**NLP ASSIGNMENT 3:**

VARUN NANDU

Q 3)

In [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), Brown clustering or IBM clusteringis a form of [hierarchical clustering](https://en.wikipedia.org/wiki/Hierarchical_clustering) of words based on the contexts in which they occur, proposed by Peter Brown, Vincent Della Pietra, Peter deSouza, Jennifer Lai, and [Robert Mercer](https://en.wikipedia.org/wiki/Robert_Mercer_(businessman)) of [IBM](https://en.wikipedia.org/wiki/IBM) in the context of [language modeling](https://en.wikipedia.org/wiki/Language_model). The intuition behind the method is that a class-based language model (also called cluster *n*-gram model), i.e. one where probabilities of words are based on the classes (clusters) of previous words, is used to address the [data sparsity](https://en.wikipedia.org/w/index.php?title=Data_sparsity&action=edit&redlink=1) problem inherent in language modeling.

Word2vec can utilize either of two model architectures to produce a [distributed representation](https://en.wikipedia.org/wiki/Distributed_representation) of words: [continuous bag-of-words](https://en.wikipedia.org/wiki/Continuous_bag-of-words) (CBOW) or continuous [skip-gram](https://en.wikipedia.org/wiki/Skip-gram). In the continuous bag-of-words architecture, the model predicts the current word from a window of surrounding context words. The order of context words does not influence prediction ([bag-of-words](https://en.wikipedia.org/wiki/Bag-of-words) assumption). In the continuous skip-gram architecture, the model uses the current word to predict the surrounding window of context words. The skip-gram architecture weighs nearby context words more heavily than more distant context words. According to the authors' note, CBOW is faster while skip-gram is slower but does a better job for infrequent words.

Thus, word2Vec and brown clustering are both implemented differently. Because of the difference in implementation the outputs for both methods are different. Also running time for brown clustering exceeded 24 hours for me and the running time for word2vec was about 10-15 min.

Refrences:

1. <https://en.wikipedia.org/wiki/Word2vec>
2. <https://en.wikipedia.org/wiki/Brown_clustering>
3. Collaborators: Rohan Phadake