CSE 256 Paper 1 Review

Review the following paper according the the prompts.
When you are done, submit the form. You will get a copy of your answers.

Make sure to print out the form in PDF format and submit on Gradescope.

Paper: Distributed Representations of Words and Phrases
and their Compositionality
https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

* Required

Full name *

Le Yang

UCSD student ID *

A59011332

Official UCSD email *

ley004@ucsd.edu

What are the key contributions (< 10 sentences) *

In this paper, the authors present several extensions of the original Skip-gram Model.

The authors present how to train distributed representations of words and phrases with the Skip-gram model and demonstrate that these representations exhibit linear structure that makes precise analogical reasoning possible.

They trained models on several orders of magnitude more data than previously models and get improvement in the quality of the learned word and phrase representations.

They found that subsampling of the frequent words can results in faster training and better representations of uncommon words.

They also present the Negative sampling algorithm(a simplified variant of Noise Contrastive Estimation for training Skip-gram model) which is an extremely simple training method that learns accurate representations especially for frequent words.

Discuss three(3) strong points of the paper (< 3 sentences each) *

1. Hierarchical Softmax: the authors utilize a binary Huffman tree, which assigns short codes to the frequent words which results in fast training.

2. Negative Sampling: the authors simplified Noise Contrastive Estimation (NCE) to Negative sampling (NEG). The authors did experiments for both NCE and NEG on every task and also tried to including language modeling for them.

3. Subsampling of Frequent Words: the authors used a simple subsampling approach and found that subsampling of the frequent words can results in faster training and better representations of uncommon words.

Discuss three weak point of the paper (< 3 sentences each) *

1. In the Empirical Results section, the authors did not explain why Negative Sampling(NEG) has even slightly better performance than the Noise Contrastive Estimation(NCE).

2. In the Hierarchical Softmax section, the author should show a graph and details of their binary Huffman tree structure.

3. The paper did not discuss much about the situation the the word is not in the vocabulary.

Rate the paper's originality and creativity *

	1	2	3	4	5	
Nothing new	○	○	○	◉	○	Very novel

Rate the paper's technical depth *

	1	2	3	4	5	
Not very technical	○	○	◉	○	○	Highly technical

Rate the paper's readability *

	1	2	3	4	5	
Well-written and easy to read	○	◉	○	○	○	Difficult to follow

What are the main questions you are left with after reading this paper? These can be technical questions or motivation related *

Based on the property of Skip-gram model, the authors simplified NCE to NEG. Can Negative Sampling(NEG) also be applied to other models?

Anything else you want to say about this paper

This paper is easy to read and greatly structures. And also the author smade the code for training the word and phrase vectors based on the techniques described in this paper available as an open-source project, which is very nice.

A copy of your responses will be emailed to ley004@ucsd.edu.

Submit