# EN.601.769 Assignment 1: Semantic Role Labeling

Isabel Cachola

March 5, 2021

All code for this homework can be found here.

## 1 Data collection

**Why did you choose the semantic roles you did?** In addition to AGENT and PATIENT, I also chose EXPERIENCER, THEME, and RECIPIENT as semantic roles to model. I chose EXPERIENCER and THEME because I personally have difficulty understanding the difference between these two semantic roles, so I chose them for the project to give me the opportunity to better understand these roles. I chose RECIPIENT because thought defining this role in terms of data collection would present an interesting challenge.

**Motivate the definitions you developed for each semantic role.** I define the following semantic roles:

$$AGENT := ((volition > 0) \text{ OR } (instigation > 0)) \text{ AND } (existed > 0) \tag{1.1}$$

$$PATIENT := ((volition < 0) \text{ OR } (instigation < 0)) \text{ AND } (existed\_before > 0) \tag{1.2}$$

$$\begin{aligned} THEME := (change\_of\_location > 0) \text{ AND } (volition < 0) \\ \text{AND } (existed\_before > 0) \end{aligned} \tag{1.3}$$

$$\begin{aligned} EXPERIENCER := (change\_of\_state\_continuous > 0) \text{ AND } (volition < 0) \\ \text{AND } (awareness > 0) \end{aligned} \tag{1.4}$$

$$RECIPIENT := (change\_of\_possession > 0) \text{ AND } (existed\_before > 0) \text{ AND } (volition < 0) \tag{1.5}$$

I define PATIENT as the opposition of AGENT, except for the $existed\_before$ condition. I differentiation THEME and EXPERIENCER by defining THEME as requiring a $change\_of\_location$

| | AGENT | PATIENT | EXPERIENCER | RECIPIENT | THEME |
|---|---|---|---|---|---|
| train | 2827 | 1906 | 345 | 133 | 123 |
| dev | 380 | 245 | 43 | 14 | 22 |
| test | 373 | 219 | 30 | 10 | 15 |

Table 1: Counts of positive examples per role per split of the dataset.

but not necessarily awareness. For example, in the sentence "Mary gave Susan a cookie," the THEME is "cookie", which is not aware of its change of location. In contrast, my definition of EXPERIENCER requires a *change_of_state_continuous* and *awareness*. In the example "Mary smelled cookies," Mary is the EXPERIENCER and is aware of her change of state. For both THEME and EXPERIENCER, the $volition < 0$. RECIPIENT was the trickiest role to define. An event that involves a RECIPIENT involves something changing possession. I use the *volition* proto-role value to differentiation between the giver and the recipient.

**How often does each semantic role occur in your training sets? Why might some roles occur much more frequently or rarely than other roles?** The exact counts of positive examples in my dataset are shown in Table 1. As we can see form the table, AGENT and PATIENT are significantly more common than EXPERIENCER, RECIPIENT, THEME. This makes sense as AGENT and PATIENT have the broadest definitions, which means they are more likely to apply to more sentences. Also, per Dowty's proto-role hypothesis, semantic roles should fall on the spectrum between proto-agent and proto-patient.

# 2 Modeling

**Describe your model architecture and intuition behind your major design decisions.** The modeling architecture is a simple feed forward perception. I first embed the text of the example, then concatenate the span of embeddings between the argument and predicate, before feeding into a feed forward perceptron with a hidden dimension of 128. Concatenating the embedded span of text between the argument and predicate forces the model to give extra weight to that portion of the text. I train for 5 epochs with an Adam optimizer, and choose the best checkpoint based on the cross entropy loss on the dev set. I also trained model on both the original training datasets and balanced training sets, in which I randomly upsampled the positive labels.

**If you included additional model inputs or outputs, discuss what they are and why you used them.** Described above.

**How well do your models perform on the test sets? Is performance correlated with the frequency of the role in the training set?** As shown in Table 2, my model performed very well on the AGENT semantic role, but not well on the other roles. In particular, the more rare labels did not perform well. We can see that balancing the training data improved performance on PATIENT, THEME, and RECIPIENT, but did not make a difference for EXPERIENCER. However, for the rare SRLs, there were so few examples in the test set that it makes sense the model would have difficulty doing well on the test sets.

| | AGENT | PATIENT | THEME | EXPERIENCER | RECIPIENT |
|---|---|---|---|---|---|
| Original | 0.760248 | 0.163569 | 0.000000 | 0.0 | 0.0000 |
| Balanced | 0.733427 | 0.378223 | 0.166667 | 0.0 | 0.0625 |

Table 2: F-1 scores for each semantic role, trained on either the original dataset or the balanced dataset.

| | AGENT | PATIENT | THEME | EXPERIENCER | RECIPIENT |
|---|---|---|---|---|---|
| Mary ran down the street . | 0 | 1 | 1 | 0 | 1 |
| Mary was chased by a dog . | 0 | 0 | 1 | 0 | 1 |
| Mary gave John a dog . | 0 | 0 | 1 | 0 | 1 |
| Mary smelled her dog . | 0 | 0 | 1 | 0 | 1 |
| Mary gave John a dog . | 0 | 0 | 1 | 0 | 1 |

Table 3: Predictions for example sentences from models trained on balanced data

# 3 Exploration

**Examine the training data you collected for one of the roles. Do any of the coarse semantic role labels seem counter intuitive or wrong? How would you modify your definitions to better capture the semantic roles? Can any set of UDS criteria perfectly capture the characteristics of a semantic role?** Examining the RECIPIENT role more closely, the most common error I found is my definition tagging an argument as RECIPIENT when it is more accurately the THEME. For example, in the sentence "I sent[PRED] this[ARG] to Mary last week", my definition tags the argument "this" as a RECIPIENT, when it is more accurately the THEME and Mary is the RECIPIENT. The problem with using the proto-role values in my definitions is that both THEME and RECIPIENT are closer to a proto-patient, and therefore harder to differentiate. A more accurate definition would likely need to involve using syntax, in addition to the proto-role values. Additionally, there were examples where $change\_of\_possession > 0$ but confidence was low so the sentence was incorrectly given a positive label. For example, the sentence "I have[PRED] a 1 1 / 2 year old female calico[ARG] ." had a $change\_of\_possession > 0$ with $confidence = 0.201$. For more accurate tagging, it would be better to take into account confidence. Finally, I don't think it's possible to perfectly capture a semantic role using a set of UDS criteria.

**Write a minimal pair of sentences that use similar words but differ in the expected semantic roles. Do your models correctly predict the difference?** I chose the following sentences as examples:

1. AGENT: Mary[ARG] ran[PRED] down the street .
2. PATIENT: Mary[ARG] was[PRED] chased by a dog .
3. THEME: Mary gave[PRED] John a dog[ARG] .
4. EXPERIENCER: Mary[ARG] smelled[PRED] her dog .
5. RECIPIENT: Mary gave[PRED] John[ARG] a dog .

The models trained on unbalanced data predicted negative for all examples for all labels, which

indicates a negative bias. As shown in Table 3, the models trained on the balanced dataset had slightly more interesting results. From the results, the only examples with a their labels correctly prediction are examples (3) and (5), however it is clear from the results that these models have a positive label bias.

# 4 Extension

**After building a separate model for each semantic role, you might wonder whether the roles are independent. What are two ways by which you could determine if any of the semantic roles you chose are correlated?** Given a labeled dataset, I would look at the frequency at which two semantic roles occur in the same example. Another method to test the independence of semantic roles is to create a single model, that predicts multiple labels rather than multiple models that each predict a single label. The assumption here being that if the roles are correlated, the model will improve performance by having access to the information from the other labels.

**It seems wasteful to build a separate classifier for each semantic role. Besides saving memory, why would a multi-class or multi-label classifier be a reasonable model for this task?** As mentioned above, it is unlikely the roles are completely independent. It is likely that a multiclass classifier would perform better, by virtue of having access to information across labels.

**Would a bag-of-words model using type-level word embeddings (e.g., GloVe embeddings) do well on this task? Why or why not?** I don't think it would because a bag of words model does not take into account the different senses of a word, which is *very* important for semantic role labeling. As I showed in the examples above, the same word can play different semantic roles, depending on the context of the sentence.

**The models you developed included (at a minimum) only text and predicate/argument indicator in-formation. What are two other linguistically motivated inputs that may be useful for the task? Why would you expect them to help?** In addition to the inputs described above, I would expect encoding syntactic information would help improve the performance, as the combination of words and syntax often give clue to the semantic role of a word. Additionally, I would expect using more modern, context embeddings (such as ELMo or BERT) would improve the performance, as these embedding take into account the context of a word, which as mentioned above, is important to the semantic role the word plays.