Machine Comprehension Using Robust Rule-Based Features and Multiple-Sentence Enhancing

{ Wei Chen, Danyang Wang and Xiaoshi Wang } Stanford University

Introduction

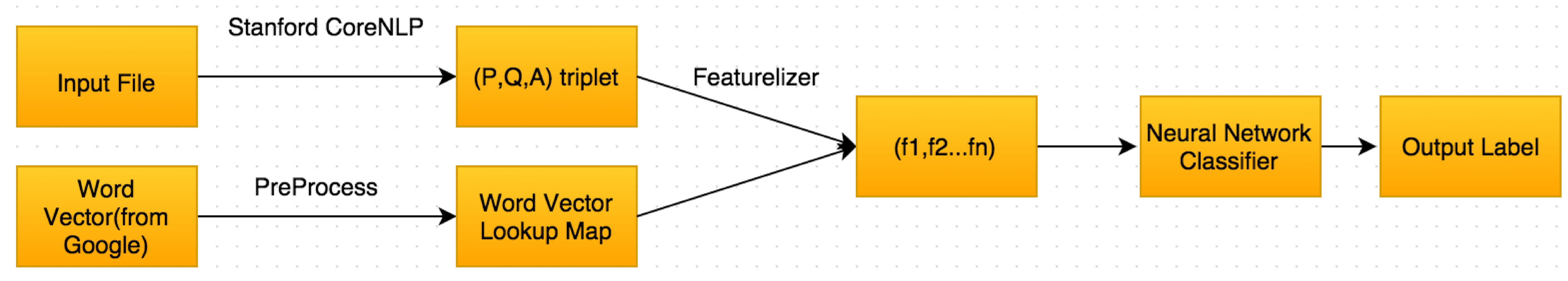
Machine comprehension is a raising research field which attracts interest from both industry and academia. Our project focuses on extracting important information from text input to answer questions based on the text. Specifically, we extract features from the passages and questions to train a neural network classifier. We use MC500, a dataset proposed by Microsoft Research, to evaluate the effectiveness of our classifier. The final accuracy it can achieve on dev set 60%.

Sample Input

Passage: ...John asked Tim if he could play on the slide. Tim said no. John was very upset and started crying. A girl named Susan saw him crying. Susan told the teacher Ms. Tammy. ...

Question: Who saw John crying and told Ms. Tammy?
A) Tim
B) Susan*
C) John
D) Ms. Tammy

Data Pipeline



Features and Model

Features

- Enhanced Sliding Window Feature:we first concatenate the question and answer to form a string s. Within a sliding window in P of size k, we count the number of word matches to s and score the answer by the maximum sliding window count. To prevent counts boosted by trivial words, we weighted the count of each word w by its inverse frequency across the passage.We used the sum of sliding window scores from size-2 to size-30.

- Distance Features: Features based on distance of key words between the question and the answer. The intuition is the part of the passage representing the answer is usually not far away from the one represents the question.

- Syntax Features: In syntax features, we represent similarities between question-answer statement and sentences in passages using dependency tree parsing. We generate question-answer statements based on rules proposed by (Wang et al., 2015). We modified their rules to obtain more reliable results.

  Q: What did he do on Tuesday?
  A: He went to school.
  Generated: He went to school on Tuesday.

- Word Embeddings: We compare cosine similarity of word-vector summation of words in question-answer pairs and word-vector summation of words in sentences from passages. We then return the largest value as a feature for each candidate answer.

- Coreference Resolution: We use the Stanford CoreNLP library to find the nouns and pronouns that refers to the same object in each passage. Then replace the nouns and pronouns with the most representative mention. From example, She->Mary,They->Tom and Jerry,etc.Then rerun the above features with the new passage text.

Model

- Shallow Neural Network: We use a one hidden layer perceptron. We set 12 node in the hidden layer, learning rate be 0.3 and momentum be 0.2.

Result

Each individual feature can capture some aspect of the problem and thus improve the classifier accuracy compared with the baseline model.

Model	Accuracy
Baseline(SW+D)	55.13%
Improved SW+D	56.5%
SW+D+Coref	55.5%
SW+D+Syntax	57%
SW+D+WordEmbed	59.5%

The classifier with all featurelizers can achieve a accuracy of 60% on the dev set. Since many of the features can synthesize information from multiple sentences across the passage, our system demonstrates even stronger performance over questions which requires clues from multiple sentences. Specifically, the overall performance on such questions in dev is 62% which outperforms the reported baseline model by 7%.

References

[1] M.Richardson, C.J.C.Burgers, and E.Renshaw *MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text*, EMNLP, 2013.

[2] H.Wang, M.Bansal, K.Gimpel, and D.McAllester *Machine Comprehension with Syntax, Frames, and Semantics*, ACL, 2015

[3] E.Smith, N.Greco, M.Bosnjak, and A.Vlachos *A Strong Lexical Matching Method for the Machine Comprehension Test* EMNLP, 2015

Future Research

1. We can explore more feature related to word-embedding. In future we can change the simple summation to weighted summation or connect word-embedding to syntax or semantic features.
2. We can add more multi-sentence features to our classifier. The intuition is, a complete answer of a question may lie in more than one sentence. We can add some features here to capture this kind of situation. Currently we only have one multi-sentence feature and we think we can explore more on this.