Exploring the Generalizability of Sequence-to-Sequence Architecture
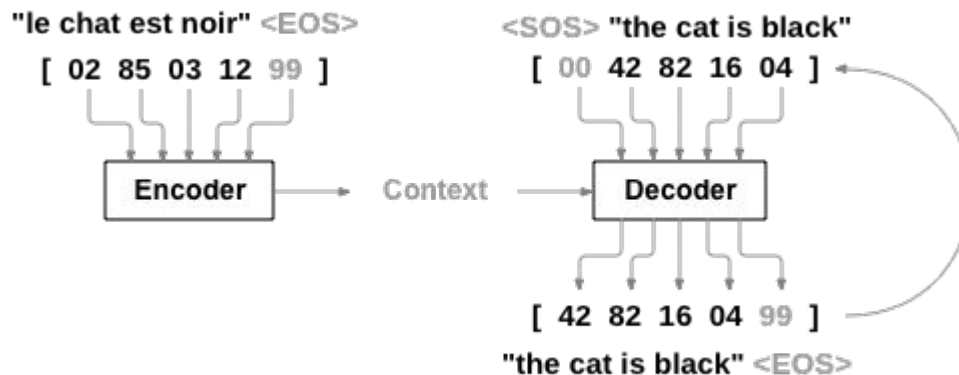
Kumar Abhishek and **Nishant Kambhatla**

Motivation

➢ The ultimate aim of AI is to reach Artificial General Intelligence (AGI).

➢ Deep learning has improved the performance on many NLP tasks individually.

➢ But the generalization of NLP models remains a hard problem within an approach that focuses on the particularities of a single metric, dataset, and task.
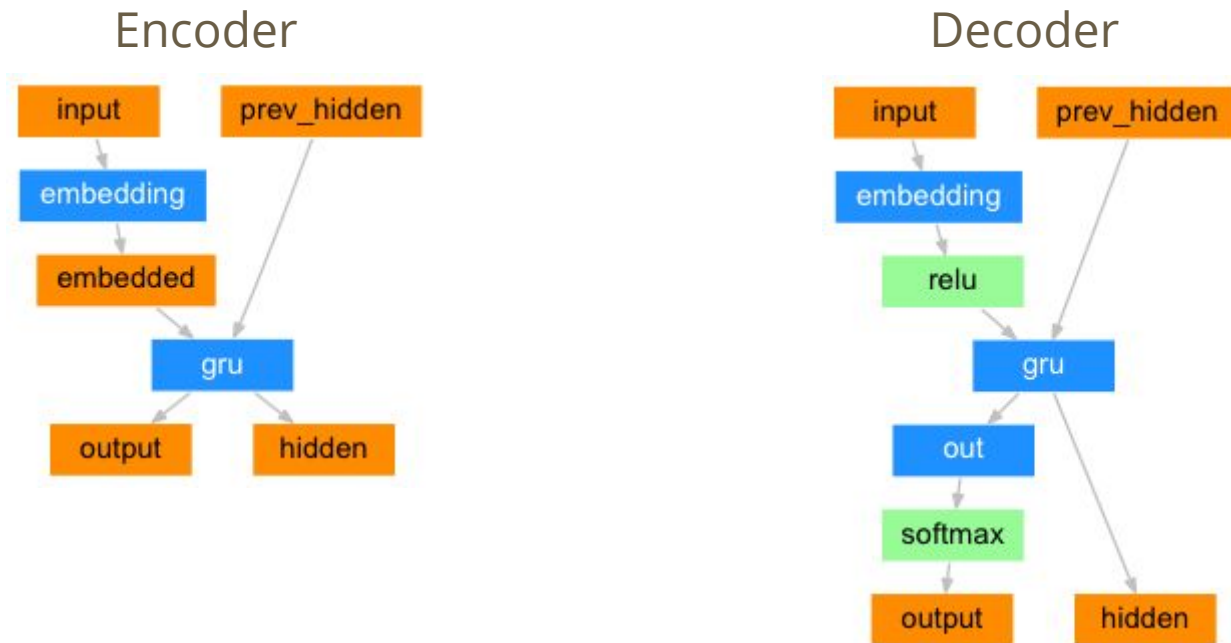
Generalizability

➢ The extent to which research findings can be applied to settings other than that in which they were originally tested.

➢ ~~One task → many architectures~~ One architecture → many tasks

The Encoder-Decoder Architecture

Kyunghyun Cho et al., *Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation*, EMNLP, 2014.

The Encoder-Decoder Architecture

Encoder

Decoder

Sean Robertson, *Translation With A Sequence To Sequence Network And Attention*, PyTorch Tutorials

Encoder-Decoder with Attention

Attention Decoder

Sean Robertson, *Translation With A Sequence To Sequence Network And Attention*, PyTorch Tutorials
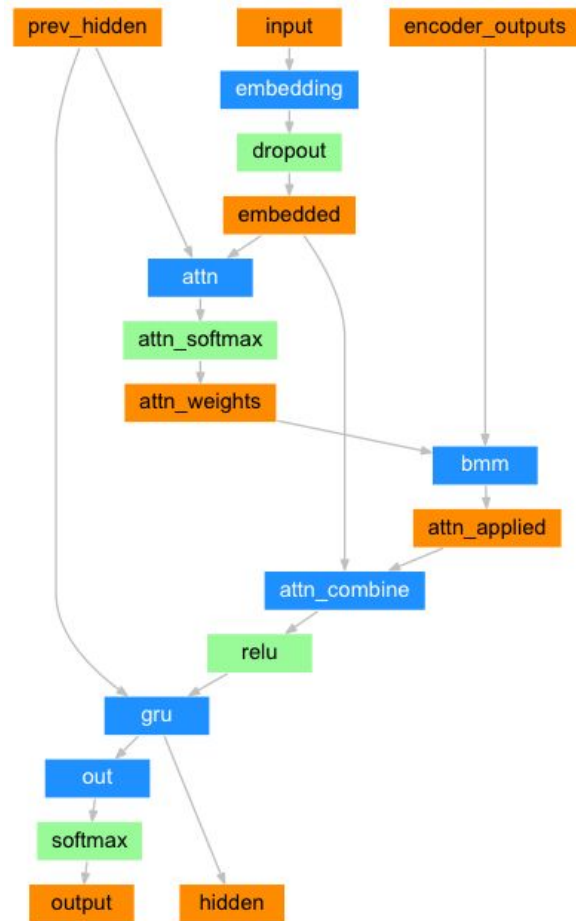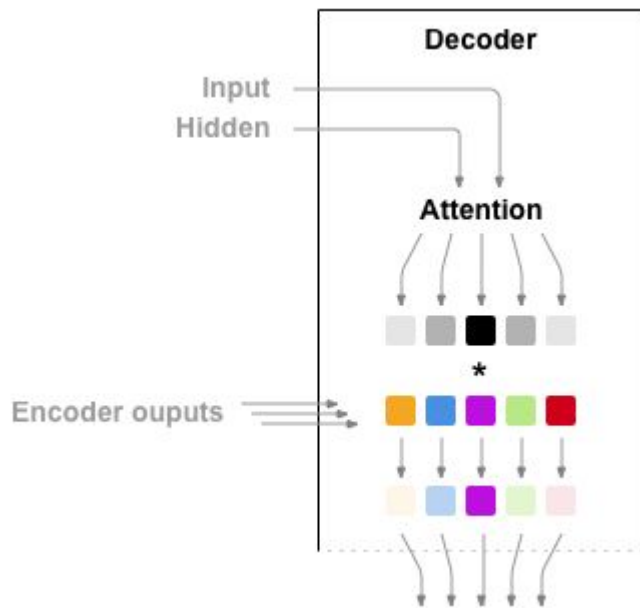
Encoder-Decoder with Attention

Translation Example:
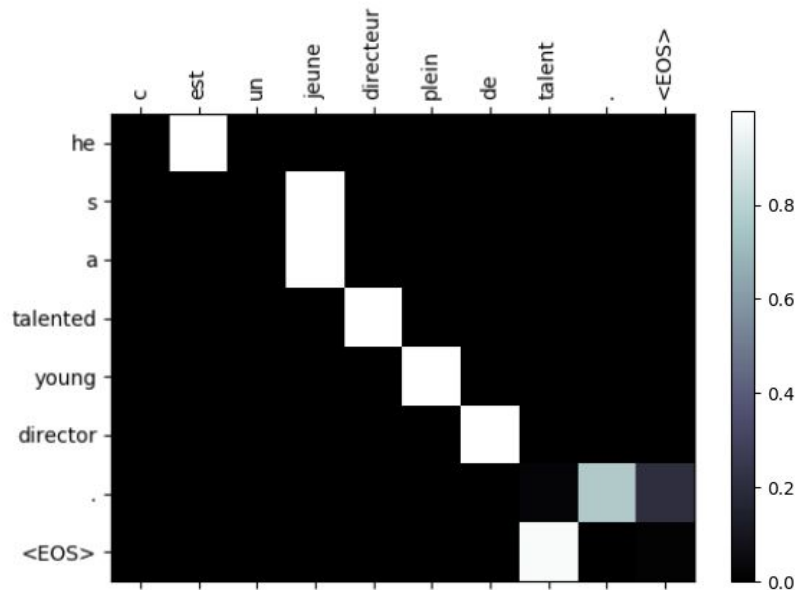
Datasets

Neural Machine Translation

IWSLT '14 dataset

- **De-En**
  - 63.9k sentence pairs for training, 930 for validation, and 1660 for testing

text	#sent.	German		English	
		$|W|$	$|V|$	$|W|$	$|V|$
parallel	63.9k	1.16M	63.1k	1.22M	35.5k
dev2010	930	19.1k	4.2k	20.2k	3.4k
tst2010	1660	30.3k	5.2k	32.0k	3.9k

IWSLT 2014 Evaluation Campaign, *http://workshop2014.iwslt.org/59.php*

Text Summarization

CNN/Daily Mail dataset.

```
Online news articles paired with multi-sentence (average
3.75 sentences) summaries.
```

	CNN			Daily Mail		
	train	valid	test	train	valid	test
# months	95	1	1	56	1	1
# documents	90,266	1,220	1,093	196,961	12,148	10,397
# queries	380,298	3,924	3,198	879,450	64,835	53,182
Max # entities	527	187	396	371	232	245
Avg # entities	26.4	26.5	24.5	26.5	25.5	26.0
Avg # tokens	762	763	716	813	774	780
Vocab size	118,497			208,045		

Karl Moritz Hermann et al., *Teaching Machines to Read and Comprehend*, NeurIPS 2015

Text Summarization

Example:

Text: The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "TopGear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack."

Summary: Producer Oisin Tymon will not press charges against Jeremy Clarkson, his lawyer says.
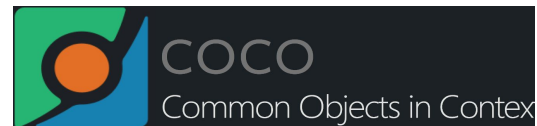
Dialogue Systems

Movie subtitles in English from OpenSubtitles

```
446,612 documents, 3.2G tokens
```

Examples:

- "Yeah , I want to tell him I 'm okay ."
- "I can 't make you believe it ."
- "We are going to the hospital ."

Jörg Tiedemann et al., *Parallel Data, Tools and Interfaces in OPUS*, LREC 2012

Image Captioning

Microsoft COCO dataset

`164k images, 5 captions per image`

The man at bat readies to swing at the pitch while the umpire looks on.

A large bus sitting next to a very tall building.

Tsung-Yi Lin et al., *Microsoft COCO: Common Objects in Context*, ECCV 2014

The Plan

The Roadmap

1. Data collection and preprocessing
2. Building an initial architecture
3. Seq2Seq experiments
4. Model optimization
5. Evaluation of the model across tasks

Evaluation

Tasks	NMT	Summarization	Dialog	Image Captioning
Arch1				
Arch2				

Comparing Performance Across Architectures and Tasks
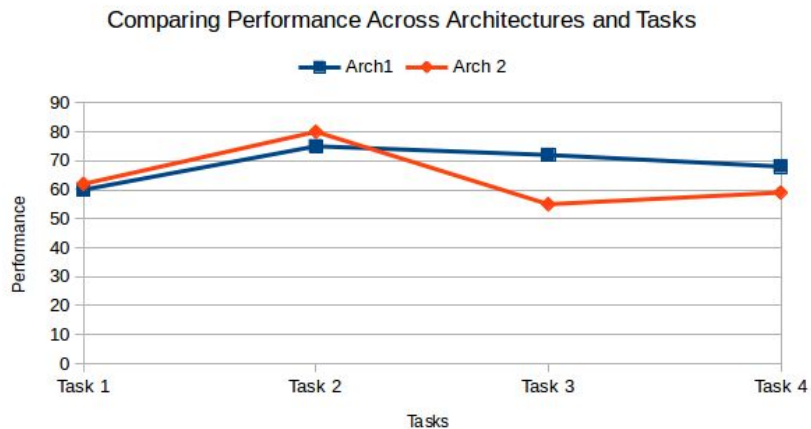
Progress So Far

Initial Architecture

```
[2019-03-27 22:34:08,586 INFO]  * src vocab size = 26725
[2019-03-27 22:34:08,586 INFO]  * tgt vocab size = 43572
[2019-03-27 22:34:08,586 INFO] Building model...
[2019-03-27 22:34:17,268 INFO] NMTModel(
  (encoder): RNNEncoder(
    (embeddings): Embeddings(
      (make_embedding): Sequential(
        (emb_luts): Elementwise(
          (0): Embedding(26725, 500, padding_idx=1)
        )
      )
    )
    (rnn): LSTM(500, 250, num_layers=2, dropout=0.3, bidirectional=True)
  )
  (decoder): InputFeedRNNDecoder(
    (embeddings): Embeddings(
      (make_embedding): Sequential(
        (emb_luts): Elementwise(
          (0): Embedding(43572, 500, padding_idx=1)
        )
      )
    )
    (dropout): Dropout(p=0.3)
    (rnn): StackedLSTM(
      (dropout): Dropout(p=0.3)
      (layers): ModuleList(
        (0): LSTMCell(1000, 500)
        (1): LSTMCell(500, 500)
      )
    )
    (attn): GlobalAttention(
      (linear_in): Linear(in_features=500, out_features=500, bias=False)
      (linear_out): Linear(in_features=1000, out_features=500, bias=False)
    )
  )
  (generator): Sequential(
    (0): Linear(in_features=500, out_features=43572, bias=True)
    (1): Cast()
    (2): LogSoftmax()
  )
)
[2019-03-27 22:34:17,269 INFO] encoder: 16370500
[2019-03-27 22:34:17,269 INFO] decoder: 49373572
[2019-03-27 22:34:17,269 INFO] * number of parameters: 65744072
[2019-03-27 22:34:17,271 INFO] Starting training on GPU: [0]
```

18

Initial Results: NMT

- En-Fr dataset from Tatoeba
  - 164,327 sentence pairs.
  - Example: "I am cold." ↔ "J'ai froid."

[Training]

```
[2019-03-26 20:57:58,586 INFO] Extracting features…
[2019-03-26 20:57:58,593 INFO] Building `Fields` object…
[2019-03-26 20:57:58,593 INFO] Building & saving training data…
[2019-03-26 20:57:58,593 INFO] Reading source and target files: data/en.train data/fr.train.
[2019-03-26 20:57:58,639 INFO] Building shard 0.
[2019-03-26 20:58:01,335 INFO]  * saving 0th train data shard to data/fr-en.train.0.pt.
[2019-03-26 20:58:04,846 INFO] Building & saving validation data…
[2019-03-26 20:58:04,846 INFO] Reading source and target files: data/en.val data/fr.val.
[2019-03-26 20:58:04,849 INFO] Building shard 0.
[2019-03-26 20:58:04,861 INFO]  * saving 0th valid data shard to data/fr-en.valid.0.pt.
[2019-03-26 20:58:04,900 INFO] Building & saving vocabulary…
[2019-03-26 20:58:06,094 INFO]  * reloading data/fr-en.train.0.pt.
[2019-03-26 20:58:07,972 INFO]  * tgt vocab size: 43572.
[2019-03-26 20:58:08,014 INFO]  * src vocab size: 26725.
```

Initial Results: NMT

[Demo]

```
SENT 696: ['You', 'should', 'apologize.']
PRED 696: Tu devrais présenter tes excuses.
PRED SCORE: -0.8661

SENT 697: ['I', 'want', 'you', 'to', 'be', 'my', 'friend', 'again.']
PRED 697: Je veux que vous soyez à nouveau mon ami.
PRED SCORE: -0.8589

SENT 698: ['It', 'looks', 'as', 'if', "you're", 'right.']
PRED 698: On dirait que tu as raison.
PRED SCORE: -0.9138

SENT 699: ['Show', 'me', 'how', 'to', 'do', 'that.']
PRED 699: Montre-moi comment faire ça !
PRED SCORE: -0.3248

SENT 700: ['When', 'he', 'came', 'to,', 'he', 'was', 'tied', 'to', 'a', 'chair', 'in', 'the', 'basement.']
PRED 700: Lorsqu'il reprit ses esprits, il était ligoté à une chaise dans la cave.
PRED SCORE: -0.0473
PRED AVG SCORE: -0.1249, PRED PPL: 1.1330
```

Overview

✓   Data for all the tasks:
- ✓   Collected and pre-processed

➢   Training:
- ✓   NMT
- ✓   Summarization
- ✓   Dialog
- ➢   Image Captioning

➢   Hyperparameter optimization

➢   Model evaluation

Thank You