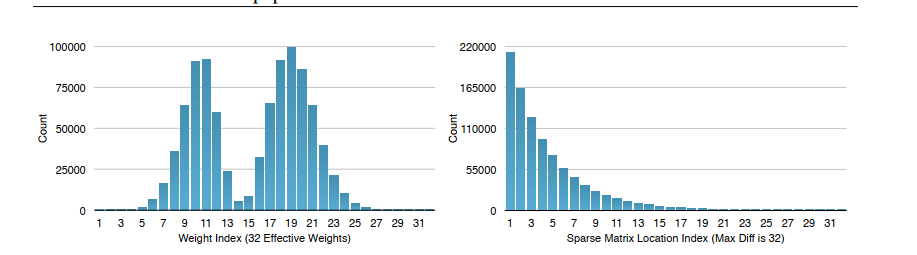
阅读韩松发表在 ICLR 2016 上的文献，请回答：怎么在剪枝和量化训练的基础上进一步压缩网络模型？



答：使用哈夫曼编码，因为如论文中的图所示，量化权重索引和稀疏矩阵位置索引的分布都呈现非均匀分布，而对于非均匀分布的数据的编码可以使用哈夫曼编码，比如对于出现概率最大，即图中计数最多的权重或稀疏矩阵位置，就可以使用位数最少的0或1表示，计数较少的就可以用两位、三位以致更多位表示，具体要使用哈夫曼树或者哈夫曼算法编码，这样像论文中所说可以节省了20%-30%的网络存储。