. Он включает в себя удаление переменной с максимальными значением p-значения больше порога 0.05. Итерации продолжаются, пока возможно добавить или удалить признак.

Таблица 1. Результаты линейной регрессии

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 |
| *MSE* | 16.19 | 627.24 | 190.36 | 20.91 | 16.33 |
| *MAE* | 3.24 | 20.83 | 12.38 | 3.52 | 3.31 |
|  | 0.92 | -4.23 | -1.39 | 0.67 | 0.85 |

# Градиентный бустинг (GBM)

Параметры

max\_depth = 3

min\_child\_weight = 3.01

Таблица 1. Результаты градиентного бустинга

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 |
| *MSE* | 24.13 | 39.66 | 71.19 | 9.96 | 12.44 |
| *MAE* | 3.95 | 4.32 | 7.16 | 2.58 | 2.86 |
|  | 0.88 | 0.67 | 0.10 | 0.84 | 0.89 |

# Приложение

## Приложение 1

Таблица 1. Признаки, привлекаемые для разрабатываемой модели

|  |  |  |
| --- | --- | --- |
| № | имя | Описание |
| 1 | advRh2 | Адвекция относительной влажности на высоте 2м |
| 2 | advRh850 | Адвекция относительной влажности на высоте 850 гПа |
| 3 | advT2 | Адвекция температуры на высоте 2м |
| 4 | advT850 | Адвекция температуры на высоте 850 гПа |
| 5 | advTdd2 | Адвекция дефицита точки росы на высоте 2м |
| 6 | advTdd850 | Адвекция дефицита точки росы на высоте 850 гПа |
| 7 | cape | Доступная потенциальная энергия конвекции |
| 8 | cin | Энергия подавления конвекции |
| 9 | defSh500 | Деформация сдвига поля ветра на высоте 500 гПа |
| 10 | defSt500 | Деформация сжатия поля ветра на высоте 500 гПа |
| 11 | defV500 | Полная деформация поля ветра на высоте 500 гПа |
| 12 | dudx500 | Частная производная компоненты скорости ветра на высоте 500 гПа |
| 13 | dudy500 | --//-- |
| 14 | dvdx500 | --//- |
| 15 | dvdy500 | --//- |
| 16 | lcl | Уровень конденсации |
| 17 | lfc | Уровень свободной конвекции |
| 18 | rainc | Количество жидких осадков из конвективных схем |
| 19 | rainnc | Кол-во жидких осадков |
| 20 | rh2 | Относительная влажность на высоте 2 м. |
| 21 | rh500 | Относительная влажность на высоте 500 гПа |
| 22 | rh850 | Относительная влажность на высоте 850 гПа |
| 23 | rh925 | Относительная влажность на высоте 925 гПа. |
| 24 | rot500 | Ротор поля ветра на высоте 500 гПа |
| 25 | slptend | Барическая тенденция на уровне моря |
| 26 | slp | Давление на уровне моря |
| 27 | snownc | Количество твердых осадков |
| 28 | t2 | Температура воздуха на высоте 2 м. |
| 29 | t500 | Температура воздуха на высоте 500 гПа |
| 30 | t850 | Температура воздуха на высоте 850 гПа |
| 31 | t925 | Температура воздуха на высоте 925 гПа. |
| 32 | td2 | Точка росы на высоте 2 м. |
| 33 | td500 | Точка росы на высоте 500 гПа |
| 34 | td850 | Точка росы на высоте 850 гПа |
| 35 | td925 | Точка росы на высоте 925 гПа. |
| 36 | tdd2 | Дефицит точки росы на высоте 2 м. |
| 37 | tdd500 | Дефицит точки росы на высоте 500 гПа |
| 38 | tdd850 | Дефицит точки росы на высоте 850 гПа |
| 39 | tdd925 | Дефицит точки росы на высоте 925 гПа. |
| 40 | totalprecip | Общее количество осадков |
| 41 | u10 | Зональная составляющая скорости ветра на высоте 10 м |
| 42 | u500 | Зональная составляющая скорости ветра на высоте 500 гПа |
| 43 | u850 | Зональная составляющая скорости ветра на высоте 850 гПа |
| 44 | v10 | Меридиональная составляющая скорости ветра на высоте 10 м |
| 45 | v500 | Меридиональная составляющая скорости ветра на высоте 500 гПа |
| 46 | v850 | Меридиональная составляющая скорости ветра на высоте 850 гПа |
| 47 | w500 | Вертикальная составляющая скорости ветра на высоте 500 гПа |
| 48 | w850 | Вертикальная составляющая скорости ветра на высоте 850 гПа |
| 49 | wdir10 | Направление ветра на высоте 10 м |
| 50 | wdir500 | Направление ветра на высоте 500 гПа |
| 51 | wdir850 | Направление ветра на высоте 850 гПа |
| 52 | wspd10 | Скорость ветра на высоте 10 м |
| 53 | wspd500 | Скорость ветра на высоте 500 гПа |
| 54 | wspd850 | Скорость ветра на высоте 850 гПа |

## Приложение 2

Параметры конфигурации модели

// globalattributes:

:TITLE = " OUTPUT FROM WRF V3.4.1 MODEL" ;

:START\_DATE = "2012-12-28\_12:00:00" ;

:SIMULATION\_START\_DATE = "2012-12-28\_12:00:00" ;

:WEST-EAST\_GRID\_DIMENSION = 150 ;

:SOUTH-NORTH\_GRID\_DIMENSION = 150 ;

:BOTTOM-TOP\_GRID\_DIMENSION = 28 ;

:DX = 5000.f ;

:DY = 5000.f ;

:GRIDTYPE = "C" ;

:DIFF\_OPT = 1 ;

:KM\_OPT = 4 ;

:DAMP\_OPT = 0 ;

:DAMPCOEF = 0.2f ;

:KHDIF = 0.f ;

:KVDIF = 0.f ;

:MP\_PHYSICS = 8 ;

:RA\_LW\_PHYSICS = 1 ;

:RA\_SW\_PHYSICS = 1 ;

:SF\_SFCLAY\_PHYSICS = 2 ;

:SF\_SURFACE\_PHYSICS = 2 ;

:BL\_PBL\_PHYSICS = 2 ;

:CU\_PHYSICS = 5 ;

:SURFACE\_INPUT\_SOURCE = 1 ;

:SST\_UPDATE = 0 ;

:GRID\_FDDA = 0 ;

:GFDDA\_INTERVAL\_M = 0 ;

:GFDDA\_END\_H = 0 ;

:GRID\_SFDDA = 0 ;

:SGFDDA\_INTERVAL\_M = 0 ;

:SGFDDA\_END\_H = 0 ;

:HYPSOMETRIC\_OPT = 2 ;

:SF\_URBAN\_PHYSICS = 1 ;

:SHCU\_PHYSICS = 0 ;

:MFSHCONV = 0 ;

:FEEDBACK = 1 ;

:SMOOTH\_OPTION = 0 ;

:SWRAD\_SCAT = 1.f ;

:W\_DAMPING = 0 ;

:MOIST\_ADV\_OPT = 1 ;

:SCALAR\_ADV\_OPT = 1 ;

:TKE\_ADV\_OPT = 1 ;

:DIFF\_6TH\_OPT = 0 ;

:DIFF\_6TH\_FACTOR = 0.12f ;

:OBS\_NUDGE\_OPT = 0 ;

:BUCKET\_MM = -1.f ;

:BUCKET\_J = -1.f ;

:PREC\_ACC\_DT = 0.f ;

:OMLCALL = 0 ;

:ISFTCFLX = 0 ;

:ISHALLOW = 0 ;

:DFI\_OPT = 0 ;

:WEST-EAST\_PATCH\_START\_UNSTAG = 1 ;

:WEST-EAST\_PATCH\_END\_UNSTAG = 149 ;

:WEST-EAST\_PATCH\_START\_STAG = 1 ;

:WEST-EAST\_PATCH\_END\_STAG = 150 ;

:SOUTH-NORTH\_PATCH\_START\_UNSTAG = 1 ;

:SOUTH-NORTH\_PATCH\_END\_UNSTAG = 149 ;

:SOUTH-NORTH\_PATCH\_START\_STAG = 1 ;

:SOUTH-NORTH\_PATCH\_END\_STAG = 150 ;

:BOTTOM-TOP\_PATCH\_START\_UNSTAG = 1 ;

:BOTTOM-TOP\_PATCH\_END\_UNSTAG = 27 ;

:BOTTOM-TOP\_PATCH\_START\_STAG = 1 ;

:BOTTOM-TOP\_PATCH\_END\_STAG = 28 ;

:GRID\_ID = 1 ;

:PARENT\_ID = 0 ;

:I\_PARENT\_START = 1 ;

:J\_PARENT\_START = 1 ;

:PARENT\_GRID\_RATIO = 1 ;

:DT = 30.f ;

:CEN\_LAT = 45.f ;

:CEN\_LON = 134.5f ;

:TRUELAT1 = 45.f ;

:TRUELAT2 = 45.f ;

:MOAD\_CEN\_LAT = 45.f ;

:STAND\_LON = 134.5f ;

:POLE\_LAT = 90.f ;

:POLE\_LON = 0.f ;

:GMT = 12.f ;

:JULYR = 2012 ;

:JULDAY = 363 ;

:MAP\_PROJ = 1 ;

:MMINLU = "MODIFIED\_IGBP\_MODIS\_NOAH" ;

:NUM\_LAND\_CAT = 20 ;

:ISWATER = 17 ;

:ISLAKE = -1 ;

:ISICE = 15 ;

:ISURBAN = 13 ;

:ISOILWATER = 14 ;

# Литература

1. Отчет о научно-исследовательской работе по теме внутреннего плана "Разработка информационной системы и технологии оперативного прогноза метеорологических параметров на станциях Приморского края на основе подхода MOSи выходных данных модели WRF", 2014г.