# **A Critical review of “Distributed Representations of Sentences and Documents”**

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.**

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

**Strong points:**

1. Introduction of vector representation of paragraph using Neural networks model where they have looked at the current context words of that paragraph.
2. Proposed model outperforms state of the art models.
3. Tried to solve the information retrieval problem using neural network based distributed vector representation of paragraph.

**Weak Points:**

1. Objective function is not well defined.
2. In the empirical analysis, they have not mentioned how the size of paragraph embedding might influence the results.
3. How they user the paragraph vector D for creation of new paragraphs is not clear. Also, how the paragraph might have some relation between each other is not clear.

**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:** learning vector representations of words and use of binary Huffman code for the hierarchy.
2. **Neural probabilistic language models:** learning vector representations of words using neural networks
3. **Distributed representations of phrases and their compositionality:** learning vector representations of words using neural networks

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

1. **BERT**: Unsupervised feature-based paragraph embedding.
2. **LINE: Large-scale Information Network Embedding:** advanced document embedding approaches.
3. **Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks: well** as distributed representations of longer bodies of text such as paragraphs and documents

**Identify 3 possible future direction.**

1. How this paragraph embedding might be used for entity tagging might be a good area to look at.
2. How to efficiently measure similarity between different paragraph is also a good area.
3. Solving a different IR problem – Rank top k document from given N document and query Q.

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

1. Stanford Sentiment Treebank Dataset
2. IMDB dataset