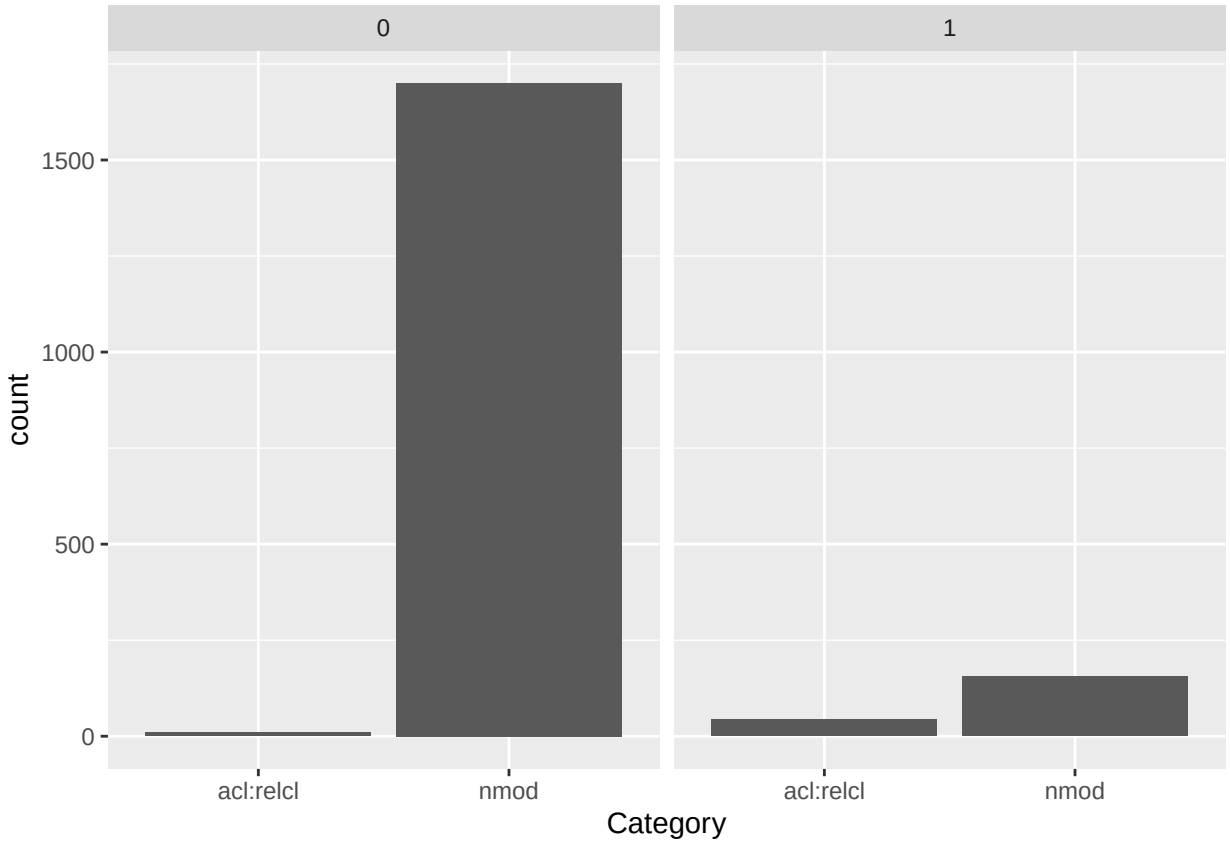
Right Extraposition

2024-05-07

Dialgoue Data

1. Processing the Data

```
##                                                                                      File
## 1 ../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 2 ../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 3 ../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 4 ../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 5 ../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
## 6 ../Dialouge Corpus Filttered/parse_gold_filttered\\Phase1\\Phase1_gold\\hi_1385_gold.conllu
##   Sent_ID                                                         Sentence
## 1      29
## 2      73                             [incomprehensible]
## 3      94
## 4      94
## 5     124
## 6     157
##   Length Category dep_Length Dependency.Length Type
## 1      6     nmod -0.8087416        -0.7003707    0
## 2      7     nmod -0.1268124        -0.6255117    0
## 3     16     nmod -0.8087416        -0.5399585    0
## 4     16     nmod  1.2370460        -0.5399585    0
## 5      9     nmod  0.5551168         0.3155728    0
## 6     12     nmod -0.8087416         1.0213862    0
```



3. Fitting generalized lin- ear mode

```
m_com <- glm(Type~dep_Length*Dependency.Length,
             data=Combined_data,
             family = "binomial")
summary(m_com)
```

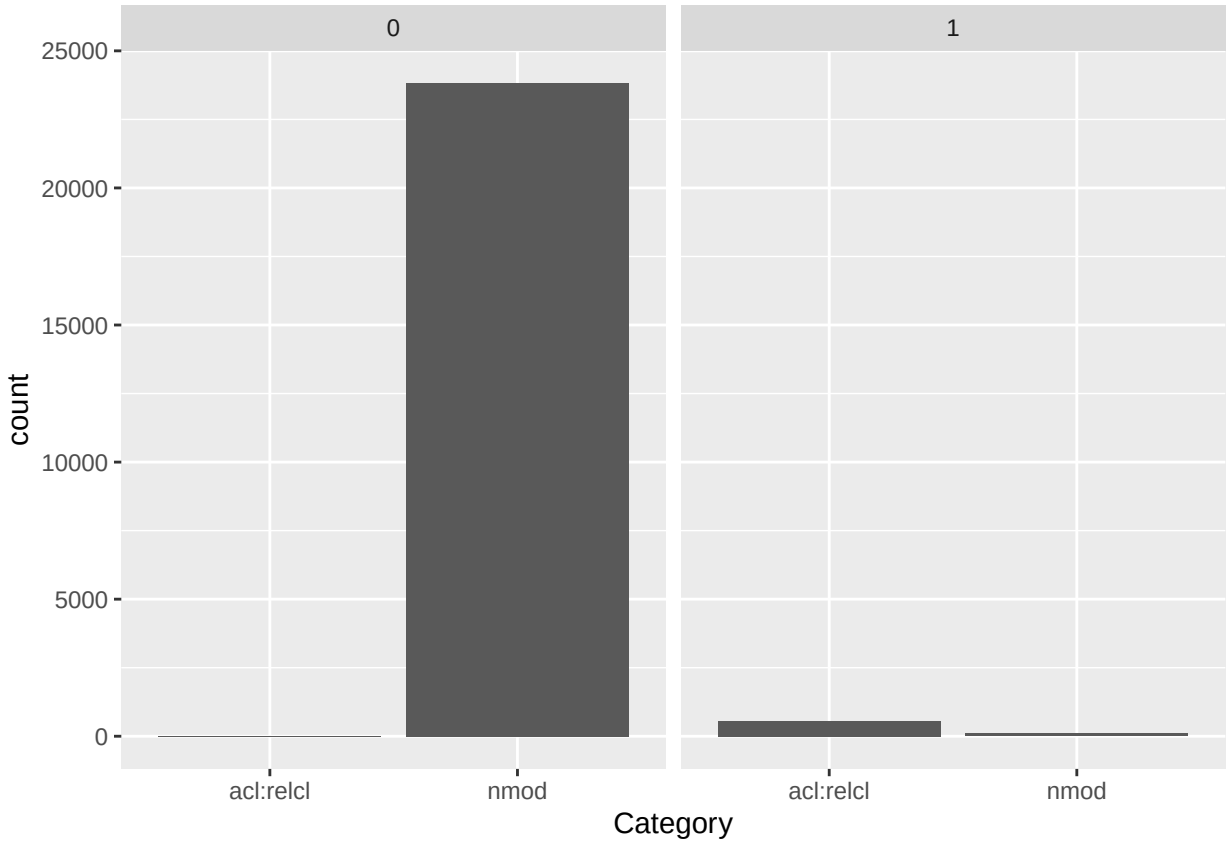
```
##
## Call:
## glm(formula = Type ~ dep_Length * Dependency.Length, family = "binomial",
##     data = Combined_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8203  -0.4422  -0.4230  -0.3667   2.6763
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -2.27837    0.08108 -28.099  < 2e-16 ***
## dep_Length                    0.50721    0.06728   7.539 4.74e-14 ***
## Dependency.Length            -0.03435    0.08288  -0.414   0.6786
## dep_Length:Dependency.Length  0.16061    0.06732   2.386   0.0171 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1276.4  on 1908  degrees of freedom
## Residual deviance: 1184.9  on 1905  degrees of freedom
## AIC: 1192.9
##
## Number of Fisher Scoring iterations: 5
```

Written Text Data (HDTB)

1. Processing the Data

```
##                                                    File Sent_ID
## 1 ../UD_Hindi-HDTB/Phase1\\gold\\hi_hdtb-ud-dev.conllu  dev-s1
## 2 ../UD_Hindi-HDTB/Phase1\\gold\\hi_hdtb-ud-dev.conllu  dev-s1
## 3 ../UD_Hindi-HDTB/Phase1\\gold\\hi_hdtb-ud-dev.conllu  dev-s1
## 4 ../UD_Hindi-HDTB/Phase1\\gold\\hi_hdtb-ud-dev.conllu  dev-s1
## 5 ../UD_Hindi-HDTB/Phase1\\gold\\hi_hdtb-ud-dev.conllu  dev-s1
## 6 ../UD_Hindi-HDTB/Phase1\\gold\\hi_hdtb-ud-dev.conllu  dev-s2
##                                                                                      Sentence
## 1                                  483
## 2                                  483
## 3                                  483
## 4                                  483
## 5                                  483
## 6
##   Length Category dep_Length Dependency.Length Type
## 1     24     nmod -0.1899218         0.2595706    0
## 2     24     nmod  0.1104248         0.2595706    0
## 3     24     nmod  0.7111179         0.2595706    0
## 4     24     nmod  1.0114645         0.2595706    0
## 5     24     nmod -0.7906150         0.2595706    0
## 6     15     nmod -0.4902684        -0.8028540    0
```



3. Fitting generalized linear mode

```
m_com_txt <- glm(Type~dep_Length*Dependency.Length,
             data=Combined_data_txt,
             family = "binomial")
summary(m_com_txt)
```

```
##
## Call:
## glm(formula = Type ~ dep_Length * Dependency.Length, family = "binomial",
##     data = Combined_data_txt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1795  -0.1745  -0.1332  -0.1284   3.1781
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -4.29007    0.05787 -74.136  < 2e-16 ***
## dep_Length                      0.97522    0.02626  37.139  < 2e-16 ***
## Dependency.Length              -0.02918    0.05605  -0.521    0.603
## dep_Length:Dependency.Length -0.10705    0.01886  -5.677 1.37e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5956.6  on 24481  degrees of freedom
## Residual deviance: 4425.3  on 24478  degrees of freedom
## AIC: 4433.3
##
## Number of Fisher Scoring iterations: 7
```