Multivariate\_Analysis

Ishana Rana

2025-07-10

# Multivariate Analysis

The goal of this part is to analyze potential inter-metric correlations.

## Loading datasets:

bertscore <- read.csv("C:\\Users\\babus\\OneDrive\\Documents\\uni uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-ASR\\data\\eval\_results\\BERTScore\_scores.csv", skip=1)  
str(bertscore)

## 'data.frame': 6 obs. of 3 variables:  
## $ File : chr "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...  
## $ with\_punct : num 84.7 88 84 91.1 94 ...  
## $ without\_punct: num 86 89.3 86 93.7 95.8 ...

head(bertscore)

## File with\_punct without\_punct  
## 1 EN2009c 84.66 85.98  
## 2 EN2009d 87.98 89.34  
## 3 ES2016a 84.01 86.02  
## 4 ES2016b 91.10 93.74  
## 5 ES2016c 93.98 95.83  
## 6 ES2016d 89.66 90.12

summary(bertscore)

## File with\_punct without\_punct   
## Length:6 Min. :84.01 Min. :85.98   
## Class :character 1st Qu.:85.49 1st Qu.:86.85   
## Mode :character Median :88.82 Median :89.73   
## Mean :88.56 Mean :90.17   
## 3rd Qu.:90.74 3rd Qu.:92.83   
## Max. :93.98 Max. :95.83

bleu <- read.csv("C:\\Users\\babus\\OneDrive\\Documents\\uni uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-ASR\\data\\eval\_results\\BLEU\_scores.csv")  
str(bleu)

## 'data.frame': 6 obs. of 3 variables:  
## $ File : chr "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...  
## $ with\_punct : num 38.8 36.9 43.6 54.3 50 ...  
## $ without\_punct: num 38.1 38.5 49.1 59.4 55.1 ...

head(bleu)

## File with\_punct without\_punct  
## 1 EN2009c 38.77 38.14  
## 2 EN2009d 36.93 38.54  
## 3 ES2016a 43.61 49.07  
## 4 ES2016b 54.26 59.40  
## 5 ES2016c 49.96 55.11  
## 6 ES2016d 30.65 31.81

summary(bleu)

## File with\_punct without\_punct   
## Length:6 Min. :30.65 Min. :31.81   
## Class :character 1st Qu.:37.39 1st Qu.:38.24   
## Mode :character Median :41.19 Median :43.80   
## Mean :42.36 Mean :45.34   
## 3rd Qu.:48.37 3rd Qu.:53.60   
## Max. :54.26 Max. :59.40

# splitting dataset as it contains scores of ROUGE-1 and ROUGE-L  
lines <- readLines("C:\\Users\\babus\\OneDrive\\Documents\\uni uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-ASR\\data\\eval\_results\\ROUGE\_scores.csv")  
split\_index <- grep("ROUGE-L", lines)  
  
rouge1\_lines <- lines[2:(split\_index - 1)]  
rougel\_lines <- lines[(split\_index + 1):length(lines)]  
  
rouge1 <- read.csv(text = rouge1\_lines)  
rougel <- read.csv(text = rougel\_lines)  
  
str(rouge1)

## 'data.frame': 6 obs. of 3 variables:  
## $ File : chr "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...  
## $ with\_punct : num 74.9 72.9 80.1 85 83 ...  
## $ without\_punct: num 74.8 73 80 84.9 82.8 ...

head(rouge1)

## File with\_punct without\_punct  
## 1 EN2009c 74.92 74.76  
## 2 EN2009d 72.91 72.97  
## 3 ES2016a 80.14 80.04  
## 4 ES2016b 84.99 84.93  
## 5 ES2016c 83.00 82.82  
## 6 ES2016d 70.03 69.89

summary(rouge1)

## File with\_punct without\_punct   
## Length:6 Min. :70.03 Min. :69.89   
## Class :character 1st Qu.:73.41 1st Qu.:73.42   
## Mode :character Median :77.53 Median :77.40   
## Mean :77.67 Mean :77.57   
## 3rd Qu.:82.28 3rd Qu.:82.12   
## Max. :84.99 Max. :84.93

str(rougel)

## 'data.frame': 6 obs. of 3 variables:  
## $ File : chr "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...  
## $ with\_punct : num 39.4 35.1 39.4 37.2 46.9 ...  
## $ without\_punct: num 39.2 35.1 39.3 37.2 46.8 ...

head(rougel)

## File with\_punct without\_punct  
## 1 EN2009c 39.35 39.22  
## 2 EN2009d 35.10 35.10  
## 3 ES2016a 39.37 39.28  
## 4 ES2016b 37.24 37.22  
## 5 ES2016c 46.86 46.83  
## 6 ES2016d 31.62 31.30

summary(rougel)

## File with\_punct without\_punct   
## Length:6 Min. :31.62 Min. :31.30   
## Class :character 1st Qu.:35.63 1st Qu.:35.63   
## Mode :character Median :38.30 Median :38.22   
## Mean :38.26 Mean :38.16   
## 3rd Qu.:39.37 3rd Qu.:39.27   
## Max. :46.86 Max. :46.83

wer <- read.csv("C:\\Users\\babus\\OneDrive\\Documents\\uni uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-ASR\\data\\eval\_results\\WER\_scores.csv", skip=1)  
str(wer)

## 'data.frame': 6 obs. of 3 variables:  
## $ File : chr "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...  
## $ with\_punct : num 89.1 89.9 88.7 82.9 77 ...  
## $ without\_punct: num 86.2 87.5 86.4 78.3 71.6 ...

head(wer)

## File with\_punct without\_punct  
## 1 EN2009c 89.12 86.16  
## 2 EN2009d 89.94 87.55  
## 3 ES2016a 88.73 86.38  
## 4 ES2016b 82.87 78.33  
## 5 ES2016c 77.02 71.61  
## 6 ES2016d 86.95 83.21

summary(wer)

## File with\_punct without\_punct   
## Length:6 Min. :77.02 Min. :71.61   
## Class :character 1st Qu.:83.89 1st Qu.:79.55   
## Mode :character Median :87.84 Median :84.69   
## Mean :85.77 Mean :82.21   
## 3rd Qu.:89.02 3rd Qu.:86.33   
## Max. :89.94 Max. :87.55

## 

## Analysis:

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.4.3

#install.packages("corrplot")  
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.4.3

## corrplot 0.95 loaded

# specifying directory where the created plots shall be saved at  
plot\_dir = "C:\\Users\\babus\\OneDrive\\Documents\\uni uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-ASR\\data\\eval\_results\\R\_code\\analysis inbetween"  
dir.create(plot\_dir, showWarnings = FALSE)

### Correlation Matrix

#### Sidenote

**Because of the small sample size these results also lack statistical power.**

#### with punctuation:

metrics\_df\_with\_punct <- data.frame(  
 File = wer$File,  
 WER = wer$with\_punct,  
 BLEU = bleu$with\_punct,  
 ROUGE1 = rouge1$with\_punct,  
 ROUGEL = rougel$with\_punct,  
 BERTScore = bertscore$with\_punct  
)  
metrics\_df\_with\_punct

## File WER BLEU ROUGE1 ROUGEL BERTScore  
## 1 EN2009c 89.12 38.77 74.92 39.35 84.66  
## 2 EN2009d 89.94 36.93 72.91 35.10 87.98  
## 3 ES2016a 88.73 43.61 80.14 39.37 84.01  
## 4 ES2016b 82.87 54.26 84.99 37.24 91.10  
## 5 ES2016c 77.02 49.96 83.00 46.86 93.98  
## 6 ES2016d 86.95 30.65 70.03 31.62 89.66

# Compute correlation matrix (Spearman is more robust for small samples)  
cor\_matrix\_with\_punct <- cor(metrics\_df\_with\_punct[,-1], method = "spearman")  
print(cor\_matrix\_with\_punct)

## WER BLEU ROUGE1 ROUGEL BERTScore  
## WER 1.0000000 -0.6000000 -0.6000000 -0.37142857 -0.77142857  
## BLEU -0.6000000 1.0000000 1.0000000 0.65714286 0.37142857  
## ROUGE1 -0.6000000 1.0000000 1.0000000 0.65714286 0.37142857  
## ROUGEL -0.3714286 0.6571429 0.6571429 1.00000000 0.02857143  
## BERTScore -0.7714286 0.3714286 0.3714286 0.02857143 1.00000000

# save output to image:  
png(file.path(plot\_dir, "with\_punct\_corr\_matrix\_circle.png"), width = 1000, height = 800, res = 150)  
corrplot(cor\_matrix\_with\_punct, method = "circle", type = "upper", tl.cex = 0.8)  
dev.off()

## png   
## 2

png(file.path(plot\_dir, "with\_punct\_corr\_matrix\_number.png"), width = 1000, height = 800, res = 150)  
corrplot(cor\_matrix\_with\_punct, method = "number", type = "upper", tl.cex = 0.8)  
dev.off()

## png   
## 2

#### without punctuation:

metrics\_df\_without\_punct<- data.frame(  
 File = wer$File,  
 WER = wer$without\_punct,  
 BLEU = bleu$without\_punct,  
 ROUGE1 = rouge1$without\_punct,  
 ROUGEL = rougel$without\_punct,  
 BERTScore = bertscore$without\_punct  
)  
metrics\_df\_without\_punct

## File WER BLEU ROUGE1 ROUGEL BERTScore  
## 1 EN2009c 86.16 38.14 74.76 39.22 85.98  
## 2 EN2009d 87.55 38.54 72.97 35.10 89.34  
## 3 ES2016a 86.38 49.07 80.04 39.28 86.02  
## 4 ES2016b 78.33 59.40 84.93 37.22 93.74  
## 5 ES2016c 71.61 55.11 82.82 46.83 95.83  
## 6 ES2016d 83.21 31.81 69.89 31.30 90.12

cor\_matrix\_without\_punct <- cor(metrics\_df\_without\_punct[,-1], method = "spearman")  
print(cor\_matrix\_without\_punct)

## WER BLEU ROUGE1 ROUGEL BERTScore  
## WER 1.0000000 -0.4285714 -0.5428571 -0.31428571 -0.77142857  
## BLEU -0.4285714 1.0000000 0.9428571 0.54285714 0.54285714  
## ROUGE1 -0.5428571 0.9428571 1.0000000 0.65714286 0.42857143  
## ROUGEL -0.3142857 0.5428571 0.6571429 1.00000000 0.08571429  
## BERTScore -0.7714286 0.5428571 0.4285714 0.08571429 1.00000000

# save output to image:  
png(file.path(plot\_dir, "without\_punct\_corr\_matrix\_circle.png"), width = 1000, height = 800, res = 150)  
corrplot(cor\_matrix\_without\_punct, method = "circle", type = "upper", tl.cex = 0.8)  
dev.off()

## png   
## 2

png(file.path(plot\_dir, "without\_punct\_corr\_matrix\_number.png"), width = 1000, height = 800, res = 150)  
corrplot(cor\_matrix\_without\_punct, method = "number", type = "upper", tl.cex = 0.8)  
dev.off()

## png   
## 2

# Testing Correlation Significance

#### with punctuation:

cor.test(metrics\_df\_with\_punct$WER, metrics\_df\_with\_punct$BERTScore, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$WER and metrics\_df\_with\_punct$BERTScore  
## S = 62, p-value = 0.1028  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## -0.7714286

cor.test(metrics\_df\_with\_punct$WER, metrics\_df\_with\_punct$ROUGE1, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$WER and metrics\_df\_with\_punct$ROUGE1  
## S = 56, p-value = 0.2417  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## -0.6

cor.test(metrics\_df\_with\_punct$WER, metrics\_df\_with\_punct$ROUGEL, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$WER and metrics\_df\_with\_punct$ROUGEL  
## S = 48, p-value = 0.4972  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## -0.3714286

cor.test(metrics\_df\_with\_punct$WER, metrics\_df\_with\_punct$BLEU, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$WER and metrics\_df\_with\_punct$BLEU  
## S = 56, p-value = 0.2417  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## -0.6

The results show that WER has: - strong negative relationship with BERTScore but is statistically not significant - has moderate negative correlations with ROUGE-1, ROUGE-L and BLEU but is statistically also not significant

cor.test(metrics\_df\_with\_punct$BLEU, metrics\_df\_with\_punct$ROUGE1, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$BLEU and metrics\_df\_with\_punct$ROUGE1  
## S = 0, p-value = 0.002778  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 1

cor.test(metrics\_df\_with\_punct$BLEU, metrics\_df\_with\_punct$ROUGEL, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$BLEU and metrics\_df\_with\_punct$ROUGEL  
## S = 12, p-value = 0.175  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.6571429

These results show that BLEU has:

- perfect positive correlation with ROUGE-1 that is also statistically significant

- moderate strong positive relation with ROUGE-L but not statistically significant

cor.test(metrics\_df\_with\_punct$ROUGEL, metrics\_df\_with\_punct$BERTScore, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_with\_punct$ROUGEL and metrics\_df\_with\_punct$BERTScore  
## S = 34, p-value = 1  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.02857143

This test shows that there is no correlation at all, indicating they capture different linguistic phenomena.

#### Without punctuation:

For this condition, the pairs that showed some difference to the “with\_punct” ones were tested. None of them showed any statistically significant correlation.

cor.test(metrics\_df\_without\_punct$WER, metrics\_df\_without\_punct$BLEU, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_without\_punct$WER and metrics\_df\_without\_punct$BLEU  
## S = 50, p-value = 0.4194  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## -0.4285714

cor.test(metrics\_df\_without\_punct$BLEU, metrics\_df\_without\_punct$ROUGEL, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_without\_punct$BLEU and metrics\_df\_without\_punct$ROUGEL  
## S = 16, p-value = 0.2972  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.5428571

cor.test(metrics\_df\_without\_punct$BLEU, metrics\_df\_without\_punct$BERTScore, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_without\_punct$BLEU and metrics\_df\_without\_punct$BERTScore  
## S = 16, p-value = 0.2972  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.5428571

cor.test(metrics\_df\_without\_punct$ROUGEL, metrics\_df\_without\_punct$BERTScore, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: metrics\_df\_without\_punct$ROUGEL and metrics\_df\_without\_punct$BERTScore  
## S = 32, p-value = 0.9194  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.08571429