CASE STUDY: Word Embedding Techniques

Date: 01/01/2025

Completed By:
Nishanth Devabathini

CONTENTS:

1.) What is Word Embedding?

- Word Embedding can be defined as a way to categorize similar words in the form of a numerical vector, so that the computer can easily process it.

- In technical terms, we can say that word embedding is a numeric representation of words in a lower-dimensional space, capturing semantic and syntactic information.

- They play a vital role in natural language processing (nlp) , and are used as an input to the machine learning models.

- There are multiple ways to do word embedding but, 2 of the important ones are as follows:

2.) Word2Vec Model

- It is a word embedding model , which aims to capture the semantic relationships between words by mapping them to high-dimensional vectors.

- It achieves this by analysing the words nearby to the target word.

- It works on the idea that similar words should have a similar vector representation.

- In simple terms, we can say that it works on the principle that “words that appear in similar contexts have similar meanings”.

- This model has 2 approaches:

I. Continuous Bag of Words (CBOW):

- With this approach the main objective of the model is to predict the target word based on its context.
- It achieves this by analysing the surrounding words.

II. Skip-Gram:

- With this approach the main objective of the model is to predict the context (i.e, the surrounding words), based on the target word.

3.) GloVe Model

- It is a word embedding model , which aims to capture the semantic relationships between words by using the concept of co-occurrence.
- In this , the model iterates through the entire dataset and then checks the co-occurrence of words , through which a co-occurrence matrix is created and a value is assigned as an element in the matrix based on the probability of co-occurrence of the 2 words.

4.) References

- Chatgpt
- https://www.geeksforgeeks.org/word-embeddings-in-nlp/