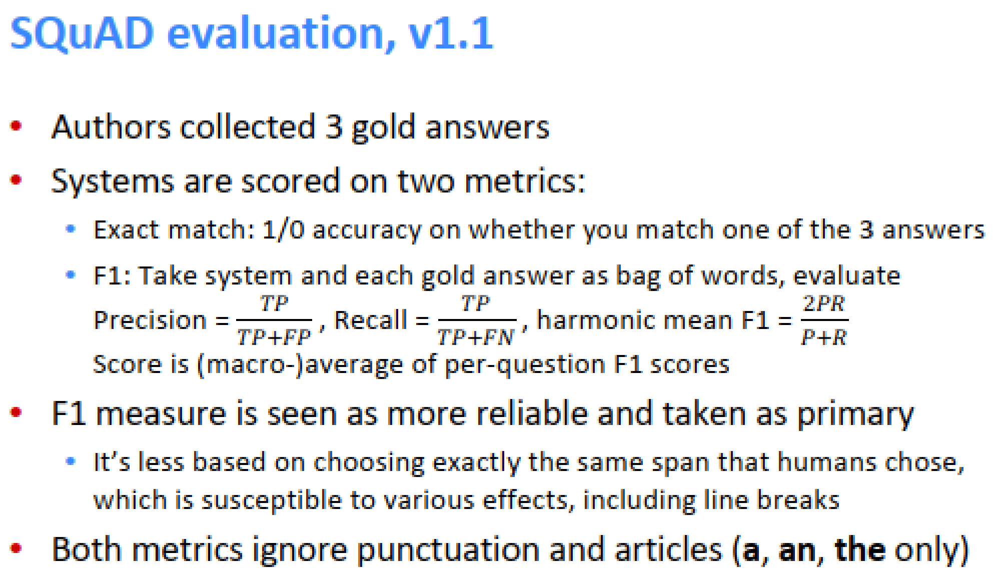
Textual Question Answering

SQuAD 1.1

输入T和Q，输出回答A。A必须是T中一个span，即extractive question answering。



SQuAD evaluation, v1.1

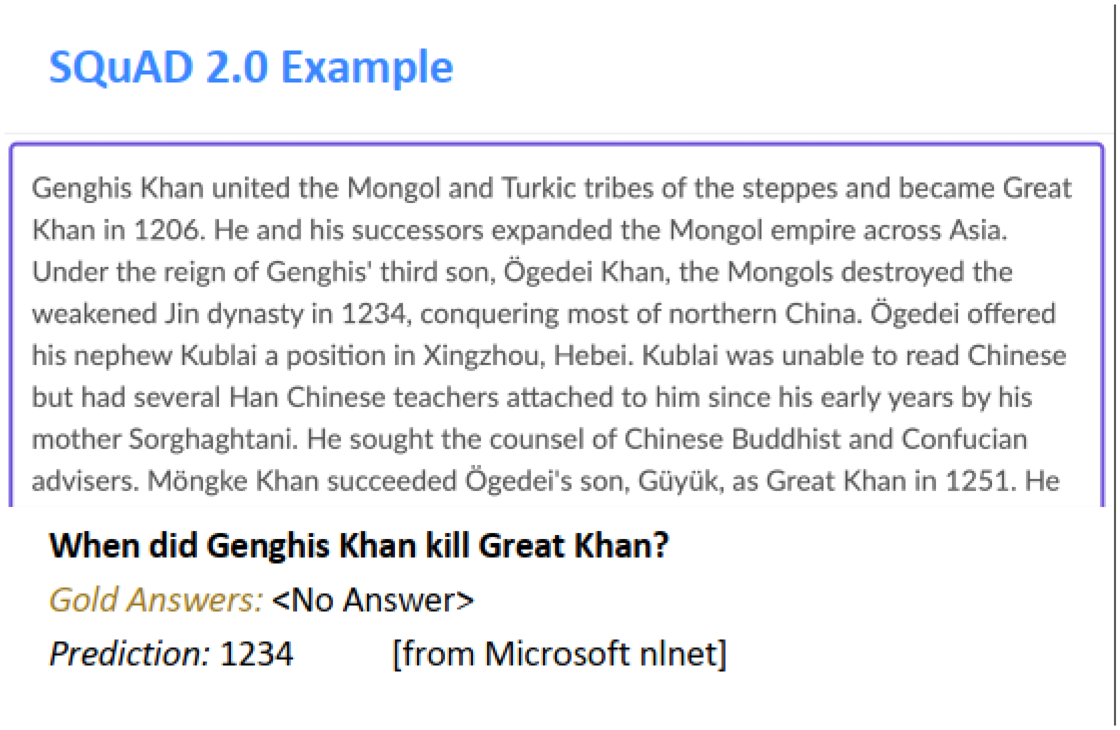
- Authors collected 3 gold answers
- Systems are scored on two metrics:
  - Exact match: 1/0 accuracy on whether you match one of the 3 answers
  - F1: Take system and each gold answer as bag of words, evaluate Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$, harmonic mean F1 = $\frac{2PR}{P+R}$
    Score is (macro-)average of per-question F1 scores
- F1 measure is seen as more reliable and taken as primary
  - It’s less based on choosing exactly the same span that humans chose, which is susceptible to various effects, including line breaks
- Both metrics ignore punctuation and articles (**a**, **an**, **the** only)

缺点

- 所有问题都有答案
- 系统只需要对span排序，并取分数最高的

SQuAD 2.0

train中1/3是no answer；dev/test中1/2是no answer。



SQuAD 2.0 Example

Genghis Khan united the Mongol and Turkic tribes of the steppes and became Great Khan in 1206. He and his successors expanded the Mongol empire across Asia. Under the reign of Genghis' third son, Ögedei Khan, the Mongols destroyed the weakened Jin dynasty in 1234, conquering most of northern China. Ögedei offered his nephew Kublai a position in Xingzhou, Hebei. Kublai was unable to read Chinese but had several Han Chinese teachers attached to him since his early years by his mother Sorghaghtani. He sought the counsel of Chinese Buddhist and Confucian advisers. Möngke Khan succeeded Ögedei's son, Güyük, as Great Khan in 1251. He

When did Genghis Khan kill Great Khan?

Gold Answers: <No Answer>

Prediction: 1234 [from Microsoft nlnet]

缺点

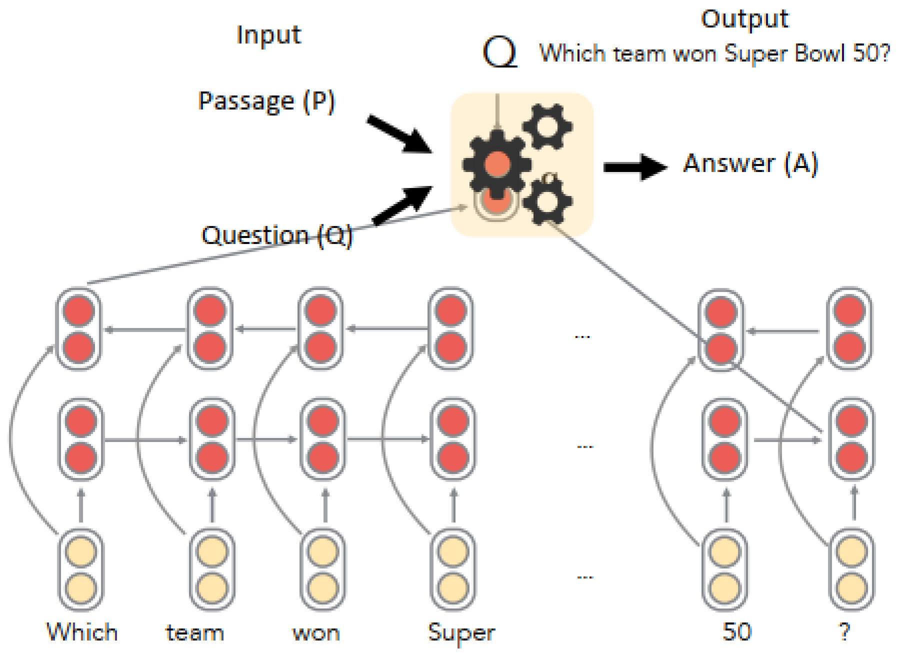
- 答案必须是一段span，不能yes/no、计数等
- 不能有隐含意义，像"作者xxx的举动反应他的什么情绪"

但SQuAD仍是最广泛使用的QA数据集。

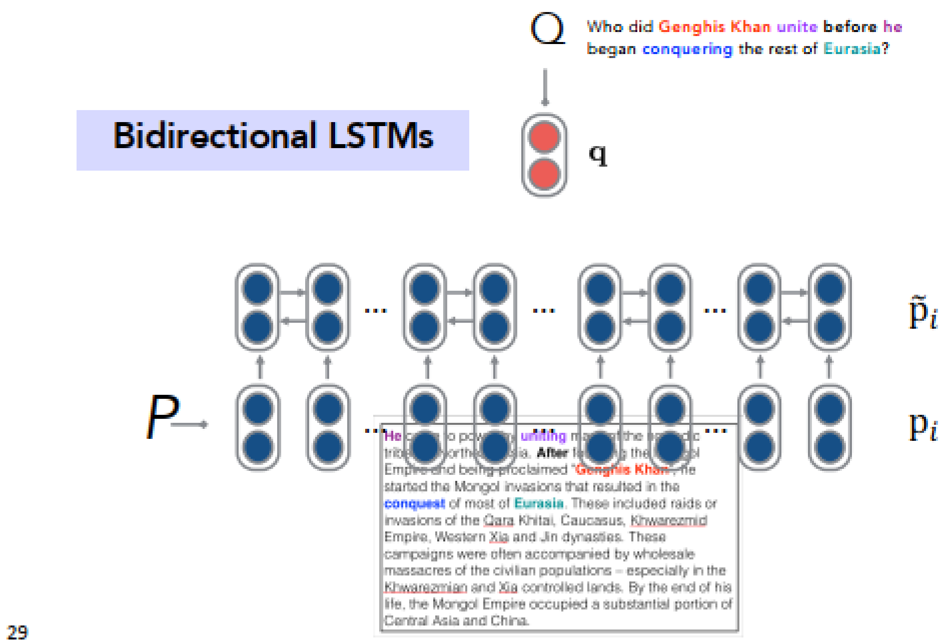
Stanford Attentive Reader

具体算法流程:

1. 通过bi-lstm计算Q的向量:

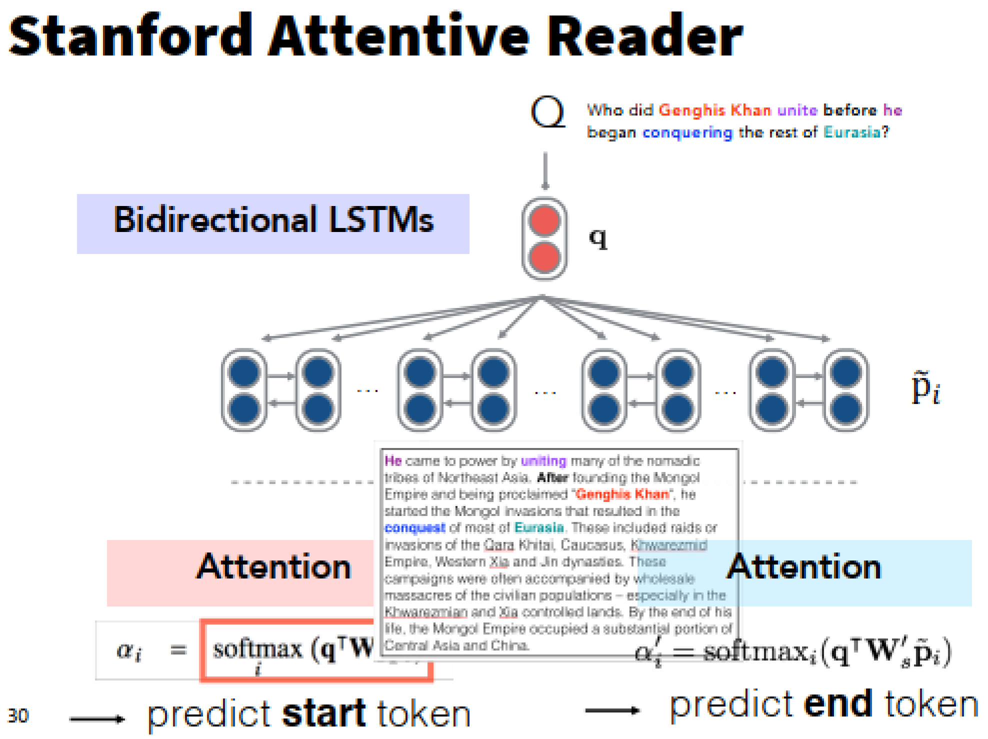


2. 通过另一个bi-lstm计算每个P中单词的向量:



Q
Who did Genghis Khan unite before he began conquering the rest of Eurasia?
Bidirectional LSTMs
q
$\tilde{\mathbf{p}}_i$
P
$\mathbf{p}_i$
29

3. Q向量和P中单词向量计算attention分数，并得到A的<START>和<END>位置:



Stanford Attentive Reader

Q
Who did Genghis Khan unite before he began conquering the rest of Eurasia?
Bidirectional LSTMs
q
$\tilde{\mathbf{p}}_i$
Attention
Attention
30

He came to power by uniting many of the nomadic tribes of Northeast Asia. After founding the Mongol Empire and being proclaimed "Genghis Khan", he started the Mongol invasions that resulted in the conquest of most of Eurasia. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezmid Empire, Western Xia and Jin dynasties. These campaigns were often accompanied by wholesale massacres of the civilian populations – especially in the Khwarezmian and Xia controlled lands. By the end of his life, the Mongol Empire occupied a substantial portion of Central Asia and China.

$\alpha_i = \text{softmax}_i(\mathbf{q}^\top \mathbf{W}$

$\alpha'_i = \text{softmax}_i(\mathbf{q}^\top \mathbf{W}'_s \tilde{\mathbf{p}}_i)$

→ predict **start** token

→ predict **end** token

Stanford Attentive Reader++

2点改进:

- SAR用正向lstm和反向lstm的最后一层的hidden state作为Q向量。SAR++用bi-lstm每一层hidden state的加权和作为Q向量。
- wordvec加入manual特征，包括POS tags/ frequency/exact match等

BiDAF (Bi-Directional Attention Flow)

Attention Flow

之前各个context word对query的attention，而attention可以是双向流动的。BiDAF考虑context2query和query2context的attention。

Dynamic Coattention Networks

C2Q+Q2C+second-level attention

FusionNet

multi-level attention