

ML, Assignment #3

Summary Report about MIT online lectures

CAU SW 20184286 Lee Donghwa

2023-06-01

- https://www.youtube.com/watch?v=NmLK_WQBxB4 [CNN]
 - https://www.youtube.com/watch?v=ySEx_Bqxvvo [RNN, Attention, and Transformers]
-

[CNN]

Convolutional Neural Networks (CNNs) are a type of deep learning model commonly used for image processing tasks. CNNs leverage convolution and filters to extract useful information from images. CNNs can detect and extract features. This makes them capable of recognizing and differentiating objects in images. First of all, Computers perceive images as numerical data rather than visual content. Through the use of neural networks, computers learn important patterns and features from images, enabling them to recognize and differentiate objects. To extract useful information from images, convolution and filters are employed for feature extraction. Convolution applies kernels to small regions of the image, detecting and extracting features. Convolutional Neural Networks (CNNs) are deep learning models specifically designed for image processing tasks.

CNNs consist of multiple convolutional layers, allowing them to learn complex features. Over time, they have evolved into advanced architectures such as R-CNN (Region-based CNN) and Fully Convolutional Networks, which further enhance object detection capabilities. Pooling operations are utilized to reduce the size of the image and summarize the extracted features effectively. By decreasing spatial resolution through pooling, important information is preserved while reducing dimensions.

These concepts in computer vision form the foundation for processing and analyzing images. Utilizing various techniques and models in computer vision, such as image

recognition, object detection, segmentation, feature extraction, and image classification, enables the implementation of diverse applications. Ultimately, computers gain the ability to understand the meaning of image data and achieve visual perception at a level comparable to humans.

[RNN, Attention, and Transformers]

Sequence models are designed to process data with a sequential nature, such as music or sentences. They utilize neurons with recurrence, allowing them to maintain information from previous steps while making predictions for the current input. Recurrent Neural Networks (RNNs) are a type of sequence model composed of neurons with a recurrent structure. This structure enables RNNs to capture temporal dependencies and context in the data.

The intuition behind RNNs lies in their ability to process sequential data by maintaining a hidden state that serves as memory. At each step, the hidden state is updated based on the input and the previous hidden state. This allows RNNs to capture and learn patterns in sequential data.

Unfolding refers to the concept of expanding the RNN over multiple time steps to visualize its recurrent nature. It reveals how information flows through the network and how the hidden state evolves over time.

Building an RNN from scratch involves initializing the model with random weights and training it on a dataset using techniques like backpropagation. Backpropagation is an algorithm used to compute the gradients of the model's parameters with respect to the loss function, enabling the optimization of the model through gradient descent.

Sequential modeling refers to the process of using sequence models, such as RNNs, to model and understand sequential data. RNNs are particularly effective in tasks such as language modeling, speech recognition, and machine translation, where the order of the data is crucial.

Gradient issues can arise in RNN training due to the vanishing or exploding gradient problem. The vanishing gradient problem occurs when the gradients diminish as they propagate backward through time, making it challenging for the model to learn long-term dependencies. On the other hand, the exploding gradient problem occurs when the gradients become extremely large, leading to unstable training. Techniques

like gradient clipping, weight initialization strategies, and using specialized RNN variants like LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit) help mitigate these gradient issues and improve the training stability of RNNs.

Long Short-Term Memory (LSTM) is an improved variant of the basic RNN structure designed to address the challenge of capturing long-term dependencies in sequential data. It incorporates a memory cell and gate mechanisms to selectively retain or forget information over time, allowing the model to effectively process and remember relevant information for longer sequences.

RNNs and LSTM have found applications in various domains, including natural language processing, speech recognition, machine translation, time series analysis, and handwriting recognition. Their ability to model sequential data and capture contextual information makes them well-suited for tasks where the order of data is crucial.

Attention is a fundamental concept in machine learning that allows models to focus on relevant parts of the input when making predictions. In the context of sequence modeling, attention mechanisms enable the model to attend to specific elements of the input sequence that are more informative or important for the task at hand. It assigns weights or importance scores to different parts of the input sequence, allowing the model to allocate more attention to relevant elements.

The intuition behind attention is inspired by human perception and the idea of selectively focusing on specific parts of the input. By attending to relevant elements, the model can make more informed predictions or decisions.

There is a strong relationship between attention and search. Attention mechanisms can be seen as a form of internal search within the input sequence, where the model dynamically selects and attends to different parts of the sequence based on their relevance. This enables the model to capture dependencies and relationships across the sequence.

Neural networks can learn attention mechanisms through training. By jointly learning the attention weights and the model parameters, the network can optimize both the attention mechanism and the task-specific objective. This allows the model to effectively use attention to improve performance on various tasks.

Scaling attention refers to extending the application of attention mechanisms beyond sequential data to other domains and tasks. Attention has been successfully applied to tasks such as image recognition, machine reading comprehension, and reinforcement learning. Scaling attention involves adapting and integrating attention

mechanisms into different architectures and models to enhance their ability to focus on relevant information, leading to improved performance and interpretability.