第二阶段报告

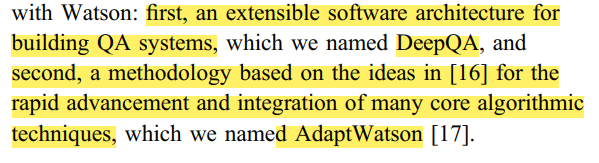
伍锦添 2018200532（第12小组）

1. 阶段进程及问题

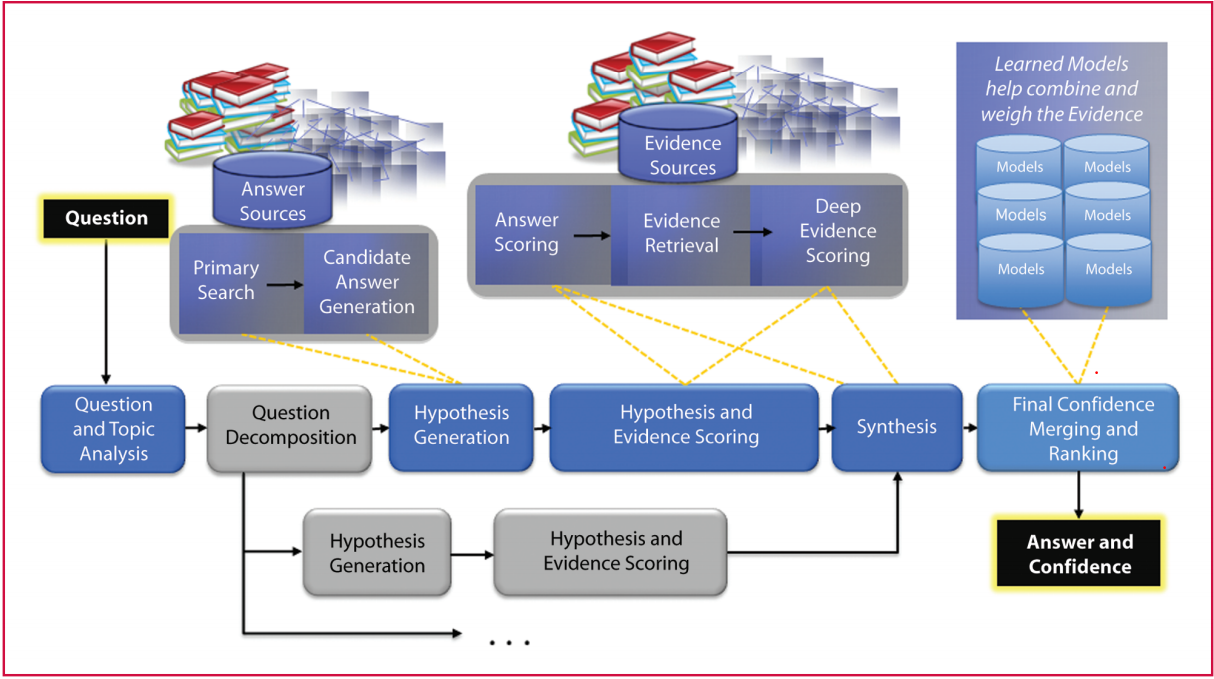
考虑到难度问题，本人计划仅完成项目中的“自然语言识别”部分。本阶段主要内容为收集论文，探究实现项目操作的方法。不足是未按照计划进行编码，还需要进一步学习python技术和理论知识

1. 项目介绍（来自官方文件）

·Foundation of Watson: DeepQA + AdaptWatson



·DeepQA实现流程及要素整理



1）Question and Topic Analysis：自然语言识别技术

-ESG (English Slot Grammar parser)：句式分析

-PAS (predicate-argument structure)：更加抽象的分析

2）Answer and evidence sources

3）Discovering candidate answers: classification——提高召回率

Set of candidate answers——each become independent process of proving hypothesis

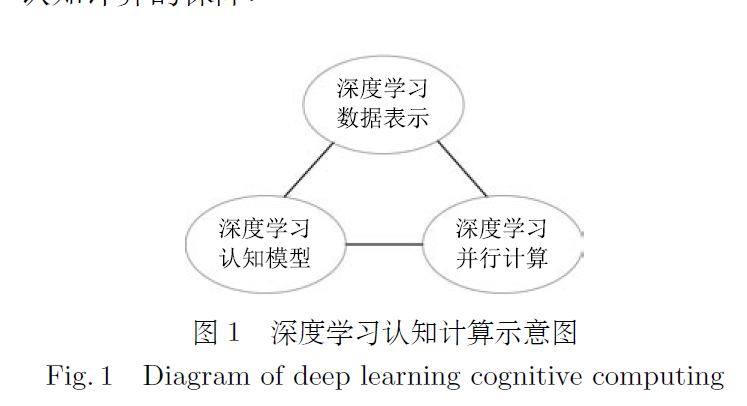
4）Collecting and scoring evidence

1. 论文查找

本部分选择了查找中较为有用的两篇论文，对内容做大致梳理。

1. 陈伟宏、安吉尧、李仁发、 李万里：《深度学习认知计算综述》

·深度学习认知计算的要素：



首先，在数据表示部分介绍了基于多模态的数据表示和基于张量的数据表示。在深度学习认知模型部分介绍了较为传统的卷积神经网络、栈式自编码器、递归神经网络，同时引入了基于张量的深度学习模型。而并行计算则是将一个计算问题分为多个子问题，用多个处理器同时处理的技术手段，方法包括GPU加速、数据与模型并行以及搭建计算集群。

1. 奚雪峰、 周国栋：《面向自然语言处理的深度学习研究》

本篇的重点主要集中在“数据表示”部分，首先介绍了one-hot representation的方法，并说明其不足在于无法表示词与词之间的相关性，并在后文提出“词向量”的表示方法，解决“词”一级的问题。而在“文本”一级则介绍了词袋模型，同时也介绍了向量空间模型及其应用。