Can Dependency Aid Natural Language Generation?

Cases from English and Japanese

Gyu-min Lee
(Korea University, gyuminlee@korea.ac.kr)

July 14 2020

Outline

Who am I?

- Gyu-min Lee
- Graduated from Handong University (Pohang, ROK)
  - B.A. in Studies of English & Counseling Psychology
  - Minored in Computer Science
- Student at Korea University Graduate School (Seoul, ROK)
  - For M.A. in Linguistics
  - Majoring in Computational Linguistics
  - Under the guidance of Prof. Sanghoun Song

Who am I?

- Gyu-min Lee
- Graduated from Handong University (Pohang, ROK)
    - B.A. in Studies of English & Counseling Psychology
    - Minored in Computer Science
- Student at Korea University Graduate School (Seoul, ROK)
    - For M.A. in Linguistics
    - Majoring in Computational Linguistics
    - Under the guidance of Prof. Sanghoun Song

Who am I?

- Gyu-min Lee
- Graduated from Handong University (Pohang, ROK)
  - B.A. in Studies of English & Counseling Psychology
  - Minored in Computer Science
- Student at Korea University Graduate School (Seoul, ROK)
  - For M.A. in Linguistics
  - Majoring in Computational Linguistics
  - Under the guidance of Prof. Sanghoun Song

Who am I?

Who am I?

- A little background
  - "I would like to help others to understand the information written in other languages through (automatic) translation." – ME (2003)
  - "Programming is so fun!" – also ME (2008)
  - Science track at high school
  - Started as Computer Science major
  - Eventually, English & Counseling Psychology
  - "Deep understanding of syntax can help others (including machines) to understand the language!" – ME as well (2015)

Who am I?

- Research Interests
  - Syntactic analysis of human language
  - Machine natural language generation
  - Introduction of rich grammar to NLP
- Project Participation
  - In writing a proposal for deep analysis of KFL learners
  - In writing a proposal for Korean NER development
  - In building a library for syntactic experiments for Korean language using neural LMs (present)

Who am I?

- Research Interests
  - Syntactic analysis of human language
  - Machine natural language generation
  - Introduction of rich grammar to NLP
- Project Participation
  - In writing a proposal for deep analysis of KFL learners
  - In writing a proposal for Korean NER development
  - In building a library for syntactic experiments for Korean language using neural LMs (present)

Who am I?

- Why this topic?
  - "NLP is so successful that it does not require feature engineering"
  - But NLP still requires a lot of resources and still "sounds" computer...
  - What if we "teach" the neural language model grammar?
- Where to go from here?
  - More ways to teach neural LMs grammar
  - Neural LMs are learning grammars – how? How can we even facilitate that?
  - Probing neural LMs' language acquisition
  - See how much grammar we can teach for the best performance and efficiency

Who am I?

- Why this topic?
  - “NLP is so successful that it does not require feature engineering”
  - But NLP still requires a lot of resources and still “sounds” computer...
  - What if we “teach” the neural language model grammar?
- Where to go from here?
  - More ways to teach neural LMs grammar
  - Neural LMs are learning grammars – how? How can we even facilitate that?
  - Probing neural LMs’ language acquisition
  - See how much grammar we can teach for the best performance and efficiency

What Is This?

- Master's Thesis
  - Offering a (yet another) case where *linguistics matters to NLP*
- The ultimate goal
  - To probe the "neural language acquisition"
  - The understanding of how neural networks acquire a language

The Current Status of NLG

- Natural language generation (NLG) is important from summarization to MT
- Introduction of neural networks greatly improved the surface performance of NLG
- Introduction of Seq2Seq structure further improved the NLG
- ...and once again with Attention-only Seq2Seq (a.k.a. Transformer)

Seq2Seq

- Encoder
  - Makes a sentence embedding vector out of source sentence vectors
  - Similar to sentence classification RNN models
- Decoder
  - A conditional neural network language model
  - Current time-step word is based on the sentential embedding vector from the Encoder and the words generated so far through translation.
- Generator
  - Make probability out of the vector resulting from the decoder per time-step with Softmax function

Seq2Seq

- Encoder
  - Makes a sentence embedding vector out of source sentence vectors
  - Similar to sentence classification RNN models
- Decoder
  - A conditional neural network language model
  - Current time-step word is based on the sentential embedding vector from the Encoder and the words generated so far through translation.
- Generator
  - Make probability out of the vector resulting from the decoder per time-step with Softmax function

Seq2Seq

- Encoder
  - Makes a sentence embedding vector out of source sentence vectors
  - Similar to sentence classification RNN models
- Decoder
  - A conditional neural network language model
  - Current time-step word is based on the sentential embedding vector from the Encoder and the words generated so far through translation.
- Generator
  - Make probability out of the vector resulting from the decoder per time-step with Softmax function

Problems with current NLG

- Room for improvement for E2E neural NLG
  - Domain-general
  - Grammatical appropriateness and naturalness
  - More predictability

A Fundamental Issue with NLG

- *Human language is hierarchical*
  - Yet, neural LMs treat language to be linear

Various Methods...

- Revisiting the template-based method with information extraction
- Teaching the machine the *un*likely tokens
- Feature engineering, like “surprise” for pun generation

Various Methods...

- Revisiting the template-based method with information extraction
- Teaching the machine the *un*likely tokens
- Feature engineering, like “surprise” for pun generation

Various Methods...

- Revisiting the template-based method with information extraction
- Teaching the machine the *un*likely tokens
- Feature engineering, like “surprise” for pun generation

A Fundamental Issue with NLG

- *Human language is hierarchical*
  - Yet, neural LMs treat language to be linear

Let's teach deep learning syntax!

- Not all the syntax – rich but not efficient
- The argument structure – perhaps a good balance between efficiency and linguistic richness?
- In the form of dependency
- i.e. *Grammatical Relation*
- Minimal Recursion Semantics?

Didn't They Learn Syntax Already?

- Recent findings showing that neural language models have acquired some syntax
- If language models can perform that well with that little bit of syntactic knowledge. . .
- It would do much better with explicit syntax!

Didn't They Learn Syntax Already?

- Recent findings showing that neural language models have acquired some syntax
- If language models can perform that well with that little bit of syntactic knowledge...
- It would do much better with explicit syntax!

Didn't They Learn Syntax Already?

- Recent findings showing that neural language models have acquired some syntax
- If language models can perform that well with that little bit of syntactic knowledge...
- It would do much better with explicit syntax!

Data

- Gold Data
  - Redwoods Treebank
  - Hinoki Treebank
- Silver Data
  - Following Hajdik et al. (2019) that found using silver data as well improved NLG compared to gold only
  - MRS generated with ERG/JACY
  - Brown...?
  - Japanese corpora...?

Data

- Gold Data
  - Redwoods Treebank
  - Hinoki Treebank
- Silver Data
  - Following Hajdik et al. (2019) that found using silver data as well improved NLG compared to gold only
  - MRS generated with ERG/JACY
  - Brown...?
  - Japanese corpora...?

Model

- Baseline from Hajdik et al. (2019) for English
    - From MRS
    - Encoder-decoder design
    - Bi-LSTM Encoder
    - Attention Decoder
- Transformer
- BERT/XLNet/ELECTRA. . . ?

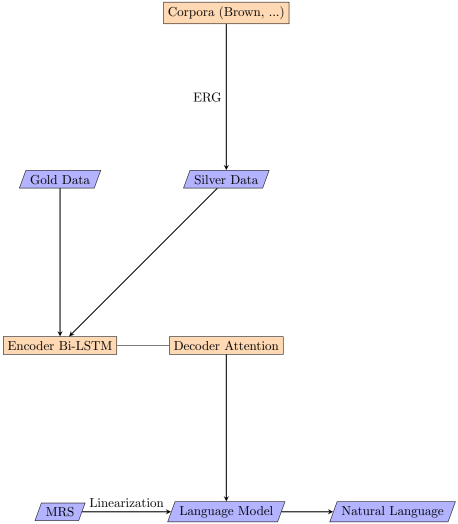
Model

- Baseline from Hajdik et al. (2019) for English
  - From MRS
  - Encoder-decoder design
  - Bi-LSTM Encoder
  - Attention Decoder
- Transformer
- BERT/XLNet/ELECTRA...?

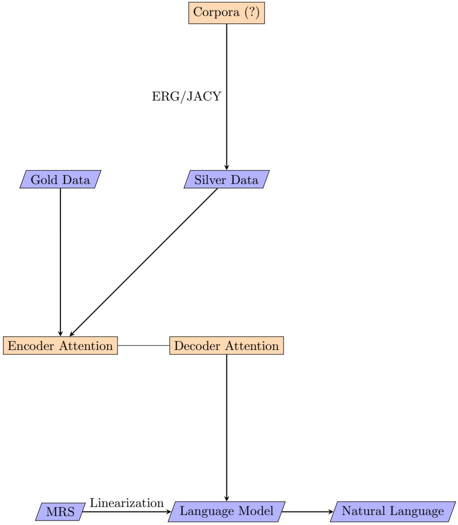
Model

- Baseline from Hajdik et al. (2019) for English
  - From MRS
  - Encoder-decoder design
  - Bi-LSTM Encoder
  - Attention Decoder
- Transformer
- BERT/XLNet/ELECTRA. . . ?

Model from Hajdik et al. (2019)

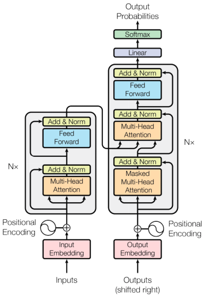


Suggested Model



Transformer

- Vaswani et al. (2017)
- Seq2Seq with only Attention mechanism
- Without any variant of recurrent neural network
- Thus, *Attention is All You Need*



Attention Mechanism

- Because long term dependency even with LSTM & GRU
- To get a good result at Google, you should QUERY well
- Given the vectors of *QUERY, KEY, & VALUE*,
- Attention helps us to perform the correct queries.
- i.e., attention governs how “important” each word are.

Why Transformer?

- This task is a translation between the source language of MRS representation and the target language of English and Japanese
- Linearized MRS is not free from long term dependency problem
- Using attention would help in letting the machine to understand which grammatical feature matters the most
- In other words, each attention heads will “look at” certain linguistic features, thus “learning” the syntax better

Why Not BERT and others?

- Their strength comes from being pre-trained
- Models pre-trained with MRS representation of ERG/JACY?
- If there isn't one, we can train MRS BERT, etc.
- But is it really beneficial for now?

Expected result

- Would score higher BLEU
- Would show better qualitative result
  - Hajdik et al. (2019) already showed an improvement of around 15 BLEU by implementing MRS
  - Because of the implementation of the newer model, my model would perform better
  - As Transformer outperforms RNN-based Seq2Seq *in general* in MT tasks
  - Transformer would benefit from syntax as well
- Would have learned more syntax than the general Transformer models

Implications

- Significance of rule-based, hard-coded computational grammar
- Importance of syntactic information for neural NLG
- Extension of Hajdik et al. (2019) to more languages
  - Extension to another language
  - Increased language independency of the finding
  - Implementation of a newer technology

State of the research

- In early steps
  - Literature reviews on NLG
  - Literature reviews on neural LM's acquisition of syntax
  - Literature reviews on machine translation
- Collecting data
  - Getting Redwoods
  - Getting Hinoki
  - Deciding on the corpora for silver dataset
- reproducing Hajdik et al. (2019)
  - Translating from their data to English
  - Using Transformer via OpenNMT

State of the research

```
**SENTENCE**
The Cathedral and the Bazaar
**PREDICTED**
A bunch of the view .
**LINEARIZED MRS**
(|_ unknown|mood=INDICATIVE|perf=-|sf=PROP-OR-QUES ARG-NEQ|_ (|_ _and_c|num=PL|pers=3
L-INDEX-NEQ|_ (|_ _cathedral_n_1|ind=+|num=SG|pers=3 RSTR-H-of|_ (|_ _the_q|_ )|_ )|_
R-INDEX-NEQ|_ (|_ _bazaar_n_1|ind=+|num=SG|pers=3 RSTR-H-of|_ (|_ _the_q|_ )|_ )|_ )|_ )|_

**SENTENCE**
Permission is granted to copy , distribute and / or modify this document under the terms of the
Open Publication License , version card0 .
**PREDICTED**
Fortunately , the view , " ' " ' " " ' " " ' " ' " " ' " " ' " " ' " ' " ' " ' " ' " ' " ' " ' " '
" ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' " ' "

**LINEARIZED MRS**
(|_ _in+order+to_x|mood=INDICATIVE|perf=-|sf=PROP ARG1-H|_ (|_
_grant_v_1|mood=INDICATIVE|perf=-|sf=PROP|tense=PRES ARG2-NEQ|_ (|_
_permission_n_1|num=SG|pers=3 )|_ ARG1-EQ-of|_ (|_ parg_d|mood=INDICATIVE|perf=-|sf=PROP
ARG2-NEQ|_ _permission_n_1|num=SG|pers=3 )|_ )|_ ARG2-H|_ (|_
implicit_conj|mood=INDICATIVE|perf=-|sf=PROP L-HNDL-HEQ|_ (|_
_copy_v_1|mood=INDICATIVE|perf=-|sf=PROP ARG1-NEQ|_ _permission_n_1|num=SG|pers=3 ARG2-NEQ|_ (|_
_document_n_1|ind=+|num=SG|pers=3 RSTR-H-of|_ (|_ _this_q_dem|_ )|_ )|_ )|_ L-INDEX-NEQ|_
_copy_v_1|mood=INDICATIVE|perf=-|sf=PROP R-HNDL-HEQ|_ (|_ _and_c|mood=INDICATIVE|perf=-|sf=PROP
L-HNDL-HEQ|_ (|_ _distribute_v_to|mood=INDICATIVE|perf=-|sf=PROP ARG1-NEQ|_
_permission_n_1|num=SG|pers=3 ARG2-NEQ|_ _document_n_1|ind=+|num=SG|pers=3 )|_ L-INDEX-NEQ|_
_distribute_v_to|mood=INDICATIVE|perf=-|sf=PROP R-HNDL-HEQ|_ (|_
_modify_v_1|mood=INDICATIVE|perf=-|sf=PROP ARG1-NEQ|_ _permission_n_1|num=SG|pers=3 ARG2-NEQ|_
_document_n_1|ind=+|num=SG|pers=3 )|_ R-INDEX-NEQ|_ _modify_v_1|mood=INDICATIVE|perf=-|sf=PROP )
|_ R-INDEX-NEQ|_ _and_c|mood=INDICATIVE|perf=-|sf=PROP ARG1-EQ-of|_ (|_
_under_p|mood=INDICATIVE|perf=-|sf=PROP ARG2-NEQ|_ (|_ _term_n_of|ind=+|num=PL|pers=3 ARG1-NEQ|_
(|_ _license_n_1|num=SG|pers=3 RSTR-H-of|_ (|_ _the_q|_ )|_ ARG1-EQ-of|_ (|_
```

State of the research

- Trained with a small portion of the data from Hajdik et al. (2019)
- Using OpenNMT implementation of Transformer
- for 40 steps
- In short, Transformer is applicable

Agenda

- Write out the literature reviews
- Get the data
- Make the preprocessor (i.e. linearizer for MRS representation)
- Run the experiment with bigger epochs

Problems

- Where exactly do I get the Redwoods & Hinoki?
- Any model pre-trained with MRS representation of English/Japanese?
- Is OpenNMT the best choice for the task?
- Baseline model for Japanese?