Word Embeddings

Agenda

- Intro to NLP and Word Embeddings Quiz
- NLP and its applications
- Introduction to Word Embeddings
- Working of Word2Vec
- Working of GloVe
- Applications of Word Embeddings

Let’s begin the discussion by answering a few questions on introduction to NLP and word embeddings

Intro to NLP Quiz

What is the primary goal of Natural Language Processing (NLP)?

- A. To translate human languages into computer programming languages
- B. To enable computers to understand, interpret, and generate human language
- C. To convert speech into text documents
- D. To create artificial languages for communication between computers

Intro to NLP Quiz

What is the primary goal of Natural Language Processing (NLP)?

A. To translate human languages into computer programming languages

B. To enable computers to understand, interpret, and generate human language

C. To convert speech into text documents

D. To create artificial languages for communication between computers

Natural Language Processing

Branch of artificial intelligence (AI) that deals with the interaction between machines and human languages

Aims to **automate the reading, interpretation, and understanding of human language**, also called natural language

Processing → Understanding → Generation

Intro to NLP Quiz

Which of the following are the applications of Natural Language Processing?

- A — Sentiment Analysis
- B — Machine Translation
- C — Chatbot
- D — Document Summarization

Intro to NLP Quiz

Which of the following are the applications of Natural Language Processing?

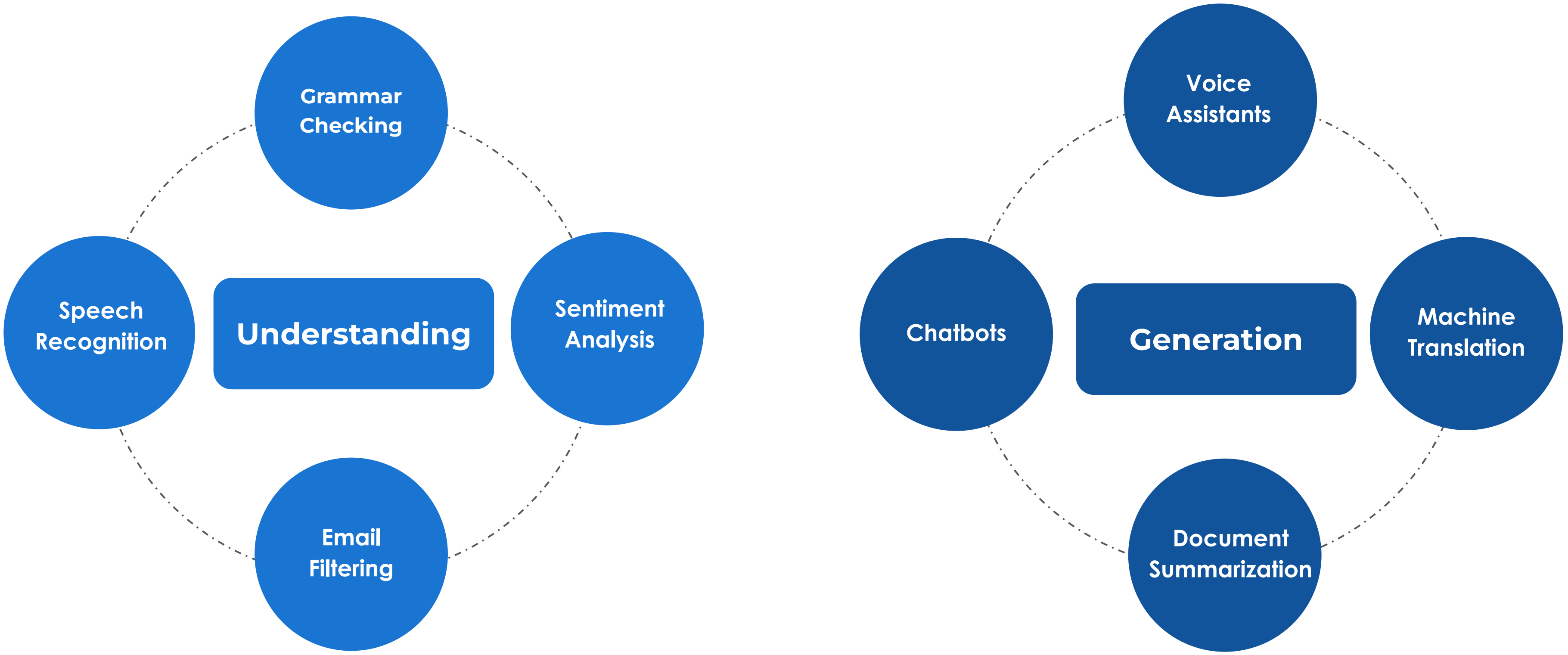
A Sentiment Analysis

B Machine Translation

C Chatbot

D Document Summarization

Applications of NLP



Word Embeddings Quiz

Which one of the following is a key advantage of using word embeddings?

A. Word embeddings preserve semantic relationships between words

B. Word embeddings are easier to implement

C. Word embeddings work better with punctuation

Word Embeddings Quiz

Which one of the following is a key advantage of using word embeddings?

A Word embeddings preserve semantic relationships between words

B Word embeddings are easier to implement

C Word embeddings work better with punctuation

Word Embeddings

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.

Word Embeddings Quiz

How is the task of learning word embeddings addressed by the CBOW model?

A. By predicting a target word given its surrounding context words

B. By predicting surrounding context words given a target word

C. By generating word embeddings based on co-occurrence matrices

D. By using different normal distribution samples as dense vectors for each word in the vocabulary

Word Embeddings Quiz

How is the task of learning word embeddings addressed by the CBOW model?

A. By predicting a target word given its surrounding context words

B. By predicting surrounding context words given a target word

C. By generating word embeddings based on co-occurrence matrices

D. By using different normal distribution samples as dense vectors for each word in the vocabulary

Word2Vec

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

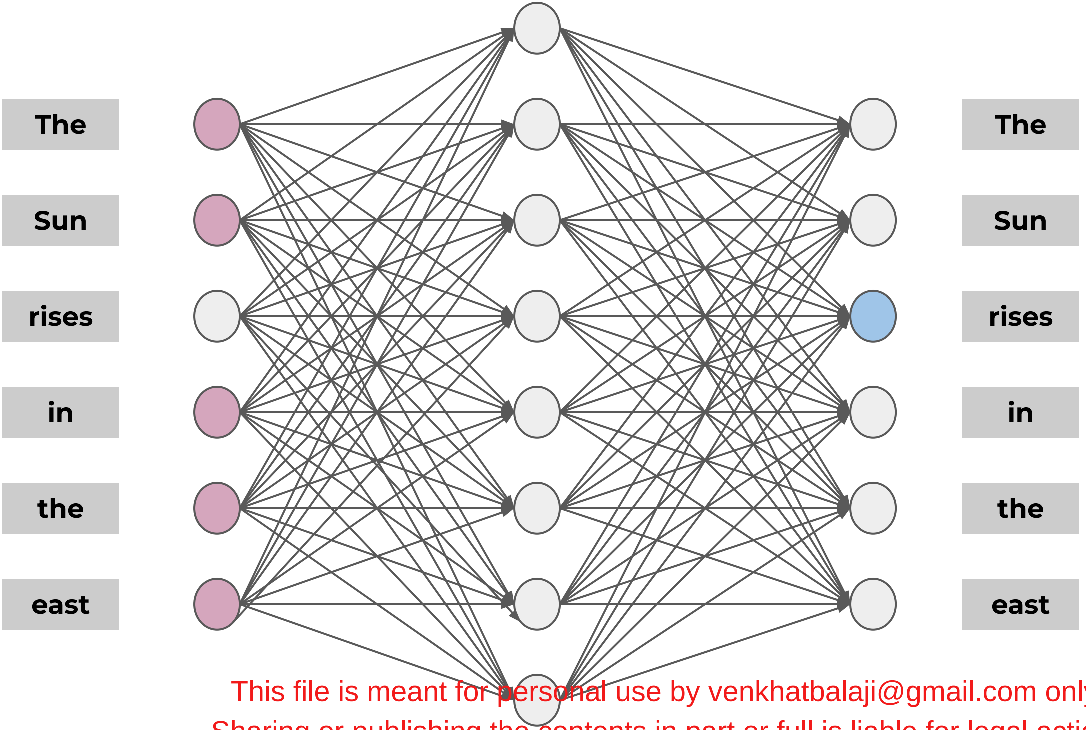
Was developed in 2013 by Tomas Mikolov et al. at Google - and has since become the de facto standard for building pre-trained word embeddings

Takes a text corpus as input and returns a set of vectors known as feature vectors that represent the words in that corpus

Continuous Bag of Words (CBOW) model

The	Sun	rises	in	the	east

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



Word Embeddings Quiz

What is the purpose of the context window size?

A. To control the dimensionality of the word embeddings

B. To specify the number of neurons in the hidden layer of the neural network

C. To specify the number of layers in the neural network

D. To define the range of neighboring words considered for prediction

Word Embeddings Quiz

What is the purpose of the context window size?

A. To control the dimensionality of the word embeddings

B. To specify the number of neurons in the hidden layer of the neural network

C. To specify the number of layers in the neural network

D. To define the range of neighboring words considered for prediction

Window Size for Word Embedding Generation

Determines the range of neighboring words considered for prediction in word embedding models like CBOW (Continuous Bag of Words)

Larger context window => includes more surrounding words in the context => potentially captures more semantic information, increases computational complexity.

Example:

In CBOW, the context window specifies how many words before and after the target word are taken into account as context for predicting the target word.

Word Embeddings Quiz

Consider the below sentence:

'The Sun rises in the east.'

The word 'rises' is passed as an input word to a neural network, which predicts the context words for the input word as the output.

Which of the following models works using this mechanism?

A CBOW

B Skip-gram

C Negative Sampling

D n-gram

Word Embeddings Quiz

Consider the below sentence:

'The Sun rises in the east.'

The word 'rises' is passed as an input word to a neural network, which predicts the context words for the input word as the output.

Which of the following models works using this mechanism?

A CBOW

B Skip-gram

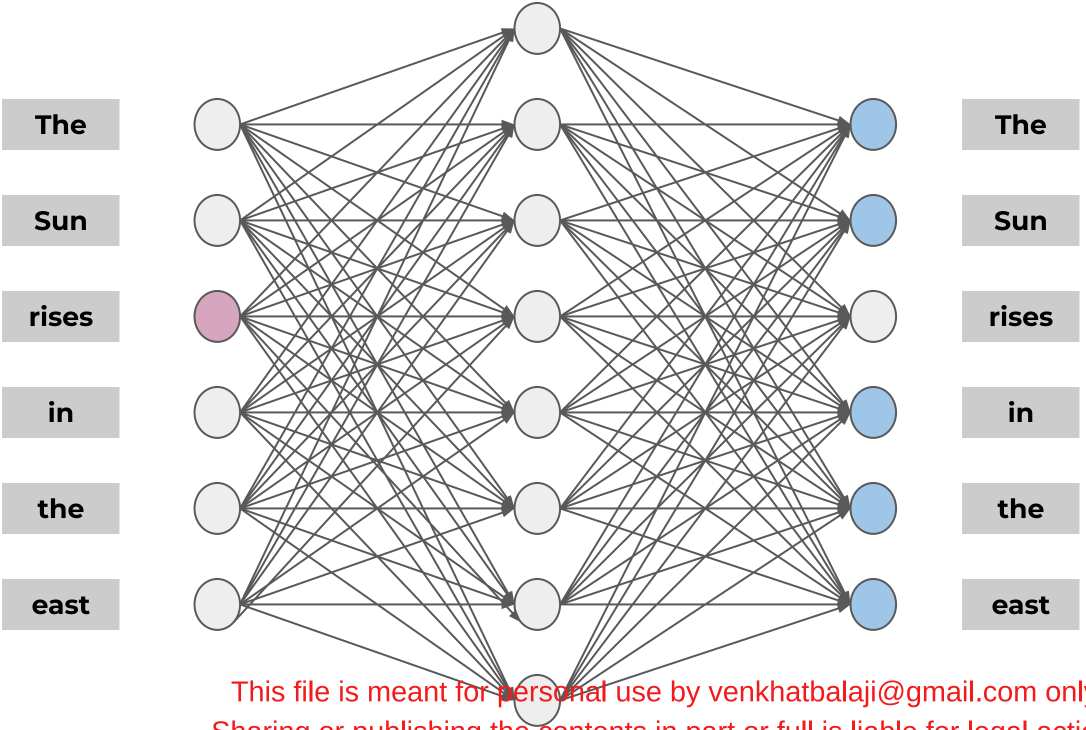
C Negative Sampling

D n-gram

Skip-gram model

The	Sun	rises	in	the	east

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



Word Embeddings Quiz

What is the primary objective of negative sampling in the Word2Vec model?

A — To increase the size of the context window for word prediction

B — To improve the accuracy of word embeddings by increasing the training dataset

C — To speed up the training process by reducing the computational overhead of training word embeddings

D — To penalize the model for incorrectly predicting context words during training

Word Embeddings Quiz

What is the primary objective of negative sampling in the Word2Vec model?

A. To increase the size of the context window for word prediction

B. To improve the accuracy of word embeddings by increasing the training dataset

C. To speed up the training process by reducing the computational overhead of training word embeddings

D. To penalize the model for incorrectly predicting context words during training

Negative Sampling

A technique used in Word2Vec to speed up the training process by reducing the computational overhead

Instead of considering all possible words during training, negative sampling randomly selects a small subset of words as negative samples for each target word.

Makes the training process more efficient while still providing meaningful updates to the word embeddings

Negative Sampling

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

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

Word Embeddings Quiz

How does the GloVe model differ from word2vec?

- **A** GloVe uses a neural network architecture, while word2vec uses a co-occurrence matrix
- **B** GloVe incorporates negative sampling to train word embeddings, while word2vec uses skip-gram
- **C** GloVe optimizes word embeddings to capture global word co-occurrence frequencies, while word2vec focuses on local context

Word Embeddings Quiz

How does the GloVe model differ from word2vec?

- **A** GloVe uses a neural network architecture, while word2vec uses a co-occurrence matrix
- **B** GloVe incorporates negative sampling to train word embeddings, while word2vec uses skip-gram
- **C** GloVe optimizes word embeddings to capture global word co-occurrence frequencies, while word2vec focuses on local context

GloVe model

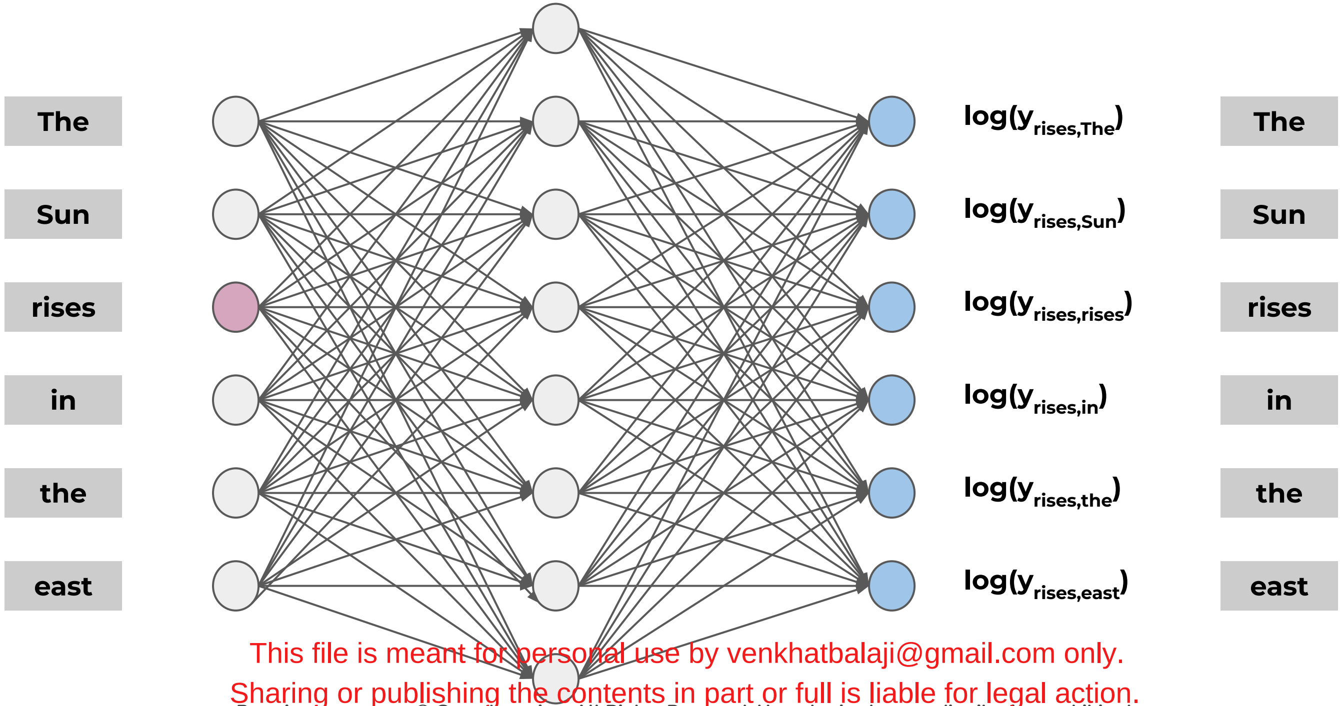
Stands for Global Vectors is a word embedding model designed to capture global and local word-word co-occurrence statistics from a large corpus of text.

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

GloVe model

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.



Word Embeddings Quiz

How can word embeddings contribute to improving the performance of text classification models?

A. By directly serving as labels for classifying text documents

B. By capturing semantic relationships between words and enhancing feature representations

C. By automatically generating textual descriptions of classified categories

Word Embeddings Quiz

How can word embeddings contribute to improving the performance of text classification models?

A. By directly serving as labels for classifying text documents

B. By capturing semantic relationships between words and enhancing feature representations

C. By automatically generating textual descriptions of classified categories

Applications of Word Embeddings

In text classification tasks, word embeddings are utilized to represent words as dense vectors in a continuous semantic space

By capturing semantic relationships between words, word embeddings provide richer feature representations compared to traditional one-hot encodings or bag-of-words representations

This enhanced representation can help text classification models better understand the underlying semantics of the text and improve classification accuracy

Applications of Word Embeddings

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.

Power Ahead!