Introduction to State Space Models

UG BTech - 2nd Year

Vinay , Vaibhav Mahore , Shubhadeep Sing , Snehal Biswas

Indian Institute of Science, Bangalore

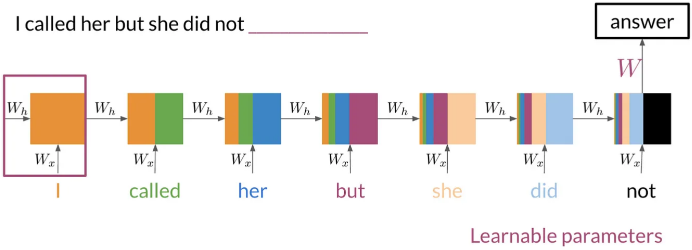
March 18, 2025

The Need for Sequential Data Modeling

- **What is Sequential Data?**
  - **Ans :** It's the data that comes in a specific order, where the arrangement of the data points matters.
- **Example(NLP):** The dog bites the man vs The man bites the dog.
- **Need for Sequential Data Modeling:** It's crucial because many datasets have an inherent order e.g Language, Time series in which the sequence and context between data points are essential for accurate analysis and predictions.
- **Challenge:** Traditional models(e.g. Feedforward Networks) treat inputs independently and fail to capture such temporal dependencies.

Recurrent Neural Networks (RNNs)

- **Why RNNs?**
  - A key characteristic that sets RNN apart from traditional FFN is their **recurrence**, which allows them to maintain an internal state over time .



$$h_t = \sigma(W_{hh} \cdot h_{t-1} + W_{xh} \cdot x_t + b_h) \tag{1}$$

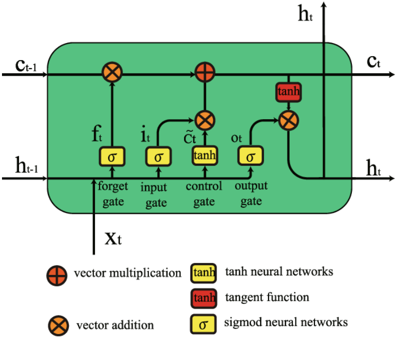
$$o_t = W_{hy} \cdot h_t + b_y \tag{2}$$

RNNs: Structure and Challenges

- **Benefits:**
  - **Temporal Contextualization:** RNNs maintain an internal state that carries information from previous time steps, allowing them to capture the temporal dependencies inherent in sequential data.
  - **Efficient Weight Sharing:** RNNs share weights across different time steps, reducing the number of parameters and improving generalization.
- **Challenges:**
  - **Vanishing and Exploding Gradients:** Gradients may decay over time making early data less influential or become excessively large leading to training instability.
  - **Sequential Processing Bottleneck:** RNNs process data sequentially, making them computationally inefficient, especially for long sequences.

Long Short-Term Memory (LSTM)

- **Introduction of LSTM:**
  - LSTMs overcome standard RNN limitations by using a dedicated cell state to store and retain long-term information.
  - They use three specialized gates—input, forget, and output—to precisely control how data enters and exits the cell state.



$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right),$$

$$i_t = \sigma\left(W_i[h_{t-1}, x_t] + b_i\right),$$

$$\tilde{C}_t = \tanh\left(W_C[h_{t-1}, x_t] + b_C\right),$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t,$$

$$o_t = \sigma\left(W_o[h_{t-1}, x_t] + b_o\right),$$

$$h_t = o_t \times \tanh\left(C_t\right).$$

Limitations of LSTMs

- **Scalability and Computational Complexity:**
  - Despite improvements, LSTMs remain computationally intensive for very long sequences.
- **Long-Range Dependency Challenges:**
  - Research such as Bengio et al. (1994) and Pascanu et al. (2013) has demonstrated that LSTMs can still struggle with vanishing gradients when modeling very long sequences.
- **Sequential Processing Bottleneck:**
  - The inherent sequential nature of LSTM processing limits parallelism, resulting in slower training compared to more modern architectures.

Methodology

- Data collection and preprocessing
- Model architecture
- Training approach
- Evaluation metrics

Implementation

- Technologies used
- Key algorithms
- Technical challenges
- Solutions implemented

Results

- Model performance
- Key findings
- Comparative analysis
- Visualizations

Future Work

- Potential improvements
- Scalability considerations
- Additional features
- Research directions

Thank You

any Questions?