Package ‘psmineR’

October 14, 2022

**Type** Package

**Title** Performance Spectrum Miner for Event Data

**Version** 0.1.0

**Date** 2022-10-06

**Description** Compute detailed and aggregated performance spectrum for event data. The detailed performance spectrum describes the event data in terms of segments, where the performance of each segment is measured and plotted for any occurrences of this segment over time and can be classified, e.g., regarding the overall population. The aggregated performance spectrum visualises the amount of cases of particular performance over time. Denisov, V., Fahland, D., & van der Aalst, W. M. P. (2018) <doi:10.1007/978-3-319-98648-7_9>.

**Depends** R (>= 3.5.0)

**License** MIT + file LICENSE

**Encoding** UTF-8

**Imports** bupaR (>= 0.5.1), dplyr, data.table, forcats, ggplot2, tidyr, rlang (>= 1.0.0), cli (>= 3.2.0), glue, stringi

**RoxygenNote** 7.2.1


**Suggests** knitr, eventdataR, rmarkdown, covr, testthat (>= 3.1.3)

**BugReports** https://github.com/bupaverse/psmineR/issues

**Config/testthat/edition** 3

**NeedsCompilation** no

**Author** Greg Van Houdt [aut], Gert Janssenswillen [ctb, cre], Gerard van Hulzen [ctb]

**Maintainer** Gert Janssenswillen <gert.janssenswillen@uhasselt.be>

**Repository** CRAN

**Date/Publication** 2022-10-10 17:00:02 UTC
R topics documented:

- plot ......................................................... 2
- psmineR ..................................................... 2
- ps_aggregated ............................................ 3
- ps_detailed ................................................ 5

Index

---

plot                  *Plot Methods*

Description

Visualize performance spectrum.

Usage

```r
## S3 method for class 'ps_aggregated'
plot(x, ...)

## S3 method for class 'ps_detailed'
plot(x, ...)
```

Arguments

- `x` Object of class `ps_aggregated()` or `ps_detailed()`.
- `...` Additional variables to pass further.

Value

A `<code>ggplot2</code>` object, which can be customised further.

---

psmineR         psmineR

Description

Performance Spectrum Miner For Event Data
ps_aggregated

Descriptions

Description

Plots the aggregated performance spectrum. The performance spectrum describes the event data in terms of segments, i.e., pairs of related process steps. The performance of each segment is measured and plotted for any occurrences of this segment over time and can be classified, e.g., regarding the overall population. The aggregated performance spectrum visualises the amount of cases of particular performance over time (Denisov et al., 2018). See References for more details.

Usage

```r
ps_aggregated(
  log,
  segment_coverage,
  n_segments,
  classification = NULL,
  grouping = c("start", "complete"),
  scale = NULL,
  bins = 30
)
```

```r
## S3 method for class 'log'
ps_aggregated(
  log,
  segment_coverage,
  n_segments,
  classification = NULL,
  grouping = c("start", "complete"),
  scale = NULL,
  bins = 30
)
```

```r
## S3 method for class 'grouped_log'
ps_aggregated(
  log,
  segment_coverage,
  n_segments,
  classification = NULL,
  grouping = c("start", "complete"),
  scale = NULL,
  bins = NULL
)
```
Arguments

log: Object of class log or derivatives (grouped_log, eventlog, activitylog, etc.).

segment_coverage, n_segments
numeric: Provide either segment_coverage or n_segments. If neither is provided, segment_coverage = 0.2 will be used.
segment_coverage: The percentage of cases (default 0.2) in which each segment must be present to be visualised in the spectrum. Ignored if n_segments is specified.
n_segments: Visualise only the top n segments based on frequency.

classification character (default NULL): The variable defining the colour legend. This variable should be present in log.
If NULL (default) when log is a grouped_log, the first grouping variable will be used as classification.
If NULL (default) or "quartile" when log is an eventlog or activitylog, a quartile variable dividing the durations of the segments in quartiles is calculated.

grouping character (default "start"): The timestamps, "start" or "complete", which are binned in the histogram.

scale ggplot2 scale function (default scale_fill_discrete_bupaR): Set color scale. Defaults to scale_fill_discrete_bupaR.

bins numeric (default 30): The number of bins in the aggregated performance spectrum.

Value

A ggplot2 object describing the aggregated performance spectrum.

Methods (by class)

• ps_aggregated(log): Plot aggregated performance spectrum for a log.
• ps_aggregated(grouped_log): Plot aggregated performance spectrum for a grouped_log.

References


See Also

ps_detailed()

Examples

library(psmineR)
library(eventdataR)
sepsis %>%
  ps_aggregated(segment_coverage = 0.2,
                classification = "quartile",
                grouping = "start",
                bins = 15)

---

**ps_detailed**

*Detailed Performance Spectrum*

**Description**

Plots the detailed performance spectrum. The performance spectrum describes the event data in terms of segments, i.e., pairs of related process steps. The performance of each segment is measured and plotted for any occurrences of this segment over time and can be classified, e.g., regarding the overall population. The detailed performance spectrum visualises variability of durations in a segment across cases and time (Denisov et al., 2018). See **References** for more details.

**Usage**

```r
ps_detailed(
  log,
  segment_coverage,
  n_segments,
  classification = NULL,
  scale = NULL
)
```

```r
## S3 method for class 'log'
ps_detailed(
  log,
  segment_coverage,
  n_segments,
  classification = NULL,
  scale = NULL
)
```

```r
## S3 method for class 'grouped_log'
ps_detailed(
  log,
  segment_coverage,
  n_segments,
  classification = NULL,
  scale = NULL
)
```
Arguments

log: Object of class log or derivatives (grouped_log, eventlog, activitylog, etc.).

segment_coverage, n_segments
numeric: Provide either segment_coverage or n_segments. If neither is provided, segment_coverage = 0.2 will be used.
segment_coverage: The percentage of cases (default 0.2) in which each segment must be present to be visualised in the spectrum. Ignored if n_segments is specified.
n_segments: Visualise only the top n segments based on frequency.

classification character (default NULL): The variable defining the colour legend. This variable should be present in log.
If NULL (default) when log is a grouped_log, the first grouping variable will be used as classification.
If NULL (default) or "quartile" when log is an eventlog or activitylog, a quartile variable dividing the durations of the segments in quartiles is calculated.

scale ggplot2 scale function (default scale_color_discrete_bupaR): Set color scale. Defaults to scale_color_discrete_bupaR.

Value

A ggplot2 object describing the detailed performance spectrum.

Methods (by class)

• ps_detailed(log): Plot detailed performance spectrum for a log.
• ps_detailed(grouped_log): Plot detailed performance spectrum for a grouped_log.

References


See Also

ps_aggregated()

Examples

library(psmineR)
library(eventdataR)
sepsis %>%
  ps_detailed(segment_coverage = 0.2,
              classification = "quartile")
Index

activitylog, 4, 6
character, 4, 6
eventlog, 4, 6
ggplot2, 2, 4, 6
grouped_log, 4, 6
log, 4, 6
NULL, 4, 6
numeric, 4, 6
plot, 2
ps_aggregated, 3
ps_aggregated(), 2, 6
ps_detailed, 5
ps_detailed(), 2, 4
psmineR, 2
scale_color_discrete_bupaR, 6
scale_fill_discrete_bupaR, 4