Package ‘mwa’
March 24, 2014

Type Package
Title MWA package for causal inference in spatiotemporal event data
Version 0.1
Date 2014-03-21
Author Sebastian Schutte and Karsten Donnay
Maintainer Sebastian Schutte <schuttesebastian@gmail.com>
Description MWA grants insights into causal relationships in spatiotemporal event data.
License LGPL-3
Depends R (>= 2.6), cem (>= 0.3), rJava (>= 0.9), MASS (>= 7.0)
SystemRequirements Java (>= 6.0)
NeedsCompilation no
Repository CRAN
Date/Publication 2014-03-24 13:26:47

R topics documented:

mwa-package .................................................. 2
matchedwake .................................................. 3
mwa_data .................................................... 6
plot.matchedwake ............................................ 7
print.matchedwake .......................................... 8
slideWakeMatch ............................................. 9
slidingWake ................................................ 10
summary.matchedwake .................................. 11

Index 13
Description

MWA analyzes spatially and temporally referenced data. Specific types of events might affect subsequent levels of other events. To estimate the corresponding effect, treatment, control, and dependent events are selected from the empirical sample. Treatment effects are established through automated matching and a diff-in-diffs regression design. The analysis is repeated for various spatial and temporal offsets from the treatment events.

Details

The full functionality of MWA is given through `matchedwake`, which relies on a small set of auxiliary methods. Note that `print()`, `summary()` and `plot()` commands are overloaded to return outputs specific to class `matchedwake`.

Author(s)

Sebastian Schutte and Karsten Donnay

References


See Also

`matchedwake`, `slidingWake`, `slideWakeMatch`, `print`, `summary`, `plot`

Examples

```r
# Loading sample data
data(mwa_data)

# Specify required parameters:
# - 2 to 10 days in steps of 2
t_window <- c(2,10,2)
# - 2 to 10 kilometers in steps of 2
spat_window <- c(2,10,2)
# - column and entries that indicate treatment events
treatment <- c("type","treatment")
# - column and entries that indicate control events
control <- c("type","control")
# - column and entries that indicate dependent events
dependent <- c("type","dependent")
# - columns to match on
matchColumns <- c("match1","match2")
```
# Specify optional parameters:
# - use weighted regression (default estimation method is "lm")
weighted <- TRUE
# - temporal units
t_unit <- "days"
# - match on counts of previous treatment and control events
TCM <- TRUE
# - maximal Java heap space to be used in the analysis
memory <- 1 # in GB

## Not run:
# Execute method:
results <- matchedwake(mwa_data, t_window, spat_window, treatment, control, dependent,
                        matchColumns, weighted = weighted, t_unit = t_unit, TCM = TCM,
                        memory = memory)

## Plot results:
plot(results)

## Return detailed summary of results:
summary(results, detailed = TRUE)

## End(Not run)

### matchedwake

**estimate treatment effect for sliding spatiotemporal windows**

### Description

This function performs the MWA analysis consisting of two steps: counts for previous and posterior events are established for different spatial and temporal offsets from treatment and control events. After that, the treatment effect is estimated in a difference-in-differences regression design.

### Usage

```r
matchedwake(data, t_window, spat_window, treatment, control, dependent, 
             matchColumns, weighted = weighted, t_unit = t_unit, TCM = TCM, 
             memory = memory)
```

### Arguments

- **data**
  - data.frame containing the observations. See Details.
- **t_window**
  - specification of temporal windows in t_units. See Details.
- **spat_window**
  - specification of spatial windows in kilometers. See Details.
- **treatment**
  - vector of Strings identifying which type of events serve as treatments. See Details.
control vector of Strings identifying which type of events serve as controls. See Details.
dependent vector of Strings identifying which type of events are affected by treatment. See Details.
matchColumns vector of Strings indicating the columns to match on. See Details.
t_unit String specifying the temporal units to be used, either "days", "hours", "mins" or "secs". Default = "days". See Details.
estimation String specifying method used for estimation, "lm", "att" or "nb". Default = "lm". See Details.
weighted Boolean specifying whether regression is weighted (only affects estimations using "lm" or "att"). Default = FALSE.
estimationControls vector of Strings indicating additional control dimensions to be included in the estimation. See Details.
TCM Boolean to select whether the method should match on counts of previous treatment and control instances. Default = FALSE.
deleteSUTVA Boolean to select whether overlapping treatment and control episodes are deleted. Default = FALSE.
alpha1 first significance level used for the analysis and plots. Default = 0.05.
alpha2 second significance level used for the analysis and plots. Default = 0.1.
memory size of the Java heap space for the counting procedure in gigabytes. Default = 1.
match.default Boolean to select whether observations are matched using cem. Default = TRUE.
... optional parameters that can be passed to the methods used for matching and estimation. See Details.

Details

The method expects data to be a data.frame. Dates must be given in column timestamp and formatted as a date string with format "YYYY-MM-DD hh:mm:ss". Alternatively, a POSIX Date can be specified using the same format. data must also contain two entries called lat and long for the geo location of each entry.

t_window specifies the minimal and maximal temporal window sizes and corresponding steps used in the iteration. Required syntax is c(min_window, max_window, step_size) with step_size in units of t_unit. The spatial window spat_window is specified in the same way with kilometers as units.

treatment, control and dependent define which category of events is considered to be treatment, control and dependent cases respectively. The required syntax is c(column_name, value) where column_name must be entered as String and value can be Numeric, Boolean, or a String.

matchColumns selects the columns in data used for matching. Matching variables are expected to be coded together with every treatment and control type event and are assumed to reflect a set of suitable matching variables (what is suitable will, of course, vary from case to case).

The optional argument t_unit specifies the temporal resolution for which the analysis is to be conducted, one of either "days", "hours", "mins" or "secs". If the time stamps provided in data are more precise than the resolution they are truncated accordingly.
MWA estimates treatment effects using a diff-in-diffs regression design, i.e. \( dependent\_post \sim dependent\_pre + treatment \) (where “pre” and “post” refer to pre and post intervention). Three different methods can be chosen using estimation: a linear model ("lm", stats), all models available through ("att", cem) or a count dependent model ("glm.nb", MASS). For regressions using "lm" or "att" weighted sets whether or not the regression is weighted by the number of treatment vs. control cases. Additional control variables can be specified via estimationControls. For example, if estimationControls = c("covariate1"), the package automatically modifies the estimation formula to "dependent\_post \sim dependent\_pre + covariate1 + treatment".

The package supports full inheritance for optional arguments of the following methods: cem and att (cem), lm (stats), glm.nb (MASS). To guarantee unique inputs for each method, options have to entered into matchedwake() using a prefix that consists of the method name separated by ".". For example, in order for cem to return an exactly balanced dataset simply add cem.k2k = TRUE as optional argument.

Value

Returns an object of class matchedwake, which is a list of objects with the following slots:

- estimates: data.frame with estimates and p values for all spatial and temporal windows considered. For estimation = "lm" it also returns a pseudo \( R^2 \) value.
- matching: data.frame with detailed matching statistics for all spatial and temporal windows considered. Returns the number of control and treatment episodes, L1 metric, percent common support. All values are given both pre and post matching.
- SUTVA: data.frame with detailed statistics on the degree of overlaps of the spatiotemporal cylinders. Returns the fraction of cases in which two or more treatment (or control) episodes overlap ("SO": same overlap) and the fraction of overlapping treatment and control episodes ("MO": mixed overlap). All values are given pre and post matching and for the full time window.
- wakes: data.frame providing the information for the spatiotemporal cylinders (or wakes) for all spatial and temporal windows considered. Returns the eventID (i.e. the index of the event in the time-ordered dataset), treatment (1: treatment episode, 0: control episode), counts of dependent events, overlaps ("SO" and "MO") pre and post intervention, and the matching variables.
- parameters: list of all arguments passed to the method.
- call: the call.

Author(s)

Sebastian Schutte and Karsten Donnay.

References


See Also

mwa-package, slidingWake, slideWakeMatch, print.matchedwake, summary.matchedwake, plot.matchedwake
Examples

# Loading sample data
data(mwa_data)

# Specify required parameters:
# - 2 to 10 days in steps of 2
  t_window <- c(2,10,2)
# - 2 to 10 kilometers in steps of 2
  spat_window <- c(2,10,2)
# - column and entries that indicate treatment events
  treatment <- c("type","treatment")
# - column and entries that indicate control events
  control <- c("type","control")
# - column and entries that indicate dependent events
  dependent <- c("type","dependent")
# - columns to match on
  matchColumns <- c("match1","match2")

# Specify optional parameters:
# - use weighted regression (default estimation method is "lm")
  weighted <- TRUE
# - temporal units
  t_unit <- "days"
# - match on counts of previous treatment and control events
  TCM <- TRUE
# - maximal Java heap space to be used in the analysis
  memory <- 1 # in GB

## Not run:
# Execute method:
results <- matchedwake(mwa_data, t_window, spat_window, treatment, control, dependent, 
                      matchColumns, weighted = weighted, t_unit = t_unit, TCM = TCM, 
                      memory = memory)

# Plot results:
plot(results)

# Return detailed summary of results:
summary(results, detailed = TRUE)

## End(Not run)

---

mwa_data  
data to illustrate the functionality of MWA

Description

This artificial data set illustrates how MWA can be used to identify causal effects. Treatment, control, and dependent events are referenced in time and space. Increased levels of dependent events following treatments can be visually and numerically analyzed with MWA.
**Usage**

```r
data(mwa_data)
```

**Format**

A data.frame containing observations.

**Source**

Monte Carlo Simulations. See supplementary information of Schutte and Donnay (2014).

**References**


---

**Description**

Overloads the default `plot()` for objects of class `matchedwake`. Returns a contour plot: The lighter the color the larger the estimated treatment effect. The corresponding standard errors are indicated by shading out some of the estimates: No shading corresponds to \( p < \alpha_1 \) for the treatment effect in the diff-in-diffs analysis. Dotted lines indicate \( p \)-values between \( \alpha_1 \) and \( \alpha_R \) and full lines indicate \( p > \alpha_2 \). The cells indicating effect size and significance level are arranged in a table where each field corresponds to one specific combination of spatial and temporal sizes.

**Usage**

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

**Arguments**

- `x` object of class `matchedwake`.
- `...` further arguments passed to or from other methods.

**Author(s)**

Sebastian Schutte and Karsten Donnay.

**References**

See Also

matchedwake

print.matchedwake

print function for objects of class matchedwake

Description

Overloads the default print() for objects of class matchedwake.

Usage

## S3 method for class 'matchedwake'
print(x, ...)

Arguments

x            object of class matchedwake.
...

further arguments passed to or from other methods.

Value

Returns a data.frame with all significant results (significance level is alpha1 as retrieved from x$parameters).

Author(s)

Sebastian Schutte and Karsten Donnay.

References


See Also

matchedwake
slideWakeMatch

**auxiliary function to match data and estimate treatment effects**

**Description**
Method takes the output of `slidingWake`, matches observations using `cem` and estimates treatment effects using linear models (`lm` or `att`) or a count dependent variable model (`glm.nb`).

**Usage**

```
slideWakeMatch(wakes, alpha1, matchColumns, estimation, weighted, estimationControls, TCM, match.default, ...)
```

**Arguments**
- `wakes` data.frame. See “wakes” in the description of `matchedwake` for details.
- `alpha1` significance level used for the analysis and plots. Default = 0.05.
- `matchColumns` vector of Strings indicating the columns to match on.
- `estimation` String specifying method used for estimation.
- `weighted` Boolean specifying whether regression is weighted.
- `estimationControls` vector of Strings indicating additional control dimensions to be included in the estimation.
- `TCM` Boolean to select whether the method should match on counts of previous treatment and control instances.
- `match.default` Boolean to select whether observations are matched using `cem`.
- `...` optional parameters that can be passed to the methods used for matching and estimation.

**Details**
See the description of `matchedwake` for details.

**Value**
Returns a list with the following slots:
- `estimates` data.frame with estimates and p values for all spatial and temporal windows considered.
- `matching` data.frame with detailed matching statistics for all spatial and temporal windows considered.
- `SUTVA` data.frame with detailed statistics on the degree of overlaps of the spatiotemporal cylinders.
- `wakes` data.frame.

See the description of `matchedwake` for details.
Author(s)

Sebastian Schutte and Karsten Donnay.

References


See Also

matchedwake, slidingWake

slidingWake  auxiliary function to iterate through sliding spatiotemporal windows

Description

Method iterates through all spatial and temporal window sizes specified and counts dependent events with a given spatial window and for a given temporal window (symmetrically in forward and backward direction in time). For performance reasons, the iterative counting is done in Java using the rJava interface.

Usage

slidingWake(data, t_unit, t_window, spat_window, treatment, control, dependent, matchColumns, estimationControls, memory)

Arguments

data  data.frame containing the observations.
t_unit  String specifying the temporal units to be used.
t_window  specification of temporal windows in t_units.
spat_window  specification of spatial windows in kilometers.
treatment  vector of Strings identifying which type of events serve as treatments.
control  vector of Strings identifying which type of events serve as controls.
dependent  vector of Strings identifying which type of events are affected by treatment.
matchColumns  vector of Strings indicating the columns to match on.
estimationControls  vector of Strings indicating additional control dimensions to be included in the estimation.
memory  size of the Java heap space for the counting procedure in gigabytes.

Details

See the description of matchedwake for details.
**Value**

Returns a data.frame. See “wakes” in the description of `matchedwake` for details.

**Author(s)**

Sebastian Schutte and Karsten Donnay.

**References**


**See Also**

`matchedwake`, `slideWakeMatch`

---

**summary.matchedwake**  
*summary function for objects of class matchedwake*

**Description**

Overloads the default `summary()` for objects of class `matchedwake`.

**Usage**

```r
## S3 method for class 'matchedwake'
summary(object, detailed = FALSE, ...)  
```

**Arguments**

- `object`  
  object of class `matchedwake`.

- `detailed`  
  Boolean indicating whether or not a detailed summary should be returned. Default = TRUE.

- `...`  
  further arguments passed to or from other methods.

**Value**

Returns a data.frame with an overview of all significant results (significance level is `alpha1` as retrieved from `x$parameters`). If `detailed` = TRUE this overview includes a number of matching statistics and statistics on overlaps of the spatiotemporal cylinders.

**Author(s)**

Sebastian Schutte and Karsten Donnay.

**References**

See Also

matchedwake
Index

att, 5, 9

cem, 5, 9

glm.nb, 5, 9

lm, 5, 9

matchedwake, 2, 3, 7–12
mwa-package, 2
mwa_data, 6

plot, 2
plot.matchedwake, 5, 7
print, 2
print.matchedwake, 5, 8

slideWakeMatch, 2, 5, 9, 11
slidingWake, 2, 5, 9, 10, 10
summary, 2
summary.matchedwake, 5, 11