

An Intelligent Question and Answering Framework Based on Chinese Corpus

Advisor Zhao Zhou

Member Zhang Feiyu, Chen Yifei, Luo Wenqing

Contents

Overview

Advisor and Team

Timeline

Expected ResultExpected Result

Research Plan

Overview

Recently, the intelligent Q&A system has achieved a lot of accomplishment applying specific models on specific question matching and answering problem, which consists of text preprocessing, retrieval and deep semantic matching, automated testing and result ranking, but the research still lacks big idea of overall Q&A system.

Considering this situation, the project will utilize the massive data in Chinese corpus, proposed to come up with a modularized Q&A framework, which provides a unified interface for different text matching models. By this, developers can freely config or modify the models according to actual needs. And we will dig from the theoretical construction level to the engineering practice level, exploring the industry standards of the applied Q&A system.

Our Advisor



Zhao Zhou is the associate professor of College of Computer Science and Technology in Zhejiang University. He received his PHD in Hong Kong University of Science and Technology. Zhou Zhao's main research areas are intelligence question answering and social media mining, and has published more than 30 top conference/journal papers. Among them, as the first author of the paper, Zhou Zhao has published 12 CCF-A conference/journal papers.

In 2018, Zhou Zhao was awarded 30 under 30 by Forbes China.

Team Member

Zhang Feiyu, student of ZJU-UIUC Institute major in Computer Engineering, has previous experience in deep learning and pattern recognition.

Luo Wenqing, student of ZJU-UIUC Institute major in Computer Engineering, has previous experience in machine learning, cloud computing and distribute service.

Chen Yifei, student of Chu Kochen Honors College of Zhejiang University major in Applied Mathematics, has previous experience with econometric model design.

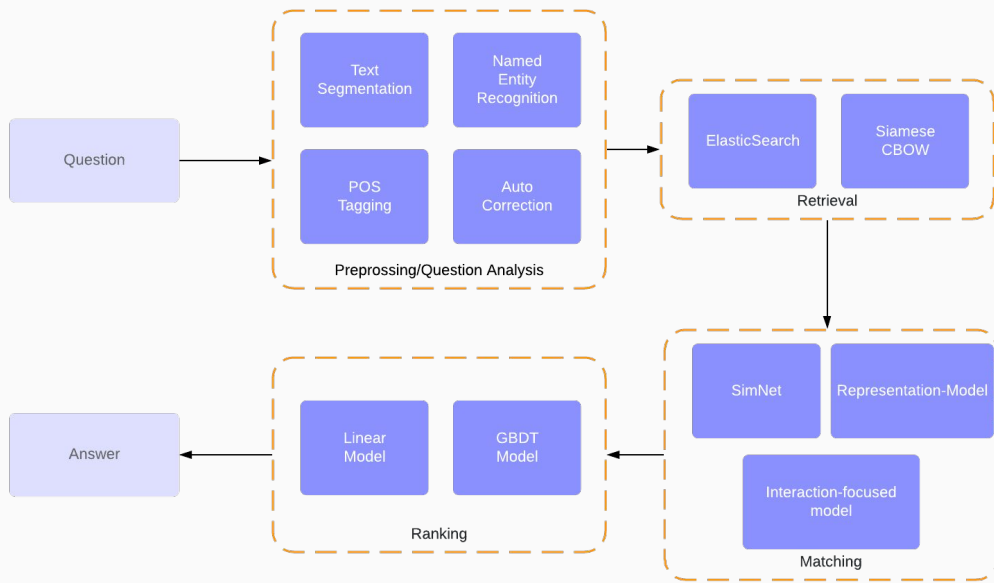
Research Plan

Architecture

Here we show the architecture of our Q&A framework.

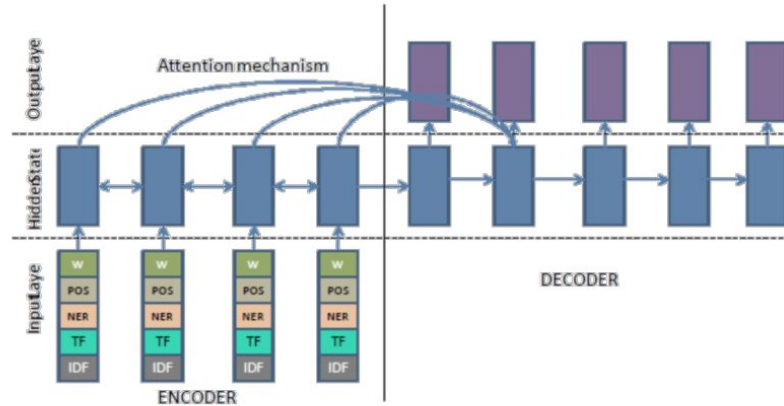
There are some features of our architecture:

- Each module connect in an unified interface
- All the models like deep text matching model, Chinese text segmentation model will be treated as a plugin, which will be very easy to be added and removed.
- Using RPC or message queue for data transmitting.



Deep Text Matching Model

Besides convenient interface for developers to modify their own models, this framework will also implement its native deep text matching models. The project intend to include two types of text matching models: One is representation based, and the other is interaction based. Developers can use these models like “black-box”, and of course, they can also use the interface to modify their own models and contribute to future open source project.



Seq2Seq with Attention

A text matching model example (seq2seq)

Timeline



Expected Result

First Step

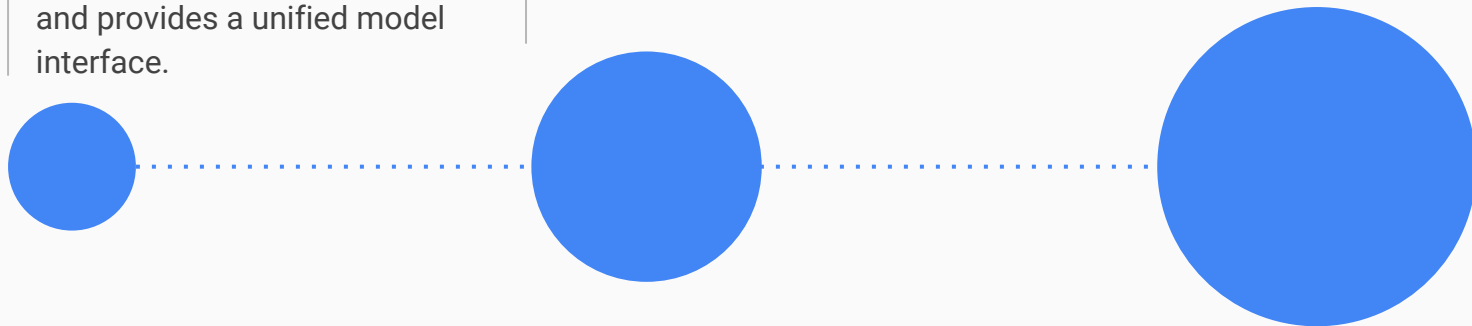
Build a basic framework which support several NLP models, and provides a unified model interface.

Second Step

Conclude our practice and publish a paper

Third Step

Making our framework open source, test and deploy it in production environment.



Innovation

Innovation will consist of three aspects:

- For research object, we are based on the chinese corpus, and doing some specifical optimization for that.
- From framework level, we introduce a toolkit that makes text matching model plugin like.
- Finally, we focus on both scientific study and industrial application.

Chinese Corpus

Model Configurable and
Plugin like

Combination of Scientific
Study and Industrial
Application

Thanks!
Q&A Section