

Reading Comprehension Datasets

- CBT(Children's Book Test): Predicting a blanked-out word of a sentence given the 20 previous sentences
- CNN/DailyMail: cloze style questions, blanking out entities in abstractive summaries. Fill in the entity based on the original article
- BAbi: find a word to answer the question

- MCTest: 4 answer choices per question. need reasoning. small
- WikiQA: open-domain, sentence selection, small
- MS MARCO: real-world Bing/Cortana queries, real web docs, 10 passages each question, complete sentence answer generated by human, unanswerable queries, large

SQuAD Dataset

Structure

- large, enable data-driven model
- wiki articles, 500 articles, 20k paragraphs, 100k questions, a wide range of topics

Structure

- high-quality, made by human
- answers are span/phrases within a sentence, not a single word or sentence, no choices
- need to extract answer, harder than previous datasets

Collecting Process

- crowdworkers give 5 questions and answers for each paragraph
- avoid using the same words/phrases as in the paragraph
- encouraged to ask hard questions requiring reasoning
- more close to real world scenario

Questions

- answer diversity

| Answer type | Percentage | Example |
|--------------------|------------|-------------------------|
| Date | 8.9% | 19 October 1512 |
| Other Numeric | 10.9% | 12 |
| Person | 12.9% | Thomas Coke |
| Location | 4.4% | Germany |
| Other Entity | 15.3% | ABC Sports |
| Common Noun Phrase | 31.8% | property damage |
| Adjective Phrase | 3.9% | second-largest |
| Verb Phrase | 5.5% | returned to Earth |
| Clause | 3.7% | to avoid trivialization |
| Other | 2.7% | quietly |

| Reasoning | Description | Example | Percentage |
|-------------------------------------|--|--|------------|
| Lexical variation (synonymy) | Major correspondences between the question and the answer sentence are synonyms. | Q: What is the Rankine cycle sometimes called ? Sentence: The Rankine cycle is sometimes referred to as a <u>practical Carnot cycle</u> . | 33.3% |
| Lexical variation (world knowledge) | Major correspondences between the question and the answer sentence require world knowledge to resolve. | Q: Which governing bodies have veto power? Sen.: <u>The European Parliament and the Council of the European Union</u> have powers of amendment and veto during the legislative process. | 9.1% |
| Syntactic variation | After the question is paraphrased into declarative form, its syntactic dependency structure does not match that of the answer sentence even after local modifications. | Q: What Shakespeare scholar is currently on the faculty ? Sen.: Current faculty include the anthropologist Marshall Sahlins, ..., Shakespeare scholar <u>David Bevington</u> . | 64.1% |
| Multiple sentence reasoning | There is anaphora, or higher-level fusion of multiple sentences is required. | Q: What collection does the V&A Theatre & Performance galleries hold? Sen.: The V&A Theatre & Performance galleries opened in March 2009. ... They hold the UK's biggest national collection of <u>material about live performance</u> . | 13.6% |
| Ambiguous | We don't agree with the crowdworkers' answer, or the question does not have a unique answer. | Q: What is the main goal of criminal punishment? Sen.: Achieving crime control via <u>incapacitation and deterrence</u> is a major goal of criminal punishment. | 6.1% |

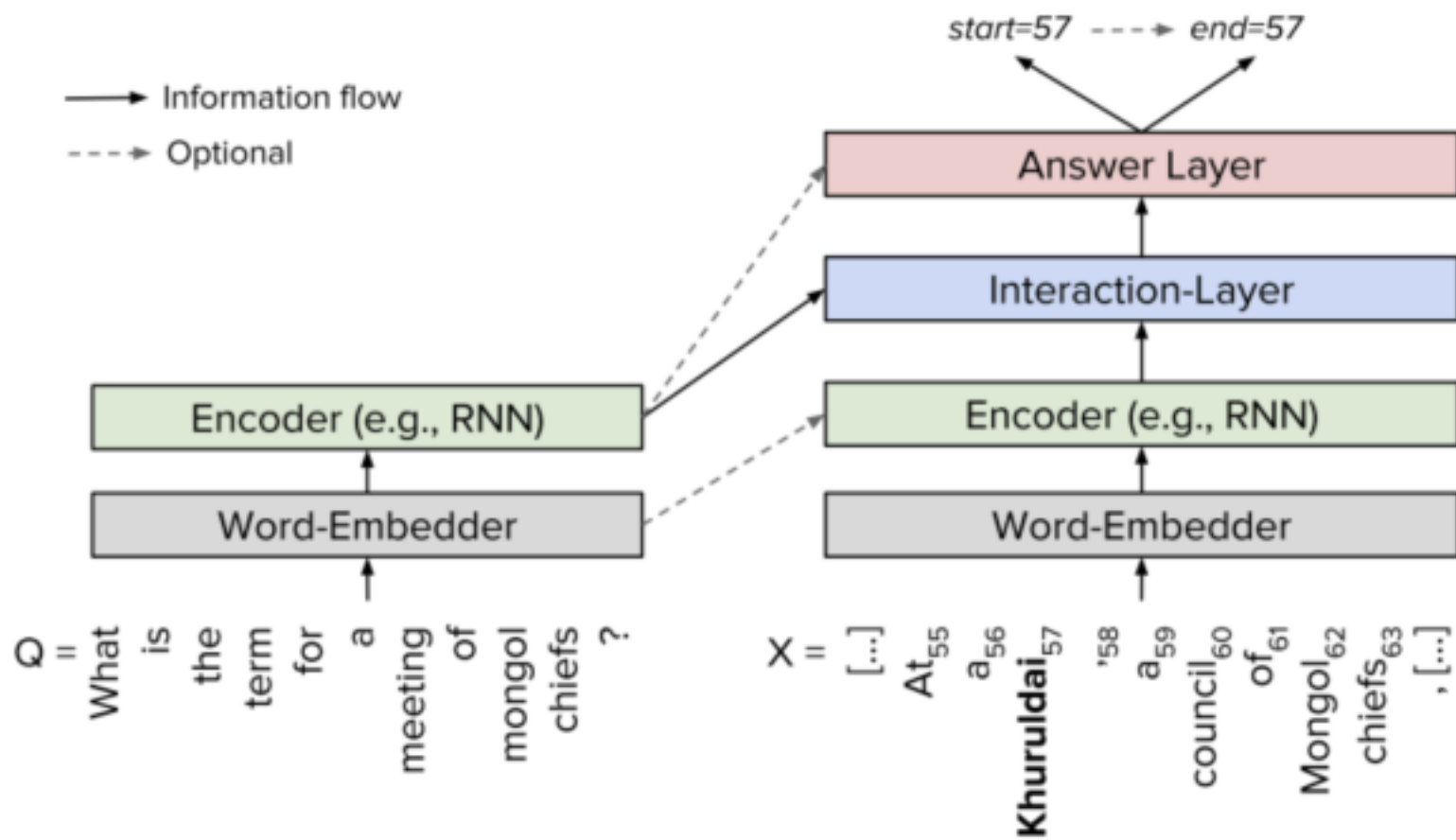
Baseline

- candidate answer + sentence lexical feature

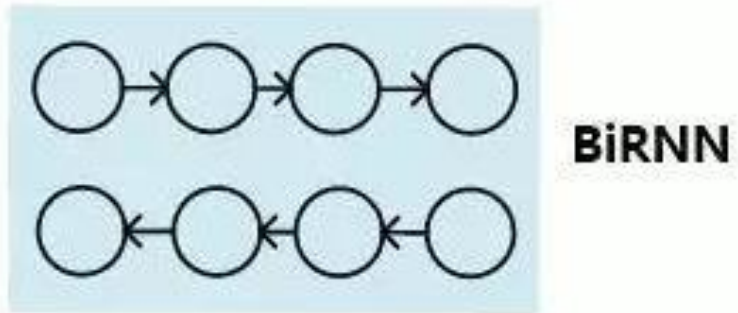
| | Exact Match | | F1 | |
|----------------------|-------------|-------|-------|-------|
| | Dev | Test | Dev | Test |
| Random Guess | 1.1% | 1.3% | 4.1% | 4.3% |
| Sliding Window | 13.2% | 12.5% | 20.2% | 19.7% |
| Sliding Win. + Dist. | 13.3% | 13.0% | 20.2% | 20.0% |
| Logistic Regression | 40.0% | 40.4% | 51.0% | 51.0% |
| Human | 80.3% | 77.0% | 90.5% | 86.8% |

End-to-End RC Solution

Work Flow

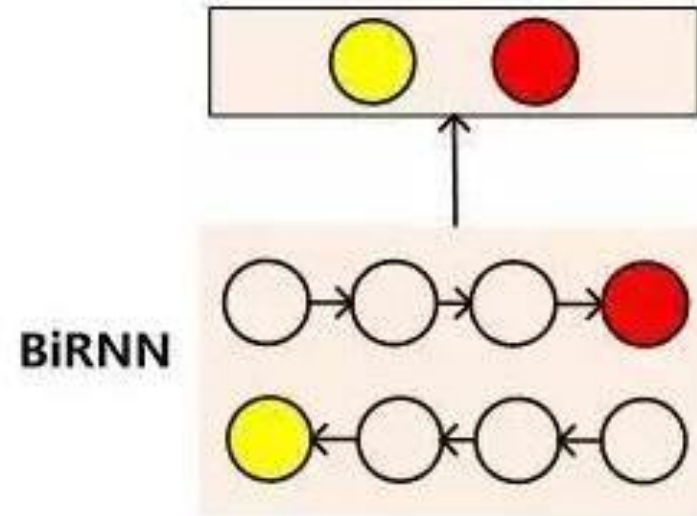


Representation: BiRNN



Document

图4.文档表示方法：模型一



Question

图8.问题表示方法：模型三

- char level & word level & sentence level

Attention

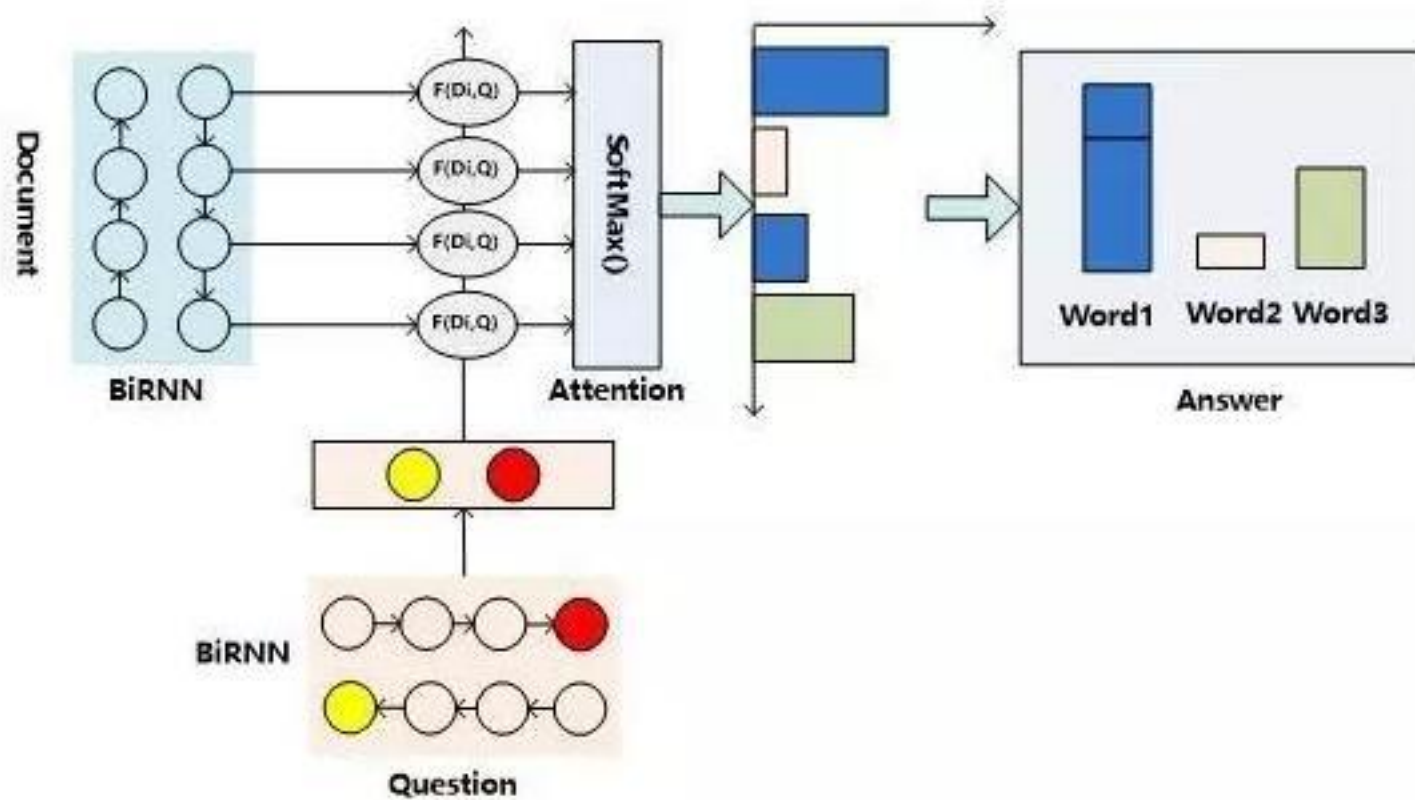


图9. 机器阅读理解的一维匹配结构

- e.g. Match-LSTM
- BiDAF

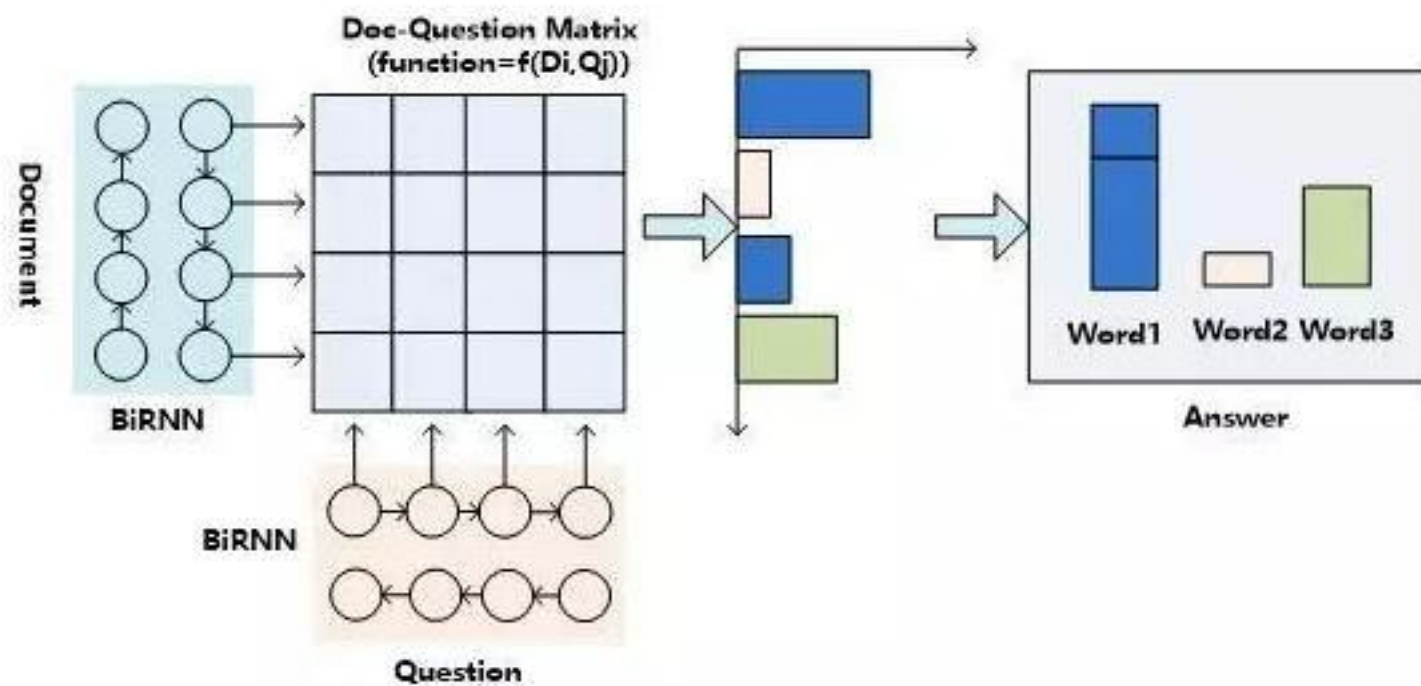


图10. 机器阅读理解的二维匹配结构

Multi-Iteration

- imitate human inference
- e.g. Memory Networks
- ReasonNet

展望

- attention匹配模型能否做得更复杂
- multi-hop推理模型能否更深入

- 处理更复杂的数据集
- 外部世界知识的引入

references

- Rajpurkar et al. SQuAD: 100,000+ Questions for Machine Comprehension of Text
- Nguyen et al. MS MARCO: A Human Generated Machine Reading Comprehension Dataset
- 张俊林. 深度学习解决机器阅读理解任务的研究进展 – 知乎专栏 <https://zhuanglan.zhihu.com/p/22671467>