Evolutionary Language Development in Multi-Agent Cooperative Learning Games

Jayashree Raman
UC Berkeley
MIDS Student
jayashreeraman@berkeley.edu

Divya Sriram
UC Berkeley
MIDS Student
divyasriram@berkeley.edu

Eric Tsai
UC Berkeley
MIDS Student
pizzic@berkeley.edu

Abstract

Lazaridou., et al 2017 proposed a framework for language learning that relies on multi-agent communication. The agents in the framework were setup in a referential game where they communicated about many images. In this paper, we propose an experiment where agents develop a private language for referring to specified sentences given a set of sentences. The challenge is for the agents to learn a method of distinguishing differences between sentences and to develop a shared language to be able to refer to particular sentences by those distinguishing features. We will evaluate the agents' ability to accurately identify and differentiate the sentences. In addition, we will identify patterns in the methods that the agents develop to refer to the different types of sentences.

Keywords: Reinforcement learning, multi-agent coordination

1 Introduction

A fascinating goal of AI is the creation of agents that can communicate with each other to fulfill a certain defined purpose. Though conversational agents could be trained through supervised learning, multi-agent coordination models allow the model to learn functional features of communication, like those of using words to convey meaning and arouse action (Lazaridou, 2017).

The multi-agent models have been implemented in a variety of ways: in the context of traditional game theory, to perform simple tasks of distinguishing between image inputs, achieving optimal outcomes in negotiations, and coordinating actions.

In general, a unifying goal of a multi-agent system is effectively designing the system to use the provided environment so agents can solve a specifically defined problem through a rewards based implementation. The key to a multi-agent model is the sharing of knowledge in some form between the two agents within the prescribed constraints of the communication protocol (Kottur, 2017).

The challenge is to create an environment where agents are incentivized to develop a language. The long-term goal is for the developed language to be portable to new contexts and new communication entities (particularly with humans).

Our goal was to create a system where one agent, the sending agent, must communicate information about a sentence to the second agent, the receiving agent, so the agents can successfully complete a referential game. In a referential game, the agents are presented with the same set of objects and must use language to refer to one of them in the manner that they can understand which object is being referred to.

Communication in this paper refers to a the output of the sending agent, a one-hot message vector. The message vector, along with the actual sentences, is taken as an input for the receiving agent. The size of the message vector is a parameter, and it defines the number of unique messages that the agents can send and represents the size of their private vocabulary.

The importance of the experiment is to show how a goal-oriented approach to information sharing can be evolved through machine learning. In the experiment, the agents developed a method to distinguish between two sentences and a way to communicate that distinction with other agents.

2 Background

One of the problems with performing language generation from language models trained on textual corpuses and dialog models is that the models lack control over the meaning of the generated text. Often times, the model can simulate the style

of a text (the grammar and wording) but has no control over the meaning of the generated text.

A recent approach toward natural language generation is in multi-agent coordination communication games, where multiple agents must communicate with each other to accomplish a shared goal. With this goal-oriented approach to language generation, the agents are motivated to use communication to relay information to the other agents to accomplish their goal. The agents use an evolutionary approach to developing a shared language to coordinate their actions. In this way, the agents take a ground-up approach to language development where they begin with information to share and must negotiate a language in which to pass that information to other agents.

The referential game is a version of the signaling game (Lewis, 1969). These games have been the subject of extensive studies and have been referred to as "cheap talk" as a framework for understanding the evolution of language (Crawford, 1998; Blume et al., 1998; Crawford and Soebel, 1982). The focus of the studies is whether the language that develops is specific, vague, or nonexistent. Lazaridou showed that in a multi-agent setting, constraining the vocabulary size of the evolved language can control how precise of a language develops.

3 Model Framework

We parallel the multi-agent framework in Lazaridou et al 2017:

- Two agents
- A specific set of tasks that each agent must perform
- Communication protocol allowing the two agents to communicate
- A reward or payoff system aligned with the defined objective of the set up

Specifically, the multi-agent game was designed with the following features:

1. A set of sentences, depicted by the vectors $s_1, s_2, s_3, ...s_N$ will be created where two sentences are randomly drawn, say (s_T, s_W) and one of these two sentences will be identified at the target sentence. Target t such that t exists in T, W

2. Each of the two agents will be classified as either the sender or the receiver. The pair of sentences (s_T, s_W), will be fed into both the sender and the receiver. However, the senders input will also include the tag for the target sentence (s_T, s_W, t)

3. A vocabulary V of size M will be established, allowing one symbol to be sent from the sender to the receiver

4. The sender chooses one symbol from V to send the receiver. This is called the sender's policy s($\theta_S(s_L, s_R, t)) \in V$

5. The target sentence remains unknown to the receiver. The receiver will attempt to determine the target sentence using the symbol sent by the sender. This is the receiver's policy r($s_L, s_R, \theta_S(s_L, s_R, t))) \in \{L, R\}$

6. If the receiver correctly identifies the target sentence, the payoff will be attributed to both agents, allowing a win for the game. This is represented as r($s_L, s_R, \theta_S(s_L, s_R, t)))$) = t

4 Data

As noted in Lazaridou et al 2017, because the referential game requires pairs of sentences for training, a large dataset can be generated from a relatively small corpus. The number of sentence pairs that can be generated from a corpus with s number of sentences is: $s(s-1)/2$

For our investigation, we will be randomly drawing pairs of sentences from the Brown corpus, with one of the sentences randomly assigned as the target sentence. The Brown corpus contains 57,340 sentences, so the total number of possible sentence pairings is 1.644 billion. For the project, a dataset of 1 million sentence pairs was drawn from the corpus.

To create target/distractor pairs for training and testing, we randomly sample two categories then randomly select a sentence from each of those categories, then randomly select of the sentences to serve as the target. The goal was to allow the difference in categories to distinguish the sentences as one possible method for the agents to distinguish the sentences.

Because our model has a fixed-size word window, the sentences in the dataset were truncated and padded to a length of 30 words.

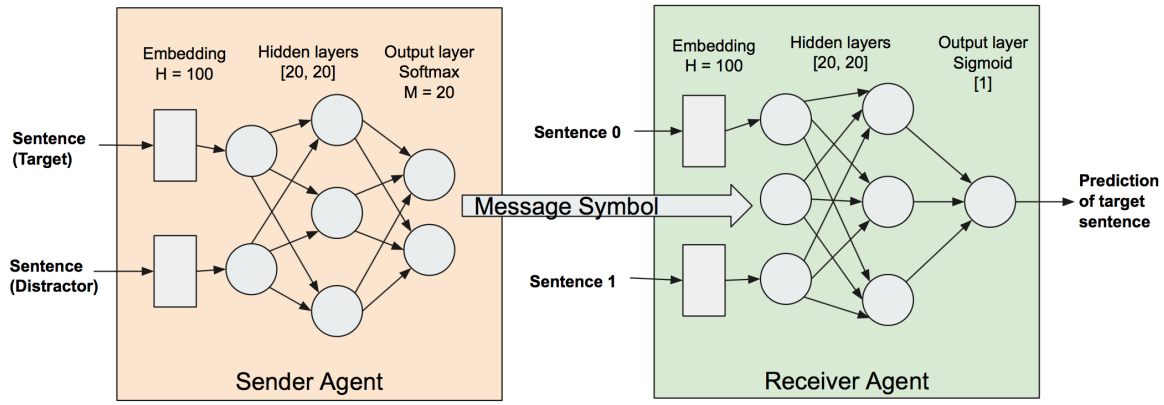


Figure 1: *Diagram of the model architecture.*

To prevent the agents from using sentence position as the inter-agent message, the receiving agent receives a shuffled version of the dataset where half of the sentences pairs have the order swapped. A binary vector indicating which of the sentence pairs had their sentences swapped acts as the labels to use for scoring the output of the receiving agent.

5 Experimental Setup

Figure 1 shows an architecture diagram of the model. The sending agent receives both sentences and passes the words of both sentences through an embedding layer. The embeddings are passed into a feed forward neural net and the output from the neural net is converted into a vector the size of the agent's vocabulary with a softmax.

The receiving agent receives the same two sentences, possibly in a different order than the sending agent did, and also receives the message from the sending agent. It also passes the sentences through an embedding layer, then feeds the word vector and the inter-agent message into a feed forward neural network. The receiving agent's output is a single sigmoid value, 0 or 1, as to whether sentence 0 or sentence 1 is the target sentence.

The agent players are both feed forward neural networks. Lazaridou's paper experiments with two architectures for the sender - the agnostic and the informed sender. In our case, since we are using sentences instead of images, we've to model with the agnostic sender.

6 Results

Mean error	0.0117954
Mean prediction for y = 0	0.0135364
Mean prediction for y = 1	0.9899450
Error percentage	0.8195%

Table 1: *Predictions after training. The error percentage is calculated based on the rounded prediction value.*

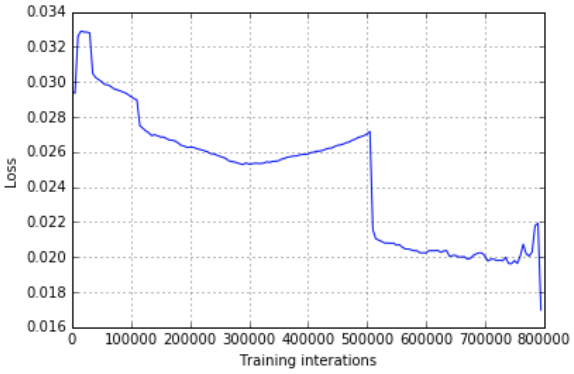
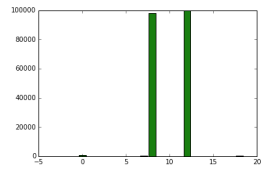


Figure 2: *Loss over training run*

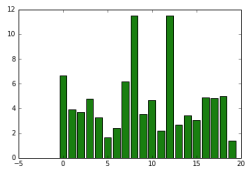
After training the agents with 8 million sentence pairs, the agents are able to select the correct sentence at a rate significantly higher than chance. The predictions of the target sentence on the test set are shown in Table 1.

The mean error of the prediction is 0.0118, a result that significantly improves upon chance. The mean prediction when the expected label is zero is 0.0135 and is 0.990 when the expected label is one. When the prediction value is rounded, the error rate is 0.82%, demonstrating that the trained model is able to predict the correct sentence for a very significant portion of the test examples.

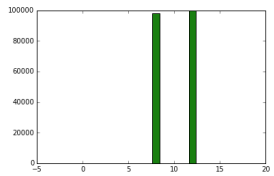
Figure 2 shows how the training loss changed



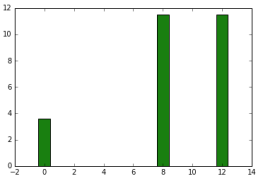
(a) *Sum of the message vectors.*



(b) *Log-sum of the message vectors.*



(c) *Sum of the rounded message vectors.*



(d) *Log-sum of the rounded message vectors.*

Figure 3: *Inter-agent message usage. The rounded messages captures the dominant index in the message. Note, some messages do not have a value that exceeded 0.5.*

over ten epochs of the 800,000 example training dataset.

To understand the basis that the agents were using to distinguish between sentences, we analyzed the inter-agent messages used. Figure 3 shows the messages that were used by the agents for the test set. Two messages were predominantly used by the agents. In the log-sum chart (figure 3b, we observe that the other indices in the vocabulary are used far less frequently. By rounding the message vectors, only values greater and equal to 0.5 remain. We observe that the same two indices dominate, but the log-sum shows that a third index does remain.

We had an intuition that the two dominant messages referred to whether the target sentence was longer or shorter than the distractor sentence. Therefore, we split the dataset by whether the target sentence is longer, shorter, or of equal length to the distractor sentence (table 2). One of the dominant messages is used primarily when the target sentence is longer, and the other dominant message is used primarily when the target sentence is shorter. This suggests that the agents learned to ignore the padding used to make the sentences equal lengths. The third message is used primarily when the two sentences have the same length. However when the messages are the same length, the two dominant messages are still predominantly used.

When the sentences are the same length, the error rate is greater than when the sentences are

Message	Longer	Equal	Shorter
0	1	32	4
8	595	6635	90,879
12	91,341	7468	920
Mean error	0.00524	0.09196	0.00531
Error percent	0.325%	6.670%	0.333%

Table 2: *Analysis of the inter-agent messages. The message counts are based on rounding the message vector to observe the dominant index for each message. There exist cases where no index was greater than 0.5 in the message vector. The error percentage is calculated based on the rounded prediction value.*

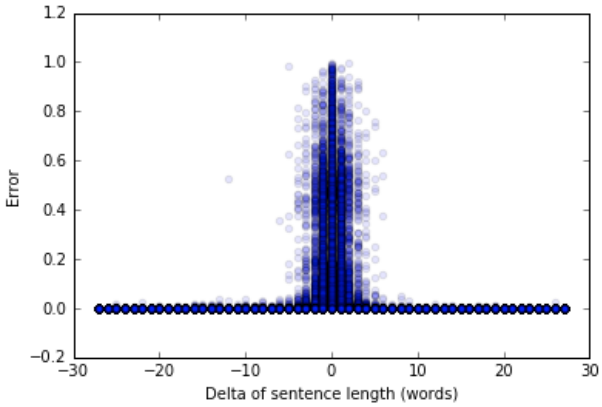


Figure 4: *Error over delta of sentence length in words: length(target) - length(distractor)*

different lengths. However, the error rate still remains significantly lower than chance, so the agents are using some unknown metric to distinguish between those sentences. In figure 4, we can see that the error only occurs when the absolute value of the sentence length difference is small.

Examination of the sentence pairs that the agents failed on was done to determine whether there was any obvious pattern to them. The only pattern that was observed is that the lengths of the sentences is similar (after truncation). The mean difference in sentence length for pairs that the agents failed and are unequal length is 4.6% (See the appendix for a sample of the sentence pairs where the agents failed.) This supports the assertion that the agents developed some metric for scoring the sentences that is correlated with– though not solely reliant upon– the sentence length.

7 Discussion and Next Steps

The results support the promise that feed forward cooperative learning networks can train agents to develop inter-agent communication to achieve a goal. The agents were able to correctly predict the target sentence with an accuracy of 99.2%. The binary nature of the messages used suggests that the agents developed a metric to score the sentences and the two dominant messages used indicate whether the target sentence had a higher or lower score on the metric.

Though these initial results of the model are promising, we recognize that there several avenues that we would like to expand upon for future research.

- Reinforcement learning

  The message vector that sender agent passes only approximates a one-hot vector. As is, the message vector *conveys* extra information through the partial values of the vector. This information leakage between the agents does not hold true in real life and is not natural. In a future model, the sender would emit a true one-hot vector message.

  Lazaridou's setup is modeled with reinforcement learning. Using a reinforcement learning model, we could solve the one-hot message issue while continuing to be able to perform backwards propagation for training.

- Other datasets

  It would be useful to test the model on different datasets, such as one of the paraphrase datasets (PPDB and Paralex). The results from training with the Brown corpus suggests that segregating the sentence pairs by category may not have provided assistance on the performance of the agents. Running the model on a paraphrase corpus would help to verify whether topic or style is relevant to the method that the agents used distinguish the sentences. It would also be instructive to test the model with sentence pairs that differ by controlled amounts to measure what types of changes define the decision boundary of the trained model.

- Constrain the sender to use more of the available vocabulary

  Currently, the model appears to be using some scoring metric to score the sentences then does a comparison of the scores to distinguish between the sentence pairs. We suspect that the messages passed between the agents indicate whether the target sentence is has the higher or lower score in the agent's metric. It would be interesting to find a method to constrain the agents so that they have to use a more diverse set of messages to win the referential game. One possible method is truncating each pair of sentences so they are the same length. Another possibility would be to alternate performing supervised learning on the sender alone and training both agents together, as per Lazaridou., et al 2017. The sender training would teach the sender to use more of the available vocabulary.

References

Andreas Blume, Douglas V DeJong, Yong-Gwan Kim, and Geoffrey B Sprinkle. 1998. *Experimental evidence on the evolution of meaning of messages in sender-receiver games*. The American Economic Review, 88(5):13231340

Vincent Crawford. 1998. *A survey of experiments on communication via cheap talk*. Journal of Economic theory, 78(2):286298

Vincent Crawford and Joel Sobel. 1982. *Strategic information transmission*. Econometrica: Journal of the Econometric Society, pp. 14311451

Satwik Kottur, Jos M.F. Moura, Stefan Lee, Dhruv Batra. 2017. *Natural Language Does Not Emerge Naturally in Multi-Agent Dialog*. arXiv:1706.08502v3 [cs.CL].

Angelika Lazaridou, Alexander Peysakhovich, Marco Baroni 2017. *Multi-Agent Cooperation and the Emergence of (Natural) Language* arXiv:1612.07182v2 [cs.CL]

David Lewis. 1969. *Convention*. Harvard University Press, Cambridge, MA, 1969.

Appendices

A Examples of failed sentence pairs

Table 3 contains a sample of the sentence pairs that the agents failed to predict correctly. Each sentence is preceded by the length of the sentence.

Table 3: *A selection of sentences pairs that the agents failed to correctly predict. Each sentence is preceded by the length of the sentence.* ***Note:*** *The sentences and sentence lengths presented in table 3 are presented prior to being truncated to 28 words (30, including the <s> and </s> tags).*

19 : He did not , as far as I can gather , find the South " worse " ; ;	18 : Notice that this man had a threefold conception of God which is the secret of his faith .
37 : This time B'dikkat smiled pleasantly at the little head which had grown out of Mercer's thigh – a sleeping child's head , covered with light hair on top and with dainty eyebrows over the resting eyes .	33 : The board of suspension of the Interstate Commerce commission has ordered a group of railroads not to reduce their freight rates on grain , as they had planned to do this month .
40 : In the early days of this controversy over the theater one of the interested parties , Stephen Gosson , published a little tract in which he objected mildly to the abuses of art , rather than the art itself .	40 : And although there was plenty of vigor in the performance , the ensemble was at its best when the playing was soft and lyrical , yet full of the suppressed tension that is one of the hallmarks of Beethoven .
29 : Mr. Ailey's " Roots Of The Blues " , an earthy and very human modern dance work , provided strong contrast to the ballet selections of the evening .	35 : The soiled fabrics used for rapid testing of detergent formulations are made in such a way that only part of the soil is removed by even the best detergent formulation in a single wash .
40 : A long evolution in an oral tradition caused the poetic language of the heroic age to be based upon formulas that show the important qualities of things , and these formulas are therefore potentially rather than always actually accurate .	31 : Adjusted sales that month were up a relatively steep 2.5% from those of the month before , which in turn were slightly higher than the January low of $17.8 billion .
19 : I want you to find Monsieur Prieur at once and give him this money for the boy's purchase .	19 : In Newark , for example , this gain was put at 26 per cent above the year-earlier level .
36 : Kerr , who set the world record earlier this month in New York with a clocking of 1.09.3 , wiped out Mills's early pace and beat the young Big 10 quarter-mile king by 5 yards .	30 : Our comment was that this was " featherbedding " in its ultimate form and that sympathy for the railroad was misplaced since it had entered into such an agreement .
10 : " You haven't dressed for the occasion " ! !	10 : Moreover , even getting this across would be difficult .
27 : It may be fostered by frustration , depression , insecurity – or , in children , simply by the desire to stop an anxious mother's nagging .	39 : Just as in the case of every prodigy child , we must watch for the efficacy of my teaching to show up in the future – if he should master all the strenuous exercises I inflicted on him .
20 : Newcomers are Ernie Kemm on piano , Wes Robbins , bass and trumpet , and Jack Kelly on drums .	25 : She showed her surprise by tightening the reins and moving the gelding around so that she could get a better look at his face .
27 : They seemed then to have had a single mind and body , a mutuality which had been accepted with the fact of their youth , casually .	25 : while the Yin , or female principle , flourished in darkness , cold , and quiet inactivity , and was associated with the Moon .
11 : And did he appreciate my efforts on his behalf ? ?	11 : Why do we let the Germans do this " ? ?
16 : the presence of other members of similar social and economic level is the sufficient condition .	16 : Elaine St. Johns may fly in from the West Coast for the editorial staff meetings .
5 : The hope was vain .	5 : A voice spoke near-at-hand .