# Critical review of “Distributed Representations of Words and Phrases and their Compositionality”

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.**

This paper presents several extensions of the continuous skip-gram model which presents an efficient way of distributed representation of ward. This extension improves both the quality of the vector and the training speed. They have introduced two concepts as an extension which are sub-sampling and negative sampling. Both techniques lead to significant speedup and more regular word representation.

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

**Strong points:**

* This paper provides a huge speedup from previous attempt of distributed representation of word through vector. This work provides a simple alternative to the computationally expensive hierarchical softmax and used negative sampling instead and used binary classification as learning problem. This provides a tremendous speed up in training the vector.
* So that the frequent word does not dominate the training process they provide the concept of subsampling which helps them to get regular word representation.
* They have also provided a way to calculate the embedding of a phrase that is not done in previous works.

**Weak Points:**

* Although they provide a way to get phrase embedding, the process they have described is ambiguous. Also, they have not generalized the way of getting a vector representation of a phrase.
* When they get a word from the unigram language model, they have raised the model's distribution to the ¾th term. They have not provided any valid reasoning behind it.
* Their unigram language model also uses the Z value. They have not provided the preferred value or any distribution of this variable.

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

1. ‘**Efficient Estimation of Word Representations in Vector Space**’: Introduction of the Skip-gram model.
2. ‘**Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics**’: Introduction of alternative to the hierarchical softmax is Noise Contrastive Estimation (NCE).
3. ‘**A neural probabilistic language model**’: Introduction of statistical language modeling.

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

1. **GloVe: Global Vectors for Word Representation**: The results of word analogy task on skip-gram model is compared to it as a baseline.
2. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding** widely applicable representations of words using neural models.
3. **Distributed Representations of Sentences and Documents:** skip-gram model of word vectors.

**Identify 3 possible future direction.**

1. This work does not provide a well generalized version of distributed representation of phrases which could be possible future direction.
2. This work provides word vectors according to the whole corpus of english language which will be appropriate in general. But sometimes a context based embedding might also be needed. So representation of word based on current context might be a good future direction.
3. How this embedding will be related to the co-occurrence matrix can be proved theoretically. Like how the SVD of the co-occurrence matrix can be represented with word embedding.

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

* An internal Google dataset with one billion words.