**Transformer**

1.文本为什么要用向量表示？拓展到词向量算法

2.为什么transformer加进去成为bert，在通用知识库做训练，没在领域做训练

深入transformer ，bert

3.为什么需要多头注意力机制？代码如何实现？推荐阅读：attention is all you need

4.聚焦：bert+知识图谱

5.残差链接:防止过拟合，跳过一层，训练时判断是不是要走那条路

6.如果不用decoder，会出现什么情况?

**DSSM**

1.有关激活函数的选择问题，如sigmod tanh

2.有关余弦相似度，平面以及多维情况下的选择？

3.为什么需要进行梯度更新？

4.有关adam优化器的原理，有多少种优化器？

5.推荐选修课—于祥春 数据挖掘

**其他**

**补基础：**

数理 代码实践 英语

**关于论文：**

越早开始写论文越好

研究什么样的问题

怎么研究问题

模型是什么

结果是什么

怎么论证

有什么意义

提问题

讲论文

文献管理工具