

Self-compassion, post-traumatic growth, and PTSD

Outlier detection

Corrado Caudek

Last modified 2020-11-30

Motivation

The purpose of the present script is to remove the outliers according to isolation forest algorithm.

Prelims

```
knitr::opts_chunk$set(  
  echo = TRUE,  
  message = FALSE,  
  strip.white = TRUE  
)  
  
library("here")  
  
## here() starts at /Users/corrado/Dropbox/papers/self_compassion  
  
suppressPackageStartupMessages(library("tidyverse"))  
library("ggthemes")  
suppressPackageStartupMessages(library("lavaan"))  
suppressPackageStartupMessages(library("brms"))  
suppressPackageStartupMessages(library("viridis"))  
library("tidyr")  
suppressPackageStartupMessages(library("mice"))  
suppressPackageStartupMessages(library("corrplot"))  
suppressPackageStartupMessages(library("bayesplot"))  
suppressPackageStartupMessages(library("semPlot"))  
suppressPackageStartupMessages(library("rio"))  
suppressPackageStartupMessages(library("outForest"))  
suppressPackageStartupMessages(library("semTools"))  
suppressPackageStartupMessages(library("semoutput"))  
suppressPackageStartupMessages(library("isotree"))  
  
options(max.print = 99999999)  
  
source(here("libraries", "self_compassion_fnc.R"))
```

```

# Read data
temp <- rio::import(
  here("data", "processed", "rescue_workers.Rds"),
)

# remove uninformative variable
temp$is_rescue_worker <- NULL

# remove missing data on all questionnaires data (for convenience,
# I use the variable 'friends')
mydata <- temp[!is.na(temp$friends), ]
# summary(mydata)

mydata$years_experience <- as.numeric(as.character(mydata$years_experience))

# Multiple imputation

# MI for age
# select only the numeric columns of 'temp1'
nums <- unlist(lapply(mydata, is.numeric))
dat_num <- mydata[, nums]

# one value of 'years_experience' is miscoded: 2011;
# it will be recoded as NA and the multiply imputed and
# added again to the 'mydata' data.frame
dat_num$years_experience <- ifelse(
  dat_num$years_experience == 2011, NA, dat_num$years_experience
)

imp <- mice(
  dat_num,
  method = "midastouch",
  m = 1,
  maxit = 1,
  seed = 12345
)

##
## iter imp variable
##   1   1 age years_experience
## Warning: Number of logged events: 3

d_imp <- complete(imp)
# summary(dat_num$age)
# summary(d_imp$age)

# add imputed 'age' to mydata

```

```

mydata$age_imp <- d_imp$age
mydata$years_experience <- d_imp$years_experience

rm(d_imp, dat_num, temp)

# Data wrangling
# Compute total score for MSPSS
mydata$mspss <- mydata$family + mydata$friends + mydata$significant_other

# Recode rate of activity as a numeric variable
mydata$activity_rate <- as.numeric(
  as.character(fct_recode(
    mydata$rate_of_activity,
    "5" = "Più di 1 turno a settimana",
    "4" = "1 turno a settimana",
    "3" = "1 turno ogni due settimane",
    "2" = "1 turno al mese",
    "1" = "Meno di 1 turno al mese"
  ))
)

# Recode education as a numeric variable
mydata$edu <- as.numeric(
  as.character(fct_recode(
    mydata$education,
    "1" = "Scuola media primaria",
    "2" = "Diploma",
    "3" = "Laurea breve",
    "4" = "Laurea magistrale",
    "5" = "Dottorato"
  ))
)

# Recode last_training as a numeric variable
mydata$training_time <- as.numeric(
  as.character(fct_recode(
    mydata$last_training,
    "1" = "Meno di 2 mesi fa",
    "2" = "Più di 2 mesi fa",
    "3" = "Meno di 6 mesi fa",
    "4" = "Più di 6 mesi fa",
    "5" = "Meno di 1 anno fa",
    "6" = "Più di 1 anno fa"
  ))
)

mydata$job_qualification <- fct_recode(

```

```

mydata$job_qualification,
  "Soccorritore" = "Infermiere",
  "Soccorritore" = "Medico"
)

df <- mydata %>%
  drop_na(ptg)

df <- df %>%
  mutate(
    spirituality_changes = spirituality_changes * 3
  )

df <- df %>%
  mutate(
    red_cross_commeetee_location =
      recode(red_cross_commeetee_location,
        `Comitato del Medio Verbano` = "Comitato di Varese",
        `Comitato del Medio-Verbano` = "Comitato di Varese",
        `Comitato del Agrate` = "Comitato di Monza",
        `Comitato del Agrate Brianza` = "Comitato di Monza",
        `Comitato di Asola` = "Comitato di Mantova",
        `Comitato di Beragamo Hinterland` = "Comitato di Beragamo",
        `Comitato di Beragamo hinterland` = "Comitato di Beragamo",
        `Comitato di Bergamo Ovest e Valle Imagna` = "Comitato di Bergamo",
        `Comitato di Bresso` = "Comitato di Milano",
        `Comitato di Casorate Primo` = "Comitato di Pavia",
        `Comitato di Castiglione della Pescaia` = "Comitato di Grosseto",
        `Comitato di Cinisello balsamo` = "Comitato di Cinisello Balsamo",
        `Comitato di Desenzano` = "Comitato di Brescia",
        `Comitato di Empoli` = "Comitato di Firenze",
        `Comitato di Garbagnate Milanese` = "Comitato di Milano",
        `Comitato di Lentate Sul Seveso` = "Comitato di Monza",
        `Comitato di Lodigiano Ovest` = "Comitato di Lodi",
        `Comitato di Mede e Valle Lomellina` = "Comitato di Pavia",
        `Comitato di Palazzolo sull'oglio` = "Comitato di Palazzolo sull'Oglio",
        `Comitato di Palazzolo sull'Oglio` = "Comitato di Palazzolo sull'Oglio",
        `Comitato di Palazzolo Sull'Oglio` = "Comitato di Palazzolo sull'Oglio",
        `Comitato di Pistoia` = "Comitato di Firenze",
        `Comitato di Pitigliano` = "Comitato di Grosseto",
        `Comitato di Ponte a Egola` = "Comitato di Pisa",
        `Comitato di Ribolla` = "Comitato di Grosseto",
        `Comitato di Rignano sull'Arno` = "Comitato di Firenze",
        `Comitato di San donato milanese` = "Comitato di San Donato Milanese",
        `Comitato di San Giovanni alla Vena` = "Comitato di Pisa",
        `Comitato di Scarlino` = "Comitato di Grosseto",
        `Comitato di Seriate` = "Comitato di Bergamo",

```

```

    `Comitato di Strada in Chianti` = "Comitato di Firenze",
    `Comitato di Suvereto` = "Comitato di Livorno",
    `Comitato di Uliveto Terme` = "Comitato di Pisa",
    `Comitato di Valle Imagna` = "Comitato di Bergamo",
    `Comitato di Venturina Terme` = "Comitato di Livorno",
    `Comitato di Voghea` = "Comitato di Pavia",
    `Comitato di Volterra` = "Comitato di Pisa",
    `Comitato di Rosignano` = "Comitato di Firenze",
    `Comitato Di Bresso` = "Comitato di Milano",
    `Comitato di Beragamo` = "Comitato di Bergamo",
    `Comitato di Bergamo hinterland` = "Comitato di Bergamo",
    `Comitato di Agrate Brianza` = "Comitato di Monza",
    `Comitato di Agrate` = "Comitato di Monza",
    `Comitato di Vigevano` = "Comitato di Pavia",
    `Comitato di San Marcello Pistoiese` = "Comitato di Firenze",
    `Comitato di Pontasserchio` = "Comitato di Pisa",
    `Comitato di Mortara` = "Comitato di Pavia",
    `Comitato di Montorfano` = "Comitato di Como",
    `Comitato di Menaggio` = "Comitato di Como",
    `Comitato di Groane` = "Comitato di Monza",
    `Comitato di Follonica` = "Comitato di Grosseto",
    `Comitato di Cermenate` = "Comitato di Como",
    `Comitato di Castelfranco di Sotto` = "Comitato di Pisa",
    `Comitato di Buccinasco` = "Comitato di Milano",
    `Comitato di Buccinasco` = "Comitato di Milano",
    "Comitato di Beragamo" = "Comitato di Bergamo",
    `Comitato di Bagni di Lucca` = "Comitato di Lucca",
    `Comitato di Albiano Magra` = "Comitato di Firenze"
  )
)

df$red_cross_commeetee_location <- factor(df$red_cross_commeetee_location)

```

Select important variables.

```

vars <- c(
  # ptg
  "life_appreciation", "new_possibilities", "personal_strength",
  "spirituality_changes", "interpersonal_relationships",
  # ptsd
  "avoiding", "intrusivity", "iperarousal",
  # coping
  "social_support", "avoiding_strategies", "positive_attitude",
  "problem_orientation", "transcendent_orientation",
  # perceived social support
  "family", "friends", "significant_other",
  # self-compassion
  "self_judgment", "isolation", "over_identification",

```

```

"self_kindness", "common_humanity", "mindfulness",
# neuroticism
"negative_affect", "self_reproach", "where", "age_imp",
"rescue_worker_qualification", "rate_of_activity",
##
"id", "where", "gender", "age_imp", "edu", "training_time",
"activity_rate", "employment", "red_cross_commeetee_location",
"years_experience", "last_training", "job_qualification",
"is_job_qualification_invariant", "is_team_invariant"
)

temp1 <- df %>%
  dplyr::select(vars)

```

Outliers detection.

```

# temp <- temp1[temp1$age_imp > 20 &
#           temp1$rescue_worker_qualification == "Si" &
#           temp1$rate_of_activity != "Meno di 1 turno al mese", ]

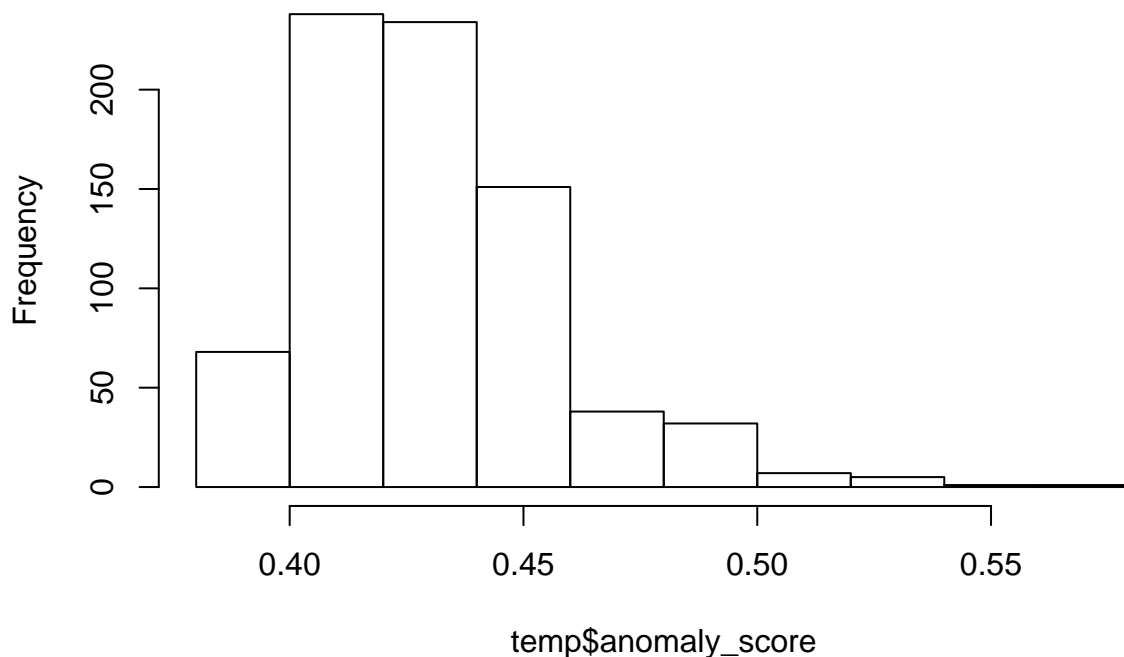
temp <- temp1[temp1$rescue_worker_qualification == "Si", ]

as_df <- isolation.forest(temp, nthreads = 1)
temp$anomaly_score <- predict(as_df, temp)

hist(temp$anomaly_score)

```

Histogram of temp\$anomaly_score



Remove extreme observations.

```
temp$outlier <- as.factor(ifelse(temp$anomaly_score >= 0.5, "outlier", "normal"))
```

```
clean_dat <- temp %>%  
  dplyr::filter(outlier == "normal")
```

```
clean_dat %>%  
  group_by(when) %>%  
  summarise(  
    n = n()  
  )
```

```
## # A tibble: 2 x 2  
##   when      n  
##   <fct>    <int>  
## 1 lombardia  577  
## 2 toscana   184
```

```
clean_dat[clean_dat$red_cross_commeetee_location == "Comitato di Beragamo", ]$red_cross_commeetee_location
```

```
table(clean_dat$red_cross_commeetee_location)
```

```
##  
##          Comitato di Varese          Comitato della Piana Pistoiese  
##                41                5  
##          Comitato di Monza          Comitato di Firenze  
##                20                56  
##          Comitato di Alte Groane    Comitato di Area Nord Milanese  
##                7                9  
##          Comitato di Area Sud Milanese    Comitato di Arezzo  
##                7                8  
##          Comitato di Mantova          Comitato di Lucca  
##                8                10  
##          Comitato di Bagno a Ripoli    Comitato di Basso Lario  
##                15                4  
##          Comitato di Beragamo          Comitato di Bergamo  
##                0                24  
##          Comitato di Bergamo Hinterland    Comitato di Bibbiena  
##                27                3  
##          Comitato di Brescia          Comitato di Milano  
##                6                23  
##          Comitato di Brugherio    Comitato di Busto Arsizio  
##                11                8  
##          Comitato di Calvisano          Comitato di Cantù  
##                11                7  
##          Comitato di Caravaggio    Comitato di Casatenovo  
##                5                5  
##          Comitato di Pavia          Comitato di Casteggio  
##                19                9
```

##	Comitato di Pisa	Comitato di Grosseto
##	20	12
##	Comitato di Castiglione delle Stiviere	Comitato di Como
##	3	28
##	Comitato di Certaldo	Comitato di Cinisello Balsamo
##	3	5
##	Comitato di Codogno	Comitato di Colico
##	4	3
##	Comitato di Crema	Comitato di Cremona
##	9	9
##	Comitato di Desio	Comitato di Donoratico
##	3	4
##	Comitato di Gallarate	Comitato di Gardone Val Trompia
##	15	4
##	Comitato di Ghedi	Comitato di Grandate
##	5	6
##	Comitato di Incisa Valdarno	Comitato di Linate
##	5	5
##	Comitato di Lecco	Comitato di Legnano
##	9	24
##	Comitato di Lentate sul Seveso	Comitato di Livorno
##	3	10
##	Comitato di Lodi	Comitato di Lomazzo
##	25	4
##	Comitato di Luino	Comitato di Lurate Caccivio
##	7	4
##	Comitato di Massa	Comitato di Medio Verbano
##	4	8
##	Comitato di Merate	Comitato di Milano Est
##	9	5
##	Comitato di Morbegno	Comitato di Muggiò
##	6	3
##	Comitato di Orbetello	Comitato di Paderno Dugnano
##	4	11
##	Comitato di Palazzolo sull'Oglio	Comitato di Pieve Emanuele
##	10	5
##	Comitato di Piombino	Comitato di Pizzighettone
##	2	3
##	Comitato di Prato	Comitato di San Donato Milanese
##	5	10
##	Comitato di San Fermo della Battaglia	Comitato di San Frediano a Settimo
##	4	3
##	Comitato di Sansepolcro	Comitato di Saronno
##	3	15
##	Comitato di Scandicci	Comitato di Sesto San Giovanni
##	4	3
##	Comitato di Settimo Milanese	Comitato di Siena
##	6	1

##	Comitato di Sondrio	Comitato di Stradella
##	7	5
##	Comitato di Suzzara	Comitato di Treviglio
##	3	7
##	Comitato di Triangolo Lariano	Comitato di Uggiate Trevano
##	3	8
##	Comitato di Valceresio	Comitato di Varedo
##	3	3
##	Comitato di Vernio	Comitato di Viareggio
##	3	4
##	Comitato di Villasanta	Comitato di Voghera
##	7	7

```
clean_dat <- clean_dat[!duplicated(clean_dat$id), ]
```

Save cleaned data.

```
saveRDS(clean_dat, here("data", "processed", "rescue_workers_cleaned_data.rds"))
```