Искусственный интеллект глазами разработчика

Конференция
для IT-специалистов,
интересующихся темой
Artificial Intelligence

МИР

МИЛАНДР
ГРУППА КОМПАНИЙ

Module
НАУЧНО-ТЕХНИЧЕСКИЙ ЦЕНТР

НовГУ им.
Ярослава Мудрого

Искусственный интеллект глазами разработчика

29 мая 2021

МИР

Мастерская инструментов разработки

mir.dev

Распознавание потребителей в распределенной энергосистеме

Как скрытые Марковские модели распознают устройства по данным о потребляемой мощности

МИЛАНДР
ГРУППА КОМПАНИЙ

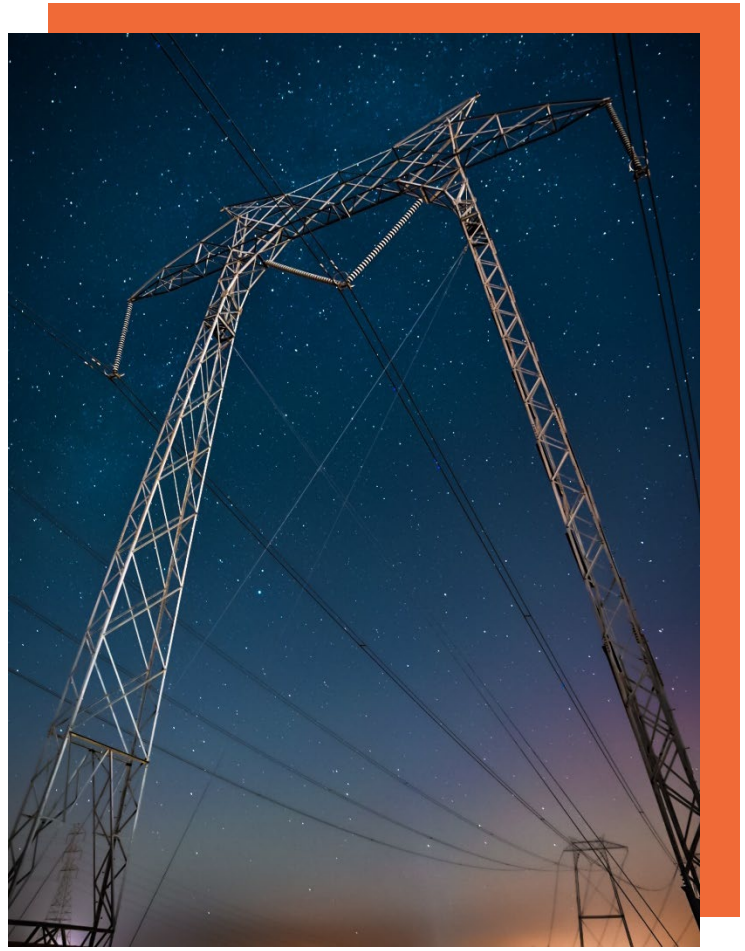
Андрей Обухов

Стажер-исследователь центра проектирования программного обеспечения

Зачем все это нужно

Экономия электроэнергии важна как для предприятий, так и для домохозяйств

Чтобы потреблять меньше, нужно понимать, какие приборы больше всего "едят"



Данные

С высокой частотой

- несколько кГц
- распознавание по периоду
- нейросети

Со средней частотой

- порядка 1 Гц
- моделируем состояния устройств (вкл/выкл)
- скрытые Марковские модели

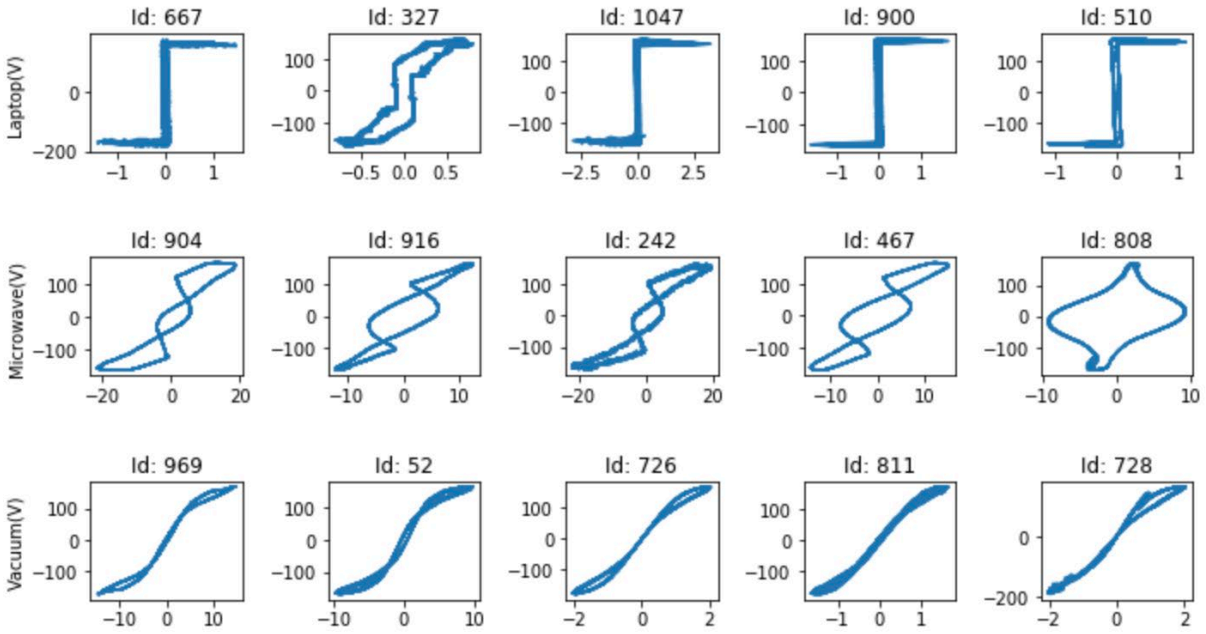
Со низкой частотой

- период порядка минуты
- больше подходит для предсказаний трендов
- статистические модели

Высокочастотные данные

Например, датасет PLAID

- Данные собраны на частоте 30 кГц
- Распознавание на основе данных за период (600 точек)



https://github.com/jingkungao/PLAID/blob/master/ParseData.ipynb

Сверточные нейросети

1. Строится график за несколько периодов
2. Сжимается до изображения 50x50
3. Подается на вход нейронной сети

Layer	Kernel/Pooling Window	Layer Size
Input	-	1@50X50
Convolution—stride 1 (C1)	[4@3X3]	4@48X48
Pooling—stride 2 (P1)	[4@2X2]	4@24X24
Convolution—stride 1 (C2)	[6@3X3]	6@22X22
Pooling—stride 2 (P2)	[6@2X2]	6@11X11
Convolution—stride 1 (C3)	[18@3X3]	18@9X9
Full out (F1)	-	11

researchgate.net/publication/327710486_Implementation_Strategy_of_Convolution_Neural_Networks_on_Field_Programmable_Gate_Arrays_for_Appliance_Classification_Using_the_Voltage_and_Current_V-I_Trajectory

Точность для различных классов устройств

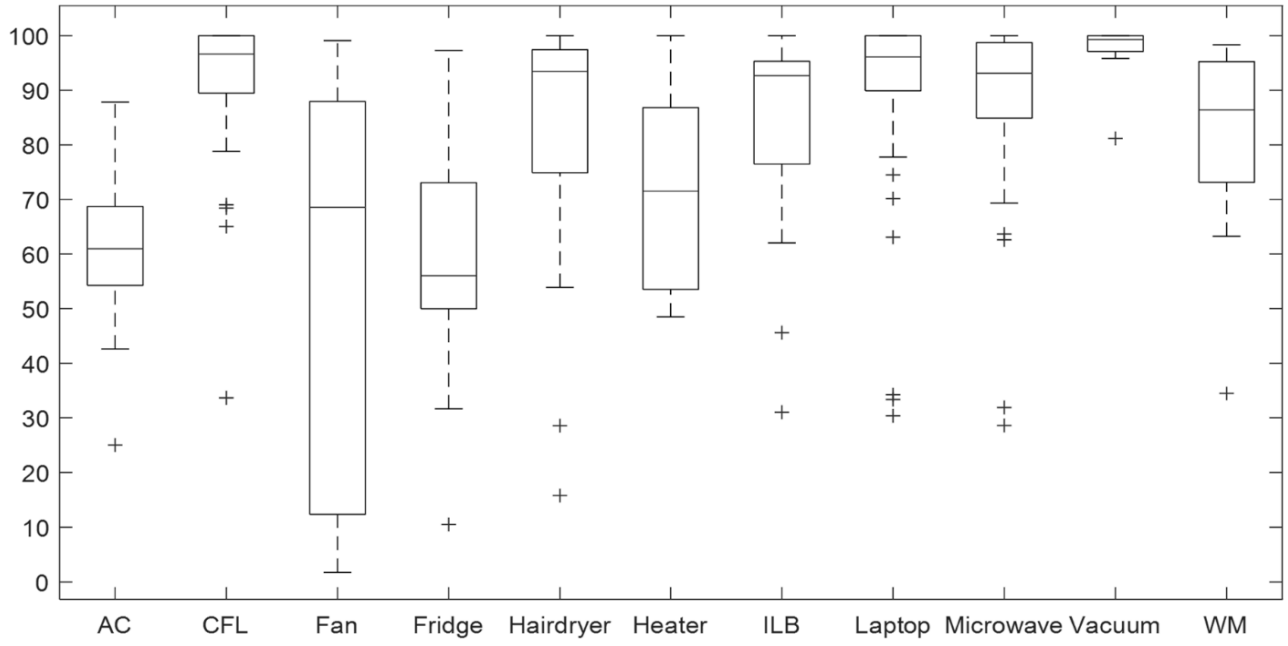


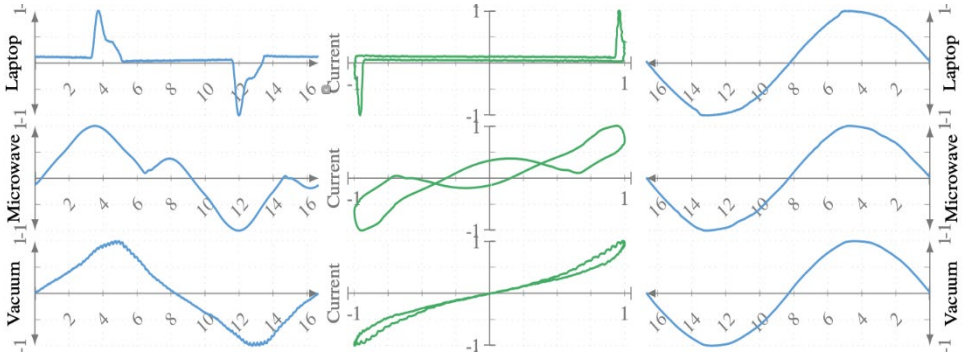
Figure 10. Box and whisker plot for PLAID 1: x axis indicating the different appliances and y axis is the F-score in percentage (%).

researchgate.net/publication/327710486_Implementation_Strategy_of_Convolution_Neural_Networks_on_Field_Programmable_Gate_Arrays_for_Appliance_Classification_Using_the_Voltage_and_Current_V-I_Trajectory

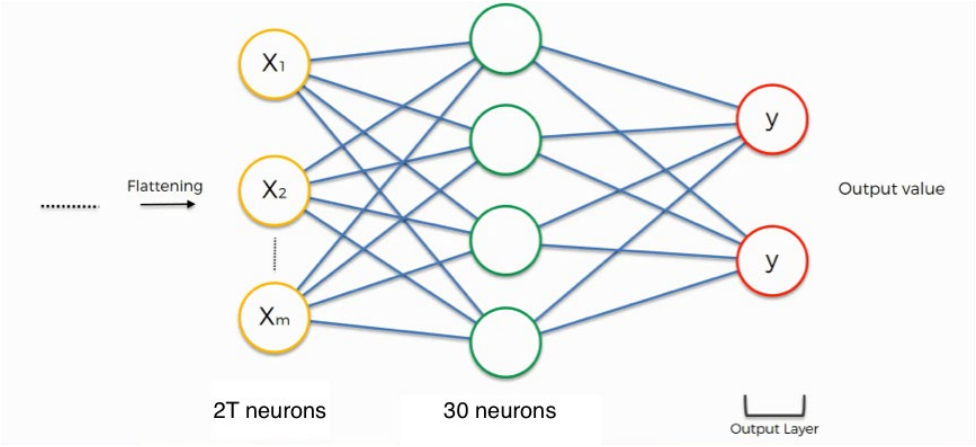
Теперь без картинок и сверток

На вход нейросети подаются
сырые данные тока
и напряжения за период

Ансамбль из двухслойных
полносвязных нейросетей,
каждая из которых решает
задачу бинарной
классификации, дает точность
88% (на датасете PLAID).



nilmworkshop.org/2016/proceedings/Paper_ID09.pdf



superdatascience.com/blogs/convolutional-neural-networks-cnn-step-4-full-connection

В чем проблемы

1

Большой объем данных из-за высокой частоты сбора
30000 * 2 = 60000 значений/сек

2

Что будет, если в розетку-измеритель включены
N > 1 потребителя (например, через удлинитель)?

Данные с низкой частотой

1

Частота сбора данных – минута и больше

2

Из статей видно, что используются для долгосрочных прогнозов потребления крупных предприятий/районов/ целых государств

Об ARMA-моделях (AR)

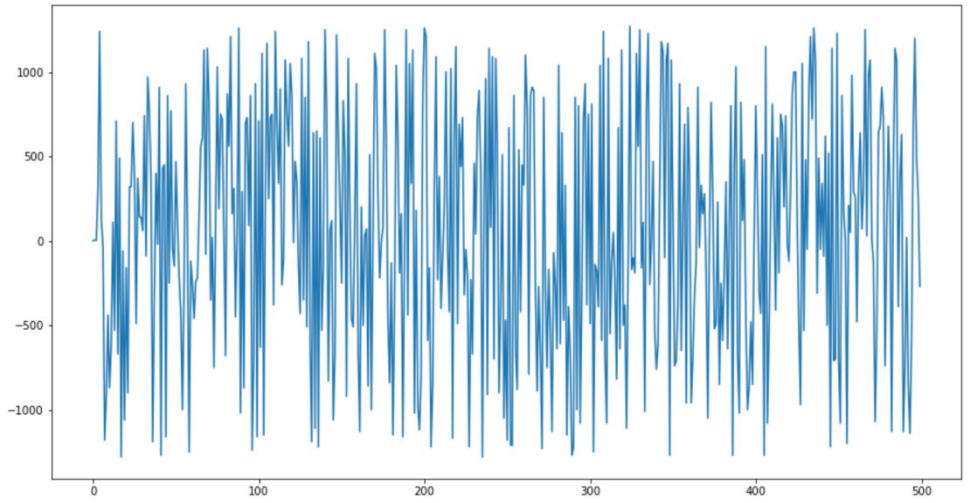
AR - авторегрессия - описываем текущий член ряда предыдущими, взятыми с некоторыми весами

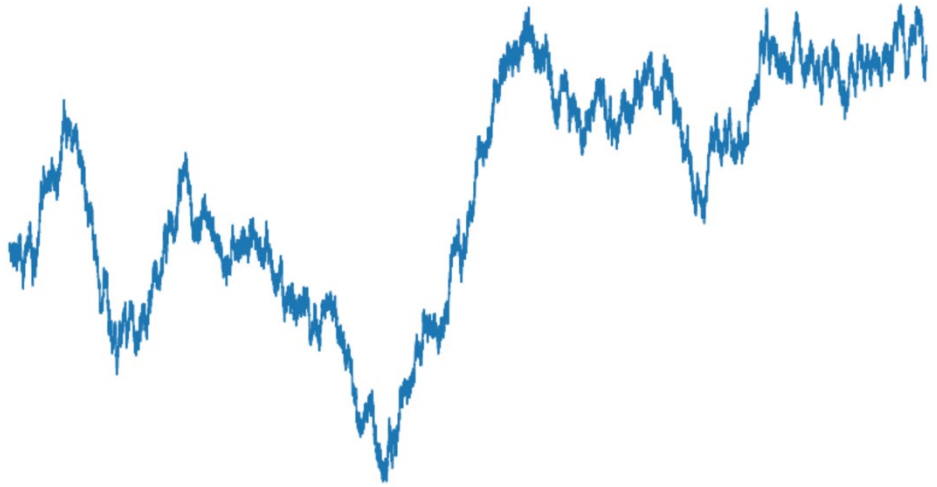
$$X_t = c + a_1 X_{t-1} + a_2 X_{t-2} + \varepsilon_t$$

Об ARMA-моделях (MA)

MA - скользящее среднее (moving average) – учитываем предыдущие отклонения с некоторыми весами

$$X_t = \sum_{j=0}^{q} b_j \varepsilon_{t-j}$$





Об ARMA-моделях

Каждый член ряда - сумма:

- константы c (что-то вроде матожидания)
- случайного отклонения e
- предыдущих p членов Y (с некоторыми весами) - AR
- предыдущих q случайных отклонений e (с весами) - MA

$$ARMA(p,q) : Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

Для каких задач хороши

- Понятны - модели интерпретируемы (можно объяснить и настроить зависимость от прошлых членов), в отличие от нейросетей
- Сравнительно легки вычислительно
- Предсказание долгосрочных трендов потребления крупных объектов (заводов, районов города, стран)

В чем проблемы

1

Если снимать мгновенные значения, устройства могут включиться и выключится, и это никак не зафиксируется в показаниях

2

Если снимать сумму или среднее значение за период, невозможно понять, что включалось (пылесос на 2 минуты или микроволновая печь на одну)

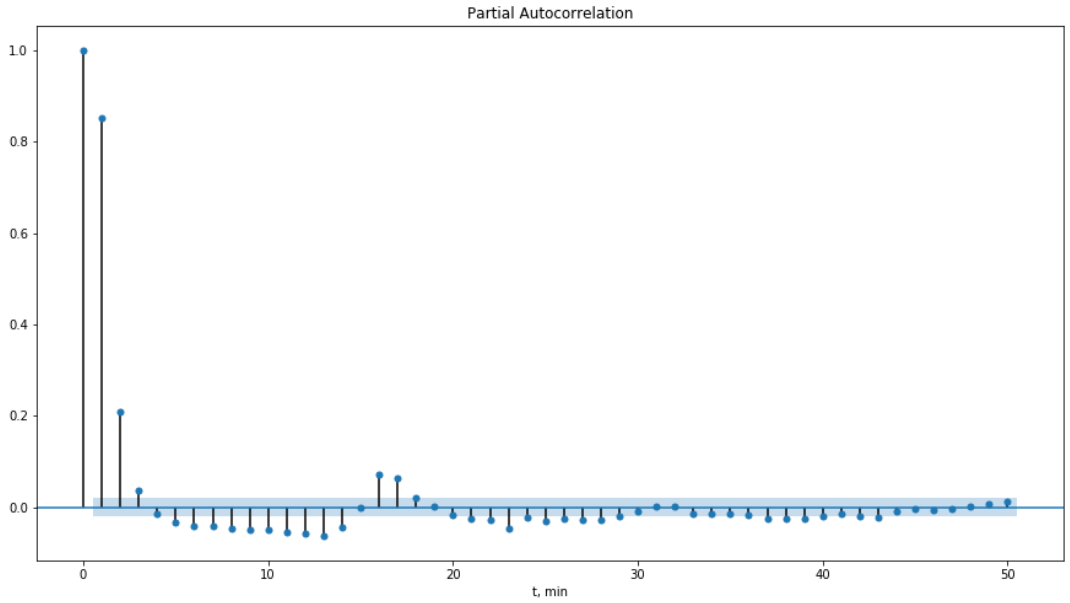
Данные со средней частотой

Показания снимаются
с периодом порядка 1 секунды

Пример: датасет REDD (1-3 сек)

Почему не применимы:

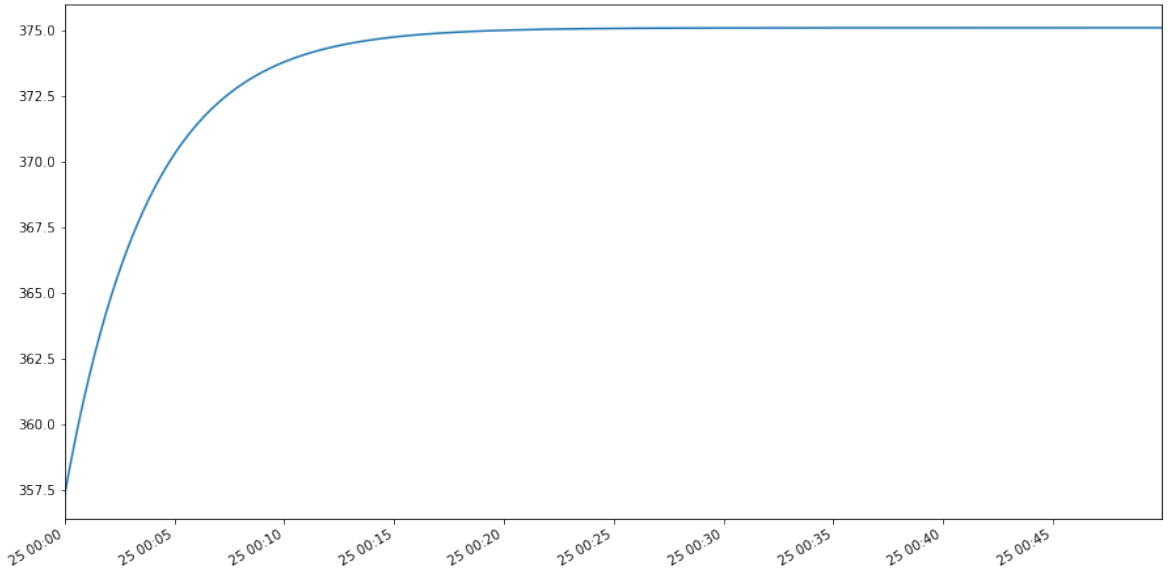
- нейросети: невозможно уловить структуру за период
- авторегрессии: слишком тяжеловесные модели



Прогнозирование авторегрессионной моделью

Какой бы ни была модель, картина предсказаний примерно следующая:

Не очень подходит для нашей задачи, т.к. у нас есть данные об общем потреблении, а статистические модели предсказывают только на основе прошлого



Данные со средней частотой

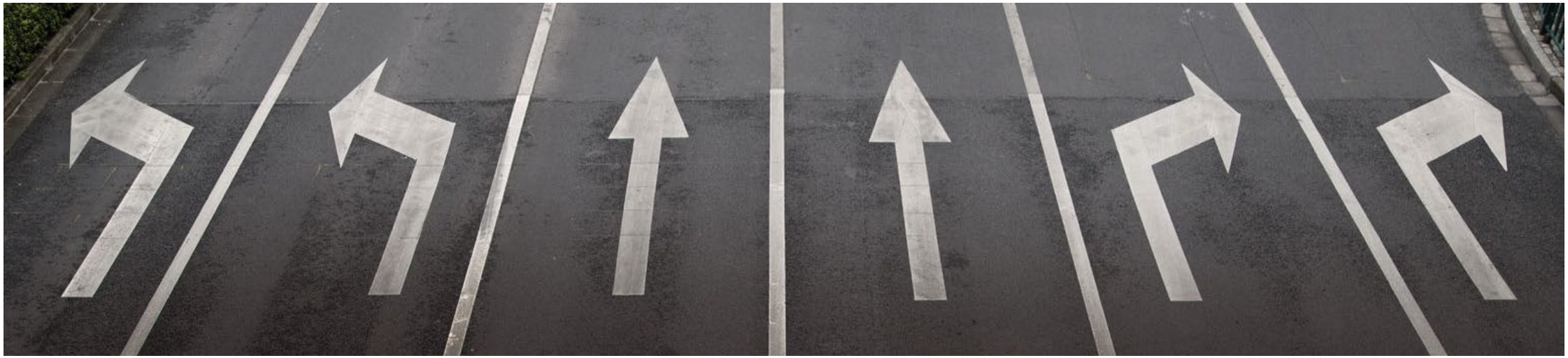
Идея: помнить не предыдущие N показаний устройств,
а просто состояние: включено/выключено

Скрытые Марковские модели

Марковские модели

Основное свойство: состояние на следующем шаге зависит только от текущего состояния; не важно, как попали в него

Например, то, куда машина поедет на перекрестке, зависит от ряда, в котором она стоит, и не зависит от предыдущих перестроений



Марковские модели

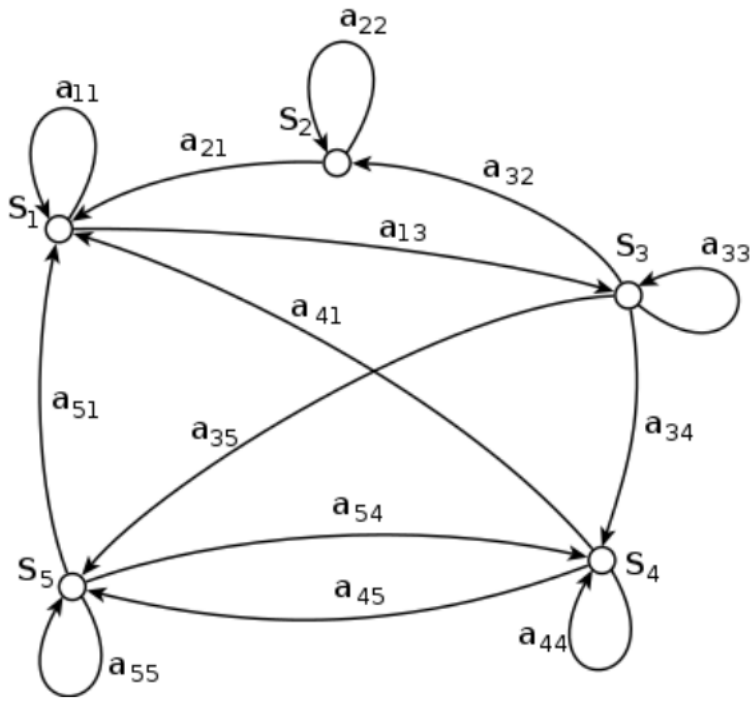
Идея: если долго наблюдать за перекрестком, начнем замечать, из какого ряда куда едут машины

Можно говорить о вероятностях того, из какого ряда в каком направлении едут машины

Марковские модели

Состояние на следующем шаге зависит только от текущего состояния; не важно, как попали в него

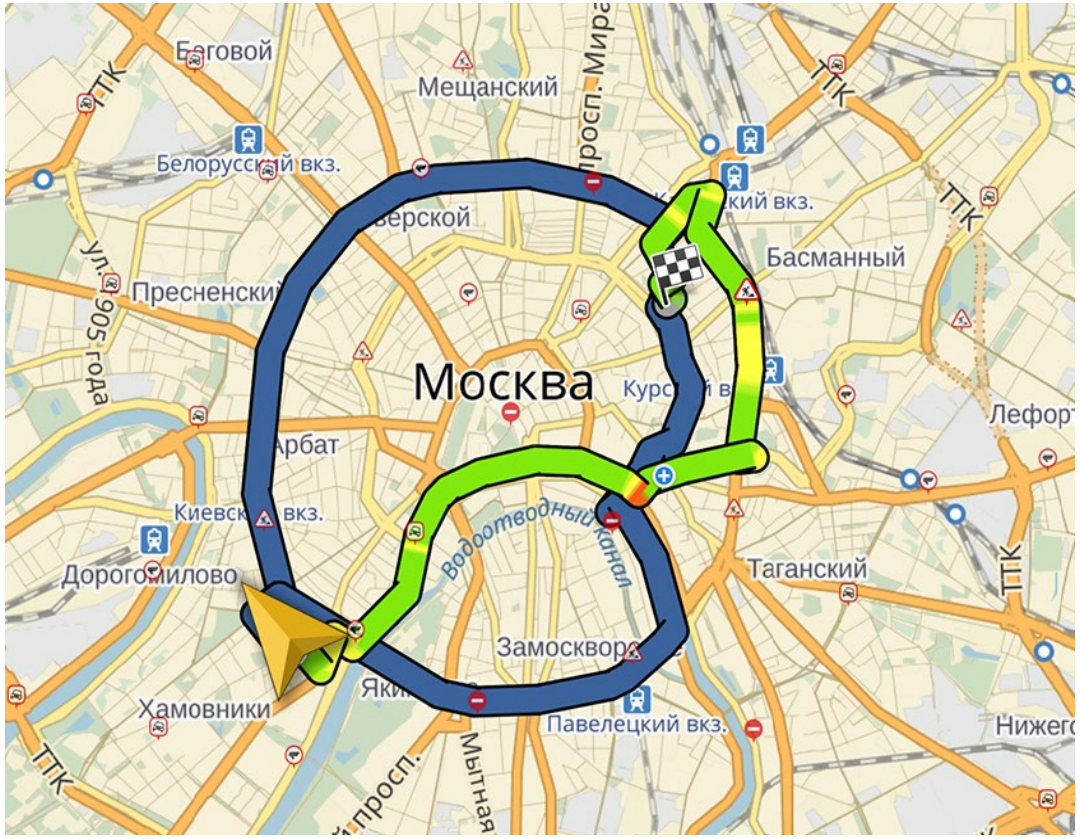
Переходы между состояниями i, j определяются вероятностями a{i, j}



Скрытые Марковские модели

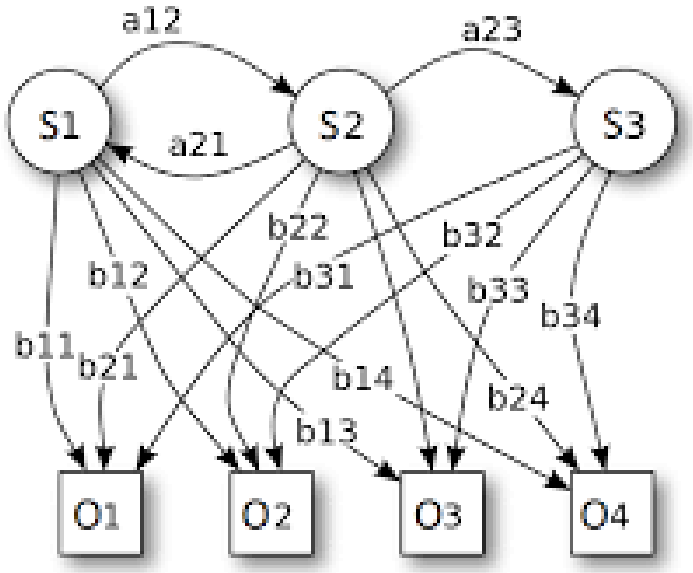
Теперь не видим машину, знаем, что она проехала сколько-то перекрестков и приехала в конечную точку

Наблюдение – проезд перекрестка в каком-то направлении – знаем
Состояние – нахождение в ряду на дороге – только догадываемся



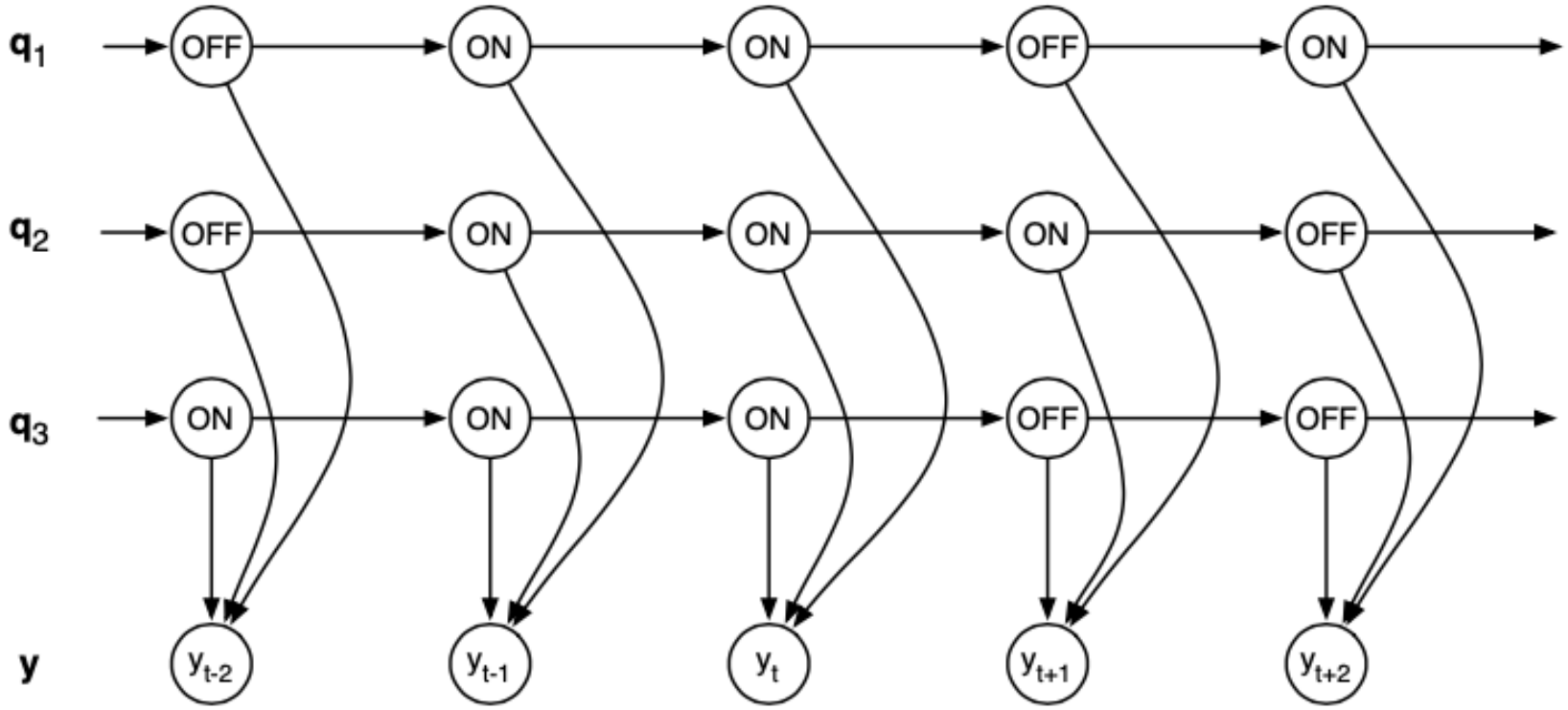
Скрытые Марковские модели

- Не видим систему (ее состояния S{i})
- Наблюдаем только O{k}
- Каждое состояние состояния S{i} с некоторой вероятностью b{i, k} может "выдавать" наблюдение O{k}



https://en.wikipedia.org/wiki/Hidden_Markov_model#/media/File:HiddenMarkovModel.svg

Применимо к устройствам

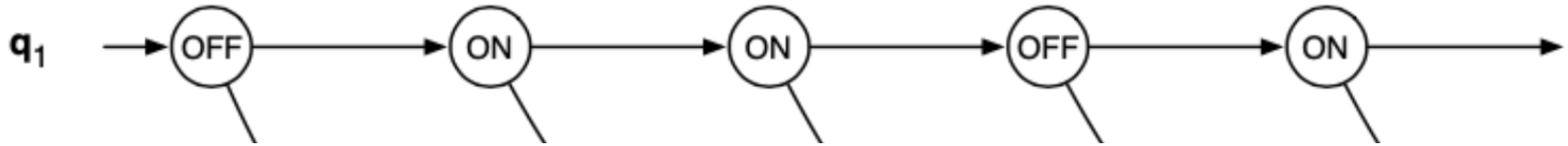


http://hanj.cs.illinois.edu/pdf/sdm11_hkim.pdf

Скрытые Марковские модели

Состояния системы =
= совокупность состояний устройств =
= уровни потребления энергии
в зависимости от режима работы

Алгоритм обучения и распознавания
был нами реализован



Важное предположение:
Каждое устройство - отдельная скрытая Марковская модель
Совместное потребление - просто сумма

Иначе: 10 устройств, каждое вкл/выкл -
получим 1024 возможных состояний

Такое предположение упрощает модель (нужно расчитывать на несколько порядков меньше состояний и вероятностей переходов) и позволяет добавлять устройства без сверхусилий

Описание устройств в модели

1

Количество состояний (об этом далее)

2

Среднее значение потребления электроэнергии в каждом из состояний

3

Матрица вероятностей переходов между состояниями

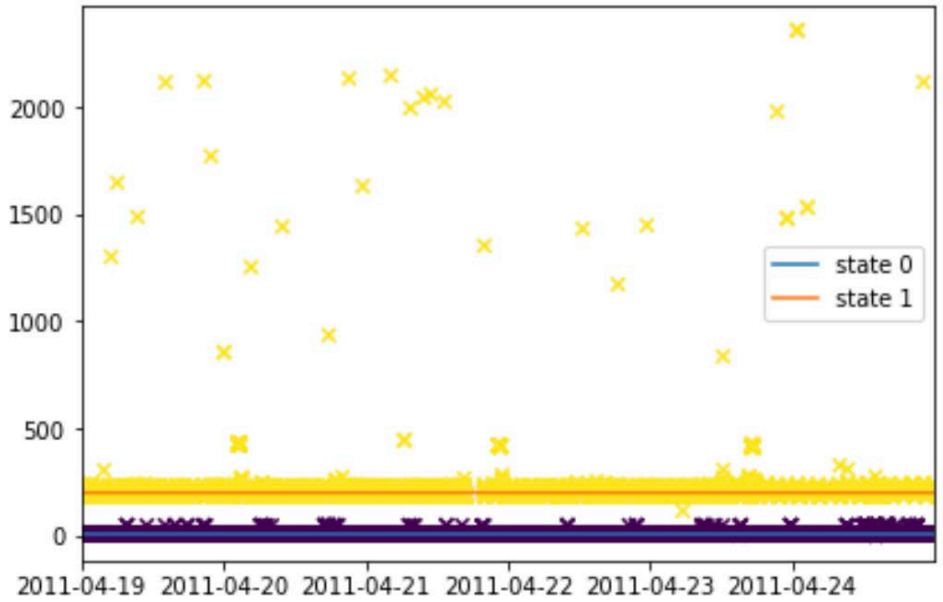
Регистрация устройств

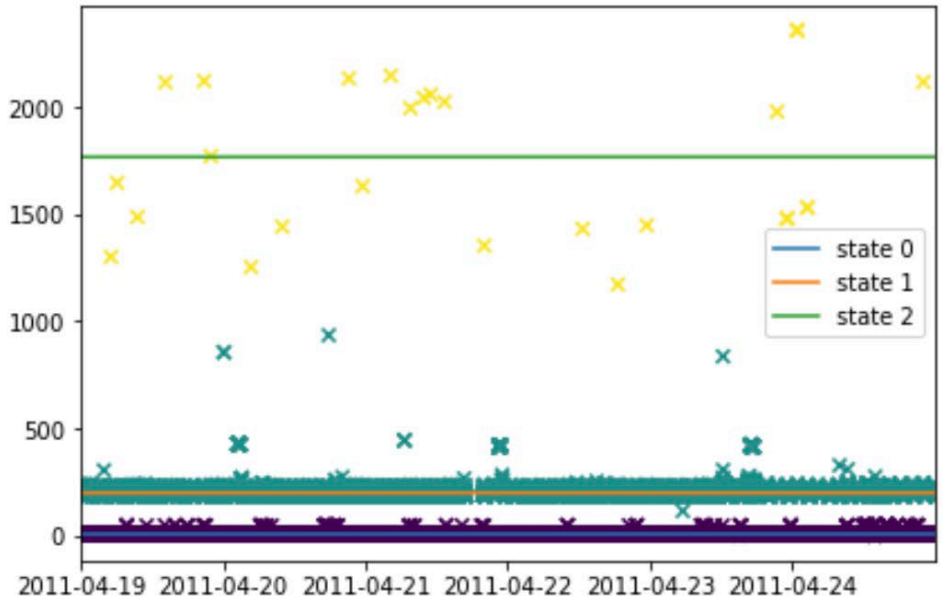
Алгоритм регистрации новых устройств:

1. Включаем устройство в "специальную" розетку
2. Моделируем работу во всех режимах работы
   (за это время собираем информацию о состояниях)
3. Добавляем модель с рассчитанными параметрами
   в словарь устройств
4. Используем устройство как обычно

Количество состояний

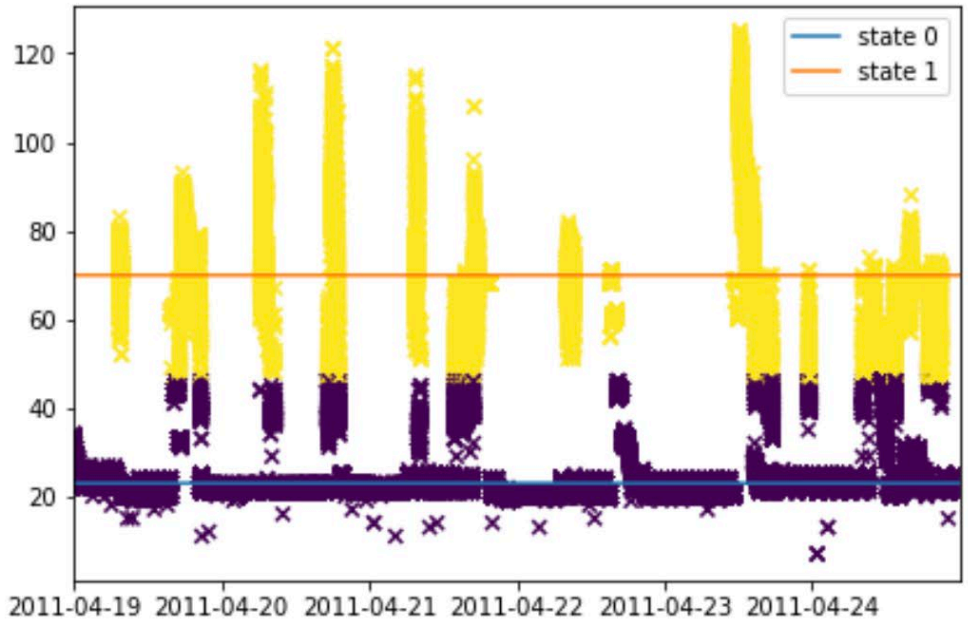
Данные для холодильника, алгоритм k-means,
число кластеров: 2, 3

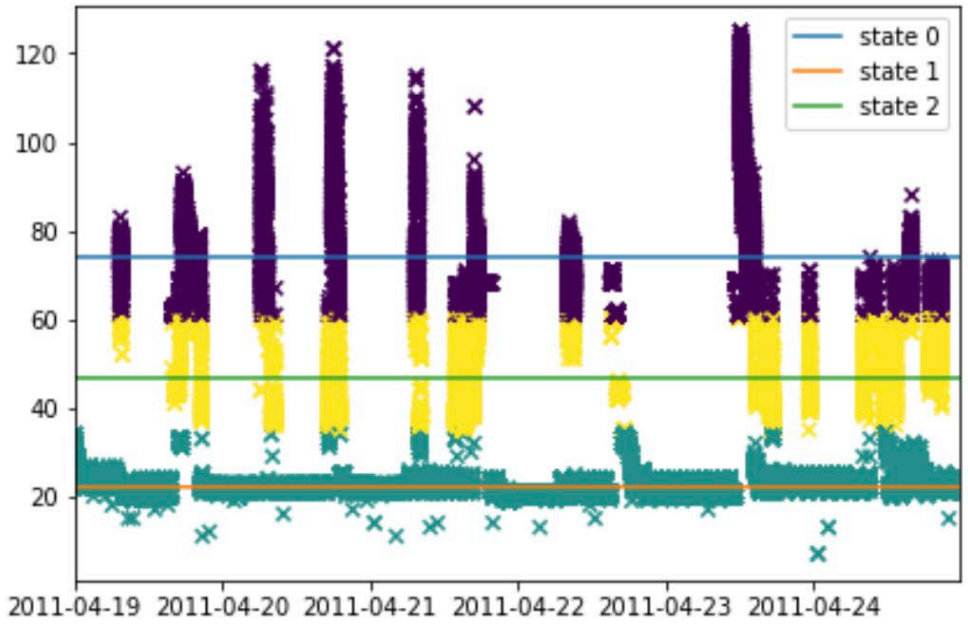




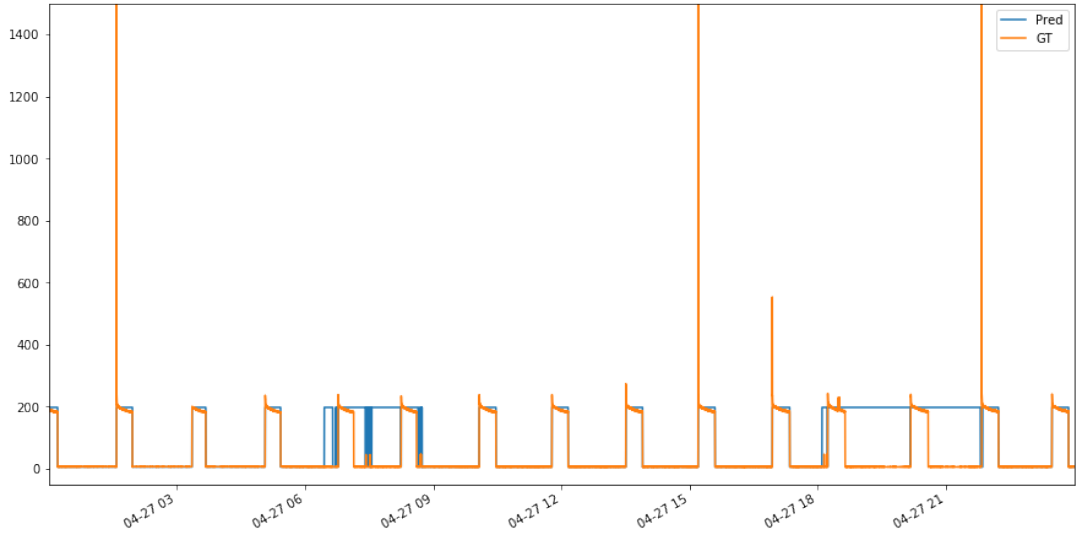
Количество состояний

Данные для розетки, алгоритм k-means,
число кластеров: 2, 3

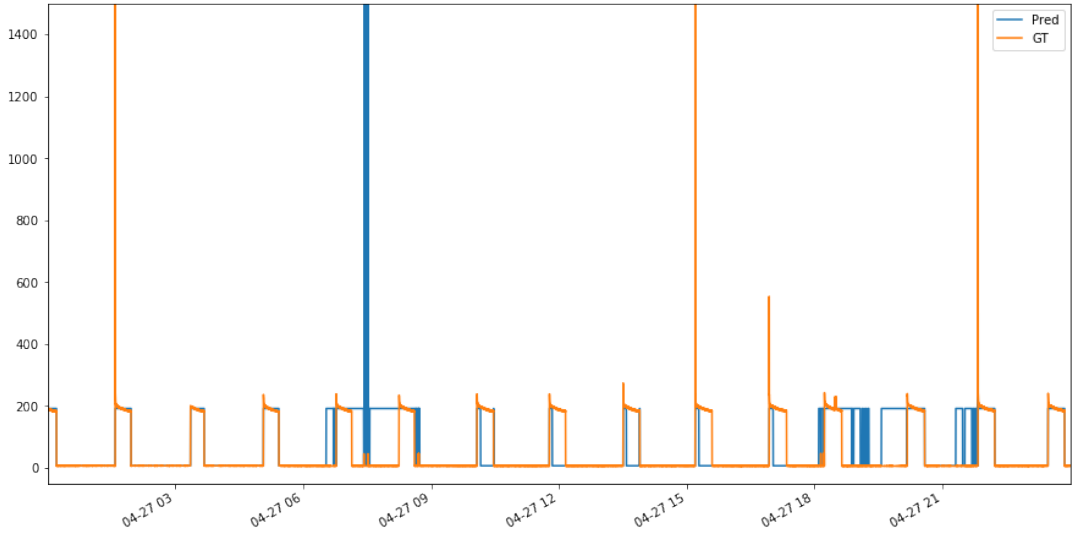




Количество состояний

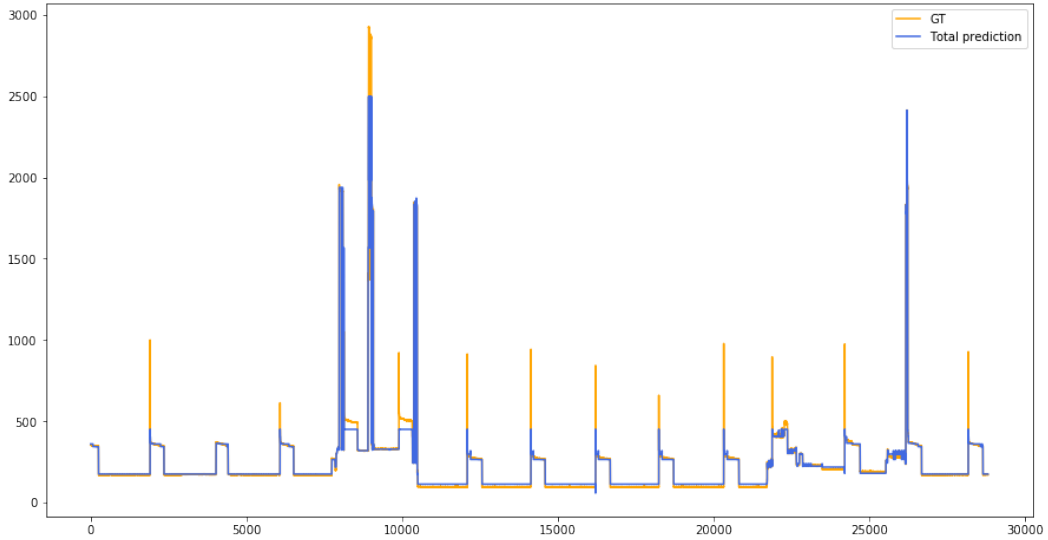


Предсказание потребления холодильника, модель обучена на 2 состояния

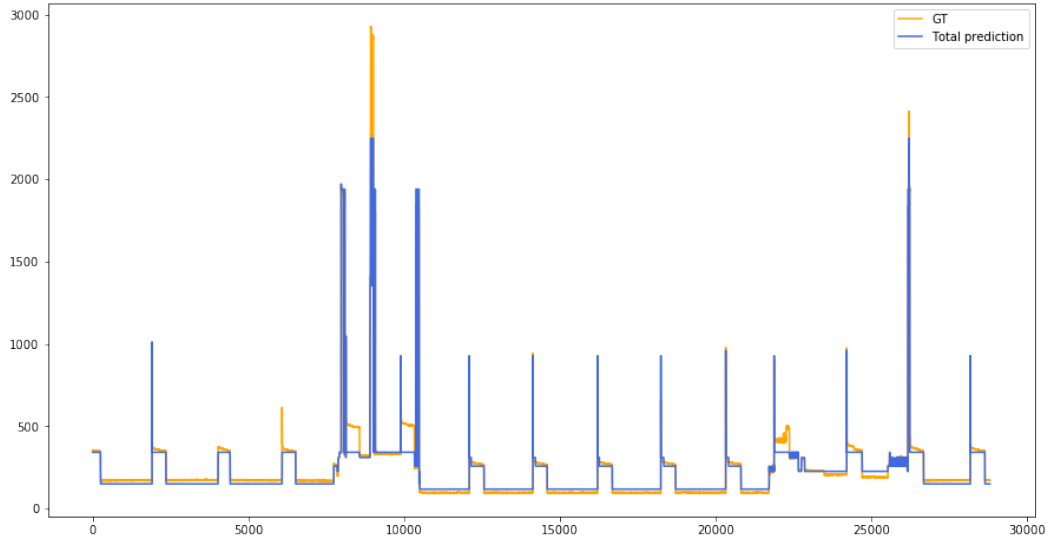


Предсказание потребления холодильника, модель обучена на 3 состояния

Проблема количества состояний



Модель обучена на 2 состояния на каждое устройство



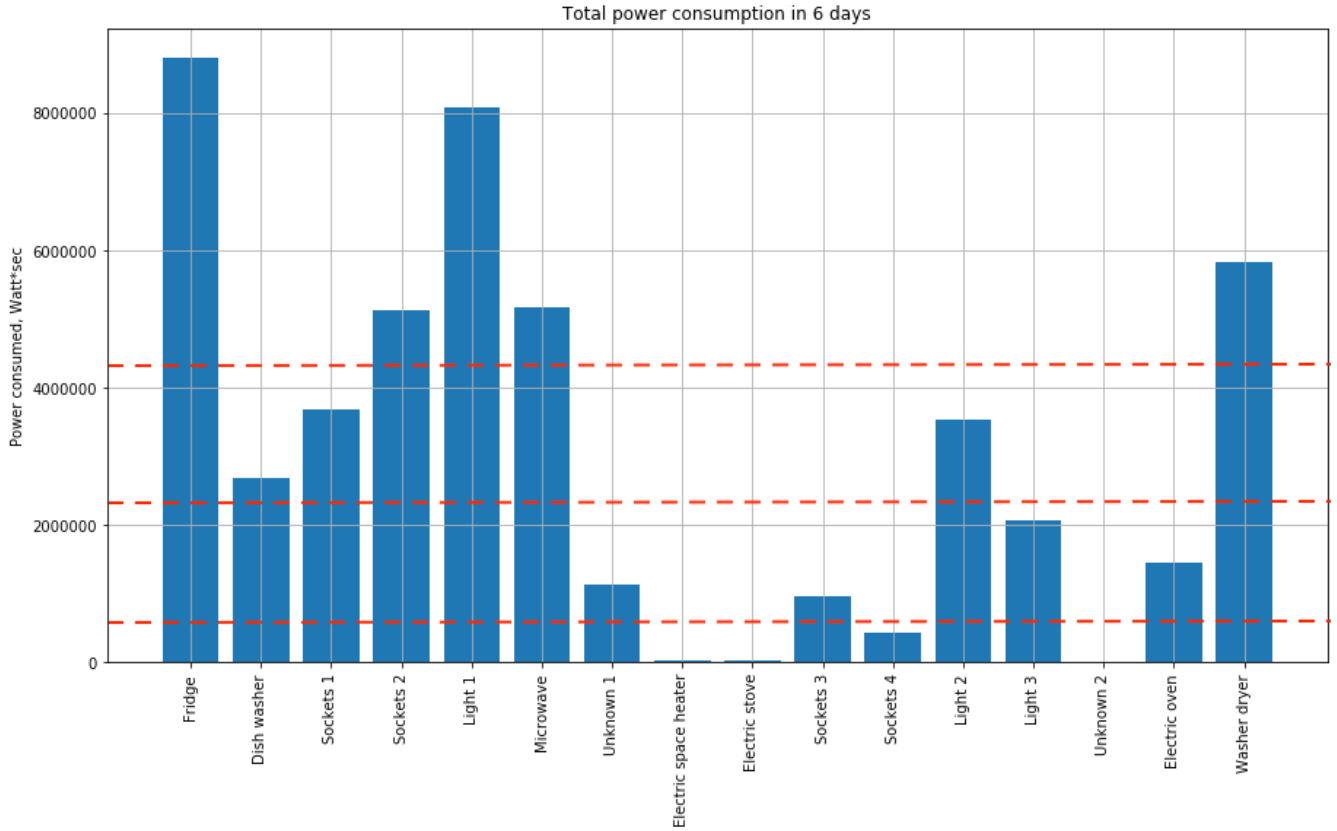
Модель обучена на 3 состояния на каждое устройство (Лучше «подгоняет» пики)

Проблема количества устройств

А сколько устройств стоит учитывать?

Меньше устройств – больше ошибок из-за «незнания»

Больше устройств – больше ошибок из-за сложности модели



Выявление аномалий в работе

1

Модель хранит информацию о типичных параметрах потребления энергии в различных режимах работы

2

Также можно посчитать примерное время нахождения в этих "правильных" состояниях, и учитывать его, если разброс не слишком велик

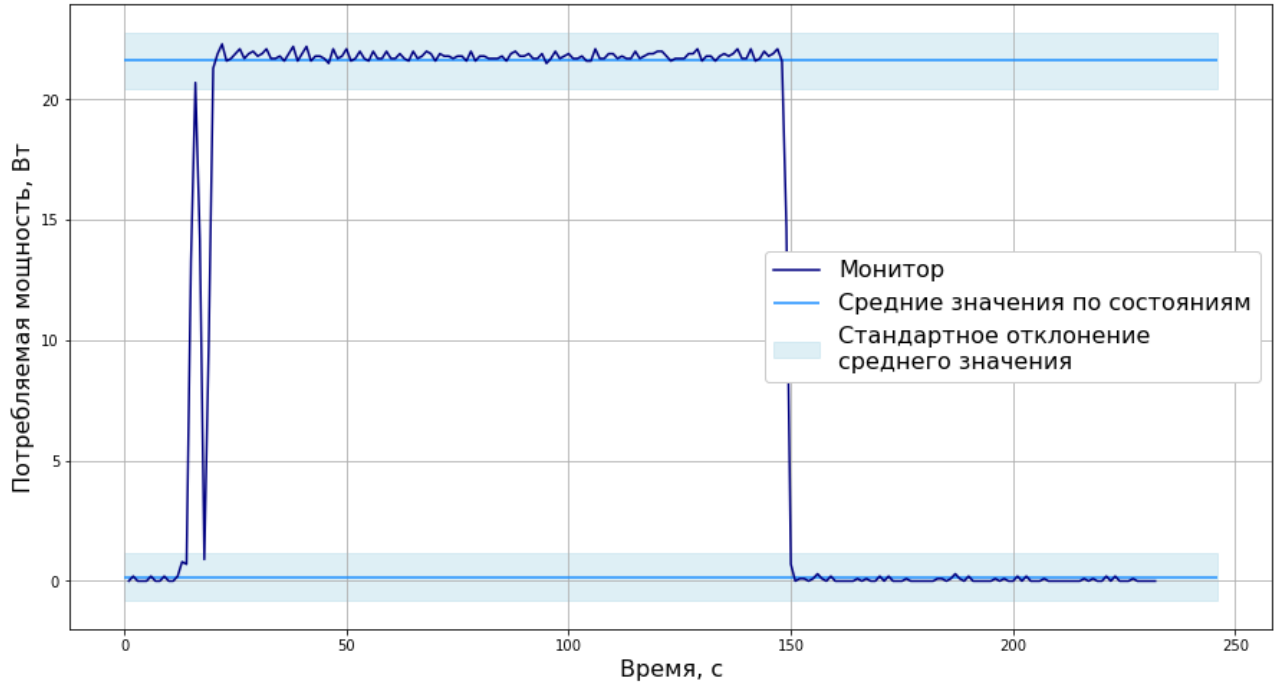
3

Потребление устройства сильно отличается от состояний или работает слишком долго (забыли выключить телевизор или сбой работы станка) => повод насторожиться

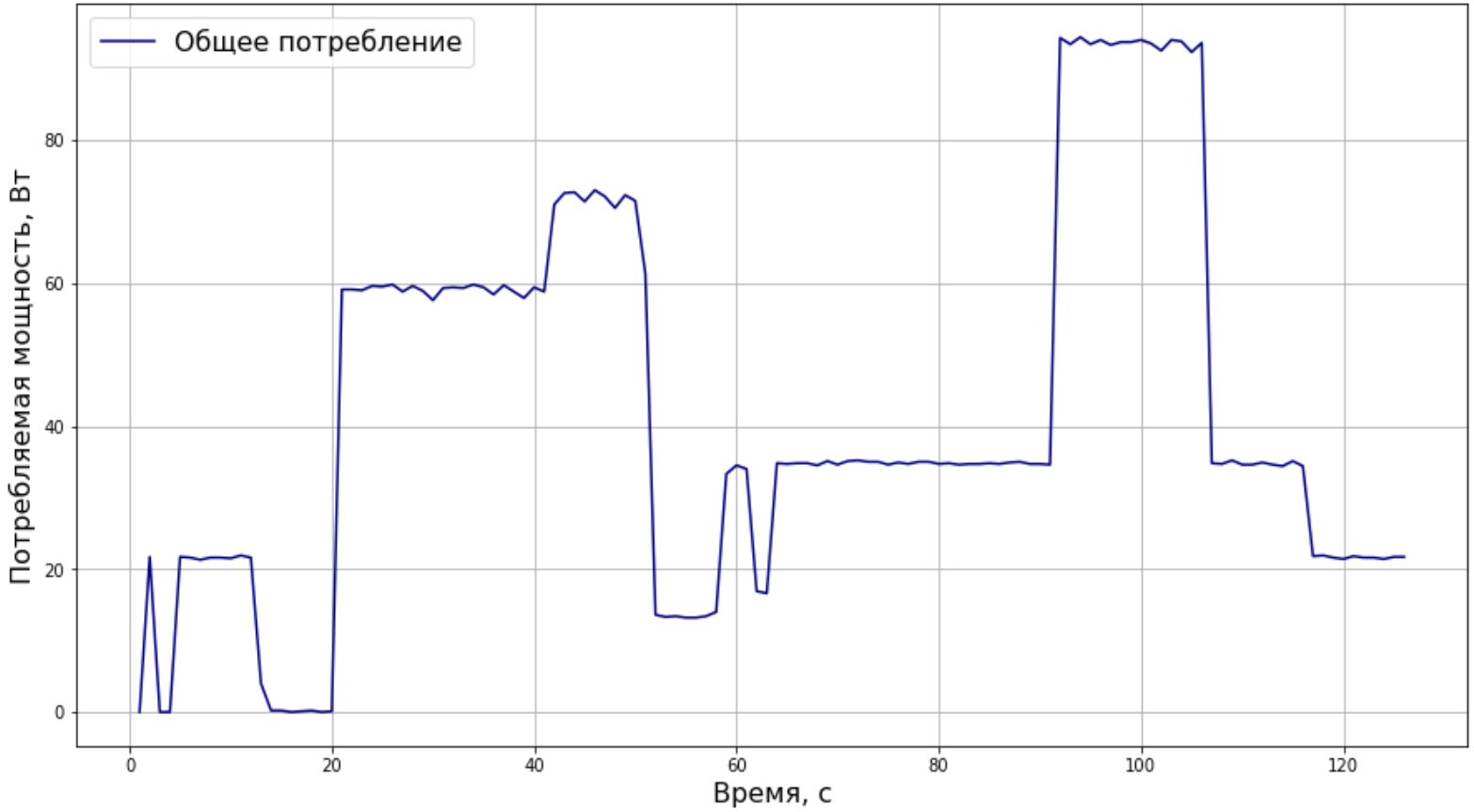
Данные с «умного удлинителя»: обучение

Данные с частотой 1 Гц

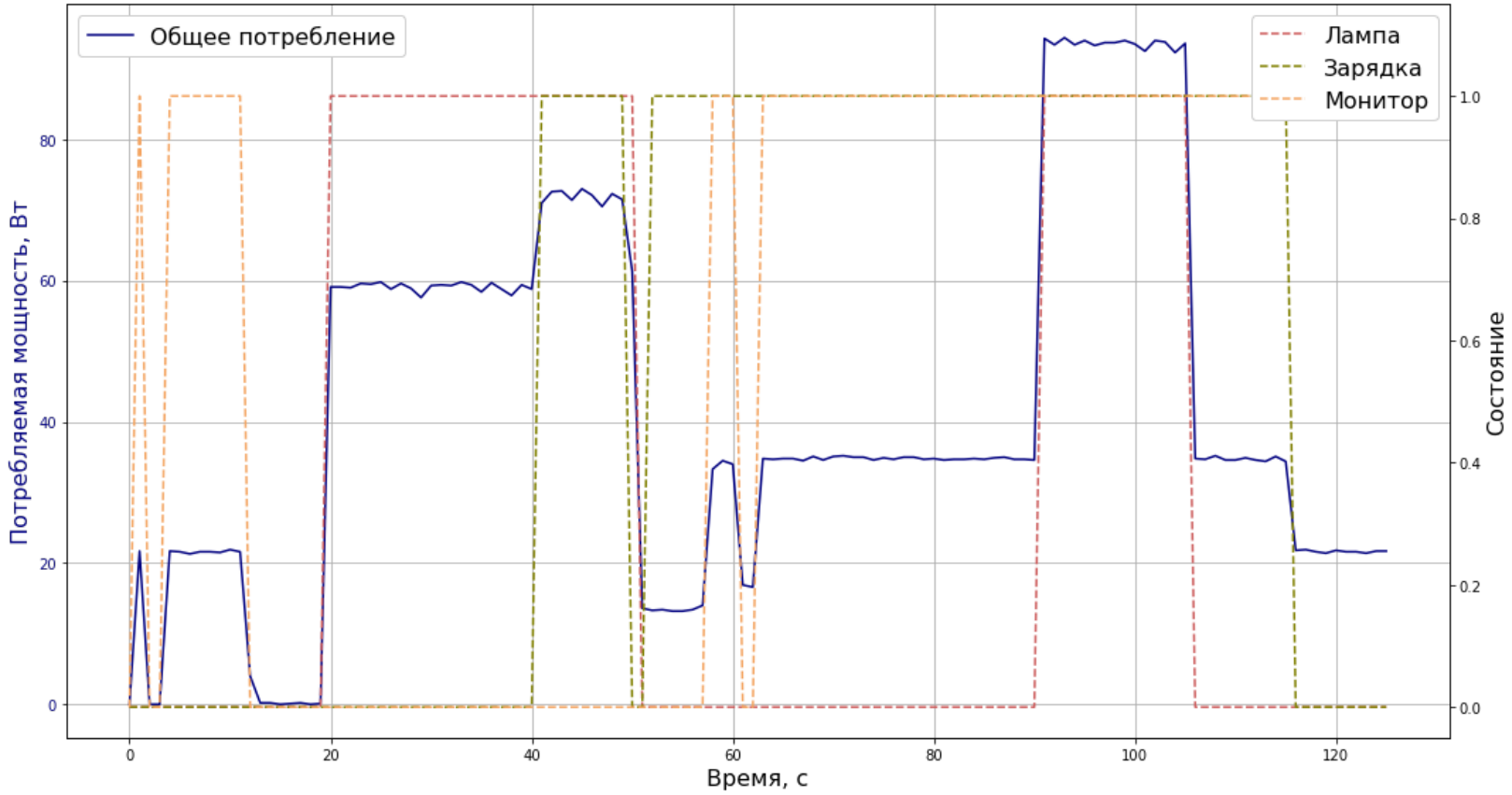
Моделировали стенд с настольной лампой, монитором и зарядкой



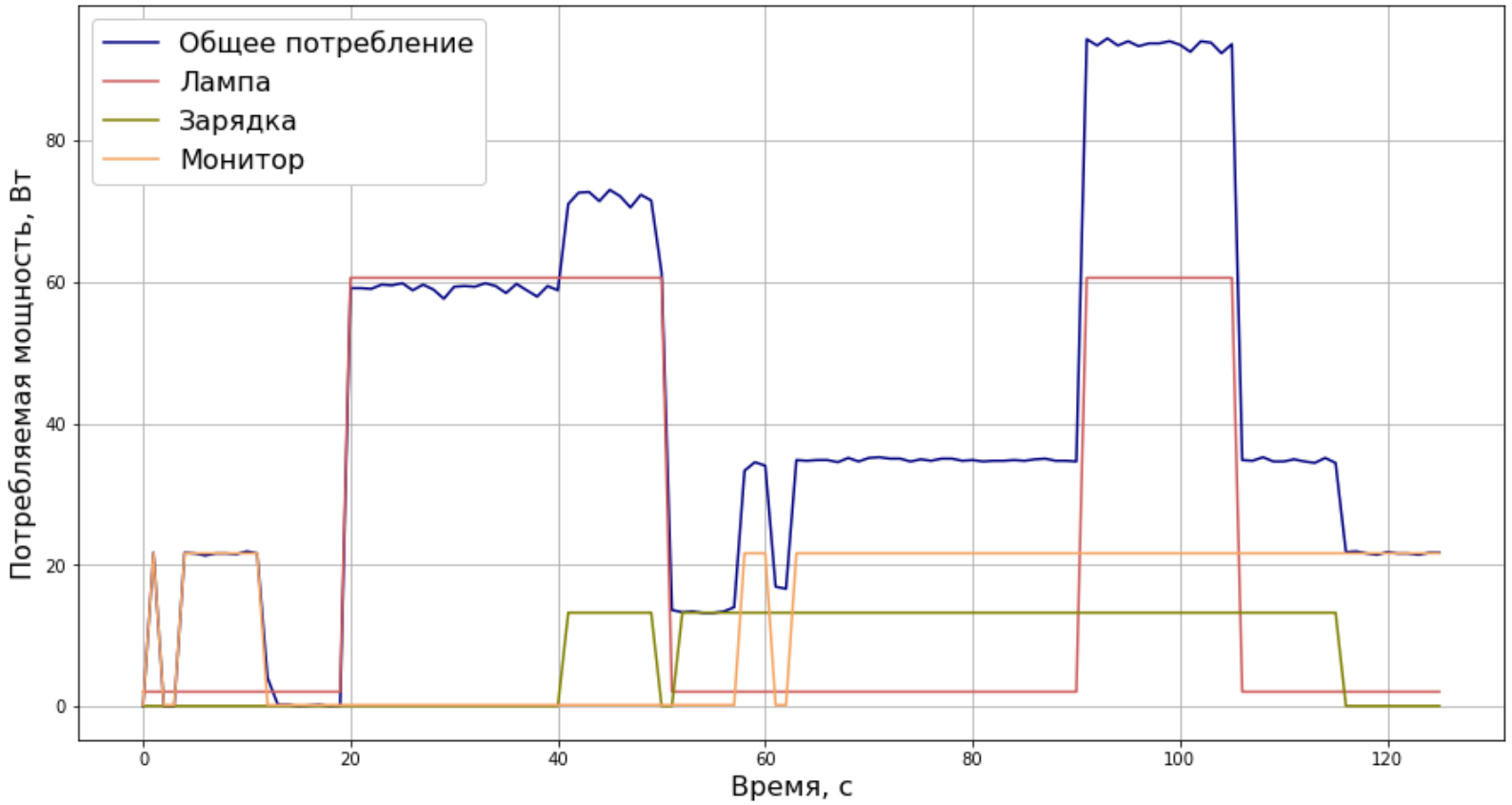
Данные с «умного удлинителя»: общее потребление



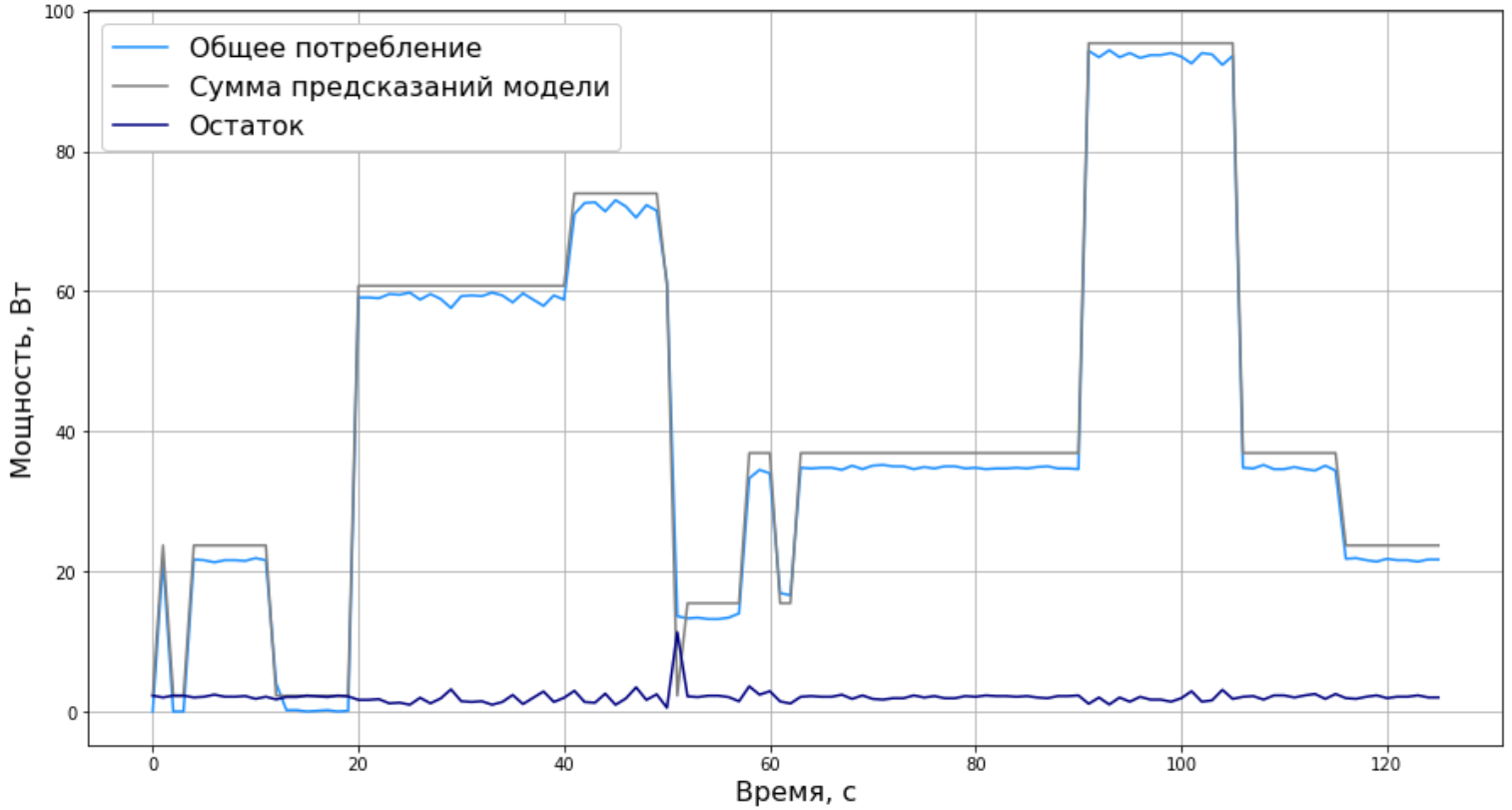
А теперь распознаем устройства



А теперь распознаем устройства



Точность



Порядка нескольких Ватт необъясненного потребления

Искусственный интеллект глазами разработчика

29 мая 2021

Мастерская инструментов разработки

mir.dev

Вопросы?