Представление

«Защищается студентка группы Б8303а Потоцкая Анастасия Эдуардовна по теме Комплексирование небольшого количества прогнозов с помощью градиентного бустинга; руководитель научная степень должность Олейников Игорь Сергеевич».

Слайд 2

Существуют модели прогноза погоды, но их точность не приемлема и не соответствует точности моделей, действующие на территории Северной Америки. Поэтому нужно попробовать другие методы прогноза, с целью улучшения результата.

Например, примеры моделей и методик прогноза ICON - европейсяка, GRAPES – японская и методы бэггинга вместо бустинга

Слайд 3

Данная курсовая работа посвящена исследованию путей улучшения качества составления прогнозов элементов погоды на метеостанциях Приморского края.

Для прогноза погоды используется численная модель прогноза WRF – региональная система прогноза погоды.

Прогнозы погоды рассчитанные WRF на прямую не используются синоптиками. Гораздо эффективнее использовать данные после статистической обработке.

~~Из-за недостатка точности и несовершенного начального состояния WRF, они подвержены систематическим и случайным ошибкам модели, которые имеют тенденцию расти со временем.~~

Слайд 4

Поэтому для составления прогнозов зачастую применяются методы комплексирования различных моделей и наблюдений на более ранние даты.

Одним из методов комплексирования является статистическая интерпретация гидродинамического прогноза (MOS model output statistic), в основе которого лежит предположение, что поля воспроизводимые моделями и данные наблюдений связаны линейной зависимостью, что в общем случае может быть не верно.

Слайд 5

Не смотря на огромное распространения и доказанную эффективность метода MOS, он требует не менее двух лет архивных результатов и наблюдений модели, в течении которых сама численная модель должна оставаться неизменной или почти неизменной. Это требование необходимо для того, чтобы полностью охватить статистику ошибок модели.

К сожалению, метеорологические центры Приморья такими данными не обладает. Поэтому применения технологии MOS в данный момент не эффективно.

Слайд 6 - Цель

Целью работы является исследование возможности комплексирование небольшого количества прогнозов с помощью градиентного бустинга для улучшения качества результатов модели WRF над территорией Приморского края.

Слайд 7 - Задачи

1. Изучить метод градиентного бустинга и его различные формы
2. Построить выборки для исследования
3. Выбрать библиотеку, реализующею градиентный бустинг
4. Построить прогноз и сравнить его с существующим решением – многомерной линейной регрессией
5. Опробовать различные методы коррекции значений рядов до применения алгоритмов бустинга

Слайд 8 – Градиентный бустинг

 Градиентный бустинг — это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Обучение ансамбля проводится последовательно. На каждой итерации вычисляются отклонения предсказаний уже обученного ансамбля на обучающей выборке. Следующая модель, которая будет добавлена в ансамбль будет предсказывать эти отклонения. Новые деревья добавляются в ансамбль до тех пор, пока ошибка уменьшается, либо пока не выполняется одно из правил "ранней остановки".

Слайд 9 – Выбор существующих библиотек

Искалась самая распространённая и простая в понимания библиотеке. Были рассмотренный следующие варианты

* CatBoost

Уступает в скорости

Sklearn

уступает в функциональнсти

* XGBoost

XGBoost была выбрана как самая быстрая из представленных.

Слайд 10 – Выборка

Прогнозируемое значение - минимальная температура воздуха на высоте 2 метров, измеренные на метеостанции Чумикан (номер 31286) в 21 ВСВ.   
В качестве признаков взяты среднее значение из показателей, рассчитанных моделью WRF за полу сутки с заблаговременностью 69

Слайд 11 – 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.

Для каждого тренировочного набора рассчитывалось Средняя абсолютная ошибка и Средняя квадратичная ошибка

Слайд 12 – Испытуемые модели

Были испробовано следующие модели:

1. Линейная регрессия
2. Бустинг без модификации выборки
3. Бустинг с удалением коррелирующих признаков
4. Бустинг с признаками на основании их вклада в деревья решений
5. Бустинг с преобразованием Фурье целевого значения
6. Бустинг с DCT целевого значения
7. Бустинг с вейвлет-преобразованием целевого значения

Слайд 12 – 16 – Результат эксперимента

Простой градиентный бустинг показал лучшие оценки среди всех испробованных методов. Так же он лучше справился с зашумленными и ошибочными данными.

Линейная регрессия – хорошо предсказывается данные с большим смещением, на скученных данных может сильно ошибаться. Может неправильно предсказывает тенденцию в течении долгого времени.

Простой бустинг – На данных с большим разбросом не справляется. На маленьких данных работает очень плохо.

Вейвлет – становиться меньше сильных выбросов. Иногда значения за несколько дней слипаются вместе, что не хорошо (из алгоритма).

Фурье – на хороших данных работает плохо (небольшое смещение), но первой кармане ему удается немного справиться с похими данными.

DCT – смещение меньше по сравнению с Фурье, но появляется слипание в несколько дней

После удаление корр – разброс уменьшается (что хорошо), особенно на последних карманов. На плохих данных становиться очень мало информации.

Import – на плохих данных справляется хуже всех – по сравнению с корре. Разброс больше и

Слайд 13 – Заключение

Исследована методика коррекции прогноза погоды по результатам численной модели WRF.

Выполнен анализ различных средств реализации MOS. – какие средства выбраны

Произведен сравнительный анализ различных методов градиентного бустинга с линейной регрессией. - результаты

В дальнейшем исследованный подход может быть внедрен в оперативный прогноз погоды.