PPINot4Py description

Alejandro Gómez Caballero

December 3, 2019

1 Measures

Defining a measure is the way to indicate wich kind of operation we want to do in our dataframe, we have several tipes:

1.1 CountMeasure

Most basic measure, used to count the times a condition occurs.
It can receive as parameter a String condition o a TimeInstantCondition

```
countStateTimeA =
    DataObjectState('lifecycle_transition == "In Progress"')
countConditionTimeA =
    TimeInstantCondition(countStateTimeA)
countMeasureTimeA =
    CountMeasure(countConditionTimeA)

#----------------------------'OR'----------------------------

countMeasureTimeA =
    CountMeasure('lifecycle_transition == "In Progress"')
```

1.2 DataMeasure

With this measure we can obtain the first or last appareance of a value when we apply a TimeInstantCondittion

1.3 TimeMeasure

In this measure, we want to count how much time has happened between 1 event and other, in a linear form or in a cyclic form

If we call "A" to the first condition and "B" to the second, we want the appareances of the pairs $A \rightarrow B$

We have the following parameters:

FromCondition: The first TimeInstantCondition

ToCondition: The second TimeInstantCondition

TimeMeasureType: Tipe of calculation, Linear or Cyclic.

SingleInstanceAggFunction: Tipe of operation we want to apply.

Precondition: In case we want to apply a previous filter to our dataframe

FirstTo: Only works for the linear type, is True if you want the first "B" and False if you want the last "B"

1.3.1 Linear type

In this case, we are counting the first appareance of A condition and the first/last appareance of B condition, then we calculate the time difference between B and A

1.3.2 Cyclic type

In cyclic case, we count each apparition of $A \rightarrow B$ apparitions and calculate an operation to this apparitions.

1.3.3 AggFuntions

We have 5 possible operations by defect in the program:

SUM: The sum of the time values of all $A \rightarrow B$ apparitions

MAX: Max time value between the $A \rightarrow B$ pairs

MIN: Min time value between the $A \rightarrow B$ pairs

AVG: The average of the time values of all $A \rightarrow B$ apparitions

GROUPBY: Raw grouped dataframe with no operation applied

1.4 aggregatedMeasure