

Embeddings

↗ Parent item IA

Los embeddings son representaciones vectoriales de elementos, típicamente palabras, en un espacio continuo de múltiples dimensiones. Son una técnica utilizada en procesamiento de lenguaje natural (NLP) y aprendizaje automático para convertir palabras o frases en vectores de números, de manera que estos puedan ser procesados por algoritmos de aprendizaje automático.

Características Principales de los Embeddings

1. **Dimensionalidad Reducida:** En lugar de representar palabras como vectores muy largos y dispersos (como en el modelo bag-of-words), los embeddings las representan en un espacio de menor dimensión.
2. **Captura de Semántica:** Los embeddings son capaces de capturar relaciones semánticas entre palabras. Por ejemplo, las palabras "rey" y "reina" estarán más cerca en el espacio vectorial que "rey" y "gato".
3. **Aprendizaje a Partir de Datos:** Los embeddings se entrenan usando grandes corpus de texto. Modelos populares como Word2Vec, GloVe y fastText son ejemplos de técnicas para generar embeddings.
4. **Usos en Tareas de NLP:** Los embeddings se utilizan en una amplia variedad de tareas de NLP, como análisis de sentimientos, traducción automática, y generación de texto.

Ejemplos de Embeddings

1. **Word2Vec:** Introducido por Google, aprende representaciones vectoriales de palabras usando redes neuronales.
2. **GloVe:** Introducido por Stanford, crea embeddings basados en el conteo de palabras en un corpus.
3. **fastText:** Desarrollado por Facebook, similar a Word2Vec pero tiene en cuenta sub-palabras, lo que mejora la representación de palabras raras o desconocidas.

Visualización de Embeddings

Los embeddings se pueden visualizar utilizando técnicas de reducción de dimensionalidad, como t-SNE o PCA, para proyectar los vectores en un espacio de 2 o 3 dimensiones. Esto permite observar cómo las palabras relacionadas están agrupadas entre sí.

En resumen, los embeddings son una herramienta fundamental en NLP para representar y procesar texto de manera eficiente y efectiva.