

# 注意力机制调研

SA23011253 任永文 SA23011255 陶祥志

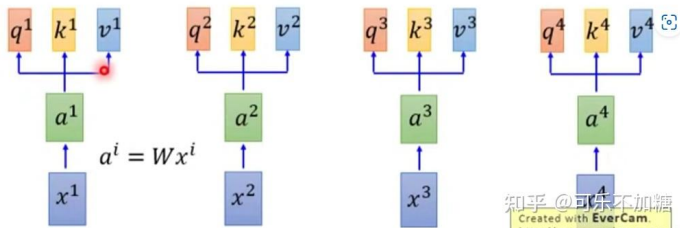
中国科学技术大学计算机学院

2023 年 11 月 14 日

深度学习中的注意力机制（Attention Mechanism）是一种模仿人类视觉和认知系统的方法，它允许神经网络在处理输入数据时集中注意力于相关的部分。通过引入注意力机制，神经网络能够自动地学习并选择性地关注输入中的重要信息，提高模型的性能和泛化能力。注意力机制从本质上讲和人类的选择性注意力机制类似，核心目标也是从众多信息中选出对当前任务目标更加关键的信息。深度学习中，注意力机制通常应用于序列数据（如文本、语音或图像序列）的处理。其中，最典型的注意力机制包括自注意力机制、空间注意力机制和时间注意力机制。这些注意力机制允许模型对输入序列的不同位置分配不同的权重，以便在处理每个序列元素时专注于最相关的部分。

# 自注意力机制

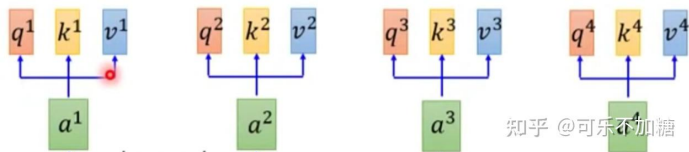
自注意力机制的基本思想是，在处理序列数据时，每个元素都可以与序列中的其他元素建立关联，而不仅仅是依赖于相邻位置的元素。它通过计算元素之间的相对重要性来自适应地捕捉元素之间的长程依赖关系。



具体而言，对于序列中的每个元素，自注意力机制计算其与其他元素之间的相似度，并将这些相似度归一化为注意力权重。然后，通过将每个元素与对应的注意力权重进行加权求和，可以得到自注意力机制的输出。

# 多头自注意力机制

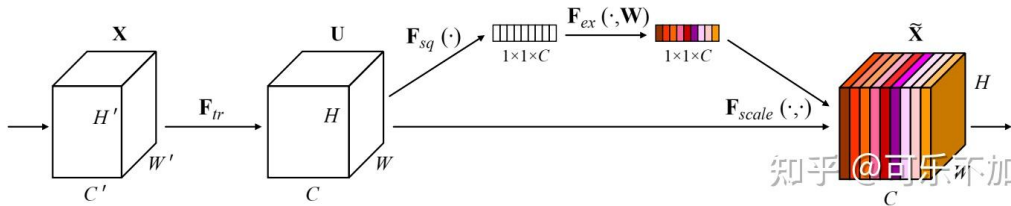
自注意力机制的基本思想是，在处理序列数据时，每个元素都可以与序列中的其他元素建立关联，而不仅仅是依赖于相邻位置的元素。它通过计算元素之间的相对重要性来自适应地捕捉元素之间的长程依赖关系。



具体而言，对于序列中的每个元素，自注意力机制计算其与其他元素之间的相似度，并将这些相似度归一化为注意力权重。然后，通过将每个元素与对应的注意力权重进行加权求和，可以得到自注意力机制的输出。

# 通道自注意力机制

顾名思义，通道注意力机制是通过计算每个通道 channel 的重要性程度；因此，常常被用在卷积神经网络里面。目前，比较经典的通道注意力机制方法就是 SENet 模型，SENet 通过学习通道间的关系（每个通道的重要性），提升了网络在特征表示中的表达能力，进而提升了模型的性能。



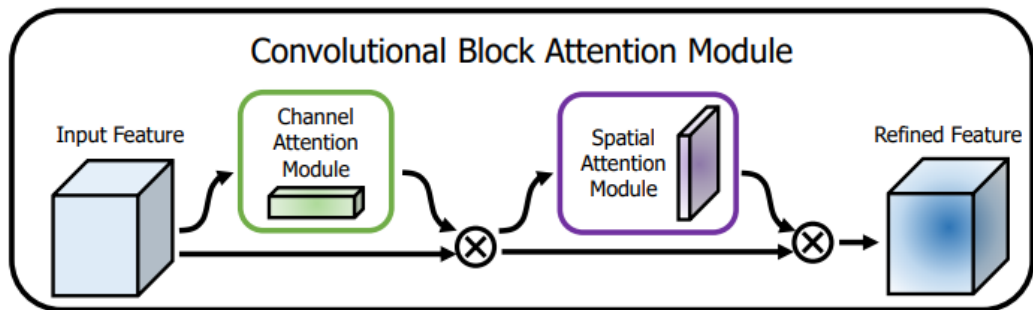
# 通道自注意力机制

如上图所示，数据  $X$  经过卷积操作后，得到  $U$ ， $U$  的通道数用  $C$  表示， $H \times C$  表示一个通道上的长和宽；此后，SENet 引入了一个 Squeeze 模块和一个 Excitation 模块。通过全局平均池化操作将每个通道的特征图转化为一个标量值，简单地说，就是用全局平均池化将每个通道上的数据进行压缩，压缩成一个标量值，即得到一个  $1 \times 1 \times C$  的矩阵。然后，通过激活函数（如 sigmoid 或 ReLU）对  $1 \times 1 \times C$  的矩阵进行操作， $W$  表示的就是激活函数，得到带有颜色的  $1 \times 1 \times C$  的矩阵，用来学习每个通道的权重。最后，经过将这些权重应用于原始特征图上，将带有颜色的  $1 \times 1 \times C$  的矩阵和  $U$  进行点乘，以得到加权后的特征图。最后，将加权后的特征图输入到后续的卷积层进行分类或检测任务。

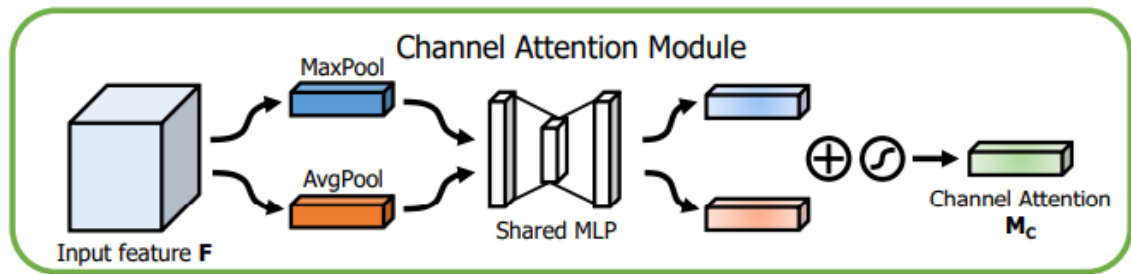
值得注意的是，SENet 并不是一个单独的网络结构，而是可以与其他卷积神经网络结构（如 ResNet、Inception 等）相结合，以增强它们的表达能力。通过在现有网络结构中添加 SENet 模块，可以更容易地将 SENet 应用于现有的深度学习任务中。说的更加简答粗暴点，可以直接在每个卷积之后都可以添加 SENet，当然这样也有可能带来过拟合的问题。

# 空间自注意力机制

空间注意力机制和通道注意力机制具有异曲同工之妙，通道注意力机制旨在捕捉通道的重要性的程度，空间注意力机制旨在通过引入注意力模块，使模型能够自适应地学习不同区域的注意力权重。其中，最为典型的是 CBAM (Convolutional Block Attention Module)，CBAM 是一种结合了通道注意力和空间注意力的模型，旨在增强卷积神经网络对图像的关注能力。从下图可以看出，CBAM 模块由两个注意力模块组成：通道注意力模块 (Channel Attention Module) 和空间注意力模块 (Spatial Attention Module)。



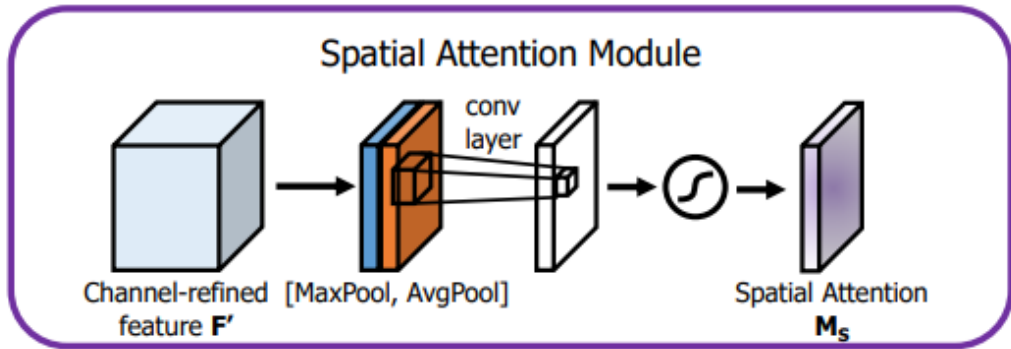
# 空间自注意力机制



CBAM 中通道注意力这边主要简单说一下，它使用全局平均池化和全局最大池化分别来获取每个通道的全局统计信息（SENet 仅使用全局平均池化），并通过两层全连接层来学习通道的权重。然后，会将处理后产生的两个结果进行相加，通过使用 Sigmoid 函数将权重归一化到 0 到 1 之间，对每个通道进行缩放。最后，将缩放后的通道特征与原始特征相乘，以产生具有增强通道重要性的特征。



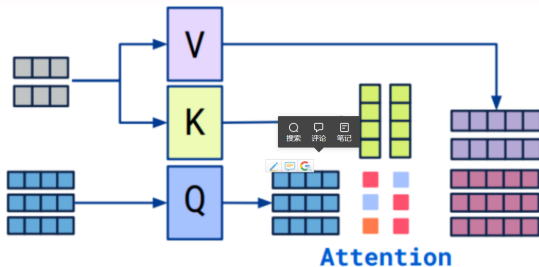
# 空间自注意力机制



CBAM 中空间注意力模块是使用最大池化和平均池化来获取每个空间位置的最大值和平均值。具体地说，由于卷积之后会产生多个通道，CBAM 中空间注意力会在每一个特征点的通道上进行最大池化和平均池化操作，得到两个 matrix 后，将两个 matrix 进行拼接，并通过一个卷积层和 Sigmoid 函数来学习每个空间位置的权重。最后，将权重应用于特征图上的每个空间位置，以产生具有增强空间重要性的特征。

# 交叉注意力机制

在 Transformer 中，CrossAttention 实际上是指编码器和解码器之间的交叉注意力层。在这一层中，解码器会对编码器的输出进行注意力调整，以获得与当前解码位置相关的编码器信息。在 Transformer 的编码器-解码器架构中，编码器负责将输入序列编码为一系列特征向量，而解码器则根据这些特征向量逐步生成输出序列。为了使解码器能够对当前生成位置的上下文进行有效的建模，CrossAttention 层被引入其中。



# 交叉注意力机制

- 编码器输入 (通常是来自编码器的输出): 它们通常被表示为 `enc_inputs`, 大小为 `(batch_size, seq_len_enc, hidden_dim)`。
- 解码器的输入 (已生成的部分序列): 它们通常被表示为 `dec_inputs`, 大小为 `(batch_size, seq_len_dec, hidden_dim)`。
- 解码器的每个位置会生成一个查询向量 (`query`), 用来在编码器的所有位置进行注意力权重计算。
- 编码器的所有位置会生成一组键向量 (`keys`) 和值向量 (`values`)。使用查询向量 (`query`) 和键向量 (`keys`) 进行点积操作, 并通过 `softmax` 函数获得注意力权重。
- 注意力权重与值向量相乘, 并对结果进行求和, 得到编码器调整的输出。