Efficient Grammatical and Semantic Sentence Embeddings Framework

Huy Thanh Vu, Thu Thi Nguyen, Supreeth Narasimhaswamy

Abstraction

- Proposing a framework for sentence embedding that captures both semantic and grammatical information.
- Designing model to learn 2 tasks: relevant sentence detection (semantic task) and POS tags prediction (grammar task).
- Designed to decrease number of parameters, save computation cost.
- Can be considered unsupervised learning.
- Proposed model is an improvement from the original paper [1] – Quick-Thought model.

Introduction

- Sentence embedding is the task of representing a sentence as a numeric vector. This representations are usually used for transfer-learning tasks.
- The goal is to have a representation that captures as much information of the sentence as possible.
- Can be done in an unsupervised, or supervised manner.

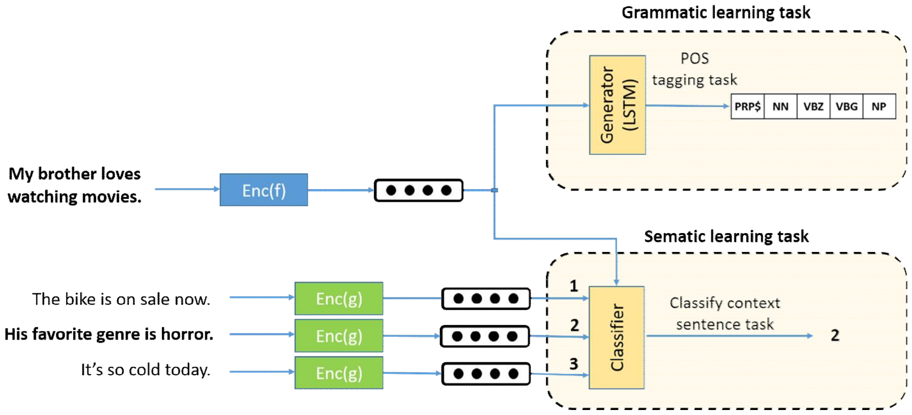
Proposed method

We design our model to solve 2 tasks:

❑ ***Semantic learning task - Classifying relevant sentences:***

Given the input sentence s and a set of candidate sentences S_{cand}. The model has to point out the relevant context sentence s_{ctxt}.

$$loss_{semantic} = -\sum_{s\in D}\sum_{s_{ctxt}\in S_{ctxt}} \log p(s_{ctxt}|s, S_{cand})$$



❑ ***Grammatical learning task – Predicting POS tags:***

The model learns to predict POS tags for each word in the sentence. The loss is the cross entropy between predictions and POS tags labels.

$$loss_{grammar} = \sum_{s\in D} cross\ entropy(target_{POS}(s), prediction_{POS}(s))$$

❑ ***Computing Total Loss:***

Total loss is the weighted sum of losses from two learning tasks. The weight α is a hyper-parameter used to regularize semantic and grammartical learning task.

$$total_{loss} = \alpha.\,loss_{semantic} + (1-\alpha).\,loss_{grammatical}$$

Experiments

We test our encoder on 4 downstream tasks:

✓ SICK: Sentence relatedness score prediction
✓ CR: Movie review sentiment classification
✓ MSRP: Parapharse identification
✓ TREC: Question classification task

Results and Discussion

	SICK			MSRP		TREC	CR
	r	ρ	MSE	Acc	F1		
Original paper [1]	0.76	0.70	0.41	0.71	0.80	0.74	0.75
Our model	0.76	0.71	0.42	0.73	0.81	0.77	0.75

- Our model performs as good as the original one on 2 tasks SICK and CR. On MSRP and TREC tasks, our model performs 2-3% better.
- This increase in accuracy maybe thanks to the ability of our encoder can capture both semantic and grammatical information of a sentence.
- *Caveat: This performance is not as high as the model in original paper [1] due to our limited of resources.*

Future Works

- Testing on downstream grammatical tasks (object number prediction, word order analysis, top constituents prediction).
- Analyzing effects of regularizor α (weights of grammar and semantic learning task) on trained encoder's quality.
- Simplifying decoder to stress more responsibility on encoder to encode more useful information.

References

[1]. Lajanugen Logeswaran, Honglak Lee, *"An efficient framework for learning sentence representations"*. In ICLR, 2018