

# Modeling Progress in Speech Recognition Models

2024-09-13

# R Markdown

```
library(rstatix)
```

```
##
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':
##
##     filter
```

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
url="https://raw.githubusercontent.com/hamlel/Carreras-con-Impacto/main/Speech%20Recognition%20Models%20
speech<-read_csv(url(url))
```

```
## Rows: 181 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr  (8): Nombre, Link, Task, Benchmark, Modelo, CER, Arquitectura, FLOPS
## dbl  (3): WER, Parametros M, Datos de entrenamiento en horas
## lgl  (1): ExtraTraining Data
## date (1): Fecha
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
str(speech)
```

```
## spec_tbl_ [181 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Nombre : chr [1:181] "A Comparative Study on Transformer vs RNN in Speech
## $ Link : chr [1:181] "https://arxiv.org/pdf/1909.06317v2" "https://arxiv.
## $ Task : chr [1:181] "Speech Recognition" "Speech Recognition" "Speech Re
## $ Benchmark : chr [1:181] "AISHELL-1" "LibriSpeech test-clean" "LibriSpeech te
## $ Modelo : chr [1:181] "CTC/Att" "Transformer" "Transformer" "AmNet" ...
```

```
## $ WER : num [1:181] 6.7 2.6 5.7 8.6 4.46 1.75 6.7 8.2 1.5 1 ...
## $ CER : chr [1:181] NA NA NA NA ...
## $ ExtraTraining Data : logi [1:181] FALSE TRUE TRUE FALSE FALSE FALSE ...
## $ Parametros M : num [1:181] NA NA NA 3.3e+07 NA NA NA NA NA ...
## $ Arquitectura : chr [1:181] NA "Transformer" "Transformer" NA ...
## $ Fecha : Date[1:181], format: "2019-09-13" "2019-09-13" ...
## $ FLOPS : chr [1:181] NA NA NA "2.53E+07" ...
## $ Datos de entrenamiento en horas: num [1:181] 170 960 960 960 960 ...
## - attr(*, "spec")=
## .. cols(
## .. Nombre = col_character(),
## .. Link = col_character(),
## .. Task = col_character(),
## .. Benchmark = col_character(),
## .. Modelo = col_character(),
## .. WER = col_double(),
## .. CER = col_character(),
## .. `ExtraTraining Data` = col_logical(),
## .. `Parametros M` = col_double(),
## .. Arquitectura = col_character(),
## .. Fecha = col_date(format = ""),
## .. FLOPS = col_character(),
## .. `Datos de entrenamiento en horas` = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
speech<- speech %>% filter(Benchmark %in% c("AISHELL-1","Common Voice French",
"Common Voice German","Common Voice Spanish",
"LRS2","LibriSpeech test-clean",
"LibriSpeech test-other","WSJ eval92"))
```

*#estadísticas descriptivas por Benchmark*

```
speech%>%
  group_by(Benchmark) %>%
  get_summary_stats(WER, type = "common")
```

```
## # A tibble: 8 x 11
##   Benchmark      variable      n   min   max median   iqr  mean    sd    se    ci
##   <chr>          <fct>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 AISHELL-1      WER        14  1.29  18.7   4.72  1.90   5.41  4.18  1.12  2.41
## 2 Common Voice ~ WER         8  8.13  13.9   9.91  1.78  10.5   1.83  0.647  1.53
## 3 Common Voice ~ WER        14  3.64  12.1   6.5   2.82   6.62  2.39  0.639  1.38
## 4 Common Voice ~ WER         8  5.5   10.5   7.18  4.34   7.70  2.15  0.761  1.8
## 5 LRS2           WER         7  1.5   10.1   6.7   5.65   6.11  3.45  1.30  3.19
## 6 LibriSpeech t~ WER        57  1.34   8.6    2.3   1.5    3.00  1.72  0.227  0.455
## 7 LibriSpeech t~ WER        48  2.48  16.5   4.68  3.15   5.89  3.33  0.481  0.967
## 8 WSJ eval92     WER        16  1.3    6.9    3.3   0.942  3.32  1.40  0.351  0.748
```

*#Descriptivos por Arquitectura*

```
speech%>%
  group_by(Arquitectura) %>%
  get_summary_stats(WER, type = "common")
```

```
## Warning: There were 3 warnings in `mutate()`.
## The first warning was:
```

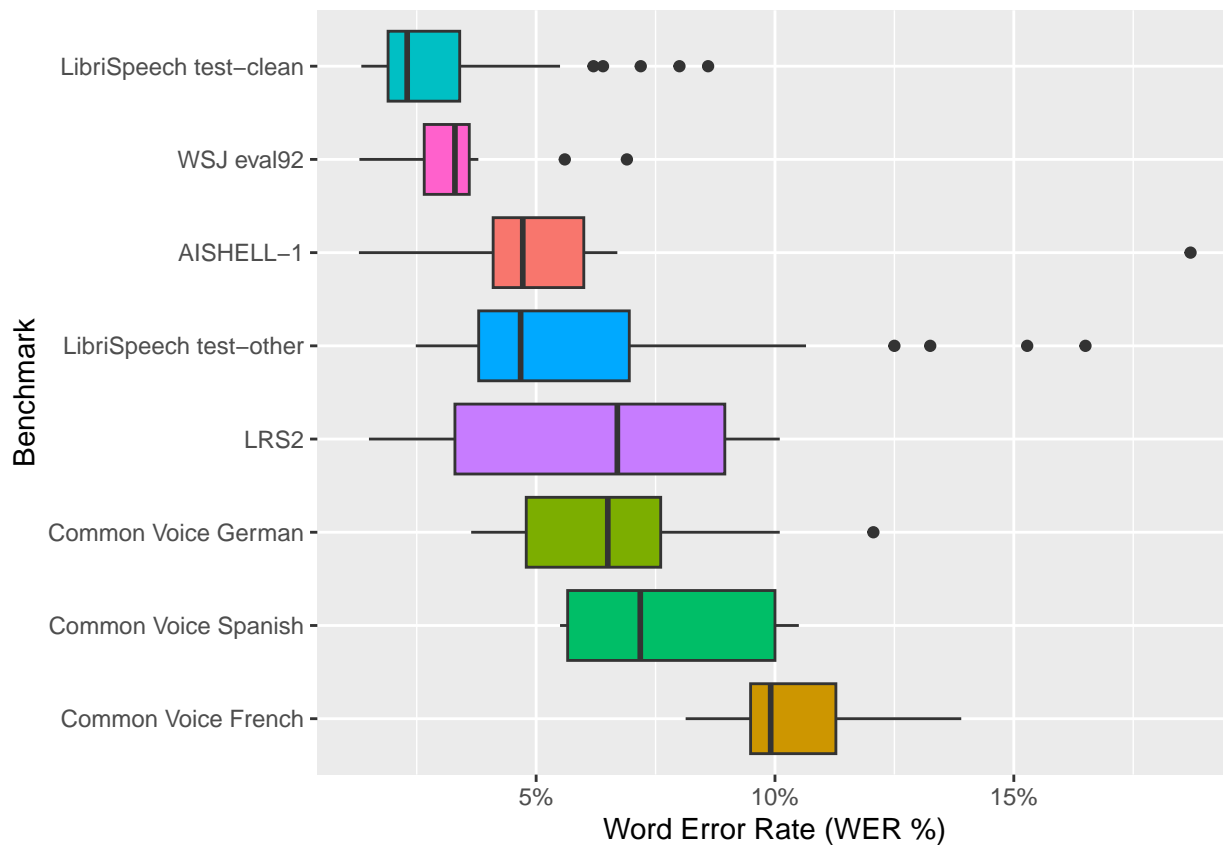
```
## i In argument: `ci = abs(stats::qt(alpha/2, .data$n - 1) * .data$se)`.
## Caused by warning:
## ! There was 1 warning in `mutate()`.
## i In argument: `ci = abs(stats::qt(alpha/2, .data$n - 1) * .data$se)`.
## Caused by warning in `stats::qt()`:
## ! NaNs produced
## i Run `dplyr::last_dplyr_warnings()` to see the 2 remaining warnings.

## # A tibble: 11 x 11
##   Arquitectura variable      n   min   max median   iqr  mean    sd    se
##   <chr>          <fct>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 CNN           WER          3  1.75 10.5   5.8  4.36  6.01  4.36  2.52
## 2 CNN-LSTM      WER          4  3.4   5.5   4.3  0.825 4.38  0.877 0.439
## 3 Conformer     WER         13  1.7   5.97  2.6  2      3.17  1.36  0.377
## 4 E-Branchformer WER          1  1.81  1.81  1.81  0      1.81  NA    NA
## 5 HMM           WER          4  4.3   8      5.15 1.45  5.65  1.64  0.821
## 6 Knowledge Destil~ WER          1  4.1   4.1   4.1  0      4.1  NA    NA
## 7 LSTM          WER          8  1.7   7.63  3.35 1.41  3.84  1.92  0.678
## 8 Pretrained     WER          1  1.9   1.9   1.9  0      1.9  NA    NA
## 9 Transformer    WER         15  1.5   5.7   2.26 0.88  2.68  1.16  0.299
## 10 Transformer-CNN WER          4  3.1   4.11  3.7  0.853 3.65  0.529 0.264
## 11 <NA>          WER        118  1.29 18.7   4.9  4.57  5.68  3.53  0.325
## # i 1 more variable: ci <dbl>
```

*#Por Benchmark*

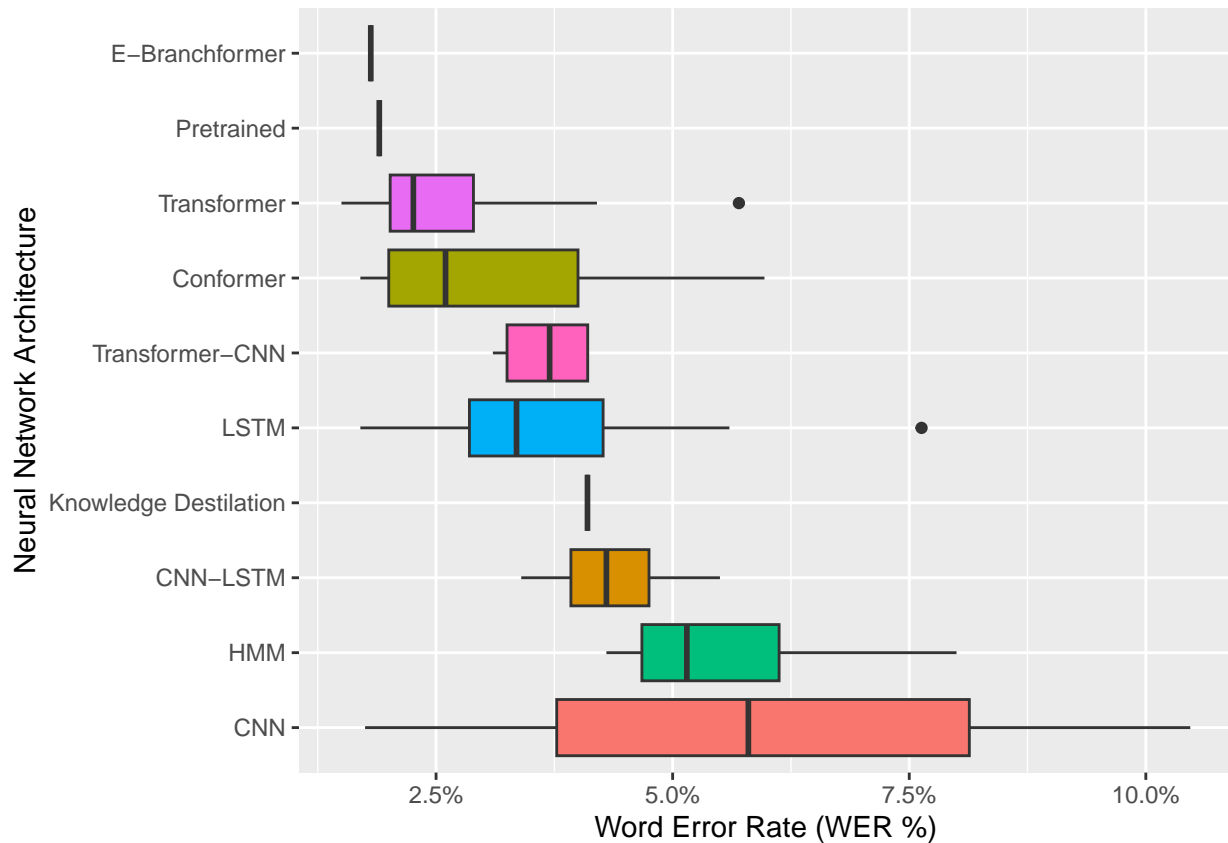
```
ggplot(speech,aes(reorder(Benchmark,-WER),WER/100,fill=Benchmark))+
  geom_boxplot()+coord_flip()+guides(fill=FALSE)+
  labs(x="Benchmark",y="Word Error Rate (WER %)")+
  scale_y_continuous(labels = scales::percent)
```

```
## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as
## of ggplot2 3.3.4.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



*#Por arquitectura*

```
ggplot(speech%>%filter(Arquitectura != "NA"),aes(reorder(Arquitectura,-WER),WER/100,fill=Arquitectura))
  geom_boxplot()+coord_flip()+guides(fill=FALSE)+
  labs(x="Neural Network Architecture",y="Word Error Rate (WER %)")+
  scale_y_continuous(labels = scales::percent)
```



```
#linear regression
```

```
ggplot(speech%>%filter(Benchmark %in% c("AISHELL-1", "LibriSpeech test-clean",
                                         "LibriSpeech test-other", "WSJ eval92")), aes(Fecha, WER/100, fill=
                                         geom_smooth(method=lm)+geom_point()+
                                         facet_wrap(~Benchmark)+guides(fill=FALSE, shape=FALSE)+
                                         scale_y_continuous(labels = scales::percent)+
                                         labs(x="Date", y="Word Error Rate (WER %)"))
```

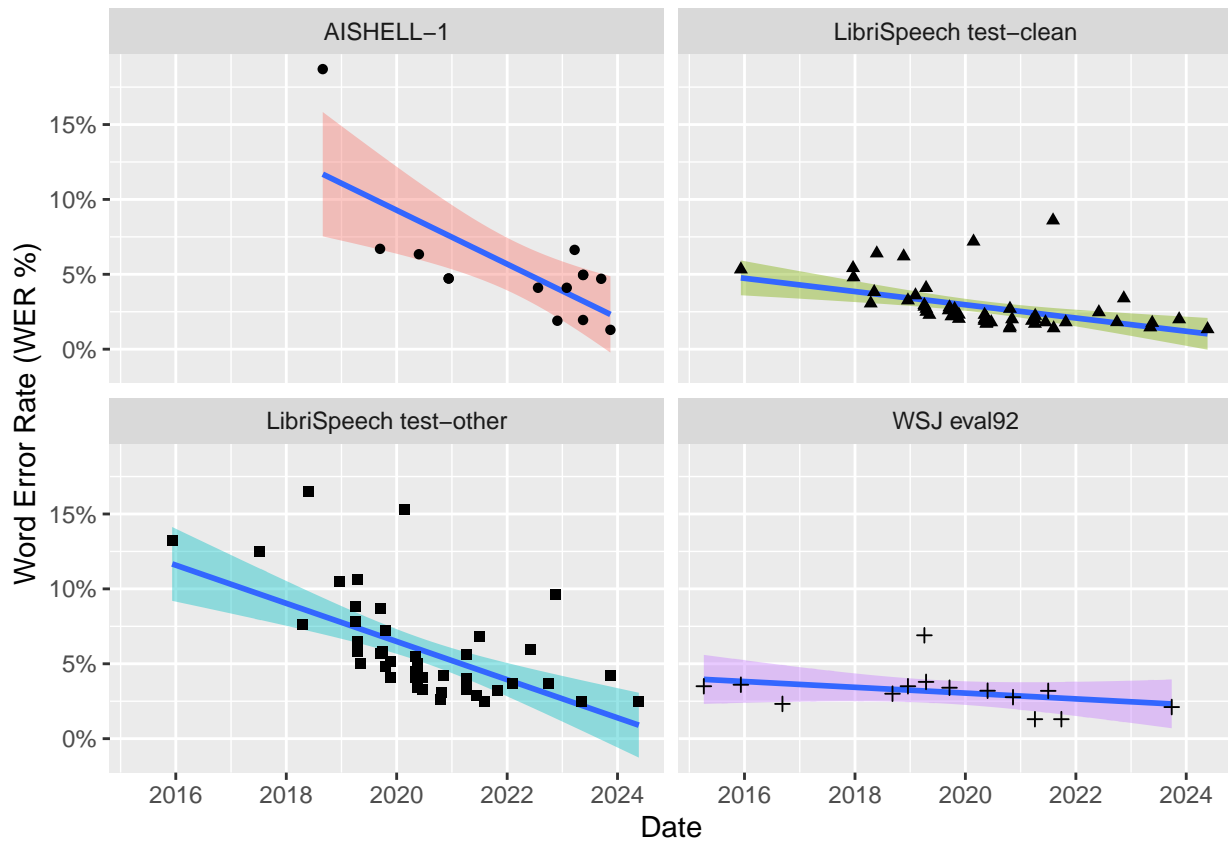
```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
```

```
## (`stat_smooth()`).
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
```

```
## (`geom_point()`).
```



```
ggplot(speech%>%filter(Benchmark %in% c("Common Voice French", "Common Voice Spanish",
                                         "Common Voice German", "LRS2")), aes(Fecha, WER/100, fill=Benchmark)) +
  geom_smooth(method=lm) + geom_point() +
  facet_wrap(~Benchmark) + guides(fill=FALSE, shape=FALSE) +
  scale_y_continuous(labels = scales::percent) +
  labs(x="Date", y="Word Error Rate (WER %)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

