Package ‘geosimilarity’

November 8, 2022

Type Package

Title Geographically Optimal Similarity

Date 2022-11-08

Version 2.2

Maintainer Yongze Song <yongze.song@outlook.com>

Description Understanding spatial association is essential for spatial statistical inference, including factor exploration and spatial prediction. Geographically optimal similarity (GOS) model is an effective method for spatial prediction, as described in Yongze Song (2022) <doi:10.1007/s11004-022-10036-8>. GOS was developed based on the geographical similarity principle, as described in Axing Zhu (2018) <doi:10.1080/19475683.2018.1534890>. GOS has advantages in more accurate spatial prediction using fewer samples and critically reduced prediction uncertainty.

Imports stats, SecDim, DescTools, ggplot2, dplyr, ggrepel

Depends R (>= 4.1.0)

License GPL-2

RoxygenNote 7.2.1

LazyData true

Encoding UTF-8

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Yongze Song [aut, cre] (<https://orcid.org/0000-0003-3420-9622>)

Repository CRAN

Date/Publication 2022-11-08 16:00:02 UTC
**R topics documented:**

- `bestkappa`  
- `gos`  
- `grid`  
- `zn`

**Index**

| bestkappa | Function for the best kappa parameter |

---

**Description**

Function for determining the best kappa parameter for the optimal similarity

**Usage**

```r
bestkappa(formula, data = data, kappa = seq(0.05,1,0.05), nrepeat = 10)
```

**Arguments**

- `formula` A formula of GOS model
- `data` A data.frame of observation data
- `kappa` A numeric vector of the optional percentages of observation locations with high similarity to a prediction location. kappa = 1 - tau, where tau is the probability parameter in quantile operator. kappa = 0.25 means that 25 location are used for modelling.
- `nrepeat` A numeric value of the number of cross-validation training times. The default value is 10.

**Value**

A list of the result of the best kappa and the computation process.

**Examples**

```r
data("zn")  
# log-transformation  
hist(zn$Zn)  
zn$Zn <- log(zn$Zn)  
hist(zn$Zn)  
# remove outliers  
require(SecDim)  
k <- rmvoutlier(zn$Zn, coef = 2.5)  
dt <- zn[-k,]  
# determine the best kappa  
system.time(  
  b1 <- bestkappa(Zn ~ Slope + Water + NDVI + SOC + pH + Road + Mine,
```
gos

```r

data = dt,
kappa = c(0.01, 0.1, 1),
nrepeat = 1)
```

```r
b1$bestkappa
b1$plot
```

---

gos

**Geographically optimal similarity**

**Description**

Function for geographically optimal similarity (GOS) model

**Usage**

```r
gos(formula, data = NULL, newdata = NULL, kappa = 0.25)
```

**Arguments**

- `formula`: A formula of GOS model
- `data`: A data.frame of observation data
- `newdata`: A data.frame of prediction variables data
- `kappa`: A numeric value of the percentage of observation locations with high similarity to a prediction location. `kappa = 1 - tau`, where `tau` is the probability parameter in quantile operator. The default `kappa` is 0.25, meaning that 25 location are used for modelling.

**Value**

A list of predictions and uncertainties.

**Examples**

```r
data("zn")
# log-transformation
hist(zn$Zn)
zn$Zn <- log(zn$Zn)
hist(zn$Zn)
# remove outliers
require(SecDim)
k <- rmvoutlier(zn$Zn, coef = 2.5)
dt <- zn[-k,]
# split data for validation: 70% training; 30% testing
split <- sample(1:nrow(dt), round(nrow(dt)*0.7))
train <- dt[split,]
test <- dt[-split,]
```
system.time({
  # 0.33s
g1 <- gos(Zn ~ Slope + Water + NDVI + SOC + pH + Road + Mine,
             data = train, newdata = test, kappa = 0.25)
})
test$pred <- g1$pred
plot(test$Zn, test$pred)
cor(test$Zn, test$pred)
Index

* dataset
  grid, 4
  zn, 4
* elements
  grid, 4
  zn, 4
* trace
  grid, 4
  zn, 4

bestkappa, 2

gos, 3
grid, 4

zn, 4