Pragmatically Informative Frame Identification

1,2Xudong Hong, 1Frances Yung
1Saarland University, 2Max Planck Institute for Informatics
xhong@coli.uni-saarland.de

Frame identification is a semantic role labelling task based on FrameNet (FN) [1]. Given the context around a predicate (e.g., ‘the woman is preparing a meal.’), the task is to assign a frame (here, ‘Cooking_creation’) which is a label for a prototypical situation. State-of-the-art frame identification systems trained on FN dataset obtain strong performance on in-domain text [2] but the coverage of the current FN lexicon is very limited. Performance of frame identification systems drops significantly when an ambiguous predicate is not in the FN lexicon [3] or not in the domain of FN, e.g. image captions.

To leverage this problem, we consider a frame identifier trained on a balanced corpus like FN dataset as a model of general semantic meanings to understand a predicate. When an out-of-domain context is given, humans cannot fully understand the predicate purely based on semantic information and require pragmatic inference to disambiguate the predicate. The recent proposals claim that this inference process can be modelled with recursive Bayesian inferences [4, 5]. We hypothesize that Bayesian pragmatic inference can improve out-of-domain frame identification.

We model the frame identification task as a communication game with two rational agents using Rational Speech Acts framework [6, 7]. The speaker describes a situation with a predicate given the context. The listener interprets the sentence and chooses a frame for the predicate. We consider the frame identifier [8] trained on FN 1.5 as the literal listener (L0): \( P_{L0}(f | p) \) which predicts a frame \( f \) given a predicate \( p \). The pragmatic speaker (S1) produce a predicate given a frame with a prior \( P_{S1}(p) \) in the speaker’s domain: 

\[
P_{S1}(p | f) = P_{S1}(p)P_{L0}(f | p) / \sum_{p \in P_{S1}} P_{S1}(p)P_{L0}(f | p)
\]

The pragmatic listener (L1) reasons about the predicate that the pragmatic speaker would generate and infer the most possible frame given this predicate: 

\[
P_{L1}(f | p) \sim P_{S1}(p | f)P_{S1}(f)
\]

We then annotate 126 verbal predicates in MS COCO captions [9] that are not in FN as a test set. Finally, we test our models by performing zero-shot frame identification. Results show that the pragmatic listener (accuracy=6.72%) outperforms the literal listener (accuracy=5.47%). We believe that our method can be generalised to other out-of-domain language understanding tasks.