

Estimación eficiente Word2Vec

¿Qué es?

Consiste en la conversión de un corpus o palabras a su representación vectorial, con el fin de dar una representación gráfica a las palabras y la relación con sus aledaños.

La estimación del lugar que debería ocupar una palabra en el espacio vectorial dado su contexto o su sintaxis es otro tema que se aborda, esto se implementa por medio de varios algoritmos que utilizan redes neuronales;

¿Cómo funciona?

La idea principal de ambas arquitecturas (CBOW y Skip-gram) es usar un modelo de redes neuronales simple para aprender sobre las representaciones vectoriales de las palabras y después se utiliza N-gram NNLM con las representaciones de la anterior red.

¿Para qué sirven?

En general lo que se busca es poder representar las palabras no solo con su valor, si no con las relaciones que pueden tener entre otras palabras, ya sean semánticas o sintácticas; esto con el fin de poder traer equivalencias respecto a relaciones o simplemente generar o buscar palabras aritméticamente.

Aplicaciones

En el caso de CBOW, al utilizar palabras pre o post a la palabra que se busca predecir, pienso que se podría utilizar en reescritura de documentos o servir de soporte en trabajos de decodificación parcialmente avanzados. Además de que en general poder buscar símil de relaciones podría servir como apoyo en escrituras literarias o educativas.