

① Out of Vocabulary o OOV para cuando una palabra del test set no se encuentra en el vocabulario del modelo. Si una palabra no se encuentra en el vocabulario, cualquier n-grama que la incluya tendrá una frecuencia de aparición de cero, esto resulta en una probabilidad de cero para esa frecuencia.

Para modelar la probabilidad de palabras OOV existen diversas estrategias como suavizado de laplace, modelado explícito de OOV, interpolación de N-gramas, modelos basados en redes neuronales, estas son diferentes soluciones para el problema