

Предсказание преступлений в городе Чикаго

Тлямов Анвар, Галоев Илькин

ИТМО

2022



MASTER OF
SOFTWARE
ENGINEERING



ITMO UNIVERSITY

Актуальность темы

- Рост преступности в городе Чикаго в некоторых районах города
- Хотим прогнозировать место, тип и время преступности
- Релевантность предсказания для департамента полиции города Чикаго



Цели и задачи

Цель работы

Обучение модели предсказания типа, места и времени преступления

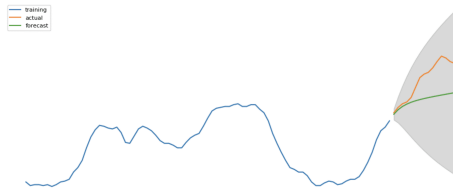
Задачи:

- Анализ и обзор существующих моделей
- Поиск датасета преступности в каком-то городе
- Обучение модели
- Подсчет метрик получившейся модели

Анализ существующих моделей

Найденные модели:

- класс статистических моделей для анализа и прогнозирования данных временных рядов.
- ARIMA – AutoRegressive Integrated Moving Average.
- AutoReg, ARMA
- Prophet от Meta(Facebook)
- LSTM



Сбор датасета

- Данные о преступлениях в Сент-Луисе с веб-сайта Департамента столичной полиции Сент-Луиса
- 449К данных
- данные собраны примерно с 2000 года

	DateOccured	Neighborhood	UCRCrime
0	2000-01-01	53	Sex Offenses
1	2003-01-01	17	Larceny-theft
2	2004-01-01	68	Fraud
3	2006-01-01	16	Fraud
4	2007-01-01	57	Other assaults
...
477636	2013-12-01	5	Larceny-theft
477637	2013-12-20	65	Forcible Rape
477638	2014-12-22	14	Other assaults
477639	2014-12-26	60	Robbery
477640	2014-12-29	1	Other

449000 rows × 3 columns

Модель prophet

- Prophet – это процедура прогнозирования данных временных рядов, которая учитывает влияние праздников на прогноз.
- Основана на аддитивной модели

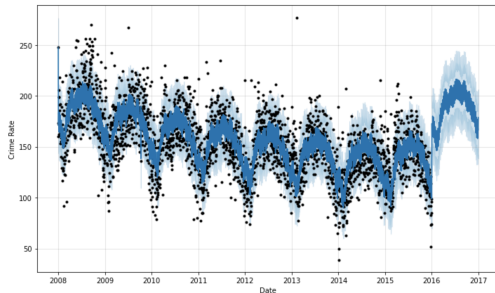
	DS	Y
0	2007-12-10	9.590761
1	2007-12-11	8.519590
2	2007-12-12	8.183677
3	2007-12-13	8.072467
4	2007-12-14	7.893572



Обучение модели

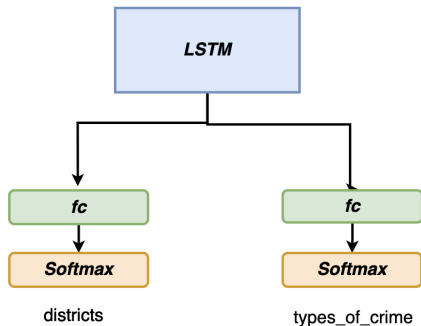
- разделили на train/val по последнему году.
- По последнему году делали тестирование
- В качестве target пытались засунуть one hot
- Кодирование места и типа в число

	DS	Y
0	2007-12-10	9.590761
1	2007-12-11	8.519590
2	2007-12-12	8.183677
3	2007-12-13	8.072467
4	2007-12-14	7.893572



Модель

- Вход LSTM подаем его параллельно на два полносвязных слоя
- Первый слой длины количества районов
- Второй от количества типов преступлений

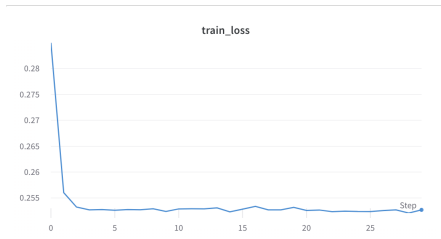


Второй датасет

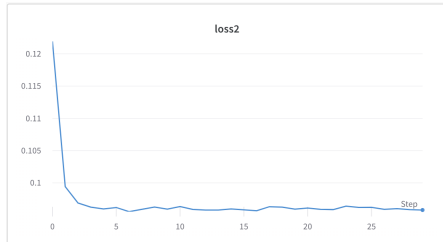
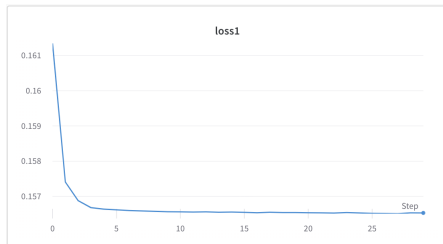
- собран с 2001 года
- 7 миллионов данных
- постоянное обновление датасета
- Взяли время, тип и место преступления из всего датасета

	Date	Primary Type	District
0	2015-09-05	BATTERY	9.0
1	2015-09-04	THEFT	15.0
2	2018-09-01	THEFT	6.0
3	2015-09-05	NARCOTICS	14.0
4	2015-09-05	ASSAULT	15.0
...
7470112	2022-01-06	WEAPONS VIOLATION	5.0
7470113	2022-01-06	ASSAULT	22.0
7470114	2022-01-06	MOTOR VEHICLE THEFT	14.0
7470115	2022-01-06	THEFT	19.0
7470116	2022-01-06	BATTERY	4.0

Результаты обучения

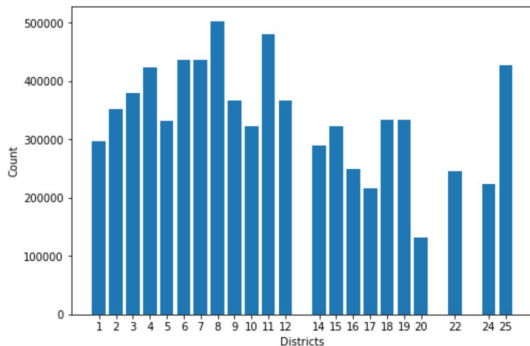


- Результаты двух loss, отвечающих за район и тип преступления соответственно.
- trainLoss – суммарный результат



Балансировка датасета

- Некоторые районы имеют б'ольшую преступность по сравнению с другими
- Решили сбалансировать датасет по количеству преступлений на район.



	Date	Primary Type	District	UCRCrime_mapping
0	2001-01-01	THEFT	9.0	0
1	2001-01-01	SEX OFFENSE	14.0	1
2	2001-01-01	OTHER OFFENSE	24.0	2
3	2001-01-01	OFFENSE INVOLVING CHILDREN	8.0	3
4	2001-01-01	BATTERY	1.0	4
...
2199995	2022-01-06	THEFT	17.0	0
2199996	2022-01-06	BATTERY	22.0	4
2199997	2022-01-06	THEFT	18.0	0
2199998	2022-01-06	MOTOR VEHICLE THEFT	1.0	7
2199999	2022-01-06	CRIMINAL DAMAGE	8.0	9

2200000 rows x 4 columns

Результаты

- На картинке выход из softmax
- Мы превращаем его в one-hot вектор и получаем нужный район
- Также мы превращаем в one-hot вектор нужный тип преступления, а потом тип вытаскиваем через специальную mapку

```
tensor([0.0109, 0.0462, 0.0424, 0.0389, 0.0440, 0.0493, 0.0468, 0.0411, 0.0452,  
        0.0401, 0.0447, 0.0407, 0.0452, 0.0122, 0.0410, 0.0461, 0.0429, 0.0407,  
        0.0406, 0.0430, 0.0415, 0.0120, 0.0447, 0.0107, 0.0432, 0.0460],  
        device='cuda:0', grad_fn=<SelectBackward0>)
```

```
[61] reverse_mapping[int(torch.argmax(prob2[0][-1]))]
```

```
'NARCOTICS'
```

Итог

Итого:

- Провели анализ и обзор существующих моделей
- Нашли датасеты преступности для обучения
- Обучили модель prophet и модель LSTM
- Подсчитали метрики для модели prophet