

# Классификация космических аппаратов на основе анализа их поведения

К. С. Рогов

Санкт-Петербургский государственный  
электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)  
Санкт-Петербург, Россия  
pluskika@gmail.com

И. И. Холод

Санкт-Петербургский государственный  
электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)  
Санкт-Петербург, Россия  
iiholod@mail.ru

**Аннотация.** В данной статье описывается классификация каталогизированных данных о космических объектах по открытой информации об их движениях. Анализируются, какие из доступных атрибутов являются необходимыми и достаточными для классификации без потери ее точности. Также исследуется возможность решения данной задачи с помощью нормального байесовского алгоритма, который позволяет построить классификатор за один проход по данным, что является критичным в случае с информацией, поступающей в режиме реального времени.

**Ключевые слова:** космические аппараты; кеплеровские параметры орбиты; классификация; нормальный байесовский классификатор

## I. ВВЕДЕНИЕ

Каждый год с момента начала космической гонки и по сегодняшний день правительства разных стран, сотовые операторы, телерадиокоммуникационные и GPS компании запускают на орбиту сотни космических аппаратов (КА). Они представляют собой специально сконструированные системы, которые могут функционировать в агрессивных средах. Их сложность и возможности сильно различаются, а назначение разнообразно. Выделяют основные классы КА:

- пилотируемые – предназначены для полета людей в космос;
- научно-исследовательские – орбитальные;
- астрономические обсерватории;
- геодезические и геофизические спутники;
- прикладные – метеорологические и навигационные;
- технические спутники связи;
- спутники для исследования земельных ресурсов.

Для определения класса новых и ранее неизвестных аппаратов могут быть использованы алгоритмы классификации [1], позволяющие производить анализ доступной информации.

Необходимо учитывать, что информация о движении аппаратов поступает постоянно в реальном времени. Для

корректной работы алгоритм должен обрабатывать данные за один проход и дообучаться на новых данных. Наиболее популярным алгоритмом классификации, удовлетворяющим данным требованиям является нормальный байесовский классификатор. Для сравнения в статье также рассмотрены модели деревьев решений и k-ближайших соседей. Также в статье исследуется необходимый и достаточный состав атрибутов, характеризующих движение КА для корректной классификации.

## II. КАТАЛОГ КОСМИЧЕСКИХ АППАРАТОВ

Для анализа были использованы данные проекта Space Track [2]. Это научно-исследовательский проект ВВС США, направленный на создание системы слежения за всеми искусственными спутниками Земли и космическими зондами.

Проект Space Track был запущен в исследовательском центре ВВС США в Кембридже, после запуска Спутника-1. В настоящее время находится в Бедфорде, штат Массачусетс. В 1960 году данные о космических аппаратах получались от 150 датчиков по всему миру, предоставлены заинтересованным сторонам. Тем самым обеспечивались ежедневные прогнозы орбит [3].

Space Track был единственной организацией, которая использовала информацию от всех типов измерительных средств: радиолокационных, оптических, радио и визуальных. Все несекретные наблюдения были переданы Смитсоновской астрофизической обсерватории. В 1961 году система была объявлена работоспособной и была названа SPADATS [3].

Данные в каталоге хранятся в формате 3LE (Three Line Element) и представляют собой параметры временного ряда, где на конкретную дату есть набор значений. Он представляет собой набор строк параметров КА, где на каждый объект отведено 3 строки по 80 символов (рис. 1). Объем данных за один год составляет в среднем около 40 Mb.

```
0 EXPLORER 11
1 22675U 93036A 09001.77226482 -.00000017 00000-0 33537-5 0 7107
2 22675 074.0367 089.9971 0016969 179.3358 180.7818 14.31133166811878
```

Рис. 1. Текстовое представление параметров КА

Из такого сжатого формата можно получить следующие атрибуты:

1. номер спутника в каталоге (номер);
2. микросекунда запуска (мсек);
3. наклонение (накл.);
4. долгота восходящего узла (ДВУ);
5. тип набора элементов (тип НЭ);
6. номер витка на момент эпохи (номер витка);
7. тип эфемерид (тип эфем.);
8. эксцентриситет (эксц.);
9. аргумент перигея (АП)
10. средняя аномалия (серд. ан-л);
11. среднее движение (сред. дв-н);
12. первая производная среднего движения по времени (1 ПСД);
13. вторая производная среднего движения по времени (2 ПСД);
14. баллистический коэффициент (бал. коэф.).

Итого данные за один год содержат около 330 тысяч записей для всех 8 классов.

- Navigation (содержит 38 КА) – навигация
- Weather (содержит 37 КА) – метеорологические

- Orgcomm (содержит 60 КА) – промышленный Интернет
- Human Spaceflight (HS) (содержит 2 КА) – пилотируемый аппарат
- Iridium (содержит 105 КА) – всемирный оператор спутниковой телефонной связи
- Intelsat (содержит 95 КА) – спутниковая связь
- Globalstar (содержит 84 КА) – спутники предоставляющие спутниковую телефонную связь и передачу данных
- Amateur (содержит 71 КА) – частные спутники

Для того чтобы ускорить обучение классификатора без существенной потери точности классификации будут использованы только значимые атрибуты, которые описывают параметры орбиты конкретного спутника, позволяющие определить его местоположение: наклонение, долгота восходящего узла, эксцентриситет, аргумент перигея, средняя аномалия, среднее движение, первая и вторая производная среднего движения по времени, баллистический коэффициент, тип эфемерид. После удаления атрибутов размер данных за год уменьшился на 9,6 МВ.

Для выбора наиболее значимых атрибутов рассмотрим таблицу корреляции Пирсона (табл. 1). В ней исследуем степень линейных связей между переменными и затем исключим зависимые атрибуты.

ТАБЛИЦА 1 КОРРЕЛЯЦИЯ АТТРИБУТОВ

	Сред. дв-н	Эксц.	Накл.	ДВУ	АП	Серд. ан-л	Тип эфем.	Бал. коэф.	1 ПСД	2 ПСД
Сред. дв-н	1.0	-0.10	0.70	0.13	-0.17	0.10	NaN	0.01	0.02	0.01
Эксц.	-0.10	1.0	-0.07	-0.01	-0.02	0.04	NaN	-0.001	0.006	0.002
Накл.	0.70	-0.07	1.0	0.13	-0.24	0.19	NaN	0.007	0.02	0.003
ДВУ	0.13	-0.01	0.13	1.0	-0.05	0.05	NaN	0.005	0.013	0.0007
АП	-0.17	-0.02	-0.24	-0.05	1.0	-0.49	NaN	-0.002	0.0002	0.008
Серд. ан-л	0.10	0.04	0.19	0.05	-0.49	1.0	NaN	-0.004	-0.007	-0.004
Тип эфем.	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN
Бал. коэф.	0.01	-0.001	0.007	0.005	-0.002	-0.004	NaN	1.0	0.58	0.002
1 ПСД	0.02	0.006	0.02	0.01	0.0002	-0.007	NaN	0.58	1.0	0.15
2 ПСД	0.01	0.002	0.003	0.0007	0.008	-0.004	NaN	0.002	0.15	1.0

Как видно из таблицы корреляции атрибуты среднее движение и наклонение, имеют высокую степень корреляции, но недостаточную для исключения из выборки, так как их значение не превышает 0.75. При этом атрибут «тип эфемерид» показал значение NaN. Это связано с тем, что он не изменяется во времени и изменения других параметров не влияет на изменение его самого. Его можно не использовать в конечной выборке атрибутов. Все остальные атрибуты являются независимыми.

Для определения влияния атрибутов на результат классификации были построены таблицы математического ожидания (табл. 2), дисперсии каждого атрибута (табл. 3) для каждого класса и значение атрибутов Gini index.

Gini index или Gini impurity измеряет степень принадлежности элементов к одному классу. Он варьируется от 0 до 1, где 0 обозначает, что все элементы принадлежат одному классу, а 1 обозначает, что элементы случайным образом распределены по различным классам.

ТАБЛИЦА II GINI INDEX

Сред. дв-н	Эксц.	Накл.	ДВУ	АП	Сред. ан-л	Бал. коэф.	1 ПСД	2 ПСД
0.32	0.05	0.59	0.01	0.009	0.01	0.00004	0.00008	0.00003

Как видно из табл. II, III и IV последние три атрибута (Бал. Коэф, 1 ПСД, 2 ПСД), имеют слабоизменяющиеся значения (одинаковое мат. ожидание, малую дисперсию и низкое значение gini index), поэтому при классификации их можно не использовать.

ТАБЛИЦА III МАТОЖИДАНИЕ

	Сред. дв-н	Эксц.	Накл.	ДВУ	АП	Сред. ан-л	Бал. коэф.	1 ПСД	2 ПСД
Amateur	4.4	0.00756	516.2	10722.6	10948.1	11293.1	0.00001	0.00001	0.00001
Globalstar	0.26	0.00001	0.0002	10710.3	8815.2	13283.5	0.00001	0.00001	0.00001
HS	0.01	0.00001	18.2	10604.2	12617.6	10842.5	0.00001	0.00002	0.00001
Intelsat	0.006	0.00268	32.5	14486.3	11332.0	9978.3	0.00001	0.00001	0.00001
Iridium	0.03	0.00003	0.02	6485.6	2821.3	3481.5	0.002	0.00001	0.00001
Navigation	0.001	0.00005	1.22	10303.7	12566.7	12569.4	0.00001	0.00001	0.00001
Orbcomm	0.02	0.00001	199.7	10920.9	10547.7	10408.9	0.00001	0.00001	0.00001
Weather	40.7	0.00002	1676.1	12354.2	11732.7	12249.2	0.00001	0.00001	0.00001

ТАБЛИЦА IV ДИСПЕРСИЯ

	Сред. дв-н	Эксц.	Накл.	ДВУ	АП	Сред. ан-л	Бал. коэф.	1 ПСД	2 ПСД
Amateur	13.8	0.01	84.9	179.8	182.5	174.8	0.00006	0.000004	0.000008
Globalstar	12.1	0.0004	51.9	180.0	144.4	179.2	0.00008	-0.00000	0.000002
HS	15.5	0.0005	48.3	182.5	180.1	167.5	0.000007	0.0002	0.000003
Intelsat	1.0	0.005	5.2	171.7	194.1	180.0	0.000009	-0.00001	0.000005
Iridium	14.4	0.001	86.38	134.0	102.7	255.0	-0.0004	-0.00001	0.000002
Navigation	1.9	0.008	54.9	172.0	157.0	191.0	0.000001	-0.00003	0.000002
Orbcomm	14.4	0.0006	49.4	181.5	196.1	166.5	0.00007	0.000005	0.000006
Weather	6.6	0.001	54.1	146.4	160.5	196.3	0.00001	-0.00005	0.000007

В результате в дальнейшем анализе будем использовать следующие атрибуты:

1. среднее движение;
2. эксцентриситет;
3. наклонение;
4. долгота восходящего узла (ДВУ);
5. аргумент перигея (АП);
6. средняя аномалия.

### III. ЭКСПЕРИМЕНТЫ

Для классификации данных были выбраны наиболее популярные модели классификации [4]

Деревья решений (DT – Decision tree) – представляют собой иерархические древовидные структуры, состоящие из решающих правил вида «если ..., то ...». Правила автоматически генерируются в процессе обучения на обучающем множестве и, поскольку они формулируются практически на естественном языке, деревья решений как аналитические модели более вербализуемы и интерпретируемы [5].

Классификатор K-ближайшего соседа (Knn – K-nearest neighbours) – задача поиска ближайшего соседа заключается в поиске среди множества элементов, расположенных в многомерном метрическом

пространстве, элементов близких к заданному, согласно некоторой функции близости. Объект классифицируется большинством голосов своих соседей. «K» всегда является положительным целым числом.

Соседи отбираются из набора объектов, для которых известна корректная классификация [6].

Случайный лес (RF – Random forest) – основан на построении большого числа деревьев решений, каждое из которых строится по выборке, получаемой из исходной обучающей выборки с помощью выборки с возвращением. В отличие от классических алгоритмов построения деревьев решений в методе случайных лесов при построении каждого дерева на стадиях расщепления вершин используется только фиксированное число случайно отбираемых признаков обучающей выборки и строится полное дерево, т.е. каждый лист дерева содержит наблюдения только одного класса [6].

Нормальный байесовский классификатор (NBC – Normal Bayes Classifier) классификатор основан на применении теоремы Байеса со строгими предположениями о независимости атрибутов. Его достоинством является построение классификатора за один проход по данным и возможность последующего дообучения модели при поступлении новых данных. [7].

Каждый классификатор был построен на данных за 2018 г. и для проверки их корректности были использованы данные за 2019 г. Для оценки качества

работы классификатора на каждом из классов по отдельности были использованы метрики precision, recall и f1-score.

Precision рассчитывается как доля объектов, названных классификатором положительными и при этом действительно являющихся положительными. Recall показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм. F1-score представляет собой гармоническое среднее между precision и recall. Она стремится к нулю, если precision или recall стремятся к нулю [8].

Первым шагом мы классифицировали выборку, используя все четырнадцать атрибутов. Значение метрики f1-score для полученных результатов приведены в табл. V в колонках «14».

ТАБЛИЦА V РЕЗУЛЬТАТЫ ДО И ПОСЛЕ ПРЕОБРАЗОВАНИЙ

Class	DT		K-nn		RF		NBC	
	14	6	14	6	14	6	14	6
Amateur	0.95	0.96	0.56	0.66	0.97	0.96	0.67	0.53
Globalstar	<b>1.00</b>	<b>1.00</b>	<b>0.70</b>	<b>0.81</b>	<b>1.00</b>	<b>1.00</b>	<b>0.88</b>	<b>1.00</b>
HS	0.26	0.26	0.09	0.02	0.29	0.39	0.25	0.70
Intelsat	<b>0.99</b>	<b>0.96</b>	<b>0.84</b>	<b>0.90</b>	<b>0.99</b>	<b>0.96</b>	<b>0.90</b>	<b>0.90</b>
Iridium	<b>0.99</b>	<b>1.00</b>	<b>0.72</b>	<b>0.94</b>	<b>1.00</b>	<b>1.00</b>	<b>0.91</b>	<b>0.96</b>
Navigation	0.89	0.89	0.69	0.74	0.89	0.89	0.73	0.89
Orbcomm	<b>0.98</b>	<b>0.99</b>	<b>0.29</b>	<b>0.79</b>	<b>1.00</b>	<b>1.00</b>	<b>0.64</b>	<b>0.82</b>
Weather	0.86	0.78	0.30	0.34	0.88	0.79	0.15	0.12

По результатам классификации, можно выделить как сильные классы (чьи результаты у каждой модели наивысшие), так и слабые классы (с самыми низкими показателями). К сильным классам относятся: Globalstar, Intelsat, Iridium и Orbcomm и (в табл. V выделены жирным шрифтом). К слабым Human Spaceflight (HS), Weather (в табл. V выделены курсивом)

После обучения моделей на уменьшенном наборе данных с 6ю атрибутами, получены результаты классификации, представленные в табл. V в колонках «6». Для большинства моделей результаты f1-score изменились в положительную сторону.

Исключение составляет классы Amateur и Weather для алгоритма Normal Bayes Classifier, где показатель f1-score уменьшился с 0,67 до 0.53 и с 0,15 до 0.12 соответственно (рис. 2). Это можно объяснить наличием в выборке двух атрибутов (сред. дв. и накл.) с высоким уровнем корреляции (0,7), что противоречит предположению о независимости атрибутов для алгоритма Normal Bayes Classifier. Удалить эти атрибуты нельзя, так как они имеют самую высокую информативность из всех (табл. II). При этом можно заметить, что для классов Amateur и Weather эти атрибуты имеют наиболее сильную зависимость (т.к. мат. ожидание у этих атрибутов наиболее сильно выражено), что приводит к ухудшению классификации.

Также необходимо заметить, что по остальным классам исключение атрибутов привело к существенному улучшению классификации именно для алгоритма Normal Bayes Classifier (особенно для класса Human Spaceflight

(HS). Это объясняется тем, что исключенные атрибуты имели существенную зависимость между собой (табл. 1). Это особенно чувствительно для байесовского алгоритма.

Время построения деревьев решений (Decision tree и Random forest) для набора данных за 1 год значительно больше, чем для K-nn и Normal Bayes Classifier. Это объясняется особенностью последних алгоритмов, которые строят классификаторы за один проход. Это их существенное преимущество при применении к временным данным.

Исключение восьми атрибутов уменьшило общий объем информации примерно в половину, что способствует уменьшению времени работы классификаторов на 12 % (табл. VI).

ТАБЛИЦА VI ВРЕМЯ ОБУЧЕНИЯ МОДЕЛЕЙ

Вариант	DT	K-nn	RF	NBC
14	5.27 сек	0.68 сек	42.56 сек	0.48 сек
6	4.73 сек	0.59 сек	38.21 сек	0.42 сек

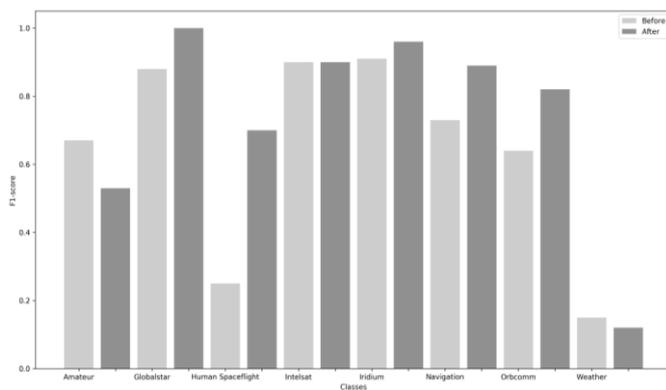


Рис. 2. F1-Score классификации модели Normal Bayes

#### IV. ЗАКЛЮЧЕНИЕ

В ходе работы был исследован набор параметров космических аппаратов за период с 2018 по 2020 год. Произведен анализ влияния атрибутов на решение задачи классификации типа КА. В результате из 14 атрибутов, значимыми являются 6. Экспериментальные исследования для разных алгоритмов показали улучшение классификации при исключении незначимых атрибутов.

Среди исследуемых алгоритмов классификации, наиболее «быстрым» является алгоритм Normal Bayes Classifier. Он выполняет построение классификатора за один проход, что особенно важно при анализе постоянно поступающих данных. Еще одним преимуществом данного алгоритма является возможность дообучения классификатора при поступлении новых данных.

Уменьшение набора данных практически в два раза, позволило сократить время обучения на 12 %. Построение классификатора по представленным данным, без их существенного преобразования, не позволяет корректно

классифицировать все классы. Для Normal Bayes Classifier слабыми классами остаются Amateur и Weather. В дальнейших исследованиях, для улучшения результатов классификации планируется использовать два метода преобразование данных слабых классов: поиск шаблонов поведения и аппроксимация параметров.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Интеллектуальный анализ данных в распределенных системах. МС Куприянов, ИИ Холод, ЗА Каршиев, ИА Голубев - 2012
- [2] Space Track. [Электронный ресурс]. Режим доступа: <https://www.space-track.org/>
- [3] Cotter, L[awrence] R.: Sputniks and Space, The National Space Surveillance Control Center. Presentation before a joint meeting of the New England Society for Paint Technology and The New England Paint, Varnish, and Lacquer Association. Boston: 19 January 1961
- [4] Обзор самых популярных алгоритмов машинного обучения [Электронный ресурс]. Режим доступа: <https://tproger.ru/translations/top-machine-learning-algorithms/>
- [5] Деревья решений: общие принципы. [Электронный ресурс]. Режим доступа: <https://loginom.ru/blog/decision-tree-p1>
- [6] Реализация и обзор метода k-ближайших соседей [Электронный ресурс]. Режим доступа: [http://www.cyberguru.ru/algorithms/algorithms-theory/algorithm-k\\_nearest.html](http://www.cyberguru.ru/algorithms/algorithms-theory/algorithm-k_nearest.html)
- [7] Domingos, Pedro & Michael Pazzani «On the optimality of the simple Bayesian classifier under zero-one loss». Machine Learning 1997
- [8] Accuracy, Precision, Recall or F1? [Электронный ресурс]. Режим доступа: <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>