Process Recognition of QA system using Joint Inference

Kristen Stewart
knstewart@cs.stonybrook.edu

Kumari Shalini
kumari.shalini@stonybrook.edu

Abstract

The goal of this project is to perform semantic role labeling for a particular data set that is used for question answering. The sentences that we are working to label are used as the knowledge base for answering questions that can be found on a 4th grade science exam. Each sentence is related to a process and should contain each of the roles: *Undergoer*, *Enabler*, *Trigger*, and *Result*. We will develop a system that will perform role labeling on a group of sentences simultaneously instead on individual samples. Each process will have multiple sentences in the knowledge base that describe it. These sentences that describe the same sentence will be jointly labeled.

1 Motivation

Question answering is an important field because of the applications it has. Good question answering systems can make Internet searches more intuitive. For this particular domain, question answering can help teachers create answer sheets for their exams. This project aims to improve the overall performance of the question answering system. By focusing on this particular domain instead of general question answering we hope to get better results than applying general question answering solvers to these particular questions.

2 Related Work and Gap

There is a lot of previous work on role labeling. Much of the work done has focused on the labeling of individual sentences using supervised machine learning techniques. Others have designed unsupervised methods using clustering. These methods can get fairly good results, but as the data set we are using is particular, we think that we can do better than the more general techniques. What is unique about our data set is the similarities between sentences that need to be labeled and the small number of roles that we have where each sentence must contain each role. Each process that is in the knowledge base has multiple sentences associated with it. Each of the sentences should be similar because they describe the same process, and so each role in the sentences should be similar. Current role labeling techniques do not provide a way to take these similarities into account when labeling roles. Current SRL systems also do not label the same roles as our system does.

3 Your idea

We believe that labeling sentences associated with the same process jointly can achieve a better performance than labeling these sentences individually. The sentences in the knowledge base will be divided by which process they describe. We will run semantic role labeling on the individual sentences to obtain scores for each label for each argument span. We will also obtain similarity scores for all the argument span pairs across sentences. So each argument span will have a similarity score to all other argument spans that are not in the same sentence. We will then construct an ILP *(integer linear program)* to assign roles to each argument span. The ILP will maximize the score for each label assigned to an argument span. It will also maximize the similarity score between argument spans assigned to the same role. So for example each argument span labeled as the Trigger for a process should be as similar as possible. The ILP will also assure that no label is assigned to two argument spans within the same sentence and that each label is assigned once per sentence.

For the similarity score, we plan on using an entailment scorer that is already implemented. This however may be improved in various ways, especially since a argument span does not need to fully entail the other in order to be similar enough to be labeled with the same role. It may produce better results to instead just use WordNet hypernyms instead of full entailment. We will need to use current SRL implementations trained to recognize the unique labels for this dataset to obtain the role scores.

4 Experimental Setup

We aim to compose knowledge by explicitly searching for sentences that express roles in expected ways. For this task, we are creating dataset manually using sentences that involve processes mentioned in questions. The data will be retrieved from Google by using script which will filter all possible types of sentences involving that particular process. We use the textual entailment tool from *Allen Institute of Artificial Intelligence* to compute the similarity score. We will compare our system against the performance of the initial SRL system. If our system improves the performance then we can assume that our idea is successful.

5 Baseline Results

Results showing the performance of Semantic Role Labeling.

Method	Precision	Recall	F1
Standard	0.4323	0.3325	0.3758
Per Process	0.4225	0.2556	0.3185
Dist Supervision	0.5614	0.2642	0.3594
Dom Adaptation	0.4386	0.3351	0.3799

Table 1: SRL Performance

Results showing the performance of Question Answering system using SRL

Method	Accuracy
BOW	63.12
Manual SRL	67.38
BOW + Manual SRL	70.92
Standard	55.32
Per Process	46.80
Dom Adaptation	55.32
Dist Supervision	51.77
BOW + Standard	65.24

Table 2: QA Performance

6 Analysis

Using automatic role labeling on a single sentence showed errors arouse out of failure to identify the proper predicate. The most dominant pattern was where the predicate is a verb that is directly attached to the ROOT. In this paper, instead of using automatic role labeling techniques on an individual sentences, we will use the collection of sentences for each process and assign most accurate roles to each argument using *Integer Linear Programming*.

We aim to improve following error.

- *Entailment Issues*
  We analyzed that textual entailment was noisy and provided inconsistent scores for even some simple cases. We can improve the score by correctly identifying the roles for each argument. For this we will use Integer Linear Programming discussed below.

  Constraints in ILP :-
  1. A sentence can contain multiple or no values for a single role.
  2. Text spans are not allowed to overlap - that is, the same word or phrase cannot be used for multiple roles.

  Our *ILP Formulation* is :-

$$\sum_k \left[\sum_i \sum_j z_{ijk} \Big(\phi_{rol}\Big(a_{ij}, r_k\Big)\lambda_1 + \lambda_2 \sum_l \sum_m z_{lmk} \Big(\phi_{sim}\Big(a_{ij}, a_{lm}\Big)\Big)\right]$$

  Here *k* is all possible roles, in our case its four. Variable *i* denotes number of arguments in sencentence *j*. ϕ_{role} denotes the role classifier score. ϕ_{sim} denotes the role similarity score. Variable *l* denotes the arguments of all other sentences *m* excluding the sentence *j*.

  We will use the following procedure which explains our ILP formulation to assign the roles to each argument.

1. Use a classifier which will compute probability of each roles assigned to argument span. Suppose we have multiple sentences $S_1, S_2, ...S_M$. And each sentence contains

arguments $a_1, a_2 ... a_N$. Now based on the constraint that each sentence can have only one *undergoer* or so, we will assign label to every argument in the sentence.

2. Now we want to maximize the score for each role assigned to arguments. So, we will look across the sentences and compute following.
   - Calculate role classifier score for every argument in the sentence with each possible roles it can be classified into.
   - Then calculate role similarity score for each argument in one sentence with the argument of the same role in another sentence.
   - Last step is to maximize the summation of role classifier score and role similarity score for every argument across all possible roles.

Hence using this, we will be able to assign more accurate roles to every argument.

References

Samuel Louvan, Chetan Naik, Veronica Lynn, Ankit Arun, Niranjan Balasubramanian and Peter Clark. 2014. *Semantic Role Labeling for Process Recognition Questions* Stony Brook University, Allen Institute of AI.