**Аналитика**

**Расчетный модуль**

Модуль %optimization.py%:

*Суть: дать возможность использовать методы для нахождения “истинных” значений при некоторых условиях уравнения связи, распределений этих значений.*

Структура модуля:

* Функция МНК(используем scipy.py)
  + Принимает некоторый объект, содержащий необходимые поля(например уравнения связи) и возвращает “оптимизированные” значения и значение функционала ошибки(в виде словаря/объекта)
* Функция Bayesian(используем PyMC3)
  + Аналогично(+объект из PyMC)
* Линейный алгоритм(линейное программирование)
  + Аналогично МНК
* Хранение файла с априорными распределениями величин

Библиотеки в python: scipy, PyMC3, pandas, sklearn + (другие дополнения по анализу данных)

**Генерация измерений**

Модуль %generation.py%

*Суть: на основе имеющихся данных сконструировать новые.*

1. Функция generation\_on\_distribution

На основе результатов расчетного модуля, а именно полученных апостериорных распределений, случайным образом выбирать значения (вероятность будет задаваться отдельно).

Библиотеки: PyMC, pandas.

1. Функция make\_some\_noise (для тестирования)

Думаю, она здесь не нужна. Она нужна в модуле-продюсере. Но во время разработки будет продублирована, чтобы упростить тестирование

Есть возможность задавать “шум” для найденных значений, чтобы смоделировать ситуацию неидеальных условий. (добавлять шум будем из соображений вероятности данного значения). Зашумление с помощью pymc3 можно сделать

1. Функция ML\_generation (возможно это даже целый модуль)

Применение машинного обучения. Цель в том, чтобы на основе пункта 1 заполнить 70-75% пропусков в данных (значений датчиков), остальные предсказать с помощью различных методов. Новые данные не обязательно схожи, но и не должны сильно выделяться от имеющихся.

Используем pandas, sklearn + (другие дополнения по машинке)

Дополнения и важные замечания к модулю:

* Учитывание важных признаков для генерации новых измерений. Например, время совершения измерения!

Учитывание режимности работы ТЭЦ

Библиотеки в python: pandas, sklearn + (другие дополнения по машинке)

**Модуль маршрутизации**

Модуль %routing.py%

*Суть: Модуль, принимающий, анализирующий, обрабатывающий и отправляющий данные(в виде временных рядов). Принимает данные в виде временного ряда из соответствующих очередей, загружает данные в объекты, используя API расчетного модуля находит “истинные” значения. При необходимости(недостаток данных) генерирует значения из модуля генерации измерений. Далее отправляем сырые и обработанные данные(преобразованные в csv) в БД*

Библиотеки python: pandas + прием данных (возможно через сервер и онлайн)

Комментарий по коммуникациям

Пока делаем так, что модули бизнес логика, маршрутизация, генерация не разделены в разные сервисы, а маршрутизация просто вызывает их, как библиотеки. Однако, скорее всего, чтобы добавить распределенности, тем самым ускорив общее время работы, разнести и их тоже по разным докерам