Part-of-Speech Tagging with Word Embeddings

CS 9875 Final Project

Paul Moore Juhani Dickinson

December 2, 2025

Outline

Section 1 Longname

Slide Title

Slide text

Slide Title 2

Slide text 2

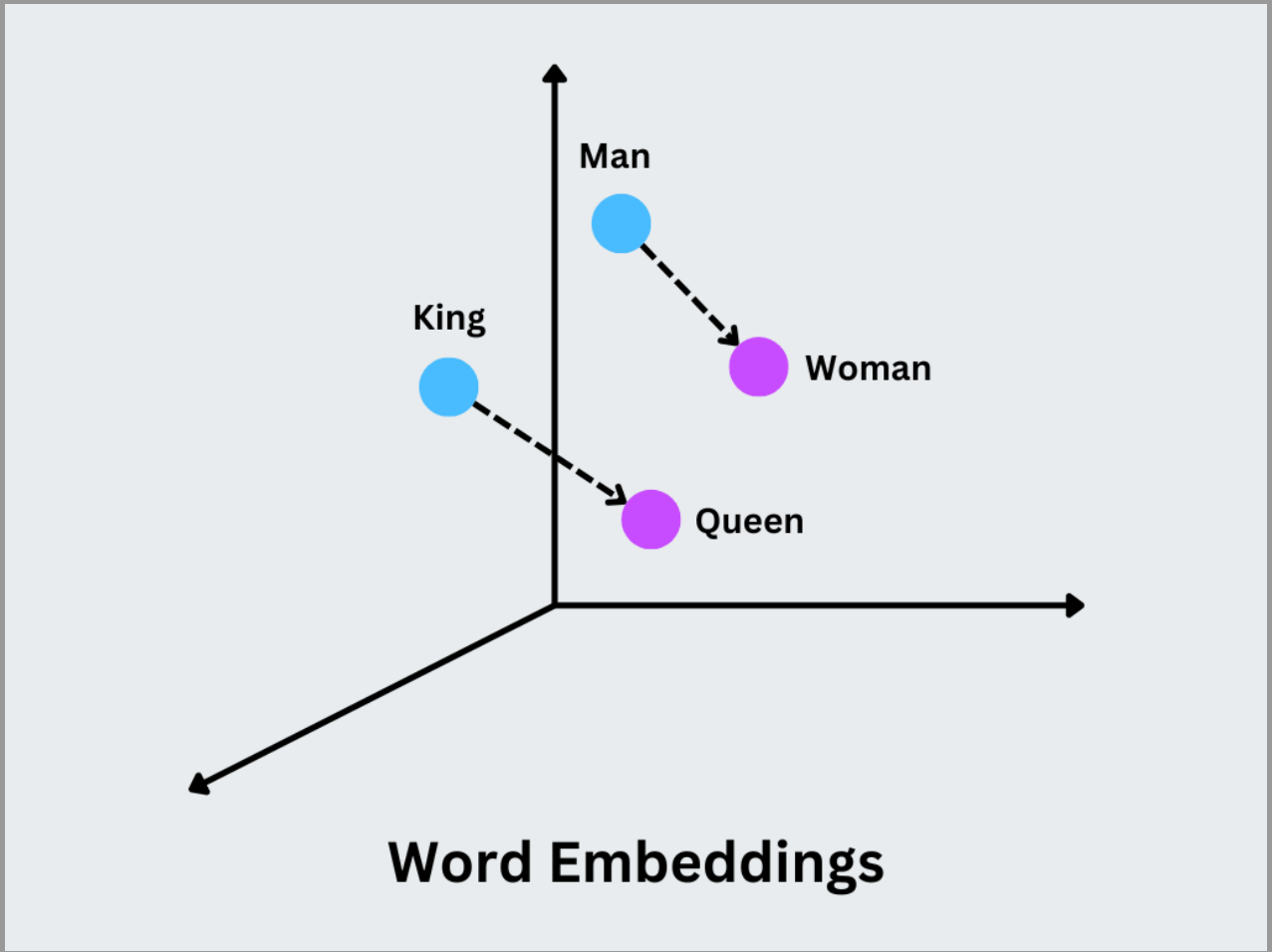
Slide text

Slide Title

Slide text

Word Embedding Models

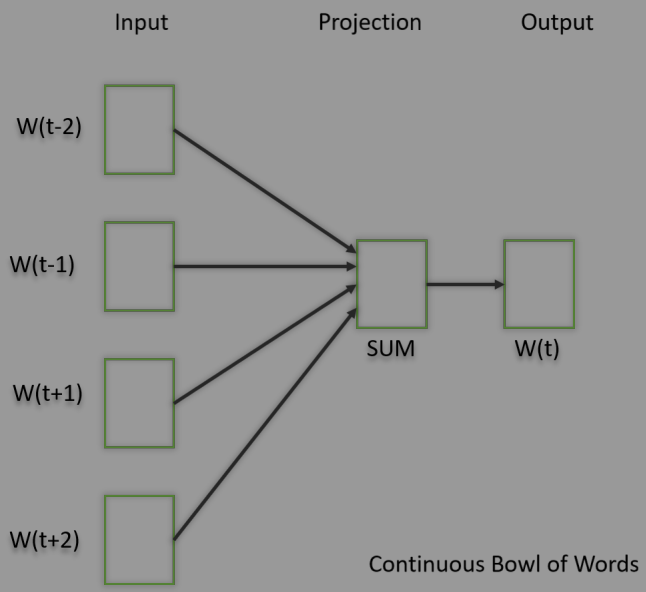
What are word embeddings?



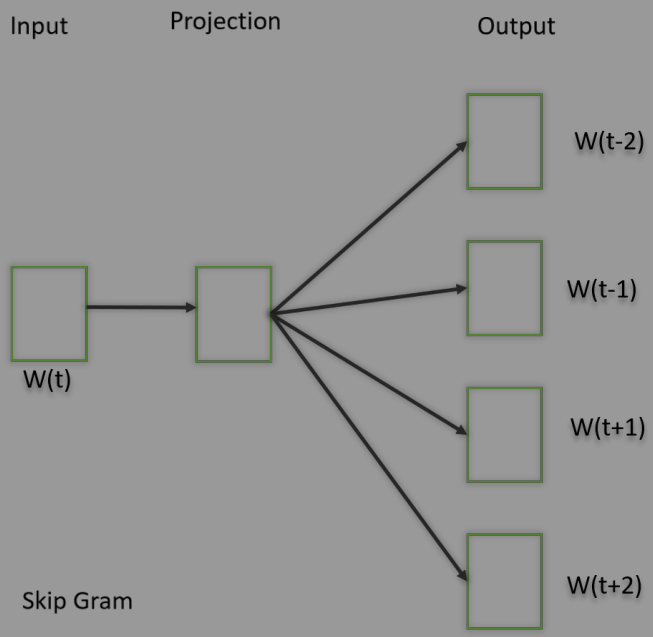
Static Word Embeddings

What are static word embedding models?

Word2Vec



CBOW



Skip Gram

Word2Vec Training Procedures

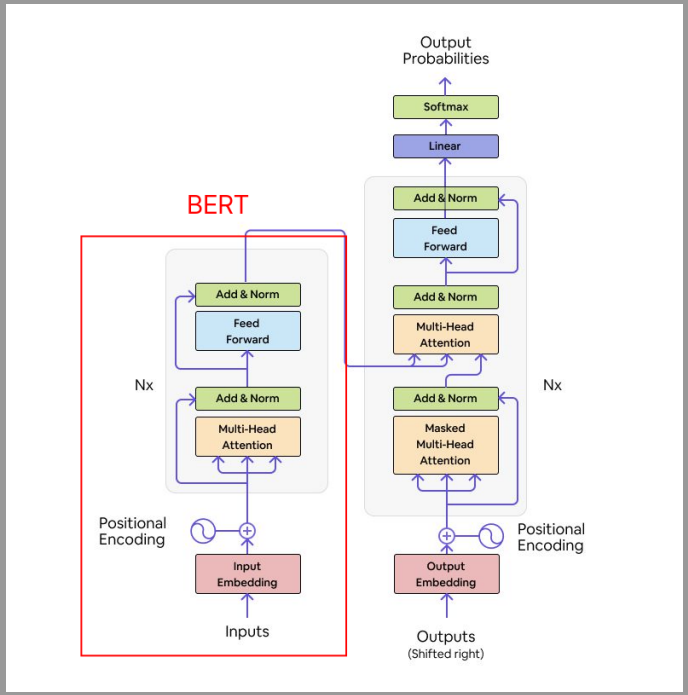
Word2Vec

$$\begin{bmatrix} W_{00} & W_{01} & W_{02} & \dots \\ W_{10} & W_{11} & W_{12} & \dots \\ W_{20} & W_{21} & W_{22} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

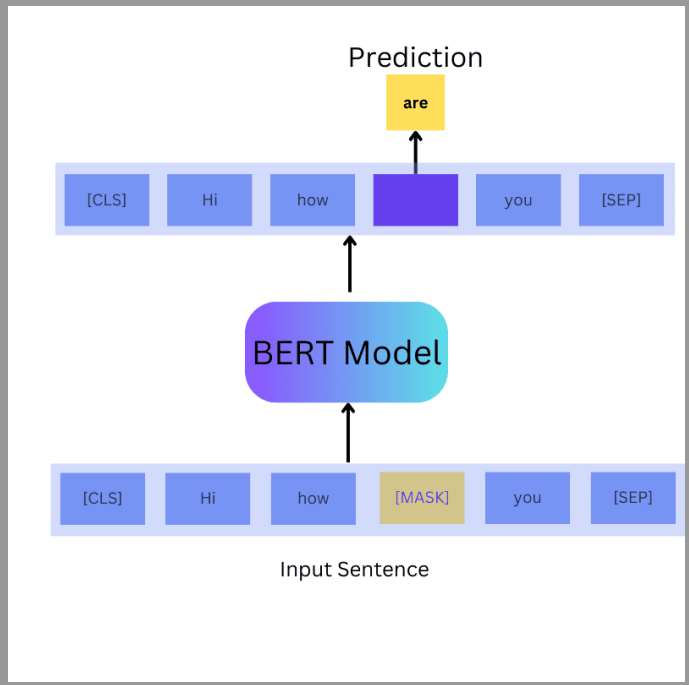
Contextual Word Embeddings

What are contextual word embeddings?

BERT

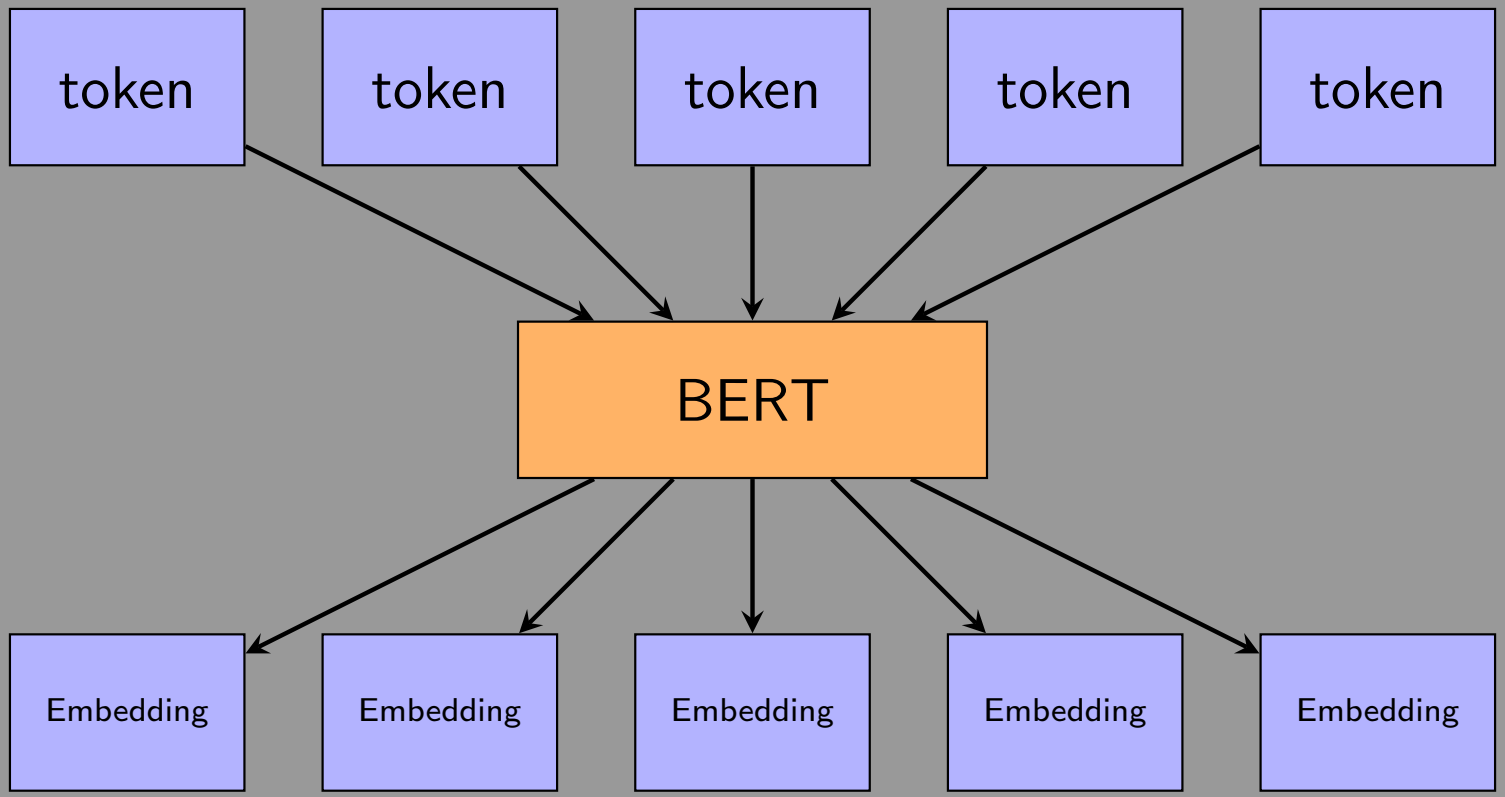


Transformer architecture



MLM

BERT


token
token
token
token
token
BERT
Embedding
Embedding
Embedding
Embedding
Embedding

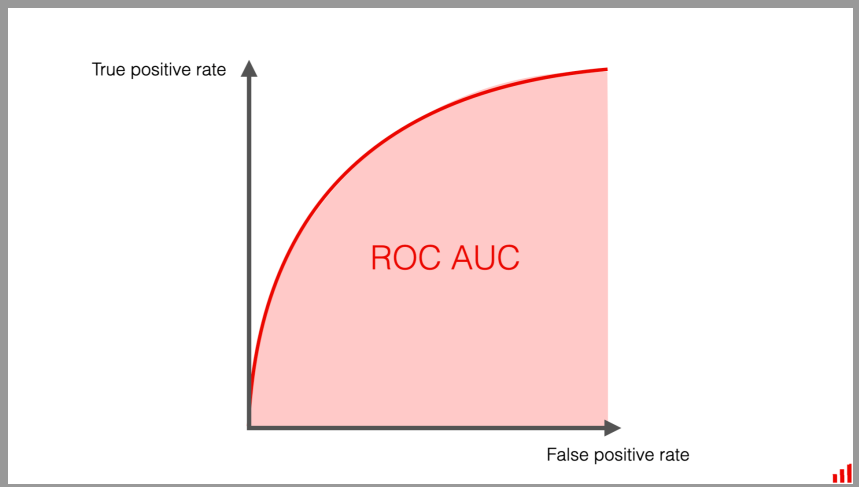
How do we use these embeddings?

How do we use these embeddings?
As input for a downstream model: SVM, Boosting, CNN, etc.

Evaluation and Theoretical Analysis

Evaluation Metrics

We will evaluate each of our models using F1 score and AUC-ROC adapted for this multi-class classification problem.



AUC-ROC

Theoretical Analysis

We will conduct a theoretical analysis on the word embedding models, and the downstream classifier models.

Anticipated Results

We anticipate achieving near perfect scores in both evaluation metrics using contextual embeddings from BERT and a simple classification model like an SVM.