Class project: Amortized RSA

Justin T. Chiu
Department of Computer Science
Cornell Tech
jtc257@cornell.edu

Abstract

The Rational Speech Acts inference procedure has been demonstrated to be effective as a model of human behaviour in small reference games. However, the inference procedure is expensive as it requires considering all possible utterances. White et al. (2020) amortize the cost of inference by training a language model perform to maximize the objective the RSA procedure optimizes. The resulting amortized model displays symptomps of language drift, where in optimizing the RSA objective produced utterances are unhuman-like. We build upon their amortization procedure by proposing a change to the objective that prevents language drift.

1 Introduction

The Rational Speech Acts (RSA) framework is an inference procedure based on Gricean maxims for generating and understanding language in a collaborative setting. RSA has had empirical success in small domains, in particular simple reference games with very limited vocabularies. RSA is potentially useful beyond these domains as a heuristic for utility maximization. Extensions of RSA to nontrivial vocabularies requires approximations: incremental RSA (Cohn-Gordon et al., 2018), reranking, and/or variational inference (White et al., 2020).

We focus on amortization for approximating RSA speakers, which has had empirical success in applications such as variational autoencoders and reinforcement learning. We highlight two broad classes of errors in text generation, and examine their effects on the SHAPEWORLD dataset: model error and search error.

2 Related Work

The approach of White et al. (2020) is the basis of our project, and the apply the amortization approach to both the SHAPEWORLD and COLORS dataset. Unfortunately we only have time to fit in the SHAPEWORLD dataset, on which we find a small contradiction to one of claims: we find that RSA does not help with task performance.

A similar amortization approach can also be found in Gu et al. (2017), which attempt to train a greedy decoder to mirror the performance of beam search in a teacher model for machine translation.

Amortization also shows up in variational autoencoders, and one could interpret all parameterized policies trained through via policy gradient as instances of amortization as well, where the alternative would be to explicitly perform policy iteration.

Incremental RSA (Cohn-Gordon et al., 2018) is another approach for approximating the expensive inference procedure in RSA, although that methods has no principled interpretation.

Approximations for RSA listeners, rather than speakers, have been explored as well. Other works use a margin-based approximation of pseudolikelihood (Gulordava et al., 2020), and learn through approximate RSA procedure via importance sampling over utterances (McDowell and Goodman, 2019).

3 Problem Setup

The SHAPEWORLD dataset consists of games where two players, a speaker and a listener, collaborate to identify a target object out of a given set. Given a set of three objects, $o = \{o_1, o_2, o_3\}$, the goal of the speaker is to accurately describe the object at index $t \in [3]$, to the listener through a single utterance $u = \langle u_1, \dots, u_L \rangle$. Utterances are limited to a fixed vocabulary of size 14 consisting of colors and shapes. The data generating process only uses at most $L \leq 2$ utterances, consisting of a color and/or shape. The literal speaker is a distribution over utterances $s_0 = p(u \mid t, o)$, while the literal listener is a distribution over targets $l_0 = p(t \mid u, o)$.

4 Method

We are interested in decoding an utterance u that refers to target t by optimizing the following program:

$$\operatorname*{argmax}_u F(u, t, o), \tag{1}$$

where F is only able to score full utterances u. This prevents efficient inference, which would require that F decomposes over prefixes of u from left to right. Such a decomposition would allow beam search, or other approximate inference techniques. Instead, we must optimize F with the following procedure:

1. Enumerate all possible utterances u
2. Score all utterances according to $F(u, t, o)$
3. Choose the highest scoring utterance.

When the number of possible utterances is infinite or exponentially large in the length of the longest allowed utterance, this procedure is infeasible.

Rather than decoding by enumerating all possible sequence for each context, White et al. (2020) propose to amortize the cost of inference. This is accomplished by training a left-to-right model to approximate the argmax of F by optimizing the following program:

$$\operatorname*{argmax}_\theta \mathbb{E}_{p(t,o)} \left[\mathbb{E}_{q_\theta(u|t,o)} \left[F(u, t, o)\right]\right], \tag{2}$$

where $q_\theta(u \mid t, o)$ is the amortized speaker. This objective can be optimized via stochastic gradient descent using the score function gradient estimator, or a biased gradient estimator in the Gumbel-softmax family of estimators.

White et al. (2020) explore a single objective, given by

$$F_{\text{length}}(t, u, o) = \log l_0(t \mid u, o) + \log \lambda|u|. \tag{3}$$

This corresponds to MAP inference in a noisy channel model with a relatively uninformative prior that is constant given a length. The hyperparameter λ controls how much influence the prior has on the objective. We refer to this objective as length.

We explore two more objectives, the first of which uses an informative prior in MAP inference:

$$F_{\text{MAP}}(t, u, o) = \log l_0(t \mid u, o) + \log p(u \mid o), \tag{4}$$

where $p(u \mid o)$ is a speaker that does not condition on the target t. We refer to this objective as MAP.

The last objective introduces a slightly larger change. Rather than approximate MAP inference, the goal is to fully approximate the posterior of the following noisy channel model

$$p(u \mid t, o) = \frac{p(t \mid u, o)p(u \mid o)}{p(t \mid o)}.$$

This is accomplished by minimizing the KL divergence between the posterior and the amortized approximation as follows:

$$\begin{aligned}
&\min_\theta D_{\text{KL}} \left[q_\theta(u \mid t) \mid\mid p(u \mid t)\right] \\
&= \min_\theta \mathbb{E}_{q_\theta(u)} \left[\log q_\theta(u \mid t) - \log p(u \mid t)\right] \\
&= \min_\theta \mathbb{E}_{q_\theta(u)} \left[\log q_\theta(u \mid t) - \log p(t \mid u)p(u)\right],
\end{aligned} \tag{5}$$

where conditioning on the full context o has been dropped for brevity, and the denominator $p(t \mid o)$ from the posterior is constant wrt θ and can be safely ignored. We refer to this objective as Bayes.

The difference between the MAP and Bayes objectives is the addition of the negative entropy term, $\mathbb{E}_q\left[\log q\right]$, in Equation 5. This entropy regularization encourages q to not reduce to a point mass, unlike the other objectives which have no such pressure.

5 Experiments

We evaluate amortized speakers obtained by optimizing the three objectives: length, MAP, and Bayes, on the SHAPEWORLD dataset. We compare the three amortized speakers against each other, as well as against the base literal speaker, on three metrics: how accurately a collection of literal listeners trained on separate data can infer the correct target, the likelihood of the produced utterances under a marginal language model, and the length of the produced utterances. The last two metrics, likelihood and length, are strongly correlated in the

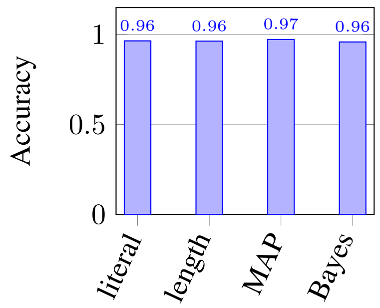

Figure 1: The speaker accuracies using an ensemble of listeners trained on held out data for SHAPEWORLD. The literal speaker was trained directly via MLE, while the amortized length, MAP, and Bayes speakers were trained to optimize their eponymous objectives using the straight-through Gumbel-softmax estimator.

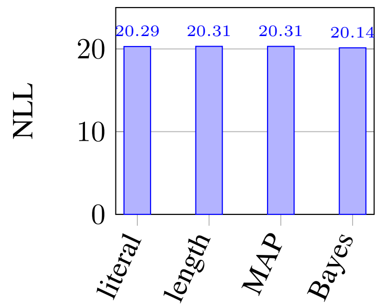

Figure 2: The average negative log likelihood (NLL) of utterances produced from each speaker in SHAPEWORLD as measured by a marginal language model $p(u)$ that does not see the objects or target. The lower the NLL, the better.

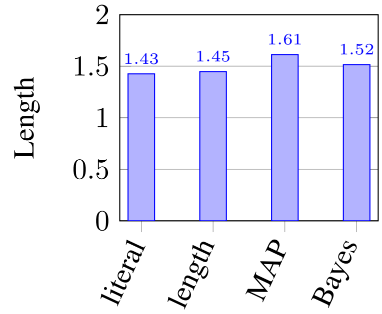

Figure 3: The average utterance lengths from each speaker using an ensemble of listeners trained on held out data for SHAPEWORLD.

case of SHAPEWORLD, where there is no strong preference for particular utterances due to the small size of the vocabulary.

Literal listeners and speakers are trained on the same splits of randomly sampled SHAPEWORLD contexts as in (White et al., 2020). The speaker $p(u \mid o)$ that was used to train the amortized speakers and does not conditional on targets was trained on the same training splits. Amortized speakers are trained via stochastic gradient descent using the straight-through Gumbel-softmax gradient estimator.

Ten test literal listeners were trained on held-out generated data and used for evaluation.

6 Results

Before discussing speaker performance, we find that the base literal listener is quite accurate, obtaining 96.94% accuracy on the ground truth test data. This is expected, as the data in SHAPEWORLD is very simple and artificially generated.

As for speaker performance, we find that all speakers are very accurate and obtain at least 96% accuracy when evaluated with test literal listeners trained on held-out data, as shown in Figure 1. The MAP speaker outperformed the other variants very slightly. We additionally outperform the speakers from White et al. (2020), where the literal speaker obtained 73% accuracy.

The likehoods of the utterances produced by the speaker also is very close, as shown in Figure 2. Interestingly, the Bayes speaker has slightly better likelihood. This is surprising, as there is little difference between the MAP and Bayes objectives. It is likely that this is not statistically significant.

The average lengths of the utterances of the speakers is given in Figure 3. The MAP utterances are on average longest. This may account for the higher accurace of the MAP speaker, which could be producing longer utterances resulting in less ambiguity. Since the MAP objective has an informative utterance prior, this leads to less pressure to produce shorter utterances.

The observation that the MAP speaker has higher accuracy as well as longer utterances implies that the literal speaker underestimates the listener, and occasionally does not provide enough information in an utterance. This could happen in a scenario where both the color and shape is necessary to successfully communicate a shape, but only one of the two qualities is uttered.

7 Discussion

Unfortunately, the SHAPEWORLD dataset itself is too simple to observe language drift, leaving our study inconclusive as to whether the proposed method fixes that particular issue. However, we may return shortly with results on the COLORS dataset, which did exhibit speakers with language drift.

Additionally, it noteworthy that application of the straight through Gumbel-softmax estimator may be leading to extremely unstable learning dynamics. Although the objective improves for the first few epochs, performance on the training objective eventually plummets.

References

Reuben Cohn-Gordon, Noah Goodman, and Christopher Potts. 2018. Pragmatically informative image captioning with character-level inference.

Jiatao Gu, Kyunghyun Cho, and Victor O. K. Li. 2017. Trainable greedy decoding for neural machine translation.

Kristina Gulordava, Thomas Brochhagen, and Gemma Boleda. 2020. Which one is the dax? achieving mutual exclusivity with neural networks.

Bill McDowell and Noah Goodman. 2019. Learning from omission. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 619–628, Florence, Italy. Association for Computational Linguistics.

Julia White, Jesse Mu, and Noah D. Goodman. 2020. Learning to refer informatively by amortizing pragmatic reasoning.