Package ‘TPSQR’

May 25, 2019

Title  Temporal Poisson Square Root Graphical Models
Version 1.0.0
Date 2018-11-01
Author Sinong Geng [aut, cre],
Zhaobin Kuang [aut]
Peggy Peissig [aut]
David Page [aut]
Maintainer Sinong Geng <sgeng@cs.princeton.edu>

Description The Package builds temporal Poisson square root graphical models (TPSQRs) for longitudinal event data (LED), and implements the Poisson pseudo-likelihood (PPL) method for the parameter estimation. An example using the Marshfield clinic data are included. The Marshfield clinic dataset itself is not provided in the package.

Imports data.table,
      Matrix,
      pROC,
      glmnet,
      parallel

Depends R (>= 3.3.0)
License MIT + file LICENSE
Encoding UTF-8
LazyData TRUE
RoxygenNote 6.1.0

R topics documented:

    TPSQR-package ........................................... 2
eventGen .................................................... 2
exampleMCD .................................................. 3
funGenerateNodes .......................................... 4
funPseudoMatrixlog ........................................ 4
runPseudolog ............................................... 5

Index 6
TPSQR-package  TPSQR: Temporal Poisson Square Root Graphical Model

Description

TPSQR builds temporal Poisson square root graphical models (TPSQRs) for longitudinal event data (LED), and implements the Poisson pseudo-likelihood (PPL) method for the parameter estimation. An example using the Marshfield clinic data is included. The Marshfield clinic dataset itself is not provided in the package.

References

http://proceedings.mlr.press/v80/geng18a.html

Examples

eventGen(admitL = 1000, exposure = 1500, nG = 3, mS = 200, alphaSet = 1, ambiguity = 175, futureDiscount = 0.1)

---

eventGen  Preprocessing for the Marshfield dataset

Description

The function preprocesses the Marshfield clinic dataset and generates a list containing LEDs of patients. This function can only be applied on linux.

Usage

eventGen(admitL = 1000)

Arguments

admitL  The minimal length of the stay of a patient to be included into the model. Only the patients with a timespan longer than admitL in the dataset are incorporated into the model.

Value

eventList eventList is saved as eventList.RData under the working directory. The generated eventList is a list containing multiple matrices, each of which represents the observations of one patient. The matrices contain two columns: one for the observed time, and the other one for the observed event types.
Description

This function implements an example that builds TPSQRs for the Marshfield Clinic data, and uses PPL for the parameter estimation. The function can only be used on linux.

Usage

```r
eampleMCD(lambda = c(10^(seq(-1, -13, length.out = 71))),
          admitL = 1000, exposure = 1500, nG = 3, mS = 200, alphaSet = 1,
          ambiguity = 175, futureDiscount = 0.1)
```

Arguments

- `lambda`: Considered regularization parameters.
- `admitL`: The minimal length of the stay of a patient to be included into the model. Only the patients with a timespan longer than `admitL` in the dataset are incorporated into the model.
- `exposure`: The drug exposure time. We assume that the two events with a time interval longer than `exposure` have no effects to each other.
- `nG`: Number of groups in the dyadic influence function.
- `mS`: Maximal length of the data squashing. We will squash the events with the time intervals less than `mS`.
- `alphaSet`: Considered ridge regression parameters.
- `ambiguity`: The parameter controlling the ambiguity when dealing with the occurrence time in the data. Since the observed time is often inaccurate, we use `ambiguity` to include some randomness into the occurrence time of events.
- `futureDiscount`: The discount made on the effects of future events to the previous events. In practice, the performance of the algorithm can be improved by discounting the effects of future events.
- `v20.txt`: `v20.txt` is the Marshfield clinic data that should be saved under the working directory.

Details

The details of this example are discussed in the cited paper. The calculation of the example may take several hours.

Value

- `aucTPSQR`: The AUCs for the structure recovery and the used parameters.

References

- [http://proceedings.mlr.press/v80/geng18a.html](http://proceedings.mlr.press/v80/geng18a.html)

Examples

```r
eampleMCD()
```
funGenerateNodes  

*Node generation for TPSQRs*

**Description**

This function builds TPSQRs for LED. Also, the data squashing is implemented. This function can only be used on linux.

**Usage**

```r
funGenerateNodes(nG = 3, exposure = 1000, mS = 200, ifMinus1 = F, ambiguity = 175, futureDiscount = 0.1)
```

**Arguments**

- **nG**: Number of groups of the dyadic influence function.
- **exposure**: The drug exposure time. We assume that the two events with a time interval longer than `exposure` have no effects to each other.
- **mS**: Maximal length of the data squashing. We will squash the events with the time intervals less than `mS`.
- **ifMinus1**: Whether to delete 1 for x.
- **ambiguity**: The parameter controlling the ambiguity when dealing with the occurrence time in the data. Since the observed time is often inaccurate, we use `ambiguity` to include some randomness into the occurrence time of events.
- **futureDiscount**: The discount made on the effects of future events on the previous events. In practice, the performance of the algorithm can be improved by discounting the effects of future events.
- **eventList**: eventList should be saved as `eventList.RData` under the working directory. The event List contains multiple lists, each of which contains the observations of one patient. The matrices contain two columns: one for the observed time, and the other one for the observed event types.

**Value**

- `ijt`: The information of the edges saved as indexes.

---

funPseudoMatrixlog  

*Design matrix generation for the PPL method*

**Description**

The design matrix for the PPL method is generated to achieve the parameter estimation for TPSQRs.

**Usage**

```r
funPseudoMatrixlog(nG = 3, exposure = 1000, pnodes = 20)
```
Arguments

\texttt{runPseudolog} \hspace{1cm} \textit{The implementation of PPL.}

\textbf{Value}

- \texttt{dM} The built design matrix. \texttt{dM} is saved as \texttt{dMlog.RData} under the working directory.
- \texttt{fix} The built fixed effects. \texttt{fix} is saved as \texttt{fix.RData} under the working directory.

\textbf{Description}

This function implements the PPL Method. This function can only be used on linux.

\textbf{Usage}

\texttt{runPseudolog(lambda, nG = 3, alphaSet = (0:1) * 0.2, admitL, exposure, ifAUC = T, pnodes = 20, ifAIC = F)}

\textbf{Arguments}

- \texttt{lambda} The array of the values of the regularization parameter considered. The \texttt{lambda} relies heavily on the warm start.
- \texttt{nG} Number of groups of the dyadic influence function.
- \texttt{alphaSet} Considered ridge regression parameters.
- \texttt{admitL} The minimal length of the stay of a patient to be included into the model. Only the patients with a timespan longer than \texttt{admitL} in the dataset are incorporated into the model.
- \texttt{exposure} The drug exposure time. We assume that the two events with a time interval longer than \texttt{exposure} have no effects to each other.
- \texttt{ifAUC} A Boolean variable indicating whether to return the AUCs.
- \texttt{ifAIC} A Boolean variable indicating whether to return the AUCs selected by AIC.
- \texttt{mS} Maximal length of the data squashing. We will squash the events with the time intervals less than \texttt{mS}.
- \texttt{gtMatrix} The ground truth result saved in an index form. \texttt{gtMatrix} should be be saved as \texttt{gtMatrix.RData} under the working directory.
Index

eventGen, 2
exampleMCD, 3

funGenerateNodes, 4
funPseudoMatrixlog, 4

package–TPSQR (TPSQR-package), 2

runPseudolog, 5

TPSQR (TPSQR-package), 2
TPSQR-package, 2