Milestone 2: Error Analysis

(return of the milestone)

Team 3
Hector Liu
Da Teng
Guoqing Zheng
Xiawen Chu
Ryan Carlson

Overview

- (Quick) Progress Update
- Errors
- Proposed Fixes to ^Errors

Training Data Preprocessing

- Data is messy.
- Let’s clean it up

- Write regexs to identify citations, years, (some) names
- Annotate only the “relevant” data so it’s easy for us to index only the important stuff

Question Types

- From description:
  - factoid, causal, method, purpose, true/false

- Those are bad. We want to describe the **type of answer** these questions should lead us to.
- Let's make some better labels
  - Quantity, Cause, Factoid, Relation, Time, Binary

- Also: "Multiple" and "Negated" labels

Error Analysis 1

Empty or answer not included in candidate sentences

- Candidate sentences are used to narrow down the answer search space
- The recall here is essential to ensure successful answer selection

Error Analysis 1

- Causes
  - term mismatch
    - Morphologically changed term
    - e.g. neurodegenerative <-> neurodegeneration
  - phrase search only
    - If the noun phrase is not matched as a whole, the sentence will not be retrieved
    - e.g.
    - How many persons ...
    - the noun phrase is “many persons”

Error Analysis 1

- Solution
  - For term mismatching
    - We adopt a morphology analyzer
    - We also store long terms with their prefix (currently length 12, an ad-hoc decision)
  - For phrase search problem
    - We adopt a back-off model, when there are too little candidate sentences, we use token search instead of phrase search

Error Analysis 2

Background Corpus

- The baseline system use an external background corpus on a provided server
  - We cannot control the fields in it
  - It do not include some information given in the text
- In this task, answers are from the given document
  - We should retrieve answers from it

Error Analysis 2

- Solution
  - Supplement the background corpus with a local index
  - Add in relevant fields that we are interested in

Error Analysis 3

Incorrect answer due to hard voting

- Sometimes we can see incorrect answers due to hard voting strategy
  - Voting using the first ranked answer
  - Seems to be to harsh, do not consider the PMI score

Error Analysis 3

- Solution
  - We use soft voting, i.e., use average scores of answer choice to all candidate sentences as the final score of the answer choice

Error Analysis 4

Clear irrelevant answers

- Some answers are clearly irrelevant
  - Q: How many residues does the CLU2 protein sequence have?
    - A: 449
    - A: protein
    - A: 82.3
    - A: 52.5
    - A: 6
  - The second answer is not even a number

Error Analysis 4

- Solution
  - We adopt question type information to filter out irrelevant answers
  - We have built annotators that find relevant information, such as number annotators
  - We are building methods to use such annotations
  - Attempting to implement a classifier next step

Iteration results

System	Baseline	Local background corpus	Soft vote	Soft vote + backoff	Hard vote + backoff
Doc 1	0.22	0.22	0.22	0.30	0.30
Doc 2	0.55	0.55	0.30	0.30	0.40
Doc 3	0.33	0.22	0.44	0.40	0.30
Doc 4	0.20	0.40	0.44	0.40	0.50
Average	0.325	0.3475	0.35	0.35	0.375

Thanks!

Questions?

Suggestions?

Thoughts?