

Моделирование действий пилота с целью предсказания траектории движения самолета

Над проектом работала: Филатова Анастасия

Куратор проекта: Шпильман Алексей Александрович

Глобальная цель

- Методами машинного обучения научиться предсказывать, как будет меняться скорость и траектория движения самолета в зависимости от окружающих факторов

Глобальные задачи

1. Проанализировать и понять, какие факторы оказывают влияние на принятие пилотом/автопилотом решений об изменении скорости и траектории движения самолета в каждый момент времени, и насколько сильным оказывается это влияние
2. Методами машинного обучения решить задачу моделирования действий пилота в конкретный момент времени в зависимости от окружающих условий

Локальная цель

- В каждый момент времени полета по окружающим данным определить – увеличит, уменьшит или оставит прежней пилот/автопилот скорость самолета
- В терминах машинного обучения – решить задачу классификации по трем классам: 0 – увеличит скорость, 1 – уменьшит скорость, 2 – оставит прежней

Исходные данные

received	callsign	altitude	groundspeed	latitudedegrees	longitudedegrees	flighthistoryid
12.11.2012 1:00	AIP5138	23000	313	46.4799995422363	-111.800003051758	280326212
12.11.2012 1:01	AIP5138	23000	312	46.4500007629395	-111.680000305176	280326212
12.11.2012 1:02	AIP5138	23000	315	46.4199981689453	-111.550003051758	280326212
12.11.2012 1:03	AIP5138	22900	317	46.4000015258789	-111.419998168945	280326212
12.11.2012 1:04	AIP5138	22900	317	46.3699989318848	-111.300003051758	280326212
12.11.2012 1:05	AIP5138	23000	315	46.3499984741211	-111.169998168945	280326212
12.11.2012 1:06	AIP5138	23000	313	46.3300018310547	-111.050003051758	280326212
12.11.2012 1:07	AIP5138	23000	314	46.2999992370605	-110.930000305176	280326212
12.11.2012 1:08	AIP5138	23000	315	46.2799987792969	-110.819999694824	280326212
12.11.2012 1:09	AIP5138	23000	315	46.25	-110.699996948242	280326212
12.11.2012 1:34	AAAY9871	15000	370	35.5800018310547	-114.819999694824	280415619
12.11.2012 1:10	AIP5138	23000	313	46.2299995422363	-110.580001831055	280326212
12.11.2012 1:11	AIP5138	23100	313	46.2000007629395	-110.449996948242	280326212
12.11.2012 1:12	AIP5138	22900	313	46.1800003051758	-110.330001831055	280326212
12.11.2012 1:13	AIP5138	23100	309	46.1500015258789	-110.220001220703	280326212
12.11.2012 1:14	AIP5138	22300	308	46.1300010681152	-110.099998474121	280326212
12.11.2012 1:15	AIP5138	21000	316	46.0999984741211	-109.970001220703	280326212
12.11.2012 1:16	AIP5138	19700	318	46.0699996948242	-109.849998474121	280326212
12.11.2012 1:17	AIP5138	18400	318	46.0499992370605	-109.730003356934	280326212
12.11.2012 1:18	AIP5138	17500	320	46.0299987792969	-109.620002746582	280326212
12.11.2012 1:19	AIP5138	16200	317	46.0	-109.480003356934	280326212
12.11.2012 1:20	AIP5138	14800	310	45.9700012207031	-109.370002746582	280326212
12.11.2012 1:21	AIP5138	13100	311	45.9500007629395	-109.25	280326212

Обработка данных

1. Отбор из исходных данных тех, которые подходят нам для текущей задачи
2. Составление из отобранных данных признаков, которые влияют на изменение скорости
3. Всего получилось 12 признаков. Их можно поделить на три типа:
 - Связанные с текущим местоположением (высота, координаты, текущая скорость, ...)
 - Связанные с погодными условиями (давление, скорость и направление ветра, порывы ветра, ...)
 - Связанные с планом полета (предполагаемое время прибытия, расстояние до аэропорта прибытия, ...)

Погодные данные

- Необходимо соотнести метеосводки и сведения о текущем местоположении самолета

metar_reports_id	weather_station_code	date_time_issued	wind_direction	wind_speed	wind_gusts	visibility	temperature
209957316	KTBN	2012-11-13 08:55:00+00:00	230.0	3.4500000000000002		10.0	28.0
209957315	KMUI	2012-11-13 08:55:00+00:00	280.0	10.359999999999999		8.0	42.0
209957314	KHRT	2012-11-13 08:55:00+00:00	360.0	16.109999999999999	21.870000000000001	10.0	50.0
209957313	KHOP	2012-11-13 08:55:00+00:00	0.0	0.0		10.0	30.0
209957312	KFRI	2012-11-13 08:55:00+00:00	0.0	0.0		10.0	24.0
209957309	KSZL	2012-11-13 08:55:00+00:00	200.0	4.5999999999999996		10.0	26.0
209957308	KRND	2012-11-13 08:55:00+00:00	30.0	10.359999999999999		10.0	48.0
209957307	KSSC	2012-11-13 08:55:00+00:00	150.0	3.4500000000000002		7.0	59.0
209957306	KLHW	2012-11-13 08:55:00+00:00	0.0	0.0		10.0	59.0
209957305	KLFI	2012-11-13 08:55:00+00:00	200.0	4.5999999999999996		10.0	62.0
209957304	KHFF	2012-11-13 08:55:00+00:00	160.0	3.4500000000000002		5.0	57.0
209957303	KGUS	2012-11-13 08:55:00+00:00	250.0	9.2100000000000009		10.0	30.0
209957302	KCOF	2012-11-13 08:55:00+00:00	20.0	4.5999999999999996		9.0	69.0
209957301	KBAD	2012-11-13 08:55:00+00:00	0.0	0.0		10.0	35.0
209865096	MUCM	2012-11-12 09:05:00+00:00	30.0	3.4500000000000002			68.0
209865095	MUCL	2012-11-12 09:05:00+00:00	60.0	14.960000000000001			75.0
209864177	ETHL	2012-11-12 09:05:00+00:00	60.0	3.4500000000000002		1.8600000000000001	44.0

Подготовка погодных данных

1. Получена информация о местоположении метеостанций и приведена в нужный формат
2. Разработан алгоритм, который для каждого самолета в каждый момент времени находит ближайшую метеостанцию и получает актуальный прогноз
3. Давление пересчитано с учетом перепада высот
4. Заполнены пропуски в графах «порывы ветра» и «давление»

Данные из плана полета

- Три признака на основе этих данных:
 - Оставшееся время до посадки
 - Оставшееся расстояние до аэропорта прибытия
 - Признак-индикатор, показывающий хватит ли нам времени, чтобы долететь до аэропорта прибытия с учетом оставшегося времени и текущей скорости
- Для нахождения оставшегося расстояния потребовалось найти информацию о координатах аэропортов и также привести ее в нужный формат

Результаты подготовки данных

- В результате получилась матрица объекты-признаки следующего вида:

	altitude	groundspeed	latitudedegrees	longitudedegrees	wind_dir	wind_speed	wind_gusts	visib	pressure	remaining_time	remaining_distance	is_avg_speed_bigger	action
0	37000	497	42.3699989319	-82.5800018311	290.0	02.24	02.24	06.21	225.82548	538	870.60069	-1	2
1	37000	497	42.3499984741	-82.4000015259	290.0	02.24	02.24	06.21	225.82548	537	861.967276	-1	1
2	37000	492	42.3499984741	-82.2200012207	290.0	02.24	02.24	06.21	225.82548	536	853.523074	-1	2
3	37000	492	42.3499984741	-82.0299987793	210.0	02.24	02.24	05.09	227.78677	535	844.531593	-1	2
4	37000	492	42.3300018311	-81.8199996948	210.0	02.24	02.24	05.09	227.78677	534	834.196126	-1	0
5	37000	498	42.3300018311	-81.6299972534	210.0	02.24	02.24	05.09	227.78677	533	825.017394	-1	2
6	37000	498	42.3199996948	-81.4499969482	210.0	02.24	02.24	05.09	227.78677	532	816.079727	-1	2
7	37000	498	42.3199996948	-81.2699966431	210.0	02.24	02.24	05.09	227.78677	531	807.219324	-1	2
8	37000	498	42.2999992371	-81.0800018311	210.0	02.24	02.24	05.09	227.78677	530	797.428601	-1	2
9	37000	498	42.2799987793	-80.9000015259	210.0	02.24	02.24	05.09	227.78677	529	788.024547	-1	2
10	37000	498	42.2799987793	-80.6999969482	0.0	0.0	0.0	05.09	227.55921	528	777.860377	-1	2
11	37000	498	42.2700004578	-80.5500030518	0.0	0.0	0.0	05.09	227.55921	527	769.972825	-1	1
12	37000	493	42.2299995422	-80.3700027466	0.0	0.0	0.0	05.09	227.55921	526	759.850423	-1	1
13	37000	487	42.2200012207	-80.1800003052	0.0	0.0	0.0	05.09	227.55921	525	749.667793	-1	1
14	37000	482	42.1800003052	-80.0	0.0	0.0	0.0	05.09	227.55921	524	739.259352	-1	0
15	37000	487	42.1699981689	-79.8000030518	0.0	0.0	0.0	05.09	227.55921	523	728.262276	-1	2
16	37000	487	42.1300010681	-79.6299972534	0.0	0.0	0.0	05.09	227.55921	522	718.086283	-1	2
17	37000	487	42.0999984741	-79.4499969482	0.0	0.0	0.0	05.09	227.55921	521	707.429976	-1	1
18	37000	486	42.0800018311	-79.2799987793	0.0	0.0	0.0	05.09	227.55921	520	697.414203	-1	2
19	37000	486	42.0499992371	-79.0999984741	0.0	0.0	0.0	05.09	227.55921	519	686.439212	-1	2
20	37000	486	42.0299987793	-78.9000015259	0.0	0.0	0.0	05.09	227.55921	518	674.37773	-1	1
21	37000	480	42.0	-78.7300033569	0.0	0.0	0.0	05.09	227.55921	517	663.620235	-1	2
22	29800	469	34.6800003052	-105.029998779	0.0	0.0	0.0	0.0	270.63756	540	874.608272	-1	2

Задача классификации

- Для получения результатов необходимо было применить методы машинного обучения на подготовленных данных
- Был выбран алгоритм Random Forest, а именно его реализация из библиотеки scikit-learn – RandomForestClassifier
- Random Forest алгоритм – один из довольно простых, но в то же время действенных алгоритмов, показывающих хорошие результаты

Результаты

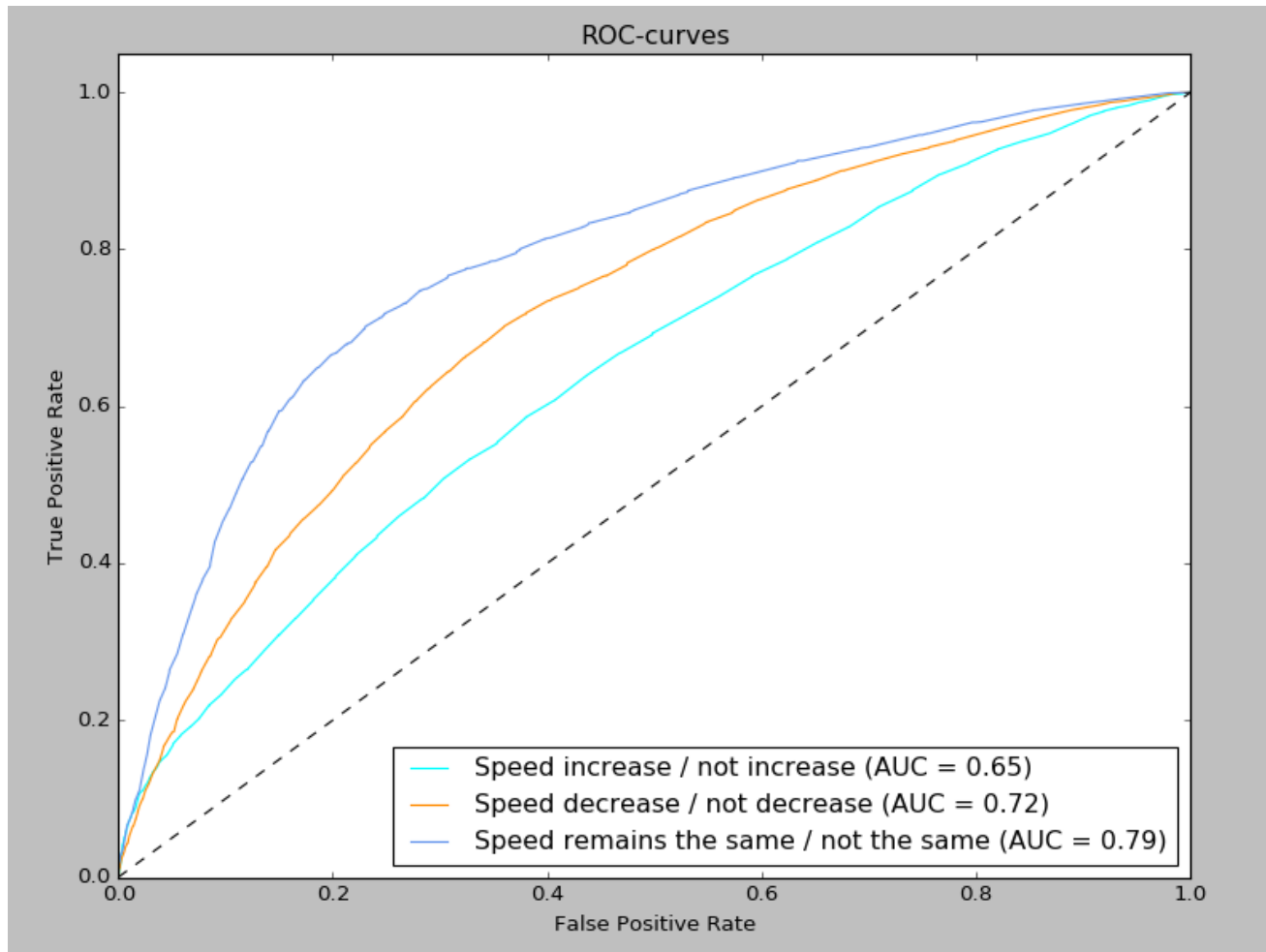
- Алгоритм обучался на тренировочной выборке, состоящий из 300.000 объектов и тестировался на выборке, состоящей из 10.009 объектов

Предсказание \ Реальные данные	Скорость увеличилась	Скорость уменьшилась	Скорость осталась прежней
Скорость увеличилась	1273	1458	356
Скорость уменьшилась	1124	3371	296
Скорость осталась прежней	584	690	857

Точность и полнота результатов

	Скорость увеличилась	Скорость уменьшилась	Скорость осталась прежней
Точность	42.7%	61%	56.8%
Полнота	41.7%	70.4%	40.2%

Результаты бинарных классификаторов



Планы на следующий семестр

- Улучшить текущие результаты, доработать признаки, влияющие на изменение скорости
- Продолжить решение глобальной задачи

ССЫЛКИ

- Git-репозиторий с кодом и подробным описанием:
<https://github.com/nast1415/machine-learning>