**# Encoder–Decoder框架**

**1. 概念**

* Encoder-Decoder框架可以看作是一种深度学习领域的研究模式，它并不特指某种具体的算法，而是一类算法的统称。
* 这个框架很好的诠释了机器学习的**核心思路**，即: 将现实问题转化为一类可优化或者可求解的数学问题，利用相应的算法来实现这一数学问题的求解，然后再应用到现实问题中，从而解决了现实问题。
* 分为Encoder和Decoder两部分:

1. Encoder: 将现实问题转化为数学问题

图形用户界面, 应用程序

描述已自动生成

1. Decoder: 求解数学问题，并转化为现实世界的解决方案

图形用户界面

描述已自动生成

1. 合并图解

对于文本领域的Encoder-Decoder框架的实际模型案例比如，输入一个句子序列，经过encoder进行非线性编码，获得一个中间向量(中间语义)，decoder根据这个向量和之前生成的历史信息去生成另外一个句子。

注: 除了受中间向量的影响外，通常还受前序逐步生成的历史信息影响，即:

黑白色的标志

中度可信度描述已自动生成

图示

描述已自动生成

**2. Encoder-Decoder框架特点**

* 它是一个end-to-end的学习算法。
* 不论输入和输出的长度是什么，中间向量的长度都是固定的(导致存在信息缺失问题)。
* 根据不同的任务可以选择不同的编码器和解码器(可以是CNN、RNN、 LSTM、 GRU等)，即可以任意选取不同的组合。
* 缺陷:

基础Encoder-Decoder框架存在的最大问题在于信息缺失。

Encoder将输入编码为固定大小的向量的过程是一个"信息有损的压缩过程"，信息量越大，转化得到的固定向量中信息的损失就越大，这就得Decoder无法直接无关注输入信息的更多细节。输入的序列过长，先输入的内容携带的信息可能会被后输入的信息稀释掉或被覆盖了，那么解码的时候一开始就没有获得输入序列足够的信息，可能会导致模型效果比较差。

**3. 不同Encoder-Decoder模式**

* 最简单的解码模式

图示

描述已自动生成

* 带输出回馈的解码模式

**图示

描述已自动生成**

* 带编码向量的解码模式

**图示, 工程绘图

描述已自动生成**

* 带注意力的解码模式

**图示, 工程绘图, 示意图

描述已自动生成**