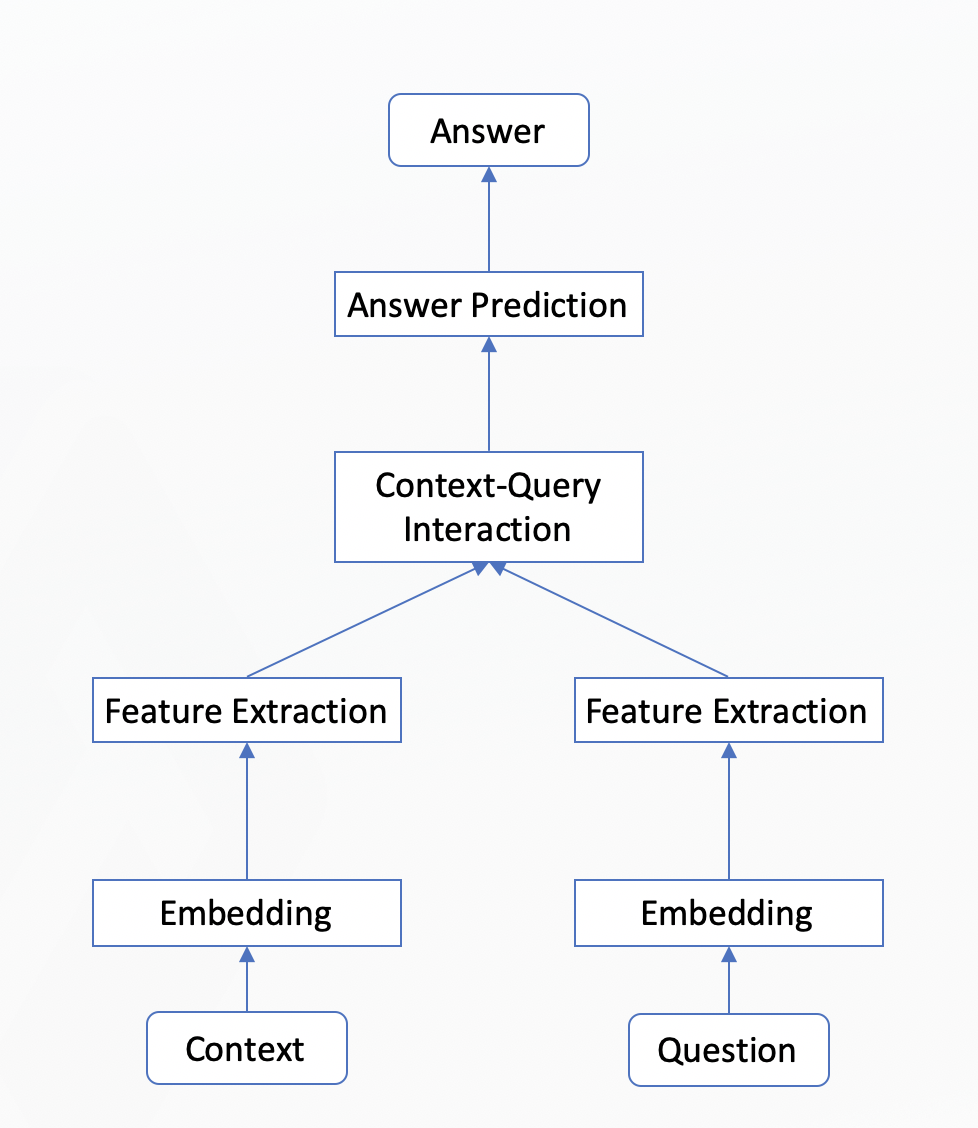
本部分主要介绍了机器阅读理解整个项目总结，机器阅读理解比较偏学术方向，整个过程是围绕如何提升算法效果，其基础架构如下：



了解结构之后学习了MRC的评估指标

GLOVE的方式是通过相除的概率比值来确定是否相关。

基础部分还介绍了ATTENTION，其本质是计算权重的方法，描述哪些重要，哪些不重要，权重越大则特征越重要，其权重是通过QUERY和KEY进行相似度计算，然后归一化得到，将权重和value进行加权求和，得到value的重要性。一般采用训练的方式计算attention。项目主要描述的self attention。

TextCNN和CharCNN来自于图像CNN的算法，TEXTCNN和CharCNN都是一种文本分类的算法，只是粒度大小的区别，CharCNN主要是解决OOV的问题。

在BERT之前主要有BiDAF、RNet（引入了Pointer Network,注意力机制融入RNN，门控循环神经网络）、FusionNet(把之前的各种框架大一统，进行全关注注意力以及单词历史)、QANet（CNN+Self-Attention，翻译来实现数据扩增）

在BERT之后开始对机器阅读理解模型进行了统一架构，通过海量语料先验知识进行预训练，然后利用自身业务数据进行finetune模型中，从而调整模型权重。BERT有两个预训练任务（Mask-LM和NSP），其输入是基于位置的编码信息。

ERNIE引入了百度新闻数据等中文数据进行优化，并且在2.0进行连续预训练框架，从词级别到结构级别的预训练任务整合进去，是从任务上改进BERT，没有改变BERT结构。

Transformer-XL 通过循环机制和相对位置编码的方式优化了编码的方式，XLNet通过排列组合的方式建立语言模型开展双流注意力构造模型。ALBERT是采用了两种减少参数方法（矩阵分解和参数共享），使用SOP替换掉NSP，采用n-gram MASK 进行学习。Retro-Reader是基于BERT和其他策略的组合阅读理解模型。

在讲完所有预训练模型之后，介绍了对不同模型进行权重得到相应的综合效果，这就是模型集成放啊放，有bagging/boosting/stacking/blending的方法。模型蒸馏是采用小的模型去学习大模型的预测结果以及泛化能力。

后续学习路线需要利用统计学习方法来夯实基础（例如LR， GBDT的变体在面试比较多，统计学习方法和西瓜书），在项目应用方面要多参加实战和比赛进行参数调整，在跟进前沿上学习大厂论文和开源项目，不仅仅是MRC。