REF: https://taksw222.tistory.com/187

# ELMO(Embeddings from Language Model)

ELMo, which stands for Embeddings from Language Models, is a deep contextualized word representation model. It generates word embeddings **that capture the meaning of words in context by considering the entire input sentence.** Unlike traditional methods like Word2Vec or GloVe, which assign a single vector to each word, ELMo creates dynamic word representations based on the surrounding words. This allows it to better handle polysemy (multiple meanings of a word) and coreference (multiple expressions referring to the same entity) in NLP tasks

### Key Features:

**Contextualized Embeddings:**
ELMo's primary strength is its ability to produce embeddings that vary depending on the context of the sentence.

**Deep Bi-directional LSTM:**
It utilizes a deep bi-directional LSTM (Long Short-Term Memory) network trained on a large language modeling task

**Task-Specific Weights:**
The model learns task-specific weights for the different layers of the LSTM, allowing for adaptation to various NLP tasks

### Key Benefits:

**Improved NLP Performance :**
ELMo has shown significant improvements in various NLP tasks, including sentiment analysis, named entity recognition, and question answering.

**Better Understanding of Language:**
By considering context, ELMo can better capture the nuances of language and resolve ambiguitie

**Adaptability:**
The ability to learn task-specific weights makes ELMo adaptable to different NLP applications.

### Limitations:

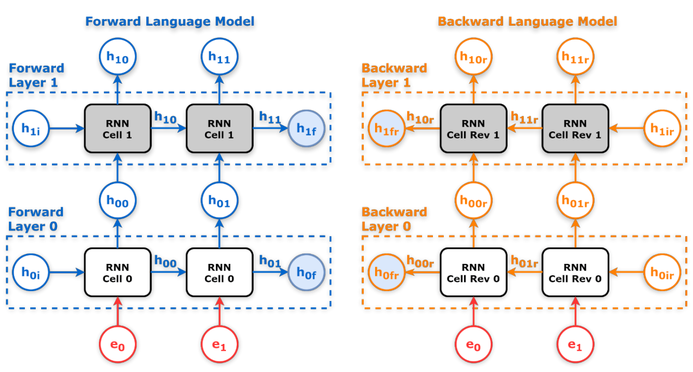
- **Computational Intensity:** Training and using ELMo can be computationally expensive.
- **Potential for Overfitting:** In some cases, ELMo might overfit to the training data, especially if the training corpus is not diverse enough.
- **Limited Interpretability:** While ELMo provides contextualized embeddings, the internal workings of the LSTM can be difficult to interpret.

In essence, ELMo was a significant advancement in NLP, demonstrating the power of contextualized word embeddings and paving the way for more advanced models like those based on the Transformer architecture.

### Use Cases

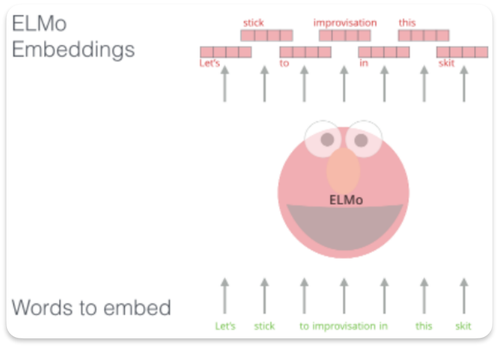
- ELMo has been successfully applied to various NLP tasks, including:
  - **Sentiment Analysis**: Determining the emotional tone of text.
  - **Named Entity Recognition**: Identifying and classifying named entities (e.g., people, organizations, locations).
  - **Machine Translation**: Improving the accuracy of translations.
  - **Question Answering**: Generating answers to questions based on given text.

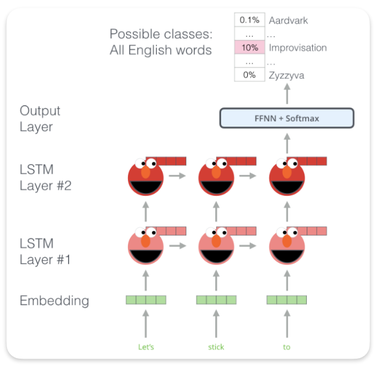
## Architecture




### Process

(1) Connect the embedding layers of the forward and backward models, and two hidden layers.
(2) Multiply weights to the embedding layer, the first hidden layer, and the second hidden layer, respectively.
(3) Add all three vectors.




The structure is designed to produce a final output after passing through two LSTM layers. However, as mentioned earlier, ELMo has a **bidirectional** structure. That is, it also utilizes a backward LSTM