DMRS to text generation with neural networks

Valerie Hadijk
Swabha Swayamdipta
Michael Goodman
Jan Buys

DEPARTMENT OF LINGUISTICS
UNIVERSITY of WASHINGTON

W PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Text Generation

- Recently neural networks have become the predominant focus of research in natural language generation in the NLP community.
- Models are typically trained end-to-end: Given input to condition on, generate the output without using any explicit intermediate representations.
- Applications:
    - Machine translation
    - Summarization
    - Dialogue systems
    - Tabular data to text generation
    - Auto-reply
    - Image captioning

Neural generation

Strengths

- Generate (mostly) fluent text, handle long-distance dependencies much better than previous approaches
- Provides a general framework for many generation tasks and different input modalities (seq2seq)
- Works really well when
  - The task is well defined
  - There is lots of training data available

Neural generation

Weaknesses

- Lack of interpretability (prone to adversarial attacks)
- Lack of guarantees of correctness or convergence
- Requires lots of training data
- Long-form generations lack coherence - really good at capturing surface patterns (even long-distance), but lacks understanding of what is being generated.

Neural generation using linguistic structure

- Separate what we want to say from how to say it
- Use semantic representation as intermediate step
- This talk: Focus on generating from a semantic representation

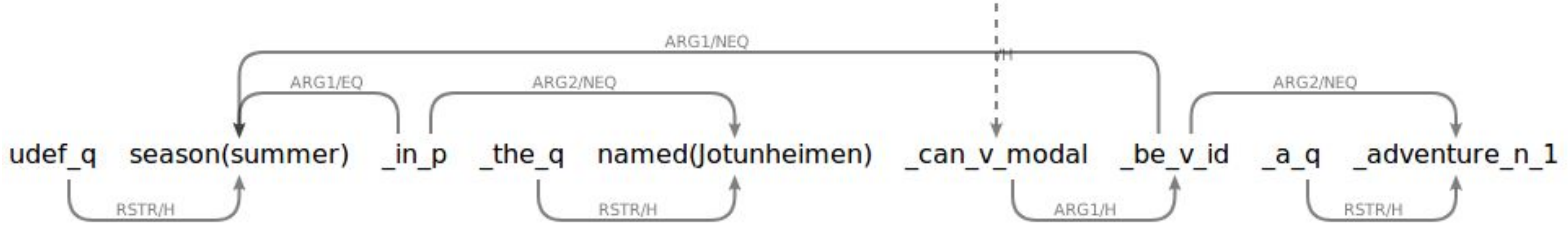
Why DMRS-to-text generation?

- Represents sentence meaning
- High-precision grammar available
- Detailed annotations
- Graph-based representation
- Large annotated corpora available for training (Redwoods)
- Not AMR

Why **Neural** DMRS-to-text generation?

- Robust generation: Full sentence coverage, OOV handling
- Improve output fluency through RNN language model
- Facilitates integration with other neural network models

DMRS to text generation



Summer in the Jotunheimen can be an adventure.

DMRS to text generation

Approach

- Linearize the DMRS graphs
  - Represent graphs in PENMAN format (Michael Goodman)
  - Simplify the representation :Remove nodes ids, remove spans, merge properties
  - Anonymize entities
- Tokenize the output sentences
- Use neural sequence-to-sequence models

DMRS to text generation

```
(10005 / _can_v_modal
        :lnk "<26:29>"
        :sf PROP
        :tense PRES
        :mood INDICATIVE
        :perf -
        :ARG1-H (10006 / _be_v_id
                :lnk "<30:32>"
                :sf PROP
                :tense UNTENSED
                :mood INDICATIVE
                :perf -
                :ARG1-NEQ (10001 / season
                        :lnk "<0:6>"
                        :carg "summer"
                        :pers 3
                        :num SG
                        :gend N
                        :ind -
                        :ARG1-EQ-of (10002 / _in_p
                        :lnk "<7:9>"
                        :sf PROP
                        :tense UNTENSED
                        :mood INDICATIVE
                        :perf -
                        :ARG2-NEQ (10004 / named
                                :lnk "<14:25>"
                                :carg "Jotunheimen"
                                :pers 3
                                :num SG
                                :ind +
                                :RSTR-H-of (10003 / _the_q
                                        :lnk "<10:13>"))))
        :ARG2-NEQ (10008 / _adventure_n_1
                :lnk "<36:46>"
                :pers 3
                :num SG
                :RSTR-H-of (10007 / _a_q
                        :lnk "<33:35>"))))
```

DMRS to text generation

```
( _can_v_modal sf=PROP|tense=PRES|mood=INDICATIVE|perf=-
  :ARG1-H ( _be_v_id sf=PROP|tense=UNTENSED|mood=INDICATIVE|perf=-
      :ARG1-NEQ ( season :carg "summer" pers=3|num=SG|gend=N|ind=-
      :ARG1-EQ-of ( _in_p sf=PROP|tense=UNTENSED|mood=INDICATIVE|perf=-
            :ARG2-NEQ ( named :carg "Jotunheimen" pers=3|num=SG|ind=+
      :RSTR-H-of ( _the_q ) ) ) )
      :ARG2-NEQ ( _adventure_n_1 :pers=3|num=SG
      :RSTR-H-of ( _a_q ) ) ) )
```

DMRS to text generation

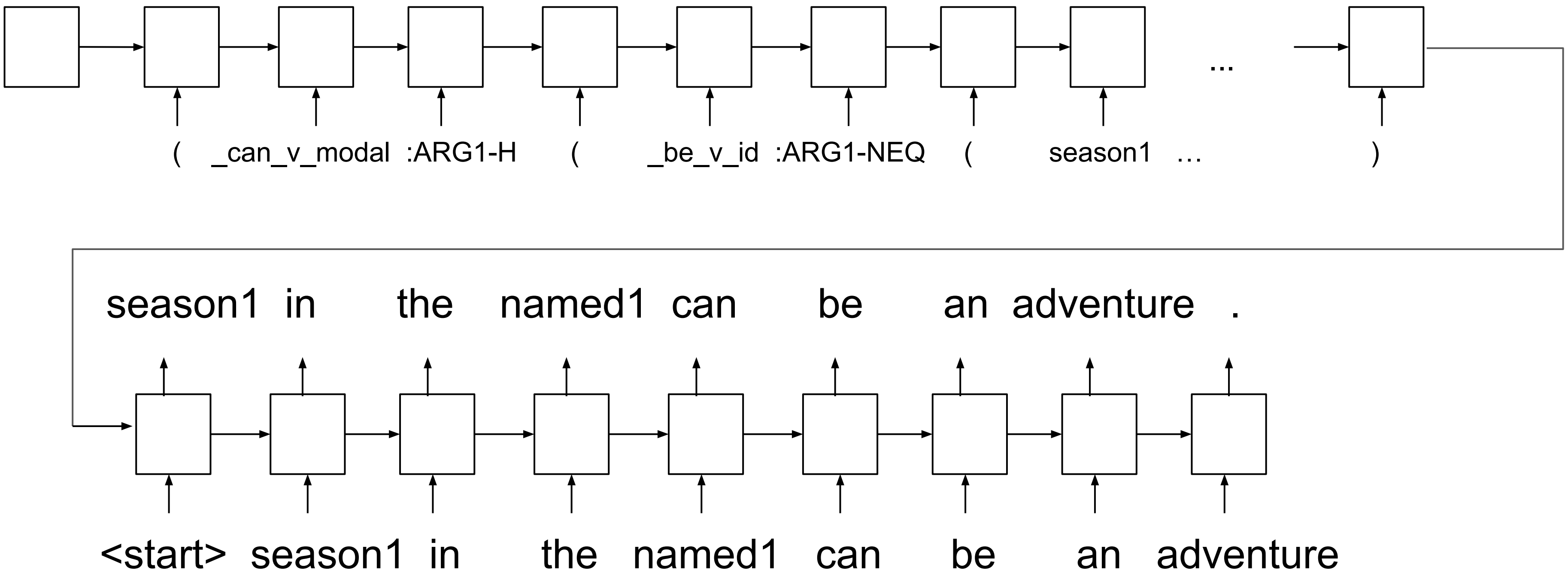
```
( _can_v_modal sf=PROP|tense=PRES|mood=INDICATIVE|perf=-
 :ARG1-H ( _be_v_id sf=PROP|tense=UNTENSED|mood=INDICATIVE|perf=-
     :ARG1-NEQ ( season1 pers=3|num=SG|gend=N|ind=-
     :ARG1-EQ-of ( _in_p sf=PROP|tense=UNTENSED|mood=INDICATIVE|perf=-
            :ARG2-NEQ ( named1 pers=3|num=SG|ind=+
     :RSTR-H-of ( _the_q ) ) ) )
     :ARG2-NEQ ( _adventure_n_1 :pers=3|num=SG
     :RSTR-H-of ( _a_q ) ) ) )
```

season1 in the **named1** can be an adventure .

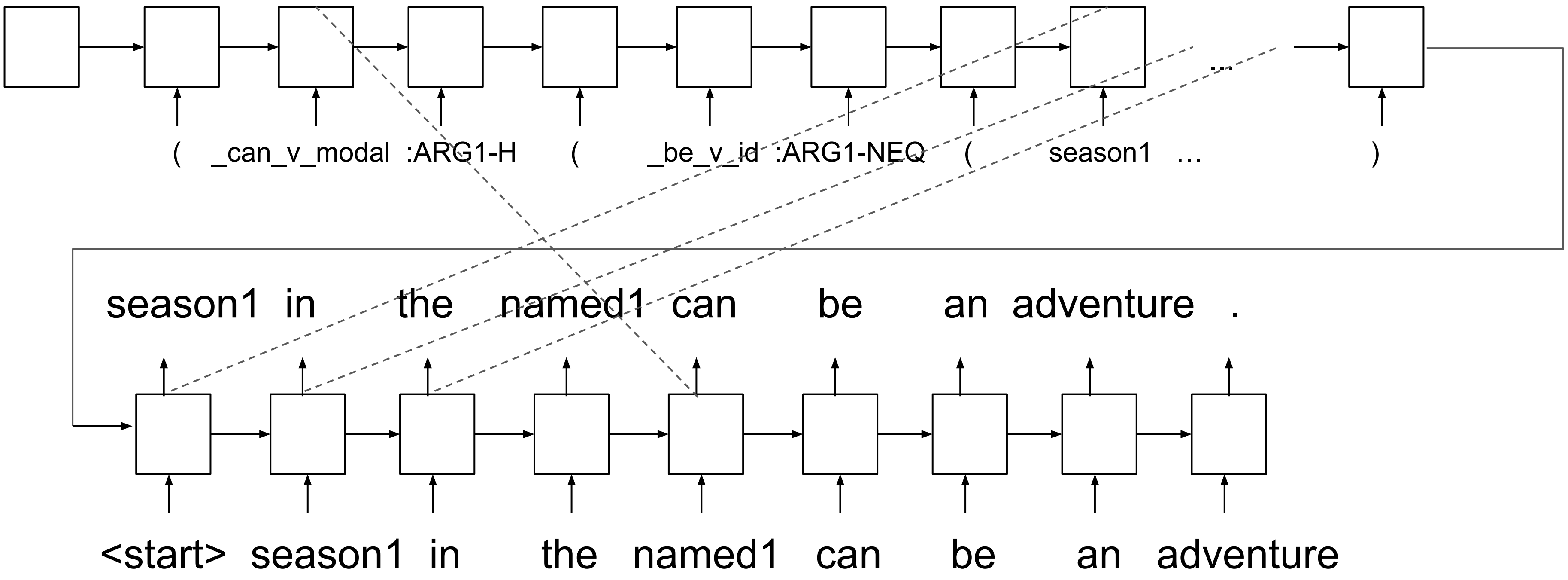
Sequence to sequence models

- Recurrent Neural Networks, “encoder-decoder” models
- Used for Neural Machine Translation, as well as other language generation tasks
- Use version with attention mechanism that models the alignment between the output and input sequences
- Trained end-to-end, decoding is greedy or with beam search

Sequence to sequence models



Sequence to sequence models (with attention)



Experiments

Sources of training data

- Gold: Redwoods (ERG 1214) - 72k training sentences
- Silver: Gigaword corpus subset - 1M sentences

Evaluate on held-out Redwoods data

Evaluation metric: BLEU

Use OpenNMT implementation of seq2seq models

Results

Hyperparameters (best found so far)

- 2 Layer BiLSTM,
- 0.5 dropout,
- embed size 500,
- hidden size 80,
- starts learning rate decay at epoch 25
- Train 30 epochs

Results

	Full Redwoods (BLEU)	WSJ only (BLEU)
Gold (70k)	65.90	70.12
Gold 70k + Silver 60k	70.78	74.11
Gold 70k, Silver 600k	**75.74**	**78.57**
Silver only 1M	71.68	73.77

In comparison, for AMR-to-text generation, comparable model gets 33.8 BLEU (although on a different dataset).

Results

Gold constant argument strings, larger vocabulary

	Vocab size	Full Redwoods (BLEU)	WSJ only (BLEU)
Gold (70k)	38K	67.59	72.11
Gold 70k + Silver 600k	100K	**77.76**	**81.37**

Results

On sentences covered by both generators

	Coverage	BLEU
ERG/ACE	80.3	61.98
NN Gold 70k + Silver 600k	99.9	**81.40**

Example generations

Reference	The Paris-based group said its earlier projection **--** that group profit for all of 1989 would be close to the 322.7 million francs posted for 1988 **--** remains valid.
NN generator	The Paris-based group said its earlier projection that group profit for all of 1989 would be close to the 322.7 million francs posted for 1988 remains valid.
ERG generator	The Paris based group said that its earlier projection **like** group profit for all of 1989 would be close to the 322.7 million franks, posted for 1988, remains valid.

Example generations

Reference	Am looking for a radar detector for my car.
NN generator	looks for a radar detector for my car.
ERG generator	Are looking for a radar detector for my car.

Example generations

Reference	I consulted your webpage for a digital cam.
NN generator	I consulted your fax for a digital cam.
ERG generator	I consulted your webpage, for a digital cam.

Example generations

Reference	But the test may prove to be more sensitive in determining whether a tumor has spread or returned following treatment, Dr. Wilson said.
NN generator	But the test may prove to be more sensitive in determining whether a tumor has spread or returned following treatment, Dr. Wilson said.
ERG generator	But as though the test may prove to be more sensitive in determining if following treatment, a tumor has spread, or has returned, doctor Wilson said.

Example generations

Reference	Dr. Wyndham Wilson, a cancer treatment specialist at the National Cancer Institute, said the test is widely used in research centers but isn't having a major impact because it is only occasionally useful in choosing the most effective treatment.
NN generator	Dr. Wyndham Wilson, a cancer treatment specialist at the National Cancer Institute, said the test is widely used in research centers, but **doesn't** have a major impact because it is only occasionally useful in choosing the most
ERG generator	Doctor Wyndham Wilson, a cancer treatment specialist at the National Cancer Institute, said that the test is widely used in research centers, but is not having a major impact because it is useful, in choosing of the most effec

Work in progress / future work

- Strategies to combine gold and silver training data
- Hyperparameter turing
- Handling unknown words, scaling to larger vocabularies
- Use annotated alignments during training (“hard” attention)
- Neural network architectures that model graph structures explicitly (graph convolutional encoders)

Work in progress / future work

- Contribution of different levels of annotation to generation quality (properties, handle constraints, quantifiers, predicate senses, abstract predicate types)
- Error analysis: Compare to ERG generations, grammaticality and consistency of generations
- Controlling syntax, style or content of generations
- Downstream applications