**Critical review of: ‘Efficient Estimation of Word Representations in Vector Space’**

Md. Mahadi Hassan, banner id – 904146206

**Analyze, Synthesize and Interpret each work (in the context) in your words in 2 sentences to understand the research problem being studied.**

Proposed two novel model architectures for computing word embedding with a huge training dataset which is computationally efficient than previous existing models. Proposed model can learn word representation in such a way that, it can capture relationship between words and similar words cluster together in the embedding space.

**Identify 3 strong and weak aspects of the paper.**

**Strong points:**

1. Introduces technique (continuous bag of wards and skip-gram) to learn high-quality word vectors from huge datasets with billions of words.
2. Their trained word embedding could perform word analogy tasks with a reasonable performance which means this embedding contains meaningful context between words. Also, mathematical operations can be done to infer related words.
3. Their proposed model has far less complexity than the existing models which make it usable and scalable model.

**Weak Points:**

1. Although this model is less computationally extensive then existing models like LDA, this model’s complexity can be reduced. This model learns its embedding be training its model on skip-gram task. In this task, while training this model needs to compute a SoftMax for each word in the vocabulary. This makes the model much more computationally expensive. This problem can be solved by converting this problem into a binary classification problem which Mikolov proposed in his following work.
2. This paper does not provide the model's description. How many hidden layers were used and how the model is constructed is not clear to the reader which reduces the reproducibility of the paper.
3. How the model is learning is not clear when we read the paper. The loss function is also not specified in this paper.

**List any 3 foundational paper that is cited in the in this. For each paper, also mention the context of the citation.**

1. **A neural probabilistic language model:** Converting the learning of vector representations of word into neural network language model.
2. **Linguistic Regularities in Continuous Space Word Representations:**  Representation of word through vectors are surprisingly good at capturing syntactic and semantic regularities in language.
3. **Distributed representations. In: Parallel distributed processing: Explorations in the microstructure of cognition:** Distributed representation of words can be successfully implemented through complex models.

**List 3 papers that cites this paper and the context.**

1. **GloVe: Global Vectors for Word Representation:**  The word analogy task is inherited from this paper.
2. **Distributed Representations of Words and Phrases and their Compositionality:** Implementation of word embedding with CBOW and Skip-Gram model is referenced as the work is an improvement over this paper. Also, the word analogy task is also inherited from this paper.
3. **Distributed Representations of Sentences and Documents:** A particular implementation of a neural network-based algorithm for training the word vectors is available in this paper.

**Identify 3 possible future direction.**

1. **Improve the current model and convert into a simpler model (binary classification) which will make this model computationally more efficient.**
2. **How this embedding will be related to the co-occurrence matrix can be proved theoretically. Like how the SVD of co-occurrence matrix can be represented with word embedding.**
3. **Can we train the embedding without using the one hot encoding? For predicting the context word, we need to do SoftMax. If we re-format the problem in some way that we can produce a scalar number from the hidden variables (Regression task), then this expensive SoftMax can be avoided.**

**List down if there is/are any dataset or tools stated in the paper**

They have used a Google News corpus for training the word vectors. The test set is available at www.fit.vutbr.cz/˜imikolov/rnnlm/word-test.v1.txt