

CellPAD: Detecting Performance Anomalies in Cellular Networks via Regression Analysis

Introduction

CellPAD is a tool for KPI time-series anomaly detection based on regression analysis. It implements a pipeline to detect anomalies and update prediction models for streaming data in an automated manner. It takes into account both trend and seasonality component of a time series. It can detect two types of anomalies: sudden drops in a time series and correlation changes of two correlated time series.

Prerequisites

CellPAD is written in Python 3. To run CellPAD, please install the following packages first.

- Python 3.x;
- scikit-learn 0.19.0 or higher.

Installation

Run:

```
python setup.py install
```

Algorithm and Feature Selection

CellPAD integrates various statistical-based and machine-learning-based regression algorithms. They include:

- Statistical algorithms:
 - WMA, EWMA, and Holt-Winters (HW)
 - Local Correlation Score (LCS)
- Machine-learning-based regression regression:
 - Random Forest Regression (RF), Regression Tree (RT), Simple Linear Regression (SLR), and Huber Regression (HR).

How to call different algorithms?

```
DropController.detect(predictor)

ChangeController.detect(predictor)
```

- For **DropController**, "*predictor*" can be "RF", "RT", "SLR", "HR", "WMA", "EWMA", "HW".

- For **ChangeController**, "predictor" can be "RF", "RT", "SLR", "HR", "LCS".

How to perform feature selection?

```
DropController(feature_types=["Numerical"],
               feature_time_grain="Weekly",
               feature_operations=["Wma", "Mean", "Mean"])

ChangeController(feature_types=["Indexical"],
                 feature_time_grain="Weekly",
                 feature_operations=["Raw"])
```

- "*feature_types*" can be a subset of ["Numerical", "Indexical"]
- "*feature_time_grain*" can be a subset of ["Hourly", "Daily", "Hourly"]
- "*feature_operations*" can be a subset of ["Raw", "Mean", "Median", "Wma", "Ewma"]

How to remove any trend components?

```
DropController(to_remove_trend=True,
               trend_remove_method="center_mean")

ChangeController(to_remove_trend=False,
                 trend_remove_method="center_mean")
```

- "*to_remove_trend*" is **True** or **False** to indicate whether to remove the trend or not.
- "*trend_remove_method*" is "center_mean" or "past_mean".
 - "center_mean": the trend at time i is the mean of the points in $[i-84, i+83]$.
 - "past_mean": the trend at time i is the mean of the points in $[i-167, i]$.
 - Note that the accuracy of "past_mean" is slightly less than that of "center_mean" in general, as the latter considers the time interval closer to time i .

Examples and test data

We provide two examples on how to run CellPAD for anomaly detection of sudden drops and correlation changes. Run the following:

sudden drop

```
cd ./example
python example_drop.py
```

correlation change

```
cd ./example
python example_change.py
```