Talking Geckos QA System

Jie Cao
Boya Song
Overview

• Rule based system [Based on Ellen’s Quarc Paper]

• Stanford core nlp lib: POS, sentence split, NER, dependency relations

• Classify questions according to keyword: what, where, when, why, how, etc.

• Word Match function

• Rules that award points to each sentence

• Narrow down answer according to key word type
Work Process

Input

1. wordMatch(Common Component)
   (Question, StorySentences) -> ScoredSentences

2. reverseWordMatch(Make Assumption)
   correctAnswer -> correctSentence

3. narrowDownAnswer(Why)
   correctSentence -> correctAnswer

4. narrowDownAnswer(What)
   correctSentence -> correctAnswer

5. narrowDownAnswer(…)
   correctSentence -> correctAnswer

6. narrowDownAnswer(How)
   correctSentence -> correctAnswer

7. BestSentence(Why)
   ScoredSentence -> correctSentence

8. BestSentence(What)
   ScoredSentence -> correctSentence

9. BestSentence(…)
   ScoredSentence -> correctSentence

10. BestSentence(How)
    ScoredSentence -> correctSentence

Rule Isolation

Developing Task Parallelism

Google: Fast is better than slow
Reports and Testing

- WordMatch Rules
- Object Persistence, Reload
- Pre-annotated Story and Question
- Classify Question Type Rules
- Test By Question Type List
- Annotated Test Report
- Annotated and Scored Report For Every Question Type

151 Stories, 1200 Questions
Test all data in less than 1 minute
Precision Improvement

- Used reports and searched for general patterns to narrow down answer within sentence
- Find continuous named entity tags in answer
- Location, time, and preposition keywords
- Why: substring after “so” or “because”—simple but very effective
- What: substring after core sub, core verb, root verb, or strike of word contained in question
- Who, Where type of questions are more complicated
Sample Report

QuestionID: 1999-W38-1-9
Question: How many peacekeeping troops does Canada now have in 22 nations around the world?
Answer: about 3,900
SentenceSize: 1
CorrectSentence: Right now, Canada has about 3,900 peacekeepers on 22 missions around the world.
MyAnswer: 3,900
MyAnswerSentence: Right now, Canada has about 3,900 peacekeepers on 22 missions around the world.
MyAnswerScore: AnswerScore{recall=1.0, precise=0.5, fmeasure=0.6666666666666666, myCorrect=1, correctTotal=1, myTotal=2, matchKey='3,900'}
Difficulty: easy
Included: Yes

QuestionID: 1999-W37-5-3
Question: How much longer did Meiorin take to run 2.5 kilometres than she was supposed to?
Answer: 49 seconds
SentenceSize: 1
CorrectSentence: She took 49 seconds longer.
MyAnswer: 49 seconds
MyAnswerSentence: She took 49 seconds longer.
MyAnswerScore: AnswerScore{recall=1.0, precise=1.0, fmeasure=1.0, myCorrect=2, correctTotal=2, myTotal=2, matchKey='49 seconds'}
Difficulty: moderate
Included: Yes

rightSentence=94, length = 175, avgRecall = 0.47672550966159993, avgPrecision = 0.3014595835896379, avgFmeasure = 0.32152025726198186
Question: Why is the Sheldon Kennedy Foundation abandoning its dream of building a ranch for sexually abused children?

MyAnswerSentence: Troubled by poor business decisions, the Sheldon Kennedy Foundation has abandoned its dream of building a ranch for sexually abused children and will hand its donations over to the Canadian Red Cross.
WordMatch: TokenScore

1. Generate
   1. VerbsInQuestion, OthersInQuestion, rootVerb, subjInQuestion, dObjInQuestion

2. General Word Match Score
   1. WORD: “word_0.5”
   2. DIS_WORD: “disword_{pow(2,2-distance)}” // unimportant “Mod”, longer distance from the root

3. Continuous Word Match Bonus Score
   1. STRIKE: “strike_1” // for every continuous word (not include stopword)

4. Verb Match
   1. Verb stopwords no score. // lucene stopwords
   2. VerbsInQuestion.contains(“verb”)
      • AUX_VERB: 0.5
      • ROOT_VERB: 6
      • COM_VERB: 0.5
      • CORE_VERB: 4

5. Noun Match
   1. ROOT_VERB: 6 // Root Verb in noun form
   2. subjInQuestion != null && matched
      • SUBJ: 1 // match subj in question.
      • CORE_SUBJ: 3 // match core_subj in question
      • SEC_VERB: 3 // verb of this matched noun is also in Question
   3. dObjInQuestion != null && matched
      • DOBJ: 1 // match dobj in question
      • SEC_VERB: 3 // verb of this matched noun is also in Question

Future Work:
1. Learning weights for features
dcoref
2. Thesaurus

Time Limited
Just Extract Features, Manual tuning weights
## BestSentence: Rule Score

<table>
<thead>
<tr>
<th></th>
<th>WEAK_CLUE=1</th>
<th>CLUE=2</th>
<th>GOOD_CLUE =4</th>
<th>CONFIDENT=6</th>
<th>SLAM_DUNK=20</th>
<th>Only TokenScore in all dataset Average: CorrectSentence/All</th>
<th>TokenScore+RuleScore in all dataset Average: CorrectSentence/All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where</td>
<td>GOOD_CLUE: WHERE_LOCATION_PREP CLUE: WHERE_LOCATION_NER</td>
<td>86/155=0.5548</td>
<td>93/155=0.6000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When</td>
<td>GOOD_CLUE: WHEN_TIME_NER SLAM_DUNK: WHEN_ORDER_TOKEN SLAM_DUNK: WHEN_BEGIN_TOKEN</td>
<td>117/173=0.6763</td>
<td>119/173=0.6879</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Who/ Whose</td>
<td>GOOD_CLUE:WHO_PERSON_NER CLUE:WHO_ORGANIZATION_MISC_NER CLUE:WHOSE_PRP_POS</td>
<td>116/187=0.6203</td>
<td>118/187=0.6310</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What</td>
<td>GOOD_CLUE: WHAT_KIND_TOKEN CLUE: WHAT_DATE_NER SLAM_DUNK: WHAT_NAME_TOKEN</td>
<td>202/311=0.6325</td>
<td>202/311=0.6325</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How</td>
<td>HOW_SECKEY_K: distance from k to root GOOD_CLUE: HOW_DISTANCE_TEXT</td>
<td>141/232=0.6078</td>
<td>143/232=0.6164</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**testset1** 52/84 What, 35/57 How, 27/44 Where, 25/41 Who, 23/32 When, 16/28 Why ...

**testset2** 57/88 What, 41/50 How, 28/44 Who, 27/40 When, 24/38 Where, 21/31 Why ...
Summary

- Reporting system worked very well, helped to improve precision
- System is fairly simple and straightforward
- Unfortunately due to Stanford core nlp lib bug, we could not incorporate coreference resolution into our system
- Short on time: Tuning weights manually (ML?) Recall and precision still have room for improvement
Thanks

Q&A