IRR_yelc_finetune

2023-10-31

Data

```
data_p <- read.csv("yelc_scoring_physical.csv") # physical punishment topic
data_d <- read.csv("yelc_scoring_driving.csv") # driving topic
data_m <- read.csv("yelc_scoring_medical.csv") # medical issues topic

rating_h_g <- data_p[, c("Human_Score", "GPT35turbo")]
#_h indicate human scoring; _g indicate original ChatGPT(GPT3.5Oturbo)
rating_h_f1 <- data_p[, c("Human_Score", "FineTune_1st_Score")]
# _f1 indicate 1st scoring with a fine-tuned model
rating_h_f2 <- data_p[, c("Human_Score", "FineTune_2nd_Score")]
# _f2 indicate 2nd scoring with a fine-tuned model
rating_f1_f2 <- data_p[, c("FineTune_1st_Score", "FineTune_2nd_Score")]

rating_h_d <- data_d[, c("Human_Score", "finetune_driving")]
rating_h_m <- data_m[, c("Human_Score", "finetune_medical")]
```

Spearman Correlation

```
# Phsyical Punishment Essay
cor(rating_h_g$Human_Score, rating_h_g$GPT35turbo, method = "spearman")
```

```
## [1] 0.2670511
```

```
cor(rating_h_f1$Human_Score, rating_h_f1$FineTune_1st_Score, method = "spearman")
```

```
## [1] 0.6721604
```

```
cor(rating_h_f2$Human_Score, rating_h_f2$FineTune_2nd_Score, method = "spearman")
```

```
## [1] 0.6751483
```

```
cor(rating_f1_f2$FineTune_1st_Score, rating_f1_f2$FineTune_2nd_Score, method = "spearman")
```

```
## [1] 0.9236965
```

```r
cor(rating_h_d$Human_Score, rating_h_d$finetune_driving, method = "spearman")
```

```
## [1] 0.8228037
```

```r
cor(rating_h_m$Human_Score, rating_h_m$finetune_medical, method = "spearman")
```

```
## [1] 0.7361989
```

Percentage Agreement

```r
#install.packages("irr")
library(irr)
```

```
## Loading required package: lpSolve
```

```r
# Physical Punishment Essay

agree(rating_h_g, tolerance= 0) # tolerance = 0 : exact agreement
```

```
##  Percentage agreement (Tolerance=0)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 26
```

```r
agree(rating_h_g, tolerance= 1)
```

```
##  Percentage agreement (Tolerance=1)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 64.1
```

```r
agree(rating_h_g, tolerance= 2)
```

```
##  Percentage agreement (Tolerance=2)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 88.4
```

```r
agree(rating_h_f1, tolerance= 0)
```

```
##  Percentage agreement (Tolerance=0)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 44.8
```

```r
agree(rating_h_f1, tolerance= 1)
```

```
##  Percentage agreement (Tolerance=1)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 91.2
```

```r
agree(rating_h_f1, tolerance= 2)
```

```
##  Percentage agreement (Tolerance=2)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 100
```

```r
agree(rating_h_f2, tolerance= 0)
```

```
##  Percentage agreement (Tolerance=0)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 43.1
```

```r
agree(rating_h_f2, tolerance= 1)
```

```
##  Percentage agreement (Tolerance=1)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 91.7
```

```r
agree(rating_h_f2, tolerance= 2)
```

```
##  Percentage agreement (Tolerance=2)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 100
```

```r
agree(rating_f1_f2, tolerance= 0)
```

```
##  Percentage agreement (Tolerance=0)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 90.1
```

```r
agree(rating_f1_f2, tolerance= 1)
```

```
##  Percentage agreement (Tolerance=1)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 98.9
```

```r
agree(rating_f1_f2, tolerance= 2)
```

```
##  Percentage agreement (Tolerance=2)
##
##  Subjects = 181
##    Raters = 2
##   %-agree = 100
```

```r
agree(rating_h_d, tolerance= 0)
```

```
##  Percentage agreement (Tolerance=0)
##
##  Subjects = 100
##    Raters = 2
##   %-agree = 29
```

```r
agree(rating_h_d, tolerance= 1)
```

```
##  Percentage agreement (Tolerance=1)
##
##  Subjects = 100
##    Raters = 2
##   %-agree = 81
```

```r
agree(rating_h_d, tolerance= 2)
```

```
##  Percentage agreement (Tolerance=2)
##
##  Subjects = 100
##    Raters = 2
##   %-agree = 100
```

```r
agree(rating_h_m, tolerance= 0)
```

```
##  Percentage agreement (Tolerance=0)
##
##  Subjects = 100
##    Raters = 2
##   %-agree = 31
```

```
agree(rating_h_m, tolerance= 1)
```

```
##  Percentage agreement (Tolerance=1)
##
##  Subjects = 100
##    Raters = 2
##   %-agree = 75
```

```
agree(rating_h_m, tolerance= 2)
```

```
##  Percentage agreement (Tolerance=2)
##
##  Subjects = 100
##    Raters = 2
##   %-agree = 98
```

Cohen's Kappa

```
kappa2(rating_h_g)
```

```
##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 181
##    Raters = 2
##     Kappa = 0.0539
##
##         z = 1.5
##   p-value = 0.133
```

```
kappa2(rating_h_f1)
```

```
##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 181
##    Raters = 2
##     Kappa = 0.238
##
##         z = 6.02
##   p-value = 1.77e-09
```

```
kappa2(rating_h_f2)
```

```
##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 181
##    Raters = 2
##     Kappa = 0.221
##
##         z = 5.6
##   p-value = 2.2e-08
```

```
kappa2(rating_f1_f2)
```

```
##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 181
##    Raters = 2
##     Kappa = 0.843
##
##         z = 18.2
##   p-value = 0
```

```
kappa2(rating_h_d)
```

```
##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 100
##    Raters = 2
##     Kappa = 0.149
##
##         z = 3.54
##   p-value = 0.000396
```

```
kappa2(rating_h_m)
```

```
##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 100
##    Raters = 2
##     Kappa = 0.173
##
##         z = 4.07
##   p-value = 4.62e-05
```

Krippendorff's Alpha

```
# https://rpubs.com/jacoblong/content-analysis-krippendorff-alpha-R
for (p in c("tidyverse", "irr")) {
  if (!requireNamespace(p)) {
    install.packages(p)
  }
}
```

```
## Loading required namespace: tidyverse
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
```

```
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lpSolve)
library(irr)
```

```
# install.packages("devtools", dependencies = TRUE)
# devtools::install_github("mikegruz/kripp.boot")
library(kripp.boot)
```

```
rating_h_g_t <- t(as.matrix(rating_h_g)) # _t indicate transpose matrix

rating_h_f1_t <- t(as.matrix(rating_h_f1))
rating_h_f2_t <- t(as.matrix(rating_h_f2))
rating_f1_f2_t <- t(as.matrix(rating_f1_f2))

rating_h_d_t <- t(as.matrix(rating_h_d))
rating_h_m_t <- t(as.matrix(rating_h_m))
```

```
kripp.alpha(rating_h_f1_t, method = "ordinal") # nominal, ordinal, interval, or ratio data.
```

```
##  Krippendorff's alpha
##
##  Subjects = 181
##    Raters = 2
##     alpha = 0.643
```

```
kripp.alpha(rating_h_f2_t, method = "ordinal") # nominal, ordinal, interval, or ratio data.
```

```
##  Krippendorff's alpha
##
##  Subjects = 181
##    Raters = 2
##     alpha = 0.65
```

```
kripp.alpha(rating_f1_f2_t, method = "ordinal") # nominal, ordinal, interval, or ratio data.
```

```
##  Krippendorff's alpha
##
##  Subjects = 181
##    Raters = 2
##     alpha = 0.922
```

```r
kripp.alpha(rating_h_d_t, method = "ordinal") # nominal, ordinal, interval, or ratio data.
```

```
##  Krippendorff's alpha
##
##  Subjects = 100
##    Raters = 2
##     alpha = 0.766
```

```r
kripp.alpha(rating_h_m_t, method = "ordinal") # nominal, ordinal, interval, or ratio data.
```

```
##  Krippendorff's alpha
##
##  Subjects = 100
##    Raters = 2
##     alpha = 0.684
```

End