

量化（4）：AWQ

之前系列的blog地址：

[量化（1）：概念介绍](#)

[量化（2）：SmoothQuant](#)

[量化（3）：OmniQuant](#)

本 blog 介绍 AWQ 的思想，并在 mixtral-7x8B moe 模型上实践。