

Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский университет «Высшая школа экономики»
Факультет компьютерных наук
Образовательная программа Прикладная математика и информатика
бакалавриат
01.03.02 Прикладная математика и информатика

О Т Ч Е Т
по производственной практике

Выполнил студент гр. БПМИ 162

Сурин Денис Владимирович
(ФИО)

(подпись)

Проверил:

Устюжанин Андрей Евгеньевич,
Доцент

должность, ФИО руководителя от НИУ ВШЭ

оценка по 10 бальной шкале

подпись

Москва, 2019

Содержание.

<i>Цель практики.....</i>	<i>2</i>
<i>Задача практики.</i>	<i>2</i>
<i>Календарь план-график.....</i>	<i>3</i>
<i>Изученные материалы.</i>	<i>3</i>
<i>Примененные подходы и полученные результаты.....</i>	<i>4</i>
<i>Заключение.</i>	<i>6</i>
<i>Использованные источники.....</i>	<i>6</i>

Цель практики.

Целью практики было ознакомиться с текущими алгоритмами предсказания положения спутников и разработать / усовершенствовать их.

Задача практики.

В современном мире околоземная орбита приобретает все большее значение. Если на заре космической эры на ней располагались лишь несколько аппаратов, вначале практически не использующихся, то в настоящий момент спутники повсеместно используются как в военных целях, так и в мирной промышленности. Спектр применения довольно широк - от предсказания погоды до определения геопозиции или экстренной связи. Также в ближайшее время планируется вывод большего числа спутников для доступа к интернету. С течением времени некоторые спутники выходят из строя и становятся космическим мусором. Из-за большего числа объектов на орбите возникают столкновения, которые приводят к поломкам существующих спутников, причем в случае столкновения 2 спутников, в итоге появляется еще большее количество космического мусора. К примеру, одно лишь столкновение Iridium-Cosmos в 2009 году привело к увеличению космического мусора на 13%. При увеличении числа подобных столкновений использование околоземной орбиты может стать невозможным по причине большего количества мусора. Данный сценарий получил специальное название - эффект Кesslera.

Таким образом, предсказание положения спутника в нужный момент времени для предотвращения столкновения становится одной из наиболее актуальной задач современной астрономии. Ее решение позволит избежать затрат из-за поломок спутников, предотвратить эффект Кesslera, а также даст

возможность более рациональных перемещений, что приведет к большому сроку службы и экономии топлива.

Календарь план-график.

№ п/п	Сроки проведения	Выполненные работы	Отметка руководителя о выполнении (подпись)
1	15.11.19	1. Организационное собрание	
2	15.11.19	2. Инструктаж по ознакомлению с требованиями охраны труда, техники безопасности, пожарной безопасности, а также правилами внутреннего трудового распорядка	
3	15.11.2019	3. Экскурсия обзорная	
4	16.11-28.12.2019	4. Выполнение индивидуального задания	
5	16.11-28.12.2019	5. Консультации	
6	29.12.2019	6. Подготовка и предоставление отчета о прохождении практики	

Изученные материалы.

Данную тему сравнительно недавно стали пытаться решить методами машинного обучения, так как традиционно данная задачи считалась астрофизической и решалась специалистами из другой области. Одна из релевантных статья была издана в мае 2019 г. [2].

В качестве набора данных авторы статьи использовали 2 публично доступных космических каталога: TLE и ILRS. TLE каталог создан стратегическим командованием США (USSTRAT- COM) и является крупнейшим доступным каталогом космических объектов (<http://www.space-track.org/>). ILRS предоставляет более точные данные, но число космических объектов, данные о которых он предоставляет, меньше (https://ilrs.cddis.eosdis.nasa.gov/data_and_products/data/index.html). Целью было улучшить точность предсказаний искусственных спутников Земли.

Данную задачу не получается решить, используя традиционные физические подходы, так как нет параметров, которые необходимы для расчётов: положение космического объекта вначале вычислений, условия окружающей среды такие как атмосферное давление, давление света; точные характеристик тела, такие как масса, геометрия и материалы.

Каталог ILRS более точный, поэтому именно его данные рассматривались как истинные. Данные каталога TLE использовались в качестве признаков для получения предсказания моделью. Для получения предсказания использовался метод опорных векторов (SVM) [5]. В тестовую выборку вошел временной промежуток с 301 по 314 день. В обучающую - данные с 1 по 300 день (при подборе параметров SVM для обучения брали данные с 1 по 286 дни, и данные по последним 14 дням в качестве валидационной выборки). Эксперименты показали, что в многих случаях удалось добиться большей точности в сравнении с физической моделью, а также увеличение числа данных аналогично в большинстве случаев приводило к улучшению результата.

Примененные подходы и полученные результаты.

Для знакомства с данными и получения базовой оценки были применены модели классического машинного обучения. Для наших экспериментов данные были сгенерированы при помощи симуляторов GMAT и SGP4. В дальнейшем их количество будет увеличено. Набор данных представляет из себя координаты спутников (x , y , z) в данный момент времени и их скорость по каждой из компонент (V_x , V_y , V_z), представлена информация по 100 спутникам. Разброс значений координат примерно от -600000 до 600000. Скорости же примерно от -2 до 2. В обоих случаях данные распределены примерно симметрично относительно 0. Гистограмма похожа на бета распределение (см. рис. 1).

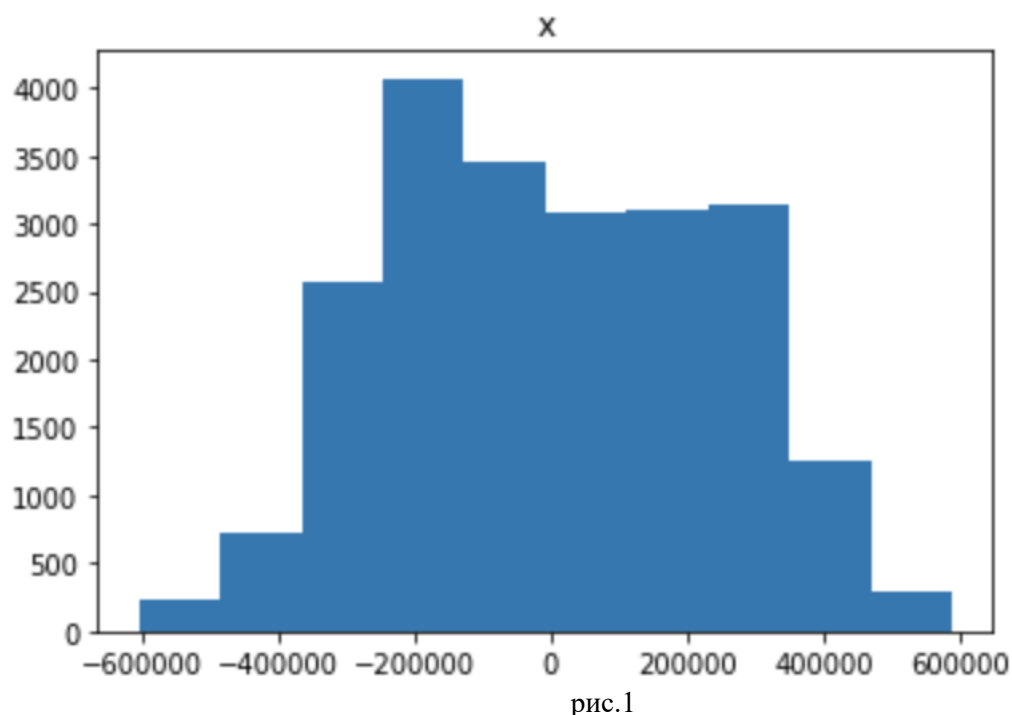


рис.1

В качестве метрики была взята MAPE[6]. После ряда экспериментов удалось получить несколько наиболее успешных результатов. В таблице представлены результаты на тестовой (**Mean MAPE test**) и обучающей (**Mean MAPE train**) выборках.

Model	Linear regression	Ridge	RandomForest Regressor	xgboost
Mean MAPE train	2.13	2.12	0.99	11.70
Mean MAPE test	2.78	2.77	2.96	102.93

Из экспериментов с моделями видно, что линейная регрессия показала наилучший результат, причем добавление L2 регуляризации улучшило результат незначительно (0.01%). Случайный лес в сравнении с линейной регрессией показал значительно лучшие результаты на обучающей выборке (0.99% против 2.13%), но уступил ей на тестовой (2.96% против 2.77%). Преимущество на обучающей выборке можно объяснить устройством случайного леса (так как он построен на деревьях, дающих большие возможности для настройки модели). Но потеря преимущества при тестировании свидетельствует, что случайный лес переобучился под обучающую выборку.

Заключение.

В своей работе я разобрал статью [2], проанализировал данные и провел ряд экспериментов с моделями: линейная регрессия, гребневая регрессия, случайный лес и градиентный бустинг. Эксперименты и анализ литературы показывают, что применение методов классического машинного обучения к данной астрофизической задаче показывает хорошие результаты. В дальнейших исследованиях планируется провести эксперименты с рекуррентными нейронными сетями, так как данные представляют из себя координаты спутника в течение последовательных временных промежутков.[3]

Использованные источники.

1. Hao Peng, Xiaoli Bai. Improving Orbit Prediction Accuracy through Supervised Machine Learning. <https://arxiv.org/abs/1801.04856>
2. Hao Peng, Xiaoli Bai. Machine Learning Approach to Improve Satellite Orbit Prediction Accuracy Using Publicly Available Data. <https://link.springer.com/article/10.1007/s40295-019-00158-3>
3. Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara. Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. <https://arxiv.org/abs/1909.00590>
4. IDAO 2020. <https://official.contest.yandex.ru/contest/16669/problems/>
5. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York, ISBN 978-1-4419- 3160-3 978-1-4757-3264-1. <https://doi.org/10.1007/978-1-4757-3264-1> (2000)
6. https://en.wikipedia.org/wiki/Mean_absolute_percentage_error
7. <https://github.com/denissurin2011/satellite-2019>