Unsupervised Natural Language Generation for Task-Oriented Dialogue System

Dian Huang
dhuang05@stanford.edu

Motivation

- Traditional e-commerce dialogue system generates response with fixed template, which can lack variety and entertainment.
- Typical machine learning needs labor-intensive and error prone labels in data collection.
- Aim to proposes a label-free, annotation-free, unsupervised keywords-to-sentence generation

Method	Labels	Text
Template	name[Zizzi], eatType[coffee shop], rating[average], location[Burger King]	The [Name] [eatType] is located near [location]. It has an [rating] customer rating.
Conventional Machine Learning	name[Zizzi], eatType[coffee shop], rating[average], location[Burger King]	The Zizzi coffee shop is located near Burger King. It has an average customer rating.
Proposed Approach		The Zizzi coffee shop is located near Burger King. It has an average customer rating

Data-Structures & Pre-processing

Dataset

- Data 1: 51426 text samples from E2E [1] dataset about the description of restaurant.
- Test Data: 4000 randomly selected sentences in E2E with less than 6 keywords.
- Only text, and no labels or annotations

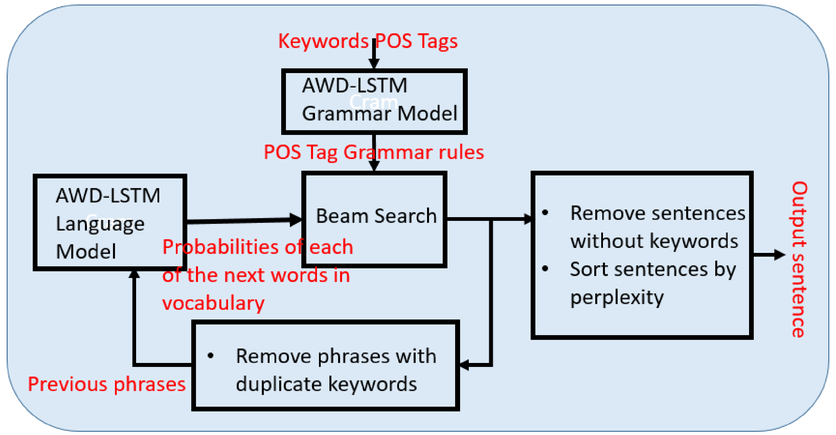
Pre-processing

- POS tagging each word in the dataset with Stanford tagger into 16 different categories.

The	Zizzi	Coffee	shop	is	located	near	Burger	King
DT	NNP	NN	NN	VBZ	VBN	IN	NNP	NNP
There	is	a	pub	called	Zizzi	near	the	Sorrento
EX	VBZ	DT	NN	VBN	NNP	IN	DT	NNP

Models & Algorithm

Goal: Want to find a sentence with low perplexity and has all the keywords



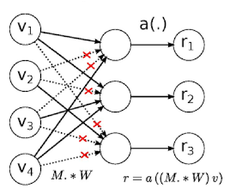
- Language model outputs the probability of each of the next word in vocabulary
- Grammar model generates the POS Tag for target sentence
- Beam search select from the words with the same POS Tags
- Filter out sentences without keywords

ASGD Weight-Dropped Language Model (AWD-LSTM)

ASGD

- For first k steps: $\omega_{k+1} = \omega_k - \gamma_k \nabla f(\omega_k)$
- Add the average of weight matrix

DropConnect



LSTM

$$i_t = \sigma(W^i x_t + U^i h_{t-1})$$
$$f_t = \sigma(W^f x_t + U^f h_{t-1})$$
$$o_t = \sigma(W^o x_t + U^o h_{t-1})$$
$$\tilde{c}_t = tanh(W^c x_t + U^c h_{t-1})$$
$$c_t = i_t \odot \tilde{c}_t + f_t \odot + \tilde{c}_{t-1}$$
$$h_t = o_t \odot tanh(c_t)$$

Issues:

- Has to select next word from the full vocabulary
- Rare words may never be predicted

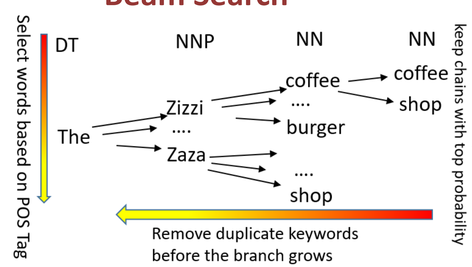
Proposed Two Models Approach

AWD-LSTM Grammar Model

- Generate grammar rules with POS Tags
- Rare words issue resolved as it belongs to one of the POS tags
- $\underset{t_i}{argmax} \log p(t_i | t_{i-1} t_{i-2} \dots t_1)$

Key words	Zizi	Coffee	Shop			
Tags	NNP	NN	NN			
DT	NNP	NN	NN	VBZ	DT	JJS
NNP	NN	NN	MD	VB	DT	JJS

Beam Search



- Combined grammar model with language model and maximize $\underset{w_i \in w_{t_i}}{argmax} \log p(w_i | t_i, w_{i-1}, w_{i-2}, w_{i-3} \dots w_1)$

AWD-LSTM Language Model

- Predict the next words given previous words by the word that maximize sentence probability: $\underset{w_i}{argmax} \log p(w_i | w_{i-1}, w_{i-2}, w_{i-3} \dots w$

Filter

- Generated senteces has the same POS Tags given by grammar model and all the keywords.
- Select sentence with lowest perplexity as output.

Reference:

[1] Jekaterina Novikova, Ondrej Dusek, and Verena Rieser. ˇ2017b. The E2E Dataset: New Challenges for Endto-End Generation. In Proceedings of SIGDIAL, pages 201–206, Saarbrucken, Germany.

[2] Simon Mille and Stamatia Dasiopoulou. 2018. FORGeat E2E 2017. In E2E NLG Challenge System Descriptions.

Evaluation & Discussion

Samples Output

Keywords	Reference	Generate
Strada restaurant average	Strada is a restaurant with an average customer rating	a strada is a family friendly restaurant that has average food
Aromi restaurant Chinese riverside .	Aromi is a restaurant which provides Chinese food in the riverside area	riverside aromi is a low customer restaurant that serves chinese customer .
The Punter English high .	The Punter is a restaurant with high prices.	the punter is english price with a high customer rating .

- Some sentence includes extra information.

Comparison

	Rouge-L	BLEU	Model
This work	**0.56**	**0.31**	**Unsupervised**
TGEN[1]	0.68	0.65	Seq2seq
FORGE3[2]	0.56	0.46	Templates

- Other word train with labels. (This work tests with only the data with less than 6 labels.

Discussion

- With dataset about only restaurant, language models achieve perplexity of 3.9. Grammar model with 16 different tags achieve a perplexity of 2.1.
- With low perplexity, the sentences generated are generally grammatically correct, but without supervised label it sometimes adds extra information.
- The extra information lowers BLEU score, which is about precision. Rouge score about recall remains to be high.

Future work

- Build grammar model on a generalized corpus rather than only on restaurant
- Adapt the generated sentences based on user's question.