Emergent Quantized Communication

Boaz Carmeli, Ron Meir, Yonatan Belinkov,

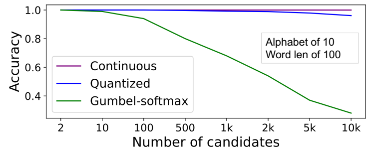
The Technion – Israel Institute of Technology

The Problem

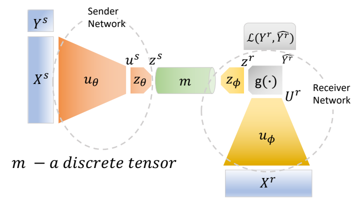
- Communication emerges from letting artificial agents solving tasks
  - Assess its characteristics
- Constraining the channel
  - Discrete communication
- NN need to be continuous for gradients to flow
  - Reinforcement learning
  - Gumbel softmax
- Quantization as a superior discretization approach
  - Can be optimized end-to-end
  - Resulted in good performance

Key Results

Quantization performance are comparable to continuous
Gumbel degrades with increased number of distractors



Network Architecture



Quantized Communication

The Channel

	Continuous	Gumbel	Quantized
Symbol	Floating point	1-hot	Integer
Word	0.10 | -0.23 | 0.69	0 | 0 | 1	4.0 | 1.0 | 7.0
Message	0.10 | -0.23 | 0.69	0 | 0 | 1	4.0 | 2.0 | 7.0
	0.72 | -0.54 | -0.21	1 | 0 | 0	7.0 | 1.0 | 2.0

Instant Channel: Single-word message
Recurrent channel: Multiple-words message

Baselines

Continuous Communication:
- Good performance with sufficient capacity
- Enables end-to-end training with back-propagation

Gumbel-softmax:
- A continuous approximation for a categorical distribution
- Discrete messages are approximated via sampling procedure

The Algorithm

Algorithm 1: Quantizing continuous communication.

```
msg: continuous message                        ▷ Floating point vector
S: scaling factor                              ▷ Set the alphabet range
procedure NORMALIZE(msg)
    min_elem ← min(msg.elements)
    max_elem ← max(msg.elements)
    msg ← (msg.elems − min_elem)/max_elem
    return msg
end procedure
procedure QUANTIZE(msg)
    msg ← Normalize(msg)
    s_msg ← msg/S                              ▷ Scale to range
    qt_msg ← quantize(s_msg)                   ▷ Integer vector
    deqt_msg ← dequant(qt_msg)                 ▷ Rounded float
    discrete_msg ← deqt_msg                    ▷ For logging
    deqt_msg ← deqt_msg ∗ S                    ▷ Scale back
    return (deqt_msg, discrete_msg)
end procedure
msg ← Quantize(msg)
```

Main Idea and Running Example

Scales continuous messages to alphabet size and rounds values to the closest integer

Example for alphabet of 4 and word length of 5

Initial Msg	-3.5	0.2	1.3	-1.3	2.7	
Normalize	0	3.7	4.8	2.2	6.2	Sender
	0	0.59	0.77	0.35	1.0	Sender
Scaling by 4	0	2.38	3.09	1.4	4.0	Sender
Rounding	0	2.0	3.0	1.0	4.0	Sender
Scaling back	0	0.5	0.75	0.25	1.0	Rcvr

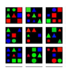
Explanation and Pytorch Implementation

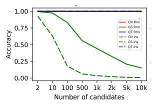
- `quantize_per_tensor` and `dequantized` methods allow gradients to flow through non differentiable rounding operation
- Implementing the straight through estimator during backward by overriding the backward methods and returning the input gradients.

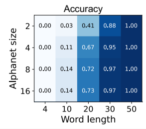
Games Datasets and Results

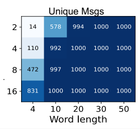
Synthetic Object Referential Game

Target: object with 4 props each with 10 values



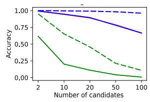


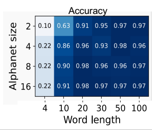


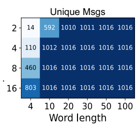


ImegeNet Image Referential Game

Target: an ImageNet image



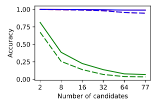


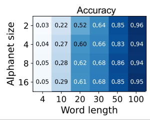


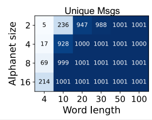
Sentence Referential Game

Target: A sentence that describes human intent

`when will I receive my new card?`







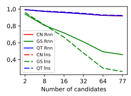
Sentence Classification Game

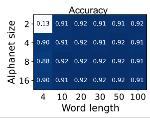
Target: A sentence that describes intent

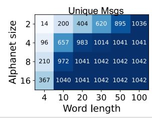
`how do I find the exchange rate?`

Candidate: The intent

`exchange rate`







Classification and Referential Games

- In the **referential game** the message sent by the sender needs to accurately describe the target.
- In **classification game** on the other hand, message need to let parties agree on the object's class.

The Role of Channel Capacity

- Plays significan role in agens ability to accomplish a task **but is not the only factor**
- Sometimes even unlimited channel capacity is not enough
  - Continuous channel has unlimited capacity, still cannot perfectly solve the classification game
  - With RNN even a Gumbel-softmax channel has enough capacity to solve the e.g., object game
- **Optimization plays significant role too**
- Limiting the capacity of a quantized channel can be achived by controllling word leangth and/or alphabet size

Open Issues

- Which inductive biases needs to be added beyond merely requiring the communication to be discrete?
- Can a non-natural, yet discrete, language that emerge from agents' communication be translated to natural language?
- Do humans represent the world using continuous or discrete elements, and when during the communication process discretization occur?

Contact: Boaz Carmeli boaz.carmeli@campus.technion.ac.il