

Описание функциональных возможностей

Глобальные переменные:

- `TRAIN_RATIO` — отношение объёма тренировочной выборки к объёму всей выборки;
- `types` — типы трафика;
- `timestamps` — все уникальные отметки времени в формате 'yyyy-mm-dd hh:mm:ss', отсортированные по возрастанию;
- `app_class` — типы приложений;
- `app_color` — цветовая легенда для типов приложений.

Функции:

- `s1 %*% s2` — попарная конкатенация множеств строк `s1` и `s2`, используется для быстрого создания названий всех временных рядов;
- `reparamGG(param)` репараметризует обобщённое гамма-распределение;
- `detect_all_anomalies(app, time, type, alpha=0.05)` выявляет аномалии в выборке объёмов трафика типа `type` по приложению `app` за время `time` на уровне значимости `alpha`;
- `plot_ggamma(app, time, type)` оценивает параметры обобщённого гамма-распределения, проводит тест хи-квадрат, строит гистограмму эмпирического распределения и кривую плотности обобщённого гамма-распределения для выборки объёмов трафика типа `type` по приложению `app` за время `time`;
- `cor.mtest(mat, ...)` проводит тест на равенство нулю коэффициентов корреляций для выборок, являющихся столбцами матрицы `mat`, возвращает матрицу р-значений;

- `rmse(x, y)` вычисляет метрику RMSE между векторами `x` и `y`;
- `calc_rmse_train(ts, kfourier)` для одномерного временного ряда `ts` вычисляет метрику RMSE между тренировочной выборкой и значениями модели регрессии на члены тригонометрического ряда Фурье до порядка `kfourier` с SARIMA остатками, используется для выбора оптимального порядка модели;
- `calc_rmse_var(ts, p)` для многомерного временного ряда `ts` вычисляет метрику RMSE между нормированной тестовой выборкой и прогнозом, полученным моделью векторной авторегрессии порядка `p` по нормированной тренировочной выборке;
- `plot_var(ts, p)` для многомерного временного ряда `ts` строит графики каждого одномерного временного ряда и наносит на них прогнозы моделью векторной авторегрессии порядка `p`, возвращает время в секундах, затраченное на обучение и прогнозирование;
- `calc_rmse_auto_arima(ts, kfourier=0)` для одномерного временного ряда `ts` по нормированной тренировочной выборке определяет лучшую SARIMA модель (если `kfourier=0`) или модель регрессии на члены тригонометрического ряда Фурье до порядка `kfourier` (`kfourier>0`) по критерию Акаике, вычисляет метрику RMSE между нормированной тестовой выборкой и прогнозом;
- `plot_auto_arima(ts, title, kfourier=0)` для одномерного временного ряда `ts` определяет лучшую SARIMA модель (если `kfourier=0`) или модель регрессии на члены тригонометрического ряда Фурье до порядка `kfourier` (`kfourier>0`) по критерию Акаике, строит график временного ряда (с заголовком `title`) и наносит на него прогноз по выбранной модели, возвращает время в секундах, затраченное на обучение и прогнозирование;

- `plot_users_hist(users, type)` для типа трафика `type` и матрицы количества уникальных пользователей за час `users` строит гистограммы и эмпирические функции распределения до и после перерыва, проводит тест Колмогорова-Смирнова на однородность выборок до и после перерыва;
- `calc_hourly_users(data, type)` для выборки `data` за каждую отметку времени из `timestamps` вычисляет количество уникальных пользователей для типа трафика `type`;
- `calc_hourly_users_all(data)` для выборки `data` за каждую отметку времени из `timestamps` вычисляет количество уникальных пользователей по всему трафику;
- `calc_hourly_sum(data, type)` для выборки `data` за каждую отметку времени из `timestamps` вычисляет суммарный трафик типа `type`;
- `collect_hourly_info(app)` для типа приложения `app` вычисляет количество уникальных пользователей, суммарный и средний трафик за каждую отметку времени из `timestamps` по отправленному, полученному и общему трафику;
- `plot_hourly_info(colname)` для каждого типа приложения из `app_class` наносит на график столбец `colname` из матрицы, возвращаемой `collect_hourly_info`, в соответствии с цветовой легендой `app_color`;
- `calc_rmse_var_after_break(ts, p, nahead, ntest)` для многомерного временного ряда `ts` вычисляет метрику RMSE между нормированной тестовой выборкой и прогнозом, полученным моделью векторной авторегрессии порядка `p` по нормированной тренировочной выборке, эта функция используется вместо функции `calc_rmse_var` в случае, когда тестовая выборка следует за тренировочной после некоторого

перерыва, **nahead** — дальность прогнозирования (длина перерыва плюс объём тестовой выборки), **ntest** — объём тестовой выборки.