Word Embeddings

Agenda

- Key Questions
- What are word embeddings? How are they useful?
- What is Word2vec? How do CBOW and Skip-gram work?
- What is Negative Sampling? How does it work better than CBOW and Skip-gram?
- What is GloVe? How does it work?
- What are the different applications of word embeddings?

Key Questions

Great Learning
POWER AHEAD

What are word embeddings? How are they useful?

What is Word2vec? How do CBOW and Skip-gram work?

What is Negative Sampling? How does it work better than CBOW and Skip-gram?

What is GloVe? How does it work?

What are the different applications of word embeddings?

What are word embeddings? How are they useful?

Type of representation for words in a vector space, where **words with similar meanings** are represented by **vectors that are close** to each other.

These **vectors** are typically **high-dimensional**, **continuous-valued representations** that **capture semantic relationships** between words based on the context

Semantic Similarity

Word embeddings group similar meanings together in a vector space, showing semantic relationships by proximity.

Contextual Information

Word embeddings capture word meanings based on their context in a given text, reflecting nuances in usage.

Dimensionality Reduction

Word embeddings provide a more compact representation of words, reducing computational complexity.

What is Word2vec? How do CBOW and Skip-gram work?

A **two-layer neural network-based method for efficiently creating word embeddings**

It was **developed in 2013 by Tomas Mikolov et al. at Google** and has since become **the de facto standard for building pre-trained word embeddings**

Word2vec **takes a text corpus as input** and **returns a set of vectors known as feature vectors** that represent the words in that corpus and there are two types of Word2vec

Continuous Bag of Words (CBOW)

uses context to predict the target word

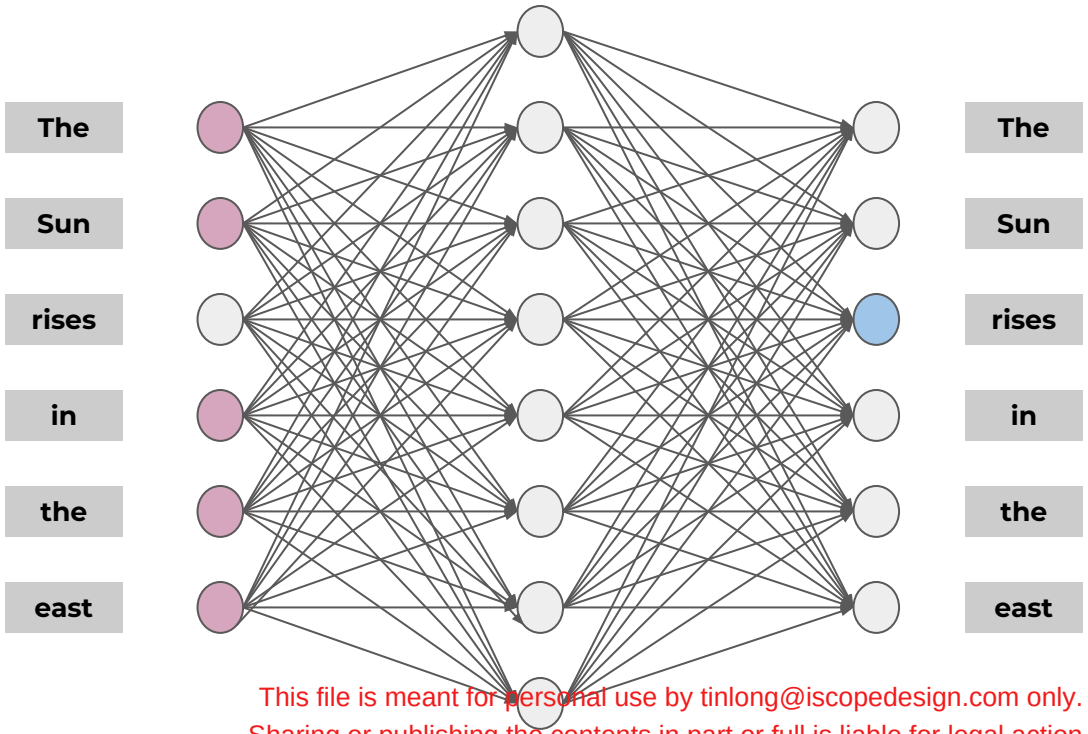
Skip-gram

uses a word to predict its context

Continuous Bag of Words (CBOW)

The	Sun	rises	in	the	east

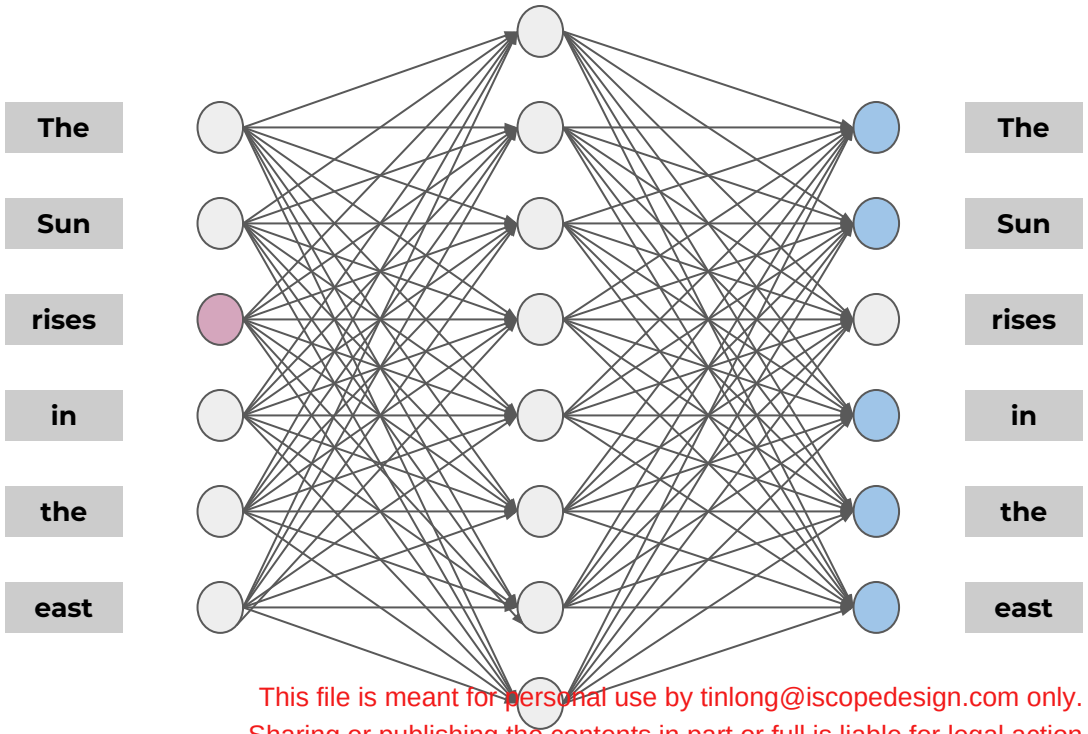
- Here the target word is *rises*
- Rest all are context words



Skip-gram

The	Sun	rises	in	the	east

- Here the target word is *rises*
- Rest all are context words



What is Negative Sampling?

Instead of predicting the actual context words (like Skip-gram), Negative Sampling transforms the task into a binary classification problem.

It randomly samples a small set of negative (non-context) words for each training instance, and the model is trained to distinguish true context words from these negatives.

Skip-gram

Change Task from Word as input → Neural Network Task – Predict Neighbour Word → Context as output

Negative Sampling

to Word 1 as input, Word 2 as input → Neural Network Task – Are the 2 words neighbors or not ? → 0.9 **Binary Classification**

How does it work better than CBOW and Skip-gram?

Great Learning
POWER AHEAD

Computationally more efficient than traditional Skip-gram and CBOW because it **reduces the number of computations** required during training.

Converges faster during training, leading to **quicker model training** compared to the more complex tasks in CBOW and traditional Skip-gram.

More scalable for large vocabularies as it doesn't require computing probabilities for all words in the vocabulary, making it suitable for handling extensive and diverse datasets

What is GloVe? How does it work?

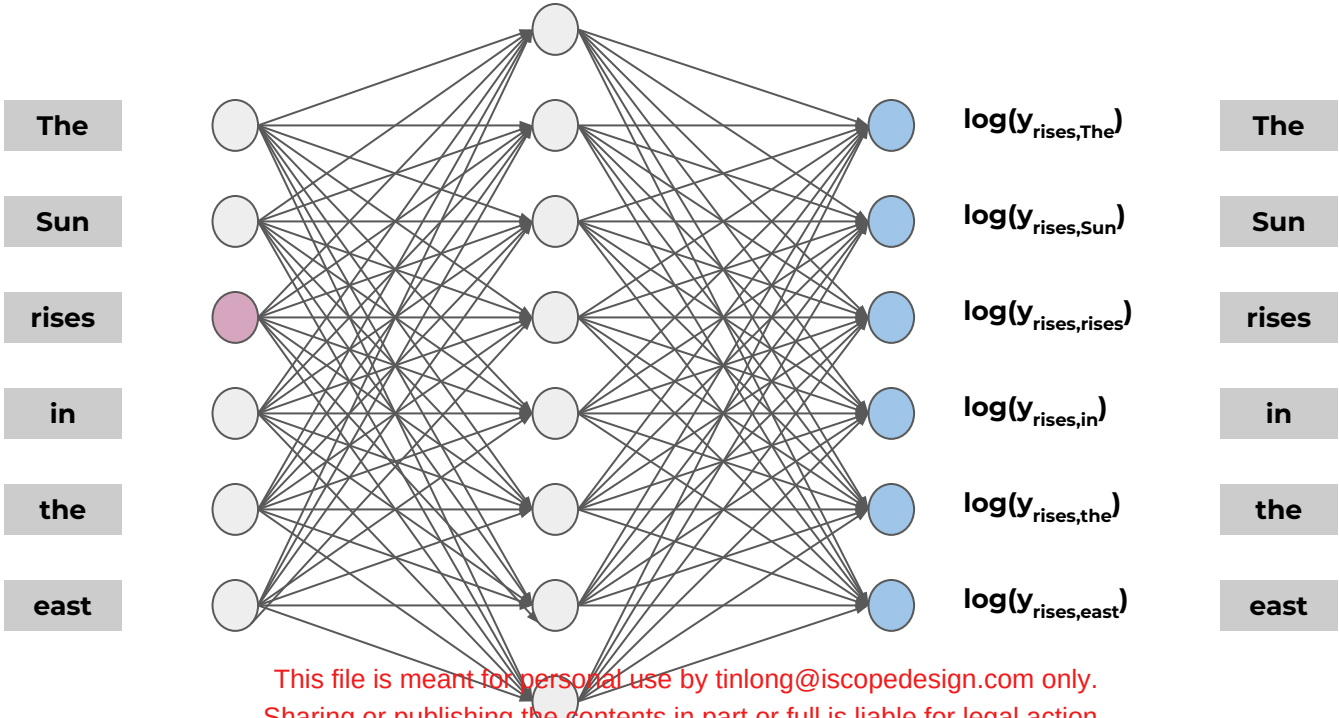
GloVe, which stands for **Global Vectors** is a word embedding model designed to capture global word-word co-occurrence statistics from a large corpus of text.

GloVe starts by constructing a word co-occurrence matrix from a large corpus of text. The matrix reflects how often words co-occur in the same context

	The	sun	rises	in	the	east
The	0	1	0	0	0	0
sun	1	0	1	0	1	0
rises	0	1	0	1	0	0
in	0	0	1	0	1	0
the	0	1	0	1	0	1
east	0	0	0	0	1	0

What is GloVe? How does it work?

From the co-occurrence matrix, GloVe calculates the probability distribution of word pairs co-occurring. This represents the likelihood of one word appearing near another.



What are the different applications of word embeddings?

Great Learning
POWER AHEAD

Text Classification

Word embeddings enhance the performance of text classification tasks, including sentiment analysis, topic classification, and spam detection, by providing more meaningful and context-aware representations of words.

Semantic Search

Semantic search involves understanding the meaning and context of search queries and documents, enabling a search engine to return results that are not just keyword-matched but also semantically relevant.

Named Entity Recognition

Named Entity Recognition systems use word embeddings to recognize and classify entities (such as names of people, organizations, and locations) in text by leveraging the semantic information encoded in the embeddings.

Great Learning

Happy Learning !