

# Tarea 6: Embeddings and CNNs

Y. Sarahi Gracia González



## Word2Vec

- Explique la estrategia de selección de palabras dentro de la ventana de contexto en w2v, ¿Por qué se hace así y cuál es la intuición?
- ¿Qué estrategia se usa para construir frases de palabras y construir un solo vector para conceptos basados en más de un token? Para aprender la representación de un sólo vector para una frase se siguen varios pasos, primero:
  1. Se encuentran las palabras que frecuentemente aparecen juntas como una frase.
  2. Se define un score 1 y un cierto umbral, de modo que todos los bigramas por encima de ese umbral se considerarán un único token.

$$score(w_i, w_j) = \frac{count(w_i, w_j) - \delta}{count(w_i) * count(w_j)} \quad (1)$$

3. Estas frases se reemplazan por un único token (en lugar los  $n$  que conformn la frase) en el dataset de entrenamiento.

Notemos que sólo las palabras que aparecen juntas en uno (o unos pocos) cotexto específico serán tomadas en cuenta como un token, de modo que *Estado de México* sería uno de éstos ejemplos, pero *esto es* no lo sería. Pues las pabras *esto* y *es* podrían aparecer juntas en cualquier texto del español, mientras que, a menos que estemos hablando del estado específico, *estado, de* y *México* no co-ocurren.

- Según el autor de w2v ¡cuáles podrían ser las ventajas/desventajas de CBOW y Skipgram?
- ¿Cuáles son las diferencias entre usar Hierarchical Softmax, Negative Sampling y NCE? ¿Cuál recomienda el autor y por qué?
- ¿Cuál diría usted que es la principal conclusión y aportación del paper de w2v? ¿Qué crítica hría usted a estos papers de w2v?

## Glove

## Otros Papers